Solution Structure of a DNA Duplex Containing Mispair-Aligned N^4C−Ethyl−N^4C Interstrand Cross-Linked Cytosines†

Mateus Webba da Silva,‡ Anne M. Noronha,§ David M. Noll,§,|| Paul S. Miller,§ O. Michael Colvin,‡ and Michael P. Gamcsik*,‡

Department of Medicine, Duke University Medical Center, Durham, North Carolina 27710, Department of Biochemistry and Molecular Biology, Bloomberg School of Public Health, Johns Hopkins University, 615 North Wolfe Street, Baltimore, Maryland 21205, and Department of Biophysics and Biophysical Chemistry, School of Medicine, Johns Hopkins University, 725 North Wolfe Street, Baltimore, Maryland 21205

Received June 27, 2002; Revised Manuscript Received October 26, 2002

ABSTRACT: The solution structure of an interstrand cross-linked self-complementary oligodeoxynucleotide containing directly opposed alkylated N^4C−ethyl−N^4C cytosine bases was determined by molecular dynamics calculations guided by NMR-derived restraints. The undecamer d(CGAAAC*TTTCG)$_2$, where C* represents directly opposed alkylated N^4C−ethyl−N^4C cytosine bases, serves as model for the cytotoxic cross-links formed by bifunctional alkylating agents used in cancer therapy. The structure of the duplex shows the cross-link protruding into the major groove. An increase in the diameter of the DNA at the pseudoplatform formed by the cross-linked residues creates an A-DNA characteristic hole in the central portion of the DNA. This results in a centrally underwound base step and a number of subsequent overwinding steps leading to an overall axis bend toward the major groove. The structure shows narrowing of both minor and major grooves in the proximity of the cross-link. The perturbation leads to preferential intrastrand base stacking, disruption of adjacent canonical (A·T) base pairing, and buckling of base pairs, the extent of which diminishes with progression away from the lesion site. Overall, the distortion induced by the cross-link spreads over three base pairs on the 5′- and 3′-sides of the cross-link.

Bifunctional alkylating agents such as cyclophosphamide, melphalan, bischloroethylnitrosourea (BCNU), and cisplatin are among the most widely used therapeutics in the clinical management of cancer. The cytotoxicity of these agents is thought to be due to formation of DNA interstrand cross-links which thereby prevent strand separation during transcription and replication (*1*). Recent studies have suggested that cancer cells resistant to these agents have an improved capacity to repair interstrand cross-links (*2*). Since the development of resistance is one of the primary reasons for treatment failure, a better understanding of the processes by which such DNA lesions are recognized and repaired could lead to the development of more effective alkylating agents.

A number of different mechanisms have been proposed to play a role in the repair of DNA interstrand cross-links (*3*−*11*). However, it is not clear what lesion-induced changes in the structure and/or dynamics trigger lesion recognition and the subsequent selection and implementation of appropriate repair pathways. These types of studies have been hindered by the difficulty in preparing purified, stable cross-links amenable to physicochemical and structural analysis. A few laboratories have been successful in preparing nitrogen mustard (*12*−*15*) and cisplatin (*16*−*20*) cross-links, and their results have provided valuable insight into cross-link-induced structural perturbations. However, studies delineating structural alterations and DNA repair in relation to systematic changes in both the nucleotide base sequence and the chemistry of the cross-link are lacking. By de novo synthesis of cross-linked duplexes employing solid phase methods, we have succeeded in preparing stable cross-links as models for therapy-induced lesions (*21*−*24*). This communication reports our initial investigations into the solution structure of one such model. In parallel studies, the susceptibilities of these model duplexes to recognition and repair are being evaluated.

Here we report an NMR solution structure study of the self-complementary undecamer deoxyoligonucleotide d(CGAAAC*TTTCG)$_2$ containing the N^4C−ethyl−N^4C interstrand cross-link (Figure 1a,b). This cross-link mimics the BCNU-induced G−C cross-link and the C−C cross-link that results from treatment with mechlorethamine (*25*). Preliminary studies suggest that this cross-link lesion can be repaired by *Escherichia coli* and *Saccharomyces cerevisiae*. Our results illustrate structural features that could allow recognition of this DNA duplex by different repair apparatus.

† This research was supported by a grant from the National Cancer Institute (CA082785). A.M.N. was supported in part by a postdoctoral fellowship from the Natural Sciences and Engineering Research Council (NSERC). D.M.N. was supported by grants from the Robert Leet and Clara Guthrie Patterson Trust and the Alexander and Margaret Stewart Trust.

* To whom correspondence should be addressed: MSRB 395, Box 3843, Research Drive, Durham, NC 27710. Phone: (919) 681-2244. Fax: (919) 684-5653. E-mail: michael.gamcsik@duke.edu.

‡ Duke University Medical Center.

§ Department of Biochemistry and Molecular Biology, Bloomberg School of Public Health, Johns Hopkins University.

|| Department of Biophysics and Biophysical Chemistry, School of Medicine, Johns Hopkins University.

10.1021/bi026368l CCC: $22.00
Published on Web 11/26/2002

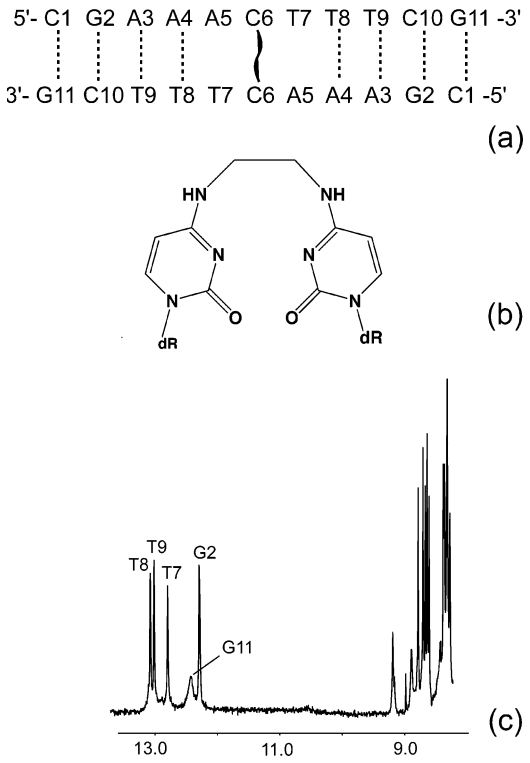

FIGURE 1: (a) Scheme of the self-complementary d(CGAAAC*TTTCG)$_2$ duplex. The dotted lines represent hydrogen bonds, and the wavy line represents the interstrand cross-link. (b) Chemical structure of cross-linked cytosines showing mutually *cis* alkylated cytosines. (c) One-dimensional ^{1}H NMR spectrum of d(CGAAAC*TTTCG)$_2$ in a ^1H_2O solution containing 100 mM NaCl (pH 7.4) showing the assigned imino proton signals. The imino proton resonance of G11 is broad due to increased water accessibility to this terminal residue.

MATERIALS AND METHODS

The synthesis by solid phase methods of short DNA duplexes that contain an N^{4}C−ethyl−N^{4}C interstrand cross-link has been described (*24*). The NMR sample (2 mM in single-strand concentration) was prepared by dissolving it in 0.3 mL of buffer containing 10 mM sodium phosphate (pH 7.4) and 100 mM NaCl.

NMR Spectroscopy. Standard NMR experiments were carried out on Varian Inova 500, 600, and 800 MHz spectrometers. We recorded NOESY (200 and 60 ms) spectra in ^1H_2O at 2 °C and NOESY (*26*) (50, 100, 150, 250, and 600 ms), DQF-COSY (*27*), TOCSY (*28*) (spin-lock time of 60 ms), and ^{1}H−^{31}P gradient HSQC (*29*) spectra in ^2H_2O at 20 °C. In ^1H_2O, data were acquired with a jump-and-return pulse sequence (*30*) and in ^2H_2O, with Watergate suppression of the residual water signal (*31*). All data sets were acquired in a phase sensitive mode (TPPI). In ^1H_2O, NOESY data sets were collected with 8K complex points over a spectral width of 20 kHz with 300 t_1 increments. In ^2H_2O, NOESY data sets were collected with 3412 complex points and 300 t_1 free induction decays covering a spectral width of 9 kHz. TOCSY spectra were recorded with 3798 t_2 complex points and 256 t_1 increments, and the DQF-COSY spectrum was acquired with 3468 t_2 complex points and 512 t_1 increments. All proton chemical shifts were referenced to internal sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS). All data sets were processed using VNMR (Varian Instruments) and FELIX 2000 (Accelrys Inc.). Structures were visualized and figures prepared with Insight II (Accelrys Inc.).

Distance and Torsion Restraints. NOE cross-peaks involving exchangeable protons in NOESY spectra (50 and 200 ms mixing times) in H_2O buffer were classified as strong (strong intensity at 50 ms), medium (barely observable at 50 ms), and weak (not observable at 50 ms and observable at 200 ms), and the observable proton pairs were restrained to distances of 3.0 ± 0.6, 4.0 ± 1.0, and 5.0 ± 1.2 Å, respectively. NOE buildups for nonexchangeable protons were derived from NOESY spectra in ^2H_2O buffer recorded as a function of missing time (50, 100, 150, 250, and 600 ms). Distances were estimated from the initial buildup rates within FELIX 2000. The cytosine H5−H6 interproton distance of 2.46 Å was used as a reference. The upper and lower bounds were allowed to vary ±20%. Overlapping cross-peaks were given generous bounds (up to ±40%). Atoms participating in experimentally identified canonical base pairing (based on NOE patterns) were restrained with distances corresponding to ideal hydrogen bond geometry (*32*).

The appearance of strong H1′−H2′ and very weak to no H2″−H3′ cross-peaks in the DQF-COSY spectrum indicates that the most populated conformations are of the S type. Thus, δ and the endocyclic $\nu(0)-\nu(4)$ torsion angles for all nonterminal residues were moderately constrained, leaving the sugar free to take any conformation without an energy penalty between C4′-*endo* and O1′-*endo*, including C2′-*endo*. Phosphorus chemical shifts span only the very narrow region that is typically found for undistorted DNA. Furthermore, no residues showed cross-peaks resulting from $^3J_{H5'-P}$ or $^3J_{H5''-P}$, representing couplings smaller than 5 Hz. But $^4J_{H4'-P}$ was readily detectable. Thus, β was loosely constrained to the *trans* conformation (maintaining both B and A helix geometry), 180 ± 20°, and γ kept at 36 ± 30° (*33*). Dihedral angle constraints for the ϵ torsion angle were derived from the observation of $^3J_{H3'-P}$ coupling (*34*, *35*); i.e., for residues for which an H3′−P peak in the HSQC was clearly observed, representing a coupling greater than 5 Hz, ϵ was loosely constrained to include both *trans* and *gauche*$^-$ conformations (−120 ± 45°). The glycosidic torsion angle χ was restricted to the experimentally assigned *anti* disposition for all residues (220 ± 50°). The program *CURVES 5.2* (*36*) was used to estimate DNA conformation and helical parameters.

Distance-Restrained Molecular Dynamics Regularization. Calculations were performed with XPLOR (*37*) using the CHARMm force field (*38*) and adapted for restrained MD for nucleic acids. All calculations were executed in vacuo without explicit counterions. Sets of rMD calculations were performed starting with Insight II-generated model structures of B-DNA and A-DNA, and using random velocities fitting a Maxwell−Boltzmann distribution. The empirical energy function was developed for nucleic acids and treated all hydrogens explicitly. It consisted of energy terms for hydrogen bonding, and nonbonded interactions, bonds, bond angles, torsion angles, and tetrahedral and planar geometry, including van der Waals and electrostatic forces. The effective function included terms describing distance and dihedral restraints, which were in the form of square well potentials (*39*). Most estimated distances from NOE data analysis were incorporated as ambiguous restraints using the "SUM averaging" option of XPLOR, since they could reflect intrastrand and/or interstrand contributions. On the basis of

the 2-fold symmetry, noncrystallographic symmetry restraints were imposed on all atoms. Planarity restraints were not used in the computations. The simulated annealing procedure consisted of a total of 53 ps of rMD, including heating for 7 ps from 300 to 1000 K, a 20 ps scale-up of restraints at high temperatures, cooling to 300 K for 14 ps, and equilibration rMD for 12 ps. The temperature was controlled by coupling the molecules to a temperature bath with a coupling constant of 0.025 ps (*40*). The van der Waals term was approximated using the Lennard-Jones potential energy function, and bond lengths involving hydrogens were fixed with the SHAKE algorithm (*41*) during molecular dynamics calculations. Sets of 10 structures for each of the initial A- and B-DNA species, selected on the basis of the lowest total energy, are essentially indistinguishable from one another, with an rmsd of <0.12 Å. Coordinates (entry 1N4B) have been deposited in the Protein Data Bank.

RESULTS

The one-dimensional imino and aromatic proton spectrum in 100 mM NaCl at pH 7.4 and 2 °C is shown in Figure 1c. The spectrum shows well-resolved peaks in the imino proton and aromatic proton regions, indicating a stable structure suitable for NMR studies. The number of peaks corresponds to one conformer with complemetary base pairing.

Resonance Assignments. Using the information from through-space NOE connectivities and homo- and heteronuclear through-bond *J* coupling connectivities, exchangeable and nonexchangeable protons (except for some H5′/5″ protons) of the 5′-CGAAAC*TTTCG-3′ sequence were assigned. We did not distinguish relative assignment for H5′ and H5″ protons.

Three sharp thymine imino proton resonances, one sharp guanine imino proton resonance, and one broad guanine imino proton resonance are observed in the region of 12−13.5 ppm (Figure 1c). Assignment of the imino and amino exchangeable protons was performed following analysis of jump-and-return NOESY spectra with mixing times of 60 and 200 ms. Shown in Figure 2a is a 1-1 NOESY spectrum (200 ms) depicting selected assignments for exchangeable protons. NOEs between thymine imino protons and adenine H2 and amino protons across A·T base pairs and between imino and cytosine amino and H5 protons across G·C base pairs are also observed. The data set is consistent with canonical base pairing for G2·C10, T9·A3, and T8·A4 base pairs, but not for a T7·A5 base pair. In the latter, the NOE between T7 H3 and A5 H2 is absent. However, the dipolar connectivities to the amino protons of A5, T7 H3·A5 H61 and T7 H3·A5 H62, can still be observed with a chemical shift difference of 1.7 ppm. This could indicate that the A5 H61−T7 O4 hydrogen bond is still present, although the T7 H3−A5 N1 hydrogen bond cannot be inferred. Notably, unlike the other imino resonance of thymine, T7 H3 shows cross-peaks to C6 ethyl protons.

Due to broadening of G11 H1 resonances, G11 H1−C1 H41 and G11 H1−C1 H42 cross-peaks are only observed at very low contour levels. The associated chemical shift difference of 1.2 ppm between the amino proton resonances of C1 indicates that canonical base pairing is maintained. The broadness of the G11 H1 imino proton resonance of this terminal base can be explained with the increased

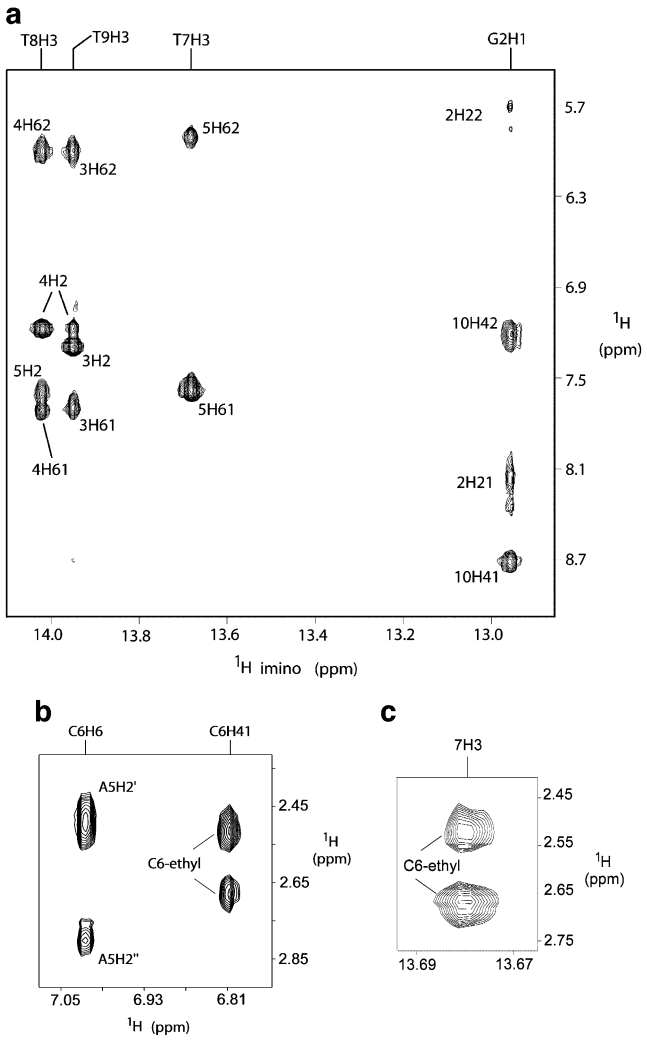

FIGURE 2: Selected regions from a NOESY spectrum (200 ms) in ^1H_2O at 2 °C and pH 7.4 with 100 mM NaCl. (a) Correlations involving imino and amino protons. The indicated assignments are relevant to establishing canonical base pairing. NOE correlations between C6 H ethyl protons and C6 H41 (b) and the imino T7 H3 (c).

accessibility of water to the hydrogen bonds forming the base pair. The imino resonances of the cross-linked cytosine, C6 H41, were assigned due to their strong cross-peak to C6 H5 and weaker cross-peak to C6 H6. This means that N4 atoms of the alkylated cytidine are mutually cis (Figure 1b). C6 H41 exhibits contacts to its C6 ethyl protons and T7 H3, an indication of their proximity (Figure 2b). The well-stacked nature of the consecutive canonical base pairs can be readily traced through several intra- and interstrand cross-peaks: G10 H41−C1 H41, T8 H3−A5 H2, T9 H3−A4 H2, T8 H3−T9 H3, T8 H3−T7 H3, and T9 H3−G2 H1. The imino and amino chemical shifts are given in Table 1.

Assignments of the H8/H6 base and H1′ sugar protons were made through analysis of the 50 and 250 ms NOESY spectra by established methods (*42*, *43*) (Figure 3). We can readily trace the sequential dipolar connectivities between the base and its own 5′-flanking sugar H1′ protons along individual strands as expected for right-handed DNA. With a short mixing time (50 ms), weak H8/H6−H1′ cross-peaks were observed, indicating that all residues are *anti* (when compared to C2 H6−C2 H5 and C11 H6−C11 H5 cross-peaks).

Table 1: Chemical Shifts for Exchangeable Protons in d(CGAAAC*TTTCG) at 2 °C and pH 7.4

assignment	C1	G2	A3	A4	A5	C6*	T7	T8	T9	C10	G11
H1−H3	−	12.96	−	−	−	−	13.68	14.02	13.95	−	13.2
H41−H61	8.36	8.2	7.69	7.70	7.53	6.81	−	−	−	8.71	*a*
H42−H62	7.16	5.7	6.0	6.00	5.91	−	−	−	−	7.21	*a*

[a] Broadened beyond recognition.

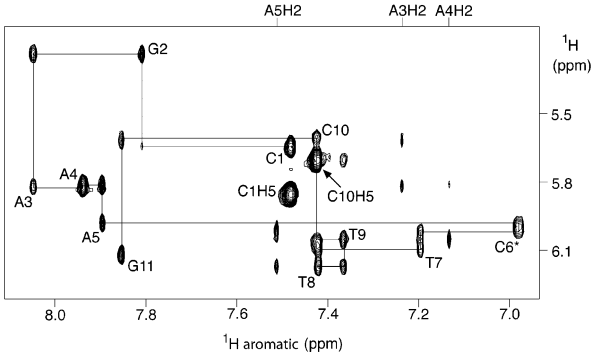

FIGURE 3: Selected region of the NOESY spectrum (250 ms) of d(CGAAACTTTCG)$_2$ in a ^2H_2O solution at 20 °C and pH 7.4 with 100 mM NaCl showing NOE correlations between H8, H6, H2, and H5 and H1′. The self-peaks involving H1′ protons have been labeled, and their sequential connectivities are indicated.

The assignment of adenine A H2 protons was derived from analysis of 250 and 600 ms NOESY spectra. When the glycosidic bond of an adenosine is in the regular *anti* domain, the intraresidue H2−H1′ distance is approximately 4.6 Å. This distance should allow for the identification of the A H2 signals in a canonical right-handed B-DNA duplex. Cross-strand (*i*) A H2−(*j* + 1) N H1′ (where *i* and *j* represent complementary bases and N is any base) and interstrand (i) A H2−(*i* + 1) N H1′ couplings may be observable with a longer mixing time. Hence, we observe A3 H2−A3 H1′, A3 H2−C10 H1′, A4 H2−A4 H1′, and A4 H2−T9 H1′ at 600 ms. The interstrand cross-peaks are observed even at 250 ms. With this mixing time, the adenine A5 shows strong interstrand cross-peaks (A5 H2−A5 H1′ and A5 H2−T8 H1′). In contrast to A3 H2 and A4 H2, the inter- and/or intrastrand A5 H2−C6 H1′ cross-peak is observed.

The stereospecific assignment of individual H2′ and H2″ protons was achieved by intensity comparison of the H1′−H2′ and H1′−H2″ intraresidue cross-peaks in a 50 ms NOESY spectrum, where the former is found to be stronger than the latter. In contrast to other nucleotides (with the exception of C10), the chemical shift of C1 H2′ appears at a lower field with respect to C1 H2″. Two-dimensional TOCSY and DQF-COSY spectral patterns indicated that H1′−H2′ coupling constants were reasonably large (>6 Hz, except for C10 in which measurement was precluded due to isochronous signals). Except for the C1·G11 terminal base pair, the deoxyribose sugars showed weak H2″−H3′ cross-peaks. This observation fixes the sugar geometries to be predominantly S-type (*44*), and consequently, the strand structures may be taken to belong to the B-DNA family. Backbone torsion angles were qualitatively examined from a ^{1}H−^{31}P HSQC correlation spectrum (Supporting Information). The phosphorus chemical shifts fall within a ∼1 ppm envelope, indicating preferential population of the canonical B_I backbone phosphate conformation.

Structural Features. A view of 10 superpositioned refined structures of the duplex d(CGAAAC*TTTCG) is plotted in Figure 4 and exhibits pairwise heavy atom root-mean-square deviations (rmsds) for all residues of 0.53 ± 0.2. The input and structure convergence parameters are listed in Table 2. A representative structure is plotted in Figure 5. In Figure 5a, a skeletal view depicts the protrusion of the cross-link into the major groove. This perturbation decreases as one moves away from the cross-link where the structure is seen to be compact, with bases lying in the interior of the helix and consequently being poorly accessible to the solvent. The right-handed self-complementary duplex with antiparallel strands consists of A3·T9, A4·T8, and C10·G2 canonical base pairs. The anticipated canonical base pairing is disrupted in bases adjacent to the A5−T7 cross-link.

Geometrical parameters for the solution structure of d(CGAAAC*TTTCG) in 100 mM NaCl (pH 7.4(are shown in Table 3. Significant distortions due to the cross-link are observed in the base pair propeller twists, buckle, cross-strand interphosphate (P−P), and base cross-strand (C1′−C1′) distances, *x* displacement, and sugar pucker. The magnitude of the base pair propeller twist is greater at the cross-linked pseudoplatform and diminishes away from it. The magnitudes of buckling between canonical base pairs follows the same trend. Concomitantly, the interphosphate (P−P) and interbase (C1′−C1′) distances increase for canonical base pairs. This has a pronounced effect on the *x* displacement that approaches the 4 Å displacement found in A-DNA compared to no displacement which is characteristic of B-DNA. While classical B-DNA features a stepwise twisting of 36° for every canonical base pair resulting in a duplex with partial intrastrand base stacking, here we observe instead a duplex with dramatically different twisting angles in the central 5′-AAC*TT-3′ segment and preferential intrastrand base stacking. The cross-link C6*−C6* mispair and the adjacent A5−T7 pseudoplatform show only intrastrand stacking (Figure 6a) with a related right-handed A5pC6 base step twist of ∼66° (underwound by ∼30°) and a C*6pT7 base step twist of ∼32° (overwound by ∼4°). The five-membered ring of A5 stacks on C6*, as does T7. For the A5−T7 pseudoplatform and the A4·T8 base pair, only the pyrimidines stack (Figure 6b), with a right-handed base step twist of ∼44°. A4·T8 and A3·T9 base pairs (Figure 6c) stack primarily through the overlap of purines with the buckled platforms related by a A3pA4 base step twist angle of ∼26°. Normally, in AT-rich sequences, Watson−Crick pairs have overwinding propeller twist parameters (*45*, *46*). However, in the present case, both small unwinding steps are necessary to accommodate the large overwinding resulting from the cross-link. For the A3·T9 to G2·C10 step, a G2pA3 base step twist angle of ∼35° results in good intrastrand stacking (Figure 6d). In summary, overwinding of a portion of a strand is met with underwinding of the adjacent portion, which results in bending toward the major groove by an angle of ∼27°.

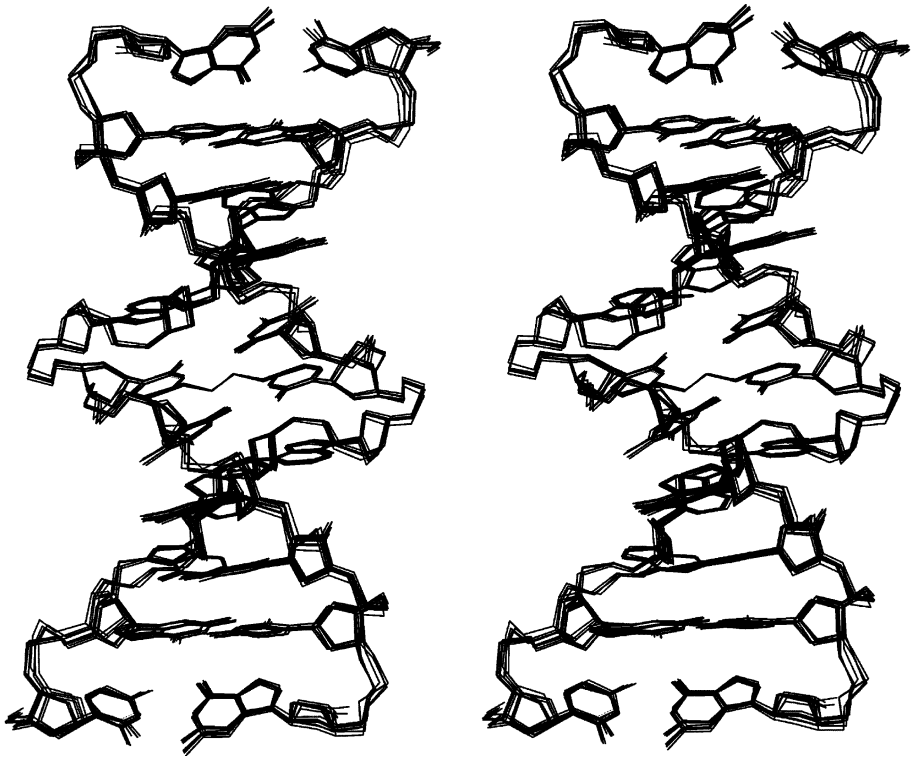

FIGURE 4: Stereoview of a superposition of 10 relaxation matrix-refined structures of the d(CGAAAC*TTTCG)$_2$ duplex.

Table 2: Restraints and Refinement Statistics for 10 Selected Structures for d(CGAAAC*TTTCG)$_2$

NMR distance restraints	
total no. of restraints	449
no. of nonexchangeable protons	243
no. of exchangeable protons	53
no. of hydrogen bond restraints (empirical)	44
no. of dihedral angle restraints	109
structural statistics	
NMR *R*-factor ($R_{1/6}$)	0.104−0.105
NOE rmsd (Å) (total)	0.012−0.013
no. of NOE violations exceeding 0.2 Å	0
heavy atom pairwise rmsd	0.5±0.2

The H2 chemical shifts of adenines A4 and A3 are upfield from that for adenine A5. From the structure, it is apparent that both A4 and A3 fall in the shielding region of anisotropic ring current fields of each other, as opposed to A5, which only experiences the influence of its own ring current. If the terminal bases are excluded, the overall base step twist angle is ∼36°, and therefore, the perturbations are localized and fully accommodated within three bases on either side of the lesion.

Notably, we observe narrowing of the width of both the minor *and* major grooves. With the reduction of both minor and major grooves, the accessibility of the solvent seems to be reasonably impaired below 30 °C. Indeed, the diameter of the DNA helix toroid (as measured from the interstrand distance between phosphorus atoms) increases at the cross-link site (Table 3). This increased diameter is the basis for the disruption of the adjacent base pairing (A5·T7). With progression away from the lesion, the diameter sharply decreases, as is normal with a bubble in a DNA stem (C6* → A5 and C6* → T7) and then increases to canonical B-DNA (A5 → A4, A3 → G2, T7 → T8, and T9 → C10). Concomitantly, the magnitude of propeller twisting diminishes, as does the magnitude of the Watson−Crick pair buckle.

DISCUSSION

Conventional refinement protocols seek a structure with a minimal weighted sum of a force field energy and a penalty term based on experimentally observed parameters. An ensemble of structures typically calculated during an NMR refinement represents the precision of determination of such an average conformation. Due to the time-average nature of NMR signals, such a structure also corresponds to an average conformation in solution (for a discussion on the issue, see ref *47*). This is illustrated in the study presented here where among the set of 10 final structures there is sugar puckering between C1′-*exo* and C3′-*exo* (Table 3), i.e., showing little deviation from the S-type puckering phase angle ($148° < P < 180°$). These have suboptimal energies and could result from torsions induced by the cross-link. This is not surprising since the conformational flexibility is to be expected for the ethyl moiety of the cross-link. Such sugar puckers have been observed in other NMR structures previously (*47*−*49*) and not only reflect the broad distribution of sugar puckers centered on C2′-*endo* but also are a reflection of the greater flexibility of B-DNA over A-DNA (*50*). Indeed, this greater flexibility of the B helix compared to the A helix makes the B helix a better candidate for recognition (*51*). The terminal bases are N-type or close to N-type (C1 is C3′-*endo*, and G11 is O1′-*endo*) since they enjoy a greater range of motion.

In a previous circular dichroism study of an N^4C−ethyl−N^4C cross-linked duplex that contained the same 5′-AAAC*TTT-3′ core sequence and 5′-complementary overhanging ends (*24*), the CD spectrum of this cross-linked duplex was similar to that of B-DNA. Our studies show that the cross-linked duplex has overall a B-type helical conformation, albeit with features that indicate deformability toward

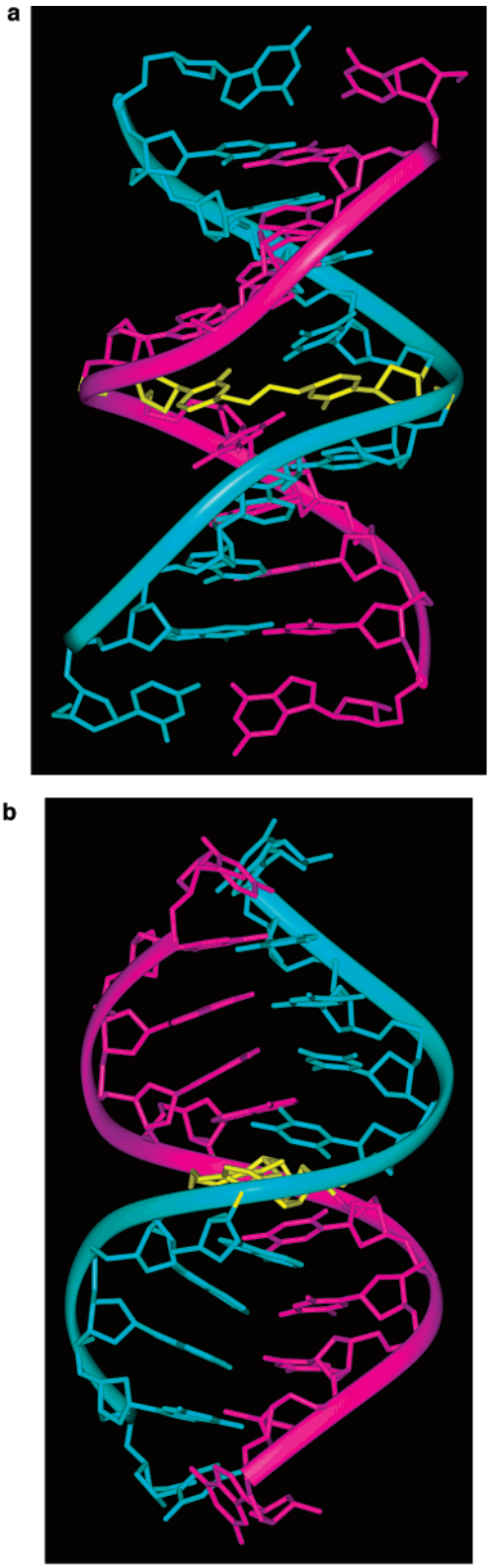

FIGURE 5: Orthogonal views of the duplex structure of d(CGAAAC*TTTCG)$_2$. The individual strands are pink and blue, and the cross-link is yellow. (a) View into the major groove. (b) Effect of the cross-link on the molecular axis. Noticeable are the bending and/or kink of the DNA stem and also the disruption of base pairing adjacent to the cross-link.

A-DNA. In general, the deformability of the B helix to the A helix is not only reflected by conversion of geometries of the deoxyribose sugars from C2′-*endo* in B-DNA to C3′-*endo* in A-DNA. In a B helix−A helix transition, the long

Table 3: Geometrical Parameters for a Selected Structure of d(CGAAAC*TTTCG)$_2$

	interstrand		base−base	
base pair	P−P (Å)	C1′−C1′ (Å)	buckle (deg)	twist (deg)
G2·C10	18.6	10.7	−2.9	3.4
A3·T9	17.9	10.2	20.6	0.2
A4·T8	17.1	9.8	34.5	−13.5
A5·T7	16.3	10.9	−3.2	−27.1
C6·C6	19.4	10.2	0.0	−44.9
base	phase	pucker	*x* displacement (Å)	
C1	17.4	C3′-*endo*	−0.2	
G2	150.5	C2′-*endo*	0.2	
A3	194.9	C3′-*exo*	0.3	
A4	118.4	C1′-*exo*	0.2	
A5	187.4	C3′-*exo*	−3.1	
C6*	133	C1′-*exo*	1.9	
T7	142.9	C1′-*exo*	−0.1	
T8	122.4	C1′-*exo*	0.5	
T9	123.1	C1′-*exo*	0.2	
C10	126.5	C1′-*exo*	0.8	
G11	88.3	O1′-*endo*	0.1	

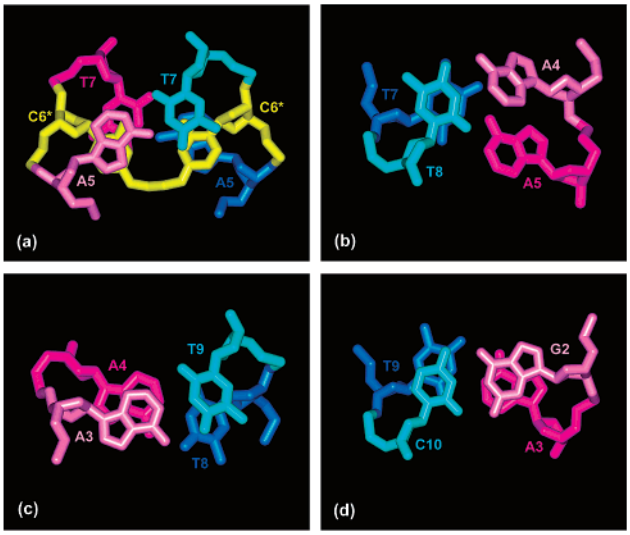

FIGURE 6: Nearest-neighbor stacking depictions of bases in the structure of d(CGAAAC*TTTCG)$_2$. Stacking interactions are solely intrastrand. (a) Between the C6*−C6* and A5−T7 cross-links, the stacking only involves pyrimidines. (b) Stacking between A5−T7 and A4−T8 cross-links is rather poor. (c) Stacking in the base pair step in connection with A4·T8 and A3·T9 base pairs involves the six-membered rings of the adenines. The intrastrand stacking between the thymines involves the methyl moiety of T9. (d) Good stacking involving the six-membered rings of the purines and between the pyrimidines in the A3·T9 to G2·C10 base pair step.

and narrow B duplex with its 10−10.5 bp repeat, 3.4 Å rise, and base pairs stacked at the center of the helix is converted to the underwound and compact A-DNA structure, characterized by an 11 bp repeat, a 2.6 Å rise, and base pairs that are inclined up to 20° and displaced by 4 Å so that they essentially wrap around the helix axis (*52*).

Inspection of the torsion angles ϵ and ξ along the sugar−phosphate backbone reveals the phosphodiester populating the B_I conformation. The ensemble of refined structures shows the possible formation of an interstrand hydrogen bond between the C*6 imino and O3′ of A4.

Bending. Our NMR-based structure shows that the ethyl cross-link induces a bend in the duplex of ∼27° toward the major groove with some conformational flexibility. However,

our previous study (*24*) of multimers in which this same cross-link was phased every 10 bp suggested little or no bending deformation. This conclusion was based upon the similar electrophoretic mobility of the duplex cross-linked to normal DNA. In theory, bending should decrease the electrophoretic mobility. The discrepancy between these results is likely due to the flexibility or bendability of the helix at the site of the cross-link. A DNA helix may be bent and relatively inflexible or, on average, straight but highly flexible. A highly flexible helix may also result in "normal" mobility on the gel. The gel studies coupled with our NMR results suggest that the bent conformation observed by NMR is one of an ensemble of conformations possible in a highly flexible helix.

Comparison with Other Cross-Linked Structures. Intuitively, distortions induced by interstrand cross-links are expected in sugar puckering with concomitant bending, and the possibility of disruption of flanking base pairs. Most of these were observed in the study presented here. A solution structure study of a trimethylene interstrand cross-link in d(CpG)*d(CpG) sites flanked by G·C Watson−Crick base pairs has been determined (*53*). As in our study, this cross-link, which in this case resides in the minor groove, results in a one-residue bulge involving the unpaired cytosines. The authors did not find significant bending of the structure possibly due to the short stem used as compared to the structural perturbations observed (only two symmetry-related base pairs were observed in an octamer). The cross-linked guanines formed a platform inducing widening of the diameter of the DNA, a feature we have also observed in directly opposing cytosines.

Conformational flexibility, thought to have induced a mixture of C2′-*endo* and C3′-*endo* conformations, due to a cross-link is also observed in the solution structure of a psoralen cross-linked DNA duplex with a d(GGGTACCC) sequence (*45*, *46*, *48*, *49*). The cross-link involves thymine residues in opposite strands at d(ApT)*d(ApT) sites, and these are flanked by G·C Watson−Crick base pairs. Just as with the 5′-AAAC*TTT-3′ sequence, the cross-link distorts the ApT base stacking.

Biological Significance. A key issue in the repair of a DNA lesion is the ability of the repair machinery to recognize and repair certain lesions while failing to repair others. A growing body of evidence suggests that both a distortion of the DNA helix and a modification to the DNA chemical structure are needed to trigger repair (*54*−*57*). Furthermore, changes in major and minor groove dimensions provide a means of regulating the accessibility of DNA to proteins. The bifunctional alkylating antitumor drugs generally alkylate in the major groove of DNA to produce interstrand cross-links. In protein−DNA interactions, the introduction of a helix into the major groove induces broad conformational changes characteristic of a B-to-A conversion (*51*). In the present structural environment, we observe features that represent a departure from canonical B-DNA toward A-DNA with some conformational flexibility. These structural and dynamic changes appear to be a general feature of the "conformation recognition" of DNA by proteins (*51*). Therefore, our cross-linked duplex may represent a protein-recognizable model for a major groove-alkylated DNA and justifies further systematic studies of sequence- and cross-link-induced changes in structure. To our knowledge, specific $N^4C-alkyl-N^4C$-type lesions have not been observed; however, Romero et al. (*25*) have characterized an interstrand cross-link formed by reaction of mechlorethamine with a C·C mispair in DNA. In this case, alkylation occurs at N3 of cytosine to produce an $N^3C-alkyl-N^3C$ interstrand cross-link. Although the structure of this cross-linked DNA has not been determined, it appears that its rate of formation is dependent upon the type of base pairs that surround the C·C mispair. As with the distortions seen in our duplex, molecular dynamics simulations with solvated duplexes containing a C·C mispair suggested local opening of the duplex when the mispair is flanked by canonical A·T base pairs.

The NMR study reported here provides the first three-dimensional structure of a DNA helix containing a direct ethyl cross-link between nucleosides. It thus expands the limited repertoire of studies on the perturbations resulting from a variety of environmental and therapeutic agents that react with DNA. Structural studies of this type are an important and necessary component in the understanding of the relation of structural perturbations to recognition and repair of interstrand cross-links. Preliminary experiments in *E. coli* show the $N^4C-ethyl-N^4C$ cross-link in a 5′-AAAC*TTT-3′ sequence in plasmid DNA can be repaired by the nucleotide excision repair pathway. The pronounced structural perturbations we observe in our cross-linked duplex may be the basis for recognition of this lesion. Both the B-to-A helix deformability and DNA bendability are reflections of the flexibility induced by the $N^4C-ethyl-N^4C$ lesion that could allow for its recognition and trigger appropriate repair pathways. Experiments for testing this hypothesis are currently in progress.

ACKNOWLEDGMENT

Duke University NMR Center was established with grants from the NIH, the NSF, and the North Carolina Biotechnology Center. We would like to thank Drs. Robert E. London and Eugene F. DeRose for their help in obtaining the proton-phosphorous HSQC data at the NMR facility of the National Institutes of Environmental Health Sciences.

SUPPORTING INFORMATION AVAILABLE

A figure showing a $^1H-^{31}P$ HSQC spectrum of $d(CGAAAC*TTTCG)_2$ at 20 °C and a table of chemical shifts for the nonexchangeable protons. This material is available free of charge via the Internet at http://pubs.acs.org.

NOTE ADDED AFTER ASAP POSTING

This article was inadvertently released ASAP on 11/26/02. The sequence in the abstract is now correct. The correct version of this paper was posted 12/17/02.

REFERENCES

1. Colvin, M., and Chabner, B. A. (1990) in *Cancer Chemotherapy: Principles and Practice* (Collins, J. M., Ed.) pp 276−313, J. B. Lippincott Co., Philadelphia.
2. Dong, Q., Johnson, S. P., Colvin, O. M., Bullock, N., Kilborn, C., Runyon, G., Sullivan, D. M., Easton, J., Bigner, D. D., Nahta, R., Marks, J., Modrich, P., and Friedman, H. S. (1999) *Cancer Chemother. Pharmacol. 43*, 73−79.
3. Dronkert, M. L., and Kanaar, R. (2001) *Mutat. Res. 486*, 217−247.
4. Sladek, F. M., Munn, M. M., Rupp, W. D., and Howard-Flanders, P. (1989) *J. Biol. Chem. 264*, 6755−6765.

5. Cheng, S., Sancar, A., and Hearst, J. E. (1991) *Nucleic Acids Res. 19*, 657−663.
6. Berardini, M., Foster, P. L., and Loechler, E. L. (1999) *J. Bacteriol. 181*, 2878−2882.
7. Berardini, M., Mackay, W., and Loechler, E. L. (1997) *Biochemistry 36*, 3506−3513.
8. McHugh, P. J., Sones, W. R., and Hartley, J. A. (2000) *Mol. Cell. Biol. 20*, 3425−3433.
9. Broomfield, S., Hryciw, T., and Xiao, W. (2001) *Mutat. Res. 486*, 167−184.
10. Mu, D., Bessho, T., Nechev, L. V., Chen, D. J., Harris, T. M., Hearst, J. E., and Sancar, A. (2000) *Mol. Cell. Biol. 20*, 2446−2454.
11. Li, L., Peterson, C. A., Lu, X., Wei, P., and Legerski, R. J. (1999) *Mol. Cell. Biol. 19*, 5619−5630.
12. Grueneberg, D. A., Ojwang, J. O., Benasutti, M., Hartman, S., and Loechler, E. L. (1991) *Cancer Res. 51*, 2268−2272.
13. Ojwang, J. O., Grueneberg, D. A., and Loechler, E. L. (1989) *Cancer Res. 49*, 6529−6537.
14. Rink, S. M., and Hopkins, P. B. (1995) *Biochemistry 34*, 1439−1445.
15. Rink, S. M., Solomon, M. S., Taylor, M. J., Rajur, S. B., Mclaughlin, L. W., and Hopkins, P. B. (1993) *J. Am. Chem. Soc. 115*, 2551−2557.
16. Bellon, S. F., and Lippard, S. J. (1990) *Biophys. Chem. 35*, 179−188.
17. Brabec, V., Sip, M., and Leng, M. (1993) *Biochemistry 32*, 11676−11681.
18. Paquet, F., Boudvillain, M., Lancelot, G., and Leng, M. (1999) *Nucleic Acids Res. 27*, 4261−4268.
19. Sip, M., Schwartz, A., Vovelle, F., Ptak, M., and Leng, M. (1992) *Biochemistry 31*, 2508−2513.
20. Coste, F., Malinge, J. M., Serre, L., Shepard, W., Roth, M., Leng, M., and Zelwer, C. (1999) *Nucleic Acids Res. 27*, 1837−1846.
21. Noll, D. M., Noronha, A. M., and Miller, P. S. (2001) *J. Am. Chem. Soc. 123*, 3405−3411.
22. Noronha, A. M., Noll, D. M., and Miller, P. S. (2001) *Nucleosides, Nucleotides Nucleic Acids 20*, 1303−1307.
23. Noronha, A. M., Wilds, C. J., and Miller, P. S. (2002) *Biochemistry 41*, 8605−8612.
24. Noronha, A. M., Noll, D. M., Wilds, C. J., and Miller, P. S. (2002) *Biochemistry 41*, 760−771.
25. Romero, R. M., Rojsitthisak, P., and Haworth, I. S. (2001) *Arch. Biochem. Biophys. 386*, 143−153.
26. Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546−4554.
27. Piantini, U., Sorensen, O. W., and Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 6800−6801.
28. Braunschweiler, L., and Ernst, R. R. (1987) *J. Magn. Reson. 53*, 521−528.
29. Kay, L. E., Keifer, P., and Saarinen, T. (1992) *J. Am. Chem. Soc. 114*, 10663−10665.
30. Gueron, M., and Plateau, P. (1982) *J. Am. Chem. Soc. 104*, 7310−7311.
31. Piotto, M., Saudek, V., and Sklenar, V. (1992) *J. Biomol. NMR 2*, 661−665.
32. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
33. Chou, S. H., Zhu, L., Gao, Z., Cheng, J. W., and Reid, B. R. (1996) *J. Mol. Biol. 264*, 981−1001.
34. Varani, G., Abouela, F., and Allain, F. H. T. (1996) *Prog. Nucl. Magn. Reson. Spectrosc. 29*, 51−127.
35. Smith, J. S., and Nikonowicz, E. P. (2000) *Biochemistry 39*, 5642−5652.
36. Lavery, R., and Sklenar, H. (1989) *J. Biomol. Struct. Dyn. 6*, 655−667.
37. Brunger, A. T. (1992) Ph.D. Thesis, Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT.
38. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) *J. Comput. Chem. 4*, 187−217.
39. Clore, G. M., Gronenborn, A. M., Carlson, G., and Meyer, E. F. (1986) *J. Mol. Biol. 190*, 259−267.
40. Berendsen, H. J. C., Postma, J. P. M., Vangusteren, W. F., Dinola, A., and Haak, J. R. (1984) *J. Chem. Phys. 81*, 3684−3690.
41. Ryckaert, J.-P. P., Ciccotti, G., and Berendsen, H. J. C. (1977) *J. Comput. Phys. 23*, 327−341.
42. Wijmenga, S. S., Mooren, M. M. W., and Hilbers, C. W. (1993) in *NMR of Macromolecules, A Practical Approach* (Roberts, G. C. K., Ed.) pp 217−288, Oxford University Press, New York.
43. Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.
44. Hosur, R. V., Govil, G., and Miles, H. T. (1988) *Magn. Reson. Chem. 26*, 927−944.
45. Kerwood, D. J., Zon, G., and James, T. L. (1991) *Eur. J. Biochem. 197*, 583−595.
46. Shatzky-Schwartz, M., Arbuckle, N. D., Eisenstein, M., Rabinovich, D., Bareket-Samish, A., Haran, T. E., Luisi, B. F., and Shakked, Z. (1997) *J. Mol. Biol. 267*, 595−623.
47. Tjandra, N., Tate, S., Ono, A., Kainosho, M., and Bax, A. (2000) *J. Am. Chem. Soc. 122*, 6190−6200.
48. Hwang, G. S., Kim, J. K., and Choi, B. S. (1996) *Biochem. Biophys. Res. Commun. 219*, 191−197.
49. MacDonald, D., Herbert, K., Zhang, X., Pologruto, T., Lu, P., and Polgruto, T. (2001) *J. Mol. Biol. 306*, 1081−1098.
50. Dickerson, R. E. (1999) in *Oxford Handbook of Nucleic Acid Structure* (Neidle, S., Ed.) pp 145−197, Oxford University Press, New York.
51. Olson, W. K., Gorin, A. A., Lu, X.-J., Hock, L. M., and Zhurkin, V. B. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 11163−11168.
52. Arnott, S. (1999) in *Oxford Handbook of Nucleic Acid Structure* (Neidle, S., Ed.) pp 1−38, Oxford University Press, New York.
53. Dooley, P. A., Tsarouhtsis, D., Korbel, G. A., Nechev, L. V., Shearer, J., Zegar, I. S., Harris, C. M., Stone, M. P., and Harris, T. M. (2001) *J. Am. Chem. Soc. 123*, 1730−1739.
54. Sugasawa, K., Okamoto, T., Shimizu, Y., Masutani, C., Iwai, S., and Hanaoka, F. (2001) *Genes Dev. 15*, 507−521.
55. Hess, M. T., Schwitter, U., Petretta, M., Giese, B., and Naegeli, H. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 6664−6669.
56. Zou, Y., Luo, C., and Geacintov, N. E. (2001) *Biochemistry 40*, 2923−2931.
57. Gunz, D., Hess, M. T., and Naegeli, H. (1996) *J. Biol. Chem. 271*, 25089−25098.

BI026368L