JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

p53 Searches on DNA by Rotation-Uncoupled Sliding at C-Terminal Tails and Restricted Hopping of Core Domains

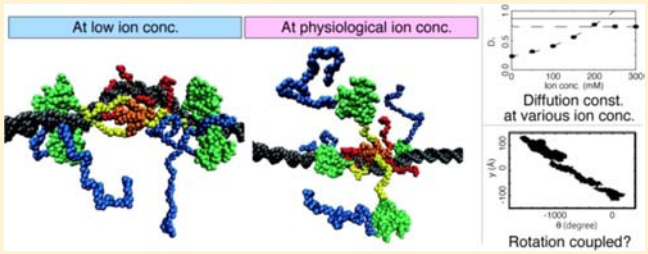
Tsuyoshi Terakawa,[†] Hiroo Kenzaki,[†] and Shoji Takada*[,†,‡]

[†]Department of Biophysics Graduate School of Science, Kyoto University, Kyoto 606-8502 Japan

[‡]CREST Japan Science and Technology Agency, Saitama 332-0012 Japan

Supporting Information

ABSTRACT: The tumor suppressor p53 is a transcription factor that searches its cognate sites on DNA. During the search, the roles and interplay of its two DNA binding domains, the folded core domain and the disordered C-terminal domain (CTD), have been controversial. Here, we performed molecular simulations of p53 at various salt concentrations finding that, at physiological salt concentration, p53 diffuses along nonspecific DNA via rotation-uncoupled sliding with its CTD, whereas the core domain repeats dissociation and association. This is in perfect agreement with a recent single molecule experiment. In the simulation of tetrameric full-length p53, two DNA binding domains both bound to nonspecific DNA in a characteristic form at low salt concentration, whereas at physiological salt concentration, only CTD kept bound to DNA and the core domain frequently hopped on DNA. Simulations of a construct that lacks the core domain (TetCD) clarified rotation-uncoupled diffusion on nonspecific DNA. At low salt concentration, the diffusion constant due to sliding was dependent on the salt concentration, which differs from the prediction of a classic theory of transcription factors. At physiological salt concentration, it was independent of the salt concentration, in harmony with experiments. Moreover, we found that the sliding via the CTD follows the helical pitch of DNA (i.e., rotation-coupled sliding) at low salt concentration while it is virtually uncoupled to the helical pitch, a hallmark of rotation-uncoupled sliding at physiological salt concentration.



INTRODUCTION

The tumor suppressor p53 is a multifunctional transcription factor that plays vital roles in maintaining the genome integrity by controlling the transcription of various genes, of which products induce apoptosis, cell cycle arrest, DNA repair, or others.[1] p53 is a homo tetrameric protein, and each subunit is composed of four domains [the N-terminal (NTD), the core, the tetramerization (TET), and the C-terminal (CTD) domains] and one linker region which connects the core and TET domains (Figure 1A).[2] The core and TET domains have folded structures, which were well characterized by X-ray crystallography and NMR spectroscopy,[3−9] whereas the NTD and CTD and the linker region are known to be intrinsically disordered.[10] Interestingly, p53 contains two DNA binding domains: the core domain and the CTD. The core domain recognizes and is responsible for specific binding to its cognate sites on DNA.[3,8] Conversely, the disordered CTD binds to DNA nonspecifically.[11] A recent experiment showed that the p53 affinities to nonspecific and specific DNA can be similar or different, depending on salt concentration and on an acetylation in the core domain.[12] Exact roles of two DNA binding domains for nonspecific DNA have been controversial.

As a transcription factor, p53 needs to search its cognate sites on long DNA, which requires efficient search mechanisms. von Hippel et al. proposed a few distinct search mechanisms for general transcription factors, and one suggested mechanism is sliding on DNA:[13] A protein diffuses along DNA without dissociation. For various DNA binding proteins, single molecule and biochemical experiments gave support for the sliding along DNA.[14−21]

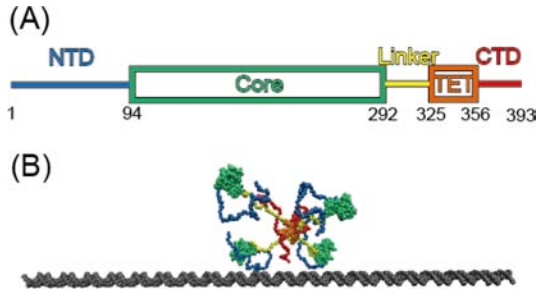


Figure 1. (A) The domain composition and (B) the initial structure of flp53 with dsDNA. (A) flp53 consists of an NTD (blue), followed by a central DNA binding core domain (Core, green), a linker region (Linker, yellow), a TET (orange), and the CTD (red). The boxes and lines represent folded domains and disordered regions, respectively. (B) The color scheme for flp53 is same as in (A), and dsDNA is in gray.

Received: June 3, 2012
Published: August 10, 2012

   dx.doi.org/10.1021/ja305369u | J. Am. Chem. Soc. 2012, 134, 14555−14562

Specifically for p53, biochemical experiments suggested that p53 actually diffuses along DNA and that the sliding requires its CTD.[22,23] In addition, recent single molecule experiments for p53 showed that the diffusion constant of sliding along DNA depends on the ion concentration for a construct that lacks CTD but independent of the ion concentration for constructs that contain CTD.[14,24] From these results together with the von Hippel theory, they proposed that the CTD mediates sliding of p53, while the core domain diffuses on DNA by frequent dissociation and association.[14] We note that the interpretation is largely based on the classic theory of von Hippel et al., which uses an electrolyte rod DNA model lacking atomic resolutions. Thus, we need to re-examine this with finer resolution, which we address here for the case of p53.

The sliding can be further classified into two types: the sliding along DNA helical pitch[15] (we call it rotation-coupled sliding hereafter) and the sliding not tightly coupled to the rotation (sometime called two-dimensional sliding; we call it rotation-uncoupled sliding hereafter).[25] In the former, a protein diffuses along one helical rail on DNA irrespective of the course: major groove, minor groove, or phosphate backbone. In the rotation-uncoupled sliding, a protein diffuses wherever on DNA molecular surface without coupling to the helicity. For some DNA binding proteins, single molecule experiments suggested that the sliding couples the rotation based on the size dependence of the diffusion constant.[15,19,21] However, the rotary movement around DNA has not been directly observed so far due to low spatial and temporal resolution of the single molecule experiments.[26] In addition, they set ion concentration quite low ($\leq$20 mM) in order to increase the affinity between a protein and DNA. Thus, sliding mechanism at physiological ion concentration has not been well characterized. As pointed out by Kampmann, the rotation-uncoupled sliding has its unique advantage that a protein can bypass some barriers on DNA that are made of other bound macromolecules, enhancing DNA search in cellular environment.[25] Thus, the distinction between the two sliding modes has clear biological relevance.

The purpose of this work is to reveal the detailed dynamics of p53 sliding along nonspecific DNA at high spatial and temporal resolution. Dynamic nature of nonspecific protein−DNA binding has been successfully characterized on the other proteins using NMR spectroscopy.[27,28] However, it is difficult to address highly dynamic processes, such as protein diffusion on DNA, using such a spectroscopic approach. Alternatively, molecular dynamics simulation may compliment this caveat, but the system size (1572 residues in p53 and 200 bps in DNA) and the long time scale involved make an atomic molecular dynamics simulation difficult at the moment. Recent studies by Levy et al. with coarse grained (CG) models for various transcription factors, including p53, have shown great promise,[29] revealing target searching[30,31] and recognition mechanisms, including the fly casting mechanism,[32,33] and the roles of the disordered region in transcription factors in such mechanisms.[34,35]

Here, motivated by recent experiments,[14] we addressed the p53 dynamics on nonspecific DNA at various ion concentrations. Technically, we extended the method of Levy et al. in three points in order to incorporate atomic physical property and flexibility into CG model. First, we used the model in which double-strand (ds) DNA were treated dynamically.[36] In this model, CG sites are placed so that realistic major as well as minor grooves are well represented. In addition, intramolecular interaction strength is parametrized to reproduce accurately the persistence length of DNA. Second, we used an atomic interaction-based coarse grained (AICG) model[37] for folded domains of p53. In this model, the nonlocal interactions were weighted according to the contact energies calculated from atomic structures. Besides, the fluctuations of the CG model were optimized to reproduce the fluctuations calculated from atomic simulations in the previous work.[37] Third, we constructed statistical potentials for disordered region that composes 40% of p53 based on our previous study, in which we showed that such a statistical approach reasonably captures structural characteristics of the disordered region.[38]

Using this extended CG simulation, we demonstrate that at physiological ion concentration, p53 slides along DNA by its CTD, whereas the core domain repeats dissociation and association, consistent with the mechanism proposed by the single molecule experiment based on the ion concentration dependence of the diffusion constant.[14] In addition, we propose that the mechanism of the sliding mediated by the CTD is the rotation-uncoupled sliding at physiological salt concentration. It is indicated that p53 CTD achieves a maximal one-dimensional diffusion constant on DNA at physiological salt concentration.

EXPERIMENTAL SECTION

Protein Model for Folded Domains. For a CG model of folded domains of p53, we used an AICG model developed by Li et al.[37] Each CG particle located on C_α atom represents an amino acid. See Supporting Information (SI) Appendix text for the complete description. AICG was used for the core domain (residues 91−289) and for the TET (residues 326−356), of which the PDB codes of the reference structures are 2XWR[9] and 1AIE,[6] respectively. For the core domain, the solution structure of the DNA-unbound state is very similar to the DNA-bound form.[37] The contact interactions between different subunits were imposed for the TET domains but not for the core domains. The assumption that contact interactions between the core domains are negligible is based on the previous experimental result revealing that there are indeed few contacts between the core domains in the full length p53 (flp53) in solution and that the interactions of these contacts are weak.[39] In the initial structure of the simulations, the TET domains form a tetramer.

Protein Model for Disordered Regions. For a CG model of disordered regions of p53, i.e., the NTD (residues 1−90), the linker (residues 290−325), and the CTD (residues 357−393), we used a model containing the statistical potentials for virtual bond angles and dihedral angles. The statistical potentials were constructed from generic loop structures in PDB. In our previous work, the model reproduced the profiles of small-angle X-ray scattering and NMR residual dipolar coupling of the intrinsically disordered p53 NTD reasonably well.[38] Here we extended this approach for which complete description is in SI Appendix text.

DNA Model. For a DNA model, we used the 3SPN.1 model developed by de Pablo group.[36] This model was calibrated to reproduce the ionic strength dependence of the melting temperature, the persistence length, and the heat capacity profiles of double-strand (ds)DNA. Each nucleotide is represented by three CG particles corresponding to sugar, phosphate, and nitrogenous base. The electrostatic interaction is treated in the same way as in the protein model and is modeled by the Debye−Hückel theory. The sequence-specific interactions in DNA are only modestly included in the current model via sequence-specific base-pair and stacking interactions in the B-type dsDNA. However, we believe that the model accuracy is adequate to investigate searching dynamics on "nonspecific" dsDNA. See SI Appendix text and the original paper for complete description.

The parameters in the three different models above are independently decided so that the individual CG models reproduce some physical properties at room temperature (300 K). Without any recalibration, we can integrate these models. A posteriori comparison with experiments provides validation of the way of integration.

Generic Interchain Interactions. Between different subunits of p53 and between p53 and dsDNA, we included the excluded volume effect and electrostatic interaction. The latter was modeled by the Debye–Hückel theory.

Coarse Grained Simulations. First, we prepared the initial structure of the flp53 and the TetCD construct that contains only TET and CTD with DNA. The initial structure of flp53 was prepared using Modeler.[40] We used 2XWR[9] and 1AIE[6] as the template structure for the core (residues 91–289) and the TET (residues 326–356) domains, respectively, and modeled NTD (residues 1–90), CTD (residues 357–393), and linker region (residues 290–325) as random coil. The initial structure of TetCD was prepared by truncating the N-terminal 325 residues from the full-length model. The random sequence of DNA is prepared: AGTGCAATTGGCAATACATA-CAACTGTCGTCATAATATGCGCGGCCTTGT-CATCGCGGTGCCCCAAGCTCCGCCCTACACGTATGCTCT-GAGTGTGTCTGACTCCTGTCTCAAAATTCATGCG-TAGTCTGGGCCTCTAAGGACTGCAGGGCGCCGAAGTCG-CAACCCGGTGTGCGTGCGACTAGTTGTGCCACGTATGGG.

The protein and the 200bp dsDNA were placed in a box with dimensions of 400 × 750 × 400 Å with the DNA being placed at the center of the box along its *y*-axis. Two ends of dsDNA were constrained. In the initial structure, flp53 and TetCD do not bind to DNA.

The potential energy functions described above were used for simulations. The production runs of the CG simulations were conducted by Langevin dynamics for 1×10^9 MD steps in the case of flp53 and for 3×10^8 MD steps in the case of TetCD with friction coefficient $\gamma = 0.25$ and temperature $T = 300$K. The ion concentration was set to 50, 100, 150, and 200 mM in the case of flp53 and to 1, 50, 100, 150, 200, 250, and 300 mM in the case of TetCD. We recorded the coordinates of p53 and DNA at every 10^4 MD steps and finally collected 1×10^5 frames for flp53 and 3×10^4 frames for TetCD. We neglected first 10^4 frames in posterior analysis because these frames contained initial binding and relaxation processes. We performed all the CG simulations by CafeMol.[41] The structural graphics were prepared with VMD.[42]

Analysis. As the contact definition in Figure 2C, we defined that DBDs contact with DNA if the smallest distance between CG particles of DBDs and DNA is within 7 Å. We also tried 6 and 8 Å as the criteria for a contact. The result, however, did not significantly affected by the criterion.

In the analysis of p53 TetCD rotation around DNA, we prepared local coordinate axis (lx_i, ly_i, and lz_i) (Figure S1); lx_i is defined by $lx_i = r^S_{10i-9} - r^C_{10i-9}$, where r^S_I is the position vector of *I*-th sugar bead of a DNA strand and r^C_I is the position vector of the center of mass of *I*-th basepairing two sugar beads; ly_i is defined by $ly_i = r^C_{10i-1} - r^C_{10i-9}$; and lz_i is defined by the cross product of lx_i and ly_i. We use *i*-th local coordinate axis when *i* minimizes the value $\sum_{k=0}^{9} \| r^C_{pro} - r^C_{10i-9+k} \| /10$, where r^C_{pro} is the position vector of the center of mass of p53 TetCD. The position vector of origin of *i*-th local coordinate is r^C_{10i-9}. Then, we can obtain local coordinates of p53 TetCD centroid (lx, ly, and lz) by a coordinate transformation. Then, the rotation around origin on lx–lz plane approximates the rotation movement of p53 TetCD around DNA.

In order to analyze DNA bending upon p53 TetCD nonspecific binding, we calculated the value $s = \langle \hat{r}_{i+10} - \hat{r}_i \rangle$, where $\hat{r}_i$ is unit vector of $r_i = R_{i-10} - R_i$, where R_i is the position vector of *i*-th sugar bead. If the DNA has extended (bend) conformation in particular local region, the *s* value in this region is high (low).

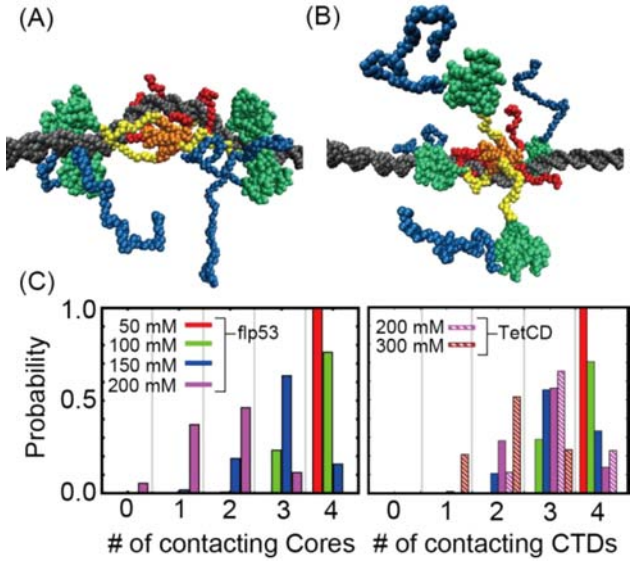


Figure 2. (A,B) Representative snapshots of flp53 in the simulation at (A) 50 and (B) 200 mM ion concentrations. The color scheme is the same as in Figure 1. (C) The probabilities of the number of domains contacting with DNA in the case of the core domain (left) and CTD (right) at 50–200 mM ion. The probability distribution from flp53 simulation is depicted by the solid bars and that from TetCD simulation is depicted by the shaded bars.

RESULTS

Full-Length p53 (flp53) . To characterize search mechanisms of tetrameric flp53 on DNA, we designed binding and sliding simulations of flp53 with nonspecific dsDNA at various salt concentrations. The simulation system contains a tetrameric flp53 complex and a 200 bp dsDNA. The dsDNA was sustained linearly (designated as *y*-axis) by pinning the two ends. The flp53 was initially placed near, but not in contact with, DNA (Figure 1B). In the simulations, flp53 was modeled by a CG protein model, in which one CG particle represents one amino acid. We employed a previously developed AICG model[37] for each of the folded domains (the core and TET domains) and developed a knowledge-based statistical potential for disordered regions (N-terminal domain, linker, and CTD). Four subunits of flp53 were bound in the TET domains by the AICG contact terms. For DNA, we employed a CG model, in which each nucleotide is represented by three particles, each for sugar, phosphate, and base.[36] This DNA model has been developed so that it approximates physical property of the B-type dsDNA. Between the four subunits of flp53 as well as between flp53 and dsDNA, general electrostatic and excluded volume interactions were taken into account. No specific structure-based interactions were included between flp53 and DNA (see the Experimental Section and SI for more details).

Using this setup, we conducted binding and sliding simulations at 50, 100, 150, and 200 mM salt (monovalent ion) concentrations. At all salt concentrations tested, we observed that flp53 spontaneously binds to DNA and starts sliding on it.

At 50 mM, after the binding and subsequent relaxation, flp53 took a characteristic structure binding DNA at four core domains and four CTDs: four CTDs bound DNA at a center, while each pair of core domains flanked the CTD (Figure 2A). The linkers (yellow) between the core (green) and the CTD (red) took rather straight conformations (Figure 2A). Notably, this form is very similar to one of the forms found in a previous simulation study.[31] Close agreement of completely independent simulations supports this particular form at relatively low salt concentration. The linkers themselves are not very stiff, and thus straight conformations found here are probably caused by electrostatic repulsions between four core domains and CTDs. As we mentioned in the Experimental Section, we did not impose any contact interactions between core domains.

dx.doi.org/10.1021/ja305369u | J. Am. Chem. Soc. 2012, 134, 14555–14562

Therefore the dimer-like conformation observed is not due to the direct attraction. We interpret that this conformation is stabilized by the electrostatic repulsion between the core domains and the CTDs and the electrostatic attraction between the core domains and the DNA. At 50 mM, once bound, all core domains and CTDs remained bound to DNA all the time (Figure 2C, red bars, and movie S1).

At higher salt concentrations, we found the binding modes are gradually modulated. Figure 2C shows statistics of the numbers of DNA-bound core domains and CTDs, both of which decreased with the salt concentration, as expected. Specifically, at 200 mM, on average ~3 CTDs were bound to DNA, while only 1 or 2 core domains remained bound to DNA for most of time (Figure 2B,C, purple bars). There is a small, but non-negligible, probability of all core domains dissociated from DNA. In contrast, at least one CTD was bound to DNA at anytime. This result indicates that nonspecific DNA binding affinity of CTD is higher than that of the core domain and that p53 might completely dissociate from DNA without CTD. We speculate that the reason why the CTD has higher affinity to DNA is that this region is intrinsically disordered. In general, the disordered tail has an extended interface of positively charged residues, which makes the tighter binding to DNA possible.[43]

Importantly, the DNA binding of the core domain at 200 mM is highly dynamic. We plotted the time series of the distance between a core domain and DNA (Figure 3A, green).

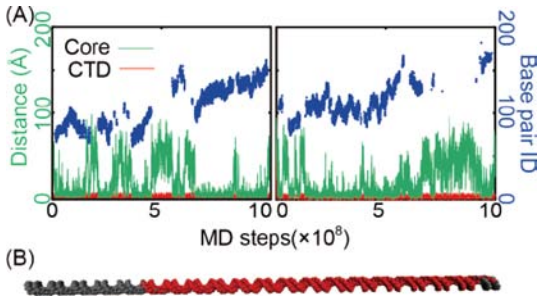


Figure 3. Distances and contacts of flp53 with DNA. (A) Two time series of the distance between the core and DNA (green) and between CTD and DNA (red), and the base pair index which is the nearest from the core domain (blue). The index is plotted only when the core is in contact with DNA. (B) Nucleotide CG beads that contacted (within 7 Å) at least once with any of the p53 core domains during a trajectory is colored red.

Clearly, the core domain repeated dissociation from and association to DNA. Once dissociated, the core domain went as much as 60 Å away from DNA. The dissociation and association of four core domains did not look correlated. After dissociation, the core rebinds to DNA at the site no more than 15bp away from the dissociation site (Figure 3A, blue). In contrast, the CTD remained close to DNA even when it transiently lost direct contact with DNA. Since the CTD is next to the TET domain, four of the CTDs are unavoidably close to each other which seems to prevent them far from DNA. This is different from the core domain which is connected to the TET domain via the linker region making the four core domains well separated.[44] From these, we conclude that flp53 slides along DNA by its CTDs, whereas core domains repeat dissociation and association at 200 mM ion. The binding mode found here matches perfectly with a recent single molecule experiment[14] as well as many other experimental observations.[22−24] Core domains showed dissociation and association repeatedly, which is reminiscent of hopping (movie S2). However, since at least one CTD is always bound to DNA, the hopping of the core domain is restricted by the length of the linker, and thus we call it "restricted hopping". Importantly, during the restricted hopping of the core, most of the DNA bases were "touched" by at least one core domain (Figure 3B), suggesting that even though the core domains often dissociate from DNA, they do not overlook their cognate binding sites.

We next investigated the nonspecific DNA binding surfaces of the core domain and CTD. Figure 4 plots the probability of

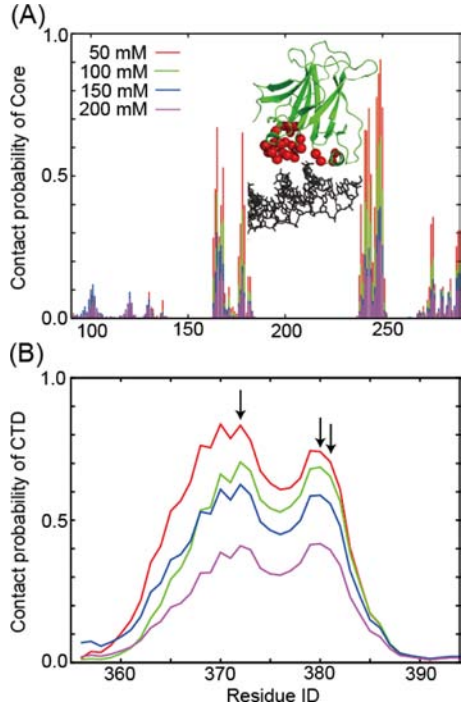


Figure 4. Contact probabilities of flp53 residues with DNA at 50−200 mM ion. (A) Residues in the core domain. (inset) Residues that contact with DNA at a probability higher than 0.2 in simulations are shown by red beads in the crystal structure (3KMD) of the core domain (green) in complex with DNA (gray). (B) Residues in the CTD. The black arrows indicate highly conserved lysine residues 373, 381, and 382. We consider that each residue contacts with DNA if the distance between the residue and that in DNA is within 7 Å.

each residue in the core domain and in CTD (Figure 4A,B, respectively) contacting with DNA at 50−200 mM salt concentrations. Here we defined the contact by the criterion that the smallest distance between a residue and DNA is within 7 Å. We see that, although the contact probability decreases with the ion concentration, the peak positions (i.e., key residues) do not change in both the cases of the core domain and CTD. This result shows that, although the binding affinity depends on the ion concentration, the binding surface does not.

As for the binding surface of the core domain, the residues with the high probability cluster near the specific DNA binding surface (inset of Figure 4A). This result demonstrates that the nonspecific binding surface in the core domain is similar to the specific binding surface found in the crystal structure.[8] This agreement facilitates the search of the cognate site during sliding on DNA. Note that this condition is fulfilled even at 200 mM where the core domain exhibits restricted hopping.

dx.doi.org/10.1021/ja305369u | J. Am. Chem. Soc. 2012, 134, 14555−14562

As for CTD, there are two regions with high contact probabilities (Figure 4B). This is consistent with the previous experimental result where two regions were identified to contact with nonspecific DNA using NMR.[11] This experiment also showed that the chemical shifts of highly conserved lysine residues 373, 381, and 382 are greatly perturbed by DNA. Consistent with this, these residues have a relatively high contact probability in the simulation. These results indicate that the model in this work well reproduces the nonspecific binding surface of CTD.

p53 TetCD. Since the above simulations of flp53 clarified that the CTD is responsible for the sliding at physiological salt concentration, we now focus on the tetrameric p53 TetCD, fragments containing TETs and CTDs. Due to the smaller size, we could simulate TetCD on a much longer time scale than flp53, which enabled us detailed analysis of the diffusion constant along DNA. We performed simulations of p53 TetCD with dsDNA at 1−350 mM monovalent ion concentrations. The dsDNA was sustained along y-axis in the same manner as above. The initial placement of TetCD was similar to that for flp53. We employed the same simulation models for proteins, DNA, and their interactions.

For salt concentrations of 1−300 mM, we observed spontaneous binding and sliding of TetCD on DNA [Figure 5A and movie S3 (at 50 mM) and Figure 5B and movie S4 (at 200 mM)]. At 350 mM, after binding on DNA, TetCD repeated dissociation and association. Up to 300 mM, on the other hand, no dissociation event was observed (brown shaded bars in Figure 2C). Hereafter, we focus on the sliding motions at 1−300 mM.

Figure 5C plots motions of TetCD along DNA (y coordinate of the center of the TET domains) as a function of time. We see apparently unbiased fluctuation in y axis of TetCD. We also note that the absolute value of y did not exceed 200 Å, confirming that TetCD did not approach the termini of DNA (∼ ±350 Å). Thus we can neglect terminal effects on the diffusion.

To test the one-dimensional diffusion, we calculated the mean square displacement (MSD) of the center of TET, defined by $\mathrm{MSD}(\Delta t) = \langle (y_{i+\Delta t} - y_i)^2 \rangle$ (Figure 5D) as a function of time difference Δt for every salt concentration that showed diffusion motions. For all the salt concentrations, we see clearly straight lines proving that the sliding of TetCD along y-axis is indeed diffusion. The MSD is connected to the one-dimensional diffusion constant D_1 by the equation $\mathrm{MSD}(\Delta t) = 2D_1\Delta t$. We see that D_1 depends on the salt concentration. In Figure 5E, we plotted D_1 against the salt concentration. Unexpectedly, the D_1 data (Figure 5E) exhibited two distinct behaviors: For 1−200 mM, D_1 monotonically (approximately in quadratic form) increases, whereas D_1 is fairly constant for 200−300 mM. The behavior at 200−300 mM is qualitatively consistent with the recent single molecule experiments where the diffusion constant of TetCD did not depend on the ion concentration.[14] Figure 5E also showed that the diffusion constant of p53 TetCD without DNA (black line in Figure 5E) is slightly higher than the highest diffusion constant of p53 TetCD with DNA, implying that the excluded volume effect of DNA slightly hinders the diffusion to some extent.

We next address whether the TetCD sliding is along the DNA helical pitch (i.e., rotation-coupled sliding)[15] or not (i.e., rotation-uncoupled sliding).[25] For the purpose, we first defined the rotation angle θ that monitors rotational motions of TetCD around the dsDNA long axis (see Experimental Section). If

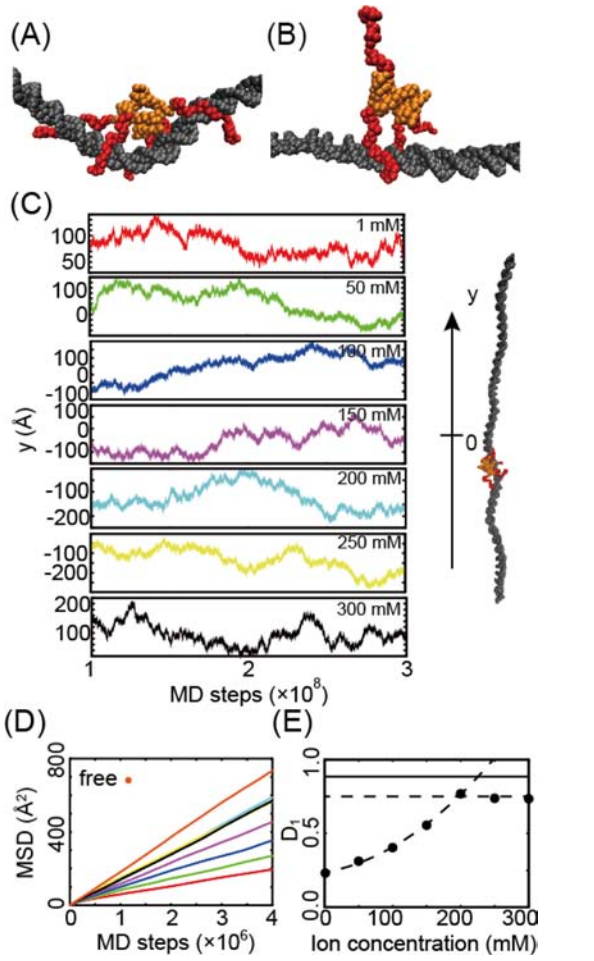


Figure 5. Sliding motions of p53 TetCD. (A,B) Representative structures in the simulation trajectory at (A) 50 and (B) 200 mM ions. The color scheme is same as in Figure 1A,B. (C) Time series of y coordinate of the center of TetCD at 1−300 mM ion. (right) The definition of y-axis is given with the representative structure of p53 TetCD and DNA. (D) The MSD of y coordinate of the center of TetCD at 1−300 mM ion. The color scheme for each ion concentration is the same as in (C). "Free" represents the result of TetCD without DNA. (E) The diffusion constant of y coordinate of the center of p53 TetCD at 1−300 mM ion. The diffusion constants at 1−200 mM ion and those at 200−300 mM are fitted by quadratic and linear functions, respectively (dashed line). The diffusion constant of p53 TetCD without DNA is depicted with a black line.

TetCD diffusion on DNA is strictly along helical pitch of DNA, such as major and minor grooves or phosphate backbone, the motions in y must be coupled with those in θ. We projected the simulation trajectory at 1 and 200 mM on the θ−y plane (Figure 6A). These figures apparently show the strong correlation between θ and y (the correlation coefficient R = −0.98) at 1 mM ion and virtually no correlation (R = −0.15) at 200 mM. In order to investigate the coupled rotation and diffusion more rigorously, we plotted the difference in rotation angles $\Delta\theta$ (for the duration of $1-10^7$ MD steps) with the difference in y coordinate (Δy) in Figure 6B for a number of pairs of snapshots in the trajectories at various salt concentrations. In the figures, the dashed line corresponds to the helical pitch of the ideal B-type DNA, 34 Å displacement in 360° rotation (the slope of the regression line is −0.094). We see that, at 1 mM, $\Delta\theta$ and Δy are well coupled (R = −0.680 and the slope is −0.106), while at higher salt concentrations,

dx.doi.org/10.1021/ja305369u | J. Am. Chem. Soc. 2012, 134, 14555−14562

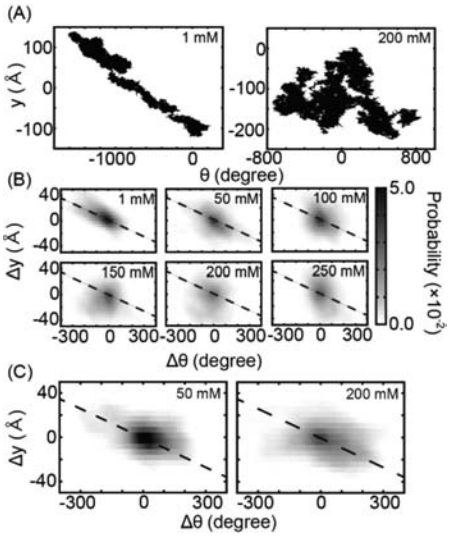


Figure 6. Correlation between lateral (y) and rotational (θ) motions of p53. (A) Representative trajectories of TetCD at 1 and 200 mM ion on $\theta-y$ plane. (B,C) The two-dimensional probability distribution of change in θ ($\Delta\theta$) and in y (Δy) for TetCD simulations at 1−250 mM ion (B) and for flp53 simulations at 50 and 200 mM ions (C). The dashed line corresponds to the helical pitch of the ideal B-type DNA (34 Å displacement coupled to 360° rotation).

such a coupling is quickly diminished. At 200 mM, we see the reduced correlation between $\Delta\theta$ and Δy ($R = -0.130$ and the slope is -0.024), suggesting that the TetCD sliding at this condition is not strongly affected by the helicity of DNA. The situation is similar in the case of flp53 (Figure 6C). At 50 mM, $\Delta\theta$ and Δy are modestly coupled ($R = -0.379$ and the slope is -0.03926), while at 200 mM, such a coupling is reduced ($R = -0.249$ and the slope is -0.02645). Thus, we conclude that the sliding of p53 at very low ion concentration is rotation coupled, while that at physiological condition is rotation uncoupled.

Finally, we address dsDNA bending caused by the binding of p53 TetCD. It has been suggested that specific or nonspecific protein binding tends to bend DNA.[45,46] For example, Bustamante et al. proposed that DNA bent by nonspecific binding of Cro can facilitate recognition of its target site.[46] We defined and computed at 200 mM a bending score S which takes unity for a straight dsDNA and decreases as dsDNA bends (Figure 7). When no protein is bound to dsDNA, the S values were 0.93−0.95. In contrast, when p53 TetCD bound to the DNA, the S values clearly decreased near the protein bound region. Thus, we confirmed that DNA is bent by the binding of p53 TetCD. Note that not absolute but relative values of S do have meaning because both ends of the DNA are fixed in order

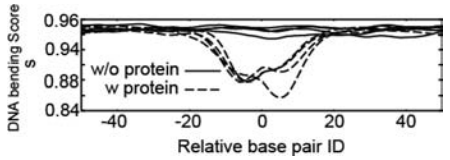


Figure 7. The DNA bending score S with and without bound p53. The horizontal axis is the base pair ID relative to the base pair that is the nearest from p53 TetCD. In each of four dashed lines, the S value is averaged over the snapshots from 1.0 −1.5 × 10^8, 1.5 −2.0 × 10^8, 2.0−2.5 × 10^8, and 2.5−3.0 × 10^8 MD steps, respectively.

to investigate one-dimensional diffusion of p53, and so the absolute values are affected by this setup.

■ DISCUSSION AND CONCLUSION

Since the classic work of von Hippel and Berg, it has been thought that one-dimensional diffusion can be classified into two mechanisms (sliding and hopping) and that the one-dimensional diffusion constant D_1 for protein sliding on DNA is independent of the salt concentration,[26] while the so-called hopping mechanism leads to salt-dependent search rates. Often, observation of salt-independent diffusion constant is interpreted as the evidence of sliding motions. Our simulations here suggested this is not always right. In Figure 5E, we showed that the D_1 of p53 TetCD sliding on DNA depends on ion concentration of 1−200 mM, though TET domain kept bound to DNA in this ion concentration range. We note that von Hippel and Berg's theory is based on a mesoscopic model where DNA is a simple electrolyte rod without molecular structural details. At an atomic level, of course, even along any of helical rails of dsDNA (major and minor grooves or phosphates), the interaction surface has some ruggedness which modulates electrostatic interactions. Thus, it is not too surprising even if sliding-based D_1 depends on ion concentration. We speculate that this salt-dependent D_1 in the limit of low salt concentration is a general feature for many DNA-binding proteins, although more works are necessary to prove it. This result also suggests that the diffusion along DNA becomes quite slow when the interaction between a protein and DNA is strong.

In Figure 5E, we also showed that the D_1 of p53 TetCD sliding along DNA does not depend on the ion concentration at a middle range of monovalent ion concentration (200−300 mM). This result is qualitatively consistent with the recent single molecule experiments measured at 2 mM divalent ion and sub-100 mM monovalent ion.[14] This finding indicates that the ion concentration range of 200−300 mM in this simulation corresponds to a physiological ion concentration. Here, the following two technical problems in simulations make quantitative comparison difficult, leading to inconsistent values of ion concentration between the simulation and the real systems: First, our treatment of salt via Debye−Hückel approximation does not allow us to include divalent ions, which is known to have very large impact. Second, even for monovalent ions, the Debye−Hückel approximation is less accurate for a high ion concentration. Qualitatively, in experiments, a monovalent ion higher than 100 mM together with a 2 mM divalent ion led to the TetCD dissociation from DNA. Similarly, in simulations, monovalent ions higher than 300 mM led to dissociation from DNA. Thus, we speculate that these two situations may be qualitatively similar, corresponding to just below the threshold of the dissociation. In this ion concentration range, p53 slid along DNA with its CTD, whereas the core domains exhibited restricted hopping in the simulation (movie S2). This sliding mechanism is in harmony with that proposed from the single molecule experiment.[14] Experimentally, the hopping at the core domains was proposed based on the ion concentration-dependent diffusion constant. However, as above, the ion concentration dependence alone is not sufficient to prove the hopping mechanism. Yet, our simulations visualized the restricted hopping of the core domains, supporting the view suggested.

As salt concentration increased, the one-dimensional diffusion constant D_1 of TetCD along DNA increased up to

dx.doi.org/10.1021/ja305369u | J. Am. Chem. Soc. 2012, 134, 14555−14562

200 mM, at which it reached a plateau. Salt concentration higher than 300 mM led to the dissociation of TetCD from DNA. Thus the plateau, corresponding to 200−300 mM, is the maximal sliding-based D_1. Comparison of simulations and experiments[14] suggests that TetCD achieves this maximal diffusion constant near physiological salt concentration condition. Too strong interactions between a protein and DNA make diffusion slow. Too weak interactions make a protein dissociated from DNA. We speculate that p53 TetCD has evolved to realize the fastest-possible diffusion along DNA.

The correlation analysis of the one-dimensional diffusion along DNA and rotation around it (Figure 6BC) showed that the diffusion of p53 is virtually uncoupled to the rotation at 200 mM and higher. These results suggest that the mechanism of sliding mediated by its CTD is the rotation-uncoupled sliding at physiological ion concentration. Since p53 CTD is disordered and is not the domain that specifically binds to DNA, we should not readily generalize this finding to other DNA binding proteins. It is very interesting to uncover whether DNA-searching proteins, in general, diffuse on DNA via rotation-coupled or rotation-uncoupled sliding. The latter has its unique advantage that they can bypass some barriers on DNA made by other proteins, thus enhancing the search efficiency.[25]

The average number of CTD contacting to DNA in flp53 simulation is slightly smaller than that in TetCD simulation (Purple solid bars and purple shaded bars in Figure 2C). This result indicates that the existence of core domains slightly weakens the binding of the CTD. Besides, although we could not observe significant correlation between the number of CTDs contacting DNA and the number of contacting cores (Figure S2), we found weak correlation between four CTDs. That is, if we assume the binding of each of four CTDs is independent, the probability of each CTD in TetCD contacting to DNA at 300 mM is calculated to be 0.487. Under this assumption, the probabilities of 0, 2, and 4 CTDs contacting to DNA are calculated to be 0.069, 0.375, and 0.057, respectively. However, the probabilities observed in the simulation are actually 0.000, 0.511, and 0.000, respectively (brown shaded bars in Figure 2C). Therefore, the probabilities of 0 and 4 CTDs contacting to DNA are under expressed, and the probability of 2 CTDs contacting to DNA is overexpressed. This result indicates the four CTDs coordinately inhibit too strong and too weak binding. This could facilitate the efficient sliding along DNA.

The fact that p53 has two DBDs and that each of them has the distinct role, i.e., the core domain for specific and the CTD for nonspecific binding, implies the importance of the flexible linker region present between the two DBDs. Only the linker can control relative position and orientation of two DBDs. Moreover, the linker is between the core and TET domains, which makes four core domains moving independently. This is in contrast with the four CTDs directly linked by the TET domains. Since the linker region has not had much attention, more works on it will be important.

Several recent studies suggest that transcription factors may have two distinct modes (search and recognition modes) in order to solve the paradox between search speed and recognition stability.[47,48] In the snapshot from flp53 simulation at physiological ion concentration (Figure 2B and movie S2), the arrangement of core domains is quite different from that in the specific complex crystal structure,[8] invoking the search mode. Much of the work is desired to clarify this point. Future studies should confirm this binding mode and the quite large conformational change needed upon recognizing specific sequence on DNA.

Future studies should also improve the treatment of electrostatic interaction, especially, for the protein−DNA interaction. Specifically, the usage of the invariant dielectric constant of 78 and the Debye−Hückel approximation may need to be refined. Also, unit positive charges in all Lys, Arg, and His residues and unit negative charges in all Asp and Glu are a rather simple approximation. We can possibly improve it by rearranging the charge position and amount so that the electrostatic field calculated from atomic structure is best reproduced.[49,50] Fortunately, in this work and a series of works in the Levy group, the simple treatment does not seem to cause any critical problem. However, in order to overcome the limit of its application, more sophisticated treatment is desired.

Albeit much room for improvement, CG simulation methods used here as well as a series of works in the Levy group[43] are quite generic and can be applied to many other DNA binding proteins.[51,52] For such a purpose and others, we have been developing a generic CG biomolecular simulation software CafeMol.[41] Indeed, all the simulations here were performed with CafeMol.

■ ASSOCIATED CONTENT

Supporting Information

Detailed simulation method and procedure. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

takada@biophys.kyoto-u.ac.jp

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Sir Alan Fersht for advice and valuable comments. This work was supported by Grant-in-Aid for Scientific Research, by Research and Development of the Next-Generation Integrated Simulation of Living Matter of the Ministry of Education Culture, Sports, Science, and Technology (MEXT) and by the Global COE Program "Formation of a Strategic Base for Biodiversity and Evolutionary Research: from Genome to Ecosystem" of the MEXT.

■ REFERENCES

(1) Levine, A. J.; Oren, M. *Nat. Rev. Cancer* **2009**, *9*, 749−58.
(2) Joerger, A. C.; Fersht, A. R. *Annu. Rev. Biochem.* **2008**, *77*, 557−82.
(3) Cho, Y.; Gorina, S.; Jeffrey, P. D.; Pavletich, N. P. *Science* **1994**, *265*, 346−55.
(4) Jeffrey, P. D.; Gorina, S.; Pavletich, N. P. *Science* **1995**, *267*, 1498−502.
(5) Clore, G. M.; Omichinski, J. G.; Sakaguchi, K.; Zambrano, N.; Sakamoto, H.; Appella, E.; Gronenborn, a M. *Science* **1994**, *265*, 386−91.
(6) Mittl, P. R. E.; Chène, P.; Grütter, M. G. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1998**, *54*, 86−9.
(7) Cañadillas, J. M. P.; Tidow, H.; Freund, S. M. V.; Rutherford, T. J.; Ang, H. C.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 2109−14.
(8) Chen, Y.; Dey, R.; Chen, L. *Structure* **2010**, *18*, 246−56.
(9) Natan, E.; Baloglu, C.; Pagel, K.; Freund, S. M. V.; Morgner, N.; Robinson, C. V.; Fersht, A. R.; Joerger, A. C. *J. Mol. Biol.* **2011**, *409*, 358−68.

(10) Bell, S.; Klein, C.; Müller, L.; Hansen, S.; Buchner, J. *J. Mol. Biol.* **2002**, *322*, 917−927.
(11) Weinberg, R. L.; Freund, S. M. V.; Veprintsev, D. B.; Bycroft, M.; Fersht, A. R. *J. Mol. Biol.* **2004**, *342*, 801−11.
(12) Arbely, E.; Natan, E.; Brandt, T.; Allen, M. D.; Veprintsev, D. B.; Robinson, C. V.; Chin, J. W.; Joerger, A. C.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 8251−6.
(13) Berg, O. G.; Winter, R. B.; von Hippel, P. H. *Biochemistry* **1981**, *20*, 6929−48.
(14) Tafvizi, A.; Huang, F.; Fersht, A. R.; Mirny, L. A.; van Oijen, A. M. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 563−8.
(15) Blainey, P. C.; Luo, G.; Kou, S. C.; Mangel, W. F.; Verdine, G. L.; Bagchi, B.; Xie, X. S. *Nat. Struct. Mol. Biol.* **2009**, *16*, 1224−9.
(16) Kabata, H.; Kurosawa, O.; Arai, I.; Washizu, M.; Margarson, S. A.; Glass, R. E.; Shimamoto, N. *Science* **1993**, *262*, 1561−3.
(17) Harada, Y.; Funatsu, T.; Murakami, K.; Nonoyama, Y.; Ishihama, A.; Yanagida, T. *Biophys. J.* **1999**, *76*, 709−15.
(18) Wang, Y.; Austin, R.; Cox, E. *Phys. Rev. Lett.* **2006**, *97*, 048302.
(19) Gorman, J.; Chowdhury, A.; Surtees, J. A.; Shimada, J.; Reichman, D. R.; Alani, E.; Greene, E. C. *Mol. Cell* **2007**, *28*, 359−70.
(20) Kim, J. H.; Larson, R. G. *Nucleic Acids Res.* **2007**, *35*, 3848−58.
(21) Dikic, J.; Menges, C.; Clarke, S.; Kokkinidis, M.; Pingoud, A.; Wende, W.; Desbiolles, P. *Nucleic Acids Res.* **2012**, *40*, 4064−70.
(22) McKinney, K.; Mattia, M.; Gottifredi, V.; Prives, C. *Mol. Cell* **2004**, *16*, 413−24.
(23) Liu, Y.; Kulesz-Martin, M. F. *Cell Death Differ.* **2006**, *13*, 881−4.
(24) Tafvizi, A.; Huang, F.; Leith, J. S.; Fersht, A. R.; Mirny, L. A.; van Oijen, A. M. *Biophys. J.* **2008**, *95*, L01−3.
(25) Kampmann, M. *J. Biol. Chem.* **2004**, *279*, 38715−20.
(26) Gorman, J.; Greene, E. C. *Nat. Struct. Mol. Biol.* **2008**, *15*, 768−74.
(27) Iwahara, J.; Schwieters, C. D.; Clore, G. M. *J. Am. Chem. Soc.* **2004**, *126*, 12800−8.
(28) Kalodimos, C. G.; Biris, N.; Bonvin, A. M. J. J.; Levandoski, M. M.; Guennuegues, M.; Boelens, R.; Kaptein, R. *Science* **2004**, *305*, 386−9.
(29) Vuzman, D.; Hoffman, Y.; Levy, Y. *Pac. Symp. Biocomput.* **2012**, *17*, 188−199.
(30) Givaty, O.; Levy, Y. *J. Mol. Biol.* **2009**, *385*, 1087−97.
(31) Khazanov, N.; Levy, Y. *J. Mol. Biol.* **2011**, *408*, 335−55.
(32) Levy, Y.; Onuchic, J. N.; Wolynes, P. G. *J. Am. Chem. Soc.* **2007**, *129*, 738−9.
(33) Marcovitz, A.; Levy, Y. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 17957−62.
(34) Tóth-Petróczy, Á.; Simon, I.; Fuxreiter, M.; Levy, Y. *J. Am. Chem. Soc.* **2009**, *131*, 15084−5.
(35) Vuzman, D.; Levy, Y. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 21004−9.
(36) Sambriski, E. J.; Schwartz, D. C.; de Pablo, J. J. *Biophys. J.* **2009**, *96*, 1675−90.
(37) Li, W.; Wolynes, P.; Takada, S. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 3504−9.
(38) Terakawa, T.; Takada, S. *Biophys. J.* **2011**, *101*, 1450−8.
(39) Tidow, H.; Melero, R.; Mylonas, E.; Freund, S. M. V.; Grossmann, J. G.; Carazo, J. M.; Svergun, D.; Valle, M.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 12324−9.
(40) Mart, M. A.; Stuart, A. C.; Roberto, S.; Melo, F.; Sali, A. *Biochemistry* **2000**, *29*, 291−325.
(41) Kenzaki, H.; Koga, N.; Hori, N.; Kanada, R.; Li, W.; Okazaki, K.; Yao, X. Q.; Takada, S. *J. Chem. Theory Comput.* **2011**, *7*, 1979−89.
(42) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.
(43) Vuzman, D.; Levy, Y. *Mol. Biosyst.* **2012**, *8*, 47−57.
(44) Shoemaker, B. A.; Portman, J. J.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8868−73.
(45) Wu, H. M.; Crothers, D. M. *Nature* **1984**, *308*, 509−13.
(46) Erie, D. A.; Yang, G.; Schultz, H. C.; Bustamante, C. *Science* **1994**, *266*, 1562−6.
(47) Slutsky, M.; Mirny, L. A. *Biophys. J.* **2004**, *87*, 4021−35.
(48) Mirny, L.; Slutsky, M.; Wunderlich, Z.; Tafvizi, A.; Leith, J.; Kosmrlj, A. *J. Phys. A* **2009**, *42*, 434013.
(49) Gabdoulline, R. R.; Wade, R. C. *J. Phys. Chem.* **1996**, *100*, 3868−78.
(50) Beard, D. A.; Schlick, T. *Biopolymers* **2001**, *58*, 106−15.
(51) Hyeon, C.; Thirumalai, D. *Nat. Commun.* **2011**, *2*, 487−97.
(52) Takada, S. *Curr. Opin. Struct. Biol.* **2012**, *22*, 130−7.