PROCTOR, NATHANAEL K., M.S. Studying the Effects of DNA Methylation on Adduct Formation Using Molecular Modeling. (2011)
Directed by Dr. J. Phillip Bowen and Dr. Norman Chiu. 65 pp.

The goal of this study was to use molecular modeling to compare and analyze the molecular structure of a double-stranded DNA fragment, and the effects of DNA methylation to adduct formation, which may eventually lead to disease-related genetic mutations. Specifically, the research work in this thesis focuses on using molecular modeling to simulate the experimental results in a recent report in which DNA adduction occurs with BPDE (benzo[a]pyrene diol epoxide) within a specific double-stranded DNA fragment that contained various methylation patterns and was quantitatively measured. The ability to use molecular modeling to correlate the pattern of DNA methylation and the locations of the most frequent adduction sites with genotoxic compounds can be very useful to further advance the study of genetic mutation, prevention of diseases, and so on. In this study the MMFF94s force field was used to run molecular dynamics simulations on dsDNA, and the results were analyzed to determine changes in the rotation of specific base pairs and the distances between the base pairs as a result of DNA methylation. The results show that there is a significant change in those two characteristics between non-methylated DNA and methylated DNA which might lead to adduct formation with BPDE.

STUDYING THE EFFECTS OF DNA METHYLATION ON ADDUCT FORMATION USING MOLECULAR MODELING

by

Nathanael K. Proctor

A Thesis Submitted to
the Faculty of the Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Greensboro
2011

Approved by

_____
Committee Co-Chair

_____
Committee Co-Chair

APPROVAL PAGE

This thesis has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Co-Chair _____

Committee Co-Chair _____

Committee Member _____

_____
Date of Acceptance by Committee

_____
Date of Final Oral Examination

ACKNOWLEDGMENTS

The author would like to thank my Lord and Savior, Jesus Christ, without whom nothing is possible.

The author wishes to express deep appreciation for the guidance and patience of Dr. J. Phillip Bowen and Dr. Norman Chiu as they directed this research project.

The author also wishes to thank Dr. Will Taylor for serving on the thesis committee and for giving direction throughout the course of this project.

Special thanks to my wife, Heather A. Proctor, for her unending encouragement, support, and love.

Thanks also to the Proctor and Loupe families for their encouraging words and support.

Finally, financial support in the form of a teaching assistantship from the Department of Chemistry and Biochemistry is gratefully acknowledged and appreciated.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

Human cells can regulate themselves by using a process known as DNA methylation. This methylation is vitally important in cell differentiation and expression of genes; however, when the methylation is uncontrolled or does not occur in the correct place, serious consequences such as cancer may occur. Recently, computer models have been used to simulate a wide variety of biological processes including protein function and binding. It is our goal to use computer-based molecular modeling and molecular dynamics simulations in the computer program SYBYL to determine slight changes in DNA structure such as distance between base pairs, torsion angles, and distance between complementary strands that occur with methylation that would give rise to higher chances of cancer developing. The specific sequence of DNA was already tested using LC mass spectrometry to determine where adducts occurred on methylated DNA, and our goal is to show that the same results would be predicted using SYBYL's simulated dynamics.

In terms of its molecular structure, human DNA is a very sensitive macromolecule. Slight changes in the DNA can have disastrous effects on human health. When nucleotides are modified or changed, the resulting DNA sequence can lose its genetic information, or it can become a problem for certain

cellular processes such as replication [1]. In one case, after an *N*-acetyl-2-aminofluorene or an *N*-2-aminofluorene was added to a guanine, the DNA sequence was changed in very specific ways [2]. It has long been known that DNA methylation is an important part of the cell cycle, and healthy cells have many methylated base pairs. For example, sixty percent of cytosine residues in CpG islands contain a methylation at the 5 position of cytosine [3]. This can lead to problems since these CpG islands contain about seventy-five percent of the promoters for genes [4]. The problems occur where there are excessive methylation events and the DNA becomes subject to adduction of different types, such as a carcinogenic diol epoxide metabolite, anti-benzo[a]pyrene diol epoxide, BPDE. BPDE is an adduct that can occur at multiple DNA bases. For example, it can occur on both the 3′ and 5′ sides of the adenine base although one side is more stable than the other [5]. Once the nucleobase has this adduction on it, the stability of the DNA double helix is changed, and it is possible that the change in stability causes the base to be recognized by a repair enzyme.

It is believed that an excess of methylation causes changes in the conformation of the DNA double helix. A study that was done previously showed that the adduct 4-OHEN forms at certain cytosine residues due to both the sequence and the secondary structure of DNA [6]. The secondary structure of the DNA is highly dependent upon the sequence; therefore, if the sequence changes slightly the secondary structure can change as well. These slight changes will then cause the double-stranded DNA to be more open and available

in some places where large adductions can gain access and react with the DNA base pairs. Once there is a large adduct in the DNA helix, other conformational changes can occur that have the possibility to cause major problems in the cell. For example, the gene expression in the particular region may become repressed, the DNA might not be transcribed properly, and the gene might not be expressed. In one study, it was found that the major cause of the P53 tumor formation in lung cancer is not caused by an endogenous pathway, but instead, methylation of the CpG sites of the gene is believed to cause the chemical carcinogen adduct, BPDE, to occur more often [7]. Therefore, in many cases the over-methylation of DNA can lead to diseases.

In a mass spectroscopic study, Paul Vouros and his associates at Northeastern University studied the effects of DNA methylation on a double-stranded DNA helix (5′-ACCCG$_5$CG$_7$TCC G$_{11}$CG$_{13}$C-3′/5′-GCGCGGGCGC GGGT-3′) [8]. In this study, they methylated several cytosine residues on one strand, known as strand 1, and incubated the DNA with BPDE. Then, they repeated the same experiment with methylation of cytosine residues on the opposite strand, strand 2, and finally with both strands methylated on the same cytosine residues. The next step was to determine if the BPDE adducts occurred, and if they occurred more often in one place relative to the others. Each BPDE adduct increased the molecular mass of the DNA, and therefore a method using an on-line nanoLC/MS/MS was used to determine the adduct formation. It is known that when a BPDE adduct binds to the DNA, it does so at

the C_8 or N_2 position of the guanine [9]. When the results were analyzed, it was found that adduction occurred on all of the guanine residues with the guanine at the fifth position, G_5, being the most common site for BPDE adduction. Their results have also shown that with only one strand methylated, G_{11} formed BPDE adduction more often than G_{13}, which in turn had a higher frequency of adduction than G_7. However, when both of the strands are methylated the sequence of adduction changes so that G_5 had a higher frequency of adduction than G_7 which was higher than G_{11} which was higher than G_{13}. Therefore with excess methylation, a difference was seen in the trend of BPDE adduction. They concluded that DNA methylation significantly increases the chances of BPDE adduction.

The goal of molecular modeling is to predict the behavior of chemicals using thermodynamic and quantum mechanical rules prior to experimentation. When applying these rules to a biological environment, many factors influence the results of the simulated dynamics. Some of the advances and the areas of struggle that still exist are discussed by Cheatham and Young [10].

In their discussion, they point out that great advances have been made in the area of ion interaction. Ion interaction is important for proper DNA molecular dynamics because the negative charges on the backbone of the DNA will repel each other and lead to splitting of the DNA strands or the two strands separating from each other. Usually Na^+ ions are used to interact with the backbone, and a common setup is to place the sodium ions about 6 Å away from the backbone in

the groove [11]. The positive charges on the sodium ions will interact and balance with the negative ions of the phosphate groups of the DNA backbone.

Another success that was mentioned is the ability of models to allow the DNA to bend. Large DNA molecules will have a curvature, and presently some molecular dynamics simulations can allow that to happen in the simulation as well. One final success that was noted is the ability to use molecular dynamics on varied DNA structures. Computer-based modeling of DNA is now able to demonstrate each of the different conformations of DNA. For example, in living cells DNA can adopt an A-form, B-form, or Z-form. The differences between these conformations is due to the puckering of the sugar ring. Both A- and B-form DNA have a right-hand helix conformation; however, Z-form DNA is in a left-handed helix. Molecular modeling can be used to build these different forms of DNA.

However, even with all the advances that are seen, there are still some problems that arise when running molecular dynamics simulations. The first of these is the conformational sampling. When a dynamics simulation runs, the desired result is the one that has the lowest energy trajectory or the one that traces the path of movement in a low energy conformer. It has been shown that the molecule might reach a local minimum and get stuck in that conformation even though it is not representative of what is really happening [12]. An analogy can be made to a valley between two mountains. At the bottom of the valley is the lowest energy conformer. When the conformers are checked for energy, any

conformer with energy higher than the previous conformer is discarded. Therefore, the last result will hopefully be the one that is the lowest in energy. However, perhaps there is a small perturbation in the energy surface that behaves like the energy well. If a conformer is found to be in that energy well, then it will be in a local minima; however, it has not reached the overall minimum. This can then lead to a misrepresentation of reality.

Another weakness that is apparent with nucleic acid modeling is the energy calculations which is a function of the force field that is applied. Each force field has different equations and parameters along with its strengths and weaknesses. For example, some force fields will allow the study of different types of DNA, such as the Hoogsteen base pairing that is not commonly seen in the majority of DNA but that still occurs in native DNA [13]. Therefore, in order to be able to model this observed structure, different force fields must be applied that have a different set of parameters. The application of which energy is best will be discussed later.

Another struggle that has faced molecular modeling with DNA is the anomeric effect in the nucleosides. Nucleoside sugars have different conformations, and there is an equilibrium between the North and South conformations. The anomeric effect is due to lone pair electrons arranging *anti* to electronegative heteroatoms. With DNA, the sugar, a furanose, contains an oxygen, and an aromatic nitrogen is in the base, either a purine or a pyrimidine. These two atoms lead to the anomeric effect occurring in the sugar which will

affect the puckering of the sugar ring. There are two conformations, the East and West, which are the barriers that must be passed through to convert from North to South or vice versa. One study that has been done showed that with DNA molecular modeling the eastern barrier is too low and allows for interconversion between the two conformations much easier than it should be [14]. Therefore, the representation of molecular dynamics with DNA would not be true to reality.

The development of solvent models is also still a challenge. Solvation is applying a solvent to the biomolecule that is being studied. In the human cell, all of the biomolecules are solvated with water. Therefore, in order to have proper molecular dynamics simulations that mimic nature, the biomolecules must be solvated. Problems arise when the DNA double helix becomes solvated. The cost on the machine increases significantly when large portions of nucleic acids are solvated. This refers to the amount of resources used by the computer. If too much is used, the computer can slow down or become unresponsive. The authors mention two ways that help avoid these problems, and that is by using an implicit solvation model. Cheatham and Young [10] mention two types of implicit solvent models that are commonly used. One is a Born methodology [15], and the second is using a Poisson-Boltzmann method [16]. These two types of implicit solvation have helped cut down on the cost of running a molecular dynamics simulation. Another problem arises with the lack of structural water. In native DNA, water will align so that the hydrogen is in the backbone of the DNA. In simulations, this is still a step that is hard to

accomplish, although some force fields and parameters do allow this to occur [13]. If these problems are not addressed, the resulting simulations might then show bends or twists that are not representative of the true nature of DNA.

There are many different sets of molecular mechanics parameters that can be used to run a simulation. The objective is to apply a set of parameters that will give the best results for the specific problem. MMFF94s simulation parameters, for example, will be different than the original MMFF94 parameters. One difference between the two models is that the MMFF94s parameters change the out-of-plane bending with planar geometries using nitrogen [17]. A study was carried out by Halgren that compared the MMFF94, MMFF94s, CFF95, CVFF, MSI CHARMm, AMBER*, OPLS*, MM2*, and MM3* force fields in determining conformational energies of a large set of known molecules [18]. His findings showed that the MMFF94 and MMFF94s force fields had the best results and were the most consistent; however, there were problems with these as well. This should be expected since no computer-based modeling scheme or mathematical model is perfect.

Also, when modeling double-stranded DNA, it is important to remember how DNA behaves in living cells. There is a breathing of DNA that occurs naturally when the two strands come apart and then hybridize back together. Studies have been done showing how well computers are at modeling this behavior when comparing that to *in vitro* experiments [19]. In these studies, it was found that the time-scale for reliable results when using molecular dynamics

with DNA is less than nanoseconds. The reason for the unreliability is that the double-stranded DNA helix will fall apart after a period of time close to the nanosecond range. A newer study has been carried out and showed that a 5 nanosecond molecular dynamics simulation has been observed that did not fall apart [20].

As can be seen, significant advances have been made in the area of computer modeling. There are still problems that are being addressed, and adjustments are being made constantly in order to make molecular dynamics simulations reflect experimental observations. Molecular dynamics simulations for double helix DNA has been a challenging area, but there have been vast improvements in certain areas that allow us to have confidence in obtaining reliable results when trying to replicate experimental work that has already been carried out.

CHAPTER II

GOALS

It was our goal to model the same DNA sequence that was used in the study carried out by Prof. Paul Vouros [8]. The first goal was to create the methylated 14-mer double stranded oligonucleotides, as was used in the previous study by Vouros, in a computer modeling program. This was accomplished in a way to prevent the DNA from adopting a conformation that is unrepresentative of DNA in nature.

The next goal was to generate counterions to balance out the negative charges of the DNA backbone. The backbone contains many negative charges associated with the phosphate groups. Therefore, placement of the counterions is important so that the negative charges will not repel each other and cause the DNA to fall apart in a molecular dynamics simulation. The ions are typically placed in the minor groove 6.0 Å away from the phosphate groups.

The third goal was to solvate the DNA with water in a box with periodic boundary conditions. Solvation is very important, as mentioned before, with all DNA modeling to have results that are reliable. The periodic box was set up so that the dsDNA could be contained in a molecular area with the water that is solvating it. Also, when setting up periodic boundary conditions, multiple boxes were generated which also contain the molecules that were being studied.

These added solvated molecules could then be used as well to collect more data during a molecular dynamics simulation. The box is set up with boundaries so that the DNA in one box will not interact with the DNA in another box. Otherwise, these interactions would generate additional complications.

The fourth goal was to minimize the energy of the oligonucleotide so that the conformation would be most similar and representative of natural DNA. With minimization, it is important to already have solvated the DNA with water so that no unnatural bending would occur that would lead to interaction with other parts of the DNA that would cause the results to be unreliable.

The next goal was to run a molecular dynamics simulation of the different DNA double helices using the MMFF94 force field. The final goal was to analyze the results obtained to determine whether or not something happens around the G_5 position of the methylated strands of the DNA double helix when compared to the unmethylated DNA double helix. Changes that would occur in the torsion angle or the distance between successive base pairs could lead us to believe that structural changes are occurring in the methylated DNA that would allow an adduct such as BPDE more accessibility.

CHAPTER III

METHODS

The preliminary data collection was not actually for comparison with the previous work which would consist of running the molecular dynamics on the four different methylated patterned DNA strands. Instead, it was to familiarize ourselves with the SYBYL program. Second, it was also to calibrate our work to previous work. The first objective to be carried out was to create the sequence used by Vouros of 5′-ACCCGCGTCCGCGC-3′ [8]. This was accomplished by using the biopolymer build command in SYBYL. All of the work that was done was with SYBYL version 7.2.2. Previous DNA molecular modeling studies have been done using SYBYL; therefore, we believed this program would be appropriate to use [21]. The B-form of double stranded DNA was created, and the complementary strand was automatically generated based on the strand sequence that was entered in. Then, in our initial work, each of the cytosines was methylated at the five position independently so that each double strand had only one methylated cytosine. So a double strand of DNA was created that had the C_2 methylated, one that had the C_3 methylated, and so on until eight different DNA molecules of the same sequence were generated but with different cytosines methylated in each of the positions. An unmethylated DNA sequence

was also created which was available for comparison of our results. Therefore, nine different DNA molecules were created and modified in all.

Each of these molecules was then solvated with water using SYBYL's implicit command for solvation. The DNA was then energy minimized with MMFF94s, and the resulting structures were analyzed. The distance from the guanine to the complementary cytosine on the opposite strand was measured. No molecular dynamics simulations were carried out on these molecular structures, but while the data may not be significant since it was not directly related to the Vouros study, the experience gained by learning the commands was invaluable.

The first step in running the computer simulations was to build the DNA double helix and modify the DNA according to the study by Vouros. This was accomplished by using SYBYL's biopolymer – build command. B-form DNA was selected as the conformation used to build the duplexes since it is the most abundant form in living cells. The sequence 5′-ACCCGCGTCCGCGC-3′ was built. Since the double stranded DNA had been selected, the complementary strand was automatically prepared by SYBYL.

Initially, to make modifications to the DNA, the build/edit command in SYBYL was used. Unexpectedly, as soon as the changes were completed the dsDNA would split apart drastically. This showed a problem in some of the parameters when trying to modify DNA in this way. An alternative way to make changes to the atoms was used to convert the hydrogen on the five position to a

methyl group. This consisted of using the command – “change atom”, and changing the hydrogen of the carbon in the fifth position on the appropriate cytosine to a methyl group. This spot on the cytosine is the position that is methylated in human DNA. In DNA duplex 1, strand 1 was methylated at four sites corresponding to the Vouros study. These are shown by a superscript M in front of the corresponding cytosine, 5′-ACC^{M}CG^{M}CGTC^{M}CG ^{M}CGC-3′. Similarly, the DNA duplex 2 had four methylation sites on strand 2 of the double-stranded DNA, 3′-TGGG^{M}CG^{M}CAG G^{M}CG-^{M}CG-5′. DNA duplex 3 was created such that each of the two complementary methylated strands was together. So the total sequence for duplex 3 was 5′-ACC^{M}CG^{M}CGTC^{M}CG-^{M}CGC-3′/3′-TGGG^{M}CG^{M}C AGG^{M}C G^{M}CG-5′. Finally, a fourth DNA duplex was created with no methylation sites at all, and this was considered to be the control.

Once all of the DNA duplexes were created, they needed counter ions placed in the groove for ion interactions to balance the negative charges of the phosphate backbone. There is no command in SYBYL for adding counterions; therefore, a script was used that would add the counterion of our choice to a position that corresponds to the dihedral of the phosphate backbone at a distance that was specified. Using the script, Na^{+} ions were placed along the backbone at a one-to-one ratio with the negative charge of the phosphate groups. The Na^{+} ions were not just point charges as is the case in some studies. Each Na^{+} ion had a van der Waals radius of 1.2; therefore, the ions were more representative of actual ions in human cells. The script can be found in Appendix

A. The distance that was selected was 6.0 Å. Each of the DNA duplexes now had a charge balance for the backbone negativity.

The DNA duplexes were then solvated with water. An implicit solvation model within SYBYL was used to place water molecules around the DNA. This model will add water molecules to fill up a periodic boundary condition box. The number of water molecules to be added is calculated by SYBYL. For DNA duplex 1, 3068 water molecules were added to the box to solvate the DNA. DNA duplex 2 had 2547 water molecules added, DNA duplex 3 had 2521, and finally DNA duplex 4 had 2972 water molecules added to the box containing the DNA.

The next step was to minimize the DNA in order to reach a minimum energy conformation as the starting point for the molecular dynamics simulation phases of the studies. All of the DNA duplexes were geometry optimized using the MMFF94s force field with charges being assigned with the MMFF94s parameters. The maximum iterations was set to 100,000. Everything else for the initial setup was left in the default mode. The energy minimization was then run for each of the four duplexes.

After the energy minimization was completed, a visual comparison was made of the different duplexes in regard to the guanine at the fifth position which in the study by Vouros was the position where the majority of the adduction occurred. The biopolymer command of compare structures was used to carry out this task. Only the guanine of the fifth position was selected for the comparison. In the options for comparison, there were different choices for the fit

and alignment of the two guanines. The first that was done was a fit with the sidechain which in this case would be the nucleotide base. The second comparison made was with a fit of the backbone which would be the phosphate groups aligned with each other. The final visual comparison was made with the whole molecule aligned and fit to the other as best as possible. To differentiate the two guanines, the guanine of the unmethylated strand was colored purple, and the guanine of the methylated strand was colored green. A visual comparison could be made of the two guanines and how the position at G_5 changes with an energy minimization with methylated DNA. However, this is not conclusive evidence since DNA *in vivo* is constantly breathing and moving.

Once the energy minimization was completed, a molecular dynamics simulation was initiated. The setup of the dynamics simulation was done using the MMFF94 force field, and the charges were automatically assigned by MMFF94. The length of the dynamics run was set to be 10,000 fs with a snapshot taken every 50 fs. This would generate 201 samples that would be analyzed. Everything else for the setup conditions was left in the default mode.

When the dynamics simulations were finished, the resulting DNA duplexes were analyzed using the Analyze Dynamics command. For each of the duplexes, the torsion angle was measured at the backbone of the specified guanine base pair and in either direction on the base pair. It was anticipated that the twisting of the guanine might be a cause for the BPDE adduction by making the guanine more accessible. The measurements were made by using a

spreadsheet formulated by SYBYL. A graph was also produced by plotting the torsion angle versus time to show how the guanine behaved during the dynamics simulation.

The distance was also measured from the C_8 position of the guanine residue, where the BPDE adduct will occur, to the nitrogen directly above it on the cytosine 4 residue. Along with this the distance from the C_8 of the guanine to the carbon directly below on the cytosine 6 residue was monitored. These distances were also measured using the same spreadsheet as before and plotted against time to show the movement of the guanine residue.

After further consultation, it was thought that the distances that had been measured and the torsion angles were probably correlated; therefore, the data would not be two independent effects. It would be biased to say that the distance increased as well as the torsion angle because they are so closely correlated with the atoms that were chosen to measure the distances between them. If, however, the distance was measured from one glycosidic nitrogen to the next glycosidic nitrogen, then the results would not be as strongly correlated and would be much more significant; therefore, the distances were then measured between the nitrogens of the glycosidic bond above and below the guanine base.

The final analysis and comparison that was done was that the torsion angles of each of the different guanine positions of interest in the different methylated strands were measured. These angles could then be compared with each other to observe the effects of methylation within the DNA molecule. The

distances between the glycosidic nitrogens above and below were also measured for each of those guanines as well to see the changes within each molecule and to make comparisons with those in the study done by Vouros.

CHAPTER IV

RESULTS AND DISCUSSION

The results of the initial studies with each cytosine independently methylated can be seen in Appendix B. Figure 1 shows the two complementary nucleotides that were used for the distance measurements. The specific atoms that were used have been labeled as well.

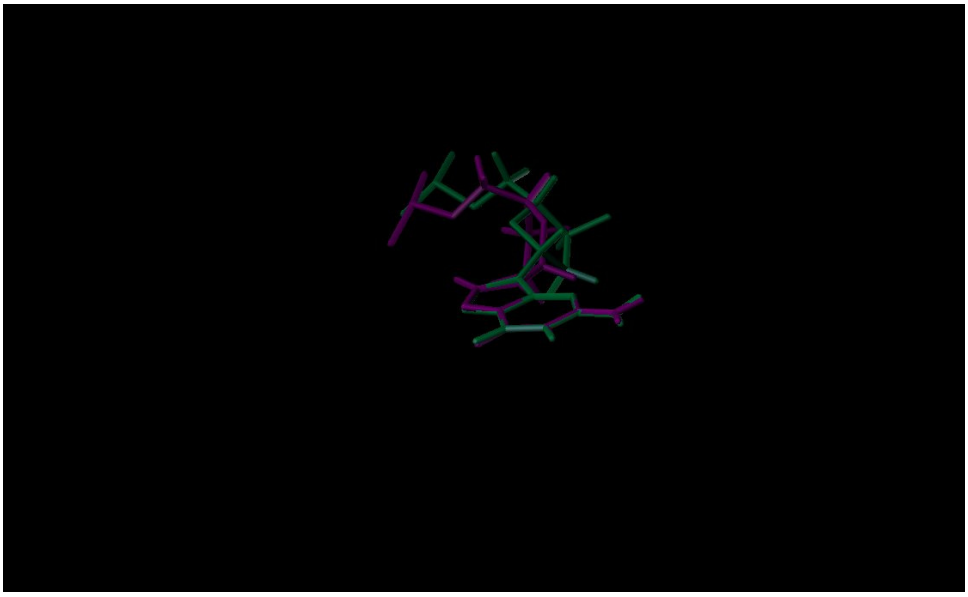
Figure 1. Guanine paired to complementary cytosine.

The first measurements that were taken were from the middle of the two rings, the N_3 of cytosine and the H_1 of guanine. This measurement, however,

would not have taken into account the distance that could be due to the twisting of the bases. Therefore, the distance was measured from the oxygen atoms of the rings to the opposite hydrogen atoms. Again, the data were not significant to the overall goals and results of this study, but these results show that there are observable changes that occurred with methylation even if only one cytosine was methylated.

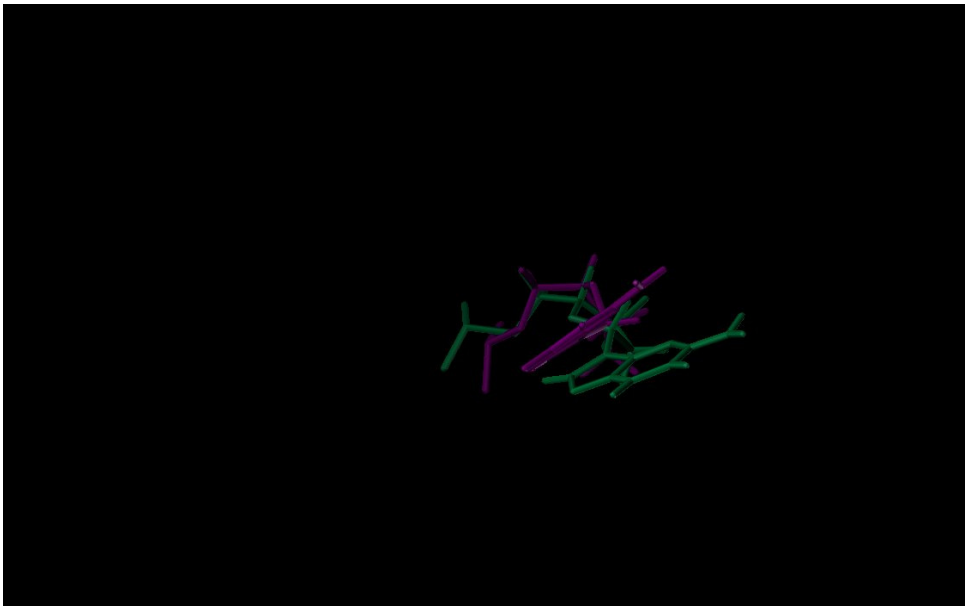
A visual comparison of the guanine bases of the methylated duplex and the unmethylated duplex was carried out. Figures 2, 3, and 4 show the G_5 of the unmethylated duplex when compared to the G_5 of the duplex with both strands being methylated.

Figure 2. Guanines aligned by base. (purple = unmethylated, green = methylated)



When comparing the two guanines, if the bases are fit to each other, not much difference was seen between the two bases. This was to be expected since the bases have not had any direct changes to themselves. Looking at the backbone, there were quite a few differences that occurred due to increased methylation. The backbones were quite different from each other, and therefore conformation changes were happening due to the methylation.

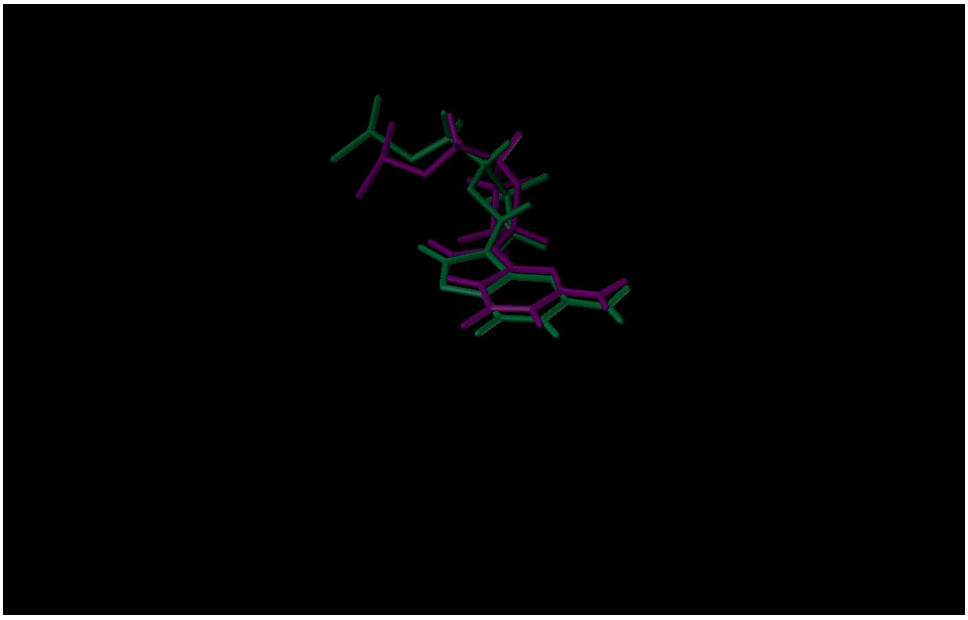
Figure 3. Guanines aligned by backbone. (purple = unmethylated, green = methylated)



When the backbones were aligned and the bases compared to each other, there was a significant change observed between the two bases. The base of the methylated strand showed a torsion shift as well as a shift in position compared to the guanine of the unmethylated strand. These changes could be

what leads to space around the guanine opening up and more adduction occurring at this site.

Figure 4. Guanines aligned by all atoms. (purple = unmethylated, green = methylated)



With the guanines aligned by all of the atoms in the backbone and the base, the differences between the unmethylated strand and the methylated strand were not observed as clearly as before. It should be noted that there was still a change in the torsion of the base. The change in distance was even less obvious here since the bases appear to be on top of each other. It could be seen that the site of adduction was actually more accessible in the guanine of the methylated strand; therefore, the changes that are observed could possibly lead to adduction occurring more frequently with methylation. In order to get more

accurate and reliable results, the results of the molecular dynamics simulations needed to be analyzed.

The results of the rest of the first part of the project showed that there was a difference between DNA strands that occurs with the different methylation patterns of DNA. Graphs were created of the data that were obtained and can be seen in the Appendix D. Each graph shows the torsion angle versus the time elapsed of the molecular dynamics simulation. In the first set, the torsion angle around the glycosidic bond of the G_5 of each of the duplexes may be seen. Figure 5 shows the atoms that were used to find the torsion.

Figure 5. Torsion angle atoms.

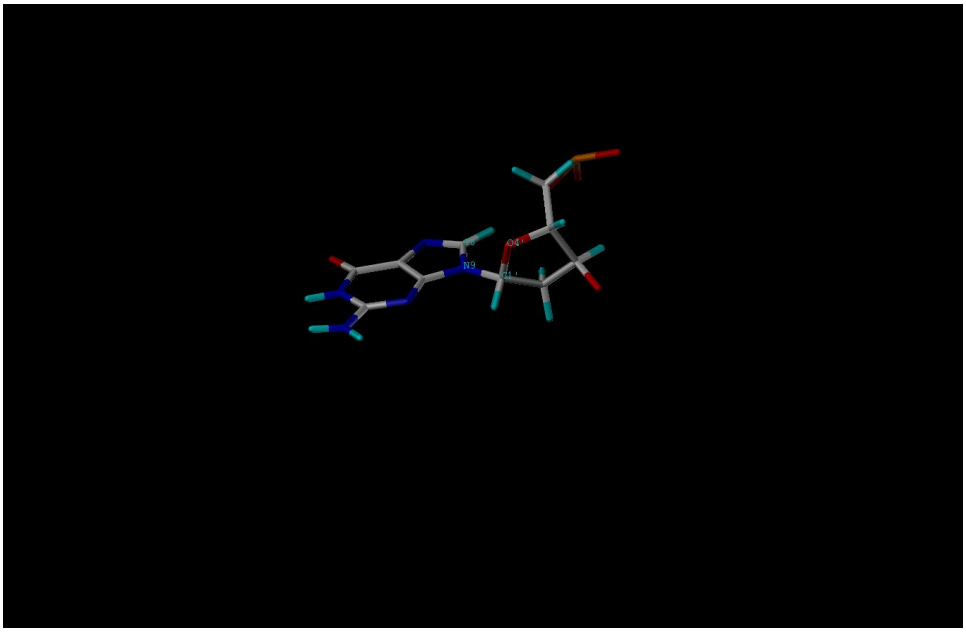


Figure 6. View of planar base.

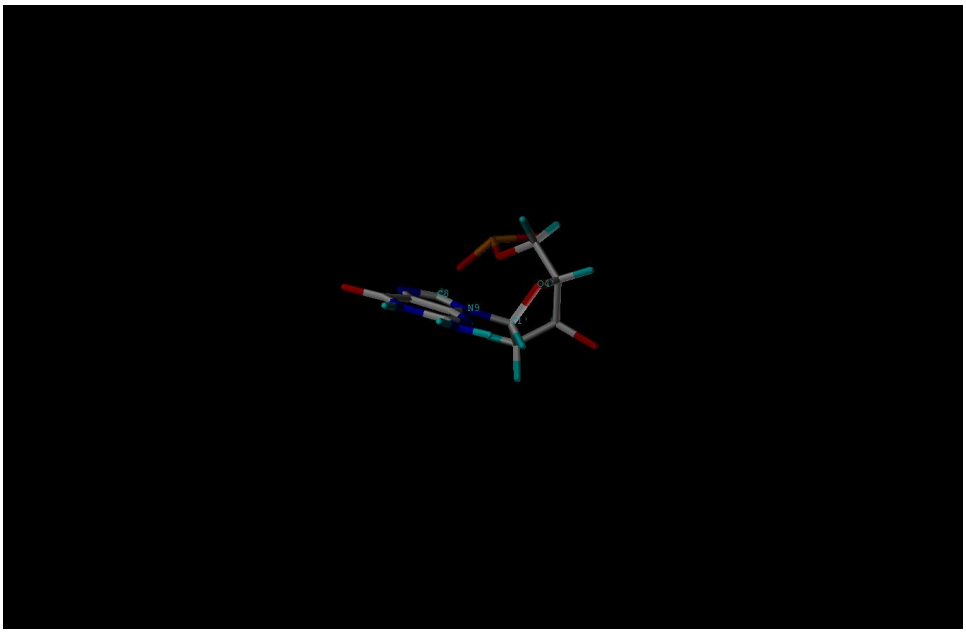


Figure 6 shows the planar base that is part of the guanine that would be influenced by changes in the torsion angle The fluctuations could be observed over time, but the average and standard deviation are reported for each duplex in Table 1. This table shows that the torsion angle increased slightly with added methylation. Although the standard deviations do allow some overlap, the data show that over a period of time, in methylated strands the torsion angle was greater than in the unmethylated strands. This could be the result of the base twisting out of the normal plane which would in turn allow for more opportunity of the adduct occurring at the G_5 position.

Table 1. Torsion angles, standard deviations, standard errors of the mean, and P-value compared to unmethylated DNA around G_5 of each oligonucleotide.

	Average Torsion Angle	Standard Deviation	Standard Error of the Mean	P-value
Unmethylated DNA	39.46º	22.27º	1.570º	
Methylated Strand 1	61.65º	12.09º	0.8528º	<0.0001
Methylated Strand 2	59.92º	11.97º	0.8443º	<0.0001
Methylated Strand 1 & 2	64.15º	14.89º	1.050º	<0.0001

When looking at the data obtained from measuring the torsion angles, it can be observed that there was considerable overlap between the angles when the standard deviation was taken into account. In order to determine if the data are significant and if the changes that are observed are not just minor changes and movements, simple statistical tests were run on the data. The t-test was run on the torsion angles to see if the data were significant. The p value was less than 0.0001 which means the data are indeed significant. The data that were generated are provided in Appendix C.

Distances were also measured above and below the G_5 base to get an idea of the space around the base. The graphs of these data can also be found in the appendix. The measured distances are from the C_8 position of the G_5 base to the C_5 position on the C_6 base and can be seen in Figure 7. The C_5 position on the C_6 base was also a site for methylation. The average distances and standard deviations for the measurements can be seen in Table 2.

Figure 7. Picture of the two nucleotide, G_5 and C_6, between which the distance was measured using the C_5 and C_8 atoms.

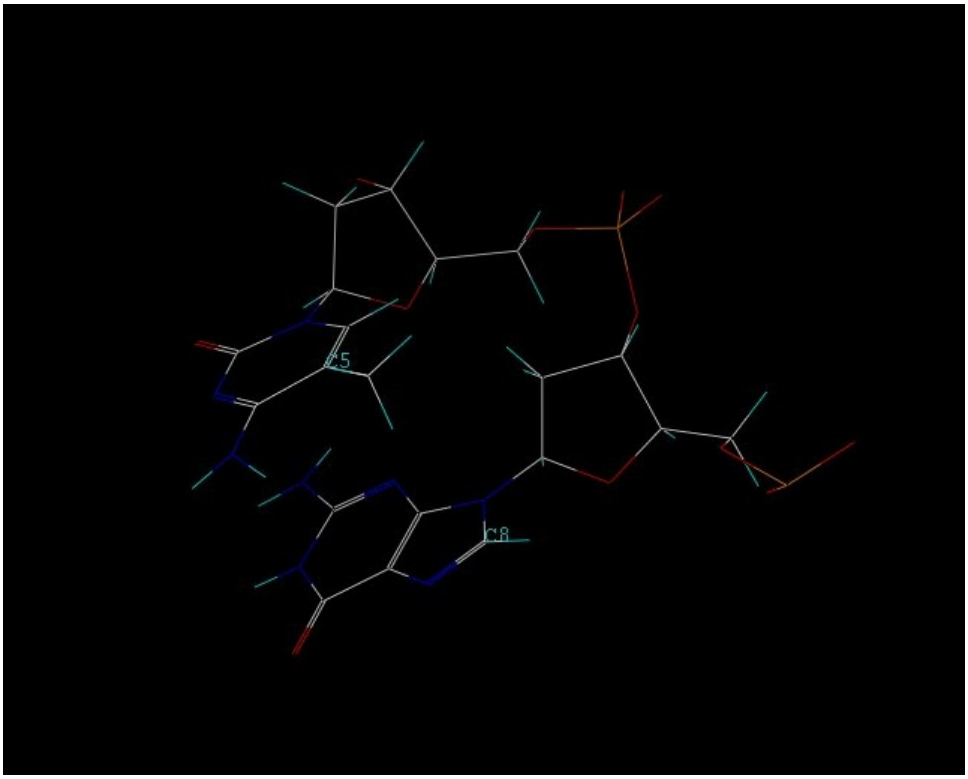


Table 2. Distances, standard deviations, standard errors of the mean, and P-value compared to unmethylated DNA between G_5 and C_6 for each oligonucleotide.

	Average Distance	Standard Deviation	Standard Error of the Mean	P-value
Unmethylated DNA	3.78 Å	0.23 Å	0.0162 Å	
Methylated Strand 1	4.12 Å	0.25 Å	0.0176 Å	<0.0001
Methylated Strand 2	4.04 Å	0.31 Å	0.0219 Å	<0.0001
Methylated Strand 1 & 2	4.72 Å	0.40 Å	0.0282 Å	<0.0001

The data show that there was an increase in the distance between the G_5 and C_6 bases. When comparing the unmethylated duplex with a duplex that has only one strand methylated, the distances are not significantly different because there was overlap when the standard deviations were taken into account. When the unmethylated duplex was compared with the duplex that has both strands methylated, there was no overlap with the standard deviations. The t-test was also carried out on this set of data, and the p value was found to be less than 0.0001. Again, this signifies that the actual data obtained and the changes occurring are significant. The results are also significant because the measurements were taken from the region of the DNA where the adduction will occur. Therefore, if the distance was increasing at that point it could be that there was more space in that area, and that would presumably lead to an increase in availability for the adduct to gain access and bind to the DNA. Taking into account both the torsion angle and the distance between the bases both increasing with increased methylation, it was reasonable to suspect both of these for possible explanations of why adduction will occur at the G_5 position.

These results, however, could easily be correlated since the measured distance between the guanine and cytosine would change as the torsion angle of the guanine changed. Therefore, in order to get distance data that was not coupled to the change in the torsion angle, the distance between the glycosidic nitrogens of the bases above and below the guanine in the fifth position was

measured. Figure 8 shows the atoms that were used for the distance measurements. The calculation results obtained are displayed in Table 3.

Figure 8. Glycosidic nitrogen atoms that were used for distance measurements.

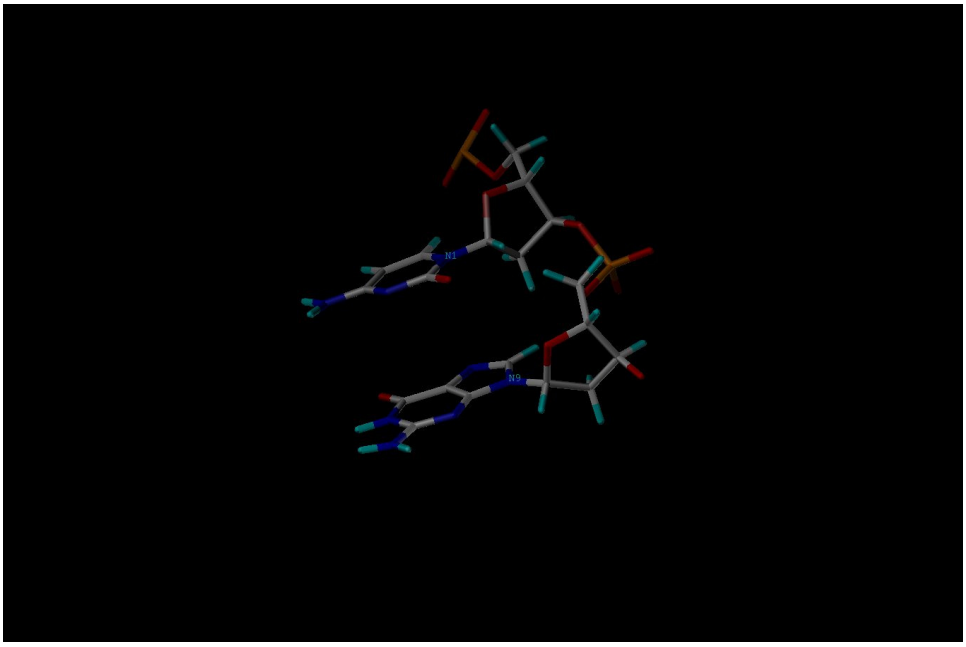


Table 3. Distances, standard deviations, standard errors of the mean, and P-value compared to unmethylated DNA between glycosidic nitrogens of adjacent base pairs.

	Average Distance	Standard Deviation	Standard Error of the Mean	P-value
Unmethylated DNA	4.40 Å	0.23 Å	0.0162 Å	
Methylated Strand 1	4.29 Å	0.25 Å	0.0176 Å	<0.0001
Methylated Strand 2	4.08 Å	0.20 Å	0.0141 Å	<0.0001
Methylated Strand 1 & 2	4.74 Å	0.32 Å	0.0226 Å	<0.0001

The data do not show a significant increase with methylation on just one strand. In fact, the distance actually decreases when only one strand was methylated, as seen in methylated strand 1 and methylated strand 2 which was not expected. When both strands were methylated the data show an increase in the average distance, so over the given period of time the distance was greater when there was excess methylation compared to when there was no methylation. Once again the t-test was run, and it gives a p value of less than 0.0001 when comparing the unmethylated strand to the duplex with both strands methylated which shows that the data are significant. However, it seems that the major factor affecting the adduction rate would be the torsion angle. The hypothesis is still supported that perhaps this excess methylation leads to conformational changes in the DNA that opens it up more around the guanine in the fifth position to allow adduction to occur.

The second part of the study was again to look within each DNA strand to see if our results are comparable to Vouros' study with respect to the adduction rate at $G_5 > G_{11} > G_{13} > G_7$ when analyzing the DNA with only one strand methylated. It was also another goal to see the trend of $G_5 > G_7 > G_{11} > G_{13}$ when looking at the DNA with both strands methylated. The same analyses were done, and the data generated may be found in Appendix D. When reviewing the data, it was important to look at the differences that occur between the unmodified DNA and the different methylated DNA strands. Therefore, the data do not seem to follow the trend when comparing the angles and distances of

each guanine in the specific strand being studied. However, when the data obtained from normal DNA are subtracted out the resulting data are more reliable and actually match the expected trend better. The results can be seen in Table 4.

Table 4. Torsion angles and (differences from unmethylated DNA)

	G_5	G_7	G_{11}	G_{13}
Unmethylated	39.46º, (0º)	67.80º, (0º)	84.14º, (0º)	76.49º, (0º)
Methylated Strand 1	61.65º, (22.14º)	46.13º, (21.67º)	80.05º, (4.09º)	79.68º, (3.19º)
Methylated Strand 2	59.92º, (20.46º)	24.89º, (42.91º)	58.18º, (25.96º)	79.50º, (3.01º)
Methylated Strands 1 & 2	64.95º, (25.49º)	78.77º, (10.97º)	49.65º, (34.44º)	72.45º, (4.04º)

Disappointingly, the data obtained with molecular dynamics do not show the same trends that were seen in Vouros' study. In none of the cases was the trend followed. When observing the methylated strand 1 and methylated strands 1 and 2, the torsion angle was greater at guanines other than the fifth position. In the Vouros study, the guanine in the fifth position was the site of the greatest adduction. Interestingly, a general trend was seen that the least change

occurred in the guanine at position thirteen which supports the idea that the guanine at this position would receive the least amount of adduction occurring. This would follow the trend when both strands contain methylation, but not when only one strand is methylated. Therefore, the data are not conclusively supportive of our hypothesis.

The data for the distance measurements above and below can be seen in Appendix E as well. Table 5 shows the results of the total distance change around the specified guanine base. This means the change occurred due to methylation above and below the guanine. The distance was obtained by subtracting the unmethylated DNA distance from the methylated strands above and below and then summing those two numbers. Figure 9 shows the area surrounding the guanine, and how it is possible that changes in the distance both above and below the base occur with methylation. For example, if the base tilts up, then the space below will be greater, but if the base tilts down, the space above will be greater. Therefore, the total change in distance is reported in Table 5.

Figure 9. Space above and below guanine.

Table 5. Differences in distance above and below guanine base from unmethylated strand.

	G_5	G_7	G_{11}	G_{13}
Unmethylated	0.0 Å	0.0 Å	0.0 Å	0.0 Å
Methylated Strand 1	0.44 Å	-0.22 Å	0.03 Å	0.17 Å
Methylated Strand 2	0.13 Å	0.27 Å	0.73 Å	0.61 Å
Methylated Strand 1 & 2	0.21 Å	-0.08 Å	0.40 Å	0.22 Å

None of the distance data when subtracted from the unmethylated DNA data shows the trends that were observed in Vouros' study. In some cases, the G_7 of methylated strand 1 and methylated strand 1 and 2, the data actually show a decrease in the total distance above and below the guanine. The raw data

seen in Appendix E show that both increases and decreases occur. Perhaps, the base is tilting up or down more in the case of methylated DNA than in the unmethylated DNA, but it is hard to show that definitively for every case. For example, in comparing the distance data for the unmethylated strand and the methylated strand 1, for G_5 there was an increase in the distance from C_4 to G_5 when going from unmethylated to methylated. In G_{11} there was a decrease in the distance from C_{10} to G_{11}, but there was an increase in G_{11} to C_{12} when going from unmethylated to methylated. This means that each base behaves differently, and in some cases the area above the guanine opens up more with methylation, and with other guanines in the same strand the area below the base opens up more with methylation. Table 6 shows the change in distance from the unmethylated DNA that is the greatest, either above or below the guanine base. It should be noted that the adduction should have more chances to occur, regardless of whether the space opens up above or below the guanine.

Table 6. Change in distance (Å) above or below guanine base.

	G_5	G_7	G_{11}	G_{13}
Unmethylated	0.0	0.0	0.0	0.0
Methylated Strand 1	0.55	-0.03	0.10	0.21
Methylated Strand 2	0.45	0.39	0.23	0.32
Methylated Strand 1 & 2	0.34	0.31	0.28	0.28

The change in distance either above or below the base that occurs due to methylation can be seen to increase in most cases. The one exception is the G_7 of the methylated strand 1. Once again, the trend that was expected was not clearly observed. In all of the methylated strands the guanine in the fifth position had the most change which does support the trends seen in the Vouros study. In the last case of the methylated strand 1 and 2, the trend was most closely followed. The only difference in only difference in this case was that the guanine in the eleventh and thirteenth positions had the same change. Overall, the data do not support the trends seen in the previous work. There are increases that are seen in both the torsion angle and the distance for all of the methylation cases, but it was difficult to say definitively which one is more important in causing increased susceptibility for adduction.

CHAPTER V

CONCLUSIONS

Many of the goals defined in this study were able to be completed. First, a dsDNA molecule was created using the SYBYL program, and then it was possible to modify the DNA to generate methylated DNA strands similar to the Vouros study. This was easily accomplished by the SYBYL program that was used. The commands of building biopolymers were simple, and the sequence of DNA was very quickly generated.

Second, it was possible to successfully place Na^+ ions to counteract the negative charges of the phosphate backbone in order to stabilize the DNA for molecular dynamics simulations. While there is no implicit command for counterion placement, a script was provided that enabled SYBYL to place counterions. The placement of the ions actually matched a method that was reported elsewhere with the ion placed at the vertex of the phosphate group of the backbone 6 Å away. This action stabilized the dsDNA for energy minimizations and dynamics simulations.

Another goal that was accomplished was the ability to solvate the DNA in a periodic boundary box with water to simulate reality in a cell. Since DNA *in vivo* is surrounded by water, energy minimizations and dynamics simulations need to incorporate water to be reliable. The problem with solvation that is faced

by molecular modeling with DNA is that the water will not arrange itself along the DNA in such a way that the positive pole of the water molecule is pointing toward the negative DNA molecule. However, even though the water molecules in the SYBYL system exhibited this problem, it is better to have the problems associated with solvation than to try to run a dynamics simulation *in vaccuo* because excessive problems will occur when running a dynamics simulation in a vacuum.

Another goal that was accomplished was running a molecular dynamics simulation and analyzing the results to generate data that could be compared to the Vouros study. The dynamics simulation allowed the molecule of DNA to move and, theoretically, behave as it does in living cells. The analyses then allowed for measurement of the specific parts of the DNA that were of interest in order to compare how the DNA behaves and moves differently with added methylation. One of the specific characteristics able to be measured was the torsion angles of the glycosidic bonds of the guanines in each of the different duplexes. The second characteristic measured was the distances between adjacent base pairs. At first, the distance was measured from the spot on the guanine where adduction occurs, the C_8 position, to the atom directly above or below. However, since correlation is likely between the torsion angle and this distance, another distance was measured between the glycosidic nitrogen of the guanine to the glycosidic nitrogen of the base above or below the guanine. This measurement allowed the ability to compare distances that were not as strongly

correlated to each other and therefore would be a more reliable characteristic to report.

The comparison between the G_5s of the unmethylated and methylated duplexes showed encouraging visual results. A clear difference could be seen between the two bases showing that methylation has a definite effect on the energy minimization. The methylation causes a change in the position of the guanine from the unmethylated strand to the methylated strands that might allow the DNA to be more open and accessible at that point. Differences in both the height of the base compared with the backbone and the twisting of the base were seen when the two guanines were superimposed. These results would presumably lead to the belief that based solely on minimization, differences could be observed that support the previous study. Since there was no dynamics simulation run at that point, the data might not be as significant as it first seemed.

The results of our dynamics simulation show that it is possible to run a molecular dynamics simulation on a dsDNA molecule and obtain data that are reasonable. Our hypothesis that excess methylation will increase the space around a specific base therefore opening it up more and making it more accessible for an adduction is supported by the first part of the data. The first encouraging result was that the DNA did not split apart after 10,000 fs of simulated movement. This shows that the stability provided by the counterions as well as the solvation with water allowed the dynamics simulation to run to completion. Although 10,000 fs is a short time period, the energy of the system

did reach a minimum. This means that the molecules were able to move around in a low energy state which is more representative of reality than if the molecules were at a high energy state or switching rapidly back and forth between high and low with no reason to do so. It was possible to see a definite change in both the distance and the torsion angle when comparing the guanine at the fifth position between the different methylated strands, implying that with methylation, the area around the base opens up more. This could easily lead to more adduction occurring. Statistically, our data were sound as well; therefore, it is believed that this part of the study strongly supports our hypothesis that the methylation will cause changes in the DNA that will make it more accessible to adduction. It was also felt that the molecular modeling technique used was mimicking reality to a reliable degree.

In the second part of the study, the data do not clearly and definitively support the trends that were seen in previous work. Some of the data from the torsion angles were somewhat close to what was expected. The clear trend that was hoped for was not present, and none of the data fits the trends that were seen by Prof. Vouros. Along the same lines, it was difficult to determine if the distance data follow a trend or not. Looking at the total change in area around the guanines did not lead to any trends being seen; but in most cases, there was more space opening above or below the guanines with increased methylation. It was clear that there was an increase in either the distance above or below the guanine in almost all of the cases with excess methylation. When observing the

guanine in the fifth position, it was seen that the distance increased the most at this position. This follows the trends noted by Vouros. The rest of the data do not follow the trends that were expected. The results that most closely follow the trends that were seen in the previous study were those of the duplex that had both of its strands methylated. The trend was followed except for the guanine in the thirteenth position which in our study was observed to have the same change as the guanine in the eleventh position. While the results were along the line of what was expected with the increases in both the torsion angles and distances at each of the guanines when there is methylation, the data did not clearly mimic what was seen in the previous study. Therefore, this is not a technique that as is could be trusted to find "hot spots" for adduction in DNA based on methylation patterns.

One of the main problems that might be causing this is the anomeric effect. Since the conformation of the furanose sugars can easily switch from North to South in computer-based modeling schemes which is not expected in reality, the data might be reflecting this switch if it is occurring on the guanines that were analyzed. When a random sample was taken from close to the end of our dynamics simulation and a cursory glance at the sugars was taken, it was found that several of the nucleotide sugars had switched from the North to the South conformation. This switching of conformations would then have an effect on our results. This is expected since the measured distances contained the glycosidic nitrogen which is the same nitrogen that is involved with the anomeric

effect. The oxygen and nitrogen where the anomeric effect will occur are also the same ones used in measuring the torsion angles. Since these two atoms were so heavily involved with our analyses, any problems that would generate with them would cause our analyses to be skewed as well.

Future studies could be done that would use a different molecular force field with additional parameters that would not allow the pseudorotation of the sugar from North to South conformation to occur as easily, thereby resolving one of the main problems in this study. The results would then be more representative of reality while at the same time giving data that might show the expected trends from the Vouros' study. Another potential problem that was not taken into account was the extent or the effects of the association of water in solvation. The implicit SYBYL command does not align water molecules in the grooves of the DNA; therefore, other unknown problems might have occurred. In future studies, perhaps, it would be possible to use a different solvation command to allow the DNA to be properly solvated with structural water as it is *in vivo*. Also, it might be beneficial to use other molecular force fields such as AMBER to study the DNA molecular dynamics since these force fields were created to deal specifically with biological macromolecules. Another future study might be a docking experiment with the BPDE adduct at each of the guanine positions to measure the affinity and see if that changes with increases in methylation.

REFERENCES

[1] B. Singer, B. Hang. Nucleic acid sequence and repair: role of adduct, neighbor bases, and enzyme specificity. *Carcinogenesis* **2000**, 21, 6, 1071-1078.

[2] X. Tan, N. Suzuki, A. P. Grollman, S. Shibutani. Mutagenic Events in *Escherichia coli* and Mammalian Cells Generated in Response to Acetylaminofluorene-Derived DNA Adducts Positioned in the *Nar* I Restriction Enzyme Site. *Biochemistry* **2002,** 41*,* 14255-14262.

[3] T. Bestor, J. R. Edwards, J. Ju, X. Li. Universal Methylation Profiling Methods. IPN WO 2010/011312 A9. **2010**.

[4] R. A. Rollins, F. Haghighi, J. R. Edwards, et al. Large-scale structure of genomic methylation patterns. *Genome Res.* **2006**, 16, 157-163.

[5] S. Yan, R. Shapiro, N. E. Geacintov, and S. Broyde. Stereochemical, Structural, and Thermodynamic Origins of Stability Differences between Stereoisomeric Benzo[*a*]pyrene Diol Epoxide Deoxyadenosine Adducts in a DNA Mutational Hot Spot Sequence. *J. Am. Chem. Soc.* **2001,** 123*,* 7054-7066.

[6] M. F. Denissenko, J. X. Chen, M. Tang, and G. P. Pfeifer. Cytosine methylation determines hot spots of DNA damage in the human *P53* gene. *Proc. Natl. Acad. Sci. USA* **1997,** 94, 3893–3898, Genetics.

[7] A. Kolbanovskiy et al. Base Selectivity and Effects of Sequence and DNA Secondary Structure on the Formation of Covalent Adducts Derived from the Equine Estrogen Metabolite 4-Hydroxyequilenin. *Chem. Res. Toxicol.* **2005,** 18*,* 1737-1747.

[8] J. Glick, W. Xiong, Y. Lin, A. M. Noronha, C. J.Wilds, P. Vouros. The influence of cytosine methylation on the chemoselectivity of benzo[a]pyrene diol epoxide-oligonucleotide adducts determined using nanoLC/MS/MS. *J. Mass. Spectrom*. **2009**, 44, 1241–1248.

[9] S. C. Cheng, B. D. Hilton, J. M. Roman, A. Dipple. DNA adducts from carcinogenic and noncarcinogenic enantiomers of benzo[a]pyrene dihydrodiol epoxide. *Chemical Research in Toxicology* **1989**, *2*, 334.

[10] T. E. Cheatham, III, M. A. Young. Molecular Dynamics Simulation of Nucleic Acids: Successes, Limitations, and Promise. *Biopolymers (Nucleic Acid Sciences)*, **2001,** 56, 232–256.

[11] A. Naome´, P. Schyman, A. Laaksonen, D. P. Vercauteren. Molecular Dynamics Simulation of 8-oxoguanine Containing DNA Fragments Reveals Altered Hydration and Ion Binding Patterns. *J. Phys. Chem. B* **2010,** 114*,* 4789–4801.

[12] T. E. Cheatham, III, P. A. Kollman. *J Am Chem Soc* **1997**, 119, 4805–4825.

[13] T. E. Cheatham, III. Simulation and modeling of nucleic acid structure, dynamics and interactions. *Current Opinion in Structural Biology* **2004**, 14, 360–367.

[14] R. K. Jalluri, Y. H. Yuh, and E. W. Taylor. O-C-N Anomeric Effect in Nucleosides. The Anomeric Effect and Associated Stereoelectronics Effects. American Chemical Society. **1993**.

[15] Zacharias, M. *Biophys J* **2001**, 80, 2350–2363.

[16] Gilson, M. K.; Davis, M. E.; Luty, B. A.; McCammon, J. A. *J Phys Chem* **1993**, 97, 3591–3600.

[17] T. A. Halgren. MMFF VI. MMFF94s Option for Energy Minimization Studies. *J Comput Chem* **1999,** 20: 720-729.

[18] T. A. Halgren. MMFF VII. Characterization of MMFF94, MMFF94s, and Other Widely Available Force Fields for Conformational Energies and for Intermolecular-Interaction Energies and Geometries. *J Comput Chem* **1999** 20: 730-748.

[19] S. A. Harris. Modelling the biomechanical properties of DNA using computer simulation. *Phil. Trans. R. Soc. A* **2006** 364, 3319–3334.

[20] M. A. Young, G. Ravishanker, D. L. Beveridge. A 5-Nanosecond Molecular Dynamics Trajectory for B-DNA: Analysis of Structure, Motions, and Solvation. *Biophysical Journal* **1997,** 73, 2313-2336.

[21] T. G. Gantchev and D. J. Hunting. Probing the interactions of the solvated electron with DNA by molecular dynamics simulations: bromodeoxyuridine substituted DNA. *J Mol Model* **2008,** 14, 451–464.

APPENDIX A. SYBYL SCRIPT

```
UIMS DEFINE MACRO ADD_COUNTERIONS SYBYLBIOPOLYMER YES
#
# add_counterions.spl
# Usage: BIO ADD_COUNTERIONS a_type distance
#
# Description: Adds counterions to phosphate groups (usually DNA and RNA)
# opposite the charged oxygens.  Works on the default area.
#
#
# CWA 5-3-91
#
# CWA 10-4-91 the original procedure was augmented with the %midpoint and
# %unitvec expression generators created by Malcolm Cline specifically for
# this project
#
#
# this prompt determines which counterion is to be added
setvar ci %promptif("$1" "string" "Na" "enter counterion symbol" \
        "an atom type like Li, Na, K, Ca, or Al")
#
```

```
# this prompt determines how far (in angstroms) the counterion is to be
# placed from each phosphate atom
setvar distance %promptif("$2" "REAL" "6.0" \
   "enter P...COUNTERION distance" "in angstroms")
#
# this next determines atom numbers for all phosphorus atoms and extracts
# coordinates
for i in %atoms(M1((<P.3>)))
#    echo $i
   setvar PX %atom_info($i X)
   setvar PY %atom_info($i Y)
   setvar PZ %atom_info($i Z)
   setvar count "0"
# the next two lines determine the atom numbers of the two anionic oxygens
# (atom type O.co2 in version 5.5, modify to O.2 in 5.4 and earlier) on each
# phosphorus
### Note added by MAC:  you could modify this part of the macro to correctly
### orient the counterion for terminal phosphates (i.e. along the vector
### from the phosphate through the midpoint of the plane of the THREE
       oxygens)
   for j in %atom_info($i NEIGHBORS)
      if %streql(%atom_info($j TYPE) "O.co2")
```

```
            setvar count %math($count + 1)
#           echo $count
#           echo $j
#           echo %atom_info($j TYPE)
# the next three lines create atom labels and extracts coordinates for the
# oxygens
            setvar %cat("O" $count "X") %atom_info($j X)
            setvar %cat("O" $count "Y") %atom_info($j Y)
            setvar %cat("O" $count "Z") %atom_info($j Z)
#           echo %atom_info($j Z)
        endif
    endfor
#   echo $PX
#   echo $PY
#   echo $PZ
#   echo $O1X
#   echo $O1Y
#   echo $O1Z
#   echo $O2X
#   echo $O2Y
#   echo $O2Z
# this section calculates the midpoint (avX, avY, avZ) of the line between
```

```
# the two O.2 oxygens of every phosphate
   setvar midpoint %midpoint($O1X $O1Y $O1Z $O2X $O2Y $O2Z)
# this next section calculates the normal vector (delta values for X,Y,Z)
# for the vector between the P atom and the midpoint between the two O.2
      oxygens
   setvar normal %unitvec($PX $PY $PZ $midpoint)
# here the delta values for X,Y,Z are multiplied by $distance to position
# the counterion $distance angstroms (unit vectors) from the phosphate atom.
# 6.0 is the default value of $distance from the prompt above.
   setvar delX %math(%arg(1 $normal) * $distance)
   setvar delY %math(%arg(2 $normal) * $distance)
   setvar delZ %math(%arg(3 $normal) * $distance)
# the coordinates for the new counterion can now be calculated
   setvar NAX %math($delX + $PX)
   setvar NAY %math($delY + $PY)
   setvar NAZ %math($delZ + $PZ)
   echo "atom" $ci "added at coordinates" $NAX, $NAY, $NAZ
# the new counterions are now added to the molecule in M1
### NOTE added by MAC:  the name of the atom, "SOD" which reflects the
### choice of sodium above, could be replaced by a SWITCH statement
### which named the atom appropriately for the choice of counterion.
   ADD RAWATOM M1 SOD $ci $NAX $NAY $NAZ
```

endfor

.

APPENDIX B. DATA OF PRELIMINARY WORK

Table 7. Maximum distances from guanine to opposite cytosine.

	G_5	G_7	G_{11}
Unmethylated	2.2 Å	2.3Å	2.1 Å
Methylated C2	2.1 Å	2.15Å	6.0 Å
Methylated C3	2.2 Å	3.8 Å	2.1 Å
Methylated C4	2.25 Å	2.4 Å	2.0 Å
Methylated C6	2.1 Å	2.05 Å	2.1 Å
Methylated C9	2.8 Å	2.1 Å	2.5 Å
Methylated C10	2.3 Å	2.4 Å	2.4 Å
Methylated C12	2.3 Å	2.1 Å	2.2 Å
Methylated C14	3.0 Å	2.1 Å	2.3 Å

Table 8. Maximum distances from different positions on the guanine in Å.

	G5a	G5b	G7a	G7b	G11a	G11b
Unmethylated	2.4	2.2	2.8	2.1	2.6	2.3
Methylated C2	2.1	2.5	2.2	2.2	6.5	6.5
Methylated C3	2.4	2.1	5.0	3.0	2.4	2.3
Methylated C4	2.4	2.3	3.4	2.4	2.1	2.2
Methylated C6	3.0	2.2	2.3	2.2	2.6	2.1
Methylated C9	2.5	2.6	2.4	2.2	2.6	2.1
Methylated C10	2.8	2.2	3.2	2.1	2.5	2.2
Methylated C12	2.4	2.4	2.8	2.2	2.6	2.1
Methylated C14	4.8	2.8	2.2	2.2	3.8	3.6

APPENDIX C. DATA FROM PART I

Table 9. Torsion angle data and statistical information.

	Mean	SD	SEM	N	90% CI	95% CI	99% CI
Unmethylated	39.46º	22.27º	1.571º	201	36.86º - 42.06º	36.36º - 42.56º	35.37º - 43.55º
Methylated strand 1	61.65º	12.09º	0.8528º	201	60.24º - 63.06º	59.97º - 63.33º	59.43º - 63.87º
Methylated strand 2	59.92º	11.97º	0.8443º	201	58.52º - 61.32º	58.26º - 61.58º	57.72º - 62.12º
Methylated strand 1 & 2	64.15º	14.89º	1.050º	201	62.41º - 65.89º	62.08º - 66.22º	61.42º - 66.88º

Table 10. T-test and statistical significance of torsion angle data compared to unmethylated DNA.

	t	degrees of freedom	standard error of difference	P value
Methylated strand 1	12.4150	400	1.787	< 0.0001
Methylated strand 2	11.4729	400	1.783	< 0.0001
Methylated strand 1 & 2	13.0665	400	1.890	< 0.0001

Table 11. Distance data and statistical information in Å.

	Mean	SD	SEM	N	90% CI	95% CI	99% CI
Unmethylated	3.78	0.23	0.0162	201	3.75 - 3.80	3.75 - 3.81	3.73 - 3.82
Methylated strand 1	4.12	0.25	0.0176	201	4.09 - 4.15	4.09 - 4.15	4.07 - 4.17
Methylated strand 2	4.04	0.31	0.0219	201	4.00 - 4.08	3.99 - 4.08	3.98 - 4.09
Methylated strand 1 & 2	4.72	0.40	0.0282	201	4.67 - 4.76	4.66 - 4.77	4.64 - 4.79

Table 12. T-test and statistical significance of distance data compared to unmethylated DNA

	t	degrees of freedom	standard error of difference	P value
Methylated strand 1	14.1897	400	0.024	< 0.0001
Methylated strand 2	9.5494	400	0.027	< 0.0001
Methylated strand 1 & 2	28.8827	400	0.033	< 0.0001

APPENDIX D. GRAPHS

Figure 10. Torsion angle of the unmethylated DNA duplex vs. time. The average torsion angle was 39.46º and the standard deviation was 22.27º.

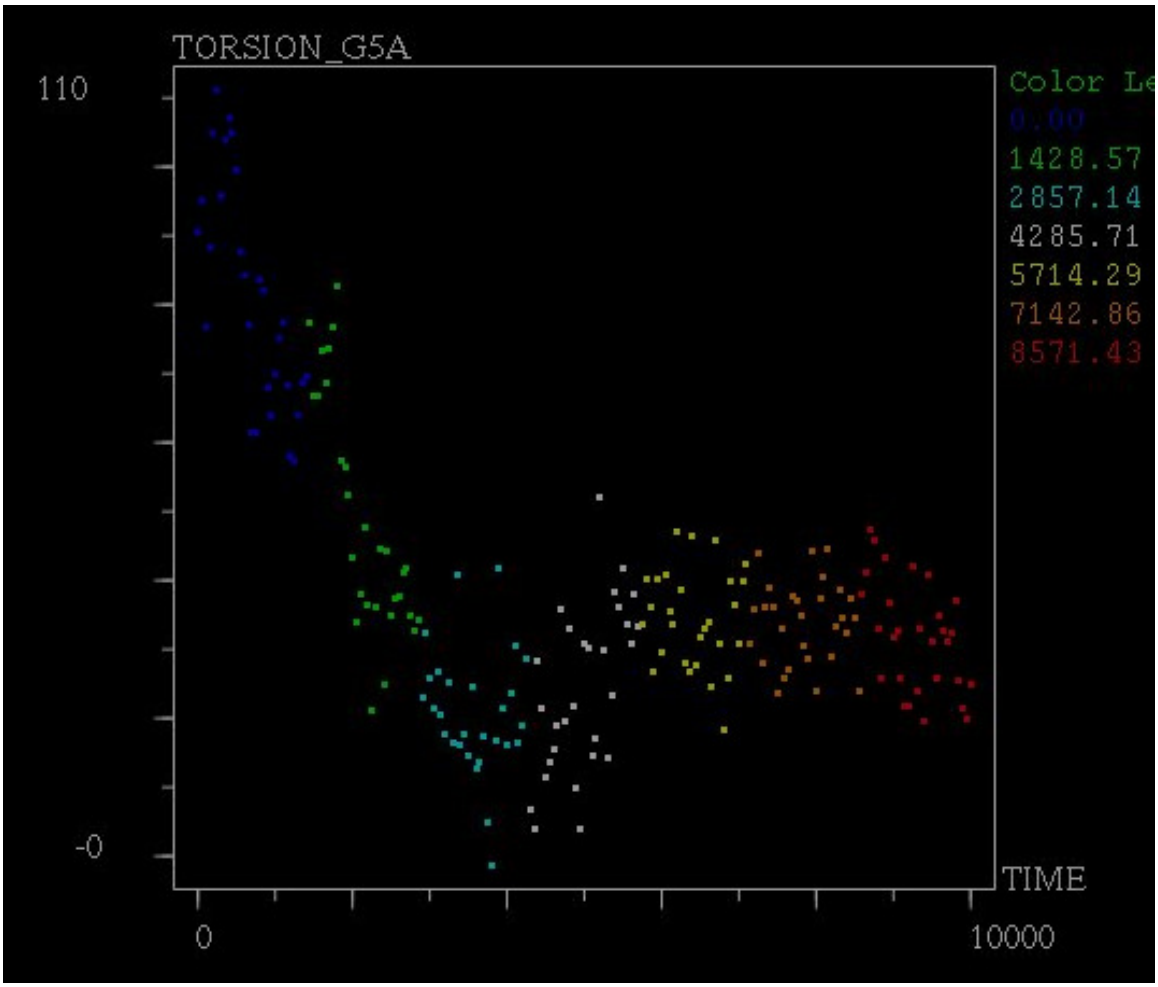


Figure 11. Torsion angle of the Methylated Strand 1 DNA duplex vs. time. The average torsion angle was 61.65º and the standard deviation was 12.09º.

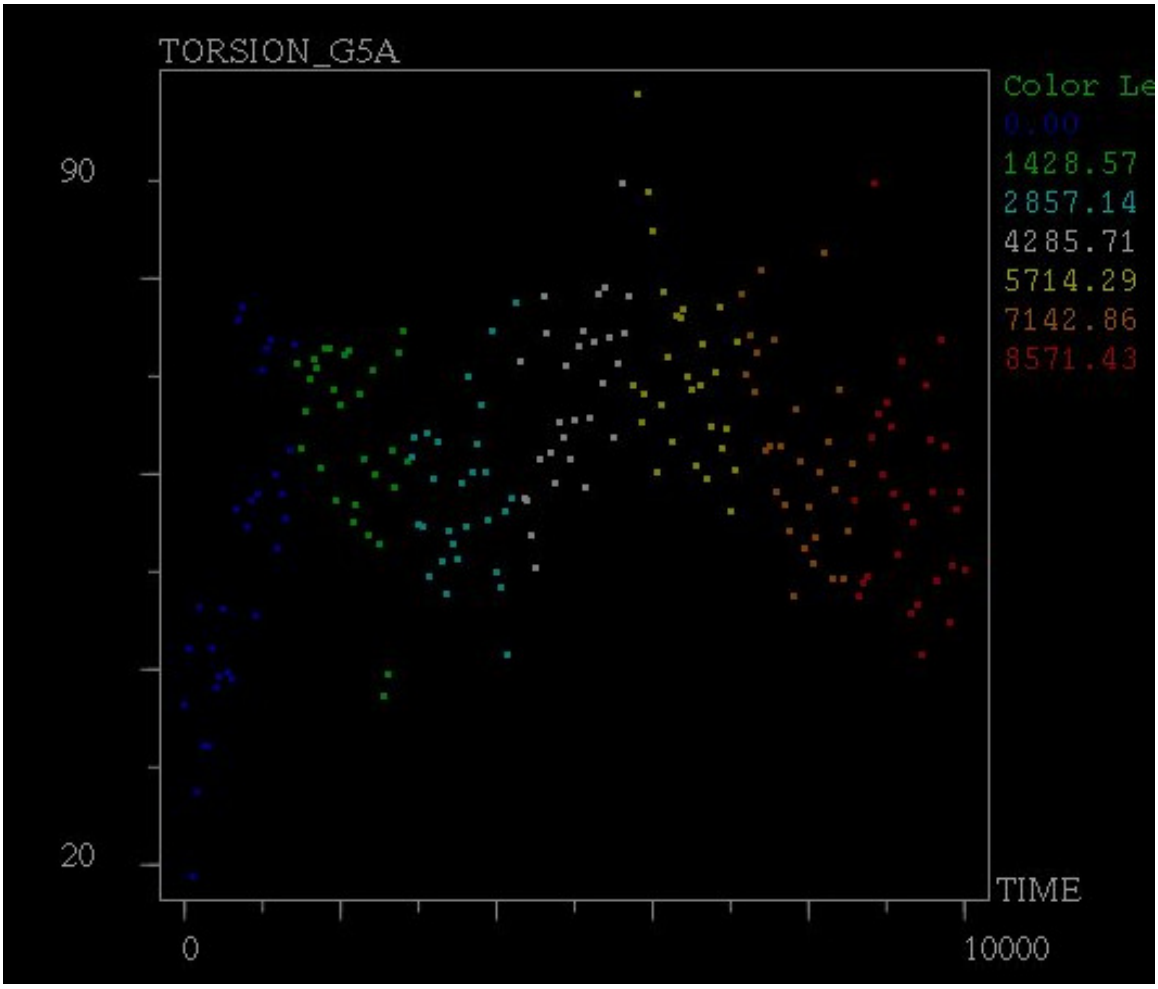


Figure 12. Torsion angle of the Methylated Strand 2 DNA duplex vs. time. The average torsion angle was 59.92º and the standard deviation was 11.97º.

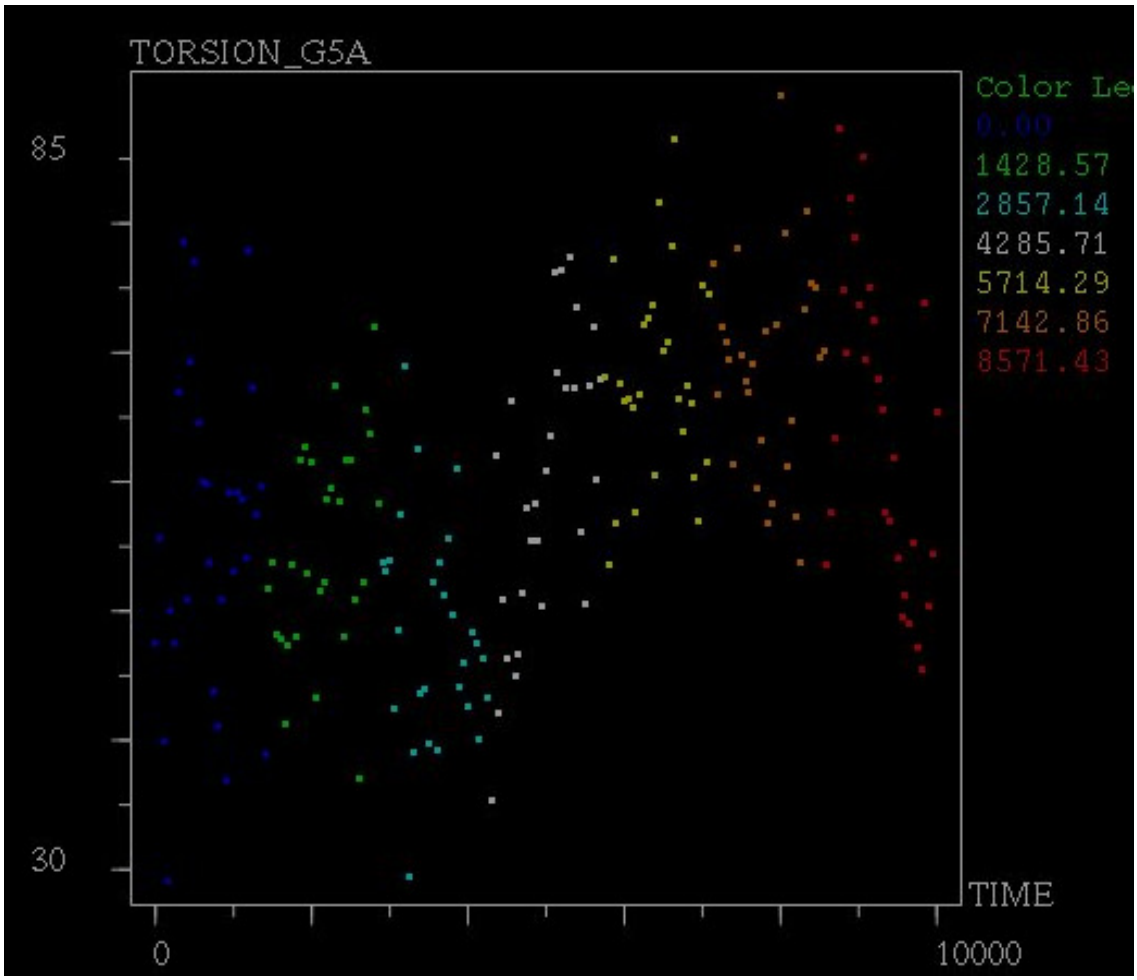


Figure 13. Torsion angle of the Methylated Strands 1 and 2 DNA duplex vs. time. The average torsion angle was 64.15º and the standard deviation was 14.89º.

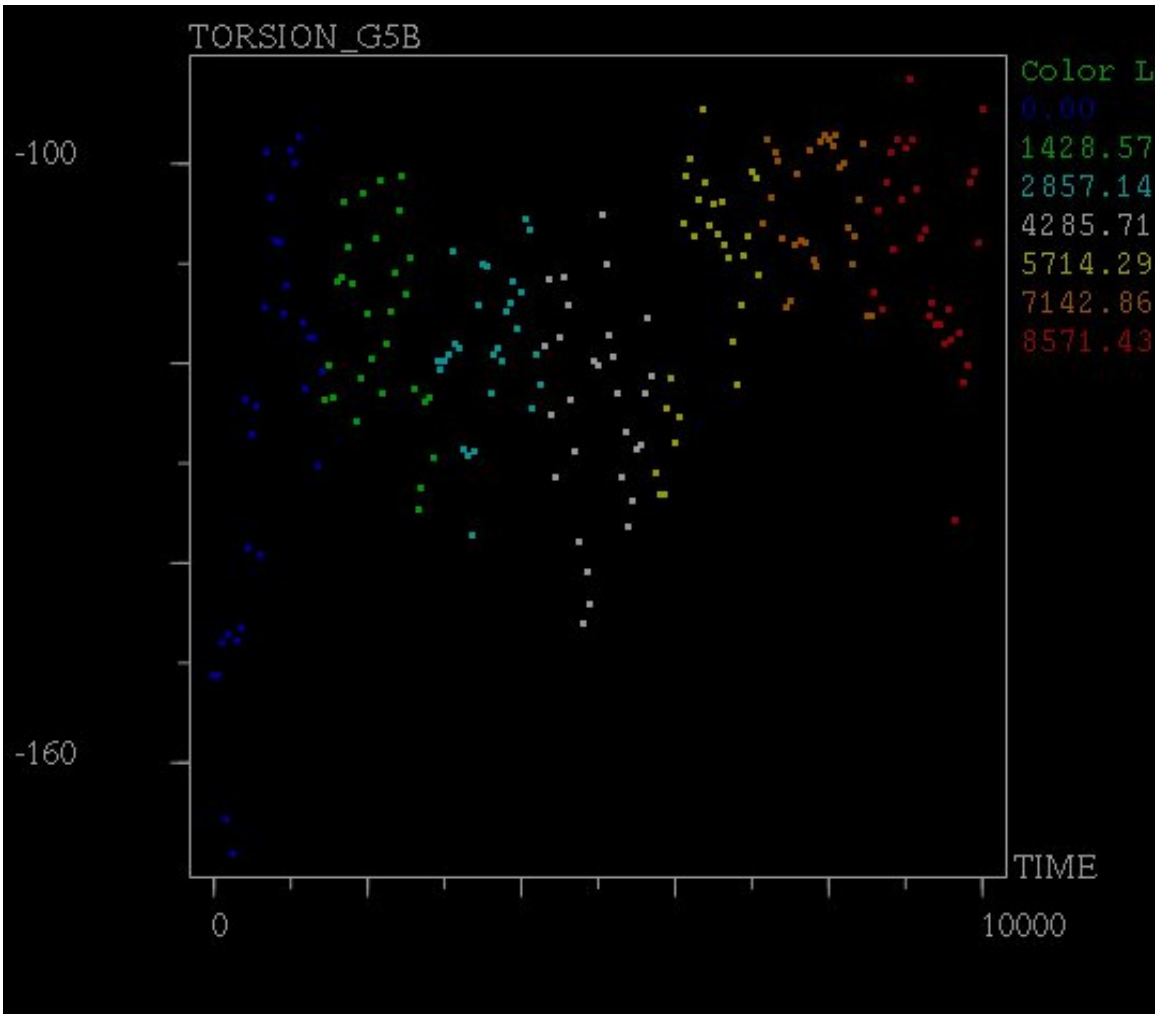


Figure 14. Distance between G_5 and G_6 bases vs. time for the unmethylated DNA duplex. The average distance was 3.78 Å with a standard deviation of 0.23 Å.

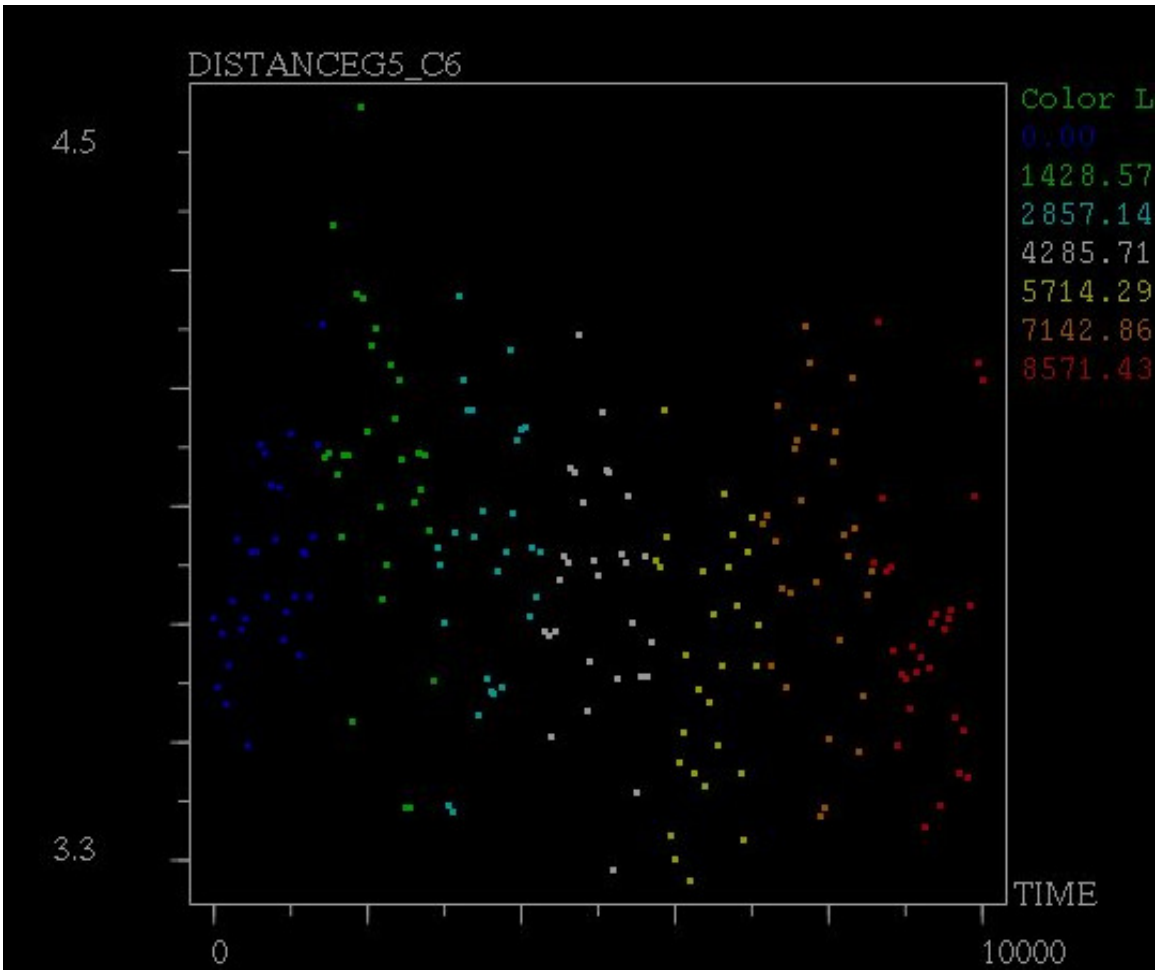


Figure 15. Distance between G_5 and G_6 bases vs. time for the methylated strand 1 DNA duplex. The average distance was 4.12 Å with a standard deviation of 0.25 Å.

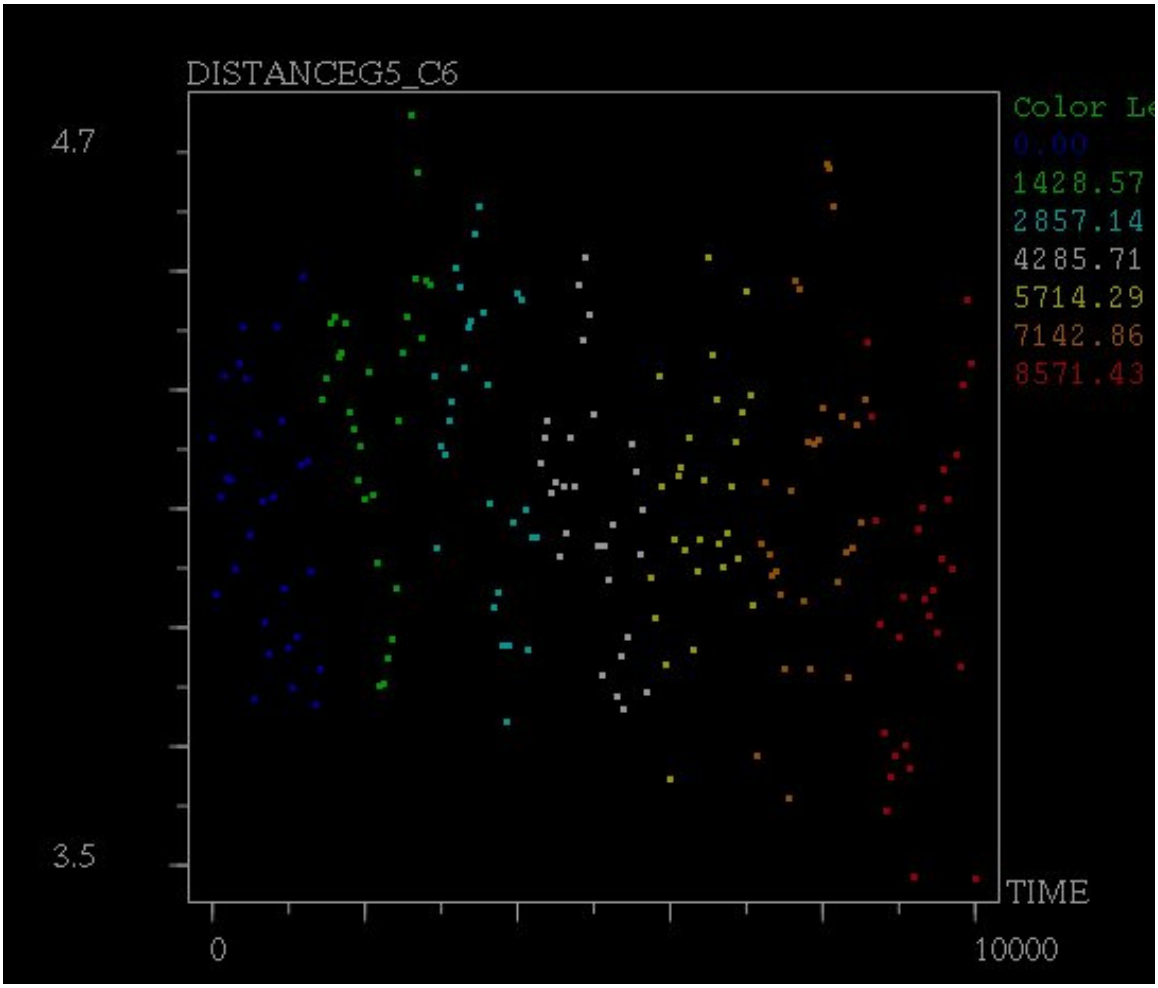


Figure 16. Distance between G_5 and G_6 bases vs. time for the methylated strand 2 DNA duplex. The average distance was 4.04 Å with a standard deviation of 0.31 Å.

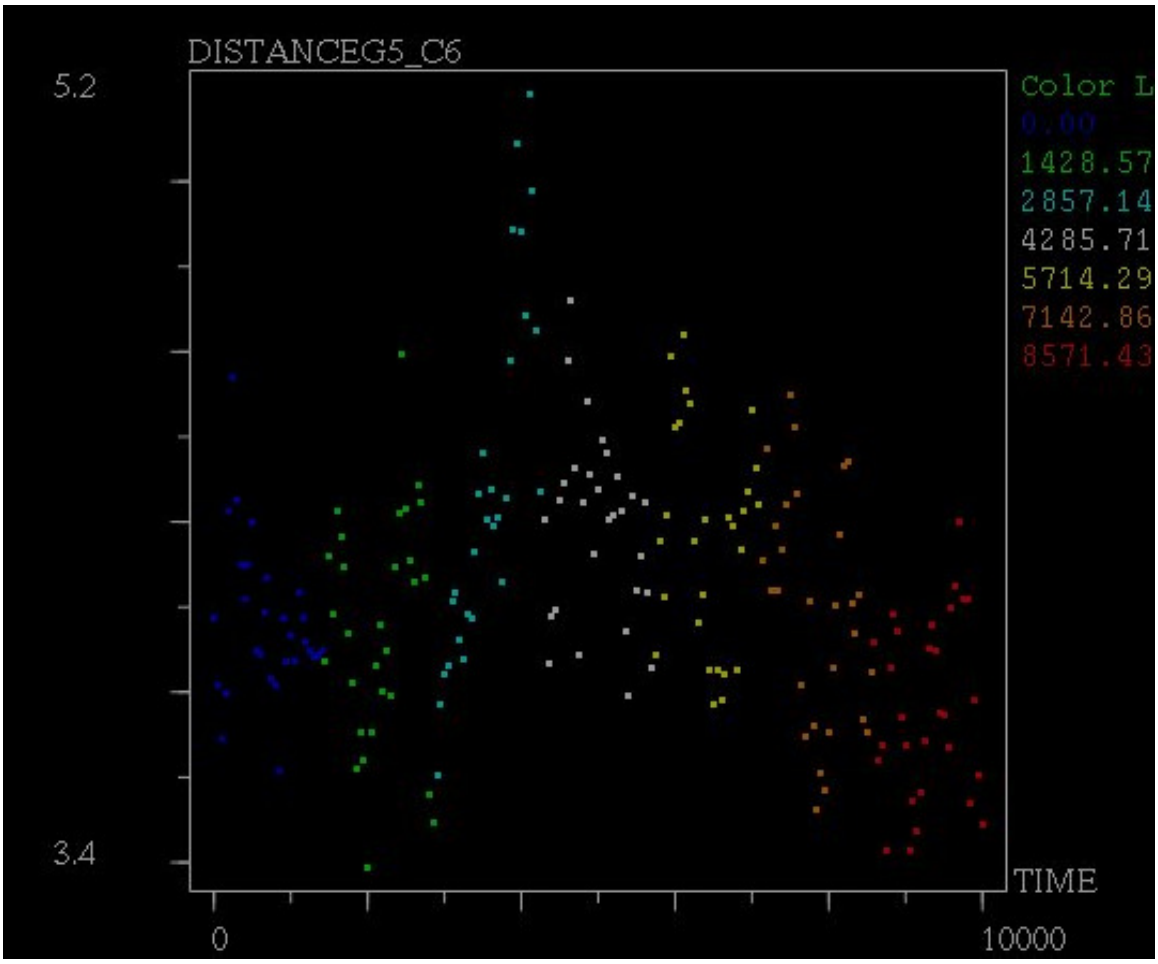
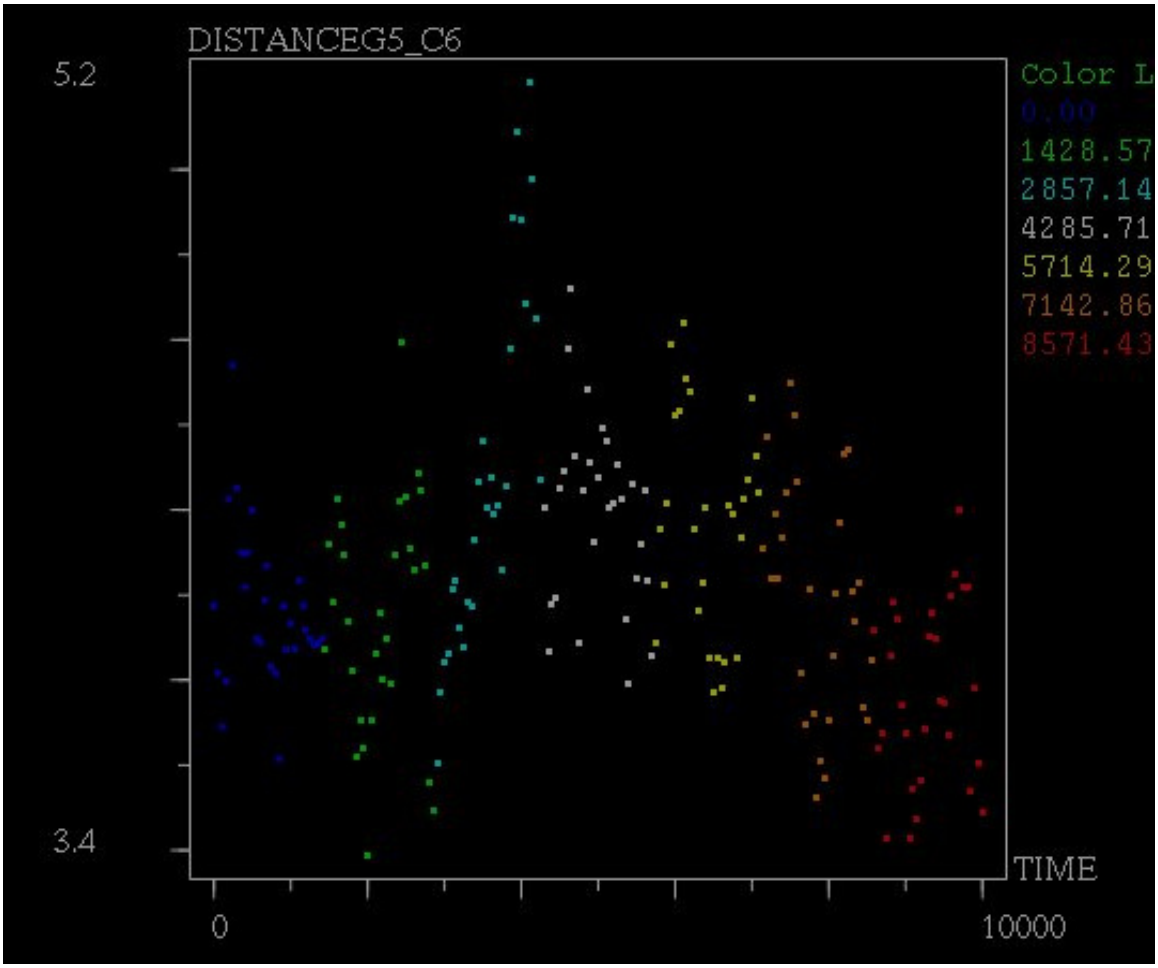


Figure 17. Distance between G_5 and G_6 bases vs. time for the methylated strand 1 and 2 DNA duplex. The average distance was 4.72 Å with a standard deviation of 0.40 Å.



APPENDIX E. DATA FROM PART II

Table 13.Torsion angle data of glycosidic bond for G_5.

	Mean	Standard Deviation
Unmethylated	39.46º	22.27º
Methylated Strand 1	61.65º	12.09º
Methylated Strand 2	59.92º	11.97º
Methylated Strand 1 & 2	64.95º	14.89º

Table 14. Torsion angle data of glycosidic bond for G_7.

	Mean	Standard Deviation
Unmethylated	67.80º	11.85º
Methylated Strand 1	46.13º	17.80º
Methylated Strand 2	24.89º	18.17º
Methylated Strand 1 & 2	78.77º	11.63º

Table 15. Torsion angle data of glycosidic bond for G_{11}.

	Mean	Standard Deviation
Unmethylated	84.14º	12.67º
Methylated Strand 1	80.05º	16.99º
Methylated Strand 2	58.18º	18.28º
Methylated Strand 1 & 2	49.65º	18.46º

Table 16. Torsion angle data of glycosidic bond for G_{13}.

	Mean	Standard Deviation
Unmethylated	76.49º	10.09º
Methylated Strand 1	79.68º	10.88º
Methylated Strand 2	79.50º	11.63º
Methylated Strand 1 & 2	72.45º	13.25º

Table 17. Glycosidic nitrogen distances between C_4 and G_5 in Å.

	Mean	Standard Deviation
Unmethylated	4.63	0.23
Methylated Strand 1	5.18	0.44
Methylated Strand 2	5.08	0.28
Methylated Strand 1 & 2	4.50	0.31

Table 18. Glycosidic nitrogen distances between G_5 and C_6 in Å.

	Mean	Standard Deviation
Unmethylated	4.40	0.23
Methylated Strand 1	4.29	0.25
Methylated Strand 2	4.08	0.20
Methylated Strand 1 & 2	4.74	0.32

Table 19. Glycosidic nitrogen distances between C_6 and G_7 in Å.

	Mean	Standard Deviation
Unmethylated	4.89	0.20
Methylated Strand 1	4.70	0.31
Methylated Strand 2	4.77	0.48
Methylated Strand 1 & 2	4.50	0.24

Table 20. Glycosidic nitrogen distances between G_7 and T_8 in Å.

	Mean	Standard Deviation
Unmethylated	4.61	0.22
Methylated Strand 1	4.58	0.22
Methylated Strand 2	5.00	0.25
Methylated Strand 1 & 2	4.92	0.31

Table 21. Glycosidic nitrogen distances between C_{10} and G_{11} in Å.

	Mean	Standard Deviation
Unmethylated	4.47	0.35
Methylated Strand 1	4.40	0.24
Methylated Strand 2	4.97	0.44
Methylated Strand 1 & 2	4.59	0.37

Table 22. Glycosidic nitrogen distances between G_{11} and C_{12} in Å.

	Mean	Standard Deviation
Unmethylated	4.23	0.15
Methylated Strand 1	4.33	0.22
Methylated Strand 2	4.46	0.36
Methylated Strand 1 & 2	4.51	0.33

Table 23. Glycosidic nitrogen distances between C_{12} and G_{13} in Å.

	Mean	Standard Deviation
Unmethylated	4.75	0.22
Methylated Strand 1	4.71	0.19
Methylated Strand 2	5.04	0.35
Methylated Strand 1 & 2	4.69	0.26

Table 24. Glycosidic nitrogen distances between G_{13} and C_{14} in Å.

	Mean	Standard Deviation
Unmethylated	3.92	0.18
Methylated Strand 1	4.13	0.20
Methylated Strand 2	4.24	0.30
Methylated Strand 1 & 2	4.20	0.26