Optimal length of conformational transitions 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, TX 77005, USA*

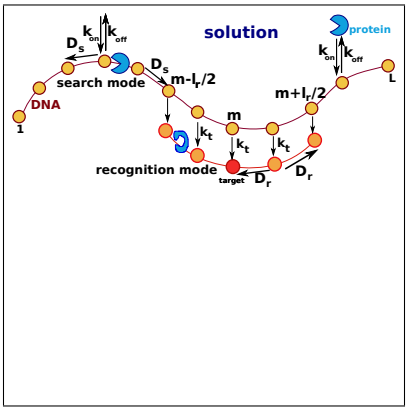
‡*Mathematical and Statistical Computing Laboratory, Division of Computational Bioscience, Center for Information Technology, National Institutes of Health, Bethesda, Maryland 20892, USA*

E-mail: tolya@rice.edu

Abstract

The beginning of most 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 transition between slow recognition and fast search conformations. In addition, the nucleotide composition near the target site is more symmetrically homogeneous, leading to stronger effective interactions of proteins with specific target sites. But theoretical calculations indicate that these effects should significantly slow down the search process, in contradiction with available experimental observations. We propose a possible resolution of this problem by suggesting that conformational transitions are taking place only along 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. It is found that there is an optimal length of the conformational transitions zone with the fastest search time. Physical-chemical mechanisms of the observed phenomena are discussed.

Graphical TOC Entry



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 starting point of most major biological processes is when protein molecules find, recognize and bind to specific sequences on DNA molecules. This generates the cascades of biochemical transitions that support 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] and theoretical[3–9,15,21,25–27,29,30,32,34] methods. A significant progress in understanding the molecular picture of the target search on DNA has been achieved, but many aspects of underlying mechanisms still 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 3D bulk solution diffusion estimates.[3,5,6,8,9] This is known as a *facilitated diffusion*. It has been argued theoretically that fast protein search is a result of combining 3D bulk solution motion with 1D sliding of non-specifically 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] Experiments also indicate that the non-specifically bound proteins are involved in conformational transitions between weaker interacting searching conformations, 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 while only in the recognition mode. At the same time, theoretical calculations indicate that for realistic conditions such conformational transitions significantly slow down 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. But this clearly contradicts to experimentally observed fast search times for typical transcription factor proteins.[17,20,21,31] Furthermore, 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 protein and DNA molecules. However, theoretical calculations

predict that such additional affinity near the target actually slows down the protein search,[15] which again does not agree well with experimental observations.

In this paper, we present a possible resolution of this problem by introducing a mechanism that can reconcile these experimental and theoretical results. Our hypothesis is that the enhanced interactions near the target sites due to symmetric homogeneous nucleotide segments might stimulate the increased conformational transitions activity 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, while the positive effect due to finding the target by sliding along the recognition mode and coming directly from the searching mode is still preserved. We developed a theoretical description of this hypothetical mechanisms using two different approaches via continuum and discrete-state stochastic models. This provides a comprehensive description of the process, and it is found that there is an optimal length of the conformational transitions zone that accelerates the protein search dynamics.

We consider a protein search process for a specific sequence on a single DNA molecule as shown in Figs. 1 and 2 for the discrete-state and continuum descriptions, respectively. In the discrete-state model (Fig. 1) the DNA molecule is viewed as having L sites, and to each of them the protein can bind non-specifically 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 transition 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 is a reasonable approximation which simplifies calculations significantly. But it can be argued also 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 by direct transition from the state m in the search mode, or via diffusion

in the recognition mode: see Fig. 1. The continuum model, presented in Fig. 2, is very similar to the discrete-state model. Here we assume that the recognition mode and the target region of length l_t are positioned symmetrically in the middle of the system, while all dynamic rules are the same as in the discrete-state model.

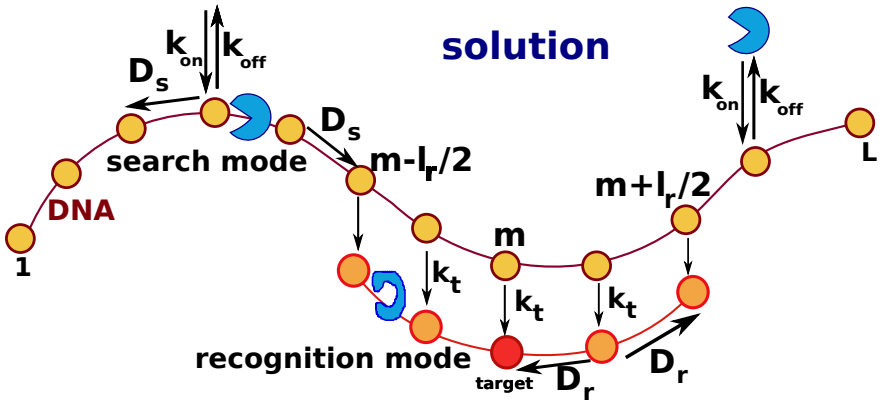


Figure 1: A general scheme of the discrete-state model for the protein target search on DNA. There are L sites on DNA. The protein molecule starts the search in the solution. It can bind to DNA with a rate k_{on} in the search conformation only. The irreversible conformational transitions with a rate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

To explain better the proposed mechanism, it is reasonable to consider a simpler, but still quite realistic, case of very fast conformational transitions, $k_t \to \infty$. In this case, assuming the discrete-state description, the problem simplifies into effective one-dimensional chain of two types of states. 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 zone, it cannot return back to the search mode. We can define functions $G_n(t)$ and $F_n(t)$ as first-passage probability density functions of reaching the target site m at time t starting from the state n at $t = 0$ in the recognition conformation (functions $G_n(t)$) or in the search conformation (functions $F_n(t)$). In addition, the function $F_0(t)$ describes the first-passage probability density function when the protein starts in the solution. These first-passage probabilities are governed by set of backward

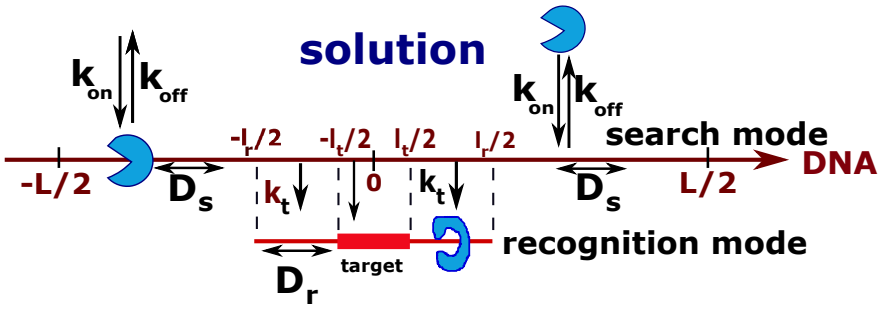


Figure 2: A general scheme for the protein target search on DNA in the continuum model framework. The length of DNA is equal to L. The target of size l_t is located (at the middle) symmetrically on DNA molecule. There is a special region of size l_r around target where protein can transit with the rate k_t. A protein molecule can slide along DNA with the diffusion rate D_s in the search mode, or D_r at 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.

master equations,[7,15,26,32] producing for the recognition conformations

$$\begin{cases} \dfrac{\partial G_n(t)}{\partial t} = D_r \left[G_{n-1}(t) + G_{n+1}(t)\right] - 2D_r G_n(t); & (1) \\ \dfrac{\partial G_{m-l_r/2}(t)}{\partial t} = D_r \left[G_{m-l_r/2+1}(t) - G_{m-l_r/2}(t)\right]; & (2) \\ \dfrac{\partial G_{m+l_r/2}(t)}{\partial t} = D_r \left[G_{m+l_r/2-1}(t) - G_{m+l_r/2}(t)\right]. & (3) \end{cases}$$

In addition, we have $G_m(t) = \delta(t)$ because if the protein starts in the target the search is immediately accomplished.

These backward master equations can be solved by utilizing the method of Laplace transformations,[7,15,26,32] where the first-passage probability functions change via $\widetilde{G_n}(s) \equiv \int_0^\infty G_n(t)e^{-st}dt$. Then the backward master equations are modified as follows,

$$\begin{cases} (s + 2D_r)\widetilde{G_n}(s) = D_r \left[\widetilde{G_{n-1}}(s) + \widetilde{G_{n+1}}(s)\right]; & (4) \\ (s + D_r)\widetilde{G_{m-l_r/2}}(s) = D_r \widetilde{G_{m-l_r/2+1}}(s); & (5) \\ (s + D_r)\widetilde{G_{m+l_r/2}}(s) = D_r \widetilde{G_{m+l_r/2-1}}(s); & (6) \end{cases}$$

with $\widetilde{G_m}(s) = 1$. These equations can be solved, yielding

$$\widetilde{G_n}(s) = \frac{x^{|n-m|} + x^{2l+1-|n-m|}}{1 + x^{2l+1}}, \tag{7}$$

for $m - l_r/2 \leq n \leq m + l_r/2$, and where a parameter x is given by

$$x = \frac{s + 2D_r - \sqrt{s^2 + 4D_r s}}{2D_r}. \tag{8}$$

A similar analysis can be done for the search conformational states. The backward master equations for the functions $F_n(t)$ in the Laplace form can be written as

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

These equations can be solved exactly. For example, for the first-passage probability function starting from the solution we have

$$\widetilde{F_0}(s) = \frac{(aS(s) + Q(s))(s + k_{\text{off}})k_{\text{on}}}{Ls(s + k_{\text{on}} + k_{\text{off}}) + (S(s) + 2l)k_{\text{on}}k_{\text{off}}}, \tag{13}$$

where the auxiliary functions are given by

$$Q(s) = \frac{(1 + x)(1 + x^{l_r/2+1})(1 - x^{l_r/2})}{(1 + x^{l_r+1})(1 - x)}; \tag{14}$$

$$a = \frac{x^l + x^{l_r/2+1}}{1 + x^{l_r+1}}; \tag{15}$$

and

$$S(s) = \frac{y(1+y)(y^{2l} - y^{2L-l_r})}{(1-y)(y^{m-l_r/2} + y^{l_r/2+1-m})(y^{m+l_r/2} + y^{2L+1-m-l_r/2})}, \tag{16}$$

with

$$y = \frac{s + 2D_s + k_{\text{off}} - \sqrt{(s + 2D_s + k_{\text{off}})^2 - 4D_s^2}}{2D_s}. \tag{17}$$

It allows us to evaluate the search time, which we identify as a mean first-passage time to reach the target starting from the solution. It can be found from $T_0 \equiv -\frac{\partial \widetilde{F_0}}{\partial s}|_{s=0}$, leading to the following final expression,

$$T_0 = \frac{Lk_{\text{off}} + (L - S(0) - l_r)k_{\text{on}}}{(S(0) + l_r)k_{\text{on}}k_{\text{off}}} + \frac{(l_r/2)(l_r/2+1)(3S(0) + 2l_r - 1)}{6D_r(S(0) + l_r)}. \tag{18}$$

The physical meaning of the search time from Eq. (18) can be easily explained. It consists of two terms, $T_0 = t_1 + t_2$, with

$$t_1 = \frac{Lk_{\text{off}} + (L - S(0) - l_r)k_{\text{on}}}{(S(0) + l_r)k_{\text{on}}k_{\text{off}}} = \left(\frac{L}{S(0) + l_r}\right)\frac{1}{k_{\text{on}}} + \left(\frac{L}{S(0) + l_r} - 1\right)\frac{1}{k_{\text{off}}}, \tag{19}$$

and

$$t_2 = \frac{(l_r/2)(l_r/2+1)(3S(0) + 2l_r - 1)}{6D_r(S(0) + l_r)}. \tag{20}$$

The first term, t_1 corresponds to the mean time it takes for the protein molecule to reach the recognition region. In Eq. (19) $S(0)$ corresponds to the distance that protein diffuses along the DNA in the searching configuration before dissociating back into the solution. Then $\left(\frac{L}{S(0)+l_r}\right)$ is the average number of association events from the solution, and $\left(\frac{L}{S(0)+l_r} - 1\right)$ 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. (20), is also simple to understand: it is just the 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.

The results of calculations for the protein search times using the discrete-state model are presented in Figs. 3 and 4. The most important observation is that there is an optimal length of the fluctuations region which leads to the fastest search times. It can be shown that this length corresponds to the situation when the protein spends comparable times to reach the recognition region and to find the target from the recognition mode, i.e., $t_1 \approx t_2$. The optimal size and the effect of the search acceleration depends on several factors. Increasing the diffusion constant D_r increases the optimal length: see Fig. 3. This is because for larger D_r the time t_2 should decrease, and to keep $t_1 \approx t_2$ the non-fluctuations region $(L - l_r)$ should shrink. The optimal length and the acceleration also depend, although to a less degree, on the diffusion constant in the search mode, D_s, as shown in Fig. 4. Here increasing the diffusion rate in the search mode lowers t_1 and to decreases 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.

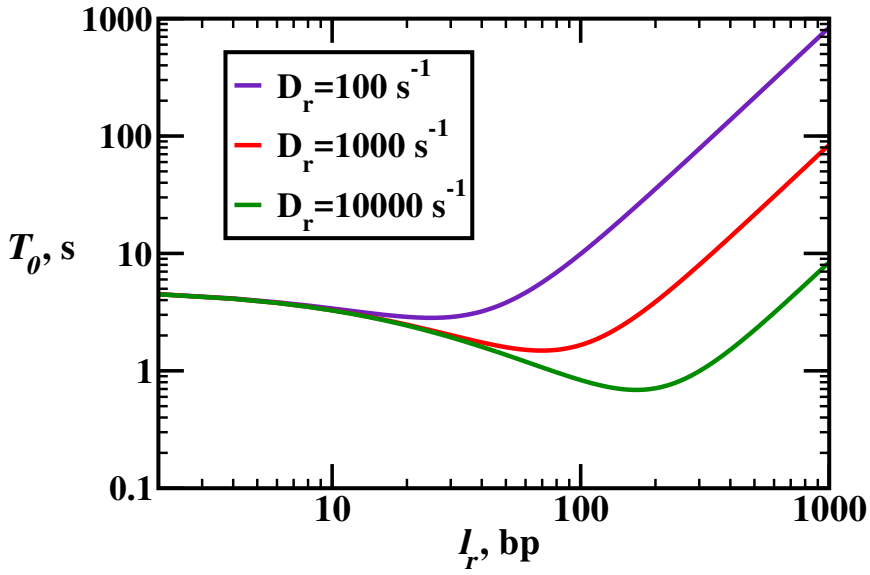


Figure 3: Dependence of the search time on the length of the fluctuation transitions zone for different diffusion constants in the recognition mode from the discrete-state model. The parameters used for calculations are the following: $D_s = 10^3 s^{-1}, k_{\text{off}} = 10 s^{-1}, m = 501, L = 1001 bp, k_{\text{on}} = 10^3 s^{-1}$.

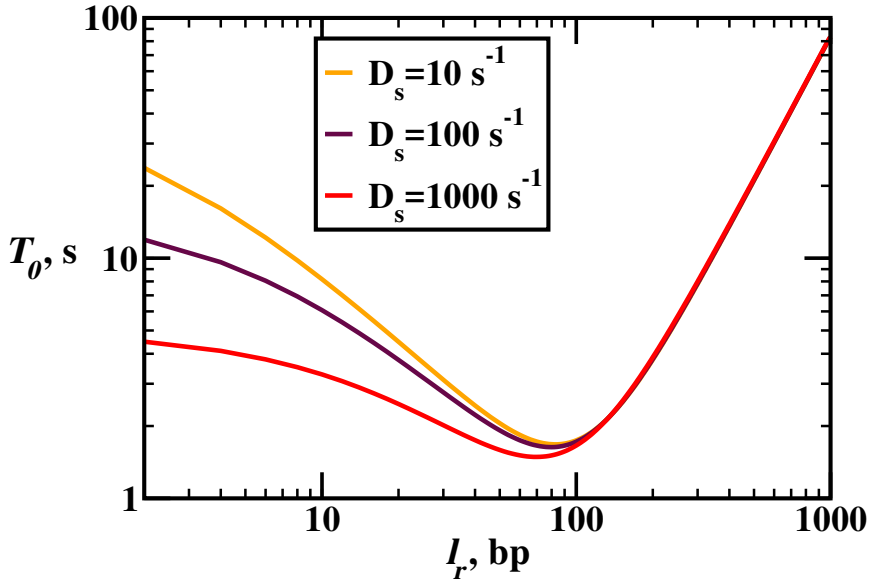


Figure 4: Dependence of the search time on the length of the fluctuation transitions zone for different diffusion constants in the search mode from the discrete-state model. The parameters used for calculations are the following: $D_r = 10^3 s^{-1}$, $k_{\text{off}} = 10$ s^{-1}, $m = 501$, $L = 1001$ bp, $k_{\text{on}} = 10^3$ s^{-1}.

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 presented in Fig. 2. To evaluate the search dynamics, we assume that the system is in the stationary state with a constant flux J_0 of proteins injected into the solution and removed at the target. Then the search time can be calculated from

$$T_0 = \frac{N_b + N_s + N_r}{J_0}, \tag{21}$$

where N_b, N_s and N_r are the steady-state numbers of proteins in the bulk, in the search conformation, and in the recognition conformation, respectively. Because, the conformational transition is irreversible, the overall search time is again can be written as a sum of two terms, $T_0 = t_1 + t_2$, corresponding to entering into the recognition mode and finding the target. They can be written as

$$t_1 = \frac{N_b + N_s}{J_0}, \quad t_2 = \frac{N_r}{J_0}. \tag{22}$$

Each of these times can be evaluated separately as explained below.

The time to reach the recognition conformations states can be calculated using the following arguments. The number of proteins in the search mode is given by

$$N_s = \int_{L/2}^{L/2} c_s(x)dx, \tag{23}$$

where $c_s(x)$ is the stationary protein concentration in the searching mode. It is reasonable to assume that the protein concentration in the region where the conformational transitions are taking place is not the same as in the other part of DNA, i.e., $c_s(x) = c_1(x)$ for $-L/2 < x < -l_r/2$ and $l_r/2 < x < L/2$, and $c_s(x) = c_2(x)$ for $-l_r/2 < x < l_r/2$. These concentrations can explicitly calculated from the following reaction-diffusion equations,

$$D_s \frac{d^2 c_1}{dx^2} - k_{\text{off}} c_1 + (k_{\text{on}}/L) N_b = 0; \tag{24}$$

$$D_s \frac{d^2 c_2}{dx^2} - (k_{\text{off}} + k_t) c_2 + (k_{\text{on}}/L) N_b = 0; \tag{25}$$

with the corresponding initial and boundary conditions. In addition, we have a stationary balance condition,

$$k_{\text{on}} N_b = J_0 + k_{\text{off}} N_s, \tag{26}$$

which can be used to get the explicit expression for t_1, namely

$$t_1 = \frac{1}{k_{\text{off}}} \left[\frac{(k_{\text{off}} + k_t)(k_{\text{off}} + k_{\text{on}})L}{k_t k_{\text{on}}(l_r + S)} - 1 \right]. \tag{27}$$

In this expression, the parameter S is the average scanning length of the protein in the searching mode before dissociating, and it is given by

$$S = \frac{2k_t}{\sqrt{k_{\text{off}}(k_{\text{off}} + k_t)}} \frac{\tanh(z_1)\tanh(z_2)}{\tanh(z_1)/\lambda_1 + \tanh(z_2)/\lambda_2}, \tag{28}$$

where

$$\lambda_1 = \sqrt{D_s/k_{\rm off}}, \quad \lambda_2 = \sqrt{D_s/(k_{\rm off}+k_t)}; \tag{29}$$

and

$$z_1 = \frac{L-l_r}{2\lambda_1}, \quad z_2 = \frac{l_r}{2\lambda_2}. \tag{30}$$

A similar analysis can be done to calculate the time needed to reach the target while being in the recognition mode. The final result is

$$t_2 = \frac{2}{D_r(l_r+S)}\left\{\frac{(l_r-l_t)^3}{3} + \frac{S\lambda_2^2}{2}\left[\frac{(l_r-l_t)^2}{2\lambda_2^2} - 1 + \frac{1}{\sinh(z_2)}\left(\sinh(\frac{l_t}{2\lambda_2}) + \frac{(l_r-l_t)}{\lambda_2}\cosh(\frac{l_t}{2\lambda_2})\right)\right]\right\}. \tag{31}$$

The results of the calculations for the search times for the continuum model are presented in Figs. 5 and 6. One can see that theory again predicts the optimal length of the fluctuations region. The dependence of the optimal length on other parameters also can be explained using the balance arguments for protein to be found comparable times in the search or in the recognition modes. Increasing the transition rate k_t makes the search faster and it shifts the optimal position to smaller values: see Fig. 5. This is due to the fact that larger transition rates decrease the time t_1, so to compensate the time t_2 should be decreased by shortening the length of the fluctuation region. The protein search also depends on the size of the target l_t as presented in Fig. 6. Increasing l_t lowers the search time, as expected. But it also increases the optimum length of the fluctuations region because the protein is diffusing shorter distances, $l_r - l_t$, in the recognition mode, yielding smaller t_2. To compensate for this, shorter search segment without fluctuations is needed to lower t_1.

We introduced a possible new mechanism of the protein search for specific sequences on DNA. It is argued that stronger interactions near the target region stimulate the fluctuation transitions only in the limited range of DNA. Two complementary theoretical approaches to quantitatively describe this mechanisms 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 search times.

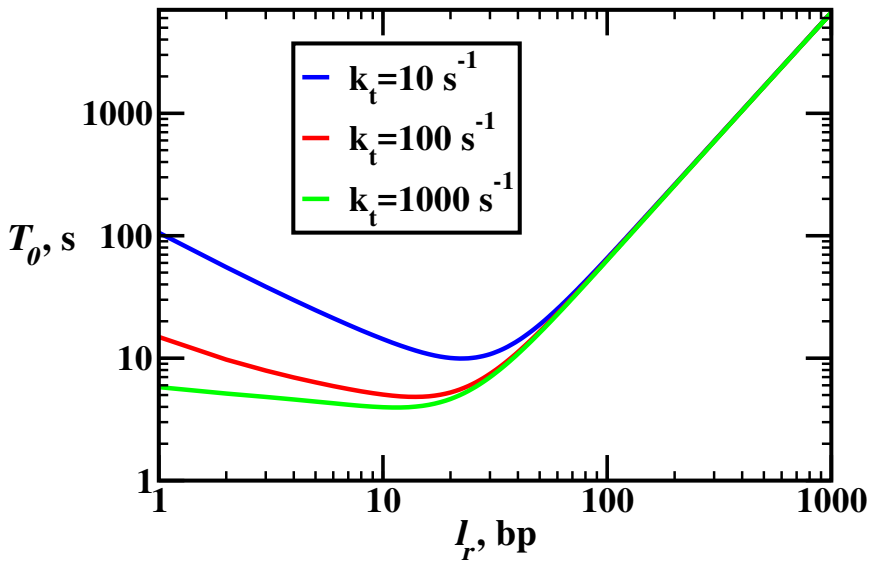


Figure 5: Dependence of the search time on the length of the fluctuation transitions zone for different transition constants to the recognition mode. The parameters used for calculations are the following: $D_r = 10^2 s^{-1}$, $D_s = 10^3 s^{-1}$, $k_{\text{off}} = 10 s^{-1}$, $l_t = 1$, $L = 1001$,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

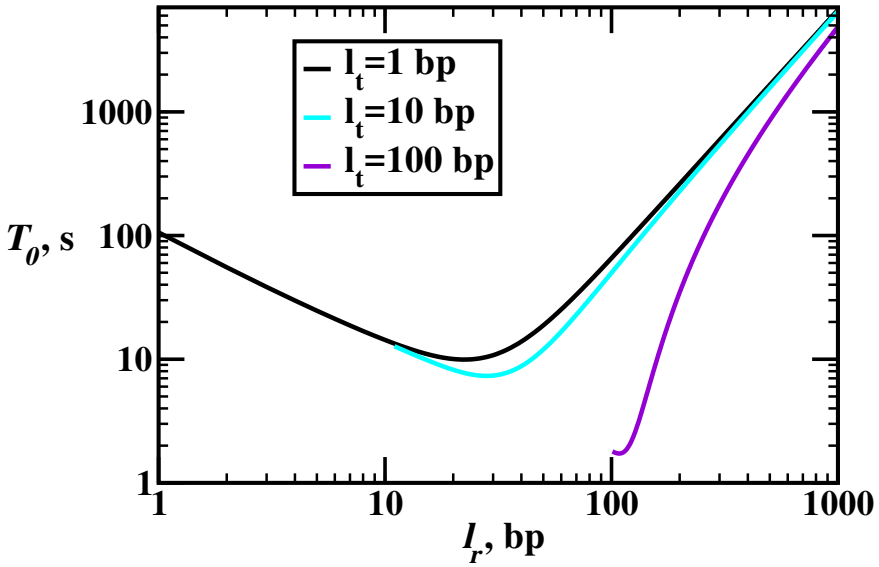


Figure 6: Dependence of the search time on the length of the fluctuation transitions zone for different target sizes. The parameters used for calculations are the following: $D_r = 10^2 s^{-1}$, $D_s = 10^3 s^{-1}$, $k_{\text{off}} = 10 s^{-1}$, $k_t = 10 s^{-1}$, $L = 1001$, $k_{\text{on}} = 10^3 s^{-1}$.

Theoretical calculations suggest that there is the optimal length of the fluctuation transitions region which accelerates the protein search. It corresponds to the balance between being in the search or in the recognition mode. It has a clear physical meaning. The diffusion in the recognition mode is usually slow, and making the fluctuations region long will slow down search due to the effective trapping in these conformational states. If the fluctuations segment is too short, the protein might slide over this region without going into the recognition mode. This also will increase the search time. Clearly, there must be an optimum length that will minimize the search times. It will be important to test our idea using experimental approaches as well as more advanced theoretical methods.

References

(1) Alberts, B., et al. *Molecular Biology of 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**, 48, 67.

(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. *Ann. 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. USA* **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. USA* **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. USA* **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? *Nucl. Acid 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. USA* **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. USA Sci.* **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. and 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. USA* **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. *Nucl. 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. *Nucl. Acids Res.* **2006**, 34 2550.