Facilitated Diffusion of DNA-Binding Proteins

Konstantin V. Klenin,[1] Holger Merlitz,[2,*] Jörg Langowski,[1,†] and Chen-Xu Wu[2]

[1]*Division of Biophysics of Macromolecules, German Cancer Research Center, D-69120 Heidelberg, Germany*
[2]*Department of Physics, Xiamen University, Xiamen 361005, People's Republic of China*

The diffusion-controlled limit of reaction times for site-specific DNA-binding proteins is derived from first principles. We follow the generally accepted concept that a protein propagates via two competitive modes, a three-dimensional diffusion in space and a one-dimensional sliding along the DNA. However, our theoretical treatment of the problem is new. The accuracy of our analytical model is verified by numerical simulations. The results confirm that the unspecific binding of protein to DNA, combined with sliding, is capable to reduce the reaction times significantly.

PACS numbers: 87.16.Ac

Introduction.—The understanding of diffusion-controlled chemical reactions has become an indispensable ingredient of present-day technological development. The optimization of catalysts, fuel cells, improved batteries using electrodes with nanostructured surfaces, or the function of semiconductive devices are just a few of countless examples where diffusive processes, often in crowded or fractal environments, are involved to define the most important system parameters. For any living organism, diffusion plays the central role in biochemical and physical reactions that keep the system alive [1,2]: the transport of molecules through cell membranes, of ions passing the synaptic gap, or drugs on the way to their protein receptors are predominantly diffusive processes. Furthermore, essentially all of the biological functions of DNA are performed by proteins that interact with specific DNA sequences [3,4], and these reactions are diffusion controlled.

However, it has been realized that some proteins can find their specific target sites on DNA much more rapidly than is "allowed" by the diffusion limit [1,5,6]. It is therefore generally accepted that some kind of facilitated diffusion must take place in these cases. Several mechanisms, differing in details, have been proposed for it. All of them essentially involve two steps. First, the protein binds to a random nonspecific DNA site. Second, it diffuses (slides) along the DNA chain. These two steps may be reiterated many times before the protein actually finds the target, since the sliding is occasionally interrupted by dissociation.

Berg *et al.* have provided a thorough (but somewhat sophisticated) theory that allows an estimation of the resulting reaction rates [5]. Recently, Halford and Marko have presented a comprehensive review on this subject and proposed a remarkably simple semiquantitative approach that explicitly contains the mean sliding length as a parameter of the theory [6].

In the present work we suggest an alternative view on the problem starting from first principles. Our theory leads to a formula that is similar in form to that of Halford and Marko, apart from numerical factors. In particular, we give a new interpretation of the sliding length, which makes it possible to relate this quantity to experimentally accessible parameters.

Theory.—To estimate the mean time τ required for a protein to find its target, we consider a single DNA chain in a large volume V. At time $t = 0$, the protein molecule is somewhere outside the DNA coil. We introduce the "reaction coordinate" r as the distance between the center of the protein and the center of the target, which is assumed to be presented in one copy. When r is large, the only transport mechanism is the 3-dimensional (3D) diffusion in space. On the contrary, at small r, the 1-dimensional (1D) diffusion along the DNA chain is more efficient.

Let us define the efficiency of a transport mechanism in more strict terms. Let $\tau(r - dr, r)$ be the mean time of the first arrival of the protein at the distance $(r - dr)$ from the target, provided it starts from the distance r. In the simple cases, when the diffusion of a particle can be fully characterized by a single coordinate, this time is given by the equation [7,8]

$$d\tau \equiv \tau(r - dr, r) = \frac{Z(r)}{D\rho(r)} dr, \qquad (1)$$

where D is the diffusion coefficient, $\rho(r)$ the equilibrium distribution function of the particle along the reaction coordinate (not necessary normalized), and $Z(r)$ the local normalizing factor

$$Z(r) = \int_r^\infty \rho(r') dr'. \qquad (2)$$

Note that the quantity $1/d\tau$ is the average frequency of transitions $r \to r - dr$ in the "reduced" system with a reflecting boundary at the position $r - dr$ (so that the smaller distances from the target are forbidden). The quantity

$$\nu \equiv \frac{dr}{d\tau} = \frac{D\rho(r)}{Z(r)} \qquad (3)$$

has the dimension of velocity and can be regarded as a measure for the efficiency of a transport process.

For 3D diffusion inside the volume V, we have $\rho(r) = 4\pi r^2 c$, where c is the protein concentration and the factor 4π is chosen to provide a convenient normalization for a system containing only one protein molecule: $Z(0) = Vc = 1$. Hence, for sufficiently small r, when $Z(r) \approx Z(0) = 1$, the transport efficiency is

$$v_{3D}(r) = 4\pi D_{3D} r^2 c. \tag{4}$$

In the case of a 1D diffusion along the DNA chain we have $\rho(r) = 2\sigma$, with σ being the linear density of a nonspecifically bound protein. The factor 2 accounts for the fact that the target can be reached from two opposite directions. We assume, again, that the distance r is sufficiently small, so that the DNA axis can be considered as a straight line. Thus, the efficiency of the 1D-diffusive transport near the target is given by

$$v_{1D} = 2D_{1D}\sigma. \tag{5}$$

Our main assumption is that, during the combined diffusion process, the probability of the (nonspecifically) bound state is close to its equilibrium value for each given value of r. Then the frequencies $1/d\tau_{3D}$ and $1/d\tau_{1D}$ are additive, and so are the efficiencies of the two transport mechanisms given by Eqs. (4) and (5). Hence, the mean time of the first arrival at the target of radius a can be found as

$$\tau = \int_a^\infty \frac{dr}{v_{3D} + v_{1D}}. \tag{6}$$

The main contribution to this integral is made by the distances close to a. For that reason, the upper limit of integration is set to infinity. Before evaluation of Eq. (6), we note that

$$1 = Z(0) = Vc + L\sigma, \tag{7}$$

where V is the volume and L is the DNA length. The meaning of this equation is that the system contains only one protein molecule. Substituting Eqs. (4) and (5) into Eq. (6) and taking into account Eq. (7), we get, finally,

$$\tau = \left(\frac{V}{8D_{3D}\xi} + \frac{\pi L\xi}{4D_{1D}}\right)\left[1 - \frac{2}{\pi}\arctan\left(\frac{a}{\xi}\right)\right]. \tag{8}$$

Here, we have introduced a new parameter

$$\xi = \sqrt{\frac{D_{1D}K}{2\pi D_{3D}}}, \tag{9}$$

with $K = \sigma/c$ being the equilibrium constant of nonspecific binding. It is easy to verify that ξ is just the distance, where the efficiencies of the two transport mechanisms [Eqs. (4) and (5)] become equal to each other.

Numerical model.—In what follows we present numerical simulations to test the accuracy of our analytical result for the reaction time given by Eqs. (8) and (9). In order to approximate the real biological situation, the DNA was modeled by a chain of N straight segments of equal length l_0. Its mechanical stiffness was defined by the bending energy associated with each chain joint:

$$E_b = k_B T \alpha \theta^2, \tag{10}$$

where $k_B T$ is the Boltzmann factor, α the dimensionless stiffness parameter, and θ the bending angle. The numerical value of α defines the persistence length, i.e., the "stiffness" of the chain [9]. The excluded volume effect was taken into account by introducing the effective DNA diameter, d_{eff}. The conformations of the chain, with the distances between nonadjacent segments smaller than d_{eff}, were forbidden. The target of specific binding was assumed to lie exactly in the middle of the DNA. The whole chain was packed in a spherical volume (cell) of radius R in such a way that the target occupied the central position.

In order to achieve a close packing of the chain inside the cell, we first generated a relaxed conformation of the free chain by the standard METROPOLIS Monte Carlo (MC) method. For further compression, we defined the center norm (c norm) as the maximum distance from the target (the middle point) to the other parts of the chain. Then, the MC procedure was continued, but a MC step was rejected if the c norm was exceeding 105% of the lowest value registered so far. The procedure was stopped when the desired degree of compaction was obtained.

The protein was modeled as a random walker within the cell with reflecting boundaries. During one step in the free 3D mode, it was displaced by the distance ε_{3D} in a random direction. Once the walker approached the chain closer than a certain capture radius r_c, it was placed to the nearest point on the chain and its movement mode was changed to the 1D sliding along the chain contour. In this mode, the step represented a displacement by the distance ε_{1D} performed with an equal probability in either direction. The ends of the chain were reflective. After each 1D step (and immediately after the capture) the walker could jump off the chain by the distance r_c and reenter the 3D mode. This operation was carried out with the kickoff probability p.

A simulation cycle started with the walker at the periphery of the cell and ended when the walker came within the distance a to the target. During all simulation cycles the chain conformation remained fixed.

Below in this Letter, one step is chosen as the unit of time and one persistence length of the DNA chain (50 nm) as the unit of distance. The following values of parameters were used. The length of one segment was chosen as $l_0 = 0.2$, so that one persistence length was partitioned into 5 segments. The corresponding value of the stiffness parameter was $\alpha = 2.403$ [9]. The effective chain diameter was $d_{eff} = 0.12$, the capture radius $r_c = d_{eff}/2$, and the radius of the active site was $a = 0.08$. The diffusion coefficients are defined as $D_{3D} = \varepsilon_{3D}^2/6$ and $D_{1D} = \varepsilon_{1D}^2/2$. The step size of the walker was $\varepsilon_{3D} = 0.04$ and $\varepsilon_{1D} = \varepsilon_{3D}/\sqrt{3}$,

yielding identical diffusion coefficients $D_{3D} = D_{1D} = 8 \times 10^{-4}/3$.

The radius R of the cell and the DNA length L were varied in different sets of simulation. For each fixed pair (R, L), the kickoff probability was initially set to $p = 1$ (no 1D transport, $\xi = 0$) and subsequently reduced to $p_i \equiv 2^{-i}$, $i = 1, 2, \ldots, 11$. For each parameter set, the simulation cycle was repeated 2000 times. The equilibrium constant K required for the calculation of the parameter ξ [Eq. (9)] has to be determined as the ratio $V\tau_{1D}/L\tau_{3D}$, where τ_{1D} and τ_{3D} are the average times the walker spent in the bound and the free states, respectively. Note that ξ depends on the choice of the probability p, but not on cell size or chain length, since $\tau_{1D} \sim L$ and $\tau_{3D} \sim V$. For each choice of p, the constant K was determined in a special long simulation run without target for specific binding.

Results.—In a first set of simulations, chains of various lengths between $L = 8$ and $L = 56$ were packed into a cell of radius $R = 2$ and volume $V_0 = 4\pi R^3/3 = 32\pi/3$. The resulting averaged reaction times τ are plotted in Fig. 1 as a function of the variable ξ [Eq. (9)]. The curves are plots of Eq. (8). It is obvious that the above relation was well able to reproduce the simulation results on a quantitative level. This good agreement between theoretical and computational model indicates that the derivation of Eq. (8), although quite simple, already contains the essential ingredients of the underlying transport process. A moderate deviation between simulation and theory is visible in case of $L = 56$ and large values of ξ. In the discussion we will shortly touch the limits of the theoretical approach if ξ becomes very large. With the present selection of chain

parameters, the results prove that a 1D sliding can speed up the reaction time significantly. If, however, the unspecific binding becomes too strong, its effect turns into the opposite and the reaction time is increasing. The most efficient transport is achieved with a balanced contribution of both 1D- and 3D-diffusion.

Figure 2 displays the results of a second set of simulations, where the longest chain of $L = 56$ was placed into cells with volumes of two, four, and eight times the initial value $V_0 = 32\pi/3$, leading to systems of rather sparse chain densities. The plots of Eq. (8) are again in good overall agreement with the simulation results, although a systematic deviation in case of large cell volumes, i.e., at low chain densities, is visible. The theoretical approach seems to underpredict the reaction time by up to 10%. A systematic investigation of the limits of our approach is part of ongoing research. For the time being we note that in crowded environments (of high chain density) Eq. (8) appears to be more accurate than in sparse environments.

Discussion.—Recently, Halford and Marko have proposed a remarkably simple semiquantitative approach to estimate the reaction time [6], yielding the expression

$$\tau = \frac{V}{D_{3D}l_{sl}} + \frac{Ll_{sl}}{D_{1D}}. \tag{11}$$

Following their argumentation, l_{sl} was interpreted as the average sliding length of the protein on the DNA contour. It is instructive to note that, for $\xi \gg a$, Eq. (8) turns into

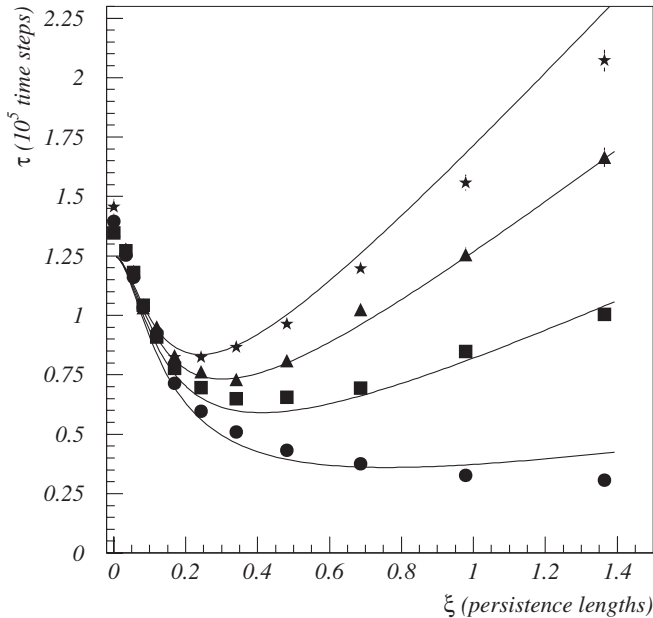

FIG. 1.   Reaction time τ as a function of the sliding parameter ξ [Eq. (9)] at a fixed cell radius $R = 2$ and chain lengths $L = 56$, 40, 24, 8 (top to bottom). The curves are plots of Eq. (8).

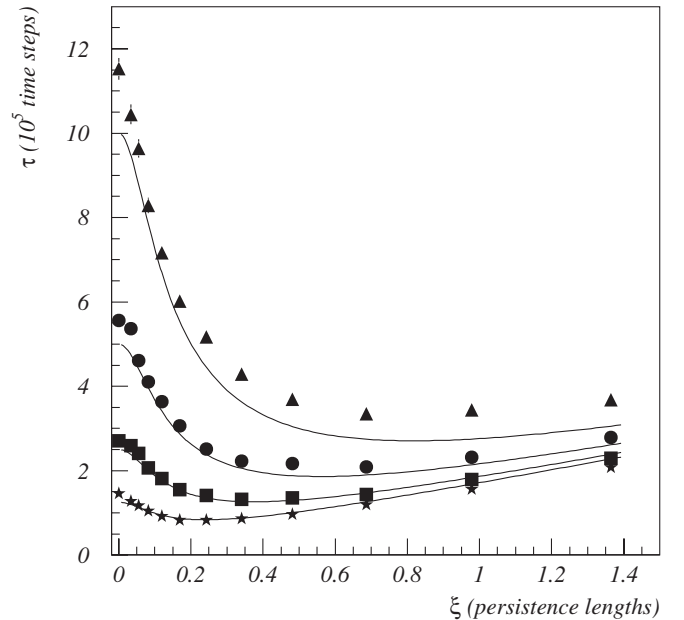FIG. 2.   Reaction time τ as a function of the sliding parameter ξ [Eq. (9)] at fixed chain length $L = 56$ and with varying cell volumes ($8x$, $4x$, $2x$, and $1x$ the original volume $V_0 = 32\pi/3$, top to bottom). The curves are plots of Eq. (8).

$$\tau = \frac{V}{8D_{3D}\xi} + \frac{\pi L \xi}{4D_{1D}}, \tag{12}$$

which is of identical functional form if we identify ξ with the sliding length of Halford and Marko. With Eq. (9) we can now express l_{sl} in terms of experimentally accessible quantities, assigning a physical meaning to a previously heuristic model parameter. Additionally, Eq. (12) contains the numerical factors which turn the initially semiquantitative approach into a model of quantitative accuracy.

Our results demonstrate (Fig. 2) that crowding decreases the optimum sliding length: the shortest reaction time is reached at lower nonspecific binding affinities. In a crowded environment the chance for the protein to bind or rebind nonspecifically is much higher, so that the period of free diffusion is shorter after each kick. In contrast, in sparse environments the chance to hit the target is increased if the protein remains in sliding mode over a rather long distance. Increasing the chain density will shift the minimum of τ to lower values of ξ (Fig. 2), while decreasing the chain length *at constant volume* will shift it to higher values (Fig. 1). The derivative of Eq. (12) allows an estimate of the optimum sliding length ξ_{opt}:

$$\xi_{opt} = \sqrt{\frac{VD_{1D}}{2\pi L D_{3D}}}. \tag{13}$$

Sliding distances have been estimated experimentally to up to 1000 bp for the restriction endonuclease EcoRV in dilute solution from the dependence of cleavage rate on DNA length [10], but from the same enzyme's processivity a much shorter sliding length of about 50 bp was estimated later [11]. The DNA concentration in the latter work was 5 nM for a 690 bp DNA, while the highest chain density used here was 0.4 nM for $L = 56$ persistence lengths, corresponding to an 8230 bp DNA. For the DNA length and concentration used in [11], $\xi_{opt} = 0.22$, or 33 bp. We thus see that the relatively short sliding lengths estimated in more recent work make good sense for the biological function of DNA-binding proteins, since they constitute the best compromise between one- and three-dimensional search.

The limits of our new approach are presently under investigation. In the derivation of Eq. (6) we assumed chemical equilibrium between the free and the nonspecifically bound states of the walker. For high affinity of the protein to the DNA, i.e., large values of ξ, this assumption may not be justified, since the protein always starts in free diffusion mode at the periphery of the cell. The violation of that assumption may become more serious if the chain density inside the cell is low, so that the protein has to search for a long time before it is able to bind to the DNA

for the first time. Additionally, in order to evaluate the efficiency of 1D diffusion [Eq. (5)], it was assumed that the DNA axis could be considered as a straight line over the distance of 1D diffusion. This is satisfied if the sliding length is smaller than the persistence length of the chain, i.e., $\xi < 1$.

In summary, the relation (8), derived from first principles, provides a quantitative estimate for the reaction time of a protein that is moving under the control of two competitive transport mechanisms in a crowded environment. Although drawing an idealized picture of the living cell, it will serve as the starting point for more realistic approaches, equipped with additional parameters that are subsequently calibrated in sophisticated simulations. The sliding parameter ξ [Eq. (9)] connects the heuristic sliding length of Halford *et al.* to experimentally accessible quantities. The simulations, although so far performed on a limited range of system parameters, confirm earlier results that an unspecific binding combined with a 1D-diffusion mode enables for a significant speedup of the reaction. The relation (8) can be used to extend the investigations to system sizes which are not easily accessible in numerical simulations such as those presented in this work: the size of a realistic cell nucleus is of the order of 10 microns and it contains DNA chains adding up to a length of the order of meters.

We thank J. F. Marko for fruitful discussions.

*Electronic address: merlitz@gmx.de
†Electronic address: jl@dkfz.de

[1] A. D. Riggs, S. Bourgeois, and M. Cohn, J. Mol. Biol. **53,** 401 (1970).
[2] P. H. Richter and M. Eigen, Biophys. Chem. **2,** 255 (1974).
[3] O. G. Berg and P. H. von Hippel, Annu. Rev. Biophys. Biophys. Chem. **14,** 131 (1985).
[4] M. Ptashne and A. Gann, *Genes and Signals* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001).
[5] O. G. Berg, R. B. Winter, and P. H. von Hippel, Biochemistry **20,** 6929 (1981).
[6] S. E. Halford and J. F. Marko, Nucleic Acids Res. **32,** 3040 (2004).
[7] A. Szabo, K. Schulten, and Z. Schulten, J. Chem. Phys. **72,** 4350 (1980).
[8] K. V. Klenin and J. Langowski, J. Chem. Phys. **121,** 4951 (2004).
[9] K. Klenin, H. Merlitz, and J. Langowski, Biophys. J. **74,** 780 (1998).
[10] A. Jeltsch, C. Wenz, F. Stahl, and A. Pingoud, EMBO J. **15,** 5104 (1996).
[11] N. P. Stanford, M. D. Szczelkun, J. F. Marko, and S. E. Halford, EMBO J. **19,** 6546 (2000).