HHS Public Access

Author manuscript

Biochemistry. Author manuscript; available in PMC 2022 January 12.

Published in final edited form as:

Biochemistry. 2021 January 12; 60(1): 41–52. doi:10.1021/acs.biochem.0c00596.

Structure of a Stable Interstrand DNA Crosslink Involving a β-*N*-Glycosyl Linkage Between an *N*^{6}-dA Amino Group and an Abasic Site

Andrew H. Kellum Jr.[1], **David Y. Qiu**[1], **Markus W. Voehler**[1], **William Martin**[2], **Kent S. Gates**[3], **Michael P. Stone**[*,1]

[1]Department of Chemistry, Vanderbilt University Center for Structural Biology, and the Vanderbilt-Ingram Cancer Center, Vanderbilt University, Nashville, TN 37235, USA

[2]Department of Biochemistry, Vanderbilt University, Nashville, TN 37232, USA

[3]Departments of Chemistry and Biochemistry, University of Missouri, Columbia, Missouri 65221, USA

Abstract

Abasic (AP) sites are one of the most common forms of DNA damage. The deoxyribose ring of AP sites undergoes anomerization between α and β configurations, via an electrophilic aldehyde intermediate. In sequences where an adenine residue is located on the opposing strand and offset 1 nt to the 3′-side of the AP site, the nucleophilic *N*^{6}-dA amino group can react with the AP aldehyde residue to form an interstrand crosslink (ICL). Here, we present an experimentally determined structure of the dA-AP ICL by NMR spectroscopy. The ICL was constructed in the oligodeoxynucleotide 5′-d(T^{1}A^{2}T^{3}G^{4}T^{5}C^{6}T^{7}A^{8}A^{9}G^{10}T^{11}T^{12}C^{13}A^{14}T^{15}C^{16}T^{17}A^{18})-3′:5′-d(T^{19}A^{20}G^{21}A^{22}T^{23}G^{24}A^{25}A^{26}C^{27}X^{28}T^{29}A^{30}G^{31}A^{32}C^{33}A^{34}T^{35}A^{36})-3′, with the dA-AP ICL forming between A^8 and X^{28}. The NMR spectra indicated an ordered structure for the cross-linked DNA duplex and afforded detailed spectroscopic resonance assignments. Structural refinement, using molecular dynamics calculations restrained by NOE data (rMD), revealed the structure of the ICL. In the dA-AP ICL, the 2′-deoxyribosyl ring of the AP site was ring-closed and in the β configuration. Juxtapositioning the *N*^{6}-dA amino group and the aldehydic C1 of the AP site within bonding distance while simultaneously maintaining two flanking unpaired A^9 and T^{29} bases stacked within the DNA is accomplished by unwinding of the DNA at the ICL. The structural data is discussed in the context of recent studies describing the replication-dependent unhooking of the dA-AP ICL by the base excision repair glycosylase NEIL3.

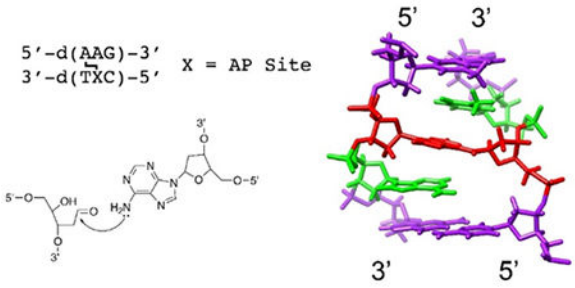
Graphical Abstract

[*]**Corresponding Author**: Michael P. Stone; Phone 615-322-2589; michael.p.stone@vanderbilt.edu.

Supporting Information

The supporting information consists of Tables S1, Chemical shifts of non-exchangeable DNA protons; S2, Chemical shifts of exchangeable protons; S3, NOE distance restraints used for rMD calculations; S4, Pseudorotation angle restraints used in rMD calculations; S5, Calculated partial charges for the dA-AP ICL; S6, Calculated bond lengths for the dA-AP ICL; and Figures S1, Expanded plot of the COSY spectrum of the dA-AP ICL duplex, showing cytosine H5 to H6 correlations; and S2, ^{1}H spectrum showing the thymine N3H and guanine N1H imino protons of the dA-AP ICL.

The Supporting Information is available free of charge on the ACS Publications website.



dA-AP Interstrand Cross-Link in DNA

INTRODUCTION

Abasic sites (AP sites) in DNA result from spontaneous depurination of canonical and chemically-modified nucleobases.[1–3] They are also formed by the removal of damaged bases by DNA glycosylases as a part of the base excision repair (BER) pathway.[4] Spontaneous depurination has been estimated to generate 10^4 AP sites per cell per day, with recent studies measuring levels of AP sites in cellular DNA in the range of 1 x 10^{-7}/nt.[2, 5]AP sites, which equilibrate between α and β anomers via the ring-opened aldehyde form of the 2′-deoxyribosyl ring,[6–9] are genotoxic, with different bases incorporated opposite the lesion depending on the polymerase and sequence context.[10, 11] AP-containing duplexes retain a B-DNA conformation.[6, 7] The presence of an AP site places equilibrium amounts of a ring opened electrophilic aldehyde in the DNA double helix. The propensity for this aldehyde to be attacked by nucleophilic amine residues on the nucleobases of DNA, leading to the formation of interstrand cross-links (ICLs) (Scheme 1), is of interest.[12–21]

Both dG and dA generate equilibrium amounts of ICLs following reaction with an AP site on the opposing strand of a DNA duplex. Once formed, these ICLs are remarkably long-lived, despite the fact that they form reversibly. The dG-AP ICL exhibited a half-life of 130 h at pH 7 and 37 °C.[16] The dA-AP ICL, which forms at higher levels,[17] exhibited a half-life of 66 or 84 h, depending on the surrounding sequences.[22] Thus, AP-derived ICLs may be persistent in cellular DNA. The question as to how these ICLs are repaired is under active investigation. In a *Xenopus* egg extract replication-dependent repair system, the DNA glycosylase NEIL3 unhooked the dA-AP ICL, while the Fanconi anemia pathway came into play only if the levels of NEIL3 were depleted.[23] NEIL3 did not unhook the dA-AP ICL in a fully-paired duplex, but unhooked the ICL in splayed structures that model the strand-separated DNA at the leading edge of a replication fork.[23, 24] In *Xenopus* extracts, ubiquination of the replicative CMG helicase by the E3 ubiquitin ligase TRAIP, activated by collision of the replication fork with the dA-AP ICL, led to recruitment of NEIL3.[25] Studies using the bacteriophage ϕ29 DNA polymerase demonstrated that if not repaired, these ICLs block replication.[26] The reduced dG-AP ICL required multiple polymerases for bypass, and was mutagenic.[27]

From a structural perspective, at the nucleoside level NMR studies of dG-AP and dA-AP ICLs revealed ring-closed aminoglycoside structures with a mixture of α and β anomers at C1 of the 2′-deoxyribosyl ring of the AP residue.[22, 28] Crystallographic analysis confirmed

the presence of the cyclic aminoglycoside linkage for the dG-AP ICL.[29] Molecular dynamics calculations comparing the dG-AP and dA-AP ICL predicted that the two lesions differ both in hydrogen bonding and base stacking patterns.[30] Given the potential biological importance of these lesions, we were keenly interested in determining structures of AP-derived ICLs in DNA.

Here, we present the structure of a dA-AP ICL determined by NMR. This ICL was constructed in the oligodeoxynucleotide 5′-d($T^1A^2T^3G^4T^5C^6T^7A^8A^9G^{10}T^{11}T^{12}C^{13}A^{14}T^{15}C^{16}T^{17}A^{18}$)-3′:5′-d($T^{19}A^{20}G^{21}A^{22}T^{23}G^{24}A^{25}A^{26}C^{27}X^{28}T^{29}A^{30}G^{31}A^{32}C^{33}A^{34}T^{35}A^{36}$)-3′ (X = AP site), with the dA-AP ICL formed between nucleotides A^8 and X^{28}. This oligodeoxynucleotide contained the local 5′-$A^8A^9G^{10}$-3′:5′-$C^{27}X^{28}C^{29}$-3′ sequence at the site of the A^8-X^{28} ICL (Chart 1). This sequence generated high yields of the ICL,[17] and has been used in studies of replication-dependent repair.[23] NMR spectra indicated an ordered duplex, and afforded detailed spectroscopic resonance assignments. The dA-AP ICL exists in the ring-closed aminoglycoside form, with the N^6-amino group of dA attached to the 2-deoxyribosyl ring in the β configuration. A total of 426 NOE restraints were obtained which, in combination with additional restraints based on expected torsion and hydrogen bonding angles, were used to restrain molecular dynamics (rMD) calculations. The structures emerging from the rMD calculations reveal stacking of the unpaired flanking A^9 and T^{29} bases within the DNA at the dA-AP ICL. Movement of the N^6-dA amino group and C1 of the AP site within bonding distance while maintaining the flanking unpaired A^9 and T^{29} bases stacked within the double helix is enabled by unwinding of the DNA near the ICL. The structural data is discussed in the context of recent studies describing an unprecedented pathway involving unhooking of this ICL by the base excision repair glycosylase NEIL3.[23–25, 31, 32]

EXPERIMENTAL SECTION

Materials.

Oligodeoxynucleotides were purchased from Midland Certified Reagents Co. (Midland, TX). They had been subjected to RP-HPLC purification by the supplier. Uracil DNA glycosylase (UDG) was purchased from New England Biolabs Inc. (Ipswich, MA).

Sample Preparation.

The dA-AP ICL formed spontaneously following incubation of the duplex [5′-dTAGATGAAC(dU)TAGAACATA-3′]:[5′-TATGTCTAAGTTCATCTA-3′] in the presence of uracil DNA glycosylase (UDG)[4] at 37 °C for 120 h in 100 mM NaCl, 50 mM HEPES (pH 7.0) (Scheme 1). The solution containing the dA-AP ICL was lyophilized. The product was dissolved in 50 μL of HEPES buffer (50 mM, pH 7) containing NaCl (100 mM) and urea (8 M) and separated from non-cross-linked DNA on a 20% denaturing TBE-polyacrylamide gel. The gel was imaged by UV shadowing. The slowly-migrating band containing the dA-AP cross-linked duplex was excised, added to 10 mL of buffer (50 mM disodium phosphate, pH 8.0), and placed in a rotator shaker overnight at 4 °C. The extracted oligodeoxynucleotide duplex containing the dA-AP ICL was lyophilized, dissolved in 1 mL

of H_2O, and desalted using Sephadex G-25 equilibrated with 1 mM disodium phosphate (pH 8.0).

NMR Spectroscopy.

Samples were prepared in 10 mM NaH_2PO_4, 100 mM NaCl, and 50 μM Na_2EDTA (pH 7). Spectra of non-exchangeable protons were obtained in 99.996% D_2O, while spectra of exchangeable protons were obtained in 95:5 H_2O:D_2O. All spectra were processed using the TOPSPIN software (Bruker Biospin Inc., Billerica, MA) and further analyzed using TOPSPIN and SPARKY[33] software. Spectra were referenced to the chemical shift of water at the corresponding temperature, which was referenced to trimethylsilyl propanoic acid (TSP).

NOESY[34, 35] and magnitude COSY[36] spectra were collected at 900 MHz. The temperature was 298 K. These were collected with 2048 real data points in the t_2 dimension and 512 real data points in the t_1 dimension. The data were zero-filled to obtain final matrices of 2048 x 1024 data points. NOESY spectra in D_2O were collected at NOE mixing times of 60, 150, 200, and 250 ms. NOESY spectra in 95:5 H_2O:D_2O were collected at 278 K. The mixing time was 250 ms. Water suppression was accomplished by excitation sculpting.[37–39] Spectra were collected with 2048 real data points in the t_2 dimension and 512 real data points in the t_1 dimension. The data were zero-filled to obtain final matrices of 2048 x 1024 data points. TOCSY[40] spectra in 95:5 H_2O:D_2O were collected at 298 K. Water suppression was accomplished with excitation sculpting.[37–39] The spin lock time was 120 ms. Spectra were collected with 2048 real data points in the t_2 dimension and 512 real data points in the t_1 dimension. The data were zero-filled to obtain final matrices of 2048 x 1024 data points.

NMR Distance Restraints.

The program SPARKY[33] was used to obtain volume integrations of NOESY cross-peaks from spectra obtained with a 250 ms mixing time at 298 K. The intrinsic error of the integrations was assigned to be half of the volume of the lowest intensity cross-peaks. The volume integration data were divided into five categories based on confidence levels arising from consideration of resonance shape, resonance intensity, degree of resonance overlap, and proximity to the water resonance. A 10% value in error was assigned to well-resolved and non-overlapping cross-peaks. A 20% value in error was assigned to strong but slightly broadened cross-peaks, overlapped cross-peaks, or peaks with moderate S/N. A 30% error value was assigned to strong, but medially broadened or overlapped cross-peaks. A 40% error value was assigned to cross-peaks with weak S/N, or that were slightly overlapped. A 50% error value was assigned to cross-peaks that were highly broadened, near the diagonal or water-suppression, or possessed moderate S/N and medial overlap or broadening. Distance restraints were generated through integration values of NOEs with assigned errors.[41–43] Square potential energy wells corresponding to the upper and lower limits of the interproton distance vectors were employed through the use of MARDIGRAS.[44] Additional distance restraints were created using canonical Watson-Crick hydrogen bonding distances for base pairs in B-DNA. Watson-Crick hydrogen bonding restraints were not employed for the A^8:T^{29} base pair, or for the A^9 base, which was opposite to X^{28}, and which did not have a base pairing partner. The square potential energy well between T^{29} H1′ and A^{30} H8 was

set from a lower bound of 5 Å to an upper bound of 7 Å because a very weak NOE was observed. The square potential energy well between T^7 H1′ and A^8 H8 was set from a lower bound of 6 Å to an upper bound of 10 Å, using an "anti-distance" restraint[45, 46] because no NOE was evident between these two protons.

Restrained Molecular Dynamics Calculations.

Canonical B-DNA values for deoxyribose pseudorotation and phosphodiester backbone were used as restraints in the rMD calculations.[47] However, pseudorotation restraints were not used for A^8, A^9, X^{28}, and T^{29} in the dA-AP ICL region. Pseudorotation restraints were also not used for the terminal and penultimate nucleotides T^1, A^2, T^{17}, A^{18}, T^{19}, A^{20}, T^{35}, and A^{36}. The complete list of pseudorotation restraints that were used is provided in Table S4 in the Supporting Information. Backbone torsion angle restraints for nucleotides were assigned potential energy well windows of ± 30°, with the exception of nucleotides A^8, A^9, X^{28}, and T^{29} neighboring the dA-AP ICL, which were assigned potential energy well windows of ± 60°. The backbone torsion angle restraints were centered at the B-DNA backbone torsion $\alpha = -60°$, $\beta = 180°$, $\gamma = 60°$, $\varepsilon = 195°$, and $\zeta = -105°$. The partial charges and bond lengths for the dA-AP ICL were calculated using the B3LYP/6-31G* basis set in the program GAUSSIAN.[48] These are provided in Table S5 and S6 in the Supporting Information. The dA-AP ICL was constructed using the program MOE.[49] A covalent bond was introduced between the N^6-amino group at A^8 and X^{28} C1, in which the *N*-glycosyl linkage was assigned as the β configuration, based upon experimental observations from NOE intensity data. The bases A^8 and T^{29} were unpaired and were inserted into the duplex, above and below the dA-AP ICL. The resulting structure was further refined using the program xLEaP[50] to include values for partial charges and bond lengths of the dA-AP ICL moiety obtained from GAUSSIAN. The .top and .inp files were generated in xLEaP for use in the program AMBER.[51]

The rMD calculations were performed using a simulated annealing protocol in the program AMBER using the parm99 force field.[52] All restraints had applied force constants of 30 kcal mol^{-1} $Å^{-2}$. Initial calculations were performed for 20 ps over 20,000 steps. The system was heated from 0 to 600 K for the first 1,000 steps with a 0.5 ps coupling. The system was then kept at 600 K for 1,000 steps and cooled to 100 K for 16,000 steps with 4 ps coupling. The generalized Born model was used for solvation.[53] The salt concentration was set at 0.1 M. NOE generated distances were compared to intensities calculated to intensities calculated from emergent structures using complete relaxation matrix analysis (CORMA).[43] The ten structures with the lowest deviations from experimental distance restraints were used to generate an average refined structure. The structural renderings were created in CHIMERA.[54]

Data Deposition.

The structure factors and coordinates have been deposited in the Protein Data Bank and Biological Magnetic Resonance Bank (www.rcsb.org). The PDB ID code is 6XAH. The BMRB ID code is 30759.

RESULTS

NMR Assignments.

(a) Non-exchangeable DNA Protons.—The duplex oligodeoxynucleotide containing the dA-AP ICL contained six C:G base pairs (Chart 1). A COSY spectrum revealed six crosspeaks emanating from scalar couplings between cytosine H5 and cytosine H6 protons. Significantly, only one COSY crosspeak was observed for C^{27}, which was the cytosine nearest to the ICL (Figure S1 in the Supporting Information). This COSY crosspeak was well resolved and of the same linewidth as the corresponding crosspeaks from the other five cytosines. This indicated that the dA-AP ICL exhibited a similar correlation time as the bases not in vicinity of the dA-AP ICL in the DNA, and the ICL formed an ordered structure in the duplex.

The assignments of the deoxyribose H1′ and nucleobase H6 or H8 resonances were accomplished by standard protocol.[5556] A continuous set of sequential NOEs was evident from T^1 H6 NOE to T^7 H1′ (Figure 1). The NOE between T^7 H1′ and A^8 H8 was weaker than normal. The NOE between A^8 H1′ and A^9 H8 was observed. The sequential NOEs continued uninterrupted from A^9 H8 to A^{18} H1′. For the complementary strand, a continuous set of sequential NOEs was evident from the T^{19} H6 to C^{27} H1′ (Figure 1). The connectivity was interrupted due to the absence of a base at X^{28}. The X^{28} H1 resonance was assigned by its NOE to T^{29} H6. A NOE was evident between T^{29} H6 and T^{29} H1′. The NOE between T^{29} H1′ and A^{30} H8 was weak. The connectivity was continuous from NOE between A^{30} H8 to A^{36} H1′. Perturbations included changes in chemical shifts as well as peak intensities when compared to the unmodified duplex. The chemical shift assignments for the non-exchangeable DNA protons are provided Table S1 of the Supporting Information.

(b) Exchangeable DNA Protons.—The imino and amino proton resonance regions of the NOESY spectrum are shown in Figure 2. The imino proton resonances were identified using established methods.[57] The imino protons for the terminal base pairs T^1:A^{36} and A^{18}:T^{19} were not assigned due to their rapid exchange with water. One imino proton resonance was shifted upfield from the thymine N3H and guanine N1H imino protons as shown in Figure 3. This resonance was assigned to the T^{29} N3H proton. The resonance was sharp at 274 K but broadened when the temperature increased and was not visible at 293 K (Figure S2 in the Supporting Information). The resonance was visible when temperature was lowered, indicating a reversible process. The NOEs between guanine N1H imino protons and cytosine N^4H protons were evident for each of the C:G base pairs. The NOEs between the thymine N3H proton and adenine H2 were evident for each base pair except for the two terminal base pairs. In addition, T^{29} N3H did not possess a NOE to A^8 H2. The chemical shift assignments for the exchangeable DNA protons are provided Table S2 of the Supporting Information.

(c) Identification of the dA-AP ICL.—The dA-AP ICL was established from TOCSY data. The X^{28} H1 proton (δ 5.68 ppm), which was unambiguously assigned from its sequential NOE connectivity to the T^{29} H6 proton (Figure 1), exhibited a through bond

correlation to the A^8 N^6H proton (δ 7.88 ppm) (Figure 4). Only one crosspeak was observed between X^{28} H1 and A^8 N^6H. This indicated that at the ICL site, the *N*-glycosyl linkage strongly favored a single configuration, as opposed to a diastereomeric mixture of α and β anomers. Lowering the signal did not reveal the presence of a minor anomer.

Ring-Closed Structure and Anomeric Configuration of the ICL.

The dA-AP ICL has the potential to exist as a ring-closed amine or a ring-opened imine structure. A correlation between the A^8 N^6H proton and X^{28} H1 in the TOCSY spectrum confirmed the ring-closed structure (Figure 4). A NOESY spectrum confirmed the anomeric configuration of the dA-AP ICL at X^{28}. Figure 5 shows the NOEs between X^{28} H2′, H2, and H1 protons, as well as between H2′, H2, and H3 protons. These data were collected at a mixing time of 60 ms to minimize spin diffusion of the deoxyribose protons at X^{28}. The stereotopic assignments of the deoxyribose X^{28} H2′ and H2 protons were accomplished by monitoring relative intensities of their respective NOEs to the deoxyribose H3 proton. The NOE between the deoxyribose H2 and H3 protons was anticipated to be stronger as compared to the NOE between the deoxyribose H2′ and H3 protons, regardless of the anomeric configuration. The X^{28} H2′ proton (δ 2.04 ppm) showed a stronger NOE to the H3 proton (δ 4.67 ppm) than did the H2 proton (δ 2.25 ppm). The intensity of the NOE between H2 (δ 2.04 ppm) and H1 (δ 5.69 ppm) was weak (Figure 5). In contrast, the NOE between the H2′ (δ 2.25 ppm) and H1 (δ 5.69 ppm) protons was of significantly greater intensity. This established that the *N*-glycosyl linkage at the dA-AP ICL was in the β anomeric configuration.

NOEs Between DNA and ICL Protons.

The dA-AP ICL was further characterized by two NOEs involving the A^8 H2 proton (Figure 6). These NOEs were somewhat unusual since in B-DNA, the adenine H8 proton is located in the minor groove and shows relatively few NOEs. The X^{28} H1′ proton (δ 5.68 ppm) showed a NOE to the A^8 H2 proton (δ 7.65 ppm) in the complementary strand. This was consistent with the notion that A^8 was in spatial proximity to X^{28}, and consistent with the formation of the dA-AP ICL. A NOE was also observed between the T^{29} H1′ proton in the complementary strand (δ 5.3 ppm) and the A^8 H2 proton. This indicated that A^8 remained in spatial proximity to the unpaired T^{29} nucleotide.

Chemical Shift Perturbations.

At the site of the dA-AP ICL the chemical shifts of the X^{28} H1 resonance (δ 5.68 ppm) and the A^8 N^6H resonance (δ 7.88 ppm) were consistent with the formation of an *N*-glycosyl linkage, as opposed to an imine linkage (Scheme 1). This was in the same chemical shift range as reported for the nucleoside level.[22] This was consistent with the notion that the *N*-glycosyl linkage should create a thermodynamically stable ICL. At the dA-AP ICL site, the A^8 and A^9 H8 resonances exhibited small upfield chemical shifts, and the A^8 and A^9 H1′ resonances shifted upfield. The A^8 and A^9 H8 resonances shifted approximately 0.2 ppm upfield, while the A^8 and A^9 H1′ resonances shifted upfield by 0.5-1.0 ppm from adenine H1′ resonances that were not in vicinity of the ICL. This was attributed to conformational perturbations at the ICL, which resulted in altered environments for these protons, compared

to B-form DNA. The upfield shift of the T^{29} N3H resonances suggested that this proton was not hydrogen bonded to a complementary base-pairing partner.

Structural Refinement.

A total of 422 NOE-derived distance restraints were obtained from non-exchangeable protons at 298 K (Table S3 of the Supporting Information). These included 175 internuclear restraints, 236 intranuclear restraints, and 11 restraints at the AP site. Of note, nucleotide A^8 participates in the formation of the dA-AP ICL, while nucleotide T^{29} is complementary to A^8. A weak NOE was evident between these two protons. A long-range restraint maintained the distance between T^{29} H1′ and A^{30} H8 between 5 to 7 Å. This conserved the base stacking of T^{29}, consistent with chemical shift data. An "anti-distance" restraint[45, 46] of 6-10 Å was employed between T^7 H1′ and A^8 H8. No NOE was observed, suggesting these protons must be consistently greater than 6 Å apart. The 2-deoxyribosyl ring of the AP site was assigned as the β anomer. Additional restraints included 160 backbone torsion angles, 53 hydrogen bonds, and 115 deoxyribose pseudorestraints, for a total of 754 restraints (Table 1). The latter were included as empirical restraints based on B-DNA.

A series of restrained molecular dynamics (rMD) calculations using a simulated annealing protocol yielded ten emergent structures, from which an average structure was calculated. An overlay of the ten emergent structures and the final average and minimized structure indicated excellent convergence (Figure 7). The ten emergent structures had a maximum rms pairwise difference of 0.864 Å. The average structure had a maximum rms pairwise difference of 0.579 Å, as compared to the ten individual structures.

The average structure was evaluated as to its accuracy by complete relaxation matrix analysis carried out using the program CORMA[43] (Figure 8 and Table 2). Most individual intranucleotide and internucleotide sixth root residuals R^X_1 were less than 0.1, with the overall R^X_1 values being 8.01×10^{-2} and 9.61×10^{-2}, respectively. The overall sixth root residual for the duplex containing the dA-AP ICL R^X_1 was 8.78×10^{-2}. The calculations suggested that the average refined structure was in accordance with the NOE data.

Structure of the dA-AP ICL.

Figures 9 and 10 show views of the dA-AP ICL based upon the lowest energy violation structure that emerged from the rMD calculations and in agreement with the average structure. The refined structure showed that with the 2-deoxyribosyl ring of X^{28} in the β anomeric configuration the ICL spanned the duplex. The unpaired bases A^9 and T^{29} were stacked in the duplex, with A^9 interacting with the 5′ face of G^{10} and 3′ face of A^8 and T^{29} interacting with the 5′ face of A^{30}. The DNA was underwound at ICL site. Bases C^{27} through T^{19} and their complements on the 5′ side and A^{30} through A^{36} and their complements on the 3′ side of X^{28} retained a structure similar to canonical B-DNA.

DISCUSSION

Unrepaired AP sites in genomic DNA are cytotoxic.[10, 11] AP-derived ICLs may impede replication[58] and contribute to cell death and senescence.[59–63] The structure of the dA-AP

ICL determined here provides a foundation for understanding the formation, stability, and repair of this lesion in DNA.

The duplex containing the dA-AP ICL was prepared in good yield and was sufficiently stable to be isolated by gel electrophoresis and characterized by NMR. The observation of a NOE between X^{28} H1 and A^{8} H2 confirms that the ICL was generated by reaction of the A^{8} exocyclic amine with the AP site. The chemical shift data show that the dA-AP ICL is an aminoglycoside attachment, with the 2-deoxyribose of the dA-AP in the ring closed form. The TOCSY correlation between X^{28} H1 and A^{8} N^{6}H provides further evidence for the presence of the dA-AP ICL. Strong NOEs between the X^{28} H2' and H1′ protons show that the non-native ICL glycosidic bond between C1 at X^{28} and A^{8} N^{6} is in the β-configuration.

This contrasts with studies of native AP sites in DNA that showed the α and β anomers existed in similar amounts.[6, 7]

The DNA containing the ICL is unwound relative to canonical B-DNA. NOEs between the unpaired A^{9} H1′ and A^{10} H8 along with the observation that the T^{29} N3H imino proton is in slow exchange with water confirm that unpaired bases T^{29} and A^{9} stack within the helix, with A^{9} stacked between the 3′ face of A^{8} and the 5′-face of G^{10} and T^{29} located on the 3′ side of X^{28} and stacked on the 5′-face of A^{30}. The slow exchange of the T^{29} N3H imino proton with water at higher temperatures suggests that this intrahelical base stacking is stable. Placement of A^{8} N^{6} within bonding distance of the C1 atom in X^{28}, while maintaining the stacking of unpaired bases T^{29} and A^{9} in the duplex, is enabled by local unwinding near the ICL.

In contrast to the unwound but otherwise modestly disrupted structure of the duplex containing the dA-AP ICL, some ICLs, such as those generated by cisplatin[64, 65] and nitrous acid[66], display more significant helical distortions, including extrusion of nucleobases from the duplex. ICLs derived from α,β-unsaturated aldehydes involve reversible reactions of these compounds with nucleobases also lead to differing conformational perturbations of the DNA helix.[67–72] However, ICLs generated from psoralen do not generate large distortions of the DNA.[73–76]

The structure determined here may provide insight regarding the preferred sequences for formation of the dA-AP ICL. The 5′-AAG-3′:5′-CXT-3′ sequence generates approximately 70% equilibrium yields of the dA-AP ICL at pH 7, 37 °C.[17] In contrast, a duplex with the sequence 5′-AAC-3′:5′-GXT-3′ at the cross-linking site gives more modest equilibrium yields of about 15%.[17] From a mechanistic perspective, favorable stacking of the three purine residues, A^{8}, A^{9}, and G^{10} in the 5′-AA<u>G</u>-3′:5′-CXT-3′ sequence (G^{10} underlined) (Figures 9 and 10) determined here may pre-organize the duplex for ICL formation, thus facilitating formation of the dA-AP ICL. The cytosine nucleotide (underlined) in the lower yielding 5′-AA<u>C</u>-3′:5′-GXT-3′ sequence may not provide the same favorable stacking interactions afforded by G^{10} in the structure presented here. Very high (85%) ICL yields were reported in a duplex containing the core sequence 5′-AGA-3′:5′-TXA-3′, with an A-A mismatch on the 3′-side of the AP site.[77] We speculate that this situation might yield an ICL containing an energetically favorable three-purine stack, similar to that seen in the

structure reported here. In the structure solved here, the 3′ neighbor of the unpaired T^{29} was an adenine. The yield decreased only moderately when the 3′ neighbor was changed to guanine.[78] Changing the 3′ neighbor to a cytosine or thymine, on the other hand, significantly reduced the yield to 17%.[78] The presence of a pyrimidine nucleotide at position 30 might not yield a favorable stacking interaction for the unpaired thymine that results from the dA-AP ICL. It seems likely that for the dA-AP ICL the three-purine stack illustrated in Figures 9 and 10 is only possible with the β-configuration at X^{28} C1. The α-configuration in this ICL probably would disrupt the three-purine stack by drawing A^8 away from the helical axis and into the major groove.

The structural effects of AP sites on duplex DNA have been studied by NMR spectroscopy using both the native AP site[6, 7, 79–82] as well as a THF analog.[83–87] The THF analog commonly used as a surrogate for the true AP site is refractory to chemical degradation and also is chemically incapable of ICL formation. One NMR study examined a true AP site in a 5′-AA-3′:5′-XA-3′ sequence that is the preferred sequence motif for the dA-AP ICL.[81] At the AP site two deoxyribose configurations were observed. The major species was assigned as the β anomer while the minor species was assigned as the α anomer. Substantial conformational differences were evident between the two species. From our vantage point this is striking because the two anomers may be anticipated to yield duplexes that are similar in conformation.[88] The conformation of the β anomer resembled those observed in other NMR studies involving AP sites. The base pairs A^5:T^{20} and A^7:T^{18} were maintained with the unpaired adenine occupying the void due to the AP site. The conformation assigned as the α anomer at the AP site was different than other reported conformations of an AP site-containing DNA duplex. In this conformation, A^6 and T^{20} were unpaired with A^5 tilted towards the AP site. The unpaired A^6 and T^{20} in the α anomer conformation are reminiscent of the unpaired bases A^9 and T^{29} in our structure. It seems possible that the conformation assigned as the α-anomer of the AP-containing duplex, in fact, may have actually contained an ICL between the X^{19} site and A^5. Indeed, we find that a 15% equilibrium yield of ICL was generated in DNA containing an AP embedded in the sequence characterized in the earlier study.

In the present sequence the dA-AP ICL is maintained with localized unwinding and stacking of the unpaired bases into the DNA double helix (Figure 9). This may have implications for its repair. For example, the mismatch repair (MMR) machinery repaired ICLs generated by trimethylene linkers in 5′-d(CpG)-3′ and 5′-d(GpC)-3′, but not those generated by nitrogen mustards.[89] The ICLs generated by nitrogen mustard were repaired only during replication. Structural analyses of the trimethylene linkers in 5′-d(CpG)-3′ and 5′- d(GpC)-3′ ICLs revealed significant helical distortion, with 5′-d(GpC)-3′ being more distorted than 5′-d(CpG)-3′.[90, 91] Molecular modeling of ICLs generated by nitrogen mustards revealed differing conformational consequences to the double helix.[92, 93] It seems plausible that global genome (replication-independent) repair pathways initiated by the binding of NER factors to DNA lesions may be modulated by specific conformational consequences unique to these different ICLs. In this context, it will be of interest to examine the recognition and repair of the dA-AP ICL by global genome versus replication- and transcription-dependent pathways of NER.

Along these lines, recent studies identified an unprecedented replication-dependent ICL repair pathway involving unhooking of the dA-AP lesion by the base excision repair glycosylase NEIL3.[23] It is believed that NEIL3 employs its N-terminal amino group to catalyze hydrolytic attack on the C1 carbon of the AP residue in the dA-AP ICL.[94] The reaction generates an “unhooked” (un-cross-linked) duplex containing the native A residue and the AP site. Importantly, NEIL3 fails to unhook the dA-AP ICL when it is located near the center of a fully paired duplex.[23, 24] The structure of the ICL described here helps explain this result. NEIL3 prefers to bind single-stranded DNA substrates and the catalytic machinery of the enzyme cannot easily access the non-native glycosidic linkage of the dA-AP ICL embedded within the DNA (Figure 9).[94–97]

On the other hand, the enzyme unhooks the dA-AP ICL when it is located at the single-strand/duplex junction of a splayed duplex that models the DNA at stalled replication fork. Interestingly, in *Xenopus* egg extracts and biochemical experiments with purified enzyme, NEIL3 selectively unhooks the dA-AP ICL in which the AP site is located on the leading strand of a replication fork and does not unhook the ICL where the AP site resides on the lagging strand. Binding of the leading strand ICL by NEIL3 positions the C1 carbon of the AP residue in the dA-AP ICL for backside attack by the active site amine group (Figure 11). The analogous binding to a lagging strand dA-AP ICL may instead lead to a futile frontside attack by the NEIL3 active site amino group at the C1 atom, which would not displace the N^6-A^8 conjugate.

The transient aldehyde formed at AP sites may form conjugates with other cellular nucleophiles, which are anticipated to modulate the chemical biology of DNA processing. For example, the 5-hydroxymethylcytosine (5-hmC) binding, ESC-specific (HMCES) protein forms a complex with PCNA during replication to scan for AP sites. It conjugates AP sites in single-stranded DNA at replication forks.[98] Crystallographic studies of a DNA-protein conjugate involving the HMCES ortholog YedK in *Escherichia coli* revealed a stable thiazolidine linkage between the ring-opened AP site and the α-amino and sulfhydryl substituents of the amino-terminal cysteine residue.[99] This DNA-protein conjugate prevents repair by endonucleases that would generate either mutations or double-strand breaks, and also prevents error-prone translesion DNA synthesis. In so doing, it both protects AP sites from nucleophilic attack, and directs cellular DNA damage responses towards pathways of resolution with less gentoxic outcomes.

CONCLUSIONS

The structure of the dA-AP ICL was determined in the 5′-AAG-3′:5′-CXC-3′ sequence. NMR indicated an ordered DNA ICL, and afforded detailed spectroscopic resonance assignments. The ICL exists in the ring-closed aminoglycoside form, with the β configuration at the non-native glycosidic linkage between A8 N^6 and X^{28} C1. Structural refinement revealed the stacking of unpaired flanking unpaired A^9 and T^{29} bases within the DNA at the dA-AP ICL. This stacking of the unpaired bases allows the dA-AP ICL to be accommodated by localized unwinding of the DNA at the ICL site. It will be interesting to examine the structures of the dA-AP ICLs in different sequence contexts as well as AP-derived ICLs with other native and unnatural bases.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding Sources

We acknowledge funding provided by NIH grants R01 CA-55678, R01 ES-029357, and P01 CA-160032 (M.P.S.). The Vanderbilt-Ingram Cancer Center is funded by NIH grant P30 CA-068485. Funding for the NMR spectrometers was provided by in part by instrumentation grants S10 RR-05805, S10 RR-025677, and National Science Foundation Instrumentation Grant DBI 0922862, the latter funded by the American Recovery and Reinvestment Act of 2009 (Public Law 111-5). Vanderbilt University assisted with the purchase of NMR instrumentation. Funding for open access charge: National Institutes of Health.

ABBREVIATIONS

AP site	abasic site
BER	base excision repair
CORMA	complete relaxation matrix analysis
COSY	correlation spectroscopy
ICL	interstrand cross-link
MMR	mismatch repair
NMR	nuclear magnetic resonance
NOESY	nuclear Overhauser effect spectroscopy
rMD	restrained molecular dynamics
RMS	root mean square
RP-HPLC	reverse-phased high performance liquid chromatography
THF	tetrahydrofuran
TOCSY	total correlated spectroscopy
UDG	uracil DNA glycosylase

REFERENCES

[1]. Nejad MI, Johnson KM, Price NE, and Gates KS (2016) A new cross-link for an old cross-linking drug: The nitrogen mustard anticancer agent mechlorethamine generates cross-links derived from abasic sites in addition to the expected drug-bridged cross-links, Biochemistry 55, 7033–7041. [PubMed: 27992994]

[2]. Chen H, Yao L, Brown C, Rizzo CJ, and Turesky RJ (2019) Quantitation of apurinic/apyrimidinic sites in isolated DNA and in mammalian tissue with a reduced level of artifacts, Anal. Chem 91, 7403–7410. [PubMed: 31055913]

[3]. Lindahl T, and Nyberg B (1972) Rate of depurination of native deoxyribonucleic acid, Biochemistry 11, 3610–3618. [PubMed: 4626532]

[4]. Lindahl T, Ljungquist S, Siegert W, Nyberg B, and Sperens B (1977) DNA N-glycosidases: Properties of uracil-DNA glycosidase from *Escherichia coli*, J. Biol. Chem 252, 3286–3294. [PubMed: 324994]

[5]. Rahimoff R, Kosmatchev O, Kirchner A, Pfaffeneder T, Spada F, Brantl V, Muller M, and Carell T (2017) 5-Formyl- and 5-carboxydeoxycytidines do not cause accumulation of harmful repair intermediates in stem cells, J. Am. Chem. Soc 139, 10359–10364. [PubMed: 28715893]

[6]. Hoehn ST, Turner CJ, and Stubbe J (2001) Solution structure of an oligonucleotide containing an abasic site: Evidence for an unusual deoxyribose conformation, Nucleic Acids Res. 29, 3413–3423. [PubMed: 11504879]

[7]. Chen J, Dupradeau FY, Case DA, Turner CJ, and Stubbe J (2008) DNA oligonucleotides with A, T, G or C opposite an abasic site: Structure and dynamics, Nucleic Acids Res. 36, 253–262. [PubMed: 18025040]

[8]. Manoharan M, Ransom SC, Mazumder A, Gerlt JA, Wilde JA, Withka JA, and Bolton PH (1988) The characterization of abasic sites in DNA heteroduplexes by site specific labeling with carbon-13, J. Am. Chem. Soc 110, 1620–1622.

[9]. Wilde JA, Bolton PH, Mazumder A, Manoharan M, and Gerlt JA (1989) Characterization of the equilibrating forms of the aldehydic abasic site in duplex DNA by oxygen-17 NMR, J. Am. Chem. Soc 111, 1894–1896.

[10]. Choi JY, Lim S, Kim EJ, Jo A, and Guengerich FP (2010) Translesion synthesis across abasic lesions by human B-family and Y-family DNA polymerases alpha, delta, eta, iota, kappa, and REV1, J. Mol. Biol 404, 34–44. [PubMed: 20888339]

[11]. Patra A, Zhang Q, Lei L, Su Y, Egli M, and Guengerich FP (2015) Structural and kinetic analysis of nucleoside triphosphate incorporation opposite an abasic site by human translesion DNA polymerase eta, J. Biol. Chem 290, 8028–8038. [PubMed: 25666608]

[12]. Dutta S, Chowdhury G, and Gates KS (2007) Interstrand cross-links generated by abasic sites in duplex DNA, J. Am. Chem. Soc 129, 1852–1853. [PubMed: 17253689]

[13]. Sczepanski JT, Jacobs AC, and Greenberg MM (2008) Self-promoted DNA interstrand cross-link formation by an abasic site, J. Am. Chem. Soc 130, 9646–9647. [PubMed: 18593126]

[14]. Sczepanski JT, Jacobs AC, Majumdar A, and Greenberg MM (2009) Scope and mechanism of interstrand cross-link formation by the C4′-oxidized abasic site, J. Am. Chem. Soc 131, 11132–11139. [PubMed: 19722676]

[15]. Sczepanski JT, Hiemstra CN, and Greenberg MM (2011) Probing DNA interstrand cross-link formation by an oxidized abasic site using nonnative nucleotides, Bioorg. Med. Chem. Lett 19, 5788–5793.

[16]. Johnson KM, Price NE, Wang J, Fekry MI, Dutta S, Seiner DR, Wang Y, and Gates KS (2013) On the formation and properties of interstrand DNA-DNA cross-links forged by reaction of an abasic site with the opposing guanine residue of 5′-CAp sequences in duplex DNA, J. Am. Chem. Soc 135, 1015–1025. [PubMed: 23215239]

[17]. Price NE, Johnson KM, Wang J, Fekry MI, Wang Y, and Gates KS (2014) Interstrand DNA-DNA cross-link formation between adenine residues and abasic sites in duplex DNA, J. Am. Chem. Soc 136, 3483–3490. [PubMed: 24506784]

[18]. Gamboa Varela J, and Gates KS (2015) A simple, high-yield synthesis of DNA duplexes containing a covalent, thermally cleavable interstrand cross-link at a defined location, Angew. Chem. Int. Ed. Engl 54, 7666–7669. [PubMed: 25967397]

[19]. Gamboa Varela J, and Gates KS (2016) Simple, high-yield syntheses of DNA duplexes containing interstrand DNA-DNA cross-links between an N^4 -aminocytidine residue and an abasic site, Curr. Protoc. Nucl. Acid Chem 65, 5.16.11–15.16.15. doi 10.1002/cpnc.3

[20]. Nejad MI, Guo X, Housh K, Nel C, Yang Z, Price NE, Wang Y, and Gates KS (2019) Preparation and purification of oligodeoxynucleotide duplexes containing a site-specific, reduced, chemically stable covalent interstrand cross-link between a guanine residue and an abasic site, Methods Mol. Biol 1973, 163–175. [PubMed: 31016701]

[21]. Nejad MI, Price NE, Haldar T, Lewis C, Wang Y, and Gates KS (2019) Interstrand DNA cross-links derived from reaction of a 2-aminopurine residue with an abasic site, ACS Chem. Biol 14, 1481–1489. [PubMed: 31259519]

[22]. Price NE, Catalano MJ, Liu S, Wang Y, and Gates KS (2015) Chemical and structural characterization of interstrand cross-links formed between abasic sites and adenine residues in duplex DNA, Nucleic Acids Res. 43, 3434–3441. [PubMed: 25779045]

[23]. Semlow DR, Zhang J, Budzowska M, Drohat AC, and Walter JC (2016) Replication-dependent unhooking of DNA interstrand cross-links by the NEIL3 glycosylase, Cell 167, 498–511 e414. [PubMed: 27693351]

[24]. Imani Nejad M, Housh K, Rodriguez AA, Haldar T, Kathe S, Wallace SS, Eichman BF, and Gates KS (2020) Unhooking of an interstrand cross-link at DNA fork structures by the DNA glycosylase NEIL3, DNA Repair (Amst) 86, 102752. [PubMed: 31923807]

[25]. Wu RA, Semlow DR, Kamimae-Lanning AN, Kochenova OV, Chistol G, Hodskinson MR, Amunugama R, Sparks JL, Wang M, Deng L, Mimoso CA, Low E, Patel KJ, and Walter JC (2019) TRAIP is a master regulator of DNA interstrand crosslink repair, Nature 567, 267–272. [PubMed: 30842657]

[26]. Yang Z, Price NE, Johnson KM, and Gates KS (2015) Characterization of interstrand DNA-DNA cross-links derived from abasic sites using bacteriophage varphi29 DNA polymerase, Biochemistry 54, 4259–4266. [PubMed: 26103998]

[27]. Price NE, Li L, Gates KS, and Wang Y (2017) Replication and repair of a reduced 2-deoxyguanosine-abasic site interstrand cross-link in human cells, Nucleic Acids Res. 45, 6486–6493. [PubMed: 28431012]

[28]. Catalano MJ, Liu S, Andersen N, Yang Z, Johnson KM, Price NE, Wang Y, and Gates KS (2015) Chemical structure and properties of interstrand cross-links formed by reaction of guanine residues with abasic sites in duplex DNA, J. Am. Chem. Soc 137, 3933–3945. [PubMed: 25710271]

[29]. Catalano MJ, Ruddraraju KV, Barnes CL, and Gates KS (2016) Crystal structure of a nucleoside model for the interstrand cross-link formed by the reaction of 2′-deoxyguanosine and an abasic site in duplex DNA, Acta Cryst. E 72, 624–627.

[30]. Bignon E, Drsata T, Morell C, Lankas F, and Dumont E (2017) Interstrand cross-linking implies contrasting structural consequences for DNA: Insights from molecular dynamics, Nucleic Acids Res. 45, 2188–2195. [PubMed: 27986856]

[31]. Yang Z, Nejad MI, Varela JG, Price NE, Wang Y, and Gates KS (2017) A role for the base excision repair enzyme NEIL3 in replication-dependent repair of interstrand DNA cross-links derived from psoralen and abasic sites, DNA Repair (Amst) 52, 1–11. [PubMed: 28262582]

[32]. Li N, Wang J, Wallace SS, Chen J, Zhou J, and D′Andrea AD (2020) Cooperation of the NEIL3 and Fanconi anemia/BRCA pathways in interstrand crosslink repair, Nucleic Acids Res. 48, 3014–3028. [PubMed: 31980815]

[33]. Goddard TD, and Kneller DG (2006) SPARKY v. 3.113, University of California, San Francisco.

[34]. Jeener J, Meier BH, Bachmann P, and Ernst RR (1976) Investigation of exchange processes by two-dimensional NMR spectroscopy, J. Chem. Phys 71, 4546–4553.

[35]. Wagner R, and Berger S (1996) Gradient-selected NOESY - A fourfold reduction of the measurement time for the NOESY experiment, J. Magn. Res. A 123, 119–121.

[36]. Aue WP, Bartholdi E, and Ernst RR (1976) Two-dimensional spectroscopy. Application to nuclear magnetic resonance, J. Chem. Phys 64, 2229–2246.

[37]. Emetarom C, Hwang TL, Mackin G, and Shaka AJ (1995) Isotope editing of NMR spectra: Excitation sculpting using Bird pulses, J. Magn. Reson. A 115, 137–140.

[38]. Hwang TL, and Shaka AJ (1995) Water suppression that works: Excitation sculpting using arbitrary wave-forms and pulsed-field gradients, J. Magn. Reson. A 112, 275–279.

[39]. Stott K, Stonehouse J, Keeler J, Hwang TL, and Shaka AJ (1995) Excitation sculpting in high-resolution nuclear magnetic-resonance spectroscopy: Application to selective NOE experiments, J. Am. Chem. Soc 117, 4199–4200.

[40]. Bax A, and Davis DG (1985) MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy, J. Magn. Reson 65, 355–360.

[41]. Keepers JW, and James TL (1984) A theoretical study of distance determination from NMR. Two-dimensional nuclear Overhauser effect spectra, J. Magn. Reson 57, 404–426.

[42]. James TL (1991) Relaxation matrix analysis of two-dimensional nuclear Overhauser effect spectra, Curr. Opin. Struct. Biol 1, 1042–1053.

[43]. Borgias BA, and James TL (1989) Two-dimensional nuclear Overhauser effect: Complete relaxation matrix analysis, Methods Enzymol. 176, 169–183. [PubMed: 2811685]

[44]. Borgias BA, and James TL (1990) MARDIGRAS--a procedure for matrix analysis of relaxation for discerning geometry of an aqueous structure, J. Magn. Reson 87, 475–487.

[45]. Bruschweiler R, Blackledge M, and Ernst RR (1991) Multi-conformational peptide dynamics derived from NMR data: A new search algorithm and its application to antamanide, J. Biomolec. NMR 1, 3–11.

[46]. Ippel JH, Lanzotti V, Galeone A, Mayol L, van den Boogaart JE, Pikkemaat JA, and Altona C (1995) Conformation of the circular dumbbell d<pCGC-TT-GCG-TT>: Structure determination and molecular dynamics, J. Biomolec. NMR 6, 403–422.

[47]. Arnott S, and Hukins DWL (1972) Optimised parameters for A-DNA and B-DNA, Biochem. Biophys. Res. Commun 47, 1504–1509. [PubMed: 5040245]

[48]. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Pomelli J, Ochterski W, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewska VG, Daniels AD, Farkas O, Rabuck AD, Raghavachari K, and Ortiz JV (2004) GAUSSIAN 03, Gaussian, Inc., Wallingford, CT.

[49]. (2015) Molecular Operating Environment (MOE), 2013.08 ed., Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7.

[50]. Schafmeister CEA, Ross WS, and Romanovski V (1995) XLEAP, University of California, San Francisco, San Francisco.

[51]. Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KM Jr., Onufriev A, Simmerling C, Wang B, and Woods RJ (2005) The AMBER biomolecular simulation programs, J. Comput. Chem 26, 1668–1688. [PubMed: 16200636]

[52]. Wang JM, Cieplak P, and Kollman PA (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?, J. Comput. Chem 21, 1049–1074.

[53]. Bashford D, and Case DA (2000) Generalized Born models of macromolecular solvation effects, Annu. Rev. Phys. Chem 51, 129–152. [PubMed: 11031278]

[54]. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, and Ferrin TE (2004) UCSF chimera: A visualization system for exploratory research and analysis, J. Comput. Chem 25, 1605–1612. [PubMed: 15264254]

[55]. Patel DJ, Shapiro L, and Hare D (1987) DNA and RNA: NMR studies of conformations and dynamics in solution, Q. Rev. Biophys 20, 35–112. [PubMed: 2448843]

[56]. Reid BR (1987) Sequence-specific assignments and their use in NMR studies of DNA structure, Q. Rev. Biophys 20, 2–28.

[57]. Boelens R, Scheek RM, Dijkstra K, and Kaptein R (1985) Sequential assignment of imino- and amino-proton resonances in ^{1}H NMR spectra of oligonucleotides by two-dimensional NMR spectroscopy. Application to a lac operator fragment, J. Magn. Reson 62, 378–386.

[58]. Yang J, Wang R, Liu B, Xue Q, Zhong M, Zeng H, and Zhang H (2015) Kinetic analysis of bypass of abasic site by the catalytic core of yeast DNA polymerase eta, Mutat. Res 779, 134–143. [PubMed: 26203649]

[59]. Rajski SR, and Williams RM (1998) DNA cross-linking agents as antitumor drugs, Chem. Rev 98, 2723–2796. [PubMed: 11848977]

[60]. Garinis GA, van der Horst GT, Vijg J, and Hoeijmakers JH (2008) DNA damage and ageing: New-age ideas for an age-old problem, Nat. Cell Biol 10, 1241–1247. [PubMed: 18978832]

[61]. Niedernhofer LJ, Lalai AS, and Hoeijmakers JH (2005) Fanconi anemia (cross)linked to DNA repair, Cell 123, 1191–1198. [PubMed: 16377561]

[62]. Grillari J, Katinger H, and Voglauer R (2007) Contributions of DNA interstrand cross-links to aging of cells and organisms, Nucleic Acids Res. 35, 7566–7576. [PubMed: 18083760]

[63]. Bergstralh DT, and Sekelsky J (2008) Interstrand crosslink repair: Can XPF-ERCC1 be let off the hook?, Trends Genet. 24, 70–76. [PubMed: 18192062]

[64]. Huang H, Zhu L, Reid BR, Drobny GP, and Hopkins PB (1995) Solution structure of a cisplatin-induced DNA interstrand cross-link, Science 270, 1842–1845. [PubMed: 8525382]

[65]. Paquet F, Perez C, Leng M, Lancelot G, and Malinge JM (1996) NMR solution structure of a DNA decamer containing an interstrand cross-link of the antitumor drug cis-diamminedichloroplatinum (II), J. Biomol. Struct. Dyn 14, 67–77. [PubMed: 8877563]

[66]. Edfeldt NB, Harwood EA, Sigurdsson ST, Hopkins PB, and Reid BR (2004) Solution structure of a nitrous acid induced DNA interstrand cross-link, Nucleic Acids Res. 32, 2785–2794. [PubMed: 15155847]

[67]. Cho YJ, Kim HY, Huang H, Slutsky A, Minko IG, Wang H, Nechev LV, Kozekov ID, Kozekova A, Tamura P, Jacob J, Voehler M, Harris TM, Lloyd RS, Rizzo CJ, and Stone MP (2005) Spectroscopic characterization of interstrand carbinolamine cross-links formed in the 5′-CpG-3′ sequence by the acrolein-derived gamma-OH-1,N^2-propano-2′-deoxyguanosine DNA adduct, J. Am. Chem. Soc 127, 17686–17696. [PubMed: 16351098]

[68]. Cho YJ, Wang H, Kozekov ID, Kurtz AJ, Jacob J, Voehler M, Smith J, Harris TM, Lloyd RS, Rizzo CJ, and Stone MP (2006) Stereospecific formation of interstrand carbinolamine DNA cross-links by crotonaldehyde- and acetaldehyde-derived alpha-CH_3-gamma-OH-1,N^2-propano-2′-deoxyguanosine adducts in the 5′-CpG-3′ sequence, Chem. Res. Toxicol 19, 195–208. [PubMed: 16485895]

[69]. Stone MP, Cho Y-J, Huang H, Kim H-Y, Kozekov ID, Kozekova A, Wang H, Lloyd RS, Harris TM, and Rizzo CJ (2008) Interstrand DNA cross-links induced by α,β-unsaturated aldehydes derived from lipid peroxidation and environmental sources, Acc. Chem. Res 41, 793–804. [PubMed: 18500830]

[70]. Minko IG, Kozekov ID, Harris TM, Rizzo CJ, Lloyd RS, and Stone MP (2009) Chemistry and biology of DNA containing 1,N^2-deoxyguanosine adducts of the α,β-unsaturated aldehydes acrolein, crotonaldehyde, and 4-hydroxynonenal, Chem. Res. Toxicol 22, 759–778. [PubMed: 19397281]

[71]. Huang H, Wang H, Lloyd RS, Rizzo CJ, and Stone MP (2009) Conformational interconversion of the trans-4-hydroxynonenal-derived (6*S*,8*R*,11*S*) 1,N^2-deoxyguanosine adduct when mismatched with deoxyadenosine in DNA, Chem. Res. Toxicol 22, 187–200. [PubMed: 19053179]

[72]. Huang H, Kozekov ID, Kozekova A, Wang H, Lloyd RS, Rizzo CJ, and Stone MP (2010) DNA cross-link induced by trans-4-hydroxynonenal, Environ. Mol. Mutagen 51, 625–634. [PubMed: 20577992]

[73]. Tomic MT, Wemmer DE, and Kim SH (1987) Structure of a psoralen cross-linked DNA in solution by nuclear magnetic resonance, Science 238, 1722–1725. [PubMed: 3686011]

[74]. Spielmann HP, Dwyer TJ, Hearst JE, and Wemmer DE (1995) Solution structures of psoralen monoadducted and cross-linked DNA oligomers by NMR spectroscopy and restrained molecular dynamics, Biochemistry 34, 12937–12953. [PubMed: 7548052]

[75]. Spielmann HP, Dwyer TJ, Sastry SS, Hearst JE, and Wemmer DE (1995) DNA structural reorganization upon conversion of a psoralen furan-side monoadduct to an interstrand cross-link: Implications for DNA repair, Proc. Natl. Acad. Sci. USA 92, 2345–2349. [PubMed: 7892269]

[76]. Spielmann HP, Chi DY, Hunt NG, Klein MP, and Hearst JE (1995) Spin-labeled psoralen probes for the study of DNA dynamics, Biochemistry 34, 14801–14814. [PubMed: 7578089]

[77]. Nejad MI, Shi R, Zhang X, Gu LQ, and Gates KS (2017) Sequence-specific covalent capture coupled with high-contrast nanopore detection of a disease-derived nucleic acid sequence, Chembiochem. 18, 1383–1386. [PubMed: 28422400]

[78]. Johnson KM (2014) Hypoxia selective DNA alkylating analogs of tirapazamine and AP-derived DNA interstrand cross-links, PhD Thesis, University of Missouri, Columbia, MO, 2014. MOSpace Institutional Repository, https://mospace.umsystem.edu/xmlui/handle/10355/44472

[79]. Cuniasse P, Sowers LC, Eritja R, Kaplan B, Goodman MF, Cognet JA, Lebret M, Guschlbauer W, and Fazakerley GV (1987) An abasic site in DNA. Solution conformation determined by proton

NMR and molecular mechanics calculations, Nucleic Acids Res. 15, 8003–8022. [PubMed: 3671070]

[80]. Goljer I, Kumar S, and Bolton PH (1995) Refined solution structure of a DNA heteroduplex containing an aldehydic abasic site, J. Biol. Chem 270, 22980–22987. [PubMed: 7559436]

[81]. Wang KY, Parker SA, Goljer I, and Bolton PH (1997) Solution structure of a duplex DNA with an abasic site in a dA tract, Biochemistry 36, 11629–11639. [PubMed: 9305952]

[82]. Beger RD, and Bolton PH (1998) Structures of apurinic and apyrimidinic sites in duplex DNAs, J. Biol. Chem 273, 15565–15573. [PubMed: 9624147]

[83]. Cline SD, Jones WR, Stone MP, and Osheroff N (1999) DNA abasic lesions in a different light: Solution structure of an endogenous topoisomerase II poison, Biochemistry 38, 15500–15507. [PubMed: 10569932]

[84]. Kalnik MW, Chang CN, Grollman AP, and Patel DJ (1988) NMR studies of abasic sites in DNA duplexes: deoxyadenosine stacks into the helix opposite the cyclic analogue of 2-deoxyribose, Biochemistry 27, 924–931. [PubMed: 3259144]

[85]. Cuniasse P, Sowers LC, Eritja R, Kaplan B, Goodman MF, Cognet JA, Le Bret M, Guschlbauer W, and Fazakerley GV (1989) Abasic frameshift in DNA. Solution conformation determined by proton NMR and molecular mechanics calculations, Biochemistry 28, 2018–2026. [PubMed: 2541770]

[86]. Kalnik MW, Chang CN, Johnson F, Grollman AP, and Patel DJ (1989) NMR studies of abasic sites in DNA duplexes: Deoxyadenosine stacks into the helix opposite acyclic lesions, Biochemistry 28, 3373–3383. [PubMed: 2545258]

[87]. Cuniasse P, Fazakerley GV, Guschlbauer W, Kaplan BE, and Sowers LC (1990) The abasic site as a challenge to DNA polymerase. A nuclear magnetic resonance study of G, C and T opposite a model abasic site, J. Mol. Biol 213, 303–314. [PubMed: 2342108]

[88]. de Los Santos C, El-Khateeb M, Rege P, Tian K, and Johnson F (2004) Impact of the C1′ configuration of abasic sites on DNA duplex structure, Biochemistry 43, 15349–15357. [PubMed: 15581347]

[89]. Kato N, Kawasoe Y, Williams H, Coates E, Roy U, Shi Y, Beese LS, Schärer OD, Yan H, Gottesman ME, Takahashi TS, and Gautier J (2017) Sensing and processing of DNA interstrand crosslinks by the mismatch repair pathway, Cell Rep. 21, 1375–1385. [PubMed: 29091773]

[90]. Dooley PA, Tsarouhtsis D, Korbel GA, Nechev LV, Shearer J, Zegar IS, Harris CM, Stone MP, and Harris TM (2001) Structural studies of an oligodeoxynucleotide containing a trimethylene interstrand cross-link in a 5′-(CpG) motif: Model of a malondialdehyde cross-link, J. Am. Chem. Soc 123, 1730–1739. [PubMed: 11456774]

[91]. Dooley PA, Zhang M, Korbel GA, Nechev LV, Harris CM, Stone MP, and Harris TM (2003) NMR determination of the conformation of a trimethylene interstrand cross-link in an oligodeoxynucleotide duplex containing a 5′-d(GpC) motif, J. Am. Chem. Soc 125, 62–72. [PubMed: 12515507]

[92]. Guainazzi A, Campbell AJ, Angelov T, Simmerling C, and Scharer OD (2010) Synthesis and molecular modeling of a nitrogen mustard DNA interstrand crosslink, Chemistry 16, 12100–12103. [PubMed: 20842675]

[93]. Mukherjee S, Guainazzi A, and Scharer OD (2014) Synthesis of structurally diverse major groove DNA interstrand crosslinks using three different aldehyde precursors, Nucleic Acids Res. 42, 7429–7435. [PubMed: 24782532]

[94]. Liu M, Doublie S, and Wallace SS (2013) NEIL3, the final frontier for the DNA glycosylases that recognize oxidative damage, Mutat. Res 743-744, 4–11. [PubMed: 23274422]

[95]. Liu M, Bandaru V, Bond JP, Jaruga P, Zhao X, Christov PP, Burrows CJ, Rizzo CJ, Dizdaroglu M, and Wallace SS (2010) The mouse ortholog of NEIL3 is a functional DNA glycosylase in vitro and in vivo, Proc. Natl. Acad. Sci. USA 107, 4925–4930. [PubMed: 20185759]

[97]. Albelazi MS, Martin PR, Mohammed S, Mutti L, Parsons JL, and Elder RH (2019) The biochemical role of the human NEIL1 and NEIL3 DNA glycosylases on model DNA replication forks, Genes (Basel) 10, 315. doi: 10.3390/genes10040315.

[98]. Mohni KN, Wessel SR, Zhao R, Wojciechowski AC, Luzwick JW, Layden H, Eichman BF, Thompson PS, Mehta KPM, and Cortez D (2019) HMCES maintains genome integrity by shielding abasic sites in single-strand DNA, Cell 176, 144–153.e113. [PubMed: 30554877]

[99]. Thompson PS, Amidon KM, Mohni KN, Cortez D, and Eichman BF (2019) Protection of abasic sites during DNA replication by a stable thiazolidine protein-DNA cross-link, Nat. Struct. Biol 26, 613–618.

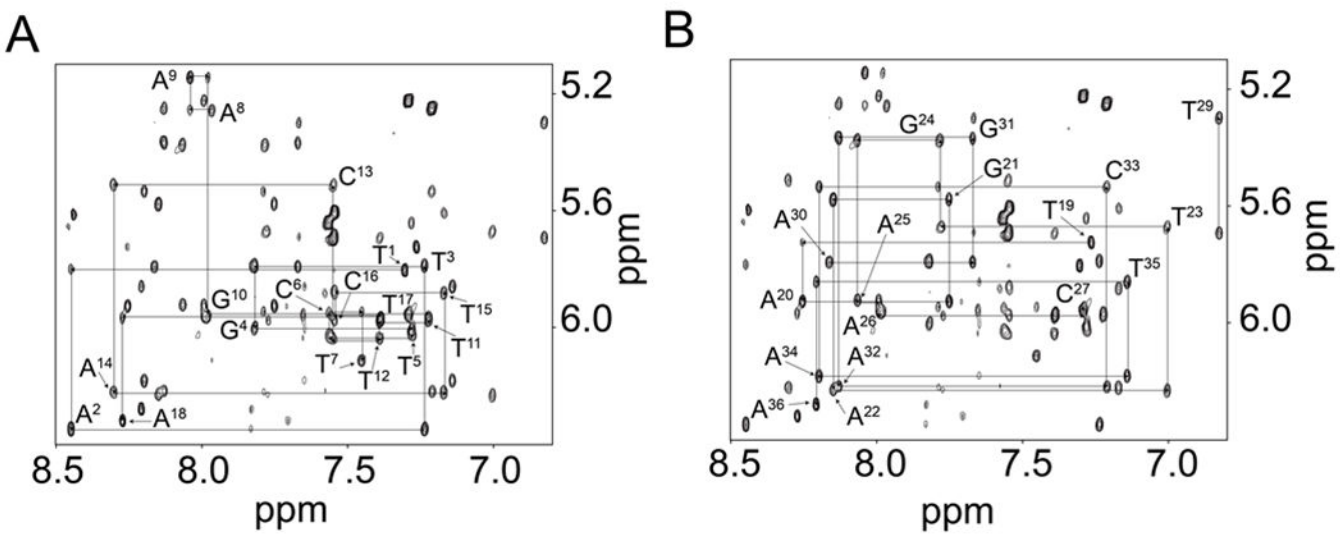


Figure 1.

Expanded plot of the NOESY spectrum of the dA-AP modified duplex, showing the sequential NOEs between aromatic H6/H8 protons and deoxyribosyl H1′ protons. (A) The strand containing the crosslinked adenine, showing bases T^{1} to A^{18}. The NOE connectivity is broken at the T^{7} H1′ → T^{7} H6 cross-peak and reinitiates at the A^{8} H1′ →A^{8} H8 cross-peak. (B) The strand containing AP site, showing bases T^{19} to A^{36}. The NOE connectivity is broken at the C^{27} H1′→ C^{27} H6 cross-peak due the absence of an aromatic proton on the AP site. The spectrum was acquired at 900 MHz and 298 K.

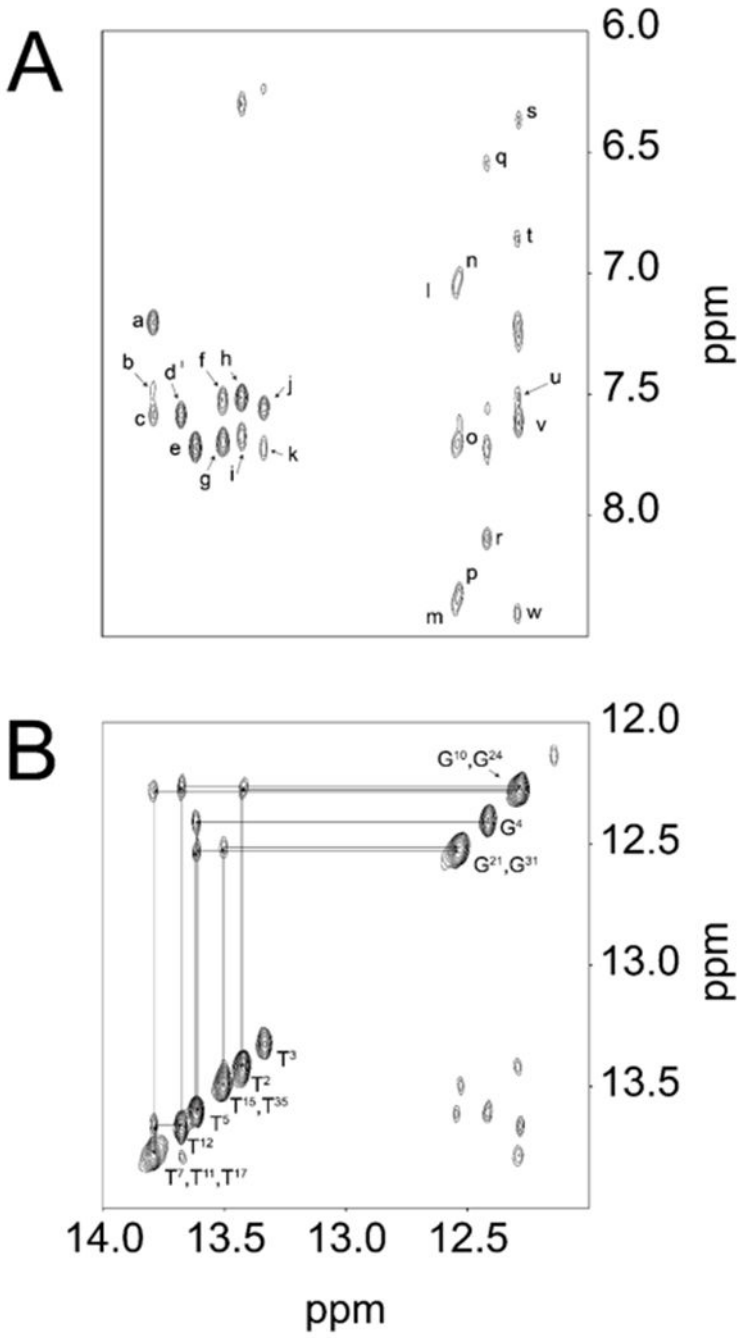


Figure 2.

A. Expanded plot showing sequential NOE connectivities for imino to amino protons. The cross peaks are assigned as: a, T^{11} N3H → A^{26} H2; b, T^{7} N3H → A^{30} H2; c, T^{11} N3H → A^{25} H2; d, T^{5} N3H → A^{32} H2; e, T^{5} N3H → A^{32} H2; f, T^{17} N3H → A^{20} H2; g, T^{15} N3H → A^{22} H2; h, T^{23} N3H → A^{14} H2; i, T^{35} N3H → A^{2} H2; j, T^{3} N3H → A^{34} H2; k, T^{23} N3H → A^{20} H2; l, G^{31} N1H → C^{6} $\mathcal{N}^{4}$ H_a; m, G^{31} N1H → C^{6} $\mathcal{N}^{4}$ H_b; n, G^{21} N1H → C^{16} $\mathcal{N}^{4}$ H_a; o, G^{31} N1H → A^{22} H2; p, G^{21} N1H → C^{16} $\mathcal{N}^{4}$ H_b; q, G^{4} N1H → C^{33} $\mathcal{N}^{4}$ H_a; r, G^{4} N1H → C^{33} $\mathcal{N}^{4}$ H_a, s, G^{10} N1H → C^{27} $\mathcal{N}^{4}$ H_a; t, G^{24} N1H → C^{13} $\mathcal{N}^{4}$ H_a; u, G^{24} N1H →

A^{14} H2; v, G^{10} N1H → C^{27} N^4 H_b; w, G^{24} N1H → C^{13} N^4 H_b. **B.** Expanded plot showing sequential NOE connectivities for the imino protons. The data were collected at 900 MHz at 250 ms mixing time, at 278 K.

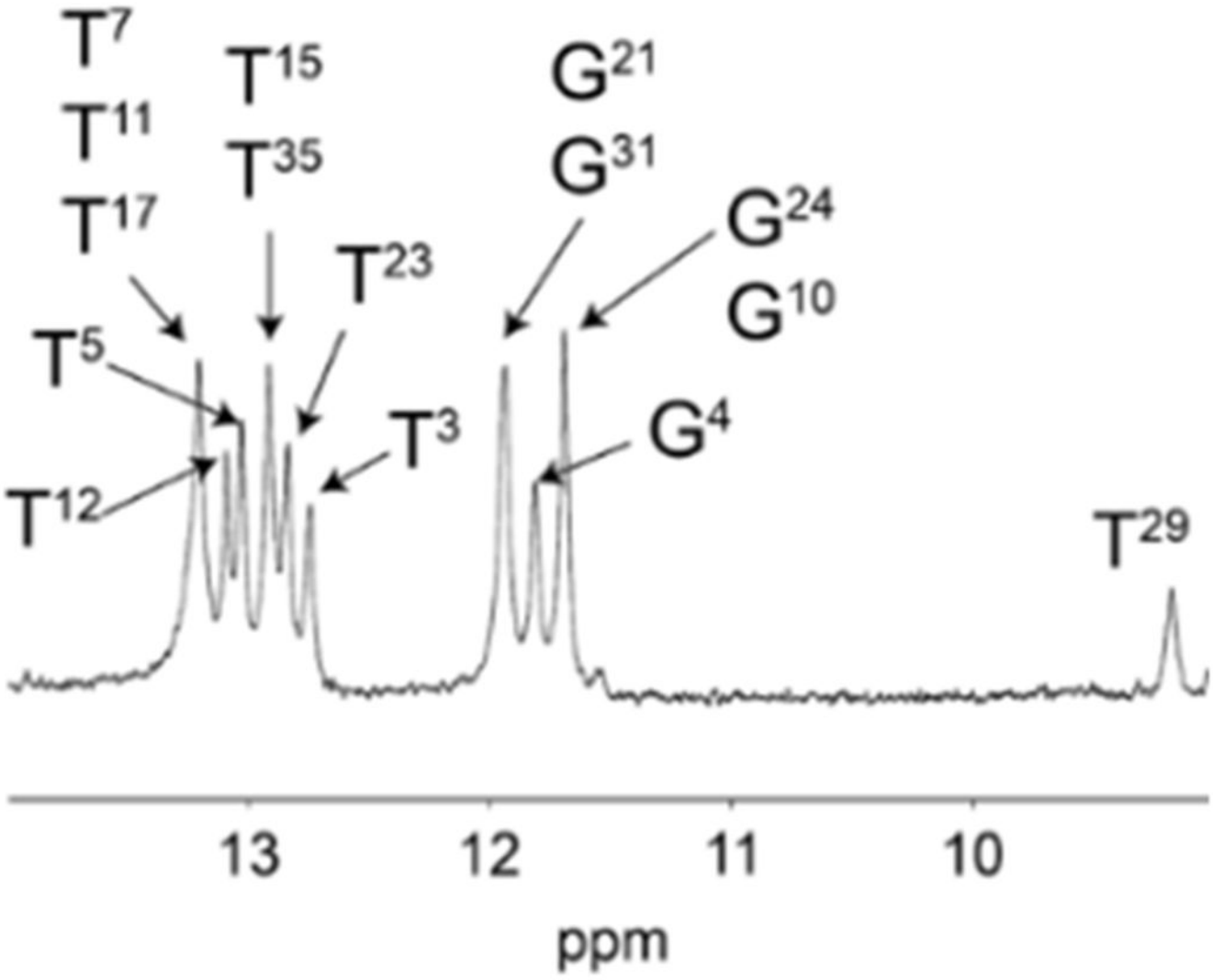


Figure 3.

^{1}H spectrum showing the thymine N3H and guanine N1H imino protons of the dA-AP interstrand crosslink. The 800 MHz spectrum was collected at 274 K in 95:5 H_2O:D_2O.

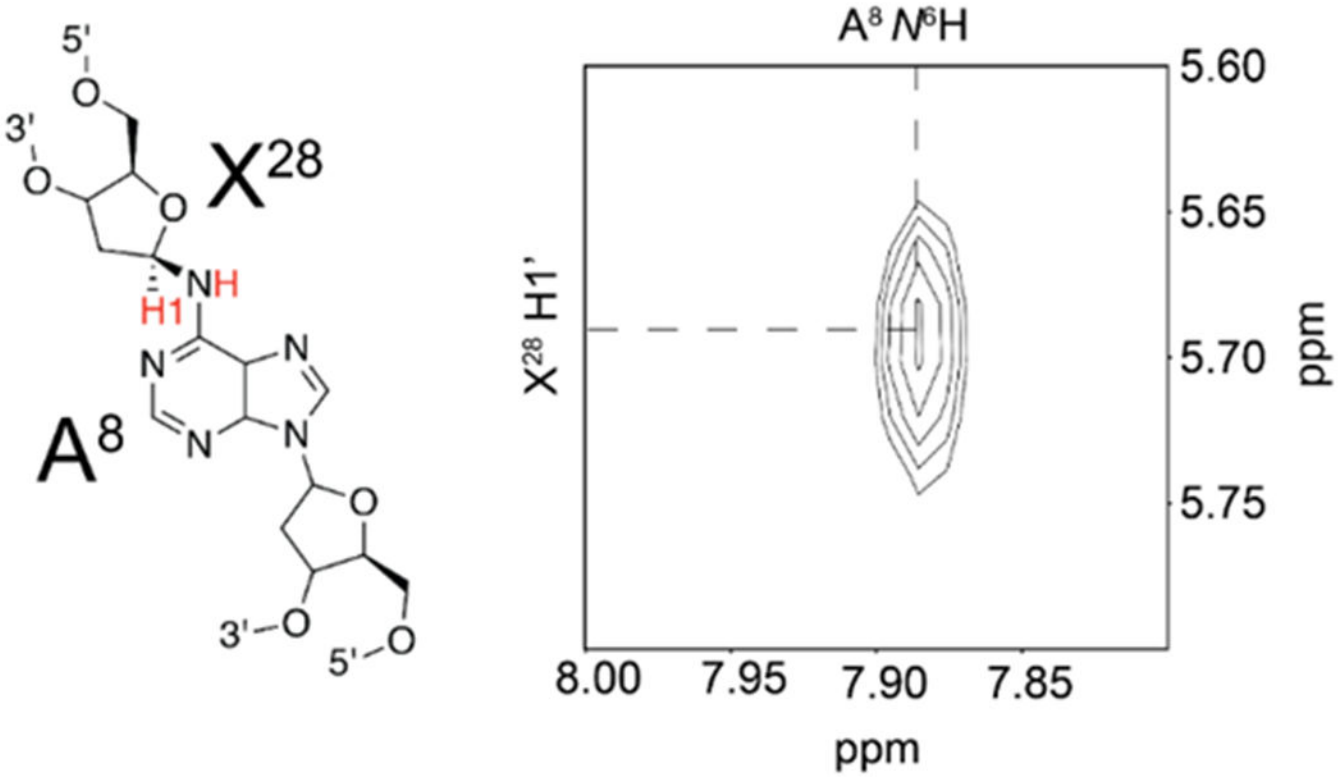


Figure 4.

Expanded plot of the TOCSY spectrum in H_2O showing the correlation between the A^8 N^6H and X^{28} H1 protons. The zpectrum was acquired at 900 MHz at 298 K

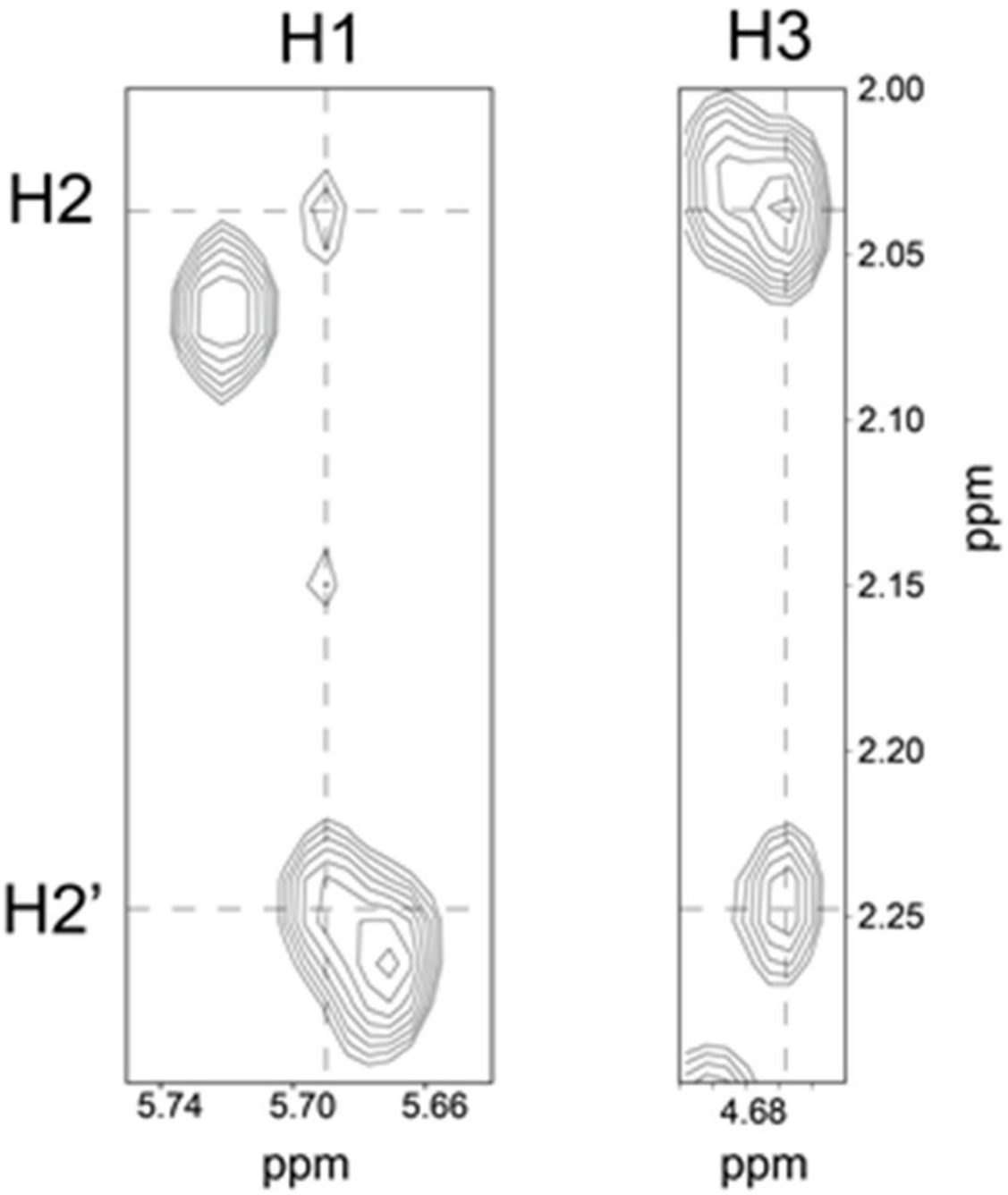


Figure 5.

Comparison of the NOE intensities for the cross-peaks arising between the X^{28} H1 and H3 deoxyribose protons to the X^{28} H2 and H2′ deoxyribose protons. The 900 MHz NMR spectra were collected at a mixing time of 60 ms at 298 K.

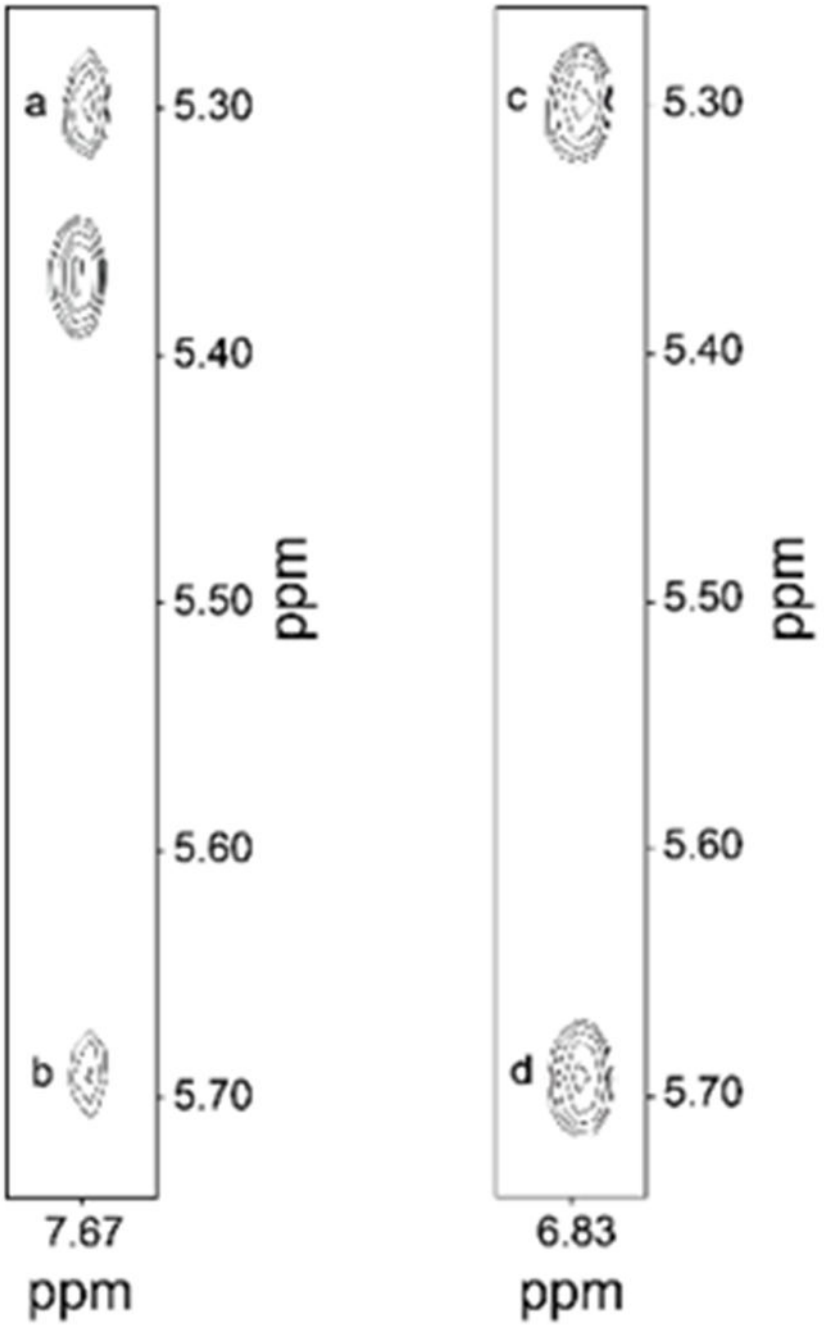


Figure 6.

Expanded plots of a NOESY spectrum of the duplex containing dA-AP ICL showing NOEs between the ICL protons and base and deoxyribose protons. The spectrum was acquired at 900 MHz and 298 K with a 250 ms mixing time. The cross-peaks are assigned as follows: a, A^{8} H2 → X^{28} H1; b, A^{8} H2 → T^{29} H1′, c, T^{29} H1′ → T^{29} H6; and d, X^{28} H1′ → T^{29} H6.

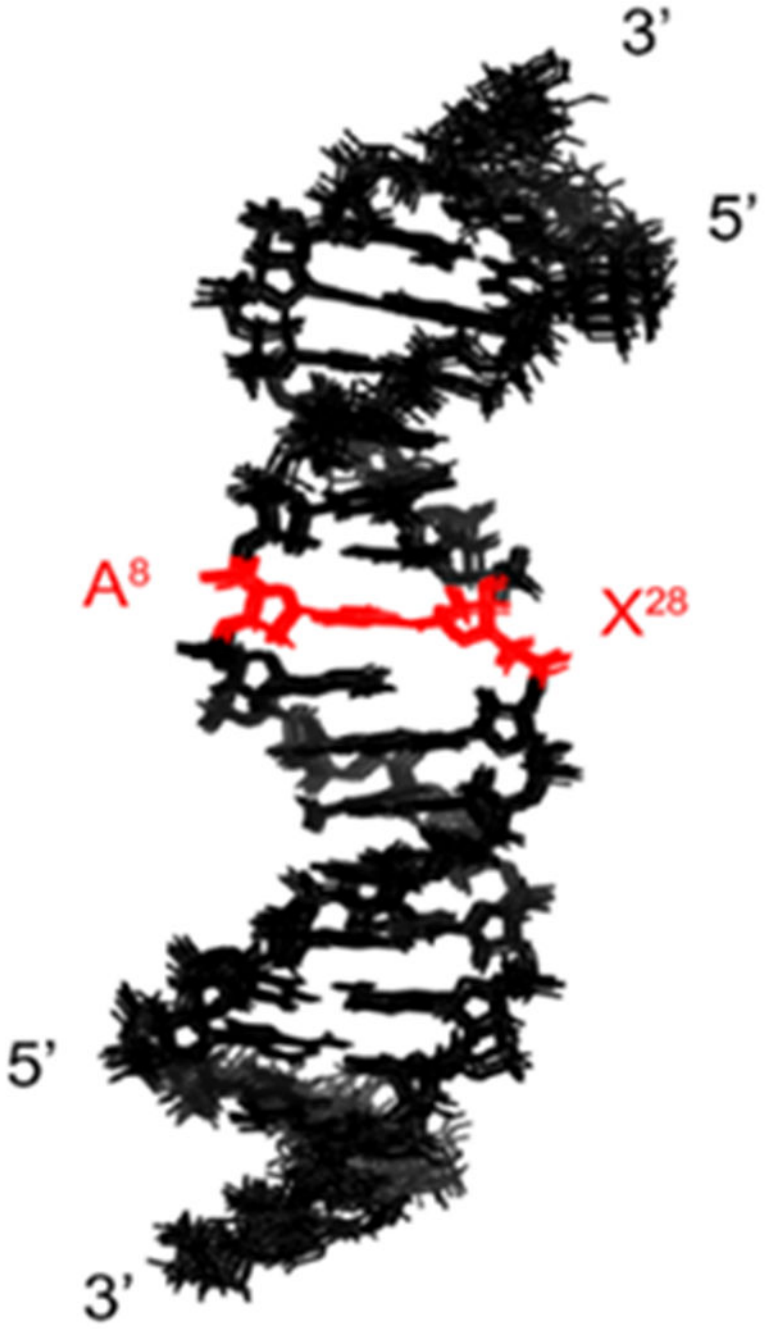


Figure 7.

Overview of the 10 lowest energy violation structures resulting from rMD calculations of duplex containing an dA-AP ICL carried out using a simulated annealing protocol and using NOE generated distance restraints.

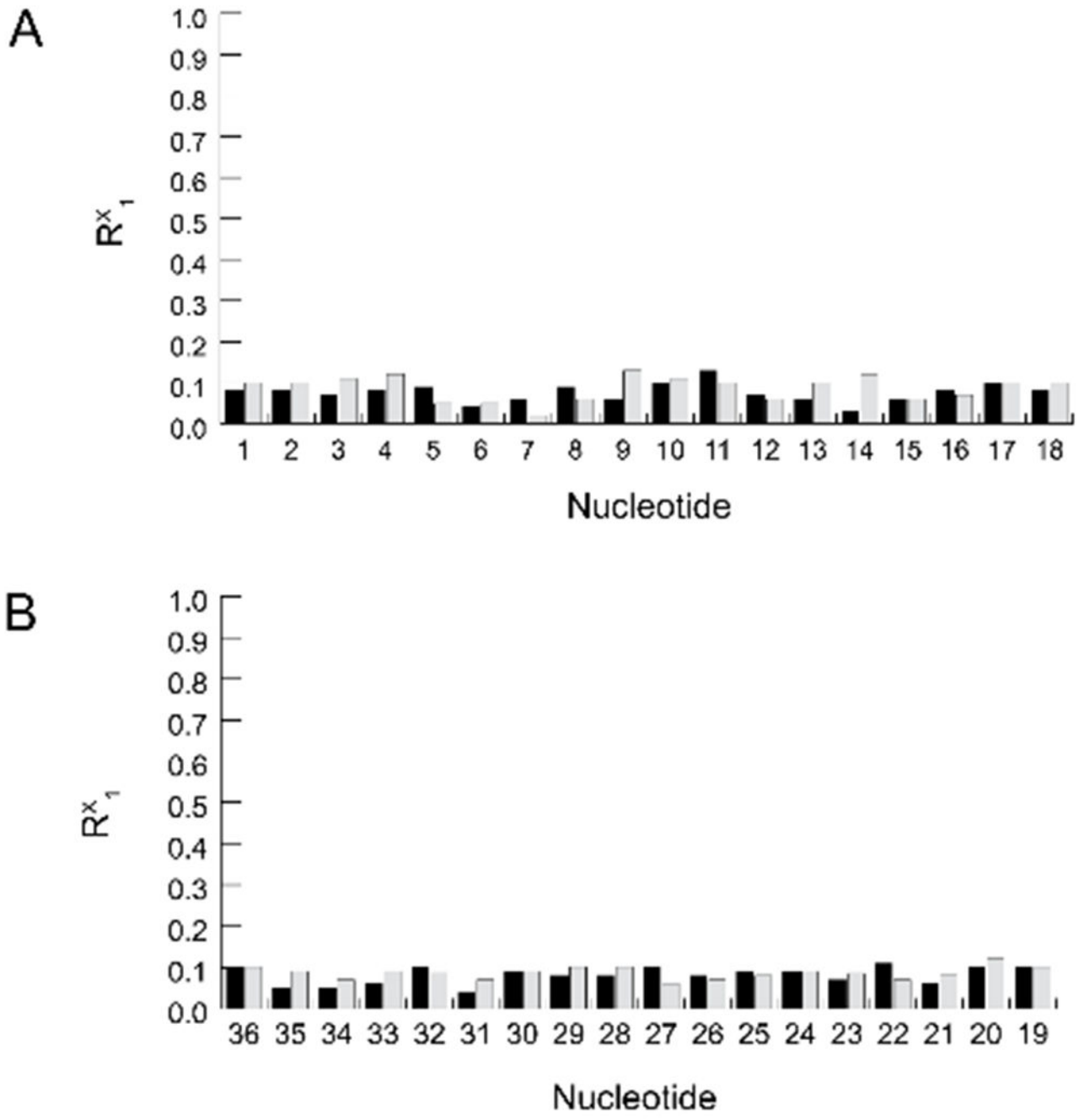


Figure 8.

Calculation of sixth root residual values (R^X_1) between theoretical NOEs predicted by complete relaxation matrix calculations and experimental NOEs for the averaged refined structure of the dA-AP ICL emergent from the rMD calculations, using CORMA. (A) The intrastrand and interstrand nucleotide R^X_1 values for nucleotides in the adenine-containing strand of the dA-AP ICL. (B) The intrastrand and interstrand nucleotide R^X_1 values for nucleotides in the AP-containing strand of the dA-AP ICL. The intranucleotide residuals are shaded in black. The DNA internucleotide residuals are shaded in gray.

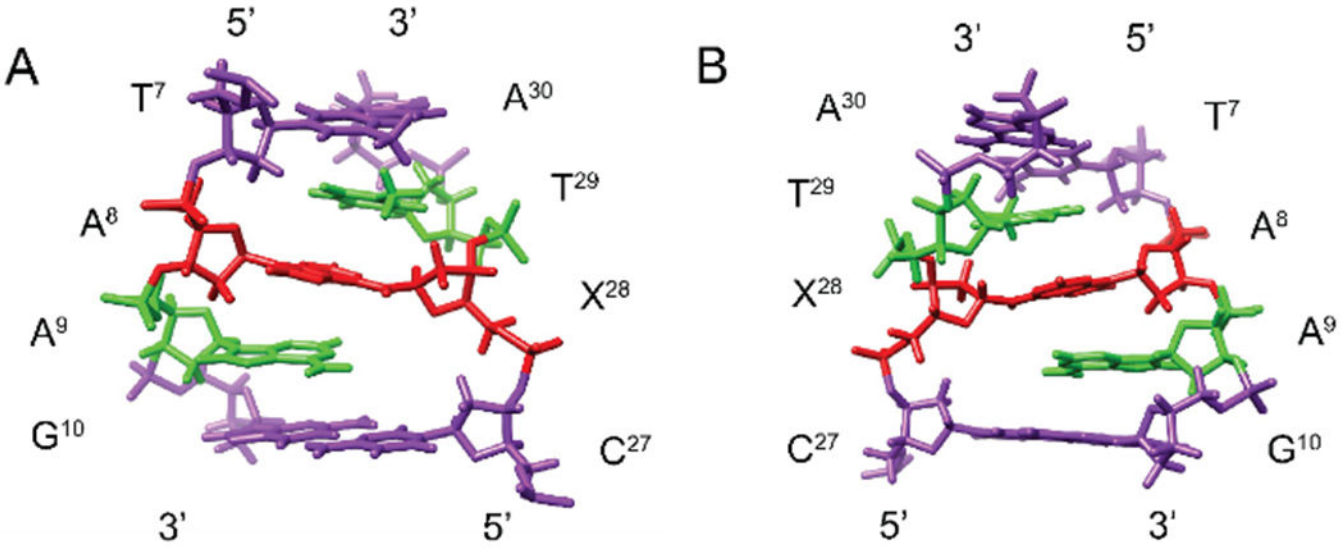


Figure 9.

Conformation of the dA-AP ICL as seen from the lowest violation structure emergent from the rMD calculations. (A) Base pairs T^{7}:A^{30}, G^{10}:C^{27}, the dA-AP ICL containing A^{8} and AP^{28}, and unpaired bases A^{9} and T^{29} as seen from the major groove. (B) Base pairs T^{7}:A^{30}, G^{10}:C^{27}, the dA-AP ICL containing A^{8} and AP^{28}, and unpaired bases A^{9} and T^{29} as seen from the minor groove. The base pairs T^{7}:A^{30} and G^{10}:C^{27} are colored purple, the unpaired bases A^{9} and T^{29} are colored green, and the dA-AP ICL containing A^{8} and AP^{28} is colored red.

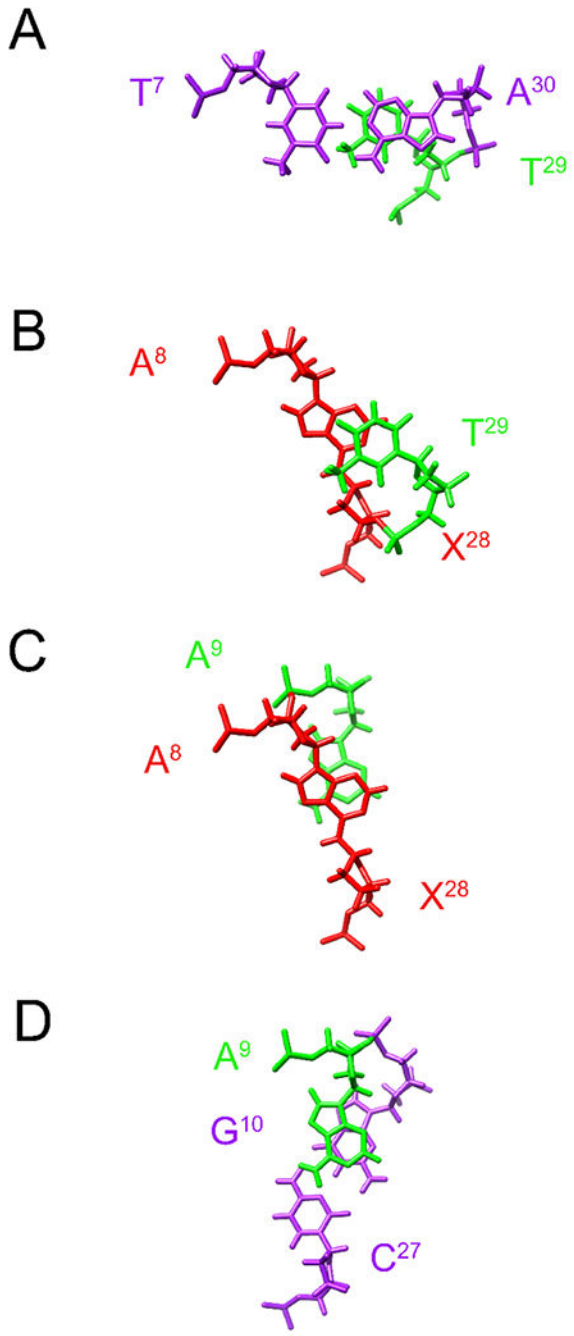


Figure 10.

Base stacking interactions of the dA-AP ICL as seen in the lowest violation structure in the rMD calculations. (A) View showing the T^{7}:A^{30} base pair in purple and the unpaired T^{29} in green. (B) View showing the unpaired T29 in green and the dA-AP ICL resulting from conjugation of A^{8} and X^{28} in red. (C) View showing the dA-AP ICL resulting from conjugation of A^{8} and X^{28} in red and the unpaired A^{9} in green. (D) View showing the unpaired A^{9} in green and the G^{10}:C^{27} base pair in purple.

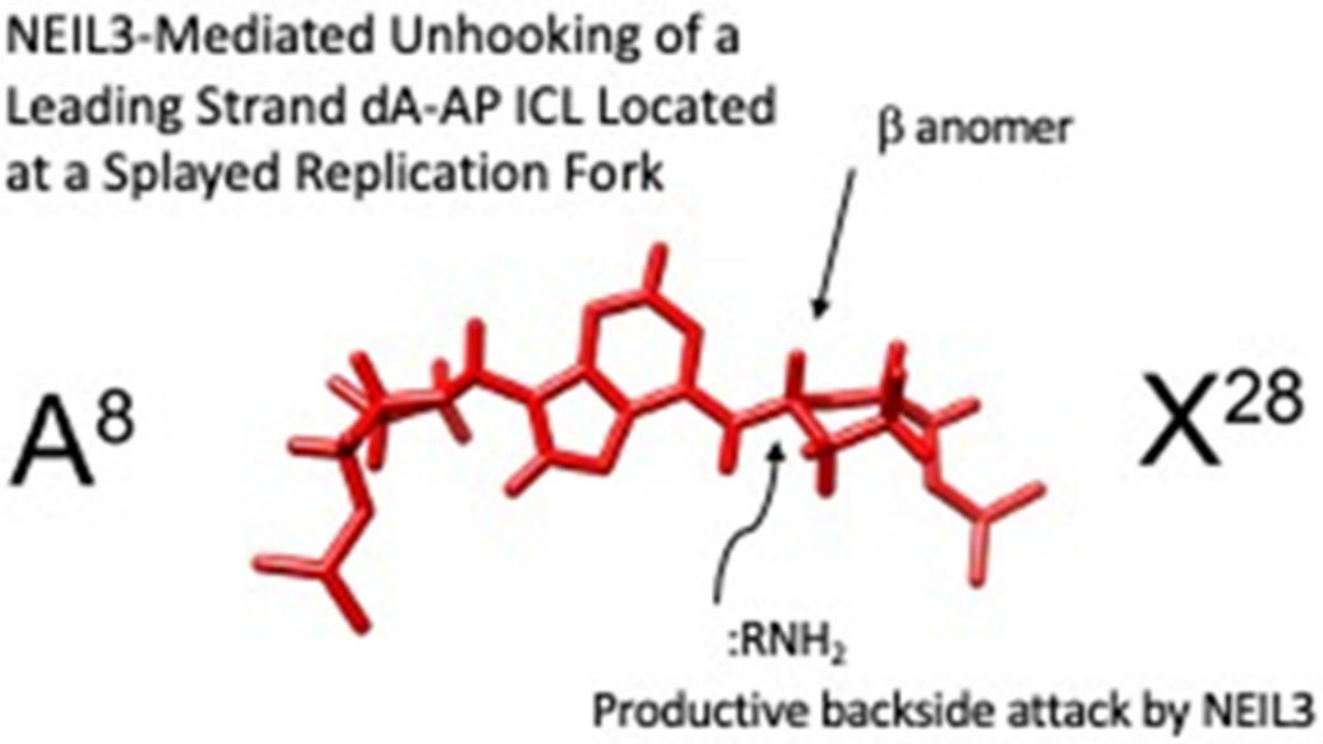
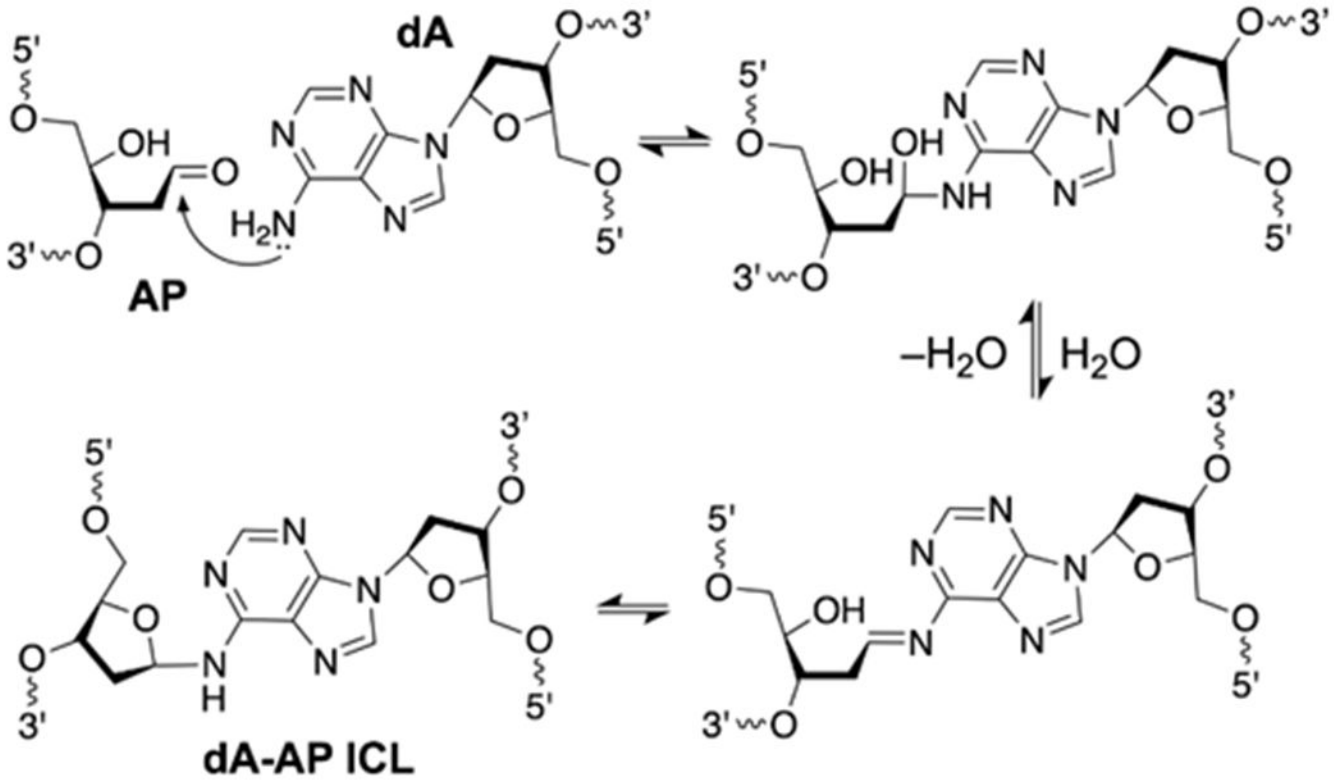


Figure 11.

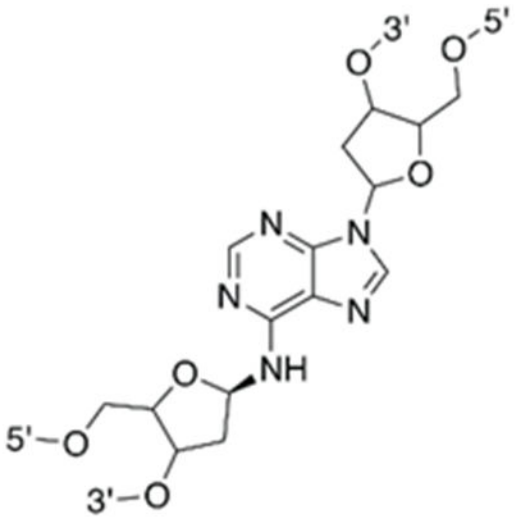
The β anomer at the dA-AP ICL may facilitate a stereo-electronically favorable backside attack by the NEIL3 active site amine when the dA-AP ICL is situated in the leading strand at a splayed replication fork.



Scheme 1.

Formation of the AP-dA ICL, following incubation of a DNA duplex containing a site-specific uracil nucleotide with the enzyme uracil deglycosylase.

A



B

5′-d(T[1] A[2] T[3] G[4] T[5] C[6] T[7] A[8] A[9] G[10]T[11]T[12]C[13]A[14]T[15]C[16]T[17]A[18])-3′
3′-d(A[36]T[35]A[34]C[33]A[32]G[31]A[30]T[29]X[28]C[27]A[26]A[25]G[24]T[23]A[22]G[21]A[20]T[19])-5′

Chart 1.

A. Structure of the dA-AP ICL. B. Oligodeoxynucleotide Sequence Used in This Work, Showing Numbering of Individual Nucleotides

Table 1.

NMR Restraints Used for the dA-AP ICL Structural Refinement.

Internucleotide Restraints	152
Intranucleotide Restraints	243
dA-AP ICL Restraints	29
Antidistance Restraints	2
Total NOE Restraints	426
Backbone Torsion Angle Restraints	160
Hydrogen Bonding Restraints	53
Deoxyribose Pseudorotation Restraints	115
Total Number of Restraints	754
Number of Distance Restraint Violations	39
Number of Torsion Angle Restraint Violations	7
Total Distance Penalty/Maximum Penalty (kcal/mol)	17.623/0.443
Total Torsion Angle Penalty/Maximum Penalty (kcal/mol)	0.652/0.063
r.ms. Distances (Å)	
r.m.s. Angles (deg)	2.305
Distance Restraint Force Field	30 kcal/mol/Å^2
Torsion Angle Restraint Force Field	30 kcal/mol/deg^2

Table 2.

RMS Differences and Sixth Root Residual (R^X_1) Values for the Average Structure of the dA-AP ICL Emergent From rMD Calculations.

Greatest RMS pairwise difference between 10 structures	0.864 Å
Greatest RMS difference from average structure	0.579 Å
Sixth root residual R^X_1 [a] calculated for intranucleotide distances in the average structure, using CORMA[b]	0.0801
Sixth root residual R^X_1 calculated for internucleotide distances in the average structure, using CORMA[a,b]	0.0961
Sixth root residual R^X_1 calculated for all distances in the average structure, using CORMA[a,b]	0.0878
average error[c]	0.00326

[a] R^X_1 is the sixth root *R* factor $\sum[((I_o)_i^{1/6}) - ((I_C)_i^{\frac{1}{6}}) \ / \ \sum((I_o)_i^{1/6})]$.

[b] Mixing time was 250 ms.

[c] Average error: $\sum(I_c - I_n)/n$, where I_C are NOE intensities calculated from the refined structure, and I_n are experimental NOE intensities.