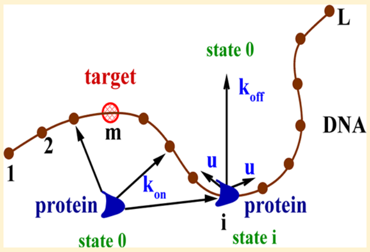
Speed-Selectivity Paradox in the Protein Search for Targets on DNA: Is It Real or Not?

Alex Veksler and Anatoly B. Kolomeisky*

Department of Chemistry, Rice University, Houston, Texas 77005, United States



ABSTRACT: Protein search for targets on DNA starts all major biological processes. Although significant experimental and theoretical efforts have been devoted to investigation of these phenomena, mechanisms of protein−DNA interactions during the search remain not fully understood. One of the most surprising observations is known as a speed-selectivity paradox. It suggests that experimentally observed fast findings of targets require smooth protein−DNA binding potentials, while the stability of the specific protein−DNA complex imposes a large energy gap which should significantly slow down the protein molecule. We developed a discrete-state stochastic approach that allowed us to investigate explicitly target search phenomena and to analyze the speed-selectivity paradox. A general dynamic phase diagram for different search regimes is constructed. The effect of the target position on search dynamics is investigated. Using experimentally observed parameters, it is found that slow protein diffusion on DNA does not lead to an increase in the search times. Thus, our theory resolves the speed-selectivity paradox by arguing that it does not exist. It is just an artifact of using approximate continuum theoretical models for analyzing protein search in the region of the parameter space beyond the range of validity of these models. In addition, the presented method, for the first time, provides an explanation for fast target search at the level of single protein molecules. Our theoretical predictions agree with all available experimental observations, and extensive Monte Carlo computer simulations are performed to support analytical calculations.

INTRODUCTION

Protein molecules are major players in all living systems, and through interactions with DNA, they support essentially all cellular activities. Protein search for target sites on DNA plays a central role in these interactions because most biological processes start when some protein molecules bind to specific sequences on DNA, initiating cascades of biochemical reactions that control them.[1−3] This fundamental aspect of protein−DNA interactions has been investigated extensively in the last 40 years by utilizing various experimental[4−29] and theoretical methods.[10,21,30−59] Although many features of protein search on DNA have been uncovered, mechanisms of these phenomena are still not well understood and it remains a controversial problem.[21,51,56,57]

A large amount of experimental observations suggests that many proteins find their targets on DNA very quickly and efficiently, and frequently the search times are much shorter than expected from limiting 3D diffusion estimates.[4,10,21,51] For example, chemical kinetic measurements on association rates of *lac* repressor proteins to specific target sequences on DNA yielded a rate constant $k_{\text{exp}} \simeq 10^{10}$ M^{-1} s^{-1}, which is approximately 100−1000 times larger than expected from maximal values for chemical rates as specified by Debye−Smoluchowski 3D diffusion theory.[21,51,56] This is known as a *facilitated diffusion* phenomenon in the protein search. Recent experimental evidence, coming mostly from single-molecule experiments that can now visualize the dynamics of single protein molecules,[10,11,14,15,24,25] indicates that the protein search is a complex dynamic process that couples 3D solution diffusion with 1D sliding of proteins bound nonspecifically to DNA, as schematically shown in Figure 1. Then, following the classical works of Berg, Winter, and von Hippel,[5−7] the total search time could be estimated as

$$\tau_{\text{s}} = \frac{L}{\lambda}(\tau_{1\text{D}} + \tau_{3\text{D}}) \tag{1}$$

with $\tau_{1\text{D}} = \lambda^2/2D_1$ and $\tau_{3\text{D}} = x^2/2D_3$, where L is the total contour length of DNA, λ is the average length of DNA that the protein molecule scans during each search cycle, x is the average distance traveled by the protein in the solution before binding to DNA, D_1 is a diffusion constant to move along the DNA, and D_3 is protein's bulk diffusion constant.[21,51]

Although current theoretical approaches have been able to explain some features of the protein search dynamics, there is an increasing number of experimental and theoretical studies that challenge existing views.[21,51,57,59] One of the most surprising and controversial observations related to the protein search is known as a "speed-selectivity paradox".[21,31,57] Since the DNA molecule is a heteropolymer, binding energies of protein molecules to the DNA chain depend on underlying sequences. The magnitude of this dependence can be described by the standard deviation of the energy distribution, σ.[21,31] It has been shown that sequence-dependent free-energy fluctuations significantly slow down protein diffusion on DNA because

Special Issue: Peter G. Wolynes Festschrift

Received: November 20, 2012
Revised: January 4, 2013
Published: January 14, 2013

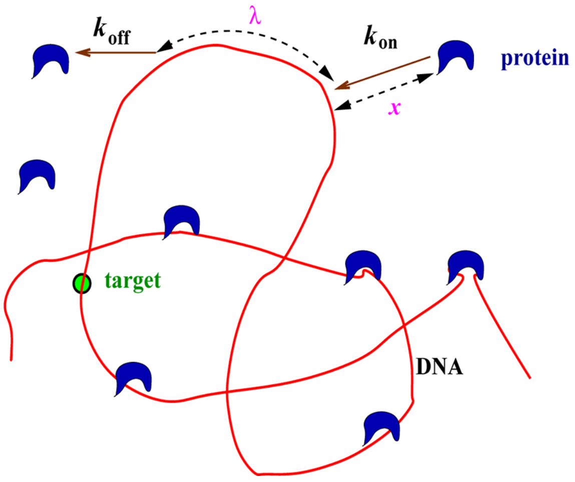


Figure 1. A general view of the protein search for targets on DNA. A protein molecule diffuses an average distance x in the solution, then binds to DNA with the rate k_{on}, and after scanning an average distance λ it dissociates with the rate k_{off}. The process is complete when any of the searching proteins can bind to the specific target site.

in this case the protein molecule moves in a random potential[21,31,60]

$$D_1 \simeq \exp\left[-\left(\frac{\sigma}{k_B T}\right)^2\right] \tag{2}$$

Fast searching rates are possible only for quick protein sliding that requires a smooth potential with $\sigma < 1-2$ k_BT, while the stability of the protein−DNA complex imposes the condition that $\sigma > 5$ k_BT. The association energy to a specific target sequence should be much larger than that to nonspecific sites to keep the protein molecule bound to DNA longer in order for other biochemical and biophysical processes to have enough time to be accomplished. To resolve this paradox, it has been proposed that during the search the protein molecule bound to DNA might exist in two different conformational states: one is a search mode where the protein quickly slides along the DNA, and another one is a slow recognition mode where the protein checks for a target sequence.[21,31,56,57,59] Multiple protein−DNA binding conformations have been observed in recent experiments.[17,24,27] However, the two-state approach to explain the speed-selectivity paradox is not without problems. In various presentations of the two-state model, the recognition state is always assumed to have a higher free energy than the searching state.[21,31,56,59] This is a surprising and probably counterintuitive assumption, since in the recognition state protein binds stronger to DNA and the electrostatic attractive contributions to the free energy are much larger in comparison with the weakly bound searching state, suggesting lower free energy for this state. Furthermore, fast search times observed in experiments require very high conformational switching rates between these two states ($>10^3$ s^{-1}), but much slower conformational changes are observed so far in the single-molecule experiments, $\simeq 1$ s^{-1}, at least for some systems.[24]

The speed-selectivity paradox is also related to another problem of the current theoretical views on protein search phenomena associated with the use of mostly continuum models for description of intrinsically discrete biochemical processes (binding/unbinding and hopping along the DNA sites). One could easily see this by analyzing eq 1 in the limit of very small diffusion constant on DNA, $D_1 \to 0$, which is the case for many experimental systems.[28] It predicts then the infinite search time, which is unphysical since the protein still can find the target site via 3D diffusion, and the search time must be finite. This is the result of the fact that continuum models are approximate and they cannot be used beyond their range of validity, which is defined for the case when the scanning length is much larger than the target size, $\lambda \gg a$, so that the effects of discreteness might be neglected. As a result, *none of current theoretical models* can explain fast protein search at the level of one or few protein molecules, the conditions which are frequently observed for both *in vitro* and *in vivo* systems.

In this paper, we present a simple physical-chemical approach based on discrete-state stochastic models that allows us to describe explicitly the protein search for targets on DNA. The method is applied to test the origins of the speed-selectivity paradox, and the analysis shows that this paradox is not real, since it is an artifact of the continuum approximation utilized in previous theoretical models. The presented theoretical method provides a first quantitative explanation of the fast target search on DNA at the level of single protein molecules. Our theoretical calculations agree with available experimental observations, and they are also fully supported by Monte Carlo computer simulations.

■ THEORETICAL METHODS

Since the continuum description of the protein search dynamics might lead to erroneous predictions, as has been argued above, a simple discrete-state stochastic model shown in Figure 2 is developed. We consider a single protein molecule that can be found in the solution, or it can associate to any state i on the DNA ($i = 1, 2, ..., L$). One of the binding sites ($i = m$) is a target for the search. On the DNA, the protein can hop along the chain with a diffusion rate u with equal probability in both directions (see Figure 2). The protein might also dissociate from the DNA with the rate k_{off}. Because 3D solution diffusion is typically fast in comparison with 1D sliding (which means that the time to diffuse the volume around the DNA molecule is shorter than the time for protein bound to DNA to slide along its contour),[14,15,51] we combine all solution states into the one (state 0), and the equal probability to reach any site on DNA from the solution is assumed (with the total rate of binding to DNA given by k_{on}); see Figure 2. The transition

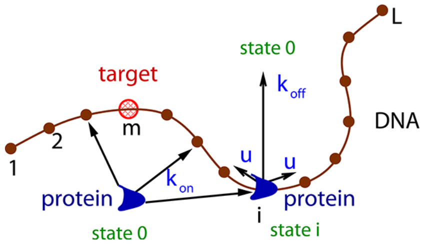


Figure 2. A general scheme of the discrete-state stochastic model for the protein search. The DNA chain has $L - 1$ nonspecific binding sites and one specific site, which is the target of the protein search. A protein molecule can diffuse along the DNA chain with the rate u, and it can dissociate into the solution with the rate k_{off}. From the solution, the protein can bind to any site on the DNA with equal probability, and the total association rate is equal to k_{on}. The search is finished when the protein binds to the target site at the position m.

rates u, k_{on}, and k_{off} can be determined from experimental measurements of protein search dynamics. For example, for *lac* repressor proteins,[14,15] the estimates for these rates yield $u \simeq 10^3$–10^6 s^{-1}, $k_{\mathrm{off}} \simeq 200$–$3000$ s^{-1}, and $k_{\mathrm{on}} \simeq 10^4$ –10^6 s^{-1}. This is probably the simplest model that captures main features of the protein search for targets on DNA. A similar discrete-state model has been discussed recently for analyzing general intermittent search problems.[61]

To describe the target search dynamics, we introduce a function $F_n(t)$, which is defined as a probability to reach the target on site m at time t for the first time if at $t = 0$ the protein was at the state n ($n = 0, 1, ..., L$). The temporal evolution of these first-passage probabilities follows the backward master equations[58,62]

$$\frac{\mathrm{d}F_n(t)}{\mathrm{d}t} = u[F_{n+1}(t) + F_{n-1}(t)] + k_{\mathrm{off}}F_0(t) - (2u + k_{\mathrm{off}})F_n(t) \tag{3}$$

for $2 \leq n \leq L - 1$, while for sites at the DNA ends ($n = 1$ and $n = L$) we have

$$\frac{\mathrm{d}F_1(t)}{\mathrm{d}t} = uF_2(t) + k_{\mathrm{off}}F_0(t) - (u + k_{\mathrm{off}})F_1(t) \tag{4}$$

$$\frac{\mathrm{d}F_L(t)}{\mathrm{d}t} = uF_{L-1}(t) + k_{\mathrm{off}}F_0(t) - (u + k_{\mathrm{off}})F_L(t) \tag{5}$$

The backward master equation is different if the protein molecule starts from the solution, $n = 0$,

$$\frac{\mathrm{d}F_0(t)}{\mathrm{d}t} = \frac{k_{\mathrm{on}}}{L}\sum_{n=1}^{L} F_n(t) - k_{\mathrm{on}}F_0(t) \tag{6}$$

Note that in this equation we used the fact that the rate to bind to any given site on DNA is k_{on}/L, and the total rate of binding to DNA is equal to k_{on}. In addition, initial conditions require that $F_m(t) = \delta(t)$ and $F_{n\neq m}(t = 0) = 0$. These equations can be analyzed by introducing Laplace transformations of first-passage probability functions, $\widetilde{F_n(s)} \equiv \int_0^{\infty} \mathrm{e}^{-st}F_n(t)\,\mathrm{d}t$. Then, backward master equations (eqs 3, 4, and 6) can be written as a set of simpler algebraic expressions

$$(s + 2u + k_{\mathrm{off}})\widetilde{F_n(s)} = u[\widetilde{F_{n+1}(s)} + \widetilde{F_{n-1}(s)}] + k_{\mathrm{off}}\widetilde{F_0(s)} \tag{7}$$

$$(s + u + k_{\mathrm{off}})\widetilde{F_1(s)} = u\widetilde{F_2(s)} + k_{\mathrm{off}}\widetilde{F_0(s)} \tag{8}$$

$$(s + u + k_{\mathrm{off}})\widetilde{F_L(s)} = u\widetilde{F_{L-1}(s)} + k_{\mathrm{off}}\widetilde{F_0(s)} \tag{9}$$

$$(s + k_{\mathrm{on}})\widetilde{F_0(s)} = \frac{k_{\mathrm{on}}}{L}\sum_{n=1}^{L}\widetilde{F_n(s)} \tag{10}$$

These equations are solved by assuming that the general form of the solution is $\widetilde{F_n(s)} = Ay^n + B$, and using boundary and initial conditions it yields

$$\widetilde{F_n(s)} = \frac{(1 - B)(y^n + y^{-n})}{y^m + y^{-m}} + B \tag{11}$$

for $1 \leq n \leq m$ and

$$\widetilde{F_n(s)} = \frac{(1 - B)(y^{1+L-n} + y^{n-L-1})}{y^{1+L-m} + y^{m-L-1}} + B \tag{12}$$

for $m \leq n \leq L$. Here, parameters y and B are given by

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

$$B = \frac{k_{\mathrm{off}}\widetilde{F_0(s)}}{(k_{\mathrm{off}} + s)} \tag{14}$$

From eq 10, one can also show that

$$\widetilde{F_0(s)} = \frac{k_{\mathrm{on}}(k_{\mathrm{off}} + s)S(s)}{Ls(k_{\mathrm{off}} + k_{\mathrm{on}} + s) + k_{\mathrm{off}}k_{\mathrm{on}}S(s)} \tag{15}$$

where the new auxiliary function $S(s)$ is given by

$$S = \frac{(1 + y)(1 - y^{L+1})}{(1 - y)(1 + y^m)(1 + y^{L+1-m})} \tag{16}$$

Explicit analytical expression for first-passage probability distribution functions in the Laplace form allows us to obtain a full dynamic description of the search process. More specifically, first-passage times to reach the target at the site m starting with equal probability on any site on the DNA chain can be computed from

$$T_m \equiv -\frac{1}{L}\frac{d}{ds}\sum_{n=1}^{L}\tilde{F}_n(s)\Bigg|_{s=0} \tag{17}$$

producing a simple expression

$$T_m = \frac{(k_{\mathrm{on}} + k_{\mathrm{off}})}{k_{\mathrm{on}}k_{\mathrm{off}}}\frac{(L - S(0))}{S(0)} \tag{18}$$

The average time to find the target starting from the solution, T_0, can be easily found using the following equality

$$\begin{aligned} T_0 &= -\frac{\partial\widetilde{F_0(s)}}{\partial s}\Bigg|_{s=0} \\ &= \frac{k_{\mathrm{off}}L + k_{\mathrm{on}}(L - S(0))}{k_{\mathrm{on}}k_{\mathrm{off}}S(0)} \\ &= T_m + 1/k_{\mathrm{on}} \end{aligned} \tag{19}$$

Any other dynamic properties of the system can be evaluated in a similar way.

■ RESULTS AND DISCUSSION

Target Position. The developed theoretical method allows us to fully explore all features of protein search dynamics. First, we investigate the role of the target position. In Figure 3, relative search times for varying target positions are calculated for different DNA lengths using parameters relevant for *lac* repressor proteins. As expected from the symmetry arguments, the fastest search is achieved when the target is in the middle of the DNA chain, as clearly shown for $L = 11$. However, increasing the length of the DNA segment makes this effect less pronounced, leading to a plateau in the search times (see results for $L = 101$ and 1001). For realistic values of DNA lengths ($L = 10^6$–10^9), one expects that the search becomes effectively independent of the target position as long as the target is not at the end sites. This might have an important biological consequence for the protein search, since it argues that target position is not a critical parameter for understanding search dynamics.

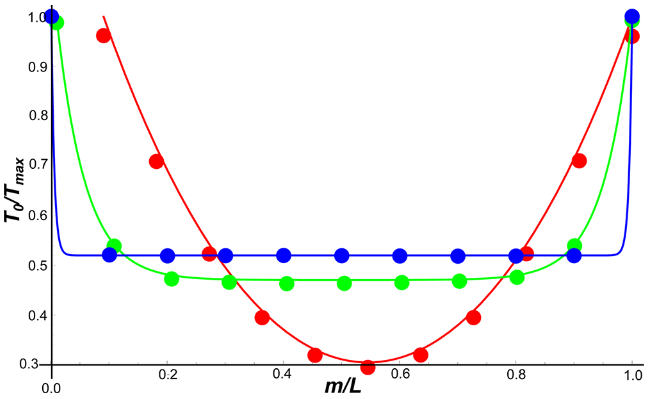


Figure 3. Relative search time as a function of the target position on the DNA chain. Solid curves are analytical results, while symbols are from Monte Carlo computer simulations. The red curve and red symbols correspond to $L = 11$, the green curve and green symbols are for $L = 101$, and the blue curve and blue symbols describe $L = 1001$. The transition rates are $k_{\mathrm{off}} = 10^3\ \mathrm{s}^{-1}$ and $u = k_{\mathrm{on}} = 10^5\ \mathrm{s}^{-1}$.

Dynamic Phase Diagram. Analytical expressions for first-passage probabilities provide a convenient way of analyzing protein search dynamics under different conditions. To understand mechanisms of the target search, we note that there are three length scales in the system: the size of the target a (for simplicity, it is taken to be equal to 1 site in our calculations), the scanning length λ (where $\lambda = (u/k_{\mathrm{off}})^{1/2}$) that corresponds to the average distance that the protein hops along the DNA chain before the detachment, and the length L of the DNA molecule. One could expect that varying scanning length λ as compared with other relevant lengths should lead to different search behaviors in the system. Our explicit calculations support these qualitative predictions, and a full dynamic phase diagram of the protein search for the target on DNA is presented in Figure 4.

Theoretical analysis predicts three dynamic search phases that we label as a random-walk regime, a sliding regime and a jumping regime; see Figure 4. The phase diagram could be understood using the following arguments. If the affinity of DNA for binding the protein molecule is strong enough or protein rapidly hops along the DNA, the scanning length could be larger than the total length of DNA, $\lambda > L$, and the protein will find the target by performing a simple random walk during one search cycle. This dynamic phase is called a random-walk regime (see Figure 4), and the average search time to reach the target should scale quadratically with the length of DNA L. One can show that, for $\lambda \geq L$, we have from eq 16

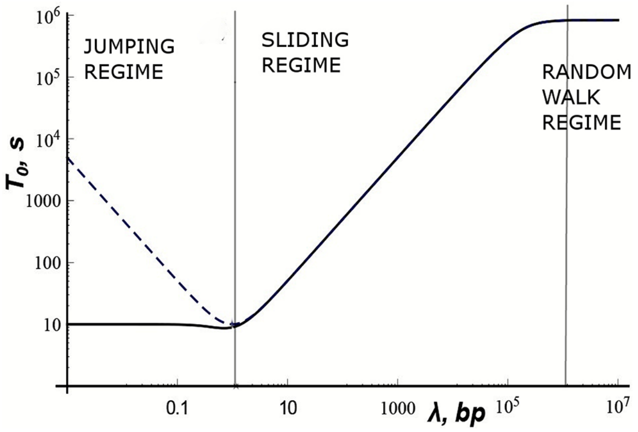


Figure 4. Average times to find the target on DNA for the protein molecule starting from the solution as a function of the scanning length $\lambda = (u/k_{\mathrm{off}})^{1/2}$. The solid curve corresponds to predictions of our discrete-state model (eq 19), and the dashed curve is the result of calculations using a continuum approach (eq 12 from ref 33). The parameters are $L = 10^6$ bp, $u = k_{\mathrm{on}} = 10^5\ \mathrm{s}^{-1}$, and $m = L/2$. The transition rate k_{off} is varied to change λ.

$$S(0) \simeq L + L[6m(1 + L - m) - (1 + 3L + 2L^2)]\lambda^{-2}/6 + O(\lambda^{-3}) \tag{20}$$

leading to $T_0 \propto L^2$. For example, for the target in the middle of the DNA chain, eq 19 yields

$$T_0 \simeq \frac{k_{\mathrm{on}} + k_{\mathrm{off}}}{k_{\mathrm{on}}k_{\mathrm{off}}}\frac{L^2}{12\lambda^2} \tag{21}$$

while for the target position at DNA ends we obtain

$$T_0 \simeq \frac{k_{\mathrm{on}} + k_{\mathrm{off}}}{k_{\mathrm{on}}k_{\mathrm{off}}}\frac{L^2}{3\lambda^2} \tag{22}$$

For constant diffusion rate u, large scanning length λ corresponds to $k_{\mathrm{off}} \ll 1$, and in this case, search times are $T_0 \simeq L^2/12u$ and $T_0 \simeq L^2/3u$ for the target in the middle and at the ends of the DNA chain, respectively.

Another dynamic search phase is observed when the scanning length is smaller than the DNA length but larger than the target size, $1 < \lambda < L$. This is called a sliding regime (Figure 4), and in this case, the protein molecule must dissociate and associate several times before finding the target. The average number of search cycles is given by L/λ and the time for each cycle is independent of the DNA length, suggesting a linear scaling for the total search time as a function of L. The target is found mostly via 1D sliding along the DNA chain. For $\lambda < L$, it can be shown from eq 19 that

$$S(0) \simeq \sqrt{1 + 4\lambda^2} \tag{23}$$

if $1 < m < L$, while for the target position at the ends ($m = 1$ or $m = L$)

$$S(0) \simeq \frac{1 + \sqrt{1 + 4\lambda^2}}{2} \tag{24}$$

Then, the average search times for targets not at the DNA ends can be well approximated as

$$T_0 \approx \frac{k_{\mathrm{on}} + k_{\mathrm{off}}}{k_{\mathrm{on}}k_{\mathrm{off}}}\frac{L}{2\lambda} \tag{25}$$

which for the fixed hopping rate u and $k_{\mathrm{off}} \ll 1$ leads to $T_0 \sim L/(uk_{\mathrm{off}})^{1/2}$.

The third dynamic search phase, called a jumping regime, is observed when the scanning length is smaller than the target size, $\lambda < 1$, so that the protein molecule effectively does not scan neighboring sites on DNA after binding. The target is found after, at average, L binding/unbinding jumping events, and not via 1D sliding as in previous regimes. In this case, it can be shown that

$$T_0 \approx \frac{k_{\mathrm{on}} + k_{\mathrm{off}}}{k_{\mathrm{on}}k_{\mathrm{off}}}L \tag{26}$$

The important observation here is that the search times do not depend on the 1D protein sliding rates. For constant hopping rate u, this regime is achieved when $k_{\mathrm{off}} \gg 1$, leading to $T_0 \simeq L/k_{\mathrm{on}}$. Theoretical calculations also show that the minimal search time is observed in the border area between jumping and sliding search regimes, although for realistic transition rates the minimum is quite shallow.

The developed phase diagram provides a full picture of protein search dynamics, and it should be generally valid for all systems where proteins are finding targets on DNA. One of the interesting predictions that the theoretical model makes is the observation of different scalings for search times as a function of DNA length for different search regimes. It is illustrated more explicitly in Figure 5 where search times are analyzed for various conditions. For these calculations, the constant 1D diffusion rate is assumed, $u = 10^5$ s^{-1}. The red curve and symbols correspond to $k_{\mathrm{off}} = 10^6$ s^{-1} which gives a scanning length less than 1, $\lambda \approx 0.32$. For all values of L, this system is in the jumping regime, and the linear scaling is clearly observed for all ranges of parameters. The green curve and symbols (Figure 5) describe $k_{\mathrm{off}} = 10^3$ s^{-1}, yielding the scanning length $\lambda = 10$. For $L < 10$, the system is in the random-walk regime ($\lambda > L$) with the quadratic scaling, and for $L > 10$, the search follows the sliding regime with linear scaling. More strongly, the crossover from $\sim L^2$ to $\sim L$ behavior is observed for the blue curve and symbols (see Figure 5), where $k_{\mathrm{off}} = 10$ s^{-1} is assumed. Here, the scanning length is equal to $\lambda = 100$. These calculations suggest that in nature tuning binding/unbinding and 1D diffusion rates might provide a convenient tool of modifying protein search dynamics.

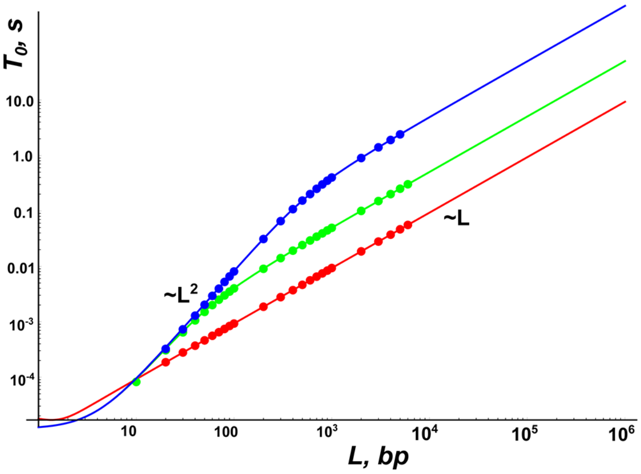


Figure 5. Target search times as a function of the DNA length. For all calculations, the parameters are $u = k_{\mathrm{on}} = 10^5$ s^{-1} and $m = L/2$. The red curve and red symbols correspond to $k_{\mathrm{off}} = 10^6$ s^{-1}, the green curve and green symbols correspond $k_{\mathrm{off}} = 10^3$ s^{-1}, and the blue curve and blue symbols correspond to $k_{\mathrm{off}} = 10$ s^{-1}. Curves are analytical predictions, and symbols are from Monte Carlo computer simulations.

Speed-Selectivity Paradox. The presented theoretical approach, in which all dynamic properties of the search processes can be obtained analytically, allows us also to fully analyze the speed-selectivity paradox. According to this paradox, due to inhomogeneous protein−DNA interactions, the protein molecule must hop along the DNA much slower, leading to shorter scanning distances and very large search times, in contradiction to experimentally observed fast times.[21,31,57] One could argue here that the increase in the search time is due to increasing the number of search cycles ($\simeq L/\lambda$) for smaller λ. Thus, the prediction is that for $\lambda \to 0$ the search time diverges, $T_0 \to \infty$. However, this is not what is observed in our discrete-state stochastic model of the protein search; see Figure 4. For small values of the scanning length ($\lambda < 1$), the search time becomes a constant value independent of the 1D hoping rate u, as indicated in eq 26. This is a physically reasonable result, since in this dynamic regime the protein molecule does not have the time to scan the DNA—it is detached before it can hop to the neighboring DNA site, and only associations and dissociations can be observed. We predict that for experimentally realistic fast binding and unbinding rates, the search rate in this jumping regime can be quite fast, only slightly slower than for the most optimal search conditions, as shown in Figure 4.

It is important to note that our model does not explicitly include heterogeneity effects that lead to eq 2, although it could be done by extending the model to include a random distribution of diffusion rates u. However, the 1D sliding diffusion coefficient D_1 is directly related to the average scanning length, $\lambda \propto (D_1 \tau_{1D})^{1/2}$. This allows us to utilize the scanning length as a critical parameter to characterize the speed-selectivity paradox.

The incorrect predictions of the speed-selectivity paradox are due to using the continuum models to analyze protein search dynamics (see Figure 4). Continuum models[21,31,33,56,59] are widely utilized in studying complex processes associated with the target search on DNA. However, one must remember that they are valid only when the scanning length is significantly larger that the size of the binding site, i.e., $\lambda > 1$. One can see from Figure 4 that in this case predictions of both discrete-state and continuum models fully agree. The problem comes when the continuum model is applied for $\lambda \leq 1$, and this is the source of the speed-selectivity paradox. Thus, this paradox is the artifact of the continuum models for protein search and most probably it does not exist for real systems.

Our theoretical method provides also explicit estimates of the search times for real biological systems. Using transition rates estimated from experiments on *lac* repressor proteins, as explained above, we find that $T_0 \simeq 10{-}50$ s^{-1}. It corresponds to the sliding regime and the border area with the jumping regime, as indicated in the dynamic phase diagram in Figure 4. Most probably, the search times are not minimal for this system, but actual times do not deviate much from the most optimal search conditions. To test our theoretical method, it is important to compare computed quantities with experimentally measured times.[21,51] In our calculations, we assumed a single protein molecule in the volume occupied by a single DNA chain. For the DNA chain of the length $L = 10^6$ bp, one could estimate the Kuhn length b from the persistence length l_p, producing $b = 2l_p = 300$ bp, and the number of the Kuhn segments in DNA N is given by $N = L/b \simeq 3 \times 10^3$. From these data, a radius of gyration of the DNA chain can be estimated, $\langle R_g^2 \rangle = Nb^2/6$, yielding $R_g \simeq 10^4$ bp, which gives the volume occupied by a single DNA molecule, $V = (4\pi R_g^3)/3 \simeq 10^{-13}$ L. The single protein molecule in this volume corresponds to protein concentration in the solution of the order of $c_p \simeq 10^{-11}$ M. Then, using the experimentally measured rate constant of association for *lac* repressor, $k_{\mathrm{exp}} \simeq 10^{10}$ M^{-1}, the experimental search time is equal to $T_0 = 1/(k_{\mathrm{exp}} c_p) \simeq 10$ s, which agrees well with theoretical predictions from our model. It is important to note that the fraction of time the protein spends on DNA can be estimated from the following expression

$$r = \frac{\tau_{1D}}{(\tau_{1D} + \tau_{3D})} = \frac{k_{on}}{k_{on} + k_{off}} \quad (27)$$

Using experimental estimates of $k_{off} \simeq 200-3000\ s^{-1}$ and $k_{on} \simeq 10^4-10^6\ s^{-1}$, we obtain that $0.77 < r < 0.997$; i.e., in all cases, we predict that the protein is mostly bound to DNA, as observed in experiments.[14,15,17] To the best of our knowledge, this is a first theoretical estimate of the target search time that explains experimental measurements on facilitated diffusion at the level of single protein molecules. Our theory suggests that short search times are due to strong coupling between 3D and 1D motions of the protein molecule that leads to fast binding and unbinding transitions, leading to rapid exploration of all sites on DNA.

■ SUMMARY AND CONCLUSIONS

We have investigated theoretically mechanisms of protein search for targets on DNA by analyzing discrete-state stochastic models that take into account relevant biochemical and biophysical transitions such as binding, unbinding, and diffusion along the DNA. Using the first-passage approach that allows for exact solution of these models at all times, the protein target search dynamics has been analyzed explicitly. It has been found that the target position is important for protein search only for short DNA segments, while for realistic large lengths it does not play any role, as long as the target is not occupying the end sites. A dynamic phase diagram for the protein search has also been constructed. Three possible search regimes are identified depending on relative values of the scanning length, the size of the target, and the DNA length. For scanning lengths larger than the DNA length, the search is dominated by simple random-walk dynamics with quadratic scaling for the search time as a function of the DNA length. When the scanning length is larger than the target but smaller than the DNA length, the system is in the sliding regime, where several search cycles must be performed by the protein before the target can be found. A linear scaling for the search time is found in this dynamic phase. For both dynamic regimes, random-walk and sliding, the same scaling behavior is observed using continuum models of protein target search. For scanning length smaller than the target size, the system follows the jumping dynamics, when the search is taking place only via binding and unbinding transitions. Again, in this dynamic phase, a linear scaling for the search time is predicted. It is important to note that the continuum approach completely fails to properly describe this regime.

An explicit analytical framework for protein search dynamics provided a convenient tool for testing the speed-selectivity paradox. Surprisingly, our theoretical analysis showed that the paradox does not exist, since in the limit of small scanning lengths the search rates might not go to zero as predicted by the paradox. This is due to the fact that the protein molecule can still find the target, although only via binding/unbinding events. It is argued that the speed-selectivity paradox is an artificial and unrealistic consequence of applying continuum models to the part of the parameter phase space where the continuum approximation does not hold anymore. A proper discrete-state stochastic analysis that is valid for all ranges of parameters corrects this error, and there is no need to invoke different protein−DNA binding conformations to explain this phenomenon. In addition, using experimental estimates of transition rates, the target search times are calculated for repressor proteins, and it is found that theoretical predictions agree well with reported experimental association rates. It is argued that this is a first theoretical calculation, consistent with basic laws of chemistry and physics and all available experimental information, that explains the facilitated diffusion at the level of single proteins.

Although the presented theoretical approach seems to be successful in capturing main features of protein search dynamics, the models discussed in this work are rather oversimplified with several approximations, and many properties of protein−DNA systems are neglected. It will be important to extend theoretical analysis to account for the sequence dependence in the protein binding to DNA. One could argue that a dynamic phase diagram similar to the one presented in this work is expected, although with increased regions of jumping and sliding phases. Another important direction of future work is to explain the existence of several protein-binding configurations and their role in the protein search dynamics. It will also be relevant to take into account correlations between 3D and 1D motions, as well as the fact that in real systems the protein molecule might not have equal probability to reach every site on DNA. The advantage of the presented discrete-state stochastic method is the fact that it provides a convenient theoretical framework which can be extended in all of these directions. Finally, for a full understanding of the protein search dynamics, it will be critically important to test presented theoretical results directly in experimental studies.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +1 713 3485672. Fax: +1 713 3485155. E-mail: tolya@rice.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The work was supported by a grant from the Welch Foundation (C-1559). The authors thank M. E. Fisher, J. Iwahara, A. G. Cherstvy, and O. Benichou for useful comments and suggestions.

■ REFERENCES

(1) Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. *Molecular Biology of the Cell*, 4th ed.; Garland Science: New York, 2002.

(2) Lodish, H.; Berk, A.; Zipursky, S. L.; Matsudaira, P.; Baltimore, D.; Darnell, J. *Molecular Biology of the Cell*, 4th ed.; W.H. Freeman and Company: New York, 2000.

(3) Phillips, R.; Kondev, J.; Theriot, J. *Physical Biology of the Cell*; Garland Science: New York, 2009.

(4) Riggs, A.; Bourgeois, S.; Cohn, M. *J. Mol. Biol.* **1970**, *53*, 401−417.

(5) Berg, O. G.; Winter, R. B.; von Hippel, P. H. *Biochemistry* **1981**, *20*, 6948−6960.

(6) Berg, O. G.; von Hippel, P. H. *Annu. Rev. Biophys. Biophys. Chem.* **1985**, *14*, 131−158.

(7) Winter, R. B.; Berg, O. G.; von Hippel, P. H. *Biochemistry* **1981**, *20*, 6961−6977.

(8) Hsieh, M.; Brenowitz, M. *J. Biol. Chem.* **1997**, *272*, 22092−22096.

(9) Stanford, N. P.; Szczelkun, M. D.; Marko, J. F.; Halford, S. E. *EMBO J.* **2000**, *19*, 6546−6557.

(10) Halford, S. E.; Marko, J. F. *Nucleic Acids Res.* **2004**, *32*, 3040−3052.

(11) Gowers, D. M.; Wilson, G. G.; Halford, S. E. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15883−15888.
(12) Iwahara, J.; Zweckstetter; Clore, G. M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 15062−15067.
(13) Kolesov, G.; Wunderlich, Z.; Laikova, O. N.; Gelfand, M. S.; Mirny, L. A. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 13948−13953.
(14) Wang, Y. M.; Austin, R. H.; Cox, E. C. *Phys. Rev. Lett.* **2006**, *97*, 048302.
(15) Elf, J.; Li, G.−W.; Xie, X. S. *Science* **2007**, *316*, 1191−1194.
(16) Blainey, P. C.; van Oijen, A. M.; Banerjee, A.; Verdine, G. L.; Xie, X. S. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 5752−5757.
(17) Tafvizi, A.; Huang, F.; Leith, J. S.; Fersht, A. R.; Mirny, L. A. *Biophys. J.* **2008**, *95*, L1−L3.
(18) Bonnet, I.; Biebricher, A.; Porte, P.−L.; Loverdo, C.; Benichou, O.; Voituriez, R.; Escude, C.; Wende, W.; Pingoud, A.; Desbiolles, P. *Nucleic Acid. Res.* **2008**, *36*, 4118−4127.
(19) van den Broek, B.; Lomholt, M. A.; Kalisch, S.−M. J.; Metzler, R.; Wuite, G.J. L. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 15738−15742.
(20) Blainey, P. C.; Luo, G.; Kou, S. C.; Mangel, W. F.; Verdine, G. L.; Bagchi, B.; Xie, X. S. *Nat. Struct. Mol. Biol.* **2009**, *16*, 1224−1230.
(21) Mirny, L. A.; Slutsky, M.; Wunderlich, Z.; Tafvizi, A.; Leith, J. S.; Kosmrlj, A. *J. Phys. A: Math. Theor.* **2009**, *42*, 434013.
(22) Rau, D. C.; Sidorova, N. Y. *J. Mol. Biol.* **2010**, *395*, 408−416.
(23) Larson, D. R.; Zenklusen, D.; Wu, B.; Chao, J. A.; Singer, R. H. *Science* **2011**, *332*, 475−478.
(24) Tafvizi, A.; Huang, F.; Fersht, A. R.; Mirny, L. A.; van Oijen, A. M. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 563−568.
(25) Normanno, D.; Dahan, M.; Darzacq, X. *Biochim. Biophys. Acta* **2012**, *1819*, 482−493.
(26) Hammar, P.; Leroy, P.; Mahmutovic, A.; Marklund, E. G.; Berg, O. G.; Elf, J. *Science* **2012**, *336*, 1595−1598.
(27) Leith, J. S.; Tafvizi, A.; Huang, F.; Uspal, W. E.; Doyle, P. S.; Fersht, A. R.; Mirny, L. A.; van Oijen, A. M. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 16552−16557.
(28) Forget, A. L.; Kowalczykowski, S. C. *Nature* **2012**, *482*, 423−429.
(29) Zandarashvili, L.; Vuzman, D.; Esadze, A.; Takayama, Y.; Sahu, D.; Levy, Y.; Iwahara, J. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, E1724−E1732.
(30) Berg, O. G. *Biopolymers* **1984**, *23*, 1869−1889.
(31) Slutsky, M.; Mirny, L. A. *Biophys. J.* **2004**, *87*, 4021−4035.
(32) Slutsky, M.; Kardar, M.; Mirny, L. A. *Phys. Rev. E* **2004**, *69*, 061903.
(33) Coppey, M.; Benichou, O.; Voituriez, R.; Moreau, M. *Biophys. J.* **2004**, *87*, 1640−1649.
(34) Belotserkovskii, B. P.; Zarling, D. A. *J. Theor. Biol.* **2004**, *226*, 195−203.
(35) Sokolov, I. M.; Metzler, R.; Pant, K.; Williams, M. C. *Biophys. J.* **2005**, *89*, 895−902.
(36) Hu, T.; Grosberg, A. Y.; Shklovskii, B. I. *Biophys. J.* **2006**, *90*, 2731−2744.
(37) Hu, T.; Shklovskii, B. I. *Phys. Rev. E* **2006**, *74*, 021903.
(38) Hu, T.; Shklovskii, B. I. *Phys. Rev. E* **2007**, *76*, 051909.
(39) Cherstvy, A. G.; Kolomeisky, A. B.; Kornyshev, A. A. *J. Phys. Chem. B* **2008**, *112*, 4741−4750.
(40) Loverdo, C.; Benichou, O.; Moreau, M.; Voituriez, R. *Nat. Phys.* **2008**, *4*, 134−137.
(41) Halford, S. E. *Biochem. Soc. Trans.* **2009**, *37*, 343−348.
(42) Fok, P.-W.; Guo, C.-L.; Chou, T. *J. Chem. Phys.* **2008**, *129*, 235101.
(43) Klenin, K. V.; Merlitz, H.; Langowski, J.; Wu, C.-X. *Phys. Rev. Lett.* **2006**, *96*, 018104.
(44) Benichou, O.; Kafri, Y.; Sheinman, M.; Voituriez, R. *Phys. Rev. Lett.* **2009**, *103*, 138102.
(45) Cherstvy, A. G. *J. Phys. Chem. B* **2009**, *113*, 4242−4247.
(46) Lomholt, M. A.; van den Broek, B.; Kalisch, S.-M. J.; Wuite, G. J. L.; Metzler, R. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 8204−8208.
(47) Givaty, O.; Levy, Y. *J. Mol. Biol.* **2009**, *385*, 1087−1097.
(48) Vuzman, D.; Polonsky, M.; Levy, Y. *Biophys. J.* **2010**, *99*, 1202−1211.
(49) Vuzman, D.; Levy, Y. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 21004−21009.
(50) de la Rosa, M. A. D.; Koslover, E. F.; Mulligan, P. J.; Spakowitz, A. J. *Biophys. J.* **2010**, *98*, 2943−2953.
(51) Kolomeisky, A. B. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2088−2095.
(52) von Hansen, Y.; Netz, R. R.; Hinczewski, M. *J. Chem. Phys.* **2010**, *132*, 135103.
(53) Murugan, R. *J. Phys. A: Math. Theor.* **2011**, *44*, 505002.
(54) Koslover, E. F.; de la Rosa, M. A. D.; Spakowitz, A. J. *Biophys. J.* **2011**, *101*, 856−865.
(55) Benichou, O.; Chevalier, C.; Meyer, B.; Voituriez, R. *Phys. Rev. Lett.* **2011**, *106*, 038102.
(56) Bauer, M.; Metzler, R. *Biophys. J.* **2012**, *102*, 2321−2330.
(57) Sheinman, M.; Benichou, O.; Kafri, Y.; Voituriez, R. *Rep. Prog. Phys.* **2012**, *75*, 026601.
(58) Kolomeisky, A. B.; Veksler, A. *J. Chem. Phys.* **2012**, *136*, 125101.
(59) Zhou, H.−X. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 8651−8656.
(60) Zwanzig, R. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 2029−2030.
(61) Benichou, O.; Moreau, M.; Suet, P.-H.; Voituriez, R. *J. Chem. Phys.* **2007**, *126*, 234109.
(62) Redner, S. *A Guide to First-Passage Processes*; Cambridge University Press: Cambridge, U.K., 2001.