Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Zixing Zheng April 14, 2020

Nanomechanics of D- and I-substituted DNA

by

Zixing Zheng

Laura Finzi

Adviser

Chemistry

Laura Finzi

Adviser

David Dunlap

Committee Member

Jose Soria

Committee Member

Vincent Conticello

Committee Member

2020

Nanomechanics of D- and I-substituted DNA

By

Zixing Zheng

Laura Finzi

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Chemistry

2020

Abstract

Nanomechanics of D- and I-substituted DNA

By Zixing Zheng

Deoxyribonucleic acid (DNA) is the molecule that carries genetic information for almost all living organisms. DNA exist in double-helical structure but there are evidence showing its polymorphism, and its conformation, as well as the physicochemical properties, can be affected by multiple factors. In this study, we focused on the relationship between specific properties of the individual bases, such as the number of hydrogen bonds they can establish in a base pair, and the biochemical characteristics of the whole double-helical structure. By substituting selected DNA sequences with either 2,6-diaminopurine(D) for adenine(A) or inosine(I) for guanosine(G), we investigated the influence of changing the number of hydrogen bonds, either increasing it to three or decreasing it to two throughout a molecule, without sacrificing sequence specificity. We used the melting temperature, CD spectroscopy and AFM imaging to characterize these effects. We found that D-substituted DNA has higher, and I-substituted DNA lower, melting temperature, compared to wild-type(WT) DNA. We found that We found that the hydrogen bonding is not the only factor influencing the stability of the DNA structure. CD spectra confirmed that both D-substituted and I-substituted DNA undergo some conformational change, but that overall, they are in the B-form. Furthermore, D-substituted DNA shows a progressive shift from a less-WT towards a WT-like spectrum with the increase of GC content, while the change is not significant for I-substituted DNA. The AFM images were used to visualize the DNA molecule under no tension. The contour length was recorded by tracing the molecules and from this parameter the axial rises per base pair was calculated. Compared to wild type DNA, D-substitution decreased the axial rise per base pair and I-substituted DNA increased it, which means D-substituted DNA is more compact than I-substituted DNA.

Nanomechanics of D- and I-substituted DNA

By

Zixing Zheng

Laura Finzi

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Chemistry

2020

Acknowledgements

I want to sincerely thank Dr. Finzi and Dr. Dunlap for their invaluable mentorship and support throughout my thesis. I am thankful to Daniel Kovari, Gustavo Borjas, Wenxuan Xu, Yan Yan, Ordy Gnewou, and all the students and faculties for their help and tireless work. I would like to express my gratitude to my committee members, Dr. Soria, Dr. Conticello, Dr. Finzi and Dr. Dunlap, for their encouragement and guidance.

Table of Contents

1. Introduction

Deoxyribonucleic acid (DNA) is a molecule that consists of two polynucleotide chains held together by hydrogen bonds. DNA contains genetic information that encodes the most important information for the characters and functions of almost all of the organisms and many viruses. Nucleotides, with nitrogenous base, a 5-carbon sugar, and phosphate groups, are the structural components of DNA. (1, 2) There are four types of nitrogenous bases commonly found in the DNA: adenine(A), guanine(G), cytosine(C), and thymine(T). The linear sequence of nucleotides composed DNA's primary structure. (1, 3)

The secondary structure of DNA is formed when the nucleotides on one strand pair with nucleotides on the other stand through hydrogen bonds. Figure 1 shows the canonical Watson-Crick base pairs where guanine forms three hydrogen bonds with cytosine and adenine forms two hydrogen bonds with thymine.

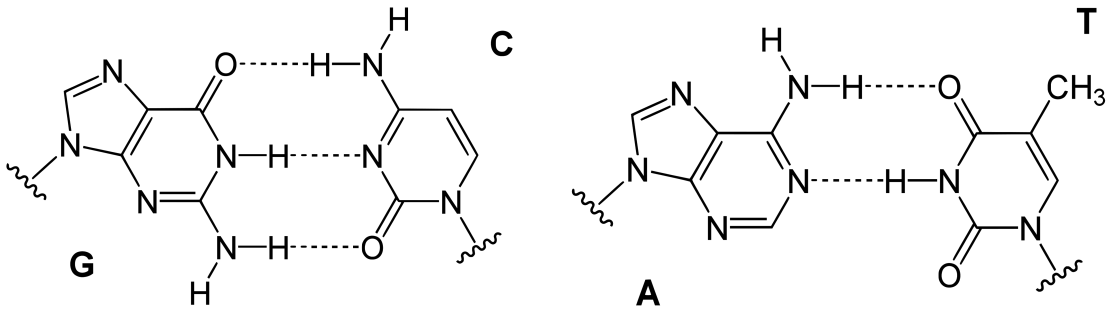

Figure 1. between Watson-Crick base pairs. On the left guanine(G) forms three hydrogen bonds with cytosine(C) and on the right adenine(A) forms two hydrogen bonds with thymine(T) (4)

The secondary structure is a fundamental component in determining DNA tertiary structure, which is the three-dimensional shape of the nucleic acid polymer. (5) The Watson-Crick base pairs confer to DNA double-helical structure, which is one of the most stable tertiary structure of DNA. Non-canonical base pairing can also occur, but it is usually found

in the secondary structure of RNA. (6) Non-canonical base pairing is less stable and may disrupt the double helix when present in the double-stranded DNA. (7)

There are three types of tertiary arrangements for the DNA double helix, namely B-, A-, and Z-DNA. (8) They differ in handedness, length of the helix pitch, number of base pairs per turn, and size of major and minor grooves. In B-DNA the major groove is wider than the minor groove, which leads to the binding of many proteins to B-DNA through the wider major groove. A-DNA and Z-DNA differ significantly from B-DNA. A-DNA and Z-DNA are also biologically active double-helical structures. A-DNA is a right-handed double helix but is shorter and more compact than B-DNA. Z-DNA is a left-handed double helix (L-form). In any of these forms, electrostatic, base stacking, and hydrogen bonding all contribute to the stiffness of the double helix and determine its conformation. Figure 2 shows the top and side views of the A-, B- and Z- DNA.

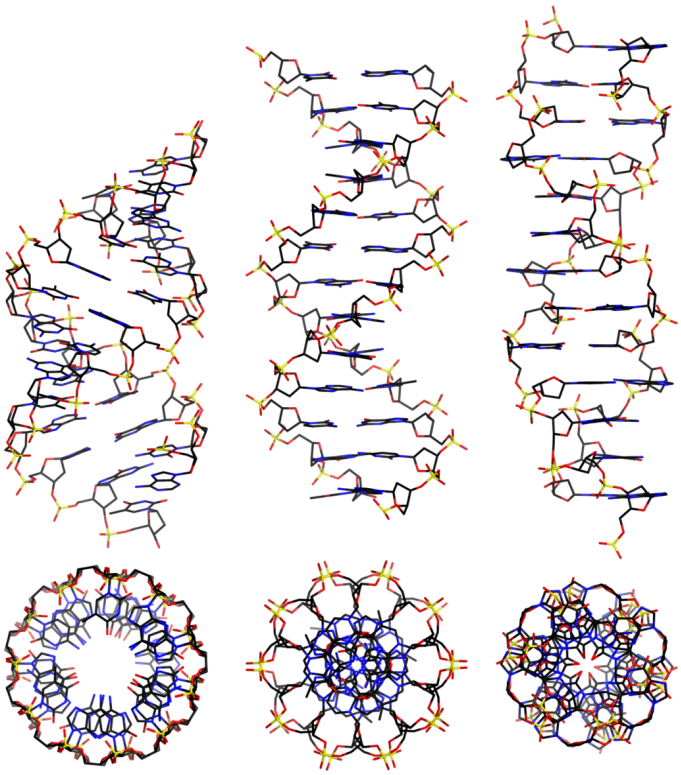

Figure 2. From left to the right is the side and top view of A-, B- and Z-form DNA conformations. (9)

Besides the canonical bases, other nucleoside analogs are also important and useful in many different ways. With distinctive structures, the nucleoside analogs can affect the tertiary structure of the DNA molecule, and therefore affect the biological function of the organism. 2'-deoxyinosine (also called inosine; I) and 2-amino-2'-deoxyadenosine (also called 2,6-diaminopurine; D) are two nucleobase analogs which can be incorporated into the DNA I in place of G and D in place of A. Compared to adenine, D has an additional amino group on the purine molecule and forms an extra hydrogen bond with thymine. In contrast, I has one less amino group compare to guanine and forms one less hydrogen bond with cytosine. The substitution of canonical bases with analogs offers a way to manipulate the physical characteristics of a DNA molecule without effectively changing the sequences. Overall, there is a wide range of applications for analogue nucleotides, including the study of the interaction between DNA and proteins or drug candidates (11-16), the investigation of RNA-related mechanisms (17, 18), and the use as dopants in DNA-based nanoelectronics. (19) Figure 3 shows the structure of 2,6 diaminopurine and inosine in the context of the base pair they form with canonical bases.

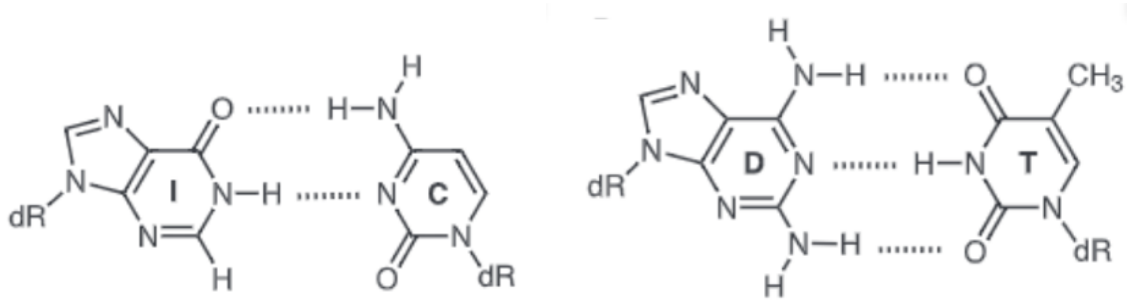

Figure 3. On the left is the structure of inosine forming two hydrogen bonds with cytosine, while on the right is the structure of 2,6 diaminopurine forming three hydrogen bonds with thymine. (35)

Both 2,6 diaminopurine and inosine can be found in nature. 2,6- diaminopurine is found in Cyanophage S-2L, which lyses some strains of unicellular cyanobacteria of the genus Synechococcus. Cyanophase S-2L, has a polyhedral and a flexible noncontractile tail. In S-2L phage DNA, 2,6-diaminopurine takes adenine's place throughout. This base and the corresponding deoxyribonucleoside have been isolated from acid and enzymatic hydrolyzates of the phage DNA, respectively, and have been identified according to optical and chromatographic behavior. (10) There is evidence showing that 2,6 diaminopurine stabilizes the secondary structure of phage DNA. Indeed, the melting temperature of the phage DNA is about 3.6 degrees higher than usual adenine-containing DNA of equivalent base composition. (20) The biological advantage or disadvantages of 2,6 diaminopurine's substitution is not clearly understood. 2,6- diaminopurine can also be useful in the medical field, while there is extensive literature indicating 2,6- diaminopurine strongly inhibits the growth of tumor cells, bacteria, and viruses. The applications may be broad, including treating leukemia. (21-23)

Compare to 2,6- diaminopurine, inosine is more prevalent in organisms and serves critical functions. Inosine has one less amino group compared to guanosine and forms two hydrogen bonds with cytosine. Even though inosine is considered a guanosine analog, it can function as a universal base and can pair with uracil, adenine, and cytosine, which are known as the wobble base pair. The wobble base pair is a pairing between two nucleotide that does not follow Watson-Crick base pair rules found in RNA molecules. With its crucial function in wobble base pairing, inosine plays a fundamental role in RNA secondary structure and the translation of the genetic code. (24) Inosine also has been used in degenerate polymerase chain reactions (PCR) primers, microarray probes, and triplexes, due to its universal pairing capability. (25) While not truly universal, the Inosine-Cytosine pairing is more stable compared to other base pairing inosine can form. (26) Inosine also has great clinical significance. Previous researches indicated that inosine, as a dietary supplement, can improve

athletes' performance, has potential benefits for patients with multiple sclerosis and may be helpful in treating Parkinson's disease, stroke and autoimmune diseases including granulomatosis with polyangiitis. (27-34)

Because of the different number of H-bonds they introduce, 2,6- diaminopurine and inosine can also be used to study the nanomechanics and polymorphism of DNA molecules. D-substituted DNA and I-substituted DNA have very different characteristics and properties from wild-type DNA. Characterizing how D-substitution and I-substitution affect the physical properties of DNA yields insight into the relationship between the specific properties of individual bases and the biochemical characteristics of the whole double-helix assembled with such bases.

At the base pair level, the substitution of 2,6-diaminopurine (D) for adenine (A) or inosine (I) for guanosine (G) is known to alter the DNA conformation in two ways. First, the number of hydrogen bonds is altered. When 2,6-diaminopurine (D) substitutes adenine(A) and base pairs with thymine, the extra amino group allows 2,6-diaminopurine (D) to form a third hydrogen bond with thymine. The additional bond strengthens the hydrogen bonding interaction between the paired backbones and therefore increases the melting temperature. Similarly, when inosine(I) substitutes guanosine, the number of hydrogen bond is decreased compared to the guanine - cytosine base pairing. The I-substituted DNA has only two hydrogen bonds between nucleosides all throughout. This significantly destabilized the DNA molecule, which is expected to have a lower melting temperature. Second, the addition or subtraction of the amino group will affect the DNA groove geometry. For 2,6-diaminopurine (D), the additional amino group can extend into the minor-groove of the B-form DNA helix, decreasing the compressibility of the groove through steric hindrance, reducing bending flexibility and biasing the molecule to deviate from the typical B-form conformation at low tension. (36-40) Contrarily, substitution of inosine for guanosine, which removes the 2-amino

group of the latter from the minor groove, may increase the DNA flexibility and leads to deviation from B-DNA. (41)

We applied three complementary biophysical techniques to measure the mechanical properties of DNA molecules containing 2,6 diaminopurine or inosine. First the melting temperature of D-substituted DNA and I-substituted DNA were measured to provide general insight into the stability of D- or I-substituted DNA.

This data allowed us to compare the compounded contribution of hydrogen bonding, base pair stacking and steric effects to DNA stability. Second, CD spectra is taken for D- and I- substituted DNA, in order to investigate these molecules are conformationally different from canonical DNA. The CD spectra can capture the difference in the chirality of molecule, which reports on the conformation. Then, in order to further investigate the conformational stability of D- and I- substituted DNA and visualize the conformation of each type of DNA under a tension free condition, we performed atomic force microscopy (AFM) imaging. The AFM image captured the shape of the DNA in a tension free condition, and the contour length of the molecular can be tracked.

2. Materials and Methods

2.1 Melting Temperature Characterization

2.1.1 Sample Preparation

Wild-type, D-substituted, and I-substituted DNA sequences were produced with polymerase chain reaction(PCR). Nucleobase analogs 2'-deoxyinosine (also called inosine; I) and 2-amino-2'-deoxyadenosine (also called 2,6- diaminopurine; D) were purchased from Thermofisher(Waltham, MA) and TriLink BioTechnologies (San Diego, CA), respectfully. D-substituted DNA was produced by replacing dATP with 2,6-diaminopurine-5'-triphosphate (dDTP) and I-substituted DNA was produced by replacing dGTP with inosine triphosphate(dITP). For wild-type and D-substituted DNA, three 155, 147 and 156 base-pair

sequences were amplified from the pBR322 plasmid, containing 40%, 54% and 65% GC-content, respectively. For I-substituted DNA, three 155, 142 and 156 base-pair sequences were amplified from pBR322 plasmid, with 40%, 57% and 65% GC-content. Because the I-substituted DNA is less stable than WT and D-substituted DNA, the 147bp DNA with 54% GC content was not successfully made. The 142 bp DNA sequence with 57% GC content was used instead.

The PCR for sild-type and D-substituted DNA was performed in 25µl reaction mixture containing 25 ng DNA template, 0.2 mM each of the appropriate dNTP (with dATP completed substituted by dDTP for the D-substituted DNA), 0.3 µM each of the forward and reverse primers, ThermoPol Buffer (New England Biolabs, Ipswich, MA), 2mM $MgCl_2$ and 5 U/µL Taq. DNA Polymerase (New England Biolabs, Ipswich, MA). Samples were initially heated to 94 °C for 1 min to denature the DNA. Seventy-five amplification cycles were then performed, each cycle consisting of the following segments: denaturation at 94 °C for 20 seconds, template-primer annealing at 43->54 °C for 30 seconds, and polymerization at 72°C for 8 seconds. After seventy-five cycles, samples were heated to 72°C for 5 minutes for final extension and preserved at 4°C. For I-substituted DNA, PCR mixture included 25 ng DNA template, 0.2 mM each of the appropriate dNTP (with dGTP completed substituted by dITP), 0.3 µM each of the forward and reverse primers, ThermoPol Buffer, 100-µg/mL bovine serum albumin (BSA, New England Biolabs, Ipswich, MA) and 5 U/µL Taq. DNA Polymerase. Samples were initially heated to 90 °C for 1 min to denature the DNA. Forty-five amplification cycles were then performed, each cycle consisting of the following segments: denaturation at 90 °C for 1 minute, template-primer annealing at 43°C for 2 minutes, and polymerization at 64°C for five minutes. After the last cycle, samples were heated to 64°C for 10 minutes for final polymerization, cooled to 37°C for 5 minutes and

preserved at 4°C. The primers used to produce DNA sequences are listed in Table S1. The DNA lengths were verified by gel electrophoresis.

2.1.2 Melting Temperature Measurement

The melting temperature was determined from fluorescently detected melting curves. The DNA products obtained as described in the previous section were purified with QIAquick PCR purification kits (Qiagen, Germantown, MD) and eluted in 10mM Tris-HCl. (pH 8.5) The concentration of the PCR product was measured by NanoDrop Lite (Thermo Scientific, Waltham, MA) by UV absorption. The PCR product was diluted to 10 ng/μL in 10 mM Tris-HCl pH 7.4, with 15 mM KCl. Syto-84 was used at a concentration of 1 μM; at that concentration, the Syto-family dyes have been shown to alter melting by no more than 0.6°C. Fluorescent intensity was recorded using a BioRad C1000 qPCR machine over a temperature range of 60-95°C in 0.5°C increments. The melting temperature was indicated by the significant decrease in fluorescent intensity.

2.2 Circular Dichroism Spectroscopy

The same DNA sequences that were used in a melting temperature measurements were also used for circular dichroism assays. The PCR products were diluted to concentrations ranging from 10-30 ng/μL in 10 mM Tris-HCl (pH 7.4) with 150 mM KCl. The CD spectra of the respective samples were measured using a Jasco 1500 810 spectrometer (JASCO Inc., Easton, MD) in a quartz cuvette and blanked against the spectrum of Tris-KCl buffer without DNA. The spectrum of each sample was measured three times at wavelengths 200–320 nm (in 0.2 nm increments), averaged and smoothed with a Savitzky-Golay filter (Mathworks, Natick, MA) of third order using a 31-step (6 nm) window width.

2.3 Atomic Force Microscopy

2.3.1 Sample Preparation

For the comparison of wild-type and D-substituted DNA, a 4642bp-long (4.6kb) DNA fragment was used. The fragment was prepared by PCR with Long Amp T^{M} (New England Biolabs) using the pKLJ12wt plasmid (47) and primers 5'-AGCGTTGGCFCCGATTGCAGAATGAATTT and 5'-TGGGATCGGCCGAAAGGGCAGATTGATAGG. The DNA lengths were verified by gel electrophoresis.

Due to the extreme low melting temperature of I-substituted DNA, it is hard to amplify 4.6kb long DNA with inosine substituted guanine. Therefore, we used a 832bp-long DNA fragment. The fragment was prepared by PCR with Taq. DNA Polymerase, ThermoPol Buffer, 100-µg/mL bovine serum albumin pUC19 plasmid and primers 5'-TAACTACGATACGGGAGG and 5'-CCGCTCATGAGACAATAA. The length of all DNA construct was verified by gel electrophoresis. The primers and plasmids used to produce DNA sequences are listed in Table S1.

2.3.2 AFM Assay

For the 4.6kbp wild-type and D-substituted DNA, nano-scale images were acquired using a Nanowizard II(JPK Instruments, Berlin). WT and D-DNA molecules were diluted in 5nM MgCl buffer to a final concentration between 0.2-0.3 ng/µl. 10ul of each solution was deposited onto freshly cleaved mica substrates and incubated at room temperature for 5-10 min. The samples were then rinsed with 0.22 µm-filtered deionized ultra-pure water and dried with a gentile nitrogen flow. 8*8 µm images were collected using a 0.5-1 Hz scan rate with 2048*2048 pixel resolution. (60)

To morphologically characterize the two samples, we traced the DNA contours using a custom-tracing routine and calculated the contour length, Lo. Only molecules in the range 1.1-1.8 µm surrounding the expected length were included.

For 832bp I-substituted DNA, the AFM images were acquired according to the following procedure. Shortly before the deposition of the sample, a 5 μl droplet of 0.01 μg/ml of poly-L-ornithine (1 kDa MW, Sigma-Aldrich, St. Louis, MO) was deposited onto freshly cleaved mica (Ted Pella, Redding, CA) and incubated for 2 min. The poly-L-ornithine-coated mica was rinsed drop-wise with 400 μl of high performance liquid chromatography grade water and dried with compressed air. 5 μl of the sample solution containing 1 nM, 832 bp DNA amplicon was deposited on the poly-Lornithine-coated mica and incubated for 2 min. This droplet was rinsed with 400 μl of high-performance liquid chromatography grade water and dried gently with compressed air.

Square, 4*4 micron images with 2048 pixels/line were acquired in intermittent contact mode in air using an RFESP-75 probe (Bruker, Camarillo, CA) in a MultiMode VIII scanning probe microscope (Bruker, Santa Barbara, CA). The oscillation amplitude was about 0.56 V and the scan rate was 2 um/s.

3. Results

3.1 Melting Temperature

The melting temperature(Tm) is reached when half of the DNA samples are denatured, which means that there are half double-stranded DNA and half single-stranded DNA. As temperature increases, the double-stranded DNA can unwind and the hydrogen bonds between base pairs will be weakened and can be broken Under Watson-Crick base pairing, guanine forms three hydrogen bonds with cytosine, while adenine forms two hydrogen bonds with thymine. Therefore, DNA fragments with high GC-content are expected to be more stable than DNA with low GC-content, and have a higher melting temperature.

In D-substituted DNA, 2,6-diaminopurine forms three hydrogen bonds with thymine. The additional hydrogen bond is expected to increase the melting temperature, in agreement

with previous experiments. Similarly, in I-substituted DNA, inosine forms only two hydrogen bonds with cytosine. I-DNA is thus expected to have lower melting temperature.

Besides hydrogen bonding, stacking interactions is also an important factor which affects the melting temperature. Base-stacking interaction dominates not only in the duplex overall stability but also significantly contributes into the dependence of the duplex stability on its sequence. (44, 45, 46) Because the stacking interaction between A/T and G/C duplexes produce sequence-dependent effects, two DNA sequence with similar length may result in different melting temperature. Stacking interactions are expected to be different in D-substituted DNA and I-substituted DNA.

In summary, the melting temperature of D-substituted DNA and I-substituted DNA can be affected in three ways: (1) the addition hydrogen bond stabilizes base-pairing for D-substituted DNA and the loss of hydrogen bond destabilized base-pairing for I-substituted DNA, (2) the stacking interaction of adenine and 2,6-diaminopurine; guanine and inosine are likely different, (3) 2,6-diaminopurine's extra amino group and inosine's lack of amino group may alter the conformational stability of B-DNA, which in turn affects stability of DNA double helix.

In order to compare the melting temperature of wild-type DNA, I-substituted DNA and D-substituted DNA, we selected three DNA sequences with various percentages of GC content (40%, 54-57%, and 65%GC), and similar length (147-156 bp). The melting temperature was derived from the fluorescence-based melting curves. In these measurements, the fluorescence intensity indicates the fraction of helical versus single-stranded DNA, and was recorded using a BioRad C1000 qPCR machine over a temperature range of 60-95°C in 0.5°C increments. The dsDNA intercalating dye Syto-84 was used, knowing that the Syto-family dyes has been shown to alter melting by no more than 0.6°C. (60) The fluorescence intensity was collected and the derivative of the intensity was calculated. The melting

temperature of the DNA sequence was indicated by the peak of the derivative curve, which indicates where fluorescence intensity dropped quickly and half of the DNA was double-stranded while one half was single-stranded.

With short DNA sequences, the melting temperature can be predicted with the SantaLucia nearest-neighbor model, which estimates the melting temperature binary units. The model measured the enthalpy and entropy of neighboring nucleosides(NN). With that model, the melting temperature can be calculated using the following formula.

$$T_M = \frac{\Delta H}{\Delta S + R ln \frac{[dsDNA]}{2}} - 273.15 \quad \text{(eqn.1)}$$

Where Tm is the melting temperature, at which 50% of double-stranded DNA molecules will be dissociated, ΔH is the sequence-dependent enthalpy, ΔS is the sequence-dependent entropy, R is the universal gas constant and [dsDNA] is the concentration of double-stranded DNA in solution. Under the SantaLucia nearest-neighbor model, ΔH and ΔS are equal to the sum of empirically determined nearest-neighbor terms and are corrected to account for salt-dependent effects (47, 48).

Although the SantaLucia model is intended for short sequences, we find that it yields predictions within ±0.8°C of our measured melting temperatures (Fig. 2 blue and well within the predicted accuracy of the model (47).

Consistent with our hypothesis, the substitution of 2,6 diaminopurine increases the melting temperature and the substitution of inosine decreases the melting temperature. The melting temperatures of I-substituted DNA, D-substituted DNA, and Wild-type DNA are shown in Figure 4. For WT DNA, the melting temperature various from 75.5°C to 84°C for DNA with 40% GC content to 65% GC content, respectively. The trend is within expectation because a higher percentage of GC content usually associate with higher stability. D-substituted DNA shows the same trend as wild-type DNA. Thus, substituting A with D and inserting a third H-bond does not eliminate the sequence-dependent effect; it does, however,

raise the the melting temperature of each sequence by 6-8 °C. (Figure 4; green) If the hydrogen bonds were the primary factor in determining the melting temperature, and not the sequence, D-substituted DNA should show approximately the same melting temperature as wild-type, contrary to what was observed.

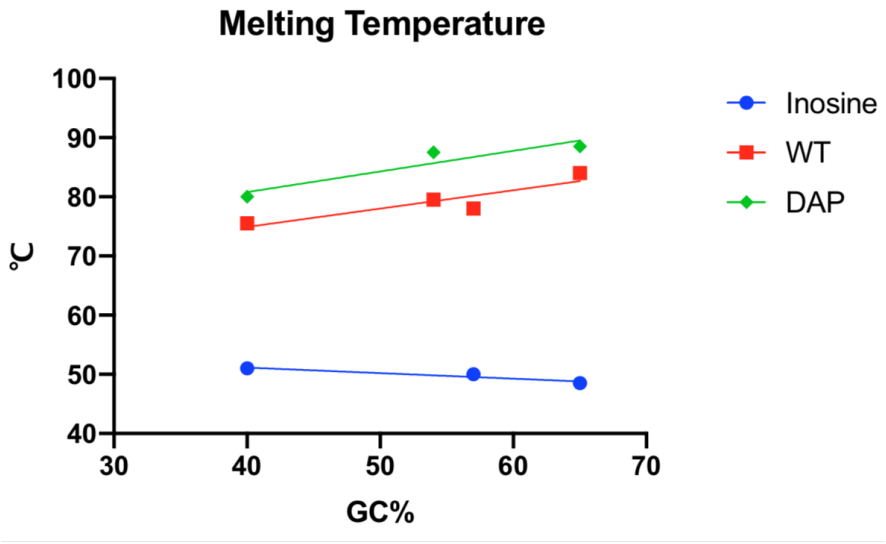

Figure 4. The melting temperature of wild-type, D-substituted and I-substituted DNA sequences with different percentage GC content. The green line indicates the data for D-substituted DNA, red for wild-type and blue for I-substituted DNA.

The melting temperature of I-substituted DNA is significantly lowered. It varies from 51°C to 48.5°C for DNA with 40% GC content to 65% GC content, indicating the significant destabilizing effect of lowering the number of H-bonds. Here, the slope of the line is much flatter, which a minimal decrease as GC content is increased, indicating a much weaker sequence dependence than in wild-type and D-DNA. This result is reasonable because, with more GC content, there are more Inosine substituted in the DNA sequence and is expected to destabilize the DNA and decrease the melting temperature. In light of these results, we speculate that the melting temperature of DNA does not only depend on the number of hydrogen bonds, but also on the sequence. In particular, we conclude that, the GC content

may contribute to the stability of the double helix in other ways besides the presence of an additional hydrogen bon and that the threshold over which the third H-bond adds significant stability to a DNA molecule is below 40% (Figure 4; green and red).

Previously, SantaLucia Model was used to predict the melting temperature of short sequences of wild-type DNA (60). It can also be useful to predict the behavior of D-substituted DNA and I-substituted DNA. Thus, the SantaLucia model was modified to account for the incorporation of non-canonical base. For D-substituted DNA, the enthalpy of each A/T pair is supplemented with an additional factor (α) accounting for both the aggregate effect of the additional hydrogen bond, changes in stacking energy, and perturbations to conformational stability. The alternative model is captured in the following equation.

$$\sum \Delta H = (\Delta H_{GA} + \alpha) + (\Delta H_{AT} + 2\alpha) + (\Delta H_{TC} + \alpha) + \Delta H_{CG} = \sum \Delta H_{WT} + 4\alpha \quad \text{(eqn. 2)}$$

A least-squares fit to the D-substituted DNA melting data yields $\alpha = -0.12 \pm 0.03$ kcal/mol. The modified model does not perfectly capture the melting temperature for each sequence but does preserve the general trend. (60) The similar model is used for I-substituted DNA.

$$\sum \Delta H = (\Delta H_{GA} + \alpha) + (\Delta H_{CG} + 2\alpha) + (\Delta H_{TC} + \alpha) + \Delta H_{AT} = \sum \Delta H_{WT} + 4\alpha \quad \text{(eqn. 3)}$$

For I-substituted DNA melting data, $\alpha = -0.7 \pm 0.14$ kcal/mol, but the model does not capture the experimental data we got. Figure 5 shows the comparison of experimental data and predicted value with SantaLucia's model.

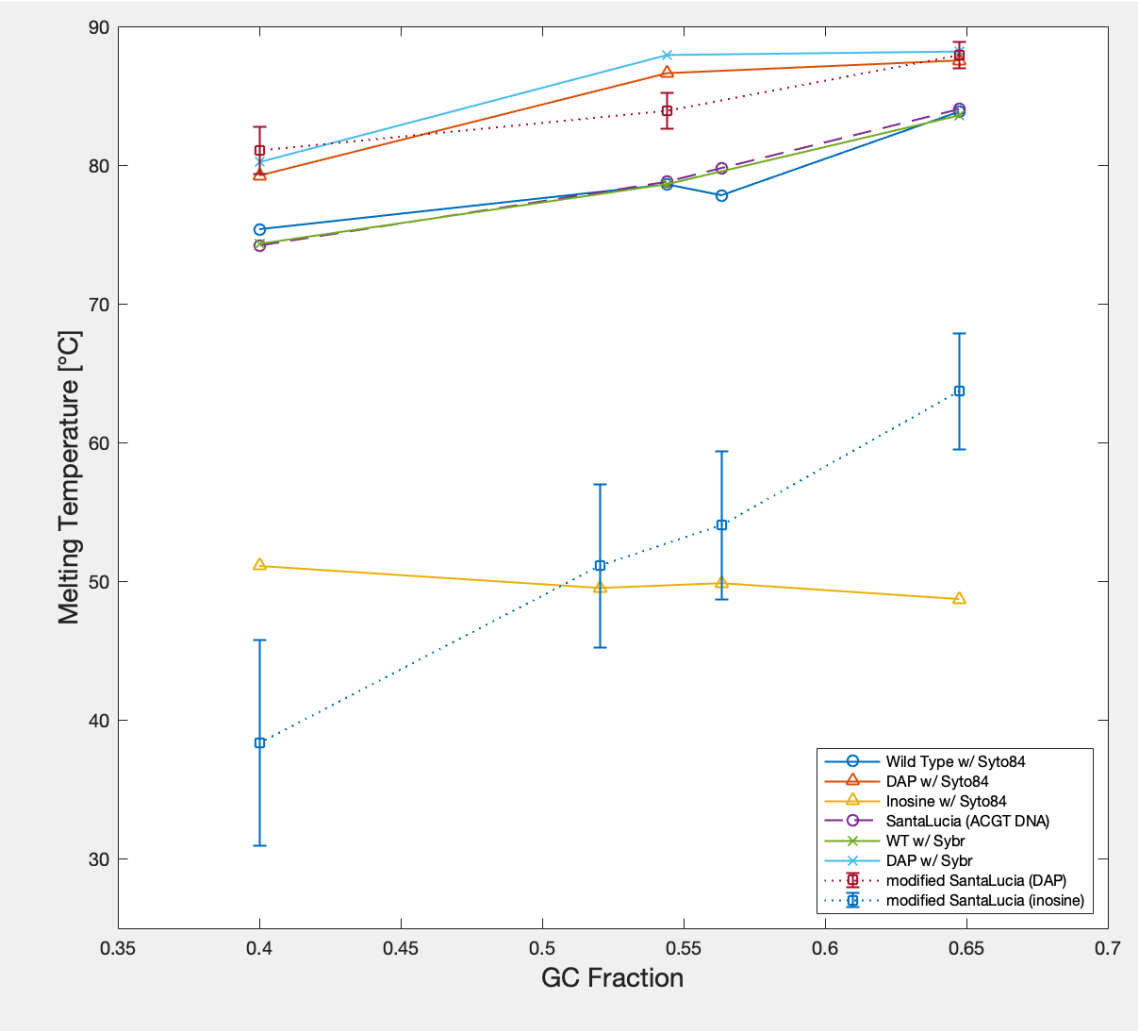

Figure 5. The comparison of experimental data and predicted value with SantaLucia's model. The solid lines are the experimental value, while the dotted lines are the predicted value.

3.2 CD Spectroscopy

As shown in the formula 4, the circular dichroism spectra measure the difference between the absorption of left- and right-handed circularly polarized light by chiral molecules, which characterizes the interactions of chiral molecules with circularly polarized light. (42, 43)

$$\Delta\varepsilon = \varepsilon_L - \varepsilon_L \; [M^{-1}cm^{-1}] \qquad \text{(eqn. 4)}$$

CD spectroscopy can provide structural information about the DNA molecule. There are three biologically active double-helical structures: A-, B-, and Z- DNA. (Figure 2) The A- and B-DNA families are right-handed helices, while the Z-DNA family has a left-handed orientation of the helix. B-form DNA is considered the most common with the caveat that it must be highly dynamic in vivo. It is 23.7 Å wide and extends 34 Å per 10bp of sequence. The double helix makes one complete turn about its axis every 10.4-10.5 base pairs in solution. (49, 50) A-DNA and Z-DNA differ significantly in their geometry and dimensions from B-DNA. A-DNA is commonly the result of dehydrating condition, while methylated DNA may exist in Z-form. Compared to B-DNA, A-DNA has a similar structure but with a shorter and more compact helical structure. Z-DNA has a left-handed double-helical structure and the helix winds to the left in a zigzag pattern.

According to previous studies, A-, B-, and Z- DNA molecules have distinct CD signatures. B-DNA is characterized by a negative peak in the wavelength range of 245-250nm and an approximately equal positive peak between 275 and 280 nm. A-form DNA is characterized by an intense negative peak at 210 nm, an intense positive in 190 nm, and a positive peak at around 260 nm. Z-form DNA usually has a negative band at 290 nm and a positive band at 260 nm.

A systematic study modifying the sequence of the DNA found that the CD spectra of the DNA depend also on the primary sequences. Poly[d(A)] poly[d(T)] adopts the B'heteronomous DNA-form (51, 52), and poly[d(G)].poly[d(C)] has some distinctly A-like features. (53) There are studies that indicate that DNA with poly [d(IC)] sequence exists in a new polynucleotide configuration: a left-handed, eight-fold helix. (54, 55) It remains unclear what features of the DNA alters the helical conformation and to what extent they do so, but we can hypothesize that the incorporation of non-canonical bases will affect the helical conformation and change the mechanical property of the molecule. If hydrogen bonds are the

main factor, the D-substituted DNA will behave similarly compared to the poly[d(G-C)] sequences, while both of them have three hydrogen bonds between base pairs throughout the sequence. And I-substituted DNA will be expected to have a similar conformation as poly[d(A-T)], which have two hydrogen bonds between base pairs throughout the sequence.

The circular dichroism spectra of wild-type, I-substituted, and D-substituted DNA was measured to test the conformation of these molecules using a Jasco 1500 810 spectrometer (JASCO Inc., Easton, MD). For these measurements, we used the same DNA sequences employed for the melting temperature measurement. Thus, three DNA sequences were used for each type of DNA with various percentages of GC content (40%, 54-57%, and 65%GC), and DNA sequences have similar lengths (147-156 bp). The spectra of each sample was measured three times in the range 200-320 nm with 1nm increment; it was then averaged and smoothed with a Savitzky-Golay filter. Figure 6 shows the CD spectrum for the different DNA sequences, where the blue lines represent the wild-type DNA, yellow lines represent the I-substituted DNA and red lines represent the D-substituted DNA.

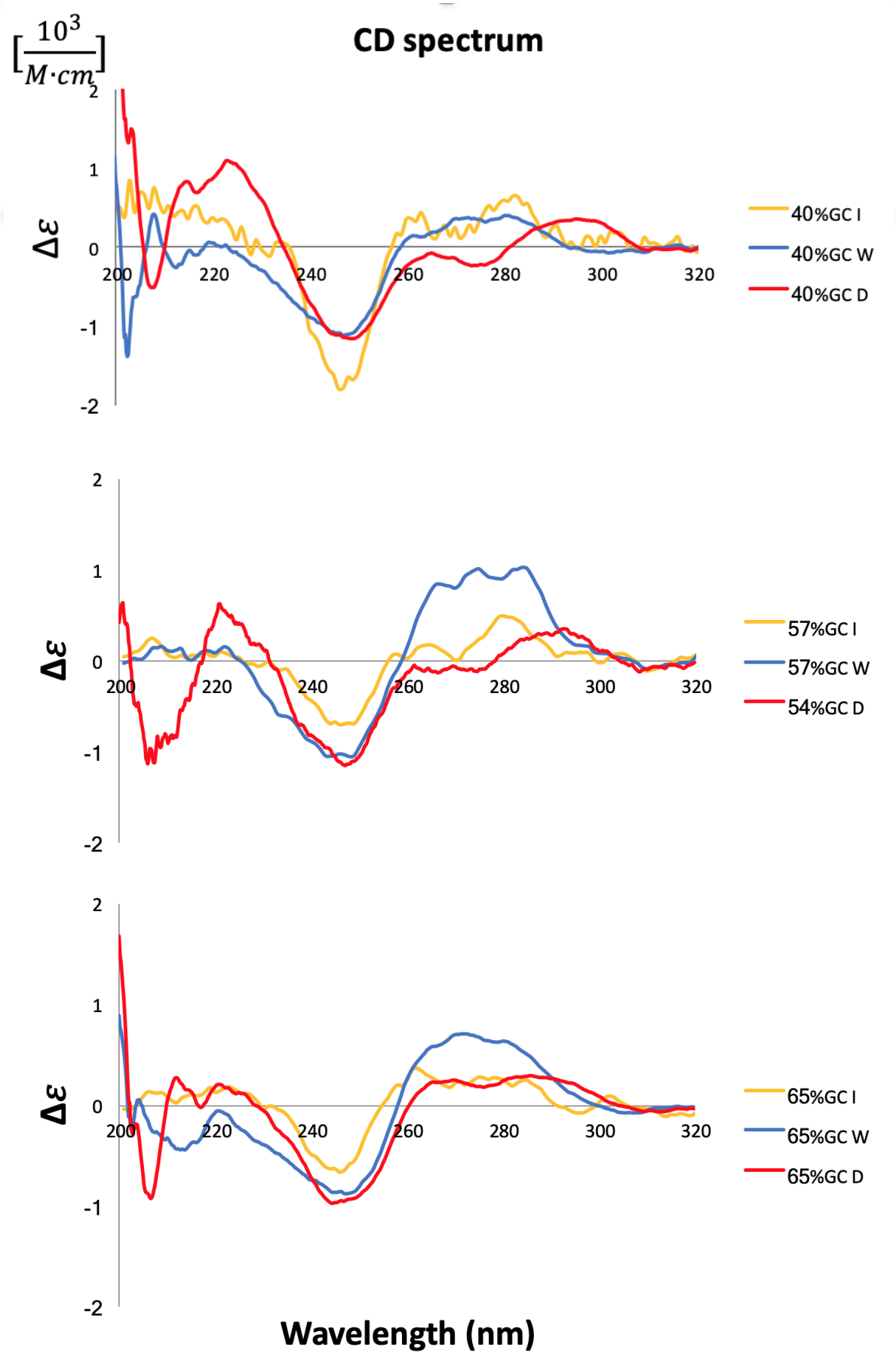
CD spectrum
$[\frac{10^3}{M \cdot cm}]$
Δε
2
1
0
-1
-2
200
220
240
260
280
300
320
40%GC I
40%GC W
40%GC D
57%GC I
57%GC W
54%GC D
65%GC I
65%GC W
65%GC D
Wavelength (nm)

Figure 6. CD spectrum of ~150bp WT, I-subsittuted, and D-substituted DNA. From top to the bottom are DNA sequences with 40%, ~55%, and 65% GC content, respectfully. The blue lines indicate the CD spectrum for wild-type DNA, yellow lines indicate I-substituted DNA and red lines indicate D-substituted DNA.

Due to the characteristics of the equipment, the absorption of the light with wavelength lower than 220 is not reliable and was not used for further analysis. As a result, the intense positive peak and negative peak below 220nm, which can be used to distinguish the conformation of DNA, cannot be observed.

The spectra of all the wild-type molecules show the expected behavior of B-DNA, specifically a local-maximum in the 270-290 nm range and a local minimum around 250 nm. Both I-substituted DNA and D-substituted DNA showed a local minimum around 250 nm, which is identical to the on in the wild-type DNA. However, instead of a wide, positive band from 265-290nm shown by wild-type DNA, both I-substituted DNA and D-substituted DNA showed two peaks separated by a low at around 265-278 nm, the exact position of which varies for different DNA sequences. However, neither D-substituted DNA nor I-substituted DNA had the signatures of A-or Z-DNA. The CD of D-substituted DNA was not similar to that of poly [d(G-C)] sequences, and analogously the CD of I-substituted DNA was not similar to that of poly[d(A-T)] sequences.

For the D-substituted DNA, there is one peak shown at 290-295 nm, one peak shown at 260-265 nm, and a low shown at 274-277 nm. Interestingly, the 40% 50% and 65% GC sequences show a progressive shift from non-WT behavior towards a WT-like spectrum, 40% molecules, which have the most DAP, show a peak at 295nm, a modest peak at 265nm, and a modest local minimum at 274nm. As the percentage of GC content increased, while the 2,6 diaminopurine content decrease, the peak at 295 nm shift slightly to 290 nm and the minimum at 274 fades and two peaks blend, which is more similar to the WT spectrum. The

I-substituted DNAs have similar behavior compared to D-substituted DNAs, which also have two peaks with one low in between, but the locations of the peaks are slighted shifted. For all of the three I-substituted DNA, there is one peak observed at around 263nm and one observed at 280-285 nm, while the modest local minimum is at around 270 nm. Contrary to the D-substituted DNA, the relationship between the shift in the position of the peaks and the percentage of Inosine content was not significant. Figure 7-9 compare the difference within each type of DNA whiling changing the percentage of GC content.

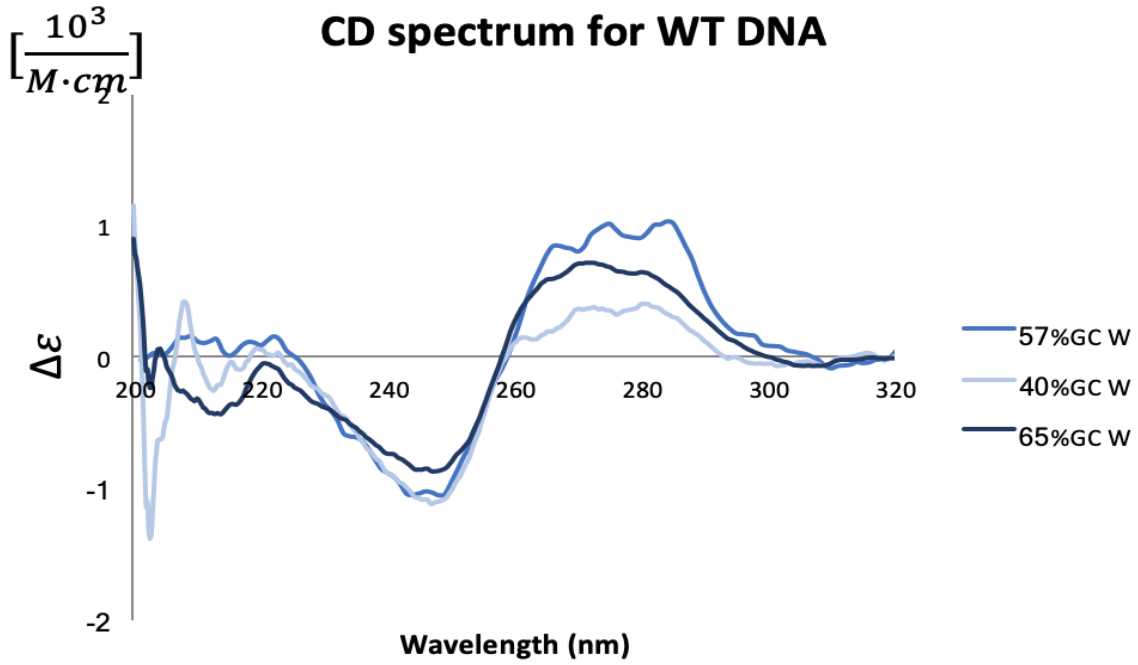

Figure 7. CD spectrum for WT DNA. Darker blue represents higher percentage of GC content in the DNA sequences.

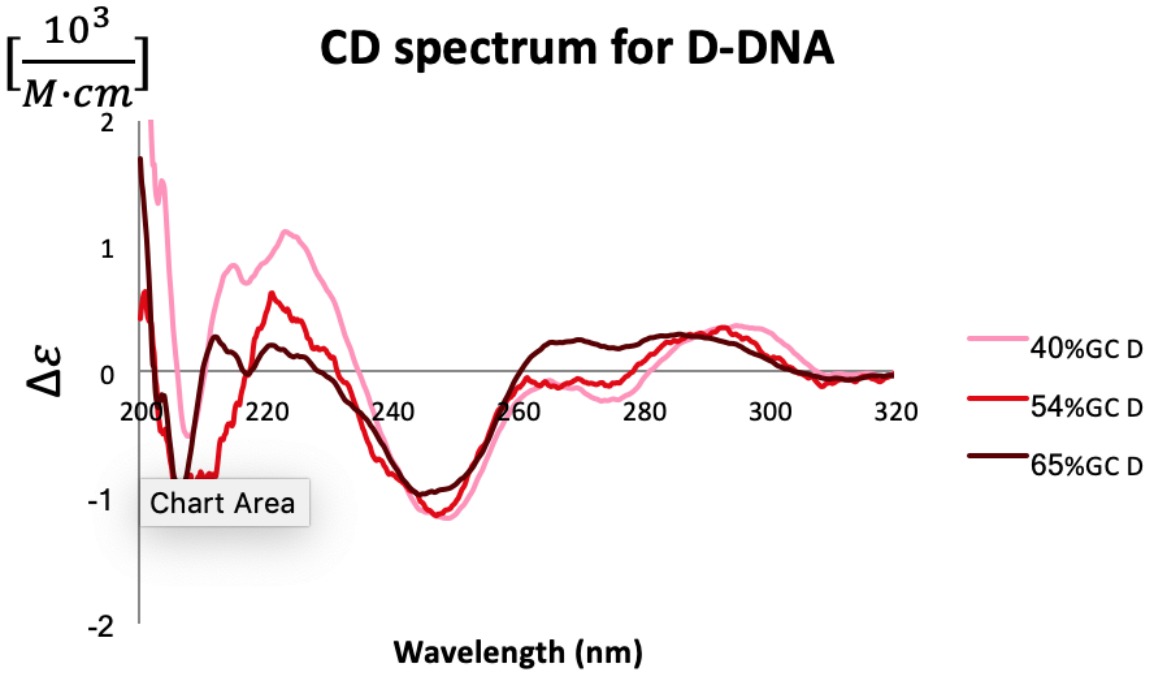

Figure 8. CD spectrum for D-substituted DNA. Darker red represents higher percentage of GC content in the DNA sequences. A progressive shift from non-WT behavior towards a WT-like spectrum can be observed with the increase of GC percentage.

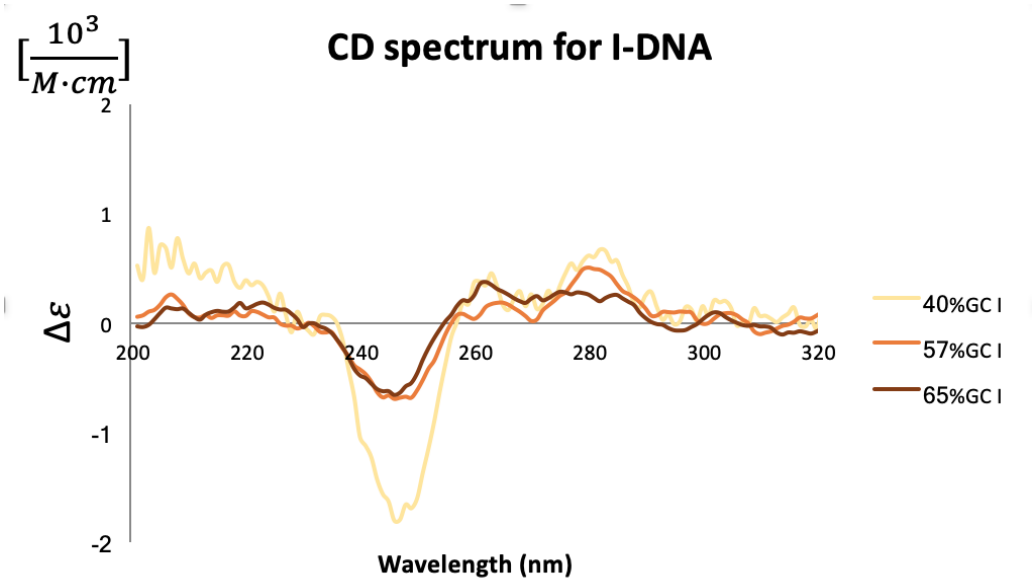

Figure 9. CD spectrum for I-substituted DNA. Darker yellow represents higher percentage of GC content in the DNA sequences.

These results revealed the significant DNA polymorphism for these substituted DNA molecules, which is consistent with previous evidence and the speculation. (51-55, 60) The number of hydrogen bonds between the two backbones of DNA is not the only factor determining DNA polymorphism and may not be the main factor. Although the CD data demonstrated that the introduction of inosine or 2,6 diaminopurine drives DNA between structurally polymorphic forms, it remains unclear what features of these alternate helical conformations are responsible for their distinct DNA mechanical properties.

3.3 AFM images

Atomic force microscopy (AFM) is a scanning probe microscopy with high resolution. Through tapping the sample with a mechanical probe, AFM can form an image of the three-dimensional shape of a sample surface. With the advantage to visualize structure of single molecules and molecular complexes in an aqueous environment, it provides a powerful and multifunctional platform to visualize and manipulate biological samples. (56-59)

4.6kbp WT and D-substituted DNA molecules were recorded and processed with a standard “flatten” filter in the JPK Data Processing software. Only the molecules that were 1.1-1.8 μm surrounding the expected length were included. For I-substituted DNA, we chose an 832bp long DNA sequence, over four hundred DNA molecules were captured and images ware processed and flattened with Nanoscope Analysis. The DNA molecules were then tracked and the contour lengths were recorded with ImageJ and Matlab. Representative images of the wild Type, D-substituted and I -substituted DNA molecules are shown in figure 10-11.

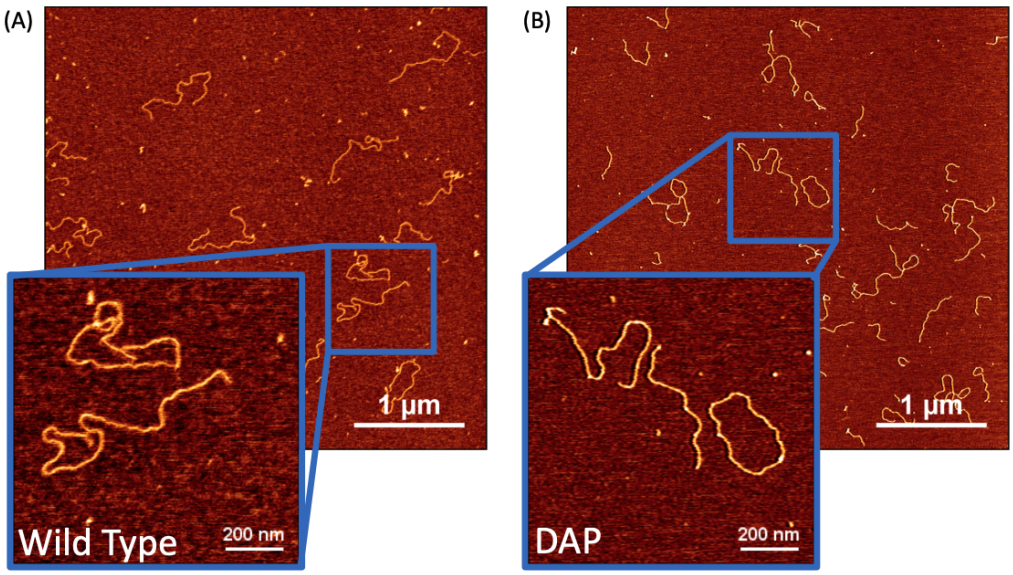

Figure 10. (A) on the left is representative picture of WT DNA and (B) on the right is the D-substituted DNA captured via AFM. The light-yellow worm-like line is the double-stranded DNA molecules.

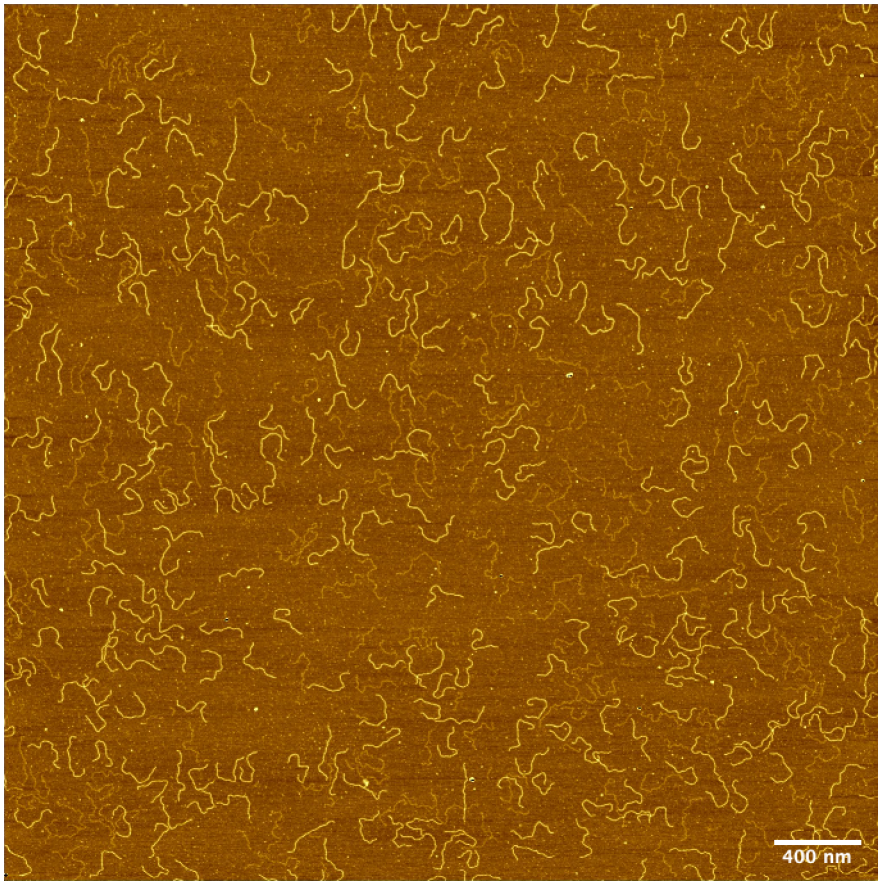

Figure 11. Representative image of I-substituted DNA Molecules captured via AFM.

Contour length was calculated by tracing individual molecules in each AFM image. Compared the wild-type and D-substituted 4.6kpb DNA, the WT DNA had an average contour length of $L_0 = 1.45 \pm 0.12\ \mu m$, while D-substituted DNA had an average contour length of $L_0 = 1.30 \pm 0.12\ \mu m$. D-substituted DNA seems to have a decreased the axial rises per base pair from 3.12 Å/bp to 2.83 Å/bp. The difference is statistically significant in a two-tail T-test with $p = 3.07\times10^{-4}$. Figure 12 summarizes the contour length of WT and D-substituted DNA.

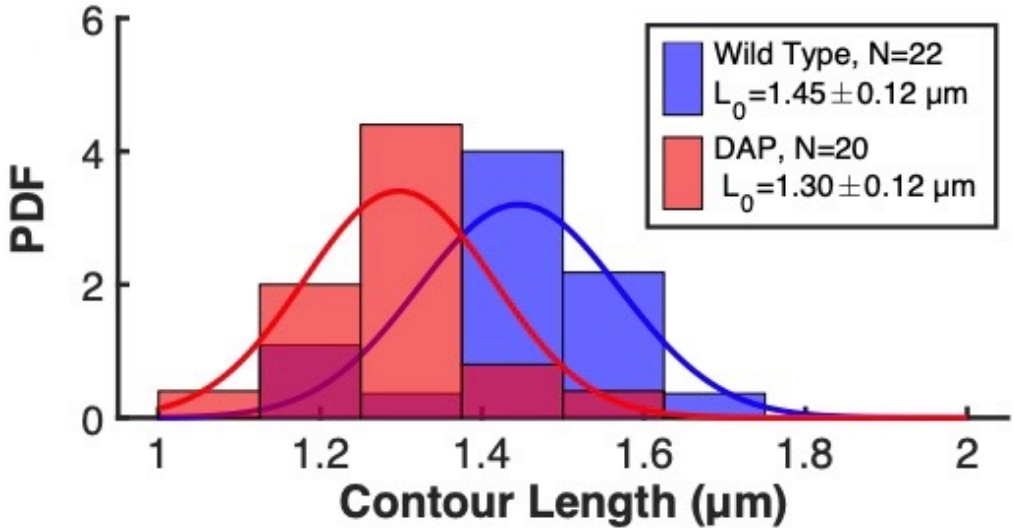

Figure 12. The blue histogram indicates the distribution of contour length for wild type DNA and orange for D-substituted DNA. The distributions are statistically different in a two-tailed T-test with $p = 3.07\times10^{-4}$.

For the I-substituted DNA, 467 DNA molecules are tracked with an average length of 289.8±21.4 nm. Compare to the theoretical contour length for 832 bp B-form DNA, which is 282.9 nm, the I-substituted DNA has increased contour length. Compare to the theoretical

value of axial rises per base pair, I-substituted DNA has decreased from 3.4 Å /bp to 3.48 Å /bp. The difference is even larger when comparing to the experimental value we got from with the 4.6kbp WT DNA, which is 3.12A/bp. The difference between I-substituted DNA and theoretical value for B-DNA in contour length is not statically significant. It might be worthy to repeat the experiment with WT 832bp DNA and compare with the I-substituted DNA to control the factors, including buffers, tension, DNA sequences, etc. And experiment with longer DNA is also important to amplify the difference in contour length. Figure 13 summarizes the contour length information for I-substituted DNA.

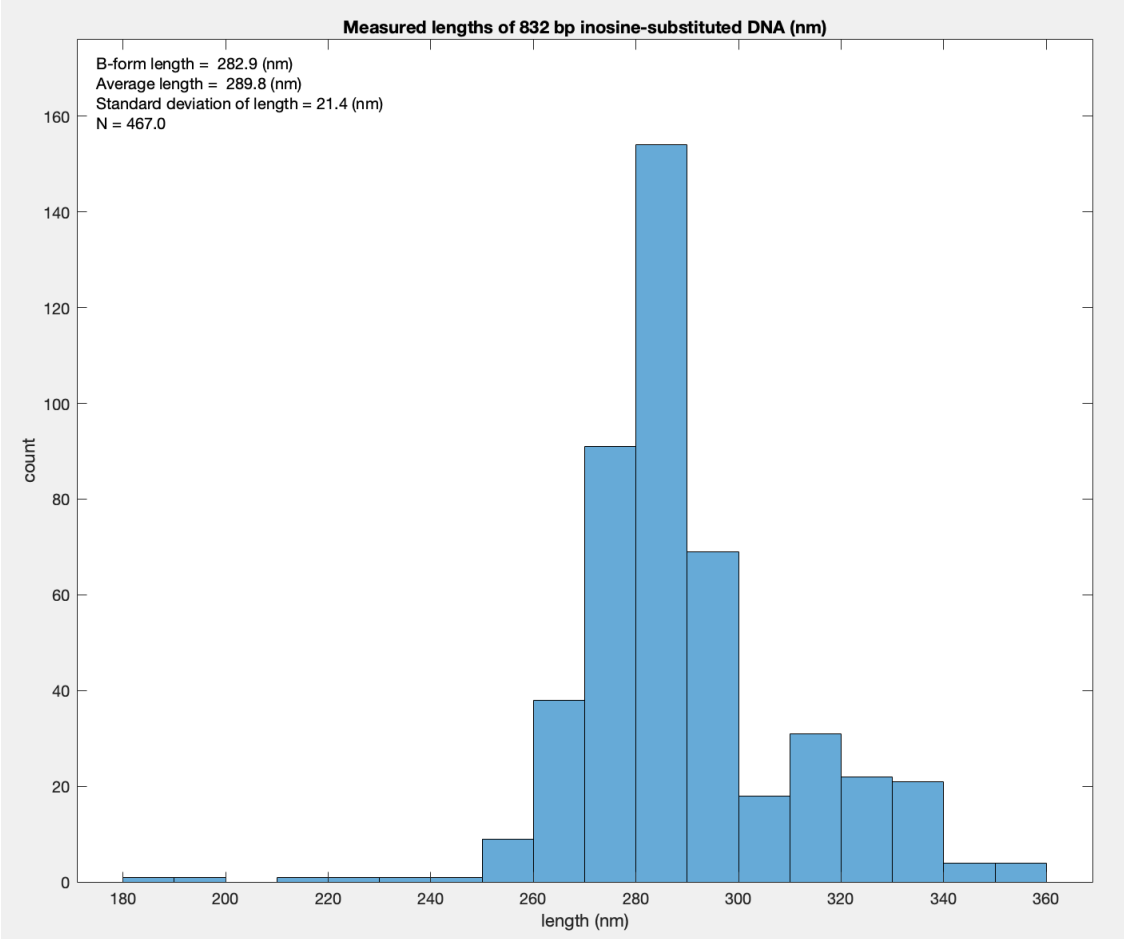

Figure 13. Histogram of contour length for 832bp I-substituted DNA. With 467 DNA molecules, 832bp I-substituted DNA has an average contour length of 289.8 ± 21.4nm, which is lower than theoretical value.

Persistence length, which quantifies the bending stiffness of a polymer, is also an important mechanical property we measured via the AFM image. The persistence length is defined as the length over which correlation in the direction of the tangent are lost. (61) In our study, persistence length was estimated from the measured mean square displacement between points along the contours of molecules in two-dimensions. (62, 63) For a worm-like chain confined to two dimensions, the average squared distance $< r^2 >$ between two points along the chain is given by

$$< r^2 > = 4L_pL_0[1 - \frac{2L_p}{L_0}\left(1 - e^{-\frac{L_0}{2L_p}}\right)] \quad \text{(eqn.4)}$$

where L_0 is the contour length, and L_p is the molecule's persistence length. Previous experiments shown the wild-type DNA yielded $L_p = 56.2 \pm 0.1\ nm$, while D-substitution increased the persistence length to $L_p = 79.9 \pm 0.3nm$ (95% confidence interval). With I-substituted DNA, we calculated the persistence length of $35.2 \pm 0.6\ nm$ which 2D worm-like chain analysis as shown in figure 14.

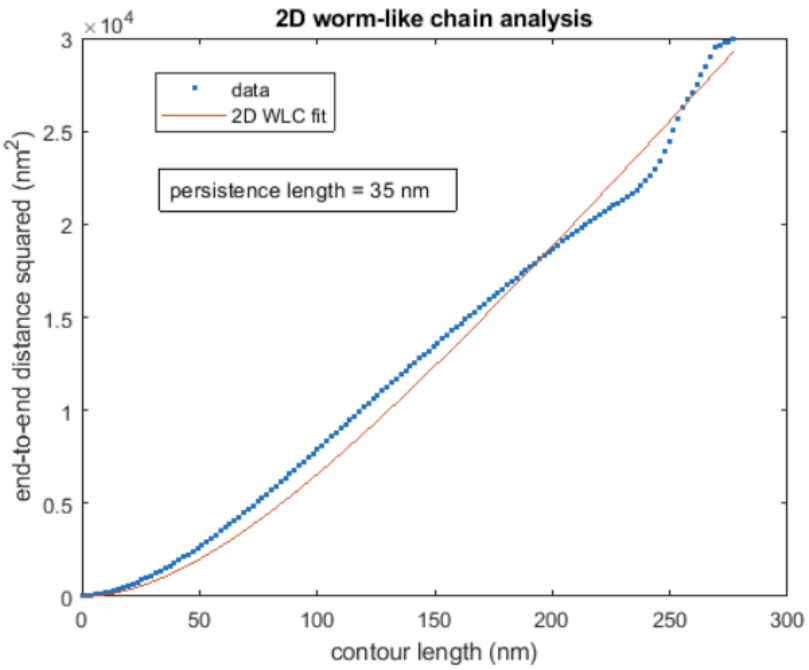

Figure 14. Persistence length of I-substituted DNA calculated with 2D worm-like chain analysis.

There is a significant increase in persistence length due to D-substitution and decrease due to I-substitution, indicating that D-substituted DNA is flexural stiffer and harder to bend, while I-substituted DNA is more flexible.

4. Discussion and Conclusion

The melting temperature measurements show that 2,6-diaminopurine substitution increases, and the inosine substitution decrease, the thermal stability of double-stranded DNA. The number of hydrogen bonds formed in the DNA molecule has a significant influence on the melting temperature; however, it is not the only factor. As for the D-substituted DNA, the melting temperature does not raise up to the level of the poly GC-fragments. Except for the fact of being translated upward by about 5-10 °C, the melting temperature of 150bp-long D-substituted DNA maintains the same trend as wild-type DNA, where the higher the GC content, the higher the melting temperature. Instead, the melting temperature of I-substituted DNA dropped over 20 °C. Thus, the destabilization resulting from this substitution of G with I is much more significant. However, the expected decrease in melting temperature with increased IC content is almost insignificant. The SantaLucia's model fitted for the wild-type sequences predicts the general trend of the D-substituted DNA. As a result, we expect that the D-substituted DNA preserved the nearest-neighbor interactions. We will soon analyze the I-substituted DNA with the SantaLucia's model to identify the nearest-neighbor effect in this type of DNA.

CD spectroscopy confirmed that the 2,6-diaminopuine, or inosine, substitution induced a conformational change in DNA. The wild-type, I-substituted and D-substituted

DNAs all displayed the characteristic features of a B-form DNA spectrum, but the D-substituted DNA had an additional local minimum at around 275nm and I-substituted DNA had a similar local minimum at around 270nm. In addition, the D-substituted DNA showed a progressive shift from less-wild-type towards a WT-like spectrum with the increase of GC (IC) content, while the change was not significant for I-substituted DNA.

The AFM images showed that the D-substitution decreased the average contour length compared to WT. On the other hand, an average contour length slightly higher than WT was measured for I-substituted DNA, at least compared to the theoretical value of B-form DNA. When the axial rise per base pair was calculated, D-substituted DNA and I-substituted DNA showed a decreased and increased value with respect to WT DNA, respectively. The results indicate the D-substituted DNA may be characterized by a more compact, while I-substituted DNA may have a more extended helix, compared to WT DNA. To confirm these results, we would need further experiments with WT DNA of the same sequences as I-substituted DNA. By calculating the persistence length, we found that D-substitution increased the persistence length and I-substitution decreased the persistence length, indicating that D-substituted DNA is more rigid and I-substituted DNA is more flexible, compared to wild-type DNA.

5. References

1. Alberts, B.; Alberts, B. In *Molecular biology of the cell*; Garland Science Taylor & Francis: New York, 2008.
2. Purcell, A. *Basic biology: an introduction*; Basic Biology Ltd: Hamilton, New Zealand, 2018.
3. Saenger, W. Principles of Nucleic Acid Structure. *Springer Advanced Texts in Chemistry* **1984**.
4. Base pair. https://en.wikipedia.org/wiki/Base_pair#/media/File:Base_pair_GC.svg (accessed Apr 5, 2020).
5. Zhurkin, V. B.; Tolstorukov, M. Y.; Xu, F.; Colasanti, A. V.; Olson, W. K. Sequence-Dependent Variability of B-DNA. *DNA Conformation and Transcription* 18–34.
6. Lemieux, S. RNA Canonical and Non-Canonical Base Pairing Types: a Recognition Method and Complete Repertoire. *Nucleic Acids Research* **2002**, *30* (19), 4250–4263.
7. Das, J.; Mukherjee, S.; Mitra, A.; Bhattacharyya, D. Non-Canonical Base Pairs and Higher Order Structures in Nucleic Acids: Crystal Structure Database Analysis. *Journal of Biomolecular Structure and Dynamics* **2006**, *24* (2), 149–161.
8. Blossey, R. DNA and the Nucleosome. *Chromatin* **2017**, 1–20.
9. By Mauroesguerroto - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=35919357
10. Weigele, P.; Raleigh, E. A. Biosynthesis and Function of Modified Bases in Bacteria and Their Viruses. *Chemical Reviews* **2016**, *116* (20), 12655–12687.
11. Cerami, A.; Reich, E.; Ward, D. C.; Goldberg, I. H. The Interaction of Actinomycin with DNA: Requirement for the 2-Amino Group of Purines. *Proceedings of the National Academy of Sciences* **1967**, *57* (4), 1036–1042.

12. Fernandez-Sierra, M.; Shao, Q.; Fountain, C.; Finzi, L.; Dunlap, D. D. E. Coli Gyrase Fails to Negatively Supercoil Diaminopurine-Substituted DNA. *Biophysical Journal* **2016**, *110* (3).

13. Bailly, C.; Payet, D.; Travers, A. A.; Waring, M. J. PCR-Based Development of DNA Substrates Containing Modified Bases: An Efficient System for Investigating the Role of the Exocyclic Groups in Chemical and Structural Recognition by Minor Groove Binding Drugs and Proteins. *Proceedings of the National Academy of Sciences* **1996**, *93* (24), 13623–13628.

14. Waring, M. J.; Bailly, C. The Purine 2-Amino Group as a Critical Recognition Element for Binding of Small Molecules to DNA. *Gene* **1994**, *149* (1), 69–79.

15. Tomasz, M.; Das, A.; Tang, K. S.; Ford, M. G. J.; Minnock, A.; Musser, S. M.; Waring, M. J. The Purine 2-Amino Group as the Critical Recognition Element for Sequence-Specific Alkylation and Cross-Linking of DNA by Mitomycin C. *Journal of the American Chemical Society* **1998**, *120* (45), 11581–11593.

16. Marco, E. Role of Stacking Interactions in the Binding Sequence Preferences of DNA Bis-Intercalators: Insight from Thermodynamic Integration Free Energy Simulations. *Nucleic Acids Research* **2005**, *33* (19), 6214–6224.

17. Yi-Brunozzi, H. Synthetic Substrate Analogs for the RNA-Editing Adenosine Deaminase ADAR-2. *Nucleic Acids Research* **1999**, *27* (14), 2912–2917.

18. Gilbert, S.; Mediatore, S.; Batey, R. Modified Pyrimidines Specifically Bind the Purine Riboswitch. **2006**.

19. Kawai, K.; Kodera, H.; Majima, T. Long-Range Charge Transfer through DNA by Replacing Adenine with Diaminopurine. *Journal of the American Chemical Society* **2010**, *132* (2), 627–630.

20. Kirnos, M. D.; Khudyakov, I. Y.; Alexandrushkina, N. I.; Vanyushin, B. F. 2-Aminoadenine Is an Adenine Substituting for a Base in S-2L Cyanophage DNA. *Nature* **1977**, *270* (5635), 369–370.

21. Weckbecker, G.; Cory, J. G. Metabolic Activation of 2,6-Diaminopurine and 2,6-Diaminopurine-2′-Deoxyriboside to Antitumor Agents. *Advances in Enzyme Regulation* **1989**, *28*, 125–144.

22. Reich, E.; Franklin, R. M.; Shatkin, A. J.; Tatum, E. L. Action Of Actinomycin Don Animal Cells And Viruse. *Proceedings of the National Academy of Sciences* **1962**, *48* (7), 1238–1245.

23. Alibek, K.; Bekmurzayeva, A.; Mussabekova, A.; Sultankulov, B. Using Antimicrobial Adjuvant Therapy in Cancer Treatment: a Review. *Infectious Agents and Cancer* **2012**, *7* (1), 33.

24. Reece, J. B.; Campbell, N. A. *Campbell biology*; Benjamin Cummings: Boston, 2011.

25. Ben-Dov, E.; Shapiro, O. H.; Siboni, N.; Kushmaro, A. Advantage of Using Inosine at the 3 Termini of 16S RRNA Gene Universal Primers for the Study of Microbial Diversity. *Applied and Environmental Microbiology* **2006**, *72* (11), 6902–6906.

26. Bass, B. L.; Weintraub, H. An Unwinding Activity That Covalently Modifies Its Double-Stranded RNA Substrate. *Cell* **1988**, *55* (6), 1089–1098.

27. Kerksick, C. M.; Wilborn, C. D.; Roberts, M. D.; Smith-Ryan, A.; Kleiner, S. M.; Jäger, R.; Collins, R.; Cooke, M.; Davis, J. N.; Galvan, E.; Greenwood, M.; Lowery, L. M.; Wildman, R.; Antonio, J.; Kreider, R. B. ISSN Exercise & Sports Nutrition Review Update: Research & Recommendations. *Journal of the International Society of Sports Nutrition* **2018**, *15* (1).

28. Liu, F.; You, S.-W.; Yao, L.-P.; Liu, H.-L.; Jiao, X.-Y.; Shi, M.; Zhao, Q.-B.; Ju, G. Secondary Degeneration Reduced by Inosine after Spinal Cord Injury in Rats. *Spinal Cord* **2005**, *44* (7), 421–426.

29. Chen, P.; Goldberg, D. E.; Kolb, B.; Lanser, M.; Benowitz, L. I. Inosine Induces Axonal Rewiring and Improves Behavioral Outcome after Stroke. *Proceedings of the National Academy of Sciences* **2002**, *99* (13), 9031–9036.

30. Koch, M.; Keyser, J. D. Uric Acid in Multiple Sclerosis. *Neurological Research* **2006**, *28* (3), 316–319.

31. Markovic-Plese, S.; Singh, A. K.; Singh, I. Therapeutic Potential of Statins in Multiple Sclerosis: Immune Modulation, Neuroprotection and Neurorepair. *Future Neurology* **2008**, *3* (2), 153–167.

32. Inosine Pranobex for Multiple Sclerosis? *InPharma* **1987**, *582* (1), 9–9.

33. Niu, P.-P.; Wu, Y.-H.; Yang, Y. Inosine for Multiple Sclerosis. *Cochrane Database of Systematic Reviews* **2019**.

34. Ascherio, A. Urate as a Predictor of the Rate of Clinical Decline in Parkinson Disease. *Archives of Neurology* **2009**, *66* (12), 1460.

35. Peters, J. P.; Mogil, L. S.; Mccauley, M. J.; Williams, M. C.; Maher, L. J. Mechanical Properties of Base-Modified DNA Are Not Strictly Determined by Base Stacking or Electrostatic Interactions. *Biophysical Journal* **2014**, *107* (2), 448–459.

36. Virstedt, J.; Berge, T.; Henderson, R. M.; Waring, M. J.; Travers, A. A. The Influence of DNA Stiffness upon Nucleosome Formation. *Journal of Structural Biology* **2004**, *148* (1), 66–85.

37. Howard, F. B.; Miles, H. T. 2NH2A.Cntdot.T Helixes in the Ribo- and Deoxypolynucleotide Series. Structural and Energetic Consequences of 2NH2A Substitution. *Biochemistry* **1984**, *23* (26), 6723–6732.

38. Lankaš, F.; Cheatham, T. E.; Špačáková, N. A.; Hobza, P.; Langowski, J.; Šponer, J. Critical Effect of the N2 Amino Group on Structure, Dynamics, and Elasticity of DNA Polypurine Tracts. *Biophysical Journal* **2002**, *82* (5), 2592–2609.

39. Travers, A. A.; Thompson, J. M. T. An Introduction to the Mechanics of DNA. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* **2004**, *362* (1820), 1265–1279.

40. Peters, J. P.; Maher, L. J. Mechanical Properties of DNA-Like Polymers. *Biophysical Journal* **2014**, *106* (2).

41. Bailly, C.; Payet, D.; Travers, A. A.; Waring, M. J. PCR-Based Development of DNA Substrates Containing Modified Bases: An Efficient System for Investigating the Role of the Exocyclic Groups in Chemical and Structural Recognition by Minor Groove Binding Drugs and Proteins. *Proceedings of the National Academy of Sciences* **1996**, *93* (24), 13623–13628.

42. Atkins, P. W.; Paula, J. D. *Elements of physical chemistry*; Oxford University Press: Oxford, United Kingdom, 2017.

43. Solomon, E. I. *Inorganic electronic structure and spectroscopy*; Wiley: New York, 2013.

44. Yakovchuk, P. Base-Stacking and Base-Pairing Contributions into Thermal Stability of the DNA Double Helix. *Nucleic Acids Research* **2006**, *34* (2), 564–574.

45. Šulc, P.; Romano, F.; Ouldridge, T. E.; Rovigatti, L.; Doye, J. P. K.; Louis, A. A. Sequence-Dependent Thermodynamics of a Coarse-Grained DNA Model. *The Journal of Chemical Physics* **2012**, *137* (13), 135101.

46. Sen, A.; Nielsen, P. E. Hydrogen Bonding versus Stacking Stabilization by Modified Nucleobases Incorporated in PNA·DNA Duplexes. *Biophysical Chemistry* **2009**, *141* (1), 29–33.

47. Santalucia, J. A Unified View of Polymer, Dumbbell, and Oligonucleotide DNA Nearest-Neighbor Thermodynamics. *Proceedings of the National Academy of Sciences* **1998**, *95* (4), 1460–1465.

48. Santalucia, J.; Hicks, D. The Thermodynamics of DNA Structural Motifs. *Annual Review of Biophysics and Biomolecular Structure* **2004**, *33* (1), 415–440.

49. Rich, A.; Nordheim, A.; Wang, A. H. J. The Chemistry and Biology of Left-Handed Z-DNA. *Annual Review of Biochemistry* **1984**, *53* (1), 791–846.

50. Ho, P. S. The Non-B-DNA Structure of d(CA/TG)n Does Not Differ from That of Z-DNA. *Proceedings of the National Academy of Sciences* **1994**, *91* (20), 9549–9553.

51. Nelson, H.; Finch, J.; Luisi, B.; Klug, A. The Structure Of An Oligo(Da).Oligo(Dt) Tract And Its Biological Implications. **1993**.

52. Wartell, R. M. The Helix-Coil Transitions of Poly (DA). Poly (DT) and Poly (DA-DT). Poly (DA-DT). *Biopolymers* **1972**, *11* (4), 745–759.

53. Trantírek Lukáš; Štefl, R.; Vorlíčková Michaela; Koča, J.; Sklenářář Vladimír; Kypr, J. An A -Type Double Helix of DNA Having B -Type Puckering of the Deoxyribose Rings 1 1Edited by I. Tinoco. *Journal of Molecular Biology* **2000**, *297* (4), 907–922.

54. Mitsui, Y.; Langridge, R.; Shortle, B. E.; Cantor, C. R.; Grant, R. C.; Kodama, M.; Wells, R. D. Physical and Enzymatic Studies on Poly d(I–C).Poly d(I–C), an Unusual Double-Helical DNA. *Nature* **1970**, *228* (5277), 1166–1169.

55. Vorlfckova, M.; Sagi, J. Transitions of Poly(Dl-DC), Poly(Dl-methyl5dC) and Poly(Dl-bromo5dC) among and within the B-, Z-, A- and X-DNA Families of Conformations. *Nucleic Acids Research* **1991**, *19* (9), 2343–2347.

56. Cai, H.-H.; Zeng, X.; Tang, X.; Cai, J. Atomic Force Microscopy: A Nanoscopic Application in Molecular and Cell Biology. *Atomic Force Microscopy in Molecular and Cell Biology* **2018**, 77–103.

57. Ricci, D.; Braga, P. C. Imaging Methods in Atomic Force Microscopy. *Atomic Force Microscopy* 13–24.

58. Pang, D.; Thierry, A. R.; Dritschilo, A. DNA Studies Using Atomic Force Microscopy: Capabilities for Measurement of Short DNA Fragments. *Frontiers in Molecular Biosciences* **2015**, *2*.

59. Lyubchenko, Y. L.; Gall, A. A.; Shlyakhtenko, L. S. Visualization of DNA and protein-DNA complexes with atomic force microscopy. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4226758/ (accessed Apr 9, 2020).

60. Cristofalo, M.; Kovari, D.; Corti, R.; Salerno, D.; Cassina, V.; Dunlap, D.; Mantegazza, F. Nanomechanics of Diaminopurine-Substituted DNA. *Biophysical Journal* **2019**, *116* (5), 760–771.

61. Flory, P. J.; Flory, P. J.; Flory, P. J.; Flory, P. J. *Statistical mechanics of chain molecules*; Interscience: New York, 1969.

62. Podestà, A.; Indrieri, M.; Brogioli, D.; Manning, G. S.; Milani, P.; Guerra, R.; Finzi, L.; Dunlap, D. Positively Charged Surfaces Increase the Flexibility of DNA. *Biophysical Journal* **2005**, *89* (4), 2558–2563.

63. Vafabakhsh, R.; Ha, T. High Flexibility of Short DNA Probed by Single Molecule Cyclization. *Biophysical Journal* **2011**, *100* (3).

6. Supplementary Table

Table S1. The summary of primers and plasmids used to produce DNA sequences, along with the general information about the DNA sequences used throughout the study.

Name	*Length (bp)*	*GC %*	*Plasmid*	*Forward primer*	*Reverse primer*	*dNTP*
40W	155	40	pBR322	5-TCTTTTACTTT CACCAGCGTTT	5- CCGCTCATG AGACAATAA	A, T, G, C
54W	147	54	pBR322	5- TTCCGGCTG GCTGGTTTATTGC	5- TTCGTTCAT CCATAGTTG	A, T, G, C
57W	142	57	pBR322	5-GATAAAGTTGC AGGACCACTTCT	5- TAACTACGA TACGGGAGG	A, T, G, C
65W	156	65	pBR322	5- GTTTCGGCG AGAAGCAGG	5- ACAGCATGGC CTGCAACG	A, T, G, C
40I	155	40	pBR322	5-TCTTTTACTT TCACCAGCGTTT	5- CCGCTCATG AGACAATAA	A, T, I, C
57I	142	57	pBR322	5-GATAAAGTTGC AGGACCACTTCT	5- TAACTACGA TACGGGAGG	A, T, I, C
65I	156	65	pBR322	5- GTTTCGGCG AGAAGCAGG	5- ACAGCATGGC CTGCAACG	A, T, I, C
40D	155	40	pBR322	5-TCTTTTACTT TCACCAGCGTTT	5- CCGCTCATG AGACAATAA	D, T, G, C
54D	147	54	pBR322	5- TTCCGGCTGG CTGGTTTATTGC	5- TTCGTTCAT CCATAGTTG	D, T, G, C
65D	156	65	pBR322	5- GTTTCGGCG AGAAGCAGG	5-ACAGCATGG CCTGCAACG	D, T, G, C
4.6 kbpW	4643	NA	pKLJ12wt	5-AGCGTTGGCGC CGATTGCAGAAT GAATTT	5-TGGGATCGG CCGAAAGGGC AGATTGATAGG	A, T, G, C
4.6 kbpD	4643	NA	pKLJ12wt	5-AGCGTTGGCGC CGATTGCAGAAT GAATTT	5-TGGGATCGG CCGAAAGGGC AGATTGATAGG	D, T, G, C
832I	832	NA	pKLJ12wt	5- TAACTACGA TACGGGAGG	5-CCGCTCATGA GACAATAA	D, T, G, C