Published online May 20, 2004

*Nucleic Acids Research, 2004, Vol. 32, No. 9* ***2785–2794***
*DOI: 10.1093/nar/gkh606*

# Solution structure of a nitrous acid induced DNA interstrand cross-link

**N. B. Fredrik Edfeldt*, Eric A. Harwood, Snorri Th. Sigurdsson, Paul B. Hopkins and Brian R. Reid**

Department of Chemistry, University of Washington, Seattle, WA 98195, USA

Received January 7, 2004; Revised and Accepted April 22, 2004

## ABSTRACT

**Nitrous acid is a mutagenic agent. It can induce interstrand cross-links in duplex DNA, preferentially at d(CpG) steps: two guanines on opposite strands are linked via a single shared exocyclic imino group. Recent synthetic advances have led to the production of large quantities of such structurally homogenous cross-linked duplex DNA. Here we present the high resolution solution structure of the cross-linked dodecamer [d(GCATCCGGATGC)]$_2$ (the cross-linked guanines are underlined), determined by 2D NMR spectroscopy, distance geometry, restrained molecular dynamics and iterative NOE refinement. The cross-linked guanines form a nearly planar covalently linked 'G:G base pair' with only minor propeller twisting, while the cytidine bases of their normal base pairing partners have been flipped out of the helix and adopt well defined extrahelical positions in the minor groove. On the 5′-side of the cross-link, the minor groove is widened to accommodate these extrahelical bases, and the major groove becomes quite narrow at the cross-link. The cross-linked 'G:G base pair' is well stacked on the spatially adjacent C:G base pairs, particularly on the 3′-side guanines. In addition to providing the first structure of a nitrous acid cross-link in DNA, these studies could be of major importance to the understanding of the mechanisms of nitrous acid cross--linking and mutagenicity, as well as the mechanisms responsible for its repair in intracellular environments. It is also the shortest DNA cross-link structure to be described.**

## INTRODUCTION

Nitrous acid is a mutagenic agent that exhibits two types of chemical reactions with DNA: firstly it converts exocyclic amino groups of DNA heterocycles to carbonyl groups (1,2) and secondly it produces interstrand cross-links in duplex DNA (3,4). Nitrous acid is formed from nitrites under acidic conditions and thus could form in the stomach, yet nitrites are common food additives used in the preparation of cured meats (5). Both deamination and cross-linking are believed to proceed by diazotization of an exocyclic amino group, followed by displacement by a nucleophile, which may be water (deamination) or the amino group of another DNA residue (cross-linking). It has been estimated that for every four deaminations one interstrand cross-link is formed (6). Interstrand cross-links are induced preferentially at d(CpG) steps (7,8), where they link the two guanines on opposite strands via a single shared exocyclic imino group (9; Fig. 1A). The observed CpG sequence preference is probably due to the close proximity of the exocyclic amino group of the guanine on one strand to the staggered guanine diazonium ion intermediate on the complementary strand (10,11).

Due to simultaneous deamination reactions with the three bases containing exocyclic amino groups, simple exposure of duplex DNA to aqueous nitrous acid is not a viable method for producing homogenous interstrand dG–dG cross-linked DNA (8). However, the synthesis of a deoxyguanosine–deoxyguanosine cross-linked phosphoramidite, which can be introduced into self-complementary oligonucleotides during normal automated solid-state synthesis, has recently led to the ability to produce large quantities of homogeneous samples containing a single dG–dG cross-link (12). Our preliminary structural characterization of such a duplex indicated that normal base pairing of the cross-linked guanines is disrupted and that these residues no longer pair with their complementary cytosine partners (12). Here we present detailed 2D NMR data of the symmetrical interstrand cross-linked dodecamer duplex [d(GCATCCGGATGC)]$_2$ (where the cross-linked guanines are underlined, Fig. 1B), together with the high resolution solution structure based on distance geometry, restrained molecular dynamics and iterative NOE refinement. Since the CCGG sequence context is the most efficient CpG context for forming cross-links *in vitro* (7), it was chosen for the present structural studies. In addition to providing the first structure of a nitrous acid DNA cross-link, these studies are of major importance to the understanding of the mechanisms of nitrous acid cross-linking and mutagenicity, as well as the mechanisms responsible for its repair in intracellular environments which are currently poorly understood.

*To whom correspondence should be addressed at present address: AstraZeneca Structural Chemistry Laboratory, AstraZeneca R&D Mölndal, 431 83, Mölndal, Sweden. Tel: +46 31 776 1604; Fax: +46 31 776 3792; Email: fredrik.edfeldt@astrazeneca.com
Present address:
Eric A. Harwood: Chiron Corporation, 201 Elliott Avenue West, Suite 150, Seattle, WA 98119, USA

## MATERIALS AND METHODS

### Oligonucleotide synthesis and purification

The cross-linked self-complementary dodecamer duplex [d(GCATCCGGATGC)]$_2$ was designed based on the results of earlier studies which indicated that the d(CCGG) sequence context is the most efficient in forming dG–dG cross-links *in vitro* (8). The self-complementary nature of the oligonucleotide is a necessity of the synthetic methodology employed, but also serves to simplify the NMR spectra due to $C_2$-symmetry. The synthesis of the duplex was performed on an Applied Biosystems 390 DNA synthesizer as described previously (12). Briefly, normal 3′ to 5′ synthesis was carried out using a heavily loaded deoxycytidine-linked solid support (Glenn Research), containing 2.5 µmol of 3′-base instead of the typical 1 µmol. A symmetrical cross-linked deoxyguanosine–deoxyguanosine phosphoramidite was introduced in the fifth coupling step, thereby joining two growing parallel strands. After normal completion of the synthesis and standard deprotection, the crude DNA was purified using 20% denaturing polyacrylamide gel electrophoresis. The sample was further purified using reverse phase HPLC, then desalted using a Sephadex G-15 column eluting with $H_2O$.

### NMR sample preparation

The purified oligonucleotide was dissolved in a buffer solution consisting of 200 mM NaCl, 10 mM sodium phosphate, 1 mM EDTA, with the pH adjusted to 7.0. The DNA concentration was in the 0.75–1.5 mM range for all experiments. Experiments were carried out in either 90% $^1H_2O$/10% $^2H_2O$ or 100% $^2H_2O$, with a sample volume of 0.4 ml. For experiments in $^2H_2O$ the samples were redissolved in 0.4 ml 99.996% $^2H_2O$ (Isotec Inc.) after repeated lyophilization from 99.96% $^2H_2O$. For low pH experiments, 1 µl aliquots of 1 N HCl were added sequentially to obtain a pH of 5.0. For experiments in the absence of buffer the sample was desalted using a Sephadex G-15 column and resuspended in a solution consisting of 200 mM NaCl, 1 mM EDTA, with the pH adjusted to 7.0.

### NMR spectroscopy

NMR data were collected on Bruker DRX-500 and DMX-750 spectrometers, and processed on SGI workstations with FELIX 95.0 software (Biosym/MSI). 750 MHz 1D $^1H$ spectra and 2D $^1H$–$^1H$ NOESY spectra were collected in 90% $^1H_2O$/10% $^2H_2O$ over a range of temperatures, and in addition both at pH 5.0 and in the absence of buffer to slow down solvent exchange. The WATERGATE pulse sequence was used for water suppression (13). $^1H_2O$-NOESY experiments were carried out with a mixing time of 150 ms, using 4096 complex points in $t_2$ and 1000 experiments in $t_1$, averaging 32 or 48 scans per $t_1$ incrementation. 2D $^1H$–$^1H$ NOESY experiments in 99.996% $^2H_2O$ were collected at several different temperatures, with 2048 complex points in $t_2$ and 800 experiments in $t_1$, averaging 32 scans per $t_1$ incrementation, and zero filled to 2048 points in both dimensions. All NOESY experiments were carried out in the phase-sensitive TPPI mode and processed with 90° phase-shifted sinebell-squared window functions in both dimensions. To measure the NOE build-up, spectra were collected contiguously at 750 MHz and 25°C using 60, 120, 180, 240 and 360 ms mixing times. The 750 MHz DQF–COSY spectrum was collected at 25°C in the phase-sensitive TPPI mode using 1024 complex points in $t_2$ and 800 experiments in $t_1$, with 32 scans, and processed using 6 Hz of Gaussian line narrowing in $t_2$ and a 90° phase-shifted sinebell-squared window function in $t_1$. A relaxation delay of 2 s was used in both the NOESY and the COSY experiments. A 500 MHz proton-detected $^1H$–$^{31}P$ HETCOR spectrum (14) was collected at 25°C in the hypercomplex mode, with 2048 complex points in the $t_2$ ($^1H$) dimension and 200 complex points in the $t_1$ ($^{31}P$) dimension, and 256 scans per $t_1$ incrementation, and processed with a 6 Hz Gaussian line narrowing window function in $t_2$ and a 90° phase-shifted sinebell function in $t_1$.

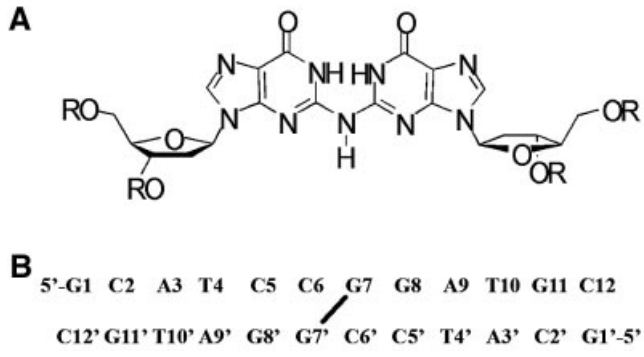



**Figure 1.** (**A**) Structure of the nitrous acid induced DNA interstrand cross-link. (**B**) Schematic representation of the cross-linked dodecamer duplex used in these studies with the residue numbering scheme. Although the duplex is symmetric the residues of one of the strands are designated with a prime (′) for convenience when describing interstrand interactions. Note in particular the cross-linked guanines (G7 and G7′), and the cytosines preceding the cross-link, C6 and C6′, which would base pair with G7′ and G7, respectively, in normal B-DNA.

### Structure determination, refinement and analysis

The Biopolymer module of the Insight II software (Biosym/MSI) was used to build a canonical B-DNA molecule and assigning AMBER ff94 potentials (15). The guanine–guanine cross-link was created by rotating the glycosidic bond of G7 and G7′ making them co-planar (with a $\chi$ angle of ~110°), then deleting the amino group of one of the guanines and introducing a new covalent bond to the other guanine. The modified shared NH-group was assigned AMBER potentials with the proper hybridization and bond order: 'N2 potential' ($sp^2$) for the N2 atom and 'H1 potential' ($sp^3$) for the H2 atom. The partial charges were assigned to –0.553 and 0.325 for N2 and H2, respectively, leaving the net charge on the molecule unchanged. The cross-link was initially modeled with both guanines in the keto form, using standard AMBER potentials and partial charges for N1, H1 and O6. The cross-link was subsequently modeled with one guanine in the keto form and one in the enol form (see Results). For the enol form H1 was removed, and the hydroxyl group was assigned AMBER potentials: 'OH potential' for the O6 atom and 'HO potential' for the H6 atom with the partial charges of –0.458 and 0.336, respectively, leaving the net charge unchanged. The bond order was changed accordingly. In addition to $sp^2$

hybridization, the N2 of the cross-link was also modeled with $sp^3$ hybridization, which had essentially no effect on the structure.

The structure was determined using NOE distance and dihedral angle restrained distance geometry and molecular dynamics and iterative NOE refinement, essentially as described previously (16–19). Distance restraints for non-exchangeable protons were derived from well resolved NOESY crosspeaks in the $^2H_2O$ spectra, which were integrated over 60, 120, 180, 240 and 360 ms mixing times. Initial distances were obtained by scaling the intensities to the average initial rates of cross-relaxation for the cytosine H5–H6 proton pairs (r = 2.5 Å), or, for distances involving methyl protons, using the thymine H6–$CH_3$ pseudoatom distance of 3.0 Å. In addition, restraints involving overlapped crosspeaks were based on visual inspection of the build-up rate in the five spectra and given bounds with a wide range (>1 Å). Restraints involving exchangeable protons were generated based on the $^1H_2O$-NOESY spectrum, also with generous bounds (1.5–2 Å wide). A set of tight hydrogen bonding distance restraints was used to maintain base pairing, where convincing evidence for such hydrogen bonds was observed (see Results). Based on experimental evidence the phosphate backbone was restrained to: non-*trans* (0° ± 150°) for α and ζ; *trans* (180° ± 30°) for β and ε (except C6 β, C5 ε and C6 ε, which were not restrained); and *gauche*$^+$ (60° ± 30°) for γ [except C6 γ, which was restrained to *trans* (180° ± 30°)] (see Results). Chirality restraints were imposed on the asymmetric sugar carbons. These restraints were used as input for distance geometry calculations using the DGII program (Biosym/MSI). Ten initial distance geometry structures were generated by embedding the smoothed initial bounds matrix followed by several cycles of simulated annealing. The structures were then subjected to a combination of restrained conjugate gradient energy minimization and restrained molecular dynamics, using the DISCOVER program (Biosym/MSI). After 1000 steps of conjugate gradient minimization, 5000 cycles (5 ps) of molecular dynamics were carried out *in vacuo* at 300 K with a step size of 1.0 fs using a distance-dependent dielectric constant of 4r, no counter ions and no cut-off distance for the non-bonded interactions. The last coordinate set was subjected to another 2000 steps of conjugate gradient minimization. The force constant was 30.0 kcal mol$^{-1}$ Å$^{-2}$ for the distance restraints and 5.0 to 60.0 kcal mol$^{-1}$ rad$^{-2}$ for the dihedral restraints. The structure with both the lowest total and restraint energy was chosen for further iterative refinement. The NOESY spectra of this structure were subsequently back-calculated using the NOESY simulation program BIRDER (20), with an empirically determined correlation time of 4.3 ns (using the cytosine H5–H6 pairs). The distance restraints were then adjusted by comparing the calculated and experimental NOE intensities in all spectral regions at the five mixing times, either quantitatively using the program FELIXBAD (L. Zhu, University of Washington, Seattle, WA) for the well resolved crosspeaks or qualitatively for the overlapped crosspeaks. New distance restraints were determined and given a range of ± 10%. Repulsive restraints were introduced as needed to maintain certain interproton distances. This iterative DG/rMD/relaxation matrix simulation refinement procedure was repeated until the back-calculated spectra matched the experimental spectra with no further improvement in the NOE *R*-factor ($R_{NOE} = 1/N\ \Sigma|I_e - I_c| / \Sigma I_e$, where $I_e$ and $I_c$ are the experimental and calculated NOE intensities, respectively) for the resolved crosspeaks, in the appearance of overlapped crosspeaks, and in the total and forcing potential energy terms for the duplex. After numerous iterations the intensities of most crosspeaks were reproduced. In the final iteration a total of 715 distance restraints were used: 226 were intraresidue; 334 interresidue; 103 repulsive; and 52 were hydrogen bonding. A set of 40 final distance geometry structures were generated, 20 with G7 in the enol form and 20 with G7′ in the enol form, which were subjected to 2000 steps of conjugate gradient minimization, 10 000 cycles (10 ps) of molecular dynamics, and finally 4000 steps of conjugate gradient minimization. Twenty-six of the 40 structures converged. The $R_{NOE}$ factor was calculated for 131 resolved crosspeaks in the five NOESY spectra for a total of 655 crosspeak intensities for all final structures. The average energy terms were calculated and the r.m.s. deviations were determined by superposition of all atoms, excluding hydrogens, for each structure onto the lowest energy structure. The dihedral angles and structure parameters were calculated using the programs NEWHELIX (R. E. Dickerson, University of California, Los Angeles, CA) and RNA (M. S. Babcock, Rutgers, The State University of New Jersey, New Brunswick, NJ), respectively. The RNA program is flexible enough to account for the guanine–guanine cross-link and the extrahelical cytidines. Statistics were calculated based on all 26 final structures.

## RESULTS

### Exchangeable proton studies

The downfield region of the 1D proton NMR spectrum of the cross-linked duplex, collected in $^1H_2O$ is shown in Figure 2. There are only five imino proton resonances, indicating that the self-complementary duplex forms a symmetrical structure with each resonance corresponding to two protons, and that the two strands are equivalent on the NMR time scale. The imino protons were assigned unambiguously based on the 2D $^1H_2O$-NOESY spectrum, shown in Supplementary Figure S1. Residues G1, T4, G8, T10 and G11 form normal Watson–Crick base pairs with their expected hydrogen bonding partners. However, the N1-imino protons of the cross-linked guanines (G7 and G7′) are not observed in either the 1D or 2D spectra, and have not been observed at any temperature, nor when solvent exchange is limited by lowering the pH to 5.0 or by removing the buffer (data not shown). This observation is both intriguing and surprising. In addition, the amino protons of the complementary cytosine partners (C6 and C6′) are not observed. These amino protons would be protected from solvent exchange and give intraresidue crosspeaks to H5, if C6 were hydrogen bonded. Thus, these observations clearly confirm the absence of G7:C6′ and G7′:C6 hydrogen bonding. Another possibility that must be considered is hydrogen bonding between the two cross-linked guanines, which is plausible given the short length of the cross-link. An imino proton in such a hypothetical 'G:G base pair' would be expected to appear as a sharp resonance at 12.5–13.5 p.p.m. if the hydrogen bond acceptor was nitrogen, or at ~11 p.p.m. if the acceptor was oxygen (21,22). The latter can immediately be ruled out in this case based on simple geometry: the

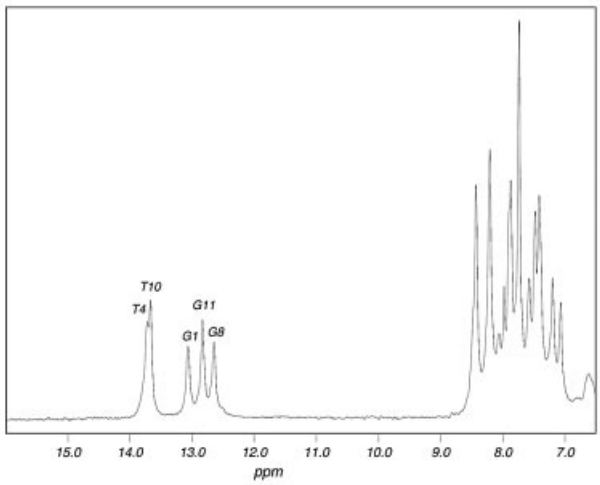



**Figure 2.** The downfield region of the 1D proton NMR spectrum, collected at 0°C, showing the five imino proton resonances corresponding to the ten stable base pairs flanking the cross-link. Note the absence of an N1-imino proton resonance for the cross-linked guanines (G7).

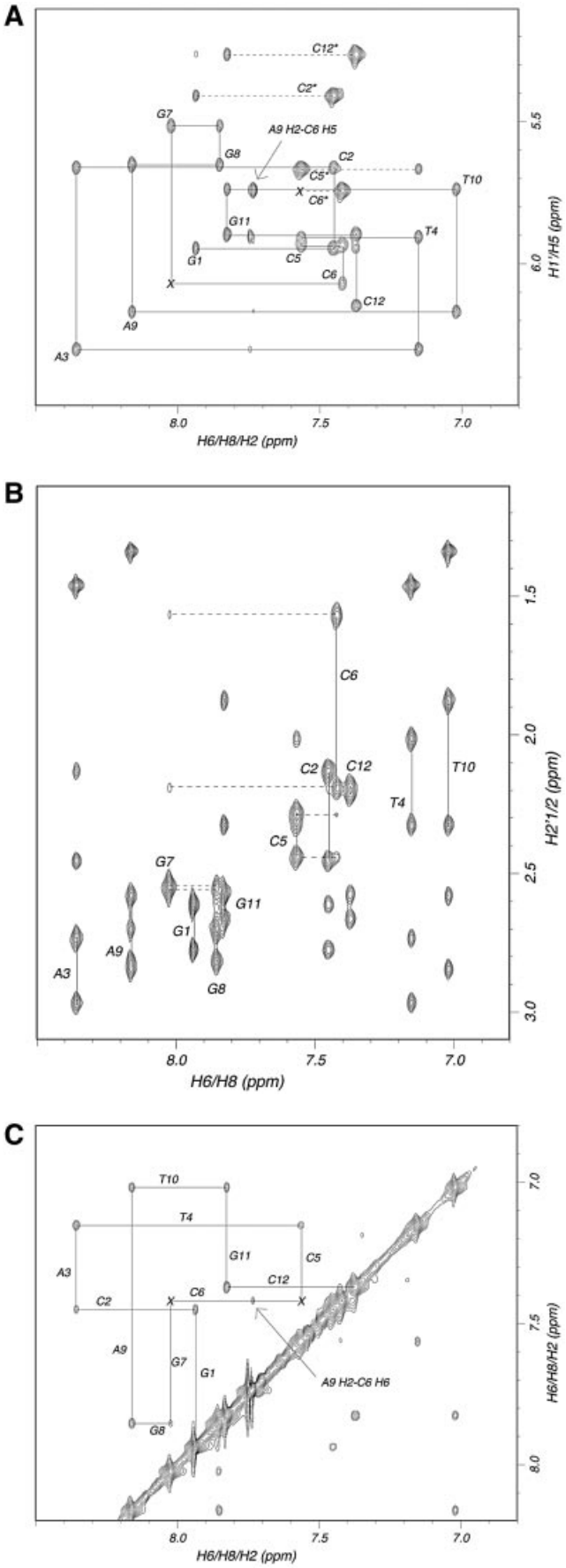
distance between H1 and the closest oxygen (O6) is always too great to allow hydrogen bonding. On the other hand, guanine imino protons that are not hydrogen bonded typically appear as broad resonances near 10 p.p.m. (23). In either case, the complete absence of a G7 imino proton resonance indicates that these protons are exchanging too rapidly with the solvent to be observed.

### Non-exchangeable proton studies

2D $^2H_2O$-NOESY spectra were initially collected over a wide range of temperatures (0–40°C), to assess temperature dependence on the structure of the duplex. No major temperature effect was observed, other than the expected line broadening at low temperature and end-fraying effects at higher temperature (data not shown). The highest quality spectra were obtained at 25°C, and thus this temperature was chosen for the structure determination. All non-exchangeable protons (with the exception of some overlapped H5′1 and H5′2 resonances) were unambiguously assigned using 2D $^2H_2O$-NOESY spectra in conjunction with DQF–COSY and $^1H$–$^{31}P$ HETCOR spectra. The chemical shifts are given in Supplementary Table 1. Figure 3 shows the H6/H8–H1′/H5 (panel A), the H6/H8–H2′1/H2′2 (panel B) and the aromatic–aromatic (panel C) regions of the $^2H_2O$-NOESY spectrum. There are twelve sets of resonances for the duplex, which means that each resonance corresponds to two symmetrical protons. This confirms that the two strands are equivalent on the NMR time scale and that there is $C_2$-symmetry about the central cross-link. The intensities of the intraresidue aromatic to H1′ crosspeaks (Fig. 3A), which are all much weaker than the cytosine H5–H6 crosspeaks, establish that all χ glycosidic torsion angles, including G7, are in the typical *anti* conformation. The relatively weak intrasugar H6/H8–H3′, H1′–H4′ and H2′2–H4′ crosspeaks in the NOESY spectrum, and the strong intrasugar H1′–H2′1 ($^3J_{H1'-H2'1}$) but weak H2′2–H3′ ($^3J_{H2'2-H3'}$) and H3′–H4′ ($^3J_{H3'-H4'}$) crosspeaks in the DQF–COSY spectrum, suggest that all residues adopt typical *C2′-endo* type sugar conformations (data not shown). The normal sequential connectivities for all residues except C5, C6 and G7, suggest that the duplex overall adopts a B-DNA type structure.

Several notable features for the residues in the cross-link region are apparent in the $^2H_2O$-NOESY spectrum. There are normal G7 H1′–G8 H8, G7 H2′1–G8 H8, G7 H2′2–G8 H8 and G7 H8–G8 H8 connectivities, which show that the cross-linked G7 is stacked well on the adjacent G8 base. These observations, in conjunction with the fact that G7 adopts the anti glycosidic conformation, suggest that the two cross-linked guanines are facing each other in the helix, in some sort of head-to-head orientation. In addition, the absence of the C6 H1′–G7 H8 and the C6 H6–G7 H8 connectivities, in conjunction with the weak C6 H2′1–G7 H8 and C6 H2′2–G7 H8 connectivities, indicate that the C6–G7 base–base stacking is disrupted. In spite of the normal C5 H1′–C6 H6 connectivity, the weak C5 H2′1–C6 H6 and C5 H2′2–C6 H6 connectivities and the absence of the C5 H6–C6 H5 and C5 H6–C6 H6 connectivities show that the C5–C6 base–base stacking is disrupted. Furthermore, there are highly unusual C6 H5–A9′ H2 (Fig. 3A) and C6 H6–A9′ H2 (Fig. 3C) connectivities which, given the overall helical structure of the molecule, must be interstrand connectivities. C6 and A9 are too far apart in sequence to give rise to intrastrand crosspeaks. The proximity of C6 to A9′ and to its hydrogen bonding partner T4 is further supported by the unusual intrastrand T4 H1′–C6 H5 connectivity (see Supplementary Figure S2). There is also a strong C5 H1′–C6 H5 connectivity. In combination, all of these observations strongly suggest that the C6 base is flipped out of the helix and adopts an extrahelical position. This extrahelical base must be located in the minor groove (close to residue A9′), with its hydrophobic (H5–H6) edge turned towards the core of the helix (and the H2 of A9′).

There are several NOEs indicating that the phosphate backbone adopts an unusual conformation at the C6 residue: the highly unusual C6 H5′1–G7 H8 and C6 H5′2–G7 H8 connectivities, which are typically not observed in B-DNA; and the unusually strong intraresidue C6 H6–H5′2 connectivity, which is stronger than the intraresidue C6 H6–H5′1 connectivity (see Supplementary Figure S3). Further information about the phosphate backbone can be derived from the $^1H$–$^{31}P$ HETCOR spectrum shown in Figure 4, in which interresidue three-bond (*n*)P – (*n* – 1)H3′ ($^3J_{P-H3'}$) and intraresidue four-bond (*n*)P – (*n*)H4′ ($^4J_{P-H4'}$) crosspeaks appear. The $^4J_{P-H4'}$ coupling is observed only when the P–O5′–C5′–C4′–H4′ linkage lies in a plane forming a W-shaped conformation (24,25), as is the case in B-DNA where the β and γ torsion angles are in the *trans* and *gauche*$^+$ conformations, respectively. There are readily detectable $^4J_{P-H4'}$ crosspeaks for all residues except C6, indicating that all other residues, including the cross-linked G7, adopt the typical β(*t*) γ(*g*$^+$) conformation. The absence of the C6 crosspeak could be due either to a deviation of its β torsion angle from the *trans* domain or of its γ torsion angle from the *gauche*$^+$ domain, or both. The γ torsion angle can be monitored via the relative $^3J_{H4'-H5'1}$ and $^3J_{H4'-H5'2}$ couplings in the DQF–COSY spectrum, and the relative intrasugar H4′–H5′1 and H4′–H5′2 NOEs (16,26). The H3′/H4′/H5′ region of the DQF–COSY spectrum (see Supplementary Figure S4A), shows that the $^3J_{H4'-H5'2}$ coupling is strong for C6 while its $^3J_{H4'-H5'1}$




coupling is absent, which is consistent only with the $\gamma(t)$ conformation. For $\gamma(g^-)$ torsion angles the $^3J_{H4'-H5'1}$ coupling would be strong and the $^3J_{H4'-H5'2}$ coupling absent, and for the normal $\gamma(g^+)$ both the $^3J_{H4'-H5'1}$ and $^3J_{H4'-H5'2}$ couplings would be absent. In addition, the 60 ms $^2H_2O$-NOESY spectrum (see Supplementary Figure S4B), shows that the intraresidue C6 H3′–H5′1 crosspeak is larger than the C6 H3′–H5′2 crosspeak, and the intraresidue C6 H4′–H5′1 crosspeak is larger than the C6 H4′–H5′2 crosspeak. This again is consistent only with the $\gamma(t)$ conformation. The β torsion angle can be determined from the relative intraresidue $^3J_{P-H5'1}$ and $^3J_{P-H5'2}$ couplings (16). As shown in the $^1H-^{31}P$ HETCOR spectrum, the C6 P–H5′1 ($^3J_{P-H5'1}$) and P–H5′2 ($^3J_{P-H5'2}$) crosspeaks are of equal intensity, which is consistent with a C6 β angle in either the *trans* conformation (~180°) or *cis* conformation (~0°). Strong $^3J_{P-H5'1}$ and weak $^3J_{P-H5'2}$ couplings would be indicative of β angles in either the low end of the *trans* (120–150°) or the *gauche*$^-$ (–30 to –60°) conformation, whereas weak $^3J_{P-H5'1}$ and strong $^3J_{P-H5'2}$ couplings would be indicative of β angles in either the high end of the *trans* (–120 to –150°) or the *gauche*$^+$ (30–60°) conformation. Additional information can be derived from the phosphorus chemical shifts, which depend on the phosphodiester torsion angles ζ and α. Both ζ and α typically adopt the *gauche*$^-$ conformation, resulting in chemical shifts in the range of –3.8 to –4.8 p.p.m. The $\alpha(t)$ and $\zeta(t)$ conformations, on the other hand, both lead to significant downfield shifts (27). As shown in Figure 4, all phosphorus resonances are in the normal range. However, these chemical shifts do not rule out the unusual $\zeta(g^+)$ and $\alpha(g^+)$ conformations, which must therefore still be considered for the extrahelical C6 residue in particular, given the numerous unusual NOEs. The ε torsion angle is difficult to determine experimentally from purely *J*-coupling observations (26), but the *gauche*$^+$ conformation can be ruled out based on the absence of any unusually large $J_{P-H3'}$ couplings (Fig. 4). This observation and the absence of four bond $J_{P-H2'}$ coupling, which has been observed for sugars in the normal *C2′-endo* conformation when ε is in the *gauche*$^-$ conformation (28), suggest that all ε torsion angles are in the normal *trans* conformation.

**Figure 3.** (**A**) Assignment of the H6/H8–H1′/H5 region of the $^2H_2O$-NOESY spectrum, collected at 25°C with a mixing time of 240 ms. The sequential aromatic to H1′ interresidue walk is indicated with lines and the intraresidue H6/H8–H1′ connectivities are labeled with the corresponding residue name and number. The absence of the C6 H1′–G7 H8 and C5 H6–C6 H5 connectivities are both marked *X*, and the interstrand C6 H5–A9′ H2 connectivity is labeled and indicated with an arrow. The unexpected G1 H8–C12 H5 and G1 H1′–C12 H6 crosspeaks are suggestive of end-to-end aggregation, although the experimentally determined correlation time of 4.3 ns is clearly indicative of a monomer dodecamer duplex (45). (**B**) Assignment of the H6/H8–H2′1/H2′2 region of the $^2H_2O$-NOESY spectrum, collected at 25°C with a mixing time of 240 ms. The intraresidue aromatic to H2′1/H2′2 connectivities are labeled and indicated with lines, and the sequential aromatic to H2′1/H2′2 interresidue walk for residues C5–G8 is indicated with dashed lines. (**C**) Assignment of the aromatic region of the $^2H_2O$-NOESY spectrum, collected at 25°C with a mixing time of 360 ms, showing the sequential interresidue walk. The absence of the C5 H6–C6 H6 and C6 H6–G7 H8 connectivities are both marked *X*, and the interstrand C6 H6–A9′ H2 connectivity is labeled and indicated with an arrow.

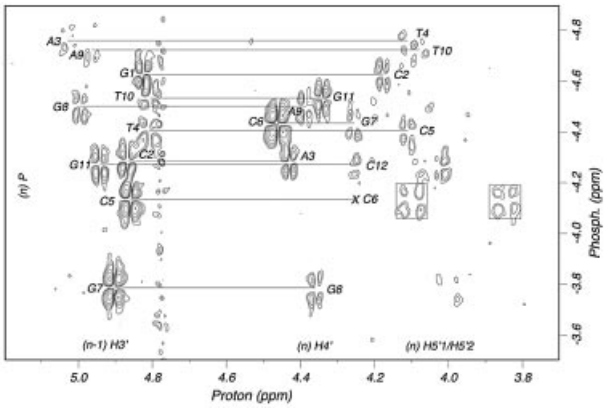



**Figure 4.** The $^{1}H–^{31}P$ HETCOR spectrum. The three bond $(n)P – (n – 1)H3'$ crosspeaks ($^{3}J_{P–H3'}$) are labeled with the names and numbers of the $(n – 1)H3'$ residue, and connected by horizontal lines to the four-bond long range $(n)P – (n)H4'$ crosspeaks ($^{4}J_{P–H4'}$). The absence of the C6 four bond intraresidue P–H4′ ($^{4}J_{P–H4'}$) crosspeak is marked *X*. The three bond intraresidue C6 P–H5′1 ($^{3}J_{P–H5'1}$) and C6 P-H5′2 ($^{3}J_{P–H5'2}$) crosspeaks are indicated with boxes.

## Structure determination

Based on the absence of an imino proton resonance for the cross-linked guanines, we initially assumed that both the G7 and G7′ N1-imino protons were present, and that G7 and G7′ were not hydrogen bonded either to their complementary cytosine partners (C6′ and C6) or to each other. The fact that only one set of resonances is observed for all residues indicates that the cross-linked guanines are symmetrical. There are hypothetically two ways to obtain $C_2$-symmetry with G7–G7′ planar: the guanines could either adopt a 'head-to-head' conformation or a 'side-by-side' conformation (Fig. 5A). There is no evidence to suggest that there is exchange between these two conformations or between other potential non-symmetrical conformations. Such interconversions should be readily detectable as they require at least one of the cross-linked guanines to make a ring flip, which would be slow on the NMR time scale. The normal G7–G8 base–base stacking, and the normal conformations adopted by the phosphate backbone in the C6–G7 and G7–G8 steps, strongly suggest that the cross-linked guanines adopt the head-to-head conformation. The side-to-side conformation appears to be sterically prohibited, as it results in an extremely narrow minor groove and a clash between the phosphate backbones of the two strands. No assumptions were made to exclude this conformation, but attempts to use it as a starting point in refinement invariably resulted in collapse of the structures. Following the procedure described in Materials and Methods, the iterative relaxation matrix-based back-calculation process consistently yielded only structures with the head-to-head conformation, confirming that the side-by-side conformation is indeed inconsistent with the experimental data. During the refinement process it also became apparent that the experimental restraints favor a planar orientation for the head-to-head conformation. The two cross-linked guanine N1-imino protons were consistently forced closer (~1.6 Å) than their combined van der Waals radii (2.2 Å). In order to prevent such a steric clash, the two guanines would have to be significantly (>50°) propeller twisted, which is not observed. Interestingly, when left unrestrained the cross-linked guanines do tend to propeller twist, suggesting that the planar conformation is not a modeling artifact but a real consequence of the experimental restraints. Thus, based on this steric constraint, the observed planar head-to-head conformation is only consistent with a single imino proton located between the two guanines. This arrangement would require one of the guanines to adopt the unusual enol tautomer, or more likely, that each guanine alternates between the keto and enol forms (Fig. 5B). However, this would suggest that there is also alternating G7:G7′ H1–N1 and G7:G7′ N1–H1 hydrogen bonding, which is an apparent contradiction to the absence of the cross-linked guanine imino proton resonance. We propose that there is a single hydrogen bonded imino proton which is exchanging unusually quickly. Interestingly, we have observed a downfield shifted imino resonance at 15.6 p.p.m. in a related cross-linked duplex (these data are presented in reference 29), which supports this hydrogen bonding scheme. Furthermore, if the enol form is present, a modest (~0.2 p.p.m.) downfield shift of the H8 would be expected (30) in the NOESY spectrum. The G7 H8 is the furthest downfield (8.0 p.p.m.) of the four guanines in the duplex. Although other factors such as base–base stacking greatly influence aromatic proton chemical shifts, this observation is consistent with the

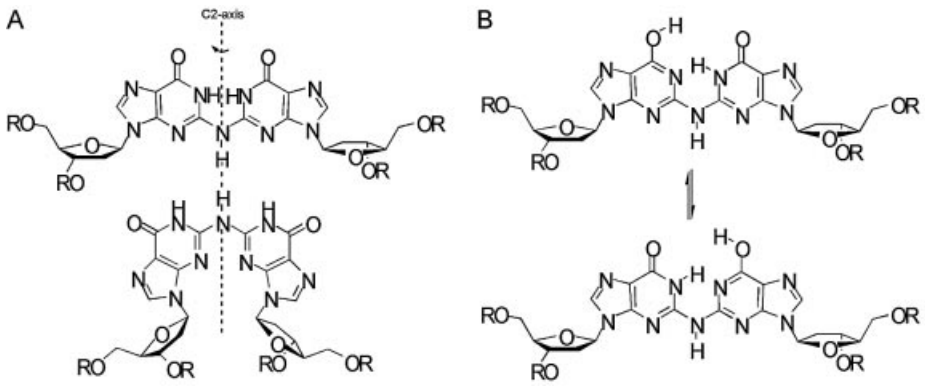



**Figure 5.** (**A**) The two possible symmetrical arrangements of the cross-linked guanines: the 'head-to-head' conformation (top) and the 'side-by-side' conformation (bottom). For each, the $C_2$-symmetry axis is indicated. (**B**) The proposed interconversion between the enol and keto forms for each of the cross-linked guanines. The two guanines are planar, with a single shared imino proton and an H1–N1 hydrogen bond.

presence of the enol form for the cross-linked guanines. We will henceforth refer to the cross-linked guanines as the G7:G7′ base pair.

In subsequent iterations of the refinement, we used a single imino proton located either on G7 or on G7′ and with either G7 or G7′ in the enol form, i.e. the scheme in Figure 5B, as described in Materials and Methods. After numerous iterations, a final set of distance restraints were obtained and used for the final distance geometry/restrained molecular dynamics run. Shown in Supplementary Figure S5 are the 26 independently generated final refined structures, which converged with an average pair-wise r.m.s. deviation of 0.13 ± 0.08 Å, an average NOE *R*-factor of 0.24 ± 0.00 and low restraint violation energies. A structural and energetic analysis is shown in Supplementary Table 2. There is excellent convergence between the 26 structures, reflecting the fact that the structure is very well defined. This can be attributed to the well resolved NMR spectra, and an abundance of distance and dihedral restraints. However, this does not reflect the explicit dynamics of the structure, since no such measurements have been made.

### Structural features

The skeletal stereo views of the lowest energy structure of the final set of refined structures are shown in Figure 6. Overall the structure is B-form DNA, except in the cross-link region (residues C5–G8). The base of the C6 and C6′ residues are flipped out of the helix and reside in the minor groove, with their hydrophobic edges turned toward the core of the helix (Fig. 6A). The cross-linked guanines form a nearly planar covalently linked G7:G7′ base pair with only minor propeller twisting (Fig. 6B). The G7:G7′ base pair is severely tilted with respect to the main helix axis, but is stacked well on the spatially adjacent C5:G8′ and G8:C5′ base pairs. The ribbon trace emphasizes the structural distortion induced by the extrahelical position of C6 and C6′ bases in the minor groove (Fig. 6A), and the unusually narrow major groove at the cross-link (Fig. 6B). Shown in Supplementary Figure S6 are the plots of the minor and major groove widths for the duplex. The minor groove widens to accommodate the extrahelical cytidine with the maximum P–P separation observed for the C6–G11′ and G11–C6′ steps (7.9 Å), and subsequently narrows at the cross-link with the minimum P–P separation for the G8–A9′ and A9–G8′ steps (4.2 Å). The major groove narrows at the cross-link with a minimum P–P separation observed for the C5–C5′ step (7.3 Å).

The rotational and translational parameters for the duplex are listed in Supplementary Table 3. The average helical twist angle for the normal Watson–Crick base pairs (G1:C12′–C5:G8′ and G8:C5′–C12:G1′) is 40° with an average rise of 3.1 Å, confirming that the duplex is overall in the B-DNA form. There are 10 residues per turn, as in B-DNA, and hence no under- or overwinding. However, as seen in Figure 6C, a base pair has in effect been removed from the helix, resulting in a reduced pitch height (28 Å). The structure is not bent as evidenced by an overall helical tilt angle of 0°, although it should be mentioned that the methods used here do not necessarily predict this parameter accurately. However, this is consistent with our preliminary electrophoretic measurements (our unpublished results), based on standard methods (31), which indicate that the duplex does not exhibit any large curvature. Detailed views of the cross-link region are illustrated in Figure 7. As seen in the top view (Fig. 7A) the intrastrand base–base stacking is particularly good between

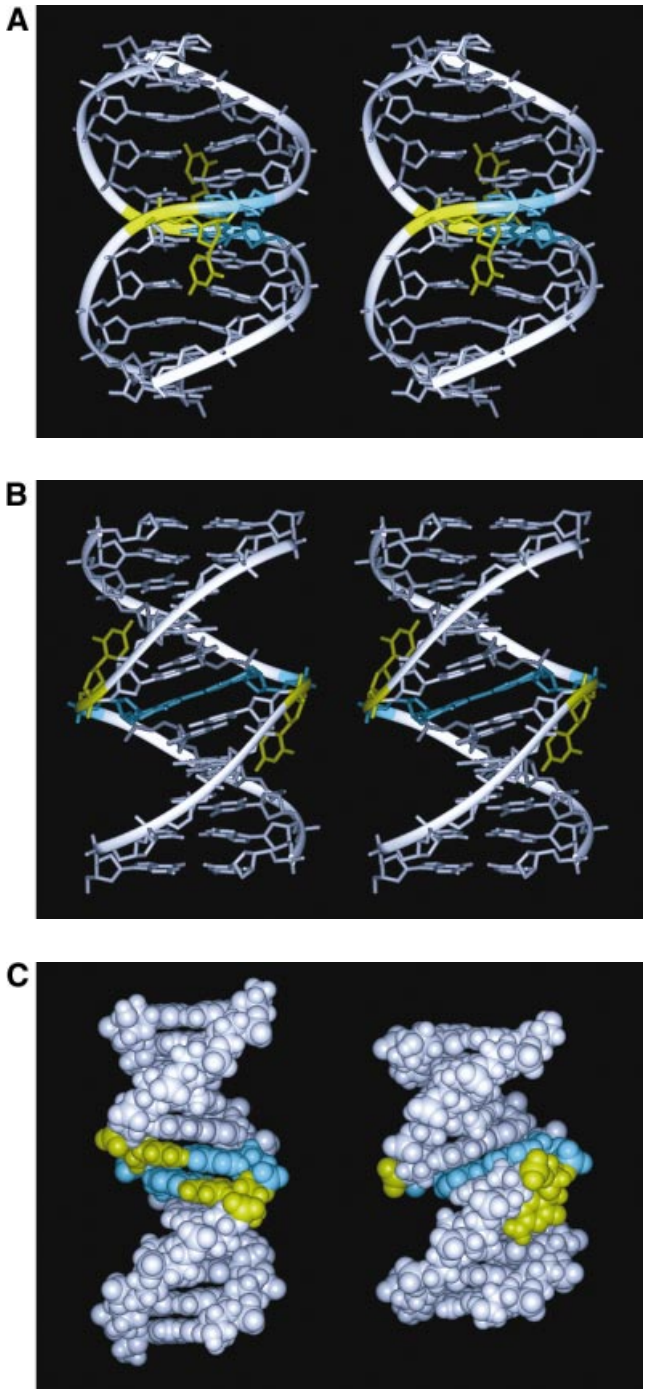



**Figure 6.** The skeletal stereo views of the lowest energy structure of the final set of refined structures. The cross-linked guanines (G7 and G7′) are colored blue and the extrahelical cytosine residues (C6 and C6′) are in yellow. The ribbon backbone trace emphasizes the unusual minor groove widths. (**A**) Looking into the minor groove at C6. The 5′-ends of the two strands are top-foreground and bottom-background, respectively. (**B**) Looking into the major groove at the cross-link. The 5′-ends of the two strands are top-right and bottom-left, respectively. (**C**) Space filling view of the of the cross-linked duplex (right) compared to idealized B-DNA (left), with G7 and G7′ in blue, C6 and C6′ in yellow.

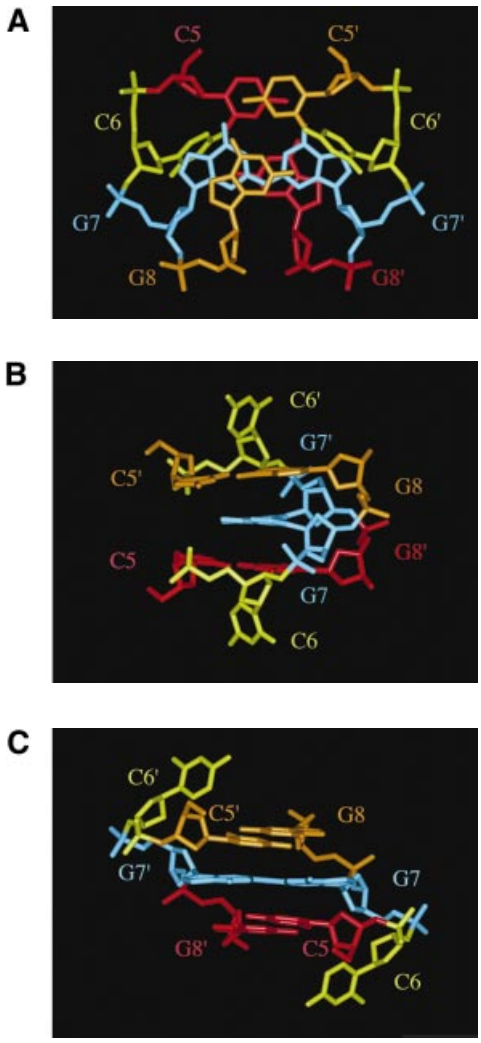



**Figure 7.** Detailed views of the cross-link region, showing the C5:G8′ base pair in red, C6 and C6′ in yellow, G7:G7′ in blue and G8:C5′ in orange. (**A**) Looking down the helix axis. (**B**) Looking into the minor groove at the cross-link. (**C**) Looking into the major groove at the cross-link.

G7 and G8 (and between G7′ and G8′), while the C5–G7 (and C5′–G7′) base–base stacking is quite poor. There is a large helical twist (54°) between the adjacent C5:G8′ and G7:G7′ base pairs (and between G7:G7′ and G8:C5′), due to the backbone segment of the intervening C6 residue. The two cross-linked guanines form a nearly planar base pair with only a minor propeller twist of –16° (Fig. 7B). The rise between G7:G7′ and the adjacent base pairs is 3.1 Å, similar to that of adjacent base pairs in normal B-DNA (Fig. 7C). Shown in Supplementary Figure S7 is a detailed depiction of the cross-link region and the structural basis for a number of the observed NOEs.

Table 1 lists the glycosidic torsion angles ($\chi$) and sugar pseudorotation angles (P). All sugars, including those of C6 and G7, adopt C2′-*endo* type conformations, with pseudorotation angles in the 114–172° range (136 and 138° for C6 and G7, respectively). The $\chi$ glycosidic torsion angles are in the *anti* range (–85 to –126°) for all residues. In spite of the extrahelical position of its base, C6 adopts a normal $\chi$ value of –118°. Also shown in Table 1 are the phosphate backbone torsion angles ($\alpha$–$\zeta$). With the exception of the C5–C6 step, but including the C6–G7 step, all interresidue backbone linkages adopt the typical $\varepsilon(t)$ $\zeta(g^-)$ $\alpha(g^-)$ $\beta(t)$ $\gamma(g^+)$ conformation. The C5–C6 step adopts the highly unusual $\varepsilon(g^-)$ $\zeta(g^+)$ $\alpha(g^+)$ $\beta(t)$ $\gamma(t)$ conformation, with C5 $\varepsilon$, C5 $\zeta$, C6 $\alpha$ and C6 $\gamma$ torsion angles of –76°, 117°, 80° and 167°, respectively. These unusual dihedral angles, along with the C6 $\beta$ torsion angle of –170°, are all consistent with the observed NOESY, DQF–COSY and $^1$H–$^{31}$P HETCOR data described above. The $\varepsilon(g^-)$ $\zeta(g^+)$ $\alpha(g^+)$ $\beta(t)$ $\gamma(t)$ conformation, highlighted in Supplementary Figure S8, causes a local strand reversal, which has the effect of flipping the C6 base out of the helix. Surprisingly, more modest deviations (35–70°) from idealized B-DNA dihedral angles for the C6–G7 linkage, all within the normal $\varepsilon(t)$ $\zeta(g^-)$ $\alpha(g^-)$ $\beta(t)$ $\gamma(g^+)$ conformation, result in normal strand continuation.

## DISCUSSION

In this study we have established that a nitrous acid induced guanine–guanine interstrand cross-link, in the $[d(CG)]_2$ sequence context, significantly alters DNA structure. Our results are in contrast to a previous model study which suggested that the cross-link could be accommodated with overall relatively minor deviations from B-DNA structure (10). In that model, the cross-linked guanines were severely tilted with respect to the adjacent bases, but still hydrogen bonded to the partner cytosines (although with less than ideal geometry). We have shown here, that the cross-linked guanines form a nearly planar covalently linked 'G:G base pair' with only minor propeller twisting, while the cytidine bases of their normal base pairing partners have been flipped out of the helix and adopt well defined extrahelical positions in the minor groove. This allows the cross-linked guanines to stack well on the adjacent base pairs, particularly on the guanines 3′ to the cross-link. The minor groove is widened 5′ to the cross-link to accommodate the extrahelical bases, whereas the major groove is quite narrow at the cross-link.

Our results indicate that the planar G:G base pair formed by the cross-linked guanines is stabilized by H1–N1 hydrogen bonding and favorable base–base stacking on the adjacent guanines. From an energetic stand point it appears that less restricted extrahelical cytidines are more favorable than poorly stacked, weakly hydrogen bonded cytidine bases in the helix. As the planar G:G base pair is sterically inconsistent with the two guanines in the typical keto form, we suggest that one of the guanines adopts the unusual enol tautomer. Since the two guanines are equivalent on the NMR time scale, each guanine must rapidly alternate between the two tautomers.

The fact that the imino proton signal of the cross-linked guanines remains undetectable under a wide range of conditions is intriguing, and we suggest that the reason is very efficient intrinsic exchange catalysis (32) by the spatially adjacent guanine O6. This exchange appears to be facilitated by the close proximity of H1 and O6, and by the fixed geometry of the cross-link. The less than ideal N1–H1–N1 hydrogen bond angle (~125°) could also be a contributing factor. The position of the cytidine bases in the minor groove appears to be relatively fixed, as evidenced by the numerous NOEs to the adjacent residues, the absence of line broadening,

**Table 1.** Backbone torsion angles (α–ζ), glycosidic torsion angles (χ) and sugar pseudorotation angles (P) for the cross-linked duplex

| Residue | α | β | γ | δ | ε | ζ | χ | P |
|---|---|---|---|---|---|---|---|---|
| G1 | – | – | 179 ± 73 | 144 ± 3 | –179 ± 0 | –103 ± 0 | –122 ± 1 | 154 ± 3 |
| C2 | –71 ± 0 | –176 ± 0 | 58 ± 0 | 134 ± 0 | –178 ± 0 | –92 ± 0 | –121 ± 0 | 145 ± 0 |
| A3 | –75 ± 0 | 180 ± 0 | 56 ± 0 | 121 ± 0 | 180 ± 0 | –94 ± 0 | –121 ± 0 | 130 ± 0 |
| T4 | –73 ± 0 | 179 ± 0 | 58 ± 0 | 111 ± 1 | –175 ± 1 | –96 ± 1 | –125 ± 1 | 114 ± 1 |
| C5 | –73 ± 2 | –176 ± 1 | 56 ± 1 | 146 ± 0 | **–76 ± 1** | **117 ± 20** | –85 ± 1 | 172 ± 1 |
| C6 | **80 ± 56** | –170 ± 20 | **167 ± 3** | 132 ± 0 | **–137 ± 1** | **–55 ± 0** | –118 ± 2 | 136 ± 2 |
| G7 | –86 ± 0 | 173 ± 0 | 74 ± 0 | 127 ± 1 | –172 ± 0 | –98 ± 1 | –110 ± 1 | 138 ± 1 |
| G8 | –89 ± 1 | –178 ± 0 | 58 ± 0 | 130 ± 0 | –170 ± 0 | –127 ± 1 | –100 ± 1 | 139 ± 1 |
| A9 | –67 ± 0 | 172 ± 0 | 61 ± 0 | 141 ± 0 | –177 ± 0 | –98 ± 1 | –98 ± 1 | 167 ± 0 |
| T10 | –71 ± 0 | 169 ± 0 | 62 ± 1 | 115 ± 1 | 180 ± 0 | –96 ± 0 | –121 ± 1 | 120 ± 1 |
| G11 | –70 ± 0 | 178 ± 0 | 56 ± 0 | 135 ± 0 | 176 ± 0 | –103 ± 0 | –110 ± 0 | 145 ± 0 |
| C12 | –67 ± 0 | 178 ± 0 | 60 ± 0 | 129 ± 0 | – | – | –126 ± 0 | 135 ± 1 |

The angles are given in degrees (°) and averaged over the final 26 refined structures. Unusual torsion angles are in bold.

and the fact that the apparent correlation time for these bases is roughly equal to that of the duplex (~4 ns). This could be explained by the fact that the hydrophobic edge of the cytidines is turned towards the core of the helix and the hydrophilic edge is in contact with the solvent. In addition, the amino proton ($N_4H$) of the extrahelical cytidine is within hydrogen bonding distance (2.0 Å) of the O6 of the 5′ ($n$ – 2) thymine. Although the cytosine amino protons are exchanging too rapidly with solvent to be observed, indicating that this hydrogen bond is at best quite weak, it could stabilize the structure and help explain the fixed location of the extrahelical cytidines.

The unusual ε($g^-$) ζ($g^+$) α($g^+$) β($t$) γ($t$) phosphate backbone conformation has, to our knowledge, not been observed previously in DNA, which is perhaps not surprising given the unusual orientation of the extrahelical cytidine bases in the minor groove. While the ε($g^-$) conformation is generally considered sterically forbidden, it is allowed in this case since the unusual ζ($g^+$) conformation redirects the backbone preventing a steric clash between the C5 H4′ and one of the C6 phosphate oxygens. The ζ($g^+$) α($g^+$) phosphodiester conformation, however, has been observed when the direction of the helix axis is disrupted in several DNA hairpin loops (16,18,28), in the cisplatin–DNA interstrand cross-link (33,34), and in single-stranded DNA complexed with a telomere end binding protein (35). The γ($t$) conformation also appears to be a common occurrence when extrahelical residues are present (33,36,37).

The role of nitrous acid induced DNA interstrand cross-links in mutagenesis, and the mechanisms responsible for its repair in intracellular environments, are poorly understood (38). It is believed that the recombination repair system is responsible for the resolution of nitrous acid induced cross-links in bacteria, but the details are undetermined (39). Equipped with the structure presented here and large quantities of homogeneous samples, these issues can now be addressed. Interstrand cross-links are generally believed to be toxic, if not lethal to cells, and if left unrepaired they can prevent strand separation during replication and transcription. Based on this structure several other modes of action seem plausible. Both the major and minor groove widths are altered and the extrahelical cytidines form a 'road block' in the minor groove, factors that all could disrupt protein binding. Furthermore, the cross-link could perhaps induce binding of proteins that recognize extrahelical bases. In fact, the cross-link structure is remarkably similar to that of the DNA found in the crystal structures of a number of modification enzymes such as bacterial methyl transferases (36,40), bacterial methylase (41), bacterial endonuclease IV (42), bacterial G:T/U-specific DNA glycosylase (43), and human uracil–DNA glycosylase (44), which all contain extrahelical bases. The cross-link structure could bind these types of enzymes with relatively high affinity, since only minor conformational rearrangements would be required for the extrahelical cytidines to fit into the binding pocket of the enzyme. Although it is unclear how any specific repair enzyme would recognize this type of structure, it is tempting to speculate that the cellular repair machinery would attempt to excise these cytosines. This could lead to a deletion mutation even if the cross-link itself is repaired, which is a leading cause of nitrous acid mutagenicity (40). If the cytosines are excised, it is likely that the cross-link still would stack well on the adjacent base pairs and be structurally similar to B-DNA. As a result, the repair machinery might not be as likely to repair the cross-link. It would be possible to study such a system, since our methodology would allow the synthesis of a cross-linked duplex lacking the extrahelical residues. It remains to be seen if any of these admittedly somewhat speculative ideas turn out to be true.

## SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

## ACKNOWLEDGEMENTS

We thank Dr José Gallego for his advice and many helpful discussions, and Mrs Elisabeth B. Golden for excellent technical assistance. Dr Leiming Zhu provided the back-calculation program BIRDER and analysis program FELIXBAD. This work was supported in part by NIH grants GM-32681 to B. R. R. and GM-45804 to P. B. H. The coordinates have been deposited in the RSCB Protein Data Bank (Piscataway, NJ) with the accession number 1S9N.

## REFERENCES

1. Schuster,H. and Schramm,G. (1958) Effects of chemical changes on the activity of the ribonucleic acid of the tobacco mosaic virus. *Naturforsch.*, **13B**, 697–704.

2. Shapiro,R. and Pohl,S.H. (1968) The reaction of ribonucleosides with nitrous acid. Side products and kinetics. *Biochemistry*, **7**, 448–455.
3. Geiduschek,E.P. (1961) 'Reversible' DNA. *Proc. Natl Acad. Sci. USA*, **47**, 950–955.
4. Becker,E.F.,Jr (1967) Structure and function of cross-linked DNA. II. Loss of reversibility of $HNO_2$-cross-linked DNA during transformation of *Bacillus subtilis. Biochim. Biophys. Acta*, **142**, 238–244.
5. Furia,T.E. (ed.) (1972) Antimicrobial food additives. *Handbook of Food Additives*. 2nd edn. CRC Press, Cleveland, OH, pp. 150–155.
6. Becker,E.F.,Jr, Zimmerman,F.K. and Geiduschek,E.P. (1964) Structure and function of cross-linked DNA.I. Reversible denaturation and *Bacillus subtilis* transformation. *J. Mol. Biol.*, **8**, 377–391.
7. Kirchner,J.J. and Hopkins,P.B. (1991) Nitrous acid cross-links duplex DNA fragments through deoxyguanosine residues at the sequence 5′-CG. *J. Am. Chem. Soc.*, **113**, 4681–4682.
8. Kirchner,J.J., Sigurdsson,S.T. and Hopkins,P.B. (1992) Interstrand cross-linking of duplex DNA by nitrous acid: covalent structure of the dG-to-dG cross-link at the sequence 5′-CG. *J. Am. Chem. Soc.*, **114**, 4021–4027.
9. Shapiro,R., Dubelman,S., Feinberg,A.M., Crain,P.F. and McCloskey,J.A. (1977) Isolation and identification of cross-linked nucleosides from nitrous acid treated deoxribonucleic acid. *J. Am. Chem. Soc.*, **99**, 302–303.
10. Kirchner,J.J., Solomon,M.S. and Hopkins,P.B. (1992) Interstrand cross-linking of duplex DNA by nitrous acid: determination of the sequence preference at nucleotide resolution. In Sarma,R.H. and Sarma,M.H. (eds), *Structure and Function*, *Volume 1: Nucleic Acids*. Adenine Press, Albany, pp. 171–182.
11. Elcock,A.H., Lyne,P.D., Mulholland,A.J., Nandra,A. and Richards,W.G. (1995) Combined quantum and molecular mechanical study of DNA cross-linking by nitrous acid. *J. Am. Chem. Soc.*, **117**, 4706–4707.
12. Harwood,E.A., Sigurdsson,S.T., Edfeldt,N.B.F., Reid,B.R. and Hopkins,P.B. (1999) Chemical synthesis and preliminary structural characterization of a nitrous acid interstrand cross-linked duplex DNA. *J. Am. Chem. Soc.*, **121**, 5081–5082.
13. Piotto,M., Saudek,V. and Sklénar,V. (1992) Gradient-tailored excitation for single quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR*, **2**, 661–665.
14. Sklénar,V., Miyashiro,H., Zon,G., Miles,H.T. and Bax,A. (1986) Assignment of $^{31}$P and $^{1}$H resonances in oligonucleotides by two-dimensional NMR spectroscopy. *FEBS Lett.*, **208**, 94–98.
15. Cornell,W.D., Cieplak,P., Bayly,C.I., Gould,I.R., Merz,K.M.,Jr, Ferguson,D.M., Spellmeyer,D.C., Fox,T., Caldwell,J.W. and Kollman,P.A. (1995) A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.
16. Zhu,L., Chou,S.-H. and Reid,B.R. (1995) The structure of a novel DNA duplex formed by human centromere d(TGGAA) repeats with possible implications for chromosome attachment during mitosis. *J. Mol. Biol.*, **254**, 623–637.
17. Chou,S.-H., Zhu,L., Gao,Z., Cheng,J.-W. and Reid,B.R. (1996) Hairpin loops consisting of single adenine residues closed by sheared A·A and G·G pairs formed by the DNA triplets AAA and GAG: solution structure of the d(GTAC*AAA*GTAC) hairpin. *J. Mol. Biol.*, **264**, 981–1001.
18. Fedoroff,O.Y., Salazar,M. and Reid,B.R. (1996) Structural variation among retroviral primer-DNA junctions: solution structure of the HIV-1 (–)-strand Okazaki fragment r(gcca)d(CTGC)·d(GCAGTGGC). *Biochemistry*, **35**, 11070–11080.
19. Gallego,J., Chou,S.-H. and Reid,B.R. (1997) Centromeric pyrimidine strands fold into an intercalated motif by forming a double hairpin with a novel T:G:G:T tetrad: solution structure of the d(TCCCGTTTCCA) dimer. *J. Mol. Biol.*, **273**, 840–856.
20. Zhu,L. and Reid,B.R. (1995) An improved NOESY simulation program for partially relaxed spectra: BIRDER. *J. Magn. Reson. Ser. B*, **106**, 227–235.
21. Lane,A.N. and Peck,B. (1995) Conformational flexibility in DNA duplexes containing single G·G mismatches. *Eur. J. Biochem.*, **230**, 1073–1087.
22. Faibis,V., Cognet,J.A.H., Boulard,Y., Sowers,L.C. and Fazakerly,G.V. (1996) Solution structure of two mismatches G·G and I·I in the K-*ras* gene context by nuclear magnetic resonance and molecular dynamics. *Biochemistry*, **35**, 14452–14464.
23. Woodson,S.A. and Crothers,D.M. (1988) Preferential location of bulged guanosine internal to a G·C tract by $^{1}$H NMR. *Biochemistry*, **27**, 436–445.
24. Sarma,R.H., Mynott,R.J., Wood,D.J. and Hruska,F.E. (1973) Determination of the preferred conformations constrained along the C-4′-C-5′ and C-5′-O-5′ bonds of β-5′-nucleotides in solution. *J. Am. Chem. Soc.*, **95**, 6457–6459.
25. Altona,C. (1982) Conformational analysis of nucleic acids. Determination of backbone geometry of single-helical RNA and DNA in aqueous solution. *Recl. Trav. Chim. Pays-Bas*, **101**, 413–433.
26. Kim,S.-G., Lin,L.-J. and Reid,B.R. (1992) Determination of nucleic acid backbone conformation by $^{1}$H NMR. *Biochemistry*, **31**, 3564–3574.
27. Gorenstein,D.G., Schroeder,S.A., Fu,J.M., Metz,J.T., Roongta,V. and Jones,C.L. (1988) Assignment of $^{31}$P resonances in oligodeoxyribonucleotides: origin of sequence-specific variations in the deoxyribose phosphate backbone conformation and the $^{31}$P chemical shifts of double-helical nucleic acids. *Biochemistry*, **27**, 7223–7237.
28. Blommers,M.J.J., van de Ven,F.J.M., van der Marel,G.A., van Boom,J.H. and Hilbers,C.W. (1991) The three-dimensional structure of a DNA hairpin in solution: two-dimensional NMR studies and structural analysis of d(ATCCTATTTATAGGAT). *Eur. J. Biochem.*, **201**, 33–51.
29. Edfeldt,N.B.F., Harwood,E.A., Sigurdsson,S.Th., Hopkins,P.B. and Reid,B.R. (2004) Sequence context effect on the structure of nitrous acid induced DNA interstrand cross-links. *Nucleic Acids Res.*, **32**, 2795–2801.
30. Evans,F.E. and Sarma,R.H. (1974) The tautomeric form of inosine in aqueous solution. *J. Mol. Biol.*, **89**, 249–253.
31. Koo,H.-S. and Crothers,D.M. (1988) Calibration of DNA curvature and a unified description of sequence-directed bending. *Proc. Natl Acad. Sci. USA*, **85**, 1763–1767.
32. Guéron,M. and Leroy,J.-L. (1992) Base-pair opening in double-stranded nucleic acids. In Eckstein,F. and Lilley,D.M.J. (eds), *Nucleic Acids and Molecular Biology*, vol. 6, Springer-Verlag, Heidelberg, pp. 1–22.
33. Huang,H., Zhu,L., Reid,B.R., Drobny,G.P. and Hopkins,P.B. (1995) Solution structure of a cisplatin-induced DNA interstrand cross-link. *Science*, **270**, 1842–1845.
34. Coste,F., Malinge,J.-M., Serre,L., Shepard,W., Roth,M., Leng,M. and Zelwer,C. (1999) Crystal structure of a double-stranded DNA containing a cisplatin interstrand cross-link at 1.63 Å resolution: hydration at the platinated site. *Nucleic Acids Res.*, **27**, 1837–1846.
35. Horvath,M.P., Schweiker,V.L., Bevilacqua,J.M., Ruggles,J.A. and Schultz,S.C. (1998) Crystal structure of the *Oxytricha nova* telomere end binding protein complexed with single stranded DNA. *Cell*, **95**, 963–974.
36. O'Gara,M., Horton,J.R., Roberts,R.J. and Cheng,X. (1998) Structures of HhaI methyltransferase complexed with substrates containing mismatches at the target base. *Nature Struct. Biol.*, **5**, 872–877.
37. Mol,C.D., Izumi,T., Mitra,S. and Tainer,J.A. (2000) DNA-bound structures and mutants reveal abasic DNA binding by APE1 DNA repair and coordination. *Nature*, **403**, 451–456.
38. Sidorkina,O., Saparbaev,M. and Laval,J. (1997) Effects of nitrous acid treatment on the survival and mutagenesis of *Escherichia coli* cells lacking base excision repair (hypoxanthine-DNA glycosylase-ALK A protein) and/or nucleotide excision repair. *Mutagenesis*, **12**, 23–27.
39. Hartman,Z., Henrikson,E.N., Hartman,P.E. and Cebula,T.A. (1994) Molecular models that may account for nitrous acid mutagenesis in organisms containing double-stranded DNA. *Environ. Mol. Mutagen.*, **24**, 168–175.
40. Klimasauskas,S., Kumar,S., Roberts,R.J. and Cheng,X. (1994) HhaI methyltransferase flips its target base out of the DNA helix. *Cell*, **76**, 357–369.
41. Reinisch,K.M., Chen,L., Verdine,G.L. and Lipscomb,W.N. (1995) The crystal structure of HaeIII methyltransferase covalently complexed to DNA: an extrahelical cytosine and rearranged base pairing. *Cell*, **82**, 143–153.
42. Hosfield,D.J., Guan,Y., Haas,B.J., Cunningham,R.P. and Tainer,J.A. (1999) Structure of the DNA repair enzyme endonuclease IV and its DNA complex: double-nucleotide flipping at abasic sites and three-metal-ion catalysis. *Cell*, **98**, 397–408.
43. Barrett,T.E., Savva,R., Barlow,T., Brown,T., Jiricny,J. and Pearl,L.H. (1998) Structure of a DNA base-excision product resembling a cisplatin inter-strand adduct. *Nature Struct. Biol.*, **5**, 697–701.
44. Slupphaug,G., Mol,C.D., Kavli,B., Arvai,A.S., Krokan,H.E. and Trainer,J.A. (1996) A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature*, **384**, 25–26.
45. Birchall,A.J. and Lane,A.N. (1990) Anisotropic rotation in nucleic acid fragments: significance for determination of structures from NMR data. *Eur. Biophys. J.*, **19**, 73–78.