Title

Revealing atomic-scale molecular diffusion of a plant-transcription factor WRKY domain protein along DNA

Permalink

https://escholarship.org/uc/item/0hr783hn

Journal

Proceedings of the National Academy of Sciences of the United States of America, 118(23)

ISSN

0027-8424

Authors

Dai, Liqiang
Xu, Yongping
Du, Zhenwei
et al.

Publication Date

2021-06-08

DOI

10.1073/pnas.2102621118

Peer reviewed

Revealing atomic-scale molecular diffusion of a plant-transcription factor WRKY domain protein along DNA

Liqiang Dai[a,b,1] , Yongping Xu[c,d,1], Zhenwei Du[c,d], Xiao-dong Su[c,d,2], and Jin Yu[e,f,g,2]

[a]Shenzhen JL Computational Science and Applied Research Institute, Shenzhen 518131, China; [b]Beijing Computational Science Research Center, Beijing 100193, China; [c]State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing 100871, China; [d]Biomedical Pioneering Innovation Center, School of Life Sciences, Peking University, Beijing 100871, China; [e]Department of Physics and Astronomy, University of California, Irvine, CA 92697; [f]Department of Chemistry, University of California, Irvine, CA 92697; and [g]NSF-Simons Center for Multiscale Cell Fate Research, University of California, Irvine, CA 92697

Transcription factor (TF) target search on genome is highly essential for gene expression and regulation. High-resolution determination of TF diffusion along DNA remains technically challenging. Here, we constructed a TF model system using the plant WRKY domain protein in complex with DNA from crystallography and demonstrated microsecond diffusion dynamics of WRKY on DNA by employing all-atom molecular-dynamics (MD) simulations. Notably, we found that WRKY preferentially binds to one strand of DNA with significant energetic bias compared with the other, or nonpreferred strand. The preferential DNA-strand binding becomes most prominent in the static process, from nonspecific to specific DNA binding, but less distinct during diffusive movements of the domain protein on the DNA. Remarkably, without employing acceleration forces or bias, we captured a complete one-base-pair stepping cycle of the protein tracking along major groove of DNA with a homogeneous poly-adenosine sequence, as individual hydrogen bonds break and reform at the protein–DNA binding interface. Further DNA-groove tracking motions of the protein forward or backward, with occasional sliding as well as strand crossing to minor groove of DNA, were also captured. The processive diffusion of WRKY along DNA has been further sampled via coarse-grained MD simulations. The study thus provides structural dynamics details on diffusion of a small TF domain protein, suggests how the protein approaches a specific recognition site on DNA, and supports further high-precision experimental detection. The stochastic movements revealed in the TF diffusion also provide general clues about how other protein walkers step and slide along DNA.

facilitated diffusion | molecular dynamics | specific recognition | WRKY | transcription factor

The search and recognition processes of transcription factors (TFs) on DNA are of fundamental importance in gene expression and regulation. To locate a target gene site sufficiently fast on a genome that is wrapped within three-dimensional (3D) space, the TFs may proceed with a facilitated diffusion process, alternating between one-dimensional (1D) diffusional movements along DNA and 3D intracellular diffusion, accompanied by occasional jumping, hopping, and intersegment transfer (1–6). Experimental detection on protein-searching motions or 1D diffusion along DNA have provided evidence on facilitated diffusion (7–12). Nevertheless, as TF protein movements of base pair (bp) distances on DNA can take place as fast as microseconds, tracking the 1D TF diffusion at such a high temporal and spatial resolution remains technically challenging (13–17).

On the other hand, high-resolution determinations of protein–DNA complex structures (18) allow one to investigate corresponding conformational dynamics by employing all-atom molecular-dynamics (MD) simulations via high-performance computing (19–21). Protein recognition on specific DNA has been actively examined in recent years using MD technologies (22–26). In comparison, protein association with nonspecific DNA has been less examined. It is commonly expected that nonspecific association and movements of protein on the DNA happen slowly for the timescale of the simulations and cannot be well sampled via the atomistic MD. Indeed, either comparatively short MD simulations (nano- or submicroseconds) were conducted (22), or external forces were added to accelerate the protein movements or enhance samplings, such as employing targeted MD or umbrella sampling simulations (14, 24, 27, 28). In cases in which comparatively long or extensive MD simulations have been conducted, one recent study concentrates on association processes of a chromatin protein with DNA (29) but not the protein movements. For exemplary all-atom simulation studies on the protein movements along DNA, however, the proteins of concern have been motor proteins such as RNA polymerases (30, 31) or the single-stranded DNA-binding protein (32) and DNA-repair proteins (33, 34). In this work, we focus on a model TF and present mainly unbiased all-atom microseconds equilibrium simulations of the diffusion dynamics of the TF protein along the double-stranded (ds) DNA, with simulation samplings accumulated over 100 microseconds. The protein

Significance

In transcription factors' search for target genes, one-dimensional diffusion of the protein along DNA is essential. Experimentally, it remains challenging to resolve the individual diffusional steps of protein on DNA. Here, we report mainly all-atom equilibrium simulations of a WRKY domain protein in association with and diffusion along DNA. We demonstrate a complete stepping cycle of the protein for one base pair on DNA within microseconds, along with stochastic motions. Processive protein diffusions on DNA have been further sampled in a coarse-grained model. We have also found preferential DNA-strand association of the domain protein, which becomes most prominent at specific DNA binding, and it can be common for small-domain proteins to balance movements on the DNA with the sequence recognition.

factor under our current investigation is a WRKY domain protein from *Arabidopsis thaliana* WRKY1.

WRKY proteins are a large family of TFs in plants playing a broad range of important roles for signal response, stress control, and disease resistance (35, 36). The number of WRKY family members in *Arabidopsis* reaches over 70, and all of them include a DNA-binding domain of about 60 amino acids that is called the WRKY domain. The WRKY domain proteins are featured by a highly conserved "WRKYGQK" region and a zinc-finger motif, both of which turn out to be indispensable for maintaining the DNA-binding function. Previously, an *apo* carboxyl-terminal domain structure of *Arabidopsis* WRKY1 had been made available (37). Recently, a high-resolution crystal structure of the N-terminal WRKY domain protein in complex with a specific DNA-binding sequence has been obtained (38). Based on this structure, we performed atomistic MD simulations on the protein–DNA complexes (with a 34-bp double-stranded DNA [dsDNA]) in explicit solvent conditions, constructed for both specific and nonspecific DNA-binding systems. We identified strong and biased association of WRKY with one strand of DNA [the preferred strand referred to as the Crick strand in the crystal structure (38)] and comparatively weak association with the other strand (the nonpreferred strand), and such preferential strand association demonstrates most prominently in the static and specific protein–DNA binding. Notably, our simulations revealed 1-bp cyclic stepping motions of the domain protein with a full set of hydrogen bonds (HBs) breaking and reforming at the protein–DNA backbone interface spontaneously, as the protein tracks along the DNA groove and frequently adjusts its orientations to align with the helical groove. Moreover, the simulations also captured events of protein sliding stochastically on the DNA (e.g., attempts at a larger step size [>1 bp], directional reversal, and moving across the DNA strand). The processive diffusion of the WRKY domain protein along DNA has been further sampled by coarse-grained (CG) MD simulations conducted at various ionic concentrations and along different DNA sequences. Accompanying single-molecule fluorescence measurements confirmed the WRKY 1D processive diffusion along DNA.

Results

Specific versus Nonspecific DNA Association of WRKY with Varied Stabilities.
We conducted microseconds equilibrium MD simulations on the WRKY–DNA complexes with a specific DNA-binding sequence (W-box) and a slightly varied but nonspecific DNA sequence, respectively. The specific protein–DNA complex had been constructed directly from the crystal structure (38) with DNA extended to 34 bp (*SI Appendix, Methods*). The nonspecific protein–DNA binding complex was modeled from the crystal structure by converting the specific core sequence of DNA (5′-CTGGTCAAAG-3′ on the preferred strand) to the slightly varied nonspecific sequence (5′-CTGATAAAAG-3′) and conducting equilibrium simulation. Using isothermal titration calorimetry (ITC), we also determined the WRKY dissociation constants with DNA on the above specific and nonspecific sequences as $K_D = 0.1 \mu M$ and $8 \mu M$, respectively (*SI Appendix, Methods* and Fig. S1).

By conducting and comparing two 10-μs MD simulations of WRKY (modeled at an ionic concentration of 150 mM) on the specific and nonspecific sequences (Fig. 1), one notices well localization of WRKY on the DNA around the specific sequence with comparatively small longitudinal ($\Delta X \sim 1.2 \pm 0.8$Å) and rotational movements ($\Delta \Theta \sim 9.8 \pm 7.4°$) of the protein center of mass (COM) after ~2 μs preequilibration. In comparison, WRKY modeled on the nonspecific DNA demonstrated more significant rotation-coupled relaxation on the DNA, with the protein COM shifted both longitudinally and rotationally ($\Delta X \sim 3.3 \pm 0.8$ Å and $\Delta \Theta \sim 30.0 \pm 9.8°$). Note, however, that the protein did not yet translocate along the DNA, since the HBs formed at the

protein–DNA interface on the nonspecific DNA sequence are still well maintained in the full simulation (e.g., Fig. 2).

The protein–DNA structural alignments according to the nearby/bound DNA segment around the protein (~10 bp; *SI Appendix*, Fig. S2A) show that the conformational rearrangements of WRKY on the nonspecific DNA are significant ($\Delta RMSD \sim 8$ Å; *SI Appendix,* Fig. S2B) and are substantially larger than those on the specific DNA ($\Delta RMSD \sim 3$ Å). Meanwhile, structural alignments according to protein core structure demonstrate almost no conformational changes of the domain protein on the nonspecific DNA versus on the specific DNA (*SI Appendix*, Fig. S2C). In addition, by measuring protein orientational angle (following the beta strands) with respect to the DNA long axis, one can see that the protein rotates from ~62 ± 5° on the specific DNA to ~81 ± 9° on the nonspecific DNA (*SI Appendix*, Fig. S2D). Hence, the rearrangements of the protein on the nonspecific DNA mainly come from the orientational changes of the protein. Besides, the nonspecific DNA in association with the protein also shows slightly larger fluctuations in its major groove size than that of the specific DNA, with detectable correlations between the groove-size variation and the protein orientation on the DNA (*SI Appendix*, Fig. S2E). Two movies are provided for viewing the WRKY relaxations on the specific and nonspecific DNA (Movies S1 and S2), respectively.

Additionally, we constructed a mutant (mt) WRKY-K122A with a lowered affinity with specific DNA as measured by the ITC experiments $K_D \sim 1 \mu M$ (*SI Appendix*, Fig. S1). Correspondingly, we performed MD simulation for this mt-WRKY modeled on the original specific DNA sequence. The results show that the mt-WRKY starts rearranging along DNA similarly to the nonspecific wild-type (wt) binding complex (Fig. 1). The positional relaxation of the mt-WRKY on DNA shows intermediate behaviors in between the specific and the nonspecific DNA-binding complexes.

WRKY Association with DNA Is Strongly Biased onto One Strand and the Bias Is Strongly Maintained in the Specific DNA-Binding Complex.
By close examinations, we identified detailed interactions at the protein–DNA interface for both the specific and the nonspecific binding systems (Fig. 2A). The schematics summarizing the HB and salt-bridge interactions between WRKY and DNA strands are provided (*SI Appendix*, Fig. S3A). In particular, we found substantial HB interactions between the protein and the preferred DNA strand (~7 to15 HBs) for both specific and nonspecific cases. In the specific binding (Fig. 2 *B, Left* and *SI Appendix*, Fig. S3A), HBs are formed via K125 to G15; K122, R131, Y133, and Q146 to G16; Y119, K144, and Q146 (water mediated) to T17; and Y119 and R135 to C18. Among them, arginine and lysine (R131, R135, and K144) can also form salt-bridge interactions with the phosphate groups on the DNA; Y119 specifically forms an HB with the C18 base, while K122 also specifically forms an HB with the G16 base. In contrast, there are much fewer HBs formed between WRKY and the other or nonpreferred DNA strand (~2 to 5 HBs). It mainly involves HB and salt-bridge interactions between R117/K118 (on the β2 strand) and the backbone DNA and three water-mediated HBs (*SI Appendix*, Fig. S3A). Note that W116, R117, K118, Y119, Q121 and K122 are from the WRKYGQK motif. Except for Y119 and K122 associating specifically with the two core DNA bases on the preferred strand, the rest interact with the DNA backbone of the nonpreferred strand.

In comparison, one sees the HB association of WRKY with the nonspecific DNA (Fig. 2 *B, Middle* and *SI Appendix*, Fig. S3B): K125 to G15; R131, Y133, and Q146 to A16; Y119 and K144 to T17; and R135 to A18 on the preferred strand; G153, Q154, and R149 (on the loop connecting β4 and β5) forming HB and salt-bridge interactions to the DNA backbone; and Q121 forming HB with the nonpreferred strand.
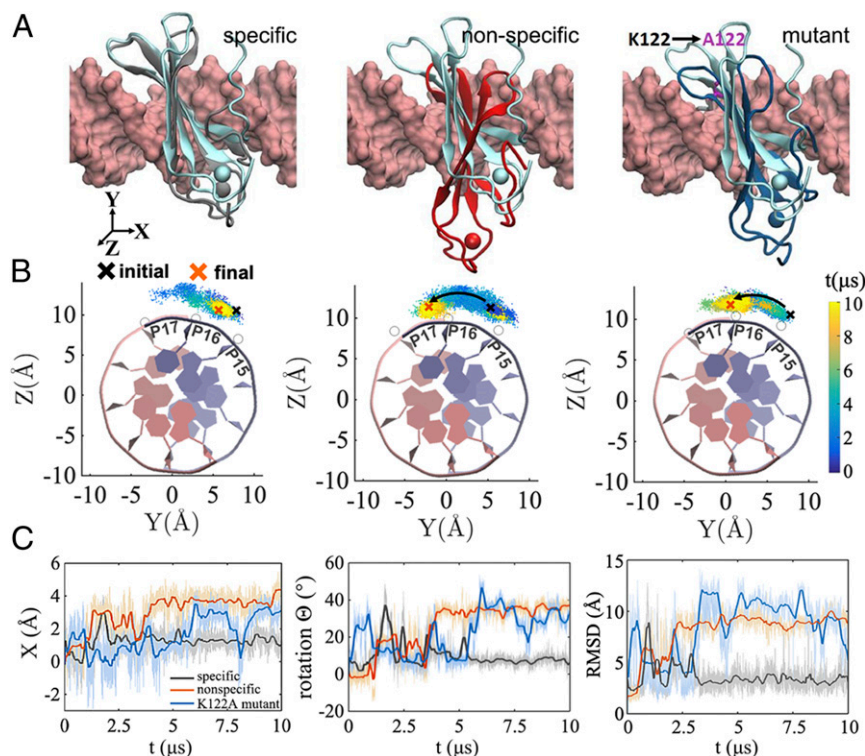
Fig. 1. Specific and nonspecific DNA association of WRKY. (*A*) Comparisons of the initial (cyan) and final (gray, red, blue) structures of the simulation of the wild-type (wt) protein binding on the specific DNA (*Left*), the nonspecific DNA (*Middle*), and the mutant (mt) protein (K122A) binding on the specific DNA (*Right*). The *x-y-z* axis is denoted, and the longitudinal axis of the DNA follows the *x* direction. (*B*) The rotational relaxation of the center of mass (COM) of the protein along DNA projected onto the *y-z* plane. The initial and final positioning (due to the protein relaxation on the DNA) are denoted. The time evolution is represented by coloring (from blue to yellow). The DNA structure is shown for reference. (*C*) The relaxation of the protein COM along X and Θ and the protein–DNA RMSDs, showing the protein rearrangements on the DNA (*SI Appendix*, Fig. S2) for respective simulation systems (wt specific, dark curves; nonspecific, orange; and K122A mt, blue). Note that the heavy lines are smoothed from the original time series [X(t), Θ(t), and RMSD(t)] over a sliding window of ~100 ns, and a similar data-smooth procedure is conducted for other plots.

Consistently, the base-specific K122 and Y119 associations with the altered core sequence become absent in the nonspecific complex. No water-mediated HBs are identified at the protein–DNA interface for the nonspecific DNA binding nor for the mt protein case (*SI Appendix*, Fig. S3C). In the mt, K122A loses contact with the specific core-DNA sequence, while most other HBs with the preferred DNA strand are preserved.

The preferred-strand HB associations also demonstrate larger fluctuations on the nonspecific DNA or for the mt protein than in the original specific DNA binding (Fig. 2C). Furthermore, WRKY associates with the nonpreferred DNA strand via the β2 strand on the specific DNA, while it has β4 and 5 strands to associate with the nonpreferred strand on the nonspecific DNA (Fig. 2B). Such alteration of the protein–DNA binding interface happens together with significant reorientation of the protein on the nonspecific DNA. In the mt, due to loss of the specific HB contact from K122, the protein also reorients on the DNA and associates with the nonpreferred DNA strand involving β4 and 5 strands.

The hydrophobic interactions between the protein and DNA have been monitored as well (*SI Appendix*, Fig. S4). The involved hydrophobic contacts with the preferred stand also appear more than those with the nonpreferred strand in both specific and nonspecific (or mt) cases. In addition, we counted water molecules around the surface of the protein or the protein–DNA interface (within 5 Å). For about the same amount of waters (slightly above 300 waters) surrounding WRKY in the specific and nonspecific DNA system, fewer waters stay close to the protein–DNA interface on the specific DNA (~36 ± 6 excluding

the waters mediating HBs) than on the nonspecific DNA (~47 ± 8). Thus, the hydrophobic interactions at the protein–DNA interface also favor the protein–DNA specific binding and bias toward the preferred DNA strand.

Atomistic Simulation of WRKY Diffusion along Homogeneous Poly-A DNA with Rotation-Coupled Protein Motions Sampled. In the above 10-μs simulation of WRKY binding on the nonspecific DNA (see Figs. 1 and 2), we have not yet detected diffusion of the protein. In order to probe essential protein translocation or displacements of protein contacts along DNA, we modeled WRKY on homogeneous poly-adenosine (poly-A) dsDNA at a length of 34 bp. It was expected that protein contacts made on homogeneous DNA sequences could be synchronized more easily to support comparatively fast protein translocation. On such poly-A DNA, we accordingly captured one complete stepping cycle of the WRKY diffusion (i.e., for 1-bp distance on the DNA) via equilibrium atomistic simulation (Fig. 3). We analyze both the COM movements of the protein along DNA and collective HB dynamics at the protein–DNA interface.

In Fig. 3*A*, representative snapshots from two MD trajectories are shown, demonstrating WRKY moving forward (+X direction or toward *Right*) and backward (toward *Left*) along DNA, respectively. The longitudinal (along X) and rotational motions (mapped on the Y–Z plane) of the protein COM along DNA are demonstrated in Fig. 3 *B* and *C*. In the forward direction, mainly four spatial states reveal for the following (labeled 1 through 4 with the initial preequilibrated state 0) according to helical motions of the protein COM on the DNA (Fig. 3*B*): In the first ~1.86 μs, WRKY tracks slightly forward along the major groove
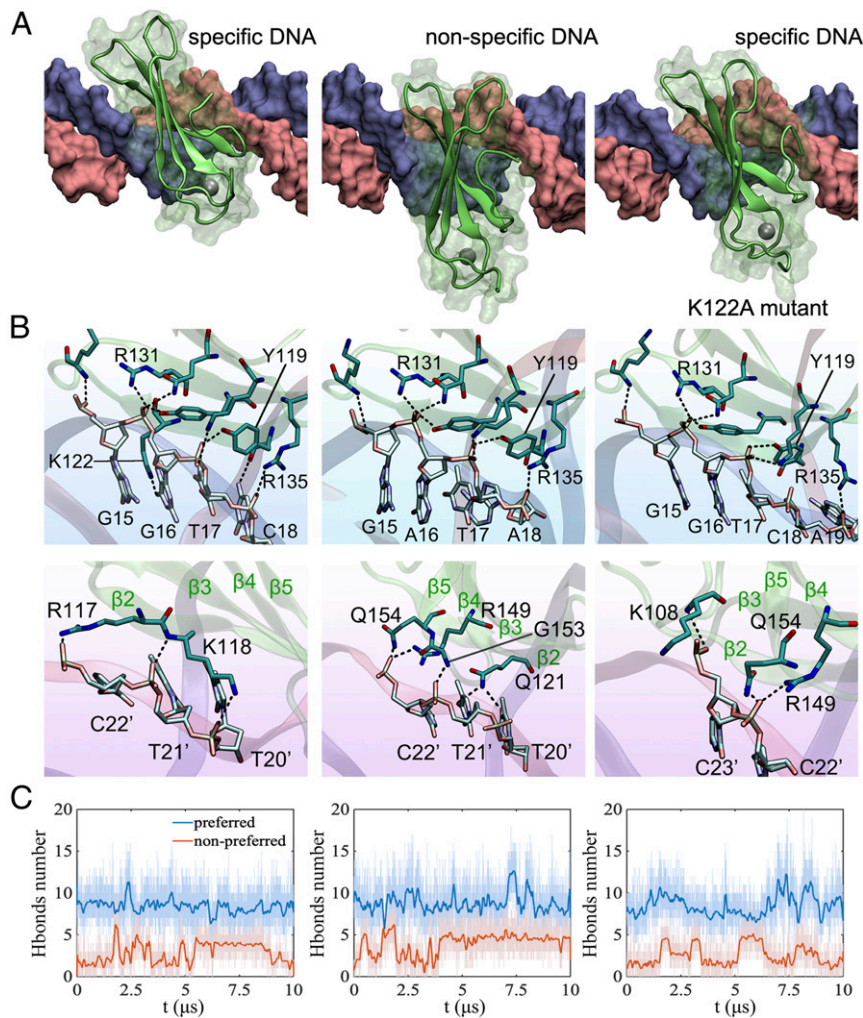
Dai et al.
Revealing atomic-scale molecular diffusion of a plant-transcription factor WRKY domain
protein along DNA

PNAS | 3 of 10
https://doi.org/10.1073/pnas.2102621118

Fig. 2. The association between WRKY and respective strands of DNA. (*A*) WRKY association on the specific DNA (core sequence: <u>GGTC</u>; *Left*), the nonspecific DNA (core: <u>GATA</u>; *Middle*), and the mt K122A on the specific DNA (*Right*). The equilibrated protein–DNA complexes are shown in surface representation with the protein in green and the DNA strands in blue (the preferred) or pink (the nonpreferred). (*B*) The HBs at the protein–DNA interface are shown on the preferred strand (*Top*) and the nonpreferred strand (*Bottom*). (*C*) Time-dependent HB statistics at the protein–DNA interface are provided on the respective DNA strands. The HBs are counted for those formed >50 ps within a sliding 1-ns simulation window.

of the DNA, closely following the preferred strand, moving from state 1 to 2 (1 → 2; Δx ~1.1 Å and ΔΘ ~21.9°); during 1.86 to 3.08 μs, however, it slightly retracks back to state 1 (2 → 1); at ~4.96 μs, the protein quickly steps forward, advancing about 1 bp within 0.2 μs (1 → 3; Δx ~1.9 Å and ΔΘ ~27.1°); after that (>7.5 μs), WRKY slides forward (3 → 4; Δx~ −0.9 Å and ΔΘ ~16.9°) but adjusts its spatial orientation on the DNA to better align with the major groove at the next location (see Movie S3 for the protein's 1-bp stepping). Comparing to the static DNA-binding case, there are still no conformational changes of the protein core during the forward movements. Nevertheless, the protein orientational changes on the DNA are substantial (the orientation angle spans from ~77 ± 10° in the first 5 μs to ~56 ± 10° in the last 5 μs; *SI Appendix*, Fig. S5) so that the domain protein can adjust and realign with the DNA helical track. Meanwhile, the DNA-groove size varies (between ~18 and 22 Å), and the variation correlates with the protein orientational change (*SI Appendix*, Fig. S5).

The protein diffusion captured in the backward direction (after a pre-equilibrated state 0) also starts with slight forward motions (0 → 1) within the first ~1.43 μs (Δx ~1.6 Å and ΔΘ ~26.0°), similar to that in the forward trajectory; then it is followed by

deviated tracking (1 → 2'; Δx ~0.7 Å and ΔΘ ~−16.9°) at ~1.43 μs and moving backward at ~3.9 μs (2' → 1'; Δx ~−2.3 Å and ΔΘ ~10.3°). There is a further sliding backward at ~5.18 μs (1' → 3' within 1.3 μs; Δx ~−0.8 Å and ΔΘ ~−36.2°). At ~7.02 μs, a strand-crossing event of WRKY on the DNA in the backward direction shows. Right after that, WRKY binds onto the minor groove of the DNA (Movie S4). A conformational change of the protein (ΔRMSD ~3 Å with broken HBs between strands β4 and β5) starts at ~1.2 μs right before the transition 1 → 2', and a new protein conformation is reached at ~4.1 μs (right upon moving backward into state 1', with new HBs formed between β5 and β2) and maintained thereafter. The protein orientational changes on the DNA (from 68 ± 12° in the first 5 μs to 46 ± 10° for the last 5 μs) are also significant and accompanied DNA-major groove-width variations show similarly (between 18 and 22 Å) to the forward trajectory.

To enhance samplings of the protein movements along the poly-A DNA, we additionally launched 10 comparatively short simulation runs (2 to 4 μs each; see individual mappings and accompanied protein–DNA interfacial HB dynamics in *SI Appendix*, Figs. S6 and S7, respectively) starting from various intermediate states along both the forward and backward diffusional
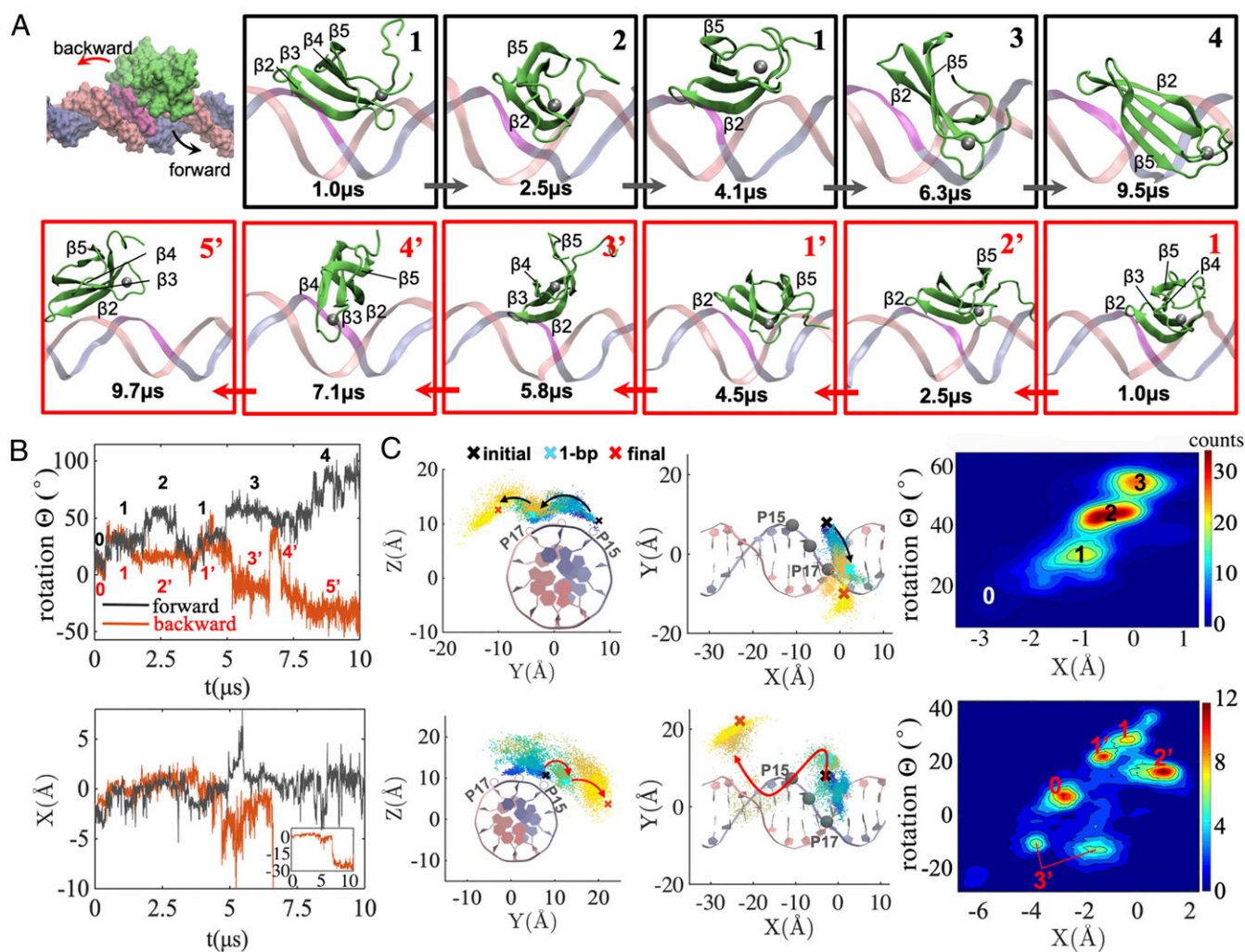
Fig. 3. The diffusion of WRKY along poly-A DNA in the forward and backward direction revealed from two 10-μs atomistic MD simulations. (A) The representative structural snapshots taken from the simulation trajectories forward (*Top*, from left to right, via state 1 → 2 → 1→ 3 → 4 according to the protein COM movements shown in B) and backward (*Bottom*, from right to left, via state 1 → 2' → 1'→ 3' → 4' →5', with primed labels to differentiate from the forward states as characterized in B), with the WRKY domain protein shown in green and two DNA strands in blue (the preferred strand) and pink (the nonpreferred strand). (B) The helical trajectories of the protein COM along the DNA, shown for the angular Θ(t) (*Top*) and the longitudinal movement X(t) (*Bottom*) from the simulation. The coordinate system is defined as in Fig. 1. Five (forward, dark line) and six (backward, orange line) states are identified along the angular coordinates. (C) The protein COM helical motions along DNA are mapped on the y-z plane (*Left*) and then on the x-y plane (*Middle*), colored by the simulation time (blue to yellow as in Fig. 1B). The dsDNA rendering is also shown for reference. The further sampled protein movements mapped on the X-Θ plane for respective forward (*Top Right*) and backward (*Bottom Right*) paths (from an original 10-μs forward/backward simulation and five additional distributed runs for 2 to 4 μs each, colored according to counts of a total of 28,412 snapshots into 200 × 200 grids on the X-Θ plane; see *SI Appendix*, Fig. S6 for individual maps and -lnP mapping with normalized counts or probability P). Note that the state 3' identified from B splits into two populations along −X(t) with a same Θ(t) due to the strand crossing.

paths. The accumulated samplings along the forward and backward paths are shown in Fig. 3 C, *Right*. One sees that that the protein COM dominantly follows a rotation-coupled path tracking along the helical DNA groove. The protein COMs are also mapped into several population states on the X–Θ plane, which accordingly reveals equilibrium distributions and hence estimated free energetics along the diffusion path (∼2 k_BT; see logarithmic probability mapping in *SI Appendix*, Fig. S6).

Furthermore, we had conducted two additional 10-μs equilibrium simulations of WRKY on the poly-A DNA, one as a repeated run and the other with a varied initial condition (from a 2-μs equilibrated protein conformation from the nonspecific DNA binding). Though protein reorientation or relaxation on the DNA persist (via the COM motions; *SI Appendix*, Fig. S8 A–C), no further protein stepping or diffusion along the poly-A DNA was sampled, as HBs at the protein–DNA interface were

stably maintained in both simulations. Mappings of all samplings obtained for WRKY on the poly-A DNA (simulation accumulated ∼70 μs; *SI Appendix*, Table S1), including static and forward/backward diffusion, are provided (*SI Appendix*, Fig. S8D). In summary, the longitudinal and rotational motions of the protein are largely coupled during the regular helical tracking motions of the protein; only occasionally, the domain protein skips the groove tracking and slides across the DNA strand.

WRKY Stepping along Poly-A DNA with Cyclic HB Breaking and Reforming at the Protein–DNA Interface Sampled in the Atomic Simulations. In the 10-μs all-atom simulations of WRKY along poly-A DNA, by close inspections on how protein individual residues break and reform HB contacts with the DNA backbone during protein diffusion, we show the representative protein-stepping schematics or HB dynamics on the DNA (Fig. 4).

Dai et al.
Revealing atomic-scale molecular diffusion of a plant-transcription factor WRKY domain
protein along DNA

PNAS | 5 of 10
https://doi.org/10.1073/pnas.2102621118

According to the HB dynamics revealed at the protein–DNA interface from the simulation (*SI Appendix*, Fig. S9A), we define different HB configurations (I to VII) and connect them to the protein COM states (i.e., states 1 to 4 from Fig. 3). Among eight key residues frequently forming HBs with the preferred DNA strand, the very front residue R135 (NH1/NH2) that initially bonds with A18 backbone (O2P; ~0.4 to 1.86 μs as configuration or config I, protein COM state 1) has the HB broken first and then fluctuates to occasionally form a new HB with A19 (config II at ~2 μs, state 2), as other contacts almost remain intact. At ~5 μs, as the protein moves forward (state 3), most of the HBs break within ~80 ns, while the central K144 shifts its HB with A17 to A18 and K122 forms a new HB with A17 (config III); four of the front HBs (but not the one by R135) reform quickly (config IV, for ~30 ns), then R131 reforms HB with A17 (config V, for ~40 ns), the backside K125 reforms HB with A16 (config VI, for ~60 ns), and, finally, R135 reforms HB with A19, which concludes the 1-bp stepping cycle (config VII or config I recovered at ~5.2 μs). Note that transitions from config III to VII happen quickly (within ~0.2 μs), as the protein COM remains at state 3 during this stepping cycle for 1 bp (~7 μs); therefore, the protein COM first oscillates back and forth (with protein orientational changes) and then moves forward (via state transitions 1 → 2 →

1 → 3 as in Fig. 3) (i.e., tracking along the DNA major groove, reorienting with the groove (~5 μs) until the majority of HBs suddenly shifted (broken and reformed). Further movements revealed in the simulation (6 to 10 μs; *SI Appendix*, Fig. S9B) account for some protein slipping (~2-bp step, incomplete): the HBs break in a slightly different way, the middle and rear contacts break and have not yet reformed, while the COM of the protein shifts ~2 bp. The schematics of protein–DNA interfacial HBs formed on the nonpreferred strand are also provided (*SI Appendix*, Fig. S9C): though there are only 2 to 3 HBs formed occasionally, one finds that R118 breaks and reforms HB with the DNA phosphates from T20′ to T23′ throughout the 10-μs simulation (across ~2 to 3 bp).

In the backward movements of WRKY along DNA, the protein also tracks along the major groove initially (<6.6 μs; *SI Appendix*, Fig. S10). It starts with R131 squeezing on the neighboring K125 (from config I′, 0.42 to 3.9 μs, to II′, ~3.9 to 5.18 μs or the protein COM state 1→ 2′ → 1′) to break the back contact K125–A15. After R131 forms stable contact with A15, the COM of protein moves backward (1′ → 3′), and the Q146–A16 HB breaks as K144 slides backward to contact A16 (config III′). The continuous movements of the COM have broken most of the HBs (config IV′). After that, the middle region
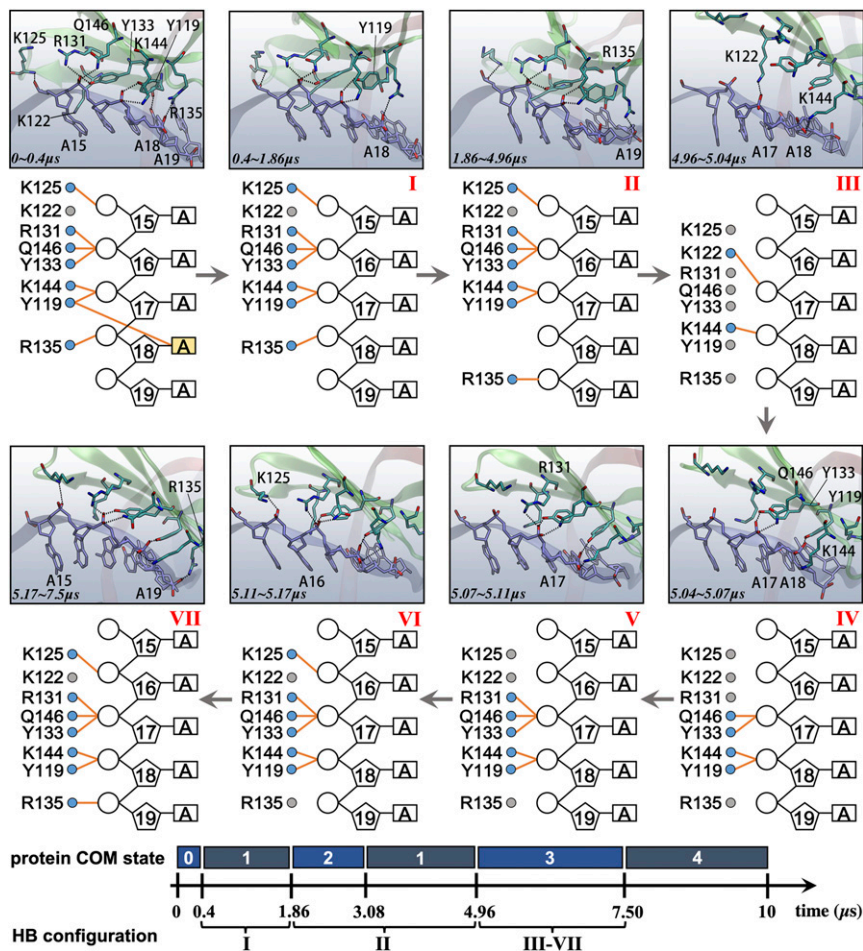

Fig. 4. The stepping schematics and structural views of WRKY moving forward along poly-A DNA during diffusion from the 10-μs all-atom equilibrium MD simulation. Since WRKY associates closely with the preferred strand of DNA, we show schematics of eight key residues (filled circle) from WRKY that make HB contacts with the preferred strand (open circle, pentagon, and rectangle for the phosphate, sugar, and base of a nucleotide, respectively). The HBs in the schematics are depicted in orange lines. The corresponding molecular views at the protein–DNA interface are provided (the preferred and nonpreferred strands in blue and pink, respectively; WRKY protein in green). The configurations I to VII are defined according to the HB dynamics at the protein–DNA interface (see *WRKY Stepping along Poly-A DNA with Cyclic HB Breaking and Reforming at the Protein–DNA Interface Sampled in the Atomic Simulations* and *SI Appendix*, Fig. S9A).

readjusts with shifted nucleotides (config V' and VI', 5.22 to 5.47μs). Finally, the edge residue R135 re-forms HB with A17 (config VII', 5.47 to 5.87μs) and the initial set of contacts almost re-form except for the one from K125 to A14. However, WRKY then seems to reduce its association with the DNA and crosses the preferred strand to move from the major groove to the minor groove (*SI Appendix*, Fig. S11A): One can find five residues (R131, Y133, K142, K144, and Q146) reestablishing contacts with the nonpreferred strand after crossing the strand and sliding a further ~2 bp backward along the DNA, while K142 and K144 keep associating with both strands (see *SI Appendix*, Fig. S11B for 6.6 to 10 μs with the protein COM state 4' → 5').

The WRKY Electrostatic Association Bias toward the Preferred DNA Strand Is Less Distinct during Protein Diffusion than in the Static DNA Binding. Furthermore, we calculated time-dependent electrostatic (*ele*) and van der Waals (*vdW*) interaction energies at the protein–DNA interface for respective DNA strands and the core protein from protein-static binding on DNA (specific and nonspecific) to the forward and backward diffusional movements (*SI Appendix*, Fig. S12A and Table S2). During the 1-bp stepping along the forward path (<5 μs), the interaction energetics with the preferred strand are weaker than in static binding; the *ele* interactions with the nonpreferred strand are nevertheless similar to those in the static binding. Later (5 to 10 μs with protein slipping and stochastic motions), the *ele* interactions weaken on the preferred strand further but are still stronger than those on the nonpreferred strand. A similar trend is revealed in the backward movements (i.e., the WRKY *ele* association with the preferred strand is weaker than that in the static binding), while the association with the nonpreferred strand strengthens from the early to late stages. With stochastic movements becoming prominent in the late stage of the forward/backward diffusion, the accompanying *ele* fluctuations also increase while the protein energetic distinctions between the two DNA strands decrease compared to the static binding systems. The average energetics obtained from various simulation systems (including those performed under the updated DNA force field) are summarized in Fig. 5A.

Since WRKY demonstrates more or less bias in association with the preferred DNA strand over the nonpreferred one, from the static binding to diffusional movements, we then analyzed the energetic disparities and correlations between the two DNA strands in association with the protein (*SI Appendix*, Fig. S12B). In order to measure how differently the protein interaction energetics are contributed to by the two DNA strands, we calculated t-values that characterize average energetic differences over the standard errors or fluctuations (see *SI Appendix*, *Methods*). The *ele* t-value in static and specific DNA binding is indeed highest (e.g., $t_s^{ele} = 172$ with $P < 10^{-5}$ for $n = 2,500$ samples) but then becomes much lower upon diffusion (*SI Appendix*, Table S2). Next, to assess whether the protein association energetics with the two DNA strands are dynamically correlated, the Pearson correlation coefficients r were calculated between the time-dependent energetic data sets. In specific and nonspecific DNA binding, the correlations are comparatively small but detectable (here $|r| > 0.05$ for $P < 0.01$) between the two-strand energetics. In the forward and backward diffusion, the correlation strength can become much larger occasionally (e.g., to $r \sim -0.52$ forward and ~ 0.18 backward (*SI Appendix*, Table S2 for full energetics, t-values, and r values for various simulation systems). The analyses thus indicate that WRKY electrostatic association shows less bias and more coordination between the two DNA strands during the protein diffusion than in the static DNA binding.

Under current technology, one cannot yet sample processive diffusion of the protein on DNA using unbiased atomic simulations; to do that, we additionally conducted CG MD simulations to the WRKY–DNA complex using CafeMol (39) (*SI Appendix*, *Methods*). In *SI Appendix*, *Results*, we show that, in the CG simulations, WRKY conducts processive diffusion along DNA and occasional dissociations from DNA at variable ionic conditions (*SI Appendix*, Fig. S13 and Movies S5–S7); in addition, WRKY also shows different stepping patterns along different DNA sequences (*SI Appendix*, Fig. S14). Furthermore, we verified the processive diffusion of WRKY on DNA experimentally with single-molecule fluorescence measurements (*SI Appendix*, *Methods* and Fig. S15), albeit current resolution is not high enough to discern bp movements of the protein. The
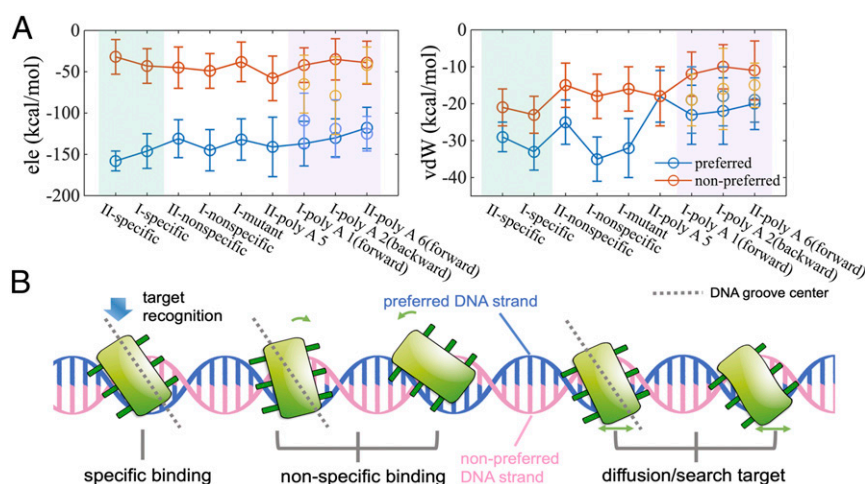
Fig. 5. The biased DNA-strand association of the WRKY domain protein in diffusion and recognition on DNA. (A) The energetic association between WRKY and two respective DNA strands (preferred in blue and nonpreferred red). The energetics were calculated from all-atom MD simulations for the specific, nonspecific DNA binding, and K122A systems and for the forward (poly-A1) and backward (poly-A2) diffusion systems (indicated with prefix I), and, additionally, under an updated DNA force field (with prefix II). The specific binding cases are colored in a light-green background, and diffusive cases are in light purple. (B) Schematics on the suggested scenario of the small TF-domain protein search and recognition of specific sequences on DNA. The domain protein reorients constantly during diffusion, following the DNA-helical track. The protein diffusional search can be facilitated via modulating the bias and coordination between the two associating DNA strands, with less bias and more coordination between the two strands to assist the protein diffusion and with more bias and less coordination between the two strands to support DNA-sequence recognition of the protein on the preferred DNA strand.

Dai et al.
Revealing atomic-scale molecular diffusion of a plant-transcription factor WRKY domain protein along DNA

PNAS | 7 of 10
https://doi.org/10.1073/pnas.2102621118

diffusion coefficient has been estimated at about 0.05 bp^2/μs, which is consistent with our computational samplings of the protein stepping on the DNA in the 10-μs all-atom MD simulations.

Discussion

In this work, based upon high-resolution structures of DNA-binding complexes of a representative TF domain protein WRKY, we demonstrated microseconds molecular dynamics of protein diffusion along DNA with unprecedented details. To avoid artifacts from external force or bias, equilibrium all-atom simulations were conducted for both static binding and protein diffusion on DNA, which were accumulated to ∼100 μs [under Amber99SB-ILDN force field (40) for proteins and Amber94 force field (41) for nucleic acids using Gromacs (42) (*SI Appendix, Methods*)], including several 10-μs–long MD simulations and multiple distributed 2- to 4-μs runs to improve samplings (*SI Appendix*, Table S1). In addition, ∼40 μs all-atom simulations were also conducted (with protein–DNA energetics and related statistics also shown in *SI Appendix*, Fig. S12) under an updated DNA force field Parmbsc1 or BSC1 (43), which slightly stabilizes DNA backbone motions. The simulations conducted under the updated force field reproduced the dominant features of WRKY–DNA binding on specific and nonspecific DNA (*SI Appendix*, Fig. S16), with biased association still toward the preferred DNA strand. The 1-bp stepping dynamics of the WRKY domain protein on the poly-A DNA has also been well captured under the updated force field (*SI Appendix*, Fig. S17 and Movie S8).

The WRKY Domain Protein Preferentially Binds One Strand of DNA, and the Protein Orientational Change on the DNA Is Significant between Specific and Nonspecific DNA Binding. For protein-static binding and recognition on DNA, both specific and nonspecific DNA (with slightly varied core sequences GGTC and GATA) binding complexes of WRKY were examined together with an mt K122A protein complex with the specific DNA. The corresponding protein–DNA binding affinities were determined via ITC measurements, with dissociation constants measured at 0.1 μM, 8 μM, and 1 μM for the specific, nonspecific, and mt K122A systems, respectively. In all these systems, one DNA strand is always preferentially bound by WRKY. The other strand, the nonpreferred one, interacts with the protein much more weakly to allow protein to associate with the DNA differently between specific and nonspecific binding modes (e.g., via variable protein β-strand regions). Consequently, the domain protein varies its orientation and affinity to different DNA sequences and recognizes certain bases on the preferred DNA strand upon the specific DNA binding. In the simulations, we have found no essential conformational change of the domain protein from the specific to nonspecific DNA binding or to regular tracking along DNA, though an occasional protein conformation transition was detected prior to directional reversal of the protein on the DNA. Meanwhile, the current study shows that relative conformational changes between the domain protein and DNA (i.e., the reorientation of the protein on the DNA) are highly significant and contribute essentially to the distinction between the nonspecific and specific DNA binding (Figs. 1, 2, and 5*B* and *SI Appendix*, Figs. S2 and S16). Such findings provide structural clues to previous work suggesting a switch of TF conformational mode in DNA search and recognition (4, 5, 44, 45).

WRKY 1-bp Stepping on Poly-A DNA Is Detected from All-Atom Equilibrium Simulations, while Its Processive Diffusion Statistics Are Obtained from Coarse-Grained Simulations. With current computing technologies, for TF-protein diffusion with an average stepping cycle lasting over tens of microseconds, it is still hard to sample the protein movements at atomic resolution. Our all-atom simulations show that protein–DNA interfacial HB contacts are constantly present, and the corresponding dynamics can

be rate limiting to hinder the protein diffusion. For homogeneous poly-A DNA, the HB dynamics seems to be facilitated along the identical DNA sequences. Consequently, we were able to identify a complete 1-bp protein-stepping cycle following the major groove of DNA, which is regulated by collective HB dynamics at the protein–DNA interface as individual HBs break and re-form throughout the cycle. In fact, three protein-stepping events were detected from six 10-μs–long atomic MD simulations of the WRKY on poly-A DNA (one captured under the updated DNA force field). Combining these simulations with multiple distributed simulations performed along the forward and backward diffusion paths, the dominant rotation-coupled DNA-helical tracking motions of the protein are demonstrated, with ∼2 k_BT diffusional free energetics estimated (*SI Appendix*, Fig. S6), consistent with previous measurements (13, 14). Furthermore, by conducting CG simulations of WRKY at the residue level on ∼200 bp DNA, processive protein diffusion along DNA were sampled at various ionic conditions and sequence patterns (*SI Appendix*, Figs. S13 and S14). Stochastic directional reversal and DNAstrand–crossing events have been well sampled in the CG simulations, while such events were captured only once or twice in the atomic simulations. The corresponding processive 1D diffusion of WRKY along DNA was also confirmed by accompanied single-molecule florescence experiments, albeit the measurements were not at a high enough resolution to detect protein-stepping motions. Interestingly, in current CG simulations, high percentiles of 1-bp stepping motions of the WRKY domain protein show along homogeneous poly-A DNA, while the 1-bp percentile drops (or more 2 to 3 bp steps show; *SI Appendix*, Fig. S14) for WRKY moving along random DNA sequences. It thus suggests that the DNA-sequence patterns make direct impacts on the protein-stepping statistics.

Stochastic Variations Revealed in Domain-Protein Stepping Provide Clues for Understanding Step-Size Variations in Other Nucleic-Acid Walkers. In the all-atom simulations, both forward and backward movements of the WRKY domain protein along DNA have been revealed. In the forward direction, right after an elementary 1-bp step of WRKY, stochasticity is noticeable as WRKY incompletely slips further for ∼2 bp, as the related HBs break and part of them re-form at the 2-bp distance. In the other case, some protein-conformational transition (HBs broken between β-strands 4 and 5) occurs right before the protein moves backward; soon after the ∼1-bp step backward, the protein shows prominent stochastic motions on the DNA as it crosses the preferred strand to move from the major to the minor groove. Such types of protein-diffusional motions along DNA have been captured currently (*SI Appendix*, Fig. S13) and in previous CG studies (46, 47), though no protein side-chain motions or protein–DNA HBs can be specified in the CG model. The WRKY domain-protein stepping on DNA is comparable to other nucleic-acid walkers or molecular walkers following a quasiperiodic track. For example, motor proteins such as DNA-packaging motors or helicases have been detected with variable stepping sizes from single-molecule measurements (48–52). The stepping motions of the motor proteins can be fast, similar to the TF proteins, though substrate binding or chemical catalysis and mechano-chemical coupling that supports directional movements of the motor proteins can be quite slow (e.g., over milliseconds). Although various models were presented to explain diverse stepping behaviors of motor proteins, current studies suggest that the multiple stepping sizes arise because of nonsynchronized motions of individual protein residues forming HB contacts on the DNA backbone. Besides, stochasticity always plays a significant role in protein stepping or sliding due to thermal fluctuations. The simulated TF-protein stepping dynamics, stochastic variations, and DNA-sequence effects await experimental validations at the sufficient high or bp resolution.

The Protein Electrostatic-Association Bias with One Strand of DNA Can Be Marginally Maintained to Assist the Domain Protein Diffusion and Maximally Employed to Support Protein DNA Sequence Recognition. To further understand how such a WRKY domain protein searches and locates specific target sequences on the DNA, we note that even though WRKY distinguishes the two strands of DNA by almost always associating tightly with the preferred strand, the disparity between protein association with the two strands varies from dynamical search to static binding or the recognition stage. During diffusion or stochastic movements of the protein on the DNA, the disparity of the *vdW* protein's association with the two DNA strands almost vanishes, while the electrostatic differentiation persists but is only marginally maintained, likely due to protein-random reorientations on the DNA. For protein-regular stepping or groove tracking along DNA, the protein–DNA energy disparity between the two strands increases, as constant reorientation of the protein happens along the DNA-helical track. It appears that some coordination between the two DNA strands in association with the protein supports the protein movements. Quasi-static protein binding on the DNA with lowered fluctuations then enhances the protein-association disparities between the two DNA strands. Such enhanced protein–DNA strand bias may contribute to fine-tuning the protein orientation and supporting specific DNA-sequence recognition on the preferred DNA strand. Due to additional electrostatic stabilization or reduced fluctuations, particularly on the preferred DNA strand, the corresponding energetic distinction between the preferred and nonpreferred strands becomes maximized for the protein on the specific DNA. The biased protein association with the preferred DNA strand can also perturb base pairing in the duplex DNA and thus assist base readout on the DNA strand for sequence recognition.

Hence, for a small-domain TF protein, our studies bring a working scenario in which the biased protein association with one strand of the dsDNA can be marginally sustained during the stochastic search process to facilitate fast protein movements, while the bias can be maximally employed into the quasi-static binding and DNA-sequence recognition as the protein reorients and stabilizes on the specific DNA. Such a scenario is related to protein geometry on the DNA-helical structure, so it seems to apply for monomeric or small TF-domain proteins that fit the DNA groove. Interestingly, for dimeric proteins with two DNA-binding domains, such as Myc-Max, which we recently studied, it is found that the two basic regions or domains bind with the two complementary strands of DNA, respectively [i.e., with each domain preferentially bound with one strand (53)]. The movements of such a dimeric TF protein on DNA then rely largely on coordination between the two protein domains. Such a perspective is supported by recent structure-based bioinformatic analyses (54), which show statistically that multispecific TFs intend to form more HBs with one strand than with the other on the DNA, while highly specific DNA-binding proteins, typically dimeric type-II restriction endonucleases, associate nonpreferentially with both DNA strands. Combining with these findings, the biased DNA-strand association scenario appears to be generic for small TF-domain proteins to balance target search and recognition on the DNA. For larger or oligomeric TF proteins, however, additional considerations of protein-internal degrees of freedom or coordination are needed.

Materials and Methods

Detailed descriptions about obtaining the crystal structure, the setup of atomic and CG simulations, the ITC experiments, and the single-molecule florescence experiments are provided in *SI Appendix, Methods*.

Data Availability. All study data are included in the article and/or supporting information.

1. O. G. Berg, P. H. von Hippel, Selection of DNA binding sites by regulatory proteins. Statistical-mechanical theory and application to operators and promoters. *J. Mol. Biol.* **193**, 723–750 (1987).
2. P. H. von Hippel, O. G. Berg, Facilitated target location in biological systems. *J. Biol. Chem.* **264**, 675–678 (1989).
3. S. E. Halford, J. F. Marko, How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* **32**, 3040–3052 (2004).
4. M. Slutsky, L. A. Mirny, Kinetics of protein-DNA interaction: Facilitated target location in sequence-dependent potential. *Biophys. J.* **87**, 4021–4035 (2004).
5. M. Bauer, R. Metzler, Generalized facilitated diffusion model for DNA-binding proteins with search and recognition states. *Biophys. J.* **102**, 2321–2330 (2012).
6. A. A. Shvets, M. P. Kochugaeva, A. B. Kolomeisky, Mechanisms of protein search for targets on DNA: Theoretical insights. *Molecules* **23**, 2106 (2018).
7. M. Ricchetti, W. Metzger, H. Heumann, One-dimensional diffusion of *Escherichia coli* DNA-dependent RNA polymerase: A mechanism to facilitate promoter location. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4610–4614 (1988).
8. P. C. Blainey, A. M. van Oijen, A. Banerjee, G. L. Verdine, X. S. Xie, A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5752–5757 (2006).
9. A. Tafvizi, F. Huang, A. R. Fersht, L. A. Mirny, A. M. van Oijen, A single-molecule characterization of p53 search on DNA. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 563–568 (2011).
10. P. Hammar *et al.*, The lac repressor displays facilitated diffusion in living cells. *Science* **336**, 1595–1598 (2012).
11. S. Redding, E. C. Greene, How do proteins locate specific targets in DNA? *Chem. Phys. Lett.* **570**, 1–11 (2013).
12. J. Gorman, E. C. Greene, Visualizing one-dimensional diffusion of proteins along DNA. *Nat. Struct. Mol. Biol.* **15**, 768–774 (2008).
13. P. C. Blainey *et al.*, Nonspecifically bound proteins spin while diffusing along DNA. *Nat. Struct. Mol. Biol.* **16**, 1224–1229 (2009).
14. E. G. Marklund *et al.*, Transcription-factor binding and sliding on DNA studied using micro- and macroscopic models. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 19796–19801 (2013).
15. A. D. Robison, I. J. Finkelstein, High-throughput single-molecule studies of protein-DNA interactions. *FEBS Lett.* **588**, 3539–3546 (2014).
16. C. Liu, Y.-L. Liu, E. P. Perillo, A. K. Dunn, H.-C. Yeh, Single-molecule tracking and its application in biomolecular binding detection. *IEEE J. Sel. Top. Quantum Electron.* **22**, 6804013 (2016).
17. M. Stracy, A. N. Kapanidis, Single-molecule and super-resolution imaging of transcription in living bacteria. *Methods* **120**, 103–114 (2017).
18. K. Nadassy, S. J. Wodak, J. Janin, Structural features of protein-nucleic acid recognition sites. *Biochemistry* **38**, 1999–2017 (1999).
19. A. D. Mackerell Jr, L. Nilsson, Molecular dynamics simulations of nucleic acid-protein complexes. *Curr. Opin. Struct. Biol.* **18**, 194–199 (2008).
20. J. L. Klepeis, K. Lindorff-Larsen, R. O. Dror, D. E. Shaw, Long-timescale molecular dynamics simulations of protein structure and function. *Curr. Opin. Struct. Biol.* **19**, 120–127 (2009).
21. J. R. Perilla *et al.*, Molecular dynamics simulations of large macromolecular complexes. *Curr. Opin. Struct. Biol.* **31**, 64–74 (2015).
22. S. Furini, P. Barbini, C. Domene, 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 (2013).
23. L. Etheve, J. Martin, R. Lavery, Protein-DNA interfaces: A molecular dynamics analysis of time-dependent recognition processes for three transcription factors. *Nucleic Acids Res.* **44**, 9990–10002 (2016).
24. M. Wieczór, J. Czub, How proteins bind to DNA: Target discrimination and dynamic sequence search by the telomeric protein TRF1. *Nucleic Acids Res.* **45**, 7643–7654 (2017).
25. M. Khabiri, P. L. Freddolino, Deficiencies in molecular dynamics simulation-based prediction of protein-DNA binding free energy landscapes. *J. Phys. Chem. B* **121**, 5151–5161 (2017).
26. B. Pandey, A. Grover, P. Sharma, Molecular dynamics simulations revealed structural differences among WRKY domain-DNA interaction in barley (*Hordeum vulgare*). *BMC Genomics* **19**, 132 (2018).
27. S. Furini, C. Domene, S. Cavalcanti, Insights into the sliding movement of the lac repressor nonspecifically bound to DNA. *J. Phys. Chem. B* **114**, 2238–2245 (2010).
28. Y. Qi *et al.*, Strandwise translocation of a DNA glycosylase on undamaged DNA. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 1086–1091 (2012).
29. M. Zacharias, Atomic resolution insight into Sac7d protein binding to DNA and associated global changes by molecular dynamics simulations. *Angew. Chem. Int. Ed. Engl.* **58**, 5967–5972 (2019).

Dai et al.
Revealing atomic-scale molecular diffusion of a plant-transcription factor WRKY domain protein along DNA

PNAS | 9 of 10
https://doi.org/10.1073/pnas.2102621118

30. D.-A. Silva *et al*., Millisecond dynamics of RNA polymerase II translocation at atomic resolution. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 7665–7670 (2014).

31. L.-T. Da *et al*., T7 RNA polymerase translocation is facilitated by a helix opening on the fingers domain that may also prevent backtracking. *Nucleic Acids Res.* **45**, 7909–7921 (2017).

32. C. Maffeo, A. Aksimentiev, Molecular mechanism of DNA association with single-stranded DNA binding protein. *Nucleic Acids Res.* **45**, 12125–12139 (2017).

33. S. Peng *et al*., Target search and recognition mechanisms of glycosylase AlkD revealed by scanning FRET-FCS and Markov state models. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 21889–21895 (2020).

34. J. Tian, L. Wang, L.-T. Da, Atomic resolution of short-range sliding dynamics of thymine DNA glycosylase along DNA minor-groove for lesion recognition. *Nucleic Acids Res.* **49**, 1278–1293 (2021).

35. P. J. Rushton, I. E. Somssich, P. Ringler, Q. J. Shen, WRKY transcription factors. *Trends Plant Sci.* **15**, 247–258 (2010).

36. T. Eulgem, P. J. Rushton, S. Robatzek, I. E. Somssich, The WRKY superfamily of plant transcription factors. *Trends Plant Sci.* **5**, 199–206 (2000).

37. M.-R. Duan *et al*., DNA binding mechanism revealed by high resolution crystal structure of *Arabidopsis thaliana* WRKY1 protein. *Nucleic Acids Res.* **35**, 1145–1154 (2007).

38. Y. P. Xu, H. Xu, B. Wang, X.-D. Su, Crystal structures of N-terminal WRKY transcription factors and DNA complexes. *Protein Cell* **11**, 208–213 (2020).

39. H. Kenzaki *et al*., CafeMol: A coarse-grained biomolecular simulator for simulating proteins at work. *J. Chem. Theory Comput.* **7**, 1979–1989 (2011).

40. K. Lindorff-Larsen *et al*., Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950–1958 (2010).

41. W. D. Cornell *et al*., A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179–5197 (1995).

42. H. J. Berendsen, D. van der Spoel, R. van Drunen, GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **91**, 43–56 (1995).

43. I. Ivani *et al*., Parmbsc1: A refined force field for DNA simulations. *Nat. Methods* **13**, 55–58 (2016).

44. L. Hu, A. Y. Grosberg, R. Bruinsma, Are DNA transcription factor proteins Maxwellian demons? *Biophys. J.* **95**, 1151–1156 (2008).

45. H.-X. Zhou, Rapid search for specific sites on DNA through conformational switch of nonspecifically bound proteins. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 8651–8656 (2011).

46. C. Tan, S. Takada, Dynamic and structural modeling of the specificity in protein–DNA interactions guided by binding assay and structure data. *J. Chem. Theory Comput.* **14**, 3877–3889 (2018).

47. M. Saito, T. Terakawa, S. Takada, How one-dimensional diffusion of transcription factors are affected by obstacles: Coarse-grained molecular dynamics study. *Mol. Simul.* **43**, 1315–1321 (2017).

48. S. Dumont *et al*., RNA translocation and unwinding mechanism of HCV NS3 helicase and its coordination by ATP. *Nature* **439**, 105–108 (2006).

49. S. Myong, M. M. Bruno, A. M. Pyle, T. Ha, Spring-loaded mechanism of DNA unwinding by hepatitis C virus NS3 helicase. *Science* **317**, 513–516 (2007).

50. E. J. Tomko, C. J. Fischer, A. Niedziela-Majka, T. M. Lohman, A nonuniform stepping mechanism for *E. coli* UvrD monomer translocation along single-stranded DNA. *Mol. Cell* **26**, 335–347 (2007).

51. J. R. Moffitt *et al*., Intersubunit coordination in a homomeric ring ATPase. *Nature* **457**, 446–450 (2009).

52. Y. R. Chemla, Revealing the base pair stepping dynamics of nucleic acid motor proteins with optical traps. *Phys. Chem. Chem. Phys.* **12**, 3080–3095 (2010).

53. L. Dai, J. Yu, Inchworm stepping of Myc-Max heterodimer protein diffusion along DNA. *Biochem. Biophys. Res. Commun.* **533**, 97–103 (2020).

54. M. Lin, J. T. Guo, New insights into protein-DNA binding specificity from hydrogen bond based comparative study. *Nucleic Acids Res.* **47**, 11103–11113 (2019).