PAPER

How conformational dynamics influences the protein search for targets on DNA

IOP Publishing Journal of Physics A: Mathematical and Theoretical
J. Phys. A: Math. Theor. **49** (2016) 444004 (15pp) doi:10.1088/1751-8113/49/44/444004

How conformational dynamics influences the protein search for targets on DNA

Maria P Kochugaeva, Alexey A Shvets and Anatoly B Kolomeisky

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

E-mail: tolya@rice.edu

Received 28 June 2016, revised 12 September 2016
Accepted for publication 19 September 2016
Published 13 October 2016

Abstract

Protein search and association to specific sequences on DNA is a starting point for all fundamental biological processes. It has been intensively studied in recent years by a variety of experimental and theoretical methods. However, many features of these complex biological phenomena are still not resolved at the molecular level. Experiments indicate that proteins can be bound non-specifically to DNA in multiple configurations. But the role of conformational fluctuations in the protein search dynamics remains not well understood. Here we develop a theoretical method to analyze how the conformational transitions affect the process of finding the specific targets on DNA. Our approach is based on discrete-state stochastic calculations that take into account the most relevant physical–chemical processes. This allows us to explicitly evaluate the protein search for the targets on DNA at different conditions. Our calculations suggest that conformational fluctuations might strongly affect the protein search dynamics. We explain how the shift in the conformational equilibrium influences the target search kinetics. Theoretical predictions are supported by Monte Carlo computer simulations.

Keywords: facilitated diffusion, discrete-state stochastic models, first-passage processes, protein search

(Some figures may appear in colour only in the online journal)

1. Introduction

Proteins and DNA are two major classes of biological molecules that essentially control and regulate all cellular processes [1, 2]. Protein–DNA interactions activate the cascades of

biochemical transitions in living systems, thus allowing them to function successfully [1, 2]. The starting step in this process is a protein finding and recognizing specific target sequences on DNA. This fundamental phenomenon has been extensively investigated via both experimental and theoretical methods [3–46]. Although we understand better now the protein search on DNA, there are still many unanswered questions on mechanisms of underlying processes [19, 36, 38].

One of the most striking observations in the protein search is the fact that some proteins can find their specific targets on DNA much faster than expected from 3D bulk diffusional estimates that follow from the famous theory developed by Smoluchovski more than 100 years ago [3, 4, 10, 36, 38]. This is known as a *facilitated diffusion*. It has been argued that the accelerated search is possible due to combining 3D bulk motion with 1D sliding of non-specifically bound protein molecules along the DNA chain [10, 36, 38]. Recent single-molecule experiments were able to visualize directly the sliding motion of proteins on DNA during the search [7, 8, 11, 13, 14, 23, 33, 46]. The surprising results in protein search dynamics have also stimulated significant theoretical efforts to explain the molecular picture behind these complex processes [10, 19, 31, 36–38, 41].

Furthermore, several experiments indicated that during the search the non-specifically bound proteins can exist in several conformations with different hopping dynamics along the DNA chain and complex transition processes between these states [6–9, 32, 35]. Studies of tumor suppressor p53 protein found that when this molecule is bound to DNA via one C-terminal domain it slides fast [6–8]. At the same time, in the conformation that involves binding of two protein domains to DNA (C-terminal and core) the motion is much slower [6–8]. Similarly, NMR experiments and MD computational studies on zing-finger Egr-1 proteins suggested that these molecules scan DNA while undergoing conformational transitions between the fast sliding and slow sliding conformational states [9, 32]. In addition, using mutational analysis it was possible to shift the dynamic conformational equilibrium between these states, influencing the target search kinetics [9]. It is reasonable to expect that many other proteins also participate in conformational fluctuations during the search process for specific sites on DNA [35].

The observations of conformational degrees of freedom during the protein search raised a question on the fundamental role of these fluctuations. It was suggested earlier that the existence of several conformations is necessary for fast finding targets on DNA in order to avoid a so-called 'speed-selectivity paradox' [34, 36, 37]. In one conformation, called the recognition mode, the protein interacts stronger with DNA and the specific site can be recognized only in this conformation. In this state the protein moves slowly along the DNA chain, and it was argued that it is not efficient for the search. The acceleration can be achieved by rapid transitioning into another conformational state, called the search mode, where the protein interacts much weaker with the DNA chain, while sliding fast. But the application of these ideas for p53 proteins was only partially successful: the predictions that the recognition state has a higher free energy and conformational transitions are very fast were not fully supported by experimental observations [6–8]. In addition, it was shown later that the speed-selectivity paradox does not exist because it is an artifact of the continuum theoretical description of the protein search process that cannot be utilized for all ranges of parameters [19]. There were also other theoretical studies of the conformational fluctuations in the protein search, but they either assumed the conformational equilibrium [12, 30], or utilized the continuum approach for calculating dynamics properties, which could not fully describe all search regimes [15, 36, 44].

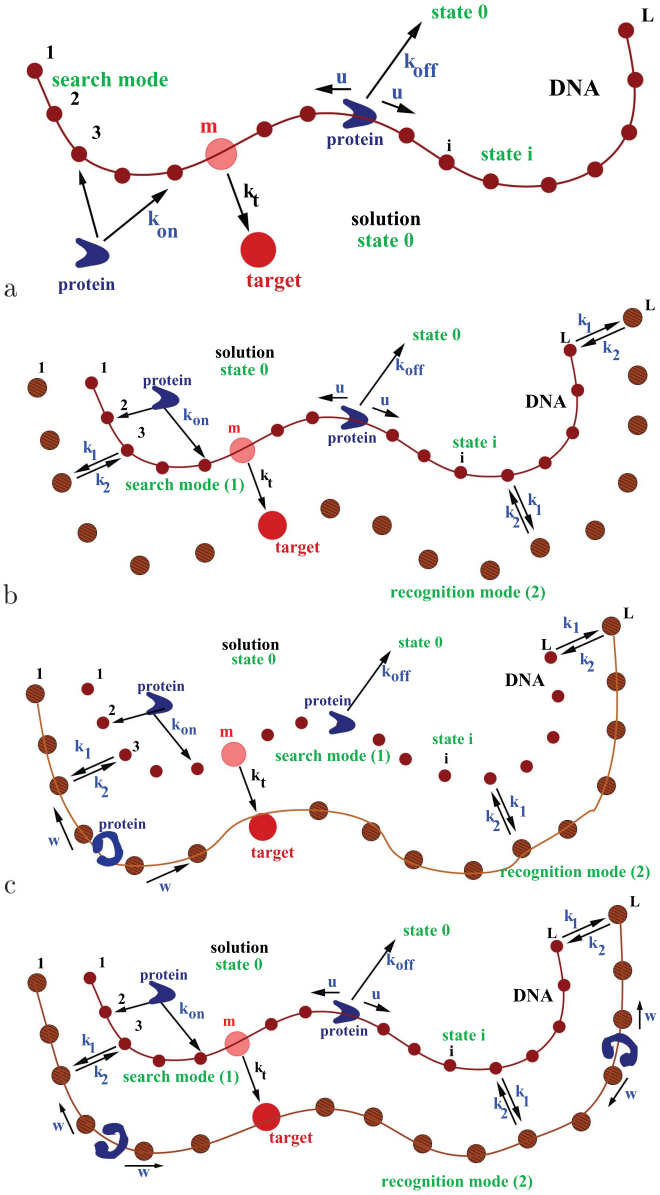

Figure 1. Schematic pictures for various models of the protein search with conformational transitions: (a) model with conformational transition only at the target; (b) model with conformational transitions at all sites and sliding in the searching mode; (c) model with conformational transitions at all sites and sliding in the recognition mode; and (d) model with conformational transitions at all sites and sliding in both searching and recognition modes.

In this paper, we present a theoretical approach for analyzing the effect of conformational transitions in the protein search for targets on DNA. It is based on the discrete-state stochastic framework that was developed earlier in our group [19]. It provides explicit expressions for dynamic properties of the system for *all* ranges of parameters, and the continuum description is a

limiting case of our theoretical method [19]. This method has been already successfully extended and generalized to account for various complex processes that are taking place during the protein search [20, 39, 40, 45]. Applying this approach, we find that the conformational fluctuations can significantly modify the protein search dynamics. Our theoretical calculations are also supported by extensive Monte Carlo computer simulations.

2. Theoretical method and results

In order to understand the effect of conformational transitions during the protein search for targets on DNA, the following strategy is employed. We consider a series of discrete-state stochastic models of increasing complexity, which are shown in figure 1. By adding step by step the relevant dynamic processes into the system, we can clarify better the role of the conformational degrees of freedom and their relations to other processes in the protein search dynamics.

2.1. Conformational transition only at the target

We start with a simplest model where the conformational transition can only take place at the target site: it is presented in figure 1(a). The DNA molecule has L sites, and the protein can diffuse along them with a rate u. At the site m the protein can undergo a conformational transition with a rate k_t into a configuration where it can bind DNA stronger, and this corresponds to a specific binding to the target (see figure 1(a)). From any non-specific site on DNA, the protein can dissociate into solution with a rate $k_{\rm off}$. Because the bulk diffusion for the protein molecule is usually very fast, we assume that from the solution the protein can bind with equal probability to any site on DNA with a total rate $k_{\rm on}$, i.e. the binding rate per site is just $k_{\rm on}/L$. It is assumed also that the protein search always starts from the solution that we label as a state 0 (figure 1(a)). This model is stimulated by experimental observations on *lac*-repressor proteins that frequently slide over the specific target before strongly binding to it [23].

The main idea of our theoretical approach is to view the protein search as a first-passage process, and this leads to a direct method of evaluating all dynamic properties in the system [19]. One can introduce a function $F_n(t)$ as a probability density to reach the specific target at time t for the first time if at $t = 0$ the protein started from the state n ($n = 0, 1, \ldots, m, m_t, m + 1, \ldots L$). Note that $n = m_t$ corresponds to the target state (see figure 1(a)), while m corresponds to the state on DNA above the target. Our main goal is to explicitly evaluate these first-passage probability functions because they fully describe the search dynamics. It can be done by analyzing the temporal evolution of $F_n(t)$ from the backward master equations [19]

$$\frac{{\rm d}F_n(t)}{{\rm d}t} = u\left[F_{n-1}(t) + F_{n+1}(t)\right] + k_{\rm off}F_0(t) - (2u + k_{\rm off})F_n(t), \quad \text{for } 1 < n < L \text{ and } n \neq m; \tag{1}$$

$$\frac{{\rm d}F_0(t)}{{\rm d}t} = \sum_{n=1}^{L}\frac{k_{\rm on}}{L}F_n(t) - k_{\rm on}F_0(t). \tag{2}$$

At DNA boundaries and at the site m the dynamics is different, and the backward master equations can be written as

$$\frac{\mathrm{d}F_1(t)}{\mathrm{d}t} = uF_2(t) + k_{\mathrm{off}}F_0(t) - (u + k_{\mathrm{off}})F_1(t); \tag{3}$$

$$\frac{\mathrm{d}F_L(t)}{\mathrm{d}t} = uF_{L-1}(t) + k_{\mathrm{off}}F_0(t) - (u + k_{\mathrm{off}})F_L(t); \tag{4}$$

$$\frac{\mathrm{d}F_m(t)}{\mathrm{d}t} = k_{\mathrm{off}}F_0(t) + u[F_{m-1}(t) + F_{m+1}(t)] \tag{5}$$

$$+k_t F_{m_t}(t) - (2u + k_{\mathrm{off}} + k_t)F_m(t). \tag{6}$$

In addition, the initial condition implies that

$$F_{m_t}(t) = \delta(t). \tag{7}$$

To solve the backward master equations (1)–(7), we use a method of Laplace transformations, with the corresponding transformations of the first-passage probability functions defined as $\tilde{F}_n(s) = \int_0^\infty \mathrm{e}^{-\mathrm{st}}F_n(t)\mathrm{d}t$ [19, 39, 40]. In this language, the backward master equations are modified into simpler algebraic expressions

$$(s + 2u + k_{\mathrm{off}})\tilde{F}_n(s) = u(\tilde{F}_{n-1}(s) + \tilde{F}_{n+1}(s)) + k_{\mathrm{off}}\tilde{F}_0(s); \tag{8}$$

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

$$(s + u + k_{\mathrm{off}})\tilde{F}_1(s) = u\tilde{F}_2(s) + k_{\mathrm{off}}\tilde{F}_0(s); \tag{10}$$

$$(s + u + k_{\mathrm{off}})\tilde{F}_L(s) = u\tilde{F}_{L-1}(s) + k_{\mathrm{off}}\tilde{F}_0(s); \tag{11}$$

$$(s + 2u + k_{\mathrm{off}} + k_t)\tilde{F}_m(s) = u(\tilde{F}_{m-1}(s) + \tilde{F}_{m+1}(s)) + k_{\mathrm{off}}\tilde{F}_0(s) + k_t. \tag{12}$$

The initial condition can be written simply as $\tilde{F}_{m_t}(s) = 1$.

The general solutions for Laplace transformations of the first-passage probability functions can be expressed as

$$\tilde{F}_n(s) = A(s)y(s)^n + B(s), \tag{13}$$

where A, y and B are unknown parameters that will be explicitly determined by substituting this ansatz into equations (8)–(12) [19]. After the corresponding substitutions, we obtain

$$\tilde{F}_0(s) = \frac{S(s)(s + k_{\mathrm{off}})k_{\mathrm{on}}}{Ls(s + k_{\mathrm{on}} + k_{\mathrm{off}}) + S(s)k_{\mathrm{on}}k_{\mathrm{off}}}, \tag{14}$$

where

$$S(s) = \frac{k_t(1 + y)(y^{2L} - 1)}{(y - 1)(y^{2L-m} + y^{m-1})\theta}, \tag{15}$$

and

$$\theta = (s + 2u + k_{\mathrm{off}} + k_t)(y^m + y^{1-m}) - u\left(y^{m-1} + y^{2-m} + \frac{y^{2m+1} + y^2 + y^{2L} + y^{2L-2m+1}}{y^m + y^{2L+1-m}}\right). \tag{16}$$

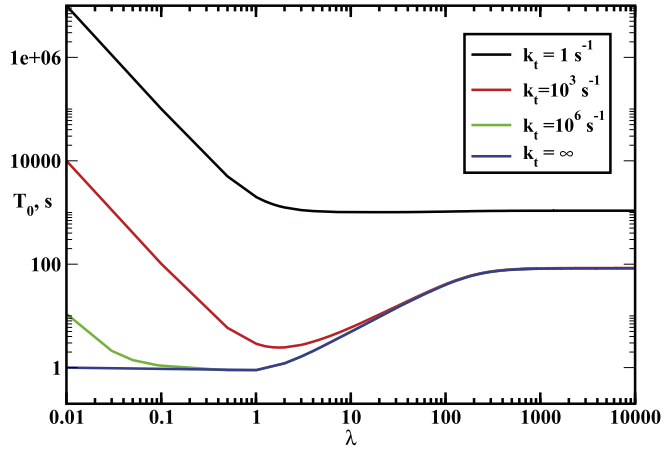

Figure 2. Dynamic phase diagram for the protein search with conformational transition only at the target. Search times as a function of the scanning length λ are shown for different conformational transition rates k_t. The scanning length is varied by changing k_{off}. The parameters used for calculations are the following: $L = 1000$, $m = L/2$, $u = 10^3\ \text{s}^{-1}$ and $k_{\text{on}} = 10^3\ \text{s}^{-1}$.

The parameter y is given by

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

The average protein search time to reach the target can be associated with a mean first-passage time (MFPT), and it can be directly computed from the Laplace transforms via [19, 39, 40]

$$T_0 \equiv -\frac{\partial \tilde{F}_0(s)}{\partial s}\bigg|_{s=0}. \tag{18}$$

The final expression for the search time is given by

$$T_0 = \frac{Lk_{\text{off}} + k_{\text{on}}(L - S(0))}{k_{\text{on}}k_{\text{off}}S(0)}. \tag{19}$$

One can see that in the limit of very fast conformational change, $k_t \gg 1$, the problem reduces to a protein search with single conformation at each site that was already fully analyzed [19, 39].

To describe the effect of conformational fluctuations we construct a dynamic phase diagram that describes the protein search for all possible conditions. The results are presented in figure 2. One can see that there are three dynamic search regimes depending on length of DNA, the average scanning length $\lambda = \sqrt{u/k_{\text{off}}}$ and the size of the target, assumed to be equal to unity in our model. This dynamic behavior is similar to the protein search without conformational transitions [19]. For large scanning lengths, $\lambda > L$, the protein binds to the DNA chain and it diffuses without dissociation until it goes into the target. Most of the time the protein spends on DNA, performing a random walk along the sites. For intermediate scanning lengths, $1 < \lambda < L$, the protein binds to DNA, slides on average a distance λ and dissociates into the solution if the target is not reached. This generally allows for the search to

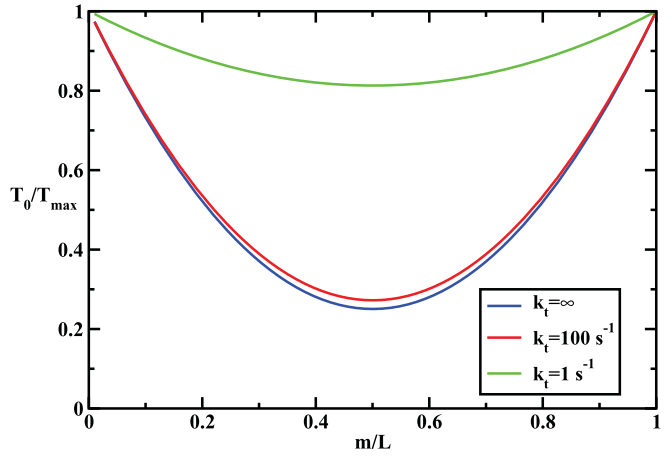

Figure 3. Normalized search times as a function of the relative position of the target along the DNA chain for different conformational transitions rates k_t. For calculations the following parameters were utilized: $L = 1000$, $u = 10^3$ s^{-1}, $k_{on} = 10^3$ s^{-1} and $\lambda = \sqrt{u/k_{off}} = 10000$. The green curve corresponds to $k_t = 1$ s^{-1}, the red curve describes $k_t = 100$ s^{-1} and the blue curve is for $k_t \to \infty$.

proceed faster because the protein is not trapped on DNA doing the slow random walk motions. In both these dynamic regimes the dependence on the conformational transition rate k_t is relatively small, although decreasing the conformational rate k_t increases the search time since this step becomes the rate-limiting for the whole process. But in the regime where $\lambda < 1$ so that there is no sliding along the DNA, the conformational effects are much more important. This is because the protein can only reach the target if it first binds to the site m. From this site it has a probability $p_t = k_t/(2u + k_t + k_{off})$ to transition into the specific configuration. Lowering the scanning length λ for the fixed diffusion rate u corresponds to increasing the rate k_{off}. Then the protein has to return many times to the site m before the successful transformation into the target can happen, and this obviously slows down the search.

It is important to note that the predictions of this model agree with experimental observations on *in vivo* search dynamics of lac-repressor proteins [23]. It was shown in these experiments that this transcription factor slides over the target without strong binding in 90% of the cases. Our theoretical picture explains this by suggesting that the protein most of the time not in the recognition mode, and the internal switching dynamics controls whether it finally gets bound or not only above the target site.

We can also investigate how the protein search dynamics depends on the position of the target m. This is illustrated in figure 3. Our results suggest that the position of the special site is important only in the random-walk regime for $\lambda > L$. For fast transitions into the specific conformation there is a strong dependence on m/L, which decreases with lowering the rate k_t. This can be easily understood if we recall that for small transition rates the conformational transition becomes the rate-limiting step in the whole search process, and the time to reach the site m, which gives the dependence on the target location for large k_t, becomes negligible.

2.2. Conformational transitions at all sites and sliding in the search mode

In the next step, we consider a case where conformational transitions can take place at all sites on DNA. The corresponding model is shown in figure 1(b). From the solution (state 0) the protein can bind to DNA in the searching conformation, which we also label as a search mode or chain 1, with the total rate $k_{\rm on}$. From the searching conformation on site i ($1 \leqslant i \leqslant L$) the protein can dissociate back into the solution with the rate $k_{\rm off}$, or it can transition into the recognition mode (labeled as a chain 2) on site i with a rate k_1. The reverse transition is taking place with a rate k_2: see figure 1(b). While being in the search mode (chain 1) the protein can diffuse with a rate u. But the sliding is not possible in the recognition mode (chain 2). Based on available experimental observations, this model seems to be the most realistic in the description of the protein search processes.

To obtain a dynamic description of this system, we again utilize the first-passage analysis as explained in detail above. We introduce here a function $F_n^{(j)}(t)$ defined as first-passage probability to reach for the first time the target, which is located at the site m on the recognition chain, at time t if at $t = 0$ the protein started at the site n of the chain j ($j = 1$ for the searching mode or $j = 2$ for the recognition mode). $F_0(t)$ is the first-passage probability to reach the target starting from the bulk solution (state 0). The temporal evolution of these probabilities follows the backward-master equations, which can be solved using the Laplace transformations, as explained for the model with conformational transition only at the target. Finally, we obtain

$$\tilde{F}_0(s) = \frac{k_{\rm on} CS(s)}{L(C(s + k_{\rm on}) - k_{\rm on}k_{\rm off}) - k_{\rm on}k_{\rm off}MS(s)}, \tag{20}$$

where

$$M = k_1 - k_t - \frac{k_1 k_2}{k_2 + s}, \quad C = s + k_1 + k_{\rm off} - \frac{k_1 k_2}{k_2 + s}. \tag{21}$$

Also, we have

$$S(s) = \frac{k_t y^{1+m}(1 + y)(y^{2L} - 1)}{(y - 1)\theta(y^{2m} + y^{2L+1})}, \tag{22}$$

with

$$\theta = (s + 2u + k_{\rm off} + k_t)(y^m + y^{1-m}) - u\left(y^{m-1} + y^{2-m} + \frac{y^{2m+1} + y^2 + y^{2L} + y^{2L-2m+1}}{y^m + y^{2L+1-m}}\right). \tag{23}$$

It can be shown that the parameter y is equal to

$$y = \frac{s + 2u + k_{\rm off} + k_1 - \frac{k_1 k_2}{k_2 + s} - \sqrt{\left(s + 2u + k_{\rm off} + k_1 - \frac{k_1 k_2}{k_2 + s}\right)^2 - 4u^2}}{2u}. \tag{24}$$

The final expression for the mean search time is given by

$$T_0 = \frac{Lk_{\rm off} + k_{\rm on}\left(1 + \frac{k_1}{k_2}\right)(L - S(0)) - \frac{k_{\rm on}k_{\rm off}k_1}{k_2 k_t}S(0)}{k_{\rm on}k_{\rm off}S(0)}. \tag{25}$$

One can see that in the limit of very large transition rates into the search mode from the recognition mode ($k_2 \gg 1$) this expression reduces to the mean search time for the model with the conformational transition only at the target as presented in equation (19).

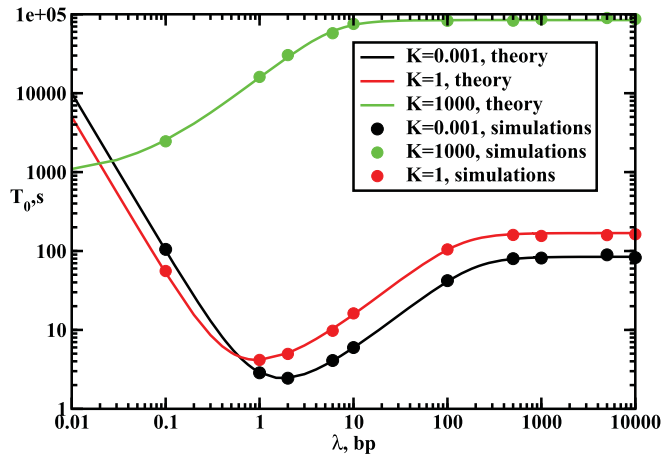

Figure 4. Dynamic phase diagram for the protein search with conformational transition at all sites and sliding in the search mode. Mean search times as a function of the scanning length are shown for different affinities K for the recognition conformational states (defined as $K = k_1/k_2$). The scanning length λ (from equation (26)) is varied by changing k_{off}. The parameters used for calculations are the following: $L = 1000$, $m = L/2$, $k_{\text{on}} = 10^3\ \text{s}^{-1}$, $u = 10^3\ \text{s}^{-1}$ and $k_1 = k_t = 10^3\ \text{s}^{-1}$. Curves are theoretical predictions, symbols are from simulations.

The important length scale for the protein search is the average scanning length λ. For this model, it is defined as an average distance that the protein moves along the DNA chain in *any* conformation, and it can be written as

$$\lambda = \sqrt{\frac{u}{k_{\text{off}}}\frac{k_2}{(k_1 + k_2)}}. \tag{26}$$

The physical meaning of this expressions is the following. Because there is no sliding in the recognition mode, there is a conformational equilibrium at every site on DNA. Then for the protein bound to DNA, the probability to be found in the search mode is equal to $\frac{k_2}{k_1 + k_2}$, and the effective diffusion rate on DNA is equal to

$$u_{eff} = u\frac{k_2}{(k_1 + k_2)}. \tag{27}$$

The scanning length is related to the diffusion and dissociation rates as $\lambda = \sqrt{u_{\text{eff}}/k_{\text{off}}}$, and this leads to equation (26).

Dynamic phase diagrams for the model with conformational transitions and sliding in the search mode are presented in figure 4. One can see that again there are three search regimes depending on the relative values of the relevant length scales in the system. For $\lambda > L$, the protein is almost always on DNA in one of two possible conformations. For $1 < \lambda < L$ the protein does multiple search cycles consisting of attaching to DNA in the search mode, occasionally reversibly transitioning into the recognition mode and moving the average distance λ along DNA in the search mode, and dissociating back into the solution. It is important to note here that the protein can scan the region where the target is located, but it might not find the target if it does not transition into the recognition mode. For $\lambda < 1$, the protein does not slide on DNA after binding to it, and the search process involves multiple

associations to DNA, reversible conformation transitions, and dissociations back into the solution.

Figure 4 also shows the effect of conformational transitions on the protein search dynamics. If we define an affinity to the recognition conformation as $K = k_1/k_2$, then our calculations show that increasing K slows down the search for not very small λ. This is consistent with a so-called speed-affinity trade-off observed recently in computational and experimental studies [9]. It was argued that the protein which is bound stronger to DNA in the recognition mode has a slower target search kinetics. This is a physically reasonable result since the protein is essentially trapped in the recognition state and this lowers its searching efficiency [9]. However, figure 4 reveals also that speed-affinity trade-off is not observed for all search regimes. The calculations indicate that for small scanning lengths the trend is reversed: stronger affinity is associated with faster search kinetics. To understand this surprising result, we note that $\lambda < 1$ corresponds to large dissociation rates $k_{\rm off}$. So the protein in the search mode will be frequently removed into the solution, while in the recognition mode this does not happen. As a result, it is faster to find the specific site if the recognition conformations are preferred.

2.3. Conformational transitions at all sites and sliding in the recognition mode

Here we consider a different model in which the conformational transitions can take place at all sites on DNA but the sliding is possible only in the chain 2 (the recognition mode). The model is illustrated in figure 1(c). All transitions rates are the same as in the model on figure 1(b), except that the diffusion rate in the recognition mode is now equal to w, and there is no sliding in the search mode ($u = 0$).

Using the first-passage method outlined above, the dynamics of the protein search in this model can be also calculated explicitly. The results are the following: the Laplace transform of the first-passage probability function of starting in the solution is equal to

$$\tilde{F}_0(s) = \frac{k_1 k_{\rm on} S(s)}{L(Ms + Mk_{\rm on} - k_{\rm on}k_{\rm off}) - \frac{k_1 k_2 k_{\rm on} k_{\rm off}(L - S(s))}{MC}}, \tag{28}$$

where

$$M = s + k_1 + k_{\rm off}, \quad C = s + k_2 - \frac{k_1 k_2}{M}; \tag{29}$$

and

$$S(s) = \frac{y^{1+2m}(1 + y)(1 - y^{2L})}{(1 - y)(y + y^{2m})(y^{1+2L} + y^{2m})}. \tag{30}$$

The parameter y is found to be

$$y = \frac{C + 2w - \sqrt{C^2 + 4wC}}{2w}. \tag{31}$$

Then, the expression for mean search time can be written as

$$T_0 = \frac{(k_1 + k_{\rm on} + k_{\rm off})L + (k_1 + k_2 + k_{\rm off})\frac{k_1 k_{\rm on}}{k_2 k_{\rm off}}(L - S(0))}{k_1 k_{\rm on} S(0)}. \tag{32}$$

For this model, we can estimate the average scanning length λ of the protein molecule on DNA in any configuration using the following arguments. Generally, there is no

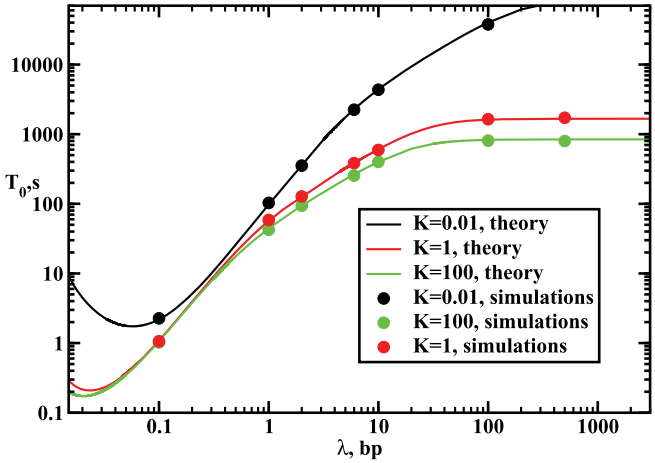

Figure 5. Dynamic phase diagram for the protein search with conformational transition at all sites and sliding in the recognition mode. Mean search times as a function of the scanning length are shown for different affinities K for the recognition conformational states (defined as $K = k_1/k_2$). The scanning length λ (from equation (34)) is varied by changing k_{off}. The parameters used for calculations are the following: $L = 100$, $m = L/2$, $k_{\text{on}} = 10^3\ \text{s}^{-1}$, $w = 1\ \text{s}^{-1}$ and $k_2 = 10^4\ \text{s}^{-1}$. Curves are theoretical predictions, symbols are from simulations.

conformational equilibrium between the states in the chain 1 and 2. The protein moves along the DNA chain only in the recognition mode with the rate w, so the average scanning length corresponds to the average sliding length on DNA while being only in the recognition mode. It can be written as $\lambda = \sqrt{w/k_{\text{off,eff}}}$ where $k_{\text{off,eff}}$ is the effective dissociation rate into the solution. It can be shown using the first passage arguments that [47]

$$k_{\text{off,eff}} = \frac{k_2 k_{\text{off}}}{(k_1 + k_2 + k_{\text{off}})}. \tag{33}$$

Then the average scanning length is given by

$$\lambda = \sqrt{\frac{w}{k_{\text{off}}} \frac{(k_1 + k_2 + k_{\text{off}})}{k_2}}. \tag{34}$$

In the limit of very large transition rates k_2 this expression simplifies into the expected result $\lambda = \sqrt{w/k_{\text{off}}}$, since in this case the unbinding rate from DNA is equal to k_{off}. At another limit, $k_{\text{off}} \gg 1$, the dissociation into the solution is essentially given just by the rate k_2 and the scanning length is equal to $\lambda = \sqrt{w/k_2}$.

Figure 5 presents the dynamic phase diagram for the system with conformational transitions at all sites and sliding in the recognition mode. Here the mean search times as a function of λ are calculated for various values of the conformational transition rates. Three dynamic search regimes can be realized depending on the relative values of the relevant length scales in the system (the size of DNA, the scanning length and the target size). In this model, in contrast to the model with sliding only in the search mode (figure 1(b)), increasing the affinity $K = k_1/k_2$ for the recognition state accelerates the search in all regimes, and the speed-affinity trade-off is never observed. One should notice that this model is probably less realistic in the description of the protein search dynamics.

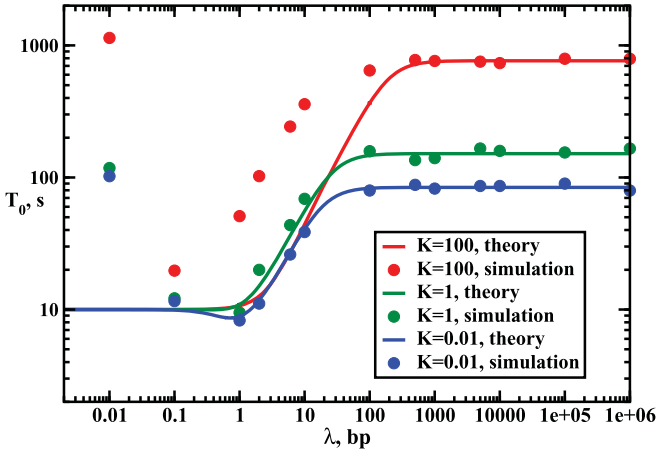

Figure 6. Dynamic phase diagram for the protein search with conformational transition at all sites and sliding in both recognition and search modes. Mean search times as a function of the scanning length are shown for different affinities K for the recognition conformational states (defined as $K = k_1/k_2$). The scanning length λ (from equation (37)) is varied by changing k_{off}. Symbols correspond to Monte Carlo computer simulations, and curves are theoretical predictions as explained in the text. The parameters used for calculations are the following: $L = 100$, $m = L/2$, $k_{\text{on}} = 10$ s^{-1}, $u = 10$ s^{-1}, $w = 1$ s^{-1} and $k_1 = 10^4$ s^{-1}.

2.4. Conformational transitions at all sites and sliding in both recognition and search modes

Finally, we consider a general model with conformational transitions at all sites and with sliding in both recognition and search states, as presented in figure 1(d). From the solution the protein can bind to DNA only in the search conformation, in which it can slide with the diffusion rate u. In the recognition mode, the protein slides with the rate w. The transition rates k_1 and k_2 describe the conformational transitions between two protein–DNA binding modes: see figure 1(d).

We were not able to solve this model for general sets of conditions, and for this reason it was mostly analyzed using Monte Carlo computer simulations. At the same time, it should be mentioned that our analytical solutions for simpler systems presented above in figures 1(b) and (c) with sliding in only one of the conformational states can approximate the more general model for corresponding conditions. For example, for $w \ll k_2$ the sliding motion in the recognition mode is negligible in comparison with the conformation transition into the search state, and our results for the model in figure 1(b) provide an excellent description for these parameters (results are not shown). Also for $u \ll (k_{\text{off}} + k_1)$ we have the explicit expressions for the dynamic properties of the system (results are also not shown). This is because in this case the sliding in the search mode is small, and our analysis for the model in figure 1(c) can be directly utilized. These arguments suggest that all dynamic features of the protein search that we already discussed for simpler models are also valid for this more general system with conformational fluctuations and diffusion in all conformational states.

Furthermore, explicit description of the dynamics of the protein search can be obtained in the situation with the conformational equilibrium between the search and recognition modes. The dynamic phase diagram for this case is shown in figure 6. In this equilibrium limit, one can define the effective diffusion rate along the DNA chain, which is an average over two

possible conformational modes

$$u_{\text{eff}} = u\frac{k_2}{k_1 + k_2} + w\frac{k_1}{k_1 + k_2}. \tag{35}$$

Similarly, the effective dissociation rate can be written as

$$k_{\text{off,eff}} = k_{\text{off}}\frac{k_2}{k_1 + k_2}. \tag{36}$$

Then the average scanning λ for this model is given by

$$\lambda = \sqrt{\frac{u\frac{k_2}{k_1+k_2} + w\frac{k_1}{k_1+k_2}}{k_{\text{off}}\frac{k_2}{k_1+k_2}}} = \sqrt{\frac{u + wK}{k_{\text{off}}}}, \tag{37}$$

with $K = k_1/k_2$. Then this system can be mapped into the problem of the protein search for targets on DNA with a single non-specifically bound state with the diffusion rate u_{eff}, the dissociation rate $k_{\text{off,eff}}$ and the binding rate k_{on}, which has been explicitly analyzed before [19]. This is also identical to analytical results obtained for the model in figure 1(a) with $k_t \to \infty$ (see equation (19)). Theoretical predictions from this approximate description are given in figure 6. There are three dynamic search regimes observed, as expected. One can also see that our conformational equilibrium approximation works quite well in the regime of $\lambda > L$ because in this case the protein is mostly bound to DNA in one of two possible conformations. It has enough time to equilibrate between the search and the recognition modes, justifying our assumptions. The approximation is reasonable even for intermediate scanning lengths λ, and it clearly breaks for small λ, which corresponds to large k_{off}, because the system spends significant amount of time outside of DNA and the conformational equilibrium cannot be reached fast.

In figure 6 the mean search times as a function of the length λ are presented for different affinities $K = k_1/k_2$. The phenomenon of speed-affinity trade off is again observed here. Shifting the conformational equilibrium to the recognition conformations (increasing K) slows down the protein search. This is because the protein moves slowly in the recognition states, and for large K it is effectively trapped in this slow mode, which makes the search inefficient. Lowering K allows the protein to explore better different regions along the DNA chain, accelerating the search kinetics.

3. Summary and conclusions

We developed a theoretical description that helped us to clarify the role of conformational transitions during the protein search for targets on DNA. By utilizing the method of first-passage processes dynamic properties for various systems with conformational fluctuations were obtained explicitly. To clarify the contributions of the conformational fluctuations several models with different contributions from conformational transitions and diffusional sliding along the DNA chain were analyzed. Our calculations show that the presence of conformations do not change the number of search regimes which is determined by the relative values of DNA length, the average scanning length of the non-specifically bound protein and the size of the target. But the conformational transitions might strongly affect the search dynamics.

In the model where the conformational change is taking place only at the target, increasing the conformational rate lowers the search time and it makes the dynamics less

dependent on the position of the specific site. The dynamical behavior is more complex in the system with the conformational transitions at all sites and the sliding in the search mode. It is found that shifting the conformational equilibrium into the recognition mode, in which the target can be recognized, slows down the search dynamics for intermediate and large values of the average scanning length. This is known as the speed-affinity trade-off when larger equilibrium constants into the recognition states is associated with larger mean search times, and it is consistent with available experimental and computational observations [9]. However, there is a surprising new result from our calculations, suggesting that for small scanning lengths the trend is reversed and there will be no speed-affinity trade-off. It is argued that this happens due to frequent dissociations into the solution from the search mode at these conditions. Similar behavior is observed if the diffusion can happen only in the recognition state: the larger affinity leads to faster search kinetics at all ranges of parameters. This can be explained by noting that transition into the search mode will not help to explore more regions on DNA, and thus staying in the recognition conformation is more beneficial for the search. Finally, we analyzed the general model of the protein search with conformational fluctuations and diffusion in all conformational states by using Monte Carlo simulations. It is shown that when the system spends most of the time being on DNA in one of two possible conformations, the equilibrium between search and recognition modes can be achieved. In this case, the dynamics can be successfully mapped into the protein search on DNA without conformational transitions for which the explicit expressions for all dynamic properties are well known.

Although our theoretical analysis provides clear physical picture on the effect of conformational dynamics in the protein search for targets on DNA, one should note that our approach is still rather simplified and many realistic processes are not taken into account. They include the sequence dependence in the protein–DNA interactions, the role of DNA conformations, the size of the specific regions and many others. It will be important to test the proposed theoretical ideas in experiments as well as in more advanced theoretical and computational 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 and Walter P 2007 *Molecular Biology of the Cell* 5th edn (New York: Garland Science)
[2] Phillips R, Kondev J and Theriot J 2012 *Physical Biology of the Cell* 2nd edn (New York: Garland Science)
[3] Berg O G, Winter R B and von Hippel P H 1981 *Biochemistry* **20** 6929
[4] Winter R B, Berg O G and von Hippel P H 1981 *Biochemistry* **20** 6961
[5] Berg O G and von Hippel P H 1985 *Annu. Rev. Biophys. Biophys. Chem.* **14** 131
[6] Tafvizi A, Huang F, Leith J S, Fersht A R, Mirny L A and van Oijen A M 2008 *Biophys. J.* **95** L01
[7] Tafvizi A, Huang F, Fersht A R, Mirny L A and van Oijen A M 2011 *Proc. Natl Acad. Sci. USA* **108** 563
[8] Tafvizi A, Mirny L A and van Oijen A M 2011 *Chem. Phys. Chem.* **12** 1481
[9] Zandarashvili L, Esadze A, Vuzman D, Kemme C A, Levy Y and Iwahara J 2015 *Proc. Natl Acad. Sci. USA* **112** E5142
[10] Halford S E and Marko J F 2004 *Nucleic. Acid Res.* **32** 3040

[11] Gowers D M, Wilson G G and Halford S E 2005 *Proc. Natl Acad. Sci. USA* **102** 15883
[12] Ivanov K L and Lukzen N N 2014 *Chem. Phys.* **435** 14
[13] Wang Y M, Austin R H and Cox E C 2006 *Phys. Rev. Lett.* **97** 048302
[14] Elf J, Li G-W and Xie X S 2007 *Science* **316** 1191
[15] Benichou O, Kafri Y, Sheinman M and Voituriez R 2009 *Phys. Rev. Lett.* **1003** 138102
[16] Hu T, Grosberg A Y and Shklovskii B I 2006 *Biophys. J.* **90** 2731–44
[17] Givaty O and Levy Y 2009 *J. Mol. Biol.* **385** 1087
[18] Kolomeisky A B and Veksler A 2012 *J. Chem. Phys.* **136** 125101
[19] Veksler A and Kolomeisky A B 2013 *J. Phys. Chem.* B **117** 12695
[20] Shvets A A and Kolomeisky A B 2016 *J. Phys. Chem. Lett.* **7** 2502
[21] Hilario J and Kowalczykowski S C 2009 *Curr. Opin. Chem. Biol.* **14** 15
[22] Coppey M, Benichou R, Voituriez R and Moreau M 2004 *Biophys. J.* **87** 1640
[23] Hammar P, Leroy P, Mahmutovic A, Marklund E G, Berg O G and Elf J 2012 *Science* **336** 1595
[24] Koslover E F, Diaz de la Rosa M A and Spakowitz A J 2011 *Biophys. J.* **101** 856
[25] de la Rosa M A D, Koslover E F, Mulligan P J and Spakowitz A J 2010 *Biophys. J.* **98** 2943
[26] Marklund E G, Mahmutovic A, Berg O G, Hammar P, van der Spoel D, Fange D and Elf J 2013 *Proc. Natl Acad. Sci. USA* **110** 19796
[27] Brackley C A, Cates M E and Marenduzzo D 2012 *Phys. Rev. Lett.* **109** 168103
[28] Brackley C A, Cates M E and Marenduzzo D 2013 *Phys. Rev. Lett.* **111** 108111
[29] Markovitz A and Levy Y 2013 *Biophys. J.* **104** 2042
[30] Hu L, Grosberg A Y and Bruinsma R 2008 *Biophys. J.* **95** 1151
[31] Zhou H-X 2011 *Proc. Natl Acad. Sci. USA* **108** 8651
[32] Zandarashvili L, Vuzman D, Esadze A, Takayama Y, Sahu D, Levy Y and Iwahara J 2012 *Proc. Natl Acad. Sci. USA* **109** E1724
[33] Gorman J and Greene E C 2008 *Nat. Struct. Mol. Biol.* **15** 768
[34] Slutsky M and Mirny L A 2004 *Biophys. J.* **87** 4021
[35] Cuculis L, Abil Z, Zhao H and Schroeder C M 2015 *Nat. Commun.* **6** 7277
[36] Mirny L A, Slutsky M, Wunderlich Z, Tafvizi A, Leith J S and Kosmrlj A 2009 *J. Phys. A: Math. Theor.* **42** 434013
[37] Bauer M and Metzler R 2012 *Biophys. J.* **102** 2321
[38] Kolomeisky A B 2011 *Phys. Chem. Chem. Phys.* **13** 2088
[39] Lange M, Kochugaeva M and Kolomeisky A B 2015 *J. Chem. Phys.* **143** 105102
[40] Lange M, Kochugaeva M and Kolomeisky A B 2015 *J. Phys. Chem.* B **119** 12410
[41] Bauer M, Rasmussen E S, Lomholt M A and Metzler R 2015 *Sci. Rep.* **5** 10072
[42] Mechetin G V and Zharkov D O 2014 *Biochemistry* **79** 496
[43] Chu X, Liu F, Maxwell B A, Wang Y, Suo Z, Wang H, Han W and Wang J 2014 *PLoS Comput. Biol.* **10** e1003804
[44] Reingruber J and Holcman D 2011 *Phys. Rev.* E **84** 020901
[45] Esadze A, Kemme C A, Kolomeisky A B and Iwahara J 2014 *Nucleic Acids Res.* **42** 7039
[46] Gilmore J L, Suzuki Y, Tamulaitis G, Siksnys V, Takeyasu K and Lyubchenko Y L 2009 *Biochemistry* **48** 10492
[47] Redner S 2007 *A Guide to First-Passage Processes* 1st edn (Cambridge: Cambridge University Press)