# Structural Consequences of the C8-Guanine DNA Adduct Formed by 3-Nitrobenzanthrone; an Environmental Carcinogen

By

Dustin A. Politica

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

August, 2016

Nashville, Tennessee

Approved:

Michael P. Stone, Ph.D.

Martin Egli, Ph.D.

Jens Meiler, Ph.D.

Carmelo J. Rizzo, Ph.D.

# Acknowledgements

I am grateful to a great many people for their assistance, encouragement, and friendship during this exciting and challenging journey. I would never have been able to get this far on my own, and to all of those that have been there along the way, you have my eternal thanks.

First and foremost I would like to thank my loving family. Thank you Astha, my wife and best friend, for the love, encouragement, support, (and distraction when necessary) you have provided to help make this all possible. To my mother, father, brother, and sister; your love, encouragement, and assistance made it possible for me to begin this journey, and has been a continued blessing. All of my friends and extended family, especially the members of the Miller family, have also been very supportive and encouraging, and I am very thankful for their part in this experience.

I would like to thank Dr. Michael P. Stone for accepting me into his research group and providing me with a wealth of opportunities. I particularly appreciate his encouragement and support in attending various conferences and meetings that have contributed to my growth as a scientist and professional. As a member of his lab, I have had the pleasure of working with some amazing individuals, which have contributed to my growth and success substantially. I would like to acknowledge Dr. Ewa Kowal and Dr. Kallie Stavros for providing exceptional support and guidance during the early stages of my graduate career. Andrew Kellum, Ken Dempster, Ryan Bowen, and other former and present members of the Stone group have also contributed significantly to the quality of my experience at Vanderbilt, and been helpful in my research, teaching, and other endeavors.

I would like to thank the other members of my committee, Dr. Martin Egli, Dr. Jens Meiler, and Dr. Carmelo Rizzo, for their helpful advice and direction. Their knowledge and expertise has helped lead me in positive directions and avoid certain pitfalls. I would like to provide special acknowledgement to Dr. Martin Egli for providing me with the opportunity to work in his lab and gain valuable experience in crystallography. He has also provided me with invaluable advice in many matters, and his support and guidance has been instrumental in my success. In addition, I would like to acknowledge Dr. Amritraj Patra for his assistance and guidance during my experience in Dr. Egli's lab.

The collaborative efforts of the Basu group at the University of Connecticut have been instrumental in the work described in this document. I would like to specifically thank Dr. Ashis Basu, Dr. Chanchal Malik, and Mr. Arindom Chatterjeee. Without their efforts in the synthesis of the modified oligonucleotides none of this work would have been possible.

Many other members of Vanderbilt's staff and faculty have been exceptional in a number of ways. I would like to thank Ms. Sandra Ford for introducing me to the graduate program at Vanderbilt and providing assistance in many things, administrative and otherwise, over the course of the past four years. I would like to thank Dr. Jennifer Ufnar for her continuing support and advice in my teaching and career endeavors, and many other matters. Dr. Tara Todd has been very helpful in providing guidance and assistance with my experiences in teaching at Vanderbilt. Dr. Ruth Schemmer has been outstanding in providing career support and guidance. I would also like to thank Dr. Steve Townsend. He has been exceedingly helpful, going above and beyond his role at Vanderbilt, and I have greatly appreciated his help and guidance. The Vanderbilt Biomolecular NMR and Mass Spectrometry Core staffs, including Dr. Markus Voehler and Dr. Wade Calcutt respectively, have provided outstanding support and guidance.

Finally I would like to thank the Vanderbilt Community. All of the other faculty, staff, and students at Vanderbilt help make it a great place to be and bring the exceptional to life.

Lastly I would like to acknowledge funding support provided by Vanderbilt University, the National Institute of Health (NIH), and the Vanderbilt Institute of Chemical Biology (VICB).

# Table of Contents

**Page**

IV. Summary

# List of Tables

**Page**

# List of Figures

**Page**

# List of Charts

**Page**

# Chapter I

## Introduction

### Deoxyribonucleic Acids (DNA)

Deoxyribonucleic acid (DNA) is one of the main biological macromolecules required for life. It has been established as the primary genetically transferable material and acts as the "blueprint" for all known living organisms. DNA provides the code for the function, growth, and development of an organism. This code is built from four nucleobases: adenine, cytosine, guanine, and thymine.[1] In unison they form a biopolymer that is organized into a hierarchical structure. Two strands of the biopolymer associate to form a double helical structure that associates with specific proteins that allow for higher levels of structure and organization into clusters of DNA known as chromosomes.[1-2] The end result is a stable molecule capable of storing large amounts of information that can be passed from generation to generation.[3]

*DNA Composition and Structure*

Despite the large size and capacity of DNA to store a tremendous amount of information, it is fundamentally composed of only four different building blocks. These blocks are the nucleosides deoxyadenosine (dA), deoxycytidine (dC), deoxyguanosine (dG), and deoxythymidine (dT). The monophosphate varieties of these are represented in **Chart 1**. The nucleosides are composed of base and deoxyribose sugar groups built from carbon, oxygen, nitrogen, and hydrogen. A phosphorus containing backbone links the nucleosides together forming a biopolymer, or nucleotide chain.[1, 4] Each chain has two ends characterized by

termination at either a nucleotide sugar's 5′ or 3′ hydroxyl group, thus labeled the 5′ or 3′ ends. The DNA chains pair with a complementary chain in an anti-parallel fashion, placing the 5′ end of one chain adjacent to the 3′ end of another chain.[5] The two chains are held together by hydrogen bonding interactions between bases from each chain. The bases pair in a consistent manner in which adenine and thymine bases form pairs and guanine and cytosine bases form pairs, or are "complements" of one another.[6]

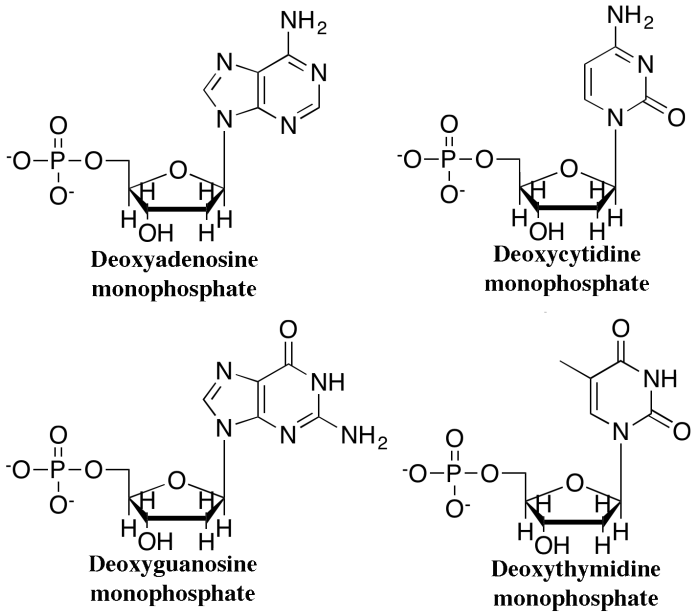



**Chart 1.** Chemical structures of the four nucleotide building blocks of DNA. The purines adenine and guanine are depicted on the left and the pyrimidines cytosine and thymine appear on the right.

The association of two chains of DNA results in a stable double-helical structure. Variations on this structure exist, but the most common form is known as B-type DNA (**Figure 1**). The structure of this form was first reported by James Watson, Francis Crick, Maurice Wilkins, and Rosalind Franklin in 1953.[5] This form is characterized by a double-helical structure

with a right-handed twist possessing two internal grooves or clefts, designated as the major or minor groove based on their consistent variation in size. Two other double-helical forms of DNA have been identified and designated as A-type or Z-type DNA[7] (**Figure 1**). The A-type DNA is also right-handed but possesses a deep, narrower major groove and shallow minor groove. Z-type DNA progresses as a left-handed double helix with similar major and minor grooves.[8] Some additional non-helical structural motifs have also been observed including G-quadruplex structures that occur at the ends of chromosomes.[9] Further variations on structure have also been observed in cases of modified DNA or DNA in complex with protein.

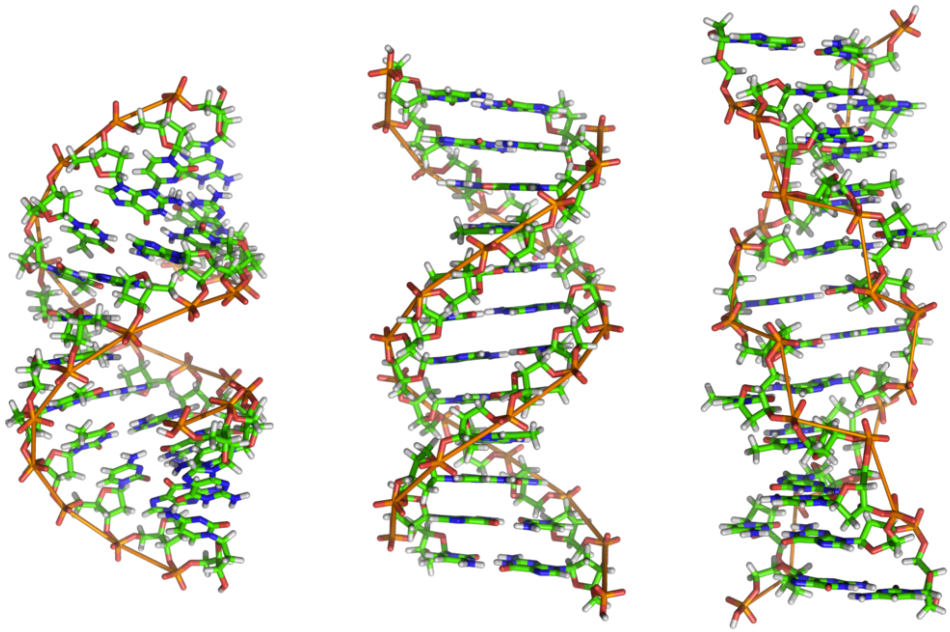


**Figure 1**. Representative structures from left to right of A-type, B-type, and Z-type DNA.[*] A-type and B-type DNA display a right-handed helical form with the A-form possessing a wider curvature and a deeper major groove. Z-type displays a left-handed curvature and is elongated in comparison to B-type.

---

[*] By Original uploader was Richard Wheeler (Zephyris) at en.wikipedia - Originally from en.wikipedia; description page is/was here., GFDL, https://commons.wikimedia.org/w/index.php?curid=2117121

*Heritability*

One of the most important properties of DNA is its ability to be passed through generations of living organisms. An organism's entire genetic code is known as its genome. To reproduce, the genome must be passed to any offspring. This occurs through sexual or asexual reproduction. In sexual reproduction an offspring's genome is composed of a partial genome provided from two parents to generate a complete genome that is a hybrid of that from both parents.[10] In asexual reproduction the offspring's genome is provided by a single parent organism and is a complete copy of the parent's genome.[11]

A process known as DNA replication is essential for reproduction in all living organisms, as well as the growth and development of multicellular organisms. Every time a cell divides its genome must be copied. This process is accomplished by a set of proteins that enact a coordinated process to generate exact copies of a cell's DNA. At the heart of this process is a protein known as a DNA-polymerase.[12] A polymerase will bind to a single strand of DNA and use it as a template to produce a copy of its complementary strand. The polymerase is capable of identifying the bases present in a strand of DNA and matching them with their complementary partner.[13] Reading from the 3′ end to the 5′ end of a strand of DNA the polymerase will synthesize a copy of the complementary strand through a series of polymerization reactions.[14] This process is incredibly accurate with a fidelity level of less than one mistake for every $10^9$ nucleotides. The insertion fidelity of replicative polymerases is very high and is further aided by a proofreading capacity that occurs before a nucleobase pair can exit the protein.[15]

In order for the polymerase to serve its function, a number of other proteins are required to perform other functions. The DNA must be unwound and separated. This is accomplished by topoisomerase[16] and helicase proteins.[17] Single strand binding proteins stabilize the single strand

form of DNA and prevent reannealing of the two strands.[18] One strand of the DNA must be copied in fragments known as Okazaki fragments. These must be joined together by a protein called DNA-ligase.[19] The DNA-polymerase requires a primer segment to initiate replication. This primer is produced by DNA-primase.[20] This complex is illustrated in **Figure 2**.

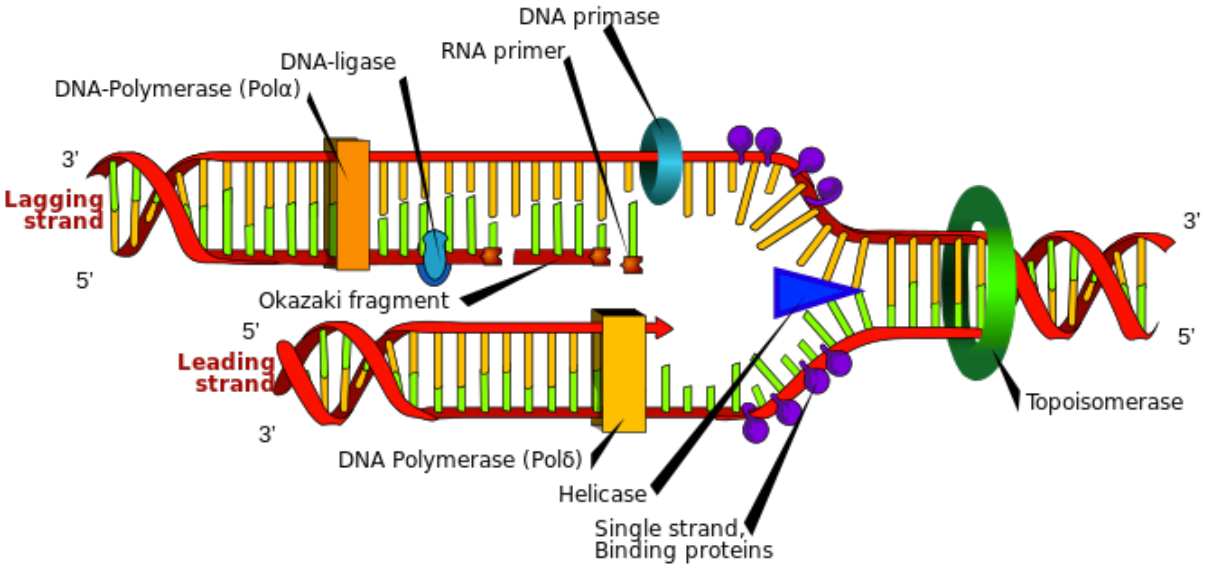



**Figure 2**. Cartoon illustrating a DNA replication complex.[‡] The parent DNA duplex is on the right and is followed by the topoisomerase responsible for relieving supercoiling. The helicase protein then aids to further unwind and separate the duplex, in which single strands are then stabilized by the single strand binding proteins. Primase inserts short primer sequences required by the DNA-polymerases. Replication is accomplished by the DNA-polymerases and is continuous for the leading strand and performed in fragments on the lagging strand. These fragments are connected by DNA-ligase.

[‡]By LadyofHats Mariana Ruiz - Own work. Image renamed from File:DNA replication.svg, Public Domain, https://commons.wikimedia.org/w/index.php?curid=691697

*DNA Damage and Repair*

For a long time it was believed that DNA was incredibly stable and highly resistive to damage. If intact DNA is required to maintain proper cellular function and to pass genetic information from generation to generation, then it must remain largely unchanged over significant lengths of time. We now know that while DNA is quite stable, it is not impervious to damage. Many types of DNA damage have been observed and cells must have mechanisms capable of repairing or otherwise managing DNA damage events.[21] The importance of DNA repair is now well recognized and was the focus of the 2015 Nobel prize in Chemistry awarded to Aziz Sancar, Paul Modrich, and Tomas Lindahl for work on "the molecular mechanisms of DNA repair processes."[22]

DNA damage results from both endogenous and exogenous sources. Cells can experience greater than 10,000 DNA damage events each day.[21] Endogenous sources of these events can result as unintended consequences of normal cellular metabolism and signaling.[23] Normal metabolic processes can produce reactive oxygen species capable of causing DNA damage.[24] One commonly found example of oxidative DNA damage is the DNA lesion 8-oxoguanine (**Chart 2**).[25] In addition to oxidation, DNA damage can result from methylation, deamination, and depurination/depyrimidification.

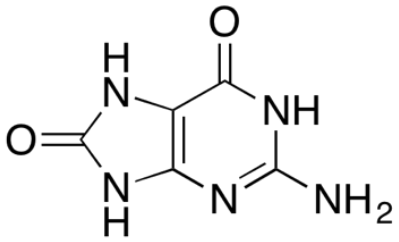



**Chart 2.** Chemical structure of the DNA lesion 8-oxoguanine.

Many sources of exogenous DNA damage exist. Ionizing radiation is a major source of concern. UV light is capable of directly and indirectly causing DNA damage. When UV radiation interacts directly with DNA it can produce radicals on purine bases, which can react with neighboring purine bases to form dimers.[26] Damage can occur indirectly when UV radiation interacts with other molecules, such as water, to produce reactive species which subsequently damage DNA. Other types of UV and radiative DNA damage exist but environmental exposure to chemicals is also of concern. DNA damage has been found to result from exposure to cigarette smoke,[27] exhaust fumes,[28] food-borne sources,[29] and many additional sources. Aflatoxin is a chemical produced by certain fungi that contaminate food sources, notably maize and peanuts, and has been directly connected with the etiology of liver cancer through the formation of DNA adducts.[30]

Given the range of potential sources of DNA damage and the frequency of damaging events, cells must possess mechanisms capable of addressing DNA damage. Repair of DNA damage is commonly enacted by either the base-excision repair (BER) or nucleotide excision repair systems (NER). Utilizing enzymes known as glycosolases, BER will repair damage to individual nucleobases.[31] The damaged base is removed leaving an abasic site in the DNA. A polymerase will then insert a nucleotide that is complementary to the opposing strand. The BER system is capable of handling many types of damage but is not effective at addressing large or bulky DNA adducts.[32] These adducts must be processed by the NER system. The NER system is composed of a set of proteins that recognize DNA damage and excise larger segments of DNA.[33] Instead of removing a single base like the BER system, NER typically removes sections of a DNA strand ranging from 10 to 30 nucleotides in length.[34] The excised section will then be

resynthesized to complete repair of the damage. This process is capable of addressing many forms of damage including both small and large DNA adducts.[35]

## Diesel Exhaust and Nitroarenes

### *Diesel Exhaust*

Complete combustion of many carbon compounds, including fossil fuels, results in products of carbon dioxide and water. However, in many cases combustion is not complete and a number of additional products can be formed. This is the case for diesel engines. In addition to carbon dioxide and water, complex mixtures of other products are formed during the combustion of diesel fuel.[28] These include certain inorganic compounds, such as sulfur dioxide and nitrogen oxide, as well as a large variety of organic compounds. Some of these compounds can pose certain risks to the environment, and plant, animal, and human health.

Carbon dioxide emissions have been implicated as a risk involving the earth's climate and global warming.[36] Sulfur dioxide and nitrogen oxide are associated with the formation of acid rain.[37] These factors can have environmental impacts and damage plant life and various ecosystems.[38] Other compounds that result from diesel combustion can be toxic and have deleterious effects on human health.

Epidemiological evidence first suggested a role for diesel exhaust in the development of cancer. A study conducted in 1982 examined the rates of cancer in coal miners occupationally exposed to high levels of diesel emissions.[39] These workers experienced increased rates of cancer, with a particular emphasis on lung cancer. Following this finding, a number of studies were performed to examine the impact of diesel exhaust on carcinogenesis. One study focusing

on railroad workers exposed to diesel exhaust showed a 40% increased risk for lung cancer.[40] Studies with dock workers, truck drivers, and other occupations have also supported these findings.[41]

*In vivo* and *in vitro* studies have also supported the role of diesel exhaust in lung carcinogenesis. Diesel exhaust has been examined as whole exhaust and as its separate gas and particulate matter phases. Rats exhibited increased tumor incidence when exposed to whole diesel exhaust.[42] Rats and mice exposed to particulate matter also experienced increased rates of tumor formation.[43] Gas phase diesel exhaust did not result in increased tumor formation. Exposure to diesel exhaust and diesel particulate matter has also been shown to result in DNA damage. Oxidative damage, bulky adducts, strand breaks, morphological cell transformation in mammalian cells, and other genotoxic effects have been observed.[44]

These studies provide strong evidence for a carcinogenic effect resulting from diesel exhaust. These and other findings have led the International Association for Research on Cancer (IARC) to classify diesel exhaust as carcinogenic to humans.[45] Additional factors do affect these results. Not all diesel exhaust is the same. Emissions control systems can have a significant effect on the composition of diesel exhaust.[46] The worst offenders tend to be in industrial settings and countries without emission control regulations. By contrast, modern diesel vehicles in areas with more stringent emission controls show marked improvement.[47] This suggests that the most significant threat is posed for workers in certain industries, and individuals exposed to diesel exhaust in areas without tight emission controls.

Since diesel exhaust is a complex mixture that can vary noticeably in its composition, the question of what within diesel exhaust is responsible for its carcinogenicity becomes important. One suspected source is a class of compounds known as nitroarenes.[48]

*Nitroarenes*

Nitroarenes are a class of related compounds that possess a polycyclic aromatic hydrocarbon (PAH) group and a nitro group. Many nitroarenes have been identified as components of diesel exhaust and are suspected contributors to its carcinogenicity.[48] Some of these compounds are 1-nitropyrene, 2-nitroflourene, 3-nitrobenzanthrone, 3-nitroflouranthene, and 6-nitrochrysene, shown in **Chart 3**.[45] Nitroarenes also have been found to originate from other sources, such as coal fly ash and cigarette smoke,[45] and have been identified as contaminants in rain water, soil samples, and other environmental sources.[49]

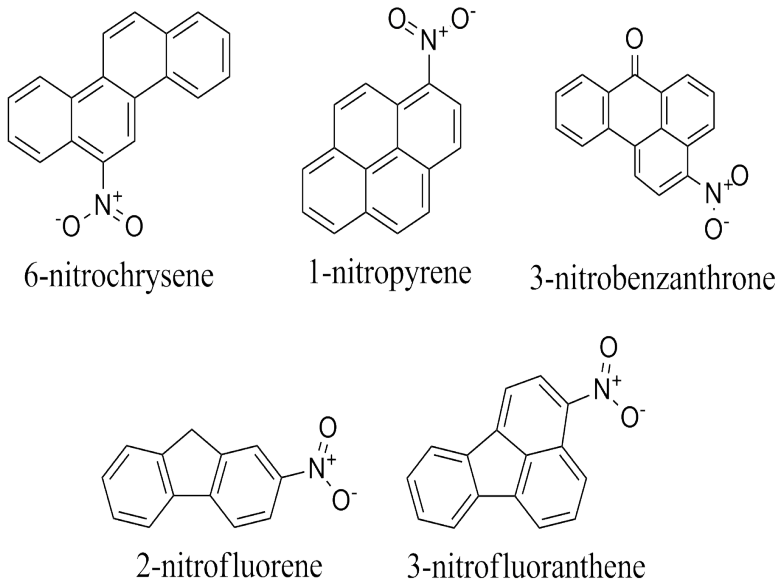
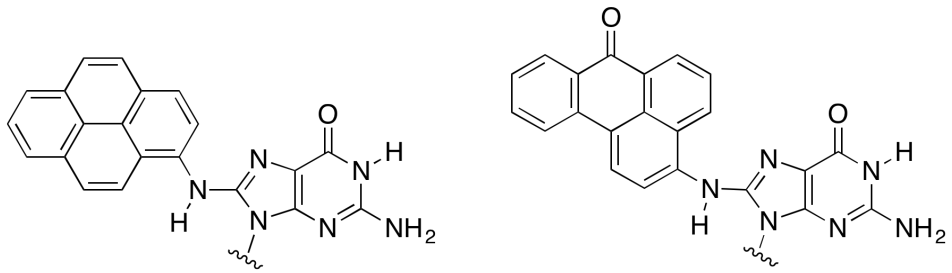



**Chart 3.** Chemical structure of five species of nitroarenes found within diesel exhaust.

Within diesel exhaust, 1-nitropyrene (1-NP) has been identified as the most abundant nitroarene present,[48] while 3-nitrobenzanthrone (3-NBA) is the most mutagenic nitroarene identified thus far.[50] Both of these compounds have been found to produce DNA adducts, and result in major adducts on the C8 position of guanine (**Chart 4**). As with other nitroarenes, this

reaction is believed to require enzymatic activation to electrophilic species.[51] This reaction occurs as part of a two-phase process in which the compounds are first converted to an *N*-hydroxy derivative followed by acetylation or sulfonation at the hydroxyl group.[52] This results in a good leaving group that, through hydrolysis, can produce reactive nitrenium ions.[53] In the case of both 1-NP and 3-NBA, these nitrenium ions can react with guanine bases at the C8 position to produce C8-guanine adducts; however, resonance of the nitrenium ion can alter the reactive electrophile in ways that allow additional products to form.[54] The types of products formed vary based on the chemical structure of the nitroarene in question.




C8-dG-AP C8-dG-ABA

**Chart 4**. Chemical structures of C8 guanine adducts formed by 1-nitropyrene (1-NP) and 3-nitrobenzanthrone (3-NBA). (Left) The aminopyrene adduct formed by 1-NP. (Right) The aminobenzathrone adduct formed by 3-NBA.

1-NP, 3-NBA and other nitroarenes have been found to be mutagenic.[55] 1-NP mutations appear to be somewhat dependent on cell type. Certain studies have indicated that +1 or -1 frameshift mutations are most frequent.[56] These mutations were commonly found at GC repeats in which a GC pair was either added or removed. Another study examined multiple cell types and found that G to T transversions were the most common mutation in all types but there were variances in the extent to which they were favored. Furthermore, when the lesion was next to a

5-methylcytosine a significant increase in G to A transitions was observed.[57] In the case of 3-NBA, either G to T transversions or G to A transitions have been reported as the primary mutations.[58] It appears that cell type and/or sequence context may play a role in determining the mutational spectrum for these adducts.[58b] Investigation into the nitroarene class of compounds is ongoing; however, current evidence has led the IARC to list a number of these compounds as probable of possible human carcinogens.[45]

## Structural Determination of Biomolecules

Identifying the structure of biomolecules can provide crucial understanding of how they function. Given the insight that can be gleaned from structural knowledge of biomolecules, considerable effort has been put forth to examine and understand many biomolecules including proteins, nucleic acids, carbohydrates, and other molecules. The small size of these molecules makes this a challenging endeavor as traditional microscopy is not sufficient to examine these molecules. A number of techniques have been established to aid in the identification of molecular structure. While no single technique is capable of addressing all structure related questions, up to this point two techniques have stood at the forefront as a result of their capacity to provide atomic level resolution in a variety of systems. These are nuclear magnetic resonance spectroscopy (NMR) and X-ray crystallography.

### *Nuclear Magnetic Resonance Spectroscopy (NMR)*

The powerful technique of NMR has been heavily utilized to identify and understand molecular structure for a wide variety of molecules. NMR has long been in use for the structural characterization of small synthetic molecules. It is a standard technique utilized in many

chemical syntheses. As NMR technology has improved, it has grown in the span of its potential applications and its use in the characterization of biomolecules is now prevalent.

The power of NMR lies in its capacity to collect data about atomic nuclei. Atomic nuclei have spins that are subject to manipulation by magnetic fields. NMR subjects nuclei to strong magnetic fields to align their spins to the Z-axis of the field indicated as $B_0$.[59] The nuclear spins can point in the direction of the magnetic field or opposite it. Nuclei pointing in the direction of the field are labeled as +1/2 and possess a lower energy level compared to those aligned opposite the field, designated as -1/2.[59] To obtain a signal, the NMR will excite nuclei with a radio frequency (RF) pulse. This will alter the direction of the spins to a desired level, most commonly 90º, to create a form of coherence between the +1/2 and -1/2 spin states. From this state the nuclei are allowed to relax back to their equilibrium states resulting in a weak, but detectible RF signal. This is recorded as a free induction decay, or FID.[60] While the signal recorded is weak, repeated acquisitions can be collected and averaged to obtain a more sensitive measurement.

The type of atomic nucleus can have a significant impact on the application of NMR.[60] The protons and neutrons in the nucleus of an atom each possess a natural characteristic known as nuclear spin. Spins can pair up and cancel each other out. Only atoms that possess a net spin other than zero are NMR active. This phenomenon presents certain limitations for NMR. Most elements have an isotope that is NMR active, however, the natural abundance of the isotope may be relatively low. Hydrogen atoms are one of the primary types of atoms examined by NMR. The $^{1}H$ isotope is NMR active and has a natural abundance of 99.9%.[60] This makes hydrogen an ideal atom type to examine with NMR. On the other hand, the isotope of carbon that is NMR active is $^{13}C$. This isotope has a natural abundance of only 1.1%.[60] This means that nearly 99% of the carbon in any sample does not produce a signal by NMR. While carbon is often examined by

NMR, the sensitivity of the measurements is greatly decreased in comparison to hydrogen, unless measures are taken to increase the abundance of $^{13}C$. Because of the compatibility of hydrogen with NMR and its presence it biomolecules, examination of hydrogen by NMR is routine. Other elements, such as carbon and nitrogen, are commonly found in biomolecules. Knowledge of these atoms can be critical in understanding the structural properties of a biomolecule through NMR.[61] Therefore, examination of these types of atoms is also common by NMR, though the issues with reduced sensitivity must be taken into account. As previously mentioned, NMR is an inherently low sensitivity technique, and preforming repeated acquisitions and averaging the signals can overcome this. When an NMR active isotope is in low abundance, a greater number of acquisitions can be completed to increase the overall signal recorded. In some cases the nuclei of interest can be enriched during sample preparation to improve the sensitivity or measurements taken.[61] In either case, the final signal obtained for the element is improved.

Resonance frequency is another property that depends on the type of element being examined by NMR.[61] The resonance frequency of a given atom is dependent upon the type of atom and the strength of the applied magnetic field. Atoms of the same element will display very similar resonance frequencies while the resonance frequencies of different elements can vary dramatically. For instance, in a magnetic field where hydrogen displays resonance at 500 MHz, carbon will have resonance at approximately 125 MHz, and nitrogen at approximately 51 MHz.[61] This is another factor that must be taken into account when performing an NMR experiment.

Biomolecular NMR. Biomolecular NMR is the use of NMR to study biomolecules, most commonly proteins and nucleic acids. The use of NMR to examine biomolecules is a relatively recent development in the history of NMR. Advancements in NMR technology and the advent of

multidimensional NMR have allowed for many applications of NMR that were previously impractical or even impossible.

As previously mentioned, atoms of the same element display very similar resonance frequencies. The chemical environment of an atom can produce small shifts in the resonance frequency for that specific atom, known as chemical shifts. In this way, atoms of a particular element will be spread out among a small range in frequency producing a spectrum. As the number of atoms for a specific element increase, the probability for overlapping and unresolved signals also increases. Biomolecules are typically quite large compared to most synthetic molecules and often result in NMR spectra that are highly overlapped and not well resolved. Improved strength of the magnetic field can help address this issue as a stronger field increases the separation of signals and improves resolution. Modern spectrometers can now exceed magnetic field strengths of 1 GHz, which is a tremendous improvement over field strengths of 30 MHz found in instruments present in 1950. The development of multidimensional NMR experiments has also been key in the application of NMR to biomolecules.[61] Transitioning from one-dimensional to two-dimensional experiments can greatly improve the resolution of atomic signals in an NMR spectrum and allow for assignment of a spectrum that could not be accomplished in only one dimension.[61] These improvements have greatly benefited the field of biomolecular NMR; however, limitations still exist in the size of systems that can be practically examined. In general, the current size limit is around 40 kDa.

Two-dimensional NMR experiments fall into one of two categories: homonuclear, or heteronuclear. Homonuclear experiments focus on observing a single atom type, most commonly $^{1}H$. Some examples of homonuclear experiments are correlation spectroscopy (COSY) or nuclear Overhauser effect spectroscopy (NOESY). COSY is utilized to identify spins that are coupled to

each other through covalent bonds.[61] Each atom observed in a COSY experiment will display a characteristic chemical shift. When two atoms are coupled to each other, they will produce a cross-peak in the COSY spectrum. The cross-peak occurs in the spectrum at a location where the characteristic chemical shift of one atom is matched to the X-axis and the characteristic chemical shift of the other atom is matched to the Y-axis. This allows chemical shifts and through-bond interactions to be identified. A NOESY experiment can be extremely valuable in determining biomolecular structure. This is due to the fact that signals in a NOESY experiment result from through-space interactions of atomic nuclei.[62] Atomic nuclei experience dipolar interactions with other nearby nuclei. NOESY experiments are designed to observe these interactions and can provide data about relative distances between nuclei. The functional limit for measuring these interactions is approximately 5 Å. The spectrum of a NOESY experiment is similar to a COSY experiment in that cross-peaks occur in the spectrum at a location where the characteristic chemical shift of one atom is matched to the X-axis and the characteristic chemical shift of the other atom is matched to the Y-axis. The relative intensities of cross-peaks in a NOESY spectrum are related to the distance between atoms. Measuring the intensities of cross-peaks, and comparing them to an interaction in the spectrum with a known distance, can allow for distance determinations between atoms to be made. These distances can then be used to guide structural elucidation of a biomolecule.

Heteronuclear experiments observe interactions between two different types of nuclei, such as $^{1}H$ and $^{13}C$ or $^{1}H$ and $^{15}N$. An example of a common heteronuclear experiment is the heteronuclear single-quantum correlation spectroscopy (HSQC) experiment. This experiment detects through-bond correlations of two different nuclei separated by one bond.[63] A HSQC will provide cross-peaks in a spectrum that represent these correlations. Like the previous

experiments mentioned, the location of the peak will match the chemical shift for one atom on the X-axis and the other atom on the Y-axis. Heteronuclear experiments differ from homonuclear experiments in that the two axes belong to two types of atoms and possess different scales, while homonuclear spectra have matching axes.

Use of these experiments, or a combination of these experiments and other techniques, can provide valuable structural information for many biomolecules. In many cases they have allowed for the determination of a three-dimensional structure of a biomolecule. This is accomplished through a series of computational refinements. Acquisition of NMR data is typically performed in aqueous solution, which is believed to provide an environment that resembles the environment that soluble biomolecules experience inside of cells. In addition, information regarding dynamics of some biomolecules can be gained through NMR. This can provide additional insight into changes in conformation of the biomolecules alone or in the presence of other molecules such as small organic molecules. Thus, the structural information garnered is valuable in understanding the biological context of many biomolecules.

*X-ray Crystallography*

No other technique has provided the same extent of information regarding biomolecular structure as X-ray diffraction. X-ray diffraction was instrumental in identifying the now canonical structure of DNA.[5] In addition, it has been applied to identify the structure of proteins, RNA, carbohydrates, viruses, and other important biomolecules. In 1914, Max von Laue was awarded the Nobel Prize for the discovery that crystals could diffract X-rays.[64] Some 50 years later, James Watson and Francis Crick reported the structure of DNA using X-ray diffraction, for which they were awarded a Nobel Prize in 1962. Fast forward to modern day and X-ray

diffraction has contributed to work leading to many more Nobel Prizes, and a great many more important scientific discoveries.[65]

X-rays are a form of electromagnetic radiation with a wavelength ranging from approximately 0.1 Å to 100 Å. X-ray wavelengths used in crystallography are typically close to 1 Å. Upon contact with a molecularly ordered crystal, X-rays have been found to scatter in a pattern that correlates to the arrangement of atoms within the crystal. This is what is known as a diffraction pattern (**Figure 3**). The scattering of X-rays is a result of interactions with the electron clouds that surround atoms. Thus, the diffraction pattern is a function of the electron density as it occurs in a crystal.[65] The diffraction pattern appears as a series of spots with molecular information represented by the spacing and relative intensities of the spots. Spacing of spots correlate with molecular size while intensities are dependent upon the arrangement of molecules. Intensities are related to the phase difference and the amplitude of the diffracted waves. Amplitudes can be calculated but phase determination is an underlying problem in diffraction and presents an obstacle to solving crystal structures. This is known as the phase problem.[66]

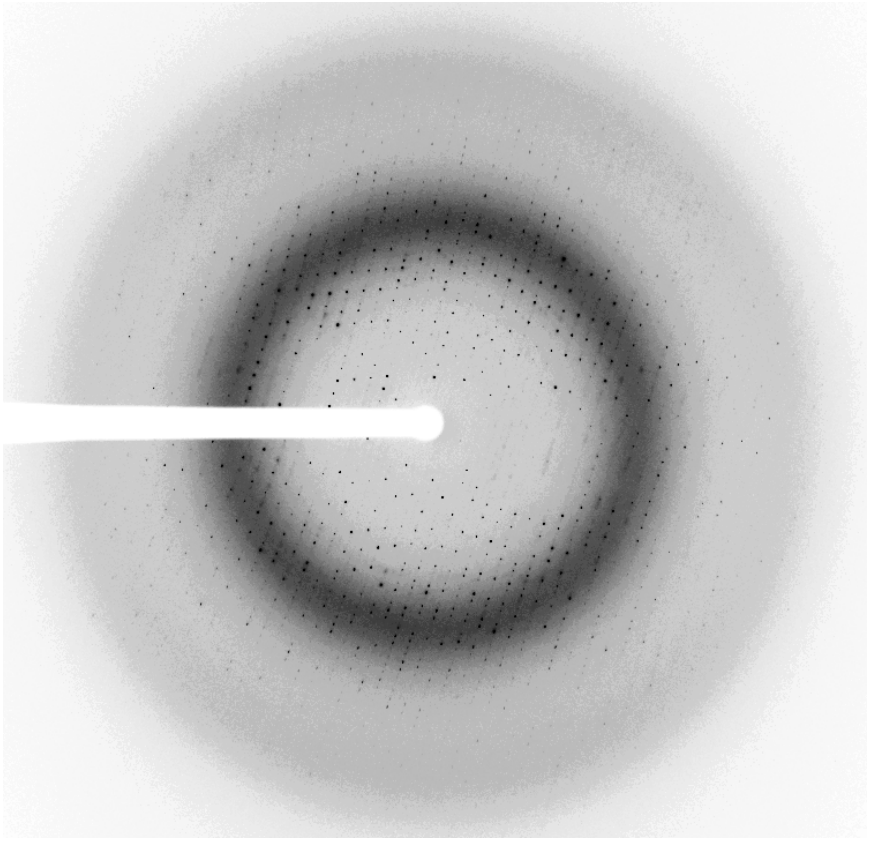


**Figure 3**. Example of a biomolecular X-ray diffraction pattern.

In relation to biomolecules, there are three common methods used to address the phase problem. These are anomalous X-ray scattering (MAD/SAD), molecular replacement (MR), and multiple isomorphous replacement (MIR). MAD and SAD are similar methods that are utilized to solve the phase problem when other homologous structures are not available. With multi-wavelength anomalous diffraction (MAD), data is collected for a crystal at a series of different wavelengths.[67] Dispersive differences can then be used to solve the phase problem. Single-wavelength anomalous diffraction (SAD) focuses on a single wavelength to collect a single dataset and is used when the stability of a crystal is a concern.[68] Replacement of a molecular atom in the crystal with a heavier atom, such as replacing sulfur with selenium, improves both of these methods.

MR is the simplest of these methods to utilize, however, it requires a homologous structure to have been previously determined. The phase data from a previously determined structure is applied to the current dataset to generate the electron density map. The molecule of interest is then fit to the electron density map to solve the structure.[69] An example of a molecule fit to electron density is shown in **Figure 4**.

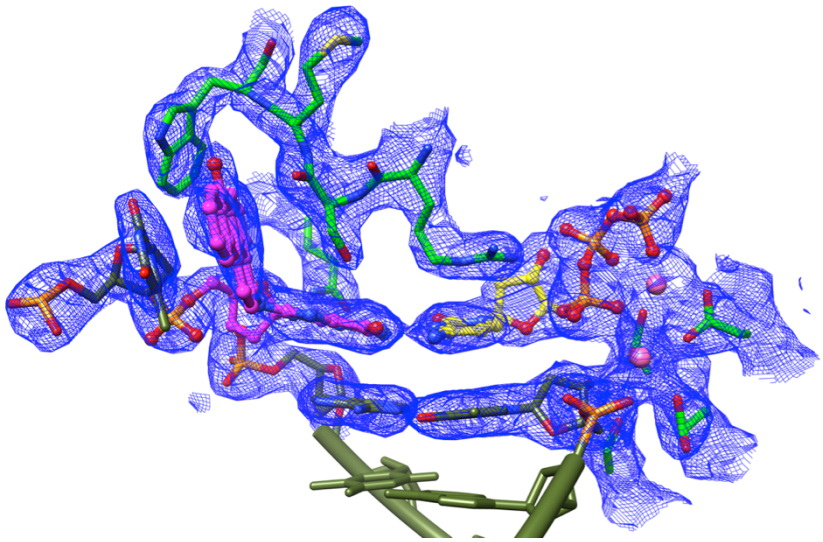


**Figure 4**. Molecular model fit to an electron density map produced through molecular replacement. Electron density is shown as the blue mesh. Molecular structure is shown as the ball and stick model within the blue mesh.

MIR requires the presence of a heavy atom within the crystal. Crystals can be grown in the presence of the heavy atom or the crystal can be soaked with the heavy atom to introduce it after the crystal has formed. The heavy atom will introduce differences in scattering amplitudes.[65] This can allow the phase problem to be solved. This method requires data to be

obtained from crystals with and without the heavy atom, and requires that the presence of the heavy atom does not result in a change in crystal structure.

Additional parameters that can be determined from diffraction data include unit cell dimensions, space group, and crystal system. A crystal's unit cell has defined three-dimensional lengths and angles. These are defined as lengths a, b, and c, and angles α, β, and γ. The unit cell will have a characteristic shape which defines the crystal system. The relationship of these is shown in **Table 1**. The symmetry of the diffraction pattern correlates to the space group. In the case of biomolecules, which possess chirality, 65 space groups are possible. 230 space groups are possible when chirality may or may not be present.

**Table 1**. Relationship of cell geometries to crystal systems and space groups.

| Cell Geometry | Crystal System | Space Groups Possible |
|---|---|---|
| a=b=c; α, β, γ = 90º | Cubic | 36 |
| a=b=c; α=β=γ≠90º | Rhombohedral | 25 |
| a=b≠c; α, β, γ = 90º | Tetragonal | 68 |
| a=b≠c; α, β=90º γ = 120º | Hexagonal | 27 |
| a≠b≠c; α,γ = 90º β,≠90º | Monoclinic | 13 |
| a≠b≠c; α=β=γ≠90º | Orthorhombic | 59 |
| a≠b≠c; α≠β≠γ | Triclinic | 2 |

To complete the crystal diffraction and subsequent structural determination, a crystal or set of crystals must first be produced. To accomplish this, a highly purified sample of the protein or biomolecule of interest must be obtained. This can be challenging and often requires multiple

steps in expression and purification. Additionally, there are no set conditions in which all biomolecules will form crystals. Often it is necessary to examine a large set of conditions to identify the conditions that are conducive to crystal formation for a specific biomolecule. In many cases this is the rate-limiting step in obtaining a crystal structure. Some parameters that can be varied to identify crystallization conditions include solution pH, salt concentration and type of salt present, protein or biomolecule concentration, temperature, addition of polyethylene glycols (PEG), and others.[65] The method utilized to form crystals is also important.

Modern crystallography frequently utilizes either the hanging drop or sitting drop method. Schematics of these methods are shown in **Figure 5**. In both cases, droplets containing the biomolecule of interest and other solution components are suspended above a reservoir containing a more concentrated crystallization solution. Water will diffuse from the droplets to the reservoir increasing the concentration of the biomolecule within the drop. As time progresses, the concentration will continue to increase and favor nucleation of a crystal. Resulting crystals can be identified under a microscope in a matter of days to weeks. Any crystals formed need to be placed in cryoprotectant and flash-frozen in liquid nitrogen prior to examination.

Not all crystals will diffract, however, high quality crystals subjected to a powerful X-ray source, such as a synchrotron insertion beamline, are capable of providing resolution on the order of 1 Å. This is very high resolution data and is not obtained in most cases, though resolution in the 2-3 Å range is more common and generally is capable of providing an excellent structural understanding of a biomolecule. One point of caution is that crystals do not mimic the natural environment of biomolecules well, and in some instances artifacts have been identified. This should be considered when evaluating the quality of a resultant crystal structure.

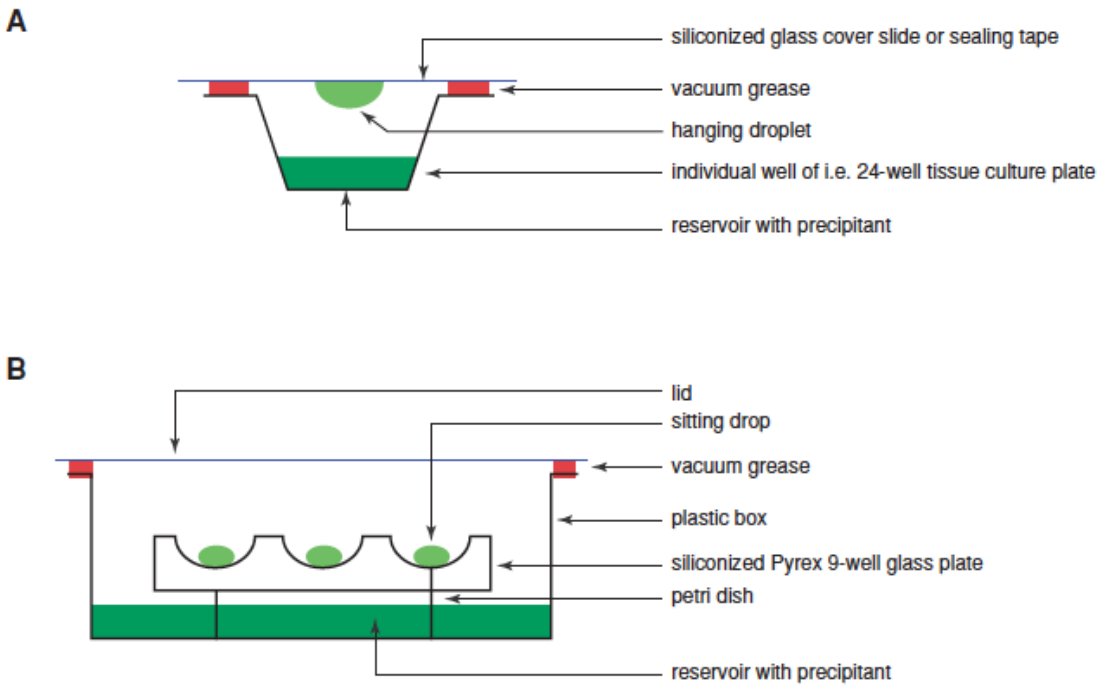



**Figure 5**. Two related methods for growing single crystals of biomacromolecules. Schematic depictions of the **(A)** hanging drop and **(B)** sitting drop vapor diffusion techniques.[65]

## Dissertation Statement

The goal of this project was to identify significant structural effects that result from the C8-guanine-aminobenzanthrone (C8-dG-ABA) DNA adduct. It was of interest to examine the impact of C8-dG-ABA on a typical B-type DNA duplex. It was expected that the bulky nature of this adduct could significantly disrupt the normal duplex, and that intercalation of the adduct was likely due to similarities in structure to other bulky C8-guanine adducts. It was also of interest to investigate the conformation of C8-dG-ABA in complex with a DNA polymerase. Of the potential polymerases to examine, polymerase eta (Pol η) was selected due to the previously reported capacity of Pol η to bypass the adduct in both an error-free and mutagenic fashion.

Chapter II reports the NMR solution structure of the C8-dG-ABA duplex and includes relevant materials and methods. Chapter III reports the crystal structure of C8-dG-ABA in complex with Pol η in the error-free insertion state. Relevant materials and methods are also included. Chapter IV discusses the relationship between the reported structures and proposes avenues for future study.

# Chapter II*

## Base-Displaced Intercalated Structure of the *N*-(2´-Deoxyguanosin-8-yl)-3-aminobenzanthrone DNA Adduct[†]

### Introduction

The nitroarene 3-nitrobenzanthrone (3-NBA, 3-nitro-7*H*-benz[*de*]anthracen-7-one) (**Chart 5**), a byproduct of incomplete combustion,[70] is found in diesel exhaust[70b, 71] and as an environmental contaminant.[70a, 72] It has one of the highest reported levels of mutagenesis in the Ames assay.[73] 3-NBA is approximately three orders of magnitude more mutagenic than the polycyclic aromatic hydrocarbon benzo[*a*]pyrene (B[*a*]P)[74] or the most abundant nitroarene found in diesel exhaust, 1-nitropyrene[75] (1-NP). 3-NBA causes DNA damage[76] and has exhibited mutagenicity in mammalian cells as well.[76c, 76d, 77] It produces micronuclei in human B-lymphoblastoid cells,[78] human hepatoma cell lines,[79] and mouse peripheral blood reticulocytes.[80] Furthermore, cells exposed to 3-NBA possess increased levels of reactive oxygen species.[81] Human exposure to 3-NBA has been documented through identification of metabolites in urine.[82] These and other findings have led the International Agency for Research on Cancer (IARC) to classify 3-NBA as a class 2B (possibly carcinogenic to humans) chemical and whole diesel exhaust as a class 1 compound (carcinogenic to humans).[83]

---

[†] Reproduced with permission from Politica, D. A.; Malik, C. K.; Basu, A. K.; Stone, M. P., Base-Displaced Intercalated Structure of the N-(2′-Deoxyguanosin-8-yl)-3-aminobenzanthrone DNA Adduct. *Chemical Research in Toxicology* **2015,** *28* (12), 2253-2266. Copyright 2015 American Chemical Society. Available online at http://pubs.acs.org/doi/abs/10.1021/acs.chemrestox.5b00277

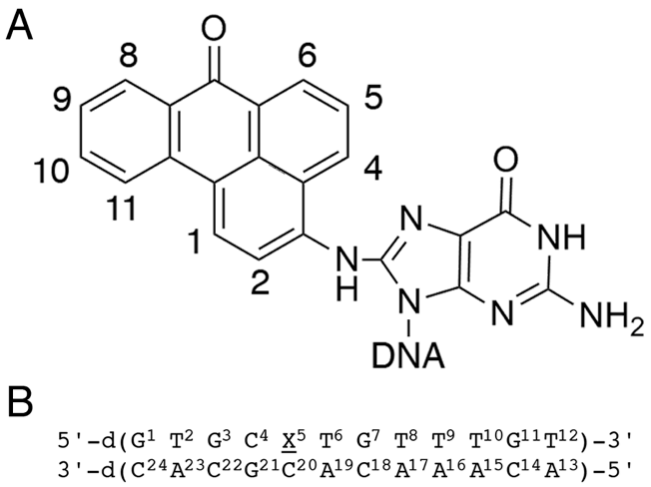



**Chart 5. A**. Chemical structure of the C8-dG-ABA adduct including numbering scheme of the adduct protons for NMR. **B**. The oligodeoxynucleotide sequence used in this work, showing numbering of the individual nucleotides.

3-NBA undergoes enzymatic nitroreduction to 3-aminobenzanthrone (3-ABA) *in vivo*.[82, 84] Activated intermediates of this conversion include the nitrenium ion, the proximate mutagenic species. This ion exists as a hybrid of two resonance structures, which can alkylate DNA at either the exocyclic nitrogen or the C2-position of the benzanthrone ring (**Chart 6)**.[71c, 80, 85] Three DNA adducts have been identified. The major alkylation product, *N*-(deoxyguanosin-8-yl)-3-aminobenzanthrone (C8-dG-ABA), forms at the C8 position of guanine. In addition, alkylation may occur at the $N^2$-dG position and the $N^6$-dA position, to form 2-(deoxyguanosin-$N^2$-yl)-3-aminobenzanthrone ($N^2$-dG-ABA) and 2-(deoxyadenosin-$N^6$–yl)-3-aminobenzanthrone ($N^6$-dA-ABA), respectively (**Chart 7)**.[76d, 86]

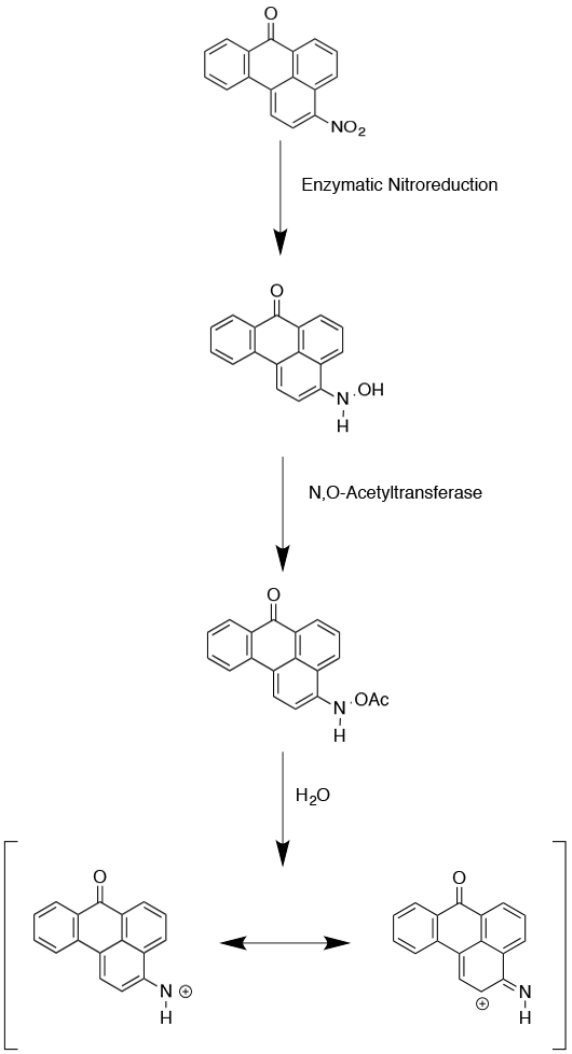



**Chart 6**. Activation of 3-NBA into reactive electrophiles. Phase I activation occurs through nitroreduction. Phase II activation can occur through acetylation or sulfonation of the hydroxyl group (acetylation shown). Final activation occurs through solvolysis to form a nitrenium ion.

The three adducts are repaired with differing efficiencies by the nucleotide excision repair (NER) proteins.[76d] In cultured human cells, the $N^2$-dG-ABA adduct is the least efficiently repaired while the C8-dG-ABA and $N^6$-dA-ABA adducts are repaired with approximately 3.5 and 4.5 times greater efficiencies, respectively.[76d] If not repaired, each of these three adducts blocks replicative DNA polymerases but is bypassed to different degrees by Y-family

polymerases during translesion synthesis (TLS).[76d, 87] The C8-dG-ABA adduct presents the most significant block to TLS, allowing only 17% bypass, as compared to 33% for the $N^2$-dG-ABA adduct and 43% for the $N^6$-dA-ABA adduct after a 72 h incubation period.[76d] Bypass of these adducts during TLS is mutagenic.[76d, 77d, 88] Depending on the type of assay, both G to T transversions[36] and G to A transitions[14] have been reported as the primary mutations in mammalian cells. The types of cells, the DNA sequence context, and whether the adducts are situated in a normal duplex or a bubble region are factors thought to modulate the mutagenic spectra of these adducts.[87]

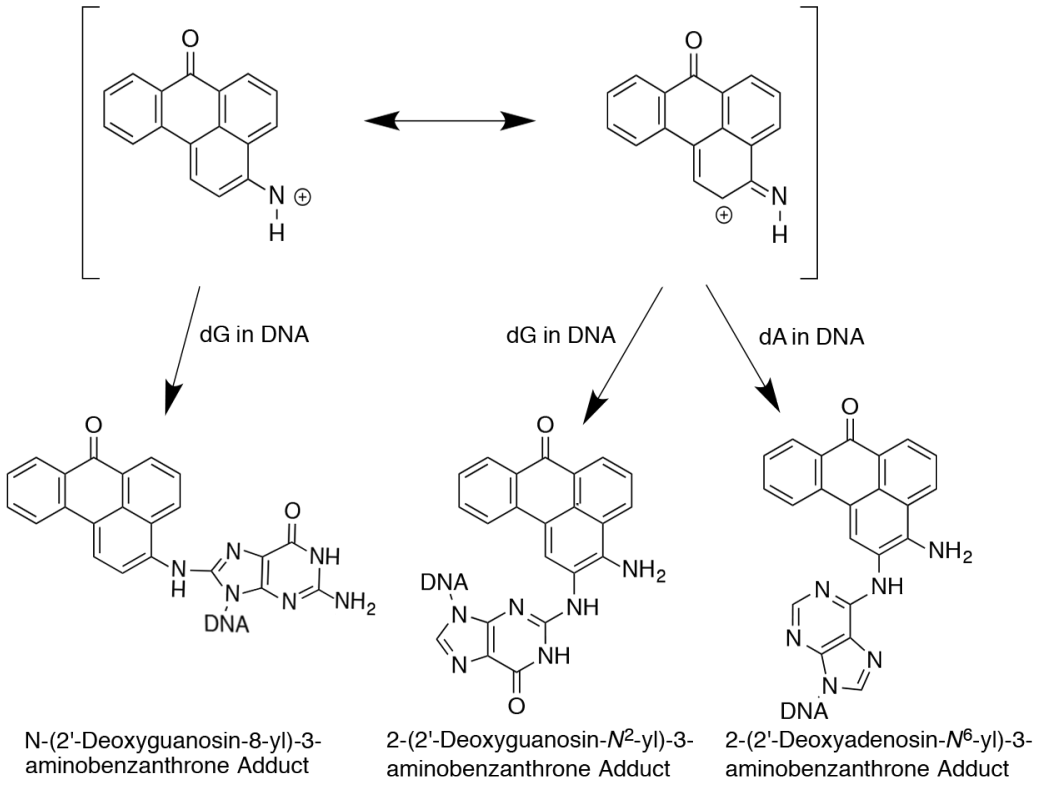



**Chart 7**. Reaction of the electrophilic nitrenium ion of 3-NBA with DNA to form three DNA aminobenzanthrone adducts. These occur at the C8 position of guanine, or the $N^2$ position of guanine and the $N^6$ position of adenine to produce the three adducts shown.

The conformations of the $N^2$-dG-ABA, C8-dG-ABA, and $N^6$-dA-ABA adducts have been of interest; it has been anticipated that differences in their conformations and thermodynamic properties in DNA might correlate with differences in their biological processing, both with respect to DNA repair and translesion replication. Lukin *et al.*[89] reported the site-specific synthesis of the $N^2$-dG-ABA adduct and its conformation in an oligodeoxynucleotide duplex. They showed that the 3-ABA moiety resided in the minor groove and was oriented toward the 5′-end of the modified strand. Of note, the $N^2$-dG-ABA adduct increased the thermal melting temperature ($T_m$) of the duplex, which, in principle, is consistent with its slower repair and its persistence *in vivo*.[76d] The observation that the C8-dG-ABA adduct is repaired with greater efficiency by NER[76d] suggested that it might instead decrease the $T_m$ of an oligodeoxynucleotide duplex. Indeed, like other bulky C8-dG arylamine adducts, such as those formed by aminopyrene (AP),[90] *N*-acetyl-aminofluorene (AAF),[91] aminofluorene (AF),[92] 2-amino-1-methyl-6-phenylimidazo [4,5-*b*] pyridine (PhIP),[93] and 2-amino-3-methylimidazo [4,5-*f*] quinoline (IQ),[94] it seemed conceivable that the C8-dG-ABA adduct might be oriented in either a base-displaced intercalated or minor groove external conformation in this sequence context. Previous reports indicated that these bulky C8-dG adducts tended to adopt one of these conformations when placed in sequence between two pyrimidine nucleotides.[89]

We report here our findings for the three-dimensional structure and conformation of the C8-dG-ABA adduct in DNA. We examined conformational perturbations induced by this DNA adduct in the dodecamer 5'-d(GTGCXTGTTTGT)-3':5'-d(ACAAACACGCAC)-3'; X = C8-dG-ABA adduct. This sequence includes codons 272-275 of *p*53 gene, in which 3-NBA-induced G to T mutations have been reported.[58c] We employed NMR spectroscopy to elucidate the effect of the C8-dG-ABA adduct on the conformation of this duplex. We also employed UV/vis

spectroscopy to investigate its effect on the duplex melting temperature ($T_m$). We show that in this sequence context the C8-dG-ABA adduct forms a base-displaced intercalated conformation.

## Materials and Methods

*Sample Synthesis and Characterization*

The synthesis and purification of the C8-dG-ABA modified oligodeoxynucleotide 5′-d(GTGCXTGTTTGT)-3′, X = C8-dG-ABA, was performed as previously reported.[77d] The unmodified single strand oligodeoxynucleotides 5′-d(GTGCGTGTTTGT)-3′ and 5′-d(ACAACACGCAC)-3′ were obtained from the Midland Certified Reagent Co. (Midland, TX). Their purity was verified with capillary gel electrophoresis and reverse-phase HPLC (Phenomenex Gemini C-18 250 × 10 mm column) utilizing a gradient of acetonitrile in a mobile phase containing 0.1 M ammonium formate (pH 7). In all cases, following a 5 min equilibration period with 5% acetonitrile, a gradient from 5% to 16% acetonitrile was employed over 30 min. The flow rate was 2.0 mL $min^{-1}$. Samples were purified by HPLC under the same conditions. The samples were deemed to be ≥ 99% pure following this procedure. Purified oligomers were desalted by passing over G-25 Sephadex. They were dried by evaporation using a centrivap apparatus (Labconco, Kansas City, MO). Oligodeoxynucleotide concentrations were determined by absorbance spectroscopy measurements at 260 nm. The calculated extinction coefficients[95] were 1.09 x $10^5$ L $mol^{-1}$ $cm^{-1}$ for the modified strand and 1.21 x $10^5$ L $mol^{-1}$ $cm^{-1}$ for the complementary strand. The calculated extinction coefficient of the modified strand was not corrected for the presence of the ABA adduct. The modified and unmodified oligodeoxynucleotides were each annealed with the complementary strand in 10 mM $NaH_2PO_4$,

100 mM NaCl, and 50 μM $Na_2EDTA$ (pH 7). The samples were heated to 85 °C for 10 min and then cooled to room temperature to form either the unmodified DNA duplex or the duplex containing the C8-dG-ABA adduct. To establish the precise 1:1 strand stoichiometry, the duplex samples were subjected to DNA grade hydroxyapatite chromatography.[96] A gradient from 10 mM to 100 mM $NaH_2PO_4$ (pH 7) over 120 min was employed. The final strand stoichiometry was verified both by capillary gel electrophoresis and HPLC chromatography.

*Mass Spectrometry*

Single strand oligodeoxynucleotides were characterized using a Voyager MALDI-TOF spectrometer (Applied Biosystems, Grand Island, NY). A matrix consisting of 3-hydroxypicolinic acid in ammonium hydrogen citrate was used. Mass spectra were recorded in the negative ion mode and recorded to ± 1 *m/z*. Mass spectrometry of the duplex containing the C8-dG-ABA adduct was performed using a SYNAPT LC-TOF mass spectrometer (Waters Corp, Milford, MA). UHPLC was performed using an Acquity UPLC BEH C18 2.1 × 50 mm column (Waters Corp) with a gradient from 5% to 15% acetonitrile in 0.1 M ammonium acetate (pH 7) over 15 min. The data collection window was set to observe masses between 300-2000 *m/z* in the negative ion mode.

*DNA Melting Temperatures*

Melting temperature determinations for the modified and unmodified duplexes were obtained using a Cary UV/vis spectrometer (Varian Associates, Palo Alto, CA). The sample absorbance was monitored at 260 nm. The sample concentrations were 2 μM in 10 mM $NaH_2PO_4$, 100 mM NaCl, and 50 μM $Na_2EDTA$ (pH 7). The thermal scan was performed from 5

°C to 85 °C in 1 °C $min^{-1}$ increments. $T_m$ values were determined using the first derivatives of the experimentally obtained absorbance vs. temperature plots.

*NMR Spectroscopy*

The duplexes were prepared in 10 mM $NaH_2PO_4$, 100 mM NaCl, and 50 μM $Na_2EDTA$ (pH 7), in volumes of 180 μL or 500 μL, utilized in either 3 mm or 5 mm NMR tubes, respectively. For experiments to examine non-labile protons samples were prepared in 99.996% $D_2O$. Examination of labile protons was performed in 9:1 $H_2O$:$D_2O$. Water suppression was accomplished using the Watergate pulse program.[97] All spectra were processed using the TOPSPIN software package (Bruker Biospin Inc., Billerica, MA) and further analyzed using TOPSPIN and SPARKY[98] software. Spectra were referenced to the chemical shift of the water resonance at the corresponding temperature, with respect to trimethylsilyl propanoic acid (TSP).

Unmodified Duplex. Spectra were recorded at 1 mM concentration. NOESY[99] and COSY experiments in $D_2O$ were performed on a 900 MHz spectrometer (Bruker Biospin, Inc., Billerica, MA) at 15 °C with 2048 real data points in the $t_2$ dimension and 512 real data points in the $t_1$ dimension. The NOESY spectrum was obtained using a mixing time of 250 ms. The NOESY and COSY spectra were zero-filled during processing to obtain final matrices of 2048 × 1024 data points.

C8-dG-ABA Duplex. NOESY and magnitude COSY spectra were obtained at 900 MHz at a sample concentration of 350 μM in 500 μL. The temperature was 25 °C. NOESY and COSY spectra were completed with 2048 real data points in the $t_2$ dimension and 512 real data points in the $t_1$ dimension. The NOESY spectra were obtained using mixing times of 80, 150, 200, and 250 ms. All spectra were zero-filled during processing to obtain final matrices of 2048 × 1024

data points. The NOESY spectra in 9:1 $H_2O$:$D_2O$ were obtained on Bruker 800 MHz and Bruker 900 MHz spectrometers at 5 °C and 15 ºC respectively. The mixing time was 250 ms. The spectra were obtained with 2048 real data points in the $t_2$ dimension and 512 real data points in the $t_1$ dimension and then zero-filled during processing to obtain a matrix of 2048 × 1024 data points. Additional NOESY and magnitude COSY spectra were performed at 900 MHz at a concentration of 520 μM in 180 μL. The temperature was 15 °C. Other experimental conditions and spectra processing were the same as those used at 25 ºC.

*NMR Distance Restraints*

The program SPARKY[98] was utilized to determine volume integrations of NOESY cross-peaks from 15 °C and 25 °C spectra with 250 ms mixing times. These integrations were also performed using 150 ms mixing time spectra. The intrinsic error of integrations was assigned as half of the volume of the lowest intensity cross-peaks. The confidence levels of the integrations were divided into five categories based on peak shape, peak intensity, degree of overlap, and proximity to the water resonance. A 10% error value was assigned to well-resolved and non-overlapping cross-peaks. Strong but slightly broadened or overlapped cross-peaks or peaks with moderate S/N were given a 20% error value. A 30% error value was assigned to strong but medially broadened or overlapped cross-peaks. Cross-peaks with weak S/N or slightly overlapped cross-peaks with moderate S/N were assigned a 40% error value. A 50% error value was assigned to cross-peaks near the diagonal or water suppression, highly broadened or overlapped cross-peaks, and cross-peaks with moderate S/N and medial overlap or broadening. Integration values of NOEs with assigned errors were used to generate distance restraints.[100] Square potential energy wells corresponding to the upper and lower bounds of interproton

distance vectors calculated by the program MARDIGRAS[101] were utilized. Additional distance restraints were generated using Watson-Crick hydrogen bonding distances for base pairs, but excluding the $X^5$:$C^{20}$ base pair. Ten anti-distance restraints,[102] characterized by square potential energy wells maintaining inter-proton distances from a lower bound of 5 Å to an upper bound of 10 Å, were used.

*Restrained Molecular Dynamics Calculations*

Distance restraints, obtained as described above, were used in restrained molecular dynamics (rMD) calculations. Deoxyribose pseudorotation and phosphodiester backbone restraints derived from canonical B-DNA values were included in rMD calculations.[103] Pseudorotation restraints were not included for bases $C^4$, $X^5$, $T^6$, $A^{19}$, and $C^{20}$ as NMR data indicated perturbations for deoxyribose protons associated with these bases that may be associated with alternative deoxyribose puckers. Perturbations included changes in chemical shifts and/or peak volume integrations when compared to the unmodified duplex. Additionally, pseudorotation restraints were not included for terminal and penultimate nucleotides ($T^{12}$, $A^{13}$, $C^{14}$, $C^{24}$). The complete list of pseudorotation restraints that were utilized is provided in **Table A1** of the appendix. Backbone torsion angle restraints for nucleotides $T^2$, $G^3$, $G^7$, $T^8$, $T^9$, $T^{10}$, $G^{11}$, $C^{14}$, $A^{15}$, $A^{16}$, $A^{17}$, $C^{18}$, $C^{22}$, and $A^{23}$ were assigned potential energy well windows of ±30˚. Nucleotides $C^4$, $T^6$, $A^{19}$, and $G^{21}$, which neighbored the $X^5$:$C^{20}$ base pair, were assigned potential energy well windows of ±60˚. No backbone torsion angle restraints were used for the terminal and penultimate nucleotides $G^1$, $T^{12}$, $A^{13}$, $C^{24}$, the modified base $X^5$, or its complementary base $C^{20}$. The backbone torsion angle restraints were centered at the B-DNA backbone torsion angles $\alpha$ = -60º, $\beta$ = 180º, $\gamma$ = 60º, $\varepsilon$ = 195º, and $\zeta$ = -105º. These were assigned potential energy wells

as described above. The partial charges and bond lengths for the C8-dG-ABA adduct were calculated using the B3LYP/6-31G* basis set in the program GAUSSIAN,[104] and are provided in **Figure A1** of the Appendix. The structure of the modified duplex used to initiate the rMD calculations was generated in PDB format using the program MOE.[105] At the modified nucleotide $X^5$ the C8-dG-ABA moiety was built manually and added to an unmodified duplex. The C8-dG-ABA moiety was manually positioned to intercalate in the duplex. The face of the ABA moiety containing the ABA H4, H5, H6 protons was positioned to face the minor groove and the bay area of the ABA moiety containing the H2, H1, H11, H10 protons was positioned to face the major groove. At nucleotide $X^5$, the glycosidic bond was placed into the *syn* conformation. The complementary base $C^{20}$ was displaced from the duplex into the major groove. The resulting structure was further modified using the program xLEaP[106] to include values for partial charges and bond lengths of the ABA moiety as obtained from GAUSSIAN. The .top and .inp files were then generated in xLEaP for use in the program AMBER.[107]

The rMD calculations were performed using a simulated annealing protocol in the program AMBER[107] using the parm99 force field.[108] All restraints had applied force constants of 32 kcal $mol^{-1}$ $Å^{-2}$. The generalized Born model was used for solvation.[109] The salt concentration was set at 0.1 M. Initial calculations were performed for 20 ps over 20,000 steps. The system was heated from 0 to 600 K for the first 1,000 steps with a 0.5 ps coupling. This was followed by 1,000 steps at 600 K then 16,000 steps of cooling to 100 K with 4 ps coupling. Cooling to 0 K was completed during the final 2,000 steps with 1 ps coupling. Final calculations were performed over 100,000 steps for 100 ps. The system was heated from 0 to 600 K for the first 5,000 steps with a 0.5 ps coupling. This was followed by 5,000 steps at 600 K then 80,000 steps of cooling to 100 K with 4 ps coupling. Cooling to 0 K was completed during the final 10,000 steps with 1

ps coupling. Structure coordinates were saved after each cycle. NOE generated distances were compared to intensities calculated from emergent structures using complete relaxation matrix analysis[100b] (CORMA). The ten structures with the lowest deviations from experimental distance restraints were used to generate an average refined structure. This structure was then subjected to energy minimization. A second set of calculations was completed in a similar manner without the use of the anti-distance restraints. These calculations were performed to check that the utilization of anti-distance restraints did not alter the course of the rMD calculations. For constant temperature rMD calculations, the $X^5$ base was placed manually into the *anti* conformation about the glycosidic bond using the program MOE.[105] This structure was then energy minimized. All restraints used in the constant temperature rMD calculations remained the same as those used for the simulated annealing rMD calculations, as did the applied force constants of 32 kcal $mol^{-1}$ $Å^{-2}$. The emergent structures were evaluated as to sixth root residuals ($R^X_1$) between the calculated NOE volumes and experimental NOE volumes[100c, 110] in the same manner as was calculated for the average refined structure emergent from the simulated annealing rMD calculations, using the program CORMA.

*Data Deposition*

The structure factors and coordinates have been deposited in the Protein Data Bank and Biological Magnetic Resonance Bank (www.rcsb.org). The RCSB ID code for the duplex containing the C8-dG-ABA adduct is rcsb104415. The PDB ID code is 2n4m. The BMRB Id code is 25672.

## Results

### *Characterization*

Mass spectrometry of the single strand oligodeoxynucleotides confirmed the expected masses. The expected masses were 3941.5 g $mol^{-1}$ for the modified strand, 3698.5 g $mol^{-1}$ for the unmodified primary strand, and 3592.4 g $mol^{-1}$ for the complementary strand, and the masses were recorded at 3942.1 g $mol^{-1}$, 3697.4 g $mol^{-1}$, and 3591.2 g $mol^{-1}$, respectively.

### *DNA Melting Temperature*

The melting temperature ($T_m$) for the C8-dG-ABA duplex was 44 °C. Compared to the measured melting point of 55 °C for the unmodified duplex, this represented a decrease in $T_m$ of 11 °C.

### *NMR Spectroscopy*

DNA non-Exchangeable Protons. The base aromatic and deoxyribose anomeric protons were assigned using established procedures (**Figure 6**).[111] For the modified strand (**Figure 6A)**, the modified base $X^5$ is alkylated at the C8 position of the guanine imidazole ring, and hence, there is no $X^5$ H8 proton. Consequently, the sequential pattern of NOE connectivity between the deoxyribose H1' protons and base aromatic protons was terminated. Additionally, the NOE between $C^4$ H6 and $C^4$ H1' was weak. The sequential pattern of NOE connectivity was re-initiated at the NOE between $X^5$ H1' and $T^6$ H6, which was also weak. The corresponding sequential pattern of NOE connectivity in the complementary strand (**Figure 6B**) was also interrupted. There was no NOE observed between $A^{19}$ H8 and $C^{20}$ H1'. The $C^{18}$ H1' to $A^{19}$ H8

NOE was weak. This indicated a perturbation in the complementary strand opposite the modified base $X^5$. At 25 °C the $T^6$ H6 resonance was overlapped with the $G^{21}$ H8 resonance but at 15 ºC these two resonances could be resolved (**Figure A2** in the Appendix). This allowed for the unequivocal identification of the NOEs associated with these two protons, including the NOE between $X^5$ H1' and $T^6$ H6. The adenine H2 protons were assigned based upon NOEs to the thymine imino protons of the respective A:T base pairs and cross-peaks to their respective H1′ protons. The remaining deoxyribose protons were assigned from a combination of NOESY and COSY data. The assignments of the non-exchangeable DNA protons for the C8-dG-ABA modified duplex and the corresponding unmodified duplex are summarized in **Tables A2 and A3** of the appendix.

DNA Exchangeable Protons. The imino and amino proton resonance regions of the NOESY spectrum are shown in **Figure 7**. The imino proton resonances were identified for all base pairs except for the terminal base pairs $G^1$:$C^{24}$ and $T^{12}$:$A^{13}$, and modified base pair $X^5$:$C^{20}$. The inability to assign the imino proton resonances of the terminal base pairs was attributed to their rapid exchange with water; two broad and unassignable resonances were observed in the imino proton region of the spectrum between 12.6 and 12.9 ppm, which may correspond to the imino protons of the terminal $G^1$:$C^{24}$ and $T^{12}$:$A^{13}$ base pairs. The anticipated NOEs between guanine N1H imino and cytosine $N^4$H amino protons were observed for each of the remaining C:G base pairs. The anticipated NOEs between thymine N3H imino and adenine H2 protons were observed for each of the A:T base pairs, with the exception of the terminal base pair $T^{12}$:$A^{13}$. The $T^6$ N3H imino resonance was broad, and its NOE to $A^{19}$ H2 was weak as compared to the other NOEs between thymine N3H imino protons and adenine H2 protons. The chemical shift assignments for the exchangeable DNA protons of the C8-dG-ABA modified duplex are

provided in **Table A4** of the Appendix with corresponding chemical shifts for the unmodified duplex shown in **Table A5** of the Appendix.

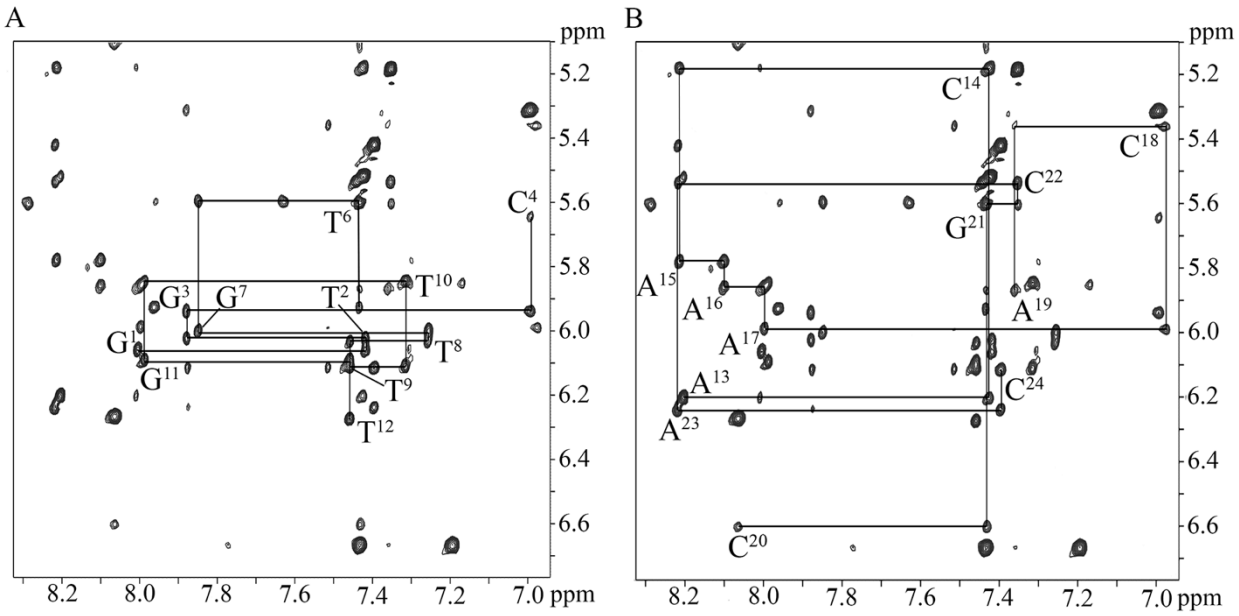



**Figure 6.** Expanded plot of a NOESY spectrum of the C8-dG-ABA modified duplex, showing the sequential NOE connectivity between aromatic H8/H6 protons and deoxyribose H1' protons. **A**. The modified strand, showing bases $G^{1}$ through $T^{12}$. The NOE connectivity is broken at the $C^{4}$ H1′-H6 cross peak and reinitiates at the $X^{5}$ H1' - $T^{6}$ H6 cross peak since the modified $X^{5}$ base does not have an H8 proton. **B**. The complementary strand connectivity showing assignments for bases $A^{13}$ through $C^{24}$. The connectivity is broken at the $A^{19}$ H1′→$A^{19}$ H8 NOE and reinitiates at the $C^{20}$ H1′→$C^{20}$ H6 NOE; no NOE cross peak is observed between $A^{19}$ H1′ and $C^{20}$ H6. The 900 MHz spectrum was acquired at 25 ºC using a 250 ms mixing time.

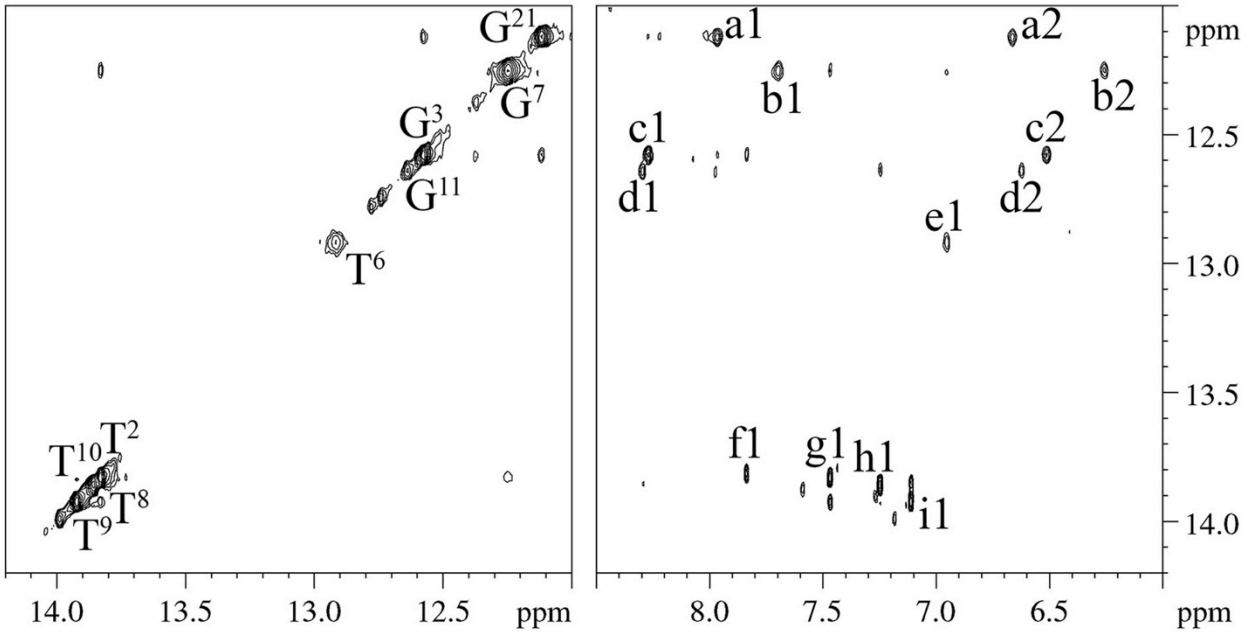



**Figure 7**. Expanded plots of the imino and amino regions of a NOESY spectrum of the C8-dG-ABA modified duplex. **Left Panel**: The imino proton region of the spectrum, showing NOEs between guanine N1H and thymine N3H imino protons. The broadening of the $T^6$ N3H imino proton resonance is evident, **Right Panel**: NOEs between the guanine N1H and cytosine $N^4$H amino protons and between thymine N3H and adenine H2 protons. The NOEs are assigned as follows: a1, $G^{21}$ N1H → $C^4$ $N^4H_b$; a2, $G^{21}$ N1H → $C^4$ $N^4H_a$; b1, $G^7$ N1H → $C^{18}$ $N^4H_b$; b2, $G^7$ N1H → $C^{18}$ $N^4H_a$; c1, $G^3$ N1H → $C^{22}$ $N^4H_b$; c2, $G^3$ N1H →$C^{22}$ $N^4H_a$; d1, $G^{11}$ N1H → $C^{14}$ $N^4H_b$; d2, $G^{11}$ N1H → $C^{14}$ $N^4H_a$; e1, $T^6$ N3H → $A^{19}$ H2; f1, $T^2$ N3H → $A^{23}$ H2; g1, $T^8$ N3H → $A^{17}$ H2; h1, $T^{10}$ N3H → $A^{15}$ H2; i1, $T^9$ N3H → $A^{16}$ H2. Note that cross-peak e1, arising from base pair $T^6$:$A^{19}$, the 3'-neighbor with respect to the modified base pair $X^5$:$C^{20}$, is weaker than the cross peaks f1, g1, h1, and i1. The 900 MHz spectrum was acquired at 15 ºC using a 250 ms mixing time.

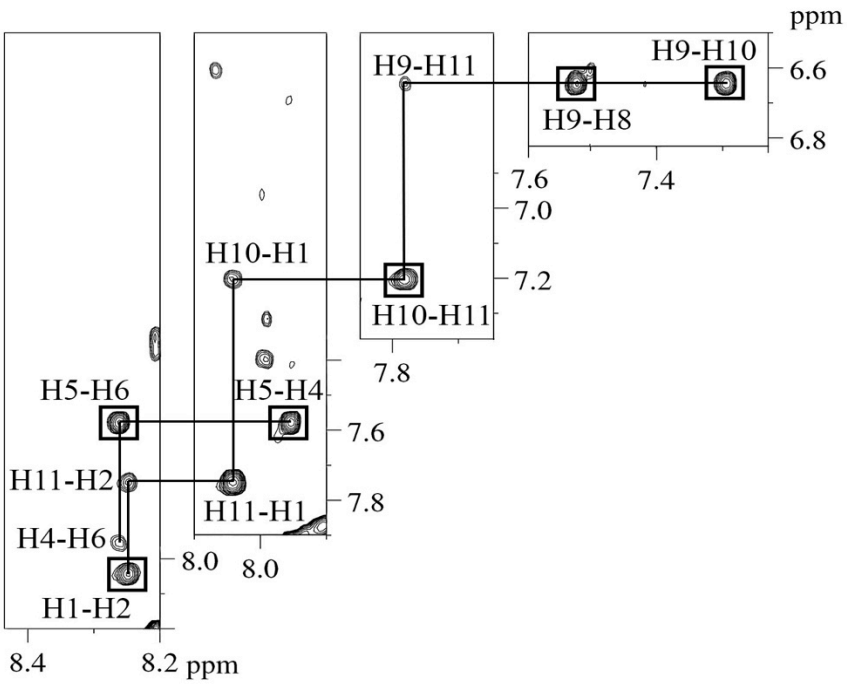



**Figure 8**. Expanded plots of regions of the NOESY spectrum of the C8-dG-ABA modified duplex, showing NOEs associated with the ABA cross-peaks (Chart 1). The boxed cross-peaks represent cross-peaks that are simultaneously observed in COSY spectra. In all cases, the cross-peaks are labeled to indicate proton resonances associated with the vertical (t1) axis first and proton resonances associated with the horizontal (t2) axis second. The 900 MHz spectrum was acquired at 15 º C using a 250 ms mixing time.

Adduct Protons. The ABA moiety of the C8-dG-ABA adduct exhibits three distinct 1H spin systems, consisting of the H1 and H2 protons, the H4, H5, and H6 protons, and the H8, H9, H10, and H11 protons (**Chart 5**). The COSY spectrum was consistent with the presence of these three spin systems (**Figure 8**). Six COSY cross peaks were identified. The chemical shifts at 15 ºC of the ABA H1 and H2 proton resonances were identified at 8.04 ppm and 8.25 ppm on the basis of a strong COSY cross-peak. The ABA H4, H5, and H6 resonances were identified in the COSY spectrum at 7.95 ppm, 7.61 ppm, and 8.26 ppm, respectively. Strong scalar couplings

were observed between the adjacent protons H4 and H5, and H5 and H6; a weaker scalar coupling was observed between H4 and H6. A comparison of NOESY and COSY data allowed for identification of the ABA H8, H9, H10, and H11 protons. In the NOESY data an NOE was observed between the H1 proton and the H11 proton, allowing H11 to be assigned at 7.78 ppm. With this assignment in hand, H8, H9, H10 could be assigned from the COSY spectrum, at 7.42 ppm, 6.64 ppm, 7.20 ppm, respectively. The assignments of the ABA H4 and H6 resonances were corroborated by NOEs between H4 and H6 and deoxyribose protons of the modified base as shown in **Figure 9**. The assignments of the C8-dG-ABA adduct protons are provided in **Table A6** of the Appendix. The ABA H4 proton displayed multiple NOEs with the deoxyribose protons of nucleotide $X^5$, indicating that it was closer in space to the deoxyribose as compared to the ABA H6 proton, which displayed only a single NOE to the complementary strand. A total of 25 NOEs between adduct protons and deoxyribose and base protons for both the modified and complementary strands were identified (**Figure 9)**. Of these, ABA protons H2, H4, and H5 showed nine interactions with the modified strand and ABA protons H6, H8, H9, and H10 showed sixteen interactions with the complementary strand. No NOEs to either the modified or complementary strand of the duplex were observed for either ABA protons H1 or H11. A resonance was observed at 9.24 ppm in the $H_2O$ spectrum of the modified duplex. This may represent the ABA 3-NH proton present at the point of attachment between the adduct and the C8 of the modified base. However, the assignment remained equivocal because no NOEs were observed for this proton. This was attributed to the rapid exchange of the ABA 3-NH proton with solvent.

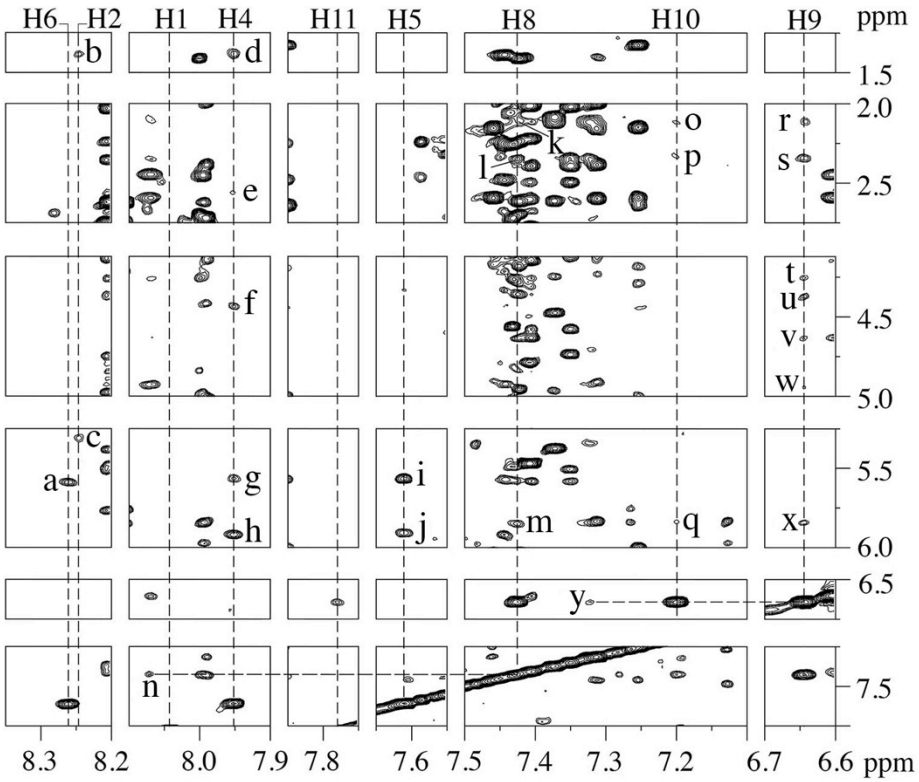



**Figure 9**. Expanded plots of a NOESY spectrum of the C8-dG-ABA modified duplex showing NOEs between adduct protons (**Chart 1**) and base and deoxyribose protons. The ABA H2, H4, and H5 protons presented NOEs to protons of the modified stand of the duplex. The ABA H8, H9, and H10 protons presented NOEs to the complementary strand of the duplex. The ABA H1 and H11 protons displayed no NOEs with either strand of the duplex. The cross-peaks are assigned as follows: ABA H6: a → $G^{21}$ H1′; b, ABA H2 → $T^{6}$ $CH_3$; c ABA H2 → $C^{4}$ H5; ABA H4: d → $T^{6}$ $CH_3$, e → $X^{5}$ H2′′, f → $X^{5}$ H4′, g → $T^{6}$ H1′, h → $X^{5}$ H1′; ABA H5: i → $T^{6}$ H1′, j → $X^{5}$ H1′; ABA H8: k → $A^{19}$ H2′′, l → $A^{19}$ H2′, m → $A^{19}$ H1′, n → $C^{20}$ H6; ABA H10: o → $A^{19}$ H2′′, p → $A^{19}$ H2′, q → $A^{19}$ H1′; ABA H9: r → $A^{19}$ H2′′, s → $A^{19}$ H2′, t → $C^{20}$ H5′, u → $C^{20}$ H5′′, v → $C^{20}$ H4′, w → $A^{19}$ H3′, x → $A^{19}$ H1′, y → $A^{19}$ H8. The 900 MHz spectrum was acquired at 15 º C using a 250 ms mixing time.

Unmodified Duplex. The NMR data from the unmodified duplex was unremarkable, and the spectra of the unmodified duplex were fully assignable. The NOESY spectrum displayed two separate sets of continuous connectivity in the H6/H8 to H1' section of the spectrum which could be assigned to the primary and complementary strands of the DNA duplex (**Figure A3** in the Appendix). The complete assignment of the spectrum did not reveal significant chemical shift perturbations, indicating a duplex conformation in the B-DNA[112] family.

*Chemical Shift Perturbations*

Chemical shift perturbations observed in the NMR spectra for the adducted nucleotide and surrounding nucleotides of the modified duplex as compared to the unmodified spectra are shown in **Figure 10.** Significant chemical shift perturbations were observed in the region of the adducted nucleotide, particularly for nucleotides $C^4$, $X^5$, and $T^6$ and $A^{19}$ and $C^{20}$. The $C^4$ nucleotide exhibited upfield changes in chemical shifts for the H6, H1′, H2′, H2′′, and H3′ protons of 0.33, 0.09, 1.23, 0.36, and 0.60 ppm, respectively. The $X^5$ H8 proton is replaced by the adduct but the H1′, H2′, H2′′, and H3′ deoxyribose protons exhibited changes in shifts of 0.06 ppm upfield, 0.87 ppm downfield, 0.12 ppm upfield, and 0.13 ppm downfield, respectively. The $T^6$ protons displayed downfield changes in chemical shifts of 0.27, 0.10, and 0.03 ppm for the H6, H2′, and H3′ protons and upfield shift changes of 0.26 and 0.05 ppm for the H1′, and H2′′ protons, respectively. For bases in the complementary strand, the $A^{19}$ nucleotide protons displayed upfield changes in chemical shift of 0.84 ppm for H8, 0.30 ppm for H1′, 0.52 ppm for H2′, 0.50 ppm for H2′′, and 0.08 ppm for H3′. The $C^{20}$ nucleotide protons exhibited downfield changes in chemical shift of 0.89, 1.05, 0.50, 0.28, and 0.28 ppm, for protons H6, H1′, H2′, H2′′, and H3′, respectively. The $G^{21}$ nucleotide protons displayed upfield changes in chemical

shift for the H8, H1′, H2′, H2′′, and H3′ protons of 0.49, 0.28, 0.23, 0.22, and 0.40 ppm respectively. Nucleotides distal to the adducted nucleotide displayed minimal chemical shift perturbations, predominantly less than 0.1 ppm, suggesting that these regions maintained a conformation similar to the unmodified duplex. Exceptions to this were the $C^{18}$ H6, H1′, H2′, and H2′′ protons, which exhibited upfield changes in shifts of 0.19, 0.16, 0.31, and 0.30 ppm, respectively, and $A^{17}$ H2′′, which exhibited an upfield change in chemical shift of 0.12 ppm.

*Structural Refinement*

A total of 320 NOE derived distance restraints were obtained from analyses of NOESY spectra for non-exchangeable protons collected at 15 ºC or 25 ºC (**Table A7** in the Appendix*)*. These included 125 internucleotide restraints, 171 intranucleotide restraints, and 24 adduct restraints. Additionally, proximal to the lesion site, 10 anti-distance restraints (**Table A8** in the Appendix), placing a lower bound of 5 Å and an upper bound of 10 Å upon interproton distances, were employed. These were utilized in instances in which specific NOEs were not observed in the spectra. In such instances it was concluded that the failure to observe an NOE indicated that the protons in question were a minimum of 5 Å apart. Additional restraints included 90 backbone torsion angles, 42 hydrogen bond restraints, and 75 deoxyribose pseudorotation restraints, for a total of 537 restraints (**Table 2**). The latter were included as empirical restraints based on canonical B-DNA,[103] which was consistent with NMR data.

A series of restrained molecular dynamics (rMD) calculations employing a simulated annealing protocol yielded ten emergent structures, from which an average structure was calculated. An overlay of the ten emergent structures and the final average and minimized structure indicated excellent convergence, and is shown in **Figure 11**. The ten emergent

structures had a maximum RMS pairwise difference of 0.59 Å and the average structure had a maximum RMS pairwise difference of 0.40 Å, as compared to the ten individual structures. This indicated that there were sufficient experimental restraints to allow the rMD calculations to define a well-converged structure. The refinement statistics are shown in **Table 2**.

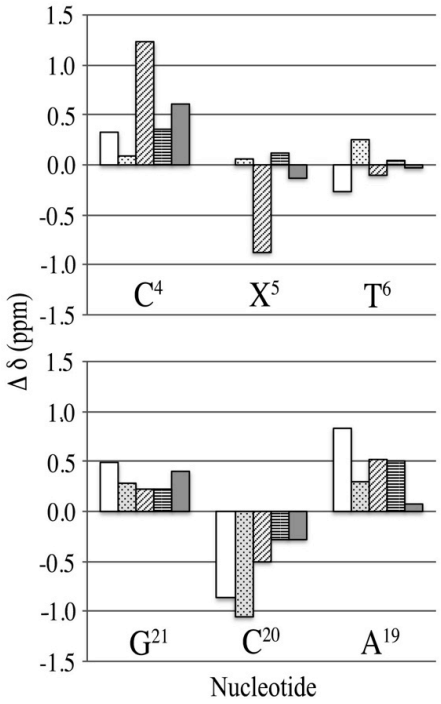



**Figure 10**. Changes in chemical shifts for the protons of the ABA modified and surrounding nucleotides of the C8-dG-ABA modified duplex as compared to the unmodified duplex. **Top Panel**: Nucleotides of the modified strand, showing nucleotides $C^4$ through $T^6$. **Bottom Panel**: Nucleotides of the complementary strand showing nucleotides $A^{19}$ through $G^{21}$. The base aromatic H6 or H8 protons are shown in white. The deoxyribose H1′ protons are shown in dots, the H2′ protons are shown in diagonal lines, the H2′′ protons are shown in horizontal lines, and the H3′ protons are shown in gray. The Δδ (ppm) values are calculated as $\delta_{\text{unmodified duplex}} - \delta_{\text{modified duplex}}$. Positive values of Δδ represent upfield shifts and negative values of Δδ represent downfield shifts, with respect to the unmodified duplex.

**Table 2.** NMR restraints used for the C8-dG-ABA structure calculations and refinement statistics.

| | |
|---|---|
| **NOE restraints** | |
| Internucleotide | 125 |
| Intranucleotide | 171 |
| C8-dG-ABA Adduct | 24 |
| Anti-distance restraints[a] | 10 |
| Total | 330 |
| Backbone torsion angle restraints | 90 |
| Hydrogen bonding restraints | 42 |
| Deoxyribose pseudorotaton restraints | 75 |
| Total number of restraints | 537 |
| **Refinement statistics** | |
| Number of distance restraint violations | 64 |
| Number of torsion restraint violations | 10 |
| Total distance penalty/maximum penalty (kcal/mol) | 5.0/0.296 |
| Total torsion penalty/maximum penalty (kcal/mol) | 0.39/0.118 |
| r.m.s. distances (Å) | 0.013 |
| r.m.s. angles (º) | 2.3 |
| Distance restraint force field (kcal/mol/Å$^2$) | 32 |
| Torsion restraint force field (kcal/mol/deg$^2$) | 32 |

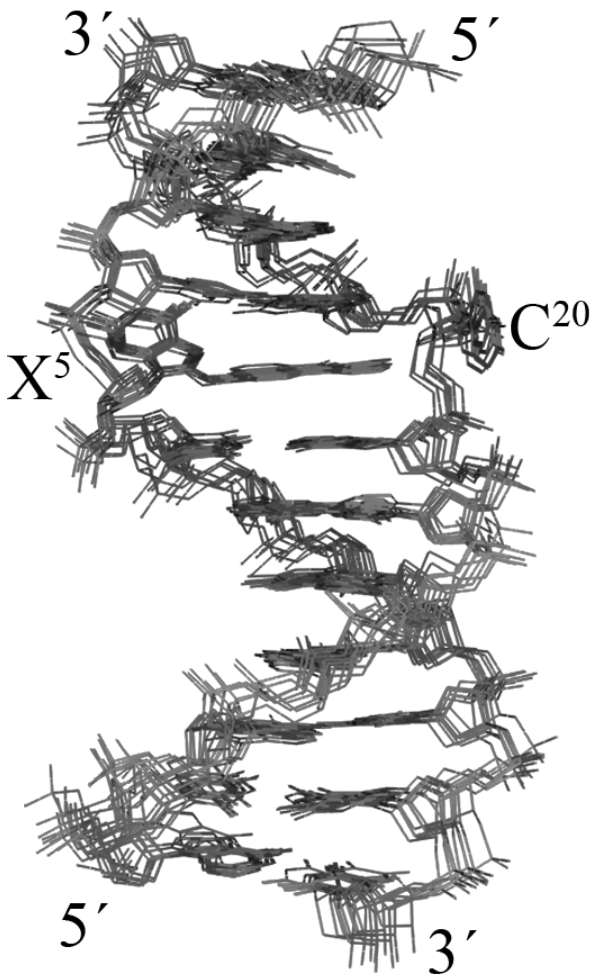



**Figure 11**. Overlay of ten lowest energy violation structures resulting from rMD calculations of the C8-dG-ABA modified duplex carried out using a simulated annealing protocol and using NOE generated distance restraints. The view is looking into the major groove.

The average structure emergent from the rMD calculations was evaluated as to its accuracy by complete relaxation matrix analysis carried out using the program CORMA[100b] (**Figure 12 & Table 3**). Most individual internucleotide and intranucleotide sixth root residual $R^X_1$ values were less than 0.1, with the overall internucleotide and intranucleotide sixth root residual $R^X_1$ values being $8.2 \times 10^{-2}$ and $9.7 \times 10^{-2}$ respectively. The overall sixth root residual

for the modified duplex $R^X_1$ value was 0.088. This indicated that the average refined structure was in agreement with NOE data.

**Table 3.** RMS differences and sixth root residual ($R^X_1$) values calculated for the average structure.

| Average structure (calculated from ten structures emergent from the simulated annealing rMD calculations) | | | |
|---|---|---|---|
| RMS pairwise difference between structures | | | 0.592 |
| RMS difference from average structure | | | 0.397 |
| | Complete relaxation matrix analysis for the calculated average structure, using the program CORMA[a] | | |
| | Intranucleotide | Internucleotide | Total |
| $R^X_1$[b] | 0.081 | 0.092 | 0.085 |
| Average error[c] | | | 0.019 |

[a]Mixing time was 250 ms

[b]$R^X_1$ is the sixth root R factor: $\sum[((I_o)_i^{1/6})-((I_c)_i^{1/6})/\sum((I_o)_i^{1/6})]$

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

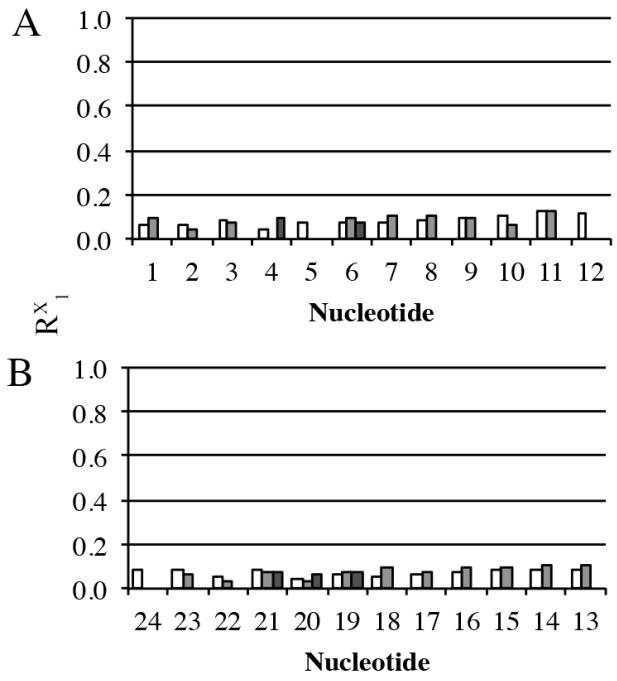



**Figure 12**. Calculation of sixth root residual values ($R^X_1$) between theoretical NOEs predicted by complete relaxation matrix calculations and experimental NOEs for the averaged refined structure of the C8-dG-ABA duplex emergent from the rMD calculations, using the program CORMA. **A.** The intranucleotide, internucleotide and nucleotide to lesion $R^X_1$ values for individual nucleotides in the modified strand. **B.** The intranucleotide, internucleotide and nucleotide to lesion $R^X_1$ values for individual nucleotides in the complementary strand. The intranucleotide residuals are shaded in white. At the modified nucleotide $X^5$, the intranucleotide residuals calculated for the C8-dG-ABA adduct are included with those of the $X^5$ nucleotide. The DNA internucleotide residuals are shaded in gray. With the exception of the modified nucleotide $X^5$, they indicate NOEs to the respective 3′-neighbor nucleotides. The internucleotide residuals between the C8-dG-ABA adduct and other nucleotides are shaded in dark gray. These involve nucleotides $C^4$, $T^6$, $A^{19}$, $C^{20}$, and $G^{21}$. In all cases, the sixth root residual factor was calculated as $R_1^x = \sum[((I_o)_i^{1/6})-((I_c)_i^{1/6})/\sum((I_o)_i^{1/6})]$, in which $I_c$ are NOE intensities calculated by complete relaxation matrix analysis of the refined structure and $I_o$ are experimental NOE intensities.

A parallel set of rMD calculations using a simulated annealing protocol was performed, in which the 10 anti-distance restraints proximate to the lesion site were not employed. The refined structures emergent from this set of calculations were similar to those emergent from the calculations that employed the anti-distance restraints, indicating that the utilization of anti-distance restraints was not significantly biasing the rMD calculations. The set of ten structures emergent from the rMD calculations performed without the anti-distance restraints had a maximum rms pairwise difference of 0.58 Å (**Table A9** in the Appendix). The structural features did not differ significantly from those of the structure determined using the anti-distance restraints, as shown in **Figure A4** in the Appendix. The RMS pairwise difference between the structures obtained with and without the use of anti-distance restraints was 0.48 Å. One notable difference was a reduced degree of displacement for the $C^{20}$ base. The C2 and N3 edge of $C^{20}$ was oriented 1.8 Å closer to the ABA adduct while the C5 and C6 edge of $C^{20}$ was oriented in a similar position compared to the structure determined with the use of anti-distance restraints. In the absence of the anti-distance restraints, the $C^{20}$ deoxyribose was also shifted 0.6 Å closer to the ABA adduct.

Another parallel set of rMD calculations was carried out at constant temperature, in which the $X^5$ base was initially placed into the *anti* conformation about the glycosidic bond. The structures that emerged from these calculations displayed greater disagreement with the NOE data, as determined by complete relaxation matrix analysis.[100b] The average intranucleotide and internucleotide $R^X_1$ values for the $C^4$, $X^5$, $T^6$, $A^{19}$, $C^{20}$, and $G^{21}$ nucleotides were calculated for both the structure with the adducted guanine in the *syn* conformation and the structure with the adducted guanine in the *anti* conformation. The structure with the adducted guanine in the *syn*

conformation about the glycosidic bond displayed an average $R^X_1$ value of $6.8 \times 10^{-2}$ for these nucleotides. The structure with the adducted guanine in the *anti* conformation about the glycosidic bond displayed an average $R^X_1$ value of $9.0 \times 10^{-2}$ for these nucleotides. Most distances between protons at the lesion site were within the range of the NOE derived restraint windows. Of the 41 restraints used, nine were outside the window with the largest violation being 0.07 Å. These values are shown in **Table A10** in the Appendix. The violations were determined to be acceptable as each of the nine values falling outside of the window possessed deviations that were small and were associated with restraints possessing higher error margins. These restraints were connected to NOEs that were weak, broadened, overlapped, or a combination of these factors.

***C**onformation of the C8-dG-ABA Adduct*

**Figures 13 and 14** show views of the C8-dG-ABA adduct, based upon the lowest energy violation structure that emerged from the rMD calculations. This structure is in good agreement with the average structure. The modified guanine was shifted into the *syn* conformation about the glycosidic bond. This placed the base in proximity to the $C^4$ deoxyribose protons and altered the deoxyribose pucker to the 3'-endo conformation for the $X^5$ deoxyribose. The ABA moiety intercalated between the $A^{19}$ and $G^{21}$ bases of the complementary strand. It was oriented such that the bay region of the ABA moiety, including protons H1, H2, H10, and H11, faced toward the major groove of the DNA. The ABA H9 proton faced toward the phosphodiester backbone of the complementary strand, between bases $A^{19}$ and $G^{21}$. The ABA H4, H5, H6, and H8 protons faced toward the minor groove of the DNA, as did the keto oxygen of the ABA moiety. The intercalation of the ABA moiety between the $A^{19}$ and $G^{21}$ bases of the complementary strand

forced the $C^{20}$ base to be displaced into the major groove (**Figure 13 A**). The alignment of the planar ABA arylamine between the $A^{19}$ and $G^{21}$ bases allowed for possible π-stacking interactions (**Figure 14)**. Bases $G^{1}$ and $T^{2}$ and their complement on the 5′-side and from $G^{7}$ through $T^{12}$ on the 3′-side and their complement retained a structure similar to canonical B-DNA.

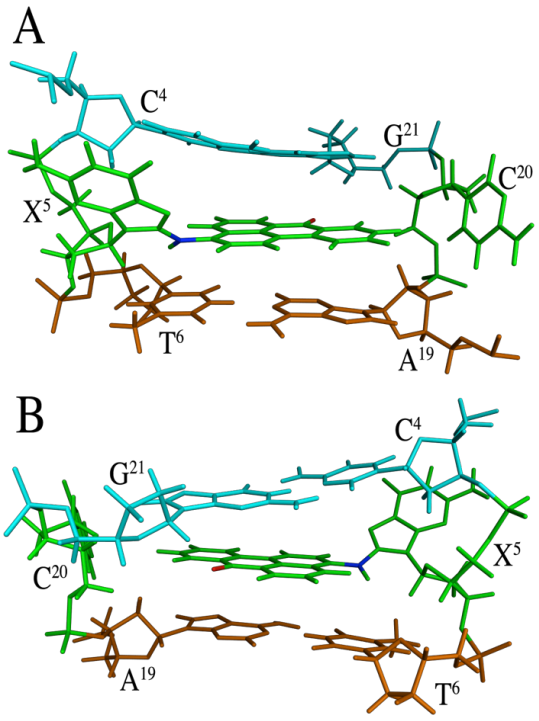



**Figure 13**. Conformation of the C8-dG-ABA adduct as seen in the lowest violation structure emergent from rMD calculations. (**A**) Base pairs $C^{4}$:$G^{21}$, $X^{5}$:$C^{20}$, and $T^{6}$:$A^{19}$ as seen from the major groove. (**B**) Base pairs $C^{4}$:$G^{21}$, $X^{5}$:$C^{20}$, and $T^{6}$:$A^{19}$ as seen from the minor groove. The $X^{5}$ base is rotated into the *syn* conformation at the glycosidic torsion angle; the C8-dG-ABA moiety is intercalated between base pairs $C^{4}$:$G^{21}$ and $T^{6}$:$A^{19}$; the complementary base $C^{20}$ is displaced into the major groove. The $C^{4}$ and $G^{21}$ base carbon atoms are in cyan. The $X^{5}$ and $C^{20}$ base carbon atoms are in green. The $T^{6}$ and $A^{19}$ base carbon atoms are in orange. The adduct oxygen atom is shown in red and the nitrogen atom at the point of adduct attachment is shown in blue.

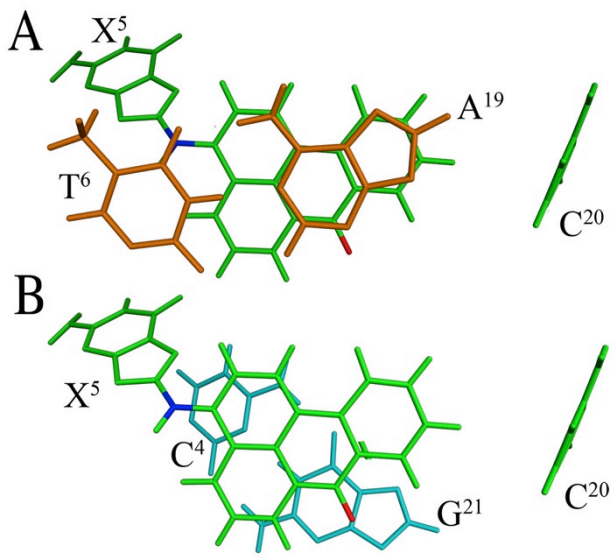



**Figure 14**. Base stacking interactions of the C8-dG-ABA adduct as seen in the lowest violation structure emergent from rMD calculations. A. View looking through the duplex from the 3′-side of the ABA adduct showing carbon atoms of bases $X^5$ and $C^{20}$ in green and 3′-neighbor bases $A^{19}$ and $T^6$ in orange. B. View looking through the duplex from the 3′-side of the adducted nucleotide showing carbon atoms of bases of $X^5$ and $C^{20}$ in green and 5'-neighbor bases $G^{21}$ and $C^4$ in cyan. The oxygen atom of the adduct is shown in red and the nitrogen atom at the point of adduct attachment is shown in blue.

## Discussion

Of the three DNA adducts known to be formed following exposures to 3-NBA, the C8-dG-ABA adduct is believed to be the most significant contributor to mutagenic outcome.[76d] Thus, an understanding of the conformation of the C8-dG-ABA adduct in DNA is important, especially in light of human exposures to 3-NBA[82] and its potential carcinogenicity in humans.[83]

*Conformation of the C8-dG-ABA Adduct*

The refined structure of the C8-dG-ABA adduct in this duplex reveals a base-displaced intercalated structure in which the modified base shifts into the *syn* conformation about the glycosidic bond. The usual method for evaluating the conformation of the glycosidic bond involves monitoring the intensity of the NOE between the guanine H8 proton and the guanine H1' proton; this NOE is strong when the purine base is in the *syn* conformation and weak when the base is in the *anti* conformation. However, this method is not applicable to C8-dG adducts, since these adducts do not possess a guanine H8 proton. A large body of work with bulky C8-adduct base-displaced intercalated structures in which the adducted base is shifted into the *syn* conformation[90, 93-94] has revealed that the chemical shifts of the deoxyribose H2' and H2" protons are also diagnostic markers of *syn* vs. *anti* conformation with respect to the glycosidic bond. In the present instance, the significant changes in chemical shifts of 1.23 ppm upfield and 0.87 ppm downfield for the $C^4$ H2' and $X^5$ H2' protons respectively, are consistent with the conclusion that the C8-dG-ABA adduct orients in the *syn* conformation about the glycosidic bond. This conclusion is also consistent with the predictions of complete relaxation matrix (CORMA) calculations[100b] performed upon structures emergent from the rMD calculations. These revealed that structures emergent from constant temperature rMD calculations, for which the $X^5$ base was placed in the *anti* conformation about the glycosidic bond, displayed greater disagreement with the NOE data. In that instance, for nucleotides $C^4$, $X^5$, and $T^6$ and $A^{19}$, $C^{20}$, and $G^{21}$, the average sixth root residual $R^X_1$ with respect to the experimental NOE data obtained from the *anti* conformation is 9.0 while that obtained from the *syn* conformation is 6.8. The orientation of the intercalated arylamine ring of the ABA moiety is established by the sixteen NOEs between ABA H6, H8, H9, and H10 protons and those of the complementary strand, including protons of

nucleotides $A^{19}$, $C^{20}$, and $G^{21}$, indicating that this face of the adduct is positioned near the $C^{20}$ deoxyribose and close to the area that would be occupied by the $C^{20}$ base in a Watson-Crick base pair. This positioning of the aryl moiety allows for π stacking between the aryl ring and the $A^{19}$ and $G^{21}$ bases. Indeed, significant upfield changes in chemical shifts for the $A^{19}$ H8 and $G^{21}$ H8 proton resonances of 0.84, and 0.49 ppm, respectively, are consistent with such stacking interactions. A comparison of **Figures 14A and 14B** suggests a greater potential for stacking between the aryl ring with the $A^{19}$ base. This may be reflected in the 0.35 ppm greater chemical shift perturbation of the $A^{19}$ H8 proton resonance as compared to the $G^{21}$ H8 proton resonance. The breaks in the sequential NOEs between base aromatic and deoxyribose H1' protons identified in the complementary strand and weak the NOEs between $C^{20}$ H1' and $C^{20}$ H6 and $C^{20}$ H1' and $G^{21}$ H8 support the conclusion that the $C^{20}$ base is displaced from the duplex into the major groove. In addition, the large downfield chemical shifts observed for the $C^{20}$ H8 and H1' proton resonances are also consistent with this conclusion. NOEs observed between the $C^{20}$ H5 proton and the aromatic H9 and H10 protons of the adduct indicate the $C^{20}$ base is displaced into the major groove and not the minor groove of the duplex.

*Thermodynamic Effects of the C8-dG-ABA Adduct*

In this sequence, the C8-dG-ABA adduct decreases the melting temperature ($T_m$) by 11 ºC. The decreased thermal stability of the duplex is reflected in the NMR spectra. For example, in **Figure 7,** broadening of the $T^6$ N3H imino resonance and in the broadening of the NOE cross-peak between the $T^6$ imino proton and $A^{19}$ H2 proton may be attributed to an increased rate of exchange between the $T^6$ N3H imino proton and water, likely the consequence of thermodynamic dynamic destabilization of the duplex. In this regard, the displacement of the $C^{20}$

base into the major groove results in the loss of one base pair and the π-stacking interactions for this base. Furthermore, transition of the $X^5$ base into the *syn* conformation about the glycosidic bond results in further loss of π-stacking interactions. Thus, despite the fact that the base-displaced intercalated conformation of the ABA moiety facilitates π-stacking between the intercalated aryl ring and neighboring bases, these new stacking interactions are evidently not sufficient to compensate for the loss in stability contributed by other sources. These may include loss of π-stacking interactions from the modified guanine base and its complementary cytosine base, as well as the loss of the hydrogen bonds contributed by this base pair. Additionally, some degree of helix unwinding may be present. As reported by Mu *et al.*[113], unwinding of the helix leads to a decrease in stability and potential alterations in other π-stacking interactions of the duplex.

*Comparisons to Other C8-dG Aryl Amine Adducts*

Other bulky C8-dG aryl amine adducts have been examined as to conformation when placed between two pyrimidine bases. The C8-dG-IQ,[92] C8-dG-AAF,[91] C8-dG-AP,[90] C8-dG-IQ,[94] and C8-dG-PhIP[93] adducts also exhibit base-displaced intercalated structures in the 5'-d(C<u>X</u>C)-3':5'-d(GCG)-3' sequence. The base-displaced intercalated conformation of the C8-dG-ABA adduct in the 5'-d(C<u>X</u>T)-3':5'-d(ACG)-3' sequence exhibits similarities to these other C8-dG arylamine adducts. In general, these bulky C8-dG adducts are destabilizing when placed in sequence between pyrimidine bases, typically exhibiting decreases in $T_m$ of between 10 ºC to 14 ºC.[114] However, the C8-dG-IQ adduct produced just a 4 ºC decrease in duplex $T_m$[114b] and the C8-dG-AP adduct also exhibited a decreased $T_m$ of 5.9 ºC, though this was in the 5′-T<u>X</u>A-3′ sequence.[115]

It has been proposed that the propensity for C8-dG aryl amine lesions to assume base-displaced intercalated conformations is, in part, related to the steric potential of the hydrophobic aryl amine moieties to intercalate into the DNA duplex.[92, 116] Shapiro *et al.*[116b] recognized the potential role of hydrophobic surface area and suggested that C8-dG adducts having ring structures possessing greater surface area provide a greater hydrophobic area favoring base-displaced intercalation. The C8-dG-ABA adduct exhibits a surface area similar to the C8-dG AAF and C8-dG-PhIP adducts, calculated using the program Vega ZZ.[117] In comparison, the C8-dG-AP adduct has a somewhat smaller surface area of about 91% that of the larger adducts. The C8-dG-AF and C8-dG-IQ adducts possess the smallest surface area, about 84% that of the large adducts. In the 5'-d(CXC)-3':5'-d(GCG)-3' sequence the C8-dG-ABA adduct forms only the base-displaced intercalated structure as does the C8-dG-AAF[91] adduct. However, the similar sized C8-dG-PhIP adduct exhibits a minor external conformation.[93] The intermediate surface area C8-dG-AP adduct[90] and the small surface area C8-dG-IQ adduct[94] display only one conformation but the small surface area C8-dG aminofluorene (AF) lesion equilibrates between the base-displaced intercalated conformation characterized by the *syn* alignment and an external conformation characterized by the *anti* alignment of the glycosidic torsion angle.[92, 116a]

Consequently, it is now recognized that factors in addition to hydrophobic effects related to surface area must contribute to the formation of base-displaced intercalated structures. DNA sequence is important. Wang *et al*.[94, 118] examined the conformation of the C8-dG-IQ adduct placed at differing positions in the *Nar*I restriction site sequence, a hotspot for frameshift mutations in bacteria. They documented the role of sequence in determining adduct conformation; the conformations of the adduct in the 5'-CXG-3' and 5'-GXC-3' sequences were similar and groove-bound,[118] whereas in the 5'-CXC-3' sequence, the base-displaced intercalated

conformation was favored.[94] These sequence effects are consistent with the contributing role of electronic π-stacking interactions. The phenyl ring of the C8-dG-PhIP adduct favors an orientation out-of-plane with respect to the IP ring resulting in a greater unwinding for the duplex. This increases the distance between the PhIP moiety and the neighboring base pairs and reduces π-stacking efficiency of the PhIP ring.[93] In the 5'-d(CXC)-3':5'-d(GCG)-3' sequence the C8-dG-IQ adduct[94] stacks with both the 3´and 5´neighbors and the adducted guanine remains in a position that allows for additional stacking interactions. Favorable electronic dipole-dipole interactions between the heterocyclic IQ ring and neighboring DNA bases may enhance stacking.

*Comparison to the $N^2$-dG-ABA Adduct*

The conformation of the C8-dG-ABA adduct in this sequence, 5'-d(CXT)-3':5'-d(ACG)-3', differs from the conformation of the $N^2$-dG-ABA adduct, elucidated by Lukin *et al.*[89] in the 5'-d(TXC)-3':5'-d(GCA)-3' sequence. The $N^2$-dG-ABA adduct oriented in the minor groove pointing towards the 5' end of the duplex and the modified guanine remained in the *anti* configuration about the glycosidic bond, enabling the complementary cytosine to remain inserted into the duplex.[89] Furthermore, in contrast to the present results indicating that the C8-dG-ABA adduct lowers the $T_m$ of this duplex, the $N^2$-dG-ABA adduct examined by Lukin *et al*.[89] increased the $T_m$ value. Thus, we conclude that regiochemistry, both with respect to resonance structures of the nitrenium ion, the proximate electrophile with regard to DNA alkylation **(Chart 6)**,[71c, 80, 85] and with regard to different nucleophilic sites on the DNA bases[76d, 86] (**Chart 7**), plays significant roles in determining the conformational and thermodynamic effects induced by DNA adducts arising from human exposures to 3-NBA.

*Structure-Activity Relationships*

Although the C8-dG-ABA and $N^2$-dG-ABA[89] adducts have been examined in differing sequence contexts [5'-d(TXC)-3':5'-d(GCA)-3' vs. 5'-d(CXT)-3':5'-d(ACG)-3'], it seems likely that thermodynamic and conformational differences may modulate the repair efficiencies of C8-dG-ABA vs. $N^2$-dG-ABA adducts. That the C8-dG-ABA adduct is a better substrate for NER, as compared to the $N^2$-dG-ABA adduct,[76d] is consistent with the observation that while the C8-dG-ABA adduct lowers the $T_m$ of this duplex, the $N^2$-dG-ABA adduct examined by Lukin *et al*.[89] increased the $T_m$ value. The thermodynamic differences between the C8-dG-ABA and $N^2$-dG-ABA adducts bear similarities with the corresponding adducts arising from the heterocyclic amine food mutagen 8-[(3-methyl-3H-imidazo[4,5-*f*]quinolin-2-yl)amino]-2'-deoxyguanosine (IQ). In rat tissues the $N^2$-dG-IQ adduct is more persistent than was the C8-dG-IQ adduct.[119] As well, the relative thermal stability of the $N^2$-dG-IQ adduct as compared to the corresponding unmodified duplex was also noted when this adduct favored base-displaced intercalated conformations in both the 5'-CXC-3' sequence,[120] and the 5'-CXG-3' sequence,[121] the C8-dG-IQ adduct reduced the thermal stability of the DNA.[114b, 122] The minimal effects of the $N^2$-dG-IQ adducts upon the $T_m$ values of the duplexes may explain the persistence of this adduct in rat tissues,[119] consistent with the notion that the increased thermal stability as compared to the C8-dG-IQ adducts may correlate with decreased NER efficiency. Consequently, it was proposed that the C8-dG-IQ adduct contributes more towards the genotoxic properties of IQ.[123] The role of DNA sequence in modulating the conformations and biological processing of bulky C8-dG adducts is established, with particular emphasis upon the *Nar*I restriction sequence, a hotspot for frameshift mutagenesis in *E. coli*.[124] In this regard, the contributions of the flanking base pairs are likely important; the modulation of both NER and polymerase bypass of C8-dG arylamine

adducts as a function of 3'-flanking sequence, attributed to differential conformational effects, including the orientation of base stacking, has been noted by Jain *et al*.[125] It will thus be of interest to examine sequence-specific conformational effects exhibited by the C8-dG-ABA and $N^2$-dG-ABA adducts, and their possible relationships to both DNA repair and error-prone lesion bypass by Y-family polymerases.

# Chapter III

## Mechanism of Error Free Bypass of the Environmental Carcinogen *N*-(2'-Deoxyguanosin-8-yl)-3-aminobenzanthrone Adduct by Human DNA Polymerase η

### Background

Synthesis of normal DNA is accomplished in a very high fidelity manner by replicative polymerases. However, when DNA has been damaged, replicative polymerases are often incapable of synthesizing the damaged region. This is the case for bulky DNA adducts such as those formed by 3-NBA.[58b] A process known as trans-lesion synthesis (TLS) is responsible for bypass of these lesions. TLS requires the involvement of bypass or TLS polymerases including the Y-family polymerases.[58a, 126]

TLS is a complicated process in which there is much yet to be revealed, however, it is clear that the process must involve several 'polymerase switches'.[127] A replicative polymerase will stall when it encounters a lesion. A TLS polymerase must then replace the stalled polymerase.[128] The TLS polymerase will then process past the lesion and insert bases opposite the lesion in an error-free or error-prone manner. In some cases the polymerase will continue to extend the synthesized strand for a number of additional base pairs before it dissociates, and the replicative polymerase steps back in to complete the synthesis.[127] In other cases, after the TLS polymerase has incorporated a base opposite the lesion, it dissociates and a second TLS polymerase completes the extension step.[129] From this point, again the replicative polymerase can displace the second TLS polymerase and complete the DNA synthesis.

Thus far, every living organism examined has been found to possess some type of TLS polymerase.[130] In many cases organisms will possess more than a single TLS polymerase. The Y-family polymerases are TLS polymerases, and in humans include four members: polymerase eta (Pol η), polymerase iota (Pol ι), polymerase kappa (Pol κ) and Rev1.[126b] Polymerase zeta (Pol ζ) is an additional TLS polymerase found in humans and resides in the B-family of polymerases. Pol η is known for its capacity to effectively bypass pyrimidine photodimers, which occur from exposure to UV radiation.[131] It possesses a large active site that is able to accommodate bulky lesions.[132] Pol ι also possesses a large active site and is capable of bypassing 6-4 pyrimidine adduct. It is unique in that it utilizes Hoogsteen base pairing instead of Watson-Crick pairing to match incoming dNTPs.[133] Pol κ is capable of bypassing many lesions,[134] however it appears to be more specialized in performing the extension step during TLS.[135] Rev 1 is only capable of inserting cytosine opposite a lesion.[136] It has the potential to associate with other polymerases and it is believed that it plays a more important role in polymerase recruitment.[137] Pol ζ is believed to play a major role in the extension step of TLS.[129, 138] The added capacity of these polymerases to handle various DNA adducts is coupled with a reduction in fidelity when compared to replicative polymerases.[126b] This has implications regarding the mutagenicity of adducts processed by TLS.

Structural investigations for members of the Y-family polymerases have previously been reported. Much of this work has been done using the archeal homolog from *S. sulfartaricus*, DNA polymerase IV (Dpo4).[139] In 2001, the first structures of a Y-family polymerase were reported.[140] These structures were ternary complexes with Dpo4, undamaged DNA, and correct or mismatch incoming dNTPs. In 2013, Kirouac *et al*. reported structures of Dpo4 in complex with DNA containing a C8-guanine-aminopyrene adduct.[139d] In a pre-insertion state the adducted

guanine was found to be in an *anti* conformation and the aminopyrene moiety was found either in a hydrophobic pocket of the protein or rotated into the major groove of the DNA. In a post-insertion state the adducted guanine was properly base paired. The adduct was in the major groove of the DNA and shielded from solvent by the little finger domain of the protein.[139d]

Since the initial report for Dpo4, a number of other structures have been reported for other Y-family polymerases, including the human polymerases.[132, 141] In all cases, the Y-family polymerases share certain characteristics including the presence of palm, thumb, and finger domains found in replicative polymerases, plus an additional little finger domain.[126b] A set of four structures were reported for human Pol η in complex with DNA containing a thymine dimer.[132] The large active site of Pol η was able to accommodate the lesion and allowed for Watson-Crick base pairing with the incoming dATP. Pol η was identified as a "molecular splint" in that it was capable of stabilizing the B-type DNA structure of the damaged DNA.

The efficiency and fidelity of TLS varies significantly for different lesions.[126b] Examination of TLS of three different lesions formed by 3-NBA identified different degrees of efficiency and varying mutational spectra for these adducts.[58a] The C8-dG-ABA adduct presented a significant block to TLS polymerases but was bypassed to some extent.[58a] Bypass of this adduct has been shown to produce mutations, with G to T transversions being the major mutation observed.[58c, 142] Despite the capacity for misincorporation opposite the lesion, the majority of bypass events were error-free. This is, perhaps, unexpected given the exceptionally high degree of mutagenicity observed for 3-NBA in the Ames assay.[50b] When TLS of the C8-dG-ABA lesion was investigated in a series of knockdown experiments using HEK293T cells, Pol η was indicated as a major culprit in the mutagenic bypass of the C8-dG-ABA adduct.[58b] However, in all conditions examined, bypass of C8-dG-ABA was still largely error-free.

The mechanism of error-free bypass was of interest and here I report the structure of a Pol η•C8-dG-ABA DNA duplex•dCTP ternary complex trapped in the insertion state. The structure identifies how correct incorporation of cytosine occurs opposite the C8-dG-ABA lesion.

## Structure of C8-dG-ABA in Complex with Pol η in the Insertion State

3-Nitrobenzanthrone (3-nitro-7H-benz-[d,e]anthracen-7-one, 3-NBA, **Chart 8**) is a product of incomplete combustion[70a, 70b] and has been identified as a component of diesel exhaust[70b, 143] and as an environmental contaminant.[49, 70a] The International Association for Research on Cancer (IARC) has classified diesel exhaust as carcinogenic to humans[45] and 3-NBA is a suspected culprit of this carcinogenicity.[45, 70a] 3-NBA has been found to be both highly mutagenic[50b, 55, 58a, 144] and genotoxic,[145] and is itself classified as a potential human carcinogen by the IARC.[45]

Enzymatic reduction *in vivo* engenders a conversion of 3-NBA to aminobenzanthrone (ABA).[146] Intermediates in the reduction pathway can form reactive nitrenium ions that alkylate DNA resulting in ABA adducts.[53, 147] Three different adducts have been identified, and they occur on the C8 (**Chart 8**) or $N^2$ positions of guanine (major adducts), or the $N^6$ position of adenine (minor adduct).[54, 148] Due to the bulky nature of these adducts, it is believed that they present strong blocks to replicative polymerases and bypass is instead accomplished by lower-fidelity translesion synthesis (TLS) polymerases.[58a, 126b, 149] In humans these include the Y-family Pols η, ι, κ and Rev1.[126b, 127, 149]

However, C8-dG-ABA, one of the two major adducts formed,[54, 58a] has previously been implicated in significant blocking of bypass by TLS polymerases in a nucleotide excision repair-

deficient human cell line.[58a] When bypass did occur, of the two dG and one dA adducts, the C8-dG-ABA adduct triggered the highest mutation frequency.[58a] Nevertheless, the majority of bypass events were non-mutagenic, with nearly 70% correct incorporation of cytosine reported.[58a] Another study specifically examined the C8-dG-ABA adduct in human embryonic kidney (HEK293T) cells, and also observed correct incorporation of cytosine opposite this lesion in 86% of bypass events.[58b] A series of siRNA knockdown experiments showed that each of the Y-family polymerases contributed to the bypass of this lesion.[58b] However, hPol η played the most significant role in the mutagenic insertion step, with a decrease in mutation frequency of 39% upon knockdown of hPol η.[58b] In all conditions tested, error-free bypass remained at levels greater than 70%.[58b]

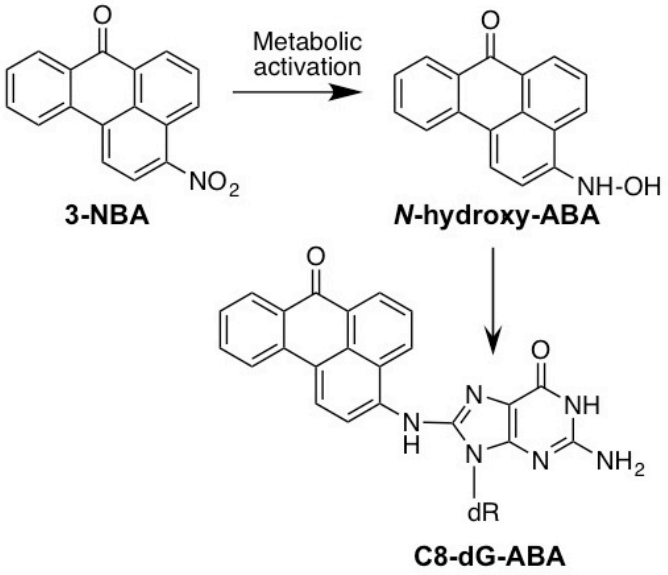



**Chart 8.** Chemical structures of 3-NBA, *N*-OH-ABA and one of the major *in vivo* ABA adducts: C8-dG-ABA.

To gain a better understanding of the mechanism of error-free bypass of the C8-dG-ABA adduct by hPol η, we determined the crystal structure of a ternary Pol•DNA•dCTP complex trapped at the insertion stage. The hPol η catalytic core encompassing residues 1-432 was expressed in *E. coli* and purified following published protocols.[132] The DNA duplex is composed of a 12-mer template with the C8-dG-ABA lesion (X) 5'-CAT XAT GAC GCT-3' and an 8-mer primer 5'-AGC GTC AT-3'. Crystals of the complex with dCTP were grown in the presence of $Ca^{2+}$ instead of $Mg^{2+}$ to prevent incorporation of the incoming nucleoside triphosphate into the primer strand. The structure was phased with molecular replacement using hPol η as the search model and refined to a resolution of 2.6 Å.

An example of the electron density around the final model of the complex is depicted in **Figure 15** and a summary of crystal data, data collection and refinement parameters is given in **Table 4**. The final model allows visualization of 10 of the template nucleotides (the 5'-terminal C and A outside the active site are disordered) and the entire primer strand. The biggest surprise is that the C8-dG-ABA adducted nucleotide adopts the *anti* conformation at the active site of hPol η (**Figures 16, Figure A10**). The conformation is stabilized by a H-bond between the ABA amino moiety and the phosphate group, with the C8-dG-ABA 2'-deoxyribose adopting a C3'-*endo* conformation. As a consequence the adducted G forms a standard Watson-Crick base pair with the incoming dCTP. The ABA adduct is tilted relative to the guanine plane by about 80° and rotated outwards into the major groove of the template-primer duplex.

**Table 4.** Selected crystal data, data collection parameters, and structure refinement statistics

| **Data Collection** | |
|---|---|
| Space group | $P6_1$ |
| Resolution [Å][a] | 50.0-2.60 (2.64-2.60) |
| Unit cell *a, b, c* [Å] | 99.11, 99.11, 81.69 |
| Completeness [%] | 100 (100) |
| I/σ(I) | 15.3 (2.0) |
| R-merge [%] | 13.3 (88.8) |
| Redundancy | 7.6 (7.5) |
| **Refinement** | |
| R-work [%] | 16.6 (21.4) |
| R-free [%][b] | 23.8 (28.1) |
| Avg. B [Å$^2$] | 39.0 |
| R.m.s.d. bonds [Å] | 0.009 |
| R.m.s.d. angles [deg.] | 1.1 |
| PDB ID | 5JUM |

[a] Statistics for the highest-resolution shell are shown in parentheses.

[b] Based on 5% of the reflections.

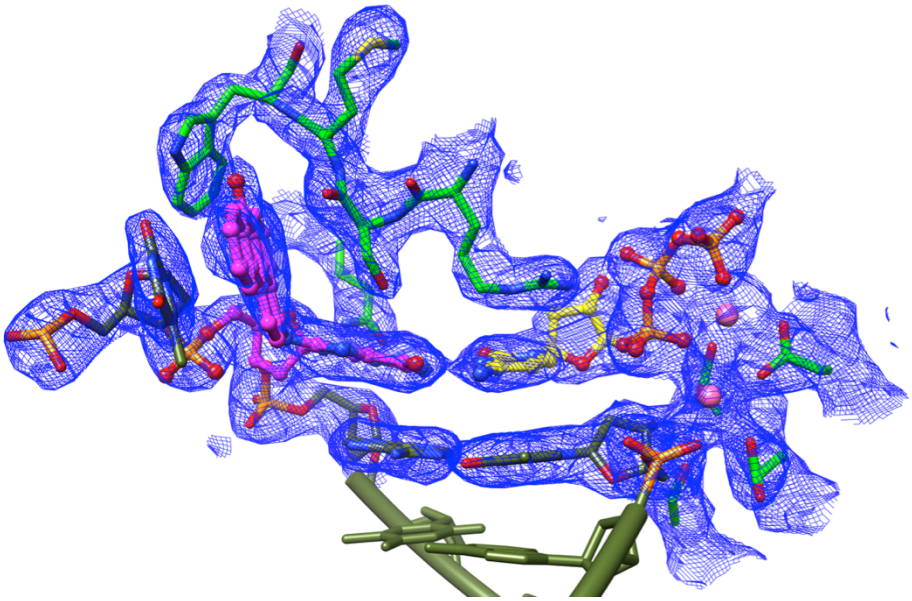


**Figure 15.** Quality of the final Fourier $2F_o$-$F_c$ sum electron density (1σ threshold) around the active site region of the ternary complex between hPol η, C8-dG-ABA adducted template-primer duplex and dCTP. Carbon atoms of hPol η, C8-dG-ABA and dCTP are colored in light green, magenta and yellow, respectively, $Ca^{2+}$ ions are pink spheres, and the remainder of the duplex is colored in olive.

However, rather than exhibiting enhanced conformational flexibility that is often associated with portions of a structure that are extruded from the duplex, the structure offers evidence that the polyaromatic hydrocarbon moiety assumes a configuration that is stabilized by multiple interactions. Thus, ABA forms a sandwich with the amide group of Ser-62 and the T3 template residue that is lodged outside the active site. The carbonyl oxygen of Ser-62 engages in an n→π* interaction with the ABA keto group and Cβ-H and C2′-H from Ser-62 and T3, respectively, engage in C-H→π interactions with aromatic rings from opposite sides of the ABA plane (**Figure 16**). Trp-64 seals the backside of this mostly hydrophobic pocket. Both, the keto

oxygens of ABA and T3 (O2) are too far removed from Trp-64 N-H (distances of 3.71 and 3.83 Å, respectively; **Figure A1**) to engage in H-bonds.

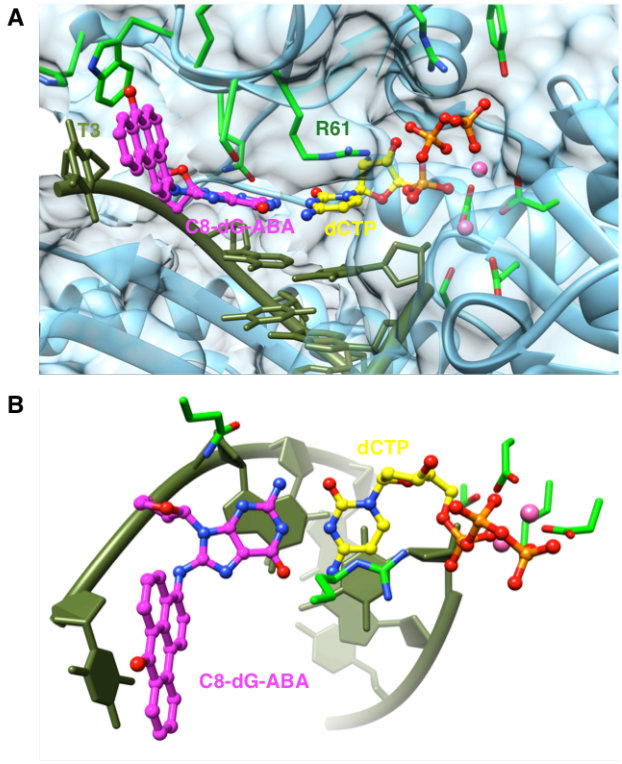



**Figure 16.** Active site configuration in the ternary hPol η insertion-step complex with dCTP opposite C8-dG-ABA. (A*)* View into the DNA major groove, and (B) rotated by ~90° around the horizontal axis and looking roughly along the normal to the nucleobase plane of the incoming dCTP. Carbon atoms of C8-dG-ABA, incoming dCTP, and selected hPol η side chains are colored in magenta, yellow, and light green, respectively, and $Ca^{2+}$ ions are pink spheres.

The extensive interactions observed for ABA at the hPol η active site extend to the base, sugar and phosphate portions of the adducted dG (**Figure 16**, **Figure A1**). On the minor groove side, the $NH_2$ group of Gln-38 from the finger domain forms H-bonds with O4′ of the 2′-

deoxyribose and N3 of the base. The Cγ2(H)$_3$ methyl group of Ile-48 stacks onto the six-membered ring of guanine. In the major groove the hydroxyl group of Ser-61 is positioned somewhat far to establish H-bonds with the $O^6$ and N7 acceptor functions of guanine (distances of 3.80 and 3.75 Å, respectively). Finally, the OH group of Tyr-39 stabilizes the template strand conformation with a H-bond to the non-bridging OP1 phosphate oxygen of C8-dG-ABA. On the opposite side of the active site, the guanidino moiety of Arg-61 from the finger domain stacks on top of the nucleobase of dCTP. The $NH_2$ moiety of the arginine forms salt bridges with the α- and β-phosphate groups and helps align the former for attack by the 3′-OH of the terminal primer nucleotide (distance O3′…Pα= 3.85 Å) together with the two divalent metal ions.

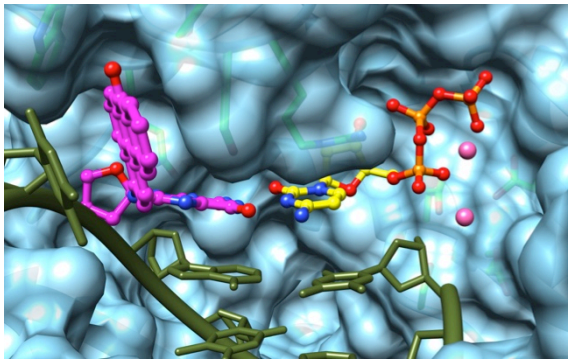


**Figure 17**. Space filling model of the hPol η active site with C8-dG-ABA (magenta carbon atoms) accommodated inside a cleft to the side of the active site that allows for extensive hydrophobic interactions between ABA moiety and surrounding residues from the polymerase finger domain. Carbon atoms of the incoming dCTP are colored in yellow.

The tight fit between the nascent C8-dG-ABA:dCTP base pair and hPol η is further illustrated in **Figure 17** which depicts a space filling model of the active site of the polymerase. Finger residues Ser-62 and Arg-61 (foreground, center) protrude from the ceiling along with

Gln-38, and bear down on the G:C base-pair plane at the split between the primer and template strands.

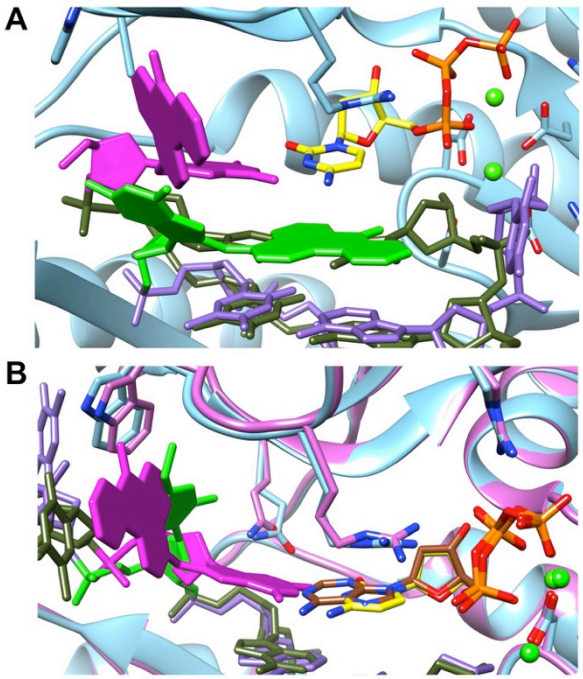



**Figure 18**. Model of error-prone bypass of C8-dG-ABA and comparison of hPol η complexes with the C8-dG-ABA and $O^4$MedT adducts. (A) Superimposition of the crystal structure of the hPol η (Pol ribbon in cyan) complex with C8-dG-ABA (magenta/olive) and dCTP (yellow) and the NMR structure of DNA (lilac) with the C8-dG-ABA (light green) lesion.[150] (B) Superimposition of the hPol η complexes with C8-dG-ABA (magenta/ olive)/dCTP (yellow; Pol ribbon in cyan) and $O^4$MedT (light green/lilac)/dATP (brown; Pol ribbon in pink).[151]

Two metal ions are accommodated in a pocket on one side of the active site, by conserved Glu and Asp residues. On the other side, where the template strand enters the active site, the ABA adduct fits snugly into a cleft formed by Ser-62, Met-63 (main chain) and Trp-64 and extended by T3 from the template strand. The structure thus offers insight into how hPol η

stabilizes a configuration of the template strand that allows error-free bypass of the bulky C8-dG-ABA lesion.

We recently reported the NMR solution structure of a DNA 12mer duplex containing a C8-dG-ABA nucleotide opposite dC.[150] In this structure, ABA was intercalated in the duplex, with the adducted G in the *syn* conformation and extruded into the major groove. The intercalated adduct thus takes the place of a base pair and thereby forces the opposing cytosine into the major groove (**Figure 13**). We superimposed this DNA duplex model with base pairs on the 5′-side of the adducted dG deleted onto the template-primer duplex of the ternary hPol η complex, such that the intercalated ABA assumes the position of the T:A pair at the -1 position at the Pol active site (**Figure 18A**). This configuration illustrates a potential basis for error-prone bypass of C8-dG-ABA, in that incoming dATP or dGTP might be preferred due to favorable stacking. Insertion could then proceed in a non-templated fashion, causing a mutation, or opposite a residue located 5′-adjacent to C8-dG-ABA, thus causing a frameshift.

hPol η inserts dCTP ca. 3.5-fold more efficiently than dATP opposite another C8 adduct of dG, 8-oxoG.[142] Incorporation of dATP occurs with the adducted dG in the *syn* orientation and stacked inside the template strand. However, in comparison to the $C_8$=$O_8$ moiety of 8-oxoG, the ABA adduct is far too bulky to be accommodated within the minor groove.

Finally, the C8-dG-ABA:hPol η complex constitutes the second example of an insertion-stage bypass structure with an adduct lodged in the hydrophobic cleft to the side of the active site. A somewhat similar orientation was found for $O^4$MedT (**Figure 18B**) that is replicated by hPol η in an error-prone fashion.[151] However, unlike in the case of C8-dG-ABA where only the ABA portion is inserted into the cleft, the entire $O^4$MedT nucleobase can be accommodated inside the pocket, with the lesion reaching all the way to the back of the pocket.

## Materials and Methods

### *hPol η Catalytic Core Protein Expression and Purification*

The hPol η plasmid (pET28a) comprising residues 1-432 was a generous gift from Dr. Wei Wang, NIDDK, NIH. The polymerase was expressed in *E. coli* and purified as previously described.[132] Briefly, *E. coli* transfection of the plasmid was performed using a heat shock method and cultures were produced on agarose plates in the presence of Kanamycin. Single colonies were isolated and introduced to 100 mL volumes of LB broth also containing Kanamycin. Cultures were incubated overnight at 25 ºC and transferred to 1 L LB broth containing flasks after culture absorbance was recorded at 0.8 OD. Protein expression was induced with isopropyl β-D-1-thiogalatopyranoside (IPTG) and cells were harvested after 24 hours. Cellular protein was isolated by centrifugation and subjected to further purification.

Cellular protein was purified by $Ni^{2+}$-affinity, MonoS and Superdex75 chromatography. The His tag was removed by PreScission protease. Purity was confirmed using SDS-Page. The purified protein solution was concentrated to 5 mg/mL and stored in 20% glycerol at -80 ºC and used within 24 hours of being thawed.

### *Oligonucleotide Synthesis and Annealing*

C8-dG-ABA modified 12mer template 3′-TCG CAG TAX TAC-5′ (X= C8-dG-ABA) and unmodified 8mer primer 5′-AGC GTC AT-3′were used in the crystallization experiments. The synthesis of site-specifically modified oligonucleotides containing the C8-dG-ABA lesion was performed as previously reported and placed in sequence.[152] Single strand oligodeoxynucleotides were characterized using a Voyager MALDI-TOF spectrometer (Applied

Biosystems, Grand Island, NY). A matrix consisting of 3-hydroxypicolinic acid in ammonium hydrogen citrate was used. Mass spectra were recorded in the negative ion mode and recorded to ± 1 *m/z*.. Purity was confirmed by capillary gel electrophoresis and HPLC as previously described in Chapter II. Unmodified DNA primer was purchased from Integrated DNA Technologies (Coralville, IA). Template and primer strands were annealed at 85 ºC in a 1:1 molar ratio in the presence of 10 mM sodium HEPES buffer (pH 8.0), 0.1 mM EDTA, and 50 mM NaCl. Following 10 minutes at 85 ºC the solution was slowly cooled to room temperature. Oligodeoxynucleotide concentrations were determined by absorbance spectroscopy measurements at 260 nm. The calculated extinction coefficients[95] were 1.18 x $10^5$ L $mol^{-1}$ $cm^{-1}$ for the modified template strand and 7.89 x $10^4$ L $mol^{-1}$ $cm^{-1}$ for the primer strand. The calculated extinction coefficient of the modified strand was not corrected for the presence of the ABA adduct.

*Crystallization of the hPol η•C8-dG-ABA DNA•dCTP Insertion Complex*

DNA template-primer duplex was added to solution with the protein in a 1.2:1 molar ratio. The buffer solution was comprised of 50 mM Tris-HCl, 450 mM KCl, and 3 mM DTT at pH 7.5. 5 mL of 100 mM $CaCl_2$ was added to the complex and the solution was concentrated to a final concentration of ~2-3 mg/mL by ultrafiltration. dCTP was added following ultrafiltration to form the ternary complex. Crystallization experiments utilized the hanging drop vapor diffusion technique at 18°C using a sparse matrix screen (Hampton Research, Aliso Viejo, CA). One µL of the complex solution was mixed with 1 µL of reservoir solution and equilibrated against 500 µL reservoir wells. Crystals appeared in droplets containing 0.1 M MES (pH 5.5), 5 mM $CaCl_2$, and 25% (w/v) PEG 2000 MME within one day and were harvested after a week (**Figure 19**).

Crystals were mounted in nylon loops, cryo-protected in reservoir solution containing 25% glycerol (v/v), and frozen in liquid nitrogen.

**Figure 19**. Crystal of hPol η•C8-dG-ABA DNA•dCTP complex used to collect diffraction data.

*X-ray Diffraction Data Collection, Structure Determination and Refinement*

The 21-ID-D beamline of the Life Sciences Collaborative Access Team (LS-CAT) at the Advanced Photon Source, Argonne National Laboratory (Argonne, IL) was used to collect diffraction data. The program HKL2000[153] was used to integrate and scale all data. Structures were determined by the Molecular Replacement technique with the program MOLREP.[154] The hpol η structure PDB entry 4O3N (protein only)[142] was used as the search model for molecular replacement. Structure refinement and model building were carried out with Phenix[155] and COOT,[156] respectively. Illustrations were prepared with the program UCSF Chimera.[157]

# Chapter IV

## Summary

*Summary*

In Chapter II I reported the base-displaced intercalated structure of a C8-dG-ABA containing duplex. In Chapter III I reported the structure of a ternary Pol η•C8-dG-ABA DNA•dCTP complex trapped at the insertion stage.

The C8-dG-ABA duplex was significantly altered from an undamaged B-type DNA duplex. The ABA moiety of the lesion intercalated in the duplex and displaced the opposing cytosine into the major groove, thereby eliminating the base pairing at the lesion site. In order to allow for intercalation, the adducted guanine was rotated about the glycosidic bond into a *syn* conformation. This resulted in the adducted guanine being placed in the major groove. The base pairs neighboring the lesion site remained intact. The ABA moiety was positioned with the bay region of the adduct facing the major groove. The ring structure of the adduct was found to occupy the position of the complementary cytosine in the undamaged duplex. This allowed π-stacking interactions to occur between the adduct and the bases neighboring the complementary cytosine. Regions of the duplex beyond the lesion area (the lesion, direct neighbors, and their complements) resembled natural B-type DNA and were unremarkable. The stability of the duplex was reduced in comparison to the undamaged duplex. These findings were similar to other bulky C8-guanine adducts when placed in sequence between two pyrimidine bases and opposite cytosine.

The ternary Pol η•C8-dG-ABA DNA•dCTP complex captured at the insertion stage indicated a mechanism for error-free insertion. The adducted guanine was found to be in the *anti* conformation about the glycosidic bond and Watson-Crick base paired with an incoming dCTP in the active site of Pol η. The *anti* conformation was stabilized by a π-bond between the adduct amino group and the phosphate backbone of the DNA. The ABA moiety was rotated into the major groove of the duplex and resided in a hydrophobic cleft of the protein. This position was stabilized by multiple interactions, including interactions with Ser-62 of the protein and the T3 template residue. In this position, the ABA moiety was sandwiched between protein residues and the T3 base. This positioning did not present a clear path for the adduct to move into the active site of the protein. This offers a potential explanation as to why C8-dG-ABA is not efficiently bypassed during TLS. It may require that the polymerase dissociates from the duplex and subsequently Pol η, or another TLS polymerase, associates with the duplex in a post-insertion orientation to continue TLS.

In tandem, these structures demonstrate that Pol η altered the conformation of the C8-dG-ABA adduct to enact the error-free bypass of the lesion. In comparison to the duplex structure, the adducted guanine was rotated about the glycosidic bond from a *syn* conformation into the natural *anti* conformation. This enabled a Watson-Crick base pair to be formed with an incoming dCTP and repositioned the ABA moiety from an intercalated position into the major groove. Modeling of C8-dG-ABA in an intercalated position in complex with Pol η revealed a possible mechanism for the mutagenic bypass of the lesion. The intercalated ABA moiety extended into a position beneath the incoming dNTP. The aromatic ring system of the adduct could form favorable π-stacking interactions with an incoming dATP. This is similar to the π-stacking

interaction observed in the duplex structure between the ABA ring system and $A^{19}$ of the complementary strand (**Figure 14**).

The results reported herein provide insight into how the C8-dG-ABA adduct is bypassed in an error-free fashion. A potential mechanism is also provided for the mutagenic bypass of the adduct. Understanding these mechanisms helps in assessing the true risk this type of adduct poses to human health. In the case of 3-NBA this can have a direct impact on diesel emission controls and regulations.

*Future Directions*

The structure of C8-dG-ABA within a DNA duplex was a base-displaced intercalated structure. This is similar to other structures reported for C8-guanine adducts which are formed by nitroarenes, when placed between two pyrimidine nucleotides. Variations in this type of structure have been reported when one or two purine bases are placed adjacent to the lesion, or when a mismatch is placed opposite the lesion.[158] Examination of the C8-dG-ABA adduct in different sequence contexts could reveal if sequence context plays a role in the structural changes induced by C8-dG-ABA. Furthermore, mutational spectra resulting from C8-dG-ABA appears to be affected by sequence context.[58b] Identifying any structural differences in the context of a DNA duplex as a function of sequence context could help glean insight into this phenomenon. Additionally, a mismatch at the lesion site could alter the structure of the duplex and examination of a mismatch could uncover a potential variation in structure.

The structure of the Pol η•C8-dG-ABA DNA•dCTP complex was captured in the non-mutagenic insertion state. In this state the ABA moiety was positioned in a hydrophobic pocket with no clear path for translocation. Further studies examining the +1 insertion state could help

identify how the polymerase processes beyond the lesion. Examination of C8-dG-ABA in complex with Pol κ or Pol ζ would be of further interest as they have been implicated in the extension step for this adduct.[58b]

While Pol η bypasses the lesion in a primarily error-free manner, a significant level of mutations have been observed. Therefore, future studies examining mismatches during the insertion stage are of interest. G to T transversions have been reported as the major mutation. Thus, examination at the insertion state with a dATP opposite the lesion could shed light on how mutagenic bypass occurs.

## References


1. Seidler, R. J.; Mandel, M., Quantitative Aspects of Deoxyribonucleic Acid Renaturation: Base Composition, State of Chromosome Replication, and Polynucleotide Homologies. *J. Bacteriol.* **1971**, *106* (2), 608-614.

2. Giaever, G. N.; Wang, J. C., Supercoiling of Intracellular DNA can Occur in Eukaryotic Cells. *Cell* **1988**, *55* (5), 849-856.

3. Watson, J. D.; Crick, F. H., Genetical Implications of the Structure of Deoxyribonucleic Acid. *JAMA, J. Am. Med. Assoc.* **1993**, *269* (15), 1967-1969.

4. Wilkins, M. H. F.; Stokes, A. R.; Wilson, H. R., Molecular Structure of Nucleic Acids: Molecular Structure of Deoxypentose Nucleic Acids. *Nature* **1953**, *171* (4356), 738-740.

5. Watson, J.; Crick, F., Molecular Structure of Nucleic Acids. *Nature* **1953**, 171.4356, 737-738.

6. Levine, C.; Chargaff, E., Phosphatide Composition in Different Liver Cell Fractions. *Exp. Cell Res.* **1952**, *3* (1), 154-162.

7. Dickerson, R. E.; Drew, H. R.; Conner, B. N.; Wing, R. M.; Fratini, A. V.; Kopka, M. L., The Anatomy of A-, B-, and Z-DNA. *Science* **1982**, *216* (4545), 475-485.

8. Wang, A. H.-J.; Quigley, G. J.; Kolpak, F. J.; Crawford, J. L.; Van Boom, J. H.; van der Marel, G.; Rich, A., Molecular Structure of a Left-handed Double Helical DNA Fragment at Atomic Resolution. *Nature* **1979**, *282*, 680-686.

9. (a) Li, J.; Correia, J. J.; Wang, L.; Trent, J. O.; Chaires, J. B., Not so Crystal Clear: the Structure of the Human Telomere G-quadruplex in Solution Differs from that Present in a Crystal. *Nucleic Acids Res.* **2005**, *33* (14), 4649-4659; (b) Wang, Y.; Patel, D. J., Solution

Structure of the Human Telomeric Repeat d [AG 3 (T 2 AG 3) 3] G-tetraplex. *Structure* **1993**, *1* (4), 263-282.

10. Barton, S. C.; Surani, M.; Norris, M., Role of Paternal and Maternal Genomes in Mouse Development. *Nature* **1984**, *311*, 374-376.

11. Cook, R. E., Asexual Reproduction: a Further Consideration. *Am. Nat*. **1979**, *113* (5), 769-772.

12. Johnson, R. E.; Klassen, R.; Prakash, L.; Prakash, S., A Major Role of DNA Polymerase δ in Replication of Both the Leading and Lagging DNA Strands. *Mol. Cell* **2015**, *59* (2), 163-175.

13. Speyer, J.; Karam, J. D.; Lenny, A., On the Role of DNA Polymerase in Base Selection. *Cold Spring Harbor Symp. Quant. Biol.* **1966**, pp 693-697.

14. Yudelevich, A.; Ginsberg, B.; Hurwitz, J., Discontinuous Synthesis of DNA During Replication. *Proc. Natl. Acad. Sci. USA* **1968**, *61* (3), 1129-1136.

15. Eckert, K. A.; Kunkel, T. A., DNA Polymerase Fidelity and the Polymerase Chain Reaction. *Genome Res.* **1991**, *1* (1), 17-24.

16. Champoux, J. J., DNA Topoisomerases: Structure, Function, and Mechanism. *Annu. Rev. Biochem.* **2001**, *70* (1), 369-413.

17. Matson, S. W.; Bean, D. W.; George, J. W., DNA Helicases: Enzymes with Essential Roles in all Aspects of DNA Metabolism. *BioEssays* **1994**, *16* (1), 13-22.

18. Wold, M. S., Replication Protein A: a Heterotrimeric, Single-stranded DNA-binding Protein Required for Eukaryotic DNA Metabolism. *Annu. Rev. Biochem*. **1997**, *66* (1), 61-92.

19. Lehnman, I., DNA Ligase: Structure, Mechanism, and Function. *Science* **1974**, *186* (4166), 790-797.

20. Frick, D. N.; Richardson, C. C., DNA Primases. *Annu. Rev. Biochem*. **2001**, *70* (1), 39-80.

21. Friedberg, E. C., DNA Damage and Repair. *Nature* **2003**, *421* (6921), 436-440.

22. Lindahl, T.; Modrich, P.; Sancar, A., The Nobel Prize in Chemistry 2015. *Nobelprize.org. Nobel Media* **2015**, http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2015/ (accessed June 20, 2016).

23. (a) Ames, B. N., Endogenous Oxidative DNA Damage, Aging, and Cancer. *Free Radical Res. Commun*. **1989**, *7* (3-6), 121-128; (b) De Bont, R.; Van Larebeke, N., Endogenous DNA Damage in Humans: a Review of Quantitative Data. *Mutagenesis* **2004**, *19* (3), 169-185.

24. Marnett, L. J., Oxyradicals and DNA Damage. *Carcinogenesis* **2000**, *21* (3), 361-370.

25. Grollman, A. P.; Moriya, M., Mutagenesis by 8-Oxoguanine: an Enemy Within. *Trends Genet*. **1993**, *9* (7), 246-249.

26. Setlow, R.; Carrier, W., Pyrimidine Dimers in Ultraviolet-irradiated DNA's. *J. Mol. Biol.* **1966**, *17* (1), 237-254.

27. Leanderson, P.; Tagesson, C., Cigarette Smoke-induced DNA-damage: Role of Hydroquinone and Catechol in the Formation of the Oxidative DNA-adduct, 8-Hydroxydeoxyguanosine. *Chem.-Biol. Interact*. **1990**, *75* (1), 71-81.

28. Benbrahim-Tallaa, L.; Baan, R.; Grosse, Y.; Lauby-Secretan, B.; El Ghissassi, F.; Bouvard, V.; Guha, N.; Loomis, D.; Straif, K.; Portier, C., Carcinogenicity of diesel-engine and gasoline-engine exhausts and some nitroarenes. *Lancet Oncol*. **2012**, *13*.7, 663-664.

29. Felton, J.; Knize, M., Heterocyclic-amine Mutagens/Carcinogens in Foods. In *Chemical Carcinogenesis and Mutagenesis I*, Springer: 1990; pp 471-502.

30. (a) Williams, J. H.; Phillips, T. D.; Jolly, P. E.; Stiles, J. K.; Jolly, C. M.; Aggarwal, D., Human Aflatoxicosis in Developing Countries: a Review of Toxicology, Exposure, Potential Health Consequences, and Interventions. *Am. J. Clin. Nutr*. **2004**, *80* (5), 1106-1122; (b) Smela, M. E.; Hamm, M. L.; Henderson, P. T.; Harris, C. M.; Harris, T. M.; Essigmann, J. M., The Aflatoxin B1 Formamidopyrimidine Adduct Plays a Major Role in Causing the Types of Mutations Observed in Human Hepatocellular Carcinoma. *Proc. Natl. Acad. Sci. USA* **2002**, *99* (10), 6655-6660.

31. (a) Krokan, H. E.; Bjørås, M., Base Excision Repair. *Cold Spring Harbor Perspect. Biol.* **2013**, *5* (4), a012583; (b) Dianov, G.; Lindahl, T., Reconstitution of the DNA Base Excision-repair Pathway. *Curr. Biol*. **1994**, *4* (12), 1069-1076.

32. Memisoglu, A.; Samson, L., Base Excision Repair in Yeast and Mammals. *Mutat. Res., Fundam. Mol. Mech. Mutagen*. **2000**, *451* (1), 39-51.

33. Sancar, A., Mechanisms of DNA Excision Repair. *Science* **1994**, *266* (5193), 1954.

34. He, Z.; Henricksen, L. A.; Wold, M. S.; Ingles, C. J., RPA Involvement in the Damage-recognition and Incision Steps of Nucleotide Excision Repair. *Nature* **1995**, *374*, 566 - 569

35. Huang, J.-C.; Hsu, D. S.; Kazantsev, A.; Sancar, A., Substrate Spectrum of Human Excinuclease: Repair of Abasic Sites, Methylated Bases, Mismatches, and Bulky Adducts. *Proc. Natl. Acad. Sci. USA* **1994**, *91* (25), 12213-12217.

36. Solomon, S.; Plattner, G.-K.; Knutti, R.; Friedlingstein, P., Irreversible Climate Change due to Carbon Dioxide Emissions. *Proc. Natl. Acad. Sci. USA* **2009**, *106* (6), 1704–1709, doi: 10.1073/pnas.0812721106

37. Likens, G. E.; Wright, R. F.; Galloway, J. N.; Butler, T. J., Acid Rain. *Sci. Am.* **1979**, *241* (4).

38. (a) Schindler, D. W., Effects of Acid Rain on Freshwater Ecosystems. *Science* **1988**, *239* (4836), 149-157; (b) Velikova, V.; Yordanov, I.; Edreva, A., Oxidative Stress and some Antioxidant Systems in Acid Rain-treated Bean Plants: Protective Role of Exogenous Polyamines. *Plant Sci*. **2000**, *151* (1), 59-66.

39. Reger, R.; Hancock, J.; Hankinson, J.; Hearl, F.; Merchant, J., Coal Miners Exposed to Diesel Exhaust Emissions. *Ann. Occup. Hyg*. **1982**, *26* (8), 799-815.

40. Garshick, E.; Schenker, M. B.; Muñoz, A.; Segal, M.; Smith, T. J.; Woskie, S. R.; Hammond, S. K.; Speizer, F. E., A Retrospective Cohort Study of Lung Cancer and Diesel Exhaust Exposure in Railroad Workers. *Am. J. Respir. Crit. Care Med*. **1988**, *137* (4), 820-825.

41. (a) Steenland, K.; Deddens, J.; Stayner, L., Diesel Exhaust and Lung Cancer in the Trucking Industry: Exposure-response Analyses and Risk Assessment. *Am. J. Ind. Med*. **1998**, *34* (3), 220-228; (b) Gustavsson, P.; Plato, N.; Lidström, E.-B.; Hogstedt, C., Lung Cancer and Exposure to Diesel Exhaust among Bus Garage Workers. *Scand. J. Work, Environ. Health* **1990**, 348-354.

42. Mauderly, J. L.; Jones, R. K.; Griffith, W. C.; Henderson, R. F.; McClellan, R. O., Diesel Exhaust is a Pulmonary Carcinogen in Rats Exposed Chronically by Inhalation. *Toxicol. Sci.* **1987**, *9* (2), 208-221.

43. Kunitake, E.; Shimamura, K.; Katayama, H.; Takemoto, K.; Yamamoto, A.; Hisanaga, A.; Ohyama, S.; Ishinishi, N., Studies Concerning Carcinogenesis of Diesel Particulate Extracts Following Intratracheal Instillation, Subcutaneous Injection, or Skin Application. *Dev. Toxicol. Environ. Sci*. **1985**, *13*, 235-252.

44. Bhatia, R.; Lopipero, P.; Smith, A. H., Diesel Exhaust Exposure and Lung Cancer. *Epidemiology* **1998**, *9* (1), 84-91.

45. Cancer, I. A. f. R. o., Diesel and Gasoline Engine Exhausts and some Nitroarenes. *IARC Monogr. Eval. Carcinog. Risks Hum*. **2013**, *105*.

46. Majewski, W. A.; Khair, M. K., *Diesel Emissions and Their Control*. Society of Automotive Engineers, 2006.

47. Johnson, T. V. Diesel Emission Control: 2001 in Review; *SAE Technical paper*: 2002-01-0285, 2002, doi:10.4271/2002-01-0285.

48. Pitts, J. N.; Lokensgard, D. M.; Harger, W.; Fisher, T. S.; Mejia, V.; Schuler, J. J.; Scorziell, G. M.; Katzenstein, Y. A., Mutagens in Diesel Exhaust Particulate Identification and Direct Activities of 6-Nitrobenzo [a] pyrene, 9-Nitroanthracene, 1-Nitropyrene and 5H-Phenanthro [4, 5-bcd] pyran-5-one. *Mutat. Res. Lett*. **1982**, *103* (3), 241-249.

49. Murahashi, T.; Iwanaga, E.; Watanabe, T.; Hirayama, T., Determination of the Mutagen 3-Nitrobenzanthrone in Rainwater Collected in Kyoto, Japan. *J. Health Sci*. **2003**, *49* (5), 386-390.

50. (a) Howard, P. C.; Heflich, R. H.; Evans, F. E.; Beland, F. A., Formation of DNA Adducts in Vitro and in Salmonella typhimurium upon Metabolic Reduction of the Environmental Mutagen 1-Nitropyrene. *Cancer Res*. **1983**, *43* (5), 2052-2058; (b) Enya, T.; Suzuki, H.; Watanabe, T.; Hirayama, T.; Hisamatsu, Y., 3-Nitrobenzanthrone, a Powerful Bacterial Mutagen and Suspected Human Carcinogen found in Diesel Exhaust and Airborne Particulates. *Environ. Sci. Technol*. **1997**, *31* (10), 2772-2776.

51. Howard, P. C.; Beland, F. A., Xanthine Oxidase Catalyzed Binding of 1-Nitropyrene to DNA. *Biochem. Biophys. Res. Commun*. **1982**, *104* (2), 727-732.

52. Arlt, V. M.; Stiborova, M.; Henderson, C. J.; Osborne, M. R.; Bieler, C. A.; Frei, E.; Martinek, V.; Sopko, B.; Wolf, C. R.; Schmeiser, H. H., Environmental Pollutant and

Potent Mutagen 3-Nitrobenzanthrone Forms DNA Adducts after Reduction by NAD (P) H:Quinone Oxidoreductase and Conjugation by Acetyltransferases and Sulfotransferases in Human Hepatic Cytosols. *Cancer Res*. **2005**, *65* (7), 2644-2652.

53. Xue, J.; Du, L.; Zhu, R.; Huang, J.; Phillips, D. L., Direct Time-Resolved Spectroscopic Observation of Arylnitrenium Ion Reactions with Guanine-Containing DNA Oligomers. *J. Org. Chem*. **2014**, *79* (8), 3610-3614.

54. Arlt, V. M.; Schmeiser, H. H.; Osborne, M. R.; Kawanishi, M.; Kanno, T.; Yagi, T.; Phillips, D. H.; Takamura-Enya, T., Identification of Three Major DNA Adducts Formed by the Carcinogenic Air Pollutant 3-Nitrobenzanthrone in Rat Lung at the C8 and N2 Position of Guanine and at the N6 Position of Adenine. *Int. J. Cancer* **2006**, *118* (9), 2139-2146.

55. Purohit, V.; Basu, A. K., Mutagenicity of Nitroaromatic Compounds. *Chem. Res. Toxicol*. **2000**, *13* (8), 673-692.

56. (a) Stanton, C. A.; Garner, C. R.; Martin, C. N., The Mutagenicity and DNA Base Sequence Changes Induced by 1-Nitroso-and 1-Nitropyrene in the cl Gene of Lambda Prophage. *Carcinogenesis* **1988**, *9* (7), 1153-1158; (b) Malia, S. A.; Vyas, R. R.; Basu, A. K., Site-specific Frame-shift Mutagenesis by the 1-Nitropyrene-DNA Adduct N-(Deoxyguanosin-8-yl)-1-Aminopyrene Located in the (CG) 3 Sequence: Effects of SOS, Proofreading, and Mismatch Repair. *Biochemistry* **1996**, *35* (14), 4568-4577.

57. Watt, D. L.; Utzat, C. D.; Hilario, P.; Basu, A. K., Mutagenicity of the 1-Nitropyrene-DNA Adduct N-(Deoxyguanosin-8-yl)-1-aminopyrene in Mammalian Cells. *Chem. Res. Toxicol.* **2007**, *20* (11), 1658-1664.

58. (a) Kawanishi, M.; Fujikawa, Y.; Ishii, H.; Nishida, H.; Higashigaki, Y.; Kanno, T.; Matsuda, T.; Takamura-Enya, T.; Yagi, T., Adduct Formation and Repair, and Translesion DNA

Synthesis Across the Adducts in Human Cells Exposed to 3-Nitrobenzanthrone. *Mutat. Res., Genet. Toxicol. Environ. Mutagen.* **2013** *753* (2), 93-100; (b) Pande, P.; Malik, C. K.; Bose, A.; Jasti, V. P.; Basu, A. K., Mutational Analysis of the C8-Guanine Adduct of the Environmental Carcinogen 3-Nitrobenzanthrone in Human Cells: Critical Roles of DNA Polymerases η and κ and Rev1 in Error-Prone Translesion Synthesis. *Biochemistry* **2014**, *53* (32), 5323-5331; (c) vom Brocke, J.; Krais, A.; Whibley, C.; Hollstein, M. C.; Schmeiser, H. H., The Carcinogenic Air Pollutant 3-Nitrobenzanthrone Induces GC to TA Transversion Mutations in Human p53 Sequences. *Mutagenesis* **2009**, *24* (1), 17-23.

59. Carrington, A.; McLachlan, A. D., *Introduction to Magnetic Resonance*; Harper and Row: New York, 1967.

60. Farrar, T. C.; Becker, E. D., *Pulse and Fourier Transform NMR: Introduction to Theory and Methods*; Elsevier, 2012.

61. Skelton, N. J.; Chazin, W., Solution Structure Determination of Proteins by Nuclear Magnetic Resonance Spectroscopy. *Drugs Pharm. Sci*. **2000**, *101*, 683-726.

62. Neuhaus, D.; Williamson, M. P., *The Nuclear Overhauser Effect in Structural and Conformational Analysis*; VCH: New York, 1989.

63. Ernst, R. R.; Bodenhausen, G.; Wokaun, A., *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*; Clarendon Press Oxford, 1987; Vol. 14.

64. von Laue, M., Concerning the Detection of X-ray Interferences. *Nobel lecture* **1915**, 13.

65. Egli, M., Diffraction Techniques in Structural Biology. *Curr. Protoc. in Nucleic Acid Chem*. **2010**, *65,* 7.13.1-7.13.41. doi: 10.1002/cpnc.4

66. Taylor, G., The Phase Problem. *Acta Crystallogr., Sect. D: Biol. Crystallogr*. **2003**, *59* (11), 1881-1890.

67. Hendrickson, W. A.; Pähler, A.; Smith, J. L.; Satow, Y.; Merritt, E. A.; Phizackerley, R. P., Crystal Structure of Core Streptavidin Determined from Multiwavelength Anomalous Diffraction of Synchrotron Radiation. *Proc. Natl. Acad. Sci. USA* **1989**, *86* (7), 2190-2194.

68. Hendrickson, W. A., Determination of Macromolecular Structures from Anomalous Diffraction of Synchrotron Radiation. *Science* **1991**, *254* (5028), 51-58.

69. Rossmann, M. G., The Molecular Replacement Method. *Acta Crystallogr., Sect. A: Found. Crystallogr*. **1990**, *46* (2), 73-82.

70. (a) Wahnschaffe, U.; Kielhorn, J.; Mangelsdorf, I., Selected Nitro- and Nitro-oxy-polycyclic Aromatic Hydrocarbons; Geneva World Health Organization, 2003. (b) Murahashi, T., Determination of Mutagenic 3-Nitrobenzanthrone in Diesel Exhaust Particulate Matter by Three-dimensional High-performance Liquid Chromatography. *Analyst* **2003**, *128* (1), 42-45.

71. (a) Schuetzle, D.; Lee, F. S. C.; Prater, T. J.; Tejada, S. B., The Identification of Polynuclear Aromatic Hydrocarbon (PAH) Derivatives in Mutagenic Fractions of Diesel Particulate Extracts. *Int. J. Environ. Anal. Chem*. **1981**, *9* (2), 93-144; (b) Paputapeck, M. C.; Marano, R. S.; Schuetzle, D.; Riley, T. L.; Hampton, C. V.; Prater, T. J.; Skewes, L. M.; Jensen, T. E.; Ruehle, P. H.; Bosch, L. C.; Duncan, W. P., Determination of Nitrated Polynuclear Aromatic-Hydrocarbons in Particulate Extracts by Capillary Column Gas-chromatography with Nitrogen Selective Detection. *Anal. Chem*. **1983**, *55* (12), 1946-1954; (c) Arlt, V. M.; Stiborova, M.; Henderson, C. J.; Osborne, M. R.; Bieler, C. A.; Frei, E.; Martinek, V.; Sopko, B.; Wolf, C. R.; Schmeiser, H. H.; Phillips, D. H., Environmental Pollutant and Potent Mutagen 3-Nitrobenzanthrone Forms DNA Adducts

after Reduction by NAD(P)H: Quinone Oxidoreductase and Conjugation by Acetyltransferases and Sulfotransferases in Human Hepatic Cytosols. *Cancer Res*. **2005**, *65* (7), 2644-2652.

72. Murahashi, T.; Iwanaga, E.; Watanabe, T.; Hirayama, T., Determination of the Mutagen 3-Nitrobenzanthrone in Rainwater Collected in Kyoto, Japan. *J. Health Sci.* **2003**, *49* (5), 386-390.

73. Enya, T.; Suzuki, H.; Watanabe, T.; Hirayama, T.; Hisamatsu, Y., 3-Nitrobenzanthrone, a Powerful Bacterial Mutagen and Suspected Human Carcinogen Found in Diesel Exhaust and Airborne Particulates. *Environ. Sci. Technol*. **1997**, *31* (10), 2772-2776.

74. (a) Ames, B. N.; McCann, J.; Yamasaki, E., Methods for Detecting Carcinogens and Mutagens with Salmonella-mammalian-microsome Mutagenicity Test. *Mutat. Res*. **1975**, *31*, 347-363; (b) Maron, D. M.; Ames, B. N., Revised Methods for the Salmonella Mutagenicity Test. *Mutat. Res*. **1983**, *113* (3-4), 173-215.

75. Fifer, E. K.; Howard, P. C.; Heflich, R. H.; Beland, F. A., Synthesis and Mutagenicity of 1-Nitropyrene 4,5-oxide and 1-Nitropyrene 9,10-oxide, Microsomal Metabolites of 1-Nitropyrene. *Mutagenesis* **1986**, *1* (6), 433-438.

76. (a) Arlt, V. M.; Sorg, B. L.; Osborne, M.; Hewer, A.; Seidel, A.; Schmeiser, H. H.; Phillips, D. H., DNA Adduct Formation by the Ubiquitous Environmental Pollutant 3-Nitrobenzanthrone and its Metabolites in Rats. *Biochem. Biophys. Res. Comm*. **2003**, *300* (1), 107-114; (b) Bieler, C. A.; Cornelius, M. G.; Stiborova, M.; Arlt, V. M.; Wiessler, M.; Phillips, D. H.; Schmeiser, H. H., Formation and Persistence of DNA Adducts Formed by the Carcinogenic Air Pollutant 3-Nitrobenzanthrone in Target and Non-target Organs after Intratracheal Instillation in Rats. *Carcinogenesis* **2007**, *28* (5), 1117-1121;

(c) Landvik, N. E.; Arlt, V. M.; Nagy, E.; Solhaug, A.; Tekpli, X.; Schmeiser, H. H.; Refsnes, M.; Phillips, D. H.; Lagadic-Gossmann, D.; Holme, J. A., 3-Nitrobenzanthrone and 3-Aminobenzanthrone Induce DNA Damage and Cell Signalling in Hepa1c1c7 Cells. *Mut. Res.-Fund. Mol. M.* **2010**, *684* (1-2), 11-23; (d) Kawanishi, M.; Fujikawa, Y.; Ishii, H.; Nishida, H.; Higashigaki, Y.; Kanno, T.; Matsuda, T.; Takamura-Enya, T.; Yagi, T., Adduct Formation and Repair, and Translesion DNA Synthesis Across the Adducts in Human Cells Exposed to 3-Nitrobenzanthrone. *Mut. Res.-Gen. Tox. En*. **2013**, *753* (2), 93-100.

77. (a) Purohit, V.; Basu, A. K., Mutagenicity of Nitroaromatic Compounds. *Chem. Res. Toxicol*. **2000**, *13* (8), 673-692; (b) Adachi, S.; Kawamura, K.; Takemoto, K.; Suzuki, H.; Hisamatsu, Y., *Relationships Between Acute and Chronic Effects of Air Pollution*. ILSI Press, 2000; p 463; (c) Arlt, V. M.; Gingerich, J.; Schmeiser, H. H.; Phillips, D. H.; Douglas, G. R.; White, P. A., Genotoxicity of 3-Nitrobenzanthrone and 3-Aminobenzanthrone in Muta $^{(TM)}$ Mouse and Lung Epithelial Cells Derived from Muta$^{(TM)}$ Mouse. *Mutagenesis* **2008**, *23* (6), 483-490; (d) Gadkari, V. V.; Tokarsky, E. J.; Malik, C. K.; Basu, A. K.; Suo, Z. C., Mechanistic Investigation of the Bypass of a Bulky Aromatic DNA Adduct Catalyzed by a Y-family DNA polymerase. *DNA Repair* **2014**, *21*, 65-77.

78. Phousongphouang, P. T.; Grosovsky, A. J.; Eastmond, D. A.; Covarrubias, M.; Arey, J., The Genotoxicity of 3-Nitrobenzanthrone and the Nitropyrene Lactones in Human Lymphoblasts. *Mut. Res.-Gen. Tox. En*. **2000**, *472* (1-2), 93-103.

79. Lamy, E.; Kassie, F.; Gminski, R.; Schmeiser, H. H.; Mersch-Sundermann, V., 3-Nitrobenzanthrone (3-NBA) Induced Micronucleus Formation and DNA Damage in Human Hepatoma (HepG2) Cells. *Tox. Lett*. **2004**, *146* (2), 103-109.

80. Arlt, V. M.; Zhan, L.; Schmeiser, H. H.; Honma, M.; Hayashi, M.; Phillips, D. H.; Suzuki, T., DNA Adducts and Mutagenic Specificity of the Ubiquitous Environmental Pollutant 3-Nitrobenzanthrone in Muta Mouse. *Env. Mol. Mutagen*. **2004**, *43* (3), 186-195.

81. Hansen, T.; Seidel, A.; Borlak, J., The Environmental Carcinogen 3-Nitrobenzanthrone and its Main Metabolite 3-Aminobenzanthrone Enhance Formation of Reactive Oxygen Intermediates in Human A549 Lung Epithelial Cells. *Tox. Appl. Pharm*. **2007**, *221* (2), 222-234.

82. Seidel, A.; Dahmann, D.; Krekeler, H.; Jacob, J., Biomonitoring of Polycyclic Aromatic Compounds in the Urine of Mining Workers Occupationally Exposed to Diesel Exhaust. *Int. J. Hyg. Env. Health* **2002**, *204* (5-6), 333-338.

83. Benbrahim-Tallaa, L.; Baan, R. A.; Grosse, Y.; Lauby-Secretan, B.; El Ghissassi, F.; Bouvard, V.; Guha, N.; Loomis, D.; Straif, K.; Workin, I. A. R. C. M., Carcinogenicity of Diesel-engine and Gasoline-engine Exhausts and some Nitroarenes. *Lancet Oncol*. **2012**, *13* (7), 663-664.

84. Borlak, J.; Hansen, T.; Yuan, Z. X.; Sikka, H. C.; Kumar, S.; Schmidbauer, S.; Frank, H.; Jacob, J.; Seidel, A., Metabolism and DNA-binding of 3-Nitrobenzanthrone in Primary Rat Alveolar Type II Cells, in Human Fetal Bronchial, Rat Epithelial and Mesenchymal Cell Lines. *Polycyclic Aromat. Compd*. **2000**, *21* (1-4), 73-86.

85. (a) Gates, K. S.; Nooner, T.; Dutta, S., Biologically Relevant Chemical Reactions of N7-Alkylguanine Residues in DNA. *Chem. Res. Toxicol*. **2004**, *17* (7), 839-856; (b) Onozato,

M.; Ohshima, S., Analysis of Mutagenicity of Nitrobenzanthrones by Molecular Orbital Calculations. *Polycyclic Aromat. Compd*. **2006**, *26* (2), 93-101; (c) Yang, Z.-Z.; Qi, S.-F.; Zhao, D.-X.; Gong, L.-D., Insight into Mechanism of Formation of C8 Adducts in Carcinogenic Reactions of Arylnitrenium Ions with Purine Nucleosides. *J. Phys. Chem. B* **2009**, *113* (1), 254-259; (d) Mizerovska, J.; Dracinska, H.; Arlt, V. M.; Hudecek, J.; Hodek, P.; Schmeiser, H. H.; Frei, E.; Stiborova, M., Rat Cytochromes P450 Oxidize 3-Aminobenzanthrone, a Human Metabolite of the Carcinogenic Environmental Pollutant 3-Nitrobenzanthrone. *Interdiscip. Toxicol*. **2008**, *1* (2), 150-154; (e) Mizerovska, J.; Dracinska, H.; Frei, E.; Schmeiser, H. H.; Arlt, V. M.; Stiborova, M., Induction of Biotransformation Enzymes by the Carcinogenic Air-pollutant 3-Nitrobenzanthrone in Liver, Kidney and Lung, after Intra-tracheal Instillation in Rats. *Mut. Res.-Gen. Tox. En*. **2011**, *720* (1-2), 34-41; (f) Stiborova, M.; Cechova, T.; Borek-Dohalska, L.; Moserova, M.; Frei, E.; Schmeiser, H. H.; Paca, J.; Arlt, V. M., Activation and Detoxification Metabolism of Urban Air Pollutants 2-Nitrobenzanthrone and Carcinogenic 3-Nitrobenzanthrone by Rat and Mouse Hepatic Microsomes. *Neuro. Endocrinol. Lett*. **2012**, *33* Suppl 3, 8-15; (g) Xue, J. D.; Du, L. L.; Zhu, R. X.; Huang, J. Q.; Phillips, D. L., Direct Time-resolved Spectroscopic Observation of Arylnitrenium Ion Reactions with Guanine-containing DNA Oligomers. *J. Org. Chem*. **2014**, *79* (8), 3610-3614.

86. (a) Kawanishi, M.; Enya, T.; Suzuki, H.; Takebe, H.; Matsui, S.; Yagi, T., Postlabeling Analysis of DNA Adducts Formed in Human Hepatoma Cells Treated with 3-Nitrobenzanthrone. *Mutat. Res*. **2000**, *470* (2), 133-139; (b) Arlt, V. M.; Schmeiser, H. H.; Osborne, M. R.; Kawanishi, M.; Kanno, T.; Yagi, T.; Phillips, D. H.; Takamura-Enya, T., Identification of Three Major DNA Adducts Formed by the Carcinogenic Air

Pollutant 3-Nitrobenzanthrone in Rat Lung at the C8 and N2 Position of Guanine and at the N6 Position of Adenine. *Int. J. Cancer* **2006**, *118* (9), 2139-46; (c) Kanno, T.; Kawanishi, M.; Takamura-Enya, T.; Arlt, V. M.; Phillips, D. H.; Yagi, T., DNA Adduct Formation in Human Hepatoma Cells Treated with 3-Nitrobenzanthrone: Analysis by the 32P-Postlabeling Method. *Mutat. Res*. **2007**, *634* (1-2), 184-191.

87. Pande, P.; Malik, C. K.; Bose, A.; Jasti, V. P.; Basu, A. K., Mutational Analysis of the C8-Guanine Adduct of the Environmental Carcinogen 3-Nitrobenzanthrone in Human Cells: Critical Roles of DNA Polymerases Eta and Kappa and Rev1 in Error-prone Translesion Synthesis. *Biochemistry* **2014**, *53* (32), 5323-5331.

88. Nishida, H.; Kawanishi, M.; Takamura-Enya, T.; Yagi, T., Mutagenic Specificity of N-Acetoxy-3-aminobenzanthrone, a Major Metabolically Activated Form of 3-Nitrobenzanthrone, in Shuttle Vector Plasmids Propagated in Human Cells. *Mutat. Res*. **2008**, *654* (1), 82-87.

89. Lukin, M.; Zaliznyak, T.; Johnson, F.; de Los Santos, C. R., Incorporation of 3-Aminobenzanthrone into 2'-Deoxyoligonucleotides and its Impact on Duplex Stability. *J. Nucleic Acids* **2011**, *2011*, 521035.

90. Mao, B.; Vyas, R. R.; Hingerty, B. E.; Broyde, S.; Basu, A. K.; Patel, D. J., Solution Conformation of the N-(Deoxyguanosin-8-yl)-1-aminopyrene ([AP]dG) Adduct Opposite dC in a DNA Duplex. *Biochemistry* **1996**, *35* (39), 12659-12670.

91. O'Handley, S. F.; Sanford, D. G.; Xu, R.; Lester, C. C.; Hingerty, B. E.; Broyde, S.; Krugh, T. R., Structural Characterization of an N-Acetyl-2-aminofluorene (AAF) Modified DNA Oligomer by NMR, Energy Minimization, and Molecular Dynamics. *Biochemistry* **1993**, *32*, 2481-2497.

92. Mao, B.; Hingerty, B. E.; Broyde, S.; Patel, D. J., Solution Structure of the Aminofluorene [AF]-intercalated Conformer of the Syn-[AF]-C8-dG Adduct Opposite dC in a DNA Duplex. *Biochemistry* **1998**, *37* (1), 81-94.

93. Brown, K.; Hingerty, B. E.; Guenther, E. A.; Krishnan, V. V.; Broyde, S.; Turteltaub, K. W.; Cosman, M., Solution Structure of the 2-Amino-1-methyl-6-phenylimidazo[4,5-b]pyridine C8-Deoxyguanosine Adduct in Duplex DNA. *Proc. Natl. Acad. Sci. USA* **2001**, *98* (15), 8507-8512.

94. Wang, F.; Demuro, N. E.; Elmquist, C. E.; Stover, J. S.; Rizzo, C. J.; Stone, M. P., Base-displaced Intercalated Structure of the Food Mutagen 2-Amino-3-methylimidazo[4,5-f]quinoline in the Recognition Sequence of the NarI Restriction Enzyme, a Hotspot for -2 bp Deletions. *J. Am. Chem. Soc.* **2006**, *128* (31), 10085-10095.

95. Cavaluzzi, M. J.; Borer, P. N., Revised UV Extinction Coefficients for Nucleoside-5'-monophosphates and Unpaired DNA and RNA. *Nucleic Acids Res*. **2004**, *32* (1), e13.

96. Bernardi, G., Chromatography of Nucleic Acids on Hydroxyapatite. I. Chromatography of Native DNA. *Biochim. Biophys. Acta* **1969**, *174* (2), 423-434.

97. Piotto, M.; Saudek, V.; Sklenar, V., Gradient-tailored Excitation for Single-quantum NMR Spectroscopy of Aqueous Solutions. *J. Biomol. NMR* **1992**, *6*, 661-665.

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

99. (a) Jeener, J.; Meier, B. H.; Bachmann, P.; Ernst, R. R., Investigation of Exchange Processes by 2-dimensional NMR Spectroscopy. *J. Chem. Phys*. **1979**, *71* (11), 4546-4553; (b) Wagner, R.; Berger, S., Gradient-selected NOESY - A Fourfold Reduction of the Measurement Time for the NOESY Experiment. *J. Magn. Res. A* **1996**, *123* (1), 119-121.

100. (a) Keepers, J. W.; James, T. L., A Theoretical Study of Distance Determination from NMR. Two-dimensional Nuclear Overhauser Effect Spectra. *J. Magn. Reson*. **1984**, *57*, 404-426; (b) Borgias, B. A.; James, T. L., Two-dimensional Nuclear Overhauser Effect: Complete Relaxation Matrix Analysis. *Methods Enzymol*. **1989**, *176*, 169-183; (c) James, T. L., Relaxation Matrix Analysis of Two-dimensional Nuclear Overhauser Effect Spectra. *Curr. Opin. Struct. Biol*. **1991**, *1*, 1042-1053.

101. Borgias, B. A.; James, T. L., MARDIGRAS--A Procedure for Matrix Analysis of Relaxation for Discerning Geometry of an Aqueous Structure. *J. Magn. Reson*. **1990**, *87*, 475-487.

102. (a) Bruschweiler, R.; Blackledge, M.; Ernst, R. R., Multi-conformational Peptide Dynamics Derived from NMR Data: A New Search Algorithm and its Application to Antamanide. *J. Biomol. NMR* **1991**, *1* (1), 3-11; (b) Ippel, J. H.; Lanzotti, V.; Galeone, A.; Mayol, L.; van den Boogaart, J. E.; Pikkemaat, J. A.; Altona, C., Conformation of the Circular Dumbbell d<pCGC-TT-GCG-TT>: Structure Determination and Molecular Dynamics. *J. Biomol. NMR* **1995**, *6* (4), 403-422.

103. Arnott, S.; Hukins, D. W. L., Optimised Parameters for A-DNA and B-DNA. *Biochem. Biophys. Res. Comm*. **1972**, *47*, 1504-1509.

104. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi,

R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adarno, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Revision B.01*; Gaussian Inc.: Wallingford, CT, **2009**.

105. *Molecular Operating Environment (MOE) 2013.08*; Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7: 2015.

106. Schafmeister, C. E. A.; Ross, W. S.; Romanovski, V. XLEAP, *University of California, San Francisco: San Francisco,* **1995**.

107. Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The AMBER Biomolecular Simulation Programs. *J. Comput. Chem*. **2005**, *26* (16), 1668-1688.

108. Wang, J. M.; Cieplak, P.; Kollman, P. A., How Well does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comput. Chem*. **2000**, *21* (12), 1049-1074.

109. Bashford, D.; Case, D. A., Generalized Born Models of Macromolecular Solvation Effects. *Annu. Rev. Phys. Chem*. **2000**, *51*, 129-152.

110. Thomas, P. D.; Basus, V. J.; James, T. L., Protein Solution Structure Determination Using Distances from Two-dimensional Nuclear Overhauser Effect Experiments: Effect of Approximations on the Accuracy of Derived Structures. *Proc. Natl. Acad. Sci. USA* **1991**, *88* (4), 1237-1241.

111. (a) Patel, D. J.; Shapiro, L.; Hare, D., DNA and RNA: NMR Studies of Conformations and Dynamics in Solution. *Q. Rev. Biophys*. **1987**, *20* (1-2), 35-112; (b) Reid, B. R., Sequence-specific Assignments and Their Use in NMR Studies of DNA Structure. *Q. Rev. Biophys*. **1987**, *20*, 2-28.

112. Saenger, W., *Principles of Nucleic Acid Structure*. Springer: New York, 1984.

113. Mu, H.; Kropachev, K.; Wang, L.; Zhang, L.; Kolbanovskiy, A.; Kolbanovskiy, M.; Geacintov, N. E.; Broyde, S., Nucleotide Excision Repair of 2-Acetylaminofluorene- and 2-Aminofluorene-(C8)-guanine Adducts: Molecular Dynamics Simulations Elucidate how Lesion Structure and Base Sequence Context Impact Repair Efficiencies. *Nucleic Acids Res.* **2012**, *40* (19), 9675-9690.

114. (a) Koehl, P.; Valladier, P.; Lefevre, J. F.; Fuchs, R. P. P., Strong Structural Effect of the Position of a Single Acetylaminofluorene Adduct within a Mutation Hot Spot. *Nucleic Acids Res*. **1989**, *17*, 9531-9541; (b) Elmquist, C. E.; Wang, F.; Stover, J. S.; Stone, M. P.; Rizzo, C. J., Conformational Differences of the C8-Deoxyguanosine Adduct of 2-Amino-3-methylimidazo[4,5-f]quinoline (IQ) within the NarI Recognition Sequence. *Chem. Res. Toxicol*. **2007**, *20* (3), 445-454.

115. Nolan, S. J.; Vyas, R. R.; Hingerty, B. E.; Ellis, S.; Broyde, S.; Shapiro, R.; Basu, A. K., Solution Properties and Computational Analysis of an Oligodeoxynucleotide Containing N-(Deoxyguanosin-8-yl)-1-aminopyrene. *Carcinogenesis* **1996**, *17* (1), 133-44.

116. (a) Mao, B.; Hingerty, B. E.; Broyde, S.; Patel, D. J., Solution Structure of the Aminofluorene [AF]-external Conformer of the Anti-[AF]-C8-dG Adduct Opposite dC in a DNA Duplex. *Biochemistry* **1998**, *37* (1), 95-106; (b) Shapiro, R.; Ellis, S.; Hingerty, B. E.; Broyde, S., Effect of Ring Size on Conformations of Aromatic Amine-DNA Adducts:

the Aniline-C8 Guanine Adduct Resides in the B-DNA Major Groove. *Chem. Res. Toxicol*. **1998**, *11* (4), 335-341.

117. Pedretti, A.; Villa, L.; Vistoli, G., VEGA--An Open Platform to Develop Chemo-bio-informatics Applications, Using Plug-in Architecture and Script Programming. *J. Comput. Aided Mol. Des*. **2004**, *18* (3), 167-173.

118. Wang, F.; Elmquist, C. E.; Stover, J. S.; Rizzo, C. J.; Stone, M. P., DNA Sequence Modulates the Conformation of the Food Mutagen 2-Amino-3-methylimidazo[4,5-f]quinoline in the Recognition Sequence of the NarI Restriction Enzyme. *Biochemistry* **2007**, *46* (29), 8498-8516.

119. Turesky, R. J.; Markovic, J.; Aeschlimann, J. M., Formation and Differential Removal of C-8 and N2-Guanine Adducts of the Food Carcinogen 2-Amino-3-methylimidazo[4,5-f]quinoline in the Liver, Kidney, and Colorectum of the Rat. *Chem. Res. Toxicol*. **1996**, *9* (2), 397-402.

120. Stavros, K. M.; Hawkins, E. K.; Rizzo, C. J.; Stone, M. P., Base-displaced Intercalation of the 2-Amino-3-methylimidazo[4,5-f]quinolone N2-dG Adduct in the NarI DNA Recognition Sequence. *Nucleic Acids Res*. **2014**, *42* (5), 3450-63.

121. Stavros, K. M.; Hawkins, E. K.; Rizzo, C. J.; Stone, M. P., Base-displaced Intercalated Conformation of the 2-Amino-3-methylimidazo[4,5-f]quinoline N2-dG DNA Adduct Positioned at the Nonreiterated G1 in the NarI Restriction site. *Chem. Res. Toxicol*. **2015**, *28* (7), 1455-1468.

122. Elmquist, C. E.; Stover, J. S.; Wang, Z.; Rizzo, C. J., Site-specific Synthesis and Properties of Oligonucleotides Containing C8-Deoxyguanosine Adducts of the Dietary Mutagen IQ. *J. Am. Chem. Soc*. **2004**, *126* (36), 11189-11201.

123. Poirier, M. C.; Beland, F. A., DNA Adduct Measurements and Tumor Incidence During Chronic Carcinogen Exposure in Animal Models: Implications for DNA Adduct-based Human Cancer Risk Assessment. *Chem. Res. Toxicol*. **1992**, *5* (6), 749-755.

124. (a) Fuchs, R. P.; Schwartz, N.; Daune, M. P., Hot Spots of Frameshift Mutations Induced by the Ultimate Carcinogen N-Acetoxy-N-2-acetylaminofluorene. *Nature* **1981**, *294* (5842), 657-9; (b) Koffel-Schwartz, N.; Verdier, J. M.; Bichara, M.; Freund, A. M.; Daune, M. P.; Fuchs, R. P. P., Carcinogen-induced Mutation Spectrum in Wild-type, uvrA and umuC Strains of Escherichia coli. *J. Mol. Biol*. **1984**, *177*, 33-51; (c) Koehl, P.; Burnouf, D.; Fuchs, R. P. P., Construction of Plasmids Containing a Unique Acetylaminofluorene Adduct Located within a Mutation Hot Spot. A New Probe for Frameshift Mutagenesis. *J. Mol. Biol*. **1989**, *207*, 355-364; (d) Koehl, P.; Valladier, P.; Lefevre, J. F.; Fuchs, R. P., Strong Structural Effect of the Position of a Single Acetylaminofluorene Adduct within a Mutation Hot Spot. *Nucleic Acids Res*. **1989**, *17* (23), 9531-9541; (e) Hoffmann, G. R.; Fuchs, R. P., Mechanisms of Frameshift Mutations: Insight from Aromatic Amines. *Chem. Res. Toxicol*. **1997**, *10* (4), 347-359; (f) Broschard, T. H.; Koffel-Schwartz, N.; Fuchs, R. P., Sequence-dependent Modulation of Frameshift Mutagenesis at NarI-derived Mutation Hot Spots. *J. Mol. Biol*. **1999**, *288* (1), 191-199.

125. (a) Jain, V.; Hilton, B.; Patnaik, S.; Zou, Y.; Chiarelli, M. P.; Cho, B. P., Conformational and Thermodynamic Properties Modulate the Nucleotide Excision Repair of 2-Aminofluorene and 2-Acetylaminofluorene dG Adducts in the NarI Sequence. *Nucleic Acids Res*. **2012**, *40* (9), 3939-3951; (b) Jain, V.; Hilton, B.; Lin, B.; Patnaik, S.; Liang, F.; Darian, E.; Zou, Y.; Mackerell, A. D., Jr.; Cho, B. P., Unusual Sequence Effects on Nucleotide Excision Repair of Arylamine Lesions: DNA Bending/Distortion as a Primary

Recognition Factor. *Nucleic Acids Res*. **2013**, *41* (2), 869-880; (c) Jain, V.; Vaidyanathan, V. G.; Patnaik, S.; Gopal, S.; Cho, B. P., Conformational Insights into the Lesion and Sequence Effects for Arylamine-induced Translesion DNA Synthesis: 19F NMR, Surface Plasmon Resonance, and Primer Kinetic Studies. *Biochemistry* **2014**, *53* (24), 4059-4071.

126. (a) Kawanishi, M.; Fujikawa, Y.; Ishii, H.; Nishida, H.; Higashigaki, Y.; Kanno, T.; Matsuda, T.; Takamura-Enya, T.; Yagi, T., Adduct Formation and Repair, and Translesion DNA Synthesis Across the Adducts in Human Cells Exposed to 3-Nitrobenzanthrone. *Mutat. Res., Genet. Toxicol. Environ. Mutagen*. **2013**, *753* (2), 93-100; (b) Sale, J. E.; Lehmann, A. R.; Woodgate, R., Y-family DNA Polymerases and Their Role in Tolerance of Cellular DNA Damage. *Nat. Rev. Mol. Cell Biol*. **2012**, *13* (3), 141-152.

127. Lehmann, A. R.; Niimi, A.; Ogi, T.; Brown, S.; Sabbioneda, S.; Wing, J. F.; Kannouche, P. L.; Green, C. M., Translesion Synthesis: Y-family Polymerases and the Polymerase Switch. *DNA Repair* **2007**, *6* (7), 891-899.

128. Shachar, S.; Ziv, O.; Avkin, S.; Adar, S.; Wittschieben, J.; Reißner, T.; Chaney, S.; Friedberg, E. C.; Wang, Z.; Carell, T., Two-polymerase Mechanisms Dictate Error-free and Error-prone Translesion DNA Synthesis in Mammals. *EMBO J.* **2009**, *28* (4), 383-393.

129. Johnson, R. E.; Washington, M. T.; Haracska, L.; Prakash, S.; Prakash, L., Eukaryotic Polymerases ι and ζ Act Sequentially to Bypass DNA Lesions. *Nature* **2000**, *406* (6799), 1015-1019.

130. Woodgate, R., A Plethora of Lesion-replicating DNA Polymerases. *Genes Dev.* **1999**, *13* (17), 2191-2195.

131. Johnson, R. E.; Prakash, S.; Prakash, L., Efficient Bypass of a Thymine-thymine Dimer by Yeast DNA Polymerase, Polη. *Science* **1999**, *283* (5404), 1001-1004.

132. Biertümpfel, C.; Zhao, Y.; Kondo, Y.; Ramón-Maiques, S.; Gregory, M.; Lee, J. Y.; Masutani, C.; Lehmann, A. R.; Hanaoka, F.; Yang, W., Structure and Mechanism of Human DNA Polymerase [eegr]. *Nature* **2010**, *465* (7301), 1044-1048.

133. Nair, D. T.; Johnson, R. E.; Prakash, S.; Prakash, L.; Aggarwal, A. K., Replication by Human DNA Polymerase-ι Occurs by Hoogsteen Base-pairing. *Nature* **2004**, *430* (6997), 377-380.

134. Ohashi, E.; Ogi, T.; Kusumoto, R.; Iwai, S.; Masutani, C.; Hanaoka, F.; Ohmori, H., Error-prone Bypass of Certain DNA Lesions by the Human DNA Polymerase κ. *Genes Dev*. **2000**, *14* (13), 1589-1594.

135. Ohashi, E.; Bebenek, K.; Matsuda, T.; Feaver, W. J.; Gerlach, V. L.; Friedberg, E. C.; Ohmori, H.; Kunkel, T. A., Fidelity and Processivity of DNA Synthesis by DNA Polymerase κ, the Product of the Human DINB1 Gene. *J. Biol. Chem*. **2000**, *275* (50), 39678-39684.

136. Nelson, J. R.; Lawrence, C. W.; Hinkle, D. C., Deoxycytidyl Transferase Activity of Yeast REV1 Protein. *Nature* **1996**, *382* (6593), 729-731.

137. Ross, A.-L.; Simpson, L. J.; Sale, J. E., Vertebrate DNA Damage Tolerance Requires the C-Terminus but not BRCT or Transferase Domains of REV1. *Nucleic Acids Res*. **2005**, *33* (4), 1280-1289.

138. Ziv, O.; Geacintov, N.; Nakajima, S.; Yasui, A.; Livneh, Z., DNA Polymerase ζ Cooperates with Polymerases κ and ι in Translesion DNA Synthesis Across Pyrimidine Photodimers in Cells from XPV Patients. *Proc. Natl. Acad. Sci. USA* **2009**, *106* (28), 11552-11557.

139. (a) Rechkoblit, O.; Kolbanovskiy, A.; Malinina, L.; Geacintov, N. E.; Broyde, S.; Patel, D. J., Mechanism of Error-free and Semitargeted Mutagenic Bypass of an Aromatic Amine Lesion by Y-family Polymerase Dpo4. *Nat. Struct. Mol. Biol*. **2010**, *17* (3), 379-388; (b) Banerjee, S.; Brown, K. L.; Egli, M.; Stone, M. P., Bypass of Aflatoxin B1 Adducts by the Sulfolobus solfataricus DNA Polymerase IV. *J. Am. Chem. Soc*. **2011**, *133* (32), 12556-12568; (c) Banerjee, S.; Christov, P. P.; Kozekova, A.; Rizzo, C. J.; Egli, M.; Stone, M. P., Replication Bypass of the trans-4-Hydroxynonenal-Derived (6 S, 8 R, 11 S)-1, N 2-Deoxyguanosine DNA Adduct by the Sulfolobus solfataricus DNA Polymerase IV. *Chem. Res. Toxicol*. **2012**, *25* (2), 422-435; (d) Kirouac, K. N.; Basu, A. K.; Ling, H., Structural Mechanism of Replication Stalling on a Bulky Amino-polycyclic Aromatic Hydrocarbon DNA Adduct by a Y-family DNA Polymerase. *J. Mol. Biol.* **2013**, *425* (22), 4167-4176.

140. Ling, H.; Boudsocq, F.; Woodgate, R.; Yang, W., Crystal Structure of a Y-family DNA Polymerase in Action: a Mechanism for Error-prone and Lesion-bypass Replication. *Cell* **2001**, *107* (1), 91-102.

141. (a) Donny-Clark, K.; Shapiro, R.; Broyde, S., Accommodation of an N-(Deoxyguanosin-8-yl)-2-acetylaminofluorene Adduct in the Active Site of Human DNA Polymerase ι: Hoogsteen or Watson− Crick Base Pairing?†. *Biochemistry* **2008**, *48* (1), 7-18; (b) Irimia, A.; Eoff, R. L.; Guengerich, F. P.; Egli, M., Structural and Functional Elucidation of the Mechanism Promoting Error-prone Synthesis by Human DNA Polymerase κ Opposite the 7, 8-Dihydro-8-oxo-2′-deoxyguanosine Adduct. *J. Biol. Chem*. **2009**, *284* (33), 22467-22480.

142. Patra, A.; Nagy, L. D.; Zhang, Q.; Su, Y.; Müller, L.; Guengerich, F. P.; Egli, M., Kinetics, Structure, and Mechanism of 8-Oxo-7, 8-dihydro-2′-deoxyguanosine Bypass by Human DNA Polymerase η. *J. Biol. Chem*. **2014**, *289* (24), 16867-16882.

143. (a) Paputa-Peck, M.; Marano, R.; Schuetzle, D.; Riley, T.; Hampton, C.; Prater, T.; Skewes, L.; Jensen, T.; Ruehle, P., Determination of Nitrated Polynuclear Aromatic Hydrocarbons in Particulate Extracts by using Capillary Column Gas Chromatography with Nitrogen Selective Detection. *Anal. Chem*. **1983**, *55* (12), 1946-1954; (b) Schuetzle, D.; Lee, F. S.-C.; Prater, T. J.; Tejada, S. B., The Identification of Polynuclear Aromatic Hydrocarbon (PAH) Derivatives in Mutagenic Fractions of Diesel Particulate Extracts. *Int. J. Environ. Anal. Chem.* **1981**, *9* (2), 93-144.

144. Bieler, C.A.; Cornelius, M.G.; Klein, R.; Arlt, V.M.; Wiessler, M.; Phillips, D.H.; Schmeiser, H.H., DNA Adduct Formation by the Environmental Contaminant 3-Nitrobenzanthrone after Intratracheal Instillation in Rats. *Int. J. Cancer* **2005**, *116* (6), pp.833-838.

145. (a) Arlt, V. M.; Gingerich, J.; Schmeiser, H. H.; Phillips, D. H.; Douglas, G. R.; White, P. A., Genotoxicity of 3-Nitrobenzanthrone and 3-Aminobenzanthrone in Muta™ Mouse and Lung Epithelial Cells Derived from Muta™ Mouse. *Mutagenesis* **2008**, *23* (6), 483-490; (b) Landvik, N.; Arlt, V.; Nagy, E.; Solhaug, A.; Tekpli, X.; Schmeiser, H.; Refsnes, M.; Phillips, D.; Lagadic-Gossmann, D.; Holme, J., 3-Nitrobenzanthrone and 3-Aminobenzanthrone Induce DNA Damage and Cell Signalling in Hepa1c1c7 Cells. *Mutat. Res., Fundam. Mol. Mech. Mutagen*. **2010**, *684* (1), 11-23; (c) Phousongphouang, P. T.; Grosovsky, A. J.; Eastmond, D. A.; Covarrubias, M.; Arey, J., The Genotoxicity of 3-Nitrobenzanthrone and the Nitropyrene Lactones in Human Lymphoblasts. *Mutat. Res.,*

*Genet. Toxicol. Environ. Mutagen*. **2000**, *472* (1), 93-103; (d) Lamy, E.; Kassie, F.; Gminski, R.; Schmeiser, H. H.; Mersch-Sundermann, V., 3-Nitrobenzanthrone (3-NBA) Induced Micronucleus Formation and DNA Damage in Human Hepatoma (HepG2) Cells. *Toxicol. Lett*. **2004**, *146* (2), 103-109; (e) Hansen, T.; Seidel, A.; Borlak, J., The Environmental Carcinogen 3-Nitrobenzanthrone and its Main Metabolite 3-Aminobenzanthrone Enhance Formation of Reactive Oxygen Intermediates in Human A549 Lung Epithelial Cells. *Toxicol. Appl. Pharmacol*. **2007**, *221* (2), 222-234.

146. (a) Borlak, J.; Hansen, T.; Yuan, Z.-X.; Sikka, H. C.; Kumar, S.; Schmidbauer, S.; Frank, H.; Jacob, J.; Seidel, A., Metabolism and DNA-binding of 3-Nitrobenzanthrone in Primary Rat Alveolar Type II Cells, in Human Fetal Bronchial, Rat Epithelial and Mesenchymal Cell Lines. *Polycyclic Aromat. Compd*. **2000**, *21* (1-4), 73-86; (b) Seidel, A.; Dahmann, D.; Krekeler, H.; Jacob, J., Biomonitoring of Polycyclic Aromatic Compounds in the Urine of Mining Workers Occupationally Exposed to Diesel Exhaust. *Int. J. Hyg. Environ. Health* **2002**, *204* (5), 333-338.

147. (a) Arlt, V. M.; Phillips, D. H.; Reynisson, J., Theoretical Investigations on the Formation of Nitrobenzanthrone-DNA Adducts. *Org. Biomol. Chem*. **2011**, *9* (17), 6100-6110; (b) Yang, Z.-Z.; Qi, S.-F.; Zhao, D.-X.; Gong, L.-D., Insight into Mechanism of Formation of C8 Adducts in Carcinogenic Reactions of Arylnitrenium Ions with Purine Nucleosides. *J. Phys. Chem. B* **2008**, *113* (1), 254-259.

148. Arlt, V. M.; Sorg, B. L.; Osborne, M.; Hewer, A.; Seidel, A.; Schmeiser, H. H.; Phillips, D. H., DNA Adduct Formation by the Ubiquitous Environmental Pollutant 3-Nitrobenzanthrone and its Metabolites in Rats. *Biochem. Biophys. Res. Commun*. **2003**, *300* (1), 107-114.

149. Prakash, S.; Johnson, R. E.; Prakash, L., Eukaryotic Translesion Synthesis DNA Polymerases: Specificity of Structure and Function. *Annu. Rev. Biochem*. **2005**, *74*, 317-353.

150. Politica, D. A.; Malik, C. K.; Basu, A. K.; Stone, M. P., Base-Displaced Intercalated Structure of the N-(2′-Deoxyguanosin-8-yl)-3-aminobenzanthrone DNA Adduct. *Chem. Res. Toxicol.* **2015**, *28* (12), 2253-2266.

151. O'Flaherty, D.; Patra, A.; Su, Y.; Guengerich, F.; Egli, M.; Wilds, C., Lesion Orientation of O 4-Alkylthymidine Influences Replication by Human DNA Polymerase η. *Chem. Sci*. [Online early access]. DOI: 10.1039/C6SC00666C Published online Apr 26, 2016. http://pubs.rsc.org/en/Content/ArticleLanding/2016/SC/C6SC00666C#!divAbstract (accessed June 15, 2016) (accepted manuscript, has not undergone final copyediting, typesetting, or proof review).

152. Gadkari, V. V.; Tokarsky, E. J.; Malik, C. K.; Basu, A. K.; Suo, Z., Mechanistic Investigation of the Bypass of a Bulky Aromatic DNA Adduct Catalyzed by a Y-family DNA Polymerase. *DNA repair* **2014**, *21*, 65-77.

153. Minor, W.; Cymborowski, M.; Otwinowski, Z.; Chruszcz, M., HKL-3000: The Integration of Data Reduction and Structure Solution–From Diffraction Images to an Initial Model in Minutes. *Acta Crystallogr., Sect. D: Biol. Crystallogr*. **2006**, *62* (8), 859-866.

154. (a) Vagin, A.; Teplyakov, A., MOLREP: An Automated Program for Molecular Replacement. *J. Appl. Crystallogr*. **1997**, *30* (6), 1022-1025; (b) Collaborative, C. P., The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1994**, *50* (Pt 5), 760.

155. Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W., PHENIX: A Comprehensive Python-based System for Macromolecular Structure Solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66* (2), 213-221.

156. Emsley, P.; Cowtan, K., Coot: Model-building Tools for Molecular Graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr*. **2004**, *60* (12), 2126-2132.

157. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem*. **2004**, *25* (13), 160d5-1612.

158. Lukin, M.; de los Santos, C., NMR Structures of amaged DNA. *Chem. Rev. (Washington, DC, U. S.)* **2006**, *106* (2), 607-686.

# Appendix

A I: Supporting Information for Chapter II

Reproduced with permission from Politica, D. A.; Malik, C. K.; Basu, A. K.; Stone, M. P., Base-Displaced Intercalated Structure of the N-(2′-Deoxyguanosin-8-yl)-3-aminobenzanthrone DNA Adduct. *Chemical Research in Toxicology* **2015,** *28* (12), 2253-2266. Copyright 2015 American Chemical Society

**Table A1.** Pseudorotation angle restraints used in AMBER structural refinement calculations.

| **ucleotide** | **Angle** | **Lower (º)** | **Upper (º)** |
|---|---|---|---|
| $T^2$ | $\nu_0$ | -44.7 | -14.7 |
| $T^2$ | $\nu_1$ | 18.1 | 48.1 |
| $T^2$ | $\nu_2$ | -37.2 | -6.7 |
| $T^2$ | $\nu_3$ | -16.9 | 24.2 |
| $T^2$ | $\nu_4$ | -1.9 | 34.0 |
| $G^3$ | $\nu_0$ | -43.9 | -13.9 |
| $G^3$ | $\nu_1$ | 22.2 | 52.2 |
| $G^3$ | $\nu_2$ | -44.6 | -14.6 |
| $G^3$ | $\nu_3$ | -3.2 | 26.8 |
| $G^3$ | $\nu_4$ | -4.4 | 25.6 |
| $G^7$ | $\nu_0$ | -43.9 | -13.9 |
| $G^7$ | $\nu_1$ | 22.2 | 52.2 |

| | | | |
|---|---|---|---|
| $G^7$ | $\nu_2$ | -44.6 | -14.6 |
| $G^7$ | $\nu_3$ | -3.2 | 26.8 |
| $G^7$ | $\nu_4$ | -4.4 | 25.6 |
| $T^8$ | $\nu_0$ | -52.1 | -22.1 |
| $T^8$ | $\nu_1$ | 15.0 | 45.0 |
| $T^8$ | $\nu_2$ | -27.4 | 2.6 |
| $T^8$ | $\nu_3$ | -25.0 | 5.0 |
| $T^8$ | $\nu_4$ | 13.5 | 43.5 |
| $T^9$ | $\nu_0$ | -52.1 | -22.1 |
| $T^9$ | $\nu_1$ | 15.0 | 45.0 |
| $T^9$ | $\nu_2$ | -27.4 | 2.6 |
| $T^9$ | $\nu_3$ | -25.0 | 5.0 |
| $T^9$ | $\nu_4$ | 13.5 | 43.5 |
| $T^{10}$ | $\nu_0$ | -52.1 | -22.1 |
| $T^{10}$ | $\nu_1$ | 15.0 | 45.0 |
| $T^{10}$ | $\nu_2$ | -27.4 | 2.6 |
| $T^{10}$ | $\nu_3$ | -25.0 | 5.0 |
| $T^{10}$ | $\nu_4$ | 13.5 | 43.5 |
| $G^{11}$ | $\nu_0$ | -44.7 | -14.7 |
| $G^{11}$ | $\nu_1$ | 18.1 | 48.1 |
| $G^{11}$ | $\nu_2$ | -37.2 | -6.7 |
| $G^{11}$ | $\nu_3$ | -16.9 | 24.2 |

| | | | |
|---|---|---|---|
| $G^{11}$ | $\nu_4$ | -1.9 | 34.0 |
| | | | |
| $C^{14}$ | $\nu_0$ | -44.7 | -14.7 |
| $C^{14}$ | $\nu_1$ | 18.1 | 48.1 |
| $C^{14}$ | $\nu_2$ | -37.2 | -6.7 |
| $C^{14}$ | $\nu_3$ | -16.9 | 24.2 |
| $C^{14}$ | $\nu_4$ | -1.9 | 34.0 |
| $A^{15}$ | $\nu_0$ | -43.9 | -13.9 |
| $A^{15}$ | $\nu_1$ | 22.2 | 52.2 |
| $A^{15}$ | $\nu_2$ | -44.6 | -14.6 |
| $A^{15}$ | $\nu_3$ | -3.2 | 26.8 |
| $A^{15}$ | $\nu_4$ | -4.4 | 25.6 |
| $A^{16}$ | $\nu_0$ | -43.9 | -13.9 |
| $A^{16}$ | $\nu_1$ | 22.2 | 52.2 |
| $A^{16}$ | $\nu_2$ | -44.6 | -14.6 |
| $A^{16}$ | $\nu_3$ | -3.2 | 26.8 |
| $A^{16}$ | $\nu_4$ | -4.4 | 25.6 |
| $A^{17}$ | $\nu_0$ | -43.9 | -13.9 |
| $A^{17}$ | $\nu_1$ | 22.2 | 52.2 |
| $A^{17}$ | $\nu_2$ | -44.6 | -14.6 |
| $A^{17}$ | $\nu_3$ | -3.2 | 26.8 |
| $A^{17}$ | $\nu_4$ | -4.4 | 25.6 |

| | | | |
|---|---|---|---|
| $C^{18}$ | $\nu_0$ | -52.1 | -22.1 |
| $C^{18}$ | $\nu_1$ | 15.0 | 45.0 |
| $C^{18}$ | $\nu_2$ | -27.4 | 2.6 |
| $C^{18}$ | $\nu_3$ | -25.0 | 5.0 |
| $C^{18}$ | $\nu_4$ | 13.5 | 43.5 |
| $G^{21}$ | $\nu_0$ | -43.9 | -13.9 |
| $G^{21}$ | $\nu_1$ | 22.2 | 52.2 |
| $G^{21}$ | $\nu_2$ | -44.6 | -14.6 |
| $G^{21}$ | $\nu_3$ | -3.2 | 26.8 |
| $G^{21}$ | $\nu_4$ | -4.4 | 25.6 |
| $C^{22}$ | $\nu_0$ | -52.1 | -22.1 |
| $C^{22}$ | $\nu_1$ | 15.0 | 45.0 |
| $C^{22}$ | $\nu_2$ | -27.4 | 2.6 |
| $C^{22}$ | $\nu_3$ | -25.0 | 5.0 |
| $C^{22}$ | $\nu_4$ | 13.5 | 43.5 |
| $A^{23}$ | $\nu_0$ | -44.7 | -14.7 |
| $A^{23}$ | $\nu_1$ | 18.1 | 48.1 |
| $A^{23}$ | $\nu_2$ | -37.2 | -6.7 |
| $A^{23}$ | $\nu_3$ | -16.9 | 24.2 |
| $A^{23}$ | $\nu_4$ | -1.9 | 34.0 |

**Table A2.** Chemical Shifts (ppm) of the Non-exchangeable DNA Protons for the 5'-d(GTGCXTGTTTGT)-3':5'-d(ACAAACACGCAC)-3' Duplex; X = C8-dG-ABA Adduct. The temperature was 15 ºC. NA = not applicable; this proton does not exist in this nucleotide.

| | **Non-exchangeable Protons** | | | | | |
|---|---|---|---|---|---|---|
| **Nucleotide** | **H8/H6** | **H5/H2/$CH_3$** | **H1'** | **H2'** | **H2"** | **H3'** |
| $G^{1}$ | 8.00 | NA | 6.04 | 2.71 | 2.83 | 4.84 |
| $T^{2}$ | 7.42 | 1.41 | 6.01 | 2.24 | 2.62 | 4.95 |
| $G^{3}$ | 7.88 | NA | 5.92 | 2.58 | 2.69 | 5.01 |
| $C^{4}$ | 7.00 | 5.30 | 5.62 | 0.84 | 2.06 | 4.26 |
| $X^{5}$ | NA | NA | 5.92 | 3.50 | 2.68 | 5.11 |
| $T^{6}$ | 7.44 | 1.39 | 5.57 | 2.25 | 2.48 | 4.92 |
| $G^{7}$ | 7.85 | NA | 6.00 | 2.64 | 2.82 | 4.99 |
| $T^{8}$ | 7.26 | 1.33 | 6.02 | 2.14 | 2.58 | 4.88 |
| $T^{9}$ | 7.46 | 1.60 | 6.11 | 2.16 | 2.60 | 4.91 |
| $T^{10}$ | 7.32 | 1.71 | 5.84 | 2.00 | 2.39 | 4.92 |
| $G^{11}$ | 7.99 | NA | 6.06 | 2.72 | 2.73 | 5.01 |
| $T^{12}$ | 7.44 | 1.63 | 6.26 | 2.26 | 2.26 | 4.56 |
| $A^{13}$ | 8.21 | 7.99 | 6.19 | 2.61 | 2.77 | 4.84 |
| $C^{14}$ | 7.41 | 5.47 | 5.12 | 2.01 | 2.23 | 4.79 |
| $A^{15}$ | 8.21 | 7.27 | 5.76 | 2.74 | 2.84 | 5.04 |
| $A^{16}$ | 8.10 | 7.13 | 5.85 | 2.62 | 2.83 | 5.05 |

| | | | | | | |
|---|---|---|---|---|---|---|
| $A^{17}$ | 8.00 | 7.49 | 5.97 | 2.45 | 2.75 | 4.98 |
| $C^{18}$ | 6.96 | 5.01 | 5.34 | 1.57 | 2.01 | 4.70 |
| $A^{19}$ | 7.32 | 7.86 | 5.85 | 2.11 | 2.35 | 4.93 |
| $C^{20}$ | 8.07 | 6.26 | 6.60 | 2.45 | 2.59 | 5.10 |
| $G^{21}$ | 7.35 | NA | 5.58 | 2.39 | 2.50 | 4.57 |
| $C^{22}$ | 7.35 | 5.16 | 5.50 | 2.02 | 2.34 | 4.74 |
| $A^{23}$ | 8.21 | 7.855 | 6.22 | 2.61 | 2.84 | 4.96 |
| $C^{24}$ | 7.37 | 5.37 | 6.09 | 2.09 | 2.12 | 4.47 |

**Table A3.** Chemical Shifts (ppm) of the DNA Protons for the unmodified 5'-d(GTGCGTGTTTGT)-3':5'-d(ACAAACACGCAC)-3' Duplex. The temperature was 15 ºC. NA = not applicable; this proton does not exist in this nucleotide.

| | **Proton** | | | | | |
|---|---|---|---|---|---|---|
| **Nucleotide** | **H8/H6** | **H5/H2/$CH_3$** | **H1'** | **H2'** | **H2"** | **H3'** |
| $G^{1}$ | 7.80 | NA | 6.04 | 2.69 | 2.83 | 4.85 |
| $T^{2}$ | 7.41 | 1.39 | 5.95 | 2.25 | 2.60 | 4.94 |
| $G^{3}$ | 7.90 | NA | 5.93 | 2.64 | 2.75 | 5.02 |
| $C^{4}$ | 7.32 | 5.316 | 5.71 | 2.06 | 2.42 | 4.86 |
| $G^{5}$ | 7.90 | NA | 5.98 | 2.63 | 2.80 | 4.99 |
| $T^{6}$ | 7.18 | 1.48 | 5.83 | 2.15 | 2.53 | 4.89 |
| $G^{7}$ | 7.85 | NA | 5.99 | 2.62 | 2.82 | 4.96 |
| $T^{8}$ | 7.28 | 1.32 | 6.05 | 2.17 | 2.61 | 4.89 |
| $T^{9}$ | 7.49 | 1.62 | 6.15 | 2.19 | 2.63 | 4.92 |
| $T^{10}$ | 7.34 | 1.73 | 5.87 | 2.02 | 2.41 | 4.92 |
| $G^{11}$ | 8.01 | NA | 6.08 | 2.72 | 2.74 | 5.03 |
| $T^{12}$ | 7.44 | 1.64 | 6.28 | 2.27 | 2.27 | 4.57 |
| $A^{13}$ | 8.22 | 8.01 | 6.21 | 2.62 | 2.79 | 4.85 |
| $C^{14}$ | 7.42 | 5.48 | 5.13 | 2.03 | 2.24 | 4.80 |
| $A^{15}$ | 8.23 | Not Assigned | 5.78 | 2.77 | 2.86 | 5.06 |
| $A^{16}$ | 8.14 | Not Assigned | 5.90 | 2.67 | 2.87 | 5.08 |

| | | | | | | |
|---|---|---|---|---|---|---|
| $A^{17}$ | 8.07 | 7.56 | 6.07 | 2.54 | 2.87 | 5.01 |
| $C^{18}$ | 7.14 | 5.16 | 5.50 | 1.87 | 2.31 | 4.77 |
| $A^{19}$ | 8.16 | 7.61 | 6.147 | 2.64 | 2.85 | 5.00 |
| $C^{20}$ | 7.20 | 5.19 | 5.55 | 1.95 | 2.31 | 4.82 |
| $G^{21}$ | 7.84 | NA | 5.86 | 2.62 | 2.72 | 4.98 |
| $C^{22}$ | 7.36 | 5.42 | 5.62 | 2.04 | 2.40 | 4.84 |
| $A^{23}$ | 8.28 | 7.86 | 6.27 | 2.69 | 2.90 | 5.02 |
| $C^{24}$ | 7.39 | 5.43 | 6.10 | 2.10 | 2.14 | 4.50 |

**Table A4.** Chemical Shifts (ppm) of the Exchangeable DNA Protons for the 5'-d(GTGCXTGTTTGT)-3':5'-d(ACAAACACGCAC)-3' Duplex; X = C8-dG-ABA Adduct. The temperature was 15 ºC. NA = not applicable; this proton does not exist in this nucleotide.

| | **Exchangeable Protons** | | |
|---|---|---|---|
| **Nucleotide** | **N1H/N3H** | $N^4H_a$ | $N^4H_b$ |
| $G^1$ | Not Assigned | NA | NA |
| $T^2$ | Not Assigned | NA | NA |
| $G^3$ | 12.57 | NA | NA |
| $C^4$ | NA | 6.69 | 7.97 |
| $X^5$ | Not Assigned | NA | NA |
| $T^6$ | 12.95 | NA | NA |
| $G^7$ | 12.27 | NA | NA |
| $T^8$ | 13.86 | NA | NA |
| $T^9$ | 13.93 | NA | NA |
| $T^{10}$ | 13.87 | NA | NA |
| $G^{11}$ | 12.64 | NA | NA |
| $T^{12}$ | Not Assigned | NA | NA |
| $A^{13}$ | NA | NA | NA |
| $C^{14}$ | NA | 6.65 | 8.32 |
| $A^{15}$ | NA | NA | NA |
| $A^{16}$ | NA | NA | NA |
| $A^{17}$ | NA | NA | NA |

| | | | |
|---|---|---|---|
| $C^{18}$ | NA | 6.28 | 7.73 |
| $A^{19}$ | NA | NA | NA |
| $C^{20}$ | NA | Not Assigned | Not Assigned |
| $G^{21}$ | 12.12 | NA | NA |
| $C^{22}$ | NA | 6.54 | 8.29 |
| $A^{23}$ | NA | NA | NA |
| $C^{24}$ | NA | Not Assigned | Not Assigned |

**Table A5.** Chemical Shifts (ppm) of the Exchangeable DNA Protons for the unmodified 5'-d(GTGCGTGTTTGT)-3':5'-d(ACAAACACGCAC)-3' Duplex. The temperature was 15 ºC. NA = not applicable; this proton does not exist in this nucleotide.

| | **Exchangeable Protons** | | |
|---|---|---|---|
| **Nucleotide** | **N1H/N3H** | $N^4H_a$ | $N^4H_b$ |
| $G^1$ | Not Assigned | NA | NA |
| $T^2$ | 13.79 | NA | NA |
| $G^3$ | 12.74 | NA | NA |
| $C^4$ | NA | 6.39 | 8.27 |
| $G^5$ | 12.78 | NA | NA |
| $T^6$ | 13.56 | NA | NA |
| $G^7$ | 12.50 | NA | NA |
| $T^8$ | 13.95 | NA | NA |
| $T^9$ | 13.98 | NA | NA |
| $T^{10}$ | 13.91 | NA | NA |
| $G^{11}$ | 12.68 | NA | NA |
| $T^{12}$ | Not Assigned | NA | NA |
| $A^{13}$ | NA | NA | NA |
| $C^{14}$ | NA | 6.65 | 8.33 |
| $A^{15}$ | NA | NA | NA |
| $A^{16}$ | NA | NA | NA |
| $A^{17}$ | NA | NA | NA |

| | | | |
|---|---|---|---|
| $C^{18}$ | NA | 6.46 | 7.96 |
| $A^{19}$ | NA | NA | NA |
| $C^{20}$ | NA | 6.48 | 8.12 |
| $G^{21}$ | 12.86 | NA | NA |
| $C^{22}$ | NA | 6.48 | 8.38 |
| $A^{23}$ | NA | NA | NA |
| $C^{24}$ | NA | 6.85 | 8.16 |

**Table A6.** Chemical shifts (ppm) of the C8-dG-ABA Protons for the 5'-d(GTGCXTGTTTGT)-3':5'-d(ACAAACACGCAC)-3' Duplex; X = C8-dG-ABA Adduct. The temperature was 15 ºC.

| C8-dG-ABA Proton | Chemical Shift (ppm) |
|---|---|
| H1 | 8.04 |
| H2 | 8.25 |
| H4 | 7.95 |
| H5 | 7.61 |
| H6 | 8.26 |
| H8 | 7.43 |
| H9 | 6.64 |
| H10 | 7.20 |
| H11 | 7.78 |

**Table A7.** NOE Distance Restraints Used for rMD Calculations for the 5'-d(GTGCXTGTTTGT)-3':5'-d(ACAAACACGCAC)-3' Duplex; X = C8-dG-ABA Adduct.

| Nucleotide 1 | Proton | Nucleotide 2 | Proton | Lower (Å) | Upper (Å) | Window (Å) |
|---|---|---|---|---|---|---|
| $G^1$ | H8 | $G^1$ | H1' | 2.94 | 4.88 | 1.94 |
| $G^1$ | H8 | $G^1$ | H3' | 2.89 | 4.68 | 1.79 |
| $G^1$ | H1' | $G^1$ | H2' | 2.80 | 3.32 | 0.52 |
| $G^1$ | H1' | $G^1$ | H2" | 1.92 | 2.51 | 0.59 |
| $G^1$ | H1' | $G^1$ | H3' | 3.63 | 3.92 | 0.29 |
| $G^1$ | H2' | $G^1$ | H3' | 1.95 | 2.68 | 0.73 |
| $G^1$ | H2" | $G^1$ | H3' | 2.49 | 3.02 | 0.53 |
| $G^1$ | H8 | $T^2$ | H6 | 3.33 | 5.16 | 1.83 |
| $G^1$ | H8 | $T^2$ | Me | 2.50 | 3.62 | 1.12 |
| $G^1$ | H1' | $T^2$ | H6 | 2.46 | 4.64 | 2.18 |
| $G^1$ | H1' | $T^2$ | Me | 2.63 | 5.93 | 3.30 |
| $G^1$ | H2' | $T^2$ | Me | 2.66 | 3.20 | 0.54 |
| $G^1$ | H2" | $T^2$ | H6 | 2.09 | 3.75 | 1.66 |
| $G^1$ | H2" | $T^2$ | Me | 2.24 | 3.70 | 1.46 |
| $G^1$ | H3' | $T^2$ | H6 | 3.64 | 5.79 | 2.15 |
| $T^2$ | H6 | $T^2$ | H1' | 3.03 | 5.41 | 2.38 |
| $T^2$ | H6 | $T^2$ | H2" | 1.86 | 3.46 | 1.60 |
| $T^2$ | H1' | $T^2$ | H2' | 2.69 | 3.51 | 0.82 |
| $T^2$ | H1' | $T^2$ | H3' | 2.63 | 4.45 | 1.82 |

| | | | | | | |
|---|---|---|---|---|---|---|
| $T^2$ | H2' | $T^2$ | H3' | 2.16 | 3.26 | 1.10 |
| $T^2$ | H6 | $G^3$ | H8 | 3.51 | 5.53 | 2.02 |
| $T^2$ | Me | $G^3$ | H8 | 3.75 | 5.76 | 2.01 |
| $T^2$ | H1' | $G^3$ | H8 | 3.07 | 3.95 | 0.88 |
| $T^2$ | H2' | $G^3$ | H8 | 3.30 | 3.96 | 0.66 |
| $T^2$ | H3' | $G^3$ | H8 | 3.60 | 5.29 | 1.69 |
| $G^3$ | H8 | $G^3$ | H1' | 3.35 | 5.29 | 1.94 |
| $G^3$ | H8 | $G^3$ | H3' | 3.03 | 5.96 | 2.93 |
| $G^3$ | H1' | $G^3$ | H2' | 2.60 | 3.73 | 1.13 |
| $G^3$ | H1' | $G^3$ | H2" | 2.30 | 3.08 | 0.78 |
| $G^3$ | H1' | $G^3$ | H3' | 3.26 | 5.31 | 2.05 |
| $G^3$ | H8 | $C^4$ | H6 | 3.79 | 5.29 | 1.50 |
| $G^3$ | H8 | $C^4$ | H5 | 3.36 | 4.41 | 1.05 |
| $G^3$ | H1' | $C^4$ | H6 | 3.16 | 3.67 | 0.51 |
| $G^3$ | H1' | $C^4$ | H5 | 3.41 | 4.36 | 0.95 |
| $G^3$ | H2' | $C^4$ | H6 | 3.21 | 5.89 | 2.68 |
| $G^3$ | H2' | $C^4$ | H5 | 3.08 | 3.68 | 0.60 |
| $G^3$ | H2" | $C^4$ | H6 | 2.38 | 2.95 | 0.57 |
| $G^3$ | H2" | $C^4$ | H5 | 2.86 | 3.73 | 0.87 |
| $G^3$ | H3' | $C^4$ | H5 | 4.13 | 6.60 | 2.47 |
| $C^4$ | H6 | $C^4$ | H1' | 3.37 | 4.09 | 0.72 |
| $C^4$ | H6 | $C^4$ | H2' | 2.47 | 2.92 | 0.45 |

| | | | | | | |
|---|---|---|---|---|---|---|
| $C^4$ | H6 | $C^4$ | H2" | 2.89 | 4.54 | 1.65 |
| $C^4$ | H6 | $C^4$ | H3' | 3.23 | 4.25 | 1.02 |
| $C^4$ | H5 | $C^4$ | H2' | 3.85 | 4.60 | 0.75 |
| $C^4$ | H1' | $C^4$ | H2" | 2.41 | 3.15 | 0.74 |
| $C^4$ | H2' | $C^4$ | H3' | 2.50 | 3.05 | 0.55 |
| $C^4$ | H2" | $C^4$ | H3' | 2.62 | 3.92 | 1.30 |
| $C^4$ | H5 | $X^5$ | ABA 2 | 3.90 | 4.83 | 0.93 |
| $C^4$ | H2' | $X^5$ | H1' | 3.70 | 4.30 | 0.60 |
| $C^4$ | H2" | $X^5$ | H1' | 3.70 | 4.15 | 0.45 |
| $X^5$ | ABA 4 | $X^5$ | H1' | 2.30 | 4.55 | 2.25 |
| $X^5$ | ABA 4 | $X^5$ | H4' | 3.07 | 4.17 | 1.10 |
| $X^5$ | ABA 4 | $X^5$ | H5' | 2.84 | 4.54 | 1.70 |
| $X^5$ | ABA 4 | $X^5$ | H5" | 3.31 | 5.65 | 2.34 |
| $X^5$ | ABA 5 | $X^5$ | H1' | 3.47 | 5.85 | 2.38 |
| $X^5$ | H1' | $X^5$ | H2' | 2.48 | 4.16 | 1.68 |
| $X^5$ | H1' | $X^5$ | H3' | 2.79 | 4.52 | 1.73 |
| $X^5$ | H1' | $X^5$ | H4' | 2.71 | 3.32 | 0.61 |
| $X^5$ | H2' | $X^5$ | H3' | 2.15 | 2.78 | 0.63 |
| $X^5$ | H2' | $X^5$ | H4' | 3.05 | 5.44 | 2.39 |
| $X^5$ | H2" | $X^5$ | H3' | 2.40 | 3.84 | 1.44 |
| $X^5$ | H2" | $X^5$ | H4' | 3.22 | 5.09 | 1.87 |
| $X^5$ | H3' | $X^5$ | H4' | 2.55 | 3.15 | 0.60 |

| | | | | | | |
|---|---|---|---|---|---|---|
| $X^5$ | ABA 2 | $T^6$ | Me | 3.72 | 5.90 | 2.18 |
| $X^5$ | ABA 4 | $T^6$ | H6 | 3.51 | 5.06 | 1.55 |
| $X^5$ | ABA 4 | $T^6$ | Me | 3.17 | 5.79 | 2.62 |
| $X^5$ | ABA 4 | $T^6$ | H1' | 3.51 | 4.89 | 1.38 |
| $X^5$ | ABA 5 | $T^6$ | H1' | 3.13 | 3.42 | 0.29 |
| $X^5$ | H1' | $T^6$ | H6 | 2.84 | 4.37 | 1.53 |
| $X^5$ | H1' | $T^6$ | Me | 2.45 | 3.28 | 0.83 |
| $X^5$ | H2' | $T^6$ | H6 | 3.03 | 4.85 | 1.82 |
| $X^5$ | H2' | $T^6$ | Me | 2.66 | 4.46 | 1.80 |
| $X^5$ | H3' | $T^6$ | H6 | 3.11 | 4.26 | 1.15 |
| $X^5$ | H3' | $T^6$ | Me | 3.63 | 5.25 | 1.62 |
| $X^5$ | H4' | $T^6$ | Me | 4.51 | 6.77 | 2.26 |
| $X^5$ | ABA 8 | $A^{19}$ | H1' | 3.55 | 4.14 | 0.59 |
| $X^5$ | ABA 8 | $A^{19}$ | H2' | 2.93 | 4.75 | 1.82 |
| $X^5$ | ABA 8 | $A^{19}$ | H2" | 2.69 | 3.76 | 1.07 |
| $X^5$ | ABA 9 | $A^{19}$ | H8 | 3.85 | 5.85 | 2.00 |
| $X^5$ | ABA 9 | $A^{19}$ | H1' | 3.76 | 5.07 | 1.31 |
| $X^5$ | ABA 9 | $A^{19}$ | H2' | 3.06 | 4.55 | 1.49 |
| $X^5$ | ABA 9 | $A^{19}$ | H2" | 3.07 | 4.93 | 1.86 |
| $X^5$ | ABA 9 | $A^{19}$ | H3' | 3.33 | 4.97 | 1.64 |
| $X^5$ | ABA 10 | $A^{19}$ | H1' | 4.25 | 5.84 | 1.59 |
| $X^5$ | ABA 10 | $A^{19}$ | H2" | 3.38 | 5.23 | 1.85 |

| | | | | | | |
|---|---|---|---|---|---|---|
| $X^5$ | ABA 8 | $C^{20}$ | H6 | 3.77 | 4.50 | 0.73 |
| $X^5$ | ABA 9 | $C^{20}$ | H4' | 3.30 | 5.70 | 2.40 |
| $X^5$ | ABA 6 | $G^{21}$ | H1' | 3.06 | 3.45 | 0.39 |
| $T^6$ | H6 | $T^6$ | H2" | 2.33 | 4.14 | 1.81 |
| $T^6$ | H1' | $T^6$ | H2' | 2.56 | 3.05 | 0.49 |
| $T^6$ | H1' | $T^6$ | H3' | 3.19 | 4.64 | 1.45 |
| $T^6$ | H2" | $T^6$ | H3' | 2.20 | 3.54 | 1.34 |
| $T^6$ | H6 | $G^7$ | H8 | 4.05 | 4.70 | 0.65 |
| $T^6$ | H1' | $G^7$ | H8 | 3.18 | 3.70 | 0.52 |
| $T^6$ | H2' | $G^7$ | H8 | 3.50 | 4.67 | 1.17 |
| $T^6$ | H2" | $G^7$ | H8 | 2.58 | 3.45 | 0.87 |
| $T^6$ | H3' | $G^7$ | H8 | 3.40 | 5.12 | 1.72 |
| $G^7$ | H8 | $G^7$ | H1' | 2.93 | 4.78 | 1.85 |
| $G^7$ | H8 | $G^7$ | H2' | 1.95 | 2.40 | 0.45 |
| $G^7$ | H8 | $G^7$ | H2" | 2.92 | 3.85 | 0.93 |
| $G^7$ | H8 | $G^7$ | H3' | 3.61 | 4.87 | 1.26 |
| $G^7$ | H1' | $G^7$ | H2" | 2.07 | 2.45 | 0.38 |
| $G^7$ | H8 | $T^8$ | H6 | 4.30 | 5.00 | 0.70 |
| $G^7$ | H8 | $T^8$ | Me | 3.32 | 3.65 | 0.33 |
| $G^7$ | H1' | $T^8$ | H6 | 3.18 | 4.33 | 1.15 |
| $G^7$ | H1' | $T^8$ | Me | 3.32 | 5.00 | 1.68 |
| $G^7$ | H2' | $T^8$ | H6 | 2.39 | 4.14 | 1.75 |

| | | | | | | |
|---|---|---|---|---|---|---|
| $G^{7}$ | H2' | $T^{8}$ | Me | 2.49 | 4.05 | 1.56 |
| $G^{7}$ | H2'' | $T^{8}$ | H6 | 2.25 | 2.69 | 0.44 |
| $G^{7}$ | H2'' | $T^{8}$ | Me | 2.47 | 4.03 | 1.56 |
| $G^{7}$ | H3' | $T^{8}$ | H6 | 3.84 | 4.90 | 1.06 |
| $G^{7}$ | H3' | $T^{8}$ | Me | 3.39 | 4.78 | 1.39 |
| $T^{8}$ | H6 | $T^{8}$ | H1' | 2.57 | 4.84 | 2.27 |
| $T^{8}$ | H6 | $T^{8}$ | H2'' | 2.31 | 3.60 | 1.29 |
| $T^{8}$ | H6 | $T^{8}$ | H3' | 3.04 | 4.64 | 1.60 |
| $T^{8}$ | H1' | $T^{8}$ | H2' | 2.50 | 3.40 | 0.90 |
| $T^{8}$ | H1' | $T^{8}$ | H2'' | 2.12 | 2.45 | 0.33 |
| $T^{8}$ | H1' | $T^{8}$ | H3' | 2.94 | 4.91 | 1.97 |
| $T^{8}$ | H6 | $T^{9}$ | H6 | 2.90 | 5.37 | 2.47 |
| $T^{8}$ | H6 | $T^{9}$ | Me | 3.11 | 4.88 | 1.77 |
| $T^{8}$ | H1' | $T^{9}$ | H6 | 2.93 | 5.85 | 2.92 |
| $T^{8}$ | H1' | $T^{9}$ | Me | 3.13 | 5.54 | 2.41 |
| $T^{8}$ | H2' | $T^{9}$ | Me | 1.84 | 3.56 | 1.72 |
| $T^{8}$ | H3' | $T^{9}$ | Me | 3.12 | 5.19 | 2.07 |
| $T^{9}$ | H6 | $T^{9}$ | H1' | 3.31 | 3.91 | 0.60 |
| $T^{9}$ | H1' | $T^{9}$ | H2'' | 2.00 | 2.83 | 0.83 |
| $T^{9}$ | H1' | $T^{9}$ | H3' | 2.80 | 4.08 | 1.28 |
| $T^{9}$ | H6 | $T^{10}$ | H6 | 3.60 | 6.06 | 2.46 |
| $T^{9}$ | H6 | $T^{10}$ | Me | 2.58 | 3.78 | 1.20 |

| | | | | | | |
|---|---|---|---|---|---|---|
| $T^{9}$ | H1' | $T^{10}$ | H6 | 2.76 | 3.48 | 0.72 |
| $T^{9}$ | H1' | $T^{10}$ | Me | 3.18 | 4.87 | 1.69 |
| $T^{9}$ | H2' | $T^{10}$ | Me | 2.45 | 3.28 | 0.83 |
| $T^{9}$ | H2" | $T^{10}$ | H6 | 2.30 | 3.00 | 0.70 |
| $T^{9}$ | H2" | $T^{10}$ | Me | 2.49 | 3.67 | 1.18 |
| $T^{9}$ | H3' | $T^{10}$ | Me | 3.49 | 5.34 | 1.85 |
| $T^{10}$ | H6 | $T^{10}$ | H1' | 2.91 | 5.67 | 2.76 |
| $T^{10}$ | H6 | $T^{10}$ | H2' | 2.02 | 2.90 | 0.88 |
| $T^{10}$ | H6 | $T^{10}$ | H2" | 2.08 | 3.87 | 1.79 |
| $T^{10}$ | Me | $T^{10}$ | H2' | 2.30 | 4.90 | 2.60 |
| $T^{10}$ | Me | $T^{10}$ | H2" | 2.64 | 5.13 | 2.49 |
| $T^{10}$ | H1' | $T^{10}$ | H2' | 2.08 | 3.48 | 1.40 |
| $T^{10}$ | H1' | $T^{10}$ | H2" | 1.77 | 2.70 | 0.93 |
| $T^{10}$ | H1' | $T^{10}$ | H3' | 2.48 | 4.08 | 1.60 |
| $T^{10}$ | H2' | $T^{10}$ | H3' | 2.35 | 2.75 | 0.40 |
| $T^{10}$ | H2" | $T^{10}$ | H3' | 2.25 | 3.57 | 1.32 |
| $T^{10}$ | H6 | $G^{11}$ | H8 | 3.42 | 5.18 | 1.76 |
| $T^{10}$ | H2' | $G^{11}$ | H8 | 2.19 | 4.52 | 2.33 |
| $T^{10}$ | H2" | $G^{11}$ | H8 | 2.39 | 4.41 | 2.02 |
| $T^{10}$ | H3' | $G^{11}$ | H8 | 2.67 | 6.07 | 3.40 |
| $G^{11}$ | H8 | $G^{11}$ | H1' | 2.56 | 4.84 | 2.28 |
| $G^{11}$ | H1' | $G^{11}$ | H3' | 2.96 | 3.96 | 1.00 |

| | | | | | | |
|---|---|---|---|---|---|---|
| $G^{11}$ | H1' | $G^{11}$ | H4' | 2.51 | 3.39 | 0.88 |
| $G^{11}$ | H8 | $T^{12}$ | H6 | 3.01 | 5.41 | 2.40 |
| $G^{11}$ | H8 | $T^{12}$ | Me | 2.66 | 3.68 | 1.02 |
| $G^{11}$ | H1' | $T^{12}$ | Me | 3.64 | 4.54 | 0.90 |
| $G^{11}$ | H2" | $T^{12}$ | Me | 2.80 | 3.40 | 0.60 |
| $G^{11}$ | H2" | $T^{12}$ | H3' | 3.39 | 4.92 | 1.53 |
| $G^{11}$ | H3' | $T^{12}$ | H6 | 3.18 | 5.20 | 2.02 |
| $G^{11}$ | H3' | $T^{12}$ | Me | 4.07 | 6.40 | 2.33 |
| $T^{12}$ | H6 | $T^{12}$ | H3' | 2.38 | 2.78 | 0.40 |
| $T^{12}$ | Me | $T^{12}$ | H2' | 2.13 | 4.05 | 1.92 |
| $T^{12}$ | Me | $T^{12}$ | H3' | 3.81 | 6.97 | 3.16 |
| $T^{12}$ | H1' | $T^{12}$ | H3' | 3.18 | 5.09 | 1.91 |
| | | | | | | |
| $A^{13}$ | H8 | $A^{13}$ | H1' | 2.38 | 5.09 | 2.71 |
| $A^{13}$ | H8 | $A^{13}$ | H3' | 3.10 | 5.73 | 2.63 |
| $A^{13}$ | H1' | $A^{13}$ | H2' | 2.47 | 4.35 | 1.88 |
| $A^{13}$ | H1' | $A^{13}$ | H2" | 1.99 | 2.43 | 0.44 |
| $A^{13}$ | H1' | $A^{13}$ | H3' | 3.09 | 5.24 | 2.15 |
| $A^{13}$ | H2' | $A^{13}$ | H3' | 2.19 | 3.20 | 1.01 |
| $A^{13}$ | H2" | $A^{13}$ | H3' | 2.44 | 3.07 | 0.63 |
| $A^{13}$ | H2 | $C^{14}$ | H1' | 3.66 | 4.79 | 1.13 |
| $A^{13}$ | H1' | $C^{14}$ | H6 | 2.68 | 4.38 | 1.70 |

| | | | | | | |
|---|---|---|---|---|---|---|
| $A^{13}$ | H1' | $C^{14}$ | H5 | 2.57 | 4.72 | 2.15 |
| $A^{13}$ | H1' | $C^{14}$ | H1' | 3.56 | 4.95 | 1.39 |
| $A^{13}$ | H2' | $C^{14}$ | H5 | 2.76 | 3.50 | 0.74 |
| $A^{13}$ | H2" | $C^{14}$ | H6 | 2.16 | 2.89 | 0.73 |
| $A^{13}$ | H2" | $C^{14}$ | H5 | 2.25 | 5.80 | 3.55 |
| $C^{14}$ | H6 | $C^{14}$ | H1' | 2.90 | 4.10 | 1.20 |
| $C^{14}$ | H6 | $C^{14}$ | H2' | 1.74 | 2.45 | 0.71 |
| $C^{14}$ | H6 | $C^{14}$ | H3' | 2.91 | 4.41 | 1.50 |
| $C^{14}$ | H5 | $C^{14}$ | H1' | 4.04 | 5.99 | 1.95 |
| $C^{14}$ | H5 | $C^{14}$ | H2' | 3.26 | 4.82 | 1.56 |
| $C^{14}$ | H5 | $C^{14}$ | H2" | 2.97 | 6.47 | 3.50 |
| $C^{14}$ | H1' | $C^{14}$ | H2' | 2.39 | 3.45 | 1.06 |
| $C^{14}$ | H1' | $C^{14}$ | H2" | 2.06 | 2.69 | 0.63 |
| $C^{14}$ | H1' | $C^{14}$ | H3' | 3.56 | 5.09 | 1.53 |
| $C^{14}$ | H2" | $C^{14}$ | H3' | 2.42 | 3.27 | 0.85 |
| $C^{14}$ | H1' | $A^{15}$ | H8 | 2.92 | 3.78 | 0.86 |
| $C^{14}$ | H2" | $A^{15}$ | H8 | 2.38 | 3.60 | 1.22 |
| $A^{15}$ | H8 | $A^{15}$ | H1' | 3.00 | 3.98 | 0.98 |
| $A^{15}$ | H8 | $A^{15}$ | H3' | 3.07 | 4.71 | 1.64 |
| $A^{15}$ | H2 | $A^{15}$ | H1' | 3.93 | 4.87 | 0.94 |
| $A^{15}$ | H1' | $A^{15}$ | H2' | 2.20 | 3.04 | 0.84 |
| $A^{15}$ | H1' | $A^{15}$ | H2" | 1.92 | 3.67 | 1.75 |

| | | | | | | |
|---|---|---|---|---|---|---|
| $A^{15}$ | H1' | $A^{15}$ | H3' | 2.23 | 4.02 | 1.79 |
| $A^{15}$ | H2'' | $A^{15}$ | H3' | 2.31 | 3.41 | 1.10 |
| $A^{15}$ | H2 | $A^{16}$ | H2 | 3.70 | 4.10 | 0.40 |
| $A^{15}$ | H1' | $A^{16}$ | H8 | 2.81 | 3.50 | 0.69 |
| $A^{15}$ | H2' | $A^{16}$ | H8 | 3.60 | 5.83 | 2.23 |
| $A^{15}$ | H3' | $A^{16}$ | H8 | 2.86 | 5.00 | 2.14 |
| $A^{16}$ | H8 | $A^{16}$ | H1' | 3.21 | 5.60 | 2.39 |
| $A^{16}$ | H8 | $A^{16}$ | H2' | 1.91 | 2.54 | 0.63 |
| $A^{16}$ | H8 | $A^{16}$ | H3' | 2.98 | 5.20 | 2.22 |
| $A^{16}$ | H2 | $A^{16}$ | H1' | 3.90 | 4.70 | 0.80 |
| $A^{16}$ | H1' | $A^{16}$ | H2' | 2.06 | 3.85 | 1.79 |
| $A^{16}$ | H1' | $A^{16}$ | H2'' | 1.96 | 2.98 | 1.02 |
| $A^{16}$ | H1' | $A^{16}$ | H3' | 2.97 | 4.64 | 1.67 |
| $A^{16}$ | H2' | $A^{16}$ | H3' | 1.80 | 2.55 | 0.75 |
| $A^{16}$ | H2'' | $A^{16}$ | H3' | 2.28 | 4.34 | 2.06 |
| $A^{16}$ | H2 | $A^{17}$ | H2 | 3.76 | 4.47 | 0.71 |
| $A^{16}$ | H2' | $A^{17}$ | H8 | 3.09 | 4.85 | 1.76 |
| $A^{17}$ | H8 | $A^{17}$ | H1' | 3.23 | 4.10 | 0.87 |
| $A^{17}$ | H8 | $A^{17}$ | H2' | 2.12 | 4.02 | 1.90 |
| $A^{17}$ | H2 | $A^{17}$ | H1' | 3.84 | 4.79 | 0.95 |
| $A^{17}$ | H1' | $A^{17}$ | H2' | 2.62 | 4.07 | 1.45 |
| $A^{17}$ | H1' | $A^{17}$ | H2'' | 1.93 | 2.76 | 0.83 |

| | | | | | | |
|---|---|---|---|---|---|---|
| $A^{17}$ | H2' | $A^{17}$ | H3' | 2.09 | 2.73 | 0.64 |
| $A^{17}$ | H2'' | $A^{17}$ | H3' | 2.24 | 4.40 | 2.16 |
| $A^{17}$ | H8 | $C^{18}$ | H6 | 3.90 | 5.82 | 1.92 |
| $A^{17}$ | H2 | $C^{18}$ | H1' | 4.00 | 5.42 | 1.42 |
| $A^{17}$ | H1' | $C^{18}$ | H6 | 3.79 | 4.99 | 1.20 |
| $A^{17}$ | H2' | $C^{18}$ | H6 | 3.07 | 6.09 | 3.02 |
| $A^{17}$ | H2'' | $C^{18}$ | H6 | 2.70 | 4.56 | 1.86 |
| $A^{17}$ | H2'' | $C^{18}$ | H5 | 2.62 | 4.67 | 2.05 |
| $A^{17}$ | H3' | $C^{18}$ | H6 | 3.90 | 5.50 | 1.60 |
| $C^{18}$ | H6 | $C^{18}$ | H1' | 3.25 | 4.03 | 0.78 |
| $C^{18}$ | H6 | $C^{18}$ | H2' | 2.30 | 2.91 | 0.61 |
| $C^{18}$ | H6 | $C^{18}$ | H2'' | 3.50 | 4.05 | 0.55 |
| $C^{18}$ | H6 | $C^{18}$ | H3' | 3.49 | 4.53 | 1.04 |
| $C^{18}$ | H5 | $C^{18}$ | H2' | 3.28 | 5.37 | 2.09 |
| $C^{18}$ | H1' | $C^{18}$ | H2' | 2.74 | 3.16 | 0.42 |
| $C^{18}$ | H1' | $C^{18}$ | H2'' | 2.50 | 2.96 | 0.46 |
| $C^{18}$ | H1' | $C^{18}$ | H3' | 3.24 | 4.94 | 1.70 |
| $C^{18}$ | H2' | $C^{18}$ | H3' | 2.35 | 2.97 | 0.62 |
| $C^{18}$ | H2'' | $C^{18}$ | H3' | 2.60 | 3.49 | 0.89 |
| $C^{18}$ | H1' | $A^{19}$ | H8 | 3.45 | 4.00 | 0.55 |
| $C^{18}$ | H2' | $A^{19}$ | H8 | 3.40 | 4.20 | 0.80 |
| $C^{18}$ | H3' | $A^{19}$ | H8 | 4.25 | 5.25 | 1.00 |

| | | | | | | |
|---|---|---|---|---|---|---|
| $A^{19}$ | H8 | $A^{19}$ | H1' | 3.68 | 4.73 | 1.05 |
| $A^{19}$ | H8 | $A^{19}$ | H2' | 2.26 | 3.25 | 0.99 |
| $A^{19}$ | H8 | $A^{19}$ | H3' | 3.08 | 4.85 | 1.77 |
| $A^{19}$ | H1' | $A^{19}$ | H2' | 2.96 | 3.36 | 0.40 |
| $A^{19}$ | H1' | $A^{19}$ | H2" | 2.22 | 3.09 | 0.87 |
| $A^{19}$ | H8 | $C^{20}$ | H6 | 3.84 | 5.79 | 1.95 |
| $A^{19}$ | H1' | $C^{20}$ | H6 | 3.71 | 5.23 | 1.52 |
| $A^{19}$ | H2' | $C^{20}$ | H6 | 2.57 | 3.78 | 1.21 |
| $A^{19}$ | H2' | $C^{20}$ | H5 | 3.24 | 4.73 | 1.49 |
| $A^{19}$ | H2" | $C^{20}$ | H6 | 2.89 | 4.21 | 1.32 |
| $A^{19}$ | H2" | $C^{20}$ | H5 | 3.48 | 5.87 | 2.39 |
| $A^{19}$ | H3' | $C^{20}$ | H6 | 3.06 | 3.68 | 0.62 |
| $A^{19}$ | H3' | $C^{20}$ | H5 | 2.64 | 3.08 | 0.44 |
| $C^{20}$ | H6 | $C^{20}$ | H1' | 3.04 | 5.00 | 1.96 |
| $C^{20}$ | H6 | $C^{20}$ | H2' | 2.36 | 3.60 | 1.24 |
| $C^{20}$ | H6 | $C^{20}$ | H2" | 2.33 | 3.70 | 1.37 |
| $C^{20}$ | H6 | $C^{20}$ | H3' | 3.06 | 4.96 | 1.90 |
| $C^{20}$ | H6 | $C^{20}$ | H4' | 3.52 | 5.55 | 2.03 |
| $C^{20}$ | H6 | $C^{20}$ | H5' | 3.30 | 5.30 | 2.00 |
| $C^{20}$ | H6 | $C^{20}$ | H5" | 3.30 | 5.30 | 2.00 |
| $C^{20}$ | H5 | $C^{20}$ | H2' | 3.15 | 5.95 | 2.80 |
| $C^{20}$ | H1' | $C^{20}$ | H2' | 2.90 | 4.14 | 1.24 |

| | | | | | | |
|---|---|---|---|---|---|---|
| $C^{20}$ | H1' | $C^{20}$ | H2" | 1.91 | 2.49 | 0.58 |
| $C^{20}$ | H1' | $C^{20}$ | H3' | 3.27 | 4.15 | 0.88 |
| $C^{20}$ | H1' | $C^{20}$ | H4' | 3.02 | 3.84 | 0.82 |
| $C^{20}$ | H2' | $C^{20}$ | H3' | 2.40 | 2.80 | 0.40 |
| $C^{20}$ | H2" | $C^{20}$ | H3' | 2.55 | 3.30 | 0.75 |
| $C^{20}$ | H1' | $G^{21}$ | H8 | 3.36 | 3.81 | 0.45 |
| $G^{21}$ | H8 | $G^{21}$ | H1' | 3.23 | 4.05 | 0.82 |
| $G^{21}$ | H8 | $G^{21}$ | H2" | 2.85 | 4.26 | 1.41 |
| $G^{21}$ | H8 | $G^{21}$ | H3' | 3.34 | 5.25 | 1.91 |
| $G^{21}$ | H1' | $G^{21}$ | H2' | 2.51 | 3.00 | 0.49 |
| $G^{21}$ | H1' | $G^{21}$ | H3' | 3.35 | 4.94 | 1.59 |
| $G^{21}$ | H2' | $G^{21}$ | H3' | 2.30 | 2.97 | 0.67 |
| $G^{21}$ | H2" | $G^{21}$ | H3' | 2.48 | 4.18 | 1.70 |
| $G^{21}$ | H1' | $C^{22}$ | H6 | 4.30 | 5.20 | 0.90 |
| $G^{21}$ | H1' | $C^{22}$ | H5 | 4.50 | 5.38 | 0.88 |
| $G^{21}$ | H2' | $C^{22}$ | H6 | 2.64 | 4.87 | 2.23 |
| $G^{21}$ | H2' | $C^{22}$ | H5 | 2.81 | 5.10 | 2.29 |
| $G^{21}$ | H2" | $C^{22}$ | H6 | 2.50 | 3.12 | 0.62 |
| $G^{21}$ | H2" | $C^{22}$ | H5 | 2.90 | 4.58 | 1.68 |
| $G^{21}$ | H3' | $C^{22}$ | H6 | 2.95 | 4.78 | 1.83 |
| $G^{21}$ | H3' | $C^{22}$ | H5 | 2.85 | 4.71 | 1.86 |
| $C^{22}$ | H6 | $C^{22}$ | H1' | 2.97 | 4.11 | 1.14 |

| | | | | | | |
|---|---|---|---|---|---|---|
| $C^{22}$ | H6 | $C^{22}$ | H2' | 1.91 | 2.94 | 1.03 |
| $C^{22}$ | H6 | $C^{22}$ | H2" | 2.23 | 4.11 | 1.88 |
| $C^{22}$ | H6 | $C^{22}$ | H3' | 2.61 | 4.04 | 1.43 |
| $C^{22}$ | H5 | $C^{22}$ | H2' | 3.20 | 4.45 | 1.25 |
| $C^{22}$ | H1' | $C^{22}$ | H2' | 2.25 | 3.23 | 0.98 |
| $C^{22}$ | H1' | $C^{22}$ | H2" | 2.06 | 2.40 | 0.34 |
| $C^{22}$ | H1' | $C^{22}$ | H3' | 3.50 | 5.47 | 1.97 |
| $C^{22}$ | H2' | $C^{22}$ | H3' | 2.30 | 3.00 | 0.70 |
| $C^{22}$ | H2" | $C^{22}$ | H3' | 2.60 | 4.18 | 1.58 |
| $C^{22}$ | H1' | $A^{23}$ | H8 | 2.85 | 3.40 | 0.55 |
| $C^{22}$ | H2' | $A^{23}$ | H8 | 3.60 | 4.20 | 0.60 |
| $C^{22}$ | H3' | $A^{23}$ | H8 | 4.45 | 5.15 | 0.70 |
| $A^{23}$ | H8 | $A^{23}$ | H1' | 3.00 | 3.79 | 0.79 |
| $A^{23}$ | H8 | $A^{23}$ | H3' | 2.97 | 4.26 | 1.29 |
| $A^{23}$ | H2 | $A^{23}$ | H1' | 3.56 | 4.87 | 1.31 |
| $A^{23}$ | H1' | $A^{23}$ | H2' | 2.34 | 3.50 | 1.16 |
| $A^{23}$ | H1' | $A^{23}$ | H2" | 1.83 | 2.32 | 0.49 |
| $A^{23}$ | H1' | $A^{23}$ | H3' | 3.36 | 5.38 | 2.02 |
| $A^{23}$ | H2" | $A^{23}$ | H3' | 1.87 | 2.64 | 0.77 |
| $A^{23}$ | H8 | $C^{24}$ | H5 | 3.63 | 4.10 | 0.47 |
| $A^{23}$ | H1' | $C^{24}$ | H6 | 3.00 | 3.99 | 0.99 |
| $A^{23}$ | H1' | $C^{24}$ | H5 | 3.19 | 5.25 | 2.06 |

| | | | | | | |
|---|---|---|---|---|---|---|
| $A^{23}$ | H2' | $C^{24}$ | H6 | 2.36 | 3.95 | 1.59 |
| $A^{23}$ | H2' | $C^{24}$ | H5 | 2.91 | 4.69 | 1.78 |
| $A^{23}$ | H2" | $C^{24}$ | H6 | 1.98 | 2.86 | 0.88 |
| $A^{23}$ | H2" | $C^{24}$ | H5 | 2.87 | 4.59 | 1.72 |
| $A^{23}$ | H2" | $C^{24}$ | H3' | 2.28 | 5.30 | 3.02 |
| $A^{23}$ | H3' | $C^{24}$ | H6 | 2.72 | 5.46 | 2.74 |
| $C^{24}$ | H6 | $C^{24}$ | H1' | 2.68 | 4.83 | 2.15 |
| $C^{24}$ | H6 | $C^{24}$ | H2' | 1.80 | 2.20 | 0.40 |
| $C^{24}$ | H6 | $C^{24}$ | H3' | 2.62 | 3.80 | 1.18 |
| $C^{24}$ | H5 | $C^{24}$ | H1' | 3.82 | 5.92 | 2.10 |
| $C^{24}$ | H5 | $C^{24}$ | H2' | 2.98 | 5.44 | 2.46 |
| $C^{24}$ | H1' | $C^{24}$ | H2' | 2.05 | 5.06 | 3.01 |
| $C^{24}$ | H1' | $C^{24}$ | H2" | 1.62 | 2.20 | 0.58 |
| $C^{24}$ | H1' | $C^{24}$ | H3' | 2.92 | 5.14 | 2.22 |
| $C^{24}$ | H2' | $C^{24}$ | H3' | 1.91 | 2.55 | 0.64 |
| $C^{24}$ | H2" | $C^{24}$ | H3' | 2.28 | 2.75 | 0.47 |

**Table A8.** Anti-distance restraints used for rMD calculations for the 5'-d(GTGCXTGTTTGT)-3':5'-d(ACAAACACGCAC)-3' duplex; X = C8-dG-ABA adduct.

| Nucleotide 1 | Proton | Nucleotide 2 | Proton | Lower (Å) | Upper (Å) |
|---|---|---|---|---|---|
| $X^5$ | ABA H8 | $C^{20}$ | H1' | 5.00 | 9.99 |
| $X^5$ | ABA H8 | $C^{20}$ | H2' | 5.00 | 9.99 |
| $X^5$ | ABA H9 | $C^{20}$ | H1' | 5.00 | 9.99 |
| $X^5$ | ABA H9 | $C^{20}$ | H2" | 5.00 | 9.99 |
| $X^5$ | ABA H10 | $C^{20}$ | H6 | 5.00 | 9.99 |
| $X^5$ | ABA H10 | $G^{21}$ | H8 | 5.00 | 9.99 |
| $A^{19}$ | H2' | $C^{20}$ | H1' | 5.00 | 9.99 |
| $A^{19}$ | H2" | $C^{20}$ | H1' | 5.00 | 9.99 |
| $C^{20}$ | H1' | $G^{21}$ | H5' | 5.00 | 9.99 |
| $C^{20}$ | H1' | $G^{21}$ | H5" | 5.00 | 9.99 |

**Table A9.** RMS differences and sixth root residual $R^X_1$ values calculated for the average structure without including anti-distance restraints.

| Average structure (calculated from ten structures emergent from the simulated annealing rMD calculations) | | | |
|---|---|---|---|
| RMS pairwise difference between structures | | | 0.575 |
| RMS difference from average structure | | | 0.389 |
| | Complete relaxation matrix analysis for the calculated average structure, using the program CORMA[a] | | |
| | Intranucleotide | Internucleotide | Total |
| $R^X_1$[b] | 0.080 | 0.089 | 0.084 |
| Average error[c] | | | 0.019 |

[a]Mixing time was 250 ms

[b]$R^X_1$ is the sixth root R factor: $\sum[((I_o)_i^{1/6})-((I_c)_i^{1/6})/\sum((I_o)_i^{1/6})]$

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

**Table A10.** Comparison of restraints used in rMD calculations to measured distances in refined structure for lesion site.

| Nucleotide 1 | Proton | Nucleotide 2 | Proton | Lower (Å) | Upper (Å) | Measured (Å) |
|---|---|---|---|---|---|---|
| $C^4$ | H5 | $X^5$ | ABA 2 | 3.90 | 4.83 | 4.67 |
| $C^4$ | H2' | $X^5$ | H1' | 3.70 | 4.30 | 4.37* |
| $C^4$ | H2" | $X^5$ | H1' | 3.70 | 4.15 | 4.15 |
| $X^5$ | ABA 4 | $X^5$ | H1' | 2.30 | 4.55 | 3.26 |
| $X^5$ | ABA 4 | $X^5$ | H4' | 3.07 | 4.17 | 3.70 |
| $X^5$ | ABA 4 | $X^5$ | H5' | 2.84 | 4.55 | 4.55 |
| $X^5$ | ABA 4 | $X^5$ | H5" | 3.31 | 5.65 | 5.70* |
| $X^5$ | ABA 5 | $X^5$ | H1' | 3.47 | 5.85 | 5.78 |
| $X^5$ | H1' | $X^5$ | H2' | 2.48 | 4.16 | 3.03 |
| $X^5$ | H1' | $X^5$ | H3' | 2.79 | 4.52 | 3.77 |
| $X^5$ | H1' | $X^5$ | H4' | 2.71 | 3.32 | 2.69* |
| $X^5$ | H2' | $X^5$ | H3' | 2.15 | 2.78 | 2.28 |
| $X^5$ | H2' | $X^5$ | H4' | 3.05 | 5.44 | 3.84 |
| $X^5$ | H2" | $X^5$ | H3' | 2.40 | 3.84 | 2.80 |
| $X^5$ | H2" | $X^5$ | H4' | 3.22 | 5.09 | 3.85 |
| $X^5$ | H3' | $X^5$ | H4' | 2.55 | 3.15 | 2.92 |
| $X^5$ | ABA 2 | $T^6$ | Me | 3.72 | 5.90 | 5.23 |
| $X^5$ | ABA 4 | $T^6$ | H6 | 3.51 | 5.06 | 4.31 |
| $X^5$ | ABA 4 | $T^6$ | Me | 3.17 | 5.79 | 5.44 |
| $X^5$ | ABA 4 | $T^6$ | H1' | 3.51 | 4.89 | 3.95 |

| | | | | | | |
|---|---|---|---|---|---|---|
| $X^5$ | ABA 5 | $T^6$ | H1' | 3.13 | 3.42 | 3.44* |
| $X^5$ | H1' | $T^6$ | H6 | 2.84 | 4.37 | 2.83* |
| $X^5$ | H1' | $T^6$ | Me | 2.45 | 3.28 | 2.81 |
| $X^5$ | H2' | $T^6$ | H6 | 3.03 | 4.85 | 4.88* |
| $X^5$ | H2' | $T^6$ | Me | 2.66 | 4.46 | 4.40 |
| $X^5$ | H3' | $T^6$ | H6 | 3.11 | 4.30 | 4.33* |
| $X^5$ | H3' | $T^6$ | Me | 3.63 | 5.25 | 4.94 |
| $X^5$ | H4' | $T^6$ | Me | 4.51 | 6.77 | 4.69 |
| $X^5$ | ABA 8 | $A^{19}$ | H1' | 3.55 | 4.14 | 3.84 |
| $X^5$ | ABA 8 | $A^{19}$ | H2' | 2.93 | 4.75 | 4.03 |
| $X^5$ | ABA 8 | $A^{19}$ | H2" | 2.69 | 3.76 | 2.69 |
| $X^5$ | ABA 9 | $A^{19}$ | H8 | 3.85 | 5.85 | 4.15 |
| $X^5$ | ABA 9 | $A^{19}$ | H1' | 3.76 | 5.07 | 4.89 |
| $X^5$ | ABA 9 | $A^{19}$ | H2' | 3.06 | 4.55 | 3.06 |
| $X^5$ | ABA 9 | $A^{19}$ | H2" | 3.07 | 4.93 | 3.02* |
| $X^5$ | ABA 9 | $A^{19}$ | H3' | 3.33 | 4.97 | 4.84 |
| $X^5$ | ABA 10 | $A^{19}$ | H1' | 4.25 | 5.84 | 5.88* |
| $X^5$ | ABA 10 | $A^{19}$ | H2" | 3.38 | 5.23 | 4.84 |
| $X^5$ | ABA 8 | $C^{20}$ | H6 | 3.77 | 4.50 | 4.43 |
| $X^5$ | ABA 9 | $C^{20}$ | H4' | 3.30 | 5.70 | 5.53 |
| $X^5$ | ABA 6 | $G^{21}$ | H1' | 3.06 | 3.45 | 3.43 |

* indicates values outside of imposed distance restraints

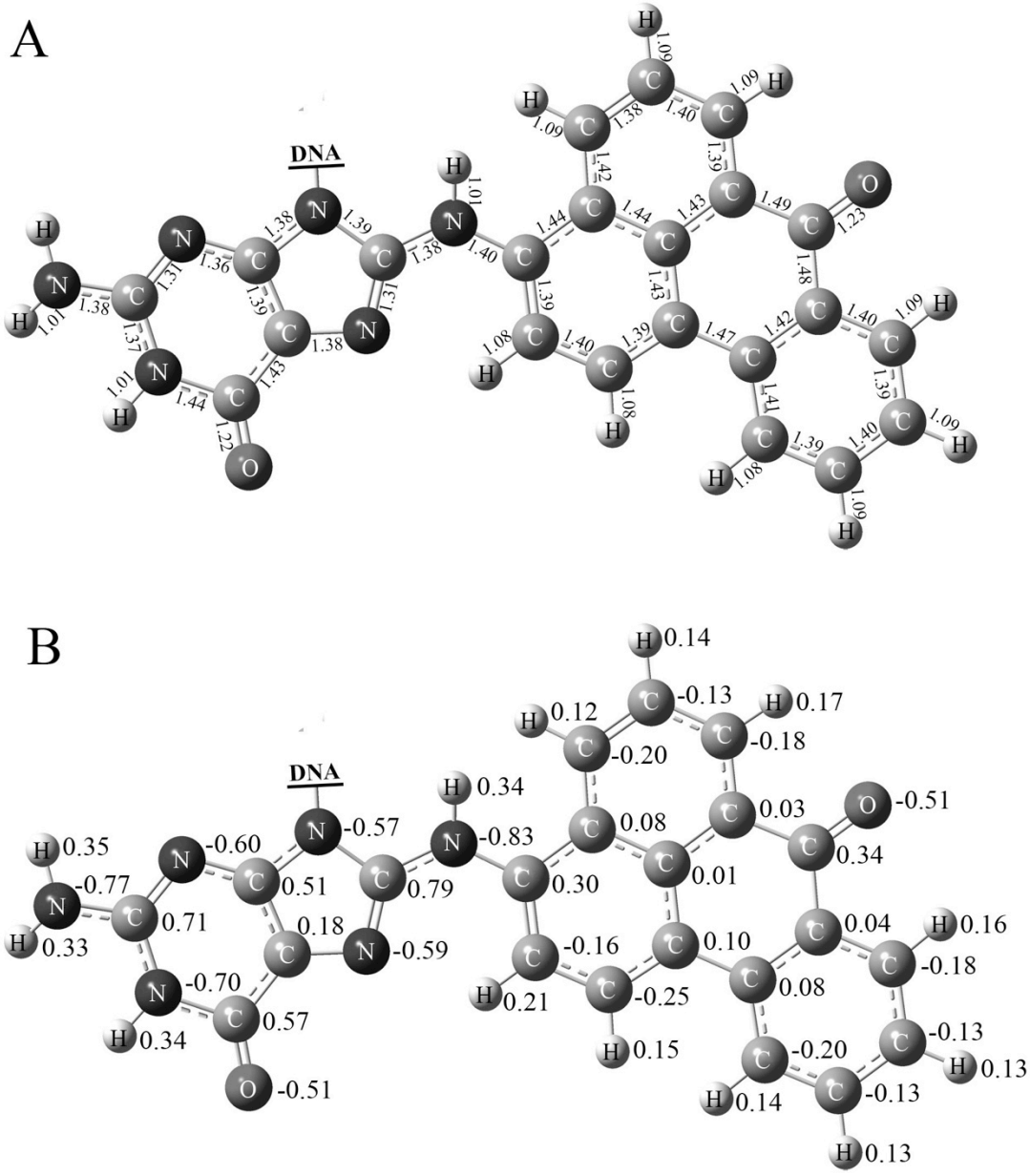



**Figure A1.** Structure of the C8-dG-ABA adduct, depicting bond lengths and partial charges used for rMD calculations. **A.** Calculated bond lengths (Å) for the C8-dG-ABA adduct. **B.** Calculated partial charges for atoms of the C8-dG-ABA adduct. All calculations were performed in GAUSSIAN.

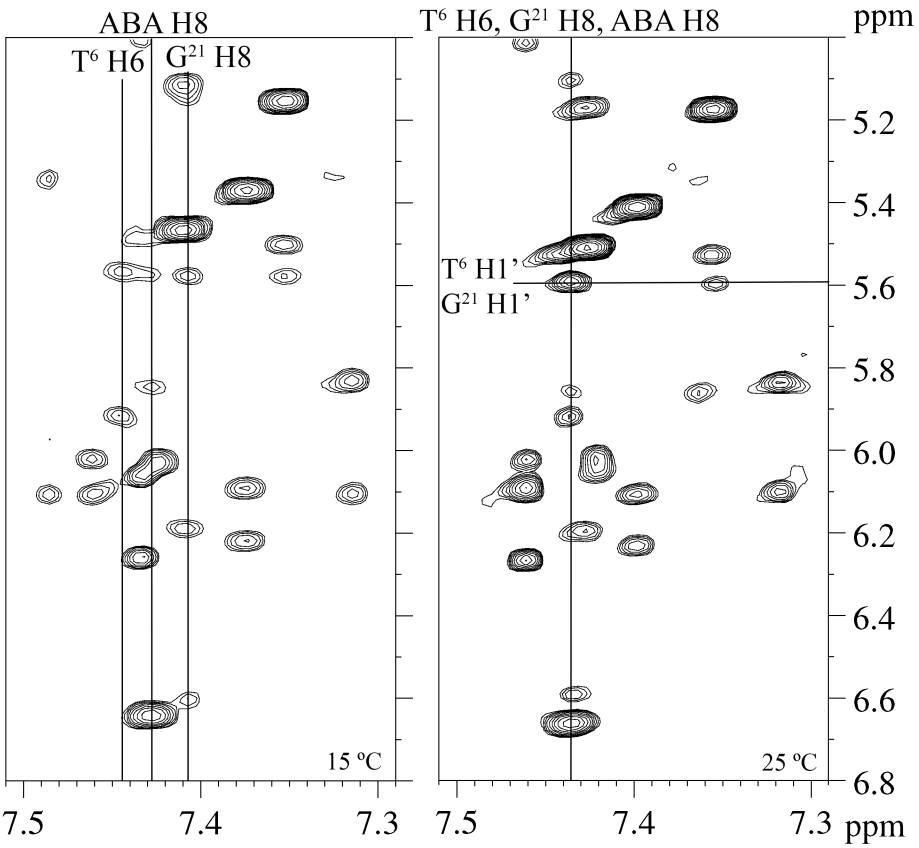



**Figure A2.** Expanded plots of NOESY spectra comparing chemical shift overlaps for the C8-dG-ABA adduct H8, $T^6$ H6 and $G^{21}$ H8 protons in the regions of the spectra showing NOEs between base aromatic and deoxyribose H1' proton resonances. The duplex is 5'-d(GTGCXTGTTTGT)-3':5'-d(ACAAACACGCAC)-3'; X = C8-dG-ABA Adduct. **Left Panel:** The NOESY spectrum at 15 ºC. **Right Panel:** The NOESY spectrum at 25 ºC. The improved spectral resolution obtained at 15 ºC spectrum allowed unequivocal assignment of these resonances.

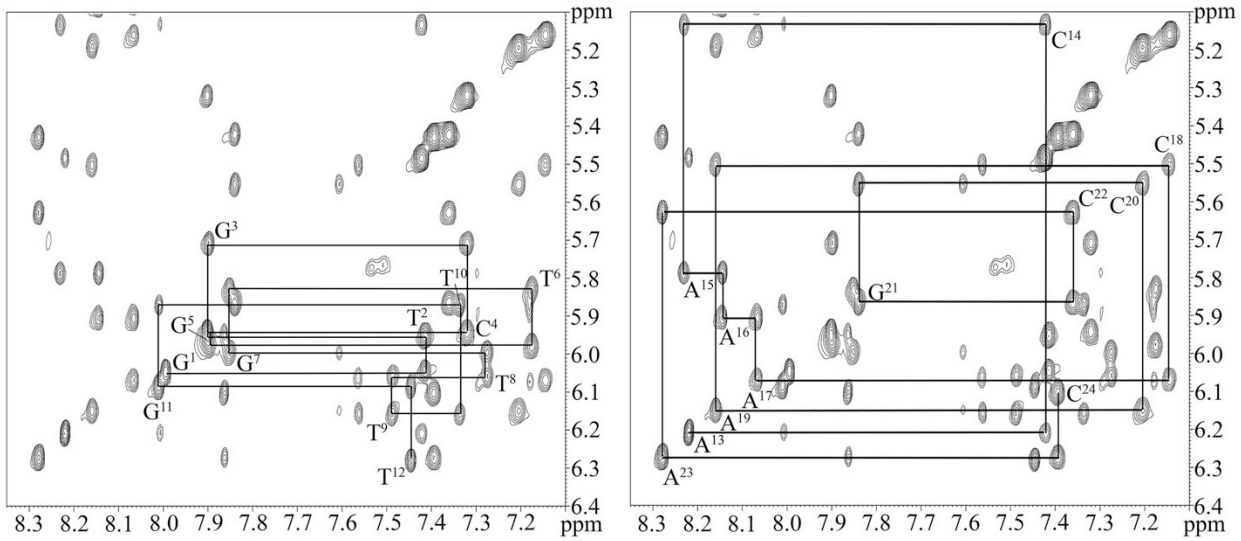



**Figure A3.** Expanded plots from a NOESY spectrum of the 5'-d(GTGCGTGTTTGT)-3':5'-d(ACAAACACGCAC)-3' duplex showing the sequential NOE connectivity between aromatic H8/H6 protons and deoxyribose H1' protons. **Left Panel:** The primary strand, showing bases $G^{1}$ through $T^{12}$. **Right Panel:** The complementary strand connectivity showing assignments for bases $A^{13}$ through $C^{24}$. The 900 MHz spectrum was acquired at 15 ºC using a 250 ms mixing time.

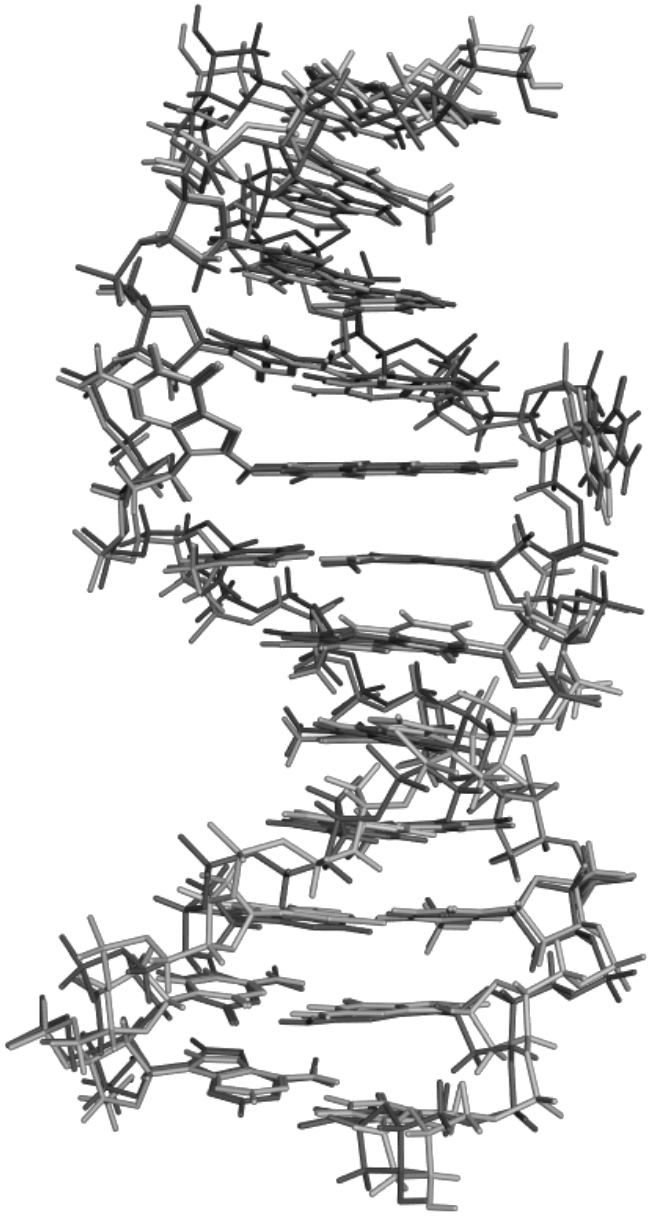


**Figure A4.** Overlay of lowest energy refined structure with anti-distance restraints in dark grey with the lowest energy refined structure without negative restraints in light grey. Major structural features are unchanged and the maximum pairwise difference of the two structures is 0.475 Å.

A II: Supporting Information for Chapter III

**Table A11**. (Expanded version of Table 4 in the main paper) Crystal data, data collection parameters, and structure refinement statistics[a].

| **Data Collection** | |
|---|---|
| Wavelength [Å] | 0.97856 |
| Space group | $P6_1$ |
| Resolution [Å] | 50.0 - 2.60 (2.64 - 2.60)[a] |
| Unit cell *a, b, c* [Å] | 99.11, 99.11, 81.69 |
| Unique reflections | 14,197 (702) |
| Completeness [%] | 100 (100) |
| I/σ(I) | 15.3 (2.0) |
| Wilson B-factor [Å$^2$] | 39.0 |
| R-merge | 0.133 (0.888) |
| Redundancy | 7.6 (7.5) |
| **Refinement** | |
| R-work | 0.166 (0.215) |
| R-free | 0.238 (0.281) |
| Number of atoms | |
| Protein/DNA | 3,377/378 |
| dCTP/water/ $Ca^{2+}$ | 28/145/2 |
| Protein residues | 430 |
| B-factor [Å$^2$] | |

| | |
|---|---|
| Average | 39.0 |
| Protein/DNA | 39.3/38.8 |
| dNTP/$M^{2+}$ /water | 29.0/37.6/39.1 |
| R.m.s. deviations | |
| bonds [Å] | 0.009 |
| angles [deg.] | 1.1 |
| Ramachandran | |
| Favored (%) | 95 |
| Allowed (%) | 4.5 |
| Outliers (%) | 0.5 |
| PDB ID Code | 5JUM |

A[a] Statistics for the highest-resolution shell are shown in parentheses.

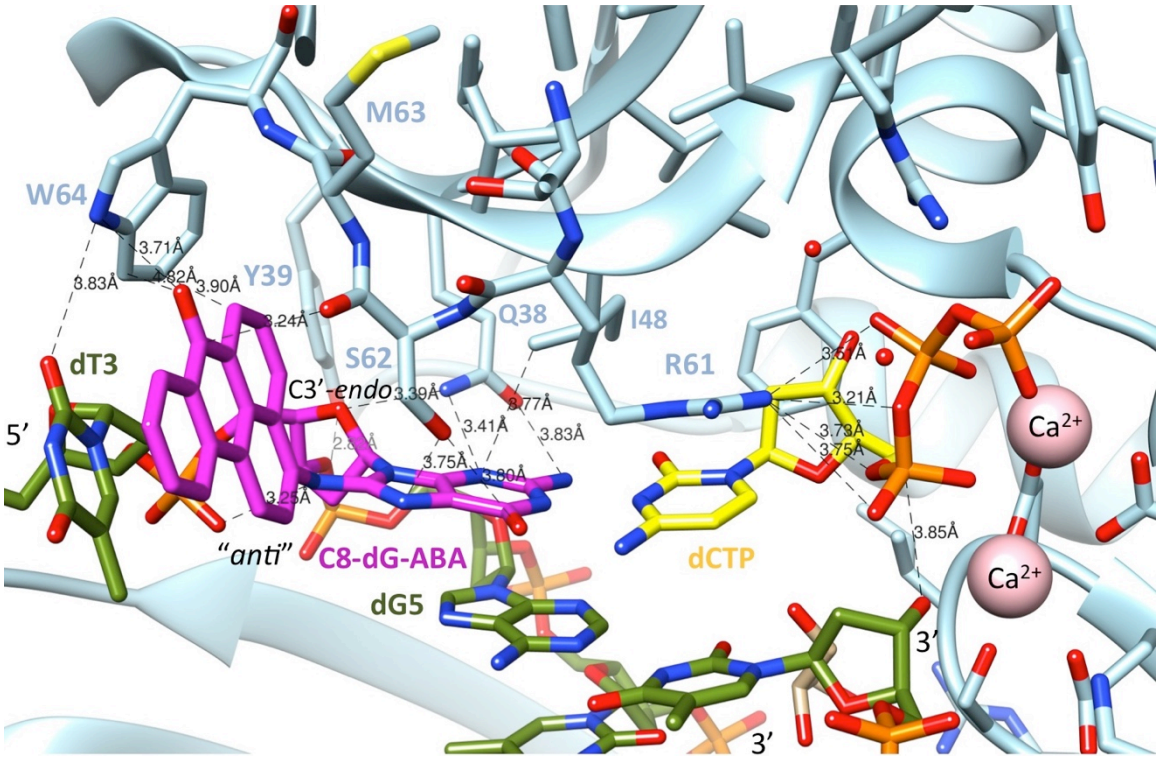



**Figure A5**: Close-up view of the active site in the crystal structure of the hPol η•C8-dG-ABA DNA•dCTP insertion complex. The color code matches that in Figure 3 (main paper) and selected distances in Å are indicated with dashed lines.