Coarse-grained modeling of protein unspecifically bound to DNA

2014 Phys. Biol. 11 026003

(http://iopscience.iop.org/1478-3975/11/2/026003)

IOP Publishing Physical Biology
Phys. Biol. 11 (2014) 026003 (11pp) doi:10.1088/1478-3975/11/2/026003

Coarse-grained modeling of protein unspecifically bound to DNA

Carlo Guardiani[1], **Massimo Cencini**[2] **and Fabio Cecconi**[2]

[1] Dipartimento di Fisica, University 'Sapienza', Piazzale A. Moro 2, 00185 Rome, Italy
[2] Istituto dei Sistemi Complessi, Consiglio Nazionale delle Ricerche, via dei Taurini 19, 00185 Rome, Italy

E-mail: fabio.cecconi@roma1.infn.it

Received 29 October 2013, revised 7 January 2014
Accepted for publication 19 February 2014
Published 1 April 2014

Abstract

There is now a certain consensus that transcription factors (TFs) reach their target sites, where they regulate gene transcription, via a mechanism dubbed *facilitated diffusion* (FD). In FD, the TF cycles between events of 3D diffusion in solution (jumps), 1D diffusion along DNA (sliding), and small jumps (hopping), achieving association rates higher than for 3D diffusion alone. We investigate the FD phenomenology through molecular dynamics simulations in the framework of coarse-grained modeling. We show that, despite the crude approximations, the model generates, upon varying the equilibrium distance of the DNA–TF interaction, a phenomenology matching a number of experimental and numerical results obtained with more refined models. In particular, focusing on the kinematics of the process, we characterize the geometrical properties of TF trajectories during sliding. We find that sliding occurs via helical paths around the DNA helix, leading to a coupling of translation along the DNA axis with rotation around it. The 1D diffusion constant measured in simulations is found to be interwoven with the geometrical properties of sliding and we develop a simple argument that can be used to quantitatively reproduce the measured values.

Keywords: facilitated diffusion, coarse-grained modeling, transcription factor, molecular dynamics

1. Introduction

Transcription factors (TFs) play a key role in the regulation of gene expression, acting as gene-transcription activators or inhibitors both in prokaryotes and in eukaryotes [1]. One of the most fundamental issues in protein–DNA recognition is the ability of TFs to selectively identify their specific target sites that are embedded among tens of millions of competing non-specific DNA sequences. A related issue pertains to the high rate of recognition of the specific target sites. As early as 1970, Riggs *et al* [2] observed that the *lac* repressor in *E. coli* can associate with the cognate operator sequence at a rate about two orders of magnitude higher than that predicted by the Smoluchowski equation for a diffusion-limited association reaction.

Berg *et al* [3] explained this paradox, suggesting that TFs do not target their sequences through pure 3D diffusion, and can also diffuse while being unspecifically associated (mainly due to electrostatic interactions [4]) with the DNA. Such a dimensional reduction, dubbed *facilitated diffusion* (FD), can make the search more efficient, speeding up the identification of target sites. More specifically, FD proceeds by means of four pathways [3]: (i) *sliding* along the DNA, (ii) *hopping*, (iii) *jumping* and (iv) *intersegmental transfer*. During sliding, the TF remains in unspecific contact with the DNA chain, performing mono-dimensional diffusion along its contour. During hopping, the TF detaches from the DNA but reassociates with it at a short distance from the dissociation point. During jumping, the TF dissociates from the DNA, undergoing free 3D diffusion, and rebounds to the DNA in a completely uncorrelated location. Finally, in the intersegmental transfer, relevant to compact DNA conformations, the TF transiently binds two non-contiguous DNA branches, allowing its transfer from one DNA segment to the other. The latter process requires the possibility for the TF to bind at multiple loci.

FD has been extensively studied through analytical models [3–6] which achieve closed-form solutions at the price of a drastic simplification in the complexity and the heterogeneity of the genome. The approximation of the TF–DNA affinity landscape, for instance, may lead to significant deviations from the experimental patterns. A more detailed level of description is based on computational stochastic models which allow large-scale simulations involving DNA stretches of the order of 10^6 bp, and tens of thousands of TFs, and can reach the time scale of a few seconds [7, 8]. This high performance, however, relies on a set of assumptions that are considered quite controversial and that need further elucidation. More specifically, the issues include: (i) the proportion of sliding and hopping during 1D diffusion; (ii) the fraction of time that the TF spends in 3D and 1D diffusion; (iii) the effects of molecular crowding related to the presence of multiple copies of the TF that prevent each other's movement, acting as moving roadblocks.

In order to clarify these issues, experimental studies can be profitably integrated with coarse-grained molecular simulations. For instance, while fluorescence experiments have allowed the direct observation of a single TF moving along DNA (confirming the FD theory) [9], the spatial resolution of the technique does not discriminate between hopping and sliding. Thus a quantitative characterization of the two types of motion still remains elusive. Another source of ambiguity concerns the values of the mono-dimensional and tri-dimensional diffusion constants. While there is a general consensus on the fact that $D_1 < D_3$, the measured values of these constants vary by several orders of magnitude according to the particular DNA sequence and the experimental setup [10, 11]. This variability is anything but irrelevant since it is closely related to the so-called *speed–stability paradox* [5, 12]. In fact, on the one hand, a high diffusion constant allows a fast scanning of non-specific sites, improving the searching for the target sequences, while on the other hand, high D_1 values can only be attained at the price of a low TF–DNA affinity that may destabilize the complex formed by the TF with its specific target site.

Another question that has not yet been met with a conclusive answer is that of the fraction of time spent by the TF in 3D diffusion and in sliding. Assuming that only sliding and jumping are at work, simple analytical arguments [4, 5] suggest that the average time necessary to reach the target is $t_s = (\tau_1+\tau_3)M/\overline{n}$, where M is the total number of sites, $\overline{n}$ is the average number of sites scanned during a single sliding event, and τ_1 and τ_3 are the average durations of individual episodes of sliding and 3D diffusion, respectively. Assuming that the search time has been to some extent optimized by evolution, t_s is minimal if $\tau_1 = \tau_3$, i.e. when the TF spends exactly the same amount of time in sliding and 3D diffusion. This hypothesis of optimality, however, contrasts with experimental studies on bacteria suggesting that the TF spends much more time in sliding than in 3D diffusion ($\tau_1/(\tau_1 + \tau_3) = 0.9$) [9]. Even though the discrepancy may be due to the absence of hopping in the above argument, of course, one cannot exclude the possibility that evolution has selected a suboptimal solution.

All of these problems can be addressed through molecular dynamics simulations, but unfortunately not in the framework of atomistic methods. The longest atomistic simulation for DNA reported to date has been a few microseconds [13], while sliding events typically involve time scales of $O(s)$ and sliding lengths of $O(100)$ bp [14]. Thus, it is clear why atomistic MD methods are not suited to the study of FD, and resorting to a coarse-grained phenomenological modeling is mandatory. Recently, Brackley *et al* [15] introduced a coarse-grained model portraying the TF as a sphere with a binding site on its surface and the DNA as a chain of beads. The model, also accounting for both DNA flexibility and sequence heterogeneity, showed that the search time could be minimized by an appropriate tuning of the TF–DNA affinity. Givaty and Levy [16] proposed a much more detailed model whereby DNA is simplified as a double-stranded helix with three beads per nucleotide, while the TF is described as a bead for each residue. Levy's simulations show that during sliding, the TF remains deeply buried in the major groove and presumably makes use of the same binding interfaces for specific and non-specific DNA interactions.

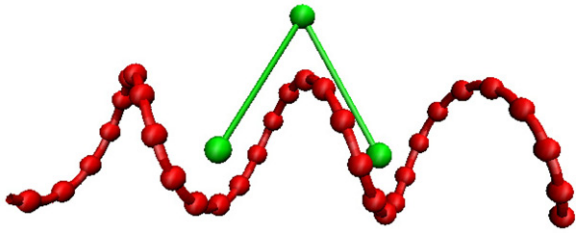

Figure 1. Cartoon representation of the TF (in green) and DNA (in red) model used in this work; see the text for details.

In this work, we introduce and study a model with a level of resolution intermediate between those mentioned above. With reference to figure 1 TF is portrayed as a triangular object with the first and last beads representing the DNA binding regions, so as to mimic the basic features of homodimeric prokaryotic TFs. The central bead models the scaffold of the protein, imparting the correct orientation to the DNA binding domains. The DNA is represented by a single helix frozen in its equilibrium conformation, as this greatly facilitates the identification of the various searching regimes.

The aim of our work is to develop a toy model including as few ingredients as possible, yet able to capture the known phenomenology of the dynamics of TFs unspecifically bound to DNA. The approach enabled us to connect the mono-dimensional diffusion coefficient to the geometrical properties of the TF trajectories, confirming previous results [17, 18]. This suggests that two minimal key elements are sufficient: the helical topology of DNA and a confining DNA–TF interaction tethering the TF in the neighborhood of the DNA.

We restricted our study to the case of purely non-specific TF–DNA interaction whose importance cannot be underestimated, as it is reasonable to assume that a TF before reaching its target spends most of the time in non-specific attraction with DNA sites. On the other hand, this is at the core of the searching process, which being 'unproductive', requires speeding up mechanisms or shortening optimal pathways.

A preliminary exploration of parameter space has been performed to obtain behaviors that reasonably match the principal features of the FD phenomenology. Then we run simulations to analyze the TF dynamics in the proximity of the DNA in order to characterize the interplay between three searching modes: sliding, hopping, and free diffusion. As we shall see, our simulations show the existence of these three different regimes whose mutual prominence depends on σ, the minimum of the Lennard-Jones (LJ) potential, which dramatically affects the energy landscape. Specifically, we found that sliding occurs via the coupling of rotation and translation along the DNA, in which the TF propagates one-dimensionally along the DNA while rotating along the DNA-helical contour. This characteristic motion is consistent with experimental observations for several proteins [19].

Finally we have also quantified how each mechanism contributes to search efficiency.

2. Materials and methods

2.1. The model

In the following we briefly describe the coarse-grained representation used in this study for the DNA chain and the TF.

The DNA helix. During the simulations the DNA was kept frozen in its initial configuration, so the DNA beads, used to represent the DNA bases, do not interact with one another. The DNA configuration is chosen to be a straight helix as, unlike [15], we want to retain the helical geometry in order to understand its effect on the FD process. However, since we adopt a very crude model for the TF (see below), we do not need the detailed description of the DNA double helix proposed in [16]. We thus consider a minimal representation in terms of a single straight helix where each bead represents a base pair (bp). In particular, to mimic the typical conformation of B-DNA [1] we consider a helix of radius $\varrho = 13.0$ Å, with $h = 10.5$ base pairs per helix turn and the distance between two consecutive bases along the DNA axis (here chosen to be along z) taken to be $b = 3.32$ Å, so the helix pitch is $P = hb = 34.86$ Å. In this way the coordinates of the nth bead are simply obtained from the parametric equations of the helix

$$\begin{aligned} x(n) &= \varrho \cos(2\pi z(n)/P) \\ y(n) &= \varrho \sin(2\pi z(n)/P), \quad n = 1, \ldots, N \\ z(n) &= bn\,. \end{aligned} \tag{1}$$

The total number of base pairs N in our simulations is $N = 1000$ which is larger than the DNA persistence length (about 100 bp in physiological conditions [1]). However, assuming a linear conformation longer than that found *in vivo* is relevant to single-molecule experiments, where DNA chains are typically stretched (see, e.g., [20]).

Transcription factor. The modeling of the TF requires some discussion. In prokaryotes, TFs are typically homodimeric, as they target palindromic DNA sequences [21]. In each monomeric subunit the DNA recognition region is a helix–loop–helix motif whereby the second helix is designed to fit into the major groove establishing hydrogen bonds and hydrophobic interactions with the nucleotide bases. Since the helix–loop–helix motifs of the two subunits must fit into two adjacent major grooves, they are located at the typical distance of one pitch $P \approx 32$–34 Å. In eukaryotes, TFs can be both homodimeric and heterodimeric so as to increase the range of DNA sequences to be recognized. For instance, steroid hormones receptors are typical homodimeric receptors while the TFs containing the leucine zipper motif are normally heterodimeric and the helix–loop–helix TF can be both homodimeric and heterodimeric [21]. Our modeling approach aims at reproducing the basic features of prokaryotic TFs. Therefore, we portray the TF as three beads arranged at the vertices of an equilateral triangle of side 32 Å, to roughly fit the distance between two major grooves. A variation of the side in the range 28–35 Å and isosceles TF conformations do not affect the essence of the results. The first and third beads can be thought of as the centers of mass of the DNA-recognizing regions of the two subunits. The third bead represents the center of mass of the portion of the TF not directly involved in DNA recognition, which typically stays away from the DNA helix.

The TF triangular structure is enforced by the following interactions. The 1–2 and 2–3 distances of the TF beads are allowed to undergo small oscillations around their equilibrium value, r_0, via a stiff harmonic potential [22]

$$V_h(r_{i,i+1}) = \frac{k_h}{2}(r_{i,i+1} - r_0)^2\,, \tag{2}$$

whereas the 1–3 distance is maintained via a bending potential

$$V_\theta(\theta) = \frac{k_\theta}{2}(\theta - \theta_0)^2\,. \tag{3}$$

Being interested in the phenomenology of FD and not in the target search time, we assume only unspecific interactions between the TF and the DNA chain, which are modeled as described below. Bead 1 and bead 3 interact with the DNA beads through a standard 12–10 LJ potential:

$$V_{\mathrm{LJ}}(r_{ij}) = 5\epsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \frac{6}{5}\left(\frac{\sigma}{r_{ij}}\right)^{10}\right], \tag{4}$$

where r_{ij} is the distance between bead $i \in \{1, 3\}$ of the TF and bead j of the DNA. The parameter ϵ determines the well depth of the LJ potential while σ tunes the position of the minimum. Thus σ determines the equilibrium distance of the TF from the helix axis: the larger σ, the farther the equilibrium position of the TF from the DNA. Conversely, bead 2 of the TF interacts with the nucleotides of DNA through a repulsive, excluded-volume potential

$$V_{\mathrm{rep}}(r_{2j}) = \epsilon_2\left(\frac{\sigma_2}{r_{2j}}\right)^{12}. \tag{5}$$

This potential forces the central bead of the TF to point away from the DNA axis, imparting the correct orientation to the TF. In our simulation, we kept ϵ fixed to set the energy scale and varied σ over a wide range of values. For the sake of clarity, the list and the values of the parameters defining the DNA–TF model are summarized in table 1.

Simulation box. Since the focus of our investigation was the sliding behavior of the TF, we introduced a cylindrical confinement potential [23]:

$$V_{\mathrm{conf}} = V_{xy} + V_z = \frac{k_B T}{(R_{xy} - r)^2} + \frac{k_B T}{(R_z - |\Delta z|)^2}\,. \tag{6}$$

Table 1. Table summarizing the parameters and their values used in the DNA–TF interaction model and in the simulations.

Parameter	Value
ϵ	1
ϵ_2	0.8ϵ
k_h	50ϵ
k_θ	20ϵ
k_BT	0.25ϵ
σ_2	5 Å
r_0	32 Å
θ_0	60°
γ	1
m	1

In this expression $r = \sqrt{x^2 + y^2}$ is the distance between a TF bead and the DNA axis that was set to coincide with the z-axis, Δz is the distance along the z-axis between the bead of the TF and the center of mass of the DNA, and $R_{xy} = 100$ Å is the radius of the cylindrical confinement region. The parameter R_z represents half the height of the confinement region, that we set equal to half the length of the DNA plus 1.5 helical turns. The V_{xy} component of the confinement potential forces the TF to remain in a circular region of radius R_{xy} centered on the DNA axis, while the V_z component prevents the TF from exceeding a distance equal to R_z from the DNA center of mass along the z-axis.

The value $R_{xy} = 100$Å for the simulation box can be justified using the following argument. The average volume available to interphasic DNA spans the range 10^{11}–10^{12} Å^3. We can assume that this is the volume of a spherical region $V = 4\pi R_g^3/3$ with R_g being the DNA gyration radius. Following Berg and Blomberg [24], we can construct around the DNA contour a coaxial cylinder with a volume equivalent to the sphere $4\pi R_g^3/3 = \pi R_{xy}^2 L$, where $L \sim 10^7$ Å is the typical DNA length. This yields values of R_{xy} in the range 60–200 Å. On not too long time scales, the TF may be reasonably assumed to be confined in a cylindrical region of radius R_{xy} around a DNA segment. Since metaphasic DNA is more condensed, it can be assumed to be confined in a cylindrical region with a smaller radius R_{xy}. In this situation, the TF can be expected to spend a smaller fraction of time in 3D diffusion, which is similar to what happens for small values of σ (see section 3.1). Moreover the TF will have a greater tendency to rebind the DNA in the neighborhood of the point of detachment. In this regime there will be only a weak interplay between sliding and 3D diffusion, leading to a low efficiency of exploration of new sites (see section 3.4). This appears to be consistent with the fact that tightly packed DNA is normally not transcribed or replicated, but rigidly transferred to daughter cell during mitosis.

We performed Langevin molecular dynamics simulations using a stochastic position Verlet integration scheme [25] with time step $h = 0.002$ and friction coefficient γ (see table 1). The simulation time unit can be converted to the physical one by using the time scale $\tau = \sigma_2\sqrt{m/\epsilon}$ [26]. With $\epsilon = 4k_BT \simeq 16 \times 10^{-21}$ J and $\sigma_2 = 5$ Å, and assuming an average mass $m \simeq 10$ kDa for each bead of the TF, we obtain $\tau \sim 9$ ps.

As is customary, the Lennard-Jones interactions were truncated at a cutoff distance $r_c = 4\sigma$ to speed up the calculations.

2.2. Statistical analysis.

2.2.1. Determination of sliding, hopping and jumping events. As discussed in the introduction, the process of FD proceeds by means of four pathways [3]: (i) sliding along the DNA, (ii) hopping, (iii) jumping and (iv) intersegmental transfer. In our model, due to the chosen conformation for the DNA chain, only the first three mechanisms are at work. In order to compute the statistics of sliding, jumping and hopping, it is necessary to define the criteria for discriminating each event, which are described below.

We consider the TF to be in the sliding regime if the closest DNA neighbor of bead 1 or bead 3 of the TF is below a distance cutoff of 1.2σ. This criterion allows the identification of a number of sliding and non-sliding windows. With this criterion it may happen that the TF is bound to the DNA with only one bead, while the other is detached. We have studied the statistics of such events and found that when sliding occurs the percentage of time spent in two-bead sliding: is higher than 90% for $\sigma \geqslant 8$ Å; is between 40% and 90% for $4\,\text{Å} \leqslant \sigma < 8$ Å; and decreases by up to a few % for $1 \leqslant \sigma < 4$ Å.

For each value of σ, the average sliding length, $\langle|\Delta Z_s|\rangle$, is measured as the average distance covered by the TF between an attachment and the first subsequent detachment. Non-sliding windows are classified as hopping events if the displacement of the TF along the DNA axis is smaller than twice $\langle|\Delta Z_s|\rangle$; otherwise the event is considered as jumping. The idea underlying this choice is that hopping implies short-range flights between dissociation and reassociation points [27].

Clearly, the discrimination between sliding and hopping and that between hopping and jumping suffer from a certain degree of arbitrariness due to the necessity of introducing specific thresholds in distances. However, upon varying the threshold values, we verified that the results are qualitatively the same except for some quantitative effects on the hopping statistics.

2.2.2. Computation of the sliding diffusion constant. The mono-dimensional diffusion process of the TF during sliding is characterized by the diffusion coefficient, D_1, along the DNA axis. The constant D_1 can be estimated through the trajectories of molecular dynamics from the mean square deviation (MSD) along the z-axis during the sliding events. First, the trajectory of the TF is segmented into sliding, hopping and jumping events as described above. Second, for each sliding window w one computes the MSD on the window as (see e.g. [28])

$$\overline{\Delta Z^2(k, w)} = \sum_{i=1}^{N_w - k} \frac{(Z_{i+k} - Z_i)^2}{N_w - k}, \tag{7}$$

where Δt is the time interval between two successive measurements, N_w the total number of measurements in w and $Z_i = z(\mathrm{i}\,\Delta t)$ indicates the z-coordinate of an attractive bead of the TF at time i Δt. The square deviations (7) are then

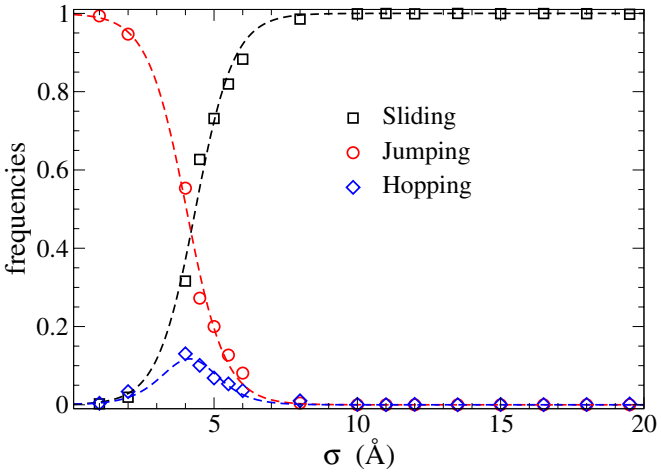

Figure 2. Transport mode statistics. Frequencies of sliding (black squares), jumping (red circles) and hopping (blue diamonds). The dashed lines are just a guide for the eye and have been obtained by fitting the data via suitable sigmoid functions. Data are obtained by averaging over 50 runs each lasting $T = 10^6$ time units (see section 2).

averaged over all the M sliding windows of all trajectories such that $N_w \geqslant k$:

$$\langle \Delta Z^2(k) \rangle = \frac{1}{M} \sum_{w=1}^{M} \overline{\Delta Z^2(k, w)} \, . \tag{8}$$

Since our TF is perfectly symmetric, the calculation is independent on the chosen bead apart from statistical fluctuations whose impact can be minimized by averaging the results of the two equivalent attractive beads. This averaged quantity provides an estimation of the mean square displacement along z over a time interval $k\,\Delta t$, which for a diffusive process behaves as

$$\langle \Delta Z^2(k) \rangle = 2D_1 \, \Delta t \, k \, . \tag{9}$$

The constant D_1 is finally obtained by linear regression.

2.2.3. Computation of the exploration efficiency. The efficiency of the DNA exploration by the TF can be estimated from the fraction of DNA sites not yet visited by the TF during the sliding ν_{sites}. The procedure that we used follows [16] and is described below. At the beginning of each run the counter for the newly probed DNA basis is set to zero. Then, at the beginning of every sliding window within the run, each DNA bead is marked with a flag 'zero'. Then if the DNA bead closest to one of the TF attractive beads is within a distance of 1.2σ from the latter, the corresponding flag is switched to '1' and the counter of probed sites increased by 1. At the end of each sliding event, when the TF detaches from the DNA, the counter is normalized to the number of DNA sites to get the fraction of sites explored in that sliding event. Then the flag vector is reset to zero. The overall fraction of visited sites is just the sum of the fractions of sites visited in all the sliding windows in each run. This quantity is then averaged over all runs to yield ν_{sites}. The strategy of resetting the flag vector to zero is motivated by the assumption that when the TF detaches from the DNA, it is likely to reassociate with a completely uncorrelated sequence exploring a completely new patch of DNA.

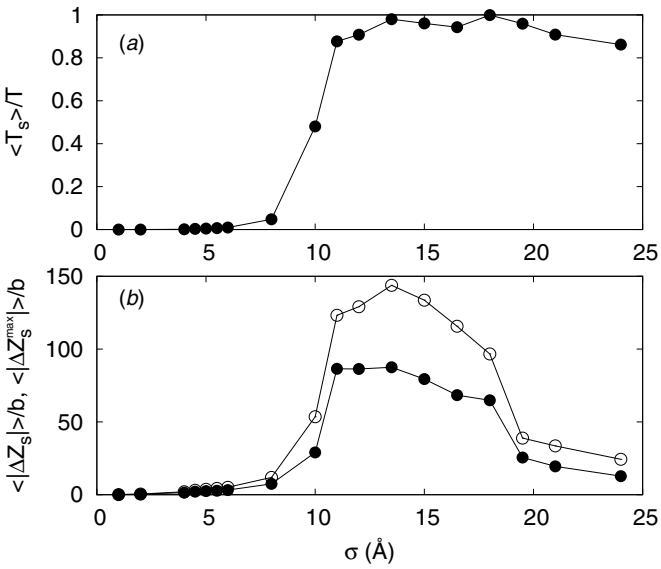

Figure 3. Sliding statistics: duration and length. (*a*) Average duration of individual sliding events as a function of σ normalized to the duration of the single-run time window $T = 10^6$ code time units. Notice that the saturation at high σ values is a consequence of the fact that sliding occurs over the full duration of the simulation run. (*b*) Average distance $\langle |\Delta Z_s| \rangle$ (full circles) and average maximal distance $\langle |\Delta Z_s^{\max}| \rangle$ (empty circles) covered by the TF in an individual sliding event as a function of σ and normalized by the base pair distance b.

3. Results

Experimental studies of FD *in vitro* are conducted by varying the salt concentration, which influences the occurrence of the different transport modes [29–31]. As we shall see, in our model a similar behavior is obtained by changing the parameter σ, the equilibrium distance between the TF and the DNA helix. An increase in the salt concentration enhances the screening of electrostatic interactions and thus increases the TF–DNA equilibrium distance, which in our model is controlled by σ.

3.1. The statistics of sliding, hopping and jumping

In order to understand the importance of the different transport modes (sliding, hopping and jumping) of TFs while interacting with the DNA, we need first to evaluate their occurrence statistics. In figure 2, we show the empirical transport mode frequencies of occurrence measured in simulations. The various modes were identified and analyzed as discussed in section 2. We find that sliding and jumping frequencies follow a sigmoidal profile with a prevalence of 3D diffusion at low σ and a dominance of sliding at high σ. Hopping events are rare for every value of σ except for a small hump around $\sigma \approx 4$ Å. This a likely consequence of the strongly confining features of the 12–10 LJ interactions. When the equilibrium distance between the TF and DNA, σ, is too small, complete dissociations with enduring jumps are the most probable events. Conversely, for large σ values, the TF tends to linger, bound to the DNA, with sliding events lasting for about the entire duration of the simulation runs (figure 3(*a*)). Hopping

events are statistically significant only at the transition between the jumping dominated and sliding dominated regimes.

Sliding is the most relevant transport mode in determining the TF–DNA interaction, as only when sliding can the TF can actually probe the sequence of nucleotides of the DNA. We thus studied the average distance along the DNA axis, $\langle|\Delta Z_s|\rangle$, explored by the TF during a single sliding event. Due to the random walk character of sliding, in the interval between the times of attachment and detachment the TF might have moved past the future point of detachment. Therefore, we also measured the maximal distance from the point of attachment reached by the TF within the time of detachment, $\langle|\Delta Z_s^{\max}|\rangle$. Both $\langle|\Delta Z_s|\rangle$ and $\langle|\Delta Z_s^{\max}|\rangle$, normalized by the base pair distance b, are shown in figure 3(b). These quantities provide a proxy for the number of bases probed between the point of association with the DNA and the subsequent point of detachment from it. As one can see, when sliding is dominating the number of bases probed by the TF can be as large as ∼150. However, for values of σ larger than 18–20 Å, while sliding remains the prevailing mechanism of transport (figure 2 and 3(a)), the number of probed bases drops dramatically. This behavior will be rationalized later while investigating the behavior of the one-dimensional diffusion constant D_1.

3.2. The geometry of sliding TF trajectories

We now focus on the geometrical properties of TF trajectories during sliding. The basic phenomenological features are illustrated in figures 4 and 5, showing the positions of one of the attractive beads (i.e. bead 1 or bead 3) during sliding, for three representative values of σ. Figure 4 shows a three-dimensional view, while figure 5 displays two-dimensional projections. Denoting with (x, y, z) the Cartesian coordinates of the bead position, the left panels show the projection onto the (x, y)-plane, transverse to the DNA axis, while the right ones show the cosine of the angle of rotation around the DNA axis as a function of the position along the DNA axis, i.e. $(z, x/\sqrt{x^2+y^2}) = (z, \cos(\theta))$.

For all values of σ we found that, during sliding, the TF traces the helical path of the DNA, as is clear from the 3D plots (figure 4), so diffusion along the DNA chains proceeds with a characteristic roto-translation as suggested by experimental studies [20, 32–35]; see also the review [19]. However, some differences are observed with varying σ, as discussed in the following.

For $\sigma = 6$ Å, the TF traces circular orbits orthogonal to the helix contour around each DNA bead, so the overall motion draws a superhelical trajectory (figure 4 and 5(a), left). In this case the envelope path is in phase with the DNA helix, as demonstrated by the behavior of the points representing $\cos(\theta)$ versus z (figure 5(a), right), which accumulate around the curve $\cos(2\pi z/P)$ (P being the DNA-helix pitch). For smaller values of σ sliding becomes less frequent, but is always in phase with the DNA helix (not reported). For larger values of σ, the TF sliding beads trace a helix in antiphase with respect to the DNA helix (figures 4 and 5(b), (c) (right)), meaning that the TF recognition domains reside in the DNA groove. As far as the distance from the DNA axis is concerned, for $\sigma = 6$ Å the TF bead moves both in and out of the DNA helix (figure 5(a), right), while it remains well inside and outside it for $\sigma = 15$ Å and 21 Å, respectively (as shown in (figure 5(b), (c) (left)). As we will show in the next section, these observations will be key to understanding the behavior of the one-dimensional diffusion coefficient D_1.

We complete the description of the geometrical aspects of TF sliding motion by discussing its orientation with respect to the DNA. Given the TF triangular geometry, its orientation can be characterized in terms of the angle ϕ between the segment joining bead 1 and bead 3 and the DNA axis. Since the dynamics is fully symmetric under the exchange of bead 1 and bead 3, we can restrict the angle to the range $[0 : 90]$ degrees. Measurements are made only when the TF is associated with DNA, i.e. in the sliding windows. Figure 6 shows that the orientation statistics depends on σ. For $\sigma < 6$ Å (figure 6(a)), sliding occurs very rarely and typically only one of the attractive beads is in contact with DNA; this can be appreciated from figures 4(a) and 5(a), left, where the spots out of the superhelical path correspond to the instants at which the TF is sliding but with one of the beads not attached to the DNA. As a consequence, the probability density, $p(\phi)$, is rather broad with no preferential orientation. The peaks observed in the figure result from the loops that the TF makes around DNA beads. For intermediate values of σ (6 Å $< \sigma <$ 18 Å; figure 6(b)), the TF describes a helical path in the interior of the DNA helix (central panel of figure 4 and figure 5(b)) and $p(\phi)$ takes on a well defined peak around $\phi = 0^\circ$ meaning that the TF slides along the DNA in parallel orientation, with beads 1 and 3 residing in two consecutive grooves of the DNA. As we shall discuss below, this appears to be the fastest TF–DNA configuration in terms of diffusive properties. For $\sigma > 18$ Å (figure 6(c)) the peak around zero broadens and a new peak appears between 70° and 80°. These features signal that now the attractive beads, while performing a helical motion outside the DNA helix, flip between a parallel orientation with respect to the DNA axis and an almost orthogonal one, where the two beads straddle the helix.

3.3. The one-dimensional diffusion coefficient

To characterize the TF sliding along the DNA, we estimated from the runs the one-dimensional diffusion coefficient D_1 from the linear behavior of the mean square displacement along the DNA axis; see equation (9) in section 2.

The results reported in figure 7(a) show the dependence of D_1 on σ. The diffusion constant displays small values for both large and small σ, while exhibiting a bump for intermediate values. Figure 7(b) shows the average distance Δ of the attractive beads (1 and 3) from the z-axis. A negative correlation between $D_1(\sigma)$ and $\Delta(\sigma)$ is apparent, indicating that high values of D_1 require a TF deeply embedded into the DNA groove, as highlighted by the horizontal line that marks the DNA radius $\Delta = \varrho$.

The observed behavior of the diffusion constant, D_1 can be rationalized by a simple phenomenological argument based on the geometrical properties of TF sliding motion that have been characterized in the previous section. Basically, during

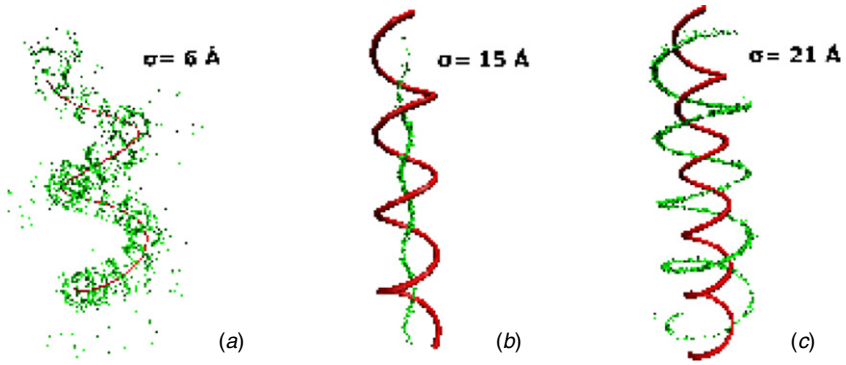

Figure 4. Roto-translation of the TF during sliding. The green dots represent the positions on one of the attractive TF beads for each frame satisfying the sliding condition. The trajectory traced by the TF is overlaid to the structure of the DNA chain (red). Panels (*a*)–(*c*) refer to three representative σ values as labeled in the figure.

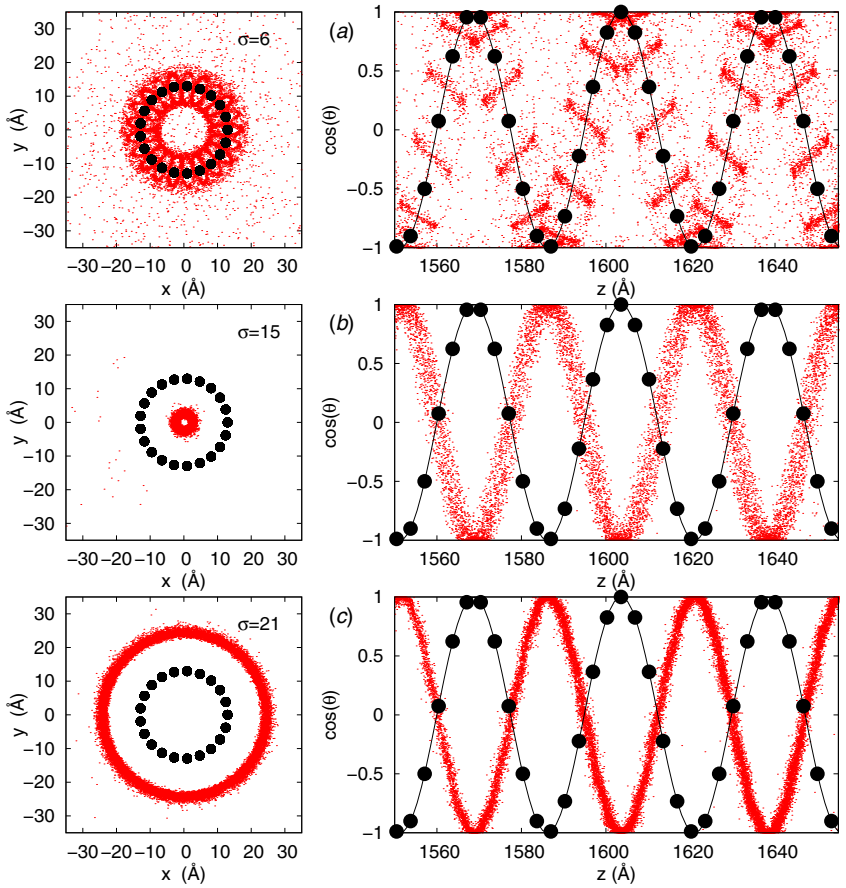

Figure 5. Two-dimensional projections (red dots) of the TF–bead trajectories of figure 4 for $\sigma = 6$ Å (*a*), 15 Å (*b*) and 21 Å (*c*). Left panels refer to the projection onto the (x, y)-plane. Right panels show the cosine of the angle of rotation of the TF around the z-axis (DNA axis) as a function of the position along the z-axis itself. The black curves correspond to the DNA helix $\cos(2\pi z/P)$. Black full circles identify the DNA beads.

sliding, TF beads diffuse, drawing a helical path (figure 4) at distance Δ from the DNA axis (figure 7(*b*)). Such a helical path has the same pitch $P = hb$ as the DNA helix with, possibly, a phase shift (figure 5(*b*), (*c*)), which is inessential for the following derivation. This scenario occurs for σ large enough (figure 4(*b*), (*c*)). For smaller σ the path drawn by the TF is slightly more complex (figure 4(*a*)), but on average still helical. Assuming an ideal helical motion, the displacement along the axis, δz, is linked to the arc-length of a curvilinear displacement, $\delta\ell$, along the helix by the formula

$$\delta z = \frac{\delta\ell}{\sqrt{1 + (2\pi\Delta/P)^2}}\,. \tag{10}$$

From the Einstein equation for a three-atom molecule, like our TF, the mean square curvilinear displacement along the helical path is $\langle\delta\ell(t)^2\rangle = 2D_3 t$ with $D_3 = k_B T/(3\gamma)$, while along the z-axis we have $\langle\delta z(t)^2\rangle = 2D_1 t$. Then, using equation (10) to

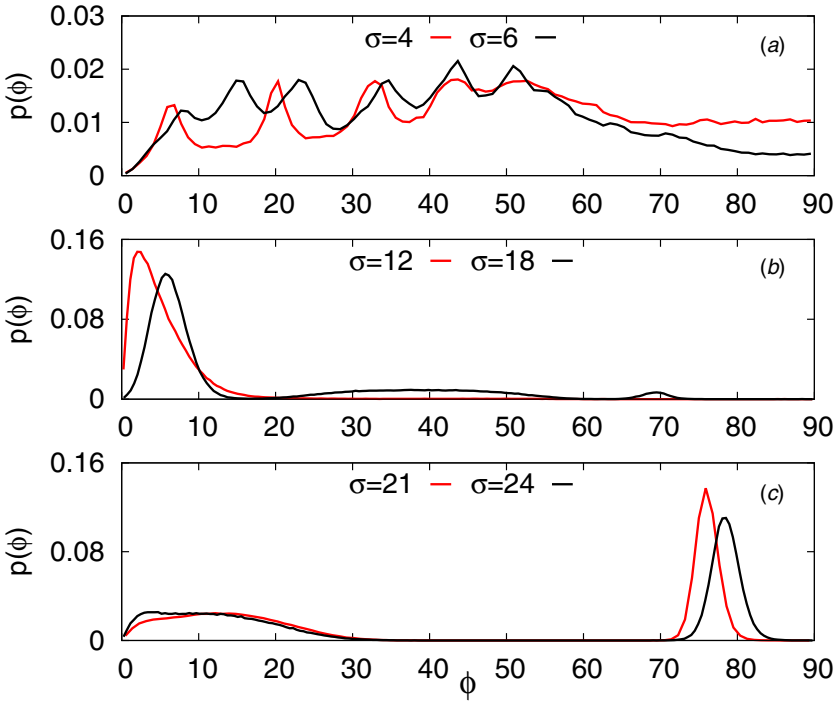

Figure 6. Probability density function $p(\phi)$ of the orientation angle ϕ of the TF with respect to the DNA axis: (*a*)–(*c*) are for small, intermediate and high values of σ as labeled. Notice the difference in *y*-axis scale between panel (*a*) and panels (*b*), (*c*).

convert displacements along the helix to displacements along the DNA axis yields

$$D_1 = \frac{D_3}{1 + (2\pi\Delta/P)^2} = \frac{k_B T}{3\gamma[1 + (2\pi\Delta/P)^2]}, \qquad (11)$$

which relates the diffusion constant along the z-axis to the geometrical properties of the helical path followed by the TF. The shaded region in figure 7(*a*) is bracketed by the upper and lower bounds of D_1 obtained using equation (11) by replacing Δ with $\Delta \pm s_\Delta$, where s_Δ is the standard deviation of the TF distance from the z-axis. The region accurately brackets the simulation data, supporting the reliability of the prediction (11). For instance, if the TF were diffusing on a helix with radius equal to that of the DNA, it would correspond to a D_1 with the value marked by the solid line in figure 7(*a*).

It is interesting to observe that equation (11) is consistent with the theoretical prediction of Bagchi *et al* [17] based on the computation of the translational friction induced by the TF helical track along the DNA. To obtain the connection, one should neglect the friction contribution of TF self-rotation predicted by Schurr [18] which is not relevant to our model.

The ability of equation (11) to quantitatively explain the behavior of the simulated sliding diffusion constant suggests that, within our model, D_1 is mainly determined by the geometrical properties of the sliding path. In other words, the DNA geometry conspires with the interaction potential to constrain the TF to diffusing along a helical path without being much influenced by possible potential barriers; indeed the derivation was based on the free diffusion coefficient D_3. Essentially the effect of the interaction potential is embodied in the fact that Δ in figure 7(*b*) turns out to be a non-trivial function of σ. In principle, the potential of interaction between TF and DNA depends on the nucleotide sequence, so diffusion is modified by the presence of a rugged energy landscape [6, 14]. This effect typically depresses the diffusion; in particular it will affect the value of D_3 used in the argument above by decreasing it by a factor that is $\propto \exp[-(E_v/k_B T)^2]$, E_v being the standard deviation of the TF–DNA (now disordered) binding energy. However, experimental data suggest that this effect, when present, is very small with $E_v \leqslant 1k_B T$–$2k_B T$ [14]. Of course, the model that we introduced can be easily generalized to include sequence heterogeneity.

3.4. Search efficiency

Even in a scenario of non-specific TF–DNA interaction it is interesting to quantify the efficiency of sequence exploration during sliding. Following [16], we estimate the exploration efficiency in terms of 'probed positions', i.e. we measure the fraction of new sites, ν_{sites}, visited by the sliding TF; see section 2.

In figure 8 we show the fraction ν_{sites} as a function of σ. The exploration efficiency displays a well pronounced peak in the range 4 Å $< \sigma <$ 5.5 Å. At first sight, this result may look surprising, as for such values of σ the diffusion constant D_1 is rather smaller than its maximum value attained at $\sigma \approx 13$ Å (figure 7). However, this behavior represents the essence of FD whereby slow sliding can be compensated by frequent jumping and hopping. Indeed, a direct comparison between figures 2 and 8 reveals that the search efficiency peaks in the region where hopping is maximal and jumping/sliding events have comparable frequencies. In other words, the possibility of realizing an efficient search through the DNA chain to localize as quickly as possible the DNA target sequence requires a

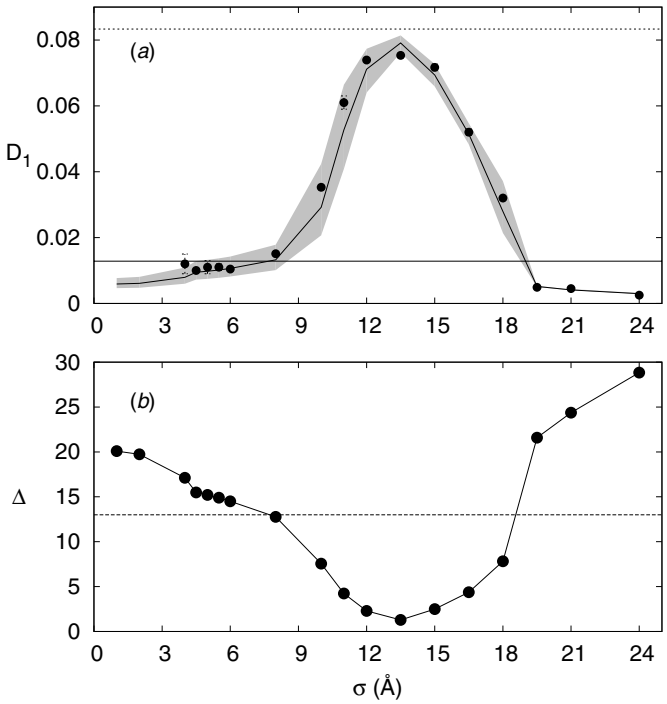

Figure 7. Sliding diffusion coefficient. (*a*) Dependence of the diffusion constant along the DNA axis, D_1, on σ. Filled circles denote the measured D_1 from simulations; the solid line interpolating data is formula (11) with Δ being the average TF–DNA distance from numerical simulations; the shaded region is bracketed by the values (11) computed by replacing Δ with $\Delta \pm s_\Delta$, where s_Δ is the standard deviation of the TF distance from the z-axis. For a comparison, the top horizontal dashed line displays the free diffusion coefficient $D_3 = k_B T/(3\gamma)$ for the three-bead TF and the bottom solid horizontal line shows formula (11) evaluated at $\Delta = \varrho$. (*b*) Average distance, Δ, of beads 1 and 3 from the DNA axis. The dashed line denotes the DNA radius ϱ; points below this line correspond to situations in which the beads are on average inside the DNA helix.

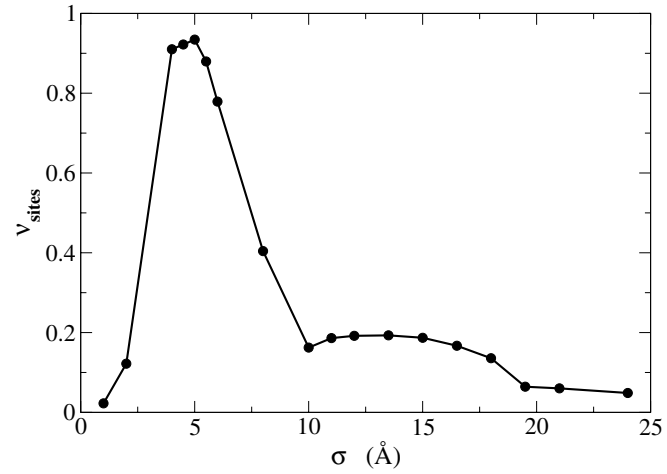

Figure 8. Fraction ν_{sites} of positions probed by the TF during sliding as a function of σ.

suitable interplay of all transport modes. The result shown in figure 8 is in qualitative agreement with those observed in [16].

4. Discussion and conclusions

In this work we performed molecular dynamics simulations of facilitated diffusion using a very simplified model. We represented DNA as a single helical chain of beads frozen in the standard conformation of B-DNA. To capture the main features of typical prokaryotic homodimeric transcription factors (TFs) that target palindromic DNA sequences [21], the TF was represented as a three-bead triangular structure, where the first and last beads correspond to the binding regions, whereas the central bead corresponds to the scaffold. Our model does not include electrostatics, and the TF–DNA interactions are modeled through a Lennard-Jones potential of well depth ϵ and equilibrium distance σ. In our simulations, ϵ is kept constant while exploring a wide range of σ values.

Our simulations show that the DNA–TF equilibrium distance σ crucially affects the dynamics of the TF. For small σ the TF spends most of its time in 3D diffusion. At intermediate values of σ a sharp transition occurs, with a drop in the jumping frequency and an abrupt increase in the sliding frequency paralleled by the appearance of a hump in the hopping frequency and an increase of the D_1 diffusion constant. The shape of the trajectory traced by the TF is also very sensitive to σ. For intermediate σ the TF forms circular orbits orthogonal to the DNA contour, creating a superhelical path. For larger σ the TF traces a helical trajectory in phase with the DNA groove. Both of the attractive beads of the TF are accommodated at the bottom of the groove, imparting a parallel orientation to the DNA axis. Finally, for even larger σ the coils of this helical path in antiphase with the DNA helix become wide enough to cause a drop in the D_1 diffusion constant. The behavior of the D_1 constant was explained by a simple geometric argument based on the projection of the mean square displacement of the TF trajectory onto the DNA axis. The expression that we derived is similar to the one introduced in [17] save for the self-rotation frictional contribution [18], which in our simulations is not relevant. We did not find any apparent dependence of D_1 on possible energetic barriers. However, the latter enter the expression for the diffusion constant in an implicit way by setting the average distance of the TF from the DNA.

An interesting feature of our model is that, despite the crude approximations, it was able to reproduce a number of known phenomenological patterns. For instance, for intermediate values of σ, the TF always remains deeply buried in the groove of the DNA molecule with a parallel orientation with respect to the DNA axis. This result is in agreement with a circular dichroism study by Johnson *et al* [36] showing that the TF interaction with unspecific DNA sequences is sufficient to induce the structuring of typical DNA binding motifs, and is confirmed by recent NMR analyses by Iwahara *et al* [37] revealing that protein HoxD9 interacts with non-specific binding sites using the same interface as was employed for the recognition of the specific target site. This result, also consistent with the Givaty and Levy findings [16], might have far reaching biological implications [38] suggesting the existence of only a very low barrier separating the search and recognition states postulated in [5].

It is also interesting to notice that in our model the translational move of the TF during sliding is always coupled to rotation around the DNA axis induced by the helical path, which is in phase either with the DNA strand or with its groove. This result is consistent with single-molecule fluorescence tracking assays performed by Blainey *et al* [20] for the calculation of D_1 for labeled human oxoguanine DNA glycosylase (hOgg1). This study was based on the observation that in the case of pure translation the D_1 coefficient depends on the TF radius R like $1/R$, while in the case of roto-translation $D_1 \propto 1/[(4/3)R^3 + R(R_{OC})^2]$ (with R_{OC} being the distance of the TF from the DNA axis, i.e. Δ in our notation; see figure 7(*b*)), so purely translational and roto-translational sliding can be discriminated from each other. This work, along with other recent studies [32–34] extending the analysis of sliding to several other proteins, supports the suggestive idea that the coupling between rotation and translation might be a feature shared by, at least, a group of TFs.

As an overall conclusion, our model, despite its crude approximations, turns out to reproduce fairly well a number of experimental patterns. This represents an *a posteriori* validation of the two key elements of our scheme, namely the helical topology of the DNA molecule and a TF–DNA interaction potential with a well localized minimum and a short tail. This extensive model validation suggests its viability in investigating more complex aspects of FD such as the influence of molecular crowding and DNA flexibility.

Acknowledgments

We thank A Vulpiani for discussions. FC and MC acknowledge support from MIUR PRIN-2009PYYZM5.

Appendix

Since detailed models of TF–DNA interaction normally include electrostatics through a Debye–Hückel potential (DH), in this appendix we will show that the Lennard-Jones potential (LJ) is flexible enough to account for relevant aspects of screened electrostatic interactions. As is customary, in order to include the excluded-volume effect and automatically remove possible singularities, the original DH potential is complemented with a short-range repulsive form (see e.g. [16])

$$V_{el}(r) = 5\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} + B\frac{\mathrm{e}^{-\lambda r}}{r}\right], \tag{A.1}$$

where λ is the inverse of the screening length. To achieve a mapping with the LJ potential, λ and B are parameters to be adjusted such that V_{el} presents the same position $x = \sigma$ and depth $V_{\min} = -\epsilon$ as the LJ minimum. It is convenient to rescale the distance $x = r/\sigma$, so that the minimum of the V_{LJ} potential lies at $x = 1$. The parameters λ and B are obtained by solving the system

$$\left.\frac{\partial V_{el}}{\partial x}\right|_{x=1} = 0 \tag{A.2}$$

$$V_{el}(x = 1) = -\epsilon. \tag{A.3}$$

Simple algebraic manipulations yield $\lambda\sigma = 9$ and $B = 6\sigma\exp(\lambda\sigma)/5$, so the interaction potential reads

$$V_{el}(r) = 5\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \frac{6}{5}\frac{\mathrm{e}^{-9(r/\sigma-1)}}{r/\sigma}\right]. \tag{A.4}$$

Since λ is known to be proportional to the inverse of the square root of the salt concentration C_s, it follows that $\sigma \propto \sqrt{C_s}$.

This simple argument shows that given a V_{el} potential, it is always possible to determine an approximating (equivalent) LJ potential characterized by a minimum with the same position and depth. A useful by-product of the employment of the LJ form consists in the possibility of readily locating the putative equilibrium position.

References

[1] Alberts B, Johnson A, Lewis J, Raff M, Roberts K and Walter P 2007 *Molecular Biology of the Cell* 5th edn (New York: Garland science)
[2] Riggs A D, Bourgeois S and Cohn M 1970 The lac repressor–operator interaction: III. Kinetic studies *J. Mol. Biol.* **53** 401–17
[3] 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
[4] Halford S E and Marko J F 2004 How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* **32** 3040–52
[5] Slutsky M and Mirny L A 2004 Kinetics of protein–DNA interaction: facilitated target location in sequence-dependent potential *Biophys. J.* **87** 4021–35
[6] Sheinman M, Bénichou O, Kafri Y and Voituriez R 2012 Classes of fast and specific search mechanisms for proteins on DNA *Rep. Prog. Phys.* **75** 026601
[7] Chu D, Zabet N R and Mitavskiy J 2009 Models of transcription factor binding: sensitivity of activation functions to model assumptions *J. Theor. Biol.* **257** 419–29
[8] Zabet N R and Adryan B 2012 A comprehensive computational model of facilitated diffusion in prokaryotes *Bioinformatics* **28** 1517–24
[9] 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
[10] 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
[11] Graneli A, Yeykal C C, Robertson R B and Greene E C 2006 Long-distance lateral diffusion of human rad51 on double-stranded DNA *Proc. Natl Acad. Sci. USA* **103** 1221–6
[12] Bénichou O, Loverdo C and Voituriez R 2008 How gene colocalization can be optimized by tuning the diffusion constant of transcription factors *Europhys. Lett.* **84** 38003
[13] Laughton C and Harris S A 2011 The atomistic simulation of DNA *Comput. Mol. Sci.* **1** 590–600
[14] 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
[15] Brackley C A, Cates M E and Marenduzzo D 2012 Facilitated diffusion on mobile DNA: configurational traps and sequence heterogeneity *Phys. Rev. Lett.* **109** 168103
[16] Givaty O and Levy Y 2009 Protein sliding along DNA: dynamics and structural characterization *J. Mol. Biol.* **385** 1087–97

[17] Bagchi B, Blainey P C and Xie X S 2008 Diffusion constant of a nonspecifically bound protein undergoing curvilinear motion along DNA *J. Phys. Chem.* B **112** 6282–4
[18] Schurr J M 1979 The one-dimensional diffusion coefficient of proteins absorbed on DNA: hydrodynamic considerations *Biophys. Chem.* **9** 413–4
[19] Gorman J and Greene E C 2008 Visualizing one-dimensional diffusion of proteins along DNA *Nature Struct. Mol. Biol.* **15** 768–74
[20] Blainey P C, Guobin L, Kou S C, Mangel W F, Verdine G L, Bagchi B and Xie X S 2009 Nonspecifically bound proteins spin while diffusing along DNA *Nature Struct. Mol. Biol.* **16** 1224–30
[21] Branden C and Tooze J 1999 *Introduction to Protein Structure* 2nd edn (New York: Garland Science)
[22] Hinchliffe A 2008 *Molecular Modelling for Beginners* 2nd edn (Chichester: Wiley)
[23] Cacciuto A and Luijten E 2006 Self-avoiding flexible polymers under spherical confinement *Nano Lett.* **6** 901–5
[24] Berg O G and Blomberg C 1977 Association kinetics with coupled diffusion. An extension to coiled-chain macromolecules applied to the lac repressor–operator system *Biophys. Chem.* **7** 33–39
[25] Melchionna S 2007 Design of quasi-symplectic propagators for Langevin dynamics *J. Chem. Phys.* **127** 044108
[26] Padding J T and Louis A A 2006 Hydrodynamic interactions and Brownian forces in colloidal suspensions: coarse-graining over time and length scales *Phys. Rev.* E **74** 031402
[27] Gowers D M, Wilson G G and Halford S E 2005 Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA *Proc. Natl Acad. Sci. USA* **102** 15883–8
[28] Qian H, Sheetz M P and Elson E L 1991 Single particle tracking. Analysis of diffusion and flow in two-dimensional systems *Biophys. J.* **60** 910–21
[29] Winter R B, Berg O G and von Hippel P H 1981 Diffusion-driven mechanisms of protein translocation on nucleic acids: 3. The Escherichia coli lac repressor–operator interaction: kinetic measurements and conclusions *Biochemistry* **20** 6961–77
[30] Winter R B, Berg O G and von Hippel P H 1981 Diffusion-driven mechanisms of protein translocation on nucleic acids: 3. The escherichia coli lac repressor–operator interaction: kinetic measurements and conclusions *Biochemistry* **20** 6961–77
[31] 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
[32] Liu S, Abbondanzieri E A, Rausch J W, Grice S F J and Zhuang X 2008 Slide into action: dynamic shuttling of HIV reverse transcriptase on nucleic acid substrates *Science* **322** 1092–7
[33] Kochaniak A B, Habuchi S, Loparo J J, Chang D J, Cimprich K A, Walter J C and van Oijen A M 2009 Proliferating cell nuclear antigen uses two distinct modes to move along DNA *J. Biol. Chem.* **284** 17700–10
[34] Lin Y, Zhao T, Jian X, Farooqui Z, Qu X, He C, Dinner A R and Scherer N F 2009 Using the bias from flow to elucidate single DNA repair protein sliding and interactions with DNA *Biophys. J.* **96** 1911–7
[35] Tafvizi A, Huang F, Leith J S, Fersht A R, Mirny L A and van Oijen A M 2008 Tumor suppressor p53 slides on DNA with low friction and high stability *Biophys. J.* **95** L01–3
[36] Johnson N P, Lindstrom J, Baase W A and von Hippel P H 1994 Double-stranded DNA templates can induce alpha-helical conformation in peptides containing lysine and alanine: functional implications for leucine-zipper and helix–loop–helix transcription factors *Proc. Natl Acad. Sci. USA* **91** 4840–4
[37] Iwahara J and Clore G M 2006 Detecting transient intermediates in macromolecular binding by para-magnetic NMR *Nature* **440** 1227–30
[38] Marcovitz A and Levy Y 2011 Frustration in protein–DNA binding influences conformational switching and target search kinetics *Proc. Natl Acad. Sci. USA* **108** 17957–62