Characterization of the 2,6-Diamino-4-hydroxy-N^5-(methyl)-formamidopyrimidine DNA Lesion

By

Stephanie N. Bamberger

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Chemistry

December 15th, 2018

Nashville, Tennessee

Approved:

Michael P. Stone, Ph.D.

Carmelo J. Rizzo, Ph.D.

Brian O. Bachmann, Ph.D.

Martin Egli, Ph.D.

Acknowledgements

I would like to thank all of the people who I have met throughout this journey for their help and support. None of this would have been possible without the guidance provided by individuals across numerous departments and career stages.

I would first like to thank my family and friends, who supported me unfalteringly throughout this process, demonstrating the great bounds patience and kindness can be stretched to. Thank you to my parents for all of the love, support, stability, and truthfulness you gave me throughout my life, and especially during this period. I could not have done it without you. Gratitude is also extended to the Writing Center at Vanderbilt University; your writing sessions and organized retreats were invaluable, as was your patient advice and quiet encouragement.

I would like to extend my thanks to Vanderbilt University for welcoming me to study and for providing me with opportunities I could only dream of. Thank you to the Biomolecular NMR Facility, Center for Structural Biology, the Office of Biomedical Research and Education Training, the Mass Spectrometry Core, and all the staff who keep these giants of science running. I especially thanks the National Institute of Health for funding and the opportunity to participate in the Training Program in Environmental Toxicology (NIH T32 ES007028)**.**

Thank you to my advisor Dr. Stone for the opportunity to work in his lab and for providing an environment in which I could grow into an independent scientist. Thank you to Dr. Rizzo for allowing me to complete a substantial amount of work in his lab, to the point where people became confused when I exited the elevator on a different floor. Thank you to Dr. Bachmann and Dr. Egli for being members of my Ph.D. committee along with Dr. Stone and Dr. Rizzo.

No words can express the depths of my gratitude to our collaborators. In particular I would like to thank Tracy Salyard-Johnson of the Rizzo lab for acting as my mentor and friend, Dr. Chanchal Malik of the Rizzo lab for his work synthesizing our samples, and Dr. Markus Voehler for his continued guidance and instruction. My thanks also go out to my wonderful undergraduate mentees, Hope Pan and Summer Brown; you are both amazingly talented and hard working. Thank you to all current and former Stone lab members. Thank you to Sandra Ford, for everything you did for me and for the Chemistry department.

Table of Contents **Page**

List of Tables

Page

List of Figures

Page

List of Schemes

Page

List of Abbreviations

DNA: deoxyribonucleic acid

dA: adenine

dG: guanine

dC: cytosine

dT: thymine

AFB_1 :aflatoxin B_1

7Me-dG: 7-methylguanine:

TLS: translesion synthesis

AFB_1-N7-dG :8,9-dihydro-8-(N7-guanyl)-9-hydroxyaflatoxin B_1

Fapy-dG: formamidopyrimidine

MeFapy-dG: 2,6-diamino-4-hydroxy-N^5-(methyl)-formamidopyrimidine

SAM: S-adenosyl-L-methionine

TMZ: temozolomide

GBM: glioblastoma multiforme

MGMT: O^6-methylguanine-DNA methyl transferase

ALL: acute lymphoblastic leukemia

AML: acute myelocytic leukemia

O^6Me-dG: O^6-methylguanine

3Me-dG: 3-methylguanine

3Me-dA: 3-methyladenine

7Me-dA: 7-methyladenine

AGT: O^6-alkylguanine transferase

BER: base excision repair

Endo IV: Endonuclease IV

hOGG1: human 8-oxoguanine DNA glycosylase

NER: nucleotide excision repair

NMR: nuclear magnetic resonance

COSY: $^{1}H-^{1}H$ Correlation Spectroscopy

TOCSY: $^{1}H-^{1}H$ Total Correlation Spectroscopy

NOESY: Nuclear Overhauser effect spectroscopy

Fpg: formamidopyrimidine DNA glycosylase

CD_2Cl_2: deuterated dichloromethane

DMT: [bis-(4-methoxyphenyl)phenylmethyl]

CGE: capillary gel electrophoresis

DEPT: Distortionless enhancement by polarization transfer

hPol η :human Pol η

Dpo4: *Sulfolobus solfataricus* P2 Pol IV

Tg: thymine glycol

Endo III: Endonuclease III

Endo VIII: Endonuclease VIII

mNth1: mouse Endo III homologue

Chapter I.

Introduction

DNA Function and Structure

All living organisms possess a genetic blueprint, deoxyribonucleic acid (DNA), responsible for encoding their unique appearances and characteristics. The amount of DNA an organism possesses varies depending on species, but the importance of DNA is universal. DNA dictates cellular function, controlling development, growth, and reproduction. Of paramount importance is the ability of DNA sequences to be passed down through generations with high fidelity, preserving genetic codes that have been refined over millions of years of evolution. Failure to faithfully replicate DNA can lead to mutations in gene expression; alterations in the genetic blueprint may be advantageous or deleterious

The structure and function of DNA is largely preserved between all organisms. DNA is composed of strands of nucleotides linked by their phosphorous containing backbones; each nucleotide is composed of a five-carbon deoxyribose sugar unit and a nitrogenous base. The number of nucleotides in each strand of DNA is vast; human chromosome 1 alone contains approximately 249 million nucleotides per strand of DNA. Genetic information is conferred by DNA utilizing a code of four nitrogenous bases attached to the deoxyribose sugar. The four bases native to DNA are adenine (dA), guanine (dG), cytosine (dC), and thymine (dT) (**Figure 1**).

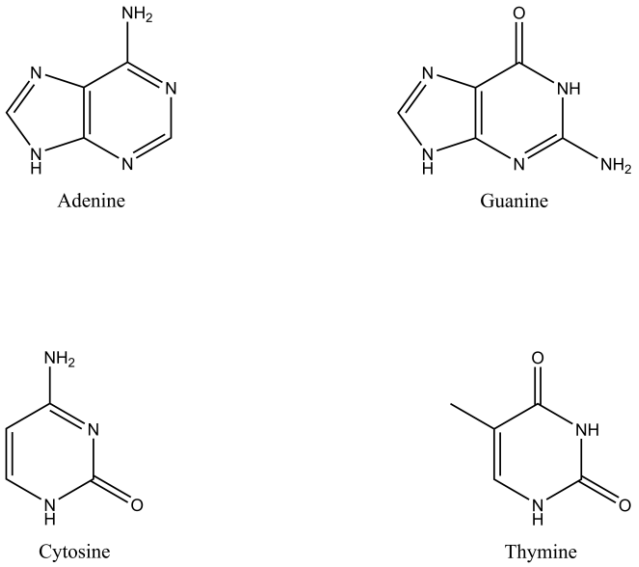


Figure 1. The four bases native to DNA are adenine (A), guanine (G), cytosine (C), and thymine (T).

The structure of DNA was elucidated in 1953 by James Watson and Francis Crick, using data collected by Dr. Rosalind Franklin, Dr. Raymond Gosling, and Dr. Maurice Wilkins.[1] Watson and Crick determined the structure of DNA to be helical, consisting of two anti-parallel strands of DNA. Unlike their earlier model, which erroneously proposed the phosphate backbone as being internally situated in a triple-strand helix,[2, 3] the model set forward in 1953 correctly proposed the phosphate backbone as being externally located to allow for favorable interactions with water molecules. The hydrophobic nitrogenous bases were relegated to the internal portion of the helix, where, based on the work of Dr. William Astbury[4], Watson and Crick predicted the bases would stack in such a way as to include ten bases per turn of the helix, with each nucleotide ~3.4 Å apart. Dr. Erwin Chargaff and Dr. Jerry Donohue advised Watson and Crick that the tautomers of thymine and cytosine would be predominantly in the keto and amine configurations.[5, 6] Based on the work of Dr. Erwin Chargaff, who determined the amount

of cytosine equals the amount of guanine and the amount of adenine equals the amount of thymine,[7] and the advisement of both Dr. Erwin Chargaff and Dr. Jerry Donohue, Watson and Crick were able to create a model of DNA, which portrayed the double strand helical structure and base pairing potential of the macromolecule (**Figure 2**).

Watson and Crick went further, noting in their land mark paper:

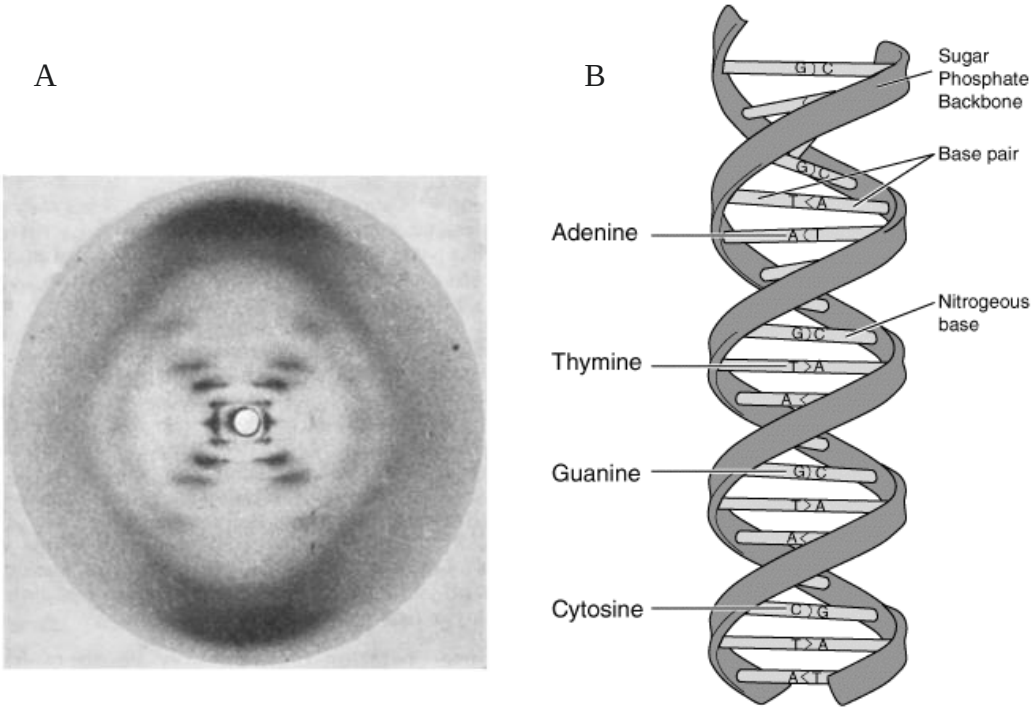


Figure 2. A) Rosalind Franklin's Photograph 51, the x-ray diffraction image of crystallized DNA.[1, 2] *B) Watson and Crick's model of DNA predicted the biological macromolecule would form a helical structure in which the bases would stack in such a way as to include ten bases per turn of the helix, with each nucleotide being placed at ~3.4 Å apart. Source: https://commons.wikimedia.org/wiki/File:DNA-structure-and-bases.png*

"It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material."

-Watson and Crick (1953)

This bold prediction, made only a year after DNA was proven to be the carrier of hereditary information, was eventually proven to be true. Meselson and Stahl provided crucial evidence of the mechanism of DNA replication in 1958, validating the semi-conservative model of DNA replication.[8, 9] Today extensive time and effort has been devoted to investigating repair and bypass mechanisms when a mismatch or damaged base is present.

Base pairs are formed between a purine and pyrimidine nitrogenous base in the DNA duplex; A:T and G:C base pairs are favorable due to their hydrogen bonding interactions. The A:T base pair forms two hydrogen bonds per pair, while the G:C base pair forms three hydrogen bonds per pair (**Figure 3**). The formation of hydrogen bonds between incorrect bases is disfavored in comparison. Disruption of the normal geometry of base pairs is often necessary to establish hydrogen bonding when a mismatch is present; this disruption can lead to unfavorable steric effects and disruption of base stacking interactions.[10, 11]

The additive strength of the hydrogen bonds and the base stacking of the nitrogenous bases gives the DNA duplex its stability. The ability of DNA bases to pair via hydrogen bonding and to form π-π stacking interactions between neighboring bases not only gives DNA it's 3-dimensional shape, but is also essential for the maintenance of DNA integrity. In the event of misincorporation during DNA replication, the complementary strand serves as a template,

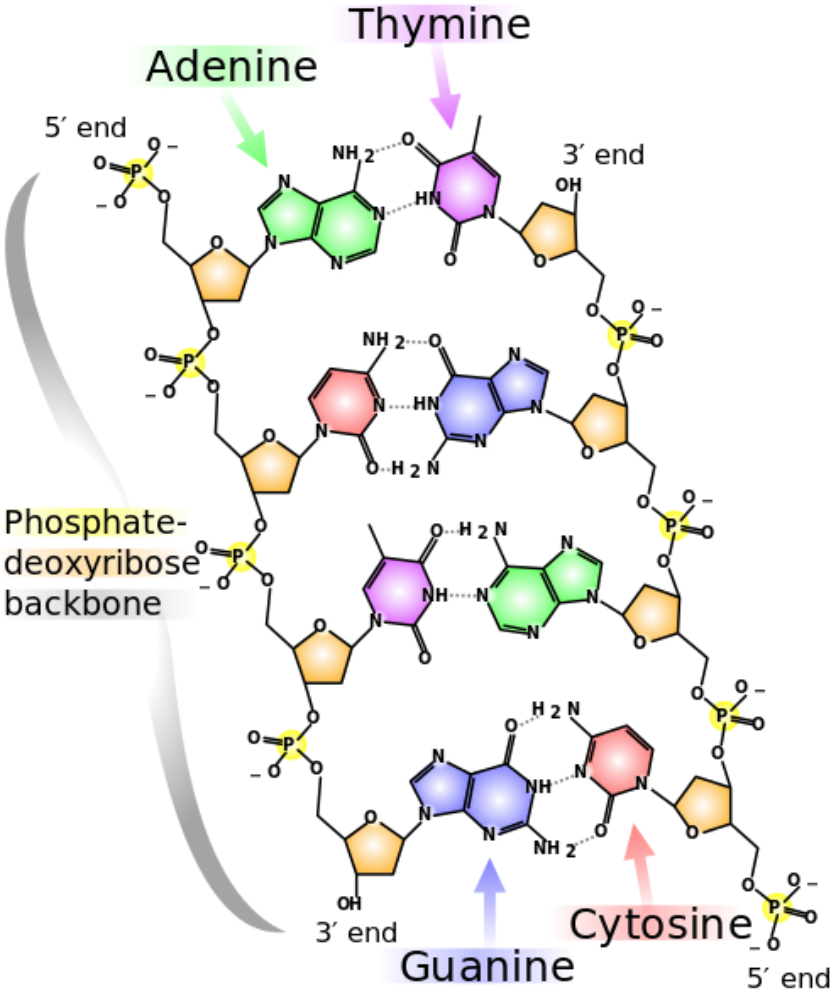


Figure 3. In DNA the A:T base pair forms two hydrogen bonds per pair and the G:C base pair forms three hydrogen bonds per pair.

Source: https://en.wikipedia.org/wiki/DNA#/media/File:DNA_chemical_structure.svg

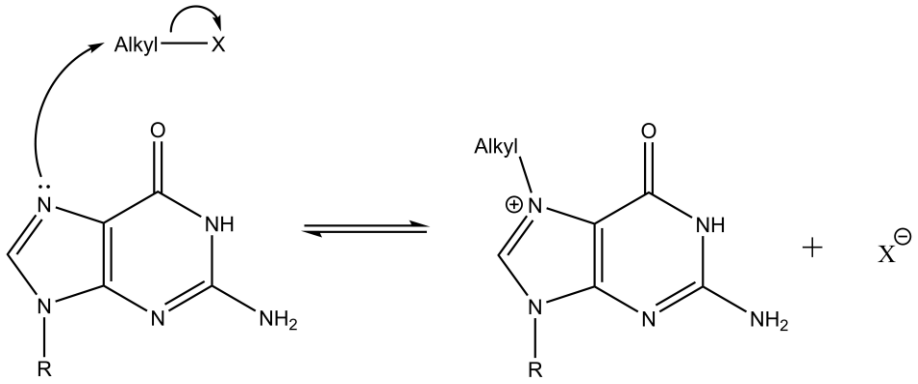
enabling enzymes to perform checks of the newly synthesized strand for incorrect base pairing and to accurately remove the erroneously incorporated base.[12] The complementary strand can also serve as a template for repair when DNA damage occurs; after the damaged portion of DNA is removed, DNA polymerase is able to accurately fill in the missing nucleotides based off the sequence of the undamaged complementary strand.[13] Failure to repair damaged DNA may result in mutations, caused by error-prone bypass of the adducts by DNA polymerases. Accumulation

of mutations is correlated with the formation of cancerous cells, which are able to rapidly proliferate compared to normal cells.[14] Indeed, one of the hallmarks of cancerous cells is their genetic instability; generally, there is a positive correlation between the number of mutations and the malignancy of the cancer.[15]

DNA damage is a daily occurrence, producing a diverse array of DNA lesions. In addition to well-known carcinogens, everyday cellular processes result in damage to DNA; a prime example of this is molecular oxygen, O_2.[16] Molecular oxygen is vital for the survival of many complex organisms, including humans. Molecular oxygen is responsible for accepting electrons from the end of the electron transport chain, a process that occurs in mitochondria and is a cells primary source of adenosine triphosphate, a source of energy. The oxygen is primarily reduced to water. Failure to fully reduce the molecular oxygen results in formation of the free radical superoxide. Superoxide is a highly reactive molecule known to cause oxidative damage to DNA. By virtue of breathing, accumulation of superoxide inevitably occurs and causes oxidative DNA damage. The inevitability of DNA damage occurring further highlights the importance of DNA damage repair.

DNA Alkylation

Alkylating agents are compounds prone to attack by nucleophiles, resulting in the transference of the alkyl group to the nucleophile (**Scheme 1**). Alkyl groups exhibit a vast diversity of sizes and chemical properties, from a simple methyl (CH_3) group to the bulky



Scheme 1. The N7 of guanine acts as a nucleophile toward an electrophilic alkyl group, resulting in the cationic species N7-Alkyl-dG.

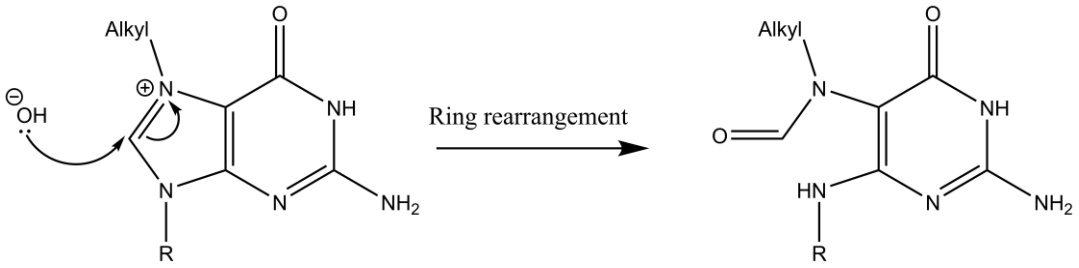
aflatoxin B_1 (AFB_1) group. Some sites in DNA are nucleophilic, and as such can react with alkylating agents when present. Alkylation of DNA is of interest due to the numerous alkylating agents present in the environment and the prevalence of their uses in the food industry and medicine.[17-23]

An important use of alkylating agents is the treatment of cancer; alkylating agents are the oldest class of chemotherapeutic agents still used today.[24-27] The effectiveness of alkylating agents in killing/inhibiting cancerous cells relies on the fast proliferation and decreased repair of DNA exhibited by cancer cells.[28, 29] These properties make cancer cells more sensitive to DNA damage by alkylating agents, ideally causing apoptosis in the cancerous cells. Unfortunately, like most chemotherapeutic agents, this class of drugs impacts not only cancerous cells but also healthy cells, leading to the nocuous side effects seen in many cancer patients undergoing treatment.

The most nucleophilic site in DNA is the N7 of guanine;[30, 31] as such, DNA alkylation products are primarily 7-alkyl-G adducts.[32-34] These cationic adducts can account for up to 80% of alkylation products, depending on the alkylating agent.[35] Mutations can occur, often due to the formation of an abasic site after depurination, a process which is promoted by acidic conditions.[36, 37] Smaller alkyl groups can be innocuous, causing no mutations or stalling during bypass. An example of this is 7-methylguanine (7Me-dG); the methyl group does not interfere with the base pairing face and is relatively small, producing minimal distortion of duplex structure and steric interference in the active sites of polymerases.[38, 39] Larger alkyl groups, which have the potential to distort normal duplex structure, may result in polymerase stalling, particularly in high fidelity polymerases, or mutagenic bypass during replication.[40] Recruitment of translesion synthesis (TLS) polymerases, which have a higher error rate when replicating unmodified DNA due to their larger active sites, may be required to bypass particularly bulky lesions.[41-43] 8,9-dihydro-8-(N7-guanyl)-9-hydroxyaflatoxin B_1 (AFB_1-N7-dG) induces significant helical distortion and is mutagenic when bypassed.[40]

Formation of Formamidopyrimidine Adducts

Alkylation of the N7 position of guanine produces a cationic adduct prone to depurination under acidic conditions, forming an abasic site.[36, 37] These cationic adducts may also undergo a secondary reaction with hydroxide.[44, 45] Nucleophilic attack by hydroxide at the C8 position, followed by spontaneous opening of the imidazole ring, forms the



Scheme 2. The cationic N7-Alkyl-dG is susceptible to nucleophilic attack by hydroxide ion, resulting in spontaneous imidazole ring-opening to produce the Alkyl-Fapy-dG adduct.

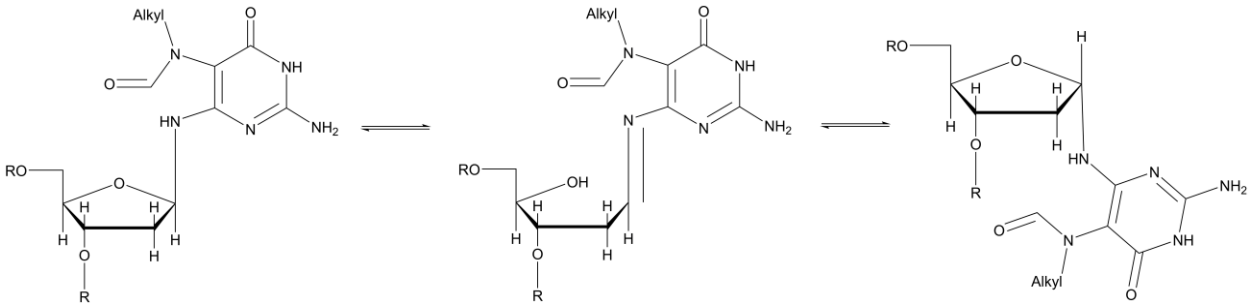
formamidopyrimidine (Fapy-dG) adduct (**Scheme 2**); this process is more favorable under alkaline conditions. The formation of alkyl-Fapy-dG adducts is of particular interest due to the prevalence of the precursor adduct, 7-alkyl-dG, compared to other alkylation adducts. The simplest of these N7 alkylation adducts is 7Me-dG, the predominant lesion formed from exposure to methylating agents; the adduct can be converted under alkaline conditions to the 2,6-diamino-4-hydroxy-N^5-(methyl)-formamidopyrimidine (MeFapy-dG) adduct.

Consequences of Formamido DNA Adducts

Investigations into the consequences of several Fapy-dG lesions have shown them to be mutagenic, cytotoxic, and persistent.[46-50] The simplest of these lesions, Fapy-dG, arises from oxidation of guanine rather than alkylation;[51, 52] Fapy-dG demonstrates increased flexibility and discrete isomers arising from rotation around bonds and from epimerization (**Schemes 3 and 5**).[53, 54] The complexity of this lesion has precluded characterization of the isomers; substitution of O4′ of the deoxyribose ring with a CH_2 group was utilized to prevent anomerization and characterize the rotational isomers.[55-57] The isomeric complexity of this class of lesions has

spurred interest, particularly when it was discovered that the mutagenic spectrum of a Fapy-dG lesion is equally complex.

The mutational spectra of Fapy-dG when bypassed in COS7 cells is a prime example of the potential mutational complexity of the Fapy-dG lesions.[47] The ratio of mutations and predominance of certain types of mutations was sequence dependent. Mutational patterns depended on the identity of the neighboring bases; the percentage of G → A transitions was 21.4% and 14.5% in the 5'-TXG-3' and 5'-CXG-3' sequence contexts, respectively. Intriguingly, the dependence on sequence context was observed when the nearest bases were identical and more distant bases were altered. No deletion mutations were detected in the 5'-GCTAGT**X**GGTCC-3' sequence, while deletion mutations accounted for 4% of the products obtained from bypass of the 5'-GCTCGT**X**GGTCC-3' sequence. The exact mechanism by which nearby bases influences the mutational spectrum is currently unknown. Proposed models include a primer/template misalignment–realignment mechanism and a direct misinsertion model, the latter of which recent studies support over the former.[47] Bypass of the enol tautomer of the α anomer of Fapy-dG has been hypothesized to occur by formation of a reverse wobble-base pair with an incoming dT; in β anomer enriched samples a decrease in G → transitions were



Scheme 3. Acid-catalyzed anomerization of Fapy-dG lesions results in the formation of the native β anomer or the non-native α anomer.

observed. Undamaged α-dG, when bypassed, directed insertion of dT ~60% of the time.[58] The isomers of Fapy-dG also have the potential to interact differently with enzymes; alterations in hydrogen bonding capabilities, base pairing face, and adduct-residue interactions at the active site could contribute to alternative base pairing with incoming dNTPs and successful bypass by enzymes.

Modulation of geometrical isomers by neighboring bases has been observed in the AFB_1-Fapy-dG adduct.[59] Hydrogen bonds formed with the exocyclic amino proton of a 3′-neighboring dA and the oxygen of the formyl group of AFB_1-Fapy-dG favor the *E* configuration of the formamido bond. Weaker hydrogen bonding between the formyl oxygen and the exocyclic amino proton of a 3′ dC results in a mixture of *E* and *Z* isomers. Conversion to the *Z* isomer when dC or dA are the 3′ neighbor is favored at higher temperatures, which weakens the hydrogen bond between the formyl oxygen and exocyclic amino proton. When no 3′ exocyclic amino groups are available for hydrogen bonding, as is the case when the 3′ neighbor is dT, the *Z* isomer, but not the *E* isomer, is observed. The ratio of configurational isomers is determined by whether the adduct is present in a single strand oligodeoxynucleotide or in duplex DNA. In single strand the adduct exists as the α anomer; annealing the single strand to its complementary strand results in conversion of the adduct to the β anomer, a process favored due to intercalation of the adduct into the DNA duplex.[60, 61]

Both Fapy-dG and AFB_1-Fapy-dG are highly mutagenic, producing G → T transversions as the predominant mutation.[46, 62-64] Formamidopyrimidine adducts formed by alkylation of DNA can possess greater mutagenic potential than their initially formed alkyl-N7-dG precursor adducts. Bypass of AFB_1-N7-dG exhibited a mutational rate of 45% in primate cells; bypass of AFB_1-Fapy-dG in primate cells exhibited a mutational rate of 97%.[40]

DNA Methylation

Another common type of modification of DNA bases, which occurs both intentionally and unintentionally, arises from methylation.[65-70] Methylation can occur from exposure to both exogenous sources and endogenous sources; the distribution of methylated bases varies depending on the source of the methylating agent and the properties of the methylating agent.[71]

Endogenous Methylation

Endogenous methylation plays a vital role in cells. DNA methylation is catalyzed by DNA methyltransferase, a conserved family of enzymes that plays a role in epigenetic regulation. Methylation occurs primarily at the C5 position of cytosine. The methylating agent in these reactions is S-adenosyl-L-methionine (SAM).[72]

One of the main functions of DNA methylation in *E. coli* is differentiation between the parent and daughter strands in a newly replicated DNA duplex; it is vital that, should a mismatch occur, the newly synthesized strand be corrected rather than the parent strand.[73, 74] While the rate of error is low, high-fidelity DNA polymerases will occasionally incorporate the wrong nucleotide opposite the parent strand, resulting in a mismatch that will be passed on to future cells if not corrected. Methylation of the parent strand provides a marker to enzymes, enabling discrimination between the parent and daughter strand. This signaling preserves the integrity of the DNA code as it is passed down through cell divisions. In eukaryotic cells, the precise mechanism by which the daughter strand is differentiated from the parent strand is currently unknown.[75, 76]

Methylation of cytosine also serves another purpose in some organisms: control of gene function and expression. Incorporation of a methyl group at the C5 of cytosine results in projection of the methyl group into the major groove of helical DNA, thereby inhibiting transcription.[77, 78] Recently, studies have found that methylation of cytosine also has the potential to promote transcription factor binding or fail to hinder binding.[79, 80] Selective inhibition of gene expression at a cellular level is highly advantageous for multicellular organisms, particularly as the complexity of the organism increases and the cells become further differentiated for specific functions; the necessary functions of a cell located in the heart will differ greatly from the functions required of a cell located in the kidney. Transcription inhibition by cytosine methylation prevents the production of unnecessary proteins for the function of individual cells, increasing their efficiency and ability to perform specific functions. Loss of the ability to methylate DNA has been found to cause inappropriate gene expression and delayed cell differentiation.[81]

Exogenous Methylation

The introduction of the methylating chemotherapeutic agent temozolomide (TMZ) in 1999 revolutionized the way we treat patients with glioblastoma multiforme (GBM), the most common malignant primary brain tumor. TMZ is a S_N1-type DNA methylating agent that can pass through the blood-brain barrier.[82, 83] Upon hydrolysis, it forms a methyldiazonium ion capable of alkylating DNA at the N7-dG, N3-dA, and O^6-dG positions, and also the N1-dA and N3-dC positions, particularly in single strand DNA.[84] Patients with GBM generally have a poor prognosis, with an median overall survival time of 7.6 months when treated with radiation.[85] The introduction of TMZ into patients' treatment regimen extends median survival time to 9.3

months. Patients with changes in the O^6-methylguanine-DNA methyl transferase gene (*MGMT*) gene of glioblastoma cells showed even greater median survival times,[86, 87] with the median survival time increasing to 13.5 months.[85] Progression-free survival rates were found to be improved in patients receiving both radiotherapy and TMZ; treatment with TMZ increased the 1 and 2-year survival rates to 37.8% and 10.4%, respectively, an improvement over the 1 and 2-year survival rates, 22.2% and 2.8%, respectively, in patients who do not receive TMZ.

The successful treatment of GBM with TMZ has expanded interest in utilizing the drug in the treatment of other types of cancer. Recruitment is underway in a clinical trial to determine the efficacy of TMZ in the treatment of mucosal melanoma.[88] Phase 2 trials on the use of TMZ in patients with advanced melanoma are also underway.[89] Results published on the efficacy of TMZ in treatment of rectal cancer suggested TMZ may be beneficial in the treatment of some patients.[90] This increased interest in treatment of other types of cancers with TMZ, some with longer median survival times, raises the question as to long term side effects in patients receiving treatment with TMZ. Although TMZ extends the median survival time of patients,[91-98] the use of TMZ has also been correlated to an increased incidence of secondary acute lymphoblastic leukemia (ALL) and acute myelocytic leukemia (AML).[99-101]

Methylation of DNA can arise from exposure to other exogenous methylating agents. DNA may be exposed to DNA methylating agents from exogenous sources[102-104] including cigarette smoke,[105-108] food,[109-111] environmental exposures,[112, 113] workplace chemicals,[114, 115] and endogenous alkylating agents.[116] Unlike the controlled enzymatic reactions that occur with endogenous methylating agents, cytosine is not the predominant target of methylation. The DNA adduct 7Me-dG is the major methylation product after exposure of DNA to exogenous

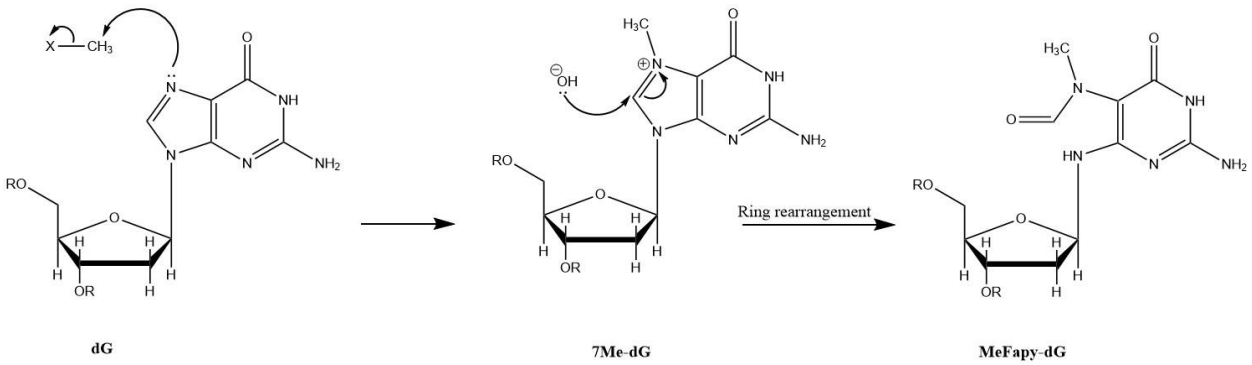
methylating agents, comprising 65-80% of all methylated products.[35] The predominance of this adduct can be attributed to the nucleophilicity of the N7 of guanine.

Other common methylated products are O^6-methylguanine (O^6Me-dG), 3-methylguanine (3Me-dG), 3-methyladenine (3Me-dA), and 7-methyladenine (7Me-dA). Unlike 5-methylcytosine, these methylated DNA adducts can be highly mutagenic (O^6Me-dG) or cytotoxic (3Me-dA). Some, like 7Me-dG, are not particularly mutagenic or cytotoxic, and as such have been regarded as relatively harmless DNA adducts.[83] Recent research provides evidence that 7Me-dG is potentially more deleterious to cells than initially thought due to the adduct undergoing a secondary reaction.

The 2,6-Diamino-4-hydroxy-N^5-(methyl)-formamidopyrimidine (MeFapy-dG) Adduct

7Me-dG, the predominant adduct formed after exposure to exogenous alkylating agents, can undergo a secondary reaction with hydroxide to form the 2,6-diamino-4-hydroxy-N^5-(methyl)-formamidopyrimidine (MeFapy-dG) adduct (**Scheme 4**). Unlike 7Me-dG, MeFapy-dG has been found to have potentially serious biological consequences. MeFapy-dG and its potential biological consequences have been investigated for decades; MeFapy-dG was first described in 1962 by Haines et. al, who identified it at the nucleoside level.[117]

The MeFapy-dG adduct has been identified in hepatic cells of rats treated with *N*,*N*-dimethylnitrosamine, 1,2 dimethylhydrazine, and *N*-methylnitrosourea; MeFapy-dG was also identified in bladder epithelial cells of rats treated with N-methylnitrosourea.[50, 118] Unlike 7Me-dG, MeFapy-dG inhibits DNA synthesis,[119-121] is mutagenic in mammalian cells,[122] and is bypassed by Y-family polymerases in an error-prone manner.[123, 124] Thus, there is a need to

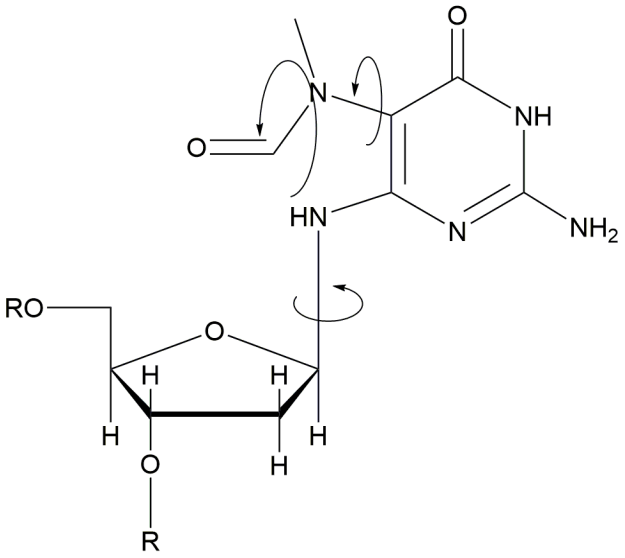


Scheme 4. Formation of the MeFapy-dG lesion occurs when the initial cationic adduct, 7Me-dG, undergoes nucleophilic attack by hydroxide, resulting in ring opening of the imidazole ring.

characterize the chemistry and biology of the MeFapy-dG lesion to better understand the structural basis for its cytotoxicity and mutagenicity.

The question as to repair of site-specific DNA methylation damage has become of paramount importance in improving the efficacy of treatments and in the development of patient-specific approaches (personalized medicine) of TMZ-based adjuvant therapies in the clinic. Some tumor cells are better able to repair TMZ-induced DNA alkylation, diminishing the efficacy of this anti-tumor agent. It is recognized that cells expressing O^6-alkylguanine transferase (AGT) encoded by the O^6-methylguanine-DNA methyl transferase gene (*MGMT*) gene[125] exhibit refractory behavior with respect to TMZ treatment, whereas epigenetic silencing of the *MGMT* gene increases cytotoxic response to TMZ treatment.[87, 91, 126] Likewise, the repair of 3-methyl-dA and AP sites by the base excision repair (BER) mechanism also mediates glioblastoma resistance to TMZ treatment.[94] Consequently, understanding mechanisms of recognition and repair, and also the error-prone replication, of regio- and site-specific DNA methylation damage induced by TMZ treatment is of interest.[83, 94]

The chemistry and biology of the MeFapy-dG lesion is complex. Anomerization, the acid catalyzed ring opening of the deoxyribose ring followed by closing to form either the



Scheme 5. The MeFapy-dG adduct has the potential to exist as a mixture of equilibrating conformational isomers arising from rotation around the formyl bond, C^5-N^5 bond, and glycosidic bond.

natural β anomer or unnatural α anomer, occurs (**Scheme 3**).[127-133] Additionally, there is potential for conformational isomerization arising from rotation around the formyl bond, C^5-N^5 bond, and glycosidic bond (**Scheme 5**).[50, 118, 128, 134-136] Tomasz et al.[128] identified C^5–N^5 rotamers in the Fapy-dG nucleoside derived from 7-ethyl-dG as well as with the mitomycin C-Fapy-dG adduct. NMR studies on AFB_1-Fapy-dG demonstrated that the R_a atropisomer is favored for the AFB_1-Fapy adduct and that intercalation of the AFB_1 moiety on the 5' face of the pyrimidine ring maintains the R_a absolute configuration of the C^5–N^5 bond.[60, 131] Raoul et al. demonstrated in 1995 that *E* and *Z* geometrical isomers occur, with respect to rotation around the formyl bond of Fapy-Ade; $^3J_{HH}$ couplings of 1.3 Hz and 11.7 Hz were attributed to the *Z* and *E* geometrical isomers, respectively.[137] Christov et al.[133] utilized an *Escherichia coli* Endonuclease IV (Endo IV)-based enzymatic assay,[138] showing that at equilibrium in duplex DNA the α- and β-MeFapy-dG anomers exist in a 20:80 α:β ratio. Christov et al. also showed that the

equilibration between the two anomers occurred over 3 days, and that the β anomer is favored in single strand and duplex DNA.[133] The β anomer is also favored in the unsubstituted Fapy-dG adduct.[129, 138]

The isomeric species may interact differently with DNA processing enzymes, leading to different biological outcomes. Both the rates of mutation and the most common mutations induced by the MeFapy-dG lesion vary depending on the 5' and 3' neighboring bases in DNA.[122, 124] Differences in hydrogen bonding abilities and steric strain between the lesion and neighboring bases have the potential to influence the populations of the isomeric species of the MeFapy-dG adduct.

Repair of Fapy-dG Lesions

Alkylated Fapy-dG lesions have proven difficult to repair, refractory compared to their N7-dG precursor adducts. AFB_1-Fapy-dG, a highly mutagenic adduct linked to hepatocellular cancer, has a much longer half-life than AFB_1-N^7-dG.[139] This relatively bulky Fapy-dG lesion was determined to be a poor substrate for the repair enzyme human 8-oxoguanine DNA glycosylase (hOGG1), an enzyme that performs BER; comparatively, the smaller MeFapy-dG lesion was determined to be a better substrate for hOGG1.[140] Repair of AFB_1-Fapy-dG has been found to occur preferentially by the nucleotide excision repair (NER) pathway in bacteria.[140] Recent research into the role of DNA glycosylase NEIL1 has revealed the enzyme offers some protection against aflatoxin-induced hepatocellular carcinoma, a surprising discovery given that human NEIL1 is part of the BER machinery.[141]

MeFapy-dG has been found to be persistent in rat bladder epithelial cells and rat liver cells.[50, 118] This smaller Fapy-dG lesion is a substrate for repair enzymes hOGG1 and human

NTH1.[142, 143] Human NEIL1 has also been found to excise MeFapy-dG.[144] The ability of cells to repair alkyl-Fapy-dG lesions is of paramount importance when planning a treatment regimen for a cancer patient; the tolerance of their specific strain of cancer to certain types of alkylation damage will affect the efficiency of the treatment. Increased production of alkylation-repair enzymes can result in an increased tolerance to alkylating chemotherapeutic agents, potentially harming the patient with little improvement to their prognosis. Improved treatment may be obtained through 'personalized medicine', in which a cancer treatment regimen is tailored to individual patients based on the probability of positive and negative responses to specific chemotherapeutic agents.

Decreased ability to repair alkyl-Fapy-dG lesions may have its own hazards. While methods of cancer treatment have improved vastly over the years, it is undeniable that many chemotherapeutic agents do not only affect cancerous cells.[145] Damage to healthy non-cancerous cells during treatment may lead to the development of a secondary cancer, whose origins cannot be traced back to the original tumor. There is some evidence that this occurs; patients treated with temozolomide for varying types of brain tumors have been observed to have an increased incidence of developing acute lymphoblastic leukemia. In one such study the patient, a young girl with no previous health problems, survived the brain tumor that originally brought her to the hospital with numbness, only to develop acute lymphoblastic leukemia several months after treatment with temozolomide concluded.[101] Attempts to treat her acute lymphoblastic leukemia with further chemotherapy failed, and she ultimately passed away from pneumonia. Off-target reactions, an inevitable outcome when employing non-selective treatment methods, are essential in the consideration of treatment plans.

The exact mechanism by which alkyl-Fapy-dG lesions are recognized is currently unknown; alkyl-Fapy-dG lesions have intact base pairing faces, and as such are able to hydrogen bond with an incoming cytosine during replication. The complexity of the mutagenic spectra observed when these lesions are present provides a possible mechanism for accumulation of a diverse array of mutations, such as those required to become cancerous, months after treatment is halted. Difficulty repairing these lesions would increase the probability of mutations occurring.

NMR Spectroscopy of DNA

Nuclear Magnetic Resonance (NMR) spectroscopy is a powerful investigative tool in the field of chemistry. This is particularly true in cases involving highly complex molecules, such as DNA and proteins. Detection and discrimination between equilibrating isomers can be achieved through careful observation of NMR active nuclei, such as ^{1}H and ^{13}C. Due to the complexity

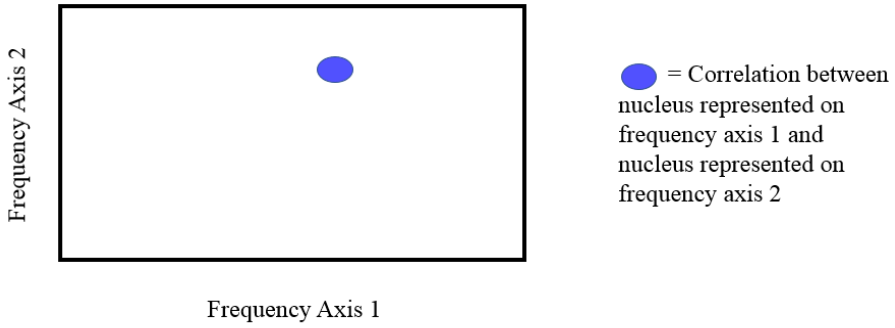


Figure 4. In a 2-dimensional NMR experiment the frequency axes are formed from 1D NMR experiments. Correlations are seen between two NMR active nuclei at the intersection of the chemical shifts of the nuclei; the correlations observed is dependent on the NMR method employed.

and size of even a relatively short DNA duplex, 2D NMR experiments are commonly used for assignments of individual protons and extrapolation of structural features. A 2D NMR experiment contains two frequency axes, with cross-peaks indicating a specific type of relationship between two nuclei of differing chemical shifts (**Figure 4**). The type of relationship observed is dependent on the NMR method used. The 2D NMR experiments employed in this work were COSY, TOCSY, HSQC, and NOESY experiments.

COSY, TOCSY, and HSQC experiments all show "through bonds" correlations, or interactions between two nuclei connected by a number of covalent bonds. The ^{1}H-^{13}C HSQC[146] experiment extends between hydrogen and carbon nuclei separated by one bond. One frequency axis corresponds to the ^{13}C nuclei; the second corresponds to the ^{1}H nuclei. Cross-peaks will arise between ^{1}H nuclei and ^{13}C nuclei only when they are directly bound to each other (**Figure**

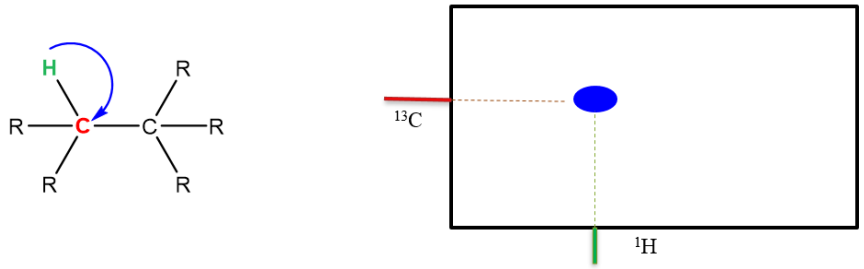


Figure 5. The HSQC experiment produces correlations between ^{1}H nuclei and ^{13}C nuclei only when they are directly bound to each other. (Left)The ^{1}H nucleus (green) is directly bound to the ^{13}C nucleus (red). (Right) A cross-peak at the intersection of the frequency axes of the ^{1}H nucleus and ^{13}C nucleus. The chemical shift of the ^{1}H nucleus and ^{13}C nucleus are shown as green and red lines, respectively.

5). This method can be used to establish the chemical shift of a proton, which may be difficult to determine otherwise due to overlap in the 1D ^{1}H-NMR spectrum. This technique can be enhanced by selective ^{13}C-labeling; ^{13}C, the NMR active isotope of carbon, has a natural abundance of 1%. Enriching a sample with a ^{13}C-label to nearly 100% at a selected carbon allows for the rapid identification of key portions of molecules; detection of minor species when the sample is an isomeric mixture is also enhanced.

A ^{1}H–^{1}H Correlation Spectroscopy (COSY)[147] experiment detects, via through-bond spin-spin coupling, interactions between two hydrogen nuclei separated by three bonds; correlations between protons separated by more than three bonds will not normally be observed. In a molecule of unmodified DNA, correlations between the H5 and H6 protons of cytosine are the only COSY peaks that will be observed between protons of the nitrogenous bases (**Figure 6**). Connectivity of the deoxyribose ring protons can also be established by the sequential connectivity observed in a COSY experiment. The ^{1}H–^{1}H Total Correlation Spectroscopy (TOCSY)[148] experiment, also known as the homonuclear Hartmann–Hahn (HOHAHA) experiment, exhibits correlations between protons separated by three or more bonds; correlations

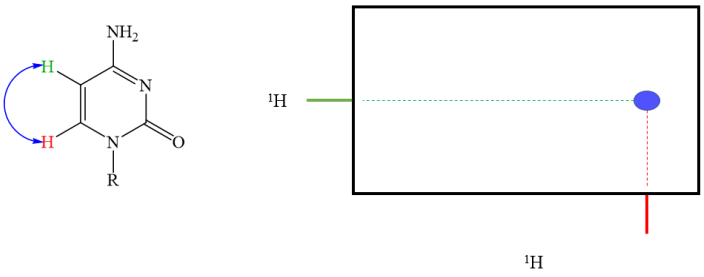


Figure 6. The COSY experiment produces correlations between ^{1}H nuclei separated by three bonds by detection of through-bond spin-spin coupling. In cytosine the H5 (green) and H6 (red) protons will have a correlation represented as a cross-peak at the intersection of the H5 frequency and H6 frequency.

are observed between all hydrogen nuclei in a 'spin system' (**Figure 7**). Similar to a COSY experiment, cross-peaks arise from through-bond spin-spin couplings. A mixing time added to the TOCSY method allows for evolution of cross-peaks beyond the three bond limit seen in a COSY experiment, producing cross-peaks between all nuclei in a spin system. The extent of magnetization transfer is dependent on the length of the mixing time; the shorter the time, the less the magnetization can travel. Overly long mixing times result in a loss of signal due to T_2 relaxation. The spin system ends when the system of sequential couplings is broken; in DNA this occurs in the form of an atom with no proton bound to it.

A Nuclear Overhauser effect spectroscopy (NOESY)[149, 150] experiment shows correlations between two nuclei that are spatially close to each other, but not necessarily bonded close together (**Figure 8**). Correlations arise from nearby nuclei that undergo cross-relaxation by

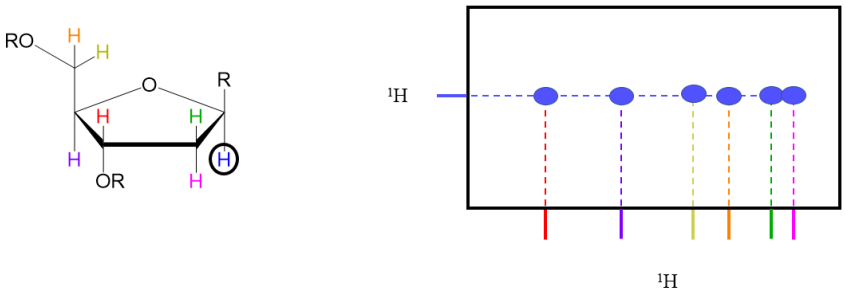


Figure 7. The TOCSY experiment produces correlations between all ^{1}H nuclei in a spin system, including those produced in a COSY experiment. (Left) The H1'proton of the deoxyribose ring is spin-spin coupled to the H2', H2'', H3', H4', H5', and H5'' protons. (Right) Cross-peaks are produced between the H1'proton of the deoxyribose ring is spin-spin coupled to the H2', H2'', H3', H4', H5', and H5'' protons. Cross-peaks will also arise at the intersection of the frequencies between all protons in the spin system.

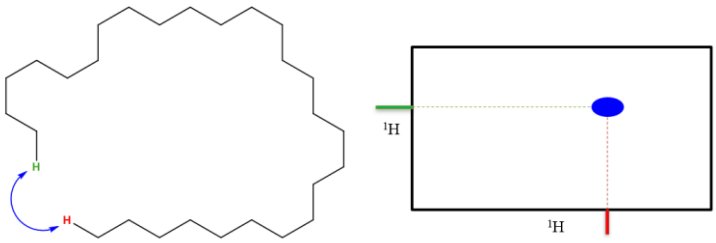


Figure 8. In a NOESY experiment correlations between two protons that are spatially close to each other, regardless of how many bonds separate them, are observed. (Left) Large molecules, such as a protein or DNA, can be held in a relatively ordered structure, resulting in proximity in two protons (red and green) separated by many bonds. (Right) Protons within ~5 Å will produce a cross-peak at the intersection of their frequencies in a NOESY experiment.

spin-lattice relaxation; unlike the other 2D techniques, no spin coupling is required between the two nuclei. The strength of a NOESY correlation is dependent on the distance between the two nuclei and on the population of each molecule or isomer. In the study of biological macromolecules, such as protein and DNA, NOESY experiments are utilized to determine structural features. Hydrogen bonding between the nitrogenous bases of DNA hold the strands in a structured double helix, allowing for the evolution of internucleotide NOEs between neighboring nucleotides. In undamaged B-DNA, sequential NOE connectivity of H8/H6 protons to H1′ protons is observed for each strand of DNA (**Figure 9**), forming what is colloquially known as the 'walking region'. This connectivity is also observed between other protons of the deoxyribose sugar and nearby H8/H6 protons. In damaged or modified DNA breaks in the NOE connectivity can be observed. Breaks in the H8/H6 to H1′ NOE connectivity are indicative of increased disorder in the duplex; the evolution of internucleotide NOEs is dependent on the relatively ordered structure of duplex DNA. Loss of order through the disruption of hydrogen

bonds or through steric interactions decreases this stability, sometimes leading to a loss, weakening, or broadening of internucleotide NOEs.

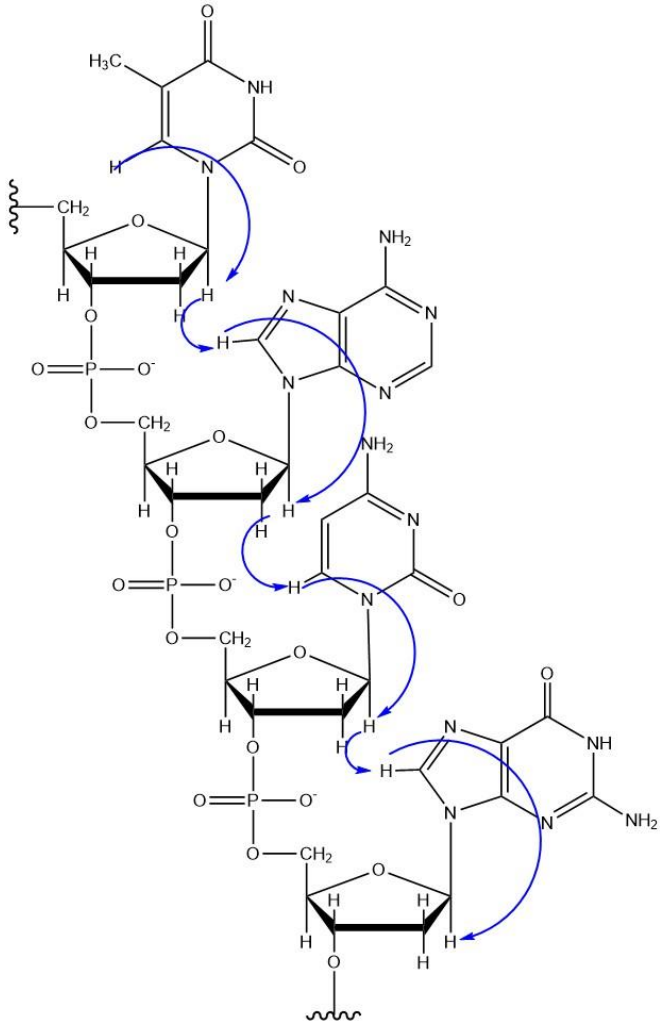


Figure 9. The sequential connectivity of H8/H6 to H1' NOEs is observed in undamaged B- form duplex DNA. Using these NOEs, one can 'walk' from the 5' end of the strand to the 3' end with no breaks in connectivity. Deviation from this pattern in damaged DNA is indicative of a perturbed duplex structure and increased disorder in the duplex.

Statement of Purpose

Alkylated Fapy-dG adducts are a diverse and complex type of DNA damage which, given the prevalence of alkylating agents in our cells and in our environment, have implications for individuals worldwide. Exposure to aflatoxin B_1, a potent carcinogen, is most common in Eastern Asia, South America, and Sub-Saharan Africa.[151] Formation of methylated DNA adducts results from exposure of DNA to exogenous sources[102-104] including cigarette smoke,[105-108] food,[109-111] environmental exposures,[112, 113] workplace chemicals,[114, 115] and endogenous alkylating agents.[116] The most predominant methylation product, 7-methylguanine, can undergo a secondary reaction with hydroxide to form the mutagenic and highly persistent MeFapy-dG lesion.[50, 118, 122]

Of particular interest is intentional exposure of cells to alkylating agents, such as chemotherapeutic agents. A cancer patient's response to some chemotherapeutic agents is regulated in part by genetics; cancer cells show variable tolerance upon exposure to alkylating agents. Patients whose cancer cells overexpress a protein known as O^6-methylguanine methyltransferase (MGMT) generally have a poor response to methylating agents such as temozolomide. Conversely, patients whose cancer cells produce low levels of MGMT respond well to treatment with temozolomide.[152, 153] Repair of MeFapy-dG has also been found to occur in the presence of formamidopyrimidine DNA glycosylase (fpg) and human 8-oxoguanine DNA glycosylase 1 (hOGG1).[154-156]

Outcomes of chemotherapy may be improved by screening patients in advance to determine whether they are likely to respond well to specific treatment regimens. This is potentially more complicated in the case of Fapy-dG lesions, which exists as a complex mixture of equilibrating isomers. Enzymes, which are often stereospecific, may have variable

interactions with these lesions, depending on which isomer it encounters. It is hypothesized that the population of isomers is sequence specific; the predominance of geometrical isomers of AFB_1-Fapy-dG was found to be modulated by neighboring nucleotides due to the presence or lack of favorable hydrogen bonds between the formamido group and neighboring bases.[59] Smaller Fapy-dG lesions such as Fapy-dG and MeFapy-dG have eluded full characterization in solution without chemical modifications. The purpose of this research is to shed further light on the mixture of isomers of MeFapy-dG present in solution, develop methods for the further study of alkyl-Fapy-dG lesions, and to investigate the sequence specificity of the isomeric populations of the MeFapy-dG adduct.

In order to quantitate the types and amounts of specific configurational and conformational isomers of the MeFapy-dG lesion in DNA, we have developed NMR based isotope-editing approach. This has permitted elucidation of the identities and populations of the MeFapy-dG isomers. Similar approaches have been used to identify isomers of the structurally similar Fapy-dG lesion.[56] In the present work the anomeric population of the MeFapy-dG phosphoramidite was determined to favor the α anomer; eight equilibrating species of the phosphoramidite were observed in solution. We unambiguously assigned the eight major isomers of the MeFapy-dG lesion in the 5'-TXT-3' sequence context. A similar distribution of species was observed in the 5'- $C^1 A^2 T^3 X^4 A^5 T^6 G^7 A^8 C^9 G^{10} C^{11} T^{12}$-3' (X = ^{13}C-MeFapy-dG) sequence context. In the G:C rich duplex, 5'-$G^1 C^2 T^3 A^4 G^5 T^6 X^7 G^8 G^9 T^{10} C^{11} C^{12}$-3' (X = MeFapy-dG), two species were identified, both of which were β anomers arising from rotation around the C^5-N^5 bond; no α anomers were identified in this system. Methods utilizing isotopic labeling of the MeFapy-dG lesion and the labile N^6H proton were developed; these methods can theoretically be applied to other, more complex, Fapy-dG lesions.

Statement of Dissertation

During this dissertation I developed methodology to rapidly identify isomers of MeFapy-dG in solutions containing complex isomeric mixtures. I proposed that rotation around the formyl bond could be detected by $^{3}J_{CH}$ values between the ^{13}C-labeled carbon atom of the methyl group and the formyl proton of the formamido moiety. This was confirmed by a non-decoupled DEPT-135 experiment, which further demonstrated a relatively large chemical shift difference between the rotamers. I determined that the anomers of MeFapy-dG were most easily identified by NOEs between the H1′, H3′, and H4′ protons of the deoxyribose ring in a trimer oligodeoxynucleotide; in a larger oligodeoxynucleotide duplex, the anomers of MeFapy-dG were determined by NOEs between the lesion and neighboring nucleotides. I developed a method of identification of rotamers arising from rotation around the C^5-N^5 bond, comparing NOE peak strengths between the N^6H, CH_3, and CHO protons in a trimer sequence, and between the CH_3, CHO, and 5′ neighboring nucleotide protons in duplex DNA. I determined the anomeric population of MeFapy-dG in a phosphoramidite sample, identifying the α anomer as the major species, and determined the isomeric populations of MeFapy-dG in a trimer and dodecamer system.

Chapter II.

2,6-Diamino-4-hydroxy-N^5-(methyl)-formamidopyrimidine (MeFapy-dG) Adduct Phosphoramidite

Materials and Methods

Oligodeoxynucleotide Synthesis. Dr. Chanchal Malik synthesized the MeFapy-dG phosphoramidite as previously described.[157]

NMR. (a) Sample Preparation. The MeFapy-dG phosphoramidite was dried *in vacuo* for 24 h. The 5-mm NMR tube was washed and initially dried under N_2; the tube was additionally dried *in vacuo* for 48 h; the canister containing the NMR tube was filled with dry N_2 upon release of the vacuum. The dried phosphoramidite was dissolved in 600 µL of deuterated dichloromethane (CD_2Cl_2) and immediately transferred to the NMR tube; the sample was placed under an atmosphere of argon gas and sealed with parafilm. Additional CD_2Cl_2 was added throughout the experiments as needed. All spectra were collected using Avance III NMR spectrometers equipped with cryogenic probes and precise temperature control (Bruker BioSpin, Inc., Billerica, MA).

(b) NMR Data Collection. COSY[147] and $^1H–^1H$ total correlated spectroscopy (TOCSY)[148] spectra in CD_2Cl_2 were recorded at a ^1H frequency of 600 MHz. Acquisition parameters for COSY experiments were as follows: 2048 real data points in the t_2 dimension and 512 real data points in the t_1 dimension, number of scans per FID = 16, sweep width = 7211.539 Hz, QF mode, relaxation delay = 1.5 s; the data was zero filled during processing to obtain matrices of 2048 x 2048 data points. Acquisition parameters for TOCSY experiments were as

follows: 2048 real data points in the t_2 dimension and 512 real data points in the t_1 dimension, number of scans per FID = 16, sweep width = 7211.539 Hz, States-TPPI mode, relaxation delay = 1.5 s, mixing time = 80 ms; the data was processed to obtain matrices of 1024 X 1024 data points. Nuclear Overhauser effect spectroscopy (NOESY)[149, 150] experiments were performed at a ^{1}H frequency of 800 MHz with 2048 real data points in the t_2 dimension and 512 real data points in the t_1 dimension. These spectra were obtained with mixing times of 0.2s to 1.8 s, 80 scans per FID, a sweep width of 9615.385 Hz, States-TPPI mode, and a relaxation delay of 2 s. The data was zero filled during processing to obtain matrices of 2048 x 1024 points.

(c) NMR Data Processing. NMR spectra were processed with the TOPSPIN software (Bruker BioSpin, Inc., Billerica, MA) and analyzed with the SPARKY software[158] and TOPSPIN. NMR spectra were integrated using TOPSPIN.

Results

TOCSY NMR: **Figure 10** shows the spectrum resulting from a TOCSY experiment conducted on the MeFapy-dG phosphoramidite. Eight spin systems of the deoxyribose ring were identified by N^6H-H1′ connections, indicating the presence of eight isomers of the MeFapy-dG phosphoramidite. Protons of the deoxyribose ring, which exhibited similar splitting, were identified by TOCSY and COSY experiments. The N^6H proton chemical shift was split into two distinct groups, separated by over 0.5 ppm and integrating in a 40:60 ratio. The source of these

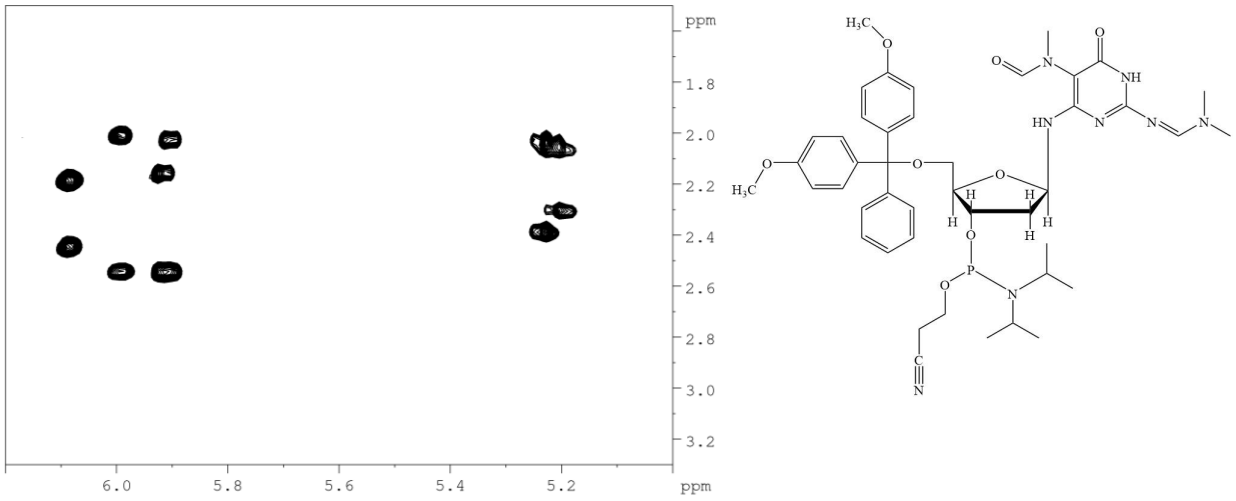


Figure 10. The TOCSY experiment indicated the presence of eight spin systems arising from the deoxyribose ring of the phosphoramidite (N^6H-H2′ and N^6H-H2′′ cross-peaks shown). The peaks were split into two distinct groups.

two groups was hypothesized to be epimerization of the phosphoramidite that occurred immediately after the initial synthesis of the phosphoramidite.

The H2′ and H2′′ protons were unequivocally identified utilizing a NOESY experiment based on proximity to the H3′ proton (**Figure 11**). Anomeric identity was determined according to NOE peak strength between the H2′ and H2′′ protons and H1′ proton (**Figure 11**). Contrary to findings of a previous study of MeFapy-dG,[133] the α anomer was the major species, accounting for 60% of the MeFapy-dG phosphoramidite. The remaining 40% was determined to be the β anomer.

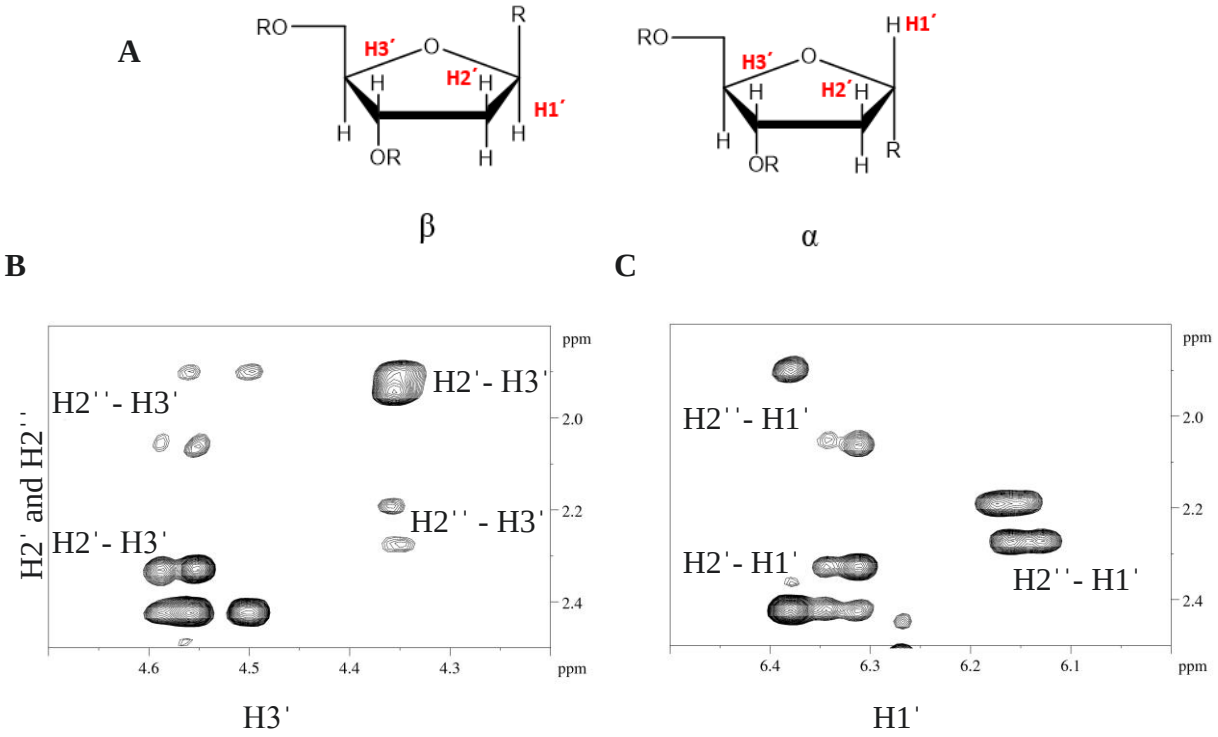


Figure 11. A) The H2' and H2'' protons can be unambiguously assigned by comparison of NOE peaks strength between these protons and the H3' proton; due to a shorter distance between the H2' and H3' protons compared to H2'' and H3' protons a stronger H2'-H3' NOE will be observed. The β anomer will exhibit a stronger H1'-H2'' NOE than H1'-H2' NOE due to a difference in distance between the protons; the α anomer will exhibit a stronger H1'-H2' NOE than H1'-H2'' NOE due to a difference in distance between the protons B) The H2' and H2'' protons of the deoxyribose ring were unequivocally assigned by comparison of NOEs between these protons and the H3' proton. C) Anomeric identity was established by NOE peak strength between H1'-H2' and H1'-H2'' protons. The downfield group of peaks (with respect to H1') were assigned as the β anomers; the upfield group of peaks were assigned as the α anomers.

Rotational isomers were detected by variation in NOE peak strength that could not be attributed to differences in population. NOE peaks between N^6H, formyl, and methyl protons were used to make these determinations (**Figure 12**). Rotation around the C^5-N^5 bond would result in R_a and S_a geometrical isomers; it was hypothesized that these isomers could be differentiated by comparing variations between N^6H-formyl and N^6H-CH_3 NOE strengths; these variations in NOE strength, should they occur, would be unattributable to population differences. N^6H-formyl and N^6H-CH_3 NOEs were present for the eight species; the NOEs showed greater differences in strength than population differences would suggest, indicating the presence of rotational isomers. Due to the relatively small amount of each isomer and to partially overlapping peaks, precise integration of NOEs could not be performed, precluding full characterization of rotational isomers in this system. Nonetheless, visual inspection of NOEs indicated the presence of rotational isomers arising from rotation around the C^5-N^5 bond (**Figure 13**). Formyl-CH_3 NOEs were observed for both the α and β anomers in this system, although

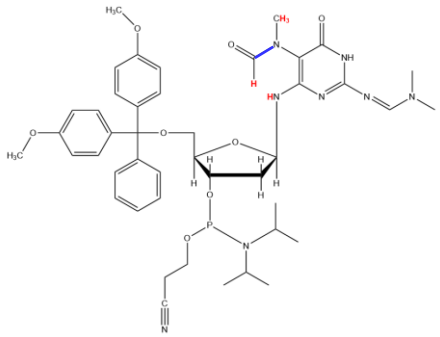


Figure 12. Comparison of the NOE peak strength between the N^6H-CH_3 and N^6H-Formyl protons was hypothesized to indicate whether rotation around the C^5-N^5 bond (shown in blue) occurred. Comparison of NOE peak strength between the CH_3-Formyl protons was hypothesized to indicate the presence of rotational isomers arising from rotation around the formyl bond.

NOE overlap prevented accurate integration. Interestingly, there were only two distinct chemical shifts of the formyl proton for each anomer, suggesting that: (1) rotation around the formyl bond does not occur in this system, (2) rotation around the formyl bond occurs rapidly in this system, leading to averaging of the formyl peaks, or that (3) rotation around the formyl bond does not cause an appreciable change in the chemical environment of the formyl proton.

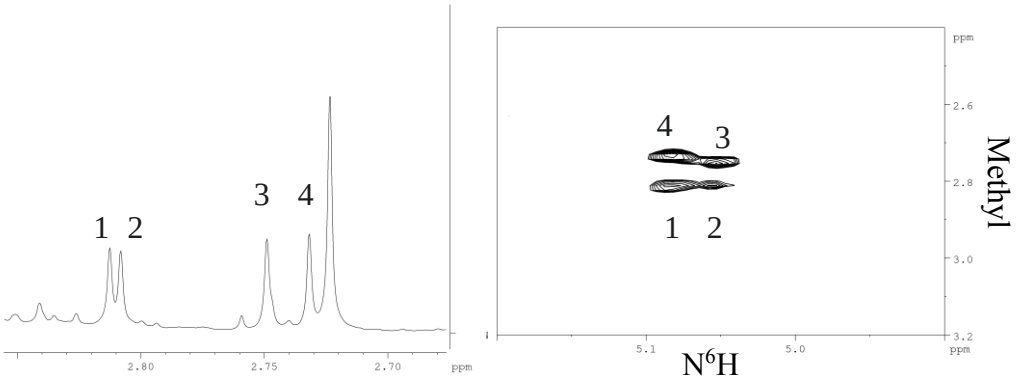


Figure 13. Comparison of the relative abundance of the β anomer and the CH_3-N^6H NOE peak strength revealed a discrepancy; while isomers 1 and 2 are clearly similar in abundance according to integration of the ^{1}H-NMR (left), the CH_3-N^6H NOE peak strength appears to be quite different (right).

Discussion

The MeFapy-dG lesion has thus far eluded full characterization in solution due to the complexity of the spectra, an observation which arises from multiple isomers of the MeFapy-dG adduct. The phosphoramidite represents the simplest system in which to identify both conformational and configurational isomers. Detection of the N^6H proton proved vital for further characterization of rotational isomers occurring in the MeFapy-dG adduct. Future research into Fapy-dG lesions in more biologically relevant samples should therefore include

experiments in a solvent that minimizes deuterium exchange, such as 9:1 H_2O:D_2O. Comparison of NOEs between the N^6H proton and other protons in the system was not sufficient for unambiguous assignment of all of the rotational isomers in the phosphoramidite system, indicating that other methods may be required in systems exhibiting substantial overlap.

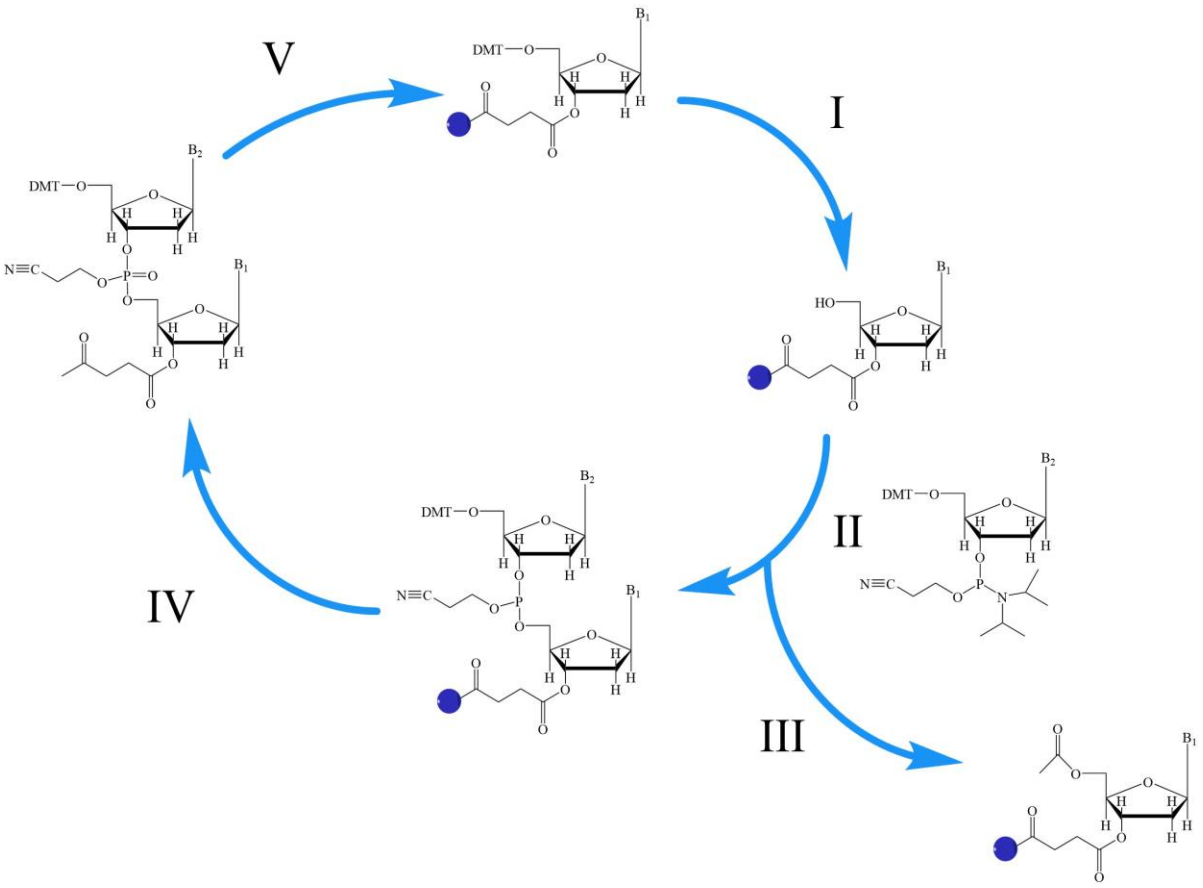
Dr. Robert Bruce Merrifield was awarded the Nobel Prize in 1984 for his development of solid-phase peptide synthesis.[159] This method of solid-phase synthesis solved a number of problems faced by scientists of the time who were working on synthesizing peptides and other biological polymers. Synthesis of these biopolymers was a long and arduous process, requiring many purification steps and often yielding very little sample. By binding one end of the biopolymer chain to a nonreactive solid support, excess reagents that failed to react during the addition step could be quickly separated from the growing chain by a simple washing step, preventing off-target reactions that formed undesired products. This drastically reduced the amount of time required to purify the product, often the most time consuming aspect of a synthesis, thereby increasing the speed at which scientists could synthesize large biopolymers. Indeed, this method was eventually automated, as Dr. Merrifield suggested it could be.

The elegant methods utilized in solid-phase oligodeoxynucleotide synthesis were pioneered by Dr. Robert Bruce Merrifield in 1963 in the solid-phase synthesis of peptides.[160] His research eventually culminated in the first synthesis of an enzyme, ribonuclease A,[161] and later evolved to form the powerful method used today to synthesize oligodeoxynucleotides: solid-phase oligonucleotide synthesis via phosphoramidites.[162]

Protecting groups are vital for high fidelity synthesis of oligodeoxynucleotides; it permits sequence specific incorporation of phosphoramidites into the growing oligodeoxynucleotide, including modified phosphoramidites. The phosphoramidite method of solid-phase

oligodeoxynucleotide synthesis relies on the selective removal of these protecting groups. Unlike biosynthesis of DNA, which proceeds in a 5′to 3′ direction, the solid-phase synthesis using the phosphoramidite method proceeds 3′ to 5′. Synthesis proceeds via a cycle comprised of deprotection, activation and coupling, capping, and oxidation steps (**Scheme 6**). The initial deprotection step is removal of the 5′ protecting group, [bis-(4-methoxyphenyl)phenylmethyl] (DMT), by treatment with acid. The resulting DMT cation is orange in colour, enabling spectrophotometric measurement of the amount of DMT released at the start of each cycle; from this measurement, quantification of the success of the preceding cycle's coupling step is estimated. After removal of the DMT group a second phosphoramidite, which is not bound to a solid support, is activated with a tetrazole catalyst, rendering the 3′ phosphorous susceptible to nucleophilic attack by the deprotected 5′ hydroxyl of the phosphoramidite bound to the solid support.

While the yield of each cycle tends to be very high, there will inevitably be some phosphoramidite bound to solid support with unreacted 5′ hydroxyl groups at the end of each coupling. Left unchecked, these chains would continue to grow, producing undesired oligodeoxynucleotides. This becomes particularly problematic when synthesizing long oligodeoxynucleotides; synthesis of a 100-nucleotide long product would inevitably produce truncated oligodeoxynucleotides comprised of numerous off-target sequences. To address this issue a capping step is required, in which unreacted 5′ hydroxyl groups are capped at the end of



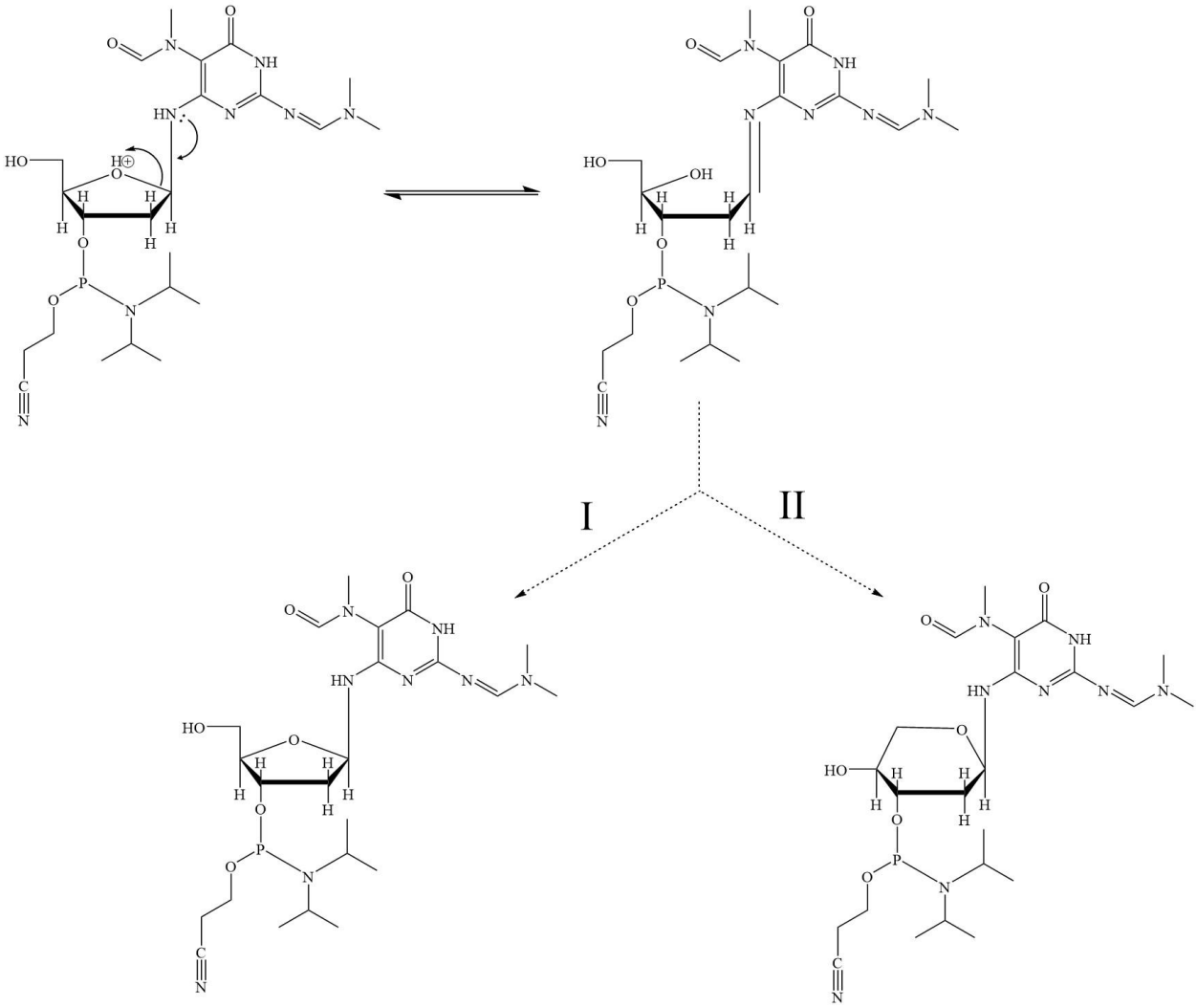
Scheme 6. The solid-phase phosphoramidite oligodeoxynucleotide synthesis cycle. I) Deprotection: The initial phosphoramidite, bound to a solid support, is acid-deprotected to yield a free 5' hydroxyl group. The DMT cation produced is bright orange in colour. II) Activation and Coupling: The second phosphoramidite is activated with tetrazole. The diisopropylamino group is displaced by the free 5' hydroxyl of phosphoramidite 1. III) Capping: Unreacted oligodeoxynucleotide 1 is capped to prevent further elongation in the next cycle, which would produce an oligodeoxynucleotide with a one nucleotide deletion. IV) Oxidation: The phosphite-triester is oxidized, converting it to a stable species suitable for further cycles. V) The cycle begins again, starting with deprotection of the 5' hydroxyl group. The cycle is repeated until the desired product is obtained.

each cycle to render them unreactive to further coupling. Introduction of acetic anhydride and *N*-methylimidazole acetylates the 5' hydroxyl group. After capping, the growing oligodeoxynucleotide can be oxidized to stabilize the phosphite-triester, which is unstable in acidic solutions. At this point the cycle can recommence, beginning with the acid-catalyzed removal of the 5' DMT group of the growing oligodeoxynucleotide. Measurement of the amount of DMT released serves as a measurement of the success of the previous coupling step; only products that were successfully coupled with a phosphoramidite in the previous cycle will release a DMT cation upon exposure to acid, producing a smaller amount of DMT at the start of the next cycle compared to the previous cycle.

Solid-phase oligodeoxynucleotide synthesis has revolutionized the manner by which scientists obtain DNA samples, enabling the production of specific sequences of high fidelity oligodeoxynucleotides with relative ease. The solid support enables 'washing' of the material between each cycle, removing unreacted material that plagued chemists synthesizing biopolymers. The synthesis of the MeFapy-dG phosphoramidite, pioneered by Dr. Plamen Christov,[157] provided a vital tool for further investigation of this lesion. Previous spectroscopic analysis of the phosphoramidite was largely limited to 1D NMR; this work further investigates the equilibrating isomers of the phosphoramidite, while also providing the first quantification of the anomeric composition of the MeFapy-dG phosphoramidite.

The sensitivity of the phosphoramidite required selection of a non-acidic solvent. In acidic environments the phosphoramidite may be deprotected by loss of the DMT protecting group; additionally, depurination of the MeFapy-dG base is also possible. Loss of the DMT group results in a rapid conversion from the furanose form of the deoxyribose ring to the pyranose form. Acid-catalyzed ring opening of the deoxyribose sugar may result in reclosure to

form either the 5-membered furanose ring or the 6-membered pyranose ring (**Scheme 7**). Reclosure to form the pyranose ring is favored, leaving the phosphoramidite unsuitable for incorporation into an oligodeoxynucleotide. Conversion from the furanose ring to the pyranose



Scheme 7. Acid catalyzed ring opening of the deoxyribose ring of the MeFapy-dG phosphoramidite occurs when the DMT protecting group is removed. Reclosure of the ring can result in the regeneration of the furanose species (I) or the formation of the pyranose species (II). The pyranose ring formation is favored over furanose ring formation.

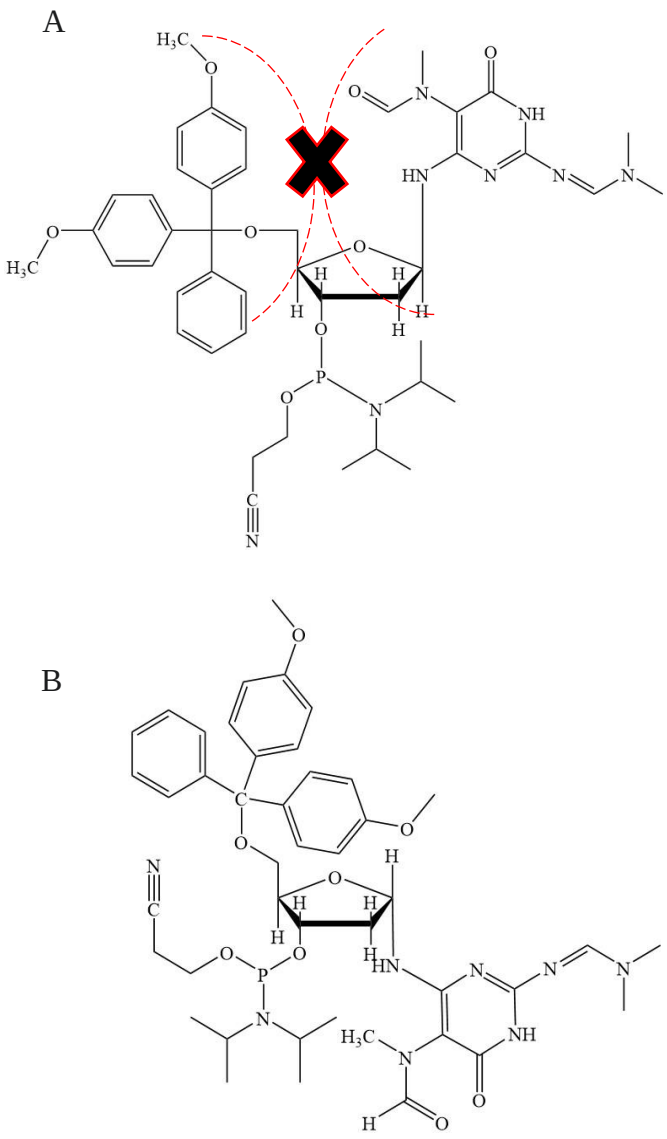
ring does not occur under biologically relevant conditions, where there is no free 5′ hydroxyl group of the MeFapy-dG lesion.

The TOCSY experiment (**Figure 10**) displayed eight spin systems, each corresponding to the deoxyribose ring of the phosphoramidite, indicating the presence of eight isomers of the MeFapy-dG phosphoramidite. The TOCSY and COSY experiments were jointly used to assign individual protons of the deoxyribose ring; a NOESY experiment was required to differentiate between the H2′ and H2′′ protons (**Figure 11**). The spin systems were split into two distinct groups separated by ~ 0.5 ppm with respect to the N^6H proton, which was visible in the spectrum due to a lack of deuterium exchange in the selected solvent. The large chemical shift difference was hypothesized to be due to anomerization of the phosphoramidite, which would have occurred when it was initially synthesized; anomerization was predicted to alter the chemical environment of the N^6H proton more drastically than the presence of rotational isomers. The two groups were present in a 40:60 ratio, similar to findings in a previous study which found a 40:60 α:β ratio in freshly annealed duplex.[133] A NOESY experiment was performed to unequivocally assign the anomeric identity of the two groups of the MeFapy-dG phosphoramidite. Identification of the H2′ and H2′′ protons was made based on the knowledge that the H2′ proton is spatially closer to the H3′ than the H2′′, resulting in a stronger H2′-H3′ NOE compared to the H2′′-H3′ NOE. Anomeric assignment was accomplished by determining whether the H1′ proton was spatially closer to the H2′ proton or H2′′ proton in the eight isomers. As was hypothesized, the splitting of the isomers into two groups was due to anomerization.

The α anomer was found to be the major species, contrary to the anomeric population determined by Christov and coworkers in oligodeoxynucleotides.[133] In the oligodeoxynucleotide the β anomer was the major species in freshly annealed duplex DNA and in duplex DNA that

had been allowed to equilibrate over several days; the β anomer was more heavily favored in duplex DNA that had been allowed to equilibrate. The difference between the experimental results obtained by Christov et al. and the experimental results obtained for the MeFapy-dG phosphoramidite are attributed to the presence of the 5′-[bis-(4-methoxyphenyl)phenylmethyl] (DMT) group in the phosphoramidite.

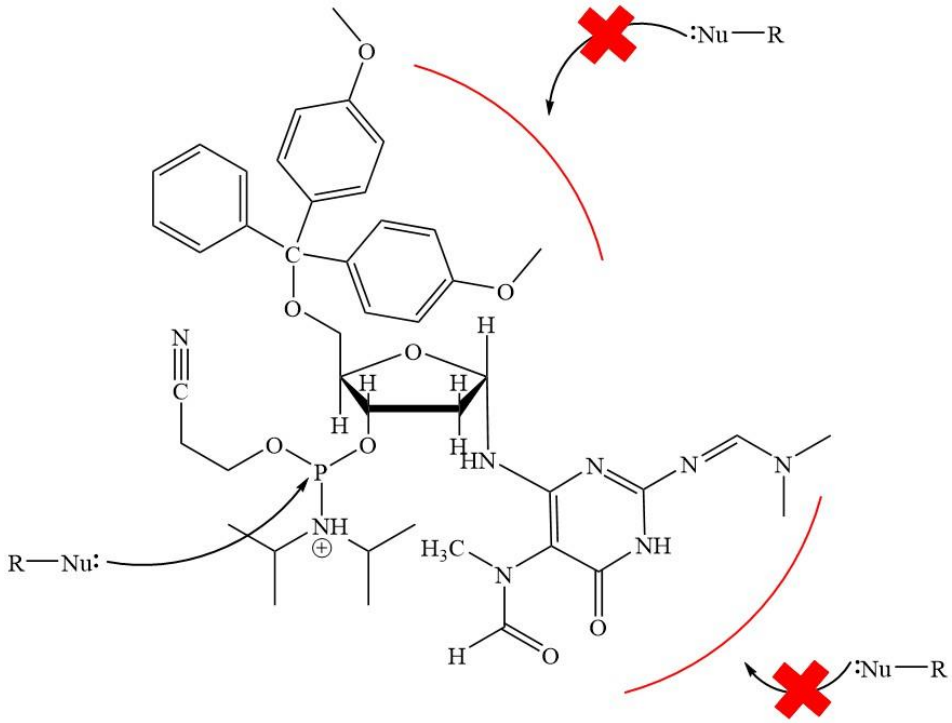
The bulky DMT group is hypothesized to be a steric strain on the phosphoramidite when the β anomer is formed, due to the placement of the MeFapy-dG base and DMT group on the same face of the deoxyribose ring (**Scheme 8**). By comparison, conversion to the α anomer is hypothesized to relieve this strain by moving the MeFapy-dG base to the opposite face of the deoxyribose ring (**Scheme 8**). In comparison to the 5′-DMT protecting group, the 3′ protecting group is relatively non-bulky, further favoring this conversion. The preference for the β anomer of MeFapy-dG previously observed in an oligodeoxynucleotide is attributed to the size of the 5′ and 3′ groups bound to MeFapy-dG; the 5′ and 3′ neighboring nucleotides in the oligodeoxynucleotide are more comparable to each other in bulk than the 5′ and 3′ protecting groups of the MeFapy-dG phosphoramidite. The preference for the β anomer in the oligodeoxynucleotide is hypothesized to be due to less steric strain compared to when the α anomer, which would impose greater steric strain on the 3′ neighboring nucleotide, is present. The increased predominance of the β anomer when the oligodeoxynucleotide is annealed to its complementary strand supports this hypothesis.



Scheme 8. A) The β anomer of the MeFapy-dG phosphoramidite contains the DMT group and MeFapy-dG base on the same face of the deoxyribose ring, increasing the steric strain on the molecule. B) In the α anomer of the MeFapy-dG the DMT group and MeFapy-dG base are located on opposite faces of the deoxyribose ring, reducing the steric strain compared to the β anomer.

Incorporation of the MeFapy-dG phosphoramidite during solid-phase oligonucleotide synthesis has proven difficult; a truncated failure sequence is consistently observed as the major product. Analysis of the most prevalent failure sequence by MALDI revealed the sequence resulted from failure to incorporate the MeFapy-dG phosphoramidite. The anomeric composition of the MeFapy-dG phosphoramidite provides a possible explanation for the low rates of incorporation. The α anomer of the MeFapy-dG phosphoramidite contains the MeFapy-dG base and the 3′ phosphorous on the same face of the deoxyribose ring. During the coupling step of the synthesis, the MeFapy-dG base blocks access to the 3′ phosphorous, sterically limiting nucleophilic attack by the 5′ hydroxyl group of the deprotected oligodeoxynucleotide chain (**Scheme 9**).

Evidence for rotational isomers was observed in the MeFapy-dG phosphoramidite. The inability to unequivocally identify rotation around the C^5-N^5 bond is largely attributed to overlap of NOEs, the number of species of the phosphoramidite, and the size of the phosphoramidite. Overlap of CH_3-N^6H and N^6H-formyl NOEs of MeFapy-dG isomers of the same anomeric identity precluded accurate measurements of NOE strength. This was exacerbated by the number of MeFapy-dG isomers; the signal of each individual isomer was decreased compared to what would have been predicted based on the amount of sample, creating a signal-to-noise issue. The relatively small amount of each isomer makes highly precise measurement of NOE strength even more crucial; even a small error in the measurement of NOE strength can cause a large variation in the perceived results.



Scheme 9. The α anomer of the MeFapy-dG phosphoramidite is predicted to generate more steric hindrance to an incoming nucleotide. The MeFapy-dG base may impede nucleophilic attack on the 3' side of the deoxyribose ring; the DMT group would act as a steric block to incoming nucleotides on the 5' side of the deoxyribose ring.

The weakness of the NOEs is also attributed to the small size of the molecule, which has a molecular weight of 856.96 g/mol, a difficulty compounded by the low viscosity of the solvent. These factors increased the rate of molecular tumbling, decreasing the molecular correlation time. The increased molecular motion results in an increased fluctuation in the magnetic field as the molecule tumbles in solution. The molecular motion is greater than υ_0, resulting in a more dominant double-quantum relaxation, ω_2. Consequentially, the off-diagonal NOEs of the phosphoramidite were positive.

Positive NOEs have a maximum strength of 50% compared to the strength that might be observed for a negative NOE. Positive NOE strength decreases as the molecule approaches a mass of ~1,000 g/mol ($\tau_c \upsilon_0 = 1$), at which point it goes through a zero point. As molecules become larger the NOEs gradually become more negative. The size of the MeFapy-dG phosphoramidite would have further weakened the NOEs; at a mass of 856.96 g/mol, the phosphoramidite was near this transition point between positive and negative NOEs.

Chapter III.

2,6-Diamino-4-hydroxy-*N*^{5}-(methyl)-formamidopyrimidine (MeFapy-dG) Adduct in Single Strand 5′-TXT-3′ Sequence

Materials and Methods

Oligodeoxynucleotide Synthesis. The unlabeled MeFapy-dG phosphoramidite was synthesized by Dr. Chanchal Malik as previously described; the ^{13}C-labeled MeFapy-dG phosphoramidite was synthesized using the same methodology with ^{13}C-labeled methyl iodide.[157] Incorporation of the MeFapy-dG phosphoramidite into the oligodeoxynucleotide 5'-$T^1X^2T^3$-3' (X = ^{13}C-MeFapy-dG or MeFapy-dG) was performed using a Millipore Expedite™ Nucleic Acid Synthesis System. The oligodeoxynucleotides were synthesized on a 1 μmol scale using Expedite reagents (Glen Research, Sterling, VA) with the standard synthetic protocols for the couplings of the unmodified bases. The couplings of the MeFapy-dG and ^{13}C-MeFapy-dG phosphoramidites were performed off-line for 30 min as was described by Elmquist et al.[163] Fapy-dG nucleotides can isomerize to the pyranose form when the 5′-hydroxyl group is deprotected during solid phase synthesis. Consequently, the MeFapy-dG and ^{13}C-MeFapy-dG phosphoramidites were incorporated into oligodeoxynucleotides by adjusting the deprotection time for removal of the 5′-dimethoxytrityl group of the MeFapy-dG or ^{13}C-MeFapy-dG nucleotides, thereby minimizing their rearrangement to the ribopyranose forms. Specifically, the DMT group of MeFapy-dG or ^{13}C-MeFapy-dG was removed on the synthesizer using 160 μL of Cl_3CCO_2H for a shorter time of 20 s to maximize yields of the desired MeFapy-dG or ^{13}CMeFapy-dG furanose isomer. Subsequent to the addition of the MeFapy-dG or ^{13}C-MeFapy-

dG nucleotides, the remainder of the nucleotide additions were performed online on the synthesizer using the standard protocols. The phosphoramidite was incorporated into the target sequence 5'-TXT-3' (X = ^{13}C-MeFapy-dG or MeFapy-dG). The completed oligodeoxynucleotides were cleaved from the solid support and deprotected by dissolving in 0.1 M NaOH (1.2 mL). The solution was stirred at room temperature for 24 h; it was then adjusted to pH 7 with 0.1 M acetic acid. Absorbance spectroscopy measurements at 260 nm were used to establish oligodeoxynucleotide concentrations. The calculated extinction coefficient of the unmodified trimer (5'-TGT-3') was 29300 L mol^{-1}cm^{-1}; the extinction coefficient of the modified trimer (5'-TXT-3'; X = ^{13}C-MeFapy-dG or X = MeFapy-dG) was not corrected for the presence of the MeFapy-dG lesion.[164]

HPLC Purification: Reversed-phase HPLC purification was performed using a Beckman Coulter system comprised of a System Gold 126 solvent module and a 168 diode array detector; 32 Karat (v. 7.0, Beckmann Coulter, Fullerton, CA) software was used for the acquisition, processing, and analysis of HPLC data. The diode array detector was configured to monitor both 254 nm and 360 nm wavelengths. Prior to injection, samples were centrifuged using Millipore Ultrafree®-MC-HV centrifugal filters (Durapore®-PVDF 0.45 µm). The samples were purified by reverse-phase HPLC utilizing a Gemini-NX 5um C18 110 Å (250 mm x 10 mm) column. The gradient was established using acetonitrile in a mobile phase of 0.1 M ammonium formate (pH = 7). Initial conditions were 3% acetonitrile for 2 min followed by an increase to 7% acetonitrile over 28 min. The final step of each run was an equilibration period of 2 min at 3% acetonitrile. The flow rate was 1.5 mL/min. The samples were re-purified via the same method as necessary. The oligodeoxynucleotides were desalted using a Sephadex C18 Light Sep-Pak

column (Waters Co., Milford, MA). Purity was verified by capillary gel electrophoresis (CGE) and MALDI.

Capillary gel electrophoresis (CGE): CGE was performed using a Beckman Coulter P/ACE MDQ CE; runs were performed at a voltage of 9.0 kV for 55 min with the UV detector set to 254 nm. The cartridge and sample storage temperatures were 30.0 °C and 25 °C, respectively.

Mass spectrometry: Mass spectrometric characterization was performed using a Voyager MALDI-TOF spectrometer. The ^{13}C-MeFapy-dG trimer (5'-TXT-3'; X = ^{13}C-MeFapy-dG) and MeFapy-dG trimer (5'-TXT-3'; X = MeFapy-dG) were prepared in a 3-hydroxypicolinic acid and ammonium citrate matrix; spectra were recorded in negative ion mode. All mass spectra were recorded to ±1 *m/z*.

NMR. (a) Sample Preparation. The trimer oligodeoxynucleotides were prepared in 100 mM NaCl and 10 mM NaH_2PO_4 (pH 8). The MeFapy-dG trimer sample was prepared at a concentration of 681 µM. The ^{13}C-MeFapy-dG trimer spectra were recorded at a concentration of 511 µM. In each instance the volume was 600 µL in a 5-mm sample tube. Spectra designed to observe non-labile protons were collected in buffer containing 99.996% D_2O. Spectra designed to observe labile protons were conducted in buffer containing 9:1 H_2O: D_2O; the Watergate pulse program was used to achieve water suppression.[165] All spectra were collected using Avance III NMR spectrometers equipped with cryogenic probes and precise temperature control (Bruker BioSpin, Inc., Billerica, MA).

(b) NMR Data Collection. Typical acquisition parameters for homonuclear experiments were as follows: 2K complex data points with 512 increments, 32 scans per FID, a sweep width

of 5250 Hz in both dimensions, relaxation delay of 2.0 s, and States-TPPI mode. The mode for the COSY and HSQC experiments was QF and Echo-Antiecho, respectively. Presaturation was sufficient for water suppression in samples dissolved in 99.996% D_2O; excitation sculpting was implemented for samples dissolved in 9:1 H_2O: D_2O. Nuclear Overhauser effect spectroscopy (NOESY)[149, 150] experiments were performed at a ^1H frequency of 900 MHz with 2048 real data points in the t_2 dimension and 512 real data points in the t_1 dimension. These spectra were obtained at mixing times of 300 ms, 500 ms, 800 ms, 1 s, and 1.3 s. The data was zero filled during processing to obtain matrices of 2048 x 1024 points. COSY[147] and $^1H-^1H$ total correlated spectroscopy (TOCSY)[148] spectra in 9:1 H_2O: D_2O were recorded at a ^1H frequency of 600 MHz. The mixing time for the TOCSY experiment was 70 ms. Spectra were obtained with 2048 real data points in the t_2 dimension and 512 real data points in the t_1 dimension; the data was zero filled during processing to obtain matrices of 2048 x 1024 data points. ^1H-^{13}C HSQC experiments were performed at a ^1H frequency of 600 MHz. HSQC spectral data[146] was zero filled to obtain matrices of 2048 x 1024 data points. Acquisition parameters for heteronuclear $^{13}C-^1H$ HSQC experiments were as follows: 2048 complex data points (sweep width = 7500 Hz), 1024 increments (sweep width = 7545.427 Hz), 50 scans per FID, garp decoupling, relaxation delay of 1.4 s, transmitter frequency offsets of 2817.7 Hz (^1H) and 5734.31 Hz (^{13}C), and Echo-and-Antiecho acquisition mode. Distortionless enhancement by polarization transfer (DEPT)[166] experiments were performed at a ^1H frequency of 600 MHz with 6,144 and 18,432 scans for the decoupled DEPT-135 and non-decoupled DEPT-135 experiments, respectively. Waltz decoupling was used in the decoupled DEPT-135. All NMR data were collected at a temperature of 278 ± 0.5 K.

(c) NMR Data Processing. NMR spectra were processed with the TOPSPIN software (Bruker BioSpin, Inc., Billerica, MA) and analyzed with the SPARKY software[158] and TOPSPIN. NMR spectra were integrated using TOPSPIN.

Results

Mass Spectrometry: Mass spectrometry of the ^{13}C-labeled trimer and unlabeled trimer confirmed the theoretical masses; the theoretical masses of the ^{13}C-labeled trimer and unlabeled trimer were 908.6 g mol^{-1} and 907.6 g mol^{-1}, respectively. The masses recorded for the ^{13}C-labeled trimer and unlabeled trimer were 908.3 g mol^{-1} and 907.3 g mol^{-1}, respectively.

^{13}C HSQC NMR. **(Figure 14)** shows the results of a ^{13}C HSQC experiment conducted on the 5'-TXT-3' (X = ^{13}C-MeFapy-dG) trinucleotide. The ^{13}C labeled methyl carbon signals of the MeFapy-dG adduct were observed as ten spectroscopically identifiable species, existing in equilibrium and in slow exchange on the NMR time scale. The higher field group of five ^{13}C labeled methyl carbons resonated in the 31-32 ppm chemical shift range, whereas the lower field group of five ^{13}C labeled methyl carbons resonated in the 35-36 ppm chemical shift range. Within each group there were four strong ^{13}C-HSQC signals and one weak ^{13}C-HSQC signal. Likewise, the protons associated with the ^{13}C-labeled methyl groups appeared in the spectrum as five methyl proton signals resonating in the 2.5-2.6 ppm range, and five methyl proton signals resonating in the 2.6-2.8 ppm range.

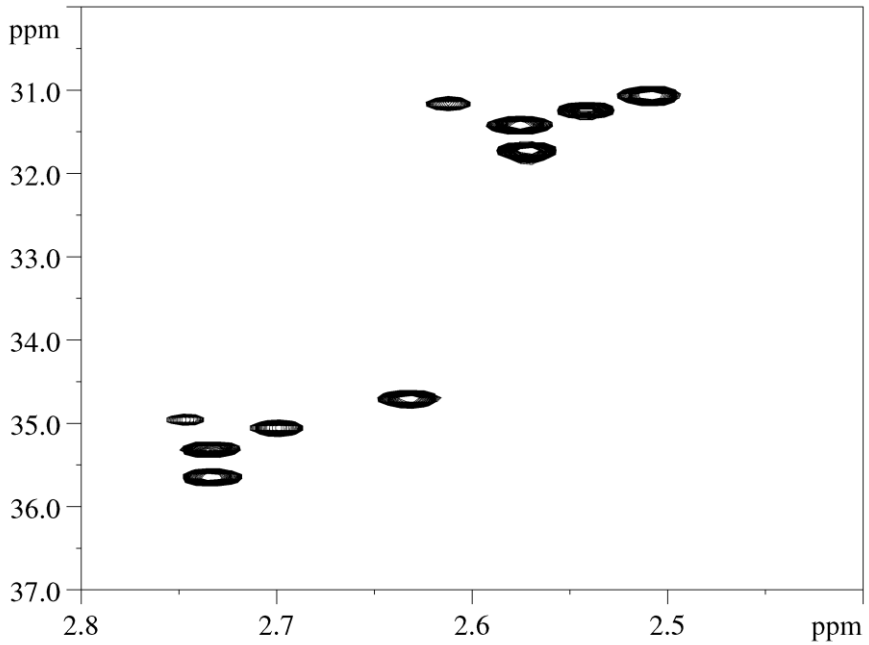


Figure 14. An HSQC experiment of 5'-TXT-3' (X = ^{13}C-MeFapy-dG) trinucleotide revealed ten isomers of MeFapy-dG separated on both the ^{1}H and ^{13}C axis. In each group of peaks there were four major and one minor species; the minor species was the most downfield peak of each group.

*Reprinted (adapted) with permission from Bamberger, S. N.; Malik, C. K.; Voehler, M. W., et al , **2018,** Chem. Res. Toxicol. Configurational and conformational equilibria of N^{6}-(2-deoxy-D-erythro-pentofuranosyl)-2,6-diamino-3,4-dihydro-4-oxo-5-N-methylformamidopyrimidine (MeFapy-dG) lesion in DNA, 31, 924-935. Copyright (2018) American Chemical Society.*

(a) Geometrical Isomerization of the Formyl Group. The 5'-TXT-3' (X = ^{13}C-MeFapy-dG) trinucleotide exhibited a less complex NMR spectra than larger oligodeoxynucleotides; as such, it was used as a model to characterize the equilibrium distribution of isomers of the MeFapy-dG lesion in DNA. The ten ^{13}C signals arising from the 5'-TXT-3' (X = ^{13}C-MeFapy-dG) trinucleotide in the ^{13}C HSQC spectrum were examined further utilizing a ^{1}H-decoupled

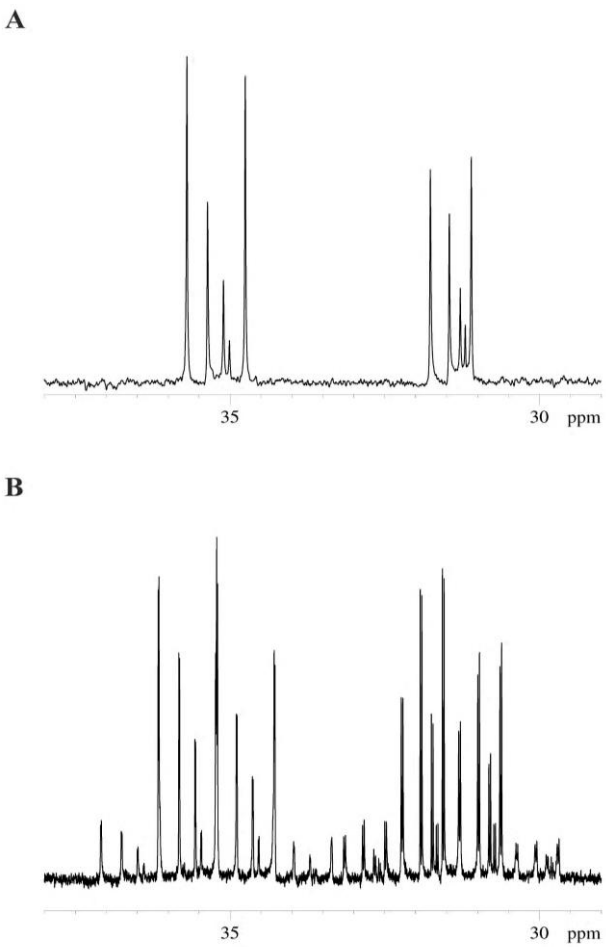
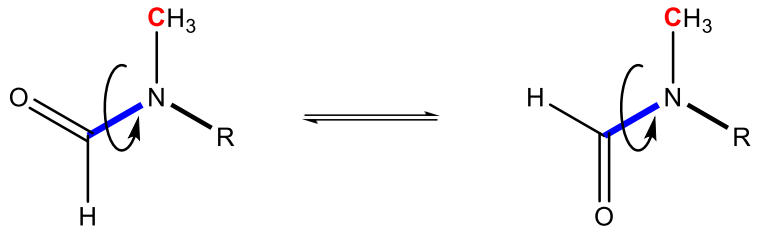


Figure 15. ***A****) A DEPT-135 experiment of the 5'-TXT-3' (X =* ^{13}*C-MeFapy-dG) trinucleotide confirmed that the MeFapy-dG adduct exists as a mixture of ten equilibrating isomers.* ***B****) Removal of the decoupling pulse during a DEPT-135 experiment on the 5'-TXT-3' (X =* ^{13}*C-MeFapy-dG) trinucleotide was used to obtain* $^{3}J_{CH}$ *values between the* ^{13}*C-labeled methyl carbon and formyl proton of MeFapy-dG; the downfield and upfield groups had* $^{3}J_{CH}$ *values of 1.7 Hz and 4.0 Hz, respectively.*

*Reprinted (adapted) with permission from Bamberger, S. N.; Malik, C. K.; Voehler, M. W., et al , **2018,** Chem. Res. Toxicol. Configurational and conformational equilibria of* N^6*-(2-deoxy-D-erythro-pentofuranosyl)-2,6-diamino-3,4-dihydro-4-oxo-5-N-methylformamidopyrimidine (MeFapy-dG) lesion in DNA, 31, 924-935. Copyright (2018) American Chemical Society.*

DEPT-135 ^{13}C NMR experiment (**Figure 15**). A ^{1}H-coupled spectrum collected under the same conditions showed that each of the ten ^{13}C signals was resolved into a quartet of doublets, arising from a combination of the larger $^{2}J_{CH}$ methyl coupling and a smaller $^{3}J_{CH}$ between the ^{13}C-labeled methyl carbon and the formyl proton. The $^{3}J_{CH}$ values of the downfield group of ^{13}C signals near 35 ppm differed from those of the upfield group near 31.5 ppm, measuring as 1.7 Hz and 4.0 Hz, respectively (**Figure 15**). These values allowed for assignment of the more upfield group of ^{13}C MeFapy-dG resonances (31-32 ppm) as *E* geometrical isomers, arising from the ^{13}C nucleus and formyl proton existing in a transoid relationship; the downfield group of ^{13}C MeFapy-dG resonances (35-36 ppm) was assigned as *Z* geometrical isomers in which the same two nuclei exist in a cisoid relationship (**Scheme 9**).



Scheme 10. Equilibrating geometrical isomers in which the same two nuclei exist in a transoid (left) or cisoid (right) relationship.

(b) Epimerization of the Deoxyribose α and β Anomers. A series of NOESY experiments established the presence of both α and β anomers of the MeFapy-dG deoxyribose ring. The α anomers were identified by their strong H1'-H3' NOEs and weaker H1'-H4' NOEs (**Figure 16**). The ^{3}J coupling between the MeFapy-dG N^{6}H proton and the deoxyribose H1' proton allowed the assignment of the former from COSY and TOCSY spectra; α anomers were also identified by NOEs between N^{6}H and deoxyribose H4' (**Figure 16**). No NOE cross-peaks were observed between the N^{6}H and deoxyribose H3' protons for the α anomers. The β anomers were identified

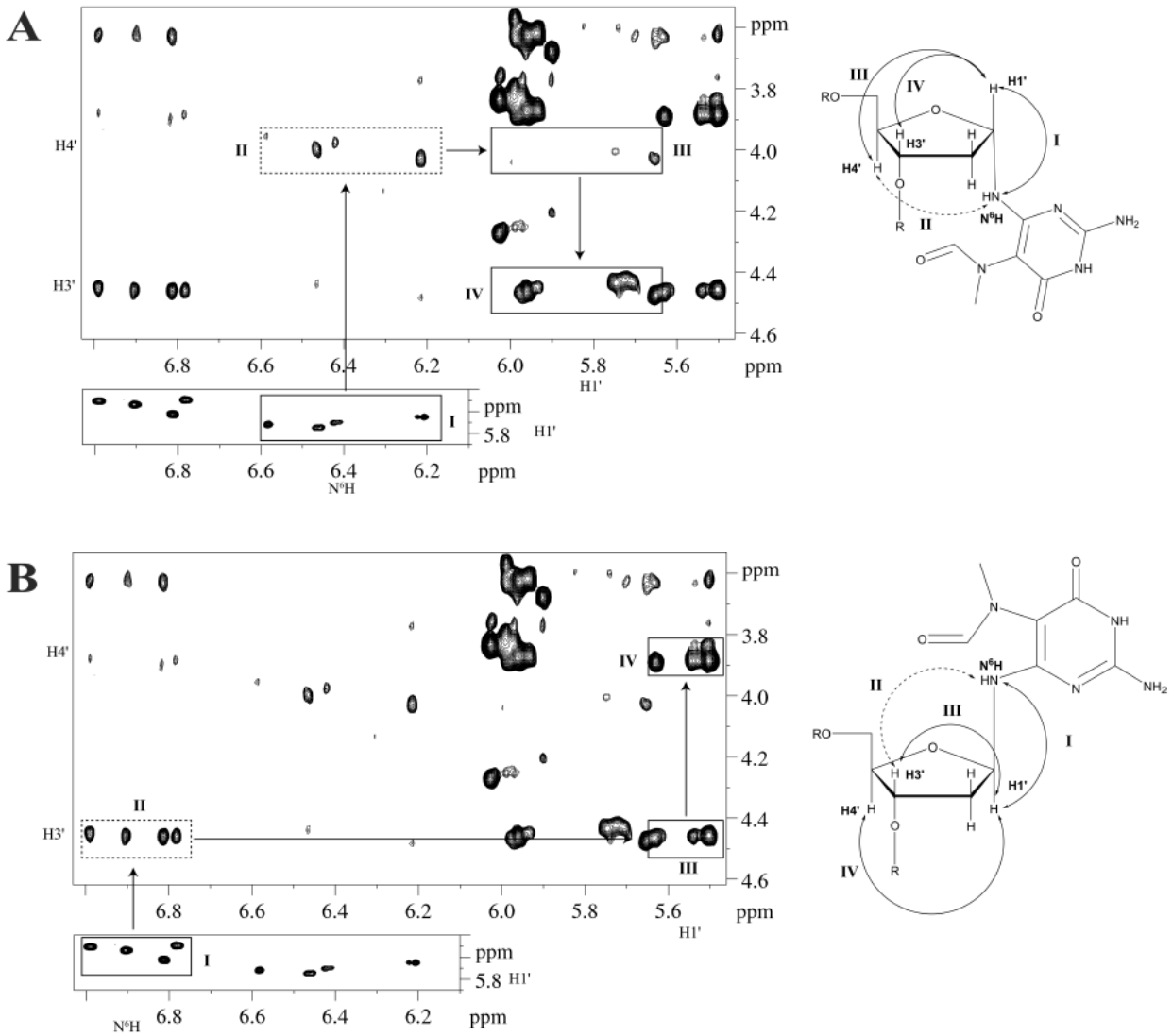


*Figure 16. A) NOESY spectrum of MeFapy-dG 5'-TXT-3' was obtained in buffer containing 9:1 H_2O:D_2O at 278 K. H1'-N^6H NOE peaks (I) of four isomers of MeFapy-dG were identified. The isomers were identified as α anomers by NOEs between N^6H and deoxyribose H4' (II), weak H1'-H4' NOEs (III), and strong H1'-H3' NOEs (IV). (B) NOESY spectrum of MeFapy-dG 5'-TXT-3' was obtained in buffer containing 9:1 H_2O:D_2O at 278 K. H1'-N^6H NOE peaks (I) of four species were identified. The isomers were identified as β anomers by NOEs between N^6H and deoxyribose H3' (II), strong H1'-H4' NOEs (III), and weak H1'-H3' NOEs (IV). Reprinted (adapted) with permission from Bamberger, S. N.; Malik, C. K.; Voehler, M. W., et al , **2018,** Chem. Res. Toxicol. Configurational and conformational equilibria of N^6-(2-deoxy-D-erythro-pentofuranosyl)-2,6-diamino-3,4-dihydro-4-oxo-5-N-methylformamidopyrimidine (MeFapy-dG) lesion in DNA, 31, 924-935. Copyright (2018) American Chemical Society.*

by their strong deoxyribose H1'-H4' NOEs and weaker deoxyribose H1'-H3' NOEs; the β anomers also exhibited strong N^6H-H3' NOEs and weak N^6H-H4' NOEs (**Figure 16**).

(c) Atropisomerism About the C^5-N^5 Bond. For the 5'-TXT-3' (X = ^{13}C-MeFapy-dG) trinucleotide assignments of R_a and S_a atropisomers arising from rotation around the C^5-N^5 bond were established by examining the relative intensities of NOEs between the N^6H, CH_3 and formyl protons. An inverse relationship was observed between N^6H-CH_3 and N^6H-formyl NOE intensities between species of the same anomers and conformation around the formyl bond; the isomer with a stronger N^6H-CH_3 NOE exhibited a weaker N^6H-formyl NOE (**Table 1**). Variations in NOE peak strength between isomers known to be the same anomer and same isomer with respect to rotation around the formyl bond, when adjusted for the population of each species, established the assignments of the atropisomers.

^{13}C isomeric species	***N^6*H-CH_3 NOE Intensity Relative to Atropomer Population[a]**	***N^6*H-CHO NOE Intensity Relative to Atropomer Population[a]**	**Integrated intensity of ^{1}H resonance[b]**	**Assignment of MeFapy-dG C^5-N^5 Atropisomer**
β *Z* isomer 1	57%	45%	50%	S_a
β *Z* isomer 2	43%	55%	50%	R_a
β *E* isomer 1	51%	56%	54%	S_a
β *E* isomer 2	49%	44%	46%	R_a
α *Z* isomer 1	62%	53%	58%	R_a
α *Z* isomer 2	38%	47%	42%	S_a
α *E* isomer 1	22%	45%	38%	R_a

α *E* isomer 2	78%	55%	62%	S_a

Table 1. The assignment of R_a *and* S_a *atropisomers about the* C^5-N^5 *bond was made on the basis of comparative NOE intensities between the* N^6H *proton and the methyl and the aldehyde protons of the MeFapy-dG adduct. It was necessary to correct these NOE intensities for the relative populations of each isomer, as determined from integration of* ^1H *spectra. For the* S_a *atropisomer, the* N^6H*-Me NOE exhibited greater intensity, whereas for the* R_a *isomer, the* N^6H*-CHO NOE exhibited greater intensity.*

[a]Determined from a NOESY experiment in buffer containing 9:1 H_2O:D_2O.

[b]Determined from integration of the N^6H *proton in a* ^1H *spectrum in buffer containing 9:1* H_2O:D_2O.

(d) Relative Abundances of Configurational and Conformational Isomers. The isomeric assignments of the MeFapy-dG lesion in the 5'-TXT-3' (X = ^{13}C-MeFapy-dG) trinucleotide enabled the calculation of the relative populations of each isomeric species (**Table 2**). Integration of the NMR spectra revealed that the α and β anomers existed in a 35:65 ratio, favoring the β anomer. For the β anomer, with respect to the C^5-N^5 bond, the R_a and S_a atropisomers were equally populated. However, the *Z* geometrical isomer of the formyl moiety was preferred. For the α anomer, the *E*-S_a isomer was present at 12%, whereas all other isomers

were present at 5-7%. The two minor species seen in the DEPT-135 and HSQC experiments were not detected in ^{1}H homonuclear experiments, presumably due to their low populations.

MeFapy-dG Isomer	**Relative Abundance in 5'-TXT-3' Sequence**
β *E* S_a	14%
β *E* R_a	13%
β *Z* S_a	21%
β *Z* R_a	21%
α *E* S_a	12%
α *E* R_a	7%
α *Z* S_a	5%
α *Z* R_a	7%

Table 2. Calculated populations of the eight major rotational and configurational isomers for the MeFapy-dG adduct in the 5′-TXT-3′trimer. The α and β anomers were present in a ratio of 35:65. The Z *and E geometrical isomers of the formyl group were favored in the β and α anomers, respectively.*

[a]Determined from a ^{1}H spectrum.

*Reprinted (adapted) with permission from Bamberger, S. N.; Malik, C. K.; Voehler, M. W., et al , **2018,** Chem. Res. Toxicol. Configurational and conformational equilibria of N^6-(2-deoxy-D-erythro-pentofuranosyl)-2,6-diamino-3,4-dihydro-4-oxo-5-N-methylformamidopyrimidine (MeFapy-dG) lesion in DNA, 31, 924-935. Copyright (2018) American Chemical Society.*

Discussion

Fapy-dG adducts are likely to exist in cellular DNA as a mixture of species,[50, 118, 127-136] which may be differentially recognized and processed by cellular enzymes, potentially impacting therapeutic outcomes in chemotherapy. The present data reveal that in the 5'-TXT-3' sequence the MeFapy-dG lesion exists as ten spectroscopically identifiable and distinct isomeric species (**Figure 14**). Eight of these may be identified unequivocally. Four of these distinct isomeric species in the 5'-TXT-3' sequence are α deoxyribose anomers, while the other four are β deoxyribose anomers. At equilibrium the β-MeFapy-dG anomer is favored, and the two anomers exist in a 35:65 α:β ratio. This corroborates the work of Christov et al. who,[133] utilizing an *Escherichia coli* Endonuclease IV (Endo IV)-based enzymatic assay,[138] showed that at equilibrium in duplex DNA the α and β MeFapy-dG anomers exist in a 20:80 α:β ratio.[133] The preference for the β anomer of the MeFapy-dG adduct is likely due to increased distortion necessary to accommodate the α anomer in DNA. There is no high-resolution structural data available for the α anomer of the MeFapy-dG adduct. However, it has been determined from molecular dynamics calculations restrained by NMR-derived distances and torsion angles that the α anomer of the AFB_1-Fapy-dG adduct distorts DNA, consistent with perturbations of the base stacking register of the DNA.[61] The α anomer of carba-Fapy-dG, in which the oxygen of the deoxyribose ring has been replaced by a CH_2 group, while highly destabilizing to duplex DNA, causes local perturbations without altering the overall structure of the duplex, presumably due to the smaller size of the adduct.[56, 132] It is established that the α anomer of the bulkier AFB_1-Fapy-dG adduct destabilizes duplex DNA when present.[61]

In addition to undergoing epimerization, the MeFapy-dG adduct may undergo rotation around the formyl bond, C^5-N^5 bond, and glycosidic bond. Isotope-edited NMR spectroscopy

provides a powerful approach toward evaluating these complex mixtures of isomeric species, *in situ*.[61, 136] Guengerich and Humphreys utilized site-specific ^{15}N isotope labeling in NMR studies to establish that MeFapy-dG existed solely as the N^5 regioisomer; it had previously been hypothesized that MeFapy-dG could also exist as the N^6 regioisomer.[136] In the present studies the CH_3 carbon of the MeFapy-dG moiety was isotope-labeled with ^{13}C, using site-specific DNA modification chemistry developed by Rizzo and co-workers.[157] This has enabled the first quantitative assessment of the levels of specific configurational and conformational isomers of the MeFapy-dG lesion in an oligodeoxynucleotide. In the 5'-TXT-3' sequence the MeFapy-dG lesion exists as eight spectroscopically identifiable and distinct isomeric species (**Table 2**). Four of these are β deoxyribose anomers, while the other four are α deoxyribose anomers. Each of the α and β anomers exists as two pairs of rotamers, one involving atropisomerism about the C^5-N^5 bond, and the other involving geometrical isomerism of the formyl bond.

Utilizing $^3J_{HH}$ coupling between the formyl proton and N^5 proton, Raoul et al. demonstrated in 1995 that *E* and *Z* geometrical isomers, with respect to rotation around the formyl bond, occur in the Fapy-Ade adduct; $^3J_{HH}$ couplings of 1.3 Hz and 11.7 Hz were attributed to the *Z* and *E* geometrical isomers, respectively.[137] The geometrical isomers of Fapy-Ade were determined to exist in a 1:5 *E*:*Z* ratio; the energy barrier to rotation from the *Z* to *E* geometrical isomer was calculated to be 75.5 kJ/mol at 370 K. Fapy-dG adducts in DNA tend to favor the *Z* isomer, presumably because the carbonyl oxygen proximal to the pyrimidine ring is stabilized by hydrogen bonding with the proton on N^6.[56, 137] More recently, we showed that, in the case of the AFB_1-Fapy-dG adduct, the equilibrium between *E* and *Z* isomers is controlled by major groove hydrogen bonding interactions.[59] The ability of the MeFapy-dG base to interconvert between *E* and *Z* geometrical isomers of the formamide is also established.[50, 128]

The $^{3}J_{CH}$ values of the *E* and *Z* geometrical isomers of MeFapy-dG were measured as 4.0 Hz and 1.7 Hz (**Figure 15**). We determined in the trimer sequence 5′-TXT-3′ the *Z* geometrical isomer of MeFapy-dG is only slightly favored over the *E* geometrical isomer. This differs from the ratio determined by Boiteux et al. for the MeFapy-dG base; they determined that the base consisted of 88% *Z* isomer and 12% *E* isomer.[135] We attribute this difference to an increase in steric restraints and stabilization of the E geometrical isomer in the trimer, compared to the free MeFapy-dG base.

Patra et al.[167] obtained crystal structures of ternary Pol:DNA:dNTP complexes between MeFapy-dG-adducted DNA template:primer duplexes and the Y-family polymerases human Pol η (hPol η) and *Sulfolobus solfataricus* P2 Pol IV (Dpo4). These structures of the hPol η and Dpo4 complexes at the insertion and extension stages, respectively, were representative of error-free replication. MeFapy-dG was in the *anti* conformation about the glycosidic bond and exhibited normal Watson-Crick base pairing with dCTP or dC. The formamido group of the MeFapy-dG adduct was perpendicular to its pyrimidine ring, rather than in the plane of the ring. Our findings corroborate the crystal structures obtained by Patra et al. The differences in distance between the N^6H-methyl protons and N^6H-formyl protons in the R_a and S_a isomers are predicted to be smaller when the formamide moiety is perpendicular to the pyrimidine ring, compared to if the formamido group was in the plane of the pyrimidine ring. This is reflected in the relative NOE peak strengths observed between isomers in the 5′-TXT-3′ trimer after adjustment for population abundance (**Table 2**).

Two additional minor species seen in the DEPT-135 (**Figure 15**) and HSQC (**Figure 14**) spectra could not be unequivocally identified. This was attributed to their low populations in the 5'-TXT-3' sequence, which precluded the observation of NOEs involving these species. One

hypothesis is that these minor species correspond to the *syn* isomers resulting from rotation around the glycosidic bond, although *syn* isomers have not been identified in other Fapy-dG lesions. The potential for *syn* rotamers about the glycosidic bond, albeit at low levels, is significant, because the *syn-anti* conformational interconversion of the glycosidic bond is relevant with respect to error-prone bypass of the MeFapy-dG lesion. In the *syn* conformation, the MeFapy-dG moiety faces into the DNA and may form hydrogen bonds with incoming dNTPs, leading to misinsertions. Wiederholt and Greenberg[168] proposed that the Fapy-dG lesion in the *syn* conformation may mis-code for dATP insertion during error prone replication, potentially leading to a G→T transversion.[138, 169] A similar mechanism could account for the ability of the MeFapy-dG lesion to cause G→T transversions. The bulkier CH_3 group of MeFapy-dG must be accommodated within the stacking of the DNA helix. This is likely disfavored thermodynamically, which may account for the low levels of the proposed *syn* isomers, as evidenced by the minor peaks in the ^{13}C-HSQC (**Figure 14**) and DEPT-135 (**Figure 15**) spectra.

To date, Fapy-dG lesions have not been found to undergo rotation around the glycosidic bond to produce *anti* and *syn* conformational isomers, although the potential exists. Elegant work by Gehrke et al. utilized a Fapy-dG analogue in which the oxygen of the deoxyribose ring was substituted with a carbon (carba-Fapy-dG), preventing epimerization.[57] The carba-Fapy-dG lesion was found to base pair in a non-Hoogsteen manner with an incoming dATP in *Bst* Pol I. Both the carba-Fapy-dG and dATP were in the *anti* conformation; a greater distortion of the DNA backbone was observed compared to the same environment with carba-Fapy-dG●dCTP. The ability of the carba-Fapy-dG to base pair with the incoming dATP while in the *anti* conformation was attributed to the greater flexibility of the lesion compared to dGTP.

Presumably, this allowed for a lower energy rotation that moved the carba-Fapy-dG●dATP out of planarity to hydrogen bond, although this also reduced base-stacking in the duplex.

The crystallographic studies of Egli and co-workers[167] found the MeFapy-dG lesion only in the *anti* conformation during error-free insertion of dCTP opposite MeFapy-dG by hpol η. The *anti* conformation of the glycosidic bond places the Fapy-dG moiety into the major groove, allowing hindered rotation about both the C^5-N^5 bond and formyl bond. Significantly, with respect to DNA replication, the MeFapy-dG formamido moiety is not involved with Watson-Crick base pairing with an incoming dNTP, providing a mechanism to explain the preference for error-free bypass.[122, 124]

Error-Prone Replication. If not repaired, MeFapy-dG lesions are mutagenic. The lesion inhibits DNA synthesis.[119-121] Earley et al.[122] determined that both the mutational rate and mutational spectrum of the MeFapy-dG lesion are sequence dependent in primate cells; alteration of the 3' and 5' neighbors of the lesion resulted in varying and diverse mutational spectra. They examined nine different site-specifically modified MeFapy-dG-containing oligodeoxynucleotides. In each instance, when replicated in primate cells, MeFapy-dG was strongly mutagenic, predominantly causing G to T transversions. Single and dinucleotide deletions were also found in specific sequence contexts. Single-nucleotide deletions occurred at the adducted site and one nucleotide downstream of the MeFapy-dG adduct.

Sequence Dependence. Sequence-specific differences in the populations of configurational and conformational species of the MeFapy-dG lesion could modulate the biological response to these adducts, accounting for the observation of variable mutations in different sequences.[122, 124] We therefore hypothesize that the origin of the diverse mutational spectra of MeFapy-dG arises from a dependence of isomeric populations on the sequence

context. Kalam et al.[169] reported that in the 5'-TXT-3' sequence, the mutational frequency of the structurally related Fapy-dG lesion was approximately 30%, whereas in the 5'-TGA-3' sequence it was approximately 8%. Sequence-dependent effects, both with respect to DNA repair and to error-prone bypass of the MeFapy-dG lesion, may be due to differences in steric strain and hydrogen bonding between the isomers of MeFapy-dG and its neighboring bases, resulting in different sequences favoring certain isomers. We have observed this for the AFB_1-Fapy-dG adduct.[59] With regard to the MeFapy-dG lesion, one can speculate that if the 3′ neighbor of MeFapy-dG is a purine, greater steric hindrance might be provided, disfavoring the α anomer more than a 3′ pyrimidine would. Potential hydrogen bonding between the α anomer MeFapy-dG base and its 3′-neighbor leading to stabilization of the anomer must also be considered. Likewise, hydrogen bonding between the β anomer and its 5′ and 3′ neighbors may also lead to certain rotational isomers being favored. These differing ratios of the MeFapy-dG isomers can, theoretically, result in different interactions between the adduct and the active sites of DNA polymerases.

Chapter IV.

2,6-Diamino-4-hydroxy-*N*^{5}-(methyl)-formamidopyrimidine (MeFapy-dG) Adduct in the 5'- C^1 A^2 T^3 X^4 A^5 T^6 G^7 A^8 C^9 G^{10}C^{11}T^{12}-3' Dodecamer Sequence

Materials and Methods

Oligodeoxynucleotide Synthesis. The ^{13}C-labeled MeFapy-dG phosphoramidite was synthesized by Dr. Chanchal Malik as previously described for the unlabeled MeFapy-dG phosphoramidite, using the same methodology and ^{13}C-labeled methyl iodide.[157] Incorporation of the MeFapy-dG phosphoramidite into the oligodeoxynucleotide 5'- C^1 A^2 T^3 X^4 A^5 T^6 G^7 A^8 C^9 G^{10}C^{11}T^{12}-3' (X = ^{13}C-MeFapy-dG) was performed using a Millipore Expedite™ Nucleic Acid Synthesis System. The oligodeoxynucleotide was synthesized on a 1 μmol scale using Expedite reagents (Glen Research, Sterling, VA) with the standard synthetic protocols for the couplings of the unmodified bases. The coupling of the ^{13}C-MeFapy-dG phosphoramidite was performed as described in the previous chapter. Subsequent to the addition of the ^{13}C-MeFapy-dG nucleotide, the remainder of the nucleotide additions were performed online on the synthesizer using the standard protocols. The completed oligodeoxynucleotide was cleaved from the solid support and deprotected by dissolving in 0.1 M NaOH (1.2 mL). The solution was stirred at room temperature for 24 h; it was then adjusted to pH 7 with 0.1 M acetic acid. Absorbance spectroscopy measurements at 260 nm were used to establish oligodeoxynucleotide concentrations. The calculated extinction coefficients of the unmodified dodecamer (5'-C^1 A^2 T^3 G^4 A^5 T^6 G^7 A^8 C^9 G^{10}C^{11}T^{12}-3') and its complementary strand were 114300 L mol^{-1}cm^{-1} and 117900 L mol^{-1}cm^{-1}, respectively; the extinction coefficient of the modified strand (5'-C^1 A^2 T^3

X^4 A^5 T^6 G^7 A^8 C^9 $G^{10}C^{11}T^{12}$-3'; X = ^{13}C-MeFapy-dG) was not corrected for the presence of the MeFapy-dG lesion.[164]

HPLC Purification: Reversed-phase HPLC purification was performed using a Beckman Coulter system comprised of a System Gold 126 solvent module and a 168 diode array detector; 32 Karat (v. 7.0, Beckmann Coulter, Fullerton, CA) software was used for the acquisition, processing, and analysis of HPLC data. The diode array detector was configured to monitor both 254 nm and 360 nm wavelengths. Prior to injection, samples were centrifuged using Millipore Ultrafree®-MC-HV centrifugal filters (Durapore®-PVDF 0.45 µm). The samples were purified by reverse-phase HPLC utilizing a Gemini-NX 5um C18 110 Å (250 mm x 10 mm) column. The gradient was established using acetonitrile in a mobile phase of 0.1 M ammonium formate (pH = 7). Separation was achieved via a method utilizing an increase from 5% to 10% acetonitrile over 30 min. The flow rate was 1.5 mL/min. The samples were re-purified via the same method as necessary. The oligodeoxynucleotides were desalted using a Sephadex G25 column (Waters Co., Milford, MA). Purity was verified by capillary gel electrophoresis (CGE) and MALDI.

Capillary gel electrophoresis (CGE): CGE was performed using a Beckman Coulter P/ACE MDQ CE; runs were performed at a voltage of 9.0 kV for 55 min with the UV detector set to 254 nm. The cartridge and sample storage temperatures were 30.0 °C and 25 °C, respectively.

Unmodified DNA oligodeoxynucleotides. Unmodified DNA oligodeoxynucleotides 5'-C^1 A^2 T^3 G^4 A^5 T^6 G^7 A^8 C^9 $G^{10}C^{11}T^{12}$-3' and 5'-$A^{13}G^{14}C^{15}G^{16}T^{17}C^{18}A^{19}T^{20}C^{21}A^{22}T^{23}G^{24}$-3' were obtained from Midland Certified Reagent Co. (Midland, Texas). Absorbance spectroscopy

measurements at 260 nm were used to establish oligodeoxynucleotide concentrations. Minor impurities were removed by HPLC, as needed.

Mass spectrometry: Mass spectrometric characterization was performed using a Voyager MALDI-TOF spectrometer. The MeFapy-dG dodecamer (5'-C^1 A^2 T^3 X^4 A^5 T^6 G^7 A^8 C^9 $G^{10}C^{11}T^{12}$-3'; X = ^{13}C-MeFapy-dG), unmodified DNA oligodeoxynucleotide 5'-C^1 A^2 T^3 G^4 A^5 T^6 G^7 A^8 C^9 $G^{10}C^{11}T^{12}$-3', and unmodified DNA oligodeoxynucleotide 5'-$A^{13}G^{14}C^{15}G^{16}T^{17}C^{18}A^{19}T^{20}C^{21}A^{22}T^{23}G^{24}$-3' were prepared in a 3-hydroxypicolinic acid and ammonium citrate matrix; spectra were recorded in negative ion mode. Mass spectra were recorded to ± 1 *m/z*.

NMR. (a) Sample Preparation. The single strand and duplex dodecamer samples were prepared in 100 mM NaCl, 10 mM NaH_2PO_4, and 50 µM Na_2EDTA (pH 7). The unmodified duplex was prepared at a concentration of 665 µM. The MeFapy-dG dodecamer was prepared at a concentration of 461 µM. Annealing of the unmodified duplex in 100 mM NaCl, 10 mM NaH_2PO_4, and 50 µM Na_2EDTA (pH 7) was achieved by heating the sample to 85 °C, followed by slow cooling to room temperature. Annealing of the modified duplex in 100 mM NaCl, 10 mM NaH_2PO_4, and 50 µM Na_2EDTA (pH 7) was achieved by mixing equimolar amounts of the modified MeFapy-dG dodecamer and its complementary strand at room temperature; the solution stood at room temperature for 24 h. Excess of either strand was removed by passage of the sample through a DNA grade hydroxyapatite column.

(b) NMR Data Collection. Acquisition parameters for the ^{1}H-^{13}C HSQC experiments were as follows: 2048 complex data points (sweep width = 7500 Hz), 1024 increments (sweep width = 4527.869 Hz and 4527.869 Hz for single strand and duplex, respectively), number of scans per FID = 24 (single strand) and 48 (duplex), Echo-Antiecho mode, relaxation delay = 1.4

s, transmitter frequency offsets of 2819.99 Hz (^{1}H) and 5433.23 Hz (^{13}C), and garp decoupling; the data was processed to obtain matrices of 2048 x 1024 data points. NMR data was collected at a temperature of 298 ± 0.5 K and 278 ± 0.5 K for single strand and duplex, respectively. COSY[147] spectra in 99.996% D_2O were recorded at a ^{1}H frequency of 800 MHz. Acquisition parameters for COSY experiments were as follows: 2048 real data points in the t_2 dimension and 512 real data points in the t_1 dimension, number of scans per FID = 16, sweep width = 9615.385 Hz, QF mode, relaxation delay = 1.5 s; the data was zero filled during processing to obtain matrices of 2048 x 1024 data points. Nuclear Overhauser effect spectroscopy (NOESY)[149, 150] experiments were performed at a ^{1}H frequency of 900 MHz. Acquisition parameters for NOESY experiments were as follows: 2048 real data points in the t_2 dimension and 512 real data points in the t_1 dimension, a mixing time of 250 ms, number of scans per FID = 128, sweep width = 9009.009 Hz, States-TPPI mode, relaxation delay = 1.5 s; the data was zero filled during processing to obtain matrices of 2048 x 2048 data points. For comparison purposes identical NOESY experiments were run with ^{13}C-decoupling.

(c) NMR Data Processing. NMR spectra were processed with the TOPSPIN software (Bruker BioSpin, Inc., Billerica, MA) and analyzed with the SPARKY software[158] and TOPSPIN. NMR spectra were integrated using TOPSPIN.

Results

Mass Spectrometry: Mass spectrometry of the unmodified dodecamer confirmed the theoretical mass of 3645.4 g mol^{-1}. The experimental mass recorded for the unmodified dodecamer was 3645.4 g mol^{-1}. Mass spectrometry of the ^{13}C-labeled dodecamer confirmed the theoretical mass of 3678.5 g mol^{-1}. The experimental mass recorded for ^{13}C-labeled dodecamer

was 3678.1 g mol^{-1}. Mass spectrometry of the complementary strand confirmed the theoretical mass of 3645.4 g mol^{-1}. The experimental mass recorded for the complementary strand was 3646.1 g mol^{-1}.

NMR Spectroscopy

Unmodified 5'-C^1 A^2 T^3 G^4 A^5 T^6 G^7 A^8 C^9 $G^{10}C^{11}T^{12}$-3' Duplex. The NOESY spectrum produced by the unmodified dodecamer duplex 5'-C^1 A^2 T^3 G^4 A^5 T^6 G^7 A^8 C^9 $G^{10}C^{11}T^{12}$-3' was unremarkable, allowing for sequential assignment of NOEs between H8/H6 to H1′ protons (**Figure 17**). No breaks in the two sets of sequential NOEs between H8/H6 to H1′ protons were observed at 298 K.

^{13}C HSQC NMR of single strand and duplex. The dodecamer 5'-C^1 A^2 T^3 X^4 A^5 T^6 G^7 A^8 C^9 $G^{10}C^{11}T^{12}$-3' (X = ^{13}C-MeFapy-dG) was examined using a ^{13}C HSQC experiment. Similar to what was observed in the MeFapy-dG trimer (5′-TXT-3′), ten isomers were observed in the HSQC of the single strand dodecamer (**Figure 18**). The ten species were also split into two distinct groups of peaks on the carbon axis, hypothetically due to rotation around the formyl bond. In the dodecamer, however, greater chemical shift dispersion was observed for the methyl proton signals, with the higher field set resonating between 2.6-2.9 ppm and the lower field set resonating between 2.8-3.1 ppm. Lowering the temperature resulted in peak broadening and the evolution of new peaks; the exception to this was the most downfield peak of each original group that was observed at 298 K. Raising the temperature to 308 K resulted in peak

strengthening of the four most upfield peaks in each group (**Figure 19**). When the dodecamer was annealed to its complementary strand 5'-A^{13} G^{14} C^{15} G^{16} T^{17} $C^{18}A^{19}$ T^{20} $C^{21}A^{22}T^{23}G^{24}$-3', the

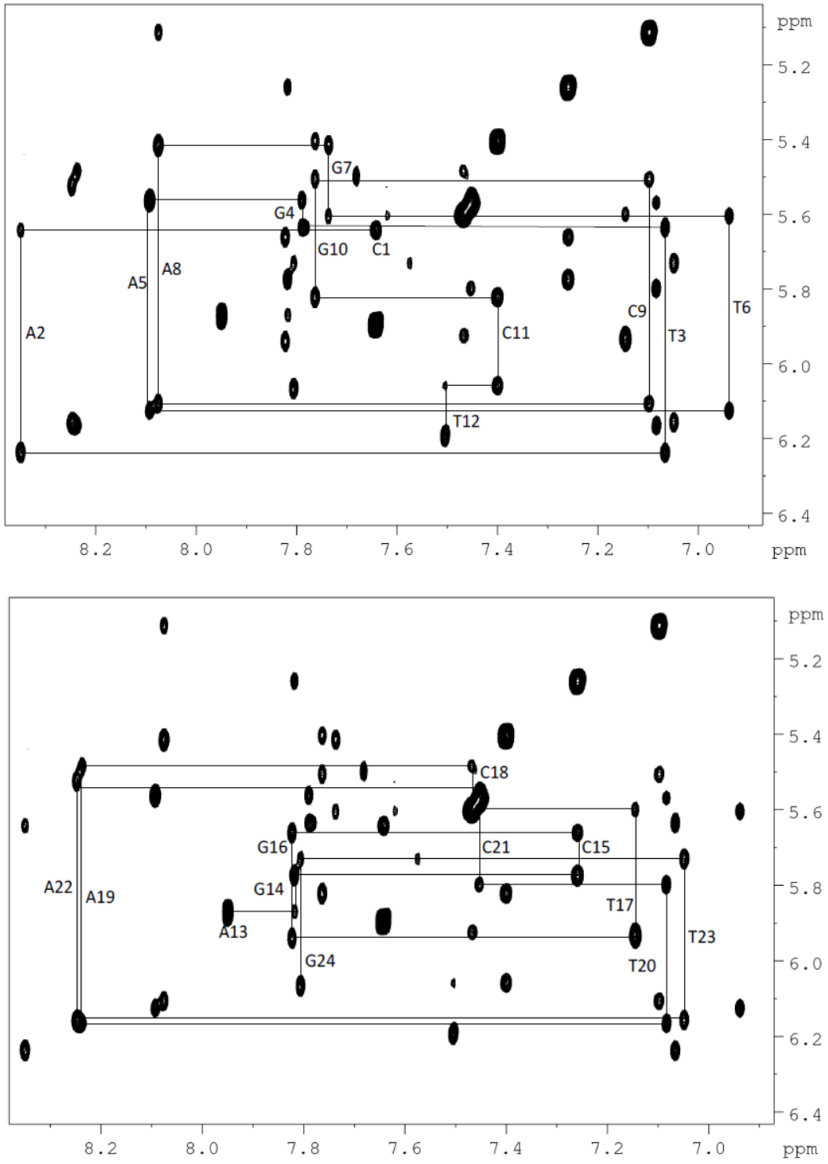


Figure 17. Sequential NOE connectivity between H8/H6 to H1' protons of the unmodified primary strand 5'- C^1 A^2 T^3 G^4 A^5 T^6 G^7 A^8 C^9 $G^{10}C^{11}T^{12}$-3' (top) and complementary strand (bottom).

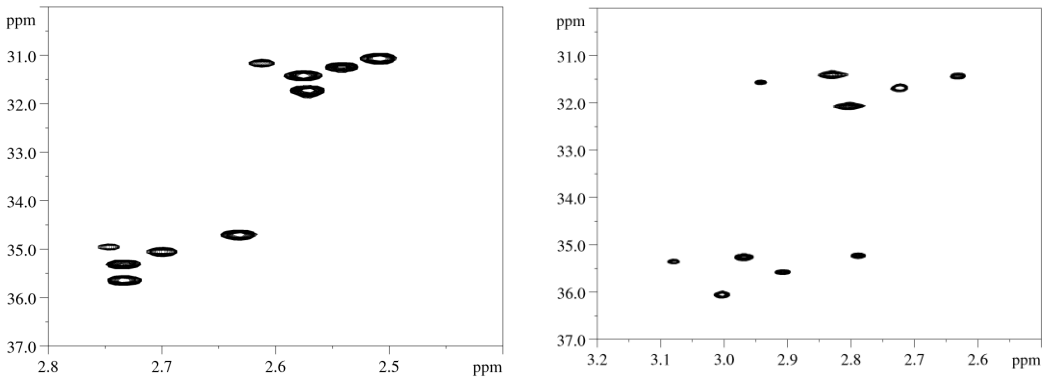


Figure 19. (Left) HSQC of the MeFapy-dG trimer 5'-TXT-3' (X = ^{13}C-MeFapy-dG) and (right) MeFapy-dG dodecamer 5'-C^{1} A^{2} T^{3} X^{4} A^{5} T^{6} G^{7} A^{8} C^{9} G^{10}C^{11}T^{12}-3' (X = ^{13}C-MeFapy-dG).

*Reprinted (adapted) with permission from Bamberger, S. N.; Malik, C. K.; Voehler, M. W., et al , **2018,** Chem. Res. Toxicol. Configurational and conformational equilibria of N^{6}-(2-deoxy-D-erythro-pentofuranosyl)-2,6-diamino-3,4-dihydro-4-oxo-5-N-methylformamidopyrimidine (MeFapy-dG) lesion in DNA, 31, 924-935. Copyright (2018) American Chemical Society.*

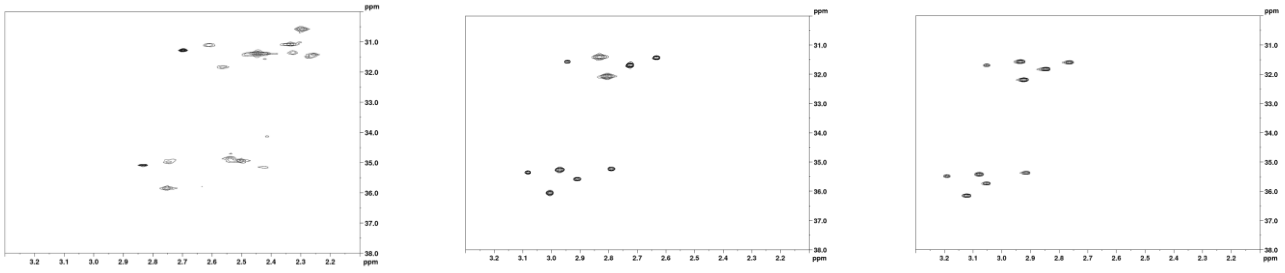


Figure 18. HSQC of the MeFapy-dG dodecamer 5'-C^{1} A^{2} T^{3} X^{4} A^{5} T^{6} G^{7} A^{8} C^{9} G^{10}C^{11}T^{12}-3' (X = ^{13}C-MeFapy-dG) at 278 K (left), 298 K (center), and 308 K (right).

*Reprinted (adapted) with permission from Bamberger, S. N.; Malik, C. K.; Voehler, M. W., et al , **2018,** Chem. Res. Toxicol. Configurational and conformational equilibria of N^{6}-(2-deoxy-D-erythro-pentofuranosyl)-2,6-diamino-3,4-dihydro-4-oxo-5-N-methylformamidopyrimidine (MeFapy-dG) lesion in DNA, 31, 924-935. Copyright (2018) American Chemical Society.*

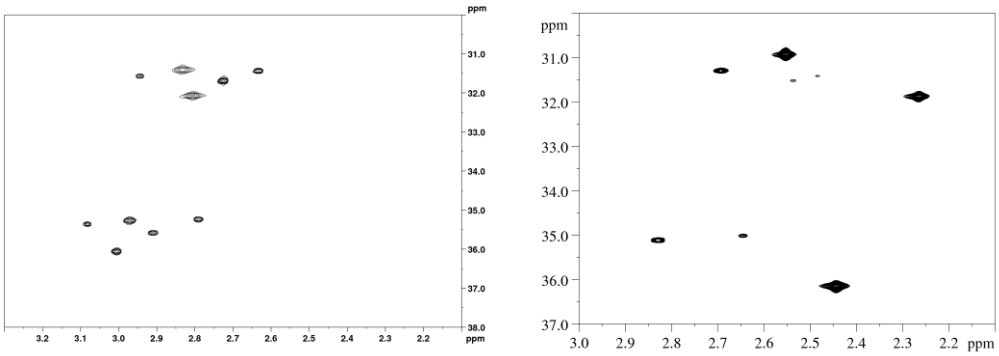


Figure 20. HSQC of the (left) single strand MeFapy-dG dodecamer 5'-C[1] A[2] T[3] X[4] A[5] T[6] G[7] A[8] C[9] G[10]C[11]T[12]-3' (X = [13]C-MeFapy-dG) and (right) of the duplex MeFapy-dG dodecamer 5'-C[1] A[2] T[3] X[4] A[5] T[6] G[7] A[8] C[9] G[10]C[11]T[12]-3' (X = [13]C-MeFapy-dG): 5'-A[13] G[14] C[15] G[16] T[17] C[18]A[19] T[20] C[21]A[22]T[23]G[24]-3'. The number of peaks was reduced from ten in single strand to eight in duplex; of the eight in duplex, two of the peaks were very weak.

*Reprinted (adapted) with permission from Bamberger, S. N.; Malik, C. K.; Voehler, M. W., et al , **2018,** Chem. Res. Toxicol. Configurational and conformational equilibria of N[6]-(2-deoxy-D-erythro-pentofuranosyl)-2,6-diamino-3,4-dihydro-4-oxo-5-N-methylformamidopyrimidine (MeFapy-dG) lesion in DNA, 31, 924-935. Copyright (2018) American Chemical Society.*

number of ^{13}C-HSQC signals was reduced to eight, two of which were very weak (**Figure 20**). Of the remaining six MeFapy-dG species, three appeared to be the predominant species.

Analysis of the 5'-C[1] A[2] T[3] X[4] A[5] T[6] G[7] A[8] C[9] G[10]C[11]T[12]-3' (X = [13]C-MeFapy-dG) Duplex. NMR analysis of the dodecamer duplex indicated six major species of MeFapy-dG were present in solution, of which three species were predominant. The methyl group of MeFapy-dG was identified by comparison of two NOESY experiments, identical in all parameters with the exception of a decoupling pulse (**Figure 21**). COSY and NOESY experiments of the dodecamer duplex at several temperatures indicated that lower temperatures were necessary for

characterization in this system; this was due to partial denaturing of the duplex at room temperature. At 298 K, NOEs between nucleotides were absent from C^1 to T^6. Due to the increased complexity of the NOESY spectra, the ^{13}C-labeled dodecamer was used to identify the methyl peaks, which were weakened by the number of isomers. To unequivocally assign the methyl groups, two NOESY experiments were run, identical in all parameters but for a decoupling pulse applied in one of the experiments. The decoupling pulse was used to decouple the ^{13}C-labeled carbon of the methyl group from its directly-bound protons, thereby eliminating the peak splitting of the methyl group observed in a non-decoupled NOESY. Comparison of the two spectra revealed multiple instances of convergence of two peaks from the non-decoupled NOESY to a single peak in the decoupled NOESY; the single peaks in the decoupled NOESY

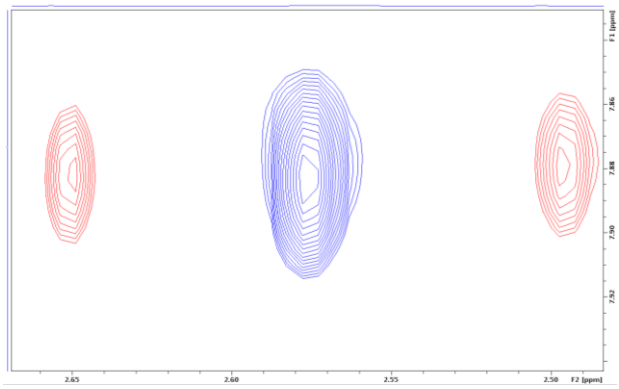


Figure 21. Overlap of two NOESY experiments on the dodecamer 5'-C^1 A^2 T^3 X^4 A^5 T^6 G^7 A^8 C^9 $G^{10}C^{11}T^{12}$-3' (X = ^{13}C-MeFapy-dG): 5'-A^{13} G^{14} C^{15} G^{16} T^{17} $C^{18}A^{19}$ T^{20} $C^{21}A^{22}T^{23}G^{24}$-3' duplex, identical in all parameters but for a decoupling pulse applied in one of the experiments, was used to unequivocally assign the methyl cross-peaks of the MeFapy-dG. The non-decoupled NOESY experiment (red) exhibited splitting of the MeFapy-dG methyl peaks due to the ^{13}C-labeled carbon. In the decoupled NOESY experiment (blue) the two peaks observed in the non-decoupled experiment resolved into a single peak of approximately twice the intensity.

were assigned as being the methyl group (**Figure 21**). Five of the six strongest methyl peaks were identified in the NOESY spectrum utilizing this method; of those five only two, the most upfield methyl signal from each group, had definitive NOE cross-peaks to neighboring nucleotides. These isomers had NOE cross-peaks between the methyl group and the 5' neighbor, T^3, and are tentatively assigned as β anomers.

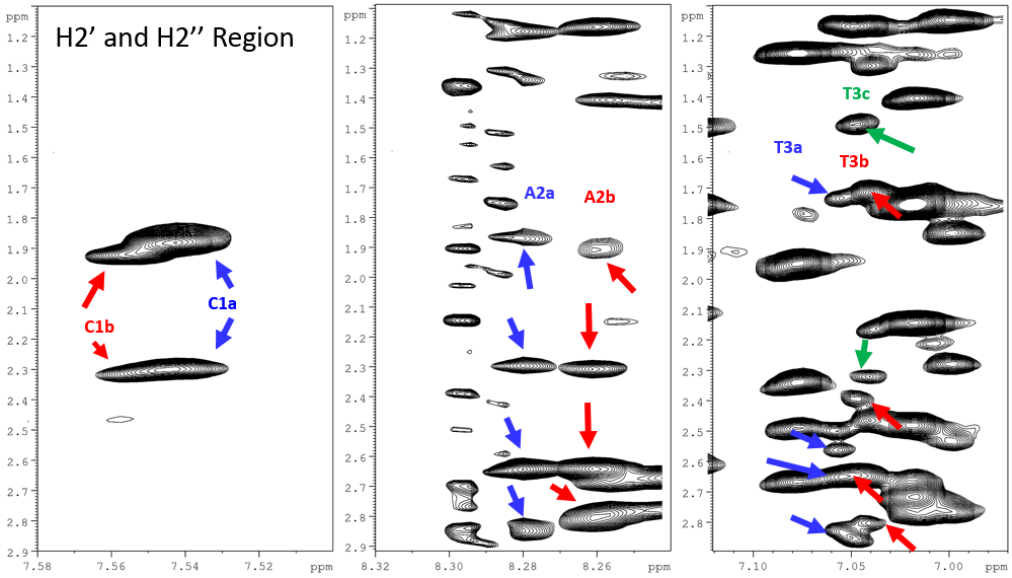


Figure 22. Peak splitting was observed in several neighboring nucleotides of the MeFapy-dG adduct in the dodecamer duplex of 5'-C^1 A^2 T^3 X^4 A^5 T^6 G^7 A^8 C^9 $G^{10}C^{11}T^{12}$-3' (X = ^{13}C-MeFapy-dG): 5'-A^{13} G^{14} C^{15} G^{16} T^{17} $C^{18}A^{19}$ T^{20} $C^{21}A^{22}T^{23}G^{24}$-3'. Doubling of cross-peaks was observed in C^1 (left panel) and A^2 (center panel). Tripling of the 5' neighbor of the MeFapy-dG adduct, T^3, was observed (right panel). This observed peak splitting indicates perturbations in normal duplex structure sufficient to cause two distinct chemical environments for nearby nucleotides. No cross-peaks between the MeFapy-dG adduct and its neighboring nucleotides were observed.

In addition to multiple species of MeFapy-dG, nearby nucleotides also exhibit peak splitting, indicating altered chemical environments for some neighboring nucleotides (**Figure 22**). Tripling of cross-peaks was also observed in a COSY experiment (**Figure 23**); the cytosine directly opposite of the lesion was split into three peaks, indicating the MeFapy-dG adduct causes three distinct chemical environments of nearby nucleotides in this system.

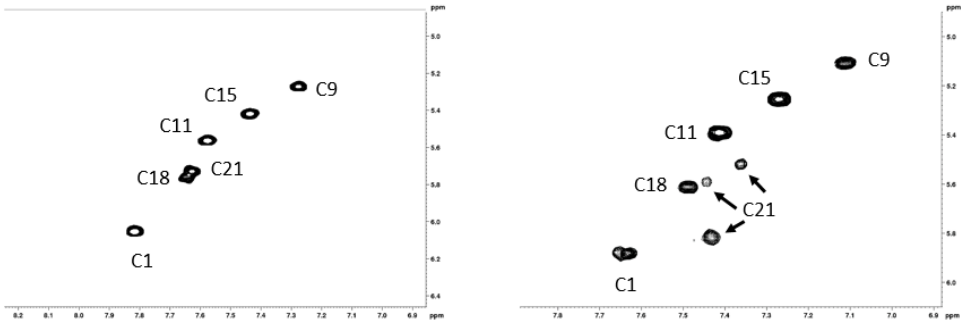


Figure 23. Six H5-H6 cross-peaks were observed in a COSY experiment of the unmodified duplex 5'- $C^1 A^2 T^3 G^4 A^5 T^6 G^7 A^8 C^9 G^{10} C^{11} T^{12}$-3': 5'-$A^{13} G^{14} C^{15} G^{16} T^{17} C^{18} A^{19} T^{20} C^{21} A^{22} T^{23} G^{24}$-3' , as expected. In the modified duplex 5'- $C^1 A^2 T^3 X^4 A^5 T^6 G^7 A^8 C^9 G^{10} C^{11} T^{12}$-3' (X = ^{13}C-MeFapy-dG): 5'-$A^{13} G^{14} C^{15} G^{16} T^{17} C^{18} A^{19} T^{20} C^{21} A^{22} T^{23} G^{24}$-3' nine H5-H6 cross-peaks were observed in a COSY experiment. The additional peaks arose from doubling of the ^{C1} H5-H6 cross-peak and tripling of the C^{21} H5-H6 cross-peak.

Partial spectroscopic assignment of non-exchangeable protons was achieved by comparison of spectra produced by the modified dodecamer duplex and spectra produced by the unmodified dodecamer duplex. A break in sequential NOE connectivity between the H6/H8 protons and H1′ protons was observed in the modified strand. NOE cross-peaks between T^3H1′-T^3H6 and T^6H1-A^5H8 were not present, indicating distortion of the duplex in the region of the adduct. NOE connectivity resumed at A^5H1′-A^5H8 and continued unbroken to the end of the strand. No NOEs between X^4H1′ and other nucleotides were observed. A break in sequential

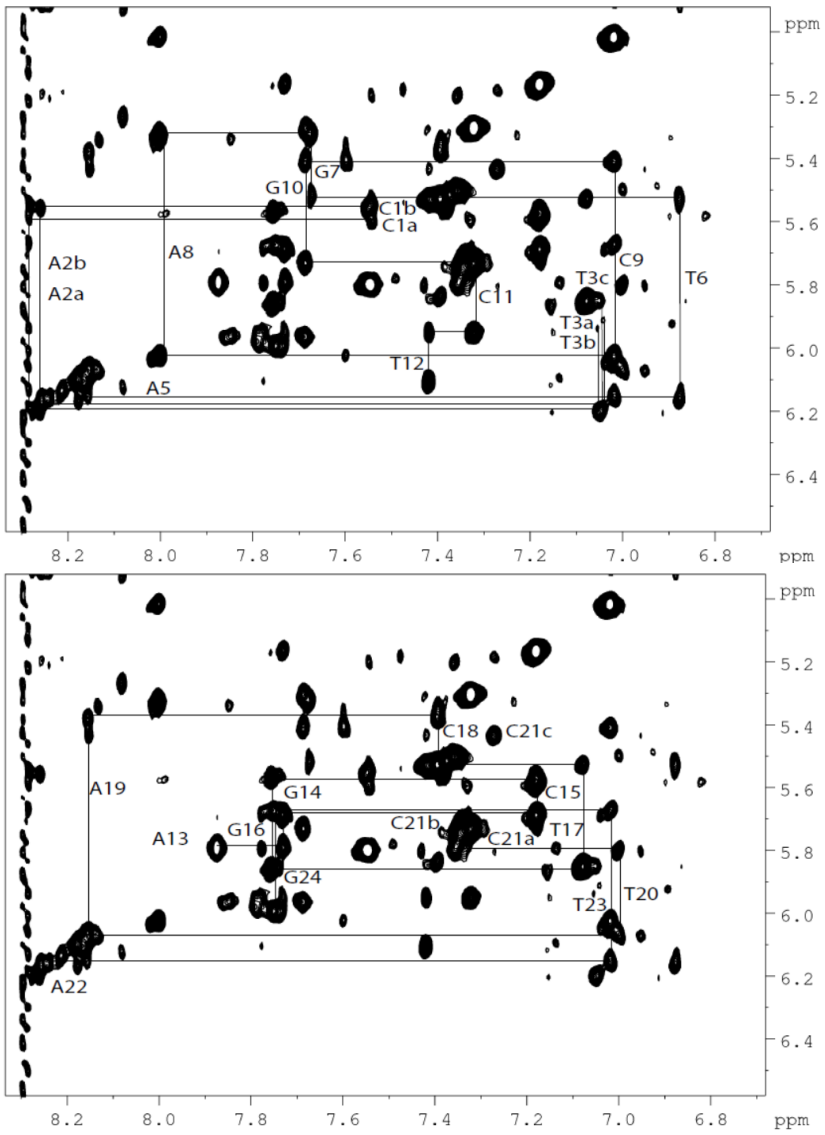


Figure 24. Sequential NOE connectivity between H8/H6 to H1' protons of the modified dodecamer 5'- C^1 A^2 T^3 X^4 A^5 T^6 G^7 A^8 C^9 $G^{10}C^{11}T^{12}$-3' (X = ^{13}C-MeFapy-dG) (top) and unmodified complementary strand 5'-A^{13} G^{14} C^{15} G^{16} T^{17} $C^{18}A^{19}$ T^{20} $C^{21}A^{22}T^{23}G^{24}$-3' (bottom). A break is observed in the primary strand at T3, the 5' neighbor of the MeFapy-dG adduct. The walk resumed with the NOE cross-peak A^5H1'-A^5H8 and continues unbroken for the remainder of the strand. Peak splitting is observed for nucleotides C^1-T^3. The complementary strand experiences a break at C^{21}, the nucleotide directly opposite of MeFapy-dG. T^3 and C^{21} exhibit the greatest peak splitting, with each nucleotide existing in three distinct chemical environments.

NOE connectivity of the complementary strand was observed at C^{21}, the nucleotide directly opposite of MeFapy-dG (**Figure 24**). C^{21} exhibited peak splitting similar to that seen for T^3, indicating three distinct chemical environments were being imposed on nearby nucleotides by MeFapy-dG. Signals of the formyl group of MeFapy-dG were unable to be definitively assigned. Coupled with the lack of NOE cross-peaks between the deoxyribose protons of MeFapy-dG and neighboring nucleotides, the weakness of the MeFapy-dG methyl group NOEs, and congested spectra, further characterization of the MeFapy-dG adduct in this system was precluded.

Discussion

In a single strand dodecamer the MeFapy-dG adduct existed as a mixture of ten isomers that were split into two distinct groups (**Figure 18**). These ten peaks are hypothesized to correspond to the isomers identified in the trimer sequence. Unlike in the trimer sequence, where the more downfield group of methyl peaks corresponded to the *Z* isomers and comprised 55% of the total isomers, the upfield group of methyl peaks in the dodecamer comprised 60% of the total isomers, indicating that in the dodecamer the *E* isomer might be favored over the *Z* isomer. This apparent preference in duplex remained unchanged when the duplex was annealed to its complementary strand; the upfield group of methyl peaks and downfield group of methyl peaks comprised 60% and 40% of the isomers, respectively. The most downfield methyl signal in each group did not exhibit the broadening observed in the other four methyl peaks; these downfield methyl peaks are hypothesized to be the two unidentified minor species observed in the trimer sequence, which were also the most downfield methyl signals in each group. The remaining four major species are tentatively assigned as arising, in part, from anomerization

(**Figure 20**). This assignment is based on NOEs observed between two of the species of MeFapy-dG and the 5' neighboring thymine. No NOEs were observed between the remaining two isomers and either neighboring nucleotides; these isomers are tentatively assigned as the α anomers. The integration of these proposed α and β anomers supports their assignments; the β:α ratio was 64:36, in agreement with previous experiments.[133]

The four observed species are also hypothesized to have originated from rotation around the formyl bond. This prediction is based on modeling of MeFapy-dG in this system, in which the 5' neighbor is a thymine and the 3' neighbor is an adenine. No favorable interactions were observed between MeFapy-dG and the 5' thymine, regardless of rotational conformation. A favorable interaction was observed only when the MeFapy-dG lesion existed as the β *Z*-R_a isomer. A favorable electrostatic interaction could be formed between the formyl oxygen of MeFapy-dG and the amino group of the 3' neighboring guanine (**Figure 25**). Rotation around the formyl bond to form the *E* isomer is not be predicted to form unfavorable interactions with neighboring nucleotides. Due to the smaller size of the formyl group, conversion between the *Z* and *E* isomers is predicted to be more favorable that rotation around the C^5-N^5 bond, which would require greater energy and fail to produce favorable interactions between MeFapy-dG and the 5' thymine.

Modified Duplex Stability. The instability of the modified duplex at 298 K indicates that the lesion is highly destabilizing when located near the terminal ends of the duplex. The lack of observed NOEs at 298 K from C^1-T^6 of the primary strand and from A^{19}-G^{24} of the complementary strand indicated the duplex was undergoing denaturing at room temperature; lowering the temperature restored the lost NOEs. At lower temperatures splitting of nucleotides was observed for the three nucleotides 5' of the lesion. This observed splitting could arise due to

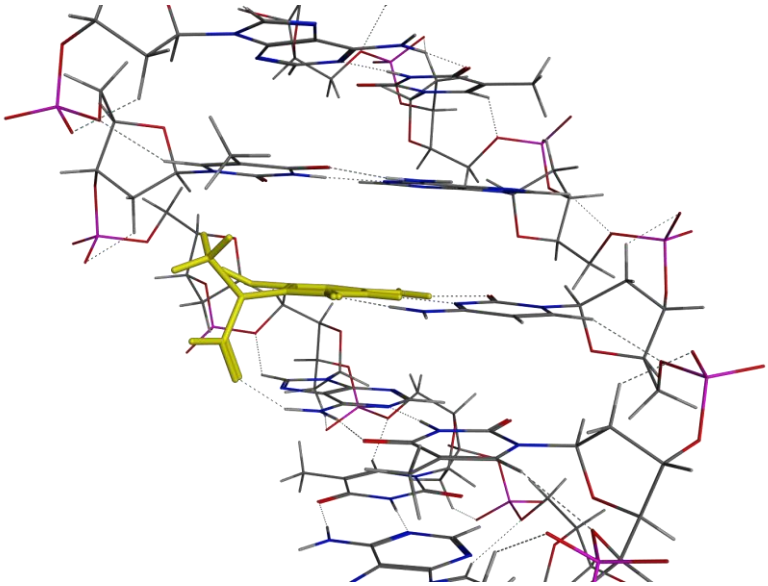


Figure 25. In the dodecamer duplex system 5'-C[1] A[2] T[3] X[4] A[5] T[6] G[7] A[8] C[9] G[10]C[11]T[12]-3' (X = [13]C-MeFapy-dG): 5'-A[13] G[14] C[15] G[16] T[17] C[18]A[19] T[20] C[21]A[22]T[23]G[24]-3' the Z-Ra *isomer is predicted to be favored when the MeFapy-dG exists as the β anomer. The* Z-Ra *isomer is capable of forming a favorable interaction between the formyl oxygen of MeFapy-dG and the amino group of the 3' neighboring guanine.*

the increased disorder at the ends of a DNA duplex; with fewer neighboring 5' hydrogen bonds in this system, in which MeFapy-dG is located near the terminal portion of the duplex, compared to a system in which MeFapy-dG is more centrally situated, disruption of the normal duplex structure is relatively favorable. Distortion of the duplex in the terminal region is more energetically favorable in this system than it would be if the lesion was located in the center of the duplex. The increase in peak splitting observed in T^3, the 5' neighbor of MeFapy-dG, suggests that the splitting is caused by multiple equilibrating isomers of MeFapy-dG in duplex DNA. The peak splitting seen for nucleotides C^1 and A^2 is hypothesized to be due to anomerization. The α and β anomers of MeFapy-dG are the isomers most likely to create a large disruption in DNA duplex structure, creating multiple chemical environments for nucleotides not

directly adjacent to the adduct. The presence of an α anomer nucleotide has been found to be highly destabilizing in duplex DNA, requiring distortion of the DNA backbone to accommodate it. Introduction of an α-adenosine into a decamer duplex induced an 18° kink of the helical axis into the major groove.[170] Indeed, a study incorporating a stabilized analogue of the α anomer of Fapy-dG (carba-Fapy-dG) found that the α-carba-Fapy-dG paired opposite of any nucleotide was destabilizing, presumably due to a lack of favorable interactions between the α anomer and any of the canonical bases.[132]

Repair of Duplex DNA Containing MeFapy-dG. Beranek and co-workers[118] established that the MeFapy-dG lesion was persistent in DNA. The amount detected increased over a period of 72 h; the amount of its precursor lesion, N7-Me-dG, decreased from the initial measurement at 12 h after exposure over the same period of time.[118] Kadlubar and coworkers determined that the MeFapy-dG lesion was persistent over a period of 21 d after exposure to *N*-methylnitrosourea.[50] The amount of N7-Me-dG decreased over 150-fold from the initial measurement at 2 h to the final measurement at 21 d. The abundance of MeFapy-dG, while low, was found to be stable for 21 d, indicating that the lesion may be difficult to repair.

The repair of MeFapy-dG by cellular glycosylases is dependent upon the identity of the opposing base in DNA. Asagoshi et al.[171] treated substrates containing MeFapy-dG:N pairs (N = A, G, C, T) as well as a thymine glycol (Tg):A pair, with the repair enzymes Endonuclease III (Endo III), Endonuclease VIII (Endo VIII), and a mouse Endo III homologue (mNth1). The activity of Endo III for MeFapy-dG was highest when opposite G and A (activities relative to Tg = 0.55) while the activity of Endo VIII was highest when opposite G (activity relative to Tg = 0.41). The lowest activity for both enzymes was with C opposite of MeFapy-dG (activities relative to Tg = 0.05 (Endo III) and 0.06 (Endo VIII)). In contrast, mNth1 recognized all

MeFapy-dG pairs equally well, with activity comparable to that observed with Tg. The results obtained in the nicking assay were further substantiated by the analysis of the Schiff base intermediate using $NaBH_4$ trapping assays. These results indicate that *Escherichia coli* and mammalian Tg glycosylases have a potential activity to recognize MeFapy-dG. However, as demonstrated for MeFapy-dG:C pairs with Endo III and Endo VIII, their distinctive activities with respect to the base opposite of MeFapy-dG indicates unequal participation in the repair of the MeFapy-dG lesion in cells. Wiederholt et al.[172] described interactions of DNA containing the structurally related Fapy-dG lesion with Fpg and MutY. They concluded that Fpg excises Fapy-dG efficiently, similar to 8-oxodG.

Chapter V.

2,6-Diamino-4-hydroxy-*N*^{5}-(methyl)-formamidopyrimidine (MeFapy-dG) Adduct in the 5'-G^1 C^2 T^3 A^4 G^5 T^6 X^7 G^8 G^9 T^{10}C^{11}C^{12}-3' Dodecamer Sequence

A former member of the Stone lab, Richard Dempster, gathered NMR data on a sample in which the MeFapy-dG lesion was incorporated into a relatively G:C rich dodecamer sequence (5'-G^1 C^2 T^3 A^4 G^5 T^6 X^7 G^8 G^9 T^{10}C^{11}C^{12}-3', X = MeFapy-dG). All experiments were conducted on modified and unmodified duplex structures; no data on the modified single strand was available.

Materials and Methods

Experimental procedures were performed as described in previous sections of this document. Sample preparation and purification was recorded by Richard Dempster to have been performed by Dr. Liang Li, another former member of the Stone lab.

Results

Unlike what was observed in the 5'- C^1 A^2 T^3 X^4 A^5 T^6 G^7 A^8 C^9 G^{10}C^{11}T^{12}-3' sequence, the 5'-G^1 C^2 T^3 A^4 G^5 T^6 X^7 G^8 G^9 T^{10}C^{11}C^{12}-3' dodecamer exhibited greater stability at 298 K, permitting experiments to be performed at temperatures ranging from 278 K to 298 K. No melting of the duplex was observed at 298 K. No ^{13}C-labeled sample was available for analysis to determine the number of isomers of MeFapy-dG in this system. The sequential connectivity

of H8/H6 to H1′ NOEs was almost fully intact in this system; connectivity of NOEs between all neighboring nucleotides could be identified, with exception of those between the H1′ proton of MeFapy-dG and its neighboring nucleotides and intranucleotide NOEs between the H1′ proton and other protons of the adduct. Connectivity of the MeFapy-dG NOEs with neighboring nucleotides was reestablished in the H2′/H2′′ region.

At 298 K the unmodified dodecamer contained eight cytosine H5-H6 cross-peaks, corresponding to the eight cytosine nucleotides in the duplex (**Figure 26**). The peaks were well distinguished, with no apparent overlap. The COSY spectrum of the modified dodecamer varied with temperature; the number of cytosine H5-H6 cross-peaks was reduced upon lowering the temperature (**Figure 27**). At 298 K the modified dodecamer contained nine cytosine H5-H6 cross-peaks; the H5-H6 cross-peaks of C^{17}, C^{18}, and C^{20} exhibited high overlap. At 288 K nine cytosine H5-H6 cross-peaks were visible; reduced overlap was observed. At 278 K the number of cytosine H5-H6 cross-peaks was reduced to seven, possibly due to overlap of the C^{17}, C^{20}, and C^{24} peaks.

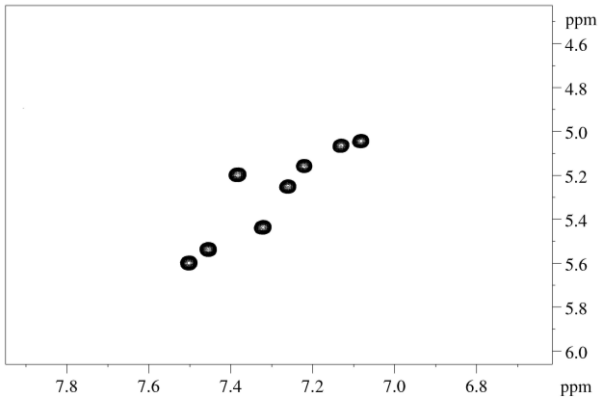


Figure 26. COSY spectrum of unmodified duplex 5′-G^1 C^2 T^3 A^4 G^5 T^6 X^7 G^8 G^9 $T^{10}C^{11}C^{12}$-3′: 3′-$C^{24}G^{23}A^{22}T^{21}C^{20}A^{19}C^{18}C^{17}C^{16}A^{15}G^{14}G^{13}$-5′. Eight cytosine H5-H6 cross-peaks were visible at 298 K.

NOESY spectra were obtained in 99.9996% D_2O and 9:1 H_2O:D_2O; experiments run at 288 K and 278 K were determined to produce the best spectra for analysis. The best NOESY spectrum in 99.9996% D_2O was obtained at 288 K; the best NOESY spectrum in 9:1 H_2O:D_2O was obtained at 278 K. Superior resolution of NOE peaks was observed in experiments run in 99.9996% D_2O. Peak splitting was observed in nucleotides located near the MeFapy-dG lesion and on the same strand as the adduct. Peak doubling was observed for nucleotides G^5, T^6, G^8, and G^9 (**Figure 28**). The sequential connectivity of H8/H6 to H1′ NOEs was disrupted in the modified strand at the H1′ proton of MeFapy-dG; no NOEs were observed between the H1′ proton of MeFapy-dG and G^8, nor between the H1′ proton of MeFapy-dG and any intraresidue protons. A NOE peak was tentatively assigned as being one of the formyl proton signals of MeFapy-dG. The NOE appeared upfield of the predicted range of a formyl proton based on characterizations of other Fapy-dG lesions, although the chemical shift of the formyl proton in this sequence was within the range of formyl protons identified in the 5'- C^1 A^2 T^3 X^4 A^5 T^6 G^7 A^8 C^9 $G^{10}C^{11}T^{12}$-3' (X = ^{13}C-MeFapy-dG) dodecamer (6.8 to 7.9 ppm). Scrutiny of the COSY spectrum indicated the proposed formyl signal was not the H1′ proton of X^7; no COSY cross-

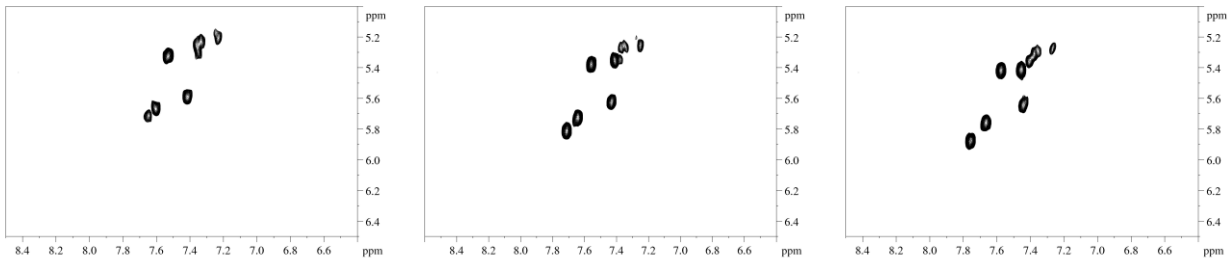


Figure 27. COSY spectra of modified duplex 5′-G^1 C^2 T^3 A^4 G^5 T^6 X^7 G^8 G^9 $T^{10}C^{11}C^{12}$-3′ (X = MeFapy-dG): 3′-$C^{24}G^{23}A^{22}T^{21}C^{20}A^{19}C^{18}C^{17}C^{16}A^{15}G^{14}G^{13}$-5′ at (left) 278 K, (center) 288 K, and (right) 298 K. Nine cytosine H5-H6 cross-peaks were observed at 298 K and 288 K; seven cytosine H5-H6 cross-peaks were observed at 278 K.

peaks were observed between the proposed formyl signal and the H2′/H2′′ protons of MeFapy-dG. No NOEs were observed between the G^5 H8 proton and the proposed formyl proton, indicating that it was unlikely to be a third isomer of T^6; the H6 proton of the two T^6 isomers both exhibited a NOE to the H8 proton of the isomers of G^5. The sequential connectivity of H8/H6 to H1′ NOEs was unbroken in the complementary strand (**Figure 29**). The peak splitting observed in the COSY experiment was not observed in the NOESY experiment; the cytosine

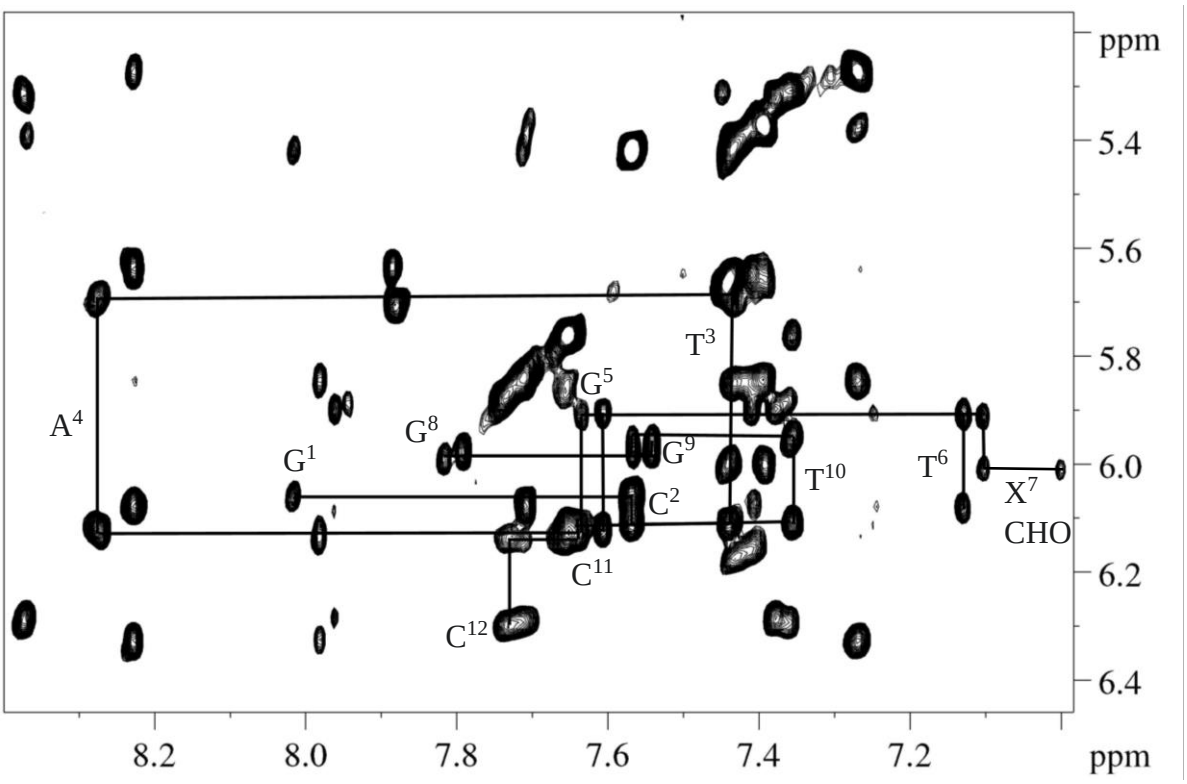


Figure 28. NOESY spectra of modified duplex 5′-G^1 C^2 T^3 A^4 G^5 T^6 X^7 G^8 G^9 $T^{10}C^{11}C^{12}$-3′ (X = MeFapy-dG): 3′-$C^{24}G^{23}A^{22}T^{21}C^{20}A^{19}C^{18}C^{17}C^{16}A^{15}G^{14}G^{13}$-5′ at 288 K in 99.9996% D_2O. Peak doubling was observed for nucleotides G^5, T^6, G^8, and G^9. A NOE peak was tentatively assigned as being between the formyl proton of MeFapy-dG (X^7)and T^6H1′. A break in the sequential connectivity of H8/H6 to H1′ NOEs was observed at the site of the lesion.

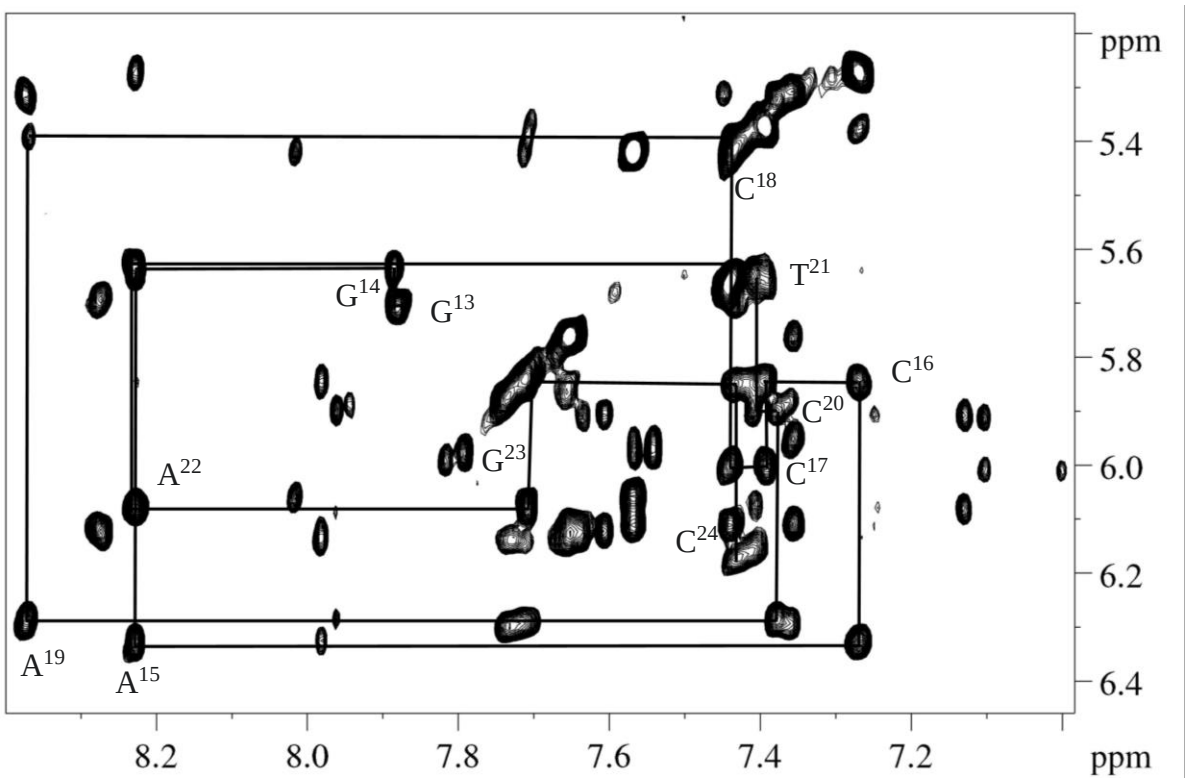


Figure 29. NOESY spectra of modified duplex 5'-G[1] C[2] T[3] A[4] G[5] T[6] X[7] G[8] G[9] T[10]C[11]C[12]-3' (X = MeFapy-dG): 3'-C[24]G[23]A[22]T[21]C[20]A[19]C[18]C[17]C[16]A[15]G[14]G[13]-5' at 288 K in 99.9996% D_2O. The sequential connectivity of H8/H6 to H1' NOEs was unbroken in the complementary strand.

NOEs did not exhibit splitting and appeared to be present as a single species. This may be due to the small chemical shift difference observed between peaks in the COSY experiment and to the small amount of sample reportedly available for these experiments.

Experiments run in 9:1 H_2O:D_2O allowed for further assignments to be made; NOEs between the N^6H proton of X^7 and other protons was achieved. Two isomers of MeFapy-dG were observed, as evidenced by the two N^6H signals observed. In an expanded view of the H1'-H6/H8 region of the NOESY experiment (**Figure 30**) a NOE between the downfield isomer of the T^6 H1' proton and the X^7 N^6H proton is present. The upfield T^6 isomer did not exhibit a NOE

between its H1′ proton and the X^7 N^6H proton. A NOE interaction was observed between the proposed X^7 CHO proton and the X^7 N^6H proton. This was further supported in the H6/H8/N^6H-H2′/H2′′ region (**Figure 31**). NOE cross-peaks are observed between the T^6 H2′/H2′′ protons and the X^7 N^6H proton. NOE correlations are also observed between the proposed formyl group and the X^7 N^6H proton. Different NOEs are observed between the two species of T^6 and the X^7

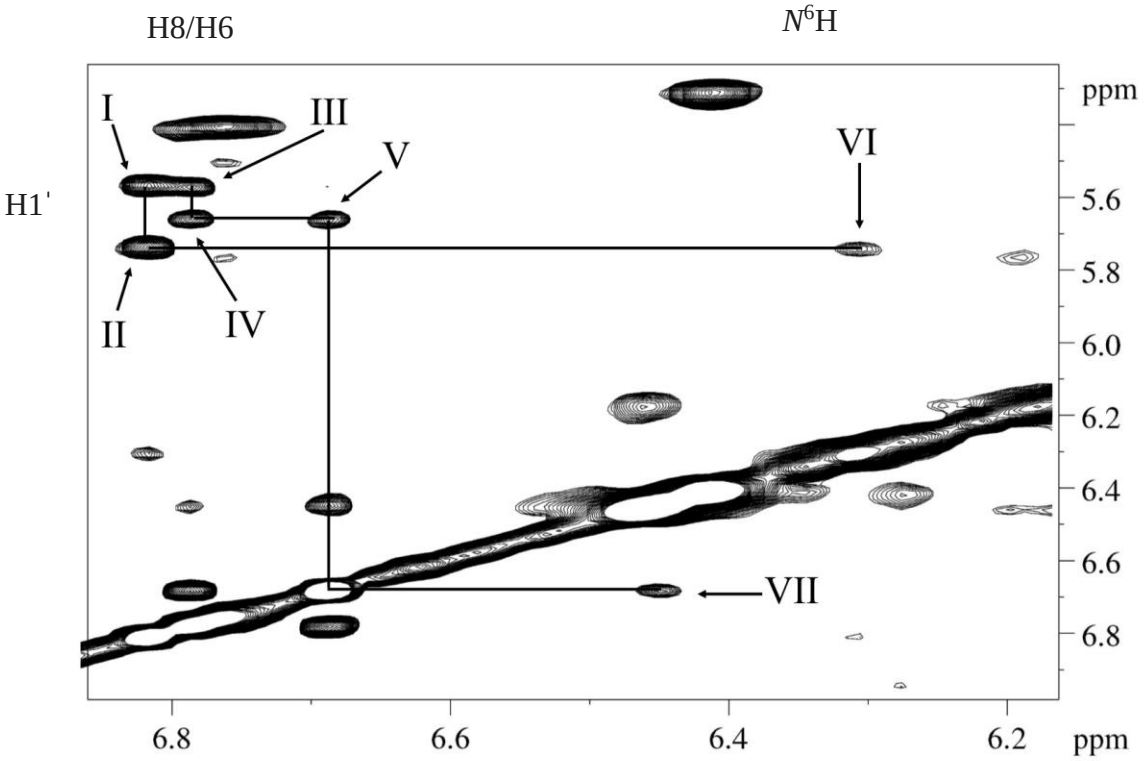


Figure 30. NOESY spectra of modified duplex 5′-G^1 C^2 T^3 A^4 G^5 T^6 X^7 G^8 G^9 $T^{10}C^{11}C^{12}$-3′ (X = MeFapy-dG): 3′-$C^{24}G^{23}A^{22}T^{21}C^{20}A^{19}C^{18}C^{17}C^{16}A^{15}G^{14}G^{13}$-5′ at 278 K in 9:1 H_2O:D_2O. Sequential NOE connectivity was observed between G^5 H1′-T^6 H6a (I), T^6 H1′-T^6 H6a (II), G^5 H1′-T^6 H6ba (III), T^6 H1′-T^6 H6b (IV), T^6 H1′b- X^7 CHO (V), T^6 H1′a- X^7 N^6H (VI), and X^7 CHO- X^7 N^6H (VII).

N^6H proton. The sequential NOE connectivity between neighboring nucleotides is restored in the H2′/H2′′ region (**Figure 32**), indicating a limited distortion of the duplex in the region of the lesion that is nonetheless sufficient to weaken NOEs, particularly between the lesion and its 3′

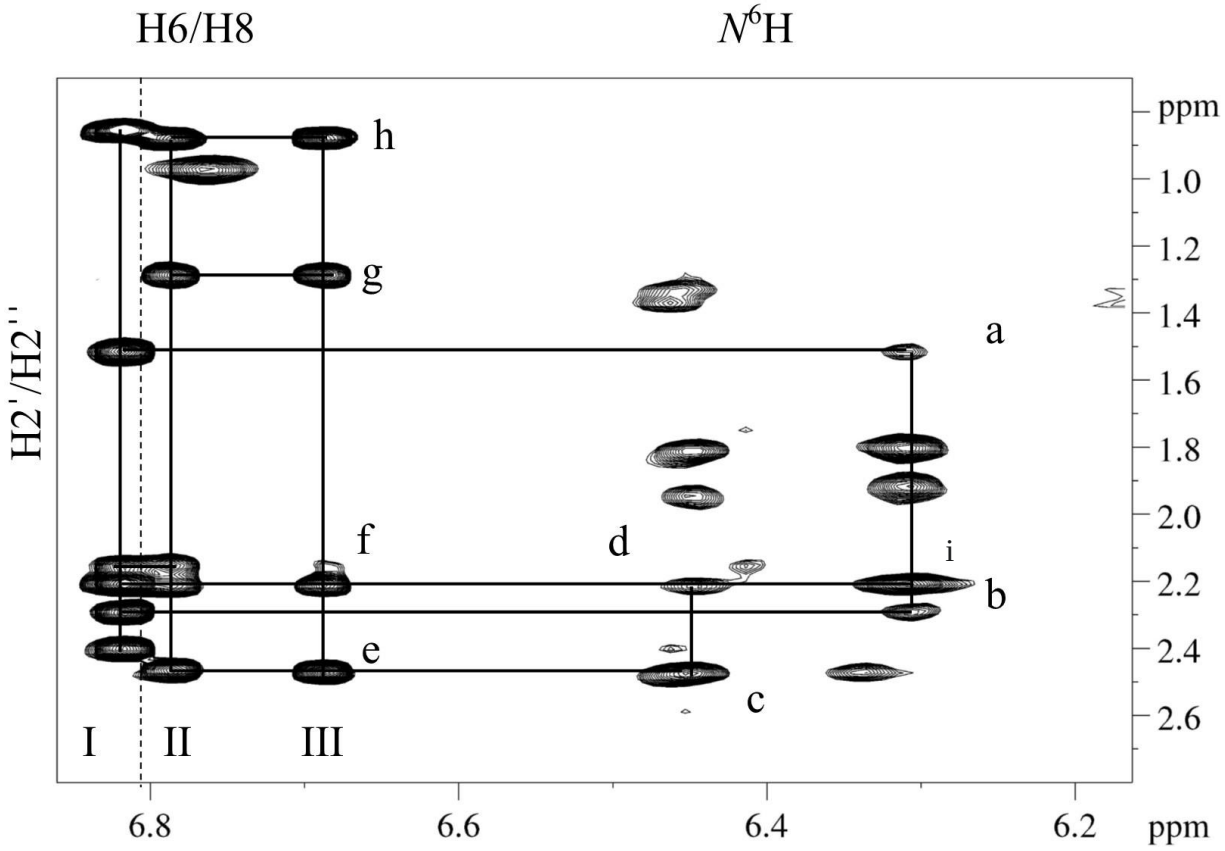


Figure 31. NOESY spectra of modified duplex 5′-G^1 C^2 T^3 A^4 G^5 T^6 X^7 G^8 G^9 $T^{10}C^{11}C^{12}$-3′ (X = MeFapy-dG): 3′-$C^{24}G^{23}A^{22}T^{21}C^{20}A^{19}C^{18}C^{17}C^{16}A^{15}G^{14}G^{13}$-5′ at 278 K in 9:1 H_2O:D_2O. NOE correlations between the downfield species of T^6 (Column I), the upfield species of T^6 (Column II), the proposed X^7 formyl proton (Column III), and the X^7N^6H proton are shown. NOE correlations are observed between I-T^6 H2′-X^7 N^6H (a), I-T^6 H2′′-X^7 N^6H (b), X^7 N^6H-X^7 CH_3 (c), II-T^6 H2′′-X^7 N^6H (d), X^7 CHO- X^7CH_3 (e), II-T^6 H2′′-X^7 CHO (f), II-T^6 H2′-X^7 CHO (g), II-T^6 CH_3- X^7 CHO, and X^7 N^6H-X^7 CH_3 (i).

neighbor, G^8. The apparently limited distortion of the duplex that occurs in this system differs vastly from what was observed in a less G:C rich dodecamer (see previous chapter). A second MeFapy-dG formyl signal was identified at a chemical shift of 7.7 ppm; a NOE was observed between this proton and the CH_3 group of MeFapy-dG. Both species of MeFapy-dG displayed

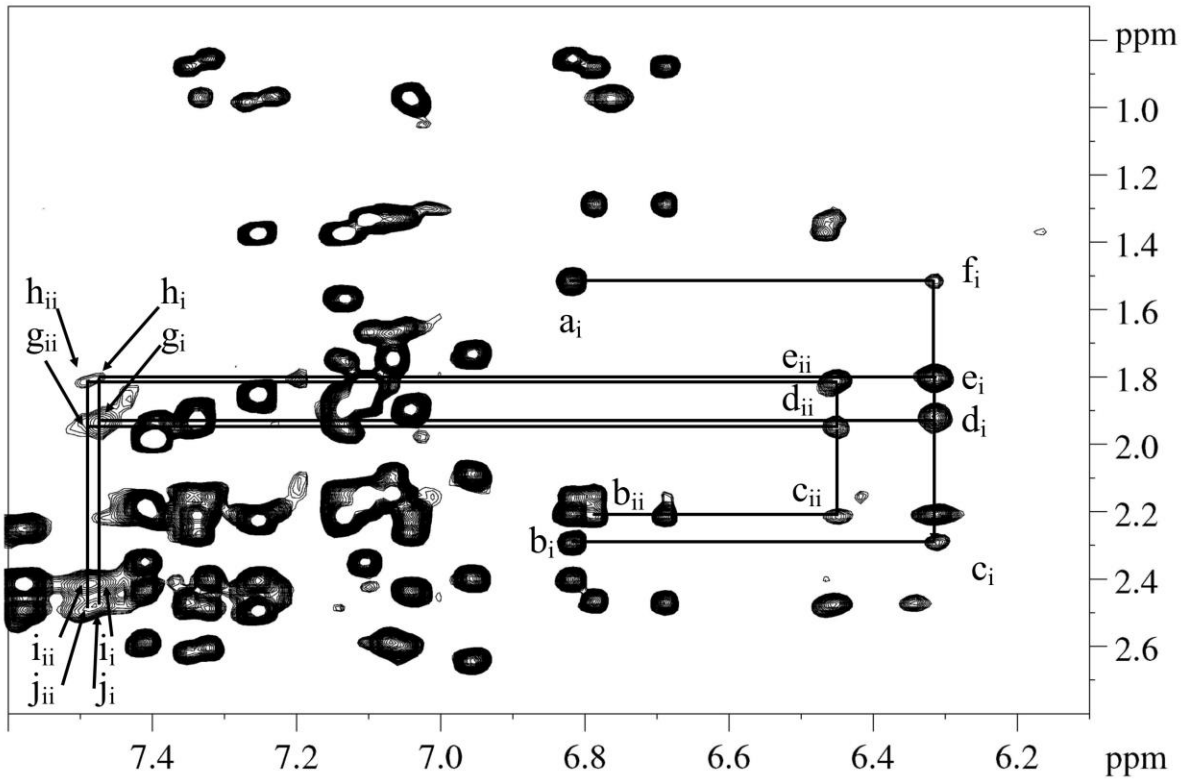


Figure 32. NOESY spectra of modified duplex 5'-G^1 C^2 T^3 A^4 G^5 T^6 X^7 G^8 G^9 $T^{10}C^{11}C^{12}$-3' (X = MeFapy-dG): 3'-$C^{24}G^{23}A^{22}T^{21}C^{20}A^{19}C^{18}C^{17}C^{16}A^{15}G^{14}G^{13}$-5' at 278 K in 9:1 H_2O:D_2O. Sequential NOE correlations are observed from T^6-G^8 : T^6 H2'- T^6 H6 ($a_{ii,}$), T^6 H2''- T^6 H6 ($b_{i,}$ and b_{ii}), T^6 H2''- X^7 N^6H ($c_{i,}$ and c_{ii}), T^6 H2'- X^7 N^6H (f_i), X^7 H2''- X^7 N^6H ($d_{i,}$ and d_{ii}), X^7 H2'- X^7 N^6H ($e_{i,}$ and e_{ii}), X^7 H2'- G^8 H8 ($h_{i,}$ and h_{ii}), X^7 H2''- G^8 H8 ($g_{i,}$ and g_{ii}), G^8 H2'- G^8 H8 ($i_{i,}$ and i_{ii}), and G^8 H2''- G^8 H8 ($j_{i,}$ and j_{ii}).

NOE correlations between the CH_3 group and formyl group, as anticipated. There are some distinctive differences between the two species that can be observed in this region. One isomer of MeFapy-dG (MeFapy-dG_i) exhibits NOE correlations between its N^6H proton and both the H2′ and H2′′ proton of T^6; the other isomer (MeFapy-dG_{ii}) shows a NOE correlation only between its N^6H proton and one of the H2′ protons of T^6, presumably H2′′. Only one isomer, MeFapy-dG_{ii}, exhibited a NOE between the CH_3 group of T^6 and the formyl proton of MeFapy-dG. A X^7 CHO- X^7 N^6H was observed in both isomers, as expected, further supporting the assignments of the formyl peaks. The MeFapy-dG_i isomer displayed a NOE correlation between the neighboring T^6 H1′ and N^6H proton, unlike the second species of MeFapy-dG. Both species of MeFapy-dG displayed a weak X^7 N^6H-G^8 H8 NOE correlation, as was predicted when modeling the lesion in MOE; the distance between the X^7 N^6H and the G^8 H8 proton is predicted to be ~4.5 Å (**Figure 33**).

Based on the NOEs observed between the two species of MeFapy-dG and neighboring nucleotides, the two major species observed in this sequence context are hypothesized to arise from β anomers undergoing rotation around the C^5-N^5 bond, forming R_a and S_a isomers. This prediction is based on the observation that one species of MeFapy-dG has a strong NOE correlation between its formyl proton and the CH_3 group of T^6, but a weak X^7 CH_3-T^6 CH_3 NOE. The second species of MeFapy-dG, in contrast, has a strong X^7 CH_3-T^6 CH_3 NOE correlation and a weak X^7 CHO-T^6 CH_3 NOE.

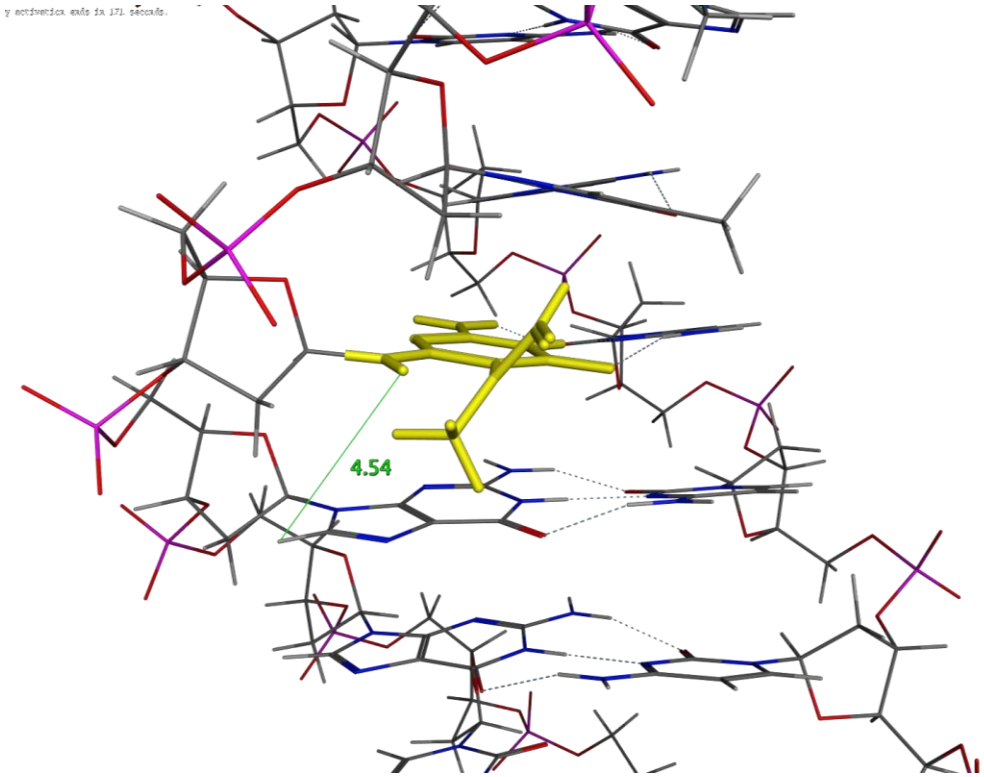


Figure 33. Modeling of the MeFapy-dG containing dodecamer in MOE indicated that in the β anomer the distance between the X^7 N^6H and the G^8 H8 proton is predicted to be ~4.5 Å, within the 5 Å limit of detection of a NOESY experiment.

Discussion

Modified Duplex Stability. The MeFapy-dG lesion, when incorporated into a G:C rich sequence, is less disruptive to duplex structure than it is in more A:T rich sequences (see previous chapter). Sequential NOE connectivity was observed in some regions of NOESY spectra; neighboring nucleotides of the lesion were identifiable in all regions. The presence of NOEs between MeFapy-dG and both neighboring nucleotides, which were missing in the 5'- C^1 A^2 T^3 X^4 A^5 T^6 G^7 A^8 C^9 $G^{10}C^{11}T^{12}$-3' (X = ^{13}C-MeFapy-dG) dodecamer sequence, suggests less disorder in the region of the lesion, as was predicted for a more G:C rich sequence. Further

stabilization may be achieved in a sequence in which both the 3' and 5' neighboring nucleotides are dG or dC; this is of interest due to the continuing inability to calculate a structure of a MeFapy-dG containing duplex utilizing NMR data. A structure of the Fapy-dG lesion in solution, generated by NMR data, could only be achieved using the β-carba-Fapy-dG analogue, which is incapable of undergoing anomerization.[55]

The two species of MeFapy-dG in this system were identified as being present in a ~62:38 ratio, according to integration of the MeFapy-dG formyl proton in a ^{1}H-NMR spectrum; these measurements were confirmed by integration of the isomers of the T^{6} H6 proton and G^{8} H8 proton, which were also split due to the two species of MeFapy-dG. Based on the presence of NOEs between the adduct and its 5' neighboring nucleotide, both observed species were determined to be the β anomer; no NOEs indicating the presence of the α anomer were observed. The major species of MeFapy-dG in this system exhibited a strong X^{7} CH_3-T^{6} CH_3 NOE correlation and a weak X^{7} CHO-T^{6} CH_3 NOE correlation. Based on these NOEs this species was tentatively assigned as the R_a isomer. The minor second species of MeFapy-dG exhibits a strong X^{7} CHO-T^{6} CH_3 NOE correlation and a weak X^{7} CH_3-T^{6} CH_3 NOE. Based on these NOEs this species was tentatively assigned as the S_a isomer (**Figure 34**).

A comparison of the X^{7} CHO- X^{7} CH_3 NOE correlations of the two isomers of MeFapy-dG did not definitively indicate the presence of rotamers of the formyl bond, producing *E* and *Z* isomers. The X^{7} CHO- X^{7} CH_3 NOE of the minor species integrated to ~31% of the total X^{7} CHO- X^{7} CH_3 NOE strength of the two isomers, in reasonable agreement with that predicted from the ~32% population of the minor species; integration of the NOEs would theoretically be identical to the population distribution when the two species do not differ with respect to rotation around the formyl bond. The X^{7} CHO- X^{7} CH_3 NOE correlations of the β *Z*-R_a and β *E*-S_a

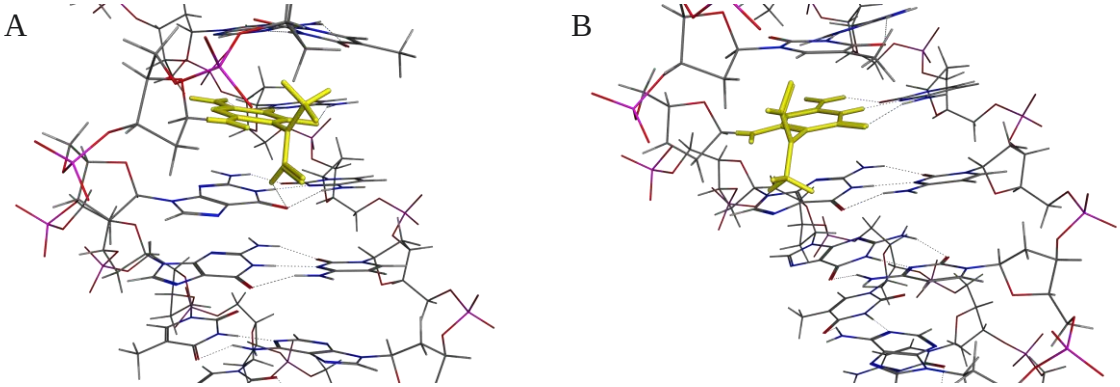


Figure 34. The major species of MeFapy-dG in the 5'-G[1] C[2] T[3] A[4] G[5] T[6] X[7] G[8] G[9] T[10]C[11]C[12]-3' (X = MeFapy-dG): 3'-C[24]G[23]A[22]T[21]C[20]A[19]C[18]C[17]C[16]A[15]G[14]G[13]-5' sequence context is tentatively assigned as the R_a *isomer (A). The minor species is tentatively assigned as the* S_a *isomer (B). Models were constructed in MOE.*

isomers identified in the trimer sequence, 5'-TXT-3' (X = MeFapy-dG), demonstrated a population ratio of 60:40 and a X^7 CHO- X^7 CH_3 NOE strength ratio of 88:12, respectively. The large difference in NOE strength between the β *Z*-R_a and β *E*-S_a isomers of MeFapy-dG in the trimer system, when compared to the population difference, does not support the relatively small difference in NOE strength observed in this dodecamer duplex system as being due to rotation around the formyl bond.

The inability to detect rotation around the formyl bond could arise from two scenarios. The first, and most likely, is that the two species have the same isomerism with respect to the formyl group; they are either both *E* isomers or both *Z* isomers. A difference in NOE strength between the methyl and formyl groups of MeFapy-dG, independent of population, would not be apparent if the isomers are either both *E* isomers or both *Z* isomers. Determination of specific isomeric identity would require a different method, such as employing isotopic labeling. The second alternative, which is deemed less likely due to the double bond character of the N-CHO

bond, is that rapid rotation around the bond is occurring, rendering the isomers unobservable on the NMR time scale.

In the trimer 5'-TXT-3' (X = MeFapy-dG) it was observed that the four *E* isomers of MeFapy-dG exhibited more upfield methyl peaks than their *Z* isomer counterparts, with respect to both the methyl carbon chemical shift and methyl protons chemical shift. Adherence to this trend in the dodecamer sequence would suggest that the major R_a isomer exists as the *E* isomer while the minor S_a isomer also exists as the *E* isomer. The major and minor species of MeFapy-dG are therefore tentatively assigned as the *E*-R_a isomer and *E*-S_a isomer, respectively (**Figure 35**). No methyl peaks were observed in the chemical shift range anticipated for the *Z* isomers, which was observed to be ~2.6-2.8 ppm in the 5'-TXT-3' (X = MeFapy-dG) system.

Inspection of models constructed in MOE revealed a favorable interaction between the MeFapy-dG lesion and its 3' neighboring base that would favor the *E*-R_a isomer over the *Z*-R_a isomer: the formation of a hydrogen bond between the proton of the formyl group and the C-6

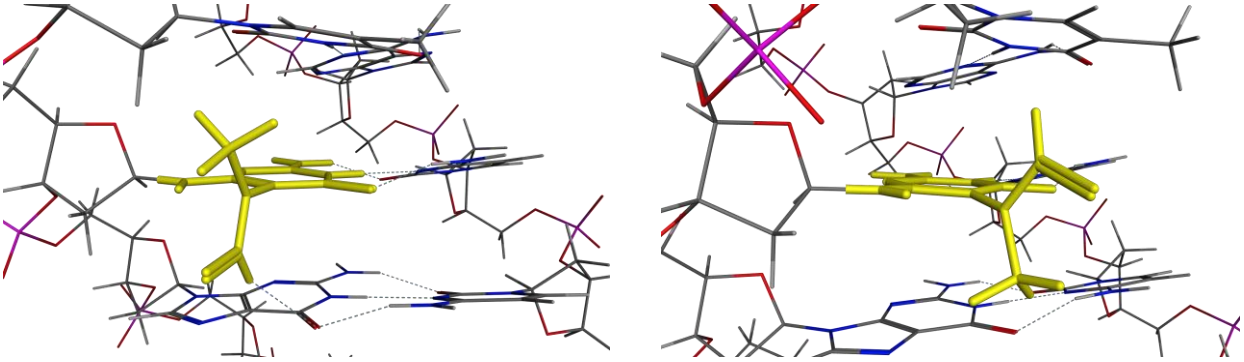


Figure 35. In the sequence context 5'-G^1 C^2 T^3 A^4 G^5 T^6 X^7 G^8 G^9 $T^{10}C^{11}C^{12}$-3' (X = MeFapy-dG): 3'-$C^{24}G^{23}A^{22}T^{21}C^{20}A^{19}C^{18}C^{17}C^{16}A^{15}G^{14}G^{13}$-5' the E-$R_a$ *isomer of MeFapy-dG (left) forms a hydrogen bond with its 3' neighbor, guanine. The* E-S_a *isomer of MeFapy-dG (right) is unable to form hydrogen bonds with either neighboring nucleotide.*

carbonyl group of the 3' neighboring guanine. No hydrogen bonding interactions were observed in the *Z*-R_a isomer. Rather, the *Z*-R_a isomer exhibits a decreased distance between the formyl oxygen of MeFapy-dG and the C-6 carbonyl oxygen of the 3' neighboring guanine, bringing them within 2 Å of each other (**Figure 36**); the proximity of the two oxygen atoms would be disfavored due to the partial negative charges of the atoms, creating repulsive forces.

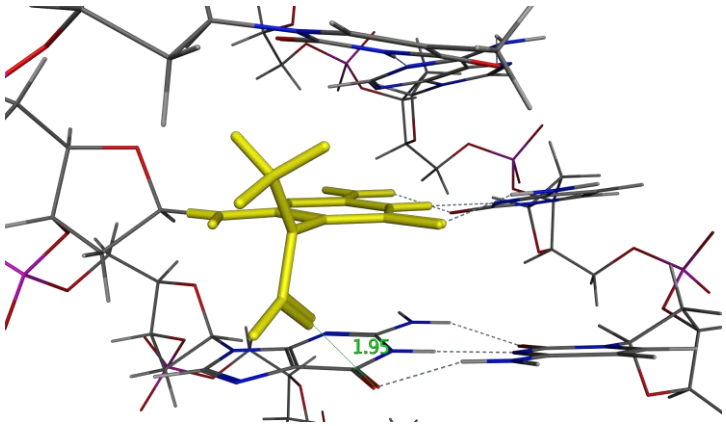


Figure 36. The Z-R_a *isomer of MeFapy-dG is hypothesized to be unfavorable in this sequence due to the close proximity of the formyl oxygen and C-6 carbonyl oxygen of the 3´neighboring guanine.*

The distance between the formyl oxygen and N^6H proton is virtually identical between the *Z*-S_a and *E*-S_a isomers, indicating that intramolecular hydrogen bonding is unlikely to have contributed to the favorability of the isomers (**Figure 37**). No favorable interactions were observed between either the *Z*-S_a or *E*-S_a isomers and their 5' or 3' neighboring nucleotides. Sterics are unlikely to have played a role in the preference for the *E*-S_a isomer or *Z*-S_a isomer; the *Z* and *E* isomers do not appear to induce any additional steric strain on the structure.

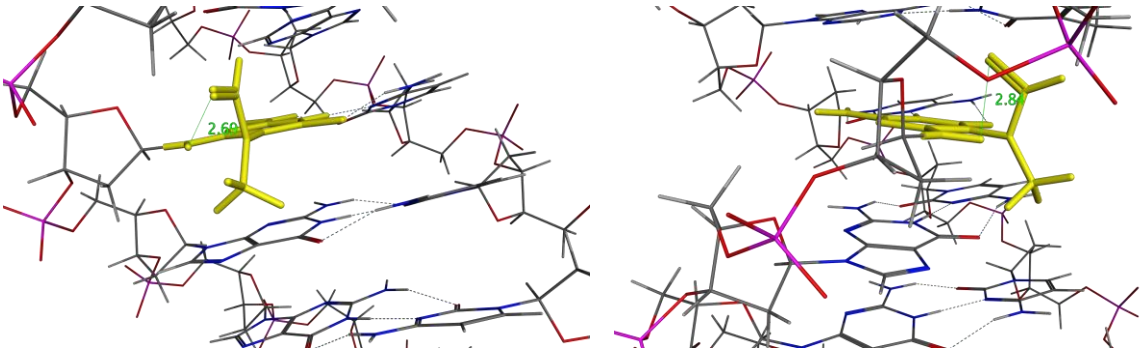


Figure 37. No significant difference in distance between the formyl oxygen and N^6H *proton of the* E-S_a *and* Z-S_a *isomers of MeFapy-dG. Intramolecular hydrogen bonding is therefore unlikely to have contributed to a preference between the* E- S_a *and* Z- S_a *isomers.*

Inspection of a space-filling model of the duplex indicated a potential cause for the preference of the *E-S_a* isomer over the *Z-S_a* isomer. When the MeFapy-dG adduct is present as the *E-S_a* isomer, the formyl oxygen of the adduct is more solvent exposed; the formyl proton is relatively shielded from solvent exposure by comparison (**Figure 38**). Solvent exposure of the formyl oxygen is predicted to be more favorable than solvent exposure of the formyl proton due to the greater partial charge of the oxygen atom, as determined in MOE. The absolute value of the formyl oxygen partial charge is almost tenfold greater than the absolute value of the formyl proton partial charge. Exposure of the oxygen to solvent may result in favorable hydrogen bonding interactions with water; exposure of the formyl proton, which occurs in the *Z-S_a* isomer, would be less likely to produce these favorable interactions.

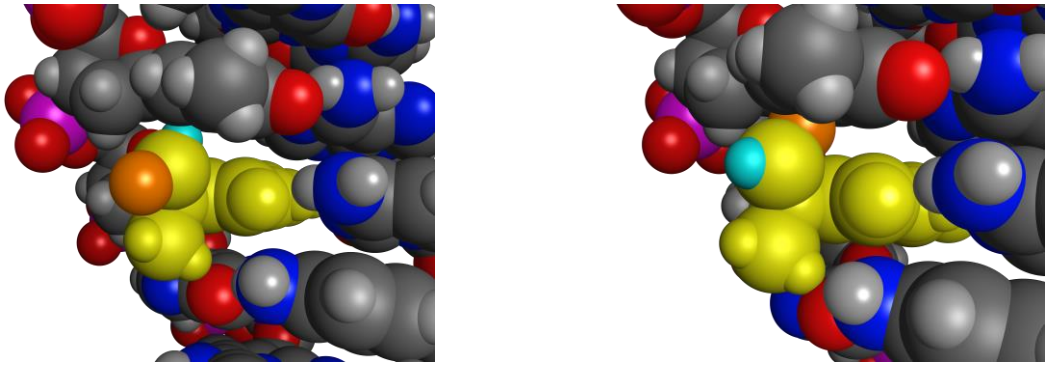


Figure 38. Space-filling models of the MeFapy-dG adduct (yellow) in a dodecamer duplex. The E-S_a *isomer of the adduct (left) projects the formyl oxygen (orange) outside of the DNA duplex structure, potentially allowing for the formation of hydrogen bonds with water molecules of the solvent. The* Z-S_a *isomer of the adduct (right) projects the formyl proton (cyan) outside of the DNA duplex structure. The formation of hydrogen bonds between the formyl proton and water molecules is projected to be less favorable than the formation of hydrogen bonds between the formyl oxygen and water molecules.*

MeFapy-dG Epimerization. Based on previous studies it was hypothesized that the anomeric ratio of MeFapy-dG would be approximately 20:80 α:β.[133] In this sequence context no NOEs indicative of the α anomer was observed. It is hypothesized that the anomeric distribution of the MeFapy-dG adduct is sequence dependent, with important factors including sterics, hydrogen bonding, and base stacking. In this sequence the 3′ neighbor is a purine, a bulkier nitrogenous base when compared to a pyrimidine. The presence of a 3′ neighboring purine is hypothesized to reduce the favorability of the MeFapy-dG adduct existing as the α anomer, due to the increased steric strain that would be induced. Conversion to the α anomer of MeFapy-dG places the MeFapy-dG base in close proximity to the 3′ neighboring base; greater steric strain and disruption of normal duplex structure is anticipated when that neighbor is a larger purine

compared to a smaller pyrimidine. A more predominant population of the α anomer of MeFapy-dG is anticipated to occur in a system in which MeFapy-dG is flanked on the 3′ side by a pyrimidine.

The apparent lack of the α anomer of MeFapy-dG in this system may also be due to the amount of sample available and the number of scans run in the experiments. A small population of the α anomer, coupled with a lack of a large quantity of sample, and potentially compounded by a deficit in the number of scans in the experiments, may have been sufficient to prevent detection of NOEs arising from the α anomer of MeFapy-dG. Synthesis of additional sample and further experiments will be required to verify or disprove this hypothesis.

Rotational Isomers of Fapy-dG. Characterization in duplex DNA of the structurally similar formamidopyrimidine (Fapy-dG) lesion, a lesion formed by oxidation of DNA, has been achieved using a stable analogue, wherein the oxygen of the deoxyribose ring is replaced with a CH_2 group (β-carba-Fapy-dG). This modification prevents anomerization of the lesion, simplifying the system, and stabilizes the lesion, which is prone to depurination.[55] The NMR solution structure of β-carba-Fapy-dG indicated that it exists as three isomers in duplex DNA: the *Z*-S_a isomer and the *E* isomer, theoretically the *E*-R_a isomer based on the majority of the computational models. The *Z*-S_a isomer of β-carba-Fapy-dG was favored over the *E* isomer. Inspection of the models indicated the major species, the *Z*-S_a isomer, was able to form a favorable interaction between the formyl oxygen of β-carba-Fapy-dG and the amino protons of its 3′ neighboring cytosine. The atoms were calculated to be separated by 3.46 Å, sufficient for favorable electrostatic interactions to occur. The minor species, the *E*-R_a isomer, placed the formyl oxygen at a distance of 4.46 Å from the amino protons of its 5′ neighboring cytosine. The electrostatic interaction in the minor species would be ~60% the strength of the electrostatic

interaction in the major species (where electrostatic interaction $\propto 1/r^2$ and the magnitude of the charges of the formyl oxygen and cytosine amino protons are assumed to be constant and the magnitude of charges of the cytosine amino protons identical).

Sequence Dependence. There is extensive precedence for the populations of the isomers of Fapy-dG adducts being dependent on the system in which the adduct is being studied; nucleoside, single strand DNA, and duplex DNA systems have been shown to exhibit strong influences on isomeric populations. Previous investigations into the β-carba-Fapy-dG adduct in duplex DNA confirmed the presence of rotational isomers occurring from rotation around both the formyl bond and the C^5-N^5 bond.[55] Similar to our observations for the MeFapy-dG adduct, the population distribution of the rotational isomers of β-carba-Fapy-dG was found to differ significantly compared to previous studies in which the adduct was incorporated into single strand DNA or studied at the nucleoside level.[55, 56, 130, 137] This precedence for significantly different populations of isomers in duplex vs. single strand has also been observed in oligodeoxynucleotides containing the AFB_1-Fapy-dG adduct; the adduct exists almost exclusively as the α anomer in single strand DNA and almost exclusively as the β anomer in duplex DNA.[60, 61] Further studies demonstrated that the population of rotational isomers of the AFB_1-Fapy-dG adduct is sequence dependent,[59] an observation we hypothesize applies to the MeFapy-dG adduct. No studies into the effect of sequence context on the population of rotational isomers of the Fapy-dG adduct have been published.

The results of this study support the hypothesis that the isomeric population of the MeFapy-dG adduct is sequence dependent. The isomeric population in this system shows evidence of being influenced by both favorable interactions, such as hydrogen bonding, and unfavorable interactions, such as repulsive forces between electronegative atoms. Sterics were

not observed to play a role in the preference for rotational isomers in this lesion, theoretically due to its small size. We anticipate that sterics will play a larger role in the favorability of bulkier alkyl-Fapy-dG adducts, such as those formed upon exposure to nitrogen mustards. In the AFB_1-Fapy-dG adduct, rotation around the C^5-N^5 bond was not observed in duplex DNA due to the higher energy barrier, a feature caused by the bulkiness of the AFB_1 substituent.[59] The loss of the potential hydrogen bond between the N^5H proton, present in Fapy-dG but not MeFapy-dG, and electronegative atoms of neighboring bases, is predicted to limit the impact of neighboring bases on isomeric populations of the MeFapy-dG adduct; as observed in this system, the CH_3 group of MeFapy-dG creates minimal steric strain, but is also unable to form hydrogen bonds with its nearest neighboring base. The full impact that neighboring bases have on the isomeric populations of MeFapy-dG will require further study of the adduct in all possible sequence contexts. Comparison of the isomeric populations with mutational spectra obtained by Earley et al. and Christov et al.[122, 124] is predicted to provide a structural basis for the observed mutational dependence of MeFapy-dG on its sequence context.

Chapter VI.

Summary and Future Directions

Summary

This dissertation has led to the development of a set of methods capable of characterizing individual isomers of the MeFapy-dG adduct in solution. As was hypothesized, the MeFapy-dG lesion exists as a mixture of equilibrating isomers in solution, arising from epimerization, rotation around the C^5-N^5 bond, and rotation around the formyl bond. The number of isomers detected in solution ranged from two to ten. The adduct was characterized in several systems, and the potential sequence dependence of isomer population was investigated in multiple sequence contexts. Further investigations will be required to clarify the full impact of neighboring nucleotide identity on isomeric population.

Due to the complex NMR spectra of DNA containing the MeFapy-dG adduct, initial characterization was achieved utilizing the MeFapy-dG phosphoramidite, a highly simplified system. Owing to the high sensitivity of this molecule to acid, great lengths were taken to ensure the sample was prepared in such a way as to minimize moisture. Protons of the deoxyribose sugar were unambiguously assigned employing COSY and TOCSY experiments. Use of a non-exchangeable solvent, deuterated dichloromethane, extended the spin system of the deoxyribose ring to include the N^6H proton, the identification of which was imperative for later characterizations. Fortuitously, the chemical shift of the N^6H proton fell between the chemical shift ranges normally observed for H8/H6 aromatic protons of the nitrogenous bases and H1′ protons of the deoxyribose sugar. The distinct chemical shift of the N^6H proton, observed in a

region of the spectrum unoccupied by other signals, permitted integration of the eight species from a ^{1}H-NMR spectrum. The population of each isomer was calculated from these values. The N^{6}H proton was observed to be split into two distinct groups; the groups were present in a 40:60 ratio. The groups were assigned as arising from the α and β anomers of MeFapy-dG; identity was determined by a NOESY experiment, in which the H1′-H3′/H4′ correlations were measured. Surprisingly, the α anomer was determined to be the predominant anomer, comprising 60% of the sample. There was no precedence for the preference of the MeFapy-dG phosphoramidite to exist as the α anomer; previous studies, in which MeFapy-dG was incorporated into an oligodeoxynucleotide, found the β anomer to be the predominant species.[133]

The α anomer is hypothesized to be favored over the β anomer in the phosphoramidite due to steric effects. When the MeFapy-dG phosphoramidite exists as the β anomer, the bulky DMT protecting group and MeFapy-dG base are located on the same face of the deoxyribose ring, increasing the steric strain. Conversion of the MeFapy-dG phosphoramidite to the α anomer relieves this strain, placing the DMT group and MeFapy-dG base on opposite faces of the ring. Incorporation of the MeFapy-dG phosphoramidite into an oligodeoxynucleotide is anticipated to disfavor the α anomer, due to the steric clashes that would occur between the MeFapy-dG base and the 3′ neighboring nucleotide. In duplex DNA this would be particularly destabilizing, presumably causing a distortion in the normal duplex structure and resulting in loss/weakening of hydrogen bonds and decreased base stacking.

The predominance of the α anomer in the MeFapy-dG phosphoramidite is hypothesized to contribute to the difficulty encountered when incorporating the MeFapy-dG phosphoramidite into an oligodeoxynucleotide during solid phase synthesis, a commonly occurring issue observed in multiple samples and sequence contexts. The yield of these reactions is low; the predominant

product is a failure sequence resulting from failure to incorporate MeFapy-dG. The MeFapy-dG base can sterically block an incoming nucleophile in the α anomer, preventing nucleophilic attack of its 3′ phosphorous. Unequivocal assignment of the rotational isomers in the phosphoramidite system was not possible due to overlapping NOE peaks; accurate integrations could not be obtained in this system.

In single strand trimer DNA (5′-TXT-3′, X = MeFapy-dG) the adduct exists as eight identifiable isomers originating from rotation around the formyl bond, rotation around the C^5-N^5 bond, and epimerization. Assignments of the deoxyribose protons were made based on COSY and TOCSY experiments. The solvents chosen for the experiments were 99.9996% D_2O and 9:1 H_2O:D_2O. In the 9:1 H_2O:D_2O solvent the N^6H proton signal was observed, permitting integration of NOEs between the formyl, methyl, and N^6H protons. This method was used to assign rotamers arising from rotation around the C^5-N^5 bond. Greater resolution of non-exchangeable protons was observed in experiments run in 99.9996% D_2O, facilitating assignment of partially overlapping peaks. Integrations of NOEs between the H1′ and H3′/H4′ protons were used for assignment of anomeric identity; the α:β ratio was determined to be 35:65. The preference for the MeFapy-dG lesion to exist as the β anomer in the trimer sequence, as opposed to the preference for the α anomer that was observed in the phosphoramidite system, is attributed to the lack of a disproportionately bulky 5′ group, present in the phosphoramidite but not the trimer; in the trimer a thymine nucleotide was present on either side of the MeFapy-dG nucleotide.

Rotation around the formyl bond was detected utilizing isotopic labeling. Incorporation of a ^{13}C-label into the methyl group of MeFapy-dG, followed by a non-decoupled DEPT-135 experiment, yielded $^3J_{CH}$ coupling constants between the methyl carbon and the formyl proton of

MeFapy-dG; the coupling constants, 1.7 Hz and 4.0 Hz, were indicative of rotation around the formyl bond, producing *E* and *Z* geometrical isomers. Similar to what was observed for the N^6H proton, the chemical shifts of the methyl protons were split into two groups; unlike in the case of the N^6H proton, which exhibited splitting due to anomerization, the source of splitting of the methyl signals was rotation around the formyl bond.

The MeFapy-dG lesion exists as ten equilibrating isomers in a single strand dodecamer system (5'-C^1 A^2 T^3 X^4 A^5 T^6 G^7 A^8 C^9 $G^{10}C^{11}T^{12}$-3'; X = ^{13}C-MeFapy-dG), mirroring what was observed in the trimer system. The two groups of methyl peaks, observed in an HSQC experiment, exhibited a different integration ratio compared to the methyl peaks of the trimer system. In the trimer system, in which the 5' and 3' neighbors of MeFapy-dG were thymine nucleotides, the two groups of MeFapy-dG methyl peaks integrated 45:55 upfield: downfield. Analysis of the dodecamer system, in which the 5' and 3' neighbors are thymine and adenine, respectively, established that the two groups of methyl peaks integrated 60:40 upfield:downfield. In the trimer sequence the *Z* isomer corresponded to the downfield group of methyl signals; theoretically the downfield group of methyl signals in the dodecamer system also corresponded to the *Z* isomer. This observation would suggest that replacing the 3' neighbor of MeFapy-dG, substituting adenine in place of thymine, has a substantial influence on the predominance of rotational isomers occurring from rotation around the formyl bond, theoretically favoring the *E* isomer in the dodecamer system. When annealed to its complementary strand the number of methyl peaks was reduced to six strong and two very weak peaks; the ratio of the two groups of methyl peaks was unchanged, indicating that duplex formation likely does not impact the favorability of *Z* and *E* rotational isomers. This would indicate that preference for *Z* or *E* isomers

is influenced predominantly by neighboring bases, presumably due to stabilization by hydrogen bonding, rather than sterics.

Favorability of the *E* geometrical isomer of MeFapy-dG was observed in a second dodecamer sequence, 5'-G^1 C^2 T^3 A^4 G^5 T^6 X^7 G^8 G^9 $T^{10}C^{11}C^{12}$-3' (X = MeFapy-dG). Data for this system, which was obtained by Richard Dempster using a sample originally prepared by Dr. Liang Li, did not include data on the single strand dodecamer. In duplex, the *E* isomer was interpreted to be favored in both the major and minor species; indeed, no data confirming the presence of the *Z* isomer could be identified, although further investigation will be required to unambiguously verify this.

The R_a and S_a isomers of MeFapy-dG were identified by NOE correlations between the X^7 CH_3-T^6 CH_3 protons and the X^7 CHO-T^6 CH_3 protons. An inverse relationship was observed, in which the major species exhibited a strong X^7 CH_3-T^6 CH_3 NOE correlation and a weak X^7 CHO-T^6 CH_3 NOE correlation; the minor species exhibited a weak X^7 CH_3-T^6 CH_3 NOE correlation and a strong X^7 CHO-T^6 CH_3 NOE correlation. The *E* isomer is hypothesized to be favored in both the R_a and S_a isomers of MeFapy-dG; the isomeric identity with respect to the formyl bond was identified based on the chemical shift of the methyl protons. In the trimer sequence the methyl protons of the *Z* and *E* geometrical isomers exhibited a distinct chemical shift difference on both the ^{13}C and ^{1}H scale; the methyl signals of the *E* isomer had a more upfield chemical shift range compared to the methyl signals of the *Z* isomer. The chemical shifts of the MeFapy-dG methyl signals of the dodecamer fell within this more upfield region. No methyl signals were observed in the chemical shift range of the *Z* isomers, which was observed to be ~2.6 to 2.8 ppm in the 5'-TXT-3' (X = MeFapy-dG) system.

Constructions of models of the *E*-R_a and *E*-S_a MeFapy-dG isomers in the dodecamer duplex provided a basis for these observations. The major species, the *E*-R_a isomer, was able to form a favorable hydrogen-bond type interaction between its formyl group proton and the C-6 carbonyl oxygen of the 3′ guanine. The *Z* isomer is hypothesized to be unfavorable due to the close proximity of the formyl oxygen and C-6 carbonyl oxygen of the 3′ guanine. No favorable or unfavorable interactions between the *E*-S_a isomer of MeFapy-dG and its neighboring nucleotides were observed in the models. Construction of space-filling models suggested the formyl oxygen of MeFapy-dG is more solvent exposed in the *E*-S_a isomer, potentially allowing for favorable interactions with the solvent in the form of hydrogen bonds. The *Z*-S_a isomer places the formyl oxygen within the duplex structure, leaving the formyl proton relatively solvent exposed. The absolute value of the partial charge of the formyl proton is approximately one tenth of that of the formyl oxygen; favorable interactions with the solvent are less likely to occur when the formyl proton is exposed.

Future Directions

Future studies would benefit from placement of the MeFapy-dG adduct in the center of the duplex and an increased G:C content of the sequence, particularly flanking the lesion, as was employed in Richard Dempster's experiments. This would stabilize the lesion, reducing the instability of the duplex and potentially reducing the number of isomers of MeFapy-dG. A more stable duplex would increase the temperature range under which the MeFapy-dG lesion could be studied in duplex DNA; the less G:C rich dodecamer duplex, in which MeFapy-dG was located near the terminal portion of the duplex, was prone to denaturing above 288 K. This might sharpen the NOE peaks; peak broadening of the MeFapy-dG lesion was observed in the single

strand, which may have also played a role in the lack of NOEs between MeFapy-dG and neighboring nucleotides when annealed to its complementary strand. Larger amounts of material will also be crucial in future experiments, in which difficulty may occur in detecting NOE peaks of the lesion, particularly when both neighboring base pairs are A:T. Due to the low incorporation rate of the MeFapy-dG phosphoramidite into an oligodeoxynucleotide, obtaining large amounts of the modified oligodeoxynucleotide has been a major difficulty. Alternative methods of obtaining the modified oligodeoxynucleotides should be considered in future studies.

Populations of the isomers of MeFapy-dG have been determined in the 5'-TXT-3' sequence context. To investigate the dependence of isomeric population of MeFapy-dG on neighboring nucleotides, incorporation of the MeFapy-dG lesion into several sequence contexts will be required; trimer sequences will be designed to model the effect of the four naturally occurring nucleotides when they are the 5' and 3' neighbors of MeFapy-dG. Characterization of the isomeric dependence of MeFapy-dG in single strand systems is anticipated to show distinct differences compared to what will be observed when the adduct is incorporated into duplex DNA. Future assignments of the isomers of MeFapy-dG can be conducted utilizing the methods contained herein.

The dependence of the isomeric populations of MeFapy-dG on neighboring bases will also be investigated in duplex DNA, in which it is projected to exist as a reduced number of isomers due to steric constraints and increased hydrogen bonding. This was observed in both dodecamer duplex systems, where annealing to its complementary strand resulted in a decrease in the number of MeFapy-dG isomers. The increased steric constraints in duplex DNA are predicted to favor the formation of hydrogen bonds between the adduct and neighboring bases, based on the decreased disorder of the biomacromolecule when annealed to its complementary

strand. The impact of neighboring bases on the isomeric populations of MeFapy-dG is therefore projected to be increased in duplex DNA compared to single strand DNA.

The methods used to determine the isomeric composition of MeFapy-dG can be extended to the study of other alkyl-Fapy lesions, such as those occurring from exposure to nitrogen mustards. Characterization of lesions arising from treatment with alkylating agents is essential to predicting how a patient's cancer will tolerate treatment, leading to more personalized medicine and improvements in the efficacy of treatment. Accurate prediction of the isomeric populations of alkyl-Fapy-dG lesions is essential for predicting interactions between various enzymes and the adducts, and may contribute to the development of personalized medical treatments.

References

(1) Franklin, R. E.; Gosling, R. G. , **(1953),** Molecular configuration in sodium thymonucleate, *Nature, 171*, 740-741.

(2) Cobb, M. In **(2015);** Basic Books: Life's Greatest Secret: The Race to Crack the Genetic Code.

(3) Watson, J. D. In **(2011);** Scribner: The Double Helix: A Personal Account of the Discovery of the Structure of DNA.

(4) Astbury, W. T. , **(1947),** X-ray studies of nucleic acids, *Symp. Soc. Exp. Biol.*, 66-76.

(5) Judson, H. F. In **(1996);** CSHL Press: The Eighth Day of Creation: Makers of the Revolution in Biology.

(6) Fry, M. In **(2016);** Academic Press: Landmark Experiments in Molecular Biology.

(7) Chargaff, E.; Lipshitz, R.; Green, C. , **(1952),** Composition of the desoxypentose nucleic acids of four genera of sea-urchin, *J. Biol. Chem.*, *195*, 155-160.

(8) Meselson, M.; Stahl, F. W. , **(1958),** The replication of DNA in *escherichia coli*, *Proc. Natl. Acad. Sci. U. S. A.*, *44*, 671-682.

(9) Pray, L. , **(2008),** Semi-conservative DNA replication: Meselson and stahl, *Nat. Educ.*, *1*, 98.

(10) Kool, E. T. , **(2001),** Hydrogen bonding, base stacking, and steric effects in DNA replication, *Annu. Rev. Biophys. Biomol. Struct.*, *30*, 1-22.

(11) Sharma, M.; Predeus, A. V.; Mukherjee, S.; Feig, M. , **(2013),** DNA bending propensity in the presence of base mismatches: Implications for DNA repair, *J. Phys. Chem. B, 117*, 6194-6205.

(12) Hsieh, P.; Yamane, K. , **(2008),** DNA mismatch repair: Molecular mechanism, cancer, and ageing, *Mech. Ageing Dev., 129*, 391-407.

(13) Clancy, S. , **(2008),** DNA damage & repair: Mechanisms for maintaining DNA integrity, *Nat. Educ., 1*, 103.

(14) Beerenwinkel, N.; Antal, T.; Dingli, D., et al , **(2007),** Genetic progression and the waiting time to cancer, *PLoS Comput. Biol., 3*, e225.

(15) Loeb, K. R.; Loeb, L. A. , **(2000),** Significance of multiple mutations in cancer, *Carcinogenesis, 21*, 379-385.

(16) Ravanat, J.; Di Mascio, P.; Martinez, G. R.; Medeiros, M. H. G.; Cadet, J. , **(2000),** Singlet oxygen induces oxidation of cellular DNA, *J. Biol. Chem., 275*, 40601-40604.

(17) Perez-Prior, M. T.; Manso, J. A.; Garcia-Santos Mdel, P.; Calle, E.; Casado, J. , **(2005),** Alkylating potential of potassium sorbate, *J. Agric. Food Chem., 53*, 10244-10247.

(18) Mamur, S.; Yüzbaşıoğlu, D.; Ünal, F.; Aksoy, H. , **(2012),** Genotoxicity of food preservative sodium sorbate in human lymphocytes in vitro, *Cytotechnology, 64*, 553-562.

(19) Ghosh, M.; Godderis, L. , **(2016),** Genotoxicity of ethylene oxide: A review of micronucleus assay results in human population, *Mutat. Res., 770*, 84-91.

(20) Li, F.; Segal, A.; Solomon, J. J. , **(1992),** In vitro reaction of ethylene oxide with DNA and characterization of DNA adducts, *Chem. Biol. Interact.*, *83*, 35-54.

(21) Solomon, J. J.; Mukai, F.; Fedyk, J.; Segal, A. , **(1988),** Reactions of propylene oxide with 2′-deoxynucleosides and in vitro with calf thymus DNA, *Chem. Biol. Interact.*, *67*, 275-294.

(22) Goodman, L.; Wintrobe, M.; Dameshek, W.; Goodman, M.; Gilman, A.; McLennan, M. , **(1946),** Nitrogen mustard therapy: Use of methyl-bis(beta-chloroethyl)amine hydrochloride and tris(beta-chloroethyl)amine hydrochloride for hodgkin's disease, lymphosarcoma, leukemia and certain allied and miscellaneous disorders, *J. Am. Med. Assoc.*, *132*, 126-132.

(23) Scott, R. B. , **(1970),** Cancer chemotherapy-the first twenty-five years, *Br. Med. J.*, *4*, 259-265.

(24) Gilman, A. , **(1963),** The initial clinical trial of nitrogen mustard, *Am. J. Surg.*, *105*, 574-578.

(25) Christakis, P. , **(2011),** The birth of chemotherapy at yale: Bicentennial lecture series: Surgery grand round, *Yale J. Biol. Med.*, *84*, 169-172.

(26) Joensuu, H. , **(2008),** Systemic chemotherapy for cancer: From weapon to treatment, *Lancet Oncol.*, *9*, 304.

(27) Corrie, P. G. , **(2008),** Cytotoxic chemotherapy: Clinical aspects, *Medicine*, *36*, 24-28.

(28) Golomb, F. M. , **(1963),** Agents used in cancer chemotherapy, *Am. J. Surg.*, *105*, 579-590.

(29) Ojwang, J. O.; Grueneberg, D. A.; Loechler, E. L. , **(1989),** Synthesis of a duplex oligonucleotide containing a nitrogen mustard interstrand DNA-DNA cross-link, *Cancer Res., 49*, 6529-6537.

(30) Reiner, B.; Zamenhof, S. , **(1957),** Studies on the chemically reactive groups of deoxyribonucleic acids. *J. Biol. Chem., 228*, 475-486.

(31) Lavery, R.; Pullman, B. , **(1981),** Molecular electrostatic potential on the surface envelopes of macromolecules: B-DNA, *Int. J. Quantum Chem., 20*, 259-272.

(32) Drabløs, F.; Feyzi, E.; Aas, P. A., et al , **(2004),** Alkylation damage in DNA and RNA—repair mechanisms and medical significance, *DNA Repair, 3*, 1389-1407.

(33) Ludlum, D. B. , **(1990),** DNA alkylation by the haloethylnitrosoureas: Nature of modifications produced and their enzymatic repair or removal, *Mutat. Res., 233*, 117-126.

(34) Mattes, W. B.; Hartley, J. A.; Kohn, K. W. , **(1986),** DNA sequence selectivity of guanine-N7 alkylation by nitrogen mustards, *Nucleic Acids Res., 14*, 2971-2987.

(35) Smith, S. A.; Engelward, B. P. , **(2000),** In vivo repair of methylation damage in aag 3-methyladenine DNA glycosylase null mouse cells, *Nucleic Acids Res., 28*, 3294-3300.

(36) Verly, W. G. , **(1974),** Monofunctional alkylating agents and apurinic sites in DNA, *Biochem. Pharmacol., 23*, 3-8.

(37) Loeb, L. A.; Preston, B. D. , **(1986),** Mutagenesis by apurinic/apyrimidinic sites, *Annu. Rev. Genet., 20*, 201-230.

(38) Ezaz-Nikpay, K.; Verdine, G. L. , **(1994),** The effects of N7-methylguanine on duplex DNA structure, *Chem. Biol., 1*, 235-240.

(39) Lee, S.; Bowman, B. R.; Ueno, Y.; Wang, S.; Verdine, G. L. , **(2008),** Synthesis and structure of duplex DNA containing the genotoxic nucleobase lesion N7-methylguanine, *J. Am. Chem. Soc., 130*, 11570-11571.

(40) Lin, Y.; Li, L.; Makarova, A. V.; Burgers, P. M.; Stone, M. P.; Lloyd, R. S. , **(2014),** Error-prone replication bypass of the primary aflatoxin B_1 DNA adduct, AFB_1-N7-gua, *J. Biol. Chem., 289*, 18497-18506.

(41) Waters, L. S.; Minesinger, B. K.; Wiltrout, M. E.; D'Souza, S.; Woodruff, R. V.; Walker, G. C. , **(2009),** Eukaryotic translesion polymerases and their roles and regulation in DNA damage tolerance, *Microbiol. Mol. Biol. Rev., 73*, 134-154.

(42) Prakash, S.; Johnson, R. E.; Prakash, L. , **(2005),** Eukaryotic translesion synthesis DNA polymerases: Specificity of structure and function, *Annu. Rev. Biochem., 74*, 317-353.

(43) Yang, W. , **(2005),** Portraits of a Y-family DNA polymerase, *FEBS Lett., 579*, 868-872.

(44) Chetsanga, C. J.; Bearie, B.; Makaroff, C. , **(1982),** Alkaline opening of imidazole ring of 7-methylguanosine. 1. analysis of the resulting pyrimidine derivatives, *Chem. Biol. Interact., 41*, 217-233.

(45) Chetsanga, C. J.; Makaroff, C. , **(1982),** Alkaline opening of imidazole ring of 7-methylguanosine. 2. further studies on reaction mechanisms and products, *Chem. Biol. Interact., 41*, 235-249.

(46) Smela, M. E.; Hamm, M. L.; Henderson, P. T.; Harris, C. M.; Harris, T. M.; Essigmann, J. M. , **(2002),** The aflatoxin B_1 formamidopyrimidine adduct plays a major role in causing the types of mutations observed in human hepatocellular carcinoma, *Proc. Natl. Acad. Sci. U. S. A.*, *99*, 6655-6660.

(47) Sha, Y.; Minko, I. G.; Malik, C. K.; Rizzo, C. J.; Lloyd, R. S. , **(2017),** Error-prone replication bypass of the imidazole ring-opened formamidopyrimidine deoxyguanosine adduct, *Environ. Mol. Mutagen.*, *58*, 182-189.

(48) Minko, I. G.; Rizzo, C. J.; Lloyd, R. S. , **(2017),** Mutagenic potential of nitrogen mustard-induced formamidopyrimidine DNA adduct: Contribution of the non-canonical α-anomer, *J. Biol. Chem.*, *292*, 18790-18799.

(49) Bauer, G. B.; Povirk, L. F. , **(1997),** Specificity and kinetics of interstrand and intrastrand bifunctional alkylation by nitrogen mustards at a G-G-C sequence, *Nucleic Acids Res.*, *25*, 1211-1218.

(50) Kadlubar, F. F.; Beranek, D. T.; Weis, C. C.; Evans, F. E.; Cox, R.; Irving, C. C. , **(1984),** Characterization of the purine ring-opened 7-methylguanine and its persistence in rat bladder epithelial DNA after treatment with the carcinogen N-methylnitrosourea, *Carcinogenesis*, *5*, 587-592.

(51) Malins, D. C.; Haimanot, R. , **(1990),** 4,6-diamino-5-formamidopyrimidine, 8-hydroxyguanine and 8-hydroxyadenine in DNA from neoplastic liver of english sole exposed to carcinogens, *Biochem. Biophys. Res. Commun.*, *173*, 614-619.

(52) Dizdaroglu, M.; Jaruga, P. , **(2012),** Mechanisms of free radical-induced damage to DNA, *Free Radic. Res., 46*, 382-419.

(53) Coste, F.; Ober, M.; Le Bihan, Y., et al , **(2008),** Bacterial base excision repair enzyme fpg recognizes bulky N7-substituted-FapydG lesion via unproductive binding mode, *Send to Chem Biol., 15*, 706-717.

(54) Piotr, C. , **(2014),** Theoretical studies on the tautomeric properties of diamino-5-formamidopyrimidines, *Z. Naturforsch., 53*, 1027-1036.

(55) Lukin, M.; Zaliznyak, T.; Attaluri, S.; Johnson, F.; de los Santos, C. , **(2012),** Solution structure of duplex DNA containing a β-carba-fapy-dG lesion, *Chem. Res. Toxicol., 25*, 2423-2431.

(56) Lukin, M.; Minetti, C. A. S. A.; Remeta, D. P., et al , **(2011),** Novel post-synthetic generation, isomeric resolution, and characterization of fapy-dG within oligodeoxynucleotides: Differential anomeric impacts on DNA duplex properties, *Nucleic Acids Res., 39*, 5776-5789.

(57) Gehrke, T. H.; Lischke, U.; Gasteiger, K. L., et al , **(2013),** Unexpected non-hoogsteen-based mutagenicity mechanism of FaPy-DNA lesions, *Nat. Chem. Biol., 9*, 455-461.

(58) Amato, N. J.; Zhai, Q.; Navarro, D. C.; Niedernhofer, L. J.; Wang, Y. , **(2015),** In vivo detection and replication studies of α-anomeric lesions of 2'-deoxyribonucleosides, *Nucleic Acids Res., 43*, 8314-8324.

(59) Li, L.; Brown, K. L.; Ma, R.; Stone, M. P. , **(2015),** DNA sequence modulates geometrical isomerism of the trans-8,9-dihydro-8-(2,6-diamino-4-oxo-3,4-dihydropyrimid-5-yl-formamido)-9-hydroxy aflatoxin B_1 adduct, *Chem. Res. Toxicol., 28,* 225-237.

(60) Mao, H.; Deng, Z.; Wang, F.; Harris, T. M.; Stone, M. P. , **(1998),** An intercalated and thermally stable FAPY adduct of aflatoxin B_1 in a DNA duplex: Structural refinement from ^{1}H NMR, *Biochemistry, 37,* 4374-4387.

(61) Brown, K. L.; Voehler, M. W.; Magee, S. M.; Harris, C. M.; Harris, T. M.; Stone, M. P. , **(2009),** Structural perturbations induced by the a-anomer of the aflatoxin B_1 formamidopyrimidine adduct in duplex and single-strand DNA, *J. Am. Chem. Soc., 131*, 16096-16107.

(62) Pande, P.; Haraguchi, K.; Jiang, Y.; Greenberg, M. M.; Basu, A. K. , **(2015),** Unlike catalyzing error-free bypass of 8-OxodGuo, DNA polymerase λ is responsible for a significant part of Fapy·dG-induced G → T mutations in human cells, *Biochemistry, 54,* 1859-1862.

(63) Patro, J. N.; Wiederholt, C. J.; Jiang, Y. L.; Delaney, J. C.; Essigmann, J. M.; Greenberg, M. M. , **(2007),** Studies on the replication of the ring opened formamidopyrimidine, Fapy·dG in *escherichia coli*, *Biochemistry, 46,* 10202-10212.

(64) Smela, M. E.; Currier, S. S.; Bailey, E. A.; Essigmann, J. M. , **(2001),** The chemistry and biology of aflatoxin B_1: From mutational spectrometry to carcinogenesis, *Carcinogenesis, 22,* 535-545.

(65) Jin, B.; Li, Y.; Robertson, K. D. , **(2011),** DNA methylation: Superior or subordinate in the epigenetic hierarchy? *Genes Cancer., 2*, 607-617.

(66) Chen, T. , **(2011),** Chapter 10 - mechanistic and functional links between histone methylation and DNA methylation, *Prog. Mol. Biol. Transl. Sci., 101*, 335-348.

(67) Liloglou, T.; Field, J. K. , **(2010),** Detection of DNA methylation changes in body fluids, *Adv. Genet., 71*, 177-207.

(68) Robertson, K. D.; A.Jones, P. , **(2000),** DNA methylation: Past, present and future directions, *Carcinogenesis, 21*, 461-467.

(69) Shuker, D. E.; Prevost, V.; Friesen, M. D.; Lin, D.; Ohshima, H.; Bartsch, H. , **(1993),** Urinary markers for measuring exposure to endogenous and exogenous alkylating agents and precursors, *Environ. Health Perspect., 99*, 33-37.

(70) De Bont, R.; van Larebeke, N. , **(2004),** Endogenous DNA damage in humans: A review of quantitative data, *Mutagenesis, 19*, 169-185.

(71) Beranek, D. T. , **(1990),** Distribution of methyl and ethyl adducts following alkylation with monofunctional alkylating agents, *Mutat. Res., 231*, 11-30.

(72) Jin, B.; Robertson, K. D. , **(2013),** DNA methyltransferases (DNMTs), DNA damage repair, and cancer, *Adv. Exp. Med. Biol., 754*, 3-29.

(73) Modrich, P. , **(1991),** Mechanisms and biological effects of mismatch repair, *Annu. Rev. Genet., 25*, 229-253.

(74) Modrich, P.; Lahue, R. , **(1996),** Mismatch repair in replication fidelity, genetic recombination, and cancer biology, *Annu. Rev. Biochem., 65*, 101-133.

(75) Modrich, P. , **(1997),** Strand-specific mismatch repair in mammalian cells, *J. Biol. Chem., 272*, 24727-24730.

(76) Modrich, P. , **(2006),** Mechanisms in eukaryotic mismatch repair, *J. Biol. Chem., 281*, 30305-30309.

(77) Boyes, J.; Bird, A. , **(1991),** DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein, *Cell., 64*, 1123-1134.

(78) Watt, F.; Molloy, P. L. , **(1988),** Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev., 2*, 1136-1143.

(79) Dantas Machado, A. C.; Zhou, T.; Rao, S., et al , **(2015),** Evolving insights on how cytosine methylation affects protein–DNA binding, *Brief Funct. Genomics, 14*, 61-73.

(80) Yin, Y.; Morgunova, E.; Jolma, A., et al , **(2017),** Impact of cytosine methylation on DNA binding specificities of human transcription factors, *Science, 356*.

(81) Sheaffer, K. L.; Kim, R.; Aoki, R., et al , **(2014),** DNA methylation is required for the control of stem cell differentiation in the small intestine, *Genes Dev., 28*, 652-664.

(82) Newlands, E. S.; Stevens, M. F.; Wedge, S. R.; Wheelhouse, R. T.; Brock, C. , **(1997),** Temozolomide: A review of its discovery, chemical properties, pre-clinical development and clinical trials, *Cancer Treat. Rev., 23*, 35-61.

(83) Fu, D.; Calvo, J. A.; Samson, L. D. , **(2012),** Balancing repair and tolerance of DNA damage caused by alkylating agents, *Nat. Rev. Cancer, 12*, 104-120.

(84) Sedgwick, B.; Bates, P. A.; Paik, J.; Jacobs, S. C.; Lindahl, T. , **(2007),** Repair of alkylated DNA: Recent advances, *DNA Repair, 6*, 429-442.

(85) National Cancer Institute Trials Produce Practice-Changing Results for Brain Cancer. *https://www.cancer.gov/news-events/cancer-currents-blog/2016/asco-temozolomide-brain*, **2018)**.

(86) Krex, D.; Klink, B.; Hartmann, C., et al , **(2007),** Long-term survival with glioblastoma multiforme, *Brain, 130*, 2596-2606.

(87) Hegi, M. E.; Diserens, A.; Gorlia, T., et al , **(2005),** MGMT gene silencing and benefit from temozolomide in glioblastoma, *N. Engl. J. Med., 352*, 997-1003.

(88) Lu, S.; Beijing Cancer Hospital HDI Versus Chemotherapy as Systemic Adjuvant Therapy for Resected Mucosal Melanoma. *https://www.clinicaltrials.gov/ct2/show/NCT03435302?term=Temozolomide&rank=7*, **2018)**.

(89) Jun, G.; Beijing Cancer Hospital A Study of Apatinib Combined With Temozolomide in Patients With Advanced Melanoma.

https://www.clinicaltrials.gov/ct2/show/NCT03422445?term=Temozolomide&rank=10, **2018)**.

(90) Jeong, J. H.; Hong, Y. S.; Park, Y., et al , **(2016),** Phase 1 study of preoperative chemoradiation therapy with temozolomide and capecitabine in patients with locally advanced rectal cancer, *Int. J. Radiat. Oncol. Biol. Phys.*, *96*, 289-295.

(91) Stupp, R.; Hegi, M. E.; Mason, W. P., et al , **(2009),** Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial, *Lancet Oncol.*, *10*, 459-466.

(92) Stupp, R.; Mayer, M.; Kann, R., et al , **(2009),** Neoadjuvant chemotherapy and radiotherapy followed by surgery in selected patients with stage IIIB non-small-cell lung cancer: A multicentre phase II trial, *Lancet Oncol.*, *10*, 785-793.

(93) Chamberlain, M. , **(2011),** Evolving strategies: Future treatment of glioblastoma, *Expert Rev. Neurother.*, *11*, 519-532.

(94) Bobola, M. S.; Kolstoe, D. D.; Blank, A.; Chamberlain, M. C.; Silber, J. R. , **(2012),** Repair of 3-methyladenine and abasic sites by base excision repair mediates glioblastoma resistance to temozolomide, *Front. Oncol.*, *2*, 176.

(95) Stupp, R.; Taillibert, S.; Kanner, A. A., et al , **(2015),** Maintenance therapy with tumor-treating fields plus temozolomide vs temozolomide alone for glioblastoma: A randomized clinical trial, *JAMA*, *314*, 2535-2543.

(96) Seystahl, K.; Wick, W.; Weller, M. , **(2016),** Therapeutic options in recurrent glioblastoma--an update, *Crit. Rev. Oncol. Hematol., 99*, 389-408.

(97) Lieberman, F. , **(2017),** Glioblastoma update: Molecular biology, diagnosis, treatment, response assessment, and translational clinical trials, *F1000Res., 6*, 1892.

(98) Stupp, R.; Mason, W. P.; van, d. B., et al , **(2005),** Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma, *N. Engl. J. Med., 352*, 987-996.

(99) De Vita, S.; De Matteis, S.; Laurenti, L., et al , **(2005),** Secondary ph+ acute lymphoblastic leukemia after temozolomide, *Ann. Hematol., 84*, 760-762.

(100) Chamberlain, M. C.; Raizer, J. , **(2009),** Extended exposure to alkylator chemotherapy: Delayed appearance of myelodysplasia, *J. Neurooncol., 93*, 229-232.

(101) Momota, H.; Narita, Y.; Miyakita, Y.; Hosono, A.; Makimoto, A.; Shibui, S. , **(2010),** Acute lymphoblastic leukemia after temozolomide treatment for anaplastic astrocytoma in a child with a germline TP53 mutation, *Pediatr. Blood Cancer, 55*, 577-579.

(102) Kyrtopoulos, S. A. , **(1998),** DNA adducts in humans after exposure to methylating agents, *Mutat. Res., 405*, 135-143.

(103) Dahlmann, H. A.; Vaidyanathan, V. G.; Sturla, S. J. , **(2009),** Investigating the biochemical impact of DNA damage with structure-based probes: Abasic sites, photodimers, alkylation adducts, and oxidative lesions, *Biochemistry, 48*, 9347-9359.

(104) Shrivastav, N.; Li, D.; Essigmann, J. M. , **(2010),** Chemical biology of mutagenesis and DNA repair: Cellular responses to DNA alkylation, *Carcinogenesis, 31*, 59-70.

(105) Hecht, S. S.; Hoffmann, D. , **(1988),** Tobacco-specific nitrosamines, an important group of carcinogens in tobacco and tobacco smoke, *Carcinogenesis, 9*, 875-884.

(106) Pfeifer, G. P.; Denissenko, M. F.; Olivier, M.; Tretyakova, N.; Hecht, S. S.; Hainaut, P. , **(2002),** Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers, *Oncogene, 21*, 7435-7451.

(107) Hecht, S. S. , **(2003),** Tobacco carcinogens, their biomarkers and tobacco-induced cancer, *Nat. Rev. Cancer, 3*, 733-744.

(108) Hecht, S. S. , **(2008),** Progress and challenges in selected areas of tobacco carcinogenesis, *Chem. Res. Toxicol., 21*, 160-171.

(109) Sen, N. P.; Miles, W. F.; Donaldson, B.; Panalaks, T.; Iyengar, J. R. , **(1973),** Formation of nitrosamines in a meat curing mixture, *Nature, 245*, 104-105.

(110) Bartsch, H.; Montesano, R. , **(1984),** Relevance of nitrosamines to human cancer, *Carcinogenesis, 5*, 1381-1393.

(111) Lindamood, C.,3rd; Bedell, M. A.; Billings, K. C.; Dyroff, M. C.; Swenberg, J. A. , **(1984),** Dose response for DNA alkylation, [^{3}H]thymidine uptake into DNA, and O6-methylguanine-DNA methyltransferase activity in hepatocytes of rats and mice continuously exposed to dimethylnitrosamine, *Cancer Res., 44*, 196-200.

(112) Sedgwick, B.; Vaughan, P. , **(1991),** Widespread adaptive response against environmental methylating agents in microorganisms, *Mutat. Res., 250*, 211-221.

(113) Bianchini, F.; Wild, C. P. , **(1994),** 7-methyldeoxyguanosine as a marker of exposure to environmental methylating agents, *Toxicol. Lett., 72*, 175-184.

(114) Swenberg, J. A.; Boysen, G.; Georgieva, N.; Bird, M. G.; Lewis, R. J. , **(2007),** Future directions in butadiene risk assessment and the role of cross-species internal dosimetry, *Chem. Biol. Interact., 166*, 78-83.

(115) Swenberg, J. A.; Bordeerat, N. K.; Boysen, G., et al , **(2011),** 1,3-butadiene: Biomarkers and application to risk assessment, *Chem. Biol. Interact., 192*, 150-154.

(116) Rydberg, B.; Lindahl, T. , **(1982),** Nonenzymatic methylation of DNA by the intracellular methyl group donor S-adenosyl-L-methionine is a potentially mutagenic reaction, *EMBO J., 1*, 211-216.

(117) Haines, J. A.; Reese, C. B.; Todd, L. , **(1962),** The methylation of guanosine and related compounds with diazomethane, *J. Chem. Soc.*, 5281-5288.

(118) Beranek, D. T.; Weis, C. C.; Evans, F. E.; Chetsanga, C. J.; Kadlubar, F. F. , **(1983),** Identification of N^5-methyl-N^5-formyl-2,5,6-triamino-4-hydroxypyrimidine as a major adduct in rat liver DNA after treatment with the carcinogens, N,N-dimethylnitrosamine or 1,2-dimethylhydrazine, *Biochem. Biophys. Res. Commun., 110*, 625-631.

(119) Boiteux, S.; Laval, J. , **(1983),** Imidazole open ring 7-methylguanine: An inhibitor of DNA synthesis, *Biochem. Biophys. Res. Commun., 110*, 552-558.

(120) O'Connor, T. R.; Boiteux, S.; Laval, J. , **(1988),** Ring-opened 7-methylguanine residues in DNA are a block to in vitro DNA synthesis, *Nucleic Acids Res., 16*, 5879-5894.

(121) Asagoshi, K.; Terato, H.; Ohyama, Y.; Ide, H. , **(2002),** Effects of a guanine-derived formamidopyrimidine lesion on DNA replication: Translesion DNA synthesis, nucleotide insertion, and extension kinetics, *J. Biol. Chem., 277*, 14589-14597.

(122) Earley, L. F.; Minko, I. G.; Christov, P. P.; Rizzo, C. J.; Lloyd, R. S. , **(2013),** Mutagenic spectra arising from replication bypass of the 2,6-diamino-4-hydroxy-N^5-methyl formamidopyrimidine adduct in primate cells, *Chem. Res. Toxicol., 26*, 1108-1114.

(123) Christov, P. P.; Angel, K. C.; Guengerich, F. P.; Rizzo, C. J. , **(2009),** Replication past the N5-methyl-formamidopyrimidine lesion of deoxyguanosine by DNA polymerases and an improved procedure for sequence analysis of in vitro bypass products by mass spectrometry, *Chem. Res. Toxicol., 22*, 1086-1095.

(124) Christov, P. P.; Yamanaka, K.; Choi, J., et al , **(2012),** Replication of the 2,6-diamino-4-hydroxy-N5-(methyl)-formamidopyrimidine (MeFapy-dGuo) adduct by eukaryotic DNA polymerases, *Chem. Res. Toxicol., 25*, 1652-1661.

(125) Tano, K.; Shiota, S.; Collier, J.; Foote, R. S.; Mitra, S. , **(1990),** Isolation and structural characterization of a cDNA clone encoding the human DNA repair protein for O6-alkylguanine, *Proc. Natl. Acad. Sci. U. S. A., 87*, 686-690.

(126) Aldape, K.; Zadeh, G.; Mansouri, S.; Reifenberger, G.; von Deimling, A. , **(2015),** Glioblastoma: Pathology, molecular mechanisms and markers, *Acta Neuropathol., 129*, 829-848.

(127) Gates, K. S.; Nooner, T.; Dutta, S. , **(2004),** Biologically relevant chemical reactions of N7-alkylguanine residues in DNA, *Chem. Res. Toxicol., 17*, 839-856.

(128) Tomasz, M.; Lipman, R.; Lee, M. S.; Verdine, G. L.; Nakanishi, K. , **(1987),** Reaction of acid-activated mitomycin C with calf thymus DNA and model guanines: Elucidation of the base-catalyzed degradation of N7-alkylguanine nucleosides, *Biochemistry, 26*, 2010-2027.

(129) Haraguchi, K.; Greenberg, M. M. , **(2001),** Synthesis of oligonucleotides containing Fapy·dG (N6- (2-deoxy-α,β-d-erythro-pentofuranosyl)-2,6- diamino-4-hydroxy-5-formamidopyrimidine), *J. Am. Chem. Soc., 123*, 8636-8637.

(130) Burgdorf, L. T.; Carell, T. , **(2002),** Synthesis, stability, and conformation of the formamidopyrimidine G DNA lesion, *Chem. Eur. J., 8*, 293-301.

(131) Brown, K. L.; Deng, J. Z.; Iyer, R. S., et al , **(2006),** Unraveling the aflatoxin-FAPY conundrum: Structural basis for differential replicative processing of isomeric forms of the formamidopyrimidine-type DNA adduct of aflatoxin B_1, *J. Am. Chem. Soc., 128*, 15188-15199.

(132) Büsch, F.; Pieck, J. C.; Ober, M., et al , **(2008),** Dissecting the differences between the α and β Anomers of the oxidative DNA lesion FaPydG, *Chem. Eur. J, 14*, 2125-2132.

(133) Christov, P. P.; Banerjee, S.; Stone, M. P.; Rizzo, C. J. , **(2010),** Selective incision of the α-N5-methyl-formamidopyrimidine anomer by *escherichia coli* endonuclease IV, *J. Nucleic Acids., 2010,* 850234.

(134) Tudek, B. , **(2003),** Imidazole ring-opened DNA purines and their biological significance, *J. Biochem. Mol. Biol., 36*, 12-19.

(135) Boiteux, S.; Belleney, J.; Roques, B. P.; Laval, J. , **(1984),** Two rotameric forms of open ring 7-methylguanine are present in alkylated polynucleotides, *Nucleic Acids Res., 12*, 5429-5439.

(136) Humphreys, W. G.; Guengerich, F. P. , **(1991),** Structure of formamidopyrimidine adducts as determined by NMR using specifically ^{15}N-labeled guanosine, *Chem. Res. Toxicol., 4*, 632-636.

(137) Raoul, S.; Bardet, M.; Cadet, J. , **(1995),** Gamma irradiation of 2'-deoxyadenosine in oxygen-free aqueous solutions: Identification and conformational features of formamidopyrimidine nucleoside derivatives, *Chem. Res. Toxicol., 8*, 924-933.

(138) Patro, J. N.; Haraguchi, K.; Delaney, M. O.; Greenberg, M. M. , **(2004),** Probing the configurations of formamidopyrimidine lesions fapy.dA and fapy.dG in DNA using endonuclease IV, *Biochemistry, 43*, 13397-13403.

(139) Croy, R. G.; Wogan, G. N. , **(1981),** Temporal patterns of covalent DNA adducts in rat liver after single and multiple doses of aflatoxin B_1, *Cancer Res., 41*, 197-203.

(140) Alekseyev, Y. O.; Hamm, M. L.; Essigmann, J. M. , **(2004),** Aflatoxin B_1 formamidopyrimidine adducts are preferentially repaired by the nucleotide excision repair pathway in vivo, *Carcinogenesis, 25*, 1045-1051.

(141) Vartanian, V.; Minko, I. G.; Chawanthayatham, S., et al , **(2017),** NEIL1 protects against aflatoxin-induced hepatocellular carcinoma in mice, *Proc. Natl. Acad. Sci. U. S. A., 114*, 4207-4212.

(142) Luna, L.; Bjørås, M.; Hoff, E.; Rognes, T.; Seeberg, E. , **(2000),** Cell-cycle regulation, intracellular sorting and induced overexpression of the human NTH1 DNA glycosylase involved in removal of formamidopyrimidine residues from DNA, *Mutat. Res. , 460*, 95-104.

(143) Eide, L.; Bjørås, M.; Pirovano, M.; Alseth, I.; Berdal, K. G.; Seeberg, E. , **(1996),** Base excision of oxidative purine and pyrimidine DNA damage in saccharomyces cerevisiae by a DNA glycosylase with sequence similarity to endonuclease III from escherichia coli, *Proc. Natl. Acad. Sci. U. S. A., 93*, 10735-10740.

(144) Katafuchi, A.; Nakano, T.; Masaoka, A., et al , **(2004),** Differential specificity of human and escherichia coli endonuclease III and VIII homologues for oxidative base lesions, *J. Biol. Chem., 279*, 14464-14471.

(145) Awasthi, R.; Roseblade, A.; Hansbro, P. M.; Rathbone, M. J.; Dua, K.; Bebawy, M. , **(2018),** Nanoparticles in cancer treatment: Opportunities and obstacles, *Curr. Drug Targets, 19*, 1696-1709.

(146) Willker, W.; Leibfritz, D.; Kerssebaum, R.; Bermel, W. , **(1993),** Gradient selection in inverse heteronuclear correlation spectroscopy, *Magn. Reson. Chem., 31*, 287-292.

(147) Aue, W. P.; Bartholdi, E.; Ernst, R. R. , **(1976),** Two-dimensional spectroscopy. application to nuclear magnetic resonance, *J. Chem. Phys., 64*, 2229-2246.

(148) Bax, A.; Davis, D. G.; Sarkar, S. K. , **(1985),** An improved method for two-dimensional heteronuclear relayed-coherence-transfer NMR spectroscopy, *J. Magn. Reson., 63*, 230-234.

(149) Jeener, J.; Meier, B. H.; Bachmann, P.; Ernst, R. R. , **(1979),** Investigation of exchange processes by 2-dimensional NMR spectroscopy, *J. Chem. Phys., 71*, 4546-4553.

(150) Wagner, R.; Berger, S. , **(1996),** Gradient-selected NOESY - A fourfold reduction of the measurement time for the NOESY experiment, *J. Magn. Reson. A, 123*, 119-121.

(151) Kew, M. C. , **(2013),** Aflatoxins as a cause of hepatocellular carcinoma, *J. Gastrointestin. Liver Dis., 22*, 305-310.

(152) McCormack, A. I.; McDonald, K. L.; Gill, A. J., et al , **(2009),** Low O6-methylguanine-DNA methyltransferase (MGMT) expression and response to temozolomide in aggressive pituitary tumours, *Clin. Endocrinol. (Oxf), 71*, 226-233.

(153) Lee, S. Y. , **(2016),** Temozolomide resistance in glioblastoma multiforme, *Genes Dis., 3*, 198-210.

(154) Hamm, M. L.; Gill, T. J.; Nicolson, S. C.; Summers, M. R. , **(2007),** Substrate specificity of fpg (MutM) and hOGG1, two repair glycosylases, *J. Am. Chem. Soc., 129*, 7724-7725.

(155) Pujari, S. S.; Tretyakova, N. , **(2017),** Chemical biology of N^5-substituted formamidopyrimidine DNA adducts, *Chem. Res. Toxicol., 30*, 434-452.

(156) Asagoshi, K.; Yamada, T.; Terato, H., et al , **(2000),** Distinct repair activities of human 7,8-dihydro-8-oxoguanine DNA glycosylase and formamidopyrimidine DNA glycosylase for formamidopyrimidine and 7,8-dihydro-8-oxoguanine, *J. Biol. Chem., 275*, 4956-4964.

(157) Christov, P. P.; Brown, K. L.; Kozekov, I. D.; Stone, M. P.; Harris, T. M.; Rizzo, C. J. , **(2008),** Site-specific synthesis and characterization of oligonucleotides containing an N6-(2-deoxy-d-erythro-pentofuranosyl)-2,6-diamino-3,4-dihydro-4-oxo-5-N-methylformamidopyrimidine lesion, the ring-opened product from N7-methylation of deoxyguanosine, *Chem. Res. Toxicol., 21*, 2324-2333.

(158) Goddard, T. D.; Kneller, D. G. , **(2006),** SPARKY v. 3.113, *University of California, San Francisco* .

(159) Merrifield, R. B. Solid Phase Synthesis. *https://www.nobelprize.org/prizes/chemistry/1984/merrifield/lecture/*.

(160) Merrifield, R. B. , **(1963),** Solid phase peptide synthesis. I. the synthesis of a tetrapeptide, *J. Am. Chem. Soc., 85*, 2149-2154.

(161) Gutte, B.; Merrifield, R. B. , **(1971),** The synthesis of ribonuclease A, *J. Biol. Chem., 246*, 1922-1941.

(162) McBride, L. J.; Caruthers, M. H. , **(1983),** An investigation of several deoxynucleoside phosphoramidites useful for synthesizing deoxyoligonucleotides, *Tetrahedron Lett., 24*, 245-248.

(163) Elmquist, C. E.; Stover, J. S.; Wang, Z.; Rizzo, C. J. , **(2004),** Site-specific synthesis and properties of oligonucleotides containing C8-deoxyguanosine adducts of the dietary mutagen IQ, *J. Am. Chem. Soc., 126*, 11189-11201.

(164) Cavaluzzi, M. J.; Borer, P. N. , **(2004),** Revised UV extinction coefficients for nucleoside-5'-monophosphates and unpaired DNA and RNA, *Nucleic Acids Res., 32*, e13.

(165) Piotto, M.; Saudek, V.; Sklenar, V. , **(1992),** Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions, *J. Biomol. NMR, 2*, 661-665.

(166) Doddrell, D. M.; Pegg, D. T.; Bendall, M. R. , **(1982),** Distortionless enhancement of NMR signals by polarization transfer, *J. Magn. Reson., 48*, 323-327.

(167) Patra, A.; Banerjee, S.; Johnson Salyard, T. L., et al , **(2015),** Structural basis for error-free bypass of the 5-N-methylformamidopyrimidine-dG lesion by human DNA polymerase ? and sulfolobus solfataricus P2 polymerase IV, *J. Am. Chem. Soc., 137*, 7011-7014.

(168) Wiederholt, C. J.; Greenberg, M. M. , **(2002),** Fapy·dG instructs klenow exo$^-$ to misincorporate deoxyadenosine, *J. Am. Chem. Soc., 124*, 7278-7279.

(169) Kalam, M. A.; Haraguchi, K.; Chandani, S., et al , **(2006),** Genetic effects of oxidative DNA damages: Comparative mutagenesis of the imidazole ring-opened

formamidopyrimidines (fapy lesions) and 8-oxo-purines in simian kidney cells, *Nucleic Acids Res., 34*, 2305-2315.

(170) Aramini, J. M.; Cleaver, S. H.; Pon, R. T.; Cunningham, R. P.; Germann, M. W. , **(2004),** Solution structure of a DNA duplex containing an α-anomeric adenosine: Insights into substrate recognition by endonuclease IV, *J. Mol. Biol., 338*, 77-91.

(171) Asagoshi, K.; Yamada, T.; Okada, Y., et al , **(2000),** Recognition of formamidopyrimidine by escherichia coli and mammalian thymine glycol glycosylases. distinctive paired base effects and biological and mechanistic implications, *J. Biol. Chem., 275*, 24781-24786.

(172) Wiederholt, C. J.; Delaney, M. O.; Pope, M. A.; David, S. S.; Greenberg, M. M. , **(2003),** Repair of DNA containing fapy.dG and its beta-C-nucleoside analogue by formamidopyrimidine DNA glycosylase and MutY, *Biochemistry, 42*, 9755-9760.