DNA search efficiency is modulated by charge composition and distribution in the intrinsically disordered tail

Dana Vuzman and Yaakov Levy[1]

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

Edited by Peter G. Wolynes, University of California at San Diego, La Jolla, CA, and approved October 6, 2010 (received for review August 7, 2010)

Intrinsically disordered tails are common in DNA-binding proteins and can affect their search efficiency on nonspecific DNA by promoting the brachiation dynamics of intersegment transfer. During brachiation, the protein jumps between distant DNA regions via an intermediate state in which the tail and globular moieties are bound to different DNA segments. While the disordered tail must be long and positively charged to facilitate DNA search, the effect of its residue sequence on brachiation is unknown. We explored this issue using the NK-2 and Antp homeodomain transcription factors. We designed 566 NK-2 tail-variants and 55 Antp tail-variants having different net charges and positive charge distributions and studied their dynamics and DNA search efficiencies using coarse-grained molecular dynamics simulations. More intersegment transfers occur when the tail is moderately positively charged and the positive charges are clustered together in the middle of the tail or towards its N terminus. The presence of a negatively charged residue does not significantly affect protein brachiation, although it is likely that the presence of many negatively charged residues will complicate the DNA search mechanism. A bioinformatic analysis of 1,384 wild-type homeodomains illustrates that the charge composition and distribution in their N-tail sequences are consistent with an optimal charge pattern to promote intersegment transfer. Our study thus indicates that the residue sequence of the disordered tails of DNA-binding proteins has unique characteristics that were evolutionarily selected to achieve optimized function and suggests that the sequence-structure-function paradigm known for structured proteins is valid for intrinsically disordered proteins as well.

intrinsically disordered proteins | protein-DNA interaction | sliding | intersegment transfer | Monkey-bar mechanism

It is generally recognized that protein segments or even whole proteins can fulfill key functions in the cell without having a well defined tertiary structure (1–3). These intrinsically disordered regions are abundant in eukaryotic organisms and are associated mostly with regulatory functions (1–9). The amino acid sequence of intrinsically disordered proteins differs from that of globular proteins as the former are rich in charged residues, deficient in hydrophobic residues, and show a low degree of complexity (10, 11). Yet, disordered proteins are not without structure (12–14) and recent studies have shown that they can adopt collapsed structures in which their degree of compaction is modulated by their net charge (15–17). The sequence-structure-function relationship is essential in structured proteins and may also be important with respect to intrinsically disordered proteins, about which much less is known. To obtain an understanding of this relationship in intrinsically disordered proteins, it is necessary to investigate the sequence determinants and evolutionary conservation of their structure and function.

We have previously shown that DNA-binding proteins tend to have disordered tails: about 70% of DNA-binding proteins have a disordered tail while only about 50% of non-DNA-binding proteins have such a tail (18). While, traditionally, many studies have focused on the folded region of DNA-binding proteins to evaluate affinity and selectivity to DNA, it is becoming clearer that the disorder tail modulates several key aspects of protein-DNA interactions (19). Binding homeodomain protein N-tails, which are highly charged and disordered in solution, in the minor groove induces the tail to fold, although it may also remain flexible or partly disordered in the complex (20). Several independent pieces of evidence point to the fact that the disordered N-tails of homeodomains contribute to specific DNA sequence selectivity (20–24). However, only a few hypotheses concerning the mechanism by which the N-tails modulate target selection have been presented (19, 24–26). It was recently shown that the disordered tails stabilize the homeodomain in the presence of DNA but destabilize the homeodomain in its absence (26, 27). The N-tail of homeodomains also facilitates the formation of specific interactions at the DNA-binding site and thereby increases kinetic specificity (27).

The disordered tail of DNA-binding proteins can be viewed as another subdomain that interacts with the DNA nonspecifically (Fig. 1). We have shown, in agreement with kinetic NMR studies (28–30), that the N-tails of homeodomain proteins can significantly facilitate DNA search (18). The attraction of the positively charged N-tail to the DNA may increase the affinity of the protein to the DNA and therefore increase the proportion of the search undertaken by means of one-dimensional sliding along the DNA. While a greater sliding propensity may improve search efficiency, one may note that the N-tail can also reduce the diffusion coefficient. Furthermore, the tail can facilitate intersegment transfer, which is known to accelerate the search for a specific target site on DNA (31–34).

The intersegment transfer of tailed DNA-binding proteins between different DNA regions is achieved by the division of the protein into two moieties (29): a structured moiety and a disordered tail. This construction supports the “monkey-bar” or “brachiation” mechanism whereby the protein jumps between two DNA molecules through an intermediate in which the recognition helix of the protein is adsorbed to one DNA fragment while the disordered N-tail is adsorbed to the other, using a motion that resembles that of a child brachiating along monkey-bars (18). Alternatively, the transfer may be described as following the fly-casting mechanism (35–38), in which the disordered tail reaches DNA regions beyond the near vicinity of the location of the structured moiety of the protein. In multidomain transcription factors, the disordered linker that connects the two folded domains plays a similar role in facilitating brachiation dynamics (34, 39).

Author contributions: D.V. and Y.L. designed research; D.V. and Y.L. performed research; D.V. and Y.L. analyzed data; and D.V. and Y.L. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1] To whom correspondence should be addressed. E-mail: koby.levy@weizmann.ac.il.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1011775107/-/DCSupplemental.

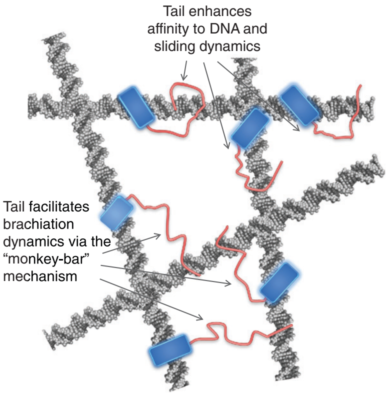
Tail enhances affinity to DNA and sliding dynamics
Tail facilitates brachiation dynamics via the "monkey-bar" mechanism

Fig. 1. A schematic illustration of the effect of the disordered tail on DNA search. The disordered tail (red) of a globular protein (blue) can be viewed as another subdomain that interacts with DNA nonspecifically. The disordered tail can promote sliding dynamics and enhance affinity to the DNA when it interacts with an adjacent segment of DNA. Alternatively, the tail can facilitate brachiation dynamics of intersegment transfer (also called monkey-bar mechanism) in which the protein jumps between distant DNA regions via an intermediate in which the tail and the folded moieties are transiently bound to different DNA segments.

For a disordered tail to promote intersegment transfer events, it must possess certain characteristics. The HoxD9, *Antennapedia* (Antp), and NK-2 *Drosophila* homeodomains search DNA quite differently although they have very similar structured regions. The existence of an N-tail in these homeodomains modulates the properties of the DNA search. These tails are of different lengths (9, 10, and 17 residues respectively) and have different net charges (+2, +4, and +5, respectively). One could conclude that the efficiency of the monkey-bar dynamics will correlate with the length of the tail and with its positive net charge. However, it is clear that the linkage between the sequence of the tail and search efficiency is more complex than that. For example, most of the charged residues of these three N-tails are clustered together and one may conjecture that this clustering serves to form an effective concentrated charge unit that is essential for jumping (18). 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 (40, 41).

In this computational study, we explore the dynamics of hundreds of specifically-designed N-tails of the NK-2 and Antp homeodomains (566- and 55-tail variants, respectively) on DNA to decipher the interplay between the sequence characteristics of disordered tails and the nature of their interactions with nonspecific DNA. In particular, we ask how the net charge of the tail and the organization of the charges along the tail affect search efficiency. Finally, we ask whether the biophysical criteria that are found essential to promoting intersegment transfer are also found in a statistical analysis of natural DNA-binding proteins.

Results and Discussion

The Effect of the Net Charge of the Disordered Tail on DNA Search. Disordered regions in proteins, and particularly the disordered tails of DNA-binding proteins, are rich with positively charged residues (10, 11). The low complexity of the sequence of disordered proteins is reflected not only in their sequence composition (i.e., abundance of some residues) but also in the pattern and periodicity of sequence repetitiveness (11, 42). For example, about 40% of the N-tails of wild-type NK-2 and Antp homeodomains comprise positively charged residues (these tails include six and four positively charged residues and zero and one negatively charged residues, respectively). Furthermore, the charges are clustered: all six positively charged residues in the tail of NK-2 are consecutively linked to form a highly charged segment and, for Antp, three of the four positively charged residues in its tail form a charged segment (Table S1). These positive charges are essential to enabling the DNA-binding proteins to interact with DNA. However, the effects of the number of consecutive charges per segment and the positions of the charges along the tail on DNA search are unknown.

To address these issues, we designed 153 and 55 variants of the N-tails of the NK-2 and Antp homeodomains, respectively. These variants differ in terms of the length of the positively charged segment (segments of 1–17 or 1–10 positive charges for NK-2 and Antp, respectively) and the position of the segments along the tail (see Fig. 2). The effect of the length and position of the charged segment on the nonspecific interactions of the homeodomain with DNA was studied by simulating the dynamics of each variant on two parallel 100 bp DNA molecules (separated by 60 Å) at a wide range of salt concentrations using a simple computational model (18, 39, 43), in which protein-DNA interactions were represented solely by electrostatic forces. We found that the salt concentration for an optimally efficient search was 0.07 M for most NK-2 variants and 0.08 M for most Antp variants.

The fractions of sliding, hopping, and three-dimensional search performed by the 153 NK-2 homeodomain variants with different consecutive positive residue segment lengths and locations on the N-tail are shown in Fig. 2 *A–C*. The colors of the scatter plots represent the percentage of the corresponding search mode used. Short positively charged segments highly populate the three-dimensional search mode, irrespective of their position along the N-tail, while tails that are highly charged make more use of the hopping mode when searching DNA. The sliding search mode is greatly occupied by variants with moderately charged segments (with 5–9 charges) located closer to the N terminus of their tails.

The strong dependence of the usage of hopping and sliding search modes on the effective charge of the N-tail implies that the one-dimensional diffusion coefficient, D_1, is also affected by the tail charges. The value of D_1 for 153 NK-2 variants when they linearly diffuse on a single 100 bp DNA molecule at a salt concentration of 0.01 M (by either sliding or hopping dynamics) increases with decreasing effective charge (Fig. S1) because the number of hopping events increases at the expense of sliding at this salt concentration, with hopping being faster than helically sliding along the DNA backbone (43). Indeed, many tail variants that perform more hopping than sliding, have larger D_1 values. Yet, we found that the rate of linear diffusion is slower for tails with a charged segment of moderate length than for long charged segments, presumably due to cross-talk between the tail and the globular part of the homeodomain that affects its interaction with the major groove and therefore the linear diffusion. Overall, not only the net charge of the tail but also the position of the charges along the tail can affect the protein motion on DNA.

A highly positive charge density on the tail may accelerate intersegment transfer, because it can act as an additional strong DNA recognition motif. Alternatively, the highly positive segment may increase the affinity of the protein for the DNA and thus inhibit intersegment transfer, as was shown by thermodynamic analysis (19, 27). Fig. 2*D* shows the number of intersegment transfers performed by the 153 variants of NK-2. A short charged segment is insufficient to promote transfer from one DNA molecule to another because of its weak attraction to DNA. A tail with a highly charged segment has a very high affinity for nonspecific interactions with DNA, which does not support jumping. Tails that consist of a moderately charged segment located closer to the N-terminal have a moderate affinity to the DNA that allows them to interact with the DNA but also to explore alternative DNA regions and, indeed, the largest number of intersegment transfer events is achieved for variants with a moderately long charged segment located closer to the N-terminal

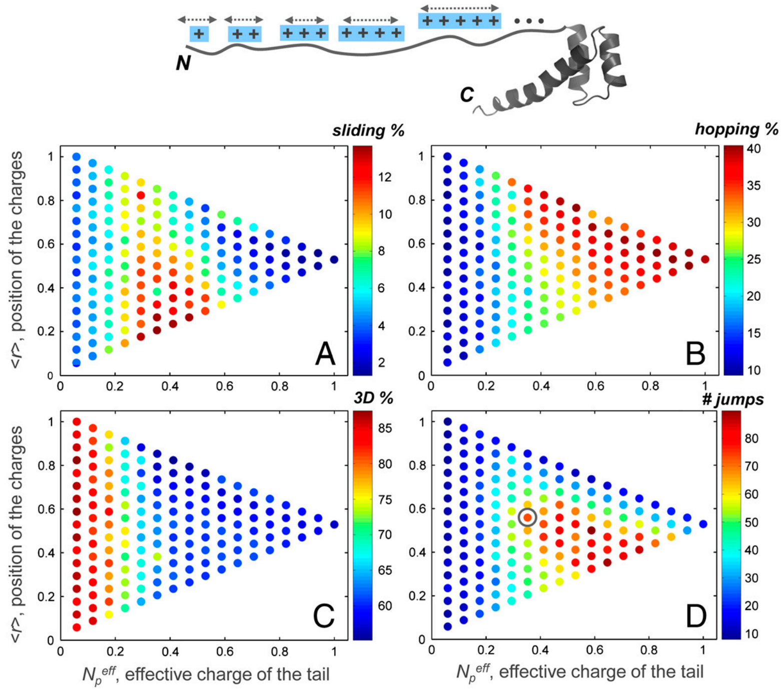

Fig. 2. The effect of the length and location of positively charged segments in the disordered N-tail on the properties of DNA search by the homeodomain. The influence of the effective charge, N_p^{eff}, and the average position, $\langle r \rangle$, of the charged segment of the tail on the interplay between: (*A*) sliding, (*B*) hopping, and (*C*) three-dimensional diffusion search modes for 153 variants of the NK-2 homeodomain on two parallel DNA molecules separated by 60 Å at a salt concentration of 0.07 M. The length of the positive segment was varied from 1–17 residues and it was positioned at all possible positions along the 17 positions of the tail (lower $\langle r \rangle$ values mean that the charged residues are centered closer to the N terminus). Each dot in the figures corresponds to a different tail variant and the color corresponds to percentage use of the relevant searching mode. Note that sliding, hopping, and three-dimensional diffusion sum to 100%. (*D*) The number of intersegment transfer events (i.e., jumps) performed by the 153 variants of the NK-2 homeodomain between two DNA fragments separated by 60 Å. The gray circle indicates the N_p^{eff} and $\langle r \rangle$ values of the N-tail of wild-type NK-2 suggesting that its sequence was evolved to promote monkey-bar mechanism.

(NK-2: Fig 2*D*; Antp: Fig. S2). Although wild-type disordered N-tails can be more complicated than the 153 modeled tails, we note that the disordered N-tails of wild-type NK-2 and Antp homeodomains have a charge pattern that resembles the designed variants that perform the largest numbers of intersegment transfers. For example, the tail of wild-type NK-2 includes an aspartic acid at position #3 and wild-type Antp has a glycine residue (which splits the segment of four positively charged residues into segments comprising three and one residue). Yet, the wild-type tails incorporate a segment of moderate length that supports enhanced intersegment transfer.

Direct transfer from one DNA fragment to another significantly improves the efficiency of target search by a DNA-binding protein (18, 31). Here, we investigated the efficiency of DNA search by 153 variants of NK-2. The search efficiency is estimated by the number of DNA positions probed during the simulations by the homeodomain using sliding (for more details see ref. 43). Optimal search efficiency is achieved in the presence of a moderate effective charge located closer to the N terminus of the variant tail (Fig. S3). This high search efficiency is linked to the high number of intersegment transfer events seen for these NK-2 variants, yet the correlation between search efficiency and number of jumps is not linear (Fig. S3). Variants with a relatively short charged segment in their tail (group I) search DNA more efficiently than other variants although they do not participate in many intersegment transfer events. Variants with a more highly charged segment (group II) tend to search DNA with lower efficiency, although they make greater use of the brachiation mechanism and perform more intersegment transfer events (Fig. S3). Group I variants, which have shorter charged segments than those in group II, have a limited ability to perform intersegment transfer because the tail (which is required for brachiation) has a weak affinity to the DNA, but they can still search efficiently because of their efficient linear diffusion on a single DNA molecule. This observation illustrates that the charged section of the tail is an important component of the search process that can modulate the quality of the search via several mechanisms.

The Effect of Charge Distribution Along the Disordered Tail on DNA Search. We found that the disordered N-tail effectively influences brachiation between different DNA regions if the cluster of charges is of medium length (i.e., 4–7 charges) and located closer to the N terminus. Yet, it is unclear how segmentation of the cluster of positive charges into smaller groups will affect the capability of the homeodomain to brachiate and consequently to search the DNA efficiently. To address these questions, we designed variants of the disordered tail of the NK-2 homeodomain in which six positively charged residues were distributed in various patterns along the tail. There are ten possible ways to segment the six charges (see *Methods* and Fig. 3) and each segmentation pattern can be further varied by introducing spacings of various lengths between the segments. Among all the possibilities, we sampled 253 variants and examined how degree of segmentation and spacing affect the number of intersegment transfer events performed.

Fig. 3 shows the number of intersegment transfer events performed by the 253 variants of NK-2 having tails composed of six positively charged residues. When the six positive charges are more clustered (i.e., less segmented, higher values of C_{indx}) and their average position is closer to the middle or the N terminus of the tail, then the protein jumps more frequently (Fig. 3*A*). This result implies that, as the length of the spaces formed by intervening neutral residues becomes shorter and the total number of spaces decreases (i.e., lower values of S_{indx}), the number of intersegment transfers increases (Fig. 3*B* and Fig. S4).

To examine how the presence of negatively charged residues affects the capability of the homeodomain to jump from one DNA molecule to another, we introduced a negatively charged bead to the tail at position 3, which is occupied by a Glu in the wild-type tail of the NK-2 homeodomain. 160 additional variants with the six positive charges placed in different segmentation and spacing arrangements in the presence of the negative charge were designed. These variants exhibit similar jumping behavior to that observed for the variants without the negative charge (Fig. 3 *C* and *D*), but the geometric options for positioning the positive charges are constrained by the presence of the negative charge. Accordingly, more jumping is seen when the positive charges are more clustered, i.e., when there are fewer and/or shorter spaces between the positive charges. It is interesting to note that wild-type NK-2, which is encircled in the scatter plots of Fig. 3, performs the largest number of intersegment transfers, suggesting that the positioning of the charged residues in the natural tail

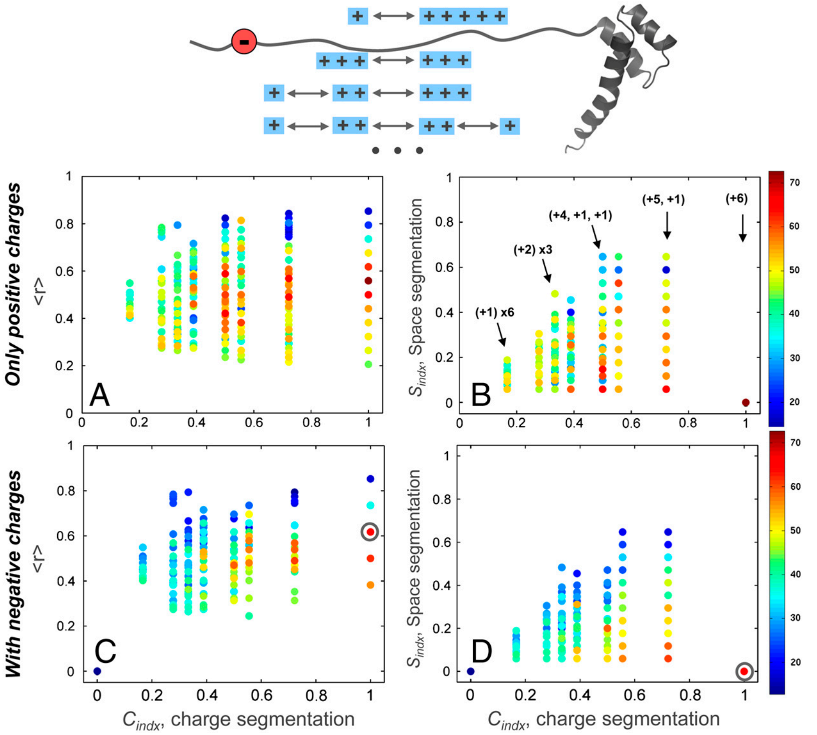

Fig. 3. The effect of charge segmentation and spacing on DNA search. Quantitative characteristics of intersegment transfer by NK-2 variants between two DNA molecules separated by 60 Å at a salt concentration of 0.07 M. All N-tail variants include six positive charges that are divided into different charge segments and positioned at different locations along the tail. There are (*A* and *B*) 253 variants that include only the six positive charges and (*C* and *D*) another 160 variants that include a fixed negative charge at the third residue position. The graphical representation illustrates the construct of the variants with and without the negative charge. All the variants are classified based on their indexes of charge segmentation ($C_{\rm indx}$), spacing ($S_{\rm indx}$), and the average position of the charges along the tail ($\langle r\rangle$). The color in all represents the number of intersegment transfer events. Some segmentation patterns of the charges are illustrated in *B*. For example, the pattern $(+1)\times 6$ corresponds to tails with six charged segments (distributed along the tail with various spacing patterns) each of a single positively charged residue and the pattern $(+6)$ corresponds to a single charged segment of six consecutive residues (which is positioned in different locations along the tail). Note that some variants may have the same $C_{\rm indx}$ and $\langle r\rangle$ values but different $S_{\rm indx}$ values (or the same $C_{\rm indx}$, and $S_{\rm indx}$ values, but different $\langle r\rangle$ values). These variants will overlap and in these cases the variants with the highest number of intersegment transfer events are shown. The gray circles indicate the $C_{\rm indx}$, $\langle r\rangle$, and $S_{\rm indx}$ of the N-tail of wild-type NK-2.

of DNA-binding proteins supports efficient DNA scanning by the brachiation/monkey-bar mechanism.

Our result that the effective charge of the N-tails significantly affects their sliding propensity and the ability of the N-tail to promote jumping via the monkey-bar mechanism expands upon the findings of recent studies by Marsh and Forman-Kay (15) and Mao et al. (16). These studies have shown, for 32 and 21 intrinsically disordered proteins, respectively, that their net charge governs their degree of compaction and structure. Fine analysis of sequence features was not practical in these two experimental studies because of the limited protein set used. While these studies indicate that the compaction of intrinsically disordered proteins depends linearly on the net charge of their tails, we show that the function carried out by these proteins (i.e., the ability of the N-tail to jump between DNA molecules) does not constantly increase with increasing net charge and one may look for a net charge at which the function is optimized. Furthermore, our study suggests that, in addition to the net charge, the organization of the charges and their distribution along the tail can modulate DNA search quite significantly. Accordingly, not only the number of positive charges, but also their positioning, degree of segmentation, and the number and lengths of the spaces between them may appreciably affect the functioning of the tail.

Bioinformatics of the Charge Segmentation in the Disordered Tail of DNA-Binding Proteins. Our observation that the features of the N-tails of wild-type NK-2 and Antp homeodomains are correlated with high numbers of intersegment transfer events may suggest that tails of DNA-binding proteins are selected to have a charge pattern consistent with efficient DNA search. To explore this aspect more thoroughly, we analyzed the sequence of the disordered N-tails of 1,384 homeodomain transcription factors. The net charge of the N-tails is positive and their positive charges are more densely arranged than their negative charges and than the positive charges in the globular parts of the homeodomains (Fig. 4 *A* and *B*). The maximal segment length for positive charges in the N-tail is 0–6 residues, while for negative charges the maximum length is 0–2 residues. In the globular part of the homeodomains, the difference between the distributions of negatively and positively charged segments is much smaller (Fig. 4 *C* and *D*). The spacing between the charged clusters is smaller in the tail than in the globular regions of the homeodomain (Fig. S5).

Conclusions

We have previously shown that DNA-binding proteins tend to have disordered tails that, functioning as another subdomain, engage in nonspecific interactions with DNA as a result of their often positive charge. The disordered tail can both increase the affinity of the globular region to the DNA and facilitate intersegment transfer from one DNA molecule to another as it supports a brachiating motion by the protein. A larger number of jumps will result in a more efficient DNA search. The ability of the disor-

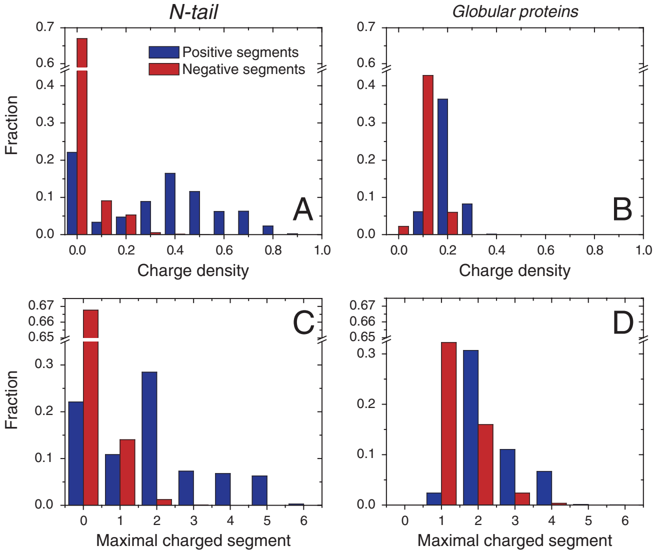

Fig. 4. Bioinformatic analysis of the charge pattern on the N-tails of homeodomains. Distribution of (*A* and *B*) effective positive (blue) and negative (red) charges and (*C* and *D*) maximal absolute charged segments in the disordered N-tails and globular parts of 1,384 wild-type homeodomain DNA-binding proteins.

dered tail to promote jumping depends on its length and net charge (18). Yet, it is clear that the details of the sequence of the tail dictate its nature and its interactions with DNA.

In this article, we report an extensive computational study aimed at dissecting the linkage between the sequence features of the tail and the efficiency of the DNA search by proteins. We designed 566 variants of the 17-residue tail of the NK-2 homeodomain and 55 variants of the 10-residue Antp tail. The variants differed in terms of the number of positive residues in the tail, their position along the tail, the segmentation of the charges, as well as number and length of the spaces between them formed by the presence of intervening neutral and negative residues. We studied the dynamics of each variant as it interacted nonspecifically with two parallel 100 bp DNA molecules (i.e., where the interactions were guided by electrostatic forces only). We find that higher numbers of intersegment transfers is achieved when the tail is significantly but moderately charged (i.e., 6–10 charged residues for the tails of NK-2) and when the charges are concentrated in the middle of the tail or closer to its N terminus. Increased clustering of charges, i.e., the absence of large or numerous spaces between charged residues, enhances the monkey-bar mechanism.

The natural tails of DNA-binding proteins may also include negatively charged residues, although the net charge is positive. We therefore designed a set of tails in which the positive charges were arranged in various positions while the third position was occupied by a negative residue, as in the N-tail of the wild-type NK-2 homeodomain. These variants behave in a similar manner to those that lack the negative charge, although the presence of the negative charge means that a higher effective positive charge is needed to achieve an enhanced jumping mechanism. Clearly, the sequence in wild-type DNA-binding protein tails can be more complex and include more negative charges whose positioning along the tails may affect interaction with the DNA. It is possible that, in such cases, the organization of the positive charges will be different to that reported based on our simple designed tails.

Interestingly, we found that the charge pattern of the N-tail of wild-type NK-2 and Antp homeodomains is consistent with a large number of intersegment transfer events. It is therefore tempting to suggest that natural tails are evolutionarily selected to achieve efficient DNA search. To more thoroughly explore the evolutionary selection of the disordered tail of DNA-binding proteins, we analyzed the charge pattern of the N-tail of 1,384 homeodomains. We found that the N-tails are highly positively charged, their positive charges are highly clustered, and the spacing between the charge clusters is much smaller than between clusters of positive charges in the structured regions of the homeodomains. Overall, our results strengthen recent reports (9, 10) that the sequences of intrinsically disordered regions have unique characteristics that govern their structure and function and suggest that the tails of DNA-binding proteins have been evolutionarily selected to have the charge pattern for optimal DNA search.

Methods

Simulation Model. The molecular and dynamic natures of DNA search by variants of the NK-2 and Antp homeodomains were studied using a reduced model (18, 43) that allows sampling of long time scale processes such as sliding, hopping, three-dimensional diffusion, and intersegment transfer (44–46). We modeled the DNA as having three beads per nucleotide, representing phosphate, sugar, and base. Each bead was located at the geometric center of the group it represents and a negative point charge was assigned to beads representing the DNA phosphate groups.

In the simulations, a 100 bp B-DNA molecule was used to study protein diffusion on a single double-stranded DNA molecule, and two parallel 100 bp B-DNA molecules were used to investigate intersegment transfer dynamics. The protein and the DNA were placed in a box with dimensions 300 × 300 × 700 Å, with the DNA being placed at the center of the box along its Z-axis. The two DNA molecules were separated by 60 Å because the largest number of jumps by homeodomains was observed at this distance (18). The DNA remained in place and rigid throughout the simulations.

The protein was represented by a single bead for each residue located at the Cα of that residue. Beads representing charged amino acids (Lys, Arg, Asp, and Glu) were charged in the model. Unlike the DNA, the protein remained flexible during the simulations and could undergo folding and unfolding events. We simulated the protein with a native topology-based model corresponding to a perfectly funneled energy landscape (47), where native protein interactions were attractive and all other interactions were repulsive (48). 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-Huckel potential, which accounts for the ionic strength of a solute immersed in aqueous solution (43). The dynamics of each protein was studied at salt concentrations in the range of 0.01–0.12 M using a dielectric constant of 80 and a temperature at which the protein is completely folded. This study indicated that the salt concentration for optimal search efficiency is 0.07 M for NK-2 and 0.08 M for Antp homeodomains (similar optimal salt concentrations were obtained for the variants of these proteins in which only the tails were mutated). We emphasize that the information from the crystal structure was used only to model the protein, while the interface between the protein and DNA was modeled solely by electrostatic and repulsive interactions. Accordingly, our model does not include any bias toward the specific binding mode. More details of the simulation can be found in refs. 18, 39, 43, 49.

Using this model, we studied the interaction and structure of several DNA-binding proteins with nonspecific DNA (18, 39, 43). The nonspecific interactions between the protein and DNA were quantified by classifying each snapshot as performing one-dimensional diffusion (either sliding or hopping) or three-dimensional diffusion. Trajectories were analyzed to quantify the percentage of protein sliding and hopping, structural features during sliding, the number of intersegment transfer events, and the linear diffusion coefficient (D_1) as described in ref. 43.

The Designed Variants of Homeodomain Proteins. Hundreds of variants of the NK-2 and Antp homeodomains (the PDB IDs of the wild-type proteins are 1NK2 and 1AHD, respectively) were designed. These variants differed from each other only in terms of the charge properties of their N-tails and were created with the aim of deciphering how the sequence details of the tail can modulate DNA search by the homeodomains. The variants were designed to address two major questions. The first question is how the number of charges on the tail (i.e., the net charge) and their positioning along the tail affect the number of times the homeodomain jumps from one DNA molecule to another. To answer this question, we designed 153 variants of the 17-residue NK-2 tail in which a single cluster composed of 1–17 consecutive positive charges was located at various positions along the tail. For the Antp homeodomain, whose tail comprises 10 residues, 55 variants with different consecutive positive residue cluster lengths and locations on the N-tail were designed as well. The second property of the sequence of the tail that may affect jumping is the distribution of the charges along the tail. To address the relationship between the degree of segmentation of the charges and DNA search, another 160 NK-2 variants were designed. All of these variants included six positive charges segmented in various patterns and separated by different numbers of spaces of different lengths.

Because the N-tail of the wild-type NK-2 homeodomain includes a negatively charged residue at position 3, we designed a third set with 253 variants of the N-tail of NK-2. In these variants, six positive charges were arrayed in different patterns along the tail by varying the number of charges per charged segment and the number and length of the spaces between segments. In all cases, a negative charge was placed at the third position of the tail. This set of variants is similar to the second set of 160 variants with the exception that one fixed negative charge is placed at the third position.

Quantitative Characterization of Intersegment Transfer Events for Tail Variants. The tail variants were classified according to the following parameters, which aimed to capture the sequence complexity of the distribution of the charges along the disordered tail.

1. The effective positive charge, N_p^{eff}, calculated as Np/L, where Np is the total number of charged residues in the N-tail and L is the length of the tail in amino acids
2. The average position of the charges on the tail, $\langle r \rangle$, calculated as $\Sigma r_i/Np \times L$, where r_i is the position of charged residue i along the tail. Values of $\langle r \rangle$ range from 0–1, where a value of zero indicates that the positive charges are located at the N-terminal, whereas a value of 1 indicates that they are located at the C-terminal.
3. The positive charge segmentation index, C_{indx}, calculated as $\Sigma N_i^2/N_p^2$, where N_i is the number of positive charges in segment i. Ten different segmentation patterns of the six charges of the N-tail of NK-2 can be

made. Values of C_{indx} that are close to unity indicate less segmentation and values close to 0 indicate extensive segmentation. The following are the ten possible segmentation patterns and their corresponding C_{indx} values: $([+1\times 6],\ C_{\mathrm{indx}}=0.17)$, $([+2\times 2,\ +1\times 2],\ C_{\mathrm{indx}}=0.28)$, $([+2\times 3],\ C_{\mathrm{indx}}=0.33)$,$([+3,\ +1\times 3],\ C_{\mathrm{indx}}=0.33)$, $([+3,\ +2,\ +1],\ C_{\mathrm{indx}}=0.39)$, $([+3\times 2],\ C_{\mathrm{indx}}=0.5)$, $([+4,\ +1\times 2],\ C_{\mathrm{indx}}=0.5)$, $([+4,\ +2],\ C_{\mathrm{indx}}=0.55)$, $([+5,\ +1],\ C_{\mathrm{indx}}=0.72)$, $([+6],\ C_{\mathrm{indx}}=1.0)$. There are many other alternative formulas to quantify the degree of charge segmentation. We have selected the current formula because it is simple and its values more equally span from 0–1.

4. The spacing index, S_{insx}, describes the spacing of the positively charged segments. $S_{\mathrm{indx}}=S^{\mathrm{eff}}\times \Sigma S_i^2/S^2$, where S^{eff} is the normalized number of spaces within the tail (i.e., $S^{\mathrm{eff}}=S/L$, where S is the total number of spaces in the tail) and S_i is the number of spaces in segment i. A residue is considered to function as a "space" between positive charges if it is negatively charged or neutral. For variants with a single cluster of positive charges (i.e., no spaces) S_{indx} is set to zero, while an S_{indx} value of close to zero indicates the presence of many short spaces between the positive charges a value approaching unity indicates the presence of a few long spaces. Note that S_{indx} gets lower values either by small number of spaces (low S^{eff}) or by high degree of segmentation (there are many spaces but each space S_i is small).

Statistical Analysis of Tail Properties in DNA-Binding Proteins. Sequences of homeodomain proteins were downloaded from the NHGRI site http://genome.nhgri.nih.gov/homeodomain/ (50). Determination of protein tails was performed using IUPred (51). A protein was considered to have a disordered 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 globular region of a homeodomain was defined by excluding the disordered tail from the whole protein sequence. The charge on the tail was calculated by assigning one positive charge (+1) to each Lys or Arg in the tail and one negative charge (−1) to each Glu or Asp. Histograms of effective charge, maximal charged segments, sequence lengths, effective spaces, and charge position were obtained from the analyzed data.

ACKNOWLEDGMENTS. We are grateful to Ariel Azia for fruitful discussions on bioinformatics analysis. This work was supported by the Kimmelman Center for Macromolecular Assemblies and the Minerva Foundation with funding from the Federal German Ministry for Education and Research. Y.L. is the incumbent of the Lillian and George Lyttle Career Development Chair.

1. Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* 293:321–331.
2. Dunker AK, et al. (2001) Intrinsically disordered protein. *J Mol Graph Model* 19:26–59.
3. Tompa P (2002) Intrinsically unstructured proteins. *Trends Biochem Sci* 27:527–533.
4. Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Bio* 6:197–208.
5. Uversky VN (2002) What does it mean to be natively unfolded? *Eur J Biochem* 269:2–12.
6. Liu JG, et al. (2006) Intrinsic disorder in transcription factors. *Biochemistry* 45:6873–6888.
7. Mittag T, et al. (2010) Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase. *Structure* 18:494–506.
8. Mittag T, Kay LE, Forman-Kay JD (2010) Protein dynamics and conformational disorder in molecular recognition. *J Mol Recognit* 23:105–116.
9. Dunker AK, Silman I, Uversky VN, Sussman JL (2008) Function and structure of inherently disordered proteins. *Curr Opin Struct Biol* 18:756–764.
10. Uversky VN, Gillespie JR, Fink AL (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* 41:415–427.
11. Romero P, et al. (2001) Sequence complexity of disordered protein. *Proteins* 42:38–48.
12. Mittag T, Forman-Kay JD (2007) Atomic-level characterization of disordered protein ensembles. *Curr Opin Struct Biol* 17:3–14.
13. Tran HT, Mao A, Pappu RV (2008) Role of backbone-solvent interactions in determining conformational equilibria of intrinsically disordered proteins. *J Am Chem Soc* 130:7380–7392.
14. Eliezer D (2009) Biophysical characterization of intrinsically disordered proteins. *Curr Opin Struct Biol* 19:23–30.
15. Marsh JA, Forman-Kay JD (2010) Sequence determinants of compaction in intrinsically disordered proteins. *Biophysical J* 98:2383–2390.
16. Mao AH, Crick SL, Vitalis A, Chicoine CL, Pappu RV (2010) Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc Natl Acad Sci USA* 107:8183–8188.
17. Muller-Spath S, et al. (2010) Charge interactions can dominate the dimsnions of intrinsically disordered proteins. *Proc Natl Acad Sci USA* 107:14609–14614.
18. Vuzman D, Azia A, Levy Y (2010) Searching DNA via a "Monkey Bar" mechanism: the significance of disordered tails. *J Mol Biol* 396:674–684.
19. Crane-Robinson C, Dragan AI, Privalov PL (2006) The extended arms of DNA-binding domains: a tale of tails. *Trends Biochem Sci* 31:547–552.
20. Gruschus JM, Tsao DHH, Wang LH, Nirenberg M, Ferretti JA (1997) Interactions of the vnd/NK-2 homeodomain with DNA by nuclear magnetic resonance spectroscopy: basis of binding specificity. *Biochemistry* 36:5372–5380.
21. Billeter M, et al. (1993) Determination of the nuclear-magnetic-resonance solution structure of an antennapedia homeodomain-DNA complex. *J Mol Biol* 234:1084–1094.
22. Joshi R, et al. (2007) Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell* 131:530–543.
23. 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.
24. Zeng WL, Andrew DJ, Mathies LD, Horner MA, Scott MP (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.
25. Privalov PL, et al. (2007) What drives proteins into the major or minor grooves of DNA? *J Mol Biol* 365:1–9.
26. Dragan AI, et al. (2006) Forces driving the binding of homeodomains to DNA. *Biochemistry* 45:141–151.
27. Toth-Petroczy A, Simon I, Fuxreiter M, Levy Y (2009) Disordered tails of homeodomains facilitate DNA recognition by providing a trade-off between folding and specific binding. *J Am Chem Soc* 131:15084–15085.
28. Iwahara J, Clore GM (2006) Direct observation of enhanced translocation of a homeodomain between DNA cognate sites by NMR exchange spectroscopy. *J Am Chem Soc* 128:404–405.
29. Iwahara J, Zweckstetter M, Clore GM (2006) NMR structural and kinetic characterization of a homeodomain diffusing and hopping on nonspecific DNA. *Proc Natl Acad Sci USA* 103:15062–15067.
30. Iwahara J, Clore GM (2006) Detecting transient intermediates in macromolecular binding by paramagnetic NMR. *Nature* 440:1227–1230.
31. Hu T, Shklovskii BI (2007) How a protein searches for its specific site on DNA: the role of intersegment transfer. *Phys Rev E* 76:051909.
32. Sheinman M, Kafri Y (2009) The effects of intersegmental transfers on target location by proteins. *Phys Biol* 6:016003.
33. Loverdo C, et al. (2009) Quantifying hopping and jumping in facilitated diffusion of DNA-binding proteins. *Phys Rev Lett* 102:188101.
34. Doucleff M, Clore GM (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.
35. Shoemaker BA, Portman JJ, Wolynes PG (2000) Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc Natl Acad Sci USA* 97:8868–8873.
36. Trizac E, Levy Y, Wolynes P (2010) Capillarity theory for the fly-casting mechanism. *Proc Natl Acad Sci USA* 107:2746–2750.
37. Turjanski AG, Gutkind JS, Best RB, Hummer G (2008) Binding-induced folding of a natively unstructured transcription factor. *PLoS Comput Biol* 4:e1000060.
38. Chen J (2009) Intrinsically disordered p53 extreme C terminus binds to S100B(betabeta) through "fly-casting". *J Am Chem Soc* 131:2088–2089.
39. Vuzman D, Polonsky M, Levy Y (2010) Facilitated DNA search by multi-domain transcription factors: a cross-talk via a flexible linker. *Biophys J* 99:1202–1211.
40. Brendel V, Bucher P, Nourbakhsh IR, Blaisdell BE, Karlin S (1992) Methods and algorithms for statistical-analysis of protein sequences. *Proc Natl Acad Sci USA* 89:2002–2006.
41. Karlin S, Altschul SF (1993) Applications and statistics for multiple high-scoring segments in molecular sequences. *Proc Natl Acad Sci USA* 90:5873–5877.
42. Wootton JC, Federhen S (1996) Analysis of compositionally biased regions in sequence databases. *Methods Enzymol* 266:554–571.
43. Givaty O, Levy Y (2009) Protein sliding along DNA: dynamics and structural characterization. *J Mol Biol* 385:1087–1097.
44. Berg OG, Winter RB, Vonhippel PH (1981) Diffusion-driven mechanisms of protein translocation on nucleic-acids .1. Models and theory. *Biochemistry* 20:6929–6948.
45. Halford SE, Marko JF (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res* 32:3040–3052.
46. Mirny L, Slutsky M (2009) How a protein searches for its site on DNA: the mechanism of facilitated diffusion. *J Phys A-Math Theor* 42:434013–434035.
47. Onuchic JN, Wolynes PG (2004) Theory of protein folding. *Curr Opin Struc Biol* 14:70–75.
48. Levy Y, Onuchic JN, Wolynes PG (2007) Fly-casting in protein-DNA binding: frustration between protein folding and electrostatics facilitates target recognition. *J Am Chem Soc* 129:738–739.
49. Marcovitz A, Levy Y (2009) Arc-repressor dimerization on DNA: folding rate enhancement by colocalization. *Biophysical J* 96:4212–4220.
50. Moreland RT, Ryan JF, Pan C, Baxevanis AD (2009) The homeodomain resource: a comprehensive collection of sequence, structure, interaction, genomic and functional information on the homeodomain protein family. *Database* 2009:1–8 bap004.
51. 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.