

Protein association with circular DNA: Rate enhancement by nonspecific binding

Ramzi Alsallaq and Huan-Xiang Zhou^{a)}

Department of Physics and Institute of Molecular Biophysics and School of Computational Science, Florida State University, Tallahassee, Florida 32306, USA

(Received 9 November 2007; accepted 5 February 2008; published online 21 March 2008)

An analytical solution for the nonspecific-binding-facilitated diffusion-controlled rate of association of a protein with a specific site on a circular DNA is derived. Nonspecific binding is modeled by a short-range attractive surface potential. The protein undergoes diffusion in the bulk solution and in the surface layer. The association rate for a circular DNA is compared to the counterpart for a linear DNA, in which the ends of the surface layer are treated as reflecting. As expected, when the DNA length is long, the shape of the DNA does not affect the association rate. For a shorter length, the association rate for the linear DNA is modestly higher than the circular counterpart. The higher rate of the linear DNA is possibly due to its more open shape, which affords it a higher ability to draw the protein from the bulk to its surface. The analytical solution is verified by Brownian dynamics simulations. © 2008 American Institute of Physics. [DOI: [10.1063/1.2888996](https://doi.org/10.1063/1.2888996)]

I. INTRODUCTION

Many biological functions involve the association of proteins with specific sites located within a very long DNA. A simple model involving only three-dimensional diffusion search fails to explain the rapid rates at which such association occurs. It is very unlikely that proteins find their specific targets directly without binding to nonspecific sites. Nonspecific binding of 434 Cro repressor to a model B-DNA has been studied by Brownian dynamics simulations.¹ Many authors^{2–6} have proposed that proteins find their specific sites on DNA through mechanisms of facilitated diffusion. In general, the proposed mechanisms have two components. The first is the three-dimensional diffusion, which can bring a protein to a random nonspecific site on the DNA. The second is the translocation from the initial random site to the final specific site. In several theoretical studies,^{3,4,6–10} the coupling between the one-dimensional motion along the DNA and the three-dimensional diffusion in the bulk solution is explicitly considered. Of these two are of particular relevance to the present study. One is our own work,^{4,6} in which nonspecific binding is explicitly accounted for by a short-range attractive potential and motion in the surface layer is modeled as anisotropic three-dimensional diffusion. This work forms the theoretical basis of the present study. The other is by Klenin *et al.*,¹⁰ who proposed a theoretical model for the reaction time and performed numeric simulations assuming that the protein moves via the two traditional modes: three-dimensional diffusion in the bulk solution and one-dimensional sliding along the DNA. After each sliding step, the protein is allowed to get off the DNA via a “kickoff” probability. This work is related to the simulation part of the present study.

In the present work, we derive an analytical solution for the rate constant of association of a protein with a specific

site on a circular DNA. The derivation is based on our previous theoretical model for linear DNA.^{4,6} We compare the results between a circular DNA and its linear counterpart and perform Brownian dynamics simulations to verify the analytical solution.

Why are we interested in modeling protein association with a circular DNA? First of all, circular DNA naturally occurs in cells and has been studied experimentally.^{11–13} Second, because circular DNA lacks the ends as compared to linear DNA, a comparison of results between circular and linear DNA can be used to isolate end effects and check various assumptions made about ends of linear DNA. For example, to circumvent the mathematical difficulty presented by a finite linear DNA modeled as a cylinder, the finite cylinder has been extended to infinite length with inert flanking tails that lack the capacity for nonspecific binding.^{6,7} Although this treatment appears reasonable, the effect of the extra excluded volume introduced is uncertain.

In our model,⁶ nonspecific binding is accounted for by a short-range attractive potential and, consequently, the borders of the surface layer at the ends of the linear DNA appear reflecting. Reflecting ends have also been assumed in other studies.^{7,9,13} Whether the ends of a DNA are merely reflecting or present additional kinetic pathways can be tested by comparing the rates of association between a linear DNA and its circular counterpart. For example, Surby and Reich¹³ in their experimental study found that the dissociation rate constant of EcoRI methyltransferase from a linear DNA is threefold higher than from the circular DNA with the same length, suggesting the possibility of dissociation via the ends of the linear DNA. In the same study, Surby and Reich also reported that the same dissociation rate constant of EcoRI endonuclease from DNA was found regardless of whether the DNA is circular or linear. From this and other experimental studies, it has been concluded that the ends of DNA may

^{a)}Electronic mail: zhou@sb.fsu.edu.

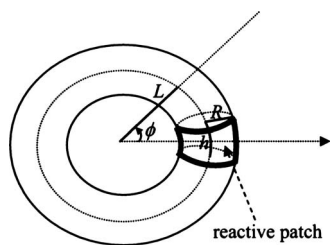


FIG. 1. Circular DNA modeled as a torus.

appear reflecting to some proteins such as EcoRI and EcoRV endonucleases^{13,14} but not merely reflecting to others such as EcoRI DNA methyltransferase,¹³ *E. coli* RNAP searching its promoter site on T7 DNA,¹⁵ and human Rad51 forming a ringlike conformation and sliding along a DNA.¹⁶ In drawing such conclusions, it is implicitly assumed that when the ends of a linear DNA are merely reflecting, one expects the association kinetics to be unchanged when the DNA is circularized.

The analytical solution for the association rate of a protein with a circular DNA, derived in this paper, allows for a direct comparison with the result for the corresponding linear DNA obtained previously.⁶ We find that the two association rates are very similar at long DNA lengths and show a modest difference at short lengths. This finding validates the implicit assumption invoked in determining whether additional kinetic pathways at the DNA ends are called for. Our analytical solution is verified by Brownian dynamics simulations.

The rest of the paper is organized as follows. In Sec. II, we present an outline of the derivation of the analytical solution for the association rate, with details relegated to two appendices. In Sec. III, the procedure for the Brownian dynamics simulations is described. Results for representative model parameters are presented in Sec. IV. Finally, a number of issues arising from the present study are discussed in Sec. V.

II. ANALYTICAL THEORY

In this section, we present our derivation for the rate of nonspecific-binding-facilitated association of a protein to a specific site on a circular DNA. The model is the same as that in our previous work,⁶ except for the shape of the DNA. The DNA is previously modeled as a cylinder; here, it is modeled as a torus on account of the circular shape (see Fig. 1). The cross section of the torus has a radius of R (as determined by the protein-DNA contact distance). The centers of all the cross sections make a circle with a radius L ; the circumference of this circle, $2\pi L$, will be taken as the length of the DNA. The specific site is modeled as a reactive patch, which spans a range of azimuthal angle from $-\phi_0$ to ϕ_0 (correspondingly, the full width of the patch is $h=2\phi_0 L$). The protein, modeled as a spherical particle, diffuses around the torus. Near the DNA surface, an attractive potential modeling nonspecific binding influences the diffusion. As before, we invoke the equilibration approximation between the surface layer and the bulk solution, which is valid in the limit of a very thin surface layer. The nonspecific binding constant

per unit surface area will be denoted as K_{ns} . Protein diffusion in the bulk solution is assumed to be isotropic, with diffusion constant D . In the surface layer, the diffusion is assumed to be anisotropic; the diffusion coefficient is D_{\parallel} in the longitudinal direction (i.e., along the length of the DNA), 0 in the transverse direction, and D in surface normal direction.

The association rate is determined by the steady-state Smoluchowski equation,

$$\nabla \cdot \mathbf{J}(\mathbf{r}) = 0, \quad (1a)$$

with the flux at position \mathbf{r} given by

$$\mathbf{J}(\mathbf{r}) = -\mathcal{D}e^{-\beta U(\mathbf{r})} \nabla e^{-\beta U(\mathbf{r})} P(\mathbf{r}), \quad (1b)$$

where $U(\mathbf{r})$ is the interaction potential giving rise to nonspecific binding, $\beta=(k_B T)^{-1}$, \mathcal{D} is the diffusion matrix, with all three principal components equal to D in the bulk solution but two of the principal components changing to D_{\parallel} and 0, respectively, in the surface layer, and $P(\mathbf{r})$ is the pair distribution function. This function approaches 1 at infinite protein-DNA separation. On the DNA surface, $P(\mathbf{r})$ satisfies the absorbing boundary condition over the reaction patch and the reflecting boundary condition elsewhere. The boundary condition is of a mixed type (involving the pair distribution function in one part of the DNA surface but its flux elsewhere on the surface). The diffusion-controlled association rate is given by

$$k_D = - \int_{\text{reactive patch}} ds \mathbf{n} \cdot \mathbf{J}(\mathbf{r}), \quad (2)$$

where ds is the element of surface area and \mathbf{n} is the unit normal vector.

A. Association rate in the absence of a surface potential

We first present the solution for the association rate in the absence of a surface potential. The solution is obtained by making the constant-flux approximation,¹⁷ which circumvents the difficulty presented by the original mixed boundary condition. In this approximation, the flux on the reactive patch is assumed to be a constant, and this constant is determined by requiring that the absorbing boundary condition is satisfied on average (instead of everywhere on the reactive patch).

In the absence of a potential (and assuming isotropic diffusion), the steady-state Smoluchowski equation becomes the Laplace equation, with boundary conditions on a torus. This problem can be solved by using toroidal coordinates (ξ, ζ, ϕ) , which are related to the Cartesian coordinates by¹⁸

$$x = \frac{c \sinh \xi \cos \phi}{\cosh \xi - \cos \zeta}, \quad (3a)$$

$$y = \frac{c \sinh \xi \sin \phi}{\cosh \xi - \cos \zeta}, \quad (3b)$$

$$z = \frac{c \sin \zeta}{\cosh \xi - \cos \zeta}, \quad (3c)$$

where $c = (L^2 - R^2)^{1/2}$. From here on $\cosh \xi$ will be denoted by χ . The surface of the torus modeling the DNA is defined by $\chi = L/R \equiv \chi_0 = \cosh \xi_0$.

Details for the solution of the association rate are presented in Appendix A. The final result is

$$\frac{k_{D0}}{4\pi Dc} = \frac{\pi L^2 R^3}{4c^5} \left(- \sum_{n=0, m=0}^{\infty} u_{nm} P_{n-1/2}^m(\chi_0) \times Q'_{n-1/2}(\chi_0) \frac{\sin m\phi_0}{m\phi_0} \right)^{-1}, \quad (4)$$

where a prime signifies derivative, $\sin m\phi_0/m\phi_0$ is 1 when $m=0$, and $P_{n-1/2}^m(x)$ and $Q_{n-1/2}(x)$ are (associated) Legendre functions of the first and second kinds, respectively (the present special cases with half-integer degrees and $x = \cosh \xi \geq 1$ are also known as toroidal functions). The determination of the coefficients u_{nm} is described in Appendix A. For the special case $\phi_0 = \pi$ (i.e., the whole torus surface is absorbing), the exact result for the diffusion-controlled rate can be obtained from the capacitance^{19,20} of a torus-shaped conductor.¹⁸ This is given by

$$\frac{k_{D0}(\phi_0 = \pi)}{4\pi Dc} = \frac{2}{\pi} \sum_{n=0}^{\infty} \varepsilon_n \frac{Q_{n-1/2}(\chi_0)}{P_{n-1/2}(\chi_0)}, \quad (5)$$

where $\varepsilon_n = 1$ when $n=0$ and 2 when $n > 0$. Comparison of the result of Eq. (4) at $\phi_0 = \pi$ with Eq. (5) shows that the constant-flux approximation leads to a slight underestimation for the association rate (also see Sec. IV B).

B. Association rate in the presence of a surface potential

We model nonspecific binding by an attractive potential confined to a thin surface layer.⁶ Outside the surface layer, the potential is absent and diffusion is assumed to be isotropic; hence, there the pair distribution function still satisfies the Laplace equation. Assuming that the potential only depends on the ξ coordinate, the steady-state Smoluchowski equation for the pair distribution function $P_s(\mathbf{r})$ in the surface layer takes the form

$$D \frac{\partial}{\partial \xi} \frac{\sinh \xi e^{-\beta U(\xi)}}{\chi - \cos \zeta} \frac{\partial e^{-\beta U(\xi)} P_s(\mathbf{r})}{\partial \xi} + D_{\parallel} \frac{1}{\sinh \xi (\chi - \cos \zeta)} \frac{\partial^2 P_s(\mathbf{r})}{\partial \phi^2} = 0 \quad (6)$$

in toroidal coordinates. It should be recalled that the diffusion coefficient in the transverse direction (i.e., along ζ) is assumed to be zero. Let the outer boundary of the surface layer be specified by $\chi = \chi_1 \equiv \cosh \xi_1$ (note that $\chi_1 < \chi_0$). At the outer boundary, the pair distribution function satisfies the continuity condition

$$[e^{\beta U(\xi)} P_s(\mathbf{r})]_{\xi=\xi_1} = [P(\mathbf{r})]_{\xi=\xi_1}. \quad (7a)$$

In the limit that the width of the surface layer is extremely small, the above equation can be assumed to be valid when the left-hand side is evaluated at any point within the surface layer. That is,

$$e^{\beta U(\xi)} P_s(\mathbf{r}) \approx [P(\mathbf{r})]_{\xi=\xi_1}, \quad \xi_0 > \xi > \xi_1, \quad (7b)$$

which is equivalent to

$$P_s(\mathbf{r}) \approx e^{-\beta U(\xi)} [P(\mathbf{r})]_{\xi=\xi_1}, \quad \xi_0 > \xi > \xi_1. \quad (7c)$$

Integrating both sides of Eq. (7c) over the width of the surface layer, we find

$$\int_{\xi_0 > \xi > \xi_1} dl P_s(\mathbf{r}) \approx \int_{\xi_0 > \xi > \xi_1} dl e^{-\beta U(\xi)} [P(\mathbf{r})]_{\xi=\xi_1} \equiv K_{ns} [P(\mathbf{r})]_{\xi=\xi_1}, \quad (8)$$

where $dl = [c/(\chi - \cos \zeta)] d\xi$ is the element of length along the normal of the torus surface.

The last result allows the effect of the potential to be accounted for as a boundary condition (see Appendix B). By applying the constant-flux approximation and following steps very similar to those for the case without a potential, we arrive at the following result for the diffusion-controlled association rate:

$$\frac{k_D}{4\pi Dc} = \frac{\pi L^2 R^3}{4c^5} \left(- \sum_{n=0, m=0}^{\infty} v_{nm} P_{n-1/2}^m(\chi_0) \times Q'_{n-1/2}(\chi_0) \frac{\sin m\phi_0}{m\phi_0} \right)^{-1}. \quad (9)$$

The determination of the coefficients v_{nm} is described in Appendix B. They depend on the dimensionless parameter

$$\lambda = \frac{D_{\parallel} K_{ns}}{DR}, \quad (10)$$

which accounts for the surface potential. When $\lambda = 0$, Eq. (9) reduces to Eq. (4). Notice that D_{\parallel} and K_{ns} appear together in their influence on the association rate. When the whole DNA surface is absorbing, the exact solution (i.e., without the constant-flux approximation) for the association rate is again given by Eq. (5).

III. BROWNIAN DYNAMICS SIMULATIONS

The main purpose of our Brownian dynamics simulations is to verify the analytical solutions for the association rate given by Eqs. (4) and (9). We now describe the procedure of our Brownian dynamics simulations.

A. Algorithm for calculating association rate

In Brownian dynamics simulations, one launches a large number of trajectories. Along each trajectory, association is allowed to occur with an appropriate probability. The association rate is then calculated from the statistics of the trajectories. Two different algorithms have been developed for this calculation. In the approach introduced by Northrup *et*

al.,²¹ trajectories of the ligand (which in our case is the protein) are started from a spherical surface, referred to as the b-surface, enclosing the receptor (which in our case is the DNA). The trajectories are continued until the ligand is absorbed by the receptor or by a distant spherical surface, referred to as the q-surface. If the fraction of trajectories terminated due to absorption by the receptor is γ , then the diffusion-controlled association rate is given by

$$k_D = \frac{k_b \gamma}{1 - (1 - \gamma)k_b/k_q}, \quad (11)$$

where k_b (k_q) is the diffusion-controlled association rate when the whole b-surface (q-surface) becomes the specific site for association. k_b and k_q are given by the well-known Smoluchowski result,

$$k_b = 4\pi D b, \quad (12a)$$

$$k_q = 4\pi D q, \quad (12b)$$

where b and q are the radii of the b-surface and q-surface, respectively. This algorithm is rigorous when the b-surface is large enough such that the pair distribution function outside it is isotropic.²²

In an alternative algorithm developed by us,²³ the reactive patch on the receptor is extended outward slightly in the normal direction into a reactive region. Trajectories are launched from this region. When a trajectory is found in this region, association occurs with a finite rate κ . If the fraction of trajectories terminated due to reaction (as opposed to escaping to infinite separation) is S , then the diffusion-controlled association rate is obtained as

$$k_D = \kappa V_{RR} \langle e^{-\beta U} \rangle_{RR} S / (1 - S), \quad (13)$$

where V_{RR} is the volume of the reaction region and $\langle \cdots \rangle_{RR}$ denotes the average over the reaction region. We implemented this algorithm for the problem of protein association to a specific site on a circular DNA without a surface potential, and the simulation results for k_{D0} are in good agreement with Eq. (4). We note that this algorithm actually gives the time-dependent rate coefficient of diffusion-controlled association, which approaches the “steady-state” result given in Eq. (13) at long times. However, the length of the transient period before reaching the steady-state value is much shorter than the time resolution of typical experiments;²⁴ hence, it is the steady-state result for the association rate that is relevant for comparing against experiments.

The presence of a surface potential posed a special challenge for faithfully propagating the Brownian trajectories. Following Merlitz *et al.*,²⁵ we adopted the following procedure for dealing with the surface potential. A surface layer, obtained by extending the DNA surface by ε along the normal direction, was explicitly introduced.²⁶ New positions were proposed according to free diffusion. Every move from the bulk solution into the surface layer was always accepted. On the other hand, a move from the surface layer to the bulk solution was accepted with a probability p . If not accepted, the old position was reused. The value of p is obviously dictated by the dimensionless parameter λ in Eq. (10), and the two are inversely related. We tested the following ansatz:

$$\lambda = \alpha(p^{-1} - 1). \quad (14)$$

This relation gives the correct result $\lambda=0$ when $p=1$ and predicts the expected inverse dependence of λ on p as $p \rightarrow 0$. The constant α was determined by minimizing the difference between simulation results for k_D and the corresponding results calculated from the analytical solution of Eq. (9).

With the surface potential accounted for in the above away, we modified the algorithm of Northrup *et al.*²¹ to calculate k_D . The need for modification arises from the particular shape of the receptor, a torus, in our problem. The b-surface, which is spherical in the algorithm of Northrup *et al.*, has to at least enclose the torus. With such a starting surface γ , the fraction of trajectories terminated due to reaction is very small, especially when the reactive patch is narrow. A small γ makes it difficult to calculate k_D accurately. Therefore, we chose as the starting surface a torus which encloses the circular DNA. Specifically, the b-surface was taken as the outer boundary of the surface layer described in the preceding paragraph. The cross section of this surface is concentric with the cross section of the enclosed DNA and has a radius $R + \varepsilon \equiv R_1$. Equation (11) was still used for calculating k_D , but now k_b was the rate for being absorbed by a toroidal surface. This rate is given by Eq. (5) with χ_0 given by L/R_1 . The q-surface was still spherical and, thus, Eq. (12b) remained valid for k_q .

B. Simulation details

The circular DNA was held fixed and the protein was modeled as a freely diffusing point particle (the finite size of the protein was accounted for by inflating the radius of the DNA cross section into R , the protein-DNA contact distance). The proposed displacement at the next time step was given by

$$\Delta \mathbf{r} = (2D\Delta t)^{1/2} \mathcal{R}, \quad (15)$$

where Δt is the time step and \mathcal{R} is a vector of random numbers with a normal distribution. As described already, the surface potential could change whether the move was accepted. The boundary conditions on the DNA surface and the q-surface also could affect the fate of a move. Reflecting boundary conditions (over the DNA surface with $|\phi| > \phi_0$) and absorbing boundary conditions (over the reaction patch, with $|\phi| < \phi_0$, and the q-surface) were treated in a simple manner. If a proposed move brought the protein across a reflecting boundary, the old position was reused. If the proposed move brought the protein across an absorbing boundary, the trajectory was terminated and labeled as being absorbed either by the reactive patch or by the q-surface. Variable time steps were used, with those near reflecting and absorbing boundaries taking very small values.

The following parameters were used. The radius of the DNA cross section was fixed at $R=30$ Å. The surface layer had a width of $\varepsilon=3$ Å. The radius of the q-surface was set to $q=13(L+R_1)$. The following values of the torus radius L were studied: 50, 100, 200, 300, and 400 Å. The width of the reactive patch varied from 0 to the full length $2\pi L$ of the DNA. The exit probability was limited to the following val-

TABLE I. Comparison of analytical results for the association rates to specific sites (with width $h=3 \text{ \AA}$) on circular and linear DNA.

$L \text{ (\AA)}$	100			200			300		
λ	0.5	5	50	0.5	5	50	0.5	5	50
$\frac{k_D(\text{circular})}{4\pi D} \text{ (\AA)}$	21.51	42.92	78.09	21.97	44.88	99.74	22.06	45.24	106.9
$\frac{k_D(\text{linear})}{4\pi D} \text{ (\AA)}$	22.14	45.24	90.09	22.14	45.50	106.7	22.14	45.52	110.7
Rel. diff. (%)	2.8	5.1	13	0.7	1.4	6.5	0.4	0.6	3.4

ues: $1, 2_{-1}, 2^{-2}, \dots$, and 2^{-14} . The number of trajectories for each set of parameters was 10 000, rendering the statistical error in the calculated rates at $\sim 0.5\%$, as estimated according to binomial statistics.

IV. RESULTS

A. Comparison of k_D between circular and linear DNA

Table I shows a comparison between the results for the association rate with a circular DNA, calculated from the analytical solution of Eq. (9), and the corresponding results for a linear DNA with the same cross section and length, calculated from the analytical solution published previously.⁶ The length of the linear DNA is equal to that for the circular DNA, as given by $2\pi L$. On the linear DNA, the reactive patch is centrally located along the length, and the ends of the surface layer are treated as reflecting. In addition, the linear DNA is extended from both ends to infinity, but the extended portions lack the capacity for nonspecific binding.

The difference in k_D between circular and linear DNA is quite small at $L=300 \text{ \AA}$ and only becomes significant when L is reduced to 100 \AA (corresponding to ~ 180 base pairs). Note that the difference is systematic in that the values of k_D for the circular DNA are consistently lower than the counterparts for linear DNA and the differences increase with increasing λ .

B. Comparison of analytical solution with Brownian dynamics simulations

To connect the association rate obtained from Brownian dynamics simulations with the analytical solution, we need to determine the precise relation between the exit probability p and the dimensionless parameter λ . We adjusted the constant α in the assumed relation between λ and p given by Eq. (14) to minimize the difference between the simulation results for k_D and the corresponding results calculated from the analytical solution of Eq. (9) for L at 50, 100, 200, 300, and 400 \AA , h from 5 \AA to the full DNA length $2\pi L$, and p at 2^{-4} , 2^{-8} , and 2^{-11} . As Fig. 2 shows, a value of 0.022 for α leads to very good agreement between simulation and the exact result given by Eq. (5) when the reactive patch covers the whole DNA surface, and agreement within $\sim 10\%$ between simulation and the analytical solution of Eq. (9) over the wide ranges of L , h , and p . Compared with the exact result at $h=2\pi L$ and the simulation results for a full range of h , the analytical solution appears to underestimate k_D modestly due

to the use of the constant-flux approximation. Underestimation by the constant-flux approximation is known.¹⁷

Figure 3 presents a comparison between simulation and the analytical solution for L at 50, 100, 200, 300, and 400 \AA , p at $1, 2^{-1}, 2^{-2}, \dots$, and 2^{-14} , and h fixed at 3 \AA . Agreement is again seen over the expanded range of p values.

V. DISCUSSION

We have derived an analytical solution for the association rate of a protein with a specific site on a circular DNA

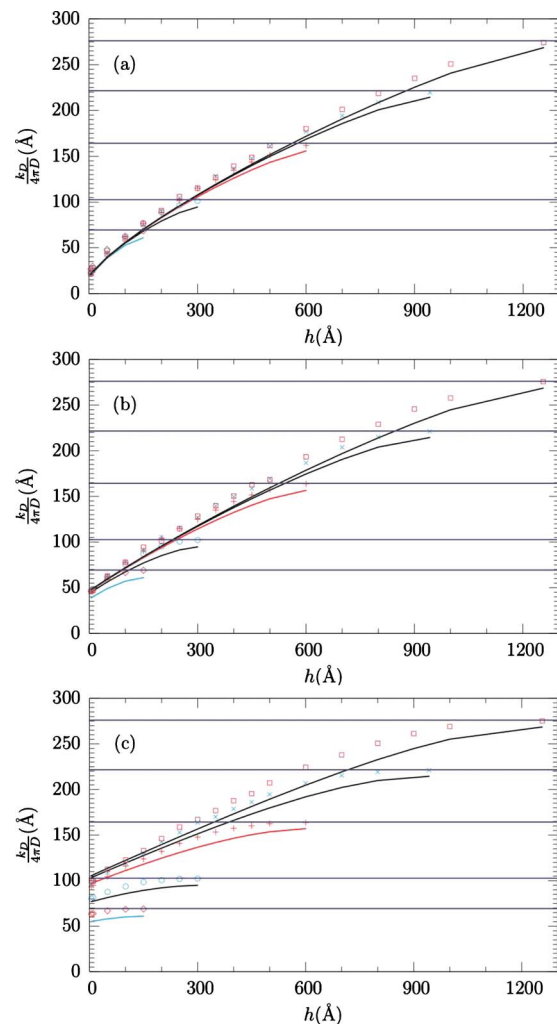


FIG. 2. (Color online) Comparison of simulation results for k_D (symbols) and the analytical solution of Eq. (9) (curves). Each curve corresponds to a particular value of L , at 50, 100, 200, 300, and 400 \AA (from bottom to top). The exact results for k_D given by Eq. (5) when the whole DNA surface is absorbing (i.e., $h=2\pi L$) are shown as horizontal lines. The exit probability p is 2^{-4} , 2^{-8} , and 2^{-11} , respectively, in (a), (b), and (c).

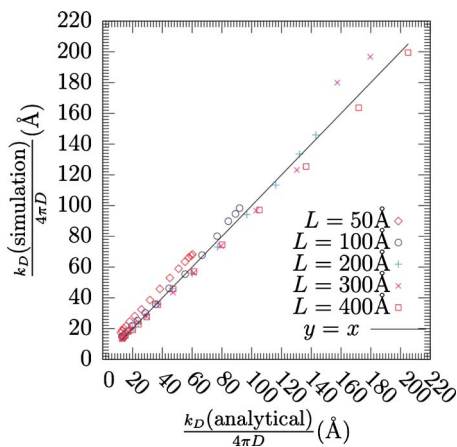


FIG. 3. (Color online) Comparison simulation results for k_D and the analytical solution of Eq. (9). The pairs of values for k_D at each L value are represented by a distinct symbol (shown in the figure). For each L value, results for 15 values of p , at $1, 2^{-1}, 2^{-2}, \dots$, and 2^{-14} , are shown (from low to high). The corresponding value for the dimensionless parameter λ is given by Eq. (14) with $\alpha=0.022$.

and found that the results are very close to those found previously for a linear DNA. In a number of experimental studies,¹³ the question of whether special end effects are at play in a linear DNA is addressed through comparing against the corresponding circular DNA. Our work reported here establishes a sound basis for these studies.

Below we further discuss several issues that arose from our study.

A. The small difference in k_D between circular and linear DNA

Though the difference in k_D between circular and linear DNA is small, it is nonetheless interesting to explore the reasons for the difference. As noted, the difference is systematic in that the values of k_D for a circular DNA are always lower than those for the corresponding linear DNA. A number of factors can possibly contribute to the lower k_D of the circular DNA.

First, in the solution of k_D for a finite linear DNA, the DNA is extended to infinity with inert flanking tails that lack the capacity for nonspecific binding. The extension was introduced by Berg and Ehrenberg⁷ to circumvent the mathematical difficulty posed by the finite DNA but its effect was never quantitatively evaluated. As pointed out by Berg and Ehrenberg, the extension effectively introduces an excluded-volume effect. It is expected to cause an increase in k_D since the extension reduces the search space.

Second, the analytical solutions of k_D for both the linear and the circular DNA are obtained by using the constant-flux approximation. It is known that this approximation can cause an underestimate of the association rate, and the amount of underestimation can be different for different cases.¹⁷ It is possible that the approximation causes a larger underestimation for the circular DNA than for the linear DNA.

Third, in the analytical solution of k_D obtained here for the circular DNA, the surface diffusion coefficient in the transverse direction (i.e., around the cross section of the DNA) is assumed to be zero. In the corresponding linear

DNA, because of axial symmetry, the value of the diffusion coefficient in the transverse direction does not affect the association rate. In the case of a circular DNA, the diffusion coefficient in the transverse direction does affect the association rate, and a nonzero diffusion coefficient is expected to increase k_D . In general, a speed up in surface diffusion will lead to an increase in k_D . A simple analogy is the rate constant for diffusive barrier crossing, which (as given by the high-friction limit of Kramers' theory²⁷) is proportional to the diffusion constant, even though diffusion can bring a particle *both* to the product energy well *and* back to the reactant energy well.

Finally, the shape of the DNA, which provides sites for nonspecific binding, seems to present opposite scenarios in affecting k_D . When binding to multiple specific sites, it is known that the different specific sites can interfere with each other, and the rate enhancement by the increase in specific sites is maximal when they are maximally separated.^{28,29} One may expect a similar behavior for nonspecific binding sites. That is, nonspecific binding sites that are more disperse could afford them a higher ability to draw the protein from the bulk toward them. The nonspecific binding sites are more disperse in the linear DNA than in the circular DNA; from this consideration, the rate enhancement by nonspecific binding in the linear DNA might be expected to be greater than in the circular counterpart. On the other hand, a protein molecule newly dissociated from the circular DNA, as opposed to the linear DNA, is also more likely to be near nonspecific binding sites and hence more likely to rebind. Which of these two opposing scenarios has the upper hand in affecting k_D cannot be resolved by our analytical theory. Brownian dynamics simulations perhaps can provide clues.

B. Caveats of the analytical solution for k_D

The model leading to the analytical solution for the diffusion-controlled rate of association of a protein with a circular DNA has a number of limitations. First, the DNA is modeled as rigid. Under physiological salt concentrations, DNA have a persistence length of ~ 500 Å. DNA can be reasonably modeled as rigid up to, say, twice the persistence length, which corresponds to $L=160$ Å. Longer DNA become flexible, which perhaps serves to blur the difference between circular and linear DNA. The main conclusion of our study, that the difference in k_D between circular and linear DNA is small when the DNA length is long and only becomes significant when L is reduced to 100 Å, is thus unaffected even when DNA flexibility is taken into consideration.

Another potential criticism of our model concerns the path of diffusion in the surface layer. It was proposed that some proteins, while nonspecifically bound, move along the major groove of the DNA, thus tracing a helical path.³⁰ However, Kampmann³¹ pointed out that no direct experimental support for this proposal has been published. Indeed, Kampmann measured the effects of obstacles and a Holiday junction on the processivity ratio of EcoRI endonuclease and interpreted his results as being consistent with two-dimensional diffusion on the DNA surface but not with dif-

fusion along a helical path. If proteins do follow a helical path on the DNA surface, an analytical solution for k_D is unlikely to be found.

The analytical solution and the Brownian dynamics simulations presented here apply to the dilute conditions studied by *in vitro* experiments. In cells, proteins become confined and crowded (see Ref. 32 for a recent review). Accounting for such complications falls outside the scope of the present study.

C. Brownian dynamics simulations for calculation of k_D

We have tested a promising procedure for handling surface potentials and modified an algorithm for calculating k_D through the use of a starting surface, which allows for a significant increase in calculation speed. Some technical details are worth further exploring. For example, the spherical shape of the original b-surface was designed to ensure the rigorous calculation of the association rate.^{21,22} There is some uncertainty regarding the rigor of the algorithm when the b-surface takes a toroidal shape, though the favorable comparison between simulation results and the analytical solution strongly suggest that whatever error introduced by the use of the toroidal b-surface is not significant. This problem can be avoided by implementing our alternative algorithm [Eq. (13)]²³ to calculate k_D .

The procedure for handling surface potentials still has one limitation: the constant α appearing in the relation between the exit probability p and the dimensionless parameter λ is not determined *a priori*. It is possible that analytical results regarding α can be obtained for simple geometries. Such analytical results and additional experimentation through Brownian dynamics simulations may lead to a determination of α . This determination may rely on an explicit account of anisotropic diffusion in the surface layer.

With these technical improvements, it will be desirable to use Brownian dynamics simulations to study analytically intractable problems, such as more realistic shapes for the protein and the DNA and the calculation of other interesting quantities like processivity ratio.^{6,9,11,31,33,34}

D. Paradox regarding an optimal sliding length

In some works,^{5,10} a sliding length along the DNA has been introduced. In our notation, the sliding length is given by

$$l_s = \left(\frac{D_{||} K_{ns} R}{D} \right)^{1/2} = \lambda^{1/2} R. \quad (16)$$

It was found that the reaction time, which is equivalent to the mean first passage time³⁵ for reaching the specific site, exhibits a minimum at an “optimal” sliding length.¹⁰ In contrast, the association rate is a monotonically increasing function of the dimensionless parameter λ (see Figs. 2 and 3) and hence of the sliding length l_s . Apparently, there is a paradox regarding whether an optimal sliding length exists.

The paradox can be resolved by the following reasoning. The reason for an increase in the reaction time at very high affinity for nonspecific binding (large l_s) is that, once reach-

ing the DNA surface, the protein gets stuck in nonspecific binding sites and thus takes an excessive amount of time before finally finding the specific site. On the other hand, as far as the association rate is concerned, what matters is the final fate of the trajectories [see Eq. (11)]; the amount of time it takes to reach the final destination is of no concern. The higher the affinity for nonspecific binding, the higher the chance the protein will make one-dimensional excursions along the DNA surface instead of three-dimensional search in the bulk solution. This reduction in dimensionality is ultimately the reason for the rate enhancement of nonspecific binding.

ACKNOWLEDGMENTS

This work was supported, in part, by NIH under Grant No. GM058187.

APPENDIX A: DERIVATION OF EQUATION (4)

In the absence of a potential, the steady-state pair distribution function $P(\mathbf{r})$ satisfies the Laplace equation and approaches 1 at infinite separation. In toroidal coordinates (ξ, ζ, ϕ) , $P(\mathbf{r})$ can be written as³⁶

$$P(\mathbf{r}) = 1 - (2\chi - 2 \cos \zeta)^{1/2} \times \sum_{n=0, m=0}^{\infty} A_{nm} P_{n-1/2}^m(\chi) \cos n\zeta \cos m\phi, \quad (A1)$$

where A_{nm} are coefficients to be determined. It is reminded that $\chi = \cosh \xi$. The mixed boundary conditions on the torus surface, specified by $\chi = \chi_0 \equiv L/R$, are treated by the constant-flux approximation,¹⁷ which takes the form

$$-[\mathbf{n} \cdot \mathbf{J}(\mathbf{r})]_{\chi=\chi_0} = -D \frac{\chi_0 - \cos \zeta}{c} \left[\frac{\partial P(\mathbf{r})}{\partial \xi} \right]_{\xi=\xi_0} = J_0 H(\phi), \quad (A2)$$

where $H(\phi)$ is 1 if $|\phi| < \phi_0$ and 0 otherwise. The constant J_0 is determined by the requirement that the original absorbing boundary condition is satisfied on average

$$\int_{\chi=\chi_0; |\phi| < \phi_0} ds P(\mathbf{r}) = 0, \quad (A3)$$

where the element of surface area on the torus is given by

$$ds = \frac{c^2 \sinh \xi_0}{(\chi_0 - \cos \zeta)^2} d\zeta d\phi.$$

Note that $\sinh \xi_0 = (\chi_0^2 - 1)^{1/2} = c/R$.

Substituting Eq. (A1) into Eq. (A2) and using the expansions³⁶

$$(2\chi - 2 \cos \zeta)^{-1/2} = \pi^{-1} \sum_{n=0}^{\infty} \varepsilon_n Q_{n-1/2}(\chi) \cos n\zeta \quad (A4)$$

and

$$H(\phi) = \frac{\phi_0}{\pi} \sum_{m=0}^{\infty} \varepsilon_m \frac{\sin m \phi_0}{m \phi_0} \cos m \phi, \quad (\text{A5})$$

we find the following relations for the coefficients $u_{nm} \equiv (D/cJ_0\phi_0)A_{nm}$:

$$a_{nm}u_{n-1,m} + b_{nm}u_{nm} + c_{nm}u_{n+1,m} = r_{nm}. \quad (\text{A6})$$

The constants in these relations are

$$a_{nm} = \begin{cases} 0, & n = 0 \\ -2P'_{-1/2}{}^m(\chi_0), & n = 1 \\ -P'_{n-3/2}{}^m(\chi_0), & n > 1, \end{cases} \quad (\text{A7a})$$

$$b_{nm} = P_{n-1/2}{}^m(\chi_0) + 2\chi_0 P'_{n-1/2}{}^m(\chi_0), \quad (\text{A7b})$$

$$c_{nm} = -P'_{n+1/2}{}^m(\chi_0), \quad (\text{A7c})$$

$$r_{nm} = \frac{2\varepsilon_n Q_{n-1/2}(\chi_0)}{\pi^2 \sinh^2 \xi_0} \frac{\varepsilon_m \sin m \phi_0}{m \phi_0}. \quad (\text{A7d})$$

For each m , Eq. (A6) is a tridiagonal system of linear equations for the set of coefficients u_{nm} with varying n , which can be easily solved for.³⁷ Once u_{nm} are found, Eq. (A3) allows the constant flux J_0 to be determined. The result is

$$\frac{J_0 \phi_0}{\pi D} = \frac{LR^2}{4c^4} \left(- \sum_{n=0, m=0}^{\infty} u_{nm} P_{n-1/2}{}^m(\chi_0) \times Q'_{n-1/2}(\chi_0) \frac{\sin m \phi_0}{m \phi_0} \right)^{-1}. \quad (\text{A8})$$

The diffusion-controlled rate constant is given by [see Eq. (2)]

$$k_{D0} = J_0 \int_{\chi=\chi_0; |\phi|<\phi_0} ds = 4\pi LRJ_0 \phi_0, \quad (\text{A9})$$

which upon substitution of Eq. (A8) leads to Eq. (4). Evaluation of the toroidal functions are done using programs presented by Segura and Gil.³⁸

$$f_{nm} = \begin{cases} P_{1/2}{}^m(\chi_0) + 2\chi_0 P'_{1/2}{}^m(\chi_0) + \frac{\lambda R(2\chi_0^2 + 3/2)m^2}{c \sinh^3 \xi_0} P_{1/2}{}^m(\chi_0), & n = 1 \\ P_{n-1/2}{}^m(\chi_0) + 2\chi_0 P'_{n-1/2}{}^m(\chi_0) + \frac{\lambda R(2\chi_0^2 + 1)m^2}{c \sinh^3 \xi_0} P_{n-1/2}{}^m(\chi_0), & n \neq 1, \end{cases} \quad (\text{B3c})$$

APPENDIX B: DERIVATION OF EQUATION (9)

The effect of a potential present in a very thin surface layer can be accounted for as a boundary condition on the outer boundary of the surface layer.⁶ Integrating both sides of Eq. (6) over the width of the surface layer, we find

$$-D \frac{\chi_1 - \cos \xi}{c} \left[\frac{\partial P(\mathbf{r})}{\partial \xi} \right]_{\xi=\xi_1} + D_{||} K_{ns} \frac{(\chi_1 - \cos \xi)^2}{c^2 \sinh^2 \xi_1} \left[\frac{\partial^2 P(\mathbf{r})}{\partial \phi^2} \right]_{\xi=\xi_1} = J_1 H(\phi), \quad (\text{B1})$$

where we have used the constant-flux approximation. Comparison of this equation with Eq. (A2) shows that the effect of the surface potential is captured by an additional term containing K_{ns} . The evaluation on the left-hand side of Eq. (B1) technically is done at the outer boundary $\chi=\chi_1$, but since this equation is derived under the condition of an extremely thin surface layer, from here on χ_1 will be replaced by χ_0 . The constant J_1 is again determined by the constraint of Eq. (A3).

The solution for k_D essentially follows the steps presented in Appendix A for the case without a potential. Again, the expansion coefficients $v_{nm} \equiv (D/cJ_1\phi_0)A_{nm}$ satisfy a set of linear equations. For $m=0$, the additional term in Eq. (B1) vanishes, so the linear equations are the same as those for the case without a potential. For $m>0$, the additional term expands the structure of the linear equations from tridiagonal to pentadiagonal. Equation (A6) becomes

$$d_{nm}v_{n-2,m} + e_{nm}v_{n-1,m} + f_{nm}v_{nm} + g_{nm}v_{n+1,m} + h_{nm}v_{n+2,m} = r_{nm}, \quad (\text{B2})$$

and the constants are given by

$$d_{nm} = \begin{cases} 0, & n = 0, 1 \\ \frac{\lambda R m^2}{c \sinh^3 \xi_0} P_{-1/2}{}^m(\chi_0), & n = 2 \\ \frac{\lambda R m^2}{2c \sinh^3 \xi_0} P_{n-5/2}{}^m(\chi_0), & n > 2, \end{cases} \quad (\text{B3a})$$

$$e_{nm} = \begin{cases} 0, & n = 0 \\ -2P'_{-1/2}{}^m(\chi_0) - \frac{4\lambda R \chi_0 m^2}{c \sinh^3 \xi_0} P_{-1/2}{}^m(\chi_0), & n = 1 \\ -P'_{n-3/2}{}^m(\chi_0) - \frac{2\lambda R \chi_0 m^2}{c \sinh^3 \xi_0} P_{n-3/2}{}^m(\chi_0), & n > 1, \end{cases} \quad (\text{B3b})$$

$$g_{nm} = -P'_{n+1/2}(\chi_0) - \frac{2\lambda R \chi_0 m^2}{c \sinh^3 \xi_0} P'_{n+1/2}(\chi_0), \quad (\text{B3d})$$

$$h_{nm} = \frac{\lambda R m^2}{2c \sinh^3 \xi_0} P'_{n+3/2}(\chi_0). \quad (\text{B3e})$$

The dimensionless parameter λ is given by Eq. (10). r_{nm} are unchanged from the case without a potential and are given by Eq. (A7d). The coefficients ν_{nm} are found by numerically solving the set of linear equations.³⁷

The relation between the constant J_1 and the coefficients ν_{nm} is identical to that between J_0 and the coefficients u_{nm} , which can be found in Eq. (A8). Similarly, the relation between the diffusion-controlled association rate k_D in the present case and J_1 is identical to that between the counterparts, k_{D0} and J_0 , respectively, in the case without a potential. This is given by Eq. (A9). We finally arrive at Eq. (9).

¹F. Yang, I. V. Ouporov, C. Fernandes, D. Motriuk, and K. A. Thomasson, *J. Phys. Chem. B* **105**, 12601 (2001).

²G. Adam and M. Delbruck, in *Structural Chemistry and Molecular Biology*, edited by N. Davidson (W. H. Freeman, San Francisco, 1968), p. 198.

³O. G. Berg, R. B. Winter, and P. H. von Hippel, *Biochemistry* **20**, 6929 (1981).

⁴H.-X. Zhou and A. Szabo, *Phys. Rev. Lett.* **93**, 178101 (2004).

⁵S. E. Halford and J. F. Marko, *Nucleic Acids Res.* **32**, 3040 (2004).

⁶H.-X. Zhou, *Biophys. J.* **88**, 1608 (2005).

⁷O. G. Berg and M. Ehrenberg, *Biophys. Chem.* **15**, 41 (1982).

⁸B. P. Belotserkovskii and D. A. Zarling, *J. Theor. Biol.* **226**, 195 (2004).

⁹M. Coppey, O. Benichou, R. Voituriez, and M. Moreau, *Biophys. J.* **87**, 1640 (2004).

¹⁰K. V. Klenin, H. Merlitz, J. Langowski, and C.-X. Wu, *Phys. Rev. Lett.* **96**, 018104 (2006).

¹¹B. J. Terry, W. E. Jack, and P. Modrich, *J. Biol. Chem.* **260**, 13130 (1985).

¹²P. T. Singer and C.-W. Wu, *J. Biol. Chem.* **262**, 14178 (1987).

¹³M. A. Surby and N. O. Reich, *Biochemistry* **35**, 2209 (1996).

¹⁴A. Jeltsch and A. Pingoud, *Biochemistry* **37**, 2160 (1998).

¹⁵C. S. Park, F. Y.-H. Wu, and C.-W. Wu, *J. Biol. Chem.* **257**, 6950 (1982).

¹⁶A. Graneli, C. C. Yeykal, R. B. Robertson, and E. C. Greene, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 1221 (2006).

¹⁷D. Shoup, G. Lipari, and A. Szabo, *Biophys. J.* **36**, 697 (1981).

¹⁸N. N. Lebedev, I. P. Skalskaya, and Y. S. Uflyand, *Worked Problems in Applied Mathematics* (Dover, New York, 1965).

¹⁹H. C. Berg and E. M. Purcell, *Biophys. J.* **20**, 193 (1977).

²⁰H.-X. Zhou, A. Szabo, J. F. Douglas, and J. B. Hubbard, *J. Chem. Phys.* **100**, 3821 (1994).

²¹S. H. Northrup, S. A. Allison, and J. A. McCammon, *J. Chem. Phys.* **80**, 1517 (1984).

²²H.-X. Zhou, *J. Chem. Phys.* **92**, 3092 (1990).

²³H.-X. Zhou, *J. Phys. Chem.* **94**, 8794 (1990).

²⁴H.-X. Zhou, *J. Phys. Chem. B* **101**, 6642 (1997).

²⁵H. Merlitz, K. V. Klenin, C.-X. Wu, and J. Langowski, *J. Chem. Phys.* **125**, 014906 (2006).

²⁶The outer boundary of the surface layer introduced here is concentric with the cross section of the enclosed DNA. This is slightly different from the surface layer used in Sec. II B for deriving the analytical solution of the association rate. There the outer boundary of the surface layer, as specified by $\chi=\chi_1$, has a slightly eccentric cross section.

²⁷H. A. Kramers, *Physica (Amsterdam)* **7**, 284 (1940).

²⁸R. Zwanzig, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 5856 (1990).

²⁹R. Zwanzig and A. Szabo, *Biophys. J.* **60**, 671 (1991).

³⁰J. Sun, H. Viadiu, A. K. Aggarwal, and H. Weinstein, *Biophys. J.* **84**, 3317 (2003).

³¹M. Kampmann, *J. Biol. Chem.* **279**, 38715 (2004).

³²H.-X. Zhou, G. Rivas, and A. P. Minton, *Annu. Rev. Biophys.* **37**, 375 (2008).

³³N. P. Stanford, M. D. Szczelkun, J. F. Marko, and S. E. Halford, *EMBO J.* **19**, 6546 (2000).

³⁴H.-X. Zhou, *Phys. Biol.* **2**, R1 (2005).

³⁵A. Szabo, K. Schulten, and Z. Schulten, *J. Chem. Phys.* **72**, 4350 (1980).

³⁶N. N. Lebedev, *Special Functions and Their Applications* (Dover, New York, 1972).

³⁷W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes* (Cambridge University Press, Cambridge, 1986).

³⁸J. Segura and A. Gil, *Comput. Phys. Commun.* **124**, 104 (2000).