Chemical and Structural Properties of DNA-Abasic Site Cross-links

A Dissertation

presented to

the Faculty of the Graduate School

University of Missouri

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

by

Nathan Eric Price

Dr Kent S. Gates, Dissertation Supervisor

December 2014

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

Chemical and Structural Properties of DNA-Abasic Site Cross-links

presented by Nathan Price,

a candidate for the degree of doctor of philosophy of Chemistry,

and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Kent Gates _____

Professor Susan Lever _____

Professor John Tanner _____

Professor Frank Schmidt _____

ACKNOWLEDGEMENTS

I would first like to thank Dr. Kent Gates for his continued guidance and support throughout my graduate degree. He has been a wonderful mentor and as a result these past few years have been an enjoyable experience resulting in a large amount of growth for me as a scientist and person.

I would like to thank my fellow chemistry graduate students from the Gates Lab. They have offered me support and advice in order to further my growth as a scientist. I look forward our continued support of one another as we progress as colleagues in our careers.

Finally I would like to thank my friends and family for all their love and support throughout my life. I am particularly grateful to my parents for their constant support, patience and understanding. I would not be where I am today without the love and guidance they have provided me.

Table of Contents

Chapter 2. Chemical and structural characterization of interstrand cross-links formed between abasic sites and adenine residues in duplex DNA

Chapter 3. Sequence effects and scope of DNA interstrand cross-links formed between abasic sites and an opposing adenine residue

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

LIST OF SCHEMES

Chapter 1

Chapter 2

Chapter 3

LIST OF TABLES

Chapter 2

ABSTRACT

The loss of a coding nucleobase from the structure of DNA is a common event that generates lesion known as an abasic (Ap) site. Ap sites exist as an equilibrating mixture of a cyclic hemiacetal and a ring-opened aldehyde. Aldehydes are electrophilic functional groups that can form covalent adducts with nucleophilic sites in DNA. Thus, Ap sites present a potentially reactive aldehyde as part of the internal structure of DNA. Here we report evidence that the aldehyde group of Ap sites in duplex DNA can form a covalent adduct with the N^6-amino group of adenine residues on the opposing strand. The resulting interstrand DNA-DNA cross-link occurs at 5'-ApT/5'-AA sequences in remarkably high yields (15-70%) under physiologically relevant conditions.

In order to rigorously characterize the structure of dA-Ap cross-links by enzymatic digestion of cross-linked DNA a putative dinucleoside cross-link remnant was synthesized by reaction of dA with 2-deoxyribose. Complete spectroscopic characterization of this material established that the anomeric carbon of the 2-deoxyribose adduct was connected to dA at the exocyclic N^6-amino group and exists as a mixture of cyclic alkoxyhemiaminals (α and β anomers of the furanose and pyranose isomers). LC-MS/MS analysis provided evidence that the synthetic compound 6 has the same chemical structure as the cross-link remnant released by enzymatic digestion of cross-linked DNA. Further experiments showed that the cross-link remnant was quite stable at neutral pH, decomposing to release unmodified dA with a half-life of 65 d at 37 °C. The inherent chemical stability of the dA-dR cross-link remnant is reflected in the stability of the cross-link in duplex DNA. For example, there is no detectable dissociation of the dA-Ap cross-link in duplex DNA over the course of 96 h at pH 7 and

22 °C. The stability of the dA-Ap cross-link offers the possibility that this may be a persistent lesion with the potential to block the action of various DNA processing enzymes. The results lay a foundation for the detection of the dA-Ap cross-link in cellular DNA.

Additionally, this work demonstrates that cross-links arising from the reaction of an Ap-site with an opposing nucleobase occur in a multitude of sequence context. Many of these cross-links can form in very high yield. Cross-link yields are affected by the local sequence composition, which most likely alters the structure of the Ap-containing duplex and in turn the amount of additional distortion required to form an interstrand cross-link. The importance of local structure and flexibility seems to be highlighted further by mispairing the cross-linking base, which improves cross-link yields and indicates a "loose" structure near the cross-linking site may promote cross-link formation. Lastly, the addition of a minor groove binder, netropsin, prevents cross-link formation only when a suitable binding site is near the cross-linking site, indicating that rigidifying the structure near the cross-link removed the necessary flexibility for the exocyclic amino group to reach across and react with an opposing Ap-site. Understanding the structure and reactivity of Ap-derived cross-links may be very important as this naturally-occurring DNA-templated reaction has the potential to generate cross-links in the genetic material of living cells.

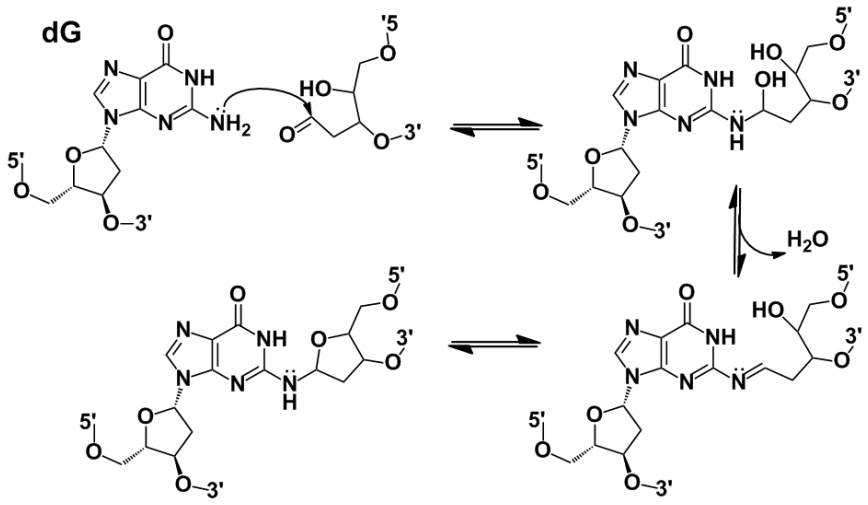
Chapter 1. Interstrand DNA-DNA Cross-link Formation Between Adenine Residues and Abasic Sites in Duplex DNA

1.1 Introduction

Interstrand cross-links are among the most deleterious types of damage that can be generated in cellular DNA.[1,2] These lesions are complete blocks to DNA transcription and replication because the covalent linkage prevents strand separation.[3,4] Concurrent modification of both strands in the duplex also presents severe challenges to cellular DNA repair systems.[5-7] Accordingly, even small numbers of interstrand DNA-DNA cross-links can have profound biological consequences.[1,2,8-10]

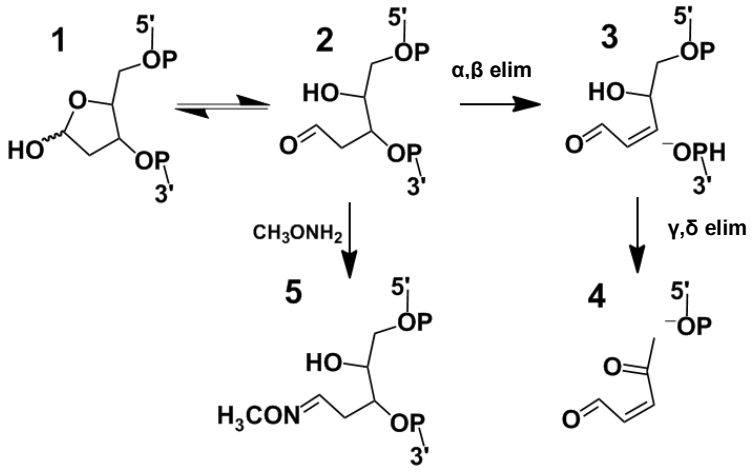
We recently reported a new structural type of interstrand cross-link derived from the reaction of a DNA abasic (Ap) site with a guanine residue on the opposing strand of the duplex (Scheme 1).[11,12] This cross-link is intriguing because Ap sites (**1**, Scheme 2) are perhaps the most common type of endogenous damage suffered by cellular DNA.[13,14] Ap sites are generated by spontaneous depurination, nitrosative stress, and enzymatic

Scheme 1



base excision repair processes.[14-18] As a result, DNA from normal mammalian tissue carries steady-state levels of 50,000-200,000 Ap sites per cell.[16,19] Ap sites are also generated by exposure of DNA to a wide variety of mutagens, toxins, and anticancer drugs.[14,18] Cross-links derived from this abundant lesion could have significance in biology and medicine.

Scheme 2



In our previous work, dG-Ap cross-links were observed at 5'-CAp/5'-AG sequences (duplex **A**, Figure 1) in equilibrium yields of 2-3%, under physiologically-relevant conditions.[11,12] The evidence supported a mechanism involving attack of the exocyclic N^2-amino group of guanine on the Ap aldehyde (Scheme 1).[11,12] Two other canonical nucleobases, adenine and cytosine, also contain exocyclic amino groups that have been shown to react with aldehydes.[20-23] However, our previous studies provided no indication regarding whether adenine or cytosine can participate in cross-link formation with an Ap site in duplex DNA.[11,12] In fact, no cross-linking was observed at the adenine residue directly opposing the Ap site in duplex **A** (Figure 1).[11,12] Cross-links involving adenine residues have been observed in DNA duplexes containing *oxidized* abasic sites;[24-27] however, it is important to recognize that the structures and reactivities of oxidized abasic sites are distinct from that of the native Ap site considered here.

In order to grasp the full potential of Ap sites to generate cross-links in duplex DNA, we felt it was necessary to examine whether canonical nucleobases other than guanine can forge cross-links with Ap sites in DNA. Here we provide evidence for a previously unknown type of interstrand DNA-DNA cross-link involving the reaction of adenine residues with Ap sites in double-helical DNA. The dA-Ap cross-links described here formed in remarkably high yields (15-70%) under physiologically-relevant conditions and we present evidence consistent with a cross-linking mechanism involving attack of the exocyclic N^6-amino group of adenine on the aldehyde group of the Ap site. The widespread occurrence of Ap sites in cellular DNA combined with the high yields of cross-links observed here, raise the possibility that this naturally-occurring, DNA-templated reaction[28-30] can generate cross-links in the genetic material of living cells.

Figure 1

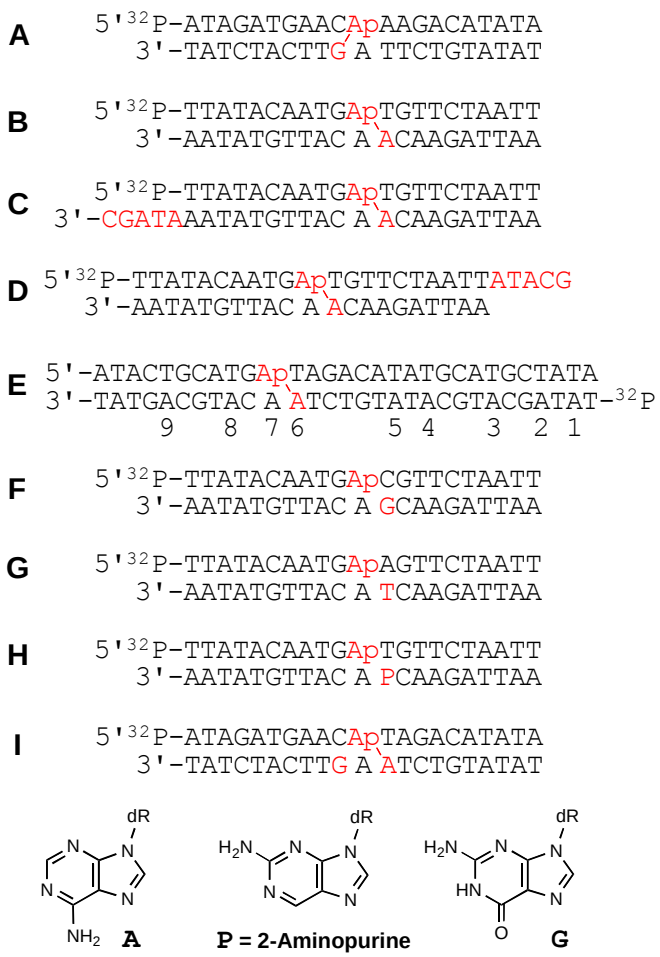


Figure 1. DNA duplexes used in this study. Ap-containing duplexes were generated from the corresponding dU-containing duplexes by treatment with uracil DNA glycosylase (UDG). Locations where cross-links form in substantial yield are indicated with a (\).

Results and Discussion

1.2 Generation and Stability of Interstrand Cross-links in DNA Duplexes Containing 5'-ApT/5'-AA Sequences

When considering nucleobase sequences in duplex DNA that might support formation of dA-Ap cross-links, we first recognized that any nucleobase involved in Ap-derived cross-link formation likely must be located near the Ap site. This is because cross-linking by a distal nucleobase presumably would generate substantial and energetically unfavorable structural distortion in the duplex.[31] With this in mind, we constructed models of Ap-containing DNA duplexes to determine sequence locations on the opposing strand that would position the N^6-amino group of adenine in close proximity to the Ap site. In these models, we employed the simplifying assumption that Ap-containing duplexes retain B-DNA-like character.[32] This approach successfully predicted cross-linking sequences in the case of dG-Ap cross-links.[11,12] Measurements within these models indicated that the exocyclic N^6-amino groups of both adenine residues opposing the Ap site in a 5'-ApT/5'-AA sequence were located near the Ap aldehyde residue (Figure 2). As noted in the introduction, however, our previous studies showed that adenine residues directly opposing an Ap site (e.g. duplex **A**) did not engage in cross-link formation.[11,12] Overall, this analysis led us to the prediction that an adenine residue offset one base to the 3'-side of the Ap site might have potential to engage the Ap-aldehyde in cross-link formation (Figure 2).

Figure 2

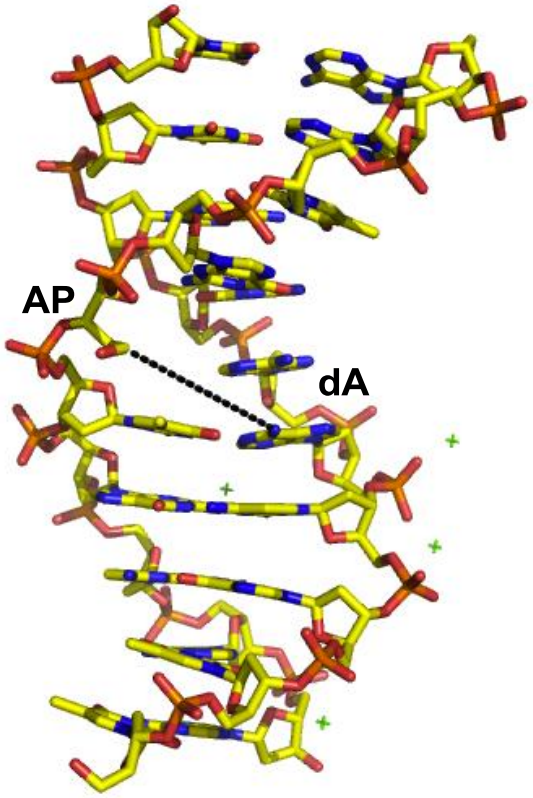


Figure 2. Molecular model illustrating the proximity of the abasic site aldehyde and N^6-amino group of deoxyadenosine in a 5'-ApT sequence in B-DNA. The image was constructed using Pymol and is based on pdb entry 3bse.

Ap sites were produced at defined locations in DNA duplexes by treatment of the corresponding 5'-^{32}P-labeled, 2'-deoxyuridine-containing duplexes with uracil DNA glycosylase (UDG).[33,34] Efficient formation of Ap sites in the DNA was confirmed by piperidine treatment to generate the cleavage product **4** (Scheme 2; Figure 3, lanes 3, 7, and 12).[14,18,35]

Figure 3

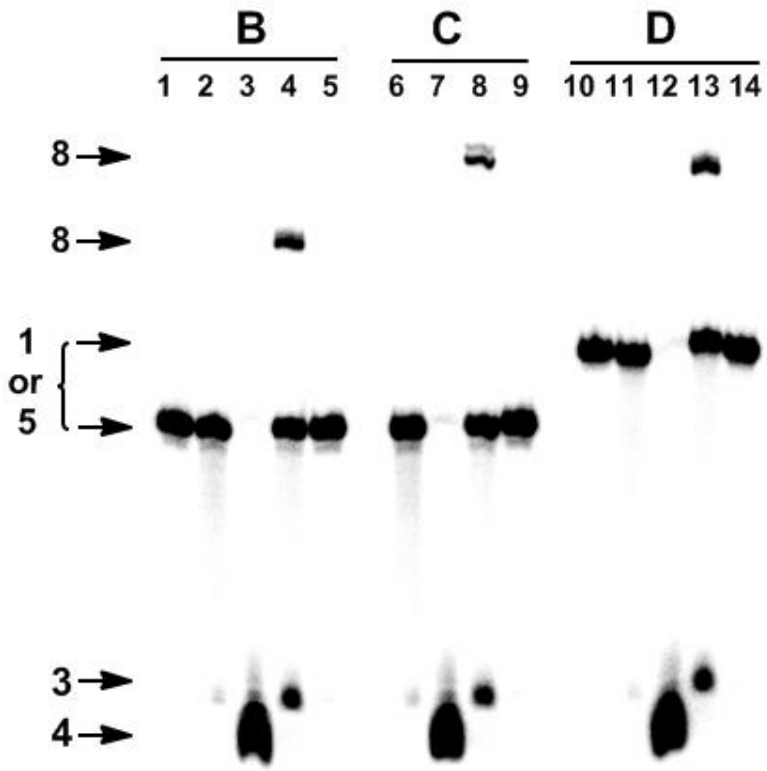


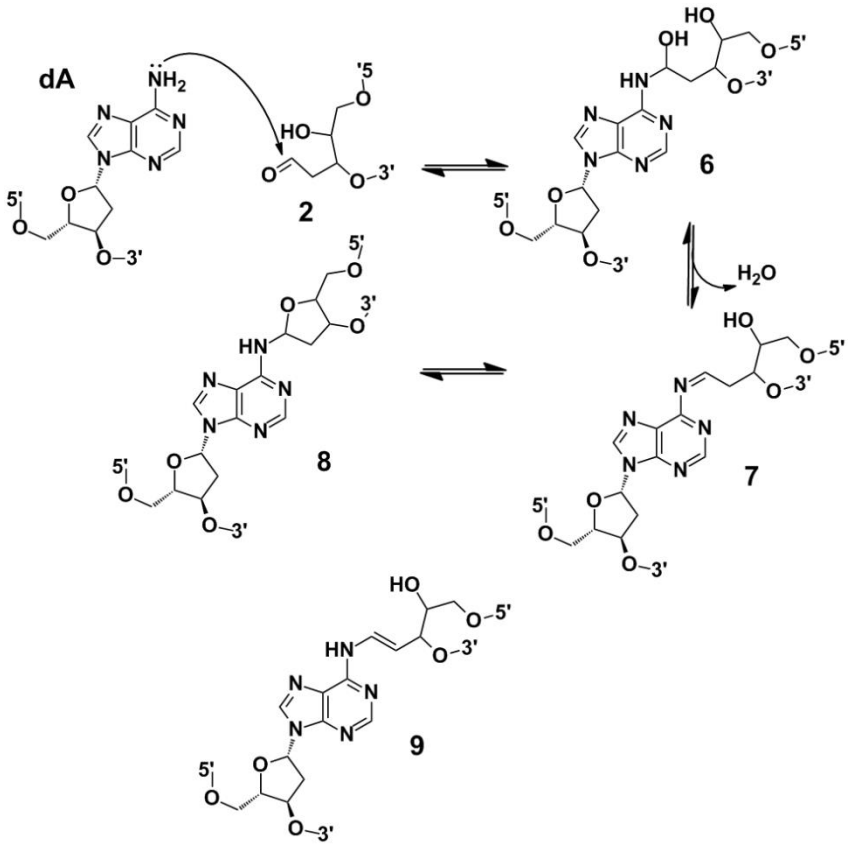
Figure 3. Interstrand cross-link formation in duplexes **A**-**C**. Lanes, 1-5, Duplex **B**; lanes 6-9, duplex **C**; lanes 10-14, duplex **D**. Lanes 1 and 10 contain the ^{32}P-labeled, uracil-containing 2'-deoxyoligonucleotides. Lanes 2, 6, and 11 contain the ^{32}P-labeled UDG-treated (abasic-site-containing) duplexes without incubation. Lanes 3, 7, and 12 contain the ^{32}P-labeled, UDG-treated (abasic-site-containing) duplexes subjected to piperidine work-up (1 M, 95 ˚C, 25 min) to cleave the Ap site. Lanes 4, 8, and 13 contain the cross-linking reactions involving incubation of the abasic-site-containing duplexes in HEPES buffer (50 mM, pH 7) containing NaCl (100 mM) at 37 ˚C. Lanes 5, 9, and 14 contain the methoxyamine capping reactions involving incubation of the abasic-site-containing duplexes in HEPES buffer (50 mM, pH 7) containing NaCl (100 mM), and CH_3ONH_2•HCl (2 mM) at 37 ˚C. The ^{32}P-labeled 2'-deoxyoligonucleotides were resolved by electrophoresis on a 20% denaturing polyacrylamide gel and the radioactivity in each band quantitatively measured by phosphorimager analysis.

Incubation of duplex **B** containing a central 5'-GApT/5'-AAC sequence in HEPES buffer (50 mM, pH 7, 100 mM NaCl) at 37 ˚C, followed by denaturing polyacrylamide gel electrophoretic analysis, revealed a 15.1 ± 0.5% yield of a slow-migrating band in the region of the gel where the cross-linked duplex was expected to appear (Figure 3, lane 4).[11,12] We also observed bands corresponding to the intact Ap-containing oligonucleotide **1** and small amounts of a product inferred to be the elimination product **3**, based upon its gel mobility.[14,18,36,37] Duplexes **C** and **D**, equipped with 3'-overhanging ends, generated slow-migrating bands whose gel mobility was retarded relative to the putative cross-link generated by duplex **B** (Figure 3, lanes 8 and 13). These results provided evidence that the slow-migrating bands were indeed cross-linked duplexes containing both full-length strands of the parent duplex and ruled out cross-link formation by the more strongly electrophilic 3'-4-hydroxy-2-pentenal-5-phosphate end group **3**. Inclusion of the aldehyde-trapping agent methoxyamine in the reactions substantially inhibited formation of the slow-migrating bands (Figure 3, lanes 5, 9, and 14), consistent with involvement of the Ap-aldehyde residue in cross-link formation (**2**, Schemes 2 and 3).[11,12,38,39] It is noteworthy that cross-link formation in duplexes **B**-**D** occurred in good yields (13-18%) at physiological pH and did *not* require conditions of reductive amination (e.g. pH 4-5 and $NaCNBH_3$) often used to stabilize the imine products resulting from the reaction of aldehydes and amines.[11,12,40-45] A time course experiment showed that easily detectable amounts of the cross-link were formed in duplex **B** within a few hours (Figure S1).

Cross-link formation was a robust reaction that was not highly sensitive to reaction conditions. For example, in duplex **B**, similar cross-link yields (16 ± 3%) were

observed in HEPES buffer (50 mM, pH 7; 100 mM NaCl), cacodylate (50 mM, pH 7, 100 mM NaCl), NaH_2PO_4/citric acid (16.5/1.8 mM, pH 7.0, 100 mM NaCl), MOPS (50 mM, pH 7, 100 mM NaCl), Bis-Tris (50 mM, pH 7.4, 100 mM NaCl), HEPES buffer (50 mM, pH 7; 100 mM NaCl, 2 mM Mg^{2+}) and in the presence of the biological thiol glutathione (1 mM, in HEPES 50 mM, pH 7, 100 mM NaCl) (Figure S2). Additionally a buffer exchange into (HEPES 50 mM, pH 7, 100 mM NaCl) was performed after treatment with UDG as an alternative to ethanol precipitation. There was no change in yield or appearance of the bands created using this alternative preparation of Ap-containing duplexes. However, it was discovered that improper removal of phenol/chloroform after an extraction of addition of phenol would prevent cross-link formation (Figure S2).

Scheme 3



The formation of low molecular weight imines and *N*-arylglycosides, analogous to those shown in Scheme 3, are reversible processes.[46,47] This made it interesting to examine the stability of the cross-link generated in duplex **B**. We found that, once formed, the cross-link in duplex **B** was stable to a variety of conditions including pH 3 (sodium acetate, 25 mM, 24 ˚C, 30 min), pH 10 (potassium phthalate, 24 ˚C, 25 mM, 30 min), or mild piperidine workup (0.1 M, 60 ˚C, 15 min) (Figure S3). On the other hand, the cross-link was decomposed by heat (90 ˚C, 15 min) or treatment with the aldehyde-trapping reagent methoxyamine (200 mM, 30 min, 60 ˚C; Figure S3).

1.3 Evidence That Cross-linking Involves the Adenine Residue Offset One Base to the 3'-Side of the Ap site in the 5'-ApT/5'-AA Sequence

To identify the site at which the Ap-aldehyde was attached to the opposing strand, we carried out hydroxyl radical footprinting of a cross-linked duplex containing the 5'-ApT/5'-AA sequence.[26,48] The longer duplex **E** (Figure 1) was used to move the DNA bands of interest away from a region of the gel where bands were obscured by a "salt flare". In this duplex, the strand opposing the Ap-containing 2'-deoxyoligonucleotide was 5'-^{32}P-labeled. The cross-link was generated as described above, isolated from the gel, and subjected to cleavage by a mixture of iron-EDTA-H_2O_2.[26,48] In this type of experiment, the site of cross-linking appears as an interruption in the "ladder" of strand cleavage products generated by the iron-EDTA-H_2O_2 reagent, because cleavages beyond the cross-link yield large, slow-migrating DNA fragments that are connected to the opposing strand.[26,48] In the present case, a clear disruption in the ladder of cleavage products was observed at the 5'-adenine residue in the 5'-GApT/5'-AAC sequence (residue A6 in duplex **E**, Figures 1 and 4). This provided evidence that

cross-linking involves attachment of the Ap-aldehyde to the adenine residue that is offset one base to the 3'-side of the Ap site.

Figure 4

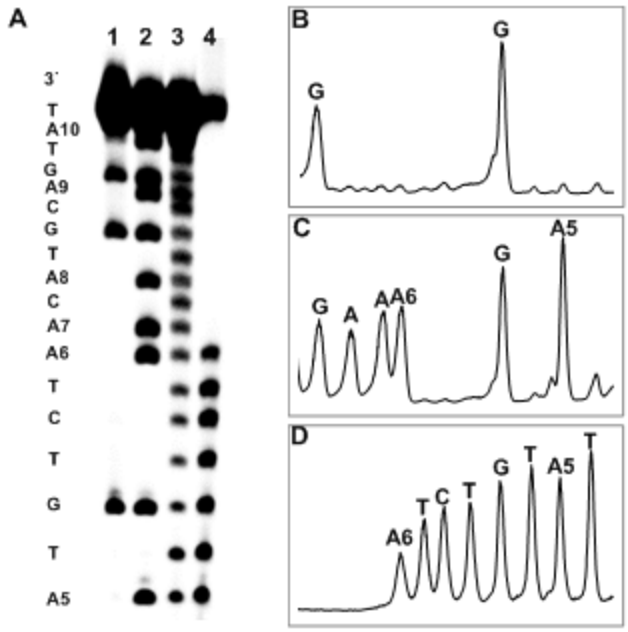


Figure 4. Hydroxyl radical footprinting of duplex **E** to locate the site of cross-link attachment. Duplex **E** contains the same core sequence as that in duplex **B**. **Panel A:** Lane 1 is a Maxam-Gilbert G-specific cleavage (sequencing) reaction of the labeled 2'-deoxyoligonucleotide strand in duplex **E**. Lane 2 is an A+G specific cleavage (sequencing) reaction of the labeled 2'-deoxyoligonucleotide strand in duplex **I**. Lane 3 is the hydroxyl radical footprinting reaction of the single stranded non-uracil containing oligonucleotide. Lane 4 is the hydroxyl radical footprinting reaction of the slow-migrating, cross-link band generated by incubation of duplex **E** HEPES buffer (50 mM, pH 7) containing NaCl (100 mM) at 37 ˚C. The ^{32}P-labeled 2'-deoxyoligonucleotides were resolved on a 20% denaturing sequencing gel and the radioactivity in each band quantitatively measured by phosphorimager analysis. **Panel B** is a Maxam-Gilbert G-specific cleavage (sequencing) reaction of the labeled 2'-deoxyoligonucleotide strand in duplex **E**. **Panel C** is an A+G specific cleavage (sequencing) reaction of the labeled 2'-deoxyoligonucleotide strand in duplex **E**. **Panel D** is the hydroxyl radical footprinting reaction of the slow-migrating, cross-link band.

We next examined the effect of substituting different nucleobases for the key adenine residue in the 5'-GApT/5'-AAC sequence. Substitution of this adenine residue in duplex **B** with a guanine or thymine residue, respectively, gave duplexes **F** and **G** that did not produce significant yields of a slow-migrating cross-link band (Figure 5). When placing the remaining canonical base, cytosine, in place of the cross-linking adenine residue also did not yield any cross-link (S15). Similarly, duplex **H** containing 2-aminopurine at this position did not efficiently generate the cross-link (Figure 5). The substitution of 2-aminopurine for the adenine residue amounts to the rather subtle relocation of a single exocyclic amino group from the major groove to the minor groove of the DNA duplex (see structures in Figure 1). Thus, the failure of duplex **H** to generate significant yields of cross-link allowed us to infer that the exocyclic N^6-amino group of dA in the 5'-GApT/5'-AAC sequence was intimately involved in the cross-linking reactions described here.

Figure 5

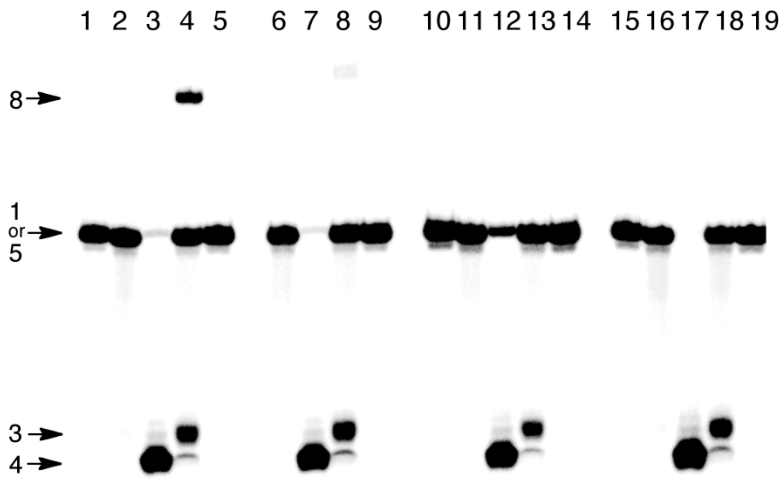


Figure 5. Nucleobase replacement experiments reveal that the underlined adenine residue in the 5'-G**Ap**T/5'-**A**AC sequence of duplex **B** is critical for cross-link formation. Replacement of this adenine residue in duplex **B** with a guanine or thymine residue in duplexes **E** and **F**, respectively, abrogates cross-link formation (lanes 13 and 18). Replacement of the crucial adenine residue with 2-aminopurine in duplex **G** yields only small amounts of a slower-moving band (lane 8). Lanes 1-5, duplex **B**, lanes 6-9, duplex **G**, lanes 10-14, duplex **E**, lanes 15-19, duplex **F**. Lanes 1, 10 and 15 contain ^{32}P-labeled, uracil-containing 2'-deoxyoligonucleotide duplexes. Lanes 2, 6, 11, and 16 contain the ^{32}P-labeled UDG-treated (abasic-site-containing) duplexes without incubation. Lanes 3, 7, 12, and 17 contain the abasic-site-containing duplexes subjected to piperidine work-up (1 M, 95 ˚C, 25 min) to cleave the Ap site. Lanes 4, 8, 13, and 18 contain the cross-linking reactions involving incubation of the abasic-site-containing duplexes in HEPES buffer (50 mM, pH 7) containing NaCl (100 mM) at 37 ˚C. Lanes 5, 9, 14, and 19 contain the methoxyamine capping reactions involving incubation of the abasic-site-containing duplexes incubated in HEPES buffer (50 mM, pH 7) containing NaCl (100 mM), and CH_3ONH_2•HCl (2 mM) at 37 ˚C. The ^{32}P-labeled 2'-deoxyoligonucleotides were resolved by electrophoresis on a 20% denaturing polyacrylamide gel and the radioactivity in each band quantitatively measured by phosphorimager analysis.

1.4 Mass Spectrometric Analysis of the Cross-linked Duplex B

The results described above were broadly consistent with a cross-linking mechanism involving reversible reaction of the exocyclic N^6-amino group on dA with the Ap aldehyde to generate hemiaminal, imine, cyclic hydroxyalkylhemiaminal, or enamine linkages at 5'-GApT/5'-AAC sequences in duplex DNA (structures **6**-**9**). MALDI-MS analysis of the cross-linked duplex **B** gave a strong $[M-H]^-$ signal at *m*/*z* 12718.66. This value is consistent with the imine **7**, cyclic hydroxylalkylhemiaminal **8**, or enamine **9** (Figure S4; calcd $[M-H]^-$ = 12719.4, 20 ppm dev). Literature precedents describing the properties of structurally related small molecules indicate that the cyclic hydroxylalkylhemiaminal structure **8** likely is favored.[21,22,49,50]

LC-MS/MS analysis, performed in collaboration with the Wang laboratory at UCR, provided additional insight regarding the chemical structure of the dA-Ap cross-link. The cross-linked duplex **B** (without the 5'-^{32}P label) was excised from a 20% denaturing polyacrylamide gel, eluted from the gel slice, digested with a four-enzyme cocktail consisting of nuclease P1, alkaline phosphatase, and phosphodiesterases I and II, and subjected to LC-MS/MS analysis using previously reported experimental conditions.[51-53] Selected-ion chromatograms revealed two peaks eluting at 24.1 and 26.1 min displaying the relevant *m/z* 368→252 transition anticipated for the neutral loss of 2-deoxyribose from the putative dA-AP cross-link (Figure 6). Further cleavage of the *m/z* 252 ion gave rise to a dominant fragment ion at *m/z* 136 in MS/MS/MS (Figure S5), which is identical to the MS/MS observed for the $[M+H]^+$ ion of dA, suggesting that the initial *m/z* 368→252 transition arises via loss of the cross-linked 2-deoxyribose adduct from dA (Figure 5C). The presence of multiple peaks in the LC-MS/MS ion

chromatogram was consistent with the expectation that the anticipated cross-link remnant resulting from the digestion of **8** will exist as an equilibrating mixture of pyranose and furanose isomers, each with α/β anomers, however the pyranose isomers will likely predominate (Scheme S1).[18,54]

Figure 6

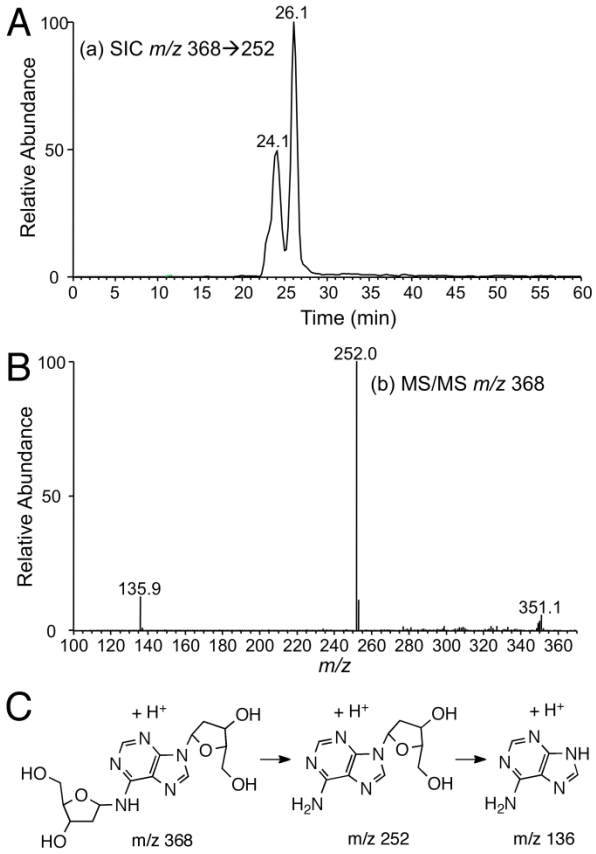


Figure 6. LC-ESI-MS and MS/MS analysis of the mixture generated by digestion of the cross-linked 2'-oligonucleotide generated by incubation of duplex **B** in HEPES buffer (50 mM, pH 7) containing NaCl (100 mM) at 37 ˚C. Panel A shows the selected-ion chromatogram (SIC) monitoring the neutral loss of 2-deoxyribose from the dA-AP cross-link. Panel B shows the MS/MS arising from the fragmentation of the ion at *m/z* 368. Panel C depicts possible structures for the ions observed at *m/z* 368, 252, and 136 in these experiments.

1.5 High Yield dA-Ap Cross-link Formation in the 5'-CApT/5'-AAG Sequence

We examined cross-link formation in duplex **I** that contained a central 5'CApT/5'-AAG sequence. This experiment was intriguing to us because, in this sequence context, the Ap site has the potential to nominally generate cross-links with *either* the guanine or adenine residue on the opposing strand (residues colored red in duplex **I** shown in Figure 1). Incubation of duplex **I** in HEPES buffer (50 mM, pH 7, 100 mM NaCl) at 37 ˚C produced a remarkable 71.5 ± 3.7% yield of a slow-migrating band in the region of the denaturing gel where the cross-linked duplexes migrate (Figure 7). Similar to the experiments described above in the context of the 5'-GApT/5'-AAC sequence, the yield of the slow-moving band in the 5'CApT/5'-AAG sequence also was greatly diminished when the aldehyde-capping reagent methoxyamine was added to the reaction mixtures. Slow-moving bands from duplexes containing 3'-overhanging ends were gel shifted, providing evidence that the slow-migrating bands were, in fact, interstrand cross-links containing both full-length strands of the duplex (Figures S6-S9). Nanospray ESI-MS further confirmed cross-link formation (Figure S10). Together the results confirmed that the slow-migrating band generated in the 5'CApT/5'-AAG sequence was an interstrand cross-linked duplex involving the Ap-aldehyde residue. Footprinting of a cross-linked 35-mer duplex containing the 5'CApT/5'-AAG core sequence indicated that the cross-link attachment was to the underlined adenine residue in the 5'CApT/5'-AAG sequence (Figure S9). The footprinting results showed *no* detectable cross-linking to the dG residue in this sequence. Thus, the remarkable cross-link yield in this sequence appears to result from efficient formation of a dA-Ap cross-link rather than simultaneous formation of dG-Ap and dA-Ap cross-links.

Figure 7

Figure 7. Interstrand cross-link formation in duplex **I**. Lane 1. 5'-^{32}P-labeled uracil-containing precursor of 2'-deoxyoligonucleotide duplex **I**. Lane 2. 5'-^{32}P-labeled abasic-site-containing 2'-deoxyoligonucleotide duplex **I** without incubation. Lane 3. 5'-^{32}P-labeled abasic-site-containing 2'-deoxyoligonucleotide duplex **I** cleaved by treatment with piperidine (1 M, 95 ˚C, 25 min). Lane 4. Duplex **I** incubated in HEPES buffer (50 mM, pH 7) containing NaCl (100 mM) at 37 ˚C. Lane 5. Duplex **I** incubated in with CH_3ONH_2 (2 mM) in HEPES buffer (50 mM, pH 7) containing NaCl (100 mM) at 37 ˚C. The ^{32}P-labeled 2'-deoxyoligonucleotides were resolved on a polyacrylamide gel and the radioactivity in each band quantitatively measured by phosphorimager analysis.

The Wang lab at URC preformed LC-MS/MS analysis of the products generated by nuclease P1 digestion[12,51-53] of the cross-linked duplex **I** and revealed the presence of a tetranucleotide with the Ap-aldehyde conjugated with the underlined adenine (Figure S10), while the corresponding tetranucleotide with the Ap-aldehyde being attached to the guanine residue was not detectable (data not shown). Additionally, LC-MS/MS analysis of the products formed from the digestion with a 4-enzyme cocktail[12] showed that the dA-Ap remnant is present at a level that is approximately two orders of magnitude higher

than its dG-AP counterpart (Figures S10-13). These results mesh with the footprinting data described above and indicate that the cross-linking occurs predominantly between the Ap-aldehyde and the underlined adenine in the 5'-CApT/5'-AAG sequence.

1. 6 Conclusions

We identified a novel cross-link involving reaction of a DNA Ap site with an adenine residue on the opposing strand of the double helix. It is now clear that both dG and dA can engage in cross-linking with an Ap site and two distinct cross-linking sequence motifs, 5'-CAp/5'-AG and 5'-ApT/5'-AA have been defined as locations for the formation of Ap-derived cross-links.[11,12] The difference in the favored locations of dA and dG residues in these cross-links, with the cross-linking nucleobase displaced to the 3'- and 5'-side of the Ap site, respectively, is a natural consequence of the helical twist of duplex DNA coupled with the fact that the N^6-amino group of adenine is located in the major groove while the N^2-amino group of guanine resides in the minor groove.

Our results were consistent with a mechanism involving attack of the exocyclic N^6-amino group of adenine on the aldehyde residue of the Ap site to generate a hemiaminal intermediate (**6**, Scheme 3). The mass spectrometric data is consistent with subsequent dehydration of **6** to give the imine (**7**), cyclic hydroxyalkylhemiaminal (**8**), or enamine (**9**) product. The detailed chemical structure of the dA-Ap cross-link using 2D-NMR will be described in chapter 2, but even before completion of that work chemical precedents involving structurally related small molecules suggested that the cyclic hydroxyalkylhemiaminal (**8**) likely will be the predominant form.[21,22,49,50] The substantial stability of the dA-Ap cross-link generated in duplex **B** is generally consistent with a

cyclic hydroxyalkylhemiaminal **8**, in which the reactive imine group of **7** has been masked by intramolecular attack of the hydroxyl group. Despite the stability of the cross-link to many of the conditions examined here, the reaction is reversible and work in chapters 2 and 3 will demonstrate whether Ap-derived cross-links have sufficient stability to block DNA processing enzymes such as helicases and polymerases.

It may be important to emphasize that the dA-Ap cross-links described here formed in good yields (13-70%) under physiologically-relevant conditions (pH 7, 100 mM NaCl) and did *not* require conditions of reductive amination (e.g. pH 4-5 and $NaCNBH_3$) often used to stabilize and increase the yield of the imine products resulting from the reaction of aldehydes and amines.[11,12,40-45] In the sequences examined to date, the yields of dA-Ap cross-links were much higher than those observed for the dG-Ap cross-links, when both reactions are conducted under the same, physiologically-relevant conditions (pH 7, in the absence of $NaCNBH_3$).[11,12] The sequence 5'-CApT/5'-AAG (duplex **I**) that, in principle, can form *both* dG-Ap and dA-Ap cross-links, in fact, generates a remarkably high yield (~70%) of the dA-Ap cross-link. It seems likely that, in some duplexes such as **I**, multiple cross-linked species exist in dynamic equilibrium. The yield of each cross-linked species at equilibrium will be determined by the inherent abilities of each nucleobase to generate imine linkages with the Ap aldehyde *and* by the ability of the DNA duplex to position the amino and aldehyde groups in a manner that favors imine formation. Thus, formation of these Ap-derived cross-links is conceptually related to template-directed synthesis with dynamic combinatorial imine libraries, in which equilibrium "sorting processes" involving reversible imine formation among a mixture of various aldehydes and amines ultimately generate higher yields of the imine

products that form the most stable ensembles with a templating molecule present in the reaction.[55,56] Our results established that the equilibrium yields of the dA-Ap cross-link can be quite high in some sequence contexts and work in chapter 3 explores the effects of sequence variation on cross-link yields.

The widespread occurrence of Ap sites[14-19] along with the facile nature of the cross-linking reactions described here make it interesting to consider the possibility that Ap-derived cross-links can form in genomic DNA. In cells, processes such as repair[13,57,58] and strand cleavage (as shown in Scheme 2)[59,60] will compete with the formation of Ap-derived cross-links. As a result, the majority of Ap sites in genomic DNA almost certainly do *not* go forward to form interstrand DNA-DNA cross-links. Nonetheless, because interstrand cross-links are especially noxious lesions,[1,2,8-10] even low yields of endogenous Ap-derived cross-links could prove significant in several diverse areas. First, endogenous DNA damage is thought to be important in the etiology of both human aging and sporadic cancers.[61] Because cross-links present severe challenges to replication, transcription, and repair, endogenously-formed cross-links could be especially significant. Unrepaired cross-links that block transcription and replication may promote cell death and senescence that contribute to the aging of tissues.[3,4,62-64] At the same time, cellular attempts to repair cross-links are error-prone and introduce cancer-causing mutations into the genome.[6,65-67]. In addition, the realization that Ap sites can generate cross-links in duplex DNA may help explain why pathways such as nucleotide excision repair (NER) and recombination that are typically associated with the repair of more complex lesions are required for eukaryotic cells to survive the introduction of Ap sites into their genome.[13,57] Finally, Ap-derived cross-

links could be relevant to difficulties inherent in the amplification and sequencing of ancient DNA.[68] Ancient DNA samples contain a plethora of lesions including oxidative base damage, Ap sites, and strand breaks,[68] but it has been suggested that a high density of interstrand cross-links ultimately limit the ability to amplify and sequence the DNA in these samples.[69] The structural nature of the cross-links in ancient DNA has not been elucidated, but Ap-derived cross-links are good candidates for these blocking lesions that hinder sample processing.[70]

Overall, these findings establish Ap-derived interstrand cross-links as a family of DNA lesions that can form in diverse sequence contexts. Studies are currently underway to define the structure, occurrence, biochemical properties, and biological consequences of these lesions.

1.7 Materials and Methods

Oligonucleotides were purchased from Integrated DNA Technologies. All enzymes were purchased from New England Biolabs (Ipswich, MA, USA). [γ-^{32}P]-ATP (6000 C*i*/mmol) was purchased from Perkin Elmer. C-18 sep-pak cartridges were from Waters. BS Polyprep columns were purchased from BioRad. Erythro-9-(2-hydroxy-3-nonyl)adenine (EHNA) hydrochloride was from Tocris Bioscience (Ellisville, MO, USA). Nuclease P1 and phosphodiesterases 1 and 2 were from Sigma-Aldrich (St. Louis, MO, USA). Alkaline phosphatase and proteinase K were from New England Biolabs (Ipswich, MA, USA). Quantification of radioactivity in polyacrylamide gels was carried out using a Personal Molecular Imager (BIORAD) with Quantity One software (v.4.6.5). All other reagents were purchased from Sigma-Aldrich.

Representative procedure for cross-link formation. Single-stranded uracil containing 2'-deoxyoligonucleotides were 5'-labeled using standard procedures.[71] Labeled DNA was annealed[1] with its complimentary strand and treated with the enzyme UDG (50 units/mL, final concentration) to generate Ap sites. UDG enzyme was removed by phenol-chloroform extraction.[71] Individual DNA duplexes were incubated in a buffer composed of HEPES (50 mM, pH 7) containing NaCl (100 mM) at 37 ˚C for 120 h. The DNA was ethanol precipitated,[71] resuspended in formamide loading buffer,[71] loaded onto a 20% denaturing polyacrylamide gel, and the gel electrophoresed for 4 h at 1000 V. Direct loading of cross-linking reactions (without ethanol precipitation) showed that ethanol precipitation does not alter the yield of the slow-migrating, cross-link band. The amount of radiolabeled DNA in each band on the gel was measured by phosphorimager analysis. The time course for the formation of the Ap site cross-link was carried out by incubating a solution containing labeled DNA (approximately 300,000 cpm), in HEPES (50 mM, pH 7) and NaCl (100 mM) at 37 ˚C. At specified time points, aliquots (10 μL) were removed and frozen at –20 ˚C, after all time points were collected samples were loaded directly on the gel and analyzed as described above.

Hydroxyl radical footprinting of cross-linked duplexes. Standard literature protocols were used in the footprinting of the cross-linked duplexes.[26,48] In these experiments, the strand opposing the uracil-containing oligonucleotide was 5'-labeled using standard procedures.[71] Labeled DNA was annealed with the uracil-containing complement[71] and treated with UDG to generate the abasic site as described above. The Ap-containing double-stranded DNA (~500,000 cpm) was incubated in HEPES (50 mM, pH 7) and NaCl (100 mM) at 37 ˚C for 120 h. The DNA was ethanol precipitated, suspended in

formamide loading buffer, and the 2'-deoxyoligonucleotides resolved on a 0.4 mm thick 20% denaturing polyacrylamide gel. The slow-migrating cross-linked duplex band was visualized using X-ray film, the band cut out of the gel, the gel slice crushed, and the gel pieces vortexed in elution buffer (NaCl 200 mM; EDTA, 1 mM) at room temperature for at least 1 h. The mixture was filtered through a poly-prep column to remove gel fragments and the filtrate desalted using a C18 Sep-pak (100 mg size). The resulting solution was evaporated using a Speed-Vac concentrator, the residue redissolved in water (24 μL), split evenly into three microcentrifuge tubes, and diluted with 2x oxidation buffer (10 μL of a solution composed of sodium phosphate, 20 mM, pH 7.2; NaCl, 20 mM; sodium ascorbate, 2 mM; H_2O_2, 1 mM). To this mixture was added a solution of iron-EDTA (2 μL, EDTA, 70 mM; $Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$, 70 mM) to start the reaction, the mixture vortexed briefly, and incubated at room temperature for 1, 2, and 3 min before addition of thiourea stop solution (10 μL of a 100 mM solution in water). Hydroxyl radical footprinting reactions, Maxam-Gilbert G, and Maxam-Gilbert A+G reactions were performed on the labeled single-strand to generate marker lanes.[35] The resulting DNA fragments were analyzed using gel electrophoresis as described above.

Static Nanospray QTOF MS. The oligonucleotide sample was analyzed in a 40 mM dimethylbutylammonium acetate (pH 7.1) buffer. Negative ion MS spectra was taken for mass range of 280-3200 Da on an Agilent6520A QTOF MS with Chip Cube source (G4240A). Monoisotopic neutral masses were calculated from the multiply charged ion spectrum present in the 500-2000 Da mass range. Sample introduction was done with New Objective Econo12 N uncoated borosilicate glass emitters. Negative ion spectrum was acquired at a capillary potential sufficient to initiate spray of the sample. The

nitrogen gas was heated to 290°C and introduced at a flow rate of 4L/min. The Fragmentor, Skimmer, and Octapole1 RF Vpp potentials were set to 200 V, 65 V, and 750 V, respectively. External Calibration was done with the Agilent ESI-Low calibration tuning mixture (cat. no. G1969-85000) and data analysis was performed with Agilent Mass Hunter Workstation QUalitative Analysis software v B.02.00, Build 2.0.197.0 with Bioconfirm Software (2008). Peptide isotope model was assumed and peak set height threshold for extraction was set to ≥ 500 counts. Deconvolution was done with a 0.1 Da step size with a result of 20 iterations of the algorithm calculation

Enzymatic digestion. Duplex **H** was digested using a 4-enzyme cocktail following the conditions described previously.[12] Briefly, nuclease P1 (5 U), phosphodiesterase 2 (0.01 U), EHNA (20 nmol) and a 50-μL solution containing 300 mM sodium acetate (pH 5.6) and 10 mM zinc chloride were added to 300 pmol of duplex **H** in a final volume of 500 μL. In this context, EHNA served as an inhibitor for deamination of 2'-deoxyadenosine (dA) to 2'-deoxyinosine (dI) induced by adenine deaminase.[72] The resulting mixture was incubated at 37˚ C for 48 h. To the digestion mixture were then added alkaline phosphatase (10 U), phosphodiesterase 1 (0.005 U) and 100 μL of 0.5 M Tris-HCl buffer (pH 8.9). The digestion was continued at 37˚ C for 2 h. Duplex **B** was digested using the 4-enzyme cocktail under similar conditions. Digestion with nuclease P1 alone was performed following previously published procedures,[12] where 0.2 unit of nuclease P1 (0.2 U) was added to 300 pmol of duplex **H** in a 500 μL water solution. The resulting mixture was incubated at 37˚ C for 2 h. The above enzymatic digestion mixture was extracted with chloroform to remove enzymes, and the aqueous layer was dried in Speed-

vac, reconstituted in water, and subjected to LC-MS/MS analyses as described previously.[12]

1.8 Supporting Information

Figure S1

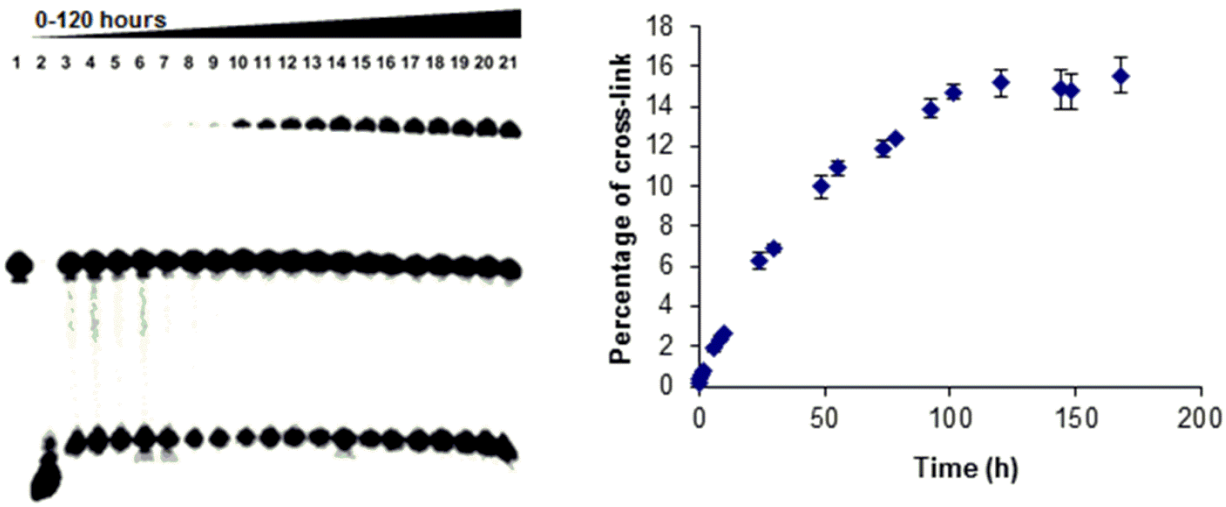


Figure S1. Time course for the formation of the dA-Ap cross-link. Duplex **B** was incubated in HEPES buffer (50 mM, pH 7) containing NaCl (100 mM) at 37 ˚C. At each time point aliquots were removed and frozen at –20 ˚C. The ^{32}P-labeled 2'-deoxyoligonucleotides were resolved on a denaturing polyacrylamide gel. **Left:** Lane 1, uracil-containing precursors of the 2'-deoxyoligonucleotide duplex. Lane 2, the abasic-site-containing duplexes cleaved by treatment with piperidine (1 M, 95 ˚C, for 25 min). Lanes 3-21 are formation time points in hours 0, 0.5, 1, 2, 6, 8, 10, 24, 30, 49, 55, 73, 78, 92, 101, 120, 144, 148, 168. The ^{32}P-labeled 2'-deoxyoligonucleotides were resolved on a denaturing polyacrylamide gel **A**. The radioactivity in each band quantitatively measured by phosphorimager analysis, and the yield of slow-migrating (cross-linked DNA) band plotted versus time. **Right:** Time = 0 was defined as the end of the 1.5 h reaction of the uracil-containing 2'-deoxyoligonucleotide duplex with UDG to generate the Ap-containing 2'-deoxyoligonucleotide duplex **B**.

Figure S2

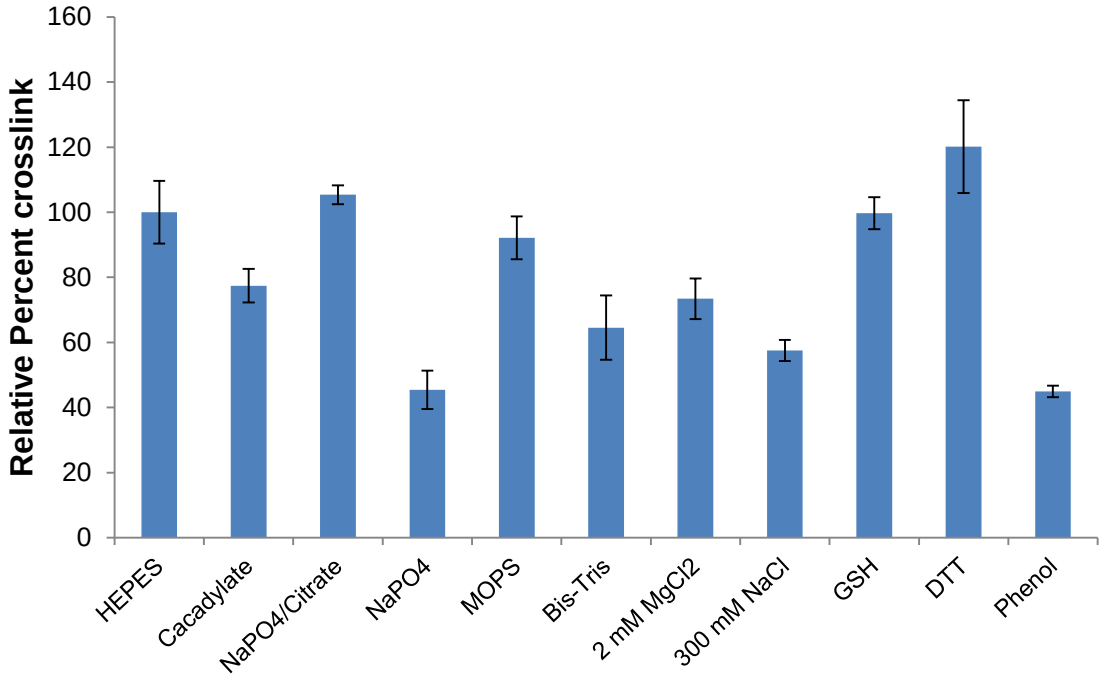


Figure S2. Cross-link formation in duplex **B** in the presence of various buffers, salt conditions, and additives. The Ap-containing duplex B was incubated for 120 h at 37 °C. The ^{32}P-labeled 2'-deoxyoligonucleotides were resolved on a denaturing polyacrylamide gel and the radioactivity in each band quantitatively measured by phosphorimager analysis. In the graph above, the yield of slow-migrating (cross-linked DNA) band was normalized against a reaction carried out under standard conditions (50 mM HEPES pH 7.0, 100 mM NaCl, bar 1, far left on the graph). The conditions employed for the other reactions, from left to right were: bar 2, 50 mM cacodylate pH 7.0, 100 mM NaCl; bar 3, 16.5 mM $NaPO_4$, 1.8 mM citric acid pH 7.0, 100 mM NaCl; bar 4, 50 mM $NaPO_4$ pH 7.2, 50 mM NaCl; bar 5, 50 mM MOPS pH 7.0, 100 mM NaCl; bar 6, 50 mM Bis, 50 mM Tris pH 7.0, 100 mM NaCl; bar 7, 50 mM HEPES pH 7.0, 100 mM NaCl, 2 mM $MgCl_2$; bar 8, 50 mM HEPES pH 7.0, 300 mM NaCl; bar 9, 50 mM HEPES pH 7.0, 100 mM NaCl, 1 mM GSH; bar 10, 50 mM HEPES pH 7.0, 100 mM NaCl, 1 mM DTT; bar 11, 50 mM HEPES pH 7.0, 100 mM NaCl, 100 mM phenol.

Figure S3

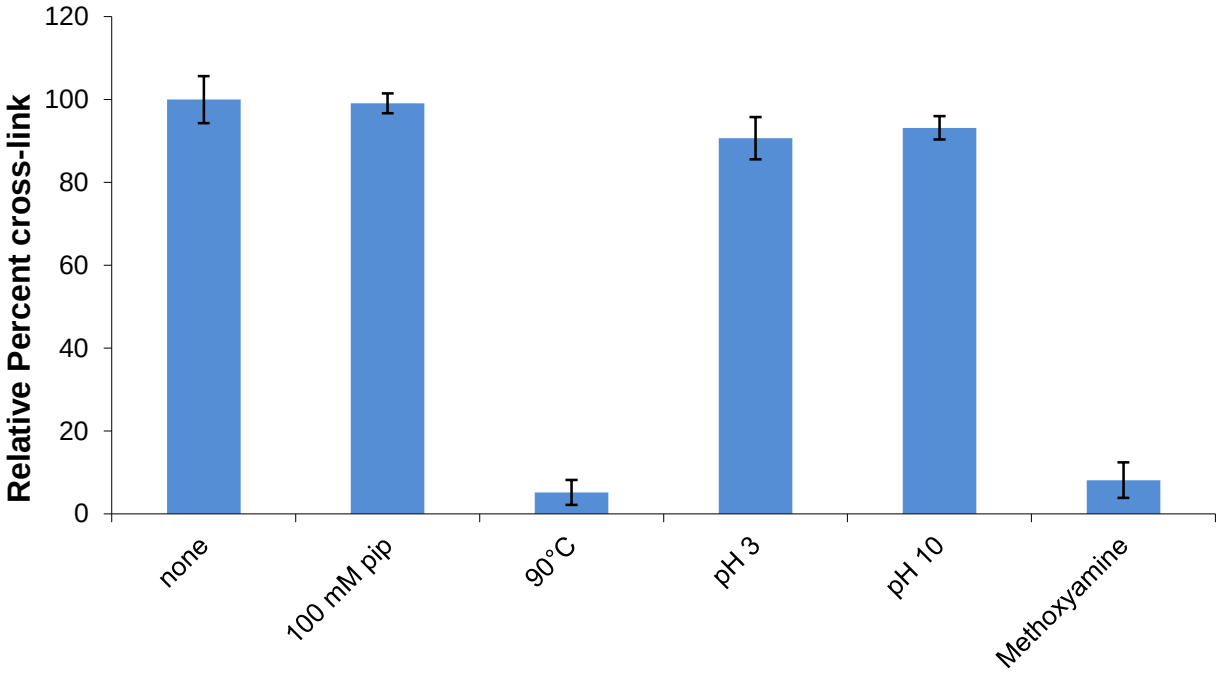


Figure S3. Stability of the cross-linked duplex B under various conditions. Duplex **B** was incubated in HEPES buffer (50 mM, pH 7) containing NaCl (100 mM) at 37 ˚C for 120 h at 37 °C, ethanol precipitated, and then subjected to various conditions. The ^{32}P-labeled 2'-deoxyoligonucleotides were resolved on a denaturing polyacrylamide gel and the radioactivity in each band quantitatively measured by phosphorimager analysis, and the yield of slow-migrating (cross-linked DNA) band was normalized against an aliquot that was precipitated and then resuspended in water (bar 1). Other aliquots were subjected to various work-up conditions: bar 2, 100 mM piperidine 60 °C for 15 min; bar 3, 90 °C for 15 min; bar 4, pH 3 (25 mM sodium acetate) 24 °C for 15 min; bar 5, pH 10 (25 mM potassium phthalate) 24 °C for 15 min, and bar 5, 200 mM methoxyamine at 60 °C for 30 min.

Figure S4

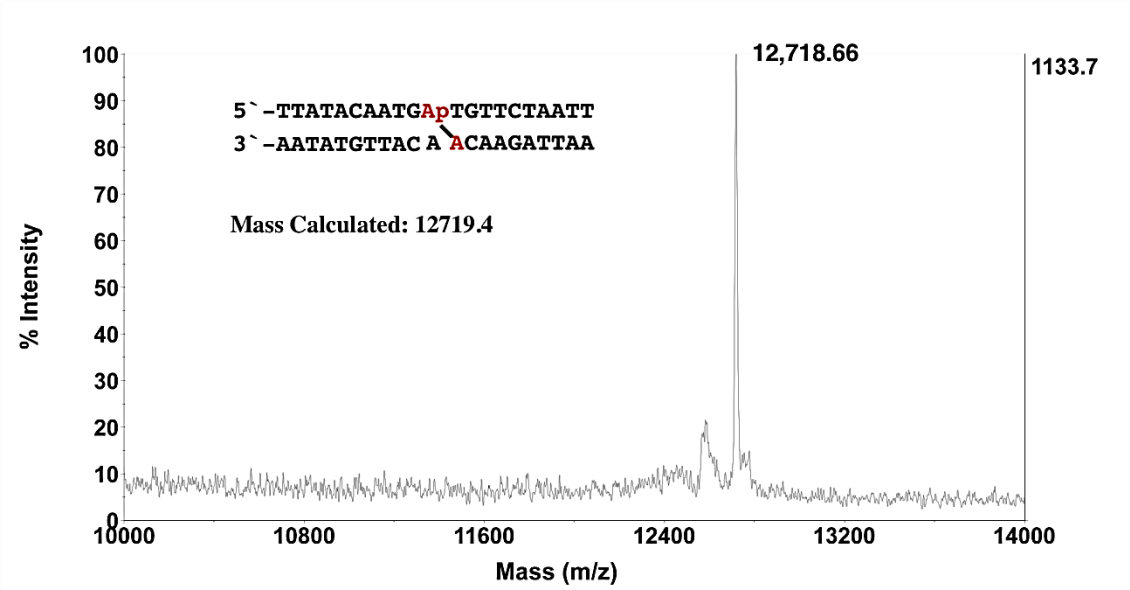


Figure S4. MALDI TOF mass spectrometric analysis of an unlabeled version of cross-linked duplex **B** generated by incubation of duplex **B** in HEPES buffer (50 mM, pH 7.0) containing NaCl (100 mM) at 37 ˚C. The cross-link was purified by preparative gel electrophoresis, desalted using a G-25 Sephadex column and ethanol precipitated. The spectrum was externally calibrated using an oligonucleotide standard with *m*/*z* of 12,378.11. MALDI TOF analysis was performed as described previously.[12]

Figure S5

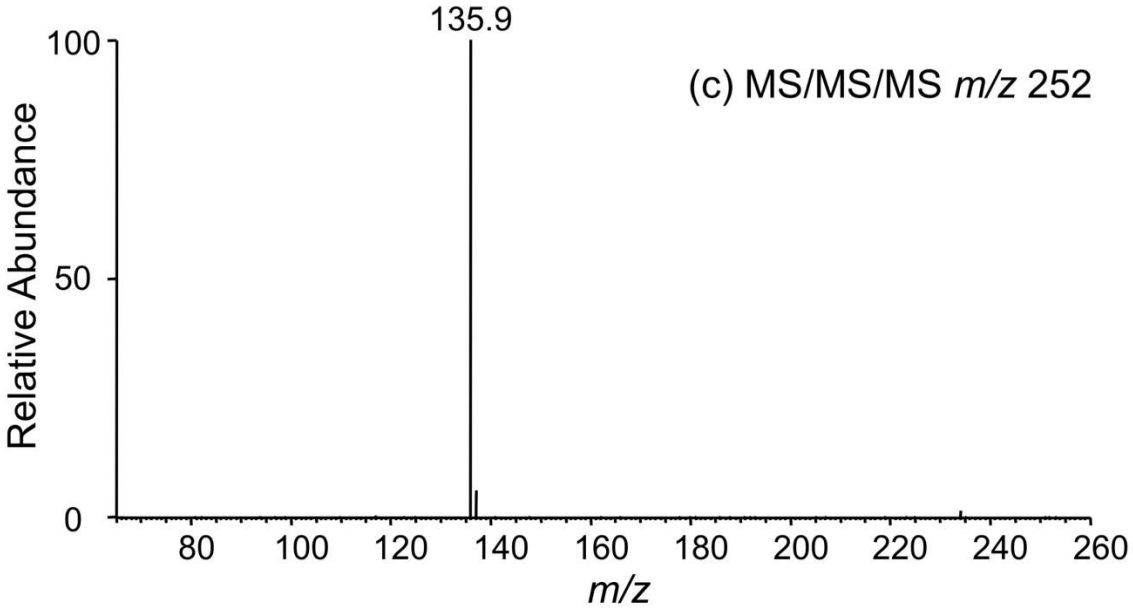
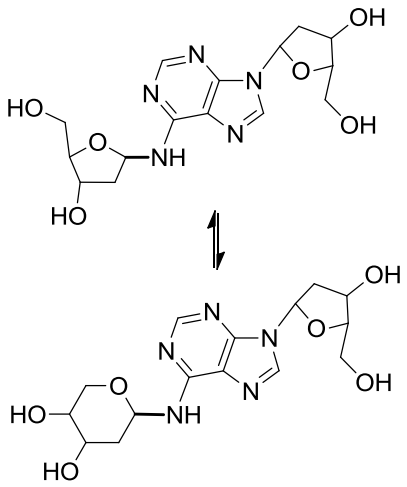


Figure S5. LC-ESI-MS/MS/MS analysis of the 4-enzyme digestion mixture of cross-linked duplex **B**. Shown here is the 252→136 *m/z* transition in positive ion mode.

Scheme S1



Scheme S1. Expected furanose (top) and pyranose (bottom) isomers of the cyclic hemiaminal arising from the reaction of deoxyadenosine with deoxyribose.

Figure S6

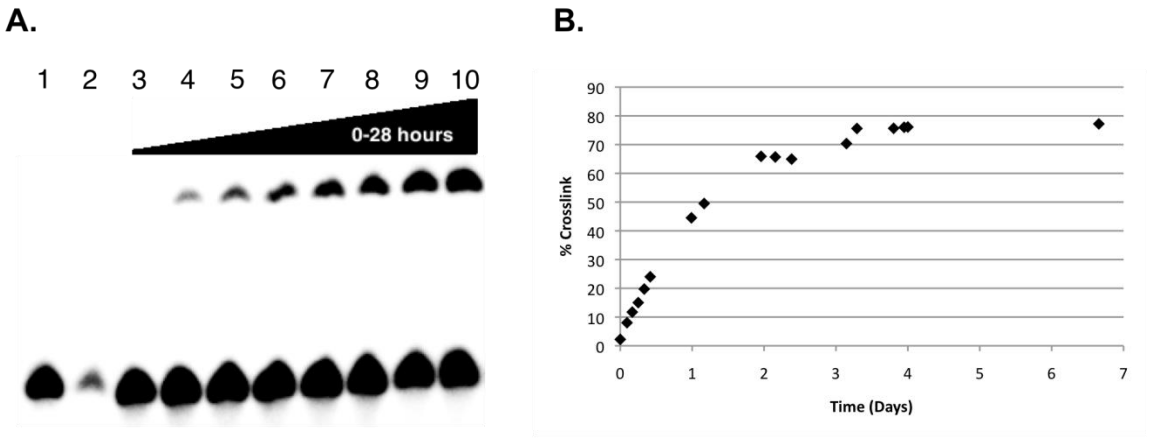


Figure S6. Time course for the formation of cross-link in duplex **I**. The abasic-site-containing duplex was incubated in HEPES buffer (50 mM, pH 7.0) and NaCl (100 mM) at 37 ˚C and aliquots were removed from the reaction and frozen prior to sequencing gel analysis (lanes 4-10). The lower bands correspond to the full length labeled 2'-deoxyoligonucleotides and the upper band cross-linked DNA. Lane 1 is the ^{32}P-labeled uracil-containing precursor 2'-deoxyoligonucleotide, lane 2 is the ^{32}P-labeled abasic-site-containing duplexes cleaved by treatment with piperidine (1 M, 95 ˚C, for 25 min), and lane 3 is the ^{32}P-labeled abasic-site-containing duplex without incubation. The ^{32}P-labeled 2'-deoxyoligonucleotides were resolved on a polyacrylamide gel and the radioactivity in each band quantitatively measured by phosphorimager analysis. The left panel shows a representative gel depicting the first 28 h of the cross-linking reaction. The right panel shows results from obtained from full time course experiments.

Figure S7

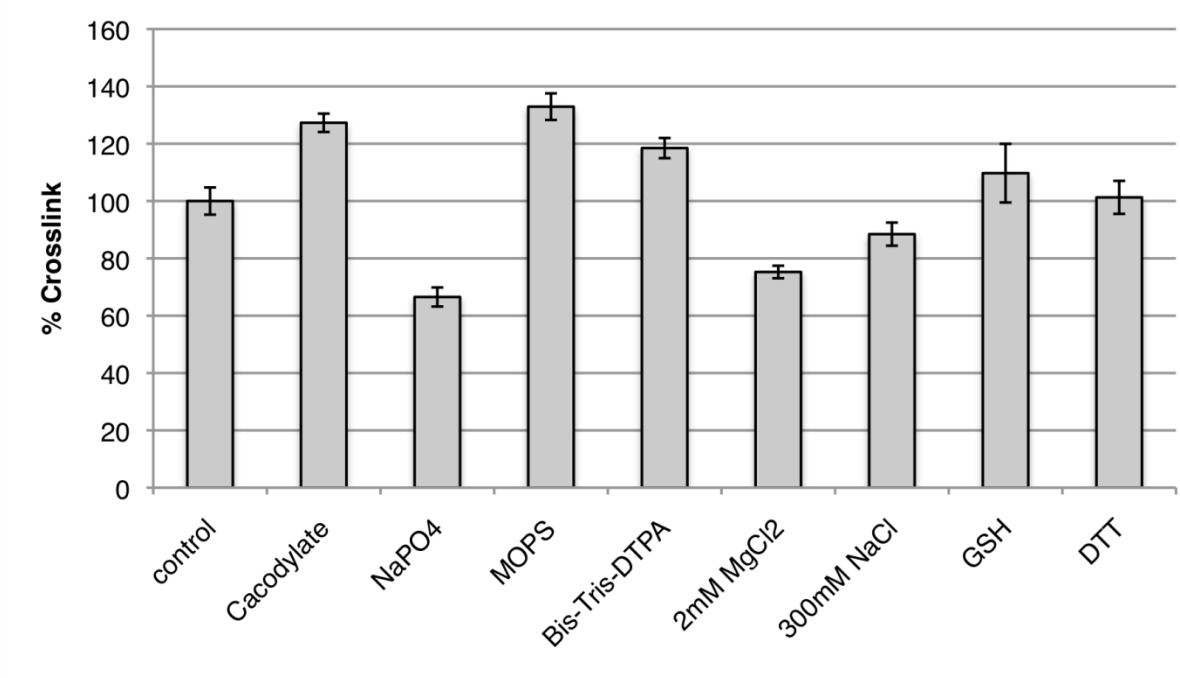


Figure S7. Relative yields of cross-linked DNA generated in duplex **I** in the presence of various buffers, salt conditions, and thiol additives. The "control" (or standard) cross-linking reaction (relative yield 100%, bar 1 on the graph) involved incubation of duplex **I** in HEPES buffer (50 mM, pH 7) containing NaCl (100 mM) at 37 ˚C for 120 h. These experiments examine the effects of different buffers, salts, or thiols present *during* the cross-linking reaction. Yields under other conditions were normalized against that obtained under standard conditions. The conditions examined were: bar 2, 50 mM cacodylate pH 7.0, 100 mM NaCl; bar 3, 50 mM $NaPO_4$ pH 7.2, 50 mM NaCl; bar 4, 50 mM MOPS pH 7.0, 100 mM NaCl; bar 5, 50 mM Bis, 50 mM Tris, 5 mM DTPA pH 7.0, 100 mM NaCl; bar 6, 50 mM HEPES pH 7.0, 100 mM NaCl, 2 mM $MgCl_2$; bar 7, 50 mM HEPES pH 7.0, 300 mM NaCl; bar 8, 50 mM HEPES pH 7.0, 100 mM NaCl, 1 mM GSH; bar 9, 50 mM HEPES pH 7.0, 100 mM NaCl, 1 mM DTT.

Figure S8

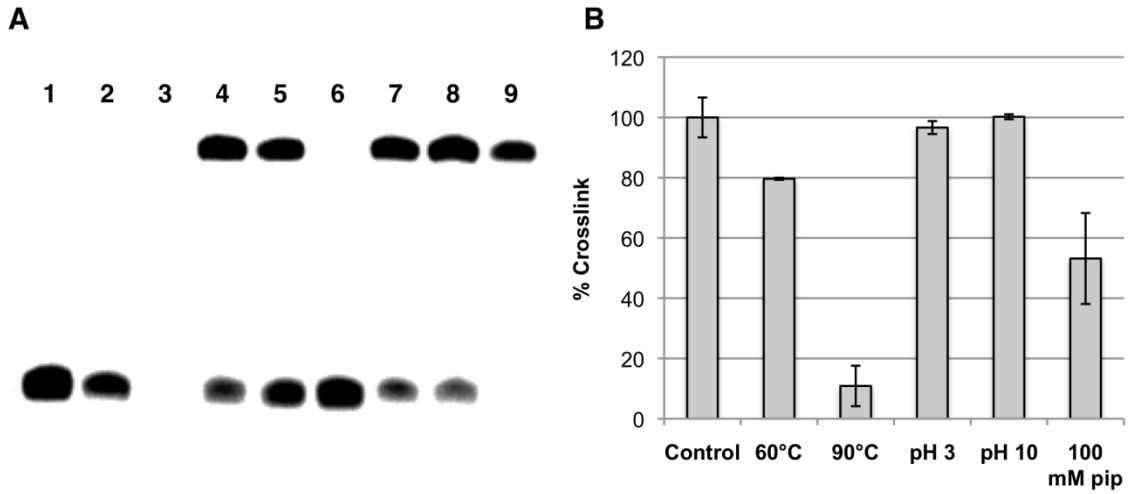


Figure S8. Stability of the cross-linked duplex **I** when subjected to various post-formation work-up conditions. Panel A displays the remaining amount of cross-linked DNA after various treatments compared with untreated cross-link. Lane 1 is the uracil-containing precursor of 2'-deoxyoligonucleotide duplex **I**. Lane 2 is the abasic-site-containing 2'-deoxyoligonucleotide duplex **I** without incubation. Lane 3 is the abasic-site-containing 2'-deoxyoligonucleotide duplex **I** cleaved by treatment with piperidine (1 M, 95 ˚C, 25 min). Lane 4 is duplex **I** incubated in HEPES buffer (50 mM, pH 7) containing NaCl (100 mM) at 37 ˚C, ethanol precipitated and stored in water for 15 min. Lanes 5-9 show cross-linked DNA generated as described for lane 4 except subjected to various work-up conditions: lane 5, 60 ˚C, 15 min; lane 6, 90 ˚C, 15 min; lane 7, pH adjusted to 3, 15 min; lane 8, pH adjusted to 10, 15 min; lane 9, 100 mM piperidine, 60˚ C, 15 min. Panel B is a bar graph with a quantitative illustration of the data from the gels. The yields of the cross-linked DNA obtained remaining after various work-ups were normalized against the amount of cross-link present in the sample that was similarly precipitated, but not subjected to any work-up (lane 4).

Figure S9

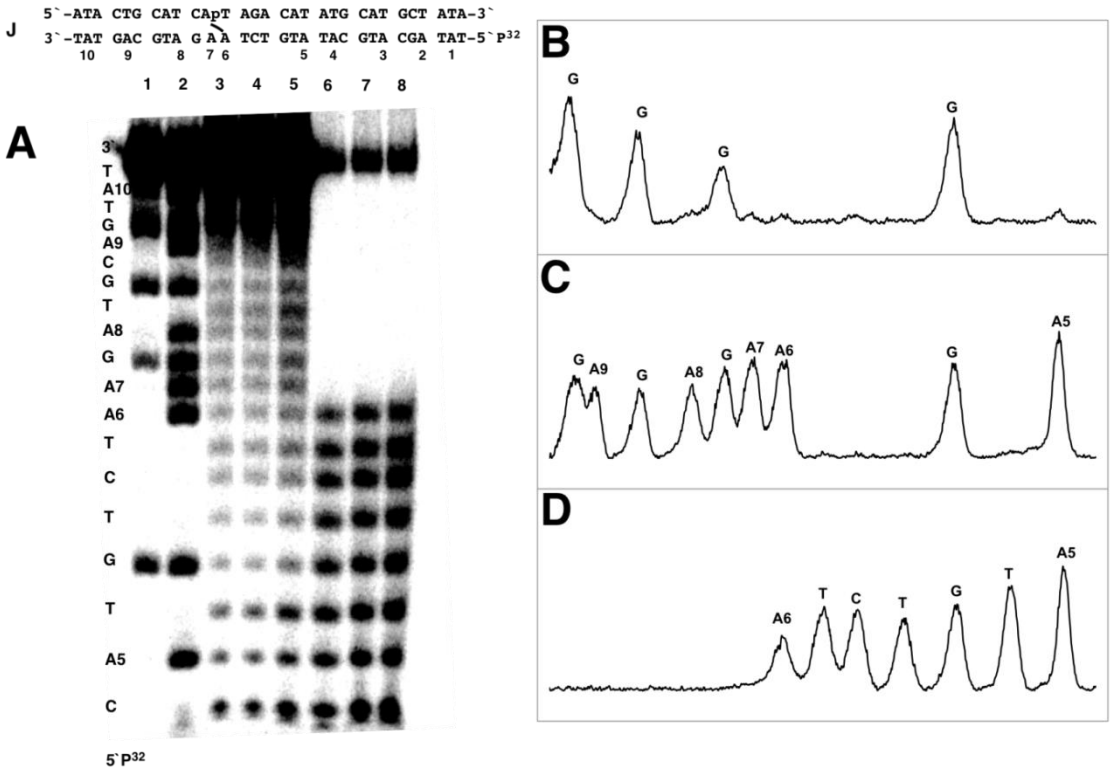


Figure S9. Hydroxyl radical footprinting of duplex **J** to locate the site of cross-link attachment (Panel A). Duplex **J** contains the same core sequence as that in duplex **H**. The slightly longer duplex **J** shown above was employed to move the bands at the cross-linking site out of a "salt-flare" zone in the polyacrylamide gel and gives clear, sharp bands at the cross-link site. Lane 1 is a Maxam-Gilbert G-specific cleavage (sequencing) reaction of the labeled 2'-deoxyoligonucleotide strand in duplex **J**. Lane 2 is an A+G specific cleavage (sequencing) reaction of the labeled 2'-deoxyoligonucleotide strand in duplex **J**. Lanes 3-5 are the hydroxyl radical footprinting reactions of the labeled 2'-deoxyoligonucleotide strand in duplex **J**. Lanes 6-9 are the hydroxyl radical footprinting reactions of the slow-migrating, cross-link band generated by incubation of duplex **J** in HEPES buffer (50 mM, pH 7.0) and NaCl (100 mM) at 37 ˚C. The ^{32}P-labeled 2'-deoxyoligonucleotides were resolved on a polyacrylamide gel visualized by phosphorimager analysis (Panel A) and used to develop densitometry traces (Panels B-D). Panel B is the Maxam-Gilbert G-specific cleavage (sequencing) reaction of the labeled 2'-deoxyoligonucleotide strand in duplex **J**. Panel C is the A+G specific cleavage (sequencing) reaction of the labeled 2'-deoxyoligonucleotide strand in duplex **J**. Panel D is the hydroxyl radical footprinting reaction of the slow-migrating cross-link band.

Figure S10

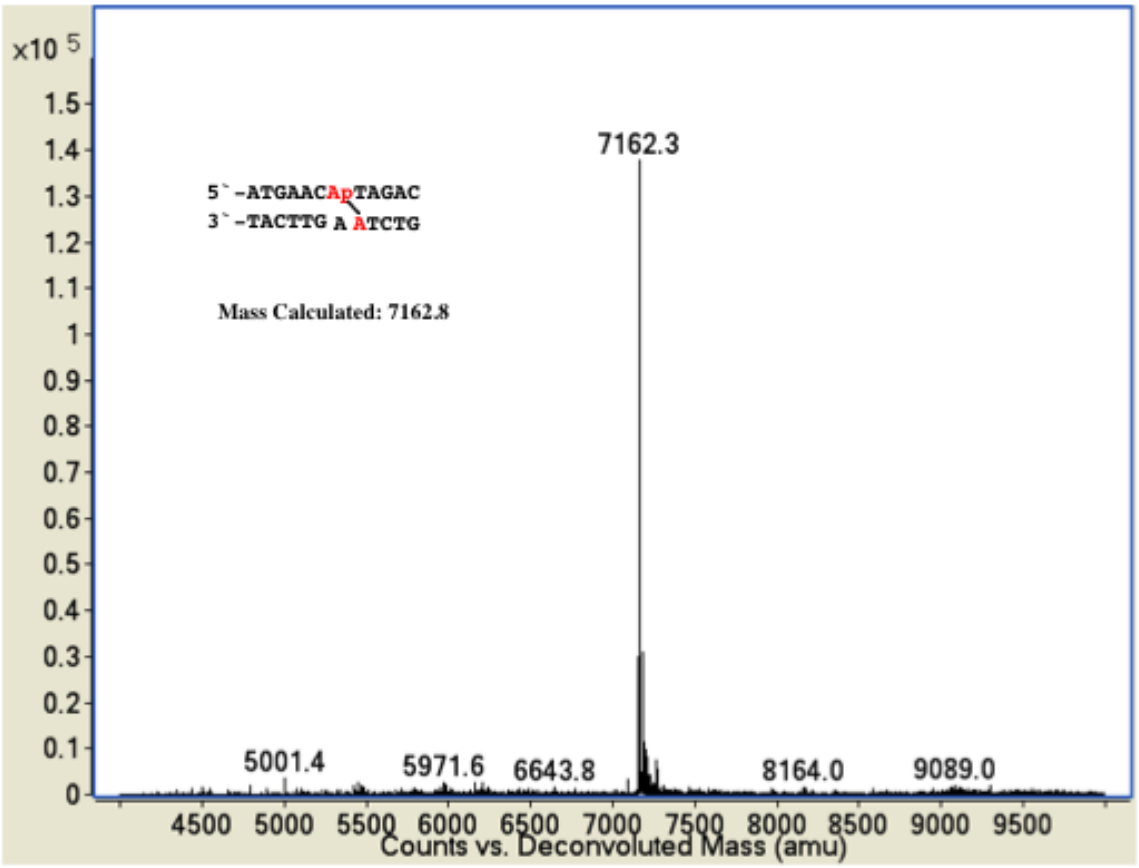


Figure S10. Nanospray QTOF MS of a cross-linked duplex containing the same core sequence found in duplex **I**.

Figure S11

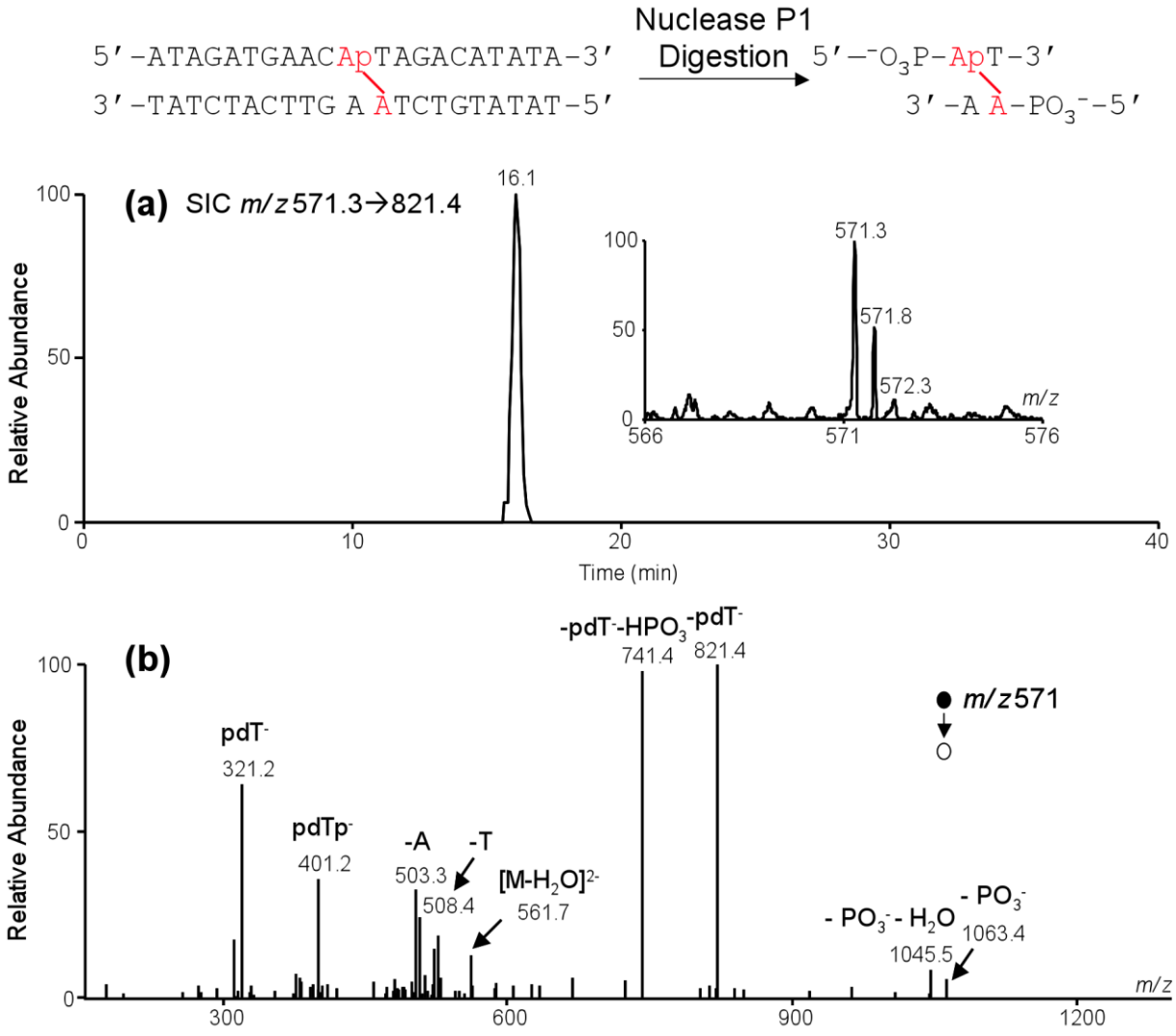


Figure S11. LC-ESI-MS and MS/MS for the analysis of the nuclease P1 digestion mixture of duplex **I** with a dA-Ap cross-link. Shown in (a) is the selected-ion chromatogram (SIC) for monitoring the loss of a thymidine-5'-phosphate from the $[M–2H]^{2–}$ ion of the tetramer shown (i.e., the *m/z* 571.3→821.4 transition). Inset gives the higher resolution "ultra-zoom" ESI-MS for the $[M–2H]^{2–}$ ion of the tetramer. Displayed in (b) is the tandem mass spectrum for the $[M–2H]^{2–}$ ion of the tetramer.

Figure S12

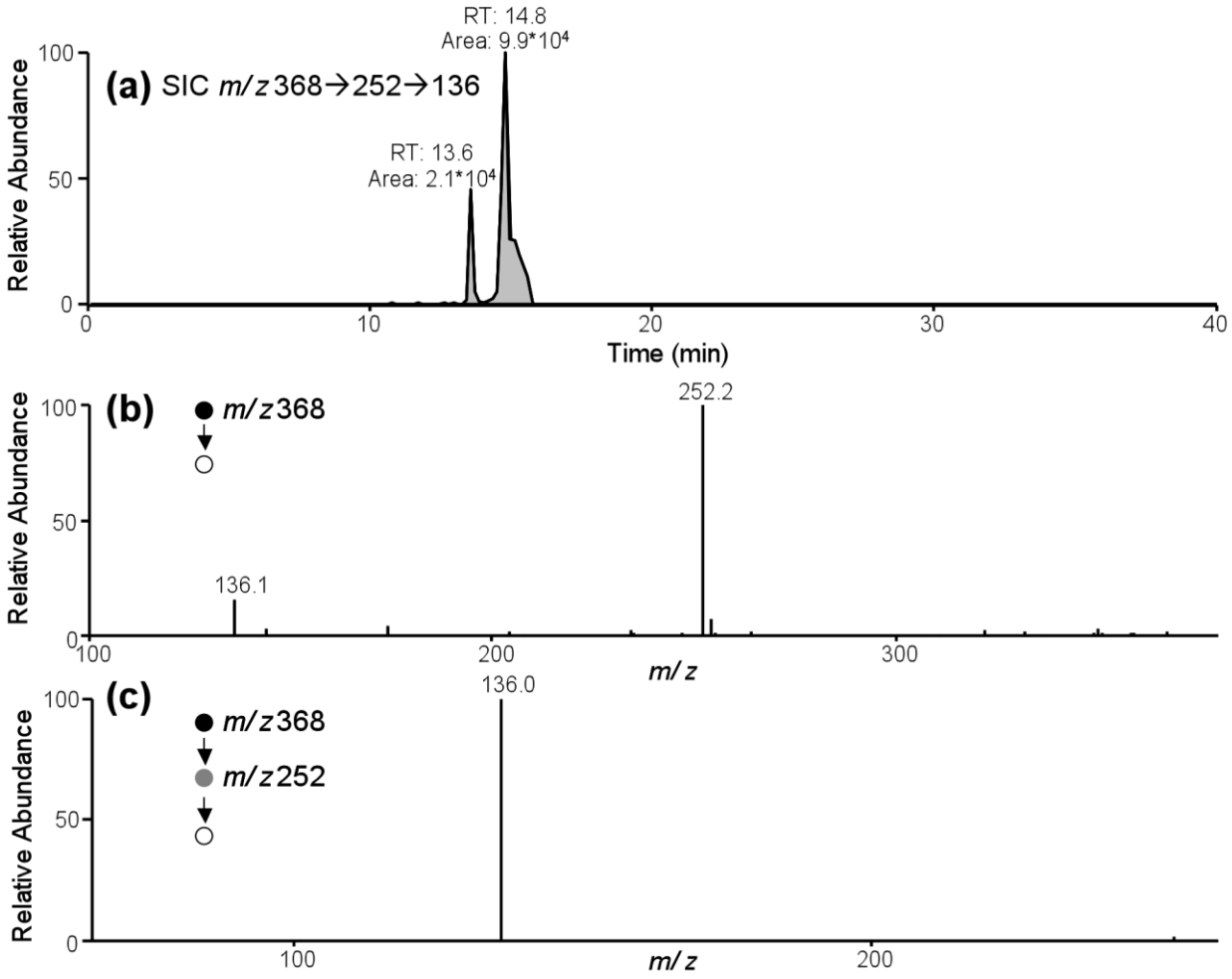


Figure S12. LC-ESI-MS/MS/MS for the analysis of the 4-enzyme digestion mixture of duplex **I** with a dA-Ap cross-link. Shown in (a) is the selected-ion chromatogram (SIC) for monitoring the m/z 368 → 252 → 136 transition. Displayed in (b) is the MS/MS from the cleavage of the $[M + H]^+$ ion of the completely digested cross-link remnant. Depicted in (c) is MS/MS/MS arising from the fragmentation of the ion of *m/z* 252 observed in (b).

Figure S13

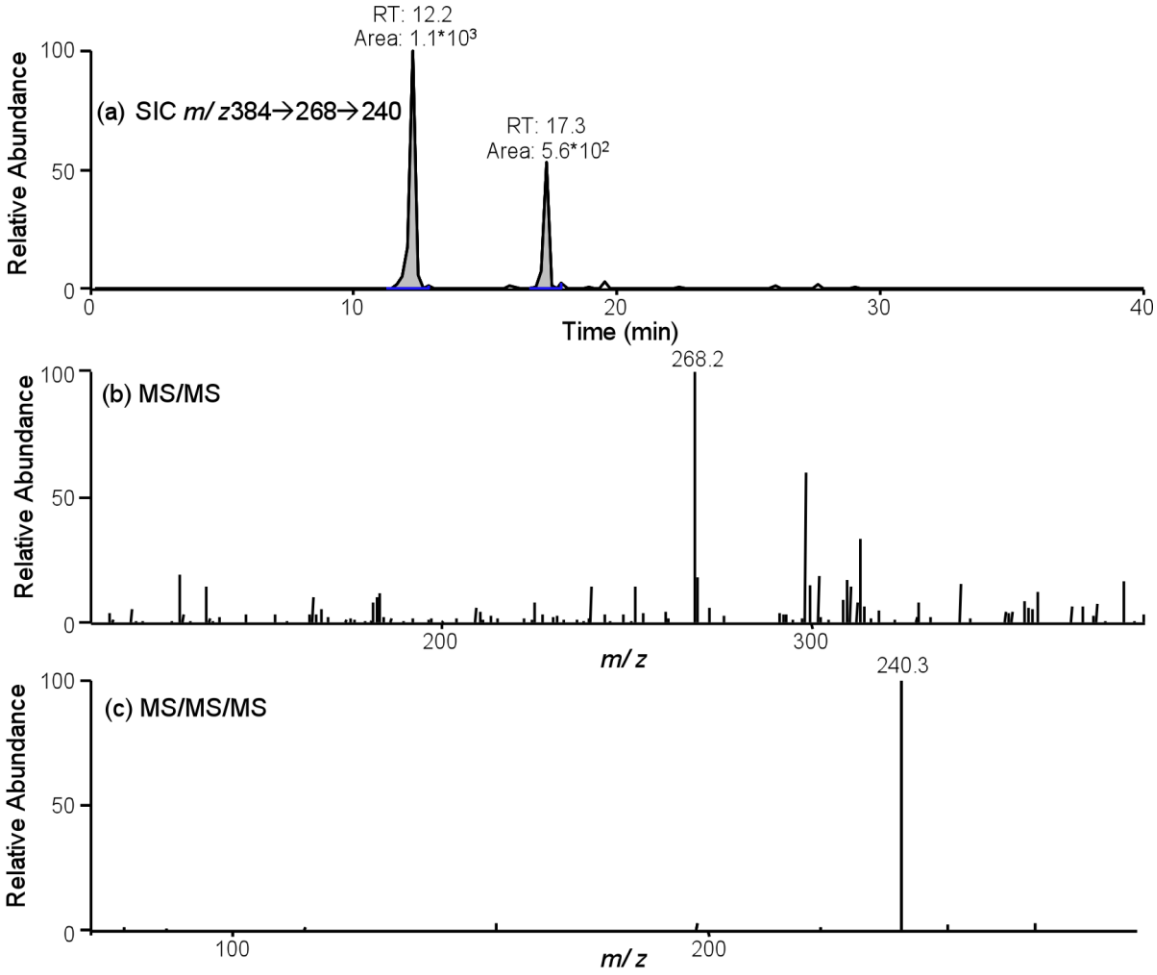


Figure S13. LC-ESI-MS/MS/MS for the analysis of the 4-enzyme digestion mixture of duplex **I**, probing whether the cross-linked duplex contains a dG-Ap cross-link. Shown in (a) is the selected-ion chromatogram (SIC) for monitoring the *m/z* 384 → 268 → 240 transition. Displayed in (b) is the MS/MS from the cleavage of the $[M + H]^+$ ion of the completely digested cross-link remnant. Depicted in (c) is MS/MS/MS arising from the fragmentation of the ion of *m/z* 268 observed in (b).

Figure S14

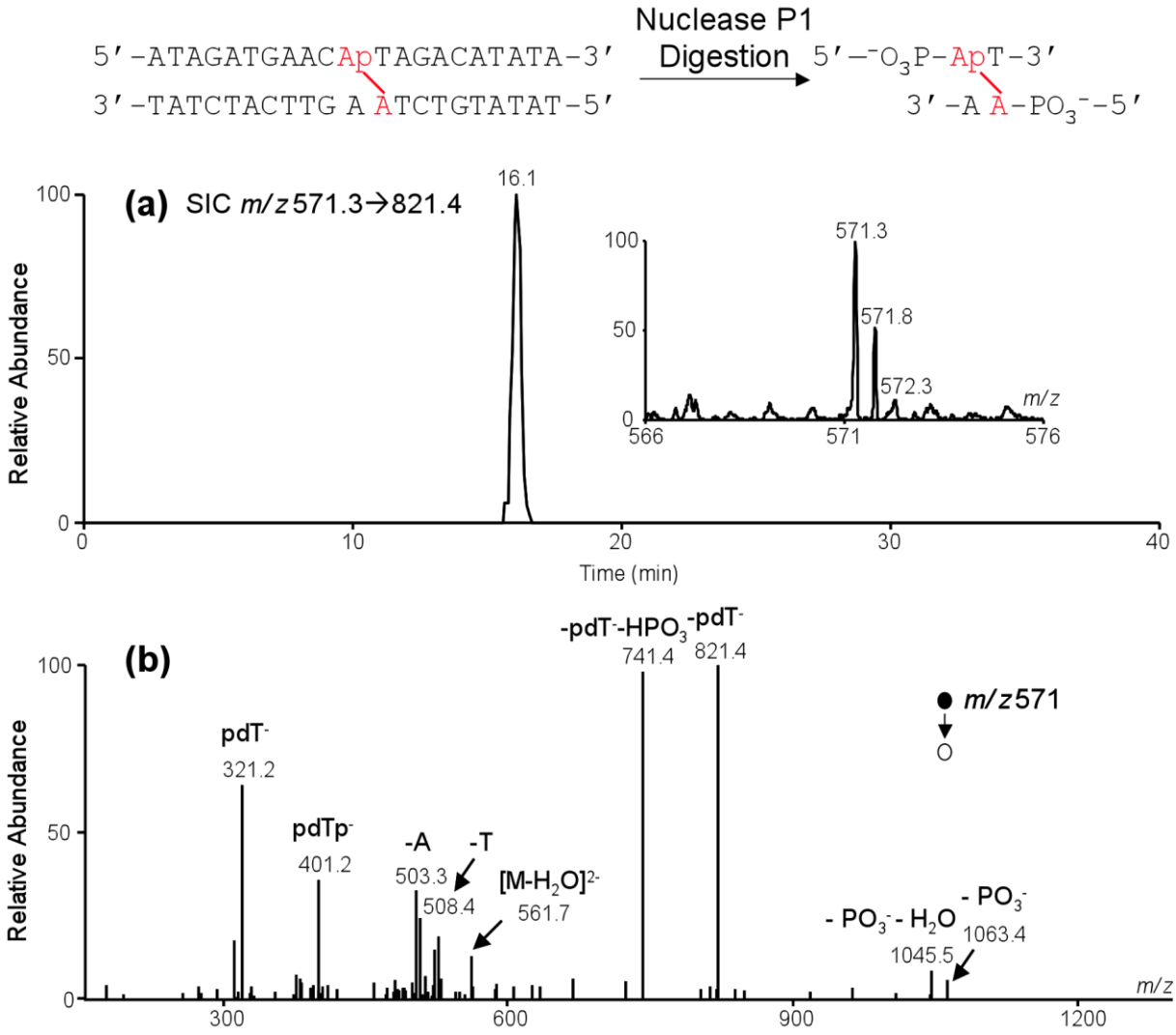


Figure S14. LC-ESI-MS and MS/MS for the analysis of the nuclease P1 digestion mixture of duplex **I** with a dA-Ap cross-link. Shown in (a) is the selected-ion chromatogram (SIC) for monitoring the loss of a thymidine-5'-phosphate from the $[M–2H]^{2–}$ ion of the tetramer shown (i.e., the *m/z* 571.3→821.4 transition). Inset gives the higher resolution "ultra-zoom" ESI-MS for the $[M–2H]^{2–}$ ion of the tetramer. Displayed in (b) is the tandem mass spectrum for the $[M–2H]^{2–}$ ion of the tetramer.

Figure S15

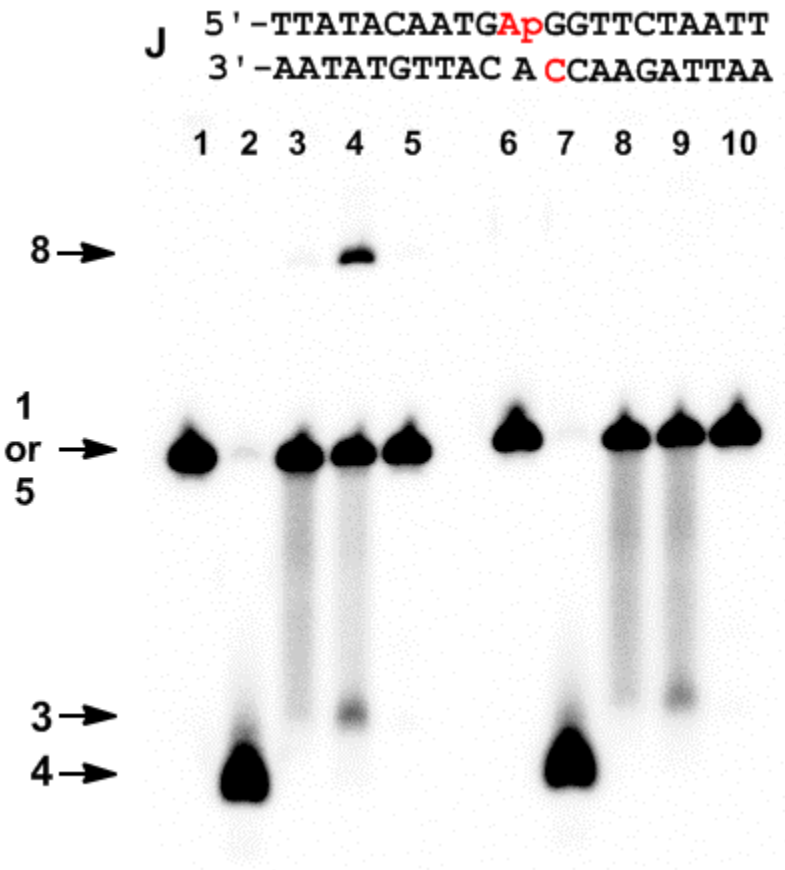


Figure S15. Lanes 1-5, duplex **B** and lanes 6-9, duplex **J**. Lanes 1 and 6 contain ^{32}P-labeled, uracil-containing 2'-deoxyoligonucleotide duplexes. . Lanes 2 and 7 contain the abasic-site-containing duplexes subjected to piperidine work-up (1 M, 95 ˚C, 25 min) to cleave the Ap site. Lanes 3 and 8 contain the ^{32}P-labeled UDG-treated (abasic-site-containing) duplexes without incubation Lanes 4 and 9 contain the cross-linking reactions involving incubation of the abasic-site-containing duplexes in HEPES buffer (50 mM, pH 7) containing NaCl (100 mM) at 37 ˚C. Lanes 5 and 10 contain the methoxyamine capping reactions involving incubation of the abasic-site-containing duplexes incubated in HEPES buffer (50 mM, pH 7) containing NaCl (100 mM), and CH_3ONH_2•HCl (2 mM) at 37 ˚C. The ^{32}P-labeled 2'-deoxyoligonucleotides were resolved by electrophoresis on a 20% denaturing polyacrylamide gel and the radioactivity in each band quantitatively measured by phosphorimager analysis.

1.9 References

(49) Amann, N.; Wagenknecht, H.-A. *Tetrahedron Lett.* **2003**, *44*, 1685-1690.

(44) Angelov, T.; Guainazzi, A.; Schärer, O. D. *Org. Lett.* **2009**, *11*, 661-664.

(36) Bayley, C. R.; Brammer, K. W.; Jones, A. S. *J. Chem. Soc.* **1961**, 1903-1907.

(63) Bergstrahl, D. T.; Sekelsky, J. *Trends in Genetics* **2007**, *24*, 70-76.

(13) Boiteux, S.; Guillet, M. *DNA Repair* **2004**, *3*, 1-12.

(52) Cao, H.; Hearst, J. E.; Corash, L.; Wang, Y. *Anal. Chem.* **2008**, *80*, 2932-2938.

(50) Capon, B.; Connett, B. E. *J. Chem. Soc.* **1965**, 4492-4497.

(23) Chaw, Y. F. M.; Crane, L. E.; Lange, P.; Shapiro, R. *Biochemistry* **1980**, *19*, 5525-5531.

(32) Chen, J.; Dupradeau, F.-Y.; Case, D. A.; Turner, C. J.; Stubbe, J. *Nucleic Acids Res.* **2008**, *36*, 253-262.

(46) Cordes, E. H.; Jencks, W. P. *J. Am. Chem. Soc.* **1963**, *85*, 2843-2848.

(70) Cross-links derived from oxidative DNA damage and lipid peroxidation products also must be considered, see: (a) Stone, M. P.; Cho, Y. J.; Huang, H.; Kim, H. Y.; Kozekov, I. D.; Kozekova, A.; Wang, H.; Minko, I. G.; Lloyd, R. S.; Harris, T. M.; Rizzo, C. J. *Acc. Chem. Res.* **2008**, *41*, 793-804, (b) Huang, H.; Solomon, M. S.; Hopkins, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 9240-9241, (c) Ding, H.; Majumdar, A.; Tolman, J. R.; Greenberg, M. M. *J. Am. Chem. Soc.* **2008**, *130*, 17981-17987, (d) Regulus, P.; Duroux, B.; Bayle, P.-A.; Favier, A.; Cadet, J.; Ravanat, J.-L. *Proc. Nat. Acad. Sci. USA* **2007**, *104*, 14032-14037, (e) Chen, H.-J. C.; Chen, Y.-C. *Chem. Res. Toxicol.* **2009**, *22*, 1334-1341.

(16) De Bont, R.; van Larebeke, N. *Mutagenesis* **2004**, *19*, 169-185.

(4) Derheimer, F. A.; Hicks, J. K.; Paulsen, M. T.; Canman, C. E.; Ljungman, M. *Mol. Pharm.* **2009**, *75*, 599-607.

(15) Dong, M.; Wang, C.; Deen, W. M.; Dedon, P. C. *Chem. Res. Toxicol.* **2003**, *16*, 1044-1055.

(11) Dutta, S.; Chowdhury, G.; Gates, K. S. *J. Am. Chem. Soc.* **2007**, *129*, 1852-1853.

(64) Garinis, G. A.; van der Horst, G. T. J.; Vijg, J.; Hoeijmakers, J. H. J. *Nat. Cell Biol.* **2008**, *10*, 1241-1247.

(14) Gates, K. S. *Chem. Res. Toxicol.* **2009**, *22*, 1747-1760.

(18) Gates, K. S.; Nooner, T.; Dutta, S. *Chem. Res. Toxicol.* **2004**, *17*, 839-856.

(28) Goodwin, J. T.; Lynn, D. G. *J. Am. Chem. Soc.* **1992**, *114*, 9197-9198.

(67) Greenberg, R. B.; Alberti, M.; Hearst, J. E.; Chua, M. A.; Saffran, W. A. *J. Biol. Chem.* **2001**, *276*, 31551-31560.

(9) Grossman, K. F.; Ward, A. M.; Matkovic, M. E.; Folias, A. E.; Moses, R. E. *Mutation Res.* **2001**, *487*, 73-83.

(3) Groth, V. D.; Carlsson, R.; Johansson, F.; Erixon, K.; Jenssen, D. *DNA Repair* **2012**, *11*, 976-985.

(27) Guan, L.; Greenberg, M. M. *J. Am. Chem. Soc.* **2009**, *131*, 15225-15231.

(17) Guillet, M.; Bioteux, S. *Mol. Cell Biol.* **2003**, *23*, 8386-8394.

(69) Hansen, A. J.; Mitchell, D. L.; Wiuf, C.; Paniker, L.; Brand, T. B.; Binladen, J.; Gilichinsky, D. A.; Rønn, R.; Willerslev, E. *Genetics* **2006**, *173*, 1175–1179.

(21) Hermida, S. A. S.; Possari, E. P. M.; Souza, D. B.; de Arruda Campos, I. P.; Gomes, O. F.; Di Mascio, P.; Medeiros, M. H. G.; Loureiro, A. P. M. *Chem. Res. Toxicol.* **2006**, *19*, 927-936.

(7) Hlavin, E. M.; Smeaton, M. B.; Miller, P. S. *Env. Mol. Mutagenesis* **2010**, *51*, 604-624.

(65) Ho, T. V.; Schärer, O. D. *Env. Mol. Mutagenesis* **2010**, *51*, 552-566.

(42) Hochgürtel, M.; Biesinger, R.; Kroth, H.; Piecha, D.; Hofmann, M. W.; Krause, S.; Schaaf, O.; Nicolau, C.; Eliseev, A. V. *J. Med. Chem.* **2003**, *46*, 356-358.

(61) Hoeijmakers, J. H. J. *J. Engl. J. Med.* **2009**, *361*, 1475-1485.

(47) Holton, S.; Runquist, O. *J. Org. Chem.* **1961**, *26*, 5193-5195.

(31) Hopkins, P. B.; Millard, J. T.; Woo, J.; Weidner, M. F.; Kirchner, J. J.; Sigurdsson, S. T.; Raucher, S. *Tetrahedron* **1991**, *47*, 2475-2489.

(12) Johnson, K. M.; Price, N. E.; Wang, J.; Fekry, M. I.; Dutta, S.; Seiner, D. R.; Wang, Y.; Gates, K. S. *J. Am. Chem. Soc.* **2013**, *135*, 1015-1025.

(45) Kozekov, I. D.; Turesky, R. J.; Alas, G. R.; Harris, C. M.; Harris, T. M.; Rizzo, C. J. *Chem. Res. Toxicol.* **2010**, *23*, 1701-1713.

(40) Kurtz, A. J.; Dodson, M. L.; Lloyd, R. S. *Biochemistry* **2002**, *41*, 7054-7064.

(51) Lai, C.; Cao, H.; Hearst, J. E.; Corash, L.; Luo, H.; Wang, Y. *Anal. Chem.* **2008**, *80*, 8790-8798.

(43) Li, X.; Zhan, Z.-Y. J.; Knipe, R.; Lynn, D. G. *J. Am. Chem. Soc.* **2002**, *124*, 746-747.

(59) Lindahl, T.; Andersson, A. *Biochemistry* **1972**, *11*, 3618-3623.

(33) Lindahl, T.; Ljunquist, S.; Siegert, W.; Nyberg, B.; Sperens, B. *J. Biol. Chem.* **1977**, *252*, 3286-3294.

(48) Luce, R. A.; Hopkins, P. B. *Methods Enzymol.* **2001**, *340*, 396-412.

(8) Magana-Schwencke, N.; Henriques, J. A.; Chanet, R.; Moustacchi, E. *Proc. Nat. Acad. Sci. USA* **1982**, *79*, 1722-1726.

(41) Manoharan, M.; Andrade, L. K.; Cook, P. D. *Org. Lett.* **1999**, *1*, 311-314.

(35) Maxam, A. M.; Gilbert, W. *Methods Enzymol.* **1980**, *65*, 499-560.

(20) McGhee, J. D.; von Hippel, P. H. *Biochemistry* **1975**, *14*, 1281-1296.

(55) Meyer, C. D.; Joiner, C. S.; Stoddart, J. F. *Chem. Soc. Rev.* **2007**, *36*, 1705–1723.

(68) Mitchell, D.; Willerslev, E.; Hansen, A. *Mutation Res.* **2005**, *571*, 265-276.

(5) Muniandy, P. A.; Liu, J.; Majumdar, A.; Liu, S.-t.; Seidman, M. M. *Crit. Rev.*

(19) Nakamura, J.; Swenberg, J. A. *Cancer Res.* **1999**, *59*, 2522-2526.

(30) Niu, J.; Hili, R.; Liu, D. R. *Nat. Chem.* **2013**, *5*, 282-292.

(2) Noll, D. M.; Mason, T. M.; Miller, P. S. *Chem. Rev.* **2006**, *106*, 277-301.

(57) Otterlei, M.; Kavli, B.; Standal, R.; Skjelbred, C.; Bharati, S.; Krokan, H. E. *EMBO J.* **2000**, *19*, 5542-5551.

(56) Ramström, O.; Lehn, J.-M. *Nat. Rev. Drug. Disc.* **2001**, *1*, 26-36.

(10) Reddy, M. C.; Vasquez, K. M. *Radiat. Res.* **2005**, *164*, 345-356.

(38) Rosa, S.; Fortini, P.; Karran, P.; Bignami, M.; Dogliotti, E. *Nucleic Acids Res.* **1991**, *19*, 5569-5574.

(71) Sambrook, J.; Fritsch, E. F.; Maniatis, T. Molecular Cloning: A Lab Manual; Cold Spring Harbor Press: Cold Spring Harbor, NY, **1989**.

(1) Schärer, O. D. *ChemBioChem* **2005**, *6*, 27-32.

Biochem. Mol. Biol. **2010**, *45*, 23-49.

(62) Schumacher, B.; Garinis, G. A.; Hoeijmakers, J. H. J. *Trends in Genetics* **2007**, *24*, 77-85.

(24) Sczepanski, J. T.; Heimstra, C. N.; Greenberg, M. M. *Bioorg. Med. Chem.* **2011**, *19*, 5788-5793.

(25) Sczepanski, J. T.; Jacobs, A. C.; Greenberg, M. M. *J. Am. Chem. Soc.* **2008**, *130*, 9646-9647.

(26) Sczepanski, J. T.; Jacobs, A. C.; Majumdar, A.; Greenberg, M. M. *J. Am. Chem. Soc.* **2009**, *131*, 11132-11139.

(6) Shen, X.; Li, L. *Env. Mol. Mutagenesis* **2010**, *51*, 493-499.

(29) Silverman, A. P.; Kool, E. T. *Chem. Rev.* **2006**, *106*, 3775-3789.

(58) Srivastava, D. K.; Vande Berg, B. J.; Prasad, R.; Molina, J. T.; Beard, W. A.; Tomkinson, A. E.; Wilson, S. H. *J. Biol. Chem.* **1998**, *273*, 21203-21209.

(37) Sugiyama, H.; Fujiwara, T.; Ura, A.; Tashiro, T.; Yamamoto, K.; Kawanishi, S.; Saito, I. *Chem. Res. Toxicol.* **1994**, *7*, 673-683.

(39) Talpaert-Borle, M.; Liuzzi, M. *Biochim. Biophys. Acta* **1983**, *740*, 410-16.

(54) Tomasz, M.; Lipman, R.; Lee, M. S.; Verdine, G. L.; Nakanishi, K. *Biochemistry* **1987**, *26*, 2010-2027.

(22) Upadhyaya, P.; Hecht, S. S. *Chem. Res. Toxicol.* **2008**, *21*, 2164-2171.

(34) Varshney, U.; van de Sande, J. H. *Biochemistry* **1991**, *30*, 4055-4061.

(53) Wang, Y.; Wang, Y. *Anal. Chem.* **2003**, *75*, 6306-6313.

(72) Wang, J.; Yuan, B.; Guerrero, C.; Bahde, R.; Gupta, S.; Wang, Y. *Anal. Chem.* **2011**, *83*, 2201-2209.

(66) Zheng, H.; Wang, X.; Warren, A. J.; Legerski, R. J.; Nairn, R. S.; Hamilton, J. W.; Li, L. *Mol. Cell Biol.* **2003**, *23*, 754-761.

(60) Zhou, C.; Sczepanski, J. T.; Greenberg, M. M. *J. Am. Chem. Soc.* **2012**, *134*, 16734-16741.

Chapter 2. Chemical and structural characterization of interstrand cross-links formed between abasic sites and adenine residues in duplex DNA

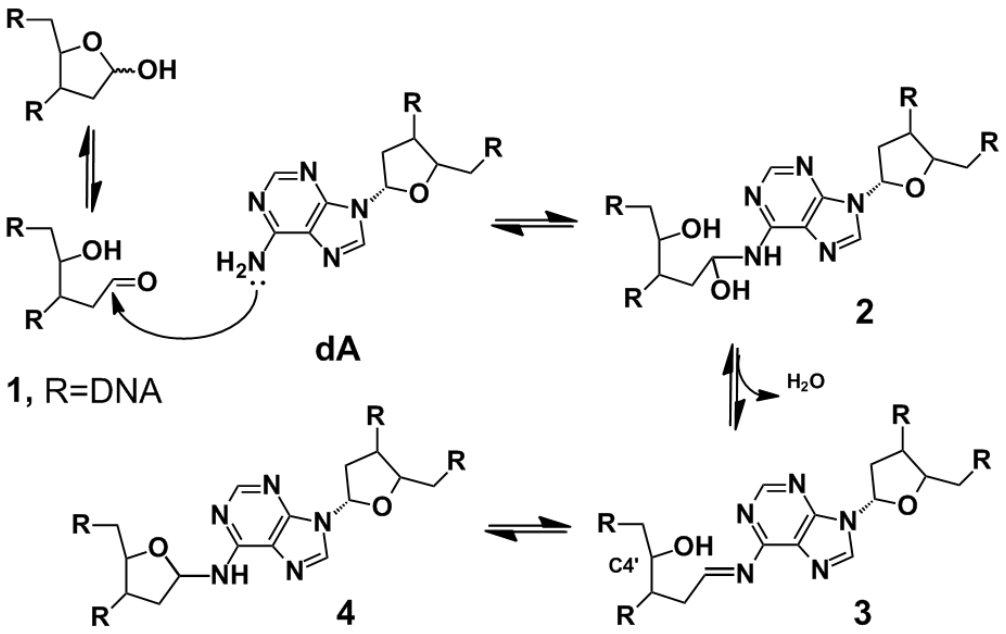
2.1 Introduction

Chemical modification of cellular DNA can have profound biological consequences. Formation of DNA lesions can either be cytotoxic or mutagenic.[1–5] The chemical structure of a DNA lesion ultimately determines the exact nature of dysfunction they cause.[6–8] Thus, careful chemical structure determination is a critical step in the overall characterization of any particular DNA lesion. Rigorous chemical structure determination of DNA adducts requires spectroscopic characterization of the lesion.[9] This may involve direct characterization of the lesion in, or released from, duplex DNA.[10,11] Alternately, one can synthesize a standard of a putative lesion and use LC/MS to compare the properties of the authentic standard to material released from duplex DNA by enzymatic digestion.[12–20]

We recently reported a new type of interstrand DNA-DNA cross-link formed by reaction of an abasic site (Ap site) with a deoxyadenosine (dA) residue on the opposing strand of the double helix.[21] These cross-links form in high yields (15-70%) in 5'ApT/5'AA sequences. We proposed that formation of dA-Ap cross-links involved initial reaction of the N^6-amino group of deoxyadenosine with the aldehyde residue of the

Ap site (**1**) to yield an imine intermediate (**3**, Scheme 1). Precedents surrounding the reactions of arylamines with glycosides further suggested that intramolecular attack of the C4"-hydroxyl group on the putative imine intermediate **3** will generate the cyclic alkoxyhemiaminal **4**.[22,23] In our previous report, we showed that digestion of double-stranded DNA containing the dA-Ap cross-link with a mixture of nuclease P1, alkaline phosphatase, and phosphodiesterases I and I released a dinucleoside cross-link "remnant". However, the chemical connectivity of the dA-Ap cross-link was not rigorously established.

Scheme 1



In the work described here, we characterized the structure and properties of the dA-Ap cross-link remnant that is released by enzymatic digestion of cross-linked DNA. An authentic standard of the putative dinucleoside cross-link remnant **5** was synthesized by reaction of dA with 2-deoxyribose. Complete spectroscopic characterization of this

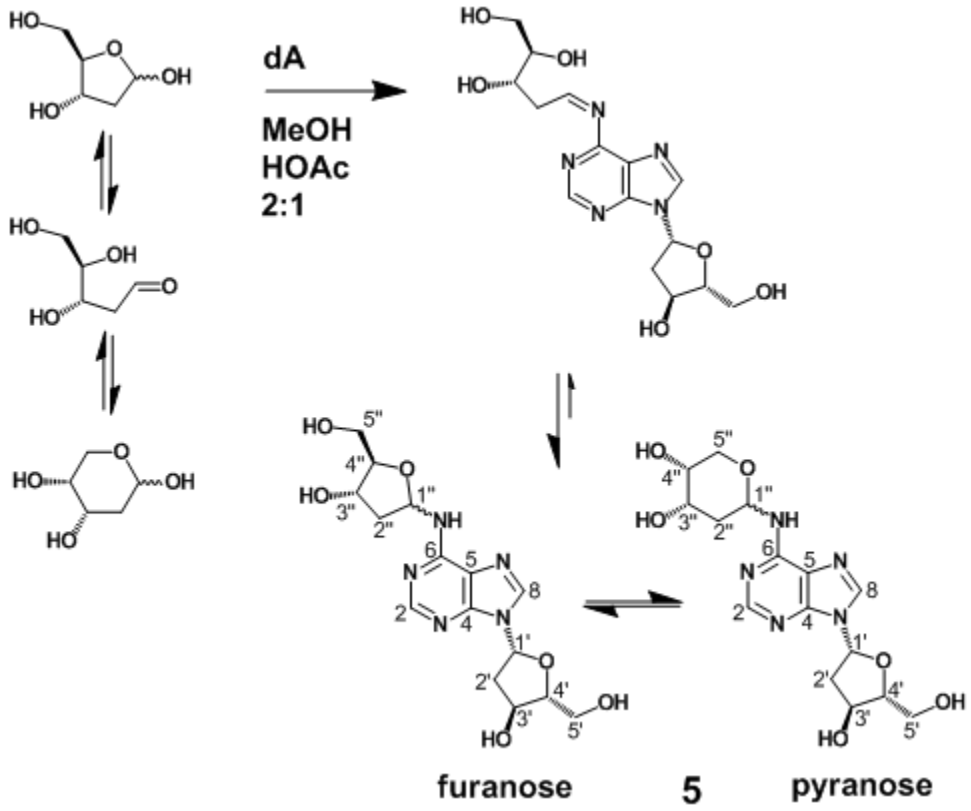
material established that the anomeric center of the 2-deoxyribose adduct was connected to dA at the exocyclic N^6-amino group and exists as a mixture of cyclic alkoxyhemiaminals (α and β anomers of both the furanose and pyranose isomers). LC-MS/MS analysis provided evidence that the synthetic compound **5** has the same chemical structure as the cross-link remnant released by enzymatic digestion of cross-linked DNA. Further experiments showed that the cross-link remnant **5** was quite stable at neutral pH, decomposing to release unmodified dA with a half-life of 65 d at 37 °C. The inherent chemical stability of the dA-dR cross-link remnant is reflected in the stability of the cross-link in duplex DNA. For example, there was no detectable dissociation of the dA-Ap cross-link in duplex DNA over the course of 96 h at pH 7 and 37 °C. The stability of the dA-Ap cross-link offers the possibility that this may be a persistent lesion with the potential to block the action of various DNA processing enzymes. Finally, the work lays a foundation for the detection of the dA-Ap cross-link in cellular DNA.

Results and Discussion

2. 2 Synthesis and Spectroscopic Characterization of the Cross-link Remnant 5

The putative cross-link remnant **5** was prepared by reaction of 2'-deoxyadenosine (0.6 M) with 2-deoxy-D-ribose (2.6 M) in a solvent mixture composed methanol/acetic acid 2:1 v/v (Scheme 2). Thin layer chromatographic analysis of the reaction mixture on silica gel revealed formation of a single new polar spot. This material was isolated by column chromatography on silica gel eluted with dichloromethane containing methanol (5-12% v/v) to give the product in 54% yield.

Scheme 2



1D-NMR analysis revealed resonances for both dA and an additional 2-deoxyribose unit in the product (Tables 1 and 2). The HMQC spectra allowed assignment of carbon-to-hydrogen connectivity. Multiple resonances were observed for the 2-deoxyribose adduct in the product. This was not unexpected, as 2-deoxyribose itself exists as an equilibrium mixture of the α-pyranose, β-pyranose, α-furanose, and β-furanose isomers (Scheme 2).[24] Similarly, *N*-aryl-2-deoxyaminoriboside analogs also exist as a mixture of cyclic forms, in which the pyranose isomers dominate. Thus, the observation of multiple resonances for the carbons and hydrogens of the 2-deoxyribose adduct in the product indicated that the product existed as a mixture of the four possible isomers (α-pyranose, β-pyranose, α-furanose, and β-furanose). An EXSY-NMR experiment that gives cross-peaks between sites that are in slow chemical exchange

Table 1

Position	Isomer	δ_c		δ_H (J in Hz)	HMBC	TOCSY	^{15}N HMBC
6	A	156.1					
	B	156					
	C	155.9					
2	A	154.9		8.26 0.29H, s	4, 6		
	B	154.8		8.25 0.55H, s	4, 6		
	C	154.8		8.25 0.08H, s	4, 6		
	D	154.7		8.24 0.08H, s	4, 6		
4		151.3					
8	A	143.5		8.29 0.63H, s	4		
	B	143.3		8.29 0.29H, s	4		
	C			8.29*			
	D			8.28 0.08H, s			
5	A	122.2					
	B	122.1					
	C	122.0					
4'		90.1		4.17 1H, m	1'	1', 3', 5'$_{a,b}$, 2'$_{a,b}$	
4''	A	88.6		4.10 0.08H, m		5''$_{A\text{-}a,b}$	
	B	87.9		4.02 0.08H, m	5''$_B$	5''$_{B\text{-}a,b}$	
	C	69.2		3.90 0.90H, m	2''$_D$	2''$_{D\text{-}a,b}$	
	D	69.3		3.90 *			
1'		87.4		6.43 1H, t(6.9)	4, 8, 4'	3', 4', 2'$_{a,b}$	
1''	A	84.5		6.28 0.08, brs		3''$_A$, 2''$_{A\text{-}a,b}$	
	B	ND		6.21 0.08H, brs		3''$_B$, 2''$_{B\text{-}a,b}$	
	C	80.3		5.48 0.55H, brs		2''$_{C\text{-}a,b}$	
	D	77.9		5.81 0.29H, brs		3''$_D$, 2''$_{D\text{-}a,b}$	
3''	A	74.3		4.47 0.08H, m		1''$_A$, 2''$_{A\text{-}a,b}$	
	B	73.8		4.44 0.08H, m		1''$_B$, 2''$_{B\text{-}a,b}$	
	C	70.3		4.07 0.55, m		1''$_C$, 4''$_C$, 2''$_{C\text{-}a,b}$	
	D	68.49		4.29 0.29H, m		1''$_D$, 5''$_{D\text{-}a,b}$, 2''$_{D\text{-}a,b}$	
3'		73.9		4.64 1H, m	1', 5'	4', 5'$_{a,b}$	
5''	C	69.6	a	3.94 0.55H, m	1''$_C$, 3''$_C$		
			b	3.79 0.55H, m	1''$_C$		
	D	66.1	a	3.84 0.29H, m	1''$_D$, 4''$_D$		
			b	3.76 0.29H, m	1''$_D$		
	A	64.7	a	3.70 0.08H, m		4''$_A$	
			b	3.65 0.08H, m		4''$_A$	
	B	63.9	a	3.69 0.08H, m		4''$_B$	
			b	3.64 0.08H, m		4''$_B$	
5'		64.4	a	3.83 1H, m	4', 3'	4', 3'	
			b	3.77 1H, m	4'	4'	
2'		41.7	a	2.80 1H, m	4', 1', 3'	1', 4'	
			b	2.55 1H, m	4', 3'	1', 4'	
2''	B	41.6	a	2.66 0.08H, m	4''$_B$, 3''$_B$	1''$_B$, 3''$_B$	
			b	2.16 0.08H, m	4''$_B$, 3''$_B$	1''$_B$, 3''$_B$	
	A	41.6	a	2.39 0.08H, m	3''$_A$	1''$_A$, 3''$_A$	
			b	2.33 0.08H, m	1''$_A$	1''$_A$, 3''$_A$	
	D	37.1	a	2.24 0.29H, m	3''$_D$	1''$_D$, 3''$_D$	
			b	2.08 0.29H, m	1''$_D$	1''$_D$, 3''$_D$	
	C	35.6	a	2.13 0.55H, m	4''$_C$, 3''$_C$	1''$_C$, 3''$_C$, 4''$_C$	
			b	2.07 0.55H, m	1''$_C$, 4''$_C$, 3''$_C$	1''$_C$, 3''$_C$, 4''$_C$	
N6	A	106.5					2''$_{A\text{-}b}$
	B	105					
	D	102					2''$_{D\text{-}b}$
	C	100.5					2''$_{C\text{-}a,b}$

* Overlapped by a larger peak

Table 1. NMR correlations for compound **5** taken in deuterium oxide.

Table 2

Position	Isomer	δ_c		δ_H (J in Hz)	HMBC	TOCSY	EXSY	NOSEY
6		153.8						
2		152.6		8.27 1H, brs	6			
4		149.4						
8		140.6		8.40 1H, s	4, 5, 1'			
5		120.2						
N6	A,C			8.15 0.46H, brs		$2''_{a,b}$, $1''_C$		
	B,D			*8.27 0.39H, brs		$2''_{a,b}$, $1''_{B,D}$		
4'		88.4		3.88 1H, m	1', 3', $2'_a$	1', 3'-OH, 5'-OH, 3', $5'_{a,b}$		
3''	A,C	86.9		3.68 0.4H, m	$1''_C$	$1''_C$, $2''_{a,b}$		
	B,D	67.9		3.85 0.5H, m				$1''_{B,D}$
			OH	5.09 1H		$2''_{a,b}$, $3''_D$		
1'		84.4		6.36 1H, t (6.9)	8, 4, 4', 3'	3', $2'_{a,b}$		
1''	A	84.2		6.26 0.1H, brs		$N6_A$, $2''_{a,b}$	$1''_C$	
	B	80.6		6.05 0.2, brs		$N6_B$, $3''_B$, $2''_{a,b}$	$1''_D$	
	C	79.4		5.83 0.3H, brs		$N6_A$, $4''_C$, $2''_{a,b}$	$1''_{A,D}$	
	D	75.5		5.63 0.4H, brs		$N6_B$, $3''_D$, $2''_{a,b}$	$1''_{B,C}$	
4''	D	71.7		4.16 0.4H, m		$2''_{a,b}$		
	A,B	67.3		4.09 0.3H, brs				
	C	66.9		3.97 0.3H, m	$3''_C$, $5''_{A-b}$, $2''_a$	$1''_C$		
			OH	4.62 1H	$4''_B$, $2''_a$	$4''_C$		
3'		71.3		4.40 1H, m	1', 4', $5'_a$	1', 3'-OH, 4', $5'_{a,b}$, $2'_{a,b}$		
			OH	5.30 1H	4', 3', $2'_a$	1', 3', 4', $2'_{a,b}$		
5''	D,B	65.7	a	3.48 0.6H, m	$4''_D$			
			b	3.41 0.6H, m	$1''_D$			$1''_{B,D}$
	C,A	64.9	a	**3.61 0.4H, m	$4''_A$, $5''_{A-b}$, $3''_A$			$1''_C$
			b	**3.51 0.4H, m	$4''_A$			
			OH	5.01 1H	$4''_A$			
5'		62.2	a	3.61 1H, m	4', 3'	5'-OH, 4'		
			b	3.52 1H, m	4', 3'	5'-OH, 4'		
			OH	5.14 1H		$5'_{a,b}$, 4'		
2'		40.1	a	2.71 1H,	1', 3'	1', 3', $2'_b$		
			b	2.27 1H,	4', 3'	1', 3', $2'_a$		
2''	A,B,C,D	37.1	a	2.02 1H, brm	$3''_D$	$N6_{A,B}$, $1''_{a,b,c,d}$, $3''_D$, $2''_b$		
		34.4	b	1.80 1H, m	$1''_D$, $3''_B$	$N6_{A,B}$, $1''_{a,b,c,d}$, $3''_D$, $2''_a$		

* Overlapped by another peak. See supporting information for D_2O exchange. **Overlapped by 5' signal

Table 2. NMR correlations for compound **5** taken in DMSO(d6).

confirmed that the multiple resonances reflected that the 2-deoxyribose adduct was present as an equilibrium mixture of isomers. For example, the H1" signals from 5.6-6.3 ppm are a dynamic isomeric mixture of four separate signals (Figure 1).

Figure 1

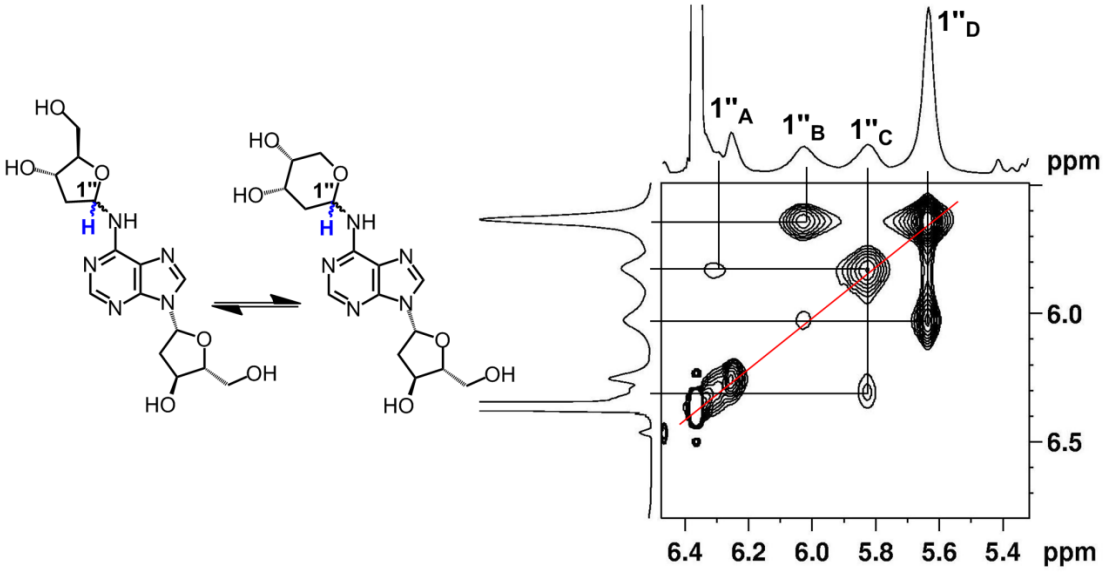


Figure 1. EXSY of the H1" region for compound **5**. The correlations between isomers of **5** can be seen by contours shown at the intersecting lines between signals for the H1" protons, indicating a dynamic equilibrium between all four isomers.

This product has the potential to exist as either an ring-opened imine or a cyclic hydroxyalkylhemiaminal; however, observation that the C1" chemical shifts fall in the range of 75-85 ppm was consistent with the cyclic structures **5**.[14] The C4" chemical shift of 69 ppm[14] and correlations between the C5" hydrogens and the C1" carbon in the HMBC experiment provided evidence that the major isomer is in the pyranose form. The NMR data provided no evidence for imine, carbinolamine, or enamine functional groups in the product. The major isomer showed a strong NOE cross-peak between the H1" and H3", consistent with the α-isomer.[25] The ratio of isomers in water was found to be

55:29:8:8 (Figure S10) which is similar to the ratio of isomers reported for unsubstituted 2'-deoxy sugars.[26]

Figure 2

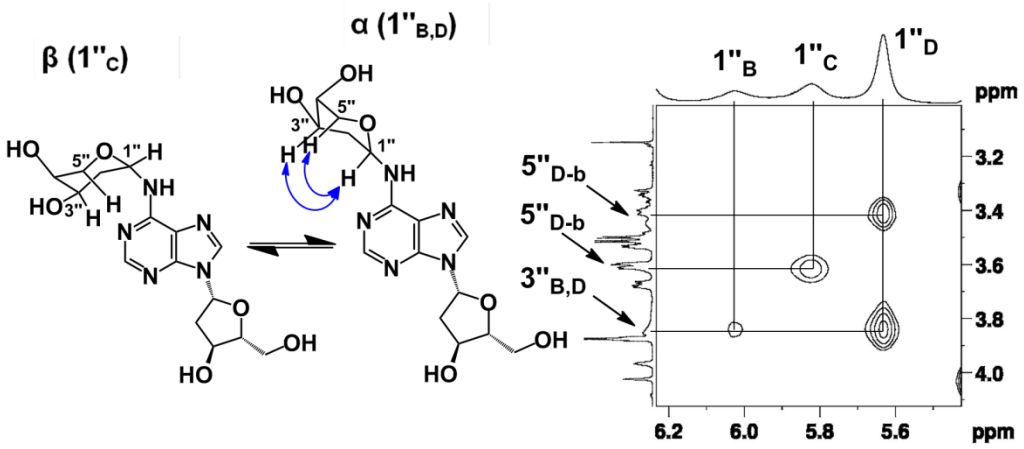


Figure 2. NOESY of the H1" region of compound **5**. The correlations between the 3" and 1" protons indicate that these are both axial and represent α-anomers in peaks 1" B and D while the 1" A and C peaks lack this correlation and represent β-anomers.

The NMR data provided evidence that the anomeric carbon of the 2-deoxyribose adduct was attached at the exocyclic N^6-amino group of the 2'-deoxyadenosine unit. For example, an ^{15}N-HMBC experiment revealed three-bond coupling between the N^6-nitrogen and the H2" protons of the 2-deoxyribose adduct (Figure 3). Furthermore, the TOCSY spectra showed that the N^6-H, the H1", and H2" protons reside in the same spin system (Figure 4 and S11). All other chemical shift assignments are shown in Tables 1 and 2 and were assigned using DEPT, HMQC, HMBC and TOCSY spectra (S2-S11). Overall the NMR analysis established that the reaction of dA with 2-deoxyribose provided the desired cross-link remnant **5**.

Figure 3

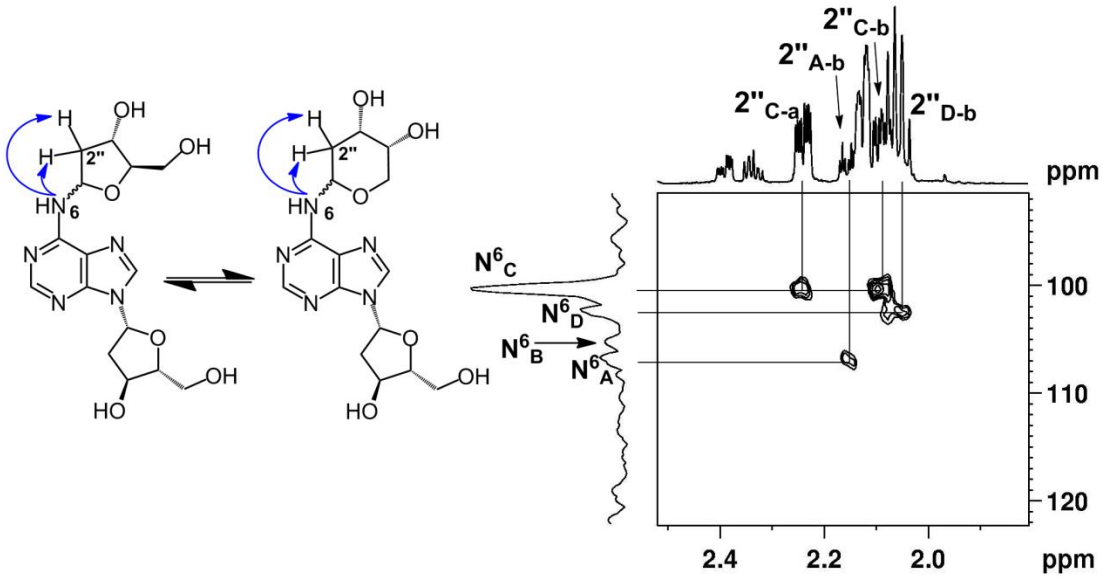


Figure 3. ^{15}N-HMBC of the H2" region of compound **5**. The correlations between the N^6 of dA and H2" of dR can be seen by contours shown at the intersecting lines between signals for the N^6 and H2" protons, indicating a 2-3 bond distance between these two atoms.

Figure 4

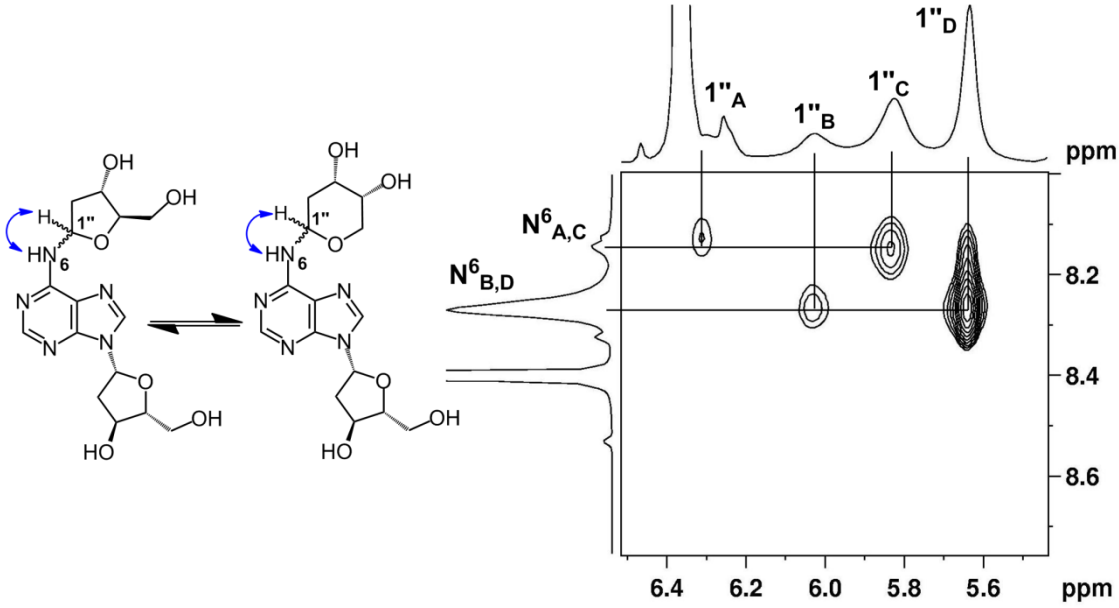


Figure 4. TOCSY of the N^6-H correlations with H1" protons for compound **5**. The correlations between the N^6-H of dA and H1" of dR can be seen by contours shown at the intersecting lines between signals for the N^6-H and H1" protons, indicating these protons are in the same spin network and not disrupted by the purine ring system.

2.3 Stability of the dA-dR cross-link remnant 5 and the dA-Ap cross-link in duplex DNA

Before endeavoring to use the synthetic cross-link remnant 5 as an analytical standard, we examined the chemical stability of this material. The structure 5 can be classified as an *N*-aryl glycosylamine. Compounds of this type can undergo hydrolytic decomposition to generate the free sugar and the unmodified arylamine.[27,28] We used HPLC analysis to examine the stability of 5 in sodium phosphate buffer (50 mM, pH 7) at both 37 °C and 75 °C. First, we noted that all four isomers of 5 detected in the NMR experiments were resolved in the reverse-phase HPLC chromatogram (Figure 5). We found that the synthetic cross-link remnant 5 was remarkably stable, decomposing to release unmodified dA with a half-life of 65 d at 37 °C and 3 d at 75 °C (Figure 6).

Figure 5

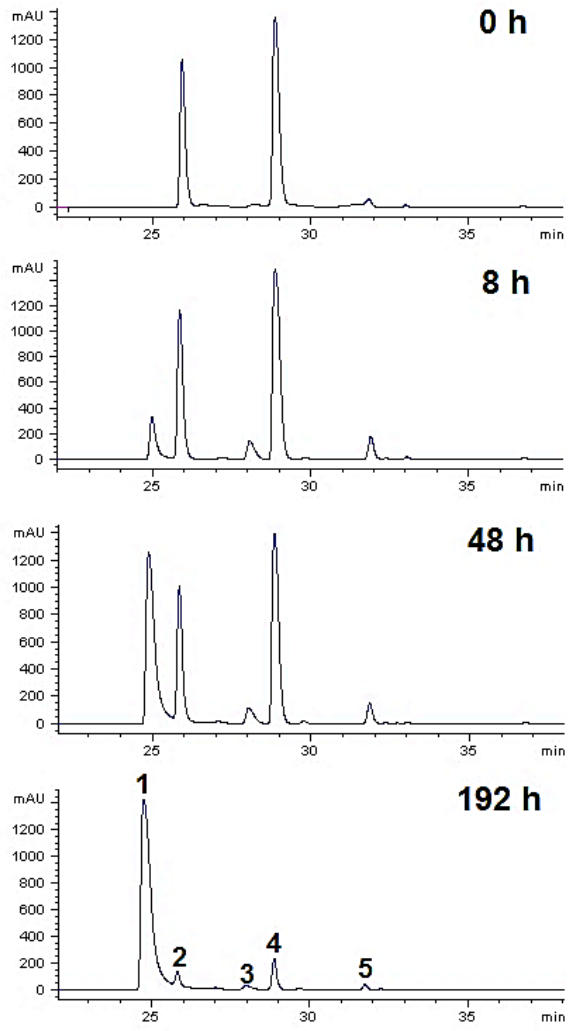


Figure 5. HPLC chromatogram panels for the dissociation of compound **5** at 75°C in 20 mM sodium phosphate (pH 7.0). Peak **1** is deoxyadenosine, peaks **2** and **4** are major isomers of compound **(5)** while peaks **3** and **5** are minor isomers of compound **(5)**. $k_{37°C}$ = $4.39^{-4} \pm 0.001^{-4}$ h^{-1}, $k_{75°C}$ = $6.47^{-3} \pm 0.0003^{-3}$ h^{-1}, E_a = 74.00 kJ/mol.

Figure 6

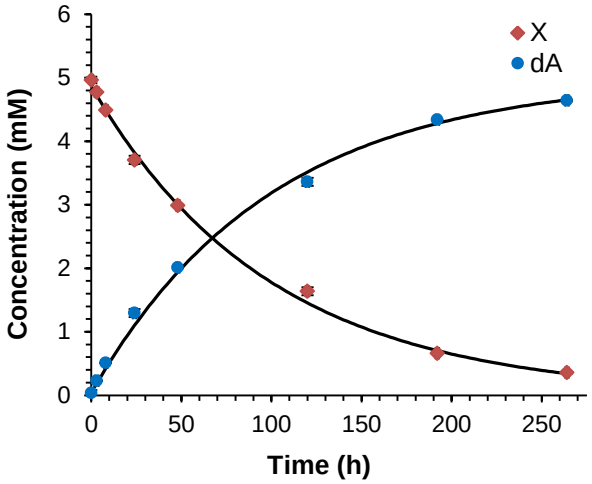


Figure 6. Plot of the concentration of compound **5** (red diamonds) at 75°C into dR and dA (blue circles) which was determined by the HPLC absorbance peaks (250 nm). Solid line represents the first order rate equation using $k_{75°C} = 6.47^{-3} \pm 0.0003^{-3}$ h^{-1} as the dissociation rate of compound **5** into dA and dR, which is the rate constant determined by non-linear regression analysis.

We also monitored the stability of the dA-Ap cross-link embedded within two DNA duplexes (Figure 7). The 5'-^{32}P-radiolabeled, cross-linked duplexes **A** and **B** were prepared as described previously,[21] gel purified, and then redissolved in HEPES buffer (50 mM, pH 7) containing NaCl (100 mM) at 37 °C. At various times, aliquots were removed, frozen, and later subjected to analysis by denaturing polyacrylamide gel electrophoresis. The cross-linked duplexes **A** and **B** are quite stable, displaying half-lives of 66 and 85 h, respectively (Figures S20 and 21). In a separate experiment, the amount of remaining cross-linked duplex **A** and **B** was monitored after incubation for 96 h at 22 °C in three different buffers with pH values of 7, 5 or 9. At pH 5 25% and 40% dissociation of cross-linked duplexes **A** and **B** into the component single strands was observed. No dissociation of the cross-linked duplexes **A** and **B** was observed over the course of 96 h at 22 °C at pH 7 or 9 (Figure S22).

Figure 7

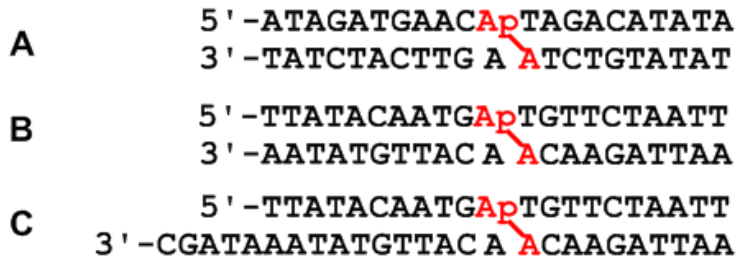


Figure 7. Oligonucleotide sequences used for enzymatic digestion and LC-MS/MS analysis and comparison of isomers within duplex DNA. The Ap-containing duplexes were generated from the corresponding dU-containing duplexes after treatment with the enzyme UDG. The major cross-linking sites are indicated with a (red \).

2.4 Evidence That Enzymatic Digestion of Duplex DNA Containing the dA-Ap Cross-link Releases the Cross-link Remnant 5

With a synthetic standard of the putative dA-Ap cross-link remnant in hand, we set out to use LC-MS/MS analysis to determine whether this material matched the material released from cross-linked DNA by enzymatic digestion.[21] We have previously reported the characterization of the dA-Ap cross-link in two different DNA duplexes **A** and **B**. These cross-linked duplexes (lacking the 5'-^{32}P-label) were prepared, gel purified and treated with a four enzyme cocktail of nuclease P1, phosphodiesterase I, alkaline phosphatase and phosphodiesterase II to release the cross-link remnant as previously described.[18] A separate control experiment showed that the synthetic standard **5** was stable under the digest conditions (Figure S18). The digested oligonucleotide mixtures and the synthetic material **5** were analyzed by LC-MS monitoring the *m/z* 368→252 transition and LC-MS/MS/MS monitoring the *m/z* 252→136 transition. The first transition is consistent with the loss of the 2-deoxyribose adduct from N^6-dA of the cross-link remnant, while the second transition corresponds to the loss of 2-deoxyribose from

the dA fragment. Again, multiple peaks displaying these transitions were observed consistent with an equilibrium mixture composed of the α-pyranose, β-pyranose, α-furanose, and β-furanose isomers. Importantly, the retention times and fragmentation pattern of the synthetic material **5** matched that of the remnant released by enzymatic digestion of the cross-linked duplexes **A** and **B** (Figure 8 and S19).

Figure 8

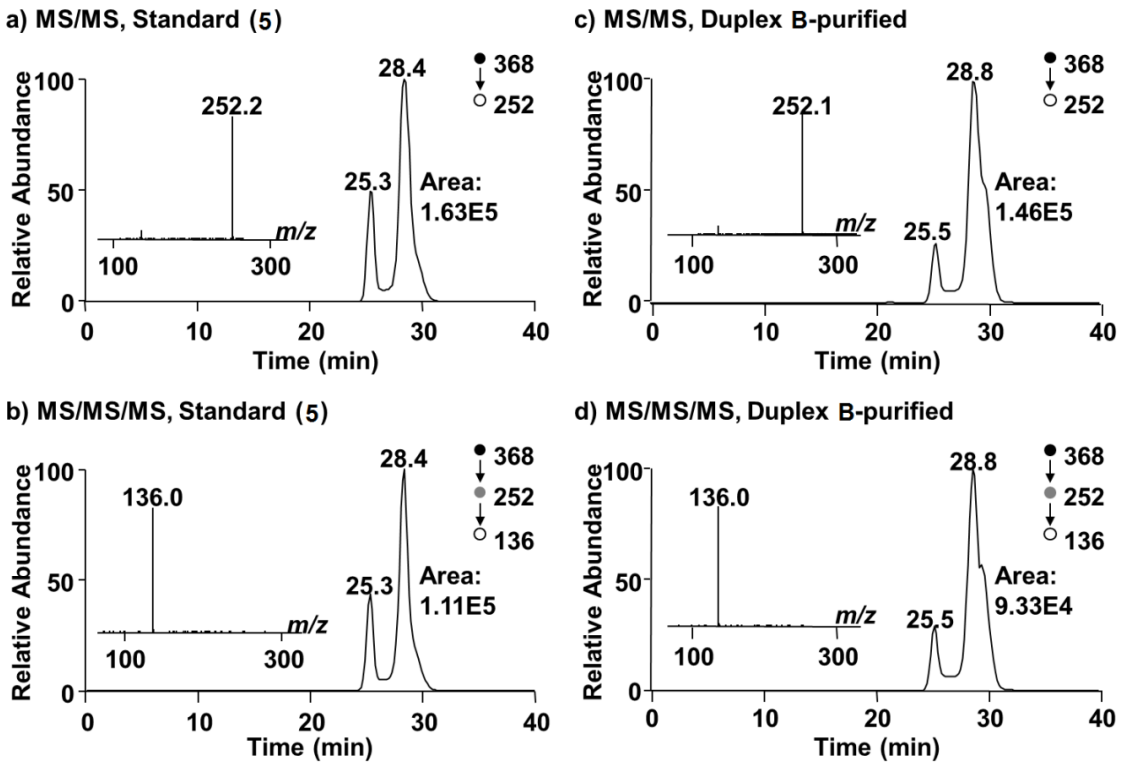
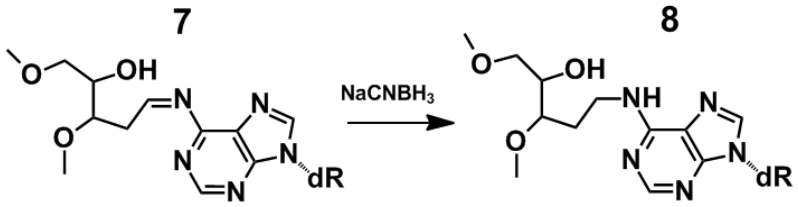


Figure 8. LC-MS/MS of the synthetic standard **5** and the enzymatically digested cross-linked duplex **B**. Panels A and B are for the synthetic nucleoside **5**, while panels C and D are for duplex **B**. Panels A and C are the chromatograms obtained for the first transition of 368→252. Panels B and D are the chromatograms obtained for the second transition of 252→136.

2.5 Chemical Reduction of the Cross-link Remnant 5

As shown in Schemes 1 and 2, the dA-Ap cross-link is an imine-derived lesion. While the NMR analysis above provided evidence that the synthetic cross-link remnant **5** exists predominantly in the ring-closed form, this material presumably exists in equilibrium with the small amounts of the imine isomer. Hydrolytic decomposition of **5** likely proceeds via attack of water on the presumed imine intermediate. Other imine and cyclic imine-derived adducts have been stabilized by hydride reduction with reagents such as sodium cyanoborohydride,[13,29,30] so we examined the reduction of the cross-link remnant **5**. Interestingly, we were not able to detect reduction of the nucleoside **5** after using common water-compatible hydride donors (sodium borohydride, sodium cyanoborohydride or sodium triacetoxyborohydride). Various solvent mixtures (methanol, water and acetic acid), pH values (3-7), and amounts of hydride donor (10-1000 equiv) and no new product could be detected by TLC, NMR or HPLC analysis. It is possible that, with this material in the stable pyranose form, there is not enough of the imine **5** available for reduction. This idea is in line with the observation that the amount of free, ring-opened aldehyde can dictate the reactivity of aldoses.[31] To probe this idea we synthesized an analog **7** of the cross-link remnant that cannot access the pyranose isomer due to protection of the 5' and 3'-hydroxyl groups of the 2-deoxyribose adduct. (Figures S12-14). We found that compound **7** (0.05 mM) can be converted to the reduced form **8** in 40% yield in acetic acid (0.12 mM) in dry methanol at 40°C in the presence of $NaBH_4$ (4.8 mM) for 68 h (Scheme 3, and Figures S15-17). Slow reductive amination has been observed for other imine adducts in DNA, including some formaldehyde adducts with dA and also Fappy-dG lesions.[14,32]

Scheme 3



2.6 Isomers of dA-Ap cross-links within duplex DNA

During PAGE analysis of dA-Ap cross-links within duplex B two closely migrating bands were seen in the cross-linking region. It is known that bent duplex structures, due to A/T tracts in kinetoplast DNA, can lead to noticeable changes in gel mobility.[33–35] Therefore, it is reasonable that isomers of cross-linked DNA may present with a gel mobility shift as well and because multiple isomers were observed in the dinucleotide remnant described above we believed there was a possibility that isomers also exist when the cross-link exist in a duplex. To study the putative isomers we used duplex **C** with a 3'-tail on the complementary strand, which enhanced the difference in gel mobility for the two bands. After separating the two bands by PAGE they were excised, desalted and resuspended in buffer. After resuspension aliquots were removed, frozen and later separated by a second PAGE (Figure 9). The second gel analysis showed no single strand or cleavage bands and after 24 hours both of the isolated bands had re-equilibrated to the same ratios (Figure 10). This experiment demonstrates the two bands can quickly re-equilibrated without completely dissociating into single stranded oligos, which is consistent with the existence of isomeric forms of the same lesion. Based on previous stability studies the two isomers are likely α/β-anomers because isomers with imine or carbinolamine structures would decompose quickly in water. To probe the structure further, altered pH and enzymatic treatments were introduced to the cross-

linking bands (Figure 11). Endonuclease IV was used on the cross-link mixture because it is an enzyme shown to remove α-nucleotides.[36–39] The results indicate there is a pH dependence of the ratio of isomers, but they cannot be processed by Endo IV. A lack of processivity by Endo IV may not be surprising, as the lesion is actually an interstrand cross-link and would not be recognized as a simple α-nucleotide. Overall, the stability and a fast re-equilibration which does not require dissociation to single stranded oligos is consistent with α/β-anomers of an interstrand cyclic-hydroxyalkylhemiaminal cross-link, which is the same structure of the synthetic dinucleotide remnant **5**.

Figure 9

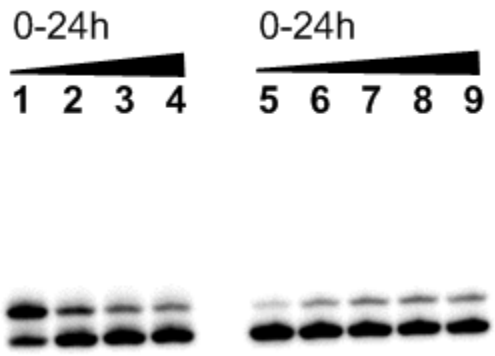


Figure 9. Isolation and re-equilibration of the top and bottom isomers from duplex **C**. After isolation the isomers were resuspended in 50 mM HEPES (pH 7.0) and 100 mM NaCl at 37°C. After resuspending in buffer aliquots were removed at 0, 4, 8 and 24 hours for the top isomer and 0, 2, 4, 8 and 24 hours for the top isomer and separated by PAGE.

Figure 10

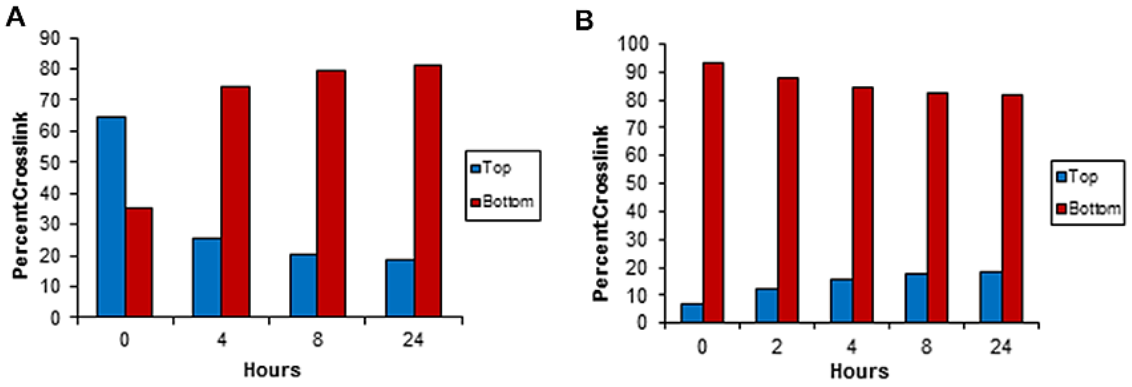


Figure 10. Isolation and re-equilibration of the top and bottom isomers from duplex **C**. After isolation the isomers were resuspended in 50 mM HEPES (pH 7.0) and 100 mM NaCl at 37°C. After resuspending in buffer aliquots were removed at 0, 4, 8 and 24 hours for the top isomer and 0, 2, 4, 8 and 24 hours for the top isomer. After 24 hours both isolated bands have re-equilibrated to the same ratio of double bands.

Figure 11

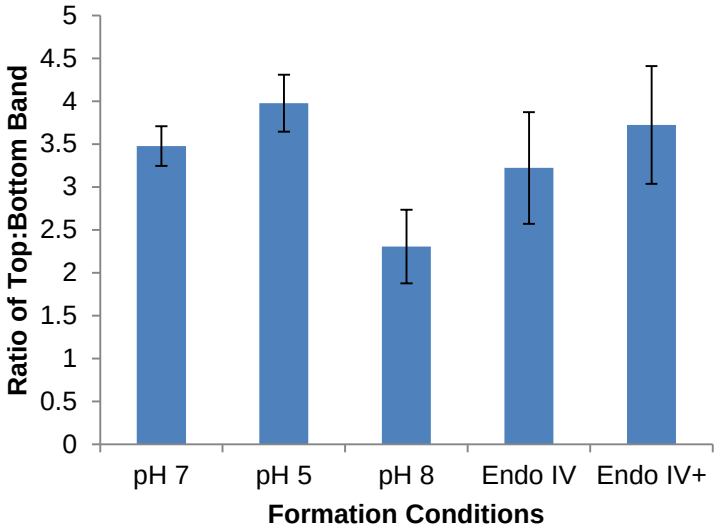


Figure 11. Formation and treatment of the putative isomers at varied pH and with Endonuclease IV. Samples were incubated at 37°C with 100 mM NaCl for 120 hours in either 16.5 mM sodium phosphate/ 1.8 mM citrate (pH 7.0), 10.3 mM sodium phosphate/ 4.9 mM citrate (pH 5.0) or in 19.5 mM sodium phosphate/ 0.3 mM citrate (pH 8.0). Additionally pH 7.0 incubations were treated with either 20 unites of endonuclease for 1 hour (Endo IV) or 40 units for 2 hours (Endo IV+).

2.7 Conclusions

Here we have described the synthesis of the cross-link remnant **5**. Complete structural characterization and stability of **5** were obtained through various spectroscopic methods. EXSY shows a dynamic mixture of four isomers, which is consistent with the expected pyranose/furanose cyclic hydroxyalkylhemiaminal anomers. Additionally, ^{15}N-HMBC and TOCSY data indicated the site of dR-dA attachment is between the $C^{1'}$ of dR with the N^6-H of dA. HPLC traces show **5** to be remarkably stable in buffered solution, even at temperatures as high as 75°C. Analysis by LC-MS indicates that enzymatically digested DNA will afford the same structure as the synthetic compound **5**. PAGE analysis of the isomers from within duplex DNA indicate that it exist in a similar cyclic-hydroxyalkylhemiaminal structure as that of **5.** Knowledge about the exact structure of this cross-link will help explain its biological function including toxicity and the ability of a cell to repair this type of damage. The stability of **5** will make future isolation and detection easier, as harsher conditions and longer experimental run times will be available while still leaving intact cross-link, even from *in vivo* systems. It will be important to detect native levels of dA-Ap cross-links from repair proficient/deficient cells and also study the repair process for these lesions as these steps could lead to understanding of their persistence and mutagenic capabilities within living organisms.

2.8 Materials and Methods

Oligonucleotides were purchased from Integrated DNA Technologies. Uracil-DNA Glycosylase (5000 units/ml) and UDG buffer were purchased from New England Biolabs. 2'-deoxyadenosine (dA) monohydrate was purchased from Tokyo Chemical Industry Co.,

Ltd. 2-deoxy-D-ribose (99%) was purchased from Sigma-Aldrich. All other reagents used were purchased from Acros Organics in reagent grade. Methanol, Dichloromethane, and H_2O were purchased from Fisher in HPLC grade. Column chromatography was performed on silica gel 60 (Sigma-Aldrich) under positive pressure. Glass-backed TLC plates with a 254-nm fluorescent indicator were purchased from Sigma-Aldrich and stored over desiccant. Developed TLC plates were visualized under a 254-nm UV lamp or by TLC staining with a 1% $KMnO_4$ solution. ^{1}H and ^{13}C NMR spectra along with the corresponding 2D experiments were generated on a Bruker Advance 800 MHz at 298°K using DMSO(d6) or D_2O (Cambridge Isotope Laboratories, Inc.) as solvents. UV-absorbance data was collected using a Hewlett Packard spectrophotometer. HPLC data was gathered using an Agilent 1100 series with binary pump, auto-injector and photo-diode array. Separations were achieved with a Varian Microsorb-MV C18 (250 x 4.6 mm) column. HPLC data analysis was performed by Chemstation for LC software Rev. A.08.03 [847] 1990-2000, followed by manual peak selection.

Preparation of Cross-linked DNA Duplexes. 2'-deoxyoligonucleotides were prepared by standard procedures for annealing, abasic site formation and precipitation. Duplexes **A** and **B** (40 µl each at 100 pmol/µl) were individually combined with H_2O, NaCl (100 mM) and MOPS buffer (25 mM, pH 7) to give three 150-µl solutions. The duplexes were annealed to 3 molar equivalents of their respective complement by incubation at 95°C for 5 minutes, followed by cooling to room temperature overnight in an insulated container. Ap sites were generated by incubation at 37°C for 2 hours in the presence of UDG enzyme at a total concentration of 400 units/ml (30 µl UDG buffer, 96 µl H_2O, and 4 µl UDG added to solutions of annealed **A** and **B**). Following incubation, UDG enzyme was

removed by phenol/chloroform extraction. The aqueous layers were pipetted off and mixed with 10% aqueous volume of 3M sodium acetate buffer (pH 5.2) and followed by 5 volume equivalents of absolute ethanol, cooled in dry ice (20 min.), and centrifuged for 30 minutes at 5°C to precipitate the DNA. The pellets were washed twice with 180 µl 80% ethanol in H_2O and then resuspended in 100 µl H_2O. Cross-link formation was achieved by incubation at 37°C for 120 hours in a buffered solution of HEPES (50 mM, pH 7.0) containing NaCl (100 mM). After the incubation time all samples were precipitated as previously described and loaded on a 2mm thick, 20% denaturing polyacrylamide gel which was ran at 200V for 4 hours. The slow moving cross-link band was visualized by UV-shadowing and excised from the gel. Oligonucleotide was eluted from the gel slice using 1 ml of a 200 mM NaCl, 1 mM EDTA solution (pH 8) and vortexing for 1 hour. The eluted sample was filtered, precipitated once again, as previously mentioned, and stored at -20°C.

Enzymatic digestion. Duplex **A** and **B** were digested using a 4-enzyme cocktail following the conditions described previously. Briefly, nuclease P1 (5 U), phosphodiesterase 2 (0.01 U), EHNA (20 nmol) and a 50-µL solution containing 300 mM sodium acetate (pH 5.6) and 10 mM zinc chloride were added to 300 pmol of duplex H in a final volume of 500 µL. In this context, EHNA served as an inhibitor for deamination of 2'-deoxyadenosine (dA) to 2'-deoxyinosine (dI) induced by adenine deaminase. The resulting mixture was incubated at 37˚C for 48 h. To the digestion mixture were then added alkaline phosphatase (10 U), phosphodiesterase 1 (0.005 U) and 100 µL of 0.5 M Tris-HCl buffer (pH 8.9). The digestion was continued at 37˚C for 2 h. Digestion with nuclease P1 alone was performed following previously published procedures, where 0.2

unit of nuclease P1 (0.2 U) was added to 300 pmol of duplex H in a 500 µL water solution. The resulting mixture was incubated at 37˚C for 2 h. The above enzymatic digestion mixture was extracted with chloroform to remove enzymes, and the aqueous layer was dried on a Speed-vac, reconstituted in water, and subjected to LC-MS/MS analyses.

N^6-(2-deoxy-D-ribos-1-yl)-2'-deoxyadenosine (5). 2'-deoxyadenosine monohydrate (154 mg, 0.57 mmol) and 2'-deoxy-D-ribose (351 mg, 2.62 mmol) were dissolved in 300 µl glacial acetic acid and 700 µl methanol. The reaction was left in a 37°C oven for 72 hours. The resulting mixture was dried under reduced pressure and purified by column chromatography; using a gradient from 5-12% methanol in dichloromethane. The fractions containing product were combined, dried under reduced pressure and subjected to three cycles of adding 2 ml chloroform and rotary evaporation to yield 115 mg (54% yield) of **5** as a white powder. ^{1}H NMR (800 MHz, D_2O) δ 8.29 (0.63H, s, H8), 8.29 (0.29 H, s, H8), 8.28 (0.08H, s, H8), 8.26 (0.29H, s, H2), 8.25 (0.55H, s, H2), 8.25 (0.08H, s, H2), 8.24 (0.08H, s, H2), 6.43 (1H, t, J=6.9 Hz, dR1), 6.28 (0.08H, brs, H1''), 6.21 (0.08H, brs, H1''), 5.81 (0.29H, brs, H1''), 5.48 (0.55H, brs, H1''), 4.64 (1H, m, H3'), 4.47 (0.08H, m, H3''), 4.44 (0.08H, m, H3''), 4.29 (0.29H, m, H3''), 4.17 (1H, m, H4'), 4.10 (0.08H, m, H4''), 4.07 (0.55H, m, H3''), 4.02 (0.08H, m, H4''), 3.94 (0.55H, m, H5''), 3.90 (0.84H, m, H4''), 3.84 (0.29H, m, H5''), 3.83 (1H, m, H5'), 3.79 (0.55H, m, H5''), 3.77 (1H, m, H5'), 3.76 (0.29H, m, H5''), 3.70 (0.08H, m, H5''), 3.69 (0.08H, m, H5''), 3.65 (0.08H, m, H5''), 3.64 (0.08H, m, H5''), 2.80 (1H, m, H2'), 2.66 (0.08H, m, H2''), 2.55 (1H, m, H2'), 2.39 (0.08H, m, H2''), 2.33 (0.08H, m, H2''), 2.24 (0.29H, m, H2''), 2.16 (0.08H, m, H2''), 2.13 (0.55H, m, H2''), 2.08 (0.29H, m, H2''), 2.07

(0.55H, m, H2''). ^{13}C NMR (200 MHz, D_2O) δ 156.1 (C6), 156.0 (C6), 155.9 (C6), 154.9 (C2), 154.8 (C2), 154.8 (C2), 154.7 (C2), 151.3 (C4), 143.5 (C8), 143.3 (C8), 122.2 (C5), 122.1 (C5), 122.0 (C5), 90.1 (C4'), 90.1 (C4'), 88.6 (C4''), 87.9 (C4''), 87.4 (C1'), 87.3 (C1'), 88.6 (C1''), 80.3 (C1''), 77.9 (C1''), 74.3 (C3''), 73.9 (C3'), 73.8 (C3''), 70.3 (C3''), 69.6 (C5''), 69.3 (C4''), 69.2 (C4''), 68.5 (C3''), 66.1 (C5''), 64.7 (C5''), 64.4 (C5'), 63.9 (C5''), 41.7 (C2''), 41.7 (C2'), 41.6 (C2''), 37.1 (C2''), 35.6 (C2'').

3,5-bis-*O*-methyl-2-deoxy-D-ribofuranose (6). 2.00 g (14.9 mmol) 2-deoxy-D-ribose was dissolved in 80 ml methanol and cooled to 0°C. Concentrated HCl (40 µl; 0.05% v/v) was added to the flask by syringe and the mixture was stirred for 1 hour while warming to room temperature, after which the starting material had disappeared on TLC. The reaction was quenched by the addition of 300 mg (1.1 mmol) Ag_2CO_3 and the resulting precipitate was removed by filtration. Evaporation of the filtrate gave a yellow oil, which was redissolved in 32 ml freshly distilled THF. The solution was cooled to 0°C before gradual addition of 1.60 g (40 mmol) NaH (60% emulsion) in 4 portions, followed by injection of 2.5 ml (40 mmol) MeI in 250-µl portions. After stirring for 2 hours at ambient temperature, the reaction was stopped by the addition of 5 ml methanol and decantation into 50 ml ice water. The aqueous mixture was extracted with CH_2Cl_2 (3 x 35 ml), and the extract washed with H_2O (50 ml), dried (Na_2SO_4) and evaporated to a yellow oil. The product was purified by column chromatography (4:1 hexane/ethyl acetate; R_f = 0.39, 0.27) to give a colorless liquid. The liquid was dissolved in 50 ml 0.01M HCl (aq.) and refluxed for 1 hour, after which the crude was neutralized with 1.5 g (5.4 mmol) Ag_2CO_3, filtered and evaporated to a grey residue. The residue was triturated in CH_2Cl_2,

then filtered again to remove residual solids. Evaporation of the filtrate gave a yellow oil, which was column chromatographed (1:1 hexane/ethyl acetate; R_f = 0.25) to afford 1.49 g (62% overall) of **6** as a colorless oil: ^{1}H NMR (500 MHz, $CDCl_3$) δ 5.53-5.48 (0.44H, m, H1), 5.46-5.41 (0.56H, m, H1), 4.29 (0.56H, ddd, J = 4.8, 4.8, 1.8, H4), 4.20 (0.44H, d, J = 7.5, C1-OH), 4.06 (0.44H, q, J = 3.8, H4), 3.97-3.92 (0.44H, m, H3), 3.83-3.74 (0.88H, m, C1-OH, H3), 3.48 (0.56H, dd, J = 10.8, 3.8, H5a), 3.44 (0.56H, dd, J = 9.8, 4.3, H5b), 3.39-3.31 (2.24H, m, H5a, *O*-C$\underline{H}_3$), 3.31-3.26 (3.56H, m, H5b, *O*-C$\underline{H}_3$), 3.23 (1.32H, s *O*-C$\underline{H}_3$), 2.18-2.08 (0.88H, m, H2), 2.07-2.01 (1.12H, m, H2)); ^{13}C NMR (126 MHz, $CDCl_3$) δ 99.0 (C1), 98.9 (C1), 82.4 (C4), 82.0 (C3), 81.8 (C4), 81.7 (C3), 74.1 (C5), 72.9 (C5), 59.2, 59.1, 56.9, 56.7 (*O*-$\underline{C}H_3$), 40.7, 38.5 (C2).

***N*^{6}-[3,5-bis-*O*-methyl-2-deoxy-D-ribofuranos-1-yl]-2'-deoxyadenosine (7).** 100 mg 2'-deoxyadenosine (0.40 mmol) and 263 mg **1** (1.62 mmol) were dissolved in 2 mL of a 1:1 mixture of 25mM ammonium phosphate buffer (pH 7) and DMSO. The mixture was stirred at 37°C for 9 days, then evaporated to a yellow oil. Purification by column chromatography (0-5% MeOH in CH_2Cl_2) gave 22 mg (14%) of **7** as a colorless oil: ^{1}H NMR (500 MHz, DMSO-d_6) δ 8.45-8.39 (1H, m, H8), 8.35-8.00 (1.5H, m, H2, N^2-H), 7.55 (0.5H, brs, N^2-H), 6.40-6.33 (1H, m, H1'), 6.24 (1H, brs, H1''), 5.35 (1H, brs, 3'-OH), 5.23-5.08 (1H, m, 5'-OH), 4.41 (1H, brs, H3'), 4.07 (0.5H, q, J = 4.3, H4''), 3.94-3.89 (0.5H, m, H4''), 3.89-3.85 (1H, m, H4'), 3.85-3.81 (1H, m, H3''), 3.64-3.58 (1H, m, H5a'), 3.55-3.48 (1H, m, H5b'), 3.36-3.30 (3.5H, m, H5'', *O*-C$\underline{H}_3$), 3.26 (1.5H, s, *O*-C$\underline{H}_3$), 3.25 (1.5H, s, *O*-C$\underline{H}_3$), 3.24 (1.5H, s, *O*-C$\underline{H}_3$), 2.71 (1H, ddd, J = 13.4, 6.6, 6.6, H2a'), 2.43-2.33 (0.5H, m, H2a''), 2.33-2.22 (1.5H, m, H2b', H2a''), 2.20-2.08 (1H, m, H2b''); ^{13}C NMR (126 MHz, DMSO-d_6) δ 153.5 (C6), 152.1 (C2), 149.1 (C4), 140.3,

140.1 (C8), 119.6 (C5), 88.0 (C4'), 83.8 (C1'), 81.7 (C3''), 81.5 (C3''), 80.9 (C4''), 80.6 (C4'', C1''), 73.3, 72.8 (C5''), 70.8 (C3'), 61.8, 61.7 (C5'), 58.5, 56.8, 56.0 (*O*-$\underline{C}H_3$), 39.4, (C2'), 35.8 (C2'').

N^6-[1-deoxo-2-deoxy-3,5-bis-*O*-methyl-D-ribofuranos-1-yl]-2'-deoxyadenosine (8). Compound **7** (18 mg, 0.05 mmol), $NaBH_3CN$ (100 mg, 1.6mmol) and acetic acid (25 ul, 0.4 mmol) were dissolved in 5 ml dry methanol. The mixture was warmed to 40°C and stirred for 48 hr, after which additional portions of $NaBH_3CN$ and acetic acid were added (200 mg and 50 ul, respectively). After a total time of 68 hr, the reaction was quenched by addition of 5 ml sat. aq. Na_2CO_3. The suspension was evaporated to dryness, redissolved in methanol, then filtered and evaporated again to give a colorless oil. The crude oil was purified by column chromatography, eluting first with 100% ethyl acetate (R_f = 0.00), then with 10% methanol in CH_2Cl_2 (R_f = 0.21) to afford 8 mg (44%) of **8** as a colorless oil: ^{1}H NMR (500 MHz, DMSO-d_6) δ 8.32 (1H, s, H8), 8.20 (1H, s, H2), 7.80 (1H, brs, *N*-H), 6.34 (1H, dd, J = 7.5, 6.5, H1'), 5.31 (1H, d, J = 3, 3'-OH), 5.25 (1H, t, J = 5.25, 5'-OH), 4.81 (1H, brs, 4''-OH), 4.43-4.38 (1H, m, H3'), 3.88 (1H, q, J = 3.3, H4'), 3.67-3.59 (2H, m, H4'', H5a'), 3.59-3.48 (3H, m H1'', H5b'), 3.37-3.31 (1H, m, H5a''), 3.29 (3H, s, *O*-C$\underline{H}_3$), 3.28-3.25 (1H, m, H5b''), 3.23 (3H, s, *O*-C$\underline{H}_3$) 3.22-3.18 (1H, m, H3''), 2.72 (1H, ddd, J = 13.1, 6.6, 6.6, H2a'), 2.25 (1H, ddd, J = 13.1, 6.1, 2.6, H2b'), 1.90-1.81 (1H, m, H2a''), 1.77-1.66 (1H, m, H2b''); ^{13}C NMR (126 MHz, DMSO-d_6) δ 154.6 (C6), 152.4 (C2), 148.0 (C4), 139.3 (C8), 119.7 (C5), 88.0 (C4'), 84.0 (C1'), 79.8 (C3''), 74.0 (C5''), 71.0 (C3'), 70.2 (C4''), 61.9 (C5'), 58.3, 57.3 (*O*-$\underline{C}H_3$), 39.4 (C2'), 36.8 (C1''), 29.6 (C2'').

HPLC analysis for the dissociation of 5. 8 ml of a 5 mM solution of compound **5** was prepared in 100 mM phosphate buffer (pH 7.0). The solution was left at 37°C and 300 µL aliquots were taken out a specific time points to be analyzed by HPLC. Solvents used were water (A) and acetonitrile (B). The following method was used; from 0-5 min 0% B, 5-50 min a gradient of 0-18% B, 50-55 min a gradient of 18-0% B and from 55-60 min 0% B. The eluent was monitored at 250 nm and the absorbance peaks were integrated and reported by the Chemstation software.

2.9 Supporting Information

Figure S1

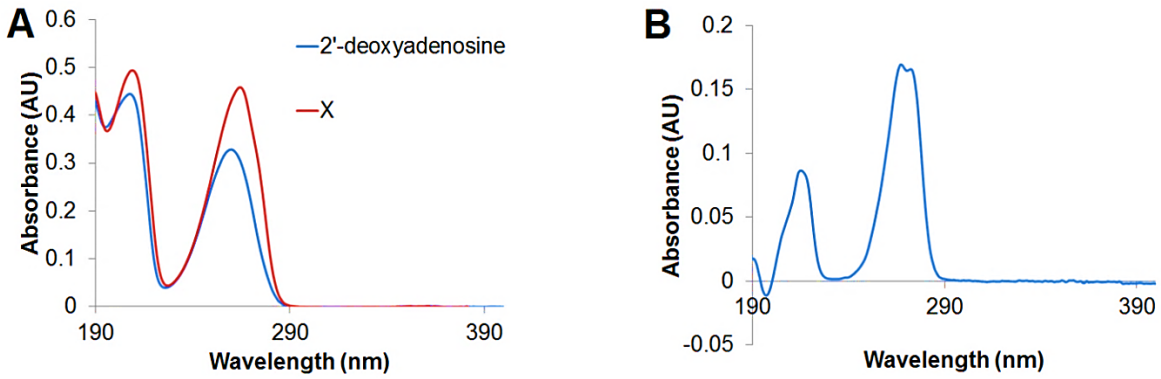


Figure S1. Panel A: UV absorption spectra for 25 µM 2'-deoxyadenosine and 25 µM of compound **5** taken in water. Panel B: difference spectra subtracting the spectra of dA from **5**. Molar absorptivity calculated for 260 nm: ε_{dA} = 13,200 L/(mol·cm), ε_5 = 18,400 L/(mol·cm).

Figure S2

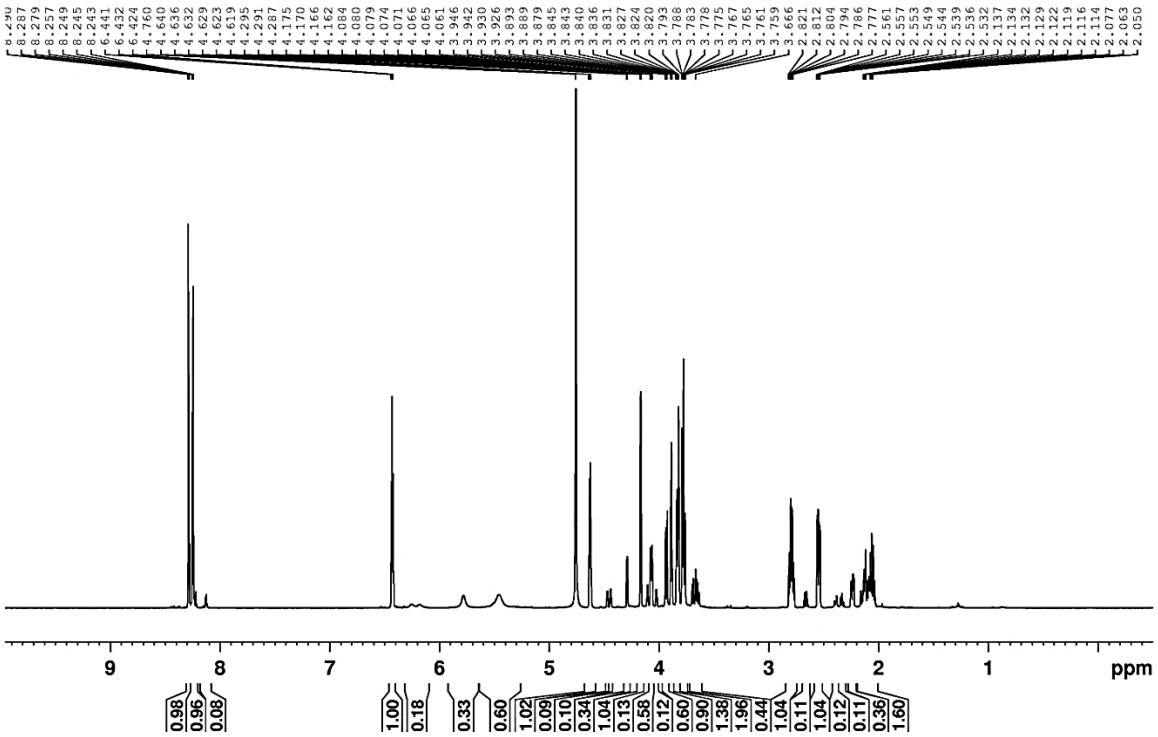


Figure S2. ^{1}H-NMR of compound **5** taken in deuterium oxide at 800 MHz.

Figure S3

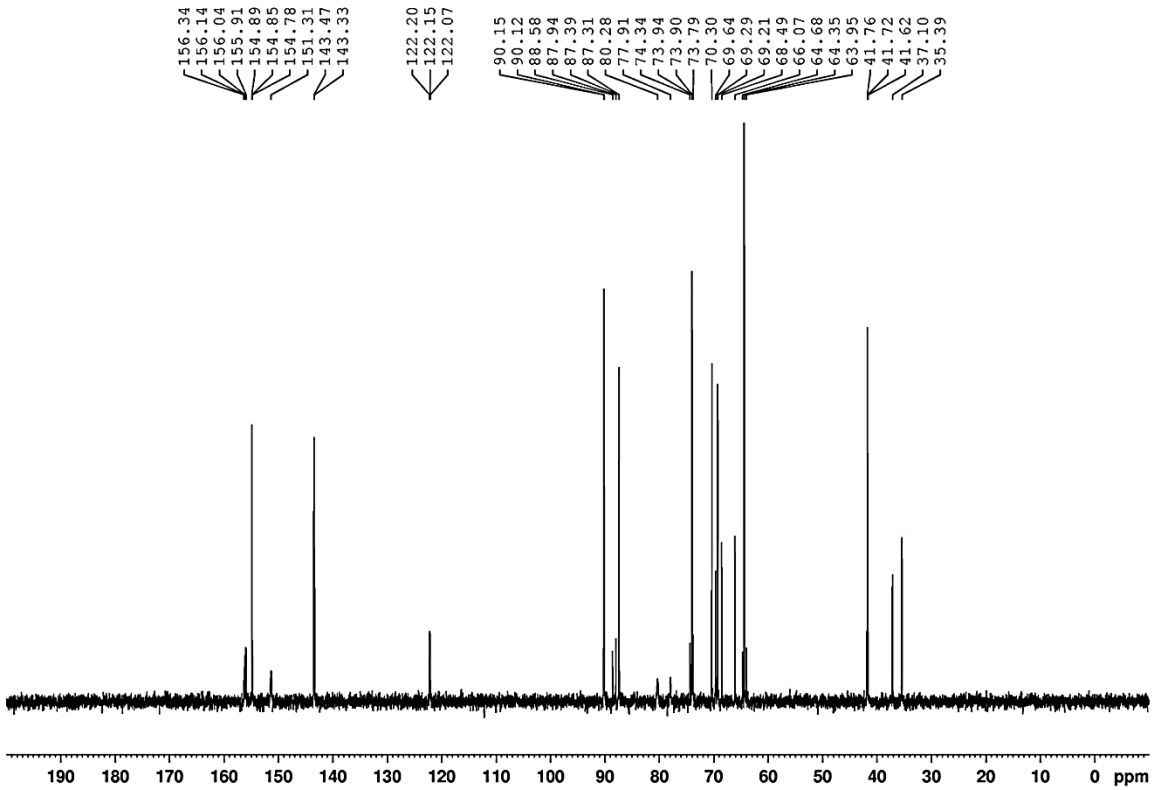


Figure S3. ^{13}C-NMR of compound **5** taken in deuterium oxide at 200 MHz.

Figure S4

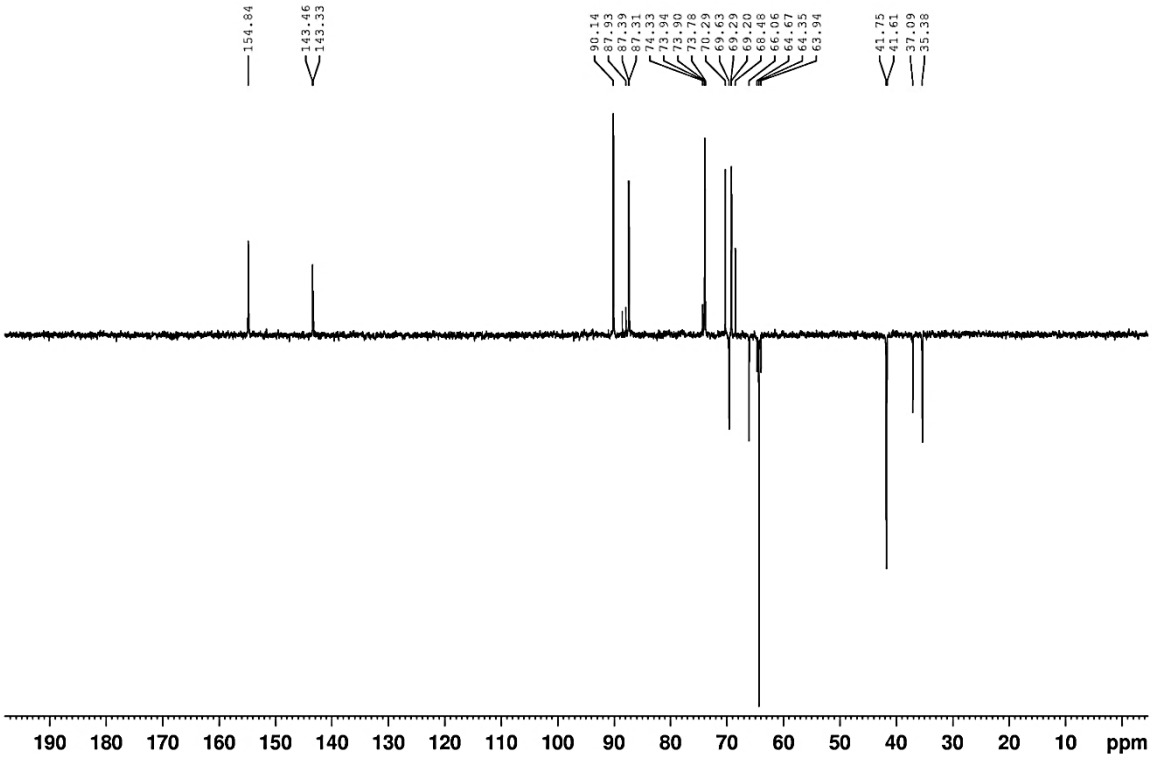


Figure S4. ^{13}C DEPT-NMR of compound **5** taken in deuterium oxide at 200 MHz.

Figure S5

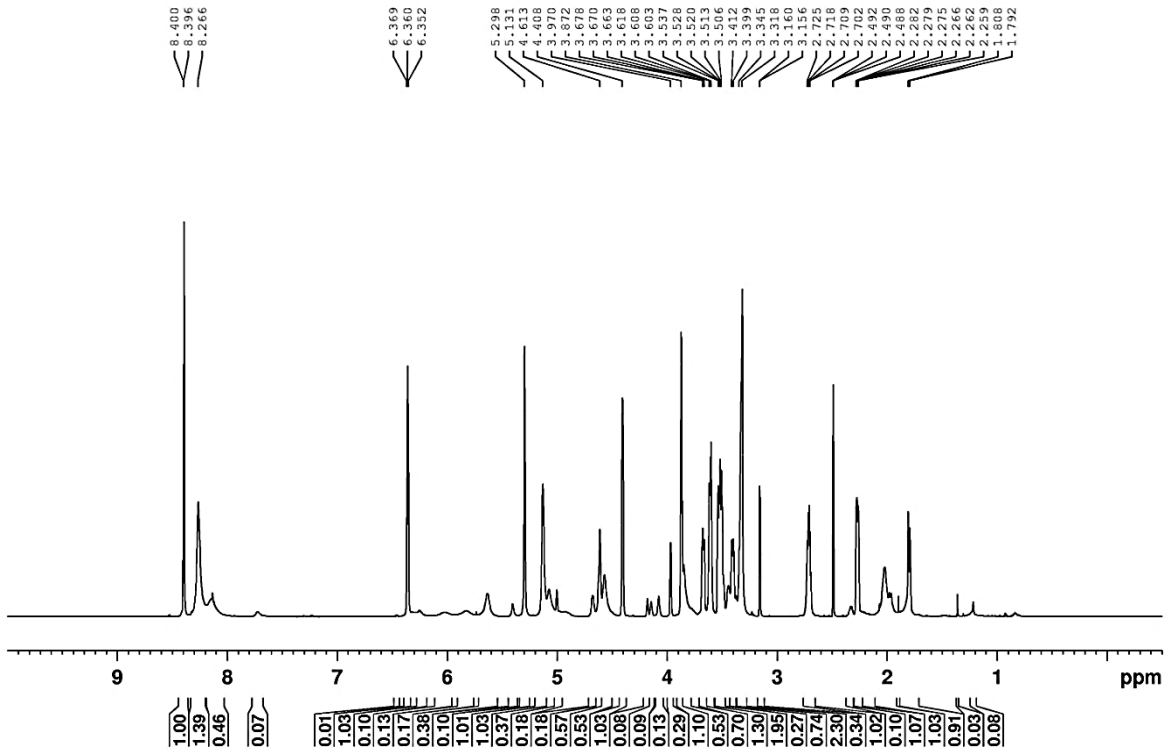


Figure S5. ^{1}H-NMR of compound **5** taken in DMSO(d6) at 800 MHz.

Figure S6

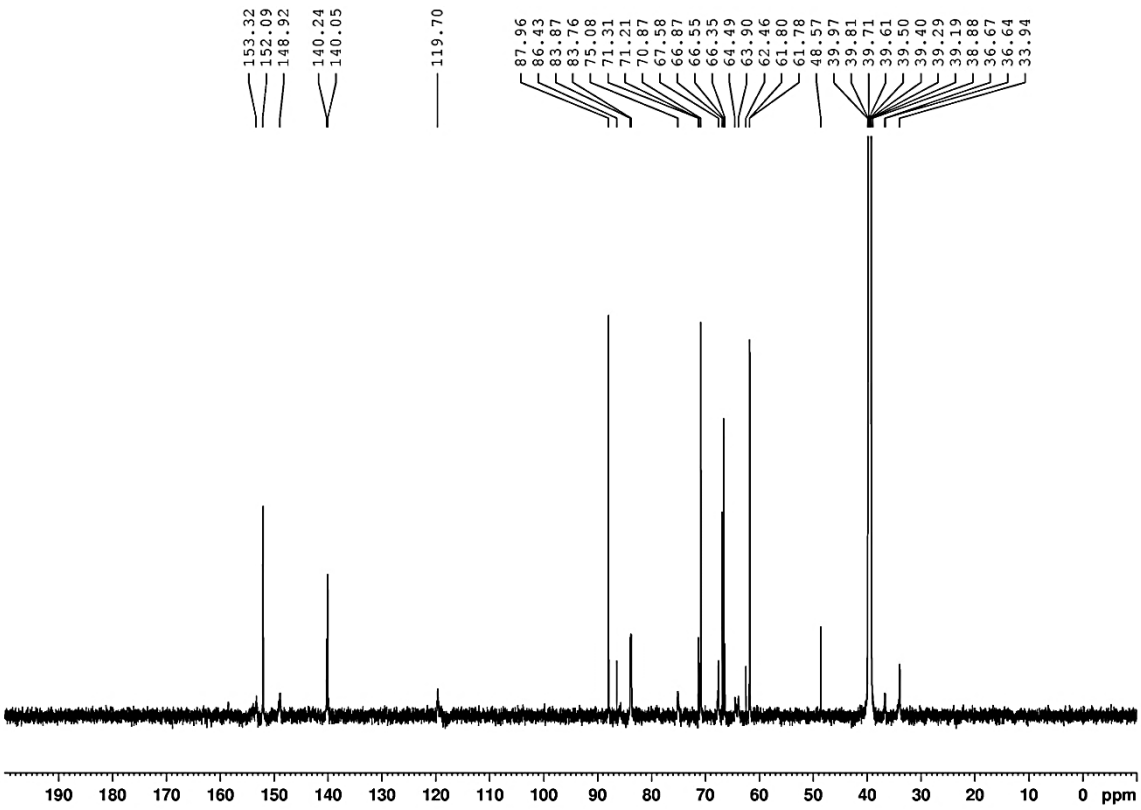


Figure S6. ^{13}C-NMR of compound **5** taken in DMSO(d6) at 200 MHz.

Figure S7

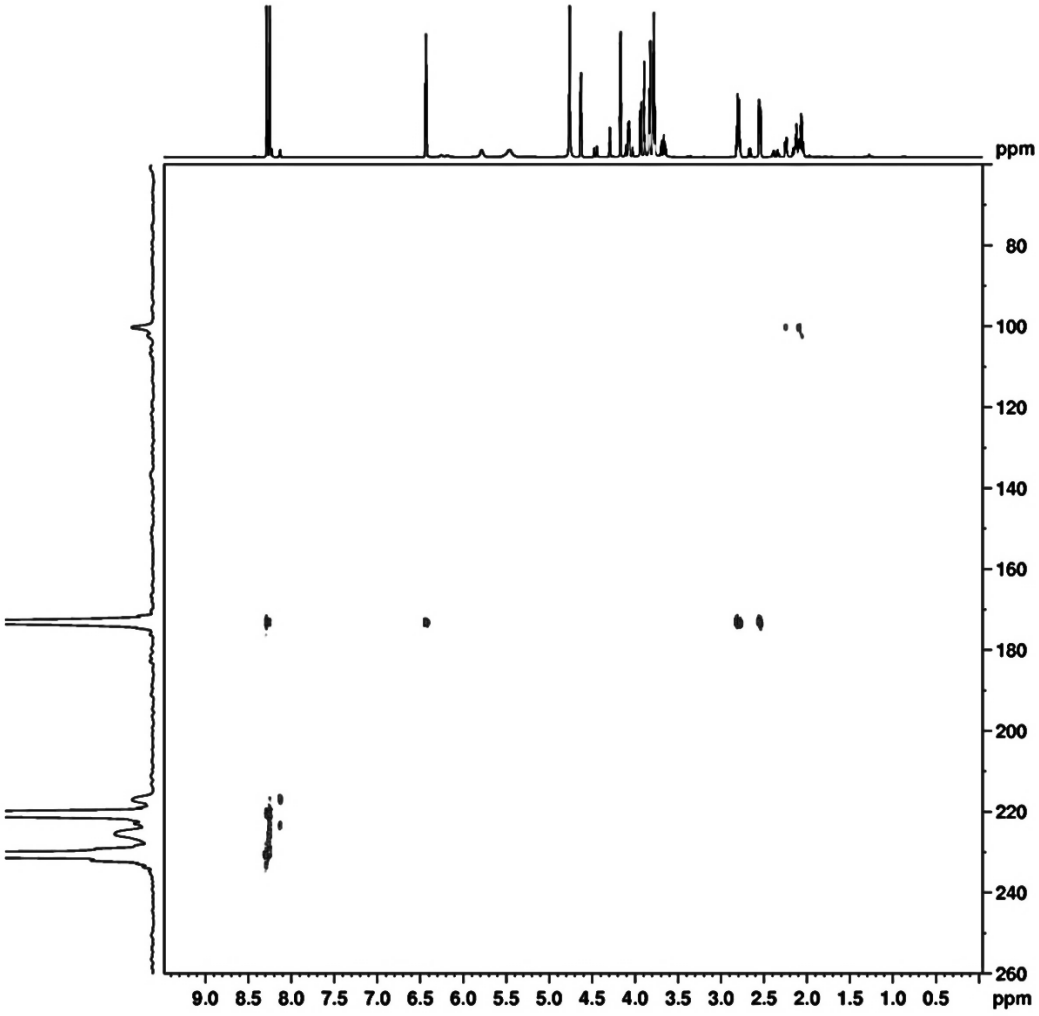


Figure S7. ^{15}N-HMBC of compound **5**, taken in DMSO(d6).

Figure S8

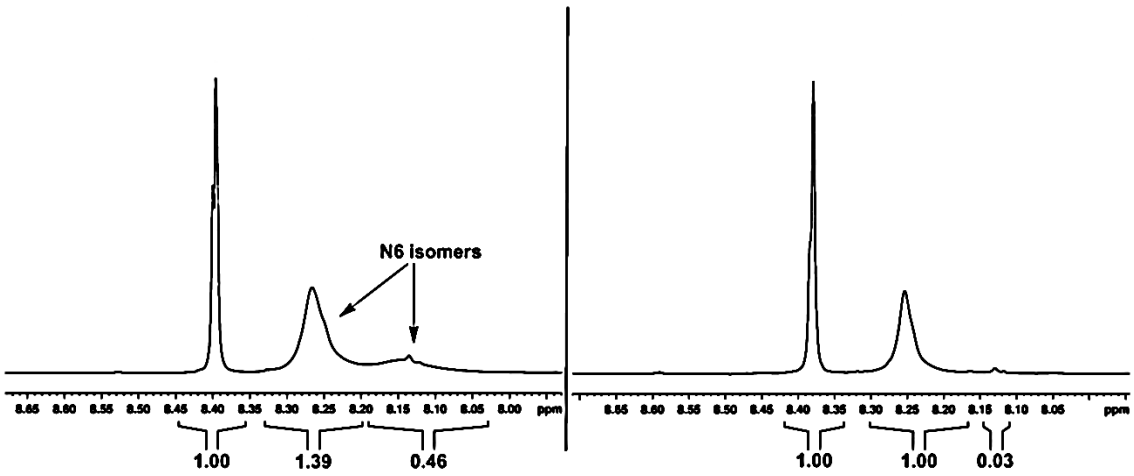


Figure S8. ^{1}H-NMR in DMSO (left) and in DMSO with D_2O exchange (right) of the N^6 protons of compound **5**. The H^2 proton of dA overlaps one isomer, however after D_2O exchange the integration is lowered by one proton and the second peak seen correlating in the ^{15}N-HMBC and TOCSY NMR disappears.

Figure S9

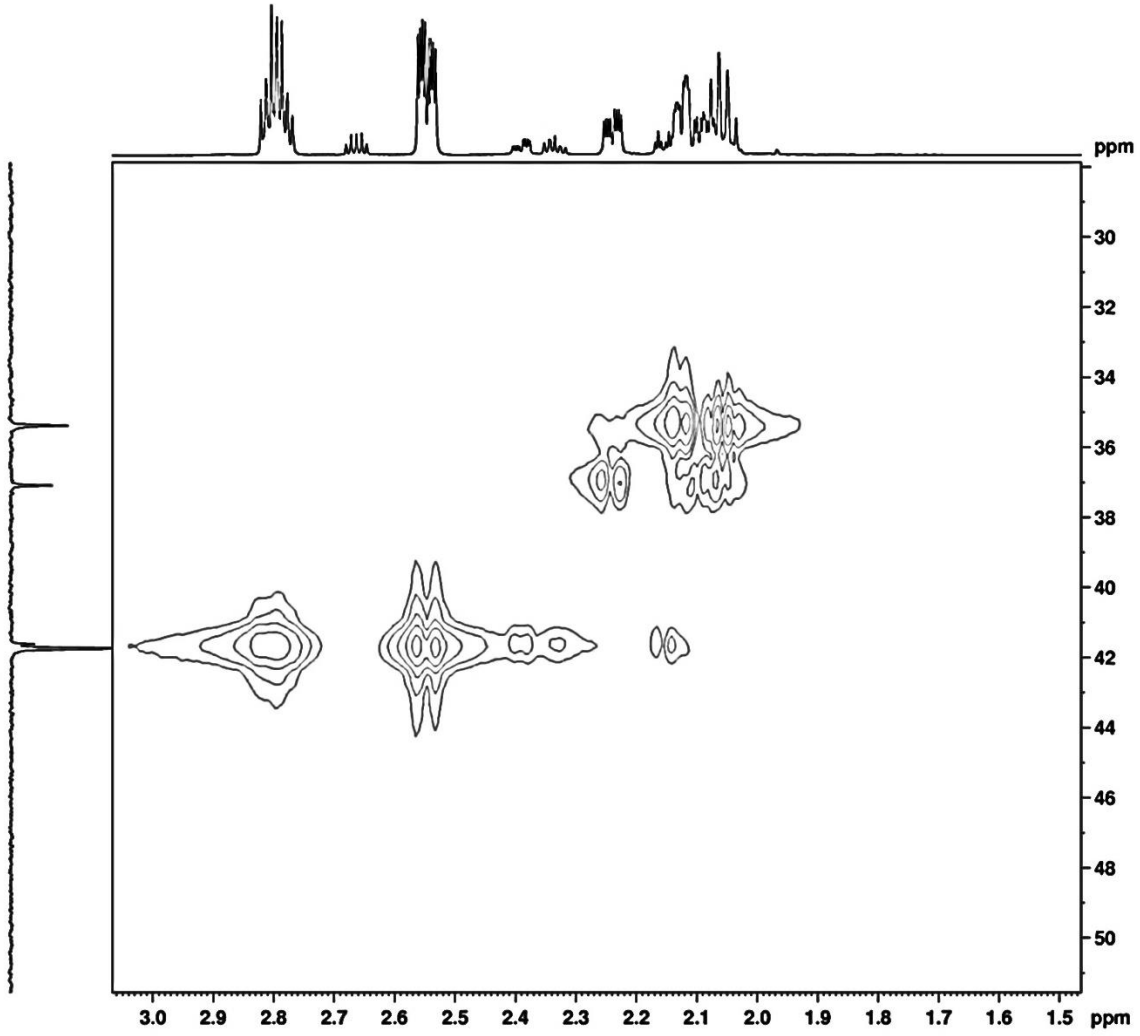


Figure S9. HMQC of the $dRH^{2'}$ and $dRH^{2''}$ region of compound **5** taken in deuterium oxide.

Figure S10

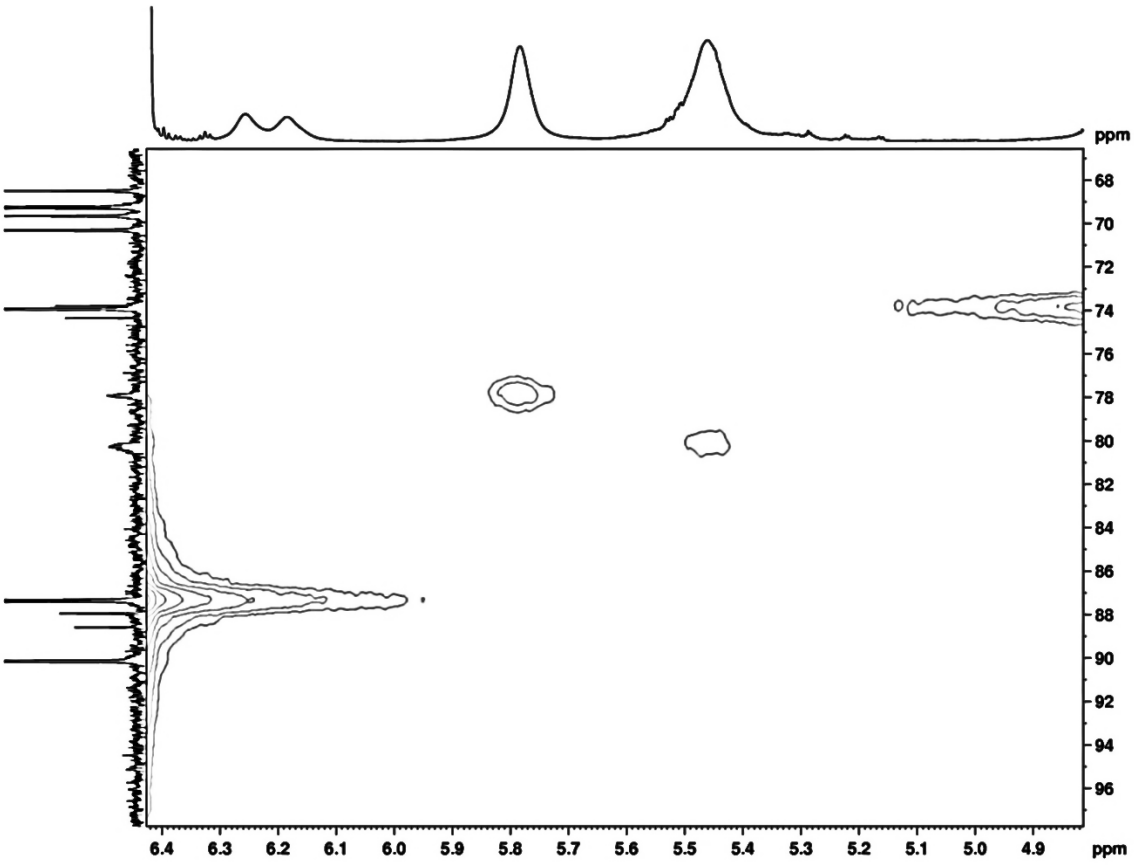


Figure S10. HMQC of the dR1’’H region of compound **5** taken in deuterium oxide.

Figure S11

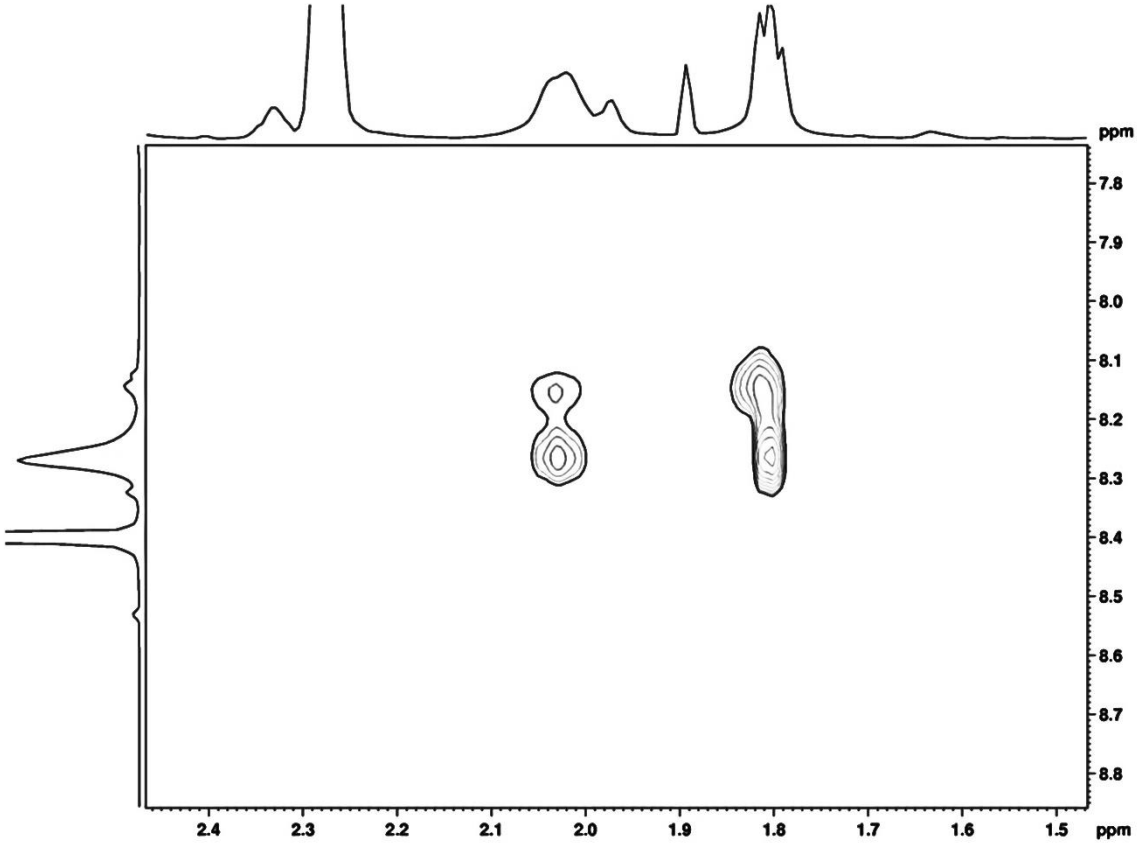


Figure S11. TOCSY of the N^6-H correlations with $dRH^{2''}$ of compound **5** taken in DMSO(d6).

Figure S12

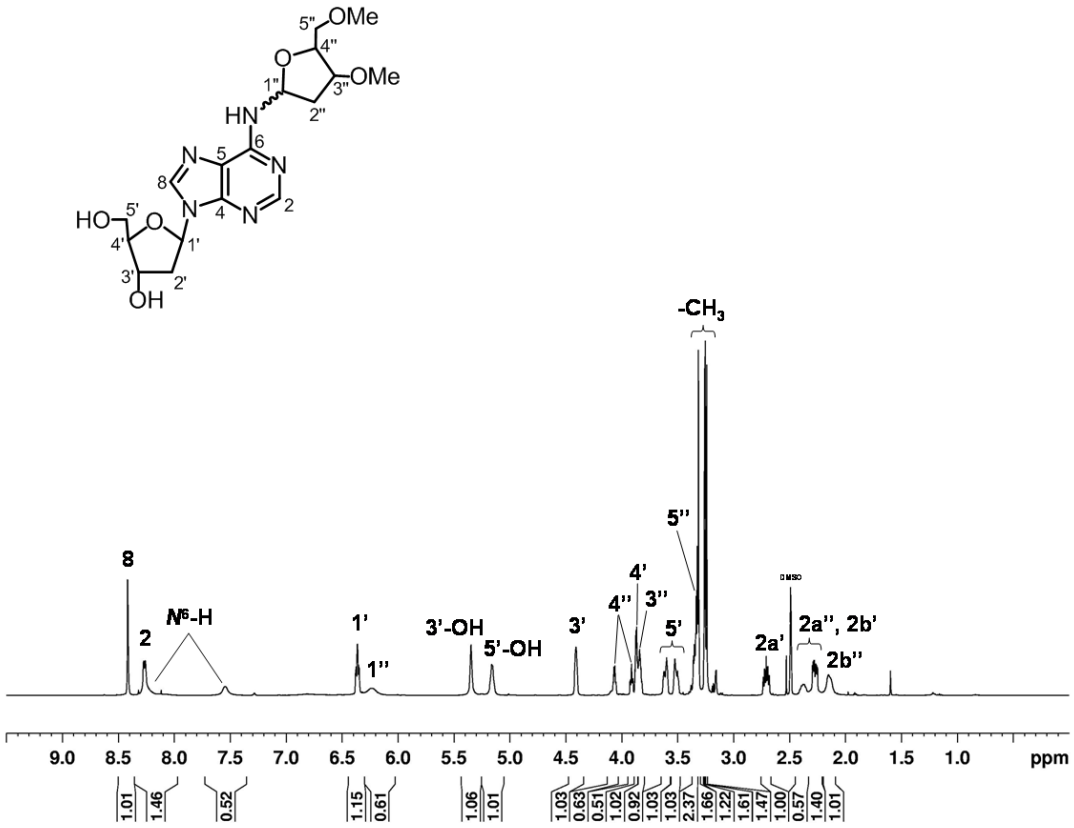


Figure S12. ^{1}H-NMR of **7** in DMSO(d6) .

Figure S13

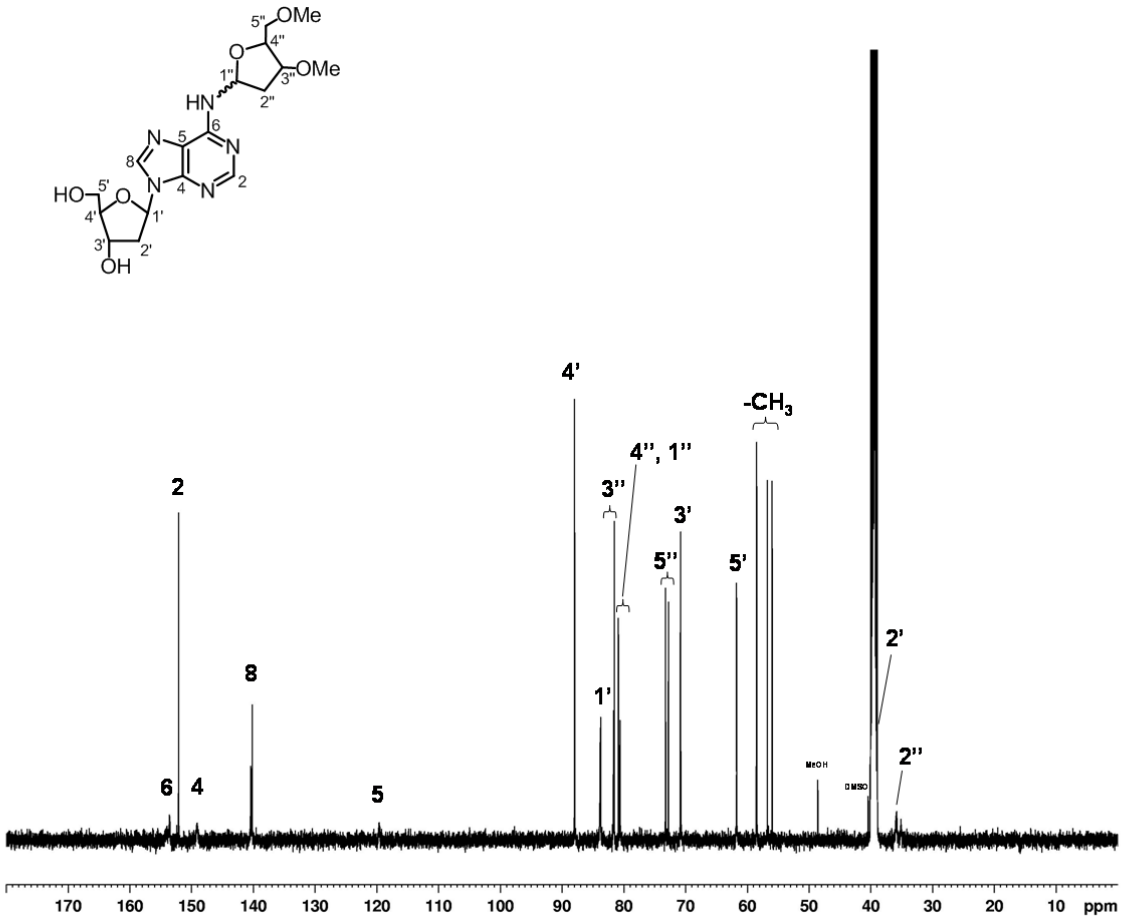


Figure S13. ^{13}C-NMR of **7** in DMSO(d6) .

Figure S14

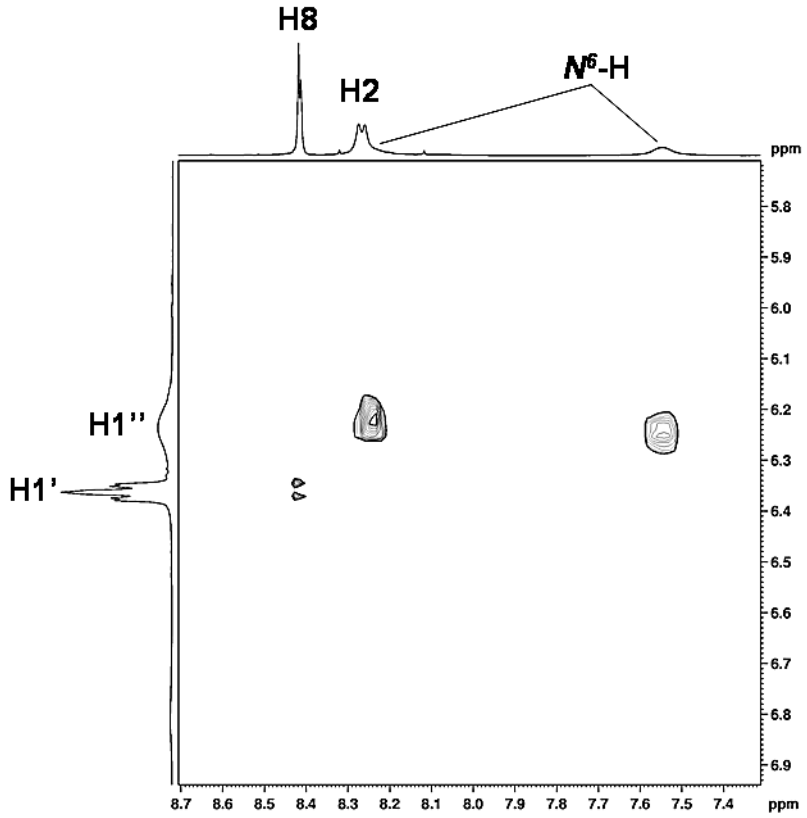


Figure S14. . Selected region of ^{1}H-^{1}H COSY spectrum of compound **7**. The crosspeaks shown represent homonuclear 3-bond correlations between the exocyclic N^6-H of dA and H1'' of the attached 2-deoxyribose moiety.

Figure S15

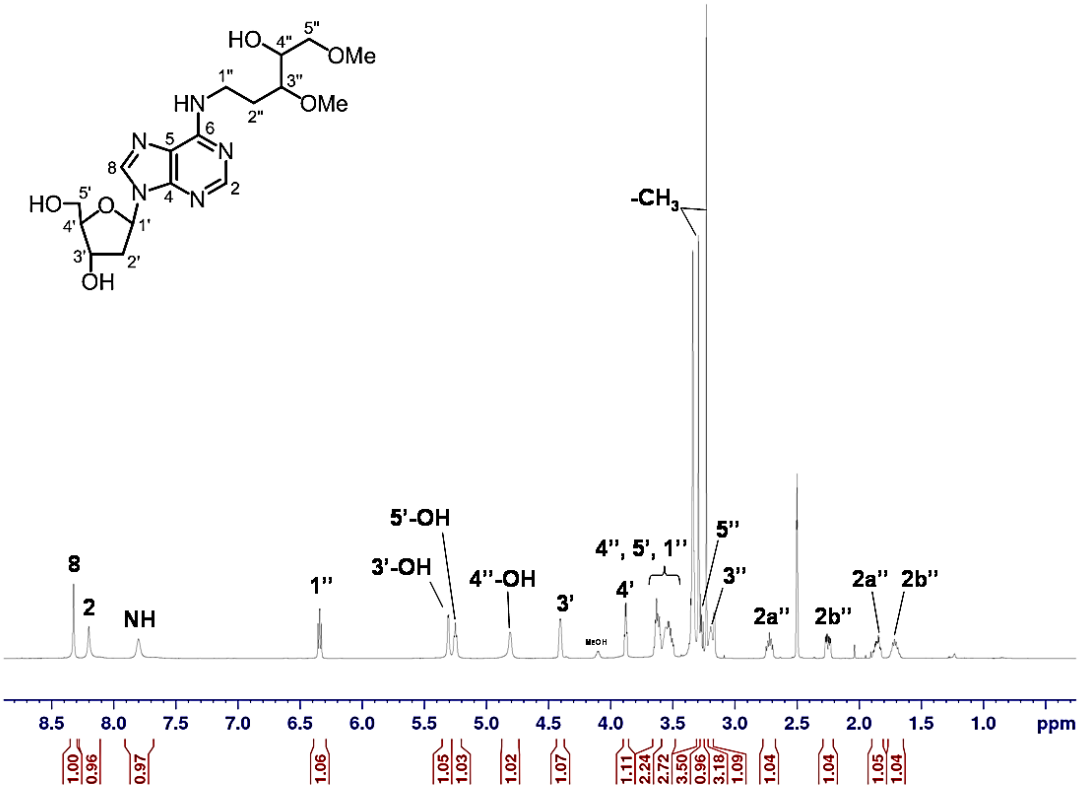


Figure S15. ^{1}H-NMR of **8** in DMSO(d6) .

Figure S16

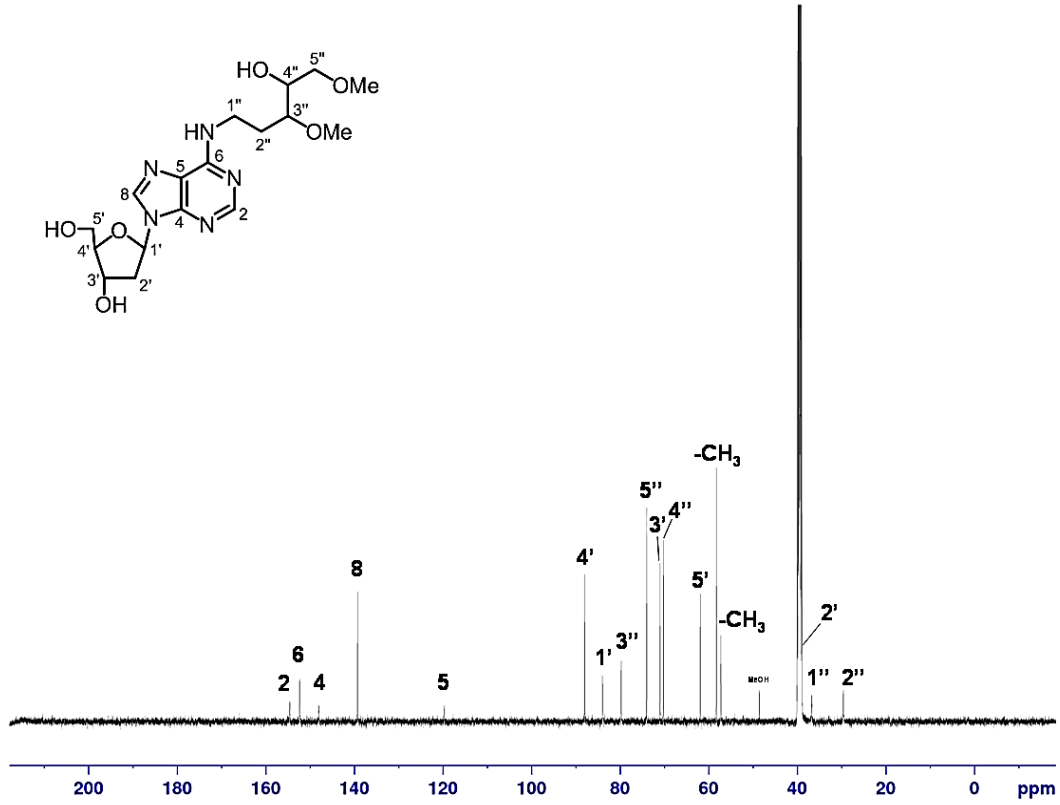


Figure S16. ^{13}C-NMR of **8** in DMSO(d6) .

Figure S17

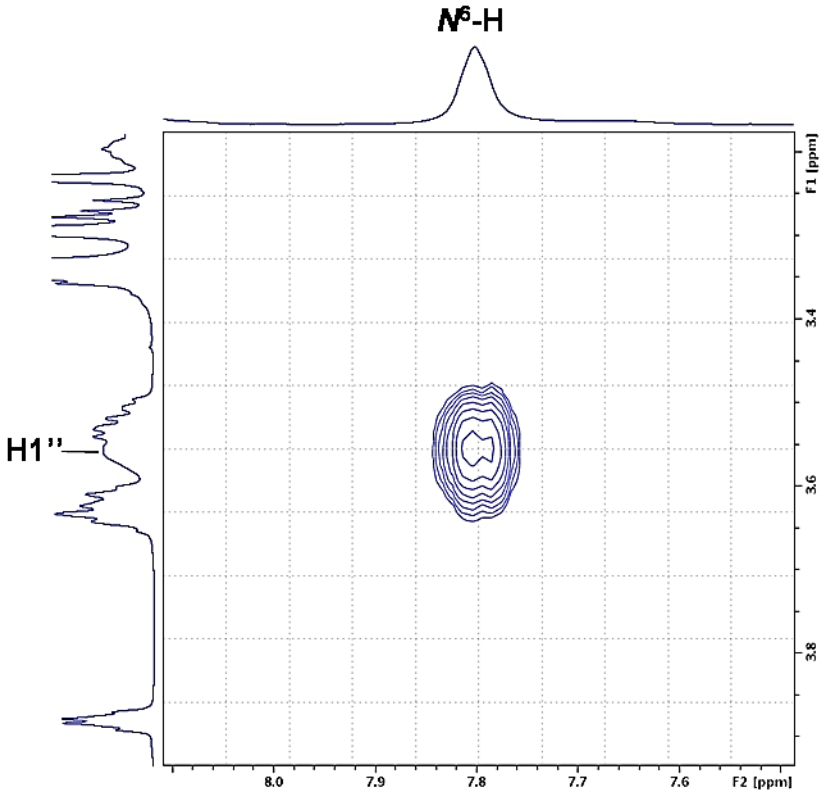


Figure S17. Selected region of ^{1}H-^{1}H COSY spectrum of compound **8**. The crosspeak shown represents the homonuclear 3-bond correlation between the exocyclic N^6-H of dA and H1'' of the reduced 2-deoxyribose moiety.

Figure S18

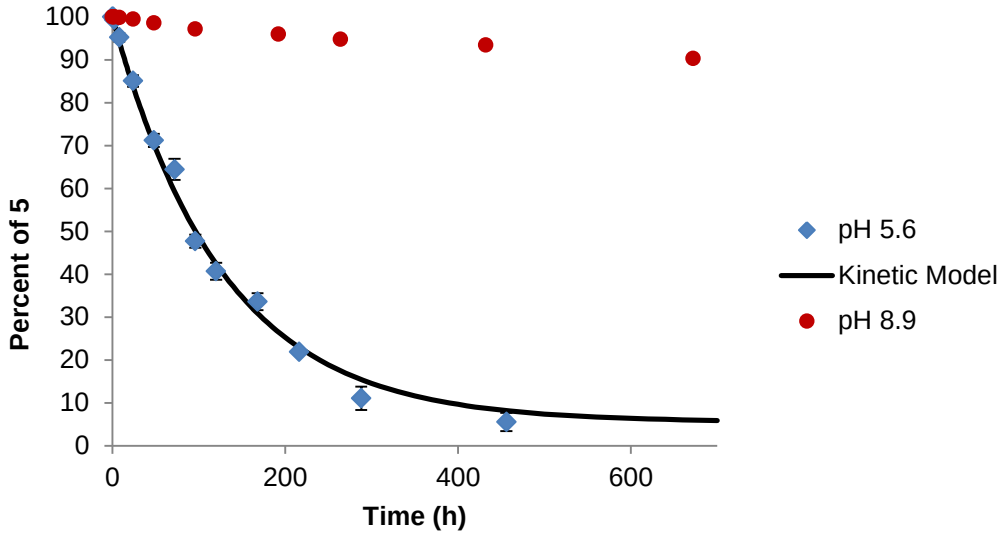


Figure S18. Dissociation of compound **5** during the conditions used to enzymatically digest duplex **A** and **B**. The conditions for pH 5.6 were 30 mM sodium acetate (pH 5.6), 40 µM EHNA (erythro-9-(2-hydroxy-3-nonyl)adenine), 1 mM zinc chloride and 1 mM **5**. The pH 8.9 conditions were the same as the pH 5.6 conditions except with the addition of Tris-HCl to a final concentration of 100 mM (pH 8.9). The dissociation was measure by the HPLC method described in the experimental section of this paper.

Figure S19

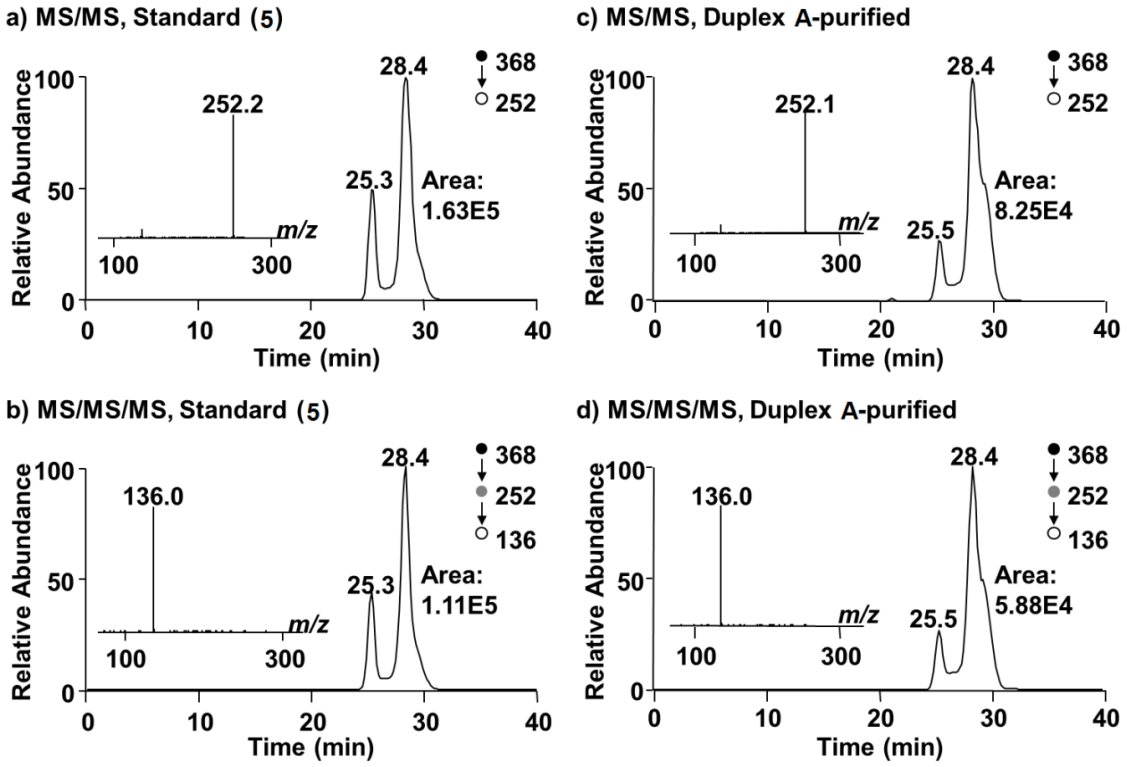


Figure S19. LC-MS/MS of the synthetic standard **5** and the enzymatically digested cross-linked duplex **A**. Panels A and B are for the synthetic nucleoside **5**, while panels C and D are for duplex **A**. Panels A and C are the chromatograms obtained for the first transition of 368→252. Panels B and D are the chromatograms obtained for the second transition of 252→136.

Figure S20

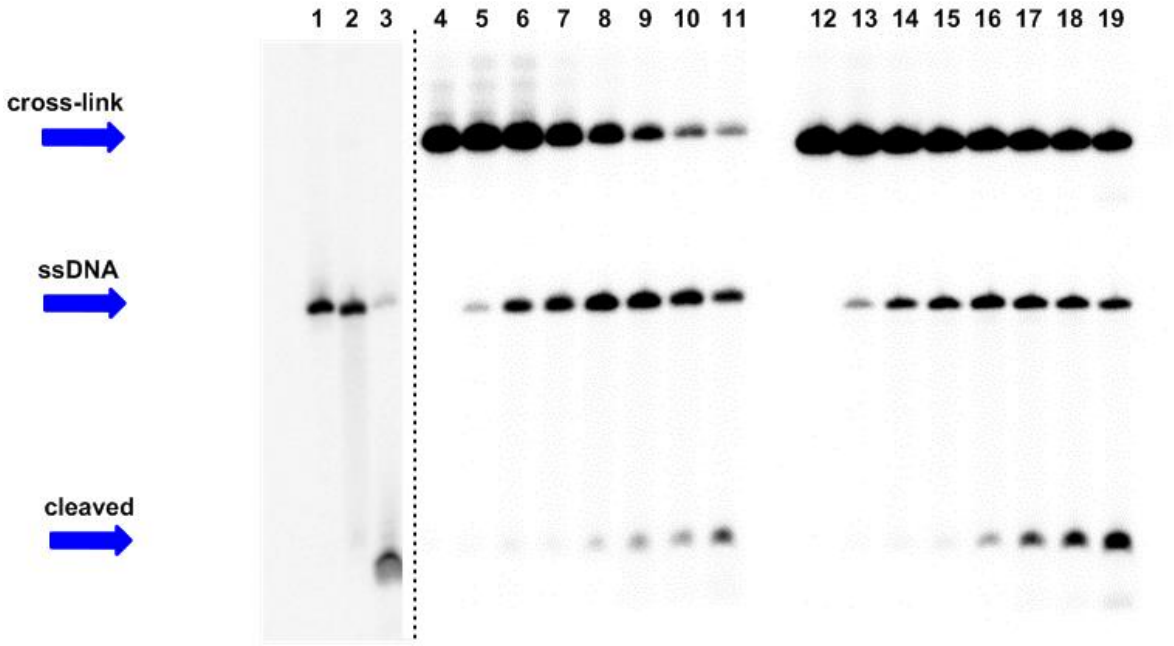


Figure S20. Dissociation of the cross-link in 50 mM HEPES (pH 7.0) and 100 mM NaCl at 37°C. Lanes 1-11 are duplex **A** and lanes 12-19 are duplex **B**. Lane 1 is the uracil containing strand of duplex F1. Lane 2 is the Ap-site containing duplex F1. Lane 3 is the piperidine treated Ap-containing duplex F1. 5 µl aliquots were taken, precipitated from ethanol and immediately frozen at the following times: 0, 8, 24 hours and then 2, 5, 10, 15 and 21 days. Lanes 4-11 are increasing time points for duplex **A**, while lanes 12-19 are increasing time points for duplex **B**. The time points were separated by polyacrylamide gel electrophoresis and the density of each band was quantitatively measured by phosphorimager analysis.

Figure S21

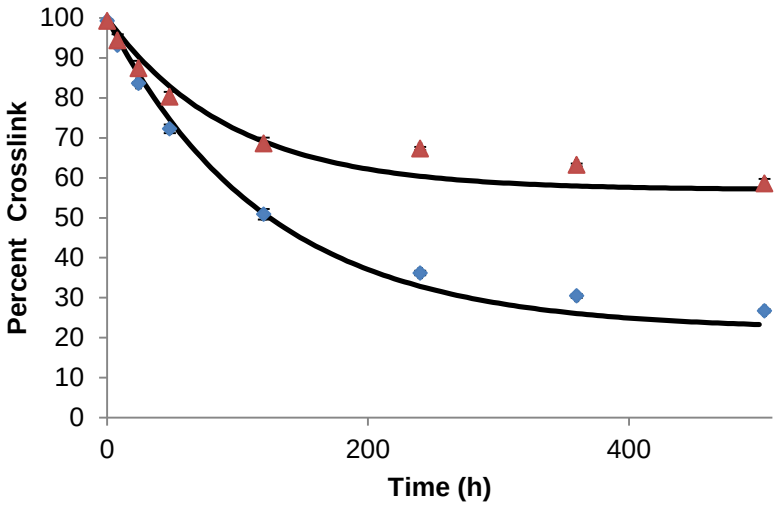


Figure S21. Dissociation of the cross-link in 50 mM HEPES (pH 7.0) and 100 mM NaCl at 37°C. Red triangles are duplex **B** and blue diamonds are duplex **A**. Aliquots were removed, immediately precipitated from ethanol and subsequently frozen. The solid line represents the model data set using the following rates k_{F1}= 8.22^{-3} h^{-1}, k_{E3}= 1.057^{-2} h^{-1}. The half-lives of dissociation are 84.31 hour for duplex **A** and 65.57 hours for duplex **B**.

Figure S22

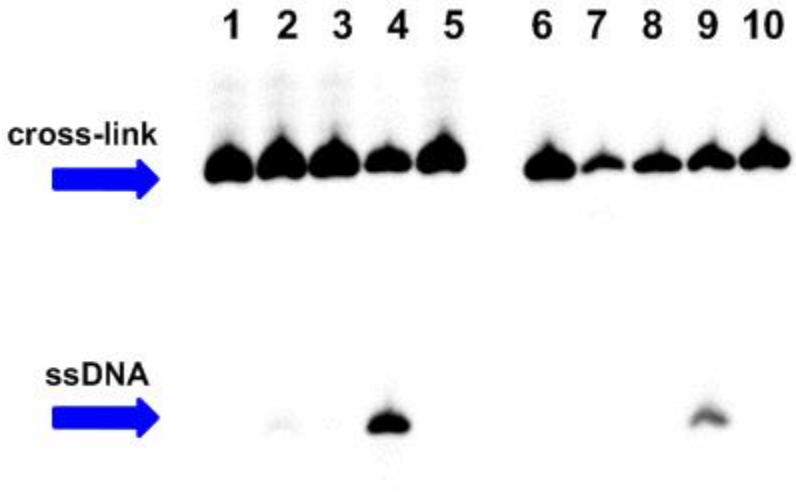


Figure S22. Cross-link was generated by incubating the Ap-containing duplexes at 37°C for 120 hours, gel purifying and excising from the gel. Lanes 1-5 are duplex **A** and lanes 6-10 are duplex **B**. Purified cross-link was incubated was redissolved and incubated in the following conditions. Lanes 1 and 6 had no incubation time. Lanes 2 and 7: 20 mM sodium phosphate (pH 7.0), 100 mM NaCl at 22°C. Lanes 3 and 8: 20 mM sodium phosphate (pH 7.0), 100 mM NaCl at 4°C. Lanes 4 and 9: 20 mM sodium acetate (pH 5.2), 100 mM NaCl at 22°C. Lanes 5 and 10: 13mM Tris, 4 mM boric acid (pH 9.2), 100 mM NaCl at 22°C. Incubations were done for 96 hours, separated by polyacrylamide gel electrophoresis and the density of each band was quantitatively measured by phosphorimager analysis.

2.10 References

(12) Antunes, A. M. M.; Duarte, M. P.; Santos, P. P.; Gamboa da Costa, G.; Heinze, T. M.; Beland, F. A.; Marques, M. M. *Chem. Res. Toxicol.* **2008**, *21*, 1443–1456.

(15) Begemann, P.; Boysen, G.; Georgieva, N. I.; Sangaiah, R.; Koshlap, K. M.; Koc, H.; Zhang, D.; Golding, B. T.; Gold, A.; Swenberg, J. A. *Chem. Res. Toxicol.* **2011**, *24*, 1048–1061.

(14) Christov, P. P.; Brown, K. L.; Kozekov, I. D.; Stone, M. P.; Harris, T. M.; Rizzo, C. J. *Chem. Res. Toxicol.* **2008**, *21*, 2324–2333.

(17) Christov, P. P.; Son, K.-J.; Rizzo, C. J. *Chem. Res. Toxicol.* **2014**, *27*, 1610–1618.

(13) Chung, F.-L.; Young, R.; Hecht, S. S. *Cancer Res.* **1984**, *44*, 990–995.

(28) Cordes, E. H.; Jencks, W. P. *J. Am. Chem. Soc.* **1963**, *85*, 2843–2848.

(24) Cortes, S. J.; Mega, T. L.; Van Etten, R. L. *J. Org. Chem.* **1991**, *56*, 943–947.

(20) Crosbie, S. J.; Murray, S.; Boobis, A. R.; Gooderham, N. J. *J. Chromatogr. B. Biomed. Sci. App.* **2000**, *744*, 55–64.

(38) Daley, J. M.; Zakaria, C.; Ramotar, D. *Mutat. Res. Mutat. Res.* **2010**, *705*, 217–227.

(35) Diekmann, S. *FEBS Lett.* **1986**, *195*, 53–56.

(31) Dworkin, J. P.; Miller, S. L. *Carbohydr. Res.* **2000**, *329*, 359–365.

(11) Essigmann, J. M.; Croy, R. G.; Nadzan, A. M.; Busby, W. F.; Reinhold, V. N.; Büchi, G.; Wogan, G. N. *Proc. Natl. Acad. Sci.* **1977**, *74*, 1870–1874.

(1) Gates, K. S. *Chem. Res. Toxicol.* **2009**, *22*, 1747–1760.

(2) Gates, K. S.; Nooner, T.; Dutta, S. *Chem. Res. Toxicol.* **2004**, *17*, 839–856.

(39) Greenberg, M. M. *Acc. Chem. Res.* **2012**, *45*, 588–597.

(16) Gong, J.; Vaidyanathan, V. G.; Yu, X.; Kensler, T. W.; Peterson, L. A.; Sturla, S. J. *J. Am. Chem. Soc.* **2007**, *129*, 2101–2111.

(3) Guengerich, F. P. *Chem. Rev.* **2006**, *106*, 420–452.

(22) Hayakawa, Y.; Nakagawa, M.; Kawai, H.; Tanabe, K.; Nakayama, H.; Shimazu, A.; Seto, H.; Otake, N. *J. Antibiot. (Tokyo)* **1983**, *36*, 934–937.

(7) Hlavin, E. M.; Smeaton, M. B.; Noronha, A. M.; Wilds, C. J.; Miller, P. S. *Biochemistry (Mosc.)* **2010**, *49*, 3977–3988.

(27) Holton, S.; Runquist, O. *J. Org. Chem.* **1961**, *26*, 5193–5195.

(23) Howard, G. A.; Kenner, G. W.; Lythgoe, B.; Todd, A. R. *J. Chem. Soc. Resumed* **1946**, 855–861.

(10) Hurley, L. H.; Reynolds, V. L.; Swenson, D. H.; Petzold, G. L.; Scahill, T. A. *Science* **1984**, *226*, 843–844.

(36) Ide, H.; Tedzuka, K.; Shimzu, H.; Kimura, Y.; Purmal, A. A.; Wallace, S. S.; Kow, Y. W. *Biochemistry (Mosc.)* **1994**, *33*, 7842–7847.

(34) Koo, H.-S.; Wu, H.-M.; Crothers, D. M. *Nature* **1986**, *320*, 501–506.

(33) Marini, J. C.; Levene, S. D.; Crothers, D. M.; Englund, P. T. *Proc. Natl. Acad. Sci.* **1982**, *79*, 7664–7668.

(32) McGhee, J. D.; Von Hippel, P. H. *Biochemistry (Mosc.)* **1975**, *14*, 1281–1296.

(37) Patro, J. N.; Haraguchi, K.; Delaney, M. O.; Greenberg, M. M. *Biochemistry (Mosc.)* **2004**, *43*, 13397–13403.

(21) Price, N. E.; Johnson, K. M.; Wang, J.; Fekry, M. I.; Wang, Y.; Gates, K. S. *J. Am. Chem. Soc.* **2014**, *136*, 3483–3490.

(5) Rouse, J.; Jackson, S. P. *Science* **2002**, *297*, 547–551.

(26) Sanderson, P. N.; Sweatman, B. C.; Farrant, R. D.; Lindon, J. C. *Carbohydr. Res.* **1996**, *284*, 51–60.

(29) Sattsangi, P. D.; Barrio, J. R.; Leonard, N. J. *J. Am. Chem. Soc.* **1980**, *102*, 770–774.

(19) Sengupta, S. K.; Blondin, J.; Szabo, J. *J. Med. Chem.* **1984**, *27*, 1465–1470.

(8) Smeaton, M. B.; Hlavin, E. M.; Noronha, A. M.; Murphy, S. P.; Wilds, C. J.; Miller, P. S. *Chem. Res. Toxicol.* **2009**, *22*, 1285–1297.

(25) Song Zhi-Wei; Ming-Hai Li; Chang-Mei Cheng; Lang-Qiu Chen; Xiao-Qiang Guo; Ru-Ji Wang; Guo-Shi Wu; Yu-Fen Zhao. *Chinese J. Struct. Chem.* **2007**, *26*, 1476–1480.

(6) Spanswick, V. J.; Lowe, H. L.; Newton, C.; Bingham, J. P.; Bagnobianchi, A.; Kiakos, K.; Craddock, C.; Ledermann, J. A.; Hochhauser, D.; Hartley, J. A. *BMC Cancer* **2012**, *12*, 436–448.

(9) Sturla, S. J. *Curr. Opin. Chem. Biol.* **2007**, *11*, 293–299.

(30) Upadhyaya, P.; Hecht, S. S. *Chem. Res. Toxicol.* **2008**, *21*, 2164–2171.

(18) Wang, J.; Yuan, B.; Guerrero, C.; Bahde, R.; Gupta, S.; Wang, Y. *Anal. Chem.* **2011**, *83*, 2201–2209.

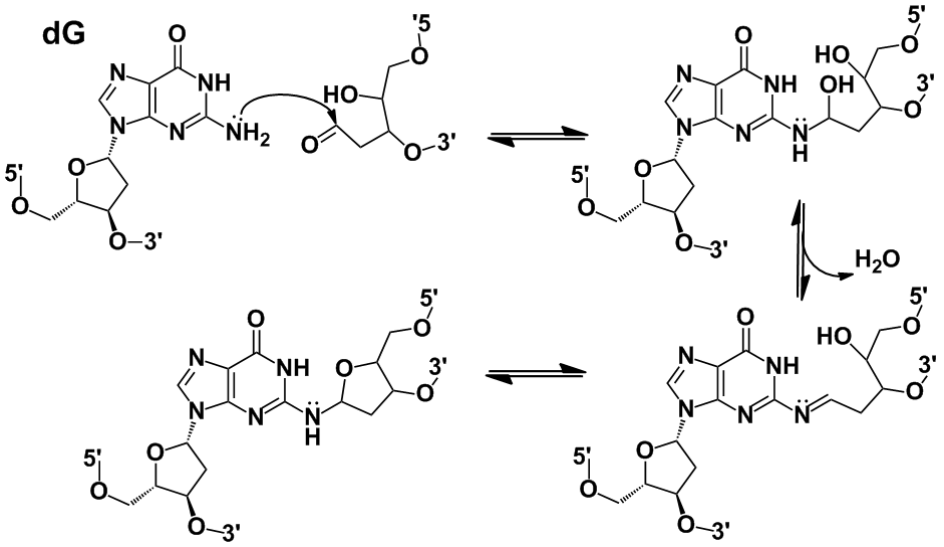
(4) Zhou, B.-B. S.; Elledge, S. J. *Nature* **2000**, *408*, 433–439.

Chapter 3. Sequence effects and scope of DNA interstrand cross-links formed between abasic sites and an opposing adenine residue

3.1 Introduction

Organisms defective in NER pathways, which typically participate in the removal of interstrand cross-links, display an enhanced rate of aging.[1–3] Additionally it has been observed that cell death and mutations arise from DNA interstrand cross-links.[4–7] Most of the previously studied cross-links occur after exposure to exogenous chemicals; however, even in the absence of environmental toxins there are thousands of oxidative and hydrolytic types of damage experienced by DNA.[5] One such form of endogenous DNA damage may be interstrand cross-links formed between abasic (Ap) sites and an opposing nucleobase within duplex DNA (Scheme 1). Our group has previously reported on cross-links derived from the reaction of Ap-sites with either opposing deoxyguanosine or deoxyadenosine residues in native duplexed DNA.[8–10]

Scheme 1



Ap-sites are generated in DNA after the glycosidic bond of a nucleoside is broken. Breakage of this bond may occur by spontaneous hydrolysis, enzymatic removal of misincorporated/damaged bases, or depurination after alkylation by chemotherapeutics/toxins.[11–14] The chemical reactivity and stability of Ap-dG/dA cross-links was characterized in our group's previous work, however these properties were only examined in a few sequences. In this study we expand the scope of Ap derived cross-links to a multitude of sequences including native and non-native duplexes with varied lengths and base compositions. The results show that Ap derived cross-links can form be at many different sequences in 5'-CAp and 5'-ApT. Additionally, the yield can be quite high in some of these sequences. We have also demonstrated that dG-Ap and dA-Ap cross-links can display several differences in rate, yield, gel mobility, reactivity with hydride and sequence preference.

3.2 Interstrand Cross-link Formation With an Opposing Deoxyguanosine Residue

Cross-links between an Ap-site and 2'-deoxyguanosine residue have been shown to form preferentially in 5'-CAp sequences, which properly aligns the N^2 amino group for imine formation with the Ap-site. Fellow Gates group member, Kevin Johnson, examined several sequences which varied the flanking and opposing bases of this 5'-CAp core. The oligonucleotides were 5'-^{32}P-labeled, separated by PAGE analysis and then the slow moving cross-link band was quantified by densitometry image analysis. The presence of $NaCNBH_3$ during the dG-Ap cross-link formation results in a reductive amination reaction and improved cross-link yields. Reduction of imine adducts can increase stability and also draw the products toward an irreversible amine-linkage.[15–17]

First, we observed the effects of alternating the base opposing the Ap-site. The opposing base might often be a dC residue because dG is known to be easily alkylated at its N7 position followed by depurination.[13] Alternately, a dA residue might occur due to misincorporation of dU in the place of dT[12]. An opposing dG or dT may occur less frequently, but will likely occur after alkylation/depurination of dA or enzymatic removal of a damaged dA/dC. Cross-link yields for duplexes which the opposing base was varied are shown in Figure **1** and may demonstrate that the structure of an Ap-containing duplex is affected by the opposing base. Pyrimidines opposing an abasic site have been shown to cause a collapse of the local duplex structure and rotate the Ap-site out of the duplex, where it would not be poised for cross-link formation, while purines may leave the helical structure more B-like.[18] The tendency toward a collapsed duplex in pyrimidine opposed duplexes is consistent with these sequences producing lower yields of cross-link.

Figure 1

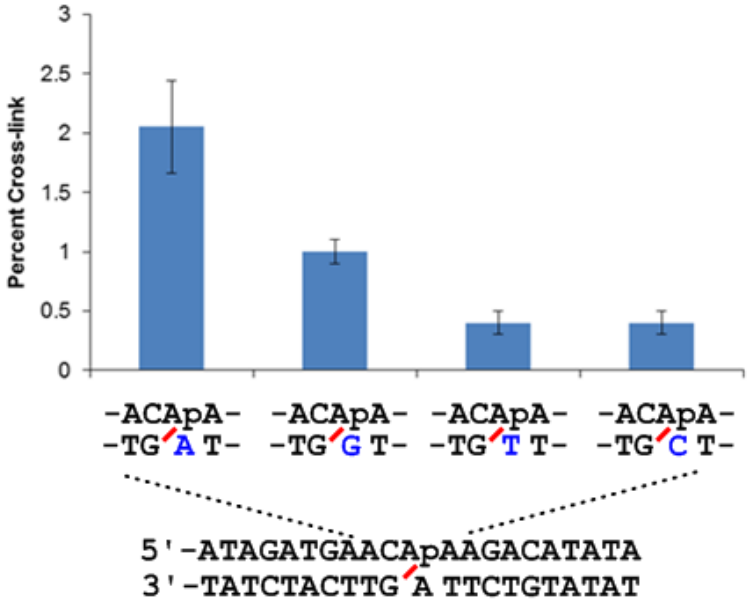


Figure 1. Percent DNA cross-links using PAGE analysis from radio-labeled duplexes. The base directly opposing the Ap-site was alternated in dG-Ap cross-linking sequences. A-opposed 2.0 ± 0.39%, G-opposed 1.0 ± 0.10%, T-opposed 0.4 ± 0.10%, and C-opposed 0.4 ± 0.10%.

Modified duplexes were also examined including inosine, 2-aminopurine, 2,6-diaminopurine and 5-methylcytosine bases. Non-native bases allow us to explore different chemical reactivity than the canonical bases provide. The results for all modified duplexes are shown in Figure 2. Inosine contains no exocyclic amino group and therefore does not offer this position for imine formation; therefore it provided no cross-link when in the dG cross-linking position. In contrast, 2-aminopurine and 2,6-diaminopurine have exocyclic amino groups in the same position as dG, however they produce much higher cross-link yields. Another sequence tested is a commonly seen endogenous modification, 5-methylcytosine, which is found in CpG islands and is an epigenetic modification of DNA.[19,20] The cross-link yields are higher for 5-

methylcytosine than for the un-methylated CpG island (Figure 3). The reduced yields were very high for 5-methylcytosine, 2-aminopurine and 2,6-diaminopurine. In fact, all of the Ap-dG cross-links showed a substantial increase in cross-link yield accompanied by a small gel shift after reduction with sodium cyanoborohydride (Figure 3). Sequences with very high reduced yields may be useful for synthesis of stable cross-links for biological or structural studies.

Figure 2

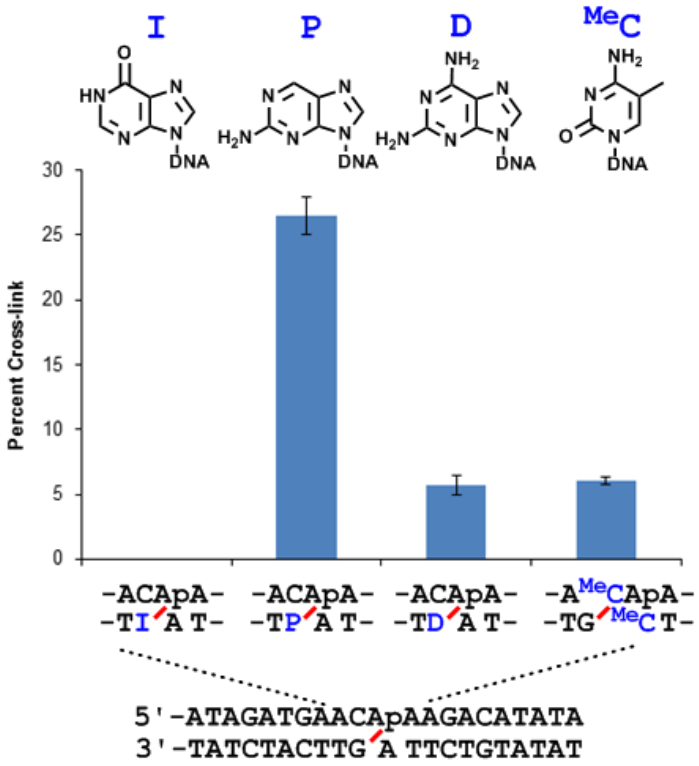


Figure 2. Percent DNA cross-links using PAGE analysis from radio-labeled duplexes. The cross-linking base was modified with non-native bases in dG-Ap cross-linking sequences. Inosine ND, 2-aminopurine 26.5 ± 1.45%, 2,6-diaminopurine 5.7 ± 0.70%, and ^{Me}CpG 6.0 ± 0.30%.

Figure 3

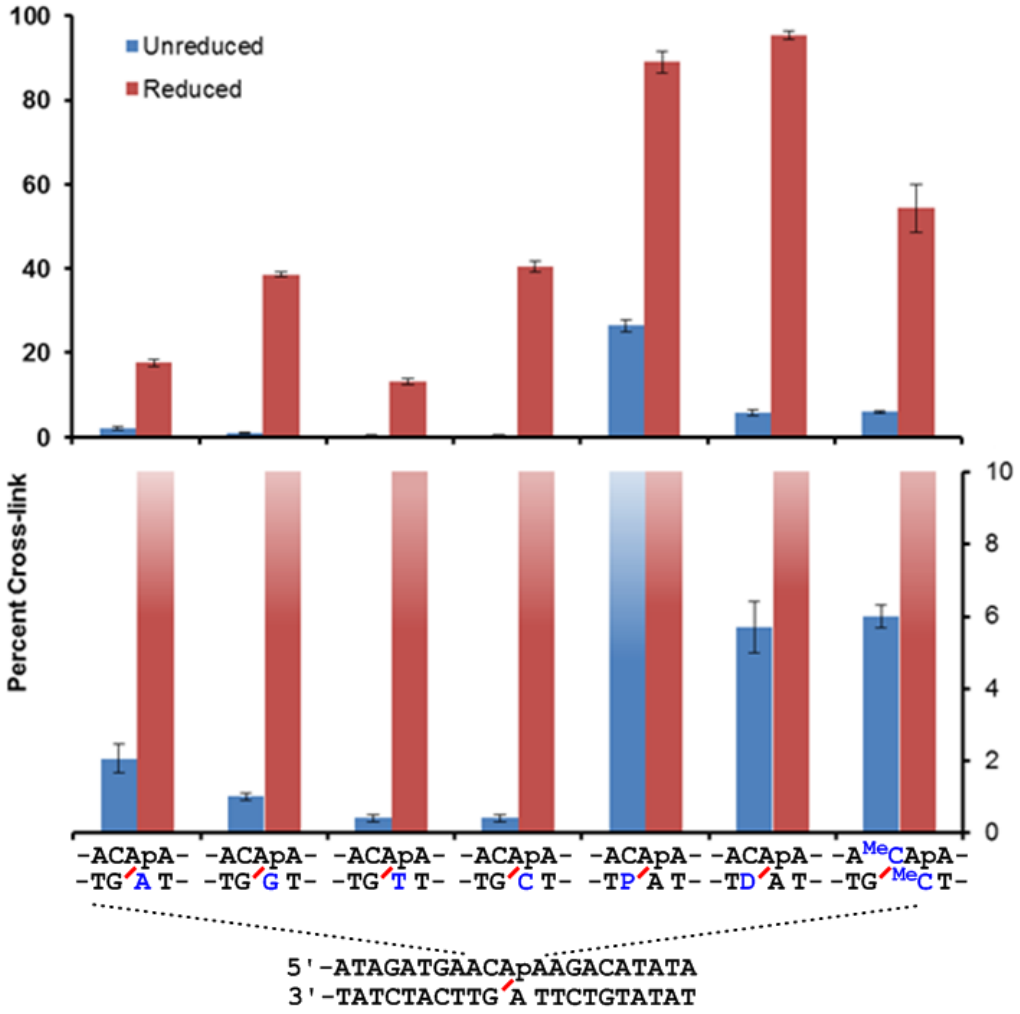


Figure 3. Percent DNA cross-links using PAGE analysis from radio-labeled duplexes. Cross-links were formed with and without the addition of sodium cyanoborohydride to promote reduction of the imine linkage in dG-Ap cross-linking sequences. Unreduced yields A-opposed 2.0 ± 0.39%, G-opposed 1.0 ± 0.10%, T-opposed 0.4 ± 0.10%, C-opposed 0.4 ± 0.10%, 2-aminopurine 26.5 ± 1.45%, 2,6-diaminopurine 5.7 ± 0.70%, and ^{Me}CpG 6.0 ± 0.30%. Reduced yields A-opposed 17.6 ± 0.86%, G-opposed 38.6 ± 0.50%, T-opposed 13.1 ± 0.70%, C-opposed 40.3 ± 1.30%, 2-aminopurine 89.1 ± 2.55%, 2,6-diaminopurine 95.5 ± 1.00%, and ^{Me}CpG 54.3 ± 5.70%.

Sequences flanking the cross-link were also varied in dG-Ap cross-linking duplexes (Figure 4). Changes to bases on the 5' side have very little impact on the reduced or unreduced cross-link yields. When the 3' position was alternated very little change was observed except when dA was put in the 3' and opposing position. This sequence is clearly an outlier and the cross-linked formed in this case was found to involve reaction of the opposing dA residue with the Ap site.[10] The possibility of a Ap-dA cross-link was previously examined and were found to occur in high yields by our group during a previous study. The details of Ap-dA cross-linking sequence and base composition effects will be explored in this chapter.[10]

Figure 4

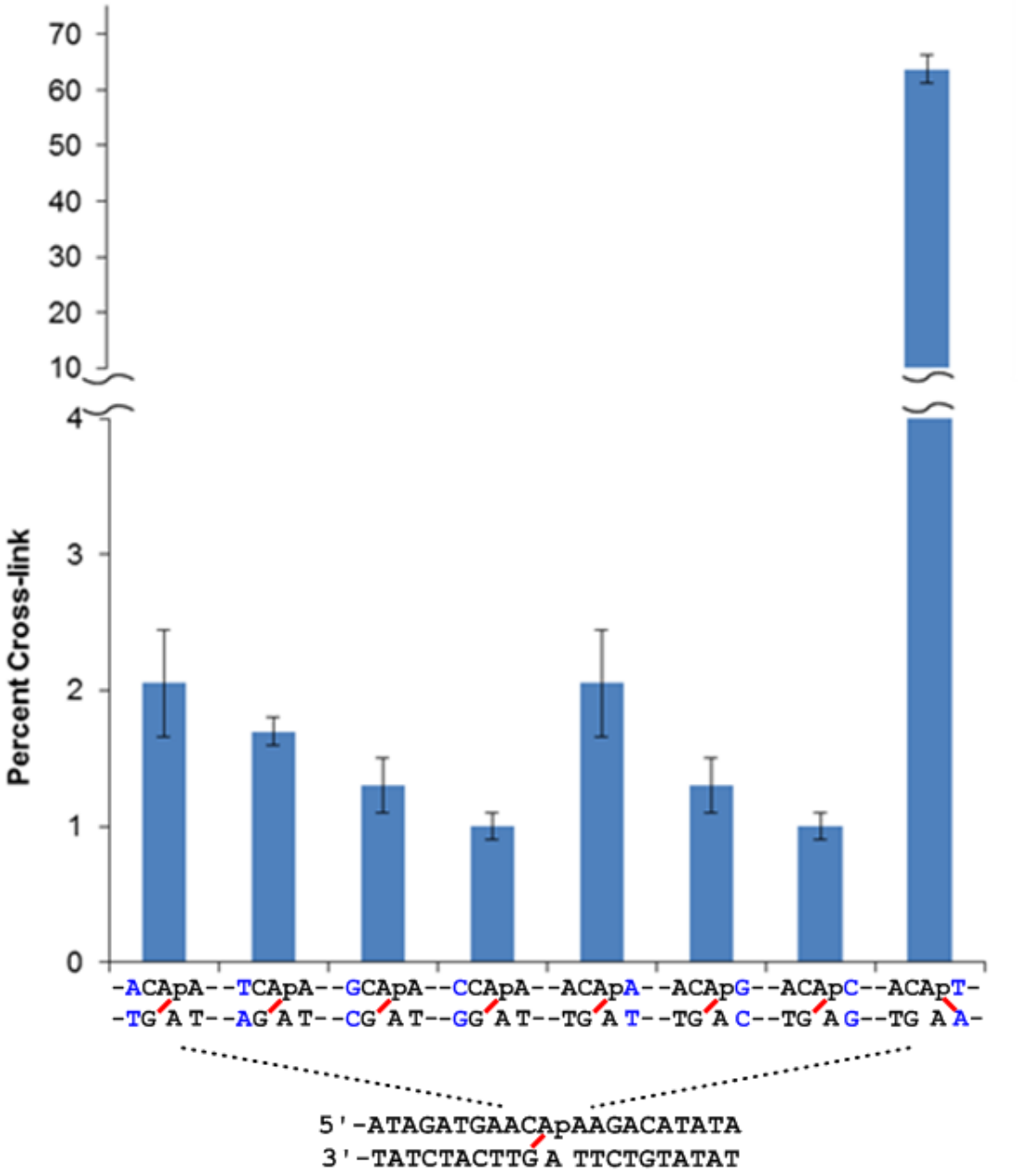


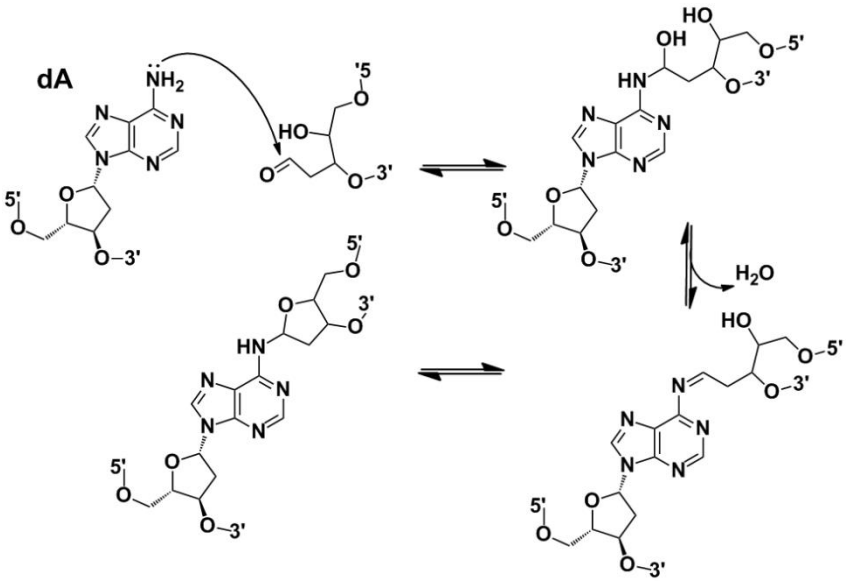
Figure 4. Percent DNA cross-links using PAGE analysis from radio-labeled duplexes. The bases 5' of the cross-link or 3' of the Ap-site where alternated in dG-Ap cross-linking sequences. 5'-A/T 2.0 ± 0.39%, 5'-T/A 1.7 ± 0.10%, 5'-G/C 1.3 ± 0.20%, 5'-C/G 1.0 ± 0.10%. 3'-A/T 2.0 ± 0.39%, 3'-G/C 1.4 ± 0.20%, 3'-C/G 1.1 ± 0.10% and 3'-T/A 63.7 ± 2.53%

Results and Discussion

3.3 Interstrand Cross-link Formation With an Opposing Deoxyadenosine Residue

Much the same as the Ap-dG cross-links, interstrand cross-links between an Ap-site and opposing adenine residues have previously been characterized in only a few (Scheme 2), but a complete understanding of sequence effects was lacking. Accordingly we set out to systematically investigate the effect of sequence on Ap-dA cross-link formation. The Ap-dA cross-links display many differences when compared to Ap-dG cross-links, which were examined throughout the following experiments.

Scheme 2



Alternating the base directly opposing the Ap-site produced higher yields with opposing purine as compared with pyrimidine bases, which is a result that mirrors the trends seen for Ap-dG cross-links (Figure 5). Changes in the 5' and 3' flanking bases revealed some clear trends (Figure 6); the first being G:C/C:G base pairs on the 5' increase the yield. Cross-link may be favored by a more stable base pairing on the 5' side of the Ap-site which might stabilize a B-like Ap-duplex. Variation of the 3' flank

indicates that pyrimidines next to the cross-linking base are favored; however the structural reason behind this is not yet understood.

Figure 5

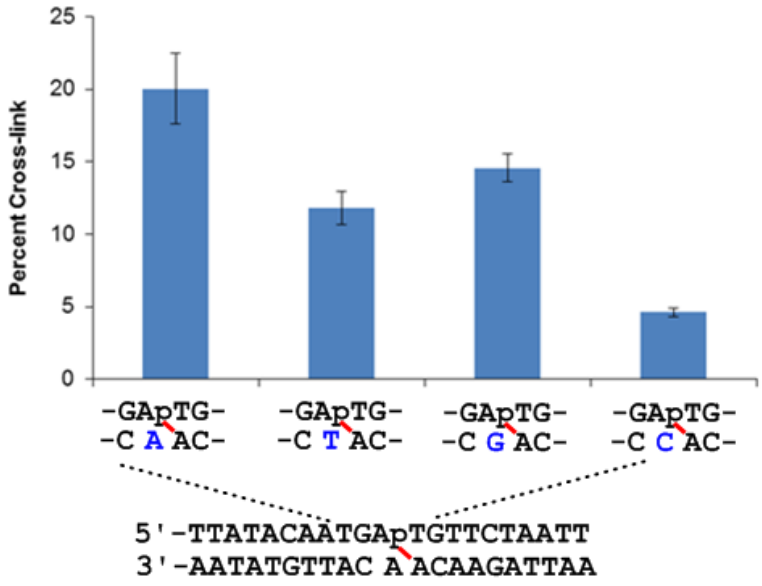


Figure 5. Percent DNA cross-links using PAGE analysis from radio-labeled duplexes. The base directly opposing the Ap-site was alternated in dA-Ap cross-linking sequences. A-opposed 20.0 ± 2.43%, T-opposed 11.8 ± 1.13%, G-opposed 14.6 ± 0.98%, and C-opposed 4.6 ± 0.29%.

Figure 6

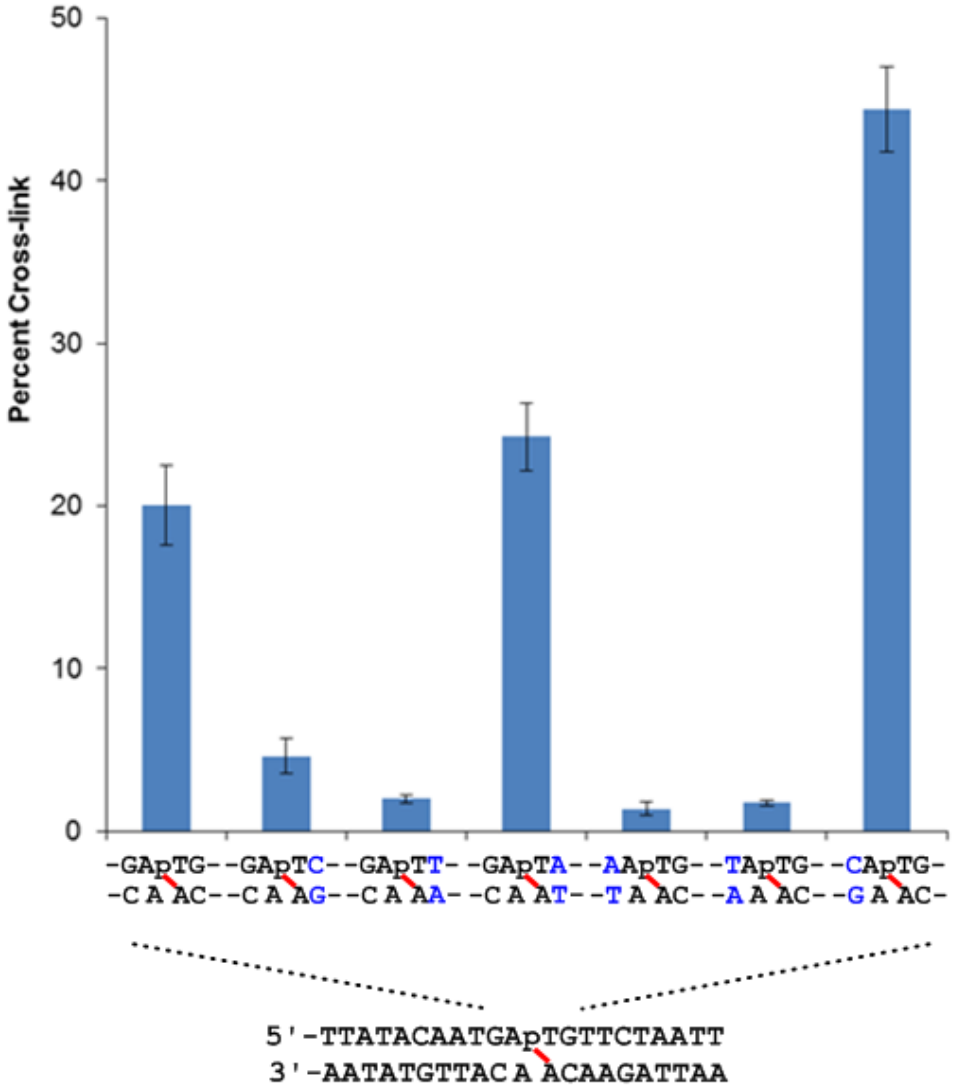


Figure 6. Percent DNA cross-links using PAGE analysis from radio-labeled duplexes. The base directly opposing the Ap-site was alternated in dA-Ap cross-linking sequences. 3'-G/C 20.0 ± 2.43%, 3'-C/G 4.6 ± 1.06%, 3'-T/A 2.0 ± 0.23%, 3'-A/T 24.2 ± 2.06%. 5'-A/T 1.4 ± 0.39%, 5'-T/A 1.7 ± 0.17% and 5'-C/G 44.4 ± 2.59%.

We carried out cross-linking reactions in the presence of $NaCNBH_3$. In the case of dG-Ap cross-links these conditions resulted in a reduction of the cross-link and correspondingly increased the cross-link yield and stability.[15–17] Therefore, in an attempt

to stabilize the Ap-dA cross-links, much the same as we had done for the Ap-dG cross-links discussed earlier, we treated them with sodium cyanoborohydride at pH 5. However once cross-link was allowed to reach equilibrium yields and then subjected to reductive conditions (750 mM NaOAc pH 5.2 and 250 mM $NaCNBH_3$ for 12 hours at 37°C), unlike the Ap-dG cross-links there was no gel shift and instead of a marked increase there was a slight decrease in cross-link yields (~5%) for Ap-dA cross-links (Figure **7**, far left). The resistance of a synthetic Ap-dA cross-link remnant to chemical reduction was discussed in chapter 2. In addition to our own studies other reports for formaldehyde derived imines and Fapy-dG imines have described these structures as being resistant to hydride mediated reduction.[21,22] The helical structure of DNA itself may increase this resistance by favoring a cyclic-hemiaminal form that is not a substrate for $NaCNBH_3$ (Scheme 4). A slight decrease in cross-linking yield during this process is likely due to reduction of the Ap-aldehyde, which causes less reactant available to form an imine cross-link.

During examination of Ap-dG cross-links we observed that 2-aminopurine and 2,6-diaminopurine produced very high yields when 5' and opposing the Ap-site. When these same two bases are placed in the position to form a dA-like cross-link, 3' and opposing the Ap-site, the importance of the exocyclic amino group's position becomes clear (Scheme 3). As described in a previous in chapter 1 a subtle relocation of the amino group from the major to the minor groove will completely prevent the cross-link.[10] We observed very low cross-link yields, even after reduction with sodium cyanoborohydride for 2-aminopurine, yet normal yields for the unreduced 2,6-diaminopurine, which retains an amino group in the major groove (Figure **7**). After examining the effect of variations

on the 5'-ApT sequences we observed the highest yields in a native duplex have appeared 5'-CApT (Figure 4 and 6, far right) sequences and we wanted to investigate cross-links from within these sequence contexts further.

Scheme 3

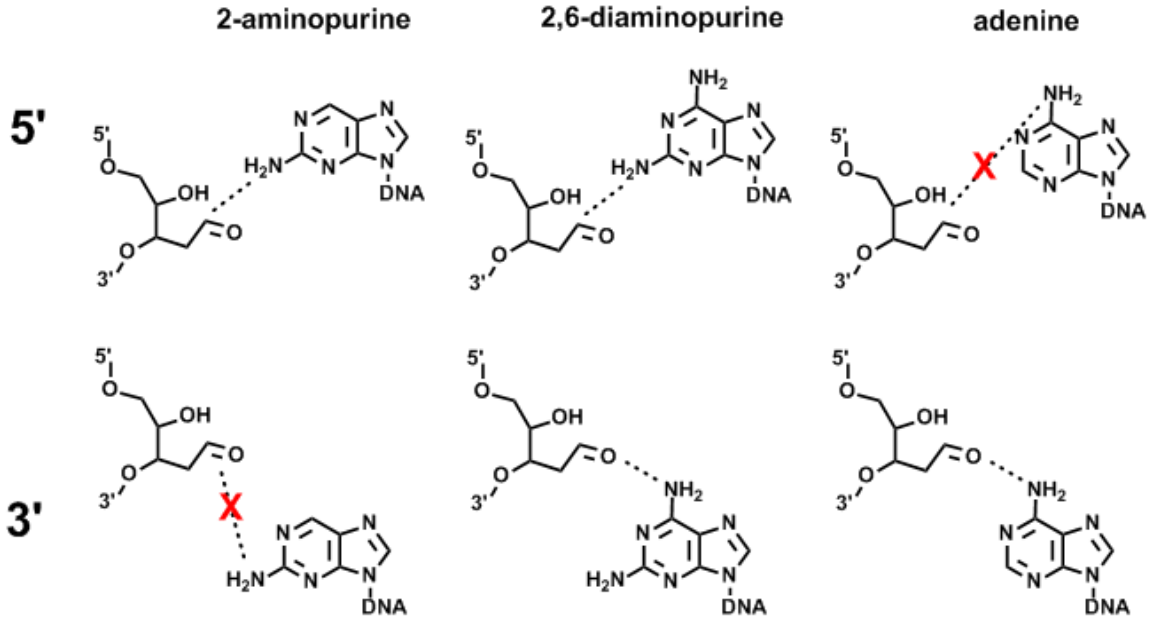


Figure 7

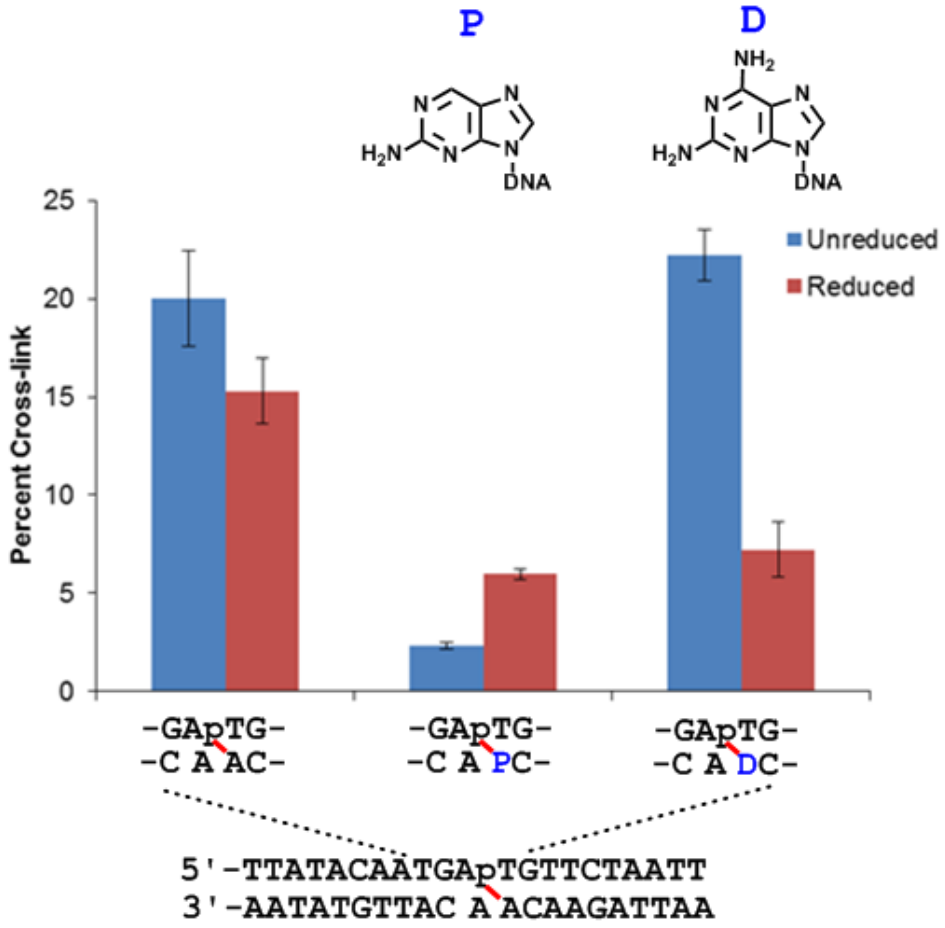


Figure 7. Percent DNA cross-links using PAGE analysis from radio-labeled duplexes. The cross-linking base was modified with non-native bases in dA-Ap cross-linking sequences and incubated under reductive or non-reductive conditions. Unreduced Adenine 20.0 ± 2.43%, 2-aminopurine 2.3 ± 0.17% and 2,6-diaminopurine 22.2 ± 1.30%. Reduced Adenine 15.3 ± 1.66%, 2-aminopurine 6.0 ± 0.27% and 2,6-diaminopurine 7.2 ± 1.40%.

3.4 High Yielding Cross-links Form Within 5'-CApT Sequences

Very high yields of Ap-site derived cross-links have been seen in duplexes with a 5'-CApT sequence. When varying the opposing base we found that dA produces the highest cross-link yields while dT and dC produce a much lower yield. One difference we noted was that dG produced the lowest cross-link yields when opposing one5'-CApT

sequence (Figure 8). Overall the same trends for native sequence variations were observed for Ap-dG and the previously mentioned Ap-dA, but all of the yields are much higher (Figure 8, 9). In addition to variations of the opposing base we examined variations of the base pair 5' of the cross-linking site, which showed slightly higher yields when an A/T pair was present as compared to a G/C pair (Figure 8). This result is again consistent with data observed for the previously discussed dA-Ap cross-linking duplex variations.

Figure 8

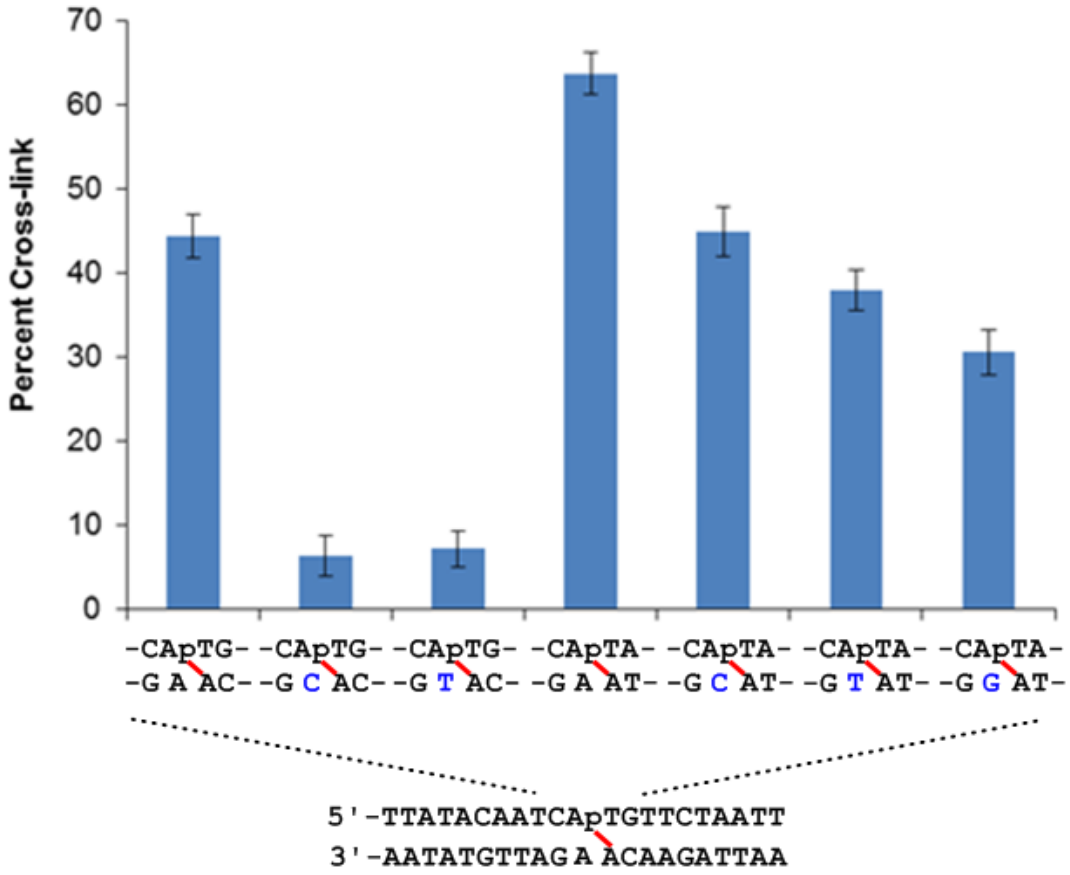


Figure 8. Percent DNA cross-links using PAGE analysis from radio-labeled duplexes. The base directly opposing the Ap-site was alternated in high yielding (5'-CApT) dA-Ap cross-linking sequences. Left to right: 44.4 ± 2.59%, 6.4 ± 2.41%, 7.2 ± 2.15%, 63.7 ± 2.53%, 44.9 ± 2.91%, 37.9 ± 2.40% and 30.6 ± 2.70%.

During a reaction of the Ap-site with an opposing base the duplex likely needs to distort relative to B-DNA geometry to allow for the new imine linkage. By introducing mismatched bases we can create local pockets of increased flexibility and also remove the penalty for breaking hydrogen bonds of the base-pairs, which may otherwise need to be broken to allow for cross-link formation. Mispairing the cross-linking base with native bases or with the universal base 5-nitroindole appears to uniformly increase the yield (Figure 9). This favorable mispair supports the notion that a "loose" duplex structure near the cross-linking site is favored.

Figure 9

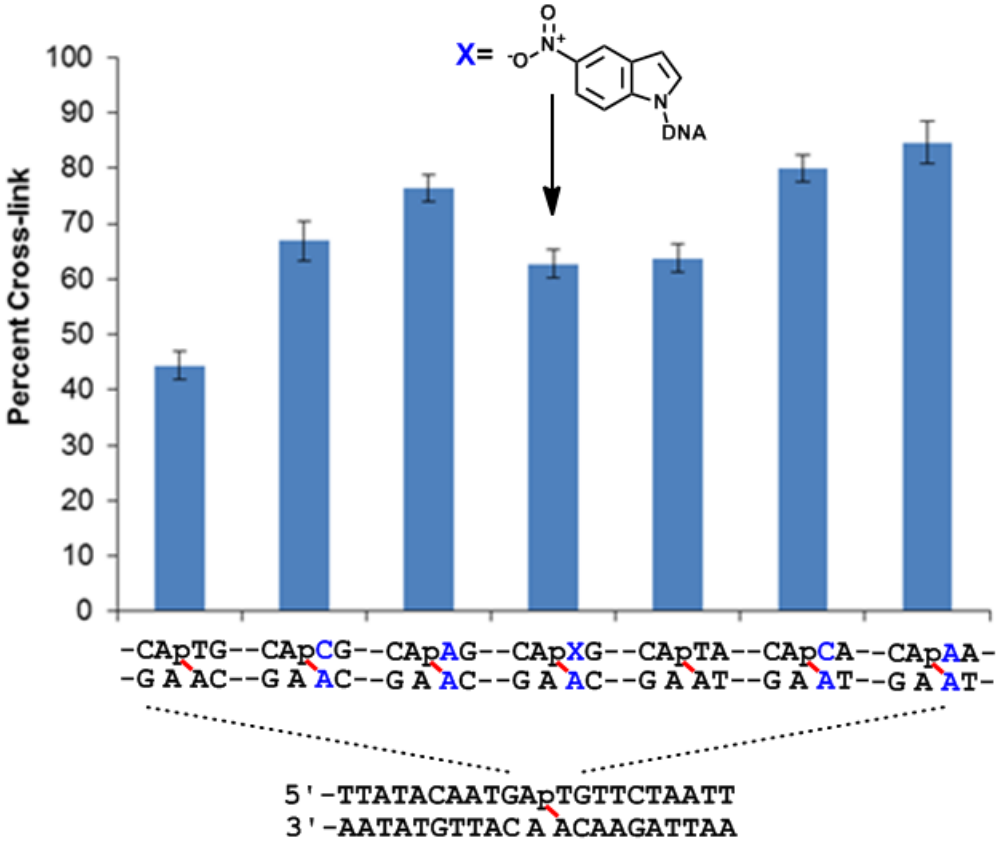


Figure 9. Percent DNA cross-links using PAGE analysis from radio-labeled duplexes. The cross-linking base was mispaired in high yielding (5'-CApT) dA-Ap cross-linking sequences. Left to right: 44.4 ± 2.59%, 66.9 ± 3.57%, 76.4 ± 2.49%, 62.7 ± 2.51%, 63.7 ± 2.53%, 80.0 ± 2.46% and 84.7 ± 3.85%.

In our earlier work we used hydroxyl radical footprinting to show the major site of cross-link attachment was to the dA and not dG,[10] however the reason for substantial increases in cross-link yield remained unknown so the role of the dG residue, 5' and opposing the Ap-site, was examined. We first explored whether the opposing G residue in this sequence can participate in cross-link formation. We examined this issue by conducting a cross-linking reaction under conditions of reductive amination that give good yields of the dG-Ap cross-link in 5'-CAp sequences. After addition of hydride to the cross-linking reaction in 5'-CApT sequences two bands with the same gel mobility as an unreduced dA and reduced dG cross-links can be observed (Figure 10). It appears that the previously seen dA cross-link is still intact, but in lower yields. Presumably the yields have been decreased due to reduction of the Ap-aldehyde as discussed earlier. Additionally, the new band displays the same gel mobility as a reduced dG cross-link. This is interesting because the weak, previously undetectable, chemistry for an Ap-dG cross-link can be driven forward by addition of a hydride source and will begin to compete with Ap-dA cross-links. Also highlighted in this reaction are the differences in Ap-dG and Ap-dA cross-links, which include chemical reactivity with hydride, gel mobility and preference for relative location in the duplex.

Figure 10

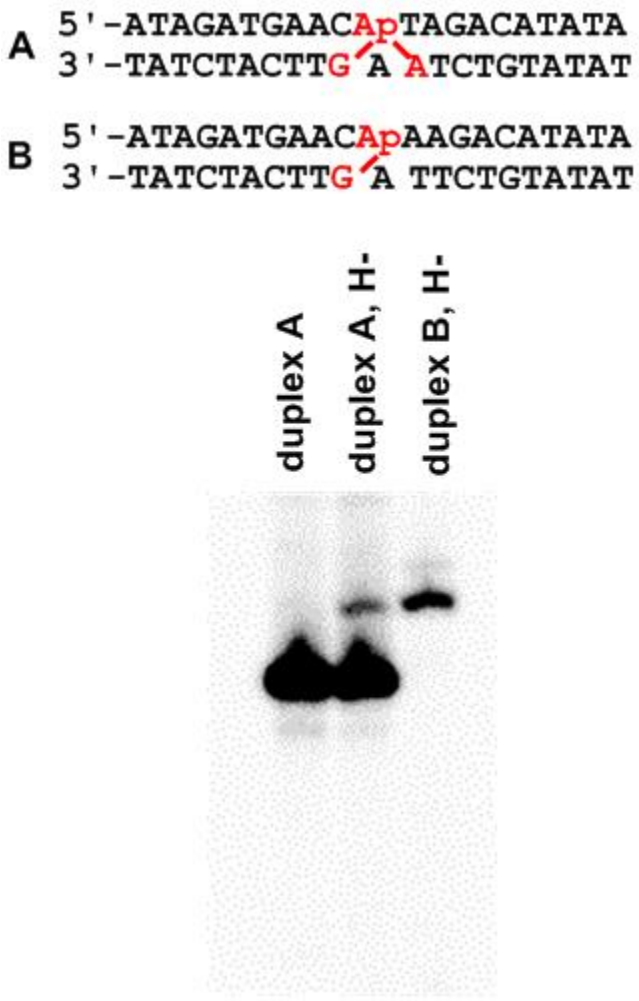
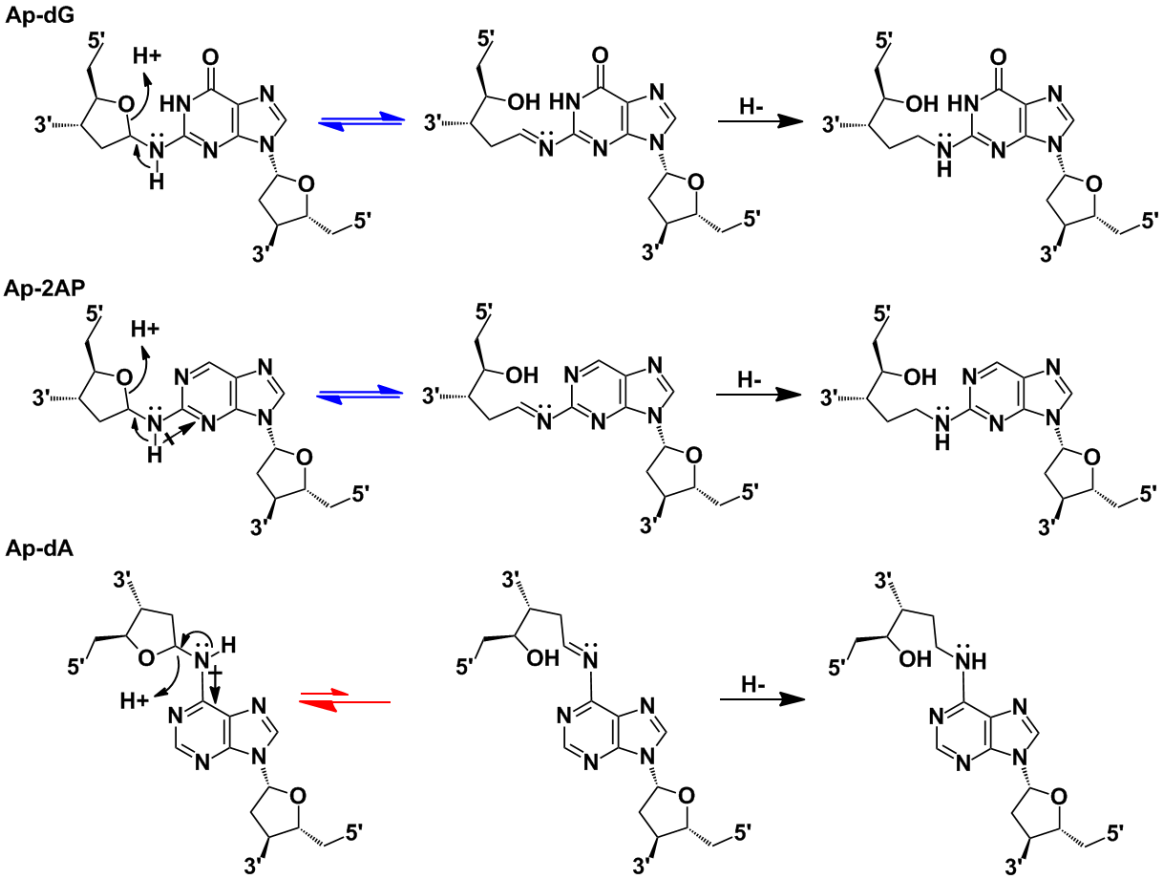


Figure 10. PAGE analysis of the 5'^{32}P-labeled duplexes **A** and **B**. The region shown is for the slow moving cross-link bands. The left lane is an incubation of duplex **A** in 50 mM HEPES (pH 7.0) and 100 mM NaCl for 120 hours and 37°C. The middle lane is duplex **A** in 750 mM NaOAc (pH 5.2) and 250 mM $NaCNBH_3$ for 16 hours and 37°C. The right lane is duplex **B** in 750 mM NaOAc (pH 5.2) and 250 mM $NaCNBH_3$ for 16 hours and 37°C.

The nucleophilicity of 2-aminopurine is likely very similar to dA, because they are positional isomers where the exocyclic amino group has just been moved from the 6 to 2 position, one might expect them to display similar resistance to reduction. However, besides the differences in the reactivities of the cross-linking base there are also environmental differences such as how exposed the imine is to solvent and how sterically

hindered the imine becomes. The interior of the DNA duplex is hydrophobic and therefore very little water will be available to act as a proton donor/acceptor, however dA and dG cross-links may be buried inside the helix to varying degrees keeping the dA cross-link more buried. In the same manner, the amount of steric hindrance for dA or dG cross-links may vary, thus allowing more or less access from water and hydride used for the reduction, where the dA cross-link may be exposed to less water and therefore less prone to ring-opening. Therefore, the cumulative effect of these differences may be that the dA cross-linking duplexes result in a buried structure with little access to water and it is difficult to drive it backward into an imine form (Scheme 4).

Scheme 4



Differences in gel mobility occur because the duplex structure is more or less bent due to local distortion at the cross-linking position. It is known that bent oligos can slow polyacrylamide gel migration, such as seen in poly dA-tracts of restriction fragments from kinetoplast DNA.[23–25] Even though our data is from denaturing polyacrylamide, structural or sequence variations still alter the gel mobility, possibly allowing different cross-linking sites or cross-link structures to be observed. Finally, the preference for relative location of the cross-linking base within the duplex is determined by the distance and angle of attack onto the Ap aldehyde (Figure 11). The distance must be short enough that the energy required to distort the duplex only a few kcals greater than that gained by forming the cross-link, while the attack must take a Bürgi-Dunitz angle onto the backside of the Ap aldehyde. In order to achieve a close enough distance for cross-link formation the duplex likely needs to pinch in the middle a pull the Ap-site closer to the reacting amino group. This pinch may simply require too much distortion and therefore too much energy when a more distal amino group is present such as that shown for dA reacting in the 5'-position (Scheme 3).

In addition to probing the reactivity of the opposing dG under conditions of reductive amination we tested the effect of replacing dG with inosine. The non-canonical base, inosine, lacks the N^2-exocyclic amino group and therefore this functional group will not be able to participate in any form of nucleophilic catalysis.[26] In fact cross-link formation was still very high for the inosine containing duplex (57.5 ± 5.10%) which indicates that the local duplex structure is primarily responsible for the observed cross-link yields in 5'-CApT sequences and not due to chemical reactivity of the exocyclic amino group.

Figure 11

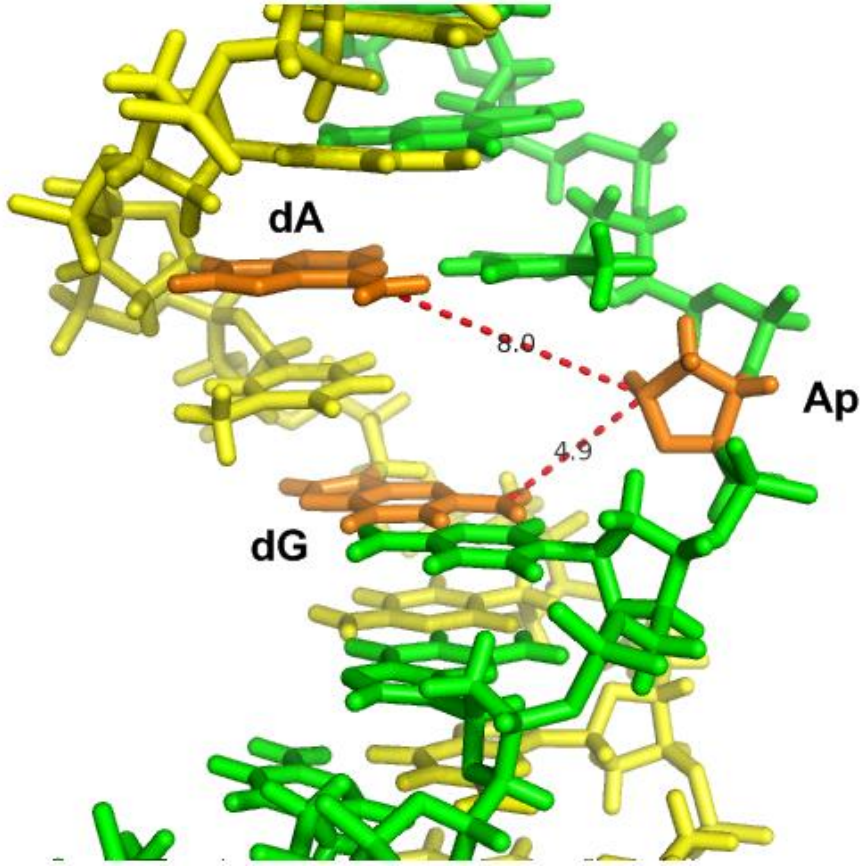


Figure 11. Pymol model depicting the distance (angstroms) between the Ap-site (orange), on the right, and the exocyclic amino groups of dA (organge) on the top left and dG (orange) on the bottom left. Struture was modified from PDB entry 2M2C.

3.5 Duplex Size and Chemical Additive Effects on dA-Ap Cross-link Formation

In addition to variations of the local sequence we examined the effect of placing these local sequence "packets" into a longer duplex. Ap-containing duplexes with 35 base pairs have a melting temperature around 53°C as opposed to 41°C for the previously studied 21mer duplexes. A partially melted duplex may contain partially hybridized "nucleation" complexes as some of the population dissociates and re-hybridizes. Therefore, using a duplex with a melting point significantly above the incubation

temperature, we queried whether there would be a change in yield or formation rate due to decreased dynamics. However, testing of several 35mer duplexes revealed similar yields and formation rates compared to their 21mer counterparts, indicating there is not a large effect on cross-link formation due to the use of a longer duplex (Figures 12 and 13).

Figure 12

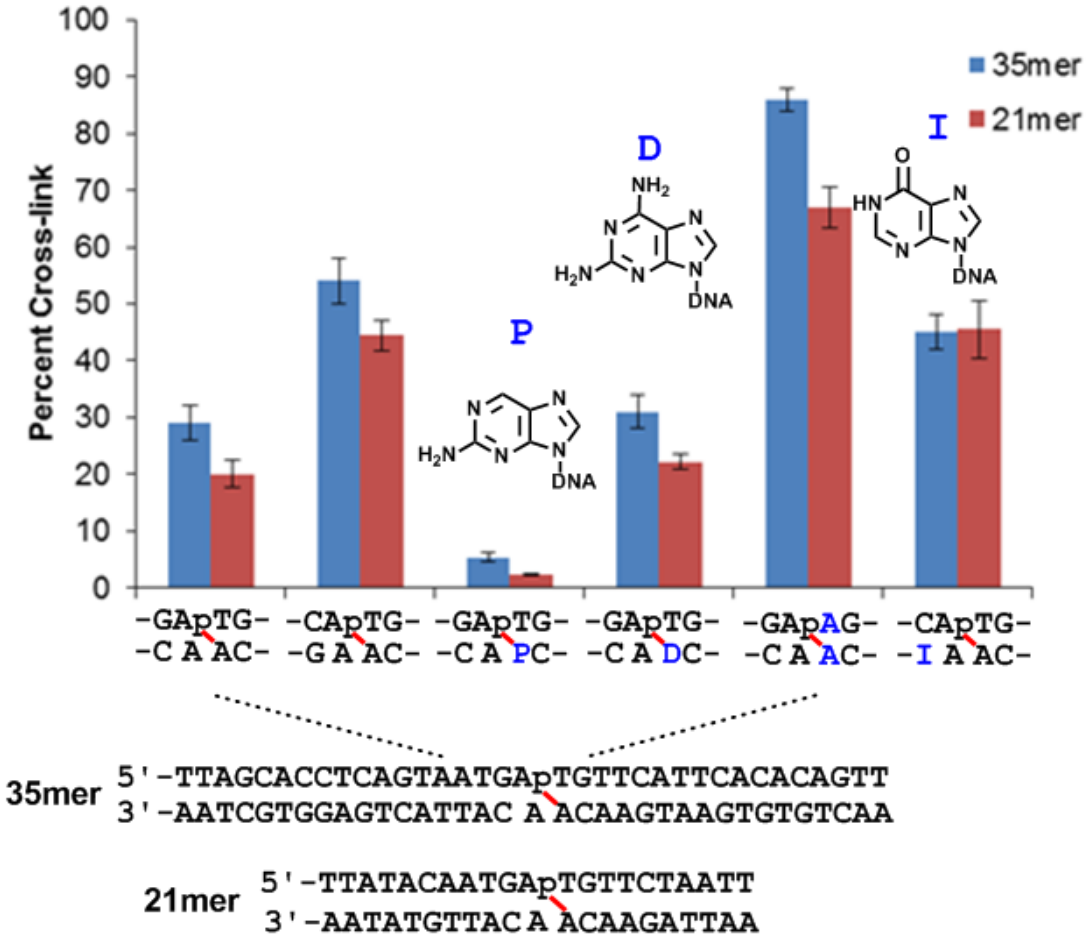


Figure 12. Percent DNA cross-links using PAGE analysis from radio-labeled duplexes. The local cross-linking sequence was varied in 35mer dA-Ap cross-linking duplexes and compared to those of its 21mer counterparts. Unmodified 30.7 ± 3.31%, 2-aminopurine 5.4 ± 0.64%, 2,6-diaminopurine 31.5 ± 3.56%, A/A mismatch 86.5 ± 1.98% and Inosine 45.4 ± 3.49%.

Figure 13

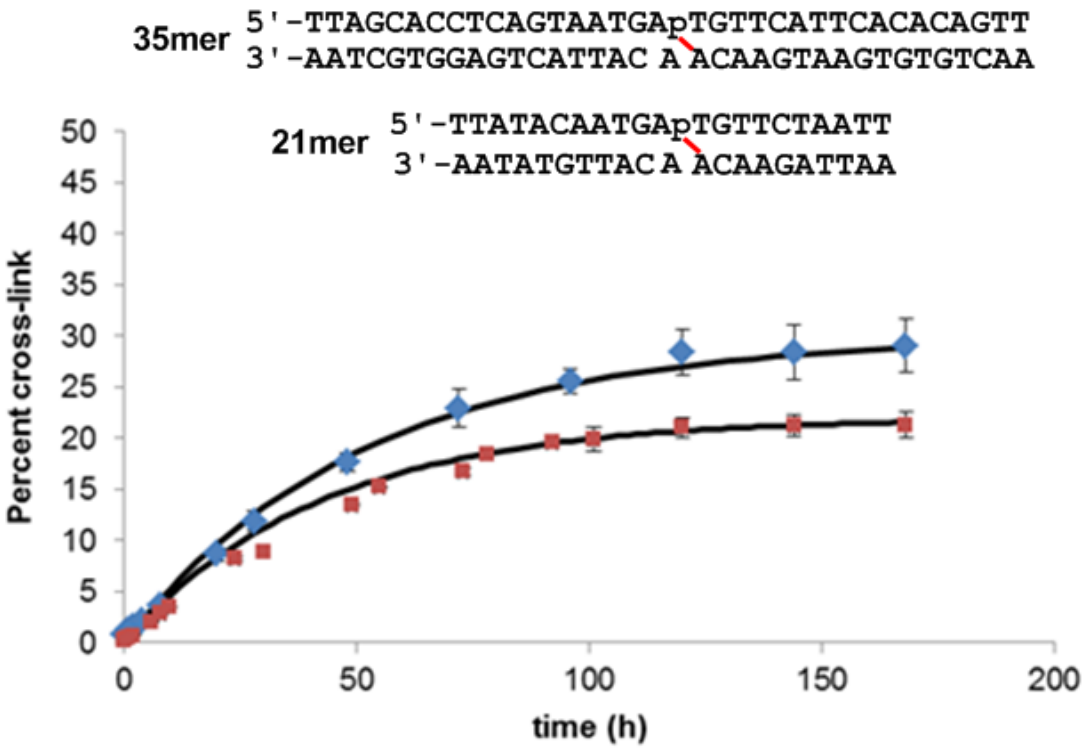


Figure 13. Time course comparing the formation of the 35mer and 21mer duplexs shown above. The duplexes were incubated in HEPES buffer (50 mM, pH 7) containing NaCl (100 mM) at 37 ˚C. At each time point aliquots were removed and frozen at -20 ˚C. The ^{32}P-labeled 2'-deoxyoligonucleotides were resolved on a denaturing polyacrylamide gel. The radioactivity in each band quantitatively measured by phosphorimager analysis, and the yield of slow-migrating (cross-linked DNA) band plotted versus time. Time = 0 was defined as the end of the 45 minute reaction of the uracil-containing 2'-deoxyoligonucleotide duplex with UDG to generate the Ap-containing 2'-deoxyoligonucleotide duplex.

We next examined the effect of duplex structure on cross-link formation by the use of chemical additives. Many small molecules can affect the structure of DNA by non-covalent binding due to hydrogen bonding, pi-stacking or hydrophobic interactions. It has previously been demonstrated that the addition of purines can improve the yield of weak low molecular weight interstrand cross-links arising from a similar reaction between a beta-eliminated C4'-oxidized abasic site and an opposing nucleobase.[27]

However, the addition of purines did not appear to increase the yield of cross-links in the intact duplex. In a similar manner, purines were not able to "rescue" low yielding dA cross-links (5-10%) such as those produced by duplexes **C** and **D** (Figure14). Earlier it was discussed that in some Ap-containing duplexes the Ap-site and its opposing base rotate out of the duplex and allow the remaining bases to collapse and fill that hole.[18] Therefore it was possible that DNA intercalators, which stack in between base pairs,[28] could fill this hole and prevent a collapse of the duplex structure. When the intercalators ethidium bromide or proflavine were added they did not improve low yielding cross-links they instead prevent cross-link formation (Figure 14).

Figure 14

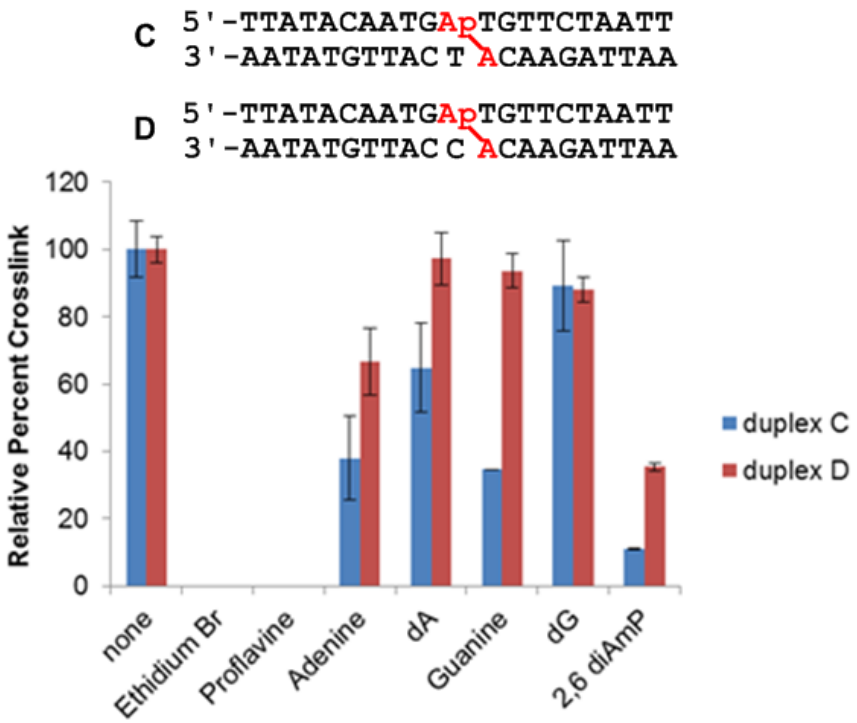


Figure 14. Relative percent cross-link is 100% when no additives are present. Additive concentrations used: 50 μM ethidium bromide, 50 μM proflavine, 100 μM adenine, 100 μM deoxyadenosine, *10 μM guanine, 100 μM deoxyguanosine, 100 μM 2,6-diaminopurine. (* poor solubility prevented a more concentrated stock solution of guanine from being prepared)

When intercalators stack between the DNA base pairs they also rigidify the duplex. This increased rigidity may prevent the proper alignment of the Ap-site and an opposing nucleobase; however it is also possible that the intercalators physically block the space between the Ap-site and this opposing base and cause a general structural distortion to the duplex.[28] To further test if increased rigidity of a duplex can prevent cross-link formation we tested another well studied DNA binding molecule. Molecules like netropsin do not stack between base pairs, but instead bind to the minor groove of DNA.[29–31] By this mode of binding these minor groove binders should not physically block the space between an Ap-site a base on the opposing strand and they do not distort the overall structure of the duplex.[28] Netropsin has a high binding constant in regions of at least four consecutive A/T base pairs (1-4x10^6 M^{-1}) and shows very weak binding in G/C regions (Scheme 5).[32,33] With this sequence specificity in mind, we designed three duplexes, one with no AT binding region, one with a binding region overlapping the cross-link site and one with a binding region adjacent to the cross-link site. Cross-link formation in these three duplexes showed striking effects when netropsin was added (Figure 15).

Scheme 5

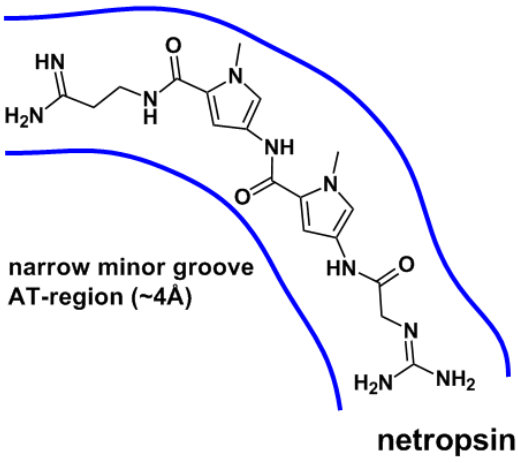


Figure 15

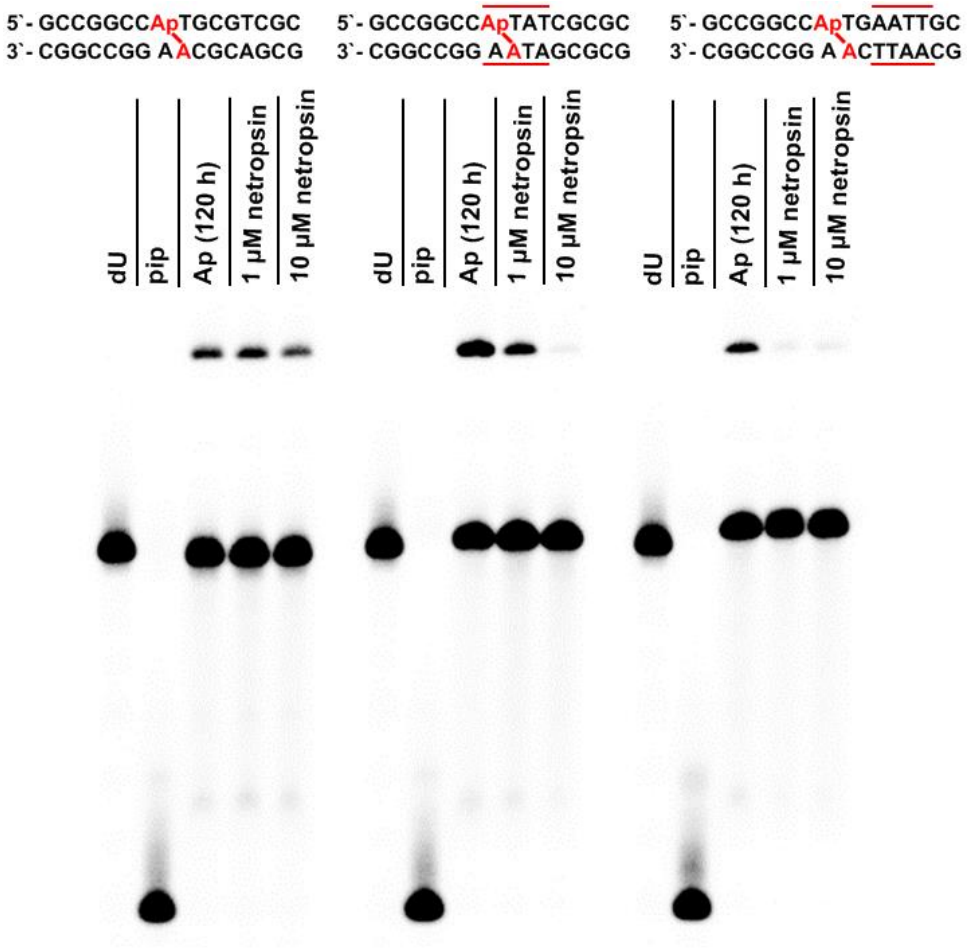


Figure 15. PAGE analysis of the 5'^{32}P-labeled duplexes shown above. In each block the first lane is the uracil containing duplex shown above. The second lane is treatment of the Ap-containing duplex with 1M piperidine. The third lane is incubation of the Ap-containing duplex in 50 mM HEPES (pH 7.0) and 100 mM NaCl for 120 hours and 37°C. The forth lane is the same incubation conditions plus the addition of 1 µM netropsin and the fifth lane is with the addition of 10 µM netropsin. The red bars around the duplex represent a netropsin binding site. Cross-link yields, **lane 3**- 14.2 ± 0.4, **lane 4**- 13.8 ± 0.4, **lane 5**- 12.5 ± 0.5, **lane 8**- 25.5 ± 0.7, **lane 9**- 11.4 ± 0.6, **lane 10**- 2.9 ± 0.2, **lane 13**- 11.4 ± 0.5, **lane 14**- 3.7 ± 0.3, **lane 15**- 3.1 ± 0.2

When there is no AT binding site there was almost no decrease in cross-link yields, however when the binding site was adjacent to the cross-linking site netropsin prevented nearly all cross-link formation. If the binding site overlaps the cross-linking region cross-link formation is still prevented by the addition of netropsin, but to a lesser

extent. Likely this region does not fit the netropsin molecule as well because it is more distorted from the typical B-DNA structure and therefore not a very good binding site. These results suggest that binding of netropsin near the cross-linking site rigidifies the DNA structure and prevents the mobility of the Ap-site and opposing nucleobase necessary for cross-link formation.

3.6 Conclusions

In this chapter we have demonstrated that interstrand cross-links between Ap-sites and an opposing nucleobase are a family of lesions and occur in a wide array of sequence context. In fact at least some dG or dA cross-links were detected in all 5'-CAp and 5'-ApT native duplexed sequences. Furthermore, we have shown that the cross-link yields are highly dependent on the local sequence, which likely affects the amount and type of distortion offered by that particular duplex. One sequence context which appears to create a favorable duplex structure is 5'-CApT sequences where the adenine residue 3' and opposing the Ap-site will form high yields of cross-link. Finally, we probed the effect of small molecules on cross-link formation. Intercalators which distort the duplex structure prevent any cross-link from forming. Minor groove binders which do not distort, but rigidify the duplex also prevent cross-link formation when they have a binding site near the Ap-site. Overall, these features of Ap-derived interstrand cross-links paint them to be an important lesion with the possibility of creating high amounts of biological dysfunction during replication, transcription and DNA repair processes.

3.7 Experimental

The same general procedure for generating cross-links in chapter one was used throughout this chapter.

3.8 Supporting Information

Below is a complete list of tables for sequences where the yield of Ap-derived interstrand cross-link was measured by 5'^{32}P radiolabeled oligonucleotide duplexes which are resolved by PAGE and the phosphorimager densitometry data is analyzed.

X = abasic site
* This data was collected by fellow Gates group member, Kevin Johnson.

Name	Sequence	Yield	Error	Red.	Error	Notes
	Vary Base Opposing Ap-site					
S2C	5`-ATA GAT GAA C**X**A AGA CAT ATA 3`-TAT CTA CTT G**a**T TCT GTA TAT	2.1	0.4	18	1	Purines opposing the Ap-site give higher yields. The reduced yield with dC-opposing is very high.
*	5`-ATA GAT GAA C**X**A AGA CAT ATA 3`-TAT CTA CTT G**g**T TCT GTA TAT	1	0.1	38.6	0.5	
*	5`-ATA GAT GAA C**X**A AGA CAT ATA 3`-TAT CTA CTT G**t**T TCT GTA TAT	0.4	0.1	13.1	0.7	
*	5`-ATA GAT GAA C**X**A AGA CAT ATA 3`-TAT CTA CTT G**c**T TCT GTA TAT	0.4	0.1	40	1	
	Vary Base 3' of the Ap-site					
*	5`-ATA GAT GAA C**Xa** AGA CAT ATA 3`-TAT CTA CTT GA**t** TCT GTA TAT	2.1	0.4	18	1	
*	5`-ATA GAT GAA C**Xg** AGA CAT ATA 3`-TAT CTA CTT GA**c** TCT GTA TAT	1.4	0.2	13.8	0.4	
*	5`-ATA GAT GAA C**Xc** AGA CAT ATA 3`-TAT CTA CTT GA**g** TCT GTA TAT	1	0.1	13.0	0.3	
*	5`-ATA GAT GAA C**X**t AGA CAT ATA 3`-TAT CTA CTT GAA TCT GTA TAT	64	3	15	1	Footprinted as a dA cross-link. When reduced the dG cross-link band appears, but there is still a strong band for the dA cross-link.
				5	1	
	Vary Base 5' of the Ap-site					
S2C	5`-ATA GAT GA**a** C**X**A AGA CAT ATA 3`-TAT CTA CT**t** GAT TCT GTA TAT	2.1	0.4	18	1	No particularly strong trends.
*	5`-ATA GAT GA**t** C**X**A AGA CAT ATA 3`-TAT CTA CT**a** GAT TCT GTA TAT	1.7	0.1	8.7	0.3	
*	5`-ATA GAT GA**g** T**X**A AGA CAT ATA 3`-TAT CTA CT**c** GAT TCT GTA TAT	1.3	0.2	18	1	
*	5`-ATA GAT GA**c** T**X**A AGA CAT ATA 3`-TAT CTA CT**G** GAT TCT GTA TAT	1.0	0.1	15.4	0.3	

Name	Sequence	Yield	Error	Red.	Error	Notes
	Mismatches					
S2C	5`-ATA GAT GAA C**X**A AGA CAT ATA 3`-TAT CTA CTT GAT TCT GTA TAT	2.1	0.4	18	1	dG mismatches 5' the AP-site or opposing the cross-linking 'G' improve the yield. However, they seem to be resistant to reductive amination?
S2/E3C	5`-ATA GAT GAA C**X**A AGA CAT ATA 3`-TAT CTA CTC GAT TCT GTA TAT	8	1	2.9	0.3	
MMG1	5`-ATA GAT GAA T**X**A AGA CAT ATA 3`-TAT CTA CTT GAT TCT GTA TAT	16	1	4.2	0.3	MMG1 footprinted as a dG cross-link
*	5`-ATA GAT GA**a** C**X**a AGA CAT ATA 3`-TAT CTA CT**t** GA**c** TCT GTA TAT	24	5	36.4	0.3	Footprinted as a dC cross-link
*	5`-ATA GAT GA**t** C**X**a AGA CAT ATA 3`-TAT CTA CT**a** GA**g** TCT GTA TAT	3.8	0.8	35	2	
*	5`-ATA GAT GA**g** T**X**t AGA CAT ATA 3`-TAT CTA CT**c** GA**t** TCT GTA TAT	3.4	0.3	19	2	
*	5`-ATA GAT GA**c** T**X**t AGA CAT ATA 3`-TAT CTA CT**G** GA**g** TCT GTA TAT	6	1	21.7	0.3	
	Non-native bases					
S6C1	5`-ATA GAT GAA T**X**A AGA CAT ATA 3`-TAT CTA CTT **P**GT TCT GTA TAT	42	4	N/A	N/A	**P** = 2-aminopurine, which is a good cross-linker (electronically similar to dA) and undergoes reductive amination very well
S6C2	5`-ATA GAT GAA T**X**A AGA CAT ATA 3`-TAT CTA CTT **P**AT TCT GTA TAT	27	1	89	3	**P** = 2-aminopurine, which is a good cross-linker (electronically similar to dA) and undergoes reductive amination very well
F1/F1mod	5`-TTA TAC AAT G**X**T GTT CTA ATT 3`-AAT ATG TTA CA**P** CAA GAT TAA	2.3	0.2	6.0	0.3	**P** = 2-aminopurine, exocyclic amine is not in the correct position, but still increases with reductive amination

Name	Sequence	Yield	Error	Red.	Error	Notes
	Non-native bases					
IsoF1	5`-TTA TAC AAT G**XC*** GTT CTA ATT 3`-AAT ATG TTA CA**G*** CAA GAT TAA	ND	N/A	ND	N/A	**C*/G*** = isodC/IsodG, which should have been placed the exocyclic amine through the center of the duplex toward the AP-site. Not sure why there is actually zero cross-link
BU1	5`-TTA TAC AAT G**X** GTT CTA ATT 3`-AAT ATG TTA CA CAA GAT TAA **A**	5	2	NA	NA	A bulge does not improve the yield of this native sequence, but even with such a "big" distortion noticeable cross-link can form
UN1	5`-TTA TAC AAT G**XL** GTT CTA ATT 3`-AAT ATG TTA CAA CAA GAT TAA	63	3	NA	NA	**L** = 5-nitroindole, a universal base, which could allow more mobility of the "paired" cross-linking base. The yield is very good, but not better than an A:A mismatch
*	5`-ATA GAT GAA C**X**A AGA CAT ATA 3`-TAT CTA CTT **I**AT TCT GTA TAT	ND	N/A	ND	N/A	**I=** inosine
*	5`-ATA GAT GAA $^{Me}$C**X**A AGA CAT ATA 3`-TAT CTA CTT G^{Me}C**T** TCT GTA TAT	6.0	0.3	54	6	^{Me}C = 5-methylcytosine, has slightly higher unreduced yields than the unmethylated duplex and much higher reduced yields
*	5`-ATA GAT GAA C**X**T AGA CAT ATA 3`-TAT CTA CTT **I**AA TCT GTA TAT	58	5	N/A	N/A	**I=** inosine
*	5`-ATA GAT GAA C**X**A AGA CAT ATA 3`-TAT CTA CTT **D**AT TCT GTA TAT	5.7	0.7	95.5	1	**D=** 2,6-diaminopurine, incredibly high reduced yield
*	5`-TTA TAC AAT G**X**T GTT CTA ATT 3`-AAT ATG TTA CA**D** CAA GAT TAA	3.8	0.1	3.4	0.1	**D=** 2,6-diaminopurine, 2 bands possibly from each of the amino groups on the **=** 2,6-diaminopurine or possibly one from an opposing dA. Interestingly none of them undergo reductive amination well, while the duplex above with **=** 2,6-diaminopurine in the 'G' cross-linking position is very strong at this reductive amination.
		22	1	7	1	

Name	Sequence	Yield	Error±%	Notes
	Vary Base Opposing Ap-site			
F1C1	5`-TTA TAC AAT G**X**T GTT CTA ATT 3`-AAT ATG TTA C**a**A CAA GAT TAA	20	2	Purines are better, though 'T' is not much worse
F1C2	5`-TTA TAC AAT G**X**T GTT CTA ATT 3`-AAT ATG TTA C**t**A CAA GAT TAA	12	1	
F1C3	5`-TTA TAC AAT G**X**T GTT CTA ATT 3`-AAT ATG TTA C**g**A CAA GAT TAA	15	1	
F1C4	5`-TTA TAC AAT G**X**T GTT CTA ATT 3`-AAT ATG TTA C**c**A CAA GAT TAA	4.6	0.3	
	Vary Base Flanking 3' of the Ap-site			
F1C1	5`-TTA TAC AAT G**X**T **g**TT CTA ATT 3`-AAT ATG TTA CAA **c**AA GAT TAA	20	2	Pyrimidines flanking the cross-linked dA on its 5' side are much better
F2C	5`-TTA TAC AAT G**X**T **c**TT CTA ATT 3`-AAT ATG TTA CAA **g**AA GAT TAA	5	1	
F3C	5`-TTA TAC AAT G**X**T **t**TT CTA ATT 3`-AAT ATG TTA CAA **a**AA GAT TAA	2.0	0.2	
F5C	5`-TTA TAC AAT G**X**T **a**TT CTA ATT 3`-AAT ATG TTA CAA **t**AA GAT TAA	24	2	
	Vary Base Flanking 5' of the Ap-site			
F1C1	5`-TTA TAC AAT **g**X**X**T GTT CTA ATT 3`-AAT ATG TTA **c**AA CAA GAT TAA	20	2	G:C pair 5' the AP-site is good. Mis-match sequences also indicate stabilization here is good
F6C	5`-TTA TAC AAT **a**X**X**T GTT CTA ATT 3`-AAT ATG TTA **t**AA CAA GAT TAA	1.4	0.4	
F7C	5`-TTA TAC AAT **t**X**X**T GTT CTA ATT 3`-AAT ATG TTA **a**AA CAA GAT TAA	1.7	0.2	
F8C	5`-TTA TAC AAT **c**X**X**T GTT CTA ATT 3`-AAT ATG TTA **g**AA CAA GAT TAA	44	3	

Name	Sequence	Yield	Error±%	Notes
	Super-sequence, Vary Base Opposing Ap-site			
E3C	5`-ATA GAT GAA C**X**T AGA CAT ATA 3`-TAT CTA CTT G**a**A TCT GTA TAT	64	3	G-opposing is actually the lowest yield. The other opposing bases follow the same trends as seen in 5'-GXT sequences
*	5`-ATA GAT GAA C**X**T AGA CAT ATA 3`-TAT CTA CTT G**t**A TCT GTA TAT	38	2	
*	5`-ATA GAT GAA C**X**T AGA CAT ATA 3`-TAT CTA CTT G**g**A TCT GTA TAT	31	3	
*	5`-ATA GAT GAA C**X**T AGA CAT ATA 3`-TAT CTA CTT G**c**A TCT GTA TAT	45	3	
	Super-sequence, Vary Base Flanking 3' of the Ap-site			
E3C	5`-ATA GAT GAA C**X**T **a**GA CAT ATA 3`-TAT CTA CTT GAA **t**CT GTA TAT	64	3	Placing a pyrimidine 5' of the cross-linked dA produces high yields
*	5`-ATA GAT GAA C**X**T **t**GA CAT ATA 3`-TAT CTA CTT GAA **a**CT GTA TAT	17.6	0.4	
*	5`-ATA GAT GAA C**X**T **g**GA CAT ATA 3`-TAT CTA CTT GAA **c**CT GTA TAT	49	2	
*	5`-ATA GAT GAA C**X**T **a**GA CAT ATA 3`-TAT CTA CTT GAA **t**CT GTA TAT	17	1	
	Super-sequence, Vary Base Flanking 5' of the Ap-site			
E3C	5`-ATA GAT GAA **c**X**c**T AGA CAT ATA 3`-TAT CTA CTT **g**AA TCT GTA TAT	64	3	No clear trends, the c/g base pair is needed to produce the high yielding "super" cross-links
*	5`-ATA GAT GAA **a**X**t**T AGA CAT ATA 3`-TAT CTA CTT **t**AA TCT GTA TAT	35.0	0.6	
*	5`-ATA GAT GAA **g**X**c**T AGA CAT ATA 3`-TAT CTA CTT **c**AA TCT GTA TAT	18	0.4	
*	5`-ATA GAT GAA **t**X**a**T AGA CAT ATA 3`-TAT CTA CTT **a**AA TCT GTA TAT	8.6	0.1	

Name	Sequence	Yield	Error±%	Notes
	Pyrimidines Opposing the Ap-site			
F8C	5`-TTA TAC AAT C**X**T GTT CTA ATT 3`-AAT ATG TTA G**a**A CAA GAT TAA	44.4	2.59	Pyrimidines opposing "good" sequences give lower yields, however the yields are still much better than a generally "bad" sequence
F8C2	5`-TTA TAC AAT C**X**T GTT CTA ATT 3`-AAT ATG TTA G**c**A CAA GAT TAA	6.4	2.41	
F8C3	5`-TTA TAC AAT C**X**T GTT CTA ATT 3`-AAT ATG TTA G**t**A CAA GAT TAA	7.2	2.15	
E3C	5`-ATA GAT GAA C**X**T AGA CAT ATA 3`-TAT CTA CTT G**a**A TCT GTA TAT	63.7	2.53	
E3C2	5`-ATA GAT GAA C**X**T AGA CAT ATA 3`-TAT CTA CTT G**c**A TCT GTA TAT	44.9	2.91	
E3C3	5`-ATA GAT GAA C**X**T AGA CAT ATA 3`-TAT CTA CTT G**t**A TCT GTA TAT	37.9	2.40	
F5C	5`-TTA TAC AAT G**X**T ATT CTA ATT 3`-AAT ATG TTA C**a**A TAA GAT TAA	24.2	2.06	
F5C2	5`-TTA TAC AAT G**X**T ATT CTA ATT 3`-AAT ATG TTA C**t**A TAA GAT TAA	4.9	1.33	
F7C	5`-TTA TAC AAT T**X**T GTT CTA ATT 3`-AAT ATG TTA A**a**A CAA GAT TAA	1.7	0.17	Pyrimidines opposing "bad" sequences don't necessarily make them worse
F7C2	5`-TTA TAC AAT T**X**T GTT CTA ATT 3`-AAT ATG TTA A**t**A CAA GAT TAA	3.3	1.04	

Name	Sequence	Yield	Error±%	Notes
	Mismatches			
F1C1	5`-TTA TAC AAT G**X**T GTT CTA ATT 3`-AAT ATG TTA CAA CAA GAT TAA	20	2	Mismatches on the 5' and 3' sides of the reaction may distort the duplex in an unfavorable way (allow Ap-site to flip out of the duplex). Though only a couple of these have been tried.
F1/F5C	5`-TTA TAC AAT G**X**T **g**TT CTA ATT 3`-AAT ATG TTA CAA **t**AA GAT TAA	15.0	0.7	
F1/F6C	5`-TTA TAC AAT **g**X**t** GTT CTA ATT 3`-AAT ATG TTA **t**AA CAA GAT TAA	ND	N/A	
F8C	5`-TTA TAC AAT C**X**T GTT CTA ATT 3`-AAT ATG TTA GAA CAA GAT TAA	44	3	Mismatching the cross-linking base improves yield. Adds flexibility to the cross-lining base and removes the need to "break" those hydrogen bonds when forming the cross-link. A:A mismatch appears to be the best, presumably gives the most freedom of movement for the cross-linking base. G-mismatches where not examined.
MMA1	5`-TTA TAC AAT C**X**c GTT CTA ATT 3`-AAT ATG TTA GA**a** CAA GAT TAA	67	4	
F12C	5`-TTA TAC AAT C**X**a GTT CTA ATT 3`-AAT ATG TTA GA**a** CAA GAT TAA	76	2	
E3C	5`-ATA GAT GAA C**X**T AGA CAT ATA 3`-TAT CTA CTT GAA TCT GTA TAT	64	3	
E5C	5`-ATA GAT GAA C**X**c AGA CAT ATA 3`-TAT CTA CTT GA**a** TCT GTA TAT	80	2	
S2/E3C	5`-ATA GAT GAA C**X**a AGA CAT ATA 3`-TAT CTA CTT GA**a** TCT GTA TAT	85	4	
*	5`-ATA GAT GAA **a**X**t** AGA CAT ATA 3`-TAT CTA CTT **g**AA TCT GTA TAT	4.6	0.1	
*	5`-ATA GAT GAA **g**X**t** AGA CAT ATA 3`-TAT CTA CTT **g**AA TCT GTA TAT	2.1	0.1	
*	5`-ATA GAT GAA C**X**T **a**GA CAT ATA 3`-TAT CTA CTT GAA **a**CT GTA TAT	2.3	0.1	
*	5`-ATA GAT GAA C**X**a **a**GA CAT ATA 3`-TAT CTA CTT GA**a** **a**CT GTA TAT	83	2	

Name	Sequence	Yield	Error±%	Notes
	"Zippered" duplexes			
Z1C1	5`-CAG TTA TAC AAG C**Xa** AGC CTA ATT CGA 3`-GTC AAT ATG TTC GA**a** TCG GAT TAA GCT	47.6	3.44	These duplexes are based off PDB structures, which indicate the non-AP sequences create zippered structures where the bases of opposing strands slip on top of one another. We do not know if this happens here because there is no crystal structure, but the yields are very high.
Z1C2	5`-CAG TTA TAC AAG C**Xa** aGC CTA ATT CGA 3`-GTC AAT ATG TTC GA**a** **g**CG GAT TAA GCT	85.6	2.25	
Z2C	5`-CAG TTA TAC AAG C**Xa** **aa**G CTA ATT CGA 3`-GTC AAT ATG TTC GA**a** **ag**C GAT TAA GCT	78.4	1.98	

3.9 References

(20) Bird, A. *Genes Dev.* **2002**, *16*, 6–21.

(22) Christov, P. P.; Brown, K. L.; Kozekov, I. D.; Stone, M. P.; Harris, T. M.; Rizzo, C. J. *Chem. Res. Toxicol.* **2008**, *21*, 2324–2333.

(17) Chung, F.-L.; Young, R.; Hecht, S. S. *Cancer Res.* **1984**, *44*, 990–995.

(25) Diekmann, S. *FEBS Lett.* **1986**, *195*, 53–56.

(26) Dirksen, A.; Dirksen, S.; Hackeng, T. M.; Dawson, P. E. *J. Am. Chem. Soc.* **2006**, *128*, 15602–15603.

(7) Dronkert, M. L. G.; Kanaar, R. *Mutat. Res. Repair* **2001**, *486*, 217–247.

(30) Drozdowska, D. In *Breast Cancer - Current and Alternative Therapeutic Modalities*; Gunduz, E., Ed.; InTech, 2011.

(8) Dutta, S.; Chowdhury, G.; Gates, K. S. *J. Am. Chem. Soc.* **2007**, *129*, 1852–1853.

(19) Ehrlich, M.; Gama-Sosa, M. A.; Huang, L.-H.; Midgett, R. M.; Kuo, K. C.; McCune, R. A.; Gehrke, C. *Nucleic Acids Res.* **1982**, *10*, 2709–2721.

(11) Gates, K. S. *Chem. Res. Toxicol.* **2009**, *22*, 1747–1760.

(13) Gates, K. S.; Nooner, T.; Dutta, S. *Chem. Res. Toxicol.* **2004**, *17*, 839–856.

(4) Greenberg, R. B.; Alberti, M.; Hearst, J. E.; Chua, M. A.; Saffran, W. A. *J. Biol. Chem.* **2001**, *276*, 31551–31560.

(12) Guillet, M.; Boiteux, S. *Mol. Cell. Biol.* **2003**, *23*, 8386–8394.

(18) Hoehn, S. T.; Turner, C. J.; Stubbe, J. *Nucleic Acids Res.* **2001**, *29*, 3413–3423.

(1) Hoeijmakers, J. H. J. *N. Engl. J. Med.* **2009**, *361*, 1475–1485.

(9) Johnson, K. M.; Price, N. E.; Wang, J.; Fekry, M. I.; Dutta, S.; Seiner, D. R.; Wang, Y.; Gates, K. S. *J. Am. Chem. Soc.* **2013**, *135*, 1015–1025.

(24) Koo, H.-S.; Wu, H.-M.; Crothers, D. M. *Nature* **1986**, *320*, 501–506.

(31) Lewis, E. A.; Munde, M.; Wang, S.; Rettig, M.; Le, V.; Machha, V.; Wilson, W. D. *Nucleic Acids Res.* **2011**, *39*, 9649–9658.

(23) Marini, J. C.; Levene, S. D.; Crothers, D. M.; Englund, P. T. *Proc. Natl. Acad. Sci.* **1982**, *79*, 7664–7668.

(21) McGhee, J. D.; Von Hippel, P. H. *Biochemistry (Mosc.)* **1975**, *14*, 1281–1296.

(5) Mitchell, J. R.; Hoeijmakers, J. H.; Niedernhofer, L. J. *Curr. Opin. Cell Biol.* **2003**, *15*, 232–240.

(14) Nakamura, J.; Swenberg, J. A. *Cancer Res.* **1999**, *59*, 2522–2526.

(29) Neidle, S. *Nat. Prod. Rep.* **2001**, *18*, 291–309.

(10) Price, N. E.; Johnson, K. M.; Wang, J.; Fekry, M. I.; Wang, Y.; Gates, K. S. *J. Am. Chem. Soc.* **2014**, *136*, 3483–3490.

(33) Ren, J.; Chaires, J. B. *Biochemistry* **1999**, *38*, 16067–16075.

(28) Rescifina, A.; Zagni, C.; Varrica, M. G.; Pistarà, V.; Corsaro, A. *Eur. J. Med. Chem.* **2014**, *74*, 95–115.

(32) Rosu, F. *Nucleic Acids Res.* **2002**, *30*, 82e – 82.

(15) Sattsangi, P. D.; Barrio, J. R.; Leonard, N. J. *J. Am. Chem. Soc.* **1980**, *102*, 770–774.

(27) Sczepanski, J. T.; Jacobs, A. C.; Majumdar, A.; Greenberg, M. M. *J. Am. Chem. Soc.* **2009**, *131*, 11132–11139.

(3) Schumacher, B.; van der Pluijm, I.; Moorhouse, M. J.; Kosteas, T.; Robinson, A. R.; Suh, Y.; Breit, T. M.; van Steeg, H.; Niedernhofer, L. J.; van IJcken, W.;

Bartke, A.; Spindler, S. R.; Hoeijmakers, J. H. J.; van der Horst, G. T. J.; Garinis, G. A. *PLoS Genet* **2008**, *4*, e1000161.

(6) Shen, X.; Li, L. *Environ. Mol. Mutagen.* **2010**, *51*, 493–499.

(16) Upadhyaya, P.; Hecht, S. S. *Chem. Res. Toxicol.* **2008**, *21*, 2164–2171.

(2) Van der Pluijm, I.; Garinis, G. A.; Brandt, R. M. C.; Gorgels, T. G. M. F.; Wijnhoven, S. W.; Diderich, K. E. M.; de Wit, J.; Mitchell, J. R.; van Oostrom, C.; Beems, R.; Niedernhofer, L. J.; Velasco, S.; Friedberg, E. C.; Tanaka, K.; van Steeg, H.; Hoeijmakers, J. H. J.; van der Horst, G. T. J. *PLoS Biol* **2006**, *5*, e2.

Vitae

Nathan Price was born to Roger and Violet Price in St. Joseph, Missouri where he lived the first eighteen years of his life. After graduating from St. Joseph Central High School in May 2003, he went on to study at the University of Missouri-Columbia. While at the University of Missouri he studied Chemical Engineering and participated in undergraduate research in the areas of chemical-biology for Professor Gates. In December 2008, Nathan graduated *Cum Laude* from University of Missouri with a B.S. in Chemical Engineering. From there he went to the University of Missouri, where he was accepted into the PhD program for Chemistry. In the Spring of 2010 he joined Professor Gates research lab to start his research towards completion of the PhD degree. After graduating with a PhD in Chemistry in December 2014, he went on to the University of California-Riverside where he accepted a post-doctoral position working for Professor Wang studying chemical toxicology.