Structure of a DNA thymine-thymine Tandem Mismatch Base Pair in Oligonucleotide Duplex

By Ashley Gadzo and Mentor Dr. Fountain
Department of Biology and Chemistry At SUNY Fredonia

INTRODUCTION

In this experiment Nuclear Magnetic Resonance spectroscopy (NMR) was used on a DNA thymine-thymine tandem mismatch base pair (TT1) and on the nitrogenous base uridine. NMR spectroscopy is a methodical chemistry technique that is used in quality control, determining the content and purity of a analyte of interest as well as its molecular structure (1). The NMR is can quantitatively examine analyte samples that contain a known and unknown. For unknown the NMR can be utilized in finding a match through the spectral libraries or infer the basic structure directly (1). Once the analyte of the sample is found, NMR can be used to determine the molecular conformation in the solution as well as the ability to study the physical properties at the molecular level. These properties consist of conformational exchange, phase changes, solubility and diffusion. The principle of using NMR is to look at the many nuclei in the analyte sample of interest. The nuclei in the analyte will have a spin and all of the nuclei are electrically charged. When a magnetic field is applied energy can be transferred between the base energy to the higher energy level (1). This transfer is usually a single energy gap. This transfer is able to happen because when it is at a wavelength that matches to the radio frequencies as the spin returns to its base level. Once it is at the base level the energy is produced at the same frequency. The signals that are corresponding to the transfer can be measured in many ways and are processed in order to bestow an NMR spectrum for the nucleus of interest. When the NMR spectra is produced there will be peaks, known as chemical shifts relates to how the atoms are oriented in the analyte to provide detail on its structure. The magnetic field can be affected by the orientation of neighboring nuclei, which is known as spin-spin coupling. The spin-spin coupling causes splitting of the signal for each type of nucleus into two or more lines (1). The size of the splitting can be measured as an absolute frequency and is independent of the magnetic field (1). The number of splitting represents the chemically bonded nuclei in close range of the observed nucleus.

When spin-spin coupling happens between two protons correlation will appear (2). If when coupling doesn't occur no correlation is expected to appear. Correlated spectroscopy (COSY) is a useful method for determining which signals arise from neighboring protons, which is usually up to four bonds (2). This method is very helpful when multiplets overlap or when there is an extensive second order coupling complicates the 1D spectrum (2). They are many various on the COSY pulse sequence. The one that was used in this experiment was the gradient enhanced double quantum coherence (DQF-COSY). The DQF-COSY has a ratio of gradient strengths is set to two yield all COSY signals but may be set to three to yield only those correlations involving three protons (2). The COSY spectrum contains diagonal and cross peaks. Also signals that are not found on the diagonal and correspond to the other signals on the same horizontal and vertical projections. The cross peaks that can be seen in the COSY spectrum indicate couplings between two multiplets up to three and can be four bonds away (2). The diagonal consists of the 1D spectrum with single peaks suppressed. The most plausible cross peaks in the spectrum is between H1 and H2 at 2.65 and 1.24 ppm (2). A cross peak that will be weaker in range will be a four bound correlation with H1 and H2 at 2.65 and 7.20 ppm (2). All the signals are antiphase, thus half of the multiplets will be positive and the other half negative. But not only can COSY produce a 1D spectrum it can create a 2D spectrum.

The COSY of a 2D spectrum is considered to be a homonuclear molecule. Homonuclear molecule are composed of only one type of element (3). Although the molecule may consist of diverse number of atoms and depending on the elements properties. However, in the 2D NMR the acquisition stage is disturbed from the excitation stage by an intermediate stages called evolution and mixing (3). In this experiment the process of evolution continued for a 24 hour period. Data acquisition covers a large number of spectra that are obtained as followed: the first time the value of time is set to zero and the first spectra is produced. Then the time is increase over the time and thus produces another spectrum. Therefore, this process is imitated until there is enough data to be used in forming a 2D Fourier transform (FID) (3). The spectrum is displayed as a topographic map. The topographic map is constructed by the frequency found in the spectrum in the time dimension on one axis and the aftermath of the evolution of the mixing stages on the other axis (3). The 2D NMR is a very helpful technique when the 1D NMR is insufficient when the signals overlap because of the resonant frequencies are similar (3). The 2D spectrum is able to save time when interested in connectivity between different types of nuclei (3). In the 2D spectrum that were produced in this experiment shows how COSY can occur though bonds to the same type of nucleus. But there is another 2 D spectra that can be produced through space which is known as NOESY which was also done in this experiment.

In COSY the magnetization is transferred by the scalar coupling (4). The protons are more three bonds away, therefore only signals of the protons that are two or three bonds are able to be visible on the COSY. There are phi torsion angles of the protein backbone that can be derived from the dipolar coupling constant between them (4). The NOESY is a crucial for the determination of the structure because it uses the dipolar coupling for the correlation of the protons (4). It is important that the protons are less than 5 Å away for each other or it will not work.

METHODOLGY

The desalted and purified DNA sample was purchased from IDT DNA. The sample was dissolved in 0.01M MES buffer pH 6.5 and 0.01M NaCl in D_2O. The DNA concentration was 1 mM. NMR data was collected on a Bruker 500 MHz Advance III. NOESY spectra were collected at 35°C with 8K data points per fid with 512 fids per experiment. DQF-COSY spectra were collected at 35°C with 4K data points per fid with 1024 fids per experiment. NMR data was processed using Bruker topspin and visualized using NMR fam SPARKY.

Assignments were made using a combination of correlation and NOESY spectra. Cross peak volumes were integrated using a Gaussian fit method and used for distance calculations. Distances were determined using the isolated spin pair approximation with the CH5-H6 distance and cross peak volumes as the references.

AMBER 14 was used to calculate the structure of the TT mismatch. NOE distance restraints were used to guide the NMR experiment. A total of 576 distance restraints consisting of NOE and Watson crick base pair restraints were used. A restrained simulated annealing calculation that started at 300K with an increase in temperature to 600K in 5 ps then gradual cooling to 0K over 20 ps was used. Restraints were applied during the entire calculation. The calculation was repeated 100 times and the resulting structures analyzed with Chimera.

2.Hoffman, Roy. "COSY." *COSY*, 26 Dec. 2018, chem.ch.huji.ac.il/nmr/techniques/2d/cosy/cosy.html.

3.Hoffman, Roy. "What Is NMR." *What Is NMR?*, 26 Dec. 2018, chem.ch.huji.ac.il/nmr/whatisnmr/whatisnmr.html.

4.Schima, Joachim. "Why 2D NMR?" *Two Dimensional NMR Spectroscopy*, 2016, www.cryst.bbk.ac.uk/PPS2/projects/schima/html/2dnmr.htm.

5'	C1	C2	G3	C4	A5	T6	T7	T8	G9	C10	G11	G12	3'
3'	G24	G23	C22	G21	T20	T19	T18	A17	C16	G15	C14	C13	5'

Table 1: This table is the schematic of the sequence being studied. The thymine-thymine mis pair highlighted in the table are observed at T6 -T19 and T7 - T18. This would created an unwanted kink in the DNA structure because it breaks the doubled hydrogen bonding rule, thus needing the proper A to satisfy the structure.

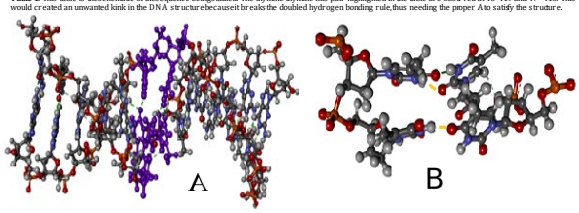

Figure 2: These structures were created on Discovery studio and are the B form of the schematic of the sequence. A is the DNA duplex that has the TT mismatch patch. The TT mismatch can be seen in the blue. B is a blow-up of TT mismatch pairs. As the top hydrogen bond is the T7 - T18 and the bottom is T6 - T19. The orange lines represent the hydrogen bonds that is happening between the nitrogenous bases. There are no distortions that were found throughout the structure.

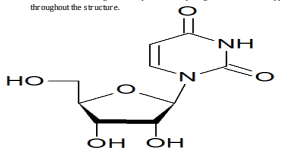

Figure 3: This Lewis structure represents the nucleotide of uridine (U) attached to a ribose sugar, which is found in RNA. U replaces thymine (T) and pairs with adenine (A). This structure was produced on Chemsketch. This Lewis structure was examined by COSY and NOESY spectra. These spectrums produced help interpret how the different hydrogens located off the uridine and ribose sugar create a three-dimensional structure in nature.

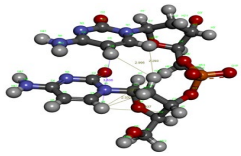

Figure 4: This C nucleotides are connected by phosphodiester bonds and was created on Discovery studios. Between the H6 on the bottom C to the H5 on the top is 3.84 Å. All the distances that were measured in the structure are less than 5 Å, thus we are able to use a NOESY spectrum to analyze the distances. The distance that are shown are an inter nucleotide proton-proton distance.

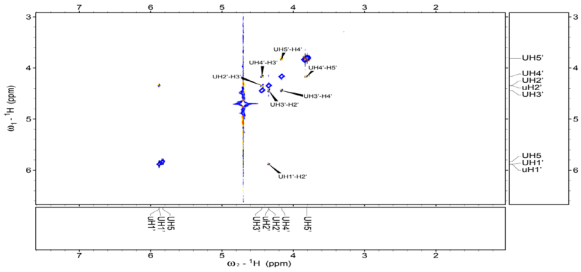

Figure 5: This spectrum shows the 2D-COSY spectrum of the U nucleotide. The characteristic peaks were assigned to each proton(H). The U correlation happens between the H2' at 4.45 ppm and the H2' and H4' protons. This correlation are able to appear since they have a distance of three bonds away. There is a symmetrical correlation that shows on the opposite side of the diagonal since it is a mirror image. The X and Y axis are used to see which correlation can be made as well as the peak found in the spectra that relates to that spot.

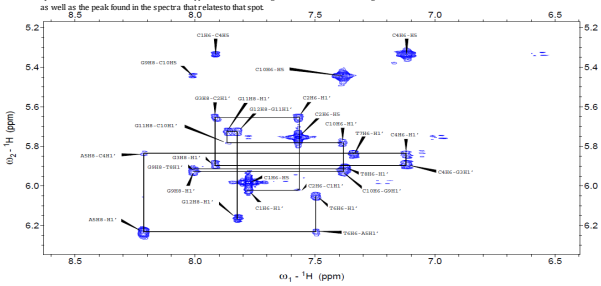

Figure 5: This spectrum represents the schematic of the sequences shown in Table 1. This 2D NOESY of the TT mismatch at the base to H1' region and CH5 to the previous nucleotide H6/H8 proton. This spectrum displays a distance in blue. The distance is 3.84 Å between H6' to H5' on the neighboring TT. The characteristic peaks that are larger will be closer in proximity. The weaker peaks help determine the 3D structure of the schematic of the sequences shown in Table 1.

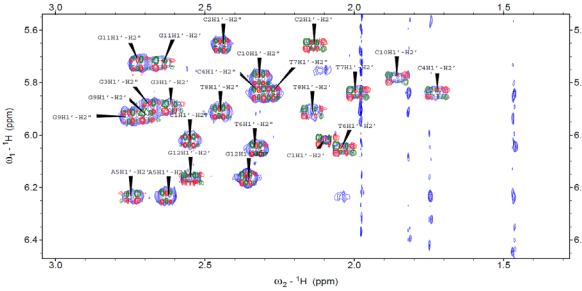

Figure 6: This spectrum displays the schematic of the sequence shown in Table 1 on the H1' to H2'/H2'' region of the COSY spectrum overlaid onto the NOESY spectrum showing the correlation and NOE data. The H2' and H2'' based off the deoxyribose and ribose sugar form. As mentioned previously the deoxyribose found in DNA lacks the OH on the second carbon whereas the ribose found in RNA contains the OH on the second carbon. The H2' is found on the DNA that lacks the second OH. The cross peaks between the H2' and H2'' is able to determine the difference by the patterns produced in the spectrum. The H2'' has a large peak which refers to the G H1', thus proving that the 3D structure of the DNA has the H2' is in a closer radius to the H1' proton on the G than the H2'. This overlay helps determine the distance of the two protons. By looking at the stress and the orientation of the protons help analyze the 3D structure of the nucleotide found in the sequence.

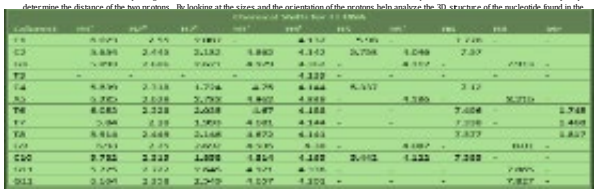

Table 2: This table shows the chemical shifts for the TT mismatch DNA. These values alludes to distances between the electrons and the electrons on the second thymine nuclei.