

The helical structure of DNA facilitates binding

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2016 J. Phys. A: Math. Theor. 49 364002

(<http://iopscience.iop.org/1751-8121/49/36/364002>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 130.238.239.78

This content was downloaded on 09/09/2016 at 13:14

Please note that [terms and conditions apply](#).

You may also be interested in:

[Diffusion properties of active particles with directional reversal](#)

R Großmann, F Peruani and M Bär

[The predicted role of steric specificity in crowding-mediated effects on reversible biomolecular association](#)

J D Powers, B T Castle and D J Odde

[Coarse-grained modeling of protein unspecifically bound to DNA](#)

Carlo Guardiani, Massimo Cencini and Fabio Cecconi

[Time scale of diffusion in molecular and cellular biology](#)

D Holcman and Z Schuss

[Role of dynamic capsomere supply for viral capsid self-assembly](#)

Marvin A Boettcher, Heinrich C R Klein and Ulrich S Schwarz

[Multiparticle collision dynamics for diffusion-influenced signaling pathways](#)

R Strehl and K Rohlf

[Control of flux by narrow passages and hidden targets in cellular biology](#)

D Holcman and Z Schuss

The helical structure of DNA facilitates binding

Otto G Berg, Anel Mahmutovic¹, Emil Marklund¹ and Johan Elf

Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, Husarg. 3, SE-75124 Uppsala, Sweden

E-mail: Johan.Elf@icm.uu.se

Received 27 April 2016, revised 1 July 2016

Accepted for publication 13 July 2016

Published 16 August 2016



CrossMark

Abstract

The helical structure of DNA imposes constraints on the rate of diffusion-limited protein binding. Here we solve the reaction–diffusion equations for DNA-like geometries and extend with simulations when necessary. We find that the helical structure can make binding to the DNA more than twice as fast compared to a case where DNA would be reactive only along one side. We also find that this rate advantage remains when the contributions from steric constraints and rotational diffusion of the DNA-binding protein are included. Furthermore, we find that the association rate is insensitive to changes in the steric constraints on the DNA in the helix geometry, while it is much more dependent on the steric constraints on the DNA-binding protein. We conclude that the helical structure of DNA facilitates the nonspecific binding of transcription factors and structural DNA-binding proteins in general.

 Online supplementary data available from stacks.iop.org/JPA/49/364002/mmedia

Keywords: reaction–diffusion equation, steric constraints, helix geometry, diffusion limited

(Some figures may appear in colour only in the online journal)

Diffusion-limited reactions play a central role in living cells since the macromolecular density is pushed close to crystalline concentrations [1]. Diffusion-limited reaction rates are well characterized in the case of reactive spheres, where the microscopic reaction–diffusion equations have been previously solved [2–12]. The binding time is in this case described by a

¹ Equal contributions.

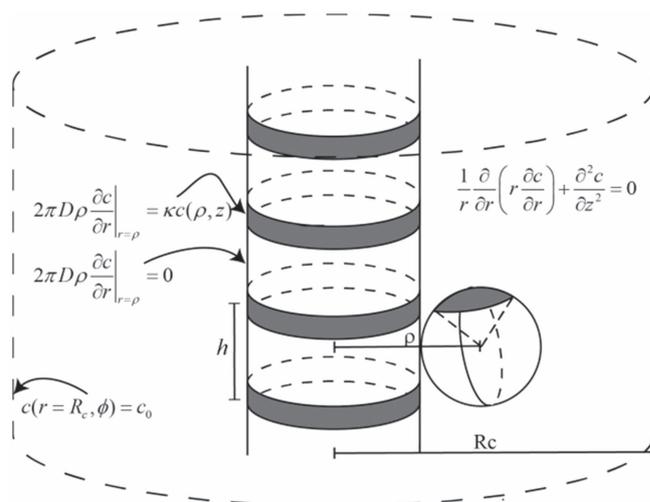


Figure 1. Reaction–diffusion equation and boundary conditions for example geometry. Only when the reactive patch on the sphere (gray) is in contact with the reactive patch on the cylinder do they react with rate κ . The sphere moves around the cylinder with translational diffusion rate D . When the sphere reaches a distance R_c away from the cylinder it is equally likely to bind another DNA segment.

sum of two parts where the first part is due to reaction at contact and the second part is due to the time for reactants to meet by diffusion [3, 6]. When the spheres are equally reactive over their entire surface, the diffusion-influenced reaction rate k_a is given by $1/k_a = 1/k + 1/4\pi\rho D$, where ρ is the reaction radius, D is the relative diffusion rate and k is the microscopic rate of binding for the local concentration of reactants in contact with the target sphere. This expression implies that when diffusion is fast $k_a = k$ and when diffusion is slow $k_a = 4\pi\rho D$. The latter relation is the diffusion limit, also known as the Smoluchowski limit.

The diffusion-limited protein binding in the case of a helical DNA-like geometry is more complex since binding to any position on a polymer extended in one-dimension (1D) is a 2D reaction–diffusion problem in which the DNA binding proteins are spatially correlated to their DNA target [13, 14]. This implies that proteins return to the same DNA segment many times before leaving their current segment and thereby ‘forgetting’ their localization. While mathematically complex, protein binding to DNA is central for life as we know it. For example, proteins need to bind DNA to turn specific genes ‘on’ and ‘off’, RNA polymerases need to initiate transcription of genes, and structural proteins need to bind DNA for efficient storage, replication and repair. The characteristic double helix structure of DNA [15] has functional importance in that the two strands enable semi-conservative replication [16] and that it contributes to structural stability in long term storage.

The question we now ask is how the helical structure of DNA contributes to the rate of diffusion-limited binding, which, in turn, is important for accessing the genetic information. In particular we ask if it matters for the rate of binding if the reactive patches are twisted into a helix like the grooves on DNA, rather than if they were straight?

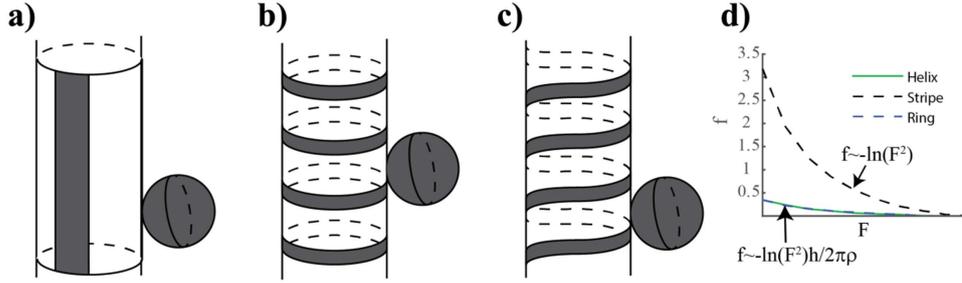


Figure 2. Illustration of the reaction geometries. (a) Stripe. (b) Ring. (c) Helix. The reaction patch covers a fraction F of the cylinder. (d) The steric factor f plotted as a function of F in the different geometries.

Results

Reactive patches on the DNA

Our approach for investigating the importance of the helical structure on the rate of protein binding to DNA is to solve the microscopic reaction–diffusion equations with specific configurations of reactive patches (see example equations in figure 1 and methods or the supplementary information (SI) for details of how they are solved). The DNA is represented by a cylinder and the protein by a sphere with reactive regions, or patches, on them.

For a periodically repeating configuration of reactive patches on the cylinder and a fully reactive sphere, the solution can be written on the simple closed form (equation (S11) in the SI)

$$k_a = \frac{2\pi D\ell}{1/\alpha F + \ln(R_c/\rho) + f(F)}, \quad (1)$$

where ℓ is the length of a base pair, F is the fraction of the cylinder surface that is reactive, ρ is the reaction radius, D is the relative translational diffusion rate constant, and R_c is the distance from the DNA where it is equally likely for the protein to bind another DNA segment. There are two particularly important parameters, $\alpha = \kappa/2\pi D = k/2\pi D\ell F$ and $f(F)$. Here, α is a unit-less measure related to the degree of diffusion control.

Under reaction-limited conditions ($\alpha F \ll 1$), i.e. where reaction upon contact is unlikely, $k_a = 2\pi D\ell\alpha F = k$. Here, the rate depends only on the size of the reactive patches and not on their geometric arrangement. In the limit of large αF , the diffusion-limited result is

$$k_a = 2\pi D\ell / [\ln(R_c/\rho) + f(F)] = 2\pi D\ell / [\ln(R_c/\rho e^{-f})], \quad (2)$$

which implies that the rate of binding is determined only by diffusion and geometric factors. The shapes and arrangements of the reactive patches all end up in the steric factor $f(F)$, where $f(1) = 0$. Thus f depends on the fractions of the surfaces that are reactive and their shapes, but not on the rates of reaction or diffusion. This separation of terms implies that the patch distribution generally can be seen as a change in the reaction radius $\rho_{\text{eff}} = \rho e^{-f}$, which simplifies the thinking about geometry in the reaction–diffusion process.

We have derived closed-form expressions for the steric factor $f(F)$ for two different geometries: a reactive stripe along a cylinder and rings of reactive regions around a cylinder (figure 2). The full expressions for the steric factor are given in the SI section for the ring case and in [17] for the stripe case. We will here first focus on approximate solutions in cases where F is small. For a reactive stripe along the cylinder, the steric factor

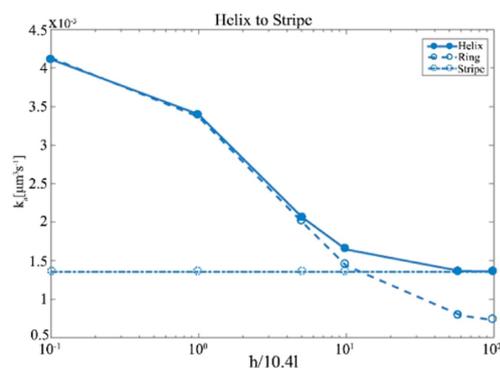


Figure 3. The diffusion-limited association rate as a function of the periodicity length with a fully reactive protein, $F_p = 1$, and a DNA cylinder with $F = 0.1$. The biological periodicity corresponds to $h/10.4\ell = 1$. When the helical reactive patch on the DNA cylinder is stretched out, it becomes the stripe case. Blue dots correspond to the simulated values.

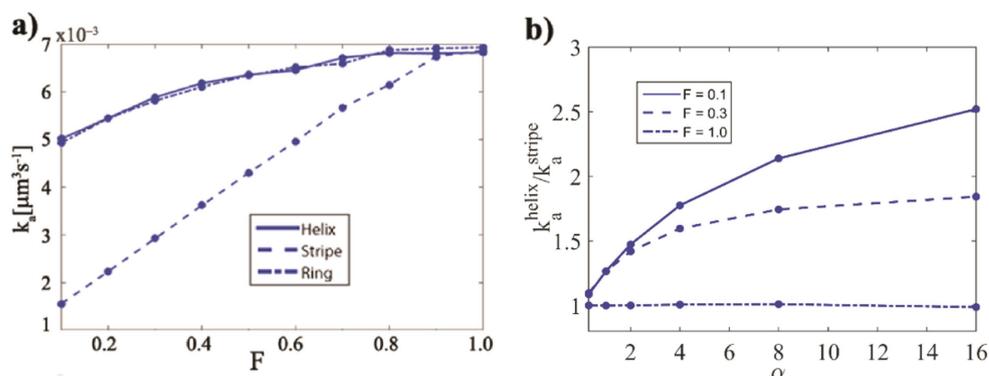


Figure 4. (a). The diffusion-limited association rate k_a plotted as a function of the reactive patch fraction on the cylinder, F , in the three different geometries shown in figures 2(a)–(c). (b) The increase in association rate to a helix compared to a stripe for different diffusion limitation (α , x -axis) and F (legend).

is $f(F) \approx -\ln(2F^2)$, equation (S13). For reactive rings around a cylinder the steric factor is $f(F) \approx -\ln(2F^2)h/2\pi\rho$, equation (S12), in the limit where the distance between rings, h , is smaller than the circumference at the reaction radius, $2\pi\rho$, which is generally the case for protein–DNA interactions. For the comparisons we have assumed that $h = 10.4\ell$, i.e. the helical repeat distance of B-DNA. The analytical solutions for the helix geometry are, however, out of our reach, but as can be seen when compared to simulations in figure 2(d), the helix case is virtually identical to the ring case (see methods for details about the simulations). The difference in the association rate between the ring and the helix geometries is expected to increase with an increasing period length where the helix case should become stripe-like. This is shown in figure 3. For the numerical calculations in the figures we have used typical values for the interaction radius ($\rho = 5.5$ nm), *in vivo* DNA density ($R_c = 14$ nm) and protein diffusion ($D = 3 \mu\text{m}^2 \text{s}^{-1}$) [17, 18].

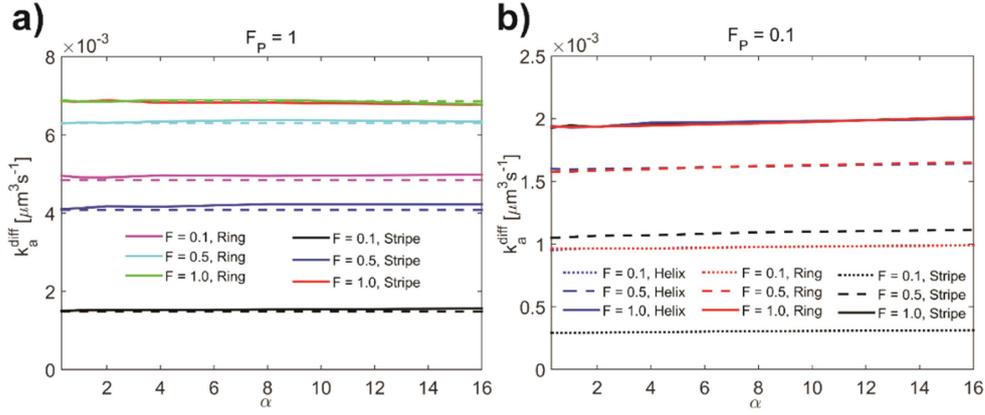


Figure 5. Diffusion-limited k_a calculated from $k_a^{\text{diff}} = k/(k/k_a - 1)$, where k_a was simulated for different values of α using different patch geometries as indicated. In panel (a), the dashed curves show the corresponding results from the analytical solutions.

Figure 4(a) shows how the overall rate k_a changes with F , i.e. the fraction of the DNA surface that is reactive, and figure 4(b) shows the ratio of association rates for the helix and stripe cases as a function of the degree of diffusion control, α , for a few different reactive surface fractions, F . The difference between the helix and stripe cases disappear both in the limit of reaction control ($\alpha F \ll 1$) and for fully reactive DNA ($F \rightarrow 1$). The ratio approaches the theoretical limit $2\pi\rho/h$ for a diffusion-limited reaction as F gets very small. However, it is significantly faster to bind to a helix than to a stripe with the same surface coverage also for conditions realistic to a DNA geometry with intermediate values of F (figure 4(a)).

Reactive patches on the protein

We now ask if the dependence of k_a on the shape of the reactive patch on DNA remains with a reactive surface patch on the protein and with differences in rotational diffusion included. Here we assume that the reactive patch on the protein is an axially symmetric region that occupies the fraction F_p of the protein surface (see figure 1). When only the fraction F_p is reactive, the association rate constant in the reaction-controlled limit can be written as

$$k = \kappa l F F_p = 2\pi D l \alpha F F_p, \quad (3)$$

which also defines α . The combination $\alpha F F_p$ is predicated by the fact that in the reaction-controlled limit, the distribution over all coordinates of the configuration space is homogeneous and the association rate must be proportional to the fraction of that space that is reactive. The analytical calculations, as well as the simulations involving a helical constraint on DNA, show that the steric constraints enter the resulting nonspecific association rate constant as an additive term to $1/\alpha F$. We conjecture that the association rate constant can be written on the same form as equation (1) also when $F_p < 1$, i.e.

$$k_a = \frac{2\pi D l}{1/\alpha F F_p + \ln(R_c/\rho) + f(F, F_p)} \quad (4)$$

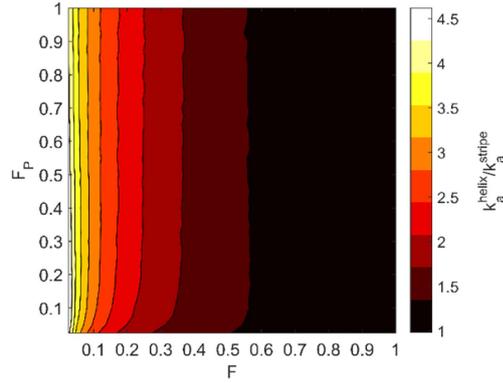


Figure 6. The ratio of association rates in the diffusion-controlled limit for helical geometry versus the stripe geometry shown as isosurfaces for different combinations of fraction of the cylinder (F) and sphere (F_p) that are reactive. The rotational diffusion is given by the Stokes–Einstein relation.

or,

$$\frac{1}{k_a} = \frac{1}{2\pi D\ell\alpha FF_p} + \frac{\ln(R_c/\rho) + f(F, F_p)}{2\pi D\ell} = \frac{1}{k} + \frac{1}{k_a^{\text{diff}}}, \quad (5)$$

When $\alpha \rightarrow 0$ (reaction limit), this gives $k_a \rightarrow k = 2\pi D\ell\alpha FF_p$, and when $\alpha \rightarrow \infty$ (diffusion limit) it gives $k_a \rightarrow k_a^{\text{diff}} = 2\pi D\ell / [\ln(R_c/\rho) + f(F, F_p)]$. The steric constraint factor, $f(F, F_p)$, may also depend on the other parameters like h , ρ and R_c , as well as on the geometric configuration of the reactive regions. It will also depend on the relationship between the diffusion constants for translation and rotation of the protein, but not on the reaction rates.

The conjecture in equation (5) implies that $k_a^{\text{diff}} = k / (k/k_a - 1)$, which should be independent of the value of α used to simulate the value of k_a . Indeed, the expected independence of α is displayed in figure 5. Figure 5(a) also shows the good agreement between the analytical solutions and the simulation results for the stripe- and ring-cases when $F_p = 1$. Figure 5(b) shows that the ring and helix geometries remain equivalent also for values of $F_p < 1$. However, as seen particularly in figure 5(b), the simulation results show a small discrepancy from the expected independence of α . This could be due to some small numerical error that increases with increasing α , where a very large number of very short-lived micro-dissociations (micro-hops) will precede a macroscopic dissociation.

In order to understand how geometry influences the diffusion-limited reaction, k_a^{diff} can be rewritten as the product $k_a^{\text{diff}} = [2\pi D\ell / \ln(R_c/\rho)] \times [1 + f/\ln(R_c/\rho)]^{-1}$, where the first factor, $2\pi D\ell / \ln(R_c/\rho)$, is the macroscopic encounter rate, i.e. the rate with which the protein and DNA come together—in arbitrary orientations. In each macroscopic encounter, the protein will leave the DNA many times in short excursions (micro-hops) before either a binding or a macroscopic dissociation occurs. During these short excursions, the protein can change orientation or location along the DNA and effectively test new orientations for reactivity. The second factor, $[1 + f/\ln(R_c/\rho)]^{-1}$, can be interpreted as the probability that a macroscopic encounter will—either directly or through micro-hops—find the reactive configuration before macroscopic dissociation occurs.

The next question is how a reactive patch on the protein influences the advantage of having a helical reactive patch on the DNA. In figure 6 we show that the relative association rate between the stripe and helix cases depends only weakly on the fraction of the sphere that

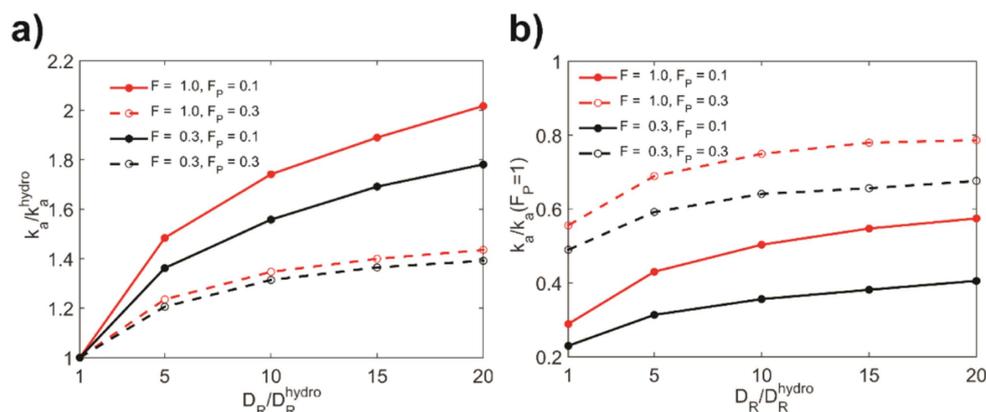


Figure 7. (a). The increase in the diffusion-limited association rate when the rotational diffusion is increased. (b) The diffusion-limited association rate for a protein with an axially symmetric reactive patch relative to the diffusion-limited association rate for a fully reactive protein ($F_p = 1$). For both plots the x -axis shows the increase of D_R relative to the hydrodynamic (Stokes–Einstein) value $D_R^{\text{hydro}} = 0.75D/R_p^2$ and the protein radius was taken as $R_p = 4.5$ nm. The values were obtained by simulations with the helix geometry.

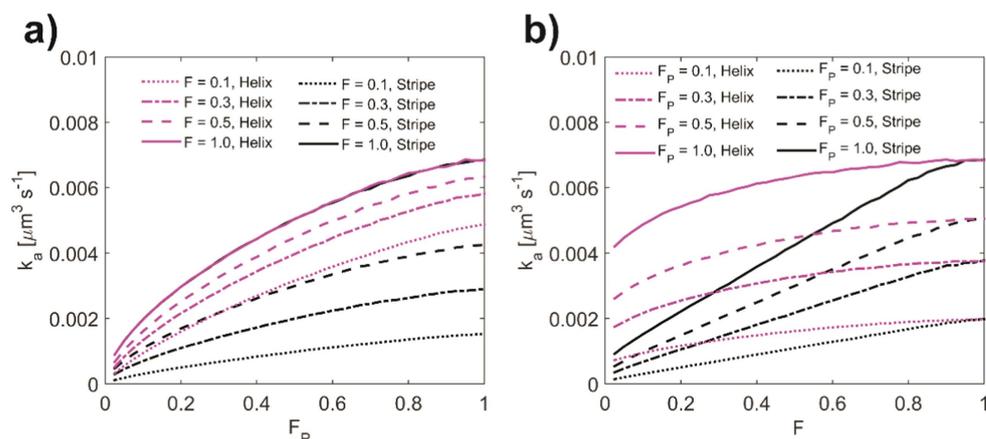


Figure 8. The dependence of the diffusion-limited k_a on the patch size of the protein (a) and on the patch size of the cylinder. (b) The magenta color corresponds to the helix geometry and the black color the stripe geometry. Note that all curves go to zero when F or F_p does.

is reactive. Thus the association rate amplification due to a particular patch distribution on the DNA does not depend on the size of the patch on the protein, although the absolute rate of binding drops with the fraction of the protein that is reactive.

The influence of the reactive patch on the protein is even smaller if we also consider that the rotational diffusion in the living cell is higher than what is expected from the Stokes–Einstein relation between translational and rotational diffusion [19]. Figure 7 shows that the effect of rotational diffusion is to broaden the reactive patch on the protein. With increasing rate of rotational diffusion the diffusion-limited association rate slowly approaches the limit of

a fully reactive protein surface ($F_p = 1$; figure 7(b)). The effects are significant but not dramatic. As expected, the largest effect is for small F_p . E.g., increasing the rotational diffusion 10 times leads to an increase of k_a by at most a factor 1.75 (if $F_p \geq 0.1$; figure 7(a)).

Two additional observations from the simulations are shown in figure 8. First, we find a nonlinear dependence of k_a on F_p for all values of F (figure 8(a)). This is similar to the sphere–sphere case when one of the spheres is fully reactive [12], figure S2 in the SI. Secondly, it is more difficult for the protein to align with the reactive patch if misaligned in the stripe geometry compared to the helical geometry. This implies that k_a is much more sensitive to changes in F for the stripe geometry than for the helical geometry (figure 8(b)). The dependence of k_a on the fraction of the cylinder that forms the reactive helical constraint is very weak in the diffusion-controlled limit; a reduction in F by 80% (from 100% to 20%) only reduces k_a by ca 20% if $F_p = 1$, or by ca 35% if $F_p = 0.2$ (figure 8(b)). In contrast, the dependence on the size of the reactive patch on a spherical protein is substantial; k_a is reduced by ca 60% when F_p is reduced by 80% (from 100% to 20%) if $F = 0.5$, or by 65% if $F = 0.2$ (figure 8(a)).

Discussion

In this work we have investigated how the non-sequence-specific association of a generic protein to DNA is influenced by restricting reactions to patches. In practice we have solved the reaction–diffusion equations for specific patch geometries and extended with simulations for the helical DNA geometry. In conclusion we have found that it is faster to bind reactive regions on DNA in the form of a helix than in the form of a straight stripe. The reason is that it is easier for a sphere to explore the nearest surface along a cylinder than to diffuse around it. As a consequence, the association rate to a helix is not sensitive to the steric constraint, F (figure 8(b)). It is also the physical reason why the helical constraint is indistinguishable from our results with periodic rings around the DNA. The increased rate of association applies directly to DNA structural proteins that bind all along the DNA, such as Fis, H-NS, and HU in bacteria [20] and histone proteins in eukaryotes. The fact that DNA is twisted into a helix rather than being a straight polymer thus makes DNA binding dynamics twice as fast at given protein concentrations. Finally, the advantage of binding to a helical patch rather than to a stripe on a cylinder is independent of the size of the patch on the protein (figure 6) but strongly dependent of the size of the patch on the DNA.

Electrostatic interactions are not explicitly considered in these calculations. Instead, some of the geometric parameters of the model are ‘effective’ and should be considered as incorporating electrostatic effects. Thus, the cylindrical capture radius, ρ , should extend as far out in the solution as determined by the reach of the electrostatic potential. Under physiological ionic conditions, the potential is strongly screened and the reach is limited [21, 22]. However, the shape of the potential will also be influenced by the pitch of the helical charge distribution along the cylinder. Thus, there may be a difference in the effective capture radius for the helical case versus straight stripe considered here. It should be noted, however, that the association rate depends logarithmically and is therefore very insensitive to the choice of the parameter ρ .

More important may be the possibility that electrostatic forces could help steer the protein into the correct orientation for binding. DNA-binding proteins often have anionic patches in locations that in the bound state end up in contact with the negatively charged phosphate backbone of the DNA [23]. Clearly this strengthens the binding constant, but electrostatic steering could also aid the association thereby reducing the effect of the steric

constraints. Thus, the parameters F and F_p would effectively become larger, particularly under low-salt conditions. Such an increase in orientational tolerance due to short-range interactions has been studied for protein–protein association [10]. The main question is how close in position and orientations the reaction partners will have to be to ensure that binding takes place. In any particular case, however, both F and F_p will be guesstimates; the main result here is to show how the association rate depends on changes in their values. In principle there is a third orientational constraint which has not been included in these calculations. That is when the reactive patch on the protein has a direction that needs to be aligned with the pitch of the helix. This may further reduce the association rate constant, but is not expected to change significantly the structure of the overall results.

Intracellular crowding [11] is considered in two ways. First it is the reason for the slow translation diffusion which is likely to push the reaction into the diffusion-limited regime. Secondly we investigate the dependence of the nonspecific association rate on the rotational diffusion, which is less hindered by crowding than the translational diffusion. Importantly, we find that the nonspecific association rate constant depends fairly weakly on the rotational diffusion constant for biologically reasonable values of F and F_p (figure 7).

For site-specific binding proteins, such as transcription factors, the situation is more complicated [18]. The first step of DNA binding is a non-specific binding event, such as described in this communication and which is made faster by the helical shape of the DNA. This can be followed by a sliding search in a 1D diffusion [14, 24–28] following the helical pitch of DNA [17, 29–33]. The helical sliding is limited hydrodynamically by the rotational motion and is therefore slower than the corresponding sliding along a straight stripe [34, 35]. Thus as described in the SI, the sliding-limited specific association rate will in general be much smaller for the helical structure than for a straight stripe. When the continuous helical track is cut off, like in the ring distribution of patches, sliding as a search mechanism will be impracticable. The situation is further complicated by crowding on DNA [36] where other proteins bound to DNA can block the search path. Only in the limit where the effective sliding length is longer than the distance between such ‘road blocks’, would the specific association be determined primarily by the non-specific rate described here. Thus, eukaryotic transcription factors that slide longer than the 50 bp between nucleosomes will have a direct advantage of the helical geometry for site specific binding.

Methods

The return-time distribution

While it is straight-forward to generate analytical solutions for the stationary association flux in the simpler geometries (fully reactive sphere and reactive stripe or rings on the DNA), in the more complex helix geometry or for a partially reactive protein, it is more convenient to use simulations of the dynamics of the dissociation process. Initially, the DNA is considered as a smooth reactive cylinder and the protein as a smooth reactive sphere; the steric constraints are accounted for later. A protein is bound to a long DNA chain. At time $t = 0$ it dissociates and starts free diffusion in solution; when and where will it rebind? The coordinate r is the distance between the cylinder axis and the center of the sphere and the angular coordinate φ is the azimuthal angle from the point on the cylinder where the protein started. The reaction radius ρ is the r -coordinate at contact, i.e. the sum of radii of the cylinder and the sphere. The DNA is considered as a straight cylinder only in the neighborhood where the protein is binding. If the protein reaches a distance $r = R_c$, it is lost to the particular DNA segment it started from and is free to rebind anywhere. R_c is defined from the density c_{ns}

(bps/volume) of DNA in solution such that $\pi R_c^2 \ell = 1/c_{ns}$, where ℓ is the length of a base pair.

If, to begin with, the motion along the cylinder axis (the z -direction) is disregarded, the free diffusion is governed by the diffusion equation in polar coordinates:

$$\frac{1}{D} \frac{\partial c}{\partial t} = \frac{\partial^2 c}{\partial r^2} + \frac{1}{r} \frac{\partial c}{\partial r} + \frac{1}{r^2} \frac{\partial^2 c}{\partial \varphi^2}. \quad (6)$$

The boundary conditions are absorbing at $r = R_c$

$$c(R_c, \varphi, t) = 0 \quad (7)$$

and reactive at $r = \rho$ (radiation boundary condition) [3]

$$2\pi D \rho \left. \frac{\partial c}{\partial r} \right|_{\rho} = \kappa c(\rho, \varphi, t). \quad (8)$$

The initial condition is

$$c(r, \varphi, 0) = \frac{1}{\rho} \delta(r - \rho^+) \delta(\varphi). \quad (9)$$

After separation of variables, the solution can be expressed as

$$c(r, \varphi, t) = \sum_{m=0}^{\infty} \sum_{n=1}^{\infty} e^{-(q_{mn}/\rho)^2 D t} \cos(m\varphi) \frac{R_{mn}(r) R_{mn}(\rho)}{\pi(1 + \delta_{m0}) \int_{\rho}^{R_c} r R_{mn}^2(r) dr}. \quad (10)$$

Here, the radial eigenfunctions satisfying the boundary conditions are expressed in Bessel functions as

$$R_{mn}(r) = J_m(q_{mn} r/\rho) Y_m(q_{mn} R_c/\rho) - Y_m(q_{mn} r/\rho) J_m(q_{mn} R_c/\rho) \quad (11)$$

and q_{mn} is the n th root of

$$\rho \left. \frac{dR_{mn}}{dr} \right|_{\rho} - \alpha R_{mn}(\rho) = 0 \quad (12)$$

from the boundary condition at $r = \rho$. Here we have introduced the parameter $\alpha \equiv \kappa/2\pi D$.

The return flux at position φ at time t is given by

$$\begin{aligned} \Phi(\varphi, t) &= \frac{\kappa}{2\pi} c(\rho, \varphi, t) \\ &= \alpha D \sum_{m=0}^{\infty} \sum_{n=1}^{\infty} e^{-(q_{mn}/\rho)^2 D t} \cos(m\varphi) \frac{R_{mn}^2(\rho)}{\pi(1 + \delta_{m0}) \int_{\rho}^{R_c} r R_{mn}^2(r) dr}. \end{aligned} \quad (13)$$

The total return probability is

$$\Phi_0 = \int_0^{\infty} dt \int_{-\pi}^{\pi} d\varphi \Phi(\varphi, t) = \alpha \rho^2 \sum_{n=1}^{\infty} \frac{1}{q_{0n}^2} \frac{R_{0n}^2(\rho)}{\int_{\rho}^{R_c} r R_{0n}^2(r) dr}. \quad (14)$$

Numerically, this is the same as our previous result [14, 18]

$$\Phi_0 = \frac{\alpha \ln(R_c/\rho)}{1 + \alpha \ln(R_c/\rho)}. \quad (15)$$

The total return flux (given return) independently of the angle φ is

$$\psi(t) = \frac{1}{\Phi_0} \int_{-\pi}^{\pi} \Phi(\varphi, t) d\varphi = \frac{\alpha D}{\Phi_0} \sum_{n=1}^{\infty} e^{-(q_{0n}/\rho)^2 Dt} \frac{R_{0n}^2(\rho)}{\int_{\rho}^{R_c} r R_{0n}^2(r) dr}. \quad (16)$$

This is the probability density for the return-time distribution from which a stochastic return time τ_1 can be generated. The corresponding distribution function is

$$\int_0^t \psi(t) dt = 1 - \frac{\alpha \rho^2}{\Phi_0} \sum_{n=1}^{\infty} e^{-(q_{0n}/\rho)^2 Dt} \frac{R_{0n}^2(\rho)}{q_{0n}^2 \int_{\rho}^{R_c} r R_{0n}^2(r) dr}. \quad (17)$$

Given return at time τ_1 , the return angle can be generated from the probability density

$$\Psi(\varphi | \tau_1) = \frac{\Phi(\varphi, \tau_1)}{\psi(\tau_1) \Phi_0} = \frac{\alpha D}{\psi(\tau_1) \Phi_0} \sum_{m=0}^{\infty} \sum_{n=1}^{\infty} e^{-(q_{mn}/\rho)^2 D\tau_1} \frac{\cos(m\varphi)}{\pi(1 + \delta_{m0})} \frac{R_{mn}^2(\rho)}{\int_{\rho}^{R_c} r R_{mn}^2(r) dr} \quad (18)$$

or the corresponding distribution function

$$\int_{-\pi}^{\varphi} \Psi(\varphi | \tau_1) d\varphi = \frac{\pi + \varphi}{2\pi} + \frac{\alpha D}{\psi(\tau_1) \Phi_0} \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{\sin(m\varphi)}{m\pi} e^{-(q_{mn}/\rho)^2 D\tau_1} \frac{R_{mn}^2(\rho)}{\int_{\rho}^{R_c} r R_{mn}^2(r) dr}. \quad (19)$$

During the free diffusion, the protein will also move along the z -axis. This motion will be independent of the motion in the r - and φ -coordinates. Assuming that the protein starts at position $z = 0$, its diffusion is given by the probability density

$$v(z, t) = \frac{1}{\sqrt{4\pi Dt}} e^{-z^2/4Dt}. \quad (20)$$

Thus, the return flux to position φ, z at time t is

$$F(\varphi, z, t) = \Phi(\varphi, t) v(z, t). \quad (21)$$

The boundary condition at $r = \rho$ (equation (8)) can be reformulated through application of the recurrence relations for the Bessel functions:

$$q_{mn} [J_{m-1}(q_{mn}) Y_m(q_{mn} \eta) - Y_{m-1}(q_{mn}) J_m(q_{mn} \eta)] - (\alpha + m) [J_m(q_{mn}) Y_m(q_{mn} \eta) - Y_m(q_{mn}) J_m(q_{mn} \eta)] = 0, \quad (22)$$

where $\eta = R_c/\rho$. From this relation the roots q_{mn} can be generated without recourse to numerical derivation. Also the normalization integral can be reformulated in a similar manner:

$$\begin{aligned} \int_{\rho}^{R_c} r R_{mn}^2(r) dr &= \frac{\rho^2 R_{mn}^2(\rho)}{2q_{mn}^2} (m^2 - \alpha^2 - q_{mn}^2) \\ &+ \frac{\rho^2 R_{mn}(\rho)}{2} \eta [J_{m-1}(q_{mn}) Y_{m-1}(q_{mn} \eta) - Y_{m-1}(q_{mn}) J_{m-1}(q_{mn} \eta)] \\ &- \frac{\rho^2 R_{mn}(\rho)}{2} \frac{m + \alpha}{q_{mn}} \eta [J_m(q_{mn}) Y_{m-1}(q_{mn} \eta) - Y_m(q_{mn}) J_{m-1}(q_{mn} \eta)]. \end{aligned} \quad (23)$$

Protein rotations

Finally, the rotations of the protein must be accounted for. Define the protein orientation as the vector from the center of the sphere to the center of the reactive patch. Based on the properties of the spherical harmonics in a spherical coordinate system where the main axis is aligned with the protein patch vector at time $t = 0$, the probability density over the spherical polar angle θ at a later time t is

$$G(\theta, t) = \sin(\theta) \sum_{n=0}^{\infty} \frac{2n+1}{2} P_n(\cos(\theta)) e^{-n(n+1)D_R t}, \quad (24)$$

where P_n is a Legendre polynomial. The probability density over the azimuthal angle ϕ is uniform at all later times. The probability that $\theta < \theta_s$ after time t is given by the distribution function

$$\begin{aligned} P(\theta_s, t) &= \int_0^{\theta_s} G(\theta, t) d\theta \\ &= \frac{1}{2} [1 - \cos(\theta_s)] + \sum_{n=1}^{\infty} \frac{1}{2} [P_{n-1}(\cos(\theta_s)) - P_{n+1}(\cos(\theta_s))] e^{-n(n+1)D_R t}. \end{aligned} \quad (25)$$

At any of the stochastically generated return times, the spherical polar rotation angle can be generated from this relation. The corresponding azimuthal rotation ϕ_s is generated from a uniform distribution. These rotation angles are then appropriately added to the coordinates defined in relation to the DNA cylinder as described below.

Overview of the event-driven simulation scheme

The idea behind the event-driven simulation (see SI for pseudocode) is to characterize the protein–DNA association kinetics by sampling a large number of rebinding attempts after the protein and DNA has just microscopically dissociated from each other. The return time distribution (equation (17)) for the protein re-associating to DNA was derived by solving the translation-rotation coupled diffusion equation for a fully reactive sphere and the fully reactive cylinder (equation (6)) with boundary conditions (equations (7) and (8)) and initial condition (equation (9)). While sampling this return time, and while also keeping track of the position of the reactive patches on the protein and DNA, rebinding simulations with arbitrary reactive patch geometries can be performed. For each sampled return time there is a corresponding probability distribution for the cylindrical azimuthal angle (equation (19)), for diffusion of the protein along the length of the cylinder axis (equation (20)) and for rotational diffusion of the protein (equation (25)).

After a microscopic dissociation event that leaves the reactants right next to each other, diffusion will carry them apart (macroscopic dissociation) with probability $1-\Phi_0$ (equation (15)). If dissociation is not successful, a re-association occurs that could lead to rebinding if the orientations allow. If rebinding does not occur, the simulation continues from the new orientation angles until either a dissociation (probability $1-\Phi_0$) or a new re-association occurs where orientations can again be tested. After a rebinding event or a successful dissociation, the simulation restarts with a uniform distribution over both reactive patches. The initial uniform distribution over the reactive patches is a consequence of microscopic reversibility and assumes that the dissociation process starts from an equilibrated bound state. The fraction of all rebinding attempts (4×10^5 or 10^6 for all points in all figures) where the protein dissociates rather than rebinds gives the probability of dissociation, P_{diss} . This is the main simulation result but we can also get the nonspecific association rate constant, k_a , by

multiplying P_{diss} with the microscopic association rate constant k giving $k_a = 2\pi D\alpha\ell F P_{\text{diss}} = kP_{\text{diss}}$. In the diffusion-controlled limit, $\alpha \rightarrow \infty$, where simulations in this way are impossible, k_a can be calculated from the simulated P_{diss} for $\alpha < \infty$ as

$$k_a^{\text{diff}} = kP_{\text{diss}}/(1 - P_{\text{diss}}). \quad (26)$$

In practice, it is inconvenient to choose a large α in the simulations as that leads to a very large number of very short-lived micro-dissociations. Except where stated otherwise, in the calculation reported here $\alpha = 16$ has been used. It can be noted that this simulation scheme allows a consistent way of accounting for the inhomogeneous boundary conditions that does not require the approximation introduced by Shoup *et al* [7], which assumes that the total association flux is homogeneous over the reactive surfaces. In this simulation scheme, a protein that returns to the cylinder is allowed a ‘virtual binding’ using the reactive boundary condition, equation (8), regardless of orientation. It will be bound only if the steric constraints are satisfied. If not, it is immediately returned to the position just outside the reaction radius and allowed to continue its attempt at macroscopic dissociation. In effect, in spite of the reactive boundary condition employed also outside the reactive regions, no reaction occurs there and the protein is treated as reflected.

Updating the simulation coordinates

To be able to perform rebinding simulations for different sphere and cylinder geometries, it is important sample and update the position of the protein sphere relative to the DNA cylinder as well as the orientation of the protein reactive patch. Let φ_D and z_D be the cylindrical azimuthal angle and height coordinate defining the contact point between the DNA cylinder and the protein sphere. The change in these coordinates is sampled (equations (19) and (20)) at each rebinding attempt of the simulation and they can be updated by addition as

$$\varphi_D(\tau_1) = \varphi_D(0) + \varphi \quad (27)$$

and

$$z_D(\tau_1) = z_D(0) + z. \quad (28)$$

A suitable coordinate system is needed to keep track of the position of the reactive patch on the protein sphere. The z -axis of the coordinate system of the sphere is defined to be the same as the z -axis of the coordinate system of the cylinder, while the x -axis of the coordinate system of the sphere is defined to be in the direction going from the center of the sphere to the contact point between the sphere and the cylinder. The center of the reactive patch on the protein sphere then defines the spherical azimuthal and polar angles ϕ_p and θ_p . With this definition of the coordinate systems, an update of the spherical azimuthal angle is required each time the cylindrical azimuthal angle is updated. More specifically, ϕ_p and φ_D are additive with reverse signs, so that

$$\phi_p(\tau_1) = \phi_p(0) - \varphi \quad (29)$$

updates ϕ_p for each return of the protein to the DNA. The spherical angles defining the orientation of the protein reactive patch are also updated when the rotational diffusion of the sphere around its own axes is sampled. To simplify the simulations and minimize the number of different probability distributions needed to be calculated, the change in spherical rotation angles has been defined in a coordinate system where

$$\theta^*(0) = \phi^*(0) = 0. \quad (30)$$

In that coordinate system the spherical angles after a time τ_1 are ϕ and θ which can be sampled with a uniform distribution and equation (25) respectively. The spherical rotation angles $\phi^*(\tau_1)$ and $\theta^*(\tau_1)$ are then transformed to the regular coordinate system of the sphere, followed by the update of ϕ_p and θ_p by setting them to be equal to the transformed $\phi^*(\tau_1)$ and $\theta^*(\tau_1)$.

In practice the transformation of ϕ^* and θ^* is done by first calculating an orthonormal base giving a set of coordinate axes defined by equation (30), and then expressing (ϕ^*, θ^*) in Cartesian coordinates in this base. The Cartesian coordinates of (ϕ^*, θ^*) are then transformed to the regular coordinate system of the sphere by summation of the three products of the base vectors and their corresponding Cartesian coordinate. These new Cartesian coordinates are converted to spherical coordinates, which gives the updated ϕ_p and θ_p .

The simulation coordinates φ_D , z_D , ϕ_p , and θ_p are periodic, and they are after each update converted to the periods

$$\begin{aligned} -\pi &< \varphi_D < \pi \\ -\frac{h}{2} &< z_D < \frac{h}{2} \\ -\pi &< \phi_p < \pi \\ 0 &< \theta_p < \pi. \end{aligned} \quad (31)$$

Reactive boundary conditions of different geometries

For each rebinding attempt of the simulations the reactivity of the DNA cylinder and protein sphere is tested to determine if binding occurs. For the cylinder the reactive patch geometries used are ring, stripe and helix. For the sphere the reactive patch is always axially symmetric, i.e. it is contained in a circle on the surface of the sphere (see figure 1).

For a cylinder with periodic ring-reactive patches, the cylinder is reactive at (r, φ_D, z_D) if and only if

$$\begin{cases} r = \rho, \\ -\frac{hF}{2} < z_D < \frac{hF}{2}. \end{cases} \quad (32)$$

For cylinder with a stripe-reactive patch, the cylinder is reactive at (r, φ_D, z_D) if and only if

$$\begin{cases} r = \rho, \\ -\pi F < \varphi_D < \pi F. \end{cases} \quad (33)$$

For a cylinder with a helix-reactive patch, the cylindrical azimuthal angle corresponding to the center of the reactive path is given by

$$\varphi_{\text{patch}} = 2\pi \frac{z_D}{h}. \quad (34)$$

The cylinder is then reactive at (r, φ_D, z_D) if $\varphi_{\text{patch}} - \pi F > -\pi \wedge \varphi_{\text{patch}} + \pi F < \pi$ while it also holds that

$$\begin{cases} r = \rho, \\ \varphi_{\text{patch}} - \pi F < \varphi_D < \varphi_{\text{patch}} + \pi F. \end{cases} \quad (35)$$

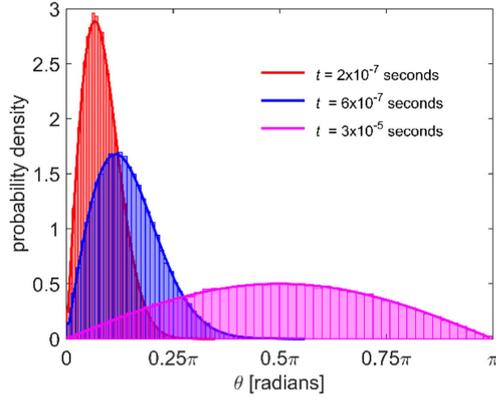


Figure 9. The probability density of different polar angles θ for a sphere which has undergone rotational diffusion for three different times with $D_R = 0.75D/R_p^2$. The distribution function in equation (25) has been calculated with 300 terms in the sum for 50 evenly spaced θ before being converted to probability densities (lines). The same probability density functions have also been estimated with Brownian dynamics simulations (normalized histograms). For every time step $\Delta t = t/1000$ in the simulations, an axis of rotation is sampled uniformly over the entire sphere surface. A rotation given by the root of the mean squared angular displacement (MSAD) $3 \cdot 2D_R\Delta t$ is then performed around this axis. After 1000 time steps θ is saved. 10^5 such samples of θ are included in each histogram.

or if $\varphi_{\text{patch}} - \pi F \leq -\pi$ while it also holds that

$$\begin{cases} r = \rho, \\ \varphi_D > \varphi_{\text{patch}} - \pi F + 2\pi \vee \varphi_D < \varphi_{\text{patch}} + \pi F. \end{cases} \quad (36)$$

or if $\varphi_{\text{patch}} + \pi F \geq \pi$ while it also holds that

$$\begin{cases} r = \rho, \\ \varphi_D > \varphi_{\text{patch}} - \pi F \vee \varphi_D < \varphi_{\text{patch}} + \pi F - 2\pi. \end{cases} \quad (37)$$

For a sphere with a symmetric reactive patch, the relationship between the area of the reactive patch and the fraction of the sphere that is reactive gives the equation

$$F_p = \frac{1}{4\pi} \int_{\phi=-\pi}^{\phi=\pi} \int_{\theta=0}^{\theta=\theta_{\text{max}}} \sin \theta d\theta d\phi, \quad (38)$$

Where θ_{max} is the polar angle between the center and border of the reactive patch. Evaluation of the integral in equation (38) gives

$$\theta_{\text{max}} = \cos^{-1}(1 - 2F_p). \quad (39)$$

If θ_{diff} is the angle between the center of the reactive patch (given by ϕ_p and θ_p) and the contact point of the sphere and cylinder (pointing along the positive x -axis), the sphere is reactive at (ϕ_p, θ_p) if and only if

$$\theta_{\text{diff}}(\phi_p, \theta_p) < \theta_{\text{max}}(F_p). \quad (40)$$

Discretizing the probability distributions

In order to calculate the return time distribution (equation (17)) and the cylindrical azimuthal angle distribution (equation (19)) we need to find the roots q_{mn} (equations (11) and (12)). This was done by making use of the recurrence relations in equation (22) and using brute force simple bisection with seeds every 0.1 step until 100 unique roots were found. Given the roots we use the recurrence relation to evaluate the integral in equation (23) before generating the probability distribution in the azimuthal angle φ for each rebinding time. The azimuthal angle φ ranges between $-\pi$ and π in steps of $\pi/360$ radians. For $R_c = 14$ nm the rebinding time ranges between 0.5 ns and 30 μ s in 600 unevenly spaced time steps. We can compare the simulation results with analytical solutions for a fully reactive protein in the case of stripe and ring geometries on the cylinder (equations (S11)–(S13)). The simulations compare well with the analytical solutions also in the diffusion-controlled limit (figure 5(a)).

For the protein rotations, the polar angle distribution (equation (25)) is in a similar way discretized for all 600 return times. For each return time, the polar angle distribution is discretized in 50 evenly spaced steps in θ so that the cumulative probability for the largest θ is smaller than $1-10^{-5}$. The analytically derived polar angle distributions (equations (24) and (25)) agrees very well with Brownian dynamics simulations sampling the same rotational diffusion process (figure 9).

Acknowledgments

This work was supported by the European Research Council and the Knut and Alice Wallenberg Foundation. AM was funded by the Centre for Interdisciplinary Mathematics, Uppsala University.

Author contributions

JE and OB designed the study, OB did the analytical calculations, AM & EM wrote simulation codes, AM, EM and OB analyzed the data, JE, EMOB and AM wrote the manuscript.

Competing financial interests

The authors declare no competing financial interests.

References

- [1] Zimmerman S B and Trach S O 1991 Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli* *J. Mol. Biol.* **222** 599–620
- [2] Smoluchowski M 1917 Versuch einer mathematischen Theorie der Koagulationskinetik kolloider Lösungen *Z. Phys. Chem.* **92** 129–68
- [3] Collins F C and Kimball G E 1949 Diffusion-controlled reaction rates *J. Coll. Sci. Imp. U. Tok.* **4** 425–37
- [4] Solc K and Stockmayer W H 1971 Kinetics of diffusion-controlled reaction between chemically asymmetric molecules: I. General theory *J. Chem. Phys.* **54** 2981–8
- [5] Schmitz K S and Schurr J M 1972 Role of orientation constraints and rotational diffusion in bimolecular solution kinetics *J. Phys. Chem.* **76** 534–45
- [6] Berg O G 1978 On diffusion-controlled dissociation *Chem. Phys.* **31** 47–57
- [7] Shoup D, Lipari G and Szabo A 1981 Diffusion-controlled bimolecular reaction-rates—the effect of rotational diffusion and orientation constraints *Biophys. J.* **36** 697–714
- [8] Zhou H X 1993 Brownian dynamics study of the influences of electrostatic interaction and diffusion on protein–protein association kinetics *Biophys. J.* **64** 1711–26
- [9] Schlosshauer M and Baker D 2002 A general expression for bimolecular association rates with orientational constraints *J. Phys. Chem. B* **106** 12079–83

- [10] Schlosshauer M and Baker D 2004 Realistic protein–protein association rates from a simple diffusional model neglecting long-range interactions, free energy barriers, and landscape ruggedness *Protein Sci.* **13** 1660–9
- [11] Zhou H X, Rivas G N and Minton A P 2008 Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences *Annu. Rev. Biophys.* **37** 375–97
- [12] Berg O G 1985 Orientation constraints in diffusion-limited macromolecular association, the role of surface diffusion as a rate-enhancing mechanism *Biophys. J.* **47** 1–14
- [13] Berg O G and Blomberg C 1976 Association kinetics with coupled diffusional flows—special application to lac repressor-operator system *Biophys. Chem.* **4** 367–81
- [14] Berg O G, Winter R B and von Hippel P H 1981 Diffusion-driven mechanisms of protein translocation on nucleic-acids: I. Models and theory *Biochemistry* **20** 6929–48
- [15] Watson J D and Crick F H 1953 A structure for deoxyribose nucleic acid *Nature* **171** 737–8
- [16] Meselson M and Stahl F W 1958 The replication of DNA in *Escherichia coli* *Proc. Natl Acad. Sci. USA* **44** 671–82
- [17] Marklund E G, Mahmutovic A, Berg O G, Hammar P, van der Spoel D, Fange D and Elf J 2013 Transcription-factor binding and sliding on DNA studied using micro- and macroscopic models *Proc. Natl Acad. Sci. USA* **110** 19796–801
- [18] Mahmutovic A, Berg O G and Elf J 2015 What matters for lac repressor search *in vivo*-sliding, hopping, intersegment transfer, crowding on DNA or recognition? *Nucleic Acids Res.* **43** 3454–64
- [19] McGuffee S R and Elcock A H 2010 Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm *PLoS Comput. Biol.* **6** e1000694
- [20] Dillon S C and Dorman C J 2010 Bacterial nucleoid-associated proteins, nucleoid structure and gene expression *Nat. Rev. Microbiol.* **8** 185–95
- [21] Kornyshev A A, Lee D J, Leikin S and Wynveen A 2007 Structure and interactions of biological helices *Rev. Mod. Phys.* **79** 943–96
- [22] Cherstvy A G and Winkler R G 2004 Complexation of semiflexible chains with oppositely charged cylinder *J. Chem. Phys.* **120** 9394–400
- [23] Cherstvy A G 2009 Positively charged residues in DNA-binding domains of structural proteins follow sequence-specific positions of DNA phosphate groups *J. Phys. Chem. B* **113** 4242–7
- [24] Mirny L, Slutsky M, Wunderlich Z, Tafvizi A, Leith J and Kosmrlj A 2009 How a protein searches for its site on DNA: the mechanism of facilitated diffusion *J. Phys. A: Math. Theor.* **42** 434013
- [25] Bauer M and Metzler R 2013 *In vivo* facilitated diffusion model *PLoS One* **8** e53956
- [26] Bauer M, Rasmussen E S, Lomholt M A and Metzler R 2015 Real sequence effects on the search dynamics of transcription factors on DNA *Sci. Rep.* **5** 10072
- [27] van den Broek B, Lomholt M A, Kalisch S M, Metzler R and Wuite G J 2008 How DNA coiling enhances target localization by proteins *Proc. Natl Acad. Sci. USA* **105** 15738–42
- [28] Hammar P, Leroy P, Mahmutovic A, Marklund E G, Berg O G and Elf J 2012 The lac repressor displays facilitated diffusion in living cells *Science* **336** 1595–8
- [29] Wang Y M, Austin R H and Cox E C 2006 Single molecule measurements of repressor protein 1D diffusion on DNA *Phys. Rev. Lett.* **97** 048302
- [30] Elf J, Li G W and Xie X S 2007 Probing transcription factor dynamics at the single-molecule level in a living cell *Science* **316** 1191–4
- [31] Pant K, Karpel R L, Rouzina I and Williams M C 2004 Mechanical measurement of single-molecule binding rates: kinetics of DNA helix-destabilization by T4 gene 32 protein *J. Mol. Biol.* **336** 851–70
- [32] Bonnet I, Biebricher A, Porté P L, Loverdo C, Bénichou O, Voituriez R, Escudé C, Wende W, Pingoud A and Desbiolles P 2008 Sliding and jumping of single EcoRV restriction enzymes on non-cognate DNA *Nucleic Acids Res.* **36** 4118–27
- [33] Sakata-Sogawa K and Shimamoto N 2004 RNA polymerase can track a DNA groove during promoter search *Proc. Natl Acad. Sci. USA* **101** 14731–5
- [34] Schurr J M 1975 The one-dimensional diffusion coefficient of proteins absorbed on DNA, hydrodynamic considerations *Biophys. Chem.* **9** 413–4
- [35] Blainey P C, Luo G, Kou S C, Mangel W F, Verdine G L, Bagchi B and Xie X S 2009 Nonspecifically bound proteins spin while diffusing along DNA *Nat. Struct. Mol. Biol.* **16** 1224–9
- [36] Li G W, Berg O G and Elf J 2009 Effects of macromolecular crowding and DNA looping on gene regulation kinetics *Nat. Phys.* **5** 294–7