Nucleic Acids Research, 2024, **52**, 6763–6776
https://doi.org/10.1093/nar/gkae333
Advance access publication date: 8 May 2024
Computational Biology

Skipping events impose repeated binding attempts: profound kinetic implications of protein–DNA conformational changes

Elena Rogoulenko and Yaakov Levy *

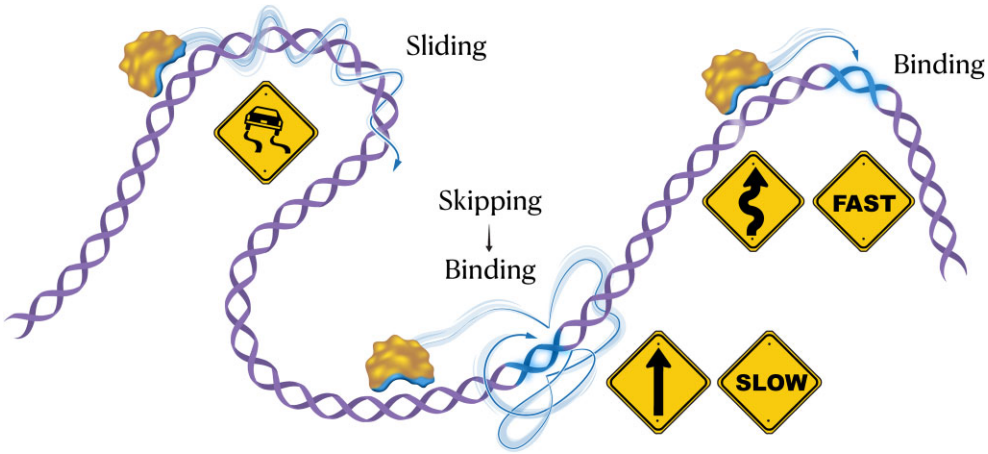
Department of Chemical and Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel

*To whom correspondence should be addressed. Tel: +972 8 9344587; Email: Koby.Levy@weizmann.ac.il

Abstract

The kinetics of protein–DNA recognition, along with its thermodynamic properties, including affinity and specificity, play a central role in shaping biological function. Protein–DNA recognition kinetics are characterized by two key elements: the time taken to locate the target site amid various nonspecific alternatives; and the kinetics involved in the recognition process, which may necessitate overcoming an energetic barrier. In this study, we developed a coarse-grained (CG) model to investigate interactions between a transcription factor called the sex-determining region Y (SRY) protein and DNA, in order to probe how DNA conformational changes affect SRY–DNA recognition and binding kinetics. We find that, not only does a requirement for such a conformational DNA transition correspond to a higher energetic barrier for binding and therefore slower kinetics, it may further impede the recognition kinetics by increasing unsuccessful binding events (skipping events) where the protein partially binds its DNA target site but fails to form the specific protein–DNA complex. Such skipping events impose the need for additional cycles protein search of nonspecific DNA sites, thus significantly extending the overall recognition time. Our results highlight a trade-off between the speed with which the protein scans nonspecific DNA and the rate at which the protein recognizes its specific target site. Finally, we examine molecular approaches potentially adopted by natural systems to enhance protein–DNA recognition despite its intrinsically slow kinetics.

Graphical abstract



Introduction

DNA transcription plays a pivotal role in shaping numerous cellular activities. The essentiality of transcription necessitates rigorous regulation, which is accomplished through the orchestrated action of many proteins, known as transcription factors (1,2). The biophysical properties of these transcription factors are fundamental for achieving the required level of regulation. The relative affinity and kinetics of their binding to DNA, both at specific binding sites and at undesired non-specific sites, are essential in achieving precise regulation.

The molecular intricacies influencing the affinities of transcription factors for DNA are sufficiently complex that predicting them is a non-trivial task (3–5). This complexity is compounded by various molecular factors, including the subtle characteristics of the DNA site. These subtleties can impact binding affinity and specificity through a DNA shape readout that complements the base readout (6). The kinetics of binding the DNA target site is also sensitive to numerous molecular properties of both the protein and the DNA, and especially the interplay between them.

Received: January 30, 2024. **Revised:** April 9, 2024. **Editorial Decision:** April 11, 2024. **Accepted:** April 16, 2024

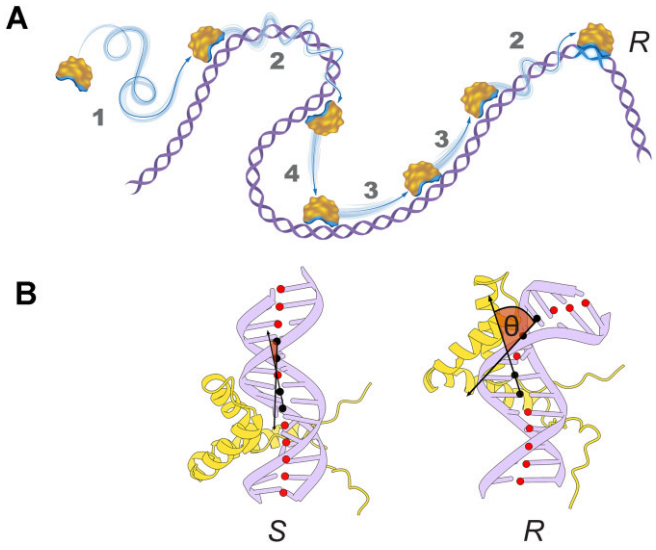


Figure 1. Illustration of the DNA search performed by a DNA-binding protein (DBP) prior to recognition of the target site. (**A**) Extensive scanning of non-specific DNA sites (colored purple) occurs before the DBP recognizes its specific DNA binding site (colored in blue). The DNA search may involve the mechanisms numbered on the figure: 1) 3D diffusion of the DBP in solution is followed by linear diffusion of the DBP along the DNA by 2) rotation-coupled translation (also called sliding), 3) rotation-uncoupled translation (also called hopping), or 4) jumping between two distant DNA sites (also called intersegmental transfer). Linear diffusion can be intermittently interrupted by 3D diffusion in solution, followed by another round of linear diffusion along the DNA. The DBP–DNA binding mode adopted can differ between the search (*S*) and recognition (*R*) states. In the *S* state, the DBP interacts with nonspecific DNA sites in a manner that is governed by electrostatic interactions. In the *R* state, the DBP interacts with a unique DNA sequence via short-range interactions (e.g. hydrogen bonds and hydrophobic interactions). In the *R* state, the conformations of the DNA and/or the DBP may differ from the conformations adopted in the *S* state, with this difference representing an energetic barrier that separates the *S* and *R* states. (**B**) The *S* and *R* binding modes of the SRY transcription factor with nonspecific and specific DNA sequences, respectively. The *S* state is approximated based on coarse-grained (CG) modelling, assuming it is stabilized solely by electrostatic interactions whereas the *R* state is based on an NMR structure of SRY with DNA (pdb: 1j46). In the *S* state, which is defined based on the coarse-grained modeling of nonspecific interactions between SRY and straight DNA (See Supplementary Figure S2), the SRY sits in the major groove of a straight B-DNA conformation, whereas in the *R* state the SRY interacts with bent DNA and sits in its minor groove. Local DNA bending was assessed based on the helical axis of the DNA molecule. The axis is defined as the line connecting the centers of each pair of phosphate beads in a CG-model (red circles). To compare the structures adopted by DNA in the *S* and *R* states, the central angle (defined by the black circles), θ, is calculated and exhibits a change from 10° to 64°. This change of 54° reflects DNA bending supported by SRY binding.

The kinetics of binding of DNA-binding proteins (DBPs) to their specific DNA target sites deviate from other bimolecular reactions that are typically governed by three-dimensional (3D) diffusion for the formation of an encounter complex (7). The presence of numerous alternative nonspecific binding sites, each with weak, but non-negligible affinity to DBPs, and the presence of long-range electrostatic interactions limit the time DBPs spend in solution. These forces, which dominant binding to nonspecific sites, enable linear diffusion along DNA (Figure 1A). Consequently, it was proposed that DBPs search DNA for their specific binding site via one-dimensional (1D) diffusion along the DNA strand, with use of a lower dimensional space (compared with 3D diffusion) explaining the fast observed kinetics (8–11). This proposal has since been supported by a plethora of experimental data across various proteins (12–19).

Computational studies reveal that two distinct modes define 1D search: sliding and hopping (20–25). In sliding, the DBP translocates along the DNA axis, akin to it being twisted around a screw, and also moves to an adjacent site within the same DNA segment. In hopping, the DBP moves to an entirely different DNA segment. The relative prevalences of sliding and hopping in 1D diffusion are determined by salt concentration in the surrounding medium and by the intrinsic molecular properties of the DBP (7). For instance, the sliding mode demands a substantially large positive patch on the DBP as well as a DNA geometry that supports interaction of the DBP with the major DNA groove. As such, DBPs lacking suitable molecular properties may exclusively use the hopping search mode when interacting with nonspecific DNA (26–30). Other molecular properties of DBPs, such as the oligomeric state (31) and the existence of multi-domains (32,33) can also impact diffusion. Particularly, disordered regions may facilitate search by allowing jumping of long DNA loops via the 'monkey-bar' mechanism (20,34,35), which enables intersegmental transfer. In addition to the evolutionary 'positive design' of DBPs to perform efficient DNA search, DBPs should incorporate aspects of 'negative design' to avoid energetic traps formed by their being overly attracted to semi-specific sites (36,37).

After engaging in an extensive search for its DNA target site within genomic DNA, the DBP must engage in specific bonding, which may further slow-down the overall kinetics of DBP–DNA recognition. The timescale associated with the recognition phase may be linked to conformational changes affecting the DBP and the DNA, which may be coupled with a large free energy barrier and thus introduce further kinetic complexity. From the perspective of the protein, the simplest molecular origin for an energetic barrier involved in binding arises from reorientation of the DBP upon the transition from the *S* mode to the *R* mode. It was suggested that, in this scenario, there is a tradeoff between the timescales of the 1D search and of binding that can be addressed by optimizing the molecular properties of the DBP (38). The recognition timescale lengthens when a conformational change, such as DNA bending or DBP distortion, is involved (37–40, 36, 41–43). The shape and flexibility of DNA are essential determinants that influence the search for the binding site and the actual binding (44–46). Different DNA sequences confer various intrinsic properties, including promoting bending (47,48) or even dictating local shape preferences that can modulate protein-DNA recognition (6,49,50) and affect the protein's affinity for its binding site (51–53).

In this study, we examine how the propensity of DNA to change conformation impacts the kinetics of protein recognition of specific binding sites and consequently the overall search time to find and bind the target site. Using a coarse-grained (CG) model of DNA capable of transitioning between structural states, we aim to understand how these transitions influence the search for specific sites and subsequent binding. We address these questions for the SRY (sex-determining region of the Y chromosome) transcription factor, which binds to a bent DNA while interacting specifically with the DNA minor groove (Figure 1B).

The SRY transcription factor, a member of the SRY-type high mobility group (HMG) box protein family, plays an essential role in testis development and male sex determination (54). Structural analyses of the proteins from this family have shown that they bind DNA in its minor groove, which causes the DNA molecule to bend away from the protein (55–58). Human SRY is comprised of three domains: a central HMG-box domain (composed of 70 residues and, for simplicity, referred to here as the SRY domain), which is highly conserved across mammals and directly involved in binding, and two additional domains, the N- and the C-terminal domains. These latter two domains exhibit low conservation, and their contribution to DNA binding is considered minimal. SRY requires a specific DNA conformation, specifically a bent conformation, to function effectively (59). This distinct characteristic qualifies it as an excellent model for studying the impact of DNA conformation flexibility on the recognition process between proteins and DNA (60).

Through computational investigation using the flexible DNA model with SRY, we aim to deepen our understanding of how binding kinetics are impacted by the ability of DNA to undergo conformational changes and the interplay between the speed with which the protein finds and binds its target DNA site. Importantly, we examine the complications arising when the protein fails to recognize its target site, adding an intriguing layer to our understanding of the complexity of gene transcription machinery.

Materials and methods

To study the mechanism by which SRY binds to DNA, we employed a CG model that was developed to address both searching of non-specific DNA sites as well as SRY recognition of its target DNA site.

SRY modeling

The structure of SRY that was implemented in the study was based on the complex formed between the HMG domain of the human SRY and DNA, as resolved by NMR (PDB: 1j46) (55). Each of the 85 residues that make up the SRY domain is represented by a single bead, positioned at the center of the Cα atom. The force-field used in the simulations implements native-topology modeling for SRY, where the experimentally obtained structure serves as a global minimum during the simulations, as was applied in the past for simulations of other proteins interacting with DNA (20,28,61). The potential energy function consists of the following terms:

$$
\begin{aligned}
V\ (\Gamma, \Gamma_0) = & \sum_{\text{bonds}} k_{ij}^{b}\left(r_{ij} - r_{ij}^{0}\right)^2 + \sum_{\text{angles}} k_{ijk}^{a}\left(\theta_{ijk} - \theta_{ijk}^{0}\right)^2 + \sum_{\text{dihedrals}} k_{ijkl}^{d} \\
& \left[\left(1 - \cos\left(\phi_{ijkl} - \phi_{ijkl}^{0}\right)\right) + \frac{1}{2}\left(1 - \cos\left(3\left(\phi_{ijkl} - \phi_{ijkl}^{0}\right)\right)\right)\right] \\
& + \sum_{\substack{contacts \\ i \neq j}} \epsilon_{ij}^{C}\left[5\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \\
& + \sum_{\substack{repulsions \\ i \\ lt;\ j-4}} \epsilon_{ij}^{NC}\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} \\
& + \sum_{ij} K_{\text{Coulomb}} B(\kappa)\left[q_i q_j \frac{e^{-\kappa r_{ij}}}{\epsilon r_{ij}}\right] \qquad (1)
\end{aligned}
$$

where the term r_{ij} represents the distance between sequential beads *i–j*, θ_{ijk} is the angle (in radians) between sequentially bonded beads *i–j–k*, and ϕ_{ijkl} is the dihedral angle (in radians) between four sequentially bonded beads *i–j–k–l*. Parameters whose values were measured in experimental structures are denoted with a superscripted 0 (e.g. r_{ij}^{0} is the distance between sequentially bonded beads *i* and *j* as measured in the experimentally determined structure). The term σ_{ij} is the distance measured between two non-sequential interacting beads *i–j* in the experimentally obtained structure. The native interactions were obtained using the CSU algorithm.

The last term in the energy function is the Debye–Hückel potential (20) with $K_{\text{Coulomb}} = 332$ kcal mol^{-1}e^{-2}. The term $q_{i/j}$ is the total charge of the residue, which can be –1 for negatively charged beads (Asp, Glu and the P beads of DNA), +1 for the positively charged beads (Arg and Lys), or 0 for neutral beads. The symbol ϵ is the dielectric constant, κ is the screening factor, and B (κ) is the salt-dependent coefficient. The Debye–Hückel potential is dependent on the distance between the charged beads, which is denoted as r_{ij}.

DNA modeling

In the current study, a DNA model was used that consisted of three beads representing the phosphate, sugar, and base groups of each nucleotide. These beads were labeled as P, S and B, respectively. The beads are bonded with harmonic and periodic potential, in a similar manner to MADna model (62) (reference). Although the original MADna model accounts for DNA sequence specificity, the model used in this study employs uniform parameters across the DNA molecule. The full description of the parameters can be found in the Supporting Information (Supplementary Tables S1–S3). The DNA utilized in the simulations was composed of 50 base pairs (bp), including 14 bp from the original NMR structure of the SRY-DNA complex, with the remaining base pairs comprised of B-DNA. Of the 28 nucleotides comprising the binding site, 13 nucleotides interact with binding residues of SRY.

The conformation of the DNA in the specific complex with SRY is bent by about 50° and thus deviates from linear B-DNA conformation. The DNA bending may be an intrinsic property of the sequence of the cognate site or induced by SRY binding. The distinct nature of the DNA conformation demands its incorporation in the computational model. Our model, therefore, was constructed to include both types of DNA conformation. The first conformation represented the B-form of DNA that corresponds to the DNA used for non-specific binding. The second DNA conformation was derived from the SRY-DNA complex and was designed by extending the 14 bp DNA extracted from the PDB 1j46 entry to a length of 50 bp using the x3dna webserver (63). To characterize the two distinct DNA conformations, we compared pairwise distances between phosphate beads in the two structures, and selected distances that exhibited a >20% change between the two structures as defining the conformational change required to switch between them. Sixteen such pairwise distances were identified. The model employed in this study permits transition between these two DNA conformations on the basis of the identified 16 pairwise distances. In the model, each of the identified pairwise contacts was assigned a dual-basin Gaussian potential (64) in which each Gaussian potential corre-

sponds to one of the distance values of that pair of phosphates (see Supplementary Figure S1). Several other approaches of dual-basin models were used in the past to study conformational changes in proteins (65–70). We note that various computational models addressed different questions on the effect of the polymeric properties of DNA (e.g. dynamics, persistence length, and structures) on various aspects of protein–DNA binding (6,71,72), however, the model used in this study concentrates on the conformational transitions of the DNA target site.

The dual-basin Gaussian consists of the following:

$$U_G = \epsilon \left[\left(1 + \left(\frac{\sigma_{NC}}{r_{ij}} \right)^{12} \right) \left(1 + G\left(r_{ij}, r_1^{ij} \right) \right) \left(1 + G\left(r_{ij}, r_2^{ij} \right) \right) - 1 \right] \quad (2)$$

where, $G(r_{ij}, r^{ij}_{n=1,2}) = -\epsilon_n \exp[-(r_{ij} - r_n^{ij})^2/2\sigma_n^2]$ and r_{ij} denotes the distance between the identified pair of phosphate beads, while r_n^{ij} represents the distance between pairs of beads in a specific state. ϵ is a global scaling factor and was set to be 1 in our simulations. Here, $n = 1$ refers to the B-DNA structure (also referred to as straight DNA state), and $n = 2$ denotes the NMR structure state (also referred to as the bent DNA state). The parameter ϵ_n indicates the depth of the energy basin, which modulates the interaction strength. The excluded volume value, represented by σ_{NC}, was set at 0.4 nm. The width of each well was adjusted to $\sigma_n = r_1^{ij} - r_2^{ij}$. We note that the model also allows additional conformations that constitute various combination of the parameters of the two distinct conformations.

The ability of the DNA to transition between the two states (namely, the straight B-DNA conformation and bent DNA conformation) is described by the parameter α, which refers to the deformability of the DNA:

$$\alpha = \frac{\epsilon_{bent}}{\epsilon_{bent}^{max}} - \frac{\epsilon_{straight}}{\epsilon_{straight}^{max}} \quad (3)$$

where ϵ_{bent} and $\epsilon_{straight}$ are the energetic strengths of the pairwise phosphate–phosphate interactions in the bent and straight DNA conformations, respectively, and the ϵ_{bent}^{max} and $\epsilon_{straight}^{max}$ terms are the maximal values used for the ε_{bent} and $\varepsilon_{straight}$ parameters, respectively. In our study, $\epsilon_{bent} = \epsilon_{bent}^{max} = 6$ and $\epsilon_{straight}^{max} = 8$. The values of $\epsilon_{straight}$ range between 0 and 8. Given these parameters, the relative strength of the straight and bent DNA conformations are tuned. At the lowest deformability value (i.e. $\alpha = 0$; which is achieved for $\epsilon_{straight} = 8$), the energy preference for the DNA's bent state is zero. Conversely, at the highest deformability value (i.e. $\alpha = 1$; which is achieved for $\epsilon_{straight} = 0$), the energy preference for the bent state reaches its maximum. We simulated the system under two scenarios: one in the absence of the SRY protein (i.e. with only the 50 bp DNA), and another in its presence. Each scenario was studied by 50 simulations with the values of α being scanned between 0 and 1.

The model incorporates two contributions to interactions between SRY and DNA: non-specific electrostatic interactions with any DNA nucleotide and specific interactions with the cognate site. The former are modeled using the Debye–Hückel equation and the latter by the Lennard-Jones function (Eq. 1). The specific interactions are defined using the NMR structure (i.e. binding to the bent DNA), in which 33 such native interactions between the SRY and the DNA were identified. The potential constants were $k_{ij}^b = 100$ kcal mol^{-1} Å^{-2}, $k_{ijk}^a = 20$ kcal mol^{-1}, $k_{ijkl}^d = 1$ kcal mol^{-1}, and $\epsilon_{ij}^C = 1$ kcal mol^{-1}. To achieve efficient sampling of binding and dissociation events, the Lennard-Jones interactions between the SRY and the DNA were modeled using a decreased strength for the specific protein–DNA interactions, with $\epsilon_{ij}^C = 0.7$ kcal mol^{-1}. The system was simulated using Langevin dynamics at a temperature of 0.4 (reduced units). The dielectric constant was set to 80 with and the salt concentration set to either 0.02 or 0.04 M. The binding of SRY to its specific site is probed by the fraction of contacts formed, $Q_{\text{Protein–DNA}}$, where complete binding is linked with $Q_{\text{Protein–DNA}} = 1$ (namely, all the 33 native interactions are formed).

Simulations and their kinetic and thermodynamic analysis

To decipher the mechanism of SRY binding to its specific target DNA, we simulated its binding to various DNA sequences that differ in their deformability (as captured by the α parameter). To obtain a complete mechanistic characterization for each DNA model, a set of 50 long simulations of 5 ms that included several association and dissociation events were collected. The SRY–DNA interactions in these trajectories were classified to one of three states: binding to nonspecific DNA sites (the search or *S* state), binding to the specific site (the recognition or *R* state), and partial binding to the specific site (the intermediate or *I* state). The three states were designated based on the fraction of contacts between SRY and the target site, $Q_{\text{Protein–DNA}}$. The *S* state is defined as $Q_{\text{Protein–DNA}} = 0$, the *R* state as $Q_{\text{Protein–DNA}} > 0.5$, and the *I* state is defined as $0 < Q_{\text{Protein–DNA}} < 0.5$.

We estimated the mean first passage times from the *S* → *I* state, *I* → *R* state and *S* → *R* state. The kinetic on-rate constants ($k_{S\to I}$, $k_{I\to R}$ and $k_{S\to R}$, respectively) were calculated as the inverse of these transition times. In a similar manner, the kinetic off-rate constants ($k_{I\to S}$, $k_{R\to I}$ and $k_{R\to S}$) were calculated by the inverse of the mean time for the corresponding transitions. The calculated kinetic rates were used to estimate the affinity of SRY for DNA (as described by their corresponding dissociation constant K_d) by measuring K_{SI}, K_{IR} and K_{SR}. These dissociation constants were estimated by $K_{IR} = k_{R\to I}/k_{I\to R}$ and $K_{SI} = k_{I\to S}/k_{S\to I}$ and $K_{SR} = k_{R\to S}/k_{S\to R}$.

To calculate the kinetic rates and skipping events, the trajectories and center of mass (COM) displacement data were initially smoothed using a convolution function to eliminate noise. The degree of bending of the DNA was quantified by the fraction of phosphate–phosphate pairs, represented by the $Q_{\text{DNA-bending}}$ term, that switch from *S* to *R* distances. The case of $Q_{\text{DNA-bending}} = 1$ occurs when all the 16 predefined phosphate–phosphate pairwise distances correspond to the values of the *R* state. The local DNA bending was assessed based on the helical axis of the DNA molecule (73). The axis is defined as the line that connects the centers of each pair of phosphate beads in a CG-model generated from the NMR structure of SRY with DNA of the same length generated by the x3dna webserver.

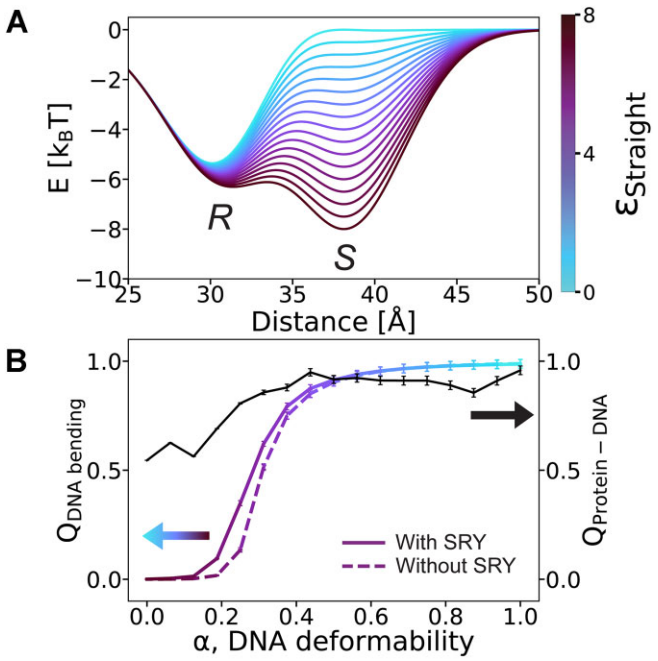


Figure 2. SRY binding to a deformable DNA studied with a dual-basin coarse-grained model. (**A**) An illustration of the dual-basin Gaussian well used, in which well depth is adjusted between the two DNA states. One state corresponds to the deformed, bent DNA conformation (taken from the specific SYR–DNA complex, PDB 1j46, i.e. the *R* state,) whereas the other represents the straight B-DNA conformation (adopted by DNA while SRY engages in nonspecific interactions during the search process; i.e. the *S* state). The *R* and *S* states are captured by 16 dual-basin Gaussians applied for pairwise distances between phosphates (see Supplementary Figure S1). The degree of DNA deformability, α, is adjusted by changing the relative strengths of the two energetic basins (see Eq. 3). In our study, the energetic strength of pairwise phosphate–phosphate interactions in the *R* basin (ε_{bent}) is kept fixed and the corresponding strength of the *S* basin ($\varepsilon_{straight}$) is scanned (see colorbar). (**B**) The mean bending of DNA ($Q_{DNA\text{-}bending}$) possessing a varying degree of sequence deformability (α) is estimated in the presence (circles) or absence (triangles) of the SRY protein. The $Q_{DNA\text{-}bending}$ parameter is the fraction of phosphate–phosphate pairs among the 16 selected pairs whose distances fit the values that define the bent conformation of the *R* state. The bendability of the DNA, as probed by the coarse-grained model of a dual-basin Gaussian potential is measured for different values of α. The solid black line indicates the fraction of specific protein–DNA interactions ($Q_{Protein\text{–}DNA}$). A value of unity corresponds to the formation of all 33 interactions that define the SRY–DNA complex in the NMR structure. Each data point is calculated as the mean from 50 simulations.

Results

SRY binding promotes DNA bending

In this study, we employed a CG model that allows conformational changes of the DNA between two major distinct conformations. In this dual-basin potential energy model (Figure 2A), the basins are modeled as Gaussian functions, with one basin corresponding to a canonical B-DNA form (i.e. an energetically-preferred straight conformation; as assumed for the *S* state; Figure 1B) and the other corresponding to the conformation DNA adopts when in complex with SRY (*i.e.* a bent conformation; as found in the *R* state; Figure 1B). Transitions between states are achieved by incrementally modulating the relative depths of the potential energy wells by systematically scanning the energy depth parameter ($\varepsilon_{straight}$) for each pair of phosphates in the straight DNA basin (Figure 2A, colorbar). The dual-basin nature of the Gaussian potentials facilitates gradual transitions between the straight and bent DNA conformations via variably bent intermediate states. These intermediate states are modeled as combinations of all the geometrical parameters of the bent and straight conformations, with the probability of each parameter varying from 0 to 1 and corresponding to its lifetime.

We use α to designate the relative energetic preference to adopt the geometric parameters of the bent over the straight DNA conformation, which describes the overall deformability of the DNA. When $\alpha = 1$, the DNA is highly deformable and therefore it exclusively adopts the bent conformation. When $\alpha = 0$, the DNA is highly non-deformable, and it adopts the straight B-DNA conformation. All intermediate deformability values correspond to various ratios of populations in both states. Intuitively, as the deformability of the DNA increases, its bending propensity increases. Figure 2B shows that the fraction of phosphate–phosphate pairwise interactions defining bent DNA, $Q_{DNA\text{-}bending}$, reaches a value of 1 as the deformability, α, approaches 1.

It is evident that DNA bending is affected by the presence of the SRY protein (Figure 2B, circles compared with triangles). Binding of SRY acts as a counterbalance to the bent conformational state of the DNA molecule. The specific interactions of SRY binding in the minor DNA groove modify the system's energy landscape, leading to a more bent conformation (Figure 2B). This introduces additional energetic contributions, enabling stabilization of the bent DNA conformation and increasing the bending propensity beyond the intrinsic bending dictated by the DNA sequence. The modeling results for partially deformable DNA support this understanding, because at intermediate α values greater DNA bending is observed in the presence of SRY compared with its absence. For such partially deformable DNA sequences ($0.1 < \alpha < 0.3$), protein binding can cause an up to 20% increase in the DNA bending tendency (Figure 2B).

DNA bendability is essential in order for SRY to recognize its specific DNA binding site. With nondeformable DNA (i.e. in the straight DNA conformation; $\alpha \leq 0.1$), SRY forms only ~60% of the specific interactions formed when it binds to its cognate site (Figure 2B; specific interactions are quantified by $Q_{Protein\text{–}DNA}$). As the DNA becomes more deformable (i.e. as α increases), SRY binds more extensively, as shown by the sharp increase in $Q_{Protein\text{–}DNA}$. This result supports that protein–DNA interactions might be highly dependent on the conformational flexibility of the DNA molecule and its ability to form a bent structure. Maximal binding of the SRY to its cognate site is achieved at $Q_{DNA\text{-}bending} \sim 0.9$, indicating that tight binding can be achieved even for DNA that slightly deviates from the DNA conformation found in the NMR structure of its complex with SRY.

DNA conformational deformability modulates SRY recognition kinetics

The dual-basin model was applied to examine how the kinetics of DNA recognition by SRY is affected by the intrinsic conformational flexibility of the DNA, which dictates its tendency to adopt the bent conformation. In these simulations, the SRY was located at a random position, either on or off the DNA, and the formation of specific interactions between SRY and its predefined specific DNA site (probed by $Q_{Protein\text{–}DNA}$) was followed over time (Figure 3, upper panel in each set). To

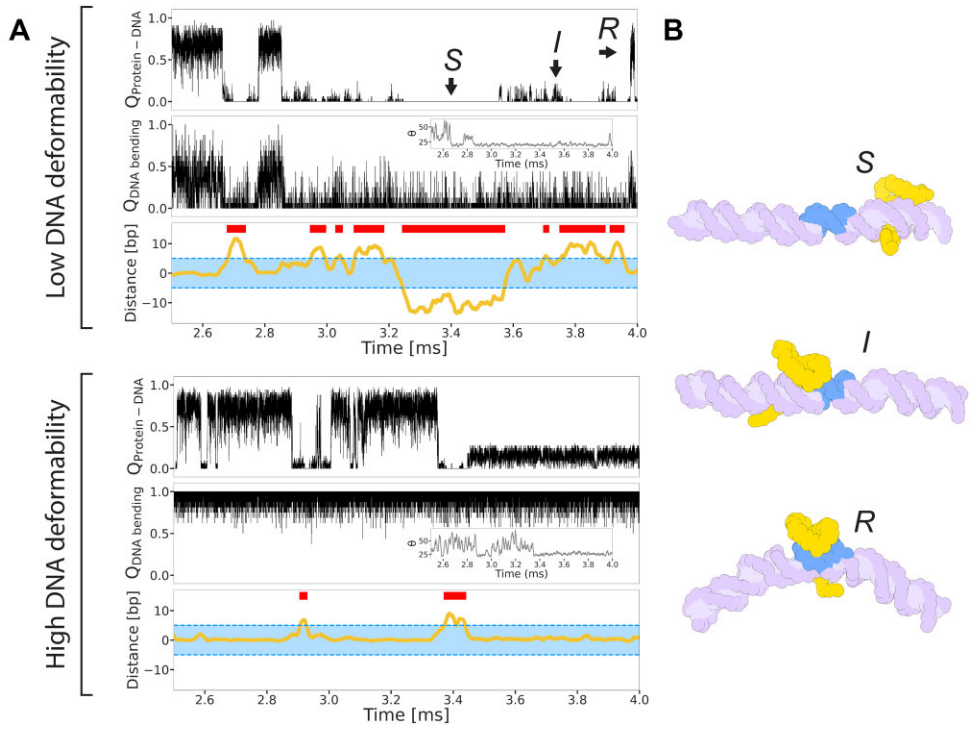


Figure 3. Effect of DNA deformability on SRY while searching for and recognizing its specific DNA binding site. (**A**) Two representative trajectories are presented for SRY interacting with two types of DNA molecules: DNA with low deformability ($\alpha = 0.2$) and DNA with higher deformability ($\alpha = 0.6$). Each of the trajectories were analyzed in terms of SRY binding and DNA bending. Each set of three trajectories shows: (upper) time evolution of the fraction of contacts between SRY and the target DNA site ($Q_{Protein-DNA}$) for DNA in the *R* state. (Middle) Time evolution of the fraction of DNA phosphate–phosphate pairs that adopt distances that fit the definition of the bent *R* state ($Q_{DNA-bending}$). The local bending angle θ is shown as an inset (note that the bending angle monitors SRY binding in the R state but is not coupled with the DNA bending). (Bottom) Time evolution of the displacement of the center of mass (COM) of SRY relative to the location of the target site on the DNA. The vicinity of the target site is marked by the blue band at distance ∼0 base pairs (bp), which indicates the region in which the COM is closest to the center of the binding site. Red bars indicate skipping events in which the SRY partially binds the target site. On the basis of the $Q_{Protein-DNA}$ plots, we classify SRY–DNA binding in three groups: SRY search of nonspecific DNA ($Q_{Protein-DNA} = 0$; the *S* state); SRY present at the binding site region but only forms partial interactions ($0.2 < Q_{Protein-DNA} < 0.5$; the *I* state); SRY recognizes the DNA target site and most of its interactions with DNA are specific ($Q_{Protein-DNA} > 0.5$; the *R* state). (**B**) Illustrative representative snapshots from the simulation of the *S*, *I* and *R* states of SRY-DNA binding states. The SRY is shown in yellow, the target DNA site in blue and nonspecific DNA sites in purple.

complement the analysis of the kinetics of SRY recognition of the specific binding site, we also probed the kinetics of the arrival of the SRY at that site by measuring the distance between the COM of SRY and the center of the binding site over time. Similarly, the evolution over time of DNA bending (probed by $Q_{DNA-bending}$; middle panel in each set) and the bending angle were also followed (Figure 3; middle panel inset).

The CG simulations indicate that SRY–DNA binding progresses through three distinct states (Figure 3): the search state (*S*), intermediate state (*I*), and the recognition state (*R*) (see Supplementary Video SV1). In the *S* state, the SRY scans nonspecific DNA sites (i.e. in the *S* state; $Q_{Protein-DNA} = 0$), while it performs hopping or sliding dynamics along the DNA, depending on the strength of the electrostatic interactions, which are dictated by the salt concentrations (see Supplementary Figure S2). In the *I* state, SRY partially binds the DNA, as indicated by the formation of, on average, ∼20% of its specific interactions with DNA. By contrast, in the *R* state, at least 80% of the specific SRY–DNA interactions are formed. The *S*, *I* and *R* states are found for SRY binding to DNA irrespective of the extent to which it is deformable (Figure 3), however, the kinetic parameters are evidently affected by the degree of deformability (i.e. by the value of α).

The kinetic rate constants associated with the transitions between the *S*, *I* and *R* states were calculated from the association/dissociation simulations of SRY and DNA. For SRY–DNA association, the kinetic rate constant for the first transition ($k_{S\to I}$) is faster than that for the second transition ($k_{I\to R}$) (Figure 4A), which can be explained by the higher energetic barrier to convert from the *I* to *R* state than from the *S* to *I* state. The transition from the partially bound *I* state to complete recognition in the *R* state occurs slowly and, in many cases, occurs only after repeated attempts because the intermediate state also dissociates due to unsuccessful binding (Figure 3; upper panel). Indeed, unsuccessful transition from the *I* state to the *R* state may result in transition back to the *S* state, in which the protein searches the DNA at sites that may be distant from the specific binding site, with particularly substantial kinetic consequences. The effect of the SRY skipping the cognate site due to unsuccessful binding will be discussed in a later section (see Results section 4). At this point, we focus on the kinetics of productive binding events.

The rates constants $k_{I\to R}$ and $k_{S\to R}$ exhibit a notable sensitivity to variations in DNA deformability (Figure 4A). Specifically, it is evident that these rates decrease with decreased DNA deformability (i.e. as the DNA becomes increasingly straight and the value of α decreases). The origin for this

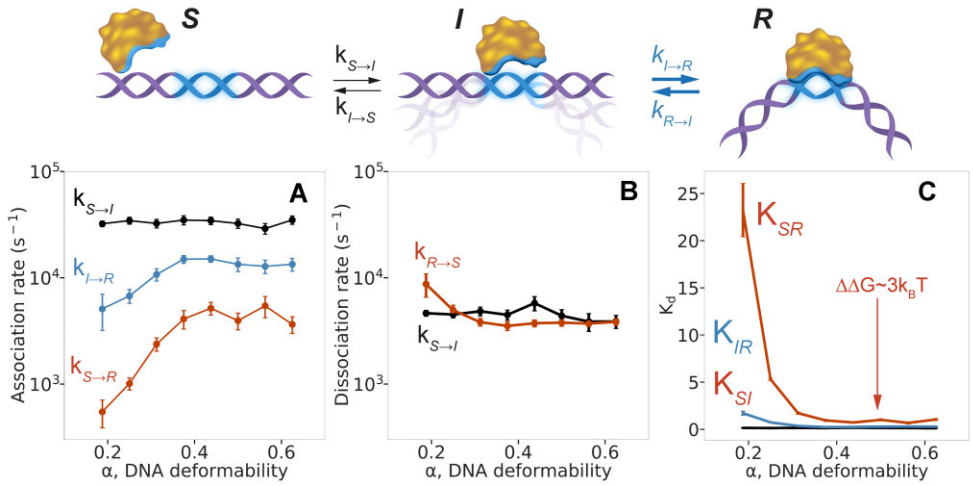


Figure 4. A kinetic scheme for SRY binding to DNA sequences of varying deformability. (Upper panel) The molecular mechanism of SRY–DNA recognition involves a conformational change to DNA that is characterized by a transition from a straight B-DNA (*S* state) to a bent DNA (*R* state). In the *S* state, the SRY interacts with the DNA nonspecifically while performing 1D diffusion, whereas in the *R* state their interactions are tighter and principally involve specific interactions. In the intermediate state, *I*, a fraction of the specific interactions are formed and the DNA may adopt fluctuating conformations. (Lower panels) The rate constants (*k*) for (**A**) the association kinetics (k_{on}) governing the transitions $S\to I$ ($k_{S\to I}$; black), $I\to R$ ($k_{I\to R}$; blue), and the overall association transition $S\to R$ ($k_{S\to R}$; red); and (**B**) the dissociation kinetics (k_{off}) governing the transition $I\to S$ ($k_{I\to S}$; black), $R\to I$ ($k_{I\to R}$; blue), and the overall dissociation transition $R\to S$ ($k_{R\to S}$; red). The kinetic parameters were obtained as the average of 50 simulations across eight DNA deformability values (α). The value of the $k_{S\to R}$ rate constant incorporates kinetic delays due to skipping events. The $k_{R\to I}$ and $k_{R\to S}$ rate constants are identical, by definition. (**C**) The equilibrium dissociation constant, K_d, calculated as k_{off}/k_{on} for the kinetic rates of the corresponding transition. The difference between the values of K_{SR} calculated for low and high α values ($\Delta\Delta G$) corresponds to the change in affinity due to DNA deformability ($\Delta\Delta G \approx 3k_BT$).

slower kinetics is the higher energetic barrier for conformational change that is involved when the straight DNA conformation is more populated than the bent conformation. A higher energetic barrier results not only in slower $I\to R$ transitions (and therefore smaller $k_{I\to R}$) but also fewer transitions (and therefore smaller $k_{S\to R}$). The first association rate $k_{S\to I}$ is least affected by α. This can be rationalized as the *I* state being stabilized by the formation of a ∼20% specific interactions between SRY and DNA and by the DNA not being required to adopt the energetically less-favorable bent conformation.

The deformability of the DNA sequence also affects the dissociation rates (Figure 4B), but to a lesser extent compared with its effect on the association rates (Figure 4A). Nonetheless, the relations between α and the off-rates are opposite to its effect on the on-rates. The binding of SRY to a less deformable DNA sequence results in higher dissociation rates, as is clear from the higher rate constants for the $R \to S$ transition (Figure 4B; and for $R \to I$ transition) at lower α values. The $I\to S$ transition is hardly affected by DNA deformability, consistently with the effect of α on the kinetics for the transition $S\to I$.

Effect of DNA deformability on binding affinity

To further investigate the transition between partial and full SRY–DNA binding and its relationship with DNA deformability, we calculated the dissociation equilibrium constants ($K_d \equiv k_{off}/k_{on}$) for several steps in the recognition process (Figure 4C). Accordingly, the K_d was calculated for the transitions $S\to I$, $I\to R$ and $S\to R$ (see Materials and Methods). The values of K_{SI} and K_{IR} for SRY binding to a bent DNA conformation (i.e. at a high α value) are close to zero and are also lower than the value of the overall dissociation equilibrium constant, K_{SR}. We focus here on the dependence of K_d on α, rather than on its absolute value.

Figure 4C shows that, while the value of K_{SI} is barely affected by DNA deformability, K_{IR} and particularly K_{SR} show a clear dependency on α. As the DNA deviates further from its preferred bent conformation for binding (i.e. at lower α values), the value of K_{SR} increases, which indicates that the propensity of SRY to dissociate from DNA increases as the DNA conformation increasingly deviates from the optimal bent conformation for binding.

The values of K_{IR} and K_{SR} increase by factors of ∼8 and ∼23, respectively, upon decreasing the value of α. Such changes in the value of K_d for the SRY–DNA complex at different α values can be employed to estimate the change in the affinity ($\Delta\Delta G$) of SRY for DNA sequences that differ in their deformability, given that $\Delta\Delta G = k_BTln\ (K_d^{\alpha\cong 0}/K_d^{\alpha\cong 1})$, where k_B is the Boltzmann constant. This approach may provide a quantitative measure of the effect of the propensity of DNA to deform on protein binding affinity to DNA. The assessment of the change in free energy resulted in $\Delta\Delta G = 1.8$ k_BT for the $I\to R$ transition, indicating that of these two states, the *R* state is the less stable, which can be attributed to the energetically costly requirement that the DNA bend to achieve *R* state, and therefore the transition $I\to R$ is less favorable as the DNA is less deformable. A value of $\Delta\Delta G = 3.1$ k_BT characterizes the overall transition $S\to R$ and includes the effect of unproductive transitions, which occur more commonly at low α values, thus making the *R* state ∼ 3 k_BT less favorable than the *S* state. These affinity differences imply that altering the propensity of the DNA molecule to deform can increase the binding affinity to DNA, where the specific interactions with the binding site remain the same. Our results thus suggest that DBP affinity for its cognate site can be increased simply by increasing the probability of the DNA to undergo conformational change without changing the strength of the interactions.

Achieving increased affinity by increasing DNA deformability is reminiscent of the effect of DNA mismatches on affin-

ity (52). It was shown that the introduction of mismatches within the specific binding site of DNA may lead to a reduction in the energetic barrier associated with the binding process because of the increased propensity of DNA to adopt a bent conformation upon the introduction of mismatches. The inherent ability of DNA to bend in response to the mismatches lowers the energetic cost required for the protein to achieve its fully bound state, for example, the change in binding affinity following the introduction of mismatches into tumor protein p53 and the TATA-Box Binding Protein was 0.4–1.8 k_BT and 0.8–1.4 k_BT, respectively. These values closely align with those calculated from our computational results, affirming the validity of our approach to modeling the system, and suggesting that modulation of DNA deformability could potentially reduce the energetic cost for specific binding.

Unproductive binding events govern overall recognition kinetics

We now consider the role of the kinetics of unproductive binding events. The precursor to both productive and unproductive binding (skipping) is the protein locating and partially binding its target DNA site, with partial bindings occurring relatively quickly once the target site is found. Transition from the intermediate state to the fully bound state at the target site (i.e. $I \rightarrow R$) is up to two-fold slower than the rate of intermediate state formation, depending on the deformability of the DNA. However, as mentioned earlier (see Results section 2), attempts to form the R state often conclude unproductively, with the protein transitioning back to the S state ($I \rightarrow S$). The SRY then continues searching non-specific DNA sites till the cognate site is found again and transition to the I state is followed by another attempt to form the R state, and so forth until the R state is achieved. Incidents of unproductive recognition of the cognate DNA site result in slower overall recognition kinetics and make $I \rightarrow R$ transitions rare events. Figure 4A shows that the overall recognition rate constants for transition from the S to the R states, revealing that $k_{S\rightarrow R}$, is about 3–10-fold slower than $k_{I\rightarrow R}$.

The value of $k_{S\rightarrow R}$ is strongly dependent on DNA deformability (Figure 4A). As the DNA becomes less deformable and less- easily populates the bent conformation, the value of $k_{S\rightarrow R}$ reduces to as little as a sixth of its original value. This reduced rate constant is driven not only by the requirement to surpass a higher free energy barrier during the transition but also by the possibility of the barrier not being overcome in a timely manner. Thus, the tendency of DNA to adopt a straight conformation (characterized by low α deformability), reduces the probability of productive transition from the I to R states. Unproductive binding events in which the SRY fails to engage in specific binding of its cognate DNA site after finding it and establishing initial binding (as indicated by formation of the I state) can be described as skipping events. In such a skipping event, the SRY slides over its intended target site and continues to explore the DNA (Figure 5).

To investigate the frequency of skipping events prior to achievement of complete binding in the R state (Figure 3, red marks), we counted the number of skipping events (N_{skip}) between two successive target-site recognition events for DNA strands with different levels of intrinsic deformability at high and low salt concentrations. At a low salt concentration of 20 mM NaCl, the SRY skips the target site on average ∼20 times when it searches highly deformable DNA compared with ∼500 times when it searches less deformable DNA. At the higher 40 mM salt concentration, similar results were obtained only for less deformable DNA, whereas for highly deformable DNA, the SRY skips the target site on average ∼100 times (data not shown).

The inverse of the number of skipping events is an estimate to the probability of recognition, P_f (Figure 6A). This probability is lowest when DNA has the highest tendency to adopt a straight conformation, indicating that the protein's ability to find and successfully bind its cognate site is hindered. However, as DNA deformability increases, the probability of recognition also increases until α values reach at least the mid-range. This relationship suggests that the conformational state of DNA plays a crucial role in facilitating or hindering protein–DNA interactions beyond simple shape readout and geometric complementarity.

The probability of recognition is also reduced at higher salt concentrations. For example, at 40 mM NaCl, $N_{skip} \approx 80$ when SRY attempts to recognize the target DNA site, which is 4-fold than the value obtained for 20 mM of NaCl (Figure 6A). This can be explained by the lower nonspecific affinity at the higher salt concentration, which reduces the residence time at a given site. Consequently, the probability of skipping the target site increases. Such an effect of salt may depend on the properties of the DNA. As the energetic barrier for recognition increases (i.e. at low α values), the effect of salt in enhancing kinetics decreases (Figure 6).

Implications of skipping the target site on the overall DNA search kinetics

Iterative attempts to recognize the target site due to unsuccessful binding events are expected to have profound consequences for the overall DNA search kinetics. In the previous section, we showed how the rate of recognition at the microscopic level is affected by the low recognition probability imposing several DNA search cycles. Here, we aim to estimate how these skipping events affect the overall macroscopic DNA search kinetics (39,74).

To address this, we used a theoretical kinetic framework for protein–DNA binding and estimated the overall recognition time for SRY to search, recognize, and specifically bind its target DNA site (τ_R). This recognition time is assumed to be the sum of the time the SRY spends in 1D and 3D diffusion as follows:

$$\tau_R = \frac{M}{\bar{n}^2 P_f}\left[\tau_{1D} + \tau_{3D}\right] \quad (4)$$

where M is the size of the genome, $\bar{n}$ is the average number of sites scanned in each 1D search round, and τ_{1D} and τ_{3D} are the 1D and 3D search times, respectively. The probability of finding and target site successfully binding it and can be estimated as (38):

$$P_f = 1/N_{skip} = \frac{k_{S\rightarrow R}}{k_{S\rightarrow R} + k_{res}} \quad (5)$$

$k_{S\rightarrow R}$ is the rate constant for transitioning from the search mode (S state) to the recognition mode (R mode), and k_{res} is the residence rate constant, which reflects the rate at which the SRY protein moves a single step while in the S mode. We used the inverse of the number of times SRY skips over its

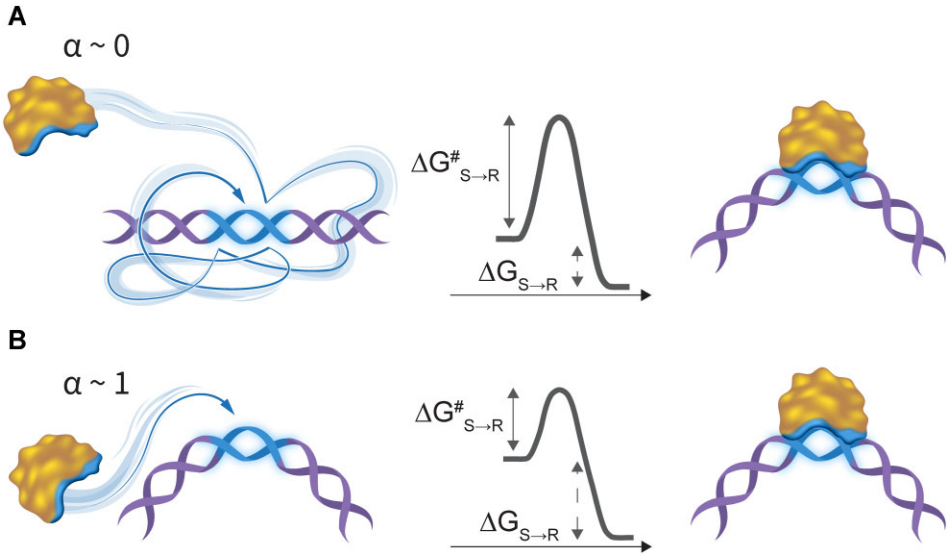


Figure 5. Schematic mechanism for protein–DNA recognition that involves DNA bending. (**A**) A DNA sequence with low deformability ($\alpha \sim 0$) has a low likelihood of populating the bent conformation and therefore a searching protein is likely to encounter DNA in a straight B-form conformation, which does not support initial protein binding to DNA. Although the protein finds the target site (marked in blue), it slides over the site several times without engaging in successful binding. Such skipping events can occur several times (blue arrow) before successful binding (right panel) is achieved. Consequently, low DNA deformability is associated with a high barrier to binding ($\Delta G^{\#}_{S\to R}$) and slow kinetics for protein–DNA recognition. (**B**) A DNA sequence with high deformability ($\alpha \sim 1$) has an intrinsic tendency to adopt a bent conformation, thus raising the likelihood of a searching protein encountering bent DNA at the target site, decreasing the likelihood of skipping events, and reducing the barrier for binding. Both scenarios A and B lead to specifically bound protein–DNA complex, yet with different kinetics rates. The change in DNA deformability also contributes to a change in affinity (see Figure 4). Thus, DNA deformability affects both the energetic barrier to binding [$\Delta G^{\#}_{S\to R}(\alpha\sim0) > \Delta G^{\#}_{S\to R}(\alpha\sim1)$] and the binding affinity [$\Delta G_{S\to R}(\alpha\sim0) < \Delta G_{S\to R}(\alpha\sim1)$].

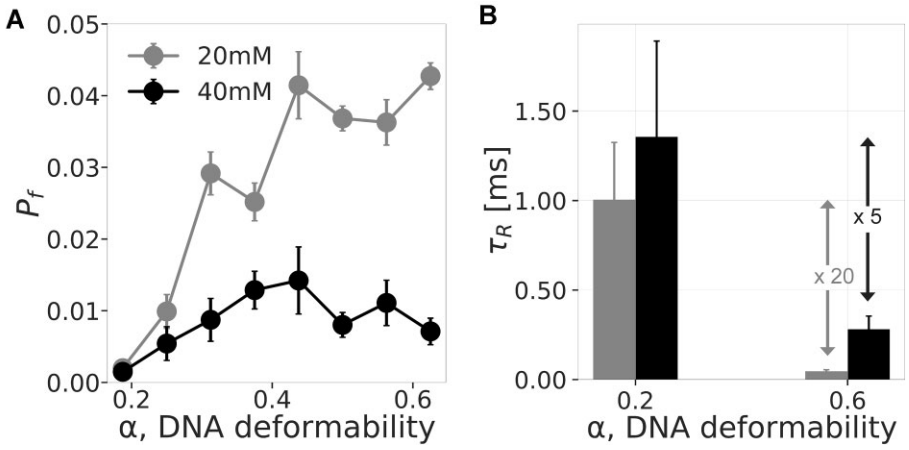


Figure 6. The effect of skipping the DNA target site on the overall recognition time. (**A**) The probability of SRY binding the target site (P_f) calculated as the inverse of the skipping frequency ($P_f = 1/N_{skip}$) obtained from the coarse-grained simulations for DNA sequences across a range of deformability (α) values and at two salt concentrations. (**B**) Overall recognition time (τ_R) for SRY binding of the target site of weakly ($\alpha = 0.2$) and highly ($\alpha = 0.6$) deformable DNA at two salt concentrations; 20 mM (grey) and 40 mM (black). The τ_R parameter is calculated using Eq. (4) and using the P_f as estimated from the simulations (Figure 6A) and $\tau_{3D} = 10^{-4}$ *s*; $\tau_0 = 10^{-8}$ *s*; $E_{ns} = 1 k_BT$; $\sigma = 0\ k_BT$.

target site (i.e. N_{skip}) to estimate the approximate value of P_f, rather than calculating it. To explore how the P_f affects the recognition time, τ_R, we now calculate it for two extreme cases: bent DNA ($\alpha = 0.6$) and straight DNA ($\alpha = 0.2$) using the P_f values that are shown in Figure 6A (*i.e.* 2×10^{-7} and 4×10^{-2} for deformability values of 0.2 and 0.6, respectively, at a salt concentration of 20 mM). Assuming the SRY protein performs a single round of 1D search ($M = <n>$; where n is the number of sites scanned in a single 1D round), we calculated τ_R by estimating τ_{1D} and τ_{3D} similar to previous studies (39,75). The typical time the SRY spends in 1D diffusion can be estimated by

$$\tau_{1D} = \tau_0 \exp\left(\frac{E_{ns}}{K_BT} + \frac{1}{2}\left(\frac{\sigma}{K_BT}\right)^2\right) \quad (6)$$

where E_{ns} is the interaction energy of SRY to nonspecific DNA sequences, σ is the ruggedness of the DNA, and τ_0 is the typical time it takes a protein to hop to a neighboring DNA site. Using the term for the 1D search, we assumed that the search landscape is smooth (namely, σ=0).

The overall recognition time depends significantly on the probability of the protein reaching the binding site and binding the specific target sequence, which (as mentioned earlier) are strongly affected by salt concentration and DNA deformability (Figure 6B). At a salt concentration of 40 mM, the recognition time is 5-fold longer for the less-deformable DNA. At a lower salt concentration of 20 mM, DNA deformability has a much greater effect on DNA recognition time, with τ_R being 20-fold higher for less-deformable compared with more-deformable DNA. We note that Eq. (4) assumes that a high number of skipping events (resulting in low P_f) necessarily lengthens τ_R, however, it is likely that in some cases skipping may result in a relatively fast return to the target site, and therefore the estimated τ_R should be treated as an upper limit.

Potential means to minimize the effect of skipping the target site on recognition kinetics

Following the direct linkage between the number of skipping events and the overall recognition time, we next address whether N_{skip} can be modulated by biophysical means in ways that might be employed evolutionarily. On the basis of Eq. (5), the recognition probability, P_f, is determined by two parameters: $k_{S\to R}$ and k_{res}. The residence rate constant k_{res} can be expressed in terms of the diffusion coefficient for SRY diffusion on DNA (as it is the residence time to stay at a single base pair (BP)), where the mean interactions energy between the diffusing SRY and the DNA is characterized by non-specific electrostatic energy, E_{ns}. The relationship between k_{res} and E_{ns} can be estimated by the following equation (38):

$$k_{\mathrm{res}} = \frac{2\mathrm{D}_{1\mathrm{D}}}{\mathrm{BP}^2} = \frac{2}{\tau_0}\exp\left(-\left(\frac{E_{ns}}{\mathrm{k_B T}}\right)\right) \quad (7)$$

Eq. (7) indicates that k_{res} increases as electrostatic interactions (represented by lower E_{ns} values) weaken, because weaker electrostatic interactions enable faster protein movement along the DNA during the 1D search and a shorter residence time at each non-specific DNA site. A simple approach to reducing E_{ns} is by raising the salt concentration of the surrounding medium, and thereby inducing the protein to utilize the hopping mode to a greater extent, at the expense of the sliding mode, and thereby engage in faster linear diffusion along the DNA (20,28). This expression hinges on an assumption that ruggedness is introduced into the potential energy landscape solely by electrostatic interactions.

The overall recognition rate constant $k_{S\to R}$, does not depend on non-specific interactions, but only on the energetic barrier to the $S\to R$ transition, $\Delta G^{\#}_{S\to R}$. It can be expressed as:

$$k_{S\to R} = \frac{1}{\tau_0}\exp\left(\frac{-\Delta G^{\#}_{S\to R}}{\mathrm{k_B T}}\right) \quad (8)$$

The relationships of skipping frequency (Figure 7A) and overall recognition time (Figure 7B) with the non-specific binding energy, E_{ns}, are affected by the height of the energetic barrier, $\Delta G^{\#}_{S\to R}$ (represented by the colorbar). When it is energetically favorable for the SRY to engage in non-specific interactions with the DNA strand (i.e. at low E_{ns} values), the residence time at any non-specific site (i.e. $1/k_{res}$) is shorter and therefore skipping events occur with a higher frequency (Figure 7A). The inherent energetic barrier for the transition between the S and R binding modes also strongly influences the number of skipping events. An increase in the $\Delta G^{\#}_{S\to R}$ for specific binding correlates with an increase in N_{skip}, implying that the protein finds the specific site but is unable to overcome the energetic barrier to transitioning from the S state to the R state and bind it.

Using Eq. 4, the effect of E_{ns} or $\Delta G^{\#}_{S\to R}$ on P_f is used to estimate τ_R. Figure 7B shows that both E_{ns} and $\Delta G^{\#}_{S\to R}$ have profound effects on τ_R while the corresponding data from Figure 7A indicate that these effects are caused by an increase in the number of skipping events. For example, for a protein–DNA system with E_{ns}= 5 k_BT, changing $\Delta G^{\#}_{S\to R}$ from 3 k_BT (blue line) to 7 k_BT (green line) increases τ_R from 0.007 s to 0.15 s (by increasing the number of skipping events ∼20-fold). A further increase in the energetic barrier from 7 k_BT to 11 k_BT (red line) results in a τ_R of 8.2 s (by increasing the number of skipping events ∼50-fold).

The dependence of τ_R on these energetic parameters shows that for any given value of $\Delta G^{\#}_{S\to R}$, there is an optimal value of E_{ns}, demonstrating a tradeoff. On the one hand, a high E_{ns} value decreases k_{res} (Eq. 7) (and therefore increases P_f; Eq. 5), but on the other hand it may slow down linear diffusion along the DNA (and therefore increase τ_{1D}; Eq. 6). The existence of an optimal E_{ns} value defines the conditions under which the DBP performs an adequately quick search yet also lingers sufficiently on the DNA to enable efficient recognition, thus reducing recognition time to a minimum. As $\Delta G^{\#}_{S\to R}$ increases, the optimal E_{ns} region is shifted toward higher values, because τ_R lengthens and greater P_f values are required to overcome the barrier.

The relationship between the recognition rate constants and the non-specific binding energy (Figure 7B) raises a question about how DBPs that must undergo a conformational change at their DNA binding site can preserve reasonably fast recognition times, τ_R, despite encountering a high energy barrier to conformational change, $\Delta G^{\#}_{S\to R}$. These proteins must navigate the challenge of efficiently reaching the target site more effectively than proteins that bind to a canonical B-form DNA. One strategy to achieve this is by modulating non-specific DNA interactions to decrease the residence time. This may be accomplished by enhanced nonspecific affinity, E_{ns}, which can be achieved in principle by making the protein surface more positively charged or by reducing frustration between the specific and nonspecific DNA binding modes (13). In such scenarios, the ruggedness of the search landscape increases. Support for this claim can be found in architectural proteins that bind to a bent DNA and that were shown to diffuse more slowly along DNA due to their intrinsic energetic ruggedness (76). Such slow diffusion may serve to reduce the number of skipping events, N_{skip}, because these are coupled to a high energetic barrier to conformational change.

Discussion

The kinetics of protein–DNA binding is affected by various factors, with one of the elementary factors being the length of the DNA molecules, which dictates the extent of nonspecific DNA sites that the protein potentially needs to scan before identifying the target site. Long search times may play a dominant role in binding kinetics. In light of this, numerous studies have focused on quantifying the mechanism and kinetics of DNA search (7,25,39,77,78). However, protein binding to the DNA target site, once identified, can pose another kinetic challenge by requiring energetically costly conformational changes from the protein and/or the DNA. The contribution of the

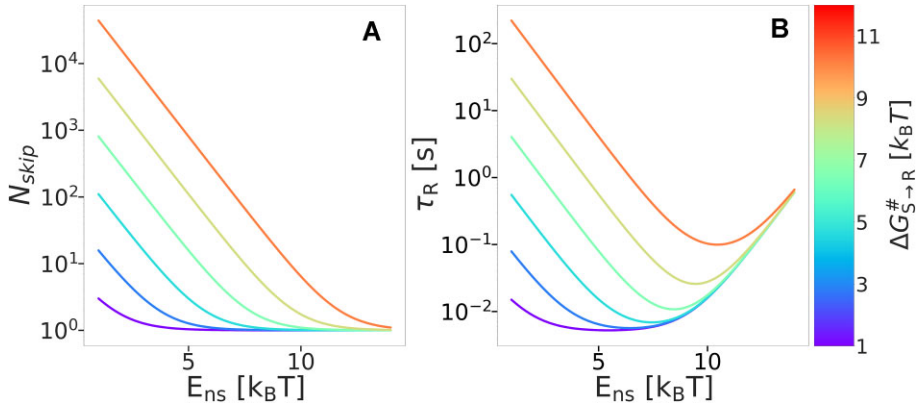


Figure 7. Trade-off between nonspecific binding and recognition kinetics. (**A**) The number of skipping events (N_{skip}; on a log scale) at different free energy barriers to the transition from searching to binding ($\Delta G^{\#}_{S\to R}$; indicated by the colorbar) is shown as a function of nonspecific binding energy (E_{ns}; calculated using Eqs. 4, 5 and 7). (**B**) The overall recognition time for SRY to recognize the target site (τ_R; on a log scale) at different $\Delta G^{\#}_{S\to R}$ values is shown as a function of E_{ns}. The strong effect of the E_{ns} (which affect k_{res} and therefore the speed of scanning nonspecific sites) on tr, indicates an increasing tradeoff between the speed of scanning nonspecific DNA (which is higher at lower E_{ns} values) and the overall recognition time (which increases with N_{skip}) as the free energy barrier increases.

search kinetics and recognition kinetics to the overall kinetics of protein–DNA binding in the genomic context has been addressed in some studies (38,39). For certain proteins, it has been suggested that a tradeoff exists between the timescales of the search and recognition kinetics, and that this tradeoff could be optimized, for example, by modifying the molecular properties of the protein (38).

In this study, we investigated how the conformational change required the DNA, which must transform from a canonical B-DNA conformation to the bent conformation necessary for SRY recognition, affects the overall kinetics for binding of the SRY transcription factor to DNA. The CG model designed and employed in the current study gradually modulates the DNA's tendency to switch between these two states, enabling us to examine how the deformability of the DNA target site affects protein binding kinetics. Through manipulation of the DNA energy landscape in our model, we observed, as anticipated, that when the DNA is less deformable (*i.e.* when it has a lower tendency to bend), recognition kinetics slow down because of a higher energetic barrier to recognition. The energetic barrier may also be associated with DNA dehydration upon transition from nonspecific to specific binding and could be influenced by the DNA sequence (79). Additionally, our findings indicate that SRY may encounter unproductive binding events when attempting to recognize its binding site on less deformable DNA molecules. Each unproductive binding event leads to SRY skipping the target site, and the frequency of such skipping events increases with decreasing DNA deformability. The occurrence of skipping events implies that SRY has to search again for the target site and again attempt to cross the energetic barrier to recognition. The number of times SRY skips over the target site is ∼20 when the target site is located in deformable DNA but increases to ∼500 on nondeformable DNA, corresponding to a recognition probability (P_f) of <0.05. This result is consistent with findings from the only other system in which skipping has been investigated to date, namely, with the experimental estimate of the number of times the Lac repressor slides over its binding site (80,81). For the Lac repressor, the recognition probability on its wild-type DNA sequence is found to be <0.1. Overall, skip events linked to conformational changes in the DNA have a profound kinetic effect. We conjecture that conformational changes in a transcription factor itself may also contribute to skipping events and, consequently, to slower recognition kinetics.

The probability of experiencing an unproductive binding event depends on various factors. From the molecular perspective, it affected by the energetic barrier for transitioning from the search to the recognition mode and by the residence time at a nonspecific site. With shorter residence times at nonspecific sites, the likelihood of crossing the energetic barrier for recognition decreases. DNA sequences that are more deformable, presenting lower energetic barriers to conformational changes, are characterized by fewer unproductive binding events. Similarly, increasing the residence time at a nonspecific site can enhance the probability of a productive binding event. However, this comes at the expense of slowing down the search of nonspecific DNA, defining a trade-off between search kinetics and recognition kinetics. We demonstrate that, indeed, at lower salt concentrations, in which nonspecific protein–DNA interactions are stronger, the number of skipping events decreases. We hypothesize that additional external factors may further reduce the number of skipping events. For example, interactions with other DBPs may prolong the residence time at the cognate site through the formation of an auxiliary interface or by an allosteric mechanism (82–84). It is also plausible that flanking DNA sequences may have a kinetic advantage via modulation of the number of skipping events (85,86).

In summary, this study reveals that DNA deformability significantly affects the biophysics of protein–DNA recognition. Beyond its expected role in specificity, such as through shape readout (4,87), our findings demonstrate that sequences with greater intrinsic deformability toward the bound conformation exhibit greater affinity. This increased affinity, stemming from enhanced deformability toward the bent conformation, resembles the enhanced affinity of transcription factors to bind to mismatches (52). Nonetheless, DNA deformability may pose a kinetic burden. The energetic barrier associated with the DNA conformational change can contribute to unsuccessful binding events, necessitating several cycles of searching nonspecific sites before successful recognition oc-

curs. Intriguingly, the strength of nonspecific protein–DNA interactions may serve as a means to modulate the recognition probability. It is likely the cell employs additional biophysical strategies to navigate and facilitate this kinetic complexity.

Data availability

The data underlying this article are available in the article and in its online supplementary material.

Supplementary data

Supplementary Data are available at NAR Online.

Funding

Funding for the open access charge was provided by the Israeli Science Foundation (grant 2072/22) and a research grant from the Estate of Gerald Alexander.

Conflict of interest statement

None declared.

References

1. Hueber,S.D. and Lohmann,I. (2008) Shaping segments: Hox gene function in the genomic age. *Bioessays*, **30**, 965–979.
2. Rezsohazy,R., Saurin,A.J., Maurel-Zaffran,C. and Graba,Y. (2015) Cellular and molecular insights into Hox protein action. *Development*, **142**, 1212–1227.
3. von Hippel,P.H. (2007) From "simple" DNA-protein interactions to the macromolecular machines of gene expression. *Annu. Rev. Biophys. Biomol. Struct.*, **36**, 79–105.
4. Rohs,R., Jin,X., West,S.M., Joshi,R., Honig,B. and Mann,R.S. (2010) Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.*, **79**, 233–269.
5. Chiu,T.P., Rao,S. and Rohs,R. (2023) Physicochemical models of protein-DNA binding with standard and modified base pairs. *Proc. Natl. Acad. Sci. U.S.A.*, **120**, e2205796120.
6. Chen,X., Tsai,M.Y. and Wolynes,P.G. (2022) The role of charge density coupled DNA bending in transcription factor sequence binding specificity: a generic mechanism for indirect readout. *J. Am. Chem. Soc.*, **144**, 1835–1845.
7. Bigman,L.S. and Levy,Y. (2023) Protein diffusion along protein and DNA lattices: role of electrostatics and disordered regions. *Annu. Rev. Biophys.*, **52**, 463–486.
8. Riggs,A., Bourgeous,S. and Cohn,M. (1970) The lac represser-operator interaction *1, *2III. Kinetic studies. *J. Mol. Biol.*, **53**, 401–417.
9. Berg,O.G., Winter,R.B. and Hippel,P.H.v. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry*, **20**, 6929–6948.
10. Halford,S.E. (2009) An end to 40 years of mistakes in DNA-protein association kinetics? *Biochem. Soc. Trans.*, **37**, 343–348.
11. von Hippel,P.H. and Berg,O.G. (1989) Facilitated target location in biological systems. *J. Biol. Chem.*, **264**, 675–678.
12. Richter,P.H. and Eigen,M. (1974) Diffusion controlled reaction rates in spheroidal geometry. Application to repressor–operator association and membrane bound enzymes. *Biophys. Chem.*, **2**, 255–263.
13. Iwahara,J. and Clore,G.M. (2006) Detecting transient intermediates in macromolecular binding by paramagnetic NMR. *Nature*, **440**, 1227–1230.
14. Iwahara,J. and 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.
15. Iwahara,J., Zweckstetter,M. and Clore,G.M. (2006) NMR structural and kinetic characterization of a homeodomain diffusing and hopping on nonspecific DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 15062–15067.
16. Sun,J., Viadiu,H., Aggarwal,A.K. and 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.
17. Blainey,P., Luo,G., Kou,S., Mangel,W., Verdine,G., Bagchi,B. and Xie,X.S. (2009) Nonspecifically bound proteins spin while diffusing along DNA. *Nature Struct. Mol. Biol.*, **16**, 1224–1229.
18. Kamagata,K., Itoh,Y. and Subekti,D.R.G. (2020) How p53 molecules solve the target DNA search problem: a review. *Int. J. Mol. Sci.*, **21**, 1031.
19. Gowers,D.M., Wilson,G.G. and 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. U.S.A.*, **102**, 15883–15888.
20. Givaty,O. and Levy,Y. (2009) Protein sliding along DNA: dynamics and structural characterization. *J. Mol. Biol.*, **385**, 1087–1097.
21. Luking,M., Elf,J. and Levy,Y. (2022) Conformational change of transcription factors from search to specific binding: a lac repressor case study. *J. Phys. Chem. B*, **126**, 9971–9984.
22. Marklund,E.G., Mahmutovic,A., Berg,O.G., Hammar,P., van der Spoel,D., Fange,D. and Elf,J. (2013) Transcription-factor binding and sliding on DNA studied using micro- and macroscopic models. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 19796–19801.
23. Mahmutovic,A., Berg,O.G. and Elf,J. (2015) What matters for lac repressor search in vivo-sliding, hopping, intersegment transfer, crowding on DNA or recognition? *Nucleic Acids Res.*, **43**, 3454–3464.
24. Felipe,C., Shin,J. and Kolomeisky,A.B. (2022) How pioneer transcription factors search for target sites on nucleosomal DNA. *J. Phys. Chem. B*, **126**, 4061–4068.
25. Shvets,A.A., Kochugaeva,M.P. and Kolomeisky,A.B. (2018) Mechanisms of protein search for targets on DNA: theoretical insights. *Molecules*, **23**, 2106.
26. Daitchman,D., Greenblatt,H.M. and Levy,Y. (2018) Diffusion of ring-shaped proteins along DNA: case study of sliding clamps. *Nucleic Acids Res.*, **46**, 5935–5949.
27. Greenblatt,H.M., Rozenberg,H., Daitchman,D. and Levy,Y. (2020) Does PCNA diffusion on DNA follow a rotation-coupled translation mechanism? *Nat. Commun.*, **11**, 5000.
28. Bigman,L.S., Greenblatt,H.M. and Levy,Y. (2021) What are the molecular requirements for protein sliding along DNA? *J. Phys. Chem. B*, **125**, 3119–3131.
29. Pal,A., Greenblatt,H.M. and Levy,Y. (2020) Prerecognition diffusion mechanism of human DNA mismatch repair proteins along DNA: Msh2-Msh3 versus Msh2-Msh6. *Biochemistry*, **59**, 4822–4832.
30. Cuculis,L., Abil,Z., Zhao,H. and Schroeder,C.M. (2016) TALE proteins search DNA using a rotationally decoupled mechanism. *Nat. Chem. Biol.*, **12**, 831–837.
31. Khazanov,N. and Levy,Y. (2011) Sliding of p53 along DNA can be modulated by its oligomeric state and by cross-talks between its constituent domains. *J. Mol. Biol.*, **408**, 335–355.
32. Zandarashvili,L., Esadze,A., Vuzman,D., Kemme,C.A., Levy,Y. and Iwahara,J. (2015) Balancing between affinity and speed in target DNA search by zinc-finger proteins via modulation of dynamic conformational ensemble. *Proc. Nat. Acad. Sci. U.S.A.*, **112**, E5142–E5149.
33. Vuzman,D., Polonsky,M. and Levy,Y. (2010) Facilitated DNA search by multidomain transcription factors: cross talk via a flexible linker. *Biophys. J.*, **99**, 1202–1211.

34. Vuzman,D., Azia,A. and Levy,Y. (2010) Searching DNA via a "Monkey Bar" mechanism: the significance of disordered tails. *J. Mol. Biol.*, **396**, 674–684.
35. Vuzman,D. and Levy,Y. (2014) The "Monkey-Bar" mechanism for searching for the DNA target site: the molecular determinants. *Isr. J. Chem.*, **54**, 1374–1381.
36. Marcovitz,A. and Levy,Y. (2013) Weak frustration regulates sliding and binding kinetics on rugged protein-DNA landscapes. *J. Phys. Chem. B*, **117**, 13005–13014.
37. Marcovitz,A. and Levy,Y. (2011) Frustration in protein-DNA binding influences conformational switching and target search kinetics. *Proc. Nat. Acad. Sci. U.S.A.*, **108**, 17957–17962.
38. Leven,I. and Levy,Y. (2019) Quantifying the two-state facilitated diffusion model of protein-DNA interactions. *Nucleic Acids Res.*, **47**, 5530–5538.
39. Slutsky,M. and Mirny,L.A. (2004) Kinetics of protein-DNA interaction: Facilitated target location in sequence-dependent potential. *Biophys. J.*, **87**, 4021–4035.
40. Wang,X., Greenblatt,H.M., Bigman,L.S., Yu,B., Pletka,C.C., Levy,Y. and Iwahara,J. (2021) Dynamic autoinhibition of the HMGB1 protein via electrostatic fuzzy interactions of intrinsically disordered regions. *J. Mol. Biol.*, **433**, 167122.
41. Tempestini,A., Monico,C., Gardini,L., Vanzi,F., Pavone,F.S. and Capitanio,M. (2018) Sliding of a single lac repressor protein along DNA is tuned by DNA sequence and molecular switching. *Nucleic Acids Res.*, **46**, 5001–5011.
42. Zhou,H.X. (2011) Rapid search for specific sites on DNA through conformational switch of nonspecifically bound proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 8651–8656.
43. Bauer,M. and Metzler,R. (2012) Generalized facilitated diffusion model for DNA-binding proteins with search and recognition states. *Biophys. J.*, **102**, 2321–2330.
44. Schnepf,M., von Reutern,M., Ludwig,C., Jung,C. and Gaul,U. (2020) Transcription factor binding affinities and DNA shape readout. *iScience*, **23**, 101694.
45. Privalov,P.L. and Crane-Robinson,C. (2018) Forces maintaining the DNA double helix and its complexes with transcription factors. *Prog. Biophys. Mol. Biol.*, **135**, 30–48.
46. Tan,C., Terakawa,T. and Takada,S. (2016) Dynamic coupling among protein binding, sliding, and DNA bending revealed by molecular dynamics. *J. Am. Chem. Soc.*, **138**, 8512–8522.
47. Haran,T.E. and Mohanty,U. (2009) The unique structure of A-tracts and intrinsic DNA bending. *Q. Rev. Biophys.*, **42**, 41–81.
48. Shatzky-Schwartz,M., Arbuckle,N.D., Eisenstein,M., Rabinovich,D., Bareket-Samish,A., Haran,T.E., Luisi,B.F. and Shakked,Z. (1997) X-ray and solution studies of DNA oligomers and implications for the structural basis of A-tract-dependent curvature. *J. Mol. Biol.*, **267**, 595–623.
49. Slattery,M., Zhou,T., Yang,L., Dantas Machado,A.C., Gordan,R. and Rohs,R. (2014) Absence of a simple code: how transcription factors read the genome. *Trends Biochem. Sci.*, **39**, 381–399.
50. Joshi,R., Passner,J.M., Rohs,R., Jain,R., Sosinsky,A., Crickmore,M.A., Jacob,V., Aggarwal,A.K., Honig,B. and Mann,R.S. (2007) Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell*, **131**, 530–543.
51. Zeiske,T., Baburajendran,N., Kaczynska,A., Brasch,J., Palmer,A.G. 3rd, Shapiro,L., Honig,B. and Mann,R.S. (2018) Intrinsic DNA shape accounts for affinity differences between Hox-cofactor binding sites. *Cell Rep.*, **24**, 2221–2230.
52. Afek,A., Shi,H., Rangadurai,A., Sahay,H., Senitzki,A., Xhani,S., Fang,M., Salinas,R., Mielko,Z., Pufall,M.A., *et al.* (2020) DNA mismatches reveal conformational penalties in protein-DNA recognition. *Nature*, **587**, 291–296.
53. Romanuka,J., Folkers,G.E., Biris,N., Tishchenko,E., Wienk,H., Bonvin,A.M., Kaptein,R. and Boelens,R. (2009) Specificity and affinity of Lac repressor for the auxiliary operators O2 and O3 are explained by the structures of their protein-DNA complexes. *J. Mol. Biol.*, **390**, 478–489.
54. Berta,P., Hawkins,J.R., Sinclair,A.H., Taylor,A., Griffiths,B.L., Goodfellow,P.N. and Fellous,M. (1990) Genetic evidence equating SRY and the testis-determining factor. *Nature*, **348**, 448–450.
55. Murphy,E.C., Zhurkin,V.B., Louis,J.M., Cornilescu,G. and Clore,G.M. (2001) Structural basis for SRY-dependent 46-X,Y sex reversal: modulation of DNA bending by a naturally occurring point mutation. *J. Mol. Biol.*, **312**, 481–499.
56. Ferrari,S., Harley,V.R., Pontiggia,A., Goodfellow,P.N., Lovell-Badge,R. and Bianchi,M.E. (1992) SRY, like HMG1, recognizes sharp angles in DNA. *EMBO J.*, **11**, 4497–4506.
57. Murphy,F.V.t., Sweet,R.M. and Churchill,M.E. (1999) The structure of a chromosomal high mobility group protein-DNA complex reveals sequence-neutral mechanisms important for non-sequence-specific DNA recognition. *EMBO J.*, **18**, 6610–6618.
58. Allain,F.H., Yen,Y.M., Masse,J.E., Schultze,P., Dieckmann,T., Johnson,R.C. and Feigon,J. (1999) Solution structure of the HMG protein NHP6A and its interaction with DNA reveals the structural determinants for non-sequence-specific binding. *EMBO J.*, **18**, 2563–2579.
59. King,C.Y. and Weiss,M.A. (1993) The SRY high-mobility-group box recognizes DNA by partial intercalation in the minor groove: A topological mechanism of sequence specificity. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 11990–11994.
60. Bouvier,B. and Lavery,R. (2009) A free energy pathway for the interaction of the SRY protein with its binding site on DNA from atomistic simulations. *J. Am. Chem. Soc.*, **131**, 9864–9865.
61. Khazanov,N., Marcovitz,A. and Levy,Y. (2013) Asymmetric DNA-Search dynamics by symmetric dimeric proteins. *Biochemistry*, **52**, 5335–5344.
62. Assenza,S. and Perez,R. (2022) Accurate sequence-dependent coarse-grained model for conformational and elastic properties of double-stranded DNA. *J. Chem. Theory Comput.*, **18**, 3239–3256.
63. Li,S., Olson,W.K. and Lu,X.J. (2019) Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. *Nucleic Acids Res.*, **47**, W26–W34.
64. Lammert,H., Schug,A. and Onuchic,J.N. (2009) Robustness and generalization of structure-based models for protein folding and function. *Proteins Struct. Funct. Bioinform.*, **77**, 881–891.
65. Okamura,K., Sakaguchi,H., Sakamoto-Abutani,R., Nakanishi,M., Nishimura,K., Yamazaki-Inoue,M., Ohtaka,M., Periasamy,V.S., Alshatwi,A.A., Higuchi,A., *et al.* (2016) Distinctive features of single nucleotide alterations in induced pluripotent stem cells with different types of DNA repair deficiency disorders. *Sci. Rep.*, **6**, 26342.
66. Okazaki,K. and Takada,S. (2008) Dynamic energy landscape view of coupled binding and protein conformational change: induced-fit versus population-shift mechanisms. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 11182–11187.
67. Whitford,P.C., Miyashita,O., Levy,Y. and Onuchic,J.N. (2007) Conformational transitions of adenylate kinase: switching by cracking. *J. Mol. Biol.*, **366**, 1661–1671.
68. Giri Rao,V.V., Desikan,R., Ayappa,K.G. and Gosavi,S. (2016) Capturing the membrane-triggered conformational transition of an alpha-helical pore-forming toxin. *J. Phys. Chem. B*, **120**, 12064–12078.
69. Jayanthi,L.P., Mascarenhas,N.M. and Gosavi,S. (2020) Structure dictates the mechanism of ligand recognition in the histidine and maltose binding proteins. *Curr. Res. Struct. Biol.*, **2**, 180–190.
70. Ramirez-Sarmiento,C.A., Noel,J.K., Valenzuela,S.L. and Artsimovitch,I. (2015) Interdomain contacts control native state switching of RfaH on a dual-funneled landscape. *PLoS Comput. Biol.*, **11**, e1004379.
71. Bhattacherjee,A. and Levy,Y. (2014) Search by proteins for their DNA target site: 1. The effect of DNA conformation on protein sliding. *Nucleic Acids Res.*, **42**, 12404–12414.
72. Tan,C. and Takada,S. (2018) Dynamic and structural modeling of the specificity in protein-DNA interactions guided by binding assay and structure data. *J. Chem. Theory Comput.*, **14**, 3877–3889.

73. Machado,M.R. and Pantano,S. (2015) Exploring LacI-DNA dynamics by multiscale simulations using the SIRAH force field. *J. Chem. Theory Comput.*, **11**, 5012–5023.
74. Kochugaeva,M.P., Shvets,A.A. and Kolomeisky,A.B. (2016) How conformational dynamics influences the protein search for targets on DNA. *J. Phys.*, **49**, 444004.
75. Mirny,L. and Slutsky,M. (2009) How a protein searches for its site ob DNA: the mechanism of facilitated diffusion. *J of Physics A-Mathematical and Theoretical*, **42**, 434013.
76. Kamagata,K., Mano,E., Ouchi,K., Kanbayashi,S. and Johnson,R.C. (2018) High free-energy barrier of 1D diffusion along DNA by architectural DNA-binding proteins. *J. Mol. Biol.*, **430**, 655–667.
77. Tafvizi,A., Mirny,L.A. and van Oijen,A.M. (2011) Dancing on DNA: kinetic aspects of search processes on DNA. *ChemPhysChem*, **12**, 1481–1489.
78. Kolomeisky,A.B. (2011) Physics of protein-DNA interactions: mechanisms of facilitated target search. *Phys. Chem. Chem. Phys.*, **13**, 2088–2095.
79. Bouvier,B., Zakrzewska,K. and Lavery,R. (2011) Protein-DNA recognition triggered by a DNA conformational switch. *Angew. Chem. Int. Ed Engl.*, **50**, 6516–6518.
80. Hammar,P., Leroy,P., Mahmutovic,A., Marklund,E.G., Berg,O.G. and Elf,J. (2012) The lac repressor displays facilitated diffusion in living cells. *Science*, **336**, 1595–1598.
81. Marklund,E., van Oosten,B., Mao,G., Amselem,E., Kipper,K., Sabantsev,A., Emmerich,A., Globisch,D., Zheng,X., Lehmann,L.C., *et al.* (2020) DNA surface exploration and operator bypassing during target search. *Nature*, **583**, 858–861.
82. Kim,S., Brostromer,E., Xing,D., Jin,J., Chong,S., Ge,H., Wang,S., Gu,C., Yang,L., Gao,Y.Q., *et al.* (2013) Probing allostery through DNA. *Science*, **339**, 816–819.
83. Morgunova,E. and Taipale,J. (2017) Structural perspective of cooperative transcription factor binding. *Curr. Opin. Struct. Biol.*, **47**, 1–8.
84. Rosenblum,G., Elad,N., Rozenberg,H., Wiggers,F., Jungwirth,J. and Hofmann,H. (2021) Allostery through DNA drives phenotype switching. *Nat. Commun.*, **12**, 2967.
85. Castellanos,M., Mothi,N. and Munoz,V. (2020) Eukaryotic transcription factors can track and control their target genes using DNA antennas. *Nat. Commun.*, **11**, 540.
86. Horton,C.A., Alexandari,A.M., Hayes,M.G.B., Marklund,E., Schaepe,J.M., Aditham,A.K., Shah,N., Suzuki,P.H., Shrikumar,A., Afek,A., *et al.* (2023) Short tandem repeats bind transcription factors to tune eukaryotic gene expression. *Science*, **381**, eadd1250.
87. Rohs,R., West,S., Sosinnsky,A., Liu,P., Mann,R. and Honig,B. (2009) The role of DNA shape in protein-DNA recognition. *Nature*, **461**, 1248–1253.

Received: January 30, 2024. **Revised:** April 9, 2024. **Editorial Decision:** April 11, 2024. **Accepted:** April 16, 2024