STRUCTURAL AND THERMODYNAMIC STUDIES OF A DNA I-MOTIF: INVESTIGATING THE MONOMER, DIMER, AND TETRAMER HUMAN TELOMERE SEQUENCE CONTEXT

A THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF MASTER OF SCIENCE

IN THE GRADUATE SCHOOL OF THE

TEXAS WOMAN'S UNIVERSITY

DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY

COLLEGE OF ARTS AND SCIENCES

BY

MIKEAL MCKIM, B.S.

DENTON, TEXAS

AUGUST 2015

TEXAS WOMAN'S UNIVERSITY
DENTON, TEXAS

June 26, 2015

To the Dean of the Graduate School:

I am submitting herewith a thesis written by Mikeal McKim entitled "Structural and thermodynamic studies of a DNA i-motif: Investigating the monomer, dimer, and tetramer human telomere sequence context." I have examined this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science with a major in Chemistry.

Richard Sheardy, Ph.D., Major Professor

We have read this thesis and recommend its acceptance:

Mary Anderson, Ph.D.

Mark Britt, Ph.D.

Department Chair

Accepted:

Dean of the Graduate School

DEDICATION

For my grandma, Hazel, I finally found my passion!
Wish you were here to see! Nevertheless, I know you are watching over me!

To my parents, Scott and Rachell, thanks for everything you have done and continue to do! I could not have done this without y'alls unconditional love and unwavering support! I love you both so much!

To John "Chad" Mathews, you have been my rock through some of my hardest times. Thank you for your love and support! I love you!

Last but not least, for my study buddies, Luke and Lucky, thanks for all the nights y'all kept me company studying on the couch, and for ensuring, I took plenty of breaks! Love you both!

ACKNOWLEDGEMENTS

First, I want to give thanks to the Lord, God Almighty for giving me the aptitude, talent and desire to study biochemistry/chemistry and to be curious about the world he designed.

I wish to express my admiration and gratitude to my mentor, Dr. Richard Sheardy, for his excellent guidance, patience, compassion, generosity and for providing me with an exceptional and fostering environment to do research. The opportunities that were laid out for me through working in the Sheardy lab have been greater than anything I could have hoped for. He has mentored me from my second semester on campus as an undergraduate through the completion of my thesis for the Master of Science degree and has taught me more than one can learn in a just a classroom setting. I am grateful to the experiences I gained through his mentorship.

Next, I want to express my appreciation to Dr. Mary Anderson, who served on my committee and unofficially advised me from the first class I attended at TWU through my master's degree. Probably unknown to her, she has been someone I strive to impress and someone I have come to count on for her straight forward, non-sugar coated, helpful advice. It is this advice, encouragement, and her high expectations that helped me to reach for the impossible. After six classes with her, four in undergraduate work and two

in my masters, while I lost a whole lot of sleep, I gained much more than just knowledge of the coarse material. She has taught me how to properly study and through hard work, I can overcome any obstacles.

I want to also thank Dr. Mark Britt, who served on my committee and has been a constant support through my undergraduate and graduate career. He has always been encouraging, compassionate, and an admirable example of unwavering determination.

Through my time at TWU, I have had the privilege of meeting many wonderful people who have impacted my education and life in such a way that they will never be forgotten. I have already thanked three of those people above. Another person who I wish to thank is Kerri Slavens, my boss for the majority of time I worked as a tutor in the Science Learning Resource Center on campus during my undergraduate degree. However, she has been so much more than my boss. She has been an excellent teacher, a dear and trusted friend, an advisor for school and life, and unselfishly helped to edit many papers throughout my time at TWU. Also deserving of my gratitude is Dr. Jack Gill, my advisor as an undergraduate. He always supported and encouraged me, as well as expressing how proud he was of me (which meant the world to me and I will always be grateful for his kind words). Mrs. Cynthia Maguire, who has been my cheerleader, a trusted friend and advisor, and someone who has encouraged me from the moment we met. A kindred soul who has taught me to never give up and how to have grace even through trials and tribulation. I am grateful for all the people who make up the Chemistry and Biochemistry

department at TWU. Even though I have not added each name here, they have all played an important part in my experiences throughout my time here. When I first began this endeavor, fearful and unsure, it was the family I gained by being in this department that helped to get me through hard times, helped encourage me when I couldn't see the light at the end, took me in and made me feel like I belonged. There will always be a place in my heart for my TWU family.

I also want to thank my friends and family who have always supported me. It is because of your love, encouragement, late night talks and/or study groups, and good advice that have carried me through this chapter of my life. I could never express to you enough how much your support has meant to me.

Lastly, I wish to thank my ex-boss from Saltgrass Steakhouse, Kendra Fische-Figueroa. Thank you for caring enough to be everything I needed you to be back then, I could never repay your kindness, compassion, and resolve through those years.

ABSTRACT

MIKEAL MCKIM

STRUCTURAL AND THERMODYNAMIC STUDIES OF A DNA I-MOTIF: INVESTIGATING THE MONOMER, DIMER, AND TETRAMER HUMAN TELOMERE SEQUENCE CONTEXT

AUGUST 2015

Telomeres are protein DNA complexes found at the ends of eukaryotic chromosomes. The human telomere sequence has a G-rich strand, (TTAGGG), complemented to a C-rich strand, (CCCTAA). Previous studies of the G-rich strand indicate that this sequence folds into a distinct conformation, designated as a G quadruplex. The C-rich strand has also been shown to form an unusual structure, under slightly acidic conditions, designated as a DNA i-motif. We are investigating the role of the loops on the structure and stability of the conformations formed from $(CCCXXX)_4$, where X is all permutations of A and /or T. To understand the significance of the presence of the loops, structure and stability will be studied for the dimer and tetramer conformations formed from $(CCCTAA)_x$, where X is either a 1 or a 2. These results indicate that the sequence context and the presence or absence of the loops, influences both the midpoint of the heat induced i-motif to single strand transition (T_m) at pH = 5.0, as well as, the midpoint of the H^+ induced single strand to i-motif transition (pH_{mp}) at 25 ^oC.

TABLE OF CONTENTS

Page

LIST OF TABLES

Tables Page

LIST OF FIGURES

Figures Page

EQUATIONS

Equation Page

CHAPTER I

INTRODUCTION

In 1951, James Watson and Francis Crick made their first attempt at publicly predicting the structure of DNA during a seminar. They proposed DNA was a triple helix that contained a sugar phosphate backbone found in the center of the molecule, with the base pairs protruding outward. As the seminar discussion continued, it was pointed out that there was a huge mistake in their calculations. They miscalculated density of the molecule and did not account for water within the structure [1]. Following the seminar, Dr. Bragg, Watson and Crick's supervisor, feeling embarrassed for them, requested that all work relating to the structure of DNA cease immediately. Instead, it was ordered that Watson was to return to studying viruses as he had done during his PhD and Crick was to focus on finishing his PhD. Undiscouraged and not willing to accept failure, in 1952, Watson and Crick decided to try once more. Fearful that someone would soon uncover DNA's secret, Watson and Crick were convinced that there should be enough information and evidence within the scientific community to discover the structure of DNA. They just needed to gather and investigate the information already out there, and then piece it together [1]. When Watson got his hands on Rosalind Franklin's X ray diffraction image of DNA, he and Crick conceded that DNA must be helical in form. Next, they would recall a dinner, from a year before, in which they had met with Erwin Chargaff. It was at this time that Chargaff had explained, "the amount of

guanine…..equaled the amount of cytosine and the amount of adenine equaled the amount of thymine". Together, with X ray diffraction, Chargaff's imparted knowledge, their tenacity, and perseverance in making the DNA model, Watson and Crick had finalized their model of DNAs structure. Two unknown scientists with very little chemistry background, who, because they cared more about innovation than their own egos, had finally figured out how to construct a DNA molecule. Their 'EUREKA!' moment came in February of 1953. Watson realized, while holding the cardboard cutout model that combining an adenine and thymine held together with hydrogen bonding was the exact same shape as a cytosine and guanine connected thru hydrogen bonding. Later that year in April, Watson and Crick published, "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid". The uncovering of DNA's helical structure and exquisite hydrogen bonding was a crucial moment in science history. Yet, in spite of the sheer magnitude of the discovery, the "chemical and structural complexity" of the DNA mystery had yet to be realized [1].

The structure that Watson and Crick discovered later came to be known as B-DNA. It is well known today, that DNA can form an assortment of conformations other than its right-handed double helix. The conformation properties of DNA are both sequence dependent and environmentally dependent. DNA can be a triple helix, double helix or a single strand. Like B-DNA, A-DNA is a right-handed helix, containing less water and is therefore more compact. Z-DNA is a double helical structure, which forms a left-handed helix that looks like an elongated mirror image of B-DNA [2]. Healthy

individuals contain 46 chromosomes composed of protein-DNA complexes, the internal storage system for genetic material [3-9]. A Eukaryotic chromosome holds a single DNA molecule containing approximately 1.7 x 10^8 base pairs. This linear DNA, which if stretched lengthwise would extend 5.8 cm long, makes up one third of the cell contents and is efficiently compacted in the cell with proteins that make up the other two thirds of the cell [10]. The protein-DNA complex, named chromatin, further compresses the DNA by a factor of 10^4 to 10^5 and is composed of two of each histone (H2A, H2B, H3, and H4) protein and 200 base pairs of DNA. The histone proteins and DNA are joined together by the linker histone (H1). As the DNA continues to supercoil and compress, it forms a higher-ordered structure called the chromosome [10]. On the end of the chromosome is where the telomere is found. Telomeres cap the chromosome ends, to avoid becoming fused together by DNA repair machinery, which repairs fragmented and/or damaged DNA, thus the telomere preserves the integrity and stability of the genome [11]. The human telomere sequence, 5'-TTAGGG-3'/ 5'- CCCTAA – 3' is repeated nearly 3,000 times and can be up to 15,000 base pairs in long [12]. Somatic cells, cells that form the body, tissues and organs, consist of 4-12 kb of double stranded DNA [12,13]. Every time somatic cells replicate the telomere is shortened, however an enzyme called telomerase, functions as a reverse transcriptase, elongating the G-rich (guanine-rich) single strand and lengthening the cell lifetime. While telomerase can lengthen the life of a cell, and thus ward off the aging process, it is also known to be upregulated in 80-85% of all cancer

cells. Through studying the telomeric repetitive single strands, it is possible a mechanism to inhibit telomerase in tumors or vice versa for aging cells could be elucidated.

A structure that could possibly be used as a target to inhibit telomerase is the intramolecular structure formed by G-rich (guanine-rich) DNA found as tandem repeats in the telomere ends of chromosomes. This structure, known as the G-quadruplex is formed from four guanine joined together to form a tetrad of guanine hydrogen bonded so that the guanine form a plane, which is stabilized by the presence of monovalent and divalent cations [3-7,9,14-18]. To help protect the chromosome from recombination and/or degradation events, single stranded telomeric DNA form intramolecular G-tetrads, which stack to form G-quadruplexes [10]. Our group has investigated the structure and stability of DNA sequences $(XXXGGG)_4$, where X = A and/ or T, as well as other sequence variations. The spectroscopic studies indicated that changing the sequence context of the loops, even if only it is only changed by one base in any one loop, the structure and fold stability is affected [19].

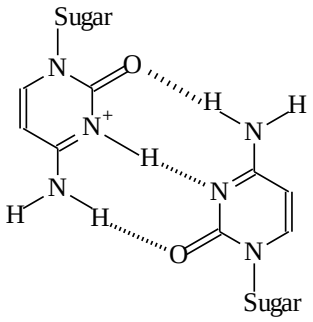

Figure 1 C:C+ Base Pair- Langridge and Rich first proposed a non-Watson-Crick hemi-protonated base pair between two cytosine nucleotides in 1963. The protonation of N3 changes the Crick face from an acceptor-acceptor-donor (C2-N3-C4) motif to an acceptor-donor-donor motif leading to the formation of three hydrogen bonds between the cytosine bases.

The C–rich complementary strand of the telomere folds into an intramolecular structure known as the i-motif. The Groundwork for the discovery of the i-motif began

with the discovery of $C:C^+$ base pairs, which was first proposed in 1963 when Langridge and Rich suggested that poly-cytosine oligos could become hemi-protonated and form hydrogen bonds at low pH. However, it was not until 1993 that Gehring, Leroy and Gueron proposed the i-motif structure. They suggested that under acid conditions when an oligo of DNA contained a stretch of cytosine, the structure formed is an intercalated helical i-motif. Since the use of NMR to determine the structure of sequence d(TCCCCC), proved difficult because of the repetitive nature of the sequence, nuclear Overhouser effects were used to first characterize its structure. Further characterization using gel filtration chromatography was used to determine the tetrameric stoichiometries. Since that first discovery, in 1993, of the i-motif structure was elucidated, supplementary research has gained momentum [4-6].

Early thoughts held there was no biological relevance for the i-motif and therefore the structure was primarily used for pH switches in nanotechnology [6,20]. However, the role of the i-motif in the regulation of gene expression, cellular aging, cancer, and in determining in vivo pH changes has recently been uncovered [4,6,18,21-24], as well as uses in nanotechnology. Further support for biological importance, is proteins that bind to tandem repeats of cytosine in the human telomere, have been reported and the structure has been seen to form at pH 7.0 in environmental conditions of 0 ºC [6,25]. Reilly et al has performed experiments that suggest C-rich DNA can fold into the i-motif under crowded cellular conditions [4]. In response to the newly acquired information, an understanding of the folding, unfolding and the environment in which there is stability must be understood

for the structure to gain a better understanding of the seemingly endless possibilities for the i-motif. The hemi–protonated C:C$^+$ hydrogen bonds are able to form once the protonation of N3 of C occurs. Upon protonation, the once Watson-Crick face of *acceptor-acceptor-donor* (C2-N3-C4) motif to an *acceptor-donor-donor* motif leading to base pairing schemes such as Hoogsteen C:CH$^+$ base pairs [26-30]. At slightly acidic pH, this protonation takes place and the molecule folds back on itself forming perpendicularly stacked C:C$^+$ pairs with all the bases in the anti-conformation. The folded structure contains two 5'-XXX-3' loops on the same end of the structure, as well as a 5'-XXX-3' loop on the other end and a 3'-XXX tail.

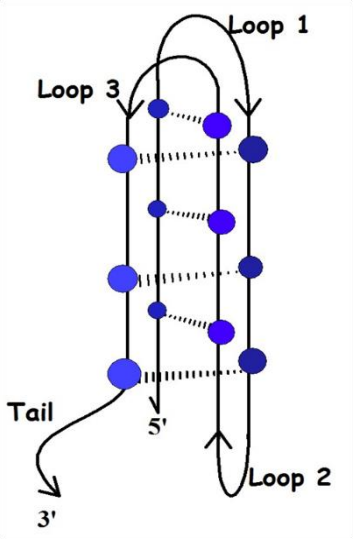

Figure 2 I-Motif conformation of the monomer sequence (CCCXXX)4 where X = A and/ or T. Under slightly acidic conditions, the C-rich monomer forms an intramolecular structure which is stabilized by 6 C:CH+ base pairs that intercalate and stack orthogonally to one another. The folded structure contains three 5'-XXX-3' loops and a 3'-XXX tail.

Similar to studies previously carried out to determine the structure and stability of the g-quadruplexes by changing the loop sequence context, our lab investigated structure and stability of the i-motif structure, first by changing the loop sequence, then by changing the number of loops formed from three to two to zero. By investigating the conformational properties of all permutations of $(CCCXXX)_4$, where X = A and/or T, the

influence that the loop sequence context has on the structure and stability of the monomolecular i-motif conformation, formed under slightly acidic conditions, may be identified. Additionally, a bimolecular and tetramolecular structure, with sequence $(CCCTAA)_x$, where X = 2 or 1, respectively, will be studied to pin point the role, if any, dictated by the presence, number, or absence of loop(s) has on the stability of the structure.

CHAPTER II

METHODOLOGY

Preparation of Buffers

Potassium phosphate buffers (10 mM phosphate, 0.1 mM EDTA, 21.4 mM K^+, pH 7.0 and 10 mM phosphate, 0.1 mM EDTA, 34.2 mM K^+, pH 5.0) were prepared under standard conditions using KH_2PO_4 (VWR International Lot # 46032627), K_2HPO_4 (VWR International # 46205641), and EDTA (EMD Chemicals Lot # 45166714). The final concentration of K^+ was brought to 115 mM by adding KCl. This solution was then filtered through a 0.45 µm Millipore filter and degassed before storing it for use. Each DNA sequence was prepared in standard 10 mM phosphate buffer with 115 mM K^+ and different pHs.

DNA oligomers

All DNA oligomers were purchased from Biosynthesis Inc. (Lewisville, TX) and used without further purification. Each sequence was reconstituted in 1 mL of buffer at either a pH of 7.0 or a pH of 5.0, heated to 95 °C, and slowly cooled to room temperature and stored at 4 °C. Extinction coefficients, provided by the supplier, are shown in Table

UV/Vis Spectroscopy

All UV/Vis spectroscopy was carried out on a Varian Cary 100 Bio model (Varian Associates, Palo Alto, CA) and was used to determine the concentration of the samples used in the circular dichroism. After each optical measurement was obtained, a dilution of each sample was prepared by pipetting 50 uL of the sample directly from the cuvette into a labeled Eppendorf tube, then diluted with 950 uL reverse osmosis (RO) water. The baseline was subtracted and the runs were made using 10 mm square quartz cuvettes from 320 – 220 nm at 25 and 95 °C. Concentrations were then calculated using the appropriate extension coefficient as reported above.

Circular Dichroism Spectroscopy

An Olis RMS 1000 CD spectrophotometer (Olis, Inc. Athens, GA) was used to carry out all circular dichroism studies. A 1 mm circular quartz cuvette was used to run all samples including the baseline. Initially, the CD spectra of a particular DNA sequence in buffer of pH 7.0 and 5.0 were determined. Different ratios of the DNA samples in pH 7.0 and 5.0 were mixed to obtain samples of different pH values. Continuing in this fashion, DNA solutions in buffers of pH ranging from 7.0 to 5.0 at 0.1 pH unit increments were generated. Since the volumes of these solutions were typically around 400 uL, a micro pH electrode was used to measure the exact pH of each solution before placing that sample in the CD spectrometer. CD spectra were determined at 25 °C from 320 nm to 220 nm with 1 nm intervals. Upon completion of the scan, an aliquot of

the DNA solution was diluted for UV/Vis determination (see below) of its concentration. For those samples at pH 5.0, CD spectra were also obtained from 320 nm to 220 nm every 5 °C from 25 °C to 95 °C to generate a two dimensional melt.

CHAPTER III

A DNA I-MOTIF: AN EXPLORATION OF STRUCTURE AND STABILITY IN THE MONOMER

To Be Submitted For Publication in the

Biopolymers Journal

Mikeal McKim, Alexander Buxton, Courtney Johnson,
Amanda Metz and Richard D. Sheardy

Spectroscopic studies of DNA oligomers of general sequence $(CCCXXX)_4$, where X = A and/or T, were carried out to investigate their structures and stabilities under different pH and temperature conditions. Figure 3 shows the CD spectra of $(CCCTAA)_4$ at different pH and T values. At pH 7.0, the oligomer is single stranded at both 25 and 95 °C as seen in the upper left hand panel of Figure 3. In the lower right hand panel of Figure 3 the oligomer is single stranded at 95 °C at either pH 7.0 or 5.0. However, the CD spectra seen in the lower left hand panel of Figure 3 of the oligomer at 25 °C and pH 5.0 indicates that the oligomer has undergone a conformational transition to the i-motif, which denatures to the single strand at pH 5.0 and 95 °C as seen in the upper right hand panel of Figure 3.

These observations are confirmed by examination of the UV/Vis spectra shown in Figure 4. The increase in molar absorptivity (i.e., extinction coefficient) at λ_{max} observed at pH 5.0 and 25 °C in the lower left hand panel of Figure 4 is consistent with the stacking of the transition dipole moments of the $C:CH^+$ base pairs that form upon

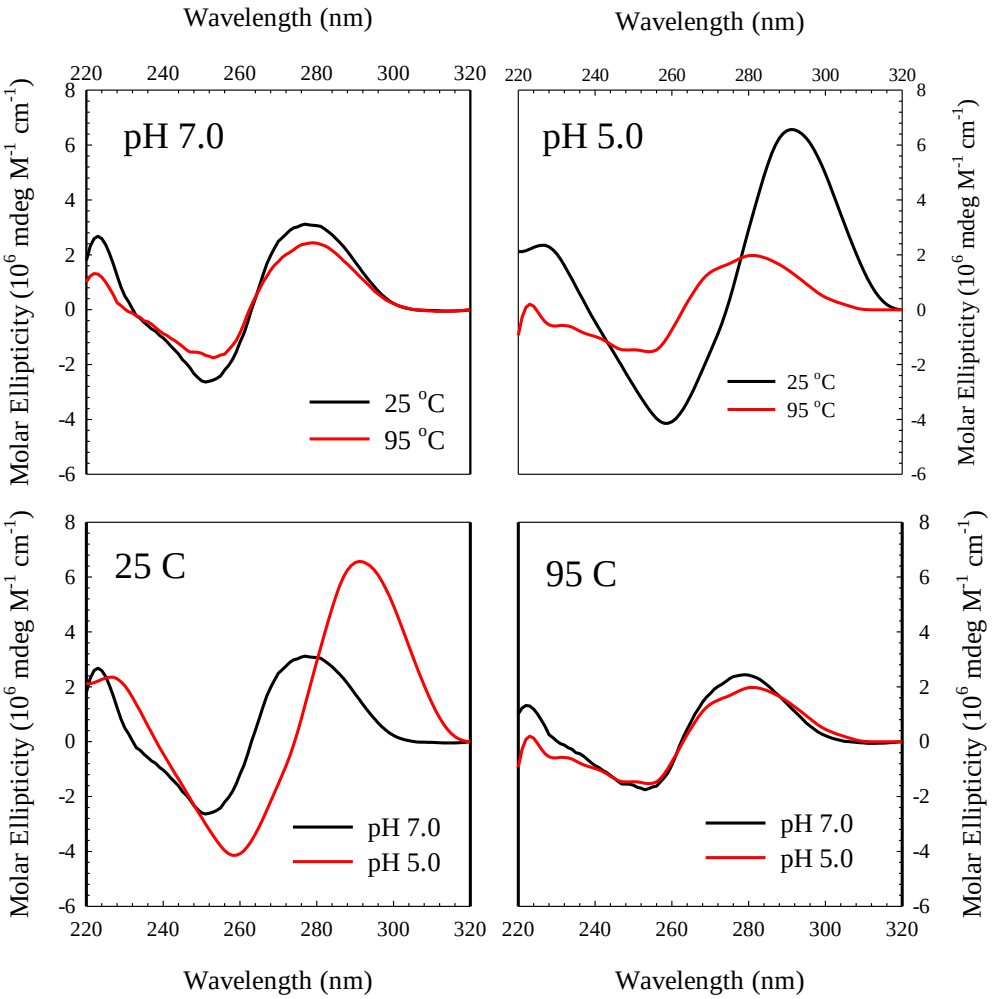

Figure 3 CD spectra graphs for the sequence $(CCCTAA)_4$ showing single strand graph representation and i-motif conformation representation graph. Upper panels- pH value is held constant and the oligo is scanned at both 25 °C and 95 °C. Left panel- At pH 7.0 the oligo is single stranded at both 25 °C and 95 °C. Right panel- At pH 5.0 the oligo is folded into an i-motif conformation at 25 °C and single stranded at 95 °C. Lower panels- Temperature is held constant while the pH of the buffer is changed between 7.0 and 50. Left panel- At 25 °C is single stranded at pH 7.0 and folded into i-motif conformation at pH 5.0. Right panel – At 95 °C oligo is single stranded at both pHs.

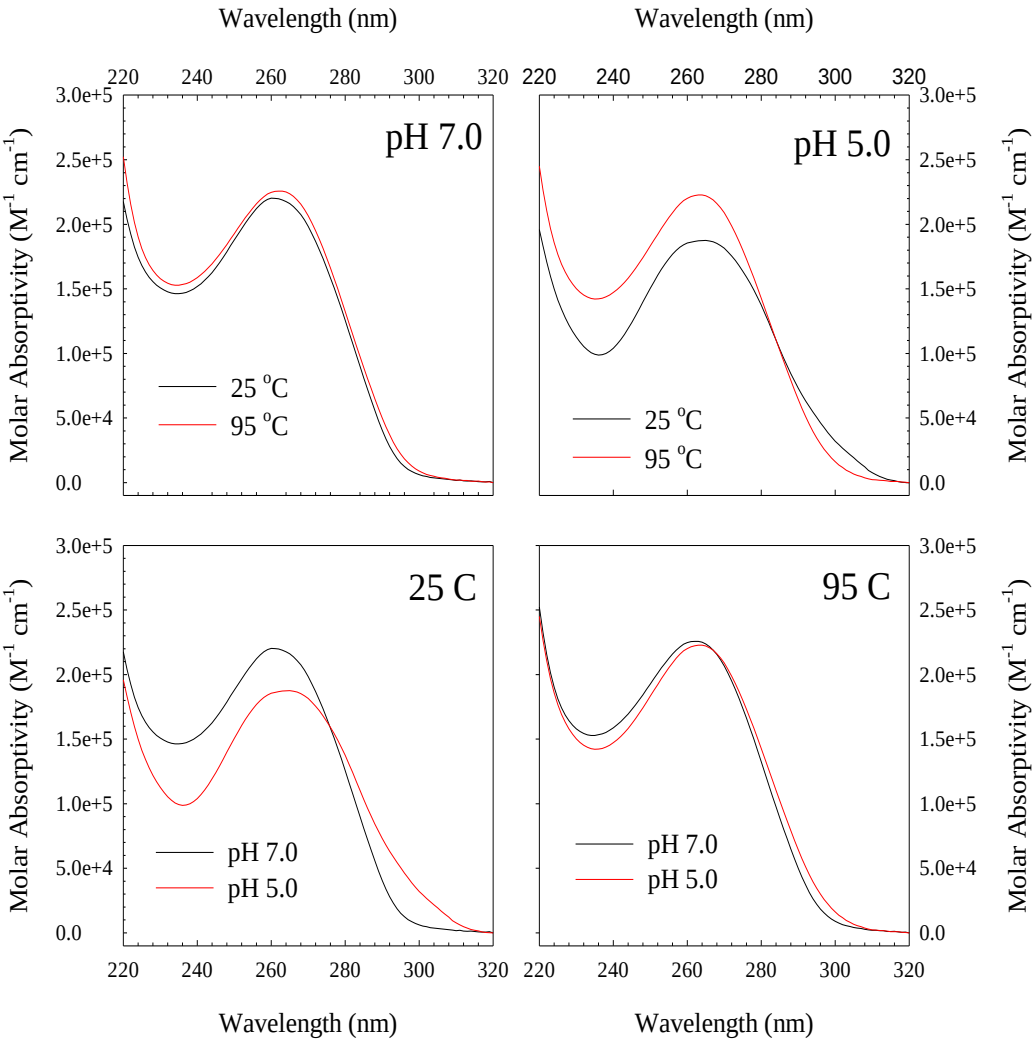

Figure 4 UV/Vis spectra for sequence $(CCCTAA)_4$ showing single strand to folded i-motif conformation change. Upper panels- pH value is held constant and the oligo is scanned at both 25 °C and 95 °C. Left panel- At pH 7.0 the oligo is single stranded at both 25 °C and 95 °C. Right panel- At pH 5.0 the oligo is folded into an i-motif conformation at 25 °C and single stranded at 95 °C. Lower panels- Temperature is held constant while the pH of the buffer is changed between 7.0 and 50. Left panel- At 25 °C is single stranded at pH 7.0 and folded into i-motif conformation at pH 5.0. Right panel- At 95 °C oligo is single stranded at both pH values.

Table 1 Spectral properties of the i-motif forming oligomers at 25 °C.

Sequence	UV/Vis		CD					
			pH 7.0		pH 5.0		Isoelliptic Point	
	λ_{max} (nm)	ε_{260} (M^{-1} cm^{-1})	λ_{max} (nm)	θ_{max} (10^6 mdeg M^{-1} cm^{-1})	λ_{max} (nm)	θ_{max} (10^6 mdeg M^{-1} cm^{-1})	λ_{max} (nm)	θ_{max} (10^6 mdeg M^{-1} cm^{-1})
$(CCCAAA)_4$	259	233,600	278	1.992	292	5.726	280	1.96
$(CCCAAT)_4$	262	220,300	276	2.161	291	5.043	279	2.11
$(CCCATA)_4$	262	226,000	277	2.530	291	6.143	279	2.45
$(CCCATT)_4$	264	204,700	278	2.665	290	6.422	278	2.66
$(CCCTAA)_4$	260	220,400	279	1.890	291	6.358	278	1.88
$(CCCTAT)_4$	265	207,100	279	2.393	290	6.860	278	2.40
$(CCCTTA)_4$	265	204,800	278	3.775	289	6.984	NO	NO
$(CCCTTT)_4$	267	204,700	279	2.529	289	8.875	278	2.48

*NO = Not Observed

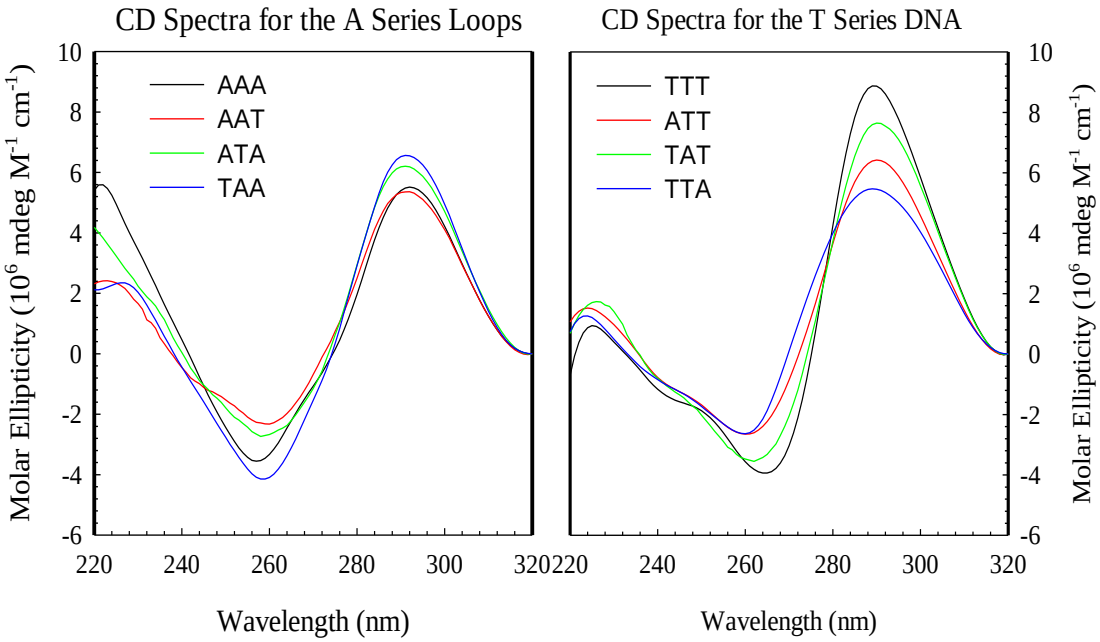

Figure 5 CD spectra overlays for the A Series Loops and the T Series Loops in pH 5.0 buffer at 25 °C. Left panel- A series spectra with loop sequence AAA in black, AAT in red, ATA in green and TAA in blue. Right panel- T series spectra with loop sequence TTT in black, ATT in red, TAT in green, and TTA in blue.

lowering the pH from 7.0 to 5.0. The decrease in molar absorptivity observed at pH 5.0 and 95 °C is consistent with the unstacking of the C:CH^+ base pairs as the i-motif denatures to the single stranded structure as observed in the upper right hand panel of Figure 4. All oligomers studied show similar behavior in their CD and UV/Vis spectra. All oligomers studied fold into the i-motif conformation as determined by their CD spectra at 25 °C and pH 5.0. The UV/Vis and CD spectral characteristics are found in Table 1. As can be seen from Figure 5 and Table 1, there are slight differences in λ_{max} and extinction coefficients in the UV/Vis spectra due to the sequence context variations

At pH 7.0, all oligomers have a peak in the range of 276 to 281 nm and molar ellipticities ranging from 1.890 to 3.775 x 10^6 mdeg M^{-1} cm^{-1}. At pH 5.0, the peak shifts to 289-292 nm with much higher molar ellipticities ranging from 5.043 to 8.875 x 10^6 mdeg M^{-1} cm^{-1}. Further, the T rich sequence have, in general, higher ellipticities at λ_{max} than the A rich sequences. The variations in peak location and molar ellipticities at either pH is due to variations in sequence context.

The overlay of the CD spectra of $(CCCTAA)_4$ at 25 °C as a function of pH is shown in Panel A of Figure 6. These data demonstrate the pH induced transition from the single strand at pH 7.0 to the i-motif at pH 5.0. The presence of the isoelliptic point at 278 nm is consistent with a two state transition. The pH titration curve in Figure 6 Panel C was obtained using the molar ellipticity at 292 nm:

$$\text{Fraction I-Motif} = \frac{(\theta_{292,\,pH} - \theta_{292,\,pH\ 5.0})}{(\theta_{292,pH\ 5.0} - \theta_{292,\,pH\ 7.0})} \qquad (1)$$

where $\theta_{292,\,pH}$ is the molar ellipticity at 292 nm for any pH solution, $\theta_{292,\,pH\ 7.0}$ is the molar ellipticity at 292 nm at pH 7.0 and $\theta_{292,\,pH\ 5.0}$ is the molar ellipticity at 292 nm at pH 5.0. The sigmoidal nature of the titration curve suggests a highly cooperative transition. The point in the titration when there are equal concentrations of both the single stranded species and the i-motif species is represented by the midpoint of the transition, pH_{mp}. Assuming a two state transition means, 50% of the C bases in a single molecule have been protonated. This should then relate the value of the pH at the midpoint of the

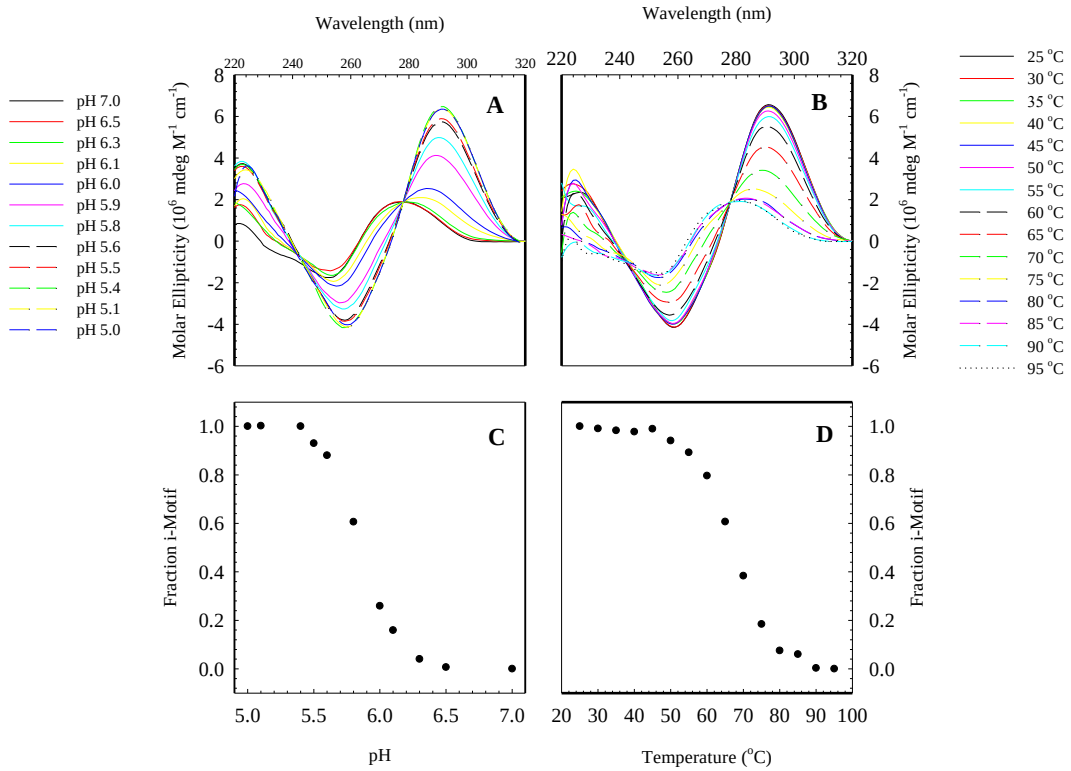

Figure 6 CD spectra of $(CCCTAA)_4$ i-Motif to single strand transition as a function of pH and T. Upper Panel- CD spectra overlays for the sequence $(CCCTAA)_4$. Left Panel- pH induced single strand to i-motif transition from pH 7.0 to 5.0 at a constant 25 °C. Right Panel- Temperature induced i-motif to single strand transition from 25 °C to 95 °C at a constant pH of 5.0. Lower Panels- Fraction of i-Motif for the pH and T transitions. Left Panel- The fraction of i-motif of the pH induced transition, pH_{mp} = 0.5 at 292 nm which represents 50% folded i-motif conformations and 25% protonated C. Right Panel- The fraction of i-motif of the T induced transition, T_m = 0.5 at 292 nm which represents 50% unfolded i-motif conformations and 25% unprotonated C.

transition to the pK_a of N3 of cytosine. For this particular oligomer, pH_{mp} occurs at pH = 5.86 $\pm$ 0.03. Although it should be mentioned, that at the midpoint of the transition, only 25% of all protonatable Cs have been protonated.

Panel B of Figure 6 show the overlay of the CD spectra of $(CCCTAA)_4$ at pH 5.0

as a function of T. These data demonstrate the thermally induced transition from the folded i-motif at 25 °C to the unfolded single stranded structure at 95 °C. The optical melting curve in Figure 6D was obtained using the molar ellipticity at 292 nm:

$$\text{Fraction i-motif} = \frac{(\theta_{292,\ T} - \theta_{292,\ 95\ ^\circ C})}{(\theta_{292,\ 25\ ^\circ C} - \theta_{292,\ 95\ ^\circ C})} \qquad (2)$$

where $\theta_{292,\ T}$ is the molar ellipticity at 292 nm at any temperature, $\theta_{292,\ 25\ ^\circ C}$ is the molar ellipticity at 292 nm at 25 °C and $\theta_{292,\ 95\ ^\circ C}$ is the molar ellipticity at 292 nm at 95 °C. For the thermally induced denaturation of this oligo, the presence of an isoelliptic point in the CD spectra and the sigmoidal shape of the melting profile suggest a cooperative, two state transition. Thus, assuming a two state transition, the midpoint of the transition, T_m, represents the point in the transition where there are equal concentrations of both the single stranded species and the i-motif species. Furthermore, the stability of the i-motif conformation should be related to the value of the T at the transition midpoint. For this particular oligomer, T_m occurs at T = 67.1 + 0.3 °C. All oligomers studied, with the exception of $(CCCTTA)_4$, gave similar pH and temperature dependent CD spectra, as well as similar pH titration and optical melting curves (see Supplemental Data for six of the remaining oligomers) yet with different pH_{mp} and T_m values as seen in Table 2. As stated above, a sharp isoelliptic point at 278 nm for both the pH titration and the thermal melt is seen in the spectral overlays presented in Figure 6. The plot of molar ellipticity as a function of wavelength was obtained from the raw CD data by correcting

Table 2 Loop sequence effects on the i-motif to single strand transition [1],

Loop Sequence	**T_m (°C)**		**ΔT_m (°C)**	**pH_{mp}**		**ΔpH_{mp}**
	X = A	*X= T*		*X = A*	*X = T*	
A-X-A	55.6	59.2	3.6	5.82	5.84	0.02
T-X-T	65.9	>75	>9	6.12	6.15	0.03
A-X-T	62.2	75.6	13.4	5.85	6.35	0.50
T-X-A	66.2	67.1	0.9	5.86	6.07	0.21

[1] T_m is the melting temperature determined at the midpoint of the Fraction i-Motif vs T plot and pH_{mp} is the pH at the midpoint of the single strand to i-motif transition determined from the Fraction i-Motif vs pH plot. The delta values are calculated as: (X = T) – (X = A).

for concentration. A sharp isoelliptic point was revealed at a particular wavelength when these temperature dependent spectra were studied, for all but one oligomer with the sequence $(CCCTTA)_4$. Consequently, for each DNA sequence displaying an isoelliptic point, the average value of the molar ellipticities at that particular wavelength was used to normalize all spectra, resulting in the observed isoelliptic points (seen in table 1). All obtained normalized ellipticities at the isoelliptic points, of the averaged values used to construct the plots, were within ±2%. The variations between the concentration corrected ellipticities came from noise, and not from differences in the DNA concentrations. Therefore, by applying the normalizations to the data, some of the noise is "smoothed" out. Normalization of the pH titration spectra was then carried out using the molar

ellipticity at the isoelliptic point. Through this normalization process, errors made in obtaining DNA concentration are minimized. Isoelliptic points and ellipticities for each oligo are summarized in Table 1. Contributing to the pH titration and thermal melt CD spectra resemblances, is that both the initial states (i.e. unfolded) and final states (i.e. folded) of the oligomers are similar for both transitions.

This normalization process worked very well for all DNA oligomers except $(CCCTTA)_4$. New samples were purchased and reconstituted under the exact same conditions as stated above, however, in all five of the new samples no isoelliptic point was observed. As the temperature increases from 25 °C to 55 °C, the molar ellipticity at 292 nm increases and then decreases as the temperature continued to increase to 95 °C as seen in Figure 7. In Panels A and B of Figure 7, before the oligomer unfolds, it appears as if there is a non-denaturational conformational change between 35 and 55 °C. Typically, molar ellipticities increase with base stacking. If this is the case, the increased molar ellipticity may be due to the transient formation of an A:T base pair that can stack with the first or last $C:CH^+$ base pair in the folded structure. Hence, the temperature induced unfolding of the i-motif conformation for $(CCCTTA)_4$ is not two state. Nonetheless, in Panel C of Figure 7 the pH induced folding does appear to be two state transition.

T_m and pH_{mp} values obtained for all eight oligomers are summarized in Figure 8. The Left Panel is the T_m values that range from 56.5 °C to over 80 °C and are listed in Table 2. It is obvious that the loop sequence context of the XXX segments influence the

stability of the folded i-motif conformation. As seen by the list in Table 2, sequences with the lowest T_m values an A in the first and third positions and sequences with a T in the third position have the highest T_m values. Additionally, comparison of the second

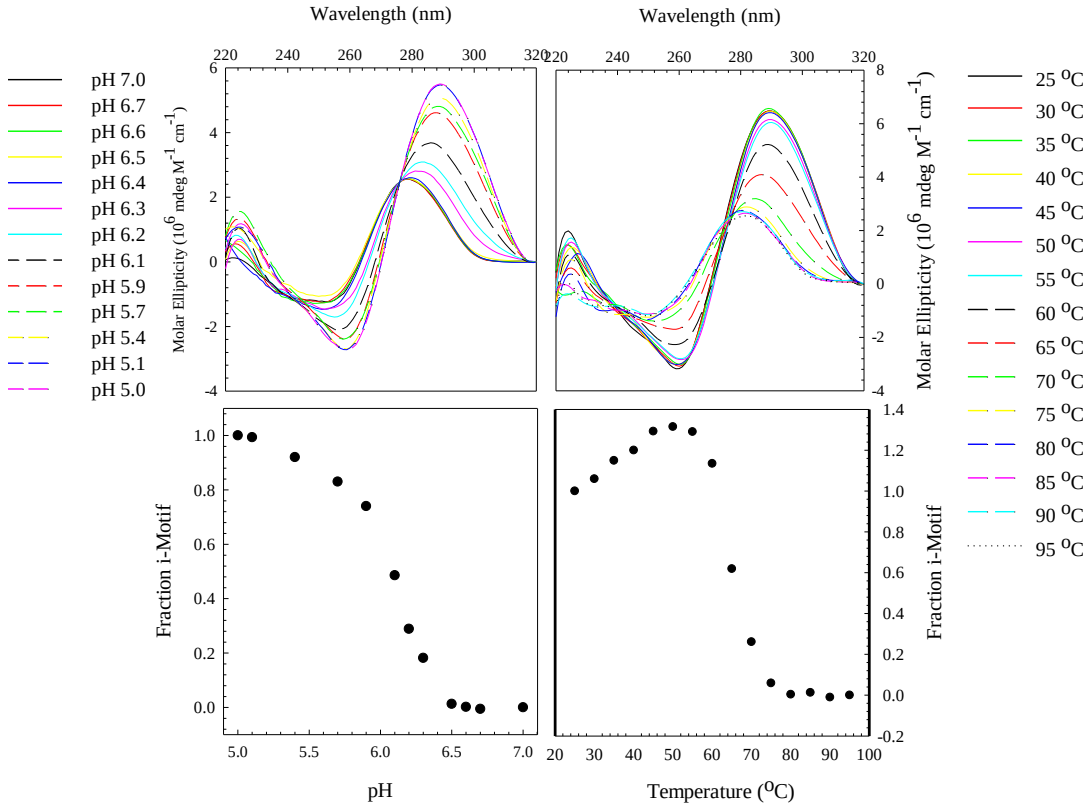

Figure 7 CD spectra of $(CCCTTA)_4$ i-Motif to single strand transition as a function of pH and T. Upper Panel- CD spectra overlays for the sequence $(CCCTTA)_4$. Left Panel- pH induced single strand to i-motif transition from pH 7.0 to 5.0 at a constant 25 °C. Right Panel- Temperature induced i-motif to single strand transition from 25 °C to 95 °C at a constant pH of 5.0. For this particular graph, a well-defined isoelliptic point is not observed. Lower Panels- Fraction of i-Motif for the pH and T transitions. Left Panel- The fraction of i-motif of the pH induced transition, pH_{mp} = 0.5 at 289 nm which represents 50% folded i-motif conformations and 25% protonated C.. Right Panel- The fraction of i-motif of the T induced transition, T_m = 0.5 at 289 nm.

position, T vs A, reveals a greater stability for the folded i-motif conformation. As seen in Figure 2 the i-motif folded conformation contains three XXX loops and a 3' tail with

XXX. Base stacking plays a significant role in the stability of DNA secondary structures [26] so the question is what are the bases in the loops and tail doing: Are they simply waving in the breeze or do they have some structure as well? Clearly, the bases have some sort of structure, stacking either with themselves or with the C:CH$^+$ base pairs.

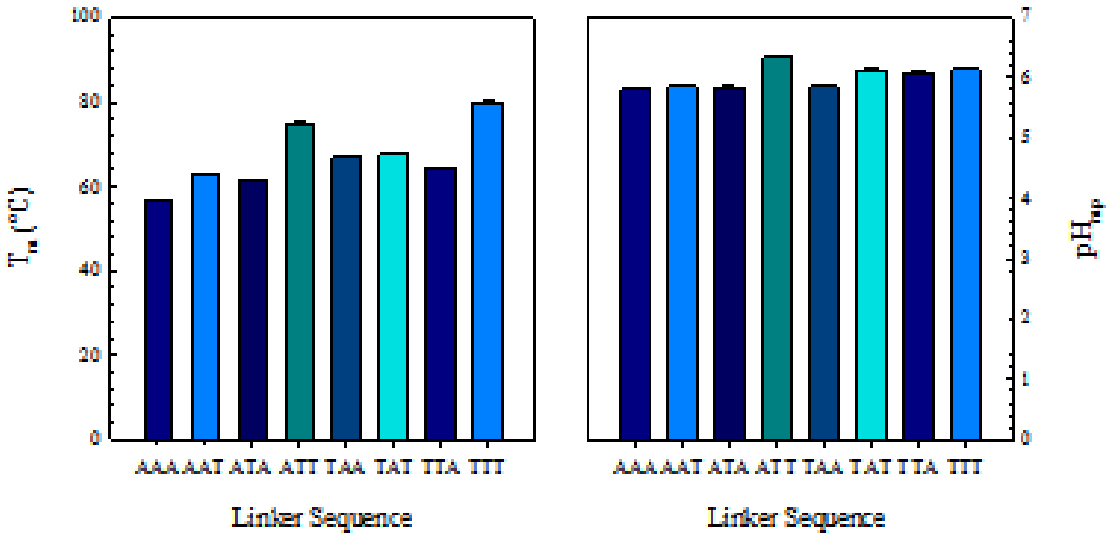

Figure 8 Comparison of the pH_{mp} and T_m values for all monomeric oligomer sequences. Left Panel- A summary of T_m values for all monomer sequences, ranging from 56.5 to 80 °C. Right Panel- A summary of pH_{mp} values for all monomer sequences, ranging from 5.82 to 6.35.

Furthermore, an A:T base pair adjacent to a C:CH$^+$ base pair, could result when and A is in the first position and a T in the third position (or T in first position and an A in the third). Thus, the variation in T_m values observed are due to sequence context effects whereby different sequences have slightly different structures, stacking interactions and possible base pairing interactions and therefore, different stabilities.

Further analysis of the right hand panel of Figure 8 and Table 2 can be done on the pH_{mp} values. In spite of the reduced changes between the pH_{mp} values from that of the T_m values, some correlations to the XXX linkers sequence context effects seen are: 1) those oligomers containing more A bases in the XXX linker have nearly identical pH_{mp} values ranging between 5.82 and 5.86; and, 2) a higher pH_{mp} value with a broader range from 6.07 to 6.35 is seen in the oligomers with more T bases in the XXX linker. In addition, for loop sequences A-X-T and T-X-A, the pH_{mp} is considerably higher when X = T rather than A. Hence, the sequence context of the bases flanking the CCC segment influences the pK_a of the protonatable N3.

Thermodynamic properties were determined through the van't Hoff enthalpic ($\Delta H°$) and entropic ($\Delta S°$) contributions to the total free energy ($\Delta G°$) of thermally induce unfolding [27].

$$K_{eq} = \frac{\text{(Fraction i-Motif)}}{(1 - \text{Fraction i-Motif})} \quad (3)$$

$$\ln K_{eq} = -\frac{\Delta H^{\circ}}{RT} + \frac{\Delta S^{\circ}}{R} \quad (4)$$

$$\Delta G^{\circ} = \Delta H^{\circ} - T\Delta S^{\circ} \quad (5)$$

The data presented in Table 3 demonstrate a sequence context effect on the stability of these i-motif conformations. With the exception of $(CCCTTT)_4$, there is generally good correlation between T_m values and the respective $\Delta G°$ values, i. e., those with lower melting temperatures have lower free energies. It is interesting to note that those

oligomers with at least two T bases have the highest T_m values, and, in general, highest $\Delta G°$ values. Further, for the oligomers with one of the two T bases found in the middle of the loop sequence, the free energies are much higher than for the other oligomers. Since the thermal denaturation of $(CCCTTT)_4$ is incomplete at 95 °C, the data in Table 3 for this sequence may not reflect the actual thermodynamic parameters. Finally, the T_m

Table 3 Thermodynamic parameters (van't Hoff) for the temperature induced i-motif to single strand transition determined using T = 298K.

Oligomer	T_m (K)	ΔH^o (kcal/mol)	$T\Delta S^o$ (kcal/mol)[1]	ΔG^o (kcal/mol)
$(CCCAAA)_4$	55.6	37.2	33.7	3.5
$(CCCAAT)_4$	62.2	36.6	32.5	4.1
$(CCCATA)_4$	59.2	38.3	34.4	3.9
$(CCCATT)_4$	75.6	74.3	63.5	10.8
$(CCCTAA)_4$	66.2	41.9	36.8	5.1
$(CCCTAT)_4$	65.9	40.3	35.5	4.8
$(CCCTTA)_4$	67.1	103.5	90.7	12.8
$(CCCTTT)_4$	>75	16.6	13.7	2.9

[1]determined using T = 298K. Values were determined using Eqs. 3 -6. van't Hoff plots and associated line parameters are provided in Supplementary Materials.

and $\Delta G°$ values listed in Table 3 are in the range of values reported recently for intramolecular i-motif forming sequences with T loops of various length and location within that sequence [31].

Kaushik et al studied the unfolding of $(CCCTAA)_4$ in cacodylate buffer at various pH values using differential scanning calorimetery (DSC). Their findings revealed a

biphasic thermal transition that they attribute to the coexistence of two different complexes with different molecularities. For the i-motif complex (molecularity of 1), the thermodynamic parameters were determined with a T_m of 39.9 °C and ΔG° of 2.4 kcal/mol at pH = 5.2. In comparison to our values of 66.2 °C and 5.1 kcal/mol, respectively, and which were obtained through van't Hoff analysis of the CD optical melting profile in phosphate buffer at pH 5.0, reveals a significantly higher melting

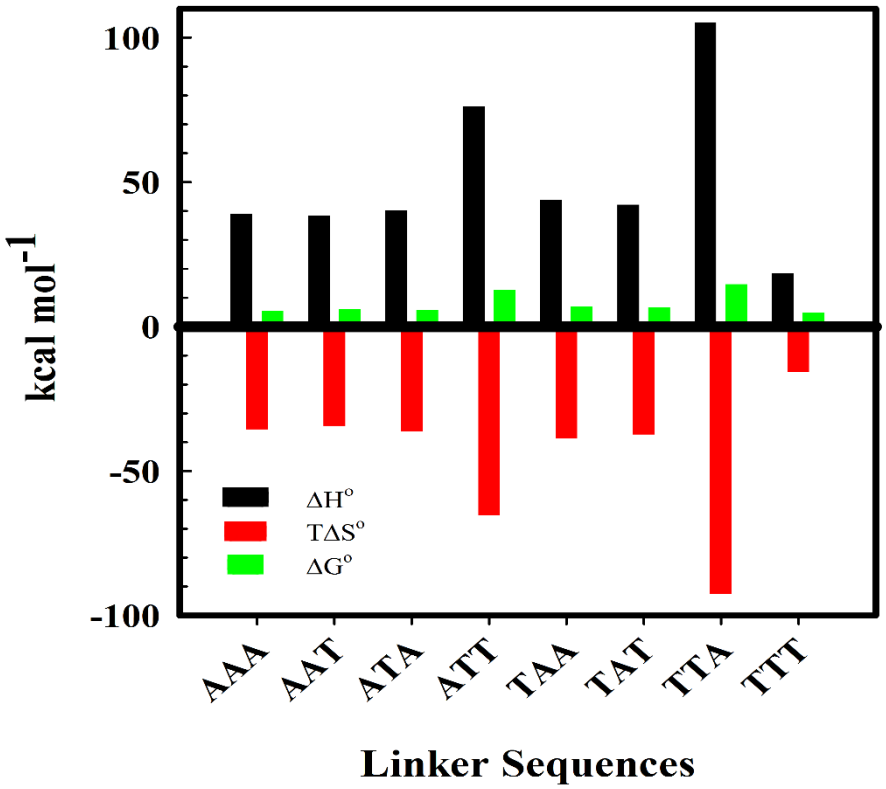

Figure 9 The enthalpy-entropy compensation, for the thermal denaturation of the monomer sequences, resulting in a small range of free energies.

temperature and slightly higher free energy. Further, our spectroscopic data were consistent with a two state transition for $(CCCTAA)_4$ for both the pH titration at 25 ^oC and thermal melt at pH 5.0. Although the difference in free energies is only slight, the other observed differences in melting temperature and statedness of the transition may simply be due to differences in sample preparation or experimental techniques (CD vs DSC) employed. We attempted DSC studies at pH of 5.0 but could not get reproducible scans even when using a sample preparation protocol similar to that reported in their paper.

The graphical representation of the data in Figure 9 reveals the enthalpy-entropy compensation resulting in a small range of free energies. Although this curve analysis approach is subject to up to 10% error, the general trends in T_m and calculated free energies support a sequence context effect on stability.

Supplemental Data

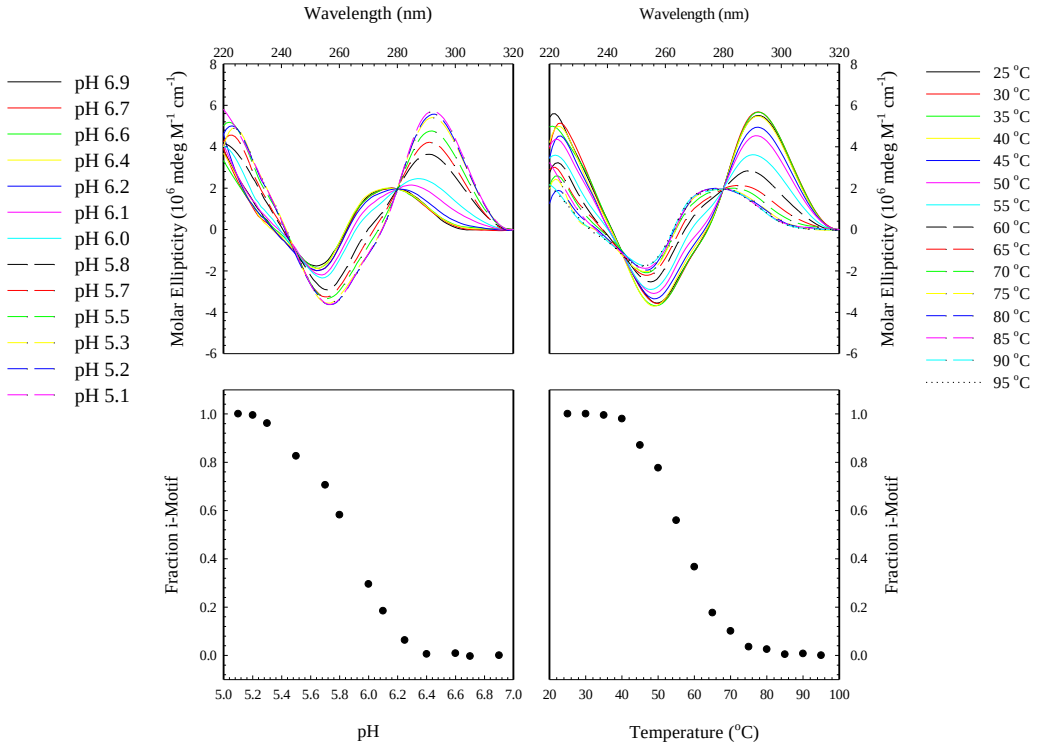

Figure S 1 CD spectra of $(CCCAAA)_4$ i-Motif to single strand transition as a function of pH and T. Upper Panel- CD spectra overlays for the sequence $(CCCAAA)_4$. Left Panel- pH induced single strand to i-motif transition from pH 7.0 to 5.0 at a constant 25 °C. Right Panel- Temperature induced i-motif to single strand transition from 25 °C to 95 °C at a constant pH of 5.0. Lower Panels- Fraction of i-Motif for the pH and T transitions. Left Panel- The fraction of i-motif of the pH induced transition, pH_{mp} = 0.5 at 292 nm which represents 50% folded i-motif conformations and 25% protonated C.. Right Panel- The fraction of i-motif of the T induced transition, T_m = 0.5 at 292 nm which represents 50% unfolded i-motif conformations and 25% unprotonated C.

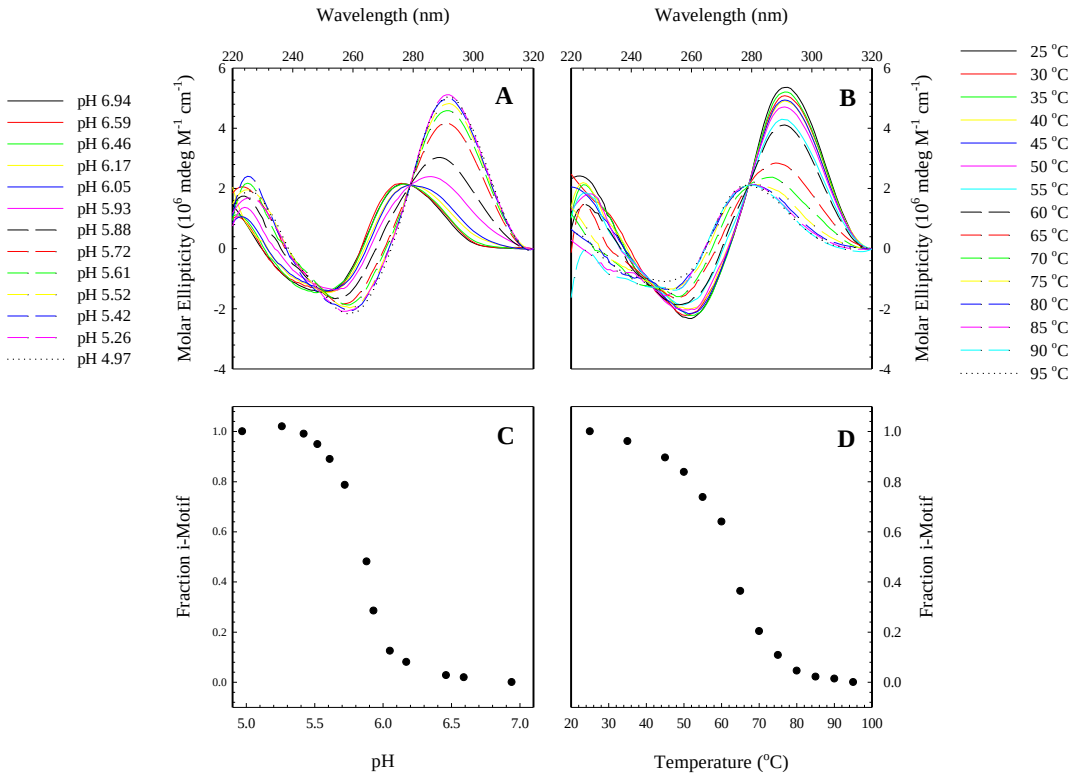

Figure S 2 CD spectra of $(CCCAAT)_4$ i-Motif to single strand transition as a function of pH and T. Upper Panel- CD spectra overlays for the sequence $(CCCAAT)_4$. Left Panel- pH induced single strand to i-motif transition from pH 7.0 to 5.0 at a constant 25 °C. Right Panel- Temperature induced i-motif to single strand transition from 25 °C to 95 °C at a constant pH of 5.0. Lower Panels- Fraction of i-Motif for the pH and T transitions. Left Panel- The fraction of i-motif of the pH induced transition, pH_{mp} = 0.5 at 292 nm which represents 50% folded i-motif conformations and 25% protonated C.. Right Panel- The fraction of i-motif of the T induced transition, T_m = 0.5 at 292 nm which represents 50% unfolded i-motif conformations and 25% unprotonated C.

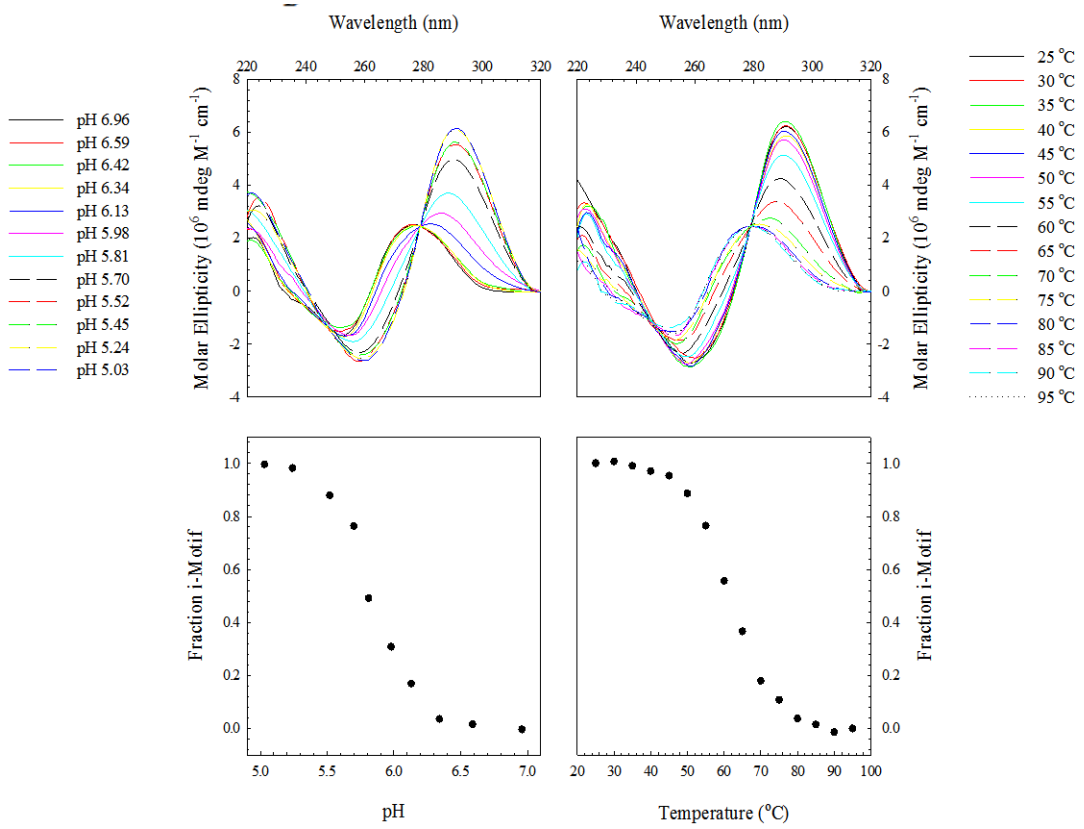

Figure S 3 CD spectra of $(CCCATA)_4$ i-Motif to single strand transition as a function of pH and T. Upper Panel- CD spectra overlays for the sequence $(CCCATA)_4$. Left Panel- pH induced single strand to i-motif transition from pH 7.0 to 5.0 at a constant 25 °C. Right Panel- Temperature induced i-motif to single strand transition from 25 °C to 95 °C at a constant pH of 5.0. Lower Panels- Fraction of i-Motif for the pH and T transitions. Left Panel- The fraction of i-motif of the pH induced transition, $pH_{mp} = 0.5$ at 292 nm which represents 50% folded i-motif conformations and 25% protonated C.. Right Panel- The fraction of i-motif of the T induced transition, $T_m = 0.5$ at 292 nm which represents 50% unfolded i-motif conformations and 25% unprotonated C.

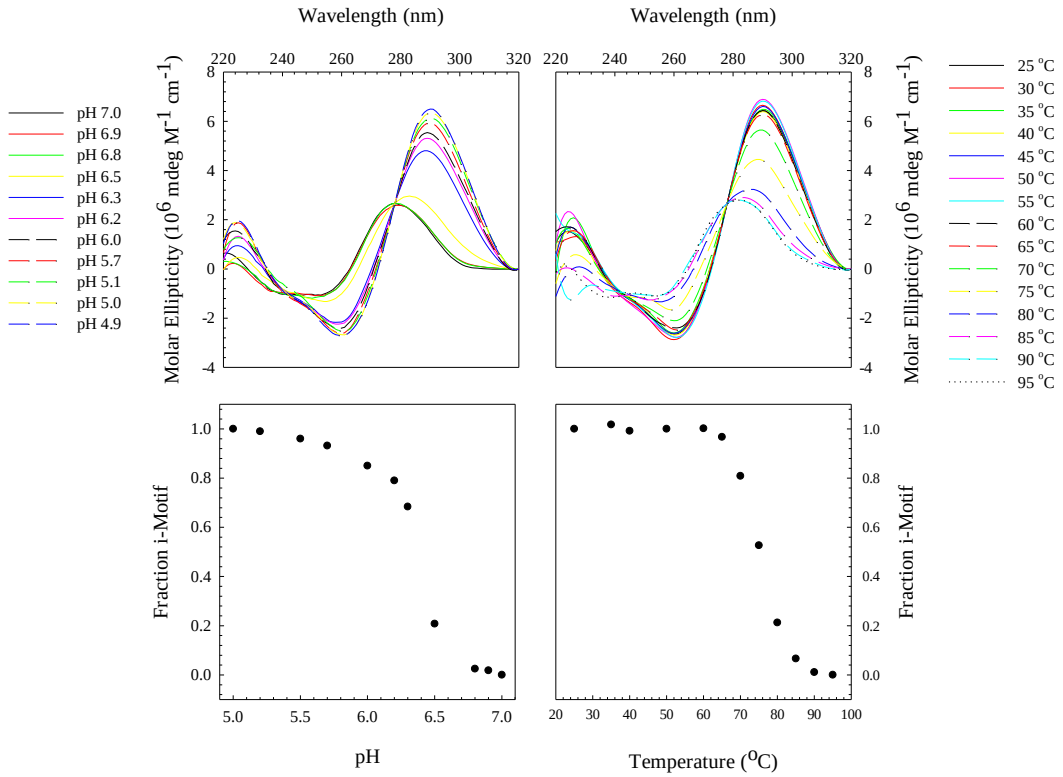

Figure S 4 CD spectra of $(CCCATT)_4$ i-Motif to single strand transition as a function of pH and T. Upper Panel- CD spectra overlays for the sequence $(CCCATT)_4$. Left Panel- pH induced single strand to i-motif transition from pH 7.0 to 5.0 at a constant 25 °C. Right Panel- Temperature induced i-motif to single strand transition from 25 °C to 95 °C at a constant pH of 5.0. Lower Panels- Fraction of i-Motif for the pH and T transitions. Left Panel- The fraction of i-motif of the pH induced transition, pH_{mp} = 0.5 at 292 nm which represents 50% folded i-motif conformations and 25% protonated C.. Right Panel- The fraction of i-motif of the T induced transition, T_m = 0.5 at 292 nm which represents 50% unfolded i-motif conformations and 25% unprotonated C.

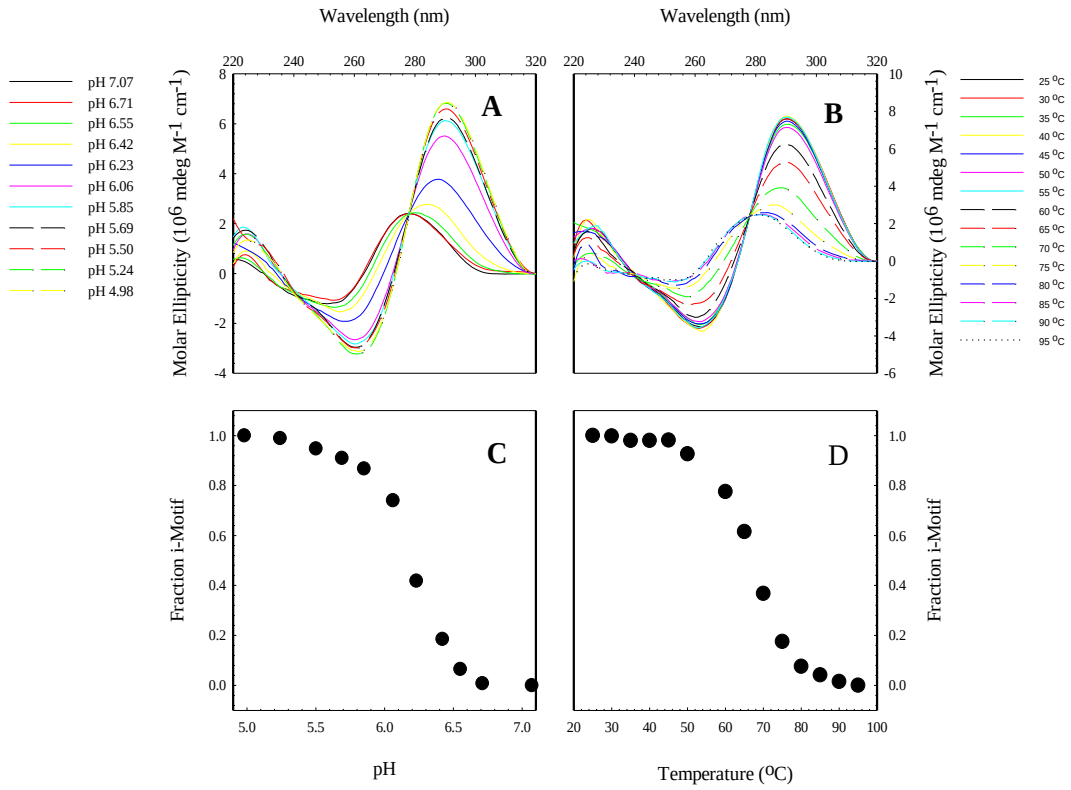

Figure S 5 CD spectra of $(CCCTAT)_4$ i-Motif to single strand transition as a function of pH and T. Upper Panel- CD spectra overlays for the sequence $(CCCTAT)_4$. Left Panel- pH induced single strand to i-motif transition from pH 7.0 to 5.0 at a constant 25 °C. Right Panel- Temperature induced i-motif to single strand transition from 25 °C to 95 °C at a constant pH of 5.0. Lower Panels- Fraction of i-Motif for the pH and T transitions. Left Panel- The fraction of i-motif of the pH induced transition, $pH_{mp} = 0.5$ at 292 nm which represents 50% folded i-motif conformations and 25% protonated C.. Right Panel- The fraction of i-motif of the T induced transition, $T_m = 0.5$ at 292 nm which represents 50% unfolded i-motif conformations and 25% unprotonated C.

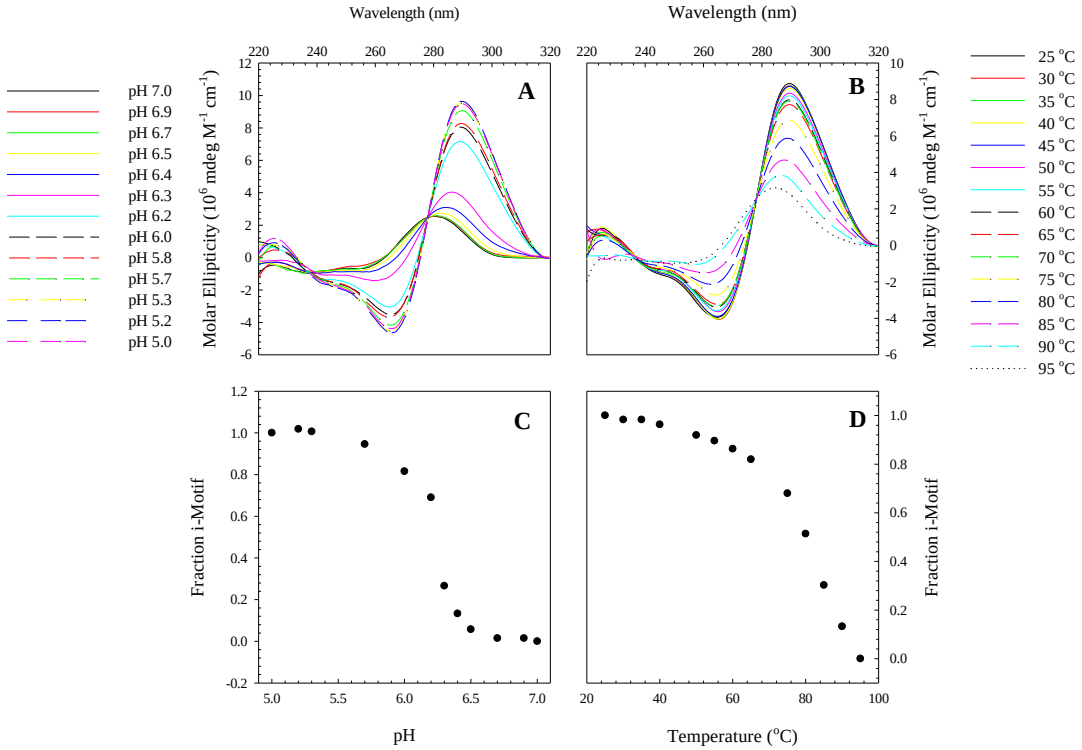

Figure S 6 CD spectra of $(CCCTTT)_4$ i-Motif to single strand transition as a function of pH and T. Upper Panel- CD spectra overlays for the sequence $(CCCTTT)_4$. Left Panel- pH induced single strand to i-motif transition from pH 7.0 to 5.0 at a constant 25 °C. Right Panel- Temperature induced i-motif to single strand transition from 25 °C to 95 °C at a constant pH of 5.0. Lower Panels- Fraction of i-Motif for the pH and T transitions. Left Panel- The fraction of i-motif of the pH induced transition, pH_{mp} = 0.5 at 292 nm which represents 50% folded i-motif conformations and 25% protonated C.. Right Panel- The fraction of i-motif of the T induced transition, T_m = 0.5 at 292 nm which represents 50% unfolded i-motif conformations and 25% unprotonated C.

van't Hoff Plots for $(CCCAXX)_4$, where X = A and/or T

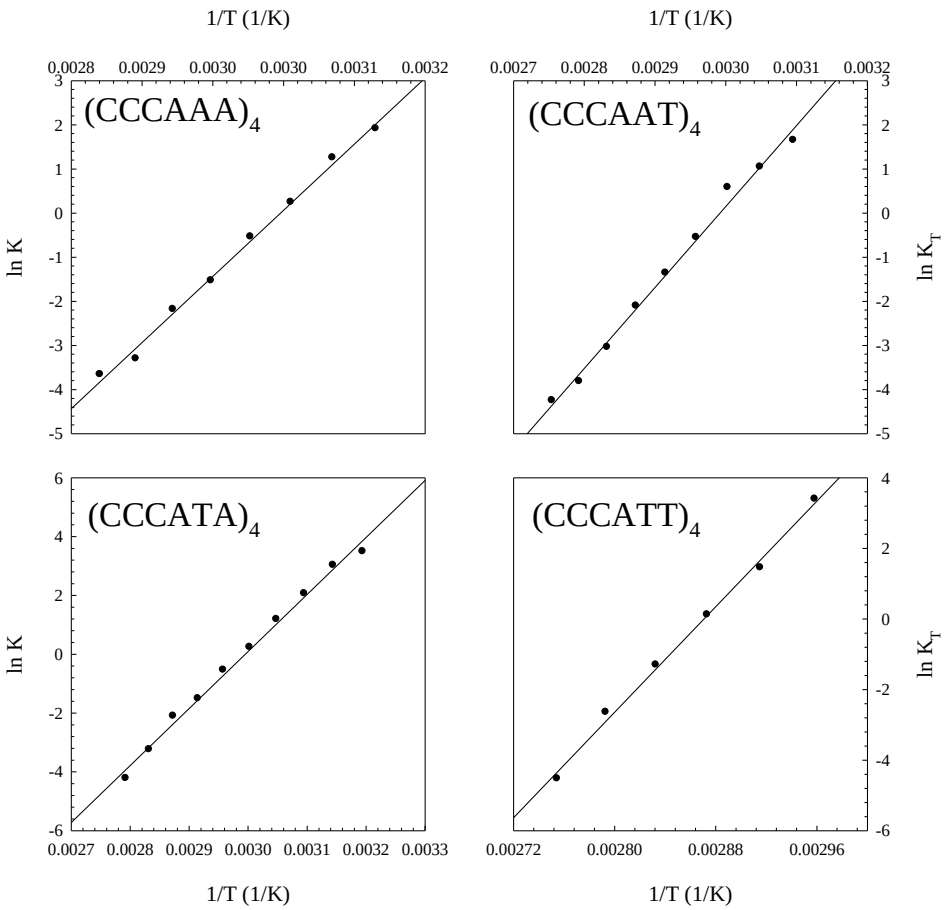

Figure S 7 van't Hoff plots for DNA sequences $(CCCAXX)_4$, where X = A and/or T. Line parameters are given in Table 1S.

van't Hoff Plots for $(CCCTXX)_4$, where X = A and/or T

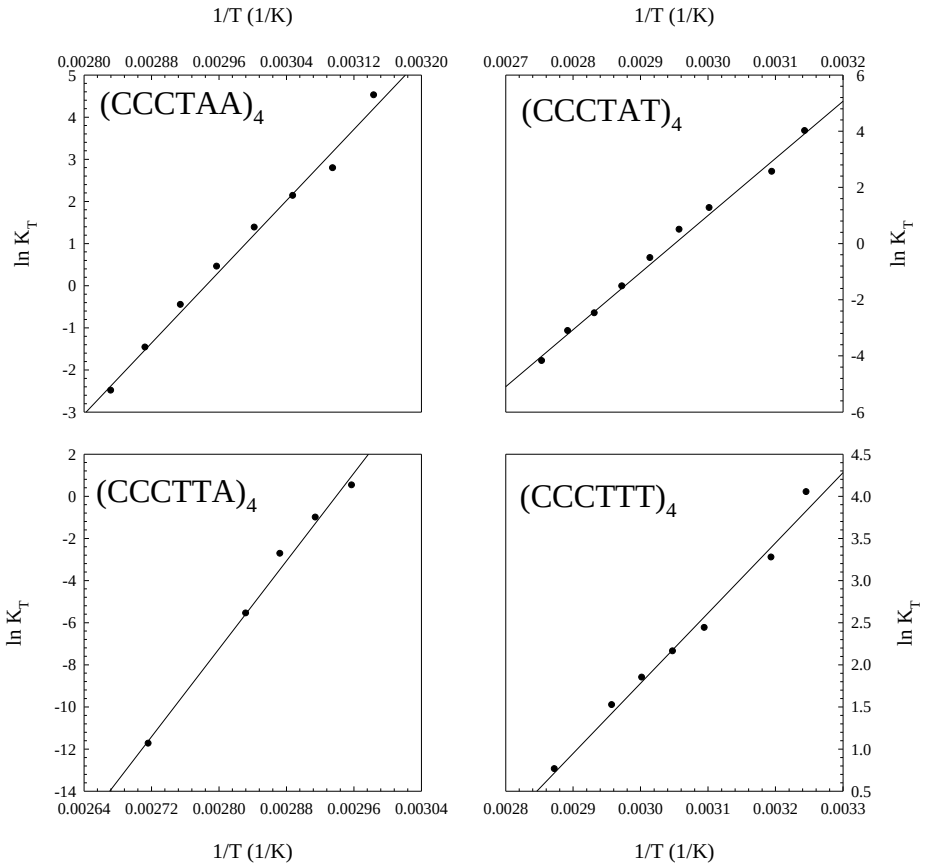

Figure S 8 van't Hoff plots for DNA sequences $(CCCTXX)_4$, where X = A and/or T. Line parameters are given in Table 1S.

Table S 1 Line parameters for the van't Hoff plots in Figures S7 and S8.

Sequence	ln K_T vs 1/T Plots (at pH 5.0)		
	slope	y-int	r^2
$(CCCAAA)_4$	1.87×10^4	-56.9	0.995
$(CCCAAT)_4$	1.84×10^4	-54.9	0.993
$(CCCATA)_4$	1.93×10^4	-58.1	0.993
$(CCCATT)_4$	3.74×10^4	-107.3	0.996
(CCCTAA)	2.11×10^4	-62.21	0.990
$(CCCTAT)_4$	2.03×10^4	-59.9	0.993
$(CCCTTA)_4$	5.21×10^4	-153.2	0.991
$(CCCTTT)_4$	8.33×10^3	-23.2	0.987

CHAPTER IV

STRUCTURAL AND THERMODYNAMIC PROPERTIES OF THE HUMAN TELOMERIC SEQUENCE: A DNA I-MOTIF DIMER $(CCCTAA)_2$ AND TETRAMER $(CCCTAA)_1$

To gain a better understanding of the how the loops in the monomeric structure play a role in the stability of the i-motif folded structure, spectroscopic studies were carried out on the sequence $(CCCTAA)_X$, were X = 2 or 1. The $(CCCTAA)_2$ sequence, has a molecularity of two and forms an i-motif structure from two intercalated strands of DNA. The dimer is capable of forming multiple possible topologies, some of which are shown in Figure 10. Motif A and B, of Figure 10, are held together by interstrand $C:C^+$ base pairs. In motif A the loops are on the same side as one another which could allow for base pair interactions within the loops. For motif B the two strands are opposite each other allowing for each loop to be on opposite sides and for loop interaction to be nearly impossible. Motif C is an intrastrand conformation where the $C:C^+$ base pairs are formed by a single strand folding back on itself, two intrastrand molecules then intercalate with each other to form the i-motif [32].

Figure 11 shows the overlays of the CD spectra of $(CCCTAA)_2$ at 25 °C as a function of pH and at pH 5.0 as a function of temperature. The pH induced transition, seen in the upper left hand panel, goes from the single stranded DNA at pH 7.0 to the i-motif conformation at pH 5.0. The spectral properties for this sequence, listed in Table 5,

show a λ_{max} between 278 at pH 7.0, shifting 279 at pH 5.0, with θ_{max} of 1.57 x 10^6 mdeg M^{-1} cm^{-1} shifting to 3.59 x 10^6 mdeg M^{-1} cm^{-1}, respectively.

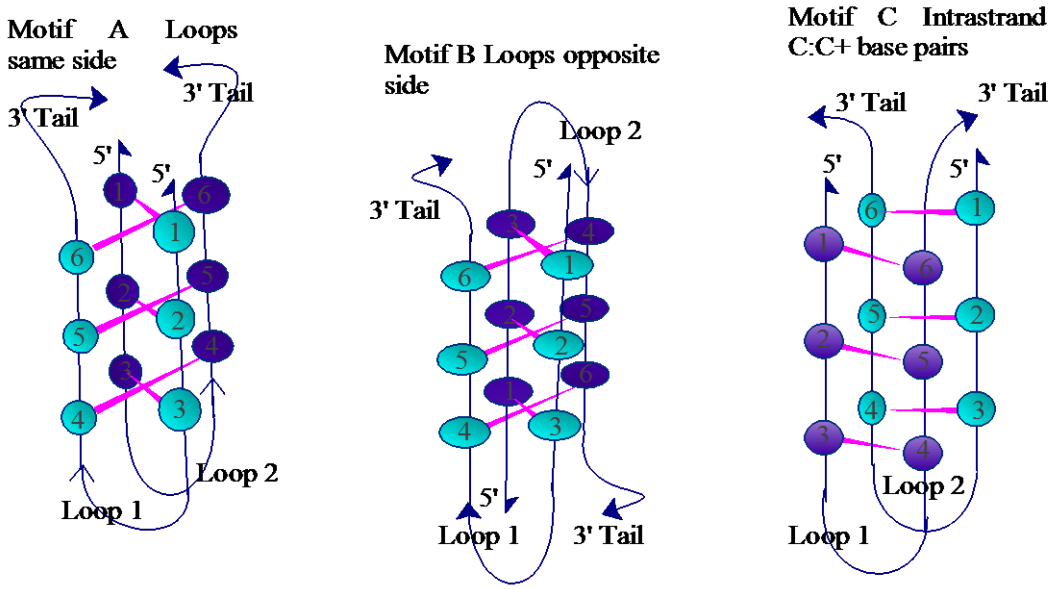

Figure 10 I-motif Dimer Conformation possibilities. For all motifs, the teal circles make up one molecule of DNA and the dark blue circles the second strand of DNA. Motif A and B are held together by interstrand C:C+ base pairs and their loops are on the same side and opposite sides, respectively. Motif C is held together by intrastrand C:C+ base pairs with loops of the same side of the molecule.

Lower left panel of Figure 11 is the fraction of i-motif pH titration curve obtained using eq. 1 at 292 nm. The midpoint of the transitions, pH_{mp}, when there are equal concentrations of the single strand and i-motif species is when the fraction of i-motif equals 0.5. For this oligomer, the pH_{mp} appears to be 5.5, however for this dimer sequence, it is possible that the structure has not completely transitioned to the i-motif species at pH 5.0 due to the lack of a well-defined curve at where the i-fraction = 1.

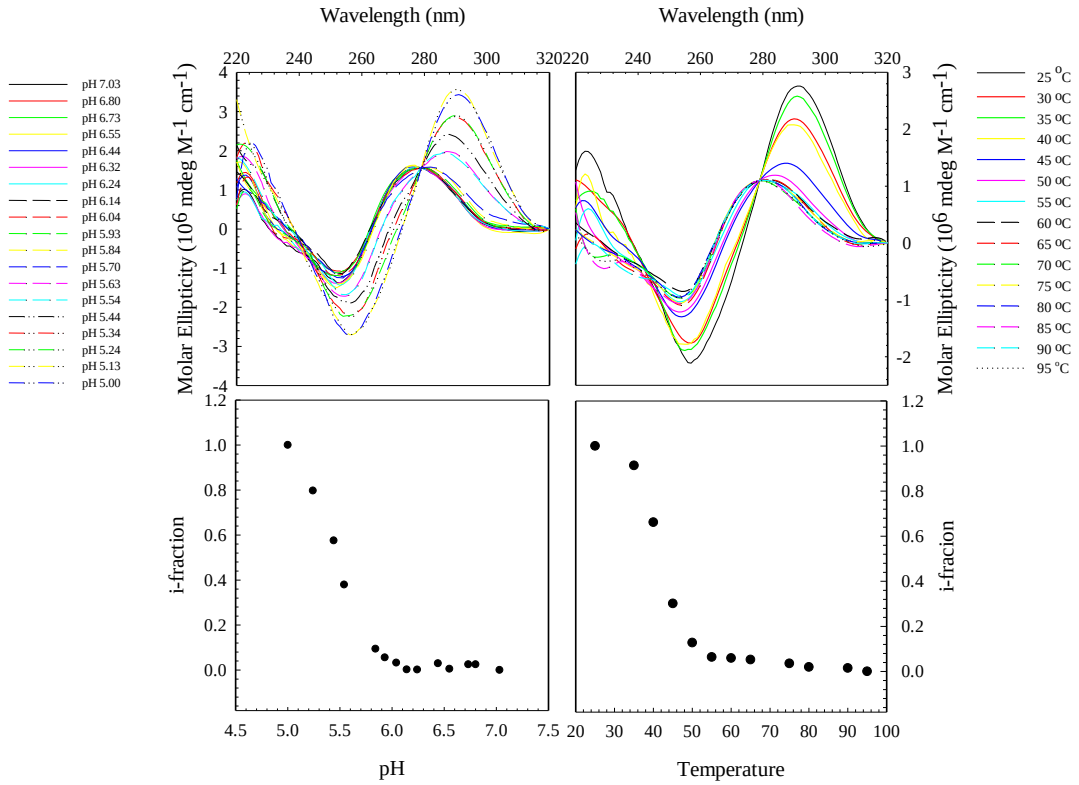

Figure 11 CD spectra for sequence $(CCCTAA)_2$ i-Motif to single strand transition as a function of pH and T. Upper Panel- CD spectra overlays for the sequence $(CCCTAT)_2$. Left Panel- pH induced single strand to i-motif transition from pH 7.0 to 5.0 at a constant 25 °C. Right Panel- Temperature induced i-motif to single strand transition from 25 °C to 95 °C at a constant pH of 5.0. Lower Panels- Fraction of i-Motif for the pH and T transitions. Left Panel- The fraction of i-motif of the pH induced transition, $pH_{mp} = 0.5$ at 292 nm which represents 50% folded i-motif conformations and 25% protonated C.. Right Panel- The fraction of i-motif of the T induced transition, $T_m = 0.5$ at 292 nm which represents 50% unfolded i-motif conformations and 25% unprotonated C.

More research is needed to determine if the reaction has gone to completion at this point In the upper right panel the thermal denaturation graph of the i-motif to single strand transition at pH 5.0 from 25 °C to 95 °C. The optical melting curve seen in the lower right panel, was obtained using the molar ellipticity at 292 nm in eq. 2 in order to obtain the T_m of the thermally induced transition to a single strand. Assuming a two state

transition, the T_m represents the point in the transition when there are equal concentrations of i-motif and single stranded species. The T_m for the dimer sequence occurs at T = 41.9 °C. The presence of a curve at i-fraction = one and the steep slope of the curve after that suggests that the transition is complete. Furthermore, the presence of an isoelliptic point at 279 nm with molar ellipticity of 1.58×10^6 mdeg M^{-1} cm^{-1} suggests that the folding and unfolding transitions are two state transition.

The normalization process discussed in Chapter III was carried out on the dimer sequence. First, the raw optical melt data was corrected for concentration to smooth out the noise created during the experimentation. This correction revealed a sharp isoelliptic point, which was then used to normalize the pH transition spectra.

This normalization process was attempted on the CD spectra for the $(CCCTAA)_1$ sequence, forming an i-motif molecule with the molecularity of four. However, after correction for the concentration, a defined isoelliptic point was not observed. The data are shown in Figure 12 where the upper left panel is of the CD spectra at 25 °C as a function of pH from 5.0 to 7.0. The upper right panel is of the CD optical melt spectra at pH 5.0 as a function of temperature. The lower panels, from left to right, are the fraction of i-motif versus pH and T. It is obvious, that neither the pH titration nor the optical melt have any clearly defined points, in fact the consecutive order of scans seen in previous spectra, as well as, a hypochromaticity of the thermally induced spectra of the i-motif unfolding spectral characteristics are absent from this data. The fraction of i-motif plots were obtain using eq. 1 and eq. 2 using the λ_{max} of 279 nm to obtain the molar ellipticities

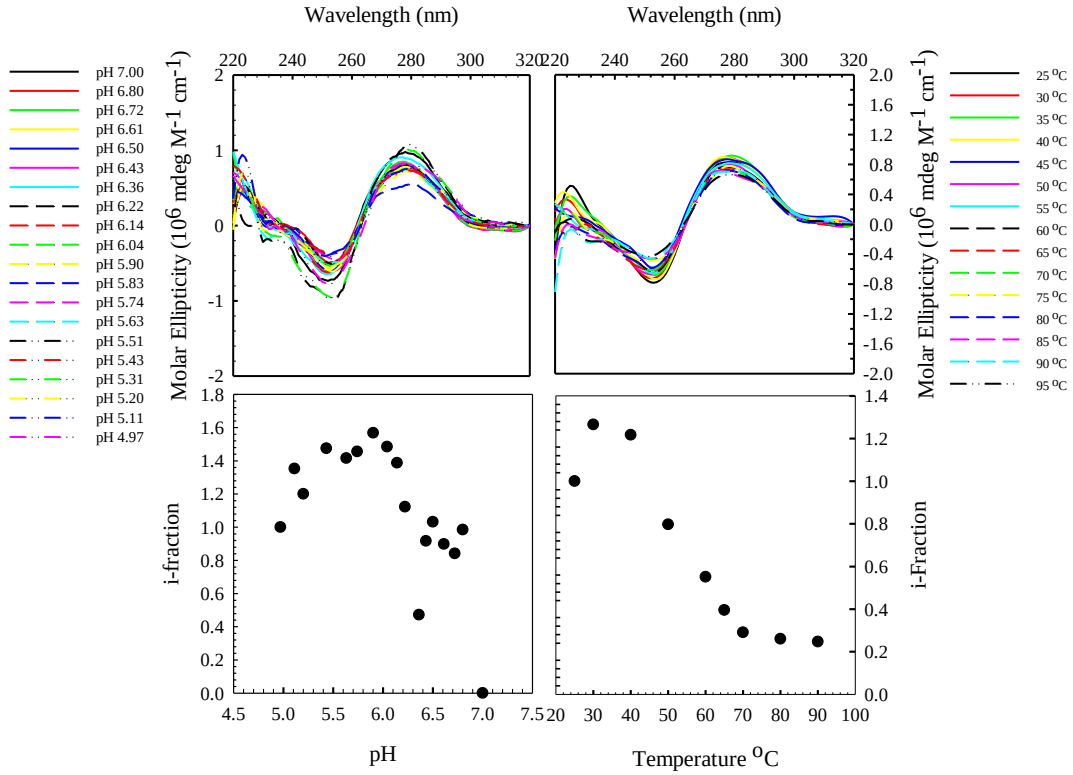

Figure 12 CD spectra for sequence $(CCCTAA)_1$ i-Motif to single strand transition as a function of pH and T. Upper Panel- CD spectra overlays for the sequence $(CCCTAT)_1$. Left Panel- pH induced single strand to i-motif transition from pH 7.0 to 5.0 at a constant 25 °C. Right Panel- Temperature induced i-motif to single strand transition from 25 °C to 95 °C at a constant pH of 5.0. Lower Panels- Fraction of i-Motif for the pH and T transitions.

for any pH solution and for any T solution, respectively as used in the equations. Clearly, from observing the fraction of i-motif versus pH plot the absence of a sigmodal curve, the assumption of a two state transition would be incorrect. Instead, it is likely that there are multiple steps involved for the hemi protonation of two single stranded oligos, and for the ultimate formation of a four-stranded structure. Furthermore, in the lower right panel, the hook at the top of the graph starts at one and then goes as high as 1.3 before coming back

down, thus leading to the assumption once again that the i-motif to single strand denaturation of this sequence is multi state and not a two state transition.

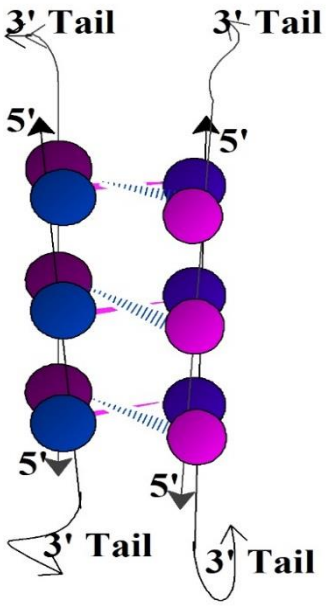

Figure 13 Tetramer conformation formed from four single strands of (CCCTAA) where the C:C+ base pair form from two parallel strands, which then intercalate with another pair of parallel strands, in an antiparallel fashion.

Gehring et al. initial discovery of the i-motif structure was with the sequence (TCCCCC), which forms an intercalated four stranded i-motif structure [33]. Through Overhouser NMR, the structure was elucidated. Thus, through NMR and gel electrophoresis the structure formation has been suggested for many other sequences as well. The structure of the sequence $(CCCTAA)_1$ studied here and seen in Figure 13, is two duplexes with parallel strands forming interstrand $C:C^+$ base pairs which then intercalate in an antiparallel fashion, with one another [34]. Studies must be performed in future work to gain an understanding of the folding and unfolding mechanism for this structure.

Table 4 Thermodynamic parameters (van't Hoff) for the temperature induced i-motif to single strand transition.

Oligomer	**T_m (K)**	**ΔH° (kcal/mol)**	**$T\Delta S^\circ$ (kcal/mol)[1]**	**ΔG° (kcal/mol)**
$(CCCTAA)_2$	41.9	21.1	19.4	1.7
$(CCCTAA)_1$	60.3	27.3	24.2	3.1

[1]determined using T = 298K. Values were determined using Eqs. 3 -6. van't Hoff plots and associated line parameters in the supplemental data.

Interestingly, the T_m for the dimer, at 41.9 °C comes out to be a difference of 13.7 degrees from the lowest T_m of the monomers studied above. The tetramer T_m = 61.7 °C which is higher than two of the monomer sequences, suggesting a greater stability in the tetramer structure than in that of the monomer sequences $(CCCAAA)_4$ and $(CCCATA)_4$.

Supplemental Data

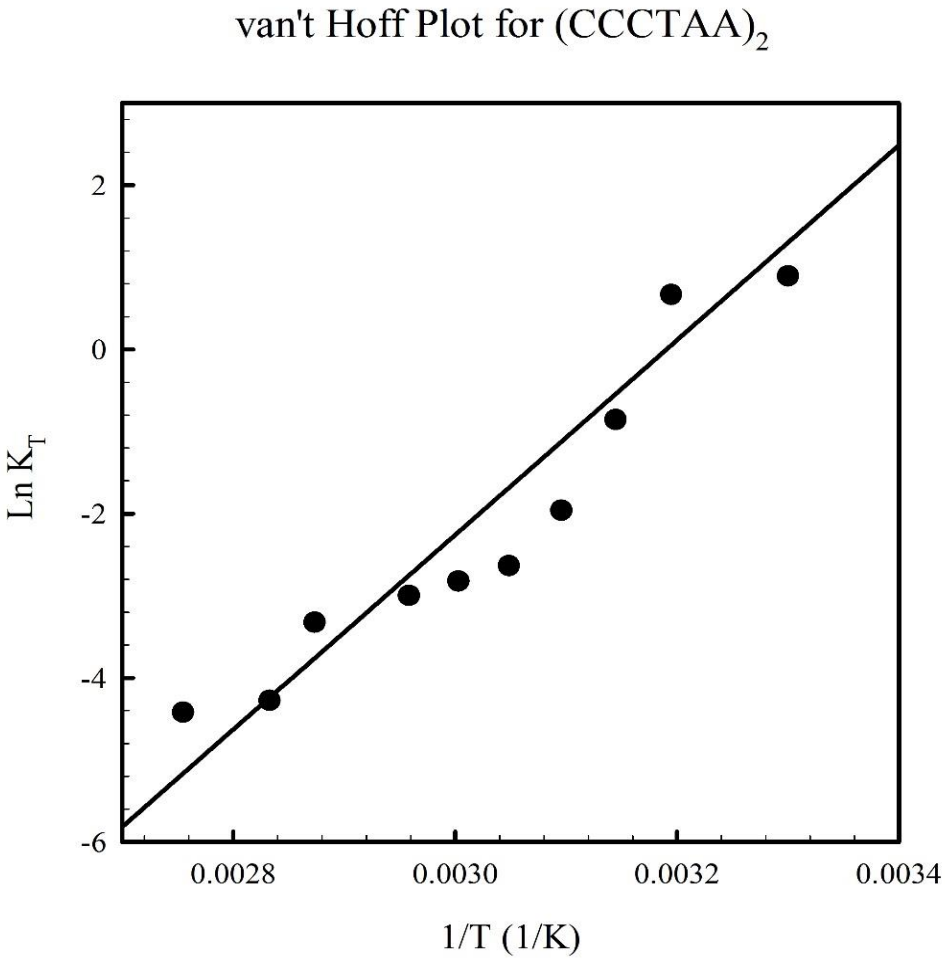

Figure S 9 van't Hoff plots for DNA sequences $(CCCTAA)_2$. Line parameters are given in Table S2.

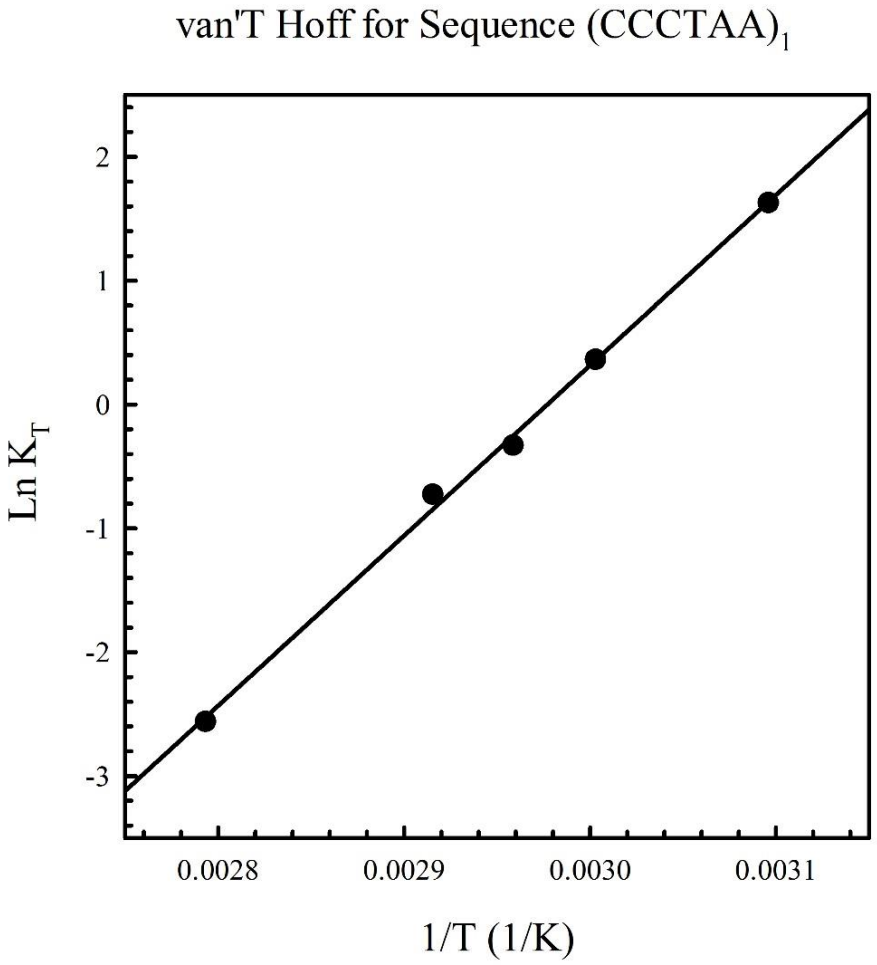

Figure S 10 van't Hoff plots for DNA sequences $(CCCTAA)_1$. Line parameters are given in Table S2.

Table S 2 Line parameters for the van't Hoff plots in Figures S9 and S10.

Sequence	ln K_T vs 1/T Plots (at pH 5.0)		
	slope	y-int	r^2
$(CCCTAA)_2$	1.06×10^4	-32.7	0.915
$(CCCTAA)_1$	1.37×10^4	-40.9	0.998

CHAPTER V

CONCLUSION

In order for these C rich strands to fold into the i-motif structure, protons must be present in solution. This proton induced folding is described below by three equilibria designated as Scheme I shown for the formation of the first hydrogen bond. For these experiments, all DNA oligomers were dissolved in standard phosphate buffer of 10 mM phosphate, 115 mM K^+, 0.1 mM EDTA and pH ranging from 7.0 to 5.0.

$$HBuf + H_2O \rightleftharpoons Buf^- + H_3O^+$$

$$C + H_3O^+ \rightleftharpoons CH^+ + H_2O$$

$$C + CH^+ \rightleftharpoons C{:}CH^+$$

Scheme I

For the first equilibrium, we only need be concerned with the second ionization of phosphoric acid because the pH range we are concerned with is between pH 7.0 and pH 5.0. Thus, at pH 7.0, the $[H_2PO_4^-]$ is 6.1 mM, the $[HPO_4^{-2}]$ is 3.9 mM and the $[H_3O^+]$ is 1.0×10^{-7} mM and at pH 5.0, those concentrations are 9.9 mM, 0.1 mM and 1.0×10^{-5} mM, respectively. Further, the total DNA concentration is around 1.0×10^{-2} mM.

Hence, at any pH between 7.0 and 5.0, the major proton donor is most likely $H_2PO_4^{-1}$ (i.e., HBuf in the above scheme).

The second equilibrium above is the protonation of N3 of any C in the strand. The use of hydronium rather than HBuf as the proton source as seen in the equilibrium equation above, allows for the use of the pK_a of N3 of C. The pK_a of free cytosine is around 4.3 [27], which more than likely is the pK_a of the first C to be protonated in solution. However, when the protonation of the first C takes place, what effect does this have on the adjacent and surrounding C's in solution?

For the third equilibrium, where following the protonation of the first C, a $C:CH^+$ base pair is formed. Assuming a cooperative process, at the time the first C is protonated, the structure begins to fold and form. Thus, both the base pair formation and the commencement of the conformational folding are the third equilibrium. This folding process is a favorable entropic process because it releases water, thus compensating for the unfavorable conformational entropy. The pH_{mp} values listed in Table 2 reflect the pK_a values for N3 at the midpoint of the transition for each particular oligomer. It is likely that the actual pK_a for any N3 depends upon several factors in addition to flanking sequence effects.

The net reaction for the formation of the first base pair is shown below:

$$2C + HBuf \rightleftharpoons C:CH^+ + Buf^-$$

A simple application of Hess's law allows deconstruction of the total enthalpy (ΔH_t) for the above reaction:

$$\Delta H_t = \Delta H_{HBuf} + \Delta H_{prot} + \Delta H_{bp} \quad (6)$$

where ΔH_{HBuf} is the enthalpy of deprotonation of the buffer, ΔH_{prot} is the enthalpy of protonation of N3 of C and ΔH_{bp} is the enthalpy of $C:CH^+$ base pair formation - each corresponding to the respective equilibria in Scheme I. The values listed in Table 3 for the enthalpy can be considered the ΔH_{bp} per oligomer, since the formation of the hydrogen bond and the folding occur simultaneously. ITC protocols are currently being developed in our lab, to evaluate the ΔH_t and ΔH_{HBuf} from eq. 6.

The net reactions taking place during the reaction are:

$$\text{Single Strand} \underset{-H^+}{\overset{+H^+}{\rightleftharpoons}} \text{i-Motif} \underset{-q}{\overset{+q}{\rightleftharpoons}} \text{Single Strand}$$

The first equilibrium depicts the protonation induced folding to the i-motif while the second equilibrium is the thermally induced unfolding back to the single strand. The enthalpy for the thermally induced i-motif to single strand transition at pH 5.0 is not equal to the enthalpy for the *de*protonation induced unfolding of the i-motif to single

strand at 25 °C. Therefore, these events must be evaluated separately through ITC experiments.

The transitions from single stranded on one end of the equilibrium to the other end becoming once again single stranded, through a pH change and then a temperature change, appears to be a cooperative process for seven of the eight monomer sequences and for the dimer sequence as well. Through the examination of Table 2 the pH at the midpoints of the protonation induced transitions and the temperature at the midpoints of the thermally induced transitions it is easy to reason there is a sequence context dependence. However, after conducting the CD spectral experiments of the dimer and tetramer, it was seen that it is possible that stability is effected by more than just the sequence context. Additionally, the molecularity and the total number of loops present could contribute to the overall stabilities of the i-motif formation.

Through a comparison of the dimer pHmp and Tm to that of the eight monomers, seen in Figure 13, it is clear that the dimer is far less stable than any of the monomer conformations, with a ΔT_m of 13.7 degrees from the lowest Tm point , and a ΔpH_{mp} of 0.32 from the lowest pH_{mp}. Major players in the stability of the i-motif formation seem to be the molecularity of the structure, (the monomer with a molecularity of one is more stable than both the dimer with a molecularity of two, and the tetramer with a molecularity of four), not only the presence of the loops but also the number of loops and thus greater opportunity for base pairing and interaction. As well as, the sequence

context within the loop regions, the more T present the more stable the structure, and the more A present the quicker the structure will deconstruct.

Unfortunately, with the tetramer sequence, not enough information could be gathered from the experiments conducted at this point. Further experiments must be carried out in order to gain enough information to draw conclusions to its contributing factors of stability. Additionally, as stated above, ITC experiments are currently being conducted to determine the ΔH_t and ΔH_{HBuf}.

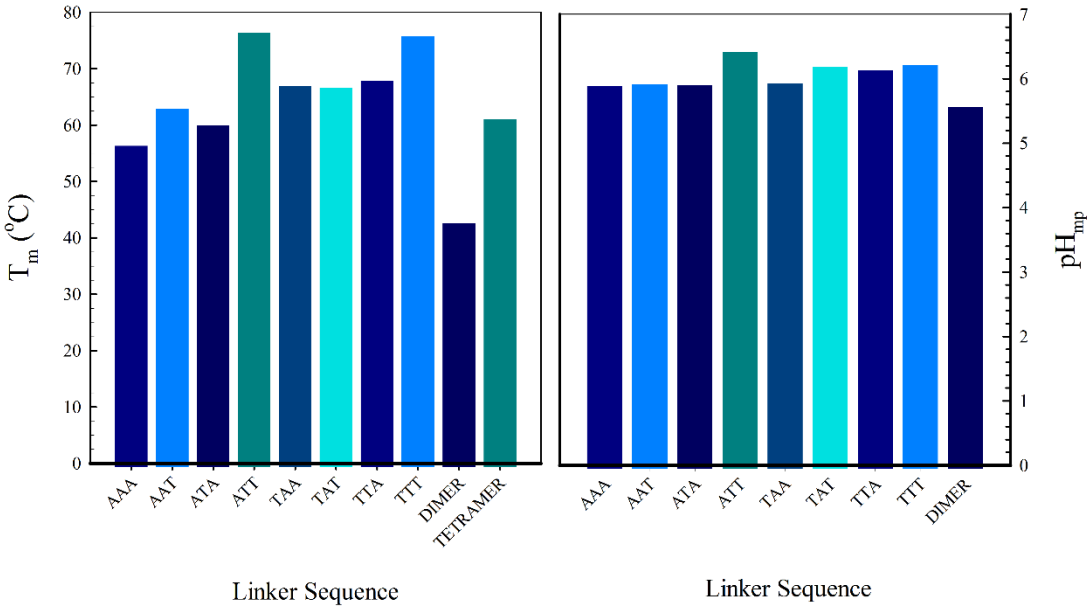

Figure 14 Comparison of the pH_{mp} and T_m values for all monomer, dimer and tetramer sequences. Left Panel- A summary of Tm values for all monomer, dimer and tetramer sequences, ranging from 41.9 to 80 °C. Right Panel- A summary of pH_{mp} values for the monomer and dimer sequences, ranging from 5.50 to 6.35. For the tetramer, a pH_{mp} value was not observed.

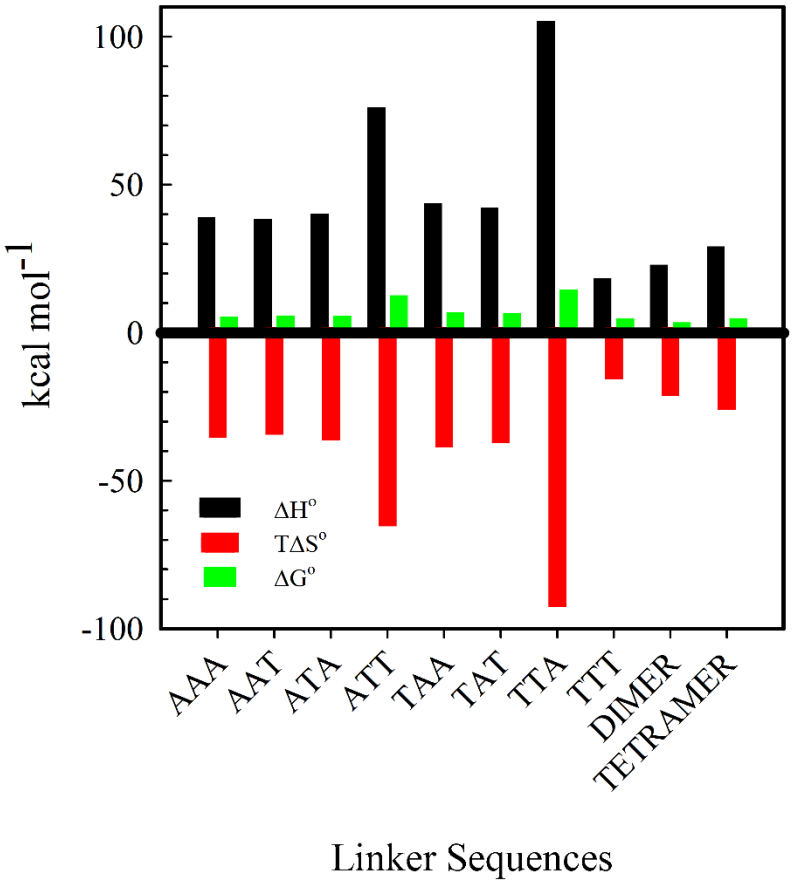

Figure 15 The enthalpy-entropy compensation, for the thermal denaturation of all oligos studied, resulting in a small range of free energies.

REFERENCES

1. Meyers, S. *Signature in the Cell;* Harper Colins: 2009; Vol. 1, pp 624.

2. Voet, D.; Voet, J. G.; Pratt, C. *Fundamentals of Biochemistry;* John Wiley & Sons, Inc.: Hoboken, NJ, 2013; .

3. Anh Tuan Phan,Vitaly Kuryavyi, Dinshaw J.Patel DNA architecture: from G to Z. **2006**.

4. Reilly, S. M.; Morgan, R. K.; Brooks, T. A.; Wadkins, R. M. Effect of interior loop length on the thermal stability and p K a of i-motif DNA. *Biochemistry (N. Y.)* **2015**, *54*, 1364-1370.

5. Phan, A. T.; Mergny, J. L. Human telomeric DNA: G-quadruplex, i-motif and Watson-Crick double helix. *Nucleic Acids Res.* **2002**, *30*, 4618-4625.

6. Henry A. Day,Pavlos Pavlou, Zoë A.E.Waller i-Motif DNA: Structure, stability and targeting with ligands. *Bioorg. Med. Chem.* **2014**, *22*, 4407–4418.

7. Li, X.; Peng, Y.; Ren, J.; Qu, X. Carboxyl-modified single-walled carbon nanotubes selectively induce human telomeric i-motif formation. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 19658-19663.

8. Anonymous.

9. Xiaogang Han,Jean-Louis Leroy and Maurice GueÂron An Intramolecular i-Motif: the Solution Structure and Base-pair Opening Kinetics of d(5mCCT3CCT3ACCT3CC). *J. Mol. Biol.* **1998**, *278*, 949-965.

10. Crothers, D. M.; Bloomfield, V. A.; Tinoco, I. *Nucleic acids: structures, properties, and functions;* University science books: 2000; .

11. Bhattacharyya, M. K.; Lustig, A. J. Telomere dynamics in genome stability. *Trends Biochem. Sci.* **2006**, *31*, 114-122.

12. Cesare, A. J.; Reddel, R. R. Telomere uncapping and alternative lengthening of telomeres. *Mech. Ageing Dev.* **2008**, *129*, 99-108.

13. De Cian, A.; Lacroix, L.; Douarre, C.; Temime-Smaali, N.; Trentesaux, C.; Riou, J.; Mergny, J. Targeting telomeres and telomerase. *Biochimie* **2008**, *90*, 131-155.

14. Yan Xu,a Yuuka Hirao,b Yoshifumi Nishimurab and Hiroshi Sugiyama I-motif and quadruplex-based device that can control a protein release or bind and release small molecule to influence biological processes. *Bioorg. Med. Chem.* **2007**, *15*, 1275-1279.

15. Voet, D.; Voet, J. G.; Pratt, C. *Fundamentals of Biochemistry;* John Wiley & Sons, Inc.: Hoboken, NJ, 2013; .

16. Anh TuaÃn Phan,Maurice GueÂron and Jean-Louis Leroy The Solution Structure and Internal Motions of a Fragment of the Cytidine-rich Strand of the Human Telomere. *J. Mol. Biol.* **2000**, *299*, 123-144.

17. Sharma, V. D.; Aifuwa, E. O.; Heiney, P. A.; Ilies, M. A. Interfacial engineering of pyridinium gemini surfactants for the generation of synthetic transfection systems. *Biomaterials* **2013**, *34*, 6906-6921.

18. Fernández, S.; Eritja, R.; Aviñó, A.; Jaumot, J.; Gargallo, R. Influence of pH, temperature and the cationic porphyrin TMPyP4 on the stability of the i-motif formed by the 5'-(C3TA2)4-3' sequence of the human telomere. *Int. J. Biol. Macromol.* **2011**, *49*, 729-736.

19. Arquès, D. G.; Fallot, J.; Michel, C. J. Identification of several types of periodicities in the collagens and their simulation. *Int. J. Biol. Macromol.* **1996**, *19*, 131-138.

20. Yan Xu,a Yuuka Hirao,b Yoshifumi Nishimurab and Hiroshi Sugiyama I-motif and quadruplex-based device that can control a protein release or bind and release small molecule to influence biological processes. *Bioorg. Med. Chem.* **2007**, *15*, 1275-1279.

21. Kaushik, M.; Prasad, M.; Kaushik, S.; Singh, A.; Kukreti, S. Structural transition from dimeric to tetrameric i-motif, caused by the presence of TAA at the 3′-end of human telomeric C-rich sequence. *Biopolymers* **2010**, *93*, 150-160.

22. Gurung, S. P.; Schwarz, C.; Hall, J. P.; Cardin, C. J.; Brazier, J. A. The importance of loop length on the stability of i-motif structures. *Chemical Communications* **2015**, *51*, 5630-5632.

23. Lieblein, A. L.; Buck, J.; Schlepckow, K.; Fürtig, B.; Schwalbe, H. Time-resolved NMR spectroscopic studies of DNA i-motif folding reveal kinetic partitioning. *Angewandte Chemie - International Edition* **2012**, *51*, 250-253.

24. Mikeal McKim, Alexander Buxton, Amanda Metz, Courtney Johnson, and Richard D. Sheardy. Structural and Thermodynamic Studies of the DNA i-Motif: Sequence Context Effects. **2014**.

25. Xiaogang Han,Jean-Louis Leroy and Maurice GueÂron An Intramolecular i-Motif: the Solution Structure and Base-pair Opening Kinetics of d(5mCCT3CCT3ACCT3CC). *J. Mol. Biol.* **1998**, *278*, 949-965.

26. Edwards, E. L.; Patrick, M. H.; Ratliff, R. L.; Gray, D. M. A·T and C·C+ base pairs can form simultaneously in a novel multistranded DNA complex. *Biochemistry (N. Y.)* **1990**, *29*, 828-836.

27. Pack, G. R.; Wong, L.; Lamm, G. pKa of cytosine on the third strand of triplex DNA: Preliminary poisson-boltzmann calculations. *International Journal of Quantum Chemistry* **1998**, *70*, 1177-1184.

28. Pack, G. R.; Wong, L.; Lamm, G. pKa of cytosine on the third strand of triplex DNA: Preliminary poisson-boltzmann calculations. *International Journal of Quantum Chemistry* **1998**, *70*, 1177-1184.

29. Nikolava, E. N.; Goh, G. B.; Brooks III, C. L.; Al-Hashami, H. M.
Characterizing the Protonation State of Cytosine in Transient G:C Hoogsteen Base Pairs in DNA. *J Am Chem Soc* **2013**, *135*, 6766-6769.

30. Segers-Nolten, G. M. J.; Sijtsema, N. M.; Otto, C. Evidence for Hoogsteen GC base pairs in the proton-induced transition from right-handed to left-handed poly(dG-dC)·poly(dG-dC). *Biochemistry (N. Y.)* **1997**, *36*, 13241-13247.

31. Kaushik, M.; Prasad, M.; Kaushik, S.; Singh, A.; Kukreti, S. Structural transition from dimeric to tetrameric i-motif, caused by the presence of TAA at the 3′-end of human telomeric C-rich sequence. *Biopolymers* **2010**, *93*, 150-160.

32. Nonin, S.; Phan, A. T.; Leroy, J. Solution structure and base pair opening kinetics of the i-motif dimer of d(5mCCTTTACC): a noncanonical structure with possible roles in chromosome stability. *Structure* **1997**, *5*, 1231-1247.

33. Day, H. A.; Huguin, C.; Waller, Z. A. E. Silver cations fold i-motif at neutral pH. *Chemical Communications* **2013**, *49*, 7696-7698.

34. Day, H. A.; Pavlou, P.; Waller, Z. A. E. i-Motif DNA: Structure, stability and targeting with ligands. *Bioorg. Med. Chem.* **2014**, *22*, 4407-4418.