9722 *Biochemistry* **2009,** *48,* 9722–9733
DOI: 10.1021/bi900695e

BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

The *cis*-(5*R*,6*S*)-Thymine Glycol Lesion Occupies the Wobble Position When Mismatched with Deoxyguanosine in DNA[†]

Kyle L. Brown,[‡] Ashis K. Basu,[§] and Michael P. Stone*[,‡]

[‡]*Department of Chemistry, Center in Molecular Toxicology, and Vanderbilt-Ingram Cancer Center, Vanderbilt University, Nashville, Tennessee 37235, and* [§]*Department of Chemistry, University of Connecticut, Storrs, Connecticut 06269*

Received April 22, 2009; Revised Manuscript Received August 7, 2009

ABSTRACT: Oxidative damage to 5-methylcytosine in DNA, followed by deamination, yields thymine glycol (Tg), 5,6-dihydroxy-5,6-dihydrothymine, mispaired with deoxyguanosine. The structure of the 5*R* Tg·G mismatch pair has been refined using a combination of simulated annealing and isothermal molecular dynamics calculations restrained by NMR-derived distance restraints and torsion angle restraints in 5′-d($G^1T^2G^3C^4G^5\underline{Tg}^6G^7T^8T^9T^{10}G^{11}T^{12}$)-3′·5′-d($A^{13}C^{14}A^{15}A^{16}A^{17}C^{18}\underline{G}^{19}C^{20}G^{21}C^{22}A^{23}C^{24}$)-3′; Tg = 5*R* Tg. In this duplex the *cis*-5*R*,6*S*:*trans*-5*R*,6*R* equilibrium favors the *cis*-5*R*,6*S* epimer [Brown, K. L., Adams, T., Jasti, V. P., Basu, A. K., and Stone, M. P. (2008) *J. Am. Chem. Soc. 130*, 11701−11710]. The *cis*-5*R*,6*S* Tg lesion is in the wobble orientation such that Tg^6 *O*^{2} is proximate to G^{19} N1H and Tg^6 N3H is proximate to G^{19} *O*^{6}. Both Tg^6 and the mismatched nucleotide G^{19} remain stacked in the helix. The Tg^6 nucleotide shifts toward the major groove and stacks below the 5′-neighbor base G^5, while its complement G^{19} stacks below the 5′-neighbor C^{20}. In the 3′-direction, stacking between Tg^6 and the G^7·C^{18} base pair is disrupted. The solvent-accessible surface area of the Tg nucleotide increases as compared to the native Watson−Crick hydrogen-bonded T·A base pair. An increase in T_2 relaxation rates for the Tg^6 base protons is attributed to puckering of the Tg base, accompanied by increased disorder at the Tg·G mismatch pair. The axial vs equatorial conformation of the Tg^6 CH_3 group cannot be determined with certainty from the NMR data. The rMD trajectories suggest that in either the axial or equatorial conformations the *cis*-5*R*,6*S* Tg lesion does not form strong intrastrand hydrogen bonds with the imidazole N7 atom of the 3′-neighbor purine G^7. The wobble pairing and disorder of the Tg·G mismatch correlate with the reduced thermodynamic stability of the mismatch and likely modulate its recognition by DNA base excision repair systems.

The common thymine oxidation product in DNA, 5,6-dihydroxy-5,6-dihydrothymine, known as thymine glycol (Tg)[1], is formed by exposure to ionizing radiation, as well as a variety of chemical oxidizing agents (*1*, *2*). Tg is also formed by oxidation of 5-methylcytosine to 5-methylcytosine glycol, followed by hydrolytic deamination (*3*, *4*). Once formed, the C5 and C6 atoms in Tg are chiral, and thus Tg exists in DNA as two diastereomeric pairs of epimers, the 5*R* *cis*, *trans* pair (5*R*,6*S*;5*R*,6*R*) and the 5*S* *cis*, *trans* pair (5*S*,6*R*; 5*S*,6*S*) (*5*−*7*) (Chart 1). The 5*R* pair is more abundant and more stable (*6*). Tg has been detected in animal and human urine (*8*, *9*). It is estimated that human cells repair hundreds of thymine glycol lesions per day (*8*, *9*).

For both the 5*R* and 5*S* pairs of epimers, the *cis* isomers predominate at the nucleoside level (*6*). For the 5*R* pair of epimers, the rate of epimerization at the nucleoside level is 5.8×10^{-3} min^{-1} (*6*). In duplex DNA, the *cis*−*trans* equilibrium of the 5*R* Tg isomer is modulated by the identity of the complementary base (*10*). When the 5*R* Tg isomer is base paired opposite dA in the 5′-GTgG-3′ sequence, a 7:3 *cis* (5*R*,6*S*):*trans* (5*R*,6*R*) mixture exists at equilibrium at 25 °C (*10*). However, when the 5*R* Tg lesion is mismatched with dG in the same sequence, only the *cis* (5*R*,6*S*) epimer is observed by NMR (*10*). The structure of the 5*R* Tg lesion has been examined in the 5′-ATgA-3′ sequence, paired opposite dA (*11*). This study concluded that Tg induced a localized structural perturbation and, based upon calculation of solvent-accessible surface area, concluded that Tg was partially extrahelical (*11*). It was also reported that the structure of the 5*R* Tg lesion in the 5′-GTgC-3′ sequence was disordered (*12*). Neither of the earlier studies (*11*, *12*) considered the *cis*−*trans* epimerization of the Tg lesion.

In the present work, the 5*R* Tg lesion has been incorporated site-specifically into 5′-d($G^1T^2G^3C^4G^5Tg^6G^7T^8T^9T^{10}G^{11}T^{12}$)-3′·5′-d($A^{13}C^{14}A^{15}A^{16}A^{17}C^{18}\underline{G}^{19}C^{20}G^{21}C^{22}A^{23}C^{24}$)-3′, forming the mismatched Tg^6·$\underline{G}^{19}$ base pair. This corresponds to the

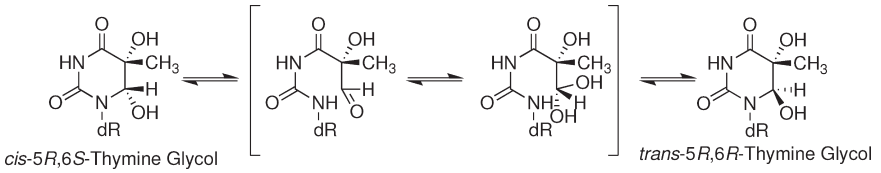
[†]This work was funded by NIH Grants CA-55678 (M.P.S.) and ES-013324 (A.K.B.).

*To whom correspondence should be addressed. Telephone: 615-322-2589. Fax: 615-322-7591. E-mail: michael.p.stone@vanderbilt.edu.

[1]Abbreviations: BER, base excision repair; COSY, correlation spectroscopy; EDTA, ethylenediaminetetraacetic acid; EOS, electronegativity of substituent; ESP, electrostatic potential; N, northern sugar pucker; HPLC, high-pressure liquid chromatography; NER, nucleotide excision repair; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; PME, particle mesh Ewald; rMD, restrained molecular dynamics; RESP, restrained electrostatic potential; rmsd, root-mean-square deviation; S, southern sugar pucker; SAS, solvent-accessible surface; Tg, thymine glycol, 5,6-dihydroxy-5,6-dihydrothymine; TPPI, time-proportional phase increment; Tg·G, the underlined Tg·G base pair in the oligodeoxynucleotide 5′-d($G^1T^2G^3C^4G^5Tg^6G^7T^8T^9T^{10}G^{11}T^{12}$)-3′·5′-d($A^{13}C^{14}A^{15}A^{16}A^{17}C^{18}\underline{G}^{19}C^{20}G^{21}C^{22}A^{23}C^{24}$)-3′. The oligodeoxynucleotides discussed in this paper do not have terminal phosphate groups; we abbreviate the nomenclature by leaving out the phosphodiester linkage. A, C, G, T, and Tg refer to mononucleotide units. A right superscript refers to numerical position in the oligodeoxynucleotide sequence starting from the 5′-terminus of chain A and proceeding to the 3′-terminus of chain A and then from the 5′-terminus of chain B to the 3′-terminus of chain; C2, C5, C6, C8, C1′, C2′, C2″, etc. represent specific carbon nuclei. H2, H5, H6, H8, H1′, H2′, H2″, etc. represent the protons attached to these carbons.

 Published on Web 09/22/2009

Chart 1: Interconversion of the *cis*-5*R*,6*S*- and *trans*-5*R*,6*R*-Tg Lesions



situation resulting from oxidative damage to 5-methylcytosine in DNA, followed by deamination to form Tg. In this oligodeoxynucleotide when the 5*R* Tg lesion is mismatched with dG, only the *cis* (5*R*,6*S*) epimer is observed (*10*), thus facilitating structural refinement. The structure of the *cis*-5*R*,6*S* Tg lesion in the Tg·G mismatch pair reveals that both Tg^6 and the mismatched G^{19} remain inserted into the duplex such that the Tg^6 base assumes a wobble orientation with G^{19} and shifts toward the major groove. The solvent-accessible surface area of the Tg^6 nucleotide increases as compared to the native T·A base pair. An increase in T_2 relaxation rates for the Tg^6 base protons suggests puckering of the Tg base, accompanied by increased disorder at the $Tg^6 \cdot G^{19}$ mismatch pair. The *cis*-5*R*,6*S* Tg lesion does not form strong hydrogen bonds to the 3′-neighbor purine G^5 N7 (*13*). The wobble orientation and disorder of the $Tg^6 \cdot G^{19}$ mismatch likely contribute to its recognition by base excision repair.

MATERIALS AND METHODS

Sample Preparation. The oligodeoxynucleotides 5′-d(GTGCGTGTTTGT)-3′ and 5′-d(ACAAACGCGCAC)-3′ were purchased from the Midland Certified Reagent Co. (Midland, TX). The oligodeoxynucleotide 5′-d(GTGCGTgGTTTGT)-3′, containing the 5*R* Tg lesion, was synthesized as reported (*14*). Oligodeoxynucleotides were purified by reverse-phase HPLC using a Phenomenex Gemini C-18 column (250 mm × 10 mm). The mobile phase was 0.1 M ammonium formate (pH 7.0) with a flow rate of 2 mL/min; elution was performed using a linear gradient from 6% to 30% CH_3CN over 25 min. Oligodeoxynucleotide concentrations were measured by UV absorbance at 260 nm. Complementary oligodeoxynucleotides were annealed in 20 mM sodium phosphate buffer, containing 100 mM NaCl, 10 mM NaN_3, and 50 μM Na_2EDTA (pH 7.0), then heated to 70 °C for 10 min and subsequently allowed to cool slowly to ambient temperature.

Mass Spectrometry. The 5*R* Tg modified oligodeoxynucleotides were characterized by mass spectrometry using a Voyager-DE MALDI-TOF spectrometer (PerSeptive Biosystems, Inc., Foster City CA). Samples were suspended in a matrix consisting of 0.5 M 3-hydroxypicolinic acid in 1:1 CH_3CN:H_2O and spotted onto sample plates. Mass spectra were recorded in the reflector mode using a laser attenuation setting of 3030 units. The accelerating voltage was 20 kV, with a grid voltage of 85.0%, guide wire voltage of 0.050%, and a delay of 100 ns. The mass spectra were averaged from 256 scans.

Capillary Gel Electrophoresis. CGE was performed using approximately 1 nmol of desalted sample. The gels and buffers were prepared using Beckman Coulter ssDNA 100-R kits (Beckman-Coulter, Inc., Fullerton, CA). An injection voltage of 10.0 kV was used for 8 s, followed by a separation voltage of 14.1 kV for 35 min, using a 33 cm capillary. The electropherogram was recorded at λ 254 nm. Elution times were referenced to an internal standard (Beckman "Orange G" product number 241524).

NMR Spectroscopy. NMR experiments were performed at a ^1H frequency of 800 MHz. The duplex concentration was 1.5 mM, and the samples were suspended in 20 mM sodium phosphate, 100 mM NaCl, and 50 μM Na_2EDTA (pH 7.0). To examine nonexchangeable protons, samples were suspended in 99.996% D_2O. The nonexchangeable proton experiments were performed at 30 ± 0.5 °C. NOESY spectra were recorded using States−TPPI phase cycling. Typical acquisition parameters were 512 real data points in the d_1 dimension with 32 scans per FID, 2K real data points in the d_2 dimension, sweep width of 10 ppm, and a relaxation delay of 2.0 s. The residual water resonance was suppressed using presaturation. For observation of exchangeable protons, samples were suspended in 9:1 H_2O:D_2O. The experiments were performed at 5 ± 0.5 °C. NOESY spectra were obtained using watergate H_2O suppression (*15*) and a sweep width of 20 ppm. The binomial water suppression delay of 188 ms at 800 MHz avoided suppression of amino resonances. The T_1 spin−lattice relaxation experiments were collected using the inversion recovery method (*16*, *17*). The inversion recovery profile for T_1 fitting included delay times of 0.01, 0.05, 0.1, 0.25, 0.5, 1, 2, 3, 4, 5, 7, and 10 s. The T_2 transverse relaxation experiments were collected using the Carr−Purcell−Meiboom−Gill (CPMG) method (*16*, *17*). The CPMG spin echo time increment profile included values of 4, 6, 8, 10, 12, 14, 16, 18, and 20 s. ^1H spectra were referenced to internal 3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid sodium salt (3-TMSP). The programs XWINNMR (version 3.5, patch level 6; Bruker Inc., Billerica, MA) and NMRPipe (*18*) were used for data processing. Skewed sine-bell-squared apodization functions were used.

Distance Restraints. NOE intensities were determined from integration of peak volumes by a Gaussian fit function using the program SPARKY v3.11 (*19*). Experimentally determined intensities were combined with intensities generated from complete relaxation matrix analysis of a B-form starting structure of DNA to produce a hybrid intensity matrix (*20*, *21*). The program MARDIGRAS (*22*−*24*), using the RANDMARDI function, was used to iteratively refine the hybrid intensity matrix and to optimize the agreement between the experimental and calculated intensities. Calculations at mixing times of 80, 150, 200, and 250 ms were each run at isotropic correlation times of 2, 3, 4, and 5 ns with a canonical B-form starting structure, producing 16 sets of distances. Each set was a product of 50 RANDMARDI iterations. A refined structure taken from initial rounds of simulated annealing rMD calculations was used as a starting structure for the second round of MARDIGRAS interproton distance calculations. The results were indexed to short- and long-range distances typical for both A- and B-form DNA as a means to assess the accuracy of resultant distance sets. The data that best matched canonical DNA distances were averaged and used in subsequent rMD calculations. The distance restraints were divided into five categories indicative of the restraint confidence level. In addition, empirical distance restraints were used to define hydrogen bonding of complementary bases. Hydrogen-bonding restraints were not used for the $Tg^6 \cdot G^{19}$ or

$G^5 \cdot C^{20}$ base pairs because wobble or Watson−Crick hydrogen bonding could not be confirmed by NMR experiments (*10*).

Torsion Angle Restraints. Deoxyribose pseudorotation was determined by fitting ^3J ^{1}H coupling constants for deoxyribose protons by amplitude-constrained multiplet evaluation (ACME) of fully relaxed COSY data (*25*). Electronegativity of substituent (EOS) Karplus curves were generated and converted to phase angle space assuming a maximum pucker amplitude (Φ) of 44 (*26*, *27*). $J_{1'2'}$, $J_{1'2''}$, and $J_{1'3'}$ were fit to the curve to determine phase angle ranges (r) for deoxyribose rings. The sugar pseudorotation and amplitude ranges were converted to upper and lower bound restraints for the five dihedral angles ν_0 to ν_4. Deoxyribose pucker conformations were confirmed to be N or S by an approximate measurement of the mol fraction of sugar pucker in the S configuration (X_S) determined from the sum of $J_{1'2'}$ and $J_{1'2''}$ (*27*). Residues that had less than 50% X_S or indicated potential for the C3′ endo conformation from EOS Karplus curves were constrained such that they were allowed to explore both N and S conformations during rMD calculations (ρ = 0−210). Residues that had X_S greater than 50% were restrained such that ρ = 125−210. Backbone torsion angles were restrained with experimental data where available. The remaining angles were restrained based on canonical A-form/B-form ranges.

Tg Parametrization. Partial charges for the Tg nucleotide were generated using GAUSSIAN 03 (*28*). Geometry optimization and molecular electrostatic potential refinement were performed using the Hartree−Fock method with the 6-31G* basis set (*10*). Potential points were calculated with a density of six points per unit area in the electrostatic potential (ESP) fit. Calculations were performed with a fixed net charge of −1. Convergence was verified by frequency analysis. The ESP output from the GAUSSIAN calculation was formatted using the program ANTECHAMBER (*29*).

Structural Refinement. Thirty starting structures of the oligodeoxynucleotide duplex containing the $Tg^6 \cdot G^{19}$ mismatch were generated using NAB (Nucleic Acid Builder) (*30*). These were energy minimized with the steepest descent algorithm for 100 steps and the conjugate gradient algorithm for 900 steps using the SANDER module of AMBER 9 (*29*). During energy minimization, heavy atoms of the nucleic acid bases were held fixed. The coordinate and topology files were generated with xLEaP (*29*) using ff99 parameters (*31*). The restraint energy function included terms that defined distance and dihedral restraints as square-well potentials (*32*). The generalized Born solvent model was used for simulated annealing, with a salt concentration of 0.2 mM (*33*, *34*). Twenty picosecond simulated annealing rMD simulations were conducted with a nonbonded interaction cutoff at 15 Å. The structure coordinates were recorded every 1 ps. Temperature was maintained using the Berendsen algorithm (*35*). The protocol utilized a starting temperature of 600 K. From 5 to 18 ps the temperature was decreased to 100 K. The temperature was decreased to 0 K during the final 2 ps of the simulations. A time constant of 0.4 ps was used for heat bath coupling during the first 0.5 ps of the simulation. From 0.5 to 18 ps the heat bath coupling was decreased to 4 ps; from 18 to 19 ps the coupling was increased to 1 ps. During the final 19 to 20 ps the coupling was increased to 0.05 ps. During the first 3 ns the relative weights of the NMR restraint energy terms were increased from 0.1 to 1; this was maintained for the remaining 17 ns.

The average structure from the final round of simulated annealing calculations was neutralized with the addition of 24 sodium ions and placed in a truncated octahedral TIP3P water box with periodic boundaries at a distance of 8 Å from the solute. Twenty-two sodium ions were restrained to the backbone with a lower bound of 3.0 Å and an upper bound of 8.0 Å while the remaining two sodium ions were not restrained. The solvated system was energy minimized using 500 steps of the steepest descent algorithm followed by 500 steps of the conjugate gradient algorithm at constant volume, with the solute held fixed by positional restraints. Next, the system was energy minimized using 2500 steps of the conjugate gradient algorithm with no positional restraints at a constant volume. The system was subsequently heated to 300 K over 100 ps; the weight of the NMR and empirical restraints was slowly increased during the heating period. A 10 ns rMD simulation at 300 K was then performed. Throughout the equilibration and production periods, the temperature was held constant using the Langevin thermostat (*36*, *37*) with a collision frequency of 1 ps^{-1}. Electrostatic interactions were treated with the particle mesh Ewald (PME) method (*38*). A 15 Å cutoff for nonbonded interactions was used. Bond lengths involving hydrogen were held fixed using the SHAKE algorithm (*39*).

Structural Analysis. The heavy atoms of the output structures from the isothermal rMD calculations were subjected to pairwise rmsd comparisons using SUPPOSE. Trajectories were analyzed with PTRAJ. Complete relaxation matrix analysis (CORMA, v.4) (*20*, *21*) was performed on the structural ensembles, assuming equal occupancy, to determine agreement with experimental ^{1}H NOESY data. Helicoidal analysis was performed using the program CURVES (*40*) on an average of each structural ensemble. Graphic representations of structures were rendered using the program CHIMERA (*41*). Solvent-excluded and solvent-accessible surface areas of nucleotides were measured using the program Maximal Speed Molecular Surface (MSMS) (*42*). A series of calculations employing probe radii of 1.5, 2.5, 3.5, and 4.5 Å were used. The percentages of solvent-accessible surfaces were rendered as a function of probe radii.

RESULTS

Characterization of the Tg·G Mismatch Pair. The modified dodecamer 5′-d(GTGCGTgGTTTGT)-3′, Tg = 5*R* Tg, was subjected to mass spectrometric analysis, which yielded the anticipated molecular ion peak with mass 3732 (m/z). Further capillary gel electrophoretic and HPLC analyses showed that the modified oligodeoxynucleotide consisted of a chromatographically separable species of high purity. NMR data were collected upon sample preparation and were repeated after 4 weeks; changes in the spectrum were not observed, suggesting that the sample had achieved equilibrium with respect to the *cis*-5*R*,6*S*: *trans*-5*R*,6*R* equilibrium. The *cis*-5*R*,6*S* epimer was favored, with the *trans*-5*R*,6*R* epimer remaining below the level of detection (*10*).

NMR Spectroscopy. (*a*) *Nonexchangeable DNA Protons*. The NMR resonances for the oligodeoxynucleotide duplex containing the Tg·G mismatch pair were assigned using standard strategies (*43*, *44*) (Figure 1). The deoxyribose protons were assigned from a combination of COSY and NOESY data. With the exception of several of the H4′ protons, and the stereotopic assignments of the H5′ and H5″ sugar protons, assignments were made unequivocally.

(*b*) *Exchangeable DNA Protons*. At the mismatched $Tg^6 \cdot G^{19}$ base pair, the Tg^6 N3H imino resonance was not

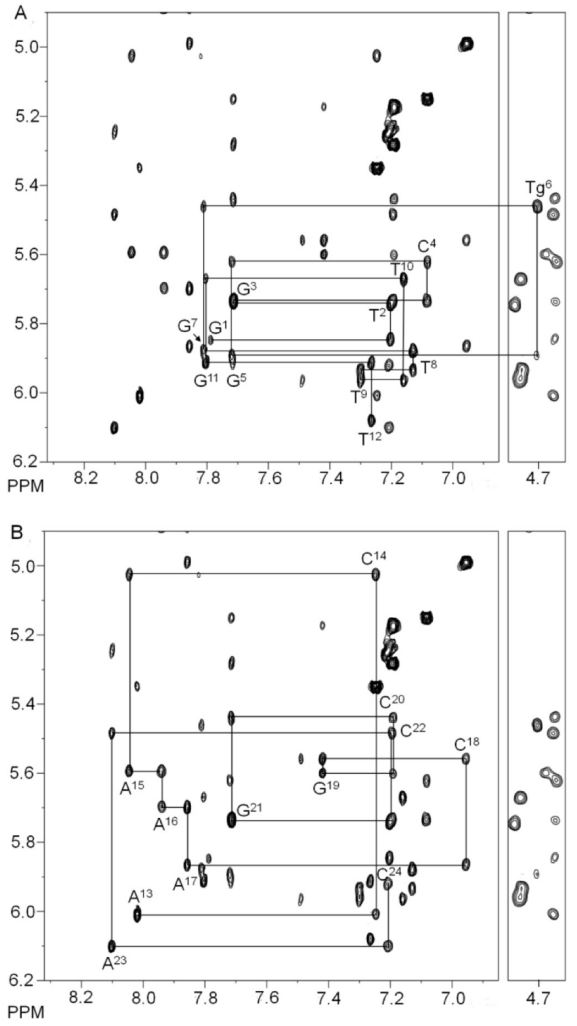


FIGURE 1: Sequential NOESY assignments of purine H8 and pyrimidine H6 protons to deoxyribose H1′ protons for the duplex containing the $Tg^6 \cdot G^{19}$ base pair. (A) The modified strand of the duplex containing the $Tg^6 \cdot G^{19}$ base pair. (B) The complementary strand of the duplex containing the $Tg^6 \cdot G^{19}$ base pair. The intranucleotide aromatic to H1′ cross-peaks are labeled.

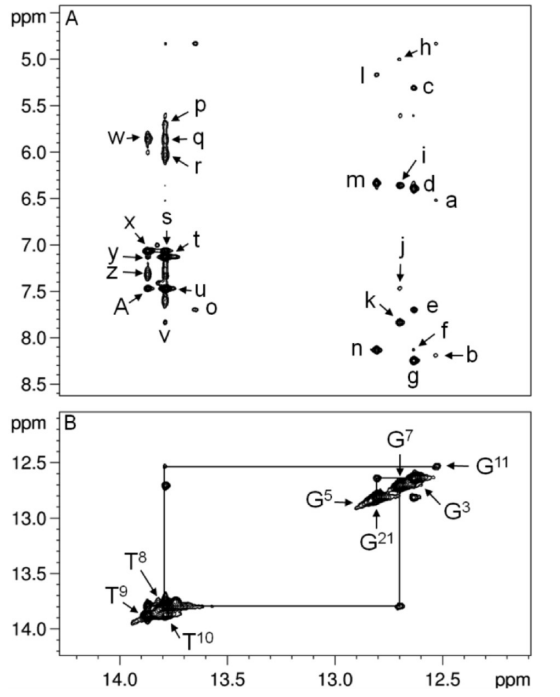


FIGURE 2: (A) Expanded plot showing NOEs from the imino protons to amino protons for the duplex containing the $Tg^6 \cdot G^{19}$ base pair. The cross-peaks are assigned as follows: a, G^{11} N^2H2→G^{11} N1H; b, C^{14} N^4H2→G^{11} N1H; c, C^{22} H5→G^3 N1H; d, G^3 N^2H2→G^3 N1H; e, A^{23} H2→G^3 N1H; f, C^4 N^4H2→G^3 N1H; g, C^{22} N^4H2→G^3 N1H; h, C^{18} H5→G^7 N1H; i, G^7 N^2H2→G^7 N1H; j, A^{17} H2→G^7 N1H; k, C^{18} N^4H2→G^7 N1H; l, C^4 H5→G^{21} N1H; m, G^{21} N^2H2→G^{21} N1H; n, C^4 N^4H2→G^{21} N1H; o, A^{23} H2→T^2 N3H; p, A^{17} N^6H2→T^8 N3H; q, A^{16} N^6H2→T^8 N3H; r, A^{15} N^6H2→T^{10} N3H; s, A^{16} H2→T^{10} N3H; t, A^{15} H2→T^{10} N3H; u, A^{17} H2→T^8 N3H; v, C^{18} N^4H2→T^8 N3H; w, A^{16} N^6H2→T^9 N3H; x, A^{16} H2→T^9 N3H; y, A^{15} H2→T^9 N3H; z, A^{16} N^6H1→T^9 N3H; A, A^{17} H2→T^9 N3H. (B) Expanded plot showing the sequential NOE connectivity for the imino protons of the duplex containing the $Tg^6 \cdot G^{19}$ base pair. The data were collected at 800 MHz at a 250 ms mixing time and a temperature of 7 °C.

identified (Figure 2). This was attributed to rapid exchange with solvent. The assignments of the remaining Watson−Crick hydrogen-bonded imino and amino protons were made using standard methods (*45*). The G^5 N1H imino resonance was broad at 5 °C and disappeared when the temperature was increased to 15 °C. In contrast, for the unmodified sample, the G^5 N1H imino resonance was sharp and was observed at temperatures as high as 40 °C (*10*). There was no cross-peak between the broad G^5 N1H resonance and G^{21} N1H, located at base pair $C^4 \cdot G^{21}$. This was attributed to its exchange with solvent. The imino resonances for base pairs $T^2 \cdot A^{23}$, $G^3 \cdot C^{22}$, $C^4 \cdot G^{21}$, $G^7 \cdot C^{18}$, $T^8 \cdot A^{17}$, $T^9 \cdot A^{16}$, $T^{10} \cdot A^{15}$, and $G^{11} \cdot C^{14}$ were observed. The imino resonances for the terminal base pairs $G^1 \cdot C^{24}$ and $T^{12} \cdot A^{13}$ were not observed, attributed to exchange broadening with water.

(*c*) *Tg Protons*. Analysis of NOESY data obtained at multiple mixing times did not reveal chemical exchange cross-peaks for either the Tg^6 CH_3 and Tg^6 H6 protons, consistent with the conclusion that only one epimer was significantly populated in the sample (*10*) (Figure 3). For the *cis*-5*R*,6*S* configuration, Tg^6 H6 and Tg^6 CH_3 are spatially proximate, yielding a strong Tg^6 H6→Tg^6 CH_3 NOE even at the short mixing time of 150 ms. Likewise, the G^5 H1′→Tg^6 H6 and G^5 H8→Tg^6 H6 NOEs were

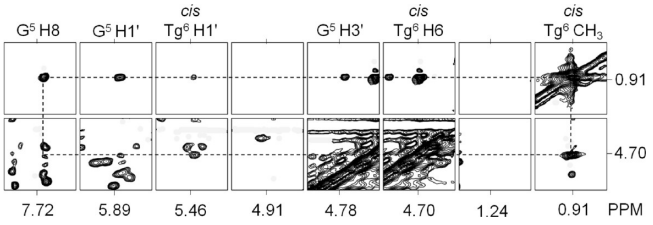


FIGURE 3: NOESY data collected for the duplex containing the $Tg^6 \cdot G^{19}$ base pair showing NOEs between Tg^6 CH_3 and DNA protons. The 800 MHz NMR data were collected at a NOE mixing time of 250 ms. The spectra were collected at 30 °C.

diagnostic of the *cis*-5*R*,6*S* configuration. The Tg^6 H3′ and Tg^6 H6 resonances were observed at 4.53 and 4.70 ppm, respectively. The Tg^6 CH_3 protons exhibited a chemical shift of 0.91 ppm. A total of 9 NOE cross-peaks were assigned between Tg^6 H6 and DNA, and 11 were assigned between Tg^6 CH_3 and DNA. These restraints used in the rMD calculations are summarized in Table 1.

Spin−Lattice Relaxation. The T_1 relaxation times of the T^2, T^8, T^9, T^{10}, and T^{12} thymine and Tg^6 CH_3 groups were compared to those of the corresponding unmodified duplex (Figure 4). The T_1 relaxation time of Tg^6 CH_3 was 1.2 s faster than was the corresponding relaxation time in the unmodified $T^6 \cdot G^{19}$ context.

Table 1: NOESY Cross-Peaks between Tg and DNA Protons

Tg H6	G^5 H1′, G^5 H2′, G^5 H2″, G^5 H8, G^7 H8
Tg CH_3	G^5 H1′, G^5 H2′, G^5 H2″, G^5 H3′, G^5 H4′
Tg H1′	G^5 H1′
Tg H2′	
Tg H2″	G^7 H8
Tg H3′	G^7 H8
Tg H4′	G^7 H8
Tg H5′	

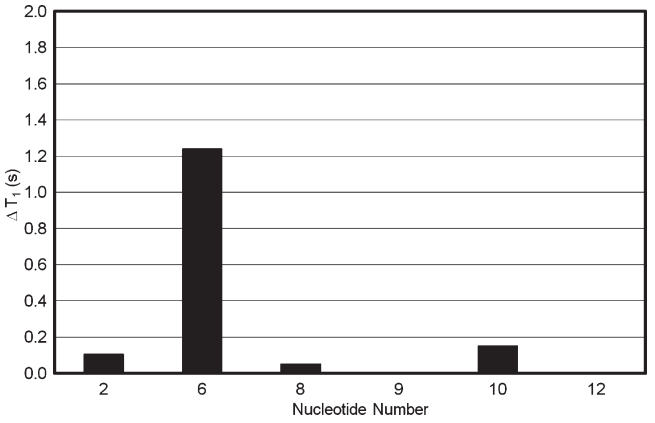


FIGURE 4: Changes in spin−lattice T_1 relaxation times of thymine and Tg^6 CH_3 of Tg·G (■) relative to unmodified T·A duplex.

The Tg^6 CH_3 relaxed an average of 1.0 s faster than other thymine CH_3 groups in the Tg^6·G^{19} duplex.

Spin−Spin Relaxation. The T_2 relaxation times of the T^2, T^8, T^9, T^{10}, and T^{12} thymine and Tg^6 CH_3 groups were compared to those of the corresponding unmodified duplex (Figure 5). The Tg^6 CH_3 of the Tg^6·G^{19} duplex relaxed 80 ms faster than the corresponding CH_3 of the unmodified sample. The Tg^6 CH_3 relaxed an average of 50 ms faster than did other CH_3 groups in the Tg^6·G^{19} duplex.

Restrained Molecular Dynamics. The experimental restraints used in the rMD calculations are summarized in Table 2. The potential for interstrand hydrogen bonding between Tg^6 and G^{19} was not enforced with distance restraints during the rMD simulation, consistent with the experimental NMR data showing that the Tg^6 N3H and G^{19} N1H imino resonances were not observed. Thirty starting structures were generated. Fifteen of these had the Tg CH_3 group in the axial position relative to the Tg base; in the other half the Tg CH_3 group was equatorial. These starting structures exhibited a maximum pairwise rmsd of 4.3 Å. After simulated annealing rMD refinement, all but 2 of the 30 starting trajectories resulted in Tg^6 CH_3 being in an axial conformation. A representative structure from simulated annealing with an axial Tg^6 CH_3 was placed in a truncated octahedron TIP3P water box for additional refinement at 300 K.

An overall CORMA $R_1{}^X$ value of 9.25×10^{-2} indicated that the input coordinates for explicit solvent rMD calculations were in agreement with NMR distance restraints. At 0.88 ns into the isothermal rMD calculations in explicit solvent, the Tg^6 CH_3 shifted from the axial orientation to the equatorial conformation and remained in the equatorial conformation for the remainder of the trajectory (Figure 6). The Tg CH_3−C5−C6−H6 torsion angle shifted from −40° to 50° (Figure 6, panel A). The glycosyl Tg^6 O4′−H1′−N1−C2 torsion angle fluctuated between −100° and −160° during the first 0.88 ns; upon repuckering, the lower limit of fluctuation decreased from −100° and −120° (Figure 6, panel B). To examine the potential for intrastrand hydrogen bonding between the Tg^6 hydroxyl groups and G^7 N7 (*13*), the Tg^6 HO5→G^7 N7 (Figure 6, panel C) and Tg^6 HO6→G^7 N7 (Figure 6, panel D) distances were tracked during the rMD trajectory; hydrogen bond occupancy was defined as occurring when the hydrogen bond donors and acceptors were within 3.5 Å with an angle cutoff of 120°. During the first 0.88 ns of the rMD trajectory the Tg^6 CH_3 axial ring pucker resulted in a 6 Å Tg^6 HO5→G^7 N7 distance and prevented hydrogen bond formation. When Tg^6 CH_3 reoriented to the equatorial conformation, the Tg^6 OH5 became axial and exhibited improved geometry for hydrogen bond formation with G^7 N7. The trajectory analysis indicated 16% occupancy for this motif over 10 ns. In contrast, the Tg^6 HO6→G^7 N7 hydrogen bond was only satisfied during the first 0.88 ns of the simulation when Tg^6 CH_3 was axial (Figure 6, panel D). Its occupancy was 2.5%. This hydrogen bond induced altered propeller twist at the lesion site. As a measure of base−base opening, the Tg^6 N3→G^{19} N1 distance was monitored during the rMD calculations. During the first 0.88 ns while Tg^6 CH_3 was in the axial conformation, the distance between Tg^6 H3→G^{19} N1 was approximately 7.5 Å. Upon repuckering the Tg^6 H3→G^{19} H1 distance decreased to approximately 5.5 Å. This indicated that the Tg^6·G^{19} base−base opening decreased upon repuckering to the Tg^6 CH_3 equatorial conformation and was consistent with helicoidal analysis (Table S1 in the Supporting Information). The root-mean-square deviation of DNA backbone heavy atoms was approximately 2.5 Å for the simulation (Figure 6, panel F).

Two ensembles of structures were extracted from the rMD data (Figure 7): one in which the Tg^6 CH_3 was in the axial conformation was extracted from the trajectory at approximately 120 ps before the axial−equatorial conformational shift; this ensemble

Table 2: Restraint Summary for Solution Structure Determination

resonances	220
restraints	
Watson−Crick	40
backbone torsion angle	134
sugar torsion angle	120
distance	384
internucleotide	189
intranucleotide	195
total restraints	678
avg restraint per nucleotide	28

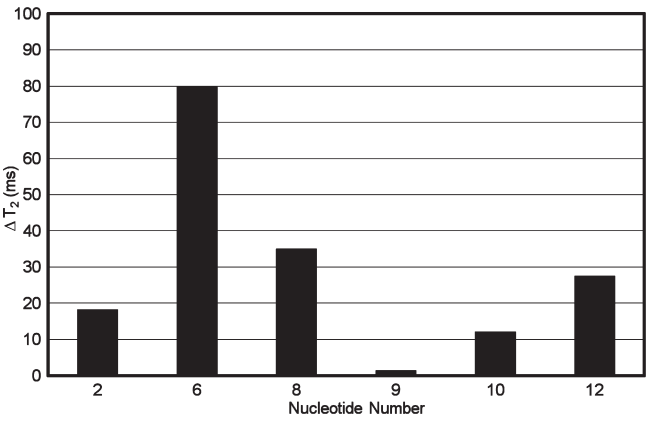


FIGURE 5: Changes in T_2 relaxation of thymine and Tg^6 CH_3 of Tg·G (■) relative to unmodified T·A duplex.

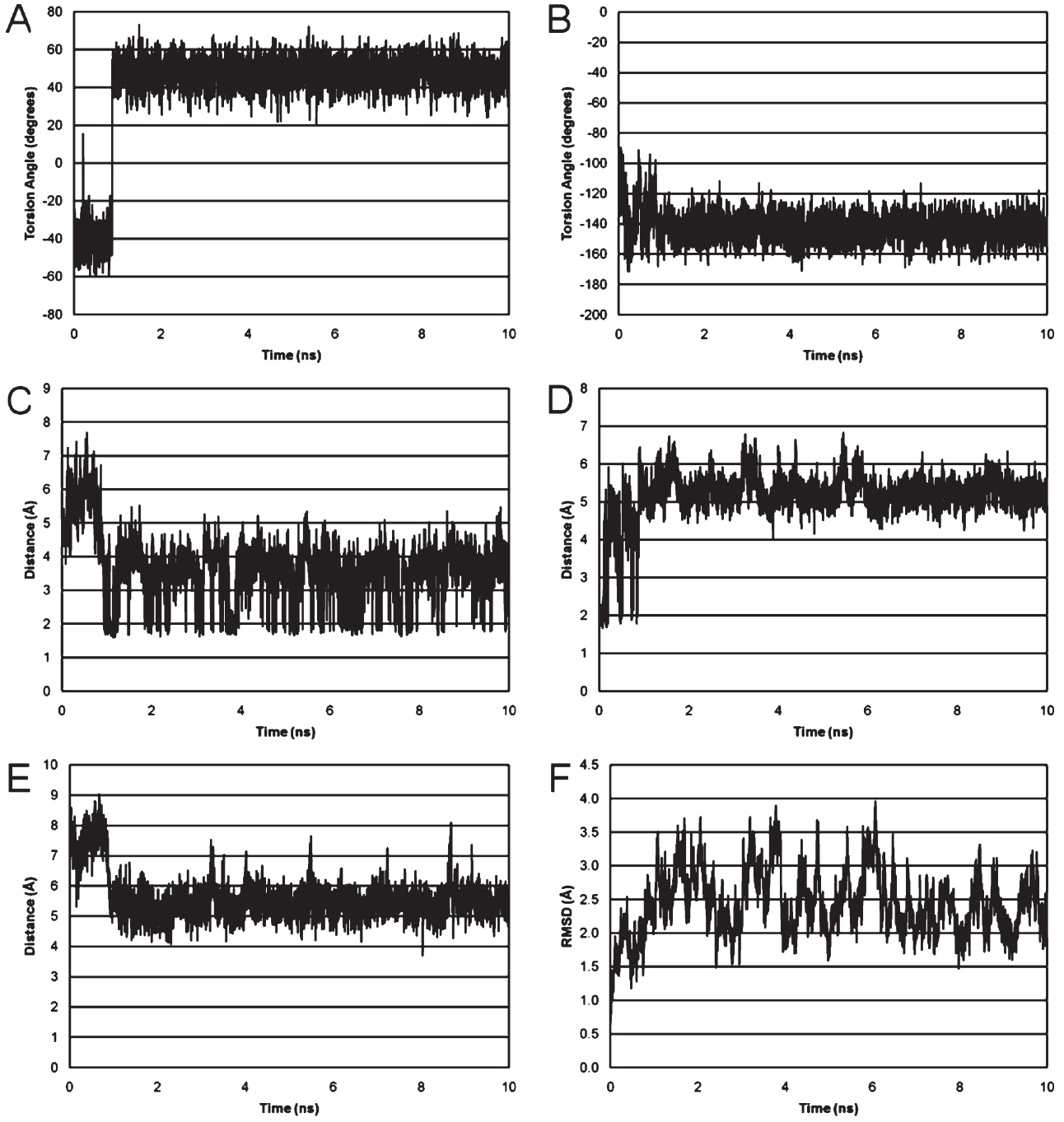


FIGURE 6: Analysis of selected Tg^6 torsion angles and internucleotide distances during the 10 ns rMD trajectories carried out in explicit solvent. (A) The Tg^6 CH_3−C5−C6−H6 torsion angle. (B) The Tg^6 O4′−H1′−N1−C2 torsion angle. (C) The Tg^6 HO5→G^7 N7 distance. (D) The Tg^6 HO6→G^7 N7 distance. (E) The Tg^6 H3→A^{19} N1 distance. (F) The rmsd of backbone heavy atoms for the entire duplex.

had an rmsd value of 0.9 Å for the core six base pairs (PDB ID 2KH7). A second ensemble of structures in which the Tg^6 CH_3 was in the equatorial conformation was extracted approximately 1 ns after the axial−equatorial conformational shift. The latter ensemble had an rmsd value of 0.8 Å for the core six base pairs (PDB ID 2KH8). Thus, both the axial and equatorial conformations of the Tg^6 CH_3 group converged to single ensembles of structures.

To evaluate the agreement of the two structural ensembles emergent from rMD calculations with the NOESY data, complete relaxation matrix analyses (CORMA, v.4) (*20*, *21*) were performed (Figure 8). The calculated sixth root residual R_1^X values indicated satisfactory agreement between both refined ensembles of structures and the ^{1}H NOESY data. The ensemble with Tg^6 CH_3 in an axial position had an average R_1^X value of 1.03×10^{-1} for the entire structure. The Tg^6 lesion itself exhibited an R_1^X value of 8.80×10^{-2} for interresidue cross-peaks with G^5 and 1.04×10^{-1} for intraresidue cross-peaks. The ensemble with Tg^6 CH_3 in the equatorial position had an average R_1^X value of 9.96×10^{-2} for the entire structure. The Tg^6 lesion itself exhibited an R_1^X value of 9.57×10^{-2} for interresidue cross-peaks with G^5 and 1.22×10^{-1} for intraresidue cross-peaks.

Refined Structure of the Tg·G Duplex. Irrespective of the axial vs equatorial conformation of the Tg^6 CH_3 group, the *cis*-5*R*,6*S* Tg lesion was positioned in a wobble orientation such that Tg^6 O^2 was proximate to G^{19} N1H and Tg^6 N3H was proximate to G^{19} O^6 (Figures 9 and 10). The *cis*-5*R*,6*S* Tg lesion shifted toward the major groove. The complementary dG remained in

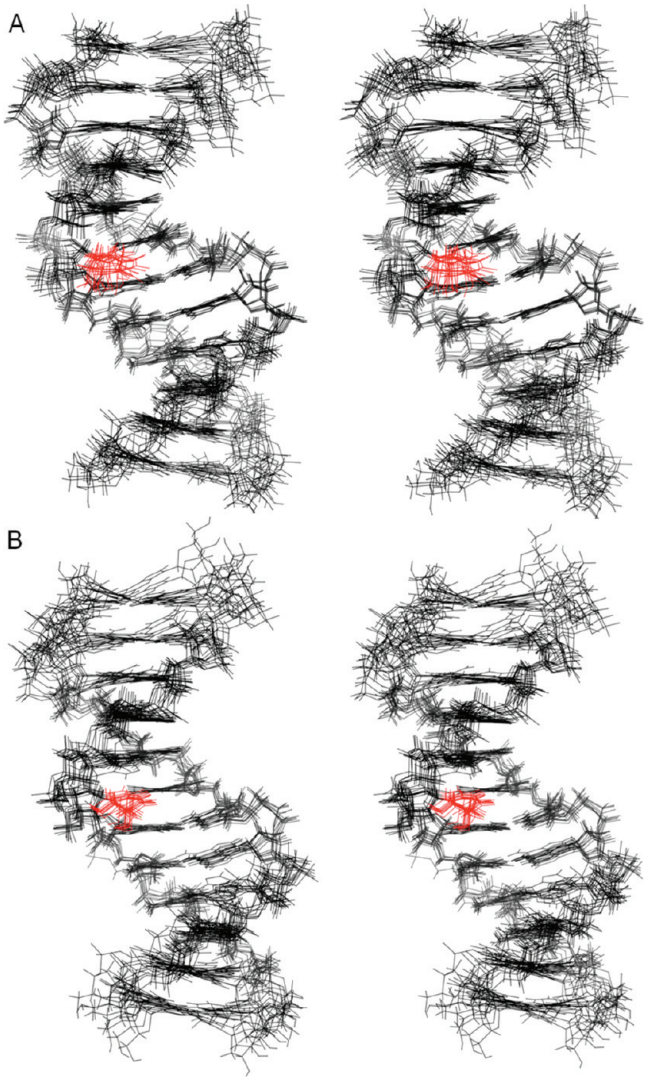


FIGURE 7: Stereoviews of 10 superimposed structures emergent from rMD calculations of the *cis*-5*R*,6*S*-Tg-adducted DNA duplex. (A) An ensemble of structures in which Tg^6 CH_3 is in the axial conformation (PDB ID 2KH7). (B) An ensemble of structures in which Tg^6 CH_3 is in the equatorial conformation (PDB ID 2KH8).

the *anti* conformation about the glycosyl bond and stacked into the duplex. The *cis*-5*R*,6*S* Tg base stacked below the 5′-neighbor base G^5, while its complement G^{19} stacked below the 5′-neighbor base C^{20}. In the 3′-direction, base stacking between the *cis*-5*R*,6*S* Tg^6 and the $G^7 \cdot C^{18}$ base pair was disrupted (Figure 10). The base pair roll, stagger, and propeller twist helicoidal parameters were disturbed at the mismatched Tg·G pair. There was also disruption in the base−base stretch, shear, and opening helicoidal parameters. The helicoidal analysis data are provided in Table 2 of the Supporting Information.

Solvent-Accessible Surface Area of the Tg Base. The extent to which the wobble position of *cis*-5*R*,6*S* Tg^6 in the $Tg^6 \cdot G^{19}$ mismatch pair altered the solvent-accessible surface of the Tg^6 lesion was evaluated for both the equatorial and axial conformations of the Tg^6 CH_3 group (*46*) (Figure 11). Irrespective of the axial vs equatorial conformation of the Tg^6 CH_3 group, when the solvent-accessible surface of the Tg^6 nucleotide relative to that of a Tg nucleoside was compared with nucleotides A^{23}, T^9, and T^{12}, the Tg^6 lesion exhibited an increased solvent-accessible surface relative to T^9 and A^{23} but was not as exposed to solvent as the terminal T^{12} nucleotide. The average accessible surface area of the Tg^6 base was 28% when the Tg^6 CH_3 group was in the axial conformation and 21% when the Tg^6 CH_3 group was in the CH_3 equatorial conformation.

DISCUSSION

Normally in duplex DNA thymine is paired opposite adenine. Consequently, oxidative damage to thymine results in the formation of Tg lesions opposite adenines in genomic DNA. However, when 5-methylcytosine is oxidatively damaged, forming 5-methylcytosine glycol, facile hydrolytic deamination of the latter yields Tg mismatched with G in duplex DNA (*3*). Since >70% of cytosines in the 5′-CpG-3′ sequence context are believed to be methylated in mammalian cells (*3*), the formation of Tg arising from oxidative damage to 5-methylcytosine is anticipated to be biologically significant. If not recognized and repaired by the cell, the Tg·G mismatch is anticipated to result in a $C^{5\text{-Me}} \rightarrow T$ mutation. Consequently, the structure of the Tg·G mismatch is of considerable interest, particularly with regard to its recognition by base excision repair enzymes (*47*). The base excision repair of Tg in mammals is mediated by at least two DNA *N*-glycosylase/AP lyases that are influenced by the nature of the diastereoisomer of dTg, the rate of *cis*−*trans* epimerization of each diastereoisomer, and the identity of the complementary purine base (*14*).

Structure of the $Tg^6 \cdot G^{19}$ Mismatch Pair. The *cis*-5*R*,6*S* Tg base stacks into the duplex and assumes the wobble position opposite G^{19} (Figure 10), similar to the unmodified T·G mismatch pair (*48*, *49*). Tg^6 shows poor stacking interactions with 3′-neighbor base pair $G^7 \cdot C^{20}$. The structural perturbation is localized to base pairs $G^5 \cdot C^{20}$, $Tg^6 \cdot G^{19}$, and $G^7 \cdot C^{18}$. The observation that there is no disruption of sequential NOE connectivity for either the complementary strand or the modified strand (Figure 1) suggests that wobble pair orientation involving the nonplanar Tg^6 base is accommodated with minimal distortion of the helical backbone at the lesion site (*10*). The wobble orientation is reflected in perturbations in base pair shear, opening, stagger, and stretch (Figures 9 and 10; Table S1 in the Supporting Information). Despite being oriented in the wobble position, neither the Tg^6 N3H imino or the G^{19} N1H imino resonances are observed in the ^{1}H NMR spectrum (Figure 2) (*10*), which is attributed to their rapid exchange with solvent. Furthermore, thermodynamic measurements indicated that the $Tg^6 \cdot G^{19}$ base pair exhibited a 13 °C decrease in thermal melting temperature and exhibited a 3 kcal/mol $\Delta\Delta G$ value. These effects were attributed to enthalpy-driven reductions in duplex stability (*10*). The unmodified T·G mispair also exhibits reduced stability (*50*).

In duplex DNA, the axial conformation of the Tg^6 CH_3 group in the *cis*-5*R*,6*S* Tg lesion orients it in the 5′ direction in the major groove, facing toward base pair $G^5 \cdot C^{20}$, whereas in the equatorial conformation, steric interactions in the 5′-direction between the Tg^6 CH_3 and G^5 are reduced (Figure 9). Quantum mechanical calculations performed on the *cis*-5*R*,6*S* Tg base predict that the axial conformation is favored (*10*), and this has been observed in the crystal structure of the nucleoside (*51*). We proposed that in the $Tg^6 \cdot G^{19}$ mismatch steric interaction between the axial conformation of the Tg^6 CH_3 and the 5′-neighboring G^5 would be reduced by wobble pairing (*10*). The present data show that in this mismatch pair the Tg^6 nucleotide shifts to the wobble orientation opposite G^{19} (Figure 10), but the determination of axial vs equatorial conformations of the Tg^6 CH_3 group remains

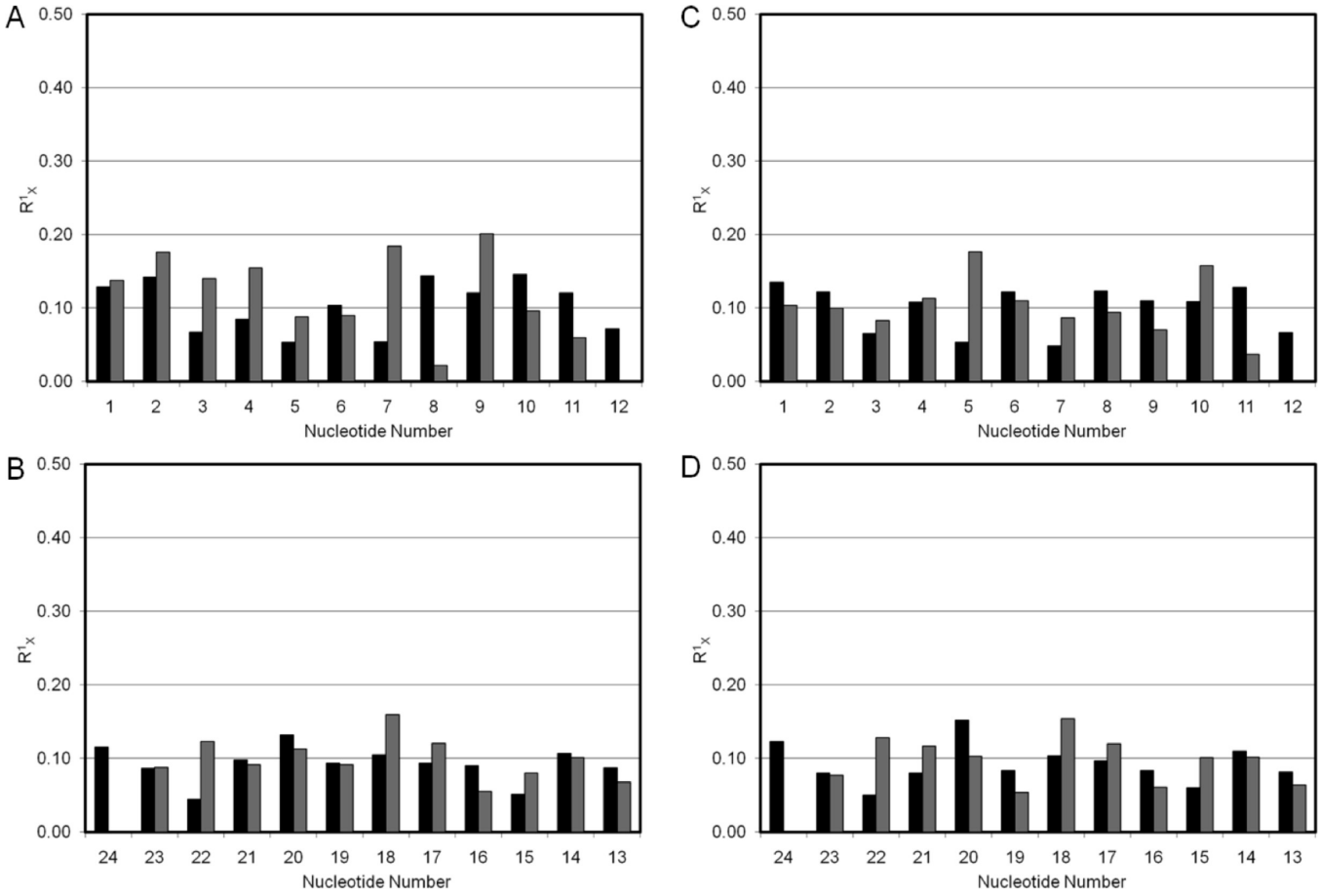


FIGURE 8: Relaxation matrix analysis of average structures emergent from rMD calculations of the *cis*-5*R*,6*S*-Tg-adducted DNA duplex, calculated using the program CORMA, showing intranucleotide (■) and internucleotide (■) values of R_1^X. (A) The modified strand with Tg^6 CH_3 in the axial conformation showing (B) the complementary strand with Tg^6 CH_3 in the axial conformation, (C) the modified strand with Tg^6 CH_3 in the equatorial conformation, and (D) the complementary strand with Tg^6 CH_3 in the equatorial conformation.

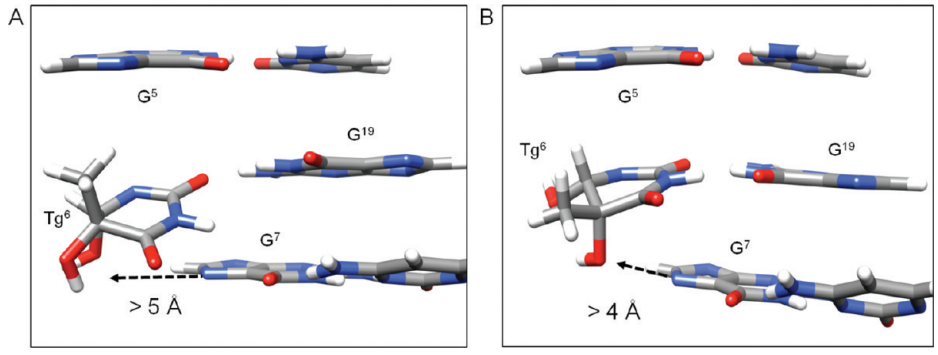


FIGURE 9: The wobble orientation of Tg^6 in the $Tg^6 \cdot G^{19}$ mismatch pair does not favor the formation of stabilizing intrastrand hydrogen bonds between the hydroxyl groups of the *cis*-5*R*,6*S* Tg^6 lesion and the imidazole ring N7 atom of the 5′-neighbor purine, as determined from analyses of rMD trajectories in explicit solvent. (A) When Tg^6 CH_3 is in the axial conformation, the distance between the C6 OH group and G7 N7 is > 5 Å. (B) When Tg^6 CH_3 is in the equatorial conformation, the distance between the C6 OH group and G7 N7 is > 4 Å.

equivocal: both conformations are observed in the rMD calculations and both show agreement with the NOE data (Figure 8). However, the axial conformation of the Tg^6 CH_3 group places it proximate to the major groove G^5 H8 proton of the $G^5 \cdot C^{20}$ base pair (Figures 9 and 10). Significantly, in the crystal structure of the RB69 polymerase involving a template containing the *cis*-5*R*,6*S* Tg lesion and an incoming dATP nucleotide, the Tg CH_3 group remained in the axial conformation, consistent with the nucleoside data (*51*), despite hindering stacking of the adjacent 5′-template guanine (*52*). In this $Tg^6 \cdot G^{19}$ mismatch the *cis*-5*R*,6*S* Tg lesion does perturb the 5′-neighbor base pair $G^5 \cdot C^{20}$. The imino resonance attributed to base pair $G^5 \cdot C^{20}$ broadens due to solvent exchange (Figure 2) and disappears from the spectrum ~35 °C lower as compared to the corresponding unmodified duplex (*10*). The observation that the rMD trajectories yield ensembles of structures in which the Tg^6 CH_3 was in either the axial or equatorial conformations (Figure 7) suggests that both conformers of the Tg^6 ring may be populated in the Tg^6G^{19} mismatch and that while shifting Tg^6 to the wobble orientation probably reduces steric interference between the axial conformation of the Tg^6 CH_3 group with the 5′-neighbor base pair $G^5 \cdot C^{20}$, the conformational dynamics of the Tg^6 lesion remain disordered

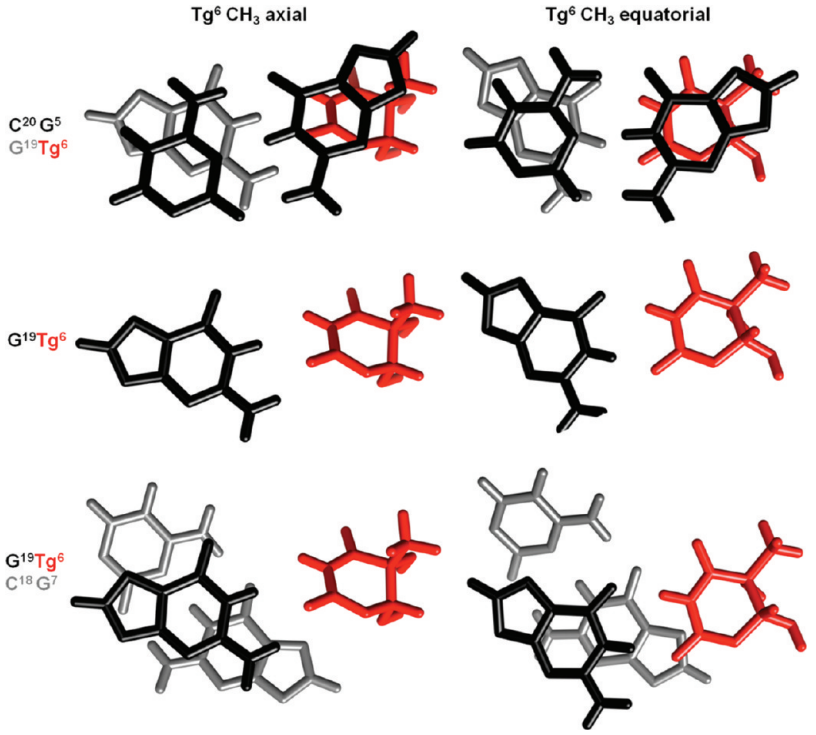


FIGURE 10: Base pair stacking interactions determined from structural refinements using rMD calculations restrained by NMR-derived distance and torsion angles for the *cis*-5*R*,6*S*-Tg-adducted duplex containing the $Tg^6 \cdot G^{19}$ mismatch pair. The left panel shows stacking interactions when Tg^6 CH_3 is in the axial conformation (PDB ID 2KH7). The right panel shows stacking interactions when Tg^6 CH_3 is in the equatorial conformation (PDB ID 2KH8). In both instances, the Tg^6 base shifts toward the major groove, into a wobble orientation with G^{19} in the complementary strand, but both Tg^6 and G^{19} remain stacked into the duplex.

at this mismatch site. The modest occupancies of the Tg^6 OH6 to G^7 N7 hydrogen bond, associated with the axial conformation of the Tg^6 CH_3 group, and the Tg^6 OH5 to G^7 N7 hydrogen bond, associated with the equatorial conformation of the Tg^6 CH_3 group, suggest that intrastrand hydrogen bond formation between the Tg C6 hydroxyl and the N7 position of the 3′ purine (*13*) does not particularly stabilize either the axial or equatorial conformations of the Tg^6 base in this mismatched pair (Figure 9). Calculations of solvent-accessible surfaces (Figure 11) indicate that Tg is more exposed to solvent in the $Tg^6 \cdot G^{19}$ mismatch pair, as compared to a Watson−Crick T·A base pair. This is in agreement with the findings of Kung and Bolton (*11*) on the basis of their solvent-accessible surface area calculations in the 5′-ATgA-3′ sequence when the 5*R* Tg lesion was placed opposite dA. It is consistent with the wobble orientation of Tg^6, which shifts the modified base toward the major groove as compared to a canonical T·A base pair. However, it is important to note that in the wobble orientation the Tg^6 lesion does remain stacked into the duplex and it is not flipped out of the helix and into the major groove (Figure 10).

Dynamics of the $Tg^6 \cdot G^{19}$ Mismatch Pair. The formation of the wobble base pair results in increased disorder at the $Tg^6 \cdot G^{19}$ mismatch. The T_2 spin−spin relaxation data are indicative of an increase in transverse relaxation rate for Tg^6 CH_3 as compared to unmodified thymine CH_3 protons (Figure 4). This is attributed to puckering of the Tg^6 six-membered ring between axial and equatorial conformations of the Tg^6 CH^3 group. However, the possibility of increased backbone and deoxyribose dynamics occurring in tandem with the puckering of the Tg ring cannot be excluded (*50*, *53*, *54*). The shift of the Tg^6 base further toward the major groove, accompanied by disruptions in hydrogen bonding at the modified base pair and its 5′-neighbor $G^5 \cdot C^{18}$ and accompanied by poor stacking interactions with 3′-neighbor base pair $G^7 \cdot C^{20}$ (Figure 10), contributes to the reduced thermal stability of the mismatch (*10*) and probably contributes to increased disorder at the $Tg^6 \cdot G^{19}$ mismatch.

Structure−Activity Relationships. If not repaired, the 5*R* Tg lesion is generally lethal to cells (*55*−*59*), although several DNA polymerases lacking 3′→5′ exonuclease activity do bypass the 5*R* Tg lesion with limited efficiencies (*13*, *60*, *61*). The bypass of Tg by Y-family DNA polymerases is stereospecific, with pol η bypassing the 5*R* lesion more efficiently (*62*) and pol κ bypassing the 5*S* lesion more efficiently (*63*). While the Tg lesion is a substrate for the UvrABC NER repair system of *Escherichia coli* (*64*) and is excised from DNA *in vitro* by human NER enzymes (*65*), it is primarily recognized as a substrate for base excision repair, both in *E. coli* and in mammalian cells (*49*). In *E. coli*, BER of this lesion is initiated by endonuclease III (Nth) (*66*) and endonuclease VIII (Nei) (*67*). Yeast (*68*), mammalian (*69*, *70*), and human orthologues (*71*−*73*) of Nth have been characterized. Likewise, human orthologues of Nei have been characterized (*74*, *75*). The human hNTH1 exhibits a 13:1 preference for excising the 5*R* epimer, whereas hNEIL1 (*74*, *76*) shows a 1.5:1 preference for excising the 5*R* epimer (*77*). Similar observations have been made for prokaryotic, yeast, and murine glycosylases (*78*).

Ocampa-Hafalla et al. (*14*) reported that, in contrast to hNth1, base excision repair of 5*R* Tg by hNEIL1 depends upon the identity of the opposing base, with Tg·G pairs being excised much more rapidly than Tg·A pairs. This is significant because when 5-methylcytosine is oxidatively damaged, forming 5-methylcytosine glycol, hydrolytic deamination yields Tg mismatched with dG (*3*). Modeling and dynamics studies suggest that good substrates for

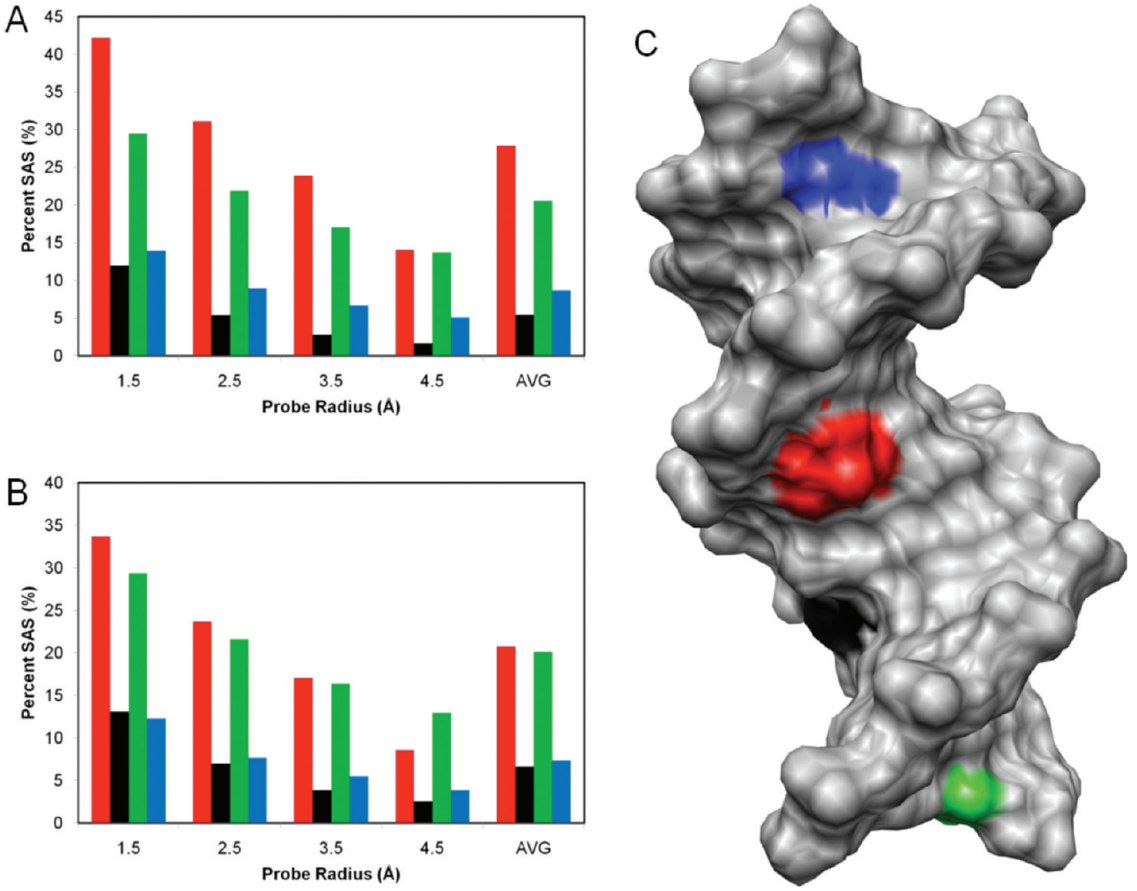


FIGURE 11: Measurement of solvent-accessible surface (SAS) areas of the Tg^6 nucleotide in the $Tg^6 \cdot G^{19}$ mismatch base pair (red) at a series of probe radii using the program MSMS. The SAS was compared to the SAS of the stacked nucleotides T^9 (black), T^{12} (green), and A^{23} (blue). (A) Results for the structure with Tg^6 CH_3 in an axial orientation. (B) Results for the structure with Tg^6 CH_3 in an equatorial orientation. (C) The solvent-excluded (SES) surface rendering of a full duplex indicates the locations of T^9 (black), T^{12} (green), and A^{23} (blue).

hNEIL1 possess in common a pyrimidine-like ring and hydrogen bond donor−acceptor properties, allowing them to be accommodated within the enzyme's binding pocket (*79*). The present structural data suggest two reasons why hNEIL1 preferentially excises 5*R* Tg lesions from the Tg·G pair, both based upon studies with the uracil DNA glycosylase enzyme, which conclude that it detects uracil as it emerges from the helix due to intrinsic thermal breathing motions in DNA, as opposed to active participation by the enzyme (*80*). In this regard, the wobble orientation of the *cis*-5*R*,6*S* Tg^6 base in the Tg·G pair shifts it toward the major groove, which is reflected in an increased solvent-accessible surface (Figure 11). While the 5*R* Tg lesion remains stacked into the helix and is not flipped out into the major groove, the wobble orientation is accompanied by poorer stacking with the 3′-neighbor base pair $G^7 \cdot C^{20}$ (Figure 10). Thus, it seems likely that the wobble orientation reduces the thermal barrier for intrinsic breathing of the lesion, thus facilitating the "flipping out" mechanism associated with the BER process. In support of this notion, the present data are indicative of increased structural disorder at the lesion site, reflected in the T_2 spin−spin relaxation data (Figure 4) accompanied by disruptions in hydrogen bonding at the modified base pair and its 5′-neighbor $G^5 \cdot C^{18}$. Thus, it seems reasonable that the rate of intrinsic thermal breathing is increased for the *cis*-5*R*,6*S* Tg^6 when paired opposite guanine, as has been reported for the unmodified G·T mismatch pair (*54*). It will now be of interest to compare the structure of the *cis*-5*R*,6*S* Tg^6 lesion when paired opposite guanine to that of the same lesion paired opposite adenine.

Summary. Oxidative damage to 5-methylcytosine in DNA, followed by deamination, yields Tg mispaired with guanine. In this duplex containing the *cis*-5*R*,6*S* Tg lesion mismatched with G, the mismatched pair exists in the wobble orientation such that Tg^6 O^2 is proximate to G^{19} N1H and Tg^6 N3H is proximate to G^{19} O^6. Both Tg^6 and the mismatched nucleotide G^{19} remain stacked in the helix. The Tg^6 nucleotide shifts toward the major groove and stacks below the 5′-neighbor base G^5, while its complement G^{19} stacks below the 5′-neighbor C^{20}. In the 3′-direction, stacking between Tg^6 and the $G^7 \cdot C^{18}$ base pair is disrupted. The solvent-accessible surface area of the Tg nucleotide increases as compared to the native T·A base pair. An increase in T_2 relaxation rates for the Tg^6 base protons is attributed to puckering of the Tg base, accompanied by increased disorder at the $Tg^6 \cdot G^{19}$ mismatch pair. The wobble pairing and disorder of the Tg·G mismatch correlate with the reduced thermodynamic stability of the mismatch and likely modulate its recognition by DNA base excision repair systems.

ACKNOWLEDGMENT

Dr. Markus Voehler and Dr. Don Stec assisted with NMR spectroscopy. We thank Dr. Nicholas Uylanov and Dr. Jarrod A. Smith for guidance on structural refinement. We also thank Drs. Thomas M. and Constance M. Harris for constructive communications.

SUPPORTING INFORMATION AVAILABLE

Table S1, helicoidal analysis of the oligodeoxynucleotide duplex containing the $Tg^{6}\cdot G^{19}$ mismatch. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Teoule, R., Bonicel, A., Bert, C., Cadet, J., and Polverelli, M. (1974) Identification of radioproducts resulting from the breakage of thymine moiety by gamma irradiation of *E. coli* DNA in an aerated aqueous solution. *Radiat. Res. 57*, 46–58.
2. Frenkel, K., Goldstein, M. S., and Teebor, G. W. (1981) Identification of the cis-thymine glycol moiety in chemically oxidized and gamma-irradiated deoxyribonucleic acid by high-pressure liquid chromatography analysis. *Biochemistry 20*, 7566–7571.
3. Zuo, S., Boorstein, R. J., and Teebor, G. W. (1995) Oxidative damage to 5-methylcytosine in DNA. *Nucleic Acids Res. 23*, 3239–3243.
4. Pfeifer, G. P. (2000) p53 mutational spectra and the role of methylated CpG sequences. *Mutat. Res. 450*, 155–166.
5. Vaishnav, Y., Holwitt, E., Swenberg, C., Lee, H. C., and Kan, L. S. (1991) Synthesis and characterization of stereoisomers of 5,6-dihydro-5,6-dihydroxy-thymidine. *J. Biomol. Struct. Dyn. 8*, 935–951.
6. Lustig, M. J., Cadet, J., Boorstein, R. J., and Teebor, G. W. (1992) Synthesis of the diastereomers of thymidine glycol, determination of concentrations and rates of interconversion of their cis-trans epimers at equilibrium and demonstration of differential alkali lability within DNA. *Nucleic Acids Res. 20*, 4839–4845.
7. Wang, Y. (2002) HPLC isolation and mass spectrometric characterization of two isomers of thymine glycols in oligodeoxynucleotides. *Chem. Res. Toxicol. 15*, 671–676.
8. Cathcart, R., Schwiers, E., Saul, R. L., and Ames, B. N. (1984) Thymine glycol and thymidine glycol in human and rat urine: A possible assay for oxidative DNA damage. *Proc. Natl. Acad. Sci. U.S.A. 81*, 5633–5637.
9. Adelman, R., Saul, R. L., and Ames, B. N. (1988) Oxidative damage to DNA: Relation to species metabolic rate and life span. *Proc. Natl. Acad. Sci. U.S.A. 85*, 2706–2708.
10. Brown, K. L., Adams, T., Jasti, V. P., Basu, A. K., and Stone, M. P. (2008) Interconversion of the *cis*-5*R*,6*S*- and *trans*-5*R*,6*R*-thymine glycol lesions in duplex DNA. *J. Am. Chem. Soc. 129*, 11701–11710.
11. Kung, H. C., and Bolton, P. H. (1997) Structure of a duplex DNA containing a thymine glycol residue in solution. *J. Biol. Chem. 272*, 9227–9236.
12. Kao, J. Y., Goljer, I., Phan, T. A., and Bolton, P. H. (1993) Characterization of the effects of a thymine glycol residue on the structure, dynamics, and stability of duplex DNA by NMR. *J. Biol. Chem. 268*, 17787–17793.
13. Clark, J. M., Pattabiraman, N., Jarvis, W., and Beardsley, G. P. (1987) Modeling and molecular mechanical studies of the cis-thymine glycol radiation damage lesion in DNA. *Biochemistry 26*, 5404–5409.
14. Ocampo-Hafalla, M. T., Altamirano, A., Basu, A. K., Chan, M. K., Ocampo, J. E., Cummings, A., Jr., Boorstein, R. J., Cunningham, R. P., and Teebor, G. W. (2006) Repair of thymine glycol by hNth1 and hNeil1 is modulated by base pairing and *cis-trans* epimerization. *DNA Repair* (*Amsterdam*) *5*, 444–454.
15. Piotto, M., Saudek, V., and Sklenar, V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR 2*, 661–665.
16. Carr, H. Y., and Purcell, E. M. (1954) Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys. Rev. 94*, 630.
17. Vold, R. L., Waugh, J. S., Klein, M. P., and Phelps, D. E. (1968) Measurement of spin relaxation in complex systems. *J. Chem. Phys. 48*, 3831–3832.
18. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR 6*, 277–293.
19. Goddard, T. D., and Kneller, D. G. (2006) SPARKY v.3.113, University of California, San Francisco.
20. Keepers, J. W., and James, T. L. (1984) A theoretical study of distance determinations from NMR—Two-dimensional nuclear Overhauser effect spectra. *J. Magn. Reson. 57*, 404–426.
21. James, T. L. (1991) Relaxation matrix analysis of two-dimensional nuclear Overhauser effect spectra. *Curr. Opin. Struct. Biol. 1*, 1042–1053.
22. Borgias, B. A., and James, T. L. (1989) Two-dimensional nuclear Overhauser effect: Complete relaxation matrix analysis. *Methods Enzymol. 176*, 169–183.
23. Borgias, B. A., and James, T. L. (1990) MARDIGRAS—a procedure for matrix analysis of relaxation for discerning geometry of an aqueous structure. *J. Magn. Reson. 87*, 475–487.
24. Liu, H., Spielmann, H. P., Ulyanov, N. B., Wemmer, D. E., and James, T. L. (1995) Interproton distance bounds from 2D NOE intensities: Effect of experimental noise and peak integration errors. *J. Biomol. NMR 6*, 390–402.
25. Delaglio, F., Wu, Z., and Bax, A. (2001) Measurement of homonuclear proton couplings from regular 2D COSY spectra. *J. Magn. Reson. 149*, 276–281.
26. Van De Ven, F. J. M., and Hilbers, C. W. (1988) Nucleic acids and nuclear magnetic resonance. *Eur. J. Biochem. 178*, 1–38.
27. van Wijk, J., Huckriede, B. D., Ippel, J. H., and Altona, C. (1992) Furanose sugar conformations in DNA from NMR coupling constants. *Methods Enzymol. 211*, 286–306.
28. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Montgomery, J. A., Vreven, T., Kudin, K. N., Burant, J. C., Millam, J. M., Iyengar, S. S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G. A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J. E., Hratchian, H. P., Cross, J. B., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Pomelli, J., Ochterski, W., Ayala, P. Y., Morokuma, K., Voth, G. A., Salvador, P., Dannenberg, J. J., Zakrzewska, V. G., Daniels, A. D., Farkas, O., Rabuck, A. D., Raghavachari, K., and Ortiz, J. V. (2004) GAUSSIAN 03, Gaussian, Inc., Wallingford, CT.
29. Case, D. A., Cheatham, T. E., III, Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The AMBER biomolecular simulation programs. *J. Comput. Chem. 26*, 1668–1688.
30. Macke, T., and Case, D. A. (1998) Modeling unusual nucleic acid structures, in Molecular Modeling of Nucleic Acids (Leontes, N. B., and SantaLucia, J., Jr., Eds.) American Chemical Society, Washington, DC.
31. Wang, J. M., Cieplak, P., and Kollman, P. A. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem. 21*, 1049–1074.
32. Clore, G. M., Brunger, A. T., Karplus, M., and Gronenborn, A. M. (1986) Application of molecular dynamics with interproton distance restraints to three-dimensional protein structure determination. *J. Mol. Biol. 191*, 523–551.
33. Tsui, V., and Case, D. A. (2000) Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers 56*, 275–291.
34. Bashford, D., and Case, D. A. (2000) Generalized Born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem. 51*, 129–152.
35. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J. Phys. Chem. 81*, 3684–3690.
36. Loncharich, R. J., Brooks, B. R., and Pastor, R. W. (1992) Langevin dynamics of peptides: The frictional dependence of isomerization rates of N-acetylalanyl-N′-methylamide. *Biopolymers 32*, 523–535.
37. Izaguirre, J. A., Catarello, D. P., Wozniak, J. M., and Skeel, R. D. (2001) Langevin stabilization of molecular dynamics. *J. Chem. Phys. 114*, 2090–2098.
38. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys. 103*, 8577–8593.
39. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of cartesian equations of motion of a system with constraints: Molecular dynamics of N-alkanes. *J. Comput. Phys. 23*, 327–341.
40. Lavery, R., and Sklenar, H. (1996) CURVES 5.1. Helical analysis of irregular nucleic acids, Laboratoire de Biochimie Theorique CNRS, Paris, France.
41. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF chimera: A visualization system for exploratory research and analysis. *J. Comput. Chem. 25*, 1605–1612.
42. Sanner, M. F., Olson, A. J., and Spehner, J. C. (1996) Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers 38*, 305–320.
43. Reid, B. R. (1987) Sequence-specific assignments and their use in NMR studies of DNA structure. *Q. Rev. Biophys. 20*, 2–28.

44. Patel, D. J., Shapiro, L., and Hare, D. (1987) DNA and RNA: NMR studies of conformations and dynamics in solution. *Q. Rev. Biophys. 20*, 35–112.
45. Boelens, R., Scheek, R. M., Dijkstra, K., and Kaptein, R. (1985) Sequential assignment of imino- and amino-proton resonances in ^{1}H NMR spectra of oligonucleotides by two-dimensional NMR spectroscopy. Application to a *lac* operator fragment. *J. Magn. Reson. 62*, 378–386.
46. Connolly, A. (1983) Analytical molecular surface calculation. *J. Appl. Crystallogr. 16*, 548–558.
47. Hazra, T. K., Das, A., Das, S., Choudhury, S., Kow, Y. W., and Roy, R. (2007) Oxidative DNA damage repair in mammalian cells: A new perspective. *DNA Repair* (*Amsterdam*) *6*, 470–480.
48. Patel, D. J., Kozlowski, S. A., Marky, L. A., Rice, J. A., Broka, C., Dallas, J., Itakura, K., and Breslauer, K. J. (1982) Structure, dynamics, and energetics of deoxyguanosine·thymidine wobble base pair formation in the self-complementary d(CGTGAATTCGCG) duplex in solution. *Biochemistry 21*, 437–444.
49. Hare, D., Shapiro, L., and Patel, D. J. (1986) Wobble dG·dT pairing in right-handed DNA: Solution conformation of the d(C-G-T-G-A-A-T-T-C-G-C-G) duplex deduced from distance geometry analysis of nuclear Overhauser effect spectra. *Biochemistry 25*, 7445–7456.
50. Allawi, H. T., and SantaLucia, J., Jr. (1997) Thermodynamics and NMR of internal G·T mismatches in DNA. *Biochemistry 36*, 10581–10594.
51. Hruska, F. E., Sebastian, R., Grand, A., Voituriez, L., and Cadet, J. (1987) Characterization of a γ-radiation-induced decomposition product of thymidine. Crystal and molecular structure of the (−)*cis*(5*R*,6*S*) thymidine glycol. *Can. J. Chem. 65*, 2618–2623.
52. Aller, P., Rould, M. A., Hogg, M., Wallace, S. S., and Doublie, S. (2007) A structural rationale for stalling of a replicative DNA polymerase at the most common oxidative thymine lesion, thymine glycol. *Proc. Natl. Acad. Sci. U.S.A. 104*, 814–818.
53. Roongta, V. A., Jones, C. R., and Gorensteia, D. G. (1990) Effect of distortions in the deoxyribosephosphate backbone conformation of duplex oligodeoxyribonucletide dodecamers containing GT, GG, GA, AC, and GU base-pair mismatches on ^{31}P NMR spectra. *Biochemistry 29*, 5245–5258.
54. Moe, J. G., and Russu, I. M. (1992) Kinetics and energetics of base-pair opening in 5′-d(CGCGAATTCGCG)-3′ and a substituted dodecamer containing G·T mismatches. *Biochemistry 31*, 8421–8428.
55. Achey, P. M., and Wright, C. F. (1983) Inducible repair of thymine ring saturation damage in phi X174 DNA. *Radiat. Res. 93*, 609–612.
56. Moran, E., and Wallace, S. S. (1985) The role of specific DNA base damages in the X-ray-induced inactivation of bacteriophage PM2. *Mutat. Res. 146*, 229–241.
57. Laspia, M. F., and Wallace, S. S. (1988) Excision repair of thymine glycols, urea residues, and apurinic sites in *Escherichia coli*. *J. Bacteriol. 170*, 3359–3366.
58. Hayes, R. C., Petrullo, L. A., Huang, H. M., Wallace, S. S., and LeClerc, J. E. (1988) Oxidative damage in DNA. Lack of mutagenicity by thymine glycol lesions. *J. Mol. Biol. 201*, 239–246.
59. Kow, Y. W., Faundez, G., Melamede, R. J., and Wallace, S. S. (1991) Processing of model single-strand breaks in phi X-174 RF transfecting DNA by *Escherichia coli*. *Radiat. Res. 126*, 357–366.
60. Hayes, R. C., and LeClerc, J. E. (1986) Sequence dependence for bypass of thymine glycols in DNA by DNA polymerase I. *Nucleic Acids Res. 14*, 1045–1061.
61. McNulty, J. M., Jerkovic, B., Bolton, P. H., and Basu, A. K. (1998) Replication inhibition and miscoding properties of DNA templates containing a site-specific cis-thymine glycol or urea residue. *Chem. Res. Toxicol. 11*, 666–673.
62. Kusumoto, R., Masutani, C., Iwai, S., and Hanaoka, F. (2002) Translesion synthesis by human DNA polymerase eta across thymine glycol lesions. *Biochemistry 41*, 6090–6099.
63. Fischhaber, P. L., Gerlach, V. L., Feaver, W. J., Hatahet, Z., Wallace, S. S., and Friedberg, E. C. (2002) Human DNA polymerase kappa bypasses and extends beyond thymine glycols during translesion synthesis in vitro, preferentially incorporating correct nucleotides. *J. Biol. Chem. 277*, 37604–37611.
64. Lin, J. J., and Sancar, A. (1989) A new mechanism for repairing oxidative damage to DNA: (A)BC excinuclease removes AP sites and thymine glycols from DNA. *Biochemistry 28*, 7979–7984.
65. Reardon, J. T., Bessho, T., Kung, H. C., Bolton, P. H., and Sancar, A. (1997) In vitro repair of oxidative DNA damage by human nucleotide excision repair system: Possible explanation for neurodegeneration in xeroderma pigmentosum patients. *Proc. Natl. Acad. Sci. U.S.A. 94*, 9463–9468.
66. Weiss, B., and Cunningham, R. P. (1985) Genetic mapping of Nth, a gene affecting endonuclease III (thymine glycol-DNA glycosylase) in *Escherichia coli* K-12. *J. Bacteriol. 162*, 607–610.
67. Jiang, D., Hatahet, Z., Melamede, R. J., Kow, Y. W., and Wallace, S. S. (1997) Characterization of *Escherichia coli* endonuclease VIII. *J. Biol. Chem. 272*, 32230–32239.
68. Roldan-Arjona, T., Anselmino, C., and Lindahl, T. (1996) Molecular cloning and functional analysis of a *Schizosaccharomyces pombe* homologue of *Escherichia coli* endonuclease III. *Nucleic Acids Res. 24*, 3307–3312.
69. Hilbert, T. P., Boorstein, R. J., Kung, H. C., Bolton, P. H., Xing, D., Cunningham, R. P., and Teebor, G. W. (1996) Purification of a mammalian homologue of *Escherichia coli* endonuclease III: identification of a bovine pyrimidine hydrate-thymine glycol DNAse/AP lyase by irreversible cross linking to a thymine glycol-containing oligoxynucleotide. *Biochemistry 35*, 2505–2511.
70. Sarker, A. H., Ikeda, S., Nakano, H., Terato, H., Ide, H., Imai, K., Akiyama, K., Tsutsui, K., Bo, Z., Kubo, K., Yamamoto, K., Yasui, A., Yoshida, M. C., and Seki, S. (1998) Cloning and characterization of a mouse homologue (mNthl1) of *Escherichia coli* endonuclease III. *J. Mol. Biol. 282*, 761–774.
71. Hilbert, T. P., Chaung, W., Boorstein, R. J., Cunningham, R. P., and Teebor, G. W. (1997) Cloning and expression of the cDNA encoding the human homologue of the DNA repair enzyme, *Escherichia coli* endonuclease III. *J. Biol. Chem. 272*, 6733–6740.
72. Aspinwall, R., Rothwell, D. G., Roldan-Arjona, T., Anselmino, C., Ward, C. J., Cheadle, J. P., Sampson, J. R., Lindahl, T., Harris, P. C., and Hickson, I. D. (1997) Cloning and characterization of a functional human homolog of *Escherichia coli* endonuclease III. *Proc. Natl. Acad. Sci. U.S.A. 94*, 109–114.
73. Ikeda, S., Biswas, T., Roy, R., Izumi, T., Boldogh, I., Kurosky, A., Sarker, A. H., Seki, S., and Mitra, S. (1998) Purification and characterization of human NTH1, a homolog of *Escherichia coli* endonuclease III. Direct identification of Lys-212 as the active nucleophilic residue. *J. Biol. Chem. 273*, 21585–21593.
74. Hazra, T. K., Izumi, T., Boldogh, I., Imhoff, B., Kow, Y. W., Jaruga, P., Dizdaroglu, M., and Mitra, S. (2002) Identification and characterization of a human DNA glycosylase for repair of modified bases in oxidatively damaged DNA. *Proc. Natl. Acad. Sci. U.S.A. 99*, 3523–3528.
75. Hazra, T. K., Kow, Y. W., Hatahet, Z., Imhoff, B., Boldogh, I., Mokkapati, S. K., Mitra, S., and Izumi, T. (2002) Identification and characterization of a novel human DNA glycosylase for repair of cytosine-derived lesions. *J. Biol. Chem. 277*, 30417–30420.
76. Bandaru, V., Sunkara, S., Wallace, S. S., and Bond, J. P. (2002) A novel human DNA glycosylase that removes oxidative DNA damage and is homologous to *Escherichia coli* endonuclease VIII. *DNA Repair* (*Amsterdam*) *1*, 517–529.
77. Katafuchi, A., Nakano, T., Masaoka, A., Terato, H., Iwai, S., Hanaoka, F., and Ide, H. (2004) Differential specificity of human and *Escherichia coli* endonuclease III and VIII homologues for oxidative base lesions. *J. Biol. Chem. 279*, 14464–14471.
78. Miller, H., Fernandes, A. S., Zaika, E., McTigue, M. M., Torres, M. C., Wente, M., Iden, C. R., and Grollman, A. P. (2004) Stereoselective excision of thymine glycol from oxidatively damaged DNA. *Nucleic Acids Res. 32*, 338–345.
79. Jia, L., Shafirovich, V., Geacintov, N. E., and Broyde, S. (2007) Lesion specificity in the base excision repair enzyme hNeil1: Modeling and dynamics studies. *Biochemistry 46*, 5305–5314.
80. Parker, J. B., Bianchet, M. A., Krosky, D. J., Friedman, J. I., Amzel, L. M., and Stivers, J. T. (2007) Enzymatic capture of an extrahelical thymine in the search for uracil in DNA. *Nature 449*, 433–437.