Optimal Length of Conformational Transition Region in Protein Search for Targets on DNA

Maria P. Kochugaeva,[†] Alexander M. Berezhkovskii,[‡] and Anatoly B. Kolomeisky*,[†]

[†]Department of Chemistry and Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States
[‡]Mathematical and Statistical Computing Laboratory, Division of Computational Bioscience, Center for Information Technology, National Institutes of Health, Bethesda, Maryland 20892, United States

S *Supporting Information*

ABSTRACT: The starting point of many fundamental biological processes is associated with protein molecules finding and recognizing specific sites on DNA. However, despite a large number of experimental and theoretical studies on protein search for targets on DNA, many molecular aspects of underlying mechanisms are still not well understood. Experiments show that proteins bound to DNA can switch between slow recognition and fast search conformations. However, from a theoretical point of view, such conformational transitions should slow down the protein search for specific sites on DNA, in contrast to available experimental observations. In addition, experiments indicate that the nucleotide composition near the target site is more symmetrically homogeneous, leading to stronger effective interactions between proteins and DNA at these locations. However, as has been shown theoretically, this should also make the search less efficient, which is not observed. We propose a possible resolution of these problems by suggesting that conformational transitions occur only within a segment around the target where stronger interactions between proteins and DNA are observed. Two theoretical methods, based on continuum and discrete-state stochastic calculations, are developed, allowing us to obtain a comprehensive dynamic description for the protein search process in this system. The existence of an optimal length of the conformational transition zone with the shortest mean search time is predicted.
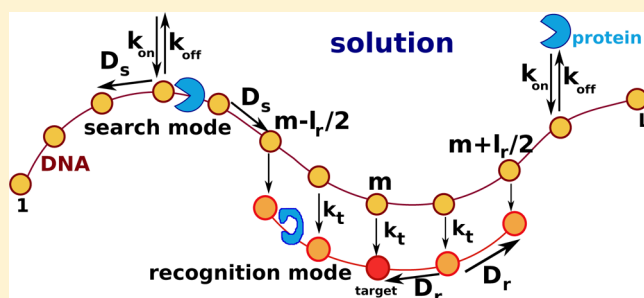
Nucleic acids and proteins are two main building blocks of all living systems, and interactions between them are responsible for maintaining, transferring, and modifying all genetic information.[1] Generally, the beginning of major biological processes is associated with protein molecules finding, recognizing, and binding to specific sequences on DNA molecules. This generates cascades of biochemical transitions that support the functioning of the living systems. Because of its fundamental importance, the protein search phenomena have been intensively studied in the last 40 years using multiple experimental[2,6,10−12,16−24,31,33] and theoretical[3−9,15,21,25−27,29,30,32,34,35] methods. Significant progress in understanding the molecular picture of the target search on DNA has been achieved, but many aspects of underlying mechanisms remain not fully explained.[8,9,25]

Multiple experimental studies show that many proteins associate to specific sites on DNA much faster than predicted from three-dimensional (3D) bulk solution diffusion estimates.[3,5,6,8,9] It has been argued theoretically that fast protein search is a result of combining 3D bulk solution motion with one-dimensional (1D) sliding of nonspecifically bound proteins along the DNA chain,[5,6,8,9] which is confirmed by directly visualizing the protein motion in single-molecule experiments.[10,12,16−18,22,24,31] This is known as a *facilitated diffusion*.

Experiments also indicate that the nonspecifically bound proteins are involved in conformational transitions between searching conformations with weak protein−DNA interactions, when the proteins slide quite fast along DNA, and stronger interacting recognition conformations, when the proteins move much slower.[17,20,21,23,31] The protein molecule can identify the specific target site only in the recognition mode. However, it seems that these conformational fluctuations should make the search less efficient. Theoretical calculations clearly show that for realistic conditions such conformational transitions significantly slow the search dynamics, and the so-called "speed−affinity trade-off" is observed.[21,26] The stronger the interactions in the recognition mode, the slower the association rate to the specific target because the protein molecules become effectively trapped in the recognition conformation where they move slowly. Clearly, there is a challenge for biological systems to achieve both rapid search and high affinity.[17,20,21,31] In addition, bioinformatics analysis of nucleotide composition near the specific sites on DNA indicates that targets are surrounded by more symmetric homogeneous nucleotide

segments.[13,14,28] This leads to the enhanced interactions between the proteins and DNA molecules. However, theoretical calculations predict that such additional affinity near the target actually should slow the protein search,[15] which again does not agree with experimental observations. Thus, during the target search proteins can be found in different molecular conformations, but the role of these conformational transitions remains not fully understood.

In this Letter, we present a resolution of these controversial observations by introducing a mechanism that can reconcile experimental and theoretical results. Our hypothesis is that the enhanced interactions near the target sites due to symmetric homogeneous nucleotide segments stimulate the conformational transitions only at some limited range around the specific sequence, and not everywhere along the DNA chain. The logic here is that cutting the size of the recognition mode will decrease the trapping effect. At the same time, the positive effects due to finding the target by sliding along DNA in the recognition mode and coming directly to the target site from the searching mode are still preserved. We developed a theoretical description of this hypothetical mechanism using two different approaches via continuum and discrete-state stochastic models. This provides a comprehensive description of the search process, and our main finding is that there is an optimal length of the conformational transition zone that leads to the shortest mean search time.

We consider a protein search process for a specific sequence on a single DNA molecule, as shown in Figures 1 and 2, for the
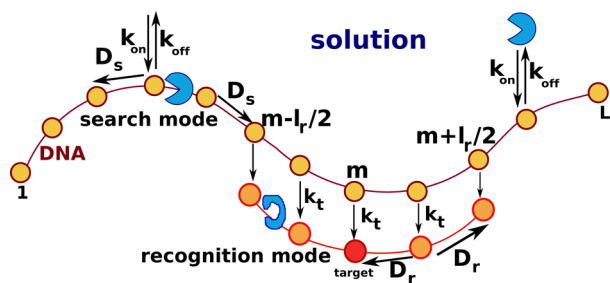

Figure 1. General scheme of the discrete-state model for the protein target search on DNA. There are L sites on DNA. The protein molecule starts the search in the solution. It can bind to DNA with a rate k_{on} in the search conformation only. The irreversible conformational transitions with a rate k_t to the recognition state can take place at sites $m - l_r/2, ..., m + l_r/2$. The protein in the recognition conformation can find the target at the site m. Alternatively, the protein can come to the target site directly from the site m in the search conformation. The protein diffusion rates in the searching and recognition states are D_s and D_r, respectively.

discrete-state and continuum descriptions, respectively. In the discrete-state model (Figure 1) the DNA molecule is viewed as having L sites, and to each of them the protein can bind nonspecifically from the solution with a rate k_{on}/L. After the association, the protein molecule is found in the searching conformation where it interacts weakly with DNA and it can diffuse along the chain with a rate D_s. The protein can also dissociate back into the solution with a rate k_{off}. There are l_r binding sites ($m - l_r/2, ..., m + l_r/2$), from which the protein can switch into the recognition conformation with a rate k_t. Because in the recognition state the protein interacts much stronger with the DNA molecule, we assume that this transition is effectively irreversible. This essentially means that once the
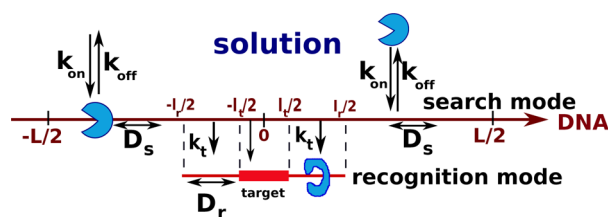

Figure 2. General scheme for the protein target search on DNA in the continuum model framework. The length of DNA is equal to L. The target of length l_t is located symmetrically in the middle of the DNA molecule. There is a special region of length l_r around the target where protein can transit from the search to recognition mode with the rate k_t. A protein molecule can slide along DNA with the diffusion rate D_s in the search mode, or D_r in the recognition mode. It also might dissociate into the solution with the rate k_{off} from search mode. From the solution the protein can associate in the search mode on DNA with the rate k_{on} per chain.

protein enters the recognition state it can neither dissociate from the DNA chain nor leave this region. This is a reasonable approximation which significantly simplifies the calculations. But it can also be argued that relaxing the irreversibility of the conformational transition will not affect main results of our analysis. In the recognition state, the protein can diffuse with a rate D_r. When the protein reaches the state m in the recognition mode, it is defined as the end of the searching process. It can be achieved either by direct transition from the state m in the search mode or via diffusion in the recognition mode (see Figure 1). The continuum model, presented in Figure 2, is very similar to the discrete-state model. Here we assume that the recognition mode and the target region of lengths l_t and $l_r - l_t$ are positioned symmetrically in the middle of the system, while all dynamic rules are the same as in the discrete-state model.

One should note that the model postulates that the protein after unbinding from DNA is equally likely to reassociate to any site on the strand. This assumption is quite realistic as one can judge from experimental observations. Physically, this is a consequence of fast bulk diffusion of the protein. Typical protein diffusions constants in the bulk medium are $D_3 \sim 1$ $\mu m^2/s$, while 1D protein diffusion is 2−3 orders of magnitude slower.[16] Then a single protein can explore a bacterial cell of size $x \sim 1$ μm in about $t \sim x^2/D_3 = 1$ s, which is much smaller than typical transcription factor search times 10−100 s in bacteria.[18] These arguments also suggest that $k_{on} < D_3/x^2$, i.e., the protein must spend enough time in the solution in order to explore the whole DNA region.

Our main goal is to calculate the mean time, T_0, for a protein molecule, which starts in the solution, to reach the target at the site m in the recognition mode (see Figure 1). Because the search ends as soon as the protein arrives to the target site for the first time, T_0 is associated with the mean first-passage time that allows us to explicitly analyze this process as described below.

To explain better the proposed mechanism, it is reasonable to consider a simpler, but still quite realistic, case of very fast conformational transitions, $k_t \rightarrow \infty$. In this case, assuming the discrete-state description, the problem simplifies and reduces to that for an effective one-dimensional chain of two types of sites. If the protein is found on any site n such that $1 \leq n < m - l_r/2$ or $m + l_r/2 < n \leq L$, it is in the search mode with the diffusion rate D_s. But on sites $m - l_r/2 \leq n \leq m + l_r/2$ the protein is in the recognition conformation with the diffusion rate D_r. Once the protein reaches the recognition mode, it cannot return to

the search mode. We define functions $G_n(t)$ and $F_n(t)$ as the first-passage time probability density functions of reaching the target site m at time t starting from the site n at $t = 0$ in the recognition conformation [functions $G_n(t)$] or in the search conformation [functions $F_n(t)$]. In addition, $F_0(t)$ is the first-passage time probability density function when the protein starts in the solution. Then the mean search time can be calculated as

$$T_0 = \int_0^\infty t F_0(t) \, dt \qquad (1)$$

These first-passage time probability densities are governed by a set of backward master equations,[7,15,26,32] and for the search conformations we have

$$
\begin{cases}
\dfrac{\partial F_n(t)}{\partial t} = D_s[F_{n-1}(t) + F_{n+1}(t)] \\
\qquad + k_{off}F_0(t) - (2D_s + k_{off})F_n(t) & (2) \\[2mm]
\dfrac{\partial F_1(t)}{\partial t} = D_s F_2(t) + k_{off}F_0(t) \\
\qquad - (D_s + k_{off})F_1(t) & (3) \\[2mm]
\dfrac{\partial F_L(t)}{\partial t} = D_s F_{L-1}(t) + k_{off}F_0(t) \\
\qquad - (D_s + k_{off})F_L(t) & (4) \\[2mm]
\dfrac{\partial F_0(t)}{\partial t} = \dfrac{k_{on}}{L}\Big[\displaystyle\sum_{n=1}^{m-l_r/2-1} F_n(t) \\
\qquad + \displaystyle\sum_{n=m+l_r/2+1}^{L} F_n(t) \\
\qquad + \displaystyle\sum_{m-l_r/2}^{m+l_r/2} G_n(t) \Big] - k_{on}F_0(t) & (5)
\end{cases}
$$

Similar expressions can be written for the recognition region

$$
\begin{cases}
\dfrac{\partial G_n(t)}{\partial t} = D_r[G_{n-1}(t) + G_{n+1}(t)] \\
\qquad - 2D_r G_n(t) & (6) \\[2mm]
\dfrac{\partial G_{m-l_r/2}(t)}{\partial t} = D_r[G_{m-l_r/2+1}(t) \\
\qquad - G_{m-l_r/2}(t)] & (7) \\[2mm]
\dfrac{\partial G_{m+l_r/2}(t)}{\partial t} = D_r[G_{m+l_r/2-1}(t) \\
\qquad - G_{m+l_r/2}(t)] & (8)
\end{cases}
$$

In addition, we have initial conditions $F_n(t = 0) = 0$, $G_n(t = 0) = 0$ ($n \neq m$), and $G_m(t) = \delta(t)$, because if the protein starts in the target site the search is instantly accomplished.

These backward master equations can be solved by using the method of Laplace transformations,[7,15,26,32] where the first-passage time probability density functions are transformed as $\widetilde{F_n}(s) \equiv \int_0^\infty F_n(t)e^{-st} \, dt$ and $\widetilde{G_n}(s) \equiv \int_0^\infty G_n(t)e^{-st} \, dt$. After the Laplace transformation, the original backward master equations take the form

$$
\begin{cases}
(s + 2D_s + k_{off})\widetilde{F_n}(s) = D_s[\widetilde{F_{n+1}}(s) \\
\qquad + \widetilde{F_{n-1}}(s)] + k_{off}\widetilde{F_0}(s) & (9) \\[2mm]
(s + D_s + k_{off})\widetilde{F_1}(s) = D_s\widetilde{F_2}(s) \\
\qquad + k_{off}\widetilde{F_0}(s) & (10) \\[2mm]
(s + D_s + k_{off})\widetilde{F_L}(s) = D_s\widetilde{F_{L-1}}(s) \\
\qquad + k_{off}\widetilde{F_0}(s) & (11) \\[2mm]
(s + k_{on})\widetilde{F_0}(s) = \dfrac{k_{on}}{L}\Big[\displaystyle\sum_{n=1}^{m-l_r/2-1} \widetilde{F_n}(s) \\
\qquad + \displaystyle\sum_{n=m+l_r/2+1}^{L} \widetilde{F_n}(s) + \displaystyle\sum_{m-l_r/2}^{m+l_r/2} \widetilde{G_n}(s) \Big] & (12)
\end{cases}
$$

for the search region, while for the recognition region it can be shown that

$$
\begin{cases}
(s + 2D_r)\widetilde{G_n}(s) \\
\qquad = D_r[\widetilde{G_{n-1}}(s) + \widetilde{G_{n+1}}(s)] & (13) \\[2mm]
(s + D_r)\widetilde{G_{m-l_r/2}}(s) = D_r\widetilde{G_{m-l_r/2+1}}(s) & (14) \\[2mm]
(s + D_r)\widetilde{G_{m+l_r/2}}(s) = D_r\widetilde{G_{m+l_r/2-1}}(s) & (15)
\end{cases}
$$

and $\widetilde{G_m}(s) = 1$.

It is convenient first to solve eqs 13−15 for the recognition conformations. Here we find that

$$\widetilde{G_n}(s) = \frac{x^{|n-m|} + x^{2l_r+1-|n-m|}}{1 + x^{2l_r+1}} \qquad (16)$$

for $m - l_r/2 \leq n \leq m + l_r/2$, where the auxiliary parameter x is defined as

$$x = \frac{s + 2D_r - \sqrt{s^2 + 4D_r s}}{2D_r} \qquad (17)$$

Substituting $\widetilde{G_n}(s)$ in eq 16 into the last sum in eq 12, we then can find a solution for the function $\widetilde{F_0}(s)$, which is our main goal, by solving eqs 9−12. The result is

$$\widetilde{F_0}(s) = \frac{[aS(s) + Q(s)](s + k_{off})k_{on}}{Ls(s + k_{on} + k_{off}) + (S(s) + 2l)k_{on}k_{off}} \qquad (18)$$

where we have introduced the following notations:

$$Q(s) = \frac{(1 + x)(1 + x^{l_r/2+1})(1 - x^{l_r/2})}{(1 + x^{l_r+1})(1 - x)} \qquad (19)$$

$$a = \frac{x^{l_r} + x^{l_r/2+1}}{1 + x^{l_r+1}} \qquad (20)$$

and

$$S(s) = \frac{y(1 + y)(y^{2l_r} - y^{2L-l_r})}{(1 - y)(y^{m-l_r/2} + y^{l_r/2+1-m})(y^{m+l_r/2} + y^{2L+1-m-l_r/2})} \qquad (21)$$

with

$$y = \frac{s + 2D_s + k_{off} - \sqrt{(s + 2D_s + k_{off})^2 - 4D_s^2}}{2D_s} \quad (22)$$

The explicit expression for the Laplace transform of the first-passage time probability density allows us to fully describe the search dynamics. Let us concentrate on determining the mean search time T_0. It can be found from the Laplace transform $\widetilde{F_0}(s)$, using the relation $T_0 = -\partial \widetilde{F_0}/\partial s|_{s=0}$. The result is

$$T_0 = \frac{Lk_{off} + (L - S(0) - l_r)k_{on}}{(S(0) + l_r)k_{on}k_{off}} + \frac{(l_r/2)(l_r/2 + 1)(3S(0) + 2l_r - 1)}{6D_r(S(0) + l_r)} \quad (23)$$

The physical meaning of the mean search time in eq 23 can be easily explained. This time is the sum of two terms, $T_0 = t_1 + t_2$, with

$$t_1 = \frac{Lk_{off} + (L - S(0) - l_r)k_{on}}{(S(0) + l_r)k_{on}k_{off}}$$
$$= \frac{L}{S(0) + l_r}\frac{1}{k_{on}} + \left(\frac{L}{S(0) + l_r} - 1\right)\frac{1}{k_{off}} \quad (24)$$

and

$$t_2 = \frac{(l_r/2)(l_r/2 + 1)(3S(0) + 2l_r - 1)}{6D_r(S(0) + l_r)} \quad (25)$$

The first term, t_1, corresponds to the mean time it takes for the protein molecule to reach the recognition region. In eq 24, $L/(S(0) + l_r)$ is the average number of association events from the solution, and $L/(S(0) + l_r) - 1$ is the average number of dissociations. The number of dissociations is less by one than the number of associations because the last binding is successful and it leads the protein molecule to the recognition mode. The second term, t_2 (see eq 25), is also simple to understand: it is just the mean searching time when the protein is in the recognition mode with a quadratic scaling on the size of the recognition region l_r, as expected.

Protein search times calculated using the discrete-state model at various conditions are presented in Figures 3 and 4. The parameters used in calculations are arbitrarily chosen just to illustrate the discussed ideas. The most important result here is
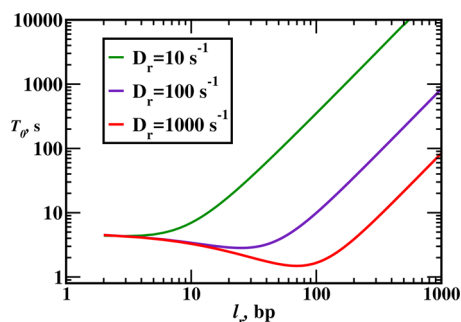
Figure 3. Dependence of the mean search time on the length of the recognition region for different diffusion rates in the recognition mode given by the discrete-state model. The parameters used for calculations are the following: $D_s = 10^3$ s^{-1}, $k_{off} = 10$ s^{-1}, $m = 501$, $L = 1001$ bp, $k_{on} = 10^3$ s^{-1}.
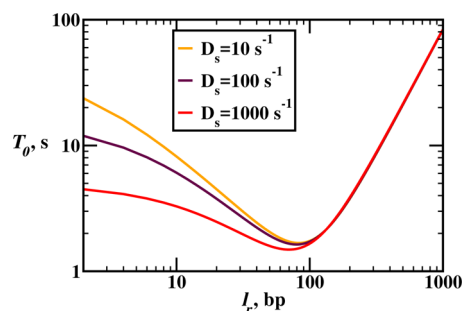
Figure 4. Dependence of the mean search time on the length of the recognition region for different diffusion rates in the search mode given by the discrete-state model. The parameters used for calculations are the following: $D_r = 10^3$ s^{-1}, $k_{off} = 10$ s^{-1}, $m = 501$, $L = 1001$ bp, $k_{on} = 10^3$ s^{-1}.

that there is an optimal length of the recognition region which leads to the shortest search times. This is a consequence of the l_r-dependencies of the mean times t_1 and t_2: the former is a monotonically decreasing function of the length of the recognition region, while the latter monotonically increases with l_r. It can be shown that the optimal length corresponds to the situation when the protein spends comparable times to reach the recognition region and to find the target after entering the recognition mode, i.e., $t_1 \approx t_2$. Both the optimal size and the effect of the search acceleration depend on several factors. Increasing the diffusion rate D_r increases the optimal length (see Figure 3). This is because for larger D_r the time t_2 should decrease, and to keep $t_1 \approx t_2$ the search region $(L - l_r)$ should shrink. The optimal length and the acceleration also depend, although to a less degree, on the diffusion rate in the search mode, D_s, as shown in Figure 4. Here, increasing the diffusion rate in the search mode lowers t_1, and to decrease t_2 correspondingly one should shrink the recognition region. The effect of D_s is smaller than the effect of D_r because in the search mode the protein can dissociate into the solution and rebind back, while in the recognition mode the diffusion is the only process.

Our analysis can be extended to the case when the conformational transition rate k_t is comparable to other rates and the size of the target might also vary. In this situation, it is convenient to analyze the continuum model shown in Figure 2. One can find the details of the calculations in the Supporting Information. Here we just give the main results.

Because the conformational transition is irreversible, the overall mean search time is again the sum of two terms, $T_0 = t_1 + t_2$, which are the mean times required for entering into the recognition mode and for finding the target after reaching this mode. It can be shown that the first mean time is equal to

$$t_1 = \frac{1}{k_{off}}\left[\frac{(k_{off} + k_t)(k_{off} + k_{on})L}{k_t k_{on}(l_r + S)} - 1\right] \quad (26)$$

In this expression, the parameter S is given by

$$S = \frac{2k_t}{\sqrt{k_{off}(k_{off} + k_t)}}\frac{\tanh(z_1)\tanh(z_2)}{\lambda_1\tanh(z_1) + \lambda_2\tanh(z_2)} \quad (27)$$

where

$$\lambda_1 = \sqrt{k_{off}/D_s}, \quad \lambda_2 = \sqrt{(k_{off} + k_t)/D_s} \quad (28)$$

and

$$z_1 = (L - l_r)\lambda_1/2, \quad z_2 = l_r\lambda_2/2 \quad (29)$$

The expression obtained for time t_2 needed to reach the target while being in the recognition mode has the form

$$t_2 = \frac{2}{D_r(l_r + S)}\left\{\frac{(l_r - l_t)^3}{3} + \frac{S}{2\lambda_2^2}\left[\frac{(l_r - l_t)^2\lambda_2^2}{2} - 1\right.\right.$$
$$\left.\left. + \frac{1}{\sinh(z_2)}\left(\sinh\left(\frac{l_t\lambda_2}{2}\right) + (l_r - l_t)\lambda_2\cosh\left(\frac{l_t\lambda_2}{2}\right)\right)\right]\right\}$$
$$(30)$$

where l_t is the target size (see Figure 2).

The results of the calculations of the mean search times for the continuum model are presented in Figures 5 and 6. One
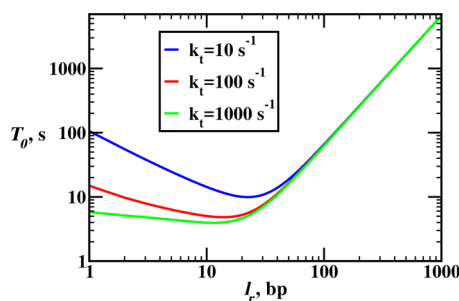

Figure 5. Dependence of the mean search time on the length of the recognition region for different transition rates to the recognition mode. The parameters used for calculations are the following: $D_r = 10^2$ s^{-1}, $D_s = 10^3$ s^{-1}, $k_{off} = 10$ s^{-1}, $l_t = 1$, $L = 1001$, $k_{on} = 10^3$ s^{-1}.
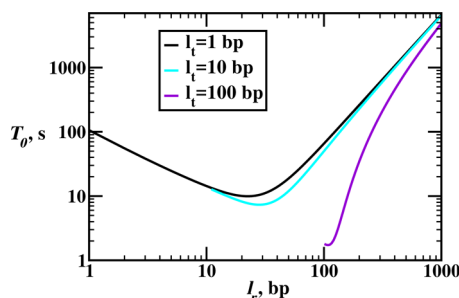

Figure 6. Dependence of the mean search time on the length of the recognition region for different target sizes. The parameters used for calculations are the following: $D_r = 10^2$ s^{-1}, $D_s = 10^3$ s^{-1}, $k_{off} = 10$ s^{-1}, $k_t = 10$ s^{-1}, $L = 1001$, $k_{on} = 10^3$ s^{-1}.

can see that the theory again predicts the existence of an optimal length of the conformational transition region. The dependence of the optimal length on the parameters can also be explained using the balance arguments for proteins to be found comparable times in the search and in recognition modes. Increasing the transition rate k_t makes the search faster and shifts the optimal length l_r to smaller values (see Figure 5). This is due to the fact that higher transition rates decrease the time t_1, so to compensate the time t_2 should be decreased by shortening the length of the recognition region. The protein search also depends on the size of the target l_t as presented in Figure 6. Increasing l_t lowers the search time, as expected. But it also increases the optimum length of the recognition region because the protein diffuses shorter distances in the recognition mode, which yields smaller t_2. To compensate for this, a shorter

search segment without conformational transitions is needed to lower t_1.

One of the assumptions in our theoretical model is the irreversibility of the transition from the search mode into the recognition mode. Relaxing this assumption should not change our main conclusion about the existence of the optimal length of the recognition region. It might be interesting to explore this issue more to understand how the strength of protein−DNA interactions influences the search dynamics.

To summarize, we propose a new mechanism of the protein search for specific sequences on DNA with conformational fluctuations. It is argued that stronger interactions near the target region stimulate the fluctuation transitions only in the limited range of DNA around the target site. Two complementary theoretical approaches to quantitatively describe this mechanism are developed. The first method is based on the discrete-state stochastic model that analyzes the protein search dynamics as a first-passage problem. The second method is the continuum model that utilizes the diffusion-reaction equations to evaluate the mean search time. Theoretical calculations predict that there is an optimal length of the conformational transition region, which accelerates the protein search. It corresponds to the balance between being in the search and in the recognition mode. It has a clear physical meaning. The diffusion in the recognition mode is usually slow, and making the recognition region long will slow the search because of the effective trapping in these conformational states. If the recognition segment is too short, the protein might slide over this region without switching to the recognition mode, and this is not efficient for the overall search. Clearly, there must be an optimum length that minimizes the mean search time. It will be important to test our idea using experimental approaches as well as more advanced theoretical methods.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpclett.7b01750.

> Derivation of the expression for the mean search time in a more general case of arbitrary k_t in the continuum model framework, presented in Figure 2, assuming that the size of the target may also vary (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: tolya@rice.edu.

ORCID Ⓞ

Anatoly B. Kolomeisky: 0000-0001-5677-6690

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Alberts, B. et al. *Mol. Biol. Cell*, 6th ed.; Garland Science: New York, 2014.

(2) Riggs, A. D.; Bourgeois, S.; Cohn, M. The lac repressor-operator interaction. 3. Kinetic studies. *J. Mol. Biol.* **1970**, *53*, 401.

(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.

(4) Berg, O. G.; von Hippel, P. H. Diffusion-controlled macromolecular interactions. *Annu. Rev. Biophys. Biophys. Chem.* **1985**, *14*, 131−60.

(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.

(6) Halford, S. E.; Marko, J. F. How do site-specific DNA-binding proteins find their targets? *Nucl. Acid Res.* **2004**, *32*, 3040.

(7) Kolomeisky, A. B. Mechanism of protein binding to DNA: statistical interactions are important. *Biophys. J.* **2013**, *104*, 966.

(8) 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.

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

(10) 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*, 15883.

(11) 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. U. S. A.* **2007**, *104*, 13948.

(12) 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.

(13) 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.

(14) Afek, A.; Lukatsky, D. B. Nonspecific protein-DNA Binding is Widespread in the Yeast Genome. *Biophys. J.* **2012**, *102*, 1881.

(15) Shvets, A. A.; Kolomeisky, A. B. Sequence heterogeneity accelerates protein search for targets on DNA. *J. Chem. Phys.* **2015**, *143*, 245101.

(16) 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.

(17) 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*, L1−L3.

(18) 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.

(19) Mahmutovic, A.; Berg, O. G.; Elf, J. What matters for lac repressor search in vivo-sliding, hopping, intersegment transfer, crowding on DNA or recognition? *Nucleic Acids Res.* **2015**, *43*, 3454.

(20) Zandarashvili, L.; Vuzman, D.; Esadze, A.; Takayama, Y.; Sahu, D.; Levy, Y.; Iwahara, J. Asymmetrical roles of zinc fingers in dynamic DNA-scanning process by the inducible transcription factor Egr-1. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, E1724.

(21) Zandarashvili, L.; Esadze, A.; Vuzman, D.; Kemme, C. A.; Levy, Y.; Iwahara, J. Balancing between affinity and speed in target DNA search by zinc-finger proteins via modulation of dynamic conformational ensemble. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E5142.

(22) Gorman, J.; Greene, E. C. Visualizing one-dimensional diffusion of proteins along DNA. *Nat. Struct. Mol. Biol.* **2008**, *15*, 768.

(23) Cuculis, L.; Abil, Z.; Zhao, H.; Schroeder, C. M. Direct observation of TALE protein dynamics reveals a two-state search mechanism. *Nat. Commun.* **2015**, *6*, 7277.

(24) Gilmore, J. L.; Suzuki, Y.; Tamulaitis, G.; Siksnys, V.; Takeyasu, K.; Lyubchenko, Y. L. Single-molecule dynamics of the DNA-EcoRII protein complexes revealed with high-speed atomic force microscopy. *Biochemistry* **2009**, *48*, 10492.

(25) 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.

(26) Kochugaeva, M.; Shvets, A. A.; Kolomeisky, A. B. How conformational dynamics influences the protein search for targtes on DNA. *J. Phys. A: Math. Theor.* **2016**, *49*, 444004.

(27) Marcovitz, A.; Levy, Y. Obstacles may facilitate and direct DNA search by proteins. *Biophys. J.* **2013**, *104*, 2042.

(28) Afek, A.; Lukatsky, D. B. Positive and negative design for nonconsensus Protein-DNA Binding Affinity in the vicinity of functional binding sites. *Biophys. J.* **2013**, *105*, 1653.

(29) Koslover, E. F.; de la Rosa, M. A. D.; Spakowitz, A. J. Theoretical and computational modeling of target-site search kinetics in vitro and in vivo. *Biophys. J.* **2011**, *101*, 856.

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

(31) Tafvizi, A.; Huang, F.; Fersht, A. R.; Mirny, L. A.; van Oijen, A. M. A single-molecule characterization of p53 search on DNA. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 563.

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

(33) 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. *Nucleic Acids Res.* **2014**, *42*, 7039.

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

(35) Flyvbjerg, H.; Keatch, S. A.; Dryden, D. T. F. Strong physical constraints on sequence-specific target location by proteins on DNA molecules. *Nucleic Acids Res.* **2006**, *34*, 2550.