Fall 2019

Effects of the Dihydrouracil Lesion on DNA Using 1H/31P 1D and 2D Solution NMR

Benjamin M. Boyd
Missouri State University, Boyd666@live.missouristate.edu

EFFECTS OF THE DIHYDROURACIL LESION ON DNA USING $^{1}H/^{31}P$ 1D AND 2D SOLUTION NMR

A Master's Thesis

Presented to

The Graduate College of

Missouri State University

In Partial Fulfillment

Of the Requirements for the Degree

Master of Science, Chemistry

By

Benjamin Michael Boyd

December 2019

EFFECTS OF THE DIHYDROURACIL LESION ON DNA USING $^{1}H/^{31}P$ 1D AND 2D SOLUTION NMR

Chemistry

Missouri State University, December 2019

Master of Science

Benjamin Michael Boyd

ABSTRACT

The effects of the dihydrouracil lesion in DNA were studied using two dimensional NMR spectroscopy. The sequence used was based off of the Drew-Dickerson Dodecamer, with the cytosine in the three position replaced by a dihydrouracil. All of the nonexchangeable proton chemical shifts, with the exception of the H2, H5', and H5'', of the lesioned DNA were identified using NOESY spectra and then compared to the chemical shift values of the Drew Dickerson Dodecamer. The largest differences in chemical shifts were observed in the nucleotides neighboring the lesion, both within the strand and on the opposite strand. The imino exchangeable protons were also studied, and it was found that the lesion significantly decreases the local stability of double stranded DNA. The phosphorous backbone was also studied via HSQC and 1D Phosphorous NMR. The lesion was found to have a significant effect on the conformation of the phosphorous backbone of the DNA molecule. In the lesioned sequence, the 5' and 3' phosphates of the lesion site had a difference in % BII from the control sequence by 36.7 % and -27.2 %, respectively, at 25 ˚C. This trend was observed in all temperatures studied. In addition, the lesioned sequence also saw a large difference in the % BII between the 5' and 3' phosphates of the lesioned site. In the lesioned sequence, this difference was 85.2 % at 25 ˚C, compared to only 21.3 % in the control sequence. These results imply that changes in the phosphodiester backbone could be a potential area of indirect readout used by repair enzymes to identify the dihydrouracil lesion.

KEYWORDS: DNA, DNA damage, NMR spectroscopy, structural biology, dihydrouracil

EFFECTS OF THE DIHYDROURACIL LESION ON DNA USING $^{1}H/^{31}P$ 1D AND 2D SOLUTION NMR

By

Benjamin Michael Boyd

A Master's Thesis
Submitted to the Graduate College
Of Missouri State University
In Partial Fulfillment of the Requirements
For the Degree of Master of Science, Chemistry

December, 2019

Approved:

Gary Meints, PhD, Thesis Committee Chair

Mathew Siebert, PhD, Committee Member

Keiichi Yoshimatsu, PhD, Committee Member

Christopher Lupfer, PhD, Committee Member

Julie Masterson, PhD, Dean of the Graduate College

ACKNOWLEDGEMENTS

I would like to thank my research advisor, Dr. Gary Meints, for accepting me into his research group and teaching me everything I know about NMR spectroscopy. His mentorship has made me a more knowledgeable, competent, and confident scientist. I would also like to thank my thesis committee, Dr. Mathew Siebert, Dr. Keiichi Yoshimatsu, and Dr. Christopher Lupfer. Without their input, this thesis would not be nearly as good as it is. I would also like to acknowledge Dr. Gautam Bhattacharyya, for his encouragement and mentorship, particularly in my first semester, when I came into the department knowing nobody. I would also like to thank the other members of the Meints research group, particularly Koby Ljunggren and Jackie Becker, for their help and support. I would also like to thank my friends who were in the Chemistry Masters Program with me, particularly Molly Duszynski, Greg Illy, Michael Bakker, and Victoria Decocq. Without their support and comradery, I don't know if I could have made it to the end. And finally, I would like to thank my wife, Emily, who's love and support carried me to the very end.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: OBJECTIVE AND HYPTHESIS

The effects of the dihydrouracil lesion on DNA will be studied through the use of various NMR spectroscopic methods. The dihydrouracil lesion in this thesis is defined as the conversion of a cytosine residue into a dihydrouracil residue. Previous work implies that his lesion has similar effect as the uracil lesion, which is the spontaneous deamination of cytosine into uracil, particularly in the base pairing behavior of the two lesions.[1,2,3] In addition, the lesion has been described as highly mutagenic, both at the transcriptional and translation al level.[2,3] At time of writing, there is not much NMR-specific research on this particular lesion. The ultimate goal is to determine what repair factor repair enzymes recognize when they recognize the lesion. The hypothesis is that changes in the phosphodiester backbone of DNA are responsible for recognition by repair enzymes.

Changes in the phosphodiester backbone will be observed and quantified as a measure of %BII, which will be explained in detail later in the thesis. Previously, the majority of research utilizing %BII calculations has been done on non-lesioned DNA, which may be partially due to large, bulky lesions disrupting the local B form structure of DNA.[4] However, previous studies of the dihydrouracil lesion show that it retains a B form tertiary structure, making it eligible for %BII calculations.[5] In addition, %BII can be relatively straightforwardly calculated from experimental results, specifically the ^{31}P chemical shift of the individual phosphate.[4] Overall, the process begins with solving for the proton chemical shifts using a NOESY spectrum, then determination of the phosphorous chemical shifts from the proton chemical shifts using a HSQC spectrum. From there, the assignments can be made using just a 1D ^{31}P NMR spectrum. The hypothesis of the thesis is supported if there is significant difference in the %BII character of the

phosphates in the lesioned sequence versus the control sequence. In particular, the phosphates of interest are the ones close to the lesion site. The hypothesis is falsified if there is not a significant difference observed in the %BII character between phosphates of the control and lesioned sequences.

CHAPTER 2 DNA AND ITS BIOLOGICAL SIGNIFICANCE

2.1 Introduction

A cell's genetic information is stored using deoxyribonucleic acid, or DNA. This makes DNA maintenance a critical function for an organism's survival. Within the DNA strand, the information is stored using different nucleotide sequences that will ultimately code for different proteins and enzymes that the cell needs. However, this information can be altered through forces such as alkylating agents, free radicals, and errors in DNA replication. These events can happen as frequently as 10^6 times in a human cell in a single day.[1] To combat this, cells hve many different pathways for repairing damaged DNA sites.[1] An aspect of DNA repair that is still unclear is to how the repair enzymes are able to recognize when a DNA base is damaged. The focus of this research is what effect the dihydrouracil lesion has on the structure of DNA, and how these effects may be used by the repair enzyme to recognize this lesion.

2.2 DNA Structure

DNA is a polymer composed of a long strand of nucleotides.[6] An individual nucleotide is composed of three parts: the nitrogenous base, a pentose sugar ring, and a phosphate group.[6] The nitrogenous base is the part of the nucleotide that is used to store information.[6] There are two broad categories of DNA bases, purines and pyrimidines.[6] Purines are nine membered bicyclic rings, while pyrimidines are six membered rings.[6] In each of these groups, there are two specific bases typically found within healthy DNA; Guanine and adenosine are the major purines found in DNA, while cytosine and thymine are the major pyrimidines.[6] The numbering scheme for pyrimidines is one through six, starting with the carbon attached to the pentose sugar, and

continuing towards the ring nitrogen atom.[7] The numbering scheme for purines also starts with the six membered ring, with each of those atoms being labelled one through six and the remaining three atoms in the second ring being labelled as seven, eight, and nine (Figure 1).[7]

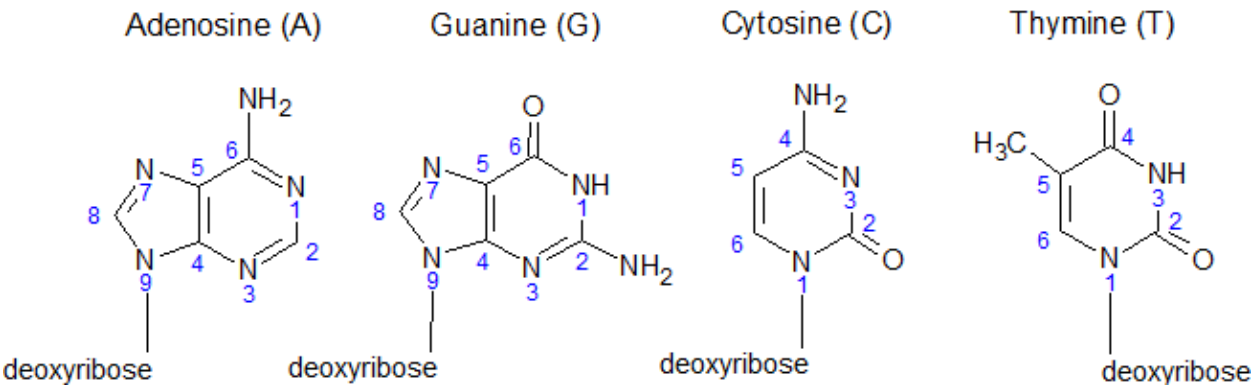

Figure 1: DNA bases in healthy DNA with numbering scheme

The pentose sugar ring of DNA is a 2-deoxyribose.[6] The deoxyribose distinction is important, as the pentose sugar for RNA contains a hydroxyl on the 2' ribose.[6] The numbering system for the pentose sugar begins with the carbon attached to the base as 1', and continues around the carbon chain to the carbon attached to the phosphate, which is labelled as 6'.[7] One thing that should be pointed out is that 2-deoxyribose has two protons on the 2' carbon, and as such they are labelled as 2' and 2", to the proton facing up and to the one facing down, respectively.[7] The bond that connects the 1' carbon of the deoxyribose to the 1 nitrogen of pyrimidines or the 9 nitrogen of purines is called the glycosidic bond (figure 2).[6] It is through the phosphate connected to the 5' carbon that allows nucleotides to chain together (figure 3).[6]

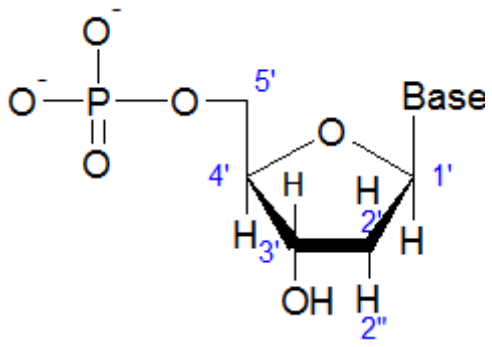

Figure 2: Chemical structure and numbering scheme of deoxyribose

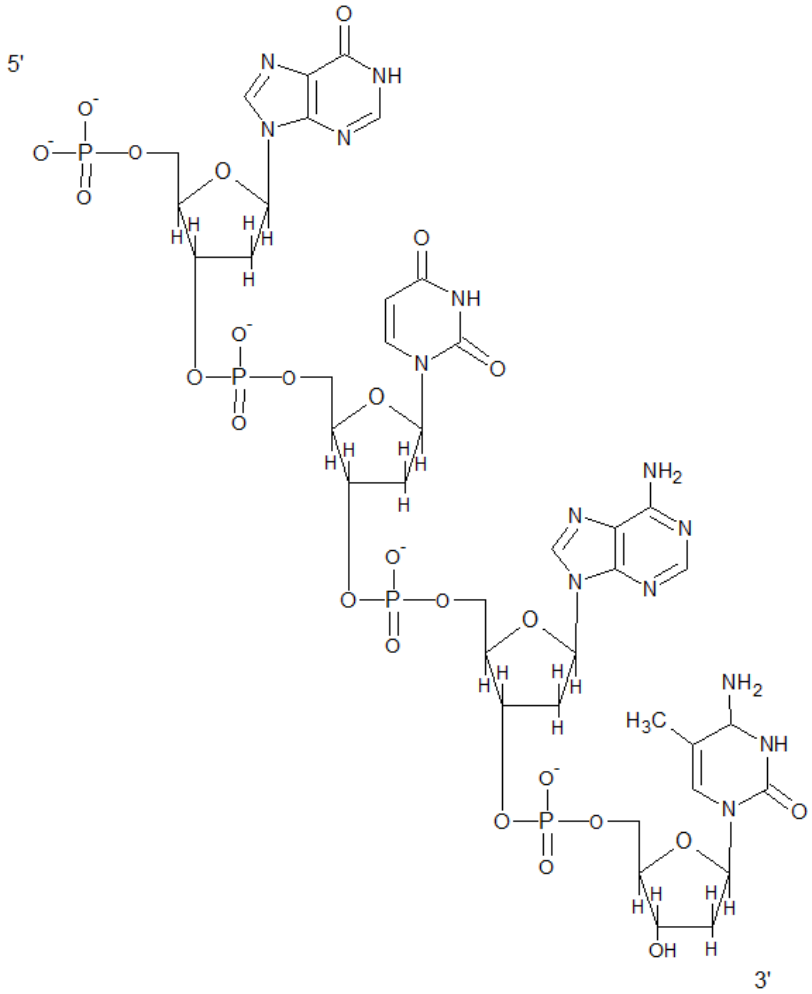

Figure 3: Depiction of a 4 nucleotide DNA strand, depicting the phosphodiester backbone

Successive nucleotides are connected via a phosphate group between the 5' hydroxyl of one nucleotide and the 3' hydroxyl of the next nucleotide (see figure 3).[6] This arrangement is called the phosphodiester linkage, as a phosphate group is connecting two monomers through two esters.[6] As these bonds are responsible for holding the individual nucleotides of a DNA strand together, it is of critical importance that these bonds be difficult to hydrolyze. Interestingly, the thermodynamics of the hydrolysis reaction of the phosphodiester linkage is fairly favorable, as it has a ΔG of 25 kilojoules per mole.[6] However, while DNA wouldn't be considered thermodynamically stable, they react slowly inside of the cell.[6] They are kinetically stable.[6]

Inside of cells, DNA typically exists as a double stranded helix, with the opposite strands running *anti*parallel to each other.[6] The dominant force that holds the two strands together is the hydrogen bonds between bases on the opposite strands of each other.[6] DNA bases also have a specific pairing partner base; adenine forms a base pair with thymine, and cytosine forms a base pair with guanine (Figure 4).[6] One may notice that the G:C base pair has three hydrogen bonds holding the bases together, while the A:T base pair only has two.[6] One consequence of this is that the C G base pair is more thermodynamically stable than the A:T base pair, and DNA strands with a large amount of C:G base pairs retain their double stranded character at higher temperatures than DNA strands with a large amount of A:T base pairs.[6] This base pairing scheme, with each pair comprising of one purine and one pyrimidine, makes the distances between base paring nucleotides relatively constant throughout the DNA molecule, with a rough distance of 1.1 nm between the 1' protons of base pairing deoxyriboses, which in turn means that DNA has a regular diameter throughout the molecule.[6]

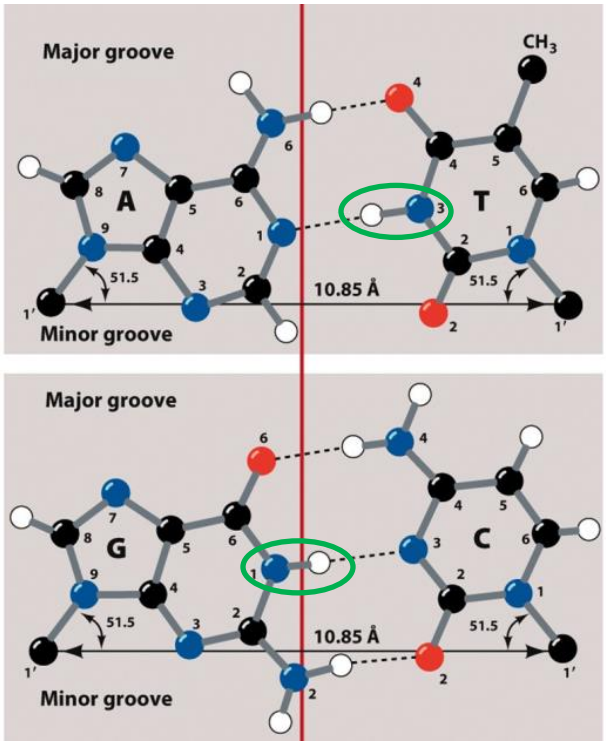

Figure 4: Watson-Crick base pairing between adenine and thymidine (top), and guanine and cytosine (bottom). Imino protons circled in green [8]

Another force that contributes to the stability of DNA's tertiary structure is the vertical stacking of bases within the structure.[6] Base stacking allows there to be strong van der Waals forces between the stacked bases, holding them together.[6] It should also be pointed out that one species of exchangeable protons, the imino protons, each contribute one hydrogen bond to their respective base pairs and are circled in green in the above figure.[9,10] When discussing DNA, an imino functional group is defined as a secondary amine, or HNR_2.[9] These protons are useful when studying DNA using NMR spectroscopy, as they can be used to assess the amount of fraying occurring at the ends of DNA and the stability of the double stranded structure.[9,10]

In addition to the Watson-Crick base pairing found in double stranded DNA, there are other modes of base pairing found in nucleic acids, ranging from base pairing schemes found in double stranded RNA or lesioned DNA to less common base pairing modes found between

canonical bases in double stranded DNA.[1,6,11,12] One of the non-Watson-Crick base pairing schemes of interest is wobble base pairing. A wobble pair is a base pair between a guanine and a pyrimidine, for example uracil, where instead of forming three hydrogen bonds from the guanine's 1N-H, $2NH_2$, and 6C=O, only two hydrogen bonds are formed, one from the 1N-H and one from the 6C=O (see Figure 5).[12]

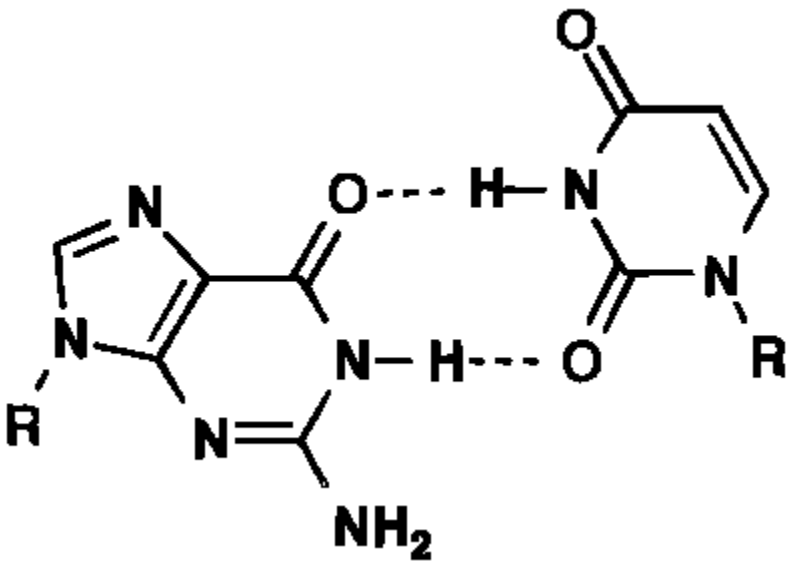

Figure 5: Guanine-Uracil Wobble Base Pairing [13]

This type of base pairing occurs due to the mismatch in dipoles that would occur if a Watson-Crick base pair was attempted between these two bases; this causes the pyrimidine to move, or "wobble", in order to create the correct angle to form two hydrogen bonds with the guanine.[6,12,13] In addition to different base pairing modes caused by mismatched bases, the way bases pair together can also be influenced by the conformation of DNA between traditional Watson-Crick base pairs.[14] Typically, the base exists in the *anti* conformation in relationship to

the deoxyribose sugar; however, it is possible for the base to be in an *syn* conformation to the pentose (see Figure 6).[14,15]

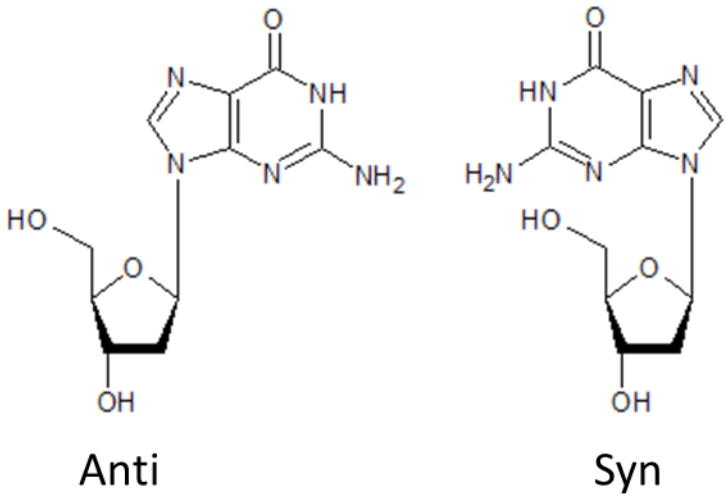

Figure 6: *Syn* and *anti* conformations of deoxyguanosine

The *syn* and *anti* conformations are defined by the torsion angle χ, which is defined as O5'-C1'-N9-C4 for purines and O5'-C1'-N1-C2 for pyrimidines.[15] The *anti* conformation of nucleotides is more stable than the *syn* conformation, and as such a nucleotide is more likely to exist in the *anti* conformation.[14,15] Watson-Crick base pairing occurs between a purine and pyrimidine in the *syn* conformation.[14,15] However, it is still possible to form a base pair between a pyrimidine in the *anti* position and a purine in the *syn* position, known as Hoogsteen base pairing (Figure 7).[14,15]

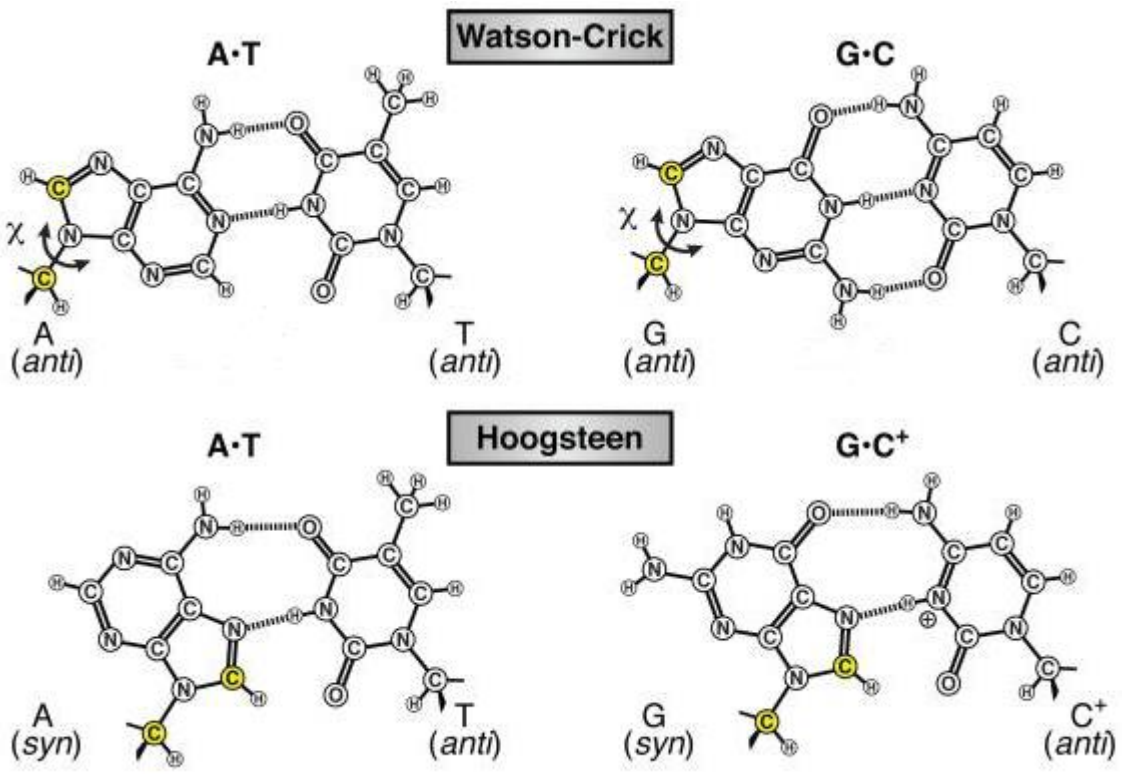

Figure 7: Depiction of Watson-Crick *vs* Hoogsteen base pairing, χ angle highlighted [15]

One should notice that in a Hoogsteen base pair between guanine and cytosine, the N3 atom must become protonated to form the necessary partial positive charge to form a hydrogen bond with guanine N7.[15] This necessity makes the guanine cytosine Hoogsteen base pair dependent on pH, while the adenine thymine Hoogsteen base pair is not.[15] In addition, this type of base pairing only occurs with two hydrogen bonds, even if it is a base pair between a guanine and a cytosine, which typically has three hydrogen bonds through Watson-Crick base pairing.[6,15] Another consequence of this type of base pairing is Z-DNA, which has a helix with a left hand turn instead of the typical right hand turn of other forms of DNA.[6]

In cells, DNA typically exists in one of two forms, either A or B form helices.[6] B form is the most common form for double stranded DNA inside of cells (see Figure 8).[6]

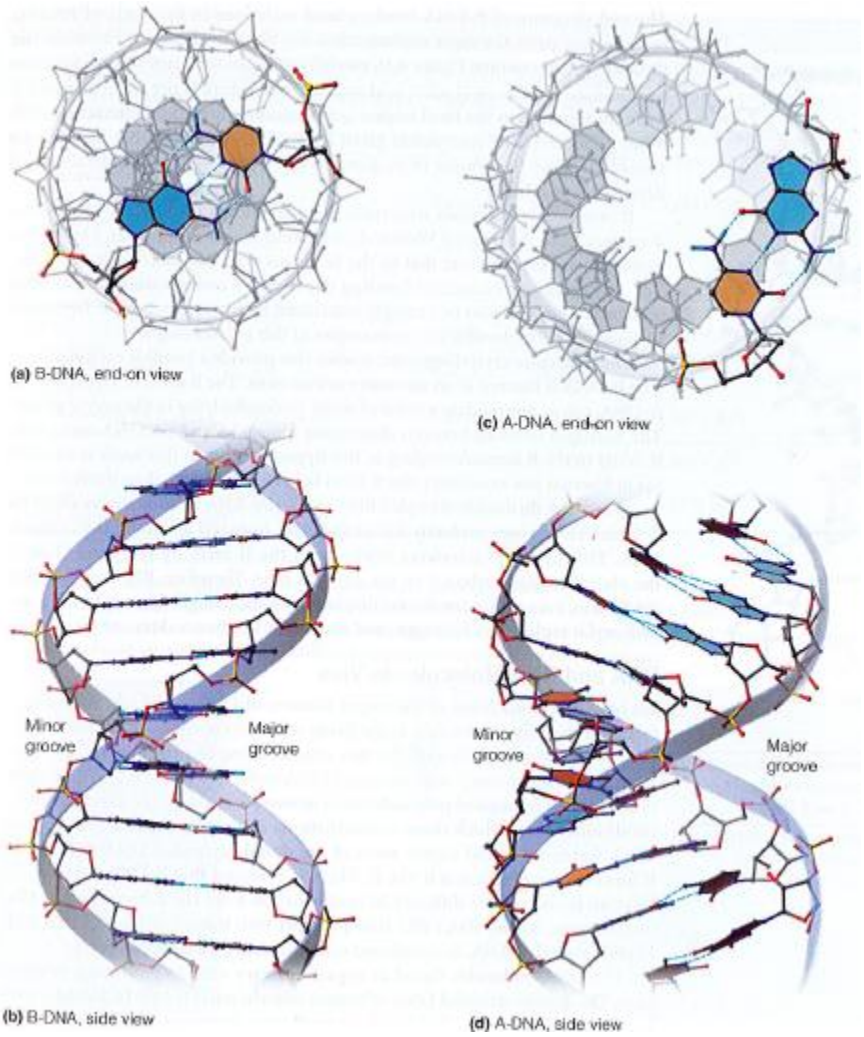

Figure 8: Visual depiction of B DNA end on (top left) and side view (bottom left), as compared with A DNA end on (top right) and side view (bottom right) [6]

While double stranded DNA is not normally present in the A form in cells, double stranded RNA always exists in the A form, as do RNA-DNA hybrid molecules.[6] In B form DNA, the bases lie closely to the helix axis, and the major and minor groove are of radically different widths.[6] In A form DNA, the bases lie further out from the helix axis, and the major and minor groove are very similar in size.[6] Within B form DNA, the conformation of DNA can

actually be further defined into the BI and BII conformations, which will be expanded upon in the Local Structure and Dynamics section.

2.3 DNA Damage

In living human cells, DNA is under a near constant assault from mutagenic stimuli, whether it be from outside factors such as free radicals created from UV radiation or methylating agents, or from consequences of the natural conditions inside of the cell, such as deamination or hydrolysis.[1] Hydrolysis of the phosphodiester bond of purine residues occurs about 18,000 times per day within the cell, causing the nucleotide to be lost.[1] Deamination occurs between 100 to 500 times per day within mammalian cells.[1]

A damaged nucleotide in DNA (either by being chemically altered, replaced or deleted all together) is called a lesion.[1] Lesions represent a threat to the healthy function of a cell. If an RNA molecule is transcribed from that lesioned DNA, and a protein is translated from that RNA, it can result in a protein that is unable to perform its intended function, wasting a cell's resources and in some cases actively damaging the cell.[1,6] If the lesion isn't corrected, it can be incorporated into the daughter strand of the DNA when it is replicated.[1,6] When a lesion is replicated, it becomes a mutation.[1]

One type of DNA damage that is of interest is the incorporation of uracil into a DNA molecule (Figure 9).

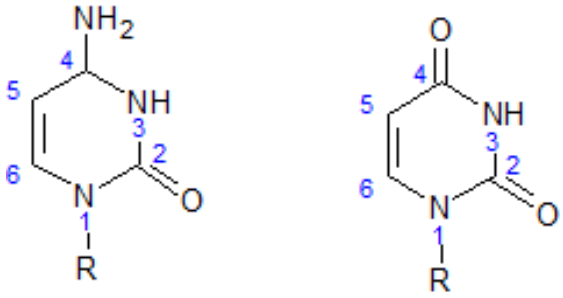

Figure 9: Structure and numbering scheme of cytosine (left) and uracil (right)

As shown above, the main difference in structure between uracil and cytosine is the replacement of the amine with a ketone. Typically, uracil is found in RNA molecules where uracil replaces thymine in its naturally occurring bases.[6] There are multiple mechanisms through which uracil can spontaneously be formed, one of which is attack by a hydroxide ion (Figure 10).[1]

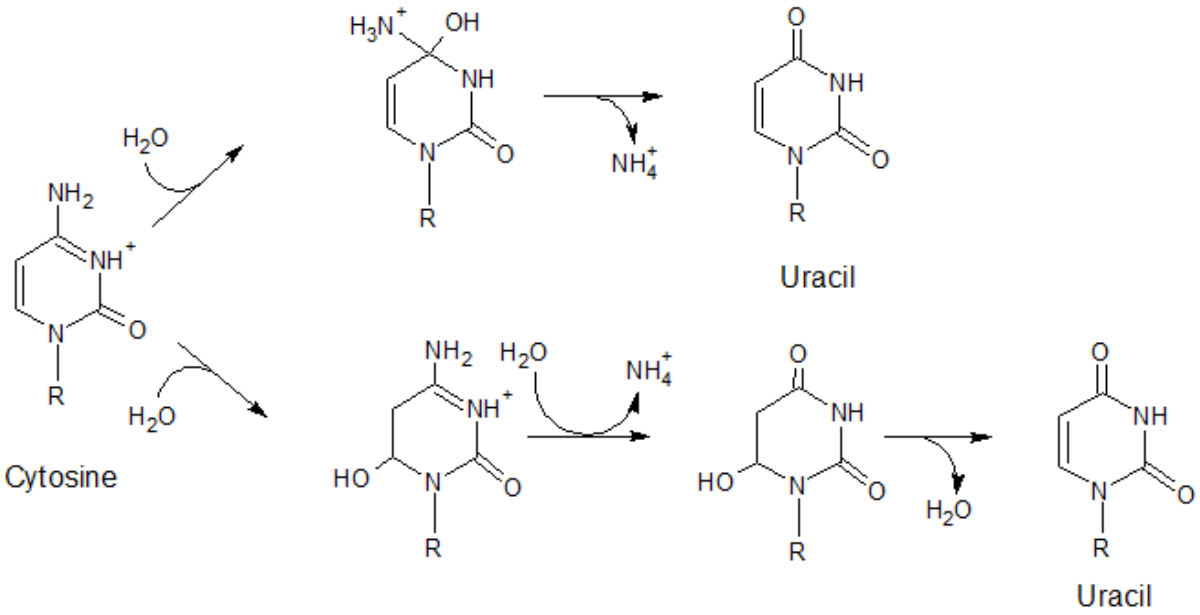

Figure 10: Formation of Uracil from Cytosine via attack by hydroxyl

As shown above, the hydroxyl ion can either attack at the 4 position or the 6 position of the pyrimidine ring.[1] The attack at the 4 position is a direct attack, and happens relatively quickly so as the intermediate does not exist for a significant amount of time.[1] The other method outlined involves an attack at the 6 position on the pyrimidine, and through the course of the reaction passes through dihydrocytosine as an intermediate.[1]

As stated above, in RNA, uracil replaces thymine as a canonical base.[6] One consequence of this is that the base pairing partner of uracil is adenine.[6] However, as shown above, uracil typically arises in DNA by the deamination of cytosine. When a DNA strand containing a uracil lesion is replicated, the daughter strands each have a different base pair at the location the lesion was at (Figure 11).[16]

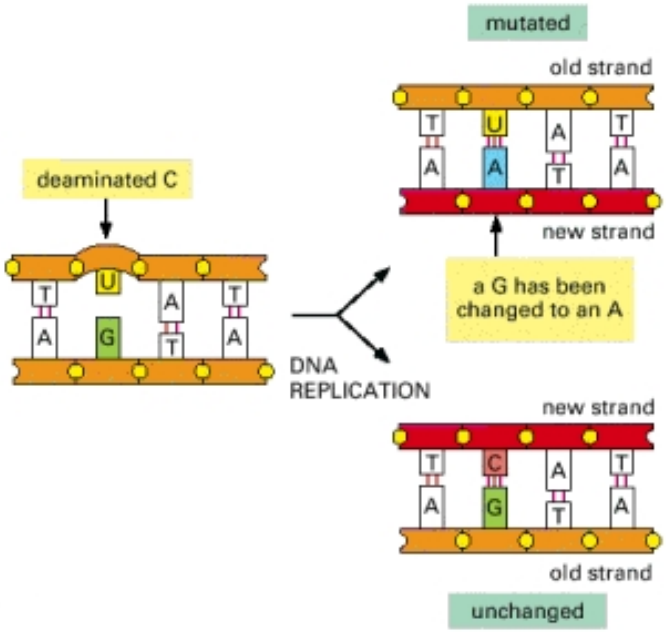

Figure 11: Depiction of a uracil lesion becoming a mutation. The daughter strand containing the guanine codes for an unchanged DNA strand, while the daughter strand containing the uracil now has an adenine where there was previously a guanine. [16]

As shown above, only one of the daughter strands of DNA is mutated, while the other strand is unchanged, as its lesioned pair has separated to form another double stranded DNA molecule.[16] However, the presence of the uracil in the mutated DNA strand causes it to be base paired with adenine, rather than its usual base pairing partner, guanine.[16]

In addition to uracil, the formation of dihydrouracil is of interest to this research. Dihydrouracil lesions can arise naturally in cells if the deamination of cytosine occurs in the presence of ionizing radiation, the mechanism of which is shown below (Figure 12).[17]

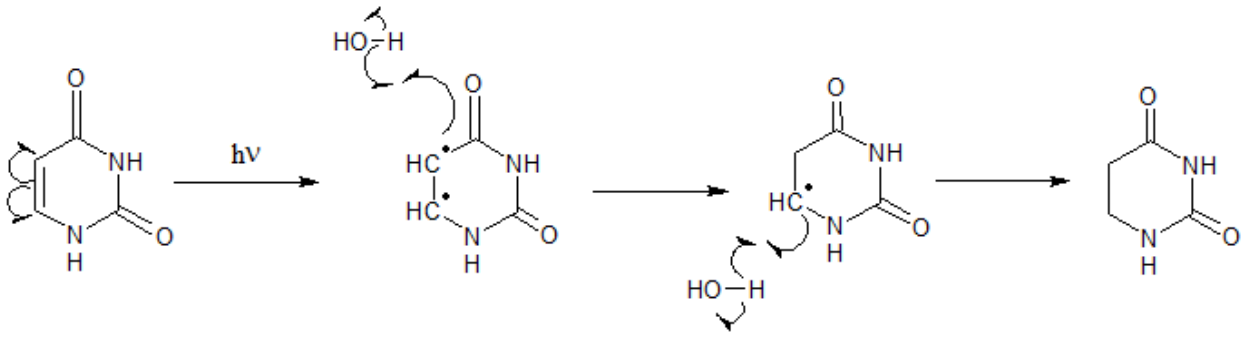

Figure 12: Formation of Dihydrouracil from Uracil

Previous research has hypothesized that dihydrouracil forms do to uv radiation exciting the pi electrons from the singlet ground state to a singlet excited state.[17] From there, intersystem crossing occurs, producing a uracil in the triplet excited state.[17] From there, a free radical addition occurs between two proton donors, in this case it is assumed to be water, leaving the dihydrouracil base.[17,18]

Like uracil, dihydrouracil's Watson-Crick base pairing partner is adenine, though occasionally it will pair with guanine using a wobble base pair mechanism.[2] Because of its base pairing with adenine, it is considered highly mutagenic, both during transcription and during replication.[2,3,5]

2.4 DNA Repair

Because of the constant stress DNA is under, DNA repair mechanisms are of critical importance to an organism's survival. Because of this, cells have multiple ways to repair a lesion site, including direct repair, base excision repair, nucleotide excision repair, and mismatch repair.[1,6] Here, the primary focus will be on base excision repair (BER), as base excision repair is the primary mode of correcting uracil and dihydrouracil damage in humans.[1,5]

Base excision repair is initiated by DNA glycosylases, a class of enzymes that catalyzes the hydrolysis of the *N*-glycosidic bond of lesioned bases.[1] Typically, each of these glycosylases only recognize a specific type of DNA damage, usually one particular lesion base or mismatched base pair.[1] The specific enzyme that catalyzes this bond in uracil is called uracil-DNA-*N*-glycosylate, or UNG.[1,6] DNA glycosylases move along the minor grove of the DNA helix until they recognize their specific type of damage that they address, in this case the uracil lesion.[1] Once uracil is recognized by UNG, the enzyme kinks the DNA and causes the uracil lesion to base flip, and places the uracil base into the uracil-binding pocket of UNG(Figure 13).[1] This binding pocket is extremely selective towards uracil; one means of selection is the exclusion of purines through size, as the pocket is only large enough to fit a pyrimidine.[1] 5-methylated pyrimidines, for example thymine, are excluded by the Tyr147 residue, and uracil affinity is greatly strengthened through the hydrogen bonding between the base and the Asn204 residue.[1] The analog to this in the BER of dihydrouracil is NTH1.[19]

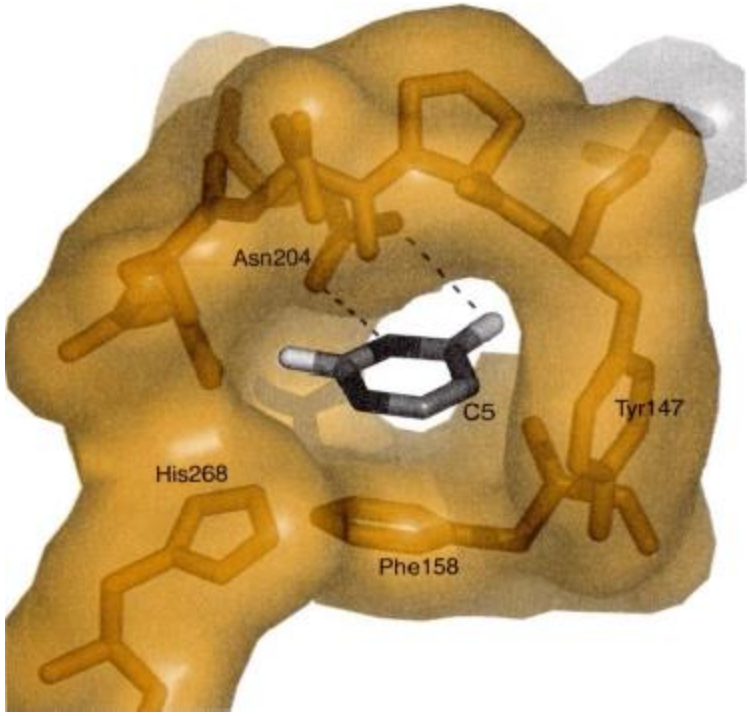

Figure 13: Uracil-binding pocket of human UNG with uracil base inside [1]

Once the uracil has been flipped into the active site of UNG, the *N*-glycosidic bond is then hydrolyzed.[1] In UNG, this reaction is thought to be catalyzed specifically by the Asp145 residue, which activates a water molecule by removing a proton from it[1]. In addition, the hydrolysis reaction is further stabilized by the His268, which helps stabilize the uracil when it is in its transition state as the *N*-glycosidic bond is being cleaved.[1] Once this bond is cleaved, the uracil base is released into the solution, leaving behind an abasic site in the DNA strand.[1,6] In addition to the ability to cleave the *N*-glycosidic bond, often times these enzymes can also function as a DNA lyase, able to cleave the phosphodiester bond 3' to the abasic site.[1,19]

After the abasic site is formed, the BER process can proceed by one of two different paths: short patch or long patch (Figure 14).[19] In short patch repair, only the lesioned nucleotide is replaced, while in long patch repair anywhere from 2 to 10 nucleotides are

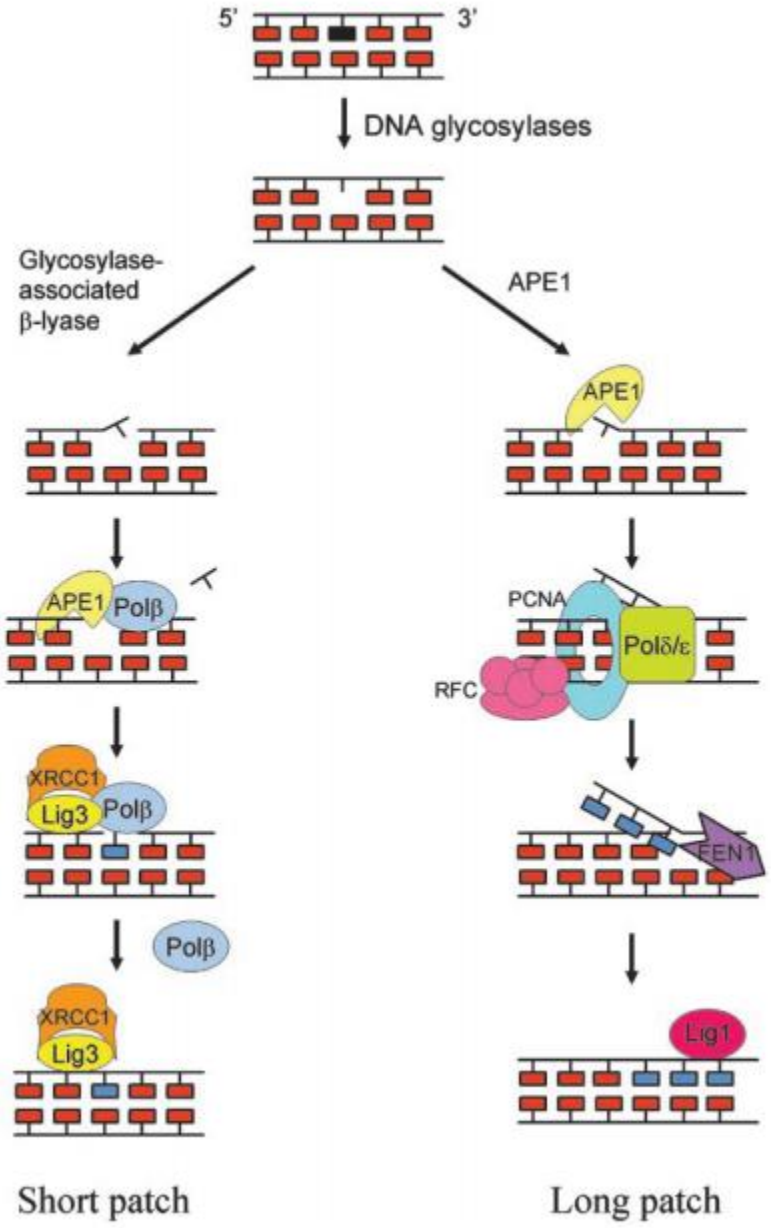

Figure 14: BER, long and short patch [19]

replaced.[19] In short patch repair, if the DNA glycosylase lacks lyase activity, after the abasic sight is formed, an incision in the DNA strand is made by hydrolyzing the phosphodiester bond

in the 5' direction of the abasic sight.[1,19] From there, DNA polymerase β is recruited and cleaves the phosphodiester bond 3' to the abasic site and then inserts the correct nucleotide into the gap, which is then sealed by DNA ligase.[1,19,20] If the DNA glycosylase does in fact have β-lyase activity, then the 5' phosphodiester bond is cleaved by AP endonuclease, creating a gap which is then filled by DNA polymerase β, which from there proceeds in the same way. In long chain BER, the process begins the same way with AP endonuclease cleaving the 5' phosphodiester bond.[19] However, after the initial cut the processes begin to differ. DNA Polδ/ε, PCNA, and FEN1 are recruited to displace the nucleotides 3' of the nick, producing a flap of DNA between two to ten nucleotides long.[19] From there, FEN1 will cleave the DNA at the single to double stranded DNA junction, releasing the flap.[19] DNA Polδ/ε then synthesizes the replacement DNA patch, which is then incorporated into the DNA by DNA ligase.[19]

2.5 DNA Local Structure and Dynamics

When studying DNA, it is important to keep in mind that it is not a static structure; rather, DNA is a dynamic structure that has its own dynamic conformation, similar to proteins.[6] As stated above, double stranded DNA has two conformations it can exist in inside of the cell, A form and B form.[6] However, within B form DNA, there are two backbone conformations that the DNA can have, defined as BI and BII.[21] The conformational states are defined through the torsion angles of the phosphodiester backbone.[21] In particular, the $C_{2'}$-$C_{3'}$-$O_{3'}$-P and the $C_{3'}$-$O_{3'}$-P-$O_{5'}$ torsion angles, named ε and ζ, respectively, are important as they are used to define the BI and BII conformational states. In the BI conformational state ε is trans while ζ is gauche, while in the BII conformational state ε is gauche while ζ is trans (Figure 15).[21]

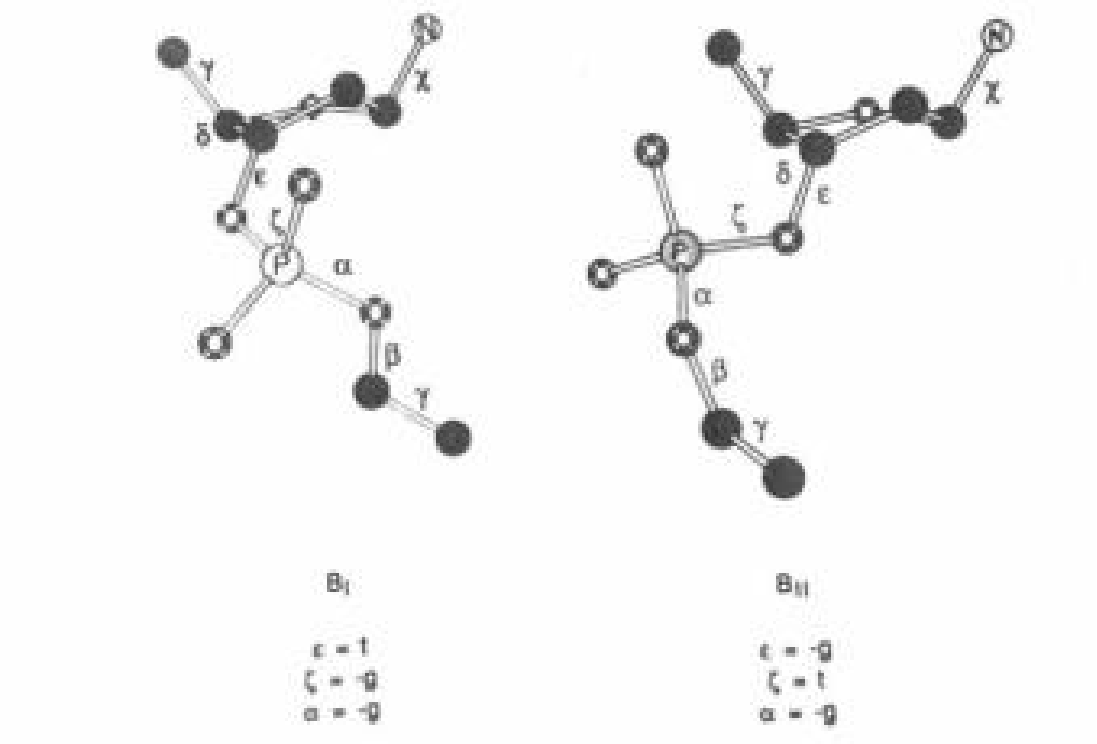

Figure 15: Visual depiction of BI(left) and BII(right) conformational states [21]

In addition to the difference in torsion angles, there is also a difference in the overall shape of the DNA molecule between BI and BII conformations.[22] The minor grove of BI DNA is fairly narrow, while the minor grove of BII DNA is larger (Figure 16).[22]

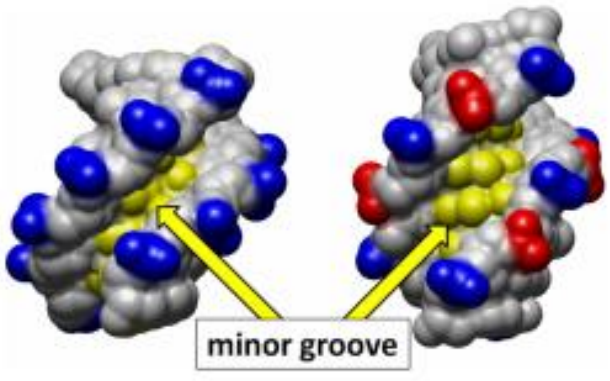

Figure 16: BI (left) and BII (right) DNA, with the minor grove highlighted in yellow[22]

It should be said that very rarely will DNA exist fully in either the BI or BII conformation; rather, DNA exists as an equilibrium between the two conformational states.[21] This equilibrium can be calculated as a percentage of BII character, or %BII[4,24]. How this is done will be explained in the 1D NMR section of this thesis.

CHAPTER 3: NUCLEAR MAGNETIC RESONANCE

3.1 Introduction to NMR

NMR, or Nuclear Magnetic Resonance, spectroscopy is a powerful characterization tool utilized by chemists to ascertain chemical and structural information[25] Two dimensional NMR spectroscopy is particularly useful to biochemists, as it can be used to identify the tertiary structure of macromolecules. Also, NMR spectroscopy offers significant advantages over crystallography methods, the other major technique used to study the structure of macromolecules. Crystallographic studies, while they are often easier to solve than NMR experiments in terms of solving tertiary structures, must take the macromolecule of interest out of its native environment.[26] This introduces the risk that the structure presented by the crystal is not accurate to the structure of the macromolecule in solution.[26] NMR offers the advantage of being able to study biological molecules in aqueous solution, which is much closer to how these molecules exist in their natural state.[27]

NMR spectroscopy is possible because atoms behave as if they are tiny magnets.[25] When atoms are placed inside of a strong external magnetic field, they will try to align with it.[25] The internal magnetic moments of the nuclei are being manipulated by the external magnetic field.[25] A radio frequency is then applied to the sample and varied until the nuclei in the sample absorb the radio energy.[25]

3.1.1 Nuclear Spin NMR spectroscopy is dependent on the concept of nuclear spin.[25] The nucleus of an atom contains protons, giving it a positive charge, and can be thought of as to be spinning on an axis.[25] This rotational property is what causes the atom to have its own magnetic field.[25] In addition to this, an atoms nuclear spin(I) is essentially how a nucleus

responds to a magnetic field, and also determines the number of quantum states available to an atom[25,28].

While all atoms have spin, only certain spin numbers produce signals in NMR spectroscopy.[25,28] Atoms with an even mass number and an even number of protons and neutrons have a spin number of 0, and as such have no magnetic properties or available quantum states.[25,28] Nuclei with an odd mass number tend to have a spin number of 1/2, and have two quantum states: 1/2 and -1/2.[25,28] Atoms with a 1/2 spin number tend to give strong, resolved signals when compared to atoms with other spin numbers, and as such much of the research performed using NMR are done using spin-1/2 nuclei.[25,28] Atoms with an even mass number and an odd number of protons and neutrons tend to have spin numbers greater than 1/2, and as such will have more than two quantum states available for the nuclei.[25,28] It is necessary to point out that this is just a trend, and not a law. For example, ^{17}O has a spin number of +5/2, which violates this rule.[25] Nuclei with these larger spin numbers have nuclear quadrupole moment, which causes them to produce very broad peaks, making their spectra difficult to interpret.[25,28]

3.1.2 Static Magnetic Field and Larmor Frequency Because of the inherit magnetic properties of the nucleus, when placed inside of a large magnetic field, the nucleus will want to orient itself with the magnetic field.[25] However, because nuclei have angular momentum, they are unable to shift their angle to be fully aligned with the magnetic field.[25,28] Instead, the nucleus acts like a spinning top or gyroscope, where torque applied on the nucleus from the external magnetic field causes the nucleus's spin axis to precess around the general magnetic field direction (Figure 17).[25]

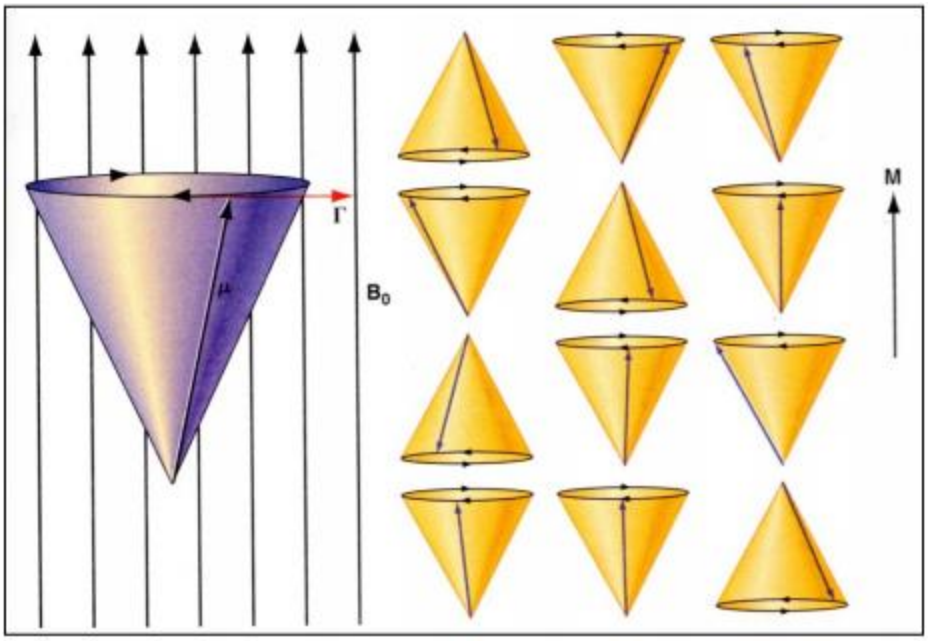

Figure 17: Larmor Precession, aligned with, α, or against, β, an external magnetic field [29]

The precession rate of the nucleus, also known as the Larmor Frequency when expressed in Hertz, can be related to the external magnetic field through the equality

$$v_0 = \frac{\gamma \beta_0}{2\pi}$$

where v_0 represents the Larmor frequency, β_0 represents the external magnetic field strength in Tesla, and γ represents the strength of the nuclear magnet, also known as the magnetogyric ratio.[25] It should be pointed out that because of this relationship, the Larmor frequency of a particular nucleus is proportional to the strength of the magnetic field interacting with it.[25] However, within a certain B_0, this Larmor frequency is the same for all nuclei of a certain type in isolation.[25,28] This phenomenon is well studied enough that the strength of an NMR is expressed through what the Larmor frequency of an isolated proton would be within the instrument.[25] For example, for a 9.76 Tesla NMR, the Larmor frequency of a proton would be 400 MHz.[25,28]

As shown in Figure 17 above, nuclei with a spin of 1/2 can only exist in certain orientations with respect to the magnetic field; they can exist in the α alignment (aligned with the magnetic field) or in the β alignment (aligned against the magnetic field).[29]

3.1.3 Nuclear Magnetic Resonance and Transverse Magnetism The alignments, or spin states, described above are of the upmost importance to NMR spectroscopy, as it is through exciting a nuclei from the α to the β spin state that an NMR spectrum is created.[29] The energy needed to promote a nucleus from the α state to the β state is defined as

$$\Delta E = h\nu_0$$

where h is Planck's constant and ν_0 is the Larmor Frequency as defined above.[25] Typically, this energy comes from a radio wave photon.[25,28,30] Another caveat to NMR is that it is a quantum phenomenon, so the energy of the photon must match the energy required to transition exactly for the transition to occur.[25] The ratio of nuclei in the α spin state and nuclei in the β spin state is given by the Boltzmann distribution,

$$\frac{N_\alpha}{N_\beta} = e^{-\Delta E/kT}$$

where N_α and N_β are the number of atoms in the higher and lower energy spin states, respectively, T is temperature, k is the Boltzmann constant, and ΔE is the difference in energy between the two spin states.[28] The larger this population difference, the stronger the signal will be[28].

3.2 NMR Interactions

NMR spectroscopy relies on many different interactions that are applied to the system in order to produce a spectrum.[25,28,31] These interactions can be broken up into two categories, external and internal interactions, depicted by the nuclear spin Hamiltonian, shown below

$$\hat{H} = \hat{H}_{ext} + \hat{H}_{int}$$

where $\hat{H}_{ext}$ represents the external interactions and $\hat{H}_{int}$ represents the internal interactions.[31] The external interactions hamiltonian can be broken up further, shown below

$$\hat{H}_{ext} = \hat{H}_{Zeeman} + \hat{H}_{RF}$$

where $\hat{H}_{Zeeman}$ represents the Zeeman Hamiltonian and $\hat{H}_{RF}$ represents the induced radio frequency field, both of which will be expanded on below.[31] In addition, the internal interactions Hamiltonian can be broken up into multiple interactions, shown below

$$\hat{H}_{int} = \hat{H}_{Dipolar} + \hat{H}_{J-coupling} + \hat{H}_{Chemical\ Shift} + \hat{H}_{Quadropolar}$$

where $\hat{H}_{Dipolar}$ is the dipolar coupling Hamiltonian, $\hat{H}_{J\text{-}coupling}$ is the J-coupling Hamiltonian, $\hat{H}_{chemical\ \text{shift}}$ is the chemical shift Hamiltonian, and $\hat{H}_{Quadropolar}$ is the quadropolar coupling Hamiltonian.[31] All of these interactions will be explained below.

3.2.1 Zeeman Hamiltonian The Zeeman Hamiltonian represents the interactions of the external magnetic field with the nuclear spin of the nuclei, and is represented mathematically below.[25]

$$\hat{H}_{Zeeman} = -\gamma B_0 \hat{I}$$

In this equation, $\hat{H}_{Zeeman}$ represents the Zeeman Hamiltonian, γ represents the strength of the nuclear magnet, $\hat{I}$ represents the nuclear spin of the nucleus, and B_0 represents the strength of the external magnetic field.

3.2.2 Radio Frequency Hamiltonian In addition to the external magnetic field, the other external force being exerted on the NMR sample is the radio frequency, or RF pulse.[25,28] These RF pulses create a new magnetic field, B_1, that is weaker and perpendicular to the B_0 magnetic field.[25] In addition, RF pulses are used to manipulate spin states and population ratios between the two spin states that is outlined above.[25,28]

3.2.3 Dipolar Coupling Hamiltonian In addition to the interactions between nuclei and the external magnetic field, there can also be interactions between nuclei within the NMR sample, typically in the form of an energy transfer.[28] This energy takes the form of electromagnetism, and when this energy is transferred through space it is called dipolar coupling.[28] As a corollary, this means that the magnetic environment of any individual nucleus is affected by the magnetic environments of any other nuclei around it.[25,28] In addition to the external magnetic field generated by the NMR instrument, individual nuclei also have their own magnetic field.[25] While this magnetic field is much less powerful than the external field, the magnetic field of one nucleus can affect the magnetic field of another, if the two nuclei are physically close enough.[25] Despite the constant tumbling of the molecules in solution, the orientation of each nuclei's magnetic field will remain aligned with the external magnetic field.[25] This gives some directionality to how the magnetic fields of different nuclei affect each other; if one nucleus has another nucleus either above or below it, it will add its magnetic field to the external magnetic field.[25] However, if the two nuclei are arranged side by side, they will detract from the external magnetic field.[25] This type of interaction becomes stronger the closer two nuclei are to each other in space, and forms the basis of the NOESY technique, which will be explained later.[28]

3.2.4 J-Coupling Hamiltonian When the energy transfers between two or more nuclei travel through bond, it is called J-coupling.[25,28] J-coupling can be easily observed in a 1D proton spectrum, in the form of peak splitting (Figure 18).[25]

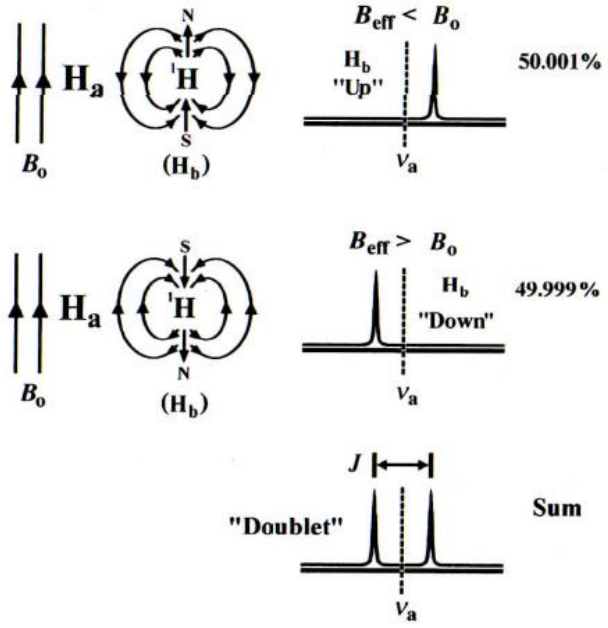

Figure 18: Visual depiction of splitting and J-coupling between two protons[25]

The example in the previous figure depicts two protons, H_a, whose magnetic field is being affected by H_b .[25] However, the effect exerted by H_b depends on its orientation; it can either subtract from the external magnetic field if it is in the "up" orientation, or add to it if it is in the "down" orientation.[25] Because each H_b nuclei in the entire sample has a roughly 50/50 chance of being in either the up or down orientation, the H_a peak is split into two peaks of equal intensity.[25] The separation between the peaks in Hz is known as the coupling constant.[25] This effect is felt mutually between both protons; that is, H_b will also experience splitting because of the effect H_a is having on its magnetic environment.[25] Typically, J-coupling only occurs over a relatively short range of 3 bonds or fewer, though there are some instances where this range can be expanded.[25,28]

The example above only took into account a two proton system. If there are multiple protons within coupling range of each other, more complicated splitting patterns will appear.[25] The splitting pattern can be predicted mathematically through the equation

$$splitting = 2nI + 1$$

where *I* is the spin number of the nuclei in question and n is the number of nuclei that are within J-coupling range, though for protons and other *I*=1/2 nuclei, this is often simplified to the n+1 rule, where the splitting is equal to the number of neighboring protons plus one.[25,32]

3.2.5 Chemical Shift Hamiltonian As stated above, all isolated nuclei of the same species should have the same Larmor frequency when inside of a constant magnetic field.[25] However, this isn't true for nuclei inside of a chemical system or molecule. The magnetic field experienced by a particular nucleus can be changed by its chemical or magnetic environment outside of the external magnetic field.[25] This change in the experienced magnetic field is caused by the electron clouds that circulate throughout the molecule; electron clouds can circulate within the external magnetic field, in turn creating their own magnetic field.[25] This new magnetic field can either add to or subtract from the external field, though typically it is thought to subtract from, which is why this phenomenon is known as shielding.[25] This new field experienced by the nucleus is known as the effective field, which can be expressed mathematically as

$$B_{eff} = B_0(1 - \sigma)$$

where B_{eff} is the effective magnetic field, B_0 is the external magnetic field, and σ is the shielding constant.[25] B_{eff} can then be plugged in to the Larmor frequency equation, giving the following equality

$$v_0 = \frac{\gamma B_{eff}}{2\pi} = \frac{\gamma B_0(1 - \sigma)}{2\pi}$$

where v is the new resonance frequency of the nucleus.[25,30] However, NMR spectra are expressed as a function of chemical shift, not as a function of frequency. Chemical shift can be defined mathematically as

$$\delta = \frac{v - v_0}{v_0} * 10^6$$

where δ is the chemical shift, v is the resonance frequency of the nucleus in question, and v_0 is the resonace frequency of some reference compound, typically tetramethylsilane, or TMS.[28,30] While the chemical shift should technically be unitless when looking at the mathematics, it is given the unit of parts per million, or ppm, because the value is multiplied by 10^6, or one million.[28] The advantage of using chemical shift instead of resonance frequency when creating an NMR spectrum is that the resonance frequency of a specific nucleus will be different depending on the strength of the external magnetic field, while the chemical shift of the nucleus will remain constant, no matter the strength of the magnetic field used.[28]

3.2.6 Quadrupolar Coupling Hamiltonian As stated above, the nuclear spin is what determines a nucleus's quantum properties when placed inside of a large magnetici field.[25] If a nucleus's spin number is greater than 1/2, it is considered a quadrupolar nucleus.[33] While spin 1/2 nuclei typically evenly distribute their positive charge in a spherical shape, quadrupolar nuclei have an uneven, asymmetrical charge distribution.[33] Quadrupolar nuclei also have more than two quantum energy levels available to them when placed inside of the NMR's magnetic field.[25] While typically quadrupolar nuclei are able to be efficiently relaxed in solution NMR, this causes the peaks to become broader than if the quadrupolar nuclei were absent, making these spectra more difficult to interpret.[33]

3.2.7 Isotropic Hamiltonian Summary In an isotropic sample, the effects of the quadrupolar coupling Hamiltonian will average out to zero, so the Hamiltonian interactions that are of importance can be simplified to

$$\hat{H} = \hat{H}_{Zeeman} + \hat{H}_{RF} + \hat{H}_{Dipolar} + \hat{H}_{J-Coupling} + \hat{H}_{CS}$$

where the Hamiltonians are the same as the ones outlined above.[33]

3.3 One Dimensional NMR

Arguably the most common elements analyzed using NMR are carbon and hydrogen. Proton NMR spetra are relatively easy to perform, as protons have a spin ½ and a natural abundance of near 100% of all hydrogen atoms, meaning that a sample can be analyzed using proton NMR without any isotopic enrichment.[25] The NMR active nuclei of carbon, ^{13}C, only has a natural abundance of 1.1% of all carbon atoms, causing carbon NMR spectra to typically have much weaker signals than proton NMR spectra if not isotopically enriched.[25] This study also utilizes phosphorous NMR. ^{31}P, the NMR active isotope of phosphorous, has a natural abundance of near 100%, so no isotopic enrichment was needed.[25]

Phosphorous NMR is particularly useful for this study as the %BII character of a phosphate within the DNA backbone can be calculated directly from the chemical shift.[4,24] In this study, the method outlined by Tian et al was utilized. For this method, it is assumed that there is a two state equilibrium between the BI and BII conformations.[4] The potential energy diagram of this equilibrium is shown in Figure 19.[4]

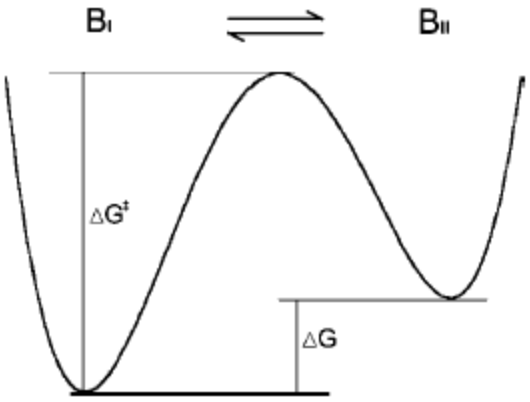

Figure 19: Energy diagram of transition between BI and BII conformations [4]

This method also assumes that the pure BI and pure BII conformations each have a single known ^{31}P chemical shift, independent of the sequence it is in.[4] This assumption leads to the following equation

$$\%\frac{BII}{100} = \frac{1}{\omega_{BII} - \omega_{BI}}\langle\omega\rangle - \frac{\omega_{BI}}{\omega_{BII} - \omega_{BI}}$$

where %BII/100 is the percent BII character divided by 100, ω_{BII} is the chemical shift of the pure BII conformation, ω_{BI} is the chemical shift of the pure BI conformation, and <ω> is the chemical shift of the phosphorous in question.[4] From there, it is assumed from experimental data that the chemical shift of the pure BI and BII conformations are as follows

$$\omega_{BI} = 0.002031T - 1.345$$

$$\omega_{BII} = -0.003079T + 1.023$$

where T is the temperature in Kelvin.[4] From there, the temperature dependent chemical shifts of the pure BI and BII conformations are plugged into the equation to find %BII, giving the following equation

$$\%BII = \frac{100}{2.368 - 0.005110T}\delta P + \frac{134.5 - 0.2031T}{2.368 - 0.005110T}$$

where δP is the chemical shift in ppm of the phosphorous in question.[4]

3.4 Two Dimensional NMR

While 1D NMR is useful for analyzing smaller molecules, or for observing specific nuclei in a compound, these spectra are often insufficient for larger biomolecules such as proteins and DNA molecules.[28] These molecules are too large to produce a readable spectrum, as the peaks will be crowded and difficult to determine where one ends and another begins.[25,28] For these molecules, 2D NMR techniques are more appropriate, as these techniques will have much better resolution as peaks can be spread out across two dimensions instead of just one.[25,28]

Because the spectrum is now in two dimensions, there must be two frequency scales; on the horizontal axis, the frequency is directly measured by Fourier transformation of the FID, similar to 1D NMR, while the vertical axis is an indirect frequency measurement, made by taking hundreds of 1D spectra and increasing the time delay in the pulse sequence.[25] A Fourier transform is performed on the incremented time delay, which in turn becomes vertical axis of the 2D spectrum.[25]

Every two dimensional NMR experiment generally has a four step pulse sequence: preparation, evolution, mixing, and detection.[25] To explain, we will only consider two nuclei, A and B.[25] In the preparation phase, where nucleus A is excited[25]. In the evolution phase, the chemical shift of A is indirectly measured.[25] This period may also be used to alter the magnetic environment of the nuclei if one desires a particular interaction to be enhanced or suppressed.[25] After A is indirectly measured, the mixing phase occurs, where magnetization is allowed to transfer from A to B.[25] This transfer can occur either through J-coupling or the NOE effect, depending on what is being measured.[25] Finally, in the detection phase, the chemical shift of B is

measured.[25] It should be said that in each step of the pulse sequence, all nuclei are experiencing it at the same time.[25] To tie it back to the example above, both A and B would become excited in the preparation phase, for example.[25]

In terms of interpreting a 2D *vs* a 1D spectrum, the biggest difference is that the peaks are no longer tied directly to one particular nuclei or group of equivalent nuclei; rather, in 2D NMR spectra each crosspeak represents an interaction between two or more nuclei, either through bond or through space.[25,28]

3.4.1 Correlation Spectroscopy (COSY) The first 2D NMR experiment designed was correlation spectroscopy, or COSY.[25] COSY measures through-bond interactions between J-coupled protons, typically with a range of three, four or, very rarely, five bonds.[25] The pulse sequence utilized to produce a COSY spectrum is 90°-t_1-90°-FID (Figure 20).[25]

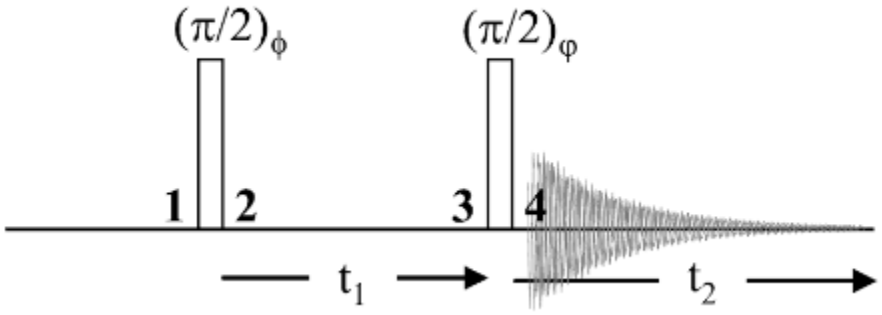

Figure 20: Pulse sequence of COSY experiment[34]

For an example, we will consider two J-coupled protons, H_a and H_b. The first 90° RF pulse will rotate the magnetization of H_a from the z-axis into the x-y plane.[25] This causes H_a's magnetization to precess in the x-y plane during t_1.[25] H_b will experience the same change in magnetization due to the RF pulse.[25,28.33] Additionally, J-coupling evolution occurs during t_1,

causing magnetization that is antiphase to the J-coupling to be produced.[25] Once the t_1 phase has been completed, a second 90° RF pulse is applied, which transfers the antiphase magnetization between the two protons.[25] During the detection phase, both protons will then precess normally, giving rise to the COSY spectrum.[25] A useful metaphor for visualizing the transfer of magnetization *via* J-coupling is to think of nuclei as bells and of J-coupling as a spring connecting the two bells together; when one bell is struck, the vibrations will travel through the spring, thus vibrating the other bell.[28] It should also be stated that COSY is a homonuclear experiment.[25] This is due to both RF pulses being tuned to protons.[25]

From a practical standpoint, COSY spectra show which protons are close to each other through bonds.[25,28] Consider the following hypothetical COSY spectrum of a halogenated, conjugated hexene (Figure 21).

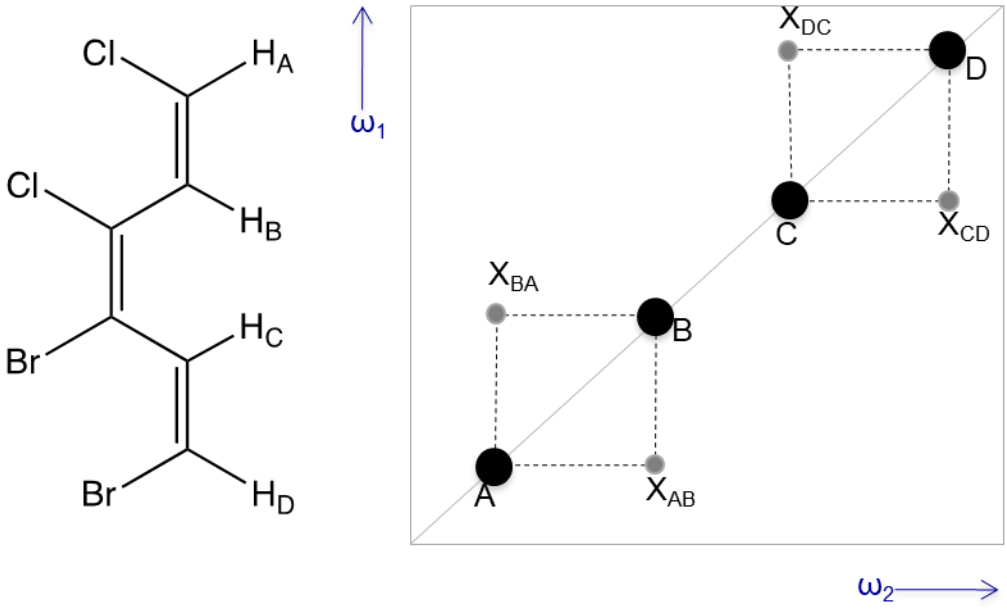

Figure 21: Hypothetical COSY spectrum of halogenated hexene[35]

The peaks along the diagonal line of the spectrum can be thought of as just the 1D NMR spectrum depicted from an aerial angle. From there, four crosspeaks can be seen in the spectrum. Two of these represent interactions between protons A and B, while the other two represent interactions between protons C and D. When actually reading a COSY spectrum, usually it is sufficient to interpret only one of the two halves of the diagonal, as the information is equivalent on either side; the only difference is the directionality of which proton transferred magnetization and which proton received magnetization.[25,33] It should be stated that even though protons B and C are located near each other in space, they do not produce a crosspeak in the COSY spectrum because their through bond distance apart is five bonds, which is outside the normal range of a COSY experiment.[25]

3.4.2 Total Correlation Spectroscopy (TOCSY) Conceptually, TOCSY experiments are similar to COSY experiments, as both have crosspeaks representing through-bond interactions, and both utilize J-coupling in order to generate a spectrum.[25] However, unlike COSY which can only produce a single "jump" when transferring magnetization, TOCSY experiments can cause magnetization to be transferred through multiple jumps along a chain of J-coupled protons, with only the transfer between the initial nucleus and the final nucleus of the jumps being recorded.[25] This long range transfer allows a TOCSY spectrum to show crosspeaks between protons up to five bonds away from each other, a significant increase in the interaction range from a COSY.[25]

For pulse sequence, a TOCSY experiment begins the same as a COSY, with an initial 90° RF pulse, followed by the t_1 evolution period.[25] After the t_1 evolution time, there is a spin-lock sequence in place of the second 90° RF pulse found in the COSY experiment.[25] The spin-lock sequence, instead of just being a single RF pulse, is a period of high power RF pulses repeated

close together in time, which favors transfer of magnetization through J-coupling within an entire spin system (Figure 22).[25]

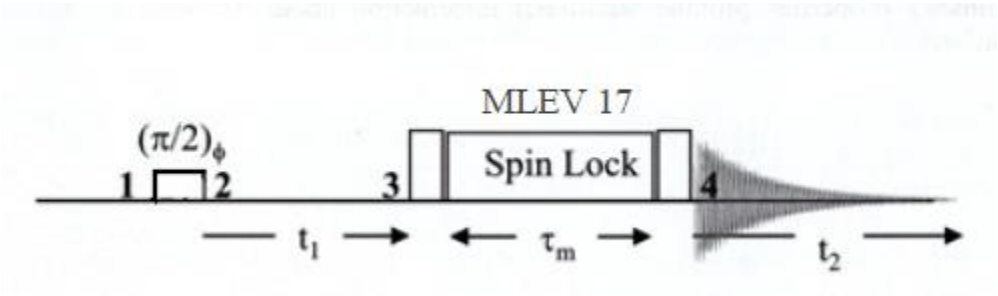

Figure 22: Pulse Sequence of TOCSY experiment[34]

Because the transfer of magnetization can be shared through a whole spin system *vs* between atoms three or fewer bonds away, a TOCSY spectrum will often have more crosspeaks than a COSY spectrum of the same molecule.[25,28] For example, take the TOCSY and COSY spectra of 3-heptanone (Figure 23).

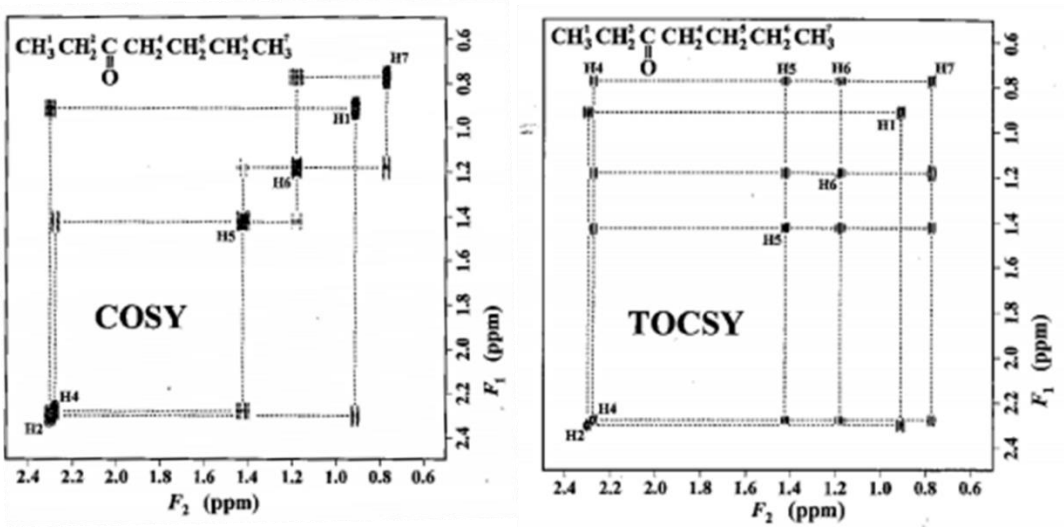

Figure 23: Comparison of hypothetical COSY and TOCSY of 3-heptanone[25]

The additional crosspeaks are visible in the spectra above. It should be noted that despite the increased range of the J-couping, the protons on carbons 1 and 2 still only show crosspeaks with each other. This is due to the ketone separating the carbon chain into two distinct spin systems, as the transfer of J-coupling is blocked by the ketone.[25]

3.4.3 Nuclear Overhauser Effect Spectroscopy (NOESY) Where COSY and TOCSY experiments utilize J-coupling to generate a spectrum depicting through bond interactions, NOESY experiments utilize dipolar coupling to generate a spectrum depicting through space interactions.[25] A useful metaphor for visualizing the transfer of magnetization via the NOE is to think of nuclei as bells.[28] If two bells are struck, each will produce their own sound, but when the bells are brought physically close together, distortion can be heard in the ringing.[28]

Magnetization in a NOESY experiment is transferred using the Nuclear Overhauser Effect.[25,28] In essence, the NOE is when the spin state of one population of nuclei changes the spin states of another population of nuclei via dipolar coupling.[36] To explain, consider two nuclei, each of which is the same nuclear species with spin 1/2, which we will call I and S, where I is the nuclei of interest and S is a nucleus that I experiences dipolar coupling with.[36] We will also assume that S and I have different chemical shifts but are not J-coupled.[36] With that, the energy diagram for the nuclei is shown below (Figure 24). Unlike an isolated spin-1/2 nucleus, there are four possible energy states that the system can be: αα, αβ, βα, or ββ, referring to A and B, respectively.[36] Initially, the system will exist in the Boltzmann distribution between the α and β spin states.[36] However, if S is excited by an outside RF pulse, this equilibrium will be disturbed, causing more pairs of nuclei to enter into the higher energy states until the system is saturated, meaning that there are an equal number of nuclei of each type in the α and β states.[36]

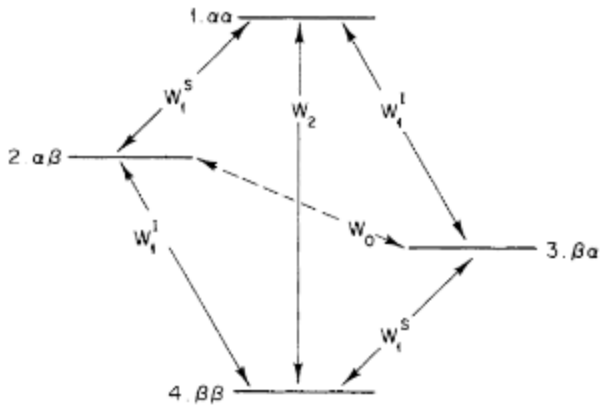

Figure 24: Energy spin diagram for two hypothetical spin-1/2 nuclei [36]

In order to return to equilibrium, there are multiple pathways the system can take.[36] First, there are the W_1^I and the W_1^S transitions, which represent the relaxation of either I or S, respectively, from the β state to the α state without any change to the other nucleus.[36] Next, there is the W_0 transition, in which S relaxes from β to α, causing I to be excited from α to β, causing an overall change of αβ to βα (or *vice versa*).[36] This type of relaxation is known as spin flipping, leading to a decrease in the population difference of I, and causes the negative NOE, this spin flipping weakens the signal of I.[36] Positive NOE comes from the W_2 transition, where S relaxing from the β state to the α state in turn causes I to also relax from the β state to the α state.[36] This causes the population difference of I to increase, leading to a stronger signal[36]. The pulse sequence utilized by the 2D NOESY, unlike other 2D NMR experiments, is actually composed of three pulses (Figure 25).[25]

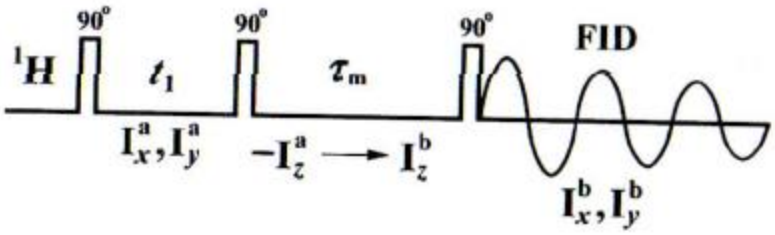

Figure 25: Pulse sequence of 2D NOESY experiment[25]

The first RF pulse is the same as in the COSY experiment, converting z-magnetization into magnetization in the x-y plane.[25] After the t_1 phase, the mixing phase begins with another RF pulse, creating a population difference by converting x'-y' magnetization into z magnetization.[25] By converting magnetization into z magnetization, it allows there to be cross relaxation between nearby nuclei during the mixing delay, τ_m.[25] After the transfer of magnetization, the final RF pulse converts the z magnetization into x'-y' magnetization, and induces the FID.[25]

From a practical standpoint, NOESY spectra can be thought of as a molecular yardstick, a tool to show how close two protons are to each other in space.[28] Consider the following hypothetical NOESY spectrum of a halogenated, conjugated hexene (Figure 26). Once again, the peaks along the center can be thought of as just the 1D proton NMR spectrum. Six crosspeaks can be seen in the spectrum. Two of these crosspeaks represent interactions between protons A and B, two between protons C and B, and two between protons C and D. The reason crosspeaks between B and C can be observed in the NOESY is that while the two protons have multiple bonds between them, the NOE is a through space interaction, so the two protons being close enough in space is enough to produce a crosspeak in the NOESY. Typically, all of the crosspeaks present in the COSY will be present in the NOESY, which makes the COSY a useful tool to help in order to interpret a NOESY spectrum.

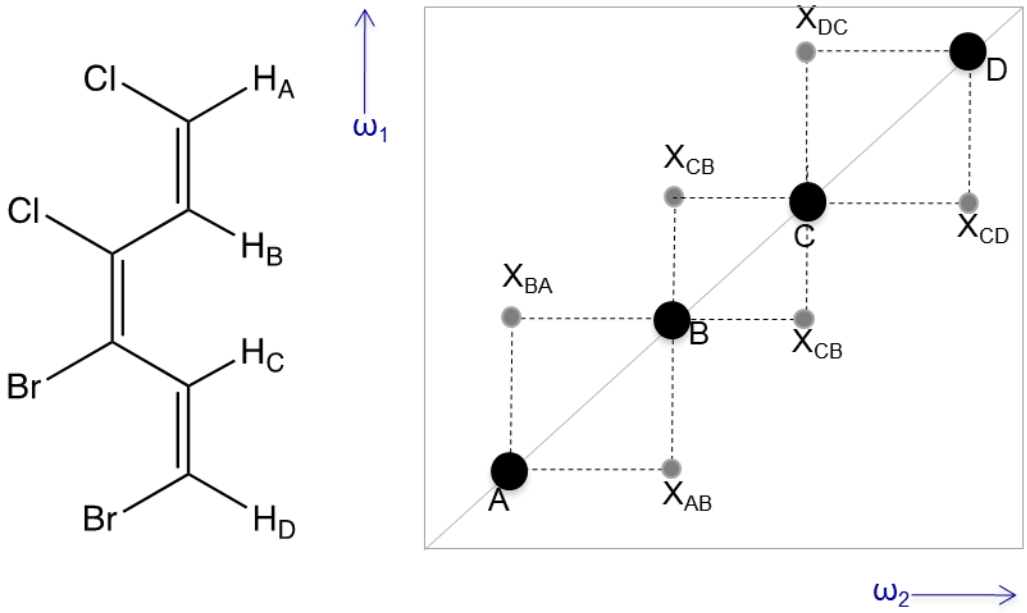

Figure 26: Hypothetical NOESY spectrum of halogenated hexene[35]

3.4.4 Heteronuclear Single Quantum Coherence (HSQC) Among the NMR experiments outlined in this thesis, HSQC is unique in that it is the only 2D heteronuclear experiment performed. Typically, it manipulates interactions between protons and nitrogen-15 that are observed using this experiment, as this is a popular technique for studying peptide bonds within proteins.[25,28] In this thesis, the nuclei observed will be protons and phosphorous-31, as we are using it to study the phosphodiester backbone of DNA. Similar to the COSY and TOCSY experiments, magnetization will be transferred from ^{1}H to ^{31}P via J-coupling.[25]

The basic pulse sequence for an HSQC experiment is illustrated in the figure below (Figure 27). It begins with a 90° RF pulse on the proton channel, followed by a delay to cause an INEPT transfer.[25] The initial pulse generates a ^{1}H doublet, which is then converted into an *anti*phase doublet by the INEPT transfer.[25] Next, an RF pulse is applied to both the proton and ^{31}P channels simultaneously.[25] After t_1 delay period, mixing is accomplished by applying another 90° pulse on

both channels, causing a back transfer, encoding the phosphorous chemical shift in the proton signal.[25]

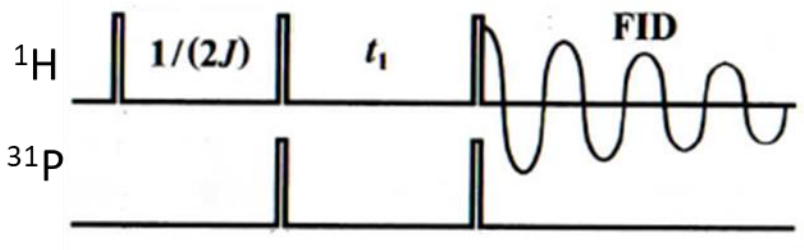

Figure 27: Pulse sequence for HSQC[25].

The spectrum produced by an HSQC will lack a diagonal line and the symmetry found in the other experiments discussed in this thesis.[25,28] This is due to it being a heteronuclear experiment.[25,28]

CHAPTER 4: MATERIALS AND METHODS

4.1 DNA Preparation

The DNA sequences used in this research were purchased from Integrated DNA Technologies, Coralville, Iowa. The sequences used were the Drew-Dickerson Dodecamer ds-5'-(C_1-G_2-C_3-G_4-A_5-A_6-T_7-T_8-C_9-G_{10}-C_{11}-G_{12})-3' and the lesioned sequence, a Dickerson sequence where the C_3 residue is replaced with a dihydrouracil nucleotide ds-5'-(C_1-G_2-dhU_3-G_4-A_5-A_6-T_7-T_8-C_9-G_{10}-C_{11}-G_{12})-3', and will be referred to as UdhU3D throughout the rest of this thesis (see Figure 28 for structures). The Drew Dickerson Dodecamer was selected because not only is its NOESY spectrum solved, it is historically important as it is the first DNA strand to be crystallized in the B-form, and as such its physical properties are very well studied.[7,37]

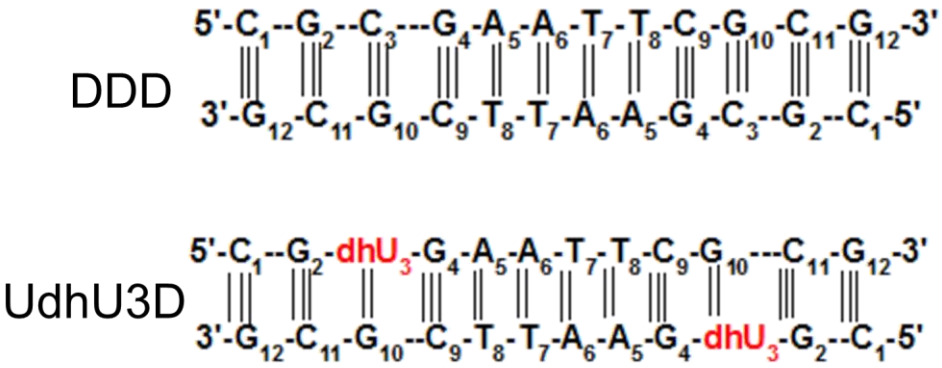

Figure 28: Double stranded DNA sequences, lesion highlighted in red

The DNA sequences were purchased as single stranded DNA. They were then brought up in 25 mM pH 7.39 phosphate buffer so that the double stranded DNA concentration would have a final concentration of approximately 0.5 mM. After being dissolved in buffer, the single

stranded DNA was then annealed by placing the DNA solution into a 90 °C water bath for fifteen minutes, and then was allowed to cool to room temperature. Once cooled, the now double-stranded DNA was lyophilized and redisolved in D_2O at least three times to remove excess H_2O, adjusting the pH in between lyophilizations with 0.1 M Phosphoric acid or 0.1 M sodium hydroxide, depending on the need. This drying and redissolving process continued until the pH of the solution was stable at approximately 7.4 between dryings. If an imino experiment was needed, the DNA was lyophilized to remove it from the D_2O solvent and then redisolved in 0.9 mL of deionized water and 0.1 mL of D_2O.

4.2 NMR Studies

All NMR experiments were performed using a Varian INOVA 400 MHz NMR located at Missouri State University.

4.2.1 NOESY All NOESY spectra were performed on DNA prepared in the conditions outlined above. The nuclei observed were protons. 128 transient scans were used with a one second recycle time at 25 °C and two second recycle time at 10 °C. 1024 complex points were used.

4.2.2 1D Imino All 1D imino experiments were performed on DNA prepared in the conditions outlined above in the 90% H_2O solvent. The pulse sequence used was wet1D to account for water suppression. 1024 transient scans were used with a two second recycle time. The spectral window was set to 8000 Hz and 16384 complex points were used. The receiver gain was set to 30 and the offset was set to 1842.1121 Hz. Spectra were taken from 5 to 25 °C in five degree intervals to acquire a melting profile of the lesioned DNA.

4.2.3 HSQC HSQC was performed on DNA prepared in the conditions outlined above in D_2O solvent. The spectral width was set to -20 to 20 ppm. 256 t_1 increments were utilized with 64 scans per t_1 increment. The one bond H-P coupling constant was set to 15. In the f2 axis, the spectral window was set was to 4001.6 Hz and 2048 complex points were utilized. The receiver gain was set to 30 and the recycle time was set to four seconds. In the f1 axis, the spectral window was set to 6460 Hz. The one bond coupling was set to 15 Hz. In the wet tab, the recycle time was set to two seconds, the proton offset was set to -150 Hz, and the power was set to -13. In the Parameters tab, the proton offset was set to -150 Hz, the observed power was set to 9.5 and the observe power was set to 55. For the phosphorous channel, the x pulse width was set to 10.8 Hz, the x pulse power was set to 56, and the phosphorous offset was set to 3988 Hz. The decoupling power was set to 0, the modulation mode used was ccg, the modulation frequency was set to 200 Hz, and the decoupling shape was set to garp1. The sample spin was turned off, and the spectrum was collected at 25 °C.

4.2.4 1D Phosphorous All 1D phosphorous spectra were performed on DNA prepared in the conditions outlined above. 1024 transient scans were utilized with a two second recycle time, a spectral window of 10,000 Hz, 16384 complex points, and a spectrum offset of 138.6329 Hz. Temperature ranged from 5-35 °C, taken in 5 °C increments in order to create a temperature study of the phosphate backbone. One experiment took approximately one hour and fifteen minutes, and a full temperature study of a particular DNA sequence took approximately 9 to 10 hours to perform, taking into account the time to change and allow the sample to adjust to the new temperature and to set up each new experiment.

4.3 Sequential Walking Technique on Dickerson Sample

Data from the NMR experiments was processed and analyzed using ACD Labs NMR software. All NOESY experiments were referenced to their water peak, with the water peak being referenced to 4.77 ppm at 25 °C and 4.94 at 10 °C.[38] From there, the NOESY was analyzed in four specific regions as outlined by Hare et al: the base-2'/2" region, the base-1' region, the 1'-2'/2" region, and the 3'/4'-2'/2" region[3] (Figure 29).

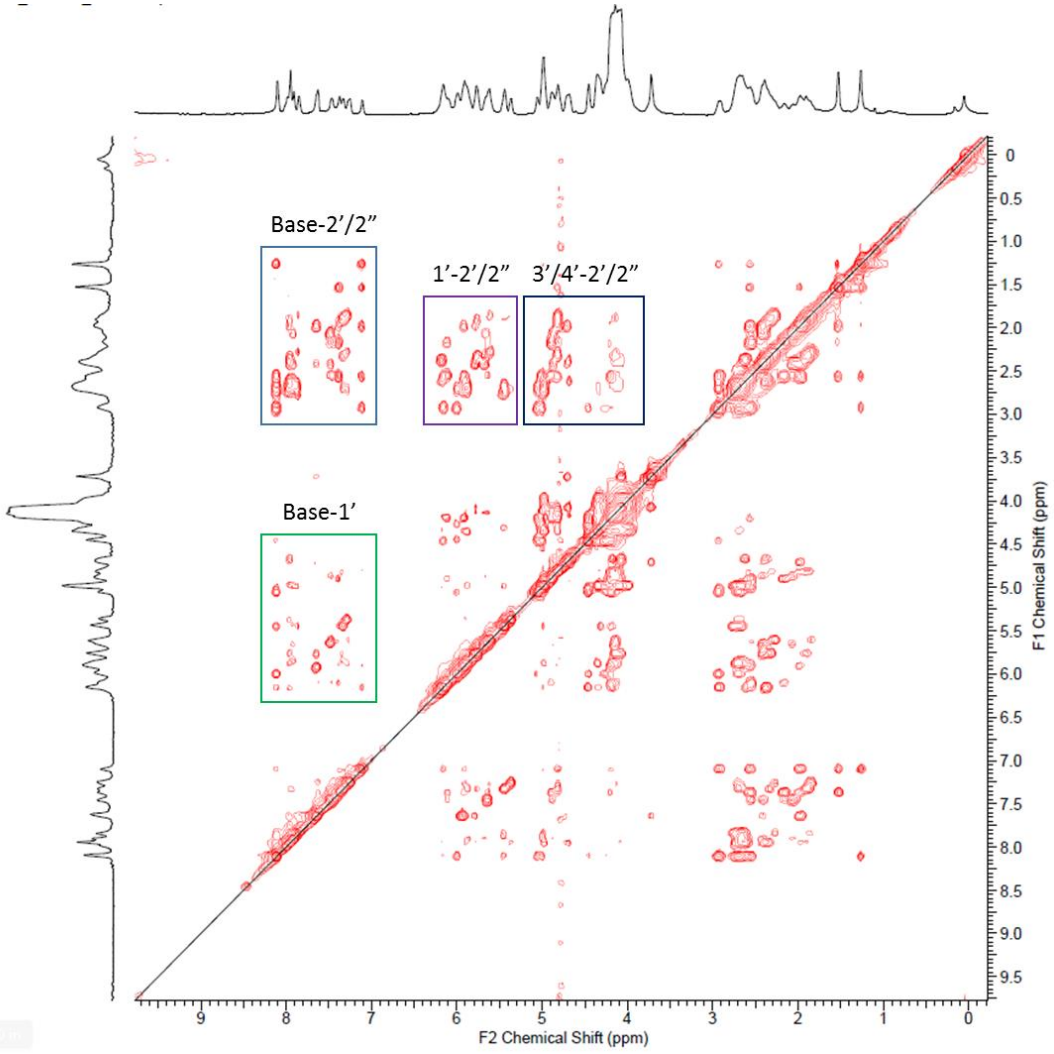

Figure 29: NOESY of DDD at 25°C with labelled spectrum regions

Each of these regions depict a different crosspeak between two or more different types of proton, on either the DNA base or the ribose sugar.[7] Consider the aromatic-s'/2" region of the spectrum (Figure 30).

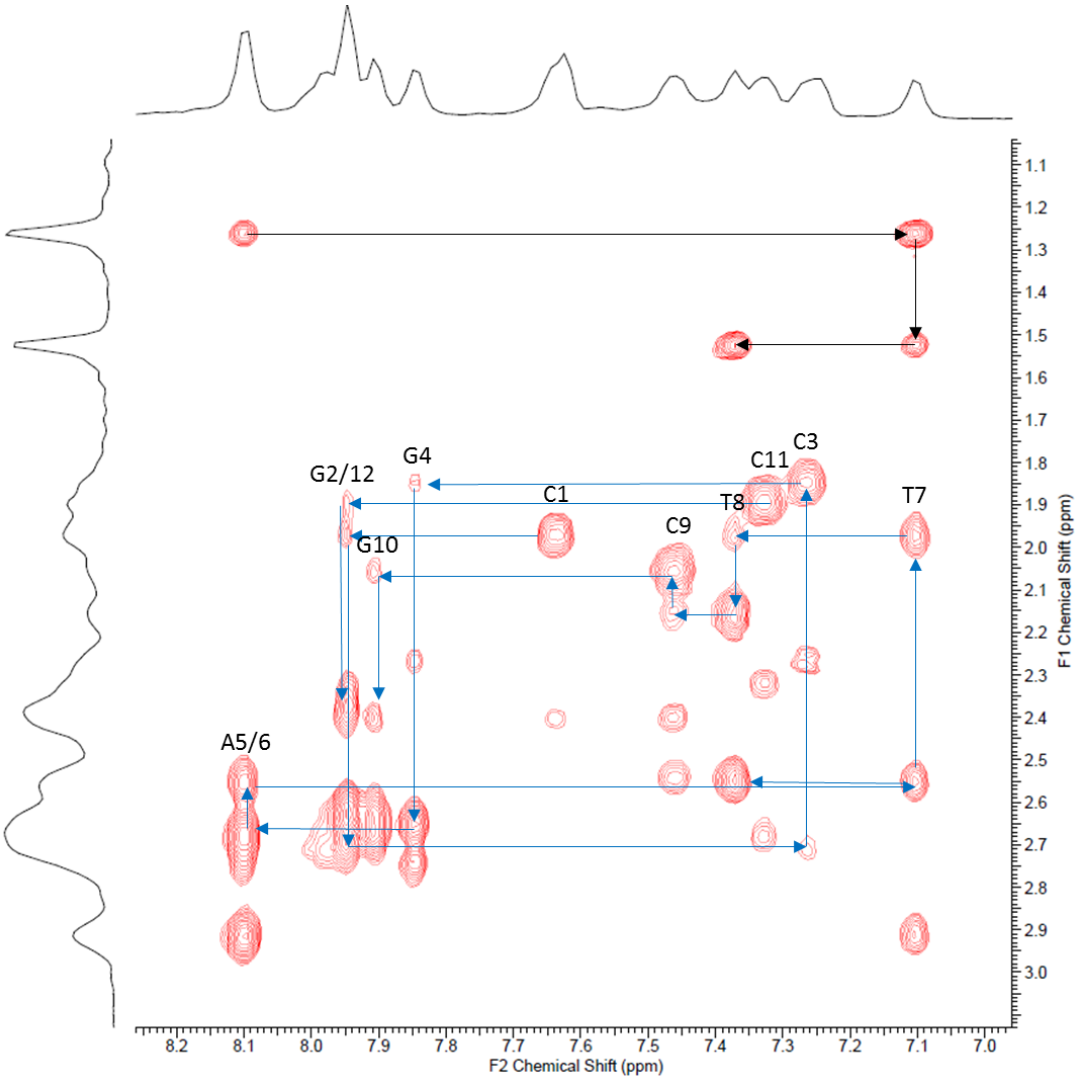

Figure 30: The aromatic-2'/2" region of the DDD NOESY spectrum. Blue arrows represent the 2' walk, black arrows represent the methyl walk. The G10 to C11 walk is missing, however it is still interpretable as the 2" walk is clear.

The aromatic-2'/2" region lies roughly between 7 and 8.3 ppm on the horizontal axis and 1 and 3 ppm on the vertical axis. It depicts the crosspeaks between the 2'and 2" protons on the ribose sugar of DNA to the 6 proton on pyrimidines and the 8 proton on purines. Theoretically, a walk can be performed using either the 2' or 2" protons with the aromatic protons. However, the 2' crosspeaks are typically weaker than the 2" crosspeaks, and as such the 2" crosspeak walks are typically used to assist with the 2' walk (see Figure 31).[7]

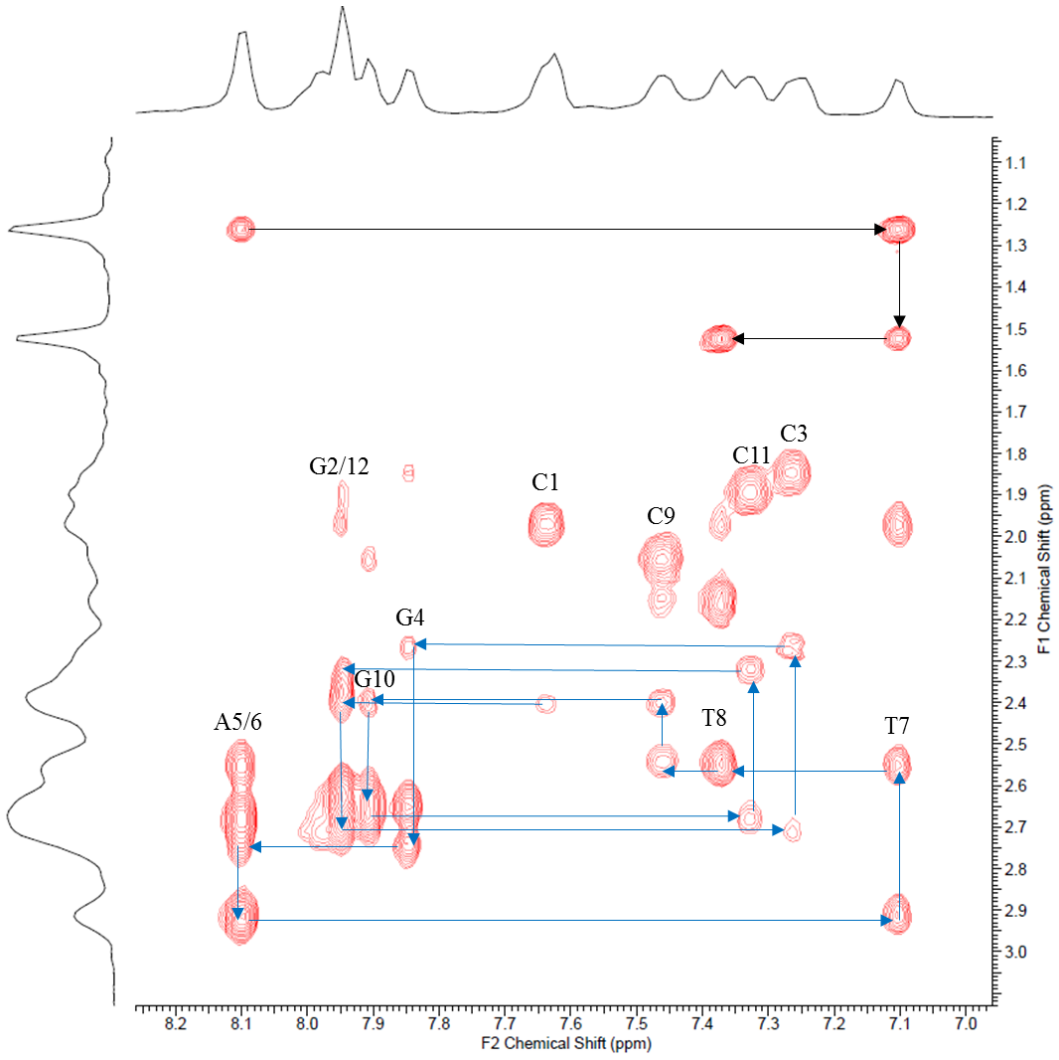

Figure 31: The aromatic-2'/2" region of the DDD NOESY spectrum. Blue arrows represent the 2" walk, black arrows represent the methyl walk. Unlike the 2' walk, the 2" walk can be made in one long, unbroken walk.

Both of these walks are shown in the two figures above. In addition to the 2’ and 2” crosspeaks, the crosspeaks of the thymidine methyl are also present in the aromatic-2’/2” region, meaning that we can also perform a walk involving the methyl crosspeaks. This walk allows us to easily identify the chemical shifts of the A6, T7, and T8 bases, giving a solid starting point to perform the 2’ and 2” walks. Next, we will consider the aromatic-1’ region of the spectrum (Figure 32).

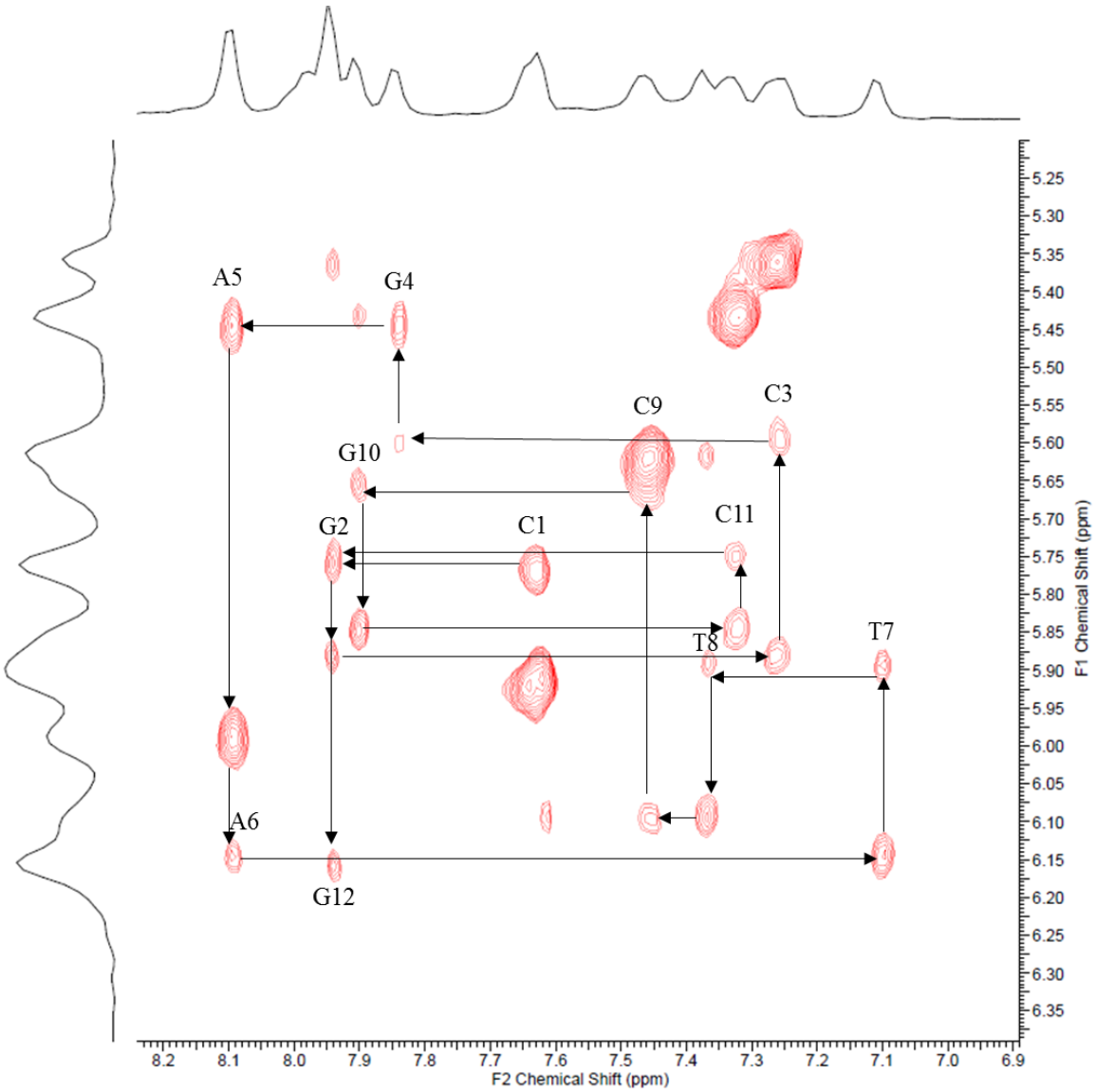

Figure 32: The aromatic-1’ region of the DDD NOESY spectrum. The arrows represent the 1’ walk

The aromatic-1’ region lies on the horizontal axis roughly between 7 and 8.3 ppm and on the vertical axis between 5 and 6.3 ppm. It should be noted that the aromatic-2’/2” and the aromatic-1’ regions have the same spread over the horizontal axis, so it is often helpful to solve these regions concurrently with each other. In addition, solving one region’s walk will give the ppm of the aromatic protons, which will be consistent throughout the spectrum. One unique feature of the aromatic-1’ region is that the crosspeaks between the cytosine/urail 5 and 6 base protons are typically located there. These crosspeaks are easily identifiable, as they are typically the strongest crosspeaks in the region, due to the protons having a close, fixed distance from each other. These crosspeaks can be found by raising the floor of the display so that only the four most intense peaks are visible. In the Drew Dickerson Dodecamer, these four peaks are the cytosine 5-6 aromatic proton crosspeaks. A walk utilizing the 1’ crosspeaks can also be performed in the aromatic-1’ region of the NOESY spectrum, pictured below (Figure 32).

Once the assignments are made in both the aromatic-2’/2” region and the aromatic-1’ region, analysis can begin of the 1’-2’/2” region. This region lies between 5 and 6.3ppm on the horizontal axis and 1 and 3ppm on the vertical axis. No new assignments are made in this region; rather, it is used as a check of the assignments made in both preceding regions. This region contains the crosspeaks that represent the interactions between the 2’ or 2” protons with the 1’ protons of the same nucleotide. If all of the peaks are assigned correctly, the ppm assignments of the 1’ protons in this region should match the ppm assignments made in the aromatic-1’ region. While a traditional walk cannot be made in this region, it is a useful region for checking one’s work, and as such can be referred to as the ‘check prime’ region (Figure 33).

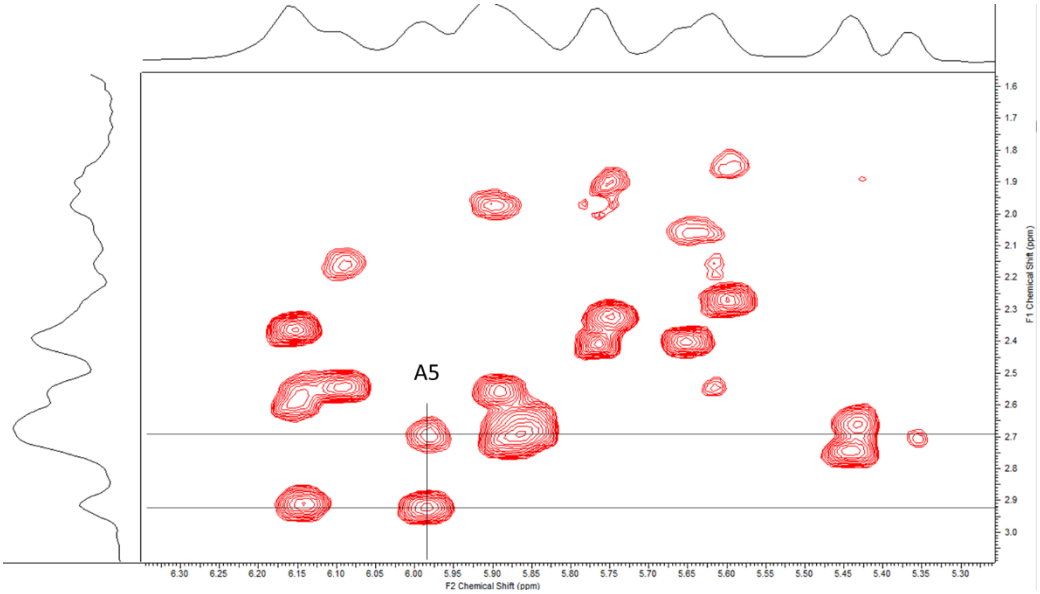

Figure 33: The 1'-2'/2" region of the DDD NOESY Spectrum. The crosspeaks of the 1' and the 2'/2" protons of the A5 residue are shown by the black lines

Once assignments in the check prime region are confirmed, the 3' and 4' assignments can be determined by analyzing the 3'/4'-2'/2" region. This region lies between 4.0 and 5.1 ppm on the horizontal axis and 1 and 3ppm on the vertical axis. In this region, two types of crosspeaks are present: there are the 2'/2" to 3' crosspeaks, and the 2'/2" to 4' crosspeaks. Like the check prime region, a traditional walk cannot be made, rather two lines are drawn over from the aromatic-2'/2" region, as both the 3' and 4' protons should have a crosspeak with both the 2' and 2" protons of the same nucleotide. Typically, the 2'/2" to 3' crosspeaks are much more intense than the 2'/2" to 4" crosspeaks, and often the weakness of the 4' crosspeaks makes assignments incredibly difficult to assign. In addition, the water peak resides in this region, which must be suppressed to make this region readable. However, as a consequence, any peaks that have the same chemical shift as the water peak end up being suppressed along with it (Figure 34).

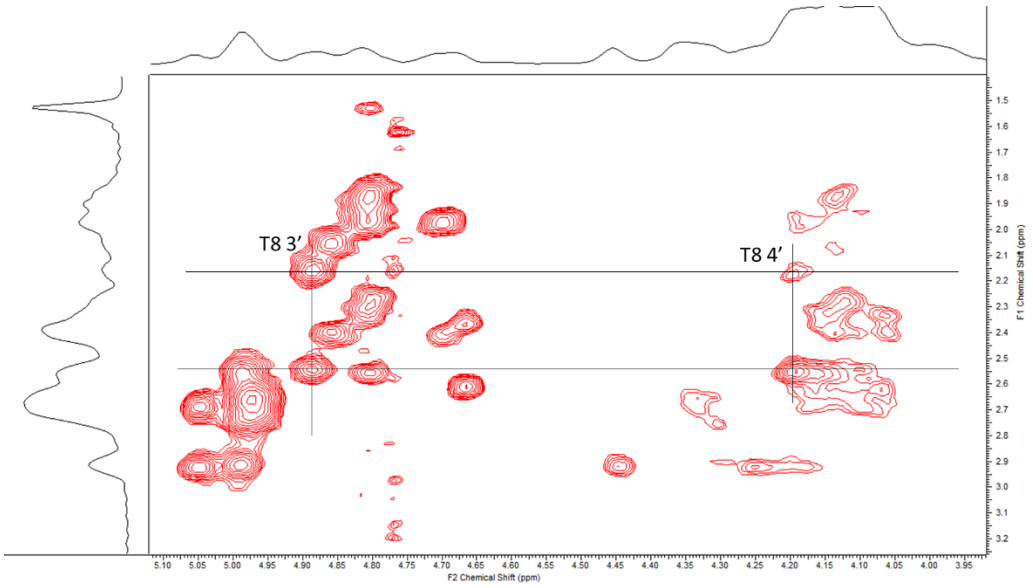

Figure 34: 3'/4'-2'/2" region of DDD NOESY Spectrum. 3' and 4' crosspeaks of T8 denoted with the intersecting black lines. Residual peaks from the suppressed water peak can be seen at 4.77ppm on the horizontal axis

Through this method, the base, 1', 2', 2", 3' and 4' proton chemical shifts can be found. This method works not only for our control sequence, but for any lesioned sequences as well. The table of Drew Dickerson Dodecamer chemical shifts is shown below (see Table 1), and the chemical shifts of the lesioned sequence will be compared against it.

Table 1: Chemical Shifts of the Drew Dickerson Dodecamer at 25°C

Position	Aromatic	1’	2’	2”	3’	4’	5/Me
C1	7.64	5.78	1.98	2.42	4.71	4.07	5.93
G2	7.95	5.89	2.71	2.71	4.95	4.34	
C3	7.27	5.60	1.85	2.27	4.80	4.12	5.37
G4	7.86	5.45	2.66	2.74	4.98	4.32	
A5	8.12	6.00	2.69	2.93	5.06	4.45	
A6	8.12	6.15	2.56	2.93	4.99	4.45	
T7	7.12	5.90	1.98	2.56	4.81	4.20	1.27
T8	7.39	6.11	2.18	5.57	4.89	4.20	1.53
C9	7.48	5.65	2.07	2.41	4.87	4.14	5.63
G10	7.91	5.85	2.68	2.68	4.98	4.34	
G11	7.33	5.75	1.90	2.32	4.81	4.15	5.43
C12	7.95	6.16	2.36	2.60	4.67	4.18	

CHAPTER 5: RESULTS

The following results are for the Dickerson and UdhU3D DNA sequences. All samples were prepared with an approximate concentration of 0.5 mM.

5.1 NOESY

The NOESY spectrum of UdhU3D was taken using the parameters outlined above. Then the 2', 2" and methyl walks were performed as they were above (Figures 35, 36).

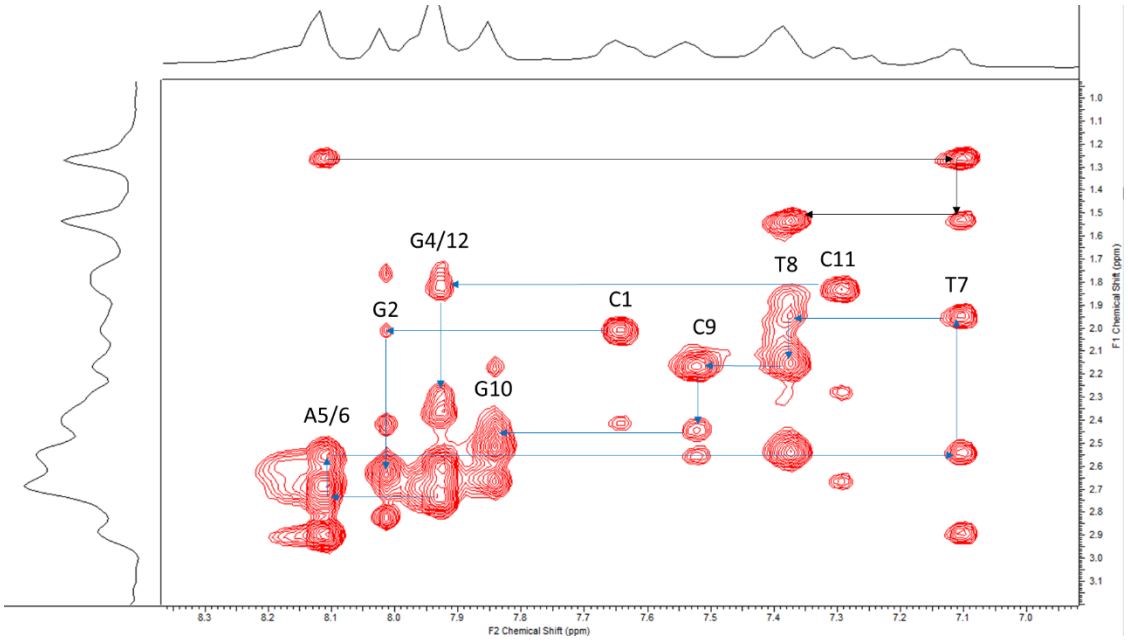

Figure 35: Aromatic-2'/2" region of UdhU3D NOESY spectrum. The methyl walk is outlined in black, and the 2' walk is outlined in blue

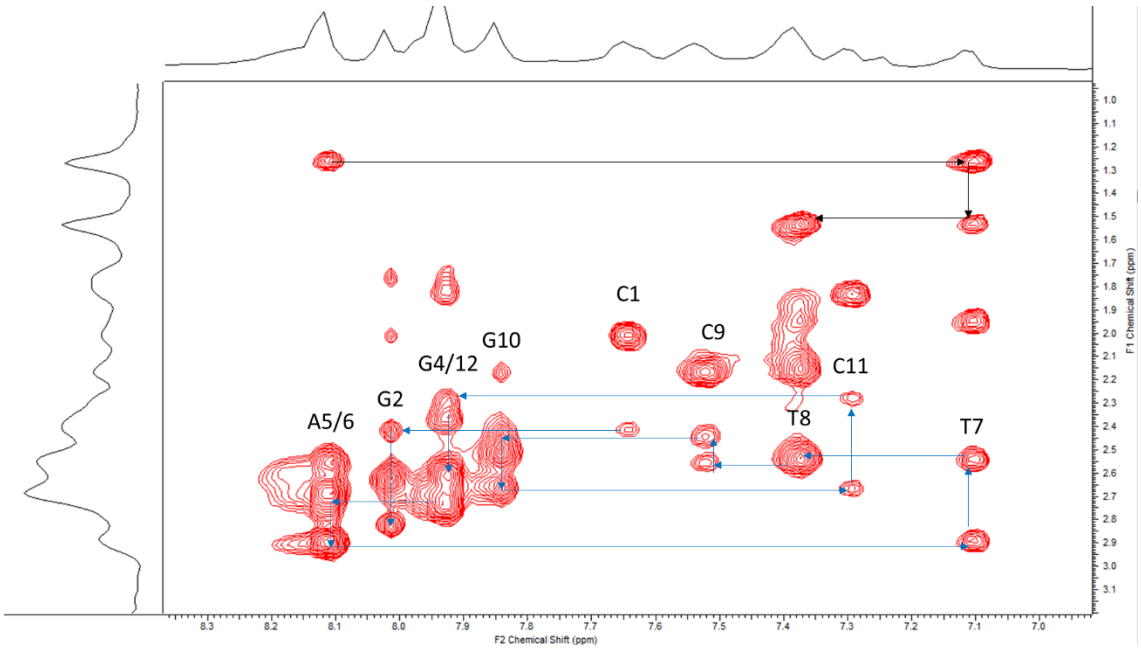

Figure 36: Aromatic-2'/2" region of UdhU3D NOESY spectrum. The methyl walk is outlined in black, and the 2" walk is outlined in blue

One may notice that the Udh3 residue is absent from both the 2' and 2" walk in the above figures. This is because the base protons are not aromatic, unlike the base protons for the other residues. This causes the base protons to have a chemical shift between 1 and 3 ppm, meaning any potential 2' or 2" walk must go across multiple different regions of the spectrum (Figure 37).

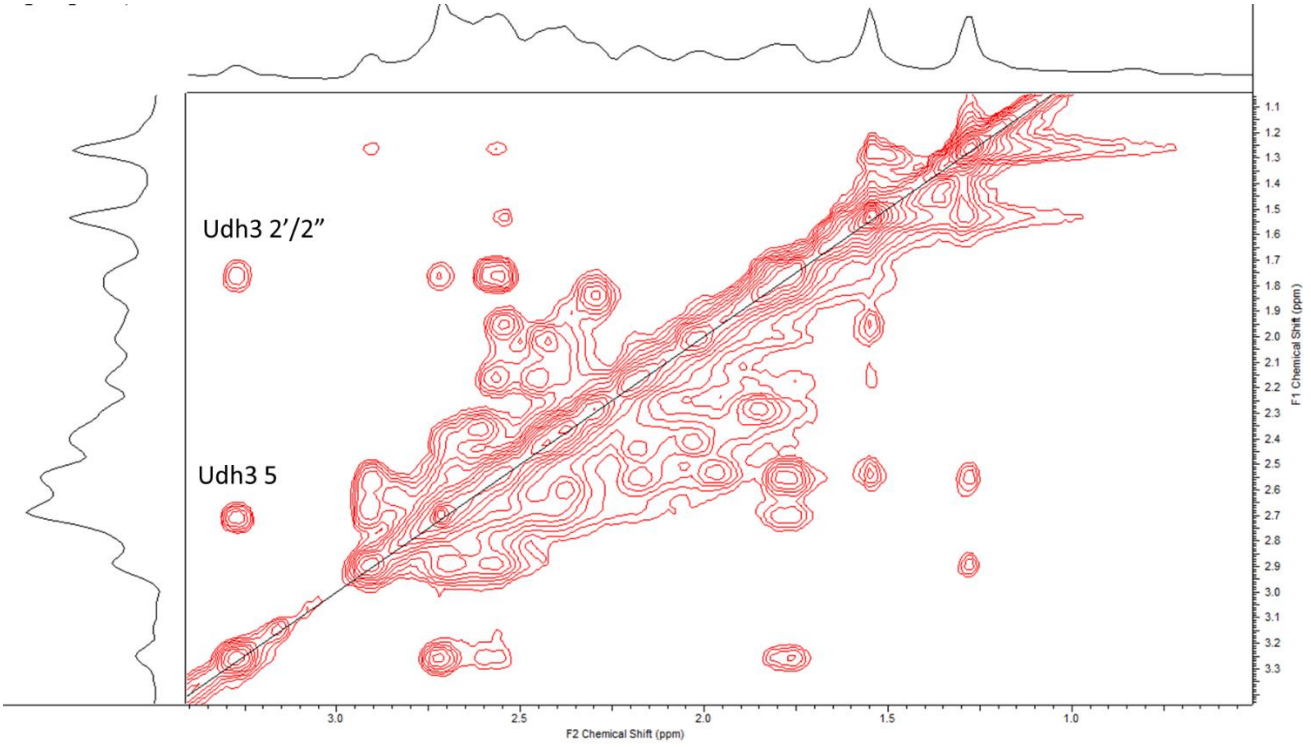

Figure 37: 2'/2"-2'/2" region of UdhU3D NOESY spectrum. The 6-2' and 6-2" crosspeaks are overlapped. In addition, the 6-5 crosspeak of Udh is visible in this region.

It should also be noted that the 2' walk is broken when going from the G10 2'/Base crosspeak to the G10 2'/ C11 Base crosspeak. 2' interactions are often weaker than the 2" interactions, and this break in the walk is also seen in the control sequence, so it is to be expected.

Once the 2' and 2" walks were completed, the 1' walk was performed (Figure 38).

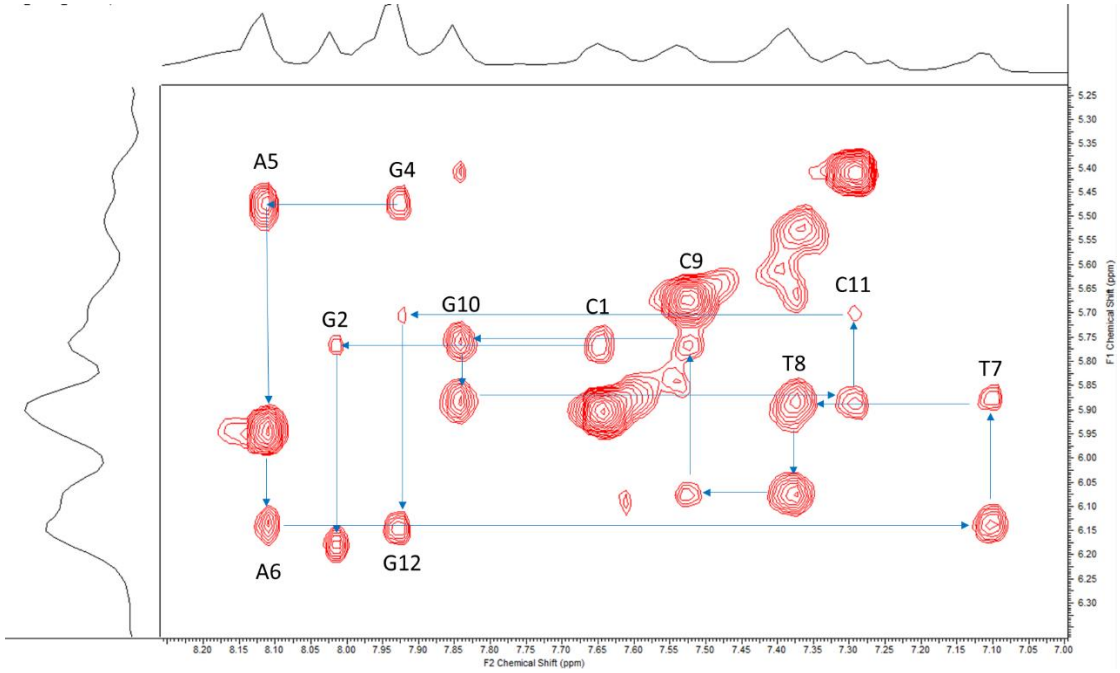

Figure 38: Aromatic-1’ region of UdhU3D NOESY Spectrum with the 1’ walk outlined in blue

Like in the 2’ and 2” walks, the Udh3 residue is absent from the 1’ walk in this region. The 1’ proton chemical shifts for the Udh3 residue were identified and confirmed using the 2’/2”-1’ region. While the walk can be completed, the crosspeaks linking the C11 and G12 residues are very weak.

Once all walks were completed and the crosspeaks for every proton of interest were identified, the chemical shifts of each proton were recorded (Tables 2, 3).

Table 2: Chemical Shift of UdhU3D at 25 °C

Position	Base	1’	2’	2”	3’	4’	5/Me
C1	7.64	5.77	2.01	2.41	4.68	4.06	5.90
G2	8.02	6.18	2.63	2.82	4.93	4.37	
Udh3	3.26	5.18	1.76	1.76	4.61	3.88	2.71
G4	7.94	5.47	2.73	2.73	4.97	4.31	
A5	8.12	5.95	2.90	2.90	5.04	4.16	
A6	8.12	6.14	2.90	2.90	4.98	4.43	
T7	7.10	5.87	2.54	2.54	4.78	4.19	1.27
T8	7.38	6.08	2.53	2.53	4.88	4.17	1.54
C9	7.52	5.76	2.44	2.44	4.86	4.14	5.67
G10	7.85	5.88	2.66	2.66	4.95	4.33	
C11	7.29	5.70	2.28	2.28	4.78	4.09	5.41
G12	7.92	6.14	2.61	2.61	4.66	4.14	

Table 3: Chemical Shift of UdhU3D at 10 °C

Position	Base	1’	2’	2”	3’	4’	5/Me
C1	7.65	5.74	2.05	2.42	4.68	4.04	5.89
G2	8.03	6.18	2.64	2.81	4.97	4.03	
Udh3	3.27	5.11	1.75	1.75	4.61	3.28	2.67
G4	7.94	5.46	2.73	2.73	4.98	4.32	
A5	8.14	5.95	2.69	2.90	5.04	4.45	
A6	8.14	6.15	2.58	2.90	5.00	4.16	
T7	7.04	5.89	1.96	2.56	4.82	4.20	1.25
T8	7.36	6.08	1.97	2.56	4.80	4.18	1.52
C9	7.54	5.75	2.18	2.45	4.87	4.14	5.65
G10	7.87	5.87	2.50	2.67	4.98	4.13	
C11	7.30	5.66	1.83	2.28	4.79	4.09	5.39
G12	7.94	6.14	2.33	2.62	4.65	4.17	

5.2 Phosphorous Studies

The 1D Phosphorous spectra of both the control sequence and lesioned sequence were solved from the HSQC and then analyzed for their % BII character (see Figure 39).

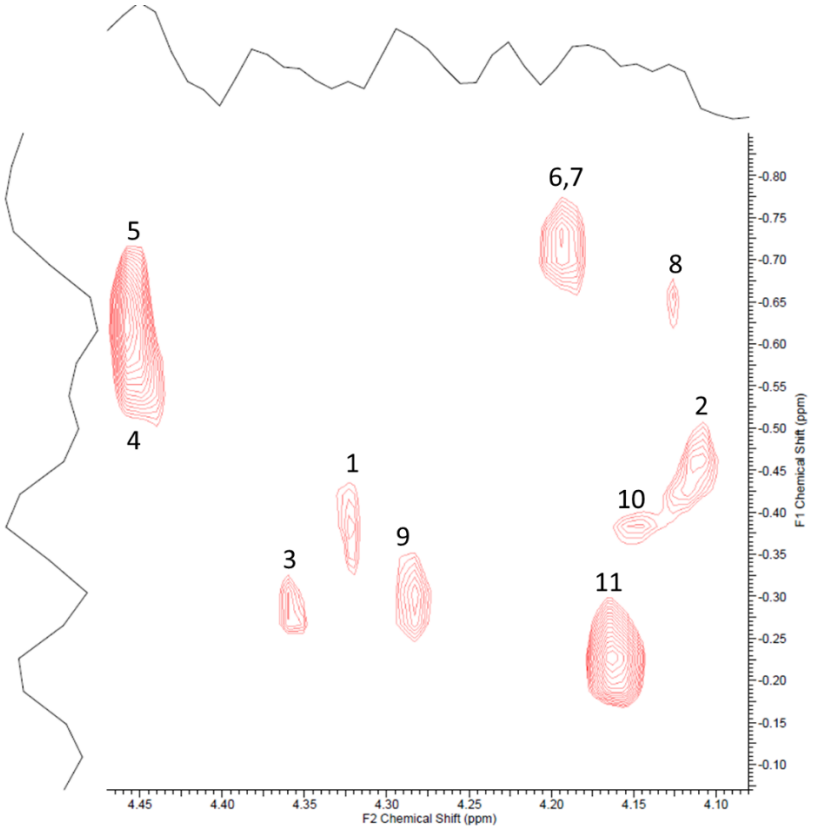

Figure 39: HSQC of the Drew-Dickerson Dodecamer at 10 ˚C

The order of which the Drew-Dickerson Dodecamer peaks appear in the 1D phosphorous spectra is known at 19 ˚C, from there assignments were extrapolated to each of the temperatures studied[39].

The chemical shifts of the phosphate peaks were referenced to phosphoric acid using phosphoric acid, with phosphoric acid being set to 0 ppm, using a coaxil insert (Figure 40).

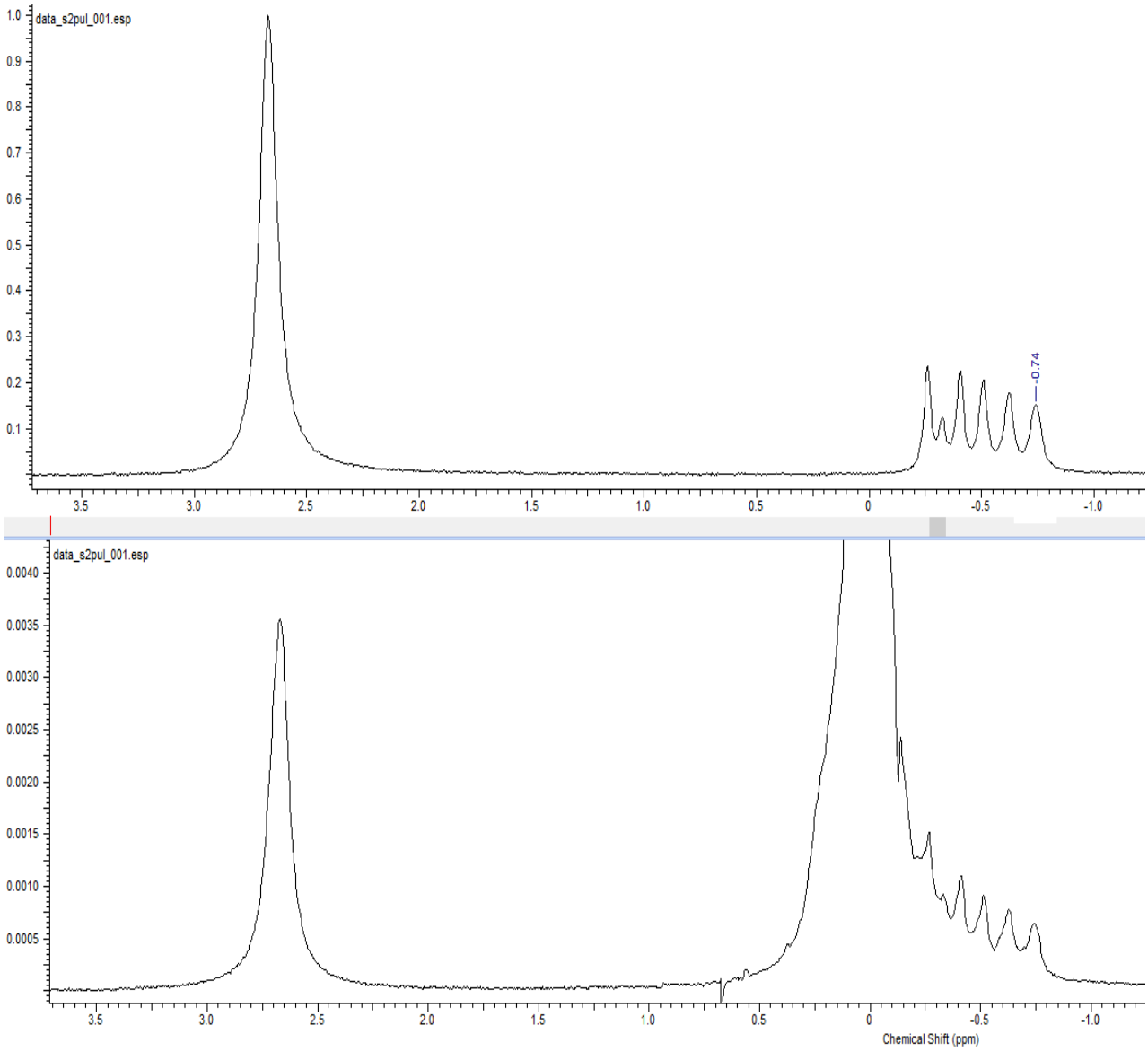

Figure 40: DDD ^{31}P spectra at 25 ˚C. The top image shows the spectrum of just the DDD sample, and the bottom shows the spectrum with the coaxil insert

A temperature study of the DDD control sequence was performed using the parameters outlined in materials and methods, shown below (Figure 41).

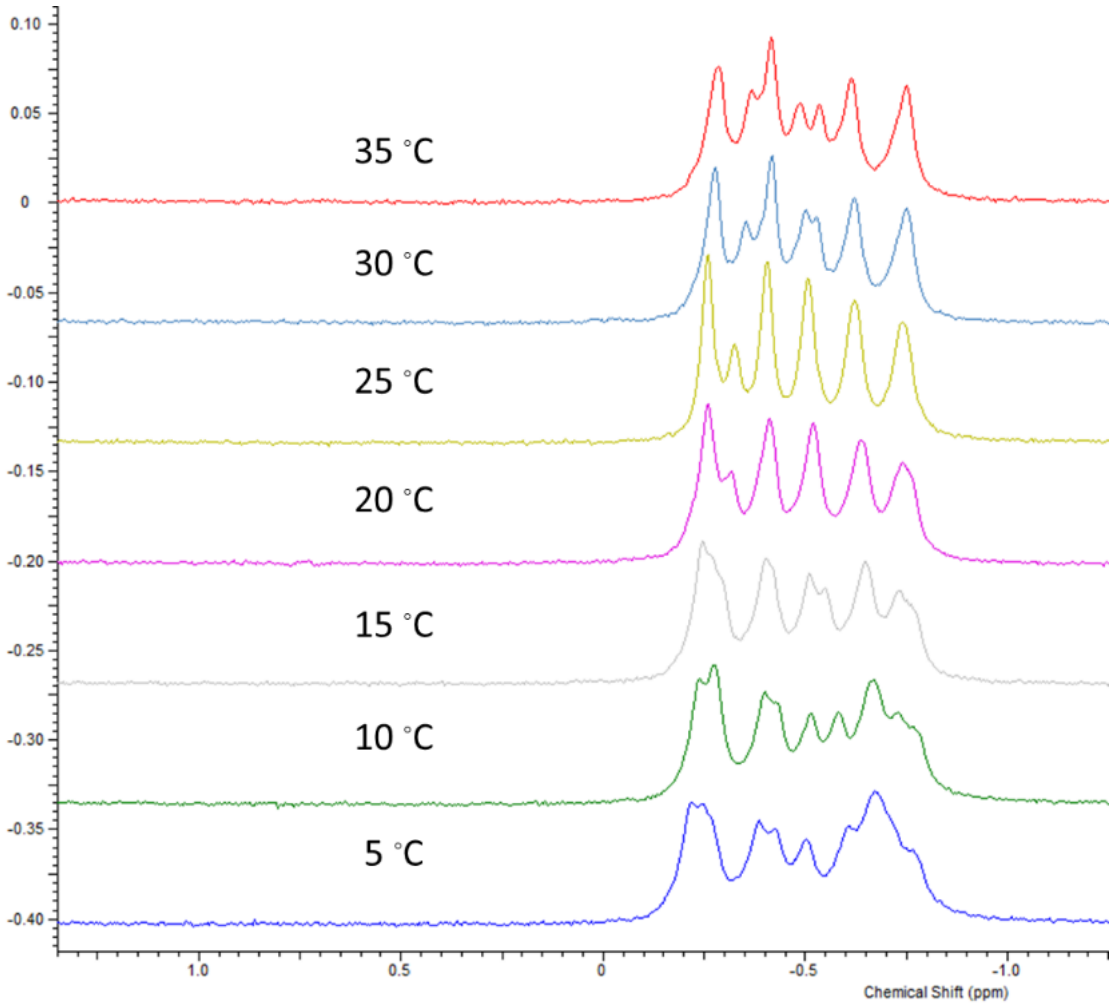

Figure 41: 1D Phosphorous Spectra of DDD overlaid, temperatures range from 5-35 ˚C

From there, the chemical shift of each phosphate was recorded, shown below (Table 4). The naming convention used for phosphorous is to number it by where it appears in the sequence, taking its number from the preceding nucleotide in the 5' direction. For example, the C1pG2 phosphorous is named 1, the G2pC3 phosphorous is named 2, and so on.

Table 4: Chemical Shifts of DDD Phosphate Backbone

^{31}P Chemical Shifts							
Position	35 ˚C	30 ˚C	25 ˚C	20 ˚C	15 ˚C	10 ˚C	5 ˚C
1	-0.41	-0.41	-0.41	-0.41	-0.40	-0.42	-0.42
2	-0.54	-0.53	-0.51	-0.52	-0.51	-0.50	-0.50
3	-0.37	-0.35	-0.33	-0.32	-0.30	-0.26	-0.25
4	-0.49	-0.50	-0.51	-0.52	-0.55	-0.57	-0.61
5	-0.62	-0.62	-0.62	-0.64	-0.65	-0.66	-0.67
6	-0.75	-0.74	-0.74	-0.76	-0.76	-0.76	-0.77
7	-0.75	-0.75	-0.74	-0.74	-0.73	-0.72	-0.72
8	-0.62	-0.62	-0.62	-0.64	-0.65	-0.66	-0.67
9	-0.29	-0.28	-0.26	-0.26	-0.25	-0.26	-0.25
10	-0.41	-0.41	-0.41	-0.41	-0.40	-0.39	-0.39
11	-0.29	-0.28	-0.26	-0.26	-0.25	-0.23	-0.22

From these chemical shifts, the % BII character of each phosphate was calculated using the method outlined by Tian *et al.* (Table 5).

Table 5: % BII Character of DDD Phosphate Backbone

	%BII Character						
Position	35 ˚C	30 ˚C	25 ˚C	20 ˚C	15 ˚C	10 ˚C	5 ˚C
1	39.0%	39.0%	39.0%	39.0%	40.2%	38.0%	38.0%
2	22.6%	24.4%	27.2%	26.4%	27.9%	29.3%	29.6%
3	44.0%	46.3%	48.5%	49.4%	51.3%	55.3%	56.0%
4	28.9%	28.0%	27.2%	26.4%	23.4%	21.7%	18.0%
5	12.5%	13.4%	14.2%	12.6%	12.3%	12.0%	11.7%
6	-3.8%	-1.3%	0.0%	-1.2%	0.0%	1.1%	1.1%
7	-3.8%	-2.5%	0.0%	1.1%	3.4%	5.4%	6.4%
8	12.5%	13.4%	14.2%	12.6%	12.3%	12.0%	11.7%
9	54.1%	54.9%	56.8%	56.3%	56.9%	55.3%	56.0%
10	39.0%	39.0%	39.0%	39.0%	40.2%	41.2%	41.2%
11	54.1%	54.9%	56.8%	56.3%	56.9%	58.6%	59.1%

Some % BII character values were calculated to be a negative percent, which is impossible. However, in the control sequence the most negative value is -3.8%, which is fairly close to zero, which could be attributed to experimental error. However, this could also potentially be due to an error in the method; the true chemical shift values for BI and BII could be outside of the bounds derived by Tian et al.

Once the % BII character of the control sequence was determined, the same process was performed again on the UdhU3D sequence. Original assignments of the lesioned phosphorous peaks was done by interpreting the HSQC spectrum (Figure 42).

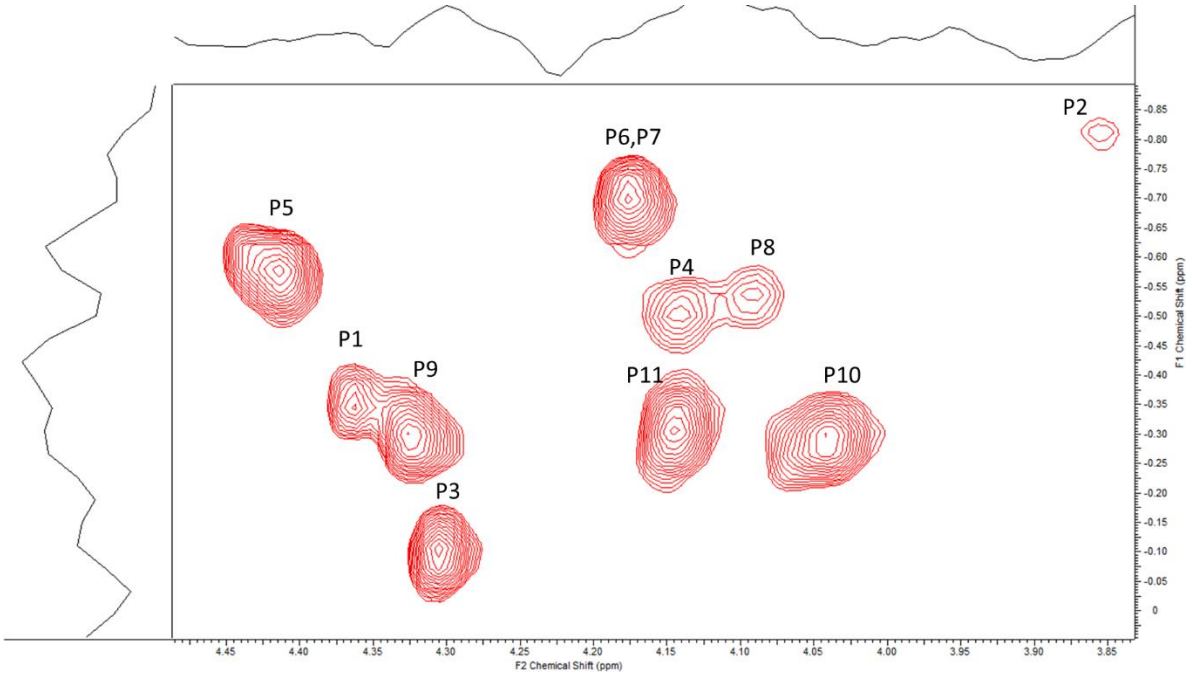

Figure 42: 4’ region of HSQC spectra of lesioned sequence with labelled phosphorous crosspeaks. The x-axis corresponds to the proton chemical shifts, while the y-axis corresponds to the phosphorous chemical shifts

The HSQC was solved by comparing the 3’ and 4’ assignments made in the NOESY to the crosspeaks found in the spectrum. Each phosphorous should have one crosspeak with the 3’ proton of the residue preceding it and one with the 4’ proton of the residue proceeding it. For example, P1 should have a crosspeak with the C1 3’ and the G2 4’ protons. The HSQC was performed and solved at 25 ˚C. It should be noted that the figure above only depicts the 4’ half of the HSQC, as it is difficult to have both the 3’ and 4’ regions of the spectrum be in phase at the same time.

Next, a temperature study of the lesioned sequence was performed using the parameters outlined in materials and methods, shown below (Figure 43). Assignments of each temperature were extrapolated from 25 ˚C.

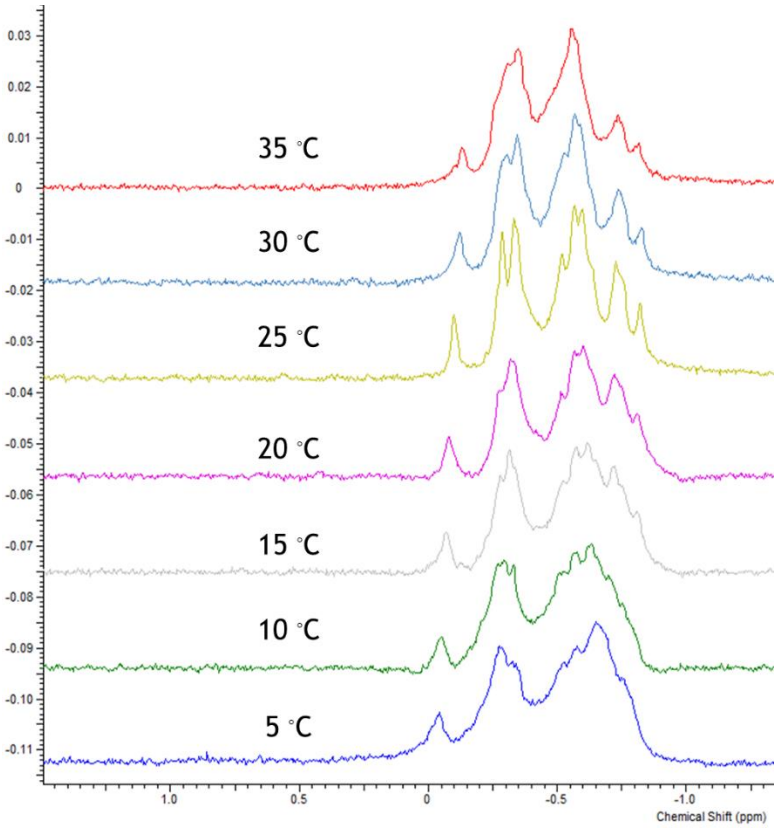

Figure 43: 1D Phosphorous Spectra of UdhU3D overlaid, temperatures range from 5-35 ˚C

Once the 1D phosphorous spectra for each temperature were solved, the chemical shift of each phosphate was recorded, shown below (Table 6).

Table 6: Chemical Shifts of UdhU3D Phosphate Backbone

^{31}P Chemical Shifts (ppm)							
Position	35 ˚C	30 ˚C	25 ˚C	20 ˚C	15 ˚C	10 ˚C	5 ˚C
1	-0.35	-0.35	-0.33	-0.32	-0.32	-0.29	-0.28
2	-0382	-0.83	-0.82	-0.81	-0.81	-0.77	-0.76
3	-0.13	-0.12	-0.10	-0.08	-0.07	-0.05	-0.04
4	-0.52	-0.53	-0.52	-0.51	-0.52	-0.52	-0.53
5	-0.57	-0.59	-0.60	-0.60	-0.62	-0.63	-0.65
6	-0.75	-0.75	-0.75	-0.75	-0.75	-0.75	-0.74
7	-0.74	-0.74	-0.73	-0.72	-0.72	-0.70	-0.69
8	-0.56	-0.57	-0.57	-0.57	-0.57	-0.57	-0.58
9	-0.31	-0.30	-0.29	-0.28	-0.28	-0.27	-0.28
10	-0.35	-0.35	-0.34	-0.33	-0.34	-0.33	-0.34
11	-0.31	-0.30	-0.29	-0.28	-0.28	-0.27	-0.28

From these chemical shifts, the % BII character of each phosphate was calculated using the method outlined by Tian *et al.* (Table 7). This was done using the equation

$$\%BII = \frac{100}{2.368 - 0.00511 0T}\delta P + \frac{134.5 - 0.2031T}{2.368 - 0.00511 0T}$$

where δP is the chemical shift of the phosphate in question, and T is the temperature in Kelvin. By assuming a two state fast-exchange model, we can use the chemical shift of each phosphate as a direct correlation to the % BII character of each individual phosphate.

Table 7: % BII of UdhU3D Phosphate Backbone

	%BII Character						
Position	35 ˚C	30 ˚C	25 ˚C	20 ˚C	15 ˚C	10 ˚C	5 ˚C
1	46.5%	46.3%	48.5%	49.4%	49.1%	52.1%	52.8%
2	-12.7%	-12.2%	-9.5%	-6.9%	-5.6%	0.0%	2.2%
3	74.2%	74.4%	75.7%	76.9%	77.0%	78.1%	78.1%
4	25.1%	24.4%	26.0%	27.6%	26.8%	27.1%	26.4%
5	18.8%	17.0%	16.5%	17.2%	15.6%	15.2%	13.8%
6	-3.8%	-2.5%	-1.2%	0.0%	1.1%	2.2%	4.3%
7	-2.6%	-1.3%	1.2%	3.4%	4.5%	7.6%	9.5%
8	20.1%	19.5%	20.1%	20.7%	21.2%	21.7%	21.2%
9	51.6%	52.4%	53.2%	54.0%	53.6%	54.3%	52.8%
10	46.5%	46.3%	47.3%	48.2%	46.9%	47.8%	46.5%
11	51.6%	52.4%	53.2%	54.0%	53.6%	54.3%	52.8%

One may notice that there are significant negative % BII character values for the P2 phosphate. These values were determined using a commonly accepted literature method, however, the bounds were determined experimentally, and this data could imply that the area between the pure BI and BII chemical shifts is greater than originally approximated by Tian *et al*. In addition, it is unlikely that the DNA structure has departed from the typical B-form conformation, as previous research has found that DNA strands containing this lesion do exist in the traditional B-form.[5]

5.3 Imino Studies

Imino studies were performed using the methods outlined above. The imino assignments of the control sequence were determined from literature.[5] Once assignments had been determined, the imino temperature study from between 5 and 25 °C was performed (Figure 44).

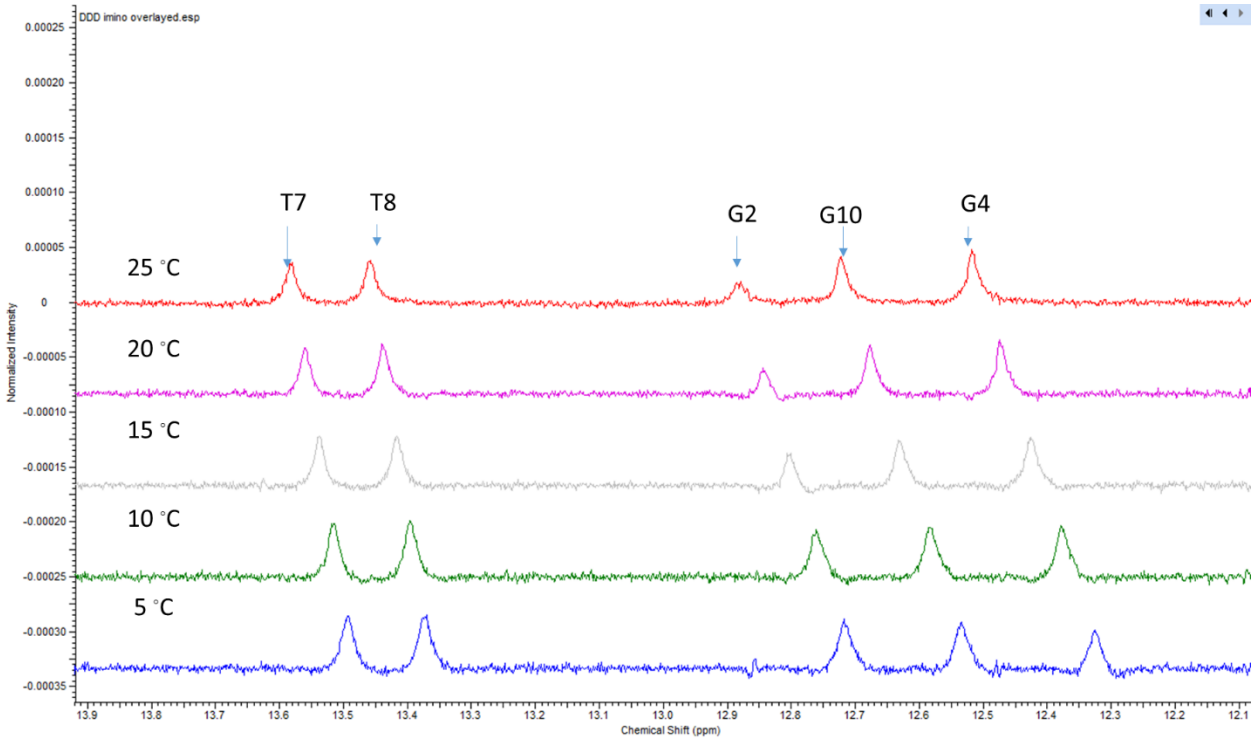

Figure 44: 1D Imino Temperature Study of DDD

One may notice that the 1 base pair is missing from the assignments above. This is due to the ends of the DNA strand being frayed and having very small double strand character.[9]

Once the assignments of the control sequence for all temperatures studied were acquired, the temperature study of the lesioned sequence was performed (see Figure 45).

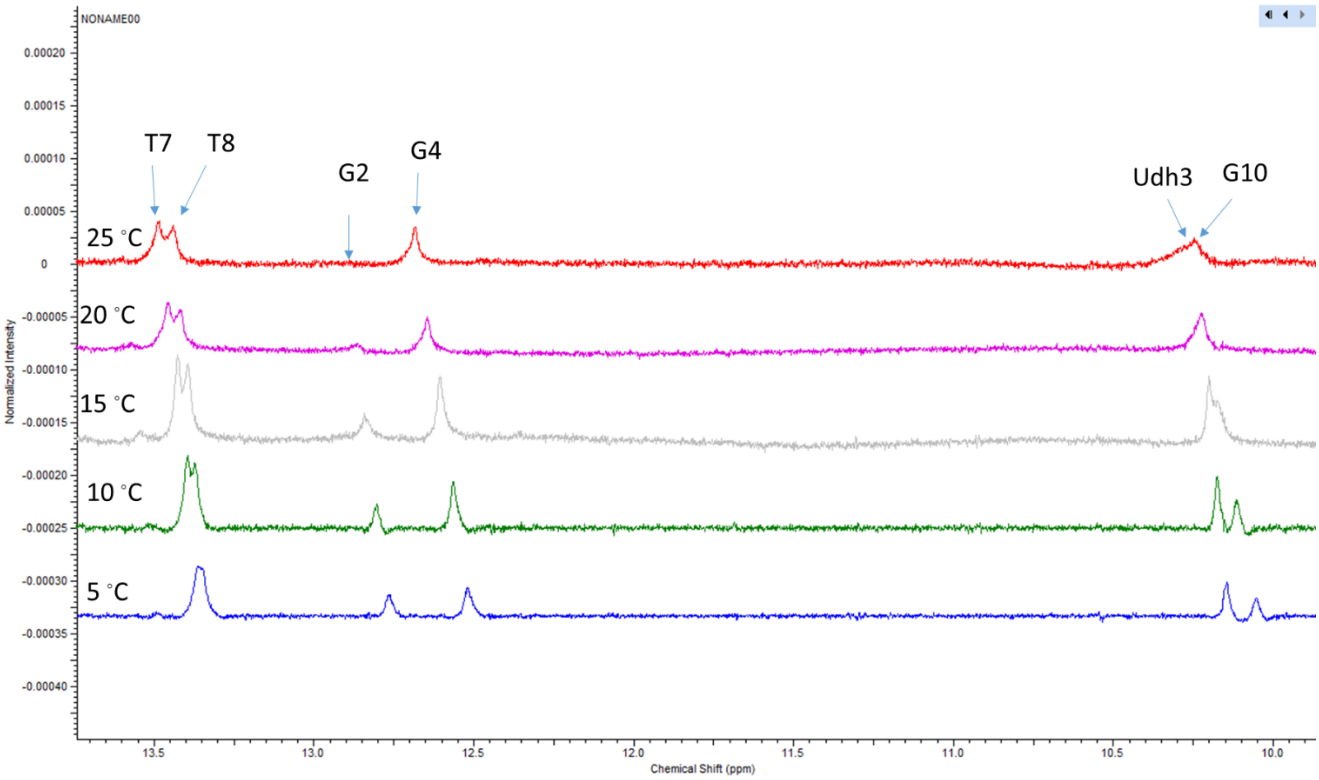

Figure 45: 1D Imino Temperature Study of UdhU3D

One may notice that the range of the imino peaks in the lesioned sequence is much larger than the control sequence. This is due to the G10 imino proton being base paired to a dihydrouracil instead of a cytosine, placing the proton in a different magnetic environment than the G10 in the control sequence.[25,28,40] There is also an additional peak seen in the lesioned sequence, as that dihydrouracil also has an imino proton in addition to the imino proton of the guanine it base pairs with. The G10 and Udh3 imino peaks were differentiated by analyzing an imino NOESY spectrum and consulting previous literature (Figure 46).[40]

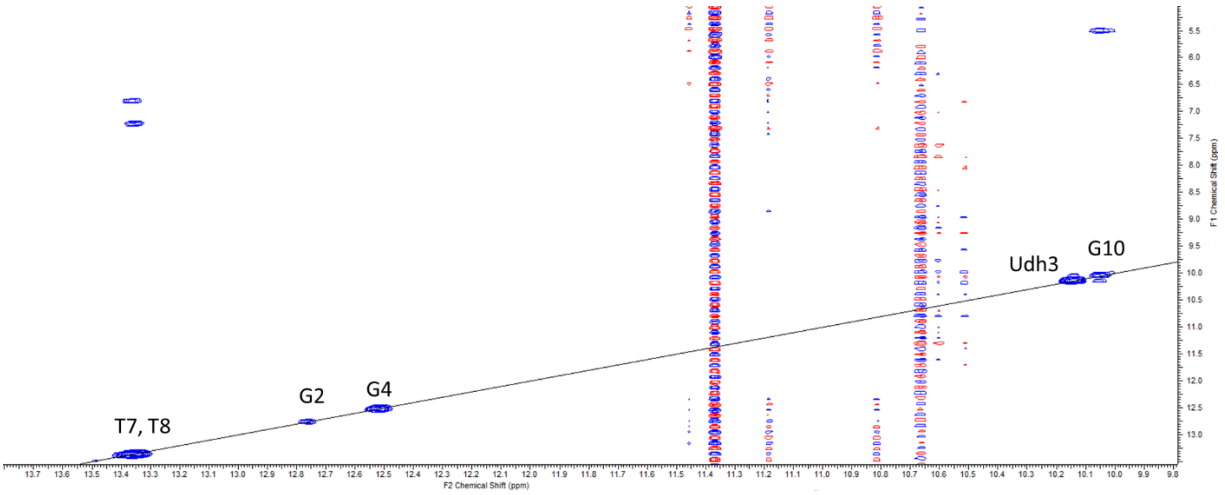

Figure 46: 2D NOESY spectrum of UdhU3D prepared in imino sample conditions.

One may notice that the phasing in this spectrum is poor when compared to others in this thesis, as evidenced by the horizontal lines through the middle of the spectrum. This is due to high quantities of water present in the sample. However, the peaks of interest are still interpretable. The identities of the Udh3 and G10 iminos were determined using the single crosspeak at 10.05 ppm on the x axis and 5.49 ppm on the y-axis. As of now, it is thought that this represents a through-space interaction between the G10 imino proton and the protons of the amine on the adjacent carbon on G10.

CHAPTER 6: DISCUSSION

In this section, the differences between the lesioned and control sequence will be analyzed and discussed. The first aspect of the two spectra to be analyzed is the difference in chemical shifts of the non-exchangeable protons (see Figure 47).

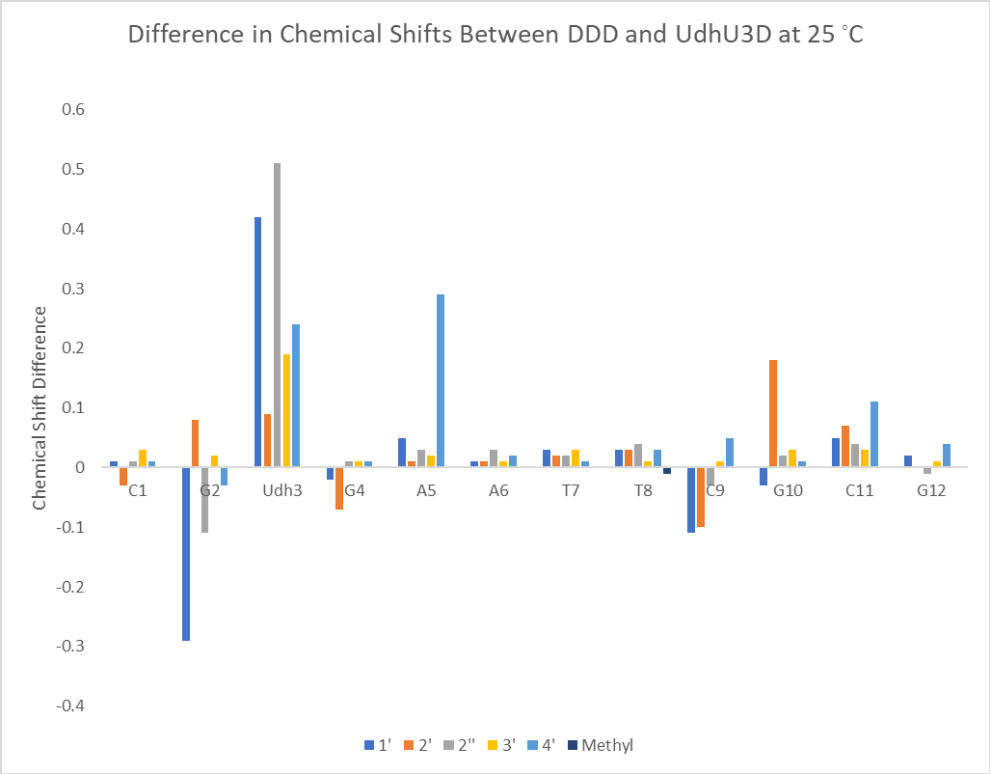

Figure 47: Difference in chemical shifts between the control and lesioned sequence at 25 ˚C. The base protons have been excluded as the difference is so great as to completely overshadow any other differences

For the purposes of this study, a significant difference for a non-exchangeable peak is roughly a difference of 0.1 ppm or higher. As shown above, the most significant differences are seen at the Udh3 position. This is to be expected as that is the lesion site, so it follows reason that the greatest differences would be seen there. There are also obvious significant differences at the

G2 position, which is to be expected as this residue is directly adjacent to the lesion. Interestingly, however, this is not true in the opposite direction, as the differences in chemical shift seen at the G4 position would not be considered significant. In addition, there were also significant differences seen at C9, G10, and C11 positions of the DNA strand. While this may seem counterintuitive, as these residues are far from the lesion site on the strand, they are fairly close to the lesion on the opposite strand. G10 is the base pairing partner of Udh3, so the observation that there are significant differences at this residue and its neighbors implies that the lesion may have an effect not only its own strand, but on the opposite strand as well. This same analysis was also performed at 10 ˚C (Figure 48).

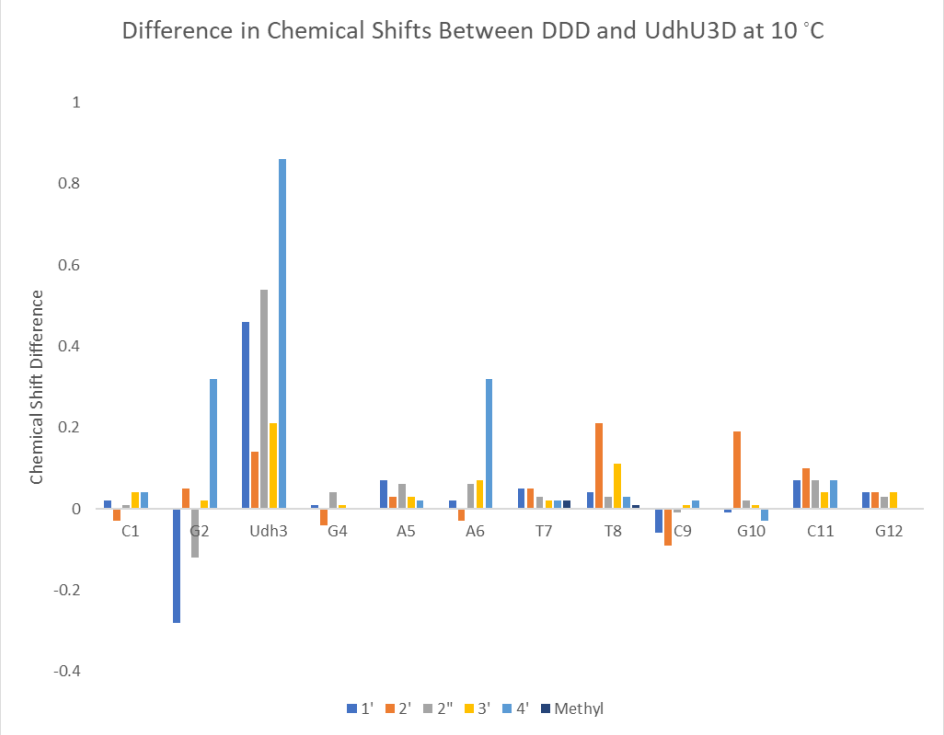

Figure 48: Differences in chemical shift between the control and lesioned sequence at 10 ˚C. The base protons have been excluded as the difference is so great that as to completely overshadow any other differences

The trend in the differences between the control and lesioned sequence seems to be consistent between the two temperatures sampled, with significant changes being seen at the Udh3, G2, G10, C11, and to a lesser extent C9 position. However, there are significant differences at the T8 position at 10 ˚C, but not at the 25 ˚C.

In addition to comparing the chemical shift, NOE intensities can also be compared between the control and lesioned sequence. This is done by looking for peaks that are either severely reduced in intensity or absent in the lesioned sequence that are present in the control sequence. At 25 ˚C, there weren't many changes visible; the only changes that could be observed in NOE intensities were observed in the Base-1' region (Figure 49).

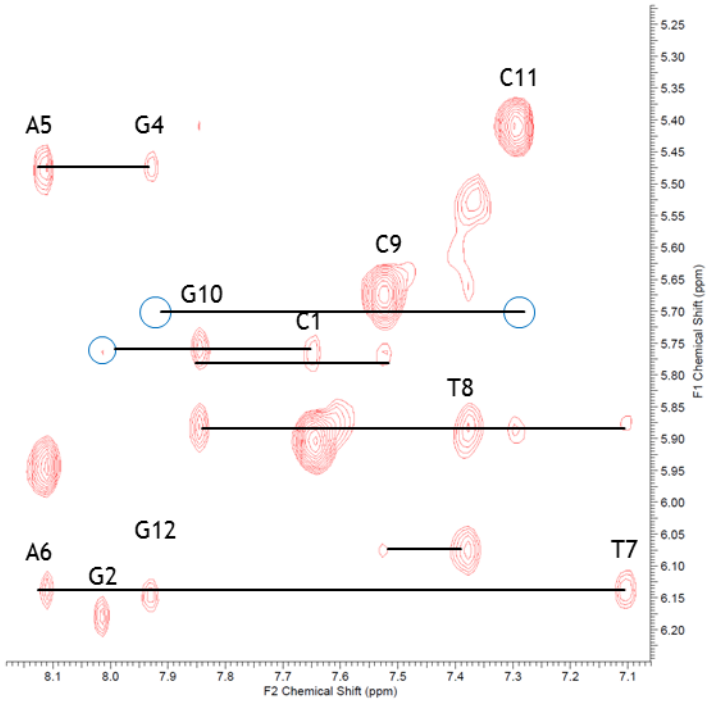

Figure 49: Base-1' region of UdhU3D NOESY Spectrum at 25 ˚C, circled areas represent missing NOE peaks when compared to the control

As shown above, there are missing crosspeaks observed between the C1 and G2 residues, as well as between the C11 and G12 residues. One may notice that there is not a visible walk either to or from the Udh3 residue.

This is due to the Udh base protons being non-aromatic and sp3 hybridized, causing the chemical shifts for these protons to appear in an entirely different region of the spectrum, making it difficult to determine if the walks to or from this lesion have been affected or not.

More significant differences in NOE intensity were seen at 10 ◦C (Figure 50).

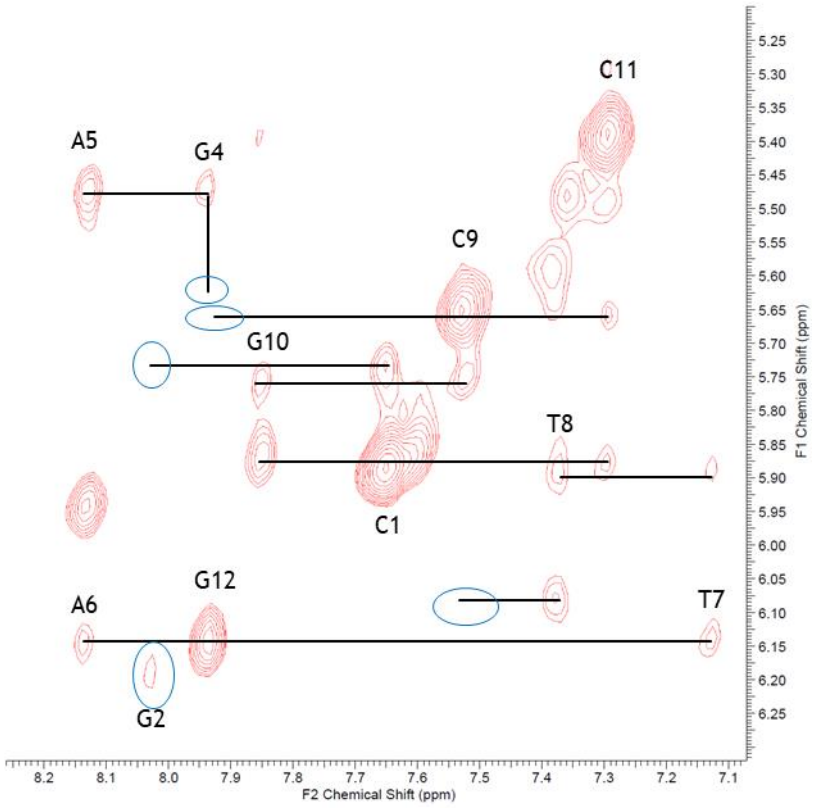

Figure 50: Base-1' region of UdhU3D NOESY Spectrum at 10 ◦C, circled areas represent missing NOE peaks when compared to the control

In addition to the missing crosspeaks present there are now additional missing peaks in the walk between C1 and G2, as well as a missing crosspeak between T8 and C9. The crosspeak

between the G2 base proton and G2 1’ proton, while technically present, is much weaker than it is in the control sequence. Due to this, it is included in the count of missing crosspeaks. Unlike at 25 ˚C, there are missing crosspeaks found within the Base-2’/2” region at 10 ˚C (see Figure 51).

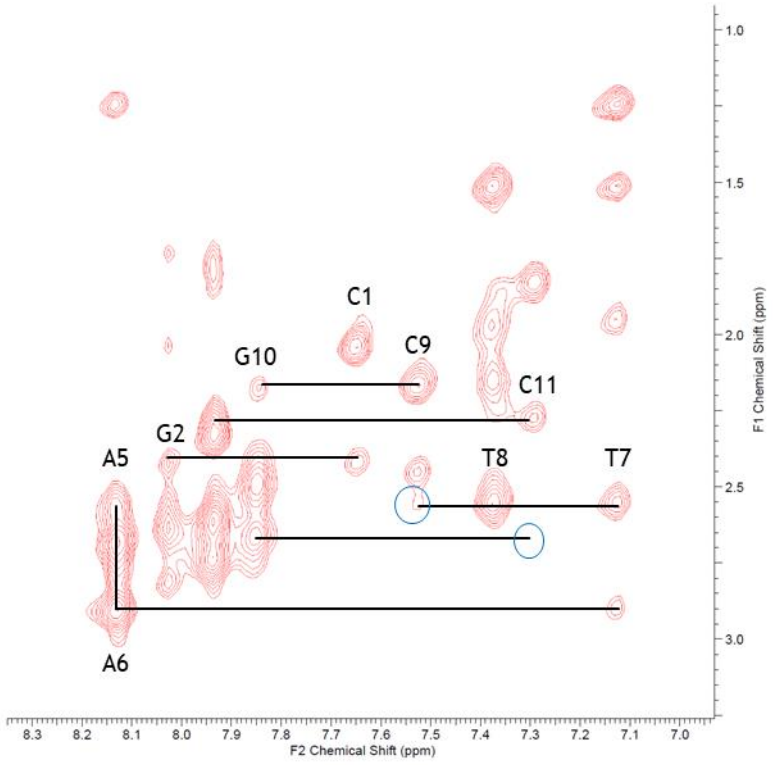

Figure 51: Base-2’/2” region of UdhU3D NOESY Spectrum at 10 ˚C, circled areas represent missing NOE peaks when compared to the control

In this region, missing crosspeaks are observed in the 2” walk between T8 and C9 and the 2” walk between G10 and C11. Interestingly, at this temperature there are no missing crosspeaks associated with the residues that neighbor the lesion.

The imino temperature study of the lesioned sequence vs the control sequence implies that dihydrouracil lesion has a destabilizing effect on the double stranded DNA. Double stranded character can be observed through the imino peaks; as a general rule, the stronger the imino

peak, the more double stranded character the base pair associated with that imino peak has.[9] This can be seen with the missing peak associated with the 1 base pair, as in literature, that peak doesn't appear above -5 ˚C in the Drew Dickerson Dodecamer.[9] In the temperature study performed for this thesis, the temperature was taken in 5 ˚C increments from 5 ˚C to 25 ˚C. In the control sequence, the intensities of all peaks remained mostly consistent throughout all temperatures studied, with only slight decrease in intensity for 2 at 25 ˚C. However, the lesioned sequence saw a significant drop off in intensity for 2 between 15 ˚C and 20 ˚C, with the peak completely vanishing at 25 ˚C. This implies that the fraying has spread to include the 2 base pair as well in the lesioned sequence. In addition, peak broadening was seen on 3 and the Udh imino proton peak as well at 15 ˚C, where the two peaks begin to overlap, with the broadening causing them to be completely overlapped at 20 ˚C. It should also be stated that peak broadening has also been implicated in increased DNA fraying.[9] In addition, a drop in intensity for the overlapped peaks is also seen at 25 ˚C. This shows that the dihydrouracil lesion does have a significant effect on the stability of the DNA strand, in the form of a dramatic increase in the amount of fraying occurring.

Once the % BII character of each phosphate was determined, the overall trends of each phosphate over the temperatures studied were analyzed (Figures 52 and 53).

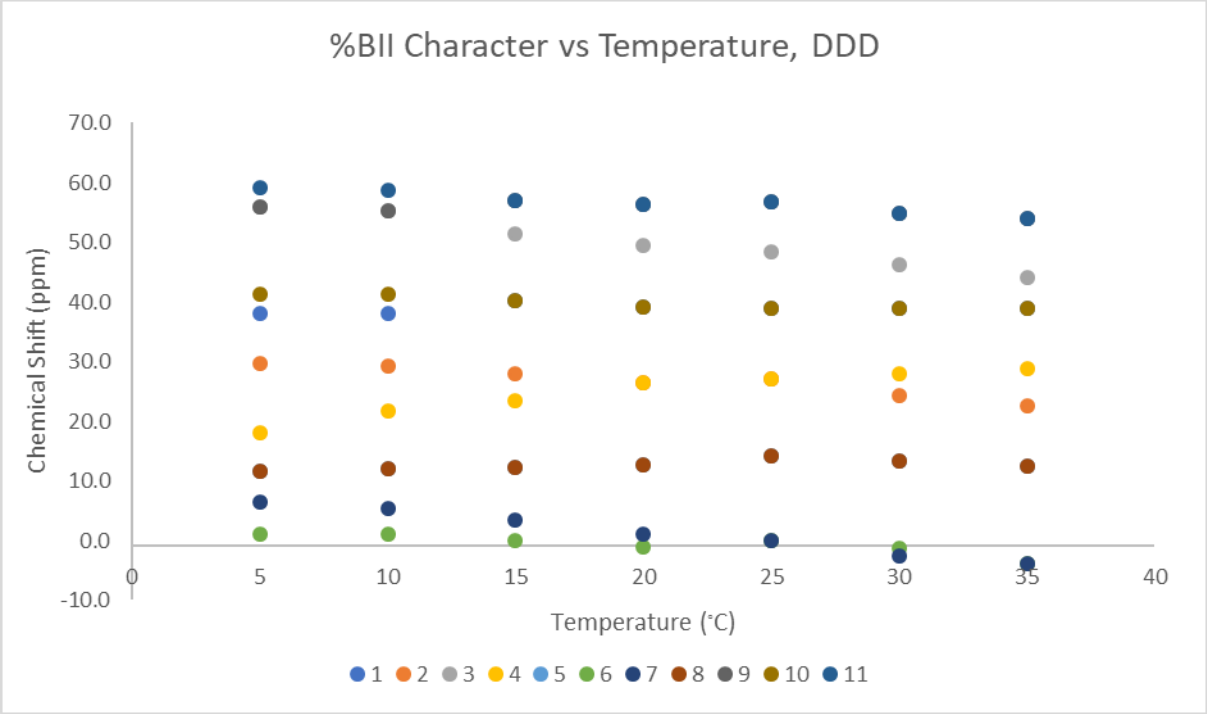

Figure 52: The % BII Character of DDD Phosphates *vs* Temperature

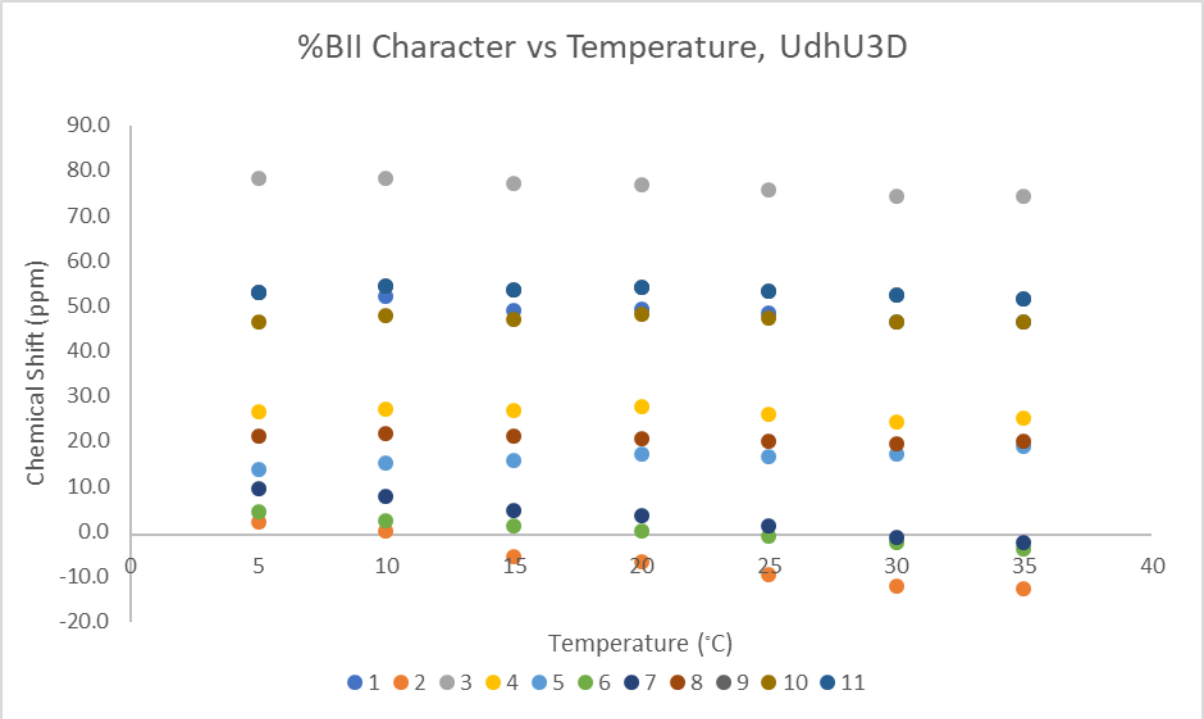

Figure 53: The % BII Character of UdhU3D Phosphates *vs* Temperature

One interesting difference between the lesioned and control sequences is that in the control sequence, P5 and P8 are overlapped and have the same % BII character no matter the temperature studied. However, in the lesioned sequence these two phosphates have resolved peaks, and thus resolved % BII characters. In addition, while the P5 trend shows a steady increase in % BII character with temperature in both sequences, P8 only shows this increase in the control, and in the lesioned sequence remains relatively constant across the temperatures studied.

The most obvious differences in % BII character are seen in the P2 and P3, with differences of 36.7 % and -27.2 %, respectively at 25 ˚C (Figure 54).

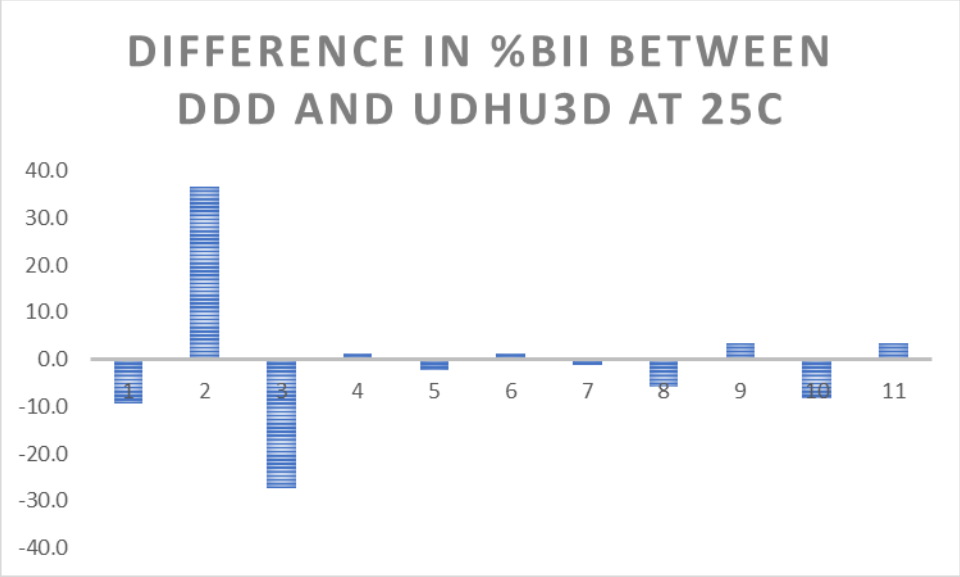

Figure 54: Difference in % BII between DDD and UdhU3D at 25 ˚C

In addition to the differences in P10, while not as extreme as the differences seen in P2 and P3, are still significant, with a difference of -8.3% at 25 ˚C. While the P10 phosphate is

relatively far away from the lesion on the same strand as it, it is very near the lesion on the sister strand of the double stranded DNA. This implies that the lesion is having some cross-strand effect on the phosphate backbone, which is also seen in the NOESY spectra. These trends are consistent across all temperatures observed (Figure 55).

These results for % BII are significant as they offer a quantitative way to determine how much of an effect a lesion is having on a DNA molecule; with the NOESY spectra alone, one can make assumptions about how the lesion is affecting the local tertiary structure, but it is difficult to make any conclusive statements about it without the help of computational studies or some other NMR experiments.

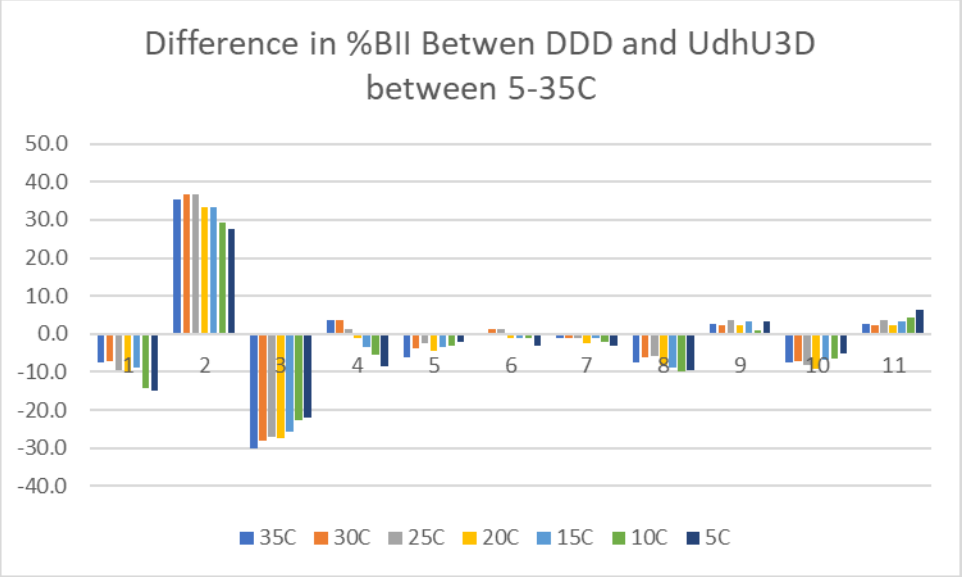

Figure 55: Difference in %BII across all temperatures studied

Earlier research into this lesion has hypothesized that the method of recognition for the dihydrouracil lesion is a base-flipping mechanism brought about when the repair enzyme comes into contact with the base, and that structural differences when the repair enzyme is not in contact

with the DNA is not likely to be a key element in recognition[3,5]. While the lesion does not cause the DNA to lose its B-form structure, the data from this thesis shows that there is significant changes in the % BII character caused by the lesion, specifically in phosphates 2 and 3, the two phosphates connected to the lesion site. This supports the idea that % BII character is a possible mode of indirect readout used to identify lesioned areas of DNA. As stated above, the most dramatic differences in % BII character were observed at the two phosphates attached to the lesioned base. In addition, there is a directionality to the difference, as the 3' phosphate of the Udh3 saw a dramatic decrease in % BII, and the 5' phosphate saw a dramatic increase. Not only that, but in the lesioned sequence, these two phosphates consistently differ in % BII character by at least 75% BII character, as opposed to the two phosphates in the control sequence (Figure 56 and 57).

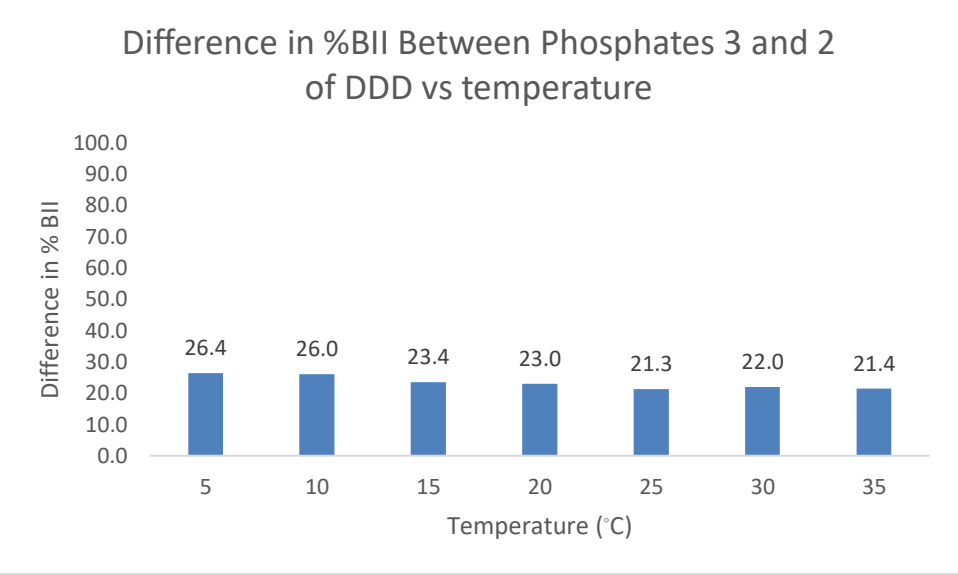

Figure 56: Difference in % BII Between Phosphates 3 and 2 of DDD *vs* Temperature (°C)

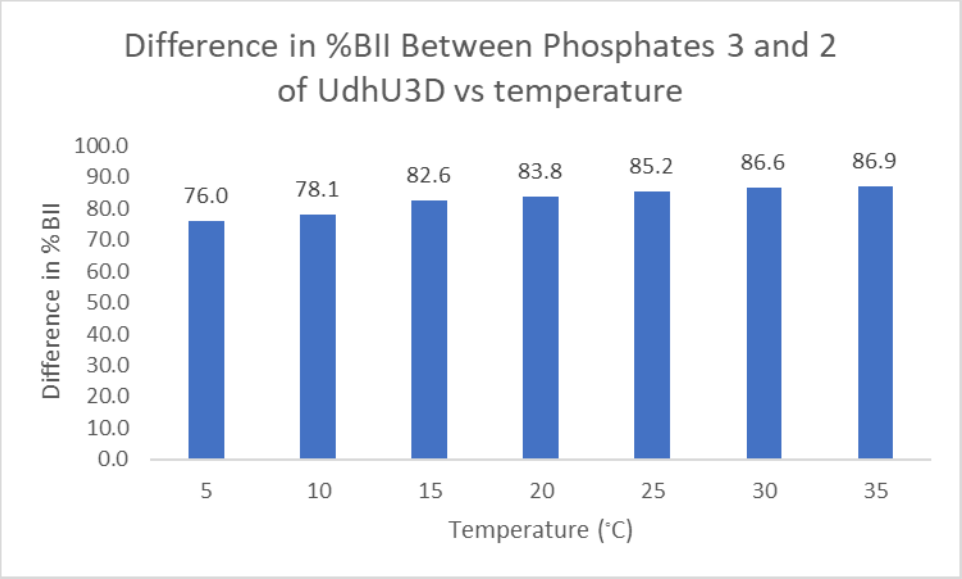

Figure 57: Difference in % BII Between Phosphates 3 and 2 of UdhU3D *vs* Temperature (°C)

It can be argued that the change in tertiary structure being detected could be a large decrease then dramatic increase in the width of the minor groove of the DNA, as it has been shown previously that %BII character has a positive correlation with the width of the minor groove.[22]

CHAPTER 7: CONCLUSIONS

In conclusion, in this thesis the effect that the dihydrouracil lesion has on the % BII character of double stranded DNA was studied using various NMR spectroscopic techniques. The process began with determination of the proton chemical shifts from the NOESY spectrum, followed by determination of the phosphorous chemical shifts at 25 ˚C from the HSQC spectrum. The chemical shifts of each phosphate across multiple temperatures was determined by interpreting a series of 1D ^{31}P spectra of the DNA sequence. The phosphate chemical shifts were then analyzed in the method outlined by Tian *et al*. Our data shows that the dihydrouracil lesion has a significant effect on the % BII character on the phosphodiester backbone of DNA, in particular in phosphates 2 and 3, the neighboring phosphates of the lesion. These phosphates in the lesioned sequence at 25 ˚C had a % BII character of -9.5% and 75.7%, respectively. This gave a difference in % BII character between the lesioned and control sequence of 36.7 % and -27.2 %, respectively. In addition, the difference in % BII character between phosphates 2 and 3 within the same sequence were calculated for both the lesioned sequence and the control sequence. In the lesioned sequence, this difference was 85.2 % at 25 ˚C, versus only 21.3 % in the control sequence. In addition, the trends in % BII seen at 25 ˚C held true for all temperatures studied. This data led to the conclusion that changes in the conformation of the phosphodiester backbone, particularly changes in % BII character, could be a possible recognition factor for repair enzymes for the dihydrouracil lesion.

In addition to the % BII studies, the effect the lesion has on the thermal stability of double stranded DNA was also studied using the imino NMR experiments. The data saw a significant decrease in peak intensities at both 20 ˚C and 25˚C for the lesioned sequence that was

not present in the control sequence. This decrease in peak intensity implies a decrease in double stranded character of the ends of the DNA molecule, or an increase in fraying. This in turn suggests that the dihydrouracil lesion may have a destabilizing effect on the thermal stability of double stranded DNA.

While the analysis of the phosphodiester backbone can give quantifiable result about the tertiary structure of DNA, it does not give a complete picture. Future work for this project would include computational studies utilizing the NOESY crosspeaks in order to quantify the distance between individual protons. In addition, it would be interesting to observe if the effects of this lesion change in response to where it occurs within the DNA strand.

CHAPTER 8: REFERENCES

(1) Friedberg, E. C.; Walker, G. C.; Siede, W.; Wood, R. D.; Schults, R. A.; Ellenberger, T. *DNA Repair and Mutagenesis,* 2nd ed. ASM Press: Washington, D.C., 2006.

(2) Liu, J.; Zhou, W.; Doetsch, P.W. *Mol. Cell. Biol.* **1995**, *15* (2), 6729-6735.

(3) Liu, J.; Doetsch, P. W. *Nucleic Acids Research* **1998**, *26* (7), 1707-1712.

(4) Tian, Y.; Kayatta, M.; Shultis, K.; Gonzalez, A.; Mueller, L. J.; Hatcher, M. E. *J. Phys. Chem. B* **2009**, *113*, 2596-2603.

(5) Villanueva, J. M.; Pohl, J.; Doetsch, P. W.; Marzilli, L. G. *J. Am. Chem. Soc.* **1999**, *121*, 10652-10653.

(6) Mathews, C. K.; Van Holde, K. E.; Appling, D. R.; Anthony-Cahill, S. J. *Biochemistry*, 4th ed. Pearson: Toronto, 2013.

(7) Hare, D. R.; Wemmer, D. E.; Chou, S.; Drobny, G. *J. Mol. Biol*. **1983**,*171*, 319-336.

(8) Voet, D.; Voet, J. G.; Pratt, C. W. *Fundamentals of Biochemistry*. John Wiley & Sons.Inc, 2013.

(9) Patel, D. J.; Kozlowski, S. A.; Marky, L. A.; Broka, C.; Rice, J. A.; Itakura, K.; Breslauer, K. J. *Biochemistry* **1982**, *21*, 428-436.

(10) Michalczyk, R.; Russu, I. M. *FEBS* **1993**, *331* (3), 217-222.

(11) Patel, D. J.; Kozlowski, S. A.; Marky, L. A.; Rice, J. A.; Broka, C.; Dallas, J.; Itakura, K.; Breslauer, K. J. *Biochemistry* **1982**, *21*, 437-444.

(12) Mamian-Lopez, M.; Temperini, M. L. A. *Anal. Chem.* **2018**, *90*, 14165-14172.

(13) Zhang, X.; Gaffney, B. L.; Jones, R. A. *J. Am. Chem. Soc.* **1997**, *119* (27), 6432-6433.

(14) Nair, D.T.; Johnson, R.E.; Prakash, S.; Prakash, L.; Aggarwal, A. K. *Nature* **2004**, *430*, 377-380.

(15) Nikolova, E. N.; Kim, E.; Wise, A. A.; O'Brien, P. J.; Andricioaei, I.; Al-Hashimi, H. M. *Nature* **2011**, *470* (7335), 498-502

(16) Alberts, B.; Johnson, A.; Lewis, J. et al. *Molecular Biology of the Cell, 4th ed.* New York: Garland Science, 2002.

(17) Jellinek, T.; Johns R. B. *Photochemistry and Photobiology*. **1970**, *11* (5), 349-359.

(18) Pitts, J. N. Jr[2]; Letsinger, R. L.; Taylor, R. P.; Patterson, J. M.; Recktenwald, G.; Martin, R. B. *J. Am. Chem. Soc.* **1958**, *81*, 1068-1077.

(19) Schermerhorn, K. M.; Delaney, S. *Acc. Chem. Res.* **2014**, *47* (4), 1238-1246.

(20) Sancar, A.; Lindsey-Boltz, L. A.; Unsal-Kacmaz, K.; Linn, S. *Annu. Rev. Biochem.* **2004**, *73*, 39-85.

(21) Beard, W. A.; Wilson, S. H. *Biochemistry* **2014**, *53*, 2768-2780.

(22) Gorenstein, D. G. *Chem. Rev.* **1994**, *94*, 1315-1338.

(23) Xu, X.; Imeddourene, A. B.; Zargarian, L.; Foloppe, N.; Mauffret, O.; Hartmann, B. *Biochemistry* **2014**, *53*, 5601-5612.

(24) Heddi, B.; Foloppe, N.; Bouchemal, N.; Hantz, E.; Hartmann, B. *J. Am. Chem. Soc.* **2006**, *128* (28), 9170-9177.

(25) Jacobson, N. E. *NMR Spectroscopy Explained; Simplified Theory, Application and Examples for Organic Chemistry and Structural Biology*. John Wiley and Sons, Inc.: Hoboken, NJ, 2007.

(26) Miessler, G. L.; Fischer, P. J.; Tarr, D. A. *Inorganic Chemistry 5th ed.* Pearson Education, Inc.: Toronto, 2014.

(27) Freifelder, D. *Physical Biochemistry 2nd ed*. Freeman and Company: San Francisco, CA, 1982.

(28) Doucleff, M.; Hatcher-Skeers, M.; Crane, N. J. *Pocket Guide to Biomolecular NMR*. Springer: New York, 2011.

(29) Engel, T.; Reid, P. *Physical Chemistry*, 2nd ed. New York: Prentice Hall, 2010.

(30) Cooksy, A. *Physical Chemistry: Quantum Chemistry and Molecular Interactions.* Pearson Education, Inc.: Toronto, 2014.

(31) Levitt, M. H. *Spin Dynamics: Basics of Nuclear Magnetic Resonance*. John Wiley and Sons: West Sussex, England, 2001.

(32) Morrison, J. D. *Organic Chemistry*. Wadsworth Publishing Company: Belmont, CA, 1979.

(33) Drago, R. S. *Physical Methods for Chemists, 2nd ed.* Saunders College Pub, Gainesville, FL, 1992.

(34) Chary, K. V. R.; Govil, G. *NMR in Biological Systems*. Dordrecht: Springer 2008.

(35) Meints, Gary. Personal communication

(36) Noggle, J. H.; Schirmer, R. E. *NuclearOverhauser Effect: Chemical Applications* New York and London: Academic Press, 1971.

(37) Dickerson, R. E.; Wing R. M.; Takano T.; Broka C.; Itakura K. Dickerson R.E. *PNAS* **1981**, *78* (4), 2179-2183

(38) Gottlieb, H.; Kotlyar, V.; Nudelman, *J. Org. Chem.* **1997**, *62*, 7512-7515

(39) Ott, J.; Eckstein, F. *Biochemistry* **1985**, *24*, 2530-2535

(40) Uesugi, S.; Nakagawa, E.; Ohtsuka, E.; Ikehara, M.; Watanabe, M.; Kobayashi, Y.; Kyogoku, Y.; Kainosho, M. *J. Am. Chem. Soc.* **1982**, *104*, 7340-7341