doi:10.1016/j.jmb.2008.11.016 J. Mol. Biol. (2009) 385, 1087–1097

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Protein Sliding along DNA: Dynamics and Structural Characterization

Ohad Givaty and Yaakov Levy*

Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel

Received 24 July 2008; received in revised form 7 October 2008; accepted 11 November 2008
Available online 20 November 2008

Efficient search of DNA by proteins is fundamental to the control of cellular regulatory processes. It is currently believed that protein sliding, hopping, and transfer between adjacent DNA segments, during which the protein nonspecifically interacts with DNA, are central to the speed of their specific recognition. In this study, we focused on the structural and dynamic features of proteins when they scan the DNA. Using a simple computational model that represents protein–DNA interactions by electrostatic forces, we identified that the protein makes use of identical binding interfaces for both nonspecific and specific DNA interactions. Accordingly, in its one-dimensional diffusion along the DNA, the protein is bound at the major groove and performs a helical motion, which is stochastic and driven by thermal diffusion. A microscopic structural insight into sliding from our model, which is governed by electrostatic forces, corroborates previous experimental studies suggesting that the active site of some regulatory proteins continually faces the interior of the DNA groove while sliding along sugar–phosphate rails. The diffusion coefficient of spiral motion along the major groove of the DNA is not affected by salt concentration, but the efficiency of the search can be significantly enhanced by increasing salt concentration due to a larger number of hopping events. We found that the most efficient search comprises ~20% sliding along the DNA and ~80% hopping and three-dimensional diffusion. The presented model that captures various experimental features of facilitated diffusion has the potency to address other questions regarding the nature of DNA search, such as the sliding characteristics of oligomeric and multidomain DNA-binding proteins that are ubiquitous in the cell.

Keywords: nonspecific protein–DNA interactions; 1D diffusion; protein sliding; facilitated diffusion

Edited by D. Case

Introduction

Interactions between proteins and nucleic acids are ubiquitous and central to the life of cells. DNA-binding proteins perform many different biological tasks such as controlling transcription, unwinding DNA, repairing damaged DNA, and supercoiling DNA. The remarkable efficiency and specificity of protein–DNA recognition present a major theoretical puzzle given the size of the genome, the large number of molecular species *in vivo*, and the crowded environment they inhabit. Deciphering the molecular principles and detailed dynamics of protein interactions with nucleic acids is fundamental to understanding the network of chemical reactions in cellular regulatory processes and is expected to give meaning to raw genomics data. One of the most fundamental issues in protein–DNA recognition is how proteins achieve a high degree of selectivity among millions of competing nonspecific DNA sequences of very similar overall structures containing exquisitely fine differences in their chemical properties.

Another important, and no less perplexing, aspect of protein–DNA interactions is the rapid recognition of the DNA target sequence. A transcription factor can find its target site among 10^6–10^9 possible sites on the DNA within ~1–10 s. The association rate of a transcription factor and its operator on DNA is about 10^{10} M^{-1} s^{-1}, which is 2 orders of magnitude

*Corresponding author. E-mail address: koby.levy@weizmann.ac.il.

Abbreviations used: 3D, three-dimensional; 1D, one-dimensional.

faster than the maximal rate of a bimolecular reaction controlled by three-dimensional (3D) diffusion. A simple 3D search for the target sequence by proteins is, thus, not sufficient to resolve the discrepancy between the experimentally measured target search rates and the maximal rate allowed by diffusion. It was suggested relatively early[1,2] that the actual efficiency of the DNA search process undertaken by binding proteins is achieved not solely through random diffusion but also using additional search mechanisms such as one-dimensional (1D) sliding along the DNA,[3–7] during which the protein binds the DNA nonspecifically.[8] Recent modeling suggests that sliding roughly 100 bp is optimal for a typical protein and would speed binding by a factor of 30.[4]

Since the pioneering work of Berg *et al.*, it has become well accepted that a protein search for its target sequence comprises both 1D search (sliding) and 3D search along the DNA, whereby the protein dissociates from the DNA and reassociates at a point uncorrelated to the dissociation point.[9] In addition to these two search mechanisms, a protein can search the DNA via hopping (also known as correlated transfer) or via intersegmental transfer. In hopping, the protein dissociates from the DNA and reassociates after a short period at a point in the vicinity of the dissociation point. In intersegmental transfer, the protein directly transfers from one DNA fragment to another, presumably via a doubly bound intermediate. Figure 1 schematically shows the four different mechanisms in use during DNA search by proteins.

The structural details of nonspecific protein–DNA interactions, which govern the search for cognate-specific sites on DNA, are elusive due to their relatively low affinity. Yet, a few crystal structures of nonspecific or semispecific protein–DNA complexes have been resolved. In nonspecific protein–DNA complexes (e.g., EcoRV,[10] BamHI,[11] lac repressor,[12,13] and BstYI[14] systems), the proteins interact mostly with negatively charged phosphate groups,[15] and the DNA remains in canonical B form.[12] The importance of electrostatic interactions in dominating nonspecific protein–DNA interactions[16,17] is supported by a salt concentration dependence stronger than that in specific interactions[18] and by observations that the protein–DNA interface is much more hydrated in the nonspecific complex than in the specific complex.[19] In recent work using NMR measurements, successfully characterized the structural and kinetic aspects of the nonspecific interaction of the HoxD9 homeodomain with DNA.[20–22] They established that the HoxD9 homeodomain makes use of identical binding interfaces and orientations in both specific and nonspecific DNA interactions. This may suggest that, during sliding, the HoxD9 homeodomain performs a helical motion by probing the major groove, as was previously conjectured for EcoRI[23] and BamHI[24] description.

Recently, a few single-molecule experiments[25–27] have been designed to capture the translocation of a DNA-binding protein along a double-stranded DNA molecule. These experiments elegantly illustrated the translocation of proteins along the DNA, verified their Brownian nature, and indicated the broad distributions of the 1D diffusion coefficient (D_1) and the diffusion length. The D_1 coefficient ranges from 2.3×10^2 nm^2 s^{-1} to 1.3×10^5 nm^2 s^{-1}, which is much smaller than the 3D coefficient of about 10^8 nm^2 s^{-1}. The relationship between the 1D search and the 3D search of DNA has been recently studied for the heterodimeric restriction enzyme BbvCI to distinguish whether the translocation of a protein from one specific site to a second specific site, separated by a different number of base pairs, includes dissociations.[28] Under physiological conditions, the translocation of a protein over a distance of 30 bp or more always includes at least one dissociation event.

In the current study, we use computational tools to explore the molecular details of sliding, its driving forces, and its generalizability to various DNA-binding proteins. Furthermore, we investigate the interplay between DNA search by sliding, DNA search by hopping, and DNA search by 3D diffusion, and how each mechanism contributes to the efficiency with which the specific binding site is targeted. Due to the elusive nature of nonspecific protein–DNA interactions, which are central to the DNA search process by proteins, we use simple models in which the interactions between proteins and DNA are governed solely by electrostatic forces.

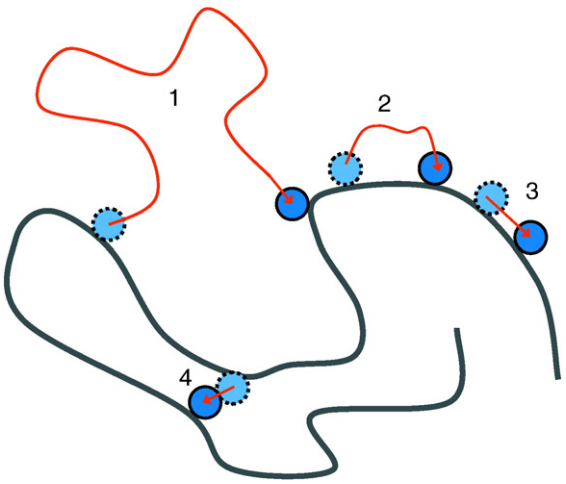

Fig. 1. Schematic illustration of the four different mechanisms by which a protein searches DNA: 3D diffusion (1); hopping or correlated transfer (2); 1D sliding (3); intersegmental transfer (4).

Results and Discussion

Structure of protein as it slides along DNA

The properties of protein sliding along B-DNA were studied for three α-helical DNA-binding proteins (Engrailed homeodomain HoxD9, Sap1, and Skn1) and for an RNA-binding protein

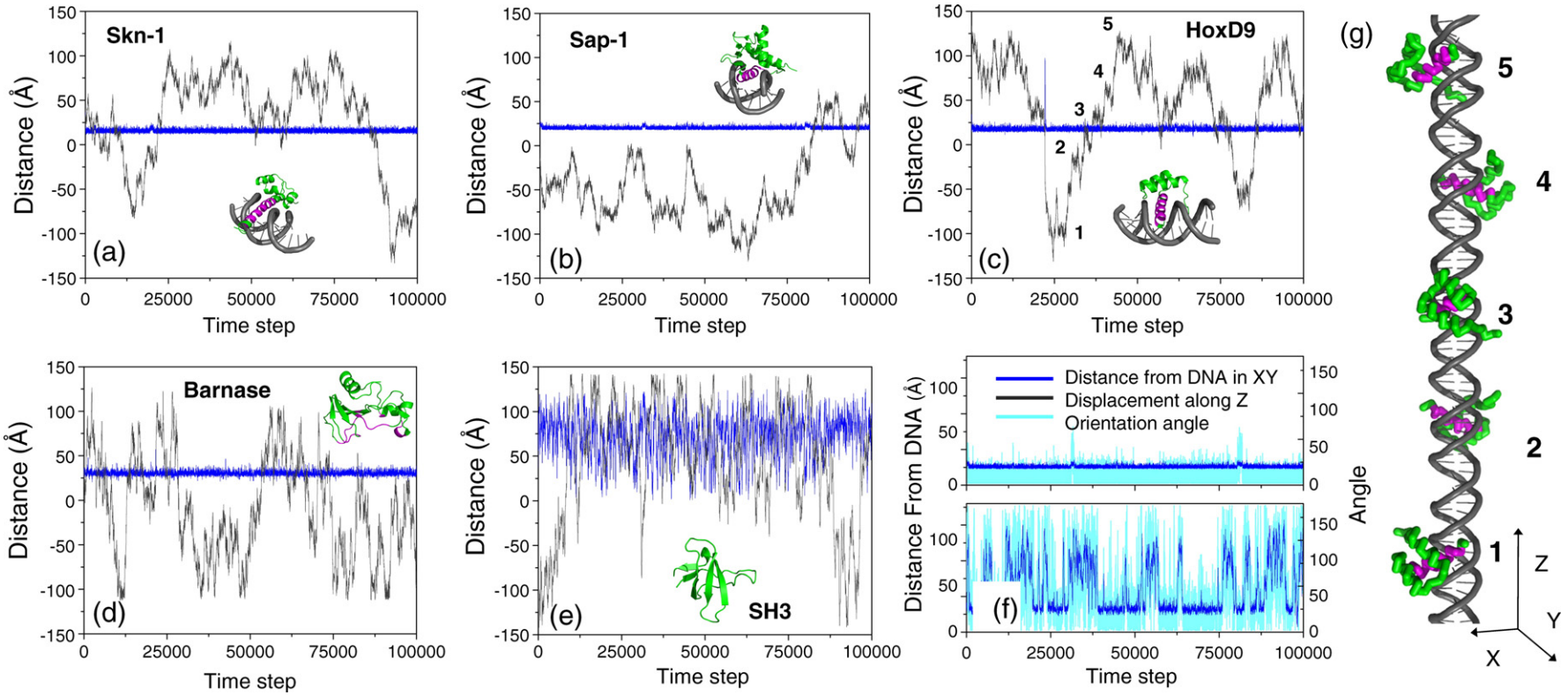

Fig. 2. Structural characteristics of protein sliding on DNA. (a–c) Typical sliding trajectories of the Skn1 (a), Sap1 (b), and HoxD9 (c) DNA-binding proteins on DNA at a salt concentration of 0.01 M and at $T \leq 0.7T_F$. Typical trajectories of the sliding dynamics of Barnase (an RNA-binding protein) and an SH3 domain (not a nucleic-acid-binding protein) are shown in (d) and (e), respectively. The blue line represents the distance of the center of mass of the protein from the main axis of the DNA in the *XY* plane. The gray line represents the translocation of the protein along the DNA (along the *Z*-axis; the center of the DNA molecules is placed at the origin). The nucleic-acid-binding proteins remain near the DNA throughout the simulation, while the SH3 domain shows no attraction to the DNA and randomly moves around it. (f) The trajectory of the Sap1 protein at two different salt concentrations: top frame, 0.01 M salt; bottom frame, 0.06 M salt. The light-blue line indicates the orientation angle of the protein towards the DNA (see Materials and Methods). Notice that at the higher salt concentration, the protein dissociates from the DNA, performs 3D diffusion, and is able to adopt various orientation angles due to the decrease in the nonspecific association constant that occurs at higher salt concentrations. (g) A graphical representation of five frames from the Engrailed homeodomain trajectory corresponding to the locations indicated in (c). This figure shows the spiral path performed by the protein in its 1D diffusion along the DNA.

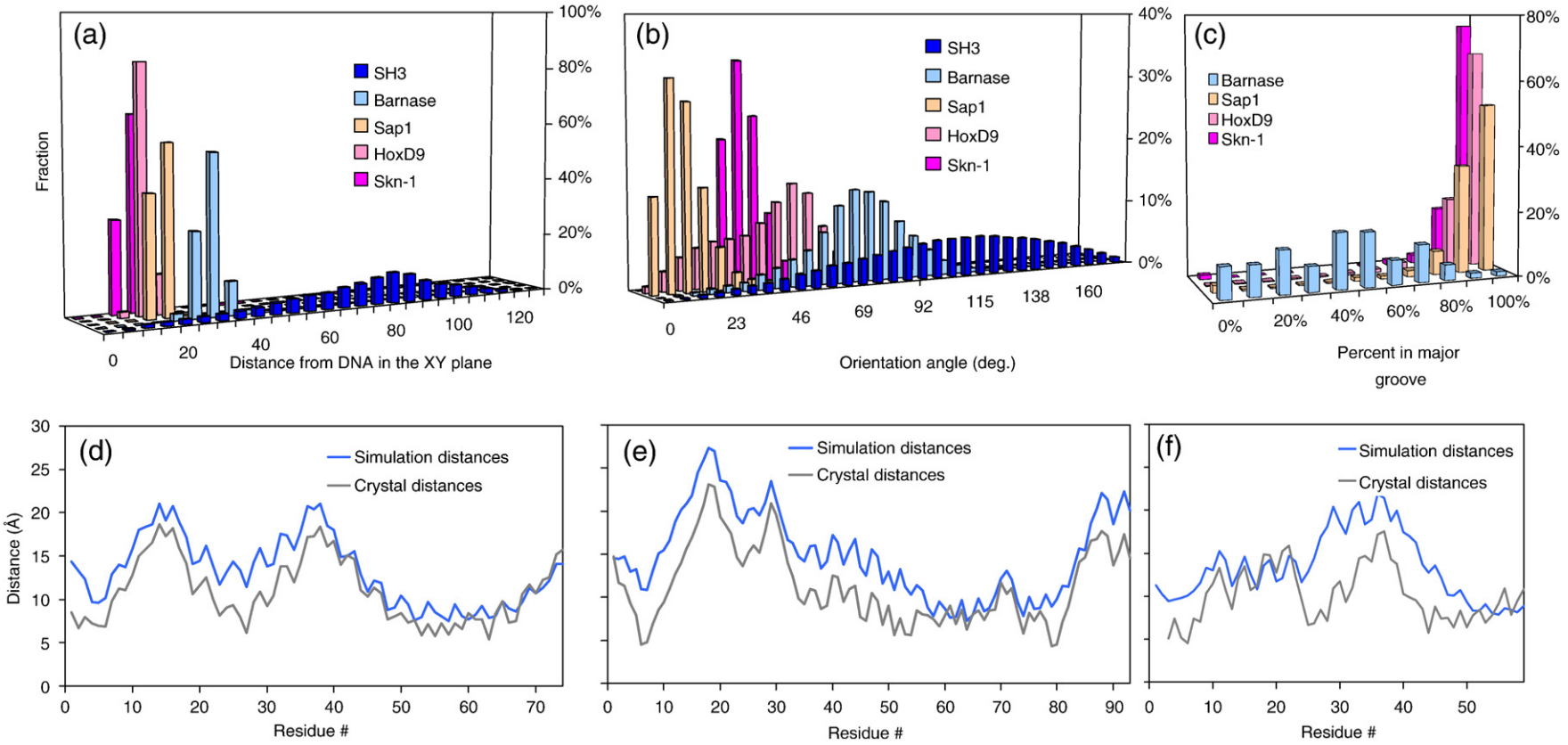

Fig. 3. Protein structural characteristics during sliding on DNA. (a) Histograms of the distances of the center of mass of the protein from the DNA in the *XY* plane. While the nucleic-acid-binding proteins are close to the nucleic acid, the SH3 domain is mostly detached from the DNA, as indicated by its wide distance distribution. (b) Histograms of the orientation angles between the protein and the DNA. The distributions of angles exhibited by the DNA-binding proteins indicate that the proteins scan the DNA using a rather unique interface. Barnase and the SH3 domain exhibit an orientation angle distribution much wider than those of other proteins, with a significant population of angles larger than 90° (~1.6 rad). (c) Percentage of recognition region situated in the major groove. During the simulation, more than 80% of the DNA-binding proteins' recognition region is located at the major groove of the DNA, while Barnase shows no clear tendency for any groove. (d–f) Average distance of each protein residue from the closest DNA bead (blue line) compared with the distances in the crystal structure of the specific complex (gray line) for Skn1 (d), Sap1 (e), and HoxD9 (f) proteins.

(Barnase). As control, we also studied the dynamics of a non-DNA-binding protein (SH3) around helical DNA. In these simulations, the 100-bp DNA molecule was kept rigid. To ensure folded proteins and to minimize protein–DNA dissociations, sliding simulations were performed at low salt concentration and at temperatures well below the protein's folding temperature.

Figure 2 presents the dynamic behavior of the five proteins with respect to their distance from the DNA (in the *XY* plane) and their displacement along the DNA molecule (along the *Z*-axis, the center of the DNA molecules is placed at the origin). The DNA-binding proteins remain close to the DNA main axis throughout the simulations while they show significant motion along the DNA, indicating that the translocation along the DNA main axis is mainly 1D and displays properties consistent with a 1D random walk driven by thermal diffusion. In contrast, the SH3 domain shows no attractions to the DNA molecule, around which it moves randomly. This reflects that electrostatic interactions, which are the only attractive force between the proteins and the DNA in our model, are sufficient to dominate nonspecific interactions. The ability of Barnase, an RNA-binding protein, to bind nonspecifically to DNA is therefore explained by the charge distribution on its surface. The distribution of the distances of each protein from the DNA (Fig. 3a) clearly illustrates that the four nucleic-acid-binding proteins are close to the DNA, while the SH3 domain is not attracted to the DNA under the same conditions.

To further quantify the structural properties of nonspecific protein–DNA interactions that are governed solely by electrostatic forces, the orientation angle of the proteins relative to the DNA was probed (see the definition of this angle in Materials and Methods). When the protein slides on the DNA at low salt concentration, the orientation angle does not change significantly; however, upon dissociation at high salt concentration, the specificity of the orientation angle is lost (Fig. 2f). The orientation angle indicates that the DNA-binding proteins interact with DNA via the recognition region that is oriented towards the DNA during most of the simulation. DNA-binding proteins accordingly exhibit a narrow distribution of angles, implying high structural preference at nonspecific binding mode (Fig. 3b). In comparison to the three DNA-binding proteins, the distribution of the orientation angle of Barnase (an RNA-binding protein) is larger, and the corresponding distribution of the SH3 domain covers all possible angle values. We note that, although the distribution of the orientation angle of the HoxD9 protein is narrow, it includes two distinct binding conformations with the DNA. While these two nonspecific binding modes could be relevant, they may also originate from the simplicity of our model.

In addition to investigating the distance and orientation of the three DNA-binding proteins relative to the DNA, we examined whether the residues constituting the binding site are located in the major groove as in their specific complex with DNA. Figure 3c shows that more than 80% of the recognition region of the three DNA-binding proteins is situated in the DNA major groove throughout the simulations. The reference protein Barnase shows no clear tendency to any groove. To probe the configuration of the DNA-binding proteins in the simulations at low salt concentration, we calculated the distance between the C^{α} bead of each protein and their closest DNA bead, and we compared the profile of the average distances during the simulation to the distances in the crystal structure of the specific complex. The distance profiles for the Sap1, Skn1, and HoxD9 proteins (Fig. 3d–f) display high similarity to those of the specific complex, indicating that nonspecific DNA association at low salt concentration is similar to that adopted for the formation of the specific complex.

Retention of recognition structural elements in sliding along the DNA was previously inferred by a CD study of Johnson *et al.*, which showed that the structure of helical DNA-binding motifs is induced by nonspecific DNA sequences.[29] This observation led the authors to propose that the lack of major

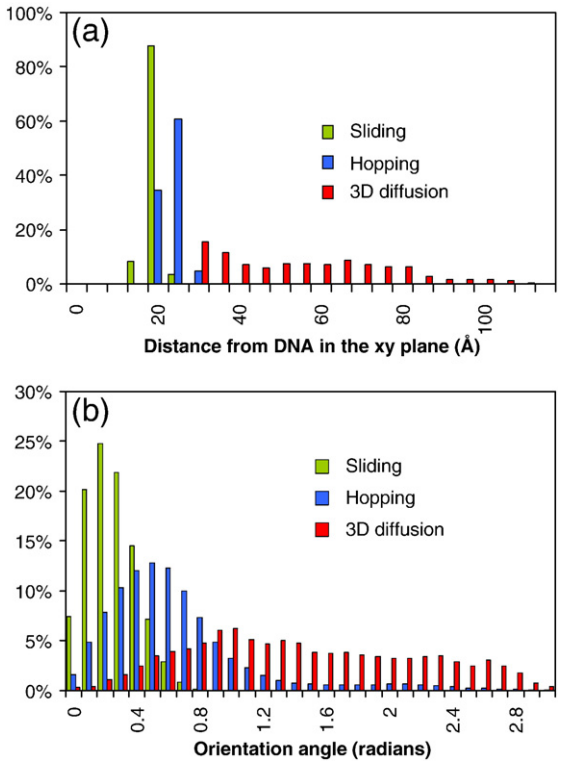

Fig. 4. Protein structural properties during hopping and 3D diffusion. The distance distributions of the center of mass of the Sap1 protein from the main axis of the DNA (a) and its orientation angle distributions (b) during sliding, hopping, and 3D diffusion. The analysis was performed on simulations sampled at a salt concentration of 0.06 M and at a temperature of $0.5T_F$. Note that, during hopping, the protein maintains a general orientation in which the binding region faces the DNA, in contrast to the nonspecific orientation of the protein during 3D diffusion.

conformational rearrangements when moving from sliding to specific recognition is essential to achieving fast binding. Recent NMR studies of Iwahara *et al.* and Iwahara and Clore support these structural features of sliding by directly showing that the HoxD9 protein binds DNA nonspecifically using the same interface as the one used in their specific complex.[20,21]

Protein–DNA structural characterization during hopping and dissociation

Characterizing the structural properties of proteins during hopping and 3D search of DNA is essential to determining the extent to which these search modes are governed by the specific protein–DNA interface found in the crystal structure. Figure 4a and b describes protein configuration relative to DNA for these two search mechanisms, during which the protein dissociates from the DNA, and during sliding. These analyses were performed on simulations performed at a salt concentration of 0.06 M and at a temperature of $0.5T_F$ of the corresponding protein (a high salt concentration is required to sample hopping and dissociation events; see Fig. 5a). During hopping, the protein adopts an orientation in which the specific binding region faces the DNA, but this orientation is much less strict than that displayed during sliding. In hopping, the protein can be viewed as performing "gliding" dynamics, where it is attracted to the DNA but does not follow the helical backbone rails as occurs during sliding. No orientation preference is displayed during 3D diffusion. Hopping durations range from a few time steps to more than 100 time steps. Most detected hopping events occur on a short timescale and span only a few DNA base pairs. Protein flipping during hopping (while maintaining its interface with the DNA) is more probable for long-lived hopping events (Fig. 6).

To address the interplay between protein flexibility and the properties of the mechanism used to search the DNA, we quantified the fraction of sliding, hopping, and 3D diffusion undertaken by Sap1 protein using a model in which the protein is completely rigid and at a temperature at which it is folded but flexible ($T=0.9T_F$). Contrary to our expectations, the usage ratios for the three DNA

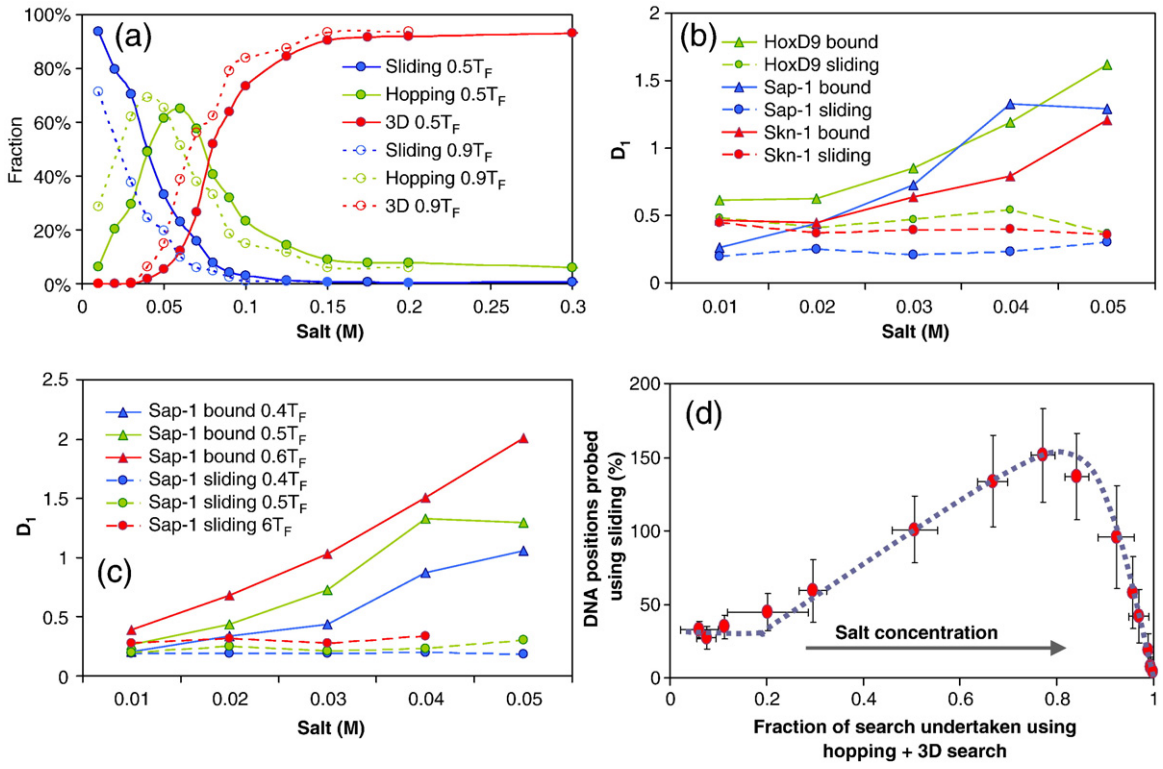

Fig. 5. The effect of salt concentration and temperature on the interplay between sliding, hopping, and 3D diffusion. (a) The proportion of use of each of the three search mechanisms for Sap1 protein at different salt concentrations and at temperatures of $0.5T_F$ (full circles) and $0.9T_F$ (empty circles). (b) Diffusion coefficient D_1 calculated for the portions of the simulation during which the protein scanned the DNA via pure sliding (circles) and for the portions of the simulation during which the protein was bound to the DNA and undertook either sliding or hopping (triangles). We see that hopping greatly increases the diffusion constant at high salt concentrations. (c) Diffusion coefficients for sliding and bound conformations of the Sap1 protein at three different temperatures. As expected, the D_1 value increases with temperature for both pure sliding and bound conformation, but with a more pronounced effect in the bound conformation, which includes hopping events whose probability increases with increasing temperature. (d) The percentage of positions probed by the protein using sliding during the simulation as a function of the fraction of nonsliding conformations adopted (i.e., as a fraction of the sum of hopping and 3D diffusion).

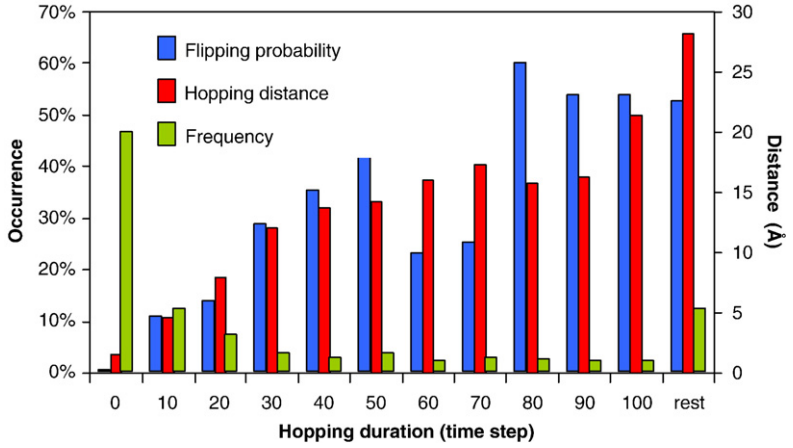

Fig. 6. Characterization of hopping dynamics. Hop instances during five simulations of 20 million steps each were binned according to the duration of the hop. Most hopping events last only a few time steps; however, longer hops have greater influence on the target search, since they propagate the protein much farther along the DNA molecule, as can be seen by the red bars. The blue bars indicate the probability of a protein to flip and change its orientation relative to the DNA for different hop durations. One can see that for hops longer than 80 steps, the probability of flipping reaches 50%.

search modes at a given salt concentration are similar for both the flexible model and the rigid model of the proteins (Fig. 7).

Efficiency of DNA search: Effects of temperature and salt concentration

The balance between DNA sliding, hopping, and 3D dissociation might be affected by either salt concentration or temperature, as both can decrease the electrostatic strength that governs nonspecific protein–DNA association. Figure 2f illustrates the different structural properties of the interactions that occur under low *versus* high salt concentration regimes; at a salt concentration of 0.01 M, the protein is close to the DNA, while at a concentration of 0.06 M, the protein is detached from the DNA during a significant part of the simulation. To characterize the dynamics of nonspecific protein–DNA recognition, we simulated the Sap1 protein at a wide range of salt concentrations and at two different temperatures (Fig. 5a). As salt concentration increases, the protein becomes more dissociated from the DNA, such that usage of the sliding search mode decreases and that of 3D diffusion increases. The decrease in sliding and the increase in 3D diffusion are accompanied by a sharp peak in the number of hopping events occurring at moderate salt concentrations. Hopping along the DNA requires that the protein not be too tightly bound at the major groove, but still be close to the DNA. In our simulation model, we found that a significant population of hopping events occurs at salt concentrations of 0.04–0.07 M. The existence of an optimal salt concentration for hopping stems from the need to balance the strength of the electrostatic attraction between the protein and the DNA. For hopping, electrostatics has to be strong enough to enable mild attraction of the protein to the DNA, but it also has to be weak enough to allow the protein to leave the major groove. At a higher temperature, the relative proportions of sliding, hopping, and dissociation at different salt concentrations are similar to the behaviors observed at a lower temperature, with the exception that the tradeoff between sliding and dissociation occurs at a lower salt concentration (Fig. 5a). Accordingly, at this high temperature ($0.9T_F$), the maximal amount of hopping is observed at a salt concentration of 0.04 M (in comparison to 0.06 M at $0.5T_F$).

The strong dependence of DNA search via sliding or hopping mode on salt concentration suggests that the 1D diffusion coefficient D_1 will also be affected by salt concentration. We estimated the values of D_1 for two different types of protein–DNA recognition: (1) the protein is bound to the DNA (i.e., engages in either sliding or hopping dynamics); and (2) the protein performs a purely helical sliding motion when its recognition region is located at the major groove. Figure 5b displays the results of this comparison (since hopping events are rather short, it is impractical to estimate the D_1 value of a hopping motion alone). The small D_1 value for the sliding motion originates from tight interactions between the protein and sugar–phosphate rails that may introduce hydrodynamic resistance into the rotational motion.[30] While the D_1 diffusion coefficient for sliding is not affected by salt concentration, the

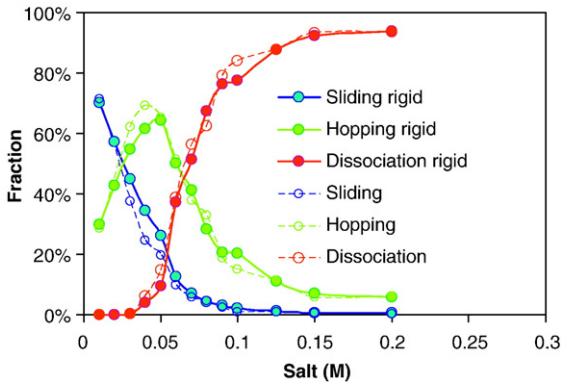

Fig. 7. The effect of protein flexibility on the interplay between sliding, hopping, and 3D diffusion. The balance between the three search mechanisms for the protein Sap1 in different salt concentrations at $0.9T_F$ (T_F is the folding temperature of flexible Sap1): empty circles represent a flexible protein with 10–12 van der Waals potentials for all native contacts, and full circles represent a rigid protein with harmonic constraints applied to all native contacts to eliminate their fluctuations.

D_1 diffusion coefficient for the combination of sliding and hopping greatly increases with salt concentration, independently of the protein system studied. This phenomenon is likely to stem from the increase in the number of hopping events that occur up to a salt concentration of 0.06 M at $T = 0.5T_F$ (Fig. 5a). During the hopping events, the protein adopts an orientation in which the specific binding region faces the DNA but its motion does not necessarily perform a helical trajectory by following the major groove of the double- stranded DNA, as occurs in sliding. Hopping, therefore, is characterized by faster dynamics. The different D_1 values of the three proteins may stem from their different geometries and electrostatic potentials (inhibiting protein flexibility does not affect the diffusion coefficient; data not shown). To examine the effect of temperature on the 1D diffusion coefficient, the D_1 coefficients for the bound and sliding states were measured at different temperatures (Fig. 5c). As expected, the D_1 coefficients increased with temperature for both pure sliding and bound conformation, but with a more pronounced effect for the bound conformation.

Our next goal was to explore search efficiency and to compare it with the prediction made by Berg *et al.* for the association rate of a specific site at different nonspecific association constants.[9] To address the efficiency of the DNA search, we define a measure, called Probed Positions, that indicates the number of new DNA sites visited by the protein during the simulation when the helical recognition site is located at the DNA's major groove. This measure is calculated by dividing the DNA into sections having a length of 3.3 Å (the length of a DNA base pair). Any frame in which the protein is situated at the DNA's major groove is added to the probed position measure, unless this position has already been scanned by the sliding mode. When the protein disassociates from the DNA, all marked positions are wiped, and the number of Probed Positions is left unchanged. Thus, the next time the DNA is probed, all positions will be unmarked. Wiping of the marked positions during dissociation is performed to simulate the fact that, upon dissociation, the probability of the protein returning to the same DNA section is negligible; thus, upon reassociation, the protein would be probing an unprobed piece of the DNA. Since our model does not include a specific site, the Probed Positions measure serves as a measure for the specific site association rate. For the purpose of recapturing the predicted behavior, we use the fraction of nonsliding instances (i.e., the percentage of instances during which the protein did not slide along the DNA in the simulation; namely, the total fraction of hopping and dissociation) as a measure of the nonspecific binding constant. To sample different sliding fractions, we simulated the system at salt concentrations varying between 0.005 M and 0.2 M. Figure 5d shows the search efficiency as a function of the salt concentration and is reminiscent of the trend predicted by Berg *et al.*[9] When the protein scans the DNA in sliding mode (at a low salt concentration), the search efficiency is limited, since the diffusion is slow when scanning the major groove. At a high salt concentration, the efficiency of the DNA search is very poor, since the electrostatic interactions are very weak and the protein mostly diffuses three-dimensionally around the DNA but rarely slides on it. At a moderate salt concentration, the DNA-binding protein searches the DNA by the three mechanisms, and this combination yields an efficient search.

In our simulation, optimal search efficiency is achieved when sliding constitutes about 20% of the total search. This result not only supports previous theoretical models[4,31] showing that an optimal search strategy requires a combination of the different search mechanisms but also indicates the importance of hopping and dissociation for an efficient search. Measurements of the translocation of the restriction enzyme BbvCI along the DNA[28] indeed occur for short distances of about 30 bp and suggest that sliding is crucial to locally scanning the DNA, while hopping and dissociation play important roles in reaching distant DNA regions.

Conclusions

Deciphering the physicochemical mechanisms that underlie the protein–DNA recognition process requires us to understand the nature of the interactions between regulatory proteins and nonspecific DNA-binding sites. These interactions dictate the speed of the DNA search and act as an intermediate step in the process of recognizing the specific DNA target. In this study, we used computational models to characterize the structural and dynamic features of the recognition of DNA by proteins during DNA search. In our simple models, the protein was flexible but the DNA remained rigid, and the protein–DNA interactions were modeled by electrostatic forces only. This modeling is in accord with structural and thermodynamic studies indicating that there are no intimate interactions between the protein–DNA interactions, and that electrostatics governs the nonspecific binding mode.

Using the molecular model, we characterized the sliding dynamics of three helical DNA-binding proteins. The sliding is bidirectional and displays properties consistent with a 1D random walk driven by thermal diffusion. In 1D sliding, the proteins perform a helical motion utilizing the same protein–DNA interface and the same protein–DNA orientation as found in the crystal structures of the specific complexes. This result is in agreement with recent NMR measurements on the HoxD9 protein indicating the signature of the specific protein–DNA recognition in nonspecific sliding.[20,21]

The efficiency of the DNA search, which can be measured by the number of new DNA sites scanned over a certain timescale, strongly depends on the strength of the electrostatic interactions. At a low salt concentration (i.e., tight nonspecific protein–DNA interactions), the protein mostly probes the

DNA's major groove, and its sliding dynamics is characterized by a small diffusion coefficient. At a higher salt concentration, the protein is more detached from the DNA, either moving away from the DNA (undertaking 3D search) or remaining in the vicinity of the DNA (undertaking a linear search via hopping). The hopping search mode is accompanied by a higher diffusion coefficient and enhances the efficiency of DNA scanning. At a very high salt concentration, the efficiency of the DNA search by proteins diminishes as the protein mostly engages in 3D diffusion. Accordingly, we suggest an optimal salt concentration at which the DNA-binding protein combines sliding, hopping, and 3D diffusion to search the DNA target. This optimal interplay between the various search mechanisms of DNA may be different for proteins having different binding affinities for DNA.

While we show that key structural and dynamic properties of protein sliding can be studied by relating only to the electrostatic interactions between proteins and DNA, it is clear that higher-resolution models are essential to understanding the role of water molecules in sliding and in the transition from nonspecific to specific recognition,[32,33] which is supposed to be accompanied by the release of water molecules from the protein–DNA interface. Furthermore, the role of DNA flexibility[34–37] in sliding and the generalizability of the sliding properties reported here for the helical proteins to other DNA-binding proteins have to be explored in the future to achieve a better understanding of regulatory processes at the molecular level.

Materials and Methods

The molecular and dynamic nature of the protein search of DNA was studied using a reduced model that allows long timescales to be simulated and thus captures sliding, hopping, and 3D diffusion (intersegmental transfer cannot be studied in the current study, since the simulations are performed using a single straight DNA). We modeled the DNA with three beads per nucleotide, representing phosphate, sugar, and base. Each bead was located at the geometric center of the group it represented. The protein was represented by one bead for each residue located at the C^{α} of that residue. Beads representing charged amino acids (Arg, Lys, Asp, and Glu) and the DNA phosphate groups were charged in the model. In the simulations, a 100-bp-long B-DNA molecule was used. The protein and the DNA were placed in a box with dimensions 200 Å×200 Å×500 Å, and the DNA was placed at the center of the box along its *Z*-axis. While the DNA remained frozen throughout the simulations, the protein was flexible and could undergo folding and unfolding events. The protein was simulated by a native topology-based model that excludes nonnative interactions (details of the Hamiltonian of the protein can be found in previous publications[38–40]). The dynamics of the systems was simulated using the Langevin equation:[41]

$$m_i \dot{v}_i = F_i - \gamma m_i v_i + R_i(t) \qquad (1)$$

where m_i, v_i, and $\dot{v}_i$ are the mass, velocity, and acceleration of the ith bead, respectively. F_i is the force applied on the ith bead, and R_i is a Gaussian stochastic variable with zero mean and with variance $\langle R_i(t)R_j(t+\tau)\rangle = 2m_i\gamma k_B T\delta(\tau)$. In this study, γ was set to 0.01. We note that the random walk motion of the proteins during sliding along the DNA is not achieved when random noise is excluded from the simulations (i.e., the dynamics is studied using the Newton equation).

The electrostatic interactions were represented by the Debye–Huckel model, which mimics the effects of salt concentration:[42]

$$U_{\text{Debye-Huckel}} = K_{\text{Coulomb}} B(\kappa) \sum_{i,j} \frac{q_i q_j \exp(-\kappa r_{ij})}{\varepsilon r_{ij}} \qquad (2)$$

The Debye–Huckel theory predicts the range of electrostatic influences of an ion to be the Debye screening length κ^{-1}. Linearization of the Poisson–Boltzmann equation yields the following relation $\kappa^2 = \frac{8\pi N_A e^2 \rho_A}{1000\varepsilon k_B T}$, where N_A is Avogadro's number, ρ_A is the solvent density, e is the proton charge, ε is the solvent dielectric constant, and C_s is the ionic concentration in molar units ($C_s = 0.5\sum_{i=1}^{n_i} c_i q_i^2$, where C_i is the molar concentration of ion with charge q_i). For monovalent salt at room temperature and with $\varepsilon = 80$, $\kappa \approx 0.32\sqrt{C_s}$ Å^{-1}. In Eq. (2), q_i is the point charge of the ith bead, r_{ij} is the distance between two charged beads (with either same or opposite charges), and $K_{\text{Coulomb}} = 4\pi\varepsilon_0 = 332$-kcal mol^{-1}. Due to the simplicity of our model and, in particular, the large distances between the charged beads of the protein and the charged beads of the DNA (since, in the reduced model, not all atoms are represented and the charges are placed at the phosphate and C^{α} beads), the salt concentrations reported in this study are several fold smaller than the experimental ionic strength. $B(\kappa)$ is a salt-dependent coefficient that equals $\exp(\kappa a_r)/(1+\kappa a_r)$, where a_r is the ion radius [for dilute solutions, $B(\kappa) \approx 1$]. The sliding dynamics was explored at a salt concentration in the range of 0.01–0.3 M and using a dielectric constant of 80. The Debye–Huckel model has been previously used to study the energetics and dynamics of various biomolecular systems such as RNA folding,[43] conformational stability of long DNA,[44] chromatin assembly,[45] protein–DNA binding,[46] and DNA–DNA association.[47] While the Debye–Huckel model is powerful in introducing the salt effect of screening electrostatic interactions to the Coulomb potential, one should be aware of its approximations.[48] The Debye–Huckel model is valid for low salt concentration, as it approximates that the potential energy of an ion is determined by pairwise interactions with other ions, which is valid mainly for dilute solution. The detailed effects of higher salt concentrations and of ion condensation on DNA have to be studied using the nonlinearized Poisson–Boltzmann equation, as well as atomistic simulations[49] that can elucidate the dynamics of the ionic layer during sliding on the DNA.

In addition to the electrostatic forces between all charged residues and phosphate beads, each bead of the protein had a repulsion potential from all of the DNA beads modeled by (σ_{ij}/r_{ij}),[12] where σ_{ij} equals 5.7 Å (the shortest distance between C^{α} atoms and the DNA beads found in a survey of crystallographic protein–DNA complexes when an α-helix served as the recognition helix). In addition, a larger repulsion distance between the protein and the 8-Å DNA, which mimics the existence of a water layer in the protein–DNA interface, was tested. A similar sliding behavior was observed for the two repulsion distances.

We have studied the sliding of three DNA-binding proteins Engrailed homeodomain HoxD9 (1HDD), Sap1 (1BC8), and Skn1 (1SKN), as well as an RNA-binding protein Barnase (1BRS). As control, we studied the dynamics of the tyrosine kinase SH3 domain (1SRL), which is not a DNA-binding protein, in the vicinity of a DNA molecule. The dynamics of each protein on long DNA was studied at various temperatures and at different salt concentrations. All the temperatures in this work are reported in terms of the folding temperature T_F of the studied protein (T_F is the temperature at which the free energies of the unfolded and folded states are equal).

Trajectories were used to explore the following structural and dynamic features of protein sliding on DNA:

1. Percentage of recognition via major groove: To determine whether a protein bead was located in the DNA's major groove, we first found the closest phosphates from each DNA chain. If the distance between these two phosphate beads fitted the characteristic distance of a B-DNA's major groove (larger than 15 Å), this protein bead was assigned to be in the major groove. The percentage of the protein's recognition region that was located in the major groove was estimated by applying the described procedure to all the protein beads.
2. Orientation angle of the protein relative to the DNA: Orientation angle to DNA was defined as the angle between the geometric center of the recognition region of the protein, the geometric center of the protein, and the point on the DNA's main (Z) axis perpendicular to the protein's geometric center.
3. Diffusion coefficient D_1: Qian *et al.* derived an expression with which to evaluate whether a diffusion trajectory is Brownian and to extract the diffusion constant D_1 of this dynamics.[50] This method calculates the mean square displacement MSD(n,N) for all time intervals of a single diffusion trajectory:

$$\mathrm{MSD}(n,N) = \sum_{i=1}^{N-n} \frac{(Z_{i+n} - Z_i)^2}{N-n} = 2D_1 n \Delta t \quad (3)$$

where N is the number of time steps measured, n is the measurement window ranging from 1 to N, Δt is the time interval between two consecutive steps, and D_1 is the 1D diffusion coefficient. In Eq. (3), the projected distance along the Z-axis is used. For Brownian diffusion, the MSD (n,N) at n below a cutoff n_c will be linear with a slope of $2D_1\Delta t$.

Structural classification of protein sliding, hopping, and 3D diffusion

To differentiate *in silico* between protein sliding, hopping (transient dissociations after which the protein is likely to reassociate in the vicinity of the dissociation point), and free 3D diffusion, it is necessary to provide a clear definition of each of them (we note that although we categorize the dynamics of the protein relative to the DNA as performing sliding, hopping, or 3D search, one can imagine a continuous transition from sliding to hopping, or from hopping to 3D search). We defined a simulation frame as being part of 3D diffusion if the protein was farther than 32 Å from the main DNA axis, since, at this distance, the average electrostatic energy drops to about 2% of the energy in the sliding conformations at low salt concentration. A snapshot was defined as being part of the sliding search procedure if three criteria were met: at least 70% of its recognition region was in contact with the correct groove, the distance of the center of mass of the recognition region from the DNA was up to 10 Å longer than that in the crystal structure, and the orientation angle was less than 90°. If the protein was at a distance of less than 32 Å from the DNA and did not meet the criteria for the sliding mode, the frame was defined as representing protein hopping along the DNA.

Acknowledgements

We are grateful to Peter von Hippel for reading of the manuscript prior for its publication. This work was supported, in part, by the Kimmelman Center for Macromolecular Assemblies, the Center for Complexity Science (Y.L.), and the Foundation Fernande et Jean Gaj. Y.L. is the incumbent Lilian and George Lyttle Career Development Chair.

References

1. Adam, G. & Delbruck, M. (1968). In *Structural Chemistry and Molecular Biology* (Rich, A. & Davidson, N., eds), W. H. Freeman, San Francisco.
2. Richter, P. H. & Eigen, M. (1974). Diffusion controlled reaction-rates in spheroidal geometry—application to repressor–operator association and membrane-bound enzymes. *Biophys. Chem.* **2**, 255–263.
3. von Hippel, P. H. & Berg, O. G. (1989). Facilitated target location in biological-systems. *J. Biol. Chem.* **264**, 675–678.
4. Halford, S. E. & Marko, J. F. (2004). How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* **32**, 3040–3052.
5. Bruinsma, R. F. (2002). Physics of protein–DNA interaction. *Phys. A*, **313**, 211–237.
6. Dahlquist, F. W. (2006). Slip sliding away: new insights into DNA–protein recognition. *Nat. Chem. Biol.* **2**, 353–354.
7. Pingoud, A. & Wende, W. (2007). A sliding restriction enzyme pauses. *Structure*, **15**, 391–393.
8. von Hippel, P. H. (1994). Protein–DNA recognition: new perspectives and underlying themes. *Science*, **263**, 769–770.
9. Berg, O. G., Winter, R. B. & von Hippel, P. H. (1981). Diffusion-driven mechanisms of protein translocation on nucleic acids: 1. Models and theory. *Biochemistry*, **20**, 6929–6948.
10. Winkler, F. K., Banner, D. W., Oefner, C., Tsernoglou, D., Brown, R. S., Heathman, S. P. *et al.* (1993). The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *EMBO J.* **12**, 1781–1795.
11. Viadiu, H. & Aggarwal, A. K. (2000). Structure of BamHI bound to nonspecific DNA: a model for DNA sliding. *Mol. Cell*, **5**, 889–895.
12. Kalodimos, C. G., Biris, N., Bonvin, A. M., Levandoski, M. M., Guennuegues, M., Boelens, R. & Kaptein, R. (2004). Structure and flexibility adaptation in nonspecific and specific protein–DNA complexes. *Science*, **305**, 386–389.

13. von Hippel, P. H. (2004). Biochemistry. Completing the view of transcriptional regulation. *Science*, **305**, 350–352.
14. Townson, S. A., Samuelson, J. C., Bao, Y., Xu, S. Y. & Aggarwal, A. K. (2007). BstYI bound to noncognate DNA reveals a "hemispecific" complex: implications for DNA scanning. *Structure*, **15**, 449–459.
15. Mossing, M. C. & Record, M. T. (1985). Thermodynamic origins of specificity in the lac repressor-operator interaction: Adaptability in the recognition of mutant operator sites. *J. Mol. Biol.* **186**, 295.
16. Savelyev, A. & Papoian, G. A. (2006). Electrostatic, steric, and hydration interactions favor Na(+) condensation around DNA compared with K(+). *J. Am. Chem. Soc.* **128**, 14506–14518.
17. Misra, V. K., Hecht, J. L., Yang, A. S. & Honig, B. (1998). Electrostatic contributions to the binding free energy of the lambda cl repressor to DNA. *Biophys. J.* **75**, 2262–2273.
18. Record, M. T., Ha, J. H. & Fisher, M. A. (1991). Analysis of equilibrium and kinetic measurements to determine thermodynamic origins of stability and specificity and mechanism of formation of site-specific complexes between proteins and helical DNA. *Methods Enzymol.* **208**, 291–343.
19. Jen-Jacobson, L. (1997). Protein–DNA recognition complexes: conservation of structure and binding energy in the transition state. *Biopolymers*, **44**, 153–180.
20. Iwahara, J., Zweckstetter, M. & Clore, G. M. (2006). NMR structural and kinetic characterization of a homeodomain diffusing and hopping on nonspecific DNA. *Proc. Natl Acad. Sci. USA*, **103**, 15062–15067.
21. Iwahara, J. & Clore, G. M. (2006). Detecting transient intermediates in macromolecular binding by paramagnetic NMR. *Nature*, **440**, 1227–1230.
22. Iwahara, J. & Clore, G. M. (2006). Direct observation of enhanced translocation of a homeodomain between DNA cognate sites by NMR exchange spectroscopy. *J. Am. Chem. Soc.* **128**, 404–405.
23. Jeltsch, A., Alves, J., Wolfes, H., Maass, G. & Pingoud, A. (1994). Pausing of the restriction endonuclease EcoRI during linear diffusion on DNA. *Biochemistry*, **33**, 10215–10219.
24. Sun, J., Viadiu, H., Aggarwal, A. K. & Weinstein, H. (2003). Energetic and structural considerations for the mechanism of protein sliding along DNA in the nonspecific BamHI–DNA complex. *Biophys. J.* **84**, 3317–3325.
25. Wang, Y. M., Austin, R. H. & Cox, E. C. (2006). Single molecule measurements of repressor protein 1D diffusion on DNA. *Phys. Rev. Lett.* **97**, 048302.
26. Graneli, A., Yeykal, C. C., Robertson, R. B. & Greene, E. C. (2006). Long-distance lateral diffusion of human Rad51 on double-stranded DNA. *Proc. Natl Acad. Sci. USA*, **103**, 1221–1226.
27. Shimamoto, N. (1999). One-dimensional diffusion of proteins along DNA—its biological and chemical significance revealed by single-molecule measurements. *J. Biol. Chem.* **274**, 15293–15296.
28. Gowers, D. M., Wilson, G. G. & 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**, 15883–15888.
29. Johnson, N. P., Lindstrom, J., Baase, W. A. & 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**, 4840–4844.
30. Schurr, J. M. (1979). The one-dimensional diffusion coefficient of proteins absorbed on DNA. Hydrodynamic considerations. *Biophys. Chem.* **9**, 413–414.
31. Slutsky, M. & Mirny, L. A. (2004). Kinetics of protein–DNA interaction: facilitated target location in sequence-dependent potential. *Biophys. J.* **87**, 4021–4035.
32. Fuxreiter, M., Mezei, M., Simon, I. & Osman, R. (2005). Interfacial water as a "hydration fingerprint" in the noncognate complex of BamHI. *Biophys. J.* **89**, 903–911.
33. Levy, Y. & Onuchic, J. N. (2006). Water mediation in protein folding and molecular recognition. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 389–415.
34. Hu, T., Grosberg, A. Y. & Shklovskii, B. I. (2006). How proteins search for their specific sites on DNA: the role of DNA conformation. *Biophys. J.* **90**, 2731–2744.
35. Zhang, Y., Xi, Z., Hedge, R. S., Shakked, Z. & Crothers, D. M. (2004). Predicting indirect readout effects in protein–DNA interactions. *Proc. Natl Acad. Sci. USA*, **101**, 8337–8341.
36. Siggers, T., Silkov, T. & Honig, B. (2005). Bending in the right direction. *Structure*, **13**, 1400–1401.
37. Rohs, R., Sklenar, H. & Shakked, Z. (2005). Structural and energetic origins of sequence-specific DNA bending: Monte Carlo simulations of Papillomavirus E_2-DNA binding sites. *Structure*, **13**, 1499–1509.
38. Clementi, C., Nymeyer, H. & Onuchic, J. N. (2000). Topological and energetical factors: what determines the structural details of the transition state ensemble and "en-route" intermediate for protein folding? An investigation of small globular proteins. *J. Mol. Biol.* **298**, 937–953.
39. Levy, Y., Cho, S. S., Onuchic, J. N. & Wolynes, P. G. (2005). A survey of flexible protein binding mechanisms and their transition states using native topology based energy landscapes. *J. Mol. Biol.* **346**, 1121–1145.
40. Levy, Y., Onuchic, J. N. & Wolynes, P. G. (2007). Fly-casting in protein–DNA binding: frustration between protein folding and electrostatics facilitates target recognition. *J. Am. Chem. Soc.* **129**, 738–739.
41. Mor, A., Ziv, G. & Levy, Y. (2008). Simulations of proteins with inhomogeneous degrees of freedom: the effect of thermostats. *J. Comput. Chem.* **29**, 1992–1998.
42. Schlick, T. (2000). Molecular Modeling and Simulation: An Interdisciplinary Guide Springer, New York.
43. Hyeon, C. & Thirumalai, D. (2005). Mechanical unfolding of RNA hairpins. *Proc. Natl Acad. Sci. USA*, **102**, 6789–6794.
44. Schlick, T., Li, B. & Olson, W. K. (1994). The influence of salt on the structure and energetics of supercoiled DNA. *Biophys. J.* **67**, 2146–2166.
45. Beard, D. A. & Schlick, T. (2001). Computational modeling predicts the structure and dynamics of chromatin fiber. *Structure*, **9**, 105–114.
46. Beard, D. A. & Schlick, T. (2001). Modeling salt-mediated electrostatics of macromolecules: the discrete surface charge optimization algorithm and its application to the nucleosome. *Biopolymers*, **58**, 106–115.
47. Savelyev, A. & Papoian, G. A. (2007). Inter-DNA electrostatics from explicit solvent molecular dynamics simulations. *J. Am. Chem. Soc.* **129**, 6060.
48. Levin, Y. (2002). Electrostatic correlations: from plasma to biology. *Rep. Prog. Phys.* **65**, 1577–1632.
49. Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A. & Case, D. A. (1998). Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate–DNA helices. *J. Am. Chem. Soc.* **120**, 9401–9409.
50. Qian, H., Sheetz, M. P. & Elson, E. L. (1991). Single particle tracking. Analysis of diffusion and flow in two-dimensional systems. *Biophys. J.* **60**, 910.