Nonspecific DNA-Protein Interaction: Why Proteins Can Diffuse along DNA

Vincent Dahirel,[1,*] Fabien Paillusson,[2] Marie Jardat,[1] Maria Barbi,[2] and Jean-Marc Victor[2]

[1]*UPMC Université Paris 06, UMR 7195, PECSA, F-75005 Paris, France*

[2]*CNRS, UMR 7600, LPTMC, F-75005 Paris, France*

(Received 9 February 2009; published 2 June 2009)

Recent single molecule experiments have reported that DNA binding proteins (DNA-BPs) can diffuse along DNA. This suggests that interactions between proteins and DNA play a role during the target search even far from their specific site on DNA. Here we show by means of Monte Carlo simulations and analytical calculations that there is a counterintuitive repulsion between the two oppositely charged macromolecules at a nanometer range. For the concave shape of DNA-BPs, and for realistic protein charge densities, we find that the DNA-protein interaction free energy has a minimum at a finite surface-to-surface separation, in which proteins can easily slide. When a protein encounters its target, the free energy barrier is completely counterbalanced by the H-bond interaction, thus enabling the sequence recognition.

DOI: 10.1103/PhysRevLett.102.228101 PACS numbers: 87.15.ak, 82.70.Dd, 87.15.kj

DNA stores the genetic material of all living cells and viruses. This huge amount of information can be used effectively only if DNA binding proteins (DNA-BPs) manipulate DNA in specific locations. When a protein finds its DNA target site, the shape complementarity of DNA binding proteins and their specific DNA sequences maximize the number of hydrogen bonds, thus leading to a strong protein-DNA association [1–6]. The rate of protein-DNA association is, however, not controlled by the association step itself, but by the whole search process. It is well established now that DNA-BPs diffuse along DNA before they reach their specific site [7,8]. During this search, nonsequence-specific interactions are the only interactions between the protein and DNA which can play a role. Those nonspecific interactions remain poorly documented. Although the predominance of electrostatics is unquestionable [1–6], it remains unclear how the protein structure comes into play [5–7]. Does the typical concavity of DNA-BPs which favors the specific association also influence the nonspecific electrostatic interaction? In DNA-protein complexes, the mean charge of the protein residues located at the interface is positive [1,2]. Nevertheless, structural studies of nonspecific complexes have shown that the protein atoms and the DNA atoms are weakly packed together at the interface [1–3,5,6], thus suggesting that a force counterbalances the electrostatic attraction. In this Letter, our purpose is to establish the general mechanisms that control the mean force between proteins and DNA and that are applicable to a wide variety of DNA-BPs. With that goal in mind, we design coarse-grained DNA and protein models, rather than detailed atomic models and investigate their interactions.

The most characteristic aspect of DNA-BPs is their shape complementarity with DNA. The concave DNA-BPs can cover the convex DNA with up to 35% of their surface [1]. At close contact, solvent molecules are excluded from the interface and the protein forms numerous weak bonds with the DNA (mainly H bonds [1]). Initially, we artificially switch off these H-bond interactions. To probe the influence of protein shape in controlling the nonspecific electrostatic interaction, we monitor changes in the potential of mean force upon modifying the curvature of smooth model proteins along the DNA direction (noted $C_{\parallel}$) and in the perpendicular direction ($C_{\perp}$) [see Fig. 1(a)]. The charge of all model proteins is taken as $+5e$ placed at a single site 0.7 nm under the protein surface facing DNA. The direct electrostatic force in vacuum is therefore the same for any protein shape investigated here. The DNA is modeled as a hard cylinder with divalent charged sites. The water and the electrolyte ions are described by the primitive model of electrolyte solutions [9]. The relative permittivity of water ϵ_r is taken as 78.25, and the radius of the salt ions 0.15 nm.

The potential of mean force between a protein and a DNA molecule separated by a distance L is equal to the free energy of the global system (protein, DNA, and ions in water). At a fixed surface-to-surface distance L, this energy only depends on the ion distribution. We compute the free energy using canonical Monte Carlo (MC) simulations that sample the ion configurations [10–12]. We freeze the rotational degrees of freedom of the protein, and study the interaction for the most attractive orientation, when the protein cavity points toward the DNA. This orientation is the one always observed for specific and nonspecific complexes, and we observed that the free energy becomes abruptly more repulsive when the protein rotates. The protein and DNA are placed in a parallelepipedical simulation box ($275 \times 275 \times 150$ nm) with periodic boundaries. The results are reported in Fig. 1(b).

The curvature $C_{\parallel}$ slightly influences the range of the interaction, as illustrated by the comparison of spherical and cylindrical proteins. However, the effect of the curvature $C_{\perp}$ is remarkably more pronounced. The free energy as a function of L, which is monotonic for $C_{\perp} > 0$, be-

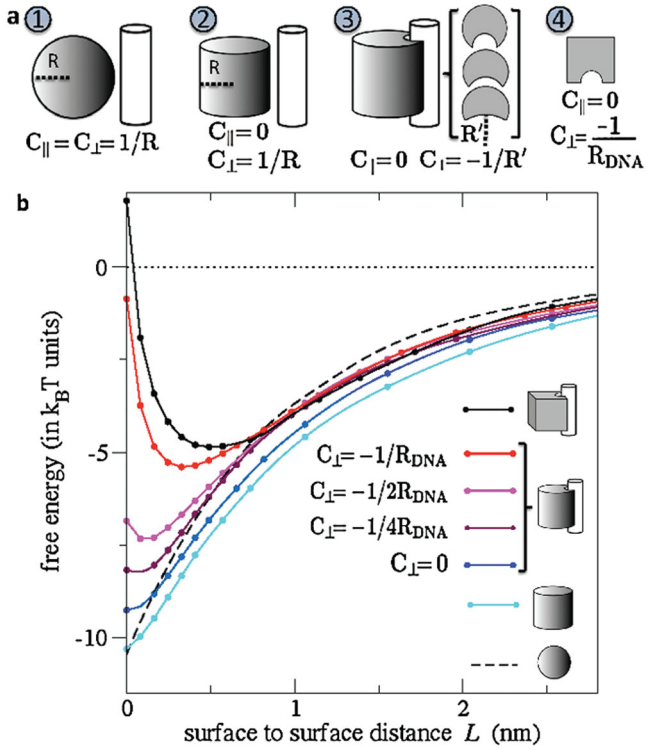

FIG. 1 (color). Influence of protein shape on the interaction. (a) Schematic view of the model proteins. The height and diameter of the cylindrical proteins (2,3) are both 5 nm, as well as the side of the cubic protein (4) and the radius of the sphere. The hollow cylindrical proteins (3) have a cylindrical cavity, of curvature $C_\perp = 0, -0.25, -0.5$, or -1 nm^{-1}. (b) Free energy of the DNA-protein systems computed by MC simulations. The protein and DNA are immersed in a monovalent salt whose Debye length $\lambda_D = 1$ nm [27] corresponds to physiological conditions. The standard deviation of the free energy is $0.2k_BT$.

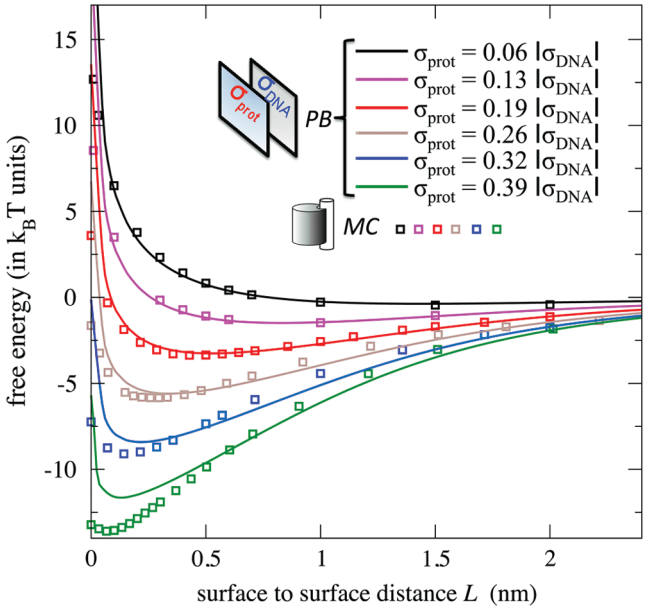

FIG. 2 (color). Influence of protein charge on the interaction. Free energy of the DNA-protein system for a set of protein charge densities obtained by PB theory (curves) and by MC simulations (squares). The area of the concave protein surface S_{int} is 15 nm^2. The charge density is $\sigma_{prot} = Z_{prot}/S_{int}$, where Z_{prot} is the charge of the protein at the interface. The charge density of DNA is $\sigma_{DNA} = -1.0e$ nm^{-2}. In the MC simulations, the shape of the model DNA-BP is a cylinder of height 5 nm, with a concave interface ($C_\perp = -1/R_{DNA}$, $C_\parallel = 0$). The protein charges are distributed on a pattern of 16 sites, 0.1 nm below the surface of the cylindrical cavity.

comes nonmonotonic for $C_\perp < 0$ and exhibits a minimum F_{min} at a distance L_{min}. For $L < L_{min}$, there is an unexpected repulsive free-energy barrier between the oppositely charged bodies, that reaches $\sim 5k_BT$ in the case of perfectly matching surfaces ($C_\perp = -1/R_{DNA}$). This behavior is weakly influenced by the shape of the remaining surface of the protein: F_{min} varies from, e.g., $-4.9k_BT$ with a cubic protein to $-5.4k_BT$ for a cylindrical one.

Once the role of the protein curvature is established, we perform simulations of concave DNA-BP models with various charge patterns to assess the influence of the protein charge on the interaction. Changing the pattern of charges while keeping the interface charge density σ_{prot} constant, the free energy exhibits only minor variations (data not shown). Conversely, changing σ_{prot} strongly modulates the free-energy profile (Fig. 2). For an interface of, e.g., 15 nm^2, if we change σ_{prot} from $0.13|\sigma_{DNA}|$ to $0.39|\sigma_{DNA}|$, F_{min} dramatically decreases from $-2k_BT$ to $-14k_BT$ and L_{min} decreases from 0.75 to 0.1 nm.

To provide a rational basis for the MC results, we carry out statistical mechanical calculations within the Poisson Boltzmann (PB) framework. The complementary interacting surfaces of the protein and the DNA are described by a minimal model of two charged parallel plates separated by a distance L. In agreement with the MC results, this model predicts a minimum of the free energy, whose depth and position can be expressed analytically [13,14]. Moreover, we introduce corrections to the plate-plate model to account for the actual curvature of protein and DNA by rescaling both the interface area S_{int} and the charge density. More precisely, the PB free energy is integrated over S_{int} after projection of each surface element on the plane orthogonal to the L axis [15]. If R and h are the radius and height of the cylindrical interface, the interaction free energy is given by

$$F(L) = \int_{-h/2}^{h/2} dx \int_{-R}^{R} dy E(L)\sqrt{1 - y^2/R^2} = E(L)S_{int}/2,$$

where $E(L)$ is the interaction free energy per unit area for two parallel plates and z is the distance between two surface elements of the curved bodies facing each other. The effective charge densities used in the PB calculation are obtained by fitting all the MC results simultaneously ($\sigma^{eff}_{DNA} \simeq 0.6\sigma_{DNA}$ and $\sigma^{eff}_{prot} \simeq 1.2\sigma_{prot}$). Despite the nanometer size of the interface, the PB results agree remarkably well with the results of the MC simulations for the concave

DNA-BP model (Fig. 2). Furthermore, the PB results shed light on the two physical mechanisms inducing an attraction and a repulsion between oppositely charged bodies. The N_+ cations and N_- anions between the plates are in equilibrium with a bulk reservoir (μVT ensemble). Here, this equilibrium displays two regimes: a counterion-dominated regime, for which the number of ions between the plates is dominated by the counterions neutralizing DNA ($N_+ \gg N_-$), and a salt-dominated regime ($N_+ - N_- \ll N_-$). It is well known that the salt-dominated regime is attractive, because salt release is favorable [16]. As expected, the ionic density decreases as the charged plates approach each other in the particular case of $\sigma_{\text{prot}} = |\sigma_{\text{DNA}}|$ (i.e., $N_+ = N_-$) representative of this regime [Fig. 3(a)]. Nevertheless, if $\sigma_{\text{prot}} < |\sigma_{\text{DNA}}|$, a constant number of neutralizing counterions remains confined between the plates in order to maintain electroneutrality. As L decreases, these cations become more and more concentrated. Below a given distance, this counterion trapping dominates the salt release (counterion-dominated regime). As shown in Fig. 3(a), the ionic density increases as L decreases for $\sigma_{\text{prot}} = -0.2\sigma_{\text{DNA}}$. The resulting enhancement of the osmotic pressure exceeds the salt-mediated attraction and results in a global repulsion.

To visualize how this mechanism applies to a more realistic interface, we compute the ionic density by MC simulations. As shown in Fig. 3(b), the two regimes are similar to those observed with the two-plate model. This highlights the significance of electroneutrality effects for the nanometric interfaces of biopolymers. Indeed, since the Debye length λ_D (i.e., the range of charge inhomogeneities in solution) is of the order of a nanometer, strong electric fields can appear locally and trap ions in a very confined space. Moreover, this physical picture explains the influence of shape complementarity: The interface is then large enough (relative to λ_D) and the gap thin enough to trap cations within a small volume.

To what extent do real DNA-BPs trap ions between their surface and DNA? To answer this question, we perform a statistical analysis on a data set of 77 proteins. The charge densities of those proteins are not directly available, but DNA-BPs are characterized by conserved propensities of residues at the interface region (see [17]). For each protein in the data set, we evaluate the total number of residues $N_{\text{tot}}^{\text{prot}}$, and the number of i residues N_i^{prot} for the charged residues (i = Arg, Lys, Asp, and Glu). References [2,17] provide $N_{\text{tot}}^{\text{int}}$, the number of residues at the interface. We estimate the charge densities of the proteins by approximating the propensity of i by $(N_i^{\text{int}}/N_{\text{tot}}^{\text{int}})/(N_i^{\text{prot}}/N_{\text{tot}}^{\text{prot}})$, and this leads to the number of i residues at the interface N_i^{int} and thus the number of charges. We take a mean interface area per residue of 0.70 nm^2 [18] to derive the mean charge density σ_{prot}. In the case of sequence-specific DNA-BPs such as transcription factors and restriction enzymes, we obtain $\sigma_{\text{prot}} = (0.17 \pm 0.03)|\sigma_{\text{DNA}}|$. We notice that the less specific DNA-BPs (polymerases, DNA-repair proteins, histones) are more charged [$\sigma_{\text{prot}} = (0.27 \pm 0.05)|\sigma_{\text{DNA}}|$]. The area of the fitting interface $S_{\text{prot}} = 15 \pm 5$ nm^2 is similar for all DNA-BPs [1]. According to these structural features, DNA-BPs should thus be repelled by DNA (cf. Fig. 2). This repulsion obtained with a coarse-grained model is in agreement with simulations of atomic models of BamHI [19], showing a repulsion when the concave surface of the protein approaches DNA.

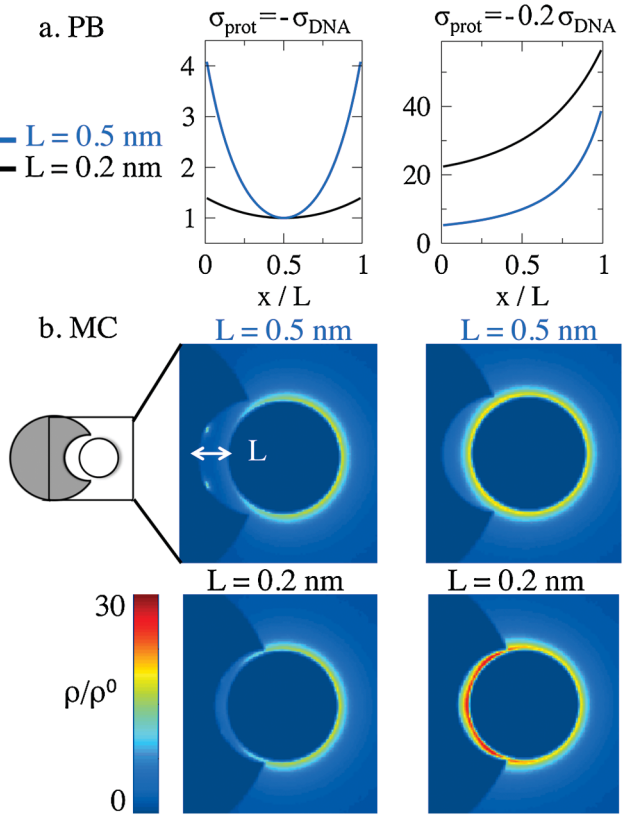

FIG. 3 (color). Ionic density fields. The density is obtained by PB theory (a), and by MC simulations (b) for two protein charge densities and two distances L. The unit is $\rho^0 = 0.2$ mol $\cdot$ L^{-1} In the PB treatment, the density along the x direction, perpendicular to the plates, is plotted ($x = 0$ on the protein and $x = L$ on DNA). In the simulations, the ionic density in the plane perpendicular to the DNA axis is plotted.

Finally, we include H-bond interactions and study the resulting free energy as a function of the protein position z along the sequence and the distance L between the surfaces. We consider a DNA-BP model of charge $\sigma_{\text{prot}} = 0.17|\sigma_{\text{DNA}}|$ with a fitting shape. We account for each H bond by a Morse potential term $V_M(L) = D[(e^{-\alpha L} - 1)^2 - 1]$ with $D = 0.5k_BT$ [20] and $\alpha = 20$ nm^{-1} [21]. Crystal structures of protein-DNA complexes provide a value of the number of H bonds n_{spec} at the specific site (30 H bonds for $S_{\text{int}} = 20$ nm^2 [2]). We assume that the number n of H bonds that the protein can make on non-specific DNA follows a Gaussian distribution of average $\langle n \rangle = n_{\text{spec}}/3$, and standard deviation $\sigma_n = \sqrt{n_{\text{spec}}}$. The

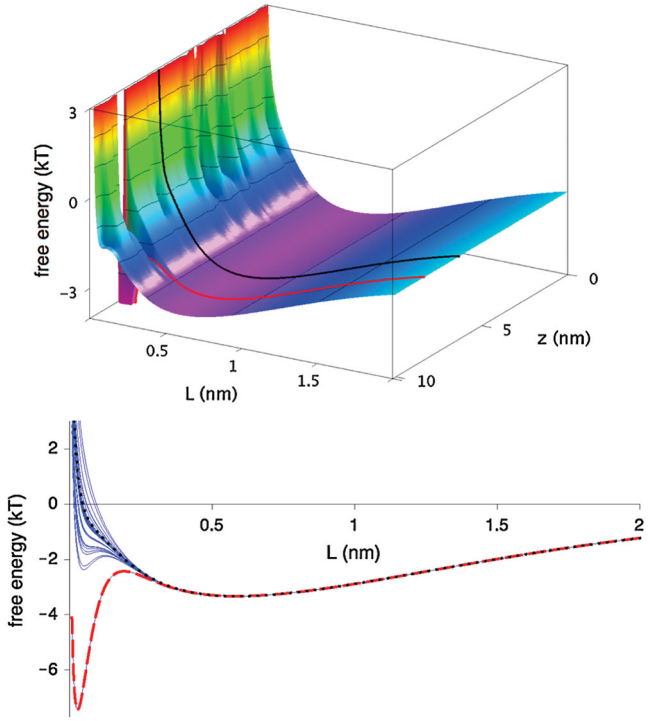

FIG. 4 (color). Free-energy landscape. The free energy is computed along a 30 bp DNA sequence, as a function of L and of the protein coordinates along DNA (z), for $\sigma_{\text{prot}} = 0.17|\sigma_{\text{DNA}}|$. The gap between level lines is k_BT. The lower graph displays the free energy as a function of L for each z value. In both graphs, the black curve corresponds to a randomly chosen nonspecific coordinate, while the red curve corresponds to the specific site.

value of $\langle n \rangle$ is low because the number of H bonds dramatically decreases for nonspecific sequences [5].

Remarkably, the osmotic repulsion between sequence-specific DNA-BPs and DNA dominates along nonspecific sequences (Fig. 4). The equilibrium gap distance of nearly 0.5 nm is in agreement with the distance observed in the complexes of EcoRV (0.51 nm [1]) with nonspecific sequences. Interestingly, along the equilibrium valley, the roughness of the sequence-dependent part of the potential is screened out: The protein can therefore easily slide along DNA. At the target site, the large H-bond interaction significantly reduces the barrier, allowing the protein to approach the DNA. Moreover, in the case of a model protein of charge $\sigma_{\text{prot}} = 0.27|\sigma_{\text{DNA}}|$ representative of weakly specific DNA-BPs, the interaction free energy is attractive whatever the position along DNA. The osmotic barrier is not high enough to counterbalance the H-bond attraction. This is consistent with the function of those proteins, which weakly depends on the DNA sequence.

From a dynamical perspective, our model provides a new reconciliation of high protein mobility and high sequence sensitivity [22–24]. The latter is usually assumed to slow down the protein diffusion [25,26]. However, according to our results, the DNA-BP freely diffuses along nonspecific DNA, confined to a valley in the free energy. The free-energy barrier, which keeps the protein at a distance from DNA, is also a signature of the sequence: Transverse thermal fluctuations enable the protein to cross the barrier only at the specific site or at highly homologous sequences. This recognition mechanism is efficient because it does not require the protein to probe the molecular details of nonspecific DNA sequences.

We gratefully acknowledge Pierre Desbiolles for stimulating discussions.

*vincent.dahirel@upmc.fr

[1] S. Jones *et al.*, J. Mol. Biol. **287**, 877 (1999).
[2] K. Nadassy, S. J. Wodak, and J. Janin, Biochemistry **38**, 1999 (1999).
[3] P. H. Von Hippel, Annu. Rev. Biophys. Biomol. Struct. **36**, 79 (2007).
[4] Y. Takeda, P. D. Ross, and C. P. Mudd. Proc. Natl. Acad. Sci. U.S.A. **89**, 8180 (1992).
[5] H. Viadiu and A. Aggarwal. Mol. Cell **5**, 889 (2000).
[6] C. G. Kalodimos *et al.*, Science **305**, 386 (2004).
[7] J. Gorman and E. C. Greene, Nat. Struct. Mol. Biol. **15**, 768 (2008).
[8] I. Bonnet *et al.*, Nucleic Acids Res. **36**, 4118 (2008).
[9] J.-P. Hansen and H. Löwen, Annu. Rev. Phys. Chem. **51**, 209 (2000).
[10] V. Dahirel *et al.*, Phys. Chem. Chem. Phys. **10**, 5147 (2008).
[11] V. Dahirel *et al.*, Phys. Rev. E **76**, 040902(R) (2007).
[12] V. Dahirel *et al.*, J. Chem. Phys. **127**, 095101 (2007).
[13] H. Ohshima, Colloid Polym. Sci. **253**, 150 (1975).
[14] F. Paillusson, M. Barbi, and J. M. Victor, arXiv:0902.1457v1.
[15] S. Bhattacharjee and M. Elimelech, J. Colloid Interface Sci. **193**, 273 (1997).
[16] D. Ben-Yaakov *et al.*, Europhys. Lett. **79**, 48002 (2007).
[17] S. Jones *et al.*, Nucleic Acids Res. **31**, 7189 (2003).
[18] E. W. Stawiski, L. M. Gregoret, and Y. Mandel-Gutfreund, J. Mol. Biol. **326**, 1065 (2003).
[19] J. Sun *et al.*, Biophys. J. **84**, 3317 (2003).
[20] D. Tareste *et al.*, Biophys. J. **83**, 3675 (2002).
[21] Y. Chen *et al.*, Nucleic Acids Res. **32**, 5147 (2004).
[22] I. Eliazar, T. Koren, and J. Klafter, J. Phys. Condens. Matter **19**, 065140 (2007).
[23] B. van den Broek *et al.*, Proc. Natl. Acad. Sci. U.S.A. **105**, 15738 (2008).
[24] O. Bénichou *et al.*, arXiv:0901.4185v2.
[25] M. Slutsky and L. A. Mirny, Biophys. J. **87**, 4021 (2004).
[26] M. Barbi *et al.*, J. Biol. Phys. **30**, 203 (2004).
[27] The Debye length λ_D is defined by $\lambda_D^{-2} = 4\pi\ell_B \sum_{i=1}^{2} \rho_i Z_i^2$, where ρ_i is the concentration of i microions, Z_i the corresponding valence, and the Bjerrum length ℓ_b equals $e^2/(4\pi\epsilon_0\epsilon_r k_B T)$.