



## Cytosine Unstacking and Strand Slippage at an Insertion–Deletion Mutation Sequence in an Overhang-Containing DNA Duplex

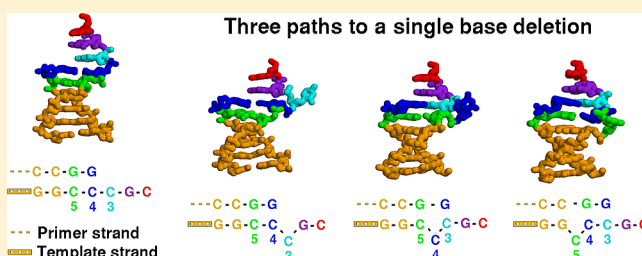
Swati R. Manjari,<sup>†</sup> Janice D. Pata,<sup>†,‡</sup> and Niles K. Banavali<sup>\*,†,‡</sup>

<sup>†</sup>Laboratory of Computational and Structural Biology, Division of Genetics, Biggs Laboratory, Wadsworth Center, New York State Department of Health, Empire State Plaza, PO Box 509, Albany, New York 12201-0509, United States

<sup>‡</sup>Department of Biomedical Sciences, School of Public Health, University at Albany, Center for Medical Sciences, PO Box 22002, Albany, New York 12201-02002, United States

### Supporting Information

**ABSTRACT:** Base unstacking in template strands, when accompanied by strand slippage, can result in deletion mutations during strand extension by nucleic acid polymerases. In a GCCC mutation hot-spot sequence, which was previously identified to have a 50% probability of causing such mutations during DNA replication by a Y-family polymerase, a single-base deletion mutation could result from such unstacking of any one of its three template cytosines. In this study, the intrinsic energetic differences in unstacking among these three cytosines in a solvated DNA duplex overhang model were examined using umbrella sampling molecular dynamics simulations. The free energy profiles obtained show that cytosine unstacking grows progressively more unfavorable as one moves inside the duplex from the 5'-end of the overhang template strand. Spontaneous strand slippage occurs in response to such base unstacking in the direction of both the major and minor grooves for all three cytosines. Unrestrained simulations run from three distinct strand-slipped states and one non-strand-slipped state suggest that a more duplexlike environment can help stabilize strand slippage. The possible underlying reasons and biological implications of these observations are discussed in the context of nucleic acid replication active site dynamics.



Newly synthesized DNA single strands, generated by catalysis of nucleotide addition by DNA polymerases, are sometimes not completely complementary to the DNA single strands used as templates.<sup>1</sup> If the lack of complementarity is due to a frameshift involving insertion or deletion of bases in the new (primer) strand, and the frameshift is not in multiples of three bases, then the protein encoded by the genetic code past the initial frameshift position is completely different. DNA replication errors occur in all DNA polymerases *in vitro*, but their frequency can vary exponentially among different polymerase types.<sup>2</sup> Replicative DNA polymerases are intrinsically more accurate but also cannot process templates with lesions or damage,<sup>3</sup> while lesion-repair polymerases can process damaged templates but do not have good intrinsic fidelity.<sup>4</sup>

Multiple mechanisms have been proposed to explain the origin of indel mutations, including DNA strand slippage,<sup>5</sup> misinsertion–misalignment,<sup>6</sup> melting–misalignment,<sup>7</sup> and dNTP-stabilized misalignment.<sup>8,9</sup> All of these mechanisms include a base unstacking process that changes the probability of the right template:primer base pair at the polymerase active site. Experimental studies of error-prone lesion-bypass polymerases have provided important insights into the structural details of indel mutation mechanisms. For example, in a crystal structure of the error-prone Y-family polymerase Dpo4, the correct template base was observed to be unstacked with its neighboring base acting as the template instead.<sup>10</sup> Unstacking

of a primer base was also observed in this polymerase for templates containing abasic<sup>11</sup> or O6-benzyl-dG modifications.<sup>12</sup> In a crystal structure of the Y-family polymerase Dbh, an unstacked template base was found three base positions from the active site.<sup>13</sup> For this polymerase, it was also shown using 2-aminopurine fluorescence that base unstacking at the primer terminus can result in template slippage that restores pairing at the primer terminus but results in a single-base deletion mutation.<sup>14</sup>

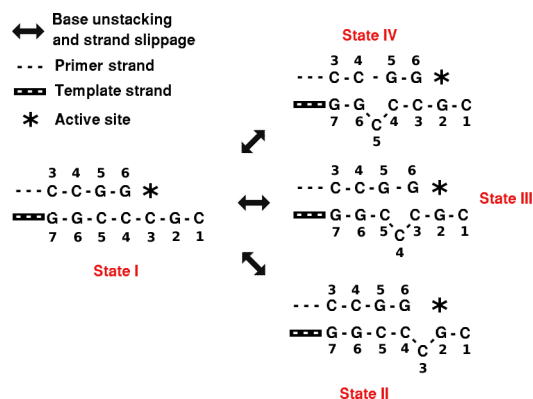
A specific GCCC sequence motif was found to be an indel mutation hot spot for the Y-family polymerase Dbh, with the exceptionally high error probability of ~50%.<sup>15</sup> As shown in Figure 1, unstacking of one of the three template cytosines is likely involved in the presentation of an incorrect template base and the occurrence of single-base deletion mutations. However, the relative propensities of unstacking of each of these cytosines are not clear. A series of nuclear magnetic resonance studies on a hairpin DNA model system have shown that the neighboring sequence affects slippage propensity.<sup>16–20</sup> In this study, explicit solvent molecular dynamics (MD) simulations are used to characterize the intrinsic probability of unstacking of the three cytosines in the context of a duplex representing the scenario at

**Received:** February 12, 2014

**Revised:** April 24, 2014

**Published:** May 22, 2014





**Figure 1.** Schematic depiction of the three cytosine unstacked states that could result in a single-base deletion mutation during DNA replication of the template strand. State I is the cytosine stacked state that would result in proper replication because an incoming guanine nucleotide would be incorporated into the primer strand at the active site (indicated by the asterisk). States II–IV are cytosine unstacked states that would all result in an incoming cytosine nucleotide being incorporated in the primer strand because they would present a guanine template base at the active site.

a DNA replication site where the 5' hot-spot cytosine does not yet have a primer strand partner base. Free energy profiles of unstacking of the three cytosines starting from state I are examined. Unrestrained MD simulations are used to characterize the unrestrained behavior of the four states depicted in Figure 1. The dynamics of the overhang template strand is also analyzed during the unrestrained simulations to understand its contribution to the unstacking process. The implications for the observed dynamics of these unstacking processes in affecting DNA polymerase fidelity are discussed.

## METHODS

The MD simulations were performed using CHARMM<sup>21,22</sup> (version c33b1) with the CHARMM27 force field,<sup>23,24</sup> the TIP3P water model,<sup>25</sup> and sodium and chloride parameters from Beglov and Roux.<sup>26</sup> Models for each state were constructed on the basis of a ternary complex structure of Dbh (Protein Data Bank entry 3BQ1) showing DNA elongation after the creation of a single-base deletion.<sup>13</sup> This complex has an extrahelical template base located three nucleotides to the 3'-side of the templating base. For state IV, the sequence of the DNA was altered to position the 5'-G of the deletion hot-spot sequence (5'-GCCC-3') in the nascent base pair binding pocket. For states II and III, the extrahelical nucleotide in state IV was moved to positions one and two nucleotides to the 3'-side of the templating base, maintaining the same extrahelical conformation. For state I, the duplex DNA in the state II model was translocated away from the active site by 1 bp, and the extrahelical base in state II was moved to stack between the 5'-G of the deletion hot-spot sequence and the template base paired with the primer terminus. This left a gap between the incoming nucleotide and primer terminal base. All models were energy minimized by conjugate gradient minimization with no experimental energy terms in CNS.<sup>27,28</sup> The minimized models for the four states of the B-form DNA duplex with the 9-mer strand CGCCGGCT pairing with a complementary strand 6-mer AGCCGG were solvated in a 54 Å dimension cubic box with a random distribution of 21 sodium ions and 7 chloride ions obtained

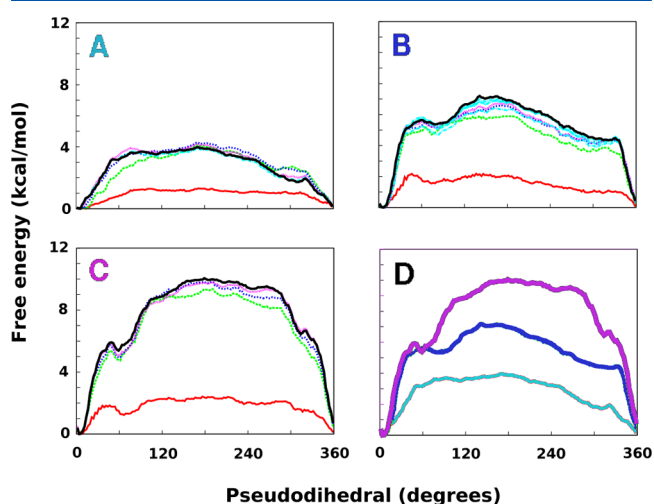
through Monte Carlo optimization. For the analogy with a DNA replication site, the 9-mer strand is considered the template strand and the 6-mer strand the primer strand. The final systems consisted of 15221 atoms for state I, 15188 atoms for state II, 15191 atoms for state III, and 15209 atoms for state IV. All DNA non-hydrogen atoms were harmonically restrained with a force constant of 10 kcal mol<sup>-1</sup> Å<sup>-2</sup>, and the full system was minimized using 1000 steps of steepest descent (SD) and 500 steps of adopted-basis Newton–Raphson (ABNR) algorithms. This minimized system for state I was used as the starting point for all the umbrella sampling simulations of base unstacking. For all simulations, long-range electrostatic interactions were treated using the particle mesh Ewald (PME) approach<sup>29</sup> with a B spline order of 4, a Fast Fourier Transform grid of one point per angstrom, and a real-space Gaussian width ( $\kappa$ ) of 0.3 Å<sup>-1</sup>. Real-space and Lennard-Jones (LJ) interaction cutoffs of 12 Å were used with nonbond interaction lists heuristically updated to 16 Å. For the unrestrained simulations, the entire system was minimized and the solvent environment was equilibrated for 20 ps using a constant-pressure and -temperature (NPT) ensemble<sup>30</sup> with the same harmonic restraint on the solute non-hydrogen atoms, and a weak center-of-mass restraint on all DNA non-hydrogen atoms (1 kcal mol<sup>-1</sup> Å<sup>-2</sup>) to prevent a drift to the edges of the solvent box. The harmonic force constant maintaining the internal structure of the DNA non-hydrogen atoms was gradually lowered in the next five 20 ps increments to 5.0, 2.0, 1.0, 0.5, and finally 0.0 kcal mol<sup>-1</sup> Å<sup>-2</sup>. Each simulation was then continued without any structural harmonic restraints for up to 10 ns, yielding a total of 40 ns for the four states that were studied.

The umbrella sampling free energy profiles for cytosine unstacking were obtained using a previously described periodic pseudodihedral restraint,<sup>31,32</sup> which is illustrated in Figure 1 of the Supporting Information. The full 360° range of this coordinate was covered in 72 windows spaced 5° apart. The initial structures in each window were generated starting from the stacked state by imposing a pseudodihedral restraint of 25000 kcal mol<sup>-1</sup> rad<sup>-2</sup> (or 7.6 kcal mol<sup>-1</sup> deg<sup>-2</sup>) and enforcing base unstacking in 5° increments through the minor and major grooves using 500 steps of SD minimization and 500 steps of NPT dynamics for each window. To obtain a better approximation of a relaxed solvent environment for the starting structures in each window, all water molecules were deleted and the DNA and ions were resolvated in the previously equilibrated 54 Å water box. The DNA non-hydrogen atoms were then harmonically restrained with a force constant of 10 kcal mol<sup>-1</sup> Å<sup>-2</sup>, which was gradually lowered in the next five 20 ps increments to 5.0, 2.0, 1.0, 0.5, and 0.0 kcal mol<sup>-1</sup> Å<sup>-2</sup>. The initial 0.1 ns of the umbrella sampling simulations in each window therefore includes gradually decreasing harmonic restraints in addition to the pseudodihedral restraint, which results in an artificially lowered free energy profile. The simulations in each window were continued for up to 0.8 or 1.4 ns, yielding a total of 232 ns for all three cytosine unstacking potential of mean force (PMF) profiles. Strand slippage subsequent to cytosine unstacking, which is required for complete transition to states II–IV from state I, was not enforced through a restraint. The pseudodihedral coordinate value, saved every 0.2 ps of the dynamics, was used for calculating the PMF using a periodic version of the weighted histogram analysis method as described elsewhere.<sup>31,32</sup> The slippage coordinate consists of two component distance cutoffs

between (1) G2:C4 and C1:C4 bases for C3 unstacking, (2) C3:C5 and G2:C5 bases for C4 unstacking, and (3) C4:G6 and C3:G6 bases for C5 unstacking. The slippage coordinate is increased by 0.4 for the first distance being  $<9 \text{ \AA}$  and by 0.6 for the second distance being  $<6 \text{ \AA}$  to yield a slippage coordinate value of 1.0 for the fully slipped state. Molecular movies were produced using VMD version 1.9,<sup>33</sup> and molecular pictures were produced using Rasmol version 2.7.<sup>34</sup> Graphs were made using gnuplot version 4.4 (<http://www.gnuplot.info>), and all figures were compiled using GIMP version 1.2 (<http://www.gimp.org>).

## RESULTS

**Free Energy Profiles for Cytosine Unstacking.** Base unstacking and flipping in DNA duplex contexts can occur through the major or minor groove of the DNA duplex and can be simplified into a one-dimensional coordinate using a center-of-mass pseudodihedral definition<sup>31</sup> (Figure 1 of the Supporting Information). Although improvements in the original coordinate have been explored,<sup>35</sup> they are applicable only for base flipping in a stacked duplex context, and not at duplex termini. The potential of mean force (PMF) or free energy profiles for cytosine unstacking were therefore obtained using explicit solvent umbrella sampling MD simulations with the original pseudodihedral restraint and are shown in Figure 2.



**Figure 2.** Free energy profiles and their convergence for cytosine unstacking in the template strand sequence CGCCCGCT: (A) cytosine 3, (B) cytosine 4, (C) cytosine 5, and (D) a comparison among cytosines 3–5. In panels A–C, the black line represents the overall free energy profile while the other lines represent the profiles for different extents of sampling per window as follows: 0.16 ns (red), 0.32 ns (dotted green), 0.48 ns (dotted blue), 0.64 ns (dotted cyan), 0.80 ns (dotted yellow), 0.96 ns (dotted black), and 1.12 ns (dotted orange). All profiles are mostly converged at 0.4 ns per window, but sampling was extended to at least 0.8 ns in each window for confirmation. In panel D, all lines represent overall free energy profiles with data for cytosines 3–5 shown as cyan, blue, and purple bold lines, respectively.

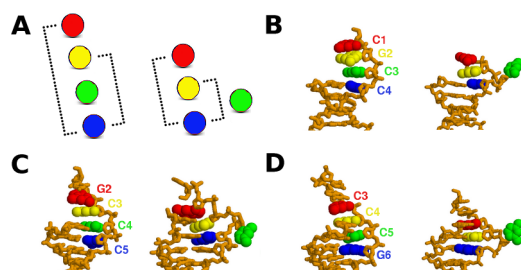
The starting state for all these umbrella sampling simulations was the fully stacked state I shown in Figure 1. The convergence of the free energy profiles was tested by calculating progressive free energy profiles for an increasing amount of sampling in 0.16 ns increments. As shown in panels A–C of Figure 2, the profiles were mostly converged after the first two

increments, i.e., in  $<0.4 \text{ ns}$  of sampling per window, and extension of sampling to 0.8 ns per window did not change the converged profiles. For cytosine 4 unstacking, the sampling was extended even further to 1.4 ns per window, which also did not change the profile significantly. The overall base unstacking profiles for the three cytosines in the hot-spot GCCC sequence near the template terminus are shown in Figure 2D. These profiles are similar to those obtained in previous studies,<sup>31,32</sup> with a well-defined minimum for the lowest-energy stacked state and a relatively flat landscape for the fully unstacked states. The stacked state energy well is located in the periodic pseudodihedral range of  $300^\circ$  to  $0^\circ$  to  $60^\circ$ , with the rest of the range occupied by unstacked states. There are clear differences among these unstacked states for the three profiles, with the free energy increases from the lowest-energy stacked state showing the following trend: cytosine 5 (C5)  $>$  cytosine 4 (C4)  $>$  cytosine 3 (C3). This suggests that unstacking is easiest for C3. The flat parts of the energy landscape of the unstacked states for the three cytosines are separated by  $\sim 3 \text{ kcal/mol}$ . These observations agree with the expectation that the lack of nearby stacked bases at duplex termini would decrease stability and result in easier unstacking.

For C4 and C5 unstacking, the first  $30^\circ$  in the pseudodihedral coordinate around the stacked state minimum are very similar through the major and minor groove pathways. However, these two free energy profiles begin to diverge from each other at a pseudodihedral of  $\sim 40^\circ$  in the minor groove pathway and  $\sim 340^\circ$  in the major groove pathway. In the case of C3, both pathways show a much more gradual free energy increase, which suggests that the base pairing interaction with a partner guanine (absent for C3) is involved in the steepness of the energy well near the cytosine fully stacked state. A distinct energy well can be observed at  $\sim 60^\circ$  for C5, which has previously been ascribed to a noncanonical trans Watson–Crick:sugar edge<sup>36,37</sup> hydrogen bonding interaction with an opposite strand guanine<sup>38</sup> on the 5'-base pair. This distinct energy well is not visible in C3 or C4 unstacking, which is consistent with the absence of the opposite strand guanine for these two cytosines. A broader energy well past the minor groove unstacking barrier is visible for C4 unstacking at a pseudodihedral value of  $\sim 90^\circ$ . These three profiles clarify the intrinsic location-dependent energetic effects for an unstacking base in an overhang-containing duplex terminus that resembles DNA at a polymerase active site.

**Template Strand Slippage in Response to Cytosine Unstacking.** The umbrella sampling MD simulations looking at cytosine unstacking do not enforce the strand slippage required to precisely convert from state I to the other three states shown in Figure 1. To attain states II–IV, additional strand slippage in the template strand by one base position is required. Whether these transitions can occur spontaneously in response to base unstacking can be examined by additionally monitoring a strand slippage coordinate. A simple measure of strand slippage (shown in Figure 3), which can be applied to all three cytosine unstacking scenarios, can be obtained by combining two center-of-mass distances between template strand non-hydrogen base atoms: (1) G2:C4 and C1:C4 for C3 unstacking, (2) C3:C5 and G2:C5 for C4 unstacking, and (3) C4:G6 and C3:G6 for C5 unstacking. In each case, the distances between the two neighboring 5'-bases and the base 3' to the unstacking base are combined. For example, if C5 is the unstacking base, C4 coming within  $6 \text{ \AA}$  of G6 increases the slippage coordinate by 0.6. If, in addition, C3 also comes within





**Figure 3.** Depiction of the slippage coordinate and its two component center-of-mass distances. (A) Base centers of mass for nonslipped and fully slipped states with dotted lines showing the two slippage coordinate component distances. (B–D) Before and after states in unstacking and strand slippage for cytosines 3–5, respectively.

9 Å of G6, the slippage coordinate value increases by 0.4 to a total value of 1. The unequal weighting for the two individual distances allows them to be distinguished in the free energy profiles and has no physical basis. The two-dimensional (2D) profiles for unstacking and strand slippage are created from unrestrained evolution of the strand slippage as previously described.<sup>32,39</sup> It is therefore possible that sampling in this unrestrained degree of freedom is not fully converged.

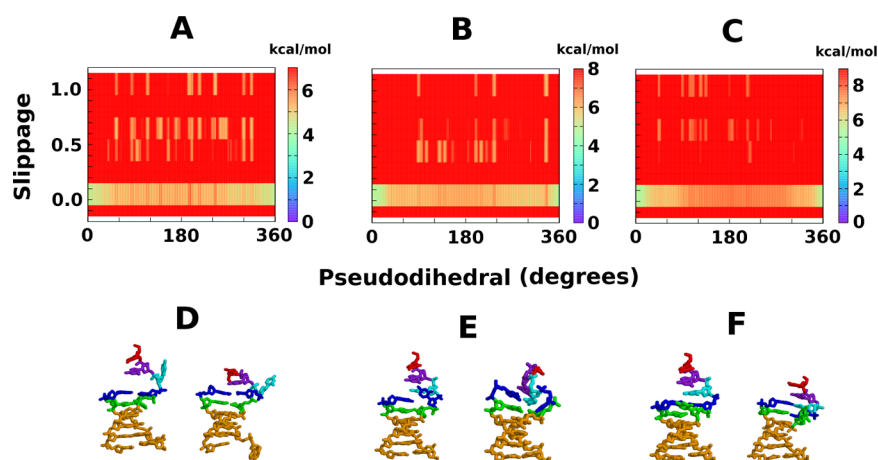
Figure 4 shows a greater population of the slippage coordinate value of 0.6 for C3 and C5 unstacking, which suggests that slippage of the immediate neighboring 5'-base occurs to a larger extent than slippage of the next base, or both 5'-bases together. For C4 unstacking, the greater population of the slippage coordinate value of 0.4 indicates that slippage of the immediate neighboring 5'-base occurs to a lesser extent than the next base, suggesting a weaker tendency for C3 to stack onto C5. As shown in Figure 4A–C and Table 1, full strand slippage (indicated by a slippage coordinate value of 1) accompanies unstacking of all three cytosines through either the minor (5° to 90° to 180°) or the major grooves (180° to 270° to 5°). The number of windows showing strand slippage are greater for C3 unstacking (18 windows) than for C4 or C5 unstacking (10 windows). For C3 unstacking, there is parity between strand slippage in the minor groove and the major

**Table 1.** Proportions of Strand Slippage in Umbrella Sampling Windows Enforcing Base Unstacking through a Pseudodihedral Restraint

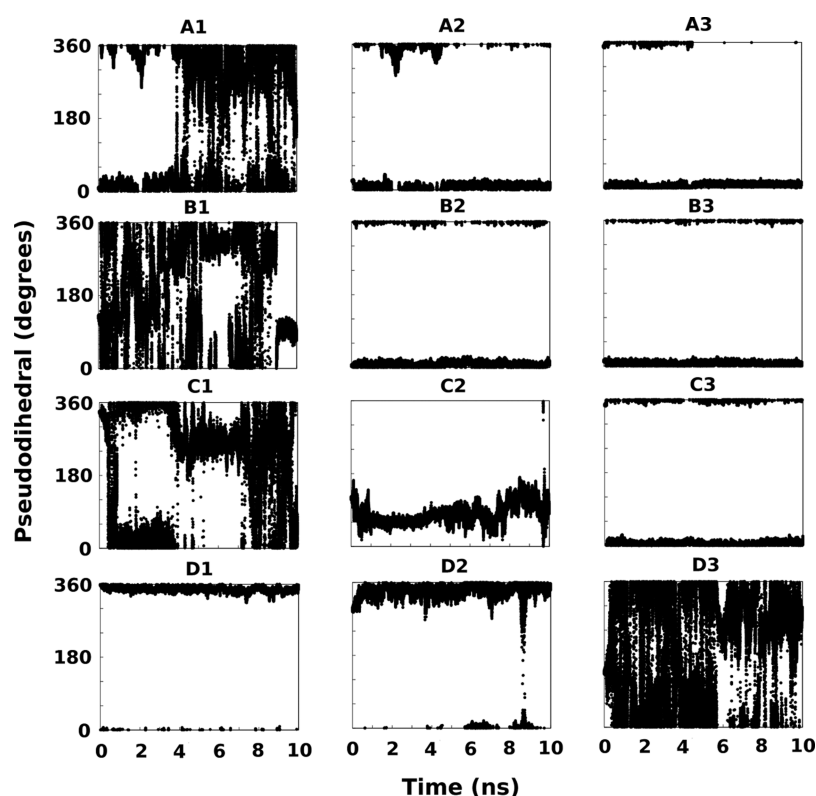
C3		C4		C5	
window <sup>a</sup>	%	window <sup>a</sup>	%	window <sup>a</sup>	%
55	16.6	90	50.2	45	1.6
65	0.0 <sup>b</sup>	95	2.4	90	4.4
85	1.0	105	9.1	105	17.2
90	1.0	145	0.0 <sup>b</sup>	110	21.3
105	0.2	155	1.7	125	35.0
115	36.2	185	0.0 <sup>b</sup>	135	19.4
135	14.4	190	0.0 <sup>b</sup>	180	2.6
155	3.0	200	28.8	215	30.5
170	0.4	235	61.3	230	0.0 <sup>b</sup>
195	81.6	335	59.2	260	0.0 <sup>b</sup>
200	25.3				
215	40.1				
225	0.2				
240	1.3				
245	83.8				
265	7.1				
300	8.1				
315	12.4				

<sup>a</sup>Pseudodihedral restraint minimum values in degrees. <sup>b</sup>Windows with non-zero but less than 0.05% sampling of strand-slipped states.

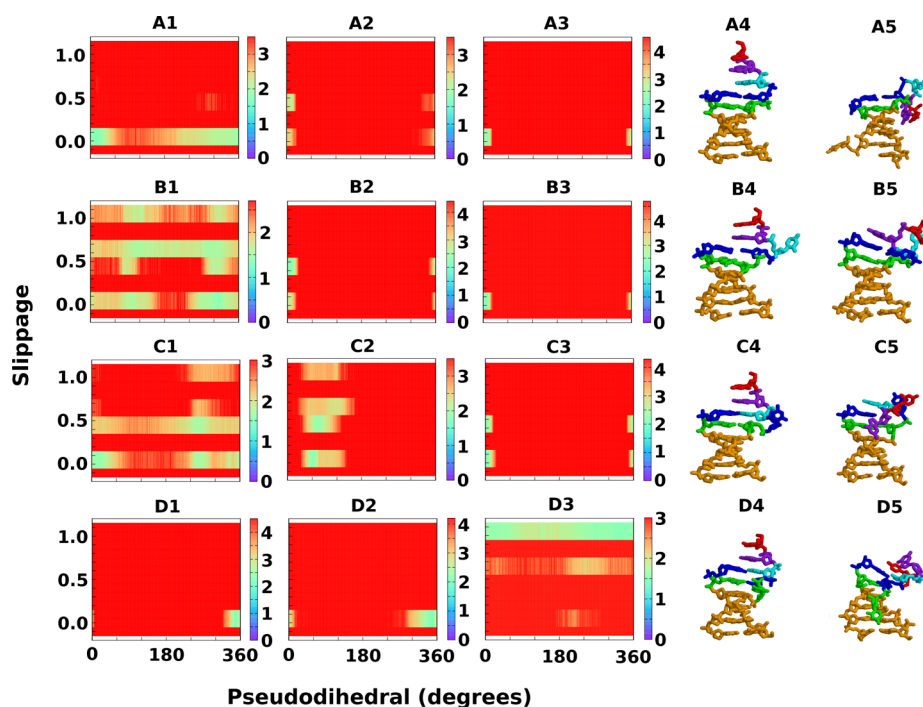
groove (9 windows each), which also exists for C4 unstacking (5 windows each). For C5 unstacking, 6 minor groove windows show strand slippage as opposed to only 3 in the major groove (of which 2 have <0.05% sampling). These observations may be due to the fact that all three cytosines are associated with a duplex terminus and an overhang and do not have the usual groove environments that they would encounter in the central regions of a DNA duplex. Especially for C3, which is beyond the duplex region, and for C4, which is just at the duplex region terminus, the number of intramolecular groove interactions is limited.



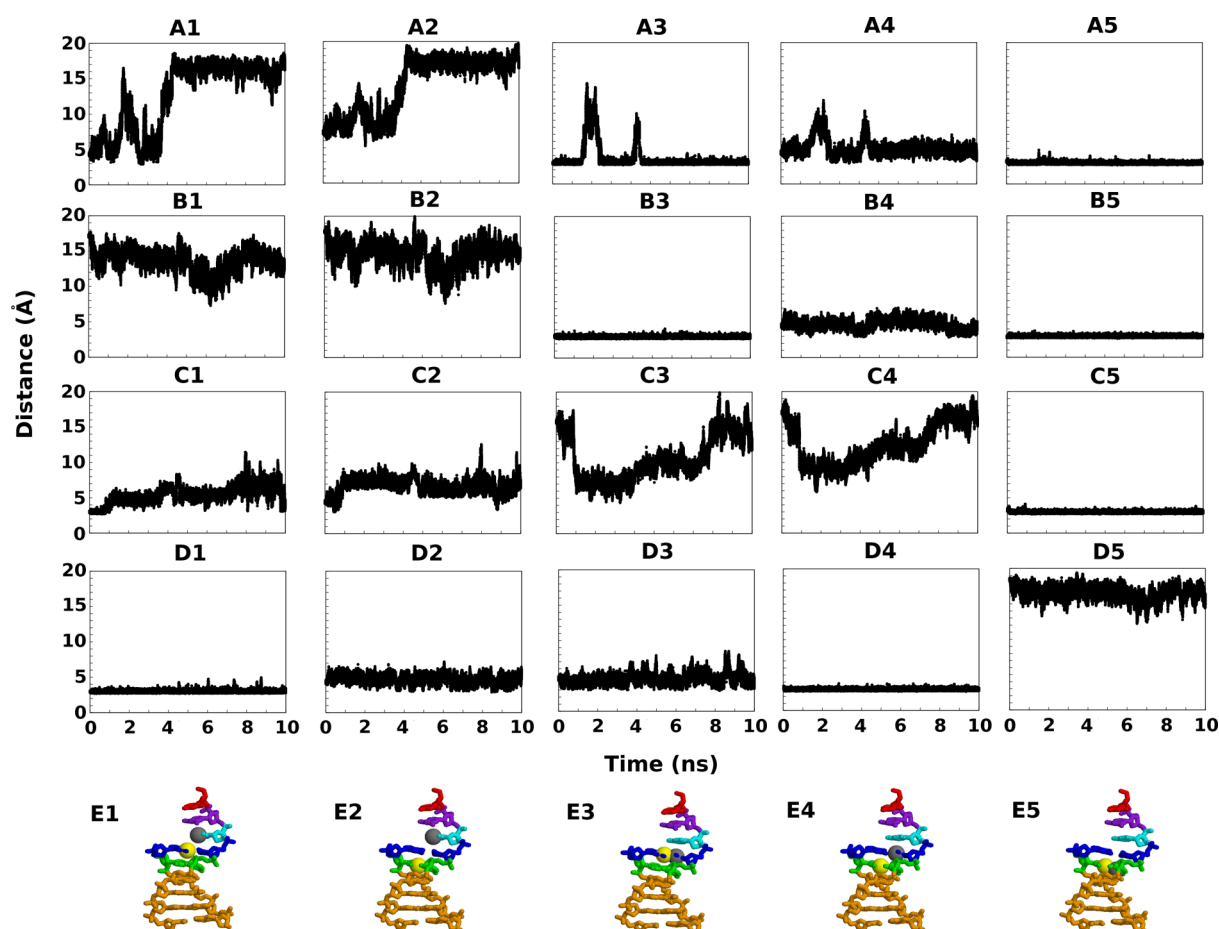
**Figure 4.** Two-dimensional free energy profiles showing the possibility of strand slippage in response to enforced base unstacking of three template cytosines in an overhang containing DNA duplex. The 2D profiles for unstacking and slippage of cytosines 3–5 are shown in panels A–C, respectively. The base unstacked geometries prior to and after strand slippage are shown in panels D–F on the left and right, respectively. The structures belong to umbrella sampling simulation windows with the following pseudodihedral values: (D) 115°, (E) 90°, and (F) 110°. The color scheme for the structures in panels D–F for the template strand is as follows: C1 in red, G2 in purple, C3 in cyan, C4 in blue, C5 in green, and the rest in orange. For the primer strand, G5 is colored green, G6 is colored blue, and the rest are colored orange.



**Figure 5.** Base unstacking pseudodihedral behavior for the three template strand cytosines in unrestrained 10 ns simulations of states I–IV shown in Figure 1. The pseudodihedral for cytosines 3–5 are shown in panels A1–D1, A2–D2, and A3–D3, respectively. Values for states I–IV are shown in panels A1–A3, B1–B3, C1–C3, and D1–D3, respectively. The definition of these pseudodihedrals is illustrated in Figure 1 of the Supporting Information.



**Figure 6.** Two-dimensional effective free energy profiles showing strand slippage and template strand cytosine base unstacking in unrestrained 10 ns simulations of states I–IV shown in Figure 1. States I–IV are shown in panels A1–A5, B1–B5, C1–C5, and D1–D5, respectively. Unstacking and corresponding slippage for C3–C5 are shown in panels A1–D1, A2–D2, and A3–D3, respectively. Panels A4–D4 show the starting conformations and panels A5–D5 the final conformations for states I–IV, respectively. The free energies indicated by the color bars are in kilocalories per mole. The color scheme in panels A4–D4 and A5–D5 is the same as in Figure 4.

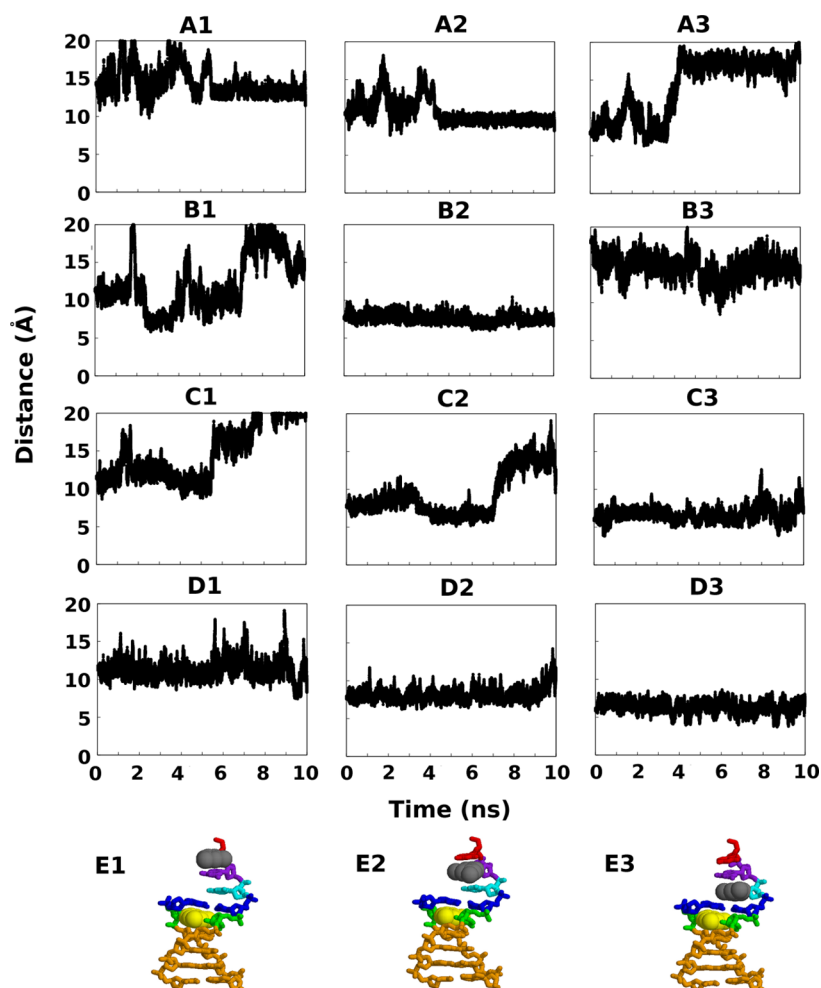


**Figure 7.** Base pairing of the three template strand cytosines in unrestrained 10 ns simulations of states I–IV shown in Figure 1. Data for states I–IV are shown in panels A1–A5, B1–B5, C1–C5, and D1–D5, respectively. The distance between the N3 atom in template strand C3 and the N1 atom in primer strand G6 is shown in panels A1–D1. The distance between the N3 atom in template strand C3 and the N1 atom in primer strand G5 is shown in panels A2–D2. The distance between the N3 atom in template strand C4 and the N1 atom in primer strand G6 is shown in panels A3–D3. The distance between the N3 atom in template strand C4 and the N1 atom in primer strand G5 is shown in panels A4–D4. The distance between the N3 atom in template strand C5 and the N1 atom in primer strand G5 is shown in panels A5–D5. These atoms are illustrated in panels E1–E5 as yellow and gray spheres, respectively. The color scheme in panels E1–E5 is the same as in Figure 4.

**Stability of Strand-Slipped States and Template Strand Overhang Dynamics.** The DNA model used in this study includes a template overhang sequence consisting of three bases (Figure 1). To assess the localized dynamics of states I–IV and this three-base template strand overhang in the absence of added restraint forces, unrestrained MD simulations were conducted on these models for 10 ns each. The base unstacking behavior of the three cytosines in the mutation hot-spot sequence during these unrestrained simulations is assessed in Figure 5 using pseudodihedral definitions shown in Figure 1 of the Supporting Information. Pseudodihedral values near  $0^\circ$  are indicative (but not necessarily confirmative) of the base being stacked on its 3'-base pair, and this is the starting point for at least two of these three cytosines in all four states. The C3 base in state II, the C4 base in state III, and the C5 base in state IV start out in their unstacked state. The C3 base, which is part of the overhang, remains unstacked in state II and tends to become unstacked in states I and III, with large fluctuations in its unstacked state geometries in these three states. It seems to remain stably stacked in state IV. The C4 base tends to remain unstacked in state III but with relatively smaller unstacked state fluctuations and tends to remain stacked in states I, II, and IV. The C5 base remains stably stacked in states I–III and remains

unstacked in state IV with large fluctuations. Even when the cytosines seem to remain stacked, they show a greater range of reversible fluctuations in the direction of the major groove ( $180^\circ$  to  $270^\circ$  to  $5^\circ$ ) than the minor groove ( $5^\circ$  to  $90^\circ$  to  $180^\circ$ ), which is consistent with the more gradual increase in free energy in the major groove in the PMF profiles (Figure 2). Except for the C4 base in state III, once a cytosine is unstacked, it seems to be able to cover the entire expanse of unstacked state pseudodihedral values on a subnanosecond time scale. This is also consistent with the relatively accessible landscape of the unstacked states seen in the PMF profiles once the major or minor groove barriers for unstacking are crossed (Figure 2). This suggests that a limited pseudodihedral range of  $\sim 140^\circ$  between and including the minor and major groove barriers probably has the strongest influence on the overall rate of base unstacking.

Figure 6 shows the effective 2D free energy profiles of strand slippage and base unstacking derived from these unrestrained simulations. Spontaneous strand slippage does not seem to occur stably in simulations started in state I (panels A1–A3), even though the C3 base is completely unstacked after 5 ns. The strand slippage in state II is also not stable as the stacking between G2 and C4 is lost (panel B1), but this does not affect



**Figure 8.** Center-of-mass distances to primer strand G5 for overhang template strand bases in unrestrained 10 ns simulations of states I–IV shown in Figure 1. Data for states I–IV are shown in panels A1–A3, B1–B3, C1–C3, and D1–D3, respectively. The distance for C1 is shown in panels A1–D1, that for G2 in panels A2–D2, and that for C3 in panels A3–D3. The base atoms used for the distances are shown as yellow and gray spheres in panels E1–E3, with the rest of the atoms colored as in Figure 4.

the stacking of its neighboring duplex region C4 and C5 bases (panels B2 and B3). The strand slippage in state III is also not fully stable (panel C2), and possible C3 base unstacking and corresponding slippage accompanying this unstacking are observed (panel C1). In contrast, the strand slippage in state IV is much more stable (panel D3) and is not accompanied by unstacking of the C4 and C3 bases (panels D1 and D2). It should be noted that a change in the slippage coordinate from 1 to 0 is indicative of departure from the slipped state, but not necessarily proper restacking. An example is the slippage coordinate value of 0 seen in the major groove near 180° for C5 unstacking in state IV (panel D3), which cannot be due to full reversal of slippage as this requires restacking of the C5 base. Unstacking of the C3 and C4 bases seems to precipitate dynamic variability in their 5'-bases, but unstacking of the C5 base does not do so. This difference is readily explained by the presence of two Watson–Crick base pairing interactions between C3:G6 and C4:G5 that stabilize the strand slippage in state IV, which prevent its loss on the 10 ns time scale.

The pseudodihedral measure of C3 unstacking in state III and C4 unstacking in state IV, in addition to being coarse-grained, is complicated by the base pair forming the first center of mass starting off disrupted. To analyze Watson–Crick base pairing directly, the N3–N1 or N1–N1 distances between

specific C:G pairs were monitored as shown in Figure 7. The C3:G6 pairing interaction (panel E1) is absent in states I and II and remains that way (panels A1 and B1, respectively). It is present in state III and could stabilize the strand slippage but is lost within 1 ns of sampling and is not regained (panel C1). It is also present in state IV and remains stable on the 10 ns time scale in this case (panel D1). The C3:G5 pairing interaction (panel E2) is not present, nor does it dynamically form, in states I–IV (panels A2–D2, respectively). The C4:G6 interaction (panel E3) is present and persistent in states I and II and is only lost twice transiently in state I for durations of <1 ns (panels A3 and B3, respectively). It is absent in states II and IV and never formed in those states (panels C3 and D3, respectively). The C4:G5 interaction (panel E4) is absent in states I–III (panels A4–C4, respectively) and is present persistently in state IV (panel D4). The C5:G5 interaction (panel E5) is present and persistent in states I–III (panels A5–C5, respectively) and is absent in state IV (panel D5). The lack of C5:G5 base pairing in state IV (panel D5) provides an example in which pseudodihedral values transiently close to 0° (Figure 5, panel D3) are not representative of restacking. These results also suggest that the most stable strand-slipped state is state IV, where C5 is unstacked, and template strand slippage



allows C3 and C4 to pair with primer strand G6 and G5, respectively.

The template strand overhang is not restrained by Watson–Crick base pairing and is therefore expected to be mobile. Its dynamics is examined using center-of-mass distances among the C1, G2, and C3 overhang bases and the primer strand G5 base in states I–IV, as shown in Figure 8. In its standard B-form orientation (starting conformation for state I), the center-of-mass distances between C1:G5, G2:G5, and C3:G5 should be around 15.2, 11.6, and 8.3 Å, respectively. A simple swinging of the overhang with little internal change in stacking is expected to maintain this distance trend, whereas an internal distortion of the overhang is not. In state I, the trend is maintained for the C1:G5 and G2:G5 distances, but not for C3:G5, which agrees with the C3 base unstacking seen in panel A1 of Figure 5. In state II, C3 starts out unstacked and remains so, and C1 also seems to become unstacked. In state III, the trend is disrupted by possible unstacking of C1, followed by G2 unstacking. In state IV, the trend is mostly maintained with transient fluctuations, which mirrors the maintenance of slippage seen in Figure 6 and its associated base pairing in Figure 7. This suggests that the stability of the strand-slipped state that accompanies C5 unstacking also extends to the overhang bases.

## DISCUSSION

Even within the confines of a polymerase active site, nucleic acid strand extension is a multistep process involving incoming nucleotide binding, conformational transitions in the enzyme, phosphoryl transfer, pyrophosphate release, and nucleic acid translocation.<sup>40,41</sup> In the midst of these is the internal dynamics of the DNA, in which the template strand is present in a duplex with the existing primer strand and as an overhang for its parts that are still to be replicated. This study provides insights into the intrinsic susceptibility of such an entity to strand slippage by examining the energetics and dynamics of base unstacking and strand slippage for an indel hot-spot sequence in an overhang-containing DNA duplex. The model used in this study places the second of the three cytosines in the hot-spot sequence (C4) at the duplex region terminus such that its 5'-cytosine (C3) is in the overhang and its 3'-cytosine (C5) is inside the DNA duplex. The PMF profiles of base unstacking for these three cytosines show that the unstacking likelihood decreases as the environment starts resembling the center of a DNA duplex (C3 > C4 > C5). Strand slippage can accompany unstacking of all three cytosine bases through the major or minor groove, albeit with a decreased frequency of strand slippage in the major groove unstacking of C5. In the strand-slipped states, the unstacked C3 and C5 bases show fast traversal of a large range of pseudodihedral values, suggesting that they can diffusively access a variety of unstacked states. Such motion is weaker for the C4 base, which is consistent with a shallow minimum in its unstacking PMF profile around 90°. Unrestrained simulations starting from four states (one non-strand-slipped and three distinct strand-slipped) suggest that spontaneous strand slippage is unlikely on a multnanosecond time scale, except in the overhang, where it is short-lived even when it does occur. On the other hand, if the strand-slipped state is achieved for cytosines within the DNA duplex, it is likely to be persistent on the multnanosecond time scale or a longer time scale, which is likely due to the base pairing interactions that form on the 5'-side of the unstacked base in the template strand. These base pairing interactions also seem to stabilize the rest of the 5'-overhang. In the context of a

polymerase active site, this increases the probability of a single-base deletion mutation because it stabilizes the wrong template base at the active site.

Starting from a correct template base at the active site of a polymerase, a single-base deletion mutation could occur due to unstacking of this base followed by strand slippage, which does not need disruption of any Watson–Crick base pairing in the overhang template. It could also occur due to unstacking of any of the paired template bases in its 3'-duplex region, followed by strand slippage, but this does require disruption of Watson–Crick base pairing, for the initial unstacking or strand slippage. These results suggest a balance between three trends that are at work the further one moves away from the active site and into the duplex region along the template strand: (1) a greater initial energetic cost of unstacking (up to a maximum unstacking cost similar to that in the center of a duplex region), (2) a greater barrier for strand slippage because of the larger number of 5'-bases that need to disrupt their base pairing to slip, and (3) a greater stabilization of the slipped state because of a larger number of rearranged 5'-base pairs. In other words, even though the initial unstacking and subsequent strand slippage can be less favored, once they occur, the strand-slipped state can be further stabilized by the increased number of pairing 5'-template bases. In this study, this trend is explicitly studied only up to the first nonterminal duplex region base, but while the increase in the cost of initial base unstacking may have a limit (~20–30 kcal/mol), both the increase in barrier height for slippage and the stabilization of the strand slippage by 5'-base pairs are likely to be progressively greater as the number of 5'-base pairs increases for the base to be unstacked.

If the barrier height for reversal of base unstacking is small (Figure 2), restacking is expected to be thermally accessible. Restacking has been observed within a few nanoseconds in MD simulations starting from an unstacked base conformation.<sup>38</sup> The barrier for strand slippage subsequent to base unstacking will be progressively higher as the number of already paired 5'-template bases increases. However, once such strand slippage occurs, each additional strand-slipped 5'-template base pair could substantially increase the likelihood of introduction of a single-base deletion mutation by increasing the barrier for reversal of slippage and restacking. A repetitive sequence in the 5'-region of an unstacking base would provide the rearranged base pairing required to stabilize the strand-slipped state, while a nonrepetitive sequence might not. The observation of stabilization of unstacking and subsequent strand slippage in the present repetitive hot-spot sequence captured in crystal structures of the Dbh Y-family polymerase<sup>13</sup> for the C5 base, and not the C4 or C3 bases, agrees with this scenario. The protein and solvent environment could also greatly influence the probability of such conformational transitions by altering the underlying energy landscapes through specific interactions. To maintain intrinsic fidelity, nucleic acid polymerases might therefore need to prevent base unstacking and strand slippage not only at the active site but also in the regions in the immediate 3'-duplex region. Future analysis of the relationship between structural architectures of nucleic acid polymerases and their fidelity needs to account for this possibility revealed by the MD simulations presented here.

## ASSOCIATED CONTENT

### Supporting Information

Three videos showing spontaneous strand slippage in response to base unstacking and one figure explaining the geometric



coordinate used to characterize base unstacking. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

\*CMS 2008, Biggs Laboratory, Wadsworth Center, NYS Department of Health, Empire State Plaza, PO Box 509, Albany, NY 12201-0509. E-mail: [nilesh.banavali@health.ny.gov](mailto:nilesh.banavali@health.ny.gov). Telephone: (518) 474-0569. Fax: (518) 402-4623.

### Funding

J.D.P. acknowledges support from National Institutes of Health Grant GM080573.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work primarily used computing resources provided by the Wadsworth Center, New York State Department of Health. Additional resources were provided through the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation Grant OCI-1053575, and CCNI housed at the Rensselaer Polytechnic Institute were also used.

## ABBREVIATIONS

indel, insertion–deletion; MD, molecular dynamics; PMF, potential of mean force; SD, steepest descent; ABNR, adopted-basis Newton–Raphson; PME, particle mesh Ewald; LJ, Lennard-Jones.

## REFERENCES

- (1) Kunkel, T. A., and Bebenek, K. (2000) DNA Replication Fidelity. *Annu. Rev. Biochem.* 69, 497–529.
- (2) Kunkel, T. (2004) DNA replication fidelity. *J. Biol. Chem.* 279, 16895–16898.
- (3) Albà, M. M. (2001) Replicative DNA polymerases. *Genome Biol.* 2, reviews3002.
- (4) Nohmi, T. (2006) Environmental stress and lesion-bypass DNA polymerases. *Annu. Rev. Microbiol.* 60, 231–253.
- (5) Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E., and Inouye, M. (1966) Frameshift mutations and the genetic code. *Cold Spring Harbor Symp. Quant. Biol.*, 77–84.
- (6) Kunkel, T., and Soni, A. (1988) Mutagenesis by transient misalignment. *J. Biol. Chem.* 263, 14784–14789.
- (7) Fujii, S., Akiyama, M., Aoki, K., Sugaya, Y., Higuchi, K., Hiraoka, M., Miki, Y., Saitoh, N., Yoshiyama, K., Ihara, K., Seki, M., Ohtsubo, E., and Maki, H. (1999) DNA Replication Errors Produced by the Replicative Apparatus of *Escherichia coli*. *J. Mol. Biol.* 289, 835–850.
- (8) Kunkel, T. (1986) Frameshift mutagenesis by eucaryotic DNA polymerases in vitro. *J. Biol. Chem.* 261, 13581–13587.
- (9) Efrati, E., Tocco, G., Eritja, R., Wilson, S., and Goodman, M. (1997) Abasic translesion synthesis by DNA polymerase  $\beta$  violates the A-rule. *J. Biol. Chem.* 272, 2559–2569.
- (10) Ling, H., Boudsocq, F., Woodgate, R., and Yang, W. (2001) Crystal structure of a Y-family DNA polymerase in action: A mechanism for error-prone and lesion-bypass replication. *Cell* 107, 91–102.
- (11) Ling, H., Boudsocq, F., Woodgate, R., and Yang, W. (2004) Snapshots of replication through an abasic lesion: Structural basis for base substitutions and frameshifts. *Mol. Cell* 13, 751–762.
- (12) Zhang, H., Eoff, R., Kozekov, I., Rizzo, C., Egli, M., and Guengerich, F. (2009) Versatility of Y-family *Sulfolobus solfataricus* DNA polymerase Dpo4 in translesion synthesis past bulky N2-alkylguanine adducts. *J. Biol. Chem.* 284, 3563–3576.

(13) Wilson, R., and Pata, J. (2008) Structural insights into the generation of single-base deletions by the Y family DNA polymerase Dbh. *Mol. Cell* 29, 767–779.

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

(15) Potapova, O., Grindley, N., and Joyce, C. (2002) The mutational specificity of the Dbh lesion bypass polymerase and its implications. *J. Biol. Chem.* 277, 28157–28166.

(16) Chi, L., and Lam, S. (2006) NMR investigation of DNA primer-template models: Structural insights into dislocation mutagenesis in DNA replication. *FEBS Lett.* 580, 6496–6500.

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

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

(19) Chi, L., and Lam, S. (2009) NMR investigation of DNA primer-template models: Guanine templates are less prone to strand slippage upon misincorporation. *Biochemistry* 48, 11478–11486.

(20) Chi, L., and Lam, S. (2012) Sequence Context Effect on Strand Slippage in Natural DNA Primer-Templates. *J. Phys. Chem. B* 116, 1999–2007.

(21) Brooks, B., Bruccoleri, R., Olafson, B., Swaminathan, S., and Karplus, M. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4, 187–217.

(22) 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. (2009) CHARMM: The biomolecular simulation program. *J. Comput. Chem.* 30, 1545–1614.

(23) Foloppe, N., and MacKerell, A., Jr. (2000) All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.* 21, 86–104.

(24) MacKerell, A., and Banavali, N. (2000) All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J. Comput. Chem.* 21, 105–120.

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

(26) Beglov, D., and Roux, B. (1994) Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations. *J. Chem. Phys.* 100, 9050–9063.

(27) Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., et al. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* 54, 905–921.

(28) Brunger, A. T. (2007) Version 1.2 of the Crystallography and NMR system. *Nat. Protoc.* 2, 2728–2733.

(29) Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An  $N \log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.* 98, 10089.

(30) Feller, S., Zhang, Y., Pastor, R., and Brooks, B. (1995) Constant-pressure molecular-dynamics simulation: The Langevin piston method. *J. Chem. Phys.* 103, 4613–4621.

(31) Banavali, N., and MacKerell, A. (2002) Free energy and structural pathways of base flipping in a DNA GCGC containing sequence. *J. Mol. Biol.* 319, 141–160.

(32) Banavali, N., and MacKerell, A. (2009) Characterizing Structural Transitions Using Localized Free Energy Landscape Analysis. *PLoS One* 4, e5525.

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

- (34) Sayle, R., and Milner-White, E. (1995) RASMOL: Biomolecular graphics for all. *Trends Biochem. Sci.* 20, 374.
- (35) Song, K., Campbell, A. J., Bergonzo, C., de los Santos, C., Grollman, A. P., and Simmerling, C. (2009) An improved reaction coordinate for nucleic acid base flipping studies. *J. Chem. Theory Comput.* 5, 3105–3113.
- (36) Leontis, N., and Westhof, E. (1998) Conserved geometrical base-pairing patterns in RNA. *Q. Rev. Biophys.* 31, 399–455.
- (37) Leontis, N., and Westhof, E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA* 7, 499–512.
- (38) Banavali, N. K. (2013) Partial base flipping is sufficient for strand slippage near DNA duplex termini. *J. Am. Chem. Soc.* 135, 8274–8282.
- (39) Banavali, N., and Roux, B. (2005) Free energy landscape of A-DNA to B-DNA conversion in aqueous solution. *J. Am. Chem. Soc.* 127, 6866–6876.
- (40) Dahlberg, M. E., and Benkovic, S. J. (1991) Kinetic mechanism of DNA polymerase I (Klenow fragment): Identification of a second conformational change and evaluation of the internal equilibrium constant. *Biochemistry* 30, 4835–4843.
- (41) Joyce, C. M. (2010) Techniques used to study the DNA polymerase reaction pathway. *Biochim. Biophys. Acta* 1804, 1032–1040.