Solution Structure of a DNA Duplex Containing the Exocyclic Lesion 3,N^4-Etheno-2′-deoxycytidine Opposite 2′-Deoxyguanosine†

David Cullinan,‡ Francis Johnson,§ Arthur P. Grollman,§ Moisés Eisenberg,§ and Carlos de los Santos*,§

Departments of Pharmacological Sciences and Physiology and Biophysics, State University of New York at Stony Brook, Stony Brook, New York 11794-8651

Received March 12, 1997; Revised Manuscript Received July 22, 1997⊗

ABSTRACT: Vinyl chloride reacts with cellular DNA producing 3,N^4-etheno-2′-deoxycytidine (ϵC) along with other exocyclic adducts. The solution structure of an oligodeoxynucleotide duplex containing an ϵC·dG base pair was determined by high-resolution NMR spectroscopy and molecular dynamics simulations. NMR data indicated that the duplex adopts a right-handed helical structure having all residues in *anti* orientation around the glycosylic torsion angle. The ϵC adduct has a sugar pucker in the C3′-endo/C4′-exo region while the rest of the residues are in the C2′-endo/C3′-exo range. NOE interactions established Watson−Crick alignments for canonical base pairs of the duplex. The imino proton of the lesion-containing base pair resonated as a sharp signal that was resistant to water exchange, suggesting hydrogen bonding. Restrained molecular dynamics simulations generated three-dimensional models in excellent agreement with the spectroscopic data. The refined structures are slightly bent at the lesion site without major perturbations of the sugar-phosphate backbone. The adduct is displaced and shifted toward the major groove of the helix while its partner on the complementary strand remains stacked. The ϵC(*anti*)·dG(*anti*) base pair alignment is sheared and stabilized by the formation of hydrogen bonds. The biological implications of structures of ϵC-containing DNA duplexes are discussed.

Vinyl chloride, used in the production of synthetic polymers, vinyl carbamate, and ethyl carbamate, a component of certain foods and alcoholic beverages, induce hepatic carcinomas in rodents (Singer & Grumberger, 1983) and humans (Purchase *et al.*, 1987). These compounds are oxidized by the hepatic P_{450} enzyme system to the reactive metabolites, chloroacetaldehyde, and epoxyethyl carbamate (Guengerich *et al.*, 1979; Barbin *et al.*, 1985; Scherer *et al.*, 1968), and, as bifunctional alkylating agents, react with cellular DNA to produce 3,N^4-etheno-2′-deoxycytidine (ϵC),[1] 1,N^2-etheno-2′-deoxyguanosine (1,N^2-ϵG), 3,N^2-etheno-2′-deoxyguanosine (3,N^2-ϵG), and 1,N^6-etheno-2′-deoxyadenosine (ϵA) (Barbin & Bartsch, 1986; Kusmierek & Singer, 1992). These exocyclic DNA adducts have been detected as endogenous lesions in human and rat liver DNA (Nath *et al.*, 1994; Nair *et al.*, 1995), contributing to the muatgenic burden of mammalian cells and, thereby, potentially to carcinogenesis. DNA repair activities that remove these lesions from DNA have been identified in human cells (Dosanjh *et al.*, 1994; Saparbaev *et al.*, 1995; Hang *et al.*, 1996).

The mutagenic properties of the ϵC adduct have been investigated in bacteria using *in vitro* and *in vivo* systems. Replication of a DNA template containing ϵC by the Klenow fragment of DNA polymerase I leads to ϵC→T transitions and ϵC→A transversions (Singer & Spengler, 1986; Simha *et al.*, 1991; Zhang *et al.*, 1995a). Replication past the adduct by mammalian DNA polymerases α and δ generates ϵC→T transitions and ϵC→A transversions while polymerase β leads to ϵC→G transversions (Shibutani *et al.*, 1996). Transformation of *Escherichia coli* with vectors containing a single ϵC adduct generates mostly ϵC→T transitions and ϵC→A transversions along with some deletion mutations. In bacteria, mutation frequency is low, indicating that the lesion is a weak mutagen (Palejwala *et al.*, 1991, 1993; Basu *et al.*, 1993; Moriya *et al.*, 1994). In contrast, ϵC is highly mutagenic in mammalian cells; ϵC→A transversions and ϵC→T transitions predominate (Moriya *et al.*, 1994). Thus, it appears that during DNA replication ϵC pairs productively with all four deoxynucleotides, depending on the polymerase involved.

The solution structures of oligodeoxynucleotide duplexes containing exocyclic adducts have been investigated by NMR spectroscopy. 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; Kouchakjdian *et al.*, 1990) or opposite a stable abasic site analog (Kouchakjdian *et al.*, 1991b), and on the ϵA lesion opposite T and dG (Kouchakjdian *et al.*, 1991a; de los Santos *et al.*, 1991). The adducts are readily incorporated into B-form DNA duplexes and stacked into the helix. Confor-

† This research was supported by Grant CA47995 from the National Institutes of Health. The NMR facility at SUNY Stony Brook is supported by Grants CHE8911350 and CHE9413510 form the NSF and 1S10RR554701 from the NIH.

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

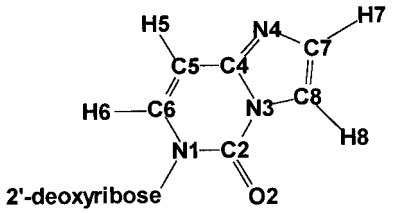
‡ Department of Physiology and Biophysics.

§ Department of Pharmacological Sciences.

⊗ Abstract published in *Advance ACS Abstracts,* September 15, 1997.

[1] Abbreviations: NMR, nuclear magnetic resonance; EDTA, disodium ethylenediamine tetracetate; TSP, (2,2,3,3-d_4) sodium 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; HMQC, heteronuclear multiple quantum correlation spectroscopy; 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; MD, molecular dynamics; RMSD, root mean square deviations.

S0006-2960(97)00572-2 CCC: $14.00



3,N4-etheno-2'-deoxycytosine

C1 G2 T3 A4 C5 ϵC6 C7 A8 T9 G10 C11
G22 C21 A20 T19 G18 G17 G16 T15 A14 C13 G12

ϵC•dG duplex

FIGURE 1: Chemical structure of 3,N^4-etheno-2′-deoxycytidine; ϵC•dG duplex sequence and numbering scheme.

mational changes are limited to the residues neighboring the lesion.

Recently, we have reported the solution structure of 11-mer oligodeoxynucleotide duplexes containing an ϵC positioned opposite dA and T (Korobka *et al*., 1996; Cullinan *et al*., 1996). In both structures, the adduct is incorporated into a B-form DNA helix but is stabilized by different mechanisms. Positioned opposite dA, a staggered ϵC(*anti*)•dA(*anti*) alignment allowed partial intercalation within the base pair, stabilized by hydrophobic interactions. Positioned opposite T, ϵC adopted a *syn* conformation around the glycosydic torsion angle, resulting in a ϵC(*syn*)•T(*anti*) alignment stabilized by hydrogen bonding. These structures are mutagenic intermediates during DNA synthesis. Although in mammalian cells, ϵC-induced mutations are common, the most frequent outcome of DNA replication (Moriya *et al.*, 1994) and feasible substrate of repair enzymes (Dosanjh *et al.*, 1994; Saparbaev *et al.*, 1995; Hang *et al.*, 1996) is DNA containing the lesion positioned opposite dG. We describe here the solution structure of a d(C-G-T-A-C-ϵC-C-A-T-G-C)•d(G-C-A-T-G-G-G-T-A-C-G) oligodeoxynucleotide duplex (referred to throughout this paper as the ϵC•dG duplex), as determined by high-resolution NMR spectroscopy and restrained molecular dynamics simulations.

The chemical structure of 3,N^4-etheno-2′-deoxycytidine and the numbering scheme for the duplex sequence employed in this paper are shown in Figure 1.

MATERIALS AND METHODS

Synthesis and Purification of Oligodeoxynucleotide Duplexes. 3,N^4-etheno-2′-deoxycytidine 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. Sequences containing an O-5′-dimethoxytrityl group were isolated by treatment of the crude synthesis product with concentrated ammonia for 46 h at room temperature and purified 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 desired sequence eluted as a main fraction at 33 min. The O-5′-dimethoxytrityl group was cleaved by treatment with 80% acetic acid for 30 min and the solution extracted with ether three times before purification by HPLC. Oligodeoxynucleotide sequences were desalted by passing them through a Sephadex G-25 column and converted to the sodium salt using a Dowex 50W cation exchange column.

Duplex Formation and Sample Preparation. A 1:1 stoichiometry of the duplex was obtained by monitoring intensity of NMR proton signals during gradual addition of the unmodified strand to the ϵC-containing strand at 55 °C. NMR samples consisted of 400 OD_{260} of the duplex dissolved in 0.6 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 5.5 mM. Samples were degassed under nitrogen before collection of the NMR data.

NMR Experiments. One- and two-dimensional NMR spectra were recorded on Bruker (AMX) and Varian (Inova) spectrometers operating at 600 and 500 MHz, respectively. Proton chemical shifts were referenced relative to TSP at 0.0 ppm. Phase-sensitive (States *et al.*, 1982) NOESY (50, 90, 150, 200, and 300 ms mixing times), COSY, DQF-COSY, and TOCSY spectra in D_2O buffer were collected with a repetition delay of 1.5 s, during which the residual water signal was suppressed by saturation. A NOESY (150 ms mixing time) spectrum in H_2O buffer was recorded using a jump-return reading pulse (Plateu & Gueron, 1982). Time domain data sets consisted of 2048 × 300 complex data points in the t_2 and t_1 dimensions, respectively. A COSY45 spectrum was recorded utilizing 4096 complex points × 1200 increments, resulting in a digital resolution of 1.2 Hz/point in the t_2 dimension. A ^{15}N-HMQC spectrum was recorded in H_2O buffer applying jump-return pulse sequences to suppress the water signal (Bax *et al.*, 1990) with 1024 complex points × 128 increments in the proton and nitrogen dimensions, respectively. NMR data were processed using Felix (Biosym Technologies, Inc.) running on Silicon Graphics computers. Time domain data sets were multiplied by shifted sinebell window functions prior to Fourier transformation. No base line correction was applied to the transformed spectra.

Molecular Dynamics Simulations. Molecular dynamics (MD) simulations were run on Silicon Graphics and Sun computers using X-PLOR 3.1 (Brunger, 1993). The structures were visualized with Midas Plus (UCSF, Computer Graphics Laboratory) and structural parameters obtained with Curves (Lavery & Sklenar, 1988; Lavery & Sklenar, 1989).

Simulations were performed *in vacuo* using an all-atom force field derived from CHARMM (Brooks *et al.*, 1983). Partial atomic charges on phosphate groups were not reduced, resulting in deoxynucleotide residues with a net charge of −1. The value of the dielectric constant was set to 4 (Friedman & Honig, 1992). Cross-peak volumes measured from NOESY (50, 90, 150, 200, and 300 ms mixing times) spectra collected in D_2O buffer were input into the program Mardigras (UCSF) to calculate interproton distances taking into consideration spin diffusion effects (Borgias & James, 1990). A single set of distance bounds was obtained by adding ±0.8 Å to the averaged value determined by Mardigras and used during MD without any further readjustment. Experimental distance bounds were enforced using an empirical square-well energy potential functions with a force constant of 30 kcal/(mol Å^2). Watson−Crick hydrogen bond alignments were enforced by distance restraint potentials on all canonical base pairs of the duplex. Backbone

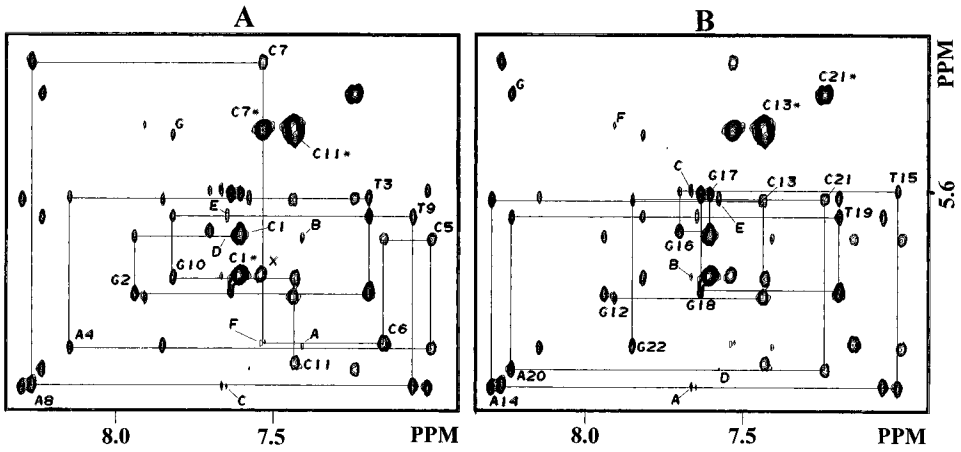


FIGURE 2: Duplicate contour plots of a portion of the NOESY (200 ms mixing time) spectrum recorded in D_2O buffer, pH 6.9, 25 °C. The figure depicts distance connectivities between base and H1′ sugar protons in the (A) modified and (B) unmodified strands of the ϵC·dG duplex. Solid lines connect each base (purine H8/pyrimidine H6) proton to its own (peaks labeled on the figure) and 5′-flanking H1′ sugar protons. Labeled peaks are assigned as follows: (A) A, A4(H2)-A4(H1′); B, A4(H2)-C5(H1′); C, A8(H2)-A8(H1′); D, A8(H2)-G16(H1′); E, A8(H2)-T9(H1′); F, C7(H6)-ϵC6(H1′), and G, G10(H8)-C11(H5). (B) A, A14(H2)-A14(H1′); B, A14(H2)-G10(H1′); C, A14(H2)-T15-(H1′); D, A20(H2)-A20(H1′); E, A20(H2)-C21(H1′), and/or A20(H2)-T3(H1′); F, G12(H8)-C13(H5); G, A20(H8)-C21(H5). Asterisks indicate the cytosine H5-H6 cross peaks. X, unknown impurity.

dihedral angles were restrained by an empirical square-well potential function with a width encompassing a range for both A- and B-form DNA. Sugar puckers and glycosylic torsion angles were not restrained during the simulations.

Six different initial models containing the ϵC adduct were built by adding an etheno bridge to the N^4- and N^3-nitrogens of a deoxycytidine residue. Two of the initial models were canonical A- and B-form 11-mer duplexes having a coplanar alignment on the ϵC6·G17 base pair. The other four models (two from B- and two from A-form DNA) were derived from 12-mer duplexes by deleting, in a staggered way, one residue from each strand at consecutive C·G base pairs located at the center. The resulting 11-mer structures were termed 3′-staggered or 5′-staggered depending whether G17 was facing the 3′- or 5′-side of the lesion, respectively. To relieve steric clashes and optimize covalent geometry, the initial models were energy minimized by the conjugate gradient method prior to MD. Simulations consisted of an equilibration step during which the temperature was increased from 115 to 350 K in 10 ps, and the scale of the NOE distance restraint potential function was gradually augmented. Once the equilibration temperature was obtained, the simulation continued at 350 K for 5−8 ps, followed by cooling to 300 K in 2.5 ps, and 80 ps of restrained dynamics at constant temperature. For each initial model, seven structures were calculated using either a different set of initial velocities (three structures) or by cooling to 300 K at different times of the equilibration period (5−8 ps). Coordinates of the last 10 ps of the simulation were averaged and subjected to 2500 steps of energy minimization, yielding distance-refined structures. Covalent bond lengths of all hydrogens were kept constant during the simulations by using the SHAKE algorithm (Ryckaert *et al.*, 1977).

Distance-refined structures within each initial model were averaged and further optimized by 30 ps of MD at 300 K restrained by the full relaxation matrix method (Yip & Case, 1989; Nilges *et al.*, 1991). During the first 3.5 ps, distance restraint potentials were gradually reduced to zero, and the scale of a potential function based on the intensity difference between experimental and back-calculated NOESY spectra was increased simultaneously. A total of 1440 experimental volumes measured from NOESY (50, 90, 150, 200, and 300 ms mixing times) spectra were input as restraints with a 30% error range. A grid search determined that an isotropic correlation time of 2.25 ns produced the closest initial fit to the experimental data, and this value was used during the refinement. Atom coordinates stored during the last 2 ps of the simulation were averaged and energy minimized yielding final NMR-refined structures.

RESULTS

Assignment of the Nonexchangeable Protons. The one-dimensional proton spectrum of the ϵC·dG duplex in D_2O buffer, pH 6.9, 25 °C, is characterized by the presence of sharp signals and the absence of minor resonances (Figure 1S, Supporting Information). Assignment of the nonexchangeable base and H1′ sugar protons followed the analysis of NOESY and COSY spectra using standard procedures (Hare *et al.*, 1983; Wüthrich, 1986; van de Ven & Hilbers, 1988). Figure 2 shows duplicate contour plots of a NOESY (200 ms mixing time) spectrum recorded in D_2O buffer, pH 6.9, 25 °C, depicting distance connectivities between the base (purine H8/pyrimidine H6) (6.9−8.4 ppm) and the sugar H1′ (5.1−6.3 ppm) proton regions. Characteristic of a right-handed helix, each base proton exhibits NOE peaks to its own and 5′-flanking H1′ sugar protons, as shown from C1 to C11 in the ϵC-containing strand (Figure 2A) and from G12 to G22 in the unmodified strand (Figure 2B). Also indicative of a right-handed helix, sequential NOE peaks are observed between A4(H2)-C5(H1′), A8(H2)-T9(H1′), A14-(H2)-T15(H1′), and A20(H2)-C21(H1′) protons (peaks B and E, Figure 2A, and C and E, Figure 2B, respectively) and between G10(H8)-C11(H5), G12(H8)-C13(H5), and A20-(H8)-C21(H5) protons (peaks G, Figure 2A, and F and G, Figure 2B, respectively). The sequential ϵC(H1′) to C7(H6) NOE cross-peak (Figure 2A, peak F) is very weak, suggesting a longer distance between these protons at the ϵC6-C7 step of the duplex.

The H5, H7, and H8 protons of ϵC, as well as the H5 proton of C5, at the 5′-side of the adduct, resonate outside the limits of Figure 2. A contour plot showing distance connectivities between the base (6.6−8.4 ppm) and the sugar H1′/H3′ (5.1−7.0 ppm) proton regions of a NOESY (300

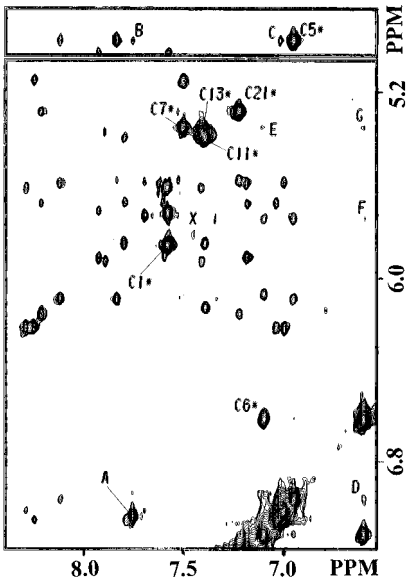


FIGURE 3: Contour plot of a portion of the NOESY (300 ms mixing time) spectrum recorded in D_2O buffer, pH 6.9, 15 °C. The spectrum shows distance interactions between the base and H1′ sugar protons. Labeled peaks are assigned as follows: (A) A, ϵC(H8)-ϵC(H7); B, ϵC(H8)-C5(H5); C, ϵC(H7)-C5(H5); D, C5(H6)-ϵC(H5); E, ϵC-(H6)-C7(H5), and F, ϵC(H5)-C5(H1′). Asterisks indicate the cytosine H5-H6 cross peaks. X, unknown impurity.

ms mixing time) spectrum collected on the same duplex sample at 15 °C is displayed in Figure 3. Seven strong NOE peaks, originated in the short (H5-H6) interproton distance (2.45 Å), account for all cytosine residues of the duplex. The C5(H5) signal, absent on the NOESY spectrum at 25 °C, is now observed near to the water at 4.62 ppm, while ϵC(H5) proton has moved in the opposite direction and resonates in the base proton region (Figure 3, peaks C5* and C6*, respectively). A contour plot of an expanded region of a phase-sensitive COSY spectrum recorded in D_2O buffer, pH 6.1, 13 °C displayed in Figure 2S (Supporting Information), confirm these assignments. ϵC(H7) and ϵC-(H8) protons were identified by their strong distance interaction observed on all NOESY spectra (Figure 3, peak A). As it was previously observed (Korobka *et al*., 1996; Cullinan *et al*., 1996), these protons have a small *J* coupling constant and did not show a cross-peak between them in the phase-sensitive COSY experiment recorded in D_2O buffer at 13 °C (Figure 2S, Supporting Information) nor in the phase-sensitive COSY45 spectrum recorded at 27 °C. ϵC(H7) and ϵC(H8) exhibited distance interactions to the flanking C5-(H5) proton on the NOESY spectrum recorded in D_2O buffer, pH 6.9, at 15 °C (Figure 3, peaks B and C).

Assignment of the H2′, H2″, H3′, and H4′ sugar protons was based on NOESY (50 and 200 ms mixing time), COSY, DQF-COSY, and HOHAHA spectra following established procedures (Hare *et al.*, 1983; Wüthrich, 1986; van de Ven & Hilbers, 1988). The chemical shifts of the nonexchangeable protons detected in D_2O buffer, pH 6.9, 25 °C, are listed in Table 1.

Assignment of the Exchangeable Protons: The imino and base/amino regions (5.8−14.0 ppm) of a one-dimensional proton spectrum recorded in H_2O buffer, pH 6.9, 2 °C is shown in Figure 3S (Supporting Information). Nine partially resolved signals between 12.0 and 14.0 ppm and one sharp resonance at 9.7 ppm account for all the imino protons of the ϵC·dG duplex. Assignment of the exchangeable protons

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

	H8/H6	H5/CH_3/H2	H1′	H2′	H2″	H3′	H4′
C1	7.63	5.90	5.78	2.00	2.41	4.69	4.02
G2	7.96		5.95	2.66	2.75	4.96	4.32
T3	7.22	1.49	5.65	2.01	2.38	4.84	4.15
A4	8.17	7.43	6.13	2.56	2.73	4.96	4.36
C5	7.02	4.62[c]	5.78	1.86	2.33	4.94	4.17
ϵC6[d]	7.17	6.61	6.11	2.02	2.29	4.84	3.96
C7	7.65	5.44	5.22	2.24	2.38	4.82	4.10
A8	8.29	7.67	6.24	2.69	2.91	5.01	4.38
T9	7.08	1.43	5.71	1.92	2.32	4.83	4.10
G10	7.84		5.90	2.58	2.67	4.94	4.32
C11	7.46	5.45	6.17	2.14	2.14	4.47	3.98
G12	7.93		5.96	2.59	2.76	4.82	4.12
C13	7.46	5.42	5.65	2.14	2.46	4.87	4.15
A14	8.32	7.69	6.24	2.69	2.90	5.02	4.38
T15	7.03	1.39	5.63	1.76	2.18	4.80	4.15
G16	7.73		5.75	2.57	2.68	4.93	4.35
G17	7.63		5.63	2.52	2.50	4.91	4.33
G18	7.66		5.94	2.52	2.76	4.88	4.32
T19	7.22	1.34	5.71	2.08	2.44	4.86	4.07
A20	8.26	7.60	6.20	2.66	2.83	5.00	4.41
C21	7.26	5.32	5.65	1.84	2.26	4.76	4.10
G22	7.88		6.12	2.56	2.34	4.63	4.14

[a] Chemical shift values given in parts per millon (ppm) relative to TSP. [b] Phosphate buffer (10 mM), pH 6.9, containing 50 mM NaCl. [c] Chemical shift value observed at 15 °C. [d] ϵC6(H7,H8) protons resonate at 7.00 ppm and 7.71 ppm, respectively.

followed the analysis of a NOESY (150 ms mixing time) spectrum recorded in H_2O buffer, pH 6.1, 2 °C. A contour plot of an expanded region showing distance connectivities between imino (12.2−13.9 ppm) and base/amino (5.1−8.7 ppm) protons is shown in Figure 4A. Upon duplex formation, deoxyadenosine H2 protons showed strong NOE interactions to the thymidine imino proton across A·T base pairs. The independent assignment of all deoxyadenosine H2 protons, from NOESY spectra recorded in D_2O, allowed identification of this interaction on the T3·A20, T9·A14, T15·A8, and T19·A4 base pairs of the duplex (Figure 4A, peaks A−D, respectively). Deoxycytidine amino protons were specifically assigned following observation of their intraresidue connectivity to the H5 proton, detected on a NOESY (150 ms mixing time) spectrum recorded in H_2O buffer, pH 6.1, 2 °C (peaks A−F, Figure 4S, Supporting Information). On formation of G·C base pairs, deoxyguanosine imino protons exhibited strong NOE peaks to both the hydrogen-bonded and non-hydrogen-bonded deoxycytidine amino protons across the G22·C1, G2·C21, G18·C5, G16·C7, G10·C13, and G12·C11 base pairs of the duplex (Figure 4A, peaks E, E′; F, F′; G, G′; H, H′; I, I′, and J, J′, respectively).

Indicative of base pair stacking, deoxyadenosine H2 protons displayed distance interactions to imino protons of flanking base pairs 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 4A, peaks K−P, respectively). Further evidence of base pair stacking is obtained on the symmetrical imino proton region (12.2−13.9 ppm) of the same NOESY spectrum plotted in Figure 4B. Distance interactions are detected between G2(N1H)-T3(N3H), T19-(N3H)-G18(N1H), G16(N1H)-T15(N3H), and T9(N3H)-G10(N1H) imino protons of sequential base pairs of the duplex (Figure 4B, peaks A, B, E, and F, respectively).

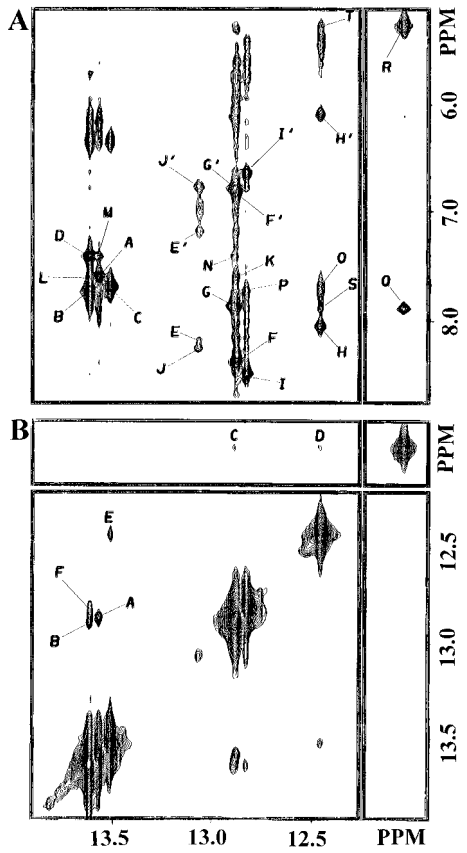


FIGURE 4: Contour plots of portions a NOESY (150 ms mixing time) spectrum recorded in H_2O buffer, pH 6.1, 2 °C. The figure displays distance connectivities between (A) imino and base/amino proton regions and (B) on the symmetrical imino proton region. Labeled peaks are assigned as follows: (A) A, T3(N3H)-A20(H2); B, T9(N3H)-A14(H2); C, T15(N3H)-A8(H2); D, T19(N3H)-A4-(H2); E, G22(N1H)-C1(N4H)$_{hb}$; E′, G22(N1H)-C1(N4H)$_{nhb}$; F, G2-(N1H)-C21(N4H)$_{hb}$; F′, G2(N1H)-C21(N4H)$_{nhb}$; G, G18(N1H)-C5(N4H)$_{hb}$; G′, G18(N1H)-C5(N4H)$_{nhb}$; H, G16(N1H)-C7(N4H)$_{hb}$; H′, G16(N1H)-C7(N4H)$_{nhb}$; I, G10(N1H)-C13(N4H)$_{hb}$; I′, G10-(N1H)-C13(N4H)$_{nhb}$; J, G12(N1H)-C11(N4H)$_{hb}$; J′, G12(N1H)-C11-(N4H)$_{nhb}$; K, G2(N1H)-A20(H2); L, T19(N3H)-A20(H2); M, T3(N3H)-A4(H2); N, G18(N1H)-A4(H2); O, G16(N1H)-A8(H2); P, G10(N1H)-A14(H2); Q, G17(N1H)-ϵC(H8); R, G17(N1H)-G17-(N2H1/N2H2), and S, G16(N1H)-ϵC(H8). (B) A, G2(N1H)-T3-(N3H); B, T19(N3H)-G18(N1H); C, G17(N1H)-G18(N1H); D, G17(N1H)-G16(N1H); E, G16(N1H)-T15(N3H), and F, T9(N3H)-G10(N1H). (N4H)$_{hb}$ and (N4H)$_{nhb}$ refer to the hydrogen-bonded and non-hydrogen-bonded cytidine amino protons, respectively.

The sharp resonance at 9.7 ppm was only observed with the sample dissolved in H_2O buffer and, as a result, has to originate from an exchangeable proton of G17. In the NOESY (150 ms mixing time) spectrum recorded in H_2O buffer, pH 6.1, 2 °C, this signal showed weak NOE peaks to the imino protons of G18 and G16 (Figure 4B, peaks C and D, respectively) and strong interactions to the ϵC(H8) proton of the adduct, as well as an exchangeable proton resonating at 5.3 ppm, near the water signal (Figure 4A, peaks Q and R, respectively). Its assignment was established following the analysis a ^{15}N-HMQC spectrum, recorded in H_2O buffer, pH 6.1, 5 °C. Figure 5 shows an expanded region of the spectrum depicting correlations between imino proton (12.0−13.9 ppm) and imino nitrogen (139.1−168.5 ppm) regions. The 9.7 ppm signal interacts with a nitrogen at 142.0 ppm, within the imino nitrogen region of the spectrum, and far from the range observed for exocyclic amino nitrogens (70−100 ppm). Subsequently, it was assigned to the G17(N1H) proton. Chemical shifts of the exchangeable protons and imino nitrogens are listed in Table 2.

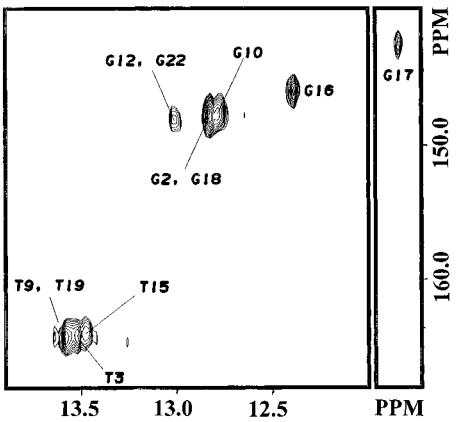


FIGURE 5: Contour plot of an expanded region of a ^{15}N-HMQC spectrum recorded in H_2O buffer, pH 6.1, at 5 °C. The figure shows correlations between the proton and nitrogen imino regions.

Table 2: Nitrogen and Exchangeable Proton Chemical Shifts[a] in H_2O buffer[b]

	T(N3H)	G(N1H)	C(N4H)$_{hb}$[c]	C(N4H)$_{nhb}$[c]	A(H2)	G(N1), T(N3)
C1·G22		13.06	8.19	7.12		147.9
G2·C21		12.88	8.38	6.83		147.6
T3·A20	13.57				7.60	164.3
A4·T19	13.62				7.40	164.7
C5·G18		12.89	7.86	6.79		147.6
ϵC6·G17[d]		9.70				142.0
C7·G16		12.45	8.05	6.11		145.5
A8·T15	13.51				7.68	164.0
T9·A14	13.61				7.71	164.7
G10·C13		12.82	8.50	6.64		147.3
C11·G12		13.06	8.25	6.77		147.9

[a] Values in ppm relative to liquid ammonia (^{15}N) and DSS (^{1}H). Proton and nitrogen chemical shifts measured at 2 and 5 °C, respectively. [b] Buffer phosphate (10 mM), pH 6.1, 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] G17(N2H) protons resonate at 5.31 ppm.

Temperature Dependence of the Imino Protons. The one-dimensional spectra of the ϵC·dG duplex displayed in Figure 6, depict the behavior of the imino protons on raising the temperature from 0 to 40 °C. At 0 °C, all imino protons, including G17 on the lesion containing base pair and terminal G12(N1H) and G22(N1H), are detected as sharp signals. On raising the temperature to 20 °C, the imino proton signals of terminal base pairs broaden as a consequence of an increase of the water exchange rate, but G17(N1H) is unaffected. At 30 °C, the imino proton signals of terminal base pairs have almost disappeared, while G17(N1H) and, to a lesser extent, G16(N1H) but not G18(N1H) signals started to broaden. At 40 °C, all imino proton signals are broad indicative of the melting temperature of the ϵC·dG duplex.

Sugar Coupling Constants. $J_{H1'-H2'}$ and $J_{H1'-H2''}$ coupling constants values were determined from the analysis of a phase-sensitive COSY45 spectrum recorded in D_2O buffer, pH 6.9, 27 °C. An expanded region showing *J* interactions between the H1′ (5.1−6.4 ppm) and H2′, H2″ (1.8−2.9 ppm) sugar proton regions is shown in Figure 5S (Supporting Information). The pattern of the H1′/H2′ and H1′/H2″ cross

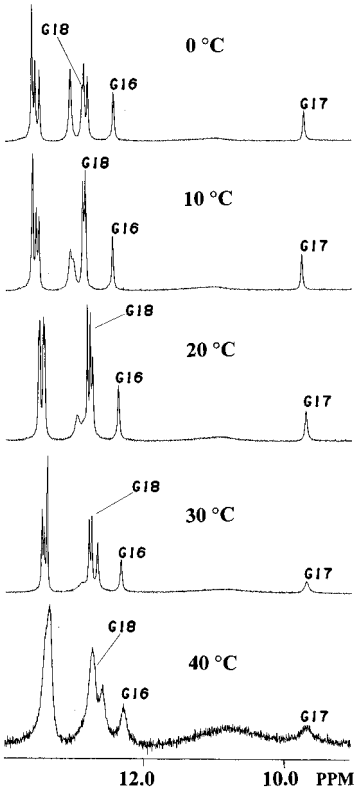


FIGURE 6: Imino proton region of one-dimensional spectra recorded in H_2O buffer, pH 6.1, from 0 to 40 °C. Assignment of G16(N1H), G17(N1H), and G18(N1H) protons is given on the spectra.

peaks establishes that, with the exception of ϵC, $J_{H1'-H2'} > J_{H1'-H2''}$ on all sugars of the duplex (Majumdar & Hosur, 1992). Although measured *J* values do not correspond to a unique sugar conformation, they establish major sugar populations on the C2′-endo/C3′-exo range (Rinkel & Altona, 1987). In the case of the adduct, its H1′-H2′ cross-peak is clearly skewed toward a different direction (Box A, Figure 5S, Supporting Information), indicating a $J_{H1'-H2''} > J_{H1'-H2'}$. Thus, ϵC has a major population with sugar conformations in the C3′-endo/C4′-exo range. $J_{H1'-H2'}$ and $J_{H1'-H2''}$ coupling constants values are listed in Table 1S, Supporting Information.

Chemical Shifts. Assignment of the base and sugar protons on the ϵC·dG duplex allowed us to identify large variations in the proton chemical shifts. These values may be compared with those recently reported for ϵC·T and ϵC·dA duplexes, whose sequences only differ in the residue positioned opposite the adduct, T and dA, respectively (Cullinan *et al.*, 1996; Korobka *et al.*, 1996). Figure 7 depicts chemical shift differences observed for the nonexchangeable base and sugar protons of the three central base pairs of the duplexes, and Tables 2S and 3S (Supporting Information) list differences for all the protons of the sequence. In general, small chemical shift differences are detected for protons on the outer base pairs of the duplex (Tables 2S and 3S, Supporting Information). Closer to the lesion, the major groove protons of C5 and, to a lesser extent, of G18 are shielded on the ϵC·dG duplex. On the contrary, major groove protons of ϵC and C7 have moved in the opposite direction, indicating that they are deshielded (Figure 7A). The same is true for the exocyclic H7 and H8 protons, which are deshielded by 0.79 and 1.22 ppm and by 0.20 and 0.81 ppm when compared to the ϵC·dA and ϵC·T duplexes, respectively (Tables 2S and 3S, Supporting Information). Chemical shift differences among minor groove protons at the lesion site are smaller and more scattered. In general, shielding is observed on the ϵC·dG duplex for the H1′ and H2″ protons of C5, ϵC, and C7 in the lesion-containing strand, while deshielding is detected on protons of the complementary strand (Figure 7B).

Molecular Dynamics Simulations. Figure 6S (Supporting Information) shows an overlap view of the central five base pair segment of the distance-refined structures and Table 4S (Supporting Information) lists their RMSD. The structures obtained when MD was started from A- and B-form coplanar and A- and B-form derived 3′-staggered initial models (28 total) converged to a structural motif with RMSD smaller than 0.9 Å from a mean structure (top structures, Figure 6S, Supporting Information). In these models, ϵC is partially displaced toward the major groove of the helix and, to some extent, coplanar with G17, its partner on the opposite strand. A second structural motif, which differs with the previous by more than 2.7 Å in the RMSD of the atom positions, was obtained when the refinement started from A- and B-form derived 5′-staggered models (Table 4S, Supporting Information). In this case, the refined structures are highly bent at the lesion site and have G17 positioned in the major groove and almost parallel to the helical axis (bottom structures, Figure 6S, Supporting Information). Despite the small RMSD among them (<0.6 Å with reference to their average), these structures are aberrant, being incompatible with the spectroscopic characteristics of the G17 imino proton and the chemical shift variations observed at the lesion site. Therefore, additional refinement was not performed.

Mean structures, calculated from the distance-refined models in agreement with the NMR data (one for each group), were further optimized by 30 ps of MD retrained by the full relaxation matrix approach. Figure 8 shows an overlap view of the final NMR-refined structures and Table 3 list the violations from experimental data and idealized geometry. All four models reproduce the experimental NOE intensities with similar NMR *R*-factors (7.1−7.4%) without significant distortions of the covalent geometry (Table 3). The structures belong to the B-form DNA family, are slightly bent, and have a local perturbation of the sugar-phosphate backbone at the lesion site. With the exception of ϵC·G17, all base pairs of the duplex are properly stacked in the helix. ϵC and terminal C11 residues have C3′-endo sugar puckers while all other sugars are in the C2′-endo range.

Figure 9 shows the middle of the duplex as viewed from the minor groove of the helix, and Figure 10 displays the same segment seen from the top. ϵC is displaced into the major groove of the helix, and G17 has moved in the opposite direction, resulting in a base pair which is sheared by 4.0 Å and propeller twisted 25°. On the lesion-containing strand, residues are over- and undertwisted at C5-ϵC6 and ϵC6-C7 steps of the duplex, respectively, while values close to normal are observed on the partner strand. The adduct is shifted 2.7 Å and −2.2 Å at C5-ϵC6 and ϵC6-C7 steps, respectively, but again values close to normal are observed on the partner strand. Base pairs adjacent to the lesion are fully Watson−Crick hydrogen bonded and C5·G18 has 22° of base pair buckle. A mean distance of 2.5 Å separates ϵC(O2) and G17(N1H), indicating the formation of a hydrogen bond across the ϵC·G17 base pair. In addition, G17(N2H1) and G17(N2H2) are within hydrogen bond distance of C7(O4′) and G18(O4′), respectively, suggesting the presence of

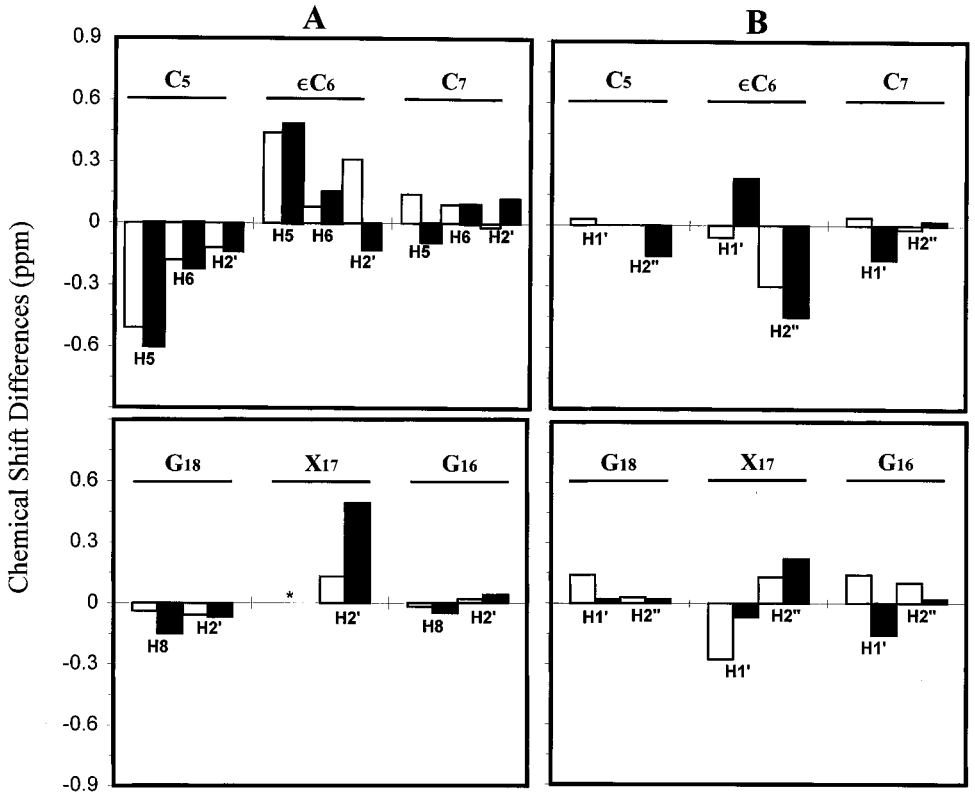


FIGURE 7: Proton chemical shift differences observed on the central three base pair segment of duplexes containing an ϵC adduct at 25 °C. (A) Base (purine H8/pyrimidine H6/H5) and sugar H2′ protons; B: H1′ and H2″ sugar protons. Open bars depict (ϵC·dG − ϵC·dA) differences (Korobka *et al.*, 1996); filled bars (ϵC·dG − ϵC·T) differences (Cullinan *et al.*, 1996). Negative values indicate shielding of the resonance on the ϵC·dG duplex. The asterisk (*) means that differences between the base proton of the residue complementary to ϵC are not informed.

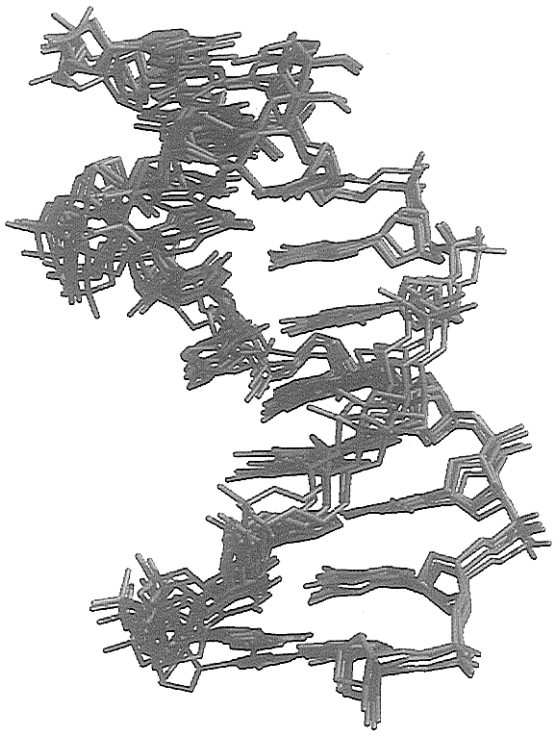


FIGURE 8: Three-dimensional models of the ϵC·dG duplex. The picture shows four converging structures obtained after 127.5 ps of NMR-restrained molecular dynamics starting from coplanar and 3′-staggered initial models (derived from A- and B-form DNA). The helix is slightly bend and the ϵC adduct is displaced in the major groove.

additional hydrogen bonds at the lesion site. Table 4 lists structural parameters measured from NMR-refined structures and Table 5S (Supporting Information) compares experimental and refined distances on the central segment of the duplex.

DISCUSSION

NMR spectra. Nonexchangeable Protons. The NMR spectra recorded in D_2O and H_2O buffer are characterized by the presence of sharp peaks (Figure 6; Figures 1S and 3S, Supporting Information). Their subsequent assignment to only one set of proton resonances (Tables 1 and 2) supports the existence of a single averaged structure for the ϵC·dG duplex in solution. The directionality of sequential distance interactions observed in the NOESY spectra between the base (purine H8/pyrimidine H6) and the sugar H1′, H2′, H2″, H3′, deoxycytidine H5, and thymine methyl protons clearly indicates that the ϵC·dG duplex adopts a right handed helical structure (Hare *et al.*, 1983; Wüthrich, 1986; van de Ven & Hilbers, 1988). Furthermore, the intensity of intraresidue NOE peaks between the base and H1′ sugar protons (Figures 2 and 3) establish glycosylic torsion angles in *anti* conformation for all residues of the duplex.

Unlike our previous experience (Korobka *et al.*, 1996; Cullinan *et al.*, 1996), all four base protons of the adduct, ϵC(H5), ϵC(H6), ϵC(H7), and ϵC(H8), resonate within 1.1 ppm in the base proton region of the spectrum, suggesting increased solvent exposure in the ϵC·dG duplex. The stereospecific assignment of ϵC(H7) and ϵC(H8) protons was based on distance interactions observed in a NOESY spectrum recorded in H_2O buffer and structural considerations. When the glycosylic torsion angle of the adduct is *anti*, its H8 proton is located toward the center of the helix

Table 3: Structure Calculations

	initial A-form		initial B-form	
	coplanar	3′-staggered	coplanar	3′-staggered
deviations from experimental restraints[a]				
RMSD NOE-distances[b] (Å)	0.016 (0.45)	0.015 (0.63)	0.013 (0.33)	0.018 (0.50)
NOE energy[b] (kcal/mol Å^2)	9.1 ± 3.0	5.8 ± 2.0	4.1 ± 0.6	8.7 ± 0.5
R-factors[c]	0.074 (0.212)	0.071 (0.220)	0.073 (0.108)	0.073 (0.112)
van der Waals energy (kcal/mol)	−353	−354	−353	−354
RMSD from idealized geometry[a]				
bond lengths (Å)	0.0062	0.0065	0.0063	0.0064
bond angles (°)	3.4	3.5	3.4	3.5
improper angles (°)	0.28	0.26	0.29	0.30

[a] Measured on each of the final refined structures. [b] Average values determined on distance-refined structures. Total of 265 experimental distances. [c] *R*-factor ($^1/_6$) as defined by X-PLOR. Total of 1440 NOE volumes. Between brackets are the values measured on the initial models.

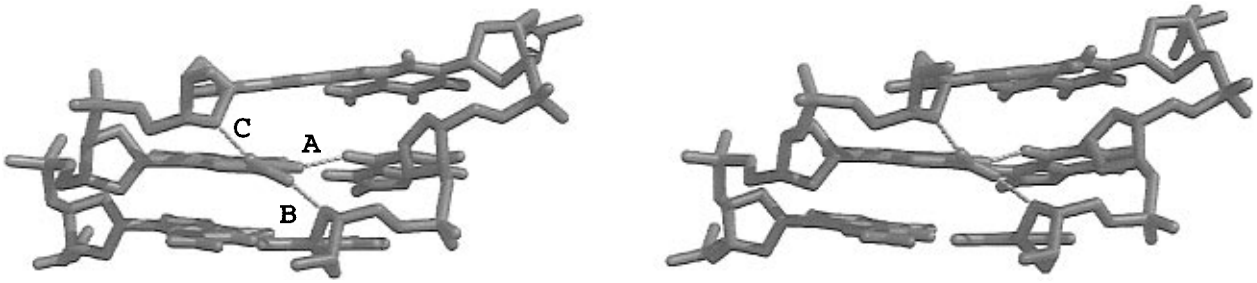


FIGURE 9: View of the central d(C5-ϵC6-C7)·d(G16-G17-G18) segment of the ϵC·dG duplex structure seen from the minor groove side. The picture depicts hydrogen bonds formed at the lesion site: A, G17(N1H)-ϵC(O2); B, G17(N2H1)-C7(O4′); C, G17(N2H2)-G18(O4′). Only the base protons are displayed.

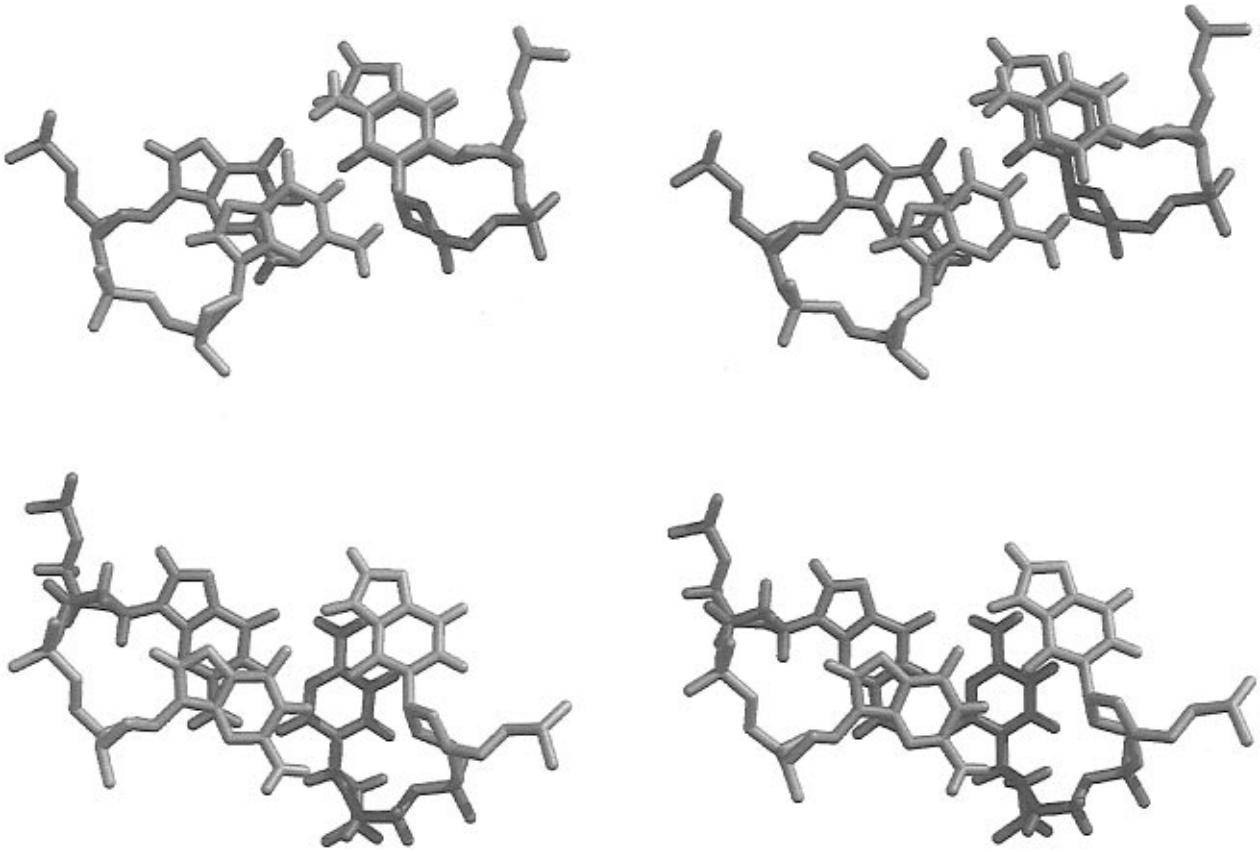


FIGURE 10: View from the top of the helix showing stacking interactions between (top) the C5·G18 and ϵC6·G17 and (bottom) the ϵC6·-G17 and C7·G16 base pairs of the duplex.

and shows NOE peaks to the imino and amino protons of G17 on the opposite strand (peak Q, Figure 4A, and peaks H and I, Figure 4S, Supporting Information). On the contrary, ϵC(H7) is positioned in the major groove of the helix and does not exhibit such distance interactions.

Exchangeable Protons. The chemical shift of the imino protons (12−14 ppm) and the distance interactions observed between T(N3H) and A(H2) as well as G(N1H) and C(N4H) protons (Figure 4A, peaks A−J,J′) confirm canonical Watson−Crick alignment for all base pairs on either side of the

Table 4: Structural Parameters[a]

helix bend (°)	23 ± 9
ϵC·G17 Base Pair Alignment	
ϵC X-displacement (Å)	0.5 ± 0.2
G17 X-displacement (Å)	−3.5 ± 0.5
shear (Å)	4.0 ± 0.2
propeller twist (°)	−25 ± 10
opening (°)	18 ± 5

Interbase Parameters

	twist (°)	shift (Å)
C5-ϵC	57 ± 5	2.7 ± 1.0
ϵC-C7	10 ± 6	−2.2 ± 1.0
G18-G17	42 ± 5	−0.6 ± 0.4
G17-G16	32 ± 5	0.7 ± 0.4

Hydrogen Bonding at the Lesion Site

	distance (Å)	energy (kcal/mol)[b]
ϵC(O2)-G17(N1H)	2.5 ± 0.2	−0.7 ± 0.5
G17(N2H1)-C7(O4′)	1.9 ± 0.1	−3.4 ± 0.5
G17(N2H2)-G18(O4′)	2.6 ± 0.6	−1.5 ± 0.9

[a] Average values and bounds measured on the refined structures. Parameters as defined by Curves (Lavery & Sklenar, 1989). Canonical B-form values are X-displacement, 0.0 Å; base pair shear, 0.0 Å; base pair propeller twist, 3.7°; base pair opening, −4.1°; interbase twist, 36.0°; interbase shift, 0.0 Å. [b] Calculated using the explicit hydrogen bond energy term of X-PLOR. For comparison, energies of G16(N1H)-C7(N3) and G18(N1H)-C5(N3) hydrogen bonds were −2.6 and −3.0 kcal/mol, respectively.

lesion. In addition, sequential distance connectivities observed between imino and A(H2) protons (Figure 4A, peaks K−P) and among the imino protons (Figure 4B, peaks A−F) indicates favorable base pair stacking throughout the ϵC·dG duplex (Wüthrich, 1986).

Alignment of the ϵC·dG Base Pair. With ϵC and G17 residues adopting an *anti* conformation around the glycosylic angle, a coplanar alignment across the base pair results in severe steric clashes that are relieved by adjustment of the local structure. The intensity of sequential NOE peaks established a long distance between ϵC(H1′) and C7(H6), while normal distances are observed for the C5-ϵC6, G16-G17 and G17-G18 steps of the duplex (Figure 2). This observation rules out a staggered ϵC·G17 base pair alignment, which requires long distances between base and H1′ protons on both strands of the duplex, as established for the ϵC·dA duplex (Korobka *et al.*, 1996).

The G17(N1H) proton showed comparable NOE interactions with G16(N1H) and G18(N1H) (Figure 4B, peaks C and D), suggesting similar distances between imino protons on the G16-G17-G18 stretch of the duplex. Likewise, ϵC-(H8) proton displayed distance connectivity with G17(N1H) on the unmodified strand (peak Q, Figure 4A) and with C5-(H5), C5(N4H), and C7(N4H) located at both sides of the lesion (peaks K, J′, and J, Figure 4S, Supporting Information). In addition, the line width of G17(N1H) signal is similar to other imino protons of the duplex, implying equivalent rates of exchange with water. Indeed, G17(N1H) signal is noticeably broad only above 30 °C, when "melting" of the duplex has started (Figure 6). Hence, we conclude that the ϵC·dG base pair is coplanar and G17(N1H) is hydrogen bonded.

ϵC·dG Duplex Structure. To extend the three-dimensional space explored by the computer simulation and evaluate the precision of the structure determination, we ran MD starting from very different DNA conformations and calculated several structures for each of the initial models. The RMSD in the atom positions along with the visual inspection of the refined structures established the existence of two different motifs determined by restrained MD (Figure 6S and Table 4S, Supporting Information). Calculations starting from the 5′-staggered models (derived from A- and B-form DNA) converged to structures with RMSD <0.6 Å with respect to their mean, but incompatible with the NMR data. Specifically, experimental distances determined at the lesion site, chemical shifts variations, and alignment of the ϵC·dG base pair are not explained by these structures. On the other hand, MD calculations starting from coplanar and 3′-staggered initial models (derived from A- and B-form DNA) converged to distance refined structures having RMSD <0.9 Å, with respect to their average, and in excellent agreement with the experimental data (Figure 6S and Table 4S, Supporting Information). Further refinement of averaged structures by the full relaxation matrix approach decreased the NMR *R*-factors to 0.07 (Table 3) increasing the accuracy of the final NMR-refined models. Our discussion focuses on this structural motif.

The ϵC·dG duplex structure is a right-handed helix within the B-form family, with an average bend of 25° at the lesion site and no major perturbations of the sugar-phosphate backbone. The steric clashes produced at the lesion-containing base pair are relieved by displacing ϵC and G17 into the major and minor groove of the helix, respectively. In addition, on the C5-ϵC6 step, the adduct is shifted toward the major groove and, on the following step, C7 is back into the helix. The complementary strand does not follow this trend, and residues are not shifted (Figures 9 and 10; Table 4). The net result is a sheared ϵC(*anti*)·G17(*anti*) alignment which retains stacking of G17 with the flanking residues and allows hydrogen bonding across the base pair. This is a novel structural motif for exocyclic adducts, differing from the staggered alignment observed for the ϵC·dA duplex (Korobka *et al.*, 1996) as well as the partially intercalated structure of the ϵA adduct (Kouchakdjian *et al.*, 1991a; Huang *et al.*, 1993).

The chemical shift differences observed between the exchangeable and nonexchangeable protons of the ϵC·dG and ϵC·dA duplexes (Korobka *et al.*, 1996) are in full support of the structure. The downfield shift detected for ϵC (H5, H6, H7, and H8) proton signals (Figure 7) reflects their increased exposure to water that results from the adduct displacement and shift present in the structure (Figures 8 and 10). In addition, the upfield shift of flanking major groove proton signals, namely C5(H5, H6, and H2′) and C7-(N4H1 and N4H2) (Figure 7A and Table 2S, Supporting Information), is the result of their shielding by ϵC which is positioned in the major groove of the helix (Figure 10).

All the refined models have a strong hydrogen bond between C7(O4′) and G17(N2H2), as well as a weak one between ϵC(O2) and G17(N1H) across the lesion-containing base pair. In addition, G17 can be stabilized further by the presence of a hydrogen bond between G17(N2H1) and G18-(O4′) (Figures 9 and 10; Table 4). Notably, this network of hydrogen bonds at the lesion site do not perturb flanking C5·G18 and C7·G16 base pairs, which retained Watson−Crick alignment. Hydrogen bonding at the lesion site is fully supported by the experimental data which showed the G17-

(N1H) proton resonating as a sharp signal in slow water exchange. Furthermore, this proton was remarkably resistant to water exchange having a rate comparable to G16(N1H) (Figure 6). Considering that proton exchange is only possible when the base pair is transiently opened and that C7·G16 is fully hydrogen bonded, similar water exchange rate strongly argues for additional hydrogen bonding at the lesion site and supports the existence in solution of the G17(N2H)-C7(O4′) and G17(N2H)-G18(O4′) hydrogen bonds.

Interproton distances measured in the refined structure are within the NMR-determined distance bounds (Table 5S, Supporting Information), giving further support to the three dimensional model presented here.

Biological Implications. DNA duplexes containing ϵC form stable complexes in aqueous solution (Korobka *et al*., 1996; Cullinan *et al*., 1996; this paper); these structures provide clues to key structural intermediates in translesional synthesis. In *E. coli*, the mutagenic frequency of ϵC is remarkably low (2%), indicating that ϵC(*anti*)·dG(*anti*) pairing is principally involved. Undoubtedly, the hydrogen bonds formed by the exchangeable protons of dG contribute to the stability of the lesion-containing base pair. The partially stacked ϵC(*anti*)·dA(*anti*) alignment and the ϵC-(*syn*)·T(*anti*) base pair, the latter with a single hydrogen bond, permit translesional synthesis in mammalian and SOS-induced *E. coli* cells.

The relative frequency of insertion of dNTPs opposite ϵC (F_{ins}) and the frequency of extension from the 3′-terminus of the replication fork (F_{ext}) have been determined in primer extension reactions catalyzed by the Klenow fragment of pol I (Zhang *et al.*, 1995). dAMP and TMP are both more efficiently inserted and extended by pol I than dGMP. Possibly, the substrate-binding pocket of pol I cannot accommodate a displaced adduct required to form an ϵC-(*anti*)·dG(*anti*) pair, as reported in this paper.

Lesion structure is not the only factor determining mutagenic specificity; the proofreading function of DNA polymerase and the frequency with which dAMP and dTMP are inserted opposite the lesion and then extended from the 3′-terminus may be involved in the selection process. These factors were studied in a mutant strain of *E. coli* in which DNA polymerase III lacks a proofreading function. Following SOS induction, there is a marked increase in base substitutions (dA and dT for ϵC) compared with wild-type strains (Moriya *et al.*, 1994). Thus, proofreading appears to play a role in the selection process. In the ϵC·dA intermediate, it is noteworthy that stacking forces allows translesional synthesis.

In contrast to results in *E. coli*, the mutagenic frequency of ϵC in simian kidney cells is very high (35−83%). Thus, all three duplex structures serve as models for intermediates in translesional synthesis, underscoring the critical role played by cellular DNA polymerases in determining mutagenic events that takes place during replication in mammalian cells (Moriya *et al.*, 1994; Shibutani *et al.*, 1996). Similar conclusions apply to our studies of the exocyclic adduct, 1,N^2-propano-2′-deoxyguanosine (Moriya *et al.*, 1994) and the oxidized base, 8-oxo-2′-deoxyguanosine (Shibutani *et al.*, 1991).

ACKNOWLEDGMENT

We thank Mr. Robert Rieger and Ms. Cecilia Torres for the preparation and purification of modified oligodeoxynucleotides. Atomic coordinates of the NMR refined structures have been deposited in the Protein Data Bank.

SUPPORTING INFORMATION AVAILABLE

Tables listing: $J_{H1'-H2'}$ and $J_{H1'-H2''}$ coupling constant values (Table 1S); proton chemical shift differences between ϵC·dG − ϵC·dA and ϵC·dG − ϵC·T duplexes (Table 2S and Table 3S, respectively); RMSD of atom positions among distance-refined models (Table 4S) and comparison of experimental and refined distance bounds (Table 5S). Figures showing one-dimensional proton spectra of the ϵC·dG duplex recorded in D_2O buffer, pH 6.9, 25 °C (Figure 1S) and in H_2O buffer, pH 6.9, 2 °C (Figure 3S). Contour plots of an expanded region of a phase-sensitive COSY spectrum recorded in D_2O buffer, pH 6.1, 13 °C (Figure 3S), of a phase-sensitive NOESY spectrum (150 ms mixing time) recorded in H_2O buffer, pH 6.1, 2 °C (Figure 4S) and of a phase-sensitive COSY45 spectrum recorded in D_2O buffer, pH 6.9 at 27 °C (Figure 5S). Three-dimensional structures of the central five base pair segment of the distance-refined models (Figure 6S) (11 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*, No. 70, pp 345−358, International Agency for Research on Cancer, Lyon, France.

Barbin, A., Besson, F., & Perrard, M. (1985) *Mutat. Res. 152*, 147−156.

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

Bax, A., Ikura, M., Torchia, D. A., & Tschudin, R. (1990) *J. Magn. Reson. 86*, 304−318.

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

Brooks, B., Bruccoleri, R., Olafson, B., States, D., Swaminathan, S., & Karplus, M. (1983) *J. Comput. 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.

Cullinan, D., Korobka, A., Grollman, A. P., Patel, D. J., Eisenberg, M., & de los Santos, C. (1996) *Biochemistry 35*, 13319−13327.

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.

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

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

Hang, H., Chenna, A., Rao, S., & Singer, B. (1996) *Carcinogenesis 17*, 155−157.

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.

Korobka, A., Cullinan, D., Cosman, M., Grollman, A. P., Patel, D. J., Eisenberg, M., & de los Santos, C. (1996) *Biochemistry 35*, 13310−13318.

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.

Kusmierek, J. T., & Singer, B. (1992) *Chem. Res. Toxicol. 5*, 634−638.

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.

Plum, G. E., & Breslauer, K. J. (1994) *Ann. NY Acad. Sci. 726*, 45−56.

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

Rinkel, L. J., & Altona, C. (1987) *J. Biomol. Struct. Dyn. 4*, 621−649.

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

Saparbaev, M., Kleibl, K., & Laval J. (1995) *Nucleic Acids Res. 23*, 3750−3755.

Scherer E., Winterwerp, H., & Emmelot, P. (1986) *The role of Cyclic Nucleic Acid Adducts in Carcinogenesis and Mutagenesis*, No. 70, pp 109−125, International Agency for Research on Cancer, Lyon, France.

Shibutani S., Takeshita M., & Grollman A. P. (1991) *Nature 349*, 431−434.

Shibutani, S., Suzuki, N., Matsumoto, Y., & Grollman, A. P. (1996) *Biochemistry 35*, 14992−14998.

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

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

Singer, B., & Spengler, S. J. (1986) *The Role of Cyclic Nucleic Acid Adducts in Carcinogenesis and Mutagenesis*, No. 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) *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.

BI9705725