doi:10.1016/j.jmb.2009.11.056 J. Mol. Biol. (2010) 396, 674–684

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Searching DNA via a "Monkey Bar" Mechanism: The Significance of Disordered Tails

Dana Vuzman, Ariel Azia and Yaakov Levy*

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

Received 23 October 2009; received in revised form 23 November 2009; accepted 24 November 2009
Available online
1 December 2009

The search through nonspecific DNA for a specific site by proteins is known to be facilitated by sliding, hopping, and intersegment transfer between separate DNA strands, yet the driving forces of these protein dynamics from the molecular perspective are unclear. In this study, molecular features of the DNA search mechanism were explored for three homologous proteins (the HoxD9, Antp, and NK-2 homeodomains) using a simple computational model in which protein–DNA interactions are represented solely by electrostatic forces. In particular, we studied the impact that disordered N-terminal tails (N-tails), which are more common in DNA-binding proteins than in other proteins, have on the efficiency of DNA search. While the three homeodomain proteins were found to use similar binding interfaces in specific and nonspecific interactions with DNAs, their different electrostatic potentials affect the nature of their sliding dynamics. The different lengths and net charges of the N-tails of the homeodomains affect their motion along the DNA. The presence of an N-tail increases sliding propensity but slows linear diffusion along the DNA. When the search is performed in the presence of two parallel DNA molecules, a direct transfer, which is facilitated by the protein tail, from one nonspecific DNA to another occurs. The tailed proteins jump between two DNA molecules through an intermediate in which the recognition helix of the protein is adsorbed to one DNA fragment and the N-tail is adsorbed to the second, suggesting a "monkey bar" mechanism. Our study illustrates how the molecular architecture of proteins controls the efficiency of DNA scanning.

Keywords: nonspecific protein–DNA interactions; sliding; homeodomain proteins; intersegment transfer; disordered tails

Edited by M. Levitt

Introduction

One of the most fundamental aspects of protein–DNA recognition is how proteins find their binding sites so rapidly given the millions of competing nonspecific sites on the DNA and the packed structure of the DNA in the cell. The association rate of a transcription factor and its operator on the DNA is two orders of magnitude faster than the maximal rate of a bimolecular reaction controlled by three-dimensional (3D) diffusion.[1] Following the study of Berg *et al.*,[2] it has become widely accepted that the search by a protein for its target DNA sequence is achieved via facilitated diffusion comprising four mechanisms: 1D sliding, hopping (also known as correlated transfer), 3D search, and intersegment transfer.

The notion of the facilitated target-search mechanism whereby diffusion occurs in a lower dimensionality space[1,3–7] was originally developed from both theoretical and experimental perspectives[2,8] and was later supported by various bulk experiments that provide indirect support for the concept of linear diffusion on DNA and for the importance of the combination of 1D and 3D diffusion to speed protein–DNA binding. Furthermore, in recent years, several studies at the single-molecule level have visualized the 1D diffusion of protein along DNA.[9] For example, sliding along DNA has been directly observed for RNA polymerase,[10] the *lac* repressor,[11,12] DNA repairs,[13] and the p53 transcription factor.[14] While the characteristics of

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

Abbreviations used: N-tail, N-terminal tail; 3D, three-dimensional; DNAB, DNA binding; NDNAB, non-DNA binding; WT, wild type.

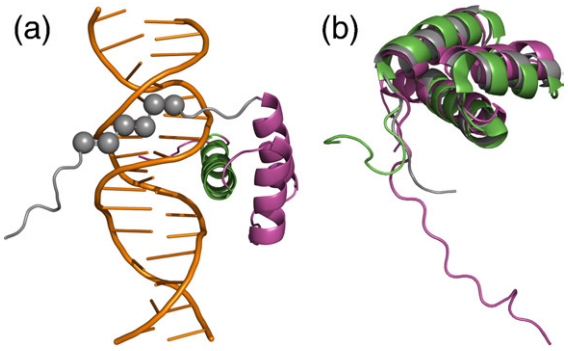

Fig. 1. The complex between a homeodomain and its specific DNA and the diversity found in the disordered tails of homologous homeodomains. (a) Complex of the NK-2 homeodomain with its specific DNA site. The 17-residue N-tail appears in gray, and the six positively charged residues of the tail, which constitute a segment of six consecutive positive charges, are shown as spheres. The recognition helix appears in green. (b) Alignment of the structures of the HoxD9 (gray) [Protein Data Bank (PDB) ID 1HDD], Antp (green) (PDB ID 1AHD), and NK-2 (magenta) (PDB ID 1NK2) homeodomains (RMSD = 1.24, 1.28, and 1.46 Å for Antp–HoxD9, Antp–NK-2, and NK-2–HoxD9, respectively). HoxD9, Antp, and NK-2 have N-tails of 9, 10, and 17 residues, respectively.

protein diffusion on DNA (e.g., the value of the diffusion coefficient and the existence of jumping or hopping events) obtained from ensemble and single-molecule measurements are consistent with the facilitated diffusion model, many aspects of the molecular characteristics of protein search on DNA are unclear. In particular, little is known about the interplay between the structural architecture of proteins and the efficiency of the DNA search. One may ask whether the dimensions and structural details of the recognition region on the proteins that directly interact with the DNA in the specific binding interaction affect the mode and efficiency of the DNA search when the proteins interact with the DNA nonspecifically. One may also hypothesize that the nature of the search will be different for tandem-linked domains compared with that for separated domains that independently diffuse on the DNA.

Recent works on sliding, hopping, and 3D search have started to provide insights into the structural and molecular properties of the four complementary search mechanisms,[15,16] as well as the interplay between them.[15,17] An extensive computational study of the structural and dynamic features of proteins on a single DNA molecule using a coarse-grained model[16] showed that sliding is a stochastic motion during which the proteins follow a helical path along the major groove. During sliding, the protein makes use of the interface that defines its specific binding to DNA. The spiral motion of the protein along the phosphate–sugar rails was hypothesized about 30 years ago by Schurr[18] and recently was elegantly confirmed by a series of NMR studies of the HoxD9 protein by Clore *et al.*[15,19,20] While the sliding motion is unambiguous and distinctive, the hopping dynamics is less strictly characterized. Often, hopping is described as consisting of dissociation events of a short lifetime in which the protein reassociates with the DNA at a point in the vicinity of the dissociation position, whereas in 3D search, the protein reassociates at a point uncorrelated with the original location of the dissociation event. During hopping, as in sliding, the protein faces the DNA, but its orientation relative to the DNA is more diverse as it is not placed at the major groove.[16] Not being helically constrained, linear diffusion along the DNA is therefore faster during hopping than during sliding, when the dynamics is restricted by the backbone rails. One can therefore say that a protein "glides" along the DNA during hopping.

The existence of intersegment transfer has been confirmed by a number of *in vitro* experiments for various proteins, such as the homeodomain HoxD9,[19] the glucocorticoid receptor GRdbd,[21] and the multidomain Oct-1,[22] which can adsorb two segments of DNA, allowing the protein to transiently form a double-bound state with two DNA fragments. This double-bound state breaks up faster than the dissociation of protein to bulk from the fully formed protein–DNA complex, and the protein has a chance to be transferred to the newly adsorbed DNA. Intersegment transfer by *Escherichia coli* RNA polymerase along DNA has been directly observed through tapping-mode scanning force microscopy.[10] It has been suggested that intersegment transfer significantly accelerates the search for a specific target site on DNA under conditions where the protein is adsorbed to the DNA most of the time, as it is *in vivo*.[23–25] However, not much is known about the molecular components that dictate the mechanism and efficiency of intersegment transfer.

To explore the effects of the molecular architecture of DNA-binding (DNAB) proteins and in particular the presence of disordered regions[26–31] on the mechanism of DNA search, in this study, we focused on three homeodomain proteins (Fig. 1). Homeodomains are conserved DNAB motifs that are associated with versatile physiological functions.[32] They play a central role in transcription regulation[22,33] and thus in many developmental processes.[34] They all exhibit a similar fold, which consists of an approximately 60-amino-acid helix–turn–helix structure that is preceded by a flexible N-terminal region[35,36] (Fig. 1). Most DNA interactions are established by the "recognition helix" that forms specific contacts with several base pairs in the major groove.[37]

The N-terminal tail (N-tail) of homeodomains is highly charged and is disordered in solution. Binding of the N-tail in the minor groove induces its folding, although it may also remain flexible or partly disordered in the complex.[38] Several independent pieces of evidence point to the fact that the disordered N-tails of homeodomains contribute to specific DNA sequence selectivity.[38–42] However, only a few hypotheses concerning the mechanism by

Table 1. Statistics of disordered tails in human DNAB proteins, human NDNAB proteins, and a random group (RAND) of proteins of the same size as the DNAB proteins that was selected from all the proteins in the previous two groups (ALL)

	Total	Proteins with no tail	Proteins with tails	Proteins with two tails as percentage of tailed proteins
ALL	20,334	48.4%	51.6%	10.2%
DNAB	864	**33.2%**	**66.8%**	**20.7%**
		<1.16e−09	*>4.14e−09*	*>3.43e−21*
NDNAB	19,470	49.1%	50.9%	9.7%
RAND	864	50.6%	49.4%	10.3%
		<0.66	*>0.67*	*>0.99*

Bold entries are statistically significant (χ^2 test, $p<0.05$), the sign indicates the direction of the statistical test (whether the number of tails is lower or higher than expected), and italic entries give the exact p value of the χ^2 test.

which the N-tails modulate target selection have been presented.[42–45] It was recently suggested that the disordered tails stabilize the homeodomain in the presence of DNA but destabilize the homeodomain in its absence.[45,46] The disordered tail of homeodomains also facilitates formation of native interactions with the specific DNAB site via the fly-casting mechanism[47] and thereby increases kinetic specificity.[46]

In this study, we examined the role of the N-tails in nonspecific DNA recognition with the aim of understanding how they contribute to search efficiency in general and to intersegment transfer in particular. Here, we describe the molecular details of 1D diffusion and intersegment transfer of the HoxD9, *Antennapedia* (Antp), and NK-2 *Drosophila* homeodomains, whose N-tails are composed of 9, 10, and 17 residues, respectively (Fig. 1). Using coarse-grained molecular dynamics simulations where nonspecific protein–DNA interactions are modeled solely by electrostatic interactions, we studied the microscopic mechanism of and criteria for intersegment transfer and tested the significance that proteins with multiple domains that interact with DNA may have on search efficiency.

Results and Discussion

Bioinformatics of the disordered tail in DNA binding proteins

As a first step toward understanding the role of disordered tails in nonspecific searches of DNA, the presence of such tails in proteins was tested in three groups of proteins: human DNAB proteins, human non-DNAB (NDNAB) proteins, and a random group of proteins of the same size as the DNAB proteins that was selected from all the proteins in the previous two groups (ALL). DNAB proteins have a high propensity to have such tails; 66.8% of DNAB proteins have a flexible tail, while only 50.9% of NDNAB proteins have tails (Table 1). DNAB proteins are longer on average than ALL proteins, and the relative length (i.e., length of tail/length of protein) of DNAB protein tails was found to be 0.40% longer than that of ALL protein tails, which increases the absolute length of the tail by about seven residues. Next, we affirmed that tails contain a significant number of charged residues, where the total charge of each protein tail has similar mean values for DNAB and ALL proteins (Fig. 1).

The three homeodomain proteins examined in this study (i.e., Antp, NK-2, and HoxD9) show an interesting pattern in the amino acid sequence of their N-tail. Most of the charged residues of their tails are clustered together and form short segments

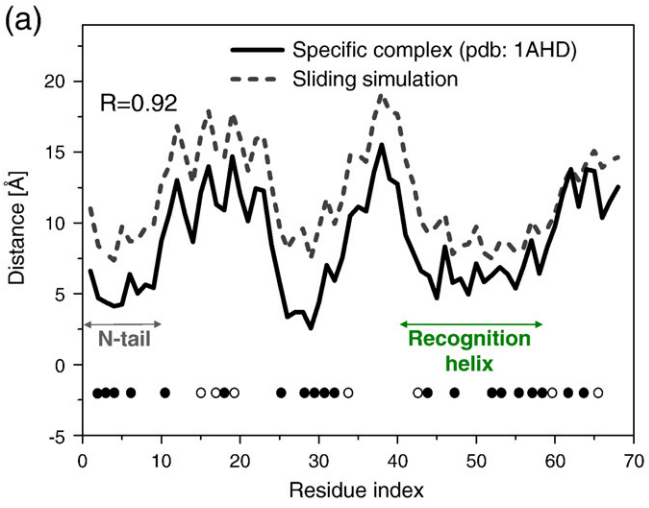

Fig. 2. Comparison of protein–DNA interfaces for specific binding and nonspecific binding. (a) Average distance of each Antp residue from the closest DNA bead when it slides on DNA at a 0.01 M salt concentration (dashed line) compared with the distances in the crystal structure (full line) ($R=0.92$). The full and empty circles at the bottom indicate positively and negatively charged residues in the sequence, respectively. Note that the N-tail includes a segment of four charged residues (three of them form a consecutive segment). (b) Snapshots of Antp as it slides along DNA. The snapshots illustrate that the recognition helix and the N-tail are bound to the DNA and, in particular, that the recognition helix is located at the major groove. For illustration, several snapshots are shown on a single DNA molecule, but in the simulations, the diffusion of a single protein molecule was studied.

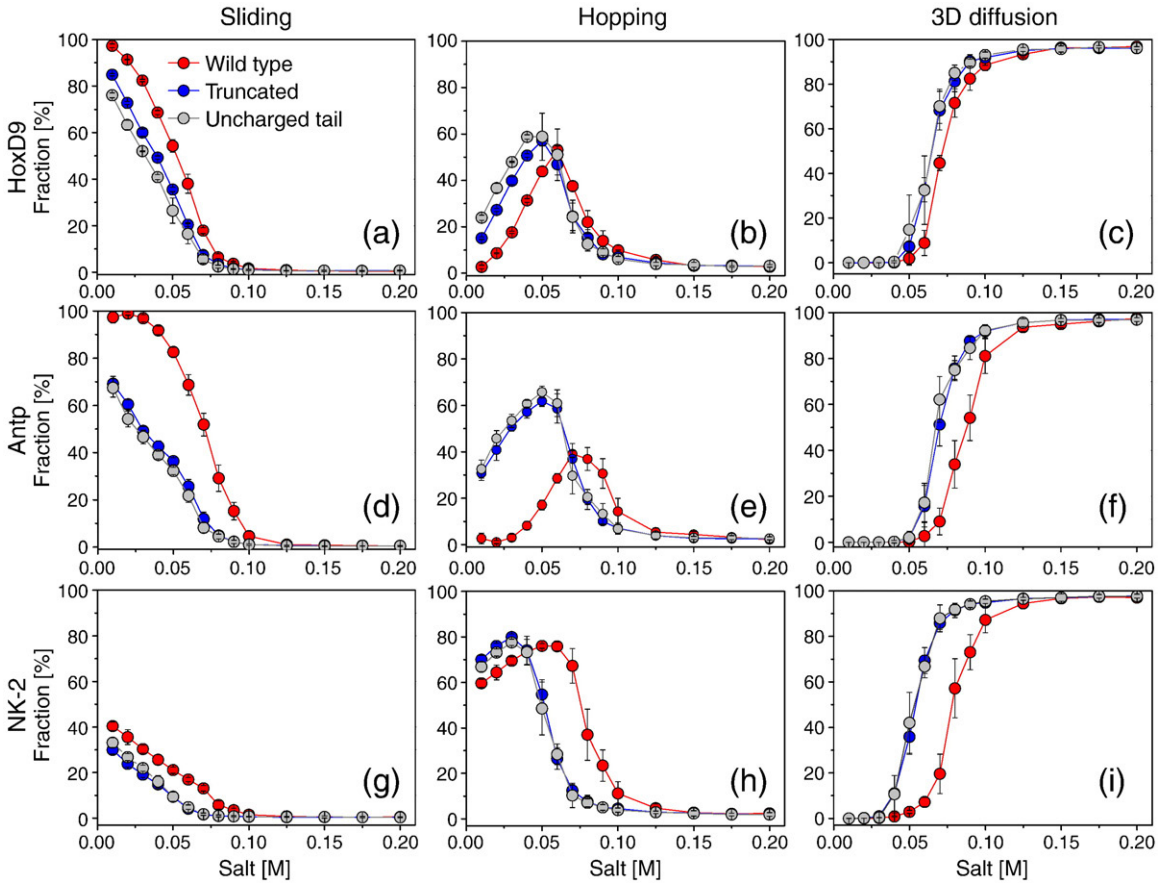

Fig. 3. The effect of the tail on the interplay between sliding, hopping, and 3D diffusion of homeodomain proteins at different salt concentrations. For each of the three studied homeodomains, HoxD9 (a–c), Antp (d–f), and NK-2 (g–i), the DNA search mechanism was investigated for three variants: the WT protein (red), a variant with a truncated N-tail (blue), and a protein with an uncharged tail (gray).

with a high charge density (Fig. 2a and Table S1). For example, all the six positively charged residues of the tail of NK-2 are consecutively linked and form a highly charged segment (for Antp, three of the four positively charged residues in its tail define a charged segment). Previous works by Karlin *et al.* analyzed patterns in protein sequences and showed that charged residues have a statistically significant tendency to group together, either in an alternating sign sequence or as a short consecutive sequence of same-sign residues.[48,49]

Effects of the disordered tail on sliding, hopping, and 3D diffusion on a single B-DNA

To address the effect of an N-disordered tail on the mechanism used to search DNA, we quantified the molecular characteristics of sliding, hopping, and 3D diffusion undertaken by the HoxD9, Antp, and NK-2 homeodomains. For each homeodomain, the search of a 100-bp DNA molecule by the wild-type (WT), N-tail-truncated, and uncharged-tail variants was simulated at a wide range of salt concentrations using a coarse-grained model in which the interactions between the protein and the DNA were modeled by the Debye–Hückel potential.

To examine the configuration of the homeodomain proteins during sliding (simulated at a low salt concentration), we measured the distance between the beads of each protein and their closest DNA bead during sliding and plotted this alongside the equivalent distances in the crystal structure of the specific complex. These distance profiles indicate a high degree of correlation between the protein binding mode adopted during sliding and the mode adopted in the specific complex ($R = 0.92$, 0.71, and 0.69, respectively) by the Antp homeodomain (Fig. 2a) and the NK-2 and HoxD9 homeodomains (Fig. S1). Such a high degree of similarity to the crystal structure indicates analogous DNA association for specific binding and nonspecific binding and was previously demonstrated by NMR for HoxD9[15,19] and by molecular dynamics studies for other DNAB proteins.[16] Although nonspecific binding and specific binding to DNA during sliding share similar features, some differences between the two binding modes are expected and found. Complete structural characterization of nonspecific protein–DNA interactions is difficult to achieve due to the low affinity and highly dynamic nature of such interactions, yet a recent study has suggested that water molecules and cations are pivotal in mediating the electrostatic

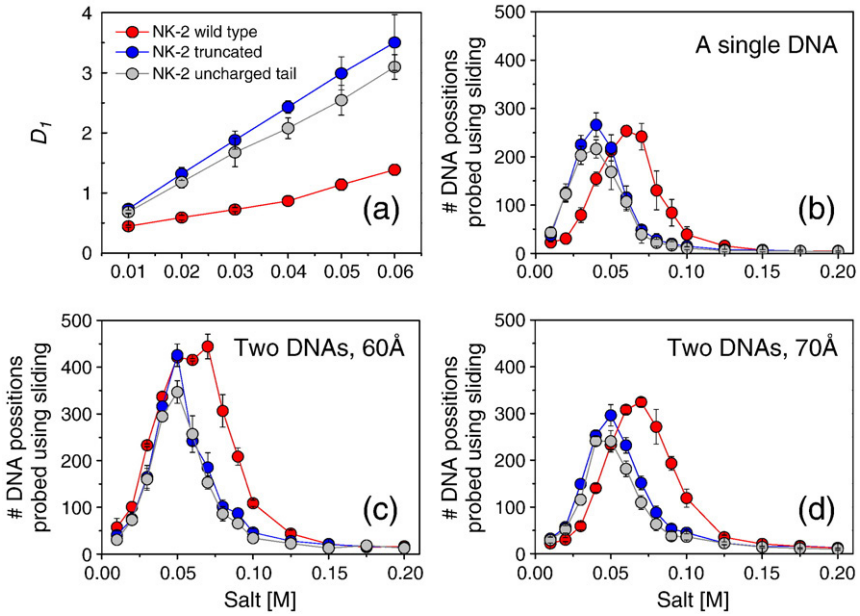

Fig. 4. The effect of the tail on the speed of linear diffusion along the DNA and on search efficiency. (a) Linear diffusion coefficient D_1 as a function of salt concentration calculated for the sliding and hopping dynamics of WT (red), N-tail-truncated (blue), and uncharged-tail (gray) NK-2's. (b–d) The number of DNA positions that were probed during the simulation by NK-2 using sliding as a function of salt concentration for a single DNA molecule (b), two DNA molecules positioned 60 Å apart (c), and two DNA molecules positioned 70 Å apart (d).

interactions that support the sliding of DNAB proteins.[50]

The fractions of sliding, hopping, and 3D search performed by the three homologous homeodomains at a salt concentration in the range of 0.01–0.2 M are shown in Fig. 3. As salt concentration increases, the electrostatic attraction between the protein and the DNA weakens and therefore the use of the sliding mechanism decreases and that of 3D diffusion rises. The hopping search mode, which requires that the protein not be bound too tightly at the major groove but still be close to the DNA,[16] is highly populated at moderate salt concentrations. The presence of a tail improves sliding at the expense of hopping and 3D diffusion. For example, at a low salt concentration, WT Antp scans the DNA only via sliding, while Antp with a truncated or uncharged tail searches the DNA 70% by sliding and 30% by hopping. It is likely that the larger fraction of sliding in the presence of the disordered tail is due to the higher affinity of the native protein for the DNA caused by the positive charges on the N-tail. Indeed, Fig. 2b illustrates that during sliding the tail is close to the DNA and leads to an extended protein–DNA interface. In addition, the high degree of similarity between proteins having truncated tails and those having uncharged tails suggests that the effects of the tail are due to favorable electrostatic interactions between the tail and the DNA. The sliding durations of the WT proteins are often longer (i.e., they probe longer segments of the DNA) than the truncated variants (Fig. S2). We also note that the balance between the three search mechanisms is different for the WT and truncated homeodomains (Fig. S3). For truncated homeodomains, an optimally efficient DNA search (measured by the number of base pairs scanned via sliding) involves 10%–20% sliding (namely, 80%–90% hopping and 3D search), while for the WT homeodomain, an efficient search is 30%–50% sliding. Furthermore, the three truncated homeodomains exhibit similar search behaviors, yet the diversity of their search behavior is greater in the presence of their tail, indicating the importance of the tail in modulating DNA kinetics.

The strong dependence on salt concentration of the percentage of usage of a hopping search mode implies that the 1D diffusion coefficient, D_1, will also be affected by salt concentration, as was recently shown,[16] and that the diffusion rate will be different for the WT and N-tail-truncated homeodomains. The value of D_1 for NK-2 variants that diffuse on the DNA (by either sliding or hopping dynamics) increases with increasing salt concentration because of the increased number of hopping events, with hopping being faster than helically sliding along the DNA backbone[16] (Fig. 4a; similar D_1 characteristics were observed for the HoxD9 and Antp homeodomains as well). The N-tail-truncated and uncharged-tail homeodomains show similar behavior when they interact nonspecifically or specifically[46] with DNA. Yet we found that the rate of linear diffusion is slower for the uncharged tail, presumably due to collisions between the disordered tail and the DNA. This slower sliding may cause slight differences in the search efficiency of the DNA by the N-tail-truncated and uncharged-tail proteins (we point out that it is not trivial to experimentally design a homeodomain with an uncharged disordered tail since the excess of charges is an intrinsic property of being disordered). Importantly, we found that the D_1

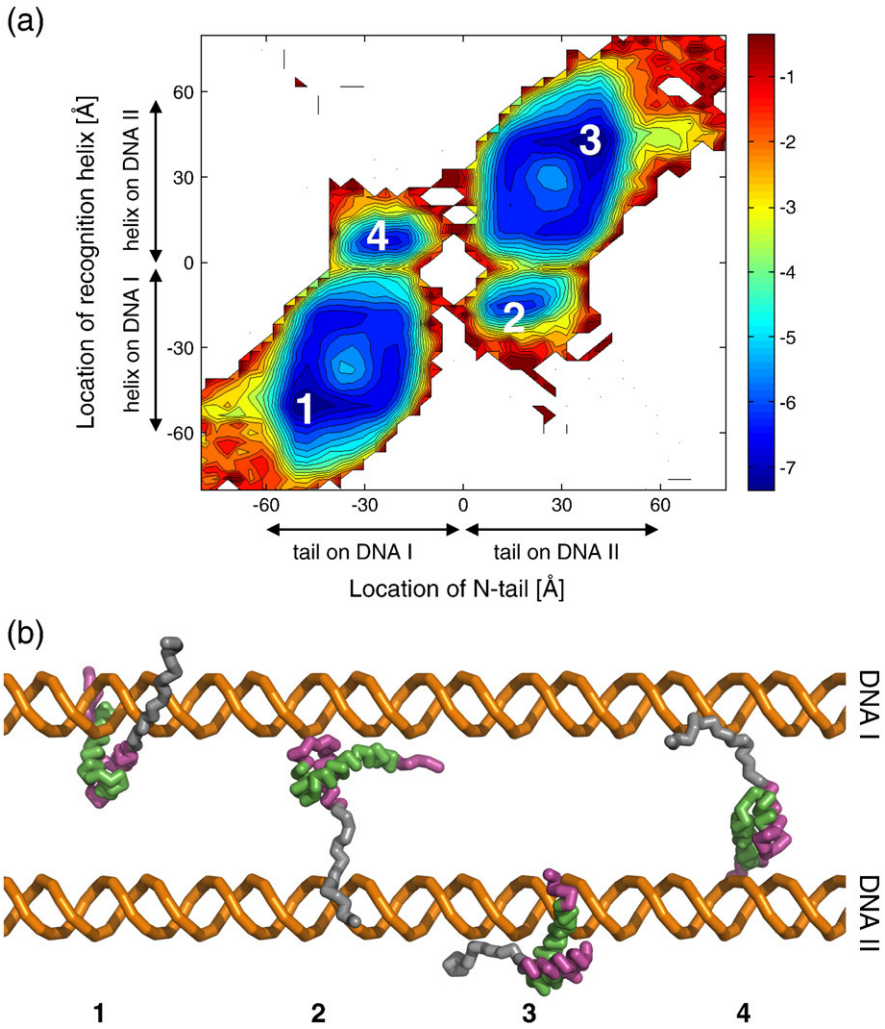

Fig. 5. Free-energy surfaces of intersegment transfer. (a) The free energy is projected along the locations of the recognition helix and the N-tail of the NK-2 homeodomain for the two DNA molecules. The location of the center of mass of the recognition helix *versus* the center of mass of the tail for WT NK-2 at a salt concentration of 0.06 M suggests a bridged intermediate for an intersegment transfer mechanism (the centers of the DNA molecules are placed at [30, 0, *z*] and [−30, 0, *z*]). (b) Graphical representation of four frames of NK-2 illustrating the mechanism of intersegment transfer. Snapshots 1 and 3 correspond to diffusion on DNA I and that on DNA II. Snapshots 2 and 4 correspond to intersegment transfer events. This figure shows that the intersegment transfer can be initiated by the tail (2) or by the recognition helix (4). For illustration purposes, several snapshots are shown, but in the simulations, a single copy of the protein was used to study the mechanism of intersegment transfer.

values of the N-tail-truncated proteins and the uncharged-tail proteins are significantly higher than the D_1 value of the WT protein. This can be explained by the lower affinity of the N-tail-truncated and uncharged-tail proteins for the DNA since they lack positive charges at the protein–DNA interface that may significantly increase the fraction of hopping dynamics even at low salt concentrations (Fig. 3b, e, and h).

Effects of the disordered tail on search of two B-DNAs: intersegment transfer via the "monkey bar" mechanism

The positively charged tail may, on the one hand, accelerate intersegment transfer since it can act as an additional DNA recognition motif. On the other hand, the existence of the tail increases the affinity of the protein for the DNA, as was shown above and by thermodynamic analysis,[44] and therefore may inhibit intersegment transfer by inducing tighter binding to the DNA. In order to study the mechanism of intersegment transfer and the influence of the positively charged N-tail, we studied the dynamics of the homeodomains in the presence of two parallel DNA molecules (each of 100 bp).

Free-energy surfaces projected onto the position of the center of mass of the recognition helix and the tail in the *xy* plane highlight the mechanism of intersegment transfer. The energy landscape for protein jumps is composed of four minima: Two minima correspond to cases in which both the recognition helix and the N-tail are bound to the same DNA molecule (either DNA I or DNA II).

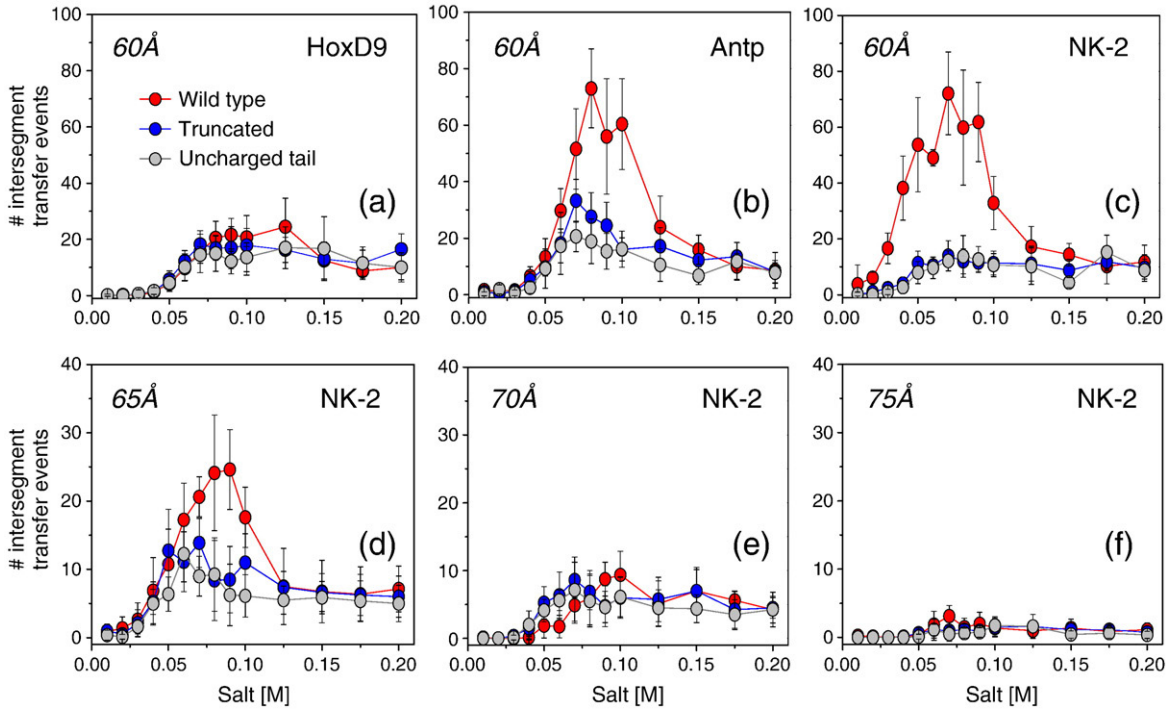

Fig. 6. Quantitative characteristics of intersegment transfer. (a–c) The effect of the tail on intersegment transfer events by the HoxD9 (a), Antp (b), and NK-2 (c) homeodomains for salt concentrations of 0.01–0.2 M when the two DNA molecules are 60 Å apart. (d–f) The effect of the distance between the two DNA molecules on the number of intersegment transfer events was studied for the NK-2 homeodomain for two DNA molecules separated by 65 Å (d), 70 Å (e), and 75 Å (f). The red line corresponds to the WT proteins, the blue line corresponds to the N-tail-truncated proteins, and the gray line corresponds to the uncharged-tail proteins. The number of intersegment transfer events and the difference between tailed and truncated proteins decrease as the length of the tail decreases (a–c) and as the distance between DNA molecules increases (d–f).

The other two minima correspond to cases in which the recognition helix and the N-tail interact with different DNA molecules. The transfer from one DNA to the other proceeds through a bridged intermediate in which the tail is located on one DNA, while the recognition helix is on the second DNA (Fig. 5). This mechanism resembles the motion of children as they swing along "monkey bars". Similar to the way a child transfers one hand at a time when swinging from bar to bar, the homeodomains cross from one DNA to another by transferring first a single domain, followed, after a certain lag time, by the transition of the other domain. This description of a homeodomain protein as composed of two subdomains, the globular part and the N-tail, that assist intersegment transfer is supported by NMR measurements, which indicated that direct translocation of HoxD9 occurs between different DNA fragments, leading the authors to conjecture that the tail and globular subdomains behave independently.[20] Intersegment transfer via the monkey bar mechanism is also observed for NK-2 with an uncharged tail, yet the free energy of the intermediates is higher compared with that of WT NK-2 and the four minima that comprise the free-energy landscape are much broader due to the lower affinity for the DNA (data not shown). Originally, intersegment transfer was hypothesized for multiple-domain proteins, and, indeed, an NMR study has recently shown that the two-domain protein Oct-1 extensively forms the bridged intermediate.[22] However, our study suggests that other features of the proteins, such as a flexible tail, can serve to stabilize the intermediate and thus facilitate intersegment transfer.

Intersegment transfer takes place via an intermediate state in which the N-tail interacts with one DNA and the recognition helix interacts with the other DNA (Fig. 5b). The intersegment transfer events can be initiated either by the tail or by the recognition helix of NK-2, with a quantitative characterization revealing similar probabilities for both intersegment pathways. To further quantify the structural properties of intersegment transfer, we explored the effects of salt concentration, tail properties, and distance between the DNA molecules on the number of these events (Fig. 6). The number of jumps performed by the three homeodomains between two DNA molecules that are 60 Å apart is depicted in Fig. 6a–c. While the tail scarcely affects the number of jumps by HoxD9, it unequivocally increases the number of intersegment transfer events at moderate salt concentrations for the Antp and NK-2 homeodomains. The small number of intersegment transfers made by HoxD9 (which has a 9-residue tail) compared with the numbers made by Antp and NK-2 (which have tails of 10 and 17 residues, respectively) suggests that the

length of the tail is an important factor in intersegment transfer. However, the small difference between the number of jumps by Antp and that by NK-2, despite the fact that the tail of NK-2 is about twice as long as that of Antp, indicates that other factors (e.g., the number of charges and their location along the tail) are important as well. We note that even the truncated variants of the homeodomains can jump, although the number of these events is quite small. The capability of the N-tail to facilitate intersegment transfer is affected by the distance between the two DNAs. The number of intersegment transfer events and the difference between WT and truncated homeodomains reduce as the distance between the DNA molecules increases (or as tail length decreases). The maximal number of intersegment transfer events and the greatest difference between the WT and truncated proteins occur for the NK-2 homeodomain, which has the longest N-tail, at a DNA separation distance of 60 Å (Fig. 6c).

In our simulation model, we found that a significant population of intersegment transfer events occur at salt concentrations of 0.08–0.1 M. The existence of an optimal salt concentration for this mechanism stems from the need to balance the strength of the electrostatic attraction between the protein and the DNA. For intersegment transfer, electrostatics has to be strong enough to enable mild attraction between the protein domain (either the recognition helix or the N-tail) and the second DNA, but not so strong as to prevent dissociation of the protein from the first DNA.

Efficiency of DNA search: effects of the disordered tail and the distance between the DNA molecules

A recent theoretical study predicted that direct transfer from one DNA fragment to another would significantly improve the efficiency of target search by a DNAB protein.[23] In this study, we investigated the efficiency of DNA search by three homeodomain proteins as a function of salt concentration, the presence of the N-tail, and the distance between the two DNA molecules. To do so, we calculated the number of DNA positions probed during the simulations by the homeodomains using sliding following the method presented by Givaty and Levy.[16]

For a single DNA molecule, optimal search efficiency is achieved at moderate salt concentrations, consistent with previous studies.[2,16] While maximal search efficiency is similar for the uncharged and truncated NK-2's, the curve for the WT protein is shifted to higher salt concentrations (Fig. 4b). This shift is due to the WT protein making greater use of sliding at the expense of hopping and 3D diffusion because of its enhanced affinity for the DNA due to its N-tail. Search efficiency on two parallel DNA molecules separated by 60 Å is almost twice that achieved during search on a single DNA molecule (Fig. 4c). This high search efficiency at a salt concentration range of 0.05–0.10 M is linked to the larger number of intersegment transfer events seen for WT NK-2 in this salt concentration range (Fig. 6c). Increasing the distance between the two DNA molecules results in lower search efficiency, which corresponds to a significant decrease in the number of intersegment transfer events (Fig. 6d–f), and in a shift of the WT NK-2 curve to higher salt concentrations (Fig. 4d), resulting in a search efficiency similar to that achieved on a single DNA molecule. Increasing the distance between the two DNA molecules is expected to be quite equivalent to shortening the tail length. These results suggest that intersegment transfer significantly facilitates the search process.

Conclusions

Various studies using a variety of techniques have strongly supported the ability of proteins to slide along DNA and have indicated that this linear diffusion, in cooperation with other search mechanisms, is essential for rapid target search of the cognate site. NMR studies,[15,19,20,22] single-molecule experiments,[9] and computational studies[16] strongly support the notion that during sliding the protein moves in a helical fashion along the phosphate–sugar rail; however, direct experimental detection of helical sliding is still to be achieved. As was predicted by Schurr,[18] helical sliding yields a slower diffusion rate than does a simple linear movement along the DNA.[16] While the facilitated diffusion mechanism and, in particular, the sliding dynamics along DNA seem to be prerequisite to many specific protein–DNA recognitions, their universality is not clear. It is obvious that specific protein–DNA interactions are subject to many subtle molecular details (known as the direct and indirect readouts) that affect the formation of hydrogen-bond networks between the two biomolecules. Is the strength of the nonspecific protein–DNA interaction also very diverse among DNAB proteins? Do all DNAB proteins slide during the search, irrespective of whether they specifically bind at the major or minor groove? Is search efficiency mostly determined by the structure of the interface formed with the DNA or does it depend on further chemical details?

In this study, we have addressed the question of the linkage between the molecular architecture of DNAB proteins and the mechanism and efficiency of their DNA search by focusing on the DNA interaction dynamics of three structurally homologous homeodomain proteins (HoxD9, Antp, and NK-2) that have different electrostatic potentials and include disordered tails that are highly positively charged but that are of different lengths. Using coarse-grained models, where the protein–DNA interactions were represented by electrostatic forces only and where the protein was flexible but the DNA molecules remained rigid, we characterized the molecular features of sliding and intersegment transfer for the homeodomain proteins. The

disordered N-tail contributes to a higher propensity to slide by increasing the affinity of the protein for the DNA in nonspecific binding. The tighter protein–DNA interaction induced by the N-tail results in a smaller 1D diffusion coefficient.

In intersegment transfer, direct relocation from one nonspecific DNA to another occurred without accumulation of free protein. The tailed proteins translocate on DNA molecules through an intermediate in which the recognition helix of the protein is adsorbed to one DNA fragment and the N-tail is adsorbed to the second, suggesting a "monkey bar" mechanism. Intersegment transfer is facilitated by the disordered tail, whose length can significantly affect the number of jumping events. A longer tail or a shorter distance between the two DNA molecules results in a greater number of intersegment transfers and therefore in a more efficient DNA search. In addition, the number and distribution of the charged residues[51] of the disordered tail are expected to affect intersegment transfer events.

The disordered tails, which are common in DNAB proteins, are important not only for increasing the affinity of the specific protein–DNA complex but also for facilitating search kinetics. Proteins with disordered tails may therefore be treated as multi-subdomain proteins. Formulating how the molecular features (e.g., degree of flexibility, size, and structure) of the two subdomains or the linkers that connect various multi-domain DNAB proteins (e.g., zinc finger proteins) affect search is essential.

Methods

Simulation model

The molecular and dynamic nature of protein search of DNA was studied using a reduced model that allows sampling of long timescale processes such as sliding, hopping, 3D diffusion, and intersegment transfer. We modeled the DNA as having three beads per nucleotide, representing phosphate, sugar, and base. Each bead is located at the geometric center of the group it represents. A negative point charge is assigned to beads representing the DNA phosphate groups. In the simulations, a 100-bp B-DNA molecule was used to study protein dynamics on a single double-stranded DNA molecule and two 100-bp B-DNA molecules were applied for investigation of intersegment transfer. The protein and the DNA were placed in a box with dimensions of 300 Å×300 Å×700 Å, with the DNA being placed at the center of the box along its *z*-axis.

The protein is represented by a single bead for each residue located at the C^{α} of that residue. Beads representing charged amino acids (Arg, Lys, Asp, and Glu) are charged in the model. Unlike the DNA, the protein remains flexible during the simulations and can undergo folding and unfolding events. We simulated the protein with a native-topology-based model, which corresponds to a perfectly funneled energy landscape,[52] where native protein interactions are attractive and all other interactions are repulsive.[53] The native contacts (defined by the CSU algorithm[54]) were modeled using the Lennard–Jones potential and made the same energetic contribution at their optimal distance irrespective of the identity of the residues participating in the contact. In addition to the native interactions, electrostatic interactions between all charged residues of the protein and the phosphate bead of the DNA were included and were modeled by the Debye–Hückel potential, which accounts for the ionic strength of a solute immersed in aqueous solution.[55] We would like to emphasize that the native-topology-based model was applied only to the protein but the protein–DNA interactions were modeled solely by electrostatic and excluded-volume interactions.

The Debye–Hückel potential has been previously used to study the energetics and dynamics of various biomolecular systems, such as RNA folding,[56] chromatin assembly,[57] and protein–DNA binding.[58] While the Debye–Hückel approximation is a powerful means of introducing the electrostatic screening effect of salt atoms into the Coulomb potential, one should be aware of the approximations it makes. In particular, the Debye–Hückel model, while valid for dilute solutions, does not take into account ion condensation on the DNA. The dynamics of each protein was studied at various salt concentrations in the range of 0.01–0.2 M and using a dielectric constant of 80 on one and two B-DNAs at the temperature at which the protein is completely folded during the simulations. Due to the simplicity of our model and in particular the long distances between the charged beads of the protein and the charged beads of the DNA (since in the reduced model not all the atoms are represented and the charges are placed at the phosphate and C^{α} beads), the salt concentrations that are reported in this study are a few-fold smaller than the experimental ionic strength. More details on the simulation can be found in the work of Givaty and Levy.[16]

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

Trajectories were analyzed to quantify the percentage of protein sliding and hopping, structural features during sliding, and the linear diffusion coefficient (D_1) using the prescriptions by Givaty and Levy.[16] To obtain significant statistics, we performed eight simulations of 10^8 steps for each protein at various salt concentrations. The simulation frame was considered as being part of 3D diffusion if the center of mass of the recognition helix was farther than 30 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 taking part in sliding search if 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 recognition helix was found at a distance of less than 30 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.[16]

Intersegment transfer was studied by following the dynamics of the homeodomain in the presence of two parallel DNA molecules (each of 100 bp) that are aligned along the *z*-axis with a corresponding offset in the *x*-axis of 60, 65, 70, or 75 Å (the DNA molecules remain in place and rigid throughout the simulations). At a shorter distance, the spiral sliding of the protein along the DNA is prohibited; at longer distances, the tails hardly contribute to jumps. One can obviously imagine that two stretches of the DNA in the cell are not necessarily parallel but that

they can adopt various architectures and geometries. The number of intersegment transfers accordingly will be affected by the relative orientation of the two DNAs; however, for a clearer comparison, we have focused in this study on parallel arrangement.[59] Intersegment transfer events were probed using the following definition: hopping or sliding of protein along a DNA molecule for 20 consecutive frames followed by dissociation from that DNA molecule and hopping or sliding along the other DNA molecule. When this event occurred through an intermediate in which the recognition helix was located closer than 30 Å to the first DNA and the N-tail was found nearer than 30 Å to the second DNA, the intersegment transfer was counted as having been initiated by the N-tail. For the opposite case, where the N-tail was closer than 30 Å to the first DNA and the recognition helix was less than 30 Å from the second DNA, the intersegment transfer was counted as having been initiated by the recognition helix. The free-energy surface for intersegment transfer was calculated by the logarithm of the probability to find the recognition helix and the tail on either DNA I or DNA II.

Statistical analysis of tail properties in DNAB proteins and all other proteins

Sequences of annotated human proteins were downloaded from the Uniprot site. The AMIGO gene ontology site was used to obtain a protein subset with the "DNA binding" gene ontology (GO) annotation. Determination of protein tails was performed using IUPred.[60] A protein was considered to have a tail if an unstructured segment of at least five amino acids was predicted at either its N-terminus or its C-terminus (if the protein was predicted to have unstructured segments at both ends, two tails were counted separately). The charge on the tail was calculated as follows: for each Lys or Arg present in the tail, one positive charge was counted (+1), and for each Glu or Asp present, one negative charge was counted (−1).

Acknowledgements

This work was supported by the Kimmelman Center for Macromolecular Assemblies, the Center for Complexity Science (Y.L.), and the Foundation Fernande et Jean Gaj. Y.L. is incumbent of the Lilian and George Lyttle Career Development Chair. Fruitful discussions with Ohad Givaty are gratefully acknowledged. We thank Paul Whitford for sharing with us his molecular dynamics code.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2009.11.056

References

1. Halford, S. E. (2009). An end to 40 years of mistakes in DNA–protein association kinetics? *Biochem. Soc. Trans.* **37**, 343–348.
2. 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.
3. Halford, S. E. & Marko, J. F. (2004). How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* **32**, 3040–3052.
4. Slutsky, M. & Mirny, L. A. (2004). Kinetics of protein–DNA interaction: facilitated target location in sequence-dependent potential. *Biophys. J.* **87**, 4021–4035.
5. Klenin, K. V., Merlitz, H., Langowski, J. & Wu, C. X. (2006). Facilitated diffusion of DNA-binding proteins. *Phys. Rev. Lett.* **96**, 018014 (018014-1 - 018014-4).
6. Wunderlich, Z. & Mirny, L. A. (2008). Spatial effects on the speed and reliability of protein–DNA search. *Nucleic Acids Res.* **36**, 3570–3578.
7. 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.
8. von Hippel, P. H. & Berg, O. G. (1989). Facilitated target location in biological-systems. *J. Biol. Chem.* **264**, 675–678.
9. Gorman, J. & Greene, E. C. (2008). Visualizing one-dimensional diffusion of proteins along DNA. *Nat. Struct. Mol. Biol.* **15**, 768–774.
10. Bustamante, C., Guthold, M., Zhu, X. S. & Yang, G. L. (1999). Facilitated target location on DNA by individual *Escherichia coli* RNA polymerase molecules observed with the scanning force microscope operating in liquid. *J. Biol. Chem.* **274**, 16665–16668.
11. Wang, J., Lu, Q. & Lu, H. P. (2006). Single-molecule dynamics reveals cooperative binding–folding in protein recognition. *PLoS Comput. Biol.* **2**, e78.
12. Elf, J., Li, G. W. & Xie, X. S. (2007). Probing transcription factor dynamics at the single-molecule level in a living cell. *Science*, **316**, 1191–1194.
13. 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.
14. Tafvizi, A., Huang, F., Leith, J. S., Fersht, A. R., Mirny, L. A. & van Oijen, A. M. (2008). Tumor suppressor p53 slides on DNA with low friction and high stability. *Biophys. J.* **95**, L01–L03.
15. Iwahara, J. & Clore, G. M. (2006). Detecting transient intermediates in macromolecular binding by paramagnetic NMR. *Nature*, **440**, 1227–1230.
16. Givaty, O. & Levy, Y. (2009). Protein sliding along DNA: dynamics and structural characterization. *J. Mol. Biol.* **385**, 1087–1097.
17. 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.
18. Schurr, J. M. (1979). The one-dimensional diffusion coefficient of proteins absorbed on DNA. Hydrodynamic considerations. *Biophys. Chem.* **9**, 413–414.
19. 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.
20. 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.
21. Lieberman, B. A. & Nordeen, S. K. (1997). DNA intersegment transfer, how steroid receptors search for a target site. *J. Biol. Chem.* **272**, 1061–1068.

22. Doucleff, M. & Clore, G. M. (2008). Global jumping and domain-specific intersegment transfer between DNA cognate sites of the multidomain transcription factor Oct-1. *Proc. Natl Acad. Sci. USA*, **105**, 13871–13876.
23. Hu, T. & Shklovskii, B. I. (2007). How a protein searches for its specific site on DNA: the role of intersegment transfer. *Phys. Rev. E*, **76**, 051909.
24. Sheinman, M. & Kafri, Y. (2009). The effects of intersegmental transfers on target location by proteins. *Phys. Biol.* **6**, 1–17.
25. Loverdo, C., Benichou, O., Voituriez, R., Biebricher, A., Bonnet, I. & Desbiolles, P. (2009). Quantifying hopping and jumping in facilitated diffusion of DNA-binding proteins. *Phys. Rev. Lett.* **102**, 188101.
26. Wright, P. E. & Dyson, H. J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure–function paradigm. *J. Mol. Biol.* **293**, 321–331.
27. Dyson, H. J. & Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6**, 197–208.
28. Uversky, V. N. (2002). What does it mean to be natively unfolded? *Eur. J. Biochem.* **269**, 2–12.
29. Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S. *et al.* (2001). Intrinsically disordered protein. *J. Mol. Graphics Modell.* **19**, 26–59.
30. Liu, J. G., Perumal, N. B., Oldfield, C. J., Su, E. W., Uversky, V. N. & Dunker, A. K. (2006). Intrinsic disorder in transcription factors. *Biochemistry*, **45**, 6873–6888.
31. Tompa, P. (2002). Intrinsically unstructured proteins. *Trends Biochem. Sci.* **27**, 527–533.
32. Gehring, W. J., Affolter, M. & Burglin, T. (1994). Homeodomain proteins. *Annu. Rev. Biochem.* **63**, 487–526.
33. Scott, M. P., Tamkun, J. W. & Hartzell, G. W. (1989). The structure and function of the homeodomain. *Biochim. Biophys. Acta*, **989**, 25–48.
34. Dorn, A., Affolter, M., Gehring, W. J. & Leupin, W. (1994). Homeodomain proteins in development and therapy. *Pharmacol. Ther.* **61**, 155–184.
35. Wolberger, C., Vershon, A. K., Liu, B. S., Johnson, A. D. & Pabo, C. O. (1991). Crystal structure of a MAT alpha-2 homeodomain–operator complex suggests a general model for homeodomain–DNA interactions. *Cell*, **67**, 517–528.
36. Tsao, D. H. H., Gruschus, J. M., Wang, L. H., Nirenberg, M. & Ferretti, J. A. (1995). The 3-dimensional solution structure of the NK-2 homeodomain from *Drosophila*. *J. Mol. Biol.* **251**, 297–307.
37. Otting, G., Qian, Y. Q., Billeter, M., Muller, M., Affolter, M., Gehring, W. J. & Wuthrich, K. (1990). Protein–DNA contacts in the structure of a homeodomain–DNA complex determined by nuclear magnetic resonance spectroscopy in solution. *EMBO J.* **9**, 3085–3092.
38. Gruschus, J. M., Tsao, D. H. H., Wang, L. H., Nirenberg, M. & Ferretti, J. A. (1997). Interactions of the vnd/NK-2 homeodomain with DNA by nuclear magnetic resonance spectroscopy: basis of binding specificity. *Biochemistry*, **36**, 5372–5380.
39. Billeter, M., Qian, Y. Q., Otting, G., Muller, M., Gehring, W. & Wuthrich, K. (1993). Determination of the nuclear magnetic resonance solution structure of an *Antennapedia* homeodomain–DNA complex. *J. Mol. Biol.* **234**, 1084–1094.
40. Joshi, R., Passner, J. M., Rohs, R., Jain, R., Sosinsky, A., Crickmore, M. A. *et al.* (2007). Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell*, **131**, 530–543.
41. Damante, G. & Dilauro, R. (1991). Several regions of *Antennapedia* and thyroid transcription factor-I homeodomains contribute to DNA-binding specificity. *Proc. Natl Acad. Sci. USA*, **88**, 5388–5392.
42. Zeng, W. L., Andrew, D. J., Mathies, L. D., Horner, M. A. & Scott, M. P. (1993). Ectopic expression and function of the Antp and Scr homeotic genes—the N-terminus of the homeodomain is critical to functional specificity. *Development*, **118**, 339–352.
43. Privalov, P. L., Dragan, A. I., Crane-Robinson, C., Breslauer, K. J., Remeta, D. P. & Minetti, C. A. S. A. (2007). What drives proteins into the major or minor grooves of DNA? *J. Mol. Biol.* **365**, 1–9.
44. Crane-Robinson, C., Dragan, A. I. & Privalov, P. L. (2006). The extended arms of DNA-binding domains: a tale of tails. *Trends Biochem. Sci.* **31**, 547–552.
45. Dragan, A. I., Li, Z. L., Makeyeva, E. N., Milgotina, E. I., Liu, Y. Y., Crane-Robinson, C. & Privalov, P. L. (2006). Forces driving the binding of homeodomains to DNA. *Biochemistry*, **45**, 141–151.
46. Tóth-Petróczy, A., Simon, I., Fuxreiter, M. & Levy, Y. (2009). The role of disordered tails in specific DNA binding of homeodomains. *J. Am. Chem. Soc.* **131**, 15084.
47. Shoemaker, B. A., Portman, J. J. & Wolynes, P. G. (2000). Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc. Natl Acad. Sci. USA*, **97**, 8868–8873.
48. Brendel, V., Bucher, P., Nourbakhsh, I. R., Blaisdell, B. E. & Karlin, S. (1992). Methods and algorithms for statistical analysis of protein sequences. *Proc. Natl Acad. Sci. USA*, **89**, 2002–2006.
49. Karlin, S. & Altschul, S. F. (1993). Applications and statistics for multiple high-scoring segments in molecular sequences. *Proc. Natl Acad. Sci. USA*, **90**, 5873–5877.
50. Dahirel, V., Paillusson, F., Jardat, M., Barbi, M. & Victor, J. M. (2009). Nonspecific DNA–protein interaction: why proteins can diffuse along DNA.*Phys. Rev. Lett.* **102**, 228101 (228101-1-22801-4).
51. Florescu, A. & Joyeux, M. (2009). Dynamical model of DNA–protein interaction: effect of protein charge distribution and mechanical properties. *J. Chem. Phys.* **131**, 105102.
52. Onuchic, J. N. & Wolynes, P. G. (2004). Theory of protein folding. *Curr. Opin. Struct. Biol.* **14**, 70–75.
53. 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.
54. Sobolev, V., Wade, R. C., Vriend, G. & Edelman, M. (1996). Molecular docking using surface complementarity. *Proteins: Struct., Funct., Genet.* **25**, 120–129.
55. Schlick, T. (2000). *Molecular Modeling and Simulation: An Interdisciplinary Guide.* Springer, New York, NY.
56. Hyeon, C. & Thirumalai, D. (2005). Mechanical unfolding of RNA hairpins. *Proc. Natl Acad. Sci. USA*, **102**, 6789–6794.
57. Beard, D. A. & Schlick, T. (2001). Computational modeling predicts the structure and dynamics of chromatin fiber. *Structure*, **9**, 105–114.
58. Marcovitz, A. & Levy, Y. (2009). Arc-repressor dimerization on DNA: folding rate enhancement by colocalization. *Biophys. J.* **96**, 4212–4220.
59. Savelyev, A. & Papoian, G. A. (2007). Inter-DNA electrostatics from explicit solvent molecular dynamics simulations. *J. Am. Chem. Soc.* **129**, 6060–6061.
60. Dosztanyi, Z., Csizmok, V., Tompa, P. & Simon, I. (2005). IUPred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*, **21**, 3433–3434.