REVEALING ATOMIC-SCALE MOLECULAR DIFFUSION OF A PLANT TRANSCRIPTION FACTOR WRKY DOMAIN PROTEIN ALONG DNA

https://doi.org/10.1101/2020.02.14.950295

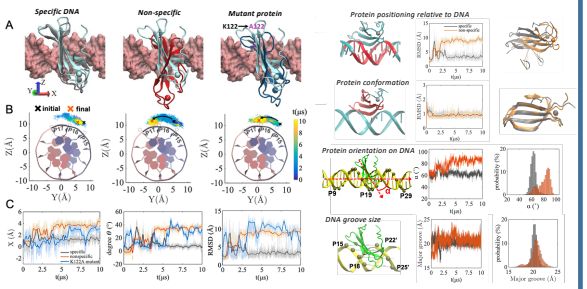
Liqiang Dai[1], Yongping Xu[2], Zhenwei Du[2], Xiao-dong Su[2], and Jin Yu[3]* (jin.yu@uci.edu)

[1] Beijing Computational Science Research Center, Beijing, China; School of Life Sciences, Peking University, Beijing, China

[3] Department of Physics and Astronomy, Department of Chemistry, NSF-Simons Center for Multiscale Cell Fate Research, University of California, Irvine, CA

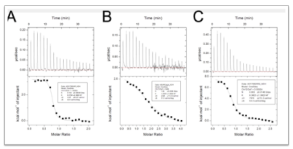
Abstract & simulation systems

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 of the plant WRKY domain protein in complex with DNA from crystallography and demonstrated microsecond diffusion dynamics of WRKY on the DNA, employing all-atom molecular dynamics (MD)simulations. Notably, we found that WRKY preferentially binds to one strand of DNA with significantly stronger energetic association than to the other strand. The preferential binding becomes highly prominent from non-specific to specific static binding to DNA, but less distinct during diffusive movements of TF along DNA. Remarkably, without employing acceleration forces or bias, we captured a complete one-base pair (bp) stepping cycle of WRKY tracking along major groove of DNA, as individual protein-DNA contacts break and reform at the binding interface. Continuous tracking of WRKY forward or backward, with occasional sliding as well as strand crossing to the minor groove of DNA, have also been captured in the simulations. The processive diffusion of WRKY had been confirmed by accompanied single-molecule fluorescence assays and coarse-grained (CG) structural simulations. The study thus provides unprecedented structural dynamics details on the TF diffusion, suggests how TF possibly approaches to target sequences, and supports further high-precision experimental follow-up. The stochastic movements revealed in the TF diffusion also provide general clues on how other nucleic acid walkers step and slide along DNA.

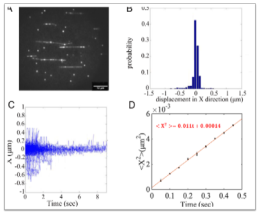


***Left*: MD simulation on specific and non-specific DNA association of WRKY.** (A) Comparisons of the initial (cyan) and final (gray, red, blue) structures of the simulation of the wild-type (wt) protein binding on the specific DNA (GGTCAA), the non-specific DNA (GATAAA), and the mutant (mt) protein (K122A) binding on the specific DNA. (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 but not translocation) are denoted. The time evolution is represented by coloring (from blue to yellow). (C) The relaxation of the protein COM on X & Θ and the RMSDs (relative to DNA), for respective simulation systems. ***Right*: Examining positioning, conformational change, orientational change of WRKY** on the specific and nonspecific DNA, along with the DNA conformational (groove size) change bound with WRKY. The results indicate almost NO conformational variation of the protein itself from specific to non-specific binding mode, while the orientational variations of the domain protein on DNA is quite significant upon the non-specific DNA binding.

Measuring protein-DNA binding affinity and processive diffusion

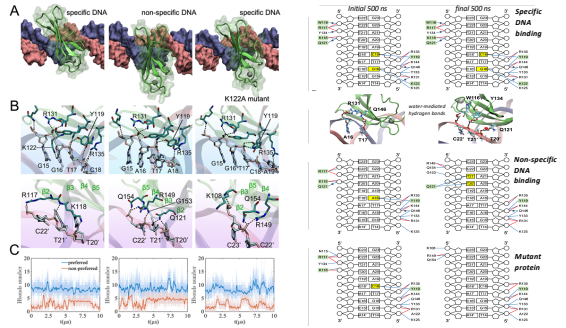


The Isothermal Titration Calorimetry (ITC) experiments (A) WRKY1-N and specific GGTC core DNA. (B) WRKY1-N and nonspecific GATA core DNA. (C) WRKY1-N mutant (K23A) and specific GGTC core DNA. The relative binding affinity (the dissociation constant KD) is determined to be 0.1μM, 8 μM and 1μM, respectively. ITC experiments were performed by titrating protein into the dsDNA.



The single-molecule fluorescence measurements of the WRKY domain protein diffusing on DNA

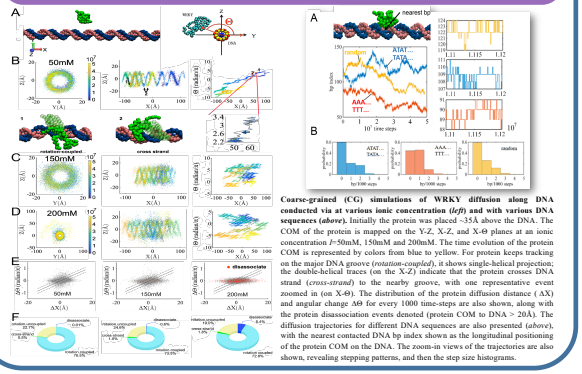
Specific vs non-specific DNA binding of the domain protein



The association between the WRKY domain protein and respective strands of DNA. (A) The equilibrated protein-DNA complexes with the DNA strands in blue (*the preferred*) and pink (*the non-preferred*). (B) The hydrogen bond (HB) associations at the protein-DNA interface on the preferred strand (*top*) and the non-preferred strand (*bottom*). (C) Time-dependent HB statistics at the protein-DNA interface.

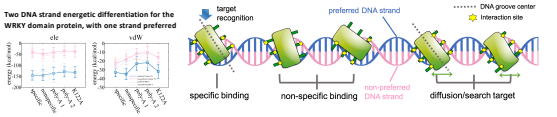
Schematics of hydrogen bonds (HBs, blue) and salt-bridge interactions (red) between WRKY and DNA. Bases forming the HBs with the protein are highlighted in yellow. The conserved protein residues from WRKYGQK motif are highlight in green.

Processive protein diffusion in the coarse-grained simulation

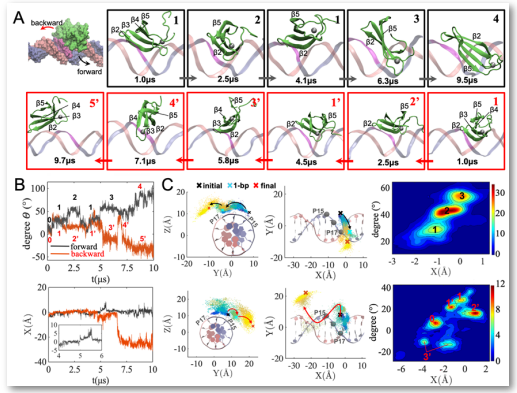


Coarse-grained (CG) simulations of WRKY diffusion along DNA conducted via at various ionic concentration (*left*) and with various DNA sequences (*above*). Initially the protein was placed ~35Å above the DNA. The COM of the protein is mapped on the Y-Z, X-Z, and X-Θ planes at an ionic concentration I~50mM, 150mM and 200mM. The time evolution of the protein COM is represented by colors from blue to yellow. For protein keeps tracking on the major DNA groove (*rotation-coupled*), it shows single-helical projection; the double-helical traces (on the X-Z) indicate that the protein crosses DNA strand (*cross-strand*) to the nearby groove, with one representative event zoomed in (on X-Θ). The distribution of the protein diffusion distance (ΔX) and angular change ΔΘ for every 1000 time-steps are also shown, along with the protein disassociation events denoted (protein COM to DNA > 20Å). The diffusion trajectories for different DNA sequences are also presented (*above*), with the nearest contacted DNA bp index shown as the longitudinal positioning of the protein COM on the DNA. The zoom-in views of the trajectories are also shown, revealing stepping patterns, and then the step size histograms.

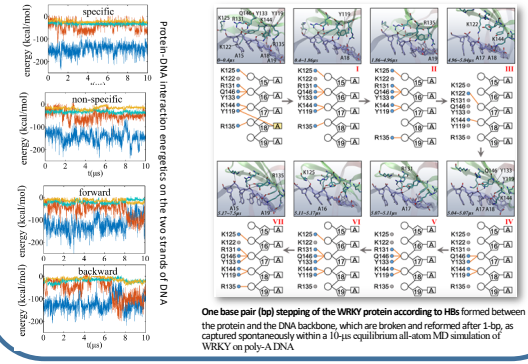
The domain protein diffusional search & recognition scenario



Spontaneous 1-bp protein stepping on DNA from simulation



The diffusion of WRKY along poly-A DNA in the forward and backward direction revealed from two 10-μs atomistic MD simulations. The WRKY domain protein is shown in green and two DNA strands in blue (the preferred strand) and pink (the non-preferred strand). The helical trajectories of the protein COM along the DNA, shown for the angular Θ(t) and the longitudinal movement X(t) from the simulation, mapped on the Y-Z & X-Y plane (*middle*), colored by the simulation time (blue to yellow). The maps on the X-Θ plane are from respective forward and backward trajectories (each with a 10-μs run and five 2-μs distributed runs). Note that state 3' splits into two substates along X(t) with the same Θ(t) due to the DNA strand crossing of the protein.



One base pair (bp) stepping of the WRKY protein according to HBs formed between the protein and the DNA backbone, which are broken and reformed after 1-bp, as captured spontaneously within a 10-μs equilibrium all-atom MD simulation of WRKY on poly-A DNA

Simulation packages and Acknowledgements

All-atom molecular dynamics (MD) simulations were performed by using the GROMACS 5.12 software under Amber99SB-ILDN force field for protein and Amber94 (and then BCS1) force field for DNA. All the coarse-grained (CG) MD simulations were performed by using the CafeMol 3.0. The molecular images were made by VMD.

This work has been supported by NSFC Grant #11775016 and #11635002. JY has been supported by the CMCF of UCI via NSF DMS 1763272 and the Simons Foundation grant #594598 and start-up fund from UCI. We acknowledge the computational support from the Special Program for Applied Research on Super Computation of the NSFC Guangdong Joint Fund (the second phase) under Grant No. U1501501 and from the Beijing Computational Science Research Center.