

# The Role of Static and Dynamic Obstacles in the Protein Search for Targets on DNA

Alexey Shvets, Maria Kochugaeva, and Anatoly B. Kolomeisky<sup>\*,†</sup>

*Department of Chemistry and Center for Theoretical Biological Physics, Rice University,  
Houston, Texas, 77005, USA*

E-mail: tolya@rice.edu

---

\*To whom correspondence should be addressed

## Abstract

Protein search for specific sequences on DNA marks the beginning of major biological processes. Experiments indicate that proteins find and recognize their targets quickly and efficiently. Because of the large number of experimental and theoretical investigations, there is a reasonable understanding of the protein search processes in purified *in vitro* systems. However, the situation is much more complex in live cells where multiple biochemical and biophysical processes can interfere with the protein search dynamics. In this study, we develop a theoretical method that explores the effect of crowding on DNA chains during the protein search. More specifically, the role of static and dynamic obstacles is investigated. The method employs a discrete-state stochastic framework that accounts for most relevant physical and chemical processes in the system. Our approach also provides an analytical description for all dynamic properties. It is found that the presence of the obstacles can significantly modify the protein search dynamics. This effect depends on the size of the obstacles, on the spatial positions of the target and the obstacles, on the nature of the search regime, as well as on the dynamic nature of the obstacles. It is argued that the crowding on DNA can accelerate or slow down the protein search dynamics depending on these factors. A comparison with existing experimental and theoretical results is presented. Theoretical results are discussed using simple physical-chemical arguments, and they are also tested with extensive Monte Carlo computer simulations.

**keywords:** protein-DNA interactions, crowding, discrete-state stochastic models, first-passage processes

# INTRODUCTION

Proteins and DNA are two main classes of biological molecules from which all living matter is made. Interactions between them control all major cellular processes involved in transfer and maintenance of genetic information, such as transcription and post-transcription modifications, translation, DNA repair and many others.<sup>1-3</sup> The starting point of these processes is a protein finding and recognizing specific target sequences on DNA that triggers the following biochemical processes. The protein search has been extensively studied using a variety of experimental and theoretical techniques.<sup>4-34</sup> Although a significant progress in clarifying search mechanisms has been achieved, many aspects of this complex biological process still remain not well understood.<sup>30,31</sup>

Experimental studies of the search process suggest that in many cases proteins associate to their targets on DNA much faster than expected from classical theories of chemical reactions.<sup>5,7,10,30,31</sup> Such surprising behavior is called a *facilitated diffusion*, and it stimulated multiple discussions on the molecular origin of this phenomenon.<sup>30,31</sup> Many experimental studies have been performed in purified *in vitro* systems in order to resolve the mechanisms of the protein search. It is now widely accepted that the facilitation is achieved because proteins search by combining motion through a bulk solution (3D mode) with hopping along the DNA chain (1D mode), and there is a fast change between these modes.<sup>20</sup> The non-specific interactions and fast inter-segment transfer rates of the protein molecule between different segments of DNA lead to effectively larger mobility for the protein molecule, accelerating the search process.<sup>20,34</sup> While these arguments probably explain reasonably well the *in vitro* experimental observations, it is not clear if they can be successfully applied for *in vivo* systems. The main reason for this is that in live cells there are many other processes taking place in parallel, and this might influence the search dynamics.<sup>30</sup> For example, due to macromolecular crowding some parts of the DNA chain are heavily covered by other proteins, preventing the sliding to the target sequence. These covering proteins serve as obstacles or road-blocks in the search for specific sites on DNA. In addition, the searching protein can be trapped by associating to other biological macromolecules in the bulk solution.

Although the presence of obstacles on DNA in live cells seems to be an important factor for the

protein search dynamics, most of theoretical investigations ignore this effect and there are only few works that addressed this issue.<sup>8,9,24,28,30,35</sup> However, the predictions from these theoretical studies are rather controversial. It was argued using an approximate theory and computer simulations that the presence of immovable obstacles always leads to larger search times.<sup>8,24,28,30</sup> At the same time, other computational studies indicated that there are conditions when the obstacle can lead to the faster search for specific targets.<sup>9,35</sup> But the molecular nature of these observations and the origin of these discrepancies have not been explained. In addition, only static obstacles have been considered so far, with the exception of the computational study of Marcovitz and Levy.<sup>9</sup> A more realistic situation in the cells is when these roadblocks can dissociate from DNA, and this might strongly affect the search.

In this paper, we develop a comprehensive theoretical approach that analyzes the effect of obstacles in the protein search dynamics. Using a discrete-state stochastic framework that accounts for most important chemical and physical processes, a fully analytical dynamic description of the protein search in the presence of static obstacles is obtained for all ranges of parameters. The analysis is extended to dynamic obstacles where approximate theoretical arguments along with extensive Monte Carlo computer simulations are utilized for describing the protein search process. By providing a microscopic picture for these complex phenomena, our analysis explains the previous controversial results by clarifying under what conditions the obstacles can facilitate or slow down the search dynamics. A comparison with available experimental observations and theoretical models is also given.

## **THEORETICAL METHODS**

We consider a simple stochastic model, presented in Fig. 1, where a single protein molecule is searching for a specific target site on a single DNA chain that consists of  $L$  binding sites. The target site is at the site  $m$  ( $1 \leq m \leq L$ ). This chain also contains one roadblock, which occupies  $\Delta$  sites on DNA with the left boundary at the site  $l_{ob}$ : see Fig. 1. This obstacle prevents the protein

from sliding to the target if the protein is bound to DNA anywhere between the sites  $l_{ob} + \Delta$  and  $L$  (Fig. 1). Our theoretical method can be extended to the case of multiple obstacles at different locations, but for simplicity we consider only a single roadblock. In addition, it is assumed for now that the obstacle is static, i.e., it can never dissociate from the DNA chain. Later, we will lift this restriction when the mobile obstacles will be considered.

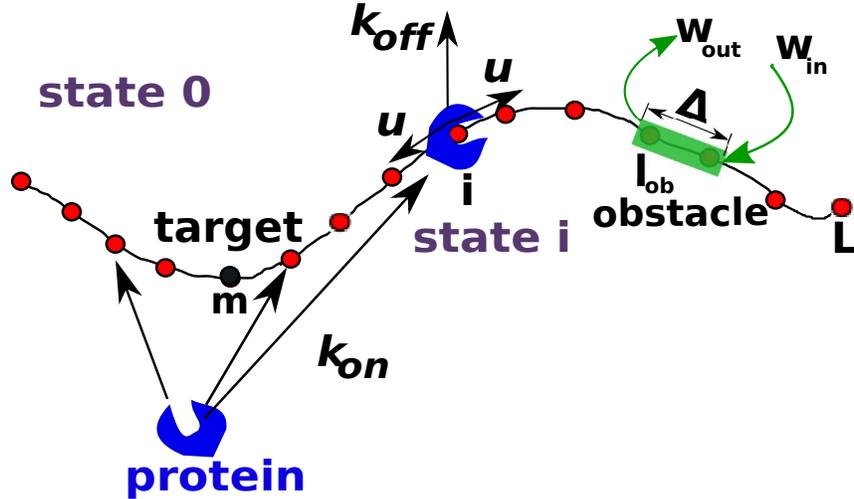


Figure 1: A general scheme for the protein target search on DNA with an obstacle. There are  $(L - 1 - \Delta)$  nonspecific sites and 1 specific site on the DNA. The target is at the site  $m$ , and the obstacle is occupying sites between  $l_{ob}$  and  $l_{ob} + \Delta - 1$ . A protein molecule can slide along the DNA chain with the 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 equal probability and the total rate  $k_{on}$ . The obstacle can dissociate from DNA with the rate  $w_{out}$  and it can return to bind to the same position with the rate  $w_{in}$ .

The protein always starts the search process from the solution that we label as a state 0. It is also assumed that the DNA chain is coiled in the solution, and the searching protein diffuses very fast in the volume around DNA. Then the protein can reach all parts of DNA with equal probability. The protein molecule can bind to any vacant site on DNA with a rate  $k_{on}$  per each site (see Fig. 1). The DNA-bound protein can diffuse along the chain with a diffusion rate  $u$  with equal probability in both directions if the motion is not blocked by the obstacle. Finally, the protein molecule can dissociate from DNA with a rate  $k_{off}$ , as shown in Fig. 1. It has been argued before that the protein search for the specific sites can be associated with first-passage processes.<sup>19,20</sup> Then we can introduce a function  $F_n(t)$ , which is defined as a probability to reach the target for the first

time, if at  $t = 0$  the protein was at the state  $n$  (where  $n = 1, 2, \dots, L$  are sites on DNA and  $n = 0$  corresponds to the bulk solution). The temporal evolution of these probabilities can be described utilizing the backward master equations,<sup>19,20</sup>

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

for  $2 \leq n \leq L - 1$  and  $n \neq m$ , or for  $n \neq l_{ob}, \dots, l_{ob} + \Delta - 1$ . At the DNA ends and around the obstacle the dynamics is slightly different,

$$\frac{dF_a(t)}{dt} = uF_{a+1}(t) + k_{off}F_0(t) - (u + k_{off})F_a(t); \quad (2)$$

where  $a = 1$  corresponds to the first site of the chain and  $a = l_{ob} + \Delta$  is the first site after the obstacle. Furthermore,

$$\frac{dF_b(t)}{dt} = uF_{b-1}(t) + k_{off}F_0(t) - (u + k_{off})F_b(t). \quad (3)$$

where  $b = L$  corresponds to the last site on DNA, while  $b = l_{ob} - 1$  is the last vacant site before the obstacle. In addition, for the bulk solution ( $n = 0$ ) we have

$$\frac{dF_0(t)}{dt} = k_{on} \left( \sum_{n=1}^m F_n(t) + \sum_{n=m+1}^{l_{ob}-1} F_n(t) + \sum_{n=l_{ob}+\Delta}^L F_n(t) \right) - k_{on}(L - \Delta)F_0(t). \quad (4)$$

At the same time, there are additional constraints in the system. If the protein molecule starts at  $t = 0$  at the target site  $m$ , the search process is immediately finished. This condition can be written as

$$F_m(t = 0) = \delta(t) \quad (5)$$

To obtain a full dynamic description of the first-passage events in this system, it is convenient to apply Laplace transformations, i.e.,  $\tilde{F}_n(s) = \int_0^\infty e^{-st} F_n(t) dt$ .<sup>20</sup> Then the set of backward master

equations can be transformed into simpler algebraic expressions,

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

for  $2 \leq n \leq L-1$ , excluding the obstacle sites and the target. Meanwhile, for the boundary sites we have

$$(s + u + k_{off})\tilde{F}_a(s) = u\tilde{F}_{a+1}(s) + k_{off}\tilde{F}_0(s), \quad (7)$$

for  $a = 1$  or  $l_{ob} + \Delta$ ; and

$$(s + u + k_{off})\tilde{F}_b(s) = u\tilde{F}_{b-1}(s) + k_{off}\tilde{F}_0(s), \quad (8)$$

for  $b = l_{ob} - 1$  or  $L$ . For the bulk solution it can be written as,

$$[s + k_{on}(L - \Delta)]\tilde{F}_0(s) = k_{on} \left( \sum_{n=1}^m \tilde{F}_n(s) + \sum_{n=m+1}^{l_{ob}-1} \tilde{F}_n(s) + \sum_{n=l_{ob}+\Delta}^L \tilde{F}_n(s) \right), \quad (9)$$

while the boundary condition leads to

$$\tilde{F}_m(s) = 1. \quad (10)$$

The solutions of these equations can be found by assuming a general form of the solution as  $\tilde{F}_n(s) = Ay^n + B$ .<sup>20</sup> This yields,

$$\tilde{F}_0(s) = \frac{k_{on}(k_{off} + s)S(s)}{s [s + k_{off} + k_{on}(L - \Delta)] + k_{on}k_{off}S(s)}, \quad (11)$$

with

$$S(s) = \frac{y(y^{-m} - y^m)}{(1-y)(y^m + y^{1-m})} + \frac{y^{m+1} - y^{2l_{ob}+m+1}}{(1-y)(y^m + y^{m+2l_{ob}-1})}, \quad (12)$$

and

$$y(s) = \frac{s + 2u + k_{off} - \sqrt{(s + 2u + k_{off})^2 - 4u^2}}{2u}. \quad (13)$$

A mean first-passage time to reach the target starting from the solution we identify as the average

search time. Then it is given by

$$T_0 = -\left. \frac{\partial \tilde{F}_0(s)}{\partial s} \right|_{s=0}. \quad (14)$$

The explicit expression for the search time  $T_0^{(ob)}$  in the presence of the static obstacle can be written as,

$$T_0^{(ob)} = \frac{k_{off} + k_{on} \{(L - \Delta) - S(0)\}}{k_{on} k_{off} S(0)}. \quad (15)$$

In the case of no obstacles on the DNA chain ( $\Delta = 0$ ), this results reduces to the average protein search time for the target on unobstructed DNA as was obtained earlier.<sup>20</sup> All other dynamic properties for the system can be easily obtained from explicit expressions for the first-passage probabilities.<sup>20,33</sup>

Unfortunately, this exactly-solvable discrete-state stochastic method cannot be easily extended to the protein search with dynamic obstacles that can bind and unbind from DNA many times. However, in the next section we present approximate theoretical arguments, supported by Monte Carlo computer simulations, that allow us to obtain a good understanding of the search dynamics in this case.

## RESULTS AND DISCUSSION

### Static Obstacle

The first question we would like to address is how the position and the size of the obstacle influence the protein search dynamics. The average search times for the static obstacle, as well as for the case without roadblocks, are presented in Fig. 2. One can see that three different search regimes can be identified depending on the range of parameters, and this qualitative behavior is independent of the presence of obstacles. This mostly reflects the fact that there are three major length scales in the system: the target length, which we assume to be equal to one, the average distance  $\lambda$  that the protein slides along the DNA chain in each encounter, and the total DNA length available for the search and not covered by the obstacle, which is equal to  $(L - \Delta)$ .<sup>20</sup>

The first regime corresponds to the case when the sliding length is less than the target size ( $\lambda < 1$ ). In this case, the protein can only bind/unbind to the available sites on DNA, but no sliding can take place. Because the sliding length can be defined as  $\lambda \simeq \sqrt{u/k_{off}}$ ,<sup>20</sup> this corresponds to the situation of very slow diffusion on DNA or very fast dissociations. Obviously, the position of the target and obstacle, and the size of the roadblock is not important in this regime since each site on DNA can be reached independently only through 3D search via the bulk solution (see Figs. 1 and 2). This dynamic phase is called a jumping regime.<sup>20</sup>

In the second regime, for  $1 \leq \lambda \leq (L - \Delta)$ , the protein in addition to 3D associations/dissociations can also slide along the DNA chain on the way to the target. We can label this dynamic phase as a sliding regime.<sup>20</sup> It can be viewed as a combined 1D+3D search process, and the proteins can come to the target from the solution (3D mode) or from DNA (1D mode). As a result, the overall search time is smaller than for the jumping phase because of the increased flux of the protein molecules to the target due to motion along DNA. However, the presence of the obstacles and their sizes become important for the search in this phase. To find the target the protein must, on average, scan the whole DNA length, not covered by the obstacles,  $(L - \Delta)$  sites. Increasing the size of the roadblock reduces the available number of sites on DNA that must be checked, and this accelerates the search: see Fig. 2a. However, the exact location of the obstacle with respect to the target influences the search dynamics, as shown in Fig 2b.

More interesting behavior is observed in the third search regime for  $\lambda > (L - \Delta)$ , which is called a random-walk dynamic phase.<sup>20</sup> Qualitatively different behavior is observed with and without the roadblocks on DNA. When there is no obstacles on DNA, the protein molecule in this regime binds to DNA and it never unbinds until the target is found. The protein is performing a 1D random-walk motion in this case. For this reason, the search time is independent of the scanning length  $\lambda$ , or, to be more precise, independent of the dissociation rate  $k_{off}$  since we keep the diffusion rate  $u$  constant for calculations in Fig. 2. Only the total free length of DNA,  $L$ , is important here. Introducing the obstacle in the system dramatically changes the search dynamics. Increasing the sliding length  $\lambda$  (lowering the dissociation rate  $k_{off}$ ) slows down the overall search time. But this

effects also depends on the position and on the size of the obstacle (Fig. 2).

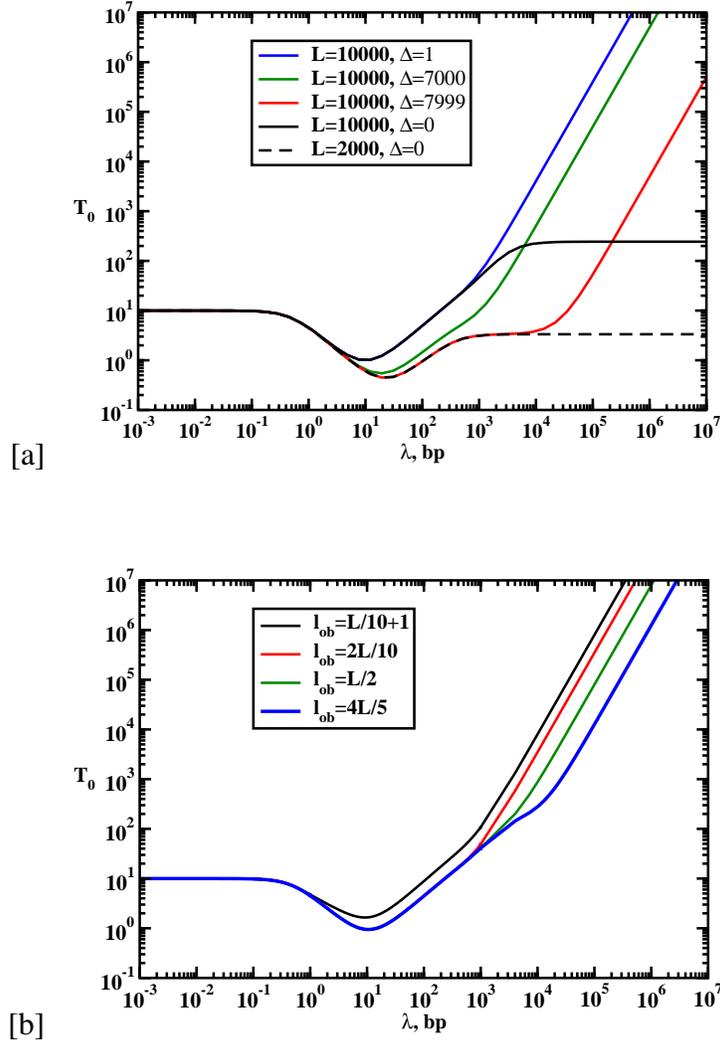


Figure 2: Average search times to find the target located at  $m = L/10$  as a function of the scanning length  $\lambda = \sqrt{u/k_{off}}$ . The transition rates utilized for calculations are:  $u = 10^5 \text{ s}^{-1}$  and  $k_{on} = 0.1 \text{ s}^{-1}$  per one DNA site. a) The length of the DNA chain is  $L = 10^4$  for solid curves and  $L = 2000$  for the dashed curve. The obstacle starts its position from  $l_{ob} = L/5$  and its size,  $\Delta$ , is varied as shown on the picture. b) The length of the DNA chain is  $L = 10^4$  and the size of the obstacle is  $\Delta = 1000$ . The position of the obstacle is varied as shown in the picture.

By analyzing the results in Fig. 2, we can also clearly determine what conditions accelerate or slow down the protein search after introducing the obstacle on DNA. Three different dynamic behaviors can be identified. In the jumping regime ( $\lambda < 1$ ) the presence of the roadblocks effectively has no effect on the search dynamics because of purely 3D search mechanism. For this set

of parameters ( $k_{off} \gg k_{on}$ ) the rate-limiting step is just to go to the target site, and the search time is  $T_0^{(ob)} \simeq 1/k_{on}$  [see Eq.(15)]. In the sliding regime, [ $1 < \lambda < (L - \Delta)$ ], finding the roadblocks on DNA lowers the search times for the fixed distance between the target and the roadblock because the protein molecule should scan smaller DNA segments during the 1D searching mode as compared to the case without obstacles (Fig. 2a). However, decreasing the distance between the target and the obstacle makes the search slower (see Fig. 2b). It also widens the sliding search regime. This happens due to the decrease of the 1D protein flux from the DNA side where the roadblock is positioned. Putting the obstacle next to the target completely shuts off this channel.

But dynamics changes dramatically in the random-walk regime, [ $\lambda > (L - \Delta)$ ], where a significant fraction of the searching trajectories can be blocked by the obstacles. This happens due to strong nonspecific interactions between the protein and DNA that keep the protein molecule bound to the DNA chain for long periods of time. Thus, our theoretical analysis provides a microscopic explanation of the controversy on the role of static obstacles in the protein search. This effect is determined by the specific sets of parameters that favor the specific dynamic behavior. The obstacles might make the search faster or slower, or they might even lead to no changes at all.

Although we have a fully analytical description for the protein search on DNA with obstacles, to understand better the molecular picture of these processes it is also convenient to consider a different approach. The search time is a mean over all possible trajectories starting from the solution. In the random-walk regime ( $\lambda > L - \Delta$ ) the largest contribution to the search time will come from trajectories that lead the protein molecule to the area on DNA between the obstacle and the DNA end (see Fig. 1). The protein cannot reach the target from these sites via sliding, and it has to redundantly visit them many times until it can dissociate back into the solution. We can define a probability  $q$  of coming to the blocked segment of DNA, and it can be found from the geometric arguments (see Fig. 1),

$$q = \frac{L - l_{ob} - \Delta + 1}{L - \Delta}. \quad (16)$$

The probability to return again to the same blocked segment for the second time is equal  $q^2$ , and, similarly, to visit the same segment after  $n - 1$  search cycles is  $q^n$ . The average time to be on

DNA during one search cycle is  $1/k_{off}$ . Then the contribution from visiting the blocked segment dominates the overall search time, and it can be found as

$$t_{ob} = \frac{1}{k_{off}}(q + q^2 + q^3 + \dots) = \frac{1}{k_{off}} \frac{q}{(1-q)}. \quad (17)$$

As for other search regimes, when  $\lambda < L$ , the contribution of sliding is minimal. Here we can view the search for the target on DNA of length  $L$  with the roadblock of size  $\Delta$  as the search on the DNA chain of the length  $L - \Delta$  but without obstacles, for which exact results are already known,<sup>20</sup>

$$T_0^{(0)}(L - \Delta) = \frac{k_{off} + k_{on} \left[ (L - \Delta) - S_0^{(0)}(L - \Delta) \right]}{k_{on} k_{off} S_0^{(0)}(L - \Delta)}, \quad (18)$$

where  $T_0^{(0)}$  is the search time for the system without roadblocks, and an auxiliary function  $S_0^{(0)}$  is given by

$$S_0^{(0)}(L - \Delta) = \frac{y(1+y)(y^{-(L-\Delta)} - y^{(L-\Delta)})}{(1-y)(y^m + y^{-m})(y^{(L-\Delta)-m+1} + y^{m-(L-\Delta)})}, \quad (19)$$

with  $y(s)$  from Eq. (13). Therefore, the total search time in all dynamic regimes can be approximated as

$$T_0^{ob}(L) \simeq T_0^{(0)}(L - \Delta) + t_{ob}. \quad (20)$$

The comparison between exact calculations and the new method is presented in Fig. 3. One can clearly see that our approximation works perfectly everywhere with only tiny deviations in the sliding regime for small-size roadblocks. But, most importantly, it allows us to understand more clearly the protein search processes on DNA with obstacles. This method suggests that in the jumping and sliding regimes [for  $\lambda < (L - \Delta)$ ] the search on DNA of length  $L$  with the obstacle of size  $\Delta$  can be effectively viewed as a search on pure DNA of length  $L - \Delta$  without any roadblocks. For the random-walk regime [ $\lambda > (L - \Delta)$ ] the search times are dominated by multiple visits to the blocked DNA segments, i.e., to the segment between the roadblock and the DNA end from which the protein cannot directly slide to the target. The stronger the interaction between the protein and

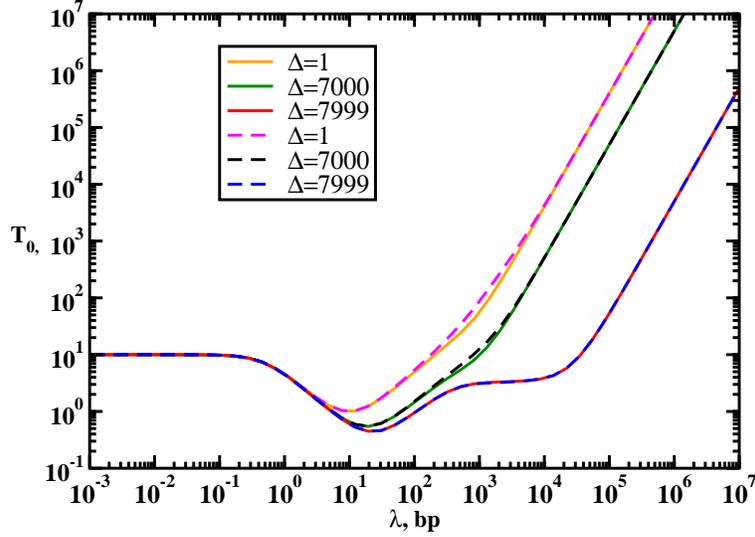


Figure 3: Comparison between exact (solid curves) and approximate solutions (dashed curves) for the search times to find the target as a function of the scanning length  $\lambda = \sqrt{u/k_{off}}$ . The length of the DNA chain is  $L = 10^4$  and the target is placed at  $m = L/10$ . The obstacle starts its position from  $l_{ob} = L/5$  and its size  $\Delta$  is varied as shown on the picture. The transition rates utilized for calculations are:  $u = 10^5 \text{ s}^{-1}$  and  $k_{on} = 0.1 \text{ s}^{-1}$ .

DNA, which effectively means very small dissociation rates  $k_{off}$ , the longer the search times.

## Dynamic Obstacle

In the next step, we investigate a more realistic situation with dynamic obstacles in the protein search for specific targets. Now we assume that the obstacle can bind to DNA with a rate  $w_{in}$  and dissociate with a rate  $w_{out}$  — see Fig. 1. As before, the roadblock can be found in one specific location on DNA, starting from the site  $l_{ob}$ , and it is not covering the target site  $m$ . For simplicity, we consider here the obstacle of size  $\Delta = 1$ , although our results can be easily extended to dynamic roadblocks of any size. As we mentioned above, it is difficult to obtain exact analytical solutions for the corresponding discrete-state model. Instead, we utilize Monte Carlo computer simulations and approximate theoretical arguments to analyze the search dynamics in this case.

We start with the jumping search regime where  $\lambda < 1$ . In this phase, the protein finds the target only via binding and unbinding events through the solution, and there is no sliding along the

DNA chain. This means that the presence of the obstacle as well as its mobility do not influence the search process at all: see Fig. 4. Similar dynamics is also observed in the sliding regime, for  $1 < \lambda < L - \Delta$ . In this phase, the search mechanism can be viewed as a combination of 3D and 1D motions. Because the obstacle is so small in size in comparison with the total DNA length ( $\Delta \ll L$ ), it does not lead to significant decrease in the search time for smaller number of the DNA free sites,  $L - \Delta \sim L$ . And because of the frequent dissociations, the protein molecule losses memory on where the obstacle is sitting. However, the dynamic nature of the roadblock can modify the range of this search regime. One can see from Fig. 4 that varying the rates  $w_{out}$  and  $w_{in}$  can shrink or widen this dynamic phase. For example, decreasing the dissociation rate of the obstacle,  $w_{out}$ , increases the rang

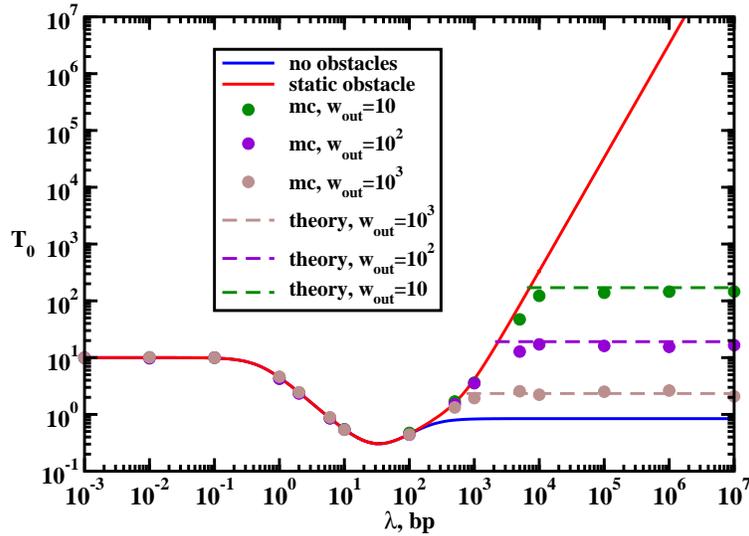


Figure 4: A dynamic phase diagram for the protein search on DNA with the dynamic obstacle. The DNA chain has the length  $L = 10^3$  bp with target at the position  $m = L/2$ , and the obstacle is at  $l_{ob} = 3L/4$ . Parameters used for calculations are:  $k_{on} = 0.1 \text{ s}^{-1}$ ,  $u = 10^5 \text{ s}^{-1}$ ,  $w_{in} = 10^6 \text{ s}^{-1}$  and different  $w_{out}$  (in units of  $\text{s}^{-1}$ ) as shown in the picture. Symbols correspond to Monte Carlo simulations whereas the dashed lines describe the approximate theory (see the text for the explanations). Solid curves correspond to exact results for DNA without obstacles and for DNA with one static obstacle.

The most interesting dynamics is observed in the random-walk regime where  $\lambda > L$ . When there is no roadblock, the search is taking place only via 1D sliding along the DNA, and the search time is independent of the sliding length (Fig. 4). If the obstacle is always present on DNA,

as we discussed above, the search time is dominated by trajectories where the protein is in the blocked segment (between the roadblock and the DNA end) without a direct access to the target and with rare dissociations into the solution. Here, the search time increases with the scanning length  $\lambda$  because it corresponds to longer times in the blocked segment. The protein search with the dynamic obstacle shows a behavior that is intermediate between these two limiting cases (see Fig. 4). The longer the obstacle sits on DNA, which is given by the time  $t_{out} = 1/w_{out}$ , the longer the search time. However, the dynamics is independent of the scanning length  $\lambda$ , similarly to the case of the search on homogeneous DNA without obstacles.

To explain this dynamic behavior we can use the following arguments. It is clear that the protein molecule performs normal search when the obstacle is dissociated from DNA. Then the overall search time can be written as

$$T_0^{(dyn)}(\lambda > L) = T_0^{(0)}(\lambda > L) + T_1, \quad (21)$$

where  $T_0^{(0)}$  is the search time on DNA without obstacles and  $T_1$  is a contribution due to the obstacle blocking the search. During each time  $t_{in} = 1/w_{in}$ , on average, the obstacle is not on DNA and the protein molecule scans a characteristic distance  $\lambda_w \simeq \sqrt{ut_{in}} = \sqrt{u/w_{in}}$ . When the roadblock binds back to DNA and sits there for the time  $t_{out}$ , the normal search cannot take place if the protein is in the blocked segment. Thus, the biggest contribution to  $T_1$  is due to the trajectories that go through the blocked segment — they are the slowest. But eventually the protein will manage to leave the blocked segment, and the average distance it moves is  $AL$ , where a coefficient  $A < 1$  reflects the mean distance to escape from the blocked segment and the fraction of proteins that will come to the blocked segment. Then the normal search will proceed as before. This suggests that it will take  $AL/\lambda_w$  obstacle binding/unbinding cycles to move out of the blocked segment, and the contribution due to obstacles can be written as

$$T_1 \simeq \frac{1}{w_{out}} \frac{AL}{\lambda_w} = \frac{1}{w_{out}} \frac{AL}{\sqrt{\frac{u}{w_{in}}}}. \quad (22)$$

Here  $1/w_{out}$  is the time that the roadblock is on DNA and the overall search is delayed. This delay is taking place until the protein moves out of the blocked segment. The predictions from this approximate theory in comparison with the computer simulations results are plotted in Figs. 4 and 5 for suitably chosen parameters  $A \simeq 0.1$ . Very good agreement suggests that our approximate theory correctly captures main features of the search dynamics in this regime.

It is important to note that the presented theoretical arguments are reasonable if the protein can scan more than one site during the time when the roadblock is not on DNA, or this mechanism will not work. This implies that our theory is valid only for  $\lambda_w > 1$ . Otherwise, the search mechanism with static obstacles as described above will be realized. These predictions are confirmed as shown in Fig. 5. When the characteristic length is not small,  $\lambda_w > 1$ , our approximate theory describes the search dynamics quite well, while for  $\lambda_w < 1$  the search time goes to plateau and a different dynamic behavior

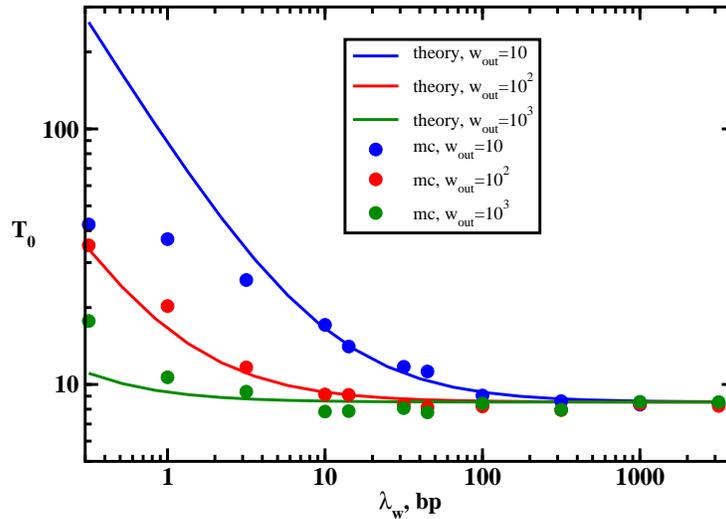


Figure 5: Average times to reach the target as a function of the characteristic length  $\lambda_w = \sqrt{u/w_{in}}$ . Parameters used for calculations are:  $m = L/2$ ,  $l_{ob} = 3L/4$ ,  $k_{on} = 0.1 \text{ s}^{-1}$ ,  $u = 10^5 \text{ s}^{-1}$ , the DNA chain length  $L = 10^4 \text{ bp}$ ,  $k_{off} = 10^{-3} \text{ s}^{-1}$  and variable rates  $w_{out}$  (in units of  $\text{s}^{-1}$ ) as indicated in the plot. Symbols correspond to Monte Carlo computer simulations, and the solid curves are theoretical predictions.

## Comparison with Existing Theories and Experiments

The effect of the static obstacles has been experimentally investigated recently for the facilitated diffusion of *lac* repressor proteins in the living cells.<sup>24</sup> By placing a roadblock protein particle next to the target it was found that the association rate to the specific sequence was reduced by a factor  $1.75 \pm 0.18$ .<sup>24</sup> Using our exact results from Eq. (15) and the parameters  $u = 7 \times 10^5 \text{ s}^{-1}$ ,  $k_{on} = 6.4 \times 10^4 \text{ s}^{-1}$  (per total DNA molecule),  $\lambda \simeq 25 \text{ bp}$  and  $\Delta \simeq 1 \text{ bp}$ , which are consistent with experimental observations,<sup>24,33</sup> we estimate that the presence of the obstacle should slow down the search in 1.80 times. This perfectly agrees with the experimental results.

Hammar et al.<sup>24</sup> have also proposed a continuum theoretical approach to account for the effect of the static obstacles in the protein search. More specifically, they found that the ratio of the association rates in the presence and absence of the single obstacle at the distance  $l$  from the target is given by<sup>24</sup>

$$r = \frac{T_0}{T_0^{(ob)}} = \frac{(\lambda + 1)^2 [1 + \lambda (1 + \tanh(l/\lambda))]}{(1 + 2\lambda) [\lambda^2 + 1 + \lambda (1 + \tanh(l/\lambda))]} \quad (23)$$

A comparison of predictions from our discrete-state stochastic model and the continuum theory of Hammar et al.<sup>24</sup> are presented in Fig. 6. One can see that small scanning length  $\lambda < 1$  both theoretical approaches agree, and the ratio of association rates is equal to one. This corresponds to the jumping regime where the search is taking place only 3D associations and dissociations. Because the obstacle is small for this system,  $\Delta \ll L$ , its presence does not modify the search times. Similar agreement is found for intermediate values of  $\lambda$  where the search is taking place in the sliding regime. Here the presence of the obstacle slows down the search by blocking the flux to the target from one side of DNA. However, for very large values of the scanning length the results from both theories start to deviate. For  $\lambda \rightarrow \infty$  the continuum model predicts  $r = 1/2$ , while our theory suggests that  $r \rightarrow 0$ . The result from the continuum theory seems to be unrealistic because increasing the scanning length corresponds to larger nonspecific interactions with DNA. This means that the protein molecule will spend more time in the blocked segment, and this should strongly decrease the association rate to the target. Our theoretical results are fully consistent with

these physical arguments. But one should also note that the continuum theory was developed under the assumption of the infinite DNA length.

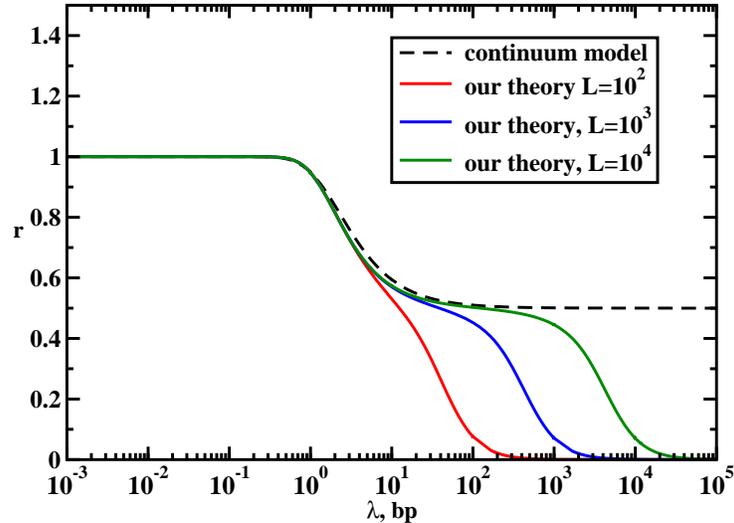


Figure 6: Ratio of association rates to the target in the presence and in the absence of the obstacle. The parameters used for calculations in the discrete-state stochastic model are:  $m = L/2$ ,  $l_{ob} = L/2 + 1$ ,  $u = 7 \times 10^5 \text{ s}^{-1}$ ,  $k_{on} = 6.4 \times 10^4 \text{ s}^{-1}$  (per total DNA molecule), and  $\Delta \simeq 1 \text{ bp}$ . Our theoretical predictions are shown by solid curves. Eq. (23) is employed for continuum results, which are shown by dashed lines.

## SUMMARY AND CONCLUSIONS

We developed a theoretical approach to analyze the role of roadblocks in the protein search for specific binding sites on DNA. Two different scenarios were considered. First, the protein search in the presence of immovable obstacles was fully analyzed using exactly solvable discrete-state stochastic model. We observed three different dynamic search phases, which are determined by the balance between the target size, the scanning length, the DNA length and the size of the roadblock. The effect of the static obstacles was different in these search regimes. In the jumping regime where the search is taking place only via 3D binding/unbinding events, effectively the obstacle does not modify the search dynamics. In the sliding regime, where the search is a combination of 3D and

1D modes, the search is faster in the presence of obstacles because of the smaller length of free DNA segments for the fixed distance between the target and the obstacle. But varying this distance might also increase the search times by making harder for proteins to slide to the target from the DNA side with the obstacle. However, the roadblock significantly decelerates the protein search in the random-walk regime where mostly 1D sliding is observed. Here the obstacle effectively blocks the approach to the target, and it takes the protein molecule many attempts to escape from the blocked segment on DNA. This analysis provides a full explanation of previous controversial theoretical results. It is argued that the previous different theoretical predictions correspond to different dynamic search regimes. We also discussed a new theoretical method that allowed us to clarify the molecular picture of the search in different dynamic phases.

Next, we developed a theoretical model for analyzing the protein search in the presence of dynamic obstacles, which supposed to be a better description of real processes in living cells. Again, three search regimes with different dynamics were identified. It was argued that the search is faster in the systems with mobile roadblocks in comparison with the static obstacles. But this effect can be observed only in the random-walk regime, where the scanning length is larger than the length of free DNA. By developing an approximate theoretical picture, we also found that in this regime the search dynamics is different for the systems with dynamic obstacles in comparison with the systems with the static roadblocks. The main idea of our approach is that the search is taking place when the obstacle is not on DNA. Extensive Monte Carlo computer simulations fully support our theoretical predictions.

In addition, our approach was compared with available experimental results. It is found that our model with the static roadblock next to the target site exactly agrees with experimental observations that show lowering the corresponding association rate. We also compared our discrete-state stochastic model with predictions from the existing continuum theory. The differences between two approaches are discussed using physical-chemical arguments. It is found that our theory is fully consistent with the fundamental views of the protein search phenomena, while the continuum approach becomes unphysical in the limit of large scanning lengths.

Although, the presented theoretical method seems to be capturing the most relevant features of the protein search in the living cells, one should note that our approach is still rather oversimplified. Many important phenomena during the protein-DNA interactions are not taken into account. They include the moving of obstacles on DNA, covering the target sequence, the intermittent interactions between the searching proteins and crowding agents in the solution, different protein and DNA conformations, the DNA chain mobility and many others. It will be critically important to test the validity of the presented method in more advanced theoretical approaches, as well as in the experimental studies.

## 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).

## 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) Lodish, H et al. *Molecular Cell Biology* 6th ed., W.H. Freeman: New York, **2007**.
- (3) Phillips R.; Kondev J.; Theriot J. *Physical Biology of the Cell*, 2nd ed., Garland Science: New York, **2012**.
- (4) Riggs, A.D.; Bourgeois, S.; Cohn, M. The *lac* repressor-operator interaction: III. Kinetic studies. *J. Mol. Biol.* **1970**, *53*, 401-417.
- (5) Berg, O.G.; Winter, R.B.; von Hippel, P.H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* **2009**, *20*, 6929-6948.

- (6) Berg, O.G.; von Hippel, P.H. Diffusion-controlled macromolecular interactions. *Ann. Rev. Biophys. Biophys. Chem.* **1985**, *14*, 131-160.
- (7) 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.
- (8) Li, G.W. ; Berg, O.G.; Elf, J. Effects of macromolecular crowding and DNA looping on gene regulation kinetics. *Nat. Phys.* **2009**, *5*, 294 - 297.
- (9) Marcovitz, A.; Levy, Y. Obstacles may facilitate and direct DNA search by proteins. *Biophys. J.* **2013**, *104*, 2042-2050.
- (10) Halford, S. E.; Marko, J. F. How do site-specific DNA-binding proteins find their targets? *Nucl. Acid Res.* **2004**, *32*, 3040-3052.
- (11) Gowers, D. M.; Wilson, G.G.; Halford, S.E. Measurement of the contributions of 1D and 3D pathways to the translocation of a protein a long DNA. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15883-158. 88.
- (12) Kolesov, G.; Wunderlich, Z.; Laikova, O.N.; Gelfand, M.S.; Mirny, L.A. How gene order is influenced by the biophysics of transcription regulation. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 13948-13953.
- (13) 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.
- (14) 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. (2007).
- (15) 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.

- (16) 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.
- (17) Hu, T.; Grosberg A. Yu.; Shklovskii B. I. How Proteins Search for their Specific Sites on DNA: the Role of DNA Conformation. *Biophys. J.* **2006**, *90*, 2731-2744.
- (18) Givaty, O.; Levy, Y. Protein Sliding along DNA: Dynamics and Structural Characterization. *J. Mol. Biol.* **2009**, *385*, 1087-1097.
- (19) Kolomeisky, A.B.; Veksler, A. How to accelerate protein search on DNA: Location and dissociation. *J. Chem. Phys.* **2012**, *136*, 125101.
- (20) 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.
- (21) Mechetin, G.V.; Zharkov, D.O. Mechanisms of Diffusional Search for Specific Targets by DNA-Dependent Proteins. *Biochemistry (Moscow)* **2014**, *79*, 633-644.
- (22) Hilario, J.; Kowalczykowski, S. C. Visualizing protein-DNA interactions at the single-molecule level. *Curr. Opin. Chem. Biol.* **2009**, *14*, 15-22.
- (23) 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. USA* **2014**, *111*, 17140-17145.
- (24) 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.
- (25) 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.

- (26) 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. USA* **2013**, *110*, 19796-19801.
- (27) Brackley, C. A.; Cates, M. E.; Marenduzzo, D. Facilitated Diffusion on Mobile DNA: Configurational Traps and Sequence Heterogeneity. *Phys. Rev. Lett.* **2012**, *109*, 168103.
- (28) Brackley, C.A.; Cates, M. E.; Marenduzzo, D. Intracellular facilitated diffusion: searchers, crowders and blockers. *Phys. Rev. Lett.* **2013**, *111*, 108111.
- (29) Gorman, J.; Greene, E. C. Visualization one dimensional diffusion of proteins along DNA. *Nat. Struct. Mol. Biol.* **2008**, *15*, 768-774.
- (30) 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.
- (31) Kolomeisky, A.B. Physics of Protein-DNA Interactions: Mechanisms of Facilitated Target Search. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2088-2095.
- (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) Lange, M.; Kochugaeva, M.; Kolomeisky, A.B. Protein Search for Multiple Targets on DNA. *J. Chem. Phys.* **2015**, *143*, 105102.
- (34) 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: Origin of the Optimal Search at Physiological Ionic Strength. *Nucl. Acid Res.* **2014**, *42*, 7039.
- (35) Lu, L.; Kaifu, L. DNA-binding protein searches for its target: Non-monotonic dependence of the search time on the density of roadblocks bound on the DNA chain. *J. Chem. Phys.* **2015**, *142*, 125101.

(36) Redner, S. *A Guide to First-Passage Processes*, 1st ed., Cambridge University Press, **2007**.

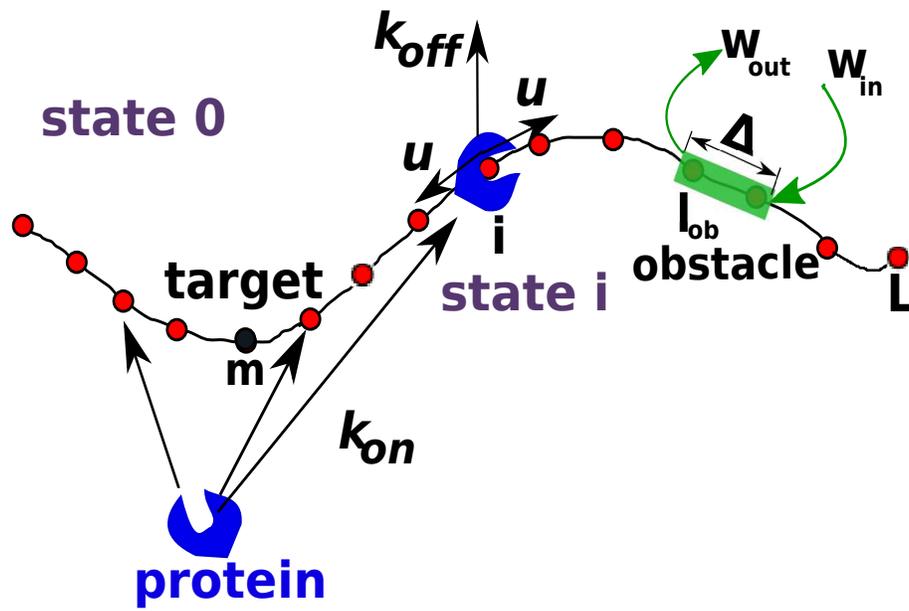


Figure 0: TOC