2018

DNA Lesions Produced from the Reaction of Diols and 5-Formylcytosine and Their Effects on DNA Replication

Brock Allen
University of the Pacific

DNA LESIONS PRODUCED FROM THE REACTION OF ENVIRONMENTAL TOXINS AND 5-FORMYLCYTOSINE AND THEIR EFFECTS ON DNA REPLICATION

by

Brock Allen

A Thesis Submitted to the

Graduate School

In Partial Fulfillment of the

Requirements for the Degree of

MASTER OF SCIENCE

Thomas J. Long School of Pharmacy and Health Sciences
Pharmaceutical and Chemical Sciences

University of the Pacific
Stockton, CA

2018

DNA LESIONS PRODUCED FROM THE REACTION OF ENVIRONMENTAL TOXINS AND 5-FORMYLCYTOSINE AND THEIR EFFECTS ON DNA REPLICATION

by

Brock Allen

APPROVED BY:

Thesis Advisor: Liang Xue, Ph.D.

Committee Member: Jerry Tsai, Ph.D.

Committee Member: Lisa Wrischnik, Ph.D.

Department Co-Chair: Jianhua Ren, Ph.D.

Dean of Graduate School: Thomas H Naehr Ph.D.

DNA LESIONS PRODUCED FROM THE REACTION OF ENVIRONMENTAL TOXINS AND 5-FORMYLCYTOSINE AND THEIR EFFECTS ON DNA REPLICATION

by

Brock Allen

DEDICATION

This thesis is dedicated to my family, especially my parents, for their constant support and encouragement throughout this process.

ACKNOWLEDGMENTS

I would like to extend my thanks to my advisor Dr. Liang Xue. He initially sparked my interest for research while I was an undergraduate in his lab and has continued to provide me with support and motivation throughout my time as a graduate. This project would not have been possible without his ongoing time, effort and guidance.

I would also like to thank Dr. Jerry Tsai and Dr. Lisa Wrischnik for agreeing to serve on my thesis committee. Their support and feedback has been a crucial factor in helping me get to this point.

I would like to thank Krege Christison for helping me to run ESI-MS samples, as well as providing guidance and insight regarding data analysis. In addition, thank you to the Pacific Mass Spectrometry facility for allowing me to use the GC/MS to run many hours of samples, as well as providing instrument support and assistance with data analysis.

I would also like to thank Dr. Hyun Joo for his time and effort assisting with the computational portions of the study. Molecular modelling can sometimes be like a different language, but Dr. Joo makes it quite easy to understand.

In addition, I would like to thank Dr. Brett Williams and Rosemarie Basi for their constant help in the stockroom, as well as Susan McCann for her help on the administrative side of things. This department would not be able to run without them.

My deepest thanks go to Siwen Wang for being the best lab mate anyone could ask for. She has been a constant source of motivation, feedback, constructive criticism, laughter and friendship throughout my years in this lab.

I would also like to thank the other members of Dr. Xue's research group; Mandeep, Vanessa, Mingheng, Danni and all of the other undergraduate students. They have always been willing to listen and provide helpful feedback.

Lastly, I would like to thank my close friends in the department: Michael Pastor and Taylor Rabara, as well as many others, for their support throughout this process as well.

DNA Lesions Produced from the Reaction of Environmental Toxins and 5-Formylcytosine and their Effects on DNA Replication.

Abstract

by Brock Allen

University of the Pacific
2018

Nucleic acids are complex macromolecules that are susceptible to both endogenous and exogenous damage. This study explored damage resulting from interactions with environmental nucleophilic toxins, such as a variety of diols and amines found in industry. These nucleophiles can react with electrophilic groups, such as 5-formylcytosine. 5-formylcytosine is an oxidation product of the epigenetic base 5-methylcytosine. It is typically removed by thymine DNA glycosylase (TDG) but is known to accumulate in the genome, making the formyl group susceptible to attack. In this study we used GC/MS and ESI-MS to show that DNA lesions from the nucleophilic addition reaction of toxins and 5-formylcytosine can be formed under physiological conditions. In addition, this formation showed a pH dependency, with lower pHs showing more product formation. Studies with a lesion formed from the reaction of 1,3-propane diol and 5-formylcytosine showed that the lesion has little effect on the

conformation of the DNA duplex. UV thermal denaturation studies showed that at a glance the lesion also has little effect on the stability of the DNA duplex, however, more extensive studies revealed a slight destabilization effect due to the lesion. Enzymatic studies showed that the presence of one lesion does not have a significant effect on the ability of DNA polymerase to efficiently complete DNA replication with high fidelity, but when the lesion was incorrectly base paired, the extension reactions resulted in deletion products or a halt in replication. Addition of a second tandem lesion to the template resulted in a decrease in fidelity, while continuing to give deletion products and replication stops in the presence of mismatched base pairs. This is particularly significant, indicating the potential for the lesion to be mutagenic or even cytotoxic. Lesions formed from other environmental toxins could be even more damaging, making them well worth future investigation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: DNA Damage and 5-Formylcytosine

Nucleic acids (DNA and RNA) are one of the four classes of major biological macromolecules. They are composed of nucleotides, each containing a nitrogenous base and 5-carbon sugar molecule linked together by a phosphate backbone. There are 5 canonical nitrogenous bases: guanine, adenine, cytosine and thymine in DNA and uracil in RNA (Figure 1-1). The pairing and unique sequence of the nucleotides constitutes the genetic code, a blueprint for the entire organism. In the average human there are 6 billion base pairs per cell and could reach almost up to six feet if unwound and laid out from end to end.

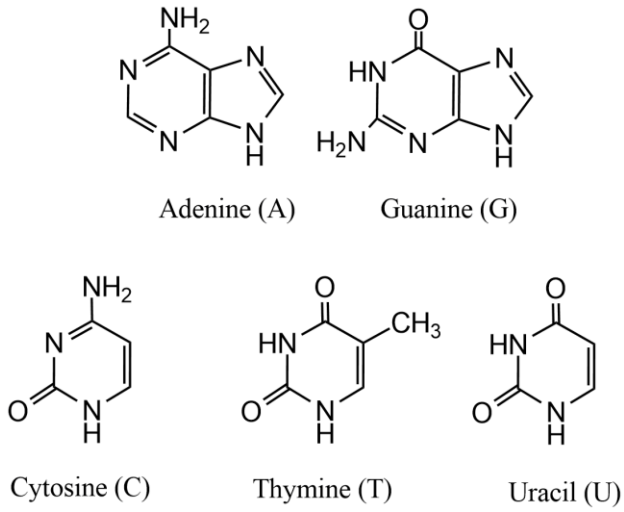


Figure 1-1. Canonical Nitrogenous Bases Found in DNA and RNA

When packed in a cell DNA and RNA can also form complex secondary and tertiary structures. These higher order structures help in packing, but also play a role in the regulation of expression. With such an intricate and large-scale design, it follows that there are numerous ways for nucleic acids to be damaged. This study explores some of the potential sources of damage to this extensive system and their potential effects on cells.

Endogenous and Exogenous DNA Damage

There are two main types of DNA damage: endogenous and exogenous. Endogenous DNA damage results from the interaction of DNA and species originating from natural sources in the cell. These natural sources include things such as reactive oxygen species and alkylating agents. Reactive oxygen species, as the name states, are chemically reactive compounds that contain oxygen and include things such as peroxides, superoxides, hydroxyl radicals and singlet oxygens. These compounds have been found to be generated in natural processes such as respiration, cell injury, hydroxylation of steroids and drugs, inflammation and phagocytosis.[1]

For example, 1-5% of the oxygen used during mitochondrial respiration undergoes a series of electron transfers and degradations to form hydroxyl radicals.[2 3] These hydroxyl radicals are extremely reactive and can have large damaging effects on DNA. Guanine bases are the most susceptible to this type of oxidative damage, resulting in the formation of 8-oxoguanine, one of the most common DNA lesions (Figure 1-2). Previous studies have shown that a single 8-oxoguanine can raise the mutation frequency of replication more than 30-fold, about 0.7%. The vast majority of these mutations are

G-T mismatches in base pairing.[4] This mismatched thymine will be paired with adenine in future rounds of replication, leading to G to T and C to A transversions in the genome.

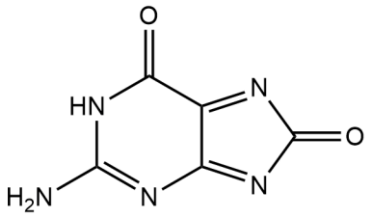


Figure 1-2. 8-oxoguanine

Another source of endogenous DNA damage is alkylating agents, reactive small molecules that cause the replacement of hydrogen with alkyl groups. These include compounds such as betaine and choline, but the most prolific alkylating agent is S-adenosylmethionine (SAM) (Figure 1-3). SAM is a common methyl donating agent that helps facilitate many sanctioned methylation processes, but it can also methylate nonspecifically to form DNA adducts.

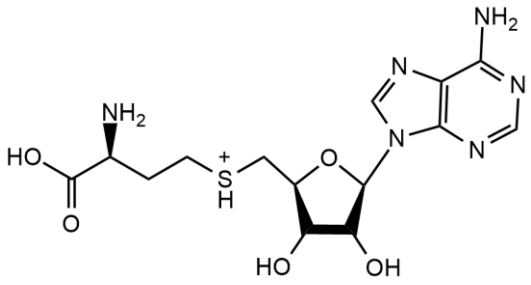


Figure 1-3. S-adenosylmethionine

Studies have shown that nonspecific methylation by SAM can generate approximately four thousand 7-methylguanine and six hundred 3-methyladenine per day in just one cell (Figure 1-4).[5] 7-methylguanine does not alter the coding specificity of the base but it does cause structural destabilization of the glycosyl bond. This strain could result in either depurination leading to the formation of a mutagenic abasic site, or the opening of the imidazole ring, forcing replication to come to a halt.[6] On the other hand, 3-methyladenine is highly cytotoxic, completely blocking replication. SAM is also able to generate more minor pyrimidine lesions such as 3-methylthymine and 3-methylcytosine (Figure 1-4), both of which block DNA replication.[7]

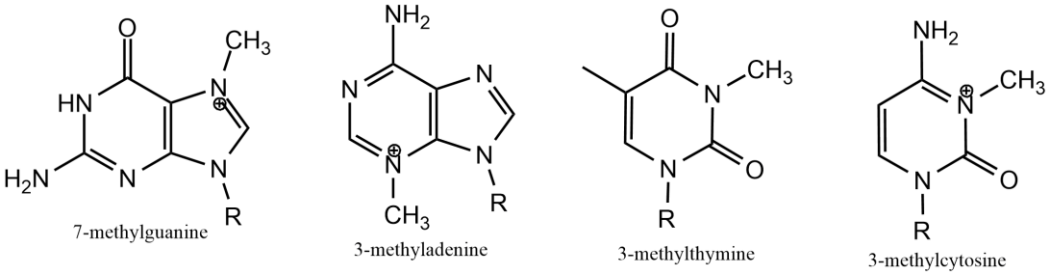


Figure 1-4. Nonspecific Methylation Products of S-adenosylmethionine

Exogenous DNA damage results from the interaction of DNA and species originating outside the organism. These external sources of damage can include things such as ionizing radiation, genotoxic chemicals and UV light. Ultraviolet radiation is known to be one of the most carcinogenic sources of exogenous damage, causing severe alterations to the genome. In particular, UV-B radiation, ranging in wavelength from 280-315 nm can be especially harmful.[8] This type of radiation is easily absorbed by the

double bonds found in pyrimidine DNA bases, causing the bonds to break open. If neighbored by another pyrimidine then these bases can covalently bond, forming a four membered cyclobutane structure (Figure 1-5).[9] These bulky thymine and cytosine dimers can lead to miscoding, making them potentially mutagenic. It is estimated that 50-100 reactions of this type happen in each cell per second of exposure to sunlight.[9] Buildup of these lesions has been shown to have a wide range of adverse effects, such as erythema, suppression of the immune system and in extreme cases even skin cancer.[10]

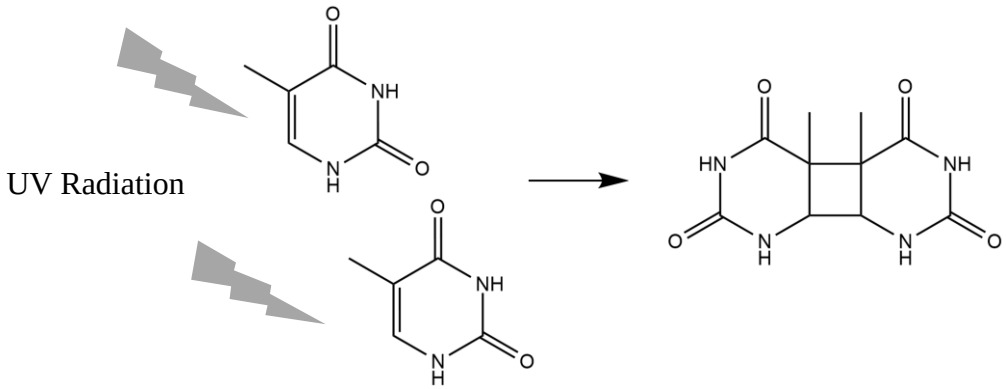


Figure 1-5. Thymine Dimers Resulting from Exposure to UV Radiation

Another source of exogenous DNA damage is ionizing radiation. This is radiation that contains enough energy to liberate electrons from a molecule. This high energy radiation includes things such as gamma rays and X rays and some of the higher end of the UV spectrum. When this radiation comes into contact with DNA, it can break apart base pairs, damage hydrogen bonding and even damage the phosphate backbone, resulting in single or double strand breaks. Hydrogen bonding and single strand breaks can typically be easily repaired by natural repair enzymes, but double strand breaks present more of a problem. If unable to be repaired, these breaks are highly cytotoxic,

resulting in cell death. If cell death is not triggered, it is possible that small fragments of the genome could be lost, or pieces joined together to form non-homologous chromosomes. These errors have drastic effects on the genome and can be highly carcinogenic.

DNA Repair. With such a high potential for mutagenic and cytotoxic DNA lesions to be formed, it becomes crucial to have a way to remove these lesions and repair the DNA. There are two main methods of DNA repair that will be discussed here; base excision repair and nucleotide excision repair.

Base excision repair is a method used for the repair of smaller DNA lesions that do not cause any distortion or conformation change to the DNA helix.[11] This can include the oxidized and alkylated lesions previously discussed. This process is catalyzed by several different groups of enzymes. The repair mechanism is initiated with recognition of the lesion by glycosylases. There are several kinds of glycosylases, each specialized to recognize a specific set of base modifications. The glycosylase will move down the DNA helix, passing each base through its binding pocket by flipping it out of the helix, scanning for particular modifications. When the enzyme locates a modification, it can then use water to initiate cleavage of the N-glycosidic bond between the base and the 2'-deoxyribose, resulting in the formation of an abasic site (AP).

An AP endonuclease can then hydrolytically cleave the phosphodiester bond between the abasic site and the rest of the sequence, leaving a 3' hydroxyl group and a 5' deoxyribose phosphate.[12] DNA polymerase can then synthesize a new segment containing the proper undamaged nucleotide. The process is then completed when DNA

ligase forms a new phosphodiester bond, binding together the new segment with the preexisting DNA sequence (Figure 1-6).

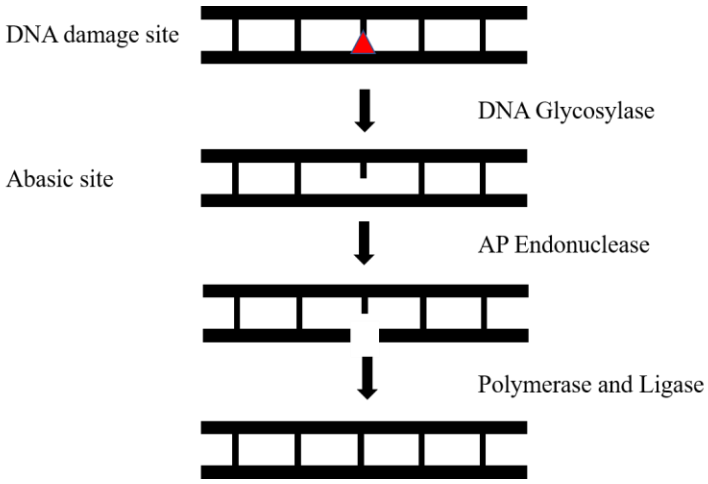


Figure 1-6. Removal of DNA Damage through Base Excision Repair

Nucleotide excision repair (NER) is a similar process but differs in the recognition and removal of the target lesion. NER is specialized for repairing bulky lesions that distort the DNA helix, such as pyrimidine dimers induced by UV light. In this mechanism, the DNA is first scanned by damage sensing proteins, such as DNA damage binding protein (DDB) and XPG-Rad23B. Once a damage site is located these proteins then recruit the repair enzymes. The main repair enzyme involved in this process is transcription factor II H (TFIIH). Two subunits of this transcription factor act as a helicase, unwinding the DNA around the damage site, creating a bubble. This enzyme is stabilized by another enzyme known as XPG, which also has endonuclease activity, allowing it to cut the sequence on either side of the lesion, leaving a gap of 25-

30 nucleotides. The small excised sequence then complexes with the transcription factor and they then dissociate from the helix together.[13] The remaining gap can then be easily resynthesized by DNA polymerase and patched back together by DNA ligase as seen in base excision repair (Figure 1-7).

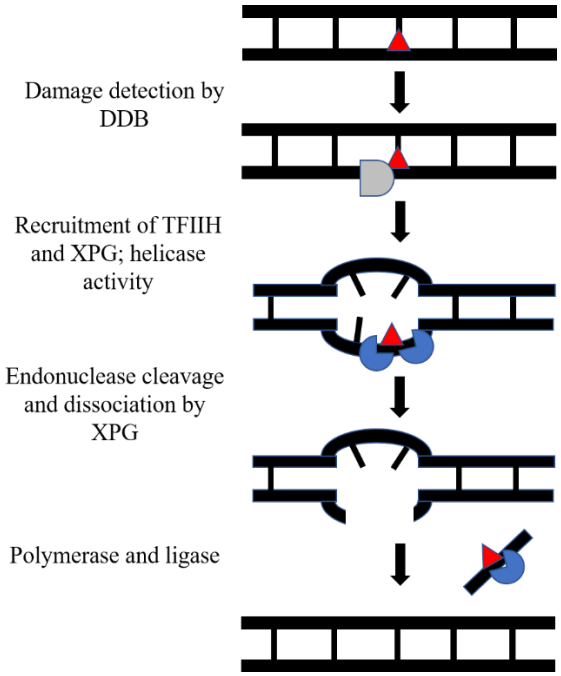


Figure 1-7. Removal of Bulky DNA Lesions through Nucleotide Excision Repair

As with all biological processes, these repair mechanisms are not full proof. Studies have shown that glycosylase activity decreases with age, correlating with an overall decline in base excision repair.[14] The lack of the ability to clear the genome of potentially mutagenic lesions results in dysfunctional cells, leading to even more increased effects of aging. The failing of these repair systems is also one of the most notable precursors to various cancers.

In addition to the natural failings of DNA repair, there are also several diseases that result in impaired repair capabilities. One such disease is xeroderma pigmentosum, a genetic disease that mutates the enzymes involved in the nucleotide excision repair pathway. This hinders and in extreme cases completely eliminates the ability to repair lesions formed from exposure to UV light, leading to symptoms such as severe sunburn, blistering and even corneal ulcerations.[15]

Another DNA damage related disease is ataxia-telangiectasia, more commonly known as Louis-Bar syndrome. This is a neurodegenerative disease caused by a mutation in the gene responsible for responding to DNA double strand breaks. A normal cell would repair breaks before replicating, but in this syndrome the repair enzymes are unable to function, allowing these double strand breaks to be propagated through cell replication. This disease has been known to cause severe disability, with large impairments to movement and coordination.[16]

Environmental Toxins: Diols and Amines

Our work focuses on exogenous DNA damage resulting from reactions with genotoxic chemicals, more specifically from reactions with environmental toxins. These toxins can be classified into diols and amines.

Diols are molecules characterized by having two hydroxyl groups in the structure. The most common diol used industrially is ethylene glycol, a small diol with only two carbons separating the hydroxyl groups (Figure 1-8A). Ethylene glycol is a viscous liquid that is both odorless and colorless. It is most commonly used in heat transfer solutions, such as those found in automobiles, liquid cooled computers and air

conditioning units. It has a specific heat capacity that is about half that of water, allowing it to both decrease the freezing point and increase the boiling point of pure water.[17] Pure ethylene glycol freezes at a temperature of about -12 °C but when it is mixed with water it freezes at about -45 °C. This makes it ideal to use as antifreeze in automobile engines, as well as a de-icing agent for windshields and aircraft. It is also used to preserve biological tissues and organs at below freezing temperatures.

Although it is quite useful, ethylene glycol has been shown to exhibit moderate toxicity upon ingestion. When digested it is oxidized to glycolic acid and in turn to oxalic acid. This compound is toxic to the central nervous system, the heart and the kidneys and can be potentially fatal if left untreated. Ethylene glycol is also sweet tasting, making it particularly dangerous for children and animals. Toxicity due to other modes of exposure such as inhalation and absorption have not been fully characterized; however, it is considered a high production volume chemical and can remain in the air for up to 10 days. This makes the potential for exposure relatively high, especially in places such as airports where ethylene glycol is frequently used.

Another diol that is commonly found in use as a heat transfer agent is propylene glycol, or 1,2-propane diol (Figure 1-8B). This compound is very structurally similar to ethylene glycol, sharing many of its physical properties as well. It also causes freezing point depression when mixed with water, making it ideal for antifreeze type solutions; however, the depression is not as significant as when using ethylene glycol, making propylene glycol a secondary compound for this purpose.

In addition to being used as a freezing agent propylene glycol is also used in polymer synthesis. It is a common precursor to many polyester resins and can also be polymerized to form plastics making it useful in the synthesis of polyurethanes. It is found in many food products, used both as a preservative as well as a sweetener in certain beverages. It is also one of the major components that is aerosolized in the cartridges of e-cigarettes and vaporizers.

In terms of safety the oral toxicity levels of propylene glycol are much lower than ethylene glycol, making it relatively safer for use in food products. Although, like ethylene glycol, the toxicity from exposure through other modes has not been fully characterized. Natural degradation of propylene glycol has been shown to require a large biochemical oxygen demand, slowing the degradation process. This means that propylene glycol can remain in soil and surface waters for up to a few weeks, increasing the potential for exposure through various modes.

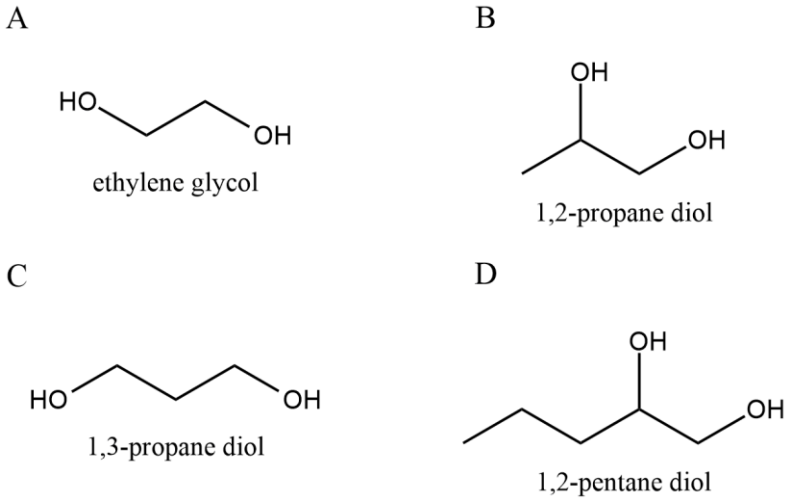


Figure 1-8. Environmental Diols

Other common diols include 1,3-propane diol (Figure 1-8C) and 1,2-pentane diol (Figure 1-8D). 1,3-propane diol is used in many industrial products and is a minor component in antifreezes, similar to the previously mentioned diols. It is also found in polyesters, adhesives, laminates, and even in wood paint. 1,2-pentane diol is used as synthetic humectant, making it optimal for use in cosmetics and beauty products. It is commonly found in lotions, skin creams and even sunblock. Despite their frequent uses, neither of these compounds have clear toxicity data.

Amines are characterized by being derived from ammonia, with various organic groups replacing one or more of the hydrogen atoms. One of the most industrially prolific families of amines are hydrazines, characterized by having two bonded NH_2 groups. Hydrazines are most commonly used as propellants and as precursors to blowing agents. For example, the main gas used in air bags in automobiles is produced from hydrazine. In addition to this, hydrazines are also main components in fuel for emergency power units in jets and even NASA space shuttles. In the space shuttles hydrazine is used in the maneuvering thrusters of the shuttle and also used in the terminal descent of the spacecraft. Hydrazines are also used as precursors to many pharmaceutical and pesticidal compounds.[18]

Hydrazine is known to be a highly toxic compound and prolonged exposure through either ingestion or inhalation could result in nausea, seizures coma and even liver and kidney damage. According to the EPA prolonged inhalation leads to an increase in lung, nasal and liver tumors in mice. Hydrazine is classified as a probable human

carcinogen. In this study phenylhydrazine, a derivative of hydrazine, was investigated (Figure 1-9A).

Another common amine is hydroxylamine (Figure 1-9B). This compound is prepared as a hydrochloride and is commonly used in the removal of hair from animal skins and also as a main component in photography developing solutions. Hydoxylamine hydrochloride is known to exhibit mild toxicity, causing skin irritation and is considered a possible mutagen if ingested. Methoxyamine (Figure 1-9C) is another derivative of this compound that is utilized in similar ways and exhibits similar toxicity.

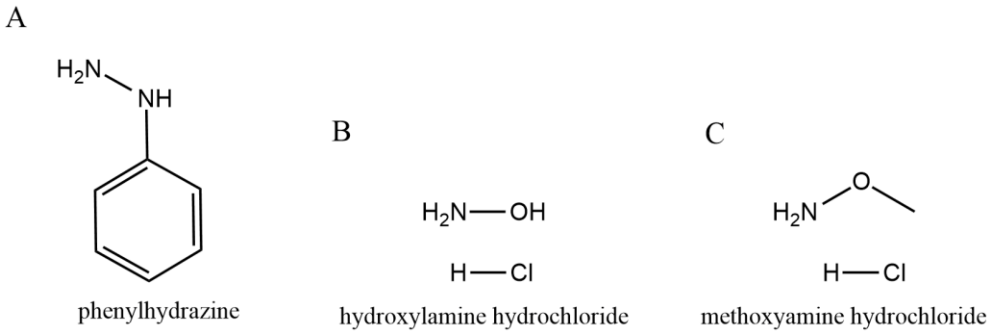


Figure 1-9. Environmental Amines

When considering diols and amines as agents of exogenous DNA damage, it can be seen that they are all nucleophilic compounds, making them potentially reactive towards electrophilic groups. This does not present much of a threat to canonical DNA, with its negatively charged phosphate backbone and nucleotides containing electron rich nitrogen and oxygen, however, natural modifications can occur to DNA, introducing new

functional groups. One such modification is the introduction of 5-formylcytosine, a nucleotide analogue involved in the epigenetic control of gene expression (Figure 1-10).

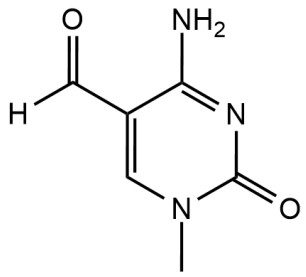


Figure 1-10. 5-Formylcytosine

Epigenetics

Epigenetics is a field of study that came about as a way to explain the phenomenon of varying cellular phenotypes. When the chromosome was first discovered, it was determined to be a mass of both nucleic acids and proteins, however, it was unclear as to which was responsible for carrying the genetic material. Despite this lack of clarity, it was clear that every cell in an organism possessed every chromosome, leading to the question how can a single cell with one set of chromosomes give rise to a complex organism containing a multitude of cellular phenotypes? Epigenetics seeks to fill in this knowledge gap and was defined early on as everything that occurs in a cell from its inception as a zygote to its completion of development into a fully differentiated organism.[19]

As technology and scientific understanding progressed, so too did the definition of epigenetics. Several important discoveries were made in the mid-twentieth century.

In 1944 the Avery-Macleod-McCarty experiment took place, proving that DNA is the material that actually caused bacterial transformations.[20] Shortly after this the Hershey-Chase experiments took place, focusing on the infection of bacterial cells by bacteriophages. They were able to show that the bacteriophage DNA enters the cell during infection, whereas the proteins remain in the bacteriophage.[21] These studies helped confirm that nucleic acids are responsible for carrying the genetic material and not proteins. Within the next few years Chargaff formulated his rules for nucleic acid base pairing; adenine with thymine and guanine with cytosine (

Figure 1-11A). This paved the way for Watson and Crick to release their discovery of the DNA double helix (

Figure 1-11B).[22]

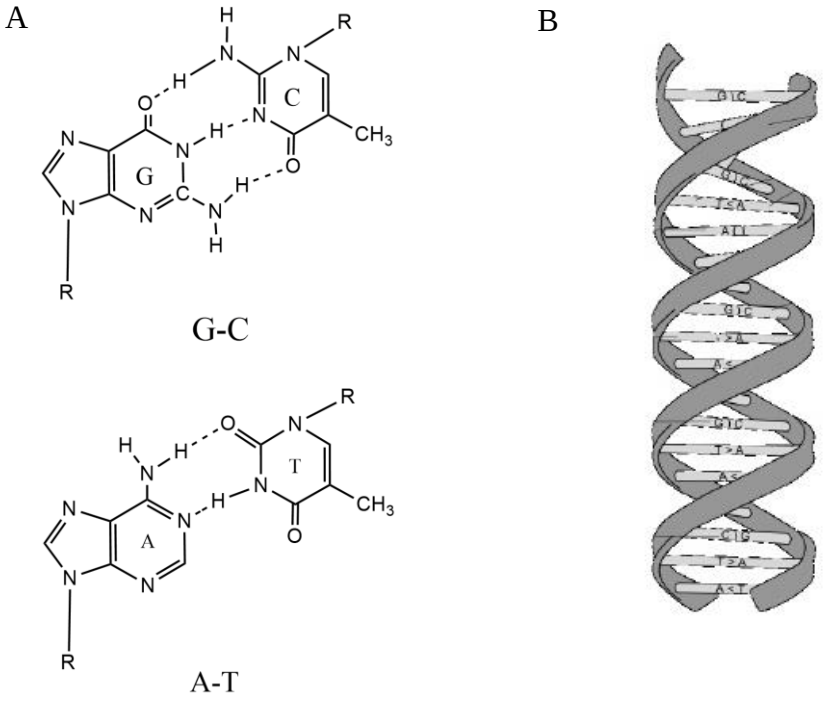


Figure 1-11. (A) Watson and Crick Hydrogen Bonding Following Chargaffs Rules of Base Pairing. (B) DNA Double Helix

With these developments the focus of epigenetics was able to be narrowed down to solely nucleic acids and a more refined definition was established: the study of heritable changes in gene function that cannot be explained by changes in DNA sequence.[19] Based on structural studies, it was determined that these changes in gene function must be due to variations in nucleotide structure and in the proteins directly complexed with DNA.

In addition to affecting the cellular phenotype, epigenetics can also have an effect on the organism as a whole. Recently, epigenetics has been thought to be affected by many external sources such as one's age, their environment and even their lifestyle. This is potentially the reason that twins can have different personalities and skill sets despite having identical genomes. Their different lifestyle choices trigger different patterns of epigenetic control, resulting in differences in overall gene expression. Epigenetics has also been thought to play a role in the development of some diseases and even in some types of cancers. With all of these potential effects and changes in gene expression, it is important to understand how exactly these changes are being implemented. There are three main modes of epigenetic control: DNA methylation, histone modification and non-coding RNA-associated gene silencing.

DNA Methylation. Of the three modes of epigenetic control, DNA methylation is thought to be the most prolific and influential of them. Out of the four canonical DNA bases only cytosine and adenine can be methylated. Cytosine methylation has been shown to occur in a wide spread of organisms, including both eukaryotes and prokaryotes although the relative rates of methylation vary greatly. Adenine methylation is also very widespread and has been seen in bacterial, plant and even mammalian DNA, though not

as frequent as the occurrence of cytosine methylation. Epigenetics focuses on the methylation of cytosine as a method of controlling gene expression.

Cytosine can be methylated at the 5' position to form 5-methylcytosine (Figure 1-12A). This methylation tends to occur in CpG sites, regions of DNA where a cytosine is directly followed by a guanine in the 5'-3' direction. These sites are self-complementary, meaning a CpG on one strand will result in the formation of a CpG on the complementary strand as well. This allows methylation to also occur in a self-complementary fashion, with both cytosines of the CpG pairs being methylated (Figure 1-12B).[23]

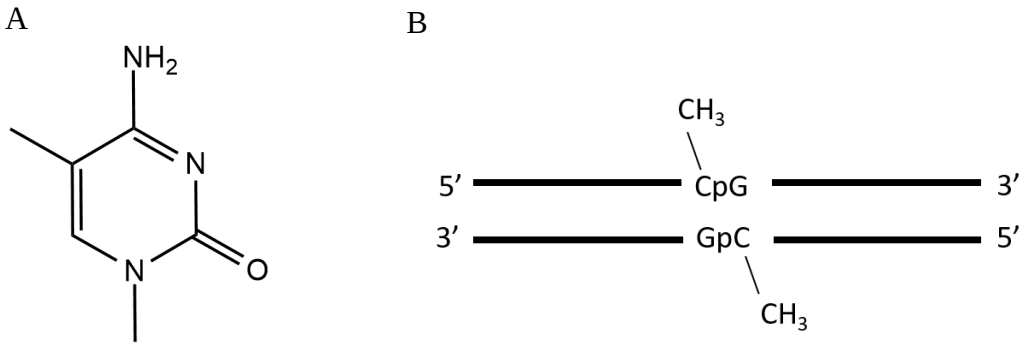


Figure 1-12. (A) 5-methylcytosine. (B) Self-complementary Methylation of CpG Dinucleotides

This pattern is crucial for passing down specific methylation patterns through cell replication. When a methylated DNA duplex is replicated, it forms a hemi-methylated duplex. The parent strand will retain the initial methylation; however, the newly synthesized strand does not contain any methylation. An enzyme called DNA

methyltransferase 1 is then able to come along and perform maintenance methylation. The enzyme can scan down the newly synthesized duplex, checking for pre-existing methylation. When it encounters a CpG that has been methylated it then copies the methylation onto the complementary side of the CpG resulting in a new fully methylated duplex (Figure 1-13).[23] This mechanism allows a specific pattern of DNA methylation to be passed down through many rounds of cell replication, thus maintaining and increasing the impact of a particular pattern of gene expression.

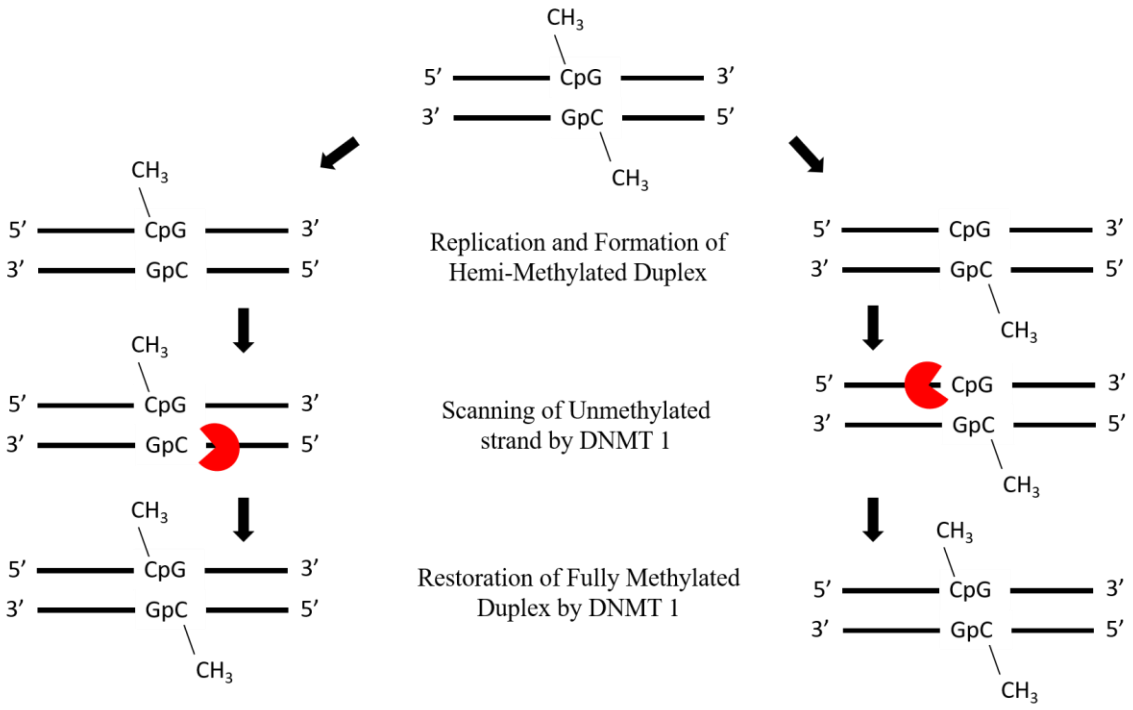


Figure 1-13. Maintenance Methylation of Hemi-Methylated DNA by DNMT 1

This raises the question of where these methylation patterns originate. When an egg is first fertilized, the zygote initially contains the methylation patterns from the parent

organisms; however, studies have shown that there is a global demethylation event immediately following fertilization. All methylation is completely wiped before the beginning of germ cell development.[24] New methylation patterns are then established through de novo methylation, a process carried out by DNA methyltransferases 3a and 3b (Figure 1-14).[25] This is needed in order for germ cells to retain totipotency. The parental gametes are already differentiated, therefore, copying their methylation pattern would not allow for a germ cell to properly undergo future differentiation. This process has been termed epigenetic reprogramming.[26]

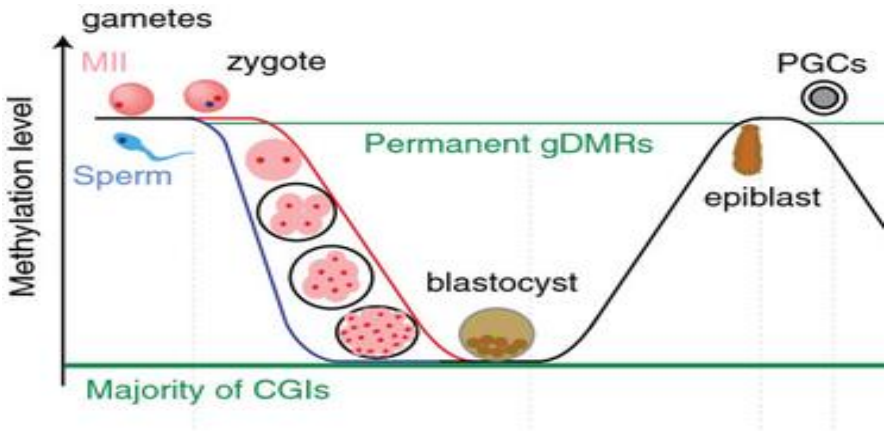


Figure 1-14. Global Demethylation upon Fertilization[27]

During this process it is estimated that 70-80% of the genomes CpG dinucleotides become methylated. Mapping has shown that some of the most methylated regions include satellite DNAs, exons of genes and repetitive elements such as transposons.[28] The methylation of these dinucleotides almost appears to occur as a default process, with all accessible CpGs becoming methylated.

The exception to this are CpG dinucleotides located within CpG islands (CGIs). These are regions of DNA at least 200bp in length that have a GC content of at least 50%. These CGIs tend to remain unmethylated after de novo methylation. Studies have shown that during de novo methylation specific binding proteins bind to the CGIs, preventing the de novo methyltransferases from gaining access to the CpGs in these regions. After de novo methylation is complete these binding proteins then dissociate, exposing the unmethylated CpGs.[29] Methylation of the CGIs then occurs in a locus specific fashion during cell differentiation. Mapping has shown that these CGIs are typically located to the 5' end of genes, in the promoter regions.

This suggests that there must be some correlation between methylation in the promoter region and expression levels of the corresponding gene. Gene knockout studies have shown that methylation of a gene results in the repression and effective inactivation of that gene.[30] This repression can occur through two potential methods. Firstly, the added methyl group protrudes into the major groove of the DNA duplex. This could create steric hindrance and repulsion with transcription factors that are recruited to the promoter region. If the transcription factors are unable to bind then transcription can not occur.[31] The other method of repression suggests that the methyl group can recruit its own repressor proteins, that can bind and sit on top of the promoter region, blocking transcription factors from binding. Both methods prevent the gene from being properly expressed, resulting in its effective inactivation.

DNA Demethylation. Both the global embryonic demethylation and the locus-specific demethylation during cell differentiation indicate that there must be some established mechanism for demethylating DNA. One method for this is known as passive

demethylation. In passive demethylation, a methylated DNA duplex is replicated, resulting in two new hemi-methylated duplexes. Normally DNA methyltransferase 1 would then perform maintenance methylation, however, in this case that does not happen. The methylation is allowed to remain at its hemi status. Upon another round of replication, the hemi-methylated duplex will generate another hemi-methylated duplex and one completely unmethylated duplex (Figure 1-15). Initially the methylation to unmethylation ratio is 1:1, but after just 8 rounds of cell division the ratio is already 1:256, diluting and effectively eliminating the effects of the methylation pattern.[32]

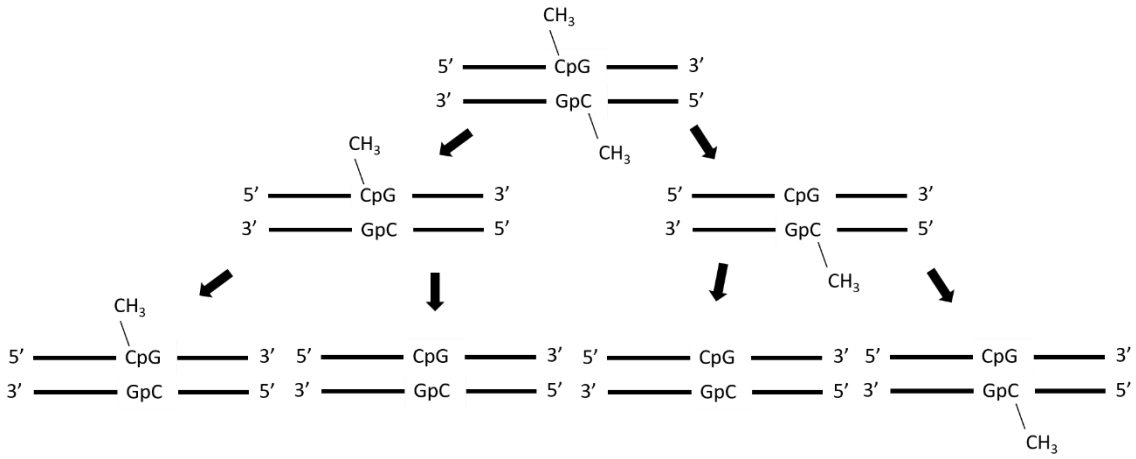


Figure 1-15. Replication Dependent Passive Demethylation

When examining embryonic global demethylation however, rates of demethylation are observed that cannot be explained by passive demethylation.[33] This mechanism is dependent on cell replication, but in the early stages of de novo methylation there is little cell division happening. To account for this discrepancy another

mechanism known as active demethylation was discovered. This active methylation is an enzymatically driven process that is not dependent on cell replication. Unfortunately, the stability of the carbon-carbon bond between the methyl group and the cytosine base is too stable to allow direct removal of the methyl group. In order to accommodate this an oxidative cycle occurs, converting 5-methylcytosine (5-mC) to 5-hydroxycytosine (5-hC) then 5-formylcytosine (5-fC) and finally 5-carboxycytosine (5-caC). 5-formylcytosine and 5-carboxycytosine can both then be removed and replaced with regular unmodified cytosine (Figure 1-16).

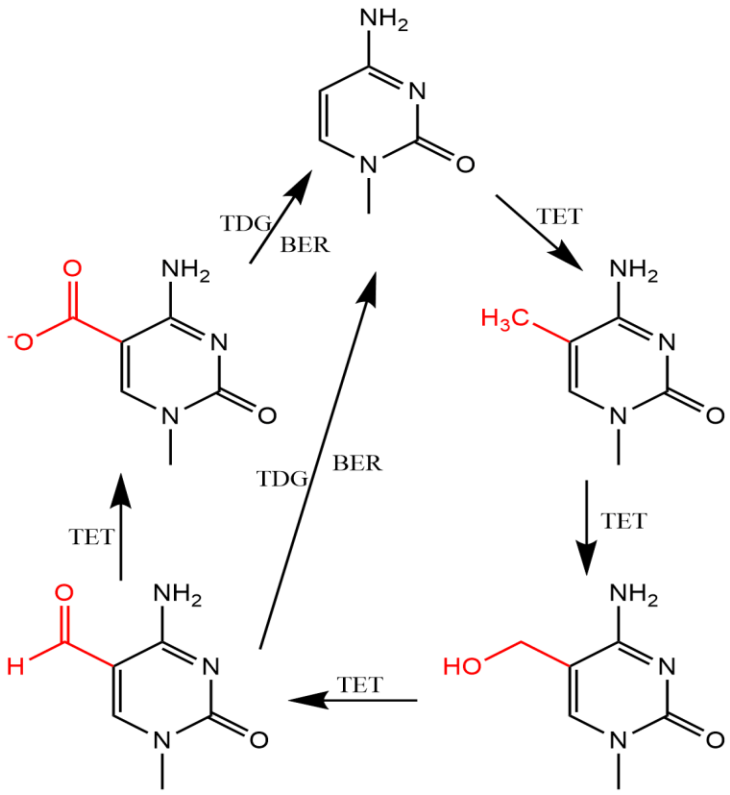


Figure 1-16. Active Demethylation through TET Oxidation and TDG Mediated Base Excision Repair

This oxidative cycle is catalyzed by a group of proteins known as TET proteins, or ten-eleven translocase proteins. These proteins are a family of iron(II) α-ketoglutarate dependent dioxygenases that target modified cytosines. Their catalytic domain consists of a double stranded beta helix (DSBH) and a cysteine rich domain. During the reaction, the DSBH brings together the iron, the α-ketoglutarate and the 5-mC in order for the oxidation to occur. The cysteine rich domain wraps around this providing additional structural stabilization.[34] These TET proteins are able to use molecular oxygen as a substrate to decarboxylate the α-ketoglutarate resulting in the formation of an iron(IV)-oxo intermediate that can then carry out the oxidation. During this process carbon dioxide and succinate are produced as byproducts.[35]

TET proteins are able to oxidize 5-mC, 5-hC or 5-fC, meaning that this catalytic domain must be flexible in order to accommodate these different substrates. Structural studies have shown that the TET-DNA contact area does not actually involve the 5 site modification, allowing the binding recognition to be maintained regardless of the modification.[36] When a TET protein binds to a modified cytosine it is capable of catalyzing either one oxidation (5mC→5hC) or completing the entire oxidation cycle (5mC→5hC→5fC→5caC). These enzymes are also not physically processive, meaning that they do not continuously scan down one DNA sequence. They will locate a modified cytosine, bind and carry out the oxidation and then dissociate away from the duplex.[37]

Although TET proteins are capable of utilizing any of the cytosine modifications as a substrate, studies have shown an increased specificity towards 5-mC over 5-hC and 5-fC.[38] This bias results in an increase in the levels of 5-hC relative to the levels of 5-fC and 5-caC. Another factor contributing to the relatively lower levels of 5-fC and 5-caC is

their ability to be removed and replaced with regular unmodified cytosine. This is done through base excision repair. As previously discussed, base excision repair is initiated by the removal of the damage site by a glycosylase. In the demethylation process this step is carried out by thymine DNA glycosylase.

Thymine DNA glycosylase (TDG) is an enzyme that initially gained attention for its role in removing thymine in T-G mismatches. It was not until later that it was discovered that unlike other glycosylases, TDG is necessary for proper embryonic development. Knockout studies showed that in the absence of TDG there was a large increase in de novo and aberrant methylation, resulting in an imbalance of chromatin towards the repressive state.[39] This indicated the significance of the role of TDG in the active demethylation process. As mentioned, the epigenetic modifications to cytosine occur at the 5 position, the same location as the methyl group in thymine. This structural similarity is thought to allow TDG to broaden its specificity to include these cytosine modifications.

Thymine DNA glycosylase actively targets both 5-formylcytosine and 5-carboycytosine during active demethylation. In both of these molecules the carbonyl oxygen is able to form hydrogen bonds with the protons from the 4-position amino group, locking in the conformation. The Tyr152 of TDG is then also able to hydrogen bond with this carbonyl oxygen, forming a stable complex. This geometry and hydrogen bonding is what separates 5-fC and 5-caC from 5-hC and 5-mC as potential TDG substrates. In addition to this crucial hydrogen bond, the Ala145 of TDG can also form a nonpolar connection with the central formyl carbon, helping to properly position the enzyme (Figure 1-17).[40]

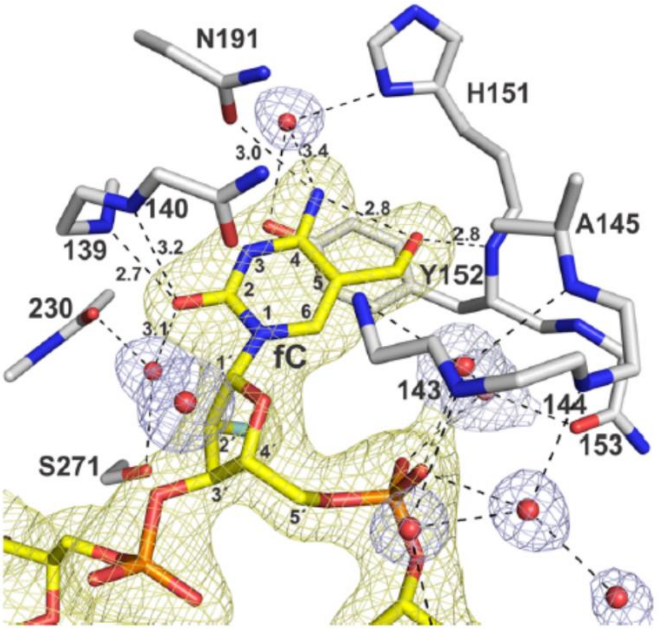


Figure 1-17. Structure of TDG bound to 5-fC[40]

Another structural feature that plays a factor in the specificity of TDG is the strength of the glycosidic bond between the base and the sugar molecule.[41] Mechanistically, base excision begins with the breaking of the glycosidic bond, resulting in an oxocarbenium ion and a pyrimidine anion. The stability of this anion can play a role in determining how likely a particular base is to be excised. In 5-fC and 5-caC the additional carbonyl provides an extension of the resonance network, helping to distribute the negative charge and making the anion more stable (Figure 1-18). The 5-hC and 5-mC bases are unable to do this, making them less favorable for excision when compared with 5-fC and 5-caC.[40]

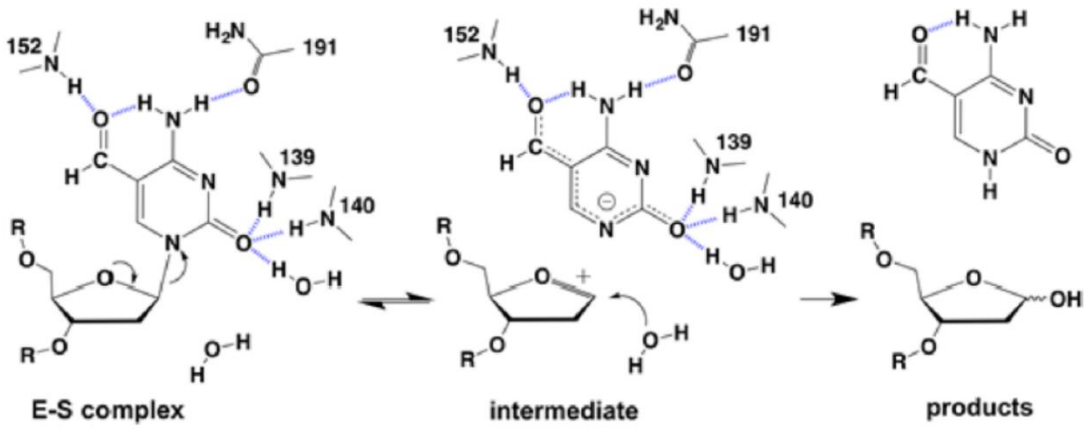


Figure 1-18. Mechanism for Cleavage of 5-fC by TDG[40]

After 5-fC and 5-caC are removed by TDG the rest of the base excision process can be carried out. Endonucleases are recruited to cut out the portion of the sequence containing the newly formed abasic site and then the gap is filled by polymerase and closed with ligase. This completes the demethylation of the strand, effectively reactivating the previously repressed gene.

5-Formylcytosine. The focus of this study was 5-formylcytosine and its potential to be reactive with environmental toxins such as the diols and amines previously described. While a large portion of 5-fC is excised by TDG, there are still significant amounts of it in the genome. Studies have shown that 5-fC is a stable modification and can remain in mammals for weeks.[42] Some of this 5-fC is leftover by TDG, due to its varying efficiencies in different regions of the genome. There are large accumulations of 5-fC in major satellite repeat DNA, regions that are mostly non-coding and tightly packed. This suggests that TDG is unable to properly associate with DNA in highly

heterochromatinized regions.[39] As a result, this 5-fC is left intact and exposed in the genome.

In addition to this, 5-fC is also thought to have its own function in epigenetic control. Structural analysis has shown that sequences containing 5-fC undergo slight conformational changes. The conformation shifts away from the traditional B-form duplex to adopt a right-handed form. This results in underwinding of the helix, moving from about 10 base pairs per turn to 13 base pairs per turn. This underwinding in turn distorts the shape of the major and minor grooves as well.[43] It is not clear what function these structural changes serve, however, it is seen that these changes are localized near the 5-fC, possibly generating some sort of binding specificity for proteins involved in the control of gene expression.

5-formylcytosine is also involved in the formation of DNA-protein crosslinks. Incubation of 5-formylcytosine and amino acid residues containing nucleophilic side chains results in the formation of an imine bond. This bond formation is reversible and pH dependent, with equilibrium lying towards the non-crosslinked form. Further investigation has shown that these crosslinks form primarily with side chains located on histones, suggesting a role in the mediation of chromatin remodeling (Figure 1-19).[44]

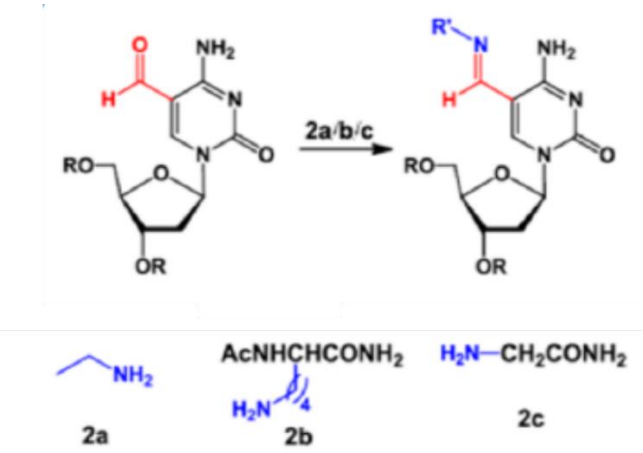


Figure 1-19. DNA-Protein Crosslinks Resulting from Incubation of 5-formylcytosine and Amino Acid Residues with Nucleophilic Side Chains[44]

Detection and mapping of 5-fC has been achieved genome wide with single base resolution by methylation assisted bisulfite-sequencing (MABS). This sequencing has shown 5-fC to be primarily found in hypomethylated promoters of highly expressed genes, and to be enriched in active transcriptional start sites and enhancers. A positive correlation was found between the level of gene expression and the level of 5-fC. In expressed genes over 50% of the cytosine modifications are those of a higher oxidation state (5-fC or 5-caC).[45] With all of this taken together, it is clear that there is a significant amount of 5-formylcytosine present in the genome at any given time.

This 5-formylcytosine is potentially reactive with a variety of molecules. As we have seen, the formyl group can react with the nucleophilic side chains of amino acids. This reactivity is also the main feature considered when designing detection methods for 5-fC. Nucleophilic primary amines, hydrazides and aminoxy derivatives have been used to form complexes with the formyl group of 5-formylcytosine.[46] While these reactions

are desirable, there is also a potential for undesirable reactions to take place. For example, if exposed to 1,3-propane diol - a nucleophilic environmental toxin – acetal formation could take place, resulting in a stable cyclic acetal lesion (Figure 1-20).

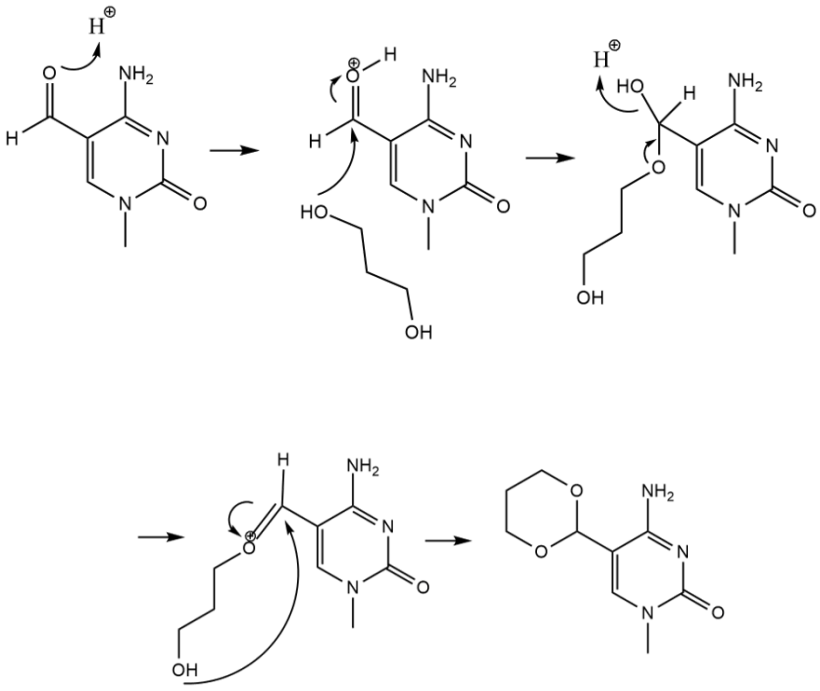


Figure 1-20. Acetal Formation Between 5-formylcytosine and 1,3-propane diol to form a Stable Cyclic Acetal DNA Lesion.

The goal of this study was to examine the potential for formation of DNA lesions formed from the reaction of 5-formylcytosine and various nucleophilic environmental toxins. Physical studies were then conducted to examine the lesions effects on DNA duplex conformation and stability, as well as enzymatic studies to determine the effects of the lesions on DNA replication.

Chapter 2: Lesion Formation and Physical Studies

Introduction

There are a wide variety of DNA lesions that can be formed upon the interaction of 5-formylcytosine and the environmental toxins discussed. While having slight variations in structure, all of these lesions are formed mechanistically through a nucleophilic addition to the carbonyl of the formyl group. Nucleophilic addition reactions are used frequently in metabolic processes and the biological synthesis of macromolecules; therefore, it is clear that this type of reaction can take place under physiological conditions. While they do readily occur, reactions of this nature can be highly specific, making it necessary to verify that they can occur between the formyl group and our environmental toxins of interest.

Lesion Formation. In order to investigate their respective reaction potentials, a model study was designed using the environmental toxins and acetaldehyde. Acetaldehyde was chosen for being the simplest form of the formyl group, allowing an isolated perspective of the addition reaction. If successful, the reactions would yield products analogous to the adducts that would be formed on 5-formylcytosine. These reactions were carried out *in vitro*. In order to simulate a physiological environment, reactions were carried out in a phosphate buffer at relevant pH values. It was noted that after incubation the compounds could easily be extracted into an organic

solvent, and as small molecules they are relatively volatile, making gas chromatography/mass spectrometry (GC/MS) an ideal method for analysis in this model study. Using this method, it was observed that addition products did form for many of the environmental toxins.

It was noted that the adduct formed from the reaction of 5-formylcytosine and 1,3-propane diol is commercially available in phosphoramidite form, omitting the need to specifically synthesize the modified nucleotide. Due to this convenience, many of the planned future studies will focus on the effects of this particular acetal-dC lesion (Figure 2-1). In order to get a better picture of its reactivity, the model study for this particular compound was conducted at a larger range of pH values, from as low as 4 up to 7. This allowed us to assess the possibility of a pH dependency on formation of the adduct.

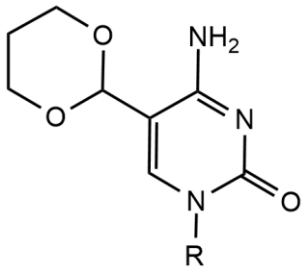


Figure 2-1 Acetal-dC Lesion Formed from the Reaction of 1,3-propane diol and 5-formylcytosine

In addition to examining this formation as an isolated reaction, it is necessary to examine the potential for these adducts to form when the formyl group is a part of 5-formylcytosine in a complete DNA sequence. After verification of the presence of the 5-

fC, the same type of incubation reaction as used in the model study could be carried out here but using a prepared 5-fC oligonucleotide rather than acetaldehyde. Due to the size of the resulting molecule, GC/MS was no longer a viable analysis method so instead ESI-MS was utilized. Using this method, it was able to be seen that the various environmental toxins do complex with the 5-fC containing oligonucleotide to form DNA adducts. Having shown that these various lesions can indeed form it was necessary to then characterize their effects on the conformation and stability of the DNA duplex. The sequences used throughout these studies are shown below in Table 2-1.

1	ACGTGCTGCACACGACGTGCTGAG
2	ACGTGCTGCAC*ACGACGTGCTGAG
3	CTCAGCACGTCGTGTGCAGCACGT
4	CCAGTCCCCCCTTTTCTTTTAAAAAGTGG
5	CCAGTCCCCCCTTTTC*TTTTAAAAAGTGG
6	CCACTTTTTAAAAGAAAAGGGGGGACTGG
7	GGCCACGAGG
8	GGCCAC*GAGG
9	CCTCGTGGCC

Table 2-1 DNA Sequences Used in Physical Studies. C* = acetal-dC Lesion

DNA Conformation. DNA lesions of various types have been known to induce structural changes on the DNA duplex. While human DNA is typically found in B form, the presence of some lesions has been shown to shift this equilibrium to favor the A form, changing the overall structure. In addition to large scale changes DNA lesions can also

cause local conformational changes. Any change in the angle of the sugar molecule could lead to changes in the backbone as well, which could in turn be propagated to neighboring bases. These conformational changes can potentially affect the stability of the duplex, or even the ability of the duplex to be involved in regular enzymatic processes.

In this study, circular dichroism (CD) was used to assess if the acetal-dC lesion caused any conformational changes to the DNA duplex. Circular dichroism is a method that involves irradiating a sample with circularly polarized light and then measuring the difference in absorptions of left and right handed light. This technique is ideal for nucleic acids due to the handedness of the DNA duplex. B-form DNA tends to give a large positive signal in the 260-280 nm range, while giving a negative signal around 245 nm. A-form DNA on the other hand, gives a positive signal around 260 nm and a negative signal near 210 nm.[47] Changes in the signal pattern, or even loss of signal could be indicative of changes to the overall conformation of the DNA duplex.

Duplex Stability. Any changes in the conformation of the DNA duplex could also result in changes of the stability of the duplex. In this study, UV thermal denaturation was used to assess any changes on the stability. DNA is held together primarily by hydrogen bonds and pi pi stacking between neighboring nucleotides. The "melting temperature" of a DNA duplex describes the transition of a sequence from the duplex form to single strands and is an indication of the strength of these interactions. The official melting point of a DNA sequence is the moment when half of the DNA has denatured into single strands. Longer sequences have a larger amount of interactions, thus requiring a larger amount of energy to separate, giving them a higher melting

temperature. Also, G:C bonds have three hydrogen bonds, while A:T bonds only have two, giving sequences higher in GC content a higher melting point. If the acetal-dC lesion affects the hydrogen bonds or stacking capabilities of the base, this could be represented by a decrease in the melting temperature.

In addition to melting studies done at one particular concentration, a set of experiments can be run varying the concentration. As the concentration of DNA fluctuates, the melting temperature fluctuates as well. This relationship between concentration and melting temperature allows us to use a concentration depending melting experiment to extrapolate the thermodynamic energy constants of the DNA duplex using the Van't Hoff Equation. This gives a clearer picture of the energetic effects of the acetal-dC lesion.

Computational Studies. While the previous studies showed that the lesion induced slight changes in the conformation and stability of the duplex, the source of these changes remained unclear. In order to get a closer look, a molecular dynamics simulation was run that allowed us to view any structural changes caused by the lesion. This study allowed us to compare hydrogen bond lengths and angle torsion of the base pairs between an unmodified and modified duplex. In addition, this study provided us additional information on the overall change in energy, reinforcing the results from the concentration dependent melting studies.

Experimental

Lesion Formation - Model Study. Incubation reactions were prepared to a total volume of 400 μL containing 500 mM acetaldehyde and 500 mM of nucleophile (diol or

amine). The reactions were buffered with a 10 mM sodium phosphate buffer with varying pH values. Buffer pH was previously adjusted from 7 using concentrated hydrochloric acid. After mixing, solutions were incubated at 37 °C for 18 hours. Reaction products were then extracted using 400 μL of diethyl ether and analyzed by an Agilent 7890A GC System with 5975C Inert XL MSD. MassHunter software was used for data analysis.

DNA Synthesis. Oligonucleotides were synthesized on an Applied Biosciences 392 DNA synthesizer using standard phosphoramidite protocols. Modified bases were also incorporated in their phosphoramidite form. All the chemicals for DNA synthesis were purchased from Glen Research (Sterling, VA). The resulting oligonucleotides were purified through gel electrophoresis and column chromatography. The concentrations of oligonucleotide solutions were determined by measuring their absorbance at 260 nm using a Varian Cary 100 Bio UV-Vis spectrophotometer (Walnut Creek, CA). Absorbance was converted into concentration using the molar extinction coefficients obtained from Oligo Analyzer 3.1 (www.idtdna.com).

Mass Spec Study. To begin, 20 nanomole of acetal-dC containing oligonucleotide was dried down to form a pellet. The pellet was then mixed with 500 μL of 80% acetic acid and left for 6 hours. Acetic acid was then removed leaving a pellet of deprotected 5-fC containing oligonucleotide. Nucleophile and buffer were added directly to the pellet to a total volume of 400 μL with 300 μM nucleophile and 10 mM sodium phosphate buffer. Phosphate buffer was previously adjusted to pH 5 using hydrochloric acid. Reaction mixtures were then left to incubate at 37 °C for 18 hours. After incubation the remaining buffer and solvent was removed to leave a pellet. Cold ethanol

precipitation was performed by dissolving the pellet in water, 5 M ammonium acetate and cold ethanol. The desalted precipitated oligonucleotide was then collected and the supernatant discarded. The remaining pellet was then dissolved in 100 μL of water to give a final concentration of roughly 100 μM. Samples were then analyzed with an Agilent ESI-MS.

Circular Dichroism. Reactions were prepared to a total volume of 2 mL containing 2 μM of either a modified or unmodified sequence and 2 μM of its complementary sequence. Solutions were also mixed with KF (50 mM) and buffer containing EDTA (1 mM) and sodium cacodylate (10 mM). Reaction mixtures were then heated to 90° C for 5 minutes and allowed to slowly cool for 2 hours. Solutions were then loaded into a quartz cuvette and analyzed with a Jasco J-810 Spectropolarimeter.

UV Thermal Denaturation. Reactions were prepared to a total volume of 1 mL containing 1 μM of either a modified or unmodified sequence and 1 μM of its complementary sequence. In the concentration dependent studies, reactions were prepared with either 5, 4, 2, 1 or 0.5 μM of each oligonucleotide. Solutions were also mixed with KCl (100 mM) and buffer containing EDTA (1 mM) and sodium cacodylate (10 mM). Reaction mixtures were then heated to 90° C for 5 minutes and allowed to slowly cool for 2 hours. Solutions were then loaded into a quartz cuvette and analyzed with a Varian Cary 100-Bio UV-Vis Spectrophotometer. UV absorbance at 260 nm was recorded while the temperature ramped at 0.3° C/min from 25 to 80° C. Melting temperatures were determined by Thermal software using the first derivative method.

Computational Studies. DNA double helices were modeled using a two level ONIOM method; combined quantum mechanical (QM) methods and molecular mechanics (MM) methods (ONION(QM:MM)) implemented in Gaussian ‘09 program package. The initial geometry of the 10mer duplexes were built using Chimera. One nucleotide on either side of the acetal-dC lesion and their hydrogen bonding partners were included in the QM level calculations. The remaining residues were treated using the low MM level calculations. Geometry optimization was carried out using hybrid B3LYP density functional with 6-31G(d) basis set for the QM level. A universal force field (UFF) was applied to the MM level, freezing them in place. To obtain the reaction energy of the nucleophilic addition reaction, cytosine and the acetal-dC lesion were treated at the B3LYP/6-31G(d) level of theory. The nucleophilic addition reaction energy was determined using the following equation:

Unmodified duplex + acetal-dC ⟶ Modified duplex + cytosine

In the case of the tandem lesion duplex, both acetal-dC lesions and one nucleotide on either side, along with their hydrogen bonding partners were included in the QM level calculations. The remaining six residues were treated using the low MM level calculations. The energy of the nucleophilic addition reaction was determined using the following equation:

Unmodified duplex + 2x(acetal-dC) ⟶ Modified duplex + 2x(cytosine)

Results and Discussion

Lesion Formation - Model Study. For the initial portion of this model study two of the nucleophilic compounds were selected, one diol and one amine. Ethylene glycol was used as the simplest of the diols. While the carbon chain lengths of the other diols vary, the main reaction mechanism is the same for all of them. Phenylhydrazine was selected for the amines. This serves as the representative compound for the amines and is much bulkier than ethylene glycol, allowing us to assess any variations in reactivity due to nucleophile size. Both reactions were carried out using phosphate buffer at a pH of 6. Chromatograms of each reaction can be seen in Figure A-1 and Figure A-2 in the appendix.

When analyzing the chromatogram of the reaction of ethylene glycol and acetaldehyde, a small peak was seen representing a compound with the mass spectrum shown below in Figure 2-2A. This compound was the expected product, formed through the acetal formation mechanism shown in Figure 2-2B. This indicated that this lesion formation mechanism was feasible in physiological conditions and also showed a potential for lesion formation upon exposure to any of the diol type compounds.

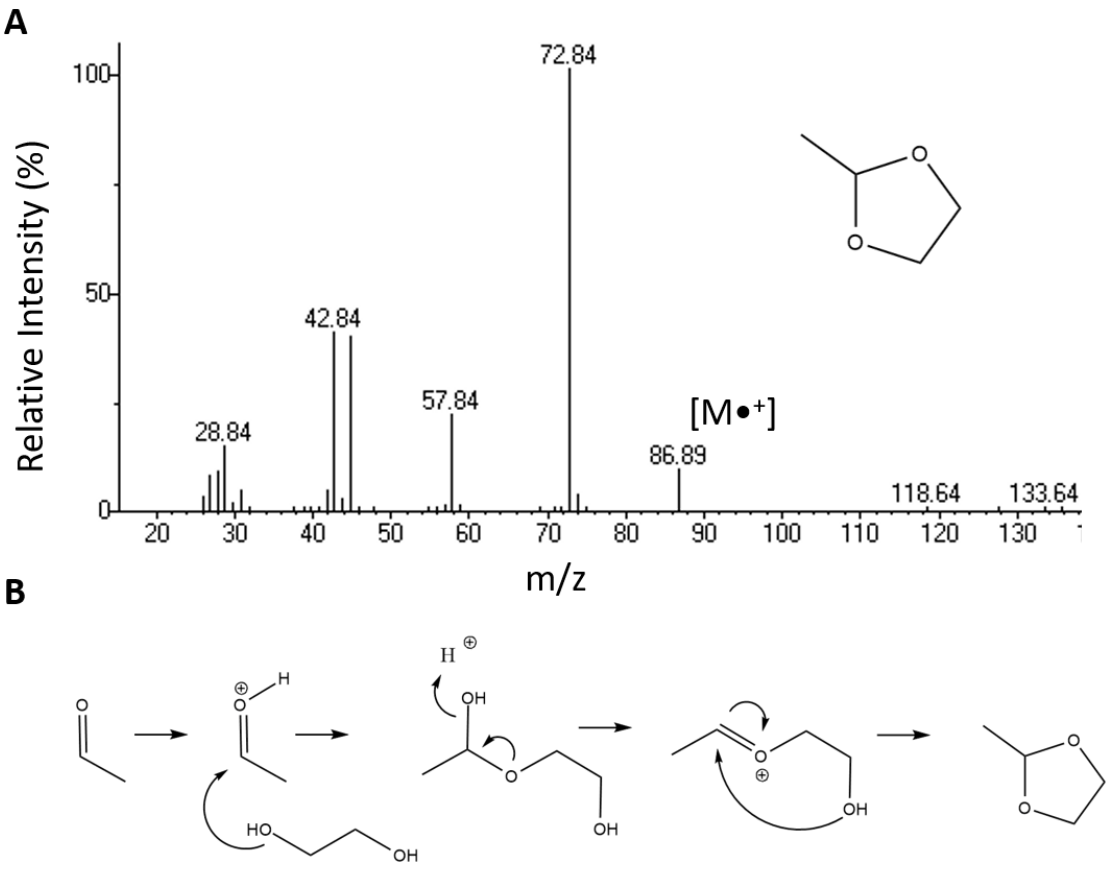


Figure 2-2 A. EI-MS of Acetal Formation Product. B. Mechanism of Acetal Formation Between Ethylene Glycol and Acetaldehyde

When analyzing the chromatogram of the reaction of phenylhydrazine and acetaldehyde, a large peak was seen representing a compound with the mass spectrum shown below in Figure 2-3A. This compound was the expected product, formed through the nucleophilic addition mechanism shown in Figure 2-3B. This indicated that this type of nucleophilic addition mechanism is feasible for the amines as well as the diols. In addition, the bulky size of the nucleophile did not hinder product formation, indicating that a wide range of compounds can possibly form lesions.

Figure 2-3 A. EI-MS of Nucleophilic Addition Product. B. Mechanism of Nucleophilic Addition Between Phenylhydrazine and Acetaldehyde

In addition to these two compounds, 1,3-propane diol was also analyzed. As mentioned, this compound was of particular interest due to its commercial availability, making it ideal to use in future studies. In order to fully understand the reaction dynamics of this compound, a series of experiments were conducted with varying pH, ranging from 4 to 7 in increments of 0.5. It was seen that at the lower pHs (4-5.5) there was a very strong peak representing the acetal formation product. As the pH increased

this product peak began to decrease, while another peak began to increase in intensity (Figure 2-4). This second peak gave a mass spectrum that identified it as unreacted 1,3-propane diol.

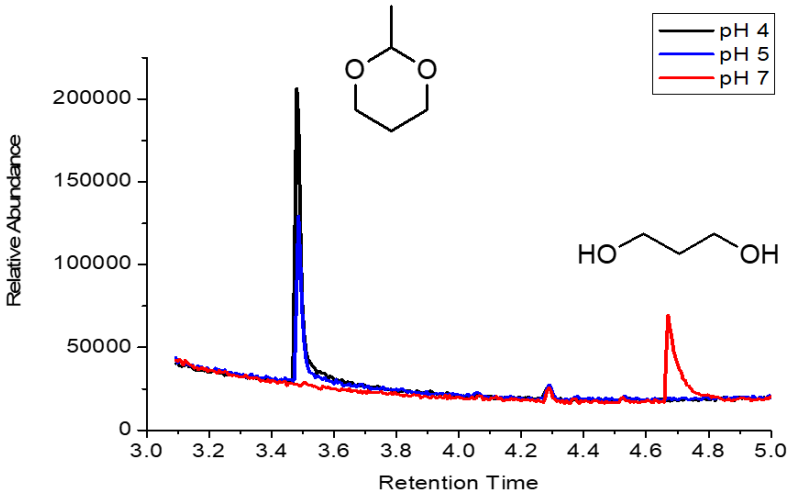


Figure 2-4 Chromatograms of the Reaction of 1,3-propane diol and Acetaldehyde at pH 4, 5 and 7

This trend indicates a pH dependency in the product formation (Figure 2-5). As seen in the proposed mechanism, the first step is an initial protonation of the carbonyl oxygen, creating a positive charge on the carbonyl carbon that the nucleophile can then attack. As an acid catalyzed mechanism, it follows that the equilibrium lies towards the product in more acidic environments. Chromatograms for pHs not shown can be seen in Figure A-3 in the appendix.

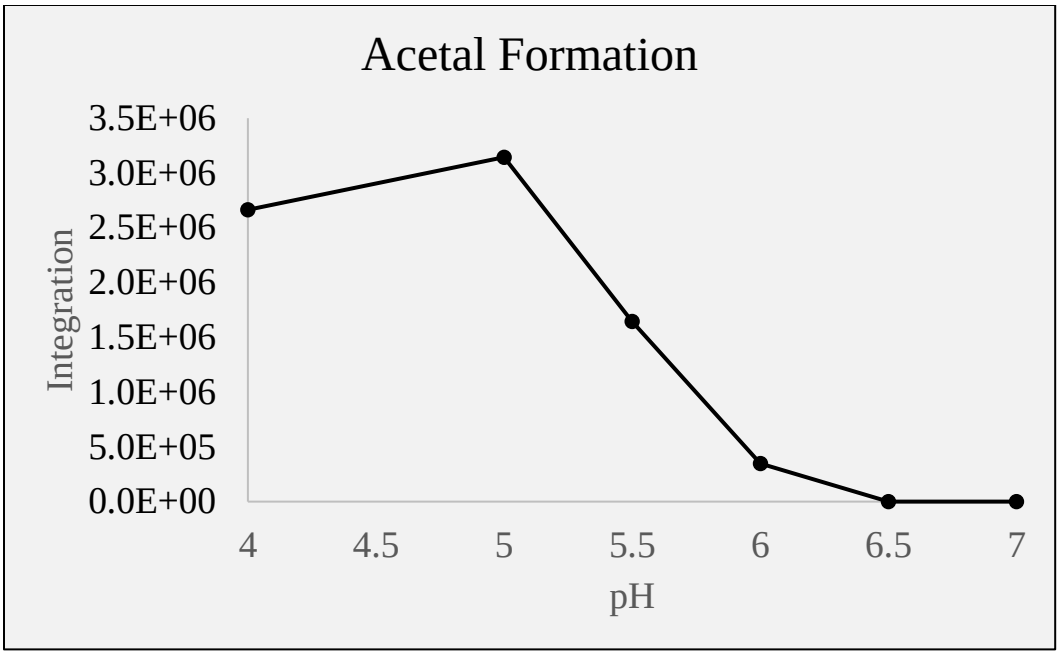


Figure 2-5 pH Dependency of Acetal Formation

While the human body does tend to be slightly alkaline, there are instances of acidic pHs. Stress, toxins and some immune reactions can all cause the body to be slightly more acidic. In addition, the digestive tract tends to be highly acidic, reaching pHs as low as 3 and 4. Highly metabolic cells also produce a large amount of hydrogen ions, making their local environments potentially acidic. Overall, the formation of this type of lesion can be considered highly physiologically relevant and a possibility warranting further investigation.

Lesion Formation – MS Study. Having verified that the most basic form of these reactions can take place, it was then necessary to verify that they happen when the formyl group is embedded in a DNA sequence as well. In order to do this, an experiment was designed taking advantage of the acetal-dC adduct. When incorporated into a

sequence, the acetal-dC adduct can easily be removed through an acidic deprotection, leaving the formyl group behind. This formyl containing sequence can then be incubated with various environmental toxins and analyzed to examine adduct formation. In this study, all of the discussed nucleophilic compounds were used, with the exception of 1,3-propane diol. From its use as a protecting group, it was clear that 1,3-propane diol is able to react with the formyl group, forming a stable lesion. As a result, it was not included in this study. In order to keep the reaction system relatively simple, a short 10mer DNA sequence (#8) was used.

The first reaction conducted was done using ethylene glycol as the environmental toxin of choice. When incubated with 5-formylcytosine ethylene glycol is expected to react through the same acetal formation mechanism as seen in the model study. A 10mer DNA sequence containing this lesion should have an exact mass of 3148.5 Da. In this study, samples were analyzed through ESI-MS in negative mode, and the majority of ions detected had a charge of -3. Following this, our expected peak can be calculated to be:

$$\frac{[3{,}148.5 - 3H]}{3} = 1{,}048.5$$

This peak was present in the mass spectrum and can be seen below in Figure 2-6.

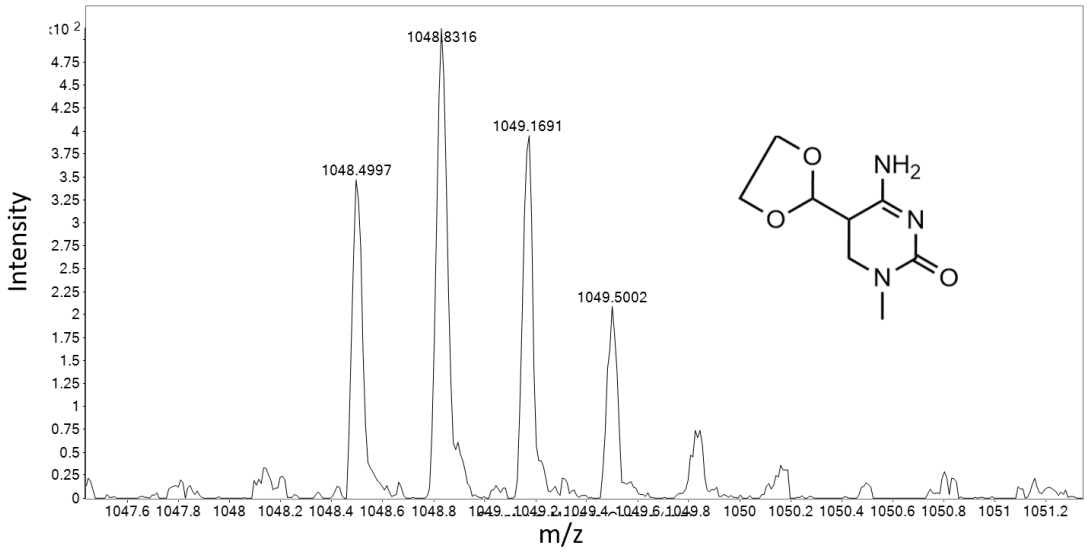
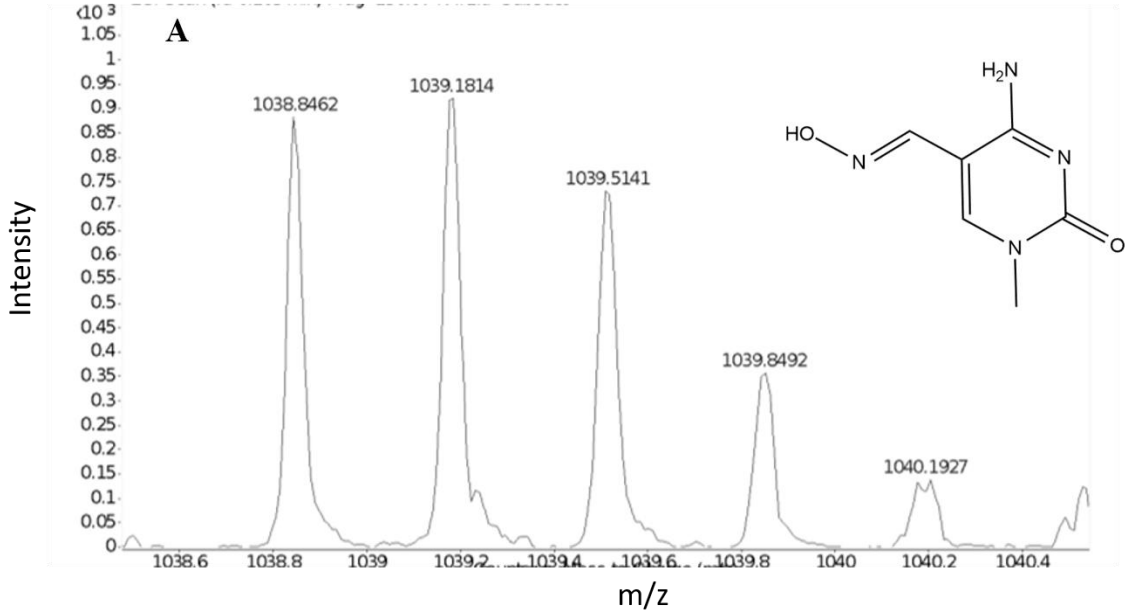
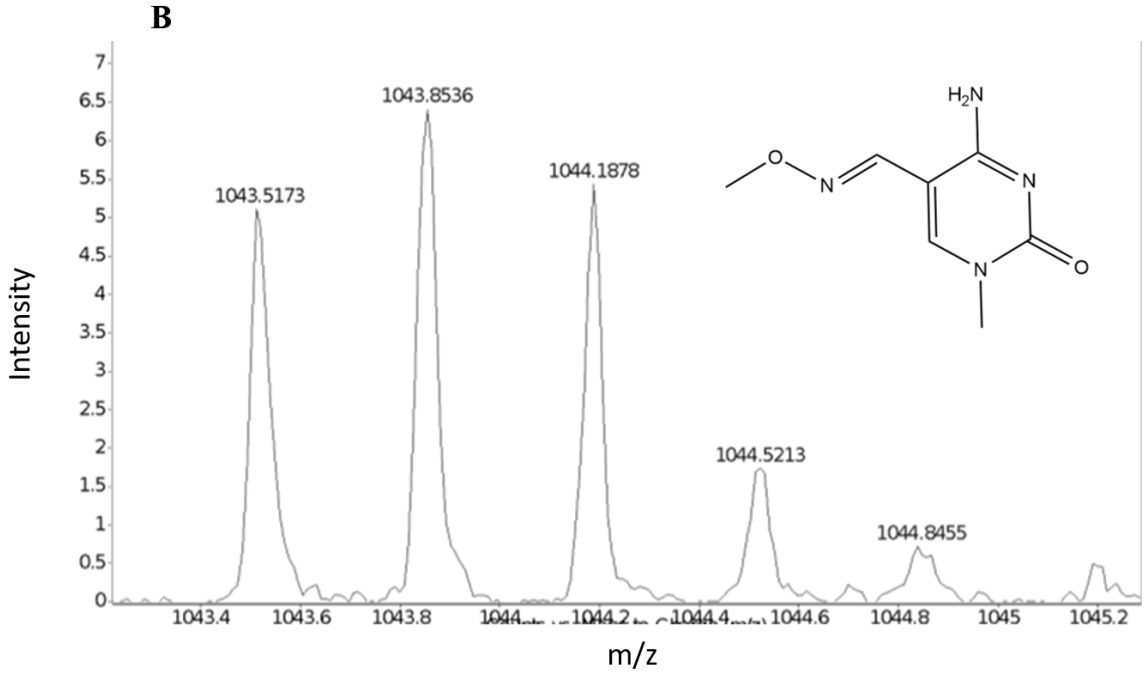


Figure 2-6 ESI-MS of the Adduct Formed from the Reaction of Ethylene Glycol and 5-formylcytosine

The presence of this species in the spectrum indicated that the acetal formation mechanism is still able to occur even when the formyl group is incorporated into a DNA sequence. In addition to the product from ethylene glycol, other experiments were able to give mass spectra showing adducts from reactions with hydroxylamine hydrochloride (Figure 2-7A), methoxyamine hydrochloride (Figure 2-7B) and phenylhydrazine (Figure 2-7C).


A
Intensity
m/z
1038.8462
1039.1814
1039.5141
1039.8492
1040.1927
HO
N
H_2N
N
N
O


B
Intensity
m/z
1043.5173
1043.8536
1044.1878
1044.5213
1044.8455
O
N
H_2N
N
N
O

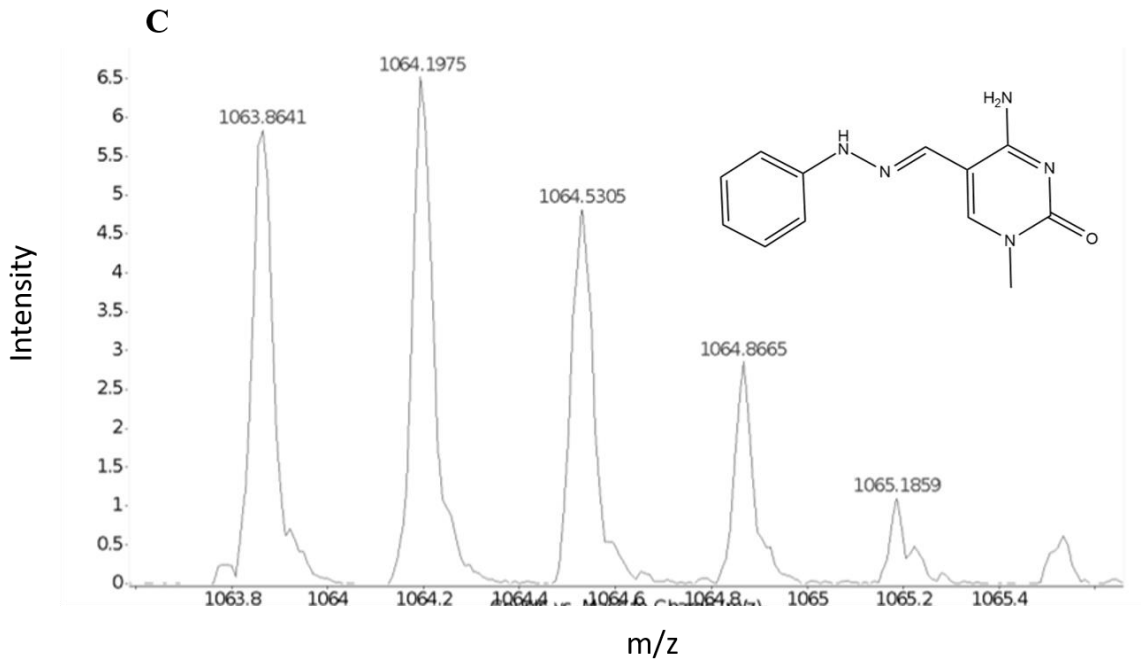


Figure 2-7 Mass Spectra of DNA Lesions Formed from the Reaction of 5-formylcytosine and A. Hydroxylamine hydrochloride B. Methoxyamine Hydrochloride and C. Phenylhydrazine

This showed that in addition to the acid catalyzed acetal formation, nucleophilic addition with amines can happen when the formyl group is incorporated into a DNA sequence as well. While the duplex can be considered to be a bulky molecule the formyl group is relatively exposed in the major groove. As discussed in chapter 1, the formyl group must be available to complex with proteins and other biological molecules, making it also available to attack from the nucleophiles described. This result serves to further increase the relevance of these types of lesions. Having confirmed the potential for these lesions to form, it was then necessary to examine what types of effects they could have on the physical characteristics of the DNA duplex. All of the remaining studies will focus on the acetal-dC lesion formed from the reaction with 1,3-propane diol.

DNA Conformation – Circular Dichroism. As previously mentioned, the presence of the acetal-dC lesion can possibly affect the conformation of the DNA duplex. In order to examine this effect two experiments were run with varying duplex lengths. The first experiment utilized sequences (#1/#2 and #3) to make a 24 nucleotide long duplex, while the second experiment utilized sequences (#4/#5 and #6) to make a 29 nucleotide long duplex. 'This allowed us to see whether the lesion had different levels of conformational impact based on the length of the duplex. Each experiment compared an unmodified and modified duplex. The results for the 24mer and 29mer experiments are seen below in Figure 2-8A and Figure 2-8B respectively.

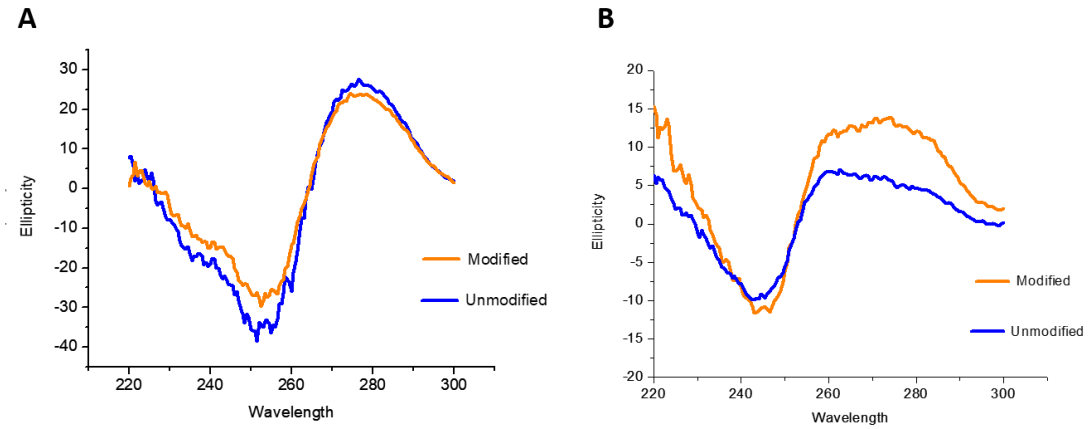


Figure 2-8 CD Spectra of an Unmodified and Modified A. 24mer and B. 29mer DNA Duplex

When examining the results of the 24mer, it was seen that the modified and unmodified curves almost exactly match each other. We saw the characteristic positive at 280 nm and negative at 245 nm, indicative of B form DNA. In addition, these positives

and negatives were maintained when comparing the modified sequence to the unmodified. This indicated that the acetal-dC lesion did not have a significant effect on the conformation of the 24mer duplex. For the 29mer, on the other hand, there was a difference between the modified and unmodified conformations. While the negative portion of the spectra was maintained, there was a discrepancy in the positive portion of the spectra. While there is a change, the overall trend is still maintained. If the acetal-dC lesion completely disrupted the duplex, the negative and positive regions would go flat. This means that while the acetal-dC lesion may have some conformational effect, it can be considered to be a minor overall change. One of the limits of CD is that it is a very qualitative technique. While we did see that there was a change of some sort, it does not provide information regarding what type of change this may be. Further studies need to be done in order to gain a clear picture of what is occurring.

DNA Stability – UV Thermal Denaturation. To gain a more quantitative measure of the effects of the acetal-dC lesion, UV melting experiments can be conducted. The first experiment conducted utilized the same 24mer duplex as used in the CD studies. The resulting melting curves from this experiment can be seen below. The absorbance of UV light by DNA increases as the duplex makes the transition to the single stranded form. This phenomenon gives rise to the S shaped melting curve. The steepest point of absorbance increase, as determined by the first derivative method, gives the melting temperature of the duplex. The results of the melting experiment done with the 24mer duplex can be seen below in Figure 2-9.

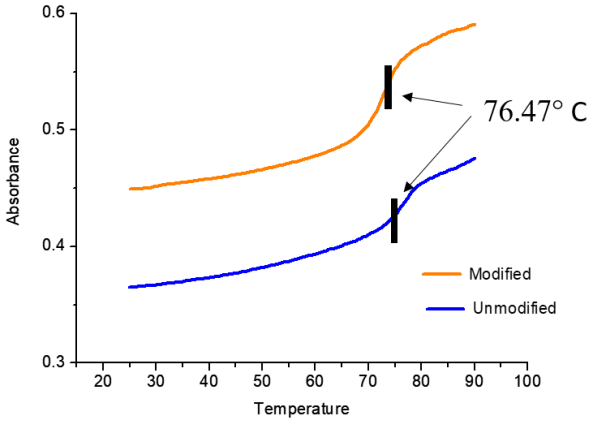


Figure 2-9 UV Thermal Denaturation Curves of an Unmodified (Bottom) and Modified (Top) 24mer DNA Duplex

In the case of the modified and unmodified 24mer duplexes it was seen that they had the same average melting temperature. This is indicative of no change in stability of the duplex as a result of the acetal-dC lesion. This result correlates with the results of the CD experiment conducted previously. As there was no change in the overall conformation, it follows that there is also no change in the stability. One potential reason for this is the structure and orientation of the acetal-dC lesion. The ring adduct projects from the 5 position, into the major groove of the duplex (Figure 2-10). In this position it does not affect the hydrogen bonding between the base pair and also does not have a large effect on the pi-pi stacking between neighboring base pairs.

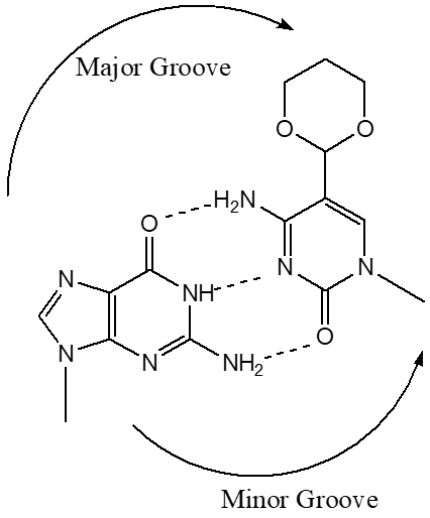


Figure 2-10 Structure and Orientation of the 5-acetaldC Lesion Base Paired with Guanine

While it was seen that the 24mer had no change, the 29mer did show a slight conformational change, meaning that there could be a slight change in stability as well. In order to investigate this the same experiment was conducted using the 29mer duplex. In this case the melting temperature did slightly decrease with the introduction of the lesion, however, this is not enough of a decrease to be considered a significant change. It is possible that the presence of one small lesion does not provide enough destabilization to affect a sequence of this length, but that in a shorter sequence the destabilization may be harder to accommodate, resulting in more significant changes in stability. In order to test this, the same melting experiment was run utilizing a 10mer duplex. The average melting temperatures for the modified and unmodified duplexes are also shown below in Table 2-2.

	Unmod	Mod	ΔT
10mer	54.65 ± 0.58	54.01 ± 0.99	-0.64
24mer	76.47 ± 0.71	76.47 ± 0.74	0
29mer	68.04 ± 0.11	67.05 ± 0.03	-0.99

Table 2-2 Average Melting Temperatures of Modified and Unmodified DNA Duplexes

Despite this theory, it was seen that there was also no significant destabilization effect in the 10mer duplex as well. This could be potentially for the same reasons as described previously. The structure and orientation of the base pair does not affect the interactions that provide the stability of the DNA duplex, thus resulting in no change in the overall melting temperature. Melting curves for the 29mer and 10mer that were not shown can be seen in the appendix (Figure A-4 and Figure A-5).

In addition to providing information regarding the melting temperature of a duplex at a particular concentration, UV thermal denaturation experiments can also be used to extrapolate the thermodynamic constants of a particular DNA duplex. This takes advantage of the concept that duplex melting is an equilibrium between single stranded and double stranded DNA that can be represented as

$$[A_{SS}] + [B_{SS}] \leftrightarrow [C_{DS}]$$

In a melting experiment the concentrations of duplex DNA and single stranded DNA can be defined as shown below, where f is the fraction of dissociation of the duplex.

$$[C_{DS}] = f\frac{C_0}{2} \qquad [SS] = (1-f)\frac{C_0}{2}$$

Given this information we are able to describe the equilibrium constant of dissociation as

$$K_{eq} = \frac{[C_{DS}]}{[SS]^2} = \frac{f\frac{C_0}{2}}{(1-f)^2(\frac{C_0}{2})^2} = \frac{2f}{(1-f)^2 C_0}$$

At the melting temperature exactly half the DNA has dissociated from the duplex into the single stranded form. By substituting f for ½ the equilibrium constant can be simplified to

$$K_{eq} = \frac{4}{C_0}$$

In thermodynamics it is known that

$$\Delta G = -RTlnK_{eq} \quad \text{and} \quad \Delta G = \Delta H - T\Delta S$$

These equations can be combined to give:

$$-lnK_{eq} = \frac{\Delta H}{RT} - \frac{\Delta S}{R}$$

This equation can then be rearranged to solve for $1/T_m$ and K_{eq} can be substituted with our previous solution to give a final equation of

$$\frac{1}{T_m} = -\frac{R}{\Delta H}\left(ln\frac{C_0}{4}\right) + \frac{\Delta S}{\Delta H}$$

By conducting a concentration dependent melting experiment, we are able obtain a range of melting temperatures corresponding to a range of DNA concentrations. This

information can be plotted in a Van't Hoff plot and the data analyzed to obtain the thermodynamic parameters for that duplex. In this study we conducted this experiment with the 10mer duplex at 5 different concentrations: 5, 4, 2, 1 and 0.5 μM. The resulting Van't Hoff plot is shown below in Figure 2-11. Each data point shown is the average of two replicate experiments. Melting curves and data for these experiments can be seen in the appendix in Figure A-6 and Table A-1, respectively.

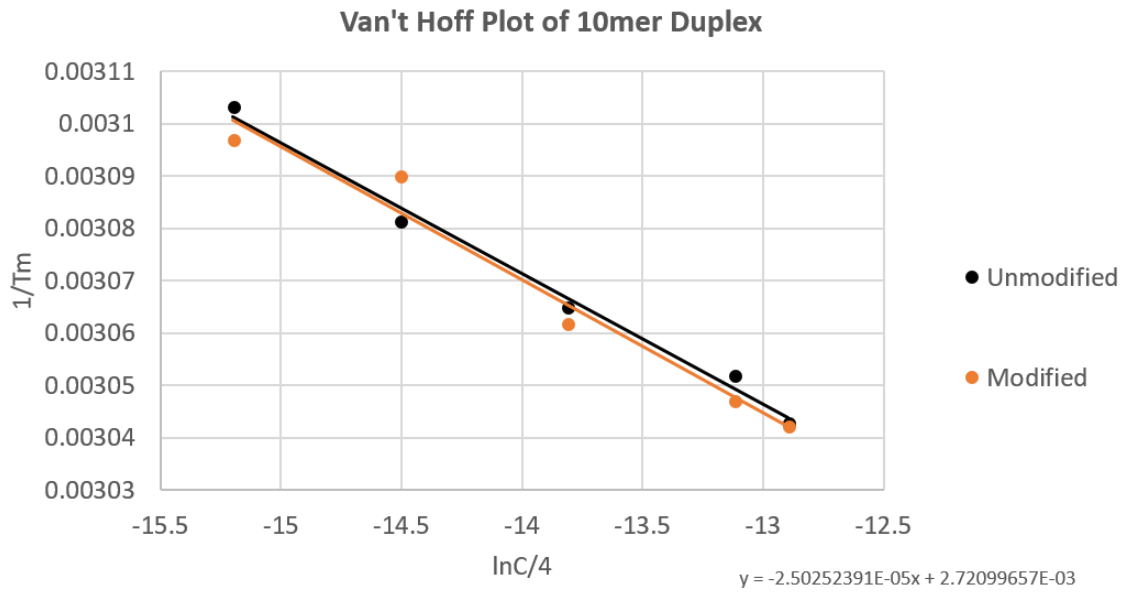


Figure 2-11 Van't Hoff Plot of a Modified and Unmodified 10mer DNA Duplex

As seen when looking at the graph, the trendlines for the modified and unmodified duplexes are nearly identical. Their parallel nature indicates the consistent correlation between concentration and the resulting melting temperature of the duplex. A more stable duplex gives higher melting temperatures, which is inversely represented in

the y axis, meaning lower trend lines represent more stable sequences. The overall trendlines shown have no significant difference, again showing that the acetal-dC lesion has no significant destabilization effect on the DNA duplex. If the examining the curves point by point it can be seen that at some concentrations the modified sequence gave a lower Tm (higher data point) while at other concentrations it gave a higher Tm (lower data point). These differences cancel each other out, resulting in the relatively similar trend lines. This similarity can also be seen in the extrapolated thermodynamic data shown below in Table 2-3. While similar, the data does show a slight destabilization in the modified sequence, represented by a lower ΔG value. This means that while there is some destabilization effect, it is not enough to manifest in a significant change in melting temperature or conformation.

	ΔH (kcal/mol)	ΔS (cal/K*mol)	ΔG (kcal/mol)
dC:dG	79.4	218.8	14.2
C*:dG	73.2	201.7	13.1

Table 2-3 Thermodynamic Constants Describing a Modified and Unmodified 10mer DNA Duplex

Overall, this concentration dependent study helps to confirm the data found in the preliminary melting experiments as well as the data from the CD experiments. As before, it is thought the structure and orientation of the acetal-dC lesion is responsible for this

lack of significant change in either conformation or stability of the DNA duplex. In order to help visualize exactly what is occurring, a computational study was conducted.

Computational Studies. In order to get a clearer picture of what is actually going on in the duplex, a molecular mechanics simulation was run with and without the lesion. The first thing we wanted to verify was our assumption that the adduct ring of the acetal-dC lesion projects into the major groove of the DNA duplex. In order to do this, unmodified and modified 10mer duplexes were constructed and geometrically optimized. The results of this simulation are shown below in Figure 2-12. The acetal-dC lesion site is shown in green.

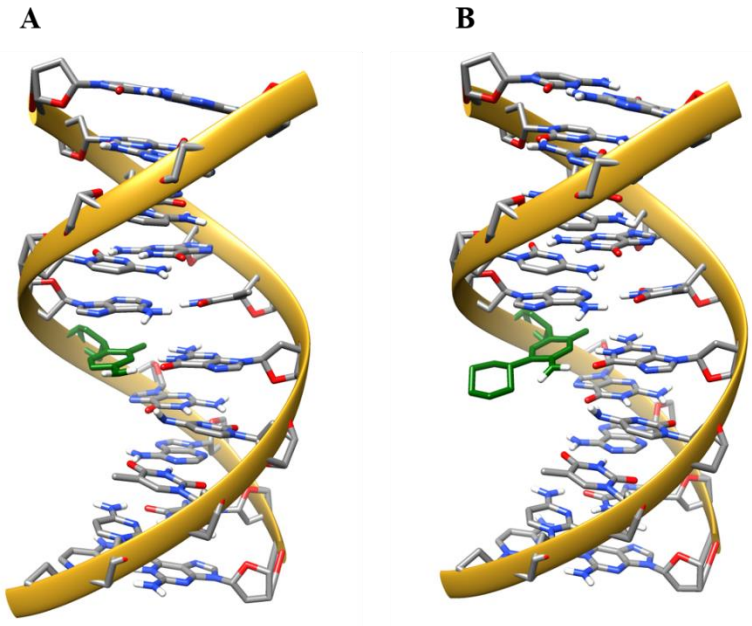


Figure 2-12 Molecular Mechanics Simulation of an Unmodified (A) and Modified (B) 10mer DNA Duplex. Acetal-dC Lesion Site is Shown in Green.

It can be seen that the adduct ring does extend into the major groove of the duplex and does not have any direct interactions with neighboring base pairs. At a glance it appears that the hydrogen bonding of the base pair and its ability to stack with neighboring base pairs remains intact, supporting the result that there is no large change in conformation or stability. While the overall shape remains intact, it can be seen that there is actually some distortion in the position of the main cytosine ring in the acetal-dC base. The extent of this distortion can be determined by examining the lengths of the base pair hydrogen bonds, as well as the angle torsion of the overall base pair. The hydrogen bond lengths in angstroms can be seen below in Figure 2-13.

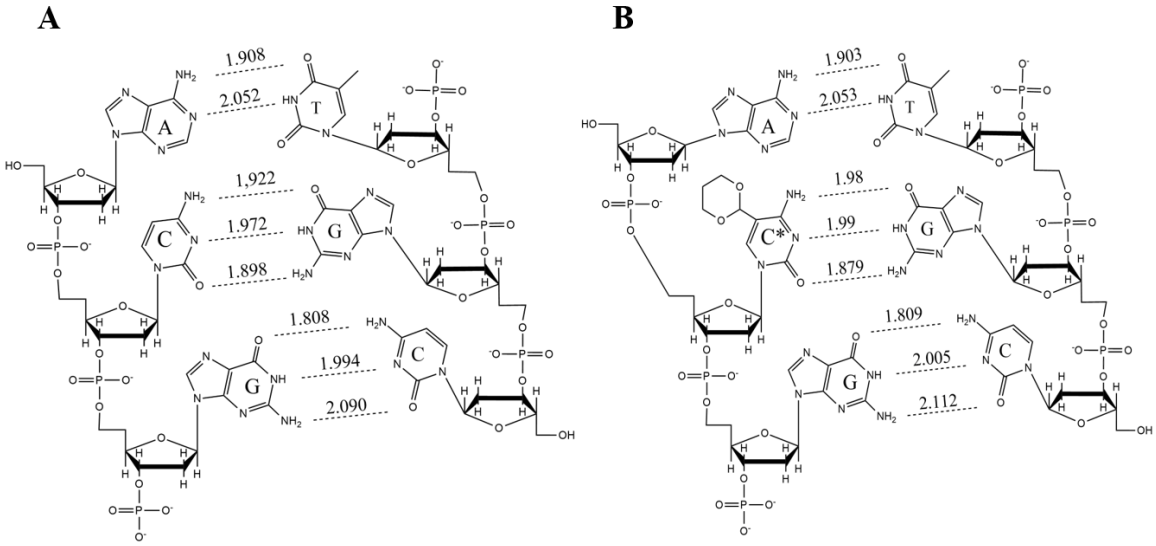


Figure 2-13 Hydrogen Bond Lengths of an Unmodified (A) and Modified (B) Region of DNA Duplex

From this data it can be seen that overall the hydrogen bonds do not significantly change when the lesion is introduced. All of the bond lengths still fall within the ideal

hydrogen bond length range of 1.5-2.5 Å, with no overall change greater than 0.1 Å. When examining the ring torsion there was more of a pronounced difference. The planes of the canonical cytosine and guanine base pair are rotated 10.8° from each other, but when introducing the lesion this torsion is increased to 13.5°. This increased torsion could affect the planarity of the base pairs, potentially impacting the pi-pi stacking interactions with neighboring base pairs. This could be the cause of the conformational change seen in the CD experiments and could also contribute to the decrease in stability seen in the UV experiments.

In addition, the molecular mechanics calculation was able to determine a 3.5 kcal/mol change in the overall energy of the system, mirroring what was seen in the UV studies. While this is a larger change than what was previously seen, it still confirms that introduction of the acetal-dC lesion to the duplex is not an energetically favorable process.

As we know, 5-formylcytosine is not isolated in the genome, but appears in clusters, especially in CpG islands. As a result, it is possible to have multiple lesions form in close proximity to each other, or even directly neighboring each other. To investigate the effects of this possibility, the same molecular mechanics simulations were ran on a duplex containing two neighboring, or tandem, lesions. As before, the duplexes were constructed and then geometrically optimized (Figure 2-14).

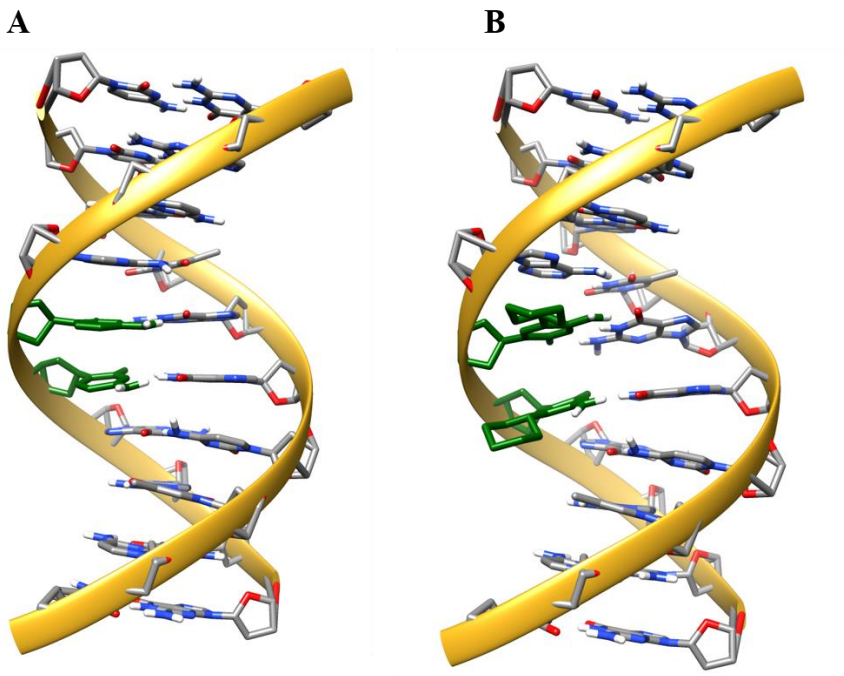


Figure 2-14 Molecular Mechanics Simulation of an Unmodified (A) and Modified (B) 10mer DNA Duplex. Acetal-dC Lesion Site is Shown in Green.

As with the single lesion, it can be seen that both acetal-dC adduct rings project into the major groove of the duplex, having no direct interferences with the base pairing or overall structure of the DNA duplex. Despite the retention of the overall conformation there does still appear to be some distortion resulting from the introduction of the lesions. The extent of this distortion can be determined by again examining the length of the hydrogen bonds and the ring torsion between each base pair. This data is summarized below in Table 2-4

Unmodified - Hydrogen Bonding			
	N-H --- O	N ---- H-N	O --- H-N
A4-T	1.86	2.22	
C5-G	1.86	2.03	2.07
C6-G	1.89	1.97	1.95
	O --- H-N	N-H --- N	N-H --- O
G7-C	1.81	2.02	2.12

Modified - Hydrogen Bonding			
	N-H --- O	N ---- H-N	O --- H-N
A4-T	1.83	2.5	
C*5-G	1.9	2.05	2
C*6-G	1.89	1.98	1.93
	O --- H-N	N-H --- N	N-H --- O
G7-C	1.75	1.97	2.11

Unmodified - Ring Torsion	
C5-G	9.5
C6-G	18.6

Modified - Ring Torsion	
C*5-G	10.1
C*6-G	15.4

Table 2-4 Hydrogen Bonding and Ring Torsion Angles of an Unmodified and Modified 10mer DNA Duplex.

When examining the hydrogen bonding, it can be seen that there is some fluctuation between the unmodified and modified values, however, all of the values still fall within the desired range of 1.5 – 2.5 angstroms. This fluctuation is also apparent in the ring torsion. In the first lesion base pair (C5-G) the ring torsion is relatively similar, but in the second base pair (C6-G) the distortion is about 3.2°. This is similar to the change in torsion seen in the presence of the single lesion. This distortion could lead to an impact in the ability of the duplex to replicate, a phenomenon that will be examined in the next chapter of this study. When examining the energy of the tandem duplex reaction, it was seen that the introduction of two lesions is actually slightly energetically favorable, with a ΔE of -3.4 kcal/mol. This differs from the reaction with a single lesion, where we saw a ΔE of +3.5 kcal/mol. This was a surprising result, showing that the presence of one lesion is energetically unfavorable, but the presence of two is potentially

favorable. More extensive physical characterization is needed in order to further investigate and understand this result.

Conclusion

It was seen in the lesion formation portion of the study that DNA lesions are able to form from the reaction of environmental nucleophilic toxins and DNA containing 5-formylcytosine. In the GC/MS model study, it was seen that both ethylene glycol and phenylhydrazine can readily react with acetaldehyde to form adducts. It was also established that there is a pH dependency when reacting 1,3-propane diol with acetaldehyde. This information was carried over to the ESI-MS study where it was shown that ethylene glycol, hydroxylamine hydrochloride, methoxyamine hydrochloride and phenylhydrazine all readily react with DNA containing 5-formylcytosine to form DNA adducts. This confirmed the relevance of studying this particular group of toxins and led us to select the acetal-dC lesion formed from the reaction of 1,3-propane diol for future studies due to its commercial availability.

In the physical studies, it was shown through CD that the acetal-dC lesion potentially has a minor effect on the conformation of the DNA duplex. It was also shown that the acetal-dC lesion has minimal effect on the stability of the DNA duplex. This was shown through single concentration UV thermal denaturation experiments, as well as a concentration dependent experiment utilizing the 10mer DNA duplex. These results were further confirmed by examining the thermodynamic constants of a modified and unmodified DNA duplex. Molecular dynamics simulations were able to shed light on the source of the conformational and stability changes, revealing that the introduction of the

acetal-dC lesion results in increased ring torsion. Molecular dynamics was also able to confirm to the destabilization effect found in the UV studies.

Overall, it was seen that the acetal-dC lesion does not drastically affect the physical properties of the DNA duplex. In regards to lesions formed from other environmental toxins, it is possible that a varying range of effects may occur. Smaller lesions, such as those formed by the methoxyamine and hydroxyamine may mirror the acetal-dC lesion. They are relatively small and short chained and could fit into the major groove as well, not largely affecting the conformation or stability. On the other hand, the lesion formed by phenylhydrazine is quite large and bulky due to the benzene ring. This may cause it to widen the major groove, in turn compressing the minor groove, resulting in a relatively large conformational and stability change in the duplex.

In addition, these studies were conducted with only one acetal-dC lesion. Upon the addition of multiple lesions, the minor effects seen in the thermodynamic constants may be compounded, manifesting in larger conformational and stability changes. While no significant change was seen in the specific experiments run, this is not enough information to completely exclude effects from environmental toxins in general.

Chapter 3: Enzymatic Studies

Introduction

It has been seen that the acetal lesion formed from the reaction of 5-formylcytoinse and 1,3-propane diol can have minor effects on the stability and conformation of the DNA duplex. This makes it necessary to investigate if these small physical changes have any potential effect on the chemical processes that the duplex is involved in. One of the most important functional traits of DNA is its ability to be replicated by DNA polymerases. In this chapter, we investigated whether the acetal-dC lesion has any effects on the ability of the cell to replicate.

Studies have shown that different types of DNA lesions can have a wide array of effects on DNA replication. While DNA is a complex and sometimes delicate molecule, it is also fairly robust. Minor lesions can be easily accommodated, and sometimes have no effect on the replication process. Certain polymerases are also specialized to manage damaged DNA by performing lesion bypass synthesis. Depending on the severity of the lesion this bypass can either be relatively error free or error prone. If error free, replication occurs normally and eventually the lesion is diluted, having no long-lasting effect on the cell as a whole. If error prone, it is possible that an incorrect base be incorporated across from the lesion. After another round of replication this mismatch can

generate a full AT to GC switch or vice versa, making these types of lesions mutagenic. The most severe lesions may be unable to be accommodated by polymerase at all, completely halting all replication processes. This could lead to cell death, making these types of lesions cytotoxic.

In order to examine the effects of the acetal-dC lesion on DNA polymerase, an assay was designed utilizing a radioactive isotope of phosphorous (^{32}P) to label several DNA primers of varying lengths. By annealing the labelled primer to a template containing the lesion we are able to observe whether or not the polymerase is able to extend past the lesion to complete the replication. The template and primer sequences are shown below in Table 3-1. Altering the dNTP content can also show whether the lesion has any effects on the fidelity of the replication process as well. This assay will allow us to determine what effect the lesion has on the genome, if any at all.

1	AGCGATGAGAGGCCACGAGGAATCGCTGGTACCG
2	AGCGATGAGAGGCCAC*GAGGAATCGCTGGTACCG
3	AGCGATGAGAGGTACCGAGGAATCGCTGGTACCG
4	AGCGATGAGAGGTAC*C*GAGGAATCGCTGGTACCG
5	CGGTACCAGCGAT
6	CGGTACCAGCGATTCCTC
7	CGGTACCAGCGATTCCTCG
8	CGGTACCAGCGATTCCTCA
9	CGGTACCAGCGATTCCTCC
10	CGGTACCAGCGATTCCTCT

Table 3-1 DNA Sequences Used in Replication Studies. C* = acetal-dC lesion

For these experiments three different kinds of polymerase were selected: Klenow Fragment with exonuclease activity (KF exo$^+$), Klenow Fragment without exonuclease activity (KF exo$^-$) and Reverse Transcriptase (RT). Klenow fragment is a β family polymerase obtained by enzymatically cleaving off the 5'-3' exonuclease domain. While this portion has been removed, KF exo$^+$ does still retain a 3'-5' exonuclease activity that acts to proofread the extension as the new strand is synthesized. Cleaving off this exonuclease activity as well results in the KF exo$^-$ fragment. By utilizing both polymerases we are able to see whether the exonuclease activity has any effect on the response of the polymerase when encountering the acetal-dC lesion. Reverse transcriptase is an enzyme generally utilized by retroviruses to convert viral RNA back into DNA; however, it can also be used to read DNA and perform replication. Reverse transcriptase also has no exonuclease activity, making it more prone to be negatively impacted by DNA lesions.

In order to test if the concentration of the polymerase has any effect on its ability to overcome the lesion, each reaction was run at two different concentrations. The lower concentration is considered the minimum amount needed to complete an extension of the DNA template. For KF exo$^+$, exo$^-$ and RT that value is 0.02 U, 0.01 U and 4 U, respectively. At these lower concentrations the polymerase may be more susceptible to interference from the lesion. The higher concentration used is considered an excess of enzyme. For KF exo$^+$, exo$^-$ and RT, these values are 0.2 U, 0.1 U and 10 U, respectively. These values serve to saturate the DNA with polymerases, possibly helping the polymerase to overcome the effects of the acetal-dC lesion.

After running the selected experiments using a template containing a single lesion, it was noted that 5-formylcytosine often appears in clusters in the genome. As a result, it is possible to have clusters of the acetal-dC lesions as well. In order to investigate this effect, another template was synthesized with two acetal-dC lesions directly neighboring each other. This sequence can also be seen in Table 3-1 (#4).

Experimental

DNA Synthesis. Oligonucleotides were synthesized on an Applied Biosciences 392 DNA synthesizer using standard phosphoramidite protocols. Modified bases were also incorporated in their phosphoramidite form. All the chemicals for DNA synthesis were purchased from Glen Research (Sterling, VA). The resulting oligonucleotides were purified through gel electrophoresis and column chromatography. The concentrations of oligonucleotide solutions were determined by measuring their absorbance at 260 nm using a Varian Cary 100 Bio UV-Vis spectrophotometer (Walnut Creek, CA). Absorbance was converted into concentration using the molar extinction coefficients obtained from Oligo Analyzer 3.1 (www.idtdna.com).

[γ-^{32}P] Labelling. DNA sequences were labelled by preparing a solution of 10 μM oligonucleotide, 20 units of T4 polynucleotide kinase, T4 buffer and 300 μCi of [γ-^{32}P]-ATP. Solutions were incubated at 37° C for 45 minutes and then stopped by heating to 70° C for 5 minutes. Excess [γ-^{32}P]-ATP was removed using a MicroSpin G-25 column from GE Healthcare. T4 polynucleotide kinase was obtained from New England Biolabs. [γ-^{32}P] was purchased from PerkinElmer (Waltham, MA).

Full Length Replication Assay. Reactions were prepared with a final concentration of 0.06 μM of template and 0.05 μM of radiolabeled primer. Templates were either unmodified or modified containing a single lesion (#1 and #2) or two tandem lesions (#3 and #4). Primers were interchanged depending on the experiment; #5 for 13mer, #6 for 18mer and #7-10 for the 19mer experiments. dNTP was added to a final concentration of 100 μM as well as buffer for either Klenow Fragment or MMLV-Reverse Transcriptase. Reaction mixtures were then annealed by heating to 90° C for 1 minute followed by 55 °C for 5 minutes. Polymerase was then added to the reactions, with either 0.02 or 0.2 U for KF exo$^+$, 0.01 or 0.1 U for KF exo$^-$ and 4 or 10 U for RT. Reactions were then incubated for 30 minutes at 37 °C for Klenow experiments and 42 °C for Reverse Transcriptase experiments. After incubation, 3 μL of a loading dye containing bromophenol blue, xylene cyanol and glycerol was added to each reaction mixture. Reactions were then heated to 70 °C for 5 minutes before being loaded into a 20% polyacrylamide gel. After electrophoresis, gels were analyzed by use of a Storm 860 phosphorimager and ImageQuant 5.1 software. All polymerases and buffers were purchased from New England Biolabs.

Single Base Extension Assay. Reactions were prepared with a final concentration of 0.06 μM of template and 0.05 μM of radiolabeled primer #6. Templates were either unmodified or modified containing a single lesion (#1 and #2) or two tandem lesions (#3 and #4). Each reaction contained only one of the four dNTPs with a final concentration of 10 μM, as well as buffer for either Klenow Fragment or MMLV-Reverse Transcriptase. Reaction mixtures were then annealed by heating to 90 °C for 1 minute followed by 55 °C for 5 minutes. Polymerase was then added to the reactions, with either

0.02 U KF exo$^+$, 0.01 U KF exo$^-$ and 4 U RT. Reactions were then incubated for 30 minutes at 37 °C for Klenow experiments and 42 °C for Reverse Transcriptase experiments. After incubation, 3 μL of a loading dye containing bromophenol blue, xylene cyanol and glycerol was added to each reaction mixture. Reactions were then heated to 70 °C for 5 minutes before being loaded into a 20% polyacrylamide gel. After electrophoresis, gels were analyzed by use of a Storm 860 phosphorimager and ImageQuant 5.1 software. All polymerases and buffers were purchased from New England Biolabs.

Results and Discussion

Single Lesion – 13mer primer. The first extension experiment ran utilized a template containing the cyclic acetal-dC lesion roughly in the center of the template. The primer used was a 13mer, annealing five bases upstream from the lesion. This design allows the polymerase to complex to the duplex and begin replication before encountering the lesion, revealing whether or not the lesion will have some effect on the processivity of the polymerase. In addition to the reaction lanes, a labelled 18mer sequence was also included as a marker for the expected product location if extension was to halt at the lesion site. The results of this experiment are shown below in Figure 3-1.

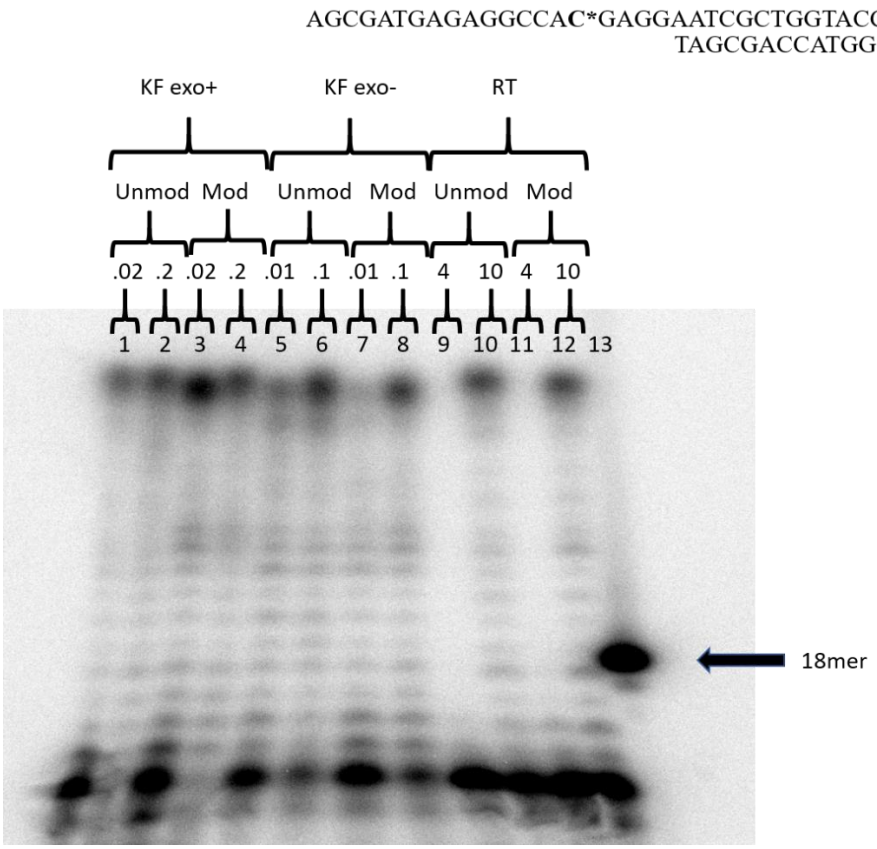


Figure 3-1 Extension from a 13mer Primer. Lanes 1, 2, 5, 6, 9 and 10: control reactions using 0.06 μM of an unmodified template, 100 mM dNTP and either 0.02 or 0.2 U KF exo$^+$ (1, 2), 0.01 or 0.1 U KF exo$^-$ (5, 6), or 4 or 10 U RT (9, 10). Lanes 3, 4, 7, 8, 11 and 12: reactions using 0.06 μM of a modified template, 100 mM dNTP and either .02 or 0.2 U KF exo$^+$ (3, 4), 0.01 or 0.1 U KF exo$^-$ (7, 8), or 4 or 10 U RT (11, 12). Lane 13: An 18 nucleotide marker.

The first four lanes of this experiment utilized Klenow Fragment exo$^+$ for the polymerase. The reactions in lanes 1 and 2 were carried out using an unmodified template that contained a regular cytosine rather than the acetal-dC lesion. In both of these lanes there are bands at the top of the gel, indicative of a full length 34 nucleotide long product. The following two lanes (3 and 4) had reactions using a modified template, containing the acetal-dC lesion. Despite this, they also both resulted in bands

corresponding with the full-length extension product. It was also seen that this extension occurs regardless of the concentration of polymerase used. In addition, it was noted that there were no product bands that aligned with the 18mer marker (lane 13), indicating that the lesion did not halt the polymerase at this location.

The next set of four lanes (5-8) utilized Klenow Fragment exo$^-$ as a polymerase. The control lanes (5 and 6) utilized an unmodified template and similar to the KF exo$^+$ lanes, showed full length product bands. When using a modified primer (lanes 7 and 8), this extension ability was retained, resulting in full length products, and no bands for the 18mer partial product. It was noted that the intensity of the band in the lower concentration lanes (5 and 7) was weaker than the bands in the high concentration lanes (6 and 8); however, because this trend is mirrored in the control and modified reactions this result is attributed to the activity of the polymerase itself rather than any effects from the lesion.

The last set of lanes (9-12) utilized Reverse Transcriptase as its polymerase. The results in this section are very analogous to the ones seen in the exo$^-$ lanes. Both the unmodified and modified reactions carried out at low concentrations (lanes 9 and 11) showed low activity, indicative of an insufficient amount of polymerase to catalyze full extension. In both of the higher concentration lanes (10 and 12) full length products are seen, as well as an absence of any 18mer partial products.

To summarize, it was seen that when using a short primer that anneals upstream of the acetal-dC lesion, all polymerases were able to continue past the lesion to complete the full extension of the primer. One possible explanation for this result is the way that

the DNA duplex interacts with the polymerase. Polymerases are large enzymes with multiple subunits and at any given time there are 6-8 DNA base pairs contained within the polymerase, interacting with a variety of residues. The active site of the polymerase is filled with positively charged residues, whose side chains are able to interact with the negatively charged phosphate backbone of DNA (Figure 3-2). These interactions serve to hold the duplex in the correct position for extension to occur and are relatively sequence non-specific.[48]

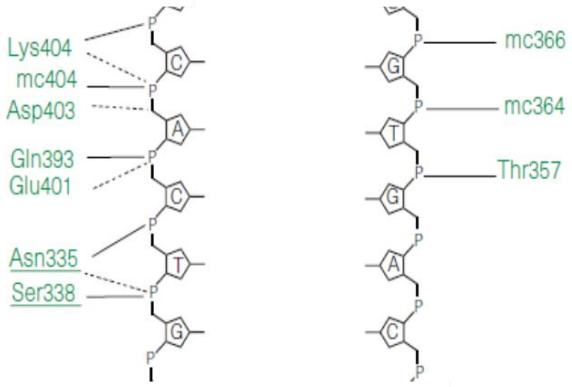


Figure 3-2 Interactions Between Positively Charged Side Chains in DNA Polymerase Active Site and the Negatively Charged DNA Phosphate Backbone.[48]

There are also more specialized interactions surrounding the 3' end of the primer and the template base that is being replicated. On the primer side there is a collection of residues that coordinate with the phosphate groups of the incoming dNTP, as well as with the magnesium cofactors needed for catalysis. Additional interactions with the sugar ring of the incoming base help to position it correctly for pairing (Figure 3-3). Collectively

these interactions serve to hold the incoming base in at the correct orientation to form a hydrogen bond with the templated base, as well as to optimize the attack of the 3' primer terminus on the phosphate groups of the incoming base.[49]

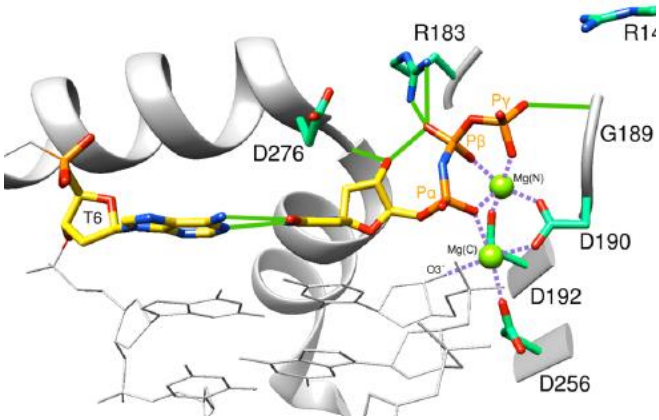


Figure 3-3 Interactions Between DNA Polymerase and the Incoming dNTP.[49]

On the template side, there is a crucial stacking interaction between the base being copied and a residue in the O-helix of the polymerase (Figure 3-4). Typically, this type of pi-pi stacking occurs between neighboring bases in a DNA sequence, but during replication the template strand is bent and this pi-pi stacking is lost. This enzyme residue serves to restore some of the stability that was lost with the pi-pi stacking. In addition, an unpaired nucleotide can undergo free rotation, which could be detrimental to forming hydrogen bonds with the incoming base. This stacking interaction holds the template base in the right orientation so the new hydrogen bonds can be formed.[50]

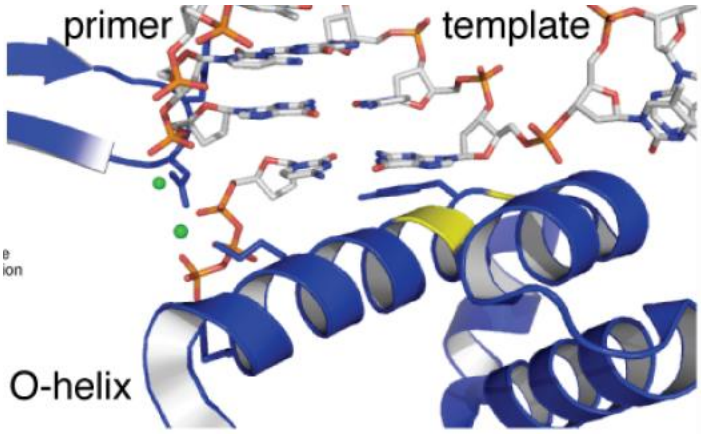


Figure 3-4 Stacking Interaction Between the O-helix of DNA Polymerase and the Template Base.[50]

When considering the acetal-dC lesion, it is important to remember that it is located on the template strand. As a result, the enzyme-phosphate backbone interactions should not be affected. The interactions between the enzyme and the incoming dNTP should remain intact as well. The only interaction potentially affected would be the stacking interaction in the template strand. If the acetal-dC lesion were to change the position of the base, this stacking interaction may be disrupted, having a detrimental effect on the formation of the nascent base pair. As seen though, there is no effect, and the polymerase is able to fully extend the template. This could be because the acetal ring projects into the major groove of the duplex, without severely altering the conformation of the main cytosine ring, allowing that crucial stacking interaction to remain intact.

Single Lesion – 18mer primer. The study with the short 13mer primer showed that the lesion has no effect on replication if the polymerase is already bound to the template and begun the extension process. We also wanted to investigate whether the lesion would have any effect on the polymerase when synthesis began at the lesion site,

rather than upstream. In order to investigate this, a reaction was designed using an 18mer primer, annealing just one base before the lesion. This brings the lesion into the binding pocket of the polymerase when it begins to associate with the DNA sequence, potentially effecting its initial recognition and binding capabilities. The results of this study are shown below in Figure 3-5.

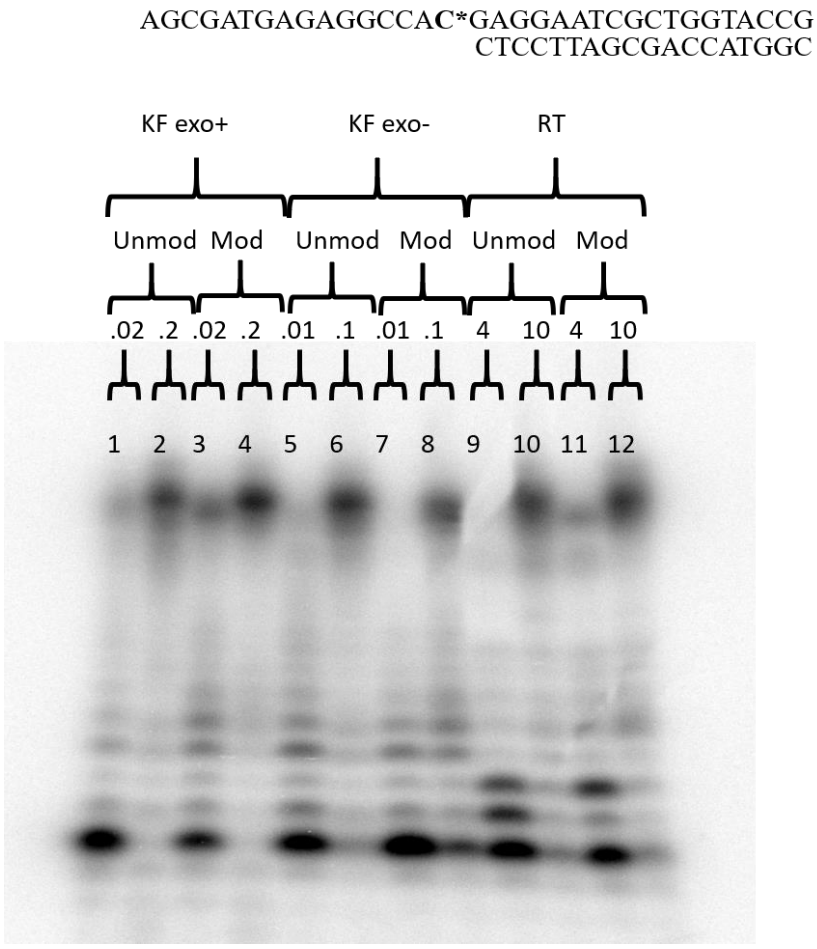


Figure 3-5 Extension from a 18mer Primer. Lanes 1, 2, 5, 6, 9 and 10: control reactions using 0.06 μM of an unmodified template, 100 mM dNTP and either 0.02 or 0.2 U KF exo$^+$ (1, 2), 0.01 or 0.1 U KF exo$^-$ (5, 6), or 4 or 10 U RT (9, 10). Lanes 3, 4, 7, 8, 11 and 12: reactions using 0.06 μM of a modified template, 100 mM dNTP and either .02 or 0.2 U KF exo$^+$ (3, 4), 0.01 or 0.1 U KF exo$^-$ (7, 8), or 4 or 10 U RT (11, 12).

This experiment followed a pattern similar to the previous, with alternating low and high concentrations of each of the three polymerases. The results were also fairly similar to the previous experiment. When using KF exo$^+$, full length products were seen for all reactions. There was a noted intensity difference between the low and high concentrations due to the differing enzyme activity. This product pattern was repeated for both KF exo$^-$ and RT.

Overall it was seen that when the primer annealed near the lesion, the lesion had no effect on the ability of the polymerase to recognize and bind to the DNA duplex. This could be for the same reasons that the polymerase is able to extend the shorter 13mer primer. While the primer length and polymerase position are different, extension of each base involves the same cycle of binding, catalysis and translocation or dissociation of the enzyme.[49] Whether the polymerase is binding to the primer for the first time, as in the case of the 18mer, or binding after dissociating from a previous base, as in the case of the 13mer, the overall binding mechanics are the same. Extension is still dependent on the interactions guiding the orientation of the incoming dNTP and the stacking interaction positioning the template base and as seen previously, these interactions are left intact, resulting in the full extension of the primer sequence.

Single lesion – 18mer primer – Single base. The previous experiments have shown that despite the presence of the acetal-dC lesion the polymerase has been able to complete replication and synthesize full length products. These results do indicate the presence of a 34 nucleotide long product; however, they do not provide any information regarding the content of these 34 bases. While the lesion does not cause enough perturbation to disrupt the enzyme stacking, it may still have some small distortion that

can affect the hydrogen bonding capacity of the base. If this is the case, an incorrect base may be incorporated across from the acetal-dC lesion, resulting in a mismatched base pair. In order to investigate this possibility, a set of reactions were designed utilizing the same 18mer primer as previous. However, instead of dNTP, each reaction only received one the of the four dNTPs. Through this method it can be observed if the expected guanine is incorporated across from the lesion, or if another dNTP is incorporated instead. The results for this experiment are shown below in Figure 3-6.

AGCGATGAGAGGCCA**C***GAGGAATCGCTGGTACCG
CTCCTTAGCGACCATGGC

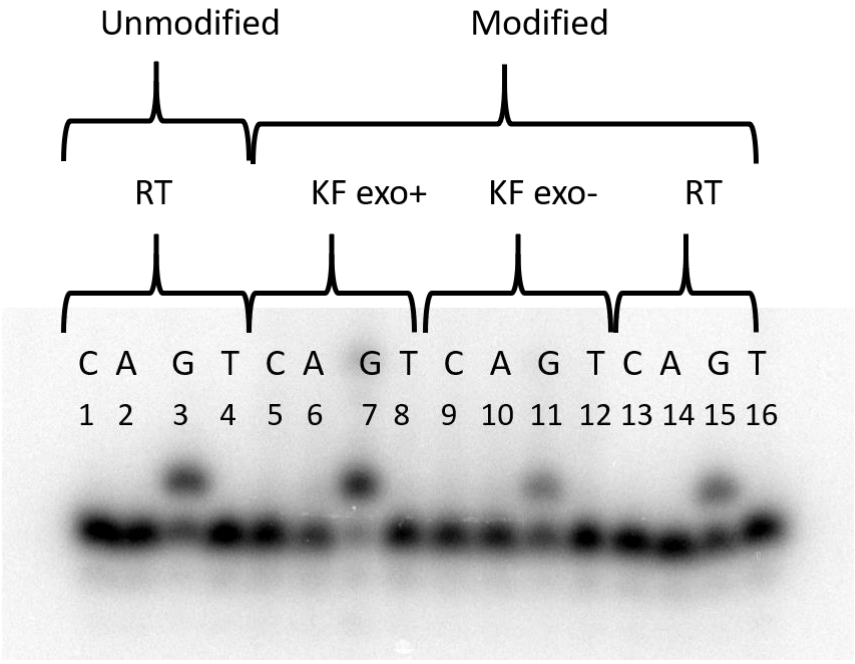


Figure 3-6 Single Nucleotide Extension of an 18mer Primer. Lanes 1, 2, 3 and 4: control reactions using 0.06 μM of an unmodified template, 4 U RT and 10 mM of either dCTP (1), dATP (2), dGTP (3), or dTTP (4). Lanes 5, 9 and 13: reactions using 0.06 μM of a modified template, 10 mM dCTP and either 0.02 U KF exo$^+$ (5), 0.01 U KF exo$^-$ (9) or 4 U RT (13). Lanes 6, 10 and 14: reactions using 0.06 μM of a modified template, 10 mM dATP and either 0.02 U KF exo$^+$ (6), 0.01 U KF exo$^-$ (10) or 4 U RT (14). Lanes 7, 11 and 15: reactions using 0.06 μM of a modified template, 10 mM dGTP and either 0.02 U KF exo$^+$ (7), 0.01 U KF exo$^-$ (11) or 4 U RT (15). Lanes 8, 12 and 16: reactions using 0.06 μM of a modified template, 10 mM dTTP and either 0.02 U KF exo$^+$ (8), 0.01 U KF exo$^-$ (12) or 4 U RT (16).

In this gel the first 4 lanes are reactions carried out using an unmodified control template. Reverse transcriptase was utilized for the controls. It was seen that the presence of dGTP (lane 3) resulted in a single base extension giving a slightly higher band. This is expected, as the unmodified version of the template contains a regular cytosine. No incorporation was seen for any of the other three dNTPs, indicating the

fidelity of the polymerase. When moving to the modified template it was also seen that dGTP is the only base incorporated across from the lesion regardless of the polymerase used (lanes 7, 11 and 15). When using KF exo$^+$, the intensity of the band appears to match that of the control, indicating that the efficiency of incorporation is not affected either. On the other hand, the bands for the KF exo$^-$ and RT catalyzed reactions do appear to be lower in intensity, indicating a possible decreased efficiency of incorporation.

Overall, it was seen that the acetal-dC lesion did not have an effect on the fidelity of the DNA polymerase when incorporating nucleotides across from the lesion. In order to understand this result, it is necessary to examine how the polymerase facilitates the correct pairing of bases. The formation of a new base pair is dependent on two main factors: the hydrogen bonding between the new bases and an interaction between the minor groove side of the base pair and the polymerase. When a new dNTP is recruited to the duplex, the polymerase is initially in what's considered an "open" conformation. The previously discussed residues will associate with the incoming dNTP and guide it into place across from the template nucleotide, allowing it to hydrogen bond. This bond facilitates a conformational change of the polymerase from the "open" form to a "closed" form where catalysis can then take place, joining the new dNTP to the primer. If the base is incorrect, the hydrogen bonds cannot properly form, and the transition to the closed conformation will not occur, in turn preventing catalysis from taking place and stalling the polymerase.[49]

In addition to these inter-nucleotide hydrogen bonds, there are also hydrogen bonds formed between the minor groove of the base pair and the polymerase.

Pyrimidines and purines contain hydrogen bond acceptors at the O2 and N3 positions respectively, making them available to receive hydrogen bonds from the enzyme.[51] If an incorrect base pair was generated, the positions of the O2 and N3 atoms may be positioned incorrectly, making this interaction impossible to form. This could result in a weakened DNA-protein complex, causing more stalling in the polymerase.

In the case of the acetal-dC lesion, the adduct does not distort the base enough to interfere with the hydrogen bonding capabilities. The bonding face is not modified, allowing it to form correct bonds with an incoming dGTP and further allowing the polymerase to transition into the closed conformation. In addition, the adduct does not significantly change the position of the O2 atom, allowing the minor groove-enzyme interaction to still form (Figure 3-7). With these interactions unaffected it follows that the fidelity of the polymerase is not affected when replicating across from the acetal-dC lesion.

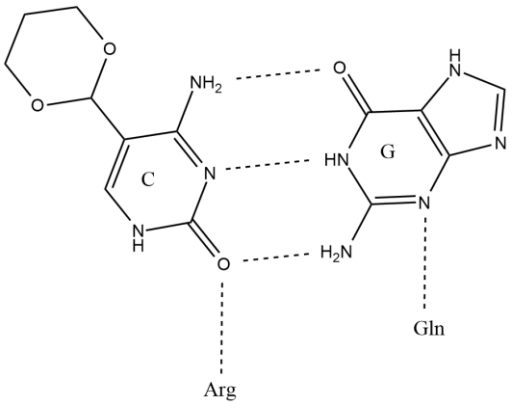


Figure 3-7 Base Pairing and Minor Groove Hydrogen Bonding of the acetal-dC Lesion and DNA Polymerase

Single Lesion – 19mer primers. Although DNA replication is largely carried out with high fidelity, polymerases can be prone to error and incorporate incorrect base pairs. This potential is magnified even further for polymerases that do not contain exonuclease activity, such as KF exo$^-$ and RT. A modified nucleotide (acetal-dC) paired with a mismatched base could make it even more difficult for DNA polymerase to bind and complete extension of the primer. In order to investigate this potential, a series of 19mer primers were designed that will anneal directly across from the lesion site in the template. Four different primers were designed, each with a different base across from the lesion, simulating each of the mismatch possibilities. The results of this experiment are shown below in Figure 3-8.

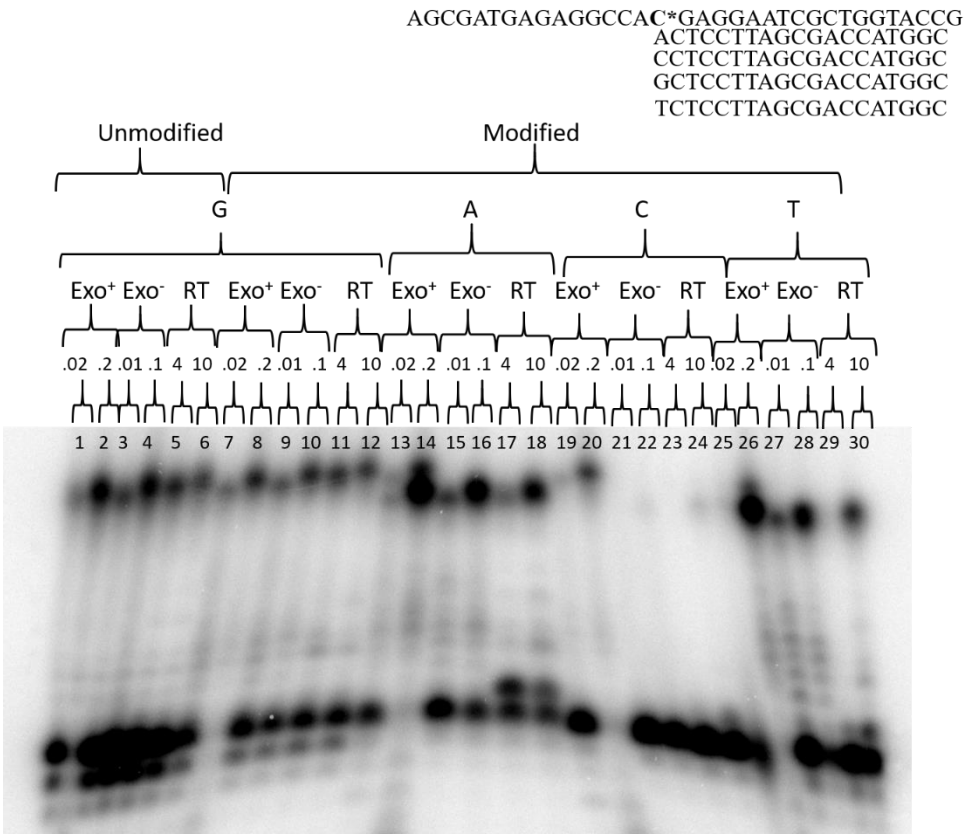


Figure 3-8 Extension of 19mer Primers Varying in Nucleotide Paired with the acetal-dC Lesion. Lanes 1-6: control reactions using 0.06 μM of an unmodified template, 100 mM dNTP, 0.05 μM of a 19mer primer with a 3' G and either 0.02 or 0.2 U KF exo$^+$ (1, 2), 0.01 or 0.1 U KF exo$^-$ (3, 4), or 4 or 10 U RT (5, 6). Lanes 7-12: reactions using 0.06 μM of a modified template, 100 mM dNTP, 0.05 μM of a 19mer primer with a 3' G and either 0.02 or 0.2 U KF exo$^+$ (7, 8), 0.01 or 0.1 U KF exo$^-$ (9, 10), or 4 or 10 U RT (11, 12). Lanes 13-18: reactions using 0.06 μM of a modified template, 100 mM dNTP, 0.05 μM of a 19mer primer with a 3' A and either 0.02 or 0.2 U KF exo$^+$ (13, 14), 0.01 or 0.1 U KF exo$^-$ (15, 16), or 4 or 10 U RT (17, 18). Lanes 19-24: reactions using 0.06 μM of a modified template, 100 mM dNTP, 0.05 μM of a 19mer primer with a 3' C and either 0.02 or 0.2 U KF exo$^+$ (19, 20), 0.01 or 0.1 U KF exo$^-$ (21, 22), or 4 or 10 U RT (23, 24). Lanes 25-30: reactions using 0.06 μM of a modified template, 100 mM dNTP, 0.05 μM of a 19mer primer with a 3' T and either 0.02 or 0.2 U KF exo$^+$ (25, 26), 0.01 or 0.1 U KF exo$^-$ (27, 28), or 4 or 10 U RT (29, 30).

In this experiment the first set of lanes (1-6) were control, utilizing an unmodified template and each of the three polymerases. In addition, these reactions were carried out with primer #7, ending in a G, the correct base to pair with the unmodified cytosine. As

expected, all reactions in this section gave full length products, with varying intensity depending on the concentration of the enzyme. The next set of lanes (7-12) continued to utilize primer #7, but with a modified template containing the acetal-dC lesion. In the reactions using KF exo$^+$ (lanes 7 and 8) full length products were seen that were analogous to those seen in the control. The lower concentration polymerase reaction gave a fainter band, while the higher concentration gave a strong band. When using KF exo$^-$ (lanes 9 and 10) and RT (lanes 11 and 12) the same pattern was seen, with full length products matching the control. This indicated that if the lesion was correctly paired with a guanine then DNA polymerases are able to bind and complete extension normally. This is very similar to what was seen when using the 18mer and 13mer primers and helps to reinforce those previous results. The acetal-dC:G base pair is correct and satisfies the necessary hydrogen bonding requirements, presenting no complication for the polymerase.

The next set of lanes (13-18) utilized a 19mer primer ending with adenine, creating a mismatched acetal-dC:A base pair. When KF exo$^+$ was used at a low concentration (lane 13), two very weak bands were seen. One matches the full-length product seen in the control, and one representing an N-1 type product. When a higher concentration was used (lane 14), these same two bands were seen, but at much higher intensities. It was also noted that at this high concentration the N-1 product seemed to be equal to or even slightly more intense that the full-length product band. When KF exo$^-$ was used the low concentration band (lane 15) only showed minor amounts of the N-1 product. At a higher concentration (lane 16) these N-1 products were amplified, while no full-length product was synthesized. When RT was used at a low concentration (lane 17)

a weak N-1 product was seen, as well as a strong single base extension product near the bottom primer line. When the concentration was increased (lane 18) the intensities of these bands were reversed; the N-1 becoming the dominant product seen.

In the case of the reactions catalyzed by KF exo$^+$, there is an equilibrium between the exonuclease activity and a mechanism involving primer misalignment followed by lesion bypass synthesis. The full-length product is a result of the exonuclease activity of the polymerase. The formation of an acetal-dC:A base pair does not allow the proper formation of hydrogen bonds between the two bases as previously described. The polymerase can recognize this discrepancy and stall, shifting the primer out of the active site into an exonuclease domain. Here the 3' end of the primer can be cleaved off, removing the mismatch and then can be moved back into the active site of the polymerase. This brings the reaction back to a position similar to when using the 18mer primer; a guanine can be added across from the lesion and then extension proceeds as normal. This results in a full-length product, giving the upper band that is seen.

Alternatively, because the polymerase is beginning after the lesion site and the mismatch is preexisting, the polymerase may not always recognize the error and repair it. In this situation the duplex must have some way to accommodate the mismatch base pair. In an unmodified sequence a mismatched base pair can be accommodated through flexible hydrogen bonding. There is no "correct" orientation, so the mismatch is constantly interchanging between different patterns of hydrogen bonding. In the case of a C:A mismatch this interchanging pattern involves the rotation of the cytosine ring around the glycosidic bond (Figure 3-9).

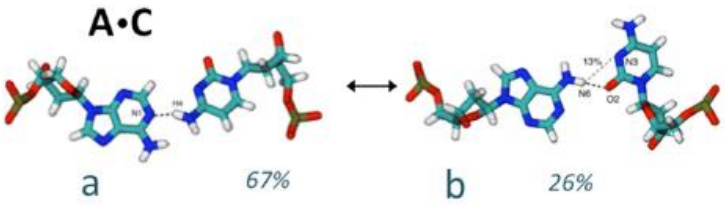


Figure 3-9 Flexible Hydrogen Bonding of a Mismatched A:C Base Pair[52]

This rotational change is tolerable in the case of an unmodified cytosine, but not with the acetal-dC lesion. The ring adduct projects out from the base and can sterically clash with the phosphate backbone if moved from the normal orientation, making this acetal-dC:A mismatch intolerable. In addition, this type of mismatch induces local structural changes, effecting the twist and roll angles of the bases directly neighboring the mismatch site, making this pairing even more unfavorable.[52]

As a result, it is more energetically favorable for the acetal-dC to be pinched out of the template. This is achieved through primer misalignment, placing the 3' terminal primer adenine across from the template adenine 5' from the lesion, bending the lesion out of the template sequence. This type of misalignment mechanism has been shown to readily occur when replicating across from abasic sites as well, making it a viable mechanism for this type of situation.[53] By excluding the lesion site the polymerase is more easily able to complete the extension, resulting in the deletion of one nucleotide (Figure 3-10).

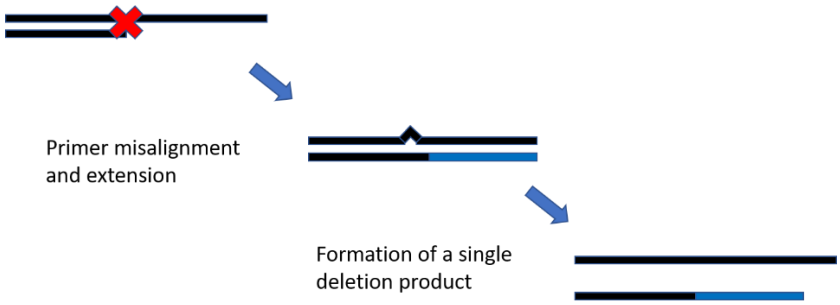


Figure 3-10 Primer Misalignment Leading to the Formation of a Single Nucleotide Deletion Product.

Although the resulting A:A base pair is still a mismatch the interchanging cycle of hydrogen bonds involves a much smaller path of movement than a C:A mismatch and also far less local distortion (Figure 3-11). In addition, the absence of the lesion makes this flexible bonding more tolerable. Due to these distortions being relatively minor the polymerase is able to accommodate the change and perform lesion bypass synthesis, completing the extension, giving rise to a final product that is one nucleotide shorter than the full-length product (N-1).

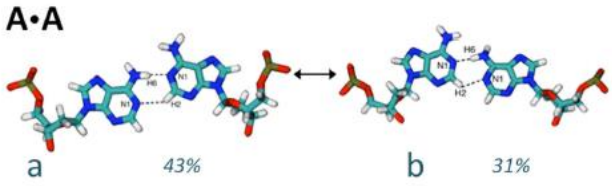


Figure 3-11 Flexible Hydrogen Bonding of an A:A Mismatched Base Pair.[52]

When using KF exo$^-$ and RT as polymerases to catalyze the reaction, the full-length product cannot be formed because these enzymes do not contain exonuclease activity. As a result, the primer misalignment mechanism is the only viable option, giving solely N-1 type products. As for the single base product seen when using RT, this could be a result of decreased lesion bypass ability. Previous studies have shown that exonuclease activity is linked to the ability to perform lesion bypass synthesis. KF exo$^+$ has this activity and is therefore very efficient at error free lesion bypass. KF exo$^-$ is lacking this activity, but studies have shown that it still has moderate lesion bypass activity.[54] The A:A base pair provides a level of distortion that is able to be managed by KF exo$^-$, resulting in the N-1 products. RT on the other hand, can be considered a more sensitive enzyme, with a much lower bypass capability. In some cases RT has even been seen to completely halt DNA replication when encountering a lesion.[55] The single base product seen could be a result of the RT being unable to overcome the distortion, and thus halting replication at that point. It does still retain some bypass activity though, as indicated by the presence of the N-1 bands.

The next set of lanes (19-24) utilized primer #9, placing a regular cytosine across from the modified acetal-dC. When KF exo$^+$ was used at a low concentration (lane 19), an extremely faint band for full length product was seen. By contrast, when the concentration was increased (lane 20), the product also increased to an intensity slightly weaker than that of the control. This indicated an effect on the overall efficiency of the polymerase. When this reaction was run with KF exo$^-$ and RT, no bands were seen in the gel except the initial primer, indicating that no extension of any kind occurred.

When considering the full-length product given by KF exo$^+$ it can be discerned that this is a result of the exonuclease activity. As seen with the previous set of reactions the exonuclease is able to cleave off the mismatch and then complete synthesis of a full-length product. The lack of N-1 bands for all three of the polymerases indicates that the primer misalignment method is unable to occur in the case of an acetal-dC:C mismatch. Similar to a C:A mismatch, the conformational change involved in the formation of a C:C mismatch cannot be accommodated in the presence of the acetal ring adduct. Naturally, the primer would attempt to misalign, placing the cytosine across from the 5' template adenine to form a C:A mismatch. Unfortunately, C:A mismatches involve a large amount of rotational change around the hydrogen bond, leading to a significant amount of local distortion. As mentioned, this distortion effects the twist angles of the neighboring base pairs and can even cause compression of the minor groove.[52] This could disrupt some of the minor groove hydrogen bonding interactions that are necessary for the polymerase to function properly. These changes may prove to be too significant to be overcome by lesion bypass synthesis, resulting in a complete halt of extension and no formation of product.

The last set of reactions (lanes 25-30) utilized primer #10, placing a thymine across from the acetal-dC lesion. The results for this reaction were very similar to those seen when the lesion was paired with adenine (lanes 13-18). When KF exo$^+$ was used at a low concentration (lane 25) no product bands were seen. When a higher concentration (lane 26) was used, two product bands were seen, one matching the full-length product seen in the control, and one representing an N-1 type product. When KF exo$^-$ was used at the low concentration (lane 27), there were only minor amounts of the N-1 product. At a

higher concentration (lane 28), these N-1 products were amplified, while no full-length product was synthesized. When RT was used at a low concentration (lane 29) no product bands were seen, but when the concentration was increased (lane 30) a strong N-1 product band was again It was noted that unlike the adenine experiment (lanes 17 and 18), only a weak single base extension product was seen at the higher concentration, while almost none was seen at the lower concentration.

As seen in previous situations KF exo$^+$ was able to give full length and N-1 bands. As discussed, this full-length band is a result of the exonuclease activity of KF exo$^+$. The N-1 bands from all three polymerases are again a result of a primer misalignment mechanism, which is extremely favorable in this particular situation. The initial acetal-dC:T mismatch provides many of the same structural challenges seen with the other types of mismatches. The adduct ring of the acetal--dC creates some steric hindrance, while also inducing some destabilizing local distortion. By pinching out the lesion, the primer thymine is able to be paired with the 5' template adenine, forming a canonical Watson-Crick base pair. This removes any conformational changes and local distortion, allowing all polymerases to successfully synthesize a N-1 product, regardless of their bypass activity.

This also explains the lack of the strong single base product seen in the RT lanes when compared to the extension with the adenine primer. Previously, the RT was unable to efficiently achieve bypass, resulting in some truncated products, but in the case of the thymine no bypass is necessary as the base pair is no longer a mismatch.

To summarize this experiment, no significant changes were noted when the acetal-dC lesion was properly matched with a guanine base. When the lesion was mismatched with either an adenine or thymine a dominant N-1 deletion product resulting from a primer misalignment and successive lesion bypass synthesis was formed. When the lesion was mismatched with a cytosine, KF exo$^+$ was able to achieve low levels of extension due to its exonuclease activity, but KF exo$^-$ and RT were unable to achieve any extension due to the structural perturbations being too great to bypass even after primer misalignment.

From these replication experiments it was seen that the lesion only has a significant effect when it is paired with a mismatched base, resulting in either deletion products or no product at all. All of the experiments ran up to this point utilized a template that contained only one lesion site; however, we know that the residual 5-fC is commonly found in CpG islands and promoter regions of genes. Due to this, it is quite possible that there are multiple neighboring 5-fC that are susceptible to attack from the nucleophilic environmental toxins, leading to more lesions and possibly an increase in the effects on the duplex. In order to investigate the effects that multiple lesions could have on replication, a tandem lesion template was synthesized, with two acetal-dC sites directly neighboring each other (#4). The same replication experiments as done previously were then carried out using this tandem template.

Tandem Lesion – 13mer primer. The first replication experiment ran utilized the shorter 13mer primer, landing in the middle of the template, upstream from the lesion sites. This allowed us to investigate whether or not the additional lesion would have any

effects on the processivity of the enzyme. The results for this experiment are shown below in Figure 3-12.

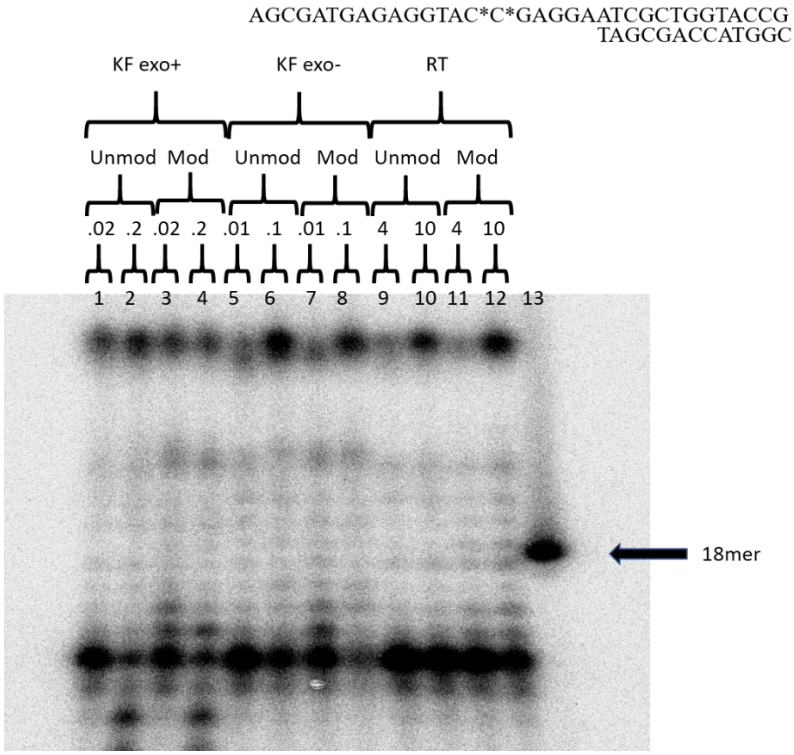


Figure 3-12 Extension of a 13mer Primer Against a Tandem Lesion Template. Lanes 1, 2, 5, 6, 9 and 10: control reactions using 0.06 μM of an unmodified template, 100 mM dNTP and either 0.02 or 0.2 U KF exo$^+$ (1, 2), 0.01 or 0.1 U KF exo$^-$ (5, 6), or 4 or 10 U RT (9, 10). Lanes 3, 4, 7, 8, 11 and 12: reactions using 0.06 μM of a modified template, 100 mM dNTP and either .02 or 0.2 U KF exo$^+$ (3, 4), 0.01 or 0.1 U KF exo$^-$ (7, 8), or 4 or 10 U RT (11, 12). Lane 13: An 18 nucleotide marker.

The results of this experiment are highly similar to those seen when using only a single lesion template. The first four lanes of this gel electrophoresis utilized KF exo$^+$ as a polymerase. The reactions in lanes 1 and 2 were carried out using an unmodified template and in both of these lanes full length product bands are seen. The following two

lanes (3 and 4) had reactions using a modified template, containing two neighboring acetal-dC lesions. They also both resulted in bands corresponding with the full-length extension bands seen in the control reactions (lanes 1 and 2). Like the single lesion reactions, there are no product bands that align with the 18mer marker.

The results seen when using KF exo$^-$ (lanes 5-8) and when using RT (lanes 9-12) were very similar. Both polymerases gave full length product bands when using the unmodified control template (lanes 5,6 and 9,10) Also similar to previous results, the lower concentration polymerase reactions gave slightly weaker product bands than the higher concentration lanes. When using a modified template (lanes 7,8 and 11,12) the extension ability was retained, resulting in full length products for all reactions. There were no notable differences between the intensity of the control reactions and that of the lesion containing reactions.

Overall it was seen that when using a short primer that anneals upstream of the acetal-dC lesion, all polymerases are able to move past the lesion to complete the full extension of the primer. The rationale for this is very similar to that used when replicating the single lesion template. To review, the main enzyme-DNA interactions that are important to replication are non-specific interactions with the phosphate backbone, interactions with the incoming dNTP and a stacking interaction with the template base. The addition of a second lesion still does not affect any of the interactions with the backbone or the incoming dNTP. As for the template side, it was seen previously that the presence of the adduct does not alter the position of the main cytosine ring enough to be detrimental to its stacking with the enzyme. If considered as isolated modified nucleotides, it would follow that neither of them is significantly displaced,

leaving the necessary stacking interactions intact. As a result, a full-length product is able to be synthesized.

Tandem Lesion – 18mer primer. In order to fully compare the tandem and the single lesion it was also necessary to carry out reactions that utilized the longer 18mer primer, annealing one base upstream from the lesion site. Similar to previously, this allows us to see whether or not the close proximity of the lesions to the initial binding site will have any effect on the ability of the polymerase to extend the sequence. The results for this experiment are shown below in Figure 3-13.

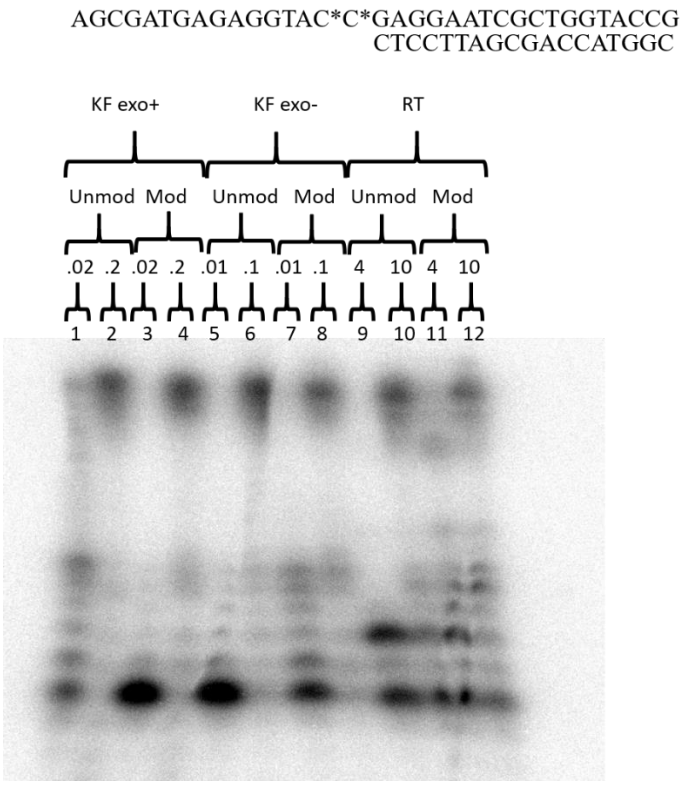


Figure 3-13 Extension of an 18mer Primer Against a Tandem Lesion Template. Lanes 1, 2, 5, 6, 9 and 10: control reactions using 0.06 μM of an unmodified template, 100 mM dNTP and either 0.02 or 0.2 U KF exo$^+$ (1, 2), 0.01 or 0.1 U KF exo$^-$ (5, 6), or 4 or 10 U RT (9, 10). Lanes 3, 4, 7, 8, 11 and 12: reactions using 0.06 μM of a modified template, 100 mM dNTP and either .02 or 0.2 U KF exo$^+$ (3, 4), 0.01 or 0.1 U KF exo$^-$ (7, 8), or 4 or 10 U RT (11, 12).

The first set of lanes (1-4) utilized KF exo$^+$ and were seen to give full length products when using both a modified and unmodified primer. The same result was seen for KF exo$^-$ and RT as well, giving full length products for both unmodified and modified sequences. There were some slight variations in efficiency and band strength in this particular experiment, but as mentioned previously we cannot draw solid conclusions from this due to the qualitative nature of this analysis method. On a molecular level this extension is happening the same way it did for the 13mer primer with the tandem lesion

and for the 13mer and 18mer primers with only the single lesion. The size and position of the acetal adduct allowed it to go unnoticed and not cause any hindrance to the polymerase.

Tandem Lesion – 18mer primer – Single base. We have now seen that a tandem lesion does not affect the ability of polymerase to synthesize a full-length product; however, it is still necessary to assess the fidelity of this full-length product. In order to do this the same 18mer primer was used, but instead of using dNTP, each reaction only contains one of the four dNTPs, allowing us to see if the correct guanine is incorporated, or some other base. The results of this experiment are shown below in Figure 3-14.

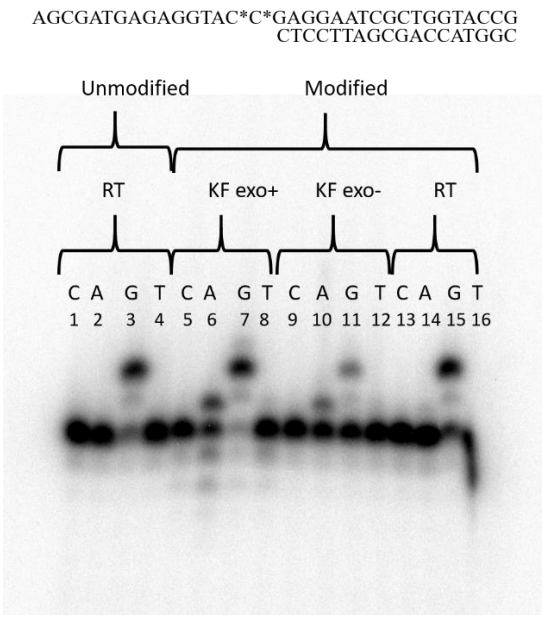


Figure 3-14 Single Nucleotide Extension of an 18mer Primer Against a Tandem Lesion Template. Lanes 1, 2, 3 and 4: control reactions using 0.06 μM of an unmodified template, 4 U RT and 10 mM of either dCTP (1), dATP (2), dGTP (3), or dTTP (4). Lanes 5, 9 and 13: reactions using 0.06 μM of a modified template, 10 mM dCTP and either 0.02 U KF exo$^+$ (5), 0.01 U KF exo$^-$ (9) or 4 U RT (13). Lanes 6, 10 and 14: reactions using 0.06 μM of a modified template, 10 mM dATP and either 0.02 U KF exo$^+$ (6), 0.01 U KF exo$^-$ (10) or 4 U RT (14). Lanes 7, 11 and 15: reactions using 0.06 μM of a modified template, 10 mM dGTP and either 0.02 U KF exo$^+$ (7), 0.01 U KF exo$^-$ (11) or 4 U RT (15). Lanes 8, 12 and 16: reactions using 0.06 μM of a modified template, 10 mM dTTP and either 0.02 U KF exo$^+$ (8), 0.01 U KF exo$^-$ (12) or 4 U RT (16).

The first four lanes on this gel are control reactions, utilizing an unmodified primer and RT as the polymerase. It was seen that there was no extension when a reaction contained dCTP, dATP or dTTP (lanes 1,2 and 4) and there is a clear band indicating a two base extension when the reaction utilized dGTP (lane 3). This two-base extension occurs because the tandem lesion template contains two cytosines directly neighboring each other allowing a guanine to be incorporated across from both of them, resulting in the two base extension product. The next four lanes (5-8) utilized KF exo$^+$ and it can be seen that while guanine was correctly incorporated, there was also

erroneous incorporation of adenine. This can be seen again in the reactions using KF exo$^-$ (lanes 9-12), but in this set the band representing incorporation of adenine is much weaker than was seen in the KF exo$^+$ experiments. The last set of lanes (13-16) utilized RT and it was seen that while guanine was still incorporated, there was no incorporation of adenine.

The incorporation of guanine by all three polymerases shows that despite the presence of the lesion the fidelity of replication is able to be maintained. The presence of adenine incorporation is a result of a lesion bypass mechanism. This can be deduced from the correlation of the intensities of the bands to the bypass ability of the particular polymerase. As discussed previously, KF exo$^+$ has strong bypass activity, KF exo$^-$ has moderate activity and RT has little to no activity, correlating with strong, medium and no incorporation of adenine respectively. Previous studies have shown that during bypass synthesis if the polymerase is unable to properly read a base it is most likely to insert an adenine across from the lesion. This is a form of error-prone lesion bypass synthesis.[53] In a reaction environment containing all dNTPs, it is possible that a competition between guanine and adenine would see guanine preferentially incorporated, but in this particular case only dATP is present, allowing its incorporation through this bypass mechanism.

Tandem Lesion – 19mer primer. In order to complete a full analysis of the tandem lesion template a set of reactions was ran utilizing the 19mer primers resulting in mismatched base pairs. This would allow us to see if the presence of the additional lesion changes the ability of the DNA to accommodate various mismatches and could possibly even destabilize the primer misalignment mechanism described previously. This is especially relevant, as it was seen in the previous single base experiment that adenine

can be incorrectly incorporated across from the lesion, presenting a possibility for extension from a mismatched base pair. The results of this experiment are shown below in Figure 3-15.

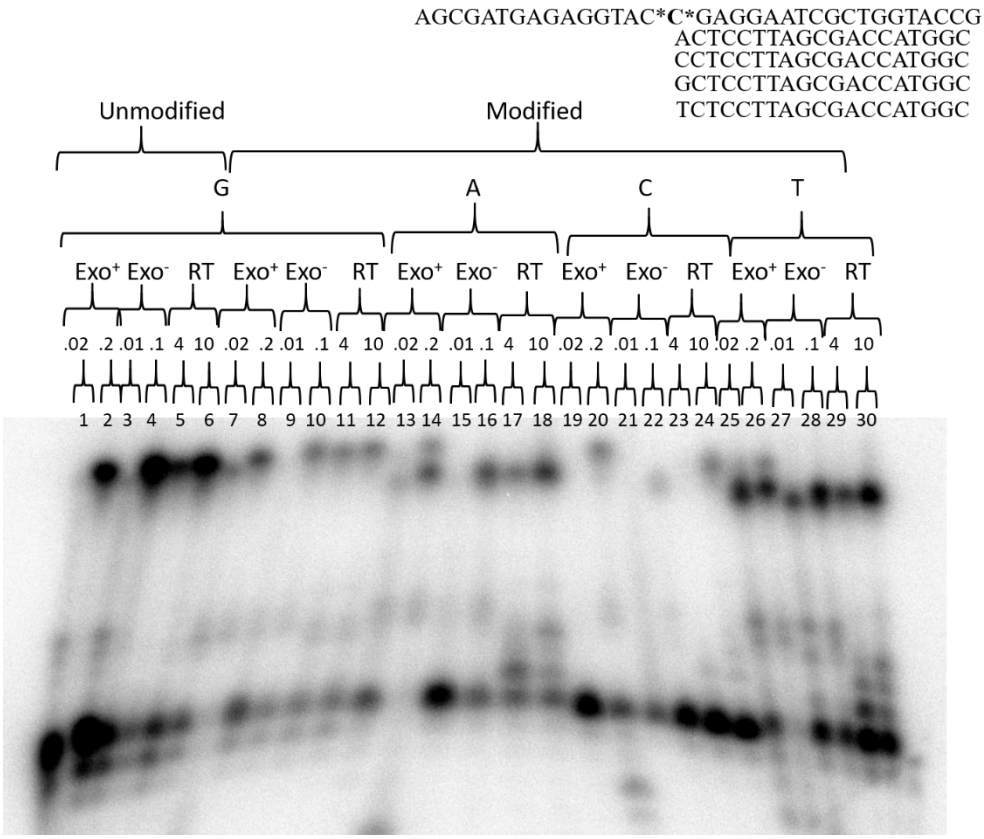


Figure 3-15 Extension of 19mer Primers Against a Tandem Lesion Template with Varying Nucleotides Paired with the acetal-dC Lesion. Lanes 1-6: control reactions using 0.06 μM of an unmodified template, 100 mM dNTP, 0.05 μM of a 19mer primer with a 3' G and either 0.02 or 0.2 U KF exo$^+$ (1, 2), 0.01 or 0.1 U KF exo$^-$ (3, 4), or 4 or 10 U RT (5, 6). Lanes 7-12: reactions using 0.06 μM of a modified template, 100 mM dNTP, 0.05 μM of a 19mer primer with a 3' G and either 0.02 or 0.2 U KF exo$^+$ (7, 8), 0.01 or 0.1 U KF exo$^-$ (9, 10), or 4 or 10 U RT (11, 12). Lanes 13-18: reactions using 0.06 μM of a modified template, 100 mM dNTP, 0.05 μM of a 19mer primer with a 3' A and either 0.02 or 0.2 U KF exo$^+$ (13, 14), 0.01 or 0.1 U KF exo$^-$ (15, 16), or 4 or 10 U RT (17, 18). Lanes 19-24: reactions using 0.06 μM of a modified template, 100 mM dNTP, 0.05 μM of a 19mer primer with a 3' C and either 0.02 or 0.2 U KF exo$^+$ (19, 20), 0.01 or 0.1 U KF exo$^-$ (21, 22), or 4 or 10 U RT (23, 24). Lanes 25-30: reactions using 0.06 μM of a modified template, 100 mM dNTP, 0.05 μM of a 19mer primer with a 3' T and either 0.02 or 0.2 U KF exo$^+$ (25, 26), 0.01 or 0.1 U KF exo$^-$ (27, 28), or 4 or 10 U RT (29, 30).

The first set of lanes in this experiment (1-6) were ran using an unmodified template and a primer ending with a guanine, giving a correct pairing with the template cytosine. The formation of full length product was observed with all polymerases, as expected. The next set of lanes (7-12) utilized a modified template containing the two neighboring lesions, along with the same primer ending in a guanine. Full length product was still seen for all polymerases; however, the efficiency was much lower than that of the control reactions. This could be due to some decrease in the efficiency of binding to the lesion sites; however, this conclusion would not be consistent with the previous experiments where changes in efficiency were not seen. This highlights the qualitative nature of these experiments, and the need for further studies to draw clear conclusions regarding changes in efficiency.

The next set of lanes (13-18) utilized a modified lesion containing template and a primer placing an adenine across from the first lesion site. Similar to the single lesion reactions, when reacting with KF exo$^+$ both full length and N-1 product bands are seen. The full-length product can be considered a result of the exonuclease activity of the enzyme. Upon detection of the mismatch the polymerase will displace the primer into the exonuclease domain, cleave off the mismatched adenine and then continue extension to obtain a full-length product. KF exo$^+$ as well as KF exo$^-$ and RT are also seen to make N-1 products. This type of deletion typically results from a primer misalignment type mechanism as previously described. In this case though, it would be assumed that both lesions would be pinched out from the template, in order to make a tolerable A:A pairing as seen previously; however, this would produce a N-2 product, which is not seen.

It is possible that some sort of distortion makes the second lesion able to pair with adenine and be bypassed, or that for some reason the second lesion is unable to be excluded along with the first one due to some type of steric interaction. Previous studies have also shown that the propensity for forming single or double nucleotide deletions in the presence of a lesion can be highly sequence dependent with regard to the bases 5' to the lesion in the template.[56] It is difficult to draw conclusions regarding what exactly is happening without further structural information and replication experiments. Despite this, it is clear that the presence of a mismatch adenine at the first lesion site results in a single deletion product extended from a acetal-dC:A mismatch.

It is also important to note that this helps to confirm our previous theory of competition between guanine and adenine in the previous single base experiment. If G and A were equally incorporated across from the lesion, then the reactions using G would give full length product, while the reactions with A would give N-1 products. When looking at the 18mer primer experiments only full-length products were seen, indicating that only G was incorporated across from the lesion when all four dNTPs are available.

The next set of lanes (19-24) utilized the modified tandem lesion template and a primer placing a cytosine across from the first lesion site. When using KF exo$^+$ at a high concentration, a very faint band was seen for the full-length extension product. As seen previously, this band results from the exonuclease activity of KF exo$^+$. It was noted that KF exo$^+$ lacked the presence of any deletion type products. In the single lesion experiments the C:C mismatch caused too much distortion to be able to overcome through bypass synthesis, and that appears to be consistent here. The polymerase cannot accommodate this type of mismatch, therefore halting synthesis.

Previously this pattern of no deletion products was carried over to KF exo$^-$ and RT as well; however, in this particular experiment an extremely faint band was seen for N-1 with KF exo$^-$ and also a faint band for a full-length product when RT was used. It is possible that KF exo$^-$ is able to overcome the distortion of the misaligned acetal-dC:C mismatch a very small percentage of the time, resulting in the faint single deletion product. The presence of a full-length product from RT is slightly more puzzling and would require further structural studies to completely understand.

The last set of lanes (25-30) utilized the modified tandem lesion template and a primer that places a thymine across from the first lesion site. It was again seen that KF exo$^+$ was able to give full length products due to the exonuclease activity; however, this is a very minor product compared to the deletion products that are formed by all three polymerases. The formation of this follows a primer misalignment mechanism as seen in the single lesion studies. Previously the primer was able to misalign by one base to achieve a canonical base pair with adenine. In this situation the primer would have to misalign by two bases, however as seen with the adenine experiments this double deletion does not readily occur, resulting in the formation of an acetal-dC:T mismatch. This pairing does not create much structural distortion, with the acetal-dC not having to undergo rotation in order to form hydrogen bonds with the thymine (Figure 3-16). This allows this pair to be easily bypassed by the polymerase, resulting in product intensities near that of the control.

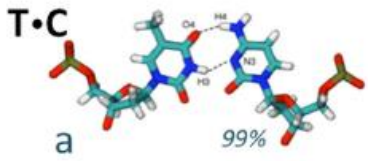


Figure 3-16 Modified Hydrogen Bonding of a T:C Mismatched Base Pair.[52]

To summarize this experiment, no significant changes were noted when a tandem acetal-dC lesion template was properly matched with a guanine base. When the lesion was mismatched with either an adenine or thymine, a dominant N-1 deletion product resulting from a primer misalignment and successive lesion bypass synthesis was formed. When the lesion was mismatched with a cytosine, KF exo$^+$ was able to achieve low levels of extension due to its exonuclease activity, KF exo$^-$ and RT produced N-1 and full-length products respectively. By incorporating a second lesion the complexity of the overall system increased, making it increasingly difficult to completely understand what is occurring on a molecular level. Additional structural as well as replication studies can be performed to shed more light on the specific mechanisms involved in this process.

Conclusion

It was seen that when replicating a DNA template containing a single acetal-dC lesion the polymerase is able to accommodate the lesion and complete a full extension of the primer. The result was the same when using both the 13mer primer that annealed several bases downstream of the lesion and the 18mer primer that anneals only one base before the primer. This was due to the position of the adduct; its placement in the major groove does not disrupt any of the crucial interactions between the DNA and the

polymerase. It was also seen that the presence of a single lesion does not affect the fidelity of this replication, allowing a normal guanine to be incorporated across from the modified cytosine.

In the situation that the lesion is mismatched with a nucleotide it is possible to still form full length products in the presence of exonuclease activity, however without this activity there is dominant formation of single base deletion products as well complete halting of replication in the case of an acetal-dC:C mismatch. These errors are potentially formed from a primer misalignment mechanism followed by lesion bypass. Overall these experiments indicate that a single acetal-dC lesion has a relatively minimal effect on the DNA but does have the potential to be both mutagenic and cytotoxic when encountered in the context of a mismatched base pair.

When performing replication using a template containing two tandem acetal-dC lesions the polymerase is still able to synthesize full length products for both the 18mer and 13mer primers. The fidelity of this replication is somewhat compromised, with both the KF polymerases inserting the correct guanine as well as an incorrect adenine. This was thought to be a result of a bypass mechanism that occurs when dATP is the only nucleotide present. Regardless of the insertion dynamics this does show that the presence of multiple lesions is potentially mutagenic.

In the case of a mismatched base paired with the tandem lesion the results are similar to that of the single lesion. There is a dominant formation of single base deletion products along with some minor full length products and also a low percentage of halting replication. The compounded effect of two lesions makes it increasingly difficult to

obtain a clear picture of the molecular interactions, however, it is still clear that the presence of multiple lesions can again be mutagenic or cytotoxic. In the future a single base extension can be ran from each of the 19mer mismatch primers in order to get a better picture of what exactly is going on.

Overall, it was seen that the acetal-dC lesion does indeed have the potential to be both mutagenic and cytotoxic when in the genome. This study only went up to two lesions, but as mentioned the 5-fC is typically found clustered in CpG islands. This could mean that there could be several more acetal-dC lesions contained within a short sequence, further compounding the effects. These lesions could also appear in different patterns, sometimes directly neighboring each other, or with normal nucleotides between each individual lesion site. All of these factors could impact the effect that the lesion has on the polymerase.

In regards to lesions formed by other environmental toxins, we have seen that the effects are highly dependent on structure, meaning that different lesions could have different effects. Smaller lesions such as those formed by the methoxyamine and hydroxyamine might not cause as much steric hindrance with the backbone, making the primer misalignment mechanism unnecessary. Larger lesion such as those formed by bulky molecules like phenylhydrazine may have even more drastic effects. The bulky ring adduct may not fit into the major groove or may cause so much local distortion that the polymerase is unable to bind to the DNA. It is clear that DNA lesions formed by nucleophilic environmental toxins can potentially have a significant impact on cell processes.

References

1. Jackson, A. L.; Loeb, L. A., The contribution of endogenous sources of DNA damage to the multiple mutations in cancer. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **2001,** *477* (1), 7-21.
2. Chance, B.; Sies, H.; Boveris, A., Hydroperoxide metabolism in mammalian organs. *Physiological Reviews* **1979,** *59* (3), 527-605.
3. Loft, S.; Poulsen, H. E., Cancer risk and oxidative DNA damage in man. *Journal of Molecular Medicine* **1996,** *74* (6), 297-312.
4. Cheng, K. C.; Cahill, D. S.; Kasai, H.; Nishimura, S.; Loeb, L. A., 8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G----T and A----C substitutions. *Journal of Biological Chemistry* **1992,** *267* (1), 166-172.
5. Rydberg, B.; Lindahl, T., Nonenzymatic methylation of DNA by the intracellular methyl group donor S-adenosyl-L-methionine is a potentially mutagenic reaction. *The EMBO Journal* **1982,** *1* (2), 211-216.
6. Barbarella, G.; Tugnoli, V.; Zambianchi, M., Imidazole Ring Opening of 7-Methylguanosine at Physiological pH. *Nucleosides and Nucleotides* **1991,** *10* (8), 1759-1769.
7. Huff, A. C.; D Topal, M., *DNA damage at thymine N-3 abolishes base-pairing capacity during DNA synthesis*. 1987; Vol. 262, p 12843-50.
8. Rastogi, R. P.; Richa; Kumar, A.; Tyagi, M. B.; Sinha, R. P., Molecular Mechanisms of Ultraviolet Radiation-Induced DNA Damage and Repair. *Journal of Nucleic Acids* **2010,** *2010*.
9. Goodsell, D. S., The Molecular Perspective: Ultraviolet Light and Pyrimidine Dimers. *The Oncologist* **2001,** *6* (3), 298-299.
10. Vink, A. A.; Roza, L., Biological consequences of cyclobutane pyrimidine dimers. *Journal of Photochemistry and Photobiology B: Biology* **2001,** *65* (2), 101-104.
11. Liu, Y.; Prasad, R.; Beard, W. A.; Kedar, P. S.; Hou, E. W.; Shock, D. D.; Wilson, S. H., Coordination of Steps in Single-nucleotide Base Excision Repair Mediated by Apurinic/Apyrimidinic Endonuclease 1 and DNA Polymerase β. *The Journal of biological chemistry* **2007,** *282* (18), 13532-13541.
12. Yun-Jeong, K.; David, M. W., III, Overview of Base Excision Repair Biochemistry. *Current Molecular Pharmacology* **2012,** *5* (1), 3-13.
13. de Laat, W. L.; Jaspers, N. G. J.; Hoeijmakers, J. H. J., Molecular mechanism of nucleotide excision repair. *Genes & Development* **1999,** *13* (7), 768-785.
14. Atamna, H.; Cheung, I.; Ames, B. N., A method for detecting abasic sites in living cells: Age-dependent changes in base excision repair. *Proceedings of the National Academy of Sciences of the United States of America* **2000,** *97* (2), 686-691.
15. Lehmann, A. R.; McGibbon, D.; Stefanini, M., Xeroderma pigmentosum. *Orphanet Journal of Rare Diseases* **2011,** *6*, 70-70.
16. Derheimer, F. A.; Kastan, M. B., Multiple roles of atm in monitoring and maintaining dna integrity. *FEBS letters* **2010,** *584* (17), 3675-3681.

17. Ethylene Glycol. In *Ullmann's Encyclopedia of Industrial Chemistry*.
18. Hydrazine. In *Ullmann's Encyclopedia of Industrial Chemistry*.
19. Felsenfeld, G., A Brief History of Epigenetics. *Cold Spring Harbor Perspectives in Biology* **2014,** *6* (1), a018200.
20. Avery, O. T.; MacLeod, C. M.; McCarty, M., STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES. *The Journal of Experimental Medicine* **1944,** *79* (2), 137.
21. Hershey, A. D.; Chase, M., INDEPENDENT FUNCTIONS OF VIRAL PROTEIN AND NUCLEIC ACID IN GROWTH OF BACTERIOPHAGE. *The Journal of General Physiology* **1952,** *36* (1), 39.
22. Watson, J. D.; Crick, F. H. C., Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953,** *171*, 737.
23. Holliday, R.; Pugh, J. E., DNA modification mechanisms and gene activity during development. *Science* **1975,** *187* (4173), 226.
24. Mayer, W.; Niveleau, A.; Walter, J.; Fundele, R.; Haaf, T., Demethylation of the zygotic paternal genome. *Nature* **2000,** *403*, 501.
25. Okano, M.; Bell, D. W.; Haber, D. A.; Li, E., DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell* **1999,** *99* (3), 247-257.
26. Smallwood, S. A.; Kelsey, G., De novo DNA methylation: a germ cell perspective. *Trends in Genetics* **2012,** *28* (1), 33-42.
27. Saadeh, H.; Schulz, R., *Protection of CpG islands against de novo DNA methylation during oogenesis is associated with the recognition site of E2f1 and E2f2*. 2014; Vol. 7, p 26.
28. Li, E.; Zhang, Y., DNA Methylation in Mammals. *Cold Spring Harbor Perspectives in Biology* **2014,** *6* (5), a019133.
29. Blackledge, N.; Thomson, J.; Skane, P., *CpG binding proteins*. 2013; Vol. 5, p a018648.
30. Mohandas, T.; Sparkes, R. S.; Shapiro, L. J., Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. *Science* **1981,** *211* (4480), 393.
31. Watt, F.; Molloy, P. L., Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes & Development* **1988,** *2* (9), 1136-1143.
32. Wu, S. C.; Zhang, Y., Active DNA demethylation: many roads lead to Rome. *Nature reviews. Molecular cell biology* **2010,** *11* (9), 607-620.
33. Wu, X.; Zhang, Y., TET-mediated active DNA demethylation: mechanism, function and beyond. *Nature Reviews Genetics* **2017,** *18*, 517.
34. Pastor, W. A.; Aravind, L.; Rao, A., TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nature reviews. Molecular cell biology* **2013,** *14* (6), 341-356.
35. Kohli, R. M.; Zhang, Y., TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **2013,** *502* (7472), 472-479.
36. Hu, L.; Li, Z.; Cheng, J.; Rao, Q.; Gong, W.; Liu, M.; Shi, Y. G.; Zhu, J.; Wang, P.; Xu, Y., Crystal Structure of TET2-DNA Complex: Insight into TET-Mediated 5mC Oxidation. *Cell* **2013,** *155* (7), 1545-1555.

37. Tamanaha, E.; Guan, S.; Marks, K.; Saleh, L., Distributive Processing by the Iron(II)/α-Ketoglutarate-Dependent Catalytic Domains of the TET Enzymes Is Consistent with Epigenetic Roles for Oxidized 5-Methylcytosine Bases. *Journal of the American Chemical Society* **2016,** *138* (30), 9345-9348.
38. Hu, L.; Lu, J.; Cheng, J.; Rao, Q.; Li, Z.; Hou, H.; Lou, Z.; Zhang, L.; Li, W.; Gong, W.; Liu, M.; Sun, C.; Yin, X.; Li, J.; Tan, X.; Wang, P.; Wang, Y.; Fang, D.; Cui, Q.; Yang, P.; He, C.; Jiang, H.; Luo, C.; Xu, Y., Structural insight into substrate preference for TET-mediated oxidation. *Nature* **2015,** *527*, 118.
39. Cortázar, D.; Kunz, C.; Selfridge, J.; Lettieri, T.; Saito, Y.; MacDougall, E.; Wirz, A.; Schuermann, D.; Jacobs, A. L.; Siegrist, F.; Steinacher, R.; Jiricny, J.; Bird, A.; Schär, P., Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* **2011,** *470*, 419.
40. Pidugu, L. S.; Flowers, J. W.; Coey, C. T.; Pozharski, E.; Greenberg, M. M.; Drohat, A. C., Structural Basis for Excision of 5-Formylcytosine by Thymine DNA Glycosylase. *Biochemistry* **2016,** *55* (45), 6205-6208.
41. Bennett, M. T.; Rodgers, M. T.; Hebert, A. S.; Ruslander, L. E.; Eisele, L.; Drohat, A. C., Specificity of Human Thymine DNA Glycosylase Depends on N-Glycosidic Bond Stability. *Journal of the American Chemical Society* **2006,** *128* (38), 12510-12519.
42. Bachman, M.; Uribe-Lewis, S.; Yang, X.; Williams, M.; Murrell, A.; Balasubramanian, S., 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nature chemistry* **2014,** *6* (12), 1049-1055.
43. Raiber, E.-A.; Murat, P.; Chirgadze, D. Y.; Beraldi, D.; Luisi, B. F.; Balasubramanian, S., 5-Formylcytosine alters the structure of the DNA double helix. *Nature Structural &Amp; Molecular Biology* **2014,** *22*, 44.
44. Li, F.; Zhang, Y.; Bai, J.; Greenberg, M. M.; Xi, Z.; Zhou, C., 5-Formylcytosine Yields DNA–Protein Cross-Links in Nucleosome Core Particles. *Journal of the American Chemical Society* **2017,** *139* (31), 10617-10620.
45. Neri, F.; Incarnato, D.; Krepelova, A.; Rapelli, S.; Anselmi, F.; Parlato, C.; Medana, C.; Dal Bello, F.; Oliviero, S., Single-Base Resolution Analysis of 5-Formyl and 5-Carboxyl Cytosine Reveals Promoter DNA Methylation Dynamics. *Cell Reports* **2015,** *10* (5), 674-683.
46. Liu, C.; Wang, Y.; Yang, W.; Wu, F.; Zeng, W.; Chen, Z.; Huang, J.; Zou, G.; Zhang, X.; Wang, S.; Weng, X.; Wu, Z.; Zhou, Y.; Zhou, X., Fluorogenic labeling and single-base resolution analysis of 5-formylcytosine in DNA †Electronic supplementary information (ESI) available. See DOI: 10.1039/c7sc03685j Click here for additional data file. *Chemical Science* **2017,** *8* (11), 7443-7447.
47. Kypr, J.; Kejnovská, I.; Renčiuk, D.; Vorlíčková, M., Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Research* **2009,** *37* (6), 1713-1725.
48. Doublié, S.; Tabor, S.; Long, A. M.; Richardson, C. C.; Ellenberger, T., Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature* **1998,** *391*, 251.
49. Beard, W. A.; Wilson, S. H., Structure and Mechanism of DNA Polymerase β. *Biochemistry* **2014,** *53* (17), 2768-2780.

50. Golosov, A. A.; Warren, J. J.; Beese, L. S.; Karplus, M., The Mechanism of the Translocation Step in DNA Replication by DNA Polymerase I: A Computer Simulation Analysis. *Structure* **2010,** *18* (1), 83-93.
51. Morales, J. C.; Kool, E. T., Minor Groove Interactions between Polymerase and DNA: More Essential to Replication than Watson–Crick Hydrogen Bonds? *Journal of the American Chemical Society* **1999,** *121* (10), 2323-2324.
52. Rossetti, G.; Dans, P. D.; Gomez-Pinto, I.; Ivani, I.; Gonzalez, C.; Orozco, M., The structural impact of DNA mismatches. *Nucleic Acids Research* **2015,** *43* (8), 4309-4321.
53. Laverty, D. J.; Averill, A. M.; Doublié, S.; Greenberg, M. M., The A-Rule and Deletion Formation During Abasic and Oxidized Abasic Site Bypass by DNA Polymerase θ. *ACS Chemical Biology* **2017,** *12* (6), 1584-1592.
54. Fazlieva, R.; Spittle, C. S.; Morrissey, D.; Hayashi, H.; Yan, H.; Matsumoto, Y., Proofreading exonuclease activity of human DNA polymerase δ and its effects on lesion-bypass DNA synthesis. *Nucleic Acids Research* **2009,** *37* (9), 2854-2866.
55. Küpfer, P. A.; Crey-Desbiolles, C.; Leumann, C. J., Trans-lesion synthesis and RNaseH activity by reverse transcriptases on a true abasic RNA template. *Nucleic Acids Research* **2007,** *35* (20), 6846-6853.
56. Shibutani, S.; Grollman, A. P., On the mechanism of frameshift (deletion) mutagenesis in vitro. *Journal of Biological Chemistry* **1993,** *268* (16), 11703-11710.

Appendix A

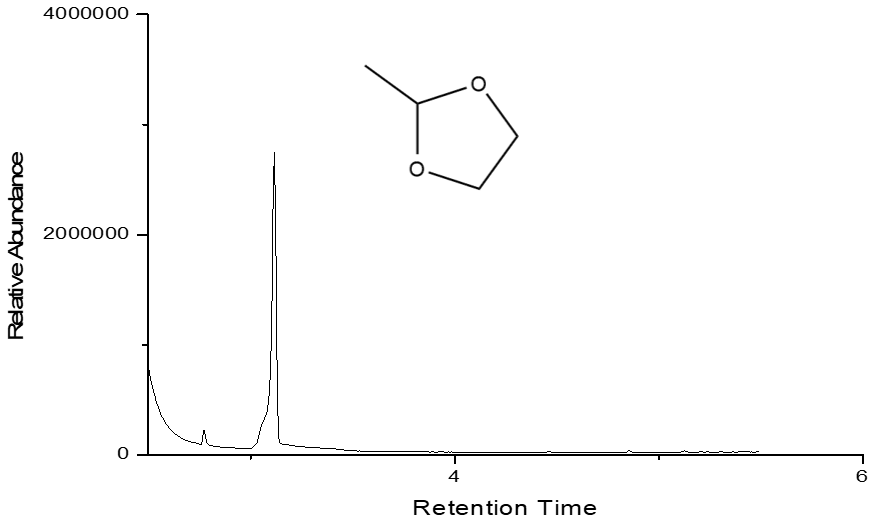


Figure A-1 Chromatogram of the Reaction of Ethylene Glycol and Acetaldehyde.

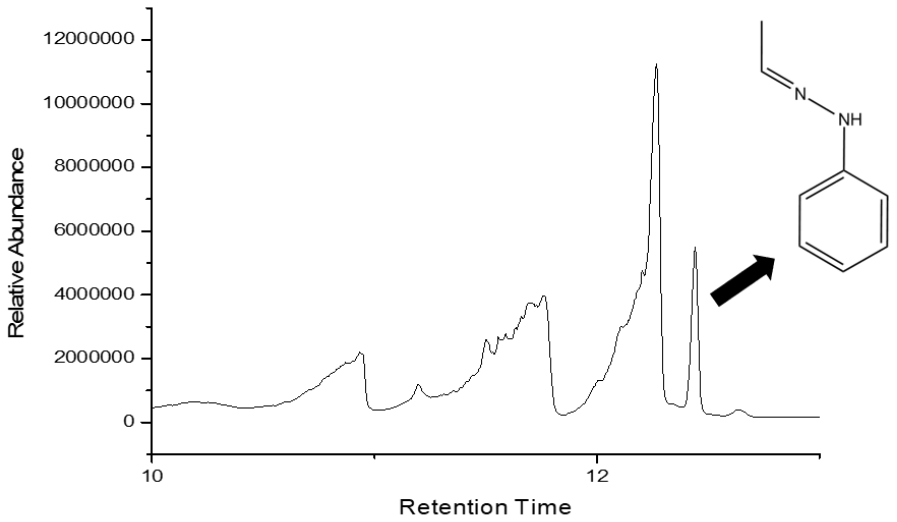
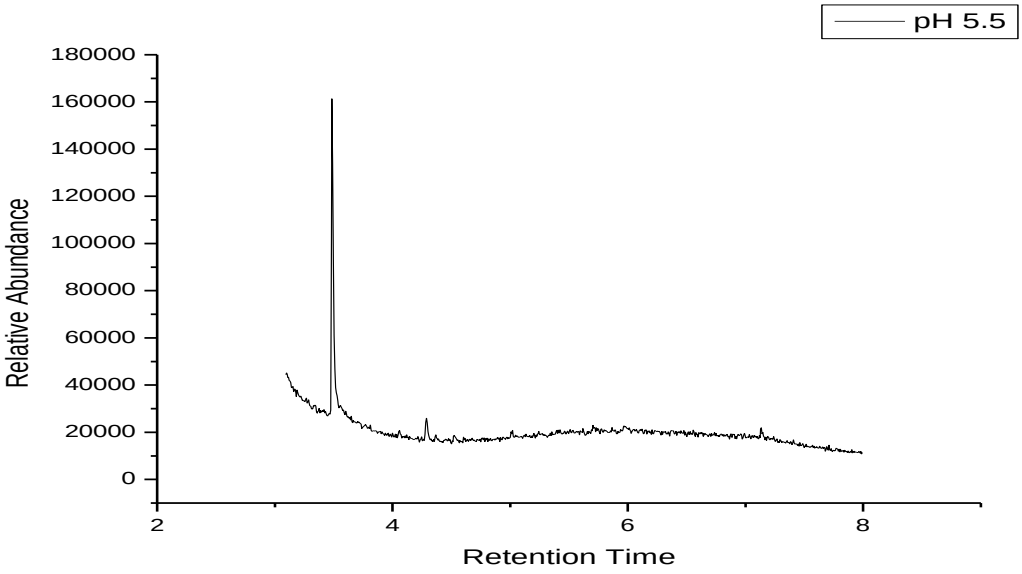


Figure A-2 Chromatogram of the Reaction of Phenylhydrazine and Acetaldehyde.



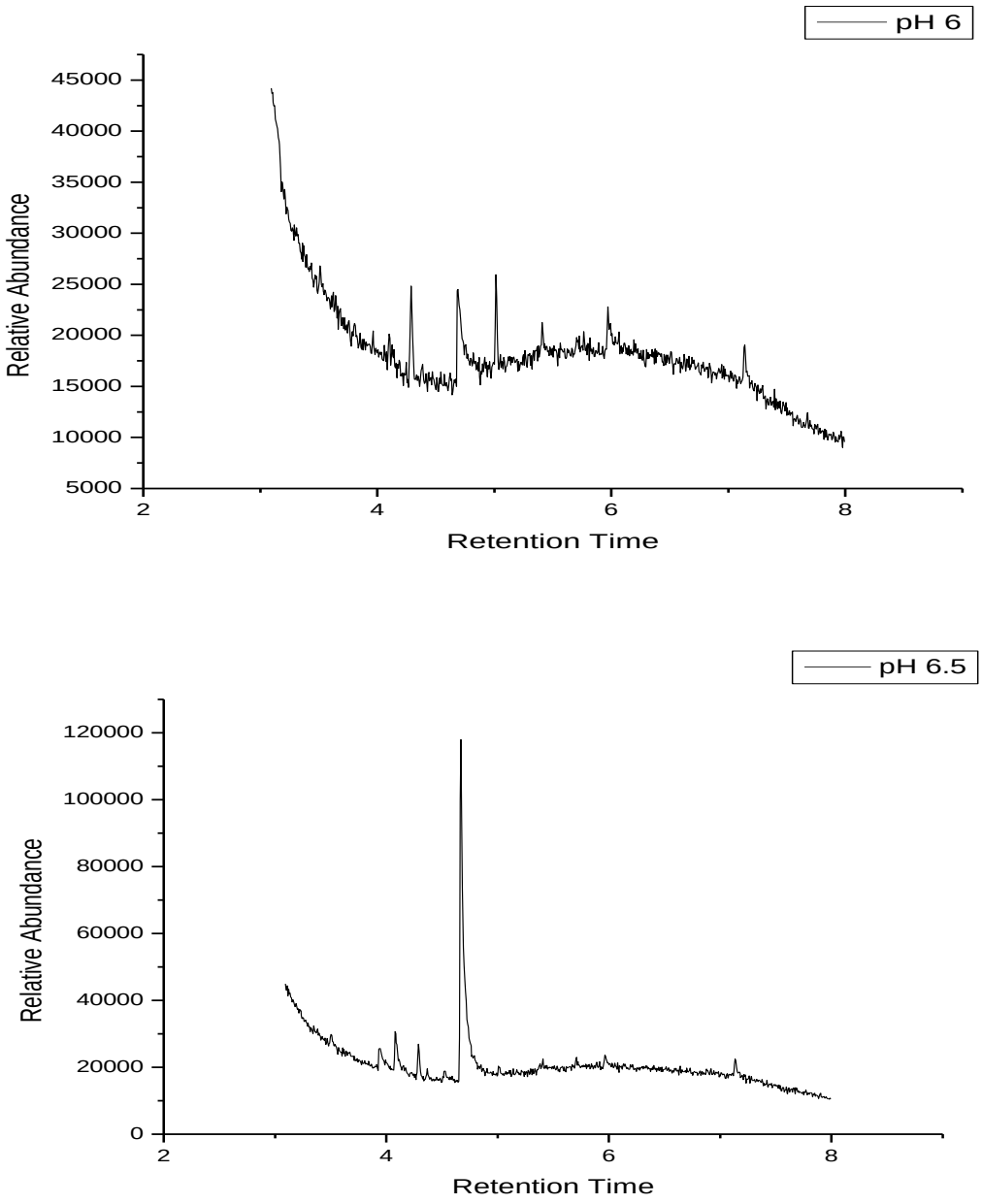


Figure A-3 Chromatograms of Reaction of 1,3-propane diol and Acetaldehyde at pH 5.5, 6 and 6.5.

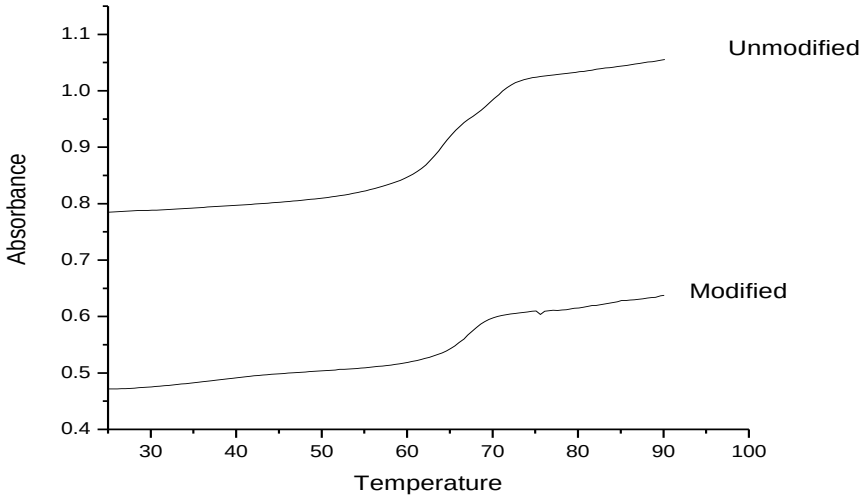


Figure A-4 UV Thermal Denaturation of an Unmodified and Modified 29mer DNA Duplex.

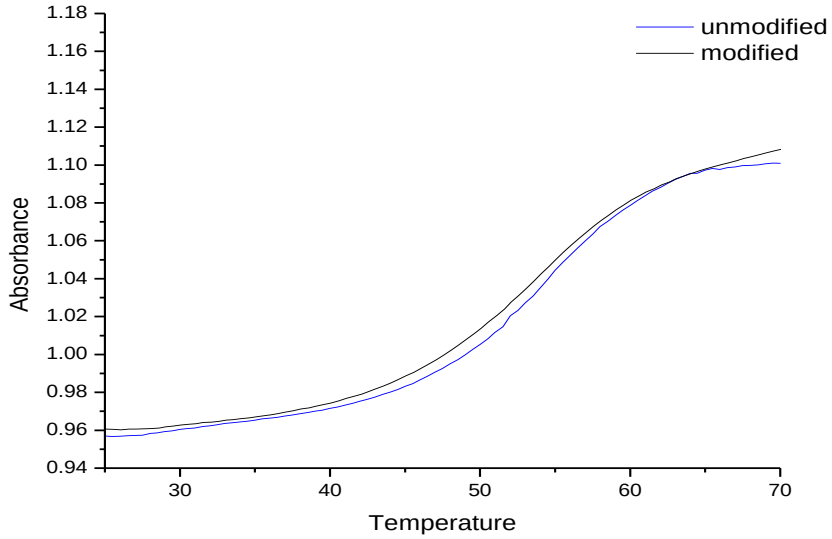
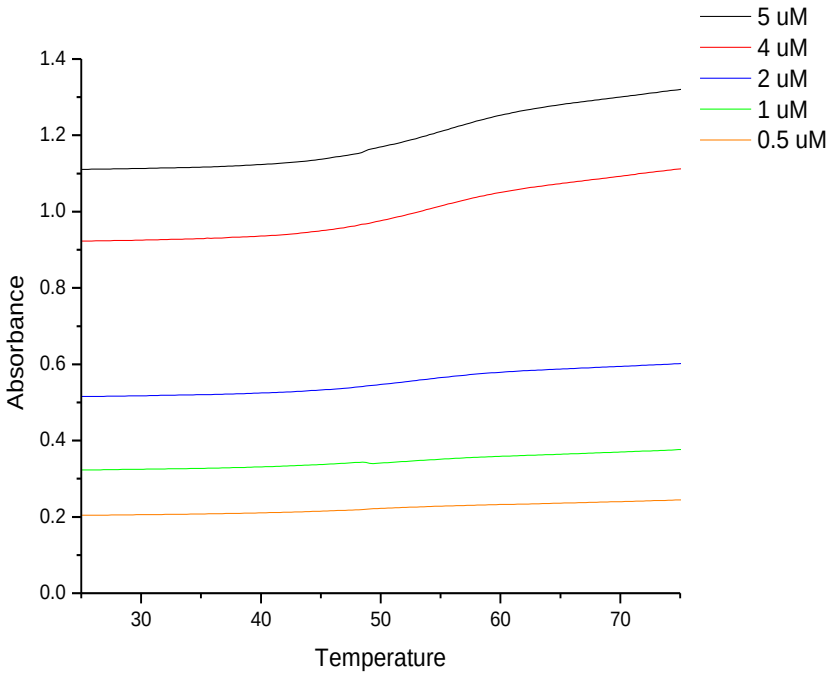


Figure A-5 UV Thermal Denaturation of a Modified and Unmodified 10mer DNA Duplex.


5 uM
4 uM
2 uM
1 uM
0.5 uM
Absorbance
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
30
40
50
60
70
Temperature

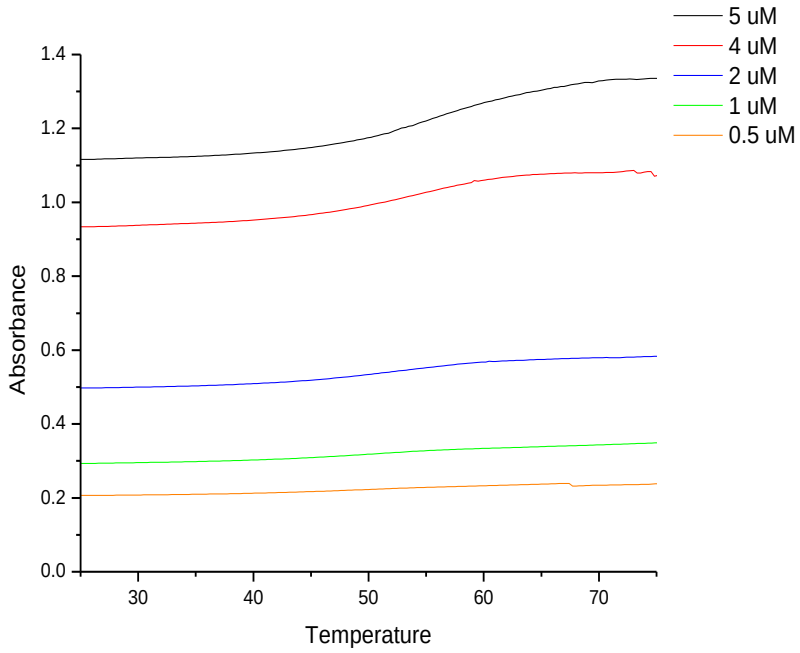


Figure A-6 Thermal Denaturation Curves at Varying Concentrations for an Unmodified (Top) and Modified (Bottom) 10mer DNA Duplex

	Unmod	Mod	ΔT
5 µM	55.47 ± 0.35	55.52 ± 0.35	+0.05
4 µM	54.49 ± 0.37	54.99 ± 0.72	+0.50
2 µM	53.09 ± 0.08	53.42 ± 0.10	+0.33
1 µM	51.35 ± 0.47	50.44 ± 0.54	-0.91
0.5 µM	49.07 ± 0.12	49.72 ± 0.42	+0.65

Table A-1 Average Melting Points of a Modified and Unmodified 10mer DNA Duplex at Varying Concentrations