

# Sequence heterogeneity accelerates protein search for targets on DNA

Alexey A. Shvets and Anatoly B. Kolomeisky<sup>a)</sup>

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

(Received 22 July 2015; accepted 2 December 2015; published online 24 December 2015)

The process of protein search for specific binding sites on DNA is fundamentally important since it marks the beginning of all major biological processes. We present a theoretical investigation that probes the role of DNA sequence symmetry, heterogeneity, and chemical composition in the protein search dynamics. Using a discrete-state stochastic approach with a first-passage events analysis, which takes into account the most relevant physical-chemical processes, a full analytical description of the search dynamics is obtained. It is found that, contrary to existing views, the protein search is generally faster on DNA with more heterogeneous sequences. In addition, the search dynamics might be affected by the chemical composition near the target site. The physical origins of these phenomena are discussed. Our results suggest that biological processes might be effectively regulated by modifying chemical composition, symmetry, and heterogeneity of a genome. © 2015 AIP Publishing LLC. [<http://dx.doi.org/10.1063/1.4937938>]

## I. INTRODUCTION

Many biological processes are initiated by proteins binding to specific target sequences on DNA. In particular, this process is responsible for transferring and maintaining the genetic information contained in DNA.<sup>1–3</sup> It was recognized long time ago that finding these specific binding sites could be quite a complicated task due to a large number of other nonspecific sites ( $\approx 10^6$ – $10^9$ ) and low concentration of relevant proteins. But experiments suggest that many proteins find their targets much faster than expected from 3D bulk diffusion estimates.<sup>4–8</sup> This surprising phenomenon is known as a *facilitated diffusion*. A significant progress in explaining facilitated diffusion processes has been achieved in recent years due to multiple experimental and theoretical advances.<sup>5–33</sup> However, the detailed mechanisms of the protein search for targets on DNA remain not well understood.<sup>7,8,23</sup>

It is now widely accepted that proteins searching for the specific binding sites on DNA at some conditions might alternate between 3D and 1D search modes.<sup>5,7–9,11</sup> This means that the protein molecule binds nonspecifically to DNA, then slides along the chain, unbinds, and repeats the scanning cycle several times until it finds the target. Recent single-molecule experiments that can visualize the dynamics of individual molecules support this picture.<sup>12,15,16,21,27,28</sup> These observations also underline the critical role of protein-DNA interactions in the facilitated diffusion. Since DNA molecule is a heterogeneous biopolymer, the sequence symmetry and its chemical composition must be an important factor in the protein search for targets. However, how specifically the sequence heterogeneity influences the protein search dynamics remains a controversial problem.

Most theoretical studies of the protein search dynamics ignore the sequence heterogeneity by assuming that the

DNA chain is a homogeneous polymer.<sup>5,8</sup> But there are several investigations that explicitly consider the search on random DNA sequences.<sup>7,34–36</sup> Comparing this process with a motion in the random potential, it was shown that the heterogeneous character of the chain leads to larger search times in comparison with a homogeneous case. But later it was argued that this result is not applicable to the protein search.<sup>23</sup> It is just an artifact of the continuum approximation, which assumed that the protein can reach the target only via DNA sliding, neglecting 3D associations and dissociations events.<sup>23</sup> A more advanced computational study of the sequence heterogeneity also found that it usually slows down the facilitated diffusion by creating traps.<sup>38</sup> However, it was also suggested that the properly positioned traps in the funnel shape near the target can accelerate the protein search.<sup>38</sup> At the same time, it is not clear if such funnel distributions are observed in real systems. Furthermore, recent theoretical studies of Lukatsky and co-workers<sup>39–42</sup> suggested that the sequence symmetry creates additional effective interactions between DNA and protein molecules. Using methods of equilibrium statistical mechanics, it was found that more homogeneous segments of DNA effectively attract proteins stronger than the heterogeneous segments. However, the role of these effective interactions in the protein search for targets on DNA has not been tested yet. Another motivation for investigating the effects of DNA sequences in the protein search in real cells comes from the recent study on nucleosome positioning in eukaryotic DNA.<sup>37</sup>

In this article, we present a theoretical approach that allows us to investigate explicitly the effect of sequence heterogeneity in the protein search for targets on DNA. It is based on a discrete-state stochastic method which takes into account the most relevant physical-chemical processes of the protein search by analyzing first-passage events in the system.<sup>23,30</sup> The advantage of this method is that it provides a full analytical description of the facilitated diffusion. One

<sup>a)</sup>tolya@rice.edu

of the main results of this approach is a development of the general dynamic phase diagram for the target search.<sup>23</sup> Three dynamic search regimes were identified depending on the different length scales in the system. For a protein sliding length  $\lambda$ , defined as an average distance that the protein scans the DNA molecule in one searching cycle, larger than the size of the DNA chain  $L$ , the protein molecule always stays on DNA and performs 1D search with a random-walk dynamics. This leads to a quadratic scaling of the search times as a function of the DNA length. When the sliding length is smaller than the length of DNA but larger than the target size, which is taken to be equal to unity ( $1 < \lambda < L$ ), the protein is searching by combining 3D and 1D motions. In this sliding regime, the linear scaling of the search times is observed. A different dynamic phase is found for the case of the sliding length smaller than the target size,  $\lambda < 1$ . This means that the protein can bind to DNA, but it cannot diffuse along the chain. Here, the search is accomplished only via 3D association and dissociation events without sliding along the DNA molecule. This also leads to the linear scaling in the search times as a function of the DNA length.

## II. THEORETICAL METHOD

In our model, we consider a single DNA molecule with  $L + 1$  binding sites and a single protein molecule, as shown in Fig. 1. It is convenient to utilize a single-molecule view of the search dynamics, while the multi-particle description of the

process with a dependence on the concentrations of proteins and DNA can be easily obtained.<sup>31,33</sup> One of the binding sites is a target, and for convenience, we put it in the middle of the chain, i.e.,  $m = L/2 + 1$ . To model the sequence heterogeneity, we assume that each monomer in the DNA chain can be in one of two chemical states,  $A$  or  $B$  (see Fig. 1). When the protein is bound to the segment  $A$  ( $B$ ), it interacts with energy  $\varepsilon^A$  ( $\varepsilon^B$ ), and  $\varepsilon = \varepsilon^A - \varepsilon^B \geq 0$ . This means that the protein attracts stronger to the  $B$  sites. The protein molecule can diffuse along DNA with the rate  $u_A \equiv u$  ( $u_B = ue^{-\varepsilon}$ , where  $\varepsilon$  is measured in  $k_B T$  units). Here, we assume that, independent of the chemical state of their neighbors, moving out of the sites  $A$  is characterized by the rate  $u_A$ , while the diffusion out of the sites  $B$  is given by  $u_B$ . The units for diffusional rates are squared length per unit time.

The protein search starts in the solution that we label as a state 0 (see Fig. 1). We assume here that the DNA molecule is coiled and the protein diffusion in the surrounding solution is very fast so that all parts of DNA can be reached with equal probability. Then, the protein molecule can bind to any site  $A$  or  $B$  on DNA with the corresponding rates  $k_{on}^A \equiv k_{on}$  or  $k_{on}^B = k_{on}e^{\theta\varepsilon}$ . Similarly, the dissociations from the DNA chain are described by the rates  $k_{off}^A \equiv k_{off}$  and  $k_{off}^B = k_{off}e^{(\theta-1)\varepsilon}$ . Here, the parameter  $0 \leq \theta \leq 1$  specifies how the protein-DNA interaction energy is distributed between the association and dissociation transitions. The physical meaning of this is that the protein molecule tends to bind faster and to dissociate slower from the sites  $B$ , as compared with  $A$  sites, because it interacts stronger with  $B$  sites. The parameter  $\theta$  quantify these effects. We also assume that the binding to the target is given by  $k_{on}^{(T)} = k_{on}$ . Since we are dealing with single molecules, the units for association and dissociation transition rates are  $s^{-1}$ . To test the effect of the sequence symmetry and heterogeneity, we consider the protein search on two different types of the DNA molecules, see Figs. 1(b) and 1(c). One of them consists of two homogeneous segments of only  $A$  and only  $B$  subunits separated by the target (Fig. 1(b)). Another one is the biopolymer with alternating  $A$  and  $B$  sites, as presented in Fig. 1(c). The block copolymer (Fig. 1(b)) has a more homogeneous sequence, while the alternating polymers (Fig. 1(c)) are more heterogeneous. It is important to note that in both cases, the overall interaction between the protein and DNA is the same (the overall chemical composition in both cases is identical), and thus, our analysis probes *only* the effect of the heterogeneity. This is different from previous theoretical studies where the effect of sequence heterogeneity was coupled with the protein-DNA interactions.<sup>38</sup>

To describe the target search dynamics, let us introduce a function  $F_n(t)$ , which is defined as a first-passage probability to reach the target, if at  $t = 0$ , the protein was at the site  $n$  ( $n = 1, 2, \dots, L + 1$  corresponds to the starting on DNA and  $n = 0$  is for the beginning of the process from the bulk solution). The temporal evolution of this quantity can be described by the backward master equations,<sup>23</sup>

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

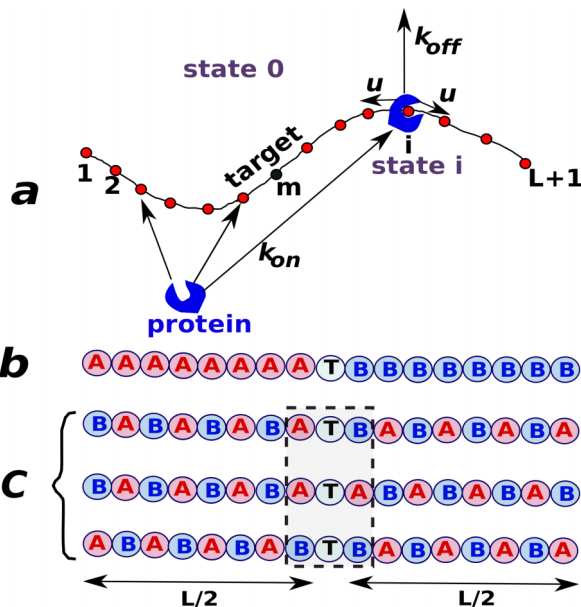


FIG. 1. (a) A general scheme of the protein search. The DNA chain consists of  $L$  nonspecific binding sites and one specific site that is a target for the search. A protein, coming from the solution, can bind to any site on DNA with the association rate per one segment given by  $k_{on}^{(i)}$  with  $i = A$  or  $B$ . When attached, the protein can diffuse along the DNA with the rate  $u_i$  ( $i = A$  or  $B$ ), and it can dissociate into the solution with the rate  $k_{off}^{(i)}$  ( $i = A$  or  $B$ ). The search is finished when the protein binds to the target site at the position  $m = L/2 + 1$ . (b) A fully symmetric AB block copolymer DNA sequence. (c) Pseudo-random alternating sequences with different compositions near the target.

for  $1 < n < L + 1$ , while at the boundaries,

$$\frac{dF_1(t)}{dt} = u_1 F_2(t) + k_{off}^{(1)} F_0(t) - (u_1 + k_{off}^{(1)}) F_1(t), \quad (2)$$

$$\frac{dF_{L+1}(t)}{dt} = u_{L+1} F_L(t) + k_{off}^{(L+1)} F_0(t) - (u_{L+1} + k_{off}^{(L+1)}) F_{L+1}(t). \quad (3)$$

In the bulk solution, we also have

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

It is convenient to analyze these equations in the Laplace space using a transformation  $\tilde{F}_n(s) = \int_0^\infty F_n(t) e^{-st} dt$ . Then, all probabilities can be found explicitly, which leads to the full dynamic description of the search process. The details of the calculations are presented in [Appendices A–C](#). More specifically, the mean first-passage time to reach the target starting from the solution is given by  $T_0 \equiv -\frac{\partial \tilde{F}_0(s)}{\partial s} \Big|_{s=0}$ , and other dynamic properties can be also written explicitly. This framework allows us to compare the search dynamics on DNA with different sequences. Similar expressions for the mean first-passage times can be found for AB alternating DNA chains, as shown in [Appendices A–C](#).

### III. RESULTS AND DISCUSSION

We start our analysis by considering more homogeneous block copolymer sequences (see [Fig. 1\(b\)](#)). In this case, the expressions for the mean search times are given by

$$T_0 = \frac{k_{off} + k_{on} [(L/2 - P^A) + e^{\varepsilon(L/2 - P^B)}]}{k_{on} k_{off} (1 + P^A + e^{\theta \varepsilon P^B})}, \quad (5)$$

where

$$P^{(i)} = \frac{x_i^{1-L/2} - x_i^{1+L/2}}{(1 - x_i)(x_i^{1+L/2} + x_i^{-L/2})}, \quad (6)$$

$$x_i = \frac{2u_i + k_{off}^{(i)} - \sqrt{(2u_i + k_{off}^{(i)})^2 - 4u_i^2}}{2u_i}, \quad (7)$$

for  $i = A$  and  $B$ . When there is no difference between sites  $A$  and  $B$  ( $\varepsilon = 0$ ), we recover the search times for the homogeneous DNA as was obtained earlier.<sup>23</sup> The results for heterogeneous block copolymer are presented in [Fig. 2](#). Again, three dynamic search phases are clearly observed. Increasing the strength of interactions with  $B$  subunits make the search in the random-walk regime much slower. This is because the protein gets effectively trapped on  $B$  sites for the sliding length larger than the DNA length, i.e., for  $\lambda > L$ . Because of the strong interactions with the  $B$  sites, the protein molecule prefers to be found on the  $B$  segment where it slowly diffuses along the DNA.

To consider the search on more heterogeneous sequences, we employ the pseudo-random alternating sequences as shown in [Fig. 1\(c\)](#). They mimic the real random situations quite well, and analytical calculations can be performed for these cases. To prove that the alternating sequences approximate well real random biopolymers, we tested this in Monte Carlo

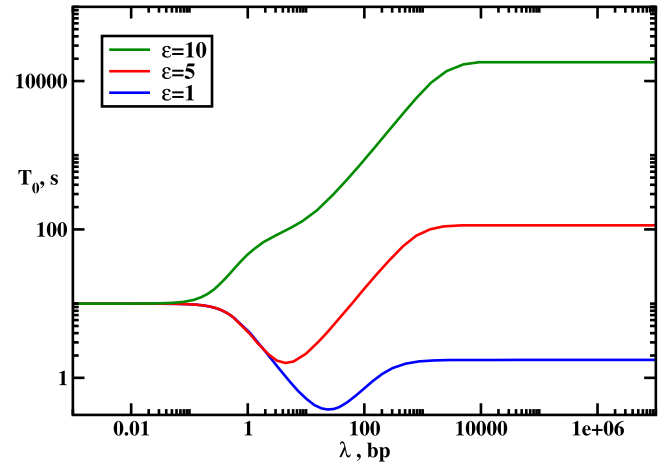


FIG. 2. Average times to find the target for block copolymer DNA sequence as a function of the scanning length  $\lambda = \sqrt{u/k_{off}}$ . The transition rates are  $u = 10^5 \text{ s}^{-1}$  and  $k_{on} = 0.1 \text{ s}^{-1}$ . The DNA length is  $L = 1000$ , and we vary the energy difference  $\varepsilon$  (in units of  $k_B T$ ) for the interaction between the protein and  $A$  and  $B$  subunits on DNA.

simulations by generating random sequences. As one can see from [Fig. 3](#), there is no difference in dynamic properties for the search on real random sequences and on alternating sequences, justifying our approximation. Another interesting observation from [Fig. 3](#) is that the chemical composition near the target might also affect the search dynamics. This can be found only for the intermediate sliding regime ( $1 < \lambda < L$ ) because in this case, the probability fluxes to the target site from the solution and from the DNA are comparable. Modifying the composition of the sites near the target can change the amount of the flux coming from the DNA chain. The flux is larger for  $BTB$  sequences (2  $B$  subunits around the target), leading to the smaller search times. This is because the protein molecule attracts stronger to  $B$  sites and it has a higher probability to be

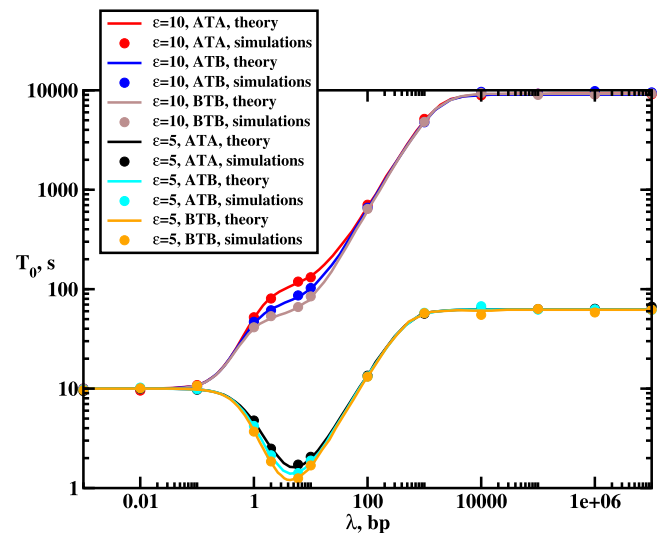


FIG. 3. Comparison of the search times for alternating sequences with random sequences generated in Monte Carlo computer simulations. The transition rates are  $u = 10^5 \text{ s}^{-1}$  and  $k_{on} = 0.1 \text{ s}^{-1}$ . The DNA length is  $L = 1000$ , the loading parameter is  $\theta = 0.5$ , and two different interaction strengths,  $\varepsilon = 10$  and  $\varepsilon = 5$ , are probed.

found here and eventually to go the target. At the same time, the flux is smaller for *ATA* sequences (2 *A* subunits around the target) with weaker interactions to *A* sites, which yields slower search dynamics. For *ATB* sequences, as expected, the intermediate dynamics is observed.

Now we can quantify the effect of sequence heterogeneity in the protein search for the specific binding sites on DNA. The results in Fig. 4 present a ratio of the search times for block copolymer sequences, which are less heterogeneous, and for various alternating sequences, which are more heterogeneous, as a function of the sliding length on DNA. One can see that the effect of the sequence heterogeneity depends on the nature of the dynamic search phase. In the jumping regime ( $\lambda < 1$ ), the symmetry of the sequence does not play any role. This is because in this case, the process is taking place only via associations and dissociations (3D search), and the structure of the DNA chain is not important. The situation is different for the intermediate sliding regime (3D+1D search,  $1 < \lambda < L$ ) where in most cases, the search on alternating sequences is faster. This can be explained by noting that the search time in this dynamic phase is proportional to  $L/\lambda$ ,<sup>23</sup> which gives the average number of cycles before the protein can find the target. In the block copolymer sequence, the protein mostly comes to the target from the *B* segment because of stronger interactions with these sites. In the alternating sequences, the protein can reach the target from both sides. It can be shown analytically (see [Appendices A–C](#)) that the scanning length on the alternating segment is larger than the scanning length for the *B* segment, i.e.,  $\lambda_{AB} > \lambda_B$ . Then, the search time is obviously faster for the alternating sequence because  $L/\lambda_{AB} < L/\lambda_B$ . The only deviation from this picture is found in *ATA* sequences where for small range of parameters, the search is slower than in the block copolymer sequence. The effect of the chemical composition near the target, as discussed above, is responsible for this.

In the random-walk regime (1D search,  $\lambda > L$ ), the effect of the sequence heterogeneity is even stronger: the protein

molecule finds the specific binding site up to 2 times faster for more heterogeneous DNA chains. To understand this behavior, we note that in this case, the mean first-passage time to reach the target is a sum of residence times on the DNA sites. Because the target is in the middle of the chain, the mean time to reach the target from the block copolymer sequence will be  $T_0 \approx (L/4)\tau_B$ , where  $\tau_B$  is the residence time at the site *B*. The protein prefers to start the search at any position on the *B* segment with equal probability, i.e., the distance to the target varies from 0 to  $L/2$ . Then, the average starting position of the protein is  $L/4$  sites away from the target. For the alternating sequences, the average distance to the target is approximately the same, but the chemical composition of intermediate sites is different, yielding,  $T_0 \approx (L/8)\tau_B + (L/8)\tau_A$ . Obviously, the protein spends much less time on *A* subunits, and this leads to faster search for the alternating DNA sequences. For  $\tau_A \ll \tau_B$ , this also explains the factor of 2 in the search speed. In this case, the *B* subunits can be viewed as traps. Thus, in the dynamic phases where the structure of DNA is important, the sequence heterogeneity almost always accelerates the protein search for targets. The stronger the contribution of 1D search modes, the stronger the effect of the sequence heterogeneity.

One could also suggest that the effects of sequence heterogeneity discussed in our work can be experimentally tested by varying the ionic strength of the solution in which the protein search is taking place. This will modify the strength of the protein-DNA attractive electrostatic interactions, leading to changes in dynamic behavior of the system.

#### IV. SUMMARY AND CONCLUSIONS

We presented a theoretical analysis of DNA sequence symmetry and heterogeneity in the protein search process. Using analytical solutions of the discrete-state stochastic models that accounts for most important physical-chemical processes in the system, we obtained a full description of the search dynamics. It is found that the sequence heterogeneity is a crucial factor in the facilitated diffusion. Unlike the previous theoretical and computational models, our approach predicts that the sequence heterogeneity mostly accelerates the search. The mechanisms of this phenomenon depend on the nature of the search regime. It is either the smaller number of search cycles or the smaller number of trapping sites on the path to the target. We also found that in the dynamic phase where the specific binding site can be reached from the solution and from the DNA chain, the chemical composition near the target might influence the search dynamics. The search is faster if the target is surrounded by the subunits which interact stronger with the protein, providing more opportunities to reach the target. Our theoretical results not only clarify the fundamental physics of the protein search dynamics but also suggest that the biological processes can be effectively regulated by modifying the sequence symmetry and heterogeneity in DNA, as well as the chemical composition near the targets. Experiments to test these predictions should provide a better understanding of the microscopic mechanisms of complex biological processes.

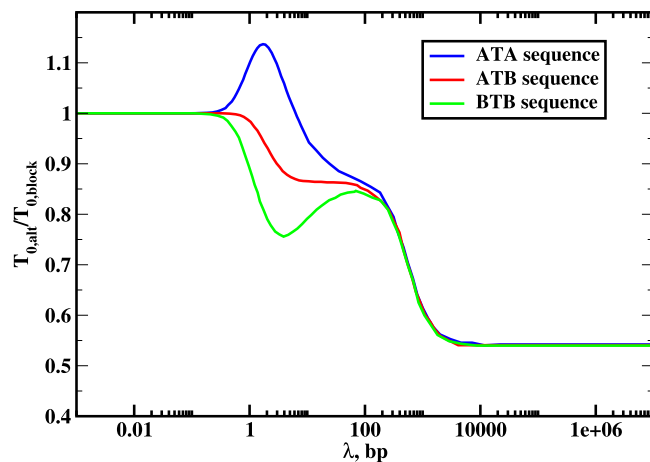


FIG. 4. The ratio of the search times for the alternating DNA sequences and for the block copolymer DNA sequences as a function of the scanning length  $\lambda = \sqrt{u/k_{off}}$ . Three different chemical compositions near the target are distinguished, namely, *ATA*, *ATB*, *BTB*. The transition rates are  $u = 10^5 \text{ s}^{-1}$  and  $k_{on} = 0.1 \text{ s}^{-1}$ . The DNA length is  $L = 1000$ , the loading parameter is  $\theta = 0.5$ , and the energy difference of interactions for the protein with *A* and *B* sites is  $\varepsilon = 5$ .



## ACKNOWLEDGMENTS

The work was supported by the Welch Foundation (Grant No. C-1559), by the NSF (Grant No. CHE-1360979), and by the Center for Theoretical Biological Physics sponsored by the NSF (Grant No. PHY-1427654).

## APPENDIX A: BLOCK COPOLYMER DNA SEQUENCE

The central quantity of our calculations is a first-passage probability defined in the main text. It evolves with time according to a set of backward master equations,<sup>23,30</sup>

$$\begin{cases} \frac{dF_n^{(A)}(t)}{dt} = u_A \{F_{n-1}^{(A)} + F_{n+1}^{(A)}\} + k_{off}^{(A)} F_0 - (2u_A + k_{off}^{(A)}) F_n^{(A)}, \\ \quad 2 < n < L/2, \\ \frac{dF_1^{(A)}(t)}{dt} = u_A F_2^{(A)} + k_{off}^{(A)} F_0 - (u_A + k_{off}^{(A)}) F_1^{(A)}, \\ F_m^{(A)}(t) = \delta(t). \end{cases} \quad (\text{A1})$$

The physical meaning of the last expression is that if we start at the target, the search process is immediately finished. Similar expressions could be written for the  $B$  segment of the DNA chain. In addition, in the bulk solution, we have

$$\begin{aligned} \frac{dF_0(t)}{dt} &= k_{on}^{(A)} \sum_{n=1}^{L/2} F_n^{(A)} + k_{on}^{(A)} F_m^{(A)} + k_{on}^{(B)} \sum_{n=1}^{L/2} F_n^{(B)} \\ &\quad - [k_{on}^{(A)} + (k_{on}^{(A)} + k_{on}^{(B)})L/2] F_0. \end{aligned} \quad (\text{A2})$$

Utilizing the Laplace transformation,  $\tilde{F}(s) = \int_0^\infty dt e^{-st} F(t)$ , we can rewrite the backward master equations as

$$\begin{cases} (s + 2u_A + k_{off}^{(A)}) \tilde{F}_n^A = u_A \{ \tilde{F}_{n-1}^{(A)} + \tilde{F}_{n+1}^{(A)} \} + k_{off}^{(A)} \tilde{F}_0, \\ \quad 2 < n < L/2, \\ (s + u_A + k_{off}^{(A)}) \tilde{F}_1^{(A)} = u_A \tilde{F}_2^{(A)} + k_{off}^{(A)} \tilde{F}_0, \\ \tilde{F}_m^{(A)} = 1. \end{cases} \quad (\text{A3})$$

For the bulk solution, it yields

$$\begin{aligned} \{s + k_{on}^{(A)} + (k_{on}^{(A)} + k_{on}^{(B)})L/2\} \tilde{F}_0 \\ = k_{on}^{(A)} \sum_{n=1}^{L/2} \tilde{F}_n^{(A)} + k_{on}^{(A)} + k_{on}^{(B)} \sum_{n=1}^{L/2} \tilde{F}_n^{(B)}. \end{aligned} \quad (\text{A4})$$

These equations can be solved by assuming that a general form of the solution is following:  $\tilde{F}_n^{(i)} = C_i x_i^n + D_i$ , where  $i = A$  or  $B$ . After the substitution of the general form of the solution into Eq. (A3) and fixing the parameter  $D_i$  via

$$D_i = \frac{k_{off}^{(i)} \tilde{F}_0(s)}{s + k_{off}^{(i)}}, \quad (\text{A5})$$

we obtain that

$$x_i = \frac{s + 2u_i + k_{off}^{(i)} - \sqrt{(s + 2u_i + k_{off}^{(i)})^2 - 4u_i^2}}{2u_i}. \quad (\text{A6})$$

Therefore, the general solution can be written as

$$\tilde{F}_n^{(i)} = C_{i1} x_i^n + C_{i2} x_i^{-n} + D_i. \quad (\text{A7})$$

Furthermore, to find the constants  $C_{i1}$  and  $C_{i2}$ , we should use the boundary conditions at the sites  $n = 1$  and  $n = m = L/2 + 1$  for  $A$  and  $B$  segments. This yields

$$C_{i2} = \frac{1 - D_i}{x_i^{L/2+1} + x_i^{-L/2}}, \quad C_{i2} = C_{i1} x_i. \quad (\text{A8})$$

Then, the solution can be presented in the form

$$\tilde{F}_n^{(i)} = (1 - D_i) \frac{x_i^n + x_i^{1-n}}{x_i^{L/2+1} + x_i^{-L/2}} + D_i, \quad (\text{A9})$$

where  $n = 1, \dots, L/2$  and  $i = A$  or  $B$ . Substituting the last expression and Eq. (A5) into Eq. (A4), we can introduce the auxiliary functions,

$$P^{(i)}(s) = \sum_{n=1}^{L/2} \frac{x_i^n + x_i^{1-n}}{x_i^{L/2+1} + x_i^{-L/2}} = \frac{x_i^{1-L/2} - x_i^{1+L/2}}{(1 - x_i)(x_i^{1+L/2} + x_i^{-L/2})}, \quad (\text{A10})$$

with  $i = A$  or  $i = B$ . Then, after some algebra on the function  $\tilde{F}_0$ , we obtain

$$\tilde{F}_0(s) = \frac{k_{on}^{(A)} + k_{on}^{(A)} P^A(s) + k_{on}^{(B)} P^B(s)}{s + (k_{on}^{(A)} + k_{on}^{(B)})L/2 + k_{on}^{(A)} + \frac{k_{on}^{(A)} k_{off}^{(A)}}{s + k_{off}^{(A)}} [P^A(s) - L/2] + \frac{k_{on}^{(B)} k_{off}^{(B)}}{s + k_{off}^{(B)}} [P^B(s) - L/2]}. \quad (\text{A11})$$

The mean first-passage time to reach the target if the initial position of the protein was in the solution can be calculated as explained in Ref. 23,

$$T_0 \equiv - \left. \frac{\partial \tilde{F}_0(s)}{\partial s} \right|_{s=0}, \quad (\text{A12})$$

which leads to

$$T_0 = \frac{1 + \frac{k_{on}^{(A)}}{k_{off}^{(A)}} [L/2 - P^A(0)] + \frac{k_{on}^{(B)}}{k_{off}^{(B)}} [L/2 - P^B(0)]}{k_{on}^{(A)} + k_{on}^{(A)} P^A(0) + k_{on}^{(B)} P^B(0)}, \quad (\text{A13})$$

or in terms of the energy difference,  $\varepsilon = \varepsilon^A - \varepsilon^B$  (see the main text), we have

$$T_0 = \frac{k_{off} + k_{on}[(L/2 - P^A(0)) + e^\varepsilon(L/2 - P^B(0))]}{k_{on}k_{off}[1 + P^A(0) + e^{\theta\varepsilon}P^B(0)]}. \quad (\text{A14})$$

This result was employed in the main text for the search times of the block copolymer DNA sequence.

## APPENDIX B: ALTERNATING DNA SEQUENCES

In this case, we distinguish three different alternating DNA sequences depending on the chemical composition around the target site. So we will classify these sequences as *ATA*, *ATB*, and *BTB* (see Fig. 1(c)). One can notice that all these sequences can be constructed from two segments:

ABA...ABT and BAB...BAT. At the beginning, let us consider the first segment. As in the previous case, we can write the set of backward master equations as

$$\begin{cases} \frac{dF_n^A(t)}{dt} = u_A\{F_{n-1}^B + F_{n+1}^B\} + k_{off}^A F_0 - (2u_A + k_{off}^A)F_n^A, \\ n = 3, 5, 7, \dots, L/2 - 1, \\ \frac{dF_{n+1}^B(t)}{dt} = u_B\{F_n^A + F_{n+2}^A\} + k_{off}^B F_0 - (2u_B + k_{off}^B)F_{n+1}^B, \\ n + 1 = 2, 4, 6, \dots, L/2, \\ \frac{dF_1^A(t)}{dt} = u_A F_2^B + k_{off}^A F_0 - (u_A + k_{off}^A)F_1^A, \\ F_m^A(t) = \delta(t). \end{cases} \quad (\text{B1})$$

Utilizing the Laplace transformation,  $\tilde{F}(s) = \int_0^\infty dt e^{-st} F(t)$ , we obtain

$$\begin{cases} (s + 2u_A + k_{off}^A) \tilde{F}_n^A = u_A\{\tilde{F}_{n-1}^B + \tilde{F}_{n+1}^B\} + k_{off}^A \tilde{F}_0, \\ n = 3, 5, 7, \dots, L/2 - 1, \\ (s + 2u_B + k_{off}^B) \tilde{F}_{n+1}^B = u_B\{\tilde{F}_n^A + \tilde{F}_{n+2}^A\} + k_{off}^B \tilde{F}_0, \\ n + 1 = 2, 4, 6, \dots, L/2, \\ (s + u_A + k_{off}^A) \tilde{F}_1^A = u_A F_2^B + k_{off}^A \tilde{F}_0, \\ \tilde{F}_m^A = 1. \end{cases} \quad (\text{B2})$$

We will be looking for the solutions in the following form:

$$\begin{cases} \tilde{F}_n^{(1)A} = Cx^n + D_1, & n = 1, 3, 5, \dots, L/2 - 1, \\ \tilde{F}_{n+1}^{(1)B} = QCx^{n+1} + D_2, \end{cases} \quad (\text{B3})$$

where the upper index “(1)” indicates the first segment. After the substitution of these solutions into Eq. (B2), we find the free constants  $D_1$  and  $D_2$  via

$$\begin{cases} (s + 2u_A + k_{off}^A) D_1 = 2u_A D_2 + k_{off}^A \tilde{F}_0, \\ (s + 2u_B + k_{off}^B) D_2 = 2u_B D_1 + k_{off}^B \tilde{F}_0. \end{cases} \quad (\text{B4})$$

This leads to

$$\begin{cases} D_1 \equiv \tilde{F}_0 f_1 = \tilde{F}_0 \frac{k_{off}^A(s + k_{off}^B + 2u_B) + 2u_A k_{off}^B}{k_{off}^A(s + k_{off}^B + 2u_B) + k_{off}^B(s + 2u_A) + s[s + 2(u_A + u_B)]}, \\ D_2 \equiv \tilde{F}_0 f_2 = \tilde{F}_0 \frac{k_{off}^B(s + k_{off}^A + 2u_A) + 2u_B k_{off}^A}{k_{off}^A(s + k_{off}^B + 2u_B) + k_{off}^B(s + 2u_A) + s[s + 2(u_A + u_B)]}, \end{cases} \quad (\text{B5})$$

where we introduced new functions  $f_1$  and  $f_2$ . Furthermore, we can easily find that

$$x = \frac{(s + 2u_A + k_{off}^A) - \sqrt{(s + 2u_A + k_{off}^A)^2 - 4(u_A Q)^2}}{2u_A Q} \quad (\text{B6})$$

and

$$Q = \sqrt{\frac{u_B(s + 2u_A + k_{off}^A)}{u_A(s + 2u_B + k_{off}^B)}}. \quad (\text{B7})$$

Utilizing the boundary conditions at  $n = 1$  and  $n = m = L/2 + 1$ , the general solution can be written as

$$\begin{cases} \tilde{F}_n^{(1)A} = C_1(x^n + ax^{1-n}) + bx^{1-n} + D_1, \\ n = 1, 3, 4, \dots, m - 2, \\ \tilde{F}_{n+1}^{(1)B} = C_1 Q(x^{n+1} + ax^{-n}) + bQx^{-n} + D_2, \end{cases} \quad (\text{B8})$$

where

$$C_1 = \frac{(1 - D_1) - bx^{-L/2}}{x^{L/2+1} + ax^{-L/2}} \quad (\text{B9})$$

and

$$a = \frac{Q - x}{1 - Qx}, \quad b = \frac{D_2 - D_1}{1 - Qx} \equiv b_0(D_2 - D_1). \quad (\text{B10})$$

Therefore, we derived the solution for the segment  $ABA \dots ABT$ . Using the same strategy, we can find the solution for the segment  $BAB \dots BAT$ . In this case, we have to make a change  $A \leftrightarrow B$  in Eq. (B1) and look for the general solution in the form

$$\begin{cases} \tilde{F}_n^{2B} = CQx^n + D_2, & n = 1, 3, 5, \dots, L/2 - 1, \\ \tilde{F}_{n+1}^{2A} = Cx^{n+1} + D_1. \end{cases} \quad (\text{B11})$$

Again using the corresponding boundary conditions at  $n = 1$  and  $n = m = L/2 + 1$ , the general solution can be written as

$$\begin{cases} \tilde{F}_n^{2B} = C_2Q(x^n + cx^{1-n}) + dQx^{1-n} + D_2, \\ n = 1, 3, 5 \dots L/2 - 1, \\ \tilde{F}_{n+1}^{2A} = C_2(x^{n+1} + cx^{-n}) + dx^{-n} + D_1, \end{cases} \quad (\text{B12})$$

where

$$C_2 = \frac{(1 - D_2) - dQx^{-L/2}}{Q(x^{L/2+1} + cx^{-L/2})} \quad (\text{B13})$$

and

$$c = \frac{1 - Qx}{Q - x} = \frac{1}{a}, \quad d = \frac{D_1 - D_2}{Q - x} \equiv d_0(D_1 - D_2). \quad (\text{B14})$$

Thus, the expressions for two main segments are determined.

Now we can calculate the probability to reach the target from the solution applying these expressions. In case when the DNA chain is made up of different segments (with  $ATB$  junction), we can write the backward master equation in the following form:

$$\begin{aligned} \frac{d\tilde{F}_0^{\text{AB}}(t)}{dt} = & k_{on}^A \sum_{n=1,3,5}^{L/2-1} \tilde{F}_n^{(1)A} + k_{on}^B \sum_{n=2,4,6}^{L/2} \tilde{F}_n^{(1)B} + k_{on}^A \tilde{F}_m^A \\ & + k_{on}^B \sum_{n=1,3,5}^{L/2-1} \tilde{F}_n^{(2)B} + k_{on}^A \sum_{n=2,4,6}^{L/2} \tilde{F}_n^{(2)A} \\ & - [k_{on}^A + (k_{on}^A + k_{on}^B)L/2] \tilde{F}_0. \end{aligned} \quad (\text{B15})$$

After utilizing the Laplace transformation, we get

$$\begin{aligned} [s + k_{on}^A + (k_{on}^A + k_{on}^B)L/2] \tilde{F}_0^{\text{AB}} \\ = k_{on}^A \left\{ \sum_{n=1,3,5}^{L/2-1} \tilde{F}_n^{(1)A} + \sum_{n=2,4,6}^{L/2} \tilde{F}_n^{(2)A} \right\} + k_{on}^A \\ + k_{on}^B \left\{ \sum_{n=2,4,6}^{L/2} \tilde{F}_n^{(1)B} + \sum_{n=1,3,5}^{L/2-1} \tilde{F}_n^{(2)B} \right\}. \end{aligned} \quad (\text{B16})$$

Using Eqs. (B8), (B12), and (B5) and performing some algebraic transformations leads to the following expression:

$$\tilde{F}_0^{\text{AB}} = \frac{k_{on}^A + M_B + N_A}{s + \{k_{on}^A(1 - f_1) + k_{on}^B(1 - f_2)\}L/2 + k_{on}^A + f_1M_B + f_2N_A + f\Delta_{AB}}, \quad (\text{B17})$$

where we introduced the auxiliary functions,

$$M_B = k_{on}^A H_B + k_{on}^B Q R_B, \quad N_A = k_{on}^A W_A/Q + k_{on}^B Y_A, \quad (\text{B18})$$

with

$$\begin{cases} H_B = \frac{x - x^{L/2+1} + a(x^{2-L/2} - x^2)}{(1 - x^2)(x^{L/2+1} + ax^{-L/2})}, \\ R_B = \frac{x^2 - x^{L/2+2} + a(x^{1-L/2} - x)}{(1 - x^2)(x^{L/2+1} + ax^{-L/2})}. \end{cases} \quad (\text{B19})$$

This corresponds to the segment that has  $B$  site before the target and

$$\begin{cases} W_A = \frac{x^2 - x^{L/2+2} + c(x^{1-L/2} - x)}{(1 - x^2)(x^{L/2+1} + cx^{-L/2})}, \\ Y_A = \frac{x - x^{L/2+1} + c(x^{2-L/2} - x^2)}{(1 - x^2)(x^{L/2+1} + cx^{-L/2})} \end{cases} \quad (\text{B20})$$

describes the segment that has  $A$  site before the target. In addition,

$$\Delta_{AB} = \Delta_{AA} + \Delta_{BB}, \quad (\text{B21})$$

where

$$\Delta_{AA} = d_0 [(k_{on}^A S_0 + k_{on}^B Q P_0) - x^{-L/2} Q N_A] \quad (\text{B22})$$

and

$$\Delta_{BB} = b_0 [x^{-L/2} M_B - (k_{on}^A P_0 + k_{on}^B Q S_0)]. \quad (\text{B23})$$

The parameters  $b_0$  and  $d_0$  are defined in Eqs. (B10) and (B14), correspondingly, and we have

$$P_0 = \frac{x^{2-L/2} - x^2}{1 - x^2}, \quad S_0 = \frac{x^{1-L/2} - x}{1 - x^2}. \quad (\text{B24})$$

Using Eq. (A12), we can find the mean first-passage time to reach the target if the initial position of the protein was in the solution. In the case of the alternating DNA sequence with  $ATB$  junction, it is given by

$$T_0^{\text{AB}} = \frac{1 - (k_{on}^A f'_1 + k_{on}^B f'_2)L/2 + f'_1 M_B + f'_2 N_A + f\Delta_{AB}}{k_{on}^A + M_B + N_A}, \quad (\text{B25})$$

where the functions  $f_1$  and  $f_2$  ( $f = f_2 - f_1$ ) are defined in Eq. (B5) and  $f'_l = \frac{df_l}{ds}(s=0)$  (for  $l = 1$  or  $2$ ). They can be written explicitly as

$$\left\{ \begin{aligned} f'_1 &= -\frac{k_{\text{off}}^B + 2(u_A + u_B)}{k_{\text{off}}^A (k_{\text{off}}^B + 2u_B) + 2k_{\text{off}}^B u_A} \\ &= -\frac{k_{\text{off}} e^{\theta\varepsilon} + 2u(1 + e^\varepsilon)}{k_{\text{off}} (k_{\text{off}} e^{\theta\varepsilon} + 2u(1 + e^{\theta\varepsilon}))}, \\ f'_2 &= -\frac{k_{\text{off}}^A + 2(u_A + u_B)}{k_{\text{off}}^A (k_{\text{off}}^B + 2u_B) + 2k_{\text{off}}^B u_A} \\ &= -\frac{k_{\text{off}} e^\varepsilon + 2u(1 + e^\varepsilon)}{k_{\text{off}} (k_{\text{off}} e^{\theta\varepsilon} + 2u(1 + e^{\theta\varepsilon}))}, \\ f' &= f'_2 - f'_1 = \frac{e^{\theta\varepsilon} - e^\varepsilon}{k_{\text{off}} e^{\theta\varepsilon} + 2u(1 + e^{\theta\varepsilon})}. \end{aligned} \right. \quad (\text{B26})$$

Similar expressions for the mean first-passage time can be found for the *ATA* junction,

$$T_0^{AA} = \frac{1 - (k_{\text{on}}^A f'_1 + k_{\text{on}}^B f'_2) L/2 + 2f'_2 N_A + 2f \Delta_{AA}}{k_{\text{on}}^A + 2N_A}, \quad (\text{B27})$$

and for the *BTB* junction,

$$T_0^{BB} = \frac{1 - (k_{\text{on}}^A f'_1 + k_{\text{on}}^B f'_2) L/2 + 2f'_1 M_B + 2f \Delta_{BB}}{k_{\text{on}}^A + 2M_B}. \quad (\text{B28})$$

### APPENDIX C: SLIDING LENGTH FOR ALTERNATING SEQUENCES

Consider a semi-infinite alternating AB sequence. The dynamics in this system is the same as in the main text for the DNA sequence of the finite length. Now we estimate the probability to reach the site  $n$  at time  $t$ . This function satisfies the forward master equation and in the case of alternating AB sequence, we can write

$$\left\{ \begin{aligned} \frac{dP_n^A(t)}{dt} &= u_B \{P_{n-1}^B + P_{n+1}^B\} + k_{\text{on}}^A P_0 - (2u_A + k_{\text{off}}^A) P_n^A, \\ n &= 1, 3, 5, \dots, \\ \frac{dP_{n+1}^B(t)}{dt} &= u_A \{P_n^A + P_{n+2}^A\} + k_{\text{on}}^B P_0 - (2u_B + k_{\text{off}}^B) P_{n+1}^B. \end{aligned} \right. \quad (\text{C1})$$

For simplicity, we concentrate on the stationary process when  $dP/dt = 0$ . Now the probability depends only on  $n$ . We can obtain  $P_n$  by solving

$$\left\{ \begin{aligned} u_B [P_{n-1}^B + P_{n+1}^B] + k_{\text{on}}^A P_0 - (2u_A + k_{\text{off}}^A) P_n^A &= 0, \\ n &= 1, 3, 5, \dots, \\ u_A [P_n^A + P_{n+2}^A] + k_{\text{on}}^B P_0 - (2u_B + k_{\text{off}}^B) P_{n+1}^B &= 0. \end{aligned} \right. \quad (\text{C2})$$

We look for the solution in the form

$$\left\{ \begin{aligned} P_n^A &= A_1 x^n + B_1, \quad n = 1, 3, 5, \dots, \\ P_{n+1}^B &= A_2 x^{n+1} + B_2. \end{aligned} \right. \quad (\text{C3})$$

The probability function is an exponentially decaying function of the distance,  $P_n \sim \exp(-n/\lambda)$ , and the corresponding correlation length  $\lambda$  can be estimated as  $\lambda_{\text{AB}} = -1/\log x$ . After the substitution of Eq. (C3) into Eq. (C2), we determine

the constants  $B_1$  and  $B_2$ ,

$$\left\{ \begin{aligned} 2u_B B_2 + k_{\text{on}}^A P_0 &= (2u_A + k_{\text{off}}^A) B_1, \\ 2u_A B_1 + k_{\text{on}}^B P_0 &= (2u_B + k_{\text{off}}^B) B_2. \end{aligned} \right. \quad (\text{C4})$$

Then, dividing the first equation in Eq. (C2) by  $x^{n-1}$  and the second one by  $x^n$ , we obtain

$$\left\{ \begin{aligned} A_2 u_B (1 + x^2) &= A_1 x (2u_A + k_{\text{off}}^A), \\ A_1 u_A (1 + x^2) &= A_2 x (2u_B + k_{\text{off}}^B). \end{aligned} \right. \quad (\text{C5})$$

Thus, the constants  $A_1$  and  $A_2$  are related with each other, and we can introduce a new function

$$Q \equiv \frac{A_2}{A_1} = \sqrt{\frac{u_A (2u_A + k_{\text{off}}^A)}{u_B (2u_B + k_{\text{off}}^B)}}. \quad (\text{C6})$$

Applying this to Eq. (C5) leads to

$$x = \frac{2u_A + k_{\text{off}}^A - \sqrt{(2u_A + k_{\text{off}}^A)^2 - 4u_B Q}}{2u_B Q}. \quad (\text{C7})$$

Recall that  $u_B = ue^{-\varepsilon}$  and  $k_{\text{off}}^B = k_{\text{off}} e^{(\theta-1)\varepsilon}$ . Also, for a homogeneous chain, the correlation length (sliding length) is equal to  $\lambda = \sqrt{u/k_{\text{off}}}$ .<sup>43</sup> Therefore, in the continuous limit, where  $\lambda \gg 1$ , using simple transformations, we derive

$$Q = e^\varepsilon \sqrt{\frac{1 + 1/2\lambda^2}{1 + e^{\theta\varepsilon}/2\lambda^2}} \stackrel{\lambda \gg 1}{\approx} e^\varepsilon \left\{ 1 + \frac{1}{4\lambda^2} (1 - e^{\theta\varepsilon}) \right\}, \quad (\text{C8})$$

and correspondingly,

$$x \approx 1 - \frac{1}{\lambda} \sqrt{\frac{1 + e^{\theta\varepsilon}}{2}}. \quad (\text{C9})$$

Finally, we obtain

$$\lambda_{\text{AB}} = -\frac{1}{\log x} \approx \frac{\sqrt{2}\lambda}{\sqrt{1 + e^{\theta\varepsilon}}}. \quad (\text{C10})$$

From this, the important conclusion is that  $\lambda \geq \lambda_{\text{AB}} \geq \lambda_B = \lambda e^{-\theta\varepsilon/2}$ .

<sup>1</sup>B. Alberts *et al.*, *Molecular Biology of Cell*, 6th ed. (Garland Science, New York, 2014).

<sup>2</sup>H. Lodish *et al.*, *Molecular Cell Biology*, 6th ed. (W.H. Freeman, New York, 2007).

<sup>3</sup>R. Phillips, J. Kondev, and J. Theriot, *Physical Biology of the Cell*, 2nd ed. (Garland Science, New York, 2012).

<sup>4</sup>A. D. Riggs, S. Bourgeois, and M. Cohn, *J. Mol. Biol.* **48**, 67 (1970).

<sup>5</sup>S. E. Halford and J. F. Marko, *Nucl. Acids Res.* **32**, 3040 (2004).

<sup>6</sup>B. van den Broek, M. A. Lomholt, S.-M. Kalisch, R. Metzler, and G. J. L. Wuite, *Proc. Natl. Acad. Sci. U. S. A.* **105**, 15738 (2008).

<sup>7</sup>L. A. Mirny, M. Slutsky, Z. Wunderlich, A. Tafvizi, J. S. Leith, and A. Kosmrlj, *J. Phys. A: Math. Theor.* **42**, 434013 (2009).

<sup>8</sup>A. B. Kolomeisky, *Phys. Chem. Chem. Phys.* **13**, 2088 (2011).

<sup>9</sup>O. G. Berg, R. B. Winter, and P. H. von Hippel, *Biochemistry* **20**(24), 6948 (1981).

<sup>10</sup>O. G. Berg and P. H. von Hippel, *Annu. Rev. Biophys. Biophys. Chem.* **14**, 131 (1985).

<sup>11</sup>R. B. Winter, O. G. Berg, and P. H. von Hippel, *Biochemistry* **20**, 6961 (1981).

<sup>12</sup>D. M. Gowers, G. G. Wilson, and S. E. Halford, *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15883 (2005).

<sup>13</sup>J. Iwahara, M. Zweckstetter, and G. M. Clore, *Proc. Natl. Acad. Sci. U. S. A.* **103**, 15062 (2006).



- <sup>14</sup>G. Kolesov, Z. Wunderlich, O. N. Laikova, M. S. Gelfand, and L. A. Mirny, *Proc. Natl. Acad. Sci. U. S. A.* **104**, 13948 (2007).
- <sup>15</sup>Y. M. Wang, R. H. Austin, and E. C. Cox, *Phys. Rev. Lett.* **97**, 048302 (2006).
- <sup>16</sup>J. Elf, G.-W Li, and X. S. Xie, *Science* **316**, 1191 (2007).
- <sup>17</sup>A. Tafvizi, F. Huang, J. S. Leith, A. R. Fersht, and L. A. Mirny, *Biophys. J.* **95**, L1-L3 (2008).
- <sup>18</sup>T. Hu, A. Yu. Grosberg, and B. I. Shklovskii, *Biophys. J.* **90**, 2731 (2006).
- <sup>19</sup>D. C. Rau and N. Y. Sidorova, *J. Mol. Biol.* **395**, 408 (2010).
- <sup>20</sup>D. R. Larson, D. Zenklusen, B. Wu, J. A. Chao, and R. H. Singer, *Science* **332**, 475 (2011).
- <sup>21</sup>P. Hammar, P. Leroy, A. Mahmutovic, E. G. Marklund, O. G. Berg, and J. Elf, *Science* **336**, 1595 (2012).
- <sup>22</sup>L. Zandarashvili, D. Vuzman, A. Esadze, Y. Takayama, D. Sahu, Y. Levy, and J. Iwahara, *Proc. Natl. Acad. Sci. U. S. A.* **109**, E1724 (2012).
- <sup>23</sup>A. Veksler and A. B. Kolomeisky, *J. Phys. Chem. B* **117**, 12695 (2013).
- <sup>24</sup>A. Marcovitz and Y. Levy, *Biophys. J.* **104**, 2042 (2013).
- <sup>25</sup>E. F. Koslover, M. A. D. de la Rosa, and A. J. Spakowitz, *Biophys. J.* **101**, 856 (2011).
- <sup>26</sup>M. Sheinman, O. Benichou, Y. Kafri, and R. Voituriez, *Rep. Prog. Phys.* **75**, 026601 (2012).
- <sup>27</sup>M. P. Landry, X. Zou, L. Wang, W. M. Huang, K. Schulten, and Y. R. Chemla, *Nucl. Acids Res.* **41**, 2416 (2013).
- <sup>28</sup>A. Tafvizi, F. Huang, A. R. Fersht, L. A. Mirny, and A. M. van Oijen, *Proc. Natl. Acad. Sci. U. S. A.* **108**, 563 (2011).
- <sup>29</sup>J. S. Leith, A. Tafvizi, F. Huang, W. E. Uspal, P. S. Doyle, A. R. Fersht, L. A. Mirny, and A. M. van Oijen, *Proc. Natl. Acad. Sci. U. S. A.* **109**, 16552 (2012).
- <sup>30</sup>A. B. Kolomeisky and A. Veksler, *J. Chem. Phys.* **136**, 125101 (2012).
- <sup>31</sup>A. Esadze, C. A. Kemme, A. B. Kolomeisky, and J. Iwahara, *Nucl. Acids Res.* **42**, 7039 (2014).
- <sup>32</sup>M. Bauer and R. Metzler, *PLoS One* **8**, e53956 (2013).
- <sup>33</sup>M. Lange, M. Kochugaeva, and A. B. Kolomeisky, *J. Chem. Phys.* **143**, 105102 (2015).
- <sup>34</sup>T. Hu and B. I. Shklovskii, *Phys. Rev. E* **74**, 021903 (2006).
- <sup>35</sup>M. Barbi, C. Place, V. Popkov, and M. Salerno, *J. Biol. Phys.* **30**, 203 (2004).
- <sup>36</sup>M. Bauer, E. S. Rasmussen, M. A. Lomholt, and R. Metzler, *Sci. Rep.* **5**, 10072 (2015).
- <sup>37</sup>D. A. Beshnova, A. G. Cherstvy, Y. Vainshtein, and V. B. Teif, *PLoS Comput. Biol.* **10**, e1003698 (2014).
- <sup>38</sup>C. A. Brackley, M. E. Cates, and D. Marenduzzo, *Phys. Rev. Lett.* **109**, 168103 (2012).
- <sup>39</sup>A. Afek, I. Sela, N. Musa-Lempel, and D. B. Lukatsky, *Biophys. J.* **101**, 2465 (2011).
- <sup>40</sup>A. Afek and D. B. Lukatsky, *Biophys. J.* **102**, 1881 (2012).
- <sup>41</sup>A. Afek and D. B. Lukatsky, *Biophys. J.* **105**, 1653 (2013).
- <sup>42</sup>A. Afek, J. L. Schipper, J. Horton, R. Gordn, and D. B. Lukatsky, *Proc. Natl. Acad. Sci. U. S. A.* **111**, 17140 (2014).
- <sup>43</sup>K. A. Dill and S. Bromberg, *Molecular Driving Forces*, 2nd ed. (Garland Science, 2010).