

# Facilitated diffusion with DNA coiling

Michael A. Lomholt<sup>a</sup>, Bram van den Broek<sup>b,c</sup>, Svenja-Marei J. Kalisch<sup>b</sup>, Gijs J. L. Wuite<sup>b</sup>, and Ralf Metzler<sup>d,1</sup>

<sup>a</sup>MEMPHYS Center for Biomembrane Physics, Department of Physics and Chemistry, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark; <sup>b</sup>Department of Physics and Astronomy, Faculty of Sciences, Vrije University, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands; <sup>c</sup>Physics of Life Processes, Leiden Institute of Physics, Leiden University, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands; and <sup>d</sup>Physics Department, Technical University of Munich, D-85747 Garching, Germany

Communicated by Joshua Jortner, Tel Aviv University, Tel Aviv, Israel, March 27, 2009 (received for review January 26, 2009)

**When DNA-binding proteins search for their specific binding site on a DNA molecule they alternate between linear 1-dimensional diffusion along the DNA molecule, mediated by nonspecific binding, and 3-dimensional volume excursion events between successive dissociation from and rebinding to DNA. If the DNA molecule is kept in a straight configuration, for instance, by optical tweezers, these 3-dimensional excursions may be divided into long volume excursions and short hops along the DNA. These short hops correspond to immediate rebindings after dissociation such that a rebinding event to the DNA occurs at a site that is close to the site of the preceding dissociation. When the DNA molecule is allowed to coil up, immediate rebinding may also lead to so-called intersegmental jumps, i.e., immediate rebindings to a DNA segment that is far away from the unbinding site when measured in the chemical distance along the DNA, but close by in the embedding 3-dimensional space. This effect is made possible by DNA looping. The significance of intersegmental jumps was recently demonstrated in a single DNA optical tweezers setup. Here we present a theoretical approach in which we explicitly take the effect of DNA coiling into account. By including the spatial correlations of the short hops we demonstrate how the facilitated diffusion model can be extended to account for intersegmental jumping at varying DNA densities. It is also shown that our approach provides a quantitative interpretation of the experimentally measured enhancement of the target location by DNA-binding proteins.**

gene regulation | search processes | random processes | intersegmental jumps | single-molecule biophysics

At any given instant of time a biological cell only uses part of its genes for production of other molecules (RNA, proteins). During the development of the cell, or the organism the cell belongs to, different genes may be turned on. A classical example is the lactose metabolism of bacteria: in absence of lactose the Lac repressor is specifically bound at the lacZ gene and prevents unnecessary production of the lactose enzyme used to digest the milk sugar; when only lactose and no glucose is present the lacZ gene is activated and lactose-digesting enzymes are produced. Equally famed is the  $\lambda$  switch in *Escherichia coli* bacteria that are infected by bacteriophage  $\lambda$ . There, the  $\lambda$  switch decides between the dormant lysogenic state or the state of lysis that leads to the production of new phages and ultimate death of the *E. coli* host cell. Molecularly, the regulation of a gene relies on the presence of certain DNA-binding proteins, so-called transcription factors, that bind to a specific binding site close to the starting sequence of the related gene. The transcription factor then promotes or inhibits binding of RNA polymerase and thus regulates the transcription of this gene. This strategy allows eukaryotic and prokaryotic cells as well as viruses to respond to different internal or external signals (1, 2).

The binding of the protein to its specific binding site (cognate site, “target”) is necessarily preceded by a search process of the protein for this target site, this search being of a purely stochastic nature. Yet, some active binding proteins can locate their target orders of magnitude faster (3) than estimated for pure 3-dimensional diffusion limited search according to the celebrated Smoluchowski result

$$k_{\text{on}} = 4\pi D_{3\text{d}} b \quad [1]$$

for the search rate  $k_{\text{on}}$ . Here, the target is assumed spherical and of radius  $b$ , whereas  $D_{3\text{d}}$  denotes the 3-dimensional diffusion coefficient  $D_{3\text{d}}$  (4). An explanation for this discrepancy is based on the linear topology of the DNA chain in which the target site is embedded. In fact, in their seminal work Berg and von Hippel analyzed the search by DNA-binding proteins and demonstrated that the combination of different search mechanisms explains the observed binding rates. According to the Berg–von Hippel-facilitated diffusion model (5) DNA-binding proteins combine 3-dimensional diffusion in the bulk and 1-dimensional diffusion along the DNA chain (6–9). This linear diffusion is mediated by nonspecific binding, the protein’s binding affinity to noncognate DNA made possible by the heteropolymeric nature of DNA (10–12). An additional search mechanism already proposed in the works of Berg and von Hippel are intersegmental transfers: some proteins with two DNA-binding sites can intermittently bind nonspecifically to 2 DNA-segments that are far apart along the DNA contour but are brought in close contact in the embedding (real) space by DNA looping. Such proteins thereby transfer across a point of close contact without actually leaving the DNA (5).

Facilitated diffusion has recently received renewed interest, following a variety of new conceptual viewpoints of gene regulation (13–18). Fresh impetus has also been given by the possibility to monitor the target search by DNA-binding proteins on the level of single DNA molecules. Various aspects of the search mechanisms of DNA-binding proteins have been studied either by fluorescence methods or indirectly by probing the response to external manipulation by optical tweezers (6–8, 19, 20). In particular, it was demonstrated that the local DNA density indeed significantly influences the target search rate (20). This observation is consistent with the Berg–von Hippel idea of intersegmental transfers mentioned above. However, in contrast to the Lac repressor protein employed in the original work by Berg and von Hippel, in the optical tweezers assay reported in ref. 20 the target search was performed by EcoRV restriction enzymes that only have one DNA-binding site. The idea of intersegmental transfers therefore has to be extended, such that the protein dissociates from one DNA segment, shortly diffuses three-dimensionally, and then immediately rebinds to another DNA segment that is close by because of DNA looping. We refer to this extended mechanism as intersegmental jumps. In what follows we develop a theoretical framework to include intersegmental jumps. This extension of the facilitated diffusion model allows us to study the influence of the local DNA density on the rate of search of DNA-binding proteins for their specific binding site. Our generalized model therefore accounts for different degrees of DNA coiling. We show that this systematic extension of the Berg–von Hippel model is quantitative and consistent with the results from our experimental study in ref. 20.

Author contributions: M.A.L., B.v.d.B., S.-M.J.K., G.J.L.W., and R.M. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

<sup>1</sup>To whom correspondence should be addressed. E-mail: metz@ph.tum.de.

## Model

Previous approaches to facilitated protein diffusion consider bulk exchange for a straight cylindrical chain (21); they invoke scaling properties in different chain configurations (22) without quantitative comparison of the search rates between these; or they reformulate the problem with uncorrelated bulk excursions (23–26). Here, we relate a multiparticle picture to single protein jumps along the DNA contour allowing us to consider DNA as a fluctuating chain.

In our description of the target search process, we use the density per length  $n(x, t)$  of proteins on the DNA as the relevant dynamical quantity ( $x$  is the distance along the DNA contour). We include 1-dimensional diffusion along the DNA with diffusivity  $D_{1d}$ , protein dissociation with rate  $k_{\text{off}}^{\text{ns}}$  and (re)adsorption after diffusion through the bulk fluid with diffusion constant  $D_{3d}$ . The dynamics of  $n(x, t)$  is thus governed by

$$\frac{\partial n(x, t)}{\partial t} = \left( D_{1d} \frac{\partial^2}{\partial x^2} - k_{\text{off}}^{\text{ns}} \right) n(x, t) - j(t) \delta(x) + G(x, t) + k_{\text{off}}^{\text{ns}} \int_{-\infty}^{\infty} dx' \int_0^t dt' W_{\text{bulk}}(x - x', t - t') n(x', t'). \quad [2]$$

Here,  $j(t)$  is the flux into the target site located at  $x = 0$ ,  $W_{\text{bulk}}(x - x', t - t')$  is the joint probability that a protein returns to the point  $x$  at time  $t$  after leaving the DNA at  $x', t'$  for a bulk excursion.  $G(x, t)$  is the flux onto the DNA of proteins that have not previously been bound to the DNA (virgin flux).  $j(t)$  is determined by imposing the boundary condition  $n(x = 0, t) = 0$  at the target. To consider the transfer kernel  $W_{\text{bulk}}$  as homogeneous in space and time we made 2 assumptions that we found to be valid for the system studied in ref. 20: First, that the DNA can be treated as being infinitely long when considering the search rate. In practice this requires that the DNA should be longer than the effective sliding length  $l_{\text{sl}}^{\text{eff}}$  introduced in Eq. 14. Note that even in bacteria the DNA length measures several millimeters, whereas DNA's persistence length is of the order of 50 nm. The second assumption is that the coiled DNA conformation fluctuates sufficiently quickly such that subsequent excursion distances  $x - x'$  can be treated as independent.

To proceed, we Laplace and Fourier transform Eq. 2:

$$\frac{n(q, u)}{W(q, u)} - n_0(q) = G(q, u) - j(u) \quad [3]$$

where we denote the Fourier and Laplace transform of a function by explicit dependence on the transformed variable. Thus  $n(q, u) = \mathcal{L}\{n(x, t)\}$ ,  $n(q, t) = \mathcal{F}\{n(x, t)\}$  and  $n_0(x) = n(x, t = 0)$ . The Fourier–Laplace transform of  $W(x, t)$  is here defined through

$$W(q, u) = \frac{1}{u + D_{1d}q^2 + k_{\text{off}}^{\text{ns}}[1 - W_{\text{bulk}}(q, u)]}. \quad [4]$$

We rewrite these equations in the form

$$n(q, u) = n^{\text{ns}}(q, u) - j(u)W(q, u) \quad [5]$$

where the density

$$n^{\text{ns}}(q, u) = W(q, u)[G(q, u) + n_0(q)] \quad [6]$$

is the solution of the diffusion problem in absence of the target, i.e., when only nonspecific binding between proteins and the DNA occurs. In position space at  $x = 0$  one obtains  $j(u) = n^{\text{ns}}(x = 0, u)/W(x = 0, u)$ .

In this study we focus on the long time behavior of  $j(t)$  for which the density  $n^{\text{ns}}(x, t)$  will reach an equilibrium value  $n_{\text{eq}}^{\text{ns}}$ . Since, by Tauberian theorems, large  $t$  correspond to small  $u$  we study the  $u \rightarrow 0$  behavior of  $j(u)$ , with  $n^{\text{ns}}(x, u) \sim n_{\text{eq}}^{\text{ns}}/u$ . For  $W_0(x = 0, u)$

we assume that the limit  $W(x = 0, u = 0)$  (time average) is finite and nonzero; this is true when the distance distribution of bulk excursions is sufficiently long tailed, as fulfilled for the transfer kernel  $W_{\text{bulk}}$  encountered below. In practice, being interested in the *first arrival* at the target, we assume that the protein concentration is sufficiently dilute such that this arrival happens after reaching  $n_{\text{eq}}^{\text{ns}}$  (“rapid equilibrium”) but still with many searching proteins. Thus  $j(u) \sim k_{1d} n_{\text{eq}}^{\text{ns}}/u$ , and therefore at long times the stationary current  $j_{\text{stat}} \sim k_{1d} n_{\text{eq}}^{\text{ns}}$  into the target is reached. Here, we introduce the constant  $k_{1d}^{-1} = W(x = 0, u = 0)$ :

$$k_{1d}^{-1} = \int_{-\infty}^{\infty} \frac{dq}{2\pi} \frac{1}{D_{1d}q^2 + k_{\text{off}}^{\text{ns}}(1 - \lambda_{\text{bulk}}(q))}, \quad [7]$$

of dimension sec over cm. We also recognize that  $\lambda_{\text{bulk}}(x) = W_{\text{bulk}}(x, u = 0)$  is the distribution of relocation lengths along the DNA after a single 3-dimensional excursion.

We express our results in terms of the association rate

$$k_{\text{on}} = j_{\text{stat}}/n_{\text{bulk}} = k_{1d} K_{\text{ns}} \quad [8]$$

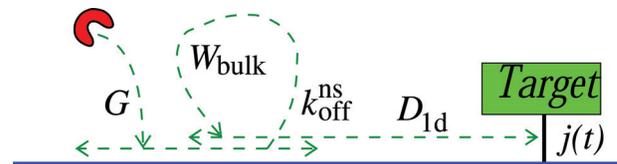
to the target where  $n_{\text{bulk}}$  is the density of unbound proteins in the bulk (at nonspecific equilibrium). The nonspecific binding constant per length of DNA is  $K_{\text{ns}} = n_{\text{eq}}^{\text{ns}}/n_{\text{bulk}}$ . In terms of the total volume concentration of proteins  $n_{\text{total}} = n_{\text{bulk}} + l_{\text{DNA}}^{\text{total}} n_{\text{eq}}^{\text{ns}}$ , where  $l_{\text{DNA}}^{\text{total}}$  is the total length of DNA divided by the volume of the entire system, we have

$$j_{\text{stat}} \sim \frac{k_{\text{on}} n_{\text{total}}}{1 + K_{\text{ns}} l_{\text{DNA}}^{\text{total}}}. \quad [9]$$

The rate  $k_{\text{on}}$  is related to the mean first arrival time for the steady-state association. Namely,  $k_{\text{on}} n_{\text{bulk}}$  is the probability per time for protein association with the target. Thus, the (survival) probability that no protein has arrived at the target yet is  $P_{\text{surv}}(t) = \exp(-k_{\text{on}} n_{\text{bulk}} t)$ , and the average target search time is  $T = 1/(k_{\text{on}} n_{\text{bulk}})$ . The flux  $j(t)$  being linear in the protein concentration  $n_{\text{total}}$  implies that for sufficiently low  $n_{\text{total}}$  the steady state is indeed reached well before the first protein actually binds to the specific target (note that in vivo protein concentrations can be as low as nanomolar).

## Results

**Straight Rod Configuration.** Consider first stretched DNA (see Fig. 1) corresponding to a cylinder of radius  $r_{\text{int}}$  (the range of nonspecific interaction), with nonspecific reaction rate  $k_{\text{on}}^{\text{ns}}$  at the boundary: the flux of proteins, per length of DNA, onto the DNA is  $k_{\text{on}}^{\text{ns}}$  times the bulk concentration next to the DNA. This implies that at equilibrium of nonspecific binding,  $n_{\text{eq}}^{\text{ns}} k_{\text{off}}^{\text{ns}} = n_{\text{bulk}} k_{\text{on}}^{\text{ns}}$  such that  $K_{\text{ns}} = k_{\text{on}}^{\text{ns}}/k_{\text{off}}^{\text{ns}}$ . (We deviate here from the conventions of ref. 20 by only letting  $n_{\text{bulk}}$  count “active enzymes.”)



**Fig. 1.** Schematic of the search mechanisms in Eq. 2 on a piece of stretched DNA. To locate the specific binding site (target) on the DNA molecule a DNA-binding protein diffuses one-dimensionally along the DNA chain with diffusivity  $D_{1d}$ , or it unbinds and performs a bulk excursion with diffusion constant  $D_{3d}$ . The virgin flux  $G$  denotes proteins that have not previously bound to DNA. Note the tendency to revisit sites when immediate rebindings and thus short hops are likely.

To find the distribution  $W_{\text{bulk}}(x, t)$  for return to the DNA of a protein released at the DNA at  $x = 0$  and  $t = 0$ , we solve the cylindrical diffusion equation for the probability density  $P(x, r, t)$  of the protein's position

$$\frac{\partial P}{\partial t} = D_{3d} \left( \frac{\partial^2}{\partial x^2} + \frac{1}{r} \frac{\partial}{\partial r} r \frac{\partial}{\partial r} \right) P \quad [10]$$

where  $r$  measures the distance to the cylinder axis. The reactive boundary condition for the probability flux out of the cylinder boundary is

$$-2\pi r_{\text{int}} D_{3d} \left. \frac{\partial P}{\partial r} \right|_{r=r_{\text{int}}} = -k_{\text{on}}^{\text{ns}} P|_{r=r_{\text{int}}} + \delta(t)\delta(x). \quad [11]$$

The initial distribution outside the cylinder is  $P(r, t)|_{t=0} = 0$ . From the solution of this problem we obtain  $W_{\text{bulk}}$  as  $k_{\text{on}}^{\text{ns}} P|_{r=r_{\text{int}}}$  in terms of modified Bessel functions  $K_\nu$ ,

$$W_{\text{bulk}}^{\text{cyl}}(q, u) = \left( 1 + \frac{2\pi D_{3d} \bar{q} r_{\text{int}} K_1(\bar{q} r_{\text{int}})}{k_{\text{on}}^{\text{ns}} K_0(\bar{q} r_{\text{int}})} \right)^{-1}, \quad [12]$$

where we abbreviate  $\bar{q} = \sqrt{q^2 + u/D_{3d}}$ .

With these ingredients one can evaluate numerically the search rate to reach the specific target on a straight DNA, as given by Eq. 7. However, some limits allow for an analytic approximation for the integral in Eq. 7: for  $k_{\text{on}}^{\text{ns}} \ll D_{3d}$  one may take  $1 - \lambda_{\text{bulk}}(q) \approx 1$ , and thus  $k_{\text{id}}^{-1} \sim (2\sqrt{D_{1d} k_{\text{off}}^{\text{ns}}})^{-1}$ . One finds the association rate

$$k_{\text{on}} = 2l_{\text{sl}} k_{\text{on}}^{\text{ns}}, \quad [13]$$

where  $l_{\text{sl}} = \sqrt{D_{1d}/k_{\text{off}}^{\text{ns}}}$  is the so-called sliding length of a protein during a single nonspecific binding period. Eq. 13 states that each time a protein, with rate constant  $k_{\text{on}}^{\text{ns}}$ , binds to the DNA within a distance  $l_{\text{sl}}$  from the target it is able to "slide" onto the target. Note that in this  $k_{\text{on}}^{\text{ns}} \ll D_{3d}$  limit the search rate  $k_{\text{on}}$  is independent of the DNA conformation. When  $k_{\text{on}}^{\text{ns}} \ll D_{3d}$  the rate of nonspecific binding will limit rebinding such that the protein will relocate by a long distance during a single 3-dimensional excursion. The part of  $\lambda_{\text{bulk}}$  corresponding to such long relocations does not contribute significantly to the integral in Eq. 7, which is the physical reason for the approximation leading to Eq. 13.

In the opposite limit  $k_{\text{on}}^{\text{ns}} \gg D_{3d}$  one may approximate  $1 - \lambda_{\text{bulk}}(q) \approx 2\pi D_{3d} / [k_{\text{on}}^{\text{ns}} / \ln(l_{\text{sl}}^{\text{eff}}/r_{\text{int}})]$  where the length

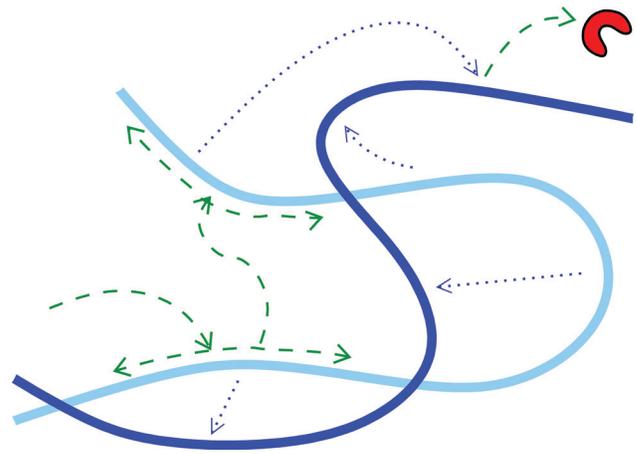
$$l_{\text{sl}}^{\text{eff}} = \sqrt{\frac{k_{\text{on}}^{\text{ns}}}{(2\pi D_{3d})}} l_{\text{sl}}. \quad [14]$$

This produces the result

$$k_{\text{on}} \sim 4\pi D_{3d} l_{\text{sl}}^{\text{eff}} / [\ln(l_{\text{sl}}^{\text{eff}}/r_{\text{int}})]^{1/2}. \quad [15]$$

In this limit the protein can rebind quickly after unbinding. This leads to oversampling and a lowered value of  $k_{\text{on}}$  compared to Eq. 13.  $l_{\text{sl}}^{\text{eff}}$  can be interpreted as an effective sliding length including immediate rebindings (21). Expression 15 generalizes Smoluchowski's result Eq. 1, replacing the naked target size  $b$  by the "antenna length" entering Eq. 15. Results 7 and 12 are identical to the classical result for straight, infinitely long DNA (21). However, the approach here allows us to explicitly consider coiled DNA.

**Coil with Persistence Length.** To obtain the transfer kernel  $W_{\text{bulk}}$  for a random coil DNA configuration (see Fig. 2) we treat the DNA as consisting locally of straight segments. We assume that the local density  $l_{\text{DNA}}$  of DNA (length per volume) around a segment is dilute such that the typical distance  $1/\sqrt{l_{\text{DNA}}}$  between segments



**Fig. 2.** Search rate enhancement on coiled DNA: a protein that in the case of straight DNA would perform a short hop can now, because of DNA looping, be captured by a foreign (remote) segment of DNA. This leads to a more efficient search since local oversampling is reduced by these intersegmental jumps. The DNA configuration is assumed to be dynamic such that when the protein is released from the foreign segment the latter has moved further away from the original segment. Dashed lines, motion of the protein; dotted lines, DNA configurational changes.

is much larger than  $r_{\text{int}}$ . We then apply a superposition technique, where the probability that the protein has reacted with a segment is obtained by treating independently the interaction with each segment, and then choose among these the interaction for which binding happens first. To obtain the dynamics of the interaction of the protein with one segment of DNA we consider a large volume  $V$  around the segment of length  $L$ . Placing the protein at a uniformly random position in space relative to the DNA segment at  $t = 0$  and denoting the probability that the protein has bound to this DNA before time  $t$  by  $J_{\text{single}}(t)$ , then the probability that it has not bound to any of  $N$  pieces of DNA is (with the assumption of independent interactions)  $P_{\text{surv}}^{\text{foreign}}(t) = [1 - J_{\text{single}}(t)]^N$ . Taking the limit of  $V, N, L \rightarrow \infty$  for fixed  $l_{\text{DNA}} = NL/V$  we get  $P_{\text{surv}}^{\text{foreign}}(t) = \exp[-J_{\text{cap}}(t)]$ , where

$$\frac{u^2 J_{\text{cap}}(u)}{k_{\text{on}}^{\text{ns}} l_{\text{DNA}}} = \left( 1 + \frac{k_{\text{on}}^{\text{ns}} K_0(\sqrt{u/D_{3d}} r_{\text{int}})}{2\pi \sqrt{u D_{3d}} r_{\text{int}} K_1(\sqrt{u/D_{3d}} r_{\text{int}})} \right)^{-1}. \quad [16]$$

The bulk transfer kernel  $W_{\text{bulk}}$  in the coiled conformation then becomes

$$W_{\text{bulk}}(q, t) = W_{\text{bulk}}^{\text{cyl}}(q, t) P_{\text{surv}}^{\text{foreign}}(t) - \frac{dP_{\text{surv}}^{\text{foreign}}(t)}{dt} \delta_{q,0} \int_t^\infty W_{\text{bulk}}^{\text{cyl}}(q=0, t') dt'. \quad [17]$$

Here, the first term on the right-hand side represents proteins that return to the original segment without being captured by foreign segments. Those are represented by the second term, where the factor  $\delta_{q,0}$  is a result of the assumption that a capture by foreign DNA segments leads to a long relocation measured along the DNA contour.\* These relocations only contribute close to  $q = 0$  in Fourier space. Exploiting the reasoning behind the approximation that led to Eq. 13 we discard them from  $W_{\text{bulk}}$  when the integral in Eq. 7 is calculated. However, even numerically this integral is difficult to handle. A good approximation can be found noting

\*Cases where this assumption is justified are given by systems in which (i) the persistence length of the DNA is large such that the chemical distance along DNA loops will necessarily be larger than the effective sliding length  $l_{\text{sl}}^{\text{eff}}$ , or (ii) the DNA attains a globular configuration, for instance, through mutual attraction bringing many chemically distant DNA segments together.

that by Tauberian theorems the Laplace inversion for  $J_{\text{cap}}(u)$  can be divided into 3 regimes when  $k_{\text{on}}^{\text{ns}} \gg D_{3\text{d}}$ :

$$J_{\text{cap}}(t) = \begin{cases} k_{\text{on}}^{\text{ns}} l_{\text{DNA}} t & , t \ll \frac{D_{3\text{d}}^2 r_{\text{int}}^2}{(k_{\text{on}}^{\text{ns}})^2} \\ 4l_{\text{DNA}} r_{\text{int}} \sqrt{\pi D_{3\text{d}} t} & , \frac{D_{3\text{d}}^2 r_{\text{int}}^2}{(k_{\text{on}}^{\text{ns}})^2} \ll t \ll \frac{r_{\text{int}}^2}{D_{3\text{d}}} \\ \frac{4\pi l_{\text{DNA}} D_{3\text{d}} t}{\ln(D_{3\text{d}} t / r_{\text{int}}^2)} & , t \gg \frac{r_{\text{int}}^2}{D_{3\text{d}}} \end{cases} \quad [18]$$

The first regime corresponds to reaction limited binding and is very brief for large  $k_{\text{on}}^{\text{ns}}$ . We discard this regime and ignore the slowly varying logarithm in the last regime. This allows us to approximate  $P_{\text{surv}}^{\text{foreign}}$  by

$$P_{\text{app}}(t) = \exp \left[ -4l_{\text{DNA}} r_{\text{int}} \sqrt{\pi D_{3\text{d}} t} - k_{\text{cap}} t \right]. \quad [19]$$

Numerical analysis shows excellent agreement between  $P_{\text{surv}}^{\text{foreign}}$  and  $P_{\text{app}}$ . The value of  $k_{\text{cap}}$  is fixed self-consistently by the condition that the total amount of proteins returning to the original segment is the same as for  $P_{\text{surv}}^{\text{foreign}}$ , i.e.,

$$\int_0^\infty dt W_{\text{bulk}}^{\text{cyl}}(q=0, t) [P_{\text{surv}}^{\text{foreign}}(t) - P_{\text{app}}(t)] = 0. \quad [20]$$

We solved this equation numerically. Once a value for  $k_{\text{cap}}$  is obtained the choice of  $P_{\text{app}}$  in Eq. 19 turns out to be very convenient. In fact, it can be written

$$P_{\text{app}}(t) = \int_0^\infty ds e^{-(s+k_{\text{cap}})t} \frac{a}{2\sqrt{\pi s^3}} e^{-a^2/(4s)}, \quad [21]$$

where  $a = 4\sqrt{\pi} r_{\text{int}} l_{\text{DNA}} \sqrt{D_{3\text{d}}}$ . When inserted into Eq. 17 in Laplace space we need to evaluate numerically only one integral to obtain  $W_{\text{bulk}}(q, u)$ , namely

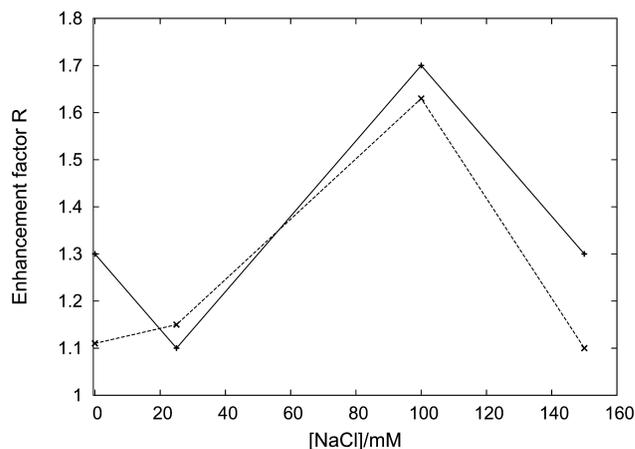
$$W_{\text{bulk}}(q, u) = \int_0^\infty ds \frac{a e^{-a^2/(4s)}}{2\sqrt{\pi s^3}} W_{\text{bulk}}^{\text{cyl}}(q, u + s + k_{\text{cap}}). \quad [22]$$

The term with  $\delta_{q,0}$  was already discarded here. Eq. 22 generalizes the straight rod result 12. In the limit  $D_{3\text{d}} \gg k_{\text{on}}^{\text{ns}}$  it is consistent with the approximation  $1 - \lambda_{\text{bulk}}(q) \approx 1$  leading to the expected result (13) for that case.

### Discussion

A single stochastic search mechanism is usually not efficient at optimizing a search process. Thus, Brownian motion, whose fractal dimension is  $d_f = 2$ , in 1-dimensional and 2-dimensional geometries leads to significant oversampling in the sense that a given site is revisited many times (27, 28). Lévy motion with a jump length distribution  $\lambda(x) \simeq |x|^{-1-\alpha}$  ( $0 < \alpha < 2$ ) has fractal dimension  $d_f = \alpha$  and thus a comparatively reduced degree of oversampling. However, the first arrival of Lévy motion is hampered by leapovers, i.e., frequent overshoots of the target (29). In 3-dimensional space both Brownian and Lévy motions are inefficient at locating a small target. Whereas typical chemical reactants in a liquid environment have no choice but to passively diffuse until mutual encounter (4) intermittent models combine different search mechanisms (28, 30). An important example of intermittent search models is the facilitated diffusion of transcription factors in gene regulation.

Here, we have developed a theoretical framework analyzing facilitated diffusion in the presence of intersegmental jumps whose relevance depends on the degree of DNA coiling. These results together with the experimental evidence from our single DNA experiments reported in ref. 20 demonstrate that the binding rate of DNA-binding proteins can be significantly enhanced



**Fig. 3.** Enhancement factor  $R$  of the search rate  $k_{\text{on}}$  in the maximally relaxed DNA configuration versus the search rate  $k_{\text{on}}$  for stretched DNA, as a function of NaCl concentration. The full line shows the values obtained from the experiment reported in ref. 20, and the dashed line represents the values from our theoretical approach. Interestingly the maximum occurs close to the physiological salt concentration of 100 mM.

when DNA is allowed to coil freely. This effect would be further amplified when the local density of DNA is increased beyond that of a relaxed coil, for instance, by packaging in a living cell. This is consistent with observations of increased cleavage rate of restriction enzymes in supercoiled over relaxed DNA plasmids (31).

In our derivation of the probability  $P_{\text{surv}}^{\text{foreign}}$  we assumed that the density of foreign segments is uniform in space. Although this is not true for a freely fluctuating coil (the density decreases on a scale of the persistence length of the DNA) we argue that this non-uniformity is not important for evaluating the search rate: in fact, if the protein diffuses far away from the original segment it will most likely perform a long relocation measured along the DNA contour. Again it does not matter for the search rate how we bookkeep these long relocations. To estimate the density  $l_{\text{DNA}}$  for a random relaxed coil we employ the Worm Like Chain model. A good approximation for the probability density to loop back, for a point on the chain a contour distance  $s$  away from another point on the chain, is (32)

$$j_M = \left( \frac{3}{4\pi |s| \ell_p} \right)^{3/2} \exp \left( -\frac{8\ell_p^2}{|s|^2} \right). \quad [23]$$

For a target in the middle of a chain of length  $L$ , this leads to the DNA density  $l_{\text{DNA}} = \int_{-L/2}^{L/2} ds j_M$  around the target.

The theory presented here quantitatively describes recent experimental data from an optical tweezers setup reported in ref. 20. In that assay the search rate  $k_{\text{on}}$  was measured at various degrees of DNA coiling for a DNA of length  $L = 6,538$  base pairs (approximately  $2.24 \mu\text{m}$  corresponding to 42 persistence lengths). The relative increase  $R$  of  $k_{\text{on}}$  in the maximally relaxed DNA configuration compared with  $k_{\text{on}}$  in the stretched state is shown in Fig. 3 as a function of the NaCl concentration.  $R$  was measured to be  $\approx 1.1$ – $1.3$  in buffers with salt concentrations of 5 mM  $\text{MgCl}_2$ , and 0–25 mM for NaCl, respectively. This relative increase  $R$  compares well with the theory presented here by using estimates of  $l_{\text{DNA}}$  based on Eq. 23 giving  $1/\sqrt{l_{\text{DNA}}} \sim 500$  bp, and a value  $l_{\text{si}}^{\text{eff}} \sim 100$  bp based on a quantitative analysis (33) of experiments (34, 35), suggesting effective sliding lengths of this size. Numerical analysis of the theory presented here shows that  $R$  depends significantly only on  $l_{\text{DNA}}$  and  $l_{\text{si}}^{\text{eff}}$  for realistic ranges of the remaining parameters. At physiological salt (100 mM NaCl) an increase of  $R \approx 1.7$  is found experimentally (20). This observation

agrees well with the increased DNA density expected because of attraction between segments induced by the 5 mM divalent  $Mg^{2+}$  ions at this NaCl concentration, see ref. 20 for details. We are not aware of a theory to calculate  $I_{DNA}$  under such attraction. However, the value  $1/\sqrt{I_{DNA}} \sim I_{sl}^{eff}$  needed to obtain  $R \approx 1.7$  at this salt condition appears reasonable. At even higher salt the ratio  $R$  drops again, which is consistent with a reduction in  $I_{sl}^{eff}$  because of weaker nonspecific binding at high salt (34, 35).

We presented a model for facilitated diffusion that is useful to calculate the search rate of DNA-binding proteins for their specific binding site on a fluctuating DNA coil with variable DNA density. This model represents a convenient way to rephrase the Berg–von Hippel theory with the additional advantage that it interpolates between stretched and coiled DNA configurations. In the development of this theory a number of assumptions were made, in particular, that of sufficiently fast local relaxation of the DNA

configuration. This assumption was found to be satisfied for the protein EcoRV used in ref. 20, but it will not hold universally. It would therefore be beneficial to generalize the theory to slowly fluctuating or static (quenched) DNA configurations, potentially leading to effects similar to those found in polymer models (36). Another interesting question for future work concerns the potential effect on in vivo gene regulation of the slow diffusion caused by molecular crowding (37, 38), including interesting consequences for the exchange dynamics of proteins with the DNA (39).

**ACKNOWLEDGMENTS.** We thank Tobias Ambjörnsson, Alexander Grosberg, Mehran Kardar, Leonid Mirny, and Roland Winkler for discussions. This work was partially funded by the Villum Kann Rasmussen Foundation, Deutsche Forschungsgemeinschaft, and by a Fundamenteel Onderzoek der Materie Projectruimte Grant, a Netherlands Organization for Scientific Research–Vernieuwings Impuls grant.

- Alberts B, et al. (2002) *Mol Biol Cell* (Garland Science, New York).
- Ptashne M (2004) *A Genetic Switch* (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY).
- Riggs AD, Bourgeois S, Cohn M (1970) The lac repressor-operator interaction. III. Kinetic studies. *J Mol Biol* 53:401.
- von Smoluchowski M (1916) Three presentations on diffusion, molecular movement according to Brown and coagulation of colloid particles. *Physikal Zeitschr* 17:557–571.
- von Hippel PH, Berg OG (1989) Facilitated target location in biological systems. *J Biol Chem* 264:675–678.
- Sokolov IM, Metzler R, Pant K, Williams MC (2005) Target search of  $N$  sliding proteins on a DNA. *Biophys J* 89:895–902.
- Elf J, Li GW, Xie XS (2007) Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* 316:1191–1194.
- Wang YM, Austin RH, Cox EC (2006) Single molecule measurements of repressor protein 1D diffusion on DNA. *Phys Rev Lett* 97:048302.
- Gowers DM, Wilson GG, Halford SE (2005) Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA. *Proc Natl Acad Sci USA* 102:15883–15888.
- Bakk A, Metzler R (2004) In vivo non-specific binding of  $\lambda$  CI and Cro repressors is significant. *FEBS Lett* 563:66–68.
- Bakk A, Metzler R (2004) Nonspecific binding of the  $O_R$  repressors CI and Cro of bacteriophage  $\lambda$ . *J Theor Biol* 231:525–533, and references therein.
- Stormo GD, Fields DS (1998) Specificity, free energy and information content in protein-DNA interactions. *Trends Biochem Sci* 23:109–113.
- Bintu L, et al. (2005) Transcriptional regulation by the numbers: Models. *Curr Opin Genet Dev* 15:116–124.
- Walczak AM, Onuchic JN, Wolynes PG (2005) Absolute rate theories of epigenetic stability. *Proc Natl Acad Sci USA* 102:18926–18931.
- Blossey R, Giuraniuc CV (2008) Mean-field versus stochastic models for transcriptional regulation. *Phys Rev E* 78:031909.
- Kolesov G, Wunderlich Z, Laikova ON, Gelfand MS, Mirny LA (2007) How gene order is influenced by the biophysics of transcription regulation. *Proc Natl Acad Sci USA* 104:13948–13953.
- Little JW, Shepley DP, Wert DW (1999) Robustness of a gene regulatory circuit. *EMBO J* 18:4299–4307.
- Ozbudak EM, Thattai M, Lim NH, Shraiman BI, van Oudenaarden A (2004) Multistability in the lactose utilization network of *Escherichia coli*. *Nature* 427:737–740.
- Bonnet I, et al. (2008) Sliding and jumping of single EcoRV restriction enzymes on non-cognate DNA. *Nucleic Acids Res* 36:4118–4127.
- van den Broek B, Lomholt MA, Kalisch SMJ, Metzler R, Wuite GJL (2008) How DNA coiling enhances target localization by proteins. *Proc Natl Acad Sci USA* 105:15738–15742.
- Berg OG, Ehrenberg M (1982) Association kinetics with coupled 3-dimensional and one-dimensional diffusion–chain-length dependence of the association rate to specific DNA sites. *Biophys Chem* 15:41–51.
- Hu T, Grosberg AY, Shklovskii BI (2006) How proteins search for their specific sites on DNA: The role of DNA conformation. *Biophys J* 90:2731–2744.
- Halford SE, Marko JF (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res* 32:3040–3052.
- Slutsky M, Mirny LA (2004) Kinetics of protein-DNA interaction: Facilitated target location in sequence-dependent potential. *Biophys J* 87:4021–4035.
- Coppéy M, Bénichou O, Voituriez R, Moreau M (2004) Kinetics of target site localization of a protein on DNA: A stochastic approach. *Biophys J* 87:1640–1649.
- Lomholt MA, Ambjörnsson T, Metzler R (2005) Optimal target search on a fast-folding polymer chain with volume exchange. *Phys Rev Lett* 95:260703.
- Hughes BD (1995) *Random Walks and Random Environments* (Oxford Univ Press, Oxford), Vol 1.
- Lomholt MA, Koren T, Metzler R, Klafter J (2008) Lévy strategies in intermittent search processes are advantageous. *Proc Natl Acad Sci USA* 105:11055–11059.
- Koren T, Lomholt MA, Chechkin AV, Klafter J, Metzler R (2007) Leapover lengths and first passage time statistics for Lévy flights. *Phys Rev Lett* 99:160602.
- Bénichou O, Loverdo C, Moreau M, Voituriez R (2008) Optimizing intermittent reaction paths. *Phys Chem Chem Phys* 10:70591–7072.
- Gowers DM, Halford SE (2003) Protein motion from non-specific to specific DNA by three-dimensional routes aided by supercoiling. *EMBO J* 22:1410–1418.
- Ringrose L, Chabanis S, Angrand PO, Woodroffe C, Stewart AF (1999) Quantitative comparison of DNA looping in vitro and in vivo: Chromatin increases effective DNA flexibility at short distances. *EMBO J* 18:6630–6641.
- Zhou H-X (2005) A model for the mediation of processivity of DNA-targeting proteins by nonspecific binding: Dependence on DNA length and presence of obstacles. *Biophys J* 88:1608–1615.
- Jeltsch A, Pingoud A (1998) Kinetic characterization of linear diffusion of the restriction endonuclease EcoRV on DNA. *Biochemistry* 37:2160–2169.
- Stanford NP, Szczelkun MD, Marko JF, Halford SE (2000) One- and three-dimensional pathways for proteins to reach specific DNA sites. *EMBO J* 19:6546–6557.
- Sokolov IM, Mai J, Blumen A (1997) Paradoxical diffusion in chemical space for nearest-neighbor walks over polymer chains. *Phys Rev Lett* 79:857–860.
- Golding I, Cox EC (2006) Physical nature of bacterial cytoplasm. *Phys Rev Lett* 96:098102.
- Dix JA, Verkman AS (2008) Crowding effects on diffusion in solutions and cells. *Annu Rev Biophys* 37:247–263.
- Lomholt MA, Zaid IM, Metzler R (2007) Subdiffusion and weak ergodicity breaking in the presence of a reactive boundary. *Phys Rev Lett* 98:200603.