LETTER TO THE EDITOR

Mean first passage time calculation for a one-dimensional random walk with random absorbing boundary

R Murugan

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

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

Abstract
In this letter, we exactly solve the MFPT problem for a one-dimensional random walk with random step size where there are two fixed reflecting boundaries and a fluctuating absorbing boundary in between. We discuss the consequences of these results in the context of site-specific DNA–protein and DNA–probe interactions.

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

The site-specific association of a protein molecule or a small stretch of a DNA molecule (the primer) with a large size DNA lattice (the template DNA) can be modelled as a random walk with random step size [1–3]. According to the two-step model [4, 5], the protein molecule or the primer molecule first non-specifically binds to the template DNA and then searches for the specific site via unbiased random jump motions on the template DNA under a non-specifically bound condition. Recently we have shown [6] that under random jump situations, when we insist the condition that the protein molecule should escape only through the specific site which is lying inside the template DNA lattice, there exists a limit for the associated mean first passage time (MFPT, denoted as T) as $\lim_{k \to \infty} T = N$ where N is the total size of the template DNA (in base pairs) and k is the jump size (in base pairs) associated with the dynamics of the protein molecule on the template DNA. The limit $\lim_{k \to \infty} T = N$ clearly indicates that by the random jump motion of the protein molecule on the template DNA the maximum achievable site-specific association rate r_m is $r_m = N^{-1}$ and to increase the rate beyond this limit either the specificity associated with specific site needs to be decreased or an external free energy input in the form of ATP (adenosine triphosphate) should be provided.

On the other hand, almost all the earlier two-step models on DNA–DNA or DNA–protein interactions invariably assumed that the position of the specific site on the template DNA is a fixed quantity which is an oversimplification of the underlying real process. This is due to the

fact that the template DNA is a long polymer chain compared to that of the dimensions of the protein molecule. Moreover, under solution conditions the template DNA is also undergoing independent conformational fluctuations that in turn cause perturbations in the relative position of the specific site on the template DNA with respect to the protein or the primer molecule. Therefore, it will be of great interest to know the effect of such fluctuations in the absorbing boundary on the MFPT associated with the escape of the protein molecule only through the specific site under random jump conditions.

Let us consider a DNA lattice of N base pairs in length which possesses the specific site at the lattice position a such that $0 < a < N$. There is a protein molecule which was at the lattice position $x = x_0$ at time $t = 0$ and currently searching for the specific site a via unbiased random jump motion with a jump size of k base pairs, i.e. starting from the lattice position x, in the next step the protein molecule can be found anywhere in the interval $[x - k, x + k]$ with equal probabilities, i.e. $1/2k$. The Fokker–Planck equation [7–11] associated with the dynamics of the protein molecule is given as $\partial_t P = 2^{-1} D_x \partial_x^2 P$. Here we should note that there are reflecting boundaries at the lattice points $\{0, N\}$, i.e. $\partial_x P|_{x=0} = \partial_x P|_{x=N} = 0$ and the only absorbing boundary at the lattice point (a), i.e. $P|_{x=a} = 0$ for the protein molecule. Here $D_x = 6[(k+1)(2k+1)]^{-1}$ is the one-dimensional phenomenological diffusion coefficient in the dimensionless form and $P = P(x, t \mid x_0, 0)$ is the probability associated with the protein molecule to be found at the lattice point x at time t where the initial condition is $P(x, 0 \mid x_0, 0) = \delta(x - x_0)$. Now the MFPT associated with the escape of the protein molecule from the interval $[0, N]$ satisfies the backward FPE as $2^{-1} D_x \partial_x^2 T_x = -1$ where the boundary conditions are $d_x T_x|_{x=0} = d_x T_x|_{x=N} = 0$ and $T_x|_{x=a} = 0$. Unfortunately, the differential equation for the MFPT (T_x) with three boundary conditions cannot be solved analytically. However, recently we have given [6] an approximate solution to this problem by the weighted averaging method. The brief idea is as follows. The entire lattice interval $[0, N]$ is broken into two subintervals $[0, a - 1]$ and $[a + 1, N]$. Now, let us consider M number of trajectories starting from the lattice point $x = 0$ at time $t = 0$. When the jump size associated with the dynamics of the protein molecule is $k > 1$, then among M number of such trajectories Q number of trajectories will hit the absorbing point $x = a$ from the subinterval $[0, a - 1]$ with a weighted MFPT of T_L and R number of trajectories will hit the absorbing boundary from the subinterval $[a + 1, N]$ with a weighted MFPT of T_R. Finally the overall MFPT (T) is computed as $T = M^{-1}[Q T_L + R T_R]$. Here $p_R = M^{-1} R$ and $p_L = M^{-1} Q$ are the splitting probabilities associated with the entry of the protein molecule from the interval $[0, a - 1]$ and from the interval $[a + 1, N]$ into the absorbing point $x = a$, respectively. Using simple arguments [6] one can show that $p_R = 2^{-1}[1 + k^{-1}]$ and $p_L = 2^{-1}[1 - k^{-1}]$.

Now let us drop the assumption that the absorbing point is fixed at $x = a$ and we assume that it fluctuates in the interval $[N_l, N_m]$ where the interval lies inside the interval $[0, N]$ such that $0 < N_l < N_m < N$. In other words the absorbing boundary (a) is undergoing a random jump motion in the interval $[N_l, N_m]$ with a jump size of $\delta = [N_m - N_l]$ with equal probabilities. To compute the MFPT associated with the escape of the protein molecule starting from the lattice point $x = 0$ at time $t = 0$, we break the interval $[0, N]$ into three subintervals namely $[0, N_l - 1]$, $[N_l, N_m]$ and $[N_m + 1, N]$. Let us assume that the jump size associated with the dynamics of the protein molecule is such that $k < [N_m - N_l]$ and we consider totally M number of trajectories starting from the lattice point $x = 0$ at time $t = 0$ among which Q number of trajectories will be absorbed at the subinterval $[N_l, N_m]$ from the subinterval $[0, N_l - 1]$ and R number of trajectories will be absorbed at the subinterval $[N_l, N_m]$ from the subinterval $[N_m + 1, N]$. First let us consider only the interval $[0, N_l - 1]$. Upon escaping from the subinterval $[0, N_l - 1]$, the protein molecule can hit any point in the subinterval $[N_l, N_m]$ and may get absorbed or may escape into the subinterval $[N_m + 1, N]$

(even when $k < [N_m - N_1]$) since the absorbing boundary fluctuates inside the subinterval $[N_1, N_m]$.

Suppose let us assume that the current position of the protein molecule is $x = N_1$. In the next step it can be found anywhere in the interval $[N_1 - k, N_1 + k]$ with equal probabilities. Among $2k$ number of such possible positions, $k - 1$ number of possibilities will lie again in the interval $[0, N_1 - 1]$ and $k + 1$ number of possibilities will lie inside the absorbing subinterval $[N_1, N_m]$ and therefore may get absorbed inside or move to the subinterval $[N_m + 1, N]$. In other words, the splitting probability associated with the entry of the protein molecule from the subinterval $[0, N_1 - 1]$ into the absorbing interval $[N_1, N_m]$ is $p_1 = (2k)^{-1}[k - 1] = 2^{-1}[1 - k^{-1}]$ and obviously the splitting probability associated with the entry of the protein molecule from the subinterval $[N_m + 1, N]$ into the absorbing interval $[N_1, N_m]$ is $p_3 = 2^{-1}[1 + k^{-1}]$. Now using these splitting probabilities one can write $Q = p_1 M$ and $R = p_3 M$ where $M = P + Q$ is the total number of trajectories started from the lattice position $x_0 = 0$ at time $t = 0$. Here we should recall the fact that when $[N_m - N_1] = \delta = 0$, i.e. no fluctuations in the absorbing boundary, then the situation will reverse and the splitting probabilities become $_0p_1 = 1 - p_1 = 2^{-1}[1 + k^{-1}]$ and $_0p_3 = 1 - p_3 = 2^{-1}[1 - k^{-1}]$ as we have shown [6] earlier in the case of a fixed absorbing boundary at the lattice position $x = a$.

Now we compute the corresponding MFPTs. The MFPT (T_1) associated with the escape of the protein molecule from the interval $[0, N_1 - 1]$ into the interval $[N_1, N_1 + i]$ is simply given as $T_{1,i} = D_x^{-1}(N_1 + i)^2$. The probability μ_i associated with the hitting of the protein molecule inside the interval $[N_1, N_1 + i]$ is simply given as $\mu_i = \frac{i}{k}$, which is the weighting factor here. Now using this weighting factor, noting the initial condition $x_0 = 0$, one can compute the MPFT associated with the escape of the protein molecule from the interval $[0, N_1 - 1]$ into the subinterval $[N_1, N_m]$ only through the lattice point $x = N_1$ as follows:

$$T_1 = D_k^{-1}N_1^2 + D_k^{-1}\sum_{i=1}^{k}\mu_i[i^2 + 2N_1 i] = D_k^{-1}N_1^2 + 2N_1 + f(k), \qquad (1)$$

where $f(k)$ is defined as $f(k) = \frac{3(k+1)}{2(2k+1)}$ and the weighting factor μ_i is defined as $\mu_i = \frac{i}{k}$. Here T_1 is the average residence time for which the protein molecule stays in the interval $[0, N_1 - 1]$ before it escapes into the subinterval $[N_1, N_m]$.

Now we compute the total time T_2 that the protein molecule takes to travel from the lattice point $x_0 = 0$ and enters inside the interval $[N_1, N_m]$ and stays there if not absorbed. Here we should note that under random jump conditions, from the interval $[0, N_1 - 1]$ the protein molecule can jump into the interval $[N_1 + i, N_m]$ where $i = 1, 2, 3 \ldots k$ with a probability of $\mu_i = \frac{i}{k}$, i.e. the initial positions associated with the dynamics of the protein molecules inside the interval $[N_1, N_m]$ are $N_1 + i$ where $i = 1, 2, 3 \ldots k$. Keeping this in mind, the average residence time associated with the protein molecule to stay in the subinterval $[N_1, N_m]$ (if it is not absorbed) can be calculated as follows:

$$T_2 = D_x^{-1}N_1^2 + D_x^{-1}\sum_{i=1}^{k}\mu_i[(N_m - N_1)i - i^2] = D_x^{-1}N_1^2 + (N_m - N_1) - f(k). \qquad (2)$$

Here the term $D_x^{-1}N_1^2$ in equation (2) is added to account for the time that is spent by the protein molecule in the interval $[0, N_1 - 1]$ before it enters into the interval $[N_1, N_m]$. Similarly starting from the lattice point $x_0 = 0$, the MFPT (T_3) associated with the escape of the protein

molecule from the interval $[N_m + 1, N]$ only through the lattice point $x = N_m$ into the interval $[N_1, N_m]$ can be calculated as follows:

$$\left.\begin{aligned}
T_3 &= D_k^{-1}N_1^2 + (N_m - N_1) + D_k^{-1}\sum_{i=1}^{k}\mu_i[2(N - N_m)i - i^2] \\
&= D_k^{-1}N_1^2 + (N_m - N_1) + 2(N - N_m) - f(k)
\end{aligned}\right\} . \qquad (3)$$

Here the term $D_k^{-1}N_1^2$ is added to T_3 to account for the time that is spent by the protein molecule in the interval $[0, N_1 - 1]$ and the term $(N_m - N_1)$ is added to account for the time for which the protein molecule stays in the interval $[N_1, N_m]$ (if not absorbed inside the interval $[N_1, N_m]$) before it enters the interval $[N_m + 1, N]$.

Now under the condition that $[N_m - N_1] > 0$ and using the splitting probabilities p_1 and p_3 one can easily compute the overall MFPT (T) associated with the protein molecule to escape only through the absorbing point which fluctuates inside the subinterval $[N_1, N_m]$ as follows:

$$T = p_1 T_1 + p_3 T_3 = D_k^{-1}N_1^2 + \frac{N_1}{2} + N - \frac{N_m}{2} + \frac{1}{k}\left(N - \frac{N_m}{2} - \frac{3N_1}{2} - f(k)\right). \qquad (4)$$

However here one should note that when $[N_m - N_1] = 0$, equation (4) becomes $T = {}_0p_1T_1 + {}_0p_3T_3$. Equation (4) is the central result of this letter from which we can compute various limiting values as follows.

Case 1. When $k = 1$, $N_m = N$ and $N_1 = 0$, i.e. when the fluctuating absorbing boundary covers the entire interval $[0, N]$, equation (4) reduces to $T \approx N$.

Case 2. Similarly when $k \to \infty$, $N_m = N$ and $N_1 = 0$, equation (4) reduces to $\lim_{k\to\infty} T \approx \frac{N}{2}$. This is an important result as far as the DNA–protein interactions are concerned. We should recall the fact that when the protein molecule searches for the specific site on the DNA lattice by a random jump method then the maximum achievable site-specific association rate is $r_m = N^{-1}$. In this context the current results clearly suggest that this limiting rate r_m can be doubled, i.e. $r_m = 2N^{-1}$ by introducing fluctuations in the relative position of the specific site on the template DNA with respect to the protein molecule.

Case 3. When $N_m = N$ and $k = 1$, then we obtain $T \approx D_k^{-1}N_1^2 + N - N_1$. Now if we define $N_1 = N - \delta$ where $\delta = N - N_1$ is the fluctuating interval associated with the absorbing boundary, then we obtain the relation for the overall MFPT as $T \approx D_k^{-1}[N^2 + \delta^2 - 2N\delta] + \delta$. Simple random walk simulations in fact prove the validity of this result (figure 1).

Case 4. When $N_m = N_1$, i.e. $[N_m - N_1] = \delta = 0$ then from the relation $T = {}_0p_1T_1 + {}_0p_3T_3$ we recover the previous [6] result as follows:

$$T = {}_1p_1T_1 + {}_1p_3T_3 = D_k^{-1}N_m^2 + N - \frac{N}{k} + \frac{2N_m}{k} + \frac{f(k)}{k}. \qquad (5)$$

From equation (5), under the condition that $N_m = N_1$, i.e. in the absence of fluctuations in the absorbing boundary we recover the limit value as $\lim_{k\to\infty} T = N$.

Case 5. Finally when $N_m = N$ and $N_1 = 0$, i.e. the absorbing boundary fluctuates in the entire interval $[0, N]$, in equation (5) we obtain the k-dependent limiting value as $T \approx N\left(\frac{1}{2} + \frac{1}{2k}\right)$ from which in the limit $k \to \infty$ we recover the result of case 2.

Here we should note that the splitting probabilities p_1 and p_3 are associated with the absorbing interval $[N_1, N_m]$ and not associated with the absorbing boundary which fluctuates inside the interval $[N_1, N_m]$. When $[N_m - N_1] = 0$, the splitting probabilities associated with
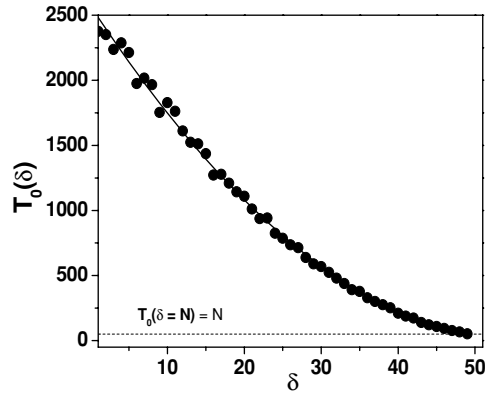
Figure 1. Random walk simulations with random boundary conditions. Here the size of the linear lattice is $N = 50$, the random walk step size k is $k = 1$, the lattice position $x = 0$ is the reflecting boundary and the lattice position y is the absorbing boundary associated with the random walker which can be anywhere in the interval $[50 - \delta, 50]$ with equal probabilities, i.e. $1/\delta$. Here δ is varied from 0 to 50 and the MFPT at each δ value was calculated over 10^5 realizations and the dotted lines are the prediction by the limiting case 3 of equation (4).

the protein molecule to enter into the absorbing boundary at $x = N_m = N_l$ from left-to-right p_L as well as from right-to-left p_R are simply given by $p_L = {}_0p_1$ and $p_R = {}_0p_3$, respectively. However, when $[N_m - N_l] > 0$ and the jump size $k = 1$, then the associated left-to-right p_L as well as right-to-left p_R splitting probabilities which are in turn functions of the fluctuating absorbing interval $\delta = [N_m - N_l]$ that can be shown as $p_L = \frac{1}{2} + \frac{3}{\delta+6}$ and $p_R = \frac{1}{2} - \frac{3}{\delta+6}$. The derivation is as follows:

Suppose when $k = 1$, $x_0 = 0$ and $\delta = 0$ then the splitting probability associated with the escape of the protein molecule through the absorbing boundary from the left-to-right (p_L), i.e. from the interval $[0, N_m - 1]$ into the point $x = N_m$ is $p_L = 1$ and obviously the splitting probability associated with the escape of the protein molecule from the right-to-left (p_R), i.e. from the interval $[N_m, N]$ is $p_L = 0$. However, under the condition that $\delta > 0$ and when the absorbing boundary fluctuates in the interval such that $[N_l, N_m]$ with equal probabilities then it is obvious to note that $p_R < 1$ and $p_L > 0$ even when the jump size is $k = 1$. This is mainly due to the fact that the randomly jumping absorbing boundary overtakes significant number of incoming trajectories of the protein molecule inside the subinterval $[N_l, N_m]$ which are coming from left-to-right. However, here one should note that the splitting probabilities p_R and p_L are in turn functions of the fluctuation interval δ of the absorbing boundary. Suppose let us assume that the current position of the protein molecule on the DNA lattice is $x = N_l$. In the next step the probability associated with the protein molecule to return back to the interval $[0, N_l - 1]$ is $1/2$. Now if the protein molecule moves from $x = N_l$ to the position $x = N_l + 1$ then at the position $x = N_l + 1$ the probability to get absorbed is $[\delta/3 + 2]^{-1}$. This is due to the fact that when the protein molecule moves from the position N_l to $N_l \to (N_l + 1)$, the absorbing boundary can be either at $x = N_l + 1$ or it already moved from $[N_l + 1, N_m]$ to N_l or moved from the position N_l into the interval $[N_l + 1, N_m]$ without meeting the protein molecule at all. Due to this reason, the effective number of absorbing points inside the interval $[N_l, N_m]$ is only $\delta/3$. However, the absorbing boundary cannot overtake the protein molecule at the lattice positions N_l and N_m both from the left-to-right and from the right-to-left directions. In other words the total number of effective absorbing points inside the interval $[N_l, N_m]$ is $\frac{\delta}{3} + 2$. Therefore when the initial position of the protein molecule on the DNA lattice is such
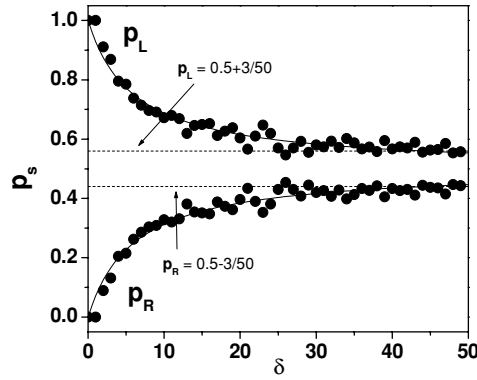
Figure 2. Splitting probabilities p_s associated with the entry of the random walker from the left-to-right p_L and from the right-to-left p_R with respect to the position of the fluctuating absorbing boundary as a function of the fluctuating interval (δ) associated with the absorbing boundary. Here the size of the lattice is $N_m = 50$, the absorbing boundary can be found anywhere in the interval $[50 - \delta, 50]$ with equal probabilities, i.e. $1/\delta$, the random walker step size is $k = 1$ and δ is varied from 0 to 50 and the splitting probabilities at each δ value were calculated over 10^5 realizations. Here the solid lines are the predictions $p_L = \frac{1}{2} + \frac{3}{\delta+6}$ and $p_R = \frac{1}{2} - \frac{3}{\delta+6}$.

that $x_0 = 0$, then the splitting probability associated with the entry of the protein molecule from left-to-right of the absorbing boundary (not the interval $[N_1, N_m]$) is $p_L = \frac{1}{2} + \frac{3}{\delta+6}$ and obviously the splitting probability associated with the entry of the protein molecule from the right-to-left of the absorbing boundary is $p_R = \frac{1}{2} - \frac{3}{\delta+6}$. This situation will be reversed when $x_0 = N$.

When $\delta \gg 6$ and also $\delta \approx N$ then we should note that $p_L \approx \frac{1}{2} + \frac{3}{N}$ and $p_R \approx \frac{1}{2} - \frac{3}{N}$ rather than $p_R = p_L = 1/2$ as in the case of $\delta = 0$. The reason is as follows. We should recall the fact that when the initial position of the protein molecule on the DNA lattice is $x_0 = 0$ and when $\delta \approx N$, the probability of the protein molecule to find the absorbing boundary at each step is N^{-1}. Due to this reason the gain in p_L in the first step from the lattice point $x_0 = 0$ is N^{-1}. On the other hand before the protein molecule moves from the lattice position $x_0 = 0$, the absorbing boundary can hit the lattice point $x_0 = 0$ with a probability of N^{-1} since $x_0 = 0$ is also a reflecting boundary for the fluctuating absorber. Similarly when the protein molecule approaches the boundary $x = N$ the probability to get absorbed at $x = N$ from left-to-right is N^{-1}. Therefore the total gain in the probability p_L is $3N^{-1}$, i.e. $p_L \approx \frac{1}{2} + \frac{3}{N}$ and obviously $p_R \approx \frac{1}{2} - \frac{3}{N}$. Nevertheless, the situation will be reversed when the initial position of the protein molecule on the DNA lattice is such that $x_0 = N$. Simple random walk simulations in fact prove the validity of these expressions for the splitting probabilities (figure 2).

In summary, in this letter we have exactly solved the MFPT problem for a one-dimensional random walk with random step size where there are two fixed reflecting boundaries and a fluctuating absorbing boundary in between. In the context of site-specific DNA–protein interactions, we have previously shown that by a random jump motion of the protein molecule on the DNA lattice the maximum achievable site-specific association rate is $r_m = N^{-1}$ where N is the size of the DNA lattice under consideration. However, the present results suggest that when there are fluctuations in the relative position of the specific site on the DNA lattice with respect to the protein molecule, then this maximum achievable site-specific association rate can be doubled, i.e. $r_m \Rightarrow 2N^{-1}$.

Acknowledgments

References

[1] Bruinsma R F 2002 *Physica* A **313** 211
[2] Von Hippel P H 1994 *Science* **263** 769
[3] Murugan R 2004 *Phys. Rev.* E **69** 011911
[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 2006 *J. Phys. A: Math. Gen.* **39** 1575–80
[7] Gardiner C W 1985 *Handbook of Stochastic Methods* (Berlin: Springer) 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 I (New York: Wiley)
[10] Redner S 2001 *A Guide to First Passage Processes* (London: Cambridge University Press)
[11] Risken H 1996 *Fokker Planck Equations* (Berlin: Springer)