BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

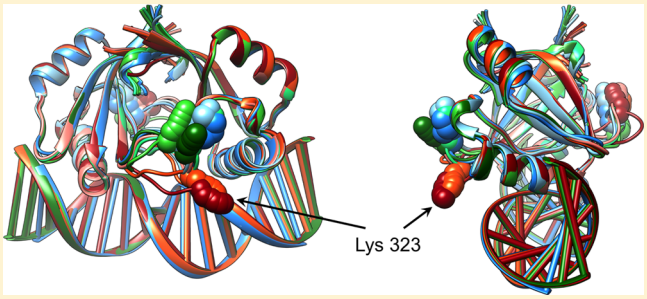
pubs.acs.org/biochemistry

Importance of $\beta_2-\beta_3$ Loop Motion for the Increased Binding and Decreased Selectivity of the ΔLL Mutant of the Human Papillomavirus Type 6 E2 Protein

Geoffrey M. Gray and Arjan van der Vaart*

Department of Chemistry, University of South Florida, 4202 East Fowler Avenue, CHE 205, Tampa, Florida 33620, United States

ABSTRACT: The binding affinity of the human papillomavirus type 6 E2 protein is strongly mediated by the sequence of the DNA linker region, with high affinity for the AATT linker and low affinity for the CCGG linker. When two terminal leucine residues are removed from the protein, the level of binding to both strands increases, but unequally, resulting in a significant decrease in selectivity for the AATT linker strand. To rationalize this behavior, we performed molecular dynamics simulations of the wild-type and mutant protein in the apo state and bound to DNA with high-affinity AATT and low-affinity CCGG linker strands. While no stable contacts were made between the $\beta_2-\beta_3$ loop and DNA in the wild type, this loop was repositioned in the mutant complexes and formed electrostatic contacts with the DNA backbone. More contacts were formed when the mutant was bound to the CCGG linker strand than to the AATT linker strand, resulting in a more favorable change in interaction energy for the CCGG strand. In addition, significant differences in correlated motions were found, which further explained the differences in binding. The simulations suggest that $\beta_2-\beta_3$ loop motions are responsible for the increased affinity and decreased selectivity of the mutant protein.



Papillomaviruses are double-stranded DNA viruses that infect mammals, and more than 100 human papillomaviruses (HPV) have been identified.[1] Some of these viruses like HPV-6 and HPV-11 cause benign warts, while others like HPV-16, HPV-18, and HPV-31 are associated with cancer and in particular cervical carcinomas.[2,3] A key regulator of the papillomavirus life cycle is the E2 protein, which controls the initiation of transcription of nearly all viral promoters and is required for viral DNA replication.[4] E2 is a homodimer, and each chain consists of three domains: an N-terminal transactivation domain, a hinge region, and a C-terminal DNA-binding domain (DBD). The DBD dimer consists of an eight-stranded antiparallel β-barrel, which binds the major groove of the 5′-ACCgNNNNcGGT-3′ consensus sequence by two α-helices, where the lowercase letters indicate preferred nucleotides and NNNN is the variable spacer sequence.[4,5]

DNA is bent in all E2−DNA complexes, and the minor groove of the spacer is compressed.[4,5] The amount of bending depends on the viral species because of differences in the electrostatic surfaces and the separations between the recognition helices among the various E2 proteins.[6] In HPV E2, no contacts are made between E2 and the spacer; however, binding affinities are strongly modulated by the spacer sequence, with high affinities for A:T rich sequences and low affinities for G:C rich sequences.[7,8] This difference in binding is thought to be due to the ease by which the DNA forms prebent structures that resemble the conformation in the complex.[5,8] The high-affinity AATT spacer sequence spontaneously forms a bent structure with a compressed minor groove; the medium-affinity TTAA sequence is bent and prefers a widened minor groove, and the low-affinity ACGT sequence is unbent and prefers a widened minor groove.[9,10] G:C sequences are less flexible than A:T sequences and do not form prebent structures.[8] For certain sequences, salts affect the stability of the prebent structure,[11,12] while very floppy spacers, like nicked DNA constructs, have weakened affinities because of entropic penalties.[8,9]

Of all characterized E2 proteins, HPV-6 E2 is the most selective for the spacer sequence. For example, HPV-6 E2 prefers AATT 1000-fold over CCGG, while for HPV-16 E2, this preference is only 10-fold.[13] This selectivity is thought to stem from differences in molecular adaptability.[7] While HPV-16 E2 shows substantial rearrangements upon DNA binding, such as a shift in the position of the recognition helices and the ordering of the $\beta_2-\beta_3$ loop, binding of HPV-6 E2 to high-affinity DNA is accompanied by minimal protein deformation. There is no reorientation of the recognition helices and no rearrangement of the $\beta_2-\beta_3$ loop (Figure 1A).[7] In addition, the C-terminal Leu 367 and Leu 368 residues have been shown to be important for the selectivity of HPV-6 E2.[14] These two residues are positioned in the core of the β-barrel (Figure 1A) and are absent in most other E2 proteins.[14,15] A study of the

Received: April 21, 2015
Revised: July 9, 2015
Published: July 14, 2015

DOI: 10.1021/acs.biochem.5b00433
Biochemistry 2015, 54, 4918−4926

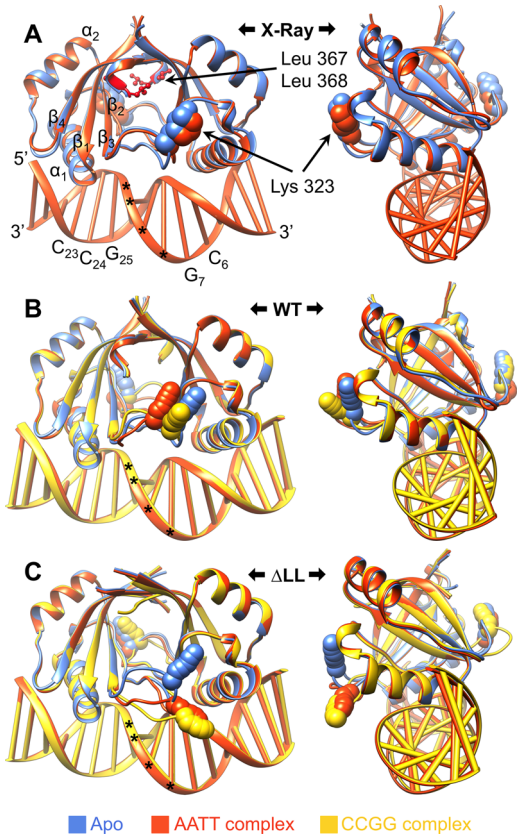


Figure 1. Structure of HPV-6 E2 protein. The apoprotein is colored blue, the AATT complex orange, and the CCGG complex yellow. The position of the linker is indicated by stars, and Lys 323 is shown by spheres. All structures were overlaid with respect to the X-ray structure of the wild-type−AATT complex; both front and side views are shown. (A) X-ray crystal structures of the wild-type protein (Protein Data Bank entries 1R8H[13] and 2AYB[7]). Leu 367 and Leu 368 of the AATT complex are colored red. (B) Averaged simulation structure of the wild-type protein. (C) Averaged simulation structure of the ΔLL mutant.

HPV-6 ΔLL DBD, a mutant in which Leu 367 and Leu 368 were removed, showed that sequence selectivity was significantly diminished in the mutant.[14] The level of binding to both the high-affinity AATT spacer sequence and the nonspecific CCGG spacer was increased for the ΔLL mutant, but because the level of binding to the nonspecific spacer was increased by a larger factor, the specificity for the high-affinity sequence decreased 6-fold. In addition, there was a significant increase in the rate of the first step of complex formation for the ΔLL mutant.[14]

While no high-resolution structures are available for the ΔLL mutant in the apo or DNA-bound states, it was hypothesized that the behavior of the ΔLL mutant is due to a change in protein motion.[14] Nuclear magnetic resonance (NMR) relaxation and hydrogen exchange studies of the apo state showed a small increase in the flexibility of the hydrophobic core and in the loop regions connecting the recognition helices. This was thought to improve protein adaptability,[14] but because no NMR data are available for the ΔLL−DNA complex, information about the motion, flexibility, and specific structural changes of the mutant complex is lacking.

To further rationalize the behavior of the HPV-6 E2 ΔLL mutant, we performed molecular dynamics (MD) simulations of the wild type (WT) and ΔLL mutant in the apo states, and in complex with DNA with high-affinity AATT and low-affinity CCGG spacers. Our simulations indicate that the increase in binding affinities and decrease in selectivity stem from the rearrangement of the $\beta_2-\beta_3$ loop, which forms transient contacts with DNA in the mutant complexes.

METHODS

Initial structures for the wild-type apoprotein and high-affinity DNA-bound complex were taken from the Protein Data Bank (entries 1R8H[13] and 2AYB,[7] respectively). To prevent excessive fraying, DNA in the complex was extended by an extra G:C base pair on each end, resulting in the 5′-GCAACCGAATTCGGTTGC-3′ sequence (Figure 2). In this sequence, the high-affinity AATT linker is underlined; we will refer to this DNA strand as "AATT". Because no experimental structure was available, modeling was used to construct the initial structure of the low-affinity CCGG linker complex. Only the linker base pairs were modified; coordinates of the protein, DNA backbone, and all other base pairs were taken from the 2AYB crystal structure. The CCGG base pairs were built with 3DNA,[16] keeping all step parameters identical to the step parameters observed for the corresponding AATT base pairs in the crystal structure. Because step parameters describe the overall orientation of two adjacent bases in terms of rigid body rotations and translations,[17] this procedure produced an initial configuration of the CCGG linker that was similar to the configuration of the AATT linker in the 2AYB crystal structure. In the following, we will refer to the CCGG linker DNA strand as "CCGG". Initial structures of the apo and complex states of the ΔLL mutant were obtained by deleting the terminal Leu residues. Starting structures of bare DNA were built in the unbent B-form using 3DNA.[16] Using Ambertools13,[18] systems were solvated in a TIP3P octahedral waterbox,[19] with a KCl concentration of 150 mM. After energy minimizations, the *NPT* ensemble was simulated using Langevin dynamics with a collision frequency of 5 ps^{-1} for temperature control[20] and weak coupling with isotropic scaling and a relaxation time of 5 ps for pressure control.[21] To avoid synchronization artifacts, the Langevin bath was initialized with a different random seed at each restart.[22,23] Bonds involving hydrogen atoms were constrained using the SHAKE algorithm,[24] allowing for a 2 fs time step, and the particle mesh Ewald was used to calculate long-range electrostatics.[25] Systems were heated from 120 to 300 K over 1 ns with 1 kcal mol^{-1} $Å^{-2}$ restraint on all protein and DNA heavy atoms, followed by restrained equilibrations over 2 ns, in which the force constant was gradually reduced from 1.0 to 0.5 to 0.25 to 0.1 kcal mol^{-1} $Å^{-2}$. Systems were subsequently equilibrated for 31 ns without restraints, followed by MD production runs of 190 ns; reported properties were calculated from the production runs. All MD simulations were performed using the Amber force field[26−28] and the Amber12 GPU code.[29] Coordinates were written every 10 ps; Stride[30] was used to calculate protein secondary structure, and DNA geometry was calculated using 3DNA[31] and MADBEND.[32] Visual analyses were performed with VMD[33] and Chimera.[34]

$$5'\text{-}G_1\ C_2\ A_3\ A_4\ C_5\ C_6\ G_7\ N_8\ N_9\ N_{10}N_{11}C_{12}G_{13}G_{14}T_{15}T_{16}G_{17}C_{18}\text{-}3'$$
$$3'\text{-}C_{36}G_{35}T_{34}T_{33}G_{32}G_{31}C_{30}N_{29}N_{28}N_{27}N_{26}G_{25}C_{24}C_{23}A_{22}A_{21}C_{20}G_{19}\text{-}5'$$

Figure 2. Simulated DNA sequence. The sequence of the NNNN spacer region is 5′-AATT-3′ for the high-affinity linker and 5′-CCGG-3′ for the low-affinity linker.

DOI: 10.1021/acs.biochem.5b00433
Biochemistry 2015, 54, 4918−4926

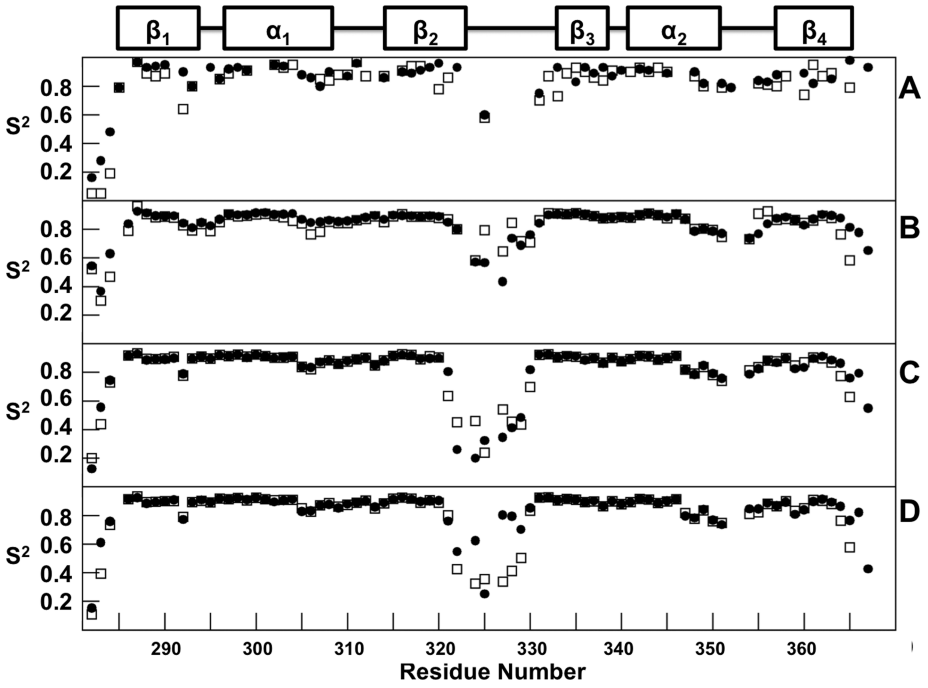


Figure 3. Protein S^2 order parameters. Values are averaged over both chains, with data for the wild type (●) and ΔLL mutant (□). (A) Apo state, experimental values.[14] (B) Apo state, simulation values. (C) AATT complex, simulation values. (D) CCGG complex, simulation values.

Table 1. Rmsd Values (Angstroms) of the Protein Backbone and DNA Phosphorus Atoms with Respect to the Averaged Simulation Structure, the Averaged Simulation Structure of the Wild-Type−AATT Complex, the X-ray Wild-Type−AATT Complex Structure,[7] and the Averaged Structure of Bare DNA or Apo State Protein (unbound) Simulations

	reference	WT−AATT	WT−CCGG	mutant−AATT	mutant−CCGG
DNA	averaged sim.	1.04 ± 0.16	1.03 ± 0.16	0.96 ± 0.16	1.02 ± 0.20
	WT−AATT sim.	1.04 ± 0.16	1.20 ± 0.17	1.01 ± 0.14	1.21 ± 0.18
	X-ray WT−AATT	1.29 ± 0.15	1.39 ± 0.16	1.31 ± 0.16	1.45 ± 0.17
	average unbound sim.	5.11 ± 0.16	7.87 ± 0.19	5.17 ± 0.15	7.89 ± 0.18
protein	averaged sim.	1.07 ± 0.14	0.74 ± 0.10	0.94 ± 0.21	0.95 ± 0.22
	WT−AATT sim.	1.07 ± 0.14	1.16 ± 0.08	1.72 ± 0.16	1.88 ± 0.23
	X-ray WT−AATT	1.57 ± 0.31	1.06 ± 0.10	1.83 ± 0.24	2.01 ± 0.23
	average unbound sim.	1.43 ± 0.34	0.94 ± 0.09	1.40 ± 0.23	1.87 ± 0.21

NMR order parameters were calculated using the IRED approach[35] as implemented in CPPTRAJ.[36] All other analyses were also performed with CPPTRAJ. Hydrogen bonds were calculated using a distance cutoff between heavy atoms of 3.0 Å and an angular cutoff of 45°. Contact analysis was performed between the DNA heavy atoms and protein heavy side chain atoms using a cutoff of 6.0 Å. The mass-weighted Pearson correlation matrices and quasiharmonic modes were calculated using standard methods.[37,38] The free energy cost of DNA bending was estimated using Olson's empirical model, which describes the DNA deformation energy in terms of step parameters.[39] In this model, the energy of each dimer step is given by $E = {}^1/_2\sum_{i=1}^{6}\sum_{j=1}^{6} f_{ij}\Delta\theta_i\Delta\theta_j$, where the i and j indices represent the six step parameters (roll, twist, tilt, shift, slide, and rise), $\Delta\theta_i$ is the deviation of step parameter i from its equilibrium value, and f_{ij} terms are force constants; the energies of all dimer steps are summed to give the total DNA deformation energy. The equilibrium values of the step parameters and the force constants were obtained from statistical analyses of X-ray structures; we used the standard, published values of these parameters for our analysis.[39]

RESULTS

NMR order parameters (S^2) were calculated from the apo state trajectories (Figure 3B) and compared to experimental values where available (Figure 3A).[14] In the NMR experiments, not all order parameters were obtained; for example, S^2 values for a large portion of the β_2−β_3 loop were unresolved, and others were missing, as well.[14] Trends were generally well reproduced in the simulations, with high S^2 values (>0.85), indicating high rigidity, for most of the protein, and low S^2 values, indicating flexibility, for the termini and the β_2−β_3 loop region. The simulation overestimated the rigidity of the N-terminus by ∼0.2 and underestimated the rigidity of the C-terminus by ∼0.2; this corresponded to a root-mean-square deviation of ∼0.5 Å. The flexibility of Gly 291 at the end of β_1 was underestimated by ∼0.2 for the mutant protein. Other discrepancies were the flexibility of Ala 320 that was underestimated by ∼0.1 in the wild type, the flexibility of His 326 that was overestimated by ∼0.2 in the mutant, and the flexibility of β_3 that was underestimated by <0.1 in the mutant. In the mutant, the flexibilities of Arg 355 and His 356 were overestimated by ∼0.1, and the flexibility of β_4 was overestimated for some residues and underestimated for others by <0.05. Despite these small

DOI: 10.1021/acs.biochem.5b00433
Biochemistry 2015, 54, 4918−4926

Table 2. Hydrogen Bonding between Protein and DNA as a Fraction of the Simulation Time, Divided into Specific Contacts with DNA Bases and Nonspecific Contacts with the Phosphate Backbone[a]

DNA	protein	WT−AATT	WT−CCGG	mutant−AATT	mutant−CCGG
		Specific			
Ade 3	Arg 302B	–	0.07	0.28	–
Ade 4	Asn 294B	0.35	0.36	0.40	0.33
Gua 13	Lys 297A	0.61	0.70	0.64	0.65
Gua 14	Lys 297A	–	0.18	–	–
Ade 21	Arg 302A	0.11	0.28	0.08	0.26
Ade 22	Asn 294A	0.39	0.41	0.35	0.34
Gua 31	Lys 297B	0.57	0.66	0.67	0.68
		Nonspecific			
Cyt 2	Arg 302B	1.00	1.00	0.93	1.00
	Lys 349B	0.26	0.17	–	0.21
Ade 3	Arg 302B	0.23	0.37	0.17	–
	Thr 353B	0.97	1.0	0.99	1.00
Ade 4	Asn 294B	0.06	0.06	0.08	–
Cyt 5	Asn 294B	0.78	0.77	0.77	0.79
Thy/Gua 11	Ser 293A	0.91	0.85	0.92	0.76
	His 318A	0.07	–	–	0.07
	Lys 327A	–	–	0.12	–
Cyt 12	Arg 300A	0.58	1.00	0.53	0.55
	Thr 316A	0.92	0.85	0.93	0.91
Gua 13	Arg 300A	1.00	1.00	1.00	1.00
Gua 14	Tyr 301A	0.12	–	–	–
Cyt 20	Arg 302A	0.87	0.92	1.00	0.87
	Lys 349A	–	–	0.19	–
Ade 21	Arg 302A	0.52	–	0.35	0.21
	Thr 353A	0.96	1.00	0.95	0.88
Ade 22	Asn 294A	0.11	–	–	–
Cyt 23	Asn 294A	0.73	0.80	0.77	0.77
Thy/Gua 29	Ser 293B	0.94	0.83	0.91	0.81
Cyt 30	Arg 300B	0.63	0.66	0.57	0.65
	Thr 316B	0.92	0.85	0.93	0.86
Gua 31	Arg 300B	1.00	1.00	1.00	1.00
Gua 32	Lys 323A	–	–	0.43	0.30

[a]A dash indicates hydrogen bonding occurred <1% of the time.

differences, the overall dynamics of the apo states were captured well by MD.

Panels C and D of Figure 3 show the order parameters for the DNA-bound complexes calculated from the MD trajectories; these have not been measured experimentally. While the S^2 parameters for the apo and complex states were similar for most regions, large differences in order parameters were observed for the β_2–β_3 loop and the termini. Compared to the apo states (Figure 3B), the β_2–β_3 loop was significantly more flexible in the wild-type and mutant complexes with high-affinity DNA (Figure 3C) and significantly more flexible in the mutant−CCGG complex (Figure 3D). The flexibility of the β_2–β_3 loop in the wild-type−CCGG complex was similar to that of the wild-type apo state, except for the motion of Lys 323, which was more flexible in the CCGG complex. The N-terminus was more flexible in the complex simulations than in the apo simulations, while the C-terminus of the mutant complexes showed increased flexibility adjacent to the deletion site (Met 365 and His 366).

Root-mean-square deviations (rmsds) for the DNA phosphorus atoms are listed in Table 1 for all systems. These rmsd values were calculated with respect to four reference states: the averaged structure of the simulated system, the averaged simulation structure of the wild-type−AATT complex, the X-ray wild-type−AATT complex structure,[7] and the averaged structure of bare DNA or apo state protein as obtained from separate simulations. The DNA rmsd values with respect to the averaged structure, the simulated wild-type−AATT complex structure, and the crystal structure are nearly identical for all systems, indicating that the structure and dynamical behavior of DNA were similar in all complexes. The rmsds with respect to bare DNA were similar for the wild-type− and mutant−AATT complexes and similar for the CCGG complexes. However, these rmsd values were higher in the CCGG than in the AATT complexes. This difference indicates that CCGG DNA undergoes larger structural changes than AATT DNA upon binding. Structural analyses showed the structural differences were mainly due to global tilt,[32] which was more similar to the protein−DNA complex values for bare AATT than for bare CCGG. Because DNA deformations are costly in terms of energy,[39,40] less energy will be spent to bend DNA in the AATT complex than in the CCGG complex. In fact, using Olson's model to estimate the energy cost of DNA bending,[39] relative to the wild-type−AATT complex, DNA was deformed by 1.2 kcal/mol in the mutant−AATT complex, by 7.6 kcal/mol in the wild-type−CCGG complex, and by 9.8 kcal/mol in the mutant−CCGG complex. This observation is in agreement with the hypothesis that AATT DNA can form prebent structures that resemble the conformation in the complex, while CCGG DNA cannot,[5,8] and helps explain the difference in affinity for the AATT and CCGG linkers.

Backbone rmsd values for the protein are also listed in Table 1. The rmsd is shown with respect to the same four reference states, except that the last reference now refers to the averaged

DOI: 10.1021/acs.biochem.5b00433
Biochemistry 2015, 54, 4918−4926

structure of the apo state as obtained in separate simulations. The rmsd values with respect to the averaged structures were low for all systems (≤1.1 Å), which indicates that the proteins are fairly rigid after equilibration. The largest rmsd values with respect to the crystal and wild-type−AATT complex simulation were observed in the mutant protein, with the rmsd in the CCGG complex being larger than the rmsd in the AATT complex. This indicates that the mutant protein had more significant structural changes than the wild type upon binding DNA, especially for the mutant−CCGG complex. The smallest rmsd with respect to the averaged simulation structure and the crystal structure was observed for the wild-type−CCGG complex, indicating that this complex had the fewest structural changes.

Figure 1 shows overlays of the averaged simulation structures and wild-type apo[13] and wild-type−AATT complex[7] crystal structures. Consistent with the low rmsds, no significant structural differences existed between the DNA of the systems. The protein structure was also very similar in all systems, except for the conformation of the β_2−β_3 loops. In the crystal structures of the wild-type apo and AATT complex, the β_2−β_3 loops are ordered and point toward the β-barrel, away from the DNA. In these structures, the side chain of Lys 323 points toward the α_1−β_2 loop, away from the DNA. While these conformations of Lys 323 and the β_2−β_3 loops were largely retained in the averaged wild-type complexes and wild-type and mutant apo proteins, the β_2−β_3 loop changed position in the mutant complexes and moved toward the DNA. Lys 323 of the loop changed its relative orientation, pointing toward the DNA and making transient contacts with the phosphates of Gua 31 and Gua 32. The loop motion and transient contacts were observed for only chain A of the mutant protein; in chain B, the β_2−β_3 loop was also displaced, but to a lesser extent (Figure 1C).

Hydrogen bonds between the protein and DNA are listed in Table 2. These are split up into specific hydrogen bonds, which are between a base and protein, and nonspecific hydrogen bonds between a phosphate and the protein. All specific and nonspecific hydrogen bonds present in the wild-type−AATT crystal structure were also prevalent in the simulation structures, with the exception of a specific contact involving Asn 294. In the wild-type−AATT crystal structure, this residue hydrogen bonds with the Cyt 5/Cyt 23 base. While the C^{α} and C^{β} positions were very similar to those observed in the crystal structure, in the simulations the amide group was rotated and formed a persistent contact with neighboring base Ade 4/Ade 22 instead. Like Cyt 5/Cyt 23, these bases are part of the cognate sequence. With a few exceptions, the hydrogen bonds that were present in the crystal and in the simulations had similar occurrences in all simulations.

Transient hydrogen bonds not present in the wild-type−AATT crystal structure were observed, as well. The most prominent of these were nonspecific hydrogen bonds between the β_2−β_3 loop and the DNA, which were observed only in the mutant (Figure 4). Lys 323A formed a nonspecific hydrogen bond with Gua 32. This hydrogen bond was observed 43% of the time in the mutant−AATT complex and 30% of the time in the mutant−CCGG complex. In addition, Lys 327A formed a nonspecific hydrogen bond in the mutant−AATT complex 12% of the time. Other transient hydrogen bonds involved Lys 349 of the α_2−β_4 loop, which formed a nonspecific hydrogen bond to Thy 11 for 12% of the time in the mutant−AATT complex and a nonspecific hydrogen bond to Cyt 2 for 17−26% of the

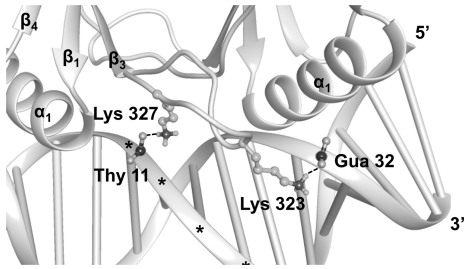


Figure 4. Hydrogen bonds among Lys 323, Lys 327, and DNA for the mutant−AATT complex. Stars indicate the position of the linker. Hydrogen bonds are shown as dashed lines.

time in all other complexes; His 318A of β_2, which formed a nonspecific hydrogen bond 7% of the time in the wild-type−AATT and mutant−CCGG complexes; and Tyr 301A of α_1, which formed a nonspecific hydrogen bond 12% of the time in the wild-type−AATT complex.

A few transient heavy atom contacts between DNA and β_2 and the β_2−β_3 loop were formed in the wild-type−AATT complex. These were between protein residues 315−322 of both chains and base pairs 10−14 and 28−32. The stable hydrogen bonds between Thr 316 and Cyt 30 observed in the crystal[7] were present, as well (Table 2). Sporadic transient contacts between residue Lys 323B and base pairs 13 and 14 also occurred, and residues His 326B, Lys 327B, and Ala 329B formed transient contacts with Thy 29. Lys 327A formed sporadic transient contacts with Thy 10, Thy 11, Gua 31, and Cyt 12, while Lys 327B formed infrequent transient contacts with Thy 28 and Thy 29. None of these transient contacts were hydrogen bonds (Table 2). Contacts in the wild-type−CCGG complex were formed between residues 315B−322B of β_2 and base pairs 10−14 and 28−32, including the Thr 316 hydrogen bonds observed in the wild-type−AATT crystal[7] (Table 2). The β_2−β_3 loop remained rigid, and no contacts between DNA and this loop were made. Formation of transient contacts with chain A occurred much less frequently in this complex.

Heavy atom contacts between DNA and the β_2−β_3 loop of the mutant are shown in Figure 5 as a function of time; DNA contacts with β_2 and β_3 residues near the loop are shown, as well. For the sake of clarity, contacts that persisted over the entire length of both simulations are omitted from the figure. In addition to these, several other persistent contacts were observed (Figure 5). Of particular interest is Lys 323A, which formed prominent hydrogen bonds with the DNA backbone in the mutant simulations (Table 2 and Figure 4). In the mutant−AATT complex, Lys 323A formed contacts with Cyt 30, Gua 31, and Gua 32 (and transiently with Thy 10), while in the mutant−CCGG complex, it formed contacts with Gua 31 and Gua 32. These contacts were persistent throughout most of the simulations, even in the absence of hydrogen bonding. The β_2−β_3 loop of chain B generally formed contacts with the linker region, whereas chain A formed contacts with the minor groove of the recognition motif adjacent to the linker. Figure 5 shows significant differences in the number of contacts and the contact frequency between the two mutant complexes. The mutant−CCGG complex formed a larger number of different contacts that involved more residues. This complex also formed more transient contacts. The mutant−CCGG complex had an approximately 1:1 ratio of contacts formed by chain A and chain B, while the mutant−AATT complex had a 2:1 ratio.

DOI: 10.1021/acs.biochem.5b00433
Biochemistry 2015, 54, 4918−4926

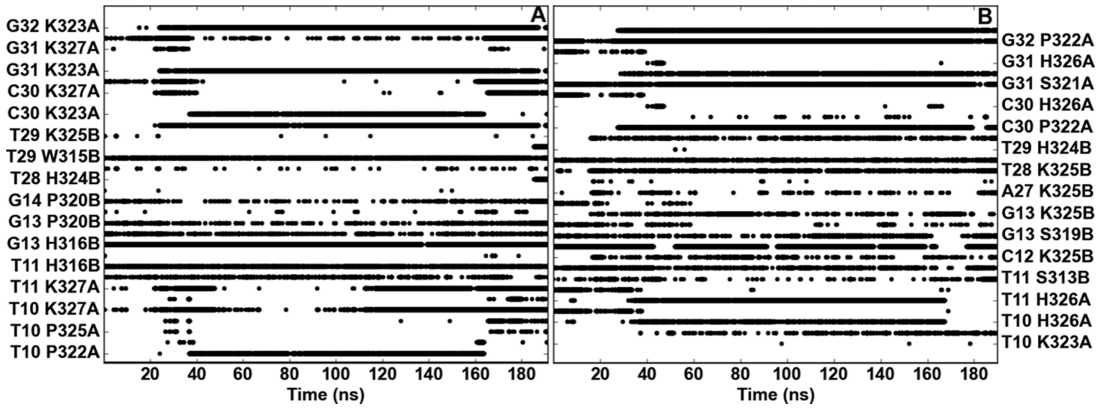


Figure 5. Heavy atom contacts between the β_2–β_3 loops and DNA for the (A) mutant–AATT and (B) mutant–CCGG complex. Contacts maintained throughout the entire simulation in both complexes have been omitted for the sake of clarity. These are between Gua 31 and Pro 322A, Cyt 12 and Val 331A, Cyt 12 and His 318B, Cyt 12 and Ala 320B, Thy 29 and Ser 315B, Thy 29 and Thr 316B, Thy 29 and His 318B, Thy 29 and Ala 329B, Thy 28 and His 318B, Gua 31 and Ser 315B, Gua 31 and Thr 316B, Thy 11 and Ala 329A, Cyt 30 and Ala 320A, Cyt 30 and Ser 315B, Cyt 30 and Thr 316B, Cyt 30 and Val 331B, and Gua 13 and Ala 320B.

The formation of additional protein–DNA contacts led to more favorable protein–DNA interaction energies in the mutant complexes (Table 3). While these are not free energies, the interaction energies show the same trend as the experimental binding affinities, with protein–DNA interactions being stronger for AATT than for CCGG, and stronger interactions for the mutant than for the wild type. Moreover, while standard deviations are large, compared to the wild types a larger increase in the total number of interactions was observed in the mutant–CCGG complex (−57.5 kcal/mol) than in the mutant–AATT complex (−56.5 kcal/mol), indicating a loss of specificity as observed in the experiments. A large portion of this change in interactions was due to β_2–β_3 loop–DNA interactions (Table 3), which were stronger for the mutant than for the wild type. Moreover, compared to the wild-type values, a larger increase in the total number of loop interactions was observed for the mutant–CCGG complex (−89.9 kcal/mol) than for the mutant–AATT complex (−63.0 kcal/mol). Further decomposing the total interaction energy at the residue level, and comparing the differences between wild-type and mutant complexes, showed that the mutant–AATT complex primarily had increases in the number of interactions between DNA and the β_2–β_3 loop of chain A, between the α_2–β_4 loop of chain A and base pairs 2 and 3, and, to a lesser extent, between α_1 of chain A and base pairs 20–22. There was also a notable loss of interaction between the α_2–β_4 loop of chain B and base pairs 2 and 3. The mutant–CCGG system had increases in the number of interactions between DNA and the β_2–β_3 loops of both chains, α_1 of chain A with base pairs 20–22, and the α_2–β_4 loop of chain B with base pairs 2 and 3, while it had a loss of interactions between α_1 of chain B and base pairs 3 and 4. Overall, the mutant had a larger number of total increases in interactions when bound to the CCGG than to AATT. This accounted for the decrease in interaction energy

Table 3. Protein–DNA Interaction Energies (kilocalories per mole)[a]

system	protein–DNA	β_2–β_3 chain A-DNA	β_2–β_3 chain B-DNA
WT–AATT	−1147.8 ± 66.3	−33.6 ± 22.2	−35.0 ± 44.0
WT–CCGG	−1083.4 ± 65.7	−3.0 ± 7.4	−4.9 ± 10.3
mutant–AATT	−1204.3 ± 76.4	−121.1 ± 65.7	−10.5 ± 7.0
mutant–CCGG	−1140.9 ± 70.2	−63.8 ± 35.2	−33.9 ± 19.9

[a]Interactions for the entire protein and the DNA are shown, along with each individual β_2–β_3 loop (on chain A and B) and the DNA.

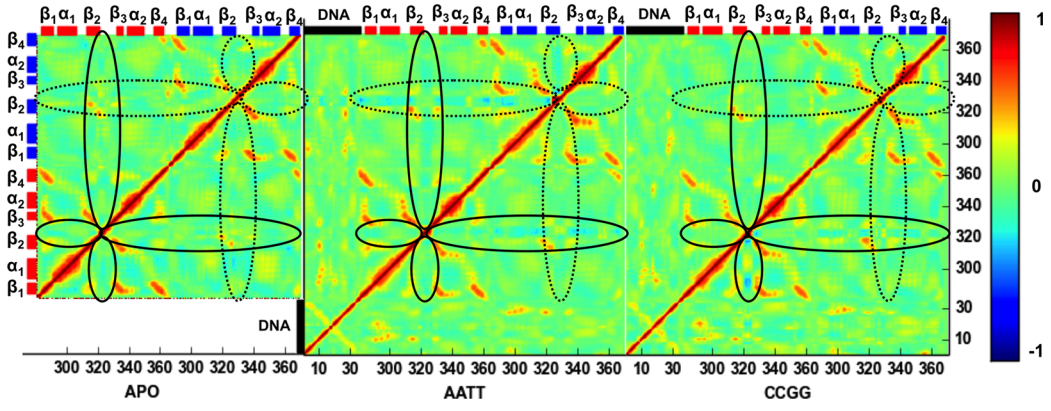


Figure 6. Normalized variance–covariance matrices. Correlations in the wild type are given in the upper triangular parts, while correlations for the mutant are given in the lower triangular parts of the matrices. Secondary structure elements for chain A are colored red and those for chain B blue. Correlations of the β_2–β_3 loop are shown as black ovals, which are continuous for chain A and dotted for chain B.

DOI: 10.1021/acs.biochem.5b00433
Biochemistry 2015, 54, 4918–4926

upon mutation that was greater for the CCGG complex than for the AATT complex.

Normalized variance−covariance matrices of protein fluctuations are shown in Figure 6. Positive values indicate in-phase correlated motions; negative values indicate anticorrelated or out-of-phase correlated motions, and zero values indicate uncorrelated or orthogonal motions. Between the apo states, differences in correlated motions were minimal, with no differences in sign. For both apo state systems, elements of the β-barrel on the same chain were positively correlated. This positive correlation was also present in the DNA-bound complexes. In the AATT complexes, a change in the sign of the correlation occurred in which the β_2−β_3 loops of both chains for the wild type (but especially chain B) and chain A of the mutant complex became negatively correlated with β-barrel elements of either chain (Figure 6, encircled). For the CCGG−mutant complex, the sign change was observed between the β_2−β_3 loop on chain A and β-barrel elements. This negative correlation did not occur in chain B of the mutant complexes, which did not have the motion of the β_2−β_3 loop. No change in the sign of correlation between the β_2−β_3 loops and the β-barrel occurred for the CCGG−wild-type complex. Overall, the variance−covariance analysis showed correlations between the motion of the β_2−β_3 loop and the β-barrel. These correlations were positive in the apo states. Upon binding, certain correlations between the β_2−β_3 loop and β-barrel became negative for all well-bound systems (the wild-type−AATT and the mutant−AATT and −CCGG complexes, with binding constants of 41, 22, and 110 nM, respectively[14]), while it remained positive for the poorly bound system (the wild-type−CCGG complex with a binding constant of 1203 nM[14]).

Quasiharmonic analyses showed that the β_2−β_3 loops of the apo state proteins were effectively rigid and did not move in the lowest-frequency modes (Figure 7). This is in contrast to the DNA-bound complexes, which showed significant motions of the β_2−β_3 loop in the lowest-frequency modes. Motions of this loop were significantly reduced in the wild-type−CCGG complex, however. In the quasiharmonic modes of the mutant complexes, the β_2−β_3 loop moved away from the protein and bent toward the DNA, making contacts. In contrast, in the quasiharmonic modes of the wild-type−AATT complex, the β_2−β_3 loop moved away from the rest of the protein but failed to move toward the DNA. In the mutant−AATT complex, the β_2−β_3 loops moved in an anticorrelated fashion, with one loop moving toward the DNA and the other toward the protein. In the mutant−CCGG complex, the motion between the loops was correlated, with both moving toward the DNA at the same time. While we observed DNA contacts with only β_2−β_3 of chain A of the mutant complexes, the quasiharmonic modes suggest that chain B could form contacts, as well. Moreover, the modes imply that in the mutant−AATT complex, only one β_2−β_3 loop will make DNA contacts at a time, while in the mutant−CCGG complex, one or both loops can contact DNA at a time. This might further explain the loss of selectivity upon mutation, that is, the reason why the mutant had a larger increase in binding affinity for CCGG than for AATT.[14]

DISCUSSION

MD simulations were performed to rationalize the increased binding affinity of the ΔLL mutant of the HPV-6 E2 protein for both high-affinity AATT and low-affinity CCGG linkers and the associated loss of selectivity for the high-affinity linker. The simulations showed that the DNA conformation was nearly

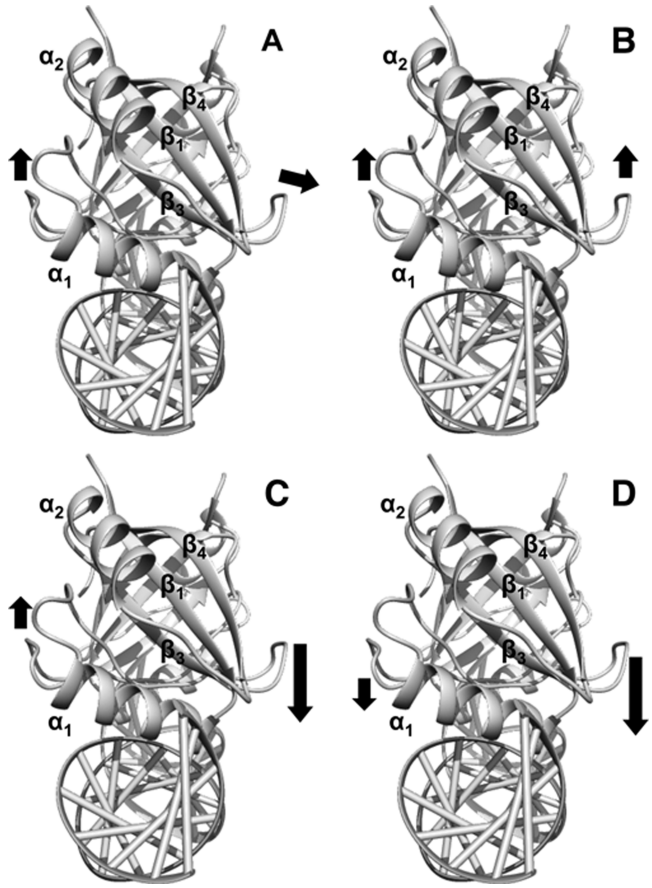


Figure 7. Lowest vibrational quasiharmonic mode for the (A) wild-type−AATT, (B) wild-type−CCGG, (C) mutant−AATT, and (D) mutant−CCGG complexes. Arrows indicate the direction of motion, with the length of the arrows corresponding to the magnitude of motion.

identical in the mutant− and wild-type−AATT and −CCGG complexes. The simulations also indicated that there is greater structural similarity between the bound and unbound AATT strand than between the bound and unbound CCGG strand. This observation supports the hypothesis that AATT is prebent in solution, resulting in greater affinity of AATT than of CCGG. This greater affinity is also due to more favorable complementary contacts between the wild type and AATT, resulting in a more favorable protein−DNA interaction energy (Table 3). Moreover, the simulations showed significant differences in the conformation of the β_2−β_3 loop. This loop was rearranged in the mutant complexes, forming new contacts with the DNA, including a nonspecific hydrogen bond. The number of new contacts was larger in the mutant−CCGG complex than in the mutant−AATT complex, and the associated change in interaction energy was also more favorable in the CCGG complex. These observations suggest that the increased affinities and decreased selectivity of the ΔLL mutant stem from β_2−β_3 loop motion.

Detailed analyses showed that the β_2−β_3 loop motion is hindered in the apo states and also in the wild-type−CCGG complex. Quasiharmonic modes showed limited motion of the β_2−β_3 loop in the wild-type−AATT complex, and in the simulation, a few transient contacts were sporadically formed between the loop and DNA. Nevertheless, these contacts would be an additional contributing factor to the selectivity of the wild

DOI: 10.1021/acs.biochem.5b00433
Biochemistry 2015, 54, 4918−4926

type for AATT. The $\beta_2-\beta_3$ loop fully moved toward the DNA in the mutant complexes, forming stable contacts with the DNA. While in the simulations these contacts were formed by only one protein chain, quasiharmonic modes suggested that the other chain could form contacts, as well. Because the motion of the $\beta_2-\beta_3$ loops was anticorrelated in the mutant−AATT complex, it is likely that only one chain forms contacts at a time in the mutant−AATT complex. In contrast, in the mutant−CCGG complex, the motion between the $\beta_2-\beta_3$ loops was correlated, and both chains could make contact at one time. This difference in motion would further decrease the selectivity of the mutant for the AATT complex.

The site of the ΔLL mutation is near the β-barrel, and in experiments, subtle changes in β-barrel dynamics were measured.[14] Our simulations elucidated an important connection between motions of the β-barrel and motions of the $\beta_2-\beta_3$ loops. In the apo state, motions between the $\beta_2-\beta_3$ loop and the β-barrel were positively correlated. The correlation was also positive in the wild-type−CCGG complex (which binds poorly), but negative for many elements in the wild-type−AATT and mutant−AATT and −CCGG complexes (which bind well). The out-of-phase motions between the β-barrel and the $\beta_2-\beta_3$ loops allowed the loop to move away from the β-barrel in the systems that bind well. These observations suggest that correlated motions between the β-barrel and $\beta_2-\beta_3$ loops are important for selectivity and binding.

AUTHOR INFORMATION

Corresponding Author

*E-mail: avandervaart@usf.edu. Phone: +1-813-974-8762.

Funding

This work was supported by National Science Foundation CAREER Award CHE-1007816.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Computer time was provided by USF Research Computing and XSEDE. We thank Ugwiyeon Lee for exploratory simulations during the initial stage of the project and Dr. Ioannis Gelis, Ning Ma, and Aleksandra Karolak for discussions.

REFERENCES

(1) de Villiers, E. M., Fauquet, C., Broker, T. R., Bernard, H. U., and zur Hausen, H. (2004) Classification of papillomaviruses. *Virology 324*, 17−27.

(2) Bosch, F. X., Lorincz, A., Munoz, N., Meijer, C., and Shah, K. V. (2002) The causal relation between human papillomavirus and cervical cancer. *J. Clin. Pathol. 55*, 244−265.

(3) Schiffman, M., Castle, P. E., Jeronimo, J., Rodriguez, A. C., and Wacholder, S. (2007) Human papillomavirus and cervical cancer. *Lancet 370*, 890−907.

(4) McBride, A. A. (2013) The Papillomavirus E2 proteins. *Virology 445*, 57−79.

(5) Hegde, R. S. (2002) The papillomavirus E2 proteins: Structure, function, and biology. *Annu. Rev. Biophys. Biomol. Struct. 31*, 343−360.

(6) Hegde, R. S., and Androphy, E. J. (1998) Crystal structure of the E2 DNA-binding domain from human papillomavirus type 16: Implications for its DNA binding-site selection mechanism. *J. Mol. Biol. 284*, 1479−1489.

(7) Hooley, E., Fairweather, V., Clarke, A. R., Gaston, K., and Brady, R. L. (2006) The recognition of local DNA conformation by the human papillomavirus type 6 E2 protein. *Nucleic Acids Res. 34*, 3897−3908.

(8) Hines, C. S., Meghoo, C., Shetty, S., Biburger, M., Brenowitz, M., and Hegde, R. S. (1998) DNA structure and flexibility in the sequence-specific binding of papillomavirus E2 proteins. *J. Mol. Biol. 276*, 809−818.

(9) Kim, S. S., Tam, J. K., Wang, A. F., and Hegde, R. S. (2000) The structural basis of DNA target discrimination by papillomavirus E2 proteins. *J. Biol. Chem. 275*, 31245−31254.

(10) Zimmerman, J. M., and Maher, L. J. (2003) Solution measurement of DNA curvature in papillomavirus E2 binding sites. *Nucleic Acids Res. 31*, 5134−5139.

(11) Blakaj, D. M., Kattamuri, C., Khrapunov, S., Hegde, R. S., and Brenowitz, M. (2006) Indirect readout of DNA sequence by papillomavirus E2 proteins depends upon net cation uptake. *J. Mol. Biol. 358*, 224−240.

(12) Xi, Z. Q., Zhang, Y. L., Hegde, R. S., Shakked, Z., and Crothers, D. M. (2010) Anomalous DNA binding by E2 regulatory protein driven by spacer sequence TATA. *Nucleic Acids Res. 38*, 3827−3833.

(13) Dell, G., Wilkinson, K. W., Tranter, R., Parish, J., Brady, L., and Gaston, K. (2003) Comparison of the structure and DNA-binding properties of the E2 proteins from an oncogenic and a non-oncogenic human papillomavirus. *J. Mol. Biol. 334*, 979−991.

(14) Brown, C., Campos-Leon, K., Strickland, M., Williams, C., Fairweather, V., Brady, R. L., Crump, M. P., and Gaston, K. (2011) Protein flexibility directs DNA recognition by the papillomavirus E2 proteins. *Nucleic Acids Res. 39*, 2969−2980.

(15) Sanchez, I. E., Dellarole, M., Gaston, K., and Gay, G. D. (2007) Comprehensive comparison of the interaction of the E2 master regulator with its cognate target DNA sites in 73 human papillomavirus types by sequence statistics. *Nucleic Acids Res. 36*, 756−769.

(16) Lu, X.-J., and Olson, W. K. (2008) 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc. 3*, 1213−1227.

(17) Olson, W. K., Bansal, M., Burley, S. K., Dickerson, R. E., Gerstein, M., Harvey, S. C., Heinemann, U., Lu, X. J., Neidle, S., Shakked, Z., Sklenar, H., Suzuki, M., Tung, C. S., Westhof, E., Wolberger, C., and Berman, H. M. (2001) A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol. 313*, 229−237.

(18) Case, D. A., Darden, T. A., Cheatham, T. E., III, Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Walker, R. C., Zhang, W., Merz, K. M., Roberts, B., Hayik, S., Roitberg, A., Seabra, G., Swails, J., Götz, A. W., Kolossváry, I., Wong, K. F., Paesani, F., Vanicek, J., Wolf, R. M., Liu, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, J., Hsieh, M.-J., Cui, G., Roe, D. R., Mathews, D. H., Seetin, M. G., Salomon-Ferrer, R., Sagui, C., Babin, V., Luchko, T., Gusarov, S., Kovalenko, A., and Kollman, P. A. (2012) *AMBER 13*, University of California, San Francisco.

(19) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys. 79*, 926−935.

(20) Levy, R. M., Karplus, M., and McCammon, J. A. (1979) Diffusive langevin dynamics of model alkanes. *Chem. Phys. Lett. 65*, 4−11.

(21) Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys. 81*, 3684−3690.

(22) Sindhikara, D. J., Kim, S., Voter, A. F., and Roitberg, A. E. (2009) Bad Seeds Sprout Perilous Dynamics: Stochastic Thermostat Induced Trajectory Synchronization in Biomolecules. *J. Chem. Theory Comput. 5*, 1624−1631.

(23) Uberuaga, B. P., Anghel, M., and Voter, A. F. (2004) Synchronization of trajectories in canonical molecular-dynamics simulations: Observation, explanation, and exploitation. *J. Chem. Phys. 120*, 6363−6374.

(24) Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys. 23*, 327−341.

DOI: 10.1021/acs.biochem.5b00433
Biochemistry 2015, 54, 4918−4926

(25) Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. *J. Chem. Phys. 98*, 10089−10092.
(26) Zgarbova, M., Otyepka, M., Sponer, J., Mladek, A., Banas, P., Cheatham, T. E., and Jurecka, P. (2011) Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput. 7*, 2886−2902.
(27) Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham, T. E., III, Laughton, C. A., and Orozco, M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Biophys. J. 92*, 3817−3829.
(28) Cheatham, T. E., Cieplak, P., and Kollman, P. A. (1999) A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn. 16*, 845−862.
(29) Salomon-Ferrer, R., Gotz, A. W., Poole, D., Le Grand, S., and Walker, R. C. (2013) Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput. 9*, 3878−3888.
(30) Frishman, D., and Argos, P. (1995) Knowledge-based protein secondary structure assignment. *Proteins: Struct., Funct., Genet. 23*, 566−579.
(31) Lu, X.-J., and Olson, W. K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res. 31*, 5108−5121.
(32) Strahs, D., and Schlick, T. (2000) A-tract bending: Insights into experimental structures by computational models. *J. Mol. Biol. 301*, 643−663.
(33) Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graphics 14*, 33−38.
(34) Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem. 25*, 1605−1612.
(35) Prompers, J. J., and Bruschweiler, R. (2002) General framework for studying the dynamics of folded and nonfolded proteins by NMR relaxation spectroscopy and MD simulation. *J. Am. Chem. Soc. 124*, 4522−4534.
(36) Roe, D. R., and Cheatham, T. E. (2013) PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput. 9*, 3084−3095.
(37) Ichiye, T., and Karplus, M. (1991) Collective motions in proteins: A covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins: Struct., Funct., Genet. 11*, 205−217.
(38) Teeter, M. M., and Case, D. A. (1990) Harmonic and quasiharmonic descriptions of crambin. *J. Phys. Chem. 94*, 8091−8097.
(39) Olson, W. K., Gorin, A. A., Lu, X.-J., Hock, L. M., and Zhurkin, V. B. (1998) DNA Sequence-Dependent Deformability Deduced from Protein-DNA Crystal Complexes. *Proc. Natl. Acad. Sci. U. S. A. 95*, 11163−11168.
(40) van der Vaart, A. (2015) Coupled binding-bending-folding: The complex conformational dynamics of protein-DNA binding studied by atomistic molecular dynamics simulations. *Biochim. Biophys. Acta, Gen. Subj. 1850*, 1091−1098.

DOI: 10.1021/acs.biochem.5b00433
Biochemistry 2015, 54, 4918−4926