Dynamics of 5R-Tg Base Flipping in DNA Duplexes Based on Simulations—Agreement with Experiments and Beyond

Shu dong Wang, Leif A. Eriksson,* and Ru bo Zhang*

ABSTRACT: Damaged or mismatched DNA bases are normally thought to be able to flip out of the helical stack, providing enzymes with access to the faulty genetic information otherwise hidden inside the helix. Thymine glycol (Tg) is one of the most common products of nucleic acid damage. However, the static and dynamic structures of DNA duplexes affected by SR-Tg epimers are still not clearly understood, including the ability of these to undergo spontaneous base flipping. Structural effects of the SR-Tg epimers on the duplex DNA are herein studied using molecular dynamics together with reliable DFT based calculations. In comparison with the corresponding intact DNA, the cis-5R,6S-Tg epimer base causes little perturbation to the duplex DNA, and a barrier of 4.9 kcal mol\(^{-1}\) is obtained by meta-eABF for cis-5S,6S-Tg base flipping out of the duplex DNA, comparable to the 5.4 kcal mol\(^{-1}\) obtained for the corresponding thymine flipping in intact DNA. For the trans-5R,6R-Tg epimer, three stable local structures were identified, of which the most stable disrupts the Watson–Crick hydrogen-bonded G5/C20 base pair, leading to conformational distortion of the duplex. Interestingly, the relative barrier height of the 5R-Tg flipping is only 1.0 kcal mol\(^{-1}\) for one of these trans-5R,6R-Tg epimers. Water bridge interactions were identified to be essential for SR-Tg flipping. The study clearly demonstrates the occurrence of partial trans-5R,6R-Tg epimer flipping in solution.

INTRODUCTION

Base flipping is a key fundamental theme in nucleic acid biophysics and biochemistry. Studies have shown that base flipping is a common strategy for enzymes such as methyltransferases, glycosylases, and endonucleases,\(^1\) to read and chemically modify bases. Base flipping may even be linked to early events in the opening and unwinding of DNA for transcription and replication processes.\(^7\) Although extensive studies have found that many DNA repair/modification proteins completely flip their target base out extrahelically, it is still under debate whether the base flipping occurs spontaneously or not.\(^8\)–\(^10\) Therefore, accurate information about base flip dynamics is of high interest and importance.\(^1,0^\)–\(^1,1^\)

Thymine glycol (5,6-dihydro-5,6-dihydroxy thymine; Tg) is the most common oxidation product of thymine. Approximately 400 Tg residues are formed in a normal cell each day, and 10–20% of genome damages have been attributed to the oxidative conversion of thymine to Tg.\(^12\)–\(^18\) Due to the chirality of the C5 and C6 atoms, Tg could exist as a mixture of the two pairs of cis- and trans- stereoisomers: the SR cis–trans pair (5R,6S: 5R,6R) and the SS cis–trans pair (5S,6R: 5S,6S).\(^19\)–\(^21\) The SR-Tg stereoisomer is thought to be the more abundant of these, with an equilibrium ratio of 7:3 between cis-5R,6S and trans-5R,6R Tg in DNA oligomers containing the Tg-A base pair, while this ratio is 87:13% at the single-nucleoside level.\(^22,23\) The epimers are suggested to induce large structural changes to duplex DNA, reflected in the fact that the SR-Tg base could be either extra-helical or coordinating to the opposing base on the complementary strand, depending strongly on the local interaction.\(^24\)–\(^28\) The exact form or distribution would thus be linked to gene translation.\(^21\)–\(^28\) Except for data from NMR spectra in solution in combination with 0.01 μs constrained molecular dynamics (rMD) simulations,\(^29\) reliable structural information related to Tg epimers is unfortunately not available. Although the Tg:adenine base pair is more biologically relevant, crystal structures are only available with cytosine opposite to 5R-Tg.\(^2,30\) It was previously thought that mismatched or damaged bases had a certain chance of spontaneously flipping out of the double helix structure of a DNA molecule because they could not form a normal and stable Watson–Crick base pair interaction\(^31\) and that the flipped base would thus function as a signal to be recognized and captured by repair
proteins. This hypothesis lacks atomistic level evidence on the effect of particular Tg epimers on the DNA supramolecular structure in biologically relevant DNA.

In this work, the cis-5R,6S- and trans-5R,6R-Tg epimer-containing DNA duplexes, respectively, (referred to as cis-DNA and trans-DNA; see Scheme 1) were modeled, and their

Scheme 1. (a) The Structures of the 5R-Tg Pair and (b) the Sequence of the Dodecamer Used in the Current Study

static and dynamic structures and energies explored using Charmm36 force field based MD simulations. Benchmark calculations on the intact dodecamer with thymine in the same position (referred to as DNA-thy) were also performed to study how the SR-Tg epimers deviate from the intact DNA duplex. The present results show that the cis-5R,6S-Tg in cis-DNA is always intrahelical and forms a Watson–Crick base pair with adenine. In contrast, three metastable conformations for trans-DNA are found. The locally most stable of these has relatively high energy and results in severe deformation of the duplex. This is attributed to the complex hydrogen-bonding network formed by trans-5R,6R-Tg with its surrounding bases, leading to loss of the classical Watson–Crick G/C and A/T pairs. The results also clearly illustrate the mechanism by which some of the trans-5R,6R-Tg conformers are capable of flipping out of the trans-DNA duplex.

■ COMPUTATIONAL METHODS AND DETAILS

The initial coordinates of the cis-5R,6S-Tg containing dodecamer DNA (denoted cis-DNA) were obtained from the NMR structure in the Protein Data Bank (PDB ID: 2KH5) in which it is noted that the C5-CH₃ group of Tg takes an axial position (Figure S1). Two additional dodecamer DNA duplexes were generated from the cis-5R,6S-Tg containing structure by mutating these in pymol to form the systems containing trans-5R,6R-Tg (trans-DNA) and T (DNA-thy), the duplex sequence shown in Scheme 1b. DNA-thy is the intact DNA duplex, included as control.

Each dodecamer duplex was immersed in ca. 9260 TIP3P water molecules, in order to ensure that the systems were completely solvated. The system was neutralized by 0.15 M NaCl to imitate the intracellular environment. For nonbonded interactions, periodic boundary conditions with a cutoff radius of 12 Å were included, and the simulation box size was 59 × 63 × 44 Å with a minimum distance of 10 Å between DNA and the edges of the box. The particle mesh Ewald (PME) algorithm was used to handle electrostatic interactions. Bonds to hydrogen atoms were constrained using the SHAKE algorithm. The water molecules were initially minimized in 1000 conjugate gradient steps with the solute molecule(s) held fixed, followed by 1000 steps of conjugate gradient minimization of the whole system. After a 500 ps heating process from 0 to 298 K in a canonical ensemble (NVT) with the solute fixed, a series of harmonic constrained isothermal–isobaric ensemble (NPT) simulations were performed to enable a controlled release of the solute degrees of freedom. The scaling used for the constraints was 5.0, 1.0, and 0.5 kcal mol⁻¹Å⁻², respectively. Under each constrained scaling, 500 ps MD simulation was carried out using an NPT ensemble. Constant temperature was maintained by the Langevin thermostat method and the pressure was maintained by the Langevin piston Nose–Hoover method (a combination of the Nose–Hoover constant pressure method and Langevin dynamics). Unconstrained MD production simulations of 1 μs were performed in NPT ensembles with time step 2.0 fs. For cis-DNA, three independent 1 μs MD simulation were performed. For trans-DNA, an initial simulation of 1 μs length was first performed, which yielded the conformer trans-DNA-1. Two independent 1 μs replicas were subsequently performed with focus on the additional stable states, trans-DNA-2 and trans-DNA-3. In addition, the native DNA-thy (Scheme 1b) was simulated in three independent 1 μs MD simulations for comparison. The total simulation time in the study is more than 9.0 μs. The trajectory of the last 0.1 μs of each simulation was used to analyze and display the results using VMD 1.9.3. The DNA conformational analyses were performed with Curves+. All MD simulations were performed using NAMD 2.13 together with the Colvar module. The Charmm36 general force field was used throughout.

For insights into the flipping process of Tg from the duplex, we also performed the enhanced sampling dynamics, using the recently developed combination of extended adaptative biased force (eABF) and metadynamics (meta-eABF). In meta-eABF, a metadynamic-like memory kernel (MtD) is incorporated into the extended system alongside the eABF biasing force, thus, leading to

\[ F_{\text{bias}}(\xi') = F_{\text{bias,eABF}}(\xi') + F_{\text{bias,MtD}}(\xi') = K(\xi' - <\xi >_M) \]

\[ + dU_{\text{MtD}}(\xi', t)/(d\xi') \]

where \( U_{\text{MtD}}(\xi', t) \) is the time-dependent MtD-like memory kernel. The extended PMF then writes

\[ \Delta A' = \Delta A'_{\text{eABF}} + \Delta A'_{\text{MtD}} \]

Through simultaneous addition of eABF biasing forces and a suitable form of the MtD Gaussian potentials, meta-eABF is particularly efficient for the rapid exploration of the free-energy landscape. The algorithm was proven to possess remarkable convergence properties over a broad range of applications including DNA, with as much as a 5-fold speedup, compared with standard ABF. The present average structures from the MD trajectories were used as initial structures for the potential of mean force (PMF) or free energy surface (FES) estimations. Meta-eABF was run under the NPT ensemble with instantaneous force values accrued in bins of width 0.1 Å × 0.1 Å × 0.1 Å. Settings for Gaussian hillWeight = 0.1 kcal mol⁻¹ and hillWidth = 5 bin width were employed in the simulations. Distance is a reaction coordinate recently proposed to study the flipping. In this study, the center-of-mass (COM) distance between the Tg (or T6 in DNA-thy) nucleotide and A19 was considered as the collective variable (CV) in the meta-eABF simulation.
DFT calculations were performed at the isolated nucleotide level in a vacuum using the Gaussian 09 code,51 to assess the interaction of the non-Watson−Crick type of hydrogen-bonding base pairs. Geometry optimizations were performed using the M06-2X/6-31+G(d,p) method.52 The optimized structures were confirmed through frequency calculations at the same level, to be real minima with no imaginary vibration frequencies. The M06-2X functional was designed in part to yield more accurate noncovalent interactions containing significant dispersion contributions, as well as reliable thermochemical data.53 The interaction energy reported in the study is defined as

\[ \Delta E_{\text{int}} = E_{\text{complex}} - (E_{\text{monomer1}} + E_{\text{monomer2}}). \]

**Tg Parametrization.** The partial atomic charges, bonds, angles, and dihedral terms were developed and fitted with the aid of the Force Field Toolkit (ffTK),54 while Lennard-Jones parameters and improper torsion parameters were taken by analogy from CHARMM’s CGenFF.55 All nonidentical atoms except hydrogens were assigned to unique atom types. We used the parametrization order as specified in the general CHARMM procedure, where the partial atomic charges were optimized first, followed by bonds and angles, and finally the dihedrals. The optimization of all parameters and vibrational analyses were done using the molecular geometries obtained by energy minimization at the MP2/6-31G* level of theory.

**RESULTS AND DISCUSSION**

The cis-5R,6S Thymine Glycol Epimer in the DNA Dodecamer. For the cis-DNA duplex, three independent 1 μs production simulations were performed and root-mean-square deviation (RMSD) with respect to the first frame of the production simulation was used to monitor the duplex structure as a measure of system stability, as displayed in Figure S2. The RMSD of cis-DNA displays only a slight fluctuation and yields very similar values to those of the intact duplex DNA-thy (Figure S3). Furthermore, the standard deviation in the RMSD over the last 0.1 μs simulation is only 0.40 Å (Figure 1a), showing that the relative structural change is very small. Therefore, in accordance with previous studies of natural and damaged DNA,56 detailed structural analysis was carried out on the last 0.1 μs simulation. In addition, root-mean-square fluctuation (RMSF) values were calculated to investigate how much the individual nucleobases moved during the simulations.57 As seen in Figure S2, the largest fluctuations occur at the terminal nucleotides of the duplex, again very similar to DNA-thy (Figure S3). The fluctuations of the cis-5R,6S-Tg base and its flanking G5 and G7 (cf. Scheme 1) are...
Figure 3. Average structures of (a) the first 25 ps simulation and (b) 25–100 ps simulation of cis-DNA. Relevant hydrogen bonds are indicated by dashed red lines. In a, Tg:O6H$_{\alpha\omega}$···N7‘:G7 = 2.03 Å; in b, Tg:O6H$_{\alpha\omega}$···O5‘ = 2.36 Å; Tg:O6H$_{\alpha\omega}$···O4‘ = 2.23 Å.

very small and positioned around RMSF values 1.17 ± 0.24, 1.74 ± 0.29, and 1.69 ± 0.29 Å, respectively. We may thus conclude that, in accordance with previous studies of DNA oligonucleotides, $^{56}$ 1 μs simulations ensure convergence of key DNA structural parameters, and reliable conclusions can be drawn.

The conformation of the C5-CH$_3$ group of thymine glycol has been thought to be a factor affecting the local structure of duplex DNA. $^{55,56,59}$ Previous NMR experiments could not discriminate between axial or equatorial conformations of the CH$_3$ group in cis-5R,6S-Tg, as both conformations showed agreement with the NOE data. $^{29,60}$ Moreover, both conformations were observed in the 0.01 μs rMD simulation. $^{29}$

The conformational selectivity of the CH$_3$ group in cis-5R,6S-Tg can be determined by local thermodynamics and controlled by the strength of the hydrogen bonds associated with the 6-OH substituent.

DFT was next used to explore the conformational selectivity. Structures of the cis-5R,6S-Tg nucleotide, including the methyl group in axial and equatorial conformations, were separately extracted from the MD simulations and optimized using the dispersion correction function M06-2X with the standard 6-31+G(d,p) basis set (Figure S5). The final single-point energy was computed at the MP2/6-311G(d,p) level (Table S1). For the system with axial conformation of the CH$_3$ group, the 5-OH can form a hydrogen bond with either the O4 or O6 atom on cis-5R,6S-Tg, while 6-OH is unbound. These two rotamers are almost isoenergetic. They are significantly higher in energy (+3.5 kcal mol$^{-1}$) than cis-5R,6S-Tg with an equatorial CH$_3$ group, for which the two intranucleotidyl hydrogen bonds of 6-OH to O5‘ and O4‘ observed on the cis-5R,6S-Tg-nucleotide are the same as those observed in the MD trajectory. These results suggest that the CH$_3$ group on cis-5R,6S-Tg either isolated or in double-stranded DNA will preferentially take an equatorial conformation. In addition, our MD simulations and DFT calculations show that the conformation of 6-OH in cis-5R,6S-Tg due to the ring puckering consistently displays the same conformation as the CH$_3$ group.

The average structure of the stable cis-DNA dodecamer overlaps well with the intact DNA-thy structure (Figure S6a), except for the terminal nucleotides. The Watson–Crick type cis-5R,6S-Tg/A19 interaction and the two flanking G/C hydrogen bond pairs were always preserved and their stacking interactions well maintained throughout the simulation (Figures S7a and S8). The centroid–centroid distance between cis-5R,6S-Tg and G7 is 4.78 ± 0.25 Å, which exceeds
the 3.90 ± 0.33 Å between T6 and G7 in DNA-thy (cf. Scheme 1 for numbering). This is due to the repulsive interaction between the axial 5-OH group and the G7 base, which is consistent with previous findings. The centroid distance between cis-5R,6S-Tg and G5 is 3.89 ± 0.15 Å.

To further explore the effect of noncovalent interactions on the affinity, interaction energy decomposition analysis (EDA) was performed on the data from the MD simulations. The interaction energies of cis-5R,6S-Tg with its adjacent G5, G7, and A19 bases were decomposed during the last 0.1 μs of simulation, as shown in Figure 4b and Table S2. First, hydrogen bonding is readily identified as a critical interaction and is considered to be a key factor in maintaining the secondary structure of DNA. The Watson–Crick type T6/A19 hydrogen bond energy is approximately −11.2 ± 1.1 kcal mol⁻¹, as obtained from the DNA-thy MD simulation (Figure 4a), in good agreement with the energy estimated at the M06-2X/6-31+G(d,p) level, −13.8 kcal mol⁻¹. The average interaction energy between cis-5R,6S-Tg and A19 is −11.5 ± 1.4 kcal mol⁻¹ (Figure 4b), which is very close to the −12.9 kcal mol⁻¹ calculated at the M06-2X/6-31+G(d,p) level. These results indicate that the hydrogen bonding energies between A19 and T6 or cis-5R,6S-Tg are almost identical. In contrast, the electrostatic interaction, Elec, between G7 and cis-5R,6S-Tg is 2.6 ± 1.6 kcal mol⁻¹ and the total interaction energy between the two is −2.5 ± 1.5 kcal mol⁻¹. This is significantly weaker than that between cis-5R,6S-Tg and G5 and provides an explanation to why the distance between the cis-5R,6S-Tg and G7 bases is extended. The sum of the dominant van der Waals (vdW) interaction energy of cis-5R,6S-Tg with its flanking G5 and G7 bases is about −11.4 kcal mol⁻¹, which is very close to the −11.5 kcal mol⁻¹ obtained for the dominant Elec interaction energy of cis-5R,6S-Tg with A19. These indicate that Elec and vDW contribute equally to the affinity of cis-5R,6S-Tg to duplex DNA. The same is found in the interactions of T6 with A19, G5, and G7 in the intact DNA-thy. Moreover, the total interaction energies of T6 or cis-5R,6S-Tg with the A19, G5, and G7 bases are −23.8 and −22.9 kcal mol⁻¹ in intact DNA-thy and cis-DNA, respectively. Thus, these comparative studies lead to the conclusion that the stability of the cis-5R,6S-Tg base in duplex DNA depends on the significant dispersion force of the cis-5R,6S-Tg nucleobase to its neighboring G5 and G7 bases, in addition to the hydrogen bond with the complementary A19 base. All replicas give highly similar results, seen in Table S3.

The trans-5R,6R Thymine Glycol Epimer in the DNA Dodecamer. 5R-Tg was determined to be present in solution as a 7:3 equilibrium mixture of the cis-5R,6S- and trans-5R,6R-Tg epimers at 298 K. It is therefore difficult to discern experimentally how each epimer affects the structural and dynamic properties of the duplex DNA. On the basis of the current analysis of cis-5R,6S-Tg binding to the recognition site of the duplex DNA, we conclude that the bases G5, G7, and A19 surrounding cis-5R,6S-Tg play a key role in the binding and that cis-DNA essentially maintains the structure of the native system. To further investigate the effect of epimers on...
the stability of double-stranded DNA, a similar set of MD simulations and analyses was performed for the system containing the trans-5R,6R-Tg base. Throughout all trajectories, the CH$_3$ group remained equatorial, and the 6-OH conformation axial. Interestingly, two relatively stable structures were identified in the 1 μs simulation. The first of these (referred to as the “metastable” state) was present in the simulation interval between 0.35 and 0.40 μs, and the RMSD and RMSF for all bases therein are shown in Figures S9a and S9b. An arc-shaped hydrogen-bonding between trans-5R,6R-Tg and A19 was noted, with an interaction energy estimated to be $-8.6 \pm 1.7$ kcal mol$^{-1}$ based on the MD trajectory. This indicates that their interaction strength was significantly lower than that of the hydrogen-bonded T6/A19 base pair in DNA. Indeed, DFT calculations show that the isolated arched hydrogen-bonded base pair is not stable on the potential energy surface of the interaction between the trans-5R,6R-Tg and A19 bases. Instead, water molecules appear necessary to keep the metastable structure with the arc-shaped trans-5R,6R-Tg/A19 base pairing intact. As seen in Figure 5a and 6a, the number of water molecules around trans-5R,6R-Tg increases relative to that of 5R,6S-Tg in the cis-DNA production trajectories. The average water number around O6 and H$_{O6}$ in trans-5R,6R-Tg is between 0.2 and 0.4 up until 0.35 μs of the simulation and increases to 1.1 and 1.2, respectively, in the metastable structure between 0.35 and 0.40 μs. In addition, around the trans-5R,6R-Tg nucleotide, the water number remains constant at 1.8 for O4’, 1.6 for O5, and 2.4 for HS during the first 0.4 μs. The results indicate that the “additional” water molecule required to maintain the metastable structure causes a weakening of the hydrogen bond strength of Tg:O6H$_{O6}$···O4’:Tg after 0.35 μs, presented in Figure 6b. It thereby becomes less capable of anchoring the trans-5R,6R-Tg base, which results in an increased rotation of the trans-5R,6R-Tg base around the N-glycosidic bond, Figure 6c. This is also consistent with the variation in RMSF values of 2.28 and 1.88 Å seen for the trans-5R,6R-Tg and A19 bases, respectively (Figure S9b).

After the metastable structure, one of the observed stable trans-DNA species, trans-DNA-1, is formed at 0.40 μs and retained during the remaining simulation. The standard deviation in the RMSD (Figure 7a) over the last 0.10 μs is
and has a disrupted Watson-S12. One of these, shown in Figure 5c, is labeled and their RMSDs and RMSFs are shown in Figures S11 and two new stable structures of trans-DNA were also observed, DNA-1 structure as discussed above. In the other two replicas, metastable DNA structure is present with lifetimes from 0.02 to 0.05 μs. After the metastable state, trans-DNA is able to reach the more stable structures referred to as trans-DNA-1, -2, and -3 barely perturb the structures of the adjacent base pairs. Superposition of the observed structures show that trans-DNA-3 overlaps very well with DNA-thy (Figure S6d and Figure S5d), while the conformation of trans-DNA-1 deviates to a large extent from the intact DNA duplex (DNA-thy; Figure S6b). The distortions for trans-DNA-1 is mainly manifested by the change of the stable C20/G5 Watson–Crick base pair to a new hydrogen bonded structure between trans-SR,6R-Tg and C20 and A19, seen in Figure 5b. This also results in further reduction of the hydrogen-bond strength between trans-SR,6R-Tg and A19. A similar situation is found for the trans-DNA-2 structure, presented in Figure 5c. These results suggest that the trans-DNA-1 and -2 assemblies should be less stable than trans-DNA-3 with its perfect hydrogen-bonded base pairs. The stabilities of the trans-DNA species were roughly estimated by calculating the total energy of the G5/C20, Tg/A19, and G7/C18 pairs. Their relative stabilization energies are 0.0, −6.4, and −29.3 kcal mol⁻¹, corresponding separately to the pairs in trans-DNA-1, -2, and -3. The stability order is consistent with the degree of deformation of trans-DNA.

**Flipping Free Energy Calculations.** Understanding the dynamic process of the thymine glycol epimer affinity to the duplex DNA can provide further insights into the recognition mechanism of the modified nucleic acid by repair enzymes and

![Figure 8](https://doi.org/10.1021/acs.jcim.1c01169)

**Figure 8.** Interaction energy decomposition for trans-SR,6R-Tg interaction with the neighboring G5, G7, and A19 bases. (a) Metastable structure, (b) trans-DNA-1, (c) trans-DNA-2, (d) trans-DNA-3.
polymerases. To address this issue, free-energy profiles were separately computed for 5R-Tg flipping out of the cis-DNA and trans-DNA supramolecular assemblies using a progressive sampling algorithm, meta-eABF. The same set of calculations was also performed for T6 flipping in DNA-thy, for comparison. In short, the center-of-mass (COM) separation distance between the Tg (or T6 in DNA-thy) nucleotide and A19 was considered as the collective variable (CV).

Previous studies of base flipping by MacKerell et al., Lavery et al., and us have shown that simulation time for each umbrella sampling window is long enough for satisfactory convergence. Herein, we also examined the free energy surface (FES) of the 5R-Tg flipping using meta-ABF simulations of lengths 30, 40, 50, 60, 100, and 120 ns, respectively, for the cis-DNA system (Figure S13). For simulation times less than 40 ns, the free energy surfaces did not converge properly. For the trajectory time of 50 ns, the peak relative to a COM separation of 12.4 Å shows a free energy barrier of 4.7 kcal mol\(^{-1}\), which is slightly higher (by 1.7 kcal mol\(^{-1}\)) than that obtained using 40 ns trajectories. The FES from the 60 ns simulation shows a free energy barrier of 4.9 kcal mol\(^{-1}\), which is very similar to that of the 50 ns simulation. In addition, comparative studies through prolonged simulation times (100 and 120 ns, respectively) show the estimated PMF to be 4.4–4.5 kcal mol\(^{-1}\), which is very close to the results obtained from the 60 ns simulation. In addition, the native thymine flipping FES from the 60 ns simulation shows a free energy barrier of 5.4 ± 0.2 kcal mol\(^{-1}\), comparable to the recent studies (5.3–7.5 kcal mol\(^{-1}\)) from our meta-eABF simulation using pseudodiagonal angle as the reaction coordinate (Figure S15). Our present results hence show that meta-eABF calculations require relatively short simulation times to meet satisfactory convergence.

At the lowest point along the PMF curve (1 of Figure 9a), cis-5R,6S-Tg is bound by the hydrogen bonds and base stacking interactions, and no water molecule is observed in the recognition region. From this first basin along the PMF curve, an energy barrier of 4.9 kcal mol\(^{-1}\) (2 of Figure 9a) must be overcome to break the hydrogen bond between A19 and cis-5R,6S-Tg. This is comparable to the 5.4 kcal mol\(^{-1}\) barrier obtained from the 60 ns simulation trajectory of T6 base flipping from DNA-thy (2 of Figure 9b). At this point, there is one water molecule forming a hydrogen-bonded bridge connecting A19 with cis-5R,6S-Tg, which further weakens the interaction between Tg and the DNA duplex. The base stacking between Tg, G5, and G7 is retained during these initial stages of the process.

As the CV distance increases, a second water enters between cis-5R,6S-Tg and A19, disrupting the canonical Watson–Crick hydrogen bonds. After breaking the hydrogen bonds, a shallow basin (3 of Figure 9a) is found in the PMF. As an increasing number of water molecules penetrate into the active region, cis-5R,6S-Tg is flipped out from the helix and surrounded by water molecules, whereby the PMF curve reaches the lowest point (4 of Figure 9a). Apparently, water-mediated hydrogen bonding...
bonding helps to reduce the activation barrier of the cis-5R,6S-Tg flipping. We conclude that the barrier height is comparable to that for T6 flipping out of the intact duplex DNA, implying that the cis-5R,6S-Tg epimer should be very stable in the duplex DNA. Interestingly, the CH3 group is observed to shift between the axial and equatorial conformation during the 60 ns process, albeit with a preference for the equatorial arrangement as seen in Figure S16. Note that the crystal data showed that A19 was mutated into cytosine in the DNA-repair enzyme interaction system, causing cis-5R,6S-Tg to lose the Watson–Crick hydrogen bonds. Such a loss of base pairing would yield a much lower barrier toward flipping. In addition, the double helix structure of DNA deviates only a little from that of the standard B-DNA. The present results can provide a reasonable implication that the recognition of cis-5R,6S-Tg in DNA by repair enzymes yields a large deformation of the double strand to facilitate relevant repair of DNA.

PMF calculations for the trans-5R,6R-Tg base flipping processes were also performed. For trans-DNA-1, the stable local structure of trans-5R,6R-Tg/C20/A19 is solvated in the starting basin 1 (Figure 10) whereby one water molecule becomes bonded to 6-OH of Tg. Along the CV, a barrier height of only 1.0 kcal mol\(^{-1}\) needs to be overcome to break the hydrogen bonds between trans-5R,6R-Tg and the C20 and A19 bases (Figure 10). At the first peak 2, a second water comes close to O4 of trans-5R,6R-Tg. The second basin 3 is very shallow, representing the loss of hydrogen bonding between trans-5R,6R-Tg and A19,C20. Thereafter, complete solvation of the trans-5R,6R-Tg base is reached, point 4 in Figure 10. The low barrier is in sharp contrast to the barrier heights of 4.0 and 5.2 kcal mol\(^{-1}\) that must be overcome for elongation of the hydrogen bonding networks in trans-DNA-2 and trans-DNA-3, respectively (Figure S17), which instead are comparable to those calculated for cis-DNA and DNA-thy.

DNA helices are flexible and can exist in multiple conformations in solution. For the presently studied trans-DNA, three stable DNA duplexes were observed. They are in all likelihood in thermodynamic equilibrium in solution, in a ratio depending on the Boltzmann distribution. Thus, the distribution of the more distorted trans-DNA-1 should be smaller than those of trans-DNA-2 and -3. Strikingly, the barrier height for trans-5R,6R-Tg base flipping in trans-DNA-1 is significantly lower than for trans-DNA-2 and -3, implying that trans-5R,6R-Tg flips out of duplex DNA very easily in trans-DNA-1, once formed. This provides an explanation to the solution NMR experiments in which it was observed that the 5R-Tg bases in DNA were only partially extrahelical. By unambiguously taking the role of the cis–trans epimers into account, our study furthermore claims that the extrahelical 5R-Tg base should originate from the trans-5R,6R-Tg epimer in trans-DNA.

The DNA structural parameters of the obtained conformations were analyzed using Curves+. As seen from Table S5, the intrabase parameter buckle, opening and interbase parameter tilt of Tg-A base pair of trans-DNA-1 are much more pronounced than those of the trans-5R,6R-Tg epimer.

Figure 10. PMF profile of trans-5R,6R-Tg flipping out of the duplex of trans-DNA-1 (the error is within ca. 0.3 kcal mol\(^{-1}\)). 1–4 illustrate the main structural changes of trans-5R,6R-Tg interaction with A19 and C20. The time evaluation plots of the examined CV in the meta-eABF simulations are shown in Figure S14.
higher than those in DNA-thy and cis-DNA. These lead to significant distortions of the grooves near the Tg site. In particular, the major groove width of trans-DNA-1 increased to ~18.6 Å compared to the major groove width of DNA-thy at ~11.6 Å. Since a wider major groove was proven more favorable for base flipping,\(^7\) this explains why only a small barrier is required for trans-5R,6R-Tg flipping in trans-DNA-1.

We adopted the pseudodihedral angle CPDb\(^7\) to study the pathway of Tg flipping, as shown in Figure 11. Native T (Figure 11a) can flip through both the major and the minor groove pathways, but the observed events for flipping through the minor groove pathway are fewer than through the major groove pathway. From Figures 11b,c and S18, we note that the base flippings of both cis-5R,6S-Tg and trans-5R,6R-Tg occur almost exclusively through the major groove pathway. A relatively larger steric barrier is found on the minor groove side, and thus the Tg base flipping in this direction seems to be largely forbidden. The present results are very consistent with previous studies.\(^7\)

- CONCLUSIONS

Using the well-known thymine glycol as an example, we have used extended molecular dynamics simulations combined with reliable DFT calculations to address the influence of epimers on the stability of DNA supramolecular assemblies. This is to our knowledge the first comparative modeling study of DNA double-stranded structures including cis-5R,6S-Tg and trans-5R,6R-Tg epimers, respectively. It is clearly demonstrated that the CH\(_3\) group of 5R-Tg is energetically more inclined to be in a pseudoequatorial conformation due to the formation of stronger hydrogen bonds between 6-OH and O\(_{5}S\),O\(_{4}S\) in the 5R-Tg nucleotide.

The duplex DNA containing cis-5R,6S-Tg has comparable stability to the corresponding intact DNA. Energy decomposition analysis shows that Elec and vdW interaction contribute equally to the binding of cis-5R,6S-Tg to the duplex DNA. Three stable duplex structures containing trans-5R,6R-Tg were observed in our MD studies, depending on their surrounding bases and the influence of water. In the replicas of trans-DNA, an arched hydrogen-bonded trans-5R,6R-Tg/A19 pair is present as a metastable structure in the trajectories, preceding each of the stable species. The most stable local structure is unambiguously found in trans-DNA-1, indicating a complex hydrogen bonded network between trans-5R,6R-Tg and the A19 and C20 bases, which points to a clear base sequence effect. The stable local structure is also demonstrated using DFT calculations.

The activation barrier for 5R,6S-Tg flipping out of the duplex DNA in cis-DNA is ca. 4.9 kcal mol\(^{-1}\). This is comparable to the 5.4 kcal mol\(^{-1}\) computed for T6 base flipping in native DNA, showing that cis-5R,6S-Tg is stably positioned in the duplex DNA and will not easily attain an extrahelical position. However, the activation barrier for trans-5R,6R-Tg to flip out of the double helix DNA ranges from 1.0 to 5.2 kcal mol\(^{-1}\), depending on its local structure. Due to the conformational equilibrium of the flexible trans-DNA species in solution, the population of trans-DNA-1 with the most stable local structure should be the smallest due to loss of the classical Watson–Crick hydrogen-bonded base pair structure. Moreover, trans-DNA-1 displayed the smallest barrier height for trans-5R,6R-Tg flipping among the studied trans-DNA species in solution. The results provide detailed structural information on the 5R-Tg epimer in a DNA duplex and can serve as a basis for understanding the recognition of the 5R-Tg epimer by repair enzymes.
In summary, the authors have shown that the base-flipping dynamics of the T/A site in DNA can be accurately modeled using meta-eABF simulations. The results provide insights into the mechanism of base flipping and can help in the development of more effective strategies for DNA repair processes.

The study highlights the potential of combining experimental and computational approaches to gain a deeper understanding of the molecular mechanisms underlying DNA repair, which is crucial for both fundamental research and the development of therapeutic interventions.
(27) Makarova, A. V.; Boldinova, E. O.; Belousova, E. A.; Lavrik, O. I. In vitro lesion bypass by human PrimPol. DNA Repair (Amst) 2018, 70, 18–24.

(28) Naldiga, S.; Huang, H.; Greenberg, M. M.; Basu, A. K. Mutagenic Effects of a 2-Deoxyribonolactone-Thymine Glycol Tandem DNA Lesion in Human Cells. Biochem. 2020, 59, 417–424.

(29) Brown, K. L.; Rognijskaya, M.; Zou, Y.; Altamirano, A.; Basu, A. K.; Stone, M. P. Binding of the human nucleotide excision repair proteins XPA and XPC/HR23B to the SR-thymine glycol lesion and structure of the cis-(SR,6S) thymine glycol epimer in the S-GTgG3‘ sequence: destabilization of two base pairs at the lesion site. Nucleic Acids Res. 2010, 38, 428–440.

(30) Imamura, K.; Averill, A.; Wallace, S. S.; Double, S. Structural characterization of viral ortholog of human DNA glycosylase NEL1 bound to thymine glycol or 5-hydroxyuracil-containing DNA. J. Biol. Chem. 2012, 287, 4288–4298.

(31) Yin, Y.; Yang, L.; Zheng, G.; Gu, C.; Yi, C.; He, C.; Gao, Y. Q.; Zhao, X. S. Dynamics of spontaneous flipping of a mismatched base in DNA duplex. Proc. Natl. Acad. Sci. U. S. A. 2014, 111, 8043–8048.

(32) Roberts, R. J. ON BASE FLIPPING. Cell 1995, 82, 9–12.

(33) Banerjee, A.; Yang, W.; Karplus, M.; Verdine, G. L. Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA. Nature 2005, 434, 612–618.

(34) Vannai, P.; Lavery, R. Base flipping in DNA: Pathways and energetics studied with molecular dynamic simulations. J. Am. Chem. Soc. 2002, 124, 7272–7273.

(35) The PyMOL Molecular Graphics System, version 1.3; Schrodinger: New York, 2010.

(36) Harrach, M. F.; Drossel, B. Structure and dynamics of TIP3P, TIP4P, and TIP5P water near smooth and atomistic walls of different hydrophilicity. J. Phys. Chem. 2014, 118, 174501.

(37) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald - an Nlog(N) method for Ewald sums in large systems. J. Chem. Phys. 1993, 98, 10089–10092.

(38) Miyamoto, S.; Kollman, P. A. SETTLE - an analytical version of the shake and rattle algorithm for rigid water models. J. Comput. Chem. 1992, 13, 952–962.

(39) Feller, S. E.; Zhang, Y. H.; Pastor, R. W.; Brooks, B. R. Constant-pressure molecular-dynamics simulation - the Langevin piston method. J. Chem. Phys. 1995, 103, 4613–4621.

(40) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant-pressure molecular-dynamics algorithms. J. Chem. Phys. 1994, 101, 4177–4189.

(41) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. J. Mol. Graph 1996, 14, 33.

(42) Lavery, R.; Moakher, M.; Maddocks, J. H.; Petkeviciute, D.; Zakrzewska, K. Conformational analysis of nucleic acids revisited: Curves+. Nucleic Acids Res. 2009, 37, 5917.

(43) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. J. Comput. Chem. 2005, 26, 1781–1802.

(44) Fiorin, G.; Klein, M. L.; Henin, J. Using collective variables to drive molecular dynamics simulations. Mol. Phys. 2013, 111, 3345–3362.

(45) 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. 2012, 8, 348–362.

(46) Lesage, A.; Lelièvre, T.; Stoltz, G.; Henin, J. Smoothed Binning Forces Yield Unbiased Free Energies with the Extended-System Adaptive Biasing Force Method. J. Phys. Chem. B 2017, 121, 3676–3685.

(47) Fu, H.; Shao, X.; Chipot, C.; Cai, W. Extended Adaptive Binning Force Algorithm. An On-the-Fly Implementation for Accurate Free-Energy Calculations. J. Chem. Theory Comput. 2016, 12, 3506–3513.
Repair of thymine glycol by hNth1 and hNeil1 is modulated by base pairing and cis-trans epimerization. DNA Repair (Amst) 2006, 5, 444−454.

Recognition of DNA adducts by edited and unedited forms of DNA glycosylase NEIL1. DNA Repair (Amst) 2020, 85, 102741.

Protein-facilitated base flipping in DNA by cytosine-5-methyltransferase. Proc. Natl. Acad. Sci. U. S. A. 2003, 100, 68−73.

Nucleic acid base pair dynamics: The impact of sequence and structure using free-energy calculations. J. Am. Chem. Soc. 2003, 125, 4998−4999.

Molecular dynamics study of the recognition of ATP by nucleic acid aptamers. Nucleic Acids Res. 2020, 48, 6471−6480.

DNA Base Pair Mismatches Induce Structural Changes and Alter the Free-Energy Landscape of Base Flip. J. Phys. Chem. B 2018, 122, 12251−12259.

Free Energy Coupling between DNA Bending and Base Flipping. J. Chem. Inf. Model. 2017, 57, 2020−2026.

An Improved Reaction Coordinate for Nucleic Acid Base Flipping Studies. J. Chem. Theory Comput. 2009, 5, 3105−3113.

Oxidatively induced DNA damage and its repair in cancer. Mutat. Res. 2015, 763, 212−245.

Oxidatively induced DNA damage: mechanisms, repair and disease. Cancer Lett. 2012, 327, 26−47.

Stereoselective excision of thymine glycol from oxidatively damaged DNA. Nucleic Acids Res. 2004, 32, 338−345.

Base-flipping mechanism in postmismatch recognition by MutS. Biophys. J. 2011, 101, 2223−2231.

Laveringe coupling between DNA bending and base pair opening. Proc. Natl. Acad. Sci. U. S. A. 1988, 85, 7231−7235.