Title

Nonspecific vs. specific DNA binding free energetics of a transcription factor domain protein

Journal

Biophysical Journal, 122(22)

ISSN

0006-3495

Authors

Al Masri, Carmen

Wan, Biao

Yu, Jin

Publication Date

2023-11-01

DOI

10.1016/j.bpj.2023.10.025

Peer reviewed

Biophysical Journal
Article

Biophysical Society

Nonspecific vs. specific DNA binding free energetics of a transcription factor domain protein

Carmen Al Masri,[1] Biao Wan,[2] and Jin Yu[1,3,*]
[1]Department of Physics and Astronomy, University of California, Irvine, California; [2]Wenzhou Institute, University of Chinese Academy of Sciences, Wenzhou, China; and [3]Department of Physics and Astronomy, Department of Chemistry, NSF-Simons Center for Multiscale Cell Fate Research, University of California, Irvine, California

ABSTRACT Transcription factor (TF) proteins regulate gene expression by binding to specific sites on the genome. In the facilitated diffusion model, an optimized search process is achieved by the TF alternating between 3D diffusion in the bulk and 1D diffusion along DNA. While undergoing 1D diffusion, the protein can switch from a search mode for fast diffusion along nonspecific DNA to a recognition mode for stable binding to specific DNA. It was recently noticed that, for a small TF domain protein, reorientations on DNA happen between the nonspecific and specific DNA binding. We here conducted all-atom molecular dynamics simulations with steering forces to reveal the protein-DNA binding free energetics, confirming that the search and recognition modes are distinguished primarily by protein orientations on the DNA. As the binding free energy difference between the specific and nonspecific DNA system slightly deviates from that being estimated directly from dissociation constants on 15-bp DNA constructs, we hypothesize that the discrepancy can come from DNA sequences flanking the 6-bp central binding sites that impact on the dissociation kinetics measurements. The hypothesis is supported by a simplified spherical protein-DNA model along with stochastic simulations and kinetic modeling.

SIGNIFICANCE How transcription factors locate their target site on the genome is essential to genetic regulation. It has been assumed that the protein switches conformations between search and recognition modes upon binding to nonspecific and specific DNA, respectively, to efficiently locate the target. By employing all-atom molecular dynamics simulations on the WRKY transcription factor domain protein bound to nonspecific and specific W-box DNA and calculating corresponding binding free energies, we demonstrate that the two DNA binding modes, i.e., search (weak binding) and recognition (strong binding), can be achieved via highly distinguishable protein domain orientations on the DNA rather than the protein internal conformational changes. We also suggest kinetic impacts from flanking DNA on measuring the relative protein-DNA binding affinity to the central DNA binding sites.

INTRODUCTION

Transcription factors (TFs) are proteins that regulate gene expression by binding to specific DNA sequences called TF binding sites (TFBSs), which are usually 6–20 bp long (1). Once bound to the TFBS, TFs can promote or block the recruitment of other proteins and enzymes such as RNA polymerase to initiate transcription, resulting in gene activation or silencing (1,2). To carry out their function, TFs must efficiently navigate through the genome to find their target binding sites among millions to billions of base-pairs, at a rate up to 100 times faster than predicted by free diffusion or random collisions with the target site (3–5). The facilitated diffusion model is a widely accepted theoretical framework for the search process of TFs (4,6–12). In this model, the TF diffuses randomly in three-dimensional (3D) space until it collides with a DNA site, which is usually a nonspecific binding site. Then, the TF can slide or hop along the DNA in one-dimensional (1D) diffusion, alternating between 1D and 3D diffusion processes until it reaches the target site on the genome. Various studies from both in vitro and in vivo settings support this model (13–25), as well as computational studies (26–32). The 1D diffusion of TFs has been further modeled by considering the sequence-dependent interaction potential energy between the protein and DNA (33,34). Studies estimate that the 1D diffusion free energy landscape fluctuates at a magnitude of 1–2 k_BT (33,34). On the other hand, once the TF locates

Submitted March 6, 2023, and accepted for publication October 20, 2023.
*Correspondence: jin.yu@uci.edu
Editor: Diego Ferreiro.
https://doi.org/10.1016/j.bpj.2023.10.025

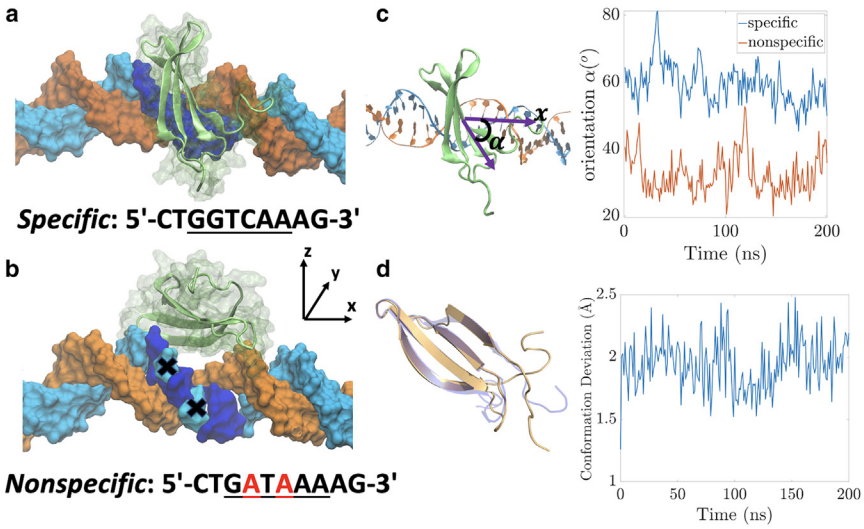

FIGURE 1 An overview of the MD simulation systems. (*a* and *b*) The WRKY domain protein bound to the specific (GGTCAA) and nonspecific (GATAAA) DNA sequences, respectively. The preferred strand is shown in cyan, while the non-preferred strand is in orange, with the binding site highlighted in dark blue. In the nonspecific system, the mutated nucleotides in the binding site are marked by an X. (*c*) The protein orientations on the DNA with respect to the x-axis (DNA long axis) for specific and nonspecific systems, with the orientation angle α defined as shown in the left diagram. (*d*) The deviation between the protein backbone conformations in the specific and nonspecific systems. The left figure shows the two protein conformations aligned. The protein from the specific system is shown in transparent blue, and that from the nonspecific one in orange. The average conformation difference was 1.95 ± 0.2 Å throughout the equilibration period. The protein conformations from the specific and nonspecific systems were aligned using pyMOL by excluding outlier atoms over five iterations, and the all-atom root mean-square deviation between the two structures was computed. To see this figure in color, go online.

the target site it is able to form a stable protein-DNA complex. Such a stability appears to require the binding free energy strength to be much larger than k_BT (33). Thus, it was proposed that TFs adopt two conformations while bound to DNA: the search mode, with relatively weak protein-DNA interactions to allow for fast protein 1D diffusion, and the recognition mode, which allows for the stable binding of the TF to the DNA once the target site is located. Such conformational changes have been extensively studied in the literature (35–38), and have been noticed in multiple TFs such as p53 (39,40), Lac repressor (41–43), and MTERF1 (44), and their structural properties and dynamics have been investigated through all-atom and coarse-grained molecular dynamics (MD) simulations (45–54). In particular, in the seminal work by Marklund et al. (45) the free-energy landscape of TF sliding on nonspecific DNA has been constructed through all-atom MD simulations, revealing a sliding energy profile roughness of $\sim 1 - 2k_BT$ and a dissociation barrier of $\sim 12k_BT$. The following all-atom MD study by Yonetani et al. on the LacI protein (50) revealed a dissociation energy barrier difference between specific and nonspecific DNA binding of $\sim 10k_BT$, which was attributed to the protein conformational variations between the protein in search vs. recognition mode. The transition between nonspecific (search mode) and specific (recognition mode) DNA binding states of TFs has been primarily attributed to large-scale protein conformational changes in the two-state model of 1D diffusion. However, recent all-atom MD simulations on the WRKY TF domain protein (55) have shown that such protein conformational changes might not be necessary to switch between the two states. Starting from a crystal structure of the WRKY domain TF bound to a specific DNA sequence (i.e., the W-box), the DNA sequence was subsequently changed to nonspecific or poly(A) sequences. Given that the protein internal or core conformation (i.e., backbone, excluding side chains and peripheral loops) in the three systems were almost identical while the protein orientations on the DNA varied significantly in the simulations, it was suggested that protein-DNA interaction energies are lowered when the protein changes orientation, allowing for the switch between nonspecific and specific modes. Nonetheless, for the nonspecific DNA binding simulation, a possibility remains that the necessary protein internal conformational change for switching from the recognition mode to the search mode has not yet been sampled within the microsecond-long MD simulations. Free energetic calculations on both nonspecific and specific DNA binding systems are therefore necessary to determine that search and recognition modes are achieved, respectively. Accordingly, we aim in this study to distinguish the protein recognition and search modes by calculating the binding free energetics of the WRKY domain protein on specific and nonspecific DNA, respectively (Fig. 1). It is expected that the protein stably bound on the specific DNA in the recognition mode would display a highly stabilized binding free energy ΔG_b^s (with $\Delta G_b^s < 0$ and $|\Delta G_b^s| \gg k_BT$), while the protein in association with the nonspecific DNA in the search mode would show a significantly smaller magnitude of the binding free energy ΔG_b^{ns} (with $\Delta G_b^{ns} < 0$ and $|\Delta G_b^{ns}| < |\Delta G_b^s|$). Starting from the microsecond equilibrated WRKY domain protein structures bound with specific and nonspecific DNA from the previous studies (55,56), we conducted all-atom MD simulations to calculate the corresponding protein-DNA binding free energies ΔG_b^s and ΔG_b^{ns}, respectively, by enforcing dissociation of the protein from the DNA binding site. The binding free energy ΔG_b was obtained using two methods: Jarzynski's equality (57) and then umbrella sampling (58), both widely used for binding free energy calculations (2,59–63). We confirmed that, although the protein internal conformations remained identical in both systems,

the protein re-orientations on the DNA resulted in well-distinguishable binding free energies, with $|\Delta G_b^{ns}| < |\Delta G_b^{s}|$. We also observed that protein dissociation dynamics were coupled with protein diffusion or horizontal motions along the DNA, accompanied by notable protein rotational motions, particularly as hydrogen-bonding (HB) interactions at the protein-DNA interface gradually diminished. Finally, we compared our computed difference in binding free energy between the specific and nonspecific DNA $\Delta\Delta G_b$ with that estimated from experimentally measured dissociation constants. Although the computed $\Delta\Delta G_b$ is comparable with the experimental measurements and consistent with the computed $\Delta\Delta G_b$ reported in previous studies on the LacI system (50), there are still some deviations between the computational and experimental estimations. While computational sampling and experimental accuracy can be contributing factors, we hypothesized that impacts from DNA sequences flanking the core (i.e., central) DNA binding site can also be a systematic source of discrepancy. Investigations through a simplified spherical protein model and stochastic dynamics simulations together showed that the flanking DNA regions impacted on measurements of protein dissociation kinetics and consequently the relative protein-DNA binding affinities.

MATERIALS AND METHODS

System setup

To compare the binding of the WRKY domain protein to specific and nonspecific DNA, we simulated two systems: the WRKY1-N in complex with a W-box DNA from the crystal structure (accession PDB: 6J4E), and the WRKY1-N in complex with a mutated W-box sequence that reduces the DNA binding affinity by ~76-fold. The former is referred to as specific DNA (5′-CTGGTCAAAG-3′), and the latter as nonspecific DNA (5′-CTGATAAAAG-3′). Following the protocol by Dai et al. (55), we extended the 15-bp DNA constructs to 34 bp to avoid DNA end-effects in the simulations. The CYS and HIS residues in the WRKY protein were changed to CYM (deprotonated state) and HIE (residue 164, with γ-nitrogen protonated) or HID (residues 138 and 166, with ϵ-nitrogen protonated), respectively, to form a stable zinc finger domain. Starting structures for our all-atom MD simulations were taken from previously generated 10-μs equilibrium trajectories (55), at $t = 8\mu$s. The simulation systems are shown in Fig. 1, *a* and *b*.

MD simulations setup

All-atom MD simulations were carried out using GROMACS 2020.4 (64) with the Amber99SB-ILDN force field (65) for protein and the Parmbsc1 (BCS1) force field for nucleic acids (66). The protein-DNA systems were solvated with TIP3P water in a rectangular box, with a minimum distance from the complex to the boundary of the simulation box of 15 Å. The system was neutralized with Na^+ and Cl^- ions at an ionic concentration of 0.15 M. The total simulation system contained ~85,000 atoms. Periodic boundary conditions were applied. Short-range electrostatic and vdW interactions used a cutoff of 10 Å, while the long-range electrostatic interactions were treated using the particle-mesh Ewald method (67). The solvated system underwent a steepest-descent minimization, followed by a 4-ns equilibration under the NVT ensemble and another 4-ns equilibration under the NPT ensemble with a time step of 2 fs. Position restraints with a force constant of 1000 kJ mol^{-1} nm^{-2} were applied to the heavy atoms of the system. The simulations used the leap-frog stochastic dynamics integrator, with the temperature set to 300 K and the inverse friction constant set to 2 ps. Pressure was set to 1 bar using the Parrinello-Rahman barostat pressure control method. Finally, the position restraints were lifted, and a final equilibration of 200 ns was conducted, with restraints applied only to the edges of the DNA (heavy atoms of the first and last residues) to keep the DNA axis aligned to the x-axis.

Equilibrium reached for the specific and nonspecific complexes

To confirm the attainment of equilibrium in the MD simulations for both specific and nonspecific systems, we evaluated the root mean-square deviation (RMSD) of the protein, as well as its orientation and radius of gyration changes. Using the final 200 ns of unconstrained equilibration, we determined that the RMSD of the protein conformation stabilized after ~80 ns for both the specific and nonspecific systems (Fig. S1 *a*). The nonspecific system exhibited a more significant change in protein orientation over time (16.6° ± 7.9° for nonspecific vs. 10.1° ± 2.8° for specific) (Fig. S1 *b*), with the protein orientation angles with respect to the DNA α found to be notably different between the specific (57° ± 8°) and nonspecific (32° ± 3°) systems throughout the equilibration process (Fig. 1 *c*). In addition, we observed small changes in the radius of gyration over time (−0.06 ± 0.16 Å for the specific system vs. −0.25 ± 0.19 Å for the nonspecific system) (Fig. S1 *c*), providing further evidence that the protein conformation achieved equilibrium in both systems.

Steered MD (SMD) simulations setup

To force the dissociation of the protein, its center of mass (COM) was steered along a designated reaction coordinate, denoted as ξ. This coordinate was aligned with the vertical dissociation pathway relative to the DNA molecule, which was oriented along the *x*-axis. Through our chosen coordinate system, we observed that the protein dissociation was most likely to occur in the $-y$-direction for the specific DNA system and in the $+z$-direction for the nonspecific DNA system (Fig. 2, *bottom*). To demonstrate this, we employed the random acceleration MD technique implemented in GROMACS (68) by randomly steering the COM of the protein (Figs. 2, *top*, and S2). To facilitate the steering process, we applied a spring force with a force constant of 3000 kJ mol^{-1} nm^{-2} to the protein's COM, guiding it along the ξ-direction at a speed of 0.1 Å/ns. Such a speed was found to be affordable while achieveing similar results on constructing potential of mean force (PMF) as reversible or quasistatic pulling in a simplified test model system (Fig. S3). To prevent the DNA molecule from being dragged along with the protein during the pulling process, we restrained the heavy atoms of the DNA using harmonic restoring forces with a force constant of 1000 kJ mol^{-1} nm^{-2}.

PMF calculations using Jarzynski's equality

Jarzynski's equality (57) enables the calculation of free energy differences from work measurements during nonquasistatic processes. It is expressed as:

$$e^{-\beta\Delta F} = \langle e^{-\beta W} \rangle \quad (1)$$

where ΔF represents the free energy difference, W is the work performed, $\langle \ldots \rangle$ denotes ensemble averaging, and $\beta = 1/k_B T$ with T the temperature of the system. To overcome computational challenges arising from the exponential term being dominated by small work values, an approximate form using cumulant expansion, assuming a Gaussian work distribution, can be employed. By applying Jarzynski's equality, the free energy profile along the reaction coordinate $\Phi(\xi)$ (or PMF) can be computed. Along a

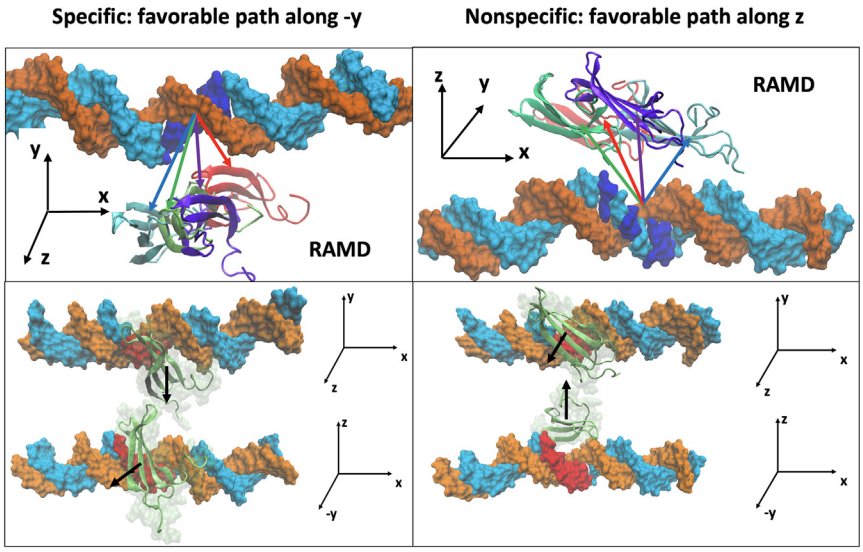

FIGURE 2 Top: utilizing random acceleration MD (RAMD) to determine the directions of protein dissociation in both specific and nonspecific systems. The COM of the protein is steered in a random direction using a constant force of $k = 2000$ kJ mol^{-1} nm^{-2}. If the COM does not move by a distance larger than 0.25 Å within 100 fs, the direction is updated to another random one until the protein reaches a user-defined maximum distance (2 nm in this study). Snapshots of four different trajectories are shown for both specific (*left*) and nonspecific (*right*) DNA binding systems, with arrows indicating the dominant directions of protein dissociation (protein and arrows colored according to the sampled trajectory). Bottom: visualization of the preferred direction of protein dissociation from different perspectives in each system. The $-y$-direction is found to be preferred in the specific binding system, while the z-direction is preferred in the nonspecific system. Supporting figures are provided in the Fig. S2. To see this figure in color, go online.

well-defined reaction coordinate ξ, when quasistatic equilibrium is reached, the protein's other degrees of freedom are fully relaxed. That is, at each reaction coordinate $\xi(\mathbf{r})$, the PMF is in principle calculated as an ensemble average:

$$\Phi(\xi') = -k_B T \ln \int \exp\left(-\frac{E(\mathbf{r},\mathbf{p})}{k_B T}\right)\delta(\xi(\mathbf{r}) - \xi')d\mathbf{r}d\mathbf{p} \tag{2}$$

Here, $E(\mathbf{r},\mathbf{p})$ is the total energy function of the system, ξ the reaction coordinate, and $\mathbf{r},\mathbf{p}$ are the phase space coordinates of the rest of the degrees of freedom, "orthogonal" to ξ in the system. In our simulations, the path was sampled via the SMD simulations (69), and the PMF was determined using the stiff-spring approximation:

$$\Phi(\lambda_t) = \Phi(\lambda_0) + \langle W\rangle - \frac{\beta}{2}\left(\langle W^2\rangle - (\langle W\rangle)^2\right) + \ldots \tag{3}$$

Where λ_t and λ_0 represent the parameter λ values (defined in the previous section) at times t and 0, respectively, and W is the work sampled between 0 and t. Biases in sampling were mitigated by employing segment-based pulling while alternating between forward and reverse pulling events, and allowing overlap or shared initial positions between neighboring segments. Additional details can be found in supporting text S-IV and Fig. S4.

PMF calculations using umbrella sampling

The free energy along a reaction coordinate ξ (or the PMF) is directly related to how ξ is populated. Umbrella sampling allows computing the PMF by applying an external potential to enhance sampling of energetically unfavorable regions (58). Multiple umbrella sampling windows, each biasing the system near a specific region parameterized by λ_i with $i = 1,\ldots,N$ were employed. Umbrella histograms $h_i(\xi)$ were recorded, representing the biased probability distributions $P_i^b(\xi)$ along ξ. The unbiased probability distribution $P(\xi)$ and the PMF $\Phi(\xi)$ were then by obtained by reweighting or unbiasing the $P_i^b(\xi)$ using the weighted histogram analysis method (WHAM) (70). In this work, the GROMACS implementation of WHAM, g_wham (71), was used to compute $P(\xi)$ and $\Phi(\xi)$. A single SMD trajectory of 200 ns was performed to sample the dissociation path, with protein-DNA dissociation occurring ~100 ns as observed from the complete breaking of HBs (Fig. S5). For both specific and nonspecific DNA binding systems, 21 umbrella sampling windows were launched with a spacing of ~0.1 nm. The COM of the protein was restrained by a harmonic potential using a force constant of 3000 kJ mol^{-1} nm^{-2} during each window simulation, and a PMF for protein-DNA dissociation was generated. Convergence of the PMFs was assessed (Fig. S6). To estimate the errors along the obtained PMF, Bayesian bootstrapping of histograms was conducted as detailed in (71), whereby each umbrella window generated 10 independent histograms, which were then chosen by sampling with replacement to generate a new PMF. Each histogram then had a corresponding probability to be sampled given by a random weight. Finally, the average PMF was obtained, with the errors being the standard deviations along the bootstrapped PMFs.

Simplified spherical protein model

To investigate how the protein binding free energetics obtained from the protein-DNA dissociation PMFs connect with the protein dissociation events and kinetics sampled randomly in a highly simplified quantitative model, the domain protein is modeled by a positively charged sphere (Fig. S7). In this model, discretized positive and negative charges were placed on the surface of the spherical protein with overall neutrality. The DNA is modeled by a line of negative charges. The protein-binding sites on the DNA were then represented by linearly aligned points with spacing of 0.34 nm. In addition, the protein-DNA interfacial HB interactions were represented by the pair-interaction between the sites on the protein hemisphere and the sites on DNA. The protein-DNA electrostatic interactions screened by a solution with ions can be approximated by the Debye-Huckel potential, and the interfacial HB interaction can be described by the Morse potential depending on the bond angle or protein orientation (supporting text S-VII and Fig. S8). More details on the implementation can be found in supporting text S-VII.

RESULTS

To assess whether the protein-DNA binding free energy difference $\Delta\Delta G_b$ between the nonspecific and specific DNA constructs are consistent with the definitions of search and recognition modes, several steps were taken. Initially, PMFs were computed using Jarzynski's equality for both systems, allowing the computation of the respective binding free energies. However, given that the SMD trajectories

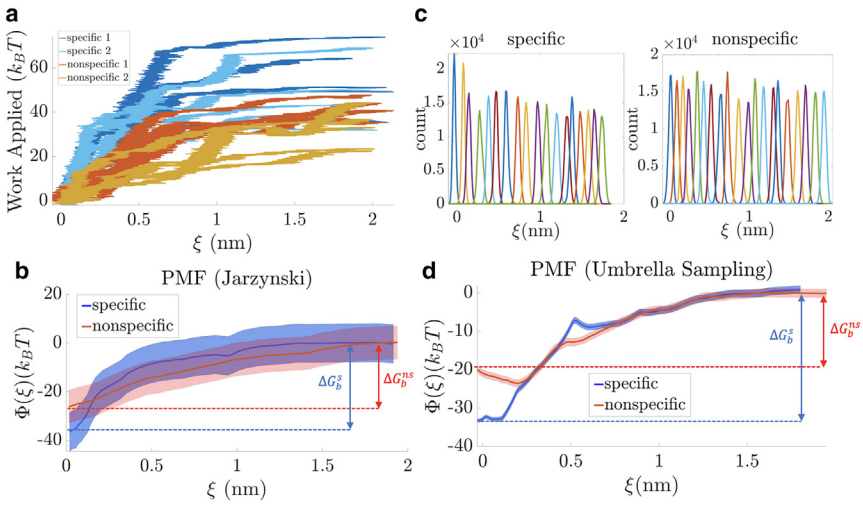

FIGURE 3 PMFs of enforced protein dissociation from DNA from implementing both Jarzynski's equality and Umbrella Sampling methods. (*a*) The work from all SMD trajectories for both specific (*blue*) and nonspecific (*orange*) systems. Dark colors correspond to setup 1 and light colors to setup 2, as detailed in supporting text S-IV. (*b*) The resulting PMFs obtained from the Jarzynski equality. (*c*) The distributions of ξ, the reaction coordinate of protein pulling, in each umbrella window at t = 200 ns for both specific and nonspecific DNA binding systems, and (*d*) the resulting PMFs from umbrella sampling. The standard deviations are shown as the shaded regions. $\Delta G_b^{s/ns}$ values are obtained from the PMF as $\Phi(\xi_0 = 0) - \Phi(\xi_\infty \approx 2\text{nm})$, the energy difference between the initial protein bound state and final dissociated state. To see this figure in color, go online.

exhibited limited sampling of the protein's rotational degrees of freedom, an alternative approach was also employed. The PMFs were then computed using the umbrella sampling method, which improved relaxation of the protein's rotation dynamics. Both methods yielded $\Delta\Delta G_b \approx 10 k_BT$, and the dissociation dynamics revealed qualitatively distinct behaviors between specific and nonspecific binding. In addition, a comparison was made between the computationally determined $\Delta\Delta G_b$ and the experimentally estimated value, which showed a difference of a few k_BT and led to an investigation of potential effects of flanking sequences on the binding affinity measurements using a highly simplified spherical protein model. In the proposed model, the PMFs for protein dissociation from a 6-bp DNA binding site were computed using protocols similar to those employed in the SMD simulations. Langevin dynamics simulations were then conducted to measure the spontaneous dissociation kinetics of the protein from a 15-bp DNA, mirroring the experimental setup. By doing so, one can compare the $\Delta\Delta G_b$ values obtained from the simplified model using both the SMD PMF and the Langevin dynamics approaches, respectively.

Binding free energy measurements using SMD and Jarzynski's equality

We first used Jarzynski's equality to calculate the PMF by pulling the protein away from the DNA in specific and nonspecific binding systems. The work applied as a function of ξ for all 20 trajectories sampled is displayed in Fig. 3 *a*. We noticed large fluctuations in the work that deviated from a Gaussian distribution (Fig. S9, *a* and *b*). Consequently, convergence of the PMFs obtained by taking only the first two cumulants of the work distributions (given by Eq. (3)) was not well achieved (Fig. S10, *a* and *b*), leading to poorly behaved PMFs (Fig. S10 *c*). To improve the results, we utilized the original exponential averaging form of Jarzynski's equality (given by Eq. (1)). This approach yielded improved convergence of the PMFs, within 4 k_BT for the specific binding system and 2 k_BT for the nonspecific binding system (Fig. S10 *d* and *e*). The resulting PMFs shown in Fig. 3 *b* showed that protein dissociation along the reaction coordinate ξ exhibited a larger energy difference between the unbound and bound states. One can then obtain the binding free energy ΔG_b from the PMF $\Phi(\xi)$ as $\Delta G_b = \Phi(\xi_0) - \Phi(\xi_\infty)$, with $\Phi(\xi_0)$ the free energy of the initial protein-DNA bound state, and $\Phi(\xi_\infty)$ the free energy after protein dissociation from the DNA. Here, $\Phi(\xi_\infty)$ is taken as the reference and set to zero, with the infinity determined by the flattening of the PMF for ξ larger than a certain distance (2 nm from current MD simulations, see Fig. 3 *b*). Accordingly, we obtain $\Delta G_b^{ns} = -26 \pm 2\ k_BT$ and $\Delta G_b^{s} = -36 \pm 4\ k_BT$, corresponding to the nonspecific and specific binding free energies, respectively. The binding free energy difference between the systems is then given by $\Delta\Delta G_b = \Delta G_b^{ns} - \Delta G_b^{s} = 10 \pm 4.5\ k_BT$. Upon analyzing the pulling trajectories used to construct the PMF via the Jarzynski's equality method, we observed highly limited protein rotation (Fig. S11). This indicated that, within the sub-microsecond pulling at 0.1 Å/ns, SMD could not provide with sufficient sampling to capture the protein's rotational degrees of freedom, which impacted on the protein free energetics and likely resulted in overestimating the absolute binding free energy ΔG_b. To address this issue, we then employed the umbrella sampling method, which allows for improved sampling on the protein's rotational degrees of freedom (see analyses below).

Binding free energy measurements using umbrella sampling

To construct the PMFs of WRKY protein dissociation from DNA, we also employed the umbrella sampling method. Convergence of the PMFs was observed at ~150 ns for

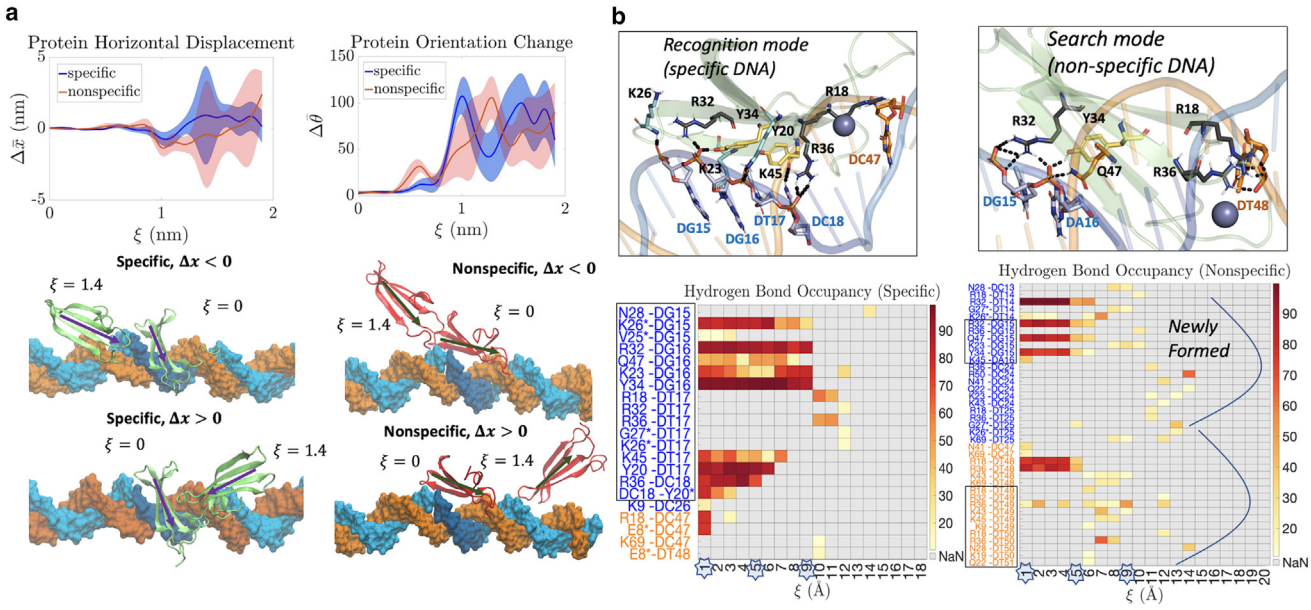

FIGURE 4 Dissociation dynamics of protein and protein-DNA interfacial HB changes during the enforced protein dissociation from DNA using the umbrella sampling method. (*a*) The protein horizontal displacements and orientational changes during the dissociation. The top left panel shows the protein horizontal displacements Δx along the DNA long axis, and the top right panel shows the protein orientation changes, $\Delta\theta$. The standard deviation of each quantity is shown via the shaded region. The bottom panel shows snapshots taken from the initial bound state ($\xi = 0$) and the intermediate dissociated state ($\xi = 1.4$ nm). The protein bound to the specific sequence is colored in green and that bound to the nonspecific sequence in red. During the enforced dissociation process, the protein can still diffuse horizontally either to the left ($\Delta x < 0$, *upper panels*) or to the right ($\Delta x > 0$, *lower panels*) in both specific and nonspecific systems. Note that the range of protein orientational change is more limited in the Jarzynski pulling method (Fig. S11) than in the umbrella sampling. (*b*) Top panels show the protein-DNA interfacial HB interactions initially formed with molecular graphics. Bottom panels display the average HB occupancy along the protein-DNA interface from umbrella sampling windows, highlighting the preferred DNA strand (sequence in *blue*) and nonpreferred strand (in *orange*) (55), for both the specific (*left*) and nonspecific (*right*) systems. To see this figure in color, go online.

both specific and nonspecific protein-DNA binding systems, with the PMF fluctuating within 1.5 k_BT, of the order of thermal fluctuations (Fig. S6). Accordingly, the last 50 ns of simulation data for each window were considered for further analyses. The overlapping histograms for the umbrella sampling windows at 200 ns are shown in Fig. 3 *c*. The final PMFs along with the error obtained through Bayesian bootstrapping analyses were generated (Fig. 3 *d*). Based on the PMF results, we determined the protein-DNA binding free energies as $\Delta G_b^s = -32 \pm 1.5\ k_BT$ for the specific system and $\Delta G_b^{ns} = -21 \pm 1.5\ k_BT$ for the nonspecific system. The difference in binding free energies between the two systems was found to be $\Delta\Delta G_b = 11 \pm 2\ k_BT$. In comparison with the PMF calculations using SMD along with the Jarzynski method, the umbrella sampling implementation exhibited reduced energetic fluctuations and improved sampling of the protein rotational degrees of freedom (Fig. 4 *a*). Interestingly, although the average value of the binding free energy decreased, the difference between the nonspecific and specific DNA binding systems remained at $\sim 10\ k_BT$. In previous experimental measurements, the dissociation constants $K_D = k_{off}/k_{on}$ for the specific and nonspecific binding systems were determined to be 0.1 and 8 μM, respectively (55). This would correspond to a binding free energy difference of $\Delta\Delta G_b = \ln(K_D^{ns}/K_D^s) \sim 4.2\ k_BT$. However, the PMF calculations in current simulations consistently indicated a slightly larger binding free energy difference of $\Delta\Delta G_b \sim 10\ k_BT$. Although both simulation and experimental measurement errors could be contributing factors, it is worth noting that the previous experiments used 15-bp DNA constructs for these measurements (55,56). To address potential systematic effects that can cause discrepancies, further investigation was conducted in the subsequent section to examine the potential impact of flanking sequences included in the DNA construct on the protein dissociation measurements.

Protein dissociation dynamics features near DNA

We analyzed the protein's structural dynamics during enforced dissociation by highlighting its horizontal diffusion along the DNA and its orientational changes relative to the DNA (Fig. 4 *a*, *top*) in specific and nonspecific binding systems. Following an initial dissociation stage ($\xi > 1$ nm), the protein exhibited a comparatively free diffusion along the DNA axis and notable rotational motion. The umbrella sampling method provided improved sampling of protein orientations, reaching up to 120° of spatial rotations (Fig. S4 *a*, *upper right*) compared with the 30° in Jarzynski's equality method (Fig. S11, *upper right panel*). Notably, in the nonspecific system, protein rotation started as early as

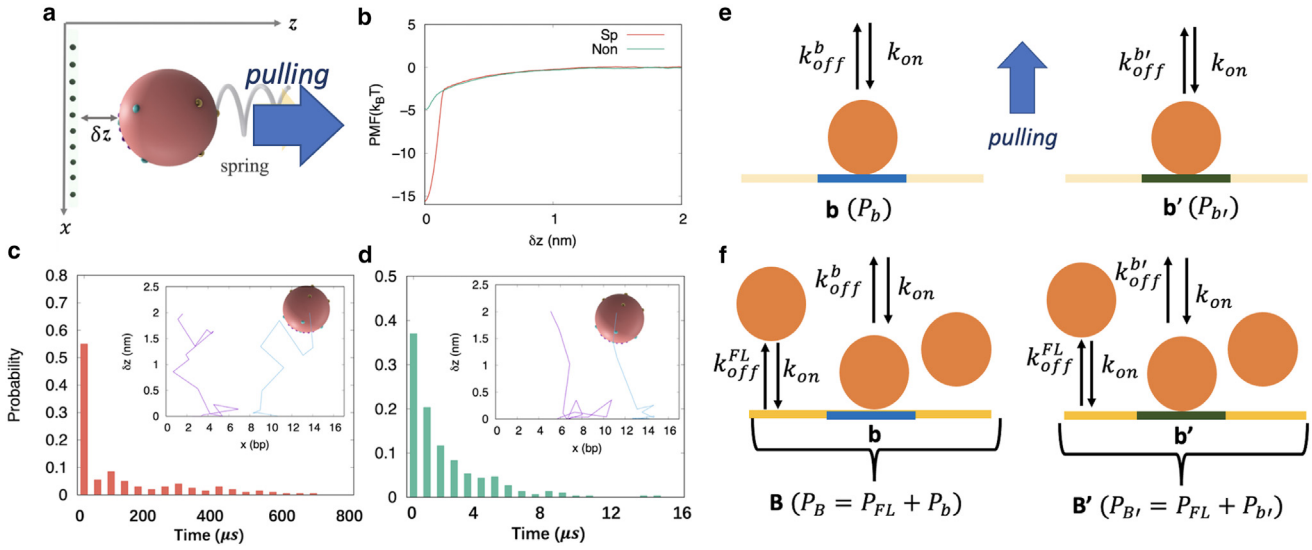

FIGURE 5 Spherical protein model comparing free energetics from steered simulations and protein-DNA dissociation kinetics, on protein binding to 15-bp DNA constructs with a 6-bp central specific/nonspecific core DNA and flanking DNA sequences. (*a*) The steering of the spherical protein using an external harmonic potential, while (*b*) compares the potential of mean force (PMF) obtained by pulling the protein away from the specific/nonspecific DNAs. Dwell-time histograms are shown in (*c*) and (*d*) for the protein bound to 15-bp DNA with specific and nonspecific sequences, respectively. The effect of flanking sequences on the relative binding free energy is depicted in (*e* and *f*), where different binding sites and their dissociation rates are indicated. The notations *b* and *b′* represent the core DNA-specific and nonspecific binding sites, respectively, while *FL* denotes the flanking sites. The association rate is assumed to be the same for all sites, denoted as k_{on}, while the dissociation rate k_{off}^{site} varies. The population on a site is denoted as P_{site}. To see this figure in color, go online.

$\xi = 0.5$ nm, indicating weaker binding and lower dissociation barriers compared with the specific DNA. In contrast, in the specific binding system, protein rotation began at $\xi \approx 1$ nm. In addition, horizontal diffusion became significant in specific DNA dissociation ($\xi \approx 1.5$ nm; Fig. 4 *a*, *upper left*), preceding nonspecific DNA dissociation ($\xi \approx 2$ nm). HB occupancy and dynamics at the protein-DNA interface were analyzed from the umbrella sampling simulations. In the specific binding system, HB contacts were mainly formed on the preferred strand (5′-CTGGTCAAAG-3′), particularly on the specific sequence. These contacts abruptly broke upon dissociation at $\xi \sim 1$ nm. After the initial breaking, only a few new HBs formed (about eight contacts), primarily with the preferred strand. In contrast, the nonspecific binding system exhibited fewer contacts with the preferred strand compared with the specific binding system. Gradual dissociation starting at $\xi \sim 0.5$ nm was observed, with the initial HBs breaking and weaker contacts emerging (up to about 35 HB contacts). After the initial breaking at $\xi \sim 0.5$ nm, no strand preference was observed as residues formed HB contacts almost equally with both strands (3 ± 1 HBs on the originally preferred strand and 2 ± 2 HBs on the nonpreferred strand).

PMF and protein dissociation kinetics from a simplified spherical protein model

In the simplified spherical protein-DNA model system, $\Delta\Delta G_b$ was estimated using two different methods. In the first method, $\Delta\Delta G_b$ was calculated using PMFs obtained from steered simulations (Figs. 5, *a* and *b*). The simulations involved applying a time-dependent harmonic potential to steer the protein away from the DNA. The Hamiltonian of the protein-DNA system was $E + 1/2k(\xi - vt)^2$, where E is the energy function of the protein-DNA system and ξ the reaction coordinate constrained under the time-dependent potential centered at vt with force constant $k = 4000$ pN/nm, at a very low velocity of $v = 1$ Å/μs. More details on the implementation can be found in Fig. S7 and supporting text S-VII. The PMFs reached a flat region very early at $\xi > 0.5$ nm. The simplified model was further calibrated with the all-atom SMD simulations by tuning the sequence-dependent protein-DNA HB strengths, so that the relative binding free energy calculated on top of the PMFs constructed from this simplified model can fit with that from the MD simulations as $\Delta\Delta G_b = 10.8\ k_BT$ (supporting text S-VII). In the second method, the free energy difference was then computed by simulating spontaneous protein dissociation from 15-bp DNA constructs, similar to the experimental setup. The free energy difference was readily obtained from $\Delta\Delta G_b = -k_BT \ln(k_{off}^{s}/k_{off}^{ns})$, where k_{off}^{s} and k_{off}^{ns} are the respective dissociation rates from the core 6-bp specific and nonspecific sequences, flanked by a nonspecific sequence. The dissociation rates can be estimated from $k_{off} = \frac{1}{\tau_d}$, where τ_d is the mean dwell time of the protein bound to the DNA. A total of 10,000 spontaneous dissociation events were recorded without considering reassociations (Figs. 5, *c* and *d*). It is important to

note that the simplified spherical protein model used in this study may lead to underestimation of the binding free energies due to the limited number of protein charges considered for calculating protein-DNA electrostatic interactions. However, this limitation is not expected to affect the binding free energy difference or the ratio of dissociation rates. The free energy difference is then measured as $\Delta\Delta G_b \approx 4.2\ k_BT$ for $\tau_d^s \sim 190\mu s$ and $\tau_d^s \sim 1.8\mu s$, which turned out to be consistent with the experimental value and was indeed significantly smaller than that obtained from the steered simulations. The result from protein-DNA dissociation rates thus underestimates the free energy difference of the core sequences due to the center core DNA sequence ($\sim$ 6 bp) being surrounded by nonspecific flanking sequences. Consequently, the protein can bind and unbind from the flanking DNA sequences, in addition to the central binding site. On average, the flanking DNA binding sites reduce the protein-DNA binding affinity compared with that being measured on the specific core DNA sequences. We therefore proceeded to model the effect of the flanking DNA binding sites on the protein-DNA binding affinity as shown in the schematics depicted in Figs. 5, *e* and *f*: we first considered DNA binding and unbinding/dissociation from only the core region of the specific (b, with protein bound population P_b) or nonspecific (b', with population $P_{b'}$) sequences (see Fig. 5 *e*). As measured from steering the protein away from the core DNA region, this gives us a difference in binding free energy $\Delta\Delta G_b^{bb'} = k_BT \ln(P_b/P_{b'}) = k_BT \ln(k_{off}^{b'}/k_{off}^{b})$. However, when b and b' are flanked by sites denoted by FL (see Fig. 5 *f*), with an approximate total dissociation rate k_{off}^{FL} and a total protein bound population P_{FL}, then the new segments $B = b + FL$ and $B' = b' + FL$ will have bound populations $P_B = P_b + P_{FL}$ and $P_{B'} = P_{b'} + P_{FL}$. The total dissociation rate will then be approximated by an average of the dissociation rates of the flanking sequences (k_{off}^{FL}) and the central sequence ($k_{off}^{b'}$ or k_{off}^{b}), weighted by the fraction of bound protein populations to FL. We denote these fractions by $\alpha \equiv \frac{P_{FL}}{P_{FL}+P_{b'}}$ and $\beta \equiv \frac{P_{FL}}{P_{FL}+P_b}$. This gives us a difference in binding free energy of:

$$\Delta\Delta G_b^{BB'} = \ln\left[\frac{\alpha \cdot k_{off}^{FL} + (1-\alpha)\cdot k_{off}^{b'}}{\beta \cdot k_{off}^{FL} + (1-\beta)\cdot k_{off}^{b}}\right] < \Delta\Delta G_b^{bb'} \quad (4)$$

A more detailed derivation and further analysis can be found in (Fig. S12 and supporting text S-IX). Note that for simplicity, we have not taken into consideration the potential DNA sequence dependence on protein-DNA association kinetics in the above analyses. If the protein binding rate to DNA k_{on} is also sequence dependent, then the relative binding free energetics would include additional effects from the association kinetics.

DISCUSSION

The TF domain protein reorients on the DNA to switch between search (nonspecific) and recognition (specific) modes

The facilitated diffusion model suggests that TFs alternate between 1D diffusion along DNA and 3D diffusion in the bulk to search for target DNA sites. To efficiently locate the target, TFs can adopt two conformations on the DNA: the search mode for rapid diffusion along nonspecific DNA and the recognition mode for stable binding to the specific target site. However, recent microsecond MD simulations on the WRKY domain protein showed that the protein core maintains the same conformation regardless of binding to nonspecific or specific DNA, but exhibits significant differences in orientation relative to the DNA (56). In this study, our focus was on elucidating the binding affinities of the WRKY domain protein on nonspecific and specific DNA to confirm that the simulated systems corresponded to the protein in the search (low DNA binding affinity) and recognition (high DNA binding affinity) modes. Our results confirmed that protein reorientation on the DNA leads to significantly less stabilized binding free energetics from specific to nonspecific DNA binding. This suggests that protein reorientation on the DNA facilitates the switch from the search to the recognition mode without requiring internal conformational changes of the protein. In particular, we determined the relative binding free energy of the WRKY domain protein between a nonspecific and specific DNA to be approximately $\Delta\Delta G_b = \Delta G_b^{ns} - \Delta G_b^{s} \approx 10\ k_BT$, consistently revealed from both the umbrella sampling and Jarzynski's equality methods. Previous predictions based on the two-state model framework indicated that the transition between TF search and recognition modes involves a change in the protein-DNA binding free energy of $\sim$5–10 k_BT (33), which is consistent with our current findings. One should nevertheless note that current findings emphasizing protein reorientations on the DNA are mainly relevant to small globular proteins or the DNA binding domain proteins. For larger or multimeric proteins, such as LacI (72–74) or p53 (75), protein internal conformational changes in the DNA binding domain have been detected upon binding to DNA specifically vs. nonspecifically.

Protein-DNA binding free energetics in the search and recognition modes

In this study, we calculated the binding free energetics of the WRKY domain protein when bound to a specific W-box sequence and a nonspecific DNA sequence that differed by two nucleotides only. Experimental measurements have shown that such minor changes in DNA sequence result in a significantly increased dissociation constant K_d ($\sim$80 times higher) compared with the W-box sequence (55,56). To calculate the binding free energy, we used SMD to

accelerate the protein-DNA dissociation by enforcing the dissociation perpendicular to the DNA long axis. It is important to note that protein spontaneous dissociation is a stochastic process in 3D space that occurs over milliseconds and can take various paths. The enforced SMD trajectories do not represent the exact protein-DNA dissociation paths but serve as an ensemble average or projection along the perpendicular reaction coordinate ξ away from DNA. Indeed, along the dissociation reaction coordinate, we observed that the protein also explores horizontal degrees of freedom along the DNA and undergoes rotational diffusion, especially during umbrella sampling simulations where the protein was allowed to relax for up to 200 ns per simulation window. Accordingly, the population sampling along the dissociation reaction coordinate effectively captures the free energetics $\Phi(\xi)$ and, in turn, the difference between the bound and unbound states of the protein with respect to the DNA, i.e., the binding free energy $\Delta G_b = \Phi(\xi_0) - \Phi(\xi_\infty)$. As a sanity check, we compared our results with the binding free energies of LacI determined in a similar study that used all-atom MD coupled with the adaptive biasing force method (50). The magnitudes of the WRKY domain protein-DNA binding free energies on specific and nonspecific DNA (21–26 vs. 32–36 k_BT) are similar to the binding free energies of LacI (27 k_BT for nonspecific DNA and 37 k_BT for specific DNA). It is conceptually important to note that a protein binding free energy of 10–30 k_BT is compatible with the diffusional search mode of the protein along DNA. Previous work has shown that the diffusional search energy barrier along DNA is ∼2–3 k_BT (55), with a rugged free energy landscape characterized by thermal fluctuations (1–2 k_BT) that agree with experimental measurements (15). This diffusional free energy profile is defined horizontally along the DNA, while the protein-DNA binding free energy ΔG_b is defined vertically, considering protein dissociation away from the DNA. Thus, a nonspecifically bound protein with a binding free energy magnitude of 10–30 k_BT at a transient binding site can still rapidly diffuse along the DNA with diffusional free energy barriers of ∼2–3 k_BT. This diffusional search relies on shifting hydrogen bonds at the protein-DNA interface, without involving protein dissociation that occurs at least 1–2 nm away from the DNA. When specific target DNA sequences are encountered during the diffusional search process, the domain protein quickly reorients on the DNA, transitioning from the search to the recognition mode. This transition is accompanied by substantial changes in protein-DNA binding free energy, such as from $\Delta G_b^{ns} \sim$ 21–26 k_BT to $\Delta G_b^{s} \sim$ 32–36 k_BT in our current system.

Coupled protein diffusion and protein-DNA HB dynamics during dissociation

To assess the sampling of protein translational and rotational degrees of freedom during enforced protein dissociation, we examined the range of protein horizontal displacements along the DNA and orientational changes. When constructing the PMF using SMD along with Jarzynski's equality, protein translational degrees along DNA were well sampled, but rotational degrees were not sufficiently explored. Limited sampling led to high fluctuations in the total work for the PMF construction. In comparison, umbrella sampling with more relaxed protein dynamics in each simulation window allowed sufficient sampling of both protein displacements and rotations, resulting in PMF convergence within ∼1–2 k_BT. Binding free energetics obtained from umbrella sampling showed decreased magnitudes compared with Jarzynski's equality method (ΔG_b^s from −36 to −32 k_BT and ΔG_b^{ns} from −26 to −21 k_BT). The relative binding free energy $\Delta\Delta G_b$ nevertheless maintains a consistent value of ∼10 k_BT, despite the different levels of protein relaxation during dissociation. Hence, $\Delta\Delta G_b$ turns out to be a comparatively robust measure between nonspecific and specific DNA sequences. By examining HBs at the protein-DNA interface, we observed that the protein formed approximately twice as many HB contacts with the six nucleotides constituting the specific site compared with the nonspecific site. In addition, HBs between the protein and DNA broke more rapidly during dissociation from the nonspecific DNA (at a distance of 5–6 Å) compared with the specific DNA (at a distance of 9–10 Å). Interestingly, during dissociation from the nonspecific DNA, the protein started reforming HB contacts with both DNA strands and was observed to randomly probe the DNA surface, consistent with the behavior expected in the search mode of a TF. In contrast, the specific complex maintained initial HB contacts at the protein-DNA interface, primarily on the preferred DNA strand, as the protein core domain (including the WRKY motif) interacted exclusively with the specific W-box sequence. After the initial HB breakage, very few new protein-DNA HB contacts reformed (approximately four times fewer than in the nonspecific system), with only two contacts formed on the nonpreferred strand. The stronger binding of the protein to specific DNA and the strong preference for the DNA strand align with the expected behavior of a protein in the recognition mode.

Simplified spherical protein model and analyses reveal flanking DNA impacts on measurements of protein-DNA affinity

The calculated protein-DNA binding free energy difference between specific and nonspecific DNA complexes in the all-atom MD simulations was ∼10 k_BT, consistent between both methods we used. In contrast, the experimentally obtained free energy difference was estimated to be ∼4.2 k_BT based on the ratio between the protein dissociation constants for nonspecific and specific DNA complexes (80 vs. 1 μM) on the 15-bp DNA constructs (56). While the $\Delta\Delta G_b$ value computed from all-atom MD is reasonable and does

not deviate far from the experimentally estimated value, it was still worth investigating potential sources of such a discrepancy beyond sampling and experimental errors. Considering that the protein binding site occupied ∼6 bp on the DNA, while the experimental DNA construct was 15 bp (56), we constructed a highly simplified spherical protein model in complex with a 15-bp DNA, with the central 6-bp core as the initial DNA binding site. In one setup, steered simulations were conducted similarly to atomic SMD simulations to enforce the protein dissociation from the 6-bp core binding site, and the relative binding free energy $\Delta\Delta G_b$ was tuned via the protein-DNA HB strength to be equal to the all-atom MD value at $\sim 10\ k_BT$. In the other setup, stochastic dynamics simulations were launched, allowing the initially bound protein to diffuse or dissociate spontaneously from the 15-bp DNA as would happen in experimental conditions. By measuring the ratio between the dissociation kinetic rates from specific and nonspecific DNA, we obtained $\Delta\Delta G_b \sim 4.2 k_BT$, highly consistent with the experimental result. These findings suggest that flanking DNA sequences have a significant impact on measuring protein-DNA dissociation kinetics or binding affinity. The protein can bind to both the core binding site and the flanking sites, leading to a variety of dissociation rates. As a result, the dissociation kinetics are influenced by all possible dissociation events from both the central and flanking DNA sites, weighted by their respective populations. Since the dissociation rates of the flanking sequences are larger than that of the central binding site with specific DNA sequences, the overall approximation or averaged dissociation rate will be larger than that of the central binding site alone. It should be noted that, in the current kinetic analyses, we have not yet considered the sequence dependence of protein-DNA association kinetics. Considering the significant role of protein-DNA electrostatic interactions in TF binding (76,77), certain sequence effects in protein-DNA associations are still possible and may further impact on measurements of protein-DNA dissociation constants. Accordingly, the relative binding free energy between the nonspecific and specific DNA constructs was measured to be lower than that of the central site alone in the absence of protein binding to the flanking DNA sequences. In other words, $\Delta\Delta G_b^{BB'} < \Delta\Delta G_b^{bb'}$ (as seen from Figs. 5, *e* and *f*), which can predict that the experimentally measured binding free energy difference is generally smaller than that determined from the pulling simulations, where the protein dissociation events are confined within the core binding sites. It is important to note that both in vitro and in silico analyses have demonstrated that DNA sequences flanking the DNA core motifs can influence the affinity of TFs to the DNA motif by altering the 3D structure and flexibility of the DNA (78,79,80,81). However, to our knowledge, no physical models have been developed to elucidate the impact of TF binding or diffusion to flanking sites on the measurement of protein-DNA binding affinities. Further experimental studies with designs of the protein binding to various lengths of DNA constructs are expected, both to validate the current model and to systematically elucidate the impact of flanking sequences on determining the protein-DNA binding affinities. Furthermore, quantitative investigations are needed to determine and understand the influence of DNA sequences on protein association kinetics and binding dynamics. Notably, such influences or effects have been reported in a recent work (82). Consequently, our current studies can be extended to include additional analyses on protein binding and rebinding effects on DNA.

SUPPORTING MATERIAL

Supporting material can be found online at https://doi.org/10.1016/j.bpj.2023.10.025.

AUTHOR CONTRIBUTIONS

J.Y., C.M., and B.W. designed the research. C.M. carried out the all-atom simulations with the corresponding data analyses. B.W. carried out the spherical protein model simulations and the corresponding data analyses. C.M. and J.Y. wrote the manuscript.

ACKNOWLEDGMENTS

J.Y. has been supported by the Center for Multiscale Cell Fate Research of UCI (University of California, Irvine) via NSF (CMCF) Division of Mathematical Sciences Grant 1763272, Simons Foundation Grant 594598, and start-up funding from UCI. C.M. has been supported by CMCF fellowship via the NSF grant (DMS1763272), from the Simons Foundation grant (594598), and from the UC Research Initiatives - Cancer Research Coordinating Committee (UC-CRCC) (C23CR5636).

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Lambert, S. A., A. Jolma, …, M. T. Weirauch. 2018. The Human Transcription Factors. *Cell.* 172:650–665.
2. Wang, G., F. Wang, …, Y. Wang. 2015. Understanding Transcription Factor Regulation by Integrating Gene Expression and DNase I Hypersensitive Sites. *BioMed Res. Int.* 2015, 757530.
3. Riggs, A. D., S. Bourgeois, and M. Cohn. 1970. The lac represser-operator interaction: III. Kinetic studies. *J. Mol. Biol.* 53:401–417.
4. Berg, O. G., R. B. Winter, and P. H. Von Hippel. 1981. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry.* 20:6929–6948.
5. Dror, I., R. Rohs, and Y. Mandel-Gutfreund. 2016. How motif environment influences transcription factor search dynamics: Finding a needle in a haystack. *Bioessays.* 38:605–612.
6. von Hippel, P. H., and O. G. Berg. 1989. Facilitated Target Location in Biological Systems. *J. Biol. Chem.* 264:675–678.
7. Halford, S. E., and J. F. Marko. 2004. How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* 32:3040–3052.

8. Klenin, K. V., H. Merlitz, …, C.-X. Wu. 2006. Facilitated Diffusion of DNA-Binding Proteins. *Phys. Rev. Lett.* 96, 018104.
9. Cherstvy, A. G., A. B. Kolomeisky, and A. A. Kornyshev. 2008. Protein DNA Interactions: Reaching and Recognizing the Targets. *J. Phys. Chem. B.* 112:4741–4750.
10. Halford, S. E. 2009. An end to 40 years of mistakes in DNA–protein association kinetics? *Biochem. Soc. Trans.* 37:343–348.
11. Kolomeisky, A. B. 2011. Physics of protein–DNA interactions: mechanisms of facilitated target search. *Phys. Chem. Chem. Phys.* 13:2088–2095.
12. Bauer, M., and R. Metzler. 2012. Generalized facilitated diffusion model for DNA-binding proteins with search and recognition states. *Biophys. J.* 102:2321–2330.
13. Sprague, B. L., R. L. Pego, …, J. G. McNally. 2004. Analysis of binding reactions by fluorescence recovery after photobleaching. *Biophys. J.* 86:3473–3495.
14. Gowers, D. M., G. G. Wilson, and S. E. Halford. 2005. Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA. *Proc. Natl. Acad. Sci. USA.* 102:15883–15888.
15. Blainey, P. C., A. M. van Oijen, …, X. S. Xie. 2006. A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proc. Natl. Acad. Sci. USA.* 103:5752–5757.
16. Wang, Y. M., R. H. Austin, and E. C. Cox. 2006. Single Molecule Measurements of Repressor Protein 1D Diffusion on DNA. *Phys. Rev. Lett.* 97, 048302.
17. Elf, J., G.-W. Li, and X. S. Xie. 2007. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science.* 316:1191–1194.
18. Bonnet, I., A. Biebricher, …, P. Desbiolles. 2008. Sliding and jumping of single EcoRV restriction enzymes on non-cognate DNA. *Nucleic Acids Res.* 36:4118–4127.
19. van den Broek, B., M. A. Lomholt, …, G. J. L. Wuite. 2008. How DNA coiling enhances target localization by proteins. *Proc. Natl. Acad. Sci. USA.* 105:15738–15742.
20. Hammar, P., P. Leroy, …, J. Elf. 2012. The lac repressor displays facilitated diffusion in living cells. *Science.* 336:1595–1598.
21. Monico, C., M. Capitanio, …, F. S. Pavone. 2013. Optical Methods to Study Protein-DNA Interactions in Vitro and in Living Cells at the Single-Molecule Level. *Int. J. Mol. Sci.* 14:3961–3992.
22. Chen, J., Z. Zhang, …, Z. Liu. 2014. Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell.* 156:1274–1285.
23. Mahmutovic, A., O. G. Berg, and J. Elf. 2015. What matters for lac repressor search in vivo—sliding, hopping, intersegment transfer, crowding on DNA or recognition? *Nucleic Acids Res.* 43:3454–3464.
24. Kamagata, K., A. Murata, …, S. Takahashi. 2017. Characterization of facilitated diffusion of tumor suppressor p53 along DNA using single-molecule fluorescence imaging. *J. Photochem. Photobiol. C Photochem. Rev.* 30:36–50, Special issue on Bioimaging.
25. Liu, Z., and R. Tjian. 2018. Visualizing transcription factor dynamics in living cells. *J. Cell Biol.* 217:1181–1191.
26. Givaty, O., and Y. Levy. 2009. Protein sliding along DNA: dynamics and structural. *J. Mol. Biol.* 385:1087–1097.
27. Das, R. K., and A. B. Kolomeisky. 2010. Facilitated search of proteins on DNA: correlations are important. *Phys. Chem. Chem. Phys.* 12:2999–3004.
28. Koslover, E. F., M. A. Díaz de la Rosa, and A. J. Spakowitz. 2011. Theoretical and Computational Modeling of Target-Site Search Kinetics In Vitro and In Vivo. *Biophys. J.* 101:856–865.
29. Marcovitz, A., and Y. Levy. 2012. Sliding dynamics along DNA: a molecular perspective. *Innovations in Biomolecular Modeling and Simulations.* 2:236–262.
30. Bauer, M., and R. Metzler. 2013. In vivo facilitated diffusion model. *PLoS One.* 8, e53956.
31. Bauer, M., E. S. Rasmussen, …, R. Metzler. 2015. Real sequence effects on the search dynamics of transcription factors on DNA. *Sci. Rep.* 5, 10072.
32. Lange, M., M. Kochugaeva, and A. B. Kolomeisky. 2015. Protein search for multiple targets on DNA. *J. Chem. Phys.* 143, 105102.
33. Slutsky, M., and L. A. Mirny. 2004. Kinetics of Protein-DNA Interaction: Facilitated Target Location in Sequence-Dependent Potential. *Biophys. J.* 87:4021–4035.
34. Slutsky, M., M. Kardar, and L. A. Mirny. 2004. Diffusion in correlated random potentials, with applications to DNA. *Phys. Rev. E - Stat. Nonlinear Soft Matter Phys.* 69, 061903.
35. Zhou, H.-X. 2011. Rapid search for specific sites on DNA through conformational switch of nonspecifically bound proteins. *Proc. Natl. Acad. Sci. USA.* 108:8651–8656.
36. Marcovitz, A., and Y. Levy. 2011. Frustration in protein–DNA binding influences conformational switching and target search kinetics. *Proc. Natl. Acad. Sci. USA.* 108:17957–17962.
37. Ferreiro, D. U., E. A. Komives, and P. G. Wolynes. 2014. Frustration in biomolecules. *Q. Rev. Biophys.* 47:285–363.
38. Leven, I., and Y. Levy. 2019. Quantifying the two-state facilitated diffusion model of protein–DNA interactions. *Nucleic Acids Res.* 47:5530–5538.
39. Melero, R., S. Rajagopalan, …, M. Valle. 2011. Electron microscopy studies on the quaternary structure of p53 reveal different binding modes for p53 tetramers in complex with DNA. *Proc. Natl. Acad. Sci. USA.* 108:557–562.
40. Leith, J. S., A. Tafvizi, …, A. M. van Oijen. 2012. Sequence-dependent sliding kinetics of p53. *Proc. Natl. Acad. Sci. USA.* 109:16552–16557.
41. Kalodimos, C. G., A. M. J. J. Bonvin, …, R. Kaptein. 2002. Plasticity in protein–DNA recognition: lac repressor interacts with its natural operator O1 through alternative conformations of its DNA-binding domain. *EMBO J.* 21:2866–2876.
42. Kalodimos, C. G., N. Biris, …, R. Kaptein. 2004. Structure and Flexibility Adaptation in Nonspecific and Specific Protein-DNA Complexes. *Science.* 305:386–389.
43. Kalodimos, C. G., R. Boelens, and R. Kaptein. 2004. Toward an Integrated Model of Protein-DNA Recognition as Inferred from NMR Studies on the Lac Repressor System. *Chem. Rev.* 104:3567–3586.
44. Hauser, K., B. Essuman, …, C. Simmerling. 2016. A human transcription factor in search mode. *Nucleic Acids Res.* 44:63–74. https://academic.oup.com/nar/article-pdf/44/1/63/25346240/gkv1091.pdf.
45. Marklund, E. G., A. Mahmutovic, …, J. Elf. 2013. Transcription-factor binding and sliding on DNA studied using micro- and macroscopic models. *Proc. Natl. Acad. Sci. USA.* 110:19796–19801.
46. Furini, S., P. Barbini, and C. Domene. 2013. DNA-recognition process described by MD simulations of the lactose repressor protein on a specific and a non-specific DNA sequence. *Nucleic Acids Res.* 41:3963–3972.
47. Bigman, L. S., H. M. Greenblatt, and Y. Levy. 2021. What Are the Molecular Requirements for Protein Sliding along DNA? *J. Phys. Chem. B.* 125:3119–3131.
48. Sun, L., M. Tabaka, …, R. Holyst. 2016. The Hinge Region Strengthens the Nonspecific Interaction between Lac-Repressor and DNA: A Computer Simulation Study. *PLoS One.* 11:e0152002–e0152012.
49. Liao, Q., M. Lüking, …, S. C. Lynn Kamerlin. 2019. Long Time-Scale Atomistic Simulations of the Structure and Dynamics of Transcription Factor-DNA Recognition. *J. Phys. Chem. B.* 123:3576–3590.
50. Yonetani, Y., and H. Kono. 2013. Dissociation Free-Energy Profiles of Specific and Nonspecific DNA–Protein Complexes. *J. Phys. Chem. B.* 117:7535–7545.
51. Terakawa, T., and S. Takada. 2015. p53 dynamics upon response element recognition explored by molecular simulations. *Sci. Rep.* 5, 17107.

52. Chu, X., and V. Muñoz. 2017. Roles of conformational disorder and downhill folding in modulating protein–DNA recognition. *Phys. Chem. Chem. Phys.* 19:28527–28539.

53. Machado, M. R., and S. Pantano. 2015. Exploring LacI–DNA Dynamics by Multiscale Simulations Using the SIRAH Force Field. *J. Chem. Theor. Comput.* 11:5012–5023.

54. Tan, C., T. Terakawa, and S. Takada. 2016. Dynamic Coupling among Protein Binding, Sliding, and DNA Bending Revealed by Molecular Dynamics. *J. Am. Chem. Soc.* 138:8512–8522.

55. Dai, L., Y. Xu, …, J. Yu. 2021. Revealing atomic-scale molecular diffusion of a plant-transcription factor WRKY domain protein along DNA. *Proc. Natl. Acad. Sci. USA.* 118, e2102621118.

56. Xu, Y.-p., H. Xu, …, X.-D. Su. 2020. Crystal structures of N-terminal WRKY transcription factors and DNA complexes. *Protein Cell.* 11:208–213.

57. Jarzynski, C. 1997. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* 78:2690–2693.

58. Torrie, G. M., and J. P. Valleau. 1974. Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. *Chem. Phys. Lett.* 28:578–581.

59. Liao, Q. 2020. Chapter Four - Enhanced sampling and free energy calculations for protein simulations. *In* Computational Approaches for Understanding Dynamical Systems: Protein Folding and Assembly. B. Strodel and B. Barz, eds Academic Press, pp. 177–213, Volume 170 of Progress in Molecular Biology and Translational Science.

60. Kästner, J. 2011. Umbrella sampling. *WIREs Comput. Mol. Sci.* 1:932–942.

61. Ho, K., D. T. Truong, and M. S. Li. 2020. How Good is Jarzynski's Equality for Computer-Aided Drug Design? *J. Phys. Chem. B.* 124:5338–5349.

62. Gumbart, J. C., B. Roux, and C. Chipot. 2013. Standard Binding Free Energies from Computer Simulations: What Is the Best Strategy? *J. Chem. Theor. Comput.* 9:794–802.

63. Chen, P.-C., and S. Kuyucak. 2011. Accurate determination of the binding free energy for KcsA-charybdotoxin complex from the potential of mean force calculations with restraints. *Biophys. J.* 100:2466–2474.

64. Lindahl, A.H., and van der Spoel, 2020. GROMACS 2020.4 Manual.

65. Lindorff-Larsen, K., S. Piana, …, D. E. Shaw. 2010. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins.* 78:1950–1958.

66. Ivani, I., P. D. Dans, …, M. Orozco. 2016. Parmbsc1: a refined force field for DNA simulations. *Nat. Methods.* 13:55–58.

67. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: An N.log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.

68. Kokh, D. B., B. Doser, …, R. C. Wade. 2020. A workflow for exploring ligand dissociation from a macromolecule: Efficient random acceleration molecular dynamics simulation and interaction fingerprint analysis of ligand trajectories. *J. Chem. Phys.* 153, 125102.

69. Izrailev, S., S. Stepaniants, …, K. Schulten. 1999. Steered molecular dynamics. *In* Computational Molecular Dynamics: Challenges, Methods, Ideas: Proceedings of the 2nd International Symposium on Algorithms for Macromolecular Modelling Springer, pp. 39–65, May 21–24, 1997.

70. Kumar, S., J. M. Rosenberg, …, P. A. Kollman. 1992. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* 13:1011–1021.

71. Hub, J. S., B. L. de Groot, and D. van der Spoel. 2010. g-wham-A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Theor. Comput.* 6:3713–3720.

72. Kelsey, D. E., T. C. Rounds, and S. S. York. 1979. lac repressor changes conformation upon binding to poly[dA-T)]. *Proc. Natl. Acad. Sci. USA.* 76:2649–2653.

73. Culard, F., and J. C. Maurizot. 1982. Binding of lac repressor induces different conformational changes on operator and non-operator DNAs. *FEBS Lett.* 146:153–156.

74. Taraban, M., H. Zhan, …, J. Trewhella. 2008. Ligand-induced Conformational Changes and Conformational Dynamics in the Solution Structure of the Lactose Repressor Protein. *J. Mol. Biol.* 376:466–481.

75. McLure, K. G., and P. W. Lee. 1999. p53 DNA binding can be modulated by factors that alter the conformational equilibrium. *EMBO J.* 18:763–770.

76. Pang, X., and H.-X. Zhou. 2017. Rate Constants and Mechanisms of Protein–Ligand Binding. *Annu. Rev. Biophys.* 46:105–130.

77. Chen, X., M.-Y. Tsai, and P. G. Wolynes. 2022. The Role of Charge Density Coupled DNA Bending in Transcription Factor Sequence Binding Specificity: A Generic Mechanism for Indirect Readout. *J. Am. Chem. Soc.* 144:1835–1845.

78. Gordân, R., N. Shen, …, M. L. Bulyk. 2013. Genomic Regions Flanking E-Box Binding Sites Influence DNA Binding Specificity of bHLH Transcription Factors through DNA Shape. *Cell Rep.* 3:1093–1104.

79. Schöne, S., M. Jurk, …, S. H. Meijsing. 2016. Sequences flanking the core-binding site modulate glucocorticoid receptor structure and activity. *Nat. Commun.* 7, 12621.

80. Yella, V. R., D. Bhimsaria, …, M. Bansal. 2018. Flexibility and structure of flanking DNA impact transcription factor affinity for its core motif. *Nucleic Acids Res.* 46:11883–11897.

81. Hörberg, J., K. Moreau, …, A. Reymer. 2021. Sequence-specific dynamics of DNA response elements and their flanking sites regulate the recognition by AP-1 transcription factors. *Nucleic Acids Res.* 49:9280–9293.

82. Marklund, E., G. Mao, …, J. Elf. 2022. Sequence specificity in DNA binding is mainly governed by association. *Science.* 375:442–445.