# DNA-protein interactions under random jump conditions

R. Murugan*

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

(Received 6 June 2003; revised manuscript received 17 July 2003; published 29 January 2004)

We model the site-specific association of a protein molecule with DNA as a random walk with random jumps. Results show that the simultaneous occurrence of processes such as sliding, hopping, and intersegmental transfer can facilitate the diffusion-controlled site-specific association rate. We have also shown that sliding would dominate at lower DNA length, whereas at higher lengths hopping and intersegmental transfer would dominate. Apart from this, we predict that the association rate is directly proportional to the size of nonspecific DNA that flanks the specific site. These results are consistent with the experimental observations.

DOI: 10.1103/PhysRevE.69.011911 PACS number(s): 87.15.Kg

## I. INTRODUCTION

Recognition of a specific sequence of DNA among a vast excess of structurally similar nonspecific sequences by a protein, by another DNA, or by a RNA is an important phenomenon in molecular biology, especially in the processes such as initiation of replication, i.e., recognition of an origin of replication by DNA polymerase and initiation of transcription, i.e., recognition of a promoter sequence by RNA polymerase (Refs. [1–13]). Earlier theoretical models suggested that a protein could find its target site on DNA in solution condition simply by a one step diffusion process as given in scheme I.

$$E+D \underset{k_2}{\overset{k_1}{\leftrightarrow}} ED_a \quad \text{(scheme I)}.$$

Here, $E$ denotes the protein, $D$ denotes its target site on DNA, and $k_1$ ($\mathrm{mol}^{-1}\,\mathrm{s}^{-1}$) and $k_2$ ($\mathrm{s}^{-1}$) are the respective rate constants. According to scheme I, the DNA-protein interaction is a kind of three dimensional diffusion controlled reaction under electrostatic potential, where DNA is negatively charged due to the presence of phosphate groups and the protein is positively charged due to the presence of basic amino acids such as lysine and arginine (Refs. [5]). But the diffusion controlled association rate has a theoretical upper limit in the order of $\sim 10^9\ \mathrm{M}^{-1}\,\mathrm{s}^{-1}$, which is $\sim$10–20 times lower than the observed rate (Refs. [14], [15]), that is, in the order of $\sim 10^{10}\ \mathrm{M}^{-1}\,\mathrm{s}^{-1}$. Moreover, experiments on *Escherichia coli* lac repressor-operator system (Ref. [15]) showed an increasing magnitude of dissociation rate constant with decreasing length of operator containing DNA and the association rate showed a turn over dependency on salt concentration. Later these paradoxes were resolved by assuming a two-step model with transient intermediate (Ref. [15]),

$$E+D \underset{k_2}{\overset{k_1}{\leftrightarrow}} ED_a \underset{k_4}{\overset{k_3}{\leftrightarrow}} ED_b \quad \text{(scheme II)}.$$

Here $ED_a$ denotes the first nonspecifically bound weak complex, $ED_b$ denotes the strongly bound complex, $K_1$ ($\mathrm{mol}^{-1}$ and $K_1 = k_1/k_2$) is the dissociation constant for the first step, and $k_3$ ($\mathrm{sec}^{-1}$) and $k_4$ ($\mathrm{sec}^{-1}$) are the respective forward and reverse rate constants for the second step. According to this model, the protein molecule nonspecifically binds to DNA to form the weak complex $ED_a$, and then searches along the DNA for the specific site by the following facilitating mechanisms (Fig. 1), to form the strongly bound complex $ED_b$.

(i) *Association and dissociation*. The protein molecule searches for the specific site by continuously attaching and detaching from DNA, and therefore it is a random diffusional search. This is a kind of macroscopic process, where the

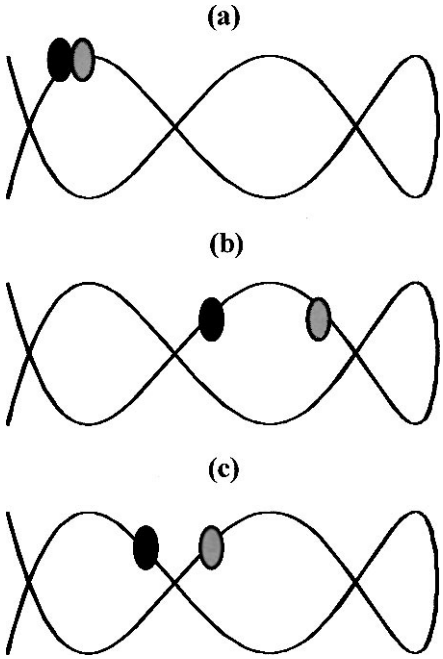


FIG. 1. Different searching modes of protein molecules for their recognition site on a supercoiled DNA, where dark-color ellipse represents the initial position of protein molecules and light-color ellipse represents the final position. (a) Sliding: step size is unit base pair. (b) Hopping: step size is few base pairs. (c) Intersegmental transfer: step size is few hundred to few thousand base pairs and this is possible only when two distal parts of the DNA come closer.

*Corresponding author. FAX: ++91-22-2280-4610/2280-4611. Email address: muruga@tifr.res.in

 

associated protein molecule dissociates completely, i.e., comes out of the electrostatic potential, and reassociates again either at the same site or at a different site with equal probabilities.

(ii) *Intersegmental transfer*. The protein molecule bound at one end jumps to another end of the same DNA lattice. Therefore, this is possible only in a supercoiled/condensed DNA, where two distal parts can come closer through a ring closure event. Since this process requires the segmental motion of DNA, the rate of transfer of protein from one segment to another is retarded by segmental diffusion. Here one should note that the protein molecule is exchanged between two distal segments of same DNA through the ring closure event without any macroscopic dissociation.

(iii) *Hopping*. Microscopic association and dissociation, where the dissociated protein molecule is still in the vicinity of DNA, i.e., in the electrostatic potential. Here the step size is few base pairs and therefore, this happens only in the linear DNA and thus the bending motion of DNA as in case of intersegmental transfer is not necessary.

(iv) *Sliding*. Transfer of the protein along the contour length of DNA, which is a one-dimensional random walk with a unit step size. This is different from hopping in a way that the protein is still under nonspecifically bound condition, i.e., it is not microscopically dissociated.

(v) *Correlated walk*. Asymmetric one-dimensional walk along the DNA lattice with unit step size, with a strong energy correlation towards the specific site (Ref. [14]). Here each movement of the protein along DNA is decided by the energy correlation between initial and final positions of protein. Positive correlations will favor the move towards the specific site, whereas negative correlations will resist the move and therefore it is a kind of energetically driven sliding.

The extent of hydrogen bonding and hydrophobic interactions between the specific sites of DNA and the protein determines the binding specificity (Refs. [16]). Here one should note that the free energy of specific binding is the sum of the free energy of nonspecific binding and the excess free energy due to specific hydrogen bonding and hydrophobic interactions. It was argued earlier (Refs. [14], [15]) that an energy correlation towards the specific site was necessary to explain the observed target-finding rate ($\sim 10^{10}$ $M^{-1}$ $s^{-1}$). This kind of energy driven searching mechanism was named as energy correlated walk (Ref. [14]). The nature and the existence of energy correlation is still under debate due to fact that so far no such free energy correlations which could drive the protein molecule towards the specific site were observed along the DNA sequences, though a sequence correlation was observed along the DNA (Refs. [17]). Other drawbacks in the earlier models are as follows. Here the processes such as association-dissociation, sliding, and hopping were treated as independent phenomena. But in the real situation it is very difficult to partition them because a sliding protein-molecule may suddenly hop or may undergo intersegmental transfer especially when it meets another protein molecule in its search path, i.e., switching from sliding to hopping or hopping to intersegmental transfer is itself a stochastic quantity. Therefore, a generalized random walk with random jump model is necessary to describe such kind of stochastic motions. In this paper we will show that the correlated motion assumption is not necessary and the random walk with random jumps itself is enough to explain the rate enhancement of specific binding of protein to DNA in solution condition. The organization of this paper is as follows. First we will describe the stochastic motion of protein molecule along DNA as a random walk with a fixed step size, from which we will prove that independent occurrence of sliding, hopping, and intersegmental transfer cannot facilitate the association rate. Then we will generalize the stochastic motion of protein along DNA as a random walk with random jumps, from which we will show that the random jump condition itself is enough to facilitate the association rate. Finally, we will discuss some of the consequences of this model in relation to the evolution of supramolecular structure of DNA.

## II. SITE-SPECIFIC ASSOCIATION OF PROTEIN WITH DNA

In the following section, we will derive the expression for the positional mean and positional variance of protein molecule when it undergoes a stochastic motion along the DNA lattice with a fixed step size.

### A. Random walk with a fixed step size

Let us assume that a protein molecule is undergoing a one-dimensional random walk with step size of $m$ base pairs along the DNA of $N$ base pairs in length and the specific site lies at $p$th base pair, where the inequality $0<x<p<N$ holds. Let us denote its present coordinate position as $x$ and the position at time $t=0$ as $x_0$. The corresponding transition probabilities in an infinitesimal time $\Delta t$ can be written as follows:

$$P(x\rightarrow x+m)=w_m\Delta t,$$

$$P(x\rightarrow x-m)=w_{-m}\Delta t,$$

$$P(x\rightarrow x)=[1-(w_m+w_{-m})\Delta t]. \tag{1}$$

Here $w_{\mathrm{m}}$ denotes the transition rate of protein molecule towards the specific site and $w_{-m}$ is its transition rate away from the specific site and $P$'s are the corresponding transition probabilities. The birth-death master equation describing the probability of finding the protein molecule at an arbitrary position $x$ in time $t$ thus becomes

$$\partial_t P(x,t)=\lim_{\Delta t\rightarrow 0}\frac{P(x,t+\Delta t)-P(x,t)}{\Delta t}=w_m P(x-m,t)+w_{-m}P(x+m,t)-(w_m+w_{-m})P(x,t). \tag{2}$$

Equation (2) takes simpler form upon introducing the generating function as $G(s,t)=\Sigma_{x=0}^{x_0}s^x P(x,t)$. Now Eq. (2) transforms to a simpler form as

$$\partial_s G(s,t) = G(s,t)\left\{\frac{w_m s^{2m} + w_{-m} - (w_m + w_{-m})s^m}{s^m}\right\}. \tag{3}$$

The solution of Eq. (3) for the initial condition $G(s,0) = s^{x0}$ [because $P(x_0,0)=1$] can be given as

$$G(s,t) = s^{x_0}\exp\left\{t\,\frac{w_m s^{2m} + w_{-m} - (w_m + w_{-m})s^m}{s^m}\right\}. \tag{4}$$

To obtain the probability distribution function $P(x,t)$, one has to expand Eq. (4) in series of powers of $s$ and then find the limit as $s \to 1$. Since Eq. (4) has an essential singularity at $s=0$, a Taylor series expansion is not possible. However, it can be expanded in a Laurent series and therefore $P(x,t)$ can be given as

$$P(x,t) = \lim_{s\to 1}\left\{\frac{1}{2\pi i}\sum_{n=0}^{x} s^n \int_{C_1}\frac{G(s',t)}{s'^{n+1}}ds + \frac{1}{2\pi i}\sum_{n=-x}^{-1} s^n \int_{C_2}\frac{G(s',t)}{s'^{n+1}}ds\right\}. \tag{5}$$

Here the contours $C_1$ and $C_2$ form a concentric shell such that $C_1$ encloses the singular point $s=0$ and $C_2$ is such that $0<s<R$, where $0<R\leqslant 1$. We are interested in the positional mean and positional variance of protein molecule, which can be obtained as follows:

$$\langle x\rangle = \lim_{s\to 1}\partial_s G(s,t) = x_0 + m(w_m - w_{-m})t, \tag{6}$$

$$\mathrm{Var}\{x\} = \lim_{s\to 1}\partial_s^2 G(s,t) - [\lim_{s\to 1}\partial_s G(s,t)]^2 + \lim_{s\to 1}\partial_s G(s,t) = m^2(w_m + w_{-m})t. \tag{7}$$

Equation (6) clearly shows that the protein molecule will move either towards the specific site ($p$) or away from it (i.e., $\langle x\rangle = x_0 + m|w_m - w_{-m}|t$ if $w_m > w_{-m}$ and $\langle x\rangle = x_0 - m|w_m - w_{-m}|t$ if $w_m < w_{-m}$) depending on the magnitude of transition rates ($w_m$ and $w_{-m}$), which is also directly proportional to the step size $m$. Moreover, Eq. (7) indicates dispersion in the probability distribution $P(x,t)$ with time $t$, which is directly proportional to square of the step size $m$. Now, using the mean first passage time (abbreviated as MFPT) formalism, we will calculate the site-specific association rate of a protein with DNA in the following section.

### B. Site-specific association rate under a fixed step-size condition

When the protein molecule is confined to a domain such that $0\leqslant x\leqslant p$ with $[\partial_t P(x,t)]_{x=0}=0$, i.e., reflecting boundary at $x=0$ and $[P(x,t)]_{x=p}=0$, i.e., an absorbing boundary at $x=p$, the mean first passage time required for the protein molecule to escape from the domain $[0,p]$, i.e., to find the target site, can be easily calculated from the Fokker-Planck equation (FPE) analog of Eq. (2) as follows. The FPE corresponding to Eq. (2) can be written as (Ref. [18])

$$\partial_t P(x,t) = -\partial_x[\alpha_1 P(x,t)] + \tfrac{1}{2}\partial_x^2[\alpha_2 P(x,t)], \tag{8}$$

where $\alpha_1 = m(w_m - w_{-m})$ and $\alpha_2 = m^2(w_m + w_{-m})$ are the corresponding drift vector and diffusion matrix. Now defining $\beta(z) = \exp\{\int_0^p (2\alpha_1/\alpha_2)dz\}$, the MFPT can be shown to be

$$T(x) = 2\int_x^\rho \beta(y)^{-1}dy \int_0^y \beta(z)\alpha_2^{-1}dz. \tag{9}$$

Evaluating the integrals in Eq. (9) we obtain

$$T(x) = \frac{\alpha_2}{4\alpha_1^2}\left\{\exp\left(\frac{4\alpha_1 p}{\alpha_2}\right) - \exp\left(\frac{2\alpha_1 p}{\alpha_2}\right) - \exp\left(\frac{4\alpha_1 x}{\alpha_2}\right) + \exp\left(\frac{2\alpha_1 x}{\alpha_2}\right)\right\}. \tag{10}$$

Equation (2) is a general case where the forward and the reverse transition rates are arbitrary. We can define the microscopic equilibrium constant as $K_{eq} = w_m/w_{-m} = e^{-\Delta G/RT}$ where $\Delta G$ (kcal/mol base pairs) is the correlation free energy along the DNA lattice. If there is no such correlation energy, then it is obvious that $K_{\mathrm{eq}}=1$. When $w_m = w_{-m}$, then $T(x) = (p^2 - x^2)/2m^2 w_m$ and thus the average rate of escape, which is simply the rate of association, becomes $r_{Kave} = p(\int_0^p T(x)dx)^{-1} = 3m^2 w_m/p^2$. Since the maximum possible diffusion controlled rate is $w_{\mathrm{m}} \sim 10^9\ \mathrm{M}^{-1}\,\mathrm{s}^{-1}$, to get a tenfold increase, a step size of $m = (10p^2/3)^{0.5} > p$ is necessary, which is clearly impossible. This is because, according to two-state model, the protein first binds nonspecifically with DNA and then searches for its specific site under nonspecifically bound condition. Therefore requirement of a step size, which is more than the length of DNA, contradicts the two-state models. Here, step size $m=1$ denotes sliding, moderate $m$ values denote hopping, and higher $m$ values denote intersegmental transfer. In the following section we will show that the simultaneous occurrence of sliding, hopping, and intersegmental transfer in the search process can itself enhance the site-specific association rate. First we will derive the expressions for the positional mean and positional variance of protein molecule on DNA under random jump condition.

### C. Random walk with random jump

Let us assume that a protein molecule is present at $x$th position of a DNA lattice of $N$ base pairs in length and $x_{01}$th position at time $t=0$, which is now undergoing a random jump with a maximum jump size of $k$, and the specific site is located at $p$th base pair, where the inequality $0<x_{01}<p<N$ holds. A random jump with size $k$ base pairs includes the possibility of all the jumps within $k$, and therefore includes the simultaneous occurrence of sliding, hopping, and intersegmental transfers. Now, the probability of finding the protein at the $x$th base pair such that $0<x<p<N$, at time $t$ can be obtained from the following generalized birth-death master equation:

$$\partial_t P(x,t)=\sum_{i=1}^{k}[w_i P(x-i,t)+w_{-i}P(x+i,t)-(w_i+w_{-i})P(x,t)]. \tag{11}$$

Here, $w_i$ is the transition rate of protein towards the specific site $p$, and $w_{-i}$ is the transition rate away from $p$ with a jump size of $i$ base pairs where $i$ is such that $1\leqslant i\leqslant k$, $k$ is the maximum jump size and $P$'s are corresponding transition probabilities. We are summing over all values of $i<k$ due to the fact that a jump size of $k$ includes all the possibilities within $k$. Now, defining the generating function $G(s,t)=\Sigma_{x=0}^{x_0}s^x P(x,t)$ as in the earlier section, Eq. (11) simplifies to

$$\partial_t G(s,t)=G(s,t)\sum_{i=1}^{k}\left\{\frac{w_i s^{2i}+w_{-i}-(w_i+w_{-i})s^i}{s^i}\right\}. \tag{12}$$

Using the initial condition as $P(x_{01},0)=1$, i.e., $G(s,0)=s^{x_{01}}$, the solution of Eq. (11) can be written as

$$G(s,t)=s^{x_{01}}\exp\left(t\sum_{i=1}^{k}\left\{\frac{w_i s^{2i}+w_{-i}-(w_i+w_{-i})s^i}{s^i}\right\}\right). \tag{13}$$

As in the preceding section, the probability distribution function can be expressed in terms of a Laurent series as

$$P(x,t)=\lim_{s\to 1}\left\{\frac{1}{2\pi i}\sum_{n=0}^{x}s^n\int_{C_1}\frac{G(s',t)}{s'^{n+1}}ds+\frac{1}{2\pi i}\sum_{n=-x}^{-1}s^n\int_{C_2}\frac{G(s',t)}{s'^{n+1}}ds\right\}, \tag{14}$$

where the contours $C_1$ and $C_2$ form a concentric shell such that $C_1$ encloses the singular point $s=0$ and $C_2$ is such that $0<s<R$, where $0<R\leqslant 1$. The positional mean and positional variance of the protein molecule on DNA lattice at time $t$ can be shown to be

$$\langle x\rangle=\lim_{s\to 1}\partial_s G(s,t)=x_{01}+t\sum_{i=1}^{k}i(w_i-w_{-i}), \tag{15}$$

$$\begin{aligned}\mathrm{Var}\{x\}&=\lim_{s\to 1}\partial_s^2 G(s,t)+\langle x\rangle-\langle x\rangle^2\\&=t^2\left\{\sum_{i=1}^{k}i^2(w_i-w_{-i})^2+\sum_{i<j,j=1}^{k}i(w_i-w_{-i})\times j(w_j-w_{-j})\right\}+t\left\{2x_{01}\sum_{i=1}^{k}i(w_i-w_{-i})+2\sum_{i=1}^{k-1}i(i-1)(w_{-i}+w_{i+1})+k(k+1)w_{-k}\right\}.\end{aligned} \tag{16}$$

From Eq. (15), we can conclude that depending on the forward and reverse transition rates the protein molecule will be driven towards or away from the specific site, which is linear with time. Equation (16) clearly shows a time dependent dispersion of the probability distribution function. Now, in the following section, we will calculate the site-specific association rate under random jump condition.

### D. Site-specific association rate under random jump condition

The MFPT taken by protein molecule to reach $p$, starting from any $x<p$, by a random jump process can be obtained from Eqs. (8) and (9) as follows:

$$T(x,k)=\frac{\alpha_2}{4\alpha_1^2}\left\{\exp\left(\frac{4\alpha_1 p}{\alpha_2}\right)-\exp\left(\frac{2\alpha_1 p}{\alpha_2}\right)-\exp\left(\frac{4\alpha_1 x}{\alpha_2}\right)+\exp\left(\frac{2\alpha_1 x}{\alpha_2}\right)\right\}, \tag{17}$$

where $\alpha_1=\Sigma_{i=1}^{k}i(w_i-w_{-i})$ and $\alpha_2=\Sigma_{i=1}^{k}i^2(w_i+w_{-i})$ are the corresponding drift and diffusion terms. The problem simplifies when one of the following conditions holds.

*Case I.* $w_i=w_{-i}$, $w_i\neq w_j$. These conditions can be understood as follows. Since we are interested in an unbiased search, i.e., no energy correlation towards the target site ($K_{\mathrm{eq}}=w_i/w_{-i}=1$), we can insist upon the condition $w_i=w_{-i}$. The second condition $w_i\neq w_j$ is true in case of a linear DNA due to the fact that without a macroscopic dissociation, hopping, and intersegmental transfers are not supported. These conditions hold well for the interaction of prokaryotic RNA polymerase with its DNA during transcription initiation where DNA is almost linear due to simultaneous occurrence of transcription and the translation (Ref. [13]). Under these conditions Eq. (12) simplifies to

$$\partial_t G(s,t)=G(s,t)\sum_{i=1}^{k}\left\{w_i\frac{s^{2i}+1-2s^i}{s^i}\right\}. \tag{18}$$

The solution of Eq. (18) for same initial condition as in the previous case becomes

$$G(s,t)=s^{x_{01}}\exp\left(\sum_{i=1}^{k}\left\{w_i\frac{s^{2i}+1-2s^i}{s^i}\right\}\right). \tag{19}$$

Now the mean and variance of $x$ can be shown to be

$$\langle x\rangle=\lim_{s\to 1}\partial_s G(s,t)=x_{01}, \tag{20}$$

$$\mathrm{Var}\{x\}=\lim_{s\to 1}\partial_s^2 G(s,t)+\langle x\rangle-\langle x\rangle^2=2t\sum_{i=1}^{k}w_i i^2. \tag{21}$$

From Eqs. (20) and (21) we can conclude that the positional mean does not evolve with time whereas the positional variance evolves linearly with time. Using Eqs. (8) and (9), the MFPT for a jump size $k$ can be given as

$$T(x,k)=\frac{1}{\alpha_2}(p^2-x^2), \tag{22}$$

where $\alpha_2=2\Sigma_{i=1}^{k} i^2 w_i$.

*Case II.* $w_i=w_{-i}$, $w_i=w_j=w$. The first condition is insisted to account for an unbiased jump. The condition $w_i=w_j$ is true when the protein molecule interacts with a closely packed/supercoiled DNA, where two distal segments of DNA come closer by a ring closure event, so that the protein molecule can either jump from one segment to another segment or slide/hop within the same segment with equal probabilities. In other words, the protein molecule can undergo sliding, hopping, or intersegmental transfer with equal probabilities, which is possible only in case of a closely packed structure. Moreover, recent studies showed that proteins themselves induced the bending motion of DNA (Bruinsma in Ref. [5]). These conditions are true in case of interaction of eukaryotic (e.g., plants and animals) RNA polymerase with its DNA in the process of transcription initiation where the DNA is closely packed/supercoiled. Under these conditions, Eq. (12) becomes

$$\partial_t G(s,t)=wG(s,t)\sum_{i=1}^{k}\left\{\frac{s^{2i}+1-2s^i}{s^i}\right\}$$
$$=wG(s,t)\left\{\frac{(1+2k)(1-s)-s^{-k}-s^{k+1}}{s-1}\right\}. \tag{23}$$

The solution of Eq. (23) with same initial condition as in case I can be given as

$$G(s,t)=s^{x_{01}}\exp\left(wt\sum_{i=1}^{k}\left\{\frac{s^{2i}+1-2s^i}{s^i}\right\}\right). \tag{24}$$

Now the mean and variance of $x$ becomes

$$\langle x\rangle=\lim_{s\to 1}\partial_s G(s,t)=x_{01}, \tag{25}$$

$$\mathrm{Var}\{x\}=\lim_{s\to 1}\partial_s^2 G(s,t)+\langle x\rangle-\langle x\rangle^2=2wt\sum_{i=1}^{k} i^2. \tag{26}$$

As in case I, Eqs. (25) and (26) clearly indicate the invariant nature of positional mean and linear evolution of positional variance with time. Now the MFPT for a jump size of $k$ can be given as

$$T(x,k)=\frac{1}{\alpha_2}(p^2-x^2)=\frac{3(p^2-x^2)}{k(k+1)(2k+1)w}. \tag{27}$$

## III. RESULTS AND DISCUSSIONS

Generally, the genomic DNA present inside a living cell is longer in length compared to the dimensions of the cell itself (e.g., in the case of *E. coli*, the DNA length is a few centimeters whereas the cell's dimension is a few micrometers). Therefore, under *in vivo* conditions the genome is closely packed via forming higher structural elements such as supercoils. Due to this fact the condition $w_i=w_{-i}$, $w_i=w_j=w$ approximately holds *in vivo*, which is simply the random walk with random jump with equal probabilities (case II). The requirement of this fact for a living cell can be visualized as follows. A nonspecifically bound protein molecule primarily can slide (jump size=1) along the DNA, but needs to hop whenever it finds a small barrier such as another protein molecule and jump to a distal part whenever there is a ring closure event of DNA (otherwise the protein molecule may get trapped in a nonspecifically bound form which is lethal to the organism). But the random jump size ($k$) depends on the amount of higher structural elements present in DNA, which varies from organism to organism. Therefore, for an arbitrary jump size $k$, which includes all the possibilities of simultaneous occurrence of sliding, hopping, and intersegmental transfer, the average (over $x$) site-specific association rate of protein with DNA under *in vivo* condition is approximately given by [from Eq. (27)]

$$r_{Kave}=w\frac{k(k+1)(2k+1)}{2p^2}, \tag{28}$$

whereas the site unspecific association rate (here it is just equal to $w$) is the diffusion controlled rate. One also should note that

$$\lim_{p\to N} r_{Kave}=w\frac{k(k+1)(2k+1)}{2N^2}. \tag{29}$$

Therefore to get a tenfold enhancement of $w$, a jump size of $k\cong 2N^{0.67}<N$ is needed (using the relation $r_{Kave}=10w$), which is clearly acceptable and suggests that at higher $N$ values [Fig. 2(a)] the contribution of sliding is negligible [e.g., for $N=100$ base pairs, sliding (i.e., a unit step size) contributes only $(100/k)=50\times 100^{-0.67}\sim 2.5\%$], whereas other mechanisms such as hopping and intersegmental transfer are the dominating ones. Variation of jump size required to enhance the diffusion controlled association rate to tenfold with respect to the DNA length is shown in Fig. 2(b), which clearly indicates that as $N$ increases, the intersegmental transfer dominates, whereas at lower $N$ values sliding and hopping dominate. These results suggest that a simultaneous occurrence of sliding, hopping, and intersegmental transfer mechanisms is necessary to facilitate the diffusion-controlled association rate. This is also true under *in vivo* conditions due to the fact that inside the living cell a large number of proteins varying in size, shape, and affinity are interacting with a single DNA molecule whose specific sites are also different. Therefore searching process of a particular protein molecule for its specific site by pure sliding will be frequently hindered by other protein molecules present in its search path thus reducing the search efficiency, which is disastrous to the organism (Ref. [15]). Thus hopping and intersegmental transfer are essential to avoid the protein-protein collisions in the course searching. If $k\sim N^{\delta}$ where $2/3\leqslant\delta\leqslant 1$, one can easily show that $r_{Kave}\cong wN^{\eta}$ (where $0\leqslant\eta\leqslant 1$), i.e., $r_{Kave}\propto N$, which is the usual observation in DNA-protein binding studies, i.e., the association rate is directly proportional to length of nonspecific DNA, which is flanking the specific site (Ref. [15] and Engler in Ref. [4]). Moreover,

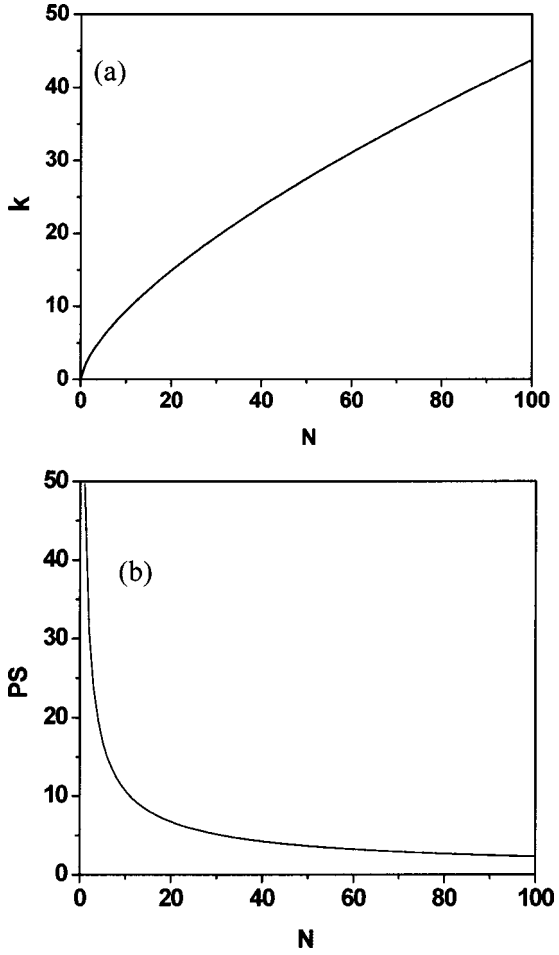


FIG. 2. (a) Variation of required jump size $k \sim 2N^{0.67}$ base pairs, i.e., to enhance the diffusion controlled association rate of DNA-protein binding to tenfold, with the size of DNA ($N$ base pairs). Here a random jump size of $k$ means after a unit jump from $x_0$th base pair of DNA, the protein can be found anywhere in the range of $x_0 \pm k$. (b) Variation of percentage occurrence ($\mathrm{PS} \sim 50N^{-0.67}$) of sliding with the size of DNA ($N$ in base pairs) which clearly indicates that at lower $N$ values sliding dominates whereas at higher $N$ values other processes such as hopping and intersegmental transfer are the dominating ones.

Fig. 3 clearly shows that when the size of DNA ($N$) and the jump size are independent, at sufficiently higher values of $N$ and $k$, $r_{Kave}$ is almost a constant quantity that is not true in the real situation and thus proves the validity of the relation

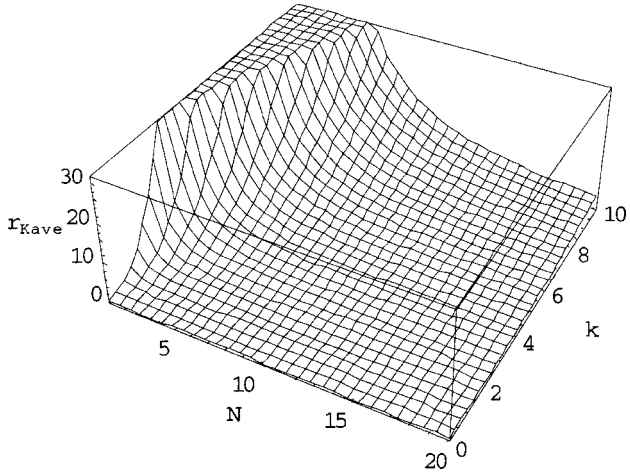


FIG. 3. Variation of site-specific association rate ($r_{Kave}$, $\mathrm{mol}^{-1}\,\mathrm{s}^{-1}$) with respect to changes in DNA length ($N$ in base pairs) and jump size ($k$, in base pairs) (when $N$ and $k$ are independent quantities), which clearly shows a constant nature of $r_{Kave}$ at higher $N$ and $k$.

$2N^{0.67} \leqslant k \leqslant N$. Though the size of genomic DNA varies from organism to organism, the rate of recognition of specific sites by the corresponding proteins is almost a constant quantity over different genomes, which can happen only when there is a kind of compensation between the genomic size ($N$) and the searching jump size $k$. In this context, our model predicts that as the genomic size ($N$) increases, the jump size $k$ also increases in order to keep the site-specific association rate constant. This compensation phenomenon suggests a positive correlation between the genomic size and its closely packed nature. Since hopping and intersegmental transfers are not much supported by a linear DNA, in due course of evolution DNA might have taken the present closely packed supercoiled and supramolecular structure. One more evidence we observe is the compartmentalization of eukaryotic DNA by a nuclear membrane. Therefore, from these arguments we can conclude that a random jump condition is itself enough to drive the target-specific association of protein with DNA and there is no need for the existence for correlation energy.

## ACKNOWLEDGMENTS

This work was supported by TIFR, Mumbai, India. The author thanks Professor S. Mazumdar for his encouragement and help.

[1] G. Adam and M. Delbruck, in *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 & Signals* (Cold Spring Harbor Laboratory New York, 2001).

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

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

[4] C. K. Reddy *et al.*, J. Mol. Biol. **314**, 619 (2001); L. E. Engler *et al.*, *ibid.* **307**, 619 (2001).

[5] J. Norberg, Arch. Biochem. Biophys. **410**, 48 (2003), and references therein; R. F. Bruinsma, Physica A **313**, 211 (2002).

[6] D. R. Lesser, M. R. Kurpiewski, and L. J. Jacobson, Science **250**, 776 (1990).

[7] P. H. Von Hippel, Science **263**, 769 (1994); Y. Jia, A. Kumar, and S. S. Patel, J. Biol. Chem. **29**, 30 451 (1996).

[8] P. H. Von Hippel, J. Biol. Chem. **15**, 675 (1989), and references therein.
[9] P. H. Von Hippel and O. G. Berg, Proc. Natl. Acad. Sci. U.S.A. **83**, 1608 (1986).
[10] J. F. Schildbach, A. W. Karzai, B. E. Raumann, and R. T. Sauer, Proc. Natl. Acad. Sci. U.S.A. **96**, 811 (1999).
[11] H. Viadiu and A. K. Aggarwal, Mol. Cells **5**, 889 (2000).
[12] P. Etchegoin and M. A. Nollmann, J. Theor. Biol. **220**, 233 (2003).
[13] B. Lewin, *Genes VII* (Oxford University Press, London, 2000), p. 243.
[14] O. G. Berg, R. B. Winter, and P. H. Von Hippel, Biochemistry **20**, 6929 (1981), and references therein.
[15] O. G. Berg, R. B. Winter, and P. H. Von Hippel, Biochemistry **20**, 6961 (1981).
[16] L. Jen-Jacobson, L. E. Engler, and L. A. Jacobson, Structure (London) **8**, 1015 (2000); G. D. Stromo and D. S. Fields, Trends Biochem.Sci. **23**, 109 (1998); L. Pardo, N. Pastor, and H. Weinstein, Biophys. J. **75**, 2411 (1998); W. Kauzmann, Adv. Protein Chem. **14**, 1 (1959); A. C. Cheng *et al.*, J. Mol. Biol. **327**, 781 (2003).
[17] F. Julicher and R. Bruinsma, Biophys. J. **74**, 1169 (1998); P. V. Benos, A. S. Lapedes, and G. D. Stormo, J. Mol. Biol. **323**, 701 (2002).
[18] C.W.Gardiner, in *Hand Book of Stochastic Methods*, edited by H. Haken (Springer-Verlag, Berlin, 1983), p. 260.