Solution Structure of an Oligodeoxynucleotide Duplex Containing the Exocyclic Lesion 3,N^4-Etheno-2′-deoxycytidine Opposite 2′-Deoxyadenosine, Determined by NMR Spectroscopy and Restrained Molecular Dynamics[†]

Alexandre Korobka,[‡] David Cullinan,[‡] Monique Cosman,[§] Arthur P. Grollman,[‡] Dinshaw J. Patel,[∥] Moisés Eisenberg,[‡] and Carlos de los Santos*,[‡]

Department of Pharmacological Sciences, State University of New York at Stony Brook, Stony Brook, New York 11794-8651, Biology & Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, California 94550, and Cellular Biochemistry and Biophysics, Memorial Sloan-Kettering Cancer Center, New York, New York 10021

Received March 7, 1996; Revised Manuscript Received June 24, 1996[⊗]

ABSTRACT: The d(C-G-T-A-C-ϵC-C-A-T-G-C)·d(G-C-A-T-G-A-G-T-A-C-G) oligodeoxynucleotide duplex containing the 3,N^4-etheno-2′-deoxycytidine adduct positioned opposite 2′-deoxyadenosine in the center of the helix has been analyzed by proton NMR spectroscopy and restrained molecular dynamics. The spectroscopic data establish a right-handed duplex, with sugar puckers in the C2′-endo/C3′-exo range, residues adopting an *anti* conformation around the glycosidic torsion angle and, with the exception of ϵC·dA, Watson−Crick hydrogen bond alignment for all base pairs. Molecular dynamics simulations, restrained by the full relaxation matrix approach, produced a three-dimensional model with an NMR *R*-factor of 7%. The duplex structure shows no significant perturbation of the sugar-phosphate backbone, which remains in B-form. The exocyclic adduct and its partner dA are incorporated into the helix without producing a noticeable kink. The ϵC·dA alignment adopts a staggered conformation with each residue displaced toward the 5′-terminus and intercalated between bases on the opposite strand, without increase of inter-phosphate distances. The partial intercalation of the ϵC(*anti*)·dA(*anti*) alignment allows stacking between the aromatic rings of ϵC and dA and with base pairs adjacent to the lesion, suggesting an important role played by hydrophobic forces in the stabilization of the solution structure.

Vinyl chloride and related compounds, such as vinyl carbamate and ethyl carbamate (urethane), are mutagenic and carcinogenic agents that cause hepatic carcinomas in rodents (Singer & Grumberger, 1983) and humans (Purchase *et al.*, 1987). The hepatic P_{450} enzyme system converts vinyl chloride to the unstable intermediate, chloroethylene oxide, which spontaneously rearranges to the more stable chloroacetaldehyde (Guengerich *et al.*, 1979). Both of these bifunctional alkylating agents can react with cellular DNA to produce 3,N^4-etheno-2′-deoxycytidine (ϵC),[1] along with several other exocyclic lesions (Barbin & Bartsch, 1986; Singer & Spengler 1986). These exocyclic adducts were recently detected as endogenous lesions in DNA isolated from humans and untreated rat livers (Nath *et al.*, 1994; Nair *et al.*, 1995). Probably, they are originated from the reaction of α,β-unsaturated aldehydes, formed endogenously during lipid peroxidation, with DNA (el Ghissassi *et al.*, 1995). A DNA glycosylase activity that removes exocyclic lesions has been identified in human cells (Dosanjh *et al.*, 1994).

The toxicity and mutagenic properties of ϵC adducts have been investigated using *in vitro* and *in vivo* systems. Studies with *Escherichia coli* DNA polymerase I established that a ϵC leads to ϵC→T transitions and, less frequently, to ϵC→A transversions (Simha *et al.*, 1991; Zhang *et al.*, 1995a). In agreement with these *in vitro* results, transformation of *E. coli* with adduct-containing phage DNA produces mostly ϵC→T transitions and ϵC→A transversions as well as deletion mutations (Palejwala *et al.*, 1991, 1993; Basu *et al.*, 1993). Recent *in vivo* studies, comparing the mutagenic properties of ϵC in bacterial and mammalian cells, showed that the mutagenic potency and spectrum of this adduct depend strongly on the type of host cell used for transfection. ϵC is a weak mutagen in bacterial cells, inducing infrequent ϵC→T transitions and ϵC→A transversions. However, the lesion is highly mutagenic in mammalian cells with the frequency of ϵC→A transversions almost twice that of ϵC→T transitions (Moriya *et al.*, 1994).

A significant body of work has been devoted to the study of structure−function relationships in DNA fragments bearing covalently modified bases (Friedberg *et al.*, 1995). NMR spectroscopy has been a powerful experimental tool for elucidating solution structures of short oligodeoxynucleotide duplexes containing exocyclic adducts. Studies have been performed on the 1,N^2-propano-2′-deoxyguanosine adduct positioned opposite dA and dG (Huang *et al.*, 1993; Kouchakjdian *et al.*, 1989, 1990) or positioned opposite a stable abasic site analog (Kouchakjdian *et al.*, 1991b), and

[†] This research was supported by Grant CA47995 (to A.P.G.) from the National Institutes of Health.

* To whom correspondence should be addressed. Phone: (516) 444-3649. FAX: (516) 444-3218. E-mail: carlos@pharm.sunysb.edu.

[‡] State University of New York at Stony Brook.

[§] Lawrence Livermore National Laboratory.

[∥] Memorial Sloan-Kettering Cancer Center.

[⊗] Abstract published in *Advance ACS Abstracts,* October 1, 1996.

[1] Abbreviations: NMR, nuclear magnetic resonance; EDTA, disodium ethylenediamine tetracetate; TSP, sodium (2,2,3,3-d_4)-3-(trimethylsilyl)propionate; NOESY, nuclear Overhauser effect spectroscopy; COSY, correlation spectroscopy; DQF-COSY, double quantum filtered correlation spectroscopy; HOHAHA, homonuclear Hartmann−Hahn spectroscopy; NOE, nuclear Overhauser effect; HPLC, high-performance liquid chromatography; ppm, parts per million; ϵC, 3,N^4-etheno-2′-deoxycytidine; dC, 2′-deoxycytidine; dA, 2′-deoxyadenosine; dG, 2′-deoxyguanosine; T, thymidine.

S0006-2960(96)00569-7 CCC: $12.00

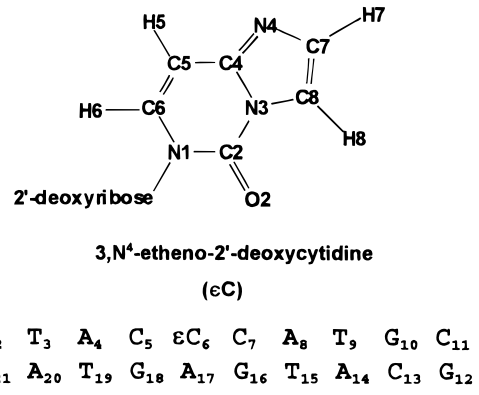

FIGURE 1: (A) Chemical structure of 3,N^4-etheno-2′-deoxycytidine (ϵC). (B) Sequence and numbering scheme of the ϵC·dA duplex.

on DNA duplexes containing a 1,N^6-etheno-2′- deoxyadenosine adduct opposite T and dG (Kouchakjdian *et al.*, 1991a; de los Santos *et al.*, 1991). These studies showed that the exocyclic adducts are incorporated readily into a B-form duplex and stacked into the helix. Local conformational changes are limited to the neighborhood of the lesion, with the rest of the helix remaining unperturbed.

We report here the structure of the d(C-G-T-A-C-ϵC-C-A-T-G-C)·d(G-C-A-T-G-A-G-T-A-C-G) oligodeoxynucleotide duplex (referred to throughout this paper as the ϵC·dA duplex), where the exocyclic 3,N^4-etheno-2′-deoxycytidine adduct is positioned opposite dA, as determined by high-resolution NMR and restrained molecular dynamics simulations. The chemical structure of the ϵC adduct and the numbering scheme of the duplex sequence employed in this study are shown in Figure 1.

MATERIALS AND METHODS

Synthesis and Purification of the Oligodeoxynucleotide Duplexes. The exocyclic 3,N^4-etheno-2′-deoxycytidine adduct was synthesized as a 5′-dimethoxytrityl-3′-β-cyanoethylphosphoroamidite derivative (Zhang *et al.*, 1995b) and incorporated into the oligodeoxynucleotide sequence by standard phosphoramidite chemistry procedures. 5′-Dimethoxytritylated sequences were isolated by treatment of the crude synthesis products with concentrated aqueous ammonia for 46 h at room temperature. Purification was accomplished by reverse phase HPLC on a preparative Dynamax (300 × 25 mm) C4 column. The mobile phase consisted of solvent A (0.1 M triethylamine acetic acid buffer, pH 6.8) and solvent B (acetonitrile). Using a linear gradient of 0% to 50% of B over 50 min, the main fraction, eluting at 33 min, was isolated. The terminal 5′-dimethoxytrityl group was cleaved by treatment with 80% acetic acid for 30 min, and this solution was extracted three times with ether before purification by HPLC. Desalting on a Sephadex G-25 column yielded pure oligodeoxynucleotide sequences that were converted to the sodium salt by passing them through a Dowex 50W cation exchange resin column.

Duplex Formation and Sample Preparation. A 1:1 stoichiometry of the duplex was established by monitoring the intensity of NMR proton signals during gradual addition of the unmodified strand to the ϵC-containing strand at 55 °C. NMR samples consisted of 300 OD_{260} of the duplex dissolved in 0.4 mL of 10 mM phosphate buffer, pH 6.9, containing 50 mM NaCl and 1 mM EDTA in either 99.96% D_2O or 90% H_2O−10% D_2O (v/v), corresponding to a concentration of approximately 7 mM. Samples were degassed under nitrogen before collection of the NMR data. pH values of D_2O buffer solutions are uncorrected pH-meter readings.

NMR Experiments. One- and two-dimensional NMR spectra were recorded on a Bruker AM spectrometer at 500 MHz field strength, while NOE buildup experiments were collected on a Bruker AMX 600 MHz spectrometer. Proton chemical shifts were referenced relative to TSP at 0.0 ppm. Phase-sensitive (States *et al.*, 1982) proton NOESY (150 ms mixing time) spectra in H_2O were recorded using a jump and return reading pulse (Plateu & Gueron, 1982) to reduce the solvent signal. Phase-sensitive NOESY spectra in D_2O (50, 90, 130, 170, and 210 ms mixing times) and phase-sensitive COSY and DQF-COSY spectra were collected with a repetition delay of 1.3 s, during which the residual water signal was suppressed by presaturation. Unless otherwise noted, time domain data sets consisted of 1024 by 256 complex data points in the t_2 and t_1 dimensions, respectively. NMR data were processed and analyzed using the program Felix (Biosym Technologies).

Molecular Dynamics Simulations. Restrained molecular dynamic simulations were performed on Silicon Graphics Indigo 2 computers using X-PLOR 3.1 (Brunger, 1993) with an all-atom force field. Structures were visualized with Midas Plus (UCSF, Computer Graphics Laboratory), and structural parameters were obtained with Curves (Lavery & Sklenar, 1988, 1989).

Molecular dynamics simulations were performed "*in vacuo*" using the all-atom force field derived from CHARMM (Brooks *et al.*, 1983), except that atomic partial charges remained unscreened, resulting in deoxynucleotide residues with a net charge of −1. The value of the dielectric constant was set to 4 (Friedman & Honing, 1992). Experimental distances were derived by measuring the initial slopes of NOE buildup curves (buildup rate) and using the cytosine H_6-H_5 distance (2.45 Å) as reference, according to $\sigma_{ij}D_{ij}^6 = \sigma_{H6-H5}D_{H6-H5}^6$, where σ_{ij} is the initial buildup rate between *i*,*j* protons and D_{ij} is their distance.

Distance restraints, defined by lower and upper bounds, were calculated by adding ±0.5 Å to the experimental distances and implemented by empirical square well energy potential functions. With the exception of ϵC6·A17, Watson−Crick hydrogen bond alignments on all base pairs of the duplex were implemented by distance restraints potentials. No restraints were used to enforce the pucker of the sugars. The backbone dihedral angles were restrained by an empirical square well potential function with a width encompassing a range for both A- and B-form DNA. Coplanarity of terminal base pairs was enforced.

Two different starting models were constructed from the canonical A- and B-form DNA. The ϵC adduct was built by adding the exocyclic ring to a deoxycytidine. Long interresidue distances were observed at the ϵC6-C7 and A17-G18 steps of the duplex; thus, steric clashes arising from the proximity of the bulky ϵC and the dA on the opposite strand were relieved by initially tilting each residue to the 5′-side on the respective strand followed by energy minimization using the conjugate gradient method. Distance-restrained molecular dynamics consisted of an equilibration step of 15 ps during which the temperature of the simulation

was increased from 115 to 350 K while the scale of the NOE distance restraint potential function was gradually augmented. This was followed by cooling to 300 K in 2.5 ps and 20 ps of restrained dynamics at constant temperature. Coordinates of the last picosecond of the molecular dynamics simulation were averaged and subjected to 2500 steps of energy minimization, yielding distance-refined structures. Distances of all covalent hydrogen bonds were kept constant during all simulations using the SHAKE algorithm (Ryckaert *et al.*, 1977).

Distance-refined structures were further optimized by 10 ps of molecular dynamics at 300 K, restrained by the full relaxation matrix method (Yip & Case, 1989; Nilges *et al.*, 1991). During the first 4.5 ps of the simulations, distance restraint potentials were reduced to zero, and the scale of a potential function based on the intensity difference between experimental and back-calculated NOE spectra was increased simultaneously. 954 experimental volumes measured from the NOESY spectra at 50, 90, 130, 170, and 210 ms mixing times were input as restraints with a 30% error range. A grid search determined that an isotropic correlation time of 4.25 ns produced the closest initial fit to the experimental data, and this value was used during the refinement. Coordinates of the last picosecond were averaged and energy minimized, yielding the final NMR-refined structures.

RESULTS

Assignment of the Nonexchangeable Protons. The one-dimensional proton spectrum of the ϵC·dA duplex in D_2O buffer, pH 6.8, 25 °C, is shown in Figure 1S of the Supporting Information. The spectrum is characterized by the presence of sharp lines and the absence of minor resonances. Assignment of the nonexchangeable base and H1′ sugar protons was accomplished by analysis of a NOESY (300 ms mixing time) spectrum, recorded in D_2O buffer, pH 6.8, 25 °C, following standard procedures (Hare *et al.*, 1983; Wüthrich, K., 1986; van de Ven & Hilbers, 1988). Figure 2 shows duplicate counterplots of an expanded region of the NOESY spectrum depicting distance connectivities between the base (purine H8/pyrimidine H6) (6.4 to 8.5 ppm) and H1′/H5 (5.0 to 6.6 ppm) proton regions. Characteristic of a right-handed DNA helix, each base proton exhibits NOE peaks to its own and 5′-flanking H1′ sugar protons, as shown from C1 to C11 in the adduct-modified strand (Figure 2A) and from G12 to G22 in the unmodified strand (Figure 2B). We note that the sequential ϵC6(H1′) to C7(H6) cross peak (Figure 2A, peak A) and, to a lesser extent, the A17(H1′) to G18(H8) connectivity (Figure 2B, peak F) are weak. This indicates a longer distance between these protons at the ϵC6-C7 and A17-G18 steps on the duplex. Sequential connectivities are detected between base (purine H8/pyrimidine H6) and the 3′-flanking deoxycytidine H5 protons on the G12-C13, A20-C21, A4-C5, C5-ϵC6, and G10-C11 but not for the ϵC6-C7 steps of the sequence (Figure 2B, peaks A and B, and Figure 2A, peaks B, C, and T, respectively).

Several characteristic NOE peaks allowed us to assign all deoxyadenosine H2 protons of the sequence. Typically this minor groove proton is within 4.6 Å both of its own and 3′-flanking H1′ protons on the same strand, and of the H1′ protons of its Watson−Crick partner and the 3′-flanking residue. We detect these connectivities for the A4, A8, A14, and A20 (peaks D−G and H−K in Figure 2A and peaks C−F and G−I in Figure 2B, respectively) residues of the duplex. Opposite the lesion, the A17(H2) proton could be identified by NOE peaks to its own H1′ proton (Figure 2B, peak J) and to the ϵC6(H1′) and C7(H1′) protons on the partner strand (Figure 2A, peaks L and M).

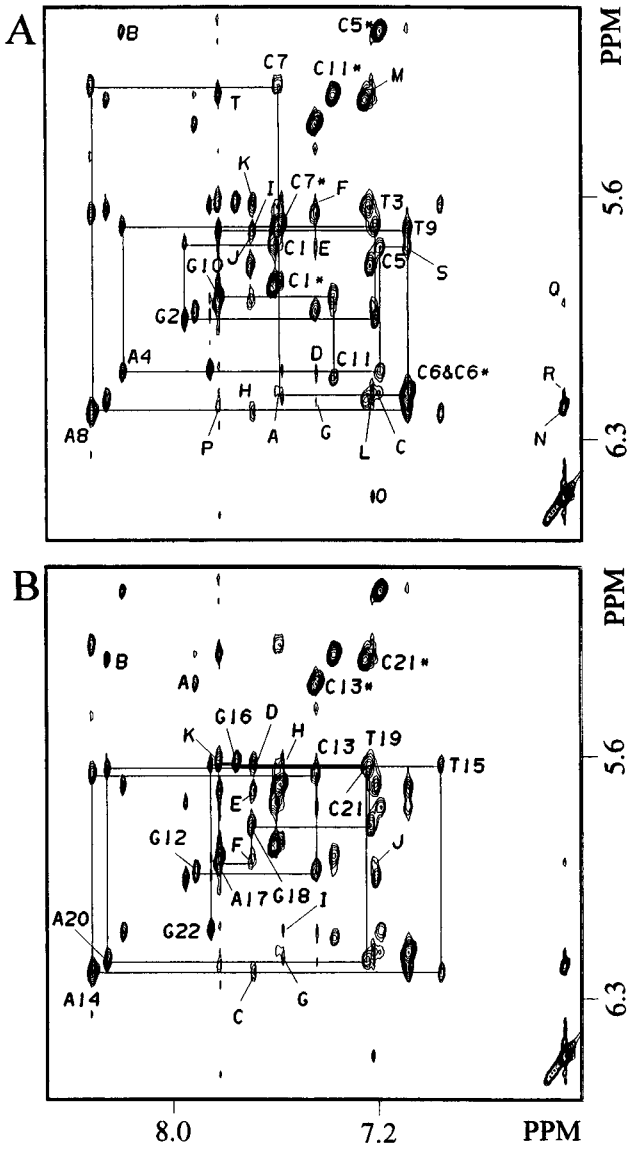

FIGURE 2: Duplicate counterplots of a portion of the NOESY (300 ms mixing time) spectrum recorded in D_2O buffer, pH 6.8, 25 °C, showing distance connectivities between base and H1′ sugar protons. Solid lines connect each base (purine H8/pyrimidine H6) to its own and 5′-flanking H1′ protons in the (A) modified and (B) unmodified strands. Labeled peaks are assigned as follows: (A) A, ϵC6(H1′)-C7(H6); B, A4(H8)-C5(H5); C, C5(H6)-ϵC6(H5); D, A4(H2)-A4(H1′); E, A4(H2)-C5(H1′); F, A4(H2)-T19(H1′); G, A4-(H2)-A20(H1′); H, A8(H2)-A8(H1′); I, A8(H2)-T9(H1′); J, A8-(H2)-T15(H1′); K, A8(H2)-G16(H1′); L, A17(H2)-ϵC6(H1′); M, A17(H2)-C7(H1′); N, ϵC6(H7)-ϵC6(H8); O, ϵC6(H8)-A17(H2); P, ϵC6(H7)-A17(H8); Q, ϵC6(H8)-A17(H1′); R, ϵC6(H8)-ϵC6(H1′); S, C5(H1′)-ϵC6(H6), and T, G10(H8)-C11(H5). (B) A, G12(H8)-C13(H5); B, A20(H8)-C21(H5); C, A14(H2)-A14(H1′); D, A14-(H2)-T15(H1′); E, A14(H2)-T9(H1′); F, A14(H2)-G10(H1′); G, A20(H2)-A20(H1′); H, A20(H2)-C21(H1′); I, A20(H2)-A4(H1′); J, A17(H2)-A17(H1′). Asterisks indicate the cytosine H5-H6 cross peak.

Assignment of the exocyclic H7 and H8 protons followed the observation of their strong NOE cross peak (Figure 2A, peak N) and was confirmed by the presence of a COSY cross peak in the phase-sensitive experiment (Figure 3, peak A).

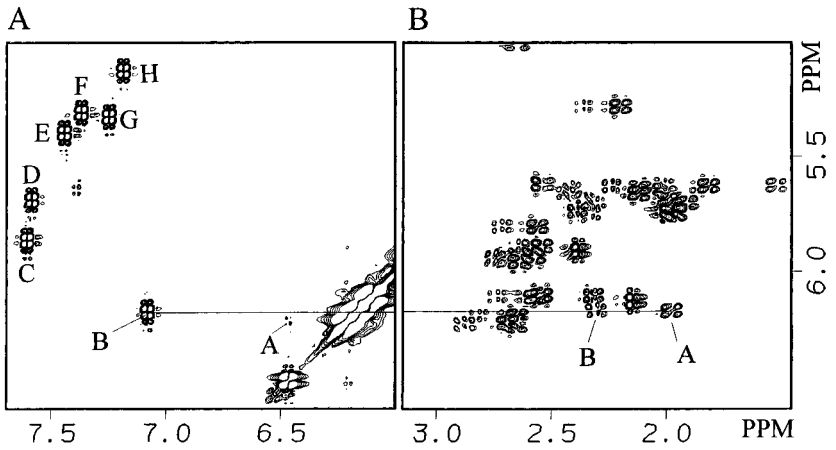

FIGURE 3: Counterplot of portions of a COSY spectrum recorded in D_2O buffer, pH 6.8, 25 °C, displaying J connectivities between (A) deoxycytidine H6-H5 and (B) sugar H1′-H2′,H2″ protons. Labeled peaks are assigned as follows: (A) A, ϵC6(H8)-ϵC6(H7); B, ϵC6(H6)-ϵC6(H5); C, C1(H6)-C1(H5); D, C7(H6)-C7(H5); E, C13(H6)-C13(H5); F, C11(H6)-C11(H5); G, C21(H6)-C21(H5), and H, C5(H6)-C5-(H5). (B) A, ϵC6(H1′)-ϵC6(H2′), and B, ϵC6(H1′)-ϵC6(H2″).

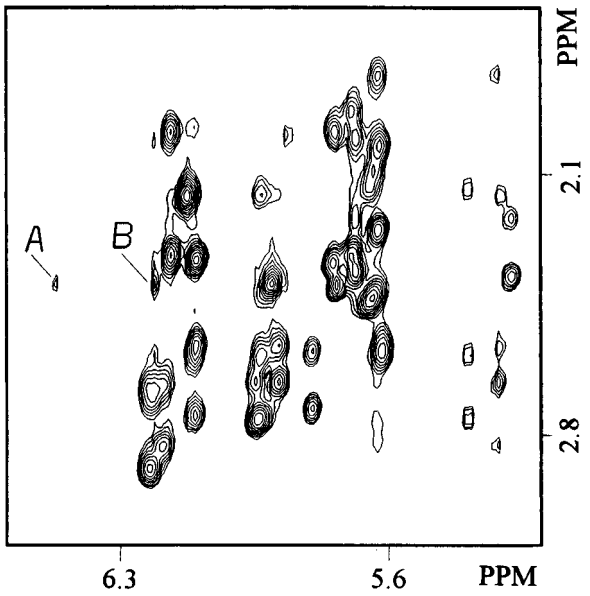

FIGURE 4: Counterplot of a portion of the NOESY (300 ms mixing time) spectrum in D_2O buffer, pH 6.8, 25 °C, showing distance connectivities between the sugar H1′ and H2′,H2″ proton regions. Each H1′ proton exhibits intra-residue NOE peaks to the H2′ and H2″ protons within the sugar ring. Labeled peaks are assigned as follows: A, ϵC6(H8)-A17(H2′,H2″); B, ϵC6(H7)-A17(H2′,H2″).

These protons showed very weak inter-strand NOE peaks to the H2, H8, and H1′ protons of A17, detected only in the 300 ms mixing time NOESY (Figure 2A, peaks O, P, Q), and moderate to the overlapping A17(H2′) and A17(H2″) sugar protons on the partner strand (peaks A and B in Figure 4). When the ϵC6 residue adopts an *anti* conformation around the glycosidic angle, its H8 proton is located within the helix while H7 is positioned toward the major groove. Thus, the ϵC6(H8) is closer and more likely to exhibit NOE peaks to the A17(H2) and A17(H1′) minor groove protons than ϵC6(H7) (Figure 2A, peaks O and Q, respectively). In agreement with this assignment, ϵC6(H7) shows a weak but clear connectivity to the major groove A17(H8) proton on the long mixing time NOESY spectrum (Figure 2A, peak P).

The COSY spectrum (Figure 3) illustrates the overlap of the ϵC6(H5) and ϵC6(H1′) proton resonances. A signal resonating at 6.17 ppm on the H1′-H5 region of the spectrum is correlated to the ϵC6(H6) and ϵC6(H2′,2″) sugar protons (Figure 3, peaks B, C, and D, respectively). Since the coupling constant between base and H1′ sugar protons is zero, only the overlap of the H5 and H1′ proton signals of the adduct can explain these observations.

Table 1: Nonexchangeable Proton Chemical Shifts[a] in D_2O[b] at 25 °C

	H8/H6	H5/CH_3/H2	H1′	H2′	H2″	H3′	H4′
C1	7.61	5.86	5.74	2.00	2.38	4.68	4.05
G2	7.95		5.96	2.66	2.76	4.95	4.34
T3	7.22	1.49	5.68	2.00	2.36	4.85	4.17
A4	8.20	7.44	6.11	2.60	2.75	4.98	4.37
C5	7.20	5.13	5.75	1.98	2.33	4.71	4.15
ϵC6[c]	7.09	6.17	6.17	1.98	2.32	4.73	4.18
C7	7.59	5.68	5.28	2.22	2.38	4.82	4.14
A8	8.34	7.69	6.22	2.69	2.90	5.01	4.36
T9	7.08	1.42	5.69	1.93	2.31	4.82	4.12
G10	7.82		5.89	2.56	2.66	4.94	4.34
C11	7.37	5.31	6.13	2.16	2.16	4.45	4.00
G12	7.92		5.94	2.59	2.75	4.82	4.22
C13	7.44	5.39	5.65	2.13	2.44	4.87	4.19
A14	8.32	7.70	6.23	2.67	2.88	5.02	4.37
T15	6.95	1.42	5.62	1.54	2.02	4.77	4.02
G16	7.75		5.61	2.55	2.58	4.93	4.28
A17	7.83	7.22	5.91	2.39	2.39	4.83	4.18
G18	7.70		5.80	2.58	2.73	4.87	4.34
T19	7.23	1.33	5.64	2.08	2.42	4.85	4.18
A20	8.26	7.58	6.19	2.67	2.83	5.01	4.36
C21	7.25	5.32	5.63	1.83	2.25	4.77	4.13
G22	7.86		6.11	2.56	2.33	4.63	4.14

[a] Values in parts per million relative to TSP at 0.0 ppm. [b] 10 mM buffer phosphate, 50 mM NaCl, pH 6.8. [c] ϵC6(H7) and ϵC6(H8) resonate at 6.21 and 6.49 ppm, respectively.

The assignment of the H2′, H2″, H3′, and H4′ sugar protons was based on the NOESY (50 and 300 ms mixing times), COSY, DQF-COSY, and HOHAHA spectra following standard procedures (Hare *et al.*, 1983; Wüthrich, K., 1986; van de Ven & Hilbers, 1988). Chemical shift values of the nonexchangeable protons detected in D_2O buffer, pH 6.8, 25 °C, are listed in Table 1.

Assignment of the Exchangeable Protons. The imino and base/amino regions (6.0−14.5 ppm) of the one-dimensional proton spectrum in H_2O buffer, pH 6.8, 2 °C, is shown in Figure 2S, Supporting Information. Eight partially resolved signals accounting for the ten imino protons present on the duplex were seen between 12.0 and 14.0 ppm, with no resonance observed between 9.0 and 12.0 ppm. Assignment

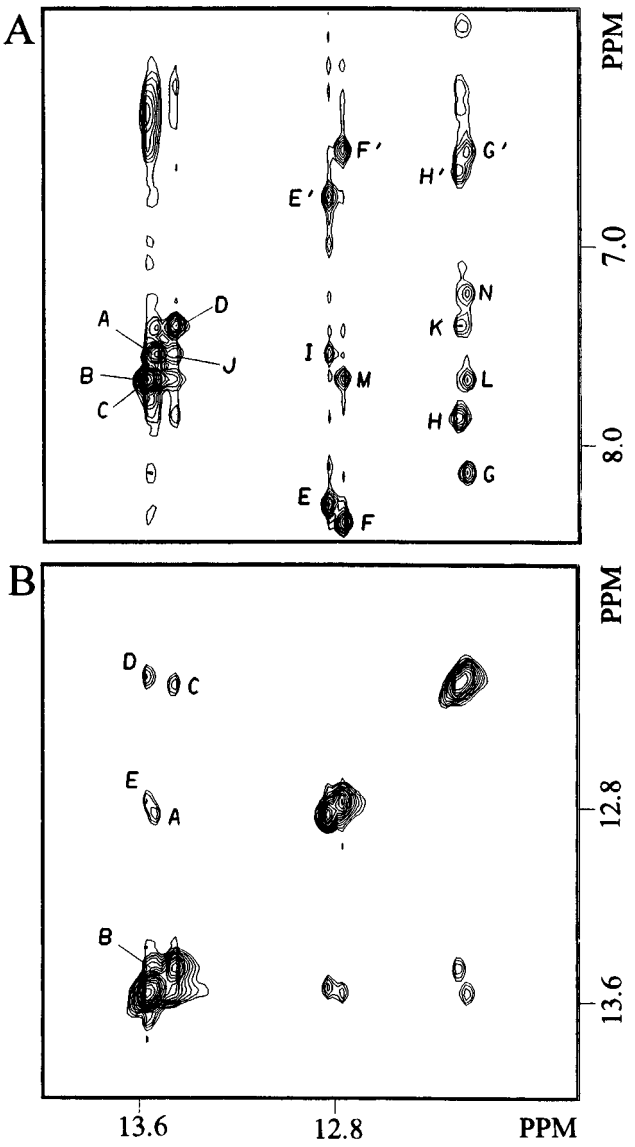

FIGURE 5: Counterplot portions of a NOESY (150 ms mixing time) spectrum recorded in H_2O buffer, pH 6.8, 2 °C, showing distance connectivities between (A) imino and base/amino proton regions and (B) on the symmetrical imino proton region. Labeled peaks are assigned as follows: (A) A, T3(N3H)-A20(H2); B, T9(N3H)-A14(H2); C, T15(N3H)-A8(H2); D, T19(N3H)-A4(H2); E, G2-(N1H)-C21(N4H)$_{hb}$; E′, G2(N1H)-C21(N4H)$_{nhb}$; F, G10(N1H)-C13(N4H)$_{hb}$; F′, G10(N1H)-C13(N4H)$_{nhb}$; G, G16(N1H)-C7(N4H)$_{hb}$; G′, G16(N1H)-C7(N4H)$_{nhb}$; H, G18(N1H)-C5(N4H)$_{hb}$; H′, G18-(N1H)-C5(N4H)$_{nhb}$; I, G2(N1H)-A20(H2); J, T19(N3H)-A20(H2); K, G18(N1H)-A4(H2); L, G16(N1H)-A8(H2); M, G10(N1H)-A14-(H2); N, G16(N1H)-A17(H2), and O, T3(N3H)-A4(H2). (B) A, G2(N1H)-T3(N3H); B, T3(N3H)-T19(N3H); C, T19(N3H)-G18-(N1H); D, G16(N1H)-T15(N3H), and E, T9(N3H)-G10(N1H). (N4H)$_{hb}$ and (N4H)$_{nhb}$ refer to the hydrogen-bonded and non-hydrogen-bonded deoxycytidine amino protons, respectively.

of the exchangeable protons followed analysis of the NOESY (150 ms mixing time) spectrum in H_2O buffer, pH 6.8, 2 °C. An expanded contour plot showing distance connectivities between the imino (11.8−14 ppm) and base/amino (6.0−8.5 ppm) proton regions is shown in Figure 5A. Strong NOE interactions are observed between the thymidine imino proton and the deoxyadenosine H2 protons across Watson−Crick base pairs. The independent assignment of the deoxyadenosine H2 protons from the NOESY spectra recorded in D_2O, permitted identification of this interaction for the T3·A20, T9·A14, T15·A8, and T19·A4 base pairs

Table 2: Exchangeable Proton Chemical Shifts[a] in H_2O[b] at 2 °C

	T(N3H)	G(N1H)	C(N4H)$_{hb}$[c]	C(N4H)$_{nhb}$[c]	A(H2)
C1·G22		nd[d]	8.11	7.01	
G2·C21		12.84	8.30	6.76	
T3·A20	13.54				7.56
A4·T19	13.46				7.40
C5·G18		12.27	7.87	6.62	
ϵC6·A17					7.25
C7·G16		12.31	8.15	6.52	
A8·T15	13.58				7.68
T9·A14	13.58				7.68
G10·C13		12.78	8.38	6.52	
C11·G12		12.97	8.17	6.56	

[a] Values in parts per million relative to TSP at 0.0 ppm. [b] 10 mM buffer phosphate, pH 6.8, 50 mM NaCl. [c] C(N4H)$_{hb}$ and C(N4H)$_{nhb}$ refer to the hydrogen-bonded and non-hydrogen-bonded cytidine amino protons, respectively. [d] nd, not detected.

of the sequence (Figure 5A, peaks A−D, respectively). Similarly, G·C base pairs were monitored by distance interactions observed between the deoxyguanosine imino proton and both the hydrogen-bonded and non-hydrogen-bonded deoxycytidine amino protons. The specific assignment of deoxycytidine amino protons through their intra-residue NOE peak to the H5 proton, previously identified in the NOESY spectrum recorded in D_2O, allowed the identification of the G2·C21, G10·C13, G16·C7, and G18·C5 base pairs of the sequence (Figure 5A, peaks E, E′; F, F′; G, G′; and H, H′, respectively). Sequential connectivities between deoxyadenosine H2 and the imino proton of flanking base pairs are detected at the G2·C21-T3·A20, T3·A20-T19·A4, T19·A4-G18·C5, G16·C7-T15·A8, and T9·A14-G10·C13 steps of the duplex (Figure 5A, peaks I−M and O, respectively). Further evidence of base pair stacking is observed in the symmetrical imino to imino (11.8−14.0 ppm) proton region of the NOESY (150 ms mixing time) experiment in H_2O (Figure 5B). Sequential connectivities are clearly observed between G2(N1H)-T3(N3H), T3(N3H)-T19-(N3H), T19(N3H)-G18(N1H), G16(N1H)-T15(N3H), and T9(N3H)-G10(N1H) imino protons (Figure 5B, peaks A−E, respectively). At the lesion site, A17(H2) exhibited an NOE peak to the imino proton of G16 on flanking G16·C7 base pair (Figure 5A, peak N) but not to the imino proton of G18 toward its 3′-side.

The exocyclic H7, H8 protons, as well as the imino proton of the terminal base pairs, did not exhibit cross peaks in the NOESY (150 ms mixing time) spectrum. Chemical shift values of the exchangeable protons recorded in H_2O buffer, pH 6.8, 2 °C, are listed in Table 2.

Molecular Dynamics Simulations. A view of the NMR-refined structures starting from A- and B-form DNA models is shown in Figure 3S, Supporting Information. The root mean square (rms) deviation in the position of the atoms was reduced from 5.9 Å between the starting models to 0.9 Å between the final structures, while the experimental restraints were fairly satisfied without major perturbation of the idealized covalent geometry. Independent of the starting structure, the refined models exhibited similar rms deviations from experimental distances for the 11-mer duplex as well as the central three-base pair fragment. Table 3 list averaged violations of the experimental restraints on the initial and refined models as well as the deviations from idealized geometry present on the refined structures. Since our assignment of the ϵC6(H7,H8) protons was based on

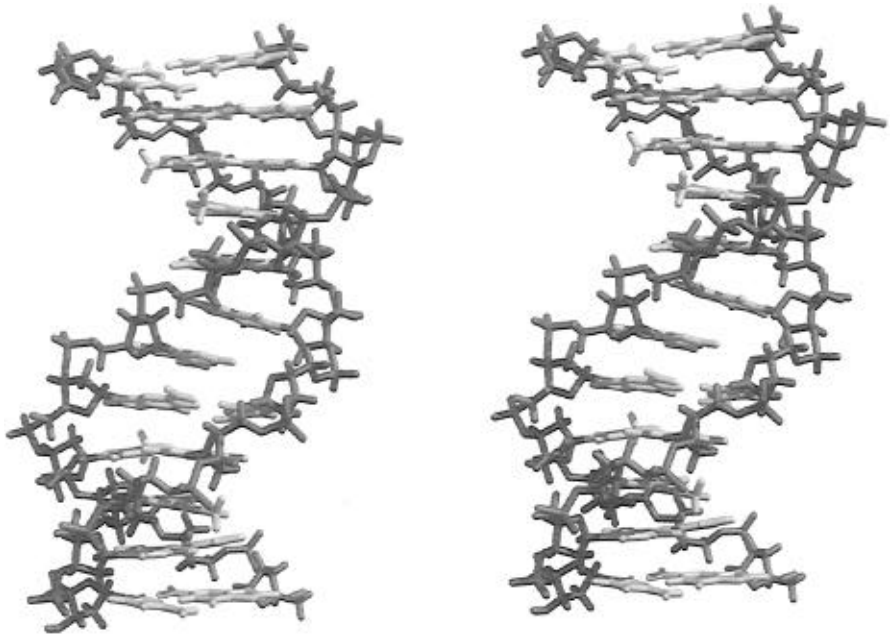

FIGURE 6: Stereopair of the three-dimensional structure, seen from the minor groove side, after 47.5 ps of NMR-restrained molecular dynamics, starting from the B-form initial model.

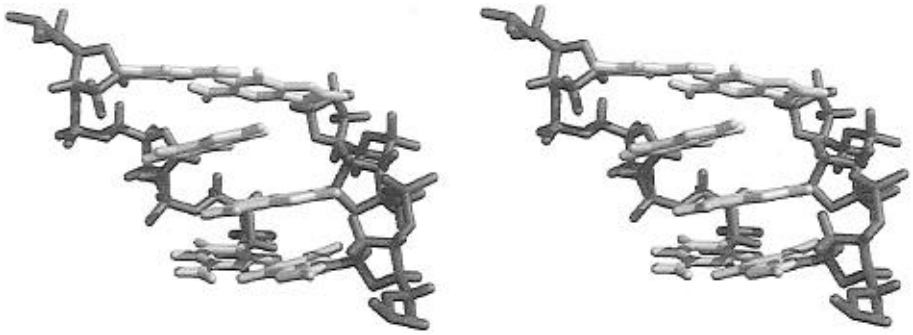

FIGURE 7: Stereoview of the central segment of the duplex structure seen from the major groove. The picture depicts increased distances between ϵC6 and C7 on the modified strand and between A17 and G18 on the complementary strand (left- and right-side strand, respectively), as well as the partial intercalation within the ϵC6(*anti*)·A17(*anti*) alignment with little perturbation of the duplex structure.

Table 3: Violations of Experimental Restraints and Idealized Covalent Geometry

RMS-Deviations from Experimental Inter-Proton Distances[a]	initial models		refined structures[b]	
	A-form	B-form	A-initial	B-initial
11-mer duplex[c]	0.19	0.11	0.07	0.07
central three-base pairs[d]	0.22	0.14	0.09	0.09
	R-factors[e]			
11-mer duplex (954 NOE-intensities)	0.113	0.097	0.071	0.071

RMS-Deviations from Idealized Covalent Geometry	refined structures[b]	
	A-initial	B-initial
bonds (Å)	0.006	0.006
bond angles (deg)	3.58	3.47
improper angles (deg)	0.18	0.18

[a] Values are given in angstroms. [b] A-initial and B-initial indicate that the refinement started from A- and B-form DNA initial structures, respectively. [c] Total of 226 distances. [d] Total of 56 distances. [e] 1/6 *R*-factor as defined by X-PLOR.

structural considerations, we interchanged the distance restraints between these protons and refined the initial models by distance-restrained molecular dynamics. As a result, the rms deviations between experimental NOE bounds and inter-proton distances measured on the central three-base pair segment of the duplex increased by 50% and four inter-proton distances fell outside the experimental bounds, as oppose to no violations observed with the original assignment.

Figure 6 shows the three-dimensional structure of the ϵC·dA duplex after 47.5 ps of NMR-restrained molecular dynamics, starting from the B-form model. The structure belongs to the B-form DNA family and has all bases stacked into the helix without major perturbation of the sugar-phosphate backbone. Figure 7 shows a stereoview of the center of the helix illustrating how the lesion is accommodated within the helix, and Table 4 lists structural parameters at the central three-base pair segment of the duplex.

The ϵC6·A17 base pair is highly staggered with a 4.80 Å separation between the ϵC and dA planes and both residues displaced inside the helix, resulting in a negative base pair stretch of 2.75 Å. Distances between bases adjacent to the lesion site also increase significantly: A17(C1′)-G18(C1′) and ϵC6(C1′)-C7(C1′) distances are 6.27 and 6.03 Å, respectively, approximately 1.1 Å longer than the averaged value measured for other C1′-C1′ distances. Similarly, distances between A17(N9)-G18(N9) and ϵC6(N1)-C7(N1)

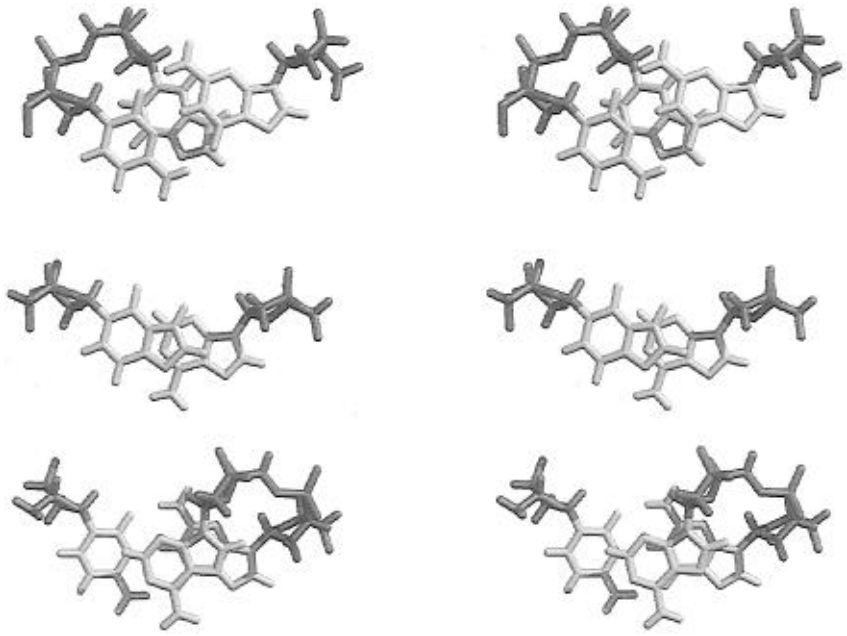

FIGURE 8: Stereoviews from the top of the helix showing stacking interactions on the central segment of the duplex. The picture displays interactions between (top) the C5·G18 base pair and ϵC adduct, (bottom) the C7·G16 base pair and A17 residue, and (center) within the ϵC6·A17 alignment.

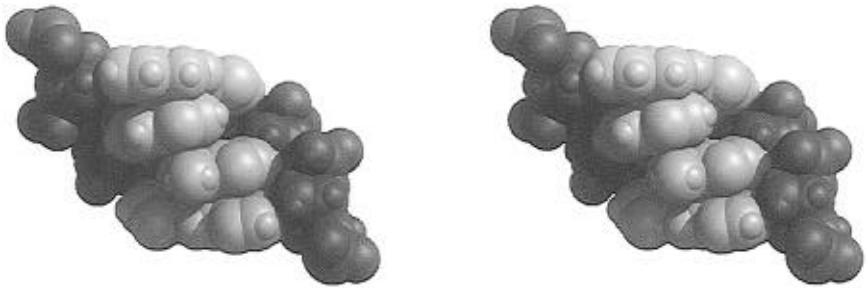

FIGURE 9: Space-filling model of the central three-base pair segment of the duplex as seen from the major groove.

Table 4: Structural Parameters[a,b]

ϵC6·A17 Alignment (Å)	
ϵC6 *Y*-displacement	−1.92
A17 *Y*-displacement	−0.84
base pair stagger	4.80
base pair stretch	−2.75
Base Rise (Å)	
C5-ϵC6	3.17
ϵC6-C7	6.91
G16-A17	3.05
A17-G18	6.89
Base Pair Buckle (deg)	
C5·G18	16
C7·G16	10

[a] Averaged values after refinement starting from A- and B-form DNA. [b] Parameters as defined by Curves. Canonical values on B-form DNA are *Y*-displacement, 0.0 Å; base pair stagger, 0.0 Å; base rise, 3.38 Å; base pair buckle, 0.0° (Lavery & Sklenar, 1989).

are 6.29 and 5.90 Å, respectively, while the averaged value observed for the rest of the sequence is 4.49 Å. However, distances between adjacent phosphorus atoms at these steps are about 7 Å, similar to the values observed throughout the duplex.

The planes of ϵC6 and A17 aromatic rings are almost perpendicular to the helical axis, with inclinations of 7.5° and 1.5°, respectively. ϵC intercalates partially between A17 and G18, and its partner A17 intercalates between ϵC6 and C7. As a result, ϵC exhibits limited stacking with the flanking C5·G18 base pair, while A17 stacks properly with G16, its 5′-flanking residue (Figure 8, top and bottom stereopairs, respectively). Stacking is also observed within the partially intercalating ϵC6·A17 alignment (Figure 8, central stereopair). The space-filling model of the structure reveals that this region has virtually no space available to accommodate water molecules (Figure 9). Base pairs flanking the lesion site adopt Watson−Crick alignment although some buckling is observed, especially for C5·G18 (Figure 7). No hydrogen bonds involving the aromatic moieties of ϵC and A17 are present in the refined structure.

Except for the terminal C11 residue whose sugar pucker is C3′-endo, the refined model has sugar rings on the south conformation, A17 being in the C1′-exo range and the rest in C2′-endo. Inter-proton distances measured on the refined structures are within the experimental bounds determined by the NOESY spectra. Table 5 compares experimental and refined distances on the central segment of the ϵC·dA duplex.

DISCUSSION

NMR Spectra. The sharp signals observed for the non-exchangeable base and sugar protons in D_2O buffer, pH 6.8, 25 °C (Figure 1S, Supporting Information), and the exchangeable imino and amino protons in H_2O buffer, pH 6.8, 2 °C (Figure 2S, Supporting Information), during the NMR

Table 5: Comparison between Experimental and Refined Inter-Proton Distances (Å)

	experimental distance bounds[a]	refined distances[b]
C5(H1′)-ϵC6(H6)	3.8−4.8	3.9
ϵC6(H1′)-C7(H6)	>4.5[c]	4.6
G16(H1′)-A17(H8)	3.0−4.0	3.4
A17(H1′)-G18(H8)	>4.5[c]	5.2
A17(H2)-C7(H1′)	>4.5[c]	4.5
ϵC6(H1′)-A17(H2)	4.0−5.2	4.0
ϵC6(H8)-A17(H2)	3.8−5.2	3.8
ϵC6(H8)-A17(H1′)	4.1−5.1	4.5
ϵC6(H7)-A17(H8)	>4.5[c]	4.5
ϵC6(H7)-A17(H2′)	3.5−4.8	3.9
ϵC6(H7)-A17(H2″)	3.5−4.8	4.1
ϵC6(H8)-A17(H2′)	3.7−5.0	4.6
ϵC6(H8)-A17(H2″)	3.7−5.0	3.7

[a] Derived from NOE buildup curves. [b] Averaged values measured on the refined models. [c] Distance interactions detected only on the 300 ms mixing time NOESY spectrum.

characterization, together with the observation of only one set of proton resonances, support the existence of a single solution structure for the ϵC·dA duplex in the NMR time scale.

The sequential connectivities detected in the NOESY (300 ms mixing time) spectrum recorded in D_2O between base (purine H8/pyrimidine H6) and H1′, H2′, H2″, H3′, deoxycytidine H5, and thymine methyl protons, and between deoxyadenosine H2 and H1′ sugar protons, indicate a right handed helical structure. Moreover, the intensity of the intra-residue base to H1′ NOE peaks observed in the long (Figure 2A and B) and short mixing time NOESY spectra establish that all residues are oriented in the *anti* conformation around the glycosidic torsion angle (Hare *et al.*, 1983; Wüthrich, K., 1986; van de Ven & Hilbers, 1988). Similarly, the absence of strong base to H3′ intra-residue connectivities in the NOESY (50 ms mixing time) spectrum (data not shown), together with the pattern of the H1′-H2′ and H1′-H2″ COSY cross peaks observed on the phase-sensitive spectrum (Figure 3), establish a south conformation for the pucker of all sugars (van de Ven & Hilbers, 1988; Majumdar & Hosur, 1992).

Distance connectivities between thymidine imino and deoxyadenosine H2 protons, and deoxyguanosine imino proton and hydrogen-bonded and non-hydrogen-bonded deoxycytidine amino protons detected in the NOESY (150 ms mixing time) spectrum in H_2O buffer, 2 °C (Figure 5A), demonstrate Watson−Crick alignment for the base pairs at either side of the lesion. Imino to imino and imino to deoxyadenosine H2 cross peaks observed between adjacent base pairs of the sequence (Figure 4A and B) indicate favorable base pair stacking throughout the duplex (Wüthrich, K., 1986).

Alignment of the ϵC·dA Base Pair. Several lines of spectroscopic evidence define structural features of the duplex at the lesion site. When both residues adopt the *anti* conformation around the glycosidic angle, alignment of the ϵC6·A17 base pair cannot be coplanar due to steric clashes between the aromatic moieties. On the modified strand, the intensity of the inter-residue ϵC6(H6) to C5(H1′) cross peak observed in all NOESY spectra implies a normal distance between these two protons at the C5-ϵC6 step (Figure 2A, peak S). At the same time, the very weak C7(H6) to ϵC6-(H1′) NOE detected only in the 300 ms mixing time NOESY spectrum recorded in D_2O (Figure 2A, peak A) implies a long distance between these protons at the ϵC6-C7 step of the duplex. We observe a similar pattern in the complementary strand, where the intensity of sequential base to sugar H1′ cross peaks on the long mixing time NOESY spectrum establishes long and normal distances between these protons at the A17-G18 and G16- A17 steps of the duplex, respectively (Figure 2B, peaks F and K). These data are consistent with a staggered ϵC6·A17 alignment, being the aromatic moiety of ϵC intercalated between A17 and G18, and the purine ring of A17 located between the ϵC and C7 residues. This orientation is further supported by inter-strand connectivities detected between the exocyclic H7, H8 and A17(H2′,H2″) sugar protons (Figure 4, peaks A and B), and between the A17(H2) and ϵC6(H1′), ϵC6(H8) and C7(H1′) protons (Figure 2A, peaks L, O, and M, respectively). A similar alignment has been reported for the 1,N^6-etheno-2′-deoxyadenosine adduct opposite T (Kouchakdjian *et al.*, 1991a) and the 1,N^2-propano-2′-deoxyguanosine adduct opposite dA (Huang *et al.*, 1993).

NMR Refined Solution Structure. As a result of the restrained molecular dynamics refinement, the initial A- and B-form DNA models converged to virtually identical B-form structures, having an rms deviation in the position of the atom of 0.9 Å. This value is not an indication of the precision of the structure determination and only suggests that the NMR-derived constraints, along with other terms of the force field, defined an averaged structural model in fair agreement with the experimental data.

In the presence of the exocyclic adduct, there is no significant perturbation of the sugar-phosphate backbone and the lesion is accommodated without any noticeable kink or bend of the helical axis (Figures 5 and 6). Duplexes are right-handed with all bases, including ϵC6 and A17, stacked into the helix. The base pair stagger measured on the ϵC6(*anti*)·A17(*anti*) alignment is 4.8 Å, and the rise at the ϵC6-C7 and A17-G18 steps is 6.9 Å. As a result, there is sufficient space for the incorporation of ϵC and its partner A17 into the helix. The negative stretch of 2.75 Å on the ϵC6(*anti*)·A17(*anti*) alignment indicates displacement of the bases toward their opposite strands, resulting in a partially intercalated structure. This arrangement, and the absence of inclination of the bases, allows stacking between ϵC6 and C5 on the modified strand, A17 and G16 on the complementary strand and within the ϵC6(*anti*)·A17(*anti*) alignment (Figure 8). These interactions observed on the three-dimensional model suggest that hydrophobic forces play an important role in the stability of the ϵC·dA duplex structure in solution.

Experimental evidence supporting the model include the NOE peaks detected between exocyclic ϵC6(H7,H8) and A17(H2′,H2″) protons (Figure 4, peaks A and B), and the weak cross peaks between A17(H2) and the ϵC6(H1′) and C7(H1′) protons observed in the 300 ms mixing time NOESY spectrum (Figure 2A, peaks L and M, respectively). On the refined structures, these distances are within NOE-determined bounds (Table 5). Further evidence of partial intercalation within the ϵC6(*anti*)·A17(*anti*) alignment is found in the chemical shifts of the exocyclic ϵC protons. When the adduct is positioned opposite T, in otherwise identical sequence context, the glycosidic torsion angle of the adduct is *syn.* As a result, the exocyclic H7 and H8 protons are placed in the major groove of the helix, where shielding by neighboring purine rings is minor [Cullinan *et al.*, 1996

(companion paper)]. The 0.59 and 0.41 ppm upfield shifts observed for the ϵC(H7) and ϵC(H8) resonances in the ϵC·dA duplex reflect increased shielding of these protons and support our interpretation of a partially intercalating model.

Formation of Watson−Crick hydrogen bonds on flanking C5·G18 and C7·G16 base pairs increase the stability of the duplex at the lesion site, but the aromatic rings of ϵC and A17 are not involved in hydrogen bonding. The refined model shows insufficient space for water molecules to surround the ϵC6(*anti*)·A17(*anti*) alignment. The presence of water-mediated hydrogen bonds stabilizing the duplex structure is under investigation.

ACKNOWLEDGMENT

We thank Mr. Robert Rieger and Ms. Cecilia Torres for preparing modified oligonucleotides and Mr. Erich Bremer for assisting in the preparation of three-dimensional figures. Initial NMR data at 500 MHz were collected at the College of Physicians and Surgeons of Columbia University, while 600 MHz experiments were recorded in the NMR Facility of the State University of New York at Stony Brook.

SUPPORTING INFORMATION AVAILABLE

Figures showing one-dimensional proton spectra of the ϵC·dA 11-mer duplex recorded in D_2O buffer, pH 6.8, 25 °C, and in H_2O buffer, pH 6.8, 2 °C (Figures 1S and 2S, respectively); three-dimensional structures of the distance-refined models starting from A- and B-form DNA (Figure 3S) (4 pages). Ordering information is given on any current masthead page.

REFERENCES

Barbin, A., & Bartsch, H. (1986) in *The Role of Cyclic Nucleic Acid Adducts in Carcinogenesis and Mutagenesis*, IARC Scientific Publications 70, pp 345−358, International Agency for Research on Cancer, Lyon, France.

Basu, A. K., Wood, M. L., Niedernhofer, L. J., Ramos, L. A., & Essigmann, J. M. (1993) *Biochemistry 32*, 12793−12801.

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

Brunger, A. (1993) X-PLOR Version 3.1, A System for X-Ray Crystallography and NMR, Yale University Press, New Haven, CT.

de los Santos, C., Kouchakdjian, M., Yerema K., Basu, A., Essigmann, J., & Patel, D. J. (1991) *Biochemistry 30*, 1828−1835.

Dosanjh, M. K., Chenna, A., Kim, E., Fraenkel-Conrat, H., Samson, L., & Singer, B. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 1024−1028.

el Ghissassi, F., Barbin, A., Nair, J., & Bartsch, H. (1995) *Chem. Res. Toxicol. 8*, 278−283.

Friedberg, E. C., Walker, G. C., & Siede, W. (1995) *DNA Repair and Mutagenesis*, ASM Press, Washington, DC.

Friedman, R. A., & Honing, B. (1992) *Biopolymers 32*, 145−159.

Guengerich, F. P., Crawford, W. M., & Hathaway, D. E. (1979) *Biochemistry 18*, 5177−5182.

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

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

Kouchakdjian, M., Marinelli, E., Gao, X., Johnson, F., Grollman, A. P., & Patel, D. J. (1989) *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., Yerema K., Basu, A., Essigmann, J., & Patel, D. J. (1991a) *Biochemistry 30*, 1820−1828.

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

Lavery, R., & Sklenar, H. (1988) *J. Biomol. Struct. Dyn. 6*, 63−91.

Lavery, R., & Sklenar, H. (1989) *J. Biomol. Struct. Dyn. 6*, 655−667.

Majumdar, A., & Hosur, R. V. (1992) *Prog. NMR Spectrosc. 24*, 109−158.

Moriya, M., Zhang, W., Johnson, F., & Grollman, A. P. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 11899−11903.

Nair, J., Barbin, A., Guichard, Y., & Bartsch, H. (1995) *Carcinogenesis 16*, 613−617.

Nath, R. G., & Chung, F.-L. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 7491−7495.

Nilges, M., Habazettl, J., Brunger, A. T., & Holak, T. A. (1991) *J. Mol. Biol. 219*, 499−510.

Palejwala, V. A., Simha, D., & Humayun, M. Z. (1991) *Biochemistry 30*, 8736−8743.

Palejwala, V. A., Rzepka, R. W., & Humayun, M. Z. (1993) *Biochemistry 32*, 4112−4120.

Plateau, P., & Gueron, M. (1982) *J. Am. Chem. Soc. 104*, 7310−7311.

Purchase, I. F. H., Stafford, J., & Paddle, G. M. (1987) *Food Chem. Toxicol. 4*, 168−202.

Ryckaert, J. P., Ciccoti, G., & Berendsen, H. J. C. (1977) *J. Comput. Phys. 23*, 327−341.

Simha, D., Palejwala, V. A., & Humayun, M. Z. (1991) *Biochemistry 30*, 8727−8735.

Singer, B., & Grumberger, D. (1983) in *Molecular Biology of Mutagens & Carcinogens*, pp 87−94, Plenum Press, New York.

Singer, B., & Spengler, S. J. (1986) in *The Role of Cyclic Nucleic Acid Adducts in Carcinogenesis and Mutagenesis*, IARC Scientific Publications 70, pp 345−358, International Agency for Research on Cancer, Lyon, France.

States, D. J., Habekorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286−292.

van de Ven, F. J. M., & Hilbers, C. W. (1988) *Eur. J. Biochem. 178*, 1−38.

Wüthrich, K. (1986) in *NMR of Proteins and Nucleic Acids*, John Wiley, New York.

Yip, P., & Case, D. (1989) *J. Magn. Reson. 83*, 643−648.

Zhang, W., Johnson, F., Grollman, A. P., & Shibutani, S. (1995a) *Chem. Res. Toxicol. 8*, 157−163.

Zhang, W., Rieger, R., Iden, C., & Johnson, F. (1995b) *Chem. Res. Toxicol. 8*, 148−156.

BI9605696