Structure of a Duplex Oligodeoxynucleotide Containing Propanodeoxyguanosine Opposite a Two-Base Deletion in the $(CpG)_3$ Frame Shift Hotspot of *Salmonella typhimurium hisD3052* Determined by ^1H NMR and Restrained Molecular Dynamics†

Jason P. Weisenseel,‡ James G. Moe,‡,§ G. Ramachandra Reddy,|| Lawrence J. Marnett,|| and Michael P. Stone*,‡

Center in Molecular Toxicology, Department of Chemistry, and A. B. Hancock, Jr., Memorial Laboratory for Cancer Research, Department of Biochemistry, School of Medicine, Vanderbilt University, Nashville, Tennessee 37235

Received July 13, 1994; Revised Manuscript Received October 16, 1994®

ABSTRACT: Structural refinement from solution ^1H NMR data was performed on the 5′-d[ATCGC(PdG)-CGGCATG]-3′•5′-d[CATGCCGCGAT]-3′ duplex, in which the adducted oligodeoxynucleotide containing the exocyclic lesion 1,N^2-propano-2′-deoxyguanosine (PdG) was annealed with the complementary strand which contained a CpG deletion. The resulting duplex required PdG and one adjacent cytosine to be unpaired. A total of 352 distances were utilized to restrain molecular dynamics calculations, of which 264 were NOE-derived. These distances were calculated using complete relaxation matrix methods from hybrid matrices, which were comprised of the experimentally determined distances and additional distances derived from either A-form or B-form DNA. A simulated annealing protocol combined with the distance restraints was able to refine a single structure with an average rms deviation of <1.35 Å. The accuracy of the refined structure was assessed using full relaxation matrix calculations, which gave good agreement with measured NOE intensities. PdG was found to be stacked into the helix below base pair C^3•G^{18}, whereas C^5 was found to be unpaired and extruded toward the major groove and parallel to base pair G^6•C^{17}. This created a localized bend in the DNA helix of approximately 20−35° at the junction between PdG and C^5. The bending corroborated previous assays performed on this modified sequence [Moe, J. G., Reddy, G. R., Marnett, L. J., & Stone, M. P. (1994) *Chem. Res. Toxicol. 7,* 319−328].

The insertion or deletion of nucleotides into the reading frame of DNA results in frame shift mutations. These sequence-dependent mutations are often associated with iterated bases, palindromes, and tandem repeats. The observed sequence dependence has led to the development of several mechanistic hypotheses to explain frame shifts. Among these are models which invoke slipped mispairing (Streisinger et al., 1966), insertion−slippage (Kunkel, 1990; Shibutani & Grollman, 1993), and the formation of quasi-palindromes (Ripley, 1982, 1990). Deletion of a cytosine by ICR-191 generated the *hisD3052* mutation, located within the histidinol dehydrogenase gene in *Salmonella typhimurium* (Oeschger & Hartman, 1970; Hartman et al., 1986; O'Hara & Marnett, 1991). Additions and deletions which restore the reading frame reverse this mutation (McCann et al., 1975). One of the most common mutations in the reiterated sequence $(CG)_4$[1] is a CG deletion (Isono & Yourno, 1974; Fuscoe et al., 1988; Bell et al., 1991; DeMarini et al., 1992).

Malondialdehyde (MDA),[2] a mutagen produced endogenously in humans during lipid peroxidation and prostaglandin biosynthesis, reverts *hisD3052* (Crawford et al., 1965; Mukai & Goldstein, 1976; Marnett & Tuttle, 1980; Basu & Marnett, 1983; Spalding, 1988). MDA exists as its tautomer β-hydroxyacrolein in aqueous solutions. Structure−activity studies (Basu & Marnett, 1984) suggest that both carbonyl equivalents react to generate a premutagenic lesion that

† This work was supported by NIH research Grants CA-55678 (M.P.S.) and CA-47479 (L.J.M.) and by NIH instrumentation Grant RR-05805 (NMR spectrometer). Additional support for laboratory core facilities, including NMR spectroscopy, was provided by the Vanderbilt Center Grant in Molecular Toxicology (ES-00267). J.P.W. was supported by a training grant in Molecular Biophysics (GM-08320).

* Author to whom correspondence should be addressed.

‡ Center in Molecular Toxicology and Department of Chemistry.

§ Present address: GENE-TRAK, 31 New York Ave., Framingham, MA 01701.

|| Center in Molecular Toxicology and A. B. Hancock, Jr., Memorial Laboratory for Cancer Research.

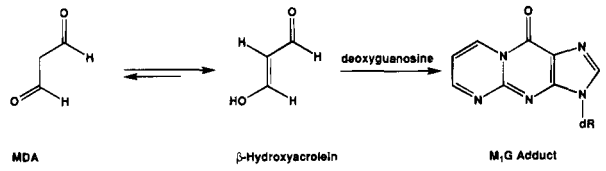
® Abstract published in *Advance ACS Abstracts,* December 1, 1994.

[1] The oligodeoxynucleotides discussed in this paper do not have terminal phosphate groups—we abbreviate the nomenclature for oligonucleotides by leaving out the phosphodiester linkage. A, C, G, T, and X refer to mononucleotide units; X is the exocyclic 1,N^2-propanodeoxyguanosine nucleotide PdG. 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 B. To keep the numbering scheme in register with previous studies on d(CGCXCGGCATG)•d(CATGCCGCGCG), the nucleotides were numbered beginning with A^{-2}, T^{-1}, C^1, etc. 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.

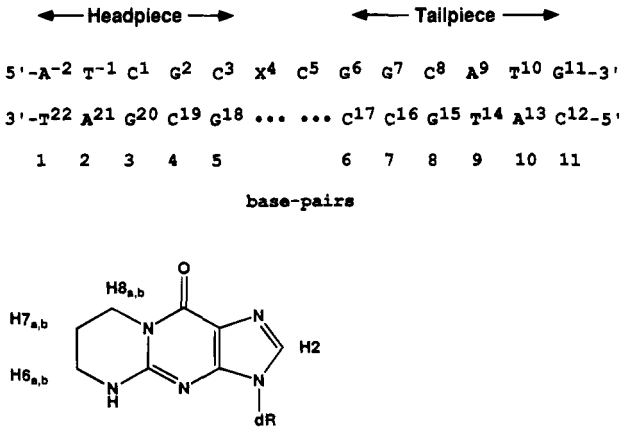
[2] Abbreviations: AAF, *N*-2-acetylaminofluorene; AF, *N*-2-aminofluorene; NOE, nuclear Overhauser enhancement; NOESY, two-dimensional NOE nuclear Overhauser enhancement experiment; ppm, parts per million; M_1G, 3-(β-D-ribofuranosyl)pyrimido[1,2a]purin-10-(3H)-one; MDA, malondialdehyde; PdG, 1,N^2-propano-2′-deoxyguanosine; DMT, dimethoxytrityl; PdG-2BD, d(ATCGC(PdG)CGGCATG)•d(CATGCCGCGAT); DSS, sodium 4,4-dimethyl-4-silapentanesulfonate; MD, molecular dynamics; MNDO, modified neglect of diatomic overlap; PEM, potential energy minimization; RMA, relaxation matrix; rMD, restrained molecular dynamics; RMSD, root mean square deviation; SCF, self-consistent field; TBE, tris borate EDTA buffer; TPPI, time-proportional phase increment; TOCSY, total homonuclear correlated spectroscopy; 1D, one dimensional; 2D, two dimensional.

0006-2960/95/0434-50$09.00/0

Scheme 1: Formation of M_1G from the Reaction between β-Hydroxyacrolein and Deoxyguanosine



Scheme 2: PdG-2BD Oligodeoxynucleotide[a]



[a] Two nucleotides were removed from the complementary strand, forcing X^4 and C^5 to be unpaired.

induces frame shifts. This is consistent with formation of the major DNA adduct, a pyrimidopurinone designated M_1G, formed from addition of 1 equiv of MDA to deoxyguanosine (Scheme 1) (Seto et al., 1983, 1986; Marnett et al., 1986; Basu et al., 1988). M_1G is not stable to the base deprotection conditions of automated DNA synthesis and has been modeled by 1,N^2-propano-2′-deoxyguanosine (PdG), a structural analog which can be readily incorporated into oligodeoxynucleotides (Marinelli et al., 1990) (Scheme 2).

Previous NMR studies established that conformational perturbation induced by PdG is dependent on the complementary base, the solution pH, and the flanking sequence. PdG(syn)•A(anti) pairing was observed at pH 5.8, simultaneous partial intercalation of the complementary PdG and A bases was observed at pH 8.9, and PdG(syn)•G(anti) pairing was observed to be pH-independent (Kouchakdjian et al., 1989b, 1990). It was also established that the exocyclic ring of PdG was inserted into the DNA duplex when positioned opposite an apurinic site (Kouchakdjian et al., 1991). Two solution structures of PdG-containing oligodeoxynucleotides refined from NMR data using molecular dynamics calculations have been reported: the pH 5.8 and pH 8.9 PdG•A adducts (Huang & Eisenberg, 1992; Huang et al., 1993). Thermodynamic measurements, obtained from UV melting assays, revealed that PdG reduced the thermal stability, transition enthalpy, and transition free energy of the duplex when positioned opposite cytosine or adenine (Plum et al., 1992).

Embedding PdG within a $(CG)_3$ repeat resulted in structural perturbation of a single dinucleotide unit, with the structural perturbation being localized to the modified base pair and the neighboring base pair in the 3′ direction (Singh et al., 1993). Subsequently, the consequences of introducing a dinucleotide deletion opposite PdG located within an oligodeoxynucleotide designated PdG-2BD were examined. PdG-2BD modeled the putative strand slippage intermediate which would precede a two base-pair deletion in the *hisD3052* sequence (Scheme 2). NMR indicated a single conformation with PdG and the 3′-neighbor cytosine unpaired. The data indicated PdG to be intrahelical and stacked, whereas the unpaired cytosine was poorly stacked and appeared to be extrahelical. Oligodeoxynucleotides containing this PdG bulge migrated anomalously on nondenaturing polyacrylamide gels, an observation consistent with the introduction of a bend or kink in the DNA helix (Moe et al., 1994).

In the present work, we have refined the solution structure of PdG-2BD using a restrained molecular dynamics/simulated annealing approach, coupled with complete relaxation matrix calculations to verify the agreement between calculated structures and NMR data. The calculations were restrained with 352 total distance constraints, of which 264 were distances derived from the NMR data, and the remainder were empirical restraints. These provided a detailed picture of the solution conformation for PdG-2BD. They provide the first refined solution structure of PdG in the context of a transient strand slippage intermediate which could be important in frame shift mutagenesis. The results demonstrate that strand slippage for PdG in the *hisD3052* $(CG)_3$ iterative repeat can be successfully accommodated by the DNA helix.

MATERIALS AND METHODS

Oligodeoxynucleotide Synthesis. The unmodified oligodeoxynucleotides were synthesized by the Midland Certified Reagent Co. (Midland, TX) and purified by anion exchange chromatography. PdG was synthesized using the methodology of Marinelli et al. (1990). The 5′-dimethoxytrityl (DMT) derivative of PdG was synthesized and purified as previously described (Benamira & Marnett, 1993) and incorporated into oligodeoxynucleotides by the Midland Certified Reagent Co. The extinction coefficients of the oligodeoxynucleotides were calculated based upon nearest neighbor analysis as 1.22×10^5 M^{-1} cm^{-1} for the modified strand 5′-d(ATCGCXCGGCATG)-3′, and 1.01×10^5 M^{-1} cm^{-1} for the complementary strand 5′-d(TAGCGCCGTAC)-3′ (Borer, 1975). Equimolar amounts of each strand were combined, and the resulting duplex was dialyzed (1000 MWCO, Spectrum) with one exchange against 0.6 M NaCl and three exchanges against deionized H_2O. The sample was lyophilized.

NMR Spectroscopy. ^1H NMR spectra were recorded at 500.139 MHz. Data processing used FELIX 2.10 (Biosym Technologies Inc. San Diego, CA) running on Silicon Graphics Iris and Indigo workstations (Silicon Graphics, Inc., Mountain View, CA). The temperature was controlled at 25 ± 0.5 °C. The modified duplex was prepared at a concentration of 2 mM. For observation of nonexchangeable protons, the sample was dissolved in 0.5 mL of 0.01 M sodium phosphate buffer containing 0.1 M NaCl and 0.05 mM Na_2EDTA at pH 7.0. The sample was exchanged three times with 99.9% D_2O, and dissolved in 99.96% D_2O. NOESY experiments utilized the TPPI phase sequence. Chemical shifts were referenced internally to DSS. For observation of exchangeable protons, the sample was dialyzed against 50 mL of 0.001 M sodium phosphate buffer containing 0.5 M NaCl, 0.05 mM Na_2EDTA, pH ~8.4, and 9:1 H_2O/D_2O containing Chelex (Bio-Rad Laboratories,

Table 1: Distribution of Experimental Restraints among Nucleotide Units of PdG-2BD[a]

restraints	A^{-2}	T^{-1}	C^{1}	G^{2}	C^{3}	X^{4}	C^{5}	G^{6}	G^{7}	C^{8}	A^{9}	T^{10}	G^{11}	C^{12}	A^{13}	T^{14}	G^{15}	C^{16}	C^{17}	G^{18}	C^{19}	G^{20}	A^{21}	T^{22}
intranucleotide	7	10	5	5	7	4	7	9	6	8	6	8	7	8	8	9	8	5	8	5	7	6	8	9
internucleotide[b]	0	5	4	3	2	3	4	3	3	4	2	8	2	0	3	7	3	3	5	2	3	3	3	8
cross-strand	0	0	1	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	3	1	0	0	1	0
total, experimental	7	15	10	8	9	12	11	12	9	12	8	16	9	8	11	16	11	8	16	8	10	9	12	17

[a] Nucleotides A^{-2}→G^{11} are in the modified strand, and nucleotides C^{12}→G^{22} are in the complementary strand. X^{4} is PdG. (See Scheme 2). [b] The internucleotide NOEs are listed in the direction $n \rightarrow n - 1$.

Richmond, CA) for three exchanges. Two-dimensional spectra were obtained by replacing the final 90° pulse of the NOESY experiment by a 1−1 binomial pulse (Bax et al., 1987; Sklenar et al., 1987). Convolution difference was used during processing to minimize the residual signal arising from water (Marion et al., 1989).

Starting Structures. The 13-mer duplex d(ATCGCGCG-GCATG)·d(CATGCCGCGCGAT) was built in the classical B-DNA geometry (Arnott & Hukins, 1972, 1973), and the two bases on the complementary strand paired with G^{4} and C^{5} were removed. The procedure was repeated starting from a classical A-DNA 13-mer (Arnott & Hukins, 1972). PdG was constructed at the fourth position by bonding a propano group to N1 and N2 of guanine to form the exocyclic ring. The partial charges on PdG were determined by performing SCF calculations on the free base (excluding the sugar−phosphate portion) using a neutral total charge on PdG. SCF calculations were carried out by the MNDO method using MOPAC (Stewart, 1983). The force field parameterization for PdG is detailed in Table 1S of the supplementary material. The A and B structures were energy-minimized by the conjugate gradients method for 200 iterations without experimental restraints to give the starting structures IniA and IniB (Figure 1) used for subsequent relaxation matrix analysis and molecular dynamics calculations.

Relaxation Matrix Analysis. A set of three NOESY experiments were performed in D_2O buffer, using mixing times of 150, 250, and 350 ms. For each of the three mixing times, two hybrid intensity matrices were constructed using MARDIGRAS (Borgias & James, 1990), which consisted of 317 experimental intensities supplemented with calculated intensities from either IniA or IniB (Arnott & Hukins, 1972, 1973). The noise level in the experimental intensities was estimated as 5% of the diagonal peak, which was input into the MARDIGRAS calculations. Complete iterative relaxation matrix analysis (RMA) calculations were performed on each matrix, yielding six sets of internuclear distances for a given value of molecular correlation time τ_c. The calculations were repeated using isotropic correlation times of 3, 4, and 5 ns, which were chosen to bracket the value for τ_c calculated assuming the rigid rod model for the oligodeoxynucleotides. This created a total of 18 distance sets, which were averaged to obtain the distance restraints used in the subsequent MD calculations. The upper and lower bounds for individual distances in MD calculations were defined to be equal to the corresponding standard deviations measured from the 18 distance sets.

Restrained MD Calculations. Energy minimization and restrained molecular dynamics (MD) calculations were carried out with X-PLOR (Brunger, 1992). All calculations were based on an energy function approach in which the total energy was the sum of the empirical energy of the molecule and effective energy, comprised of the restraint energy terms. The CHARMM force field contributed the usual terms for bonds, bond angles, torsion angles, tetrahedral and planar geometry, hydrogen bonding, and nonbonding interactions, including van der Waals and electrostatic forces. The electrostatic term used the Coulombic function based on a reduced charge set of partial charges and a distance-dependent dielectric constant of 4.0, to mimic solvent screening of charge. The van der Waals term was approximated using the Lennard-Jones potential energy function. The nonbonded pair list was updated if any atom moved more than 0.5 Å, and the cutoff radius for nonbonded interactions was 11 Å. All calculations were run *in vacuo* without explicit counterions. The effective energy function was comprised of two terms describing distance and dihedral restraints, both of which were in the form of a standard square-well potential (Clore et al., 1986). Bond lengths involving hydrogens were fixed with the SHAKE algorithm (Ryckaert et al., 1977) during molecular dynamics calculations.

Five MD calculations were performed using different random seeds on each of the IniA and IniB starting structures. IniA and IniB were further energy minimized by the conjugate gradient method for 200 iterations, with the force constant on NOE distance restraints and empirical base pair distance restraints set to 50 and 30 kcal·mol^{-1}·Å^{-2}, and 20 kcal·mol^{-1}·Å^{-2} for dihedral base pair planarity restraints. The resulting structures were subjected to 30 ps of MD using a simulated annealing protocol. The MD calculations were initialized by assigning a random set of velocities to all the atoms that fit a Maxwell−Boltzmann distribution at 1000 K. The calculation was carried out over 11 000 steps at high temperature and then cooled over 3000 steps to 300 K and equilibrated at low temperature for 16 000 steps. The molecules were weakly coupled to a temperature bath with a coupling constant of 0.05 ps (Berendsen et al., 1984). The time step of the integrator was 1 fs. Structure coordinates were archived every 0.1 ps over the final 10 ps of the MD simulation. Structure coordinates extracted from the last 5 ps of each MD calculation were averaged and energy-minimized. Convergence was monitored by calculating pairwise rmsd values for each of the final structures. Assessment of the accuracy of the calculated MD structures was achieved by performing complete relaxation matrix analysis calculations using CORMA.

Internuclear Distances. The distribution of the 264 NOE restraints which were used for the MD calculations is presented in Table 1. Table 2S in the supplementary material gives a detailed presentation of the restraints. An additional 88 empirical restraints were utilized. This provided an average of 15 restraints/nucleotide, excluding the terminal nucleotides. Nucleotides X^{4} and C^{17} were especially well-defined due to the availability of five cross-strand experimental distances for X^{4} and three cross-strand distances for

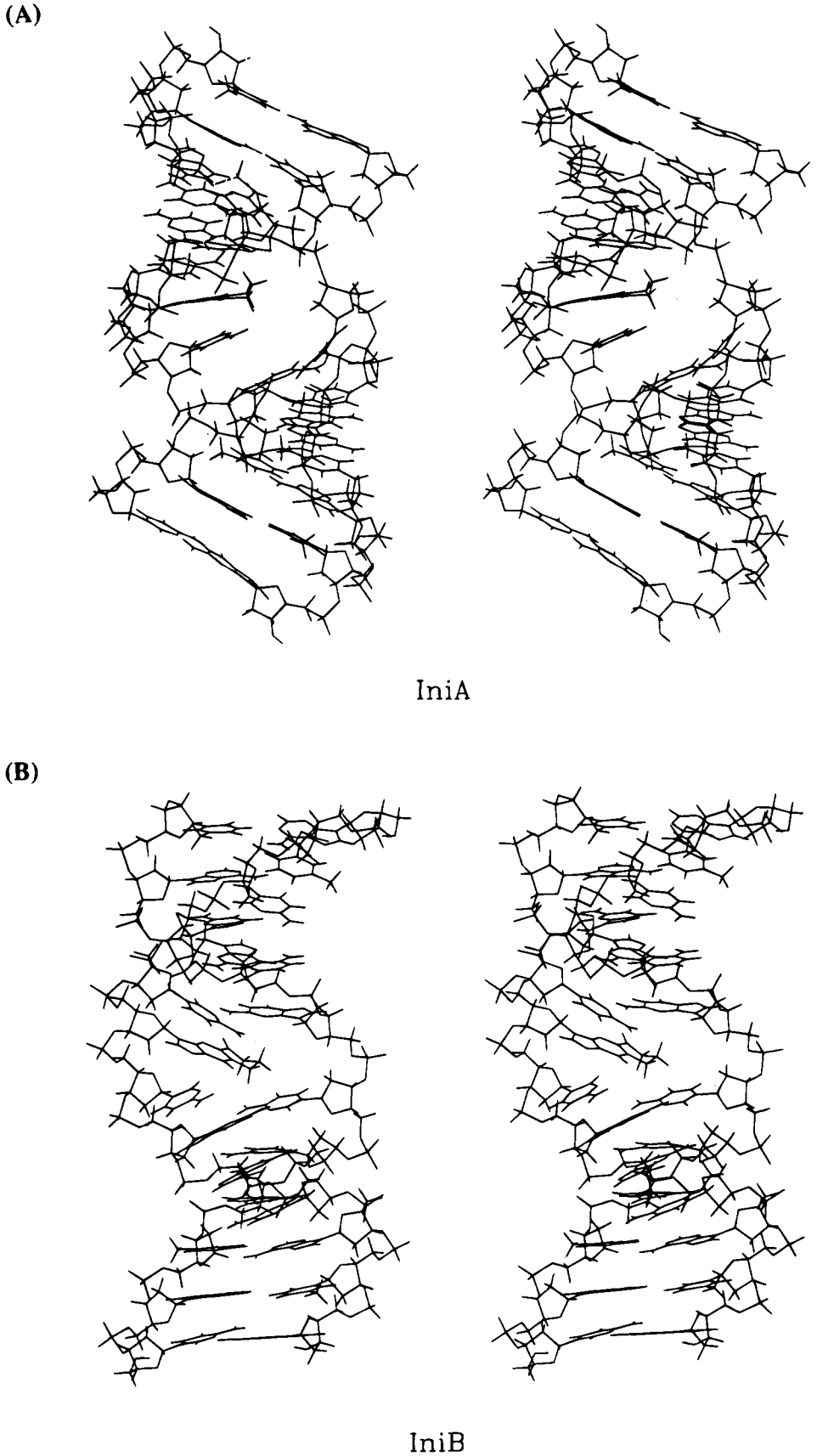


FIGURE 1: Stereoviews of (A) the IniA and (B) the IniB structures for PdG-2BD. The views are into the major groove of the DNA, and the molecules are oriented such that the 5′-terminus of the modified strand is located at the upper right. These stereoviews are meant to be viewed with stereoviewer.

C^{17} (these can be observed in Figure 7). The total number of experimentally measured distances for X^4 was 12 and for C^{17} was 16. For C^5, 11 distances were obtained, but no cross-strand interactions were observed. Distances calculated by MARDIGRAS were removed from the distance set if the values were calculated to be greater than 5 Å, or if they were

inconsistent with a reasonable structure. Inconsistent distance restraints were usually due to large errors in intensity measurements arising from spectral overlap or from cross peaks close to the water resonance whose intensities were altered by the water presaturation pulse used during acquisition of NOESY spectra. After culling poor distance restraints, 259 of the 317 distances calculated by MARDIGRAS remained.

Five additional distance restraints involving the methylenic protons of PdG were derived from the NOESY spectrum obtained in 9:1 H_2O/D_2O buffer (Moe et al., 1994). These were estimated from the size of the cross peaks in the spectrum. The final distance restraints used were as follows: $r[X^4\ H8_a-G^{18}\ NH1] = 4.50 \pm 0.50$ Å, $r[X^4\ H8_b-G^6\ NH1] = 4.75 \pm 0.25$ Å, $r[X^4\ H8_b-C^{17}\ NH4b] = 4.50 \pm 0.50$ Å, $r[X^4\ H6_{a,b}-C^{17}\ H2'] = 4.50 \pm 0.50$ Å, and $r[X^4\ H6_{a,b}-C^{17}\ H2''] = 4.50 \pm 0.50$ Å. This resulted in a total of 266 distance restraints derived from the experimental data. These were split into four classes according to the quality of the calculated distance. The quality was judged by evaluation of the standard deviation of the average distance as determined by multiple MARDIGRAS calculations, the estimated error in the volume measurement for particular cross peaks due to overlap, and the *R* factor between the experimental and calculated peak intensity. Class 1 consisted of the 138 best distance restraints, class 2 contained 54 restraints, class 3 contained 47 restraints, and class 4 consisted of the 25 worst restraints.

Empirical distance and dihedral restraints were used in MD calculations to preserve hydrogen bonding of the Watson–Crick base pairs and to prevent excessive propeller twisting between base pairs that results as an artifact of the MD calculations (Huang & Eisenberg, 1992). There were five distance restraints per G·C, four distance restraints per A·T base pair, and one torsional restraint per base pair to prevent excessive propeller twisting. Without these restraints, propeller twisting of greater than 20° occurred between the base pairs. We determined this excessive propeller twisting to be an artifact because, without the restraints, the propeller twisting occurs to the same magnitude and direction in all PEM and MD calculations, regardless of the DNA model under study.[3] Additional empirical restraints at each residue ensured that the distance between the base H8/H6 protons and the deoxyribose H5′,5″ protons remained greater than 3.5 Å. Empirical restraints are summarized in Table 3S of the supplementary material.

For PEM and MD calculations, distance restraint classes 1–4 were weighted by 100%, 80%, 60%, and 40% of the applied NOE force constant, respectively. The force constant for the distance restraints was set at 50 kcal·mol^{-1}·Å^{-2} for the first 5 ps of the simulated annealing protocol. It was scaled to 150 kcal·mol^{-1}·Å^{-2} over the next 3 ps and maintained at that level for 9 ps. At 17 ps, it was relaxed back to its starting value over 3 ps and then maintained at that value until the end of the simulation. Force constants for empirical hydrogen-bonding distance restraints and base pair planarity followed the same time path as the experimental distance restraints except they were started at 30 and 20 kcal·mol^{-1}·Å^{-2}, respectively, increased to 100 kcal·mol^{-1}·Å^{-2} and then relaxed back to 30 and 20 kcal·mol^{-1}·Å^{-2} during the equilibrium stage. The ceiling factor for NOE restraints was 1000 kcal·mol^{-1}·Å^{-2} for all PEM and MD calculations.

Helicoidal Analysis. The refined structures were analyzed using DIALS AND WINDOWS 1.0 (Ravishankar et al., 1989). The helicoidal analysis was performed separately for the two fully base-paired regions of PdG-2BD, while backbone torsion angles for the two unpaired nucleotides were recorded manually using INSIGHTII for graphical display.

RESULTS

^{1}H NMR Data. A detailed spectroscopic analysis of the NMR data was presented previously (Moe et al., 1994). The sequential NOEs (Feigon et al., 1983; Hare et al., 1983) for the modified strand were uninterrupted except at the $X^4 \rightarrow C^5$ step, where the cross peak between X^4 H1′ and C^5 H6 was missing. The cross peak between C^5 H1′ (5.77 ppm) and G^6 H8 (7.82 ppm) was weak, indicating a greater-than-normal distance between these two protons. For the complementary strand the sequential NOEs were uninterrupted. The deoxyribose spin systems were assigned using standard methods, via a combination of NOESY and TOCSY data. The complete assignment of the H2′, H2″ protons, 22 H3′ protons (C^1 and C^{17} H3′ were not assigned), and 14 H4′ protons was achieved. The H5′,5″ protons were assigned for only the 5′-terminal nucleotides A^{-2} and C^{12}. All of the cytosine amino protons were observable except for C^5 NH_2. The amino proton resonances for C^3 were closely superimposed with those arising from C^{19} but could be partially resolved at lower temperatures. All guanine and thymine imino protons were assigned. The exchangeable proton N5H of PdG was located at 6.4 ppm and exhibited NOE connectivity to the methylene protons of PdG.

Restrained Molecular Dynamics. The MD simulated annealing protocol converged toward a final structure when initiated from either IniA or IniB. A final average structure (rMD structure) was obtained by averaging the coordinates from the 10 MD calculations, followed by 200 iterations of PEM using the conjugate gradients method. This is shown in CPK representation in Figure 2. Convergence was supported by pairwise rmsd comparisons tabulated in Table 2. Table 2 illustrates the large structural difference between IniA and IniB, which was anticipated. A rmsd of 5.50 Å was measured between these two structures. Comparison of IniA and IniB with ⟨rMDA⟩ and ⟨rMDB⟩, the average structures which emerged from the MD calculations starting from IniA and IniB (Table 2), revealed that the simulated annealing protocol resulted in a significantly greater conformational change for IniA than for IniB. The rMD structure was closer to IniB than to IniA. This was consistent with the notion that the refined structure was more B-like, although, as predicted for the modified oligodeoxynucleotide, significant differences remained between IniB and rMD, evidenced by a rmsd of 2.3 Å. The rms difference between ⟨rMDA⟩ and ⟨rMDB⟩, the sets of structures emergent from the IniA- and IniB-based calculations, was observed to be 1.67 Å, while comparison of the individual sets of structures derived from IniA and IniB among themselves resulted in rms differences of 0.84 Å for ⟨rMDA⟩ and 1.07 Å for ⟨rMDB⟩ (Table 2). The conformations of the ⟨rMDA⟩ and ⟨rMDB⟩ structures were therefore individually well-refined

[3] This problem has also been noted in recent newsletters distributed to the X-PLOR user group.

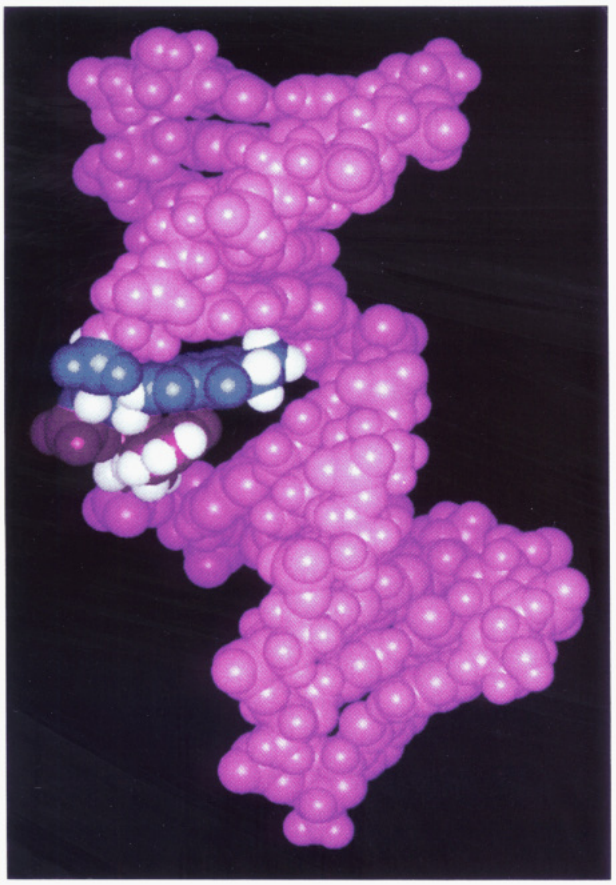


FIGURE 2: CPK representation of the final rMD structure for PdG-2BD. PdG is shown in blue, C^5 in red. Protons on PdG and C^5 are shown in white. This model structure is based upon averaging the coordinates from 10 MD calculations (see text).

Table 2: Root Mean Square (rms) Deviations, Excluding the End Base Pairs, between Various Initial Structures, Intermediate Structures, and Final Structures of PdG-2BD

	atomic rms difference (Å)
initial structures	
IniA vs IniB	5.50
rms shifts	
IniA vs ⟨rMDA⟩[a]	5.87 ± 0.16
IniB vs ⟨rMDB⟩[b]	2.33 ± 0.08
rms distributions	
⟨rMDA⟩ vs ⟨rMDA⟩	0.84 ± 0.14
⟨rMDB⟩ vs ⟨rMDB⟩	1.07 ± 0.21
⟨rMDA⟩ vs ⟨rMDB⟩	1.67 ± 0.28
⟨rMD⟩[c] vs ⟨rMD⟩	1.35 ± 0.43
⟨rMDA⟩ vs rMD[d]	0.93 ± 0.11
⟨rMDB⟩ vs rMD	1.15 ± 0.27
⟨rMDh⟩[e] vs ⟨rMDh⟩	0.99 ± 0.30
⟨rMDt⟩[f] vs ⟨rMDt⟩	0.89 ± 0.21

[a] ⟨rMDA⟩ represents the set of five structures which emerged from MD calculations starting with IniA. [b] ⟨rMDB⟩ represents the set of five structures which emerged from MD calculations starting from IniB. [c] ⟨rMD⟩ represents the set of 10 structures which emerged from MD calculations starting from both IniA and IniB. [d] rMD represents the average minimized structure from all 10 MD calculations. [e] ⟨rMDh⟩ represents the set of 10 structures for the five base pairs on the 5′-side of the lesion (the "headpiece"). [f] ⟨rMDt⟩ represents the set of 10 structures for the six base pairs on the 3′-side of the lesion (the "tailpiece").

but were in fact slightly different. Since comparisons of the headpiece and tailpiece for all 10 structures revealed convergence to below 1 Å, it was concluded that the difference between rMDA and rMDB arose from a slightly different conformation localized at the site of the lesion and unpaired nucleotides. Figure 3 shows the stereoviews and overlays of the final MD structures.

Table 3: Comparison of Sixth-Root Residual Indices R_1^x for Starting Models and Resulting rMD Structures[a]

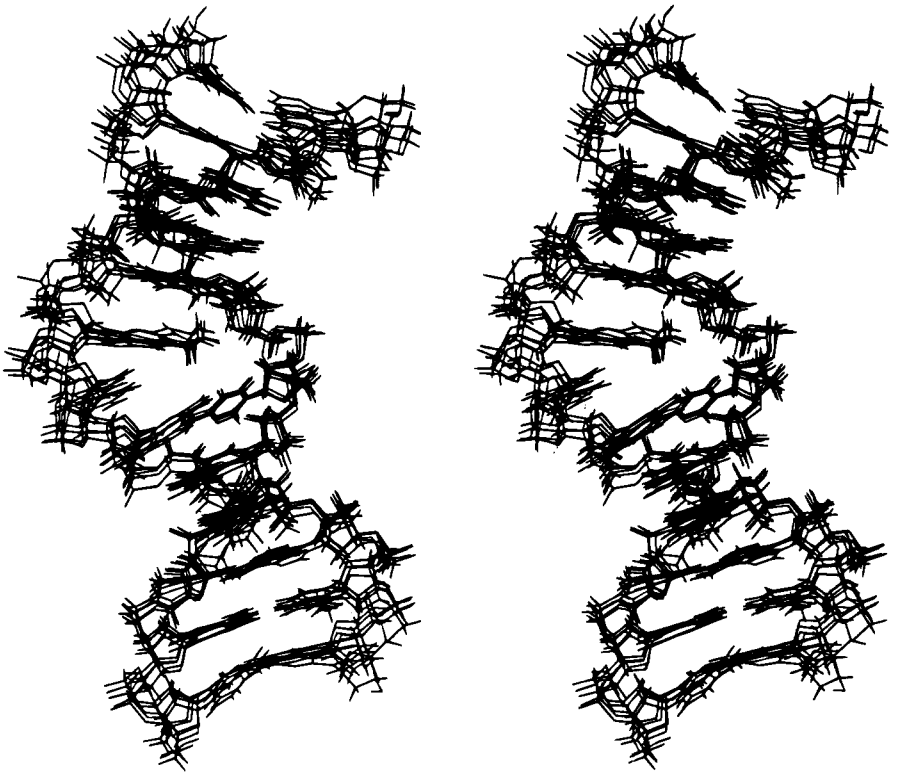
structure	intraresidue R_1^x	interresidue R_1^x	total R_1^x
IniA	0.114	0.154	0.125
IniB	0.091	0.114	0.097
rMDA	0.067	0.124	0.078
rMDB	0.083	0.084	0.084
rMD	0.068	0.119	0.072

[a] $R_1^x = \sum|(a_o)_i^{1/6} - (a_c)_i^{1/6}|/\sum|(a_o)_i^{1/6}|$, where (a_o) and (a_c) are the intensities of observed (nonzero) and calculated NOE cross peaks.

Complete Relaxation Matrix Calculations. The results of back-calculation of the intensity matrix from the converged rMD structures are tabulated in Table 3. The sixth-root R factor, R_1^x, measured the difference between NOE intensities calculated by CORMA from the model structure and the intensities measured from the NOESY spectra (Keepers & James, 1984). The inverse sixth-root deviations were calculated from the NOESY experiment measured with a mixing time of 250 ms, using an isotropic molecular correlation time of 4 ns. In general, they indicated good agreement between the calculated intensities and the NOE data. In comparing R_1^x for the two starting structures, the value for IniB was lower than the R factor for IniA, but both were on the order of 0.1 or greater. The values of the total R_1^x decreased to 0.078 and 0.084, respectively, for the emergent rMDA and rMDB structures. The sixth-root index indicated better agreement with the intraresidue NOE intensities than for interresidue intensities. The slight conformational differences between rMDA and rMDB were evidenced by the larger value of 0.124 observed for interresidue NOE connectivities of rMDA, which suggested that the rMDA structures were slightly less accurate than were the rMDB structures.

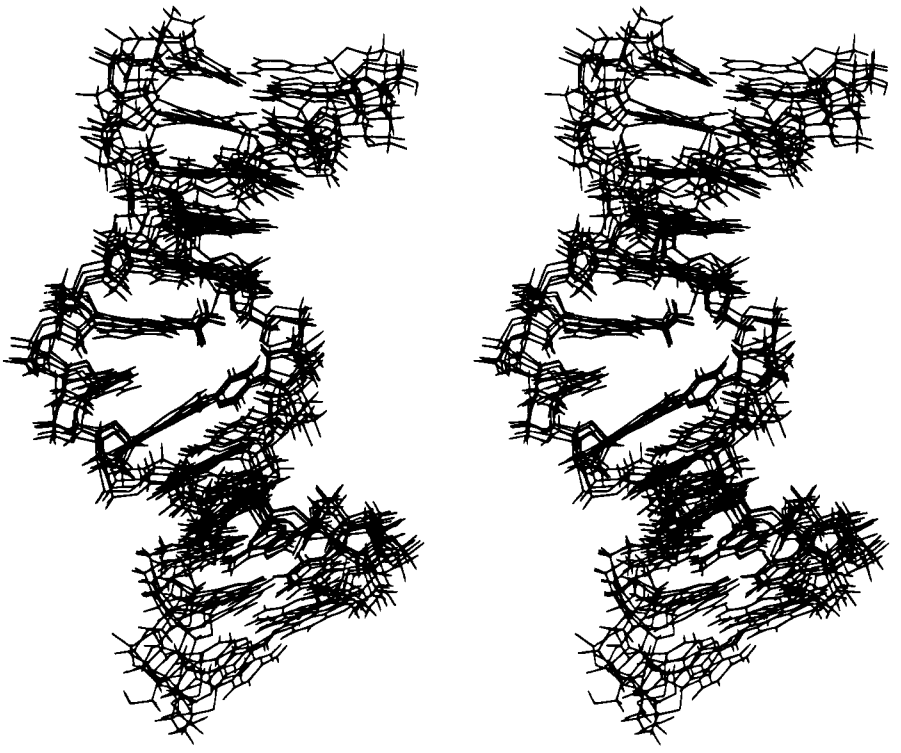
Refined Structures. (a) Backbone Torsion Angles. The calculated values for the backbone torsion angles $\alpha-\zeta$ are shown graphically in Figure 4, which also indicates the expected values for canonical A- and B-form DNA duplexes. With the exception of angles ϵ and ζ, there was poor agreement with either A- or B-form canonical values (see Discussion). The data for ζ agreed with B-form DNA, while the angles for ϵ were more consistent with A-form. Angles δ, ϵ, and ζ, which correspond to C4′−C3′, C3′−O3′, and O3′−P, were well-defined by the MD calculations, as evidenced by convergence between the values for rMDA and rMDB. In several instances, large relative changes were observed, which correlated with the site of the lesion (Figure 4). The C3′−O3′ torsion angle ϵ increased from 160−170° to 180−190° at bases C^3, X^4, and C^5. There was a large increase in the torsion angle ϵ at nucleotide C^{17} of the complementary strand, to approximately 210°. The O3′−P torsion angle ζ decreased from 270−280° to 250−260° at nucleotides C^3 and C^5, flanking the lesion. The values of ϕ, the sugar pseudorotation angle, also did not agree with either canonical B-form or A-form DNA but appeared to be reasonably well converged in the structures which emerged from the MD calculations. With the exception of nucleotides G^2, C^3, X^4, and C^5, ϕ remained >100°, consistent with an S conformation of the deoxyribose moieties. The values of χ were consistently greater than 240°, indicative of an *anti* conformation about the glycosidic bond.

(A)



<rMDA>

(B)



<rMDB>

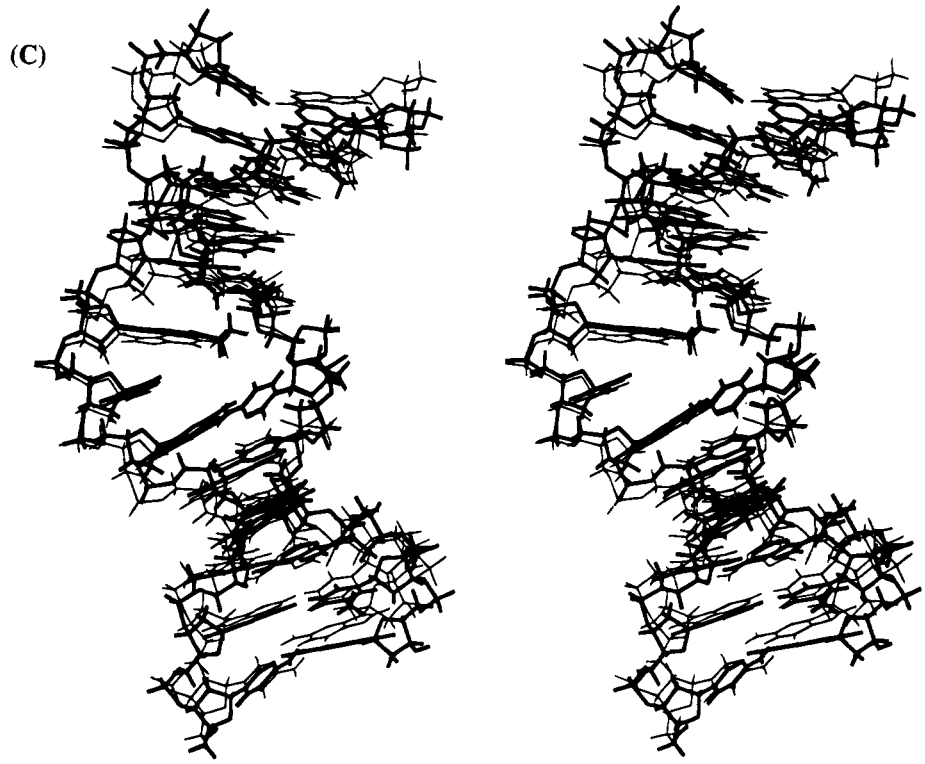


FIGURE 3: Stereoviews of rMDA and rMDB structures. (A) Superposition of five randomly seeded MD calculations based upon the IniA starting structure. (B) Superposition of five randomly seeded MD calculations based upon the IniB starting structure. (C) Superposition of the rMDA (bold) and rMDB structures, indicating that the experimental data are sufficient to drive either starting structure to a common rMD structure.

(b) Intra-Base-Pair Helical Parameters. The plots of the calculated intra-base-pair helical parameters are shown in Figure 5A. The values for stretch, shear, stagger, and buckle converged well from both IniA and IniB starting structures. They appeared to be in reasonable agreement with the canonical values for B-form DNA, although this was clearly evident only in the data for the stretching parameter, since for shear, stagger, and buckle, A-form and B-form DNA are similar. Propeller twist, opening, and inclination also converged well but did not agree well with either canonical A-form or B-form DNA. When perturbations at the terminal base pairs were discounted, the largest remaining perturbations were consistently observed at base pairs C^3•G^{18} and G^6•C^{17} (base pairs 5 and 6), adjacent to the unpaired nucleotides. The data for shear and buckling revealed opposite trends at these two base pairs, indicating that these base pairs buckled away from the two unpaired nucleotides and stretched on either side to accommodate the intrahelical position of X^4. There was a large negative shear and stagger and increased base pair opening at base pair 6 (G^6•C^{17}), which accommodated the extrusion of C^5 into the major groove while maximizing aromatic stacking interactions with base pair 7 (G^7•C^{16}).

(c) Inter-Base-Pair Helical Parameters. The calculated inter-base-pair helical parameters are shown in Figure 5B. Each of the parameters shown converged well from both IniA and IniB. While these parameters were centered on or near the canonical values for B-form DNA, large fluctuations were observed in the calculated values at individual base steps. The rise per base pair (not shown) consistently refined to a value ~3.0 Å, midway between A-form and B-form DNA. The bend in PdG was illustrated by the large negative roll at the base step between the G^2•C^{19} and C^3•G^{18} base pairs (base step 4) and significant deviations in the tilt angles at these two base pairs and between the G^6•C^{17} and G^7•C^{16} base pairs (base step 6). A decrease was not observed in the helical twist angle, as might have been predicted from the insertion of PdG into the duplex.

DISCUSSION

Refined Structure of PdG-2BD. Adduct-induced bending of the DNA duplex, as observed in Figures 2 and 3, separates PdG-2BD into two distinct "domains", designated as the "headpiece" and the "tailpiece". The headpiece consists of base pairs A^{-2}•T^{22}, T^{-1}•A^{21}, C^1•G^{20},G^2•C^{19}, and C^3•G^{18}, while the tailpiece consists of base pairs G^6•C^{17}, G^7•C^{16}, C^8•G^{15}, A^9•T^{14}, T^{10}•A^{13}, and G^{11}•C^{12}. PdG is stacked below the "headpiece". This can be observed in Figure 6A, which shows X^4 beneath base pair C^3•G^{18}. In this arrangement, the exocyclic ring of PdG is located beneath the aromatic ring of G^{18}, which accounts for the large upfield chemical shifts observed for the methylenic protons of PdG in PdG-2BD, as compared to the PdG monomer, or the PdG•C^+ Hoogsteen-like pairing interaction in the fully complementary duplex (Singh et al., 1993). The intrahelical location also explains the upfield shift observed for the NH proton of the PdG exocyclic moiety. There was no direct data to define the conformation of the exocyclic propano ring of PdG, but regardless of the starting conformation the exocyclic propano ring always converged to the same conformation after MD calculations. This would indicate that this is the low energy conformation as dictated by the energy function and the force

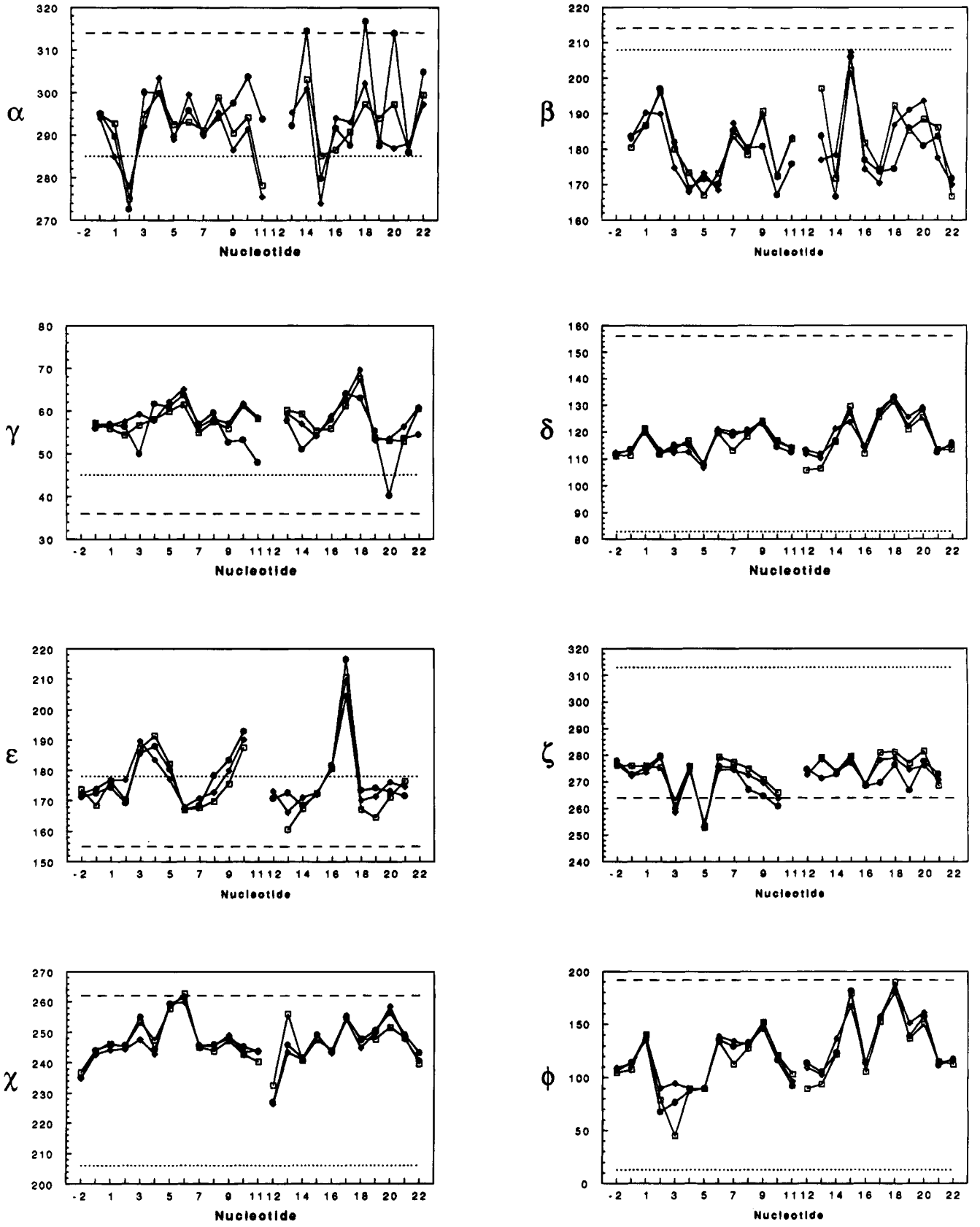


FIGURE 4: Backbone torsion angles (α−ζ), glycosidic torsion angles (χ), and sugar pseudorotation angles (φ) for the structures rMDA (◇), rMDB (□), and rMD (●). Within each graph, (− − −) represents the expected value for B-form DNA, whereas (- - -) represents the expected value for A-form DNA.

fields but may also be indirectly due to experimental restraints. At the junction of the two domains, the unpaired nucleotide C^5 is extruded toward the major groove. The "tailpiece" stacks below C^5 and is oriented at an angle of 20−35° with respect to the headpiece. The extrusion of C^5 toward the major groove is examined in Figures 6B and 7. The refined structure provides evidence that distortions involve both the modified and complementary strand. It

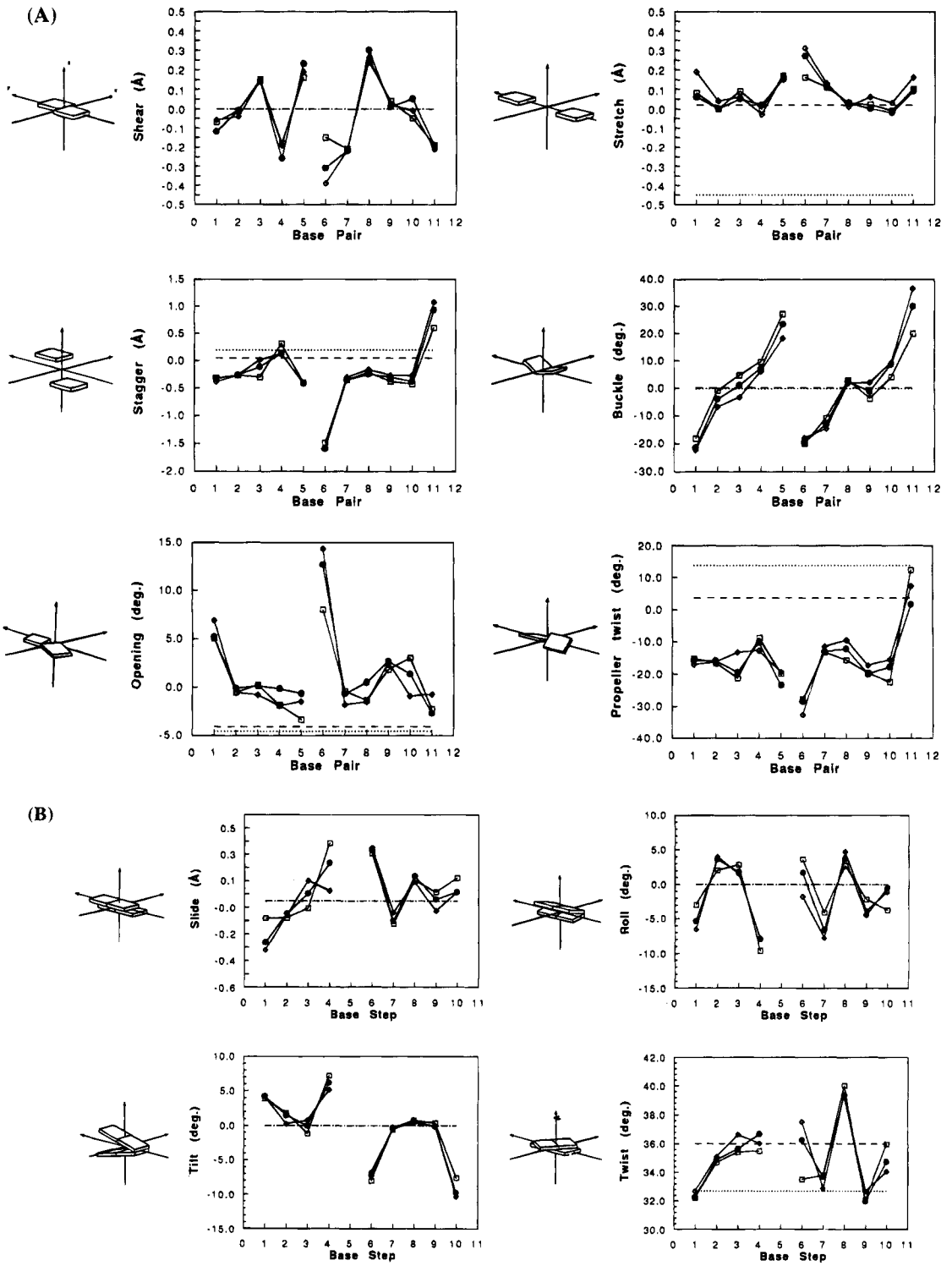


FIGURE 5: Helicoidal parameters of PdG-2BD. Graphical representations of each parameter are provided next to each of the graphs. Within each graph, (— — —) represents the expected value for B-form DNA, whereas (- - -) represents the expected value for A-form DNA. (A) Intra-base-pair helical parameters. (B) Inter-base-pair helical parameters; (◇) represents rMDA, (□) represents rMDB, and (●) represents rMD. The gaps seen in the graphs are at the site of the two base bulge.

corroborates previous spectroscopic and electrophoretic evidence which suggested that PdG was inserted opposite the two-base deletion in the complementary strand, adopted an *anti* conformation with respect to the glycosidic bond, induced DNA bending, and resulted in a localized structural perturbation involving the modified base X^4 and its 3′-neighbor C^5 (Moe et al., 1994).

Distance Restraints. Experimentally measured interproton distances provided the restraints necessary for structural refinement of PdG-2BD. In the initial stages of this work, the isolated spin-pair approximation (ISPA) was used to extract internucleotide distances from NOE volumes. The method required use of short mixing times to minimize the effects of spin diffusion in constructing build-up curves for NOE intensities. Accurate measurement of cross peak volume was impaired as a result of low signal-to-noise ratios for weak cross peaks. The problem was exacerbated by the realization that the most important distances, those between

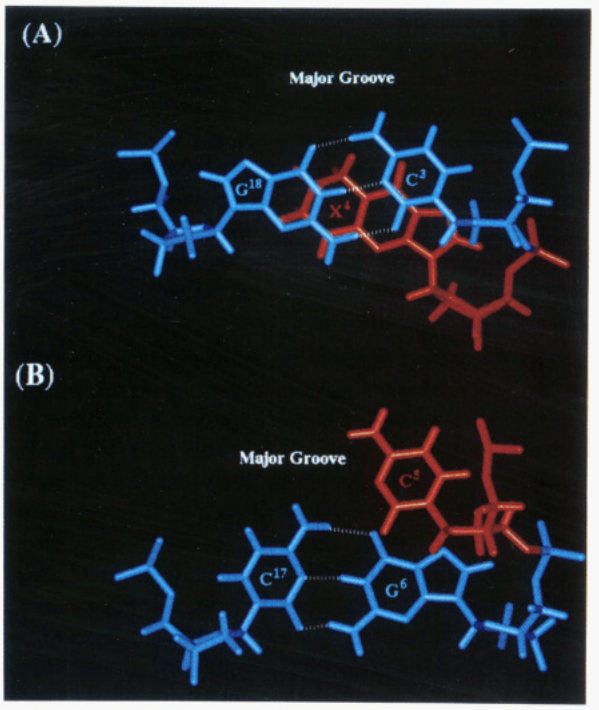


FIGURE 6: (A) Stacking pattern of X^4 beneath the C^3•G^{18} base pair. C^3•G^{18} forms the last base-pairing interaction in the "headpiece" of PdG-2BD, and PdG stacks well beneath this base pair. (B) Stacking pattern of C^5 above the G^6•C^{17} base pair. G^6•C^{17} forms the first base pair of the "tailpiece" of PdG-2BD. Note the extrusion of C^5 into the major groove.

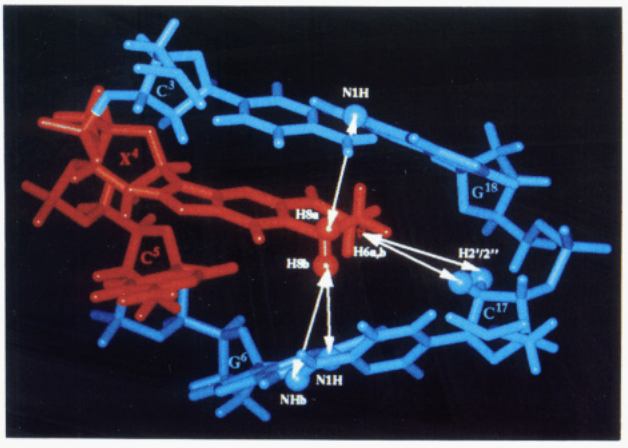


FIGURE 7: View looking into the helix from the major groove at the site of the X^4C^5 strand slippage. The NOEs which support this structure are indicated by the white arrows. In the case of X^4 $H6_{a,b}$, which could not be individually resolved in the spectrum, a pseudoatom was used in the restraint file (see text).

4−6 Å, had the greatest uncertainty. Additional distortions arose from *J*-coupled zero-quantum coherence and from *zz* terms that arose due to rf field inhomogeneity. These artifacts were most prominent at mixing times less than 100 ms (Macura et al., 1982; Rance et al., 1985; Bodenhausen et al., 1984). The structures which emerged from MD calculations based upon ISPA-derived distances were found to be qualitatively similar to the structures presented in the present work, but with a greater range of rmsd, consistent with the smaller number of input restraints.

The complete RMA provided a more rigorous method for obtaining experimental distances (Boelens et al., 1989; Kaluarachchi et al., 1991). It incorporated the effects of spin diffusion at longer mixing times, where signal-to-noise ratios were more favorable, enabling improved integration of crosspeak volumes. A hybrid intensity matrix comprised of the experimentally measured intensities and calculated intensities for missing experimental data was constructed using MARDIGRAS (Borgias & James, 1990), which iteratively calculated the full relaxation matrix and provided distances. Another advantage of RMA was that it provided a substantially larger set of distance restraints for subsequent incorporation into MD calculations than it was possible to obtain from the ISPA approach.

Molecular Correlation Times. RMA required either an experimentally measured or a calculated value for the molecular correlation time, τ_c, under the conditions in which the data were collected. In this work, τ_c was calculated based upon the assumption that the oligodeoxynucleotide approximated a rigid rod in solution.[4] The rigid rod model did not take into account the observation that PdG-2BD is bent at the site of the lesion. The degree to which the bend in PdG-2BD is flexible has also not been established; i.e., whether the bend induces a hinging motion at the lesion site. The narrow line widths of the NMR data suggest that flexing, if present, must be rapid on the NMR time scale. The single correlation time model is undoubtedly incorrect at the terminal base pairs. NOE build-up curves for terminal cytosine H5−H6 cross peaks in a typical oligodeoxynucleotide differed from the buildup rates for the internal cytosines. This was believed to be a consequence of motional averaging of the terminal base pairs due to fraying. Corroborating evidence was the failure to observe the imino protons at these sites. Data for terminal base pairs in this work should thus be interpreted cautiously. We were interested in the lesion and immediately adjacent base pairs, so error in the terminal base pairs was tolerable.

Restrained Molecular Dynamics/Simulated Annealing. The refinement protocol utilized a period of high temperature dynamics to enable the trajectory of PdG-2BD to overcome local energy barriers and, hence, to sample more conformational space (Nilsson et al., 1986). The precision of the emergent structures was measured by the ability of the experimental constraints to drive the calculations toward a common structure, irrespective of the geometry of the starting structure. The strategy involved starting the calculations from distinctly different structures, which in this case were the IniA and IniB structures derived from fiber diffraction coordinates (Arnott & Hukins, 1972). The helical portions of the molecule ("headpiece" and "tailpiece") converged well, indicating that the experimental restraints defined an internally consistent structure. The structures were less converged at the site of PdG and the unpaired C^5.

Accuracy of the Refined Structures. The accuracy of the emergent structure was asessed by monitoring the agreement between the calculated intensities of the model structure with the experimental NOE intensities. The refined structure was concluded to be accurate as judged using sixth-root *R*-factor analysis (James, 1991) (Table 3). The observed R_1^x values of ~0.08 are representative for structures refined from NOE data alone. The recently reported structure of an unmodified

[4] With a benzo[a]pyrene adduct, time-resolved measurement of fluorescence anisotropy provided an independent method for measuring molecular correlation time; the experimentally measured values were in reasonable agreement with the calculated values (Dr. Irene S. Zegar and M. P. Stone, unpublished results).

DNA sequence from the HIV-1 genome, which incorporated data for deoxyribose sugar puckering, achieved R_1^x indices as low as 0.05 (Mujeeb et al., 1993). In comparison, PdG-2BD is a modified oligodeoxynucleotide for which deoxyribose puckering restraints were not included in the refinement protocol. Thus, the observed R_1^x values appear to be reasonable.

While analysis of helicoidal parameters for the refined PdG-2BD structures revealed that the structure was well-converged, variances as compared to either canonical A-form or B-form DNA were observed in many instances (Figures 4 and 5). Differences between the canonical B-form structure and the refined PdG-2BD structure had been anticipated at and adjacent to the lesion site, resulting from the necessity of accommodating the unpaired nucleotides PdG and C^5 into the helix. The variances observed from the canonical B-DNA values in the headpiece and tailpiece of PdG-2BD for many of the helicoidal parameters perhaps illustrate limitations in structure refinement for adducted oligodeoxynucleotides, since the spectroscopic data suggested that, on average, both of these regions of the molecule probably do not deviate substantially from B-like helices (Moe et al., 1994).

The primary restraints used to define the solution structure of PdG-2BD were internucleotide $^1H-^1H$ NOEs. Due to the heterogeneous distribution of NOE data, a number of torsion and helicoidal parameters remain underdefined. In the absence of sufficient experimental restraints, the time-dependent trajectory of these torsion angles during the MD calculations is largely determined by the force field. Figure 5A reveals that the calculations consistently imparted negative propeller twist at each position of the oligodeoxynucleotide, seemingly inconsistent with either canonical A-form or B-form DNA. These were calculated independently of DNA sequence, a phenomenon observed by others (Baleja et al., 1990; Kim & Reid, 1992; Schmitz et al., 1991, 1992; Stolarski et al., 1992). The frequent disagreement between calculated values for backbone torsion angles and either A-form or B-form DNA (Figure 4) probably also reflects the inability of ^1H NOE data to adequately define these torsion angles. The present work utilized an empirical energy function derived from the CHARMM force field (Brooks et al., 1983) which was developed for nucleic acids and treated all hydrogens explicitly (Nilsson & Karplus, 1986). However, similar observations can be made from a MD-derived structure in which the AMBER force field (Weiner et al., 1984) was utilized (Mujeeb et al., 1993), which suggests that the variances in helicoidal parameters generated in the MD calculations are not specific to CHARMM. The incorporation of additional empirical restraints, particularly for the phosphodiester backbone (Huang & Eisenberg, 1992; Huang et al., 1993), might drive the calculations toward the canonical values. However, to allow the nucleotides at and adjacent to PdG to evolve in a more unrestricted manner, we elected to minimize the number of empirical restraints utilized in the MD calculations.

DNA Bending. While the NOE data does not restrain the backbone torsion angles particularly well, the MD calculations may provide insight as to which bonds could be involved in PdG-induced DNA bending. Figure 4 shows a substantial change was calculated for the C3′−O3′ torsion angle ϵ at nucleotide C^{17} of the complementary strand in PdG-2BD. This bond is directly opposite the unpaired nucleotides. The increase in ϵ resulted in the formation of a kink in the complementary strand. The calculated decrease in the angle β as well as an increase in γ at bases X^4, C^5, and G^6 and an increase in ϵ at bases C^3, X^4, and C^5 allow for a gentle roll in the backbone of the modified strand at the bulge region. The decrease in the sugar pseudorotation angle ϕ at bases G^2, C^3, X^4, and C^5 helps accommodate the roll in the backbone of the modified strand. The negative shear and stagger and increased base-pair opening at base pair G^6•C^{17} accommodate the extrusion of C^5 into the major groove while maximizing aromatic stacking interactions with base pair G^7•C^{16}. The calculations suggest that the overall effect of incorporating the modified X^4C^5 slippage unit into the DNA could be described as inducing a bend into two cylinders of B-DNA at the site of the bulge. Further insight into the PdG-induced DNA bend was provided by the perturbations observed in the intra- and inter-base-pair helical parameters (Figure 5A,B), as described under Results. There was a large increase in slide at the base steps between G^2•C^{19} and C^3•G^{18} (base step 4) and G^6•C^{17} and G^7•C^{16} (base step 6). There was a large negative roll at base step 4 and significant and opposite deviations in tilt at base steps 4 and 6 which provided an indirect measure of DNA bending.

In PdG-2BD, the PdG moiety exhibits a strong tendency to stack into the helix, whereas the 3′-neighbor C^5 is extruded into the major groove. The present results may be compared with those from a study in which PdG was placed opposite an apurinic site (Kouchakdjian et al., 1991). In that instance, PdG was also found to be intrahelical. Oligodeoxynucleotides containing unpaired bases or bulges of various sequence context and length have been examined using NMR spectroscopy (Patel et al., 1982b; Morden et al., 1983, 1990; Morden & Maskos, 1993; Rosen et al., 1992a,b; Plum et al., 1992; Aboul-ela et al., 1993). Thermodynamic measurements examined A•T rich DNA containing unpaired bases (LeBlanc & Morden, 1991; Zieba et al., 1991). DNA successfully accommodates unpaired bases in both intrahelical and extrahelical environments (Rosen et al., 1992a), the conformation being dependent both upon the identity of the unpaired base and upon the specific DNA sequence flanking the site (Morden & Maskos, 1993). As for the present example with PdG, unpaired purines generally adopt an intrahelical conformation in solution as evaluated by NMR spectroscopy (Patel et al., 1982a; Nikonowicz et al., 1990). For unpaired pyrimidines both intra- (van den Hoogen et al., 1988a,b) and extrahelical conformations (Morden et al., 1983, 1990; Kalnik et al., 1990; Morden & Maskos, 1993) have been observed. The observation that three additional bases can simultaneously adopt an intrahelical conformation illustrates the conformational flexibility of B-form DNA (Rosen et al., 1992a,b).

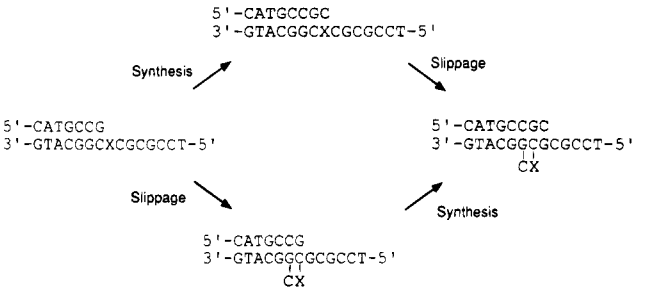
Biological Implications. The alternating $(CG)_4$ repeat sequence from the *S. typhimurium hisD3052* gene represents a hotspot for adduct-induced frame shift mutations, commonly observed to be deletions of CG (Isono & Yourno, 1974; Fuscoe et al., 1988; Benamira et al., 1992; Bell et al., 1991; DeMarini et al., 1992). When embedded within this $(CG)_3$ repeat sequence from the *hisD3052* gene, PdG equilibrates between distinctive conformations, which may lead to distinct mutagenic events. When located opposite cytosine in the complementary strand at pH 5.8, PdG is oriented such that the propano moiety faces into the major groove (Singh et al., 1993). PdG is in the *syn* conformation

at the glycosyl bond and forms a Hoogsteen-like base-pair with cytosine protonated at N3. This orientation of PdG predicts the possibility of error-free lesion bypass, via insertion of cytosine opposite PdG in the *syn* conformation. Conversely, at higher pH, PdG remains in the *anti* conformation about χ, in which case it is expected to be a noninformational lesion. Under these conditions, incorporation of adenine opposite PdG would be predicted to lead to PdG→T mutations. The apparent pK_a for this conformational transition was measured to be 7.0, which indicates that significant amounts of both structures could form at physiological pH.[5] A second pH-dependent conformational equilibrium was observed in the fully complementary duplex, in which the 3′-neighbor base pair equilibrated between Hoogsteen and Watson−Crick base-pairing. The exocyclic lesion thus perturbed two base pairs, which represented a single repeat unit in the iterated $(CpG)_3$ sequence (Singh et al., 1993). This may promote transient strand slippage of a dinucleotide repeat unit and correlate with the genesis of two base pair deletions. The present studies provide the first example of how PdG might be accommodated within a transiently slipped primer−template complex, the putative intermediate which would result in a frame shift mutation.

Conformational equilibration may be common to adducts arising from various chemical mutagens and could have significant implications in determining mechanistic pathways of adduct-induced mutagenesis. PdG adducts opposite adenine also undergo conformational exchange (Kouchakdjian et al., 1989a, 1990; Huang & Eisenberg, 1992; Huang et al., 1993). Both Fuchs (Veaute & Fuchs, 1991) and Krugh (Eckel & Krugh, 1994) reported conformational equilibria involving AAF adducts, as did Cho and Beland for 4-aminobiphenyl (Cho et al., 1992) and Cho for AAF (Cho et al., 1994). The idea was advanced in the case of benzo[a]pyrene adducts by Loechler (Rodriguez & Loechler, 1993a,b). Evidence for conformational rearrangement of the (+)-*trans*−*anti*-[BP]dG adduct following the generation of a deletion site was recently observed (Cosman et al., 1994a), using a strategy similar to the present approach we have developed for examining two-base deletions induced by PdG (Moe et al., 1994), which utilizes PdG-2BD. A similar conformational rearrangement was not observed for the (+)-*cis*−*anti*-[BP]dG adduct (Cosman et al., 1994b). Eckel and Krugh (1994) introduced the term "mutagenic switch" to describe how a *N*-2-aminofluorene (AF) adduct might be successfully bypassed under certain conditions and be mutagenic under other conditions. Using their terminology, the present results with PdG in the *hisD3052* sequence could describe a three-way "mutagenic switch", in which the possibilities are successful by-pass of PdG, miscoding leading to base substitution, or strand slippage leading to a frame shift.

The mutagenic switch hypothesis suggests that multiple mutagenic and nonmutagenic outcomes are available to PdG embedded within a DNA sequence such as the *hisD3052* sequence and argues that the actual pathway chosen by PdG will be modulated by a number of factors. These may include sequence (Burnouf et al., 1989; Koehl et al., 1989) and the identity of the particular polymerase or repair enzyme (Shibutani et al., 1993; Shibutani & Grollman, 1993). In the *hisD3052* $(CG)_3$ sequence, transient slippage of the modified dinucleotide repeat unit in the template could occur during replication, allowing utilization of the 5′-neighbor repeat unit either for nucleotide insertion or for elongation (Kunkel, 1990) (Scheme 3). Evidence that the base inserted opposite PdG during *in vitro* replication of adducted template primers correlates with the identity of the 5′-neighbor nucleotide (Shibutani & Grollman, 1993) predicts the relative frequency of two-nucleotide deletions vs successful lesion bypass would be determined by the relative rates of elongation from the strand slippage intermediate vs relaxation of the strand slippage intermediate to the correctly paired structure. Base-pair substitutions and one-base deletions occur opposite PdG in nonreiterated sequences (Shibutani & Grollman, 1993), whereas two-base deletions occur in the $(CG)_4$ repeat.[6]

Scheme 3: Potential Mechanisms of Two-Nucleotide Deletions in the *hisD3052* Gene $(CG)_3$ Repeat Sequence[a]



[a] In the top pathway, nucleotide insertion occurs opposite the lesion, followed by stalling of the replication apparatus. Transient strand slippage allows replication to continue. In the bottom pathway, the replication apparatus stalls prior to the lesion. Transient strand slippage removes the lesion from the active site and allows replication to continue.

Summary. This work provides the first observation of the behavior of a PdG-modified oligodeoxynucleotide in the context of a strand slippage structure, which models the transient intermediate leading to −2 frame shifts in a $(CG)_3$ repeat sequence. PdG-2BD was refined to a well-defined set of structures. Complete relaxation matrix analysis demonstrated that the refined structures were in good agreement with the NOE data. The PdG base is intrahelical and stacked within the duplex; C^5 is extruded into the major groove. The bulge induces a bend of approximately 20−35° in the duplex. The data suggest that formation of a two-nucleotide strand slippage intermediate, potentially leading to a two-nucleotide deletion of a CG repeat unit from the iterated $(CG)_3$ repeat sequence, could be accommodated without major distortion of the DNA helix. It seems plausible that if replication stalled at PdG in the PdG-2BD oligodeoxynucleotide, slippage of the two nucleotide unit X^4C^5 could occur without major conformational reorientation at the active site of the polymerase, enabling replication beyond the lesion.

ACKNOWLEDGMENTS

We thank Dr. K. Balasubramanian and Mr. Walter Siddiqui for their assistance with software management and workstation networking. Dr. Irene Zegar and Ms. Binbin Feng provided helpful discussions regarding the molecular

[5] J. P. Weisenseel, J. G. Moe, G. R. Reddy, L. J. Marnett, and M. P. Stone, unpublished results.

[6] M. Hashim and L. J. Marnett, manuscript in preparation.

refinement strategy. Ms. Randi Tinkham provided valuable assistance in the preparation of the manuscript.

SUPPLEMENTARY MATERIAL AVAILABLE

Three tables showing force field parameterization values used for PdG, detailing experimental distances and classes of restraints, and summarizing the empirical restraints (9 pages). Ordering information is given on any current masthead page.

REFERENCES

Aboul-ela, F., Murchie, A. I., Homans, S. W., & Lilley, D. M. (1993) *J. Mol. Biol. 229*, 173−188.

Arnott, S., & Hukins, D. W. L. (1972) *Biochem. Biophys. Res. Commun. 47*, 1504−1509.

Arnott, S., & Hukins, D. W. L. (1973) *J. Mol. Biol. 81*, 93−105.

Baleja, J. D., Germann, M. W., van de Sande, J. H., & Sykes, B. D. (1990) *J. Mol. Biol. 215*, 411−428.

Basu, A. K., & Marnett, L. J. (1983) *Carcinogenesis 4*, 331−333.

Basu, A. K., & Marnett, L. J. (1984) *Cancer Research 44*, 2848−2854.

Basu, A. K., O'Hara, S. M., Valladier, P., Stone, K., Mols, O., & Marnett, L. J. (1988) *Chem. Res. Toxicol. 1*, 53−59.

Bax, A., Sklenar, V., & Clore, G. M. (1987) *J. Am. Chem. Soc. 109*, 6511−6513.

Bell, D. A., Levine, J. G., & DeMarini, D. M. (1991) *Mutat. Res. 252*, 35−44.

Benamira, M., & Marnett, L. J. (1993) *Chem. Res. Toxicol. 6*, 317−327.

Benamira, M., Singh, U., & Marnett, L. J. (1992) *J. Biol. Chem. 267*, 22392−22400.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., & Haak, J. R. (1984) *J. Phys. Chem. 81*, 3684−3690.

Bodenhausen, G., Wagner, G., Rance, M., Sorensen, O. W., Wüthrich, K., & Ernst, R. R. (1984) *J. Magn. Reson. 59*, 542−550.

Boelens, R., Koning, T. M. G., Van Der Marel, G. A., Van Boom, J. H., & Kaptein, R. (1989) *J. Magn. Reson. 82*, 290−308.

Borer, P. N. (1975) in *Handbook of Biochemistry and Molecular Biology*, CRC Press, Cleveland, OH.

Borgias, B. A., & James, T. L. (1990) *J. Magn. Reson. 87*, 475−487.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187−217.

Brunger, A. T. (1992) in *X-PLOR. Version 3.1. A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.

Burnouf, D., Koehl, P., & Fuchs, R. P. P. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 4147−4151.

Cho, B. P., Beland, F. A., & Marques, M. M. (1992) *Biochemistry 31*, 9587−9602.

Cho, B. P., Beland, F. A., & Marques, M. M. (1994) *Biochemistry 33*, 1373−1384.

Clore, G. M., Gronenborn, A. M., Carlson, G., & Meyer, E. F. (1986) *J. Mol. Biol. 190*, 259−267.

Cosman, M., Fiala, R., Hingerty, B. E., Amin, S., Geacintov, N. E., Broyde, S., & Patel, D. J. (1994a) *Biochemistry 33*, 11507−11517.

Cosman, M., Fiala, R., Hingerty, B. E., Amin, S., Geacintov, N. E., Broyde, S., & Patel, D. J. (1994b) *Biochemistry 33*, 11518−11527.

Crawford, D. L., Sinnhuber, R. O., Stout, F. M., Oldfield, J. E., & Kaufmes, J. (1965) *Toxicol. Appl. Pharmacol. 7*, 826−832.

DeMarini, D. M., Abu-Shakra, A., Gupta, R., Hendee, L. J., & Levine, J. G. (1992) *Environ. Mol. Mutagen. 20*, 12−18.

Eckel, L. M., & Krugh, T. R. (1994) *Nature Struct. Biol. 1*, 89−94.

Feigon, J., Leupin, W., Denny, W. A., & Kearns, D. R. (1983) *Biochemistry 22*, 5943−5951.

Fuscoe, J. C., Wu, R., Shen, N. H., Healy, S. K., & Felton, J. S. (1988) *Mutat. Res. 201*, 241−251.

Hare, D. R., Wemmer, D. E., Chou, S. H., Drobny, G., & Reid, B. R. (1983) *J. Mol. Biol. 171*, 319−336.

Hartman, P. E., Ames, B. N., Roth, J. R., Barnes, W. M., & Levin, D. E. (1986) *Environ. Mutagen. 8*, 631−641.

Huang, P., & Eisenberg, M. (1992) *Biochemistry 31*, 6518−6532.

Huang, P., Patel, D. J., & Eisenberg, M. (1993) *Biochemistry 32*, 3852−3866.

Isono, K., & Yourno, J. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 1612−1617.

James, T. L. (1991) *Curr. Opin. Struct. Biol. 1*, 1042−1053.

Kalnik, M. W., Norman, D. G., Li, B. F., Swann, P. F., & Patel, D. J. (1990) *J. Biol. Chem. 265*, 636−647.

Kaluarachchi, K., Meadows, R. P., & Gorenstein, D. G. (1991) *Biochemistry 30*, 8785−8797.

Keepers, J. W., & James, T. L. (1984) *J. Magn. Reson. 57*, 404−426.

Kim, S. G., & Reid, B. R. (1992) *Biochemistry 31*, 12103−12116.

Koehl, P., Burnouf, D., & Fuchs, R. P. P. (1989) *J. Mol. Biol. 207*, 355−364.

Kouchakdjian, M., Marinelli, E., Gao, X., Johnson, F., Grollman, A., & Patel, D. (1989a) *Biochemistry 28*, 5647−5657.

Kouchakdjian, M., Marinelli, E., Gao, X., Johnson, F., Grollman, A., & Patel, D. (1989b) *Biochemistry 28*, 5647−5657.

Kouchakdjian, M., Eisenberg, M., Live, D., Marinelli, E., Grollman, A. P., & Patel, D. J. (1990) *Biochemistry 29*, 4456−4465.

Kouchakdjian, M., Eisenberg, M., Johnson, F., Grollman, A. P., & Patel, D. J. (1991) *Biochemistry 30*, 3262−3270.

Kunkel, T. A. (1990) *Biochemistry 29*, 8003−8011.

LeBlanc, D. A., & Morden, K. M. (1991) *Biochemistry 30*, 4042−4047.

Macura, S., Würthrich, K., & Ernst, R. R. (1982) *J. Magn. Reson. 47*, 351−357.

Marinelli, E. R., Johnson, F., Iden, C. R., & Yu, P. L. (1990) *Chem. Res. Toxicol. 3*, 49−58.

Marion, D., Ikura, M., & Bax, A. (1989) *J. Magn. Reson. 84*, 425−430.

Marnett, L. J., & Tuttle, M. A. (1980) *Cancer Res. 40*, 276−282.

Marnett, L. J., Basu, A. K., O'Hara, S. M., Weller, P. E., Rahman, A. F. M. M., & Oliver, J. P. (1986) *J. Am. Chem. Soc. 108*, 1348−1350.

McCann, J., Spingain, N. E., Ikobori, J., & Ames, B. N. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 979−983.

Moe, J. G., Reddy, G. R., Marnett, L. J., & Stone, M. P. (1994) *Chem. Res. Toxicol. 7*, 319−328.

Morden, K. M., & Maskos, K. (1993) *Biopolymers 33*, 27−36.

Morden, K. M., Chu, Y. G., Martin, F. H., & Tinoco, I., Jr. (1983) *Biochemistry 22*, 5557−5563.

Morden, K. M., Gunn, B. M., & Maskos, K. (1990) *Biochemistry 29*, 8835−8845.

Mujeeb, A., Kerwin, S. M., Kenyon, G. L., & James, T. L. (1993) *Biochemistry 32*, 13419−13431.

Mukai, F. H., & Goldstein, B. D. (1976) *Science 191*, 868−869.

Nikonowicz, E. P., Meadows, R. P., & Gorenstein, D. G. (1990) *Biochemistry 29*, 4193−4204.
Nilsson, L., & Karplus, M. (1986) *J. Comput. Chem. 7*, 591−616.
Nilsson, L., Clore, G. M., Gronenborn, A. M., Brunger, A. T., & Karplus, M. (1986) *J. Mol. Biol. 188*, 455−475.
O'Hara, S. M., & Marnett, L. J. (1991) *Mutat. Res. 247*, 45−56.
Oeschger, N. S., & Hartman, P. E. (1970) *J. Bacteriol. 101*, 490−504.
Patel, D. J., Kozlowski, S. A., Marky, L. A., Rice, J. A., Broka, C., Itakura, K., & Breslauer, K. J. (1982a) *Biochemistry 21*, 445−451.
Patel, D. J., Pardi, A., & Itakura, K. (1982b) *Science 216*, 581−590.
Plum, G. E., Grollman, A. P., Johnson, F., & Breslauer, K. J. (1992) *Biochemistry 31*, 12096−12102.
Rance, M., Bodenhausen, G., Wagner, G., Wüthrich, K., & Ernst, R. R. (1985) *J. Magn. Reson. 62*, 497−510.
Ravishankar, G., Swaminathan, S., Beveridge, D. L., Lavery, R., & Sklenar, H. (1989) *J. Biomol. Struct. Dyn. 6*, 669−699.
Ripley, L. S. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4128−4132.
Ripley, L. S. (1990) *Annu. Rev. Genet. 24*, 189−213.
Rodriguez, H., & Loechler, E. L. (1993a) *Biochemistry 32*, 1759−1769.
Rodriguez, H., & Loechler, E. L. (1993b) *Carcinogenesis 14*, 373−383.
Rosen, M. A., Live, D., & Patel, D. J. (1992a) *Biochemistry 31*, 4004−4014.
Rosen, M. A., Shapiro, L., Patel, D. J. (1992b) *Biochemistry 31*, 4015−4026.
Ryckaert, J.-P., Ciccotti, G., & Berendsen, H. J. C. (1977) *J. Comput. Phys. 23*, 327−341.
Schmitz, U., Pearlman, D. A., & James, T. L. (1991) *J. Mol. Biol. 221*, 271−292.
Schmitz, U., Sethson, I., Egan, W. M., & James, T. L. (1992) *J. Mol. Biol. 227*, 510−531.
Seto, H., Okuda, T., Takesue, T., & Ikemura, T. (1983) *Bull. Chem. Soc. Jpn. 56*, 1799−1802.
Seto, H., Seto, T., Takesue, T., & Ikemura, T. (1986) *Chem. Pharm. Bull. 34*, 5079−5085.
Shibutani, S., & Grollman, A. P. (1993) *J. Biol. Chem. 268*, 11703−11710.
Shibutani, S., Margulis, L. A., Geacintov, N. E., & Grollman, A. P. (1993) *Biochemistry 32*, 7531−7541.
Singh, U. S., Moe, J. G., Reddy, G. R., Weisenseel, J. P., Marnett, L. J., & Stone, M. P. (1993) *Chem. Res. Toxicol. 6*, 825−836.
Sklenar, V., Brooks, B. R., Zon, G., & Bax, A. (1987) *Febs Lett. 216*, 249−252.
Spalding, J. W. (1988) *NTP Technical Report 331*, 5−13.
Stewart, J. P. (1983) *Quant. Chem. Prog. Bull. 3*, 43.
Stolarski, R., Egan, W., & James, T. L. (1992) *Biochemistry 31*, 7027−7042.
Streisinger, G., Okada, Y., Enrich, J., Newton, J., Tsugita, A., Terzaghi, E., & Inouye, M. (1966) *Cold Spring Harbor Symp. Quant. Biol. 31*, 77−84.
van den Hoogen, Y. T., van Beuzekom, A. A., de Vroom, E., Van Der Marel, G. A., Van Boom, J. H., & Altona, C. (1988a) *Nucleic Acids Res. 16*, 5013−5030.
van den Hoogen, Y. T., van Beuzekom, A. A., Van Den Elst, H., Van Der Marel, Van Boom, J. H., & Altona, C. (1988b) *Nucleic Acids Res. 16*, 2971−2986.
Veaute, X., & Fuchs, R. P. (1991) *Nucleic Acids Res. 19*, 5603−5606.
Weiner, S. J., Kollman, P. A., Case, D. A., Singh, V. E., Ghio, C., Alagona, G., Profeta Jr., S., & Weiner, P. (1984) *J. Am. Chem. Soc. 106*, 765−784.
Zieba, K., Chu, T. M., Kupke, D. W., & Marky, L. A. (1991) *Biochemistry 30*, 8018−8026.

BI941572X