Structural Studies of an Oligodeoxynucleotide Containing a Trimethylene Interstrand Cross-Link in a 5′-(CpG) Motif: Model of a Malondialdehyde Cross-Link

Patricia A. Dooley,[§] Dimitrios Tsarouhtsis,[‡] Gregory A. Korbel, Lubomir V. Nechev, Jill Shearer, Irene S. Zegar, Constance M. Harris, Michael P. Stone, and Thomas M. Harris*

Contribution from the Department of Chemistry and Center in Molecular Toxicology, Vanderbilt University, Nashville, Tennessee 37235

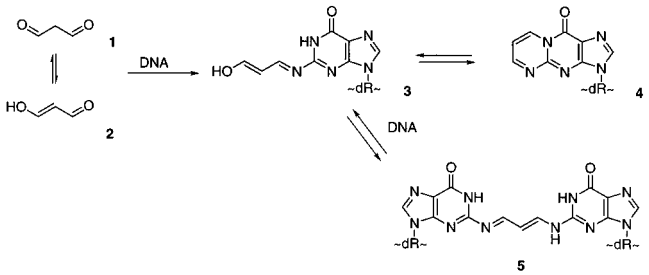
Received August 24, 2000

Abstract: Malondialdehyde (MDA), a known mutagen and suspected carcinogen, is a product of lipid peroxidation and byproduct of eicosanoid biosynthesis. MDA can react with DNA to generate potentially mutagenic adducts on adenine, cytosine, and particularly guanine. In addition, repair-dependent frame shift mutations in a GCGCGC region of *Salmonella typhimurium hisD3052* have been attributed to formation of interstrand cross-links (Mukai, F. H. and Goldstein, B. D. *Science* **1976**, *191*, 868−869). The cross-linked species is unstable and has never been characterized but has been postulated to be a bis-imino linkage between N^2 positions of guanines. An analogous linkage has now been investigated as a stable surrogate using the self-complementary oligodeoxynucleotide sequence 5′-d(AGGCG*CCT)$_2$, in which G* represents guanines linked via a trimethylene chain between N^2 positions. The solution structure, obtained by NMR spectroscopy and molecular dynamics using a simulated annealing protocol, revealed the cross-link only minimally distorts duplex structure in the region of the cross-link. The tether is accommodated by partially unwinding the duplex at the lesion site to produce a bulge and tipping the guanine residues; the two guanines and the tether attain a nearly planar conformation. This distortion did not result in significant bending of the DNA, a result which was confirmed by gel electrophoresis studies of multimers of a 21-mer duplex containing the cross-link.

Introduction

DNA damage is caused by oxidation, radiation, and chemical agents, any of which may result in DNA mutation and lead to genetic diseases, tumors, or cancer. Both exogenous and endogenous agents contribute to DNA damage. Among the latter is malondialdehyde (MDA) which arises in cells as a byproduct of thromboxane biosynthesis[1] and as a product of oxidative degradation of polyunsaturated lipids.[2] MDA (**1**) is shown in Scheme 1 along with its enol tautomer β-hydroxyacrolein (**2**). The compound is acidic (pK_a 4.46); consequently, it exists largely as the enolate anion at physiological pH. MDA is unstable, readily undergoing aldol condensations with itself and other carbonyl compounds. MDA reacts with amines including the exocyclic amino groups of the nucleic acid bases,[3−5] and, in particular, of guanine[6,7] where Schiff base **3** is formed.[8] The chemistry of the guanine adduct is complex; it undergoes a reversible ring closure to form pyrimidopurinone **4**, known as M_1G.[9] In addition, there is evidence for an interchain cross-link which apparently is also formed reversibly.[10] The structure of the cross-link is possibly **5**, but the lesion is only formed in low yield, is unstable, and has thus far eluded structural characterization.

Scheme 1



As might be expected, this plethora of adducts leads to a complex spectrum of mutations including transition and transversion point mutations, frame-shift mutations, and large-scale deletions and transpositions.[1,9,11−14] Synthetic inaccessibility of DNA containing the individual adducts has greatly complicated

* Corresponding author.
[§] Current Address: Department of Chemistry, United States Military Academy, West Point, NY 10996.
[‡] Current Address: Abbott Laboratories, Abbott Park, IL 60064.

(1) Basu, A. K.; Marnett, L. J. *Carcinogenesis* **1983**, *4*, 331−333.
(2) Marnett, L. J.; Basu, A. K.; O'Hara, S. M.; Weller, P. E.; Rahman, A. F. M. M.; Oliver, J. P. *J. Am. Chem. Soc.* **1986**, *108*, 1348−1350.
(3) Nair, V.; Turner, G. A.; Offerman, R. J. *J. Am. Chem. Soc.* **1984**, *106*, 3370−3371.
(4) Stone, K.; Uzieblo, A.; Marnett, L. J. *Chem. Res. Toxicol.* **1990**, *3*, 467−472.
(5) Stone, K.; Ksebati, M. B.; Marnett, L. J. *Chem. Res. Toxicol.* **1990**, *3*, 33−38.
(6) Moschel, R. C.; Leonard, N. J. *J. Org. Chem.* **1976**, *41*, 294−300.
(7) Seto, H.; Okuda, T.; Takesue, T.; Ikemura, T. *Bull. Chem. Soc. Jpn.* **1983**, *56*, 1799−1802.
(8) Basu, A. K.; O'Hara, S. M.; Valladier, P.; Stone, K.; Mols, O.; Marnett, L. J. *Chem. Res. Toxicol.* **1988**, *1*, 53−59.
(9) Benamira, M.; Johnson, K.; Chaudhary, A.; Bruner, K.; Tibbetts, C.; Marnett, L. J. *Carcinogenesis* **1995**, *16*, 93−99.
(10) Niedernhofer, L. J. Ph.D. Thesis, Vanderbilt University, 1996.
(11) Basu, A. K.; Marnett, L. J. *Cancer Res.* **1984**, *44*, 2848−2854.
(12) Basu, A. K.; Marnett, L. J.; Romano, L. J. *Mutat. Res.* **1984**, *129*, 39.

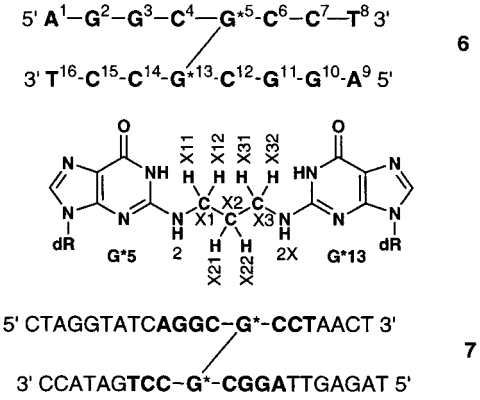
10.1021/ja003163w CCC: $20.00 © 2001 American Chemical Society
Published on Web 02/03/2001

studies of mutagenesis. Only recently has methodology been developed to prepare an oligonucleotide containing the M_1G adduct,[15,16] and the other adducts remain as synthetic challenges. An essential difference between the simple lesions and the interchain cross-link can be seen in the effect of nucleotide excision repair systems on their mutagenicity. Nucleotide excision repair is an important component of the cellular defense systems against exogenous and endogenous chemicals, free radicals, and radiation. This repair system excises a short oligonucleotide containing the damaged base(s) and then refills the gap using the other strand as a template. The nucleotide excision repair system lessens the genetic impact of simple adducts, but it potentially has the opposite effect on interchain cross-links. Interchain cross-links in DNA are in themselves absolute blocks to replications, that is, they are lethal and nonmutagenic. However, with interchain cross-links, excision of the lesion from one of the strands would leave it tethered to the other. If the gap were then refilled using this damaged template, the process would be inherently error-prone.

Mukai and Goldstein found MDA to be mutagenic in an Ames assay involving a histidine-requiring auxotroph *hisD3052*, of the bacterium *Salmonella typhimurium* sensitive to frameshift mutations which had normal nucleotide repair but MDA was not mutagenic in a variant organism in which a component of the repair system had been deleted.[17] Mitomycin C, a known cross-linking agent, exhibited the same pattern of mutagenic specificity.[18,19] The observation with MDA was subsequently confirmed in *Escherichia coli* by Yonei and Furui.[14] However, the result remains controversial; Basu et al. have concluded that the mutagenicity of malondialdehyde is independent of its ability to form interstrand cross-links.[12]

A central problem in the study of the biological effects of the putative MDA interstrand cross-link is the lack of synthetic methodology for preparing DNA duplexes in pure form containing this unstable lesion. Present methods employing MDA or its derivatives only lead to very low yields of the cross-link. As an alternative approach it should be possible to study the corresponding saturated analogue, that is, the N^2-N^2 propano cross-link, which is chemically stable. Methodology has been developed in this laboratory for the synthesis of such cross-links by reaction of the appropriate diamine with oligonucleotides containing halopurine equivalents of guanine.[20,21] This method creates the cross-link unambiguously in a defined location. The present paper describes the characterization of the structure of the cross-linked duplex in a 5′-d(CpG) sequence context by two-dimensional NMR to give the three-dimensional structure, by gel mobility to establish bending at the cross-link site and by thermal denaturation to assess the stabilization imparted to the duplex by the cross-link.

Scheme 2



Experimental Section

Synthesis and Purification of Cross-Link 6. Cross-linked oligonucleotide **6** (Scheme 2) was prepared by reaction of a halopurine-containing 8-mer (d(AGGCXCCT), X = 2-fluoro-O^6-(trimethylsilylethyl)-2′-deoxyinosine), with diaminopropane to form monoadducted 8-mer. Reaction with additional halopurine-containing 8-mer yielded cross-link **6**. Details of the synthesis and purification are supplied in the Supporting Information.

NMR Spectra. The unmodified oligodeoxynucleotide d(AGGCGCCT) was purchased from Midland Certified Reagent Co., Midland, TX. The concentration of **6** was determined assuming an extinction coefficient of 4.9×10^4 M^{-1} cm^{-1}.[22] All samples were dissolved in 10 mM sodium phosphate buffer containing 0.1 M NaCl and 50 mM EDTA (pH 7.1). For observation of nonexchangeable resonances, samples were lyophilized three times from 99.96% D_2O and finally dissolved in 0.6 mL of 99.996% D_2O, giving a 2 mM concentration of the duplex. For assignments of water-exchangeable protons, the samples were dissolved in a 9:1 H_2O/D_2O buffer of the same composition as that above. Spectra were referenced to an internal standard of TSP. Spectra were recorded at a ^1H frequency of 500.13 MHz.

Structural analysis was carried out by methodology in routine use for oligonucleotides. Details of spectral acquisition and analysis are provided in the Supporting Information. Analysis of signals of the trimethylene chain presented special problems. Classical A-DNA and B-DNA were used as the reference structures.[23] The initial models were constructed in INSIGHT II (Biosym Technologies, San Diego, CA) by bonding a propyl group to N^2 of the guanine residue in one strand with the correct orientation and partial charges.[24] Partial charges on the propyl group were obtained from those assigned to propane from the CHARMM force field.[25] The structures were energy-minimized for 100 iterations by the conjugate gradient method to give a monoadducted structure. The creation of the starting structures was completed using X-PLOR 3.1,[26] in which the free end of the propyl group was connected to N^2 of the guanine residue in the other strand with the appropriate orientation. Starting structures IniA and IniB were produced upon potential energy-minimization of the cross-linked model compounds using X-PLOR.

Bending and Melting Studies. Experimental details are provided in the Supporting Information.

Results

Melting Studies. The melting temperature of the unmodified oligodeoxynucleotide, determined from the first derivative of

(13) Fink, S. P.; Reddy, G. R.; Marnett, L. J. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 8652−8657.
(14) Yonei, S.; Furui, H. *Mutat. Res.* **1981**, *88*, 23−32.
(15) Reddy, G. R.; Marnett, L. J. *J. Am. Chem. Soc.* **1995**, *117*, 50077-5008.
(16) Schnetz-Boutaud, N. C.; Mao, H.; Stone, M. P.; Marnett, L. J. *Chem. Res. Toxicol.* **2000**, *13*, 90−95.
(17) Mukai, F. H.; Goldstein, B. D. *Science* **1976**, *191*, 868−869.
(18) Nakamura, S. I.; Oda, Y.; Shimada, T.; Oki, I.; Sugimoto, K. *Mutat. Res.* **1987**, *192*, 239−246.
(19) Mamber, S. W.; Bryson, V.; Katz, S. E. *Mutat. Res.* **1983**, *119*, 135−144.
(20) DeCorte, B. L.; Tsarouhtsis, D.; Kuchimanchi, S.; Cooper, M. D.; Horton, P.; Harris, C. M.; Harris, T. M. *Chem. Res. Toxicol.* **1996**, *9*, 630−637.
(21) Tsarouhtsis, D.; Kuchimanchi, S.; DeCorte, B. L.; Harris, C. M.; Harris, T. M. *J. Am. Chem. Soc.* **1995**, *117*, 11013−11014.
(22) Borer, P. N. *Handbook of Biochemistry and Molecular Biology*, 1st ed.; CRC Press: Cleveland, 1975.
(23) Arnott, S.; Hukins, D. W. L. *Biochem. Biophys. Res. Commun.* **1972**, *47*, 1504−1509.
(24) Hingerty, B. E.; Figueroa, S.; Hayden, T. L.; Broyde, S. *Biopolymers* **1989**, *28*, 1195−1222.
(25) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(26) Brunger, A. T. *X-Plor. Version 3.1. A System for X-ray Crystallography and NMR;* Yale University Press: New Haven, 1992.

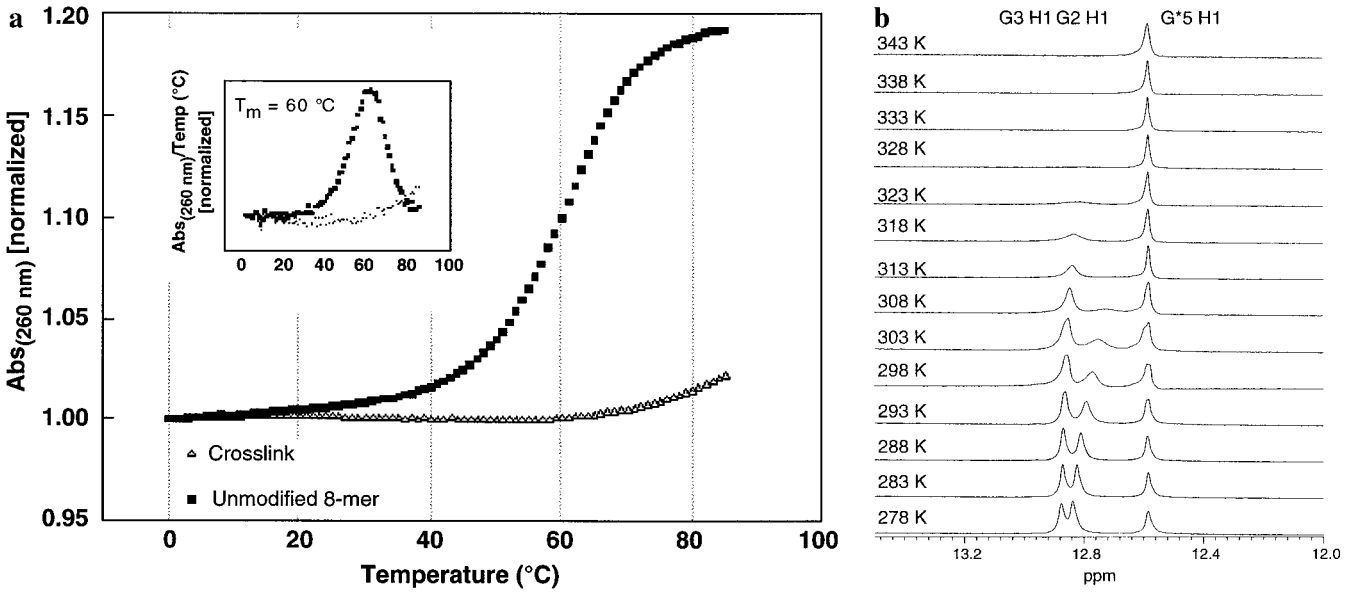


Figure 1. (A) Melting curves of the unmodified and trimethylene cross-link oligodeoxynucleotides 5′-d(AGGCG*CCT)$_2$ measured in a buffer of 1.0 M NaCl, 10 mM sodium phosphate, and 0.05 mM Na_2EDTA at pH 7.1; hyperchromicity was monitored at 260 nm. Insert: First derivative of the melting curve of the unmodified duplex, indicating $T_m = 60$ °C. (B) Expanded imino ^{1}H region of the trimethylene cross-link oligodeoxynucleotide 5′-d(AGGCG*CCT)$_2$ in 90% H_2O−10% D_2O at 5 °C increments. The imino proton of the cross-link tether G*5 does not exchange with the solvent at the measured temperatures.

the UV melting curve, was 60 ± 1 °C. Cross-linked oligodeoxynucleotide **6** did not melt over the range of temperature examined, although a small degree of hyperchromicity was observed (Figure 1A). When the imino protons of the cross-linked duplex were monitored by one-dimensional NMR spectroscopy (90% H_2O) at 5° increments from 5 to 70 °C, the terminal base-pairs of the duplex melted by 50 °C, but the imino proton of the guanine adducted to the trimethylene cross-link remained in slow exchange with the solvent and still visible at 70 °C (Figure 1B).

NMR Studies. (a) Assignments of Nonexchangeable Protons. Cross-linked duplex **6** gave well-resolved ^{1}H NMR spectra at 30 °C. A single set of cross-peaks was observed in the NOESY of the trimethylene cross-link, indicating that symmetry was retained in the self-complementary duplex. Sequential assignments of nonexchangeable protons were determined in the standard manner[27,28] and compared with the assignments of nonexchangeable protons in the unmodified deoxyoligonucleotide (Figure 2). The sequential assignment of internucleotide base aromatic protons and sugar H1′ protons showed an interruption between H8 of the adducted G*5 nucleotide and H1′ of the adjacent 5′-nucleotide, C4. A minor overlap problem occurred for G*5 and C6 H1′ resonances, but this did not prevent assignment of other proton resonances for these nucleotides, most of which were well-resolved. The assignments of the H2′ and H2″ sugar protons were made from NOEs and DQF-COSY splitting patterns with the respective H1′ protons. The stronger cross-peaks were assigned to the H2″ resonances on the basis of the assumption that the oligonucleotide conformed to a B-like DNA structure in which the H1′−H2″ distances would be shorter than the H1′−H2′ distances. The splitting patterns between H1′−H2′ and H1′−H2″ in the DQF-COSY corroborated the assignments of H2′ and H2″ protons. Some H4′, H5′, and H5″ resonances were severely overlapped, precluding unambiguous assignments. Chemical shifts of the nonexchangeable protons of the unmodified and modified duplexes are shown in Tables S1 and S2 (Supporting Information).

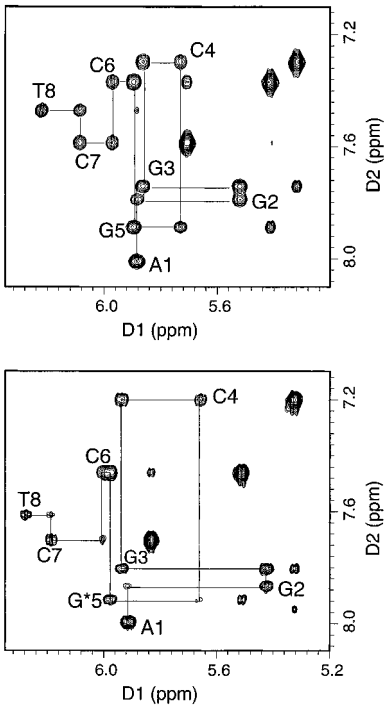


Figure 2. Expanded plot of a phase-sensitive NOESY spectrum at 350 ms mixing time showing the sequential NOEs from the aromatic to the H1′ protons. (Upper plot) Unmodified duplex. (Lower plot) Cross-linked duplex. The base positions are indicated at the intranucleotide aromatic-to-sugar cross-peak.

A notable spectral feature of the cross-linked oligodeoxynucleotide, relative to the unmodified DNA, was the upfield shift of each of the protons of nucleotide C4 (Figure 3). The other residues showed smaller, varying shifts relative to the unmodified oligomer, but the largest and most consistent difference was seen in the sugar protons of C4, the nucleotide adjacent to and base-paired with the adducted G*5.

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

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

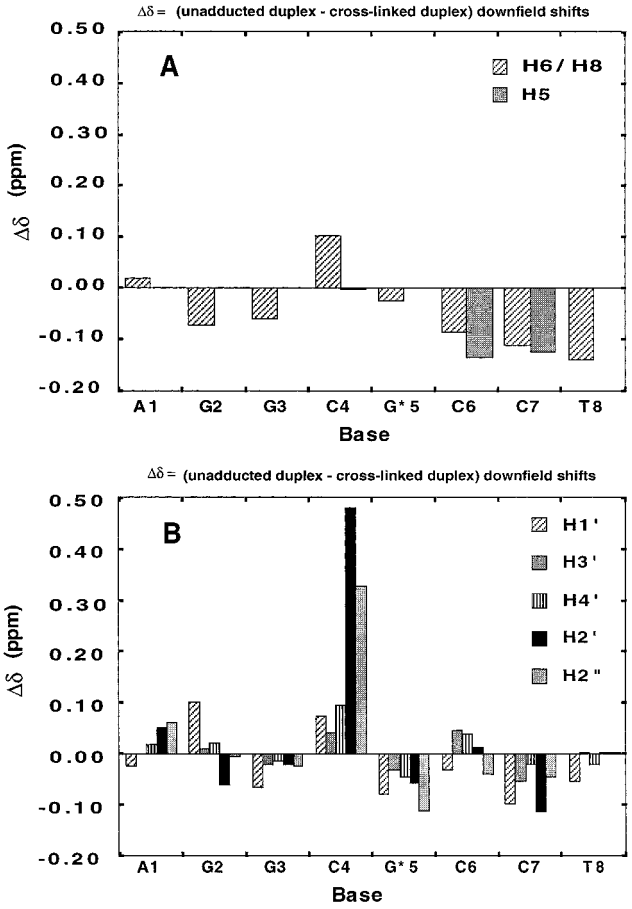


Figure 3. Chemical shift differences of nonexchangeable aromatic and deoxyribose protons of the unadducted and cross-linked oligodeoxynucleotides. (A) Aromatic H5, H6, and H8 protons. (B) Sugar protons. The majority of the structural perturbations are observed in C4, the residue adjacent to and base-paired with the adducted G*5.

(b) Exchangeable Protons. Assignments of the imino and amino protons for the unmodified and cross-linked oligodeoxynucleotides were made from NOESY spectra measured at 5 °C in 90% H_2O.[29] Cross-peaks between cytosine H5 protons and non-hydrogen-bonded cytosine amino protons (NH_{2a}) enabled the assignment of the latter. The assignment of the hydrogen-bonded cytosine amino protons (NH_{2b}) was then made on the basis of cross-peaks to NH_{2a} protons. Cross-peaks between guanine imino protons and both NH_{2a} and NH_{2b} protons of base-paired cytosine nucleotides confirmed the imino proton assignments made in the thermal melting study. Assignments based on the observed order of imino proton melting, expected to be from terminal to internal guanine residues, coincided with the imino proton assignments obtained from the NOESY experiment. No resonances were observed between the terminal base pairs A1•T16 (T8•A9) due to exchange broadening. NOE connectivity was observed from the imino protons of base pairs G2•C15 → G3•C14 → C4•G13. Since the duplex is self-complementary and only one set of cross-peaks is seen, the connectivity can be extended through base pairs G*5•C12 to C6•G11 to C7•G10.

The exocyclic amino proton of G*5 in the cross-linked oligomer was nearly isochronous with C6 NH_{2b}, 8.07 versus 8.09. The resonance did not broaden significantly at increased temperatures, similar to the observation with the imino proton of G*5.

(c) Trimethylene Cross-Link Protons. For simplification, nucleotide designations for only one strand (A^1-T^8) are used in the following discussion of the tether assignments. The resonances arising from the six protons of the trimethylene cross-link were well resolved from any sugar proton resonances. Given the sterically constrained nature of the cross-linked duplex one might expect six signals for the protons in the cross-link if the two strands were not equivalent. However, only three sets of resonances appeared, further indicating a symmetric conformation about the trimethylene cross-link. The furthest upfield resonance (1.9 ppm) was assigned to the central methylene protons (X21, X22), and the other resonances (3.88 and 3.09 ppm) to protons X11, X12, X31, and X32 of the methylene groups attached to the exocyclic amino groups. Dyad symmetry leads to equivalence of X11 and X31 and of X12 and X32. COSY and NOESY spectra confirmed that the resonances at 3.88 and 3.09 ppm arose from geminal protons. In the NOESY spectrum measured in 90% H_2O, a single set of cross-peaks between each of the trimethylene protons and both the imino proton of G*5 and the exocyclic amino proton of G*5 was found (Figure 4A). The trimethylene cross-link protons also exhibited cross-peaks to the H1′ sugar proton of the neighboring 3′-nucleotide C6 and among one another, but not to any other protons in the oligomer (Figure 4B). The intensities of the cross-peaks between the central methylene group protons (1.9 ppm) and each of the vicinal methylene protons adjacent to the exocyclic amino group of G*5 (3.88 and 3.09 ppm) were almost identical and of comparable intensity to the cross-peak of geminal H2′−H2″ sugar resonances.

DQF-COSY Splitting Patterns. For each nucleotide, the splitting patterns of the cross-peaks between H1′ and H2′/H2″ and between H3′ and H2′/H2″ observed in the ^{31}P-decoupled DQF-COSY were compared in the unmodified oligodeoxynucleotide and the trimethylene cross-link (Figure 5). Notable differences for nucleotides C4 and C6 were detected in the patterns of the H3′ and H2′/H2″ cross-peaks, and for C6 in the cross-peak between H1′ and H2′/H2″. All other patterns of cross-peaks were similar between the trimethylene cross-link (Figure 5A and C) and the unmodified oligomer (Figure 5B and D). Closer analysis of the splitting patterns was made using the spectral simulation program SPHINX and LINSHA to determine the δ torsional angle of the backbone sugar residues and the N−C−C−C torsional angle of the tether. Line widths of spectra obtained in 90% H_2O/10% D_2O precluded measurement of coupling constants between the amine proton of G*5 and the vicinal methylene protons.

^{31}P Spectroscopy. The one-dimensional ^{31}P spectrum of the trimethylene cross-link showed seven discernible phosphorus resonances, in marked contrast to the unmodified duplex (Figure 6). The clarity of the ^{31}P signals provided the opportunity to extract ϵ torsional angle information from the $^3J_{HCOP}$ coupling constants observed in the ^{31}P−^{1}H selective excitation COSY spectrum (Figure 7).[30]

NOE Restraints. For the cross-linked oligodeoxynucleotide, a total of 362 experimental intensities, supplemented with calculated intensities from IniA and IniB structures, were used in MARDIGRAS[31] calculations. Distances calculated from poorly resolved, overlapped intensities, from cross-peaks near the water resonance whose intensities were altered by presaturation of the water signal, and from those whose measurements exceeded 5 Å (as calculated by MARDIGRAS) were removed if they were inconsistent with a reasonable structure. The final

(29) Boelens, R.; Scheek, R. M.; Dijkstra, K.; Kaptein, R. *J. Magn. Reson.* **1985**, *62*, 378−386.

(30) Sklenar, V. *J. Am. Chem. Soc.* **1987**, *109*, 7525−7526.

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

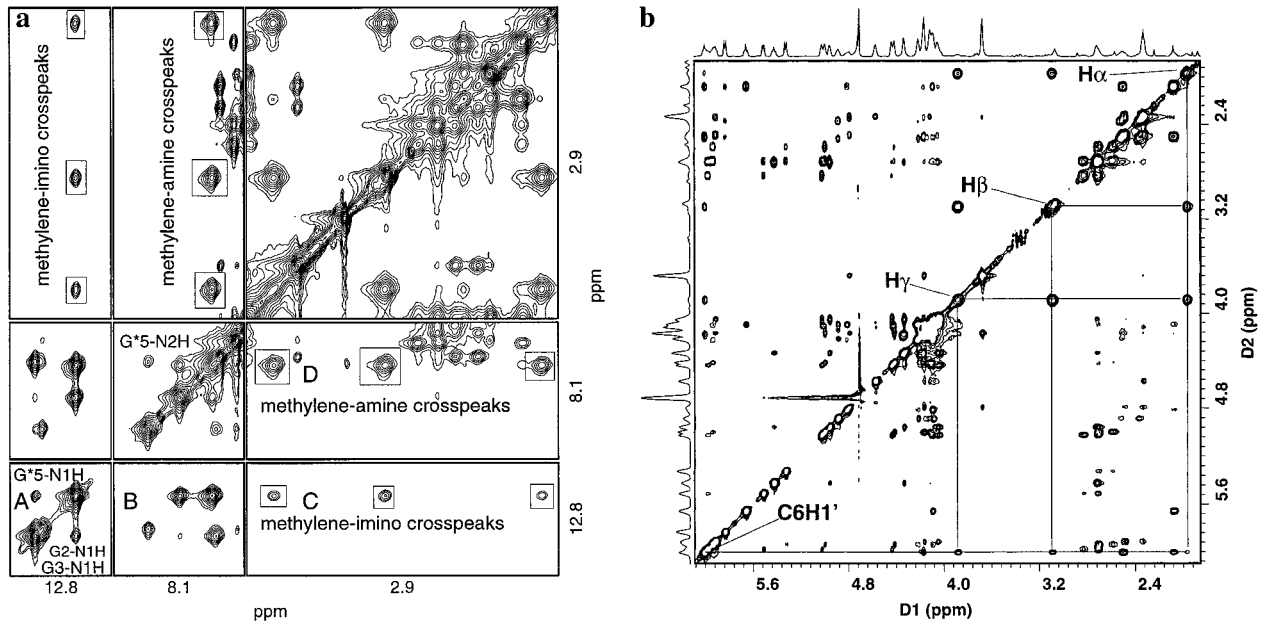


Figure 4. (a) Regions of the 10 °C water NOESY of cross-link 5′-d(AGGCG*CCT)$_2$. Tile A: Guanine imino-to-imino cross-peaks, indicative of base-pairing at this temperature. Tile B: Cross-peaks between base-pairing cytosine amino protons and guanine imino protons. Tile C: Cross-peaks between methylene protons of the cross-link and G*5 imino. Tile D: Amino protons. (b) Expanded plot of a phase-sensitive NOESY spectrum at 350 ms mixing time showing the cross-peaks between the trimethylene cross-link ^{1}H signals and the 5′-neighboring residue C6 anomeric proton. Hα, protons of the central methylene carbon; Hβ and Hγ, protons adjacent to the G*5 amino group.

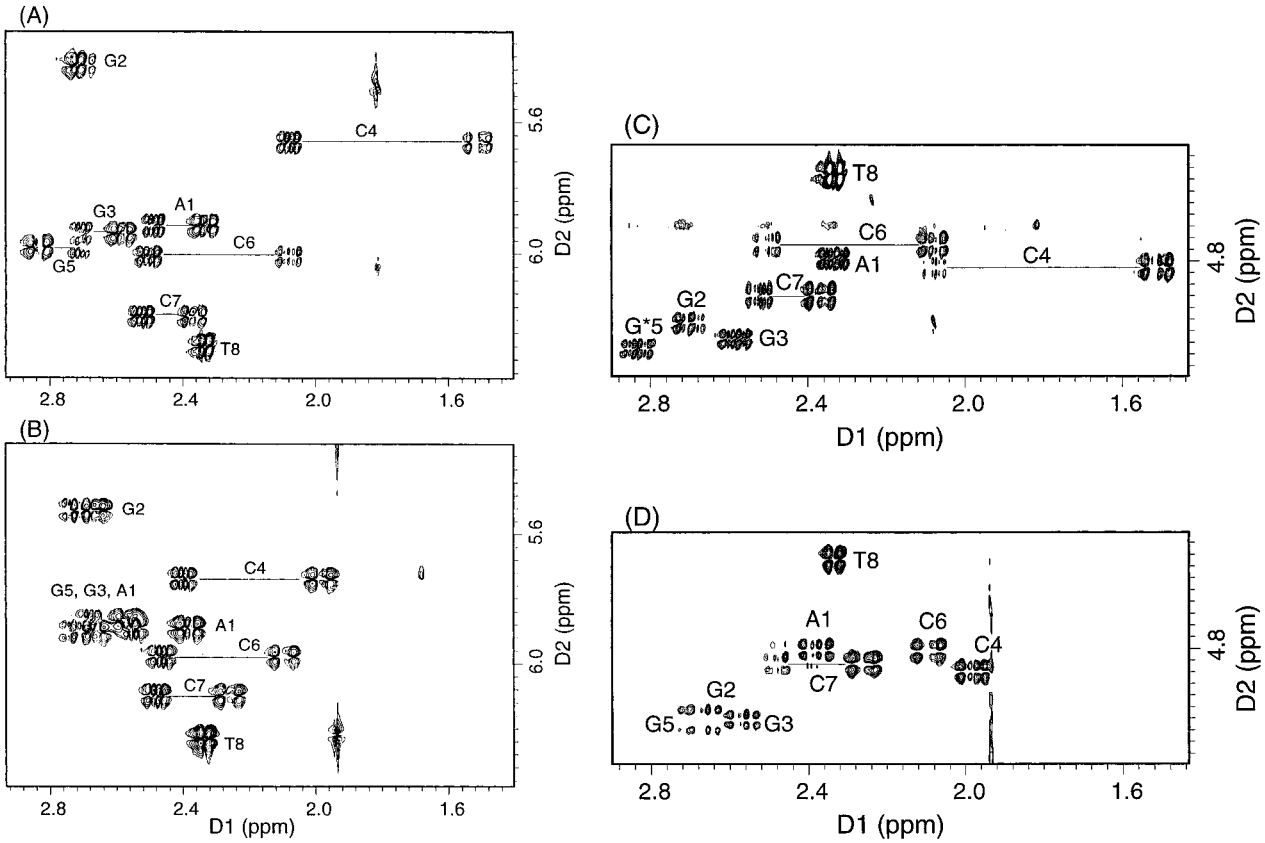


Figure 5. Expanded plot of a ^{31}P-decoupled DQF-COSY showing H1′−H2′/H2″ and H3′−H2′/H2″ regions. H1′ and H3′ signals are along the D2 axis, H2′/H2″ along D1 axis. Base positions are indicated at H1′ or H3′ downfield shift signals. (A) H1′−H2′/H2″ region of trimethylene cross-link. (B) H1′−H2′/H2″ region of unmodified duplex. (C) H3′−H2′/H2″ region of trimethylene cross-link. (D) H3′−H2′/H2″ region of unmodified duplex.

set of 294 interproton distances consisted of 196 intraresidue restraints, 86 interresidue restraints (including 6 restraints between the cross-link and the DNA) (Table S3, Supporting Information), and 12 computer-generated restraints within the trimethylene cross-link based on its W-shape (vide infra). An average of 37 intranucleotide, internucleotide, and empirical distances were obtained for each base pair.

The number of distance restraints was largest at and near the adducted nucleotide, G*5. The neighboring 5′-nucleotide was restrained by 17 intranucleotide and 10 internucleotide restraints,

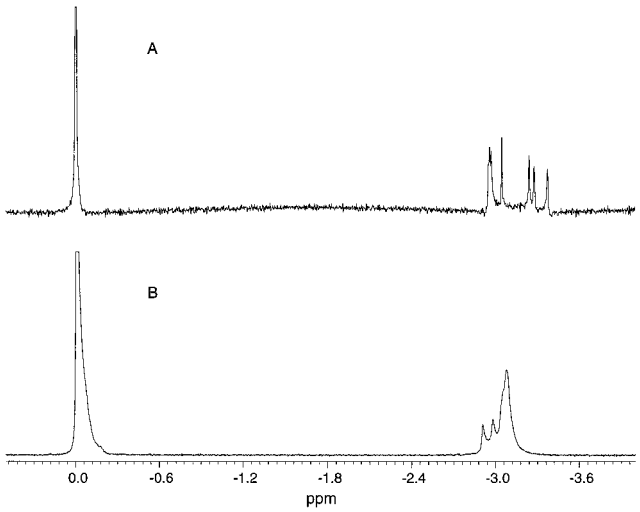


Figure 6. One-dimensional ^{31}P spectra of the oligodeoxynucleotide: (A) cross-linked duplex, showing seven discernible phosphorus resonances; (B) unmodified duplex.

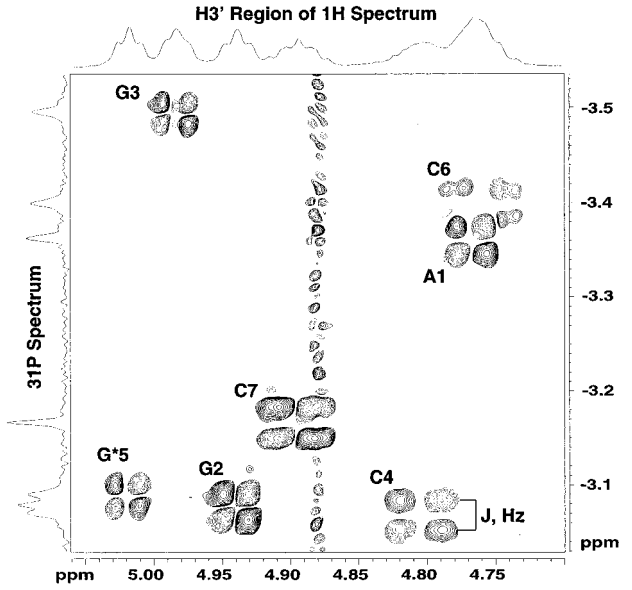


Figure 7. Two-dimensional selective excitation ^{31}P−^{1}H COSY spectrum showing cross-peaks between H3′ protons and 3′-phosphate group ^{31}P resonances. Base positions are indicated adjacent to the cross-peaks.

while the adducted G*5 showed 14 intranucleotide and 3 internucleotide restraints. The NOE distance restraints were separated into four quality sets, of which class 1 contained the 186 best restraints, including the 12 trimethylene cross-link restraints, class 2 contained 38 restraints, class 3 contained 14 restraints, and class 4, the 56 worst restraints (Table S4, Supporting Information).

Sugar Torsion Angle Restraints. (a) Dihedral Backbone Angle *δ*. Vicinal coupling constants between protons on the sugar rings were measured directly from the cross-peaks in the ^{31}P-decoupled DQF-COSY[32] and extracted from simulations. Using experimental and simulated vicinal coupling constants, the pseudorotation phase angle (P) and pucker amplitude (Φ_m) were calculated by the program PSEUROT[33] and used to

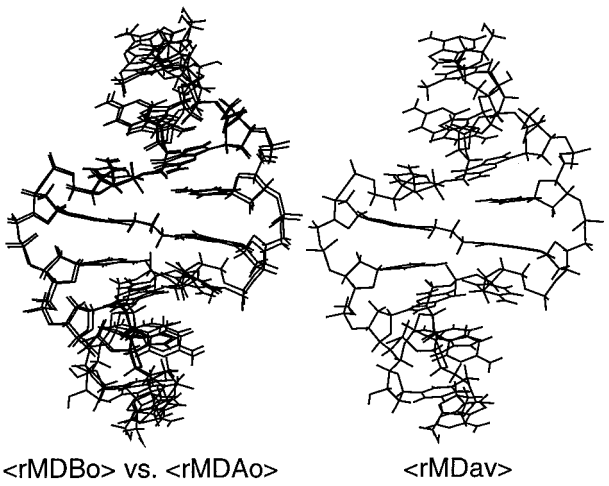


Figure 8. Molecular dynamics-generated structures of 5′-d(AGGCG*CCT)$_2$ cross-link. (A) Superposition of the energy-minimized rMDA and rMDB structures. (B) Average rMDA structure.

determine the dihedral backbone angle δ (O5′−C5′−C4′−C3′)[34] according to the following equation:

$$\delta = 120.6 + [1.1\Phi_m \cos(P + 145.2)]$$

The parameters for the cross-linked oligodeoxynucleotide 5′-d(AGGCG*CCT)$_2$ are in Table S5 (Supporting Information).

(b) Dihedral Backbone Angle ϵ. Torsion angles ϵ (C4′−C3′−O3′−P) were determined from the $^{3}J_{HCOP}$ coupling constants measured in the ^{31}P[^{1}H] selective excitation COSY spectrum. The relationship between the coupling constant $^{3}J_{HCOP}$ and torsion angle ϵ, given by the equation:[35]

$$^{3}J_{HCOP} = [15.3 \cos^2(\epsilon + 120)] - [6.2 \cos(\epsilon + 120)] + 1.5$$

gives four values of ϵ for a single coupling constant. The values for ϵ within the range of preferred nucleotide conformations (−90° to −150°)[34] were selected for use in restrained molecular dynamics calculations (Table S6, Supporting Information).

Structural Refinement. Superposition of six MD-generated structures based on IniA and IniB is shown in Figure S1 (Supporting Information). The final structures obtained by averaging the IniA structures and IniB structures, followed by potential energy minimization, are shown superimposed in Figure 8. The average restrained molecular dynamics structure is depicted in space-filling form at Figure 9. The rms differences on a per-residue basis between IniA and IniB, resulting in a 4.6 Å rms deviation between the widely divergent conformations of cross-linked A-DNA and B-DNA starting structure are shown in Figure S4 in the Supporting Information. The emergent structure was marginally more B-DNA-like than A-DNA-like, as the rmsd comparison to IniB (2.8 Å) was closer than to IniA (5.5 Å). The precision of the emergent structures is seen in the convergence to a common structure, whether starting from IniA or IniB, with an average rms deviation of 0.75 Å for the final average structure and a maximum rmsd of 1.2 Å between the two most divergent structures contributing to the average.

Structural Calculations. The NOE cross-peak intensities measured at a mixing time of 350 ms were compared to intensities derived from RMA of the refined structures[36] using

(32) Maltseva, T.; Sandstrom, A.; Ivanova, I. M.; Sergeyev, D. S.; Zarytova, V. F.; Chattopadhyaya, J. *J. Biochem.* **1993**, *26*, 173−236.
(33) van Wijk, J.; Huckriede, B. D.; Ippel, J. H.; Altona, C. *Methods Enzymol.* **1992**, *211*, 286−306.
(34) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.
(35) Lankhorst, P. P.; Haasnoot, C. A. G.; Erkelens, C.; Altona, C. *J. Biomol. Struct. Dyn.* **1984**, *1*, 1387−1405.
(36) Keepers, J. W.; James, T. L. *J. Magn. Reson.* **1984**, *57*, 404−426.

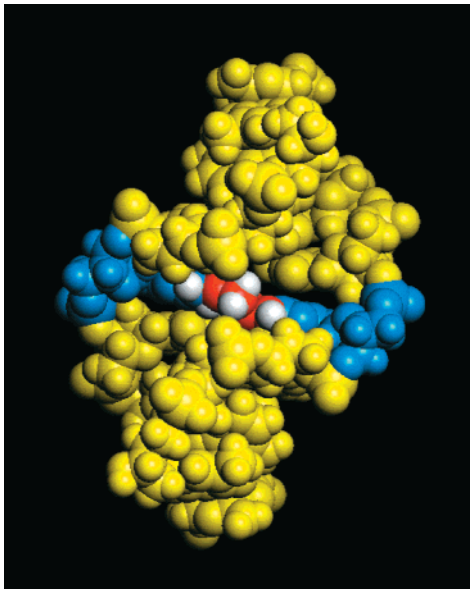


Figure 9. Molecular dynamics-generated structure of 5′-d(AGGCG*CCT)$_2$ cross-link. Space-filling model of the final structure; guanine residues tethering the cross-link are in blue, trimethylene atoms in red.

the sixth-root residual index (R_1^x), as shown in Table S7 (Supporting Information). Consistent results were obtained at each of the three mixing times, with the best values being observed for the 350 ms data. This could be a reflection of the improved sensitivity of the longer mixing time data. The *R* factor for IniA at 350 ms was 0.11, while for IniB the *R* factor was 0.12. In contrast, the *R* factors of the MD structures which emerged from IniA and IniB were 0.048 and 0.047, respectively.

Refined Structure. A comparison of the backbone torsion angles, intra-base-pair parameters, and inter-base-pair parameters for the unmodified and cross-linked oligodeoxynucleotide is available in the Supporting Information (Figures S2 and S3). Comparison was made of the per-residue values of the unmodified oligodeoxynucleotide, the trimethylene cross-link, canonical A-DNA, and canonical B-DNA.

Backbone Torsion Angles. The two torsion angles measured experimentally, δ and ϵ, were compared with the values obtained by DIALS and WINDOWS analysis of the final cross-link structure. Within the range of experimental error, the torsion angle restraints agreed well with the backbone angles of the final structure. The plots of torsion angles δ and ϵ include the experimentally determined values for each residue along with the calculated angles of the final structure.

Accommodating the trimethylene tether between G*5 and G*13 in a manner that results in planarity between the purine bases significantly distorts the backbone. Evidence for large deviations from canonical structures and from the unmodified sequence was seen in the helicoidal analysis of the backbone torsion angles (Figure S2). Among residues C4, G*5, and C6, for example, notable differences are observed between the unmodified oligomer and the cross-linked DNA due to the effect of the tether; differences for residues A1 and T8 are mostly due to fraying at the ends of the DNA.

Intra-Base-Pair Helical Parameters. When the intra-base-pair parameters calculated from the final structure of unmodified and cross-linked DNA were compared (Table S3), the parameters of the unmodified oligodeoxynucleotide more closely resembled canonical B-form DNA, but the cross-link exhibited characteristics unlike either canonical A- or B-form DNA, or the unmodified structure. The cross-link between G*5 and G*13 is accommodated by changes in stretch, stagger, opening, and *y*-displacement for base-pairs C4•G*13 and G*5•C12. The effect of distortion from canonical DNA forms was seen beyond the position of the cross-link in base pairs G3•C14 and C6•G11 in the values of base-pair parameters for stagger, buckle, and propeller twist. Evidence for the planarity of the purine residues in the G*5−G*13 cross-link was seen in the absence of propeller twist at base pairs C4•G*13 and G*5•C12.

Inter-Base-Pair Helical Parameters. In the values of the inter-base-pair parameters for slide, rise, and twist, the deviations from canonical A-DNA, B-DNA, and the unmodified oligodeoxynucleotide were notable, further indicating the distortions that the cross-link induced in the DNA. The distance between base step C4•G*14−G*5•C12, measured in the rise parameter, is very short, since the bases G*5 and G*13 are coplanar via the trimethylene tether. The amount by which base step C4•G*13−G*5•C12 slides apart is seen in the appearance of a bulge in the structure of the DNA, and the slide parameter reflects this distortion. The twist parameter of a base step, which measures the right-hand turn of A- or B-like DNA, shows large deviations in the base steps at and surrounding the cross-link. The amount of helicity in the cross-link is reduced.

Bending Studies. For the determination of bending deformations in DNA duplexes, the method of Koo and Crothers[37,38] employing electrophoretic mobilities is more reliable than NMR spectroscopy. Cross-linked 21-mer **7** containing the 8-mer cross-link sequence (boldface) and the corresponding unmodified 21-mer duplex were phosphorylated on their respective 5′-termini and then ligated to form multimers. Electrophoresis of the ligation products was carried out under nondenaturing conditions (8% polyacrylamide). No significant difference in mobilities was observed between the two samples, indicating that no significant bending was induced by the cross-link. Details of this study are provided in the Supporting Information.

Discussion

Structure of the Cross-Linked Duplex. The overall conclusion is that the DNA duplex is able to accommodate the trimethylene cross-link with minimal distortion. The cross-link, in fact, enhances duplex stabilization. Below, however, we detail the conformational modifications caused by the cross-link. The conformation of the refined structure (Figure 9) of duplex **6** is predominantly B-like DNA, determined on the basis of NOE and torsion angle restraints. Interresidue NOE connectivity between anomeric and aromatic protons is seen with the exception of residues C4 and G*5. Base-pairing interactions are observed among the inner six residues in the NMR data for the imino protons. However, the large variations in backbone torsion angles from one nucleotide to the next reflect the distortions the DNA undergoes to accommodate the cross-link. Additionally, increased shielding seen by the upfield shifts of each C4 proton relative to the unmodified duplex suggests decreased base stacking of C4. Specifically, the refined structure of the duplex shows a decrease in the twist between steps of base-pairs C4•G*13 and G*5•C12 (∼10°) compared to the unmodified duplex (∼30°). The right-handed turn of the helix is disrupted with reduced twisting of the base step, accounting for decreased ring current interactions, seen by the upfield shift of C4 proton NMR signals. The lack of NOE connectivity between C4H1′ and G*5H8 further substantiates the disruption, seen as a bulge in the DNA backbone.

(37) Koo, H. S.; Crothers, D. M. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 1763−1767.

(38) Koo, H. S.; Wu, H. M.; Crothers, D. M. *Nature* **1986**, *320*(10), 501−506.

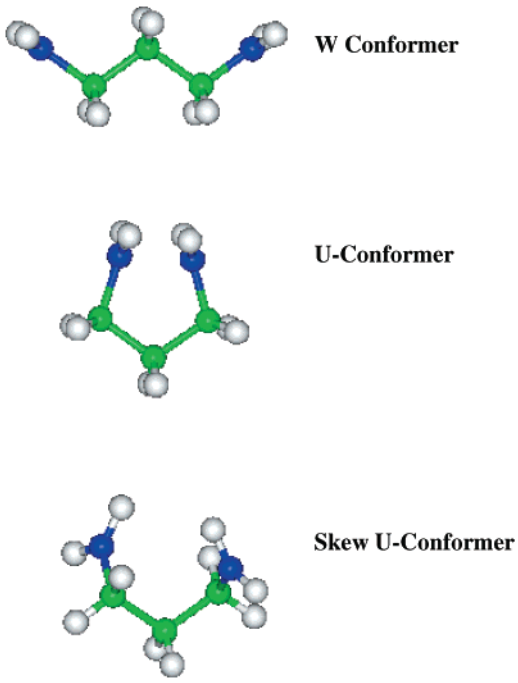


Figure 10. Possible symmetric configurations of the trimethylene tether between cross-linked guanine amino groups. The shape of the tether is defined by the five atoms in the backbone $-NH-CH_2-CH_2-CH_2-NH-$.

Conformation of the Trimethylene Cross-Link Tether. Analysis of the conformation of the tether presented unusual challenges. Dyad symmetry led to a single set of three resonances being observed for the methylene protons in all one-dimensional and two-dimensional NMR spectra. These signals arise from the terminal methylene groups being equivalent to each other but their geminal protons being nonequivalent. At the same time, the geminal protons of the central methylene are equivalent to each other; each NOE cross-peak involving the tether contains contributions from two separate and non-identical pairs of protons. As a consequence, analysis of scalar homonuclear coupling constants and computer modeling were required to determine the conformation of the $NH-CH_2-CH_2-CH_2-NH$ fragment comprising the trimethylene cross-link. The problem was made more complex by the fact that measurement of vicinal couplings to the NH proton was precluded by exchange in D_2O and excessive line width in H_2O. The dyad symmetry constraints limit the five-atom tether to only three conformations, namely planar W- and U-shapes or a helical conformation intermediate between them; these conformations are shown in Figure 10. Analysis was aided by the fact that the diastereotopic methylene protons on the cross-link carbon atom adjacent to the amino group of guanine displayed significantly different chemical shifts (δ 3.88 vs 3.09). This difference in chemical shifts reflects the fact that these protons are in different chemical environments and permitted coupling constants to be determined via a DQF-COSY spectrum. To identify the most likely conformation, the scalar coupling constants of each conformation were estimated from the vicinal Karplus correlation[39] and the appearance of the appropriate cross-peak was simulated with the program SPHINX and LINSHA. The simulated cross-peaks for a W-shape of the cross-link most resembled the experimentally observed DQF-COSY cross-peaks.

(39) Silverstein, R. M.; Bassler, G. C.; Morrill, T. C. *Spectrometric Identification of Organic Compounds*, 5th ed.; John Wiley & Sons: New York, 1991.

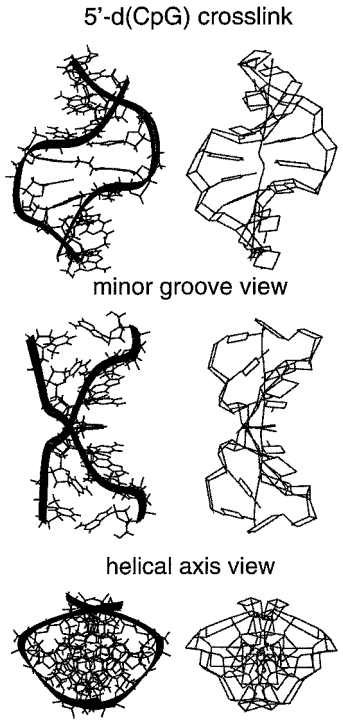


Figure 11. Superposition of the helical axis on the refined rMD structure of cross-linked 5′-d(AGGCG*CCT)$_2$. The corkscrew at the cross-link is the only bend in the axis, which straightens toward the ends of the duplex.

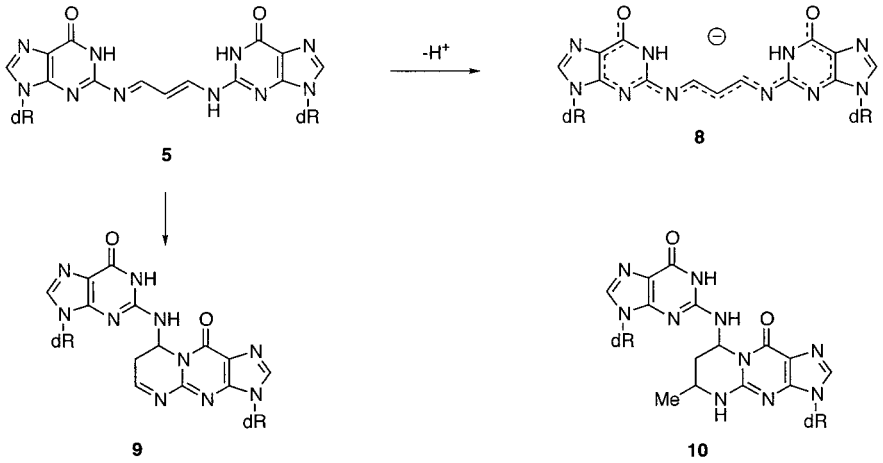
Configuration of the Residues Around the Cross-Link. The W-shaped cross-link permitted the DNA to adopt only a few conformations while still retaining symmetry. In one possible shape, the cross-linked guanines are only minimally shifted from their locations in the un-cross-linked duplex. For each chain, the single exocyclic amino proton would be rotated out of the plane of the purine in the 3′-direction. This shape was rejected because the resulting MD structure placed the tether methylene protons within NOE-observable distance of the anomeric proton of the anchoring deoxyguanosine, a feature not seen in the experimental data. Furthermore, this same structure placed H1′ of the 5′-neighboring C4 residue within NOE-observable distance of the anchoring G*5H8, but NOESY experiments showed a marked absence of a cross-peak between these two protons. Similar application of observed and absent restraints, to include δ and ϵ torsion angles derived from homo- and heteronuclear *J*-coupling constants, were applied to other potential structures bearing a W-shaped trimethylene tether. The only form that fulfilled the experimental observations tilted the anchoring guanines from their normal orientations to bring them into the same plane. Shown in Figure 11 are the helical axis of the cross-link and the rMD structure from which it was derived.

It is noteworthy that in this structure the proton of the guanine exocyclic amino group which has been replaced by the tether is *proximal* to the Watson−Crick edge. This is in contrast to two other oligonucleotides containing N^2-N^2 cross-links for which detailed three-dimensional structures have been elucidated. Patel and Tomasz have obtained the structure of a mitomycin C intrastrand cross-link[40] and Wemmer and Hopkins the interstrand cross-link formed by a distamycin−pyrrole conjugate.[41] Both of these cross-links involve four-carbon tethers

(40) Norman, D.; Live, D.; Sastry, M.; Lipman, R.; Hingerty, B. E.; Tomasz, M.; Broyde, S.; Patel, D. J. *Biochemistry* **1990**, *29*, 2861−2875.

(41) Fagan, P. A.; Spielmann, H. P.; Sigurdsson, S.; Rink, S. M.; Hopkins, P. B.; Wemmer, D. E. *Nucleic Acids Res.* **1996**, *24*, 1566−1573.

Scheme 3



and form spontaneously in CpG sequences; however, the significance of a direct comparison with these two cases is diminished by the fact that those 4-carbon tethers are severely constrained by the cyclic structures of the pyrrole moieties. Nevertheless, it is significant that both of the 4-carbon tethers involve *distal* substitution on the 2-amino groups. It will be instructive to examine the structure of the tetramethylene cross-link because of the likelihood that it will involve the *distal* positions and would not create the bulge seen with the trimethylene tether. A number of other interstrand cross-linked oligonucleotides containing dithiobis(ethane) or dithiobis(propane) links between the exocyclic amino groups of dG, dA, or dC have been synthesized and studied in the Verdine group; the duplexes containing the dithiobis(ethane) linker appear to be considerably more distorted than those with the dithiobis(propane) linker.[42−45] A crystal structure of a self-complementary 10-mer cross-linked through guanine N^2 by a dithiobis(ethane) bridge revealed two quite different structures, one fairly similar to the unmodified duplex and the other quite bent with a partially extruded cytosine; neither structure showed dyad symmetry.[46] The NMR studies indicate the trimethylene cross-linked duplex described in this report has dyad symmetry. However, we cannot rigorously exclude the possibility of rapid equilibration between asymmetric structures.

Validity of Model. Although the MDA interchain cross-link has eluded characterization, a highly probable structure involves Schiff base or enamine formation with N^2 positions of guanines (**5**). Consequently, the roughly planar structure with the tether in a W conformation that has been found for the guanine−trimethylene−guanine fragment of **6** represents a very reasonable model for the MDA cross-link. In all likelihood, the MDA cross-link would be less distorting since the tether would be shortened by the presence of two double bonds. A chemical consequence of planarity in this hypothesized structure for the MDA cross-link is that a low pK_a would be expected since the resulting anion (**8**) could delocalize over both guanine residues (Scheme 3). Imines hydrolyze readily, but in a cross-linked structure analogous to the trimethylene cross-link they might be stable. However, after denaturation of the duplex or enzymatic degradation the imines would be susceptible to hydrolytic cleavage. It is also possible that in denatured DNA or in the tethered dinucleoside the extended cross-link would cyclize to a dihydropyrimidopurinone (**9**) still retaining the guanine tether. Tetrahydropyrimidopurinone **10**, reflecting an N^2-N^2 guanine−guanine interchain cross-link, has recently been detected in enzymatic digests of DNA treated with acetaldehyde; two molecules of acetaldehyde are involved in the creation of the cross-link.[47]

As was mentioned in the Introduction, excision repair of an interstrand cross-link would be an error-prone process because, after excision of the cross-link from one chain, the tether would still be attached to a nucleoside in the template strand. We have collaborated with Sancar and co-workers to examine the repair in mammalian cell extracts of an interchain trimethylene cross-link between guanines in a CpG sequence comparable to the one described in this paper.[48] A surprising result was obtained: a futile repair process occurred in which a 22−28-mer oligonucleotide was excised 5′ to the lesion, missing the site of the lesion. This was followed by refilling the gap. However, the latter process was severely defective in carrying out the final step of ligating the nick. In no case did the repair process remove the tether.

It will be of interest to study other cross-links to see whether this failure of the excision repair process is a general phenomenon. The trimethylene cross-link between guanines in a GpC sequence would be particularly interesting because models suggest this cross-link would cause major distortions in duplex structure. The synthesis used to prepare the trimethylene cross-link is highly versatile with regard to DNA sequence and length of tether and should also be applicable to preparation of interchain cross-links of other types involving the exocyclic nitrogen atoms of adenine and cytosine.

Acknowledgment. We gratefully acknowledge support of this research by the National Institute for Environmental Health Sciences (Grants ES07781 and ES00267). Funding for the NMR spectrometer was supplied by a grant from the NIH shared instrumentation program, (Grant RR05805) and the Vanderbilt Center in Molecular Toxicology (Grant ES00267). We thank Mr. Markus Voehler for valuable assistance with NMR spectroscopy. The authors thank Dr. Stephen S. Hecht (University

(42) Ferentz, A. E.; Verdine, G. L. *J. Am. Chem. Soc.* **1991**, *113*, 4000−4002.
(43) Ferentz, A. E.; Keating, T. A.; Verdine, G. L. *J. Am. Chem. Soc.* **1993**, *115*, 9006−9014.
(44) Ferentz, A. E.; Wiorkiewicz-Kuczera, J.; Karplus, M.; Verdine, G. L. *J. Am. Chem. Soc.* **1993**, *115*, 7569−7583.
(45) Erlanson, D. A.; Glover, J. N. M.; Verdine, G. L. *J. Am. Chem. Soc.* **1997**, *119*, 6927−6928.
(46) van Aalten, D. M.; Erlanson, D. A.; Verdine, G. L.; Joshua-Tor, L. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11809−11814.
(47) Wang, M.; McIntee, E. J.; Cheng, G.; Shi, Y.; Villalta, P. W.; Hecht, S. S. *Chem. Res. Toxicol.* **2000**, *13*, 1149−1157.
(48) Mu, D.; Bessho, T.; Nechev, L. V.; Chen, D. J.; Harris, T. M.; Hearst, J. E.; Sancar, A. *Mol. Cell Biol.* **2000**, *20*, 2446−2454.

of Minnesota) for sharing his observations about the guanine N^2-N^2 cross-link of acetaldehyde prior to publication.

Supporting Information Available: Experimental details of the synthesis of cross-linked duplexes **6** and **7**, for the acquisition and analysis of NMR spectra, and for the measurement of DNA bending and melting, Table S1 and S2 (selected chemical shifts in the unadducted and adducted oligonucleotides), Table S3 and S4 (distance restraints from NOE data), Table S5 (deoxyribose coupling constants in the adducted oligonucleotide), Table S6 (deoxyribose H3′−P 3-bond coupling constants), Table S7 (comparison of sixth root residual indices R_1^x), Figure S1 (superposition of six molecular dynamics structures of the cross-link from calculations using IniB and IniA starting structures), Figure S2 (helicoidal analysis (backbone torsion angles ($\alpha-\zeta$), glycosidic torsion angles (χ), and sugar pseudorotation angles (ϕ)), Figure S3 (helicoidal analysis (intra-base-pair and inter-base-pair parameters for the rMD structures of the unmodified and cross-linked oligodeoxynucleotide)), Figure S4 (per residue comparisons between initial structures and final structures for the cross-linked oligodeoxynucleotide), Figure S5 (electrophoresis gel of ligation ladders), and Figure S6 (graphical analysis of bending study) (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

Accession Codes. The coordinates of the lowest energy structure have been deposited in the PDB (accession codes: RCSB ID is RCSB012701; PDB ID is 1HZ2) together with the restraints used in the structure calculation.

JA003163W