


Spring 5-28-2017

# 2D NMR structural study of a DNA duplex containing two unnatural base pairs

Weston Preising  
*University of San Diego*

Follow this and additional works at: [https://digital.sandiego.edu/honors\\_theses](https://digital.sandiego.edu/honors_theses)

 Part of the [Biochemistry Commons](#), and the [Structural Biology Commons](#)

---

## Digital USD Citation

Preising, Weston, "2D NMR structural study of a DNA duplex containing two unnatural base pairs" (2017). *Undergraduate Honors Theses*. 34.

[https://digital.sandiego.edu/honors\\_theses/34](https://digital.sandiego.edu/honors_theses/34)

This Undergraduate Honors Thesis is brought to you for free and open access by the Theses and Dissertations at Digital USD. It has been accepted for inclusion in Undergraduate Honors Theses by an authorized administrator of Digital USD. For more information, please contact [digital@sandiego.edu](mailto:digital@sandiego.edu).

**Weston Preising**  
**Dr. Tammy Dwyer**  
**HNRS Thesis**  
**21 May 2017**

## **2D NMR structural study of a DNA duplex containing two unnatural base pairs**

### **ABSTRACT**

---

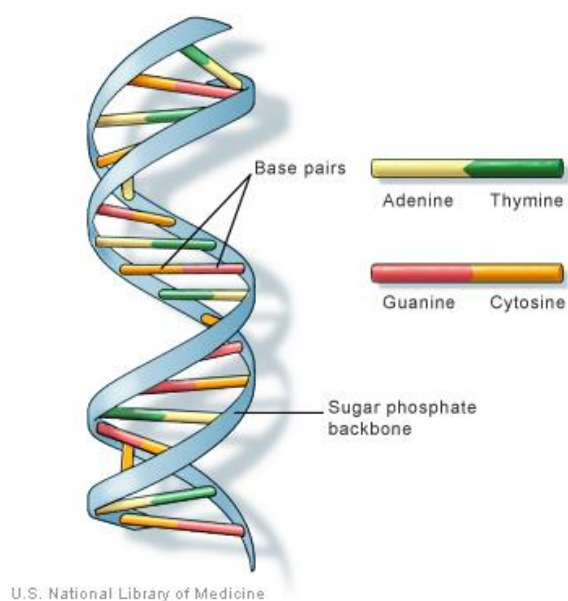
DNA is the genetic blueprint responsible for the traits of living organisms. The "genetic alphabet" consists of four molecules called nucleobases, represented by the letters A, T, G, and C strung together in a "strand." A DNA duplex consists of two "strands" held together such that A is across from T and G across from C, referred to A-T and G-C "base pairs". Recently, the genetic alphabet or code has been expanded by synthetic biologists in an effort to unveil new insights to the machinery of DNA replication, and potentially develop new medicines to cure diseases like HIV. The new unnatural base pair 5SICS-NaM is the first to be stably replicated by a semi-synthetic *E. Coli* organism. This project investigates the structure of a novel DNA duplex containing two 5SICS-NaM pairs arranged side-by-side using a technique called NMR spectroscopy. Thus far, NMR chemical shift assignments for this duplex have been largely completed which provide some information as to the structural configuration of the base pairs. Preliminary results indicate that one of the two adjacent 5SICS-NaM pairs is situated within the duplex, whereas the other may have the 5SICS base pushed out of the duplex with its paired NaM base stacked quite normally. Further analysis of the data should allow us to propose a more precise structure of the DNA duplex as well as provide insight into the versatility of DNA replication machinery to recognize altered DNA structures.

### **INTRODUCTION**

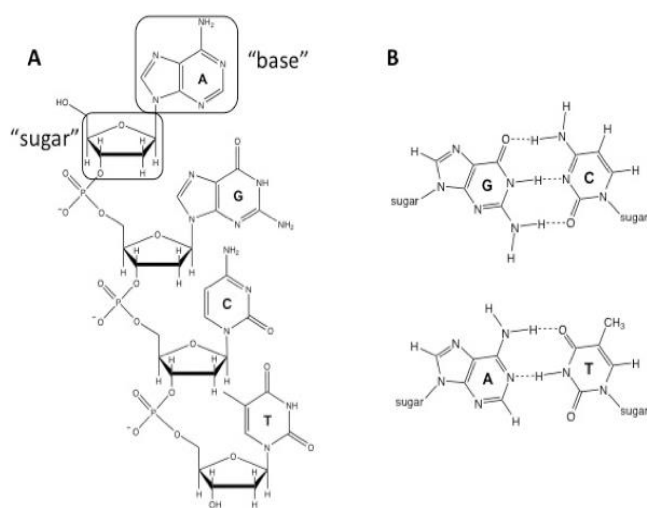
---

For as long as there has been life on Earth, there has been DNA. Deoxyribonucleic acid (DNA) is the genetic blueprint responsible for the observable and sub-visual traits of living systems. Without it, we would not be able to survive or reproduce. However, as imperative as this DNA genetic system is to life, it was only extensively investigated in 1953 by scientists James Watson and Francis Crick (1).

Watson and Crick were responsible for determining the antiparallel double helical stranded structure of DNA (Fig. 1), its connected strands held by merely four nucleotides arranged in specific pairs shown in Figure 2 (1). These strands are held together by a number of intermolecular forces including hydrogen bonds and base-stacking interactions, the latter of which plays a more prominent role in stability than originally thought. The order of these four nucleotides or bases: guanine, adenine, thymine and cytosine--or G, A, T, and C--hold the coding key for life (1). Additionally, Watson and Crick found that there are complementary rules, as purines (G and A) associate with pyrimidines (T and C) to achieve the characteristic "template-directed replication...with the high fidelity essential to life" (5).



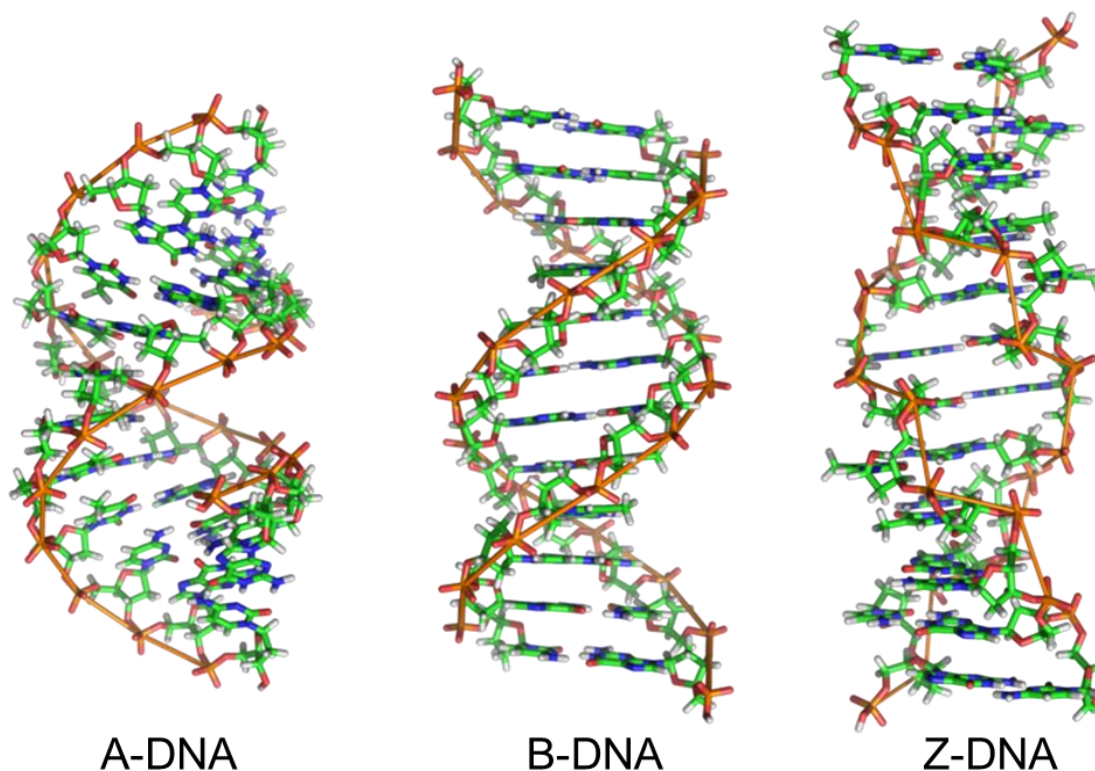
**Figure 1. Double helical structure of DNA.** Structure formation formed from nucleotide base pairs on opposite strands that are attached to a deoxyribose sugar-phosphate backbone. The four natural nucleotides of DNA are adenine (A), thymine (T), guanine (G) and cytosine (C). Sequences of these natural bases store the coding information required for creating and maintaining an organism.<sup>1</sup>



**Figure 2. Natural Watson-Crick single DNA strand and base pairing.** (A) A depiction of a single DNA strand (5' to 3') with all four natural bases: adenine, guanine, cytosine and thymine; *sugar*, is ribose; (B) Watson and Crick base pairing between purines (G & A) and pyrimidines (C & T).

The structure provided in Figure 1 is an image of dsDNA in B-form, which is the form of DNA that Watson and Crick discovered, as well as the most common form of DNA (11).

However, there are two additional forms of nucleic acids duplexes, called A-form and Z-form (Fig. 3). The nucleic acid duplexes that have A-form structure are RNA-DNA and RNA-RNA duplexes and while this form maintains the right-handed structure of B-DNA, A-form has different helical parameters, which can be seen in Figure 3 (11). A third structure of nucleic acid duplexes is called Z-form, which has a left handed helix and is thought to be necessary for certain cellular functions (11).



**Figure 3.** The three canonical double helical DNA structures.<sup>1</sup>

Recently, efforts in several labs around the world have made progress creating synthetic, analogous DNA building blocks to unveil potential secrets that nature may be hiding. This emerging field called "synthetic biology" (1), coined by Dr. Steve Benner, attempts to design and explore the intricacies of a synthetic DNA system with the hope of discovering new biological properties, as well as to make progress in biomedical efforts (2). Synthetic biology is rooted in the principle that "synthesis drives discovery in ways that analysis cannot" (3). In other words, it provides an alternative research method centered on creating synthetic systems to mimic natural biological properties (3).

One method of utilizing synthetic biology is in the introduction of unnatural base pairs (UBPs) into the existing natural system (6). The question then arises: "If we understand the biochemical machinery required to replicate DNA, can we expand the "genetic alphabet" and is there any value in doing so?" (4). Some scientists argue that there must be some concrete reason to why the four nucleotides have persisted for "eons of biological evolution" (4). However, if the genetic alphabet can be expanded, new types of amino acids could be

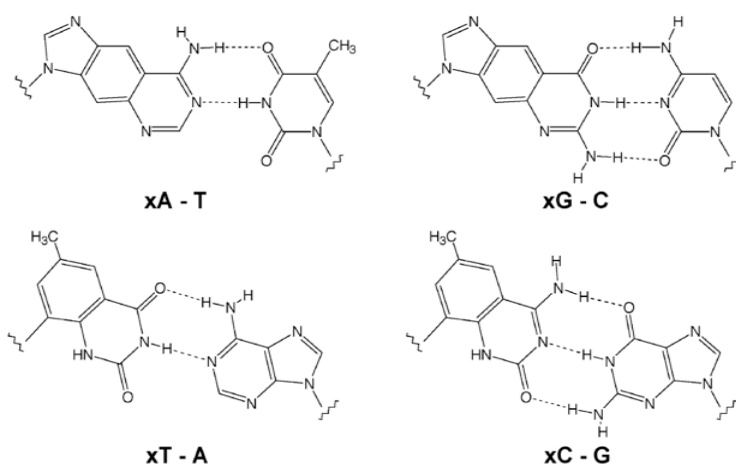
---

<sup>1</sup>Lasbury, M. DNA is As Easy As A, B, Z <http://biologicalexceptions.blogspot.com/2013/10/dna-is-as-easy-as-b-z.html> (accessed May 5, 2017).

produced in order to design new proteins and hence possibly even new medicines, catalysts, biological tools (4), and maybe even new organisms.

At the end of the 1980s, Benner created the first synthetic DNA molecule, but modified its structure to be more flexible, which allowed for the incorporation of an additional 12 synthetic nucleosides beyond the four natural nucleotides (1). The results were profound. Benner's research created a new "diagnostic tool for doctors" (1), which "helps manage the care of ~ 400,000 patients infected with HIV and hepatitis viruses each year" (3). Benner and his team hope to use this discovery to eventually create a "personalized medicine" system that is "tailored to each person's individual genetic code" (1). He believes that this would not only increase proper diagnoses, but also increase the efficacy of prescribed treatment (1).

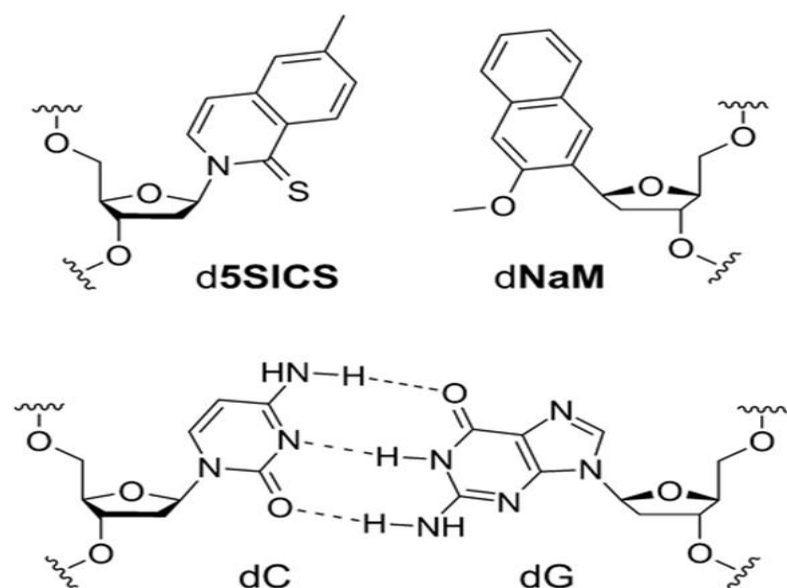
In 2006, a team at Stanford created a new functioning DNA system by incorporating additional benzene moieties into the four natural nucleotides, see Figure 4 (2). The modified nucleotides created by Kool et. al, deemed xDNA and yDNA, are fluorescent and can be useful in biological detections plus the additional benefit that an xDNA strand can complementarily bind to a natural strand and therefore provides increased specificity for certain coding regions (2).



**Figure 4. xDNA base pairing with natural canonical Watson Crick bases.** A single xDNA stand is composed of both xDNA and natural bases for a total of 8 bases with four different pairing arrangements. In contrast, natural DNA has four bases with only two different pairing arrangements (2).

Although the addition of these benzene rings elongated the nucleotides, a fully synthetic duplex had all of the characteristics of a Watson and Crick helix and was just as selective in binding as natural DNA systems (2). They conclude that this finding yields a new perspective: DNA replication might not have to follow the Watson and Crick model (2). As Ivan Amato puts it, "just why DNA and RNA use only four bases remains one of the grand mysteries of biology" (4). Voegel and others hypothesized that the reason why nature has selected its four nucleotides--A, G, C, and T—is because of certain "keto/enol tautomeric equilibrium constants for the natural bases [that] are not a function of medium polarity," however more research needs to be conducted to prove this point (5).

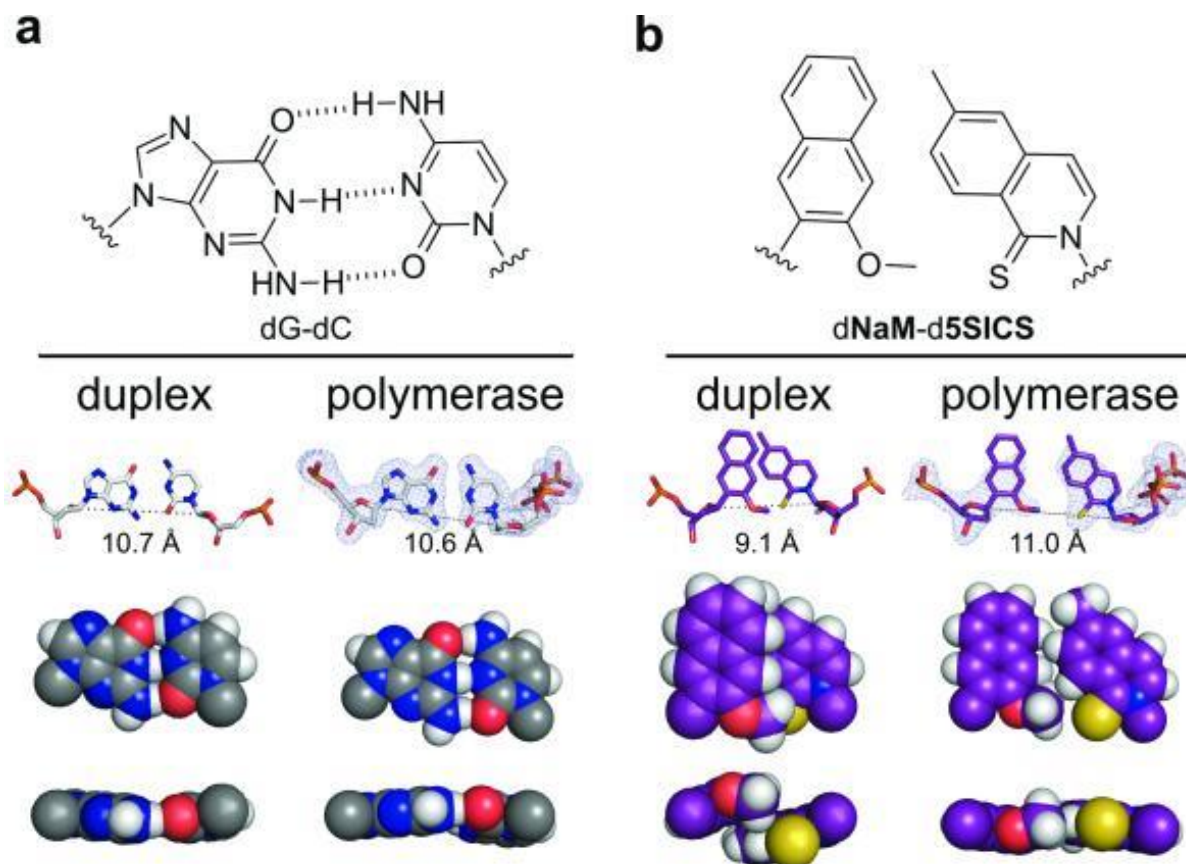
In 2009, Dr. Floyd Romesberg and his team designed and synthesized the novel d5SICS-dNaM unnatural base pair and determined the efficiency and fidelity of its replication compared to that of an A-T base pair (10). This UBP does not possess the traditional “edge-to-edge” Watson-Crick base pairing present in natural nucleotides, as shown in Figure 5 (7).



**Figure 5. The structure of a d5SICS-dNaM pairing.** A comparison of d5SICS-dNaM and pyrimidine-purine association via dC-dG. As can be seen, the d5SICS-dNaM has no hydrogen bonding between nucleosides (7).

Both 5SICS and NaM molecules are considered “hydrophobic” nucleobases meaning that they associate tightly together to exclude water. In 2014, Romesberg showed that the d5SICS-dNaM pair can be accurately and efficiently replicated through the endogenous machinery of *Escherichia coli* if unnatural triphosphates are provided (6). Moreover, the d5SICS-dNaM pair was not removed by natural DNA repair mechanisms and therefore this bacteria represents the first semi-synthetic organism “to propagate stably an expanded genetic alphabet” (6). Because these molecules have a different pairing interaction with a structure “very different from that of a normal base pair,” it is puzzling that d5SICS-dNaM are able to replicate through endogenous mechanisms *in vivo*, and through polymerase chain reaction (PCR) methods *in vitro*, with “fidelities that approach those of fully natural DNA” (7). These studies have potential to topple what is currently known about the polymerase mechanism.

To be more specific, Betz et al. (2012) found that d5SICS-dNaM pairing in a DNA duplex has an intercalated structure. The intercalative bonding between the d5SICS-dNaM pair is shown in Figure 6 with permission from Dr. Tammy Dwyer (7).

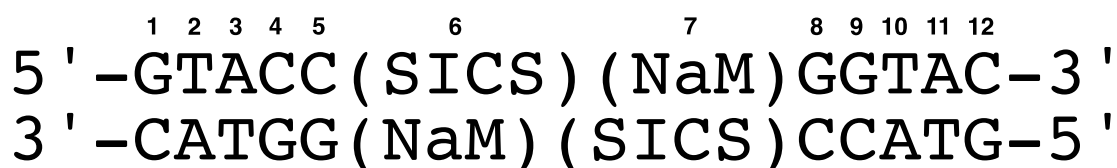


**Figure 6. KlenTaq polymerase induces the dNaM-d5SICS unnatural base pair to adopt a natural, Watson-Crick-like structure.** a) Natural dG-dC base pair. (b) Unnatural dNaM-d5SICS pair. The chemical structures of both base pairs are at the top of the panel. A comparison of the structure formed in duplex DNA versus the structure of the duplex inside the KlenTaq polymerase active site is at the bottom of the panel. Only nucleobases are shown with sugar and phosphates omitted for clarity.

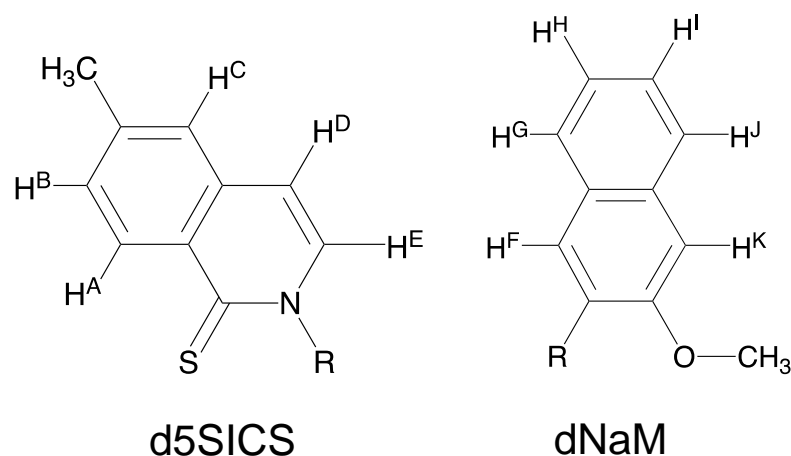
What is especially interesting about Figure 6 is that the structure of the d5SICS-dNaM pair is dynamic depending on its environment outside of a polymerase active site or inside the active site (Fig. 6b). When the UBP is outside of the KlenTaq polymerase active site it retains its intercalated duplex structure; however, while the d5SICS-dNaM pair is present in the active site of KlenTaq polymerase, the duplex mimics the edge-to-edge packing with an “internucleotide distance that is roughly the same as that of natural base pair” (Fig. 6b). Betz et al. believe that this finding indicates “complementary hydrogen bond formation is not required for the efficient and selective replication of DNA” (7). Because d5SICS-dNaM pairing can be replicated by polymerases other than KlenTaq, the structural findings of Betz and others “may be generalized to other polymerases” (7). Moreover, the findings suggest that the plasticity of the packing of hydrophobic unnatural base pairs may make them an intriguing line of study for an expanded genetic alphabet (7).

Pfaff et al. have published evidence suggesting that the structural differences of a single non-hydrogen bonding base pair in an otherwise natural DNA duplex are restored “beyond one base pair on either side” (8). This result highlights the idea that in a DNA duplex with non-natural base pairs, the structural distortions are confined to a local region immediately adjacent to the “damage” site (8). If one were to couple the findings of Pfaff and Betz, one may infer that disruptions in a DNA duplex get restored during the replication process by polymerase enzymes, which further supports the innate flexibility of natural polymerases to replicate unnatural perturbations in dsDNA.

The work discussed herein presents a snapshot of ongoing research involving two-dimensional (2D) NMR spectral analysis of a synthetic DNA duplex incorporated with two adjacent d5SICS-dNaM UBPs provided by Dr. Floyd Romesberg. The DNA sequence used for the NMR structural study is shown in Figure 7. A presentation of the d5SICS and dNaM aromatic proton identities is shown in Figure 8.



**Figure 7. DNA oligonucleotide sequence incorporated with adjacent d5SICS-dNaM base pairs.** The sequence is self-complementary<sup>2</sup> from 5' to 3'.



**Figure 8. The labeling scheme for aromatic protons of d5SICS and dNaM unnatural bases**

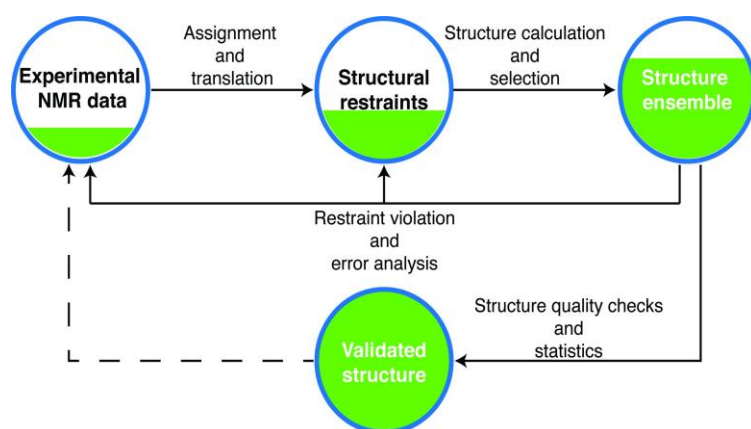
The goal of this work is to investigate the structure of a DNA duplex with two adjacent d5SICS-d5NaM pairs to determine precisely how the unnatural nucleobases are situated within the duplex, given the unusual dynamic orientation of the single pair (7). Additionally, future

<sup>2</sup> Hare states that a self-complementary duplex “halves the number of signals compared to an equivalent length of DNA without symmetry” (9). In other words, that “separate lines [on NOESY] are not seen for the two halves of the helix” (9).



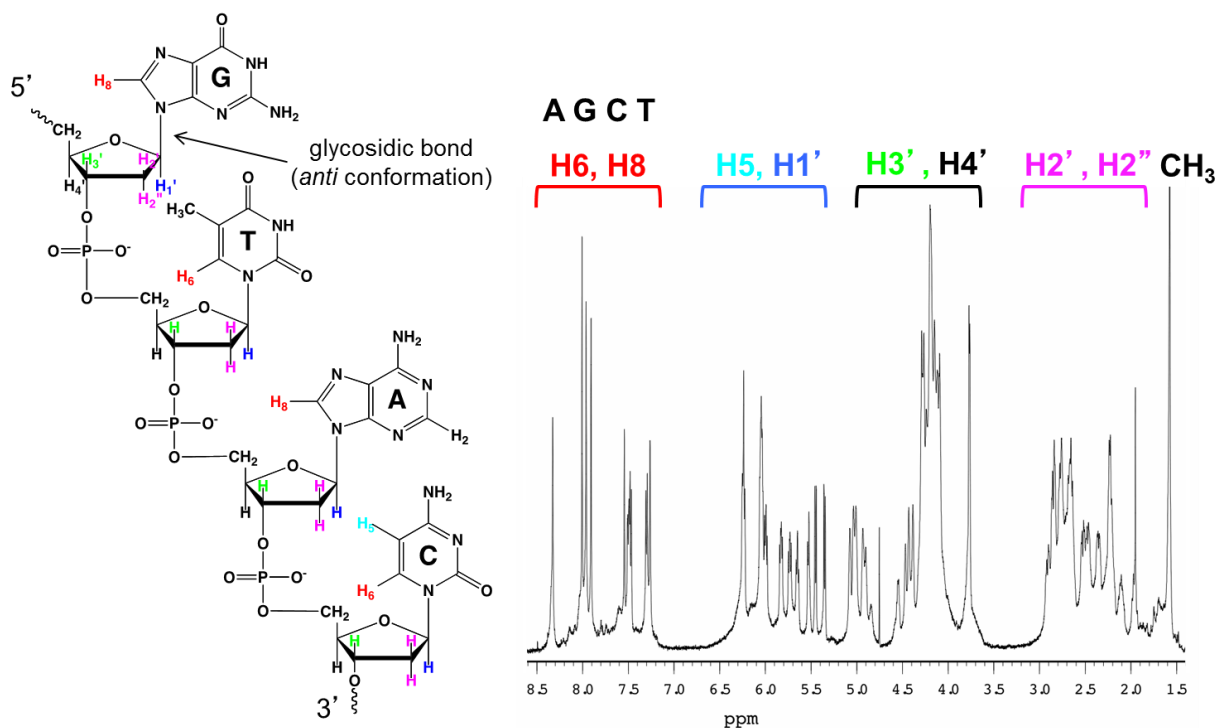
studies may be conducted to verify whether the structural disruptions of two adjacent unnatural base pairs can be detected and/or restored by natural polymerases, with the eventual end goal of investigating whether an entirely unnatural duplex may be efficiently and accurately replicated by endogenous DNA mechanisms.

This research involves the use of two-dimensional nuclear magnetic resonance (2D-NMR) technology to provide structural data of the synthetic DNA duplex of interest obtained from Dr. Romesberg. The process of NMR begins with simple resonance assignments for basic topology, followed by sequential resonance assignments and finally structure determination (12). Proton NMR is particularly useful for this research because nucleic acids have low proton densities and therefore can be quickly recognized and identified to uncover alignment of base pairs; however, finalized DNA duplex structures require assignments for sugar-phosphate backbones as well (12). An overview of structure determination by NMR is summarized in Figure 9.



**Figure 9. Schematic overview of NMR technique.** The inner cycle involving assignment, structure calculation and selection, as well as restraint violations and error analysis. The inner cycle is usually conducted several times before quality checks can be done to validate a biomolecule's structure (outer cycle).

The original protocol for assigning non-exchangeable proton resonances for a natural DNA duplex was published by Dennis Hare and others in 1983 (9). The publication presents assignment guidelines for carbon-bound protons of double-helical B-form nucleic acids. The hypothesis was that the natural ends of the duplex in Figure 7 would conserve B-form structure and hence Hare's methodology could be used. The natural nucleotide protons of interest in duplexed DNA are highlighted with their general 1D NMR chemical shifts (ppm) in Figure 10.



**Figure 10. Chemical shift assignments for small B-form DNA sequences.** The aromatic proton of interest for purines and pyrimidines is H8 and H6, respectively. The protons of interest for the deoxyribose sugar moieties are H1', H2' and H2''. In particular, the methyl of thymine and the H5 of cytosine are important for COSY spectra. The colored protons above the 1D NMR spectra connect identity to regions on 2D NMR spectra for both COSY and NOESY.

The two NMR techniques used in this research were 2D COReLATED Spectroscopy (COSY) and Nuclear Overhauser Enhanced Spectroscopy (NOESY) spectra in D<sub>2</sub>O (9). 2D NMR functions based on aligning chemical shifts from 1D spectra and dispersing the signals across two axes (or dimensions). Connections of two chemical shifts between axes are deemed “cross peaks” and represent resonance interactions between distinct protons.

Hare states that COSY spectra identify “through-bond connectivities”. COSY spectra can be used to identify scalar-coupled protons—usually 3 bonds apart, with an exception for 4-bond coupling from methyl groups on aromatic rings). For duplexed DNA, COSY spectra can group aromatic scalar-coupled protons on a single base, or protons on a one sugar, “without giving any information about the position of these residues in the sequence” (9). The only aromatic scalar-coupled protons are C H5 to C H6 which form strong cross peaks; comparatively, weaker cross peaks between T CH<sub>3</sub> to T H6 can also be seen on a COSY spectra (9). Additionally, in the sugar moieties the H1' is scalar-coupled to its own H2' and H2''. The corresponding cross peaks for these sugar proton interactions can be seen horizontally upfield from the scalar-couplings of the pyrimidines.

In contrast, NOESY spectra provide a reliable way to connect sequential nucleotides via “through-space” or through-distance connectivities within 5 Å (9). It should be noted that there

are many more non-scalar-coupled protons and therefore a NOESY spectra has much more cross peaks. A NOESY spectrum contributes to determination of DNA structure in three key ways (see Figure 10 for a visual of protons). First, the aromatic proton on one base is within 5 Å of the H1' of its own sugar *and* the aromatic proton of an adjacent base in the 5' to 3' direction. Second, the aromatic proton on one base is within 5 Å of the H2' *and* H2'' of its own sugar, where the H2'/H2'' can be further connected to the aromatic proton of the adjacent base. Third, the T CH<sub>3</sub> can be connected to the H1' of adjacent base sugar in 3' to 5' direction. Besides these three key characteristics, NOESY spectra can also verify the C H5 to C H6 and T CH<sub>3</sub> to T H6 assignments from a COSY spectra, as well as the H1' to H2'/H2'' of the sugar moieties. It is important to note that the H1' to H2'' cross peaks on NOESY spectra are stronger than the H2' because “any given H2'' is nearer to its H1' than is H2'” (9).

## METHODS

---

### *Sample Preparation*

The two self-complementary oligomers d(GTACC(SICS)(NaM)GGTAC) were synthesized in the laboratory of Dr. Floyd Romesberg (TSRI) and purified. Oligomer concentration was determined by UV absorbance at 240 nm and the extinction coefficient for 5SICS and NaM were estimated to be the same as guanine. The oligomers were mixed in a 1:1 stoichiometry, the resulting duplex was annealed at 85°C, and dialyzed against 2 L of 1.0 M NaCl, 2 L of 0.50 M NaCl, and 2 L of deionized water in a dialysis apparatus using a 1 kDa molecular weight cutoff.

The NMR sample used for structural analysis was prepared by dissolving the DNA duplex in 600 µL of 10 mM sodium phosphate buffer (pH 7.0) containing 100 mM NaCl and 0.1 mM EDTA, then lyophilizing and re-dissolving in 99.99% D<sub>2</sub>O (Cambridge Isotope Laboratories) to a 600 µL final volume. The resulting concentration of the DNA strand in the sample was 1 mM.

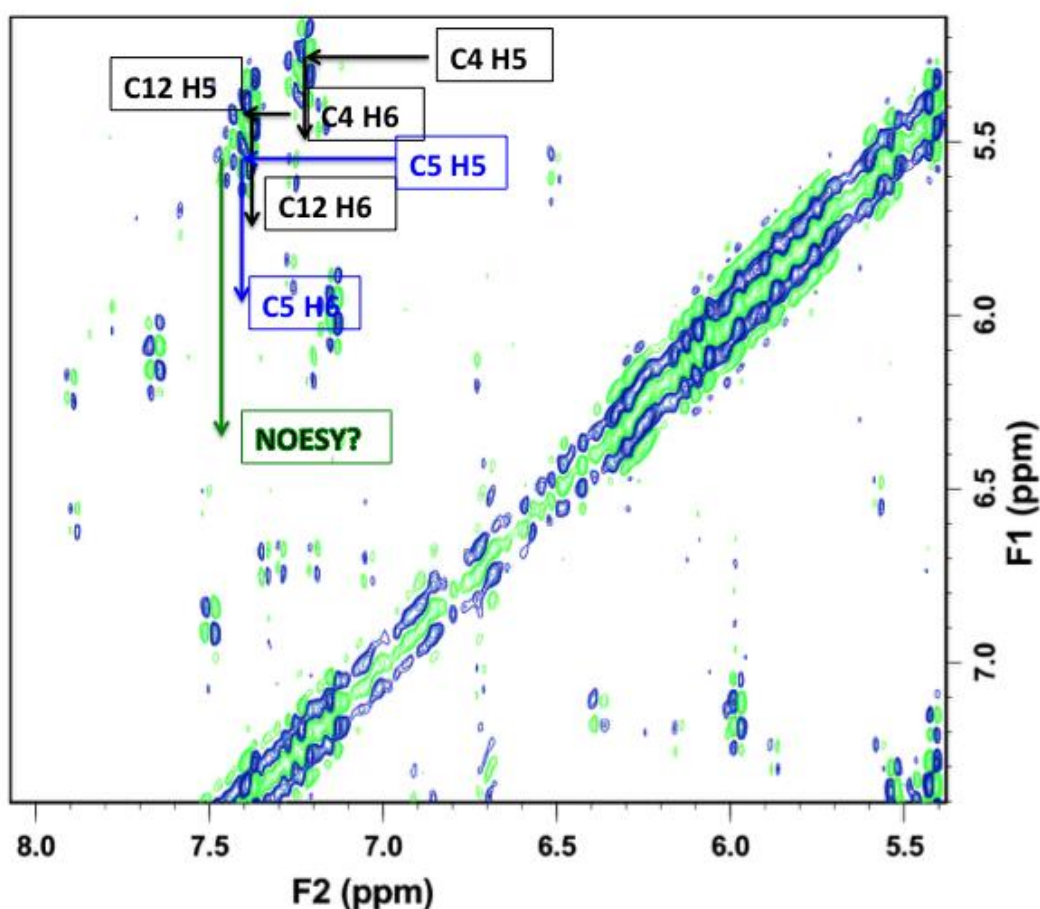
### *NMR Spectroscopy*

A Varian Inova 500 MHz spectrometer was used to obtain proton NMR spectra in D<sub>2</sub>O. NOESY and DQF-COSY spectra were acquired using the TPPI method of phase cycling, and data were collected at 20°, 25°C, 30°C, and 35°C to resolve cross-peak overlap in the spectra. NOESY spectra with a mixing time of 300 ms were collected for signal assignments, with a spectral width of 5913 Hz and 2048 complex points in t<sub>2</sub> and 512 in t<sub>1</sub> increments (zero filled to 2048 on processing. 64 scans were averaged for each t<sub>1</sub> value using a recycle delay of 2 s. Residual HOD resonance was suppressed by application of presaturation during the recycle delay and the mixing time. DQF-COSY spectra were collected with 2048 complex points in t<sub>2</sub> and 512 t<sub>1</sub> increments (zero filled to 2048) points on processing with a spectral width of 5913 Hz; 40-80 scans were collected for each t<sub>1</sub> value using a recycle delay of 2 s with presaturation of the HOD resonance. All spectra were transferred to a PC laptop and processed by Felix.

## RESULTS

The current results of this 2D NMR investigation of represent the first step of NMR study: the assignment of cross peaks to particular protons on the nucleobases and sugar moieties of the synthetic DNA duplex shown in Figure 7 following the approach and key spectral regions outlined by Hare et. al in 1983.

First and foremost it should be noted that the numbering assignments for nucleotide residues in the duplex are presented in both 2D COSY results, although numbered position assignments for the residues identified from 2D COSY were only confirmed later in 2D NOESY spectrum regions. The expanded COSY spectrum region of aromatic scalar-coupled of the H5 and H6 protons for the cytosine residues is shown in Figure 11. The chemical shift assignments for these cytosine residues are summarized in Table 1.



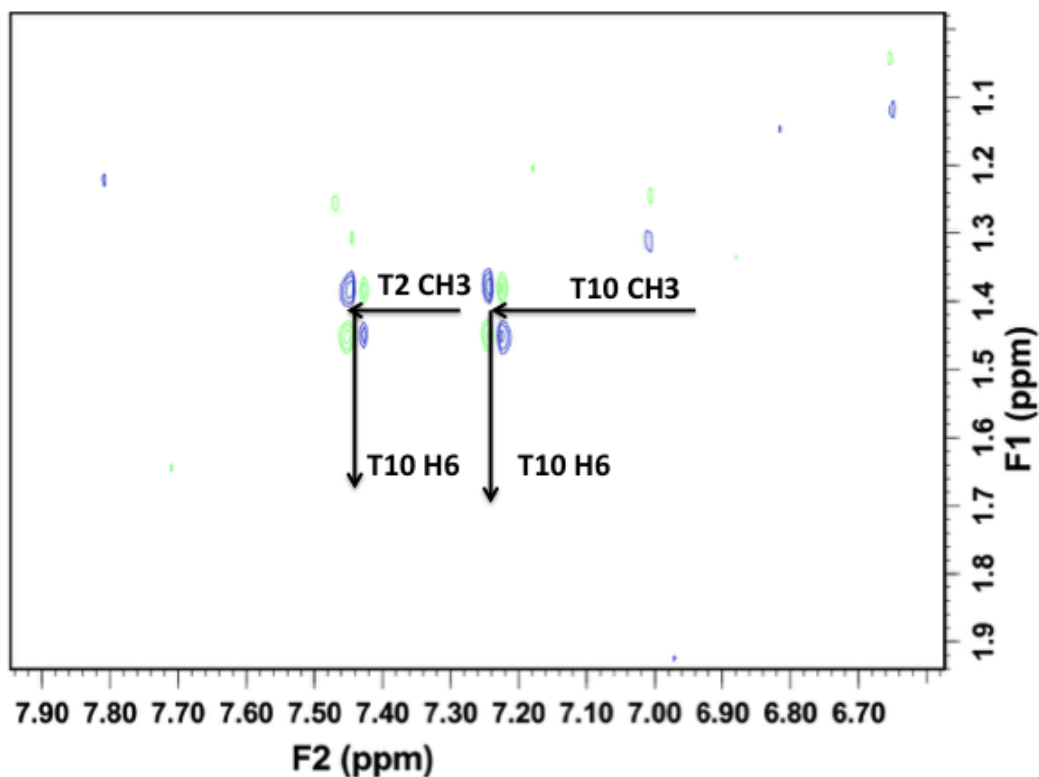
**Figure 11. Identification of H6 and H5 protons of cytosines from 2D COSY spectra of at 25°C.** *Black lines*, resonances on C4 and C12; *blue lines*, resonances on C5 with color change for clarity; *green line*, an unidentified cross peak near C5 which may be an important connection to either NaM or 5SICS in the NOESY fingerprint region. Chemical shift assignments are summarized in Table 1.

In Figure 11 there are three strong distinct aromatic cross peaks denoting three cytosine residues. The lack of aromatic scalar-coupled cross peaks for the other three cytosines—there is a total of six in the duplex—was expected due to the self-complementary nature of its sequence (see Fig. 7 for sequence). These results hint at the idea that the natural base ends of the synthetic duplex maintain 2-fold symmetry because the number of cross peaks for H6 and H5 of cytosine residues is halved (9).

**Table 1. H5 and H6 chemical shift assignments for cytosine residues from 2D COSY spectra at 25°C.**

Residue	H6	H5
C4?	7.219	5.281
C5?	7.387	5.546
C12?	7.375	5.431

The 2D COSY spectra results also serve to identify the thymine residues present in the duplex, specifically the four-bond coupling between methyl and H6 protons. The expanded COSY spectrum region for these interactions is shown in Figure 12.



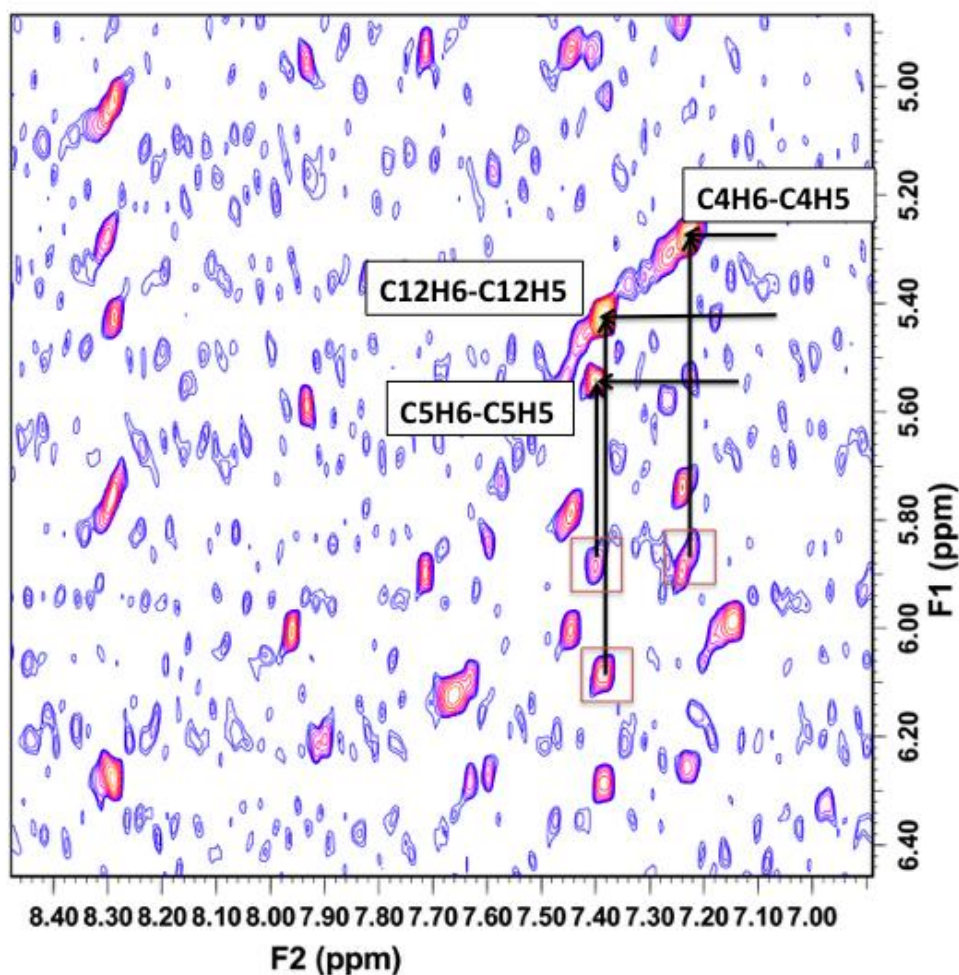
**Figure 12. Identification of H6 and methyl protons of thymine residues from 2D COSY spectra of at 25°C.** Chemical shift assignments are summarize in Table 2.

The two weak cross peaks connecting the aromatic region to the methyl region identifies only two thymine residues out of the total four in the duplex, similarly to the cytosine residues, providing further support that the natural ends of the duplex maintain 2-fold symmetry.

**Table 2. Methyl and H6 chemical shift assignments for thymine residues from 2D COSY spectra at 25°C.**

Residue	H6	H5
T2?	7.441	1.417
T10?	7.240	1.410

Following identification of the coupled proton interactions of cytosine and thymine residues, comparison of chemical shift assignments was made via 2D NOESY spectrum at 25°C. The aromatic region of the 2D NOESY spectrum for H6 and H5 protons of cytosine is shown in Figure 13.

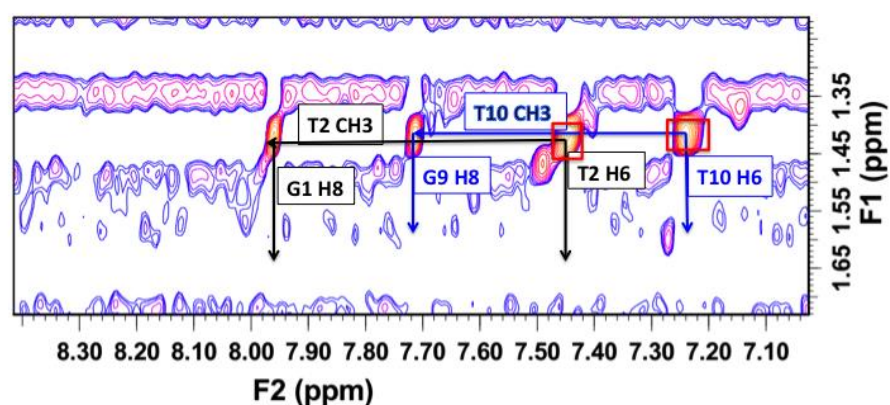


**Figure 13. 2D NOESY comparison of H6 and H5 protons of cytosine residues at 25°C.** *Black arrow intersection*, three cross peaks for H6 and H5 of cytosine corresponding to 2D

COSY results in Figure 11; *orange boxes*, denote possible sequential interactions between H6 and H1' protons between residues.

The 2D NOESY spectrum region in Figure 13 correlates the H6 and H5 protons of cytosine residues from the 2D COSY (Fig. 11), but more importantly this region is able to connect the cytosine H6 protons to either its own sugar H1' or to the H1' of the adjacent residue's sugar (9). Because there is evidence to suggest 2-fold symmetry of the duplex (Fig. 11, 12), for simplicity, position numberings will be conducted based on the 5' to 3' strand.

The top two boxed cross peaks can be horizontally connected at 5.885 ppm and this chemical shift is likely either the C4 H1' or C5 H1'. Following this numbering logic, the bottom orange boxed cross peak must be the C12 residue by elimination, with a horizontal chemical shift of 6.068 ppm. Similarly, the cross peaks and chemical shifts found in the thymine region of the 2D COSY spectrum (Fig. 12) were correlated to the same region on the 2D NOESY spectrum. This expanded 2D NOESY region is shown in Figure 14.

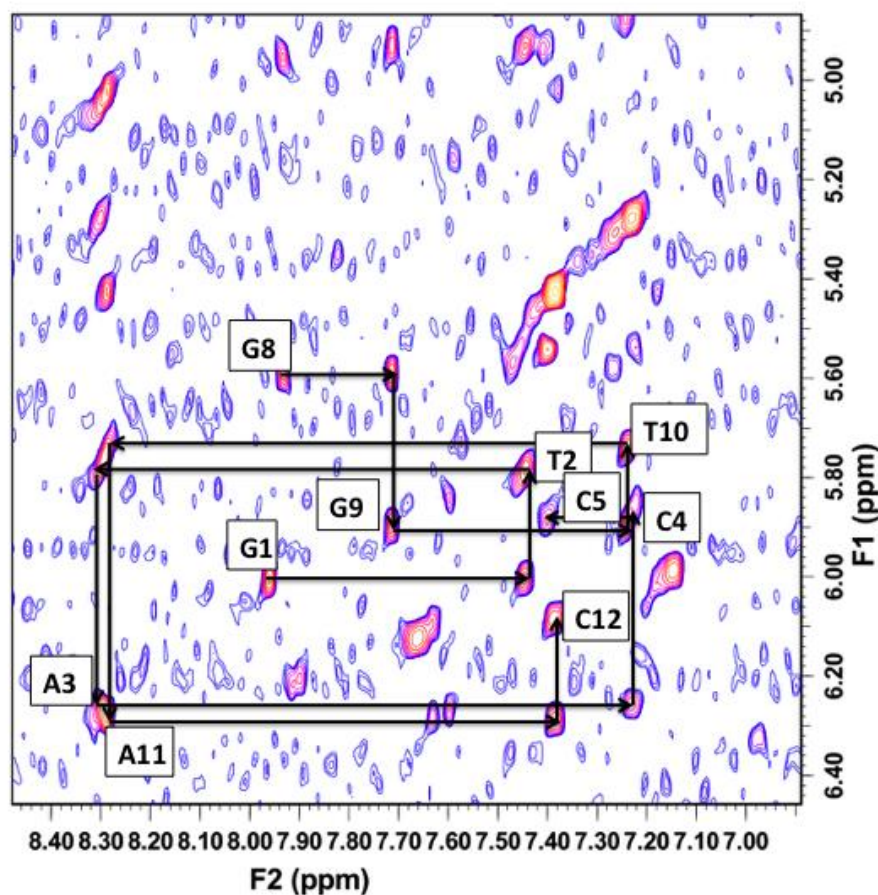


**Figure 14. 2D NOESY comparison of H6 and methyl protons of thymine residues at 25°C.** *Blue arrows*, interactions between thymine methyl and aromatic protons; *black arrows*, interactions between the other thymine methyl and aromatic protons; *red boxes*, the thymine methyl and H6 cross peaks obtained from 2D COSY spectrum in Figure 12 with chemical shift assignments made in Table 2.

In contrast to the 2D COSY spectrum region of the thymine methyl and H6 proton interactions (Fig. 12), the 2D NOESY spectrum of the same region presents more cross peaks (Fig. 14). This is because a single methyl resonance of thymine creates NOESY cross peaks to its own H6 proton (red boxes) *and* to another aromatic proton of the base on the 5' side. In the duplex, both T2 and T10 residues have adjacent G1 and G9 residues on the respective 5' sides, (Fig. 7). The *black arrow intersection* presents the interaction between a thymine methyl and aromatic H8 proton of a guanine residue vertically at 7.959 ppm; conversely, the *blue arrow intersection* presents another thymine methyl and aromatic H8 proton of another guanine at 7.715 ppm.

At this point, the H6 and H5 chemical shifts for cytosine residues and thymine H6 and methyl chemical shifts have been confirmed from Figures 11 and 12, respectively. By correlating the 2D COSY aromatic and methyl chemical shifts for cytosine and thymine residues with 2D NOESY spectrum, potential chemical shifts for cytosine H1' (Fig. 13) and guanine H8 protons (Fig. 14) has been suggested.

However, the exact numbered position of these residues requires sequential resonance assignments along all aromatic to H1' protons in either the 5' to 3' or 3' to 5' direction. This is because the aromatic to H1' connectivities “order the deoxyribose protons into their respective positions in the chain” (9). The sequential walk in the 5' to 3' direction from G1 to C12 has been determined from a 2D NOESY spectrum region and the results are presented in Figure 15 below.



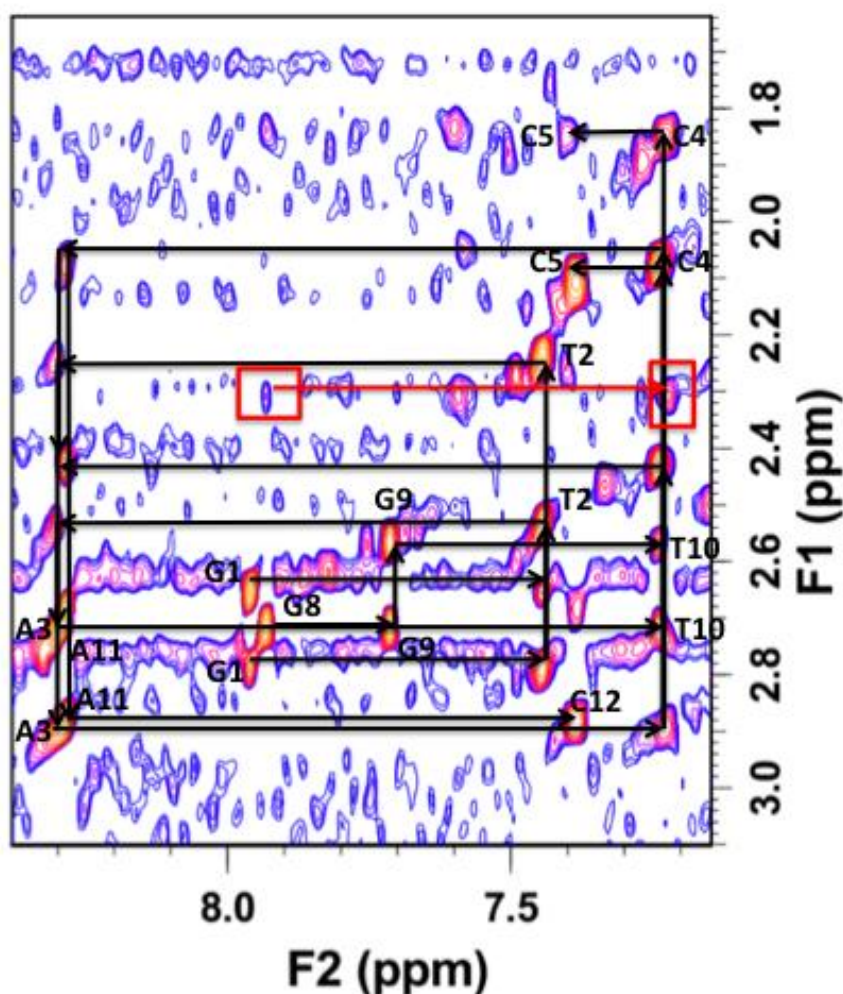
**Figure 15.** 2D NOESY spectra of H6/H8 to H1' at 25°C. Nucleoside aromatic protons to deoxyribose H1'. Chemical shift assignments are presented in Table 3.

The aromatic proton to H1' 2D NOESY spectrum in Figure 15 presents the sequential “walk” moving from G1 to G8 and G9 to C12 via cross peak alignment. There is a gap in assignments between the natural aromatic protons of C5 and G8 to 5SICS and NaM, respectively. Moreover, there is no cross peak present for the interaction between the C5 H6 and



its own H1'. The gap left by the unnatural 5SICS and NaM bases suggests a perturbed duplex structure, as the natural ends of the duplex cannot be tied into two unnatural d5SICS-dNaM base pairs.

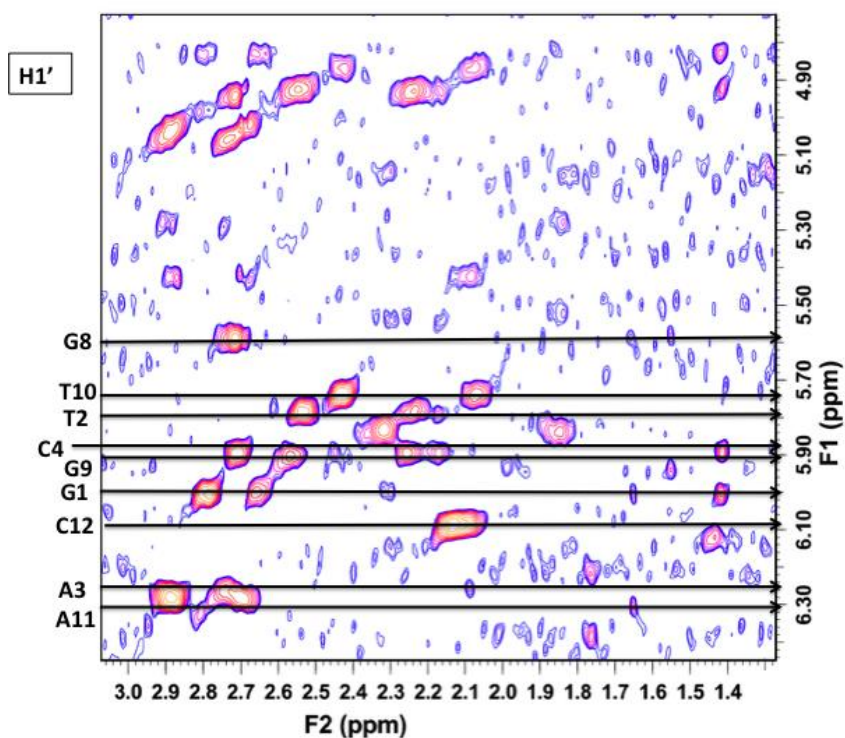
After the assignment of the aromatic to H1' protons of the natural nucleotides from Figure 15, further characterization of the natural nucleotide sugar moieties is presented in Figure 16. In this region, the aromatic proton of one nucleobases can be connected to the H2'/H2'' protons of its own sugar, where these H2'/H2'' protons can then be connected to the aromatic proton of an adjacent nucleobase.



**Figure 16.** 2D NOESY spectra of aromatic to H2'/H2'' proton resonances at 25°C. The *black arrows* sequentially connect the cross peaks between an aromatic proton to its own sugar and the sugar adjacent to it, similar to the aromatic to H1' “walk” in Figure 15. *Red box and arrows*, present unidentified cross peak interactions that may connect G8 to NaM. Chemical shift assignments are shown in Table 3.

In Figure 16, the residues directly adjacent to the unnatural NaM and 5SICS bases—C5 and G8—as well as the terminal C12, did not have all aromatic to H2'/H2'' cross peaks. First, the C5 H1' to H2' and H2'' cross peaks were both missing; second, the G8 H1' to H2'' cross peak was missing; and third, the C12 H' to H2'' cross peak was missing. On another note, the *red arrow* in Figure 16 denotes a possible connection between G8 H8 (7.935 ppm) to an unidentified peak with vertical and horizontal coordinates of 7.221 and 2.313 ppm, respectively.

The deoxyribose H2'/H2'' chemical shift assignments can be confirmed in a separate 2D NOESY region via H1' to H2'/H2'' connectivities, however these assignments have not been completed. It should be noted that “any given H2'' is nearer to its H1' than is H2', and hence gives a much stronger NOESY cross peak,” which allows H2' to be readily identified from H2'' (9). The H1' and H2'/H2'' 2D NOESY region is shown in Figure 17, however only the H1' assignments have been shown.



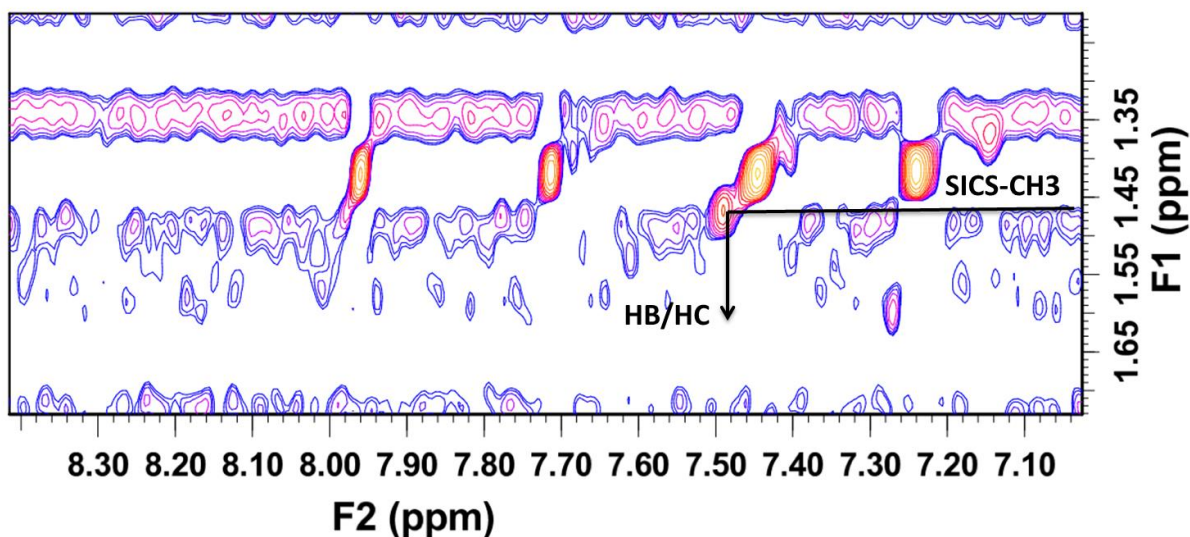
**Figure 17.** 2D NOESY spectra of H1' to H2'/H2'' at 25°C. Note: the assignment of the deoxyribose H2'/H2'' assignments has not been completed.

At this point, most natural nucleotide chemical shift assignments have been made, albeit without the confirmation of the H2'/H2'' assignments from the H1' to H2'/H2'' region (Fig. 17). A summarized chemical shift assignment from both 2D NOESY and COSY spectra is presented in Table 3.

**Table 3. Chemical shift assignments for G1 to C5 and G8 to C12 derived from 2D COSY and 2D NOESY spectra at 25°C.**

Residue	<i>Nucleobase Moiety</i>				<i>Deoxyribose Moiety</i>		
	H8	H6	H5	CH <sub>3</sub>	H1'	H2'	H2''
G1	7.959			1.418	6.007	2.656	2.788
T2		7.441		1.418	5.787	2.25	2.533
A3	8.302				6.257	2.719	2.896
C4		7.231	5.278		5.885	2.083	1.853
C5		7.402	5.543		N/A	N/A	N/A
G8	7.93				5.591	2.719	N/A
G9	7.715			1.418	5.895	2.719	2.562
T10		7.24		1.418	5.743	2.431	2.078
A11	8.288				6.286	2.881	2.69
C12		7.387	5.5425		6.086	2.132	N/A

After conducting chemical shift assignments for almost all natural nucleotides, the goal was to now try and locate the 5SICS and NaM bases. First, the methyl region of the 2D NOESY spectrum was analyzed to look for the methyl of d5SICS (see Fig. 8 for labeled 5SICS and NaM protons) and the spectra is presented in Figure 18.

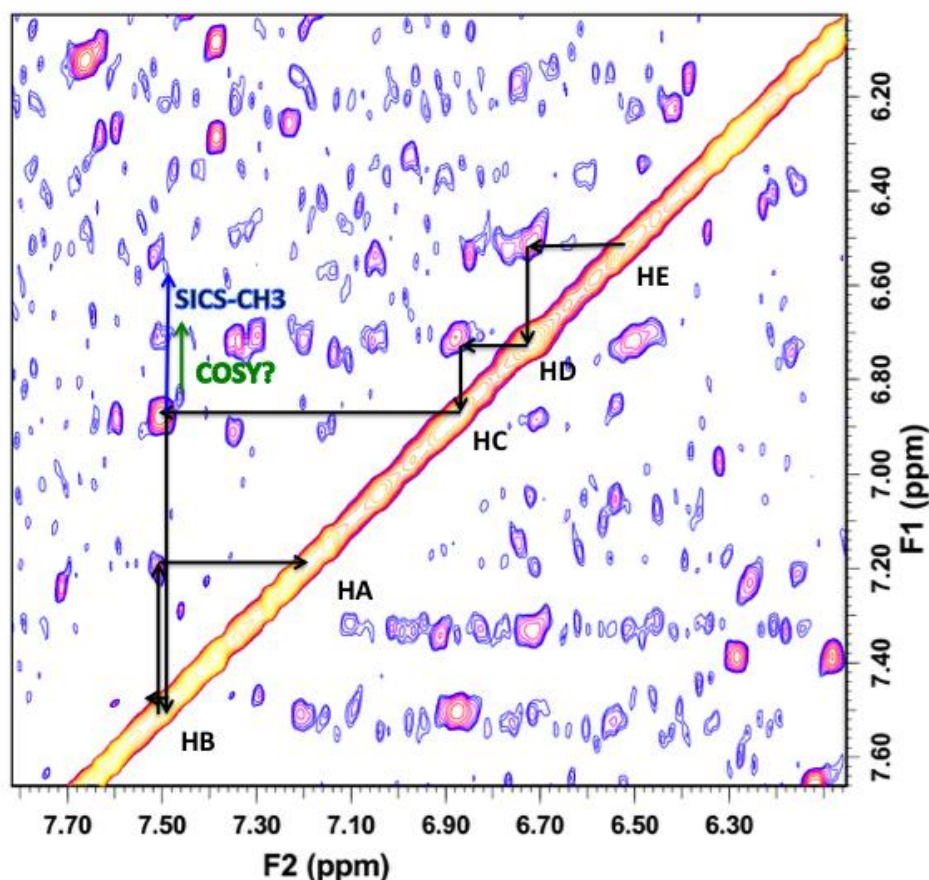


**Figure 18. 2D NOESY spectra of 5SICS CH<sub>3</sub> to its aromatic protons at 25°C.** Horizontal line, a 5SICS CH<sub>3</sub>. Vertical arrow, a 5SICS CH<sub>3</sub> to its aromatic protons, likely the cross peak of HB and HC shown in Figure 19. Chemical shift assignments are presented in Table 4.

Like thymine, 5SICS has four-bond coupling between its methyl and both H<sub>k</sub> and H<sub>B</sub> protons, however the 2D COSY spectrum does not present a cross peak in that region (Fig. 12). The NOESY spectrum in Figure 18 shows a distinct cross peak shifted downfield vertically from

the thymine residues at 1.467 ppm. This cross peak is a 5SICS CH<sub>3</sub> residue seeing its own H<sub>B</sub>/H<sub>C</sub> cross peak vertically at 7.489 ppm. The lack of a second 5SICS CH<sub>3</sub> cross peak could mean that one of the 5SICS residues is swung out of the duplex, falling outside of the 5 Å through-bond connectivity required for 2D NOESY cross peaks (9).

These observations resulted in the creation of a potential 5SICS residue aromatic “walk,” which is shown below in Figure 19 (a bit below the natural nucleotide aromatic to H1' region).



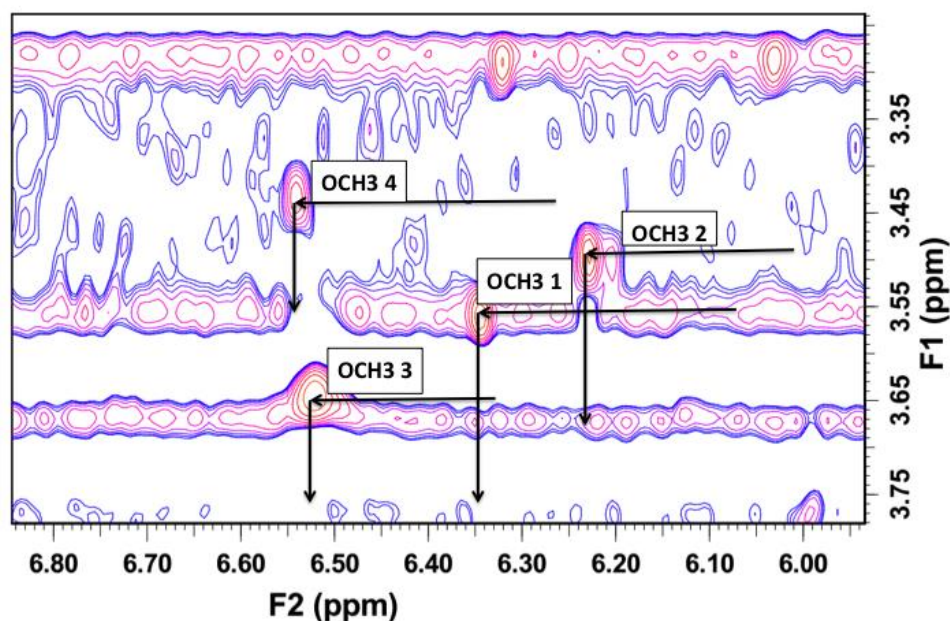
**Figure 19. 2D NOESY spectra of 5SICS Walk at 25°C.** Intranucleoside aromatic protons of the NOESY spectrum. The aromatic protons reside on the diagonal and sequential cross peak alignment from HA to HE denotes the ‘walk’ for 5SICS. Presence of a methyl group between HB and HC further supports unique 5SICS identity. *Green arrow*, denotes an small cross peak adjacent to the HB/HC that may connect to NaM. Chemical shift assignments are presented in Table 4.

The “walk” shown in Figure 19 presents the sequential aromatic proton connectivities of one 5SICS residue. The individual chemical assignments for the aromatic protons on this 5SICS residue are summarized in Table 4 below.

**Table 4. Potential chemical shift assignments for aromatic protons of a 5SICS residue from 2D NOESY spectra at 25°C.**

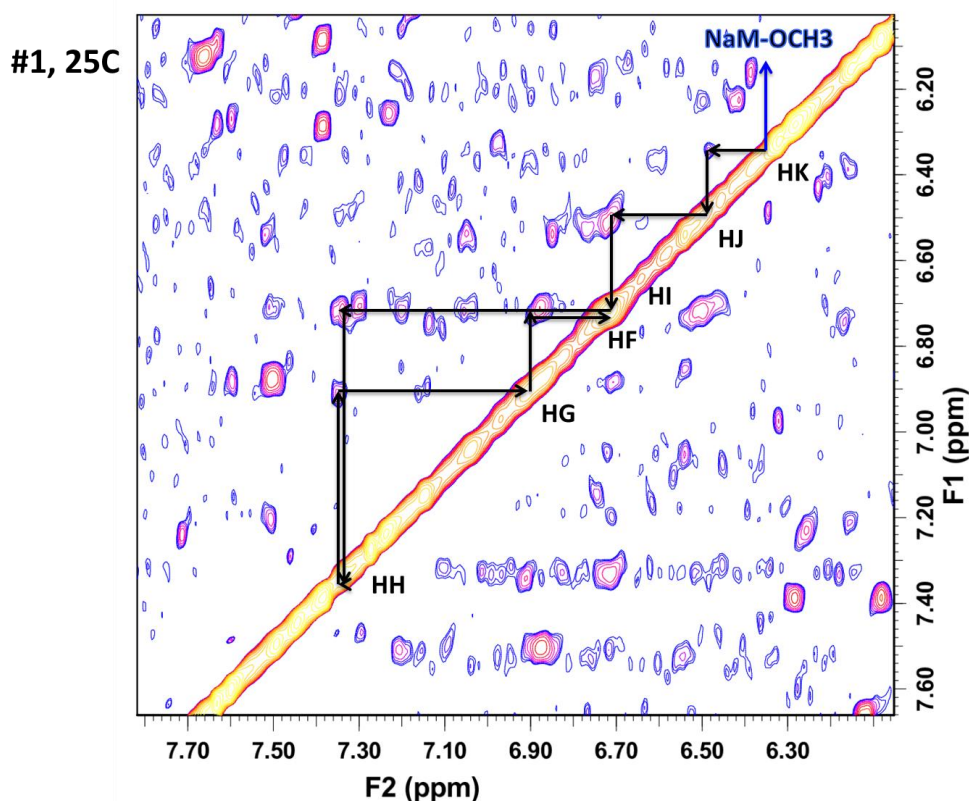
Residue	H <sub>A</sub>	H <sub>B</sub>	H <sub>C</sub>	H <sub>D</sub>	H <sub>E</sub>	CH <sub>3</sub>
5SICS	7.211	7.504	6.874	6.712	6.502	1.467

After the identification of one of the 5SICS residues in the duplex—either 5SICS6 or 5SICS18—identification of the NaM residues was explored. NaM bases have an OCH<sub>3</sub> that can be connected to its aromatic H<sub>K</sub> proton based on unpublished data from a previous research team with Dr. Tammy Dwyer as the PI. The cross peaks for the OCH<sub>3</sub> moieties of NaM residues is present near ~3.5 ppm chemical shift of the 2D NOESY spectrum. The results of OCH<sub>3</sub> assignments are shown in Figure 20.



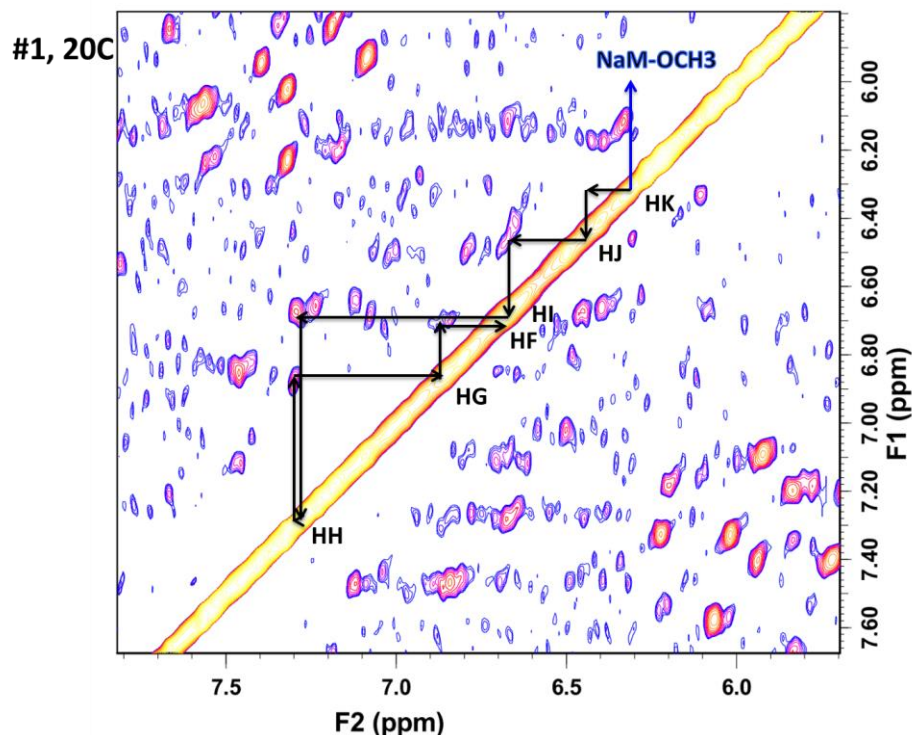
**Figure 20. 2D NOESY spectra of NaM OCH<sub>3</sub> to its aromatic protons at 25°C.** The four potential methoxy chemical shift assignments are the following: OCH<sub>3</sub> 1 to H<sub>K1</sub> (6.34 vs 3.56); OCH<sub>3</sub> 2 to H<sub>K2</sub> (6.22 vs 3.51); OCH<sub>3</sub> 3 to H<sub>K3</sub> (6.52 vs 3.65); OCH<sub>3</sub> 4 to H<sub>K4</sub> (6.54 vs 3.44). Summarized chemical shift assignments are presented in Table 5.

The 5SICS walk results presented in Figures 18 and 19, show only one potential 5SICS CH<sub>3</sub>. In contrast, the OCH<sub>3</sub> NOESY spectrum has four potential OCH<sub>3</sub> moieties, albeit that there are only two NaM residues: NaM7 and NaM19. The results from Figure 20 suggest that these two residues have two different conformations, indicating that there may be two distinct conformations of the overall synthetic duplex. The best potential aromatic NaM walk was made from the OCH<sub>3</sub> 1 and H<sub>K1</sub> cross peak from Figure 20 with chemical shifts of 6.34 and 3.56 ppm, respectively. The corresponding aromatic proton walk generated from this OCH<sub>3</sub> and H<sub>K</sub> assignment for this NaM residue is shown in Figure 21.



**Figure 21. 2D NOESY spectra of NaM Walk at 25°C.** OCH<sub>3</sub> (6.34 vs 3.56). Intranucleoside aromatic protons of the NOESY spectrum. The aromatic protons reside on the diagonal and sequential cross peak alignment from H<sub>K</sub> to H<sub>F</sub> denotes the “walk” for a NaM residue. Presence of a methoxy group near H<sub>K</sub> and H<sub>J</sub> supports unique NaM identity. Summarized chemical shift assignments are presented in Table 5.

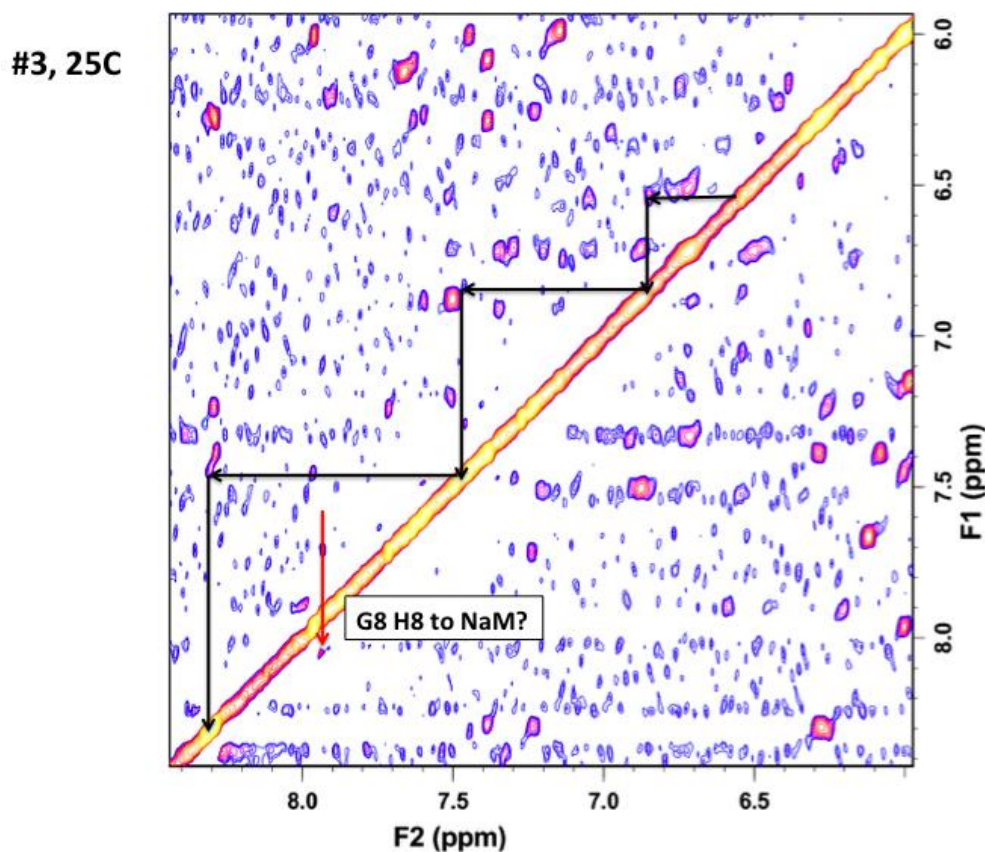
This aromatic NaM walk (Fig. 21) completes all aromatic protons from H<sub>K</sub> to H<sub>F</sub>. However, these assignments have not been correlated with a 2D COSY spectra which would be valuable in identifying the aromatic scalar-coupled H<sub>I</sub>/H<sub>I</sub>, H<sub>I</sub>/H<sub>H</sub>, and H<sub>H</sub>/H<sub>G</sub> (see Figure 8 for aromatic proton labeling). Unfortunately, full aromatic NaM walks could not be completed for the remaining three OCH<sub>3</sub> and H<sub>K</sub> cross peaks shown in Figure 20. However, the aromatic NaM walk presented in Figure 21 was confirmed from a separate 2D NOESY spectrum generated at 20°C. The corresponding NaM walk at 20°C is shown in Figure 22.



**Figure 22.** 2D NOESY spectra of NaM Walk at 20°C. OCH<sub>3</sub> (6.308 vs. 3.521). Intranucleoside aromatic protons of the NOESY spectrum. The aromatic protons reside on the diagonal and sequential cross peak alignment from H<sub>K</sub> to H<sub>F</sub> denotes the “walk” for a NaM residue. Presence of a methoxy group near H<sub>K</sub> and H<sub>J</sub> supports unique NaM identity. Summarized chemical shift assignments are presented in Table 5.

The 2D NOESY spectrum region for the NaM walk at 20C (Fig. 22) supports the previous walk created at 25°C (Fig. 21). That being said, there were four OCH<sub>3</sub> and H<sub>K</sub> cross peaks at 20°C (not presented here) which similarly could not generate full NaM walks.

Before going into a summarized table of the potential NaM walks generated from 2D NOESY spectra at both 25°C and 20°C, there is another potential walk of interest for a NaM residue shown in Figure 23.



**Figure 23.** 2D NOESY spectra of NaM Walk at 25°C. OCH<sub>3</sub> (6.52 vs 3.65). Intranucleoside aromatic protons of the NOESY spectrum. The aromatic protons reside on the diagonal and sequential cross peak alignment from H<sub>K</sub> to H<sub>F</sub> denotes the “walk” for NaM. Presence of a methoxy group near H<sub>K</sub> and H<sub>J</sub> supports unique d5NaM identity. *Red arrow*, G8 H8 to NaM @ peak below diagonal (7.933 vs 8.052 ppm).

The NaM walk shown in Figure 23 is very distinct from the NaM walks shown in Figures 21 and 22, however this walk may be a potential candidate for another NaM residue. While the sequential cross peaks from H<sub>K</sub> to H<sub>F</sub> have not been completed in Figure 23, the results suggest that another NaM walk could be shifted much farther downfield into the 8 – 9 ppm region on the horizontal axis. Another interesting finding is that the G8 H8 resonance has a cross peak interaction below the diagonal that has not been connected so far to any NaM walks (see red arrow in Fig. 23). This may be an interesting starting point for future researchers looking to finish this NMR structural study. Finally, the potential aromatic, CH<sub>3</sub> and OCH<sub>3</sub> proton assignments for NaM residues is summarized in Table 5.



**Table 5. Potential chemical shift assignments for aromatic protons on NaM from 2D NOESY spectra at 25°C and 20°C.** The numbering after residue names are used to separate the four different potential walks and are not related to residue position in sequence.

Temperature	Residue	OCH <sub>3</sub>	H <sub>K</sub>	H <sub>J</sub>	H <sub>I</sub>	H <sub>H</sub>	H <sub>G</sub>	H <sub>F</sub>
25°C	NaM #1	3.56	6.34	6.482	6.712	7.338	6.893	6.721
	NaM #2	6.22	3.51	-	-	-	-	-
	NaM #3	6.52	3.65	6.849	7.461	8.302	-	-
	NaM #4	6.54	3.44	-	-	-	-	-
20°C	NaM #1	3.521	6.308	6.458	6.678	7.288	6.867	6.7
	NaM #2	3.446	6.187	-	-	-	-	-
	NaM #3	3.59	6.521	-	-	-	-	-
	NaM #4	3.371	6.498	-	-	-	-	-

The results from Table 5 show complete potential chemical shift assignments for a d5NaM at both 25°C and 20°C.

## DISCUSSION

From NMR spectra, we can tell that the sequences preceding and following the 5SICS-NaM pairs are B-form and self-complementary. This means that although there are disturbances in the unnatural region of the sequence, these disturbances are localized and do not affect the natural ends of the duplex. However, future work involves distinguishing the chemical shift assignments of the undetermined 5SICS and NaM residues to provide insight on the overall structure. Identification of the unnatural 5SICS and NaM H1' protons would be helpful in determining the orientation of these bases, as these H1' assignments could be used to connect into their H2'/H2'' protons.

A number of possible conformations of the two unnatural base pairs could be occurring within the duplex making this particular region unclear (as shown in Fig. 20). Moreover, if the H2'/H2'' assignments for C5 and G8 deoxyribose moieties adjacent to 5SICS6 and NaM7 could be confirmed by further analysis, this may unveil new insights to the unnatural region of the duplex. On a different note, the lack of these H2'/H2'' assignments may suggest that some of these unnatural residues are swung outside of the duplex. This would support the missing 5SICS CH<sub>3</sub> group in Figure 18; however, the presence of four distinct NaM OCH<sub>3</sub> resonances in Figure 20 may suggest otherwise.

I hope that future researchers are able to carry out this 2D NMR structural investigation of this duplex based on the foundation that has been laid with this paper. It may even be useful to separate the unnatural 5SICS and NaM base pairs by a few natural residues to see if the natural nucleotides between the unnatural bases possess B-form or have distorted structures.

**REFERENCES**

1. Gramling, C. A Formula for life: For Chemistry Professor Steven Benner, Life As We Know it May Not Be The Only Alternative. Explore: Research [Online], 2005. <http://www.research.ufl.edu/publications/explore/v10n1/story1.html> (accessed Feb 17, 2017).
2. Krueger, A. T.; Lu, H.; Lee, A. H. F.; Kool, E. T. Synthesis and properties of size-expanded DNAs: Toward designed, functional genetic systems. *Accounts of Chemical Research*. DOI: 10.1021/ar068200o. Published Online: Feb 2007, 40 (2), 141–150.
3. Sismour, A. M.; Benner, S. A. Synthetic biology. *Expert Opinion on Biological Therapy*. DOI: 10.1517/14712598.5.11.1409. Published Online: Nov 2005, 5 (11), 1409
4. Amato, I. Expanding the genetic alphabet. *Science News*. DOI: 10.2307/3974601. Published Online: Feb 10, 1990, 137 (6), 88.
5. Voegel, J. J.; von Krosigk, U.; Benner, S. A. Synthesis and tautomeric equilibrium of 6-amino-5-benzyl-3-methylpyrazin-2-one. An acceptor-donor-donor nucleoside base analog. *The Journal of Organic Chemistry*. DOI: 10.1021/jo00078a038. Published Online: Dec 1993, 58 (26), 7542–7547.
6. Malyshev, D. A.; Dhami, K.; Lavergne, T.; Chen, T.; Dai, N.; Foster, J. M.; Corrêa, I. R.; Romesberg, F. E. A semi-synthetic organism with an expanded genetic alphabet. *Nature*. DOI: 10.1038/nature13314. Published Online: May 7, 2014, 509 (7500), 385–388.
7. Betz, K.; Malyshev, D. A.; Lavergne, T.; Welte, W.; Diederichs, K.; Dwyer, T. J.; Ordoukhanian, P.; Romesberg, F. E.; Marx, A. KlenTaq polymerase replicates unnatural base pairs by inducing a Watson-Crick geometry. *Nature Chemical Biology*. DOI: 10.1038/nchembio.966. Published Online: June 3, 2012, 8 (7), 612–614.
8. Pfaff, D. A.; Clarke, K. M.; Parr, T. A.; Cole, J. M.; Geierstanger, B. H.; Tahmassebi, D. C.; Dwyer, T. J. Solution structure of a DNA duplex containing a Guanine–Difluorotoluene pair: A wobble pair without hydrogen bonding? *Journal of the American Chemical Society*. DOI: 10.1021/ja7103608. Published Online: April 2008, 130 (14), 4869–4878.
9. Hare, D. R.; Wemmer, D. E.; Chou, S.-H.; Drobny, G.; Reid, B. R. Assignment of the non-exchangeable proton resonances of d(C-G-C-G-A-A-T-T-C-G-C-G) using two-dimensional nuclear magnetic resonance methods. *Journal of Molecular Biology*. DOI: 10.1016/0022-2836(83)90096-7. Published Online: Dec 1983, 171 (3), 319–336.
10. Seo, Y.; Matsuda, S.; Romesberg, F. *Journal of the American Chemical Society* 2009, 131, 5046-5047.
11. B-Form, A-Form, Z-Form of DNA  
[https://bio.libretexts.org/TextMaps/Map%3A\\_Working\\_with\\_Molecular\\_Genetics\\_\(Hardison\)/Unit\\_I%3A\\_Genes%2C\\_Nucleic\\_Acids%2C\\_Genomes\\_and\\_Chromosomes/2%3A\\_Structures\\_of\\_nucleic\\_acids/B-Form%2C\\_A-Form%2C\\_Z-Form\\_of\\_DNA](https://bio.libretexts.org/TextMaps/Map%3A_Working_with_Molecular_Genetics_(Hardison)/Unit_I%3A_Genes%2C_Nucleic_Acids%2C_Genomes_and_Chromosomes/2%3A_Structures_of_nucleic_acids/B-Form%2C_A-Form%2C_Z-Form_of_DNA) (accessed May 5, 2017).
12. Bertini, I.; McGreevy, K.; Parigi, G. *NMR of biomolecules*; 1st ed.; Wiley-VCH: Weinheim, 2012.