JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Molecular Dynamics Simulations of Ternary Complexes: Comparisons of LEAFY Protein Binding to Different DNA Motifs

Hua Wan,[†] Shan Chang,*[,‡] Jian-ping Hu,[§] Yuan-xin Tian,[∥] and Xu-hong Tian[†]

[†]College of Informatics, South China Agricultural University, Guangzhou 510642, China

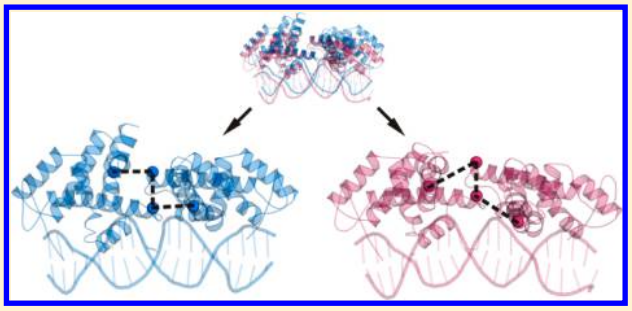
[‡]Institute of Bioinformatics and Medical Engineering, School of Electrical and Information Engineering, Jiangsu University of Technology, Changzhou 213001, China

[§]Faculty of Biotechnology Industry, Chengdu University, Chengdu 610106, China

[∥]School of Pharmaceutical Sciences, Southern Medical University, Guangzhou 510515, China

Supporting Information

ABSTRACT: LEAFY (LFY) is a plant-specific transcription factor, with a variety of roles in different species. LFY contains a conserved DNA-binding domain (DBD) that determines its DNA-binding specificity. Recently, the structures of the dimeric LFY-DBD bound to different DNA motifs were successively solved by X-ray crystallography. In this article, molecular dynamics (MD) simulations are employed to study two crystal structures of DNA-bound LFY protein from angiosperms and the moss *Physcomitrella patens*, respectively. The comparison of stabilities of the two systems is consistent with the experimental data of binding affinities. The calculation of hydrogen bonds showed that position 312 in LFY determines the difference of DNA-binding specificity. By using principal component analysis (PCA) and free energy landscape (FEL) methods, the open-close conformational change of the dimerization interface was found to be important for the system stability. At the dimerization interface, the protein−protein interaction has multiple influences on the cooperative DNA binding of LFY. The following analysis of DNA structural parameters further revealed that the protein−protein interaction contributes varying roles according to the specific DNA-binding efficiency. We propose that the protein−protein interaction serves a dual function as a connector between LFY monomers and a regulator of DNA-binding specificity. It will improve the robustness and adaptivity of the LFY-DNA ternary structure. This study provides some new insights into the understanding of the dynamics and interaction mechanism of dimeric LFY-DBD bound to DNA at the atomic level.



1. INTRODUCTION

LEAFY (LFY) transcription factor exists in all land plants and charophyte algae. Its sequence shows a high level of sequence conservation.[1] This conserved protein plays diverse roles in different species. For instance, LFY in angiosperms is an essential regulator of floral development.[2,3] In the moss *Physcomitrella patens*, LFY controls the cell division.[4] The LFY's functional variation across species is linked to the molecular evolution of LFY. Consequently, a series of studies focused on investigating the target genes of LFY,[5,6] the mode of action,[7] the interacting partners,[8,9] the structure,[10,11] the evolutionary model,[11−13] etc. The information is helpful to understand the molecular mechanisms underlying functional change of LFY, and in turn, provides valuable evolutionary insights of plants that differ from those of animals.

Early work at the genomic scale indicated that the LFY DNA-binding domain (DBD) is primarily responsible for its divergence in function.[1] The LFY-DBD encompasses the residues from the highly conserved C domain of LFY. The key amino acid substitutions in the LFY-DBD were shown to change its DNA-binding specificity. To further reveal the structural basis for sequence-specific recognition, in 2008 Hames et al. solved the crystal structures of LFY from *Arabidopsis thaliana* (AtLFY) in complex with DNA half-site sequences at 2.1 and 2.3 Å resolutions, respectively.[10] It was found that LFY-DBD binds DNA as a dimer. This cooperative DNA-binding mode of LFY was suggested to be important for binding affinity. In 2013, based on the crystallographic data of 2008, Chahtane et al. generated a LFY allele that has an altered dimerization interface in vivo.[14] This study broke new ground by revealing a novel role for LFY in enabling meristem outgrowth. They confirmed the importance of LFY monomer−monomer contacts and tested the effect of the mutations at the dimerization interface on LFY DNA-binding specificity. In 2014, Sayou et al. reported the crystal structure of PpLFY1, a LFY homologue from the moss *Physcomitrella patens*, bound to another type of DNA half-site sequence at 2.3 Å resolution.[11] PpLFY1 also dimerizes upon DNA binding, exhibiting the similar DNA-binding mode to that of AtLFY. Through EMSA

Received: November 26, 2014
Published: March 3, 2015

DOI: 10.1021/ci500705j
J. Chem. Inf. Model. 2015, 55, 784−794

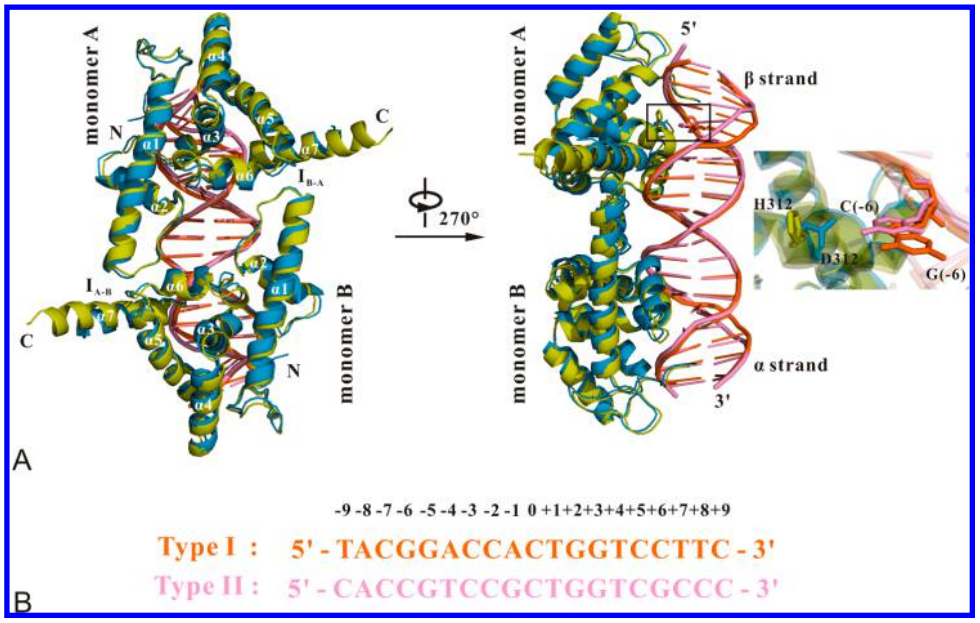


Figure 1. Structures of AtLFY-DBD (yellow) bound to DNA motif I (orange) and PpLFY1-DBD (blue) bound to DNA motif II (pink). (A) LFY dimer bound to DNA duplexes (α and β strands). Each monomer contains seven helices labeled as α1 to α7 (left). The substantially different interface residues and nucleotides are represented as sticks (right). (B) LFY-binding sites of the two DNA motifs.

analysis, AtLFY and PpLFY1, as representative LFY proteins from two species, show different preferences for two types of DNA motifs.[11] These important structural data provided some new clues to the study of LFY.

In addition to experiments, theoretical approaches were also applied to explore the molecular basis of LFY function. By using the feedback circuit analysis based on the gene network model, Mendoza et al. simulated and predicted different patterns of LFY gene expression found in the floral organs *Arabidopsis*.[15] The flowering time control network was built to capture the dynamics of the floral transition.[16] The result showed that LFY positively regulates its own expression, leading to a stable flowering state. To obtain accurately in vivo LFY binding sites in the *Arabidopsis thaliana* genome, Moyroud et al. applied a biophysical model of protein–DNA interactions to predict the binding affinity of LFY to DNA.[17] These theoretical works enhance our understanding of mechanisms of regulation by LFY at the genomic level. Nevertheless, the crystal structures of LFY in complex DNA have not been simulated systemically. The crystallographic studies[10,11] showed that LFY-DNA binding is mediated by a combination of protein–DNA and protein–protein interactions. In particular, the protein–protein interaction was suggested to govern the cooperative DNA-binding mode.[10] LFY proteins bound to different DNA motifs underwent changes of the DNA-binding specificity.[11] The LFY-bound DNA is largely B-form, and the DNA deformation was often associated with the base-specific recognition and transcription activation.[10,18,19] Thus, some detailed questions still need to be answered. What interactions at the atomic level are formed at the LFY protein–protein and protein–DNA interfaces? How does LFY change the specificity when binding to different DNA motifs? Is LFY DNA-binding specificity associated with the LFY dimerization interface? What structural changes occur along the DNA half-site sequences? Are there any influences on the structural change of DNA made by the cooperative DNA-binding mode, including both protein–DNA and protein–protein interactions?

In this paper, in order to answer the above issues, the two crystal structures of DNA-bound AtLFY and PpLFY1 were studied by molecular dynamics (MD) simulations. The principal component analysis (PCA) and free energy landscape (FEL) methods were applied to probe the functional dynamics of the two systems. To explain the conformational changes, the crucial interactions were investigated at the protein–DNA and protein–protein interfaces. We also described the difference of DNA-binding specificity between the two systems. Furthermore, the LFY-bound DNA structural parameters were analyzed at the base-pair (bp) level. By exploring the relationship between the central 3-bp spacer of half sites and the dimerization interface of LFY, we proposed dual roles of the protein–protein interaction in the cooperative DNA-binding mode.

2. SYSTEMS AND METHODS

2.1. Structures of LFY Dimer Bound to Different DNA Motifs. The two crystal structures of AtLFY-DBD bound to DNA motif I[10] (PDB code: 2VY1) and PpLFY1-DBD bound to DNA motif II[11] (PDB code: 4BHK) were obtained from the Protein Data Bank. As shown in Figure 1, the two complex structures show high similarity. LFY protein binds DNA as a dimer, and each monomer contains seven helices connected by short loops (see Figure 1A). The DNA motifs I and II both consist of 8-bp inverted half sites separated by three nucleotides (see Figure 1B). The central 3-bp spacer is associated with the protein–protein interaction at the dimerization interfaces I_{A-B} and I_{B-A} (see Figure 1A), which is required for the cooperative DNA binding in the two dimeric LFY-DBD structures.[11] The main difference between the two structures is that relative to AtLFY (yellow), PpLFY1 (blue) makes an additional base-specific contact with DNA, which is mediated by position 312 in monomer A. The different interface residues and nucleotides are highlighted in stick model (see Figure 1A). For convenience, the two systems were referred to as the type I and type II systems, respectively.

DOI: 10.1021/ci500705j
J. Chem. Inf. Model. 2015, 55, 784–794

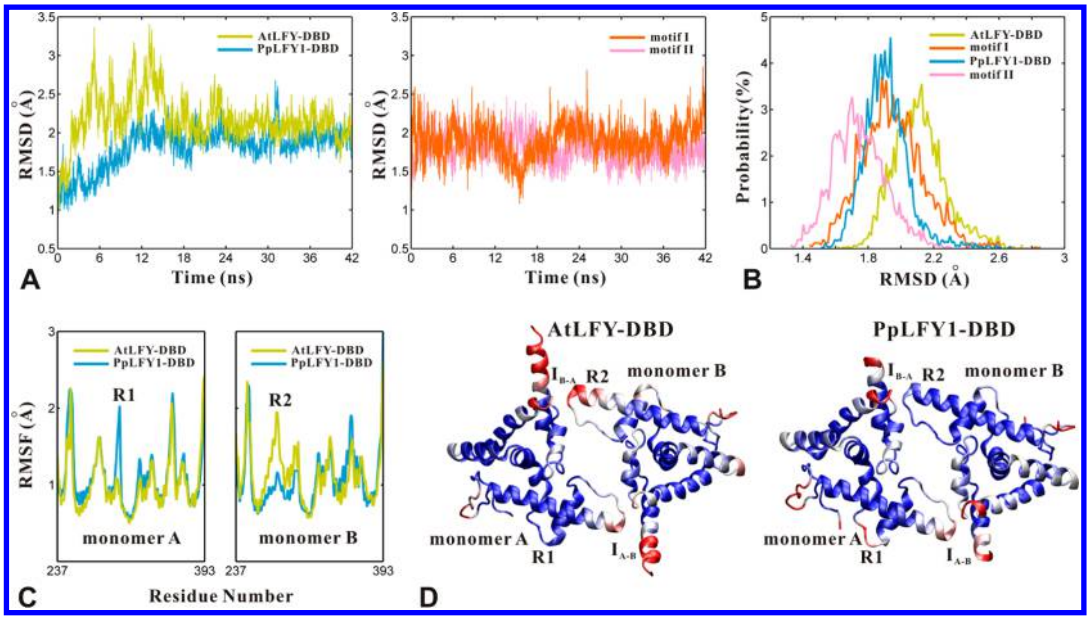


Figure 2. Comparative MD analysis of the type I system (AtLFY-DBD: yellow; motif I: orange) and the type II system (PpLFY1-DBD: blue; motif II: pink). (A) RMSD values of LFY-DBD (left) and DNA (right) backbone atoms versus simulation time. (B) Probability distribution of RMSDs calculated from the equilibrium trajectories. (C) RMSF values of the Cα atoms of monomer A (left) and monomer B (right) calculated from the equilibrium trajectories. (D) Cartoon representations of AtLFY-DBD (left) and PpLFY1-DBD (right). The residues in blue have relatively lower RMSF values (<1.0 Å) while the ones in red have relatively higher RMSF values (>1.8 Å). The other regions are shown in white.

2.2. Molecular Dynamics Simulation. The two simulation systems were prepared using VMD 1.9.[20] Each starting structure was solvated in a periodic box of TIP3P water molecules which were extended about 10 Å from the solute unit to the box wall. To obtain electroneutrality, 16 Na^+ and 48 Na^+ ions were added to the type I and type II systems, respectively. Then, the two simulations were performed using the NAMD 2.9 program[21] in conjunction with the CHARMM27 all-atom additive force field for nucleic acids.[22] All bond lengths involving hydrogen atoms were constrained by the SHAKE algorithm,[23] and the evaluation of electrostatic interactions was achieved using the particle mesh Ewald (PME)[24] method. Meanwhile, the Lennard-Jones potential was truncated at a cutoff of 12 Å. Each simulation consisted of two stages. (i) After the initial 20 000 steps of energy minimization, the systems were slowly heated from 0 to 310 K over 0.5 ns. To keep the stabilization of systems, all backbone atoms of protein and DNA were restrained with a harmonic constant of 0.1 kcal·mol^{-1}·$Å^{-2}$. (ii) The positional constraints were removed, and then the productive MD simulations were run for 42 ns under constant pressure (1 atm) and temperature (310 K) conditions. The pressure and temperature were maintained using the Langevin piston method.[25] For each system, the generated structural data was stored every 2.0 ps. Hence, 21 000 snapshots in each run were provided for further analysis.

2.3. Principal Component Analysis. Principal component analysis (PCA) is a standard method for finding global, correlated motions from MD trajectories through dimensionality reduction. The definition of PCA is based on the construction and diagonalization of the covariance matrix. The element C_{ij} in this covariance matrix is given by[26]

$$C_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle \quad (1)$$

where x_i (x_j) is the coordinate of the i th (jth) atom and $\langle \cdots \rangle$ represents an ensemble average. The eigenvectors of the matrix describe the directions of the concerted motions. Meanwhile, the eigenvalues give the magnitude of the motion for each direction. The first few principal components (PCs) are the eigenvectors with the highest eigenvalues, which usually contain the most important conformational changes of a biomolecular system.[19,27]

Additionally, the cosine content (cc_i) of PCs was introduced by Hess to measure the convergence of conformational sampling.[28] It is calculated according to

$$cc_i = \frac{2}{T}\left(\int_0^T \cos(i\pi t) p_i(t)\, \mathrm{d}t\right)^2 \left(\int_0^T p_i^2(t)\, \mathrm{d}t\right)^{-1} \quad (2)$$

where $p_i(t)$ is the amplitude of the motion along eigenvector i at time t. cc_i varies from 0 to 1. Because insufficient sampling will result in high cc_i values, the MD trajectory is not converged when the cosine content values of the first few PCs are close to 1. In this article, the PCA analysis was performed with Gromacs 4.5 package[29] to obtain the significant motions in the two systems.

2.4. Free Energy Landscape. The free-energy landscape (FEL) can provide a quantitative description of protein dynamics.[26] The FEL is obtained from the joint-probability distribution of reaction coordinates. It can be estimated by

$$\Delta G(X) = -K_B T \ln P(X) \quad (3)$$

where K_B is the Boltzmann constant, T is the temperature of the simulation, and $P(X)$ is the probability distribution along the reaction coordinate X. In the FEL, the local basins populated with high probability usually represent low-energy conformations in stable states. And the barriers populated with low probability indicate the high-energy transient states. In our study, the first and second principal components were chosen as the reaction coordinates. On the basis of the above PCA data, the two-dimensional FEL was calculated to identify the dominant conformations with lower energy.

2.5. Conformational Analysis of Nucleic Acids. Curves is a widely used software tool for analysis of nucleic acid

DOI: 10.1021/ci500705j
J. Chem. Inf. Model. 2015, 55, 784−794

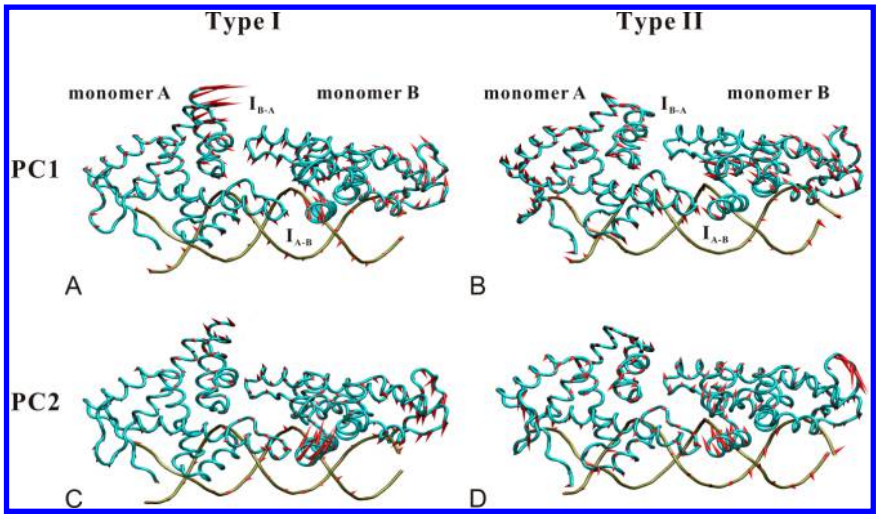


Figure 3. First and second slowest motion modes of the type I system (A and C) and the type II system (B and D). The length of the cone is positively correlated with motive magnitude, and the orientation of the cone describes motive direction.

structures. By performing the Curves program, we can obtain an entire set of DNA conformational parameters.[30] In this article, the groove widths and bend angles along the half sites were analyzed to describe the DNA structural deformations.

3. RESULTS AND DISCUSSION

3.1. Convergence Behavior of the MD Simulations. Two 42 ns MD simulations were carried out for the type I (AtLFY-DBD bound to motif I) and type II (PpLFY1-DBD bound to motif II) systems, respectively. An estimation of the error in the potential energy was evaluated using the block averaging method.[31] The values of potential energy error are plotted versus different block sizes (see the Supporting Information, Figure S1). The two systems both present a long plateau, which indicates that the simulation time is sufficient to sample enough data points.[31] The root-mean-square deviation (RMSD) values were calculated over the protein and DNA backbone atoms, and the results are shown in Figure 2A. The last 25 ns MD trajectories remain stable and then are taken as the equilibrium portions for the two systems. We compared the RMSDs with respect to the average structures for 5 ns blocks of equilibrium portion of each system (see the Supporting Information, Figure S2). They all are within a RMSD of ∼1.5 Å, indicating the convergence of simulations.[32] Figure 2B displays the distributional probability of RMSDs from the equilibrium trajectories. The RMSD values converge to about 2.11, 1.94, 1.91, and 1.74 Å for AtLFY-DBD, motif I, PpLFY1-DBD, and motif II, respectively. This finding indicates that the type II system is comparatively more stable than the type I system. The previous quantitative investigation of interaction showed that the LFY/DNA $k_{0.5}$ constant of the type I system (89 ± 22/97 ± 23 nM) is higher than that of the type II system (58 ± 17/57 ± 20 nM).[11] Therefore, our simulated result is consistent with the experimental binding affinity data. The two ternary complexes have high structural similarities, with different interaction modes at position 312 in helix α3 of monomer A (see Figure 1A). The influence on the system stability made by this interaction will be discussed below (see the section Interactions at the Binding Sites).

The flexibility of each residue is assessed by its root-mean-square fluctuation (RMSF). Figure 2C shows the RMSFs for the common 314 Cα atoms from dimeric AtLFY-DBD and PpLFY1-DBD calculated from the equilibrium trajectories. The correlation coefficient is 0.81 for the type I and type II systems. It indicates that LFY-DBD has highly similar fluctuations when binding to different DNA motifs. However, there are still some differences in two regions between the two systems. One corresponds to residues 300−303 in monomer A (labeled as R1); the other corresponds to residues 278−282 in monomer B (labeled as R2) (see Figure 2C). In order to intuitively observe the fluctuations, AtLFY-DBD and PpLFY1-DBD are colored according to the RMSFs, respectively (see Figure 2D). The red region denotes higher flexibility. By comparison, R1 of PpLFY1-DBD is more flexible than that of AtLFY-DBD. R1 is the linker of helices α2 and α3 in monomer A. Previous study indicated that helix α3 is important for the recognition by the nucleotide bases, and helix α2 has a supporting function.[10] Then, the high flexibility of R1 in PpLFY1-DBD will facilitate the favorable orientation of helix α3 to have a different conformation from that of AtLFY-DBD. It may further contribute to an additional base-specific contact between position 312 (in helix α3 of monomer A) and nucleotide bases in the type II system relative to the type I system.[11] Additionally, R2 of AtLFY-DBD has more fluctuations than that of PpLFY1-DBD. R2 locates at the dimerization interface I_{B-A}. The dimerization interface of LFY is important for the cooperative DNA binding of LFY.[10,11] We suspect that I_{B-A} of AtLFY-DBD may possess some special functional motions, which will be related to this binding cooperativity.

3.2. Motion Modes and Conformational Change. In order to inspect the functional motions of two systems, the PCA analysis was performed for Cα atoms of protein and P atoms of DNA based on the equilibrium trajectories. To judge the convergence of PCA, we further checked the cosine contents of PCs.[28] For the two systems, the cosine content values of the first eight principal components (PC1−PC8) are relatively small (see the Supporting Information, Table S1). The previous studies speculated that the threshold value of cosine content might be 0.2 for small peptides and 0.5 for proteins to discriminate the times of insufficient and sufficient sampling.[33,34] This suggests that the convergences of PCA are achieved in our study. Figure 3 compares the first and second slowest motion modes of the type I and type II systems. The first slowest slow modes of the two systems exhibit some open-

DOI: 10.1021/ci500705j
J. Chem. Inf. Model. 2015, 55, 784−794

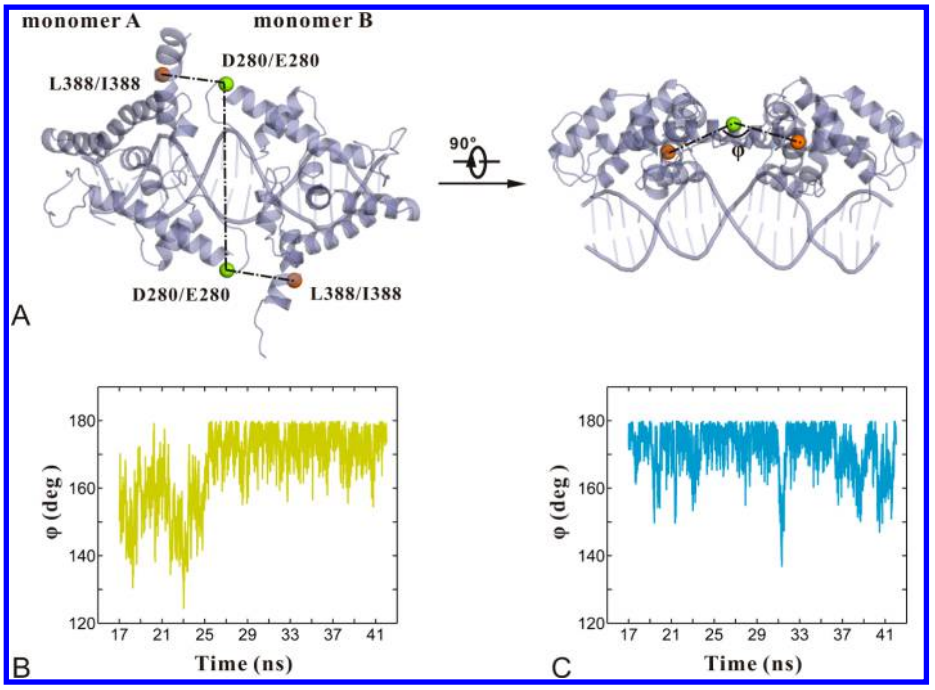


Figure 4. Representation of the dihedral angle φ (A) and changes of φ in the type I (B) and type II (C) systems. The dihedral angle φ is defined by the four Cα atoms of position 388 in monomer A (orange), position 280 in monomer B (green), position 280 in monomer A (green), and position 388 in monomer B (orange). The markedly reduced φ indicates an opening of the dimerization interface.

close movements between monomers (see Figure 3A and B). Notably, $I_{B\text{-}A}$ of the type I system has remarkable open-close motions. This motion trend is weak at the dimerization interface in the type II system. The open-close motions will lead to an opening of the dimerization interface and then are possibly associated with the monomer−monomer contacts. Meanwhile, the two systems exhibit some irregular motions in the second slowest motion mode (see Figure 3C and D). These irregular motions may help to adjust the conformations of helices for more contacts with DNA.

To assess the open-close movements in the first slowest motion mode, we used the dihedral angle φ to measure this conformational change. The dihedral angle φ is defined by four atoms at the dimerization interface, which are the Cα atoms of position 388 in monomer A, position 280 in monomer B, position 280 in monomer A, and position 388 in monomer B (see Figure 4A). Notably, the remarkable decrease of the angle φ represents an opening of the dimerization interface. In the type I system, the angle fluctuates roughly between 140° and 170° before 25 ns (see Figure 4B). Especially, the angle drops down to ∼120° at 23 ns. Meanwhile, the RMSD values increase to 2.6 Å between 22 and 24 ns (see Figure 2A). It indicates that AtLFY has a remarkable opening of the dimerization interface, and this opening causes the reduction of the stability. After 25 ns, the angle rises back and keeps stable around 170° until the end of the simulation. In the type II system, the angle has a relatively narrow fluctuation range and high values (see Figure 4C). The exception is an unexcepted decrease around 31 ns, with the angle value of ∼140°. It indicates that PpLFY1 also has an opening of the dimerization interface at 31 ns. Similarly, the RMSD markedly increases to 2.6 Å during 31−32 ns (see Figure 2A). Combining the two systems, it is found that the system stability will be decreased when the angle φ is less than 140°.

We further investigated the distribution of conformations along the PCs from equilibrium trajectories. Figure 5A and 5B display the free energy contour maps of the type I and type II systems at 310 K, respectively. The deeper color indicates lower energy. The changes of PCs versus simulation time are shown in Figure 5C and 5D for the type I and type II systems, respectively. The fluctuations of PC1 are similar to those of the dihedral angle φ (see Figure 4 and 5). In the type I system, the local minima approximately correspond to the conformations from 25−42 ns (see Figure 5A and C). Before 25 ns, the conformations are located at the negative direction of PC1. At 23 ns, the PC1 values even decrease to −3. After that, the PC1 increases from −3 to +1 and keeps positive values. Thus, the dimerization interface undergoes an open-close conformational change and the closed conformations are more stable. Relative to the type I system, the FEL of the type II system has a larger local basin (see Figure 5B), which includes most of the conformations from equilibrium trajectory (see Figure 5D). The PC1 and PC2 values show relatively low fluctuations. It indicates that the movement range of the type II system is small. Nevertheless, there is a sudden decrease of PC1 around 31 ns (see Figure 5D) and it leads to a high-energy conformational state (see Figure 5B). The above analysis is consistent with the calculation of the dihedral angle φ. Collectively, when the dihedral angle φ is higher than 160°, the ternary structure stays in low-energy conformations. Once the dihedral angle φ decreases to 140°, the dimerization interface is opened excessively and the conformations become less stable. It is possible that the opening of the dimerization interface to some extent will break the monomer−monomer contacts, which was confirmed to be crucial for binding affinity.[10,14] Therefore, the open-close movements at the dimerization interface make an important influence on the structural stability of the ternary complex.

DOI: 10.1021/ci500705j
J. Chem. Inf. Model. 2015, 55, 784−794

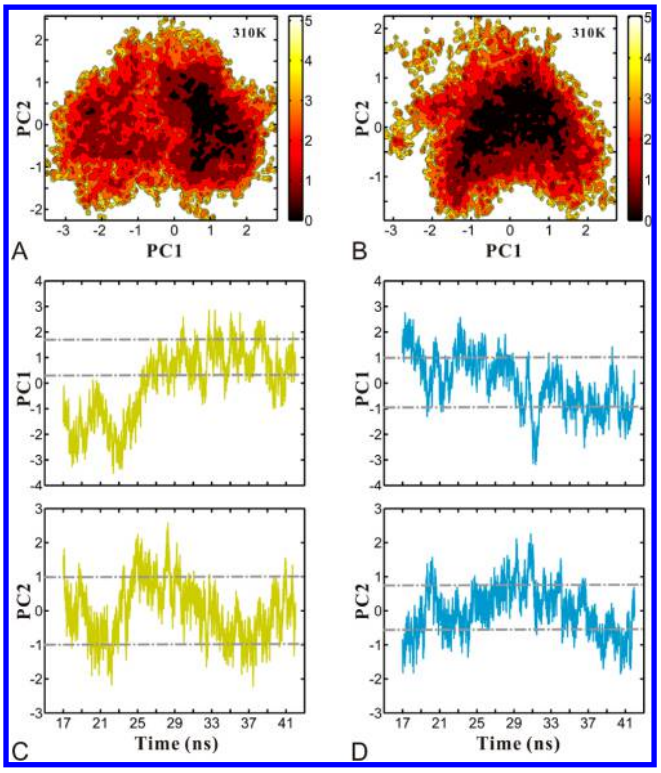


Figure 5. Free energy contour map versus motion modes and motion modes versus simulation time in the type I system (A and C) and the type II system (B and D). In the free energy contour maps (A and B), deeper color indicates lower energy. Corresponding to free-energy minima, the intervals of PC1 and PC2 are marked by dotted lines (C and D). For the free-energy minima of the type I system, the values of PC1 vary from 0.2 to 1.8 and those of PC2 from −1.0 to 1.0 (C). For the free-energy minima of the type II system, the values of PC1 vary from −1.0 to 1.0 and those of PC2 from −0.6 to 0.8 (D).

3.3. Interactions at the Binding Sites. The crystallographic study[10,11] indicated that the protein−DNA and protein−protein hydrogen bonds mediated the important interactions in the LFY-DNA binding. Furthermore, the protein−DNA hydrogen bonds can be grouped into two types: the specific interaction between amino acid and nucleotide base; and the nonspecific interaction between amino acid and DNA backbone.[19] We examined these hydrogen bonds from the equilibrium trajectories of the two systems. These hydrogen bonds were calculated by VMD 1.9[20] with a distance cutoff value of 3.5 Å and an angle cutoff value of 45°. The results are listed in Tables 1 and 2 with occupancy over 30%. The two systems have similar protein−DNA interactions (see Table 1). In the type I system, Lys307 in monomer A and monomer B forms stable base-specific hydrogen bonds with guanines −3 and +3, respectively. In the type II system, Lys307 in monomer B similarly has a stable base-specific contact with guanine +3. However, this stable base-specific interaction is not found between Lys307 in monomer A and guanine −3. Then, we further examined water-mediated hydrogen bonds between them. It is interesting that NZ of Lys307 in monomer A makes an indirect base-specific contact with N7 of guanines −3, which is mediated by water molecules with occupancy 52.23%. Therefore, Lys307 can specify the invariant guanines ±3[10] by direct or water-mediated contacts. In the type I system, NH2 of N-terminal residue Arg237 in monomer A donates a stable hydrogen bond to the atom O2 of thymine −8. In the type II system, NH1 and NH2 of Arg237 in monomer A contact the atoms O4′ and N3 of guanine −9, respectively. Meanwhile, NH1 and NH2 of Arg237 in monomer B form interactions with the atoms O4′ and O2 of cytosine +9, respectively. Then, Arg237 recognizes the terminal nucleotide bases of half sites in the two system,[10] but this base-specific interaction is comparatively weak in the type I system. Additionally, the helix α2 (Asn291 and Arg295), the helix α3 (Asn306, Lys309, Arg311 and Tyr313), the linker of helices α6 and α7 (Tyr377), and the helix α7 (Thr380) make stable contacts with the DNA phosphate backbone in the two systems (see Table 1). The nonspecific binding also contributes to the structural stability. The above observed interactions agree with the experimental results.[10]

In spite of the high similarity, there is still a substantial difference in protein−DNA interactions between the two systems. The previous study[11] revealed that His312 in monomer A of the type I system does not form any interaction with motif I, whereas Asp312 in monomer A of the type II system has an additional base-specific contact with cytosine −6 of motif II. Figure 6 describes the interaction difference at this binding site between the two systems. In the type I system, the side chain of His312 always does not contact DNA throughout the simulation process. Only Arg311 forms phosphate binding with guanine −2 in the β strand of DNA (see Figure 6A). In the type II system, Asp312 loses the direct contact with cytosine −6 in comparison with the crystallographic data. By examining the water-mediated hydrogen bonds, we found that Asp312 OD2 and OD1 make indirect contacts with the atom N4 of cytosine −6. Their occupancies are 12.00% and 11.18%, respectively. Interestingly, Arg311 adjacent to Asp312 directly interacts with adenine −4 in the β strand, and the occupancy is 50.58% (see Table 1). Instead of direct contact, Asp312 of the type II system is inclined to form water-mediated interactions with cytosine −6. It is presumably because the negatively charged Asp312 interacts unfavorably with the negatively charged DNA backbone. The charge repulsion effect leads to the weak contact between Asp312 and cytosine −6. Nevertheless, Asp312 of the type II system still adopts a different orientation of helix α3 from that of the type I system (see Figure 6B). It allows the side chain of Arg311 to be inserted into the DNA major groove and recognize the adenine −4 of the other half site. Our study reveals that position 312 in LFY determines the difference of base-specific interaction between the two systems, in direct or indirect ways. The additional base-specific contact in the type II system makes an important contribution to the higher stability relative to the type I system.

The above FEL analysis has suggested the importance of contacts between LFY monomers. Then, it is necessary to analyze the influence of protein−protein interaction on the binding in detail. We compared the base-specific and monomer−monomer hydrogen bonds calculated from the high-energy conformations (Cluster I) and the low-energy conformations (Cluster II), respectively (see Table 2). In the type I system, the interface $I_{A\text{-}B}$ maintains two stable monomer−monomer hydrogen bonds in either Cluster I (17−25 ns) or Cluster II (25−42 ns). In contrast, $I_{B\text{-}A}$ only has weak monomer−monomer contacts. Especially, $I_{B\text{-}A}$ completely loses the monomer−monomer hydrogen bonds between 22 and 24 ns. It leads to the decreased stability in the type I system (see Figure 2A). After 25 ns, two stable base-specific hydrogen bonds are formed by Arg237 in monomer B with adenine +7 and +8 of half sites. It explains that the

DOI: 10.1021/ci500705j
J. Chem. Inf. Model. 2015, 55, 784−794

Table 1. Protein−DNA Hydrogen Bonds in the Two Systems with Occupancy over 30%

type I system			type II system		
protein	DNA (id*)[a]	occupancy	protein	DNA (id*)[a]	occupancy
Tyr313-OH^{A}	C(−7)-O2P$^{\alpha}$	98.32%	Tyr313-OH^{A}	C(−7)-O2P$^{\alpha}$	97.00%
Lys309-NZ^{B}	G(+6)-O2P$^{\beta}$	97.32%	Lys309-NZ^{A}	C(−6)-O2P$^{\alpha}$	95.80%
Tyr377-OH^{B}	G(+2)-O1P$^{\alpha}$	96.84%	Arg295-NH1^{A}	C(−1)-O2P$^{\beta}$	92.12%
Lys309-NZ^{A}	G(−6)-O2P$^{\alpha}$	96.80%	Arg311-NH2^{B}	T(+4)-O2P$^{\alpha}$	91.04%
TYR377-OH^{A}	G(−2)-O1P$^{\beta}$	92.20%	Arg295-NH2^{A}	C(−1)-O1P$^{\beta}$	90.80%
Tyr313-OH^{B}	A(+7)-O1P$^{\beta}$	85.65%	Asn306-ND2^{B}	C(+6)-O2P$^{\beta}$	87.13%
Asn306-ND2^{A}	G(−6)-O2P$^{\alpha}$	76.53%	Lys307-NZ^{A}	G(−2)-O2P$^{\beta}$	86.57%
Thr380-OG1^{B}	T(+4)-O1P$^{\alpha}$	73.45%	Lys309-NZ^{B}	C(+6)-O2P$^{\beta}$	86.17%
Arg311-NH2^{A}	G(−2)-O2P$^{\beta}$	72.77%	Arg295-NH1^{B}	T(+1)-O2P$^{\alpha}$	84.21%
Asn306-ND2^{B}	G(+6)-O2P$^{\beta}$	69.17%	Tyr313-OH^{B}	G(+7)-O2P$^{\beta}$	83.57%
Arg311-NH2^{B}	G(+2)-O2P$^{\alpha}$	63.21%	Arg295-NH2^{B}	T(+1)-O1P$^{\alpha}$	80.97%
Arg295-NH1^{A}	T(−1)-O2P$^{\beta}$	62.53%	Thr380-OG1^{A}	A(−4)-O2P$^{\beta}$	80.09%
Lys307-NZ^{A}	**G(−3)-N7$^{\beta}$**	**62.50%**	Thr380-OG1^{B}	T(+4)-O1P$^{\alpha}$	77.33%
Thr380-OG1^{A}	T(−4)-O1P$^{\beta}$	60.42%	Asn306-ND2^{A}	C(−6)-O2P$^{\alpha}$	74.85%
Lys307-NZ^{B}	**G(+3)-N7$^{\alpha}$**	**47.46%**	**Lys307-NZ^{B}**	**G(+3)-N7$^{\alpha}$**	**73.37%**
Arg295-NH2^{B}	T(+1)-O2P$^{\alpha}$	44.30%	Asn291-ND2^{B}	T(+1)-O2P$^{\alpha}$	70.17%
Arg295-NH2^{A}	T(−1)-O2P$^{\beta}$	43.66%	Asn291-ND2^{A}	C(−1)-O2P$^{\beta}$	65.45%
Lys307-NZ^{B}	**G(+3)-O6$^{\alpha}$**	**36.47%**	Tyr377-OH^{B}	G(+2)-O1P$^{\alpha}$	50.74%
Arg237-NH2^{A}	**T(−8)-O2$^{\beta}$**	**32.83%**	**Arg311-NH2^{A}**	**A(−4)-N7$^{\beta}$**	**50.58%**
Asn291-ND2^{B}	T(+1)-O2P$^{\alpha}$	30.19%	**Arg237-NH2^{A}**	**G(−9)-N3$^{\beta}$**	**46.86%**
			Arg237-NH1^{B}	**C(+9)-O4′$^{\alpha}$**	**46.50%**
			Tyr377-OH^{A}	G(−3)-O1P$^{\beta}$	39.62%
			Arg237-NH1^{A}	**G(−9)-O4′$^{\beta}$**	**37.43%**
			Arg237-NH2^{B}	**C(+9)-O2$^{\alpha}$**	**30.99%**
			Arg311-NE^{B}	T(+4)-O2P$^{\alpha}$	30.35%

[a]id* is the index of a nucleotide base in the half-site sequence. The specific interactions are in bold, while nonbold denotes nonspecific interactions. ^{A}Residue belonging to monomer A. ^{B}Residue belonging to monomer B. $^{\alpha}$Nucleotide base belonging to α strand of DNA. $^{\beta}$Nucleotide base belonging to β strand of DNA.

conformations of Cluster II are more stable than those of Cluster I in the FEL analysis, since the specific recognition of DNA minor groove by Arg237 was confirmed to be important for the DNA-binding affinity.[10] Then, the open-close movements at the interface $I_{B\text{-}A}$ appears to improve the specific binding efficiency between protein and DNA. In the type II system, Cluster I (31−32 ns) and Cluster II (17−31 ns and 32−42 ns) both have stable protein−DNA binding. Cluster I even has relatively more base-specific hydrogen bonds although the Cluster I is less stable than Cluster II (see Table 2). During 31−32 ns, there is a sudden opening of the dimerization interface (see Figure 4C). This opening leads to the loss of stable monomer−monomer contacts at the interface $I_{A\text{-}B}$ (see Table 2) and, thus, causes the reduction of the stability in the type II system (see Figure 2A). Altogether, $I_{A\text{-}B}$ and $I_{B\text{-}A}$ are corporately responsible for the dimerization binding mode. In this dimerization binding mode, $I_{A\text{-}B}$ tends to form stable protein−protein interactions to connect two monomers, which directly affects the stability. Besides some monomer−monomer contacts, $I_{B\text{-}A}$ also has functional open-close movements which make an indirect contribution to the specific DNA binding of LFY. Therefore, the protein−protein interaction can regulate the cooperative binding of this ternary structure in different ways.

Since the dimerization interface is associated with the specific DNA-binding efficiency of LFY, it can be further asked whether an altered dimerization interface leads to the change of LFY DNA-binding specificity. Interestingly, two separate mutation studies using the same amino acid substitutions in the dimerization interface obtained opposite results.[11,14] In our study, the amino acid sequences of the dimerization interfaces are not identical in the AtLFY and PpLFY1 proteins. Nevertheless, except for different interaction modes at position 312 in helix α3, the two LFY proteins have similar DNA-binding specificity (see Table 1). It is worth noting that the two systems both present a protein−protein interaction surface (see Figure 2D). The surface helps the two LFY monomers to be fixed in an approximate plane, which are evaluated by the dihedral angle φ (see Figure 4). It results in a relatively constant position of the specific DNA-binding site to recognize the nucleotide base. Thus, some variations in sequence of amino acids in the dimerization interface will not considerably affect LFY DNA-binding specificity when the protein−protein interaction surface exists. Conversely, in some cases the mutations may lead to the complete loss of protein−protein interaction, thereby causing the absence of this interaction surface. The structural model showed that the dimeric LFY-DBD underwent a dramatic conformational change without the positional constraints from this interaction surface.[11] The shifts in position and orientation of LFY monomers probably change its DNA-binding specificity. It was confirmed in the previous mutation study through a SELEX (systematic evolution of ligands by exponential enrichment) experiment.[11] Therefore, we suggest that the protein−protein interaction surface is a key factor in determining whether the sequence alteration of the dimerization interface affects LFY DNA-binding specificity.

3.4. Structural Deformability of DNA. Proteins often recognize nucleotide bases by interacting with DNA which is deformed.[18,19,35] To observe the DNA distortion, the structural parameters of motifs I and II were calculated by Curves

DOI: 10.1021/ci500705j
J. Chem. Inf. Model. 2015, 55, 784−794

Table 2. Interface Interaction from the Two Clusters with Occupancy over 30%

type I system_Cluster I			type II system_Cluster I		
donor	acceptor	occupancy	donor	acceptor	occupancy
Lys307-NZ^{A}	G(−3)-N7$^{\beta}$	66.29%	Arg237-NH2^{B}	C(+9)-O2$^{\alpha}$	89.11%
Lys279-NZ^{A}	Asp362-OD1^{B}	48.06%	Lys307-NZ^{B}	G(+3)-N7$^{\alpha}$	88.12%
Lys279-NZ^{A}	Asp362-OD2^{B}	47.44%	Arg311-NH2^{A}	A(−4)-N7$^{\beta}$	58.42%
Lys307-NZ^{B}	G(+3)-N7$^{\alpha}$	43.95%	Arg237-NH2^{A}	G(−9)-N3$^{\beta}$	46.53%
Lys307-NZ^{B}	G(+3)-O6$^{\alpha}$	38.45%	Lys279-NZ^{B}	Asp362-OD2^{A}	45.54%
Arg237-NH1^{A}	T(−8)-O2$^{\beta}$	36.83%	Arg237-NH1^{B}	C(+8)-O2$^{\alpha}$	43.56%
Lys307-NZ^{B}	G(+2)-N7$^{\alpha}$	30.96%	Arg237-NH1^{A}	T(−8)-O2$^{\beta}$	39.30%
			Arg237-NH1^{A}	G(−9)-O4′$^{\beta}$	41.59%
			Lys279-NZ^{B}	Asp362-OD1^{A}	31.68%
type I system_Cluster II			type II system_Cluster II		
donor	acceptor	occupancy	donor	acceptor	occupancy
Lys279-NZ^{A}	Asp362-OD1^{B}	62.26%	Lys307-NZ^{B}	G(+3)-N7$^{\alpha}$	72.81%
Lys307-NZ^{A}	G(−3)-N7$^{\beta}$	60.73%	Arg311-NH2^{A}	A(−4)-N7$^{\beta}$	50.31%
Lys307-NZ^{B}	G(+3)-N7$^{\alpha}$	49.09%	Arg237-NH1^{B}	C(+9)-O4′$^{\alpha}$	47.43%
Lys279-NZ^{A}	Asp362-OD2^{B}	41.21%	Lys279-NZ^{B}	Asp362-OD1^{A}	46.86%
Arg237-NE^{B}	A(+7)-O4′$^{\beta}$	40.98%	Arg237-NH2^{A}	G(−9)-N3$^{\beta}$	46.85%
Arg237-NH2^{A}	T(−8)-O2$^{\beta}$	35.80%	Lys279-NZ^{A}	Asp362-OD2^{B}	39.74%
Lys307-NZ^{B}	G(+3)-O6$^{\alpha}$	35.63%	Lys279-NZ^{B}	Asp362-OD2^{A}	39.70%
Arg237-NH2^{B}	A(+8)-N3$^{\beta}$	30.34%	Arg237-NH1^{A}	G(−9)-O4′$^{\beta}$	37.31%

^{A}Residue belonging to monomer A. ^{B}Residue belonging to monomer B. $^{\alpha}$Nucleotide base belonging to α strand of DNA. $^{\beta}$Nucleotide base belonging to β strand of DNA.

program[30] along their half-site sequences. The variations of DNA groove dimensions play a key role in protein−DNA readout processes.[36] Figure 7 compares the average groove widths from the equilibrium trajectories in the two systems. The recognition sites of two motifs show remarkable variations in groove widths relative to the canonical B-DNA (major groove width 11.7 Å and minor groove width 5.7 Å). In the two motifs, the guanines at positions ±3 are recognized by Lys307 in the major groove. Accordingly, the major groove widths near by guanines ±3 are narrowed to 9−10 Å (see Figure 7A). In the two motifs, the two ends of half-site sequence are recognized by Arg237 in the minor groove. By observation, the minor groove of motif I is narrowed strikingly with a width of 1−2 Å near by positions thymine −8 and adenine +8 (see Figure 7B). Meanwhile, the minor groove of motif II is opened to have a width of more than 7 Å at positions guanine −9 and cytosine +9 (see Figure 7B). It is consistent with that A-tract and G-tract tend to compress and widen minor grooves, respectively.[37] Notably, the crystal structure data set of protein-bound DNA showed that the minor groove width values of protein-bound DNA are usually beyond 3 Å.[36] Therefore, the minor groove with the average width lower than 2 Å is probably a bit small for accommodating protein elements. It explains the relatively lower hydrogen bond occupancy at positions ±8 of the type I system than positions ±9 of the type II system (see Table 1). Additionally, the hydrogen bond calculation shows that in motif II positions −4 and −6 are recognized by Arg311 and Asp312 in the major groove, respectively (see Figure 6B). It is a substantial difference between motifs II and I. Consistent with this analysis, the major groove width of motif II is markedly deviated from normal B-form geometry at positions −4 and −6 relative to that of motif I (see Figure 7A). Altogether, except for the central 3-bp spacer, the half sites of motif II exhibit more significant groove deformation than those of motif I. Combined with calculation of hydrogen bonds, the groove width comparison indicates that the type II system shows a comparatively higher efficiency of base-specific binding between protein and DNA.

In addition to the specific protein−DNA contacts in a process known as "direct readout", the specific recognition can also be modulated by DNA structural properties without any contact between a protein and DNA bases. This phenomenon is referred to as "indirect readout".[38] The central three nucleotide bases do not form any direct specific interaction with the LFY protein in the two systems (see Table 1). Interestingly, the central 3-bp-spaced regions of the two motifs show distinctly different groove deformation. At the center of motif I, the major and minor groove widths are ∼9.3 and ∼6.5 Å, respectively (see Figure 7A and B). At the center of motif II, the major and minor groove widths are ∼11.1 and ∼5.6 Å, respectively (Figure 7A and B). Obviously, the central 3-bp spacer of motif I shows a stronger distortion from normal B-form geometry, which contributes to an additional specificity through indirect readout.[10] The previous study revealed the interdependence between the central 3-bp spacer and the LFY dimerization interface when LFY-DBD bound to motifs I and II.[11] Then, we further calculated the changes of the minor groove widths at position 0 and the monomer−monomer distances at the dimerization interface from the equilibrium trajectories. The monomer−monomer distance is measured between the Cα atom of position 280 in one LFY monomer and that of position 387 in another LFY monomer (see Figure 8A). In motif I, position 0 has a minor groove width of about 6−8 Å. The monomer−monomer distances of the type I system fluctuate between 9 and 12 Å (see Figure 8B). In motif II, the minor groove width at position 0 is about 5−6 Å. The monomer−monomer distances of the type II system roughly keep stable between 8 and 9 Å (see Figure 8C). Furthermore, widening of the minor groove is often accompanied by the compression of the major groove.[19,30] Thus, the above analysis indicates that the groove deformation of the central 3-bp spacer is linked to the opening of the dimerization interface. In view of the relationship between the structural deformation of the central 3-bp spacer and the opening of the dimerization interface, the protein−protein interaction might be required to play different roles under different conditions.

DNA bending was considered as another important structural feature in processes involving specific protein−DNA binding.[19,39] Figure 9 compares the mean bend angles of DNA from the equilibrium trajectories with the crystal values along the half sites in the two systems. In the type I system, almost all of half sites exhibit a reduced bending relative to the crystal values, except for the central 3-bp spacer with a remarkable increase of bending degree (see Figure 9A). It implies that, when the protein−DNA binding is weakened, the

DOI: 10.1021/ci500705j
J. Chem. Inf. Model. 2015, 55, 784−794

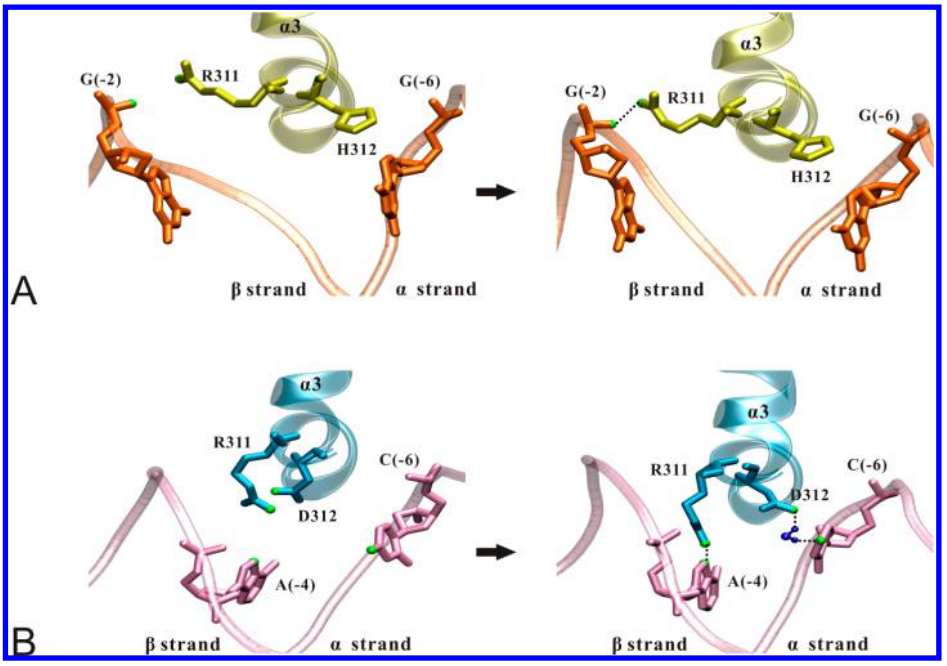


Figure 6. Protein−DNA interactions at positions 311−312 from representative structures in the type I (A) and type II (B) systems. The LFY and DNA are depicted with cartoon and tube models, respectively. Atoms involved in protein−DNA interactions are shown in CPK. Before 18 ns (left), both His312 (A) and Arg312 (B) have no contact with DNA. After 18 ns (right), His312 still keeps away from DNA. In this case, Arg311 only forms a nonspecific interaction with guanine −2 in the β strand of DNA (A). In contrast, Asp312 forms water-mediated hydrogen bonds with cytosine −6 in the α strand of DNA. It helps the helix α3 to enter the DNA major groove and then Arg311 further recognizes the adenine −4 in the β strand of DNA (B).

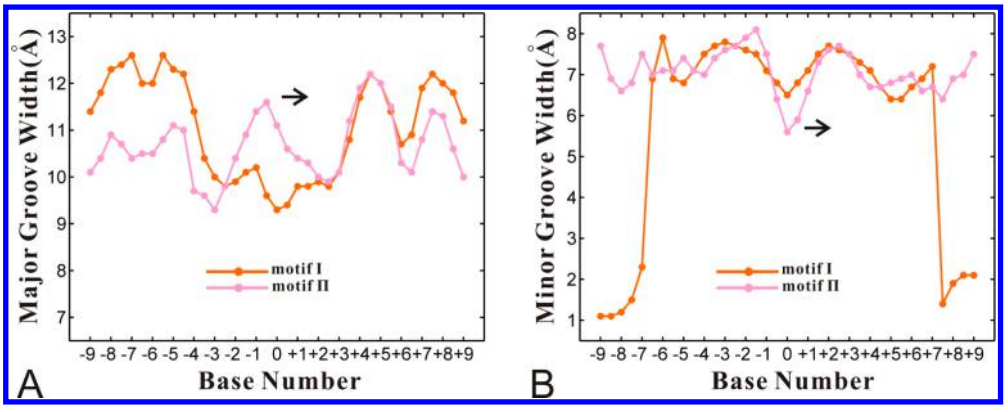


Figure 7. Average values of groove widths calculated from the equilibrium trajectories along the half sites in motif I (orange) and motif II (pink). (A) Major groove widths. (B) Minor groove widths. The arrows show the corresponding major and minor groove widths in a canonical B-DNA.

structural deformation of central 3-bp spacer will help to increase the DNA-binding specificity through indirect readout. Meanwhile, this structural deformation is related to the opening of the dimerization interface. It will cause some loss of the protein−protein interaction to achieve the additional specificity. In this case, the protein−protein interaction likely acts as a contributor of the protein−DNA binding to regulate the DNA-binding specificity. In the type II system, the LFY dimer forms highly stable specific interactions with DNA. In this case, 3-bp-spaced region shows a regular B-form conformational geometry (see Figure 7) and has a lower bending degree (see Figure 9B). Then, the protein−protein interaction only plays a role of monomer−monomer connecter when the protein−DNA binding is quite stable. The previous EMSA analysis revealed the importance of protein−protein interaction, which showed that the dimeric LFY-DNA complex has higher affinity relative to the monomeric one.[10] Our study further reveals the dual roles of the protein−protein interaction in properly keeping a cooperative DNA-binding mode during dynamic processes. By switching between different roles, the protein−protein interaction will help to promote the robustness and adaptivity of the LFY-DNA ternary structure.

4. CONCLUSIONS

In this study, comparative MD simulations were performed to investigate the binding mechanism between LFY and DNA. The stability comparison of the type I system (AtLFY bound to motif I) and type II system (PpLFY1 bound to motif II) is consistent with the experimental data. The conformational changes of the two systems were investigated by PCA and FEL methods. The two monomers show some open-close movements. The movements lead to a change of the dihedral angle at the dimerization interface. When the angle values are lower than 140°, the dimerization interface is opened excessively.

DOI: 10.1021/ci500705j
J. Chem. Inf. Model. 2015, 55, 784−794

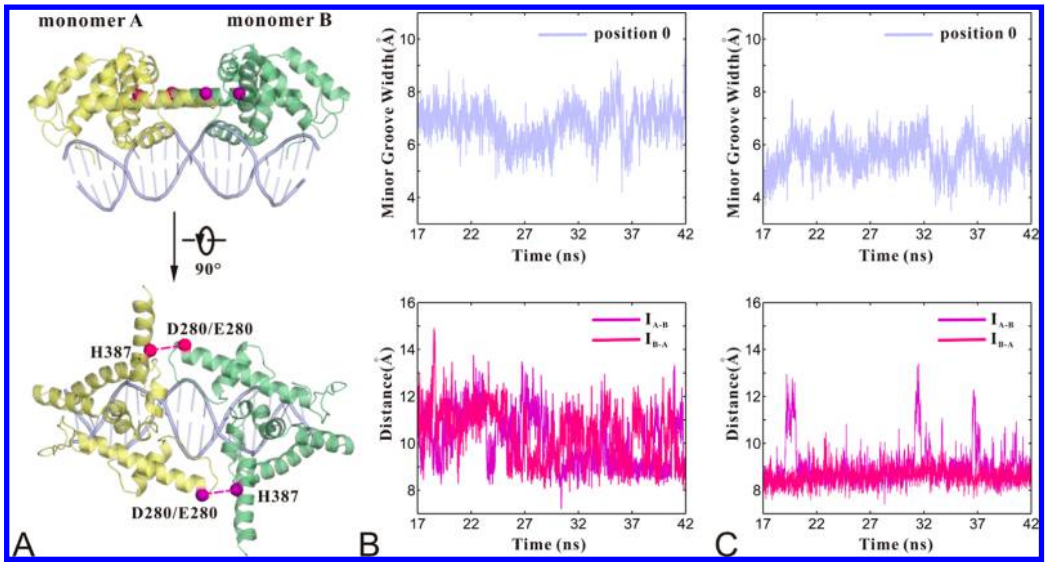


Figure 8. Time-dependent fluctuations of DNA minor groove widths at position 0 (light blue) and distances at the monomer−monomer interfaces ($I_{A\text{-}B}$ purple; $I_{B\text{-}A}$ red) calculated from the equilibrium trajectories. (A) Monomer−monomer distance defined by the Cα atoms of position 280 in one monomer and position 387 in another monomer. (B) Changes of the type I system. (C) Changes of the type II system.

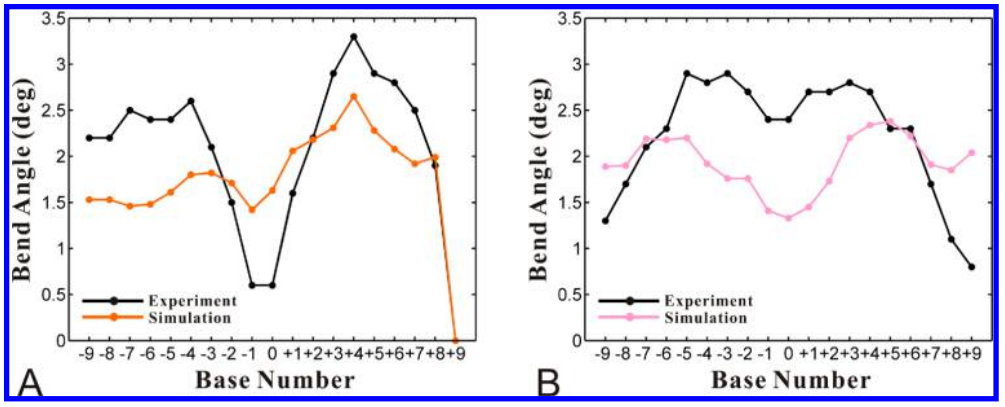


Figure 9. Comparison of the average values of the axis bend (motif I orange; motif II pink) calculated from the equilibrium trajectories along the half sites with the corresponding crystal values (black). (A) Motif I. (B) Motif II.

This opening will break the crucial monomer−monomer contacts, thereby leading to the decrease of structural stability. Then, we calculated the interactions at the protein−DNA and protein−protein interfaces. Consistent with the comparison of stability, the type II system forms more protein−DNA contacts than the type I system. In particular, position 312 in LFY determines the difference of base-specific interactions between the two systems. Combined the FEL analysis, the hydrogen bond calculation confirmed that the protein−protein interaction is important for the stability. Notably, the protein−protein interaction also appears to have an influence on the specific DNA binding of LFY. Therefore, the DNA structural parameters were analyzed to explore the roles of protein−protein interaction. Relative to motif II, the central 3-bp spacer of motif I shows a stronger groove deformation that is linked to the opening of the dimerization interface. The bend angle change further revealed that the protein−protein interaction plays dual roles as a connector between LFY monomers and a regulator of DNA-binding specificity. In conclusion, we suggest that protein−protein interaction helps this ternary structure to achieve a robust and adaptive binding by serving its dual function. This study provides the dynamics and atomic level information for the dimerization binding mode between LFY-DBD and DNA, which will facilitate a deeper understanding of the molecular mechanism underlying functional change of LFY.

■ ASSOCIATED CONTENT

Supporting Information

Error estimates for the potential energy, the RMSDs with respect to the average structures, and the cosine contents of PCs. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: schang@jsut.edu.cn.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (31200990, 11247018, 31372116) and the Foundation for Outstanding Young Teachers in Higher Education of Guangdong (Yq2013027), the Key Project of Education Department of Sichuan Province (12ZA066).

DOI: 10.1021/ci500705j
J. Chem. Inf. Model. 2015, 55, 784−794

REFERENCES

(1) Maizel, A.; Busch, M. A.; Tanahashi, T.; Perkovic, J.; Kato, M.; Hasebe, M.; Weigel, D. The Floral Regulator LEAFY Evolves by Substitutions in the DNA Binding Domain. *Science* **2005**, *308*, 260−263.

(2) Blazquez, M. A.; Soowal, L. N.; Lee, I.; Weigel, D. LEAFY Expression and Flower Initiation in *Arabidopsis*. *Development* **1997**, *124*, 3835−3844.

(3) Blazquez, M. A.; Ferrandiz, C.; Madueno, F.; Parcy, F. How Floral Meristems are Built. *Plant Mol.Biol.* **2006**, *60*, 855−870.

(4) Tanahashi, T.; Sumikawa, N.; Kato, M.; Hasebe, M. Diversification of Genie Function: Homologs of the Floral Regulator *FLO/LFY* Control the First Zygotic Cell Division in the Moss *Physcomitrella patens*. *Development* **2005**, *132*, 1727−1736.

(5) William, D. A.; Su, Y. H.; Smith, M. R.; Lu, M.; Baldwin, D. A.; Wagner, D. Genomic Identification of Direct Target Genes of LEAFY. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 1775−1780.

(6) Saddic, L. A.; Huvermann, B. R.; Bezhani, S.; Su, Y. H.; Winter, C. M.; Kwon, C. S.; Collum, R. P.; Wagner, D. The LEAFY Target LMI1 Is a Meristem Identity Regulator and Acts Together with LEAFY to Regulate Expression of CAULIFLOWER. *Development* **2006**, *133*, 1673−1682.

(7) Parcy, F.; Nilsson, O.; Busch, M. A.; Weigel, I. L. A. D. A Genetic Framework for Floral Patterning. *Nature* **1998**, *395*, 561−566.

(8) Chae, E.; Tan, Q. K. G.; Hill, T. A.; Irish, V. F. An *Arabidopsis* F-Box Protein Acts as a Transcriptional Co-Factor to Regulate Floral Development. *Development* **2008**, *135*, 1235−1245.

(9) Souer, E.; Rebocho, A. B.; Bliek, M.; Kusters, E.; de Bruin, R. A. M.; Koes, R. Patterning of Inflorescences and Flowers by the F-Box Protein DOUBLE TOP and the LEAFY Homolog ABERRANT LEAF AND FLOWER of Petunia. *Plant Cell* **2008**, *20*, 2033−2048.

(10) Hames, C.; Ptchelkine, D.; Grimm, C.; Thevenon, E.; Moyroud, E.; Gerard, F.; Martiel, J. L.; Benlloch, R.; Parcy, F.; Muller, C. W. Structural Basis for LEAFY Floral Switch Function and Similarity with Helix-Turn-Helix Proteins. *EMBO J.* **2008**, *27*, 2628−2637.

(11) Sayou, C.; Monniaux, M.; Nanao, M. H.; Moyroud, E.; Brockington, S. F.; Thevenon, E.; Chahtane, H.; Warthmann, N.; Melkonian, M.; Zhang, Y.; Wong, G. K. S.; Weigel, D.; Parcy, F.; Dumas, R. A Promiscuous Intermediate Underlies the Evolution of LEAFY DNA Binding Specificity. *Science* **2014**, *343*, 645−648.

(12) Baum, D. A.; Yoon, H. S.; Oldham, R. L. Molecular Evolution of the Transcription Factor LEAFY in Brassicaceae. *Mol. Phylogenet. Evol.* **2005**, *37*, 1−14.

(13) Jansen, R. K.; Cai, Z.; Raubeson, L. A.; Daniell, H.; Depamphilis, C. W.; Leebens-Mack, J.; Muller, K. F.; Guisinger-Bellian, M.; Haberle, R. C.; Hansen, A. K.; Chumley, T. W.; Lee, S. B.; Peery, R.; McNeal, J. R.; Kuehl, J. V.; Boore, J. L. Analysis of 81 Genes from 64 Plastid Genomes Resolves Relationships in Angiosperms and Identifies Genome-Scale Evolutionary Patterns. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 19369−19374.

(14) Chahtane, H.; Vachon, G.; Le Masson, M.; Thevenon, E.; Perigon, S.; Mihajlovic, N.; Kalinina, A.; Michard, R.; Moyroud, E.; Monniaux, M.; Sayou, C.; Grbic, V.; Parcy, F.; Tichtinsky, G. A Variant of LEAFY Reveals Its Capacity to Stimulate Meristem Development by Inducing RAX1. *Plant J.* **2013**, *74*, 678−689.

(15) Mendoza, L.; Thieffry, D.; Alvarez-Buylla, E. R. Genetic Control of Flower Morphogenesis in *Arabidopsis thaliana*: a Logical Analysis. *Bioinformatics* **1999**, *15*, 593−606.

(16) Jaeger, K. E.; Pullen, N.; Lamzin, S.; Morris, R. J.; Wigge, P. A. Interlocking Feedback Loops Govern the Dynamic Behavior of the Floral Transition in *Arabidopsis*. *Plant Cell* **2013**, *25*, 820−833.

(17) Moyroud, E.; Minguet, E. G.; Ott, F.; Yant, L.; Pose, D.; Monniaux, M.; Blanchet, S.; Bastien, O.; Thevenon, E.; Weigel, D.; Schmid, M.; Parcy, F. Prediction of Regulatory Interactions from Genome Sequences Using a Biophysical Model for the *Arabidopsis* LEAFY Transcription Factor. *Plant Cell* **2011**, *23*, 1293−1306.

(18) Frantz, B.; O'Halloran, T. V. DNA Distortion Accompanies Transcriptional Activation by the Metal-Responsive Gene-Regulatory Protein MerR. *Biochemistry* **1990**, *29*, 4747−4751.

(19) Wan, H.; Hu, J. P.; Li, K. S.; Tian, X. H.; Chang, S. Molecular Dynamics Simulations of DNA-Free and DNA-Bound TAL Effectors. *PLoS One* **2013**, *8*, e76045.

(20) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.

(21) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.

(22) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell, A. D. CHARMM General Force Field: a Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2010**, *31*, 671−690.

(23) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1997**, *23*, 327−341.

(24) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: an N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(25) Hatano, T.; Sasa, S. Steady-State Thermodynamics of Langevin Systems. *Phys. Rev. Lett.* **2001**, *86*, 3463−3466.

(26) Maisuradze, G. G.; Liwo, A.; Scheraga, H. A. Relation between Free Energy Landscapes of Proteins and Dynamics. *J. Chem. Theory Comput.* **2010**, *6*, 583−595.

(27) Wan, H.; Hu, J. P.; Tian, X. H.; Chang, S. Molecular Dynamics Simulations of Wild Type and Mutants of Human Complement Receptor 2 Complexed with C3d. *Phys. Chem. Chem. Phys.* **2013**, *15*, 1241−1251.

(28) Hess, B. Convergence of Sampling in Protein Simulations. *Phys. Rev. E* **2002**, *65*, 031910.

(29) Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.

(30) Lavery, R.; Moakher, M.; Maddocks, J. H.; Petkeviciute, D.; Zakrzewska, K. Conformational Analysis of Nucleic Acids Revisited: Curves. *Nucleic Acids Res.* **2009**, *37*, 5917−5929.

(31) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: from Algorithms to Applications*; Academic Press: San Diego, CA, 2001; Vol. 2, pp 98−100.

(32) Pieniazek, S. N.; Hingorani, M. M.; Beveridge, D. L. Dynamical Allosterism in the Mechanism of Action of DNA Mismatch Repair Protein MutS. *Biophys. J.* **2011**, *101*, 1730−1739.

(33) Maisuradze, G. G.; Leitner, D. M. Free Energy Landscape of a Biomolecule in Dihedral Principal Component Space: Sampling Convergence and Correspondence between Structures and Minima. *Proteins* **2007**, *67*, 569−578.

(34) Maisuradze, G. G.; Liwo, A.; Scheraga, H. A. Principal Component Analysis for Protein Folding Dynamics. *J. Mol. Biol.* **2009**, *385*, 312−329.

(35) Hu, J. P.; Ke, G. T.; Chang, S.; Chen, W. Z.; Wang, C. X. Conformational Change of HIV-1 Viral DNA after Binding with Integrase. *Acta Phys.-Chim. Sin.* **2008**, *24*, 1803−1810.

(36) Oguey, C.; Foloppe, N.; Hartmann, B. Understanding the Sequence-Dependence of DNA Groove Dimensions: Implications for DNA Interactions. *PLoS One* **2010**, *5*, e15931.

(37) Blanco, F. J.; Montoya, G. Transient DNA/RNA-Protein Interactions. *FEBS J.* **2011**, *278*, 1643−1650.

(38) Mauro, S. A.; Pawlowski, D.; Koudelka, G. B. The Role of the Minor Groove Substituents in Indirect Readout of DNA Sequence by 434 Repressor. *J. Biol. Chem.* **2003**, *278*, 12955−12960.

(39) Lim, F. L.; Hayes, A.; West, A. G.; Pic-Taylor, A.; Darieva, Z.; Morgan, B. A.; Oliver, S. G.; Sharrocks, A. D. Mcm1p-Induced DNA Bending Regulates the Formation of Ternary Transcription Factor Complexes. *Mol. Cell. Biol.* **2003**, *23*, 450−461.

DOI: 10.1021/ci500705j
J. Chem. Inf. Model. 2015, 55, 784−794