Target search of a protein on DNA in the presence of position-dependent bias

Xi Chen\textsuperscript{1,2}, Xiujun Cheng\textsuperscript{3}, Yanmei Kang\textsuperscript{1} and Jinqiao Duan\textsuperscript{2}

\textsuperscript{1} School of Mathematics and Statistics, Xi’an Jiaotong University, Xi’an 710049, People’s Republic of China
\textsuperscript{2} Department of Applied Mathematics, Illinois Institute of Technology, Chicago, IL 60616, United States of America
\textsuperscript{3} Center for Mathematical Science \& School of Mathematics and Statistics, Huazhong University of Science and Technology, Wuhan 430074, People’s Republic of China

E-mail: duan@iit.edu

Received 10 October 2018
Accepted for publication 28 December 2018
Published 6 March 2019

Abstract. We study the target search on DNA for proteins in the presence of non-constant drift. This search is realized by the facilitated diffusion process. Existing works on this problem focus on the case of constant drift. Starting from a non-local Fokker–Planck equation with the $\alpha$-order fractional Laplace operator and a ‘sink’ term, we obtain the possibility density function for a protein occurring at position $x$ at time $t$. Based on this, we further compute the survival probability and the first arrival density in order to quantify the searching mechanisms. The numerical results show that in the linear drift case, there is an optimal $\alpha$ index for the search to be most likely successful (searching reliability reaches its maximum). This optimal $\alpha$ index depends on the initial position–target separation. It is also found that the diffusion intensity plays a positive role in improving the search success. The nonlinear double-well drift could drive the protein to reach the target with a larger possibility than the linear drag at the initial time period, but viewed over a long time duration, the linear drift is more beneficial for target search success. In contrast to the linear drift case, the search reliability and efficiency with nonlinear double-well drift have a monotonic relationship with the $\alpha$ index, that is, the smaller the $\alpha$ index is, the higher the likelihood of a protein finding its target.
Keywords: numerical simulations, stochastic particle dynamics, stochastic search

Contents

1. Introduction 2
2. Model 4
3. Results 6
   3.1. Target search with linear drift: $v(x) = kx$ .................................6
   3.2. Target search with nonlinear drift: $v(x) = x - x^3$ ........................7
4. Discussion and conclusion 13
   Acknowledgments ........................................................................... 14
   References ...................................................................................... 15

1. Introduction

The so-called target search on DNA refers to the process in which a protein efficiently finds and then binds to its ‘target’ site on the long DNA molecule [1–11]. Many examples, including restriction enzymes in bacteria, chromatin-remodeling proteins in eukarya and gene transcription in any living organisms, have shown that the target search of proteins plays an essential role in fundamental biological functions [4, 5, 9, 10]. To keep up with the metabolic, regulatory, reproductive and defensive needs of the cell, proteins must find their target sites sufficiently fast by some specific mechanism to avoid fruitless searching of a target-less region. One strongly supporting piece of evidence was observed by Riggs et al [1] that in vitro lac repressors find their target binding sites (operator) on Escherichia coli DNA at a surprisingly high rate measured as $10^{10} \text{ M}^{-1} \text{ s}^{-1}$, which demonstrates that the protein-DNA search is about 100 times faster than the Smoluchowski diffusion limit ($10^8 \text{ M}^{-1} \text{ s}^{-1}$) and 1000 times faster than protein–protein association ($10^6$–$10^7 \text{ M}^{-1} \text{ s}^{-1}$).

Even though the precise mechanism whereby a protein finds the DNA binding site is not completely clear, many scientists have made significant efforts to unveil the binding process [2, 3, 12–17]. Berg et al [2] attributed the high protein-DNA association rates to facilitated diffusion, namely switching rounds of 3D bulk diffusion and 1D sliding diffusion along the DNA. Slutsky and Mirny [12] estimated the range of required DNA-protein binding energy and also quantified the time consumption of the two diffusion modes, respectively. Recently, with the help of the experimental data for the real DNA sequence, the facilitated diffusion mechanism has been further analyzed [18–20]. For example, Cencini and Pigolotti [18] noted the energy landscape and observed that an energetic funnel occurs around the target sequence due to the electrostatic complementarity of positive and negative charges on the protein and DNA. They used the

https://doi.org/10.1088/1742-5468/ab00e0
database for a paradigmatic example of the lac repressor to demonstrate that the funnel helps the protein to find target sites when approaching their proximity. Taking the structural plasticity of the DNA sequence into account, Beshnova et al. [19] performed a biophysical analysis in vivo to describe the nucleosome repeat length as a function of the DNA sequence, and then estimated the accessibility of DNA-protein binding using the knowledge of nucleosome position. The large amount of evidence has shown that the facilitated diffusion mechanism makes both quick finding and stable binding possible and thus might shed light on an optimal protein-DNA search mode in vivo.

In general, two strategies in facilitated diffusion can be assumed to alternatively accomplish the target search process of the protein based on the reduced linear topology of the DNA: the intersegmental jumps which lead to Lévy flights (LFs) and the sliding diffusion which yields Brownian motion [21, 22]. Intuitively, the LF with jumps can scan a large space avoiding the oversampling while Brownian motion can exactly locate the target sites preventing the overshooting. It is the combination of these two motions that improves the searching efficiency significantly within a certain parameter range [23]. Many biological processes in microscopic creatures have been observed to operate in Lévy style [24, 25], such as the movement of E. coli underlying their flagella rotation switching [26, 27] and the production of regulatory proteins in the form of transcriptional bursts [28–31]. From a mathematical point of view, the properties of the LF process can be analyzed from a spatial-fractional-order Fokker–Planck diffusion equation [32, 33]. In terms of the target search process, the protein slides on the DNA chain and combines to the target once the hydrogen bond donor and acceptor groups of the target site on the DNA match the peptide functional groups of the protein. The combination makes the total probability density of the protein occurring on the DNA non-normalized to unity and mathematical models with δ-sink are extracted to describe the dynamical properties of the DNA protein search process [34–38].

As the association–dissociation reaction and sliding motion are realized by the electrostatic interaction between the protein and the DNA backbone, a bias unavoidably occurs due to the current induced by charged molecules or the conformation transform of the chromosome in the nucleus [39, 40]. Nevertheless, the bias is generally not a simple constant, since the complex energy potential landscape resulting from the interactions among the nucleus will impose external biases on the molecules [41–43]. By using a Markov model on the phage lambda complete genome, it is shown that the distribution for each of the four nucleotides exhibits as a function of the position along this genome [41], which implicates that the electric potential induced by charged nucleotides is distributed not uniformly but position-dependently. Recently, Langer [43] added an external microelectrode to the DNA–protein complex and explored the molecular dynamics of the DNA and protein in alternating electric fields. In addition, in [42] it was pointed out that a bistable or multi-stable energy landscape with respect to the sequences emerges in the nuclear folding kinetics. Note that the bias term is given by the negative gradient of the energy [44] and the energy landscape is often complex, thus a more general position-dependent bias should be taken into account.

Based on these considerations, our purpose in this study is to investigate the effect of position-dependent bias on the target search of proteins. For simplicity, we take two kinds of typical phenomenological drifts into account: linear and nonlinear. The linear drift is related to the most simple harmonic potential, while the nonlinear drift is taken.
as the negative derivative of the standard bistable quadratic potential. Considering that the inverse Laplace transform of the solution may not work well for our spatial fractional-order Fokker–Planck equation, we resort to a numerical scheme of fast convergence and high accuracy, which was developed by Gao et al [45]. Our numerical experiment shows that this numerical scheme is successful for simulating the solution of such Fokker–Planck equations with small noise intensity.

The paper is organized as follows. In section 2, we present a brief introduction on the biological background of the target search model and the corresponding non-local Fokker–Planck equation to describe the density evolution of the proteins. In section 3, we qualify the searching mechanisms by numerically solving the Fokker–Planck equation and then calculating two important indexes: survival probability and first arrival density. In this section, both linear and nonlinear biases are taken into consideration and the roles of the parameters in the model are examined. Finally, we summarize the results and conclude with a brief discussion in section 4.

2. Model

In the description of the target search process as shown in figure 1, we use the quantity \( p(x, t) \) to represent the possibility of finding the protein at position \( x \) at time \( t \), which can be considered as a possibility density function (PDF) [35]. Assuming that at \( t = 0 \) the protein is located at \( x = x_0 \), we have the initial condition \( p(x, 0) = \delta(x - x_0) \). As referred to in the previous section, the search process involves sliding along the DNA and intersegmental jumps among the DNA segments, which can be mathematically expressed by Brownian motion and Lévy motion, respectively. Apart from these two diffusions, we take the bias (drift velocity) into consideration. Thus the dynamics of \( p(x, t) \) are derived from the Fokker–Planck equation [32, 33]

\[
\frac{\partial}{\partial t} p(x, t) = d \frac{\partial^2 p(x, t)}{\partial x^2} + \epsilon \frac{\partial^\alpha p(x, t)}{\partial |x|^\alpha} - \frac{\partial (v(x) p(x, t))}{\partial x} - \phi(t) \delta(x - x_s).
\]  

(1)

In this equation, \( d \) is the diffusion constant related to the sliding motion on the DNA. In fact, the 1D diffusion constant of the binding protein depends on the size of the protein. For instance, the relatively large protein LacI slides on the DNA with diffusion constant \( d \approx 2.1 - 4.6 \times 10^{-14} \text{ m}^2 \text{s}^{-1} \) [6], while for the smaller hOgg1, the constant is measured as \( 5 \times 10^{-13} \text{ m}^2 \text{s}^{-1} \) [46]. To adapt to general cases for various proteins, we endow the diffusion constant \( d \) with a physical dimension \( [d] = \text{cm}^2/\tilde{t} \) by fixing the length scale. Here \( \tilde{t} \) is noted as the time unit (elementary time) and can be modified to fit specific protein diffusion constants. For LacI, the variable \( \tilde{t} \approx (d/4 - d/2) \times 10^{10} \) s and for hOgg1, it can be approximated as \( \tilde{t} \approx d/5 \times 10^9 \) s. In the searching model of specific proteins finding their specific targets, \( \tilde{t} \) represents a different time scale, but it does not affect the evolution of the system when it is treated as an elementary time in the following numerical calculation. Here \( \epsilon \) is the non-negative noise intensity characterizing intersegmental jumps, the term \( v(x) \) denotes the drift and \( \phi(t) \delta(x - x_s) \) represents the target sinking at \( x = x_s \). The function \( \phi(t) \) describes the flux into the target: in terms of the whole system, once the random walker arrives at the target, it
Target search of a protein on DNA in the presence of position-dependent bias

is removed immediately. That is to say, the target is perfectly absorbing and $\phi(t)$ is physically defined as the density of first arrival. Integrating equation (1) over the state space delivers the expression of the first arrival density:

$$\phi(t) = -\frac{d}{dt} \int_{-\infty}^{\infty} p(x, t) dx.$$  \hfill (2)

Owing to the sink term, $p(x, t)$ is not normalized. So the cumulative survival probability (not hitting the target up to time $t$)

$$P(t) = \int_{-\infty}^{\infty} p(x, t) dx$$

is a deceeding function of time and $\phi(t)$ is always non-negative. Different from the consideration of constant bias in [40], the drift term $v(x)$ in equation (1) can be a position-dependent function varied by the practical matrix density. Moreover, the fractional derivative $\frac{\partial^\alpha}{\partial|y|^\alpha}$ in equation (1) is equivalent to the fractional Laplace operator $-(-\Delta)^{\frac{\alpha}{2}}$, which is defined as [47–49]

$$-(-\Delta)^{\frac{\alpha}{2}} p(x, t) = \int_{\mathbb{R}\setminus\{0\}} \left[ p(x + y, t) - p(x, t) - I_{\{|y|<1\}} yp_x(x, t) \right] \nu_\alpha(dy).$$  \hfill (4)

This is the generator of the $\alpha$-stable Lévy motion and $\nu_\alpha(dy)$ is the jump measure with definition

$$\nu_\alpha(dy) = C_\alpha |y|^{-(1+\alpha)} dy,$$  
where

$$C_\alpha = \frac{\alpha}{2^{1-\alpha} \sqrt{\pi} \Gamma(\frac{1+\alpha}{2}) \Gamma(1 - \frac{\alpha}{2})},$$

Here $\Gamma$ is the Gamma function and $\alpha \in (0, 2)$ is called the non-Gaussianity index or stable index which characterizes the non-Gaussianity of the $\alpha$-stable Lévy motion. Intuitively speaking, the smaller the $\alpha$ index is, the more prominent the non-Gaussianity or heavy-tailed property is. Specifically to the target search process, it has been proved that a Lévy search for a point-like target will never succeed for $0 < \alpha \leq 1$, so in this paper we consider non-Gaussianity in the range $1 < \alpha < 2$. When $\alpha = 2$, the symmetric $\alpha$-stable process is simply the Brownian motion. We can further rewrite equation (1) as

$$\partial_t p(x, t) = -\partial_x (v(x) p(x, t)) + d\partial_{xx} p(x, t) - \phi(t) \delta(x - x_t)$$

$$+ \varepsilon \int_{\mathbb{R}\setminus\{0\}} \left[ p(x + y, t) - p(x, t) - I_{\{|y|<1\}} yp_x(x, t) \right] \nu_\alpha(dy).$$  \hfill (5)

In the target search process, the proteins may locate arbitrary sites on the DNA. So here we consider the problem in the case of a natural boundary condition: the DNA chain is so long that one can set the boundary of stretched DNA chain at infinity, which means the PDF at infinity is zero. The technique of extrapolation is adopted here to meet the natural boundary condition in our algorithm. To reduce the calculating cost, we lay a truncation on the original infinite domain based on a convergence discussion [45]. As a theoretical support, the solution of the non-local Fokker–Planck equation has been proved to exist in $L^p$ for $p \geq 1$ when the drift is linear and double-well [50]. So in our investigation, we numerically calculate the solution of equation (5) in the cases of

https://doi.org/10.1088/1742-5468/ab00e0
linear and double-well drift terms, which cover interesting situations that evoke biases. We use the similar efficient numerical finite difference method developed by Gao et al \cite{45} to simulate the non-local Fokker–Planck equation (5). As we get the PDF of the protein located at \(x\) at time \(t\), we can further introduce the search reliability \(R(t)\) up to time \(t\) to characterize the search mechanism \cite{40}:

\[
R(t) = \int_0^t \phi(\xi)d\xi = 1 - P(t).
\]

Owing to the definition of \(R(t)\) as the integral of the first arrival density, it quantifies the probability that the protein ever arrives at the target up to time \(t\). So a large value of \(R(t)\) corresponds to a high success probability of locating the target, which has the same indication of a small survival probability \(P(t)\). Substantially, one can choose one of them to describe the search reliability alternatively. In our numerical experiments below, we use the survival probability \(P(t)\) as the reliability index. In terms of the first arrival density which depicts the variation rate of the search reliability, we can use it to quantify the search efficiency. Based on the survival probability and first arrival density, the properties of the search strategies can be analyzed from the view of quantitativity.

3. Results

3.1. Target search with linear drift: \(v(x) = kx\)

In the nucleus, the complex energy landscape results from various factors, such as the electrostatic interactions between the protein and DNA, collision of macromolecules, viscous drag in the nucleoplasm and the external induced electric field \textit{in vitro}. In this part we first assume the proteins diffusing in harmonic potential, so it imposes a linear drift on the proteins. Starting with the case of linear drift \(v(x) = -0.01x\), the PDF of the proteins locating on the DNA is numerically obtained from the Fokker–Planck equation (5) as well as other important measures to quantify the search mechanism. Since the search mechanism is substantially related to the non-Gaussianity and the initial location of the protein \cite{40}, we compare the search reliability and efficiency for different non-Gaussianity indexes and initial locations \(x_0\).

The PDF of the searching proteins with linear drift and different \(\alpha\)-stable indexes is shown in figures 2(a)–4(a). It can be seen that the possibility density of the proteins first concentrates on the initial location, which means the proteins start to move but are restricted to the region around the initial point. As time goes on, the protein diffuses and locates more and more DNA segments both in the motion of sliding and jumps, so the possibility of the proteins occurring at different locations increases. Interestingly, the PDF curve will ‘drop’ at the target site \(x_s = 1\) after some time instants, which infers that the proteins will be absorbed or removed at this location. Correspondingly, the total probability flux will flow away through the target as time evolves. The decreasing of the survival probability with time in figures 2(b)–4(b) further supports this phenomenon. The results appear to be consistent with a convection-diffusion equation with \(\delta\)-sink: when a sink is introduced into a dynamical system, the flux will flow into the sink which leads to an ‘inflection’ at the sink.

https://doi.org/10.1088/1742-5468/ab00e0
Moreover, we observe from figures 2–4 that when the initial point is fixed at $x = -0.8$, the ‘inflection’ appears earlier with a smaller non-Gaussianity index $\alpha$ than a larger $\alpha$ index case. Nevertheless, by comparing the survival probability and the first arrival density with different $\alpha$ indexes, one can infer that the relationship between the non-Gaussian $\alpha$ index and the searching reliability or efficiency is not monotonically dependent. The search strategy with more non-Gaussianity at first leads to a higher search reliability, but as time goes on, its superiority is lost to the other search mechanism with a larger $\alpha$ index. To verify the statement, we further plot the search efficiency up to time $T$, which is defined by $E = 1/(\int_0^T t \phi(t) dt)$ [34], as a function of $\alpha$ in figure 5(c). The peak point of the curve corresponds to optimal $\alpha$ where the search efficiency achieves its maximum with given $x_0$. We observe that when the starting point is close to the target, the optimal $\alpha$ approaches 2 and the search can attain a higher efficiency. For the biophysical implication, frequent jumps at first may help the protein to avoid the target-less region and lead to quick target localization. However, as the protein approaches the target, distant jumps would result in leap-overs across the target and reduce the possibility of finding the target. So for proteins distant from the target, long jumps are beneficial for the searching success, while the opposite role will be played for proteins close to the target.

To explore the influence of the noise intensity on the search mechanism, we compare the survival probability and first arrival density with different noise intensities in figure 6. An obvious difference can be seen that the intensive noise induces a much lower survival probability and higher first arrival density than the small noise. The observations illustrate that a strong noise could lead to a high searching reliability and efficiency, and we can deduce from this that when given the non-Gaussianity index $\alpha$ and initial position–target separation, an intensive Lévy motion can be more helpful for the proteins to find their targets.

The effects of initial locations on search reliability and efficiency are also considered in figure 7. It can be seen that when the initial location $x_0 = 0.3$ shifts to $x_0 = -0.8$, then to $x_0 = -1$, the survival probability and first arrival density change slightly. One cannot eliminate the contribution of weak noise to the inconspicuous difference and could observe that even a small shift of the initial location can lead to different search mechanics. It is consistent with the conclusion in [40] that the search process with constant bias is sensitive to the selection of initial locations. We could also get some inspiration from the slight disparity. A large initial position–target separation results in a high search reliability and efficiency when $\alpha = 1.5$. This tells us that when the noise intensity is small, the LF search with long-distance jumps dominantly plays a role in avoiding oversampling. Meanwhile, the occurrence of the leap-overs across the target could lead to a low search efficiency when the initial position is close to the target.

3.2. Target search with nonlinear drift: $v(x) = x - x^3$

In this part, we consider a more general case of the complex energy potential landscape with multi-stable form. As a toy model, we address the target search problem for a double-well potential, where the drift has the form $v(x) = x - x^3$.

Figure 8 shows the case with nonlinear drift $v(x) = x - x^3$. To see the different influences of the linear and nonlinear drifts on the search strategies intuitively, we
Target search of a protein on DNA in the presence of position-dependent bias

Figure 1. Target search of a protein along a fast-folding DNA chain by facilitated diffusion. The red ball represents the protein molecule and the green ball denotes the target site on the DNA. The blue line is the DNA chain.

Figure 2. The search strategy in the linear drift case $v(x) = -0.01x$ with non-Gaussian index $\alpha = 1.2$, the initial location at $x_0 = -0.8$, the target location at $x_s = 1$, noise intensity $\epsilon = 1$ and diffusion constant $d = 0.1$: (a) the curves of PDF at different time instants. (b) The survival probability up to time $t$. (c) The first arrival density $\phi(t)$. 

https://doi.org/10.1088/1742-5468/ab00e0
compared survival probabilities and first arrival densities in figure 9. It can be found that when the drift is nonlinear, the survival probability expresses a sharp decrease at first. However, as time evolves, the survival probability of the nonlinear drift case is larger than that of the linear drift. This can be explained by the fact that nonlinear drift could drive the proteins to multiple directions compared with linear drift. Once the initial direction is just towards the target, the protein can find the target site quickly, otherwise the protein will be pushed away from the target location and it will be difficult to reach it. As for the first arrival density, which is defined as the negative derivative of the survival density, a sudden reduction at the very beginning indicates a sharp decrease of the survival density. Then, the evolution of first arrival density tends to constant which corresponds with the approximately linear decrease of the survival density.

By comparing the influence of $\alpha$-stable indexes on the searching strategy with nonlinear drift in figure 10, we notice a significant disparity occurring among different
non-Gaussianity indexes both for the survival probability and first arrival density. Compared with the linear case, the $\alpha$-stable index affects the search reliability and efficiency monotonically, that is to say, the more non-Gaussian the diffusion is, the lower the survival probability becomes and the larger the first arrival density becomes. Viewed from the biophysical point, the protein moves irregularly in the presence of nonlinear bias. Once it is driven apart from the target, frequent jumps could help the protein reduce the separation from the target and localize more sites on the DNA. In a sense, it will increase the possibility of a protein finding the target to a large extent. So under the given initial location and noise intensity, a smaller $\alpha$ will lead to a more reliable and efficient search process.

Figure 4. The search strategy in the linear drift case $v(x) = -0.01x$ with non-Gaussian index $\alpha = 1.8$, the initial location at $x_0 = -0.8$, the target location at $x_s = 1$, noise intensity $\epsilon = 1$ and diffusion constant $d = 0.1$: (a) the curves of the PDF at different time instants. (b) The survival probability up to time $t$. (c) The first arrival density $\phi(t)$.
Figure 5. ((a) and (b)) The comparisons of the survival probability and first arrival density with different non-Gaussianity indexes $\alpha$ for the linear drift case $v(x) = -0.01x$ with initial location $x_0 = -0.8$. (c) Dependence of search efficiency on $\alpha$. Here the target is located at $x_s = 1$ and the noise intensity is $\epsilon = 1$.

Figure 6. The comparisons of the survival probability and first arrival density with different noise intensities $\epsilon$ for the linear drift case $v(x) = -0.01x$. Here the initial location is $x_0 = -0.8$ and the target is located at $x_s = 1$. The non-Gaussianity index is $\alpha = 1.5$. 
Target search of a protein on DNA in the presence of position-dependent bias

Figure 7. The comparisons of the survival probability and first arrival density with different initial locations by fixing $\alpha = 1.5$ and $\epsilon = 0.5$ for the linear drift case $v(x) = -0.01x$.

Figure 8. The search strategy in the nonlinear drift case $v(x) = x - x^3$ with non-Gaussian index $\alpha = 1.5$, the initial location at $x_0 = -0.8$, the target location at $x_s = 1$, noise intensity $\epsilon = 1$ and diffusion constant $d = 0.1$: (a) the curves of the PDF at different time instants. (b) The survival probability up to time $t$. (c) The first arrival density $\phi(t)$. 

https://doi.org/10.1088/1742-5468/ab00e0
4. Discussion and conclusion

In this work, we have investigated the target search mechanisms of proteins on DNA which is modeled by a non-local Fokker–Planck equation with $\delta$-sink term. Considering that the proteins moving in the nucleoplasm are inevitably biased by the complex energy potential landscape, we introduce position-dependent drift into the search dynamics to describe the external bias. As two toy models, linear and double-well drifts are employed. Here, a fast and accurate algorithm for solving the non-local Fokker–Planck equation is adopted to obtain the PDF of the proteins located at position $x$ at time $t$. The survival probability and the first arrival density, which quantify the search reliability and efficiency, are further calculated.

https://doi.org/10.1088/1742-5468/ab00e0
For the search process with a linear bias, the non-Gaussianity index $\alpha$ has a non-monotonic effect on the search reliability. Indeed, there is an optimal $\alpha$ at which the search is most likely to succeed when other parameters are fixed. Moreover, for the given parameters in our numerical experiments, the larger the noise intensity is, the more likely and faster it is that the proteins find their specific targets. With the small noise intensity, the influences of the initial protein–target distance on the search reliability are not evident, but the slight difference also implies that the proteins initially located far from the target could more likely succeed in reaching the searched target with certain $\alpha$ values. Comparing the nonlinear bias case to the linear one, we observe that the nonlinear drift can reduce the search reliability and efficiency with the given parameters, even leading to an extremely high possibility of success at the initial instant. This is consistent with the intuitive expectation that the occurrence of blocks or other molecules in nucleoplasm will be bound to hinder the search process in the long run, as the direction of particle-moving varies with position in the presence of nonlinear bias. However, if the protein initially moves towards the target without being blocked during very short times, it will quickly find the target. Furthermore, with nonlinear drift, a smaller non-Gaussianity index $\alpha$ will be more beneficial for the search process, and this is different from the linear drift case when the search reliability varies monotonically with $\alpha$.

Our results may also be explained from the viewpoint of biophysics. For a protein with fixed initial location on DNA, too long jumps missing the target and too short jumps leading to local oversampling are unfavorable. There exists a balance between the two strategies to achieve an optimal search, which varies depending on the different initial locations. When the initial location–target separation is large enough, frequent and long jumps help the protein cross uncorrelated regions and quickly find the aimed sequence. However, if the protein is initially located close to the target, long jumps can lead to leap-overs and drive the proteins far away from the target. The numerical results validate the biological fact that for different initial locations, the optimal search strategy may be obtained from different $\alpha$-stable indexes.

We would also like to present an outlook of the first-passage problem of the protein searching mechanism. Recalling that for LF long leap-overs across a point may frequently occur, the probability of the protein actually arriving at the target is significantly smaller than the passage across the site, and the practical arriving time may be much longer than the passage time. The first-passage time problem, acting as central to the kinetics of molecular biology, plays a vital role in aiding our comprehension of target search kinetics. Most available analytical results to quantify first-passage dynamics were derived for the mean first-passage time. However, due to the low copy number of the proteins in the search process, a protein encountering a shrinking target site is a relatively rare event. While the majority of trajectories are indirect and fall into the long-time region, any two such realizations will have a massively different first-passage time. Thus the average over first-passage time makes little sense and is not sufficient to describe the kinetics in this few-encounter regime.

**Acknowledgments**

We would like to thank Xu Sun and Xiaofan Li for their helpful discussions on numerical simulation. This work was partly supported by the NSF grant 1620449, the
Target search of a protein on DNA in the presence of position-dependent bias

NSFC grants 11372233, 11531006, 11771449 and 11772241, and the program of China Scholarship Council.

References

[1] Riggs A D, Bourgeois S and Cohn M 1970 The lac repressor-operator interaction: kinetic studies J. Mol. Biol. 53 401–17
[2] Berg O G, Winter R B and von Hippel P H 1981 Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory Biochemistry 20 6929–48
[3] Berg O G and von Hippel P H 1985 Diffusion-controlled macromolecular interactions Annu. Rev. Biophys. Biophys. Chem. 14 131–60
[4] Lodish H, Berk A, Zipursky S L, Matsudaira P, Baltimore D and Darnell J 2000 Molecular Cell Biology (New York: Freeman)
[5] Ptashne M and Gann A 2001 Genes and signals Cold Spring Harbor Laboratory Press (New York: Cold Spring Harbor Laboratory Press)
[6] Elf J, Li G W and Xie X S 2007 Probing transcription factor dynamics at the single-molecule level in a living cell Science 13 2088–95
[7] Halford S E 2009 An end to 40 years of mistakes in DNA-protein association kinetics? Biochem. Soc. Trans. 37 343–8
[8] Bressloff P C and Newby J M 2013 Stochastic models of intracellular transport Rev. Mod. Phys. 85 135–96
[9] Bressloff P C 2014 Stochastic Process in Cell Biology (New York: Springer)
[10] Marko J F 2015 Biophysics of protein-DNA interactions and chromosome organization Physica A 418 126–53
[11] Niranjani G and Murugan R 2016 Generalized theory on the mechanism of site-specific DNA-protein interactions J. Stat. Mech. 053501
[12] Slusky M and Mirny L A 2004 Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential Biophys. J. 87 4021–35
[13] Halford S E and Marko J F 2004 How do site-specific DNA-binding proteins find their targets? Nucl. Acids Res. 32 3040–52
[14] Cherstvy A G, Kolomeisky A B and Kornyshev A A 2008 Protein-DNA interactions: reaching and recognizing the targets J. Phys. Chem. B 112 4741–50
[15] Mirny L, Slusky M, Wunderlich Z, Tafvizi A, Leith J and Kosmrlj A 2009 How a protein searches for its site on DNA: the mechanism of facilitated diffusion J. Phys. A: Math. Theor. 42 434013
[16] Kolomeisky A B 2011 Physics of protein-DNA interactions: mechanisms of facilitated target search Phys. Chem. Chem. Phys. 13 2088–95
[17] Liu L, Cherstvy A G and Metzler R 2017 Facilitated diffusion of transcription factor proteins with anomalous bulk diffusion J. Phys. Chem. B 121 1284–9
[18] Cencini M and Pigolotti S 2018 Energetic funnel facilitates facilitated diffusion Nucleic Acids Res. 46 558–67
[19] Beshnova D A, Cherstvy A G, Vainshtein Y and Teif V B 2014 Regulation of the nucleosome repeat length in vivo by the DNA sequence, protein concentrations and long-range interactions PLoS Comput. Bio. 10 e1003698
[20] Bauer M, Rasmussen E S, Lombolt M A and Metzler R 2015 Sci. Rep. 5 10072
[21] Copy M, Benichou O, Voituriez R and Moreau M 2004 Kinetics of target site localization of a protein DNA: a stochastic approach Biophys. J. 87 1640–9
[22] Sheinman M, Benichou O, Kafri Y and Voituriez R 2012 Classes of fast and specific search mechanisms for proteins on DNA Rep. Prog. Phys. 75 026601
[23] Koslover E F, Díaz de la Rosa M A and Spakowiz A J 2011 Theoretical and computational modeling of target-site search kinetics in vitro and in vivo Biophys. J. 101 856–65
[24] Kong X and Thakor N V 1996 Adaptive estimation of latency changes in evoked potentials IEEE Trans. Biomed. Eng. 43 189–97
[25] Shen X R, Zhang H, Xu Y and Meng S X 2016 Observation of alpha-stable noise in laser gyroscope data IEEE Sens. J. 16 1998–2003
[26] Korobkova E, Emonet T, Vilar J M G, Shimizu T S and Cluzel P 2004 From molecular noise to behavioural variability in a single bacterium Nature 428 574–8
[27] Tu Y and Grinstein G 2005 How white noise generates power-law switching in bacterial flagellar motors Phys. Rev. Lett. 94 208101

https://doi.org/10.1088/1742-5468/ab00e0

15
Target search of a protein on DNA in the presence of position-dependent bias

[28] Xu Y, Li Y G, Zhang H, Li X F and Kurths J 2016 The switch in a genetic toggle system with Lévy noise Sci. Rep. 6 31505
[29] Zheng Y Y, Serdukova L, Duan J Q and Kurths J 2016 Transitions in a genetic transcriptional regulatory system under Lévy motion Sci. Rep. 6 29274
[30] Kang Y M, Chen X, Lin X D and Tan N 2017 Mean first passage time and stochastic resonance in a transcriptional regulatory system with non-Gaussian noise Fluct. Noise Lett. 16 175007
[31] Chen X and Kang Y M 2017 Switches in a genetic regulatory system under multiplicative non-Gaussian noise J. Theor. Biol. 435 134–44
[32] Metzler R and Klafter J 2000 The random walk’s guide to anomalous diffusion: a fractional dynamics approach Phys. Rep. 339 1–77
[33] Metzler R and Klafter J 2004 The restaurant at the end of the random walk: recent developments in the description of anomalous transport by fractional dynamics J. Phys. A: Math. Gen. 37 R161–208
[34] Chechkin A V, Metzler R, Gonchar V Y, Klafter J and Tanatarov L V 2003 First passage and arrival time densities for Lévy flights and the failure of the method of images J. Phys. A: Math. Gen. 36 L537
[35] Lomholt M A, Aunbörnsson T and Metzler R 2005 Optimal target search on a fast-folding polymer chain with volume exchange Phys. Rev. Lett. 95 260603
[36] Koren T, Lomholt M A, Chechkin A V, Klafter J and Metzler R 2007 Leapover length and first passage time statistics for Lévy flights Phys. Rev. Lett. 99 160602
[37] García-Garcia R, Rosso A and Schehr G 2012 Lévy flights on the half line Phys. Rev. E 86 011101
[38] Janakiraman D 2017 Lévy flights in the presence of a point sink of finite strength Phys. Rev. E 95 012154
[39] Palyulin V V, Chechkin A V and Metzler R 2014 Space-fractional Fokker–Planck equation and optimization of random search processes in the presence of an external bias J. Stat. Mech. P11031
[40] Palyulin V V, Chechkin A V and Metzler R 2014 Lévy flights do not always optimize random blind search for sparse targets Proc. Natl Acad. Sci. USA 111 2931–6
[41] Vergne N 2008 Drifting Markov models with polynomial drift and applications to DNA sequences Stat. Appl. Genet. Mol. Biol. 7 6
[42] Dhar A, Girdhar K, Singh D, Gehman H, Ebbinghaus S and Gruebele M 2011 Protein stability and folding kinetics in the nucleus and endoplasmic reticulum of eucaryotic cells Biophys. J. 101 421–30
[43] Langer A, Kaiser W, Svejda M, Schwertler P and Rant U 2014 Molecular dynamics of DNA? Protein conjugates on electrified surfaces: solutions to the drift-diffusion equation J. Phys. Chem. B 118 597–607
[44] Wu H and Noé F 2011 Bayesian framework for modeling diffusion processes with nonlinear drift based on nonlinear and incomplete observations Phys. Rev. E 83 036705
[45] Gao T, Duan J and Li X 2016 Fokker–Planck equations for stochastic dynamical systems with symmetric Lévy motions Appl. Math. Comput. 278 1–20
[46] Blainey P C, Luo G, Kou S, Mangel W F, Verdine G L, Bagchi B and Xie X S 2009 Non-specifically bound proteins spin while diffusion along DNA Nat. Struct. Mol. Biol. 16 1224–9
[47] Applebaum D 2009 Lévy Processes and Stochastic Calculus 2nd edn (New York: Cambridge University Press)
[48] Duan J 2015 An Introduction to Stochastic Dynamics (New York: Cambridge University Press)
[49] Schertzer D, Larchevêque M, Duan J, Yanovsky V and Lovejoy S 2001 Fractional Fokker–Planck equation for nonlinear stochastic differential equations driven by non-Gaussian Lévy stable noise J. Math. Phys. 42 200–12
[50] Wang M and Duan J 2017 Existence and regularity of linear non-local Fokker–Planck equations with growing drift J. Math. Anal. Appl. 449 228–43
[51] Grebenkov D S, Metzler R and Oshanin G 2018 Towards a full quantitative description of single-molecule reaction kinetics in biological cells Phys. Chem. Chem. Phys. 20 16393
[52] Pulkkinnen O and Metzler R 2013 Distance matters: the impact of gene proximity in bacterial gene regulation Phys. Rev. Lett. 110 198101
[53] Godec A and Metzler R 2016 Universal proximity effect in target search kinetics in the few-encounter limit Phys. Rev. X 6 041037