**Solution Structure of a Nitrous Acid Induced DNA Interstrand Cross-Link**

N. B. Fredrik Edfeldt[1,2,*], Eric A. Harwood[1,3], Snorri Th. Sigurdsson[1], Paul B. Hopkins[1] and Brian R. Reid[1,].

[1]Department of Chemistry, University of Washington, Seattle, WA 98195, USA.
[2]Present address: AstraZeneca Structural Chemistry Laboratory, AstraZeneca R&D Mölndal, 431 83, Mölndal, Sweden.
[3]Present address: Chiron Corporation, 201 Elliott Avenue West, Suite 150, Seattle, WA 98119, USA.
[*] To whom correspondence should be addressed at 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

**Table of Contents for Supplementary Material**

**SUPPLEMENTARY MATERIAL**

Table S1. Chemical shifts (in ppm) of the non-exchangeable protons and imino protons, and $^{31}P$ of the cross-linked duplex at 25°C.

| Residue | H6/H8 | H5/H2/Me | H1' | H2'1/2 | H3' | H4' | H5'1/2 | H1/H3[a] | $^{31}P$[b] |
|---|---|---|---|---|---|---|---|---|---|
| G1 | 7.94 | - | 5.95 | 2.61/2.78 | 4.84 | 4.26 | 3.72/3.72 | 13.10 | - |
| C2 | 7.45 | 5.41 | 5.66 | 2.13/2.45 | 4.88 | 4.20 | 4.13/4.13 | - | -4.62 |
| A3 | 8.36 | 7.75 | 6.30 | 2.73/2.97 | 5.04 | 4.45 | 4.19/4.10 | - | -4.27 |
| T4 | 7.15 | 1.46 | 5.91 | 2.02/2.33 | 4.83 | 4.13 | 4.30/4.30 | 13.74 | -4.75 |
| C5 | 7.57 | 5.67 | 5.93 | 2.29/2.44 | 4.88 | 4.13 | 4.02/3.97 | - | -4.40 |
| C6 | 7.42 | 5.74 | 6.07 | 1.57/2.19 | 4.48 | 4.27 | 4.12/3.87 | - | -4.13 |
| G7 | 8.02 | - | 5.51 | 2.55/2.56 | 4.92 | 4.27 | 4.05/3.95 | -[c] | -4.43 |
| G8 | 7.85 | - | 5.65 | 2.70/2.82 | 5.01 | 4.37 | 4.11/4.01 | 12.67 | -3.78 |
| A9 | 8.16 | 7.73 | 6.17 | 2.58/2.84 | 4.98 | 4.40 | 4.20/4.20 | - | -4.49 |
| T10 | 7.02 | 1.34 | 5.74 | 1.88/2.32 | 4.83 | 4.10 | 4.22/4.22 | 13.69 | -4.72 |
| G11 | 7.83 | - | 5.90 | 2.58/2.67 | 4.96 | 4.36 | 4.10/4.07 | 12.86 | -4.52 |
| C12 | 7.37 | 5.26 | 6.15 | 2.19/2.21 | 4.48 | 4.04 | 4.25/4.25 | - | -4.27 |

[a] Exchangeable proton spectra collected at 0°C.
[b] Referenced to an external TMP (trimethyl phosphate) standard.
[c] G7 H1 protons have not been observed under any set of conditions.

Table S2. Structural and energetic analysis averaged over the 26 final structures from distance geometry, NOESY back-calculation and restrained molecular dynamics. The units of energy are kcal $mol^{-1}$, and the units of distance Å.

| | |
|---|---|
| Total potential energy | -141.9 ± 1.7 |
| Bond energy | 8.8 ± 0.1 |
| Bond angle deformation energy | 94.2 ± 1.0 |
| Dihedral angle deformation energy | 233.7 ± 0.6 |
| Hydrogen bond energy | -16.0 ± 0.2 |
| Non-bonded energy | -268.2 ± 0.2 |
| Coulomb energy | -194.6 ± 0.7 |
| Distance forcing potential energy[a] | 5.1 ± 0.2 |
| Dihedral forcing potential energy[b] | 0.0 ± 0.0 |
| Total number of distance restraints | 715 |
| Intraresidue | 226 |
| Interresidue[c] | 334 |
| Repulsive | 103 |
| Hydrogen bonding | 52 |
| Total number of dihedral restraints | 178 |
| Backbone | 106 |
| Chiral | 72 |
| Number of distance violations | 0 |
| Maximum distance violation | 0.17 ± 0.01 |
| Pair-wise r.m.s. deviation[d] | 0.13 ± 0.08 |
| NOE R-factor[e] | 0.24 ± 0.00 |

[a] $K_{distance}$ = 30.0 kcal $mol^{-1}$ $Å^{-2}$.
[b] $K_{dihedral}$ = 5.0 to 60.0 kcal $mol^{-1}$ $rad^{-2}$.
[c] Includes 42 qualitative restraints between exchangeable protons derived from the $^{1}H_2O$-NOESY spectrum.
[d] Superimposed on the lowest energy structure, and calculated excluding hydrogen atoms.
[e] Calculated for 131 well resolved crosspeaks at five different mixing times.

Table S3. Selected rotational and translational parameters for the cross-linked duplex. The parameters were calculated using the program RNA with a (0.000, 1.808, 0.000) base pivot point and appropriate coordinate frames for the bases in the G7:G7' base pair, and are averaged over the 26 refined structures.

| Step[a] | Tilt (°) | Roll (°) | Twist (°) | Shift (Å) | Slide (Å) | Rise (Å) |
|---|---|---|---|---|---|---|
| G1-C2 | 0 ± 0 | -1 ± 0 | 44 ± 0 | -0.2 ± 0.0 | -0.8 ± 0.0 | 3.2 ± 0.0 |
| C2-A3 | -2 ± 0 | 8 ± 0 | 34 ± 0 | -0.1 ± 0.0 | -0.7 ± 0.0 | 2.9 ± 0.0 |
| A3-T4 | 0 ± 0 | 3 ± 0 | 30 ± 0 | -0.2 ± 0.0 | -0.7 ± 0.0 | 3.2 ± 0.0 |
| T4-C5 | 3 ± 0 | 9 ± 1 | 38 ± 0 | 0.1 ± 0.0 | 0.9 ± 0.1 | 3.2 ± 0.0 |
| C5-C6[b] | 66 ± 1 | 1 ± 2 | 3 ± 0 | -4.9 ± 0.1 | -2.7 ± 0.1 | -5.6 ± 0.0 |
| C5-G7[c] | 1 ± 1 | 6 ± 1 | 54 ± 0 | -0.8 ± 0.0 | 0.4 ± 0.1 | 3.0 ± 0.0 |
| C6-G7[b] | -61 ± 1 | 36 ± 2 | 41 ± 1 | 3.4 ± 0.1 | 2.9 ± 0.2 | 8.8 ± 0.1 |
| G7-G8 | -1 ± 1 | 6 ± 1 | 54 ± 0 | 0.8 ± 0.0 | 0.4 ± 0.1 | 3.0 ± 0.0 |
| G8-A9 | -3 ± 0 | 9 ± 0 | 38 ± 0 | 0.0 ± 0.0 | 1.0 ± 0.1 | 3.2 ± 0.0 |
| A9-T10 | 0 ± 0 | 3 ± 0 | 30 ± 0 | 0.1 ± 0.0 | -0.7 ± 0.0 | 3.2 ± 0.0 |
| T10-G11 | 2 ± 0 | 8 ± 0 | 34 ± 0 | 0.1 ± 0.0 | -0.7 ± 0.0 | 2.9 ± 0.0 |
| G11-C12 | 0 ± 0 | -1 ± 0 | 44 ± 0 | 0.2 ± 0.0 | -0.8 ± 0.0 | 3.2 ± 0.0 |

| Base Pair | Buckle (°) | Propeller Twist (°) | Opening (°) | Shear (Å) | Stretch (Å) | Stagger (Å) |
|---|---|---|---|---|---|---|
| G1:C12' | 3 ± 0 | -21 ± 0 | 5 ± 0 | -0.2 ± 0.0 | 0.1 ± 0.0 | 0.2 ± 0.0 |
| C2:G11' | -15 ± 1 | -20 ± 0 | 5 ± 0 | 0.7 ± 0.0 | -0.1 ± 0.0 | 0.2 ± 0.0 |
| A3:T10' | -18 ± 0 | -11 ± 1 | 4 ± 0 | 0.4 ± 0.0 | 0.0 ± 0.0 | 0.6 ± 0.0 |
| T4:A9' | -18 ± 1 | -8 ± 1 | 5 ± 0 | 0.0 ± 0.0 | 0.1 ± 0.0 | 0.7 ± 0.1 |
| C5:G8' | 0 ± 1 | -3 ± 1 | 0 ± 0 | 0.5 ± 0.0 | 0.0 ± 0.0 | 0.4 ± 0.1 |
| G7:G7' | 0 ± 1 | -16 ± 1 | 57 ± 0 | 0.0 ± 0.0 | 2.3 ± 0.0 | 0.3 ± 0.0 |
| G8:C5' | 0 ± 1 | -3 ± 1 | 0 ± 0 | -0.5 ± 0.0 | 0.0 ± 0.0 | 0.4 ± 0.1 |
| A9:T4' | 18 ± 0 | -8 ± 1 | 6 ± 0 | 0.0 ± 0.0 | 0.1 ± 0.0 | 0.7 ± 0.0 |
| T10:A3' | 18 ± 0 | -11 ± 0 | 4 ± 0 | -0.4 ± 0.0 | 0.0 ± 0.0 | 0.6 ± 0.0 |
| G11:C2' | 15 ± 0 | -20 ± 0 | 5 ± 0 | -0.7 ± 0.0 | -0.1 ± 0.0 | 0.2 ± 0.0 |
| C12:G1' | -3 ± 0 | -21 ± 0 | 5 ± 0 | 0.2 ± 0.0 | 0.1 ± 0.0 | 0.2 ± 0.0 |

[a] Calculated using covalently adjacent base pairs, except [b] and [c].
[b] Calculated using a base pair-base approach (C5:G8'-C6; C6-G7:G7') (20).
[c] Calculated using spatially adjacent base pairs (C:5:G8'-G7:G7') (20).

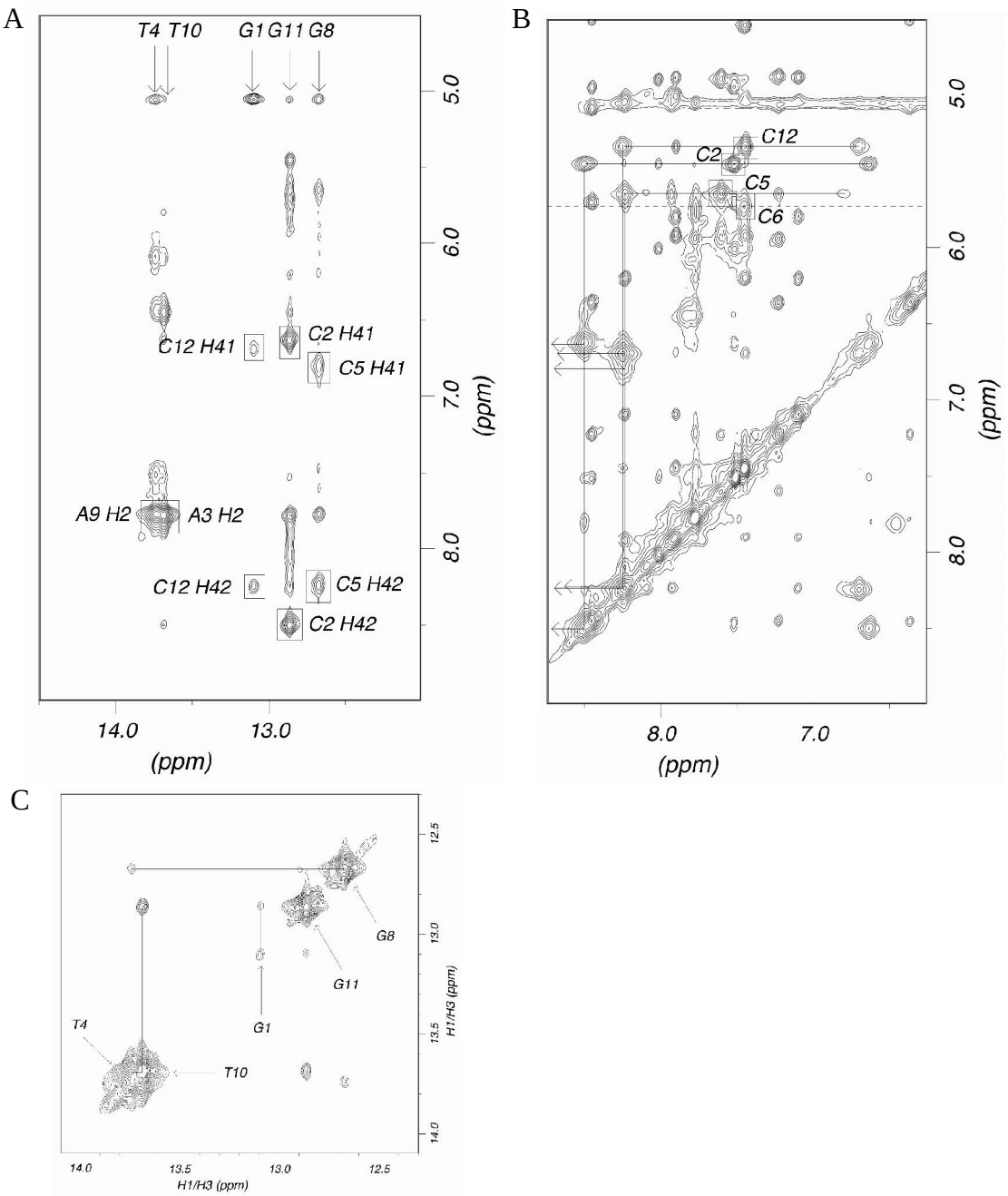



Figure S1. (A) The imino to amino and aromatic proton region of the $^{1}H_{2}O$-NOESY spectrum. The T4, T10, G1, G11 and G8 imino proton resonances are labeled and indicated with arrows. The amino protons of C12, C2 and C5, and the H2s of A9 and A3, are labeled and in boxes. Note that there are only five imino proton to solvent crosspeaks (at *ca.* 5 ppm), none of which is from G7. (B) The amino and aromatic proton region of the $^{1}H_{2}O$-NOESY spectrum. The cytosine H5-H6 connectivities are labeled, indicated with boxes and connected with lines to their respective amino proton pairs. Note the absence of a C6 amino proton pair as indicated by the dashed line. Arrows show the amino protons which give crosspeaks to imino protons in Figure S1A. (C) The imino proton region of the $^{1}H_{2}O$-NOESY spectrum. The intra- and interstrand imino proton walk is indicated, and the interstrand G1-G11, T10-T4 and T4-G8 connectivities and the intrastrand G11-T10 connectivity are labeled. The guanine imino proton resonances were

assigned by their respective interstrand crosspeaks to the amino protons of the hydrogen bonding partner cytosines, which in turn were assigned by their respective intraresidue connectivities to H5, and by the sequential imino proton walk. The thymine imino protons were assigned by their respective interstrand crosspeaks to the H2 of the hydrogen bonding partner adenines, and by the sequential imino proton walk.

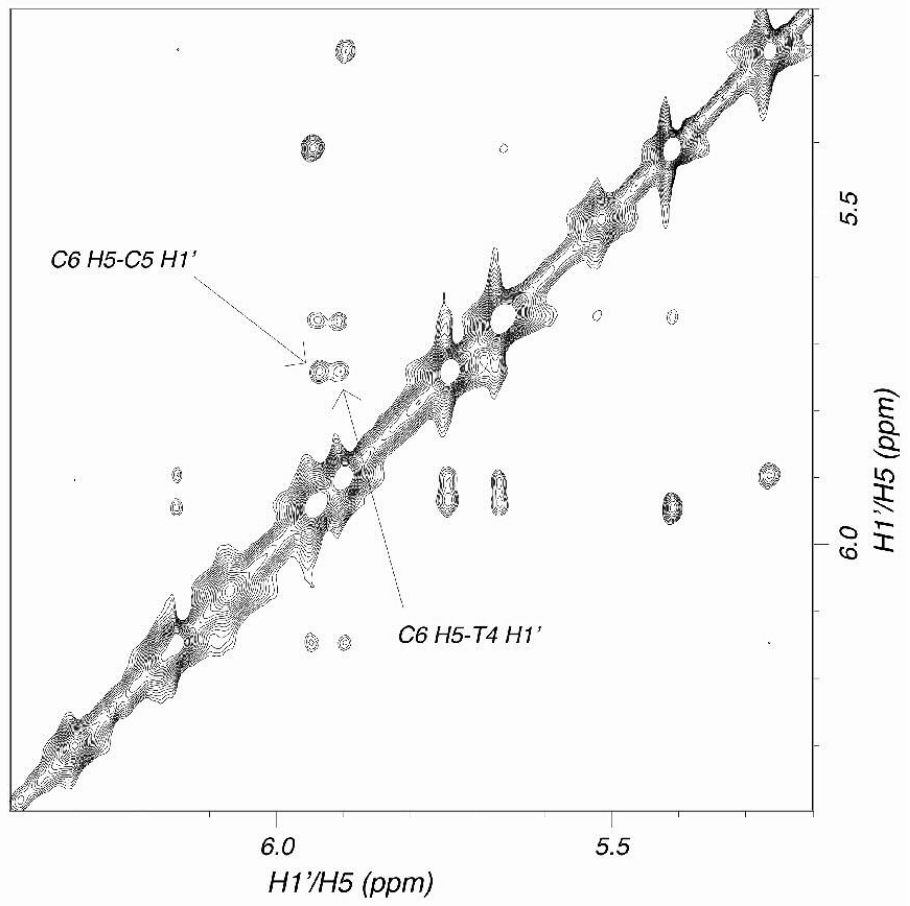



Figure S2. The H1'/H5 region of the $^{2}H_{2}O$-NOESY spectrum, collected at 25°C with a mixing time of 240 ms. The strong C6 H5-C5 H1' and unusual C6 H5-T4 H1' connectivities are labeled and indicated with arrows.

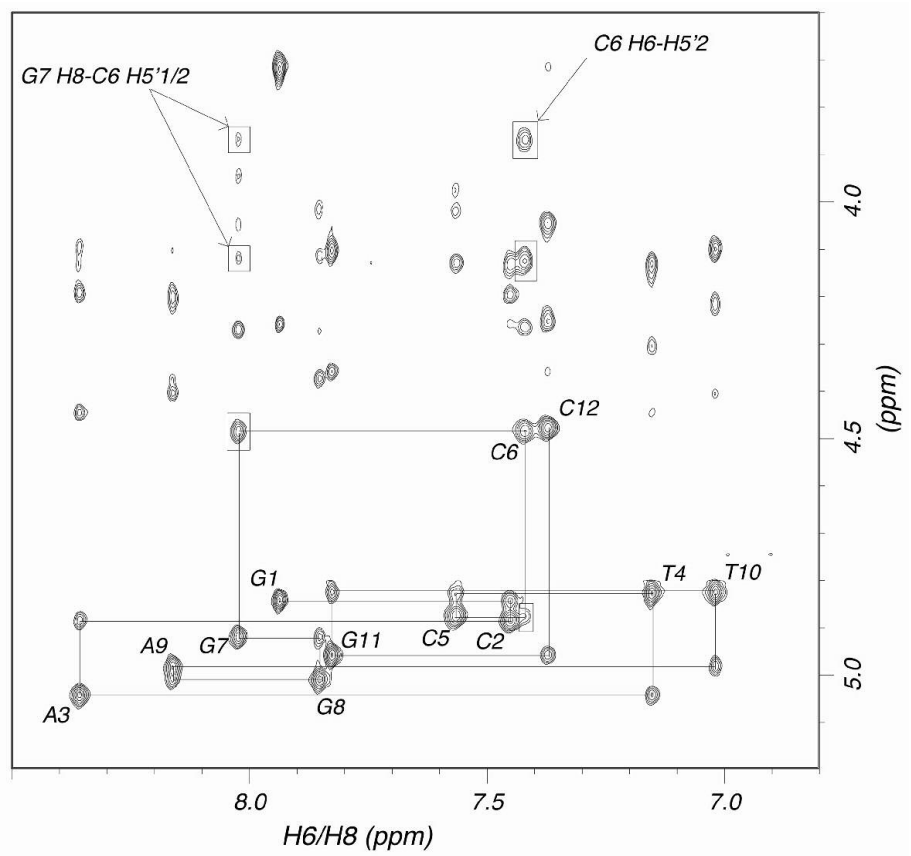



Figure S3. The aromatic to H3'/H4'/H5'1/2 region of the $^{2}H_{2}O$-NOESY spectrum, collected with a mixing time of 240 ms. The sequential aromatic to H3' interresidue walk is indicated with lines and the intraresidue H6/H8-H3' crosspeaks are labeled with the corresponding residue name and number. The C5 H3'-C6 H6 and C6 H3'-G7 H8 connectivities are in boxes. The C6 H6-C6 H5'2, C6 H6-C6 H5'1 (which is overlapped with an unusually strong C5 H4'-C6 H6 connectivity), C6 H5'1-G7 H8 and C6 H5'2-G7 H8 connectivities are labeled and indicated with boxes. The C5 H3'-C6 H6 connectivity is notably weak, which is consistent with an extrahelical location of the C6 base. On the other hand, the C6 H3'-G7 H8 connectivity is quite strong, which suggests that the C6 sugar adopts an unusual orientation relative to the G7 base. In addition, there are highly unusual C6 H5'1-G7 H8 and C6 H5'2-G7 H8 connectivities, which are typically not observed in B-DNA, indicating that the phosphate backbone adopts an unusual conformation at the C6 residue. This is further supported by the unusually strong intraresidue C6 H6-H5'2 connectivity, which is stronger than the intraresidue C6 H6-H5'1, particularly at shorter mixing times (data not shown).

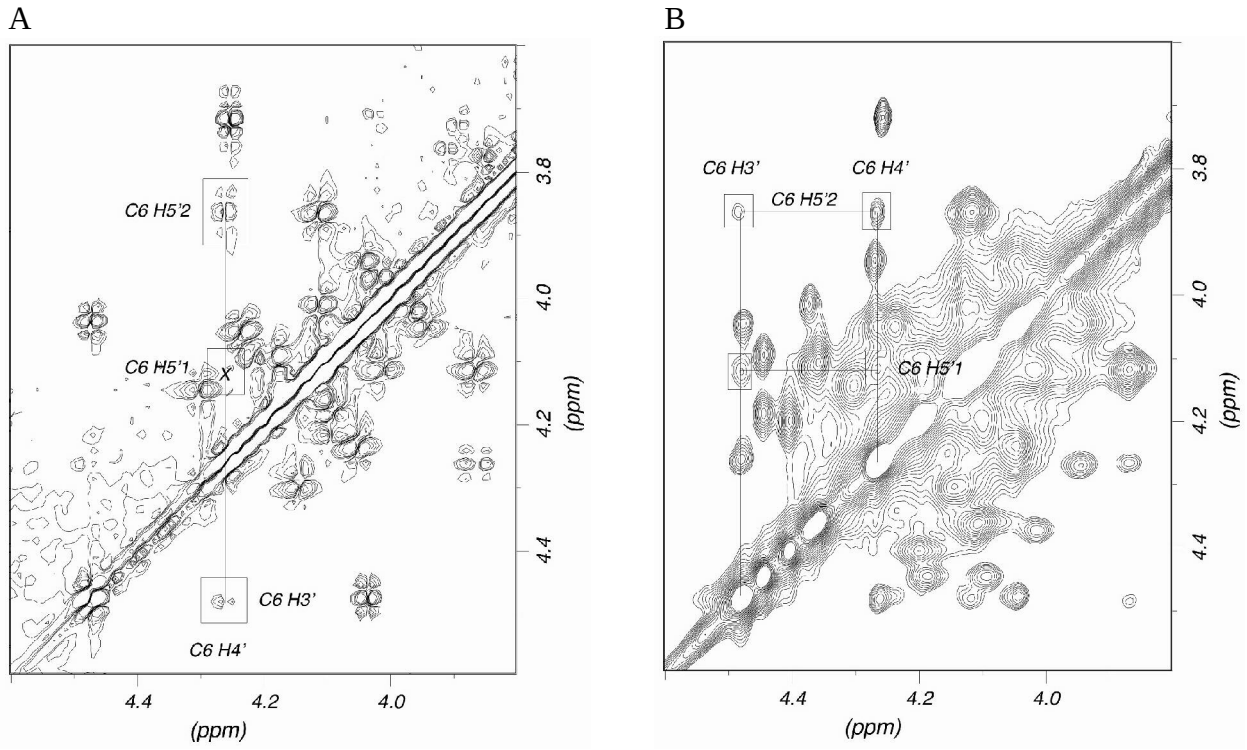



Figure S4. (A) The H3'/H4'/H5' region of the DQF-COSY spectrum. The three bond C6 H3'-H4' and C6 H4'-H5'2 crosspeaks are labeled and indicated with boxes, and the absence of the C6 H4'-H5'1 crosspeak is marked *X*. (B) The H3'/H4'/H5' region of the $^{2}H_{2}O$-NOESY spectrum, collected with a mixing time of 60 ms. The intraresidue C6 H3'-H5'1, H3'-H5'2, H4'-H5'1 and H4'-H5'2 crosspeaks are labeled and indicated with boxes.

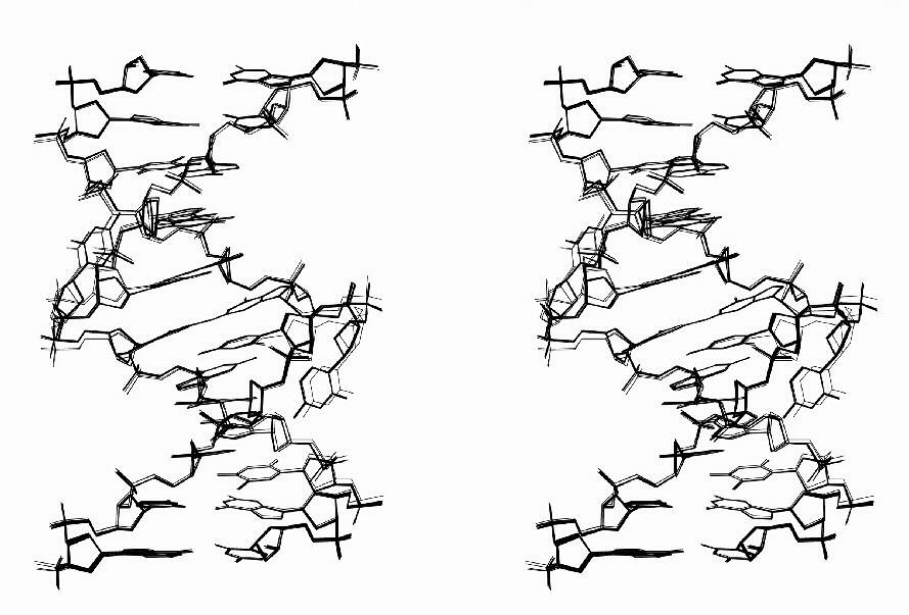
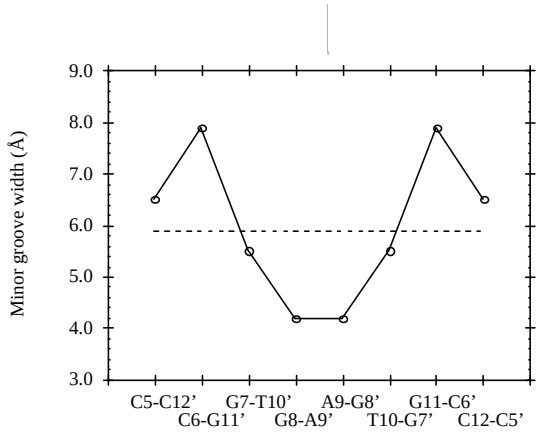


Figure S5. Superimposed skeletal stereo view of the family of 26 final refined structures.

A




B

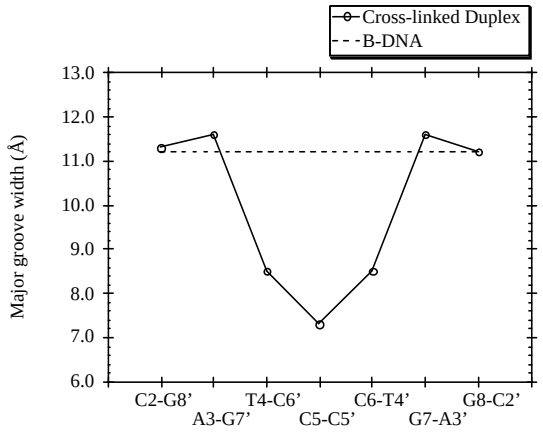
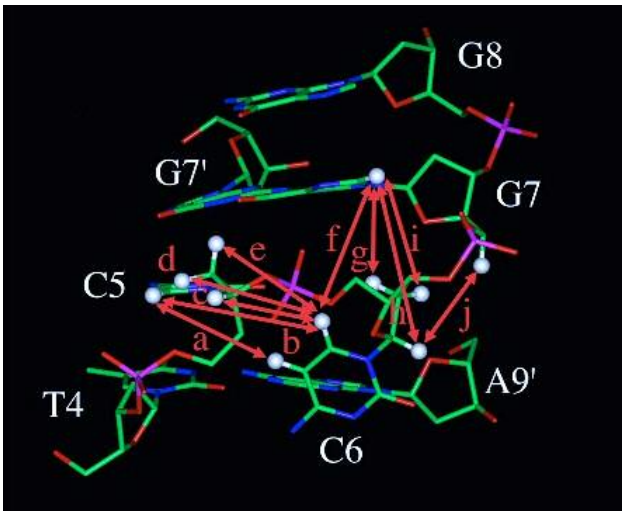
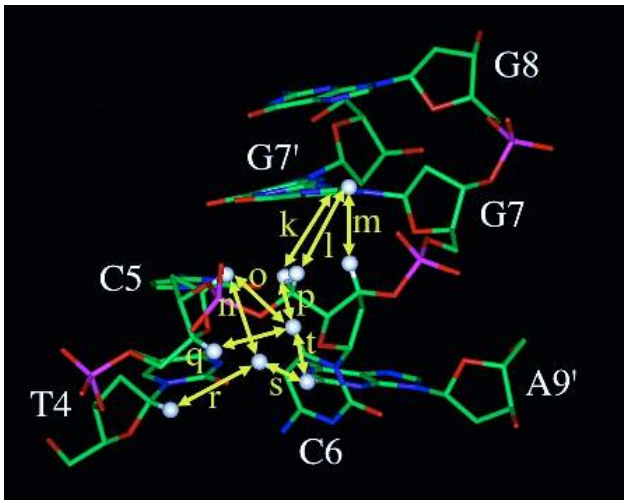



Figure S6. Plot of the minor (A) and major (B) groove widths for the cross-linked duplex, compared with idealized B-DNA. The closest P-P separations across the grooves are plotted. 5.8 Å has been subtracted from each P-P pair to account for the van der Waals radii of phosphate groups.

A




B




C

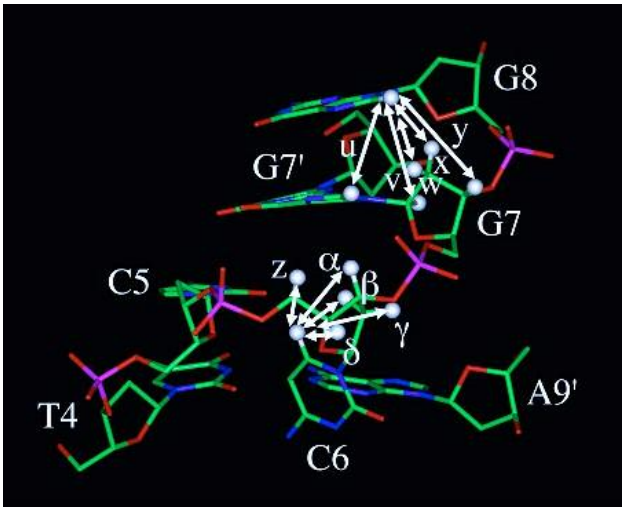



Figure S7. Structural basis for a number of the observed and absent NOEs. Residues T4-G8 of one strand and residues G7’ and A9’ of the opposite strand are shown. For simplicity only the relevant protons are shown. (A) Unusually long interproton distances, resulting in unusually weak or absent NOEs, are labeled and indicated with red arrows: (a) C5 H6-C6 H5, 5.1 Å (~4.5 Å in idealized B-DNA); (b) C5 H6-C6 H6, 6.0 Å (5.0 Å); (c) C5 H3’-C6 H6, 5.5 Å (4.9 Å); (d) C5 H2’1-C6 H6, 5.5 Å (3.9 Å); (e) C5 H2’2-C6 H6, 5.0 Å (2.4 Å); (f) C6 H6-G7 H8, 5.5 Å (4.8 Å); (g) C6 H2’1-G7 H8, 3.8 Å (3.7 Å); (h) C6 H1’-G7 H8, 6.4 Å (2.8 Å); (i) C6 H2’2-G7 H8, 4.5 Å (2.3 Å); (j) C6 H1’-G7 H5’1, 4.3 Å (1.8 Å). (B) Unusually short interproton distances, resulting in unusually strong NOEs, are labeled and indicated with yellow arrows: (k) C6 H5’2-G7 H8, 4.2 Å (7.1 Å); (l) C6 H5’1-G7 H8, 4.9 Å (6.8 Å); (m) C6 H3’-G7 H8, 3.2 Å (5.0 Å); (n) C5 H1’-C6 H5, 3.6 Å (3.7 Å); (o) C5 H1’-C6 H6, 3.0 Å (2.8 Å); (p) C6 H6-C6 H5’2, 3.2 Å (4.1 Å); (q) C5 H4’-C6 H6, 3.4 Å (5.9 Å); (r) T4 H1’-C6 H5, 3.8 Å (6.4 Å); (s) C6 H5-A9’ H2, 3.2 Å

(7.5 Å); (t) C6 H6-A9’ H2, 4.7 Å (7.3 Å). (C) Normal interproton, resulting in normal NOEs, are labeled and indicated with white arrows: (u) G7 H8-G8 H8, 4.5 Å (4.9 Å); (v) G7 H1’-G8 H8, 4.0 Å (2.8 Å); (w) G7 H2’1-G8 H8, 3.5 Å (3.7 Å); (x) G7 H2’2-G8 H8, 2.5 Å (2.3 Å); (y) G7 H3’-G8 H8, 5.1 Å (5.0 Å); (z) C6 H6-C6 H5’1, 4.8 Å (3.3 Å); (α) C6 H6-C6 H3’, 4.1 Å (3.9 Å); (β) C6 H6-C6 H2’1, 2.2 Å (1.9 Å); (γ) C6 H6-C6 H2’2, 3.7 Å (3.4 Å); (δ) C6 H6-C6 H4’, 4.6 Å (4.5 Å).

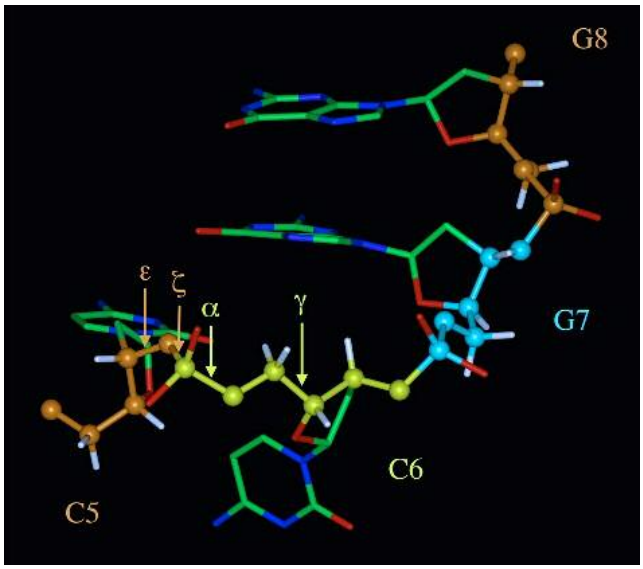



Figure S8. The phosphate backbone in the cross-link region, showing residue C5 (orange), C6 (yellow), G7 (blue), and G8 (orange). Only the unusual C5 ε and ζ, and C6 α and γ dihedral angles are labeled.