Mean first passage time calculation for the random walk with random step size

2006 J. Phys. A: Math. Gen. 39 1575

(http://iopscience.iop.org/0305-4470/39/7/004)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **39** (2006) 1575–1580 doi:10.1088/0305-4470/39/7/004

Mean first passage time calculation for the random walk with random step size

R Murugan

Department of Chemical Sciences, Tata Institute of Fundamental Research, Homi Bhabha Road, Colaba, Mumbai 400 005, India

E-mail: muruga@tifr.res.in or rmurugan@gmail.com

Received 10 October 2005, in final form 5 January 2006
Published 1 February 2006
Online at stacks.iop.org/JPhysA/39/1575

Abstract

We derive an expression for the mean first passage time (MFPT) for the random walk with random step size on a one-dimensional linear lattice. Here both ends of the linear lattice are reflecting boundaries whereas the absorbing boundary is situated anywhere in between. When the size of the lattice is N and the random step size is k, we show that the MFPT ($\overleftrightarrow{T}$) associated with the escape of the random walker only through a specific point that is situated anywhere in the interval $[0, N]$ at the limit as $k \to \infty$ is $\lim_{k\to\infty} \overleftrightarrow{T}_a = N$ which is independent of the initial position as well as the absorbing point associated with the random walker on the linear lattice under consideration. This result has potential applications in the analysis and understanding of the fundamental processes in molecular biology such as DNA–protein and DNA–probe interactions.

PACS numbers: 05.40.Fb, 87.10.+e, 87.15.Aa, 87.14.Gg

1. Introduction

Finding a specific site on a DNA lattice in the presence of an enormous amount of non-specific sites by a protein, another stretch of DNA or by a stretch of RNA is a fundamental phenomenon in molecular biology [1–5]. Here the protein molecule first non-specifically binds to the DNA lattice and then performs a one-dimensional search for the specific site by unbiased random jumps with a jump size of k base pairs (bps) where the jump size k can be directly correlated with the degree of super-coiling of the DNA lattice under consideration [6].

Let us consider a DNA lattice of N bps in length, containing the specific site at the lattice position a such that $0 < a < N$, where the set of lattice points $\{0, N\}$ constitutes (these are the helical ends) the reflecting boundaries and the lattice point $x = a$ is the only absorbing boundary (i.e. the specific site). Let us assume that the protein molecule was at the lattice position $x = x_0$ at time $t = 0$, and currently searching for the specific site by unbiased random jumps with jump size k bps i.e. starting from a position x, in the next step the protein molecule

can be found anywhere in the interval $x \pm k$ with equal probabilities (which is equal to $\frac{1}{2k}$ in the present context). Now the probability of finding the protein molecule on the DNA lattice can be described by the following birth–death master equation:

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

Here $P_{x,t}$ is the probability of finding the protein molecule at the position x at time t. The Fokker–Plank equation (FPE) associated with the master equation (1) is simply given as [7, 8] follows:

$$\partial_t P_{x,t} = \frac{D_k}{2} \partial_x^2 P_{x,t}, \tag{2}$$

where $D_k = \frac{1}{k}\sum_{i=1}^{k} i^2 = \frac{(k+1)(2k+1)}{6}$ is the one-dimensional phenomenological diffusion coefficient in the dimensionless form [6–8]. The mean first passage time (MFPT) (T_x) associated with the escape of the protein molecule through the specific site a, can be easily derived from [7–10] the backward FPE, i.e.

$$\mathrm{d}_x^2 T_x = -\frac{2}{D_k}. \tag{3}$$

When the jump size $k = 1$ (therefore $D_k = 1$) the MFPTs can be derived as follows. If the initial position x_0 is such that $0 \leqslant x_0 \leqslant a$, the MFPT is given as $T_{LR,x_0} = a^2 - x_0^2$. If the initial position x_0 is such that $a \leqslant x_0 \leqslant N$ the MFPT is given as $T_{RL,x_0} = \left(a^2 - x_0^2\right) + 2N(x_0 - a)$. Here the reflecting boundary conditions are $\mathrm{d}_x T_x|_{x=0} = \mathrm{d}_x T_x|_{x=N} = 0$ and the absorbing boundary condition is $T_x|_{x=a} = 0$ (see [7–10]). However, when $k > 1$, though the initial position of the protein molecule on the DNA lattice is such that $0 \leqslant x_0 < a$, there is a definite probability associated with the protein molecule to escape from the interval $[0, a-1]$ into the interval $[a+1, N]$ without actually getting absorbed at the lattice position $x = a$. Since we have totally three boundary conditions, neither the master equation (1) nor the corresponding MFPT equation (3) can be solved analytically. Nevertheless, such solutions are very much useful in the analysis as well as in the understanding of the fundamental processes in molecular biology such as DNA–protein interactions. In this paper, we derive a solution to this MFPT problem.

2. MFPT from the interval $[0, a-1]$ only through the point $x = a$

Now let us consider only the interval $[0, a-1]$ and let us compute the MFPT associated with the protein molecule to escape only through the point $x = a$. We consider M number of trajectories starting from the position $x = x_0$ where the initial position of the protein molecule on the DNA lattice x_0 is such that $0 \leqslant x_0 \leqslant a$. When the jump size is $k = 1$, then we should note that all the M number of trajectories will pass through the absorbing point $x = a$ and thus get absorbed. However, when the jump size is $k > 1$ only a fraction of the trajectories will pass through the point $x = a$ before the protein molecule escapes into the region $[a+1, N]$ which can be computed as follows. Let us assume that the present position of the protein molecule on the DNA lattice is $x = a - 1$. Now in the next step, the position of the protein molecule can be anywhere in the interval $(a - 1 - k) \leqslant x \leqslant (a - 1 + k)$. Here one should note that among $2k$ number of such possible positions of the protein molecule, k number of possibilities lie again in the interval $[0, a-1]$ that will not contribute to the MFPT. In the remaining k number of possible positions, the probability associated with the protein molecule to get absorbed at the position $x = a$ is $\frac{1}{k}$, i.e. the number of possibilities of getting absorbed

at the lattice position $x = a$ is unity and $k - 1$ number of possibilities will be in the interval $[a + 1, a + k]$. Since we insist the condition that the protein molecule should pass only through the lattice point $x = a$, the trajectories which end in the interval $[a + 1, a + k]$ have to travel long routes before actually they come back and get absorbed at the lattice point at $x = a$. For this reason, the time that is taken by the trajectories which hit at the interval $[a + 1, a + k]$ will simply add up to $T_{LR,0}$, however, with appropriate weighting. In other words, among M number of trajectories, $\frac{M}{k}$ number of trajectories will end at the position $x = a$ (and get absorbed) with a MFPT of $T_{LR,0} = \frac{a^2-x_0^2}{D_k}$ and $M\frac{i}{k}$ number of trajectories will end in the interval $[a + 1, a + i]$ with MFPTs of $T_{LR,i} = \frac{(a+i)^2-x_0^2}{D_k}$ which is due to the fact that the jump size i in turn includes all the possibilities of the jumps $1, 2, 3, \ldots, i$ too. When we insist the condition that the protein molecule should escape only through the lattice point $x = a$, those MFPTs ($T_{LR,i}$) associated with the intervals $[a + 1, a + i]$ where $i = 1, 2, 3 \ldots, k$ simply add up to the MFPT ($T_{LR,0}$) with the weighting factors $\mu_i = \frac{i}{k}$. Now the excess MFPT that adds up to $T_{LR,0}$ can be computed as follows. Since $T_{LR,i} = T_{LR,0} + \frac{i^2+2ai}{D_k} = T_{LR,0} + \theta_i$, the MFPT ($\vec{T}_a$) associated with the escape of the protein molecule only through the lattice position $x = a$ is given by the weighted sum $\vec{T}_a = T_{LR,0} + \sum_{i=0}^{k} \mu_i \frac{i^2+2ai}{D_k} = T_{LR,0} + \sum_{i=0}^{k} \mu_i \theta_i$, where the weighting factor is $\mu_i = \frac{i}{k}$ and $T_{LR,0} = \frac{a^2-x_0^2}{D_k}$ and θ_i is defined as $\theta_i = \frac{i^2+2ai}{D_k}$. Here, we should note that the weighting is done only on the terms which contain the jump size variable i since $T_{LR,0}$ is a constant and it is independent of the jump size variable i. Substituting the value of D_k into the sum $\sum_{i=0}^{k} \mu_i \theta_i$ one finally obtains the expression for $\vec{T}_a$ as follows:

$$\vec{T}_a = T_{LR,0} + \sum_{i=0}^{k} \mu_i \theta_i = T_{LR,0} + 2a + f(k), \tag{4}$$

where $f(k)$ is defined as $f(k) = \frac{3k(k+1)}{2(2k+1)}$.

3. Splitting probabilities and the overall MFPT

Here one should note that $\vec{T}_a$ is the mean time for which the protein molecule stays in the interval $[0, a - 1]$ before it gets absorbed at the lattice position $x = a$. Now we compute the mean time for which the protein molecule stays in the interval $[a + 1, N]$ once it escaped from the interval $[0, a - 1]$. We should note that from the position $x = a - 1$, the protein molecule can jump into the lattice positions $x = a + i$ where $i = 1, 2, \ldots, k$, without getting absorbed at a. In other words, the initial positions associated with the interval $[a + 1, N]$ are $x_0 = a + 1, a + 2, \ldots, a + k$. Here the MFPT associated with the protein molecule to escape from the interval $[a + 1, N]$ is given as $T_{RL,x_0} = D_k^{-1}\left[\left(a^2 - x_0^2\right) + 2N(x_0 - a)\right]$. We should note that the initial position x_0 falls in the interval $[a + 1, a + i]$ where $i = 1, 2, 3, \ldots, k$ with the probabilities of $\frac{i}{k}$. Suppose if we consider Q number of such trajectories, we can easily conclude that $Q\frac{i}{k}$ number of trajectories will end in the interval $[a + 1, a + i]$. Therefore to compute the overall MFPT ($\overleftarrow{T}_a$) associated with the escape of the protein molecule from the interval $[a + 1, N]$ one needs to sum T_{RL,x_0} over the initial positions $x_0 = a + 1, a + 2, \ldots, a + i, \ldots, a + k$ with the appropriate weighting factors. Noting the fact that the MFPT ($T_{RL,i}$) associated with the escape of the protein molecule from the interval $[a + 1, N]$ with the initial position at $x_0 = a + i$ as $T_{RL,x_0} = D_k^{-1}[a - (a + i)^2 + 2N(a + i - a)]$ and summing over the terms which contain the jump size variable i with the weighting factors $\mu_i = \frac{i}{k}$ similar to the derivation of equation (4), i.e. $D_k^{-1} \sum_{i=0}^{k} \mu_i [2(N - a)i - i^2]$, one obtains the overall MFPT ($\overleftarrow{T}_a$) associated

with the escape of the protein molecule from the interval $[a+1, N]$ starting anywhere from the interval $[0, a-1]$, i.e. $0 \leqslant x_0 \leqslant a$, as follows:

$$\overleftarrow{T}_a = T_{LR,0} + 2(N-a) - f(k). \tag{5}$$

Here the term $T_{LR,0} = \frac{a^2 - x_0^2}{D_k}$ in equation (5) is added to account for the time that is taken by the protein molecule to enter into the interval $[a+1, N]$ from the interval $[0, a-1]$ since the initial position of the protein molecule is still assumed to be situated in the interval $[0, a-1]$. Now we consider a total number of R trajectories starting from the interval $0 \leqslant x_0 \leqslant a$ among which M number of trajectories get absorbed at $x = a$ from the interval $[0, a-1]$ and Q number of trajectories get absorbed at $x = a$ from the interval $[a+1, N]$. It is obvious to note that though R is a constant in the present context, M and Q are functions of the jump size k. When $k = 1$, it is obvious to note that $M = R$ and $Q = 0$. However, when the jump size associated with the protein molecule on the DNA lattice is such that $k > 1$ then $M < R$ and $Q > 0$. For an arbitrary value of k one can compute the values of M and Q as follows. Let us assume that the present position of the protein molecule on the DNA lattice is $x = a - 1$. In the next step, the protein molecule can be found anywhere in the interval $(a-1-k) \leqslant x \leqslant (a-1+k)$ with equal probabilities of $\frac{1}{2k}$. Among $2k$ number of such possibilities $k-1$ number of possibilities will be positioned in the interval $[a+1, N]$ and $k+1$ number of possibilities will stay in the interval $[0, a]$. In other words, the probability of escape of the protein molecule into the interval $[a+1, N]$ without getting absorbed at the lattice point at $x = a$ is $\frac{k-1}{2k}$ and obviously the probability of the protein molecule to stay in the interval $[0, a-1]$ is $\frac{1}{2}$ and the probability to get absorbed at the lattice point $x = a$ is $\frac{1}{2k}$, i.e. the total probability associated with the protein molecule to escape into the lattice point $x = a$ from the interval $[0, a-1]$ is $\frac{k+1}{2k}$. From these results one can easily show that the values of M and Q are as $M = R\left(\frac{1}{2} + \frac{1}{2k}\right)$ and $Q = R\left(\frac{1}{2} - \frac{1}{2k}\right)$ where $\vec{p}_a = \left(\frac{1}{2} + \frac{1}{2k}\right) = \frac{M}{R}$ and $\overleftarrow{p}_a = \left(\frac{1}{2} - \frac{1}{2k}\right) = \frac{Q}{R}$ are the splitting probabilities associated with the entry of the protein molecule from the intervals $[0, a-1]$ and $[a+1, N]$ into the absorbing point $x = a$, respectively. It is obvious to note that when the jump size is $k = 1$ the splitting probabilities are given as $\vec{p}_a = 1$ and $\overleftarrow{p}_a = 0$. However, when the jump size increases as $k \to \infty$ we can easily conclude that $\lim_{k\to\infty}[\vec{p}_a = \overleftarrow{p}_a] = \frac{1}{2}$ which is independent of the initial position x_0 as well as the position of the absorbing point $x = a$. Now using these splitting probabilities one can write the expression for overall MFPT taken by the protein molecule to get absorbed at the position $x = a$ starting from the interval $0 \leqslant x_0 < a$ as follows:

$$\overleftrightarrow{T}_a = \vec{p}_a \vec{T}_a + \overleftarrow{p}_a \overleftarrow{T}_a = T_{LR,0} + N - \frac{N}{k} + \frac{2a}{k} + \frac{3(k+1)}{2(2k+1)}. \tag{6}$$

Here we should note that $\lim_{k\to\infty} \overleftrightarrow{T}_a = N$ and in the limit as $k \to 1$ we recover the relation $\lim_{k\to 1} \overleftrightarrow{T}_a \approx T_{L,x_0}$. Let us consider a circular DNA lattice of N bps (e.g. a plasmid) containing a specific site at the position $x = a$. The circular lattice is a special form of the linear lattice where the lattice position $x = 0$ and the lattice position $x = N$ are tied together. Equations (4)–(6) are still valid for a circular lattice since the probability flow from the interval $[0, a-1]$ into the interval $[a+1, N]$ through the lattice point $x = 0$ is exactly compensated by the probability flow from the interval $[a+1, N]$ into the interval $[0, a-1]$ through the lattice point $x = N$. In fact, the limit for the circular lattice $\lim_{k\to\infty} \overleftrightarrow{T}_a = N$ is consistent with the earlier estimated value [11]. Here one should note that unlike the case of a linear lattice, equation (1) (and equation (3) too) can be analytically solved for a circular lattice using periodic boundary conditions [11].

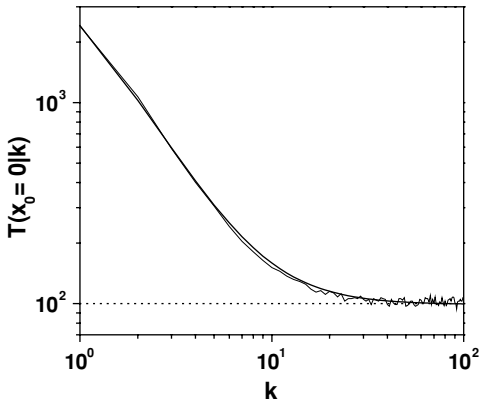

Figure 1. Variation of the MFPT as a function of the jump size k, i.e. $\overleftrightarrow{T}_{50} = T\ (x_0 = 0|k)$. Here $N = 100$, $x_0 = 0$ and $a = 50$ and the jump size k was varied in the range $1 \leqslant k \leqslant 100$. The MFPT was calculated over 10^5 trajectories. Here the solid line is the prediction by equation (6).

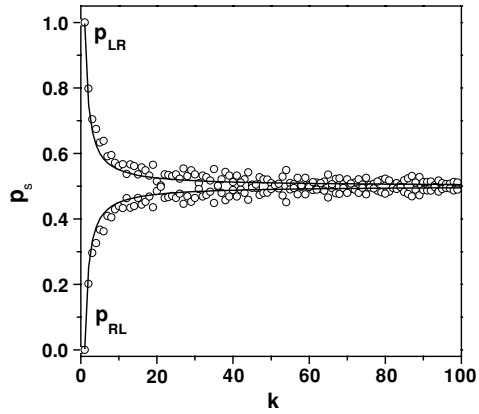

Figure 2. Variation of the splitting probabilities (p_s, hollow circles) associated with the escape of the random walker from intervals [0, 49] and [51, 100] into the absorbing point $x = 50$ as a function of the jump size k which are in agreement with the predictions (solid lines) of $p_{RL} = \overleftarrow{p}_a = \frac{k-1}{2k}$ and $p_{LR} = \vec{p}_a = \frac{k+1}{2k}$.

4. Random walk simulation results and discussion

To check the validity of equation (6), random walk simulations were carried out. Here the parameters were $N = 100$, $x_0 = 0$ and $a = 50$ and the jump size k was varied in the range $1 \leqslant k \leqslant 100$. Figure 1 shows that the mean number of steps (MFPT in dimensionless form) was calculated over 10^5 trajectories as a function of the step size k which fairly agrees with the prediction by equation (6). Figure 2 shows the splitting probabilities associated with the random walker to escape into the lattice position $x = 50$ from the intervals [0, 49] and [51, 100] as a function of the jump size k which are in agreement with the predictions as $\overleftarrow{p}_a = \frac{k-1}{2k}$ and $\vec{p}_a = \frac{k+1}{2k}$. Here the limit $\lim_{k\to\infty} \overleftrightarrow{T}_a = N$ indicates that as $k \to \infty$, the MFPT associated with the escape of the protein molecule at the lattice point $x = a$ is not only independent of the initial position $x = x_0$ but also independent of the position of the absorbing point $x = a$

itself. Suppose, instead of a single point, let us consider a stretch of absorbing points such that $a - \delta \leqslant x \leqslant a + \delta$. Here the interval δ can be thought to be inversely proportional to the specificity of the absorbing site whereas the inverse of MFPT can be thought to be inversely proportional to the affinity of the protein molecule towards the specific site. Therefore, it is obvious to note that $\lim_{k\to\infty} \overleftrightarrow{T}_{a,\delta} = \frac{N}{2\delta+1}$ which is a straightforward demonstration of the existence of affinity specificity anti-correlation in DNA–protein interactions [12] even at higher jump sizes.

5. Conclusions

In this paper, we have derived the expression for the mean first passage time for a random walk with random step size on a one-dimensional linear lattice with reflecting boundaries at the ends and an absorbing boundary anywhere inside the lattice. Here we consider a random walker undergoing an unbiased random jump motion with a random step size of k on a linear lattice with N number of points where the set of points $\{0, N\}$ constitutes the reflecting boundaries and the lattice point a such that $0 < a < N$ is the only absorbing boundary. We show that the MFPT $\overleftrightarrow{T}_a$ associated with the escape of the random walker starting from the lattice position x_0 such that $0 \leqslant x_0 \leqslant a$ at the time $t = 0$ only through the point a is $\overleftrightarrow{T}_a = T_{LR,0} + N - \frac{N}{k} + \frac{2a}{k} + \frac{3(k+1)}{2(2k+1)}$, where $T_{LR,0}$ is the MFPT associated with the jump size of $k = 1$ from which we obtain the limit $\lim_{k\to\infty} \overleftrightarrow{T}_a = N$. Unlike $\overleftrightarrow{T}_a$ this limiting value is independent of the initial position x_0 as well as the absorbing position a associated with the random walker under consideration. We generalize this result to a circular lattice. Finally, we discuss the potential applications of these results in the analysis and understanding of the fundamental processes in molecular biology such as DNA–protein and DNA–probe interactions.

Acknowledgments

This work is supported by TIFR, Mumbai. The author is a recipient of Kanwal-Rekhi fellowship for career development.

References

[1] McCammon J A and Karplus M 1982 *Annu. Rev. Phys. Chem.* **31** 29
[2] Bruinsma R F 2002 *Physica* A **313** 211
[3] Von Hippel P H 1994 *Science* **263** 769
[4] Berg O G, Winter R B and Von Hippel P H 1981 *Biochemistry* **20** 6929
[5] Berg O G, Winter R B and Von Hippel P H 1981 *Biochemistry* **20** 6961
[6] Murugan R 2004 *Phys. Rev.* E **69** 011911
[7] Gardiner C W 1985 *Handbook of Stochastic Methods* (Berlin: Verlag) p 246
[8] Van Kampen N G 1992 *Stochastic Processes in Physics and Chemistry* (Amsterdam: Elsevier) p 244
[9] Feller W 1968 *An Introduction to Probability Theory and its Applications* vol 1 (New York: Wiley)
[10] Redner S 2001 *A Guide to First Passage Processes* (London: Cambridge University Press)
[11] Lindenberg K L and Shuler K E 1971 *J. Math. Phys.* **12** 633
[12] Murugan R 2005 *Biophys. Chem.* **116** 105