Published online May 20, 2004

Nucleic Acids Research, 2004, Vol. 32, No. 9 ***2795–2801***
DOI: 10.1093/nar/gkh607

Sequence context effect on the structure of nitrous acid induced DNA interstrand cross-links

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

In the preceding paper in this journal, we described the solution structure of the nitrous acid cross-linked dodecamer duplex [d(GCATCCGGATGC)]$_2$ (the cross-linked guanines are underlined). The structure revealed that the cross-linked guanines form a nearly planar covalently linked 'G:G base pair', with the complementary partner cytidines flipped out of the helix. Here we explore the flanking sequence context effect on the structure of nitrous acid cross-links in [d(CG)]$_2$ and the factors allowing the extrahelical cytidines to adopt such fixed positions in the minor groove. We have used NMR spectroscopy to determine the solution structure of a second cross-linked dodecamer duplex, [d(CGCTACGTAGCG)]$_2$, which shows that the identity of the flanking base pairs significantly alters the stacking patterns and phosphate backbone conformations. The cross-linked guanines are now stacked well on adenines preceding the extrahelical cytidines, illustrating the importance of purine–purine base stacking. Observation of an imino proton resonance at 15.6 p.p.m. provides evidence for hydrogen bonding between the two cross-linked guanines. Preliminary structural studies on the cross-linked duplex [d(CGCGACGTCGCG)]$_2$ show that the extrahelical cytidines are very mobile in this sequence context. We suggest that favorable van der Waals interactions between the cytidine and the adenine 2 bp away from the cross-link localize the cytidines in the previous cross-linked structures.

INTRODUCTION

In the preceding paper in this journal (1), we described the solution structure of the cross-linked dodecamer duplex [d(GCATCCGGATGC)]$_2$ (the cross-linked guanines are underlined), henceforth referred to as *CCGG*. We found that the cross-linked guanines form a nearly planar, covalently linked 'G:G base pair', stacked on the 3′ side guanines of the spatially adjacent G:C base pairs. The observed planar geometry of the G:G base pair is consistent with a single H1–N1 hydrogen bond, although this could not be established experimentally as the cross-linked guanine imino protons were absent from the NMR spectra. The cytidines, which normally would base pair with the cross-linked guanines, were found to be flipped out of the helix, adopting well defined extrahelical positions in the minor groove. The phosphate backbone was found to be in the highly unusual $\varepsilon(g^-)$ $\zeta(g^+)$ $\alpha(g^+)$ $\beta(t)$ $\gamma(t)$ conformation on the 5′ side of the extrahelical cytidine, causing a local strand reversal and directing the base out of the helix. Somewhat surprisingly, more modest deviations from idealized B-DNA dihedral angles were observed for the 3′ side, all within the typical $\varepsilon(t)$ $\zeta(g^-)$ $\alpha(g^-)$ $\beta(t)$ $\gamma(g^+)$ conformational domains, resulting in normal strand continuation.

Extrahelical cytidines are well documented phenomena, and have previously been observed in cytosine bulges (2,3), C:C mismatches (4,5), a cisplatin-DNA interstrand cross-link (6,7), and in DNA bound to proteins such as bacterial cytosine-specific methyl transferase (8) and bacterial methylase (9). However, the particular location and lack of flexibility of the cytosines in the minor groove, and the unusual phosphate backbone conformation are unique to this system. This paper describes a more detailed investigation of nitrous acid cross-linked DNA. First, it was of interest to determine if the planar G:G base pair and the extrahelical cytidines are general features of cross-links in [d(CG)]$_2$, and to investigate the flanking sequence context effect on the structure. Secondly, we hoped to confirm experimentally the presence of a hydrogen bond between the two guanines within the cross-link. Thirdly, we wanted to learn more about the factors allowing the extrahelical cytidines to adopt such fixed positions in the minor groove. Therefore, we decided to determine the structure of the cross-linked dodecamer duplex [d(CGCTACGTAGCG)]$_2$ (the cross-linked guanines are underlined), henceforth referred to as *ACGT*, with the cross-link in a [d(ACGT)]$_2$ sequence and compare it to the structure of *CCGG*. The *CCGG* and *ACGT* sequences are the most susceptible to cross-linking *in vitro* (10). In addition, we performed preliminary structural characterization of the cross-linked duplex [d(CGCGACGTCGCG)]$_2$, which is analogous to *ACGT* but with flanking G:C base pairs instead of T:A base

*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

pairs (the T to G and A to C changes are underlined), henceforth referred to as *GACGTC*. The results from the study of *GACGTC* indicates that the second flanking base pairs affect the mobility of the extrahelical cytidines. Shown in Figure 1 are the schematic representations of all three duplexes used in these studies.

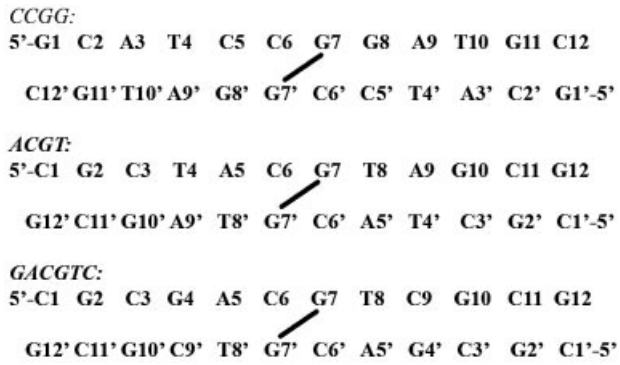

Figure 1. Schematic representation of the three cross-linked self-complementary dodecamer duplexes with their abbreviated designations and residue numbering schemes. Although the duplexes are symmetrical, the residues of one of the strands are designated with a prime (′) for convenience when describing interstrand interactions. Note, in particular, the cross-linked guanines (G7 and G7′), and the cytosines preceding the cross-link, C6 and C6′, which would normally base-pair with G7′ and G7, respectively.

MATERIALS AND METHODS

The cross-linked duplexes *ACGT* and *GACGTC* were synthesized as described previously (1,11). NMR samples were prepared, and NMR experiments were conducted, as in the previous paper in this journal (1). The solution structure of *ACGT* was determined using the distance geometry, restrained molecular dynamics and iterative NOE refinement protocol used for *CCGG* (1). Briefly, distance restraints for non-exchangeable protons were derived from ^2H_2O-NOESY spectra collected at 25°C with 60, 120, 180, 240 and 360 ms mixing times, qualitative exchangeable proton restraints, including hydrogen bonding restraints, were derived from the ^1H_2O-NOESY spectrum collected at 0°C, and phosphate backbone restraints were derived from the ^2H_2O-NOESY, DQF-COSY and ^1H-^{31}P HETCOR spectra (see Results). A total of 647 distance restraints were used, of which 215 were intraresidue, 432 interresidue, 52 hydrogen bonding, 42 derived from the ^1H_2O-NOESY spectrum, and 81 were repulsive. 104 phosphate backbone dihedral angle restraints and 72 chiral constraints were also used. The phosphate backbone was restrained to: non-*trans* ($0 \pm 150°$) for α and ζ; *trans* ($180 \pm 30°$) for β and ε (except C6 β, A5 ε and C6 ε, which were not restrained); and *gauche*$^+$ ($60 \pm 30°$) for γ [except C6 γ, which was restrained to *gauche*$^-$ ($-60 \pm 30°$)]. In each cycle the lowest energy structure was subjected to distance geometry/simulated annealing (DGII, Biosym/MSI), restrained molecular dynamics and energy minimization (DISCOVER, Biosym/MSI) resulting in a family of structures. The NOESY spectra of the resulting lowest energy structure were subsequently back-calculated using the NOESY simulation program BIRDER (12) with an empirically determined correlation time of 4.0 ns. The distance restraints were adjusted and the procedure repeated until the back-calculated spectra matched the experimental spectra, resulting in an R_{NOE} factor of 0.19 ± 0.00. The final set of 12 structures converged with pair wise RMS deviation values of 0.2 ± 0.1 Å, and low total and restraint violation energies. The cross-link was modeled with one guanine in the keto form with a single N1-imino proton and the other in the enol form, as was done previously for *CCGG* (1). As a test, in later stages of refinement, the cross-link was modeled with two N1-imino protons. However, these protons were consistently forced closer than their combined van der Waals radii (as we observed for *CCGG*). The presence of two G7/G7′ N1-imino protons is clearly inconsistent with the experimental data.

RESULTS

Exchangeable proton studies of *ACGT*

In the imino proton spectrum of *ACGT* shown in Figure 2A, there are five peaks corresponding to the hydrogen bonded residues G2, T4, T8, G10 and G12 (unambiguously assigned using the ^1H_2O-NOESY spectrum; data not shown). In addition, there is a broad peak at 15.6 p.p.m., which is only observed at very low temperature (0°C). This is most likely the G7 N1-imino proton resonance, although an unambiguous assignment could not be made in the ^1H_2O-NOESY spectrum. Rapid solvent exchange is the probable cause of such a broad peak, and it could explain why this resonance did not give any crosspeaks other than a large solvent exchange peak at the water resonance in the 0°C ^1H_2O-NOESY spectrum (data not shown). The chemical shift is highly anomalous for a guanine imino proton and ~2.5 p.p.m. further downfield than that of a typical hydrogen bonded guanine imino proton. However, it is clearly more consistent with a hydrogen bond than a lack thereof. Furthermore, this peak integrates to less than half of the others, which is expected for a single shared N1-imino proton. We suggest that this result is consistent with the G7 H1–G7′ N1 hydrogen bond we observed in the solution structure of *CCGG* (1). Such a hydrogen bond requires that one of the guanines adopt the unusual enol tautomer, or more likely that each guanine alternates between the keto and enol forms as shown in Figure 2B. For normal guanines in aqueous solutions, the keto tautomer dominates over the enol tautomer due to more favorable hydration energy in spite of nearly equal intrinsic stability (13). However, a less hydrophilic environment, such as the interior of a protein binding pocket, could shift the equilibrium to favor the enol form (14). In this case, it could be that since the two guanines are covalently linked they are less susceptible to the typical 'breathing' motions. This could make the microenvironment less hydrophilic which would indeed favor the presence of the enol form. Furthermore, since the guanine enol tautomer is fully aromatic, each guanine would be partially aromatic, and when bridged by the N2 lone pair, a large delocalized π electron cloud involving both purine rings would be created. We predict that this would result in a large deshielding ring current that could account for the large downfield shift of the G7 N1-imino proton. This type of keto-enol interconversion could also explain why, in spite of the apparent hydrogen bonding, the imino resonance is quite broad. If the imino proton is first transferred to the spatially adjacent carbonyl

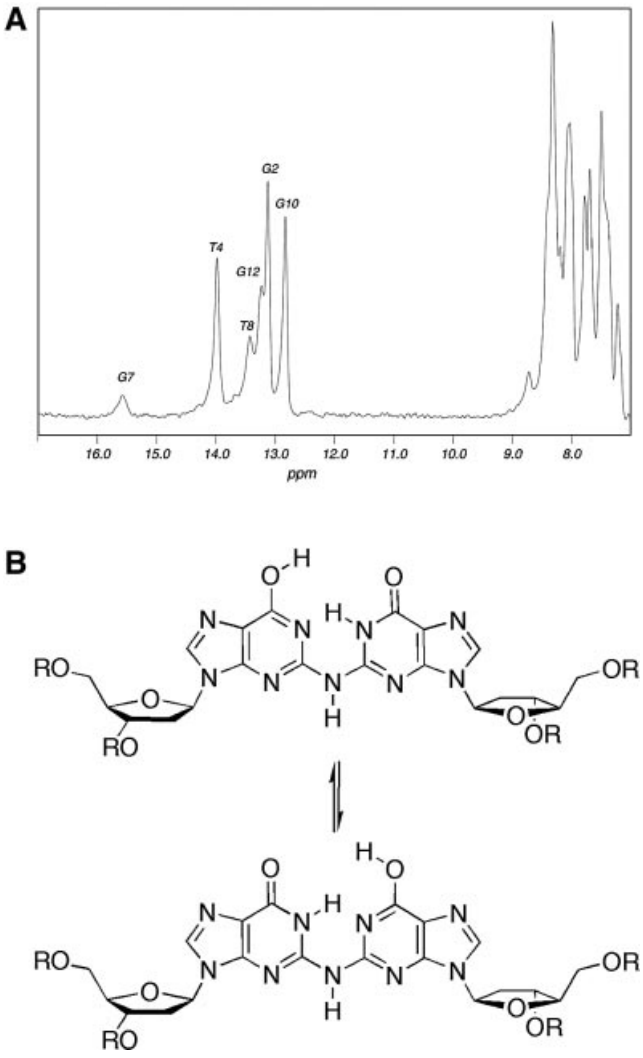

Figure 2. (**A**) The downfield region (7–17 p.p.m.) of the 1D proton 750 MHz NMR spectrum of *ACGT*, collected in 90% $^{1}H_2O$/10% $^{2}H_2O$ at 0°C. Note the broad G7 imino proton resonance. (**B**) The proposed interconversion between the enol and keto forms for each of the cross-linked guanines. The two guanines are planar, with a single shared imino proton and an H1–N1 hydrogen bond.

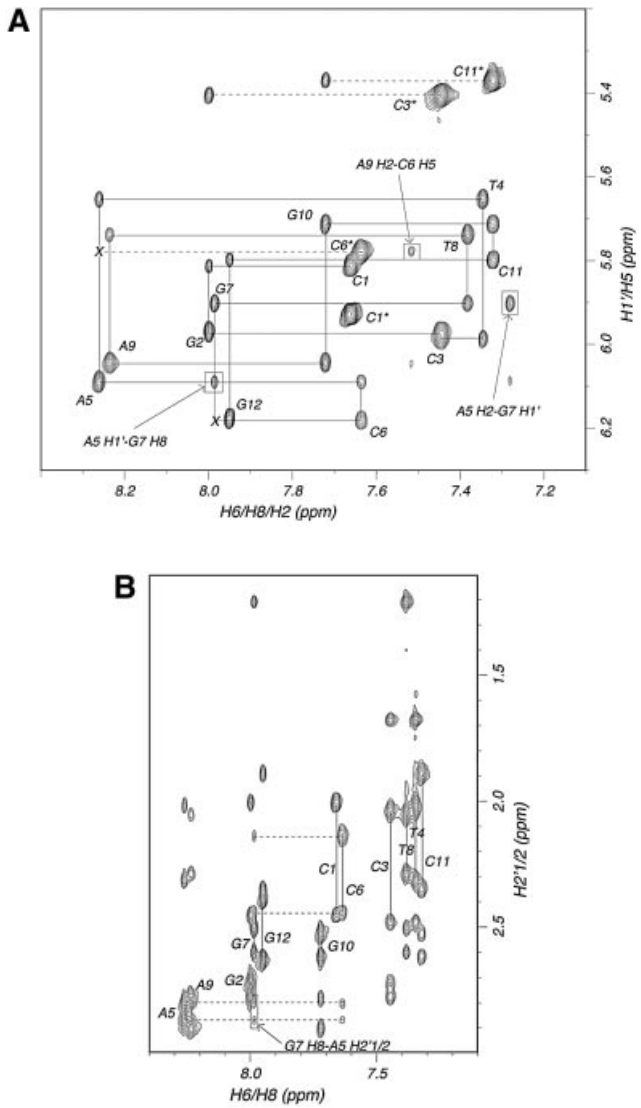

Figure 3. The 750 MHz $^{2}H_2O$-NOESY spectrum of *ACGT*, collected at 25°C with a mixing time of 360 ms. (**A**) The H6/H8–H1′/H5 region, showing the sequential aromatic to H1′ interresidue walk is indicated with lines and the intraresidue H6/H8–H1′ connectivities are labeled with the corresponding residue name and number. The absence of the C6 H1′–G7 H8 and A5 H8–C6 H5 connectivities are marked *X*, and the C6 H5–A9′ H2, A5 H1′–G7 H8 and A5 H2–G7 H1′ connectivities are labeled and indicated with arrows. (**B**) The H6/H8–H2′1/H2′2 region, showing the intraresidue aromatic to H2′1/H2′2 connectivities which are labeled and connected with lines. The sequential aromatic to H2′1/H2′2 interresidue walk for residues A5-G7 is indicated with dashed lines. The A5 H2′1–G7 H8 and A5 H2′2–G7 H8 connectivities are in a box.

oxygen (O6) as the guanine is converted to the enol form, it could subsequently be exchanged readily with the solvent from the hydroxyl position. The fixed geometry of the cross-link should facilitate this process, in which the O6 would act as an intrinsic exchange catalyst (15). While this explanation accounts for the appearance of the G7 imino proton resonance in *ACGT*, it does not explain why the G7 resonance was not observed in *CCGG*. We conclude that for some unknown reason the exchange process is simply more efficient in that duplex.

Non-exchangeable proton studies of *ACGT*

There are several features in H6/H8–H1′/H5 H6/H8–H2′1/H2′2 regions of the $^{2}H_2O$-NOESY spectrum of *ACGT*, shown in Figure 3, that indicate an overall structural similarity with *CCGG*. The intensities of the intraresidue aromatic to H1′ crosspeaks establish that all χ torsion angles 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 ($^{3}J_{H1'-H2'1}$) but weak H2′2–H3′ ($^{3}J_{H2'-2H3'}$) and H3′–H4′ ($^{3}J_{H3'-H4'}$) crosspeaks in the DQF-COSY spectrum, suggest that all residues adopt typical C2′-*endo* type sugar conformations (data not shown). The cross-linked guanines appear to form a head to head

G:G base pair that is well stacked in the helix. The normal G7 H1′–T8 H6, G7 H2′1–T8 H6, G7 H2′2–T8 H6 and G7 H8–T8 H6 connectivities (Fig. 3) indicate that the cross-linked G7 is stacked on the adjacent T8 base. As was the case in *CCGG*, the C6–G7 base–base stacking is disrupted, which is confirmed by the absence of C6 H1′–G7 H8 and C6 H6–G7 H8 connectivities, and the weak C6 H2′1–G7 H8 and C6 H2′2–G7 H8 connectivities. The A5–C6 base–base stacking is also disrupted, as evidenced by the weak A5 H2′1–C6 H6 and A5 H2′2–C6 H6 connectivities, and the lack of A5 H8–C6 H5, A5 H8–C6 H6, and A5 H3′–C6 H6 connectivities. The unusual interstrand C6 H5–A9′ H2 and C6 H6–A9′ H2 connectivities, previously seen in *CCGG*, confirm that the C6 base is flipped out of the helix and located in the minor groove. In spite of the numerous similarities with *CCGG*, a significant structural difference is evidenced by the unusual A5 H1′–G7 H8, A5 H2–G7 H1′, A5 H2′1–G7 H8 and A5 H2′2–G7 H8 connectivities. These surprising crosspeaks are indicative of A5–G7 base–base stacking, which is in contrast to *CCGG* where no C5–G7 base stacking was observed.

The phosphate backbone appears to adopt several unusual torsion angles, particularly in the A5–C6 step, and also exhibits some interesting differences compared to *CCGG*. For instance, the unusual C6 H5′1–G7 H8 and C6 H5′2–G7 H8 connectivities seen in *CCGG* are not observed, and unlike in *CCGG*, the intraresidue C6 H6–H5′1 connectivity is stronger than the intraresidue C6 H6–H5′2 connectivity (data not shown). The ^{1}H-^{31}P HETCOR (see Supplementary Figure S1) and DQF-COSY (see Supplementary Figure S2) spectra are very useful in deriving backbone torsion angles. In particular, the β and γ angles can be obtained from the intraresidue P–H4′ crosspeak, the relative H4′–H5′1 and H4′–H5′2 crosspeaks, and the relative intraresidue P–H5′1 and P–H5′2 crosspeaks (16–19). The C6 P–H4′ crosspeak is absent, indicating that one or both of the C6 β and γ torsion angles adopt unusual conformations. Based on the strong H4′–H5′1 crosspeak and absence of a H4′–H5′2 crosspeak in the DQF-COSY spectrum (Supplementary Figure S2), C6 γ adopts the unusual *gauche*$^{-}$ conformation, which is also consistent with the relative intensities of the H3′–H5′1, H3′–H5′2, H4′–H5′2 and C6 H4′–H5′1 $^{2}H_2O$-NOESY crosspeaks (data not shown). The relative intensities of the intraresidue P–H5′1 and P–H5′2 crosspeaks (P–H5′2 is more intense than P–H5′1; Supplementary Figure S1) are consistent with a C6 β angle in either the low end of the *trans* (120–150°) or the *gauche*$^{-}$ (–30 to –60°) conformation. The ζ and α angles are correlated to the phosphorus chemical shift. The C6 phosphorus chemical shift (–3.87) is at the low end of the normal range of –3.8 to –4.8 p.p.m. This appears to be inconsistent with unusual A5 ζ(*t*) and C6 α(*t*) torsion angles, as they would probably lead to a downfield shift of the C6 phosphorus resonance (20).

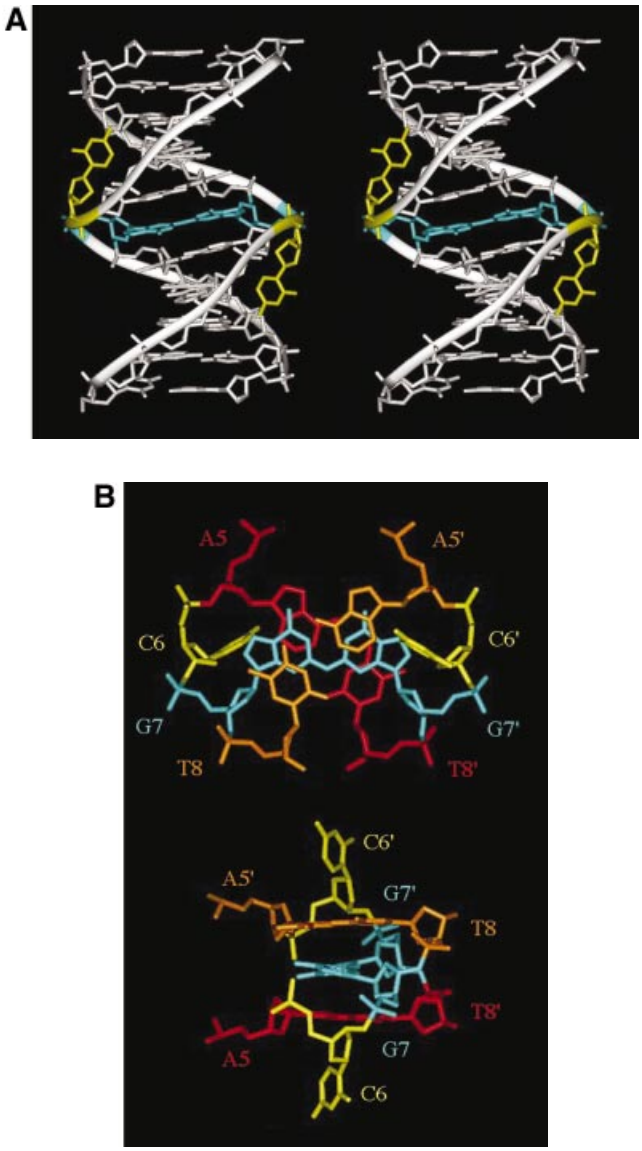

Figure 4. (**A**) The skeletal stereo view of the lowest energy structure of the final set of refined structures of *ACGT*, with the cross-linked guanines (G7 and G7′) colored blue and the extrahelical cytidine residues (C6 and C6′) in yellow. This view is looking into the narrow major groove at the cross-link, showing the nearly planar covalently linked G7:G7′ base pair. The ribbon backbone trace is shown to emphasize the unusual major and minor groove widths. The 5′-ends of the two strands are top-right and bottom-left, respectively. (**B**) Detailed representation of the cross-link region of *ACGT* showing the A5:T8′ base pair in red, the extrahelical C6 and C6′ in yellow, the G7:G7′ base pair in blue, and the T8:A5′ base pair in orange. A top view (top) and a view into the minor groove (bottom) are shown. The G7:G7′ base pair is well stacked on the spatially adjacent A5:T8′ and T8:A5′ base pairs.

Structure determination of *ACGT*

The iterative relaxation matrix and back-calculation refinement process described in Materials and Methods yielded a set of 12 independently generated final structures. The skeletal stereo view of the lowest energy structure is shown in Figure 4A. Overall the structure is very similar to that of *CCGG*, with the bases of residues C6 and C6′ flipped out of the helix and residing in the minor groove, pointing towards the 5′ end of the strand and with its hydrophobic (H5–H6) edge towards the core of the helix. The cross-linked guanines form a nearly planar covalently linked G7:G7′ base pair with only minor propeller twisting, and they are stacked well on the spatially adjacent A5:T8′ and T8:A5′ base pairs. The minor groove is widened to accommodate the extrahelical cytidines, reaching a maximum of 7.2 Å, while the major groove is narrowed to just 7.0 Å at the cross-link as the two strands are

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

Residue	α	β	γ	δ	ε	ζ	χ
C1	–	–	–179 ± 60	144 ± 0	–172 ± 0	–87 ± 0	–119 ± 0
G2	–84 ± 0	179 ± 0	52 ± 0	124 ± 0	177 ± 0	–105 ± 0	–120 ± 0
C3	–67 ± 0	180 ± 0	61 ± 0	133 ± 0	–177 ± 0	–95 ± 0	–122 ± 0
T4	–77 ± 0	176 ± 0	58 ± 0	129 ± 0	–166 ± 0	–137 ± 0	–116 ± 0
A5	–81 ± 0	172 ± 1	68 ± 0	151 ± 1	**–86 ± 0**	**136 ± 1**	–106 ± 3
C6	–83 ± 2	162 ± 3	**–35 ± 6**	141 ± 5	**–121 ± 9**	–67 ± 2	**–71 ± 5**
G7	–74 ± 0	168 ± 4	66 ± 1	**89 ± 3**	180 ± 2	–94 ± 0	–138 ± 2
T8	–66 ± 0	175 ± 0	67 ± 0	137 ± 0	–179 ± 0	–94 ± 0	–107 ± 0
A9	–70 ± 0	177 ± 0	57 ± 0	128 ± 0	–177 ± 0	–109 ± 0	–97 ± 0
G10	–72 ± 0	179 ± 0	54 ± 0	140 ± 0	–173 ± 0	–127 ± 0	–100 ± 0
C11	–67 ± 0	168 ± 0	63 ± 0	141 ± 0	–179 ± 0	–95 ± 0	–114 ± 0
G12	–75 ± 0	–176 ± 0	52 ± 0	132 ± 0	–	–	–119 ± 0

The angles are given in degrees (°) and averaged over the final 12 refined structures. Unusual torsion angles are in bold.

forced together by the short covalent linkage. A closer look at the cross-link region, shown in Figure 4B, reveals that the intrastrand base–base stacking is particularly good between the spatially adjacent purine rings of A5 and G7 (and between A5′ and G7′). There is a large helical twist of 50° between the spatially adjacent A5:T8′ and G7:G7′ base pairs (and between G7:G7′ and T8:A5′), due to the intervening backbone segment of the C6 residue (and C6′). As a result, the duplex is not overall underwound as one might expect. The two cross-linked guanines exhibit a relatively minor propeller twist of –24°. The rise between the G7:G7′ base pair and the adjacent A5:T8′ and T8:A5′ base pairs is 3.1 Å, which is typical for well stacked adjacent base pairs. All residues adopt C2′-*endo* type sugar conformations and anti glycosidic torsion angles, although the C6 χ value is somewhat unusual (–71°) compared to B-DNA. Shown in Table 1 are the torsion angles for the duplex.

Non-exchangeable proton studies of *GACGTC*

In both *CCGG* and *ACGT* the flanking T:A base pairs are in contact with the extrahelical cytidines, resulting in a number of unusual intra- and interstrand NOEs. We wanted to determine if the identity of these flanking base pairs is of importance to the location and flexibility of the extrahelical cytidines by studying *GACGTC*, which has flanking G:C base pairs. Shown in Figure 5 is the H6/H8–H1′/H5 region of the ^2H_2O-NOESY spectrum of *GACGTC*. The diagnostic A5 H1′–G7 H8 and A5 H2–G7 H1′ connectivities, which were also observed in *ACGT*, are present, indicating that there is A5–G7 base–base stacking and that the C6 base is extrahelical in this duplex as well. However, the connectivities involving C6 H5 and H6 are all weak and quite broad, particularly the intraresidue C6 H6–H5 connectivity, which is much weaker than the other cytosine H5–H6 connectivities and the intraresidue C6 H6–H1′ connectivity. This indicates that the extrahelical cytidines are more flexible in this sequence context, and subject to conformational exchange which is intermediate on the proton chemical shift timescale. The relatively slow dynamics, in turn, suggest a significant structural rearrangement. This C6-specific effect is observed both at higher and lower temperatures. We did not observe C6-specific line broadening in the *CCGG* and *ACGT* duplexes at any temperature, which suggests that there is an intrinsic difference in the interaction of the extrahelical cytidines with the minor groove in *GACGTC*. The observed line broadening and flexibility makes *GACGTC* unsuitable for structure determination.

DISCUSSION

Structural comparison between *ACGT* and *CCGG*

Although the structures of *ACGT* and *CCGG* are similar overall, there are some very interesting structural differences. Listed in Table 2 are several of the unusual interproton distances observed in *CCGG* and *ACGT*, and in Table 3 the unusual phosphate backbone conformations. These features are also shown in Figure 6, which depicts the cross-link region of the two duplexes. Most notably, the importance of purine–purine base stacking is apparent. In *CCGG* the cross-linked guanines are stacked well on the 3′ guanines of the adjacent G:C base pairs. In *ACGT* the stacking is still on the purine bases of the adjacent base pairs, but in this case that means the adenines preceding the extrahelical cytidines. Another notable difference lies in the phosphate backbone which adopts different conformations that, none the less, both result in an extrahelical cytidine. In *ACGT* the backbone adopts the unusual ε(g^-) ζ(t) α(g^-) β(t) γ(g^-) conformation in the A5–C6 step, whereas the C5–C6 step in *CCGG* adopts the unusual ε(g^-) ζ(g^+) α(g^+) β(t) γ(t) conformation. Interestingly, none of the dihedral angles in the C6–G7 step differ by more than 15° between *ACGT* and *CCGG*. They are all within the normal ε(t) ζ(g^-) α(g^-) β(t) γ (g^+) conformational domains, yet direct the next base (G7) back into the helix. The unusual A5/C5 ε and ζ angles, which direct the backbone out of the helix, are essentially the same in both duplexes, although A5 ζ (*ACGT*) is in the low *trans* domain and C5 ζ (*CCGG*) in the high *gauche*$^-$ domain. The fact that A5 α is in the very low end of the *trans* domain (136°) might explain why the ζ(g^-) α(t) conformation leads to such a minor phosphorus downfield shift in the spectrum of *ACGT*, although the C6 phosphorus chemical shift is indeed the furthest downfield of the phosphorus resonances. In both duplexes, the C6 α, β and γ angles turn the C6 base into the minor groove. In *ACGT*, only the γ angle is in an unusual conformation (*gauche*$^-$), whereas in *CCGG*, both α and γ are unusual (*gauche*$^+$ and *trans*, respectively). This difference manifests itself in the orientation of C6 H5′1 and H5′2: in *ACGT* they are pointing away

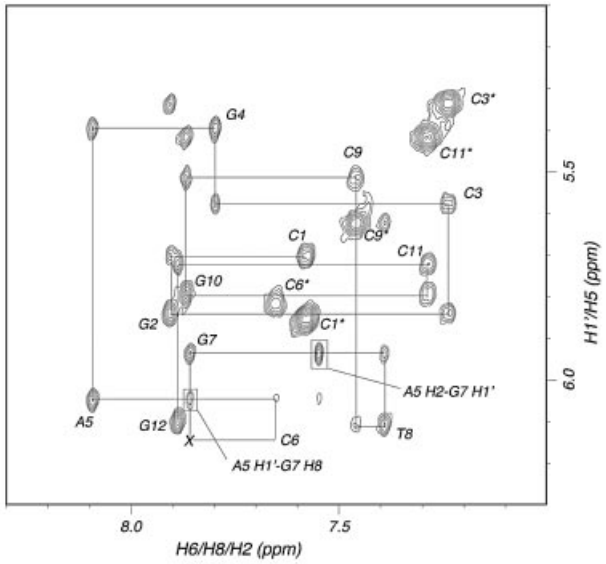

Figure 5. The H6/H8–H1′/H5 region of the ^2H_2O-NOESY spectrum of *GACGTC*, collected at 25°C with a mixing time of 360 ms. The sequential aromatic to H1′ interresidue walk is indicated with lines and the intraresidue H6/H8–H1′ connectivities are labeled with the corresponding residue name and number. The A5 H1′–G7 H8, A5 H2–G7 H1′ and A5 H1′–C6 H6 connectivities are labeled and in boxes. The absence of a C6 H1′–G7 H8 connectivity is marked *X*.

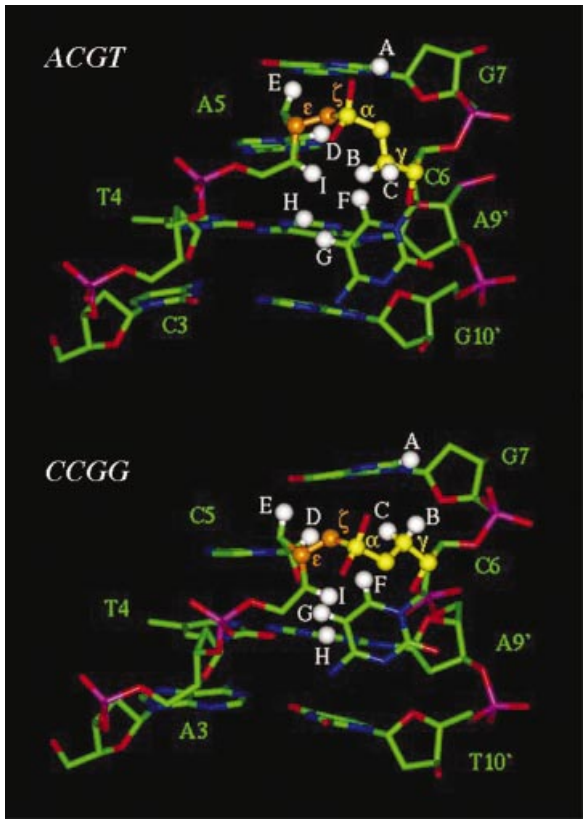

Figure 6. Comparison of the cross-link regions in *ACGT* (top) and *CCGG* (bottom), showing residues C3–G7, A9′ and G10′ of *ACGT*, and residues A3–G7, A9′ and T10′ of *CCGG*. Labeled in white are several protons giving rise to diagnostic NOEs: G7 H8 (**A**), C6 H5′1 (**B**), C6 H5′2 (**C**), C5/A5 H1′ (**D**), C5/A5 H2′2 (**E**), C6 H6 (**F**), C6 H5 (**G**), A9′ H2 (**H**) and C5/A5 H4′ (**I**). Several of the corresponding interproton distances are shown in Table 2. Also labeled are the C5/A5 ε and ζ dihedral angels (orange), and the C6 α and β angles (yellow), the values of which are shown in Table 3.

Table 2. A list of unusual interproton distances in *ACGT* and *CCGG*, compared with idealized B-DNA

Proton pair	*ACGT*	*CCGG*	Idealized B-DNA
G7 H8 (A)–C6 H5′2 (B)	6.7 Å	4.2 Å	>7 Å
G7 H8 (A)–C6 H5′1 (C)	5.9 Å	4.9 Å	>7 Å
G7 H8 (A)–C5/A5 H1′ (D)	3.8 Å	5.5 Å	>7 Å
G7 H8 (A)–C5/A5 H2′2 (E)	4.1 Å	6.5 Å	>7 Å
C6 H6 (F)–C6 H5′2 (B)	5.5 Å	3.3 Å	~4 Å
C6 H6 (F)–C6 H5′1 (C)	4.1 Å	4.8 Å	~3.5 Å
C6 H6 (F)–A9′ H2 (H)	4.5 Å	4.7 Å	>7 Å
C6 H5 (G)–A9′ H2 (H)	3.4 Å	3.2 Å	>7 Å
C6 H6 (F)–T4 H4′ (I)	4.3 Å	3.4 Å	>5.5 Å
C6 H5 (G)–T4 H4′ (I)	5.1 Å	3.8 Å	>6 Å

These protons are shown in Figure 6.

from G7 H8; in *CCGG* they are pointing toward G7 H8. This also results in very different C6 H6–H5′2 distances. The β angle is normal in both duplexes, but the low end of *trans* (162°) observed for *ACGT* is consistent with the $J_{P–H5'1}$ and $J_{P–H5'2}$ coupling data. An additional difference is that the C6 base is pushed further down into the minor groove in *ACGT*, toward the C3:G10′ base pair, although there are no NOE contacts with either of these residues. This does, however, lead to greater C6 H6–T4 H4′ and C6 H5–T4 H4′ interproton distances in *ACGT*. The particular *ACGT* backbone conformation effectively decreases the distance between the A5 and G7 bases allowing them to stack well on each other, which results in G7 H8–A5 H1′ and G7 H8–A5 H2′2 distances that are much shorter than the corresponding G7 H8–C5 H1′ and G7 H8–C5 H2′2 distances in *CCGG*. In both duplexes, the diagnostic interstrand C6 H5–A9′ H2 and C6 H6–A9′ H2 distances are virtually the same. Thus, this is an elegant

Table 3. A list of the unusual dihedral angles and the conformational domains they adopt in *ACGT* and *CCGG*, compared with idealized B-DNA

Dihedral angle	*ACGT*	*CCGG*	Idealized B-DNA
C5/A5 ε	*Gauche*$^-$ (–86°)	*Gauche*$^-$ (–75°)	*Trans*
C5/A5 ζ	High *gauche*$^+$/low *trans* (136°)	High *gauche*$^+$/low *trans* (110°)	*Gauche*$^-$
C6 α	*Gauche*$^-$ (–83°)	*Gauche*$^+$ (65°)	*Gauche*$^-$
C6 γ	*Gauche*$^-$ (–35°)	*Trans* (165°)	*Gauche*$^+$

These dihedral angles are shown in Figure 6.

example of how the phosphate backbone conformation is altered in order to accommodate the purine–purine stacking.

Effects of flanking sequence on mobility of extrahelical cytidines

In both *ACGT* and *CCGG*, the extrahelical C6 base appears to be fixed in the minor groove and lacking flexibility, as evidenced by narrow spectral line widths and an apparent correlation time similar to that of the other cytidine bases (~4 ns). In *CCGG* the C6 amino group is close enough in space to the carbonyl oxygen of T4 to form an intrastrand hydrogen bond. However, such a hydrogen bond would form with less than ideal geometry, making it very weak at best. Furthermore, there is no experimental evidence for this hydrogen bond since the C6 amino protons are exchanging too rapidly with solvent to be observed (1). A hydrogen bonding cytidine amino proton should be protected from exchange. In *ACGT* the C6 amino group is not even within hydrogen bonding distance of the T4 base, and again the amino protons are not observed (data not shown). Thus, it seems that the stability of the cytidine in the minor groove must be derived from other interactions. For example, the adenine H2 in the minor groove could create a small hydrophobic patch into which the cytosine H5 and H6 can fit snugly, resulting in favorable van der Waals interactions. As evidence for this hypothesis, we refer to *GACGTC*, which has a G4:C9′ base pair instead of a T4:A9′ base pair, and in which the extrahelical cytidines are quite mobile. It seems unlikely that this is a steric effect since the cytosine (C9′) is smaller than the adenine (A9′). On the other hand, this cytosine (C9′) has a hydrophilic carbonyl oxygen roughly in the same location as the adenine (A9′) H2. This would presumably prevent favorable van der Waals interactions with the hydrophobic edge of the extrahelical cytidine, and prevent it from fitting snugly in the groove. Furthermore, although it intuitively appears that the fixed location of the extrahelical cytidine would be entropically unfavorable, we suggest that the opposite might be true. If the C6 base were more flexible and exposed to the solvent, a network of water molecules would most likely surround it, which could result in an even more unfavorable entropic contribution. Other possible factors explaining the fixed location of the extrahelical cytidine could be that the geometry of the phosphate backbone simply will not allow this residues to move, and that the C6 χ angle rotation is sterically hindered by the sides of the minor groove (the two backbones). These latter factors, however, would not explain the flexibility observed in *GACGTC*.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

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 no. 1S9O.

REFERENCES

1. Edfeldt,N.B.F., Harwood,E.A., Sigurdsson,S.T., Hopkins,P.B. and Reid,B.R. (2004) Solution structure of a nitrous acid induced DNA interstrand cross-link. *Nucleic Acids Res.*, **32**,
2. Morden,K.M. and Maskos,K. (1993) NMR Studies of an extrahelical cytosine in an A·T rich region of a deoxyribodecanucleotide. *Biopolymers*, **33**, 27–36.
3. Moe,J.G., Reddy,G.R., Lawrence,J.M. and Stone,M.P. (1994) ^{1}H NMR Characterization of a duplex oligonucleotide containing propanodeoxyguanosine opposite a two-base deletion in the $(CpG)_3$ frameshift hotspot of *Salmonella typhimuium hisD3052*. *Chem. Res. Toxicol.*, **7**, 319–328.
4. Gao,X., Huang,X., Smith,G.K., Zheng ,M. and Liu,H. (1995) A new antiparallel duplex motif of DNA CCG repeats that is stabilized by extrahelical bases symmetrically located in the minor groove. *J. Am. Chem. Soc.*, **117**, 8883–8884.
5. Edfeldt,N.B.F. (2001) *2D NMR Studies of Unusual DNA Structures in Solution*. PhD Thesis, University of Washington, Seattle, USA.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. 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.
13. Poltev,V.I., Malenkov,G.G., Gonzalez,E.J., Teplukhin,A.V., Rein,R., Shibata,M. and Miller,J.H. (1996) Modeling DNA hydration: comparision of calculated and experimental hydration properties of nucleic acid bases. *J. Biomol. Struct. Dyn.*, **13**, 717–726.
14. Poltev,V.I., Gonzalez,E.J. and Teplukhin,A.V. (1995) The possible role of rare tautomers of DNA bases in mutagenesis: study of the effect of hydration on tautomeric equilibrium bt the Monte-Carlo method. *Molekuliarnaia Biol.*, **29**, 365–375.
15. 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*. Springer-Verlag, Heidelberg, Vol. 6, pp. 1–22.
16. 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.
17. 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.
18. 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.
19. 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.
20. 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.