ELSEVIER

Biophysical Chemistry 120 (2006) 143 – 147

Biophysical Chemistry

http://www.elsevier.com/locate/biophyschem

Critical jump sizes in DNA–protein interactions

R. Murugan *

Department of Chemical Sciences, Tata Institute of Fundamental Research Homi Bhabha Road, Colaba, Mumbai, 400005, India

Received 27 September 2005; received in revised form 29 October 2005; accepted 29 October 2005
Available online 1 December 2005

Abstract

Interaction of a protein molecule with a specific-site on the DNA lattice can be modeled as an unbiased random jump process. Here we show that there exists a critical jump size (k_c) beyond which site-specific association of a protein molecule with a DNA lattice cannot be facilitated. The maximum achievable association rate is predicted to be $\sim 10^{10}$ mol^{-1} s^{-1}. This critical jump size scales with the total length of DNA lattice (N) as $k_c \propto N^{2/3}$. Beyond k_c the mean first passage time MFPT (denoted as T) required for the protein molecule to target the specific-site follows a linear scaling law as $T \propto N$ rather than the usual $T \propto N^2$ scaling law. On the basis of these results we argue that the evolution of the super coiled structures of the genomic DNA must be a consequence of the existence of this critical jump sizes. We finally show that the random jump method of searching for the specific-site by the protein molecule on the DNA lattice itself introduce an abstract linear type potential favoring the site-specific association rate.

Keywords: DNA–protein interactions; Critical jump size; Site-specific association

1. Introduction

Recognition of a specific sequence of DNA by a protein molecule in presence of an enormous amount of non-specific sequences plays a central role in molecular biology especially in the replication (recognition of the origin of replication by DNA-polymerase) and the transcription (recognition of promoter by RNA-polymerase) of the genetic material [1–10]. Since almost all the cellular processes are taking place in a three dimensional space, it is natural to consider the mode of recognition of the specific-site of DNA by a protein molecule as a three dimensional diffusion-controlled process. However earlier in vitro studies on the interaction of the *lac* Operator (a specific DNA sequence) with the *lac* Repressor protein showed an association rate of 10^9–10^{10} mol^{-1} s^{-1} which was higher than that of the diffusion controlled rate in aqueous medium [9]. Later, aforementioned paradox was resolved by considering the DNA–protein interaction as two-step processes where the binding of the protein to the non-specific sites of DNA is the first step and subsequent recognition of the specific-site by facilitated one-dimensional searching processes such as sliding, hopping and inter segmental transfers constitute the second step [9,11]. Earlier treatments assumed an existence of a free energy correlation towards the specific-site along the DNA lattice to explain the observed site-specific association rate. Moreover aforesaid facilitating processes were assumed to be independent of each other but with cumulative effect in enhancing the site-specific association rate [9]. Nevertheless no such free energy correlations or bias pointing towards the specific-site have been identified along the DNA lattice so far. Recently we have shown that [11] the existence of the free energy correlation is not necessary when the protein molecule searches for the specific-site on DNA lattice by unbiased random jumps. Moreover, simple calculations have revealed that in order to enhance the diffusion controlled rate ($\sim 10^8$ mol^{-1} s^{-1}) to a magnitude of $\sim 10^{10}$ mol^{-1} s^{-1} or beyond, the required jump size should satisfy [11] the inequality $2N^{2/3} \leq k < N$ where N is the total length (in base pairs, *bp*) of the DNA lattice under consideration and k denotes the jump size. For example, if $N = 10^3$ bp then a minimum jump size of $k \approx 200$ bp is required to enhance the diffusion controlled rate to two orders of magnitude. Here one should note that the size of recognition site of DNA corresponding to a particular protein is

* Tel.: +91 22 2280 4545; fax: +91 22 2280 4610/2280 4611.
E-mail addresses: muruga@tifr.res.in, rmurugan@gmail.com.

0301-4622/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2005.10.014

generally very small compared to that of the genome size, e.g., in case of *E. coli*, the genome size is $N \approx 4 \times 10^6$ bp whereas the length of the recognition stretch of RNA polymerase in only 60–100 bp (consists of TATA and CAAT box [see Ref. [8]]). But the required jump size to enhance the diffusion controlled associated rate to two orders of magnitude is $k \approx 5 \times 10^4$ bp which is much higher than the length of the recognition stretch of RNA polymerase. In this article we investigate the effect of increasing jump size k on the rate associated with the site-specific interaction of protein with DNA lattice.

2. Theory and mathematical derivations

Let us consider a protein molecule which has non-specifically bound to the DNA lattice of $N+\omega$ bp in length and currently searching for the specific-site by unbiased random jump motion with a jump size of k bp. At any time t, the position of the protein molecule on the DNA lattice is denoted by the variable x where x is such that $0 \le x \le N+\omega$. Here ω is the length of the non-specific DNA that flanks the DNA stretch in the upstream of [0, N] and the set of lattice points $\{0, N+\omega\}$ are the helical ends of the DNA under consideration. We assume that starting from the lattice position $x=x_0$ the probability of finding the protein molecule after a jump anywhere in the interval $x_0-k \le x \le x_0+k$ is $1/2k$ [see Ref. [11] for a justification]. Here we consider $x=0$ as the reflecting boundary (due to the two step assumption) and let us assume that the recognition site is a set of DNA lattice points in the interval such that [N, $N+\delta$] where δ is the length of the recognition stretch and $\delta \ll \omega$ in the present context. The master equation that describes aforementioned phenomenon in dimensionless form is given as follows (contrasting to Ref. [11], where the factor w = transition rate × transition probability was used, here we set transition rate = 1 and transition probability = $1/2k$).

$$\partial_t P = \sum_{i=1}^{k} \left[P_{x+i,t} + P_{x-i,t} - 2P_{x,t} \right]. \tag{1}$$

Here P denotes the probability of observing the protein molecule at the specified position on the DNA lattice at a given time t. The corresponding Fokker–Plank equation (FPE) is,

$$\partial_t P = (D/2)\partial_x^2 P. \tag{2}$$

Where the phenomenological one dimensional diffusion coefficient can be given as,

$$D = \frac{1}{k} \sum_{i=1}^{k} i^2 = 6[(k+1)(2k+1)]^{-1}. \tag{3}$$

Under the condition that $k \le \delta$ the mean first passage time (MFPT) denoted as $T(x_0, N|k)$ taken by the protein molecule to escape from the domain [0, N] starting from the lattice position $x_0=0$ can be calculated from the [11,12] corresponding backward Fokker–Plank equation (Eq. (4)) with a reflecting boundary condition at $x=0$ where $d_x T_x|_{x=0}=0$ and an absorbing boundary condition at $x>N$ where $T_x|_{x>0}=0$ as follows.

$$d_x^2 T_x = -\frac{2}{D} \tag{4}$$

Upon integrating Eq. (4) with the appropriate boundary conditions one obtains,

$$T(x_0 = 0, N|k) = 6N^2[(k+1)(2k+1)]^{-1}. \tag{5}$$

However, when $k > \delta$ there is a definite probability for the protein molecule to escape out in to the non-specific flanking domain [$N+\delta$, $N+\omega$] without actually getting absorbed at the specific stretch [N, $N+\delta$]. In this situation Eq. (5) is clearly not valid. It is obvious to note that the MFPT will be more than that is predicted by Eq. (5) since those trajectories falling outside the interval [0, $N+\delta$] without actually getting absorbed at the specific-stretch [N, $N+\delta$] has to pass through long routes before they come back and get absorbed at the specific-stretch [N, $N+\delta$]. Here one should note that apart from $x=0$ there is a reflecting boundary at $x=N+\omega$, too. Since we have three boundary conditions i.e. reflecting boundaries at the lattice positions $\{0, N+\omega\}$ and an absorbing boundary in the specific stretch [N, $N+\delta$], neither Eq. (2) nor Eq. (4) can be solved analytically. However, one can easily compute the MFPT as follows. For simplicity we consider a situation where $\delta=0$ and $\omega=1$ which means that only the position $x=N$ is the absorbing boundary and those trajectories falling outside the interval [0, N] will be injected back into the same domain [0, N] by introducing a reflecting boundary at $x=N+1$ until all the trajectories pass through the lattice position $x=N$.

Now we consider only the interval [0, $N-1$] and let us compute the MFPT associated with the protein molecule to escape only through the point $x=N$. We consider M number of trajectories starting from the position $x=x_0$ where $0<x_0<N$. Suppose if the jump size is $k=1$ then it is obvious to note that all the trajectories will pass through and get absorbed at $x=N$. However when $k>1$ the protein molecule may hit any one of the lattice points on the DNA lattice in the set $x=\{N, N+1, N+2, \ldots, N+k\}$ among which only $x=N$ is productive and therefore the MFPTs associated with the positions at $x=\{N+1, N+2, \ldots, N+k\}$ simply add up to the MFPT associated with the position at $x=N$ with appropriate weighting factors. Here one should note that among M number of trajectories, M/k number of trajectories will end at the position $x=N$ (and get absorbed) with a MFPT that is given as,

$$T_{\mathrm{LR},x_0,0} = \frac{N^2 - x_0^2}{D}. \tag{6}$$

And $M(i/k)$ number of trajectories will end in the interval [$N+1$, $N+i$] with MFPTs which can be given as follows.

$$T_{\mathrm{LR},x_0,i} = \frac{(N+i)^2 - x_0^2}{D}. \tag{7}$$

Therefore the overall MFPT associated with the escape of the protein molecule only through the lattice position $x=N$ is given by the weighted sum,

$$T_N(x_0=0,N|k)=\sum_{i=0}^{k}\mu_i T_{\mathrm{LR},x_0,i} \quad (8)$$

where the weighting factor is given as $\mu_i=i/k$. Using the summation formula given by Eq. (8), when $x_0=0$, the overall MFPT taken by a protein molecule to escape from the domain [0, N] through only the lattice point [$x=N$] can calculated to be,

$$T_N(x_0=0,N|k)=T_N(x_0=0,N|k)+2N+\frac{3k(k+1)}{2(2k+1)}. \quad (9)$$

Here one should note that for large jump sizes Eq. (9) can be approximated as $T_N(x_0=0,N|k)\approx\frac{3N^2}{k^2}+2N+\frac{3k}{4}$ from which we find that there exists a minimum value of MFPT at the critical jump size $k=k_c=2N^{2/3}$ where the first derivative of MFPT with respect to the jump size k vanishes i.e. $d_kT_N(0,k|0)|_{k=k_c}=-\frac{6N^2}{k_c^3}+\frac{3}{4}=0$. In other words at this critical jump size the site-specific association rate is a maximum since the association rate is inversely proportional to the MFPT.

Our earlier studies [11] have shown that the jump size k_r that is the required to increase the diffusion controlled site-specific association rate of protein with DNA ($\sim10^8$ mol^{-1} s^{-1}) to two orders of magnitude satisfies the inequality $2N^{2/3}\le k_r\le N$ where N is size of the DNA lattice under consideration. In this context Eq. (9) clearly states that unlike in case Eq. (5) the MFPT is almost constant ($\approx 2N$) at the jump sizes $k\ge k_c$ (here we should note that $k<N$). Therefore, increasing the jump size beyond $k_c=2N^{2/3}$ will not enhance the association rate of protein molecule with its specific-site on the DNA lattice which means that even by the random jump method of searching the site-specific association rate cannot be enhanced beyond $\sim10^{10}$ mol^{-1} s^{-1}. It is interesting to note that most of the DNA–protein interaction systems studied so far have shown the association rate in the range of 10^9–10^{10} mol^{-1} s^{-1} which in turn aggress well with our prediction. One can easily guess that energy input (in the form of ATPs) will be necessary to enhance the site-specific association rate beyond this limit.

So far we have assumed a perfect absorbing boundary condition (irreversibility condition) for the protein molecule at the specific-site on the DNA lattice which is an over simplification of the underlying process since there is always a probability such that $P_e>0$ associated with the site-specifically bound protein molecule to escape in to the non-specific region of the DNA lattice under consideration (reversibility condition). One can easily compute the magnitude of P_e as follows. Suppose let us assume that the current position of the protein molecule on the DNA lattice is $x=N-1$. When the jump size is k bp, the probability associated with the protein molecule to find the specific-site $x=N$ in the next step is simply $1/2k$. Therefore at any time the probability associated with escape of the site-specifically bound protein molecule into the non-specific region should be proportional to $1/2k$, i.e. $P_e=\alpha/2k$ (here α is the proportionality constant), which indicates that under critical jump conditions $k\ge k_c$ and when $\alpha\ll1$, the effect of escape of the site-specifically bound protein molecule into the non-specific region of the DNA on the overall MFPT is negligible and the prediction by Eq. (9) is still valid.

In a living cell various genes are simultaneously expressed in order to carry out diverse cellular functions. For a proper cellular function a coherent and coordination control is necessary among the expression of various genes. In other words each gene should possess equal probability to get expressed. This can be achieved only if the probabilities of finding various promoters by the RNA polymerase are equal all over the genomic DNA. From our theory we can conclude that this condition is achieved only when the RNA polymerase searches for the promoters via unbiased random jumps with the jump size equal to or beyond the critical jump size i.e. $k\ge k_c$. Here we should note that as the jump size k increases, the contribution of sliding (here the step size is unit bp) to the enhancement of site-specific DNA-protein interaction rate decreases. Since the critical jump size k_c positively correlates with the total size of the DNA (N), we can conclude that the contribution of sliding to the enhancement of site-specific DNA–protein interaction rate decreases as the size of the DNA increases [11]. Moreover, existence of the critical jump size k_c in turn confirms the presence of sliding phenomenon in the site-specific association of a protein molecule with the template DNA.

3. Simulation results and discussion

In order to check the validity of the Eq. (9) we have carried out the random walk simulation studies. Here the settings are as follows. $N=100$, the position $x=0$ and $x=N+1$ are the reflecting boundary and only the position $x=N$ is the absorbing boundary and we assume $\delta=0$ and $\omega=1$. The protein molecule jump right or left with equal probabilities and the total number of trajectories used for averaging is 10^5. When the jump size is k, the probability of observing the protein molecule starting from the position x anywhere in the interval $[x-k, x+k]$ is $1/2k$. When $(x-i)<0$, the protein molecule will be reflected such that the new position is $-x+i$. Similarly when $(x+i)>N$ then the protein molecule will be put back in to the interval [0, N] such that the new position is $2N-x-i$. When $x+i=N$ the protein molecule is removed from the interval [0, N]. The simulated $T_N(x_0=0,\ N=100|k)$ as well as the theoretical prediction is shown in Fig. 1, which clearly proved the validity of Eq. (9) and therefore the arguments corresponding to DNA–protein interactions those we be put forth based on Eq. (9) are indeed valid.

From Eq. (9) we can conclude that when $k\ge k_c$ the MFPT taken by the random walker to escape from the domain [x_0, N] through the point $x=N$ scales approximately linear with the total length (N), i.e. $T_N(x_0,N|k\ge k_c)\propto N$ rather than the usual observation, i.e. $T_N(x_0,\ N|k<k_c)\propto N^2$. It is interesting to note that when a one dimensional random walker moves in linear

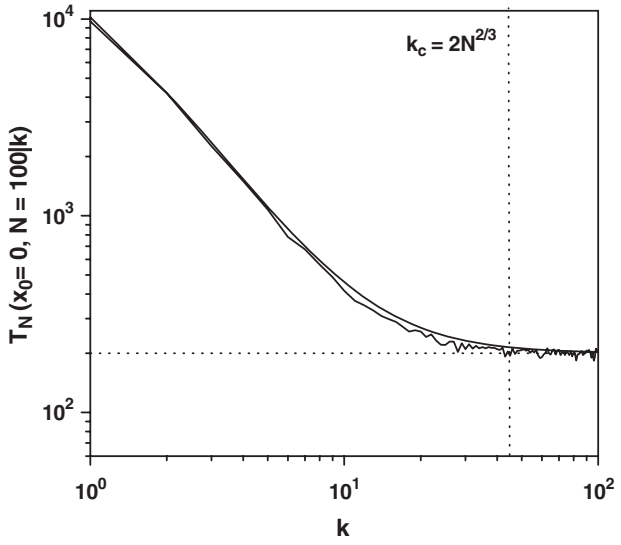

Fig. 1. MFPT of a random walker through the position $x=N$ from the domain [0, N] starting from $x_0=0$ as the function of jump size k. Here $N=100$, the position $x=0$ is the reflecting boundary and only the position $x=N$ is the absorbing boundary. Total number of random walk trajectories used to compute the MFPT is 10^5. The vertical dotted line at $k_c=2N^{2/3}\approx 44$ denotes the critical jump size beyond which the MFPT is almost independent of the jump size k. The solid line is the theoretical prediction by Eq. (9).

potential such as $f(x)=-\beta x$, the MFPT required for escaping from the domain [x_0, N] can be easily derived from the corresponding stochastic differential equation $\dot{x}=\beta+\Gamma(t)$. Here $\langle\Gamma(t)\Gamma(t)\rangle=D$ and the Fokker–Plank equation now becomes,

$$\partial_t P(x,t) = -\beta\partial_x P(x,t) + (D/2)\partial_x^2 P(x,t). \tag{10}$$

Since the MFPT (T) obeys the backward FPE that is given as $\beta d_x T + (D/2)d_x^2 T = -1$, using the reflecting boundary condition at $x=0$ as $d_x T|_{x=0}$ and absorbing boundary condition at $x=N$ as $T(N)=0$ we get,

$$T(x_0=0,N) = \frac{1}{\beta}\left(N - \frac{D}{2\beta}\left(1 - e^{-\frac{2\beta N}{D}}\right)\right). \tag{11}$$

Under the condition that $\beta=1$ and assuming a pure one step unbiased random walk (i.e. $D=1$) one can easily show that $T(x_0=0, N)\approx N$, i.e. in presence of a linear potential directed towards the absorbing boundary, the MFPT required to escape form the domain [0, N] scales approximately linear with N as in the case of random jumps with jump size $k\geq k_c$. In other words random jump method of searching for the target site itself introduces an abstract linear potential when $k\geq k_c$. Nevertheless this abstract potential does not possess the uni-directionality, i.e. similar to the form $f(x)=-\beta|x|$ and from the comparison of Eqs. ((9) and (11)) one can easily guess that $\beta=\frac{1}{2}$ in the present context. Therefore we can conclude that under unbiased random jump conditions, existence for a free energy bias towards the specific-site on the DNA lattice is not necessary since the random jump method of searching itself introduces an abstract linear potential favoring the movement of the protein molecule towards the target site.

Another interesting phenomenon that we observe from Eq. (9) is that when $k\geq k_c$ the MFPT that is required to escape from the domain [0, N] is almost independent of the initial position x_0. However the critical jump size k_c is still a function of x_0. When $x_0\neq 0$, the MFPT that is given by Eq. (9) takes the form as follows.

$$T_N(x_0,N|k) = T(x_0,N|k) + 2N + \frac{3k(k+1)}{2(2k+1)}. \tag{12}$$

Here one should note that,

$$T(x_0,N|k) = 6\left(N^2 - x_0^2\right)[(k+1)(2k+1)]^{-1}. \tag{13}$$

And therefore the modified k_c takes the form as follows.

$$k_c(\tau) = 2[\tau(2N-\tau)]^{1/3}. \tag{14}$$

Here $\tau=|N-x_0|$ is the absolute distance between the specific-site and the initial position of the protein molecule on the DNA lattice. A plot of the function given by Eq. (14) is given in Fig. 2, which clearly shows that as the protein molecule moves closer to the target site the critical jump size that is required to maximize the target finding rate diminishes, i.e. the search efficiency increases. The requirement of higher jump sizes can be positively correlated to the amount of energy that is drawn by the protein molecule from the heat bath. If we consider two different positions of the protein molecule x_1 and x_2 such that $x_1<x_2<N$, it is obvious to note that the amount of energy required to attain the critical jump size at the position x_1 is higher than that of the energy required at the position x_2, which in turn generates an energetic directionality towards the target site (N). Moreover requirement of lesser magnitude of critical jump sizes towards the target site on the DNA lattice indirectly indicates that as the protein molecule moves closer to the target site the contribution of sliding (random walk with unit step size) increases.

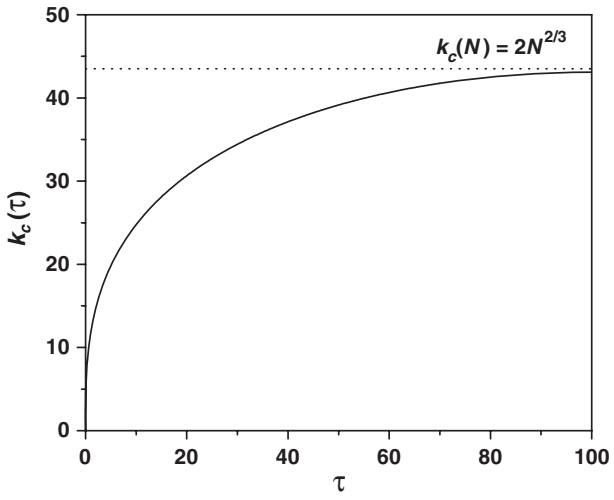

Fig. 2. The critical jump size required (Eq. (14)) to maximize the site-specific association rate of protein with DNA as function of the absolute initial distance between the protein molecule and the specific-site on the DNA lattice. Here $k_c(\tau)=2[\tau(2N-\tau)]^{1/3}$, $\tau=|N-x_0|$, and $N=100$.

4. Conclusions

The site-specific association of a protein molecule with a DNA lattice can be modeled as a one dimensional unbiased random jump process. Here we investigate the effect of the jump size k (in base-pairs) on the site-specific association rate. We show that there exists a critical jump size ($k_c \approx 2N^{2/3}$) beyond which site-specific association of a protein molecule with a DNA lattice cannot be facilitated and the maximum achievable association rate is predicted to be $\sim 10^{10}$ mol^{-1} s^{-1}. This critical jump size scales with the total length of DNA lattice (N) as $k_c \propto N^{2/3}$. Beyond k_c the mean first passage time MFPT (denoted as T) required for the protein molecule to target the specific-site follows a linear scaling law as $T \propto N$ rather than the usual $T \propto N^2$ scaling law. We also show that the random jump method of searching for the specific-site by the protein molecule on the DNA lattice itself introduce an abstract linear type potential favoring the site-specific association rate. We discuss the consequences of the existence of such critical jump sizes in the context of origin of super coiled structures of the genomic DNA in due course of evolution.

Acknowledgements

This work is supported by TIFR, Mumbai. Author is the recipient of Kanwal-Rekhi fellowship for career development. The author thanks the referees for useful suggestions and constructive comments.

References

[1] G. Adam, M. Delbruck (2001), Structural Chemistry and Molecular Biology, edited by A. Rich and N. Davidson Freeman, San Franscisco, CA, 1968, p. 198; Mark Ptashne and Alexander Gann, Genes and Signals Cold Spring Harbor Laboratory New York.

[2] J.A. McCammon, S. Harvey, Dynamics of Protein and Nucleic Acids, Cambridge University Press, Cambridge, England, 1980.

[3] J.A. McCammon, M. Karplus, Annu. Rev. Phys. Chem. 31 (1982) 29.

[4] J. Norberg, Association of protein–DNA recognition complexes: electrostatic and nonelectrostatic effects, Arch. Biochem. Biophys. 410 (2003) 48 (and references therein);
R.F. Bruinsma, Physics of protein–DNA interaction, Physica, A 313 (2002) 211.

[5] D.R. Lesser, M.R. Kurpiewski, L.J. Jacobson, The energetic basis of specificity in the Eco RI Endonuclease–DNA interaction, Science 250 (1990) 776.

[6] P.H. Von Hippel, Protein–DNA recognition: new perspectives and underlying themes, Science 263 (1994) 769;
Y. Jia, A. Kumar, S.S. Patel, Equilibrium and stopped-flow kinetic studies of interaction between T7 RNA polymerase and its promoters measured by protein and 2-aminopurine fluorescence changes, J. Biol. Chem. 271 (1996) 30451;
P.H. Von Hippel, Facilitated target location in biological systems, J. Biol. Chem. 264 (1989) 675 (and references therein);
P.H. Von Hippel, O.G. Berg, On the specificity of DNA–protein interactions, Proc. Natl. Acad. Sci. U.S.A. 83 (1986) 1608.

[7] J.F. Schildbach, A.W. Kazak, B.E. Raumann, R.T. Sauer, Origins of DNA-binding specificity: role of protein contacts with the DNA backbone, Proc. Natl. Acad. Sci. U.S.A. 96 (1999) 811;
P. Etchegoin, M.A. Nollmann, A model for protein–DNA interaction dynamics, J. Theor. Biol. 220 (2003) 233.

[8] B. Lewin, in: Genes, vol. VII, Oxford University Press, London, 2000, p. 243.

[9] O.G. Berg, R.B. Winter, P.H. Von Hippel, Diffusion-driven mechanisms of protein translocation on nucleic acids: 1. Models and theory, Biochemistry 20 (1981) 6929 (and references therein).

[10] M. Barbi, C. Place, V. Popkov, M. Salerno, Base-sequence-dependent sliding of proteins on DNA, Phys. Rev., E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics 70 (2004) 041901;
T. Tlusty, R. Bar-Ziv, A. Libchaber, High-fidelity DNA sensing by protein binding fluctuations, Phys. Rev. Lett. 93 (2004) 258103.

[11] R. Murugan, DNA protein interactions under random jump conditions, Phys. Rev., E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics 69 (2004) 011911.

[12] C.W. Gardiner, Hand Book of Stochastic Methods, in: H. Haken, (Ed.), Springer-Verlag, Berlin, 1983, p. 260.