Frustration in protein–DNA binding influences conformational switching and target search kinetics

Amir Marcovitz and Yaakov Levy[1]

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

Rapid recognition of DNA target sites involves facilitated diffusion through which alternative sites are searched on genomic DNA. A key mechanism facilitating the localization of the target by a DNA-binding protein (DBP) is one-dimensional diffusion (sliding) in which electrostatic forces attract the protein to the DNA. As the protein reaches its target DNA site, it switches from purely electrostatic binding to a specific set of interactions with the DNA bases that also involves hydrogen bonding and van der Waals forces. High overlap between the DBP patches used for nonspecific and specific interactions with DNA may enable an immediate transition between the two binding modes following target site localization. By contrast, an imperfect overlap may result in greater frustration between the two potentially competing binding modes and consequently slower switching between them. A structural analysis of 125 DBPs indicates frustration between the two binding modes that results in a large difference between the orientations of the protein to the DNA when it slides compared to when it specifically interacts with DNA. Coarse-grained molecular dynamics simulations of in silico designed peptides comprising the full range of frustrations between the two interfaces show slower transition from nonspecific to specific DNA binding as the overlap between the patches involved in the two binding modes decreases. The complex search kinetics may regulate the search by eliminating trapping of the protein in semispecific sites while sliding.

nonspecific binding | energy landscape | protein–DNA interactions

Many DNA-binding proteins (DBPs) recognize short DNA target sequences rapidly and specifically in an enormous genomic background with many sites of similar sequence to the specific target site. Various studies have suggested that proteins locate DNA target sites through a combination of multidimensional search mechanisms that allow a relatively few protein copies to activate or repress genes within reasonable response times (1, 2). DBPs bind with moderate affinity to nonspecific DNA sequences and randomly diffuse along the linear contour of the DNA ("sliding"). Sliding is interrupted by short-range dissociations to neighboring DNA segments ("hopping") as well as dissociations to the bulk and transfers to distant DNA sequences that are spatially close ("intersegment transfer").

The search kinetics for the target site may not be governed solely by the scanning efficiency of nonspecific sites along the genomic DNA: Its complexity can be increased and its pace slowed by other processes. For example, following the arrival of the protein at the target site, additional events may be required for specific DNA recognition to occur (3). This transition from the encounter complex that is stabilized by nonspecific interactions to the specific protein–DNA complex may involve conformational changes to one or both biomolecules. Most likely the transition is also coupled with expulsion of interfacial water molecules (4).

Probing the transition from nonspecific to specific protein–DNA binding is difficult because nonspecific interactions are transient in nature. Several structures of proteins bound to semispecific DNA sequences (i.e., with partial sequence similarity to the specific DNA target) have highlighted that nonspecific interactions are mostly dominated by electrostatic interactions between the positively charged protein side chains and the negatively charged DNA backbone (5–7). This notion is supported by a greater dependence of the nonspecific interactions on salt concentration in comparison with specific protein–DNA complexes (8–10). Generally, DBPs have substantial regions of positive electrostatic patches at their DNA-binding interface that complement the negatively charged DNA (11, 12). Negatively charged amino acids have a lower propensity in protein–DNA interfaces (13), though they might be observed and contribute to specificity by properly orienting the protein relative to the DNA (14) or interact with the DNA through a cation such as Mg^{2+}, as observed in EcoRV endonuclease (15). Specificity in protein–DNA interactions is obtained mostly by a formation of hydrogen bonds between donors and acceptors from protein side chains and DNA bases (10, 16), stabilized by van der Waals and hydrophobic forces, electrostatics, and water-mediated interactions between polar groups (4, 17).

Recent studies indicated a high degree of similarity between specific and nonspecific protein–DNA binding. NMR studies of the HoxD9 homeodomain sliding dynamics and kinetics showed that the protein utilizes similar interfaces for both nonspecific and specific DNA binding and that the positive patch on the protein surface maintains a similar orientation with the DNA in the two binding modes (18). In addition, single molecule studies of DBPs diffusion along DNA indicated a rotation-coupled sliding along the DNA helical path, enabling a secondary structural element of the protein binding site to probe the DNA grooves (19). Molecular dynamic simulations also showed that the electrostatic potential of DBPs is sufficient to orient the protein during sliding at the binding mode of the specific interaction (20, 21).

Although the nonspecific and specific interactions of DBPs with DNA are often very similar, they are not identical. Differences between the two binding modes may originate from imperfect overlap between the patch of positive electrostatic potential that interacts with the DNA backbone and the patch that interacts with the base pairs. The need to switch from the nonspecific upon specific recognition reflects a frustration between the two modes because the protein has to exchange some electrostatic interactions with the DNA backbone with hydrogen bonding with the bases. Higher frustration between the two modes may involve varying degrees of conformational changes in the protein or the DNA (22). For example, the contents of *lac* repressor with DNA are dramatically altered in the transition from nonspecific DNA binding to the complex with its natural operator (7). In particular, base-pair interactions formed in the specific complex at the expense of a partial loss of nonspecific charge–charge interactions induce a tilt in the protein relative to the DNA upon the transition from the nonspecific to the specific complex, as well as distorting the canonical B-DNA form of the DNA (7, 10).

Here we investigated the existence of frustration between non-specific and specific DNA binding by DBPs and its implications for DNA search mechanisms and kinetics. We asked to what extent the nonspecific binding modes of DBPs during sliding are similar to specific binding and in particular focused on whether the transition between the two introduces further kinetic complexity to the process of DNA target search by proteins.

Results and Discussion

Correspondence Between the Specific and Nonspecific Binding Patches on DBPs. Under physiological conditions, DBPs have a reasonable affinity to any DNA sequence. This trait allows the protein to alternate between one-dimensional sliding to hopping and dissociations from the DNA. Positively charged patches are common in DBPs and have unique features compared with positively charged patches found in protein–protein interfaces (11, 12). For example, positively charged patches of DBPs are larger, have a higher α-helical content, greater hydrogen-bonding potency, and a higher degree of evolutionary conservation of positively charged residues than similar patches on non-DNA binding proteins (11).

Here, we analyzed a dataset of 125 DBPs for which the crystal structure with specific DNA sequence is available (12) (see *SI Text*) and explored the dual role the positively charged patch may have as a mediator of specific contact formation at the target DNA site and of an efficient protein translocation process along the DNA. For each DBP in the dataset, a subgroup of residues constitute the patch for specific binding to DNA was defined according to its structure. The positively charged patches of the DBPs were probed using the electrostatic potential of the protein obtained from Poisson–Boltzman calculations (11).

Fig. 1 (*Upper*) shows residues forming the specific DNA-binding patch together with the electrostatic potential of the protein surface. In the DNA-binding domain of Skn-1, both the specific DNA-binding residues and the positively charged region occupy the recognition helix that interact with the DNA major groove

[Fig. 1*A*, Protein Data Bank (PDB) ID code 1skn]. The specific DNA binding residues of the chromosomal protein 7A are located in a β-sheet region, a less common DNA-binding motif, and are less localized in a region with positive electrostatic potential (Fig. 1*B*, PDB ID code 1c8c). In the monomeric lambda repressor, the positive potential at the protein surface is relatively non-uniform (Fig. 1*C*, PDB ID code 1lmb), with the specific DNA binding residues being dispersed among regions with differing electrostatic potentials. Variations in the electrostatic environment of each residue that specifically interacts with the DNA are also illustrated by highlighting neighboring positively (Arg and Lys) and negatively (Glu and Asp) charged residues in the vicinity of each residue of the specific patch (see *Middle*).

The similarity between the specific DNA-binding patch on the DBP and the positively charged nonspecific binding patch, was quantified in two approaches. First, we calculated the overlap between the specific and nonspecific binding patches (denoted as SI-PP overlap), as the percentage of protein residues that specifically bind the DNA and also belong to the DBP largest positive electrostatic patch. Second, we evaluated the surrounding electrostatic environment for each individual residue i, with a measure χ_i that may range from -1 to 1 (for residues fully surrounded by negatively or positively charged residues) (see *Methods*, Eq. 1). For each protein, we then calculated χ_{prot} as the mean of the χ_i values for the residues involved in specific DNA binding. Clearly, a protein whose specific and nonspecific DNA patches overlap or even share a mutual common identity (i.e., fully correspond) will have a higher χ_{prot} value.

Fig. 1 (*Lower*) highlights the heterogeneous χ_i values observed in three different DNA-binding proteins as well as the localization of the specific DNA-binding residues relative to the positive patch region. In Skn-1 DNA-binding domain, with a χ_{prot} of 0.42, the specific DNA binding residues entirely overlap the positive patch region (whose residues are indicated by blue markers at the baseline). In the chromosomal protein 7A, only 63% of its specific DNA binding residues overlap with the positive patch
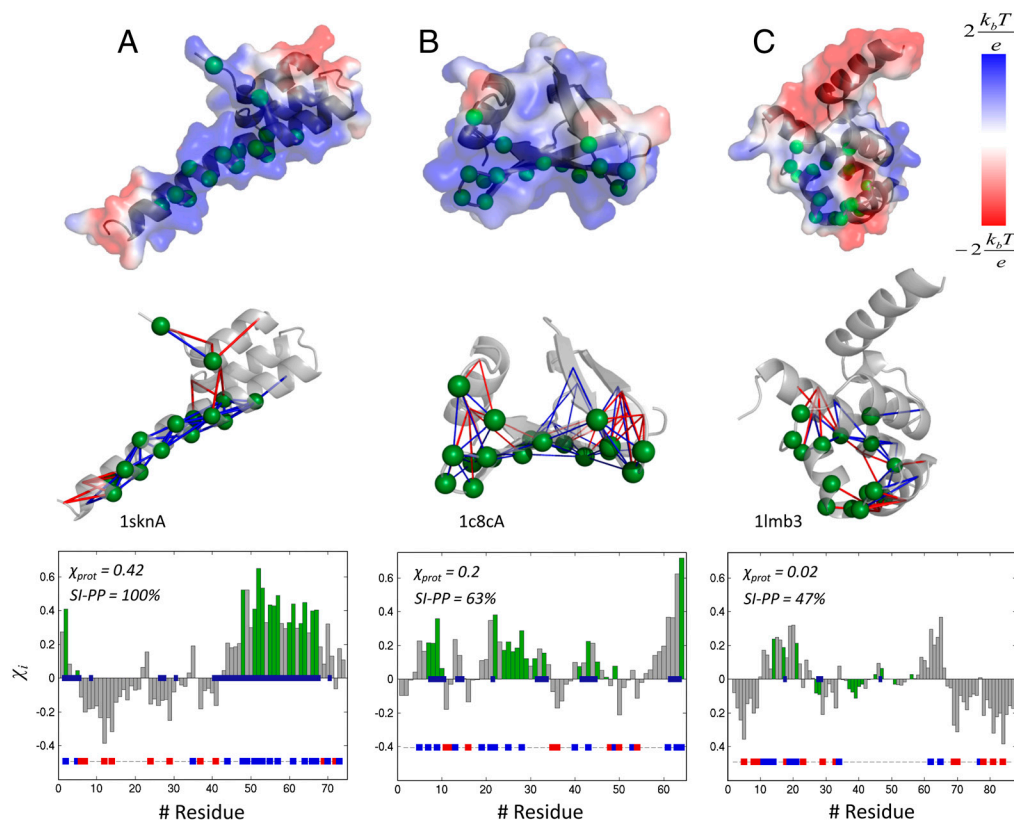

Fig. 1. Electrostatic potential of the patch for specific DNA binding. Three DBPs with varying degrees of correspondence between the patches used for specific and nonspecific binding are shown for the Skn-1 DNA-binding domain 1sknA (*A*), chromosomal protein 7A 1c8cA (*B*), and the monomeric λ-repressor 1lmb3 (*C*). The residues that participate in specific DNA binding are shown as green spheres. (*Upper*) The electrostatic potential calculated using adaptive Poisson–Boltzmann solver (APBS) (35) is shown using a scale that ranges from red to blue. (*Middle*) The positively charged (Arg and Lys) and negatively charged (Asp and Glu) amino acids that surround each of the specific DNA binding residues (with a C^α–C^α distance of <10 Å) are indicated with blue and red connecting lines, respectively. (*Bottom*) The values of χ_i (calculated according to Eq. 1), with green bars indicating the χ_i values for specific DNA binding residues (the mean of the green bars is denoted by χ_{prot}). The percentage overlap between the specific patch and the positively charged electrostatic patch [the largest blue patch (*Upper*)] is denoted by SI-PP.

and some are partially localized near negatively charged regions, giving rise to a χ_{prot} of 0.2. In the monomeric lambda repressor, many specific DNA binding residues are accommodated in the vicinity of negatively charged residues and are excluded from the positive patch region, resulting in a χ_{prot} of approximately 0 and less than 50% overlap with the positive patch.

Fig. 2 summarizes the two quantifications of the structural interplay between specific DNA binding residues and their surrounding electrostatic environment for the entire DBP dataset (divided into seven structural/functional groups). For reference purposes, we analyzed an additional dataset of 37 RNA-binding proteins (RBPs), a protein–protein dataset of 16 heterodimers in which the macromolecular interface is dominated by electrostatic interactions, and a dataset consisting of 129 homodimeric protein complexes in which the interfacial electrostatic observables discussed above are expected to be negligible (see *SI Text* for more details about the protein datasets).

Fig. 2A shows that the positive patch overlap is higher for DBPs than for RBPs and protein–protein complexes. This finding is consistent with the observation that positive patches on DBPs tend to be larger than those on the surfaces of other proteins (11). However, for many DBPs, particularly enzymes, a significant fraction of the specific DNA binding residues is excluded from the positive patch, which results in a reduced overlap.

In Fig. 2B, higher χ_{prot} values are observed for DBPs than for RBPs and protein–protein complexes, as protein interfaces for DNA binding are, on average, more enriched with positive charges than are interfaces for RNA and protein binding. However, DBPs exhibit mostly moderate χ_{prot} values (ranging from approximately 0.05 to 0.25). We find that in many DBPs the observed χ_{prot} is below the maximal χ_{prot} available from an optimal rearrangement of the charges along the sequence (Fig. S1). This indicates a partial frustration between specific and nonspecific DNA binding, as many neutral and negatively charged residues surround residues that specifically bind DNA. Although examples exist for negatively charged residues in the interface (14), these surrounding residues may introduce local electrostatic repulsion from the DNA. DBPs with higher χ_{prot} values are mostly those recognizing DNA through α-helices (for example, leucine-zipper and zinc-coordinating proteins). By contrast, DBPs with low χ_{prot} values, such as enzymes, typically represent more complex DNA-binding sites with mixed positively and negatively charged residues. These observations suggest an implicit conflict within DBPs between their specific and nonspecific DNA-binding modes, which may induce structural differences between the specific binding and sliding conformations as well as raising a kinetic barrier to transition between the two.

Differences Between the Orientations Adopted by Proteins Relative to DNA During Sliding and Specific Binding. During sliding, the protein is attracted to the DNA mostly by electrostatics and most of the interactions exist in specific DNA binding (hydrogen bonds, van der Waals) are absent (9, 10). We ask whether the different protein–DNA interactions observed for sequence specific binding compared with purely electrostatic binding, as quantified and described above, may dictate a different protein orientation relative to the DNA for nonspecific compared with specific binding.

To address this question, we studied the sliding dynamics of four proteins with varying χ_{prot} values. Sliding was studied along a nonspecific 100-bp dsDNA molecule at a low salt concentration. We used a coarse-grained model in which protein–DNA interactions were represented by electrostatic forces only and so mimicked nonspecific interactions (see *Methods*). The optimized protein configuration in sliding was captured by gradually decreasing the temperature. We then compared the orientation of the protein relative to the DNA during sliding with that seen in the experimental specific-sequence protein–DNA complex. Fig. 3 compares the relative orientations between the DBP and the DNA during specific and nonspecific DNA-binding modes. For the purpose of representation, we aligned each specific protein–DNA complex (shown in blue) with its corresponding sliding complex (protein shown in green, DNA in gray). The conformations of the DNA in the two complexes are quite similar, so discrepancies between the two binding modes make little contribution toward distorting the DNA from the canonical B-DNA form.

The nonspecific and specific interactions of the Skn-1 DNA-binding domain ($\chi_{prot} = 0.42$) with DNA are shown in Fig. 3A. The protein maintains a sliding orientation to the DNA that is very similar to the specific binding found in the crystal structure. The recognition helix can therefore efficiently sense the specific target sequence and readily form the network of hydrogen bonds that define the specific complex. We quantify the difference between the specific and nonspecific binding orientations by measuring the distance between the C^α atom of each residue that specifically bind DNA and the nearest DNA backbone phosphate atom in both the specific complex and the sliding conformation (each with its corresponding DNA). The mean difference between the distances in the specific complex and the sliding conformation (denoted as Δd^{sp-nsp}) equals 1.3 Å for Skn-1, which supports the similarity between the two binding modes. Fig. 3 B and C show a comparison between the specific and sliding protein–DNA complexes of two other DBPs: homeobox protein Hox-A9 and Smad-MH1 protein. These proteins have lower χ_{prot} values of 0.35 and 0.18, respectively, which reflect a higher frustration (lower overlap) between their specific binding residues and charged residues. In these two proteins, the sliding orientation of the protein to the DNA differs more markedly from the orientation adopted in the specific complex than is observed for the Skn1 DNA-binding protein, which has a higher value of χ_{prot}. The increasing difference between the nonspecific and specific binding modes observed for these three proteins is reflected in increasing Δd^{sp-nsp} values as χ_{prot} decreases.

A sliding conformation of a monomer from the restriction endonuclease BamHI is shown (Fig. 3D, green) together with the specific enzyme–DNA complex (blue) (23), as well as with a structure of the enzyme bound to a semispecific DNA sequence that differs by a single base pair from the specific sequence (light blue) (6). The low χ_{prot} value of 0.03 for this protein reflects a significant difference between the electrostatic qualities of the
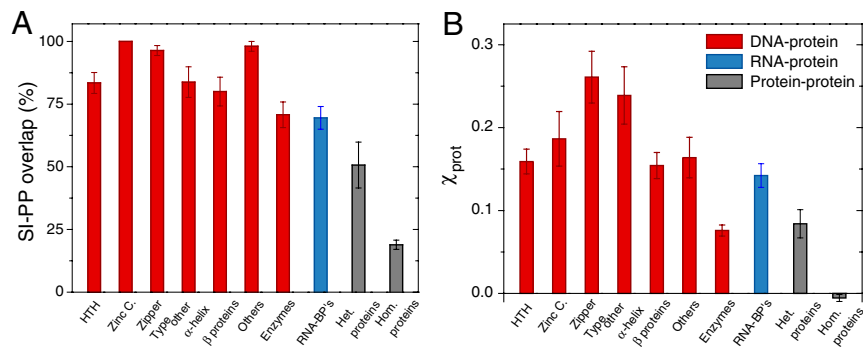
Fig. 2. Frustration between specific and nonspecific interfaces in DBPs, RBPs, and protein–protein complexes. The dataset of DBPs investigated includes 125 proteins (grouped into 7 categories based on their fold or function) and the dataset of RBPs includes 37 proteins. Protein–protein complexes are represented by 129 homodimeric and 16 heterodimeric proteins. The conflict between the specific and nonspecific interfaces is evaluated by (A) the overlap of the residues of the specific interface with the largest positive patch (SI-PP overlap) and (B) by the values of χ_{prot} (see *Methods* for additional details).
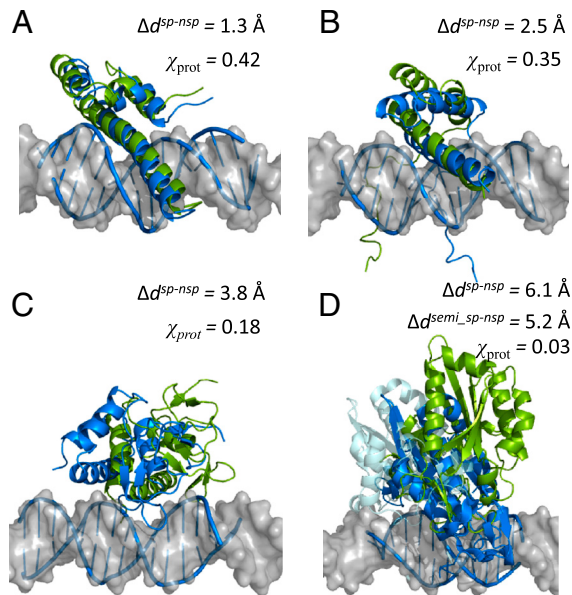
Fig. 3. Comparison of nonspecific protein–DNA binding used during sliding, obtained from coarse-grained simulations (green) to specific binding modes of DBPs obtained from the crystal structure (blue). The DNA used in the simulation is shown in gray. Four proteins that span the whole range of χ_{prot} (0.42–0) found in native DBPs are shown: (*A*) Skn-1 DNA binding (1skn); (*B*) homeobox protein Hox-A9 (1puf); (*C*) Smad-MH1 domain (1ozj); and (*D*) BamHI endonuclease. The similarity between the two binding modes is measured by $\Delta d^{sp\text{-}nsp}$, which is the average of the shortest distance of each residue of the specific interface to the DNA. For BamHI endonuclease 1bhm, additional crystal structure of the enzyme bound to semispecific DNA sequence (light blue, 1esg) is shown to compare the sliding conformation to both specific and semispecific DNA sequences.

patch and the specific protein–DNA contact map. Indeed, the electrostatically favored orientation of the protein relative to the DNA captured in the simulation deviates significantly from both the specific and semispecific protein–DNA complexes. We note, however, that the sliding conformation is more similar to the semispecific complex ($\Delta d^{sp\text{-}nsp}_{semispecific} = 5.2$ Å), than to the specific complex ($\Delta d^{sp\text{-}nsp}_{specific} = 6.1$ Å). In agreement with an earlier study of the sliding of BamH1 on DNA using Poisson–Boltzmann calculations and Brownian dynamics (24), our observation indicates that the semispecific protein–DNA complex lies partway between protein sliding on entirely random sequences and a protein binding complex with its specific sequence.

Transition Kinetics from Sliding to Specific DNA Binding. Structural differences between the nonspecific and specific interactions arising from the frustration between the patches used in each type of DNA binding may have kinetic effects on the overall search process for the target site. One can envision a time gap between localizing the target site and binding tightly to it during which the protein will have to switch its conformation. The time gap might be longer where the frustration between the two binding modes is larger, and this could reduce the likelihood of a successful switch from the nonspecific to specific binding mode.

The interplay between search kinetics and the degree of frustration between specific and nonspecific binding was studied by investigating the time scale for switching from nonspecific to specific binding using the coarse-grained model described above. We used the recognition helix of MAD (composed of 26 residues) from the heterodimer MAD-MAX that contains seven positively charged residues and three negatively charged residues. In addition to nonspecific interactions, the helix may bind a DNA target site (located at the center of the 100-bp DNA) defined based on the crystal structure, by forming specific interactions that are represented by the Lennard–Jones potential.

The helix translocation along the DNA (aligned along the *Z* axis) was monitored during the simulation, as were the formation of specific protein–DNA contacts with the target site (Fig. 4*A*). Fig. 4*B* exemplifies a typical sliding–binding trajectory for the wild-type recognition helix of MAD, for which $\chi_{prot} = 0.1$, indicating a limited overlap of the positively charged electrostatic patch that dominates the sliding mode with the patch for specific protein–DNA interactions. Hence, although the protein locates the target (at $Z = 0$) relatively fast, its orientation is inappropriate for specific DNA binding, and it only forms a partial complex with the DNA (only up to 50% of the protein–DNA contacts are formed). Indeed, Fig. 4*B* shows the helix repeatedly undergoing several dissociations from the target, engaging in additional sliding, and then reassociating with the target again, until the protein–DNA interface is fully formed (defined as >75% formation of the specific protein–DNA contacts).

To study binding kinetics, we repeated the simulation 100 times and measured two time periods in each run: the time elapsed from the start of the search until target site localization (τ_1) and the time until specific binding (i.e., formation of >75% of the specific contacts) (τ_2). Although in some of the simulations target localization was rapidly followed by target binding (i.e., $\tau_2 \sim \tau_1$), in most cases, complete binding to the DNA target significantly lagged behind target localization (i.e., $\tau_2 > \tau_1$) and the overall correlation between the two time scales was weak ($R^2 = 0.2$) (Fig. 4*B*, *Right*). A time gap between arrival at the specific site and binding to it implies the existence of a kinetic barrier in the switch from sliding to the specific binding. It further suggests that alternative protein–DNA nonspecific interfaces may have lower frustration with the specific DNA-binding residues and, thus, lower kinetic barrier that governs the transition from nonspecific to specific target binding.

To test this implication, we designed in silico a library of 150 variants of the recognition helix of MAD, in which we modified the composition and distribution of positively and negatively charged residues along the sequence while maintaining the protein–DNA specific contact map unmodified (Fig. S2). For each variant, we calculated its χ_{prot} and performed 100 sliding/binding simulations as described above. The effects of flanking nonspecific DNA on τ_1 and τ_2 were uniform for all peptides as all of the simulations started with the helix initially positioned near the DNA edge. Fig. 4*C* shows a typical sliding–binding trajectory for a variant of the recognition helix of MAD with $\chi_{prot} = 0.4$ (peptide 33; see Fig. S2). In this variant, a relatively stable partial protein–DNA complex at the target sequence precedes the formation of the complete protein–DNA interface. Values for τ_2 and τ_1 are relatively similar with only a few outliers ($R^2 = 0.6$). Fig. 4*D* demonstrates the binding kinetics a variant with $\chi_{prot} = 0.6$ (peptide 19; see Fig. S2). The high χ_{prot} value of this peptide implies that it can bind specifically to the DNA while maintaining charge–charge interactions with the DNA similar to those it has in the sliding mode. Consequently, complete binding of this peptide variant to the target DNA sequence takes place rapidly upon localization of the target by the protein, with $\tau_2 \sim \tau_1$ in all the simulations ($R^2 \sim 1$). Overall, these observations (Fig. 4) imply that lower frustration between the specific DNA-binding residues and the charged residues in the protein, i.e., increased χ_{prot} of the protein, may reduce the kinetic barrier to transition from the sliding mode to specific target binding.

The histogram in Fig. 5 summarizes the interplay between search kinetics and the degree of frustration between sliding and specific binding to DNA. It shows the average values of the Pearson correlation coefficient between τ_2 and τ_1 against the χ_{prot} values of the simulated peptide library. Higher correlation between arrival and binding times is found for peptides with lower frustration between their two DNA binding modes (i.e., higher χ_{prot} values). The decrease of time gap between target localization and binding in peptides with high χ_{prot} values is
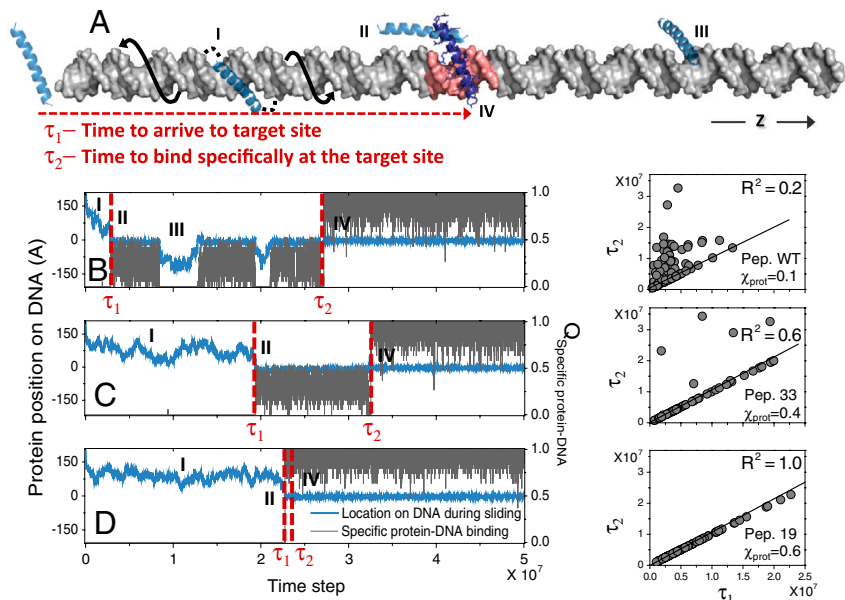
Fig. 4. Kinetics of the transition from nonspecific to specific DNA binding. (*A*) Illustration of a sliding–binding simulation for the recognition helix of MAD from the MAD-MAX heterodimer. The searching protein, shown in lucid blue, is initially positioned near the edge of the 100-bp dsDNA and starts sliding along the helical DNA path on nonspecific segments (I). It then locates the binding site, but the specific protein–DNA interface is not yet formed (II). The time required from the start of the search until the protein locates its binding site, τ_1, is then recorded. The protein can then either switch from the trapped state to form high affinity interactions with the target site (IV) or continue to diffuse to other regions of the DNA (III) until it again locates the target and eventually fully binds to it (IV). The time elapsed from the start of the search until formation of the specific interface, τ_2, is then recorded. Trajectories of three variant peptides with different χ_{prot} are shown: (*A*) wild-type peptide ($\chi_{prot} = 0.1$), (*B*) peptide 33 ($\chi_{prot} = 0.4$), and (*C*) peptide 19 ($\chi_{prot} = 0.6$) (see Fig. S1). Each trajectory is analyzed in terms of the location of the protein along the DNA (blue) and the fraction of specific contacts formed at the binding site (gray). The times τ_1 and τ_2 are indicated by red dashed lines. The correlations between τ_1 and τ_2 from 100 simulations of each peptide system are shown and indicate greater correlation as χ_{prot} increases.

reflected in Fig. 5, *Inset*. Returning to the main histogram, we note that the blue-shaded region indicates the χ_{prot} range for the natural DBP dataset described above, in which the value of the wild-type recognition helix resides (red symbol). It is therefore evident that most DBPs may experience a kinetic barrier in switching from nonspecific to specific DNA binding because of significant frustration in the DNA-binding interface between specific DNA binding residues and charged residues.

Conclusions

The observation that the interface used by DBPs to specifically recognize DNA targets is often accommodated in a patch of positive electrostatic potential used for nonspecific interactions may suggest that both binding modes share similar structural properties. Although the degree of overlap between the positively charged patch and the patch that participates in hydrogen bonding is greater in DBPs comparing to RBPs or protein–protein

Fig. 5. Interplay between frustration and the kinetics of target site search. Histogram of R^2 (the correlation coefficient between τ_1 and τ_2; see Fig. 4) for the χ_{prot} values of the library of 150 MAD recognition helix variants is shown. Target localization and binding are significantly more correlated ($R^2 \sim 1$) in proteins with higher χ_{prot} values. The region shaded in blue marks the range of χ_{prot} values obtained from the dataset of natural DBPs, and the red circle corresponds to the wild-type peptide of the MAD recognition helix. Peptides with lower χ_{prot} values (i.e., with higher frustration between their specific and nonspecific binding modes) are characterized not only by a weak correlation between τ_1 and τ_2 but also by a gap between the two time scales (*Inset*) that disappear for less frustrated peptides.

complexes with a highly charged interface, it is nevertheless imperfect. The smaller the overlap is between the residues forming specific interactions and the positively charged patch (i.e., lower χ_{prot}), the larger the difference between the sliding and specific complexes and the greater the degree of frustration between the two binding modes to DNA.

Binding to the target site may therefore impose a free energy barrier in switching from the nonspecific orientation to the specific conformation. In support of this proposition, our comparison of BamHI endonuclease monomer bound to its specific, semispecific, and nonspecific DNA sequences indicates that the orientation of the protein to the DNA when bound to the semispecific sequence lies between the orientations adopted on completely random sequences and in specific complexes. A utilization of distinct DNA-binding modes for search and recognition has been suggested earlier to resolve the so-called "speed–stability" paradox as it allows the reduction of the ruggedness of the landscape for protein sliding (25, 29). Additional evidence for the existence of a multistate model of interactions with DNA can be found in several X-ray and NMR studies (7, 26, 27) and computer simulations (24, 28). The reorientation of the protein that is governed by the electrostatic interactions with the DNA is reminiscent of the electrostatic preorganization that is found in several cases in enzymatic catalysis in which the solvent reorients its dipoles to promote the reaction (30, 31).

To quantify the effect of the degree of overlap between the specific and nonspecific patches on search kinetics, we designed in silico a library of peptides that span a wide range of degrees of frustration between the specific and nonspecific DNA-binding interactions. We show that a time gap emerges between target localization and target binding due to a conflict between the two residue sets involved in each DNA-binding mode. A high degree of overlap results in almost immediate binding in what can be envisioned as a "lock and key" mechanism. However, limited overlap usually introduces additional kinetic steps for DNA binding during which the protein can become trapped in an intermediate complex with the target, and would switch to the specific orientation by forming hydrogen bonds with the DNA bases only at the expanse of several electrostatic interactions with the sugar-phosphate backbone. An example for such multistate kinetics of DNA sequence recognition has been described earlier for the C-terminal domain of the papillomavirus E2 protein with its target DNA sequence (3). The transition from nonspecific to
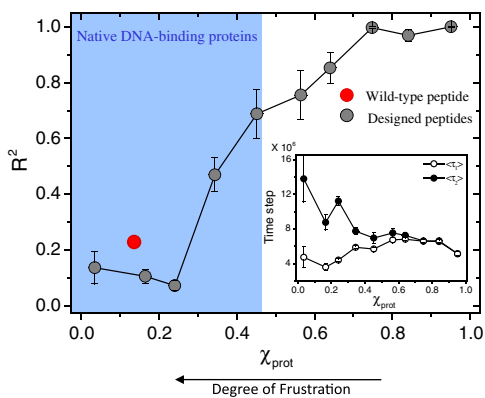
specific binding might also be delayed by subsequent sliding periods following unproductive trapping. This scenario suggests that a selectivity principle operates whereby proteins avoid "wasteful" commitments to semispecific sites as they search the genome; thus, the ruggedness of the energy landscape for sliding is moderate and the overall search is rapid (25). Commitment to the specific site requires transition from the sliding to specific binding mode across a kinetic barrier in what could be considered an "induced fit" mechanism. This may involve at least reorientation of the protein relative to the DNA as observed for Dam methyltransferase substrate recognition (27) and, in some cases, also an intramolecular conformational change to the protein (22), the DNA (32), or both (7), although these were not addressed in the current study.

Our study serves as another example of the important role of frustration in biomolecular systems (33, 34). The prevalence of frustration in DBPs results in bimodal association with DNA and suggests an additional degree of kinetic complexity to the DNA target search problem. Clearly, additional factors can enrich the complexity of the facilitated diffusion that governs the search (e.g., the existence of roadblock proteins or packing of the DNA by the nucleosomes) that, on the one hand, may result in slower kinetics but, on the other hand, may introduce functional regulation. Zero frustration between specific and nonspecific binding would result in sliding that is interrupted by trapping, whereas a high degree of frustration would result in an extremely large barrier for switching. Accordingly, weak frustration is the optimal scenario for an efficient DNA search.

Methods

Calculation of χ_{prot}. We first calculated χ_i for each residue i in the protein as follows:

$$\chi_i = \left(\sum_{j'} q_{j'} \exp\left(-a \frac{r_{ij'}}{r_c} \right) \right) \bigg/ \left(\sum_{j'} \exp\left(-a \frac{r_{ij'}}{r_c} \right) \right) \quad [1]$$

where j' denotes all the residues having their C^α atom closer than a cutoff distance of $r_c = 10$ Å to the C^α atom of residue i. $q_{j'}$ is a point charge of 1, –1, or 0, and $r_{ij'}$ is the C^α–C^α distance between residues i and j'. $a = 2$ is an exponential decay constant. A χ_i value approaching 1 is therefore expected for residues that are fully surrounded by positively charged residues. The value of χ_{prot} is the mean of the χ_i values for protein residues that belong to the interface used for specific binding to the DNA.

Simulation Model. The dynamics of sliding along DNA was studied using a coarse-grained model used previously to explore various molecular aspects of the mechanism of DNA search (20) (see *SI Text* for more details). We studied the interaction of the recognition helix of MAD from the heterodimer MAD-MAX with DNA. To study how it makes the transition from its sliding to specific binding mode, we inserted the specific sequence of the DNA for MAD at the center of the 100-bp DNA molecule. One hundred simulations were performed for the wild-type and the 150 designed peptide variants (Fig. S2), starting at the exact same initial helix position to allow unified measurements for τ_1 and τ_2 in all of the simulations.

1. von Hippel PH, Berg OG (1989) Facilitated target location in biological systems. *J Biol Chem* 264:675–678.
2. Halford SE, Marko JF (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res* 32:3040–3052.
3. Sanchez IE, Ferreiro DU, Dellarole M, de Prat-Gay G (2010) Experimental snapshots of a protein-DNA binding landscape. *Proc Natl Acad Sci USA* 107:7751–7756.
4. Lundback T, Hard T (1996) Sequence-specific DNA-binding dominated by dehydration. *Proc Natl Acad Sci USA* 93:4754–4759.
5. Winkler FK, et al. (1993) The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *EMBO J* 12:1781–1795.
6. Viadiu H, Aggarwal AK (2000) Structure of BamHI bound to nonspecific DNA: A model for DNA sliding. *Mol Cell* 5:889–895.
7. Kalodimos CG, et al. (2004) Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. *Science* 305:386–389.
8. Misra VK, Hecht JL, Yang AS, Honig B (1998) Electrostatic contributions to the binding free energy of the lambdaCI repressor to DNA. *Biophys J* 75:2262–2273.
9. Record MT, Ha JH, Fisher MA (1991) Analysis of equilibrium and kinetic measurements to determine thermodynamic origins of stability and specificity and mechanism of formation of site-specific complexes between proteins and helical DNA. *Methods Enzymol* 208:291–343.
10. von Hippel PH (2007) From "simple" DNA-protein interactions to the macromolecular machines of gene expression. *Annu Rev Biophys Biomol Struct* 36:79–105.
11. Stawiski EW, Gregoret LM, Mandel-Gutfreund Y (2003) Annotating nucleic acid-binding function based on protein structure. *J Mol Biol* 326:1065–1079.
12. Szilagyi A, Skolnick J (2006) Efficient prediction of nucleic acid binding function from low-resolution protein structures. *J Mol Biol* 358:922–933.
13. Jones S, van Heyningen P, Berman HM, Thornton JM (1999) Protein-DNA interactions: A structural analysis. *J Mol Biol* 287:877–896.
14. Lamers MH, et al. (2000) The crystal structure of DNA mismatch repair protein MutS binding to a G · T mismatch. *Nature* 407:711–717.
15. Kostrewa D, Winkler F (1995) Mg^{2+} binding to the active site of EcoRV endonuclease: A crystallographic study of complexes with substrate and product DNA at 2A resolution. *Biochemistry* 34:683–696.
16. Rohs R, et al. (2010) Origins of specificity in protein-DNA recognition. *Annu Rev Biochem* 79:233–269.
17. Luscombe NM, Laskowski RA, Thornton JM (2001) Amino acid-base interactions: A three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res* 29:2860–2874.
18. 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.
19. Blainey PC, et al. (2009) Nonspecifically bound proteins spin while diffusing along DNA. *Nat Struct Mol Biol* 16:1224–1229.
20. Givaty O, Levy Y (2009) Protein sliding along DNA: Dynamics and structural characterization. *J Mol Biol* 385:1087–1097.
21. Liu H, Shi Y, Chen X, Warshel A (2009) Simulating the electrostatic guidance of the vectorial translocations in hexameric helicases and translocases. *Proc Natl Acad Sci USA* 106:7449–7454.
22. Spolar RS, Record MT, Jr (1994) Coupling of local folding to site-specific binding of proteins to DNA. *Science* 263:777–784.
23. Newman M, Strzelecka T, Dorner LF, Schildkraut I, Aggarwal AK (1995) Structure of Bam HI endonuclease bound to DNA: Partial folding and unfolding on DNA binding. *Science* 269:656–663.
24. Sun J, Viadiu H, Aggarwal AK, Weinstein H (2003) Energetic and structural considerations for the mechanism of protein sliding along DNA in the nonspecific BamHI-DNA complex. *Biophys J* 84:3317–3325.
25. Slutsky M, Mirny LA (2004) Kinetics of protein-DNA interaction: Facilitated target location in sequence-dependent potential. *Biophys J* 87:4021–4035.
26. Friedman JI, Majumdar A, Stivers JT (2009) Nontarget DNA binding shapes the dynamic landscape for enzymatic recognition of DNA damage. *Nucleic Acids Res* 37:3493–3500.
27. Horton JR, Liebert K, Hattman S, Jeltsch A, Cheng XD (2005) Transition from nonspecific to specific DNA interactions along the substrate-recognition pathway of Dam methyltransferase. *Cell* 121:349–361.
28. Chen C, Pettitt M (2011) The binding process of a nonspecific enzyme with DNA. *Biophys J* 101:1139–1147.
29. Zhou HX (2011) Rapid search for specific sites on DNA through conformational switch of nonspecifically bound proteins. *Proc Natl Acad Sci USA* 108:8651–8656.
30. Warshel A (1978) Energetics of enzyme catalysis. *Proc Natl Acad Sci USA* 75:5250–5254.
31. Kamerlin S, Sharma P, Chu Z, Warshel A (2010) Ketosteroid isonearse provides further support for the idea that enzymes work by electrostatic preorganization. *Proc Natl Acad Sci USA* 107:4075–4080.
32. Bouvier B, Zakrzewska K, Lavery R (2011) Protein-DNA recognition triggered by a DNA conformational switch. *Angew Chem Int Ed Engl*, in press.
33. Zhuravlev PI, Papoian GA (2010) Protein functional landscapes, dynamics, allostery: A tortuous path towards a universal theoretical framework. *Q Rev Biophys* 43:295–332.
34. Ferreiro D, Hegler J, Komives E, Wolynes P (2011) On the role of frustration in the energy landscapes of allosteric proteins. *Proc Natl Acad Sci USA* 108:3499–3503.
35. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc Natl Acad Sci USA* 98:10037–10041.