JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

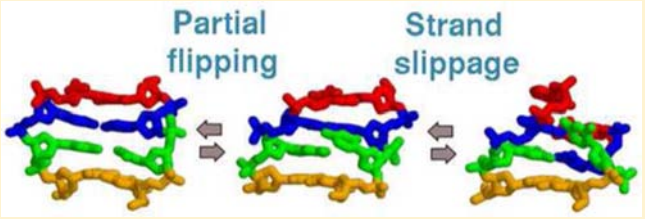
Partial Base Flipping Is Sufficient for Strand Slippage near DNA Duplex Termini

Nilesh K. Banavali*

Laboratory of Computational and Structural Biology, Division of Genetics, Wadsworth Center, New York State Department of Health, and Department of Biomedical Sciences, School of Public Health, State University of New York at Albany, CMS 2008, 150 New Scotland Avenue, Albany, New York 12208, United States

Supporting Information

ABSTRACT: Strand slippage is a structural mechanism by which insertion–deletion (indel) mutations are introduced during replication by polymerases. Three-dimensional atomic-resolution structural pathways are still not known for the decades-old template slippage description. The dynamic nature of the process and the higher energy intermediates involved increase the difficulty of studying these processes experimentally. In the present study, restrained and unrestrained molecular dynamics simulations, carried out using multiple nucleic acid force fields, are used to demonstrate that partial base-flipping can be sufficient for strand slippage at DNA duplex termini. Such strand slippage can occur in either strand, i.e. near either the 3′ or the 5′ terminus of a DNA strand, which suggests that similar structural flipping mechanisms can cause both primer and template slippage. In the repetitive mutation hot-spot sequence studied, non-canonical base-pairing with exposed DNA groove atoms of a neighboring G:C base-pair stabilizes a partially flipped state of the cytosine. For its base-pair partner guanine, a similar partially flipped metastable intermediate was not detected, and the propensity for sustained slippage was also found to be lower. This illustrates that a relatively small metastable DNA structural distortion in polymerase active sites could allow single base insertion or deletion mutations to occur, and stringent DNA groove molecular recognition may be required to maintain intrinsic DNA polymerase fidelity. The implications of a close relationship between base-pair dissociation, base unstacking, and strand slippage are discussed in the context of sequence dependence of indel mutations.



INTRODUCTION

Errors in DNA replication by DNA polymerases, if not subsequently corrected, are the basis for genetic mutations. Single base insertion–deletion (indel) errors, where one base is either inserted or deleted in the synthesized DNA strand as compared to the complementary template strand, can be particularly disruptive since they can alter the subsequent genetic code due to frameshifts.[1] Although different DNA polymerases vary widely in their fidelity, indel mutations occur in all template-dependent DNA polymerases *in vitro*.[2] There are a few proposed mechanisms for how such indel mutations can occur: DNA strand slippage,[3] misinsertion–misalignment,[4] melting–misalignment,[5] and dNTP-stabilized misalignment.[6,7] The exact structural and dynamic details by which such proposed mechanisms occur are not yet completely understood since the intermediate states involved are transient higher energy states that are difficult to characterize experimentally.

Experimental studies using X-ray crystallography have provided informative snapshots into the structural mechanisms that are involved in indel mutations. A mechanism for single base deletion due to misalignment of the template and synthesized strands was observed in the active site of an error-prone Y-family polymerase Dpo4, where the correct template base was mispositioned by unstacking, which allowed its neighboring base to assume the template role.[8] Another deletion-causing misaligned intermediate, for which Watson–Crick base-pairing at the primer terminus was maintained to yield a catalytically competent structural state, was observed in human DNA polymerase λ.[9] Structures of a Y-family polymerase Dbh showed a single-base deletion accompanied by an extrahelical template base three base positions removed from the active site in a repetitive "hot-spot" sequence.[10] In contrast, a possible single base insertion mutation caused by unstacking of a primer base, causing a +1 frameshift in the genetic code, was observed in the Y-family polymerase Dpo4 for templates containing abasic[11] or O6-benzyl-dG lesions.[12] These structural studies suggest that a primary determinant of indel mutations might be base unstacking in either the newly synthesized or the template DNA strands.

Fluorescence and NMR-based studies have also been used to characterize the dynamic nature of such processes. For the Y-family polymerase Dbh, it was shown using 2-aminopurine fluorescence that the base-pair separation at the primer terminus can be accompanied by slipping of template base to pair with the free terminal primer base, resulting in a single-base

Received: February 12, 2013
Published: May 3, 2013

dx.doi.org/10.1021/ja401573j | J. Am. Chem. Soc. 2013, 135, 8274–8282

deletion mutation.[13] A bulge-containing λ C_I frameshift mutational hot-spot sequence studied by NMR for imino proton lifetimes and helix stability showed that bulges could have both localized or delocalized effects, suggesting that the bulge can migrate along the duplex.[14] Examination of the differential exchange rates of imino protons in bulge-containing duplexes as compared to regular duplexes also indicated that the unstacked base might prefer prefers positions internal to the duplex over those at the termini.[15] A *N*2-(3-oxo-1-propenyl)dG lesion in a frameshift hotspot *Salmonella typhimurium* sequence containing a two-base complementary strand deletion also showed rapid bulge migration.[16,17] A series of recent NMR investigations have also systematically characterized the effect of base identity on strand slippage propensity in DNA hairpin primer−template models.[18−22] Propensity of strand slippage was found to depend on the template base and the incoming base, as well as the bases 5′ and 3′ to the template base. These studies highlight the dynamic nature of the base unstacking and strand slippage processes, and the likely effects of local sequence on indel mutations during genetic replication.

While computational methods have been used extensively to study dynamics of nucleic acids[23] and their interaction with proteins,[24] there are only a few such studies addressing strand slippage. Early explicit solvent molecular dynamics (MD) investigations of large-scale DNA end-to-end extension showed evidence of strand slippage and base-pair separation in response to pulling forces.[25−27] Nanosecond scale MD simulations of DNA polymerase β revealed the role of specific polymerase active site residues in stabilizing DNA mismatches at the primer terminus.[28] Template strand structural changes induced by dNTP binding during conformational rearrangements in the catalytic cycle were linked to the rate of strand slippage using crystal structures and MD simulations of mutant derivatives of DNA polymerase λ bound to a primer template.[29] Microsecond scale MD simulations of a κB sequence showed a longitudinally sheared base-pair accompanied by cross-strand intercalative base stacking and spontaneous base-pair separation within a DNA double helix.[30] Free energy analyses based on MD simulations also suggest that pol λ stabilizes misaligned DNA better than aligned DNA.[31] MD simulations have also been used to probe the mechanisms by which a dATP can be incorporated opposite a bulky dG adduct in the context of the active site of the Y-family polymerase Dpo4, and a 5′ slippage pattern producing a single base deletion mutation was identified.[32]

In the present study, restrained MD simulations are used to characterize the free energy landscape of base-pair separation in a short repetitive hot-spot sequence for indel mutations. A partially flipped metastable conformation of a cytosine three base-pairs from the DNA terminus is identified. This metastable conformation is made more energetically favorable by non-canonical triplet base-pairing with a neighboring G:C base-pair. Unrestrained MD simulations started from such partially flipped states, and carried out using multiple nucleic acid force fields, suggest that partial base-flipping can enable strand slippage at DNA duplex termini. Unrestrained MD simulations starting from a corresponding partially flipped conformation of the partner guanine base reveal that such strand slippage can occur at both the 5′- and the 3′-termini. The consequences of this possible relationship between base unstacking, dynamic non-canonical base-pairing, and strand slippage are discussed with respect to sequence dependence of indel mutations.

RESULTS AND DISCUSSION

Metastable State during Cytosine Flipping. A 9-mer sequence CCCGGCTTC containing a repetitive stretch of three cytosines at the 5′-end of a stacked DNA duplex was chosen as a model system for the present study. The three consecutive cytosines in this sequence are part of a hot-spot sequence for single base insertion and deletion mutations when replicated by a Y-family polymerase Dbh.[33] This sequence has also been crystallized in complex with this polymerase showing the third cytosine in an extra-helical conformation leading to slippage in the template DNA strand.[10] Previous studies have characterized the energetics and pathways of base-flipping leading to extrahelical base conformations in DNA duplexes.[34−38] In the present study, explicit solvent umbrella sampling MD simulations for base-flipping of the third cytosine (CCCGGCTTC) were performed using a restraint applied to a previously developed pseudo-dihedral coordinate.[34] The free energy profile for the flipping of this cytosine along this coordinate is shown in Figure 1A in red. A free energy profile for cytosine flipping in a different sequence context in the center of a 12-mer DNA duplex (GTCAGCGCATGG) is also shown for comparison in black.

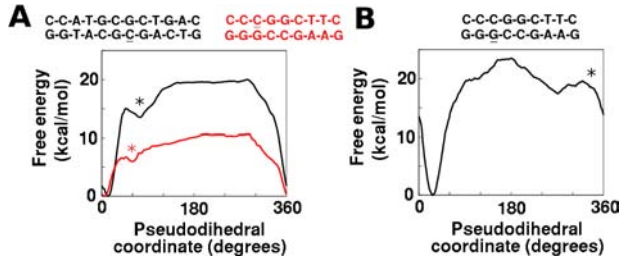


Figure 1. Free energy profiles for base-flipping in different sequence contexts. (A) Overlay of free energy profiles for cytosine (underlined) base-flipping near the center of a 12-mer DNA duplex (black) and near the terminus of a 9-mer DNA duplex (red). The 12-mer duplex is a methylation sequence for the *Hha1* cytosine-C5-methyl transferase and the 9-mer duplex is part of a repetitive hot-spot template sequence for indel mutations in the Y-family polymerase Dbh. Metastable states are indicated by asterices. (B) Free energy profile for guanine (underlined) base-flipping in the 9-mer DNA duplex. The asterisk in this panel indicates a partially flipped minor groove conformation for the guanine similar in extent of flipping to the metastable state for cytosine shown in panel A.

Both cytosine base-flipping free energy profiles have roughly the same shape with the major groove pathway showing a more gradual increase in free energy as compared to the minor groove pathway. Another notable commonality between the two profiles is the presence of a metastable state in the minor groove pathway of base-flipping (indicated by an asterisk). A distinct difference was observed between the two cytosine free energy profiles in the relative free energy of stacked versus flipped states, with the flipped states being more favorable in the hot-spot sequence by about 10 kcal/mol. This suggests that either the flipping energy landscape is dependent on the neighboring sequence, or it is influenced by proximity to the terminus, or a combination of both factors. The free energy profile for base-flipping of the paired guanine was also calculated (Figure 1B). A much higher relative free energy for the flipped states was observed as compared to its paired cytosine, which is also consistent with previous results.[34] This also suggests that proximity to the DNA duplex terminus is not

dx.doi.org/10.1021/ja401573j | J. Am. Chem. Soc. 2013, 135, 8274−8282

the sole factor for the observed stabilization of flipped states for the third hot-spot sequence cytosine.

The metastable state windows of the umbrella sampling MD simulations for the hot-spot cytosine base-flipping were examined structurally to identify the basis for their stabilization. The cytosine was partially flipped into the minor groove of its DNA duplex in these conformations (Figure 2A). It was

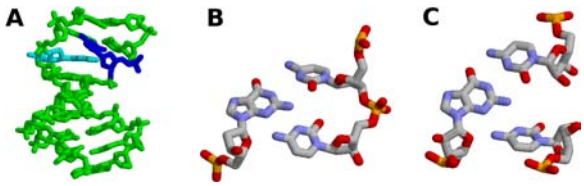


Figure 2. Metastable partially flipped minor groove conformation of a hot-spot sequence cytosine involving a non-canonical triplet base-pairing interaction. (A) The orientation of the partially flipped cytosine (blue) with respect to its partner guanine (cyan), and the rest of the DNA duplex (green). (B) The non-canonical base-paired triplet stabilizing the partially flipped conformation. (C) Similar non-canonical base-paired triplet observed in the RNA of the 2.4 Å crystal structure of the large ribosomal subunit in *Haloarcula marismortui*.[41].

stabilized through a non-canonical base-pairing interaction with the minor groove facing atoms of the GC base-pair immediately 5′ to it (Figure 2B). This non-canonical base-pairing interaction occurred primarily between the flipped cytosine and the opposite strand guanine and corresponded to a trans Watson−Crick:sugar edge C:G interaction according to the Leontis−Westhof nomenclature.[39,40] A representative example of this type of triplet base interaction between a C:G base-pair and another cytosine in the minor groove can also be seen for RNA in the 2.4 Å crystal structure of the large ribosomal subunit in *Haloarcula marismortui*[41] (Figure 2C). This suggests that such intramolecular non-canonical base-pairing is feasible and could play a role in stabilizing dynamically accessible metastable states in nucleic acids.

Dynamics in Partially Flipped Conformations. The stabilization of the minor groove flipped conformation, identified through the free energy profile, is also supported by the presence of the same non-canonical base-pairing interaction being previously recognized[40] and experimentally observed.[41] This metastable conformation was, however, observed through MD simulations carried out with an artificial external restraint. To observe the spontaneous behavior of this metastable state, multiple unrestrained MD simulations were performed starting from structures obtained from umbrella sampling windows near this conformation. A corresponding minor groove flipped starting conformation for the paired guanine was also examined using such unrestrained simulations. To avoid bias that may originate from use of a specific molecular mechanics (MM) force field, these simulations were also performed using multiple nucleic acid force fields.

Table 1 summarizes the different force fields used and the conformational dynamics observed during unrestrained MD simulations starting from partially flipped minor groove base conformations. The CHARMM27 force field,[42,43] three widely used AMBER nucleic acid force fields (AMBER94,[44] AMBER99,[45] and AMBERBSC0[46]), the BMS force field,[47] and the more recent CHARMM36 force field[48] were used. Conformational transitions in the vicinity of the flipped base were monitored and classified into four broad categories: maintenance of the partially flipped state (M); further flipping of the partially flipped base (F); restacking and proper Watson−Crick pairing of the partially flipped base (RP); and strand slippage (S), where the strand containing the partially flipped base slips by one base-pair relative to its paired strand. For the maintenance of the flipped conformation or further flipping, a positive sign in Table 1 indicates at least 1 ns of sampling in these conformational states. For restacking−pairing and strand slippage, no such residence time requirements were imposed to note presence. The detailed analysis used to obtain the data in Table 1 is in the Supporting Information text, supporting Tables 1−3, and supporting Figures 1−9. The final conformations in the 5 ns simulations for the terminal region are shown in supporting Figures 10−12, and the pseudodihedral coordinate used to characterize flipping is illustrated in supporting Figure 13. Examples of individual trajectories showing restacking-pairing and slippage for the flipped cytosine and guanine are shown in supporting videos 1−4.

The four conformational categories are not completely mutually exclusive, and multiple conformational states can be present in the same simulation. Substantial sampling of the starting partially flipped state of the cytosine was seen in 86% of the unrestrained simulations and further flipping of the cytosine was observed in about 21% of the simulations. Reversal of the partial flipping of the cytosine to a stacked and Watson−Crick paired cytosine was observed in about 6% of the simulations, suggesting that such distortions can be spontaneously corrected in the time scale of a few nanoseconds. Spontaneous strand slippage from the partially flipped cytosine state was also observed in about 10% of the simulations, suggesting that the metastable state could dynamically undergo conformational transitions not directly in a coordinate corresponding to cytosine flipping. For simulations starting from a partially flipped guanine state, substantial maintenance of the partially flipped state was observed in only about 49% of the simulations and further flipping of the guanine was observed in about 29% of the simulations. Restacking-pairing of the guanine was observed in about 49% of the simulations and strand slippage was observed in about 9% of the simulations. The stability of the strand slipped state originating from a partially flipped guanine was also lower on average than that due to a partially flipped cytosine (supporting Figures 5, 6, and 9). This could be due to intrinsic differences between reversal of guanine and cytosine base-flipping or due to the absence of specific stabilization by non-canonical base-pairing for the guanine in the starting state. The number of simulations performed is not numerous or long enough to make definitive observations about the tendencies among force fields, but the AMBERBSC0[46] and the BMS force fields[47] seem to show a greater propensity for reversion to B-form-like DNA structures, while the CHARMM27 seems to show a greater tendency for alternate conformational transitions.

Monitoring Strand Slippage. In the context of this hot-spot sequence, both primer slippage causing a single base insertion and template slippage causing a single base deletion involves loss of two existing and formation of two alternate C:G base-pairing interactions. The unrestrained simulations showing slippage due to cytosine or guanine base-flipping provide a clear look at the detailed atomic scale conformational changes involved. Two order parameters, which could simplify these complex conformational changes into one-dimensional coordinates, and could be used to monitor or enforce strand slippage are shown in Figure 3. A difference root-mean-square distance (RMSD) coordinate, earlier employed to study the A-

dx.doi.org/10.1021/ja401573j | J. Am. Chem. Soc. 2013, 135, 8274−8282

Table 1. Conformational States Induced by a Partially Flipped Base in the Minor Groove Three Base-Pairs Away from the DNA Terminus in Unrestrained MD Simulations Using Different Force Fields[a]

		Cyt60				Cyt90				Gua335			
force field	number	M	F	RP	S	M	F	RP	S	M	F	RP	S
CHARMM27	1	+				+	+			+	+		
	2	+	+			+	+		+			+	
	3	+	+			+	+					+	
	4	+									+	+	
	5	+				+	+		+	+			+
AMBER94	1	+				+				+	+	+	
	2	+			+	+				+			
	3	+				+				+			
	4	+				+				+		+	
	5	+	+			+							
AMBER99	1		+			+				+	+		
	2	+				+				+			
	3	+				+				+	+		
	4	+				+				+			
	5	+	+							+		+	+
	6		+					+				+	
	7	+										+	
	8	+			+	+						+	
	9	+				+						+	
	10	+			+	+						+	
AMBERBSC0	1	+				+						+	
	2	+				+			+	+	+		
	3	+						+		+	+		
	4	+				+						+	
	5	+				+	+				+		
BMS	1	+				+						+	
	2	+				+			+			+	
	3	+				+				+	+		
	4	+				+				+			
	5	+		+		+				+			+
CHARMM36	1	+				+	+						
	2	+						+		+			
	3	+	+									+	
	4					+	+				+		
	5	+				+	+					+	

[a]Cyt60, Cyt90, and G335 refer to unrestrained simulations with starting structure derived from the 60° cytosine, 90° cytosine, and 335° guanine windows of the restrained flipping umbrella sampling simulations, respectively. The different conformational change possibilities are labeled M, F, RP, and S for maintenance of the partially single base flipped state, further flipping, proper restacking and pairing, and slippage, respectively. Positive signs indicate presence, with the cutoff for presence in M and F scenarios being at least 20% (equivalent to 1 ns of sampling time) and derived from analysis of the pseudo-dihedral coordinate. No cutoff was used for RP and S scenarios, which were derived from the restacking−pairing and slippage coordinates, respectively. The detailed analysis from which these percentages are determined is shown in the Supporting Information. All unrestrained MD simulations were performed for 5 ns each, yielding a total sampling time of 525 ns.

form to B-form transition in DNA,[49] can be used for this conformational transition. The two reference structures in this case can be the starting partially flipped metastable conformation, and a conformation spontaneously attained after template slippage observed in the unrestrained simulations above (Figure 3C). A trajectory of template slippage in this difference RMSD coordinate, shown in Figure 3A, indicates the ability of this order parameter to clearly distinguish between the two end-point states. A possible drawback of this coordinate is that it requires prior definition of the two end-point structural states, and it is likely that the exact strand slipped structure is not known experimentally for most cases.

Another possible one-dimensional coordinate is a contact order-type parameter[50] that consists of a combination of distances that are different between the two states. Figure 3C shows six such distances that are well distinguished in the two reference state end-points for template slippage. These distances are between (1) the N3 atom in template strand C1 and the N1 atom in primer strand G9, (2) the N3 atom in template strand C2 and the N1 atom in primer strand G8, (3) the N3 atom in template strand C1 and the N1 atom in primer strand G8, (4) the N3 atom in template strand C2 and the N1 atom in primer strand G7, (5) the centers-of-mass of template strand C1 and G4 bases, and (6) the centers-of-mass of template strand C2 and G4 bases. Since the first four are standard base-pairing distances and the last two are center-of-mass distances between bases, this coordinate can be adapted to any sequence context, with no need to guess specific reference state end-point structures. The same slippage trajectory shown in Figure 3A is shown in Figure 3B for this coordinate. The

dx.doi.org/10.1021/ja401573j | J. Am. Chem. Soc. 2013, 135, 8274−8282

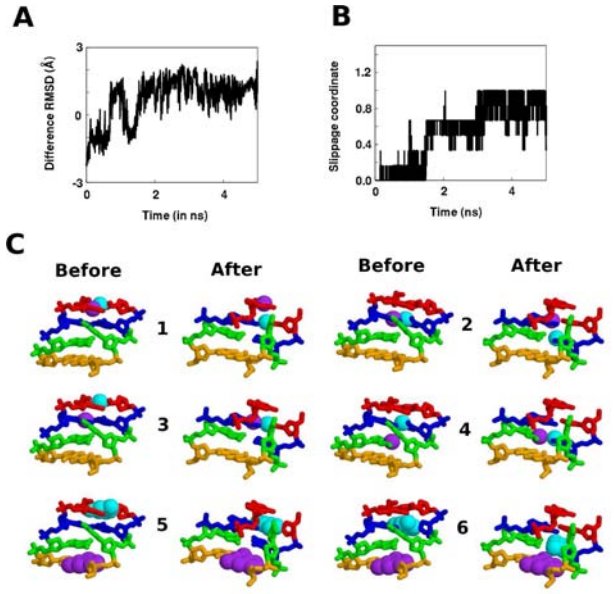


Figure 3. Feasible one-dimensional coordinate to monitor or enforce template slippage in MD simulations. (A) Strand slippage along a difference RMSD coordinate, (B) Strand slippage along a contact order type parameter composed of six distances, (C) the six distances used in the coordinate shown in B. These are between two atoms or two sets of atoms, depicted in sphere format and colored cyan and purple, shown in the conformations before or after strand slippage. The two conformations shown are also the two end-point reference conformations used for calculation of the difference RMSD coordinate. The numbering of the distances is consistent with their detailed description in the Methods section. The base-pair coloring used in all subsequent figures is as follows: red, C1:G9; blue, C2:G8; green, C3:G7; and orange, G4:C8.

stepwise changes in this coordinate during template slippage indicate the individual transitions occurring for each of these six distances.

Dynamics of the Strand-Slipped State. The behavior of the slipped intermediates was probed using longer simulations (100 ns duration with the CHARMM36 force field[48]) starting from two strand slipped conformations seen in the earlier 5 ns simulations. Simulation 1 is started from the final structure of the second AMBER94 Cyt60 simulation, and simulation 2 is started from the final structure of the tenth AMBER99 Cyt60 simulation (Table 1). Figure 4 shows the behavior during these two simulations in the pseudo-dihedral, restacking-pairing, and

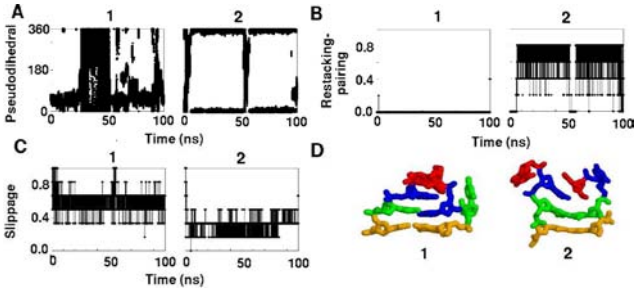


Figure 4. Dynamics of two slipped state structures examined by unrestrained 100 ns simulations using the CHARMM36[48] force field: (A) pseudo-dihedral coordinate in degrees, (B) restacking−pairing coordinate, (C) slippage coordinate, and (D) final structures from the two simulations (numbered 1 and 2).

slippage coordinates. As seen in Figure 4A, the pseudo-dihedral coordinate shows no consistent tendency to assume values corresponding to possible restacking, but does show significant further flipping in the first simulation. In the second simulation, the pseudo-dihedral coordinate suggests possible restacking, with short excursions to further flipped conformations. Figure 4B shows that no restacking-pairing occurs in the first simulation but restacking-pairing does seem to occur in the second simulation, with only one of the six distances (center-of-mass distance between C2 and C3) not meeting the cutoff criteria. Correspondingly, Figure 4C shows that the slipped state is frequented and revisited in the first simulation, whereas it is left and never reverted to in the second simulation. The final coordinates for the two simulations (Figure 4D) show that partial slippage with cytosine flipping is present at the end of the first simulation, while restacking of the cytosine with fraying of the two terminal base-pairs occurs in the second simulation. The variable behavior and relatively short residence times in metastable states seen in these simulations suggest that constructing an accurate transition probability matrix and building Markov models[51] may be easier with a larger number of shorter simulations rather than a smaller number of longer simulations for this system.

Possibility of Strand Slippage in Polymerase Active Sites. The observations made through these restrained and unrestrained simulations suggest that strand slippage can accompany partial base-flipping near DNA duplex termini, and can be made more persistent if the partially flipped conformation is stabilized through intrinsic non-canonical groove interactions. Since these simulations were performed for a solvated short DNA duplex, an important question is whether such partial base-flipping can also occur in polymerase active sites, which are likely to impose greater steric restrictions on DNA groove-related distortions. To address this question, the inclusion of strand slipped conformations was studied in two polymerase active sites for a low fidelity lesion bypass Y-family polymerase[10] and a high-fidelity replicative B-family polymerase.[52] Representative partially flipped structures observed in the solvated DNA simulations were overlaid on the duplex DNA bound to these polymerases in crystallographic structures, assuming the terminal base-pair to be the base-pair at the active site (shown in Figure 5). Primer or template strand slipped structures were then overlaid on these partially flipped structures. Following a limited optimization procedure with strong restraints on the protein structure, all of the structures could be accommodated in the protein active sites of both polymerases. A small steric overlap of the initial structures prior to optimization occurred for the B-family polymerase active site (Supporting Information Table 4), which is consistent with its greater tendency to surround the DNA, as well as its greater intrinsic fidelity. Panels A−H in Figure 5 show how partial single base-flipping near polymerase active sites could lead to either template or primer strand slippage by one position without any significant protein deformations. This suggests that the mechanisms highlighted to occur in solvated DNA by the present study could also be relevant in complex protein environments.

METHODS

The non-canonical base triplet interaction in the 2.4 Å crystal structure of the large ribosomal subunit in *Haloarcula marismortui*[41] was found using the FR3D server.[53] The restrained and shorter unrestrained MD simulations (5 ns) were performed using the program CHARMM,[54,55]

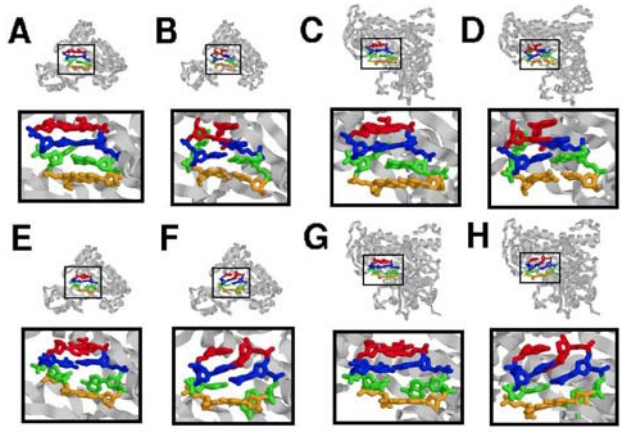


Figure 5. Overlay of representative partially flipped and strand slipped structures of a DNA duplex terminus onto DNA bound to two polymerase active sites. (A) partially flipped cytosine in Dbh active site, (B) template strand slipped structure in Dbh active site, (C) partially flipped cytosine in RB69 active site, (D) template strand slipped structure in RB69 active site; (E) partially flipped guanine in Dbh active site, (F) primer strand slipped structure in Dbh active site, (G) partially flipped guanine in RB69 active site, (H) primer strand slipped structure in RB69 active site. Dbh is a lesion-bypass Y-family polymerase,[10] and RB69 is a replicative B-family polymerase.[52] For clarity, the first 300 residues of RB69 are not shown as they occlude the DNA view.

and the longer unrestrained MD simulations (100 ns) were performed using the program NAMD.[56,57] The restrained umbrella sampling simulations were performed using the CHARMM27 force field[42,43] and the unrestrained simulations were performed using the CHARMM27, AMBER94,[44] AMBER99,[45] AMBERBSC0,[46] BMS,[47] and CHARMM36[48] force fields; with the TIP3P water model,[58] and sodium parameters from Beglov and Roux.[59] The minimized B-form 9-mer was solvated in a 55 Å dimension cubic box with randomly distributed neutralizing sodium ions. The CCCGGCTTC strand is considered the template strand and the complementary GAAGCCGGG strand is considered the primer strand to be consistent with the comparison with the Y-family polymerase Dbh hot-spot template sequence context. The final system consisted of 17839 atoms including 5752 waters and 16 sodium ions. All DNA non-hydrogen atoms were harmonically restrained with a force constant of 10 kcal/(mol·Å^2), and the full system was minimized using 5000 steps of steepest descent (SD) and 5000 steps of adopted-basis Newton−Raphson (ABNR) minimization with a energy convergence cutoff of 0.001 kcal/mol. Long-range electrostatic interactions were treated using the Particle Mesh Ewald (PME) approach[60] with a B-spline order of 4 and a Fast Fourier Transform grid of one point per Å and a real-space Gaussian width kappa of 0.3 Å^{-1}. Real space and Lennard-Jones (LJ) interaction cutoffs of 12 Å were used with non-bond interaction lists maintained and heuristically updated out to 16 Å. The entire system was minimized and the solvent environment was equilibrated for 20 ps using a constant pressure and temperature (NPT) ensemble[61] MD simulation with the same harmonic restraint on the solute. The force constant was gradually lowered in the next five 20 ps increments to 5, 2, 1, 0.5, and finally 0 kcal/(mol·Å^2). Each simulation was then continued without any restraints for the amount of time reported.

The procedure for calculation of the free energy of base-flipping using umbrella sampling with the periodic pseudo-dihedral coordinate is as previously described.[34,38] The 360° range of this coordinate was covered in 72 windows spaced 5° apart. The force constant for the pseudo-dihedral restraint was 500 kcal/(mol·radian^2) (or 0.15 kcal/(mol·deg^2)), and the simulation time in each window was 2 ns, yielding a total of 144 ns of sampling time per potential of mean force (PMF) profile shown. The pseudo-dihedral coordinate value at each step of the dynamics was saved and used for calculating the PMF using a periodic version of the Weighted Histogram Analysis Method (WHAM) as previously described.[34,38] Molecular pictures and videos were produced using VMD[62] or Rasmol,[63] and graphs were made using gnuplot version 4.4 (http://www.gnuplot.info) and compiled using GIMP version 1.2 (http://www.gimp.org) software.

To assess restacking combined with proper base-pairing (restacking-pairing) for the flipped cytosine, a coordinate composed of five distances was used: (1) between the O2 atom in template strand C3 and the N2 atom in primer strand G7, (2) between the N3 atom in template strand C3 and the N1 atom in primer strand G7, (3) between the N4 atom in template strand C3 and the O6 atom in primer strand G7, (4) between the centers-of-masses of template strand bases C2 and C3, and (5) between the centers-of-masses of template strand bases C3 and G4. For restacking-pairing of the flipped guanine, the corresponding distances were (1) between the O2 atom in template strand C3 and the N2 atom in primer strand G7, (2) between the N3 atom in template strand C3 and the N1 atom in primer strand G7, (3) between the N4 atom in template strand C3 and the O6 atom in primer strand G7, (4) between the centers-of-mass of primer strand bases C6 and G7, and (5) between the centers-of-mass of primer strand bases G7 and G8. The cutoff for the presence of base-pairing hydrogen bonds in the first three distances was 3.5 Å and the cutoff for presence of stacking interactions through the last two distances was 5 Å. Satisfaction of each of these cutoffs amounted to a value of 0.2 in the restacking coordinate, which meant complete restacking could be represented by coordinate values approaching 1.

To assess template strand slippage due to partial template strand cytosine flipping, a coordinate composed of six distances was used. These distances were between (1) the N3 atom in template strand C1 and the N1 atom in primer strand G9, (2) the N3 atom in template strand C2 and the N1 atom in primer strand G8, (3) the N3 atom in template strand C1 and the N1 atom in primer strand G8, (4) the N3 atom in template strand C2 and the N1 atom in primer strand G7, (5) the centers-of-mass of template strand C1 and G4 bases, and (6) the centers-of-mass of template strand C2 and G4 bases. For primer strand slippage due to partial primer strand guanine flipping, the corresponding distances were between (1) the N1 atom in primer strand G9 and the N3 atom in template strand C1, (2) the N1 atom in primer strand G8 and the N3 atom in template strand C2, (3) the N1 atom in primer strand G9 and the N3 atom in template strand C2, (4) the N1 atom in primer strand G8 and the N3 atom in template strand C3, (5) the centers-of-mass of primer strand G9 and C6 bases, and (6) the centers-of-mass of primer strand G8 and C6 bases. The following cutoffs for each of these distances were imposed to classify them as slipped (in the same order as above) (1) greater than 3.4 Å, (2) greater than 3.4 Å, (3) less than 3.4 Å, (4) less than 3.4 Å, (5) less than 9.0 Å, and (6) less than 6.0 Å. Satisfaction of any one of these distance cutoffs resulted in an approximate increment of 0.17 in the slippage coordinate.

The orientation of representative metastable flipped state and strand slipped cytosine structures for the sequence CCCGGCTTC onto the DNA duplexes in the polymerase active sites was based on overlaying the phosphate and sugar non-hydrogen atoms of the terminal three base-pairs using a Kabsch rigid-body least RMSD orientation.[64] The structure used for the metastable cytosine flipped state was the starting structure from the 60° umbrella sampling window for cytosine flipping used for the unrestrained simulations. The structure used for the partially flipped guanine state was similarly the unrestrained simulation starting structure from the 335° umbrella sampling window for guanine flipping. The cytosine strand slipped structure was the final structure from unrestrained Cyt60 simulation number 10 using the AMBER99 force field. The guanine strand slipped structure was the final structure from the unrestrained Gua335 simulation number 5 using the BMS force field. The 2.7 Å resolution Dbh Y-family DNA polymerase ternary complex structure (PDB ID: 3BQ1[10]) and the B-family replicative RB69 polymerase structure (PDB ID: 3RWU[52]) were used for the two polymerase active site structures. In the Y-family polymerase structure, the non-hydrogen atoms of chain T residues cytosine 3, cytosine 4, and guanine 5 were used as orientation

dx.doi.org/10.1021/ja401573j | J. Am. Chem. Soc. 2013, 135, 8274−8282

references for template strand cytosines 1−3; and the non-hydrogen sugar and phosphate atoms of chain P residues cytosine 9, guanine 10, and 2′,3′-dideoxy guanine 11 were used as orientation references for primer guanines 7−9. In the B-family polymerase structure, the non-hydrogen sugar and phosphate atoms of the chain T residues difluorotoluene 4, guanine 5, and thymine 6 were used as orientation references for template strand cytosines 1−3; and the non-hydrogen sugar and phosphate atoms of chain P 2′,3′-dideoxy cytosine residue 114, chain P cytosine residue 115, and chain A incoming ATP residue 902 were used as orientation references for primer guanines 7−9. Only the non-hydrogen sugar and phosphate atoms of the terminal three base-pairs were used to orient the terminal four base-pairs of the partially flipped structures in the protein active sites. The terminal four base-pairs of the primer and template slipped structures were also oriented onto the partially flipped structures using their non-hydrogen sugar and phosphate atoms. To remove any initial bad protein contacts with the overlaid DNA, the following optimization procedure was used: (a) 5000 steps of SD and 5000 steps of ABNR minimization to a energy change tolerance of 0.001 kcal/mol, (b) 5000 steps of Langevin dynamics at 300 K with a friction coefficient on all non-hydrogen atoms of 60 ps^{-1}, and (c) 5000 steps of SD and 5000 steps of ABNR minimization to an energy change tolerance of 0.001 kcal/mol. During this optimization procedure, all non-hydrogen atoms more than 10 Å away from any DNA non-hydrogen atom were held fixed, and all other non-hydrogen atoms had harmonic restraints with a force constant of 5 kcal/(mol·$Å^2$). There were no DNA atoms within 1.5 Å of any protein non-hydrogen atom in the final optimized structures, suggesting that these final models had no steric overlap between the DNA and the protein.

CONCLUSIONS

The present study provides a three-dimensional structural glimpse into detailed mechanisms for strand slippage that could result in single base indel mutations. The presence of a groove environment that stabilizes partially flipped states of bases near the active site can facilitate such slippage. This could conceivably be provided by protein atoms, but can also be provided intrinsically by the DNA sequence in the vicinity of the flipping base. This supports a direct mechanism for sequence dependence of strand slippage, where the flipped base establishes interactions with neighboring sequence DNA groove atoms,[34] and its stabilization causes an increased probability of strand slippage. Comparison between the cytosine and guanine flipping scenarios in the present simulations do show that increasing the stability of a partially flipped base can increase the probability of orthogonal structural changes such as slippage. Increasing the propensity for strand slippage through direct non-canonical neighboring DNA interactions may thus be one of the contributing factors to making a sequence a mutation hot-spot. This provides a plausible explanation for NMR data showing that the neighboring sequence in both the 5′- and 3′-directions can affect the propensity of strand slippage in DNA hairpin primer−template models.[18−22]

In addition to the connection between partial minor groove base-flipping and strand slippage, a similar relationship can also exist between major groove base-flipping and strand slippage. The more gradual increase in the free energy profile in the major groove pathway suggests that reversal of partially flipped states to Watson−Crick paired states might also occur with ease in the major groove. In the umbrella sampling simulations carried out with a pseudo-dihedral restraint, strand slippage was observed past the 290° window in the major groove cytosine flipping pathway, and past the 60° window in the major groove guanine flipping pathway. A two-dimensional free energy profile exploring this relationship between flipping and strand slippage can help to clarify the range of single-base flipped states that facilitate slippage.

Although this range is not explicitly narrowed down through unrestrained simulations in the present study, a primary insight provided is that the amount of flipping required for slippage to occur is not very significant. This provides a possible criterion for assessing the fidelity of a polymerase, i.e., its ability to prevent base-flipping and maintain proper primer−template base-pair interactions near its active site. Although plausible models were generated for the flipped and strand-slipped states in two polymerase active sites in this study, there is a clear need to follow up with simulations that directly assess the probability of such strand slippage in fully solvated protein−DNA environments. Such simulations can be performed for a variety of polymerase active sites with differing replication fidelity to identify the active site environmental features linked to fidelity. A restraint based on the slippage coordinates explored in this study is presently being developed to enable initial umbrella sampling simulations of strand slippage in different sequence contexts and environments. These can be refined by higher-cost, higher-accuracy methods such as transition path sampling[65] or the string method[66,67] to fully investigate the complex dynamics of these transitions.

Although replication forks have commonalities with DNA duplex termini, they also have an overhang template sequence after the active site template base that can affect the rate and propensity of strand slippage. The present simulations have not addressed the effects of such overhangs, which introduce a much larger number of structural possibilities in the solvated DNA context. They can be more efficiently explored in the protein contexts, since their structural variability is much more restricted by protein geometry. In general, the presence of such overhangs is expected to slow down or resist strand slippage by introducing friction due to overhang interactions with the protein. Partial flipping in double-helical RNA has been implicated in enzymatic deamination of the flipping base,[68] illustrating the possibility of catalysis in a partially flipped state. DNA replication errors can only occur if the strand slipped state is stable enough to allow the wrong nucleotide to be incorporated through a nucleotide addition reaction catalyzed at the polymerase active site. For the hot-spot sequence studied here, such catalytic activity in conjunction with slipped states is not in doubt for the Y-family polymerase Dbh, since indel mutations occur with about 50% probability.[33] For other DNA template sequences and protein environments, however, this connection between slippage and catalysis can be explored by modeling the entire catalytic cycle of nucleotide addition in the context of strand slipped structures.

ASSOCIATED CONTENT

Supporting Information

Tables, figures, four videos, and explanatory text. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

banavali@wadsworth.org

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

N.K.B. acknowledges computing resources provided by the Wadsworth Center of the New York State Department of Health, the Extreme Science and Engineering Discovery Environment (XSEDE, which is supported by National Science Foundation grant number OCI-1053575), and CCNI Blue Gene resources housed at the Rensselaer Polytechnic Institute. The choice of the hot-spot sequence originated from a collaboration with Dr. Janice Pata on the replication fidelity of Y-family polymerases, and this study greatly benefited from helpful discussions with her.

■ REFERENCES

(1) Garcia-Diaz, M.; Kunkel, T. *Trends Biochem. Sci.* **2006**, *31*, 206−214.
(2) Kunkel, T. *J. Biol. Chem.* **2004**, *279*, 16895−16898.
(3) Streisinger, G.; Okada, Y.; Emrich, J.; Newton, J.; Tsugita, A.; Terzaghi, E.; Inouye, M. Frameshift mutations and the genetic code. In *Cold Spring Harbor Symposia on Quantitative Biology*, Vol. *31*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1966.
(4) Kunkel, T.; Soni, A. *J. Biol. Chem.* **1988**, *263*, 14784−14789.
(5) Fujii, S.; Akiyama, M.; Aoki, K.; Sugaya, Y.; Higuchi, K.; Hiraoka, M.; Miki, Y.; Saitoh, N.; Yoshiyama, K.; Ihara, K.; Seki, M.; Ohtsubo, E.; Maki, H. *J. Mol. Biol.* **1999**, *289*, 835−850.
(6) Kunkel, T. *J. Biol. Chem.* **1986**, *261*, 13581−13587.
(7) Efrati, E.; Tocco, G.; Eritja, R.; Wilson, S.; Goodman, M. *J. Biol. Chem.* **1997**, *272*, 2559−2569.
(8) Ling, H.; Boudsocq, F.; Woodgate, R.; Yang, W. *Cell* **2001**, *107*, 91−102.
(9) Garcia-Diaz, M.; Bebenek, K.; Krahn, J.; Pedersen, L.; Kunkel, T. *Cell* **2006**, *124*, 331−342.
(10) Wilson, R.; Pata, J. *Mol. Cell* **2008**, *29*, 767−779.
(11) Ling, H.; Boudsocq, F.; Woodgate, R.; Yang, W. *Mol. Cell* **2004**, *13*, 751−762.
(12) Zhang, H.; Eoff, R.; Kozekov, I.; Rizzo, C.; Egli, M.; Guengerich, F. *J. Biol. Chem.* **2009**, *284*, 3563−3576.
(13) DeLucia, A.; Grindley, N.; Joyce, C. *Biochemistry* **2007**, *46*, 10790−10803.
(14) Woodson, S.; Crothers, D. *Biochemistry* **1987**, *26*, 904−912.
(15) Woodson, S.; Crothers, D. *Biochemistry* **1988**, *27*, 436−445.
(16) Weisenseel, J.; Moe, J.; Reddy, G.; Marnett, L.; Stone, M. *Biochemistry* **1995**, *34*, 50−64.
(17) Wang, Y.; Schnetz-Boutaud, N.; Saleh, S.; Marnett, L.; Stone, M. *Chem. Res. Toxicol.* **2007**, *20*, 1200−1210.
(18) Chi, L.; Lam, S. *FEBS Lett.* **2006**, *580*, 6496−6500.
(19) Chi, L.; Lam, S. *Biochemistry* **2007**, *46*, 9292−9300.
(20) Chi, L.; Lam, S. *Biochemistry* **2008**, *47*, 4469−4476.
(21) Chi, L.; Lam, S. *Biochemistry* **2009**, *48*, 11478−11486.
(22) Chi, L.; Lam, S. *J. Phys. Chem. B* **2012**, *116*, 1999−2007.
(23) Foloppe, N.; Gueroult, M.; Hartmann, B. *Methods Mol. Biol. (Clifton, NJ)* **2013**, *924*, 445−468.
(24) Liu, L.; Bradley, P. *Curr. Opin. Struct. Biol.* **2012**, *22*, 397−405.
(25) Konrad, M.; Bolonick, J. *J. Am. Chem. Soc.* **1996**, *118*, 10989−10994.
(26) Lebrun, A.; Lavery, R. *J. Biomol. Struct. Dyn.* **1998**, *16*, 593−604.
(27) MacKerell, A., Jr.; Lee, G. *Eur. Biophys. J.* **1999**, *28*, 415−426.
(28) Yang, L.; Beard, W.; Wilson, S.; Roux, B.; Broyde, S.; Schlick, T. *J. Mol. Biol.* **2002**, *321*, 459−478.
(29) Bebenek, K.; Garcia-Diaz, M.; Foley, M.; Pedersen, L.; Schlick, T.; Kunkel, T. *EMBO Rep.* **2008**, *9*, 459−464.
(30) Mura, C.; McCammon, J. *Nucleic Acids Res.* **2008**, *36*, 4941−4955.
(31) Foley, M.; Padow, V.; Schlick, T. *J. Am. Chem. Soc.* **2010**, *132*, 13403.
(32) Xu, P.; Oum, L.; Geacintov, N.; Broyde, S. *Biochemistry* **2008**, *47*, 2701−2709.
(33) Potapova, O.; Grindley, N.; Joyce, C. *J. Biol. Chem.* **2002**, *277*, 28157−28166.
(34) Banavali, N.; MacKerell, A. *J. Mol. Biol.* **2002**, *319*, 141−160.
(35) Varnai, P.; Lavery, R. *J. Am. Chem. Soc.* **2002**, *124*, 7272−7273.
(36) Huang, N.; Banavali, N.; MacKerell, A., Jr. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 68−73.
(37) Giudice, E.; Varnai, P.; Lavery, R. *Nucleic Acids Res.* **2003**, *31*, 1434−1443.
(38) Banavali, N.; MacKerell, A. *PLoS One* **2009**, *4*, e5525.
(39) Leontis, N.; Westhof, E. *Q. Rev. Biophys.* **1998**, *31*, 399−455.
(40) Leontis, N.; Westhof, E. *RNA* **2001**, *7*, 499−512.
(41) Ban, N.; Nissen, P.; Hansen, J.; Moore, P.; Steitz, T. *Science* **2000**, *289*, 905−920.
(42) Foloppe, N.; MacKerell, A., Jr. *J. Comput. Chem.* **2000**, *21*, 86−104.
(43) MacKerell, A.; Banavali, N. *J. Comput. Chem.* **2000**, *21*, 105−120.
(44) Cornell, W.; Cieplak, P.; Bayly, C.; Gould, I.; Merz, K.; Ferguson, D.; Spellmeyer, D.; Fox, T.; Caldwell, J.; Kollman, P. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(45) Cheatham, T., III; Cieplak, P.; Kollman, P. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845−862.
(46) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T., III; Laughton, C.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817−3829.
(47) Langley, D. *J. Biomol. Struct. Dyn.* **1998**, *16*, 487−509.
(48) Hart, K.; Foloppe, N.; Baker, C. M.; Denning, E. J.; Nilsson, L.; MacKerell, A. D., Jr. *J. Chem. Theory Comput.* **2011**, *8*, 348−362.
(49) Banavali, N.; Roux, B. *J. Am. Chem. Soc.* **2005**, *127*, 6866−6876.
(50) Shea, J.; Onuchic, J.; Brooks, C. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 16064−16068.
(51) Yang, S.; Banavali, N. K.; Roux, B. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 3776−3781.
(52) Xia, S.; Konigsberg, W.; Wang, J. *J. Am. Chem. Soc.* **2011**, *133*, 10003−10005.
(53) Sarver, M.; Zirbel, C. L.; Stombaugh, J.; Mokdad, A.; Leontis, N. B. *J. Math. Biol.* **2008**, *56*, 215−252.
(54) Brooks, B.; Bruccoleri, R.; Olafson, B.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(55) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(56) Nelson, M. T.; Humphrey, W.; Gursoy, A.; Dalke, A.; Kaĺ e, L. V.; Skeel, R. D.; Schulten, K. *Int. J. High Perf. Comput. Appl.* **1996**, *10*, 251−268.
(57) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(58) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926.
(59) Beglov, D.; Roux, B. *J. Chem. Phys.* **1994**, *100*, 9050−9063.
(60) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
(61) Feller, S.; Zhang, Y.; Pastor, R.; Brooks, B. *J. Chem. Phys.* **1995**, *103*, 4613−4621.
(62) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.
(63) Sayle, R.; Milner-White, E. *Trends Biochem. Sci.* **1995**, *20*, 374.
(64) Kabsch, W. *Acta Crystallogr. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr.* **1978**, *34*, 827−828.
(65) Bolhuis, P.; Chandler, D.; Dellago, C.; Geissler, P. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291−318.
(66) Weinan, E.; Ren, W.; Vanden-Eijnden, E. *Phys. Rev. B* **2002**, *66*, 052301.

(67) Weinan, E.; Ren, W.; Vanden-Eijnden, E. *J. Chem. Phys.* **2007**, *126*, 164103.
(68) Hart, K.; Nystrom, B.; Ohman, M.; Nilsson, L. *RNA* **2005**, *11*, 609−618.