Solution Conformation of an Abasic DNA Undecamer Duplex d(CGCACXCACGC)·d(GCGTGTGTGCG): The Unpaired Thymine Stacks Inside the Helix†

Yannick Coppel, Nathalie Berthet, Christian Coulombeau, Christiane Coulombeau, Julian Garcia,* and Jean Lhomme

LEDSS, Chimie Bioorganique, UMR CNRS 5616, Université Joseph Fourier, BP 53, 38041 Grenoble Cedex 9, France

Received October 25, 1996; Revised Manuscript Received February 20, 1997⊗

ABSTRACT: The three-dimensional structural analysis of DNA undecamer $^{5'}d(C_1G_2C_3A_4C_5X_6C_7A_8C_9G_{10}C_{11})^{3'}$, $^{3'}d(G_{22}C_{21}G_{20}T_{19}G_{18}T_{17}G_{16}T_{15}G_{14}C_{13}G_{12})^{5'}$ duplex in which the X residue is a modified abasic site [3-hydroxy-2-(hydroxymethyl)tetrahydrofuran] has been performed using NOESY, DQFCOSY, TOCSY, and $^{31}P-^{1}H$ HSQC−TOCSY spectra in relation with molecular dynamics simulations. A total of 249 distances and 224 dihedral angles were used for construction. The optimal distances were calculated using the complete relaxation matrix method from hybrid matrices which were built with the experimental NOE intensities and additional data derived from either standard A- or B-DNA. Six independent refined structures starting from canonical A- and B-DNA were determined on the basis of the NMR data, and all converged to a single family with average rms deviations below 0.6 Å and final NOE R^x factors of 0.055 ± 0.03. A satisfactory agreement was obtained between measured NOE intensities and those resulting from full relaxation matrix calculations. A single intrahelical form of right-handed DNA duplex is observed; the aromatic base of residue T17 opposite the abasic site is stacked inside the helix. No clear correlation was detected between the C5 and C7 residues, excluding their proximity and the looping out of the abasic site. The abasic site induces a kink of about 30° in the DNA duplex. This kink allows the formation of a bifurcated hydrogen bond between the amino protons of C5 and the O4 oxygen of T17. A detailed analysis of the final structures and their comparison with previous studies of abasic site lesions are described.

The loss of a nucleic base resulting from the breakage of the N-glycosidic bond is a frequent event in cells. This may occur by spontaneous hydrolysis, and from *in vitro* studies, it has been estimated that in humans about 10 000 abasic sites are produced per cell cycle by purine glycosyl bond hydrolysis (Lindahl & Nyberg, 1972).

Abasic site generation (apurinic and apyrimidinic, i.e. AP[1] sites) is markedly increased by chemical modification (or degradation) of the nucleic bases which favors the hydrolytic process or activates the DNA repair systems. In this case, highly specific glycosylases release the modified base, leaving a deoxyribose moiety (Loeb & Preston, 1986; Wallace, 1988; Lindahl, 1993; Demple & Harrison, 1994). If unrepaired, abasic sites become a potentially mutagenic or even lethal lesion (Boiteux & Laval, 1982; Loeb & Preston, 1986). Thus, abasic site repair is a critical cellular activity for cell survival. This process is initiated by enzymatic cleavage at abasic sites by various types of specific enzymes. Class II AP endonucleases cleave the phosphodiester bond on the 5′-side of abasic sites, leaving a 5′-phosphate and a 3′-OH terminus (Demple & Harrison, 1994). The repair is then completed by the release of the 5′-abasic residue, DNA repair synthesis, and ligation (Demple & Harrison, 1994). The strand cleavage at the 3′-side of the AP site is catalyzed by AP lyases (Bailly & Verly, 1989), which proceed via a syn β-elimination reaction (Manoharan *et al.*, 1988b; Mazumder *et al.*, 1991).

This process leaves an unsaturated aldehyde residue which is refractory to DNA polymerases. In this case, generating a suitable terminus requires a 3′-repair diesterase (or a δ-elimination reaction followed by the action of 3′-phosphatase) (Demple & Harrison, 1994). The effect of unrepaired abasic sites on DNA replication has been investigated. During the translesional process, the incorporation of dA by DNA polymerases occurs in 90% of the cases (Loeb & Preston, 1986; Demple & Harrison, 1994). At the present time, the reason for this preference is still unknown. Some evidence for a kinetic effect (Cai *et al.*, 1993) suggests that the local conformation of DNA at the abasic site may play an important role. This hypothesis suggests the need for precise conformational studies of abasic sites at the oligonucleotide level.

In naturally occuring AP sites, a dynamic equilibrium exists between the cyclic hemiacetal and the open chain aldehyde form of the two deoxyribose residues. The aldehydic form has a 1% abundance (Manoharan *et al.*, 1988a; Wilde *et al.*, 1989) and is easily subject to a

† This work was supported by the Association pour la Recherche sur le Cancer (ARC) and the Ligue Nationale contre le Cancer (LIGUE).

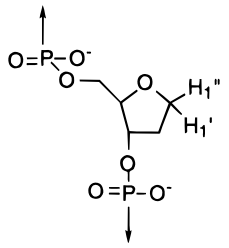
* Author to whom all correspondence should be addressed.

⊗ Abstract published in *Advance ACS Abstracts,* April 1, 1997.

[1] Abbreviations: AP, apurinic and apyrimidinic; m, multiplet; s, singlet; δppm, chemical shift in parts per million; EDTA, ethylenediaminetetraacetic acid; JUMNA, junction minimization of nucleic acids; NOE, nuclear Overhauser enhancement; NOESY, two-dimensional nuclear Overhauser enhancement experiment; DQFCOSY, double-quantum-filtered correlation spectroscopy; TOCSY, total homonuclear correlated spectroscopy; HSQC, heteronuclear single-quantum coherence; rMD, restrained molecular dynamics; MARDIGRAS, matrix analysis of relaxation for discerning geometry in solution; CORMA, complete relaxation matrix analysis; Sta-A, standard A-DNA; Sta-B, standard B-DNA; RS, restrained structures; rmsd, root mean square difference.

S0006-2960(96)02677-3 CCC: $14.00

β-elimination reaction leading to strand scission (Weiss & Grossman, 1987). This intrinsic instability of the abasic site represents a major hindrance for studying the structural and biological consequences of the presence of the lesion in the DNA strand. An elegant approach developed by several groups that avoids this instability is replacement of the 2-deoxyribose moiety with a chemically stable 3-hydroxy-2-(hydroxymethyl)tetrahydrofuran (structure **1**) (Millican *et al.*, 1984; Takeshita *et al.*, 1987). Takeshita *et al.* (1987) have shown that oligodeoxynucleotides containing this modified tetrahydrofuran group act as substrates for endonuclease IV and exonuclease III, the major AP endonucleases involved in repair in *Escherichia coli*, and also serve as effective templates for AuV reverse transcriptase and other DNA polymerases of eukaryotic and prokaryotic origin (Takeshita *et al.*, 1987; Randall, 1987). Such studies demonstrate that this analog can serve as a stable structural analog of 2′-deoxyribose in the AP site.



Structure 1

Reports concerning different NMR studies of abasic site-containing DNA duplex structures exist. These studies have established that oligonucleotides containing AP sites opposite purine residues, i.e. apyrimidinic sites, conserve a regular right-handed DNA geometry (Cuniasse *et al.*, 1987, 1990; Kalnik *et al.*, 1988, 1989a; Withka *et al.*, 1991; Goljer *et al.*, 1995) in which the base opposite the lesion stacks inside the helix. On the other hand, oligonucleotides containing AP sites opposite pyrimidine residues, i.e. apurinic sites, display additional forms where the abasic sugar and, sometimes, the opposite pyrimidine residue are extrahelical (Cuniasse *et al.*, 1990; Singh *et al.*, 1994).

Little is known about the mode of recognition of the abasic sites by AP endonuclease. Preliminary experiments indicate that cleavage at abasic sites present in synthetic oligonucleotides by exonuclease III is dependent upon the nature of the abasic site (Berthet, 1994). For a better understanding of the role of AP site-flanking sequences on helix structure, it is of great importance to precisely determine the structural features associated with this factor, including contributions from the base opposite the lesion and from the sequence of the flanking base pairs. In the present study, we focus on these conformational properties and perform a refined solution structure of an apurinic site-containing oligonucleotide, with an original DNA sequence:

$$^{5'}d(C_1\ G_2\ C_3\ A_4\ C_5\ X_6\ C_7\ A_8\ C_9\ G_{10}\ C_{11})^{3'}$$

$$^{3'}d(G_{22}\ C_{21}\ G_{20}\ T_{19}\ G_{18}\ T_{17}\ G_{16}\ T_{15}\ G_{14}\ C_{13}\ G_{12})^{5'}$$

The AP site X6, which is the chemically stable "tetrahydrofuran analog", lies in the middle of the oligomer flanked by two cytosines. The opposite thymine T17 is flanked by two guanines. Furthermore, three G·C base pairs at each end of the oligomer were added to improve its stability.

MATERIALS AND METHODS

Oligonucleotide Synthesis. The AP site analog and the corresponding 5′-dimethoxytrityl derivative were prepared as described (Takeshita *et al.*, 1987). The corresponding phosphoramidite was synthesized as follows. The 5-*O*-(4,4′-dimethoxytrityl)-1,2-dideoxy-D-ribofuranose (200 mg, 0.5 mmol) was dissolved in dry dichloromethane (5 mL) in the presence of dry diisopropylamine (415 μL, 4.3 mmol).

β-Cyanoethyl (*N*-diisopropylamino)chlorophosphite (168 μL, 0.67 mmol) was added dropwise over 5 min through a syringe under argon at room temperature with stirring. After 30 min, the reaction mixture was poured into dichloromethane (80 mL) and a solution of sodium hydrogenocarbonate (5%, 20 mL) was added. The mixture was then extracted three times with dichloromethane (100 mL), and the combined extracts were dried on sodium sulfate. After evaporation of the solvent *in vacuo*, the residual yellow oil was purified by chromatography on a column (silica gel; 60/35/5 dichloromethane/hexane/triethylamine), giving a pale yellow oil with 50% yield (155 mg, 0.25 mmol). Physical data: ^{1}H NMR (200 MHz, $CDCl_3$) δppm 7.50−6.85 (m, 13H, aromatics), 4.45 (m, 1H, C_3H), 4.10 (m, 2H, C_1Ha and C_1Hb), 4.08 (m, 1H, C_4H), 3.80 and 3.79 (2s, 6H, 2 × OCH_3), 3.69−3.60 [m, 2H, 2 × $CH(CH_3)_2$], 3.16 (m, 2H, C_5H), 2.59−2.45 (m, 2H, C_2H), 2.17 (m, 2H, C_1H), 1.19 [m, 12H, 2 × $CH(CH_3)_2$]; MS (DCI, NH_3-isobutane) [MH^+] calculated for $C_{35}H_{45}N_2O_6P$ 620.7, observed 621.

The oligomers were synthesized on a 15 μmol scale using the phosphoramidite chemistry on a Milligen/Biosearch 3700 DNA synthesizer. The DMT protective group was removed on the machine. After deprotection by heating in ammonium hydroxide (28%, 5 mL) at 50 °C for 48 h, the oligonucleotides were purified by anion-exchange chromatography on a Mono QHR 5/5 column (Pharmacia). The sample was eluted with a linear gradient of potassium chloride (from 0 to 0.6 M in 50 min) in a sodium acetate buffer (20 mM, pH 6) containing 20% acetonitrile.

The pure oligonucleotides were then desalted on Waters Sep-Pak cartridges by washing with distilled water and eluting with a mixture of water/acetonitrile (80/20). After evaporation to dryness, the oligomers were dissolved in the NMR buffer.

Sample Preparation. The heteroduplexes were formed by mixing equimolar amounts of each strand, determined on the basis of their molar extinction coefficients at 260 nm. The 1/1 stoichiometry of the duplex was confirmed by proton NMR spectroscopy. The concentration of the double-stranded DNA sample was determined to be 1.5 mM in 0.65 mL. The duplex was heated at 80 °C for 5 min, cooled to room temperature over several hours to ensure complete annealing, and then lyophilized. The undecamer was dissolved in pH 7.0 buffer containing 20 mM sodium phosphate, 0.1 M sodium chloride, and 0.1 mM EDTA, then lyophilized two times, and finally redissolved in 0.65 mL of 99.96% D_2O. For NMR experiments involving exchangeable protons, the oligomer was dissolved in 90% H_2O/10% D_2O.

NMR Experiments. NMR experiments were performed on a Varian Unity 600 spectrometer and on a Varian Unity *Plus* 500 spectrometer at temperatures of 10, 15, and 30 °C.

Chemical shifts were referenced relative to the residual HOD resonance, which was previously calibrated to 3-(trimethylsilyl)propionate-2,2,3,3-d_4 (TSP-d_4). The two-dimensional NOESY, DQFCOSY, TOCSY, and ^{31}P−^{1}H HSQC−TOCSY spectra were acquired in the phase-sensitive mode utilizing the hypercomplex method (States *et al.*, 1982). The transmitter frequency was set on the residual HOD resonance, which was presaturated during the relaxation delay and during the mixing time for the NOESY. The NOESY spectra for the oligonucleotide in D_2O were recorded with mixing times of 30, 50, 75, 150, 250, and 400 ms. Sixty-four scans for each of the 400 t_1 values were collected with 2048 points. The FID along t_1 was zero-filled to 1024 data points prior to Fourier transformation to give a final (2048 × 1024) data matrix.

In the DQFCOSY spectrum, 64 acquisitions were obtained with 4096 points for each of the 400 t_1 increments. The FID was zero-filled to 1024 points in F_1 to give a final (4096 × 1024) data matrix.

The TOCSY spectra were collected at mixing times of 40 and 90 ms. In these experiments, 64 acquisitions with 2048 points were collected for the 300 t_1 values. The data were zero-filled in F_1 to give a final (2048 × 1024) data matrix.

To study the exchangeable protons, a NOESY spectrum was recorded with a mixing time of 250 ms in 90% H_2O/10% D_2O. The first pulse was replaced by a 360 × 270 − *x* composite pulse (Freeman *et al.*, 1988), and a jump and return read pulse was used for water suppression (Plateau *et al.*, 1982).

The ^{31}P−^{1}H HSQC−TOCSY spectrum was recorded using a mixing time of 80 ms, and 2048 acquisitions with 2048 points were collected for the 40 t_1 values. The sweep widths were 474 Hz for the ^{31}P dimension and 5200 Hz for the ^{1}H dimension. The data were zero-filled to give a final (2048 × 128) data matrix.

Longitudinal relaxation times, T_1, were determined by the inversion recovery method.

Sugar Puckers. The geometry of each individual DNA sugar can be determined using the scalar-coupled NMR data. Since all five torsion angles in the sugar ring are interdependent, the overall sugar conformation can be described as a function of the pseudorotation angle, ϕ, and the pseudorotation amplitude, ϕ_m (Altona & Sundaralingam, 1972). The magnitude of the five three-bond coupling constants depends on the sugar geometry. However, it is often difficult to accurately measure these coupling constants because of spectral overlaps and line widths greater than the coupling constants. Therefore, sugar conformation is estimated through comparison of the magnitude of coupling constants from phase-sensitive DQFCOSY (Hosur *et al.*, 1986) and of the magnitude of the H1′−H4′ and H2″−H4′ distances from phase-sensitive NOESY (Chary & Modi, 1988). These sugar conformations were then used to generate five dihedral angle restraints for each sugar in the restraint molecular dynamics (rMD).

Interproton Distances. Interproton distances were derived using the MARDIGRAS program (Borgias & James, 1989, 1990) from NOESY experiments with mixing times of 50, 75, and 150 ms performed at 600 MHz. For the three intensity sets, a hybrid two-dimensional NOE relaxation matrix was set up, in which the interproton NOEs not available experimentally were computed from the geometry of the starting structure. Starting models for the MARDIGRAS calculations were standard A-DNA and standard B-DNA (Arnott *et al.*, 1980). MARDIGRAS calculates upper and lower bounds for the interproton distances (Borgias & James, 1990). Due to possible errors at the different stages (peak integration routine), the smallest range between the lower and the upper bounds was imposed at 10% of the average value.

Structural Calculations. Initial starting structures were generated by constructing the d(CGCACACACGC)·d(GCGTGTGTGCG) duplex in canonical A and B forms and then replacing the base of the A6 residue by the hydrogen H1″ using the Biopolymer module of the Biosym Insight II software. These unminimized structures are designated as Sta-A and Sta-B. Calculations were performed and structures visualized on an IBM RS 6000-39H computer running Discover version 2.95 and Insight II version 2.3.5 (Biosym/Molecular Simulations).

The empirical energy function used was AMBER (Weiner *et al.*, 1986), and the NOE and dihedral restraining functions used were standard square-well effective potentials. Calculations were performed *in vacuo* with a distance-dependent dielectric function of $\epsilon_r = 4r$. Na^+ counterions were placed at the bifurcating positions of the O−P−O angles at a distance of 5.0 Å to neutralize the phosphate groups.

In the simulations, Sta-A and Sta-B were subjected to the protocol described below with appropriate restraint weights which were modulated by multiplying the force constants with a scaling factor. In addition to the experimentally determined interproton distance restraints and sugar ring dihedral restraints, two other types of restraints were used. To prevent structural artifacts in the ill-defined backbone region, we employed torsion angle restraints to maintain right-handedness except at the abasic site (Baleja *et al.*, 1990). Additionally, base pairs were kept hydrogen bonded in the Watson−Crick mode, and distance restraints between the bases were used according to the literature (Saenger, 1984). The range for the hydrogen bond torsions was set between 170 and 190° for the inner base pairs and between 160 and 200° for the terminal base pairs to take the frayed ends into account.

The starting structures were first subjected to 2000 steps of restrained energy minimization, with NOE and empirical base pair distance force constants set to 50 kcal mol^{-1} Å^{-2} and dihedral angle force constants set to 25 kcal mol^{-1} rad^{-2}. Distance restraints were weighted by 100, 80, 60, 40, and 20% of the applied NOE force constant in the function of the quality of the calculated distance. The quality of distances was judged using the deviations observed for the distances coming from multiple MARDIGRAS calculations and from the estimation of errors in the volume measurements due to overlap or a low signal-to-noise ratio (Schmitz *et al.*, 1991). The initial velocities were assigned with a Maxwell distribution at 600 K. The system was gradually heated from 0 to 600 K during the first 2 ps and then maintained at 600 K for the next 40 ps. The time step of the integrator was set to 1 fs. During this phase of the simulation, force constants were gradually increased by a factor of 2 in the first 10 ps and then held constant at this level during the high-temperature phase. The system was slowly cooled to 300 K for 30 ps and then maintained at 300 K for an additional 40 ps. During the cooling and equilibrium phases, distance and dihedral angle force constants were gradually decreased to their initial value. The

last 10 coordinate sets, arising from the last 10 ps, were averaged and subjected to a final 2000 steps of restrained energy minimization to generate the restrained structures.

Trial runs were performed to refine the restraints. The coordinates of two restrained structures obtained from Sta-A and Sta-B structures were taken as input data for the program CORMA (Keepers & James, 1984). This program calculates the dipole−dipole relaxation matrix for a system of protons and converts them into a NOESY spectrum. The back-calculated two-dimensional NOE spectra were compared with the experimental (50, 75, and 150 ms) two-dimensional NOE spectra to determine how well the model structure fits the experimental data. Since NOE signals are proportional to r^{-6}, a small error in the distance between two protons can lead to a large change in the intensity of the cross-peak for these two protons. A deviation of a factor of 2.5 (±15% experimental error in distance) between the volumes of the calculated cross-peaks and the experimental values was considered to be a violation. In order to minimize the differences between calculated and experimental NOESY spectra, the distances corresponding to these violations were progressively adjusted to the experimental values through further refinement of the corresponding new sets of NOE restraints.

The AP site region of the duplex is not well-defined by NMR restraints. Trial runs showed that local structural perturbations in a MD simulation could result in unusual interproton contacts in the abasic site region (observed in the back-calculated spectra) that were not experimentally observed. To prevent such artifacts, we used "non-NOE constraints" (Metzler *et al.*, 1990; Wang *et al.*, 1994). For these protons, a minimal distance of 4.5 Å for proton−proton distances was imposed.

It is important to note that the range between the lower and upper bounds of distance constraints was limited to 10% of the average value. Tighter distance constraints were not used since the size of a cross-peak is also influenced by spin diffusion.

Calculations with these refined restraints were carried out from the two starting structures with three different initial velocity seeds for each, yielding six restrained structures: RS-A, RS-A1, RS-A2, RS-B, RS-B1, and RS-B2. The "crystallographic" R^c factor and the R^x factor (James, 1991) were used to measure the fit of a restrained structure to NOE data. R^c is defined by $R^c = \Sigma_i||I_o(i)|-|I_c(i)||/\Sigma_i|I_o(i)|$ and R^x by $R^x = \Sigma_i|I_o^{1/6}(i)-I_c^{1/6}(i)|/\Sigma_i|I_o^{1/6}(i)|$, where $I_o(i)$ is the experimental and $I_c(i)$ is the corresponding calculated intensity of the cross-peak i for a particular structure. These six restrained structures were averaged to generate the final structure RS-ave.

Structural Analysis. The calculations of all helical parameters were carried out with the program CURVES version 5.0 (Lavery & Sklenar, 1990) which is especially suited to the characterization of regular or irregular DNA structural features (Lavery & Sklenar, 1988, 1989).

RESULTS

The first step in determining the conformational parameters of an oligonucleotide in solution is the assignment of proton resonances. These were assigned in a sequential manner using the well-established procedure developed for right-handed DNA duplexes (Feigon *et al.*, 1983; Scheek *et al.*, 1983; Wüthrich, 1986). This strategy is based on the relative proximity of the glycosidic protons (H1′, H2′, H2″, and H3′) to the base protons of the same nucleotide and of their 3′-neighbor residues.

Nonexchangeable Proton Assignments. The DQFCOSY spectrum was used to identify the sugar protons in combination with the 75 ms mixing time NOESY spectrum. H2′ and H2″ were distinguished from one another by their coupling pattern with H1′ in DQFCOSY spectra (Figure S1 of the Supporting Information). The H1′−H2″ cross-peaks were generally characterized by a more complex coupling pattern than the H1′−H2′ cross-peaks. Furthermore, the H2′ proton resonated upfield to the H2″ proton of the same residue, except for the 3′-terminal residue, where the H2′ and H2″ had identical chemical shifts (e.g. C11) or where the order was reversed (e.g. G22). They were confirmed by a more detailed examination of the cross-peaks in the NOESY spectrum since the H1′ proton is closer to H2″ than to H2′ for all sugar conformations. The assigments of H3′ and H4′ and H5′ and H5″ were obtained through the concerted use of DQFCOSY, TOCSY, and NOESY spectra and are listed in Table 1.

An expanded NOESY spectrum which displayed connectivities correlating H8/H6 to H2′−H2″ is plotted in Figure 1. For illustration purposes, the sequential interresidue interactions are linked by continuous solid lines from C11 to C7 (Figure 1A) on the modified strand and from T15 to T19 (Figure 1B) on the other strand. Adenine H2 protons were readily distinguished from the other base protons in a one-dimensional inversion−recovery experiment (Figure S2 of the Supporting Information) due to their long relaxation time. Their assignments were performed by observing H2−H1′ cross-peaks with their own sugar and the sugar on the 3′-side.

In the strand containing the abasic site, a break was detected in the sequential assignment reflecting the missing base between C5 and C7. A less intense NOE between T17-Me and G16−H8 was consistent with T17 stacking into the duplex (Figure 1B). Furthermore, we observed other weaker NOEs than normal between the H8 proton of G18 and the H2′ and H2″ protons of T17, indicating structural perturbation at the abasic site (Figure 1B). However, no additional cross-peak indicative of an equilibrium with an extrahelical conformer could be observed even at a temperature of 30 °C, as was reported for the d(CGTGXGTGC)·d(GCACTCACG) sequence (Cuniasse *et al.*, 1990).

The proton resonances of the abasic tetrahydrofuran residue were readily assigned from the NOESY and DQFCOSY spectra, and their chemical shifts were comparable to those reported in the literature for this tetrahydrofuran moiety in different oligonucleotides (Kalnik *et al.*, 1988; Cuniasse *et al.*, 1987, 1990). The abasic H1′ and H1″ protons resonating at 3.99 and 3.90 ppm, respectively, were distinguished on the basis of cross-peak intensities between them and the abasic H3′ protons and the C7 base protons H6 and H5 in the 75 ms NOESY. All cross-peak intensities were consistent with predominantly S-type sugar conformations and with a right-handed helix, typical of a B-like DNA conformation.

Exchangeable Proton Assignments. The two-dimensional NOE spectrum acquired in an H_2O solution (Figure S3 of the Supporting Information) displayed two well-resolved thymidine imino protons at 13.91 and 13.81 ppm. They were

Table 1: Proton and ^{31}P Chemical Shift Assignments for the GTG Sequence[a]

Residue	H8	H6	H5/Me/H2	H1′	H2′	H2″	H3′	H4′	H5′/H5″	imino[b]	amino[c]	^{31}P[d]
C1		7.65	5.92	5.75	2.01	2.43	4.73	4.09	3.73		na/na	−2.91
G2	8.01			5.93	2.70	2.76	5.01	4.38	4.12	13.13		na
C3		7.41	5.49	5.67	2.07	2.42	4.87	4.22	na[h]		8.44/6.59	−2.84
A4	8.28		7.82	6.24	2.69	2.89	5.03	4.42	na			na
C5		7.32	5.25	5.92	2.20	2.20	4.86	4.21	na		8.10/6.57	−2.05
X6				3.99/3.90[e]	1.99	1.99	4.58	4.14	3.97/3.89[f]			−2.47
C7		7.66	5.91	5.56	1.98	2.40	4.83	4.22	na		8.35/6.99	−2.87
A8	8.36		7.84	6.18	2.75	2.87	5.02	4.41	4.12			na
C9		7.28	5.33	5.59	1.91	2.30	4.81	4.12	na		8.25/6.67	−2.83
G10	7.88			5.95	2.60	2.73	5.00	4.35	na	13.14		−2.84
C11		7.39	5.31	6.16	2.21	2.21	4.51	4.06	na		na/na	
G12	7.99			6.00	2.63	2.80	4.87	4.27	3.72	na		
C13		7.45	5.39	5.78	2.16	2.46	4.89	4.23	4.14		8.50/6.62	−3.10
G14	7.98			6.01	2.66	2.78	5.00	4.39	na	12.92		na
T15		7.16	1.51	5.79	1.90	2.26	4.89	4.24	4.14	13.92		na
G16	7.84			5.91	2.64	2.64	5.00	4.35	4.09	12.82		−2.71
T17		7.39	1.58	5.98	2.06	2.29	4.73	4.32	na	10.86[g]		−3.03
G18	7.94			5.98	2.71	2.81	4.94	4.41	na	12.73		−2.74
T19		7.33	1.41	5.87	2.14	2.50	4.90	4.22	na	13.82		−2.75
G20	7.93			5.90	2.64	2.69	5.00	4.39	4.11	12.82		−2.75
C21		7.38	5.45	5.78	1.92	2.35	4.84	4.20	4.11		8.50/6.69	−2.76
G22	7.97			6.17	2.64	2.39	4.70	4.09	na	na		

[a] H_2O is referenced at 4.97 ppm. Proton assignments are at 10 °C and pH 7 (at 600 MHz). [b] Assignments of H1 imino protons of guanine and of H3 imino protons of thymine. [c] Assignments of hydrogen-bonded amino protons and exposed amino protons of cytosine. [d] ^{31}P chemical shifts are referenced to H_3PO_4 at 15 °C. [e] Assignments of H1′ and H1″ protons of the abasic site. [f] Not differentiated. [g] Broad resonance. [h] Not assigned.

identified by the existence of a NOE between thymine H3 and adenine H2 protons of the complementary base. Six partially resolved guanine imino protons were detected between 12.7 and 13.1 ppm and were assigned through their characteristic NOEs with hydrogen-bonded and exposed cytidine amino protons as well as with the nonexchangeable H5 protons. The two terminal base imino protons were not observed as a result of "end fraying". The cross-peak patterns observed for all the imino protons were consistent with Watson−Crick base pairing, even for the base pairs flanking the AP site. A broad resonance was observed at 11.0 ppm for a temperature of 0 °C (Figure S4 of the Supporting Information) which disappeared when raising the temperature above 15 °C, suggesting a fast exchange with the solvent. This signal was assigned to the T17 non-base-paired imino proton. Furthermore, its chemical shift is similar to that reported in the literature for a nonpaired thymine present as a single base bulge in an oligonucleotide, a situation that shows some similarity with the abasic site (van den Hoogen *et al.*, 1992). Exchangeable proton assignments are given in Table 1.

The melting profile of the imino protons is represented in Figure 2. The exchange of the imino protons with the solvent became more rapid above 45 °C, leading to broadening of the signals that were totally absent at 50 °C in our experimental conditions. At 5 °C, The imino protons at 13.12 and 13.23 ppm were tentatively assigned to those of the terminal base pairs. They disappeared completely at a temperature of 15 °C due to rapid imino exchange with solvent via fraying. The imino protons of the G16 and G20 residues were overlapped at temperatures below 15 °C but were resolved for higher temperatures. When the temperature was increased, the imino proton of the G16 residue (indicated by an asterisk for temperatures of 25 and 35 °C) broadened before all the other imino protons (except those of the terminal base pairs). These data indicate that the C7•-G16 base pair, adjacent to the abasic site, exchanges more

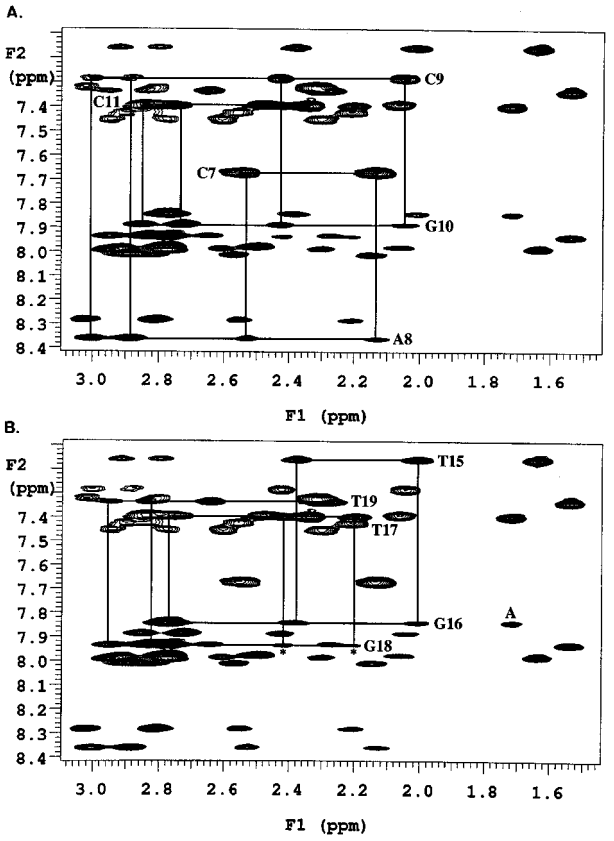


FIGURE 1: H8/H6/H2 to sugar H2′/H2″/Me regions of the 250 ms NOESY spectrum of d(CGCACXCACGC)·d(GCGTGTGTGCG) undecamer at 10 °C in 0.1 M NaCl, 20 mM sodium phosphate, and 0.1 mM EDTA at pH 7.0 dissolved in 99.96% D_2O. (A) Sequential connectivity is shown between C7 and C11 residues in the modified strand. (B) Sequential connectivity is shown between T15 and T19 residues in the complementary strand. Note that cross-peaks between G18-H8 and T17-H2′ and -H2″ (denoted by asterisks) and between T17-Me and G16-H8 (A) are of lower intensity than other comparable cross-peaks.

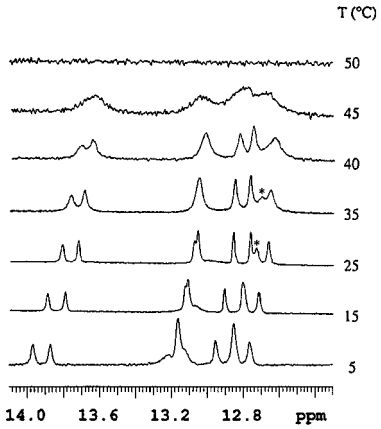


FIGURE 2: Melting profile of the imino protons of the d(CGCACXCACGC)·d(GCGTGTGTGCG) sequence acquired in H_2O from 0 to 50 °C. The asterisks denote the imino proton of the G16 residue at 25 and 35 °C.

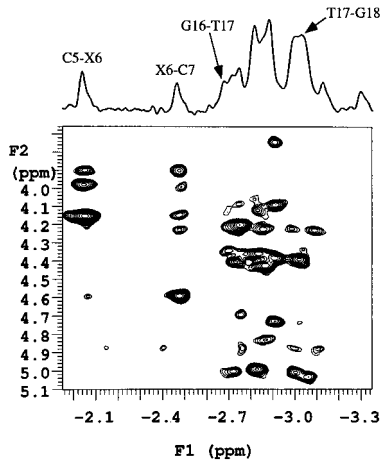


FIGURE 3: Expanded H3′, H4′, and H5′/H5″ region of two-dimensional $^{31}P-^{1}H$ HSQC−TOCSY spectrum of the undecamer at 15 °C. ^{31}P chemical shifts are reported relative to H_3PO_4. Only the most important signals are indicated on the ^{31}P one-dimensional spectrum displayed above the two-dimensional spectrum.

rapidly with the solvent than the other base pairs of the undecamer. This phenomenon was also observed to a lesser extent for the C5·G18 base pair which seemed to have an imino lifetime similar to that of the A·T base pairs.

Phosphorus Assignments. ^{31}P assignments were obtained from the two-dimensional $^{31}P-^{1}H$ HSQC−TOCSY spectrum at 15 °C and are given in Table 1. Assignment of the *i*th phosphate was achieved through connectivities with H3′ (*i*), H4′ (*i* + 1), H5′/H5″ (*i* + 1), and sometimes H2′/H2″ (*i*). The expanded H3′, H4′, and H5′/H5″ region of the two-dimensional $^{31}P-^{1}H$ HSQC−TOCSY spectrum is shown in Figure 3. The intensities of the $^{31}P-$H5′/H5″ cross-peaks generally appear to be much weaker than the those of $^{31}P-$H4′ cross-peaks as described in other ^{31}P heteronuclear experiences (Gorenstein, 1994).

All ^{31}P resonances appeared in the chemical shift range associated with the B-form DNA (between −2.7 and −3.1 ppm) except for two phosphorus resonances that were shifted downfield. The resonance at −2.05 ppm has been assigned to the phosphorus between the C5 and X6 residues given its correlations with H4′ and H5′/H5″ protons of the X6 residue. Surprisingly, its correlation with the H3′ proton of the C5 residue is very weak. The resonance at −2.47 ppm was easily assigned to the phosphorus between the X6 and C7 residues thanks to its correlations with the H1′/H1″, H2′/H2″, H3′, and H4′ protons of the X6 residue and with the H4′ proton of the C7 residue. The phosphoruses flanking the thymine T17 opposite the AP site are less pertubed as they appear at −3.03 and −2.71 ppm. These results suggest a distorsion of the backbone localized at the abasic lesion.

Sugar Puckers. A qualitative approach for determining deoxyribose sugar puckers was adopted because of the presence of appreciable overlaps and resonance line widths close to or greater than the coupling constants measured in the DQFCOSY spectrum. The magnitude of the coupling constants is proportional to the intensity of the cross-peaks (Hosur *et al.*, 1986). Through analysis of the two coupling constants most sensitive to sugar pucker, ^{3}J(H2″,H3′) and ^{3}J(H3′,H4′), the pseudorotation angles of the individual sugar puckers were estimated. The G2, A4, G10, and G18 residues do not show detectable H3′−H4′ cross-peaks (Figure S5 of the Supporting Information) with ϕ ranging from 162 to 180°. The A8, G12, G14, G16, and G20 residues show weak but detectable H3′−H4′ cross-peaks in combination with nonexistent or very weak H2″−H3′ cross-peaks, which is characteristic of a C2′ endo geometry ($\phi = 140-162°$). For the C1, C3, C5, X6, C7, C9, C11, T15, C21, and G22 residues, ϕ ranges from 98 to 144° (C1′ exo) with respect to strong intensities of H3′−H4′ cross-peaks and nonexistent or very weak H2″−H3′ cross-peaks. The C13, T17, and T19 residues exhibit medium-intensity H3′−H4′ cross-peaks with nonexistent H2″−H3′ cross-peaks constraining ϕ in the 120−162° range. These values were confirmed by comparing ^{3}J(H1′,H2′) and ^{3}J(H1′,H2″) and by estimating the H1′−H4′ and H2″−H4′ distances (Chary & Modi, 1988).

These values for the pseudorotation angles were converted into individual sugar torsion angles, ν_j, using the relationship $\nu_j = \phi_m \cos[\phi+144(j-2)]$, where j is 0−4. These torsion angles were also used in the restrained molecular dynamics calculations.

Two-Dimensional NOE Intensity Analysis. Interproton distances were obtained from two-dimensional NOE cross-peak intensities using the MARDIGRAS program (Borgias & James, 1989, 1990). Distances to rapidly rotating methyl protons were calculated using the three-site jump model, which has been demonstrated to be a good approximation (Liu *et al.*, 1992). Experimental cross-peak intensities, correlation times, and coordinates for an initial structure are the necessary input data for MARDIGRAS calculations.

Assuming isotropic motion, a single correlation time τ_c of 5.0 (±0.5) ns was estimated for the whole molecule. This value was estimated by measuring the initial buildup rate for several H6−H5 vectors as indicated by Reid *et al.* (1989). The value of 5.0 ns as the isotropic correlation time was used to calculate the two-dimensional NOE spectra.

The MARDIGRAS calculations were performed for the three two-dimensional NOE data sets with mixing times of 50, 75, and 150 ms. The numbers of cross-peak intensities were 167, 172, and 230 for the 50, 75, and 150 ms spectra, respectively. Standard B-DNA and standard A-DNA were used as starting structures, and calculations were performed with three correlation times ($\tau_c = 4.5$, 5.0, and 5.5 ns). Some restraints that were inconsistent with all others were removed as it appeared that the inconsistency was due to the lack of precision arising from spectral overlap or from peaks with a signal-to-noise ratio that was too low. The final combined set of distances from all three NOE data sets provided 191

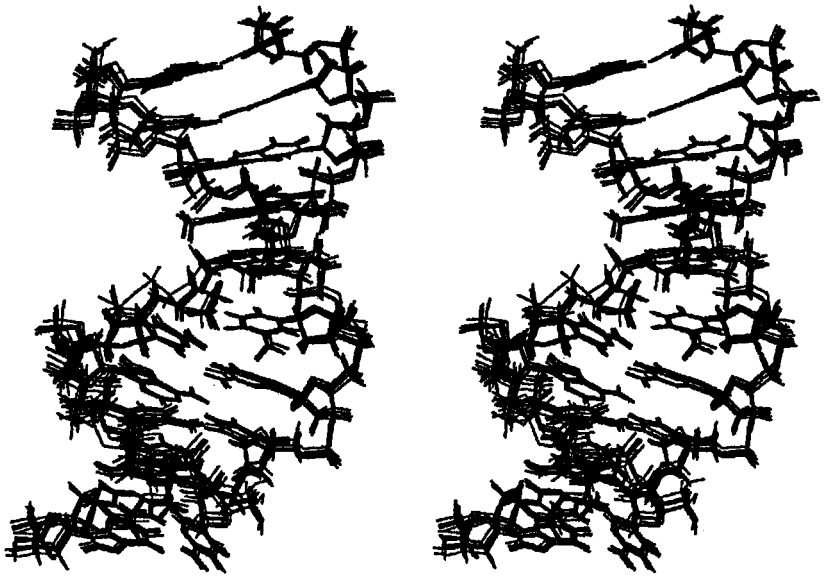


FIGURE 4: Stereoview of the superposition of the final refined structures RS-A, RS-A1, RS-A2, RS-B, RS-B1, and RS-B2.

Table 2: Atomic rms Differences[a] and R^c and R^x Factors[b]

	rmsd (Å)									
	Sta-B	RS-A	RS-A1	RS-A2	RS-B	RS-B1	RS-B2	RS-ave	$R^{c\ factor}$	$R^{x\ factor}$
Sta-A	4.94	2.78	2.55	2.87	2.63	2.60	2.49	2.73	0.74	0.157
Sta-B		2.75	3.07	2.64	2.93	2.95	3.08	2.79	0.84	0.126
RS-A			0.40	0.35	0.26	0.26	0.42	0.23	0.28	0.052
RS-A1				0.58	0.29	0.18	0.23	0.36	0.29	0.058
RS-A2					0.44	0.44	0.58	0.24	0.29	0.058
RS-B						0.19	0.20	0.25	0.27	0.052
RS-B1							0.22	0.27	0.28	0.056
RS-B2								0.38	0.29	0.055
RS-ave									0.28	0.053

[a] rmsd between starting structures (Sta-A and Sta-B) and refined structures (RS-A, RS-A1, RS-A2, RS-B, RS-B1, and RS-B2) and the average restrained structure RS-ave, excluding the end base pairs. [b] Average value of crystallographic-like NMR R^c factor and sixth-root-weighted NMR R^x factor of three mixing times (50, 75, and 150 ms).

distance restraints. It is important to note that no precise distance could be obtained between the C5 and X6 residues due to the absence of base protons at the AP site.

Structural Calculations. Restrained molecular dynamics simulations were performed using these experimental interproton distance restraints, 105 sugar dihedral restraints as well as 90 restraints which preserve the right-handed nature of the helix and 56 restraints which preserve its base pairing. The calculations were carried out from two starting structures; one is a canonical A-type conformation (Sta-A) and the other a canonical B-type conformation (Sta-B) (rms difference = 4.94 Å). Partial fraying occurred for the terminal base pairs, and consequently, an accurate rMD structure could be assumed only for the inner nine base pairs.

The AP site region is not well-defined by the NMR restraints because of the absence of base protons at the AP site which prohibited precise measurement of distances between the C5 and X6 residues. Local structural perturbations at the AP site in the MD simulations predicted some short proton−proton contacts (i.e. between C5 and X6 or C7 residues) that were not experimentally observed. This led us to include additional non-NOE restraints.

Calculations including the additional restraints and the manually modified restraints were performed until agreement between simulated and experimental NOEs was achieved. The refined restraints were applied on the different rMDs initiated from the Sta-A and Sta-B structures.

Atomic deviation between the final structures RS-A, RS-A1, RS-A2, RS-B, RS-B1, and RS-B2 (Figure 4) was less than 0.6 Å. The atomic rsmd values for the starting models and for the restrained structures are displayed in Table 2. The convergence is driven by the experimental restraints and not by the empirical force field as indicated by trial runs that were performed without experimental restraints and which did not converge (rmsd ≥ 3.0 Å). The value for rmsd between the starting structures and the final structures suggests that the latter displayed significant differences with standard A- and B-DNA.

The accuracy of the structures can be judged by comparison with the experimental two-dimensional NOE data. We calculated the theoretical two-dimensional NOE spectra for the restrained structures by complete relaxation matrix analysis using the CORMA program. We compared the agreement between the theoretical and experimental two-dimensional NOE intensities via residual indices R^c and R^x. Deviations from the three NOESY spectra (mixing times of 50, 75, and 150 ms) were calculated using an isotropic molecular correlation time of 5.0 ns and were then averaged. These R factors show how well the NOESY volumes back-calculated from a particular structure fit the experimental volumes. Because of the r^{-6} dependence of NOE intensities, the crystallographic R^c factor is dominated by errors for shorter distances; the $^1/_6$-weighted function used to calculate the R^x factor gives greater weight to longer range NOE

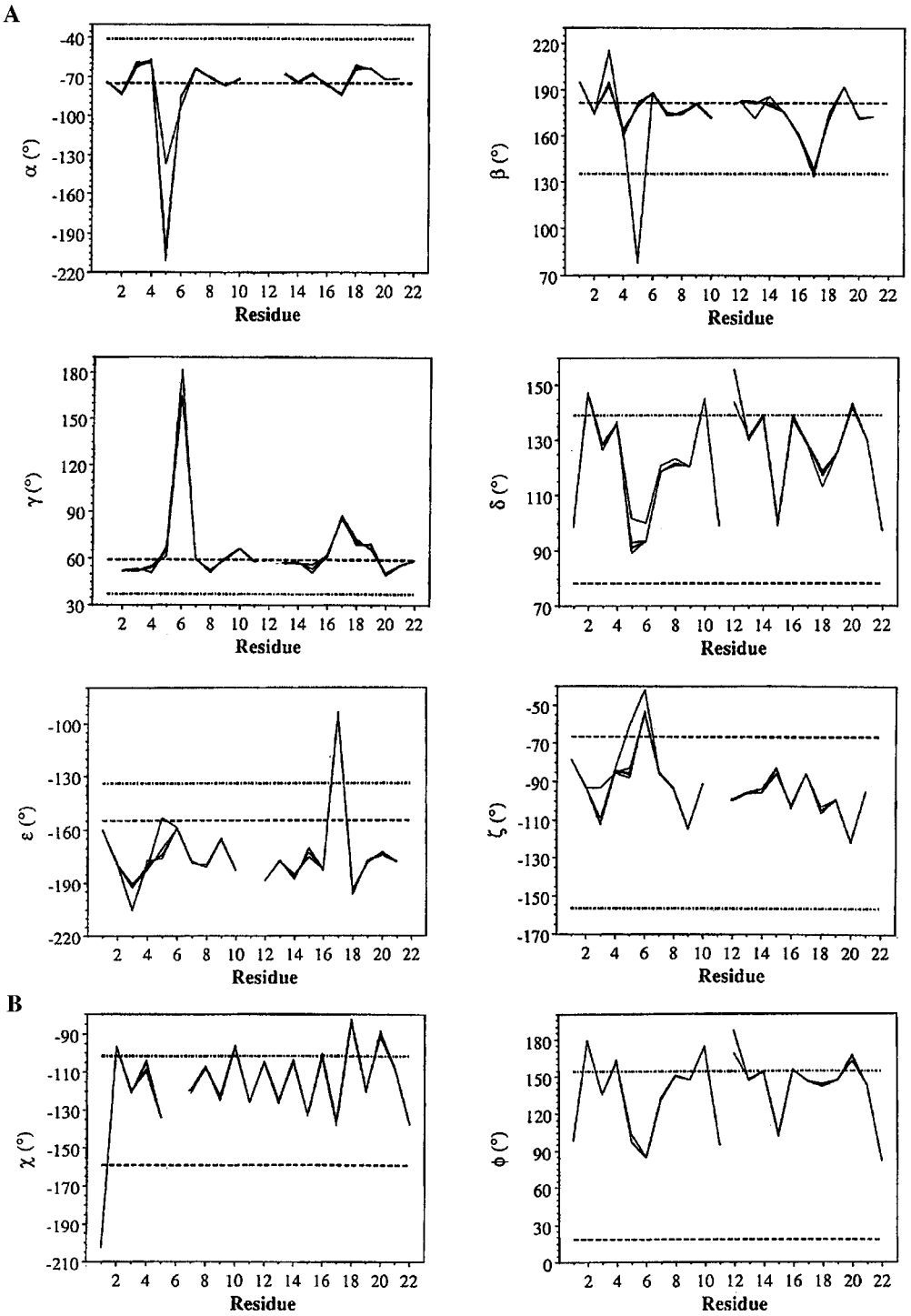


FIGURE 5: Plot of torsion angles and sugar puckers for the RS-A, RS-A1, RS-A2, RS-B, RS-B1, and RS-B2 structures (—) in comparison with standard A- (- - -) and B-DNA (- · -) (Arnott *et al.*, 1980). Parameter values were calculated using CURVES (Lavery & Sklenar, 1990): (A) backbone torsion angles (α−ζ) and (B) glycosidic torsion angles (χ) and sugar pseudorotation angles (ϕ).

interactions (James, 1991). Table 2 gives the *R* factors for Sta-A, Sta-B, RS-A, RS-A1, RS-A2, RS-B, RS-B1, and RS-B2. As can be seen for the R^x factor, the experimental values fit slightly better with the canonical B-type structure than with the canonical A-type structure. The significant reduction in the *R* factors after restrained molecular dynamics treatment indicates that all restrained structures are in excellent agreement with the experimental data (Table 2).

Structural Analysis. All backbone torsion angles, sugar conformations, and helical parameters were calculated with the CURVES algorithm (Lavery & Sklenar, 1990). The convergence to the same structure, illustrated by rmsd values and NOE *R* factors, is also seen in detail for the relationships between the base pairs.

(a) Backbone Torsion Angles. In the best final NOE fit structures, the backbone torsion angles are in the domain expected for right-handed structures. Calculated values for the backbone torsion angles are shown graphically in Figure 5 which also displays the values for canonical A- and B-form DNA (Arnott *et al.*, 1980). These angles converge well for the restrained structures except for the C5 residue (see Discussion). The values of angles α, β, and γ agree with the A-form canonical values. For angles δ, ζ, and ϵ, there is poor agreement with both the A- and B-form canonical

values. In several cases, large relative changes are observed which correspond to the site of the abasic lesion. The torsion angle α decreases to a value between −130 and −210° for the C5 residue (ill-defined). The angle β shows substantial changes for the T17 and perhaps for the C5 residues that adopt respectively the values 135 and between 80 and 180° (ill-defined). The angle γ shows a value around 170° for X6 and 85° for T17. The torsion angle ϵ increases to −95° for the T17 residue, and the torsion angle ζ adopts a value around −55° for X6.

(b) Glycosidic Torsion Angles and Sugar Puckers. The glycosidic torsion angles are well-defined by the NOE data. All values are in the antidomain except for residue G16 (Figure 5).

All sugar puckers are found in the S range (180 ± 90°) except for residues X6 and G22. The sugar puckers converge well for the restrained structures (Figure 5). With the exception of nucleotides C1, C5, X6, C11, T15, and G22, the sugar pseudorotation angles ϕ agree well with the canonical B-form DNA. The X6 sugar pucker adopts a value of 84° intermediate between S and N ranges. Due to important spectral overlaps, we could not determine if this value arises from an equilibrium mixture between the C2′ endo and C3′ endo sugar conformations (Rinkel & Altona, 1987). The terminal residues C1, C11, and G22 have sugar puckers between those of A- and B-type DNA; such puckers at terminal residues are commonly observed in DNA solution structures.

(c) Helical Parameters. Among the axis-base pair parameters, the values for *X*-displacement and *Y*-displacement (data not shown) converge well from both the Sta-A and Sta-B structures. *X*-displacement values range from −1.1 to −2.1 Å, typical of B-like structures (Poncin *et al.*, 1992). *Y*-displacement values are positive, between 0.3 and 1.3 Å. The base pair inclination values in restrained structures (Figure 6A) are closer to the canonical B-DNA value (inclination = 1.5°) than to the canonical A-DNA value (inclination = 20.7°). Tip values do not agree well with either canonical A- or B-forms and range from 5 to −7° (Figure 6A).

The values for selected intra-base pair parameters as a function of base pair position are shown in Figure 6B. These parameters are in good agreement for the restrained structures, and the C5·G18 base pair shows the most pronounced deviations. All structures exhibit rather large shears in opposite directions between base pairs except between C3·G20 and A4·T19 and C7·G16 and A8·T15. Buckle, propeller twist, and opening do not agree well with either canonical A-form or B-form DNA. Base pairs C7·G16 and A8·T15 give rise to a "positive cup" (corresponding to a negative buckle followed by a positive buckle in the 5′−3′ direction). Perturbation observed in the propeller twist of base pair C7·G16 is probably due to the absence of a nucleic base at the AP site.

The plots of the calculated inter-base pair parameters are shown in Figure 6C. All the parameters converge well from both Sta-A and Sta-B structures. Shift values are contained between −1.0 and 0.6 Å and slide values between −0.6 and 0.4 Å (data not shown). The rise per base pair consistently is refined to a value of ∼3 Å, midway between that for A-form and B-form DNA except for base pairs C3·G20 and C7·G16 which show rise values close to that of the A-form DNA. Negative tilt angles are observed between base pairs A4·T19 and C5·G18 and C7·G16 and A8·T15, and twist angles decrease to a value of 27° between base pairs C7·G16 and A8·T15, illustrating kinking of the helix at the AP site. Negative tilt angles combined with positive roll are also observed between base pairs C3·G20 and A4·T19 and C9·G14 and G10·C13.

DISCUSSION

Refined Structures. The precision of the final structures can be judged from the convergence to identical conformations starting from the two different Sta-A and Sta-B structures that are derived from fiber diffraction coordinates (Arnott *et al.*, 1980). The rmsd values, which range between 0.18 and 0.58 Å for different restrained structures, indicate that good convergence has been reached. The structures converged less satisfactorily at the abasic site due to the absence of base protons, which prohibited precise measurement of distances between the C5 and X6 residues. Thus, the relative positions of the latter were principally determined by non-NOE restraints and by the molecular force field. Notably, large differences in the backbone torsion angles of the C5 and X6 residues (principally, the torsion angle α and β of the C5 residue) are observed for one restrained structure relative to the five others. These discrepancies could also result from a more important mobility at this site so that different conformations could be in accord with the experimental data.

The final set of structures represents a well-defined region of conformational space clearly different from the initial A and B structures, as indicated by the average rmsd between the final and the initial structures (2.68 and 2.82 Å). In terms of constraint violations, a considerable improvement is apparent for the refined structures when compared with the starting structures (data not shown). Only one violation of distance among the 249 distance constraints greater than 0.2 Å and three violations of dihedral angles among the 224 dihedral constraints greater than 10° are found for the refined structures.

The accuracy of the final structures is more difficult to judge than their precision. The present structures were generated with approximations, particulary in the treatment of solvent and in the force field. Furthermore, calculations were made with nonexperimental restraints (planarity, right-handedness) and assuming an isotropic motion. Finally, all NMR data represent time-averaged phenomena. These limitations are recurrent in most of the structures of DNA duplexes determined by NMR. With this in mind, the accuracy of the final structures was measured by comparing the calculated intensities of the model structure with the experimental NOE intensities. The refined structures were considered to be accurate as judged by using sixth-root *R*-factor analysis (James, 1991) (Table 2). The observed R^x values of 0.055 ± 0.03 indicate that the structures are consistent with the measured NOE data within reasonable experimental error.

The data for the terminal base pairs should be interpreted cautiously due to fraying. However, we were principally interested in the abasic site region and in the immediately adjacent base pairs. In conclusion, the convergence of the results starting from the two different initial geometries strongly validates the final structure but cannot rigorously exclude the fact that other geometries may fit the experimental data.

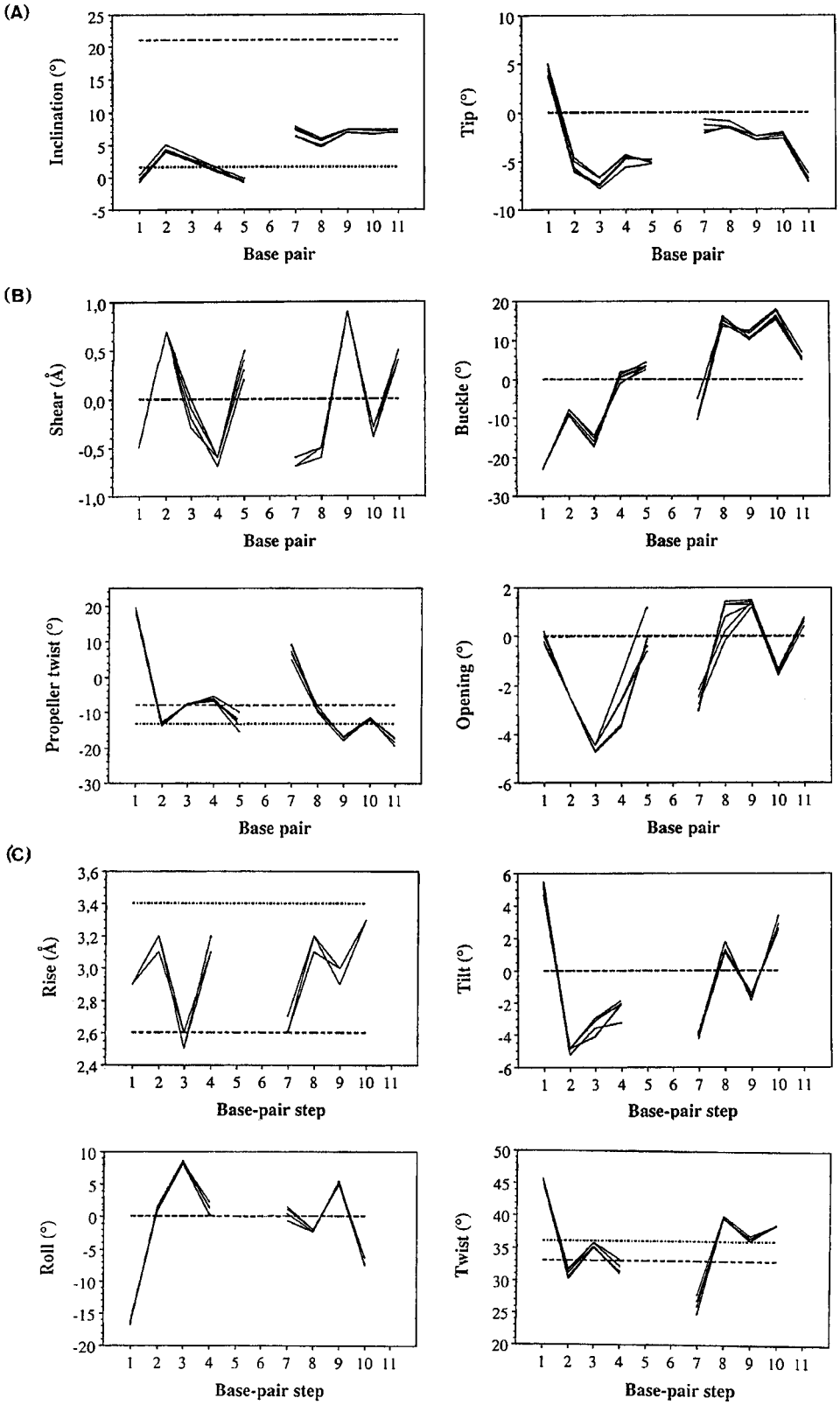


FIGURE 6: Selected helical parameters for the RS-A, RS-A1, RS-A2, RS-B, RS-B1, and RS-B2 structures (—) in comparison with canonical A- (- - -) and B-DNA (- • -) (Arnott *et al.*, 1980). Parameter values were calculated using CURVES (Lavery & Sklenar, 1990): (A) axis-base pair parameters, (B) intra-base pair parameters, and (C) inter-base pair parameters. Gaps are observed at the AP site.

T17 Stacks into the Helix. These studies establish that T17 positioned opposite the AP site remains inside the helix under the conditions of our NMR experiments but with some structural perturbations compared to canonical A- and B-DNAs. A slide value of 1.3 Å, a tilt angle of −9°, a roll angle of −24°, and a twist angle of 12° are shown between T17 and G18 residues, and a twist angle of 60° is shown between G16 and T17 residues. This is illustrated in Figures 7 and 8 which show the T17 residue located beneath the C5•G18 base pair (Figure 8A) and above the C7•G16 base pair in the RS-ave structure (Figure 8B). This position of T17 is stabilized by the formation of a weak bifurcated hydrogen bond (Taylor *et al.*, 1984) between the O4 of T17 and the amino protons of C5 (2.6 and 2.8 Å).

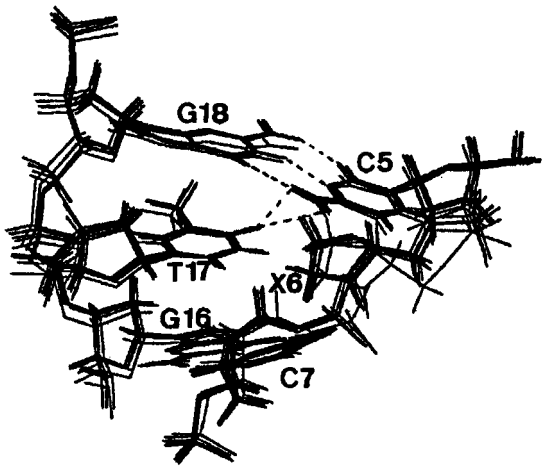


FIGURE 7: Expanded region of the AP site for the six restrained structures RS-A, RS-A1, RS-A2, RS-B, RS-B1, and RS-B2. Dashed lines represent hydrogen bonds. Note the bifurcated hydrogen bond between the O4 of the T17 residue and the amino protons of the C5 residue (2.6 and 2.8 Å). The kink is illustrated by the rotation of the C5 residue with respect to the C7 residue.

AP Site X6. The position of the sugar of the AP site is difficult to determine precisely since very few distances between the C5 and the X6 residues could be measured and no distance between the X6 and the C7 residues could be obtained. However, the calculations indicate that the sugar of the AP site lies partly in the minor groove as can be seen in Figures 7 and 8. This perturbed position is validated by the downfield-shifted resonances of the two phosphoruses, on the 3′- and 5′-sides of the AP site.

Furthermore, the presence of the abasic site destabilizes the adjacent base pairs which show a higher rate of imino proton exchange with the solvent than the other base pairs of the undecamer (except for the terminal base pairs). This effect is particularly pronounced for the C7·G16 base pair.

Sugar Puckers and Backbone Torsion Angles. In the best final NOE fit structures, most of the sugar rings adopt intermediate conformations between C1′ exo and C2′ endo except at C5, X6, and T15 and at the terminal residues C1, C11, and G22. All the glycosidic torsion angles are in the *anti* domain except for the G16 residue, and the backbone torsions appear in the domain expected for right-handed structures (Figure 5). Some significant changes are seen at the AP site. Backbone torsion angles other than δ are generally not well-defined by experimental constraints due to a limited number of protons in this region.

DNA Kinking. The AP site induces a kink in the DNA duplex, as observed in Figure 7, dividing the undecamer into two distinct stacked base-paired regions. The first one, designated as the "5′-part" relative to the position of the AP site, comprises base pairs C1·G22−C5·G18. The second, designated as the "3′-part", comprises base pairs C7·G16−C11·G12. The T17 residue is stacked with the 3′-part. The 5′-part is deflected by an angle of ∼30° with respect to the 3′-part, creating bending into the major groove (Figure 7). As well as essentially all NMR structure determinations of DNA duplexes, these results represent only the average conformation of the abasic site-containing DNA undecamer in solution. Results of molecular dynamics simulations indicate a high degree of flexibility and mobility in the structure of DNA (Miaskiewicz *et al.*, 1993). Abasic sites could be an important point of flexibility of the DNA helix axis. Additional studies to determine if DNA could be very flexible over a wide range of radii of curvature at the abasic site are underway.

The calculations may provide insight into how abasic site-induced kinking is produced. Substantial changes are seen for the torsion angle ζ of the X6 residue, for the torsion angles β and ϵ of the T17 residue, and for the torsion angle γ of these two residues (Figure 5) following the formation of a kink. Pertubations observed in the inter- and intra-base pair helical parameters (Figure 6) also play an important role in the formation of the kink. The kink goes together with a negative roll angle of −24°, a low twist angle of 12° between the T17 and G18 residues, and a high twist angle of 60° between the G16 and T17 residues. Reduction of the twist between base pairs C7·G16 and A8·T15 is also observed. Furthemore, base pairs A4·T19 and C5·G20 and base pairs C7·G16 and A8·T15 have negative tilts.

The 3′- and 5′-parts are also slightly bent as illustrated by positive rolls between base pairs C3·G20 and A4·T19 and between base pairs C9·G14 and G10·C13, by negative tilts between these base pairs and between base pairs G2·C21 and C3·G20, and by large negative and positive buckles on both sides of the AP region. In addition, this kink allows the formation of a weak bifurcated hydrogen bond between the amino protons of C5 and the O4 of T17.

Comparison of the d(CGCACXCACGC)·d(GCGTGTGTGCG) Sequence with Previous Studies of AP Site Lesions. We have performed an extensive study of an undecamer containing an apurinic site in the middle of the sequence, making use of combined NMR−molecular dynamics methods. Clear conclusions could be drawn concerning the influence of this lesion upon the conformation of the DNA helix. It is interesting to compare these results to the data reported in previous studies (Cuniasse *et al.*, 1987, 1990; Kalnik *et al.*, 1988, 1989a; Withka *et al.*, 1991; Singh *et al.*, 1994; Goljer *et al.*, 1995).

Most of this work has been devoted to the examination of oligodeoxyribonucleotides containing apyrimidinic sites (Cuniasse *et al.*, 1987, 1990; Kalnik *et al.*, 1988, 1989a; Withka *et al.*, 1991; Goljer *et al.*, 1995), which can be conveniently prepared by removal of uracil using uracil-DNA glycosylase. In all the systems described, the purine facing the abasic site stacks inside the helix independently of the flanking bases. This has been observed in duplexes in which the abasic site is the true aldehydic abasic site or the stable tetrahydrofuran analog. On the other hand only a few reports have appeared concerning the most frequent apurinic lesion, and the situation turns out to be more complex. Cuniasse *et al.* (1990) examined nonamers in which the model tetrahydrofuran apurinic site is located at the center of the sequence and is flanked by purine residues on both sides, i.e. GXG. When deoxycytidine faces the abasic site, the cytosine residue is expelled from the helix. On the other hand, when deoxythymidine is opposite the lesion, an equilibrium exists between the intra- and extrahelical location of the thymine ring. This equilibrium is sensitive to temperature. At low temperatures, the unpaired thymine stacks inside the helix. When the temperature is raised, the equilibrium is shifted toward the extrahelical conformation. Slight differences of the behavior of the dT residue have been reported by Singh *et al.* (1994) for the same d(CGTGXGTGC)·d(GCACTCACG) sequence, but in which a true AP site had been incorporated. These discrepancies presumably arise from differences in the nature of the abasic

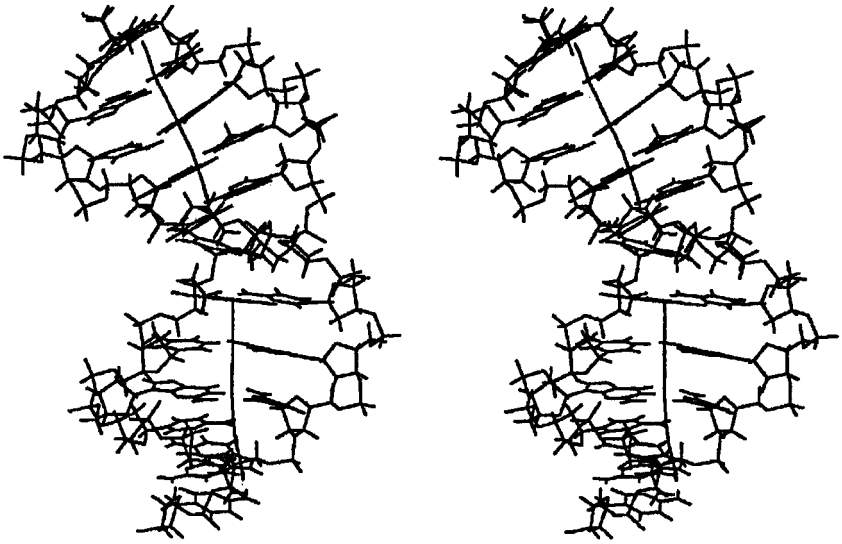


FIGURE 8: Stereoview of the average restrained structure RS-ave with the global helix surperimposed.

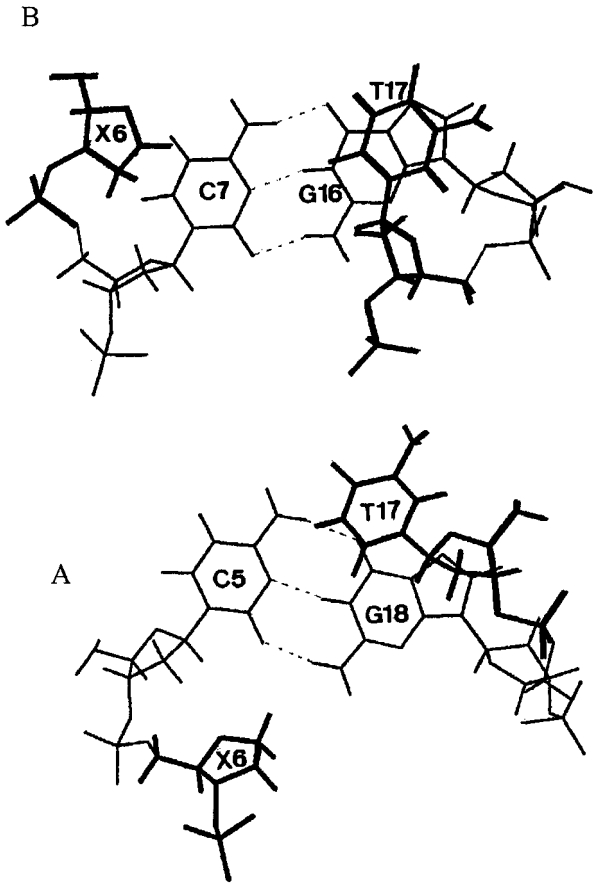


FIGURE 9: Base stacking in the RS-ave structure. The AP site and the opposite T17 residue are shown in bold lines. (A) Position of the AP site and the T17 residue relative to the last base pair C5•-G18 of the 5′-part. (B) Position of the AP site and the T17 residue relative to the first base pair C7•G16 of the 3′-part. Note that the T17 residue stacks well with the G16 residue but not with the G18 residue.

sugar residue and also in the experimental conditions used for the NMR experiments.

In the present study, we examined an apurinic duplex in which the apurinic site is flanked by two pyrimidines, i.e. CXC. Direct comparison can thus be made with the results obtained by Cuniasse *et al.* (1990) to evaluate the influence of the flanking bases upon the structure of the lesion. The NMR spectrum of the undecamer containing the CXC sequence did not exhibit, at any temperature, any signal that could be attributed to an extrahelical location of the thymine ring that faces the abasic site. It thus appears that the structure of the duplex at the apurinic site depends upon the nature of the flanking bases. A rationale for these observations could be ascribed to the differences in base−base stacking interactions involved in the different cases. Purine−purine stacking interactions have long been known to be stronger than purine−pyrimidine and pyrimidine−pyrimidine interactions. When the apurinic site is flanked by two purine residues as is the case for the GXG triplet present in the nonamer studied by Cuniasse *et al*. (1990), displacement of the base opposite X outside the helix is probably driven by the tendancy of the two "distant" guanines to approach one another and stack, with the net result being contraction of the abasic site. The gain of energy due to $\pi-\pi$ stacking of the guanines is probably high enough to disrupt the opposite stacked triplet of pyrimidine bases CTC or CCC. On the other hand, when the apurinic site is flanked by two pyrimidine bases (CXC), the tendency of the two cytosines to stack is not high enough to compete with and disrupt the opposite GTG triplet, a situation that is observed for the present undecamer. The thymine residue flanked by the two guanines in the strand opposite the lesion thus remains stacked inside the helix.

At this stage of the discussion, some analogies can be drawn between abasic sites in DNA and single base bulges, as in each case an extra base must be accommodated within the duplex. Detailed examination of the results in the literature shows a comparable behavior concerning notably the influence of the base opposite the lesion and the nature of the flanking bases. It has been shown that single purine bases in the bulge stack into duplex DNA independently of the flanking sequence (Patel *et al.*, 1982; Woodson & Crothers, 1988; Kalnick *et al.*, 1989b) as observed for abasic sites. The situation is more complex for extra pyrimidines, and it appears that pyrimidine rings can adopt looped-out and stacked-in conformations depending on the flanking sequence and on the temperature (Kalnick *et al.*, 1989c, 1990; Van den Hoogen *et al*., 1988). The equilibrium is shifted toward the stacked-in conformation when the extra pyrimidine is flanked by purines as observed here for the abasic site. In addition, kinking of DNA by base bulges has been

described (Lilley, 1995).

Biological Implications. This study performed with the best available methodology, combining NMR spectroscopy and molecular dynamics calculations, has provided new results concerning the molecular structure near an abasic site lesion of DNA. Overall conformation and local deformations are unambiguously determined, and the unpaired base is proved to be located inside the helix. However, the thymine opposite the lesion is slightly shifted in comparison with paired thymines in standard position, and the O4 atom of this thymine is hydrogen bonded to the amino protons of the C5 residue. Indeed, in the present duplex, if base stacking is retained along a right-handed helical structure, the small shift of the unpaired base and, more importantly, the kink appearing near the lesion and the concomitant alteration of the phosphodiester backbone indicate the flexibility of the system at the lesion site. The nature of the flanking sequence and of the unpaired base could consequently modify reactivity of the site by increasing or reducing recognition by repair enzymes. This could explain the sensitivity of exonuclease III to the nature of the base opposite the abasic site (Berthet, 1994).

As already noted (Kalnick *et al.*, 1988), the duplex DNAs are capable of retaining a right-handed helical structure after removal of a base, leaving a cavity that could provide a recognition site for the AP endonucleases. In the following paper, we report the binding of a synthetic AP endonuclease mimic (of the AP lyase type) to the undecamer described here. The molecule, which exhibits very high efficiency, as it cleaves abasic sites in DNA at nanomolar concentrations is constituted by a purine (adenine or diaminopurine) linked to an intercalator (aminochloromethoxyacridine) through a polyamino chain. This molecule binds selectively at the abasic site of the undecamer using the purine as the recognition unit to dock into the apurinic site pocket and hydrogen bonds to the opposite thymine in the duplex.

ACKNOWLEDGMENT

We are grateful to C. Maerschalk and Professor J. Reisse (Université Libre de Bruxelles) for assistance with NMR spectroscopy (600 MHz NMR machine) and for fruitful discussions. We thank Professor R. Lavery for critical reading of the manuscript.

SUPPORTING INFORMATION AVAILABLE

Five figures showing (1) the H1′ to H2′/H2″ region of the DQFCOSY spectrum, (2) the one-dimensional inversion recovery experiments with delays of 0.8 and 6.4 s between pulses, (3) the imino to amino, H2 and H5 region of the 250 ms NOESY spectrum acquired in H_2O, (4) the exchangeable proton region of the one-dimensional spectrum acquired at 0 °C in H_2O, and (5) the H3′ to H4′ region of the DQFCOSY spectrum (5 pages). Ordering information is given on any current masthead page.

REFERENCES

Altona, C., & Sundaralingam, M. (1972) *J. Am. Chem. Soc. 94*, 8205−8211.

Arnott, S., Chandrasekaran, R., Birdsall, D. L., Leslie, A. G. W., & Ratliff, R. L. (1980) *Nature (London) 283*, 743−745.

Bailly, V., & Verly, W. (1989) *Nucleic Acids Res. 17*, 3617−3618.

Baleja, J. D., Pon, R. T., & Sykes, B. D. (1990) *Biochemistry 29*, 4828−4839.

Belmont, P., Boudali, A., Constant, J.-F., Demeunynck, M., Fkyerat, A., Lhomme, J., Michon, P., & Serratrice, G. (1997) *New J. Chem. 21*, 47−54.

Berthet, N. (1995) Ph.D. Thesis, University Joseph Fourier Grenoble I, Grenoble, France.

Berthet, N., Boudali, A., Constant, J.-F., Decout, J.-L., Demeunynck, M., Fkyerat, A., Garcia, J., Laayoun, A., Michon, P., & Lhomme, J. (1994) *J. Mol. Recognit. 7*, 99−107.

Boiteux, S., & Laval, J. (1982) *Biochemistry 21*, 6746−6751.

Borgias, B. A., & James, T. L. (1989) *Methods Enzymol. 176*, 169−183.

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

Cai, H., Loom, B., Eritja, R., & Goodman, M. F. (1993) *J. Biol. Chem. 268*, 23567−23572.

Chary, K. V. R., & Modi, S. (1988) *FEBS Lett. 233*, 319−324.

Constant, J.-F., O'Connor, T., Lhomme, J., & Laval, J. (1988) *Nucleic Acids Res. 16*, 2691−2703.

Constant, J.-F., Fkyerat, A., Demeunynck, M., Laval, J., O'Connor, T. R., & Lhomme, J. (1990) *Anti-Cancer Drug Des. 5*, 59−62.

Cuniasse, Ph., Sowers, L. C., Eritja, R., Kaplan, B., Goodman, M. F., Cognet, J. A. H., Lebret, M., Guschlbauer, W., & Fazakerley, G. V. (1987) *Nucleic Acids Res. 15*, 8003−8022.

Cuniasse, Ph., Fazakerley, G. V., Guschlbauer, W., Kaplan, B., & Sowers, L. C. (1990) *J. Mol. Biol. 213*, 303−314.

Demple, B., & Harrison, L. (1994) *Annu. Rev. Biochem. 63*, 915−948.

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

Fkyerat, A., Demeunynck, M., Constant, J.-F., Michon, P., & Lhomme, J. (1993a) *J. Am. Chem. Soc. 115*, 9952−9959.

Fkyerat, A., Demeunynck, M., Constant, J.-F., & Lhomme, J. (1993b) *Tetrahedron 49*, 11237−11252.

Fouilloux, L., Berthet, N., Coulombeau, C., Coulombeau, Ce., Dheu-Andries, M.-L., Garcia, J., Lhomme, J., & Vatton, P. (1995) *J. Mol. Struct. 330*, 417−422.

Freeman, R., Friedrich, J., & Xi-Li, W. (1988) *J. Magn. Reson. 79*, 561−567.

Goljer, I., Kumar, S., & Bolton, P. H. (1995) *J. Biol. Chem. 270*, 22980−22987.

Gorenstein, D. G. (1994) *Chem. Rev. 94*, 1315−1338.

Hosur, R. V., Ravikumar, M., Chary, K. V. R., Shteth, A., Govil, G., Zun-Kun, T., & Miles, H. T. (1986) *FEBS Lett. 205*, 71−76.

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

Kalnick, M. W., Chang, C. N., Grollman, A. P., & Patel, D. J. (1988) *Biochemistry 27*, 924−931.

Kalnick, M. W., Chang, C. N., Johnson, F., Grollman, A. P., & Patel, D. J. (1989a) *Biochemistry 28*, 3373−3383.

Kalnick, M. W., Norman, D. G., Swann, P. F., & Patel, D. J. (1989b) *J. Biol. Chem. 264*, 3702−3712.

Kalnick, M. W., Norman, D. G., Zagorski, M. G., Swann, P. F., & Patel, D. J. (1989c) *Biochemistry 28*, 294−303.

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

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

Lavery, R. (1988) in *Structure and Expression*; *Vol 3, DNA Bending and Curvature* (Olson, W. K., Sarma, R. H., Sarma, M. H., & Sundaralingham, M., Eds.) pp 191−211, Adenine Press, New York.

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

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

Lavery, R., & Sklenar, H. (1990) *CURVES 3.0, Helical Analysis of Irregular Nucleic Acids*, Laboratory for Theoretical Biochemistry, CNRS, Paris.

Lilley, D. M. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 7140−7142.

Lindahl, T. (1993) *Nature (London) 362*, 709−715.

Lindahl, T., & Nyberg, B. (1972) *Biochemistry 11*, 3610−3618.

Liu, H., Thomas, P. D., & James, T. L. (1992) *J. Magn. Reson. 98*, 163−175.

Loeb, L., & Preston, B. (1986) *Annu. Rev. Genet. 20*, 201−230.

Manoharan, M., Ransom, S. C., Mazumder, A., Gerlt, J. A., Wilde, J. A., Withka, J. M., & Bolton, P. H. (1988a) *J. Am. Chem. Soc. 110*, 1620−1622.
Manoharan, M., Mazumder, A., Ransom, S. C., Gerlt, J. A., & Bolton, P. H. (1988b) *J. Am. Chem. Soc. 110*, 2690−2691.
Mazunder, A., Gerlt, J. A., Absalon, M. J., Stubbe, J., Cunnigham, R. P., Withka, J., & Bolton, P. H. (1991) *Biochemistry 30*, 1119−1126.
Metzler, W. J., Wang, C., Kitchen, D. B., Levy, R. M., & Pardi, A. (1990) *J. Mol. Biol. 214*, 711−736.
Miaskiewicz, K., Osman, R., & Weinstein, H. (1993) *J. Am. Chem. Soc. 115*, 1526−1537.
Millican, T. A., Mock, G. A., Chauncey, M. A., Patel, T. P., Eaton, M. A., Gunning, J., Cutbush, S. D., Neidele, S., & Mann, J. (1984) *Nucleic Acids Res. 12*, 7435−7453.
Morden, K. M., Chu, Y. G., Martin, F. G., & Tinoco, I. J. (1983) *Biochemistry 22*, 5557−5563.
Patel, D. J., Kozlowski, S. A., Marky, L. A., Rice, J. A., Broka, C., Itakura, K., & Breslauer, K. J. (1982) *Biochemistry 21*, 445−451.
Plateau, P., & Gueron, M. (1982) *J. Am. Chem. Soc. 104*, 7310−7311.
Poncin, M., Hartmann, B., & Lavery, R. (1992) *J. Mol. Biol. 226*, 775−794.
Randall, S. K., Eritja, R., Kaplan, B., Petruska, J., & Goodman, M. F. (1987) *J. Biol. Chem. 262*, 6864−6870.
Reid, B. R., Banks, K., Flynn, P., & Nerdal, W. (1989) *Biochemistry 28*, 10001−10007.
Rinkel, J. L., & Altona, C. (1987) *J. Biomol. Struct. Dyn. 4*, 621−649.
Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
Scheek, R. M., Russo, N., Boelens, R., Kaptein, R., & Van Boom, J. H. (1983) *J. Am. Chem. Soc. 105*, 2914−2916.
Schmitz, U., Pearlman, D. A., & James, T. L. (1991) *J. Mol. Biol. 221*, 271−292.
Singh, M. P., Hill, G. C., Péoc'h, D., Rayner, B., Imbach, J.-L., & Lown, J. W. (1994) *Biochemistry 33*, 10271−10285.
States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286−292.
Takeshita, M., Chang, C.-N., Johnson, F., Will, S., & Grollman, A. P. (1987) *J. Biol. Chem. 262*, 10171−10179.
Taylor, R., Kennard, O., & Versichel, W. (1984) *J. Am. Chem. Soc. 106*, 244−248.
van den Hoogen, Y. T., van Beuzekom, A. A., van den Elst, H., van der Marel, G. A., van Boom, J. H., & Altona, C. (1988) *Nucleic Acids Res. 16*, 2971−2986.
Wallace, S. S. (1988) *Environ. Mol. Mutagen. 12*, 431−477.
Wang, Y., & Patel, D. J. (1994) *J. Mol. Biol. 242*, 508−526.
Weiner, S. J., Kollman, P. A., Nguyen, D. T., & Case, D. A. (1986) *J. Comput. Chem. 7*, 230−252.
Weiss, B., & Grossman, L. (1987) *Adv. Enzymol. Relat. Areas Mol. Biol. 60*, 1−34.
Wilde, J. A., Bolton, P. H., Mazumder, A., Manoharan, M., & Gerlt, J. A. (1989) *J. Am. Chem. Soc. 111*, 1894−1896.
Withka, J. M., Wilde, J. A., & Bolton, P. H. (1991) *Biochemistry 30*, 9931−9940.
Woodson, S. A., & Crothers, D. M. (1988) *Biochemistry 27*, 8904−8914.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, pp 203−255, John Wiley & Sons, New-York.

BI962677Y