Effects of Local Sequence on the Formation of Interstrand DNA Cross-Links Derived from the Reaction of an Abasic Site with a Cytosine Residue on the Opposing Strand

---

A Thesis

presented to

the Faculty of the Graduate School

University of Missouri

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

---

by

Luke Edward Pierce

Prof. Kent S. Gates, PhD—Thesis Supervisor

DECEMBER 2019

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

Effects of Local Sequence on the Formation of Interstrand DNA Cross-Links Derived from the Reaction of an Abasic Site with a Cytosine Residue on the Opposing Strand

Presented by Luke Edward Pierce,

A candidate for the degree of Master of Science

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

Professor Kent Gates: _____

Professor Michael Harmata: _____

Professor John Brockman: _____

Professor Timothy Phillips: _____

ACKNOWLEDGEMENTS

This thesis and all results presented within would not have been possible without the constant support of my advisor, Dr. Kent Gates. Since Day 1, he has worked patiently with me in all aspects of my graduate school life, graciously accommodating me personally and all activities outside of my research. Without his guidance, knowledge, and understanding, completion of this research and my degree would not have been possible.

Furthermore, past and present members of the Gates group have contributed greatly to these results. Specifically, Nathan Price, Kevin Johnson, and Jacqueline Gamboa-Varela have had a tremendous impact on the results presented in this work. They laid the foundation for what we understand DNA interstrand cross-links to be and cytosine's role in their formation. Maryam Imani helped me get my start in the lab and taught me the safe and proper techniques and groundwork for all experiments described in this thesis. Kurt Housh, Xu Guo, and Tuhin Haldar all acted as mentors and resources throughout my research, as I was able to bounce ideas off of them and always received detailed feedback that helped better myself as a scientist.

Lastly, I would like to thank my wife and daughter. Without my wife's support throughout all of graduate school, I would not have been able to complete this research. Especially with the addition of our daughter a year into my research, she has been wonderful in taking on extra responsibilities at home when I spent many late nights in the lab. Without either of them, I would not have been able to complete any of the work that is reported in this thesis.

TABLE OF CONTENTS

Chapter 2. Formation of Interstrand DNA Cross-Link between Abasic Site (Ap) and Cytosine Residue

Chapter 3. Implications and Impacts of Interstrand DNA Cross-Links between Abasic Site (Ap) and Cytosine Residue

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

LIST OF SCHEMES

CHAPTER 1

CHAPTER 2

LIST OF TABLES

CHAPTER 2

ABSTRACT

DNA Interstrand Cross-Links are of special concern due to the difficulty of repair and cytotoxicity of cells associated with these lesions. Interstrand Cross-Links between abasic sites and native nucleobases has shown promise in recent years. A new type of Interstrand Cross-Link between an abasic site and a cytosine residue has shown capability of forming when placed under specific conditions. A mismatched cytosine residue demonstrated cross-linking capabilities up to 69% in acidic conditions. Mismatches on flanking base pairs also showed increased cross-linking capabilities. Cross-linking is also observed under human physiological conditions, demonstrating this lesion's capability of forming in vivo. Further work in this line of DNA chemistry may result in increased biological relevance and advances in understanding this type of DNA damage.

Chapter 1: Interstrand DNA Cross-Links Derived from Abasic Sites

1.1 Introduction to DNA

Deoxyribose nucleic acid (DNA) contains the genetic information in every cell of the human body. This genetic information is stored in the sequence of its four base pairs: adenine, thymine, cytosine, and guanine. The variations of these sequences code for all proteins within a cell, acting as a genetic blueprint driving all cellular operations. When the sequence of DNA is properly read, it is transcribed into messenger Ribose Nucleic Acid (mRNA)[1], which is in turn translated into the proteins[2] that carry out all cellular functions. DNA sequences must also be accurately replicated when the cell divides.

In order for the DNA sequence to be copied, its double helix must first be unwound by topoisomerase enzymes, separating its double-stranded form into single-stranded DNA. The newly separated strands can then be replicated with the correctly matched nucleotide bases, resulting in two fully functional units of double-stranded DNA. Through this replication, cell proliferation is able to occur, and the genetic code is maintained.

Issues arise when lesions in DNA cause replication to be stalled. Such blocking lesions can lead to cell death, senescence, or genetic instability. This can lead to tissue degradation, dysfunction, and aging or cancer. Lysis as a result of DNA damage can lead to regional tissue failure, differing from the healthy programmed cell death through apoptosis.[3]

DNA damage occurs through various unavoidable mechanisms.[4] Exogenous damage results from Ultra Violet (UV)-induced radiation, ionizing radiation, and daily exposure to chemicals. Endogenous damage occurs within the body through DNA methylation, abasic site

formation, and base pair mutations.[5] These lesions are responsible for the neurodegeneration, aging, and cancers that plague all living organisms.[6]

Because this damage is prevalent in all cellular life, DNA repair pathways have evolved to minimize the overall effect of damage in the cell. The type of repair is dependent upon the nature of the lesion, but mechanisms such as direct reversal, mismatch repair, excision repair, and bypass polymerases all contribute to maintaining cell health in the face of threats caused by DNA damage. These repair pathways are supplemented by so-called "checkpoints" in the cell cycle to ensure that damage is detected quickly and subsequently repaired prior to replication of defective DNA.

Due to the formidable effects of DNA damage, how lesions form as well as DNA repair mechanisms have both been large fields of research for decades. Achieving a better understanding of the mechanisms by which DNA damage occurs, its consequences, and the repair of DNA lesions are at the core of all pharmaceutical development today.[7] Advances in medicine and therapeutic agents have led to growing the field of healthcare in both research and business. Improvement of the human health span may be gained by a thorough understanding of endogenous DNA damage chemistry and how it affects us as a species.

1.2 Interstrand Cross-Links are Highly Cytotoxic DNA Lesions

One type of DNA lesion caused by both exogenous and endogenous damage is the formation of interstrand cross-links (ICL)[8]. An ICL is a covalent link between opposing strands of double-stranded DNA. ICLs are an extremely difficult lesion to repair due to the strength of the covalent bond formed, the location on the double helix, and the fact that both strands are affected simultaneously. Perhaps the key attribute of an ICL is that they cause a stalled

replication fork during cell division. Stalled replication forks, if not reactivated quickly, will cause complete DNA degradation and when left unchecked ultimately lead to cell death.[9]

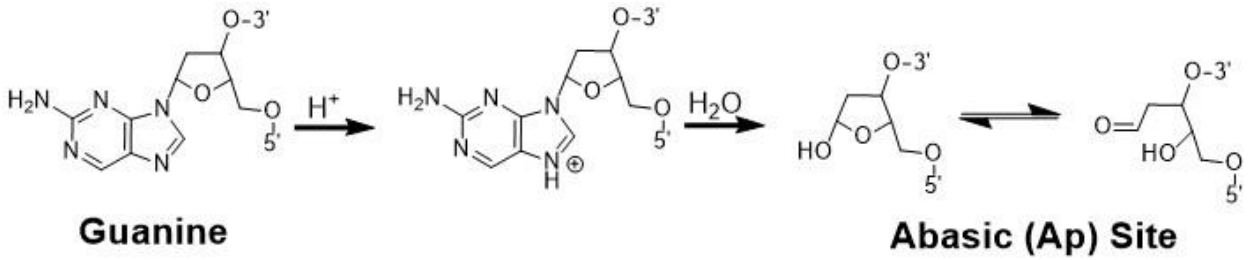
ICLs in DNA have been linked to sporadic cancers, rapid aging (progeria), and neurodegeneration. Inducing ICLs has long been an area of interest in the development of cancer pharmaceuticals, as the ability of ICLs to cause cell lysis has shown promise in fighting over-proliferative tumor cells.[10] The prevention of DNA separation at the replication fork blocks the cell's ability to carry out normal functions such as division, giving targeted introduction of ICL-forming agents potential in the future of cancer therapies.[5]

While this field of exogenously induced cross-links has expanded over the years, the mechanisms by which endogenous cross-links form remains uncertain. Formation of ICLs in the human body has caused the evolution of complex, resource-intensive repair pathways. The identity of the unavoidable ICL lesions in the living cells remains poorly understood and is a target for study to advance the field of medicine.

1.3 Abasic-Site Derived Interstrand Cross-Links in Duplex DNA

Over the past several years, the Gates group at the University of Missouri has studied the formation of potential endogenous ICLs in duplex DNA. The units of DNA participating in this specific type of ICL are abasic sites formed in place of native base pairs. Abasic sites (apurinic or apyridinic [Ap]) result in the loss of a base pair, leaving behind a sugar residue that exists in equilibrium from a ring-closed alcohol to a ring-opened aldehyde (see Scheme 1).

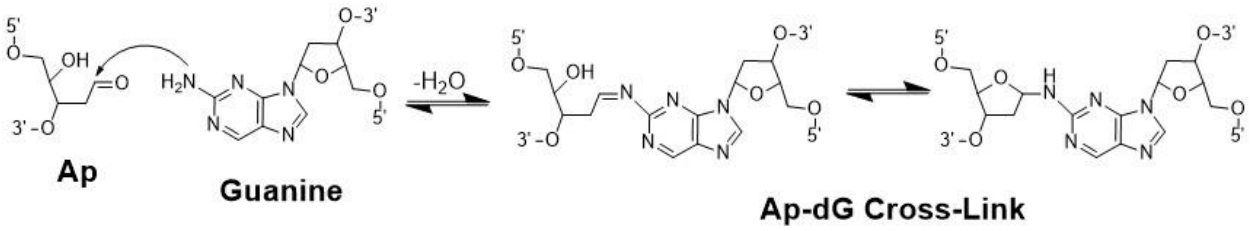
Scheme 1



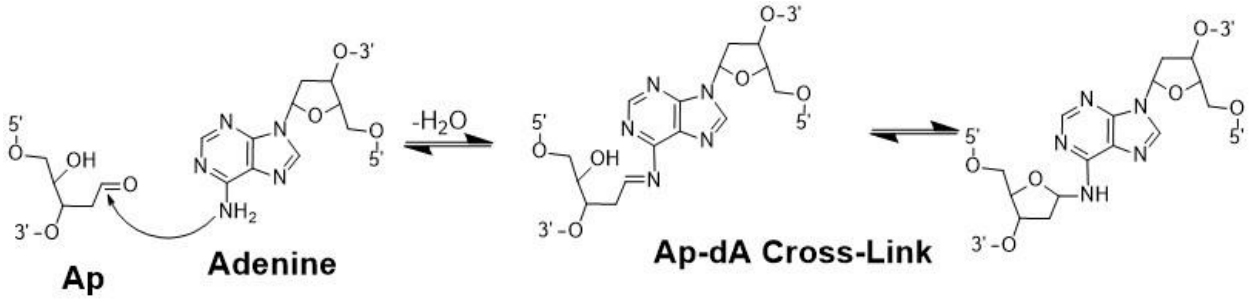
The ring-opened aldehyde is a reactive species able to react with other nucleobases on the opposing strand. Nucleophilic exocyclic amines located on the bases of the opposing strand, when in the proper position, now have the ability to attack the electrophilic aldehyde of the Ap site, creating or generating an ICL.

The Gates group has identified a series of cross-links derived from the reaction of an Ap aldehyde with exocyclic amines on guanine, adenine, and 2-aminopurine.[11-13] These reactions are able to occur under physiological conditions and proceed via imine formation to generate a stable aminoglycoside ICL, as shown in Schemes 2 and 3. These reactions are good candidates for the endogenous cross-link lesions in cellular DNA[14] and are discussed in depth in the following sections.

Scheme 2



Scheme 3



1.4 Formation of Interstrand Cross-Links Between an Abasic Site and an Opposing dG Residue

In the study of ICLs between an abasic site on one strand of duplex DNA and a guanine residue on the opposing strand, it was found that the guanine residue must be on the opposing strand, but one nucleotide offset on the 5' side of the abasic site.[11] For example, the duplex sequence A shown in Figure 1 was found to produce ICL yields in the range of 3 ± 0.3%.

Figure 1

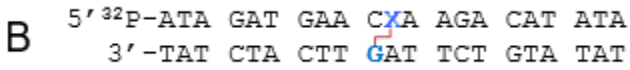
A
5′ 32P-GAT GAT CXA AGA CAT
3′-CTA CTA GAT TCT GTA

In these early experiments, a 15 base pair duplex was used, with all bases correctly matched on the opposing strand and an adenine opposing the abasic site. Reactions were carried out in NaOAc buffer with a pH of 5.5 in the presence of $NaCNBH_3$ to facilitate reductive amination, effectively trapping the ICL**s**, and after a 24-hour incubation at 37˚C. A 3% yield of ICL in this "native" duplex would create a substantial lesion in vivo, as it has been shown that a single interstrand cross-link is potent enough to trigger cell death in eukaryotic cells.[15] ICL yields were

reported at an average of ~1% when performed at pH 7 to further mimic human conditions—still substantially high to show relevance in an in vivo model.

Later studies by Johnson in the Gates group used duplex sequence B shown in Figure 2. By lengthening the sequence to a 21-oligonucleotide duplex and varying the nearest neighbor

Figure 2



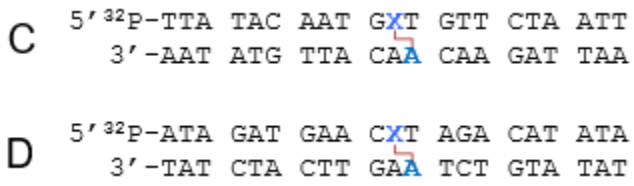
base pairs, ICL yields of ~20% were obtained. It is believed that the lengthening of the duplex may have prevented partial melting of the strands, which would have decreased the yield of the ICL. Furthermore, a sodium acetate buffer was used in the reductive amination step to trap the final cross-link, which may have resulted in a slightly catalyzed reaction that would have caused the reported higher yields.

1.5 Formation of Interstrand Cross-Links Between an Abasic Site and an Opposing dA Residue

Following successful formation and identification of the ICL generated from reaction of the ring-opened aldehyde of an abasic site with the exocyclic amino group found on the guanine residue, it was postulated that an adenine residue would be able to participate in the same chemistry with its exocyclic amino group, as shown in Scheme 3. Due to the nature of the DNA double helix, models showed that the exocyclic amino group is not within the proximity of the Ap site when offset one nucleotide to the 5’ side. Instead, the exocyclic amine was in the major groove and most sterically available when offset by one base pair to the 3’ side of the abasic site. Interestingly, during the Ap-dG studies, no cross-link was shown to form between the abasic site and the adenine directly opposing the lesion, thus the offset site was seen as the best candidate for testing.

This research was performed again using a 21-oligonucleotide sequence, as this length was shown to give a higher yield in the previous Ap-dG studies. Several sequence variations were attempted with the first, duplex C, resulting in a 15.1 ± 0.5% yield, and the highest yield of 71.5 ± 3.7% for duplex D, being displayed in Figure 3.

Figure 3



These reactions were performed under physiological conditions with HEPES buffer at a pH of 7 and incubation of 24 hours at 37˚C. Interestingly, when the duplex is carried out under conditions for reductive amination (pH 5 with $NaCNBH_3$), no increase in yield is observed. In fact, a decrease in yield is observed under these conditions for reasons that remain unclear.

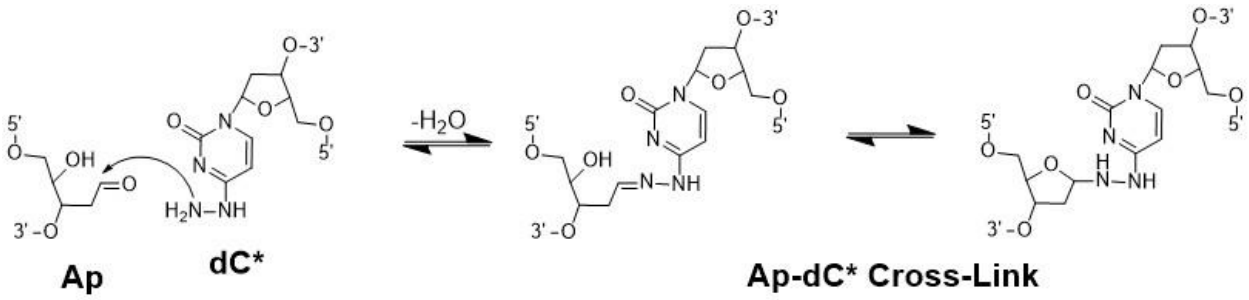
Duplex D was of particular interest, as it used the same sequence as seen in the Ap-dG crosslink of duplex B, with the exception of the presence of the adenine residue in place of the thymine on the 3' side opposing the Ap site. It was theorized that the high yield resulted from a combination of both the Ap-dG and Ap-dA cross-links. However, after identification of the cross-links, it was noted that in duplex D, only the Ap-dA cross-link was observed. The very high yields associated with the Ap-dA cross-link and the physiological relevance of the conditions in which they were formed suggest biological relevance of these ICLs.

1.6 Formation of Interstrand Cross-Links Between an Abasic Site and an Opposing Non-Canonical Nucleobase

Other ICL formations have been studied in the Gates group with non-canonical nucleobases and an abasic site. One particular non-canonical base use was a hydrazine

derivative of cytosine, designated as dC*. It was observed that an Ap-dC crosslink did not form in typical duplexes, presumably due to the spatial distance between the exocyclic amino group on a cytosine and an opposing abasic site aldehyde. The exocyclic amine of cytosine is located approximately 0.7 angstroms farther from the C1 of the Ap sugar when compared to an adenine residue in the same position.[13] Both amino groups are present in the major groove when offset one nucleotide to the 3' side of the Ap site; however, only the Ap-dA cross-link has been observed. Modification of the cytosine amino group to a hydrazine would then put the nucleophilic nitrogen in close enough proximity to take place in the cross-link chemistry as shown in Scheme 4. Additionally, equilibrium constants along the order of 10^4-10^6 are observed for the formation of hydrazones[16] while equilibrium constants of 10^1-10^3 are observed for the formation of an imine[17], as would occur in an Ap-dC ICL.

Scheme 4



The cross-link between Ap and dC* was able to be achieved with a short incubation of 1 hour at 50˚C under reducing conditions at pH of 5 with $NaCNBH_3$.[18] The cross-link was able to be successfully reversed and reformed via heat-cool cycles at 50˚C and 37˚C. This reaction was the first cross-link able to be shown involving a modified cytosine residue.

Another non-canonical base found capable of forming a DNA interstrand cross-link is the 2-aminopurine base. This base is a bicyclic residue, with an exocyclic amine present in the same

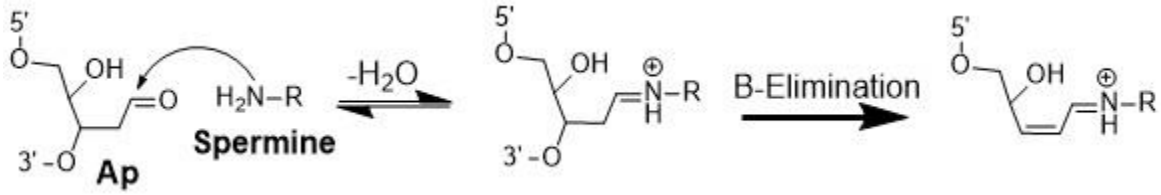
structural location as a guanine residue. Due to these similarities, the base was offset by one nucleotide on the 5' side of the abasic site for the formation of the ICL.

Through preliminary work done by Maryam Imani and Tuhin Haldar of the Gates group, this cross-link was shown to have formed when reacted in a NaOAc buffer with a pH of 5.2 and incubated for 4 hours at 37˚C.[19] This cross-linking reaction generated a 44% yield in the absence of reducing agents, but when treated with $NaCNBH_3$, the yield increased dramatically to 85%. Hydroxyl radical footprinting utilizing iron-EDTA strand cleavage confirmed that the cross-link does occur between the abasic site and the non-canonical 2-aminopurine base.

1.7 Formation of Spermine-Mediated Nicked Interstrand Cross-Links

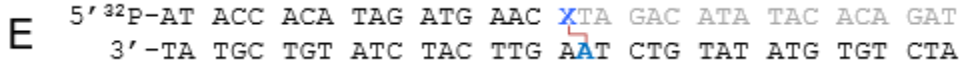
Abasic sites are also subject to DNA strand cleavage to give a 3' alkenal group. Yang of the Gates group has demonstrated the capability of the Ap sites to form ICL from this alkenal group. Spermine was introduced to an abasic legion in duplex DNA, and created a double strand break through the mechanism described in Scheme 5.

Scheme 5



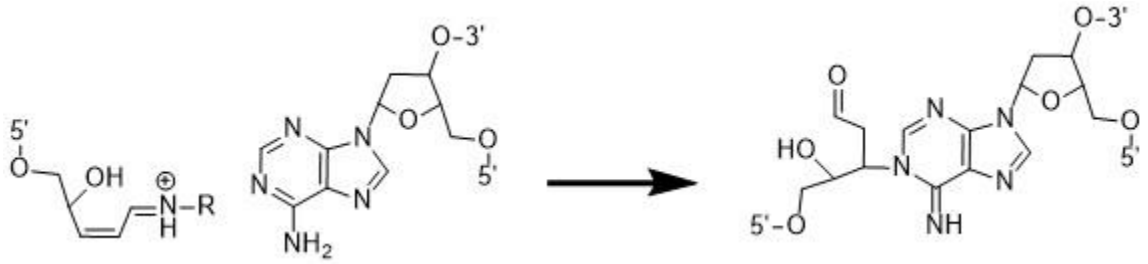
This nicked abasic site reacted with a nucleophilic amine on an adenine residue on Duplex E (shown in Figure 4) to for the ICL.[20]

Figure 4



ICL formation occurred after incubation at 37˚C in HEPES buffer with pH 7, which was then trapped through reductive amination with $NaCNBH_3$. The cross-link between the nicked abasic site and the adenine residue is theorized to have occurred though the mechanism depicted in Scheme 6.

Scheme 6



The overall yield for the cross-link formed was ~35%, which is substantial when considering the cytotoxicity of interstrand cross-links. Furthermore, formation of this cross-link was shown to inhibit DNA replication by blocking the strand separation polymerase.

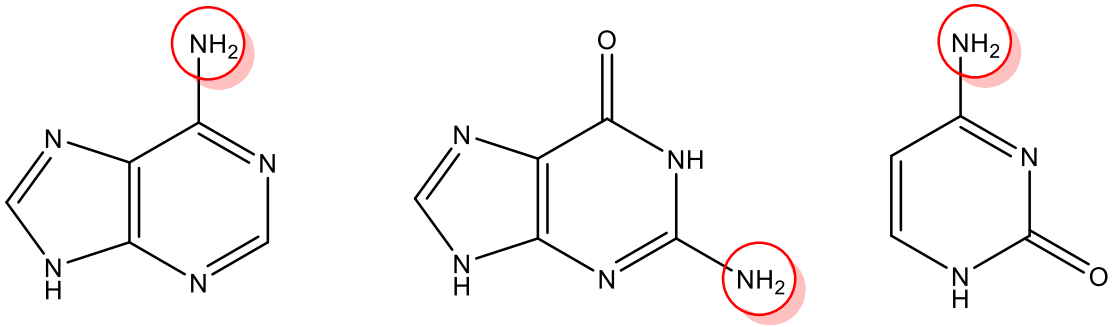
1.8 Cytosine Contains All the Necessary Tools to Form a DNA Interstrand Cross-Link

With the exception of the spermine-mediated "nicked" ICL, all DNA interstrand cross-links discussed up to this point have involved an exocyclic nucleophile attacking the abasic lesion of an opposing strand of double-stranded DNA. Most notably, the canonical bases are able to perform this chemistry in physiologically relevant conditions with their exocyclic amino groups. Cytosine is another naturally occurring nucleobase containing an exocyclic amino group (See Figure 5).

Importantly, this exocyclic amine of cytosine has the capability of reacting with aldehyde groups. In a pair of papers published by McGhee and von Hippel, it was shown that cytosine,

guanine, and adenine formed monoadducts when reacted with formaldehyde.[21-22] The reactions were all shown to have forward equilibrium constants above 1 M^{-1}, with adenine and cytosine

Figure 5



both reporting the highest values at around 12 M^{-1}. Concurrently, all reactions were exothermic with adenine and cytosine reporting the most negative enthalpies of -4 and -6 kcal/mol, respectively. Based on these results, we would expect cytosine to have the chemical potential to form ICLs similar to adenine.

Nonetheless, a cross-link formation between an abasic site and cytosine had not yet been reported. One important thing to note when looking at the differences in cytosine with other base pairs is that both adenine and guanine are purine nucleotides, while cytosine is a pyrimidine. It was theorized by Price in the Gates group that distance from the abasic site may be what is hindering the cross-linking chemistry from occurring.[13] This theory would be supported by the bicyclic nature of adenine and guanine offering more distance from the DNA backbone and making the amino group closer in proximity and able interact with the abasic site on an opposing strand with less distortion of the duplex required.

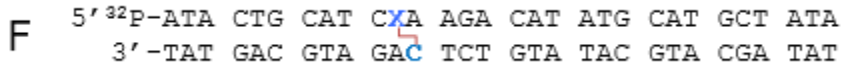
1.9 Mismatched Bases Offer New Chemical Availabilities for DNA Interstrand Cross-Linking Chemistry

All Ap-derived ICLs discussed up to this point were performed under physiologically relevant conditions in order to demonstrate the lesions' ability to form and survive in cells. Reaction conditions to achieve this goal have been targeted at a pH of 7, temperature of 37˚C, and a duplex DNA sequence with the only damage being the installed abasic site. Slight derivations in pH and temperature have also been discussed in regards to creating a higher yield or a more stable final product for the purpose of obtaining data. At this point, no changes in sequence or additions of other type of DNA damage have yet been explored.

In a study that used the Ap-dA ICL to detect a DNA mutation, Imani-Nejad of the Gates group showed that placing an A-A mismatch on the cross-linking adenine residue dramatically increased the ICL yield.[23] The cross-link yield went up from 7.3 ± 2% to 85 ± 3% when the sequence is changed from a correctly matched A-T pair to the mismatched A-A pair. The extra mobility[24] of the adenine residue may enable the amino group to participate in the cross-linking chemistry at a more successful rate.

With this concept of flexibility of the DNA double helix resulting from mismatched base pair lesions, the window of opportunity for the previously-thought-to-be-impossible Ap-dC cross-link was opened. First, Gamboa-Varela of the Gates group was able obtain a slow-moving band on an electrophoresis gel, indicating a cross-link formation, using Duplex F (Figure 6 below) containing a mismatched cytosine residue and abasic site.[25]

Figure 6



This slow moving band was observed after a 72-hour incubation at 37˚C in NaOAc buffer at pH 5.2.

1.10 Mismatched Deoxycytidine and Abasic Sites Form an Ap-dC DNA Interstrand Cross-Link

To continue research pf the Ap-dC cross-link, two duplexes were chosen as candidates for cross-link formation: Duplex F and G (see Figures 6 and 7).

Figure 7

G

5′-^{32}P-AA CAC ACT CTC CAA XAT CAC TCA CTC CCT CAC ACC
3′-TT GTG TGA GAG GTT ACA GTG AGT GAG GGA GTG TGG

Duplex F was the same duplex as discussed above in Section 1.9 of this chapter, as it had already been shown to create a slow-moving band. Duplex G was also used, as it is the same sequence in which the Ap-dC* cross-link was observed, with the exception of the canonical dC nucleotide being used in place of the modified dC*. Results from this experimenting demonstrated that the cross-link does indeed form at the position of the 3' dC residue for Duplex F in both pH of 5.2 and 7, but curiously migrated to the 5' dG residue when in the presence of $NaCNBH_3$[26]. Also interestingly, Duplex G exhibited signs of a cross-link at the dA residue directly opposing the abasic site, making evidence of the Ap-dC cross-link unclear.

1.11 Chapter 1 Summary

DNA damage by exogenous and endogenous means is unavoidable. Repair pathways have evolved to extend the lifetime of DNA. Certain lesions are more difficult to repair and the pathway is oftentimes error-prone. One such lesion is the formation of an interstrand cross-link. Gates' group has researched various methods of ICL formation with an abasic site in native duplex DNA. Ap-dG, Ap-dA, Ap-dC*, Ap-2-aminopurine, and a spermine-mediated nicked Ap-

dA have all been characterized by members of the Gates group to date. The center point of understanding this type of DNA damage is gaining a physiologically relevant model to apply to cancer therapeutics and pharmaceuticals.

To this point, crosslinks between an abasic site and three of the four canonical bases, Ap-dG, Ap-dA, and Ap-dC, have all been formed in vitro by the Gates group. While the Ap-dG and Ap-dA cross-links have several publications to their name, with full characterization of the bonds and sequences[11-13], the Ap-dC cross-link is not well understood. It is not yet fully known in what situations provide the highest Ap-dC crosslink yield and how local sequence affects the strength of the cross-link. Further insight into this area of DNA ICL chemistry is worth exploring as results may lead to the better understanding of how cross-links form in vivo and what types of specific DNA damage can facilitate cross-link formation.

1.12 Chapter 1 References

1. Wang, J. C., Lynch, A.S., Transcription and DNA supercoiling. *Current Opinion in Genetics and Development*. Volume 3, Issue 5, 1993. 764-768.

2. Kahvejian, A., Roy, G., Sonenberg, N., The mRNA closed-loop model: the function of PABP and PABP-interacting proteins in mRNA translation. *Cold Spring Harb Symp Quant Biol.* 2001. 66: 293-300.

3. Arends, M. J., Morris, R. G., Wylie, A. H., Apoptosis. The role of the endonuclease. *Am. J. Pathology*. 1990. 136(3): 593-608.

4. Gupta, R. C., Lutz, W. K., Background DNA damage for endogenous and unavoidable exogenous carcinogens: a basis for spontaneous cancer incidence? *Mutation Research*. 1999. 424 (1-2): 1-8.

5. Cox, M. M., Goodman, M. F., Kreuzer, K. N., Sherratt, D. J., Sandler, S. J., Marians, K. J., The importance of repairing stalled replication forks. *Nature*. 2000. 404: 37-41.

6. Kennedy, S. R, Loeb, L. A., Herr, A. J., Somatic mutations in aging, cancer and neurodegeneration. *Mechanisms of Ageing and Development*. 2012. 113 (4): 118-126.

7. Middaugh, C. R., Evans, R. K., Montgomery, D. L., Casimiro, D. R., Analysis of plasmid DNA from a pharmaceutical perspective. *Journal of Pharmaceutical Sciences*. 1998. 87 (2): 130-146.

8. Dronkert, M. L., Kanaar, R., Repair of DNA interstrand cross-links. *Mutation Research/DNA Repair*. Volume 486, Issue 4, 2001. 217-247.

9. Balasubramanian, B., Pogozelski, W. K., Tullius, T. D., DNA strand breaking by the hydroxyl radical is governed by the accessible surface areas of the hydrogen atoms of the DNA backbone. *Proc. Natl. Acad. Sci. USA*. 1998. 95: 9738-9743.

10. Feng, J. A., Johnson, R. C., Dickerson, R. E., Hin recombinase bound to DNA: the origin of specificity in major and minor groove interactions. *Science*. Volume 263, Issue 5145. 1994. 348-355.

11. Dutta, S., Chowdhury, G., Gates, K. S., Interstrand cross-links generated by abasic sites in duplex DNA. *J. Am. Chem. Soc*. 2007. 129(7): 1852-1853.

12. Johnson, K. M., Price, N. E., Wang, J., Fekry, M. I., Dutta, S., Seiner, D. R., Wang, Y., Gates, K. S., On the formation and properties of interstrand DNA-DNA cross-links

forged by reaction of an abasic site with the opposing guanine residue of 5'CAp sequences in duplex DNA. *J. Am. Chem. Soc*. 2013. 135(3): 1015-1025.

13. Price, N. E., Johnson, K. M., Wang, J., Fekry, M. I., Wang, Y., Gates, K. S., Interstrand DNA-DNA cross-link formation between adenine residues and abasic sites in duplex DNA. *J. Am. Chem. Soc*. 2014. 136(9): 3483-3490.

14. Sczepanski, J. T., Jacobs, A. C., Greenburg, M. M., Self-promoted DNA interstrand cross-link formation by an abasic site. *J. Am. Chem. Soc*. 2008. 130(30): 9646-9647.

15. Grossmann, K. F., Ward, A. M., Matkovic, M. E., Folias, A. E., Moses, R. E., S-cerevisiae has three pathways for DNA-interstrand crosslink repair. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*. 2002. 487(3-4): 73-83.

16. Kolmel, D. K., Kool, E. T., Oximes and Hydrazones in Bioconjugation: Mechanism and Catalysis. *Chem Rev*. 2017. 117(15): 10357-10376.

17. Crugeiras, J., Rios, A., Riveiros, E., Richard, J. P., Substituent Effects on the Thermodynamic Stability of Imines Formed from Glycine and Aromatic Aldehydes: Implications for the Catalytic Activity of Pyridoxal-5'-Phosphate (PLP). *J Am Chem Soc*. 2009. 131(43): 15815-15824.

18. Varela, J. G., Gates, K. S., A simple, high-yield synthesis of DNA duplexes containing a covalent, thermally-reversible interstrand cross-link at a defined location. *Agnew Chem Int Ed Engl*. 2015. 54(26): 7666-7669.

19. Imani Nejad, M., Price, N. E., Haldar, T., Lewis, C., Wang, Y., Gates, K.S., Interstrand DNA Cross-Links Derived from Reaction of a 2-Aminopurine Residue with an Abasic Site. *Submitted to ACS Chemical Biology*. 2019.

20. Yang, Z., Price, N. E., Johnson, K. M., Wang, Y., Gates, K. S., Interstrand cross-links arising from strand breaks at true abasic sites in duplex DNA. *Nucleic Acids Research*. Volume 45, Issue 11. 2017. 6275-6283.

21. McGhee, J. D., von Hippel, P. H., Formaldehyde as a Probe of DNA Structure. I. Reaction with Exocyclic Amino Groups of DNA Bases. *Biochemistry*. 1975. 14(6): 1281-1296.

22. McGhee, J. D., von Hippel, P. H., Formaldehyde as a Probe of DNA Structure. II. Reaction with Endocyclic Imino Groups of DNA Bases. *Biochemistry*. 1975. 14(6): 1297-1303.

23. Imani, M. N., Shi, R., Zhang, X., Gu, L. Q., Gates, K. S., Sequence-Specific Covalent Capture Coupled with High-Contrast Nanopore Detection of a Disease-Derived Nucleic Acid Sequence. *ChemBioChem*. 2017. 18: 1383-1386.

24. Wallace, R. B., Shaffer, J., Murphy, R. F., Bonner, J., Hirose, T., Itakura, K., Hybridization of Synthetic Oligodeoxyribonucleotides to Phi Chi 174 DNA: The Effect of Single Base Pair Mismatch. *Nucleic Acids Research*. 1979. 6(11): 3543-3557.

25. Johnson, L. M., Hypoxia Selective DNA Alkylating Analogs of Tlrapazamine and Ap-Derived DNA Interstrand Cross-Links. *Ph.D. Dissertation*. University of Missouri, Columbia. 2014.

26. Varela, J. G., Interstrand DNA Cross-Link Formation via Hydrazone Formation between N^4-Aminocytidine and AP-sites. *Ph.D. Dissertation*. University of Missouri, Columbia. 2017.

Chapter 2: Formation of Interstrand DNA Cross-Link between Abasic Site (Ap) and Cytosine Residue

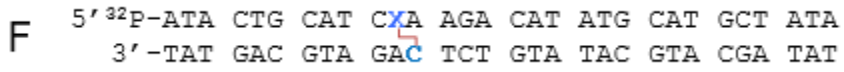
2.1 Introduction to Abasic Site and Cytosine Residue Interstrand Cross-Links

As discussed in the previous chapter, various members of the Gates group have shown the capabilities of interstrand cross-linking between an abasic site (Ap) and naturally occurring base pairs as a result of endogenous DNA damage.[27-31] While these cross-linking studied have focused primarily on the Ap-dG and Ap-dA formation, an Ap-dC cross-link has briefly been discussed to have potential, if put under the correct conditions. It became clear that the Ap-dC cross-link was unable to be formed when the cytosine residue was correctly paired with an opposing guanine. When mismatched with an adenine, though, cross-linking with a neighboring abasic site on the opposing strand was observed. This knowledge was used as the foundation of the research that will be discussed in this chapter.

2.2 Determining Optimal Conditions for the Formation of an Ap-dC Interstrand Cross-Link

With a known cross-link yield of 29% found by Gamboa-Varela in Duplex F, this was the initial target to study to determine if results were able to be repeated, and then further improved.

Figure 8



Critical points of information regarding this cross-link were still unknown, such as conditions that favor cross-link formation and which sequence variations allow for best yield of the cross-

link. The first step in understanding this cross-link was determining the reducing conditions and pH levels that best facilitated the cross-link formation.

Work began by labeling a deoxyuracil-containing parent strand from Duplex F (the Ap-containing strand) with ^{32}P by standard procedures using t4 polynucleotide kinase and ATP. A phenol-chloroform work-up was then performed to remove the kinase, and the labeled oligonucleotide was purified by gel electrophoresis. The DNA was located by UV-shadowing and the band excised from the gel. The gel slice was then placed in elution buffer to extract the DNA and the solution was desalted by passage through Sep Pak columns.

After labeling, the parent strand was annealed to its compliment from Duplex F in a solution of MOPS buffer and NaCl. Next, the abasic (Ap) site was installed by hydrolysis of the dU residue by the DNA repair enzyme uracil DNA glycosylase (UDG) in UDG buffer. A phenol chloroform work-up, followed by an ethanol precipitation, was then performed to isolate the duplex.

The next step of the experiment was to branch into four samples: pH 7.0 in reducing conditions, pH 7.0 without reducing conditions, pH 5.2 in reducing conditions, and pH 5.2 without reducing conditions. In order to achieve a pH of 7.0, HEPES buffer (50 mM) was used to facilitate the reaction, while NaOAc buffer (750 mM) was used to create an environment of pH 5.2. To control the use of reducing agent, $NaCNBH_3$ (250 mM) was added or withheld from the solution.

The four samples were prepared per the following conditions described below in Table 1:

Table 1: Initial Test Conditions

Reagent	pH 5.2 w/$NaCNBH_3$	pH 5.2 no $NaCNBH_3$	pH 7.0 w/$NaCNBH_3$	pH 7.0 no $NaCNBH_3$
NaOAc	750 mM	750 mM	0 mM	0 mM
HEPES	0 mM	0 mM	50 mM	50 mM
$NaCNBH_3$	250 mM	0 mM	250 mM	0 mM
NaCl	100 mM	100 mM	100 mM	100 mM

After preparing the samples with the concentrations listed above, all four were allowed to incubate at 37˚C (chosen for physiological relevance) for 3 days. Following the 3 day incubation, a piperidine work-up was performed in a controlled sample to show successful installation of the abasic site.

Gel electrophoresis analysis was used to analyze the controls and cross-link reactions. The reactions showed successful installation of the abasic site through the total cleavage of the strands in a piperidine work-up (Lane 5, Figure 9). Slow-moving bands for all four reactions were observed in Lanes 1-4 of Figure 5 as well, with yields of each reaction condition described in Table 2 and Figure 6.

All four test conditions did show cross-linking capability, with acidic conditions resulting from use of NaOAc buffer (at pH 5.2) and in the absence of reducing agent ($NaCNBH_3$) resulting in the highest yield (64 ± 0.3%). While these results show that the Ap-dC cross-linking reaction occurs at the greatest yield in acidic conditions (pH 5.2) rather than physiologically relevant conditions (pH 7.0), it is worth noting that the reaction does best proceed in the absence of reducing agent, which is conducive to the physiological environment.

Figure 9

Figure 9. Formation of Ap-dC cross-links in various reaction conditions. Gel electrophoresis analysis utilized a 0.4 mm thick 20% denaturing polyacrylamide gel. Quantitative measurement of the bands obtained using phosphorimager analysis. Lane 1 corresponds to pH 7.0 w/$NaCNBH_3$, Lane 2 corresponds to pH 7.0 no $NaCNBH_3$, Lane 3 corresponds to pH 5.2 w/$NaCNBH_3$, and Lane 4 corresponds to pH 5.2 no $NaCNBH_3$. Lane 5 is the Ap-containing duplex **F** treated with 1 M piperidine at 95˚C for 30 minutes. Lane 6 is the size marker consisting of the ^{32}P-labeled Ap-containing duplex **F**.

Table 2: Initial Condition Ap-dC Cross-Link Yields

Test Conditions	Average Cross-Link Yield
pH 7.0 w/$NaCNBH_3$	8.8 ± 1%
pH 7.0 no $NaCNBH_3$	35 ± 1%
pH 5.2 w/$NaCNBH_3$	15 ± 1%
pH 5.2 no $NaCNBH_3$	64 ± 0.3%

Table 2: ± values reflect standard deviation of four replicates used in this experiment.

Figure 10

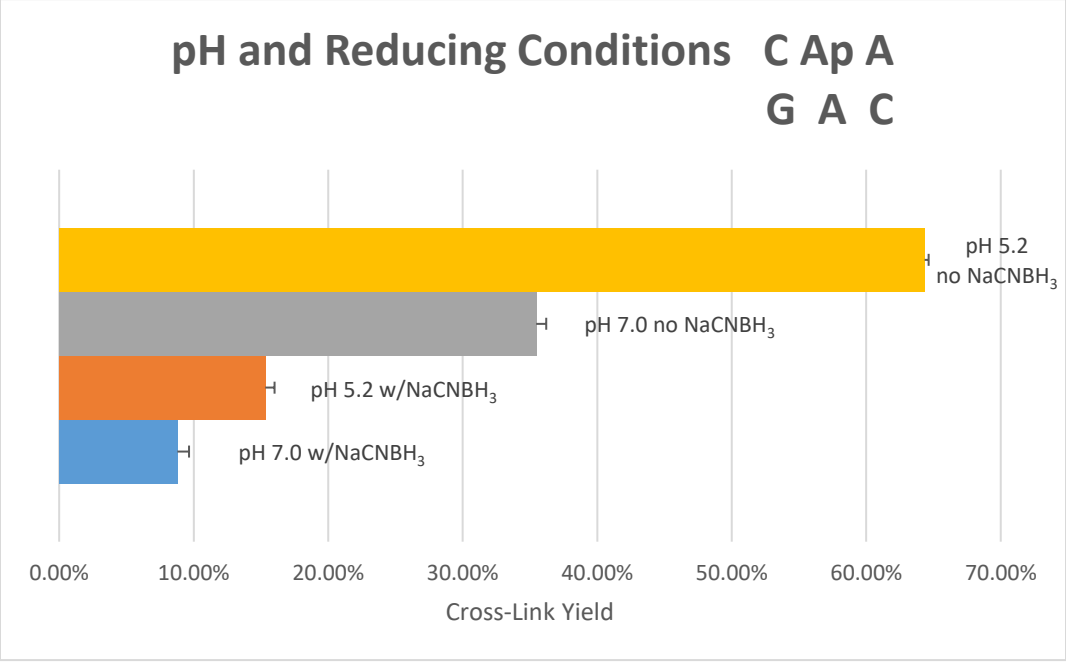


Figure 10. Bar graph of Ap-dC cross-link yields resulting from Duplex **F** subjected to various reaction conditions. Yields are qualitatively obtained from analysis of gel electrophoresis.

Furthermore, the yield of the reaction at the optimum physiologically relevant conditions (pH 7.0 in the absence of reducing agent) was shown to be 35 ± 1%, which is significantly large considering the effect a single interstrand cross-link can have on cytotoxicity of a cell.[30]

2.3 Effects of the Nucleotide Directly Opposing the dC Residue Participating in the Ap-dC Cross-Link

After establishing that a pH of 5.2 and the absence of reducing conditions results in the highest possible yield of an Ap-dC interstrand cross-link, it was of interest to determine in which sequences these reactions could occur. Previous results had clearly shown that the directly opposing residue of the dC participating in the crosslink with the abasic site had a profound effect on its ability to react, as no cross-link was shown to form when it was correctly paired with an opposing guanine.[27] When the guanine is replaced with an adenine, cross-linking is possible. With this knowledge, it was decided that the base directly opposing the dC residue would be varied to determine the effect on its ability to perform the cross-linking chemistry.

This was executed through an Opposing Base Experiment, where the base opposite the dC residue was altered. The relevant manipulations to Duplex F in order to achieve this goal are denoted with the color green in Figure 11 below. A total of four duplexes were tested, Duplexes F, H, I, and J.

Figure 11

F
5′ ^{32}P-ATA CTG CAT CXA AGA CAT ATG CAT GCT ATA
3′-TAT GAC GTA GAC TCT GTA TAC GTA CGA TAT

H
5′ ^{32}P-ATA CTG CAT CXT AGA CAT ATG CAT GCT ATA
3′-TAT GAC GTA GAC TCT GTA TAC GTA CGA TAT

I
5′ ^{32}P-ATA CTG CAT CXC AGA CAT ATG CAT GCT ATA
3′-TAT GAC GTA GAC TCT GTA TAC GTA CGA TAT

J
5′ ^{32}P-ATA CTG CAT CXG AGA CAT ATG CAT GCT ATA
3′-TAT GAC GTA GAC TCT GTA TAC GTA CGA TAT

For all four duplexes, the same reactions were ran as described in 2.2 of this chapter, including ^{32}P labeling, Ap site installation, 3 day incubation at 37˚C, piperidine work-up, and gel

electrophoresis (See Figure 12). Piperidine work-up showed full cleavage, indicating that the abasic site had been successfully installed (Figure 12, Lanes 3, 6, 9, and 12). Slow moving bands were clearly visible in all instances in which the dC residue was mismatched (Figure 12, Lanes 5, 8, and 11), but no clear slow-moving band was shown when it was correctly matched with an opposing guanine residue (Figure 12, Lane 2). An exceedingly faint band was present in Lane 2 as a yield of 2.7% cross-link was able to be quantified. Fully detailed yields are described in Table 3 and displayed in Figure 8.

Table 3: Varying Directly Opposing Base Ap-dC Cross-Link Yields

Sequence = C Ap X G A C	
Nucleotide Opposing dC Residue	**Average Cross-Link Yield**
Duplex F (X = A)	69 ± 3%
Duplex H (X = Y)	40 ± 2%
Duplex I (X = C)	53 ± 2%
Duplex J (X = G)	2.7 ± 1%

Table 3: ± values reflect standard deviation of six replicates used across two experiments.

All four duplexes did show an amount of cross-linking capability, with an opposing adenine residue resulting in the highest yield of Ap-dC cross-link (69 ± 3%). It became clear that any mismatched base pair will result in substantial cross-linking efficiency, as yields for a thymine and cytosine residue were at 40 ± 2% and 53 ± 2% yields, respectively. While a band was not clearly visible to the naked eye in Figure 12, the phosphorimager software was able to detect a small amount of cross-link in the correctly paired guanine sample (2.7 ±

Figure 12

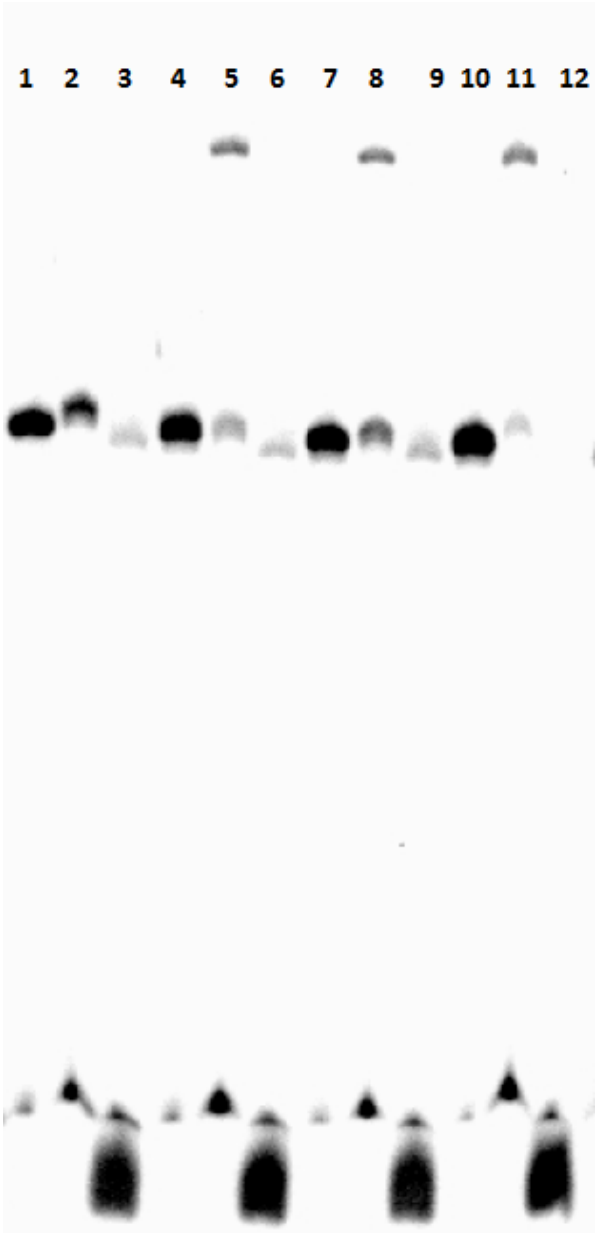


Figure 12. Formation of Ap-dC cross-links when varying the nucleotide directly opposing the dC residue. Gel electrophoresis analysis utilized a 0.4 mm thick 20% denaturing polyacrylamide gel. Quantitative measurement of the bands obtained using phosphorimager analysis. Lane 3 corresponds to Duplex **J**, Lane 4 corresponds to Duplex **I**, Lane 7 corresponds to Duplex **H**, and Lane 11 corresponds to Duplex **F**. Lanes 3, 6, 9, and 12 are the Ap-containing duplexes **J**, **I**, **H**, and **F** treated with 1 M piperidine at 95˚C for 30 minutes. Lanes 1, 4, 7, and 10 are the size marker consisting of the ^{32}P-labeled Ap-containing duplexes **J**, **I**, **H**, and **F**, respectively.

Figure 13

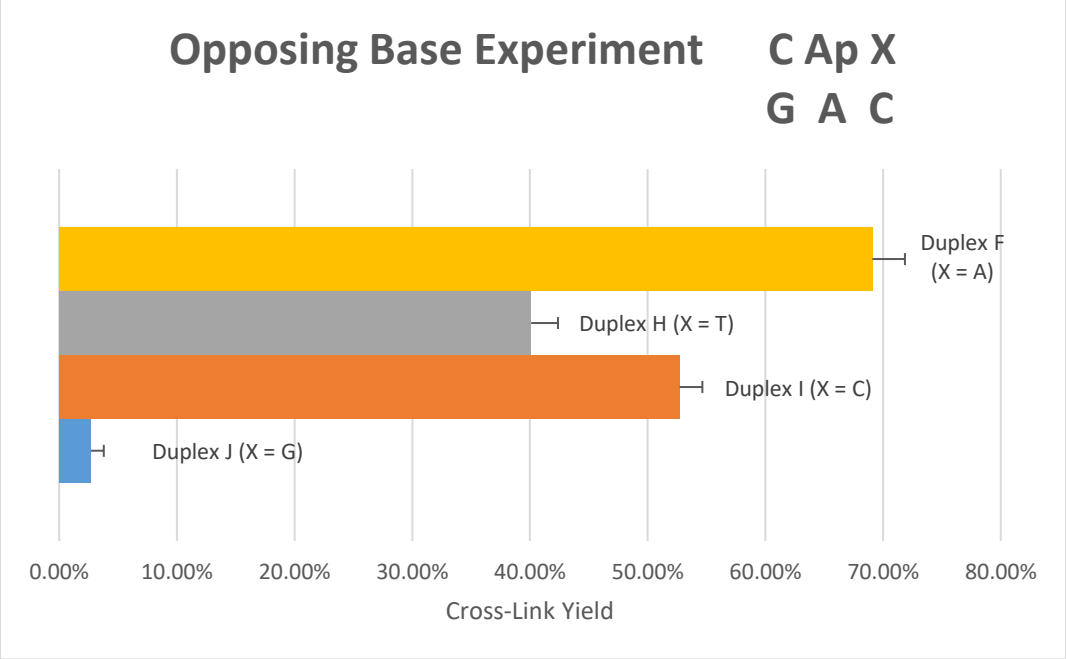


Figure 13. Bar graph of Ap-dC cross-link yields resulting from Duplexes **F**, **H**, **I**, and **J** subjected to consistent reaction conditions. Yields are qualitatively obtained from analysis of gel electrophoresis

1%). This is significant as it is the first time that cross-linking capabilities have been shown in the "natural" strand of the duplex DNA, with the only damage being the abasic site. It should be noted that the cross-link yields are similar for Duplex F in this set of experiments, shown in Table 3, as well as those displayed in Table 2. This is expected as Duplex F participated in both studies, under the same conditions.

The trend of the mismatched dC residue being a pre-requisite for cross-link chemistry was the most important information obtained from this set of experiments, as all three mismatched duplexes resulted in yields above 40%, while the correctly matched pair resulted in almost no cross-link, at 2.7% yield. Duplex F (adenine opposing the dC residue) was still shown to be the most robust of all sequences thus far as the Duplexes H and I both showed 16 and 29%

lower yields, respectively. The fact that the adenine is a purine, containing two rings in its structure, and the cytosine and thymine are pyrimidines, containing only a single ring, may affect the spatial available of the dC residue participating in the cross-link, though no hard evidence exists supporting this claim other than the trends discovered here.

2.4 Effects of the Nucleotide Directly Opposing the Abasic Site in the Ap-dC Cross-Link

With varying the base opposing dC residue resulting in vastly different cross-linking yields, it became obvious that sequence variance has a profound effect on the Ap-dC cross-link. It was shown that an adenine residue directly opposing the deoxycytidine resulting in an extremely high yield (69 ± 3%), but the effects of the nucleotide directly opposing the abasic site was not yet known. For the next set of experiments, the base opposite of the abasic site was altered from its original adenine, to a guanine, cytosine, and thymine each (marked in green in Figure 14). An adenine residue was chosen to directly opposing the dC residue participating in the cross-link as that was shown to give the highest yield. Reaction conditions were kept the same (pH 5.2, no $NaCNBH_3$, and 72-hour incubation at 37˚C) and Duplexes F, K, L, and M (see Figure 14) were utilized.

This experiment resulted in slow-moving bands on the gel electrophoresis (Figure 15) for all four nucleotides used, with each base varying in its cross-linking capability. The highest yield is seen when an adenine residue directly opposes the abasic site, as shown in Duplex F, at 60 ± 1%. This was consistent with the high yields seen previously in Duplex F. The remaining nucleotides resulted in lower yields, with cytosine (Duplex L) being the next highest at 32 ± 5%, then guanine (Duplex K) at 26 ± 1%, and the lowest yield resulting from an opposing thymine (Duplex M) at 15 ± 1%. All yields are reported in Table 4 and displayed in Figure 16. It is

reasonable that adenine resulted in the highest yield, as it is the "natural" pair to the uracil residue from which the abasic site is derived through utilization of the UDG enzyme.

Figure 14

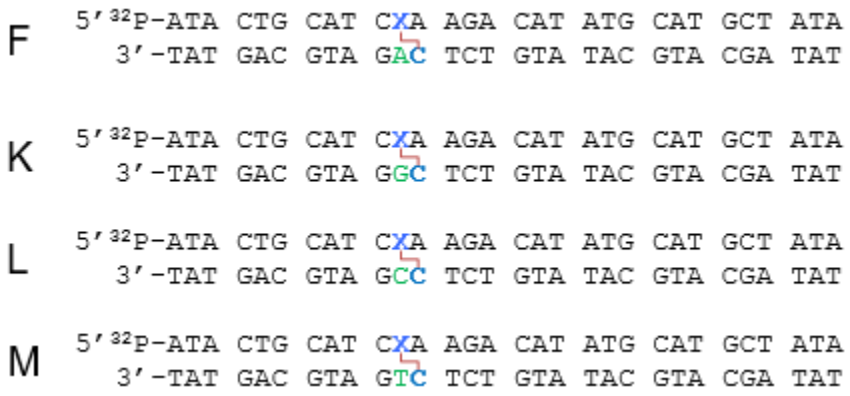


Table 4: Varying Base Opposite Abasic Site in Ap-dC Cross-Link Yields

Sequence = C Ap A G X C	
Sample	**Average Cross-Link Yield**
Duplex F (X =A)	60 ± 1%
Duplex K (X = G)	26 ± 1%
Duplex L (X = C)	32 ± 5%
Duplex M (X = T)	15 ± 1%

Table 4: ± values reflect standard deviation of four replicates used in this experiment.

Figure 15

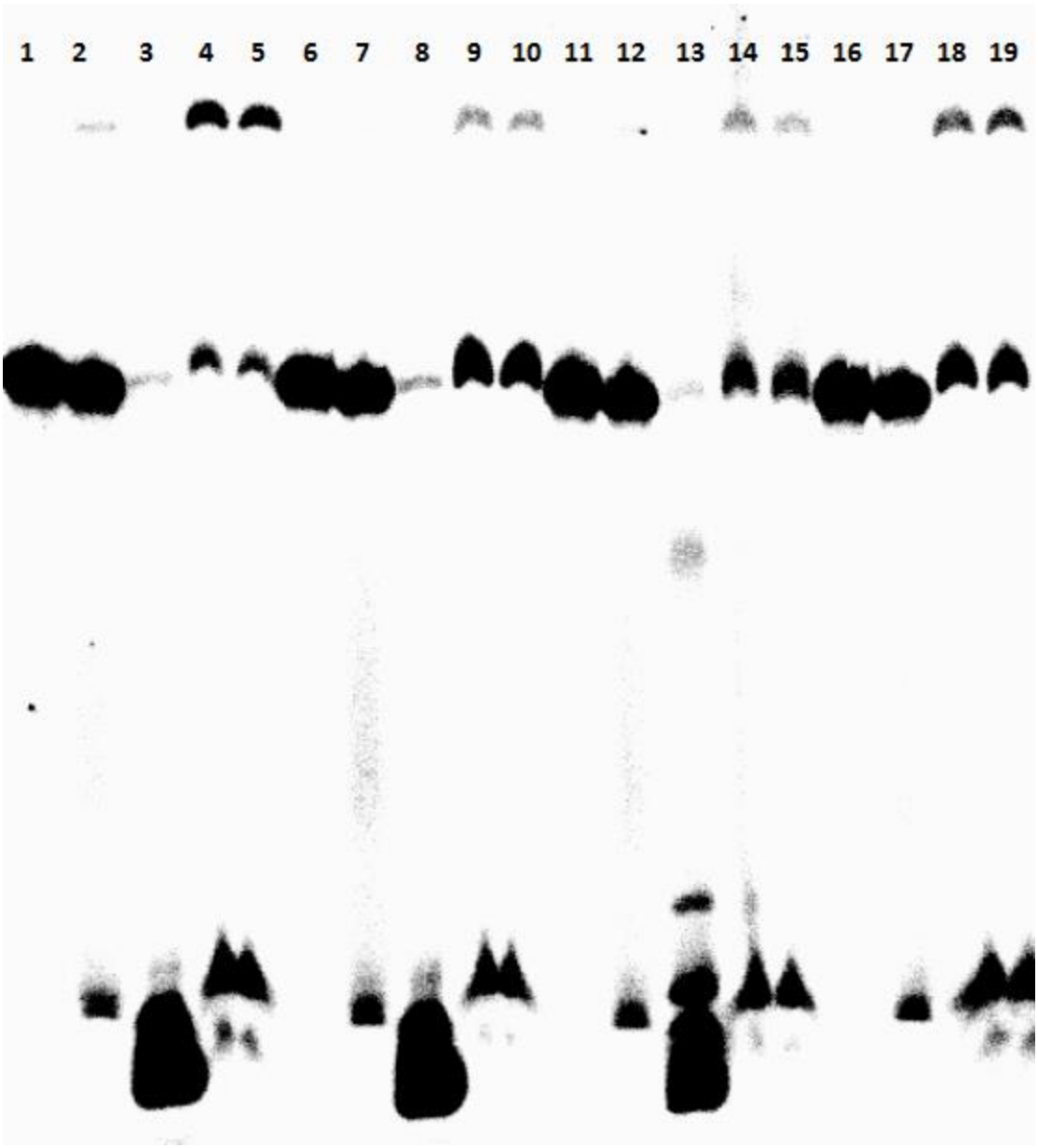


Figure 15. Formation of Ap-dC cross-links when varying the nucleotide directly opposing the abasic site. Gel electrophoresis analysis utilized a 0.4 mm thick 20% denaturing polyacrylamide gel. Quantitative measurement of the bands obtained using phosphorimager analysis. Lanes 4 and 5 correspond to Duplex **F**, Lanes 9 and 10 correspond to Duplex **M**, Lanes 14 and 15 correspond to Duplex **K**, and Lanes 18 and 19 correspond to Duplex **L**. Lanes 3, 8, 13 are the Ap-containing duplexes **F**, **M**, and **K**, respectively, treated with 1 M piperidine at 95˚C for 30 minutes. Piperidine work-up for Duplex **L** unavailable. Lanes 2, 7, 12, and 17 are the size marker consisting of the ^{32}P-labeled Ap-containing duplexes **F**, **M**, **K**, and **L**. Lanes 1, 6, 11, and 16 are the size marker consisting of the ^{32}P-labeled dU-containing duplexes **F**, **M**, **K**, and **L**, respectively.

Figure 16

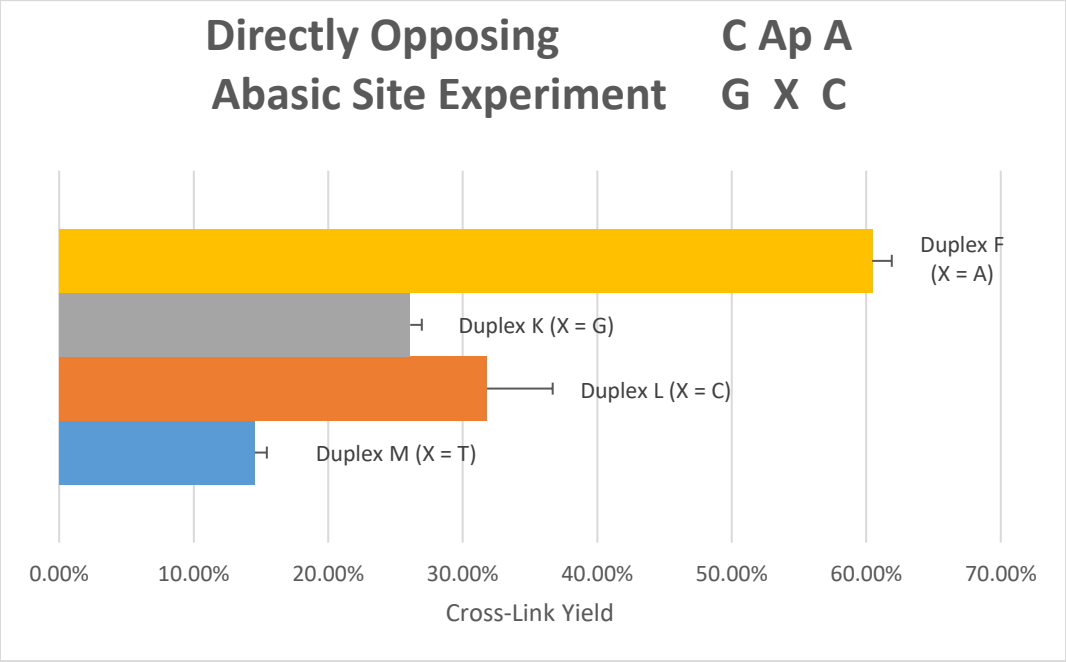


Figure 16. Bar graph of Ap-dC cross-link yields resulting from Duplexes **F**, **K**, **L**, and **M** subjected to consistent reaction conditions. Yields are qualitatively obtained from analysis of gel electrophoresis

2.5 Effects of the 5' Nearest Neighbor Base Pairs of the Ap-dC Cross-Link

To this point, it has been established that adenine residues opposing both the abasic site and the dC residue participating in the Ap-dC cross-link results in the sequence yielding the highest capability of cross-linking chemistry. Manipulating the base pairs closest to the DNA interstrand cross-link is desirable as it does not directly involve the cross-link, yet sequence variation is known to influence the capabilities, as has been shown previously in this chapter.

With this in mind, the next step was to vary the 5' nearest neighbor of the abasic site, as it is not involved directly in the Ap-dC cross-link, yet is in close enough proximity to have an effect on the chemistry. Figure 17 outlines the sequence variations that were used in this set of

Figure 17

experiments, with Duplex F participating once again, as it was shown to have the highest yield thus far. All four base pairings were used (Duplexes F, N, O, and P), but an additional base pairing (Duplex Q) was used where an adenine-adenine mismatch was included. It had become clear that mismatches in the sequence were vital to the cross-linking capability, as the mismatch to the dC residue was necessary for substantial yields. It can be theorized that further degradation of the DNA around the cross-link could create increased "flexibility" of the duplex, allowing for more interactions between the abasic site and the deoxycytidine residue.

Reaction conditions for this set of experiments remained the same (pH 5.2, no $NaCNBH_3$, and 72-hour incubation at 37˚C) and gel electrophoresis was utilized to determine cross-linking capabilities. Slow-moving bands for all five duplexes were observed, as is shown in Figure 18. Yields for Duplex F (the C-G pair) remained high at 61 ± 3%, very similar to what has been reported previously in this chapter. Yields for all other duplexes were substantially lower, though still relatively high. Duplex N (the G-C pair) reported the lowest yield at 31 ± 2%. Duplex O (A-T pair) and P (T-A pair) reported nearly identical yields at 38 ± 2% and 39 ± 4%,

Figure 18

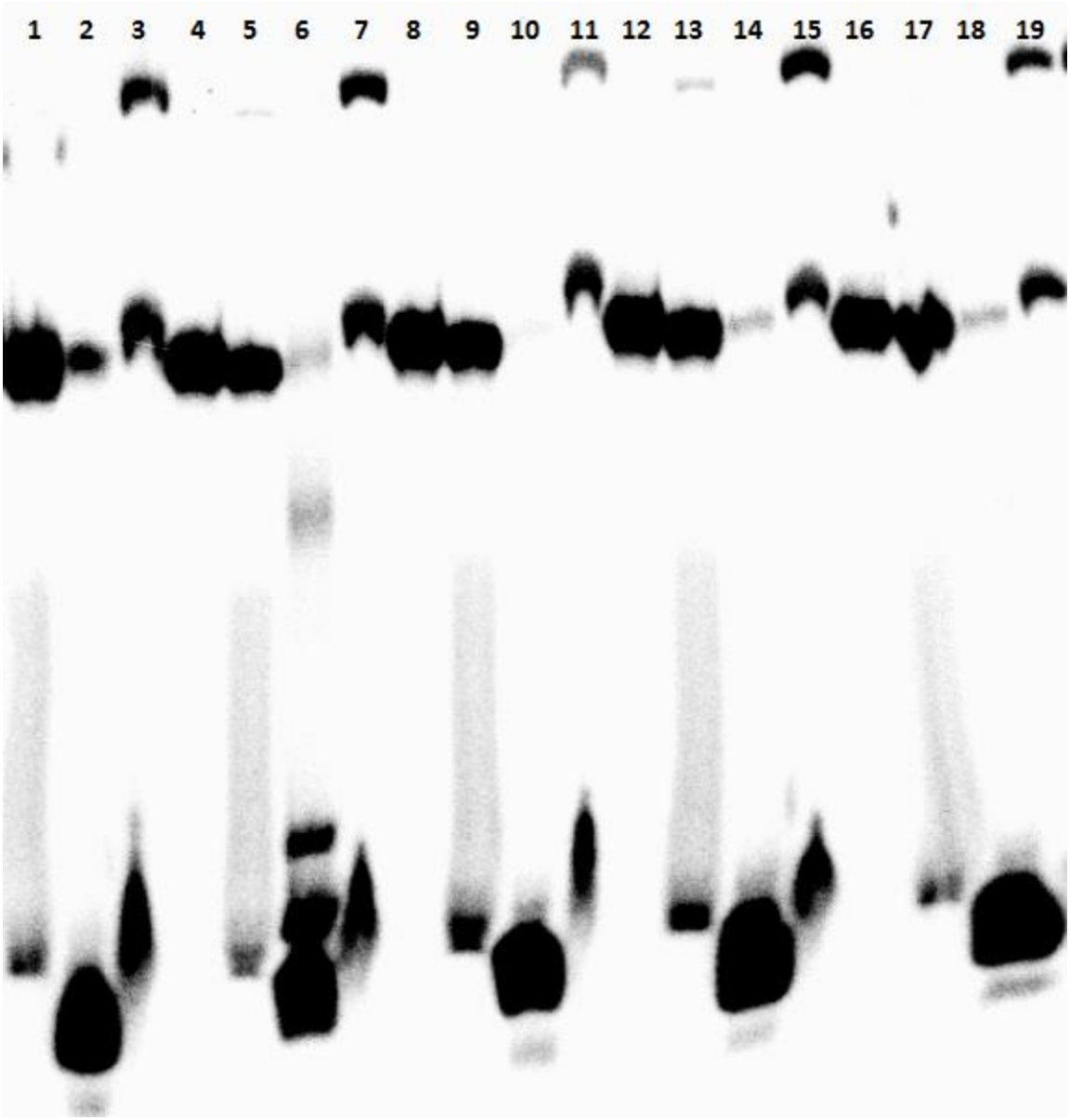


Figure 11. Formation of Ap-dC cross-links when varying the 5' nearest neighbor base pairs to the abasic site. Gel electrophoresis analysis utilized a 0.4 mm thick 20% denaturing polyacrylamide gel. Quantitative measurement of the bands obtained using phosphorimager analysis. Lane 3 corresponds to Duplex **O**, Lane 7 corresponds to Duplex **P**, Lane 11 corresponds to Duplex **N**, Lane 15 corresponds to Duplex **F**, and Lane 19 corresponds to Duplex **Q**. Lanes 2, 6, 10, 14, and 18 are the Ap-containing duplexes **O**, **P**, **F**, **N**, and **Q**, respectively, treated with 1 M piperidine at 95˚C for 30 minutes. Lanes 1, 5, 9, 13, and 17 are the size marker consisting of the ^{32}P-labeled Ap-containing duplexes **O**, **P**, **F**, **N**, and **Q**, respectively. Lanes 4, 8, 12, and 16 are the size marker consisting of the ^{32}P-labeled dU-containing duplexes **P**, **N**, **F**, and **Q**, respectively. The size marker for Duplex **O** consisting of the ^{32}P-labeled dU-containing strand was unavailable.

Table 5: Varying 5' Nearest Neighbor Base Pairs of the Abasic Site in the Ap-dC Cross-Link Yields

Sequence = X Ap A Y A C	
Sample	**Average Cross-Link Yield**
Duplex F (X = C, Y = G)	61 ± 3%
Duplex N (X = G, Y = C)	31 ± 2%
Duplex O (X = A, Y = T)	38 ± 2%
Duplex P (X = T, Y = A)	39 ± 4%
Duplex Q (X = A, Y = A)	44 ± 2%

Table 5: ± values reflect standard deviation of seven replicates used across two experiments.

Figure 19

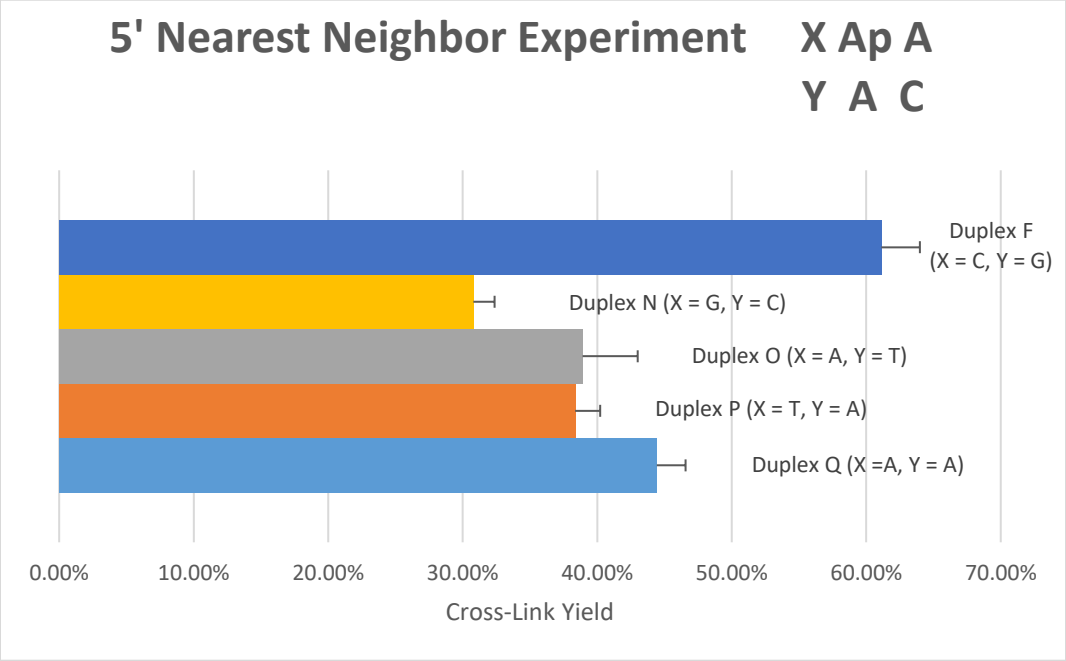


Figure 19. Bar graph of Ap-dC cross-link yields resulting from Duplexes **F**, **N**, **O**, **P**, and **Q** subjected to consistent reaction conditions. Yields are qualitatively obtained from analysis of gel electrophoresis

respectively. Duplex Q (the A-A mismatch) reported the second-highest yield at 44 ± 2%. All yields are reported in Table 5 and displayed in Figure 19.

Results from this experiment yielded interesting trends, as the reported percent yields varied quite a bit between sequences. It was expected that Duplexes O and P would have similar yields, as the difference between the duplexes is an adenine-thymine pair vs a thymine-adenine pair, with the first nucleotide listed being on the Ap-containing strand. This was observed to be true as the yields were less than 1% apart from one another. That would imply that Duplexes F and N should have similar yields as well, as the only differences are a cytosine-guanine pair versus a guanine-cytosine pair. This was far from the case as Duplex F reported the highest yield at 61± 3% and Duplex N reported the lowest yield at 31 ± 2%. This anomaly is further investigated in Section 2.6 of this chapter.

Disregarding the results of Duplex F for the time being, it is understandable that the yields obtained from Duplexes O and P are slightly higher than those reported for Duplex N, as the adenine-thymine base pair results in only two hydrogen bonds, and the guanine-cytosine pair produced three. This extra hydrogen bond in Duplex N may restrict the flexibility of the duplexes sufficiently enough to be responsible for the roughly 8% lower yield of cross-linking capability. The issue of duplex flexibility is further supported by the fact that a mismatched adenine-adenine pair in the same location (Duplex Q) results in a higher cross-linking yield than all three of the correctly-matched pairs (Duplex F excluded), with a 44 ± 2% yield. This mismatched pair is still 17% lower than the yield seen for the cytosine-guanine pair present in the Duplex F sequence.

2.6 Hydroxyl Radical Footprinting and Overhanging Tail Experiments Show the Cross-Link Takes Place Between the Ap site and the dC Residue

The unusually high yields of Duplex F when compared to others of its nature raise suspicion that a cross-link between the abasic site and the 5' guanine residue on the opposing strand may be contributing to the yields observed from the slow-moving bands. To further investigate the location of the cross-link, the ^{32}P Ap-site containing Duplex F (no cross-link) and the cross-linked form of the duplex were subjected to hydroxyl radical footprinting. In this type of experiment, the cross-link must first by purified by running the sample on an electrophoresis gel to separate the cross-linked product from the uncross-linked material. Following purification, the sample is placed in a mixture of Fe^{2+}, EDTA, and H_2O_2 is used to create controlled hydroxyl radicals that then interact with the DNA backbone, cleaving between bases at less than one site per duplex. This reaction eventually occurs until strands exist where cleavage has occurred at each point along the duplex. When a footprinting experiment like this is run on an electrophoresis gel, it can be imaged to see at exactly where the cross-link occurs, as the cross-link causes slow-moving bands due to the linkage to the complimentary strand. This creates a significant drop-off on the gel, making identification of the exact base participating in the cross-link evident.

Using this method, Figure 20 was obtained, demonstrating that the cross-link does occur at the dC residue at the 3' side of the abasic site on the opposing strand. This is consistent with data showing the cross-link does not occur with the dG residue at the 5' side of the abasic site in the absence of reducing conditions reported by Gamboa-Varela[31]. An unknown characteristic of a cytosine-guanine base pair as the 5' nearest neighbor to the abasic site may not create the

Figure 20

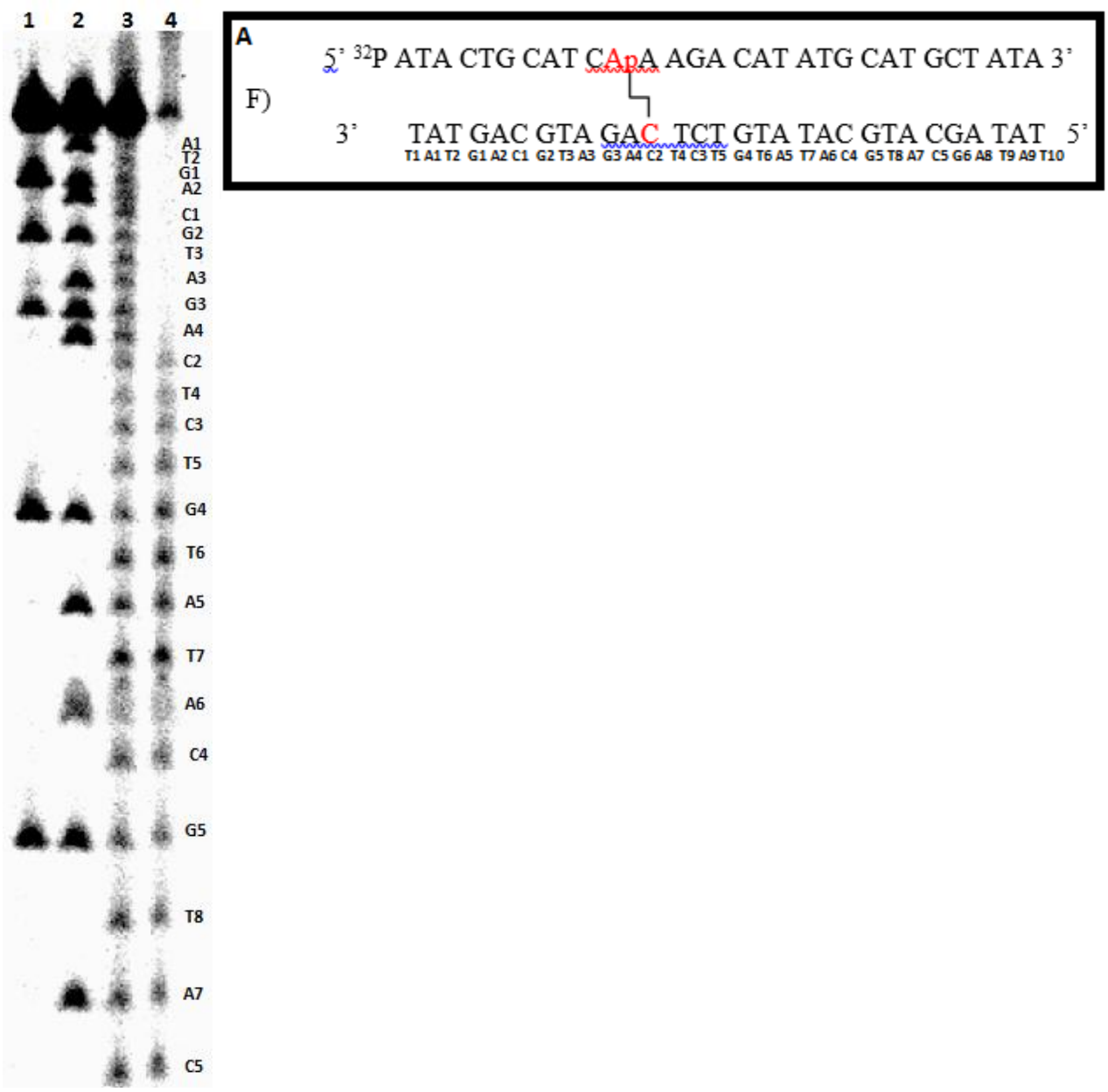


Figure 20. Hydroxyl radical footprinting of the Ap-dC cross-link formed in Duplex **F**. Lane 1, Maxam-Gilbert G sequencing of Duplex **F**. Lane 2, Maxam-Gilbert A+G sequencing of Duplex **F**. Lane 3, hydroxyl radical footprinting of a single dU-containing strand from Duplex **F**. Lane 4, hydroxyl radical footprinting of Duplex **F** incubated in sodium acetate buffer (pH 5.2) for 72 hours. Panel **A**: Duplex **A** with bases labeled.

optimum conditions for the cross-link chemistry when compared to the reactivity of the mismatched cytosine on the 3' side of the abasic site, as 100% of the cross-link was shown to occur at this site, as seen in Figure 20.

To ensure this was also not a derivative of a "nicked" cross-link as was seen by Yang in the Gates group when working with the spermine-mediated cross-links, the cross-linking experiment was performed on Duplex F including overhanging tails on both the parent and complimentary strands. In this study, Duplexes R, S, and T (Shown in Figure 21) were all used to ensure the slow-moving bands observed in Duplex F contained full strands on both sides. Figure 22 displays the gel results, showing slow moving bands of equal length for Duplexes R and S,

Figure 21

F
5′ ^{32}P-ATA CTG CAT CXA AGA CAT ATG CAT GCT ATA
3′-TAT GAC GTA GAC TCT GTA TAC GTA CGA TAT

R
5′ ^{32}P-ATA CTG CAT CXA AGA CAT ATG CAT GCT ATA CGATGAC
3′-TAT GAC GTA GAC TCT GTA TAC GTA CGA TAT

S
5′ ^{32}P-ATA CTG CAT CXA AGA CAT ATG CAT GCT ATA
3′-CAGTAGC TAT GAC GTA GAC TCT GTA TAC GTA CGA TAT

T
5′ ^{32}P-ATA CTG CAT CXA AGA CAT ATG CAT GCT ATA CGATGAC
3′-CAGTAGC TAT GAC GTA GAC TCT GTA TAC GTA CGA TAT

with an even slower band for Duplex T, indicating that the cross-link is not "nicked" at the Ap site and does contain both full strands of the duplex. The results from both the footprinting and overhanging tail experiments confirm that Duplex F is a full duplex cross-link between Ap-dC. The overhanging tail gel showed the slowest moving band with Duplex T, and same length cross-linking bands with Duplexes R and S. It is worthy to note that when Duplex S was treated with a piperidine work-up, no cleavage was noted, as is seen with Duplex R and T. This was the

Figure 22

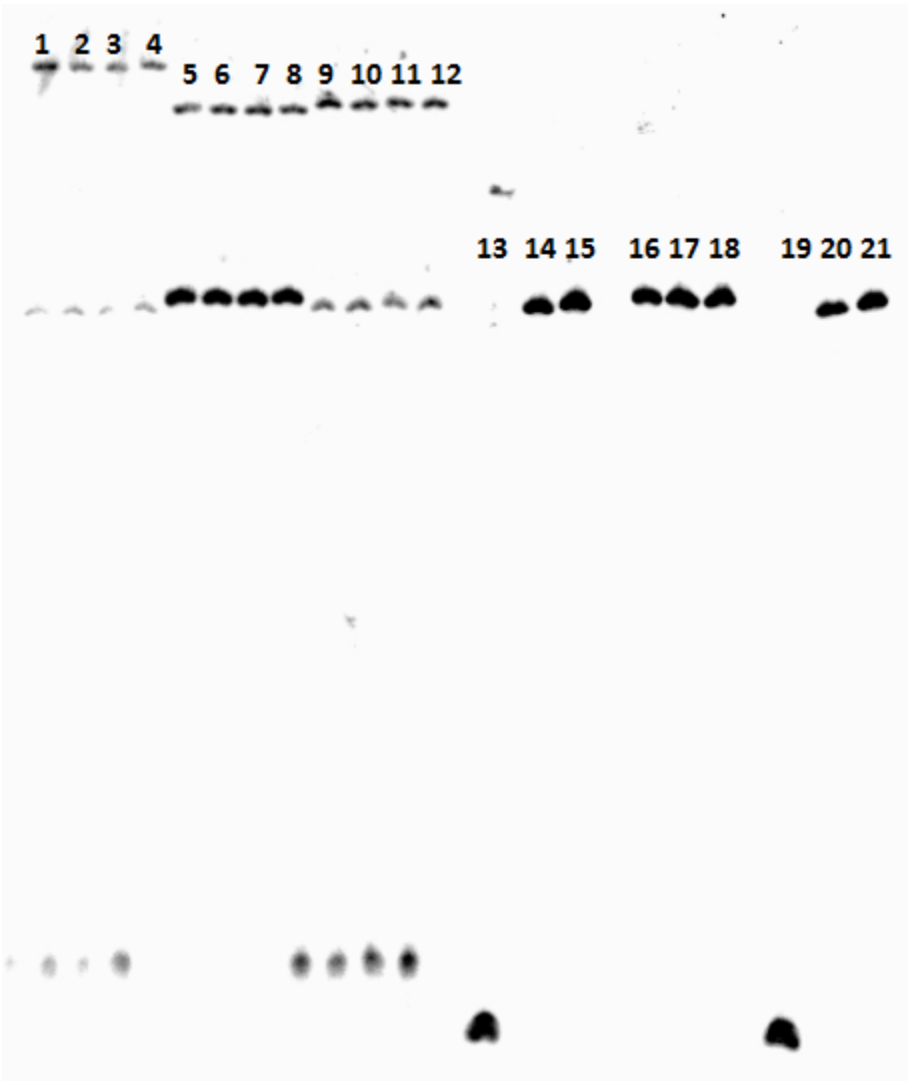


Figure 22. Formation of Ap-dC cross-links when using an overhanging tail on complimenting strands of the duplex. Gel electrophoresis analysis utilized a 0.4 mm thick 20% denaturing polyacrylamide gel. Lanes 1-4 correspond to Duplex **T**, Lanes 5-8 correspond to Duplex **S**, and Lanes 9-12 correspond to Duplex **R**. Lanes 13, 16, and 19 are the Ap-containing duplexes **T**, **S**, and **R**, respectively, treated with 1 M piperidine at 95˚C for 30 minutes. Lanes 14, 17, and 20 are the size marker consisting of the ^{32}P-labeled Ap-containing duplexes **T**, **S**, and **R**, respectively. Lanes 15, 18, and 21 are the size marker consisting of the ^{32}P-labeled dU-containing duplexes **T**, **S**, and **R**, respectively.

expected result, due to the fact that Duplex S was labeled on the complimentary, non-Ap-site containing strand, and therefore, was not affected by the piperidine work-up.

2.7 Effects of the 3' Nearest Neighbor Base Pairs of the Ap-dC Cross-Link

With a better understanding of sequence effects on the 5' side of the Ap-dC cross-link, the effect of the 3' base pairs was next studied in an attempt to better understand this specific type of DNA interstrand cross-link. Again, these manipulations are desirable, as they do not directly involve the cross-link residues, yet sequences variation in the 5' nearest neighbor base pairs resulted in marked differences in cross-linking capabilities. Figure 23 outlines the duplexes utilized for this experiment, as all four naturally occurring base pairs were used (Duplexes F, U, V, and W) as well as one adenine-adenine mismatch (Duplex X). The same reaction conditions were held constant, as with the other experiments (pH 5.2, no $NaCNBH_3$, and 72-hour incubation at 37˚C).

Slow-moving bands were observed for Duplexes F, U, V, and W, but issues with loading the gel shown in Figure 24 made cross-linking results for Duplex X unobtainable. Another gel was loaded and imaged in Figure 25 that included all five duplexes, and slow-moving bands were observed for all samples. Yields for the four correctly matched pairs were all very similar (64-69% for Duplexes F, U, and V) with the exception of the guanine-cytosine pair found in Duplex W. This duplex only had a yield of 29 ± 1%, which was 36% lower than the next yield of 65 ± 5% for the thymine-adenine residue. It is unclear why Duplex W had such a comparatively lower yield than the other correctly matched pairs, though the presence of two cytosine residues in a row on the strand opposing the abasic site may interfere with the cross-linking capability. All yields are reported in Table 6 and displayed in Figure 26.

Figure 23

Aside from this anomaly, all three other duplexes reported yields similar to what has been seen for Duplex F to this point, indicating that the 3’ nearest neighbor base pairs, when correctly matched, do not have a great impact on the cross-linking capability. The yields of the adenine-thymine and thymine-adenine base pairs were again similar, as expected at 69 ± 3% and 65 ± 5%, respectively. The adenine-thymine base pair corresponds to Duplex F, and this yield is again consistent with what was seen previously. These yields were in line with the cytosine-guanine base pair (Duplex V) yield at 68 ± 9%, which is not what was seen in the 5’ nearest neighbor set of experiments. The mismatched pair (Duplex X) did afford the highest yield of all at 87 ± 0.4%, indicating that increased flexibility due to sequence degradation is observed. This mismatch on the 3’ side of the cross-link may allow the backbones of the duplexes the necessary freedom of motion to facilitate the cross-link between the neighboring dC residue and abasic site.

Figure 24

Figure 24. Formation of Ap-dC cross-links when varying the 3' nearest neighbor base pairs to the abasic site. Gel electrophoresis analysis utilized a 0.4 mm thick 20% denaturing polyacrylamide gel. Quantitative measurement of the bands obtained using phosphorimager analysis. Lane 3 corresponds to Duplex **W**, Lane 6 corresponds to Duplex **V**, Lane 9 corresponds to Duplex **U**, and Lane 12 corresponds to Duplex **F**. Cross-link of Duplex **X** unavailable, as issues with loading occurred. See Figure **25** for cross-link of Duplex **X**. Lanes 1, 4, 7, 10, and 13 are the Ap-containing duplexes **W**, **V**, **U**, **F**, and **X**, respectively, treated with 1 M piperidine at 95˚C for 30 minutes. Lanes 2, 5, 8, and 11 are the size marker consisting of the ^{32}P-labeled Ap-containing duplexes **W**, **V**, **U**, **F**, and **X**, respectively.

Figure 25

Figure 25. Formation of Ap-dC cross-links when varying the 3' nearest neighbor base pairs to the abasic site. Gel electrophoresis analysis utilized a 0.4 mm thick 20% denaturing polyacrylamide gel. Quantitative measurement of the bands obtained using phosphorimager analysis. Lanes 1 and 2 correspond to Duplex **X**, Lane 3 corresponds to Duplex **W**, Lane 4 corresponds to Duplex **V**, Lane 5 corresponds to Duplex **U**, and Lane 6 corresponds to Duplex **F**.

Table 6: Varying 3' Nearest Neighbor Base Pairs of the Abasic Site in the Ap-dC Cross-Link Yields

Sequence = Ap A X A C Y	
Sample	**Average Cross-Link Yield**
Duplex F (X = A, Y = T)	69 ± 3%
Duplex U (X = T, Y = A)	65 ± 5%
Duplex V (X = C, Y = G)	68 ± 9%
Duplex W (X = G, Y = C)	29 ± 1%
Duplex X (X = A, Y = A)	87 ± 0.4%

Table 6: ± values reflect standard deviation of five replicates used in this experiment.

Figure 26

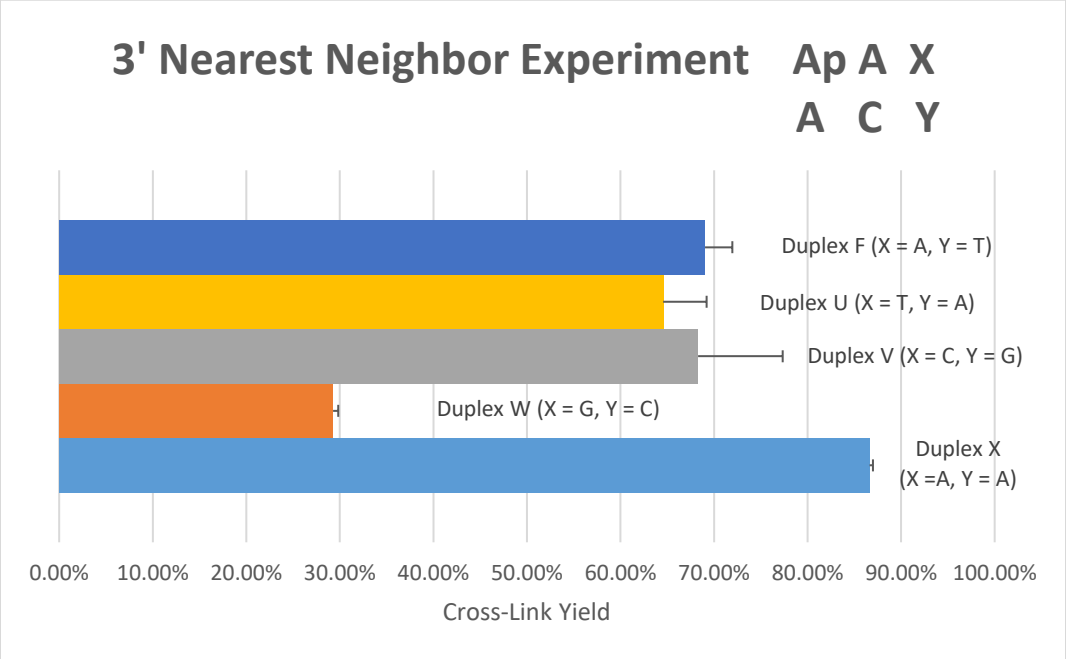


Figure 26. Bar graph of Ap-dC cross-link yields resulting from Duplexes **F**, **U**, **V**, **W**, and **X** subjected to consistent reaction conditions. Yields are qualitatively obtained from analysis of gel electrophoresis

2.8 Effects of the 3' Next Nearest Neighbor Base Pairs of the Ap-dC Cross-Link

With the knowledge of a mismatch at the 3' nearest neighbor position of the Ap-dC cross-link, the next objective was to determine what level of mismatching was necessary to create optimum conditions to facilitate the cross-linking chemistry. This was done by altering the next nearest neighbor on the 3' side of the abasic site. Performing this experiment at the next nearest neighbor of the 5' side was not deemed valuable, as a mismatch was not shown to provide the highest yield, rather a cytosine-guanine base pair (Duplex F) yielded the most cross-link. Therefore, degradation of the duplex at that location would not afford significant flexibility to the strands at the site of the Ap-dC cross-link and would not, theoretically have a great impact on the yield. As was shown in the 3' nearest neighbor experiment, there is opportunity for more duplex flexibility on the 3' end with variation in the next nearest neighbor base pairs, including a mismatch.

To this end, the duplexes in Figure 27 were prepared and tested under the same reaction conditions as discussed previously (pH 5.2, no $NaCNBH_3$, and 72-hour incubation at 37˚C).

Figure 27

X	5'^{32}P-ATA CTG CAT CXA AGA CAT ATG CAT GCT ATA 3'-TAT GAC GTA GAC ACT GTA TAC GTA CGA TAT
Y	5'^{32}P-ATA CTG CAT CXA ACA CAT ATG CAT GCT ATA 3'-TAT GAC GTA GAC AGT GTA TAC GTA CGA TAT
Z	5'^{32}P-ATA CTG CAT CXA AAA CAT ATG CAT GCT ATA 3'-TAT GAC GTA GAC ATT GTA TAC GTA CGA TAT
AA	5'^{32}P-ATA CTG CAT CXA ATA CAT ATG CAT GCT ATA 3'-TAT GAC GTA GAC AAT GTA TAC GTA CGA TAT
AB	5'^{32}P-ATA CTG CAT CXA AAA CAT ATG CAT GCT ATA 3'-TAT GAC GTA GAC AAT GTA TAC GTA CGA TAT

Figure 28

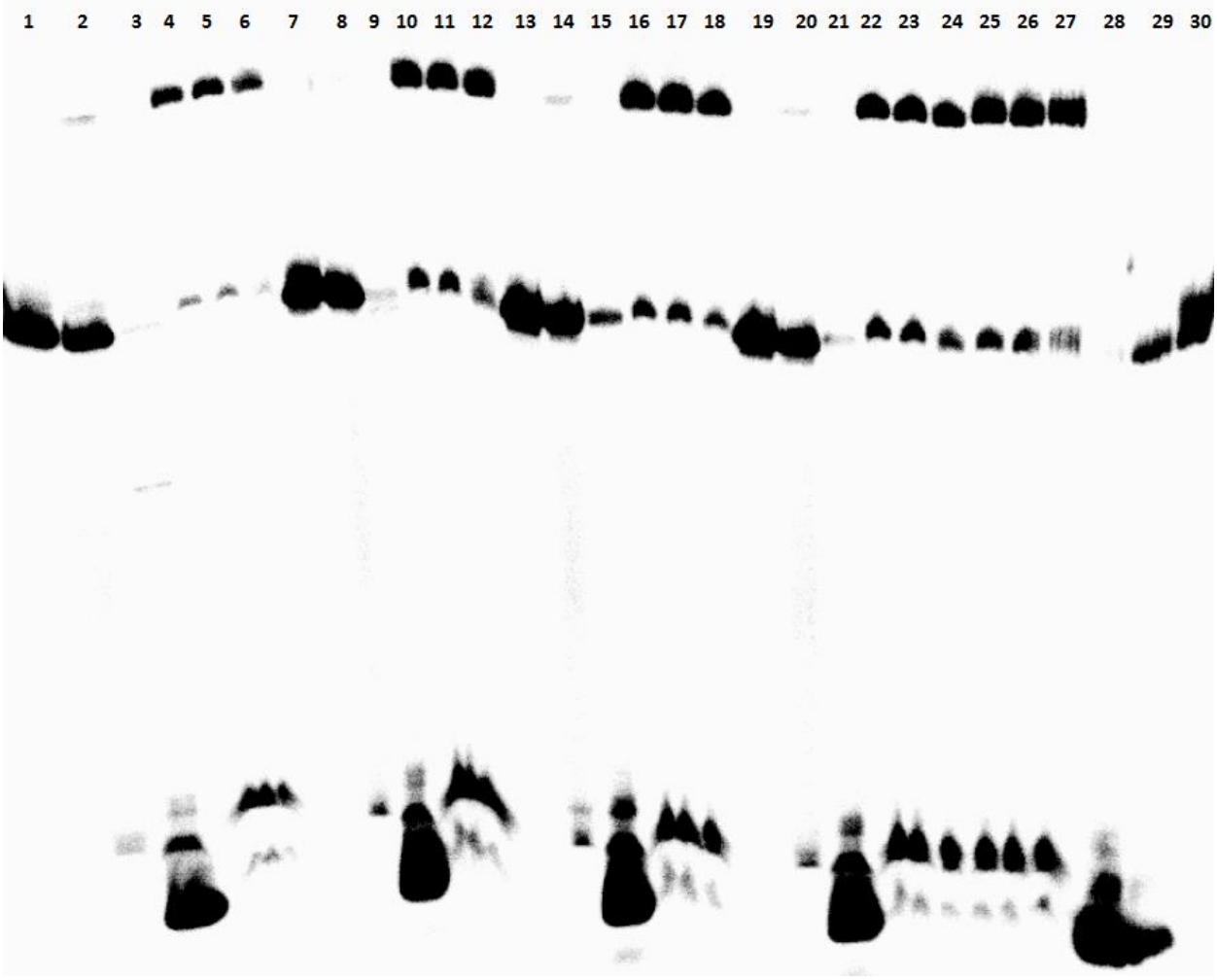


Figure 28. Formation of Ap-dC cross-links when varying the 3' next nearest neighbor base pairs to the abasic site. Gel electrophoresis analysis utilized a 0.4 mm thick 20% denaturing polyacrylamide gel. Quantitative measurement of the bands obtained using phosphorimager analysis. Lanes 4, 5, and 6 correspond to Duplex **Z**; Lanes 10, 11, and 12 correspond to Duplex **AA;** Lanes 16, 17, and 18 correspond to Duplex **X**; Lanes 22, 23, and 24 correspond to Duplex **Y**; and Lanes 25, 26, and 27 correspond to Duplex **AB**. Lanes 3, 9, 15, 21, and 28 are the Ap-containing duplexes **Z**, **AA**, **X**, **Y**, and **AB**, respectively, treated with 1 M piperidine at 95˚C for 30 minutes. Lanes 2, 8, 14, 20, and 29 are the size marker consisting of the ^{32}P-labeled Ap-containing duplexes **Z**, **AA**, **X**, **Y**, and **AB**, respectively. Lanes 1, 7, 13, 19, and 30 are the size marker consisting of the ^{32}P-labeled dU-containing duplexes **Z**, **AA**, **X**, **Y**, and **AB**, respectively.

Table 7: Varying 3' Next Nearest Neighbor Base Pairs of the Abasic Site in the Ap-dC Cross-Link Yields

Sequence = Ap A A X A C A Y	
Sample	**Average Cross-Link Yield**
Duplex X (X = G, Y = C)	78 ± 1%
Duplex Y (X = C, Y = G)	71 ± 2%
Duplex Z (X = A, Y = T)	76 ± 2%
Duplex AA (X = T, Y = A)	71 ± 0.2%
Duplex AB (X = A, Y = A)	76 ± 1%

Table 7: ± values reflect standard deviation of four replicates used in this experiment.

Figure 29

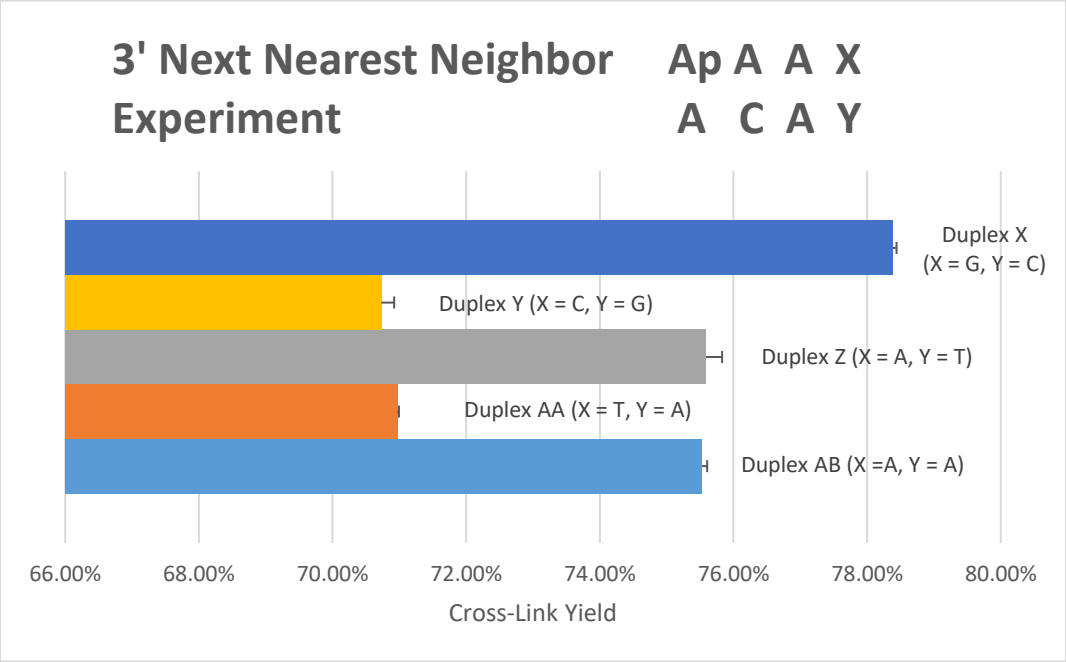


Figure 27. Bar graph of Ap-dC cross-link yields resulting from Duplexes **X**, **Y**, **Z**, **AA**, and **AB** subjected to consistent reaction conditions. Yields are qualitatively obtained from analysis of gel electrophoresis

Slow-moving bands were observed for all five duplexes shown in Figure 27, as can be observed in Figure 28. Yields for all five duplexes were fairly constant and extremely high (70-

78%). All yields are reported in Table 7 and displayed in Figure 29. As was expected, the adenine-thymine pair (Duplex Z) and the thymine-adenine pair (Duplex AA) have very similar yields at 76 ± 2% and 71 ± 0.2%, respectively. The same trend was observed for the guanine-cytosine pair (Duplex X) and the cytosine-guanine pair (Duplex Y) with yields of 78 ± 1% and 71 ± 2%, respectively. A difference between the two hydrogen bonding pairs and three hydrogen bonds pairs was not observed, perhaps due to the increased distance from the site of the Ap-dC cross-link causing these small differences to have less of an effect. The same is true for another adenine-adenine mismatch (Duplex AB), as the yield was right in the middle of the four correctly matched base pairs at 76 ± 1%. At this distance from the Ap-dC cross-link, it does not appear that any amount of additional flexibility pertinent to the cross-linking chemistry is able to be achieved, as the yields did not change significantly depending upon the sequence. This would support the theory that the 3' nearest neighbor position is the ideal location for a mismatch pair that would afford the most duplex flexibility, allowing the Ap-dC cross-link to form.

As Duplex X was shown to have the consistently highest yields over the course of the experiments discussed in this chapter (maximum yields of 87 ± 0.4%), it was desired that the cross-link have the same hydroxyl radical footprinting study as discussed in Section 2.6 of this chapter performed to ensure cross-link was indeed between the Ap site and the dC residue. Results show that this is the case and can be seen below in Figure 30. The hydroxyl radical footprinting again displayed 100% of the sample cross-linked as desired at the dC residue positon found on the 3' side of the abasic site on the opposing strand. With this information, it can be confidently stated, that with the proper sequence variation as seen in Duplex X, strikingly

Figure 30

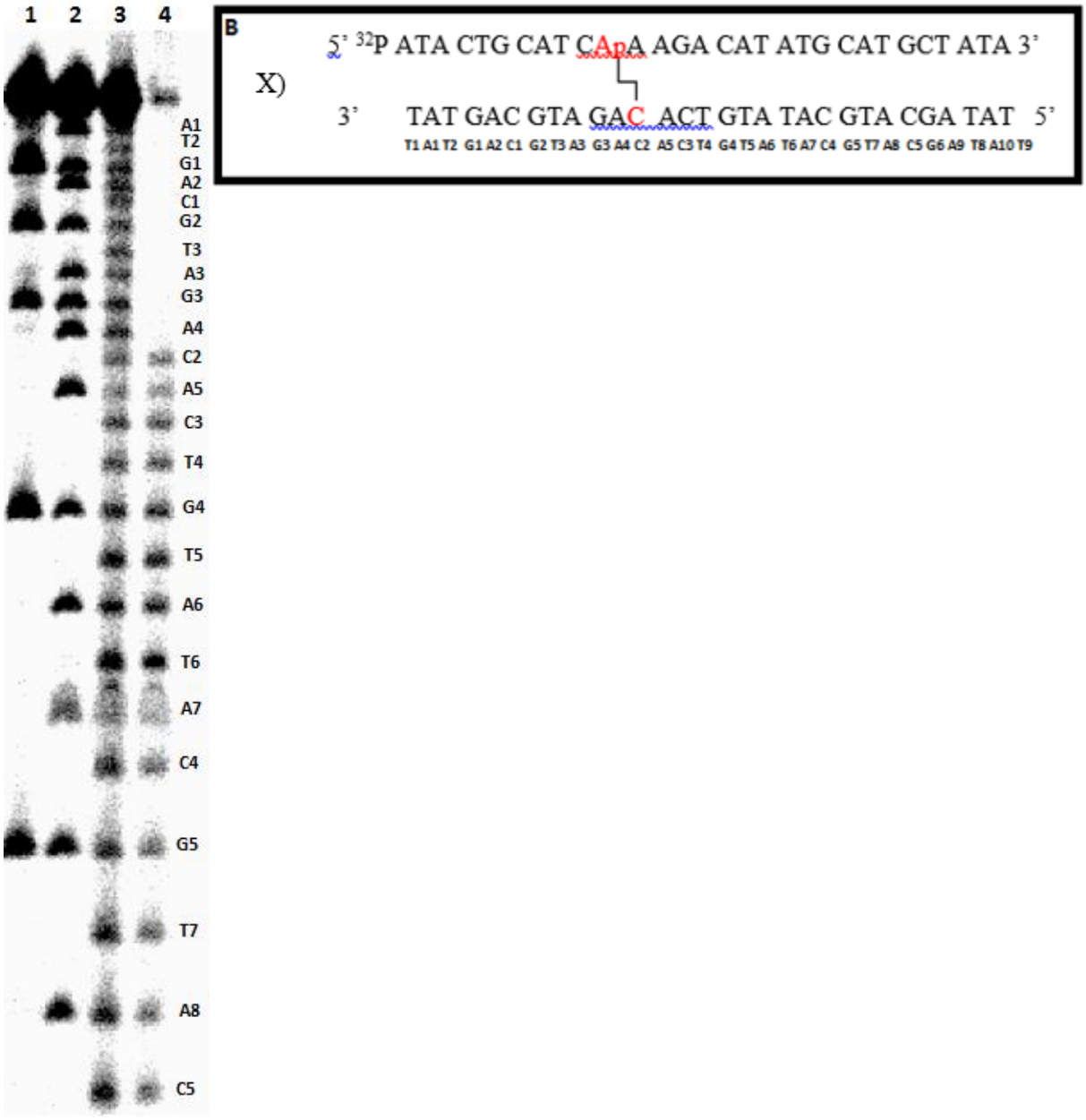


Figure 30. Hydroxyl radical footprinting of the Ap-dC cross-link formed in Duplex **X**. Lane 1, Maxam-Gilbert G sequencing of Duplex **X**. Lane 2, Maxam-Gilbert A+G sequencing of Duplex **X**. Lane 3, hydroxyl radical footprinting of a single dU-containing strand from Duplex **X**. Lane 4, hydroxyl radical footprinting of Duplex **X** incubated in sodium acetate buffer (pH 5.2) for 72 hours. Panel **B**: Duplex **X** with bases labeled.

high yields of a synthesized Ap-dC DNA interstrand cross-link around 80% or higher are able to be consistently obtained through processes described in this chapter.

2.9 Chapter 2 Summary

In this chapter, it has been demonstrated that an Ap-dC cross-link is capable of forming in physiologically relevant conditions when mismatched with a non-guanine nucleotide. The yields reported under physiologically relevant conditions (pH 7.0, no reducing conditions, 37˚C) are relatively high (35 ± 1%) when considering the potency of a single interstrand cross-link.[32] Yields are able to be increased by a substantial amount to nearly 70% when the same reaction is performed at pH 5.2. Sequence variation shows that inclusion of more mismatched pairs surrounding the Ap-dC linkage results in higher yields, likely as a result of increased duplex flexibility allowing the exocyclic amino group of the cytosine to get spatially closer to the abasic lesion on the opposing strand. Further studies showed that mismatched pairs at the next nearest neighbor position have little to no effect on the cross-linking capability, alluding that the nearest neighbor position on either side as most important when determining duplex flexibility.

2.10 Chapter 2 References

27. Dutta, S., Chowdhury, G., Gates, K. S., Interstrand cross-links generated by abasic sites in duplex DNA. *J. Am. Chem. Soc*. 2007. 129(7): 1852-1853.

28. Johnson, K. M., Price, N. E., Wang, J., Fekry, M. I., Dutta, S., Seiner, D. R., Wang, Y., Gates, K. S., On the formation and properties of interstrand DNA-DNA cross-links forged by reaction of an abasic site with the opposing guanine residue of 5'CAp sequences in duplex DNA. *J. Am. Chem. Soc*. 2013. 135(3): 1015-1025.

29. Price, N. E., Johnson, K. M., Wang, J., Fekry, M. I., Wang, Y., Gates, K. S., Interstrand DNA-DNA cross-link formation between adenine residues and abasic sites in duplex DNA. *J. Am. Chem. Soc*. 2014. 136(9): 3483-3490.

30. Johnson, L. M., Hypoxia Selective DNA Alkylating Analogs of Tlrapazamine and Ap-Derived DNA Interstrand Cross-Links. *Ph.D. Dissertation*. University of Missouri, Columbia. 2014.

31. Varela, J. G., Interstrand DNA Cross-Link Formation via Hydrazone Formation between N^4-Aminocytidine and AP-sites. *Ph.D. Dissertation*. University of Missouri, Columbia. 2017.

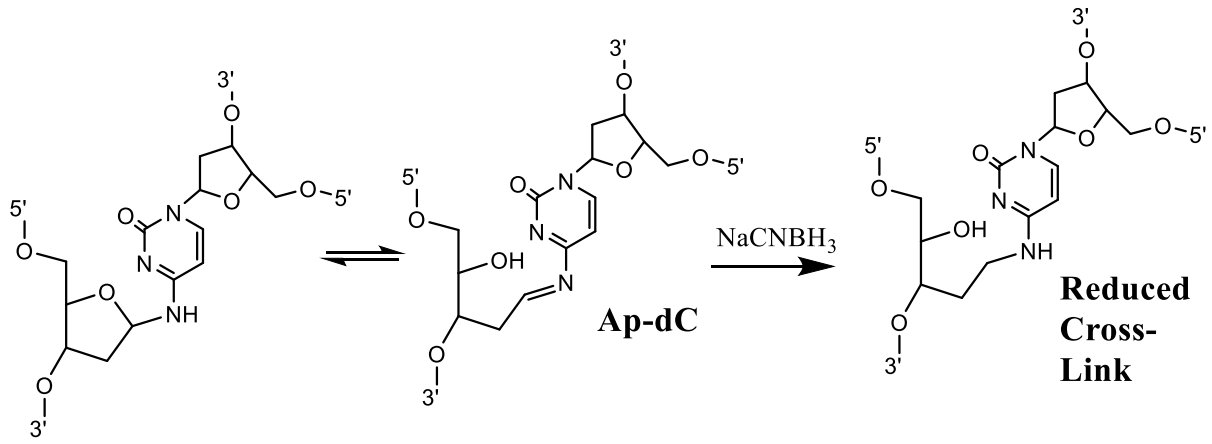
32. Grossmann, K. F., Ward, A. M., Matkovic, M. E., Folias, A. E., Moses, R. E., S-cerevisiae has three pathways for DNA-interstrand crosslink repair. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*. 2002. 487(3-4): 73-83.

Chapter 3: Implications and Impacts of Interstrand DNA Cross-Links between Abasic Site (Ap) and Cytosine Residue

3.1 Next Steps in the Ap-dC DNA Interstrand Cross-Link Line of Research

While the experiments and results reported in the previous chapter do shed quite a bit of light on the Ap-dC cross-link, there is a considerable amount of information that we do not know. The first question is determining why the addition of $NaCNBH_3$ does not increase the yield of the cross-linking reaction. The addition of this reducing agent should facilitate the irreversible formation of the ring-opened Ap-dC cross-link from the imine as shown in Scheme 7. However, this is not the case with both the Ap-dC cross-link and the Ap-dA cross-link, as reported by Price.[33] It may be possible that the conditions in which this reaction has been ran were sufficient to create an environment that would facilitate reductive amination to support the ring-open amine formation without addition of $NaCNBH_3$.

Scheme 7



Non-radioactively labeled samples of the cross-link could be subjected to Mass Spectrometry and HNMR to elucidate the exact structure of the cross-link. An issue with this would be the sensitivity of the Mass Spectrometer needed to determine the 2 atomic mass unit difference between the reduced and oxidized cross-links would be difficult to achieve with such a large sample. It has been demonstrated that performing a four-enzyme digest of these DNA duplexes is possible, and will result in the complete cleavage in between each nucleotide, yet leaving the covalent cross-link intact.[33-34] This would reduce the sample's overall mass significantly, making the needed sensitivity much easier to achieve. Knowing the exact structure of the cross-link would help to answer the question of the cross-link existing in the ring-opened or ring-closed form.

Once more is known about the exact structure and formation of the cross-link, questions can then begin to be answered about conditions. It would be interesting to determine the exact length of time necessary for the cross-link to form with the knowledge now gained about conditions and sequence to maximize yield. Running electrophoresis gels at certain time points of the reaction (e.g. 1 hour, 4 hours, 12 hours, 24 hours, etc.) would be interesting to determine if significant cross-linking is achieved on a different time scale. This work may lead to higher yields than what is reported here, as there may be an optimum incubation time outside of 72 hours. Along these same lines, it may be interesting to determine the stability of the cross-link, once an optimum incubation time has been determined, to see at what point we begin to see significant breakdown of the cross-link. This could be achieved by allowing the cross-link to continue incubation after the time of the maximum yield and take samples at time points 1, 3, and 7 days, for example, post-maximum yield to determine if a drop off does occur after a certain time period.

Finally, sequence variation has been shown to have a tremendous effect on the cross-link and its ability to form. Different variations in sequence have yet to be explored, such as those describe in Figure 31. There could be value in determining if these types of variations in sequence could result in higher yields as well. This would promote the physiological relevance of the Ap-dC cross-link, as it will no longer be limited to the sequences already discussed. Exploring these suggested sequences, and more, will provide greater context in which this Ap-dC cross-link can exist in the human genome.

Figure 31

3.2 Applications of the Synthesized Ap-dC DNA Interstrand Cross-Link

In the previous chapter, it was outlined how significantly high yields of the Ap-dC cross-link were able to be obtained. Purification of the cross-link was shown to be possible, resulting in a sample containing only the cross-linked duplex. The ability to synthesize this cross-link

with relatively good yields (~70%) from readily available reagents opens several doors for future DNA chemistry.

Perhaps the most exciting of these avenues is its application in DNA repair studies. There are several cross-link repair enzymes, such as NEIL3, that are capable of repairing in vitro synthesized DNA interstrand cross-links, with research focused on understanding the mechanism and scope of the repair.[35] Providing a new method to synthesize this specific type of Ap-dC DNA interstrand cross-link could prove to be useful in providing target duplexes for repair studies with enzymes like these. This method could provide a catalog of sequences containing Ap-dC cross-links used in these studies that could lead to better understanding of the repair pathway. Better understanding of ICL repair can lead to advancements in pharmaceuticals specifically designed to target in vivo ICLs as a method of treatment of disease.

Furthermore, cross-linking agents have been shown to be effective in the targeting of tumor cells and the induction of cell death as a means to treat cancerous tumors.[36] Further study of this specific type of Ap-dC DNA interstrand cross-link and its reaction conditions and formation could lead to new cross-linking agents. These Ap-dC inducing agents would then be available for use in this type of anti-cancer study to produce pharmaceuticals that may later be used to in this type of anti-cancer drug line.

Lastly, this type of DNA lesion has demonstrated the ability to be formed in physiological conditions. Research of common sequence motifs in the human genome that contain sections similar to the duplexes listed here or used in future studies of Ap-dC cross-link capability would then be able to be identified as prone to this type of DNA damage. Specifically, motifs already prone to a mismatched pair involving a cytosine residue would be

prone this type of cross-linking damage, as abasic site formation on a nearby nucleotide would be sufficient to create the cross-link, as displayed in this thesis. Identification of these types of motifs could lead to early detection of cancer-prone cells in the body, in the event that the possible cross-links are able to accumulate sufficiently.[37]

3.3 Chapter 3 Summary

Throughout this thesis, a method of synthesizing DNA interstrand cross-link between an abasic site and a deoxycytidine residue has been demonstrated as effective in relatively high yields. The work outlined in Chapter 2 of this thesis has laid the groundwork for future experiments in this line of research, such as time and sequence manipulation. Spectrometry of the cross-link itself may also lead to better understand of its and other cross-link's formation and stability. The capability of producing a pure Ap-dC DNA interstrand cross-link may be useful in other avenues of research relating to anti-cancer drugs and DNA repair enzymes.

3.4 Chapter 3 References

33. Price, N. E., Johnson, K. M., Wang, J., Fekry, M. I., Wang, Y., Gates, K. S., Interstrand DNA-DNA cross-link formation between adenine residues and abasic sites in duplex DNA. *J. Am. Chem. Soc*. 2014. 136(9): 3483-3490.

34. Price, N. E, Catalano, M. J., Liu, S., Wang, Y., Gates, K. S. Chemical and Structural Characterization of Interstrand Cross-Link Formed between Abasic Sites and Adenine Residues in Duplex DNA. *Nucleic Acids Research*. 2015. 43 (7): 3434-3441.

35. Imani Nejad, M, Housh, K., Rodriguez, A. A., Kathe, S., Wallace, S. S., Eichman, B. F., Gates, K. S. Unhooking of an Interstrand Cross-Link at DNA Fork Structure by the DNA Glycosylase NEIL3. *Submitted to JACS.* 2019.

36. Chen, W., Balakrishnan, K., Kuang, Y., Han, Y., Fu, M., Gandhi, V., Peng, X. Reactive Oxygen Species (ROS) Inducible DNA Cross-Linking Agents and Their Effect on Cancer Cells and Normal Lymphocytes. *Journal of Medicinal Chemistry*. 2014. 57, 4498-4510.

37. Hoy, C. A., Thompson, L. H., Mooney, C. L., Slazar, E. P. Defective DNA Cross-Link Removal in Chinese Hamster Cell Mutants Hypersensitive to Bifunctional Alkylating Agents. *Cancer Research*. 1985. 45, 1737-1743.

VITAE

Luke Edward Pierce, second of four siblings, was born to Laura Michelle Pierce and David Ervin Pierce in Bellevue, Washington. He grew up in Washington, Idaho, Kansas, before settling in Belton, Missouri for Middle and High School. He attended the University of Missouri and graduated in 2016, obtaining undergraduate degrees in Biological Sciences and Spanish Language and Literature, with a minor in Chemistry. After getting married to Adilenne Licona Gonzalez in May 2017, he started graduate school that fall to obtain his Masters in Chemistry, joining Dr. Kent Gates research group. He will (hopefully) obtain his degree in December 2019. He has worked at the University of Missouri Research Reactor (MURR) as a full-time Research Specialist developing active ingredients in the treatment of thyroid cancer throughout the duration of his graduate degree. He plans to continue his career of pharmaceutical development, and cancer/DNA research after receiving his degree.