JACS Au
AN OPEN ACCESS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Open Access
This article is licensed under CC-BY-NC-ND 4.0

pubs.acs.org/jacsau
Article

# Role of Shape Deformation of DNA-Binding Sites in Regulating the Efficiency and Specificity in Their Recognition by DNA-Binding Proteins

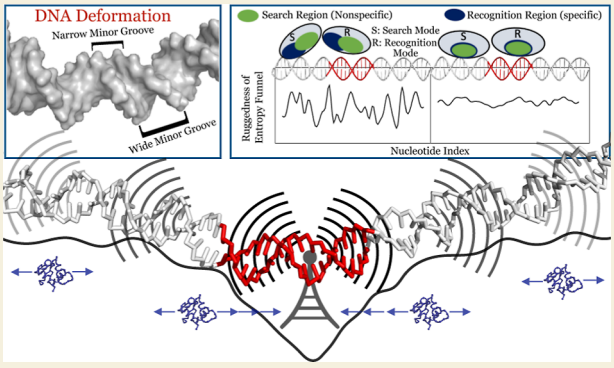
Sangeeta, Sujeet Kumar Mishra, and Arnab Bhattacherjee*

Cite This: JACS Au 2024, 4, 2640−2655

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** Accurate transcription of genetic information is crucial, involving precise recognition of the binding motifs by DNA-binding proteins. While some proteins rely on short-range hydrophobic and hydrogen bonding interactions at binding sites, others employ a DNA shape readout mechanism for specific recognition. In this mechanism, variations in DNA shape at the binding motif resulted from either inherent flexibility or binding of proteins at adjacent sites are sensed and capitalized by the searching proteins to locate them specifically. Through extensive computer simulations, we investigate both scenarios to uncover the underlying mechanism and origin of specificity in the DNA shape readout mechanism. Our findings reveal that deformation in shape at the binding motif creates an entropy funnel, allowing information about altered shapes to manifest as fluctuations in minor groove widths. This signal enhances the efficiency of nonspecific search of nearby proteins by directing their movement toward the binding site, primarily driven by a gain in entropy. We propose this as a generic mechanism for DNA shape readout, where specificity arises from the alignment between the molecular frustration of the searching protein and the ruggedness of the entropic funnel governed by molecular features of the protein and arrangement of the DNA bases at the binding site, respectively.




**KEYWORDS:** *protein−DNA interaction, DNA shape readout, transfer entropy, protein target search, facilitated diffusion*

## INTRODUCTION

Vital cellular processes, including DNA damage repair, transcription, and replication, are orchestrated by protein−DNA interactions. At the core of these interactions lies the complex process through which proteins carefully search for and pinpoint their precise binding sites on DNA. This fundamental mechanism serves as the linchpin, facilitating the transmission of genetic information and instigating a myriad of essential biological processes. The search process is thought to be facilitated by a combination of 1D protein dynamics on DNA and 3D diffusion of the searching protein within the bulk solution.[1,2] On DNA, the protein engages in "sliding" dynamics, characterized by its movement from an initial nonspecific site, often while simultaneously rotating along the DNA major groove. A prime example of sliding is observed in the binding of the lactose (lac) repressor to its operator sequences, where DNA binding relies entirely on electrostatic interactions, leading to diffusion along an equipotential surface.[3,4] Alternatively, proteins may "hop" from one site to another in three-dimensional space. This involves dissociating from the original binding site and subsequently attaching to a new site on the same DNA chain.[5] A similar but distinct mode of movement, known as intersegmental transfer, involves a jump to a site on another DNA segment, typically after the formation of a bridge complex that encompasses multiple protein domains.[6−10]

Notwithstanding the wealth of information regarding protein target search modes on DNA, the process by which proteins specifically locate their target sites remains unclear, especially considering they sample only a minuscule fraction of the transcriptionally accessible genome (roughly 2% of the 2.5-gigabase-pair genome, as indicated by single-molecule fluorescence tracking of Sox2 transcription factor (TF) in mammalian cells).[11] An in-depth analysis of the three-dimensional structures of numerous protein−DNA complexes suggests that the specific binding of protein and DNA primarily hinges on hydrophobic and hydrogen bond interactions between the amino acid side chains and the

Received: April 30, 2024
Revised: May 21, 2024
Accepted: May 22, 2024
Published: June 18, 2024

© 2024 The Authors. Published by American Chemical Society

nucleotide bases of the binding motif, a mechanism referred to as "DNA base readout".[12−14] Essentially, reading the specific bases at the binding motif is common to all protein−DNA complexes upon specific binding. Nonetheless, it is important to note that both the hydrophobic and hydrogen bond interactions are short-ranged. While they govern the binding specificity once the binding site is reached by the protein, they are incapable of regulating the target search process of proteins from a far distance. Thus, for an unbiased combination of 3D and 1D search modes where the protein scans DNA one-dimensionally approximately only 50 base pairs around each landing site after a 3D diffusion period,[15] it would still be difficult to locate the specific binding site by the proteins in a base-by-base scanning mode (sliding). Even if the protein lands on DNA near the binding motif, an equally probable 1D sliding direction along the DNA contour may lead it away from the target site. Subsequent 3D diffusion events can completely nullify the advantage of being in close vicinity to the binding motif. Consequently, the protein is expected to make multiple attempts to locate its binding motif specifically, contrary to experimental findings that suggest it achieves this feat in fewer attempts.[11]

An alternate mechanism of specific recognition of binding motifs by DNA-binding proteins, referred to as "DNA shape readout", is proposed based on the frequently observed nonstandard DNA conformations within specific protein−DNA complexes.[14,16−23] For example, it was found that the Drosophila Hox protein Sex combs reduced (SCR) senses the sequence-dependent alteration of minor groove width and corresponding variations in the electrostatic potential to distinguish small differences in the nucleotide sequences.[24] The DNA shape-dependency indeed has substantially improved the accuracy of predicting the genome-wide binding specificities of transcription factors.[14,25−27] However, how the factor affects the overall search dynamics of proteins as they strive for efficient and specific recognition of binding motifs remains a mystery. Unveiling this phenomenon also holds paramount significance in our quest to comprehend the intricate functionality of densely packed chromatin fibers. These fibers inherently display substantial DNA deformations due to their unique sequences and can further undergo induced deformations upon interacting with other proteins.[28,29] In both scenarios, these deformations are intimately linked to the precise positioning of molecular machinery responsible for detecting crucial signals within genomes, potentially regulating gene expression.

Comprehending the intricate principles that govern the orchestration of gene regulation through DNA shape readout is a dauntingly challenging task, primarily due to the lack of suitable experimental techniques capable of dissecting and evaluating the contributions of all regulating factors. In this study, we therefore delve into how local DNA shape and flexibility orchestrate the specific recognition of binding motifs by conducting extensive computer simulations using a carefully selected coarse-grained model for both protein and DNA. This model has been widely employed in previous research endeavors, including our own,[30−37] as well as by others,[9,10,38] to elucidate the mechanisms of protein search on DNA with varying topological characteristics. Our investigation focuses on two distinct scenarios: first, the search dynamics of the PU.1 protein as it seeks its target site along a linear stretch of DNA, and second, the coordinated search for their respective target sequences by a pair of proteins known to bind cooperatively, namely, Fis and Xis. In the case of PU.1, the experimentally determined complex structure with DNA reveals a notable deviation in the shape of the DNA from the B-DNA molecule owing to sequence-dependent deformation of the binding site.[39] PU.1 exhibits tissue-specific expression patterns, and its varying expression levels dictate the course of differentiation in hematopoietic stem cells.[40] Therefore, maintaining precise levels of PU.1 is essential for the differentiation process into specific blood cell lineages as even minor reductions in its expression can predispose individuals to leukemia. In the second scenario, it has been established that the binding of Fis to its target sequence induces significant changes in the minor groove geometry of the adjacent Xis binding site, resulting in cooperative binding of the latter.[41] Fis, a bacterial nucleoid protein, plays essential roles in transcription, replication, and recombination processes.[42] The bacteriophage $\lambda$ excisionase (Xis) is a DNA-binding protein with specificity for particular DNA sequences that is essential for facilitating excisive recombination. The*E. coli* Fis protein facilitates the recruitment of the phage-encoded Xis protein to the *attR* recombination site, which represents a pivotal regulatory step in the formation of the excisive intasome. We chose PU.1 and Fis-Xis systems as representatives of DNA deformation induced by sequence characteristics and deformation resulting from the association of a protein, respectively. Through a comprehensive analysis of various aspects of the target search mechanism in both scenarios, we discern that the recognition of binding motifs by these two proteins is notably enhanced when the binding sites exhibit a deformed shape. In contrast, in the absence of shape deformation at the binding motifs, the proteins compromise both efficiency and specificity in locating the target sites. A thermodynamic examination of the search path uncovers that this facilitated and specific recognition of the binding motifs using the information on DNA shape at the target DNA site is governed by an entropy-enthalpy reinforcement, with entropy making the predominant contribution. Furthermore, the influence of the deformed shape of the binding motif is not limited locally; rather, it induces a significantly broad entropic funnel encompassing adjacent flanking nucleotide bases. The funnel noticeably biases the protein's target search dynamics in the vicinity of target DNA site without altering its fundamental search modes. An exhaustive analysis of the entire information landscape along the DNA contour confirms that the effect is akin to the vibration on a string propagating through the DNA, manifested as fluctuations in the minor groove width. These fluctuations serve as a long-range driving force that explains the observed efficiency and specificity in the DNA shape readout mechanism. The findings unveil an unknown facet of the protein target search mechanism on DNA, suggesting a potential evolutionary connection between the molecular characteristics of DNA-binding proteins and the sequences of their target DNA sites.

## MATERIALS AND METHODS

### Protein and DNA Models

Appropriately tailored coarse-grained computational models offer substantial advantages for the investigation of intricate biological processes.[43−48] In this study, we employ a coarse-grained protein−DNA model, akin to those utilized in our previous investigations.[30−37] The salient features of the model are briefly outlined here, with detailed information provided in the Supporting Information. In this model, each amino acid within the protein structure is represented by


https://doi.org/10.1021/jacsau.4c00393
JACS Au 2024, 4, 2640−2655

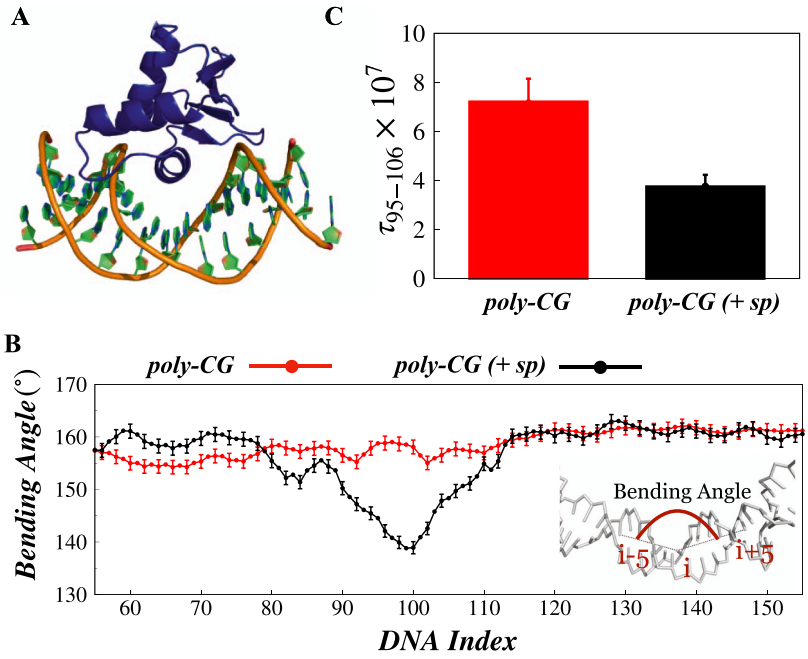



**Figure 1.** PU.1 and DNA. (A) Crystal structure of PU.1 bound to a specific DNA sequence. PU.1 is shown in blue color. (B) Bending angle of poly-CG and poly-CG (+sp) DNA sequences in the absence of PU.1. Here, poly-CG (+sp) represents poly-CG with a specific binding motif of PU.1 placed in the middle of the DNA contour. The inset shows the schematic for the calculation of bending angle for every DNA index. (C) Average MD simulation steps required by PU.1 to reach the 95−106 base pair index, where the binding motif is placed in a poly-CG (+sp) sequence.

a single bead positioned at its respective $C_\alpha$ location. The proteins employed in this investigation include PU.1, Fis, and Xis. The crystal structures for PU.1, Fis, and Xis are acquired from the RCSB Protein Data Bank [PDB ID: 1PUE[39] and 6POT,[41] serving as the native structures for the study. The energetics of the protein are delineated by a native-topology-based model, employing a Lennard-Jones potential to incorporate native contacts observed in the crystal structure. This structure-based potential manifests as a funnel-like energy landscape conducive to protein folding,[49] and it has been widely employed in the exploration of molecular details and biophysics in both protein−protein[50] and protein−DNA interactions.[30−37] For the DNA molecule, we adopt the 3SPN.2C model developed by Pablo *et al.*,[51,52] where each DNA nucleotide is represented by three beads located at the respective centers of phosphate, sugar, and nitrogenous base. This model has demonstrated success in accurately estimating melting temperature and persistence length and faithfully reproducing DNA hybridization across various compositions and ionic strengths.[53] Detailed descriptions of the protein and DNA energetics, along with reference parameters, are provided in the Supporting Information (Tables S1−S8). Electrostatic interactions are modeled between charged entities using the Debye−Hückel potential, accounting for the salt effect. A unit negative charge is assigned to Asp and Glu residues and a unit positive charge to Arg and Lys amino acid residues. To consider the impact of counterion condensation, a negative charge of 0.6 is assigned to each phosphate bead of DNA. The potency of electrostatic interactions between charged protein and DNA beads is adjusted by a factor of 1.67 to restore the local charge of phosphate beads to −1, as employed in previous studies.[35,54] It is noteworthy to highlight that this coarse-grained depiction, incorporating an implicit solvent model and accounting for salt effects, has been previously employed by us and other researchers to delve into the critical molecular intricacies including thermodynamics and kinetics of protein−DNA interactions.[8,9,30−37,43,45,55−59]

## Protein−DNA Interactions

The interaction dynamics between the protein and DNA molecules is modeled using the nonspecific interactions. These nonspecific interactions between the searching protein and DNA occur in two modes. First, an excluded volume interaction, exerted during the nonspecific encounter between the two biomolecules, is modeled by a purely repulsive potential. Second, an electrostatic interaction between the phosphate beads of DNA and charged residues of the protein influences the protein's dynamics and is represented by the Debye−Hückel potential (details in Supporting Information). It is essential to note that the applicability of the Debye−Hückel theory is limited to low salt conditions, becoming invalid for ionic concentrations exceeding 0.5 M.[44] Despite these limitations, the Debye−Hückel potential has proven successful in capturing essential features of nucleic acid biophysics.[38,60]

## Simulation Protocol

The initial configuration of a 200 bp B-DNA is generated utilizing the w3DNA web server (3D DNA structure).[61] The PU.1 and Fis & Xis target motifs, identified in the crystal structure, are positioned within the poly-CG DNA stretch at the center. The DNA is centered in the simulation box, measuring 300 × 300 × 900 Å, and subjected to periodic boundary conditions. Placing the searching proteins at a distance of 50 Å from the DNA surface, away from the target motif, ensures appropriate simulation conditions. In the simulation, proteins are initially situated at various positions along the DNA contour. The dynamics of both the protein and DNA molecules are simulated using the Langevin dynamics, incorporating a friction coefficient ($\gamma$) of value 0.05 kg/s at a temperature of $T$ = 300 K and a salt concentration of 150 mM. To ensure robust statistical analysis, 20 independent simulations, each lasting 2 × $10^8$ MD steps, are conducted for every system, employing an in-house code on a 7.74 teraflop high-performance cluster.

## Analysis

In accordance with the methodology outlined in our prior research, distinct search modes, namely, sliding, hopping, and 3D diffusion, are delineated. The protein engages in 3D diffusion within the bulk when positioned at a considerable distance from the DNA surface. Conversely, when in close proximity to DNA, the protein employs sliding dynamics to scrutinize the DNA base pairs. In a snapshot, if the protein is close enough to the DNA but does not meet sliding criteria, then the protein performs hopping dynamics. As a result, the


https://doi.org/10.1021/jacsau.4c00393
JACS Au 2024, 4, 2640−2655

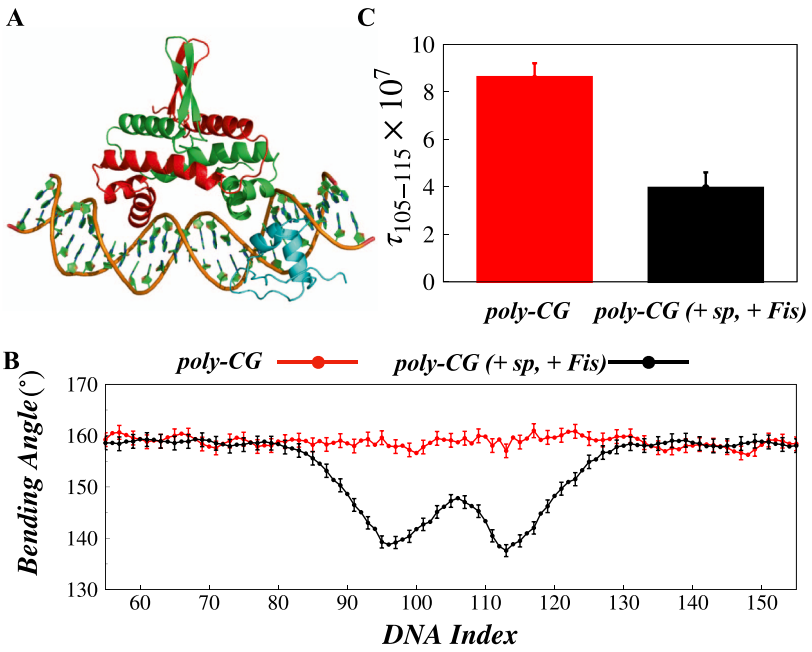



**Figure 2.** Fis, Xis, and DNA. (A) Crystal structure of Fis and Xis bound to its specific DNA motif. The Fis protein consists of a homodimer, represented by two chains (depicted in red and green colors), while Xis is illustrated in cyan. (B) Bending angle of poly-CG and poly-CG (+sp, + Fis) DNA sequences. Here, poly-CG (+sp, + Fis) represents poly-CG sequence with Fis specifically bound to its binding motif positioned in the center. (C) Average MD simulation steps required by Xis to reach its target, where the binding motif of Xis is placed in a poly-CG (+sp, + Fis) sequence.

protein can either hop or slide along the DNA contour when it searches for it nonspecifically near its surface. We monitor the distance, $R$, between the center of mass of the protein's recognition region and the closest DNA base pair from it. For the PU.1-DNA system, a snapshot is considered to be in sliding mode if the protein is at a distance of less than 20 Å from DNA ($R \leq 20$ Å). The protein is considered to perform hopping if it satisfies the condition of 20 Å < $R \leq 23$ Å. If the protein is found at a distance greater than 23 Å from the DNA (i.e., $R > 23$ Å), the protein is considered to be in 3D diffusion mode. For Xis protein in the Fis-Xis-DNA system, $R \leq 20$ Å represents sliding, 20 Å < $R \leq 26$ Å represents hopping and $R > 26$ Å is used for diffusion.

## RESULTS AND DISCUSSION

### Deformation at the DNA Binding Site Facilitates Its Recognition

Previously, it was proposed that the specificity of protein binding at the target DNA site is strongly influenced by the local bending of the respective target DNA sites.[38] Expanding on this idea, our current interest lies in investigating how DNA deformation at the binding site contributes to its specific recognition by DNA-binding proteins. Local DNA bending can result from either the inherent flexibility of the binding motif or the presence of another protein at an adjacent site. To explore these possibilities, we considered both scenarios. Accordingly, we first focus on the PU.1 system (as depicted in Figures 1A and S1A), a transcription factor belonging to the ETS family with significant roles in hematopoiesis and B-cell development.[39,62] The experimentally determined structure of PU.1 with its short DNA-binding motif AAAGGGGAAGT (PDB ID: 1PUE) suggests a significant deviation in the DNA conformation at the binding site when compared to the typical B-DNA structure (as shown in Figure 1A). PU.1 predominantly binds in the major groove of the DNA with some contact formation in the minor groove. The DNA segment exhibits the structural distortion due to the intrinsic flexibility of the DNA-binding motif.[59] Our second focus lies in the binding of Fis and Xis proteins to their respective target DNA sites. Fis forms a homodimer, comprised of 98-residue subunits derived from *E. coli* (as depicted in Figures 2A and S1B). Classified as a bacterial nucleoid protein, Fis plays a vital role in critical cellular processes, including transcription, replication, and recombination reactions.[42] Xis, shown in Figures 2A and S1C, is a winged-helix DNA-binding protein expressed at elevated levels shortly after prophage induction. However, Xis itself demonstrates low binding specificity in absence of Fis.[63−66] Fis binds to the DNA major groove, with each helix-turn-helix (HTH) unit anchored within the groove via distinct contacts with the DNA backbone. This binding necessitates the bending of DNA shape to accommodate the protein into two adjacent major grooves.[41] Consequently, the minor groove undergoes substantial deformation, expanding by approximately 50% opposite to where Fis inserts its recognition helix while narrowing down to roughly half its canonical width at the center of the Fis interface. In the present study, our motivation is to understand how Fis-induced deformation in the DNA shape influences the target search dynamics of the Xis protein. Accordingly, we setup the experiment by keeping Fis bound to its binding motif embedded in a poly-CG DNA stretch. The specific binding mechanism of Fis is represented by a straightforward Lenard-Jones potential, wherein the intensity is adjusted to replicate the global bending angle obtained from the recently resolved crystal structure of the Fis-Xis-DNA complex bound to 27 bp DNA.[67−69] Figure S2 shows the schematic representation for global bending angle calculation estimated from the starting point midpoint and end point of the target sequence labeled as P, Q, and R respectively.

Before delving into further details, we first assess whether our model adequately captures the subtle structural alterations



https://doi.org/10.1021/jacsau.4c00393
JACS Au 2024, 4, 2640−2655

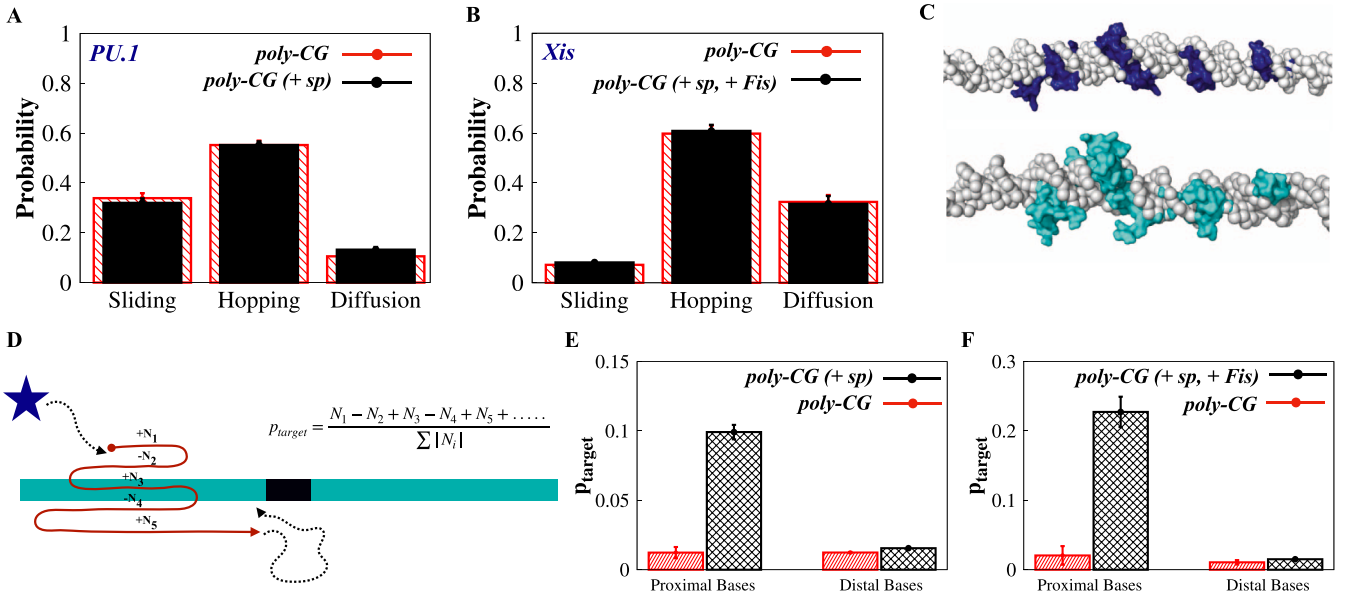



**Figure 3.** Characterization of PU.1 and Xis diffusion on DNA track. Probability of sliding, hopping, and 3D diffusion of (A) PU.1 protein on poly-CG and poly-CG (+sp) DNA sequence and (B) Xis protein on poly-CG and poly-CG (+sp, + Fis) DNA sequence. (C) Sample traces of sliding of PU.1 protein (blue color, upper panel) and Xis protein (pink color, lower panel) along the major grooves of DNA. (D) Schematic representation of directed 1D movement toward the target site, in which DNA sites are sampled several times by the protein before changing into 3D search mode. Blue color shape represents the searching protein and black color segment on the DNA represents the binding motif of the searching protein. $p_{target}$ denotes the direction of the protein moving toward the target site during each 1D scanning event. $p_{target}$ of (E) PU.1 protein on poly-CG and poly-CG (+sp) sequence and (F) Xis protein on poly-CG and poly-CG (+sp, + Fis) sequence on proximal and distant bases.

in DNA, which can be used further as a basis to probe their interactions with DNA-binding proteins (DBPs) by searching for their specific DNA sites. To do so, we consider a 200 bp DNA stretch featuring the binding motif in the middle (base pair position 95−106) flanked by poly-CG sequences. From here on, we refer to this as poly-CG (+sp) system. A separate system comprises a 200 base pair (bp) DNA segment containing Fis and Xis binding motifs positioned at site 91−117 bp, flanked by poly-CG sequences. Hereafter, we will refer to this as the poly-CG (+sp, + Fis) sequence. We analyze the structural changes of the DNA-binding motifs in the absence of searching protein for poly-CG (+sp) in the PU.1 system and the Fis-bound DNA motif (poly-CG (+sp, + Fis)) in the Fis and Xis systems, utilizing both coarse-grained and all-atom simulations (details in Supporting Information). To accomplish this, we quantify the minor groove width of the binding motif, and the results are depicted in Figures S3 and S4 for the PU.1 and Fis-Xis systems, respectively. The findings depict an excellent match between the coarse-grained and all-atom simulations, validating the suitability of our coarse-grained models. Further validation of the model is provided in the Supporting Information (Figures S5 and S6).

To investigate whether the intrinsic flexibility of DNA at the binding site plays any role in its recognition by the PU.1 protein, we examine the nonspecific binding dynamics of PU.1 with poly-CG (+sp) DNA. As a control, we consider another DNA sequence consisting of only the poly-CG DNA sequence with no specific binding motif embedded in it. This is referred to as poly-CG sequence hereafter. To quantify the conformational distortion in the respective DNA structures, we perform simulations of both DNA sequences in the absence of PU.1 under a salt concentration of 150 mM and measured the DNA bending angle at each point for both poly-CG (+sp) and poly-CG sequences.[59,70] The results presented in Figure 1B reveal that the poly-CG DNA stretch is uniformly bent, with a local bending angle of approximately 160°, while the specific DNA-binding motif, when flanked by poly-CG DNA sequences (as in poly-CG (+sp)), exhibits a greater bending of approximately 20° at the binding site, in line with its AT-rich sequence composition at the specific site region. It should be noted that the bending of DNA in the poly-CG (+sp) sequence is not limited within the binding motif; rather, it induces bending in the adjacent (up to 12 bps on both sides of the binding motif) flanking GC sequences as well. To explore whether this DNA deformation in the DNA shape at the binding site impacts the target search dynamics of PU.1, we estimate the average time required for PU.1 to reach any of the 95−106 bps nonspecifically, starting from a distance. Notably, the nonspecific search of the protein is solely governed by electrostatic interactions between charged amino acid residues and nucleotide bases. This particular stretch of DNA bases encompasses the binding motif of PU.1 in the poly-CG (+sp) sequence but only CG bases in the poly-CG sequence. Figure 1C illustrates that PU.1 reaches the 95−106 bps stretch approximately two times faster when the DNA segment features its binding motif (poly-CG (+sp) sequence) compared to the poly-CG sequence that lacks any distinct deformation in the DNA shape. The result suggests that the nonspecific search of PU.1 for its target site is kinetically facilitated by localized sequence-dependent deformation of DNA shape at the DNA-binding site. Estimation of the position-specific bending in the Fis and Xis systems highlights significant differences in DNA shape deformation between the poly-CG sequence alone and the same sequence with Fis specifically bound to its binding motif (Figure 2B) positioned in the center. To investigate how the Fis-induced deformation in the DNA influences the target search dynamics of the Xis protein, we study the dynamics of the Xis protein on a 200 bp long DNA segment with a poly-


https://doi.org/10.1021/jacsau.4c00393
JACS Au 2024, 4, 2640−2655

Table 1. Thermodynamic Parameters Calculated from the Simulation Trajectories of the DNA and PU.1 in Free Systems and PU.1-DNA Nonspecific Complex at $T$ = 300 K[a]

| systems | $H$ | $\Delta H$ | $TS_{\infty}$ | $T\Delta S_{\infty}$ |
|---|---|---|---|---|
| free systems | −2195.53 ± 2.56 | | 296.81 ± 0.14 | |
| | | −2.70 ± 3.92 | | 36.99 ± 0.20 |
| DNA−PU.1 complex | −2198.23 ± 2.98 | | 333.80 ± 0.14 | |

[a]All the values are in kcal/mol ± standard errors.

CG sequence and compare the result to poly-CG (+sp, + Fis). We refer the first sequence as poly-CG sequence. The result presented in Figure 2C illustrates that the time required for Xis to reach the 105−115 bp (binding motif of Xis) segment through nonspecific diffusion over DNA is approximately 2.2 times slower on the poly-CG sequence compared to when the Xis protein nonspecifically scans the poly-CG (+sp, + Fis) sequence. To confirm that the speed up observed in locating the binding motif by the Xis protein is due to Fis-induced deformation in the DNA shape and not through interaction with Fis protein itself, we artificially constrain the bending of the DNA segment to approximately 60−65° using a pseudo harmonic-bond at the two ends of the binding motif. This reproduces the impact of Fis-induced deformation in the absence of Fis binding. The corresponding search time of Xis shown in Figure S7 is similar to that on a poly-CG (+sp, + Fis) sequence, suggesting that Fis does not interact directly to influence the binding of Xis, as indicated in previous studies,[41,71] rather the facilitated search dynamics of Xis protein is due to deformation in DNA shape induced by Fis binding.

## Protein Exhibits a Biased Diffusion in the Close Proximity to Its Binding Motif

Having observed that deformation in DNA shape at the binding site results in the faster arrival of the searching proteins at those sites, we explore how such deformation influences the search process. A straightforward approach to assess the impact is to monitor differences in the protein's search mechanism in the presence and absence of deformation at the DNA-binding motif. Accordingly, we analyze the footprint of the searching protein throughout the trajectory until it reaches the DNA-binding site, estimating its propensities for adopting various transport modes following the prescribed method in the Materials and Methods. Figure 3A,B illustrates the propensities of various target search modes adopted by PU.1 and Xis proteins, respectively. For PU.1 searching on DNA poly-CG (+sp), the propensity of 1D sliding is 32%, hopping is 54%, and 3D diffusion is 14%. The same on DNA with poly-CG only are 34, 54, and 12%, respectively, suggesting that the target search modes are not significantly altered by the presence of binding motif-specific DNA shape deformations. For Xis searching the DNA in the presence of Fis-induced deformation at its binding site, the propensities of sliding, hopping, and 3D diffusion are 8, 60, and 32%, respectively. This profile is consistent (sliding 7%, hopping 59%, and 3D diffusion 34%) with the one when Xis scans poly-CG DNA. The results prompt a critical observation in the DNA shape readout mechanism that nonspecific electrostatic interactions may not be the dominant factor in locating the binding site by the searching proteins. Had it been the case, the propensities of various search modes would have been different on DNA featuring shape deformation at the binding site compared to DNA with poly-CG sequences only. Furthermore, the observation raises the question of what, if not search modes, causes the rapid recognition of the target site when it features distinct deformation in DNA shape.

To identify why the recognition of the target DNA site is faster on DNA featuring deformation in shape at the binding motifs, we conduct a comprehensive analysis of the one-dimensional diffusion dynamics exhibited by the searching proteins on DNA. In 1D search mode on DNA, proteins stochastically move between adjacent DNA bases in both forward and backward directions without dissociating from the DNA surface. This mode enables the protein to thoroughly scan the DNA bases along the major groove, simultaneously progressing along the length of the DNA (Figure 3C), thereby enhancing the probability of successfully recognizing the target DNA sequences compared to other search modes. During a 1D diffusion event, protein's movement is generally confined to local searches covering approximately 50 bp around each landing site on the DNA. Our objective is to characterize the 1D diffusion dynamics within 50 bp around the DNA target site (proximal site) and far from it (distal bases). We define a parameter, $p_{\text{target}}$, to track the protein's direction moving toward the target site during each 1D scanning event while scanning over proximal and distal DNA bases separately (Figure 3D). Figure 3E,F presents the results for PU.1 and Xis proteins, respectively, indicating $p_{\text{target}} \sim 0$ on distal DNA bases for both proteins, irrespective of the presence of deformation at the binding site. A $p_{\text{target}}$ close to zero indicates that the proteins stochastically search the DNA by performing back and forth motion on it. Conversely, a $p_{\text{target}} > 0$ denotes that the protein preferentially moves toward the target site while scanning on the DNA, and $p_{\text{target}} < 0$ indicates movement away from the position of the binding motif. Interestingly, our analysis suggests that when the protein scans proximal bases, its diffusion dynamics (1D) on DNA dynamics are significantly biased toward the target site ($p_{\text{target}} > 0$) in the presence of DNA deformation, either caused by intrinsic sequence flexibility (in the case of PU.1) or induced by the binding of the Fis protein (in the case of Xis). This is not the case when the binding motif does not feature any noticeable deformation in shape compared to its flanking DNA segment. The results clearly indicate that in the DNA shape readout mechanism, the alteration of the DNA shape at the binding motif can expedite its recognition by biasing the 1D diffusion of the protein in its proximity, although the overall propensities and target search mechanism remain unaltered. The result is noteworthy, as both cases involve protein diffusion governed by nonspecific electrostatic interactions. However, the observed bias is discernible only when the DNA exhibits deformed shapes at the binding sites. It is noteworthy to highlight that Fis is already bound to its target site and hence does not impede the nonspecific search of Xis protein. However, under in vivo conditions, when starting from an unbound state, both proteins can compete with each other for their target site. Consequently, their crosstalk may influence the dynamics of


https://doi.org/10.1021/jacsau.4c00393
JACS Au 2024, 4, 2640−2655

Table 2. Thermodynamic Parameters Calculated from the Simulation Trajectories of the Fis-DNA Complex and Xis in Free Systems and Fis-DNA and Xis Nonspecific Complex at $T$ = 300 K[a]

| systems | $H$ | $\Delta H$ | $TS_{\infty}$ | $T\Delta S_{\infty}$ |
|---|---|---|---|---|
| Fis−DNA and free Xis | −2146.23 ± 2.18 | | 430.62 ± 0.64 | |
| | | −2.79 ± 2.97 | | 40.83 ± 0.84 |
| Fis−DNA and Xis complex | −2149.02 ± 2.02 | | 471.45 ± 0.55 | |

[a]All the values are in kcal/mol ± standard errors.

their target search, which is, however, beyond the scope of this study.

## Thermodynamics of the DNA Shape Readout Mechanism

In order to elucidate the underlying physical principles behind the swift recognition of binding motifs characterized by altered DNA shapes compared with the ideal B-DNA geometry, we perform a thermodynamic analysis of the simulation trajectory until the protein reaches the target site. We emphasize that the analysis involves only the nonspecific search process of the proteins. The findings of this analysis for PU.1 and Xis proteins are presented in Tables 1 and 2, respectively. The change in free energy with respect to free DNA can be given by

$$\Delta G = \Delta H - T\Delta S \quad (1)$$

where $H$ is the enthalpy of the system (DNA−protein complexes or individual systems), $T$ is the temperature, and $S$ represents the configurational entropy. $\Delta H$ represents the difference between the enthalpy ($H$) of the free systems ($H_{(free)}$) and DNA−protein complexes ($H_{(complex)}$).

In the case of the PU.1 protein, enthalpic contributions are computed based on the equilibrated conformations of both free DNA and free protein and the PU.1-DNA nonspecific complex. The summarized results in Table 1 suggest that reaching the DNA-binding site nonspecifically from an unbound state is marginally enthalpically favorable ($\Delta H$ = −2.70 ± 3.92 kcal/mol). To assess the impact of configurational entropy, we estimate entropy by diagonalizing the Cartesian coordinate covariance matrix, following the methodology outlined by Schlitter.[72] It is worth noting that this approach has been previously employed in protein and DNA systems at both atomistic and coarse-grained representation to estimate configurational entropy.[56,73−77] The magnitude of configurational entropy is however lower in the coarse-grained systems compared to the all-atom description of the system with explicit solvent molecules.[76,77] The computed entropies ($S$) are dependent on the trajectory's length ($t$), indicating a convergence issue. Over a longer simulation time window, the entropy value approaches a limit ($S_{\infty}$). We find that the entropies, when calculated over a sufficiently extended time window, can be effectively fitted using an empirical relationship as described previously,[75] which can be written as

$$S(t) = S_{\infty} - \frac{\alpha}{t^{2/3}} \quad (2)$$

where $\alpha$ is the slope of eq 2. Based on the obtained values (as shown in Table 1), we compute $T\Delta S$ at 300 K to be 36.99 ± 0.20 kcal/mol, suggesting that reaching the target DNA site, which exhibits a deformed shape due to the intrinsic sequence-dependent flexibility of its sequence, is linked to a net gain in configurational entropy. It is noteworthy that the mass and moments of inertia of the binding proteins have a negligible impact on the binding process, while translational and rotational entropies depend on these parameters. Consequently, during our calculations, changes in these entropies are ignored. The result indicates that the entropic contribution is approximately 14 times higher than that of the enthalpic contribution ($T\Delta S \gg \Delta H$), suggesting a predominant role of entropy in the DNA shape readout mechanism compared to the role of electrostatic interactions between the searching protein and DNA. Additionally, it is noteworthy that $T\Delta S$ encompasses the distinct contributions of both the protein and DNA individually. With further dissection of the entropic contribution for poly-CG (+sp)($T\Delta S = T\Delta S_{PU.1} + T\Delta S_{DNA(sp)}$), we find $T\Delta S_{DNA(sp)}$ = 10.15 ± 0.98 kcal/mol and $T\Delta S_{PU.1}$ = 26.84 ± 1.82 kcal/mol. In the absence of sequence-dependent shape deformation at the binding site, such as in the poly-CG sequence, the entropic contributions of the protein and DNA are $T\Delta S_{DNA}$ = 0.72 ± 0.134 kcal/mol and $T\Delta S_{PU.1}$ = 22.49 ± 1.92 kcal/mol, respectively. This suggests that for poly-CG (+sp), the search process is assisted by change in the entropies of both DNA and protein molecules. In general, the high entropy of the protein molecules is attributed to its conformational fluctuations during various nonspecific scanning modes on and off the DNA. Such a wide range of fluctuations are missing in the free state of the protein, resulting in a high gain in protein's entropy.

The thermodynamic analysis of the interaction between Xis and the Fis-DNA complex is given in Table 2. We estimate the changes in enthalpy and entropy resulting from the nonspecific recognition of Xis at its binding site, where the Fis dimer is already bound at the two consecutive major grooves. The recognition of the DNA-binding motif by the Xis protein is primarily entropy-driven, with the entropic contribution $T\Delta S$ = 40.83 ± 0.84 kcal/mol, being approximately 15 times higher than that of the enthalpic contribution ($\Delta H$ = −2.79 ± 2.97 kcal/mol). This observation aligns with the entropy-driven recognition of PU.1 to its binding motif, indicating the generalized role of entropy in facilitating the recognition of the target DNA site by searching proteins in the DNA shape readout mechanism. Further decomposition of entropy ($T\Delta S = T\Delta S_{Pro} + T\Delta S_{DNA}$) reveals that the net entropy change of DNA ($T\Delta S_{DNA}$) is 5.29 ± 0.65 kcal/mol, which, in combination with the net change in configurational entropy of Xis (35.54 ± 2.13 kcal/mol), favors the recognition of the target site.

To further validate our findings regarding the protein's nonspecific search for the target DNA site in the DNA shape readout mechanism, which we claim is primarily an entropy-driven process, we adopt an indirect method of entropy estimation. We investigate the DNA-binding free energy landscape of PU.1 using the umbrella sampling method (details provided in the Supporting Information) and estimate the entropic contribution ($T\Delta S$) from the differences in free energy ($\Delta G$) and enthalpy ($\Delta H$) along the DNA contour (see Figure S8). The binding free energy landscape during the nonspecific target search process, for the poly-CG (+sp) sequence, is depicted in Figure 4A as a function of the distance


https://doi.org/10.1021/jacsau.4c00393
JACS Au 2024, 4, 2640−2655

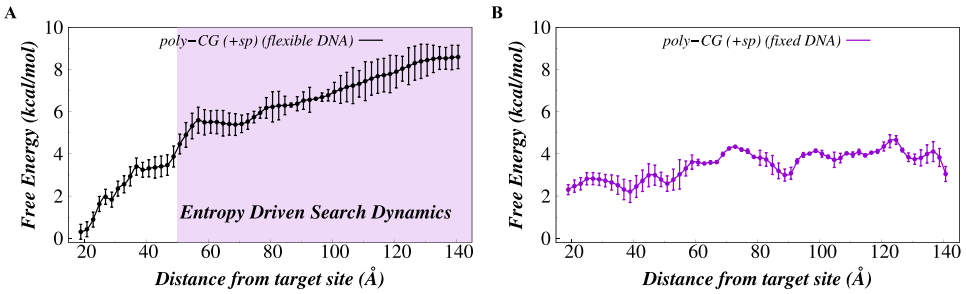



**Figure 4.** Free energy profiles of PU.1-DNA binding with poly-CG (+sp) sequence as a function of the distance between the center of mass of PU.1 and the center of mass of mid base pair of the target site of DNA for the (A) flexible and (B) fixed DNA systems. The error analysis is done by performing three independent umbrella sampling simulations.

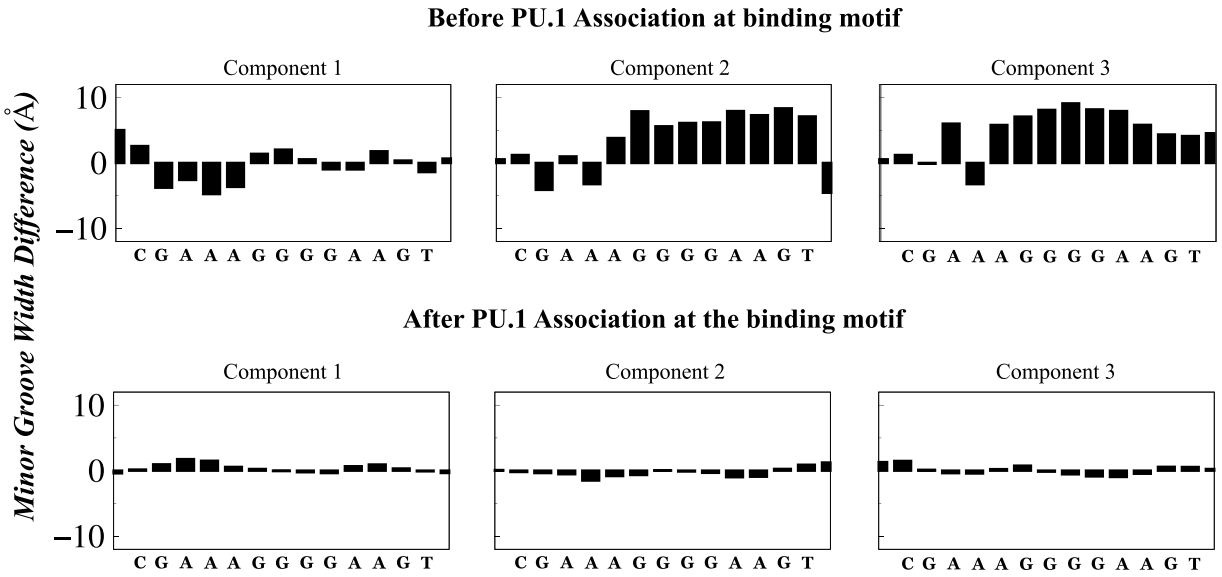



**Figure 5.** Principal component analysis for the PU.1-DNA system. Patterns of the minor groove width variations corresponding to the top 3 eigenvectors of the dynamics of the free DNA and PU.1-DNA complex.

between the searching protein and the mid base pair of DNA-binding site. The results suggest a change of approximately 8.28 ± 1.31 kcal/mol in free energy during the process, with the distance between the searching protein and the DNA-binding site, in terms of base pairs, translating to approximately 5−40 base pairs. It is important to note that a distance of ≤20 Å denotes the span of the target site (∼6 bp), indicating that when the distance is 20 Å, the protein has already reached the target span. The estimated $T\Delta S$ (see Figure S8 for individual contributions of enthalpy and entropy) along the DNA contour illustrates that the nonspecific search is predominantly entropy-driven (shaded region, 140−50 Å) until the protein is extremely close to the binding site (within 5 bps). At such close proximity, the altered protein−DNA electrostatics due to the deformed shape of the target DNA site influence the protein association much more strongly, highlighting that entropy drives the search process of the protein from a large distance far from the target DNA site. To further substantiate our findings, we repeat the umbrella sampling experiment, but this time we freeze the DNA conformation, thereby completely ceasing the entropic contribution of the DNA. The resultant free energy profile is depicted in Figure 4B, featuring a flat pattern indicating that target search by the protein. In fact, results from our kinetic experiments suggest that the nonspecific recognition of the target DNA site is approximately 3 times slower (Figure S9A) compared to when DNA entropy is considered. It is interesting to note that comparative analysis of protein search modes on flexible and fixed DNA reveals significant differences. On flexible DNA, the protein exhibits 32% sliding, 54% hopping, and 14% diffusion, whereas on fixed DNA, the protein search entails 73% sliding, 19% hopping, and 8% diffusion (Figure S9B).

To this end, we note that further clarification is needed regarding the discrepancies in the $T\Delta S$ values obtained from this indirect approach and Schlitter's method. In the latter, we assess the net change in entropy of the nonspecifically associated PU.1-DNA complex compared to the free states of the molecules. Conversely, in the indirect method, where the $T\Delta S$ estimation relies on the free-energy profile of the PU.1-DNA system calculated from umbrella sampling, we utilize the distance between the searching PU.1 and the DNA target site as the reaction coordinate. The furthest distance considered is 140 Å, which roughly corresponds to ∼40 base pairs away from the target DNA site. It is important to note that the reaction coordinate does not constrain the protein to be 140 Å away from the DNA molecule. PU.1 can still be on the DNA, performing 1D scanning, while maintaining a 140 Å distance from the DNA target site. Thus, the initial state is already a nonspecifically bound complex, resulting in a $T\Delta S$ much lower than it could be compared to completely free states of PU.1 and DNA. Consequently, for a more accurate comparison between $T\Delta S$ values obtained from Schlitter's


https://doi.org/10.1021/jacsau.4c00393
JACS Au 2024, 4, 2640−2655

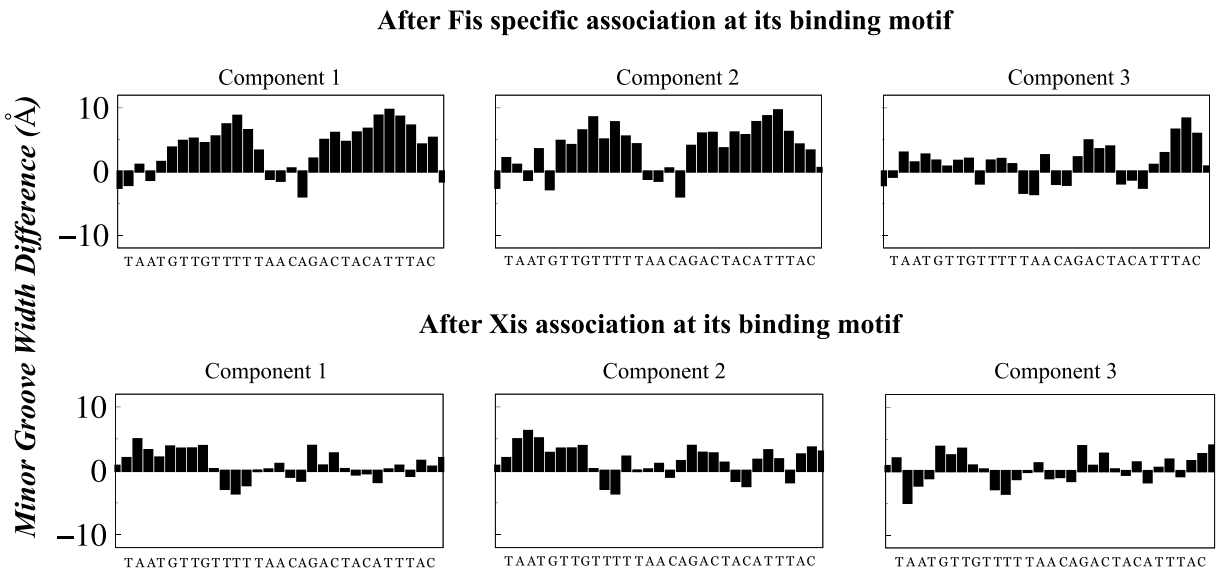



**Figure 6.** Principal component analysis for the Fis-Xis-DNA system. Patterns of the minor groove width variations corresponding to the top 3 eigenvectors of the dynamics of the Fis-DNA and Fis-Xis-DNA complexes.

method and the indirect approach, a more appropriate reaction coordinate should be considered. However, this aspect is currently beyond the scope of the present study. Nevertheless, our results obtained from two completely different approaches confirm that the nonspecific recognition of the target site on DNA in the DNA shape readout mechanism is primarily an entropy-driven process.

## Differences in DNA Configurational Entropy at the Binding Site and Distal Nonspecific Sites Induces a Funnel-like Information Transfer Landscape

To comprehend the molecular intricacies of the entropy-assisted DNA shape readout mechanism, we systematically investigate the roles played by entropic contributions from both the DNA and the searching protein. As a semiflexible polymer, the DNA molecule undergoes fluctuations around a B-DNA geometry, sampling various microstates. Any sequence-specific alteration in shape or deformation induced by the specific binding of proteins at adjacent sites significantly skews the conformational ensemble toward the microstates with constraints in DNA shape, thereby reducing the configurational entropy of the molecule. One effective way to estimate this impact is to monitor the fluctuations in minor groove width, which are highly sensitive to changes in the shape and conformational dynamics of DNA.[67,78] However, doing so is nontrivial considering the ensemble of DNA configurations with varying degrees of change in the minor groove geometry. To address this challenge, principal component analysis (PCA)[79,80] offers a method to transform the large set of variable DNA groove geometries into a smaller set that still contains most of the information in the original set. The eigenvectors obtained from PCA provide a vectorial representation of each mode of structural deformation, and the eigenvalues indicate the magnitudes of these modes of structural deformations. We performed PCA on the trajectories of both free DNA and protein−DNA complexes (PU.1-DNA, Fis-DNA, and Fis-Xis-DNA). The details of the PCA calculation and methodology are provided in the Supporting Information. Figures 5 and 6 illustrate the variations in minor groove width associated with the free DNA and protein−DNA complexes. For the top three eigenvectors, structures of DNA are generated corresponding to the minimum and maximum observed eigenvalues, as previously described.[56,75] The plotted minor groove width difference between these structures at the binding motif, as shown in Figure 5, exhibits a pattern reminiscent of the modes of vibration of a string, highly asymmetric and harmonic before protein binding. This offers a compelling insight into the conformational dynamics of the DNA molecule, where the information related to DNA shape deformation traverses between sites. Upon PU.1 reaching its cognate motif on the DNA, lower harmonics in vibration are observed. In the case of the Fis-Xis-DNA system, the differences in groove width associated with the principal eigenvectors are depicted in Figure 6. In this scenario, the modes of vibration are highly asymmetric, and higher harmonics predominate when only Fis is associated with its binding site to cause significant deformation in DNA shape in the adjacent sites, including the binding site of Xis. This is largely restored after Xis reaches its binding site nonspecifically. The results capture distinctly different dynamics of the DNA molecule at the binding motifs in the presence and absence of a nonspecifically bound protein. The associated signal is captured in the form of differences in the minor groove widths.

To further understand how the signal originated due to the deformation in shape of the binding motif influences the search process of a protein positioned at a distance from the binding motif, we probe if the signal traverses along the DNA contour. This is done by estimating the information transfer landscape of the free systems, PU.1 and DNA, and the PU.1-DNA complex using the time-delayed conditional probabilities of time-series data of minor groove width fluctuations, following Schreiber's formulation of information entropy transfer.[81] Transfer entropy serves as a statistical measure of the directed transfer of information between the two processes. The detailed method is outlined in the Supporting Information. This approach has been tested for various protein−protein and protein−DNA systems.[56,82] The advantage of this method lies in its ability to pinpoint the entropy source and entropy sink resulting from DNA shape deformation induced by protein association or the sequence-specific intrinsic flexibility of the


https://doi.org/10.1021/jacsau.4c00393
JACS Au 2024, 4, 2640−2655

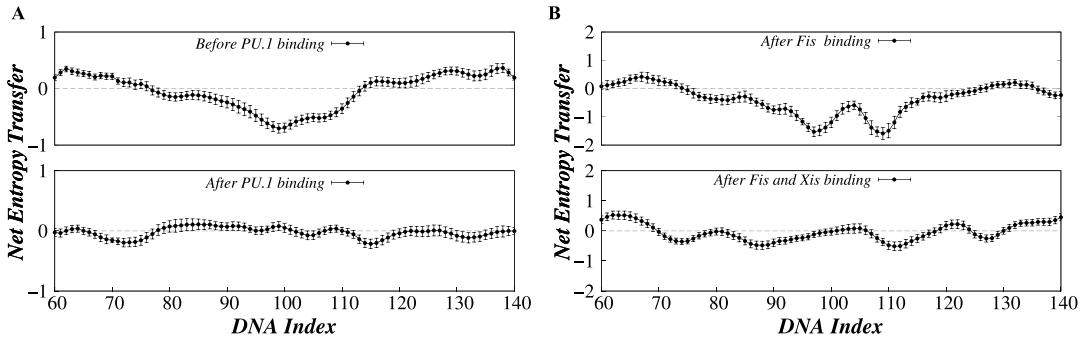



**Figure 7.** Entropy transfer. Net entropy transfer from individual bases to the entire set of DNA bases is shown for (A) poly-CG (+sp) DNA (for PU.1 protein) and (B) poly-CG (+sp, + Fis) DNA (for Xis protein). The upper and lower panel in both (A and B) represents the net entropy transfer before and after the nonspecific association of the proteins (PU.1 and Xis) on their target site, respectively. DNA indexes with a negative value of net entropy transfer are entropy sinks, whereas those with a positive value acts as the entropy sources. The error bars associated with each DNA index are defined as the standard error.

binding motif. This approach elucidates how information regarding protein binding is communicated along the DNA contour.[82−85]

The results are depicted in Figure 7, where the net information entropy transfer is plotted relative to the DNA base pair index. Figure 7A illustrates the net entropy transfer landscape along the DNA molecule in the absence of PU.1, in the apo state (poly-CG (+sp), upper panel), and when PU.1 is nonspecifically bound to its binding motif on DNA (lower panel), respectively. Likewise, Figure 7B elucidates the net entropy transfer landscape along the DNA molecule after Fis-induced shape deformation (upper panel) and when Xis reaches its target DNA site nonspecifically (lower panel), respectively. The patterns distinctly show that the shape deformation resulting either from the intrinsic flexibility of the binding motif in case of PU.1 or due to specific binding of Fis at the adjacent site of Xis protein's binding motif leads to the formation of an entropic funnel. The net entropy $T^{NET}(i \rightarrow j)$ flows toward the target site positioned at the middle of the DNA stretch from both directions. This observation is further supported by the absence of such an entropic funnel when the DNA does not feature the binding motif for PU.1 (Figure S10A) or Fis is not present to induce shape deformation in DNA (Figure S10B), indicating the role of the shape of DNA-binding sites biasing the sampled conformational space for the DNA molecule.

The observed information propagation is consistent with previous experimental and theoretical investigations that have elucidated the role of DNA allostery in facilitating the binding of a second ligand subsequent to the initial binding of the first ligand at a distant site.[86−88] In these investigations, the ligands' binding sites are separated by a variable-length spacer of DNA segments. Kim *et al.* emphasized that while electrostatic interactions between the ligands are not the primary origin of the allosteric phenomenon, allostery is predominantly influenced by the mechanical characteristics of the linker DNA.[86] They postulated that allostery through DNA mainly arises from distortion of the groove width induced by protein binding. These studies have proposed that DNA allostery engenders an oscillatory phenomenon with periodicity, facilitating the binding of two or more proteins at distinct sites through a signal conveyed by the linker DNA. These investigations have highlighted that DNA-mediated allostery and signal transmission rely on the mechanical distortions of DNA over distances. However, investigating the persistence of signal transfer induced by DNA's mechanical properties in the absence of an associating partner falls beyond the scope of these studies. Our findings provide a quantitative approach to delineate the signal induced by DNA distortion, highlighting the importance of indirect readout mechanism in the target search dynamics of the searching protein.

The width of the funnel encompasses several adjacent flanking sequences, indicating that the impact of the funnel is nonlocal due to the elastic nature of the DNA molecule. As the binding motif assumes a deformed shape due to its intrinsic flexibility in the specific sequence, the adjacent flanking DNA bases are compelled to adopt a distinct but less pronounced deformation in shape to maintain stacking and cross-stacking interactions of the DNA bases. The impact diminishes with an increasing distance from the binding motif. However, the landscape undergoes a complete transformation as PU.1 nonspecifically reaches its binding motif (lower panel, Figure 7A). The entropic funnel represents a flat landscape, indicating a return of the skewed distribution of the conformational space of the DNA molecule to normalcy. The associated entropic gain ($10.15 \pm 0.98$ kcal/mol) favors the protein search toward its DNA-binding motif. Similarly, the net entropy transfer ceases to zero (lower panel, Figure 7B) when Xis nonspecifically reaches its binding motif. The associated entropic gain is found to be $5.29 \pm 0.65$ kcal/mol. Additionally, we investigate the entropy transfer mechanism for the Fis-Xis-DNA system in the presence of a mixed flanking sequence, and the result is presented in Figure S11. The result indicates that the influence of DNA deformation, resulting from the specific binding of Fis, skews the sampled conformational space of DNA even for a scrambled sequence of the flanking DNA bases.

To further understand the role of the entropy transfer mechanism in the target search dynamics of PU.1, we monitored how the geometry of the entropic funnel altered with the change in the position of PU.1 on DNA. Figure S12 (and Video S1) presents three instances that vividly portray a dynamic information transfer landscape of the DNA, where the change in funnel width strongly correlates (see Figure S13) with the position of the searching PU.1 protein. The nonspecific binding of the protein at a DNA site within the funnel releases the shape constraints of the adjacent DNA bases at the point of the protein's probe. The effect shifts the skewed distribution of DNA conformations to regular fluctuations around a B-DNA geometry, causing an increase in the entropy of the DNA bases. Here, B-DNA refers to the


https://doi.org/10.1021/jacsau.4c00393
JACS Au 2024, 4, 2640−2655

flanking poly-CG segments around the specific target site. To support our assertion, we examine two scenarios: one where the protein is positioned on the DNA approximately 30 bps away from the target site and the other where the protein is merely 10 bps away from the target site (refer to Figure S14, upper panel). We analyze the minor groove width distributions of the flanking poly-CG segment between base pairs of 65−80 in both cases. Our findings, depicted in the lower-left panel of Figure S14, distinctly indicate that when the protein is distant from the target DNA site, the inherent deformation induced by the sequence of the target DNA site causes a significant alteration in the minor groove width distribution in the 65−80 bps segment of the poly-CG DNA. The distribution noticeably skews toward a lower minor groove width. However, when the protein is in close proximity to the target DNA site (approximately 10 bp away), it obstructs the progression of DNA deformation at the target site along the DNA contour beyond the protein. Consequently, the skewed distribution of the minor groove width within the 65−80 bps segment reverts back to that of the poly-CG B-DNA conformation. Considering the fluctuations in minor groove width as an indicator of the corresponding change in entropy, we observe higher entropy (greater fluctuations) when the skewed distribution in the minor groove width is normalized (refer to the lower-right panel of Figure S14). The entropy of the DNA bases around the binding motif of the searching protein thus acts as a driving force in directing the nonspecific search of the proteins toward their binding sites. As the protein approaches, the funnel geometry changes accordingly, explaining the directionality of the target search method of proteins in the DNA shape readout mechanism. The effect of protein−DNA electrostatic interaction is evidently highly localized and helps in neutralizing the entropy gradient at the site of the probe on DNA. In contrast, the entropic funnel centering the DNA-binding motif exerts a long-range driving force and serves as an entropic switch that regulates the 1D diffusion of a protein searching DNA bases in the vicinity of the binding site. Upon reaching the specific site, the switch is turned off.

In this context, it is important to note that a recent study has emphasized the influence of DNA mechanical properties on the binding specificity of PU.1.[59] The study revealed that the local bending stiffness of DNA is elevated when PU.1 binds at a nonspecific site, while it becomes highly flexible at the binding motif, facilitating the locking of PU.1 at this site. Subsequently, they hypothesized the role of DNA configurational entropy in determining the bending stiffness of specific and nonspecific sites. In this work, we explicitly demonstrate the role of DNA configurational entropy in regulating the recognition of the DNA-binding motif through a shape readout mechanism. Taking this a step further, we employ their theoretical model to calculate DNA bending stiffness upon PU.1 association with DNA segments to describe how semiflexible DNA responds to the association of DNA-binding proteins. Figure S15 presents the relative DNA bending stiffness as a function of the DNA index for the PU.1-DNA system. The results indicate that the bending stiffness is at its minimum when the protein is at its binding motif and increases when the protein is at the flanking region. This analysis underscores that DNA bending stiffness is not uniform at every nonspecific site, rather it displays a funnel-like pattern, much like our entropic funnel representation. The impact of the DNA-binding motif, and thus DNA shape (deformation), extends beyond the motif itself, propagating to adjacent flanking DNA bases. This impact diminishes and saturates with an increase in distance from the binding motif, as explained above based on the entropy transfer mechanism.

## Role of the Searching Protein in the DNA Shape Readout Mechanism—Origin of Specificity

While the configurational entropy of the DNA explains a rapid nonspecific recognition of the binding motif by the searching protein, what promotes specificity in their binding? Numerous studies have aimed to biophysically characterize how proteins undergo facilitated diffusion while searching for target DNA sites.[15,57,89−95] The kinetics of protein−DNA recognition interactions was found to not solely depend on search speed; rather, it also depends on how efficiently the protein establishes specific interactions upon reaching its binding motif. In the search mode (S state), a protein interacts nonspecifically with DNA through electrostatic interactions between positively charged protein residues and negatively charged phosphates on the DNA backbone.[96,97] Upon reaching the target site, the protein switches its conformation to the recognition mode (R state) to establish sequence-specific contacts with the DNA, involving hydrogen bonds, hydrophobic, and aromatic interactions between the protein residues and the DNA bases.[98] The complete search kinetics, therefore, involve transitions between the nonspecific binding mode, S, and the specific binding mode, R.[99−103] The existence of S and R binding modes is argued to be necessary to reconcile the conflict between speed and stability, commonly referred to as the speed-stability paradox, where conditions for fast search are incompatible with stable protein−DNA interactions.[2,100] Various experimental techniques, such as X-ray crystallography, NMR, and single-molecule approaches, support the two-state model. These methods reveal that DNA-binding proteins assume distinct conformations when engaged in specific interactions with DNA compared to nonspecific interactions.[3,57,100,104−107] Here, our objective is to comprehend how the entropic funnel, induced by DNA shape deformation at the binding site, modulates the transition between the S and R modes.

We identify the recognition modes for both PU.1 and Xis by analyzing their experimentally resolved specific complexes with DNA. Only amino acids within 4 Å of a nucleotide base, as determined by their side chains, are considered to be part of the R mode. Any other positively charged amino acids through which the protein nonspecifically probes the DNA, are categorized as part of the S mode. The overlap between these two states can be quantified using the protocol outlined by Leven *et al.*,[108] where a parameter $\chi_i$ is defined for each protein residue forming specific contacts with DNA in the complex, as given below

$$\chi_i = \frac{\sum_j q_j \exp\left(-a\frac{r_{ij}}{r_c}\right)}{\sum_j \exp\left(-a\frac{r_{ij}}{r_c}\right)} \tag{3}$$

Here, $j$ represents protein residues within a cutoff distance of $r_c$ = 8 Å from residue $i$. $r_{ij}$ is the distance between residues $i$ and $j$; $q_j$ denotes the point charge on residue $j$, and a = 5 is an exponential factor.[108] The overall similarity index $\chi$ is obtained by averaging over all $\chi_i$, ranging from −1 to +1. A higher $\chi$ value indicates a low frustration and greater similarity between the S and R states. Consequently, the free energy difference between the two states is low. Alternatively, a low $\chi$ value


https://doi.org/10.1021/jacsau.4c00393
JACS Au 2024, 4, 2640−2655

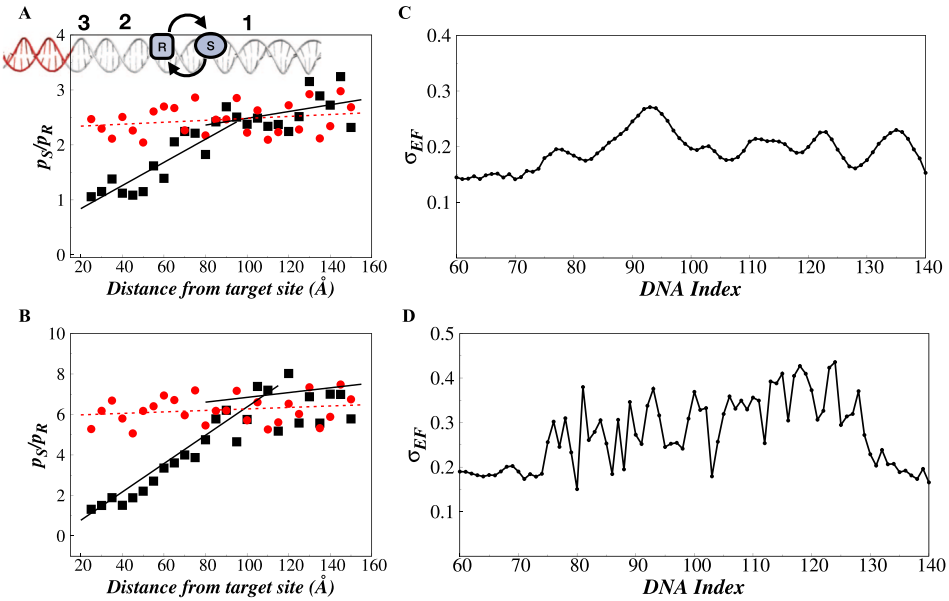



**Figure 8.** Characterization of protein dynamics. The ratio of probabilities of the proteins being in search ($p_S$) and recognition ($p_R$) modes of (A) PU.1 protein on poly-CG (red color) and poly-CG (+sp) DNA (black color) and (B) Xis protein on poly-CG (red color) and poly-CG (+sp, + Fis) DNA (black color) with the position of protein on DNA. The schematics given in the inset of (A) shows the different position of protein. Red color patch denotes the target site on DNA (shown in white color). Positions 1, 2, and 3 represent the protein position from the funnel with a distant site, near the funnel, and inside the funnel, respectively. The ruggedness of the transfer entropy landscape ($\sigma_{EF}$) of (C) poly-CG (+sp) DNA (PU.1 protein) and (D) poly-CG (+sp, + Fis) DNA (Xis protein).

signifies high molecular frustration in the protein, where two states are distinctly different from each other and are separated by a relatively higher free energy difference. Our analysis suggests that $\chi$ values for PU.1 and Xis proteins are 0.19 and 0.08, respectively. A higher $\chi$ value for PU.1 suggests a greater overlap between the two states. In comparison, a somewhat lower $\chi$ value in Xis represents distinct S and R modes. We analyze the 1D search modes of the proteins along the DNA until they reach their binding motifs, estimating the probabilities of the proteins being in search ($p_S$) and recognition ($p_R$) modes.

Variations of $p_S/p_R$ with varying protein position on DNA are shown in Figure 8A and B for PU.1 and Xis proteins, respectively. Evidently, $p_S/p_R$ for PU.1 is relatively smaller when the protein is far from the binding motif than that of Xis, indicating greater overlap between the search and recognition modes in PU.1, consistent with their respective $\chi$ values. The results further suggest relatively constant $p_S/p_R$ values for both PU.1 and Xis proteins, irrespective of their positions on DNA if there is no deformation in the shape at the binding motif (see Figure 8A,B, red circles). Conversely, in the presence of sequence-dependent shape deformation in the case of PU.1 and Fis-induced deformation at the binding motif of Xis, $p_S/p_R$ exhibits a drastic reduction (see Figure 8A,B, black squares) as the proteins enter into the entropic funnel and approach closer toward their specific binding sites. When at the close vicinity of the target site, $p_S/p_R$ reduces to ~1. The results signify that the proteins predominantly scan the DNA in their search mode while outside the entropic funnel (see position 1). On the contrary, inside the funnel (see positions 2 and 3), transitions between search and recognition mode are easier, allowing the protein to spend almost equal time in the recognition mode, thereby significantly enhancing the probability of establishing specific contacts with the binding motif.

Having observed that the entropic funnel on the DNA, originating from the shape deformation of the specific binding sites, promotes proteins to be in recognition mode in the vicinity of the binding motif, we further probe how the same is connected with the molecular frustrations $\chi$ of the respective proteins by assessing the ruggedness of the funnel ($\sigma_{EF}$) from fluctuations in the $T^{NET}$ values. The results presented in Figure 8C,D represent the ruggedness of the entropic funnel for PU.1 and Xis proteins, respectively. Notably, the PU.1 protein, which exhibits a higher similarity between the search and recognition modes (high $\chi$ value), displays a less rugged landscape (Figure 8C). The less rugged entropic funnel fits the need of the protein to minimally fluctuate to switch between its overlapped search and recognition modes. In contrast, for a protein like Xis, which features distinctly different search and recognition modes (low $\chi$ value), a more rugged funnel landscape (see Figure 8D) befits the larger conformational fluctuation required to switch between the search and recognition modes. This observation reveals the fact that a complementary relationship might exist between the molecular features of searching proteins (similarity between search and recognition modes) and the ruggedness on the DNA track around the binding motifs, which determines the overall specificity of their binding. To validate our hypothesis, we simulate the target search dynamics of Xis in the absence of the Fis protein, where the AT-rich binding motif is positioned in the middle of poly-CG sequences. Without specifically bound Fis, the binding motif displays less deformation in shape, and corresponding fluctuations in the entropic funnel are substantially lower (see Figure S16). The variation in $p_S/p_R$ for Xis under this condition shows lower efficiency in specifically recognizing the binding site (see Figure S17), indicating a less conducive environment for forming a specific complex (lower specificity) as Xis scans the binding motif primarily in its search mode.



https://doi.org/10.1021/jacsau.4c00393
JACS Au 2024, 4, 2640−2655

## CONCLUSIONS

In this study, we explored the origin of specificity in the DNA shape readout mechanism employed by proteins to recognize DNA-binding motifs. Changes in DNA shape mainly arise from the inherent sequence-specific flexibility of the binding motifs or the binding of other proteins to nearby DNA sites. In both scenarios, we demonstrate that deformation of the DNA shape at the binding motif creates an entropic funnel around it. Through an analysis of the time-delayed conditional probability of minor groove width fluctuations, we observed a net entropy flow from the adjacent site toward the binding motif, designating it as an entropy sink. As the protein scans DNA in the proximity of the binding motif, it senses the entropy gradient along the DNA contour. The protein's movement toward the binding motif relaxes the shape constraints of the DNA bases, increasing the DNA entropy, which biases the protein's 1D search dynamics toward its binding site. Such biases are absent when the protein is distant from the binding site, or the binding motif does not feature any distinct deformation compared to the flanking DNA bases. The entropy-driven biased 1D scanning of the protein near its target site thus facilitates the overall recognition process and enhances the associated kinetics. The funnel's impact is significantly long-ranged in transferring information related to alteration in shape of the binding motif and encompasses several flanking nucleotide bases. In contrast, the role of electrostatics is weaker, aiding in releasing the deformed shape-induced strain of DNA bases locally around the protein probe on DNA.

The influence of an entropic funnel resulting from DNA shape deformation is generic and, in principle, could enhance the search dynamics of any adjacent proteins. However, specificity in protein−DNA recognition necessitates compatibility between the molecular characteristics of the protein and DNA. Our study demonstrates that indeed a correspondence exists between the molecular frustration of the protein molecule and the ruggedness of the entropic landscape of DNA. It is well-established that the target search process of a protein involves search and recognition modes. The molecular frustration of a protein depends on the similarity of these two modes. If both modes overlap, then minimal fluctuation in the protein conformation is sufficient for the transition between states. In contrast, a protein has higher frustration if its search and recognition modes are distinctly different, requiring larger conformational fluctuations to switch between states. Our analysis suggests that the ruggedness of the entropic landscape of DNA is well-suited to meet this requirement; higher molecular frustration in the searching protein corresponds to greater fluctuation in the entropic funnel, facilitating the conformational transition of the protein between its two states. The associated change in the protein's entropy favors overall dynamics as it progresses toward the specific target site, increasing the probability of scanning the DNA through its recognition mode. This mechanism differs significantly outside the funnel, where the protein scans the DNA nonspecifically, mostly through its search mode. We propose this as a generic mechanism for the DNA shape readout. The discussion remains incomplete without addressing the significance of the entropic effect in facilitating protein search on DNA and maintaining specificity in recognizing the binding motif through the DNA shape readout mechanism, especially given the relatively vast genomic background the protein scans before binding specifically. In this context, we emphasize that a genome-wide analysis of transcription factor binding sequences reveals a distinct pattern of AT-rich nucleotide bases around the binding motif.[109] This pattern gradually fades within approximately 200 bp on each side of the binding motif, transitioning to a more random DNA sequence. Considering the propensity of AT-rich sequences to induce deformation in DNA shape, we hypothesize that a much wider entropic funnel exists in reality, exerting a more dominant influence on the DNA shape readout mechanism than is currently anticipated using poly-CG flanking sequences around the binding motif. Typically, TF binding sites are AT-rich,[109] capable of inducing sequence-dependent curvature.[110,111] However, there may be cases in which the bending is minimal. For such cases, the target search process can still be facilitated through DNA-mediated allostery, particularly upon the association of other proteins at slightly distant sites.[112] In fact, this mechanism is versatile and adaptable for many transcription regulators.[113] This testable hypothesis may provide essential insights into the role of flanking noncoding DNA sequences in regulating gene expression.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacsau.4c00393.

Additional description of the simulation methods, data analysis, and validation of the DNA model (PDF)

Net entropy transfer (MOV)

## AUTHOR INFORMATION


### Corresponding Author

**Arnab Bhattacherjee** − *School of Computational & Integrative Sciences, Jawaharlal Nehru University, New Delhi 110067, India;* orcid.org/0000-0002-7714-2619; Email: arnab@jnu.ac.in

### Authors

**Sangeeta** − *School of Computational & Integrative Sciences, Jawaharlal Nehru University, New Delhi 110067, India*

**Sujeet Kumar Mishra** − *School of Computational & Integrative Sciences, Jawaharlal Nehru University, New Delhi 110067, India*

Complete contact information is available at: https://pubs.acs.org/10.1021/jacsau.4c00393


### Author Contributions

CRediT: * **Sangeeta** conceptualization, data curation, formal analysis, investigation, methodology, validation, writing-original draft; **Sujeet Kumar Mishra** conceptualization, data curation, formal analysis, investigation, methodology, validation, writing-original draft; **Arnab Bhattacherjee** conceptualization, data curation, formal analysis, investigation, methodology, project administration, supervision, validation, writing-original draft.

### Notes

The authors declare no competing financial interest.



https://doi.org/10.1021/jacsau.4c00393
JACS Au 2024, 4, 2640−2655

## ■ ACKNOWLEDGMENTS

We gratefully acknowledge the financial support from the DST SERB India research grant (CRG/2023/000636), DBT India (BT/PR46247/BID/7/1015/2023) and DBT BIC research grant. A.B. gratefully acknowledges support from the Alexander von Humboldt Foundation, Germany. Sangeeta acknowledges the financial support from the Ministry of Human Resource Development (MHRD), Govt. of India for the Prime Minister's Research Fellowship (PMRF-192002-617).

## ■ REFERENCES

(1) von Hippel, P. H.; Berg, O. G. Facilitated target location in biological systems. *J. Biol. Chem.* **1989**, *264*, 675−678.
(2) Slutsky, M.; Mirny, L. A. Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential. *Biophys. J.* **2004**, *87*, 4021−4035.
(3) Hammar, P.; Leroy, P.; Mahmutovic, A.; Marklund, E. G.; Berg, O. G.; Elf, J. The lac repressor displays facilitated diffusion in living cells. *Science* **2012**, *336*, 1595−1598.
(4) Tempestini, A.; Monico, C.; Gardini, L.; Vanzi, F.; Pavone, F. S.; Capitanio, M. Sliding of a single lac repressor protein along DNA is tuned by DNA sequence and molecular switching. *Nucleic Acids Res.* **2018**, *46*, 5001−5011.
(5) Lomholt, M. A.; van den Broek, B.; Kalisch, S. M.; Wuite, G. J.; Metzler, R. Facilitated diffusion with DNA coiling. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 8204−8208.
(6) Doucleff, M.; Clore, G. M. Global jumping and domain-specific intersegment transfer between DNA cognate sites of the multidomain transcription factor Oct-1. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 13871−13876.
(7) Marius Clore, G. Exploring translocation of proteins on DNA by NMR. *J. Biomol. NMR* **2011**, *51*, 209−219.
(8) Krepel, D.; Levy, Y. Intersegmental transfer of proteins between DNA regions in the presence of crowding. *Phys. Chem. Chem. Phys.* **2017**, *19*, 30562−30569.
(9) Vuzman, D.; Polonsky, M.; Levy, Y. Facilitated DNA search by multidomain transcription factors: cross talk via a flexible linker. *Biophys. J.* **2010**, *99*, 1202−1211.
(10) Vuzman, D.; Azia, A.; Levy, Y. Searching DNA via a "Monkey Bar" mechanism: the significance of disordered tails. *J. Mol. Biol.* **2010**, *396*, 674−684.
(11) Castellanos, M.; Mothi, N.; Muñoz, V. Eukaryotic transcription factors can track and control their target genes using DNA antennas. *Nat. Commun.* **2020**, *11*, 540.
(12) Dickerson, R. E. The DNA Helix and How It Is Read. *Sci. Am.* **1983**, *249*, 94−111.
(13) Sarai, A.; Kono, H. Protein-DNA recognition patterns and predictions. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 379−398.
(14) Rohs, R.; Jin, X.; West, S. M.; Joshi, R.; Honig, B.; Mann, R. S. Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.* **2010**, *79*, 233−269.
(15) Halford, S. E. An end to 40 years of mistakes in DNA-protein association kinetics? *Biochem. Soc. Trans.* **2009**, *37*, 343−348.
(16) Fujii, S.; Kono, H.; Takenaka, S.; Go, N.; Sarai, A. Sequence-dependent DNA deformability studied using molecular dynamics simulations. *Nucleic Acids Res.* **2007**, *35*, 6063−6074.
(17) Kim, J. L.; Nikolov, D. B.; Burley, S. K. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **1993**, *365*, 520−527.
(18) Joshi, R.; Passner, J. M.; Rohs, R.; Jain, R.; Sosinsky, A.; Crickmore, M. A.; Jacob, V.; Aggarwal, A. K.; Honig, B.; Mann, R. S. Functional Specificity of a Hox Protein Mediated by the Recognition of Minor Groove Structure. *Cell* **2007**, *131*, 530−543.
(19) Gordân, R.; Shen, N.; Dror, I.; Zhou, T.; Horton, J.; Rohs, R.; Bulyk, M. L. Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. *Cell Rep.* **2013**, *3*, 1093−1104.
(20) West, S. M.; Rohs, R.; Mann, R. S.; Honig, B. Electrostatic interactions between arginines and the minor groove in the nucleosome. *J. Biomol. Struct. Dyn.* **2010**, *27*, 861−866.
(21) Stella, S.; Cascio, D.; Johnson, R. C. The shape of the DNA minor groove directs binding by the DNA-bending protein Fis. *Genes Dev.* **2010**, *24*, 814−826.
(22) Chang, Y. P.; Xu, M.; Machado, A. C. D.; Yu, X. J.; Rohs, R.; Chen, X. S. Mechanism of origin DNA recognition and assembly of an initiator-helicase complex by SV40 large tumor antigen. *Cell Rep.* **2013**, *3*, 1117−1127.
(23) Lazarovici, A.; Zhou, T.; Shafer, A.; Dantas Machado, A. C.; Riley, T. R.; Sandstrom, R.; Sabo, P. J.; Lu, Y.; Rohs, R.; Stamatoyannopoulos, J. A.; Bussemaker, H. J. Probing DNA shape and methylation state on a genomic scale with DNase I. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 6376−6381.
(24) Zeiske, T.; Baburajendran, N.; Kaczynska, A.; Brasch, J.; Palmer, A. G.; Shapiro, L.; Honig, B.; Mann, R. S. Intrinsic DNA Shape Accounts for Affinity Differences between Hox-Cofactor Binding Sites. *Cell Rep.* **2018**, *24*, 2221−2230.
(25) Luscombe, N. M.; Laskowski, R. A.; Thornton, J. M. Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res.* **2001**, *29*, 2860−2874.
(26) Seeman, N. C.; Rosenberg, J. M.; Rich, A. Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 804−808.
(27) Otwinowski, Z.; Schevitz, R. W.; Zhang, R. G.; Lawson, C. L.; Joachimiak, A.; Marmorstein, R. Q.; Luisi, B. F.; Sigler, P. B. Crystal structure of trp represser/operator complex at atomic resolution. *Nature* **1988**, *335*, 321−329.
(28) Bloomfield, V. A. DNA condensation by multivalent cations. *Biopolymers* **1997**, *44*, 269−282.
(29) Müller, B. C.; Raphael, A. L.; Barton, J. K. Evidence for altered DNA conformations in the simian virus 40 genome: site-specific DNA cleavage by the chiral complex lambda-tris(4,7-diphenyl-1,10-phenanthroline)cobalt(III). *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 1764−1768.
(30) Bhattacherjee, A.; Levy, Y. Search by proteins for their DNA target site: 2. The effect of DNA conformation on the dynamics of multidomain proteins. *Nucleic Acids Res.* **2014**, *42*, 12415−12424.
(31) Bhattacherjee, A.; Levy, Y. Search by proteins for their DNA target site: 1. The effect of DNA conformation on protein sliding. *Nucleic Acids Res.* **2014**, *42*, 12404−12414.
(32) Dey, P.; Bhattacherjee, A. Role of Macromolecular Crowding on the Intracellular Diffusion of DNA Binding Proteins. *Sci. Rep.* **2018**, *8*, 844.
(33) Dey, P.; Bhattacherjee, A. Mechanism of Facilitated Diffusion of DNA Repair Proteins in Crowded Environment: Case Study with Human Uracil DNA Glycosylase. *J. Phys. Chem. B* **2019**, *123*, 10354−10364.
(34) Mondal, A.; Bhattacherjee, A. Searching target sites on DNA by proteins: Role of DNA dynamics under confinement. *Nucleic Acids Res.* **2015**, *43*, 9176−9186.
(35) Mondal, A.; Bhattacherjee, A. Mechanism of Dynamic Binding of Replication Protein A to ssDNA. *J. Chem. Inf. Model.* **2020**, *60*, 5057−5069.
(36) Mondal, A.; Sangeeta; Bhattacherjee, A. Torsional behaviour of supercoiled DNA regulates recognition of architectural protein Fis on minicircle DNA. *Nucleic Acids Res.* **2022**, *50*, 6671−6686.
(37) Sangeeta; Bhattacherjee, A. Interdomain dynamics in human Replication Protein A regulates kinetics and thermodynamics of its binding to ssDNA. *PLoS One* **2023**, *18*, No. e0278396.
(38) Tan, C.; Terakawa, T.; Takada, S. Dynamic Coupling among Protein Binding, Sliding, and DNA Bending Revealed by Molecular Dynamics. *J. Am. Chem. Soc.* **2016**, *138*, 8512−8522.
(39) Kodandapani, R.; Pio, F.; Ni, C. Z.; Piccialli, G.; Klemsz, M.; McKercher, S.; Maki, R. A.; Ely, K. R. A new pattern for helix−turn−helix recognition revealed by the PU.l ETS−domain−DNA complex. *Nature* **1996**, *380*, 456−460.


https://doi.org/10.1021/jacsau.4c00393
*JACS Au* 2024, 4, 2640−2655

(40) Turkistany, S. A.; DeKoter, R. P. The transcription factor PU.1 is a critical regulator of cellular communication in the immune system. *Arch. Immunol. Ther. Exp.* **2011**, *59*, 431−440.
(41) Hancock, S. P.; Cascio, D.; Johnson, R. C. Cooperative DNA binding by proteins through DNA shape complementarity. *Nucleic Acids Res.* **2019**, *47*, 8874−8887.
(42) Finkel, S. E.; Johnson, R. C. The Fis protein: it's not just for DNA inversion anymore. *Mol. Microbiol.* **1992**, *6*, 3257−3265.
(43) Levy, Y.; Onuchic, J. N.; Wolynes, P. G. Fly-Casting in Protein-DNA Binding: Frustration between Protein Folding and Electrostatics Facilitates Target Recognition. *J. Am. Chem. Soc.* **2007**, *129*, 738−739.
(44) Bhattacherjee, A.; Krepel, D.; Levy, Y. Coarse-grained models for studying protein diffusion along DNA. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2016**, *6*, 515−531.
(45) Zhang, B.; Zheng, W.; Papoian, G. A.; Wolynes, P. G. Exploring the Free Energy Landscape of Nucleosomes. *J. Am. Chem. Soc.* **2016**, *138*, 8126−8133.
(46) Kapoor, U.; Kim, Y. C.; Mittal, J. Coarse-Grained Models to Study Protein-DNA Interactions and Liquid-Liquid Phase Separation. *J. Chem. Theory Comput.* **2023**, *20*, 1717−1731.
(47) Ingólfsson, H. I.; Rizuan, A.; Liu, X.; Mohanty, P.; Souza, P. C. T.; Marrink, S. J.; Bowers, M. T.; Mittal, J.; Berry, J. Multiscale simulations reveal TDP-43 molecular-level interactions driving condensation. *Biophys. J.* **2023**, *122*, 4370−4381.
(48) Rekhi, S.; Sundaravadivelu Devarajan, D.; Howard, M. P.; Kim, Y. C.; Nikoubashman, A.; Mittal, J. Role of Strong Localized vs Weak Distributed Interactions in Disordered Protein Phase Separation. *J. Phys. Chem. B* **2023**, *127*, 3829−3838.
(49) Clementi, C.; Nymeyer, H.; Onuchic, J. N. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* **2000**, *298*, 937−953.
(50) Zheng, W.; Schafer, N. P.; Davtyan, A.; Papoian, G. A.; Wolynes, P. G. Predictive energy landscapes for protein-protein association. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 19244−19249.
(51) Hinckley, D. M.; Freeman, G. S.; Whitmer, J. K.; de Pablo, J. J. An experimentally-informed coarse-grained 3-Site-Per-Nucleotide model of DNA: structure, thermodynamics, and dynamics of hybridization. *J. Chem. Phys.* **2013**, *139*, 144903.
(52) Freeman, G. S.; Hinckley, D. M.; Lequieu, J. P.; Whitmer, J. K.; de Pablo, J. J. Coarse-grained modeling of DNA curvature. *J. Chem. Phys.* **2014**, *141*, 165103.
(53) Ortiz, V.; de Pablo, J. J. Molecular Origins of DNA Flexibility: Sequence Effects on Conformational and Mechanical Properties. *Phys. Rev. Lett.* **2011**, *106*, 238107.
(54) Lequieu, J.; Córdoba, A.; Schwartz, D. C.; de Pablo, J. J. Tension-Dependent Free Energies of Nucleosome Unwrapping. *ACS Cent. Sci.* **2016**, *2*, 660−666.
(55) Dey, P.; Bhattacherjee, A. Structural Basis of Enhanced Facilitated Diffusion of DNA-Binding Protein in Crowded Cellular Milieu. *Biophys. J.* **2020**, *118*, 505−517.
(56) Mondal, A.; Mishra, S. K.; Bhattacherjee, A. Nucleosome breathing facilitates cooperative binding of pluripotency factors Sox2 and Oct4 to DNA. *Biophys. J.* **2022**, *121*, 4526−4542.
(57) Zandarashvili, L.; Esadze, A.; Vuzman, D.; Kemme, C. A.; Levy, Y.; Iwahara, J. Balancing between affinity and speed in target DNA search by zinc-finger proteins via modulation of dynamic conformational ensemble. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, E5142−E5149.
(58) Tsai, M.-Y.; Zhang, B.; Zheng, W.; Wolynes, P. G. Molecular Mechanism of Facilitated Dissociation of Fis Protein from DNA. *J. Am. Chem. Soc.* **2016**, *138*, 13497−13500.
(59) Chen, X.; Tsai, M. Y.; Wolynes, P. G. The Role of Charge Density Coupled DNA Bending in Transcription Factor Sequence Binding Specificity: A Generic Mechanism for Indirect Readout. *J. Am. Chem. Soc.* **2022**, *144*, 1835−1845.
(60) Biyun, S.; Cho, S. S.; Thirumalai, D. Folding of Human Telomerase RNA Pseudoknot Using Ion-Jump and Temperature-Quench Simulations. *J. Am. Chem. Soc.* **2011**, *133*, 20634−20643.
(61) Li, S.; Olson, W. K.; Lu, X.-J. Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. *Nucleic Acids Res.* **2019**, *47*, W26−W34.
(62) Tan, C.; Takada, S. Dynamic and Structural Modeling of the Specificity in Protein−DNA Interactions Guided by Binding Assay and Structure Data. *J. Chem. Theory Comput.* **2018**, *14*, 3877−3889.
(63) Papagiannis, C. V.; Sam, M. D.; Abbani, M. A.; Yoo, D.; Cascio, D.; Clubb, R. T.; Johnson, R. C. Fis targets assembly of the Xis nucleoprotein filament to promote excisive recombination by phage lambda. *J. Mol. Biol.* **2007**, *367*, 328−343.
(64) Bushman, W.; Yin, S.; Thio, L. L.; Landy, A. Determinants of directionality in lambda site-specific recombination. *Cell* **1984**, *39*, 699−706.
(65) Sam, M. D.; Papagiannis, C. V.; Connolly, K. M.; Corselli, L.; Iwahara, J.; Lee, J.; Phillips, M.; Wojciak, J. M.; Johnson, R. C.; Clubb, R. T. Regulation of directionality in bacteriophage lambda site-specific recombination: structure of the Xis protein. *J. Mol. Biol.* **2002**, *324*, 791−805.
(66) Sam, M. D.; Cascio, D.; Johnson, R. C.; Clubb, R. T. Crystal structure of the excisionase-DNA complex from bacteriophage lambda. *J. Mol. Biol.* **2004**, *338*, 229−240.
(67) Rohs, R.; West, S. M.; Sosinsky, A.; Liu, P.; Mann, R. S.; Honig, B. The role of DNA shape in protein-DNA recognition. *Nature* **2009**, *461*, 1248−1253.
(68) Zhou, T.; Yang, L.; Lu, Y.; Dror, I.; Dantas Machado, A. C.; Ghane, T.; Di Felice, R.; Rohs, R. DNAshape: a method for the high-throughput prediction of DNA structural features on a genomic scale. *Nucleic Acids Res.* **2013**, *41*, W56−W62.
(69) Hancock, S. P.; Ghane, T.; Cascio, D.; Rohs, R.; Di Felice, R.; Johnson, R. C. Control of DNA minor groove width and Fis protein binding by the purine 2-amino group. *Nucleic Acids Res.* **2013**, *41*, 6750−6760.
(70) Machado, M. R.; Pantano, S. Exploring LacI-DNA dynamics by multiscale simulations using the SIRAH force field. *J. Chem. Theory Comput.* **2015**, *11*, 5012−5023.
(71) Thompson, J. F.; de Vargas, L. M.; Koch, C.; Kahmann, R.; Landy, A. Cellular factors couple recombination with growth phase: characterization of a new component in the lambda site-specific recombination pathway. *Cell* **1987**, *50*, 901−908.
(72) Schlitter, J. Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem. Phys. Lett.* **1993**, *215*, 617−621.
(73) Schäfer, H.; Mark, A. E.; van Gunsteren, W. F. Absolute entropies from molecular dynamics simulation trajectories. *J. Chem. Phys.* **2000**, *113*, 7809−7817.
(74) Schäfer, H.; Daura, X.; Mark, A. E.; van Gunsteren, W. F. Entropy calculations on a reversibly folding peptide: changes in solute free energy cannot explain folding behavior. *Proteins* **2001**, *43*, 45−56.
(75) Harris, S. A.; Gavathiotis, E.; Searle, M. S.; Orozco, M.; Laughton, C. A. Cooperativity in Drug-DNA Recognition:A Molecular Dynamics Study. *J. Am. Chem. Soc.* **2001**, *123*, 12658−12663.
(76) Baron, R.; de Vries, A. H.; Hünenberger, P. H.; van Gunsteren, W. F. Comparison of atomic-level and coarse-grained models for liquid hydrocarbons from molecular dynamics configurational entropy estimates. *J. Phys. Chem. B* **2006**, *110*, 8464−8473.
(77) Debnath, A.; Schäfer, L. V. Structure and Dynamics of Phospholipid Nanodiscs from All-Atom and Coarse-Grained Simulations. *J. Phys. Chem. B* **2015**, *119*, 6991−7002.
(78) Oguey, C.; Foloppe, N.; Hartmann, B. Understanding the sequence-dependence of DNA groove dimensions: implications for DNA interactions. *PLoS One* **2010**, *5*, No. e15931.
(79) Wlodek, S. T.; Clark, T. W.; Scott, L. R.; McCammon, J. A. Molecular Dynamics of Acetylcholinesterase Dimer Complexed with Tacrine. *J. Am. Chem. Soc.* **1997**, *119*, 9513−9522.


https://doi.org/10.1021/jacsau.4c00393
*JACS Au* 2024, 4, 2640−2655

(80) Sherer, E. C.; Harris, S. A.; Soliva, R.; Orozco, M.; Laughton, C. A. Molecular Dynamics Studies of DNA A-Tract Structure and Flexibility. *J. Am. Chem. Soc.* **1999**, *121*, 5981−5991.
(81) Schreiber, T. Measuring Information Transfer. *Phys. Rev. Lett.* **2000**, *85*, 461−464.
(82) Hacisuleyman, A.; Erman, B. Entropy Transfer between Residue Pairs and Allostery in Proteins: Quantifying Allosteric Communication in Ubiquitin. *PLoS Comput. Biol.* **2017**, *13*, No. e1005319.
(83) Gourévitch, B.; Eggermont, J. J. Evaluating information transfer between auditory cortical neurons. *J. Neurophysiol.* **2007**, *97*, 2533−2543.
(84) Staniek, M.; Lehnertz, K. Symbolic transfer entropy. *Phys. Rev. Lett.* **2008**, *100*, 158101.
(85) Kamberaj, H.; van der Vaart, A. Extracting the causality of correlated motions from molecular dynamics simulations. *Biophys. J.* **2009**, *97*, 1747−1755.
(86) Kim, S.; Broströmer, E.; Xing, D.; Jin, J.; Chong, S.; Ge, H.; Wang, S.; Gu, C.; Yang, L.; Gao, Y. Q.; Su, X. D.; Sun, Y.; Xie, X. S. Probing allostery through DNA. *Science* **2013**, *339*, 816−819.
(87) Rosenblum, G.; Elad, N.; Rozenberg, H.; Wiggers, F.; Jungwirth, J.; Hofmann, H. Allostery through DNA drives phenotype switching. *Nat. Commun.* **2021**, *12*, 2967.
(88) Segers, M.; Voorspoels, A.; Sakaue, T.; Carlon, E. Mechanisms of DNA-Mediated Allostery. *Phys. Rev. Lett.* **2023**, *131*, 238402.
(89) Berg, O. G.; Winter, R. B.; von Hippel, P. H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* **1981**, *20*, 6929−6948.
(90) Bauer, M.; Metzler, R. Generalized facilitated diffusion model for DNA-binding proteins with search and recognition states. *Biophys. J.* **2012**, *102*, 2321−2330.
(91) Wunderlich, Z.; Mirny, L. A. Spatial effects on the speed and reliability of protein-DNA search. *Nucleic Acids Res.* **2008**, *36*, 3570−3578.
(92) Halford, S. E.; Marko, J. F. How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* **2004**, *32*, 3040−3052.
(93) Shvets, A. A.; Kochugaeva, M. P.; Kolomeisky, A. B. Mechanisms of Protein Search for Targets on DNA: Theoretical Insights. *Molecules* **2018**, *23*, 2106.
(94) Sokolov, I. M.; Metzler, R.; Pant, K.; Williams, M. C. Target search of N sliding proteins on a DNA. *Biophys. J.* **2005**, *89*, 895−902.
(95) Hu, T.; Grosberg, A. Y.; Shklovskii, B. I. How proteins search for their specific sites on DNA: the role of DNA conformation. *Biophys. J.* **2006**, *90*, 2731−2744.
(96) Viadiu, H.; Aggarwal, A. K. Structure of BamHI bound to nonspecific DNA: a model for DNA sliding. *Mol. Cell* **2000**, *5*, 889−895.
(97) Kalodimos, C. G.; Biris, N.; Bonvin, A. M.; Levandoski, M. M.; Guennuegues, M.; Boelens, R.; Kaptein, R. Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. *Science* **2004**, *305*, 386−389.
(98) von Hippel, P. H. From "simple" DNA-protein interactions to the macromolecular machines of gene expression. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 79−105.
(99) Zhou, H. X. Rapid search for specific sites on DNA through conformational switch of nonspecifically bound proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 8651−8656.
(100) Mirny, L.; Slutsky, M. Z. W. A. T.; Wunderlich, Z.; Tafvizi, A.; Leith, J.; Kosmrlj, A. How a protein searches for its site on DNA: the mechanism of facilitated diffusion. *J. Phys. A: Math. Theor.* **2009**, *42*, 434013.
(101) Kochugaeva, M.; Shvets, A. A.; Kolomeisky, A. How conformational dynamics influences the protein search for targets on DNA. *J. Phys. A: Math. Theor.* **2016**, *49*, 444004.
(102) Niranjani, G.; Murugan, R. Generalized theory on the mechanism of site-specific DNA-protein interactions. *J. Phys. A: Math. Theor.* **2016**, *2016*, 053501.
(103) Bauer, M.; Rasmussen, E. S.; Lomholt, M. A.; Metzler, R. Real sequence effects on the search dynamics of transcription factors on DNA. *Sci. Rep.* **2015**, *5*, 10072.
(104) Marcovitz, A.; Levy, Y. Frustration in protein-DNA binding influences conformational switching and target search kinetics. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 17957−17962.
(105) Cuculis, L.; Abil, Z.; Zhao, H.; Schroeder, C. M. Direct observation of TALE protein dynamics reveals a two-state search mechanism. *Nat. Commun.* **2015**, *6*, 7277.
(106) Zandarashvili, L.; Vuzman, D.; Esadze, A.; Takayama, Y.; Sahu, D.; Levy, Y.; Iwahara, J. Asymmetrical roles of zinc fingers in dynamic DNA-scanning process by the inducible transcription factor Egr-1. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, E1724−E1732.
(107) Tafvizi, A.; Huang, F.; Fersht, A. R.; Mirny, L. A.; van Oijen, A. M. A single-molecule characterization of p53 search on DNA. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 563−568.
(108) Leven, I.; Levy, Y. Quantifying the two-state facilitated diffusion model of protein-DNA interactions. *Nucleic Acids Res.* **2019**, *47*, 5530−5538.
(109) Cencini, M.; Pigolotti, S. Energetic funnel facilitates facilitated diffusion. *Nucleic Acids Res.* **2018**, *46*, 558−567.
(110) Hizver, J.; Rozenberg, H.; Frolow, F.; Rabinovich, D.; Shakked, Z. DNA bending by an adenine−thymine tract and its role in gene regulation. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 8490−8495.
(111) Haran, T. E.; Mohanty, U. The unique structure of A-tracts and intrinsic DNA bending. *Q. Rev. Biophys.* **2009**, *42*, 41−81.
(112) Balaceanu, A.; Pérez, A.; Dans, P. D.; Orozco, M. Allosterism and signal transfer in DNA. *Nucleic Acids Res.* **2018**, *46*, 7554−7565.
(113) Lefstin, J. A.; Yamamoto, K. R. Allosteric effects of DNA on transcriptional regulators. *Nature* **1998**, *392*, 885−888.


https://doi.org/10.1021/jacsau.4c00393
*JACS Au* 2024, 4, 2640−2655