

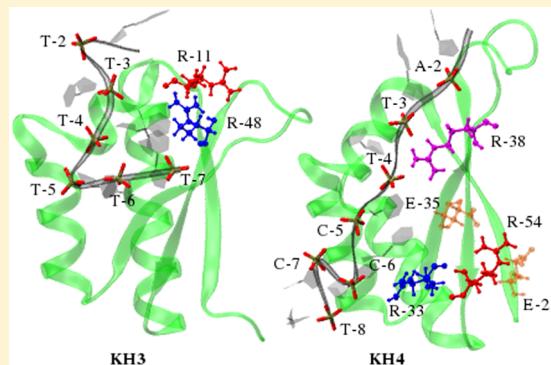
# Correlated Conformational Motions of the KH Domains of Far Upstream Element Binding Protein Complexed with Single-Stranded DNA Oligomers

Kaushik Chakraborty and Sanjoy Bandyopadhyay\*

Molecular Modeling Laboratory, Department of Chemistry, Indian Institute of Technology, Kharagpur - 721302, India

 Supporting Information

**ABSTRACT:** Single-stranded DNA binding (SSB) proteins bind with single-stranded DNA (ss-DNA) segments that are generated as intermediates during DNA metabolic processes. The primary function of an SSB protein is to protect the ss-DNA from being degraded so that other enzymes can effectively act on it. We have performed atomistic molecular dynamics simulations of the two DNA binding K homology (KH) domains (KH3 and KH4) of the far upstream element (FUSE) binding protein (FBP) complexed with two ss-DNA oligomers in aqueous solutions. Attempts have been made to study the effects of complexation on the internal motions of the protein domains and the correlated dynamics of the amino acid residue side chains. In agreement with experiments, KH3 domain has been found to be relatively more flexible in the complexed state. The calculations reveal increased long-range anticorrelated motions among several amino acid residues in the complexed forms. Compared to the KH4 domain, noticeable increase in N–H dipole ordering on complexation has been observed for the KH3 domain. Importantly, it is demonstrated that the effects of the DNA strands on the side chain orientations of the arginine and lysine residues and their ordering and dynamics play critical roles in forming the complexes and their structural stability.



## 1. INTRODUCTION

Proteins and DNA are the most abundant biomolecules which play central roles in almost all biological processes that occur in a living cell. They often interact in a cellular environment to form noncovalent protein–DNA complexes. Formation of such complexes is a key step in all aspects of genetic activities.<sup>1</sup> Proteins can bind specifically or nonspecifically with either single- or double-stranded DNA molecules through their binding domains that usually contain one or more of a relatively small group of characteristic structural motifs, such as helix-turn-helix (HTH), zinc finger, and so on. A distinct group of such proteins can specifically bind with single-stranded DNA (ss-DNA) molecules.<sup>2</sup> ss-DNAs are important intermediates in different DNA metabolic processes, such as replication, transcription, recombination, repair, and so on.<sup>3</sup> Generally, ss-DNA binding proteins (or SSB proteins) bind to the single-stranded regions of the DNA in a sequence independent manner with high affinity to prevent premature annealing of the DNA. They also protect the degradation of the ss-DNA segments by nucleases and stabilize the DNA secondary structure to allow it to exhibit its function effectively.<sup>4,5</sup>

One important point in protein–DNA binding is how the binding domain of the protein recognizes the target DNA base sequences by exploring its surface features.<sup>6</sup> Besides, both the protein and the DNA are expected to undergo some degree of structural adaptations and correlated conformational motions

during the recognition process as well as in the bound state. Such correlated motions are crucial in determining the recognition pathways and the stability of the bound complexes.

Considering the importance of the problem, various interesting aspects of protein–DNA complexations have been studied over the years using different experimental approaches. Structural aspects of different complexes formed between SSB proteins and ss-DNA oligomers have been widely studied using X-ray crystallography.<sup>7–12</sup> In an important work, Waksman and co-workers<sup>13</sup> determined the crystal structure of SSB from *Escherichia coli* bound with full-length homotetrameric ss-DNA, and showed for the first time that unlike the ordered N-terminal domain, the C-terminal domain of the SSB remains disordered in the complexed form. It is believed that the disordered C-terminal domain facilitates the binding by interacting with other components away from the binding domain during the replication process. On the other hand, by comparing the free and bound forms of replication protein A (RPA), Bochkareva et al.<sup>14</sup> showed that the DNA binding is associated with major structural changes of the protein binding

**Special Issue:** Biman Bagchi Festschrift

**Received:** February 19, 2015

**Revised:** March 27, 2015

site. In an early review, Gorenstein<sup>15</sup> discussed that <sup>31</sup>P nuclear magnetic resonance (NMR) spectroscopy can be used as an effective tool to study various conformational and dynamical properties of protein–DNA complexes. Interaction between the eukaryotic transcriptional cofactor PC4 and ss-DNA was studied using NMR by Werten et al.<sup>16</sup> By monitoring the amide resonance shift it was shown that the binding occurred through two antiparallel channels present in PC4. NMR spectroscopy has also been used successfully to understand the mechanism for displacement of the SSB protein from ss-DNA upon RecO binding.<sup>17</sup> Fluorescence spectroscopy is another important tool for investigating protein–DNA interactions.<sup>18–21</sup> The stoichiometry of binding and equilibrium binding constants of protein–DNA complexes can often be determined by probing the changes in corresponding fluorescence emission spectrum. In an important work, Kozlov and Lohman<sup>18</sup> explored the binding mechanism of the homotetrameric *E. coli* SSB protein with single-stranded deoxyoligo-nucleotides ((dT)<sub>70</sub> and (dT)<sub>35</sub>) using fluorescence stopped-flow kinetics method. It has been shown that the binding process occurs rapidly with large rate constant of about  $10^9 \text{ M}^{-1} \text{ s}^{-1}$ . Fluorescence resonance energy transfer (FRET) analysis using external probe has been carried out to study the diffusion of a tetrameric SSB protein along the ss-DNA component.<sup>19</sup> Such diffusional motion of a protein on ss-DNA showed how a SSB protein can be redistributed while remaining bound to the ss-DNA segment during DNA metabolism. Similar spontaneous diffusions of SSB proteins along ss-DNAs have also been demonstrated recently.<sup>22</sup> Several comparative studies on the binding affinities of different SSB proteins toward ss-DNA fragments are also reported in the literature.<sup>23,24</sup> In another interesting study, Nguyen et al.<sup>25</sup> probed the heterogeneous conformational dynamics of ss-DNA bound with Gene 5 protein (gSP). They demonstrated that binding with the protein results in reduced flexibility of the DNA oligomer with significant increase in unstacked base populations. In addition, different other techniques, such as atomic force microscopy (AFM),<sup>26–29</sup> electron microscopy,<sup>30,31</sup> mass spectroscopy,<sup>32,33</sup> and so on, have also been used to investigate the interactions between SSB proteins and ss-DNAs, and the structural properties of the complexes formed between those.

Computer simulation capable of providing microscopic informations at different time resolutions can be an important alternative tool to probe protein–DNA interactions. There are several reports on simulation studies of protein–DNA complexes formed by double-stranded DNAs.<sup>34–40</sup> However, despite its importance, systematic attempts have not been made to probe such complexes formed between ss-DNAs and SSB proteins from simulation studies. Recently, Carra and Cucinotta<sup>41</sup> performed MD simulations to study the interactions between the eukaryotic replication protein (RPA) and ss-DNA, and estimated the binding free energy between them. In this work, we have used atomistic MD simulations to study in detail the microscopic properties of the complex formed between the K homology (KH) binding domains of the far upstream element (FUSE) binding protein (FBP) bound to the single-stranded FUSE present 1500 base pairs upstream of the human *c-myc* promoter.<sup>42–44</sup> It is known that FBP specifically binds with single-stranded A/T rich FUSE in vitro and in vivo and regulates *c-myc* expression.<sup>45</sup> Overexpression of *c-myc* is associated with many human cancers, and thus often it is necessary to disrupt FBP activity to inhibit the proliferation of *c-myc*-dependent tumors.<sup>43</sup> In particular, attempts have been

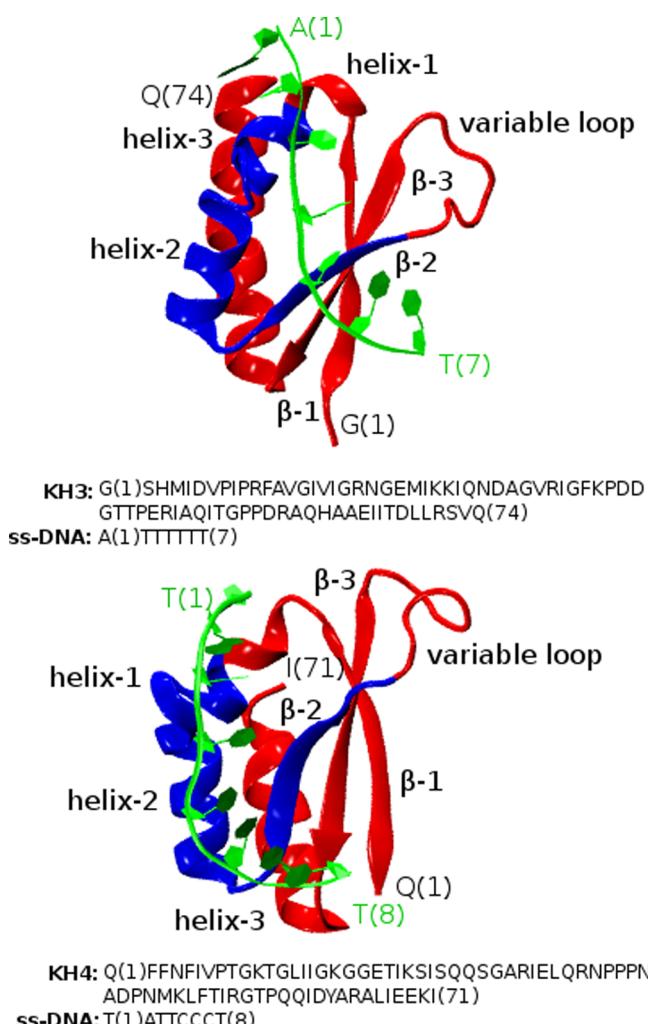
made to explore the effects of complexation on the internal motions of the protein domains and the microscopic dynamics of the amino acid residue side chains.

There are three structural domains in FBP containing 658 amino acid residues, namely, N-terminal, central, and C-terminal domains. The N-terminal domain composed of an amphipathic helix is involved in repression activity, whereas the C-terminal domain takes part in mediating the regulation of FBP and controlling its trans-activating property.<sup>46,47</sup> Further regulatory activity of FBP occurs at the central DNA binding domain. This part is also the binding region of FBP interacting repressor (FIR), and consists of four KH repeat units separated by glycine rich flexible linkers of variable lengths.<sup>44</sup> Each of these four KH domains is small (~70 amino acids) and can recognize up to four nucleotides ( $\alpha\beta$  nucleic acids) with varying affinity and specificity.<sup>48,49</sup> Clore and co-workers<sup>44</sup> from NMR studies solved the solution structure of the complex formed between the KH3 and KH4 domains of FBP and 9–10 bases each (with a separation of 5 bases) of the ss-DNA fragment present between 1525 to 1553 bases of the *c-myc* gene. The experimental structure of the two domains bound to the respective ss-DNA fragments are displayed in Figure 1. As evident from the figure, three  $\alpha$ -helices in each of the two domains are bundled on the top of a three stranded antiparallel  $\beta$ -sheet, arranged in  $\beta_1-\alpha_1-\alpha_2-\beta_2-\beta_3-\alpha_3$  topology. For convenience, we denote the different secondary structures throughout our discussion as  $\beta$ -1 (KH3: Gly-1 to Pro-10; KH4: Gln-1 to Pro-8), helix-1 (KH3: Arg-11 to Ile-18; KH4: Thr-9 to Ile-16), helix-2 (KH3: Glu-23 to Gly-33; KH4: Glu-21 to Gly-31),  $\beta$ -2 (KH3: Val-34 to Lys-39; KH4: Ala-32 to Gln-37),  $\beta$ -3 (KH3: Glu-47 to Gly-54; KH4: Lys-49 to Arg-54), and helix-3 (KH3: Pro-55 to Gln-74; KH4: Gly-55 to Ile-71). Each domain contains an invariant GXXG loop (Gly-19, Arg-20, Asn-21, Gly-22 in KH3 and Gly-17, Lys-18, Gly-19, Gly-20 in KH4) connecting helices 1 and 2, and a variable loop between  $\beta$ -2 and  $\beta$ -3 (KH3: Pro-40 to Pro-46; KH4: Arg-38 to Met-48). In each case, the ss-DNA fragment binds in the groove formed between helix-1, helix-2 and GXXG loop on one side, and the  $\beta$ -2 strand and variable loop on the other side. Primarily, the center of the groove consists of hydrophobic residues, whereas the edges contain hydrophilic polar and charged residues. The exposed bases of the ss-DNA interact with the hydrophobic center of the groove and the hydrophilic sugar–phosphate backbone binds to the more polar edges of the groove. The core binding base sequences of the DNA oligomers to KH3 and KH4 domains are 5'-TTTT and 5'-TATTC, respectively.<sup>50</sup>

This rest of the article is organized as follows. In section 2, we describe in brief the procedure adopted to setup of the systems and the simulation methods employed. The results are presented and discussed in section 3. Finally, the important findings and the conclusions reached from our study are highlighted in section 4.

## 2. SYSTEM SETUP AND SIMULATION METHODS

Six different simulations (denoted as S1–S6) were carried out using the NAMD code.<sup>51</sup> In simulations S1 and S2, aqueous solutions of the KH3–DNA and KH4–DNA complexes were studied. The initial coordinates for these two systems were taken from the NMR study of Clore and co-workers,<sup>44</sup> as reported in the Protein Data Bank (PDB ID: 1J4W). After capping the end residues of the protein and the DNA components the two complex structures (S1 and S2) were immersed separately in two large cubic cells containing



**Figure 1.** Experimental structures of the KH3 and KH4 domains of the FUSE binding protein (FBP) bound to the respective ss-DNA fragments.<sup>44</sup> The primary amino acid sequence of each of the two domains and the corresponding base sequence of the DNA oligomer (in one-letter code) are included. The binding motifs of the two domains are drawn in blue and the rest in red, and the DNA fragments in green. The secondary structures of the two domains along with their terminal residues and that of the DNA fragments are marked for convenience.

equilibrated water molecules. To avoid unfavorable contacts with the solvent, the insertion process for each system was carried out by carefully removing those water molecules that were found within 2 Å from the complex structure. To neutralize the overall charges of the two complexes, 6 Na<sup>+</sup> ions were added to system S1 and 5 Na<sup>+</sup> ions to system S2. Finally, system S1 contained the KH3–DNA complex (1372 atoms) solvated by 6240 water molecules and 6 Na<sup>+</sup> ions, and system S2 contained the KH4–DNA complex (1378 atoms) solvated by 6235 water molecules and 5 Na<sup>+</sup> ions. The initial edge length of the cubic simulation cell in each case was 60 Å.

Both the systems in S1 and S2 were first minimized using the conjugate gradient method.<sup>51</sup> The temperature of each system was then gradually increased to the room temperature of 300 K within a short simulation period of about 100 ps under the isothermal–isobaric ensemble (NPT) conditions at a constant pressure ( $P = 1$  atm). This was followed by NPT equilibration run of about 5 ns duration at 300 K for each. The system

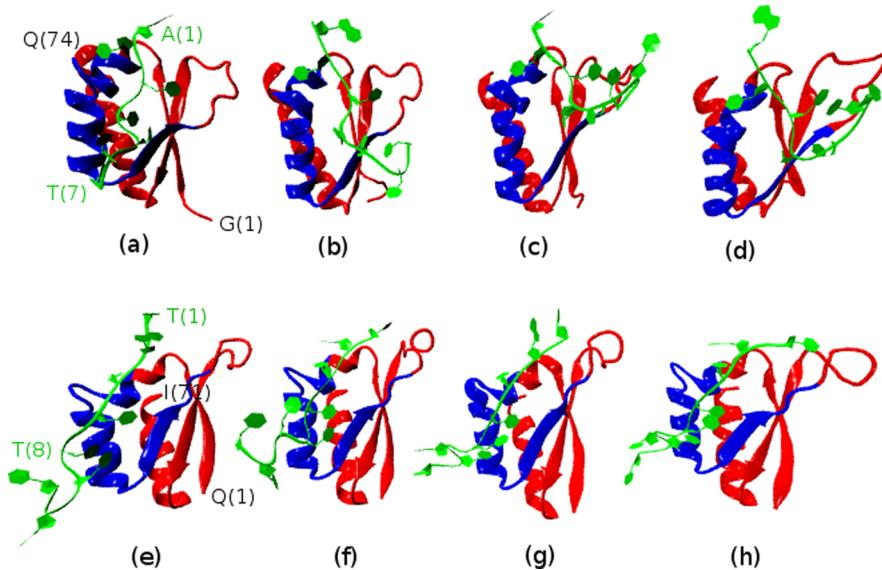
temperatures were controlled by using the Langevin dynamics method with a friction constant 1 ps<sup>-1</sup>, whereas the pressures were controlled by the Nosé–Hoover Langevin piston method.<sup>52</sup> The simulation cell volumes were allowed to fluctuate isotropically during this period to attain the appropriate densities. At the end of these two NPT runs, the cell volumes attained steady values with edge lengths 58.19 and 58.09 Å for systems S1 and S2, respectively. After this, the simulation conditions were changed from constant pressure and temperature (NPT) to that of constant volume and temperature (NVT). For each system, the NVT equilibration run was then continued further at 300 K for another 5 ns duration, followed by a long NVT production run of about 100 ns duration.

The free uncomplexed KH3 and KH4 domains of the protein starting from their structurally adapted forms were studied under identical conditions in simulations S3 and S4, respectively. Similarly, the coordinates of the ss-DNA oligomers bound to the KH3 and KH4 domains were taken from their respective complex structures to initiate the simulations of their free forms, denoted as S5 and S6, respectively. Though, the KH3 domain was charge neutral in S3, but two negative counterions (Cl<sup>-</sup>) were added to neutralize the KH4 domain charge in S4. On the other hand, the charges on the DNA strands in systems S5 and S6 were neutralized by adding 6 and 7 Na<sup>+</sup> counterions, respectively. Each of these four simulations (S3 to S6) was then carried out for about 100 ns duration following the procedure similar to that described above for systems S1 and S2. The edge lengths of the simulation cells were 58.2 and 58.13 Å, respectively, for the free KH3 (S3) and free KH4 (S4) domains, and 59.1 and 58.56 Å, respectively, for the corresponding free ss-DNA oligomers (S5 and S6).

All the simulations were carried out with a time step of 1 fs, while the trajectories were stored with a time resolution of 400 fs for subsequent analysis. The minimum image convention<sup>53</sup> was employed to calculate the short-range Lennard-Jones interactions using a spherical cutoff distance of 12 Å with a switch distance of 10 Å. The long-range electrostatic interactions were calculated by using the particle-mesh Ewald (PME) method.<sup>54</sup> We have employed the all atom CHARMM force fields for the protein<sup>55</sup> and the DNA molecules,<sup>56,57</sup> while the TIP3P model,<sup>58</sup> that is consistent with the chosen protein and DNA force fields, was employed for water.

### 3. RESULTS AND DISCUSSION

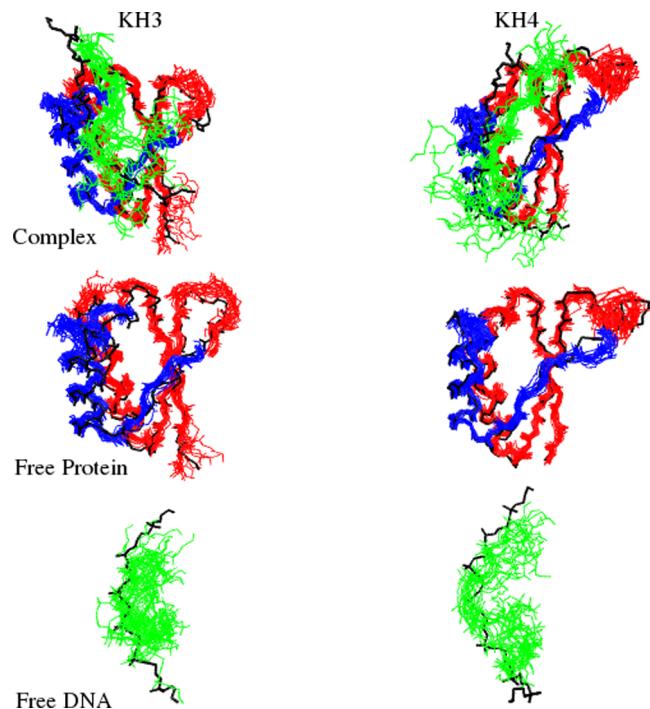
**3.1. Structural Features of the Complexes.** It is important to probe the overall structural features of the two simulated complexes formed between the KH3 and KH4 domains with the corresponding DNA strands. For that, we first show a few representative configurations as obtained from simulations S1 and S2 in Figure 2. For comparative analysis, we have included similar snapshots of configurations of the free protein and the DNA components as obtained from simulations S3 to S6 in the Supporting Information (see Figures SI-1 and SI-2). We notice that the structures of the two protein domains are significantly rigid in their complexed and free forms. Besides, comparison with Figure 1 shows that the simulated protein domains in the KH3–DNA and KH4–DNA complexes are structurally similar to that of their NMR forms.<sup>44</sup> This is particularly evident for the protein segments that are directly involved in DNA binding. It is known that the ss-DNA oligomers are in general flexible in aqueous medium,<sup>59–61</sup> which is also evident from our analysis (Figure SI-2).



**Figure 2.** Snapshots of a few representative configurations of the complex formed by the KH3 domain (a–d) and that formed by the KH4 domain (e–h) at 25, 50, 75, and 100 ns of the NVT production runs as obtained from simulations S1 and S2, respectively. The coloring scheme adopted in the figure is the same as in Figure 1.

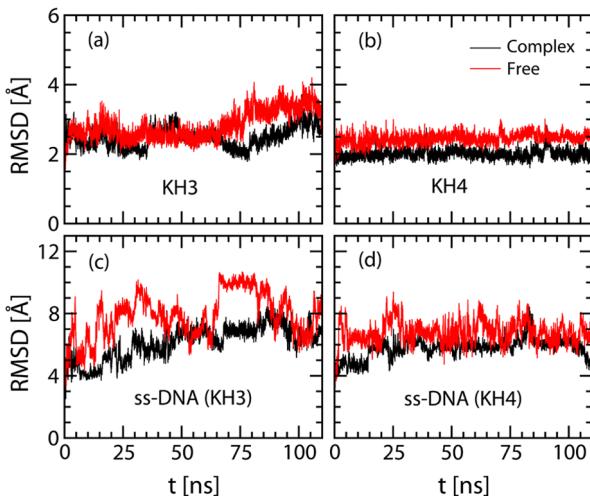
Interestingly, it is apparent from Figure 2 that compared to the protein domains the bound DNA strands exhibit greater flexibility in the two complexes. To further explore the flexibility of the protein and the DNA components in their complexed forms, in Figure 3 we have superimposed the backbone atoms of the protein ( $N$ ,  $C$ , and  $C_{\alpha}$ ) and the DNA (nonhydrogen sugar and phosphate group atoms) as obtained from simulations S1 and S2 on the corresponding initial structures. This is done by removing the translational and rotational degrees of freedom of the selected configurations with respect to the reference structures. For comparison, similar superimposed configurations of the simulated free protein (S3 and S4) and free DNA oligomers (S5 and S6) are also shown in the figure. It is evident that the backbones of the two protein domains are significantly rigid in aqueous medium. On the other hand, the backbones of the nonbinding DNA fragments are found to exhibit high degree of flexibility in the complexed forms.

To understand further the flexibility of the protein and the DNA components in the complexes (S1 and S2) studied and their structural differences with the corresponding experimental forms, we have calculated the root-mean-square deviations (RMSD). The results are shown in Figure 4. RMSDs of the simulated free protein domains (S3 and S4) and free ss-DNA fragments (S5 and S6) with respect to their structurally adapted forms in the two complexes are also calculated separately for comparison and included in the figure. The calculations are carried out by including the non-hydrogen atoms of the protein and the DNA over the entire duration along different simulated trajectories. It may be noted that in addition to knowing the deviations from experimental structures, RMSD data can provide important information on local conformational motions of flexible biomolecules. The figure provides several important features. First, it is clear that the DNA strands are more flexible than the protein domains, irrespective of whether they are present in the free or in the complexed forms. The calculated average RMSD values of the protein and the DNA components in the complexed and free forms, as obtained from 100 ns equilibrated trajectory for each system are listed in



**Figure 3.** Superpositions of the backbone  $N$ ,  $C$ , and  $C_{\alpha}$  atoms of the KH3 and KH4 domains, and the nonhydrogen atoms of the sugar and phosphate groups of the DNA strands as obtained from several simulated configurations of the two complex structures at an interval of 4 ns from the equilibrated trajectories of simulations S1 and S2. Corresponding superimposed configurations of the free forms of the two protein domains (simulations S3 and S4), and the DNA strands (simulations S5 and S6) are also shown for comparison. As a reference, the NMR structure is highlighted in black in all cases. The coloring scheme adopted in the figure is same as in Figure 1.

Table 1. It is noticed that on average there are 10–23% reduction in the RMSD values on complexation. The calculations reveal that the recognition process resulting in formation of such complexes is associated with restricted



**Figure 4.** Time evolutions of the RMSDs for all the nonhydrogen atoms of (a) KH3 and (b) KH4 domains in their complexed (black curves) and free (red curves) forms with respect to the NMR structure.<sup>44</sup> (c, d) The corresponding results for the ss-DNA fragments in the complexed and free forms.

**Table 1. Average RMSD (in Å) of the Two KH Domains and the DNA Strands in Their Complexed and Free Forms<sup>a</sup>**

system	component	complex	free
KH3–DNA	KH3	2.42(0.12)	2.56(0.1)
	DNA	5.63(0.48)	7.74(0.63)
KH4–DNA	KH4	2.00(0.05)	2.41(0.06)
	DNA	5.65(0.27)	6.8(0.37)

<sup>a</sup>The values in the parentheses are the standard deviations.

conformational fluctuations of both the protein and the DNA components, the effect being particularly more for the DNA strand complexed with the KH3 domain. Interestingly, among the two domains of the FBP protein, KH3 is found to be relatively more flexible in the complex than KH4. This is consistent with slightly higher order parameter value for the KH4 domain as observed in the NMR study by Clore and co-workers.<sup>44</sup> To probe the origin of greater flexibility of the KH3 domain in its complexed form, we have compared the average interaction energy (i.e., the binding energy) between the DNA and the KH3 domain with that between the KH4 domain and the corresponding DNA component. The calculated values are  $-368 (\pm 12)$  kcal mol<sup>-1</sup> and  $-815 (\pm 20)$  kcal mol<sup>-1</sup> for the KH3–DNA and KH4–DNA complexes, respectively. Thus, we find that the complex formed by the KH3 domain is weaker than that formed by KH4. It is clear that such relatively weaker binding allows the KH3 domain to exhibit greater flexibility in its complexed state.

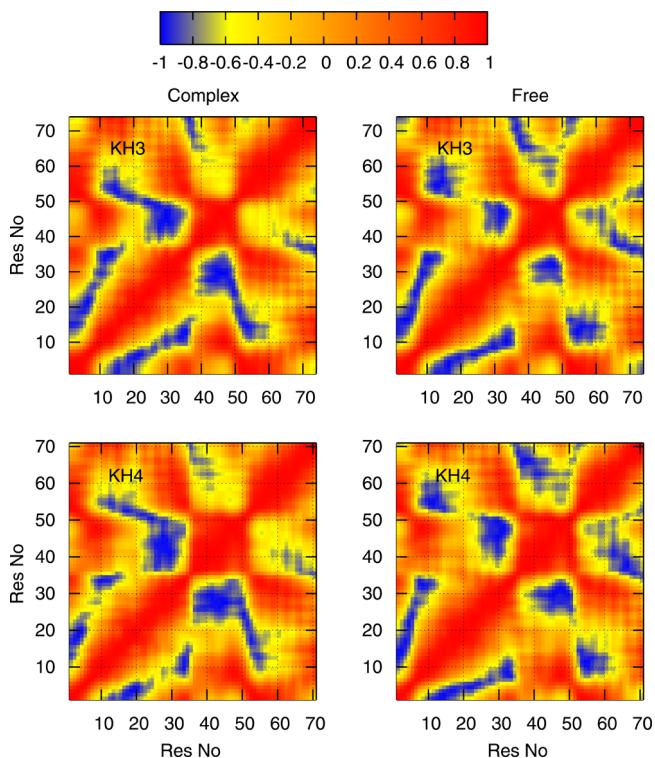
**3.2. Dynamical Features of the Complexes.** **3.2.1. Correlated Motions of Protein Segments.** The secondary structural segments forming the binding domains of the protein are expected to exhibit certain specific correlated motions during the recognition process of the target DNA bases as well as in their complexed forms. Proper analysis of such correlated motions is important to understand the microscopic properties of protein–DNA complexes. Such correlations can be short-ranged between protein residues in close proximity as well as long-ranged between well-separated residues, and are often nontrivial to probe from suitable experiments. However, simulations allow one to define

appropriate function to study such motions. Here, we probe the existence of such correlated motions, if any, by computing simple cross correlation function,  $C(i,j)$ ,<sup>62–64</sup> between the protein residues. If  $\Delta r_i$  and  $\Delta r_j$  are the displacement vectors of the  $i$ -th and the  $j$ -th  $C_\alpha$  atoms of the protein, then the corresponding function  $C(i,j)$  is defined as

$$C(i,j) = \frac{\langle \Delta r_i \cdot \Delta r_j \rangle}{\langle \Delta r_i^2 \rangle^{1/2} \langle \Delta r_j^2 \rangle^{1/2}} \quad (1)$$

where the angular brackets denote ensemble average calculations. According to the definition,  $C(i,j)$  can vary between +1 (complete correlated motion) and -1 (complete uncorrelated motion). Movements of the correlated residues occur in the same direction while the anticorrelated residues move in opposite directions. Here, it may be noted that we have used the conformation at the beginning of an analysis trajectory as the reference structure for the displacement calculations. The displacement vectors are then calculated by removing the translations and rotations of the simulated conformations with respect to the reference structure.

The  $C(i,j)$  elements in a matrix form are in general represented as a two-dimensional dynamic cross-correlation map (DCCM; with implicit time scale in the third dimension).<sup>62,63</sup> In Figure 5, we compare the DCCM for the

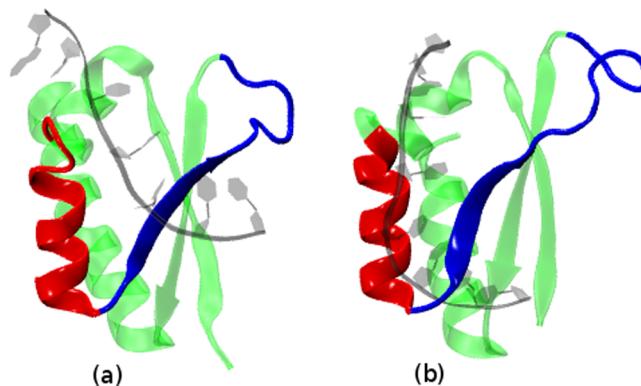


**Figure 5.** DCCM between the amino acid residues of the KH3 and KH4 domains in their complexed and free forms.  $C_{ij}$  can vary between +1 (complete correlated motion) and -1 (complete uncorrelated motion).

amino acid residues of the KH3 and KH4 domains in their complexed and free forms. It can be seen that the neighboring residues move together in the same direction and, hence, are positively correlated. Interestingly, compared to the free KH3 domain, the calculation reveals increased anticorrelated motions (with  $C(i,j)$  approaching -1) among some of the

residues (between 20 and 30) in the GRNG loop and helix-2 with the residues (between 38 and 52) in  $\beta$ -2, variable loop, and  $\beta$ -3 segment in the complexed form. On binding with the DNA strand, similar increased anticorrelated motions involving those segments have also been observed for the KH4 domain. It may be noted that these segments in the two KH domains are primarily involved in the binding process.<sup>44</sup> Thus, this is an important observation, as such long-range anticorrelated motions among the protein segments involved in binding are expected to play important roles in facilitating the approach of the flexible DNA strand during its docking within the protein's core binding groove as well as to maintain the overall structural stability of the complex formed. In contrast to the above finding, the figure further shows that upon binding with the DNA, the N-terminal helix-3 segment present in each domain exhibits small but distinctly reduced anticorrelated motions with respect to that observed among the binding segments. We believe that this is a signature of long-range cooperativity among different segments of such DNA-binding proteins in their complexed forms. However, this needs to be understood better.

In Figure 6, we show two representative conformations of the complexes formed by the KH3 and KH4 domains



**Figure 6.** Representative configurations of the complexes formed by (a) the KH3 and (b) the KH4 domains highlighting helix-2 (in red) which exhibits anticorrelated motions with  $\beta$ -2 and the variable loop segment (in blue).

highlighting helix-2 (in red) which exhibits anticorrelated motions with  $\beta$ -2 and the variable loop segment (in blue). It is known from experiments that in FBP–DNA complex the DNA strands actually bind to the grooves in the KH3 and KH4 domains formed by helices 1, 2, and GXXG loop on one side, and  $\beta$ -2 on the other side.<sup>44</sup> It is clear that increased anticorrelated motions among these segments in the complexed forms as observed in our simulations play a pivotal role during the binding process.

**3.2.2. Internal Motions and Order Parameters.** It is apparent from the above discussion that both KH3 and KH4 domains of the FBP protein undergo certain structural adaptations on binding with the ss-DNA components, and exhibit important correlated motions between the secondary structures that are involved in the process. It is also crucial to study the internal motions of the protein conformations, as the time scales associated with such motions may often determine the reaction pathways and the binding rates. NMR relaxation experiments can probe such motions by measuring the time evolutions of appropriate correlation functions and the

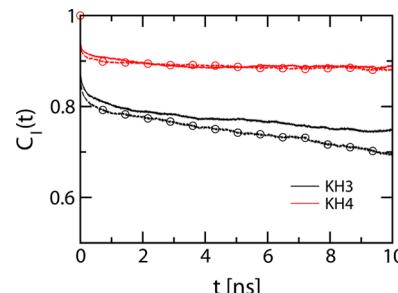
corresponding order parameters.<sup>65</sup> On the other hand, MD simulations capable of directly measuring internal motions at different time scales can provide microscopic explanation of the NMR data.<sup>66</sup> The NMR relaxation in general measures the dipolar interaction between two nuclei (protein backbone N–H dipoles in our case) that can be represented by the correlation function<sup>66,67</sup>

$$C_{\text{tot}}(t) = \langle P_2(\hat{\mu}_i(t) \cdot \hat{\mu}_i(0)) \rangle \quad (2)$$

where  $\hat{\mu}_i(t)$  is the unit dipole moment vector along the backbone N–H bond, and  $P_2(x) = (1/2)(3x^2 - 1)$  is the second order Legendre polynomial. The angular brackets denote that the calculations are carried out by averaging over all the backbone N–H bonds at different time origins. As the time scale associated with the internal motions of a large macromolecule like protein is order of magnitude faster than its overall motion, these two motions can be considered to be independent of each other. Therefore, the total correlation function  $C_{\text{tot}}(t)$  can be written as

$$C_{\text{tot}}(t) = C_O(t) \cdot C_I(t) \quad (3)$$

where  $C_O(t)$  and  $C_I(t)$  are the correlation functions corresponding to the overall and internal motions. We have calculated  $C_I(t)$  after removing the center-of-mass translational and rotational motions of the protein domains with respect to their reference structures. Relaxations of  $C_I(t)$  averaged over all the backbone N–H dipoles of the two KH domains are shown in Figure 7. To probe whether the complex formation affects



**Figure 7.** Internal correlation function ( $C_I(t)$ ) for the backbone N–H bond dipoles of the KH3 and KH4 domains of the protein in the complexed (without symbol) and in the free (with symbols) forms.

the protein's internal motions, the results are compared with the corresponding relaxation patterns of the two domains in their free forms. It can be seen that the function  $C_I(t)$  exhibits a rapid initial decay (within  $\sim 100$  ps) and then attains a plateau at longer times and does not decay further. This is true irrespective of whether the KH domains are present in the complexed or in the free forms. The rapid initial decay originates from fast vibrations/librations of the N–H bond dipoles, while the long-time plateau signifies correlated N–H dipole motions arising due to their strong interactions with the surrounding environment. Among the two KH domains, such slow long-time correlation is more evident for KH4. This is consistent with relatively more rigid conformations of KH4 in both free and complexed forms, as described earlier. Interestingly, we notice that the formation of the complex by KH3 results in distinct increase in correlations of its N–H dipole motions. However, the effect of complex formation on the N–H dipole motions in KH4 has been found to be minimum. Importantly, our calculation reveals that relatively

greater flexibility of the variable loop segment backbone (see Figure SI-3 of the Supporting Information) in KH3 contributes to its relatively faster overall N–H dipolar motions. To compare the results with NMR relaxation data,<sup>44</sup> we have fitted the  $C_1(t)$  relaxations with Lipari–Szabo model-free approach,<sup>68</sup> defined as

$$C_1(t) = S^2 + (1 - S^2)e^{-t/\tau_e} \quad (4)$$

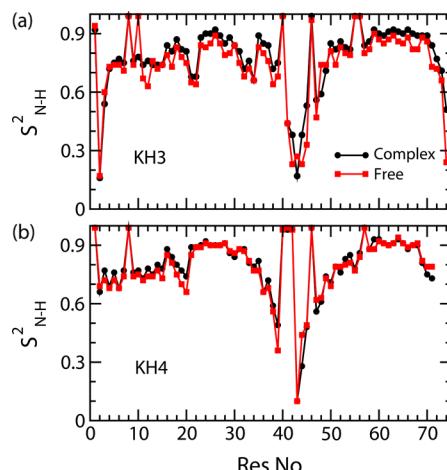
where  $S^2$  and  $\tau_e$  are the order parameter and effective internal correlation time for the protein N–H dipole. The value of  $S^2$  can vary between 0 and 1, where 1 signifies total rigid state and 0 indicates completely unrestricted motion of the N–H dipole. The calculated values of  $S^2$  and  $\tau_e$  for the two KH domains in their complexed and free forms are listed in Table 2. For

**Table 2. Order Parameter ( $S^2$ ) and the Effective Correlation Time ( $\tau_e$ ) for the Backbone N–H Bond Dipoles of the Two KH Domains in Their Complexed and Free Forms<sup>a</sup>**

domain	$S^2$		$\tau_e$ (ns)	
	complex	free	complex	free
KH3	0.77 (0.67)	0.74	1.8 (4.1)	0.96
KH4	0.89 (0.7)	0.89	2.5 (3.6)	2.3

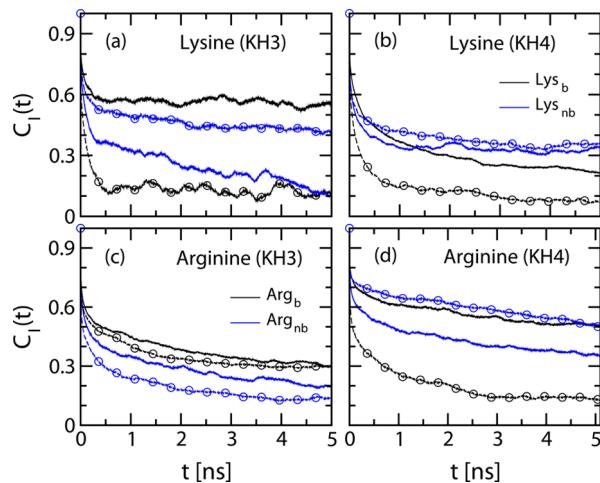
<sup>a</sup>The corresponding values as reported from NMR study<sup>44</sup> are listed in the parentheses for comparison.

comparison, the corresponding values as obtained from NMR relaxation measurements<sup>44</sup> are included in the parentheses. The data reveal that among the free forms of the two domains, the N–H dipoles in KH4 are more ordered with corresponding correlation time more than twice longer than that in KH3. In addition, significant increase in rigidity of the KH3 domain on complex formation is evident from increased  $S^2$  and  $\tau_e$  values. In particular, there is nearly a factor of 2 increase in the N–H bond correlation time for KH3 in its complexed form. On the other hand, there is no detectable evidence of increased rigidity of the KH4 domain in the complex. It may be noted that though the calculated order parameter values for the two domains in the complexed forms agree reasonably well with NMR data, but the relaxation times differ significantly. We believe that such discrepancy between the calculated and experimental results arise as in the NMR study<sup>44</sup>  $\tau_e$  values were measured for the entire FBP protein with the KH domains connected by a linker, whereas in simulations the two domains (KH3 and KH4) are studied separately. As a result, the simulated conformations of the two complexes are less rigid than the experimental structures with lower N–H bond correlation times. In Figure 8 we show the variation of backbone N–H  $S^2$  along the amino acid sequence of the two domains in their complexed and free forms. Overall, the variations for the KH4 domain residues in the complex are similar to that in the free form. In comparison, residue backbone ordering increases in the complexed form for KH3. This is consistent with the effects of complex formations on the internal motions of the two domains, as discussed before (Figure 7). Importantly, we notice that the residue backbones in the variable loop segments remain less-ordered even in the complexed forms as evident from Figure 8. In agreement with Figure 7, this is apparent for majority of the residues in the variable loop of the KH3 domain.



**Figure 8.** Residuewise backbone N–H order parameter ( $S^2$ ) of (a) the KH3 and (b) the KH4 domains of the protein in the complexed and free forms.

**3.2.3. Protein Side Chain Dynamics.** Side chain motions of specific amino acid residues often play important roles in controlling the unique three-dimensional structure of a protein and its activity. In particular, electrostatic interactions between cationic protein residues and the negatively charged DNA backbones are often crucial in determining sequence-specific protein–DNA recognition.<sup>69,70</sup> Thus, it is important to study the influence of complex formation on the side chain dynamics of cationic amino acids to obtain a microscopic understanding of protein–DNA interactions. However, despite its importance, formation of ion pairs or salt bridges between the DNA phosphate groups and the positively charged protein residues (lysines and arginines) and their role in forming protein–DNA complexes have not been understood well. In this section, we study the effect of protein–DNA interactions on the side chain motions of the lysine (Lys) and the arginine (Arg) residues in the two KH domains. This is done by calculating the relaxations of the function  $C_1(t)$  (see eqs 2 and 3) for the terminal C–N bond dipoles of the positively charged side chains of the Lys and the Arg residues. Further, to better understand the effect of complex formation on the side chain motions, we have probed  $C_1(t)$  relaxations separately for those Lys and Arg residues that are directly involved in the binding of the two protein domains with the DNA fragments (Lys-26, Arg-35, and Arg-48 for KH3, and Lys-24, Lys-49, Arg-33, and Arg-38 for KH4) and those that are not (Lys-27, Lys-39, Arg-11, Arg-20, Arg-58, and Arg-71 for KH3, and Lys-11, Lys-18, Lys-70, Arg-54, and Arg-64 for KH4). For convenience, we denote the binding Lys and Arg residues as  $Lys_b$  and  $Arg_b$ , whereas the corresponding nonbinding residues as  $Lys_{nb}$  and  $Arg_{nb}$ , respectively. The results are displayed in Figure 9. As a reference, the corresponding relaxation patterns of the function for the two KH domains in their free forms are also shown in the figure. It may be noted that compared to the protein backbone, the function  $C_1(t)$  relaxes significantly faster for the Lys and Arg side chains. This is consistent with greater flexibility and more dynamic nature of residue side chains as compared to the backbone of a protein. Besides, compared to small changes in protein backbone conformational motions (see Figure 7), the side chain dynamics of the Lys and Arg residues exhibit significant changes on complexation. However, more importantly, a closer look at the results reveal interesting effects of



**Figure 9.** Internal correlation function ( $C_i(t)$ ) for the side chain C–N bond dipoles of the lysine and arginine residues that are directly involved in binding ( $Lys_b$  and  $Arg_b$ ) and those that are not ( $Lys_{nb}$  and  $Arg_{nb}$ ) for the KH3 (a, c) and KH4 (b, d) domains of the protein in the complexed (without symbol) and in the free (with symbols) forms.

complex formation on the dynamics of the cationic side chains of these residues. First, we notice that compared to the free protein domains, the  $Lys_b$  and  $Arg_b$  residues that are directly involved in binding with the DNA strands exhibit noticeably slower relaxations in the complexed forms. This is a signature of conformational immobilization of these residue side chains due to their increased ordering (correlated local motions) on forming contacts with the DNA phosphate groups. We have extracted the order parameters ( $S^2$ ) and correlation times ( $\tau_e$ ) for these side chains by fitting the decay curves in Figure 9 following eq 4. The data are listed in Table 3 for all the systems.

**Table 3. Order Parameter ( $S^2$ ) and the Effective Correlation Time ( $\tau_e$ ) for the Side Chain C–N Bond Dipoles of the Binding and Nonbinding Lysine ( $Lys_b$  and  $Lys_{nb}$ ) and Arginine ( $Arg_b$  and  $Arg_{nb}$ ) Residues of the Two KH Domains in Their Complexed and Free Forms**

domain	residue	$S^2$		$\tau_e$ (ns)	
		complex	free	complex	free
KH3	$Lys_b$	0.56	0.13	2.13	0.48
	$Lys_{nb}$	0.20	0.42	0.53	1.05
	$Arg_b$	0.29	0.29	1.60	0.98
	$Arg_{nb}$	0.23	0.16	1.25	1.11
KH4	$Lys_b$	0.26	0.10	1.12	0.39
	$Lys_{nb}$	0.33	0.35	0.47	0.88
	$Arg_b$	0.49	0.18	1.89	1.16
	$Arg_{nb}$	0.36	0.45	1.46	4.36

Note the enhanced ordering of almost all the  $Lys_b$  and  $Arg_b$  side chains (except for the  $Arg_b$  residues in KH3) with a factor of 2–4 times increase in  $S^2$  values on complex formation. Immobilization of these residue side chains is reflected in longer C–N bond correlation times ( $\tau_e$ ) in the two complexes. The results indicate that compared to the  $Arg_b$  residues, the effect is more for the  $Lys_b$  residues in the two complexes. On the other hand, the effect of complex formation on the conformational motions of the side chains of  $Lys_{nb}$  and  $Arg_{nb}$  residues that are not directly involved in DNA binding reveals differential behavior. We notice that the side chains of  $Arg_{nb}$

residues in KH3 exhibit slower dynamics in the complex as compared to that in the free form with ~40% increase in the order parameter ( $S^2$ ) and ~12% increase in the correlation time ( $\tau_e$ ) of the corresponding C–N bond dipoles (see Table 3). However, in contrast, the side chains of  $Arg_{nb}$  residues in KH4 and  $Lys_{nb}$  residues in both KH3 and KH4 exhibit noticeably faster conformational motions in the complexed structures. This is reflected in reduced  $S^2$  and  $\tau_e$  values (see Table 3) for most of these side chains on complexation. This is a unique observation that suggests that instead of expected conformational immobilization, positively charged long side chains of Lys and Arg residues that do not directly bind to form contact pairs with DNA phosphate groups can often exhibit more flexibility in protein–DNA complexes. Such faster dynamics of these residue side chains is expected to reduce the entropic cost and hence facilitate the binding between the protein domains and the DNA strands. In a recent NMR study, Iwahara and co-workers<sup>70</sup> showed increased mobility of lysine residue side chains of a protein on complexation with DNA. Our result not only agrees well with NMR data, but it further demonstrates that formation of a protein–DNA complex may have contrasting influence on the dynamic nature of the positively charged amino acid residues depending on whether they are involved in forming direct contacts with the DNA molecule or not.

We now attempt to probe the origin of the differential effects of complex formations on the dynamics of the lysine and arginine residues of the two KH domains. For that, we first calculate the interaction energy ( $E_R$ ) of each of these residues with rest of the system averaged over the simulated trajectories of the two complexes (simulations S1 and S2). For comparison, the corresponding  $E_R$  values in the free uncomplexed forms of KH3 and KH4 domains are also determined (simulations S3 and S4). The calculated  $E_R$  values along with the corresponding changes on complexations ( $\Delta E_R = E_R^{\text{complex}} - E_R^{\text{free}}$ ) are listed in Table 4. It can be seen that in most cases  $Lys_b$  and  $Arg_b$  residues (except  $Lys_b$ -49 in KH4) interact strongly with the DNA components, as evident from noticeable lowering of the  $E_R$  values with respect to the corresponding free protein domains. This is consistent with increased ordering and restricted conformational motions of the side chains of these residues that directly form contacts with the DNA phosphate groups as discussed above and thus help stabilizing the formation of such complexes. On the other hand, most of the  $Lys_{nb}$  and  $Arg_{nb}$  residues do not exhibit any noticeable gain (or loss) in their  $E_R$  values on complexation. However, several of these become more flexible in the complexed states as discussed above, and thus also contribute in the binding process by compensating the entropy loss due to conformational immobilizations of other components of the protein domains. Importantly, in contrast to the general trend,  $Arg_{nb}$ -11 in KH3 gains noticeable interaction energy ( $\Delta E_R = -3.06 \text{ kcal mol}^{-1}$ ) on complex formation. Thus, it appears that though  $Arg_{nb}$ -11 in KH3 is not directly involved in binding with the DNA, but its orientation in the complex results in favorable interaction, which in turn leads to its restricted motions in the complex, as discussed earlier.

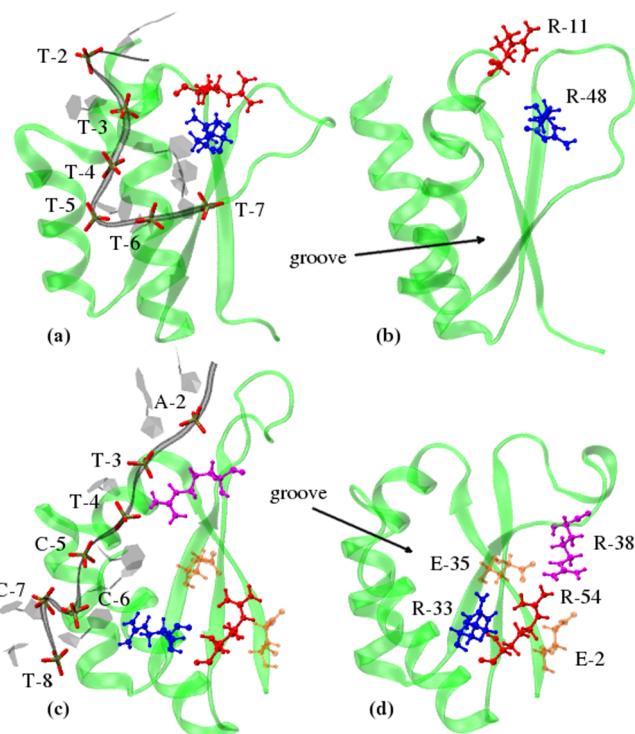
To obtain a microscopic understanding of the reason behind such unique behavior of  $Arg_{nb}$ -11 in KH3, we have probed the effect of complexation on its orientational preference, and compared that with similar effect on  $Arg_{nb}$ -54 in KH4 (which appears to lose  $E_R$  to some extent). On closely monitoring the trajectories of the two complexes (simulations S1 and S2) and

**Table 4.** Average Interaction Energy ( $E_R$ ; in kcal mol<sup>-1</sup>) of Each of the Lysine and Arginine residues Present in the Two KH Domains in Their Complexed and Free Forms; The Interaction Energy Change on Complexation ( $\Delta E_R$ ) for Each of These Residues are Also Listed<sup>a</sup>

domain	residue	$E_R$ (kcal mol <sup>-1</sup> )		$\Delta E_R$ (kcal mol <sup>-1</sup> ; $E_R^{\text{complex}} - E_R^{\text{free}}$ )
		complex	free	
KH3	Lys <sub>b</sub> -26	-14.11(0.92)	-11.72(0.72)	-2.39
	Arg <sub>b</sub> -35	-15.07(0.86)	-13.98(0.72)	-1.09
	Arg <sub>b</sub> -48	-16.48(0.87)	-14.16(0.99)	-2.32
	Lys <sub>nb</sub> -27	-10.80(0.90)	-10.85(0.89)	0.05
	Lys <sub>nb</sub> -39	-10.01(0.60)	-9.28(0.63)	-0.73
	Arg <sub>nb</sub> -11	-15.22(1.23)	-12.16(0.94)	-3.06
	Arg <sub>nb</sub> -20	-12.65(0.66)	-11.53(0.78)	-1.12
	Arg <sub>nb</sub> -58	-16.30(0.88)	-16.49(0.91)	0.19
	Arg <sub>nb</sub> -71	-14.11(0.92)	-13.98(0.72)	-0.13
KH4	Lys <sub>b</sub> -24	-16.10(0.75)	-11.36(0.92)	-4.74
	Lys <sub>b</sub> -49	-11.72(1.12)	-11.52(0.84)	-0.2
	Arg <sub>b</sub> -33	-15.94(0.91)	-13.35(0.90)	-2.59
	Arg <sub>b</sub> -38	-14.07(0.87)	-11.51(1.00)	-2.56
	Lys <sub>nb</sub> -11	-12.24(0.67)	-12.03(1.00)	-0.21
	Lys <sub>nb</sub> -18	-9.32(0.81)	-8.43(0.89)	-0.89
	Lys <sub>nb</sub> -70	-9.37(0.69)	-9.81(0.99)	0.44
	Arg <sub>nb</sub> -54	-15.05(0.83)	-15.83(0.81)	0.78
	Arg <sub>nb</sub> -64	-13.97(0.94)	-14.40(0.95)	0.43

<sup>a</sup>The values in the parentheses are the standard deviations.

that of the free uncomplexed protein domains (simulations S3 and S4), we notice interesting differential behaviors between the side chains of Arg<sub>nb</sub>-11 in KH3 and Arg<sub>nb</sub>-54 in KH4. Representative snapshots of the two complexes and the corresponding free forms of the protein domains highlighting the orientations of these Arg<sub>nb</sub> residues along with that of Arg<sub>b</sub>-48 in KH3, and Arg<sub>b</sub>-33 and Arg<sub>b</sub>-38 in KH4 are shown in Figure 10. We find that Arg<sub>nb</sub>-11 (in helix-1) and Arg<sub>b</sub>-48 (in  $\beta$ -3) in the free form of KH3 are exposed at the surface around the DNA binding groove formed by helices 1 and 2, GRNG loop, and  $\beta$ -2. As evident from the figure, on forming the complex Arg<sub>b</sub>-48 reorients itself more toward the groove to establish direct contacts with the DNA phosphate groups. This leads to favorable change in its interaction energy (see Table 4) with significant contribution in restricting the dynamics of Arg<sub>b</sub> residue side chains (see Figure 9 and Table 3). On the other hand, though Arg<sub>nb</sub>-11 is not involved in direct binding, but it too changes its orientation more closer to the groove region in the presence of the DNA. This reduces its exposure at the surface with sufficient gain in its interaction energy (see Table 4). Additionally, presence of the DNA has been found to bring Arg<sub>nb</sub>-11 and Arg<sub>b</sub>-48 closer with the distance between their side chain terminal carbon atoms within 4–5 Å to form an Arg···Arg contact pair. Such Arg···Arg pairs are found in many proteins that are often mediated by the presence of acidic residues or negatively charged bound ligands.<sup>71</sup> In this case the presence of negatively charged DNA phosphate groups bound to the protein mediates formation of the Arg<sub>nb</sub>-11···Arg<sub>b</sub>-48 pair. It is evident that such orientational preference of Arg<sub>nb</sub>-11 in the complex dominates overall increased ordering and restricted motions of Arg<sub>nb</sub> residues in KH3. Interestingly, rearrangement of the orientations of Arg<sub>b</sub>-33, Arg<sub>b</sub>-38, and Arg<sub>nb</sub>-54 with respect to each other in the KH4 domain on



**Figure 10.** (a, b) Orientational preference of the side chains of Arg<sub>b</sub>-48 (R-48) and Arg<sub>nb</sub>-11 (R-11) present around the binding groove region of the KH3 domain in its complexed and free forms. (c, d) The corresponding picture for the side chains of Arg<sub>b</sub>-33 (R-33), Arg<sub>b</sub>-38 (R-38), and Arg<sub>nb</sub>-54 (R-54) present around the binding groove region of the KH4 domain in its complexed and free forms. For clarity, the orientations of Glu-2 (E-2) and Glu-35 (E-35) stabilizing arginine pairing in free KH4 are also shown in both cases.

forming the corresponding complex present a different picture. As shown in Figure 10, Arg<sub>nb</sub>-54 in the free uncomplexed KH4 domain simultaneously forms Arg···Arg pairings with Arg<sub>b</sub>-33 and Arg<sub>b</sub>-38 with the distances between the side chain terminal carbon atom of Arg<sub>nb</sub>-54 and that of Arg<sub>b</sub>-33 and Arg<sub>b</sub>-38 being less than 5 Å. The clustering of the three arginines near the surface of the protein is stabilized by two acidic residues, namely, Glu-2 in  $\beta$ -1 and Glu-35 in  $\beta$ -2 (see Figure 10) with the distances between the side chain terminal carbon atom of Arg<sub>nb</sub>-54 and that of Glu-2 and Glu-35 varying within 4–5 Å. We notice that on forming the complex with DNA, Arg<sub>b</sub>-33 and Arg<sub>b</sub>-38 break their pairings with Arg<sub>nb</sub>-54 and reorient the side chains toward the groove and form direct contacts with the DNA phosphate groups with favorable  $\Delta E_R$  values (see Table 4). Stronger interactions with the DNA result in overall increased ordering and slower dynamics of Arg<sub>b</sub> residue side chains, as discussed before (see Figure 9 and Table 3). Note that such behavior of Arg<sub>b</sub> residues in the complex formed by the KH4 domain is similar to that observed for the KH3–DNA complex. However, unlike in the KH3–DNA complex, we do not observe any noticeable orientational change of Arg<sub>nb</sub>-54 in the KH4–DNA complex. It maintains strong contacts with Glu-2 and Glu-35, and remains exposed at the surface, as evident from Figure 10. Although Arg<sub>nb</sub>-54 tends to lose  $E_R$  to a small extent in the complex (see Table 4), but the breaking of the Arg···Arg pairs with subsequent reorientations of Arg<sub>b</sub>-33 and Arg<sub>b</sub>-38 allows Arg<sub>nb</sub>-54 to exhibit greater flexibility in an increased available volume around it. This leads to reduced ordering and faster internal motions of Arg<sub>nb</sub>-54 residues side

chain in the KH4–DNA complex, as described earlier (see Figure 9 and Table 3). As a result, Arg<sub>nb</sub>-54 gains entropy in the complexed form which appears to compensate minor loss in  $E_R$  and, therefore, contributes to stabilizing the complex.

#### 4. CONCLUSIONS

In this work we have presented results obtained from separate atomistic MD simulations of the two DNA binding domains (KH3 and KH4) of the FUSE binding protein (FBP) complexed with two single-stranded DNA (ss-DNA) oligomers in aqueous solutions. The calculated results as obtained for the two complex structures are compared with those of the uncomplexed free protein domains and the ss-DNA components by simulating their conformational evolutions starting from the respective structurally adapted forms in the two complexes.

The calculations revealed that the overall structural features of the two protein domains in their complexed states are similar to that observed in experimental study.<sup>44</sup> This is particularly true for the protein segments that are directly involved in DNA binding. It is found that in general the recognition of the DNA oligomers by the protein domains results in restricted conformational fluctuations of both the domains and the DNA components, the effect being particularly noticeable for the DNA strand complexed with the KH3 domain. Inconsistent with experiments, KH3 has been found to be relatively more flexible in the complex than KH4. An examination of the effect of complexation on the microscopic correlated motions between different segments of the protein domains reveals increased long-range anticorrelated motions among several amino acid residues in the complexed forms. This is an important finding, as these residues present around the DNA binding groove regions of the two KH domains are expected not only to play important roles during the recognition process, but also to help maintaining the overall structural stability of the FBP–ss-DNA complex.

Attempts have been made to measure the time scale of internal motions of the protein domains from the simulated trajectories. Compared to KH3, slow long-time correlation in backbone N–H dipole motions has been found to be more prevalent in the KH4 domain, which is consistent with relatively more rigid conformations of KH4 in both free and complexed forms. It is demonstrated that relatively greater flexibility of the variable loop segment in KH3 is the origin of its comparatively faster N–H dipolar motions. Calculated NMR order parameters and relaxation times suggest noticeable increase in N–H dipole ordering with a factor of 2 increase in its correlation time for KH3 in the complexed state. However, such effects of complexation on KH4 backbone N–H dipolar motions are found to be minimum in our calculations.

To further understand the microscopic details of protein–DNA interactions, we have explored the effects of complex formation on the side chain dynamics of the cationic amino acid residues (lysines and arginines) present around the binding groove regions of the two KH domains. The calculations reveal conformational immobilizations of the Lys<sub>b</sub> and Arg<sub>b</sub> residue side chains due to their increased ordering and correlated local motions on forming favorable contacts with the DNA phosphate groups in both cases. In contrast, the effect of complex formation on the conformational motions of the Lys<sub>nb</sub> and Arg<sub>nb</sub> residue side chains that are not directly involved in DNA binding reveals differential behavior for the two domains. It is noticed that inspite of overall conformational immobiliza-

tion, the side chains of Lys<sub>nb</sub> and Arg<sub>nb</sub> residues can often exhibit greater local flexibility in complexed forms. The results agree well with recent NMR data,<sup>70</sup> and indicate that such faster motions of Lys<sub>nb</sub> and Arg<sub>nb</sub> residues are likely to reduce the entropic cost and hence facilitate the binding process. Importantly, we noticed unique differential dynamics between the side chains of Arg<sub>nb</sub>-11 in KH3 and Arg<sub>nb</sub>-54 in KH4. It is observed that in the presence of DNA Arg<sub>nb</sub>-11 in KH3 reorient toward the groove region and stabilize itself by forming an Arg···Arg contact pair mediated by DNA phosphates. On the other hand, it is observed that Arg<sub>b</sub>-33 and Arg<sub>b</sub>-38 involved in forming simultaneous Arg···Arg contact pairs with Arg<sub>nb</sub>-54 (stabilized by acidic residues Glu-2 and Glu-35) in free KH4 break the pairings and reorient to establish more favorable contacts with the DNA phosphate groups in the complex. This allows Arg<sub>nb</sub>-54 to exhibit greater flexibility and relatively faster dynamics in an increased available volume around it in the KH4–DNA complex. Note that the entropy gained by Arg<sub>nb</sub>-54 due to its enhanced flexibility may contribute in stabilizing the complex. It would be interesting to probe the role of Arg<sub>nb</sub>-54 in reorienting Arg<sub>b</sub>-33 and Arg<sub>b</sub>-38 residues and thus stabilizing the KH4–DNA complex by mutating Arg<sub>nb</sub>-54 with a different residue. Detailed studies are underway in our laboratory to explore the effect of mutation on the formation of such complexes.

#### ■ ASSOCIATED CONTENT

##### S Supporting Information

Figure SI-1: Snapshots of a few representative configurations of the KH3 and KH4 domains in their uncomplexed free forms at different time intervals. Figure SI-2: Snapshots of a few representative configurations of the ss-DNA fragments that bind with the KH3 and KH4 domains in their uncomplexed free forms at different time intervals. Figure SI-3: Internal correlation function ( $C_1(t)$ ) for the backbone N–H bond dipoles of different secondary structural segments of the KH3 and KH4 domains of the protein in the complexed and in the free forms. Complete ref S5. This material is available free of charge via the Internet at <http://pubs.acs.org>.

#### ■ AUTHOR INFORMATION

##### Corresponding Author

\*E-mail: sanjoy@chem.iitkgp.ernet.in.

##### Notes

The authors declare no competing financial interest.

#### ■ ACKNOWLEDGMENTS

This study was supported in part by a grant from the Department of Science and Technology (DST; SR/S1/PC-23/2007), Government of India. Part of the work was carried out using the computational facility created under DST-FIST programme (SR/FST/CSII-011/2005). K.C. thanks CSIR, Government of India, for providing a scholarship.

#### ■ REFERENCES

- (1) Nelson, D. L.; Cox, M. M. *Lehninger Principles of Biochemistry*, 3rd ed.; Worth: New York, 2000.
- (2) Shamoo, Y. Single-Stranded DNA Binding Proteins. *Encyclopedia of Life Science*; Macmillan Publishers Ltd.: New York, 2000.
- (3) Zou, L.; Elledge, S. J. Sensing DNA Damage Through ATRIP Recognition of RPA-ssDNA Complexes. *Science* 2003, 300, 1542–1548.

- (4) Smyth, T. D.; Duncan, R. C.; Zheng, T.; Michelotti, G.; Levens, D. The Far Upstream Element-Binding Proteins Comprise an Ancient Family of Single-Strand DNA-Binding Transactivators. *J. Biol. Chem.* **1996**, *271*, 31679–31687.
- (5) Cukier, C. D.; Hollingworth, D.; Martin, S. R.; Kelly, G.; Moreno, I. D.; Ramos, A. Molecular Basis of FIR-mediated c-myc Transcriptional Control. *Nat. Struct. Mol. Biol.* **2010**, *17*, 1058–1064.
- (6) Stormo, G. D.; Zhao, Y. Determining the Specificity of Protein–DNA Interactions. *Nat. Rev.* **2010**, *11*, 751–760.
- (7) Bochkarev, A.; Bochkareva, E.; Frappier, L.; Edwards, A. M. The Crystal Structure of the Complex of Replication Protein A Subunits RPA32 and RPA14 Reveals a Mechanism for Single-Stranded DNA Binding. *EMBO J.* **1999**, *18*, 4498–4504.
- (8) Ragunathan, S.; Ricard, C. S.; Lohman, T. M.; Waksman, G. Crystal Structure of the Homo-Tetrameric DNA Binding Domain of *Escherichia coli* Single-Stranded DNA-Binding Protein Determined by Multiwavelength X-ray Diffraction on the Selenomethionyl Protein at 2.9 Å Resolution. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 6652–6657.
- (9) Kelly, T. J.; Simancek, P.; Brush, G. S. Identification and Characterization of a Single-Stranded DNA-Binding Protein from the Archaeon *Methanococcus jannaschii*. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 14634–14639.
- (10) Kerr, I. D.; Wadsworth, I. M.; Cubeddu, L.; Blankenfeldt, W.; Naismith, J. H.; White, M. F. Insights into SsDNA Recognition by the OB Fold From a Structure and Thermodynamic Study of *Sulfolobus* SSB Protein. *EMBO J.* **2003**, *22*, 2561–2570.
- (11) Haung, C.-Y.; Hus, C.-H.; Sun, Y.-J.; Wu, H.-N.; Hsiao, C.-D. Complexed Crystal Structure of Replication Restart Primosome Protein PriB Reveals a Novel Single-Stranded DNA Binding Mode. *Nucleic Acids Res.* **2006**, *34*, 3878–3886.
- (12) George, N. P.; Ngo, K. V.; Pattu, S. C.; Norais, C. A.; Battista, J. R.; Cox, M. M.; Keck, J. L. Structure and Cellular Dynamics of *Deinococcus radiodurans* Single-Stranded DNA (ssDNA)-Binding Protein (SSB)–DNA Complex. *J. Biol. Chem.* **2010**, *287*, 22123–22132.
- (13) Savvides, S. N.; Ragunathan, S.; Fütterer, K.; Kozlov, A. G.; Lohman, T. M.; Waksman, G. The C-terminal Domain of Full-Length *E. coli* SSB is Disordered Even When Bound to DNA. *Protein Sci.* **2004**, *13*, 1942–1947.
- (14) Bochkareva, E.; Belegu, V.; Korolev, S.; Bochkarev, A. Structure of the Major Single-Stranded DNA-Binding Domain of Replication Protein A Suggests a Dynamic Mechanism for DNA Binding. *EMBO J.* **2001**, *20*, 612–618.
- (15) Gorenstein, D. G. Conformation and Dynamics of DNA and Protein–DNA Complex by  $^{31}\text{P}$  NMR. *Chem. Rev.* **1994**, *94*, 1315–1338.
- (16) Warten, S.; Wechselberger, R.; Boelens, R.; Vliet, P. C. V. D.; Kaptein, R. Identification of the Single-Stranded DNA Binding Surface of the Transcriptional Coactivator PC4 by NMR. *J. Biol. Chem.* **1999**, *274*, 3693–3699.
- (17) Inoue, J.; Nagae, T.; Mishima, M.; Ito, Y.; Shibata, T.; Mikawa, T. A Mechanism for Single-Stranded DNA-Binding Protein (SSB) Displacement from Single-Stranded DNA upon SSB-RecO Interaction. *J. Biol. Chem.* **2011**, *286*, 6720–6732.
- (18) Kozlov, A.; Lohman, T. Stopped-Flow Studies of the Kinetics of Single-Stranded DNA Binding and Wrapping around the *Escherichia coli* SSB Tetramer. *Biochemistry* **2002**, *41*, 6032–6044.
- (19) Roy, R.; Kozlov, A. G.; Lohman, T. M.; Ha, T. SSB Protein Diffusion on Single-Stranded DNA Stimulates RecA Filament Formation. *Nature* **2009**, *461*, 1092–1097.
- (20) Kunzelmann, S.; Morris, C.; Chavda, A. P.; Eccleston, J. F.; Webb, M. R. Mechanism of Interaction Between Single-Stranded Binding Protein and DNA. *Biochemistry* **2010**, *49*, 843–852.
- (21) Liu, J.; Choi, M.; Stanenas, A. G.; Byrd, A. K.; Raney, K. D.; Cohan, C.; Bianco, P. R. Novel, Fluorescent, SSB Protein Chimeras With Broad Utility. *Protein Sci.* **2011**, *20*, 1005–1020.
- (22) Zhang, J.; Zhou, R.; Inoue, J.; Mikawa, T.; Ha, T. Single Molecule Analysis of *Thermus thermophilus* SSB Protein Dynamics on Single-Stranded DNA. *Nucleic Acids Res.* **2014**, *42*, 3821–3832.
- (23) Yang, S.-H.; Zhou, R.; Campbell, J.; Chen, J.; Ha, T.; Paull, T. The SOSS1 Single-Stranded DNA Binding Complex Promotes DNA End Resection in Concert With Exo1. *EMBO J.* **2013**, *32*, 126–139.
- (24) Ghalei, H.; Moeller, H. V.; Eppers, D.; Sohmen, D.; Wilson, D. N.; Loll, B.; Wahl, M. C. Entrapment of DNA in an Intersubunit Tunnel System of a Single-Stranded DNA-Binding Protein. *Nucleic Acids Res.* **2014**, *42*, 6698–6708.
- (25) Nguyen, H.-N.; Zhao, L.; Gray, C. W.; Gray, D. M.; Xia, T. Ultrafast Fluorescence Decay Profiles Reveal Differential Unstacking of 2-Aminopurine from Neighboring Bases in Single-Stranded DNA-Binding Protein Subsites. *Biochemistry* **2011**, *50*, 8989–900.
- (26) Li, B. S.; Gog, M. C. Direct Evidence of the Role of ATP $\gamma$ S in the Binding of Single-Stranded Binding Protein (*Escherichia coli*) and RecA to Single-Stranded DNA. *Langmuir* **2010**, *26*, 14755–14758.
- (27) Zhang, W.; Lü, X.; Zhang, W.; Shen, J. EMSA and Single-Molecule Force Spectroscopy Study of Interactions Between *Bacillus subtilis* Single-Stranded DNA-Binding Protein and Single-Stranded DNA. *Langmuir* **2011**, *27*, 15008–15015.
- (28) Shlyakhtenko, L. S.; Lushnikov, A. L.; Miyagi, A.; Lyubchenko, Y. L. Specificity of Binding of Single-Stranded DNA-Binding Protein to Its Target. *Biochemistry* **2012**, *51*, 1500–1509.
- (29) Japrung, D.; Bahrami, A.; Nadzeyka, A.; Peto, L.; Bauerick, S.; Edel, J. B.; Albrecht, T. SSB Binding to Single-Stranded DNA Probed Using Solid-State Nanopore Sensors. *J. Phys. Chem. B* **2014**, *118*, 11605–11612.
- (30) Chrysogelos, S.; Griffith, J. *Escherichia coli* Single-Strand Binding Protein Organizes Single-Stranded DNA in Nucleosome-Like Units. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 5803–5807.
- (31) Lefebvre, S. D.; Wong, M. L.; Morrical, S. W. Simultaneous Interactions of Bacteriophage T4 DNA Replication Proteins gp59 and gp32 with Single-Stranded (ss) DNA. *J. Biol. Chem.* **1999**, *274*, 22830–22838.
- (32) Steen, H.; Petersen, J.; Mann, M.; Jensen, O. N. Mass Spectrometric Analysis of a UV-Cross-Linked Protein–DNA Complex: Tryptophans 54 and 88 of *E. coli* SSB Cross-Link to DNA. *Protein Sci.* **2001**, *10*, 1989–2001.
- (33) Politis, A.; Park, A. Y.; Hall, Z.; Ruotolo, B. T.; Robinson, C. V. Integrative Modelling Coupled with Ion Mobility Mass Spectrometry Reveals Structural Features of the Clamp Loader in Complex with Single-Stranded DNA Binding Protein. *J. Mol. Biol.* **2013**, *425*, 4790–4801.
- (34) Kosztin, D.; Bishoop, T. C.; Schulten, K. Binding of the Estrogen Receptor to DNA. The Role of Water. *Biophys. J.* **1997**, *73*, 557–570.
- (35) Jayaram, B.; McConnell, K. J.; Dixit, S. B.; Beveridge, D. L. Free Energy Analysis of Protein–DNA Binding: The EcoRI Endonuclease DNA Complex. *J. Comput. Phys.* **1999**, *151*, 333–357.
- (36) Dolenc, J.; Gerster, S.; Gunsteren, W. F. Molecular Dynamics Simulations Shed Light on the Enthalpic and Entropic Driving Forces That Govern the Sequence Specific Recognition Between Netropsin and DNA. *J. Phys. Chem. B* **2010**, *114*, 11164–11172.
- (37) Castrignano, T.; Chillemi, G.; Desideri, A. Structure and Hydration of BamHI DNA Recognition Site: Molecular Dynamics Investigation. *Biophys. J.* **2000**, *79*, 1263–1272.
- (38) Mori, M.; Dietrich, U.; Manetti, F.; Botta, M. Molecular Dynamics and DFT Study on HIV-1 Nucleocapsid Protein-7 in Complex with Viral Genome. *J. Chem. Inf. Model.* **2010**, *50*, 638–650.
- (39) Sinha, S. K.; Bandyopadhyay, S. Conformational Fluctuations of a Protein–DNA Complex and the Structure and Ordering of Water around It. *J. Chem. Phys.* **2011**, *135*, 245104.
- (40) Sinha, S. K.; Bandyopadhyay, S. Dynamical Properties of Water around a Protein–DNA Complex from Molecular Dynamics Simulation. *J. Chem. Phys.* **2011**, *135*, 135101.
- (41) Carra, C.; Cucinotta, F. A. Accurate Prediction of the Binding Free Energy and Analysis of the Mechanism of the Interaction of Replication Protein A (RPA) with ssDNA. *J. Mol. Model.* **2012**, *18*, 2761–2783.

- (42) Liu, J.; He, L.; Collins, I.; Ge, H.; Libutti, D.; Li, J.; Egly, J. C.; Levens, D. The FBP Interacting Repressor Targets TFIIH to Inhibit Activated Transcription. *Mol. Cell* **2000**, *5*, 331–341.
- (43) He, L.; Liu, J.; Collins, I.; Sanford, S.; O'Connell, B.; Levens, D. Loss of FBP Function Arrests Cellular Proliferation and Extinguishes c-myc Expression. *EMBO J.* **2000**, *19*, 1034–1044.
- (44) Braddock, D. T.; Louis, J. M.; Baber, J. L.; Levens, D.; Clore, G. M. Structure and Dynamics of KH Domains from FBP Bound to Single-Stranded DNA. *Nature* **2002**, *415*, 1051–1056.
- (45) Liu, J.; Akoulitchev, S.; Weber, A.; Ge, H.; Chuikov, S.; Libutti, D.; Wang, X. W.; Conaway, J. W.; Harris, C. C.; Conaway, R.; Reinberg, C.; Levens, D. D. Defective Interplay of Activators and Repressors with TFIIH in Xeroderma Pigmentosum. *Cell* **2001**, *104*, 353–363.
- (46) Michelotti, G. A.; Michelotti, E. F.; Pullner, A.; Duncan, R. C.; Eick, D.; Levens, D. Multiple Single-Stranded Cis Elements are Associated with Activated Chromatin of the Human c-myc Gene in Vivo. *Mol. Cell. Biol.* **1996**, *16*, 2656–2669.
- (47) Kim, M. J.; Park, B.; Kang, Y.; Kim, H. J.; Park, J.; Kang, J. W.; Lee, S. W.; Han, J. M.; Lee, H.; Kim, S. Downregulation of FUSE-binding Protein and c-myc by tRNA Synthetase Cofactor p38 is Required for Lung Cell Differentiation. *Nat. Genet.* **2003**, *34*, 330–336.
- (48) Musco, G.; Stier, G.; Joseph, C.; Castiglione Morelli, M. A.; Nilges, M.; Gibson, T. J.; Pastore, A. Three-Dimensional Structure and Stability of the KH Domain: Molecular Insights into the Fragile X Syndrome. *Cell* **1996**, *85*, 237–245.
- (49) Valverde, R.; Edwards, L.; Regan, L. Structure and Function of KH Domains. *FEBS* **2008**, *275*, 2712–2726.
- (50) Braddock, D. T.; Baber, J. L.; Levens, D.; Clore, G. M. Molecular Basis of Sequence-Specific Single-Stranded DNA Recognition by KH Domains: Solution Structure of a Complex Between hnRNP K KH3 and Single-Stranded DNA. *EMBO J.* **2002**, *21*, 3476–3485.
- (51) Philips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorsid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- (52) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant Pressure Molecular Dynamics Simulation: The Langevin Piston Method. *J. Chem. Phys.* **1995**, *103*, 4613–4621.
- (53) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon: Oxford, 1987.
- (54) 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.
- (55) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (56) Foloppe, N.; MacKerell, A. D., Jr. All-Atom Empirical Force Field for Nucleic Acids: I. Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data. *J. Comput. Chem.* **2000**, *21*, 86–104.
- (57) MacKerell, A. D., Jr.; Banavali, N. All-Atom Empirical Force Field for Nucleic Acids: II. Application to Molecular Dynamics Simulations of DNA and RNA in Solution. *J. Comput. Chem.* **2000**, *21*, 105–120.
- (58) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (59) Rivetti, C.; Walker, C.; Bustamante, C. Polymer Chain Statistics and Conformational Analysis of DNA Molecules with Bends or Sections of Different Flexibility. *J. Mol. Biol.* **1998**, *280*, 41–59.
- (60) Mills, J. B.; Vacano, E.; Hagerman, P. J. Flexibility of Single-Stranded DNA: Use of Gapped Duplex Helices to Determine the Persistence Lengths of Poly(dT) and Poly(dA). *J. Mol. Biol.* **1999**, *285*, 245–257.
- (61) Chakraborty, K.; Mantha, S.; Bandyopadhyay, S. Molecular Dynamics Simulation of a Single-Stranded DNA with Heterogeneous Distribution of Nucleobases in Aqueous Medium. *J. Chem. Phys.* **139**, 075103.
- (62) Swaminathan, S.; Harte, W. E., Jr.; Beveridge, D. L. Investigation of Domain Structure in Proteins via Molecular Dynamics Simulation: Application to HIV-1 Protease Dimer. *J. Am. Chem. Soc.* **1991**, *113*, 2717–2721.
- (63) Luo, J.; Bruice, T. Ten-nanosecond Molecular Dynamics Simulation of the Motions of the Horse Liver Alcohol Dehydrogenase PhCH<sub>2</sub>O<sup>−</sup> Complex. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 16597–16600.
- (64) Ghosh, A.; Vishveshwara, S. A study of Communication Pathways in Methionyl-tRNA Synthetase by Molecular Dynamics Simulations and Structure Network Analysis. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 15711–15716.
- (65) Ernst, R. R.; Bodenhausen, G.; Wokaun, A. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*; Oxford University Press: New York, 2004.
- (66) Villa, A.; Stock, G. What NMR Relaxation Can Tell Us about the Internal Motion of an RNA Hairpin?: A Molecular Dynamics Simulation Study. *J. Chem. Theory Comput.* **2006**, *2*, 1228–1236.
- (67) Wong, V.; Case, D. A. Evaluating Rotational Diffusion from Protein MD Simulations. *J. Phys. Chem. B* **2008**, *112*, 6013–6024.
- (68) Lipari, G.; Szabo, A. Model-free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. *J. Am. Chem. Soc.* **1982**, *104*, 4546–4559.
- (69) Yamane, T.; Okamura, H.; Nishimura, Y.; Kidera, A.; Ikeguchi, M. Side-Chain Conformational Changes of Transcription Factor PhoB upon DNA Binding: A Population-Shift Mechanism. *J. Am. Chem. Soc.* **2010**, *132*, 12653–12659.
- (70) Anderson, K. M.; Esadze, A.; Manoharan, M.; Brueschweiler, R.; Gorenstein, D. G.; Iwahara, J. Direct Observation of the Ion-Pair Dynamics at a Protein–DNA Interface by NMR Spectroscopy. *J. Am. Chem. Soc.* **2013**, *135*, 3613–3619.
- (71) Zhang, Z.; Xu, Z.; Yang, Z.; Liu, Y.; Wang, J.; Shao, Q.; Li, S.; Lu, Y.; Zhu, W. The Stabilization Effect of Dielectric Constant and Acidic Amino Acids on Arginine–Arginine (Arg-Arg) Pairings: Database Survey and Computational Studies. *J. Phys. Chem. B* **2013**, *117*, 4827–4835.