

Base Flipping of the Thymine Dimer in Duplex DNA<sup>§</sup>Lauren L. O'Neil,<sup>†</sup> Alan Grossfield,<sup>‡,||</sup> and Olaf Wiest\*,<sup>†</sup>

Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, Indiana 46556-5670, and IBM T. J. Watson Research Center, Yorktown Heights, New York 10598

Received: May 24, 2007; In Final Form: July 26, 2007

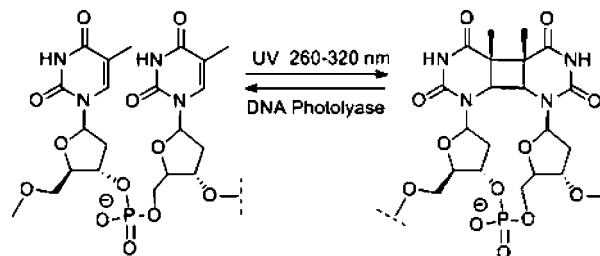
Exposure of two adjacent thymines in DNA to UV light of 260–320 nm can result in the formation of the *cis,syn*-cyclobutane pyrimidine dimer (CPD). The structure of DNA containing an intrahelical CPD lesion has been previously studied experimentally and computationally. However, the structure of the extrahelical, flipped-out, CPD lesion, which has been shown to be the structure that binds to the CPD repair enzyme, DNA photolyase, has yet to be reported. In this work the structure of both the flipped-in and the flipped-out CPD lesions in duplex DNA is reported. These structures were calculated using 8 ns molecular dynamics (MD) simulations. These structures are then used to define the starting and ending points for the base-flipping process for the CPD lesion. Using a complex, two-dimensional pseudodihedral coordinate, the potential of mean force (PMF) for the base-flipping process was calculated using novel methodology. The free energy of the flipped-out CPD is roughly 6.5 kcal/mol higher than that of the flipped-in state, indicating that the barrier to flipping out is much lower for CPD than for undamaged DNA. This may indicate that the flipped-out CPD lesion may be recognized by its repair enzyme, DNA photolyase, whereas previous studies of other damaged, as well as nondamaged, bases indicate that they are recognized by enzymes in the intrahelical, flipped-in state.

## Introduction

The movement of a DNA base from the base-stacked, hydrogen-bonded, intrahelical position to a solvent-exposed, extrahelical position is termed “base flipping”. Since flipped-out bases are more accessible to solvent or other molecules and are more prone to complexation with proteins, it is not surprising that most enzymatic DNA modification and repair reactions involve base flipping. Crystal structures of enzyme–DNA complexes in which one or more bases are flipped out of the duplex have been reported for many enzymes including M. HaeI, M. HaeIII, hOGG1, and T4 endonuclease.<sup>1–4</sup> An enzyme of particular interest is DNA photolyase.<sup>5,6</sup> DNA photolyases catalyze the photoreversion of *cis,syn*-cyclobutane pyrimidine dimers (CPDs), which are formed via a photochemical [2 + 2] reaction between two adjacent thymines in DNA as shown in Figure 1.<sup>7</sup>

A recent crystal structure of *Anacystis nidulans* DNA photolyase in complex with a model of the photodamaged DNA substrate shows that the lesion site is flipped out into the enzyme active site.<sup>8</sup> This was also suggested by earlier experimental work, computational models, and the crystal structure of *Escherichia coli* DNA photolyase without bound substrate.<sup>9–12</sup>

The CPD lesion, also known as the thymine dimer, introduces significant disruption to the DNA helix. The results from a 500 ps molecular dynamics (MD) simulation performed by Miaskiewicz et al. showed a disruption of the hydrogen bonding of the 5' thymine to its complementary adenine base.<sup>13</sup> The hydrogen bond distance was increased to 2.5 Å (N–H···N),



**Figure 1.** Formation of the *cis,syn*-cyclobutane pyrimidine dimer (CPD).

and the observed hydrogen bond angle was 125°. Another interesting feature of this structure is that the tilt and roll of the thymine bases that make up the thymine dimer are such that the base stacking is disrupted (i.e., bases are not parallel). Also, the inclusion of a thymine dimer in a DNA duplex is found to induce a bend or kink in the DNA helix. The kink angles calculated from an X-ray crystal structure of a thymine dimer containing DNA decamer and from the average structure resulting from an 800 ps MD simulation of the thymine dimer containing DNA were 30° and 22.3°, respectively.<sup>14,15</sup> In comparison, the kink angle calculated for the average structure of a nondamaged DNA duplex from an 800 ps MD simulation was found to be 8.2°. The changes in the DNA structure due to the thymine dimer have been postulated to aid in the recognition of the thymine dimer by DNA photolyases, possibly by destabilizing the DNA structure in comparison with nondamaged DNA. This destabilization could allow for the thymine dimer to undergo spontaneous base flipping, a phenomenon that is very rare for nondamaged bases.

There exist two major pathways for recognition of a damaged base: base flipping followed by binding and simultaneous base flipping and binding. In work by Verdine and co-workers on an 8-oxo-G repair enzyme, hOGG1, the temporal coupling of base flipping and enzyme recognition was studied.<sup>16</sup> For hOGG1

\* Corresponding author. E-mail: owiest@nd.edu.

<sup>†</sup> University of Notre Dame.

<sup>‡</sup> IBM T. J. Watson Research Center.

<sup>§</sup> CPD, *cis,syn*-cyclobutane pyrimidine dimer; MD, molecular dynamics; PMF, potential of mean force; WHAM, weighted histogram analysis method.

<sup>||</sup> Present address: University of Rochester Medical Center, Department of Biochemistry and Biophysics.

there is the need to recognize and repair a DNA damage that occurs in very low numbers compared to nondamaged bases. It was determined that the enzyme uses "fast sliding" along the DNA backbone and interrogates possible damaged bases in an additional pocket on the surface, allowing for the nonproductive recognition events to be fast and require small activation energies.<sup>17</sup> Although no definitive studies have been reported, it seems likely that DNA damage such as 8-oxo-G is unlikely to induce spontaneous base flipping because the damaged base has only a one atom change compared with the native base and the flipped-in conformation is virtually unchanged. Undamaged DNA bases which are also very unlikely to undergo spontaneous base flipping may also be recognized by DNA modification enzymes in this manner. By comparison, the conformation of the thymines in the thymine dimer is significantly changed compared to the nondamaged bases. This may influence the energetics of the base-flipping process and allow for DNA photolyases to recognize an extrahelical thymine dimer. Irrespective of the order of base flipping and binding, the energy required for the thymine dimer to undergo base flipping has implications for the repair by DNA photolyases. If the enzyme induces base flipping of the thymine dimer, then the energy required to do so would need to be provided by the enzyme.

Spontaneous base flipping of DNA bases had been previously studied experimentally by Dornberger et al.<sup>18</sup> Using imino proton exchange, NMR measurements of the equilibrium constant for base flipping of a guanine base in a GCGC tetramer was reported to be  $3.3 \times 10^{-7}$ , which corresponds to a free energy difference of  $\sim 9$  kcal/mol. Other experimental studies of base flipping have relied on the use of 2-aminopurine (2-ap), a nonnative DNA base that is synthetically incorporated into a DNA strand. When base stacked, the fluorescence of 2-ap is quenched, but upon base flipping the fluorescence increases.<sup>19–21</sup> There have been many other experimental studies of base flipping as it relates to enzymatic repair and modification reactions which also exploited the fluorescence behavior of 2-ap.<sup>9,22,23</sup> A persistent question in these studies has been the perturbation of the DNA structure by the nonnative 2-ap. Furthermore, 2-ap, while a close mimic to cytosine and thymine, is unable to mimic the behavior of the thymine dimer and other DNA bases, both native and nonnative. While these studies have provided insight into base flipping in the context of enzymatic reactions, they provide little insight into spontaneous base flipping. An alternative approach, which uses a small molecule to detect the flipped-out conformation of the DNA base, requires no modification of the DNA duplex and has been used to detect base flipping in free solution.<sup>24</sup>

Computational studies of base flipping have primarily focused on nondamaged DNA using potential of mean force (PMF) calculations. For a recent review of the previous computational studies of base flipping in DNA, see ref 25. The first study of the energetics of base flipping was performed by Guidice et al.<sup>26</sup> The potential of mean force for base flipping of an adenine and its base pair thymine were computed and the  $\Delta G$  values were 15 and 13 kcal/mol, respectively. In a later study, the energetics of base flipping of cytosine and guanine bases were computed and the  $\Delta G$  value for those processes was found to be 16 kcal/mol for both bases.<sup>27</sup> A study performed by Banavali et al. used a center-of-mass pseudodihedral coordinate to compute the PMF of the base flipping of the cytosine residue in the GCGC sequence recognized by M. HhaI.<sup>28</sup> The free energy required for flipping of the cytosine base was determined to be 15.3 and 17.6 kcal/mol for flipping through the major and minor grooves, respectively. For the base-pairing partner

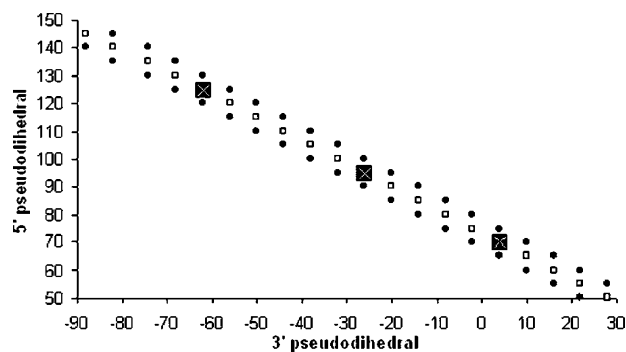
guanine the base-flipping free energy was determined to be 21.3 and 18.7 kcal/mol for flipping through the major and minor grooves, respectively.

This paper reports the calculation of the potential of mean force for the spontaneous base flipping of a thymine dimer in duplex DNA. These calculations use a pseudodihedral coordinate which has been recently reported in the literature and was used to compute a one-dimensional PMF. This pseudodihedral coordinate was extended to a two-dimensional coordinate, which is needed for the description of the base flipping of the CPD.<sup>29</sup> We will first discuss the methods used to determine the structures of the flipped-in and flipped-out conformations of the thymine dimer containing DNA. These structures will then be used to identify a coordinate and compute the two-dimensional (2D) PMF for the flipping process. Finally, we will discuss the energetics of the flipping process.

## Methods

All molecular dynamics simulations were performed using the Amber 8 suite of programs and the Cornell et al. force field with the adjustments added by Wang et al.<sup>30–32</sup> The parameters for the thymine dimer are those used by Spector et al.<sup>15</sup> This force field, like most others, has been parametrized to reproduce the canonical forms of DNA. The Amber force field has previously been used to study the flipped-in form of the thymine dimer but not the flipped-out state.<sup>15</sup> However, the use of this force field in the study of the noncanonical flipped-out structure is also appropriate. In a comparative study of the CHARMM, Amber, and Bristol-Myers Squibb (BMS) force fields in the reproduction of the experimentally determined base-pair opening rate of a GC Watson–Crick base pair, it was found that the Amber and CHARMM force fields were in good agreement with the available experimental data.<sup>33</sup>

The DNA structures, both flipped-in and flipped-out, were prepared using Insight II.<sup>34</sup> The flipped-out structures were prepared by manually rotating the base(s) of interest out of the DNA helix. The DNA was neutralized using  $\text{Na}^+$  counterions and solvated using the TIP3P water model as provided in xleap. The solvent box extended 8 Å beyond the DNA structure in each direction. The final systems consisted of 11 583 and 14 277 atoms including 3513 and 4410 water molecules and 30 sodium ions for the flipped-out and flipped-in simulations, respectively. The systems were minimized for 60 000 steps, first with restraints placed on the DNA heavy atoms to allow the water box and hydrogen atoms to equilibrate. A second round of minimization with restraints on the two base pairs to both the 5' and 3' sides of the dimer, but not on the adenine residues opposite the dimer, was conducted. Finally, a third round of minimization with no restraints was performed. The minimized system was then equilibrated in the constant-volume, isothermal (NVT) ensemble for 20 ps with restraints of 10 kcal/mol·Å<sup>2</sup> on the DNA. The system was then heated to the final temperature of 300 K over 100 ps in the constant-pressure, isothermal (NPT) ensemble with restraints of 10.0 kcal/mol·Å<sup>2</sup> on the two base pairs to both the 5' and 3' sides of the dimer, but not on the adenine residues opposite the dimer. Isotropic position scaling with a relaxation time of 2 ps was used to maintain a pressure of 1 atm, and Langevin dynamics with a collision frequency of 1.0 ps<sup>-1</sup> was used to maintain the temperature at 300 K. Over the next 300 ps, the restraints on the bases to the 5' and 3' sides of the dimer were removed, starting from the residues farthest from the dimer and moving closer. After this period of equilibration (420 ps) the production MD simulations (8 ns) were run. After 8 ns of simulation of the flipped-out structure,



**Figure 2.** Location of windows used to generate potential of mean force.

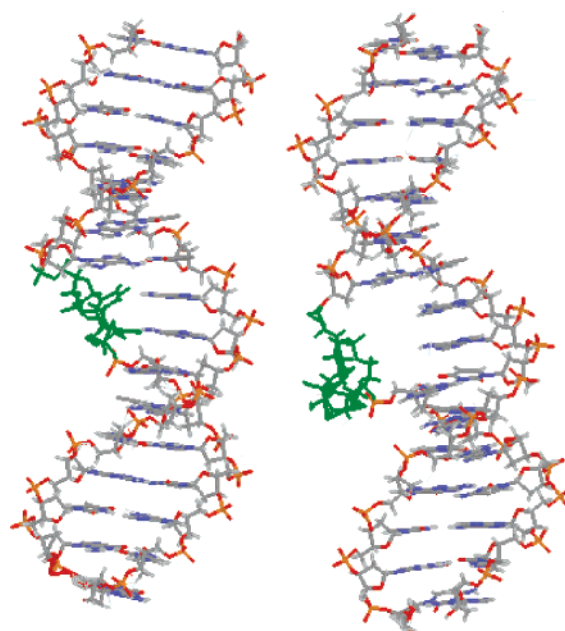
fraying of the 3' bases was observed. All calculations used SHAKE to constrain covalent bonds to hydrogen, which allowed the use of a 0.002 ps time step. Long-range electrostatic interactions were treated using PME with long-range cutoffs of 10 Å applied to the Lennard-Jones interactions.<sup>35</sup> Periodic boundary conditions were used in all calculations. Curves analysis was performed on 100 structures that were output every 80 ps from the trajectory files of the MD simulations.<sup>36</sup>

All simulations were analyzed using the ptraj module of Amber 9. The values of the two pseudodihedral angles used to define the base-flipping coordinate, as shown in Figure 5, were output as a function of time for the 8 ns production runs. These data were used to create a histogram, as shown in Figure 6. The modes of the data as shown in Table 1 were used as the starting and ending points for the flipping process.

In order to generate the potential of mean force, umbrella sampling was performed using restrained (0.05 kcal/mol·deg) harmonic potentials with the values of the pseudodihedral for each window as shown in Figure 2. The open squares show the flipping coordinate in which each window was equilibrated for 100 ps followed by 2.5 ns of sampling. The sampling for each of these windows was started from one of the extensively equilibrated structures. The large filled squares show the extensively equilibrated structures in which each window was equilibrated for 3 ns followed by 2.5 ns sampling. The filled squares show the sampling extensions used to move high-energy artifacts away from flipping coordinate; each window was equilibrated for 100 ps followed by 2.5 ns sampling.

The values of the 5' pseudodihedral move in 5° steps, and the 3' pseudodihedral moves in 6° steps. For the windows at (70°, 4°), (95°, -26°), and (125°, -62°), where the angles are in 5', 3' pseudodihedral pairs, the simulations were started from the last frame of the equilibrium simulations which had been stripped of waters and ions, resolvated and neutralized, and equilibrated for 100 ps (NVT) and 500 ps (NPT). Each of these windows was then equilibrated for 3 ns. These results of the equilibration of these windows were then used to start the equilibration for neighboring windows, which were done independently. In this way, the risk of hysteresis is minimized because each simulation does not depend on the one before it, as has been done in previous simulations.<sup>21–23</sup> Each window was equilibrated for 100 ps and sampled for 2.5 ns. The total simulation time used to generate the PMF was 159.5 ns.

The values of the pseudodihedrals were saved every 0.2 ps. The unbiased free energy was obtained by using the two-dimensional weighted histogram analysis method as implemented by Grossfield. The convergence criterion was 0.0001, and the data were placed into bins of 0.5° and 0.6° for the 5' and 3' pseudodihedrals, respectively. Convergence of the free



**Figure 3.** Average structures (8 ns) of flipped-in (left) and flipped-out (right) thymine dimer containing DNA. The thymine dimer is shown in green.

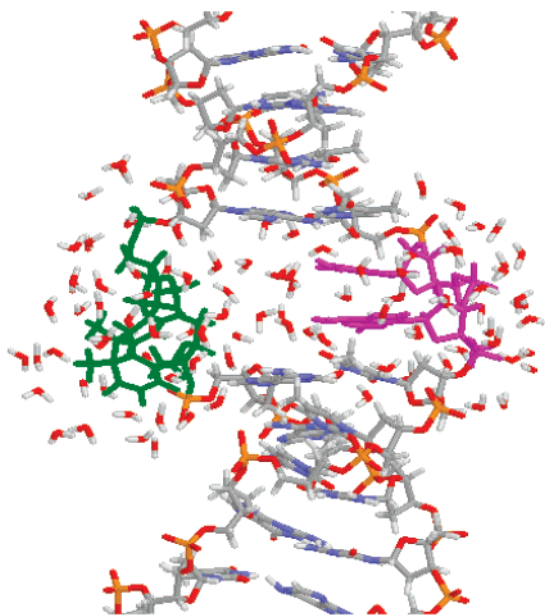
energy curves was tested using various equilibration and sampling times as shown in the Supporting Information.

## Results and Discussion

**Definition of Flipped-In and Flipped-Out States.** In order to compute the free energy needed for the thymine dimer to undergo base flipping, both the flipped-in and flipped-out states had to be properly characterized. In previous studies of base flipping, the coordinate used to define base flipping was usually defined based upon the conformation of the flipped-in DNA only. In the more complex CPD system, where formally two bases undergo base flipping, it was necessary to first define the starting and ending points of the simulation. This was done using MD simulations of the DNA duplex 5'-GCACGAAT < > TAAGCACG-3', where T < > T signifies the thymine dimer. The corresponding strand has the Watson–Crick base-pairing partners with the thymine dimer base pairing with two adenine bases. Both the flipped-in and flipped-out structures were run for 8 ns (production) after 420 ps of equilibration. The root-mean-square deviations (RMSDs) of both the flipped-in and flipped-out structures over the 8 ns production runs are shown in the Supporting Information. The CPD in the flipped-out conformation remains flipped-out over the course of the 8 ns simulation as evidenced by the  $0.5 \pm 0.21$  Å average RMSD of the CPD as well as snapshots at different time points along the simulation (see Supporting Information). The average structures are shown in Figure 3.

As previously mentioned, the inclusion of a thymine dimer in duplex DNA causes significant distortions in both the local structure and the global structure of the duplex. Base pairing of the thymine dimer to the opposite adenine bases in the flipped-in structure is disrupted due to the angle required by the cyclobutane ring. The base stacking is also disrupted as the thymine bases are no longer parallel to those above and below. Finally, there is a large bend in the overall DNA strand. Using Curves analysis, the average global bend, over 8 ns of simulation time, was found to be 30.8° and 34.4° for the flipped-in and flipped-out structures, respectively.<sup>36</sup> The global bend of the corresponding nondamaged DNA structure from a 2 ns MD



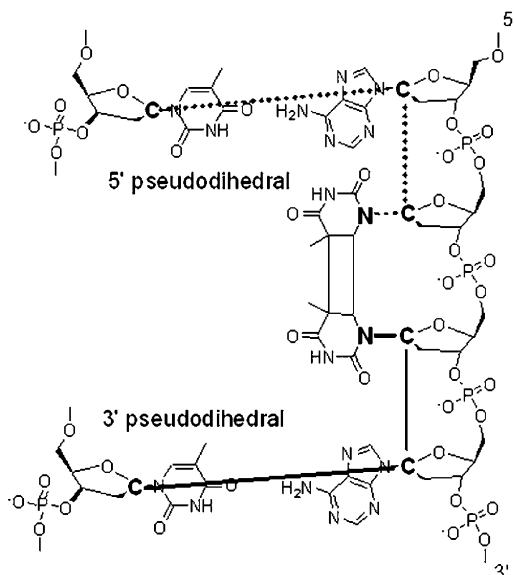


**Figure 4.** Closeup of the flipped-out structure showing the 50 waters closest to adenines 24 and 25 (opposite the two thymines of the thymine dimer) from 2000 to 4000 ps of the MD production run. Thymine dimer is highlighted in green.

simulation was found to be  $7.2^\circ$  (at 1200 ps). Analysis of the local geometry of the flipped-out structure using Curves in ways similar to those employed in previous studies does not give useful structural information. For example, the helicoidal parameters such as tilt, roll, and twist have no meaning for the open base pairs of the flipped-out structure and, as noted in the previous study, the distortions of the DNA structure occur primarily at the damaged site.<sup>15</sup> Extrahelical bases can adopt conformations that are not feasible for intrahelical bases, making the structural analysis of such structures difficult using Curves parameters.

The structures of CPD-containing DNAs have been previously determined both experimentally and computationally.<sup>14,15,37</sup> As previously mentioned, the global bend of the crystal structure of a CPD containing DNA duplex was  $30^\circ$ .<sup>14</sup> This bending angle is in close agreement with the  $27^\circ$  bend reported in a previous computational study of the CPD-containing DNA.<sup>37</sup> The structure that would be most comparable to our calculated structure is that of Spector et al., in which the global bend is reported to be  $22.3^\circ$ .<sup>15</sup> This structure looks qualitatively similar to our structure, but no direct comparison can be made because the coordinates of that structure are not available. The difference in the calculated global bend for the two structures may be due to the large difference in total simulation time, 800 ps vs 8 ns. The calculated structure of the flipped-in DNA, as shown in Figure 2, agrees well with previous structural studies of CPD-containing DNA and is a good starting point for the calculation of the PMF for base flipping.

The structure in which the thymine dimer is flipped out of the duplex has not yet been studied in the literature, even though it is clear that in complex with DNA photolyase, the dimer is flipped out. Base flipping obviously disrupts the base pairing and stacking interactions which ordinarily stabilize double helical DNA. Presumably, the thymine dimer compensates by replacing the hydrogen bonds to adenine with hydrogen bonds to one or more waters. As shown in Figure 4, there are waters positioned to hydrogen bond with the extrahelical thymine dimer and with the orphaned adenine bases. The percent occupancies of water(s) within  $5 \text{ \AA}$  of the N1 nitrogen of the orphaned

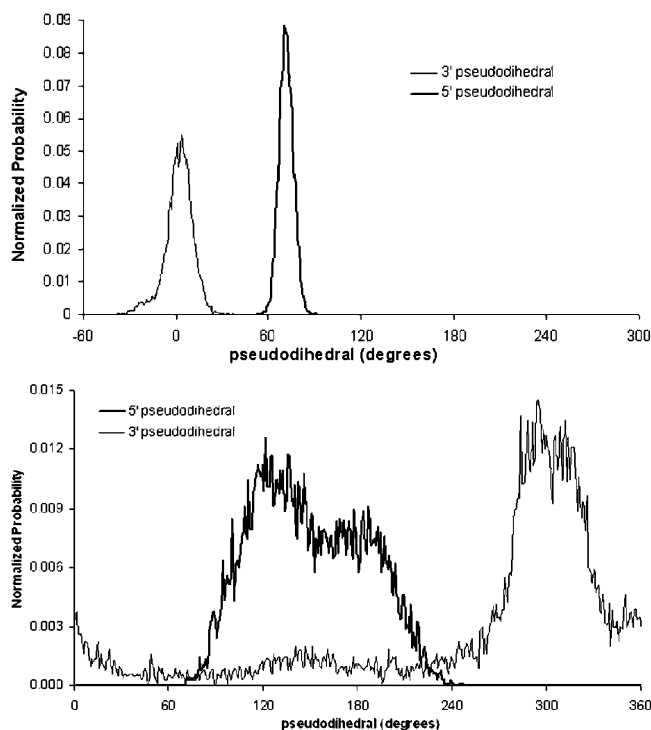


**Figure 5.** Two-dimensional pseudodihedral coordinate used to define the flipping process. Each pseudodihedral connects the N1 and C1' of the flipping base with the C1' of the adjacent base and the C1' of the base-pairing partner of the adjacent base.

adenines were found to be 122.75% and 140.54% for the 5' and 3' adenines, respectively (see Supporting Information for full analysis). This indicates that more than one water occupies the "hole" left by the CPD over the 8 ns simulation time. Also, the adenine bases opposite the thymine dimer may hydrogen bond with other bases 3' and 5' to the thymine dimer or may make contact with the DNA backbone of the thymine dimer, which has been found to be quite flexible during the MD simulations. This was also observed in the previous study of a DNA duplex containing an apurinic site in which the orphan base hydrogen bonds with neighboring bases.<sup>38</sup> The comparison of the crystal structure of DNA photolyase in complex with DNA in which the CPD is flipped out into the active site of the enzyme is limited due to the modification of the DNA that was necessary to acquire the crystal structure and the structural distortions induced by binding to the enzyme.<sup>8</sup>

**Definition of Flipping Coordinate.** The coordinate used to define the flipping process was similar to that used by Barthel and Zacharias.<sup>29</sup> In that study, the base flipping of uridine bulges in RNA was explored using a one-dimensional pseudodihedral. In our case, flipping of the thymine dimer involves two bases which are covalently linked and the motion of the two bases are, therefore, correlated. It is expected to require a more complex coordinate as shown in Figure 5.

The flipped-in and flipped-out states of the thymine dimer were defined using the pseudodihedrals shown in Figure 5. The values of the pseudodihedrals were saved every 1 ps over the course of the 8 ns MD simulations of the flipped-in and flipped-out DNAs to define the starting and ending points for the PMF. This was done in order to eliminate the uncertainty that arises from arbitrarily choosing an end point for the PMF, which would result in incorrect energy values. The values of the two pseudodihedrals for the flipped-in and flipped-out DNAs were used to construct the histograms shown in Figure 6. The modes of the histograms are listed in Table 1. It can be seen that the histograms for the flipped-in DNA have narrower distributions than those of the flipped-out DNA, indicating that the pseudodihedral angles are more flexible in the flipped-out structure. The mode of the 5' pseudodihedral distribution of the flipped-out DNA is  $122^\circ$ , but there is also a significant peak at  $180^\circ$ . The



**Figure 6.** Normalized probability histograms of the 5' (bold) and 3' (thin) pseudodihedrals over 8 ns equilibrium MD simulations of flipped-in (top) and flipped-out (bottom) DNAs. Data were saved every 1 ps. Histogram was created using  $1^\circ$  bin widths. Note that the normalized probability scales of the two plots shown are different.

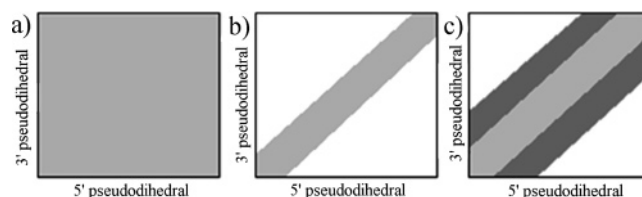
**TABLE 1.**<sup>a</sup>

system	pseudodihedral	
	5'	3'
flipped-in	71°	4°
flipped-out	122°	-65°

<sup>a</sup> The values given are the mode of histograms created using the values of the pseudodihedral angles output from the 8 ns production MD simulations. Data were saved every 1 ps.

structures that correspond to the 5' pseudodihedral angle of  $180^\circ$  differ from those at  $\sim 120^\circ$  in that the thymine dimer is flipped out and tilted to face the major groove.

**Calculation of the Potential of Mean Force.** The relative free energies of different states can be inferred directly from their relative probabilities, and as a result can in principle be computed directly from a molecular dynamics trajectory. However, in practice the existence of significant barriers slows sampling such that these quantities do not readily converge except under specific circumstances. For this reason, specialized techniques such as umbrella sampling have been developed. In this approach, the single trajectory is replaced by a series of simulations, each one biased to explore a particular region of the chosen reaction coordinate (usually, a distance or torsion angle). In this manner, one can guarantee that all relevant portions of phase space are explored, while greatly reducing the simulation time required for convergence. The potential of mean force (PMF), or relative free energy curve, can be estimated by combining the probability distributions from the biased trajectories using the weighted histogram analysis method (WHAM).<sup>39–41</sup> In the present circumstance, the relevant motions of the system are not well described by a single reaction coordinate. Rather, two separate pseudodihedral angles are necessary to adequately capture the transition mechanism. However, unlike the standard case for a two-dimensional free



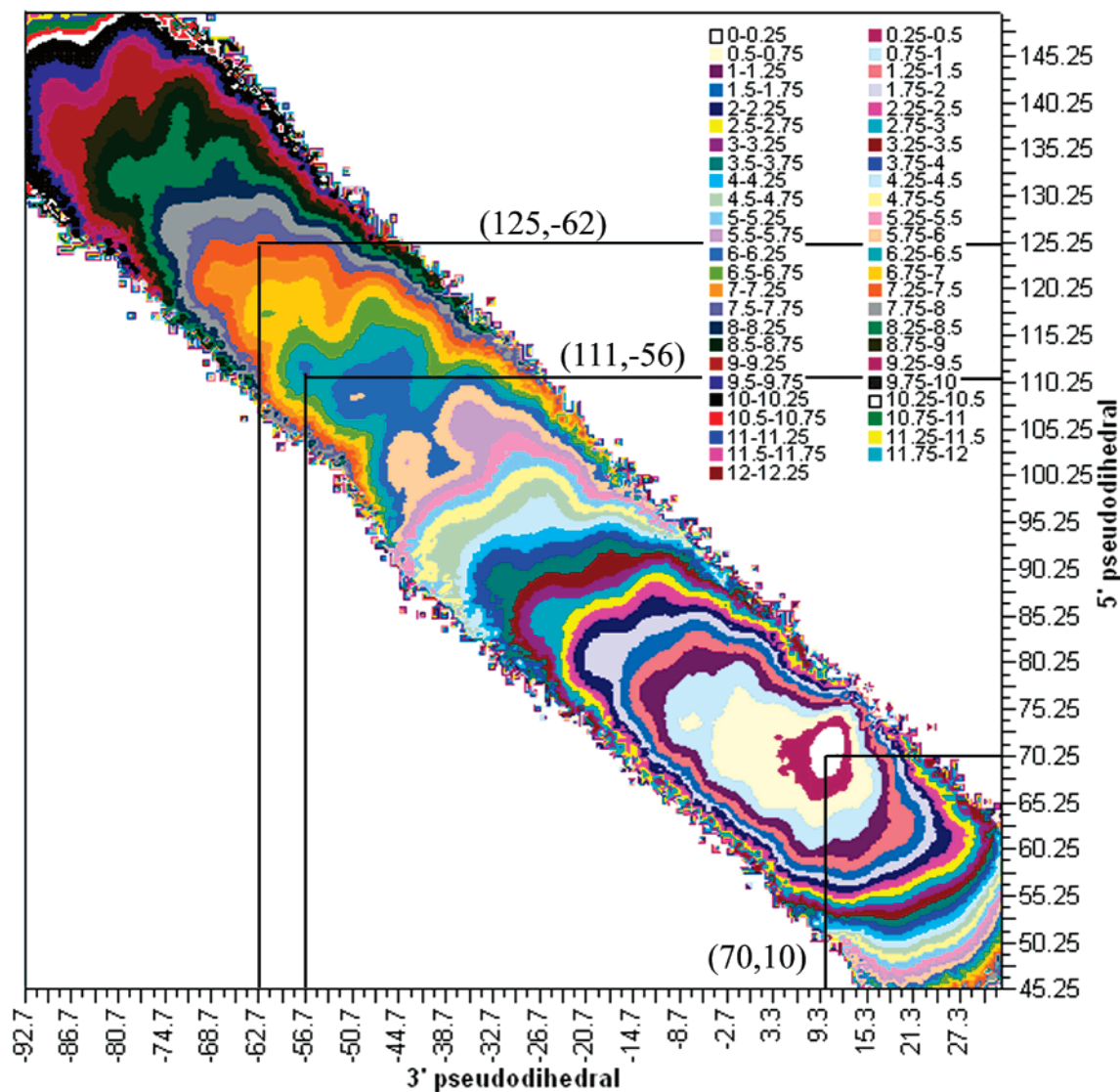
**Figure 7.** Schematic representations of sampling used to generate PMF. Sampling areas are shaded gray, and nonsampled regions are white. (a) Full sampling of 2D coordinate. (b) Sampling only the physically reasonable region. (c) Extended sampling in which physically reasonable region (light gray) is bordered by a high-energy region (dark gray).

energy calculation, where all ranges of both variables are sampled (Figure 7a), the pseudodihedral angles are coupled through the cyclobutane ring linking the two bases. As a result, only a small portion of the free energy surface is physically accessible (Figure 7b). For purposes of efficiency, we restricted our sampling to these regions and those immediately surrounding them (Figure 7c).

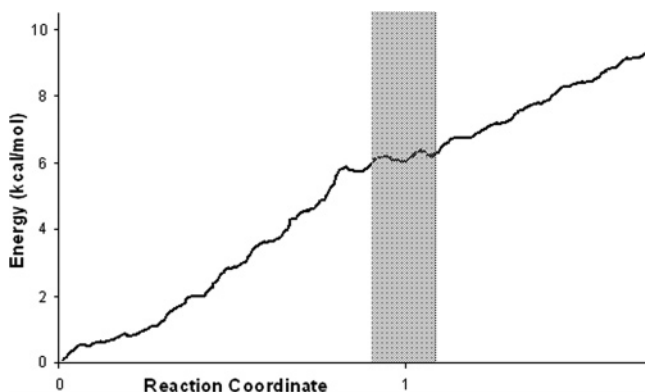
This analysis is somewhat complicated by the fact that WHAM operates on probability distributions, with the result that unsampled regions are assigned infinite free energies. Because WHAM is in effect a global fit, the presence of these infinities in physically irrelevant regions could distort the biologically critical portions of the PMF. For this reason, we used the novel sampling scheme illustrated in Figure 7c. The phase space of interest, also shown in Figure 7b, has been extended outward into the less physically realistic regions of the 2D pseudodihedral coordinate. This scheme removes the discontinuity by including a well-defined high-energy area adjacent to the area of interest.

The PMF was calculated by moving from  $50^\circ$  to  $145^\circ$  in  $5^\circ$  increments for the 5' pseudodihedral and from  $28^\circ$  to  $-88^\circ$  in  $6^\circ$  increments for the 3' pseudodihedral as shown in Table 1. The flipped-in state is defined at  $(70^\circ, 4^\circ)$  and the flipped-out state at  $(125^\circ, -62^\circ)$  with 5' pseudodihedral and 3' pseudodihedral pairs, respectively. The sampling was then performed as illustrated in Figure 7c. The final PMF, shown in Figure 8, was calculated using 2D WHAM as implemented by Grossfield using 2.5 ns sampling at each point.

The lowest energy region of the contour diagrams of the PMF is centered at  $(70^\circ, 10^\circ)$ , corresponding to the structure of the flipped-in state. The energy rises in all directions from this point, both toward and away from the flipped-out state. The previously defined flipped-out state at  $(125^\circ, -62^\circ)$  (located in the dark orange region) is 7–7.25 kcal/mol above the flipped-in state. These dihedral angle values are the modes of the data as shown in Table 1. These dihedrals were used to designate an area of phase space in which the flipped-out conformation exists. This definition of the flipped-out state is not necessarily a single structure but is an estimate of the region of the phase space because the flipped-out structure is considerably more flexible, as previously discussed. There is a lower energy structure at  $(111^\circ, -56^\circ)$  (located in the dark blue region) which is 6.25–6.5 kcal/mol above the flipped-in state, which is closer to the actual flipped-out state. A one-dimensional projection of the lowest energy pathway from the flipped-in to the flipped-out structures is shown in Figure 9. The flipped-in state, shown as the 0 point on the reaction coordinate, is at 0 kcal/mol and the energy rises to a small plateau before reaching the flipped-out state, shown as 1 on the reaction coordinate, after which the energy quickly rises once again. The limitation of using a 1D projection of a 2D surface is that information can be lost because the data are not as well represented as they are in the full 2D surface.



**Figure 8.** Contour map of the potential of mean force for the base flipping of thymine dimer. Free energy as a function of 5' and 3' pseudodihedrals. Each color represents a 0.25 kcal/mol change in energy as shown in the legend. The points indicated on the chart with cross hairs are also labeled with the corresponding pseudodihedral pairs (5',3').



**Figure 9.** One-dimensional projection of the lowest energy pathway connecting the flipped-in and flipped-out states. The complex coordinate is simplified to a reaction coordinate in which 0 is defined as the flipped-in state and 1 is defined as the flipped-out state. The highlighted region shows the plateau that is the flipped-out state.

### Conclusions

The energy required for the thymine dimer to undergo base flipping in the system studied is 6.25–6.5 kcal/mol, which corresponds to an equilibrium constant for the flipping process

on the order of  $5 \times 10^{-5}$ . This is much larger than the equilibrium constants for base flipping in the undamaged DNAs discussed earlier. While the energy required for the thymine dimer to undergo base flipping does not point definitively to extrahelical or intrahelical recognition, the energy is not prohibitively high, which would effectively rule out the possibility of extrahelical recognition. This energy value also indicates that if the enzyme is to induce base flipping of the thymine dimer the energy required to do so is low, in comparison with other, nondamaged DNA bases.

Previous computational work has determined the energy required for nondamaged DNA bases to undergo base flipping in free solution, i.e., not enzyme bound.<sup>26–28</sup> The values reported for thymine, adenine, guanine, and cytosine were 13, 15, ~19 and ~15 kcal/mol, respectively. These energy values correspond to equilibrium constants on the order of  $10^{-10}$ – $10^{-14}$ , which indicates the likelihood that nondamaged bases are flipped out in free solution is quite low. This supports the intrahelical recognition followed by enzyme-induced base-flipping mechanism of target base binding by DNA modifying enzymes for nondamaged DNA bases. However, the thymine dimer, which is formally two DNA bases, requires much less energy to undergo base flipping. This is most likely due to the structural



perturbations, both on a small scale, i.e., hydrogen bonding and base stacking, and on a larger scale, i.e., a large global bend of the DNA, that are caused by the thymine dimer. Therefore, the possibility that an extrahelical thymine dimer is recognized by DNA photolyase is greater than for nondamaged bases. The possible differences in recognition schemes may be attributed to the structural perturbations caused by the target bases. Nondamaged bases induce no structural perturbation and are therefore more difficult to flip out of the duplex. Damaged bases that contain very small changes, such as 8-oxo-G, and do not induce structural perturbations are also recognized intrahelically, as determined experimentally.<sup>17</sup> Damaged bases that induce significant perturbations to the DNA structure, such as the thymine dimer, may possibly be recognized extrahelically due to the lower energy required to undergo base flipping.

In summary, this work presents the first structure of a DNA duplex containing a thymine dimer in the flipped-out conformation. The global bend for the flipped-out structure was similar to that of the flipped-in structure. The calculated structures were used as the starting and ending points for the calculation of a two-dimensional potential of mean force for the base-flipping process. In order to calculate the two-dimensional PMF without the ability to perform complete sampling of the entire coordinate space, a novel protocol was used. The energy required for base flipping of the thymine dimer was found to be between 6.25 and 6.5 kcal/mol. This is much lower than the values reported for nondamaged DNA bases. This allows for the possibility that the thymine dimer may be recognized by enzymes in the extrahelical conformation. This work will be extended to study the sequence dependence of base flipping of the thymine dimer, which will be presented in a forthcoming publication.

**Acknowledgment.** We acknowledge the generous allocation of computer resources by the Center for Research Computing at the University of Notre Dame. L.L.O. is the recipient of a Schmitt Fellowship by the University of Notre Dame.

**Supporting Information Available:** Analysis of flipped-in and flipped-out structures, including RMSD, radial distribution functions, curves analysis and structures illustrating pseudodihedral angles, detailed sampling schemes, and convergence of the PMF. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References and Notes

- (1) Klimasauskas, S.; Kumar, S.; Roberts, R. J.; Cheng, X. *Cell* **1994**, *76*, 357–369.
- (2) Reinisch, K. M.; Chen, L.; Verdine, G. L.; Lipscomb, W. N. *Cell* **1995**, *82*, 143–153.
- (3) Banerjee, A.; Yang, W.; Karplus, M.; Verdine, G. L. *Nature* **2005**, *434*, 612–618.
- (4) Vassilyev, D. G.; Kashiwagi, T.; Mikami, Y.; Ariyoshi, M.; Iwai, S.; Ohtsuka, E.; Morikawa, K. *Cell* **1995**, *83*, 773–782.
- (5) Sancar, A. *Chem. Rev.* **2003**, *103*, 2203–2238.
- (6) Harrison, C. B.; O'Neil, L. L.; Wiest, O. *J. Phys. Chem. A* **2005**, *109*, 7001–7012.
- (7) Friedberg, E. C. *DNA Repair*; W.H. Freeman & Co.: New York, 1985; Chapters 1-5, 2-1, and 2-2.
- (8) Mees, A.; Klar, T.; Gnau, P.; Hennecke, U.; Eker, A. P. M.; Carell, T.; Essen, L.-O. *Science* **2004**, *306*, 1789–1793.
- (9) Christine, K. S.; MacFarlane, A. W., IV; Yang, K.; Stanley, R. J. *J. Biol. Chem.* **2002**, *277*, 38339–38344.
- (10) Sanders, D. B.; Wiest, O. *J. Am. Chem. Soc.* **1999**, *121*, 5127–5134.
- (11) Antony, J.; Medvedev, D. M.; Stuchebrukhov, A. A. *J. Am. Chem. Soc.* **2000**, *122*, 1057–1065.
- (12) Park, H.-W.; Kim, S.-T.; Sancar, A.; Deisenhofer, J. *Science* **1995**, *268*, 1866–1872.
- (13) Miaskiewicz, K.; Miller, J.; Cooney, M.; Osman, R. *J. Am. Chem. Soc.* **1996**, *118*, 9156–9163.
- (14) Park, H.; Zhang, K.; Ren, Y.; Nadji, S.; Sinha, N.; Taylor, J.-S.; Kang, C. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 15965–15970.
- (15) Spector, T. I.; Cheatham, T. E., III; Kollman, P. A. *J. Am. Chem. Soc.* **1997**, *119*, 7095–7104.
- (16) Banerjee, A.; Santos, W. L.; Verdine, G. L. *Science* **2006**, *311*, 1153–1157.
- (17) Blainey, P. C.; van Oijen, A. M.; Banerjee, A.; Verdine, G. L.; Xie, X. S. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 5752–5757.
- (18) Dornberger, U.; Leijon, M.; Fritzsche, H. *J. Biol. Chem.* **1999**, *274*, 6957–6962.
- (19) Ward, D. C.; Reich, E. *J. Biol. Chem.* **1969**, *244*, 1228–1237.
- (20) Guest, C. R.; Hochstrasser, R. A.; Sowers, L. C.; Millar, D. P. *Biochemistry* **1991**, *30*, 3271–3279.
- (21) Xu, D.; Evans, K. O.; Nordlund, T. M. *Biochemistry* **1994**, *33*, 9592–9599.
- (22) Holz, B.; Klimasauskas, S.; Serva, S.; Weinhold, E. *Nucleic Acids Res.* **1998**, *26*, 1076–1083.
- (23) Bjornson, K. P.; Moore, K. J. M.; Lohman, T. M. *Biochemistry* **1996**, *35*, 2268–2282.
- (24) O'Neil, L. L.; Wiest, O. *J. Am. Chem. Soc.* **2005**, *127*, 16800–16801.
- (25) Priyakumar, U. D.; MacKerell, A. D., Jr. *Chem. Rev.* **2006**, *106*, 489–505.
- (26) Guidice, E.; Várnai, P.; Lavery, R. *ChemPhysChem* **2001**, *2* (11), 673–677.
- (27) Guidice, E.; Várnai, P.; Lavery, R. *Nucleic Acids Res.* **2003**, *31*, 1434–1443.
- (28) Banavali, N. K.; MacKerell, A. D., Jr. *J. Mol. Biol.* **2002**, *319*, 141–160.
- (29) Barthel, A.; Zacharias, M. *Biophys. J.* **2006**, *90*, 2450–2462.
- (30) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A. *Amber 9*; University of California: San Francisco, 2006.
- (31) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, D. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- (32) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.
- (33) Priyakumar, U. D.; MacKerell, A. D., Jr. *J. Chem. Theory Comput.* **2006**, *2*, 187–200.
- (34) *Insight II*; Accelrys Software Inc.: San Diego, CA.
- (35) Darden, T.; York, D.; Pedersen, L. G. *J. Chem Phys.* **1993**, *98*, 10089–10092.
- (36) Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1988**, *6*, 63–91.
- (37) Pearlman, D. A.; Holbrook, S. R.; Pirkle, D. H.; Kim, S.-H. *Science* **1985**, *227*, 1304–1308.
- (38) Barsky, D.; Foloppe, N.; Ahmadi, S.; Wilson, D. M., III; MacKerell, A. D., Jr. *Nucleic Acids Res.* **2000**, *28*, 2613–2626.
- (39) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 1011–1021.
- (40) Kumar, S.; Rosenberg, J. M.; Bouzida, J.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1339–50.
- (41) Grossfield, A. <http://dasher.wustl.edu/alan/> (accessed Aug 29, 2006).