THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Analyzing the Relationship between Single Base Flipping and Strand Slippage near DNA Duplex Termini

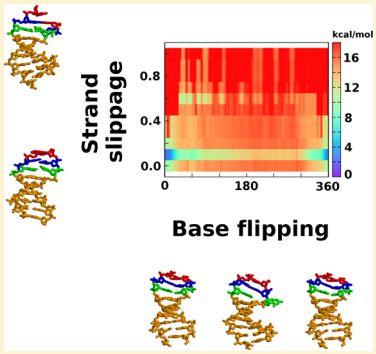
Nilesh K. Banavali*

Laboratory of Computational and Structural Biology, Division of Genetics, David Axelrod Institute, Wadsworth Center, New York State Department of Health, P.O. Box 22002, Albany, New York 12201-2002, United States

Department of Biomedical Sciences, School of Public Health, State University of New York at Albany, Center for Medical Sciences, P.O. Box 22002, Albany, New York 12201-2002, United States

Supporting Information

ABSTRACT: Insertion−deletion (indel) mutations are caused by strand slippage between pairing primer and template strands during nucleic acid strand extension. A possible causative factor for such strand slippage is base flipping in the primer strand or template strand, for insertion or deletion mutations, respectively. A simple mechanistic description is that the "hole" in the nucleic acid duplex left behind by a flipping base is occupied by a neighboring base on the same strand, resulting in slippage with respect to its paired strand. The extent of single base flipping required for occupation of its former place in the double helix by a neighboring base is not fully understood. The present study uses restrained molecular dynamics (MD) simulations along a pseudohedihedral base flipping parameter to construct two-dimensional free energy profiles along base flipping and strand slippage geometric parameters. These profiles, generated for both cytosine and guanine single base flipping in a short repetitive indel mutation hot-spot DNA sequence, illustrate the extent of single base flipping that can allow strand slippage by one base position. Relatively minor base flipping into both the major and minor grooves can result in strand slippage. Deconstruction of the collective variable strand slippage geometric parameter into its component distances illustrates the details of how strand slippage can accompany base flipping. The trans Watson−Crick:sugar edge interaction that stabilizes cytosine flipping in this hot-spot sequence is also characterized energetically. The impact of these results on understanding sequence dependence of indel errors in nucleic acid strand extension is discussed, along with a suggestion for future studies that can generalize the present findings to all nearest-neighbor sequence contexts.



INTRODUCTION

Errors in strand extension by nucleic acid polymerases prevent correct propagation of genomic information and are the underlying cause of all genetic mutations. A subset of such errors is indel mutations, where the newly synthesized nucleic acid strand either has extra nucleotides (insertion) or is missing some nucleotides (deletion) in complementarity to the template strand. Some mechanisms that have been proposed to account for such indel mutations include strand slippage,[1] dNTP stabilized misalignment,[2,3] misinsertion−misalignment,[4] and melting−misalignment.[5] The structural basis of indel mutations has been characterized in the context of polymerases by numerous crystallographic investigations,[6−10] which suggest that base flipping might be involved. NMR investigations of bulge-containing DNA duplexes suggested that bulges caused by base flipping can migrate[11] and show a nonterminal preference for localization.[12] NMR investigations on strand slippage propensity in DNA hairpin primer-template models suggest that these processes are dependent on the intrinsic DNA sequence.[13−17] Single molecule studies based on 2-aminopurine fluorescence also suggest that the template strand can slip to cause a single-base deletion mutation due to multiple base-pairing possibilities with the primer strand in a polymerase active site.[18]

DNA strand slippage has also been previously studied using computational methods based on explicit solvent MD simulations. Simulations of a solvated κB sequence showed spontaneous cross-strand intercalative base stacking and spontaneous base pair separation akin to single base slippage within a DNA double helix.[19] Crystal structure and MD simulation data on DNA polymerase λ bound to a primer-template suggests that binding of dNTP could induce template strand slippage.[20] Simulations also suggest that small variations in the DNA polymerase λ active-site environment can have larger dynamic effects on the ternary complex that could result in slippage and frameshift errors.[21] MD simulations combined with free energy analysis suggest that pol λ could even stabilize misaligned DNA better than aligned DNA.[22] 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 has also been probed by MD simulations,

Received: September 6, 2013
Revised: October 25, 2013
Published: October 25, 2013

   dx.doi.org/10.1021/jp408957c | J. Phys. Chem. B 2013, 117, 14320−14328

which indicated that a 5′ slippage pattern could produce a single base deletion mutation.[23] A recent study, which showed that partial single base flipping is sufficient for strand slippage near DNA duplex termini,[24] is followed up here by analyzing MD simulation data on restrained cytosine and guanine base flipping in a single base deletion hot-spot DNA sequence to obtain an explicit 2D free energy profile along the base flipping and strand slippage geometric parameters. The collective variable strand slippage geometric parameter is decomposed into its component distances to illustrate the detailed mechanism of slippage in response to single base flipping. The energetic stabilization of a trans Watson−Crick:sugar edge interaction that results in a metastable state for a partially flipped cytosine in this hot-spot sequence is also characterized. The relationship of these results to sequence dependence of indel mutations in nucleic acid strand extension is discussed, along with the possibilities of future studies.

METHODS

Molecular pictures were produced using Rasmol,[25] graphs were made using gnuplot version 4.4 (http://www.gnuplot.info), and composite figures were compiled using GIMP version 1.2 (http://www.gimp.org) software. All MD simulations were performed using the program CHARMM[26,27] with the nonpolarizable CHARMM27 force field[28,29] with the TIP3P water model,[30] and sodium parameters from Beglov and Roux.[31] The methods are the same as reported in the previous study[24] and are summarized below. A minimized B-form canonical duplex structure for the 9-mer hot-spot sequence CCCGGCTTC was solvated in a 55 Å cubic explicit solvent box consisting of 5752 water molecules and 16 randomly distributed neutralizing sodium ions. To remain consistent with the hot-spot sequence analysis,[32] CCCGGCTTC is considered the template strand and its pairing partner GAAGCCGGG is considered the primer strand. The system was minimized with harmonic restraints on the DNA non-hydrogen atoms and the solvent was equilibrated for 20 ps using a constant pressure and temperature (NPT) ensemble[33] MD simulation. Long-range electrostatic interactions were treated using Particle Mesh Ewald (PME)[34] with a B-spline order of 4 and a fast Fourier transform grid of one point per Å and a real-space Gaussian width κ of 0.3 Å^{-1}. A real space cutoff of 12 Å was used, which was the same as that for the Lennard-Jones (LJ) interactions. Nonbond pairwise atomic interaction lists were maintained and heuristically updated with a cutoff of 16 Å. The solute harmonic restraints were gradually reduced to zero over 100 ps of simulation.

As done previously,[24] the coordinate to measure strand slippage due to C3 flipping in the template strand was composed of six distances 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 center-of-masses of template strand C1 and G4 bases, and (6) the center-of-masses of template strand C2 and G4 bases. The coordinate to measure strand slippage due to G7 flipping in the primer strand was composed of six distances 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 center-of-masses of primer strand G9 and C6 bases, and (6) the center-of-masses of primer strand G8 and C6 bases. With the distances in the same order as above, to classify a specific distance as belonging to the strand slipped state, the following cutoffs were used: (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, (6) less than 6.0 Å. If any one of these distance cutoffs was satisfied, the collective slippage coordinate was incremented by 0.17.

The restrained MD simulation protocol to enforce single base flipping is as reported in the previous study.[24] Briefly, a pseudodihedral restraint of 500 (kcal/mol)/radian^2 (or 0.15 (kcal/mol)/deg^2) was used to enforce 2 ns of sampling in each of the 72 windows spaced 5° apart, yielding a total sampling time of 144 ns per PMF profile. The restrained pseudodihedral coordinate value (saved at 0.2 ps intervals) was used to calculate the PMF using a periodic version of the weighted histogram analysis method (WHAM) as previously described.[35,36] Two-dimensional free energy profiles along the pseudodihedral parameter and other unrestrained geometric parameters were calculated as previously described.[35] Briefly, a two-dimensional biased probability distribution, $\langle\rho_{\mathrm{bias}}(\xi_1,\xi_2)\rangle$ is first obtained from the umbrella sampling simulation trajectories, where ξ_1 is the pseudodihedral and ξ_2 is the other coordinate. The unbiased probability distribution $\langle\rho(\xi_1,\xi_2)\rangle$ is obtained using the WHAM equations,[37,38]

$$\langle\rho(\xi_1,\xi_2)\rangle = \frac{\sum_i^N n_i\langle\rho_{\mathrm{bias}}(\xi_1,\xi_2)\rangle}{\sum_j^N n_j \exp[F_j - w_j(\xi_1)/k_BT]} \quad (1)$$

where N is the number of windows, n_i and n_j are the number of time points (subscripts i and j indicating the histogram bins and simulation windows, respectively), F_j are the free energy constants for each window, and $w_j(\xi_1)$ is the biasing harmonic potential imposed along the pseudodihedral coordinate. The values of F_j can be obtained from the equation

$$\exp(F_j/k_BT) = \int\langle\rho(\xi_1,\xi_2)\rangle \exp[-w_j(\xi_1)/k_BT]\, d\xi_1\, d\xi_2 \quad (2)$$

To get optimal values of the two unknowns, F_j and $\langle\rho(\xi_1,\xi_2)\rangle$, eqs 1 and 2 are solved iteratively using a change of less than 0.0001 in successive F_j values as a convergence criterion. The free energy surface represented by $W(\xi_1,\xi_2)$ can obtained from

$$W(\xi_1,\xi_2) = -k_BT \ln(\langle\rho(\xi_1,\xi_2)\rangle) + C \quad (3)$$

where C is an arbitrary constant. In interpreting $W(\xi_1,\xi_2)$, it is important to remember that the sampling along the other coordinate is not enforced directly; therefore, the range accessible through nonthermally accessible barriers may not have been explored. Sampling of a particular range could be interpreted as an indicator of the presence of thermally accessible barriers to get to that range. The WHAM histogram bin dimensions were as follows: 0.5° for the pseudodihedral coordinate, 0.1 for the slippage coordinate, and 0.5 Å for the C3−G8 distances. The large bin width as compared to the total range for the slippage coordinate causes a pixelated appearance of the 2d free energy profiles in Figure 1 but cannot be avoided due to the possibility of only seven distinct values for the slippage coordinate.

dx.doi.org/10.1021/jp408957c | J. Phys. Chem. B 2013, 117, 14320−14328

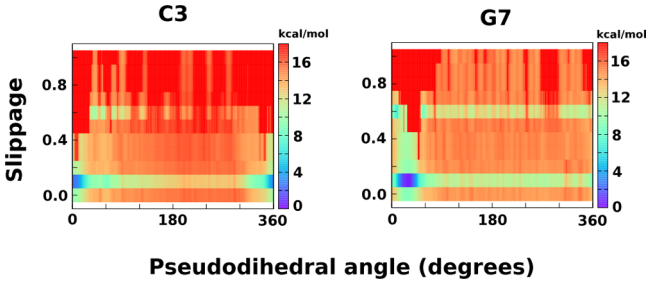


Figure 1. Two-dimensional free energy profile of single base cytosine C3 and guanine G7 flipping and accompanying single base position strand slippage.

RESULTS AND DISCUSSION

2D Free Energy Profiles of Cytosine or Guanine Single Base Flipping and Accompanying Strand Slippage. Single base flipping near DNA duplex termini can result in secondary conformational and dynamic effects such as strand slippage by one base position.[24] Because single base flipping can be simplified to a one-dimensional (1D) coordinate using a coarse-grained pseudodihedral definition (Supporting Information figure S1),[35] and strand slippage can be condensed to a composite 1D coordinate consisting of six distances,[24] the relationship between these two 1D coordinates can be explored using a combined 2D free energy profile. Figure 1 shows such 2D free energy profiles for two flipping bases from the hot-spot sequence 9-mer CCCGGCTTC: the third cytosine from strand 1 (C3), and its pairing partner, the seventh guanine from strand 2 (G7). As explained previously, these 2D profiles are generated from simulations imposing a single restraint along the pseudodihedral coordinate,[35,39] and using the saved trajectories to extract the conformational and energetic behavior along second unrestrained coordinate. Because this second coordinate is not explicitly restrained, its sampling is likely in response to the single restraint applied and cannot be guaranteed to be converged. Because the pseudodihedral is defined differently for the C3 and G7 flipping, the lowest energy state (blue region) is around 5° for C3 and around 30° for G7, the minor groove flipping pathway is in the range 5−90−185° for C3 and 210−300−30° for G7, and the major groove flipping pathway is in the range 185−275−5° for C3 and 5−95−210° for G7. Strand slippage can be judged to have occurred fully if the slippage coordinate value of 1 is populated. It can be seen in Figure 1 that strand slippage does not occur near the stacked state of the base for both C3 and G7. It increases as base flipping progresses through either groove. Along the pseudodihedral coordinate for C3, the first window that full strand slippage occurs in the minor groove is 40°, and for the major groove it is 280°. Along the pseudodihedral coordinate for G7, the first window that strand slippage occurs in the minor groove is 335°, and for the major groove is 95°. The occurrence of strand slippage in conjunction with both major and minor groove slipping suggests that formation of a vacant region in the DNA duplex, helped by other factors such as alternate Watson−Crick or noncanonical interactions, may be sufficient for strand slippage.

A significant metastable state is observed for C3 flipping through the minor groove (cyan region around 90°), in which strand slippage is also possible. This metastable state has been identified to be due to a trans Watson−Crick:sugar edge[40,41] intramolecular interaction of the flipping C3 with a neighboring guanine.[24] A total of 8 pseudodihedral coordinate windows (40, 55, 90, 130, 200, 210, 245, 280°) for C3 flipping and a total of 25 windows (95, 110, 115, 125, 130, 140, 145, 160, 170, 175, 180, 190, 200, 215, 220, 230, 235, 250, 260, 305, 310, 320, 325, 330, 335°) for G7 flipping show full strand slippage. As judged by this criterium, strand slippage is markedly more prevalent during flipping of G7, which may be related to the larger vacant region caused by its departure from the stacked state. It should be noted that this prevalence may be dependent on the restraint used to enforce flipping. In unrestrained simulations started from a structure from the 335° window for G7 flipping, restacking of the guanine was seen to be much more prevalent than strand slippage.[24] This suggests that, by preventing restacking of G7, the pseudodihedral restraint may facilitate observation of rarer or unstable transitions such as strand slippage in a shorter time scale.

Deconstructing the Relationship between Base Flipping and Strand Slippage. The parameters used to measure base flipping and strand slippage differ in one aspect: the pseudodihedral coordinate is a coarse-grained measure, whereas the strand slippage coordinate is a collective variable measure. The simplest deconstruction of the relationship between base flipping and strand slippage is therefore analysis of the individual base flipping relationships of the six distance coordinates that comprise the strand slippage coordinate. Figure 2 shows the 2D probability distributions of sampling along the base flipping pseudodihedral for C3 (panel A) and G7 (panel B) and the full slippage or its individual component distance coordinates. The specific atomic distances shown in the probability maps are illustrated for the flipped and strand-slipped states by depicting the two atoms as violet and cyan spheres in the DNA duplex and the cutoff for each individual distance is shown as a horizontal black line in the probability map. For C3 flipping, strand slippage requires sampling of the C1:G9 and C2:G8 distances above the cutoff line, and sampling of the C1:G8, C2:G7, C1:G4, and C2:G4 distances below the cutoff line. For G7 flipping, strand slippage requires sampling of the C1:G9 and C2:G8 distances above the cutoff line, and sampling of the C2:G9, C3:G8, C6:G9, and C6:G8 distances below the cutoff line. In both cases, the latter four distances are more indicative of strand slippage, because they reflect either formation of alternate Watson−Crick pairing interactions or stacking interactions, as opposed to loss of pairing interactions for the first two distances, which can occur through other transitions such as terminal base fraying. The probability maps for these individual distances reflect the relationship between base flipping and the full strand slippage coordinate; i.e., the windows that show full strand slippage also show the individual distances meeting their cutoff criteria. The additional information provided by the probability maps is that these individual distances can sample a large range of values, which is suggestive of the conformational variability within the DNA duplex that is difficult to capture in the simplified coarse-grained pseudodihedral coordinate.

To narrow down the dynamic behavior of these individual coordinates that results in strand slippage, transitions that lead from a completely nonstrand-slipped structure (slippage coordinate value = 0) to a fully strand-slipped structure (slippage coordinate value = 1) were mined from each window. The transitions were extracted as a continuous time series for each geometric variable where the slippage coordinates starts at a value of 0 and ends at a value of 1. Figure 3 illustrates five representative time series of the slippage coordinate showing such transitions distributed along the pseudohedihedral

dx.doi.org/10.1021/jp408957c | J. Phys. Chem. B 2013, 117, 14320−14328

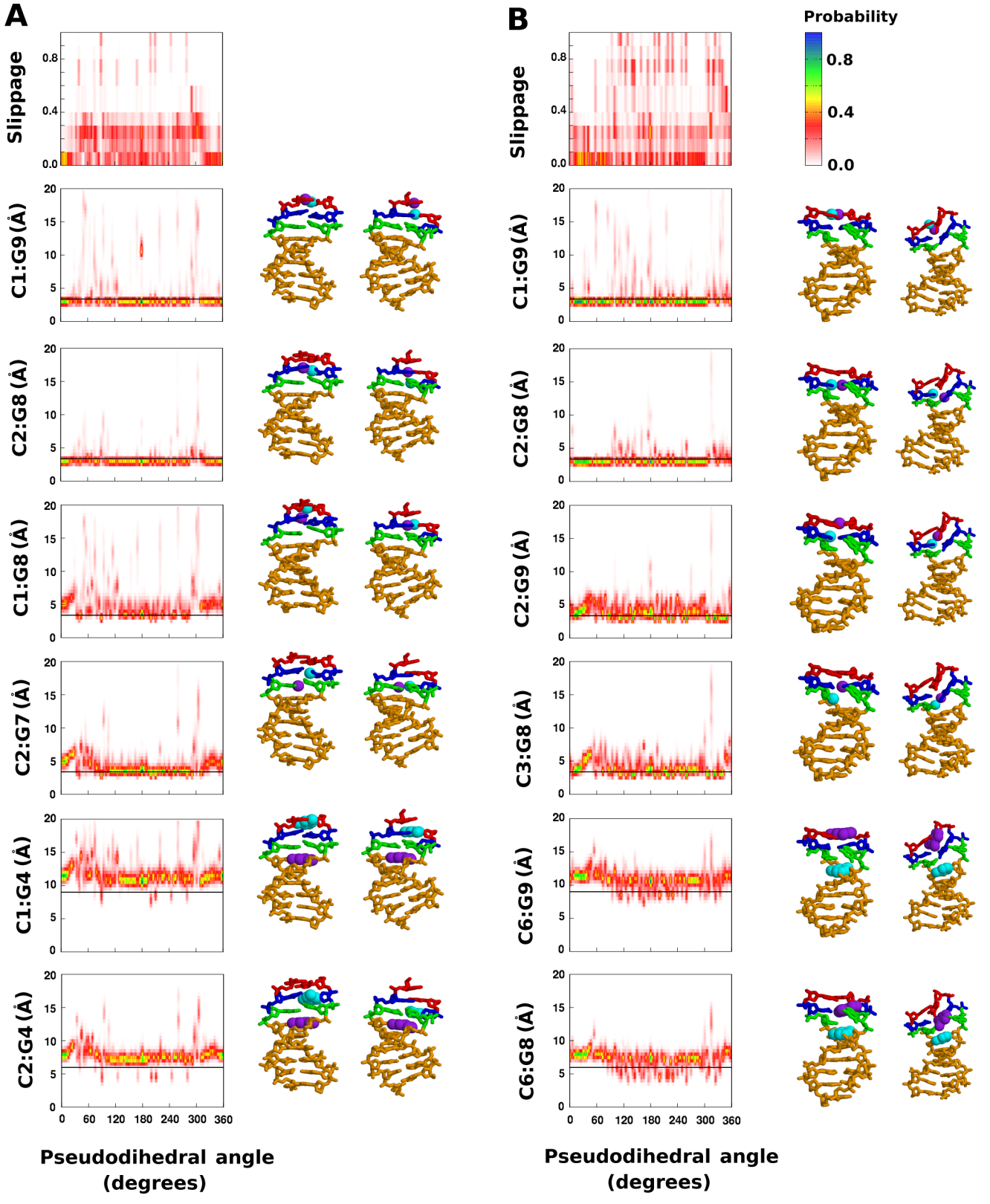


Figure 2. Probability maps of single base flipping and the full slippage coordinate or its individual component distance coordinates for (A) cytosine C3 flipping and (B) guanine G7 flipping. The distances are between pyrimidine N3 and purine N1 atoms in C1:G9, C2:G8, C1:G8, C2:G7, C2:G9, and C3:G8 coordinates. The distances are between the center-of-mass of non:hydrogen base atoms for C1:G4, C2:G4, C6:G9, and C6:G8 coordinates. The cutoffs used for the collective variable strand slippage coordinate are shown as horizontal black lines. The atoms participating in the individual distance coordinates are shown in cyan and purple spheres for the flipped and strand slipped states on the right of each probability map. In the DNA duplex illustrations, C1 and G9 are red, C2 and G8 are blue, C3 and G7 are green, and all other bases are orange.

dx.doi.org/10.1021/jp408957c | J. Phys. Chem. B 2013, 117, 14320−14328

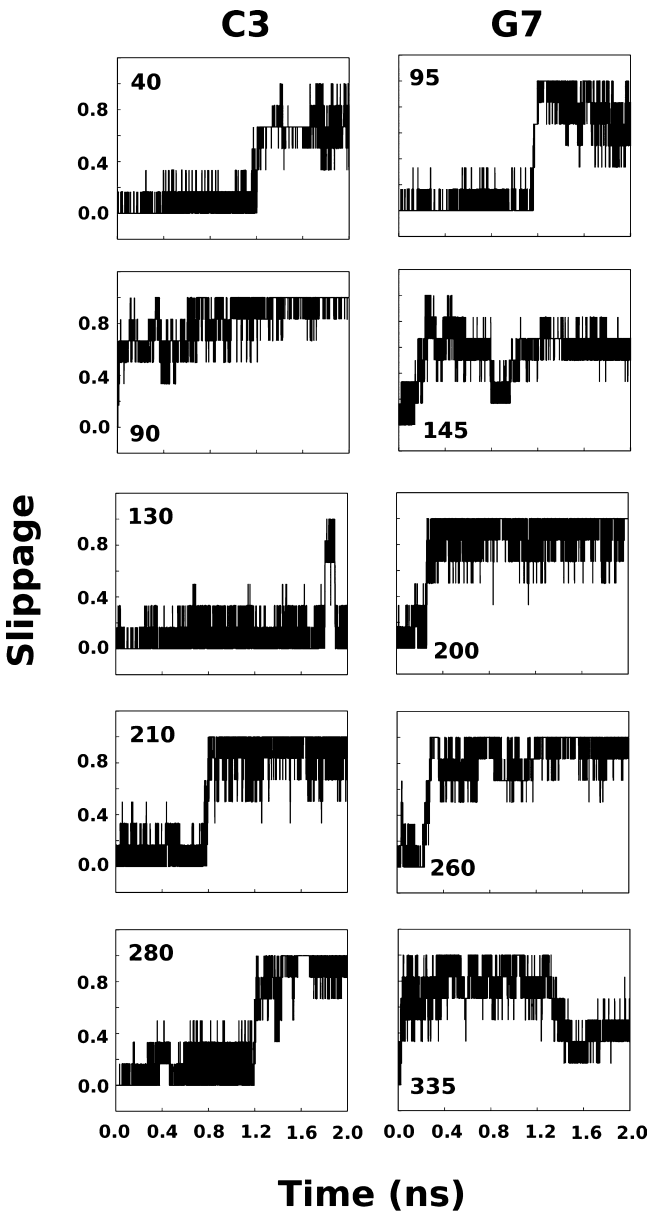


Figure 3. Five representative time series of the slippage coordinate showing full transitions that start from a slippage coordinate value of 0 and end in a slippage coordinate value of 1, for specific windows of the template strand cytosine 3 (C3) and primer strand guanine 7 (G7) base flipping restrained MD simulations. The corresponding pseudodihedral restraint minimum is shown inside each panel in degrees.

coordinate for C3 base flipping and G7 base flipping windows. In these transitions, the pairs of distance coordinates that are most related to one another are C1:G9 and C2:G8 (both C3 and G7 flipping), C1:G8 and C2:G7 (C3 flipping), C1:G4 and C2:G4 (C3 flipping), C2:G9 and C3:G8 (G7 flipping), or C6:G9 and C6:G8 (G7 flipping). The correlations between these pairs of distances in the slippage transitions during C3 flipping and G7 flipping are shown in Tables 1 and 2, respectively. The actual transition times in the mined transitions are shown in Table 3. There is a considerable variability in the slippage times, varying from 10.8 to 201.2 ps for C3 flipping windows and from 1.0 to 451.4 ps for G7 flipping windows. The shorter times recorded are possibly indicative of a state that has the base pairing and stacking of the

Table 1. Correlation Coefficients for Strand Slippage Component Distance Pairs during Strand Slippage Transitions Accompanying Cytosine C3 Flipping

pseudodihedral (deg)	C1:G9 and C2:G8	C2:G9 and C3:G8	C6:G9 and C6:G8
40	0.1	0.2	0.8
55	0.7	0.4	0.6
90	0.5	0.1	0.8
130	0.9	0.5	0.7
200	0.2	0.5	0.9
210	0.5	0.3	0.8
245	0.1	0.4	0.9
280	0.7	0.3	0.9

Table 2. Correlation Coefficients for Strand Slippage Component Distance Pairs during Strand Slippage Transitions Accompanying G7 Flipping

pseudodihedral (deg)	C1:G9 and C2:G8	C2:G9 and C3:G8	C6:G9 and C6:G8
95	0.3	0.2	0.8
110	−0.7	−0.1	0.6
115	0.3	−0.2	0.8
125	0.1	0.0	0.7
130	0.2	0.6	0.2
140	0.4	0.1	0.7
145	−0.1	0.3	0.7
160	0.5	−0.3	0.9
170	0.5	0.2	0.6
175	0.9	0.5	0.9
180	0.0	0.2	0.2
190	0.6	0.2	0.8
200	0.3	0.0	0.8
215	0.2	0.6	0.7
220	0.1	0.2	0.6
230	−0.1	0.3	0.8
235	0.0	0.0	0.8
250	0.8	0.5	0.8
260	0.5	0.4	0.5
305	0.3	0.1	0.7
310	−0.3	0.3	0.8
320	0.8	0.0	0.9
325	−0.1	0.0	0.8
330	0.1	0.2	0.7
335	0.0	0.2	0.7

nonstrand-slipped state but is particularly poised for strand slippage to occur. There is also considerable variability in the correlation coefficients (CC) between the distance pairs examined (e.g., from −0.7 to +0.9 for C1:G9 and C2:G8 distances in guanine G7 flipping). On average for C3 flipping, the CC between C1:G9 and C2:G8 distances is 0.46, between C1:G8 and C2:G7 distances is 0.34, and between C1:G4 and C2:G4 distances is 0.80. On average for G7 flipping, the CC between C1:G9 and C2:G8 distances is 0.22, between C2:G9 and C3:G8 distances is 0.18, and between C6:G9 and C6:G8 distances is 0.90. This higher CC value for the last pair of distances for both C3 and G7 flipping suggests that the motion of the two terminal bases toward the rest of the duplex occurs in a correlated fashion during slippage. A typical way in which this can happen is that the two bases do not lose their stacking with one another during slippage. Supporting Information videos S1−S4 show examples of such correlated motions that

dx.doi.org/10.1021/jp408957c | J. Phys. Chem. B 2013, 117, 14320−14328

Table 3. Times for Strand Slippage Transitions during Cytosine C3 and Guanine G7 Flipping

C3 pseudodihedral (deg)	slippage time (ps)	G7 pseudodihedral (deg)	slippage time (ps)
40	201.2	95	1.4
55	47.2	110	8.8
90	107.0	115	1.0
130	18.8	125	168.4
200	15.0	130	451.4
210	17.2	140	142.8
245	10.8	145	89.6
280	19.0	160	252.6
		170	3.6
		175	103.8
		180	90.2
		190	131.4
		200	4.2
		215	65.6
		220	9.4
		230	40.8
		235	126.4
		250	27.4
		260	50.8
		305	129.6
		310	205.6
		320	52.2
		325	126.8
		330	93.8
		335	21.8

result in strand slippage by one base position during both C3 and G7 flipping. The present results suggest that strand slippage, in addition to being affected by environmental factors, may involve an internal balance between stabilization through Watson−Crick pairing and base stacking of the slipping bases.

Relative Free Energy Cost of Full Strand Slippage. As seen in Figure 1, full strand slippage occurs spontaneously in specific umbrella sampling pseudodihedral coordinate windows for both C3 and G7 flipping, which suggests that the barriers for strand slippage in these windows are thermally accessible in the presence of the pseudodihedral restraint. Figure 4 shows the one-dimensional free energy variation of strand slippage coordinate values for specific pseudodihedral values at which strand slippage is seen to occur. The pseudodihedral coordinate value corresponding to the lowest free energy in the 2D profile shown in Figure 1 is also included for comparison (shown as filled red circles), as it corresponds to the fully stacked state of the flipping base, and therefore does not show spontaneous strand slippage. The definition of the strand slippage coordinate as a collective variable consisting of six different distance cutoff criteria, each given equal weight, has three manifestations: (a) a slippage coordinate does not represent a unique slipped state structure at values intermediate between the end-points of 0 and 1, (b) there are no small continuous variations, there is either maintenance of the slippage coordinate value or a vertical transition to one of the other six possible values, (c) some intermediate slippage coordinate values may not be populated at all, in which case, they are set to be equal to the upper boundary of the overall 2D free energy profile. For C3 flipping and G7 flipping, the slippage coordinate is not seen to extend past 0.4 for the lowest energy pseudodihedral values (0° and 29°, respectively). The fully strand slipped states for other

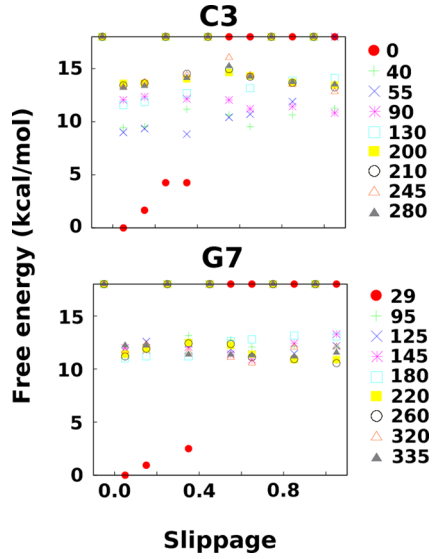


Figure 4. Free energies along the slippage coordinate for specific pseudodihedral values in which full strand slippage occurs. For comparison, the lowest energy pseudodihedral value is also included (0° for C3 flipping and 29° for G7 flipping, shown in filled red circles). The other pseudodihedral values are shown in degrees next to their symbols on the right.

specific pseudodihedral values extracted are between 10 and 15 kcal/mol higher than these lowest energy states, suggesting that it is a low probability phenomenon for this solvated DNA terminus. Within individual pseudodihedral values, however, the fully slipped state can be the lowest energy state (pseudodihedral values 90 and 245 for C3 flipping and 220, 260, and 335 for G7 flipping in Figure 1), which suggests that full strand slippage actually may reduce the energetic strain introduced by flipping to these pseudodihedral values.

Stabilization of a Partially Flipped Cytosine through a Noncanonical Interaction. During C3 flipping through the minor groove, a specific minimum is observed in the base flipping profile that is attributed to a trans Watson−Crick:sugar edge[40,41] intramolecular interaction of the flipping base with a guanine on the opposite strand.[24] This is an example of a noncanonical internal interaction that could stabilize a specific extrahelical conformation of the flipped base, and thus enable other structural changes such as strand slippage. Such stabilizing interactions can also be provided by the environment, e.g., by protein side chains in an enzyme active site,[42] but the ones resulting from the neighboring DNA atoms provide a direct mechanism for sequence dependence of strand slippage.[24] To quantify the behavior of this interaction in conjunction with flipping, the 2D free energy profiles of the C3 base flipping pseudodihedral and the N3−N2 or N4−N3 distances between cytosine C3 and guanine G8 bases were calculated and are shown in Figure 5A. A depiction of these distances in a structure where the C3−G8 trans Watson−Crick:sugar edge is present is shown in Figure 5B. The lowest energy state is located at a pseudodihedral value near 0°, but there is also a clear metastable state near 90°, at which both distances show a propensity to be near 3 Å, which is indicative of a hydrogen bond. As can be expected, this interaction is only possible while cytosine C3 flips through the minor groove, to which the N2 and N3 atoms of guanine G8 are exposed. Although this type of triplet base interaction has been observed in rRNA,[43] there the cytosine base was not adjacent in

dx.doi.org/10.1021/jp408957c | J. Phys. Chem. B 2013, 117, 14320−14328

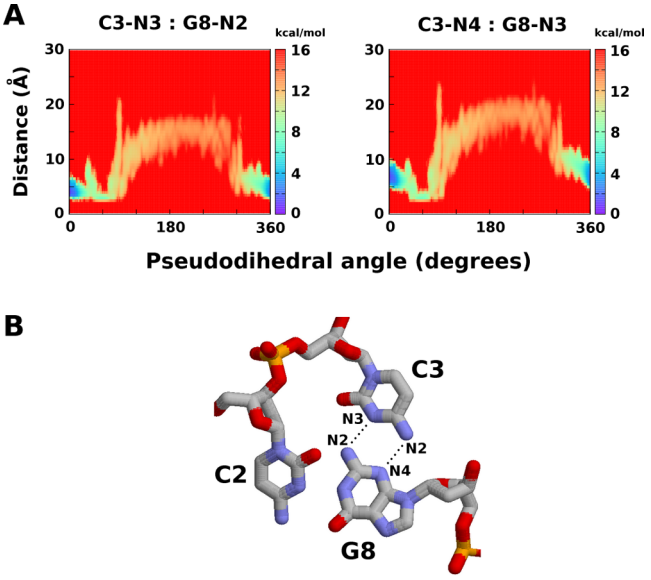


Figure 5. Two-dimensional free energy profiles of single base template strand cytosine 3 (C3) flipping pseudodihedral and the two distances representing the noncanonical trans Watson−Crick:sugar edge[40,41] hydrogen bonding with the opposite strand guanine 8 (G8). (A) Free energy profiles corresponding to the distance between atom N3 in C3 and atom N2 in G8 (left) and for the distance between atom N4 in C3 and atom N3 in G8 (right). The metastable state seen around pseudodihedral values of 90° coincides with hydrogen-bonding-like distances for both these distance variables. (B) Structural depiction of the two distances in a state where the trans Watson−Crick:sugar edge interaction is present between C3 and G8.

sequence to the G:C base pair it was interacting with. The presence of these hydrogen bonding interactions in conjunction with the observed local minimum near a pseudodihedral value of 90° further bolsters the role of this noncanonical interaction in stabilizing this partially flipped state.

CONCLUSIONS

The present study provides an analysis of the relationship between single base flipping near the duplex terminus of an indel hot-spot DNA sequence and single base position strand slippage, based on umbrella sampling MD simulations carried out with a single pseudodihedral restraint to enforce the base flipping. Strand slippage by one base position is seen to accompany base flipping in many windows where the flipped base is in either the minor groove or the major groove of the duplex. The number of windows showing full strand slippage during cytosine C3 flipping (8 windows, 11%) is lower than those seen during guanine G7 flipping (25 windows, 35%). There is a difference between the pseudodihedral restraint applied in cytosine (C3) flipping and its pairing partner guanine (G7) flipping. Both restraints use the same G4:C6 base pair as the first center-of-mass in the pseudodihedral, but this is the 3′ base pair for C3 and the 5′ base pair for G7. During the G7 flipping, this choice prevents a biasing force from being applied to the terminal C1:G9 and C2:G8 base pairs that are intimately involved in strand slippage. It is therefore possible that this restraint difference is implicated in the greater strand slippage observed during G7 flipping. Due to the lack on any biasing force on the two bases that slip, a much more likely factor seems to be the larger volume left unoccupied by a flipping guanine base in the DNA duplex. It should be noted that the likelihood of restacking for the guanine base is also probably higher because of this same factor,[24] but this restacking is explicitly prevented by the pseudodihedral restraint in the present simulations.

Slippage being seen in multiple windows of the restrained simulations suggests that it can occur due to various extents of single base flipping occurring through both duplex grooves. Probability-based analysis of the relationship between the base flipping and individual distances comprising the slippage coordinate reflect this overall slippage coordinate behavior. When the regions in which the actual slippage transitions occur are isolated, and analysis of pairwise correlations between related distance variables is performed, a significant role for intrastrand base stacking in the slippage transition is suggested. The 2D potential of mean force profiles of base flipping with the two distances involved in the trans Watson−Crick:sugar edge interaction of the flipping cytosine with a neighboring base pair guanine on the opposite strand show the clear presence of this interaction in a specific region of the minor groove pathway (pseudodihedral values near 90°). This interaction could likely be representative of a more general possibility: once the flipping base loses its Watson−Crick base pairing with its partner base, it could seek alternate interactions with exposed atoms in its flipping path, some of which could result in metastable states that facilitate other structural transitions. The intramolecular atomic partners in such interactions could belong to other base, sugar, or phosphate moieties, and the intermolecular atomic partners could belong to ligands, interacting macromolecules, ions, or solvent molecules. The flipping base is covalently attached to its backbone, but without its Watson−Crick base pairing, it has a lot of freedom in flipped states to explore either direction of the double helix, even to interact with opposite strand nucleotides up to five base pairs away (see video S1 in the Supporting Information). This posits a direct role for the sequence of the nucleic acid region in affecting alternative dynamic behavior related to base flipping.

Strand slippage in response to base flipping near DNA duplex termini has been observed in the absence of restraints with different force fields,[24] but the details of the 2D flipping−slippage profiles could be significantly force-field dependent, which could be examined by replicating the present study using different force fields. If more recent force fields are designed to give less variation from starting experimental geometries,[44−47] it is likely that they would show a lowered tendency of alternative transitions such as strand slippage. It should also be noted that base flipping in the base pair at the polymerase active site could result in single base indels without additional base-pair separation, and only formation of a new intrastrand base stacking pattern. Similarly, base flipping at a position next to the base pair at the polymerase active site could also result in single base indels, which would require breaking and re-formation of only one base pair along with new intrastrand base stacking. For the present hot-spot sequence, Y-family polymerase crystal structures seem to implicate base flipping of the third cytosine,[10] but base flipping in the third position from the terminus is not necessarily always the culprit for strand slippage.

The present study provides a template for identifying and categorizing noncanonical interactions and metastable states that lead to strand slippage by studying single base flipping in all possible local sequence contexts using restrained molecular dynamics simulations. If only base flipping at position 3 from

the terminus is to be considered, only the first three base positions from the termini are varied, and only 4 standard nucleic acid bases (adenine, guanine, thymine/uracil, cytosine) are studied, there are 64 possible trinucleotide sequence contexts (or 128 simulation sets considering the pairing partner base, i.e., both 5′ and 3′-termini). Studying these contexts for both DNA and RNA, and assuming 2 ns of sampling per pseudodihedral window spaced 5° apart, would require a total sampling time of under 40 μs. This sampling time is large but not prohibitive for the presently available computational infrastructures[48,49] and would characterize all sequence-dependent interactions possibly involved in strand slippage that are identifiable at this level. With 512 tetranucleotide sets, 2048 pentanucleotide sets, 8192 hexanucleotide sets, etc., extending the analysis to longer sequence contexts, which would include nonterminal duplex sequence variations, would get progressively more expensive but may also be feasible in the future.

■ ASSOCIATED CONTENT

Supporting Information

One figure depicting the pseudodihedral restraint and four videos (S1−S4 named jp408957c_si_002.mpg, jp408957c_si_003.mpg, jp408957c_si_004.mpg, jp408957c_si_005.mpg) showing trajectories of strand slippage in response to base flipping in the minor and major grooves. This information is available free of charge via the Internet at http://pubs.acs.org/.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: banavali@wadsworth.org. Tel: 518-474-0569. Fax: 518-402-4623. Website: http://www.wadsworth.org/resnres/bios/banavali.htm.

Notes

The author declares no competing financial interest.

■ ACKNOWLEDGMENTS

N.K.B. acknowledges computing resources provided by the Wadsworth Center, New York State Department of Health. This work also used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number OCI-1053575, and CCNI resources housed at the Rensselaer Polytechnic Institute.

■ REFERENCES

(1) Streisinger, G.; Okada, Y.; Emrich, J.; Newton, J.; Tsugita, A.; Terzaghi, E.; Inouye, M. Frameshift mutations and the genetic code. *Cold Spring Harbor Symposium on Quantum Biology, Cold Spring Harbor, NY*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1966; pp 77−84.

(2) Kunkel, T. A. Frameshift mutagenesis by eucaryotic DNA polymerases in vitro. *J. Biol. Chem.* **1986**, *261*, 13581−13587.

(3) Efrati, E.; Tocco, G.; Eritja, R.; Wilson, S. H.; Goodman, M. F. Abasic translesion synthesis by DNA polymerase β violates the A-rule. *J. Biol. Chem.* **1997**, *272*, 2559−2569.

(4) Kunkel, T. A.; Soni, A. Mutagenesis by transient misalignment. *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.; et al. DNA Replication Errors Produced by the Replicative Apparatus of Escherichia coli. *J. Mol. Biol.* **1999**, *289*, 835−850.

(6) Zhang, H.; Eoff, R. L.; Kozekov, I. D.; Rizzo, C. J.; Egli, M.; Guengerich, F. P. Versatility of Y-family Sulfolobus solfataricus DNA polymerase Dpo4 in translesion synthesis past bulky N2-alkylguanine adducts. *J. Biol. Chem.* **2009**, *284*, 3563−3576.

(7) Ling, H.; Boudsocq, F.; Woodgate, R.; Yang, W. Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. *Cell* **2001**, *107*, 91−102.

(8) Ling, H.; Boudsocq, F.; Woodgate, R.; Yang, W. Snapshots of replication through an abasic lesion: structural basis for base substitutions and frameshifts. *Mol. Cell* **2004**, *13*, 751−762.

(9) Garcia-Diaz, M.; Kunkel, T. A. Mechanism of a genetic glissando*: Structural biology of InDel mutations. *Trends Biochem. Sci.* **2006**, *31*, 206−214.

(10) Wilson, R. C.; Pata, J. D. Structural insights into the generation of single-base deletions by the Y family DNA polymerase dbh. *Mol. Cell* **2008**, *29*, 767−779.

(11) Woodson, S. A.; Crothers, D. M. Proton nuclear magnetic resonance studies on bulge-containing DNA oligonucleotides from a mutational hot-spot sequence. *Biochemistry* **1987**, *26*, 904−912.

(12) Woodson, S. A.; Crothers, D. M. Preferential location of bulged guanosine internal to a G. cntdot. C tract by proton NMR. *Biochemistry* **1988**, *27*, 436−445.

(13) Chi, L. M.; Lam, S. L. Sequence Context Effect on Strand Slippage in Natural DNA Primer−Templates. *J. Phys. Chem. B* **2012**, *116*, 1999−2007.

(14) Chi, L. M.; Lam, S. L. NMR investigation of DNA primer-template models: guanine templates are less prone to strand slippage upon misincorporation. *Biochemistry* **2009**, *48*, 11478−11486.

(15) Chi, L. M.; Lam, S. L. Nuclear Magnetic Resonance Investigation of Primer- Template Models: Formation of a Pyrimidine Bulge upon Misincorporation. *Biochemistry* **2008**, *47*, 4469−4476.

(16) Chi, L. M.; Lam, S. L. NMR investigation of primer-template models: Structural effect of sequence downstream of a thymine template on mutagenesis in DNA replication. *Biochemistry* **2007**, *46*, 9292−9300.

(17) Chi, L. M.; Lam, S. L. NMR investigation of DNA primer−template models: Structural insights into dislocation mutagenesis in DNA replication. *FEBS Lett.* **2006**, *580*, 6496−6500.

(18) DeLucia, A. M.; Grindley, N.; Joyce, C. M. Conformational changes during normal and error-prone incorporation of nucleotides by a Y-family DNA polymerase detected by 2-aminopurine fluorescence. *Biochemistry* **2007**, *46*, 10790−10803.

(19) Mura, C.; McCammon, J. A. Molecular dynamics of a κB DNA element: base flipping via cross-strand intercalative stacking in a microsecond-scale simulation. *Nucleic Acids Res.* **2008**, *36*, 4941−4955.

(20) Bebenek, K.; Garcia-Diaz, M.; Foley, M. C.; Pedersen, L. C.; Schlick, T.; Kunkel, T. A. Substrate-induced DNA strand misalignment during catalytic cycling by DNA polymerase λ. *EMBO Reports* **2008**, *9*, 459−464.

(21) Foley, M. C.; Schlick, T. Simulations of DNA Pol λ R517 mutants indicate 517′s crucial role in ternary complex stability and suggest DNA slippage origin. *J. Am. Chem. Soc.* **2008**, *130*, 3967−3977.

(22) Foley, M. C.; Padow, V. A.; Schlick, T. The Extraordinary Ability of DNA Pol λ to Stabilize Misaligned DNA. *J. Am. Chem. Soc.* **2010**, *132*, 13403.

(23) Xu, P.; Oum, L.; Geacintov, N. E.; Broyde, S. Nucleotide Selectivity Opposite a Benzo [a] pyrene- Derived N 2-dG Adduct in a Y-Family DNA Polymerase: A S-Slippage Mechanism. *Biochemistry* **2008**, *47*, 2701−2709.

(24) Banavali, N. K. Partial base flipping is sufficient for strand slippage near DNA duplex termini. *J. Am. Chem. Soc.* **2013**, *135*, 8274−8282.

(25) Sayle, R. A.; Milner-White, E. J. RASMOL: biomolecular graphics for all. *Trends Biochem. Sci.* **1995**, *20*, 374.

(26) 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.; et al. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.

(27) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy,

minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.
(28) MacKerell, A. D., Jr.; Banavali, N. K. All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J. Comput. Chem.* **2000**, *21*, 105−120.
(29) Foloppe, N.; MacKerell, A. D., Jr. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.* **2000**, *21*, 86−104.
(30) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926.
(31) Beglov, D.; Roux, B. Finite representation of an infinite bulk system: solvent boundary potential for computer simulations. *J. Chem. Phys.* **1994**, *100*, 9050−9063.
(32) Potapova, O.; Grindley, N.; Joyce, C. M. The mutational specificity of the Dbh lesion bypass polymerase and its implications. *J. Biol. Chem.* **2002**, *277*, 28157−28166.
(33) Feller, S. E.; Zhang, Y. H.; Pastor, R. W.; Brooks, B. R. Constant-pressure molecular-dynamics simulation-the Langevin piston method. *J. Chem. Pys.* **1995**, *103*, 4613−4621.
(34) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N log (N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089.
(35) Banavali, N. K.; MacKerell, A. D., Jr. Characterizing Structural Transitions Using Localized Free Energy Landscape Analysis. *PLoS One* **2009**, *4*, e5525.
(36) Banavali, N. K.; MacKerell, A. D., Jr. Free energy and structural pathways of base flipping in a DNA GCGC containing sequence. *J. Mol. Biol.* **2002**, *319*, 141−160.
(37) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. The weighted histogram analysis method for free-energy calculations on biomolecules 0.1. The method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(38) Souaille, M.; Roux, B. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comput. Phys. Commun.* **2001**, *135*, 40−57.
(39) Banavali, N. K.; Roux, B. Free energy landscape of A-DNA to B-DNA conversion in aqueous solution. *J. Am. Chem. Soc.* **2005**, *127*, 6866−6876.
(40) Leontis, N. B.; Westhof, E. Geometric nomenclature and classification of RNA base pairs. *RNA* **2001**, *7*, 499−512.
(41) Leontis, N. B.; Westhof, E. Conserved geometrical base-pairing patterns in RNA. *Q. Rev. Biophys.* **1998**, *31*, 399−455.
(42) Huang, N.; Banavali, N. K.; MacKerell, A. D., Jr. Protein-facilitated base flipping in DNA by cytosine-5- methyltransferase. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 68−73.
(43) Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2.4 A° resolution. *Science* **2000**, *289*, 905−920.
(44) Krepl, M.; Zgarbova, M.; Stadlbauer, P.; Otyepka, M.; Banas, P.; Koca, J.; Cheatham, T. E., III; Jurecka, P.; Sponer, J. Reference simulations of noncanonical nucleic acids with different χ variants of the AMBER force field: quadruplex DNA, quadruplex RNA, and Z-DNA. *J. Chem. Theory Comput.* **2012**, *8*, 2506−2520.
(45) Péerez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., III; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophys. J.* **2007**, *92*, 3817−3829.
(46) Hart, K.; Foloppe, N.; Baker, C. M.; Denning, E. J.; Nilsson, L.; MacKerell, A. D., Jr. Optimization of the CHARMM additive force field for DNA: improved treatment of the BI/BII conformational equilibrium. *J. Chem. Theory Comput.* **2011**, *8*, 348−362.
(47) Denning, E. J.; Priyakumar, U.; Nilsson, L.; MacKerell, A. D., Jr. Impact of 2-hydroxyl sampling on the conformational properties of RNA: Update of the CHARMM all-atom additive force field for RNA. *J. Comput. Chem.* **2011**, *32*, 1929−1943.
(48) Klepeis, J. L.; Lindorff-Larsen, K.; Dror, R. O.; Shaw, D. E. Long-timescale molecular dynamics simulations of protein structure and function. *Curr. Opin. Struct. Biol.* **2009**, *19*, 120−127.
(49) Freddolino, P. L.; Liu, F.; Gruebele, M.; Schulten, K. Ten-microsecond molecular dynamics simulation of a fast-folding WW domain. *Biophys. J.* **2008**, *94*, L75−L77.