2014

Spectroscopic Analyses Of Factors Affecting Formation And Stability Of I-Motif Dna

Samantha Michelle Reilly
University of Mississippi

Follow this and additional works at: https://egrove.olemiss.edu/etd

Part of the Chemistry Commons

Recommended Citation
https://egrove.olemiss.edu/etd/416

This Dissertation is brought to you for free and open access by the Graduate School at eGrove. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of eGrove. For more information, please contact egrove@olemiss.edu.
SPECTROSCOPIC ANALYSES OF FACTORS AFFECTING FORMATION AND STABILITY OF I-MOTIF DNA

A Dissertation
Presented for the
Doctorate of Philosophy
Degree in Chemistry from
The University of Mississippi

by
SAMANTHA M. REILLY
December 2014
ABSTRACT

The four-stranded i-motif (iM) conformation of DNA has importance to a variety of biochemical systems that include nanomaterials and oncogene regulation. Fundamental studies to understand iM formation and its structure in solution still remain to be done. In this dissertation, we discuss the use of the fluorescent base analog tC° to determine the structural state of DNA in solution. We also demonstrate the use of tC° to determine both the hydrodynamic properties and the folding mechanism of an iM. We also determined the loop length dependence on iM formation and stability. Our hydrodynamic studies show that the iM structure has rotational correlation times similar to unfolded single-strand DNA, but displays the same structural rigidity as duplex DNA. This combination of hydrodynamic properties is unique to the iM structure. Using kinetic rates, a mechanism is proposed for the folding of a random coil oligo into the iM. The fluorescence changes in tC° describes the hydrogen bonding of the cytosines that occur during the folding of the iM. This folding mechanism shows that all hydrogen bonds form on the same time scale as the iM forms, meaning that no one base is more important than the others in forming the iM structure. The loop length dependence studies show a distinct difference in stability. The differences in thermal stability and pKₐ observed when lengthening loops suggests a reasonable method for gaining fine control over the thermal stability and pKₐ of the iM that can be readily adapted to nanomaterial usage. Our research also shows that the optimal search algorithms for finding iMs in genomic databases should be different from the algorithms currently used.
DEDICATION

This dissertation is dedicated to everyone who helped me get through the trials and tribulations of graduate school. In particular, I thank my husband, Ryan Baucom, who was always there to support me, no matter what.
LIST OF ABBREVIATIONS AND SYMBOLS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>iM</td>
<td>i-Motif</td>
</tr>
<tr>
<td>G4</td>
<td>Guanine quadruplex</td>
</tr>
<tr>
<td>UV-Vis</td>
<td>Ultraviolet visible spectroscopy</td>
</tr>
<tr>
<td>CD</td>
<td>Circular dichroism</td>
</tr>
<tr>
<td>FDCD</td>
<td>Fluorescence-detected circular dichroism</td>
</tr>
<tr>
<td>PdC</td>
<td>Pyrrolo-deoxycytidine</td>
</tr>
<tr>
<td>tC°</td>
<td>1,3-diaza-2-oxophenoxazine</td>
</tr>
<tr>
<td>NHE</td>
<td>Nuclease Hypersensitive Element</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

I express my deepest appreciation to my advisor, Dr. Randy M. Wadkins and my committee members, Drs. James Cizdziel, Josh Gladden III, Nathan Hammer, and Susan Pedigo, who all gave wonderful advice and support, without which, I would not have been able to make it this far. I also could not have completed my studies without the assistance from the Department of Chemistry and Biochemistry and all of the wonderful staff working for the department.

In addition, I thank Dr. Alexandra Stenson of the University of South Alabama for supporting me throughout my undergraduate experience and telling me that I could and would succeed in graduate school, something I did not think was possible before working in her research lab.

I acknowledge my fellow doctoral students for always being there. Thank you for listening to my complaints and worries and for supporting me.

Lastly, I do want to thank the NIH and NSF for funding the grants that supported my research.
TABLE OF CONTENTS

ABSTRACT ........................................................................................................................................... ii
DEDICATION .......................................................................................................................................... iii
LIST OF ABBREVIATIONS AND SYMBOLS ....................................................................................... iv
ACKNOWLEDGEMENTS ...................................................................................................................... v
LIST OF TABLES ..................................................................................................................................... vii
LIST OF FIGURES ............................................................................................................................... viii
CHAPTER 1 : INTRODUCTION ............................................................................................................ 1
CHAPTER 2 : EVALUATION OF FLUORESCENT ANALOGS OF DEOXYCYTIDINE FOR MONITORING DNA TRANSITIONS FROM DUPLEX TO FUNCTIONAL STRUCTURES ........................................................................................................... 19
CHAPTER 3 : FOLDING AND HYDRODYNAMICS OF A DNA I-MOTIF FROM THE C-MYC PROMOTER DETERMINED BY FLUORESCENT CYTIDINE ANALOGS ........................................................................................................... 32
CHAPTER 4 : EFFECT OF INTERIOR LOOP LENGTH ON THE THERMAL STABILITY AND PKA OF I-MOTIF DNA ........................................................................................................................................... 55
CHAPTER 5 : SUMMARY ....................................................................................................................... 69
LIST OF REFERENCES .......................................................................................................................... 71
APPENDIX ............................................................................................................................................ 83
VITA ..................................................................................................................................................... 87
LIST OF TABLES

Table 2.1: The 113-mer DNA sequence ........................................................................................................ 22
Table 2.2: Fluorescence lifetimes of PdC and tC° in single- and double-strand DNA ................. 30
Table 3.1: T_m and pK_a values of the wild type and tC° substituted DNA folded into the iM .... 44
Table 3.2: Fluorescence lifetimes of PdC Oligo ............................................................................................ 45
Table 3.3: Fluorescence Rotational Correlation Times (θ) of tC°-containing oligos at 25 °C .... 47
Table 3.4: The rate constants of folding from circular dichroism ................................................................. 49
Table 3.5: The rate constants of folding for five different positions in the i-motif structure ...... 51
Table 4.1: Oligos used and their respective sequences, where X is the placement and number of
the thymine repeats (3, 5, 10, 15, 20) in the loop modification ................................................................. 59
Table 4.2: The pK_a, T_m, and ΔG_{37 °C} of all oligos in 30 mM sodium cacodylate ± 40% PEG-300.
.................................................................................................................................................................. 62
Table 4.3: pK_a determination for six pH indicators with two different sized PEGs at 40% (w/w)
in 30 mM sodium cacodylate buffer .......................................................................................................... 63
Table 4.4: T_m of three iMs across different sized PEGs at 40% (w/w) in 30 mM sodium
 cacodylate at pH 5.4 ........................................................................................................................................ 67
Table A.1: Fluorescence lifetimes of tC° oligos ............................................................................................ 86
Table A.2: Fluorescence Rotational Correlation Times (θ) of tC°-containing oligos at 25 °C in
150 mM potassium chloride ....................................................................................................................... 86
Table A.3: Dichroism rate constants for C20T at various concentrations ................................................. 86
LIST OF FIGURES

Figure 1.1: G-quadruplexes and i-Motifs ................................................................. 2
Figure 1.2: Loops of G-Quadruplexes. .......................................................................... 2
Figure 1.3: Capping and elongation by telomerase (large sphere) and the formation of G4s (cylindrical structure) in uncapped telomeres .......................................................... 8
Figure 1.4: Relative densities of G4s/iMs around transcription start sites (TSS) ............... 10
Figure 1.5: Schematic of an I-switch, which is composed of oligos (O1, O2, and O3). ........ 13
Figure 1.6: The G-rich oligonucleotide shown can form a G4. TMPyP4 is conjugated to this G4, which is attached to the MSP ......................................................................................... 15
Figure 1.7: DNA-pyramid protein delivery system ................................................................ 16
Figure 1.8: iM-based oligos forming Y-shaped structure .................................................. 17
Figure 2.1: Structures of PdC and tC° ........................................................................... 21
Figure 2.2: The characteristic CD spectrum of wild type ss- and ds-oligos ......................... 26
Figure 2.3: FDCD spectra with respect to temperature for tC° ........................................... 28
Figure 2.4: Excitation and emission spectra of PdC and tC° incorporated in ss- and dsDNA. ... 29
Figure 3.1: The i-motif structure of C-rich DNA ............................................................ 34
Figure 3.2: A schematic of the C20T mutant of the NHEIII promoter iM .......................... 36
Figure 3.3: Formation of i-motif structures as a function of pH ......................................... 43
Figure 3.4: Sedimentation by analytical ultracentrifugation of C20T at different pHs ........... 44
Figure 3.5: Global folding of the i-motif as monitored by CD ............................................ 47
Figure 3.6: The folding mechanism for the C20T i-motif ................................................ 49
Figure 3.7: (A) The fluorescence spectra of the tC°4 oligo................................................................. 51

Figure 4.1: Representative folding of a random coil into an iM......................................................... 58

Figure 4.2: Determination of pKₐ of Mod2 oligos by monitoring the CD signal at 298 nm ...... 63

Figure 4.3: Thermal denaturations of Mod2 oligos in (A) dilute buffer and (B) molecular crowded conditions (40% PEG-300). ........................................................................................................ 65

Figure A.1: Thermal melting of C20T at various concentrations ......................................................... 84

Figure A.2: S₂₀,ₙ vs pH of 3 µM iMotif with 10 mM KCl or 200 mM NaCl at 20°C ....................... 84

Figure A.3: Thermal melting of C20T and all five tC° substituted oligos ........................................ 85

Figure A.4: (A) CD spectra and (B) normalized fluorescence spectra of tC°4 ............................... 85
CHAPTER 1: INTRODUCTION

1.1 Discovery and Structure of G-Quadruplexes and i-Motifs

In addition to canonical Watson–Crick duplex DNA structure, DNA can exist in various non-canonical conformations such as G-quadruplexes (G4s) and i-Motifs (iMs). The G4s and iMs are formed from guanine-rich and cytosine-rich sequences, respectively. In 1962, it was found that four guanine bases can hydrogen bond with each other to form G-tetrads (Figure 1.1a) (9). Further, in 1974, it was shown that the array of G-terads stack to form G4s (Figure 1.1b) (10). The foundation of discovery of iMs was laid in 1963 when Langridge and Rich (11) proposed that two cytosines could hydrogen bond to form hemiprotontated structure (Figure 1.1c). However, it was only in 1985 that the structure of iMs was established (Figure 1.1d) using circular dichroism spectroscopy (12). Ever since these discoveries, the existence of G4s and iMs in living cells has remained a topic of great controversy, and the formation of guanine quadruplexes (G4s) in human cells has been only recently reported, quantitatively visualized by a G4-specific antibody (13). This confirmation of existence of G4s in vivo has significantly increased the importance of G4s and iMs in biological roles, especially in transcriptional regulation and drug targeting.
**Figure 1.1**: G-quadruplexes and i-Motifs. (a) G-tetrad formation that form the basis of the G4 structure; (b) a cartoon of a G4 structure; (c) C-C⁺ hydrogen bond that forms the basis of the i-motif stability; (d) a cartoon of an iM structure.

The G4s and iMs can occur in multiple conformations. They can be intramolecular (formed from single strand DNA; ssDNA) or intermolecular (formed from multiple DNA strands, generally bimolecular or tetramolecular). As discussed in the review article by Burge et al., (2), G4s can be parallel or anti-parallel. Parallel G4s are formed when all guanine rich strands are oriented in the same direction. Anti-parallel G4s are those that have at least one of the guanine-rich strands oriented in the opposite direction. In addition, Burge et al., (2) describe how G4s can vary based on the classification of the loops: propeller, lateral, and diagonal (Figure 1.2). Unlike G4s, iMs almost always occur in anti-parallel conformations.

**Figure 1.2**: Loops of G-Quadruplexes. (a) Parallel G4 with propeller loops; (b) anti-parallel G4 with both lateral and diagonal loops. Modified from a figure reused with permission from (2).
The sequences forming G4s and iMs occur abundantly near the transcription start sites (TSS) of several oncogenes (14). Due to their topological variance from duplex DNA and their function as gene regulatory elements, the G4s and iMs are perceived as unique targets for binding of drugs and transcription factors to regulate gene expression (summarized by Düchler (15)). Owing to their non-toxic nature and their biodegradability, G4 and iM-based biomaterials are gaining rapid popularity in therapeutics. Although the applicability of G4s and iMs is widely being explored in several non-physiological applications, including, nano-circuitry and nano-mechanical motors, this chapter will highlight the roles of G4s and iMs in physiological applications, such as drug targeting, biosensing, and drug delivery (16, 17). Along with the applications of G4s/iMs, several factors affecting the conformational and biophysical characteristics of G4s/iMs will be discussed in order to provide a clear understanding of the mechanism of structural dynamics of G4s and iMs.

1.2 Factors Affecting the Formation and Stability of G4

1.2.1 Cations

The G4s are composed of G-tetrads. Each G-tetrad is involved in Hoogsteen base-pairing by guanines and is stabilized by a coordination complex with a cation (2). The cations interact with the oxygens on tetrad-forming guanines and reduce the electrostatic repulsion between the oxygens. Under physiological conditions without coordinating cations, the G-tetrads lose their structural integrity due to the electrostatic repulsion between these oxygen atoms. Intracellularly, both sodium and potassium are in high abundance, but due to its ideal size and charge, the potassium ion shows preferential binding to stabilize G-tetrads. Also, the dehydration cost of coordinating sodium into the G-tetrad is greater than coordinating potassium, making potassium even more preferable (18). Campbell and Neidle have written an in-depth review on other
cations that can substitute for potassium, which include ammonium, barium, calcium, cesium, lead, lithium, rubidium, strontium, and thallium (19, 20).

1.2.2 Loop Length

The effects of loop lengths on formation of G4s have been investigated by several groups (20-25). Collectively, these studies concluded that G4s formed from short loops (with 3 or less bases) are more stable relative to those having more than 3 bases in their loops. The G4s with shorter loops showed a higher melting temperature and had lower entropic cost for folding (21, 22). Also, the G4s with longer loops require higher concentrations of cations in order to be thermodynamically as stable as G4s with shorter loops (23). Although G4s with long loops are less favorable, they can be formed under physiological conditions (25). The loop length also dictates the orientation of G4s; generally, short looped G4s adopt parallel orientation, while G4s with longer loops are more likely to have an antiparallel conformation.

1.2.3 Backbone Modifications

Studies have shown that altering the DNA backbone composition to locked nucleic acid (LNA), L-DNA (DNA made from L-deoxyribose), phosphorothioate, and 2'-O-methyl, can alter the structure and thermal stability of G4s. When the backbone of entire G4-forming sequences is modified with LNA, the formation of the G4 was inhibited due to steric hindrance by the backbone (26). The modification of the backbone to L-deoxyribose, a mirror image of naturally occurring D-deoxyribose, has been shown to decrease slightly the G4’s thermal stability when compared to the regular DNA backbone, suggesting that the chirality of the backbone has little effect on G4 formation (27). However, chemical modification of phosphorothioate, which is a sulfur substituting for 2' oxygen atom of the phosphate in the DNA backbone, shows a drop in thermal stability caused by the much larger size of the sulfur atoms compared to the oxygen
atoms they are replacing (27, 28). The 2'-O-methyl backbone modifications have shown an increase in thermal stability as well as a preference towards adopting compact parallel G4 conformation. The modification of the 2' position to O-methyl causes the ribose sugar to preferentially adopt the anti-conformation over the syn-conformation, leading to thermally stable, parallel G4s when compared to unmodified 2'-position (27, 28).

1.2.4 Molecular Crowding

Macromolecular crowding agents like proteins and nucleic acids occupy 20-40% of intracellular volume (29). Hence, it is important to understand the effects of crowding condition on G4s. In vitro, molecular crowding conditions can be mimicked by using large polymers, such as polyethylene glycol (PEG). The molecular crowding with PEG induces a conformational change of G4s from anti-parallel to parallel and lowers the concentration of cations required to stabilize the G4 structures (30, 31). In general, lowering the water activity and dielectric constant around DNA cause G4s to be more stable (32), since decrease in the dielectric constant of the solvent increases the stability of cation-guanine complex (33). However, studies indicate that crowding alone is not sufficient to enhance the stability of G4s, since using bovine serum albumin does not affect the thermal stability or conformation of the G4 structure. Contrarily, the addition of smaller co-solvents, such as acetonitrile and dimethyl sulfoxide, leads to increase in thermal stability as well as change in conformation (anti-parallel to parallel) of G4 structures. Together, these experiments suggest that dehydration of the structure is the primary requirement for enhancing the thermal stability (34).
1.3 Factors Affecting the Formation and Stability of iMs

1.3.1 pH Dependence and Molecular Crowding

For iM stabilization, the necessary cation is a proton; iMs are composed of intercalated hydrogen bonds formed between hemiprotonated cytosines. In dilute buffer conditions, the typical pK$_a$ of iMs is around pH 6.0 (35, 36). Most of the studies in the literature on iMs have been conducted at pH 5.5 in order to ensure that all DNA is in the intercalated conformation. These conditions do not mimic realistic biological systems, where the pH is approximately 7.4. However, in 2009, Zhou et al., (37) found that if the temperature was lowered to 4 °C, the pK$_a$ of the iM structure shifts to neutral pH, but this low temperature is far from physiological temperature for most organisms. Around the same time Rajendran et al., (38) found that molecular crowding with both PEG-200 and PEG-8000 at 20% (w/v) could shift the pK$_a$ of iMs to physiologically relevant pHs (7.1-7.4) at room temperature. This experiment is more in agreement with the hypothesis that the iM structure can be formed in vivo in the presence of macromolecular crowding at physiological pH.

1.3.2 Backbone Modification

Studies on iMs with backbone modifications to the DNA have shown that different modifications can have varied effects on iM stability. The iMs formed in RNA tend to be far less thermally stable than those in DNA (39). In addition, two different studies have found that having a peptide nucleic acid backbone (PNA) on iMs causes them to have lower thermal stability (40, 41). However, it was found that thermal melting temperature increased by approximately 7 °C when an intermolecular iM was formed between a strand of DNA and a strand of PNA versus between two strands of DNA, suggesting DNA/PNA iM hybrids are more stable than regular DNA iMs (40).
1.4 G4s as Drug Targets in Telomeres and Promoters

Due to their topological variance from the duplex DNA, G4-tetraplexes offer unique recognition sites for binding of drugs and transcription factors. In order to arrest cell division during cancer metastasis, G4s that are located at the ends of the telomeres have become common drug targets. The G4s within the promoters of several proto-oncogenes, such as \textit{MYC} and \textit{KIT}, are regarded as prime targets for drugs in cancer therapy because of their ability to silence or activate the expression of these genes. Over the last few years, several excellent review articles have appeared on this topic (3, 15, 42-45), and here we give brief highlights.

1.4.1 Telomeric G4 as drug targets

The length of human telomeric DNA varies depending on the cell type. In somatic cells, the telomeres span from 6-8 kb while in germline cells the telomeric length ranges between 10-20 kb (46, 47). Normally, the length of the telomeres progressively decreases due to the end effect in replication, where DNA polymerase is not able to completely replicate the lagging strand, resulting in a loss of 50-200 bases with each cell division (48). Cells with critically shortened telomeres reach the Hayflick limit, and cease to divide and apotose (49). However, in tumorous cells, the telomeric length remains unchanged (46, 50). The maintained length of telomeres in tumor cells is attributed to an enzyme called telomerase, which is a RNA-dependent DNA polymerase responsible for elongation of telomeres by adding nucleotides (51, 52). Telomerase is highly expressed in 80-85% of cancer cells (53). Due to overexpression of telomerase the cells do not attain the Hayflick limit, resulting in unrestricted number of cell divisions; thus, inhibiting telomerase activity is considered as one solution to reduce tumorigenesis (46, 50). In addition to the elongation of telomeres, telomerase also caps the ends of telomeres (54, 55). Due to this capping, the ends of telomeres are protected from being
recognized as ssDNA breaks, which would otherwise activate the DNA damage response apparatus of cells (d.r.c) leading to apoptosis \((56, 57)\). This protective mechanism is shown in Figure 1.3.

Both the catalytic and capping functions \((46)\) of telomerase are directly inhibited by the presence of G4s as telomerase does not bind when G4s are present. The ends of human telomeres are composed of tandem repeats of the ssDNA sequence \(5'-\text{TTAGGG}-3'\), which is capable of forming G4s \((58)\). Several drugs are being developed to stabilize the telomeric quadruplexes. For example, BRACO19 and BRACO20 are tri-substituted acridine derivatives, which have low toxicity and high G4-stabilizing and telomere-cap disturbing ability \((3, 46, 59, 60)\). The effects of BRACO19 were tested on human uterine carcinoma cells and showed a significant anti-cancer activity within 2 days of treatment. After 39 days of exposure to BRACO19, telomere length decreased by 17%. This activity is attributed to stabilization of G4s and inhibition of telomerase.
The BRACO20’s activity was tested in ovarian cancer cells. It was found that BRACO20 can bind to G4s with 30% higher affinity than for duplexes and could inhibit telomerase (61). Other examples of G4-binding agents are amidoanthracine-9,10-diones (e.g., BSU1051) and pentacyclic acridinium (RHPS4) (48, 62). The BSU1051 showed higher binding affinity towards G4 relative to cruciform, duplex, and ssDNA. This compound inhibits telomerase activity by binding to G4s without affecting telomerase itself. The telomerase activity is blocked in concentration-dependent manner (62). The RHPS4 showed anticancer activity in breast and vulval carcinoma; however it had no effect on ovarian carcinoma cells, which possess longer telomeres and are inherently more resistant to the cytotoxic effects of RHPS4 (48).

Natural products such as telomestatin and meridine are known to stabilize G4s and inhibit telomerase activity (63-65). Kim et al. (66) showed that telomestatin can convert linear DNA to G4s even in the absence of cations. Telomestatin exhibits 70-fold higher selectivity for intramolecular G4s than duplex DNA. The G4s stabilized by binding of telomestatin inhibit telomerase activity, leading to accelerated telomere shortening. Guittat et al., (67) found that, like telomestatin, meridine also shows selectivity towards binding to intramolecular G4s. Although meridine is regarded as inferior to telomestatin in binding to G4s and stabilizing them, it is easier (due to its compact structure relative to telomestatin) to develop meridine analogs with better binding affinities and selectivity for G4s than it is to create telomestatin analogs (67).

1.4.2 Promoter G4s as drug targets

The G4-forming sequences occur abundantly near the transcription start sites of several genes (Figure 1.4) (42). G4s located in the promoter regions regulate gene expression and hence are attractive therapeutic targets (42). The MYC, KIT, KRAS, MYB, BCL-2, VEGF, PDGFA, RB1 and TERT genes are some of the widely studied oncogenes having G4-forming sequences in
their promoter regions (42, 68). The G4s occurring in these genes adopt various conformations; however, the common features like G-tetrads and planarity are being used to develop small molecules that can bind to these planar platforms and silence genes (42). Some of the most common molecules used in regulating the gene expression by binding to G4s are TMPyP$_4$ and TMPyP$_2$ (69), trisubstituted isoalloxazines (70), Se2SAP (71), and Quarfloxin (72).

![Graph showing relative density of G4s/iMs around transcription start sites (TSS). The density of G4s/iMs is higher in upstream regions relative to intragenic regions based on G/C-rich content in the human genome.](image)

**Figure 1.4:** Relative densities of G4s/iMs around transcription start sites (TSS). The density of G4s/iMs is higher in upstream regions relative to intragenic regions based on G/C-rich content in the human genome.

### 1.5 G4s/iMs as Biosensors

Above, we discussed G4 as potential therapeutic targets. We now describe how these structures have been incorporated into diagnostic and stimuli-responsive schemes. DNA-based biosensors are useful in monitoring the changes in the environment around them via structural differences between folded and unfolded secondary structures, and are capable of detecting nM to pM concentrations of analyte. Further, the micrometer to sub-micrometer size of these sensors makes them convenient to probe relatively tight spaces like intracellular micro-environments. Here, we describe a few developing applications of G4/iM based biosensors.
1.5.1 G-quadruplex-based biosensors

**Thrombin sensor:** Thrombin is a serine protease that is an important marker in diagnosis of pulmonary metastasis. Higher concentrations of thrombin in the blood are also responsible for venous thrombosis (73). Hence, monitoring and maintaining thrombin levels in blood is important. In 1992, Bock et al., (74) found multiple thrombin binding aptamers. They started with a pool of >$10^{13}$ different 96-mers DNA sequences. These sequences were biotinylated and amplified using PCR. The biotinylated ss-oligos were then separated from the non-biotinylated strands on an agarose-avidin column and screened for their ability to bind thrombin. One of the aptamers isolated by this process was 5'-GGTTGGTGTGGTTGG-3', which is now the most prevalent thrombin binding aptamer (TBA). Using this aptamer, Sun et al., have recently reported a quantitative thrombin sensor (75). In the presence of thrombin, the TBA on the outermost layer of an electrode binds thrombin and creates a barrier for electrons and inhibits electron flow, resulting in decreased differential pulse voltammetry (DPV) signals. Thrombin concentrations in the range of 1 pM to 160 nM can be measured using this G4-based sensor.

**Ion sensors:** Potassium ions (K$^+$) are involved in many biological functions, including nerve transmission, regulation of blood pressure, and the formation of collagen or elastin. Abnormal levels of K$^+$ ions could lead to muscle cramps or weakness, nausea, diarrhoea, frequent urination, and dehydration (76). This makes it necessary to create devices that monitor the concentrations of K$^+$ ions. Huang et al., (77) have developed a G4-based probe in order to facilitate the rapid detection of K$^+$ ions in urine. For this probe, Huang et al., used an ATP-binding G4 aptamer (5'-ACCTGGGGGAGTATTGCGGAGGAAGGT-3') and the cyanine dye oligreen (OG) (77). When K$^+$ ions bind to the OG-aptamer complex, the aptamer adopts G-quadruplex structure. This leads to dissociation of OG-from the aptamer, leading to a decrease in fluorescence. This
probe is highly selective for potassium ions; the selectivity is measured to be more than 10,000-fold for K⁺ over Na⁺. The lower limit of detection of K⁺ is approximately 75 nM. Highly specific and selective G4-forming aptamers can also be used for detecting ions like Pb²⁺ and Hg²⁺ (78).

1.5.2 i-Motif based biosensors

Glucose sensors: Glucose is an ubiquitous fuel necessary for performing metabolic functions. High Glucose levels in blood/urine serve as a marker for diagnosing diabetes (79). Several different types of probes to monitor glucose levels based on iMs have been developed. Li et al., (80) have designed a probe sensitive to and selective for glucose detection. The working principle of this probe is based on two phenomena: conformational switching of iMs and aggregation of gold nanoparticles (AuNPs). In the absence of glucose, the C-rich nucleotide (5’-CCCTAACCTAACCCTAACCTACCT-3’) has an extended single-stranded conformation at neutral pH. In this ssDNA conformation, the DNA protects the AuNPs from salt-induced aggregation (pink colored solution). However, in the presence of glucose and glucose oxidase, gluconic acid is formed, which lowers the pH of the medium. At acidic pH, the C-rich strand adopts an iM conformation, deprotecting the AuNPs and leading to their aggregation. The aggregation of the AuNPs in high ionic-strength media causes the color change the AuNP solution from pink to blue. The change in the color of AuNPs is used to estimate the concentration of glucose. This probe can detect the glucose levels ranging from 0.5 µM to 5 mM. The high-sensitivity of this probe is attributed to extremely high extinction coefficients of AuNPs and efficient conformational switching of the C-rich nucleotide.
Intracellular pH sensor: The iM-based pH sensors can be used to monitor intracellular pH. The fluctuation in intracellular pH can be used to detect conditions like dysplasia as well as in monitoring other biological processes. Modi et al., have developed a nano-device known as an I-switch that can map spatial and temporal pH-changes in living cells (6). This I-switch is built from 3 different oligos (O1, O2 and O3; Figure 1.5). The O1 and O2 are hybridized to O3 to form duplex DNA. The oligos O1 and O2 have overhangs, which form one half of the bimolecular iM. The overhangs are tagged with the fluorophores Alexa-647 and Alexa-488. At pH above 6.8, the iM-forming overhangs remain as ssDNA, and at pH lower than 5.5, the overhangs from O1 and O2 form intermolecular iM. While going from pH 5.5 to 6.8, the iM unfolds, reducing the fluorescence resonance energy transfer (FRET) intensities, and thereby reporting on pH changes. Thus, the I-switch works as an efficient reporter of pH ranging from 5.5-6.8. This pH sensor was successfully employed intracellularly to detect pH changes associated with endosomal maturation.

Figure 1.5: Schematic of an I-switch, which is composed of oligos (O1, O2, and O3). Oligos O1 and O2 can hybridize to O3 and consists of overhangs tagged with fluorophores (represented by spheres). These overhangs each form one half of a bimolecular iM. At pH 7.3, the overhangs are open (not in iM conformation), while at pH 5.4, the overhangs are closed (interlocked to form iMs), leading to FRET. Reused with permission from (6).
1.6 Therapeutic agents and drug-delivery vehicles

The G4s/iMs-based nano-devices are rapidly gaining popularity as drug-delivery agents over synthetic polymers. This popularity is attributed to their ability to recognize specific targets, their ability to access intracellular spaces, and their bio-compatibility, bio-degradability and non-toxicity. Along with targeted drug-delivery vehicles, some G4-forming aptamers have exhibited therapeutic significance in cancer and HIV treatments.

1.6.1 G4-forming aptamers as therapeutic agents

The AS1411 is a 26-mer G4 aptamer that has passed Phase II clinical trials as one of the first anticancer nucleic acid aptamers to treat lymphoma (81). Initially, the aptamers were developed based on the fact that guanine-rich oligos (GROs) possess anti-proliferative properties in cancer cells (82). However, the anti-cancer activity of AS1411 was discovered via serendipity and not via systematic evolution of ligand by exponential enrichment (SELEX), which is the technique responsible for most aptamers (82). Paula Bates and co-workers proposed that the 26-mer G4 aptamer binds to nucleolin, which is a protein overexpressed in several cancer types (83). The binding of G-rich oligonucleotide to nucleolin inhibits one or more of its normal functions (ribosome production, nuclear transport, and cell entry), which suppresses tumor proliferation in cell lines. Recently, Magbanua et al., reported another 16-mer DNA aptamer AID-1 (5'-GGGTGGGTGGGTGGGT-3') that can bind to interleukin-6 receptor (IL-6R) (84). In order to locate this aptamer, Magbauna et al., performed 13 rounds of SELEX. The resulting AID-1 binds to IL-6R with K_d values in the nM range, and inhibits both HIV infection and HIV-1 integrase.
Several groups including Shieh et al., (85) and Chen et al., (7) have used G4-forming aptamers as carriers for delivering anti-cancer drugs. Chen et al., have developed light-operated vehicles comprised of photosensitizer-incorporated, G4-capped, mesoporous silica nanoparticles (MSPs), as shown in Figure 1.6 (7). The oligonucleotide 5’-GGGGTTTTGGGG-3’ forms G4s, which are conjugated to MSP and TMPyP₄. This complex of MSP-DNA-TMPyP₄ was tagged with folic acid and internalized into cells via folate-mediated endocytosis. Upon irradiating with visible or near infrared wavelengths of light, Meso-TMPyP₄ acts as a photosensitizer to generate a reactive oxygen species (ROS). The production of ROS cleaves the DNA-capping of the MSPs, releasing the cargo. The G4s in this system act both as a capping agent and as a photodynamic

**Figure 1.6:** The G-rich oligonucleotide shown can form a G4. TMPyP₄ is conjugated to this G4, which is attached to the MSP. The MSP-DNA-TMPyP₄ complex is tagged with folic acid, allowing the complex to enter cells via folate-mediated endocytosis. Upon irradiating with visible or near infrared light, the TMPyP₄ generated ROS, which uncaps MSP from DNA, releasing the cargo (shown by doxorubicin in this figure). Reused with permission from (7).
therapy drug-carrier, whereas TMPyP₄ acts as the ROS generator. Further, TMPyP₄ stabilizes the G4 DNA, protecting it from degradation. Importantly, Chen et al., successfully loaded MSPs with the anti-cancer drug doxorubicin (DOX) and demonstrated precise, spatially and temporally controlled drug release for both chemotherapy and photodynamic therapy in human cancer cells.

**Figure 1.7:** DNA-pyramid protein delivery system. The edges are composed of iM-forming sequences. Under basic conditions, the edges are sealed, forming duplex DNA. Under acidic conditions, the edges form iMs, leading to the partially open conformation of the pyramid, facilitating the cargo delivery. Reused with permission from Ref. (4).

### 1.6.3 iMs as drug-delivery vehicles

Cheng et al., have demonstrated an assembly of pH-responsive, fast-acting hydrogels based on iMs (5). They have constructed a Y-shaped DNA structure formed from 37-mer ssDNA. These ssDNA contain two functional domains: an 11-mer interlocking iM domain containing two cytosine-rich stretches, and a 26-mer, which forms the double stranded structure in the Y-shape (Figure 1.7). At basic pH, each individual Y-shaped unit is isolated from the other. However, at acidic pH, the iM-forming strands assemble to form inter-molecular iMs. This interlocked system results in the formation of three-dimensional structures in solution, and to the formation of hydrogels. Cheng et al., have suggested that these structures can be used in pH-sensitive drug-delivery systems.
Keum et al., have developed an iM-based DNA pyramid that can deliver cargo to its targets (Figure 1.8) (4). The vertices of the pyramids are connected by iM-forming DNA sequences (Figure 1.8: strands 2, 3 and 4). At basic pHs, edges of the pyramids are sealed by duplexes formed by C-rich DNA strands and a partially complementary strand (Figure 1.8: strand 1). The pyramid’s stability was modulated by introducing the mismatched bases in strand 1. Due to these mismatching bases, the melting temperature of the duplex structure forming the edges of the pyramid was reduced. The sequence of the strand 1 was optimized so that under basic pHs the strand formed duplex structure, but under acidic conditions, the strand 1 dissociated from pyramid, allowing the C-rich strands to fold into iMs, and rendering the edges of the pyramid open. In order to demonstrate the function of iM-forming DNA pyramids, Keum et al., used an enhanced green fluorescent protein (EGFP) as cargo to be delivered by the pyramid. EGFP was modified with a hexa-histidine tag, while strand 1 of the pyramid was modified with Nickel-nitrilotriacetic acid (Ni²⁺-NTA). The EGFP was attached to strand 1 via a nickel-mediated interaction between nitrilotriacetic acid (NTA) and histidine tag. At acidic pH, strands 2, 3, and 4 form iMs, opening the pyramid. At the same time, the NTA-Ni²⁺-His linkage, which is unstable in acidic conditions, is cleaved, releasing the EGFP from the open pyramid.

![Figure 1.8: iM-based oligos forming Y-shaped structure. The iM-forming ends of Y-shaped DNA (a) units form intermolecular iMs (b) in acidic conditions leading to the formation of hydrogels (c). Reused with permission from Ref. (5).](image)


1.7 Chapter Summary

This chapter provides a brief introduction to G4/iM discovery, the topological sensitivity of G4/iM to solution conditions, and applications of G4s/iMs in therapeutics and clinical research. The occurrence of G4s/iMs in the regions upstream to the TSS has made them very attractive targets for binding of drugs in order to regulate gene expression. Owing to their small sizes, G4/iM-forming oligos can easily enter intracellular micro-environments, and hence can be employed as biosensors to monitor intracellular changes. Due to their bio-compatibility, G4s/iMs-based aptamers are being developed as drug-delivery vehicles. With the recent confirmation of existence of G4s in living cells and with the increasing demands for non-toxic, bio-compatible drugs and drug-delivery vehicles, G4s/iMs will remain significantly important to therapeutics and clinical research for years to come.

PUBLICATION NOTICE

This review chapter has been published:

2.1 Chapter Abstract

Topological variants of single-strand DNA (ssDNA) structures referred to as “functional DNA,” have been detected in regulatory regions of many genes and are thought to affect gene expression. Two fluorescent analogs of deoxycytidine, Pyrrolo-dC (PdC) and 1,3-diaza-2-oxophenoxazine (tC°), can be incorporated into DNA. Here, we describe spectroscopic studies of both analogs to determine fluorescent properties that report on structural transitions from double-strand DNA (dsDNA) to ssDNA, a common pathway in the transition to functional DNA structures. We obtained fluorescence-detected circular dichroism (FDCD) spectra, steady-state fluorescence spectra, and fluorescence lifetimes of the fluorophores in DNA. Our results show that PdC is advantageous in fluorescence lifetime studies because of a distinct ~ 2 ns change between paired and unpaired bases. However, tC° is a better probe for FDCD experiments that report on the helical structure of DNA surrounding the fluorophore. Both fluorophores provide complementary data to measure DNA structural transitions.

2.2 Introduction

DNA single strands can hybridize to form higher-order functional structures, which include hairpins, triplexes, and quadruplexes (2, 86-88). The existence and physiological relevance of these secondary structures in vivo has been the subject of much controversy.
However, several *in vivo* techniques have confirmed the presence of DNA secondary structures in telomeres and regulatory regions of specific genes (e.g., BCL-2, c-myc) (89-92). Secondary structures may also serve as specific targets recognized by drugs, such as actinomycin D and PIPER, as well as transcription factors, such as Sp1, because of their topological variance from duplex Watson-Crick DNA (93, 94). On account of their occurrence in regulatory regions and their structural peculiarity, functional DNA structures may serve as biological micro-switches for altering transcription by silencing or enhancing gene expression (43, 95). Exercising control over gene expression by controlling the activation of these switches and/or introducing new switches in biological circuits could revolutionize medical research and offer new avenues of treating genetic disorders. For exercising such control, it is imperative to understand the factors affecting the mechanism of formation and maintenance of functional structures at a fundamental level.

Numerous fluorescent analogs of DNA bases have been evaluated for examining the subtleties of DNA transitions (96-98). Locating an appropriate probe that could map the mechanistic aspects of transition of double stranded (ds) DNA to single stranded (ss) secondary structure is a first step toward monitoring the formation of complex, higher-order structures since ssDNA is an intermediate in the pathway to functional structures. In the following report, we present a comparative study of two deoxycytidine analogs, PdC and tC° (Figure 2.1), to evaluate their suitability as fluorescent reporter probes for DNA transitions.
The properties of pyrrolocytosine and the effects of base stacking and hydrogen bonding on its quantum yield in nucleic acids have been previously evaluated (99). Based on this prior work, we reported the use of PdC to determine hairpin formation in short oligos (16 nucleotides) (100). The fluorophore was stable in these oligos and did not perturb DNA structure. This result implied that PdC might be a useful probe for investigating more complex DNA functional structures in detail. However, to overcome the limitations posed by the low quantum yield of PdC ($\phi_f = 0.07$), we also explored another recently characterized fluorescent base analog, tC°. The tC° analog has a quantum yield five-times greater than PdC ($\phi_f = 0.30$ for tC°; $\phi_f = 0.07$ for PdC) and a molar absorptivity maximum of 9000 M$^{-1}$ cm$^{-1}$ at 360 nm (101). The absorbance wavelength of tC° is similar to PdC, which allows both to be easily distinguishable from DNA. Both of these fluorophores have been reported to be quenched, relative to the single strand, when base-paired with guanine, making it easy to determine when the DNA is in the duplex form (96, 100, 102). In addition, unlike PdC, tC° has only minor variations in fluorescent properties caused by surrounding bases (103). However, like PdC, tC° induces little or no changes in stability upon

![Figure 2.1: Structures of PdC and tC°. R indicates the linkage to deoxyribose.](image-url)
incorporation into dsDNA. High quantum yield, retention of the original configuration of DNA, and quenching when base-paired suggested that tC° would be a useful fluorescent probe for mapping the transition of duplex DNA to a functional DNA structure spectroscopically. In this chapter, we compare spectral properties of tC° and PdC for use as reporters of DNA conformation. The DNA sequence used is known to form a cruciform structure in the cloning vector pBR322 under superhelical duress (104, 105). This sequence was chosen since the results from the work described here will provide a context for interpreting data in future studies of the PdC and tC° incorporated into this supercoiled plasmid. Hence, the effect of supercoiling on functional structure formation and conformation can be potentially probed with these fluorescent bases.

2.3 Materials and Methods

Table 2.1: The 113-mer DNA sequence that is part of the cruciform found in the terminator of the AmpR cassette in the cloning vector pBR322 under superhelical strain. The location of the substitution of cytidine by the fluorescent analogs marked by an asterisk.

| Sequence: 5′-TGA GGT TAA GGG ATT TTG GTC ATG AGA TTA TCA AAA AGG ATC TTC* ACC TAG ATC CTT TTA ATT AAA AAT GAA GTT TTA AAT CAA TCT AAA GTA TAT ATG AGT AAA CTT GGG C -3′ |

2.3.1 Materials and Equipment

All DNA oligonucleotides (sequence shown in Table 2.1), including those incorporating PdC and tC° (* indicates that the location of incorporation of the fluorophore), were obtained from Midland Certified Reagent Co. (Midland, TX, USA). All oligos were dissolved in TE buffer (10 mM Tris, 1 mM EDTA, pH 8) made from spectroscopic-grade reagents obtained from Fisher Scientific. Purity of each oligo was assessed using agarose gels, which showed no other higher or lower molecular weight contaminations. CD and FDCD experiments were performed using a circular dichroism instrument (Model 202SF, AVIV 67 Biomedical Inc. Lakewood, NJ,
USA) and fluorescence lifetime measurements were conducted on a multifrequency cross-correlation phase and modulation fluorometer (model K2, ISS Inc., Champaign, IL, USA).

2.3.2 Duplex Oligos

After ensuring the purity of the single-strand PdC, tC°, and wildtype (ssPdC, sstC°, and ssWT; the WT contains no fluorophore) oligos, each was mixed with a slight molar excess of their complementary strand. This mix was heated to 80 °C for 20 minutes, and then cooled to room temperature to anneal strands. The double-stranded oligo formation was confirmed by observing the difference in migration of ds- and ss-oligos on agarose gels. Size was determined by comparison to a DNA ladder of known lengths. In addition, duplex was the only strand observed after staining with the SYBR Gold, which stains both ss- and dsDNA.

2.3.3 Circular Dichroism (CD) and Ultraviolet (UV) absorbance

CD spectra were recorded using a quartz cuvette having an optical path length of 1 cm. All oligos analyzed were in 700 µL volume at concentration of 0.35 µM. Wild-type oligos were analyzed to ensure that the CD signals observed were not modified by the fluorophores. Each data point was averaged over an integration time of 1 s per nm. CD signals were collected from 25 °C to 90 °C at 1 °C increments with a heating rate of 1 °C/min to ensure the jacket temperature also represented the probe temperature. These spectra were collected over the range of 220 nm to 420 nm to check for any change in the global conformation of the single-stranded and double-stranded WT, PdC, and tC° oligos. While performing CD scans, the total UV absorbance of each oligo was also collected.
2.3.4 Fluorescence Detected Circular Dichroism (FDCD)

FDCD spectra were collected using a quartz cuvette of 1 cm path length. Oligos analyzed were in a total volume of 70 µL and at the concentration of 8.5 µM. Each FDCD scan was averaged over 15 s per nm in order to increase signal-to-noise ratio. The excitation wavelengths were from 300 nm to 420 nm at temperatures ranging from 25 °C to 90 °C. Emission was obtained by using a 450 nm cut-on filter.

2.3.5 Steady-State Fluorescence

Steady-state fluorescence spectra were analyzed on a K2 multifrequency cross-correlation phase and modulation fluorometer (ISS Inc., Champaign, IL, USA) for ss- and ds- (WT, PdC, and tC°) oligos. All the oligos had concentration of 0.35 µM in TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0). Excitation scans were collected over the wavelength range of 220–400 nm, with the emission wavelength fixed at 450 nm. Emission spectra were taken over the wavelength ranging from 400 nm to 580 nm with the excitation wavelength fixed at 350 nm. The ss- and dsWT oligos were used as a baseline signal and subtracted from the steady-state fluorescence data.

2.3.6 Fluorescence Lifetime Measurements

Fluorescence lifetime was measured for both ss- and ds-oligos using a frequency domain lifetime instrument. The cuvettes used were of 1 cm path length. Excitation wavelengths were 350 nm and 368 nm for PdC and tC°, respectively. Data were collected from 10–240 MHz frequencies with 100 iterations at each frequency and 40 total frequencies from which the lifetimes were calculated. Both the individual decays (fitted to 2 components, meaning 2 exponential decays) and average lifetimes are reported.
2.4 Results and Discussion

2.4.1 Incorporation of fluorescent dC analogs does not perturb DNA structure

In order to assure that no structural perturbations were introduced by insertion of the fluorophores in the oligos, changes in absorbance at 260 nm were monitored at varying temperatures to determine duplex melting temperatures as well as any transitions occurring in ssDNA (data not shown). The data indicated that the ssWT, ssPdC, and sstC° undergo similar spectral changes with temperature, indicating no perturbations to the ss-structures after replacement of cytidine with PdC or tC°. The absorbance for all ssDNA increases by ~ 10% from 25 °C to 40 °C, signifying reorientation of the strands during melting. Thus, introduction of either PdC or tC° did not lead to the formation of any additional secondary conformation in ss-oligos as predicted based on the previous literature (24, 102). No perturbation of dsDNA by PdC or tC° was observed. All duplexes melted at 45 ± 3 °C in TE buffer, indicating dsDNA containing the fluorescent bases were readily formed and stable at 25 °C. Hence, PdC and/or tC° could be used to investigate single- and double-stranded regions of DNA functional structures.

Circular dichroism (CD) spectroscopy was also used to investigate whether the overall helical structural conformation of the DNA might have been affected by the fluorophores (Figure 2.2). Our results, in agreement with melting data from absorbance and the literature, indicated that the fluorophores did not affect the structure of the duplexes or the manner in which they melted (100, 106, 107). There were only negligibly small changes in the CD spectrum with temperature for ss-oligos, which indicated little structure at any temperature as compared to dsDNA. The dsDNA oligos showed considerable CD, consistent with the common global spectral features typical of B-DNA (36). Major changes in the CD signal of dsDNA with respect to the temperature were observed at 245 nm, with relatively smaller changes that occurred at 280 nm. These CD spectra and their temperature dependence illustrated that the DNA is not affected by replacing cytidine with PdC or tC°.
Figure 2.2: The characteristic CD spectrum of wild type ss- and ds-oligos as a function of temperature. All samples were at 0.35 µM concentration.
Circular dichroism has been used for detecting the local environmental changes around PdC integrated into DNA (106). However, the direct detection of CD by PdC or other base analogs is restricted to analogs that either have an absorbance far enough removed from the DNA $\lambda_{\text{max}}$ of 260 nm to prevent spectral overlap or are incorporated into short DNA sequences. In both cases, significant amounts of DNA are required to obtain good signal-to-noise. An alternative approach to direct CD observation is to use FDCD to study the local environment of the fluorescent base analog. Since our previous data has sufficiently demonstrated that the replacement of deoxycytidine with either PdC or tC° does not affect the original DNA structure, we analyzed the fluorophores by FDCD to explore the local structural changes occurring around the base. In our studies, PdC-containing DNA at concentrations up to 8.5 µM did not show any FDCD signal over the temperature range of 25 °C to 90 °C, suggesting that PdC’s low quantum yield is insufficient for FDCD measurements at reasonable DNA concentrations with our instrument. However, tC° showed a strong FDCD signal for both ss- and ds-DNA between 300 nm and 420 nm even at submicromolar concentrations (0.5 µM; Figure 2.3). The FDCD signal for duplex DNA was significantly larger than that for the ss oligo. In addition it had a positive value throughout, indicating the base was stacked parallel to the bases in the vicinity. The lower intensity of the peak in ss-oligos is attributed to the residual structure at room temperature. Thus, tC° is a very effective probe for monitoring the localized changes occurring in DNA structure in response to variations in structural parameters.
FDCD data obtained can be used for predicting the molecular details of higher-order secondary structures. For example, in case of loop regions of i-motif and cruciform structures, tC° will exhibit lower FDCD signal since these regions are less structured. However, tC° substituting for dC in the regions where base-pairing occurs should show higher FDCD signals. These structural details could then be used to assess the stabilities of functional DNA structures after binding of transcription factors or drugs.

2.4.3 Fluorescence Properties of the Base Analogs Reveal Their Pairing State

Fluorescence intensities of PdC and tC° depend on their hydrogen bonding state, and decrease if the fluorophores are base (100, 102). Figure 2.4 clearly demonstrates this quenching of fluorescence for both PdC and tC° in duplex dsDNA versus unpaired ssDNA. Fluorescence intensities from ssDNA to dsDNA are lowered by approximately 60% and 40% for PdC and tC°,

Figure 2.3: FDCD spectra with respect to temperature for tC°. The low quantum yield of PdC did not allow for detection of FDCD at this concentration. ΔF/F is the difference in the fluorescence caused by left- and right-handed polarized light over total fluorescence signal observed for the fluorophore.
respectively. Figure 2.4 also shows the difference in the quantum yield of PdC and tC°, with PdC exhibiting significantly lower fluorescence. Both the fluorophores had an excitation maximum at 350 nm and an emission maximum of 450 nm that did not change in ssDNA versus dsDNA. These decreases in the fluorescence and the location of the maxima are consistent with the known literature (90, 102).

Like steady-state intensities, fluorescence lifetimes of PdC and tC° decrease when they are paired to their complementary base (100, 102). Since fluorescence lifetimes are independent of concentration, their decrease has the potential to be used for differentiating paired bases from
unpaired, thus they can be used for deductively mapping functional DNA structures. Lifetime data was obtained at room temperature for both ss- and ds-oligos. Two lifetimes ($\tau_1$, $\tau_2$) were fit to the phase-modulation data. The weighted average (weighted by fractions; equation in Table 2.2 description) of the two lifetimes, $\tau_1$ and $\tau_2$, were calculated and reported as $<\tau>$ in Table 2.2. These weighted averages indicate that the lifetimes for tC° decreased slightly (0.3 ns) with respect to the strandedness. In contrast, PdC showed appreciable difference (1.9 ns) between single and duplex oligos. These lifetime values are similar to those previously reported (100, 102). To unambiguously differentiate between the ss- and ds regions of the secondary structures, greater differences in lifetimes of base-paired and unpaired fluorophores are required. Thus, the lifetime data indicates that PdC is a better choice than tC° for assessing the strand base-pairing in DNA via this methodology. Similar to FDCD, lifetime measurements can facilitate the studies to observe differences between ss loop regions and ds regions in functional DNA structures like i-motif and cruciforms, and hence aid the studies to investigate formation and stability of secondary structures.

Table 2.2: Fluorescence lifetimes of PdC and tC° in single- and double-strand DNA. The $\tau_1$ and $\tau_2$ are obtained by fitting a double exponential decay curve to the data, and $f_1$ and $f_2$ are fraction of fluorescence corresponding to lifetimes $\tau_1$ and $\tau_2$, respectively. The $\chi^2$ indicate the measure of error in the fit. The value $<\tau>$ is average lifetime calculated by $\Sigma f_i \tau_i^2 / \Sigma f_i \tau_i$.

<table>
<thead>
<tr>
<th>Oligo</th>
<th>$\tau_1$</th>
<th>$f_1$</th>
<th>$\tau_2$</th>
<th>$f_2$</th>
<th>$\chi^2$</th>
<th>$&lt;\tau&gt;$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ss tC</td>
<td>5.3</td>
<td>0.9</td>
<td>1.7</td>
<td>0.1</td>
<td>2.0</td>
<td>5.1</td>
</tr>
<tr>
<td>ds tC</td>
<td>5.1</td>
<td>0.8</td>
<td>1.8</td>
<td>0.2</td>
<td>1.5</td>
<td>4.8</td>
</tr>
<tr>
<td>ss PdC</td>
<td>5.2</td>
<td>0.7</td>
<td>0.7</td>
<td>0.3</td>
<td>1.9</td>
<td>5.1</td>
</tr>
<tr>
<td>ds PdC</td>
<td>3.3</td>
<td>0.6</td>
<td>0.1</td>
<td>0.3</td>
<td>1.4</td>
<td>3.2</td>
</tr>
</tbody>
</table>
2.5 Chapter Conclusions

In this chapter, spectroscopic techniques were used to determine the appropriate fluorescent deoxycytidine analog for probing the transition of duplex DNA into topologically distinct functional nucleic acid structures. Our conclusion is that neither analog tested is a perfect reporter. Rather, we suggest the two should be used in conjunction to obtain an overall description of changes to nucleic acid structure. While tC° has a very good quantum yield that allows for FDCD analysis, it exhibits only minor changes in fluorescence lifetime between ss- or dsDNA. PdC, on the other hand, undergoes a substantial decrease in the fluorescence lifetime when base-paired, but shows no FDCD signal at concentrations appropriate for most biochemical studies. Hence, we conclude that to map the transition of duplex DNA to other functional forms, both PdC and tC° at concentrations up to 8.5 μM can be used, depending on the structural information desired. We are now investigating spectral changes incurred by these probes, when they are contained within well-known functional DNA structures, such as quadruplex and i-motif DNA.

PUBLICATION NOTICE

The work done in this chapter has been collected and published:

CHAPTER 3: FOLDING AND HYDRODYNAMICS OF A DNA I-MOTIF FROM THE C-MYC PROMOTER DETERMINED BY FLUORESCENT CYTIDINE ANALOGS

3.1 Chapter Abstract

The four-stranded i-motif (iM) conformation of cytosine-rich DNA has importance to a wide variety of biochemical systems that range from their use in nanomaterials to potential roles in oncogene regulation. The iM structure is formed at acidic pH, where hemi-protonation of cytosine results in a stable C-C$^+$ base pair. Fundamental studies to understand iM formation from a C-rich strand from the promoter of the human c-MYC gene are described in this report. We used a number of biophysical techniques to characterize both the hydrodynamic properties of the folded iM and the folding kinetics of this iM. Our hydrodynamic studies using fluorescence anisotropy decay and analytical ultracentrifugation show that the iM structure has a compact size in solution while displaying the rigidity of a double strand. Via our studies of rates of circular dichroism (CD) spectral changes and rates of quenching of fluorescent cytidine analogs, we also established a mechanism for the folding of a random coil oligo into the iM. In the course of determining this folding pathway, we established that the fluorescent dC analogs tC° and PdC can be used to monitor individual residues of an iM structure and can be used to determine the pK$_a$ of an iM. We established that the C-C$^+$ hydrogen bonding of certain bases initiates the folding of the iM structure. We also showed that substitutions in the loop regions of iMs give a distinctly different kinetic signature during folding from those bases that are intercalated. Our data reveal that the iM passes through a distinct intermediate form between the unfolded and folded form. Taken together, our work has laid the foundation for using fluorescent dC analogs
as general tools for following structural changes during iM formation, and our technique may be useful for examining folding and structural changes in more complex iMs.

3.2 Introduction

In addition to the well-known Watson-Crick DNA structures, a variety of other conformations of DNA exist that include G-quadruplexes (G4s) and i-motifs (iMs). G4s form from guanine-rich strands of DNA, whereas iMs form from cytosine-rich strands. Due to their topological variance from Watson-Crick DNA, their links to transcription and regulation (86, 87, 89, 95, 108-110), and their association with oncogene promoters (2, 43, 90, 95, 111-113), G4s and iMs have been studied as possible unique drug targets. G4s require a metal cation to stabilize the structure. As potassium cations are readily available intracellularly, G4s are viable in vivo, and their existence in vivo has been recently confirmed by Biffi et al. (13), who have developed a structure-specific antibody that binds G4s in cells. However, since iMs require protonation of cytosines to form (Figure 3.1), they have not been studied to the extent that G4s have since pH in the nucleus isn’t expected to be significantly lower than that of the cytosol (~ 7.3). Recently, it has been reported that iMs can be formed at neutral pH by addition of molecular crowding agents (38, 114). This finding suggests that iM structures are also likely to be viable in vivo and deserve further study for comparison with the better-characterized G4 structures. In addition to their potential in vivo importance, iMs have been used for both therapeutic and diagnostic purposes. Biosensors and drug delivery biomaterials have been made from iM structures that take advantage of the iMs’ pH-dependent formation, their ease of accessing intracellular spaces, their biodegradability, their biocompatibility, and their non-toxicity (4, 80). A review of these nanomaterials and their applications has recently appeared (115). These materials as well as the
biological applications of iMs have increased the necessity to better understand their structure and formation to fully exploit their properties.

![Figure 3.1](image) Figure 3.1: The i-motif structure of C-rich DNA. (A) A C-C+ hydrogen bond that forms the basis of the iM stability, with the shared proton highlighted in red. (B) An all-atom model of an iM modified from the PDB entry 1YBL (8). The bases are color coded for thymine (yellow), cytidine (red), adenine (blue), and the DNA phosphate backbone (purple).

One important aspect of iMs that dictates their final structure is their folding pathway. Kinetic studies of iM folding have been done using NMR (8, 116-119), surface plasmon resonance (SPR) (120), and fluorescence resonance energy transfer (FRET) (121). Of these studies, only NMR monitored folding by the base pairing of the cytosines. However, these NMR studies used lower temperatures (0-15 °C) and higher pH buffers (~ pH 6.0) to slow the kinetics of iM formation to monitor folding over much longer timescales (minutes to hours) (117, 122). Very few studies have monitored C-C+ pairing on short timescales at room temperature (20-25 °C). Chen et al., (35) performed one of the first studies on iM folding at room temperature using stopped-flow circular dichroism on a short timescale (sub-second). Fluorescence spectroscopy, which can also probe short timescales, uses changes in signals from fluorescent base analogs to investigate iM folding; however, most fluorescence studies to date have focused on base substitutions in the loop and terminal regions of the iM (123, 124). There is a lack of studies
concerning the kinetic mechanism through which hydrogen-bonding bases stabilize the iM. By following the kinetics of quenching of fluorescence by the C-C\textsuperscript{+} pairing interactions, a folding mechanism can be constructed for an iM. Such folding mechanisms will allow an understanding of how best to control and utilize iM-based bio- and nanomaterials, and may also have biological implications for potential iM structures in gene promoters if unique folding intermediates are formed.

The hydrodynamic properties of iM structures in solution have also not been fully characterized. Insight into their conformational dynamics, such as iM movement and shape in solution, can also give details on how to better utilize iMs in nano-devices since their hydrodynamic properties readily affect their diffusion in solution. Choi et al. (125), published one of the few papers that detailed the hydrodynamics of iMs in solution. They found that the diffusion coefficient of the iM form was much greater than that of the unfolded form, showing that the hydrodynamic radius of the unfolded form was 1.6 times that of the iM (125). However, their paper did not detail how the two forms compared to the duplex of the same strand, or how rigid the structures were in solution, aside from what could be gleaned from the diffusion coefficients. Here, we used analytical ultracentrifuge and fluorescence anisotropy decay data to gain insight into the iM’s movement and rigidity in solution compared to other DNA structures.

In this study, we examined a variant of the C-rich strand of the NHE\textsubscript{III} promoter (Wild Type Sequence: 5’-TTCCCCACCCCTCCCCACCCCTCCCCA-3’) from the human c-MYC gene, an oncogene that can promote tumor development (126). The wild-type sequence has been well characterized by Dai et al., (1) who found that it forms a complex mixture of iM structures. However, replacing the last four cytidines with an adenine, and a single C to T mutation at the sixth base results in an iM that was predominately a single species (C20T Sequence: 5’-
TTCCCTACCCTCCCCACCCTAA-3’; shown in Figure 3.2) and retains the thermal stability of the unmodified sequence. In this report, we use the C20T mutant strand labeled at five positions with individual 1,3-diaza-2-oxophenoxazine (tC°) or pyrollo-deoxycytosine (PdC) substitutions (four involved in hydrogen-bonding and one in the loop; locations shown in Figure 3.2). We have previously shown that inclusion of fluorescent dC base analogs on DNA does not significantly affect DNA structure (100, 127). The tC° fluorescent base analog has the highest quantum yield of any known cytosine analog and is quenched in duplex DNA when base-paired to its complementary guanine (φ = 0.3 in single-strand; φ = 0.2 in duplex) (102). Using these oligos, we detailed the hydrodynamics of the folded iM structure and mapped the mechanism of iM folding.

![Figure 3.2: A schematic of the C20T mutant of the NHEIII promoter iM from the c-MYC gene (1). The cytosines are shown in red. The positions of the single fluorescent cytosine analog substitutions are numbered. The equilibrium lies far towards the right structure.](image)

3.3 Materials and Methods

3.3.1 Oligos

All DNA oligos were synthesized through standard solid-phase chemistry by Midland Certified Reagent Company, Inc. (Midland, TX, USA). The oligos were stored in 10 mM Tris, 1 mM EDTA Buffer (pH 8.0) until used. The C20T oligo sequence has a folded structure at pH <
6.0 that has been determined by NMR (1). The individual tC° and PdC substitutions were made at the fourth, ninth, tenth, twelfth, and fifteenth position, which corresponds to the loop and bonded bases (Figure 3.2). The unincorporated tC° nucleoside used as a control was obtained from Dr. Marcus Wilhelmsson (Chalmers University of Technology, Sweden).

3.3.2 Circular Dichroism

In keeping with the prior literature on iMs, we define the pK$_a$ for iM formation as the pH at which 50% of the oligo is folded into the iM. To determine pK$_a$ for iM folding, DNA solutions (4 µM) were made in 10 mM Tris buffers containing 10 mM KCl with pH ranging from 8.0 to 7.0, or 10 mM phosphate buffers containing 10 mM KCl with pH ranging from 7.0 to 6.0, or 10 mM acetate buffers containing 10 mM KCl with pH ranging from 6.0 to 5.4. All reagents were spectroscopic grade and purchased from Fisher Scientific (Waltham, MA, USA). To ensure little to no intermolecular structures were formed, the solutions were heated to ~ 80 °C for ten minutes and then cooled to room temperature. CD spectra of DNA solutions were collected on an Olis DSM 20 Circular Dichroism (CD) instrument (Olis, Inc. Bogart, GA, USA) over the wavelength range of 220 to 350 nm at 20 °C, with an integration time that adjusted with absorbance. The Δε values at 350 nm for the folded (SignalFolded) and the unfolded (SignalUnfolded) were then plotted against pH and were fitted to Equation 3.1. Thermal melts from 20 °C to 60 °C were monitored via CD at 1 °C increments with a heating rate of 1 °C/min to ensure the fluorescent base analog substitutions did not induce a significant change to the iM structure or its thermal stability. Only a single transition was observed over this temperature range, and the measured T$_m$ was independent of concentration (shown in Figure A.1), indicating a unimolecular species.

**Equation 3.1:** $Signal_{Total} = \frac{Signal_{Folded} - Signal_{Unfolded}}{1 + 10^{Cooperativity \times (pH - pK_a)}} + Signal_{Unfolded}$
3.3.3 Fluorescence – Steady State and Lifetime Measurements

Steady-state fluorescence spectra were recorded on a K2 multifrequency cross-correlation phase and modulation fluorometer (ISS Inc., Champaign, IL, USA). Excitation scans were collected over the wavelength range of 220–400 nm, with the emission wavelength fixed at 450 nm. Emission spectra were taken over the wavelength ranging from 400 nm to 580 nm with the excitation wavelength fixed at 350 nm. The tC°-labeled and PdC-labeled DNA were analyzed for their fluorescent lifetimes using both a K2 fluorometer and a ChronosFD Fluorometer (ISS Inc., Champaign, IL, USA). The lifetime data were collected via forty data points taken in the frequency range of 10 to 250 MHz with an averaging time of 5 s per data point on the K2, and via ten data points taken in the frequency range of 10 to 250 MHz with an averaging time of 1 s per data point on the ChronosFD. At least three replicates were taken for each solution. All fits included an additional lifetime for a very short-lived component (τ < 1.0 ns). In addition, anisotropy decay data were collected for the tC°-labeled DNA on the ChronosFD to measure the rotational correlation times and limiting anisotropies for the DNA under select solution conditions via fifteen data points taken in the frequency range of 10 MHz to 300 MHz. Rotational correlation times were found using Equation 3.2, where $r_i$ is the limiting anisotropy of component $i$, $t$ is time, and $\theta_i$ is the rotational correlation time of component $i$.

Equation 3.2: $r(t) = \sum r_i e^{-t/\theta_i}$

For comparison, the longest rotational correlation times were computed by the SOMO (SOlution MOdeler) bead modeler within the UltraScan III program (128), using iM, unfolded, and duplex DNA structures. All DNA PDB files were introduced into the modeler without water or ions, allowing the program to solvate each DNA. All solvent accessible surface area and
SOMO parameters were left at their default settings. The software was used also to calculate the partial specific volume of each DNA. The folded iM structure was made by modifying the 1YBL PDB structure file (8) with Chimera to be the C20T sequence, followed by energy minimization. The unfolded and duplex DNA were made using a Python program based on the B-DNA coordinates provided by Arnott and Hukins (129). GROMACS (130) molecular dynamics calculations were utilized to find the various energetic minima of the unfolded DNA, which were used to sample the range of structures that the unfolded DNA has access to at room temperature (25 °C). The DNA was solvated with approximately 10000 water molecules and neutralized with 19 sodium ions. Once the system containing the DNA was minimized, molecular dynamics were calculated using the MD integrator, with a step size of 0.0005 ps and a total time of 10 ns, with snapshots recorded every 1000 steps. From these 1000 snapshots, every tenth one was processed through UltraScan III (128) to calculate rotational correlation times for the single strand DNA, which were then averaged to obtain the average rotational correlation time of the unfolded DNA at pH 8.0. These theoretical rotational correlation times are given in Table 3.3.

3.3.4 Analytical Ultracentrifugation: Experiment and Data Analysis

All sedimentation velocity (SV) experiments were performed using a Beckman XL-A analytical ultracentrifuge retrofitted to include a fluorescence detection system (AU-FDS, AVIV Biomedical) by Dr. Jack Correia at the University of Mississippi Medical Center and the resulting data collected were analyzed by Robert Wright, Daniel Lyons, and myself. The temperature of the AUC was calibrated using the method of Liu and Stafford (131). The 3 µM samples were loaded in centrifuge cells equipped with 1.2 cm SedVel60 centerpieces and sapphire windows. The centrifuge cells were placed into an An-60 rotor and temperature was equilibrated for approximately 1 hour before starting the run. All runs were performed at 60K
rpm (300,000 x g). The sedimentation was monitored by measuring one continuous scan of absorbance at 268 nm as a function of radial distance with a spacing of 0.002 cm.

The SV data was initially analyzed using a $g(s^*)$ analysis, which is a direct boundary model that finds an apparent sedimentation coefficient distribution, via the software program DCDT2+ (132, 133). The $g(s^*)$ distribution obtained was corrected to the density and viscosity of water at 20 °C ($s_{20, w}$). The SV data was also analyzed using a $c(s)$ analysis, a model using the Lamm equation that finds a sedimentation coefficient distribution, with the software program Sedfit (134). The meniscus position automatically chosen by DCDT2+ was also used in Sedfit for consistency. The $s_{20, w}$ obtained from the $g(s^*)$ distribution was consistent with the $s_{20, w}$ obtained from the $c(s)$ distribution. The weight average sedimentation coefficient from the $g(s^*)$ distribution was plotted as a function of pH and fit to a sigmoidal curve using the software program Origin (Northampton, MA). The SV was then analyzed using direct boundary fitting with the software program Sedanal (135). For direct boundary fitting, the data sets covering the pH range were divided into three pH sections representing the unfolded (pH 8.0, 7.5 and 7.0), partially folded (pH 6.5), and folded iM sequences (pH 6.0, 5.6, 5.0 and 4.5). The high pH data sets (pH 7.0 – 8.0) were globally fit to a single species model to determine the sedimentation coefficient of the unfolded iM sequence. The low pH (pH 4.5 – 6.0) data was then globally fit to a single species model to determine the sedimentation coefficient of the folded iM sequence. The partially unfolded iM sequence was then fit to an isomerization model where the sedimentation coefficient of the folded and unfolded conformations were fixed to the best fit values from the low and high pH fits, respectively, and the equilibrium constant of the isomerization was allowed to vary until the best fit to the data was obtained. Using the sedimentation coefficients ($s$) found, the diffusion and frictional coefficients ($D$ and $f_i$) were obtained from Equations 3.3 and 3.4.
(136, 137), where \(v\) is the partial specific volume of DNA (0.55 mL/g), \(M\) is molecular weight, \(N_A\) is Avogadro’s number, \(\rho_{\text{solvent}}\) is the density of the solvent (1 g/mL for dilute buffers), \(k\) is the Boltzmann constant \((1.38 \times 10^{-23} \text{ J/K})\), and \(T\) is temperature \((25 °C)\).

\[
\text{Equation 3.3: } s = \frac{M \cdot (1 - v \cdot \rho_{\text{solvent}})}{N_A \cdot f_t}
\]

\[
\text{Equation 3.4: } D = \frac{k \cdot T}{f_t}
\]

### 3.3.5 Folding Studies

An 8 \(\mu\)M solution of each oligo in pH 8.0 phosphate buffer was mixed into a cuvette with an equal amount of pH 2.4 phosphate buffer through the use of a RX-2000 Rapid Mixing Stopped-Flow Unit (Applied Photophysics, Leatherhead, Surrey, UK). The resulting pH 5.8 solution was monitored by circular dichroism (Olis DSM 20) at 298 nm or by monitoring fluorescence quenching (ISS K2 Fluorometer). Fluorescence was collected using a 370 nm LED excitation and a 395 nm long-pass filter in the emission path. The resulting decays were analyzed and fitted using multiple models with the Berkeley Madonna modeling software (Robert Macey and George Oster, University of California at Berkeley) to find the rate constants of folding. Steady state fluorescence spectra and circular dichroism spectra were taken before and after mixing to ensure that folding occurred. In addition, a control oligo, \(5'\)-CCCTAACCTAACCTAACCC-3’\), was used to ensure our data agreed with previously published values for folding rates obtained by a similar experimental setup (35).
3.4 Results and Discussion

3.4.1 Analysis of Thermal Stability and $pK_a$ of i-Motifs

Control experiments were performed to examine whether the tC° and PdC substitutions induced significant changes to the iM structure. Using each oligo, pH-dependent folding of the iM structure was monitored with circular dichroism (CD). Figure 3.3A shows the CD spectra recorded for the tC°4 oligo in multiple pH solutions, which was essentially identical to the unsubstituted oligos; the other substituted oligos behaved similarly. The $pK_a$ (~ 6.4; Figure 3B) for each oligo was found by plotting the CD signal at 298 nm at each pH and fitting the resultant graph with a sigmoidal function (Equation 3.1). As given in Table 3.1, the addition of tC° to the oligo had little effect on the $pK_a$. Our $pK_a$ values agree with that determined using NMR data reported by Dai et al., (1) who reported a $pK_a$ near 6.2. These $pK_a$ values suggest that tC° has little effect on iM formation in the C20T oligo at any location. The $pK_a$ values were also found using steady-state fluorescence quenching of tC° (Figure 3.3C and 3.3D). The $pK_a$ values obtained were identical within error to those found using CD (Table 3.1). In addition, analytical ultracentrifugation (AUC) experiments (Figure 3.4 and Appendix Figure A.2) gave a $pK_a$ value (~ 6.6 at low salt (10 mM KCl); ~ 6.2 at high salt (200 mM NaCl) near those obtained from CD (~ 6.4), showing that all methods are in agreement. The sedimentation coefficients obtained in these experiments also agree well with a previous AUC study performed on iM structures (138).

Thermal melts of the iMs were performed at pH 5.4 to ensure that the base substitutions did not also affect their thermal stability. As shown in Appendix Figure A.3 and Table 3.1, there is little change in the temperature at which 50% of the iM is melted ($T_m$) when compared to the wild-type C20T (~ 40 °C). The thermal melts and the $T_m$ obtained match well with those reported by Dai et al., (1) for the same oligo (~ 40 °C). Taken together, our results indicate that
the substitutions have little or no effect on the folded iM structure, its $pK_a$, or its thermal stability.

**Figure 3.3:** Formation of i-motif structures as a function of pH. (A) CD spectra of tC°4 recorded at each pH. (Not all CD spectra recorded are shown for clarity.) (B) Determination of $pK_a$ of tC°4 by monitoring the CD signal at 290 nm. (C) Fluorescence excitation spectra of tC°4 recorded at each pH. (D) Determination of $pK_a$ of tC°4 by monitoring fluorescence signal at 365 nm with pH.
Table 3.1: Tm and pKₐ values of the wild type and tC° substituted DNA folded into the iM form. All values are reported as mean ± one standard deviation (n ≥ 3).

<table>
<thead>
<tr>
<th>Oligo</th>
<th>Tm (°C)</th>
<th>pKₐ (CD)</th>
<th>pKₐ (Fluor.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C20T</td>
<td>43 ± 2</td>
<td>6.4 ± 0.1</td>
<td>N/A</td>
</tr>
<tr>
<td>tC°4</td>
<td>39 ± 1</td>
<td>6.6 ± 0.1</td>
<td>6.5 ± 0.1</td>
</tr>
<tr>
<td>tC°9</td>
<td>39 ± 1</td>
<td>6.3 ± 0.1</td>
<td>6.4 ± 0.1</td>
</tr>
<tr>
<td>tC°10</td>
<td>40 ± 3</td>
<td>6.6 ± 0.1</td>
<td>6.6 ± 0.1</td>
</tr>
<tr>
<td>tC°12</td>
<td>42 ± 1</td>
<td>6.3 ± 0.1</td>
<td>6.3 ± 0.1</td>
</tr>
<tr>
<td>tC°15</td>
<td>40 ± 1</td>
<td>6.3 ± 0.1</td>
<td>6.2 ± 0.1</td>
</tr>
</tbody>
</table>

**Figure 3.4:** Sedimentation by analytical ultracentrifugation of C20T at different pHs. Inset: pKa determined from sedimentation coefficient (s₂₀,ₜ) values.

3.4.2 Fluorescence Lifetimes of Substituted DNA Bases

Fluorescence lifetime data were collected for each tC°-substituted oligo (Appendix Table A.1) in three forms: folded iM (pH 5.4), unfolded single-strand (pH 8.0), and duplexed with the complementary strand (pH 8.0). The unfolded and duplexed DNA containing tC° showed only a
single lifetime of ~ 4.6 ns and ~ 3.7 ns, respectively, independent of position. The folded iM at pH 5.4 also had a lifetime of ~ 4.6 ns that accounted for ~ 60% of the fluorescence, but exhibited a second shorter lifetime of ~ 1.5 ns. This included position 15, which according to Dai et al. (1) should be unpaired in the folded form. To address this issue, we also examined folded PdC-substituted oligos since PdC shows a pronounced longer fluorescence lifetime when unpaired (100). Our lifetime results are presented in Table 3.2. Unlike tC°15, the PdC-15 lifetime is significantly longer than the oligos substituted at other positions, suggesting the 15th position is indeed in the loop of the iM (see also quenching data in Figure 3.7). Hence, fluorescence lifetimes of tC° can be used to monitor the global status (folded or unfolded) of the iM, but not whether a specific tC° is intercalated.

Table 3.2: Fluorescence lifetimes of PdC oligo. All values are reported as mean ± one standard deviation (n ≥ 3).

<table>
<thead>
<tr>
<th>Oligo</th>
<th>pH 5.4</th>
<th>pH 8.0 (Unfolded)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>τ (ns)</td>
<td>τ1 (ns)</td>
</tr>
<tr>
<td>PdC-4</td>
<td>2.2 ± 0.1</td>
<td>5.4 ± 0.1</td>
</tr>
<tr>
<td>PdC-9</td>
<td>2.1 ± 0.1</td>
<td>5.0 ± 0.1</td>
</tr>
<tr>
<td>PdC-10</td>
<td>2.6 ± 0.1</td>
<td>4.4 ± 0.3</td>
</tr>
<tr>
<td>PdC-12</td>
<td>2.8 ± 0.1</td>
<td>4.7 ± 0.1</td>
</tr>
<tr>
<td>PdC-15</td>
<td>3.5 ± 0.1</td>
<td>4.9 ± 0.1</td>
</tr>
</tbody>
</table>

3.4.3 Hydrodynamics of DNA Forms

Because many DNA-based nano-devices are being developed for use in aqueous environments, we determined the hydrodynamic properties of the C20T iM to better understand its size, shape, and rigidity in solution compared to its unfolded and duplexed counterparts. To determine its size and rigidity in solution, fluorescence anisotropy decays of tC° were collected.
for each oligo in the duplex, unfolded single-strand, and iM forms and fitted using Equation 3.2 to find the rotational correlation times (θ). A shorter 0.4 – 0.6 ns rotational correlation time was observed in fitting the data for all forms of DNA, regardless of position. We attribute these to local motions of the tC°. The duplex rotational correlation times are consistent with a very rigid structure in solution (Table 3.3). The unfolded single-strand DNAs at pH 8.0 have much faster rotational correlation times than the duplex form (Table 3.3), indicating a more compact size due to the increased flexibility of the single strand. In the iM form, the relatively short rotational correlation times indicate a more compact structure. The tC°4 and tC°9 oligos have a slightly shorter rotational correlation time compared to the other three oligos. We suggest that this arises from the position of tC°4 and tC°9 since the rotational correlation times are consistently shorter in the unfolded and duplex forms as well. The values obtained for all three forms of DNA are similar to those predicted from the molecular model of each DNA structure using Ultrascan III (Theoretical value in Table 3.3). These values did not significantly change when the ionic strength was increased by adding 150 mM KCl (Appendix Table A.2).

Additionally, AUC studies were performed to analyze the hydrodynamics of the unfolded strand and folded iM in solution. The sedimentation coefficients obtained (s; shown in Figure 3.4) suggest a change in structure from the spherical iM to the larger, freely moving random coil as pH is increased. This change is in good agreement with the rotational correlation times observed, indicating that the theoretical models of the DNA for all three structures accurately predict their behavior in solution.

Taken together, our hydrodynamics data characterized the iM form of DNA as similar to the duplex form in terms of rigidity, but more compact. It may be possible to design better nanomaterials and drug delivery vehicles by taking advantage of these features. For example, a
pH-dependent nanomaterial based on single-strand DNA iMs might be developed that is smaller than its duplex counterpart while retaining structural integrity.

**Table 3.3**: Fluorescence Rotational Correlation Times (θ) of tC°-containing oligos at 25 °C. All values are reported as mean ± one standard deviation (n ≥ 3).

<table>
<thead>
<tr>
<th>Oligo</th>
<th>pH 5.4 θ (ns)</th>
<th>pH 8.0 θ (ns)</th>
<th>Duplex θ (ns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>tC°4</td>
<td>1.9 ± 0.2</td>
<td>2.9 ± 0.2</td>
<td>8.3 ± 0.2</td>
</tr>
<tr>
<td>tC°9</td>
<td>2.2 ± 0.1</td>
<td>3.1 ± 0.1</td>
<td>8.5 ± 0.1</td>
</tr>
<tr>
<td>tC°10</td>
<td>2.6 ± 0.1</td>
<td>3.6 ± 0.1</td>
<td>9.3 ± 0.2</td>
</tr>
<tr>
<td>tC°12</td>
<td>2.9 ± 0.2</td>
<td>3.7 ± 0.2</td>
<td>8.7 ± 0.1</td>
</tr>
<tr>
<td>tC°15</td>
<td>2.7 ± 0.1</td>
<td>3.6 ± 0.4</td>
<td>8.9 ± 0.1</td>
</tr>
<tr>
<td>Theoretical</td>
<td>2.7</td>
<td>3.7 ± 0.1</td>
<td>8.8</td>
</tr>
</tbody>
</table>

**Figure 3.5**: Global folding of the i-motif as monitored by CD. (A) Example normalized CD spectra of tC°4 recorded before (pH 8.0; red) and after (pH 5.8; black) the pH jump. (B) Five kinetic traces for tC°4 determined by monitoring the CD signal at 298 nm with time after the pH jump to induce folding. The solid lines show fits to the data using a three-state sequential model.
3.4.4 Kinetic Mechanism of iM Folding

To elucidate the folding rate of the iM, CD signal traces at 298 nm were monitored with time as the pH of the DNA solution rapidly changed from 8.0 to 5.8. Typical data are shown in Figure 3.5. Spectra were taken before and after the pH change (Figure 3.5A) to ensure folding occurred. As CD monitors the overall DNA strand folding, we consider the rate constants obtained from analyses of these data as “global”. As the signal changes were not fitted well by a single exponential function, multiple models were tested, including double exponential, sequential, and equilibrium mechanisms. The single-strand $\rightarrow$ intermediate $\rightarrow$ folded sequential model was found to best fit the folding observed (lowest root mean squared deviation values; Figure 5B).

As shown in Table 3.4, all of the substituted oligos showed global folding that was (within error) similar to the wild-type (C20T), with two folding rate constants of $\sim 0.60$ s$^{-1}$ and $\sim 0.01$ s$^{-1}$, suggesting that tC° substitution does not affect the folding pathway of the structure. These rate constants did not change significantly with doubling or tripling the concentration of oligo used (Table A.3). The fractional signal amplitude ($\sim 0.90$) associated with the faster rate constant suggests that the intermediate state is very close in structure to the final state. Consecutive mechanisms can be fitted identically in the order fast-to-slow or slow-to-fast, depending on the signal arising from the intermediate structure. Since the intermediate and final forms of the iM are so similar in CD signal, we interpret this to indicate the faster folding step is first, followed by slower bond rearrangement. The amplitude change of the fast process supports a folding mechanism similar to that in Figure 3.6, where the mostly-folded intermediate structure is rapidly formed, followed by a slower rearrangement to yield the final iM conformation. However, the hydrogen bonding pattern of the intermediate structure (Figure 3.6B) cannot be
discerned by CD, thus the explicit hydrogen bonding pattern is not shown in Figure 3.6B. The global rate constants obtained for our C20T oligos are on the same time scale as those obtained from our control experiments using the i-motif oligo of Liu et al. (35), which has a rate constant of 0.15 s$^{-1}$ for pH-dependent folding.

**Table 3.4:** The rate constants of folding from circular dichroism. All values are reported as mean ± one standard deviation (n ≥ 3).

<table>
<thead>
<tr>
<th></th>
<th>C20T</th>
<th>tC°4</th>
<th>tC°9</th>
<th>tC°10</th>
<th>tC°12</th>
<th>tC°15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast CD Folding Rate Constant (s$^{-1}$) (A to B)</td>
<td>0.61 ± 0.09</td>
<td>0.68 ± 0.11</td>
<td>0.59 ± 0.13</td>
<td>2.88 ± 0.68</td>
<td>0.59 ± 0.20</td>
<td>0.51 ± 0.07</td>
</tr>
<tr>
<td>Fractional ΔCD* (A to B)</td>
<td>0.89 ± 0.07</td>
<td>0.87 ± 0.07</td>
<td>0.87 ± 0.06</td>
<td>0.93 ± 0.06</td>
<td>0.88 ± 0.05</td>
<td>0.78 ± 0.10</td>
</tr>
<tr>
<td>Slow CD Folding Rate Constant (s$^{-1}$) (B to C)</td>
<td>0.03 ± 0.02</td>
<td>0.05 ± 0.02</td>
<td>0.06 ± 0.02</td>
<td>0.03 ± 0.04</td>
<td>0.07 ± 0.02</td>
<td>0.05 ± 0.02</td>
</tr>
<tr>
<td>Fractional ΔCD* (B to C)</td>
<td>0.11 ± 0.04</td>
<td>0.13 ± 0.04</td>
<td>0.13 ± 0.05</td>
<td>0.07 ± 0.10</td>
<td>0.12 ± 0.02</td>
<td>0.22 ± 0.05</td>
</tr>
</tbody>
</table>

*The total ΔCD signal amplitude was normalized from 0.0 (initial) to 1.0 (final). Data were fit to the mechanism in Figure 3.6 (A → B → C), where the first rate is faster than the second.

**Figure 3.6:** The folding mechanism for the C20T i-motif based on CD (top; black arrows) and fluorescence quenching data (bottom; blue arrows) where the transition from unfolded DNA (A) to intermediate structures (B$_1$ and B$_2$) to the final folded iM (C) is shown. Average rate constants for each step in the mechanism are given above the arrows. The color-coding of residues and their hydrogen bonds are green for tC°4, red for tC°9, orange for tC°10, pink for tC°12, and navy for tC°15. The hydrogen bonds in B are not shown since they are not observable by CD; however, the pattern in B is likely similar to B$_2$ observed in the fluorescence quenching mechanism. Overall, both CD and fluorescence can be interpreted with the same folding mechanism that progresses through intermediate structures.
To detail the base-by-base mechanism of iM folding, traces of the quenching of each individual tC° were obtained. Rapid quenching of tC° is observed when the iM forms. Kinetic data were collected by monitoring fluorescence after a pH jump from 8.0 to 5.8. Typical initial and final fluorescence spectra collected in this experiment are shown in Figure 3.7A. The average fluorescence decays for each base substitution are shown in Figure 3.7B, where zero time is the initial fluorescence at pH 8.0 before mixing. We consider these fluorescence changes to reflect the local conditions of the individual dC residues during folding, which include stacking and C-C⁺ bond formation. The decays were fitted to the sequential model of unprotonated single-strand → protonated single-strand → intermediate → folded as this gave the lowest root-mean squared values, as well as being consistent with the global folding mechanism. As shown in Figure 7B, all fluorescence quenches nearly instantaneously (within the first 0.2 s of mixing) to at least half of its initial pH 8.0 signal (corrected for dilution). We attribute this initial quenching to an effect of protonation of nearby cytosines in the single-strand, as it occurs immediately after mixing and before the iM has folded, and is independent of position. To confirm the initial quenching is due to protonation of neighboring cytidines, we tested a cropped tC°4 sequence (5’- TTC-tC°-CTA -3’) that does not form an iM as determined by CD (Figure A.4). The cropped sequence showed a 40% decrease in fluorescence immediately upon mixing (Figure A.4). We attribute the slower, additional quenching of tC°4 oligos to iM formation. Like its related fluorescent base tC, the tC° has a very low pKₐ (~ 1.0) (139), and unincorporated tC° nucleoside does not exhibit such pH-dependent quenching (Figure 3.7A). The differences between pH effects on the free nucleoside and the tC° incorporated in the DNA strand strongly suggests that quenching arises from nearest-neighbor interactions within the
single-strand oligo, in agreement with studies of the electronic coupling interactions among hemi-protonated dC residues in iMs (140).

**Table 3.5:** The rate constants of folding for five different positions in the i-motif structure found from fluorescence quenching. All values are reported as mean ± one standard deviation (n ≥ 3).

<table>
<thead>
<tr>
<th></th>
<th>tC°4</th>
<th>tC°9</th>
<th>tC°10</th>
<th>tC°12</th>
<th>tC°15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fraction of pH 8 Signal Remaining After Initial Quenching</td>
<td>0.28 ± 0.01</td>
<td>0.36 ± 0.01</td>
<td>0.41 ± 0.01</td>
<td>0.31 ± 0.01</td>
<td>0.47 ± 0.02</td>
</tr>
<tr>
<td>Fast Fluorescence Rate Constant (s⁻¹) (A to B₁/B₂)</td>
<td>1.150 ± 0.060</td>
<td>0.920 ± 0.060</td>
<td>1.460 ± 0.310</td>
<td>0.410 ± 0.070</td>
<td>0.412 ± 0.020</td>
</tr>
<tr>
<td>Fractional ΔF/F₀ * (A to B₁/B₂)</td>
<td>-0.07 ± 0.01</td>
<td>-0.07 ± 0.01</td>
<td>-0.06 ± 0.01</td>
<td>-0.04 ± 0.01</td>
<td>0.15 ± 0.01</td>
</tr>
<tr>
<td>Slow Fluorescence Rate Constant (s⁻¹) (B₁/B₂ to C)</td>
<td>0.018 ± 0.003</td>
<td>0.012 ± 0.004</td>
<td>0.038 ± 0.036</td>
<td>0.010 ± 0.001</td>
<td>0.012 ± 0.003</td>
</tr>
<tr>
<td>Fractional ΔF/F₀ * (B₁/B₂ to C)</td>
<td>-0.01 ± 0.01</td>
<td>-0.01 ± 0.01</td>
<td>-0.02 ± 0.01</td>
<td>0.02 ± 0.01</td>
<td>-0.13 ± 0.01</td>
</tr>
</tbody>
</table>

*The total ΔF/F₀ signal amplitude was normalized from 0.0 (initial) to 1.0 (final). Data were fit to the mechanism in Figure 3.6 (A → B → C), where the first rate is faster than the second.

**Figure 3.7:** (A) The fluorescence spectra of the tC°4 oligo recorded before (pH 8.0; red) and after (pH 5.8; dashed-green) the pH jump. The fluorescence spectra of the tC° nucleoside alone before (pH 8.0; blue) and after (pH 5.8; dashed-black) is shown for comparison. The free nucleoside’s fluorescence changes very little with pH compared to when it is incorporated into DNA. (B) The average fluorescence intensity decay for each tC°-substituted oligo with time after folding of the iM structure was induced by the pH change. The color-coding in (B) is the same as that shown in Figure 3.6 (green for tC°4, red for tC°9, orange for tC°10, pink for tC°12, and navy for tC°15).
As shown in Table 3.5, following the initial quench, three of the substituted oligos (tC°4, tC°9, and tC°10) show very similar rate constants for base pairing. These three positions pair with a rate constant of ~ 1 s\(^{-1}\), which is slightly faster than the fastest rate constant for global folding via CD (~0.5 s\(^{-1}\)). As quenching is influenced by both hydrogen bonding and stacking interactions (Figure 3.7), we infer from our data that pairing of tC°4, tC°9, and tC°10 occurs before the structure in Figure 3.6B is formed. We denote this stage of folding as B\(_1\) (Figure 3.6). We denote this stage of folding as B\(_1\) (Figure 3.6). The tC°12 and tC°15 have a fast rate constant of ~ 0.4 s\(^{-1}\), which is similar to the fast global rate constant of ~ 0.5 s\(^{-1}\) determined from CD data. We interpret this to mean that the 12\(^{th}\) and 15\(^{th}\) positions move into the intermediate state (Figure 3.6B\(_2\)) simultaneously with the global folding of the iM as observed by CD (Figure 3.6B), suggesting that the base stacking occurs at the same rate as the hydrogen bonds formation. All positions exhibit a second, consistently slower rate constant of ~ 0.01 s\(^{-1}\), which is also observed in the global folding obtained from CD spectra. We interpret this as the transition from the structure in Figure 3.6B\(_2\) to the final structure (Figure 3.6C). Both fluorescence and CD can be interpreted with the same mechanism; however, fluorescence gives insight into structural details that are invisible to CD (Figure 3.6 B\(_1\) and B\(_2\)).

It is noteworthy that in Figure 3.7B, the final observed fluorescence intensity for tC° is very position dependent. We interpret these final intensities from Figure 3.7B by using Figure 3.6C. When equilibrium has been achieved, tC°4 is located in the middle of a cytosine run, with base stacking above and below it. Hence, the electronic coupling of tC° with surrounding dC residues would explain why it has the lowest fluorescence than other base-paired positions. In Figure 3.6C, both tC°9 and tC°12 are at the ends of the C-C\(^{+}\) stack, and equivalent effects on their fluorescence are observed. Finally, the tC°10 hydrogen-bond is broken in the predominant
equilibrium form of the iM (Figure 3.6C), which would reduce the effect of pairing on fluorescence. The tC°10 has the highest final fluorescence of any of the intercalated tC° residues.

In contrast to the intercalated tC° residues, following the initial pH-induced quench, the fluorescence intensity of the tC°15 increases to 60% of its pH 8 fluorescence, followed by a slower decrease to a final intensity that is 50% of the pH 8 signal (Figure 3.7B). The shape of the fluorescence trace for tC°15 is consistent with an intermediate structure where tC°15 is unpaired and unstacked. The high final fluorescence intensity observed with tC°15 is in agreement with the lifetime data from the PdC-15 substitution, both of which are consistent with the 15 position being in a loop in the fully folded iM. Our interpretation of the tC°15 data is that the 15th position is initially stacked in the single strand (Figure 3.6A), but during the folding process, becomes less stacked with the neighboring dC at the 14th position (Figure 3.6B). The subsequent restacking of tC°15 in Figure 3.6C is responsible for the slow decrease in its fluorescence.

3.5 Chapter Conclusions

Given the consistency of all the data described above, we suggest that the lower folding pathway outlined in Figure 3.6 is correct for the formation of the i-motif structure from the C20T sequence. In the course of determining this folding pathway, we established that fluorescent dC analogs can be used to monitor individual residues of an iM structure. We demonstrated the ability of tC° and PdC to track iM formation and structure from their fluorescence properties including lifetime measurements, anisotropy decay, pH-dependent intensity, and the time-dependent intensities following a pH jump. We established that the C-C°+ hydrogen bonding of certain bases (in the case of C20T, the 4 to 13, 9 to 18, and 10 to 19 bonds) initiate the folding of the iM structure. We also showed that substitutions in the loop regions of iMs, as shown by tC°15, give a distinctly different kinetic signature from those bases that are intercalated. Taken
together, this paper has laid the foundation for using fluorescent dC analogs as general tools for following structural changes during iM formation.

Our work benefitted from the prior establishment of the equilibrium structure of C20T by Dai et al. (1), such that we could correlate our combined data with a known molecular iM structure. Based on our studies using AUC, molecular modeling, CD, and fluorescent dC analogs described in this paper, this same analytical scheme can be used to determine the folding and final structure of more complex iMs that are known to exist, such as those found in the promoters of human bcl-2 and VEGF genes. Overall, this work has expanded the knowledge of iMs and their formation, and thus may be helpful to improving bio- and nano-material applications that utilize iMs, as well as helping to better understand iM formation under in vivo conditions.

**PUBLICATION NOTICE**

The work done in this chapter has been collected and published:

CHAPTER 4: EFFECT OF INTERIOR LOOP LENGTH ON THE THERMAL STABILITY AND PKₐ OF I-MOTIF DNA

4.1 Chapter Abstract

The four-stranded i-motif (iM) conformation of cytosine-rich DNA has importance to a wide variety of biochemical systems that range from their use in nanomaterials to potential roles in oncogene regulation. Fundamental studies to understand how loop length affects iM physical properties are described in this report. We characterize both the thermal stability and pKₐ of iMs with differing loop lengths in dilute solutions and with molecular crowding agents. Our work has demonstrated that intramolecular iMs with longer central loops form at higher pH and have higher thermal stability than iMs with outer loops of equal loop length. Our studies show that increases in thermal stability of iMs when molecular crowding agents are present is dependent on the loop that is lengthened in the iM structure. However, the increase in pKₐ for iMs when molecular crowding agents are present is unrelated to length. We also determined the proton activity of solutions containing PEG and note that the increase in iM pKₐ is only partially due to this effect.

4.2 Introduction

Discovered in 1962 (9) and 1993 (141), respectively, G-quadruplexes (G4s) and i-motifs (iMs) are DNA structures that have a distinct topological variance from the well-known Watson-Crick DNA structure. These structures have been linked to transcription and its regulation in some genes (86, 87, 89, 95, 108-110, 142-144). Potential G4 and iM forming sequences are highly associated with the promoter regions of many genes; thus, they are widely studied as
possible unique drug targets (2, 43, 90, 95, 111-113, 143, 144). G4s are formed from guanine-rich strands of DNA that Hoogsten hydrogen bond to form a tetrad structure, stabilized by a metal cation (usually potassium). The strand opposite of the guanine-rich one is cytosine-rich and can form a structure known as an iM, which are stabilized by multiple protonated cytosines that form stable hydrogen bonded pairs (Figure 4.1). In 2013, Balasubramanian et al., (13) published one of the first papers that confirmed G4s exist in vivo. Multiple later reports also give evidence for the existence G4s in vivo, both as DNA and RNA structures (145-149). No similar studies for iMs have been performed, so their existence in vivo is still hypothetical. Unlike G4s, iMs have not yet been considered widely as a potential drug target since the pH in the nucleus should not be much different than cytosol pH (~ 7.3) (150, 151) while the pKₐ of a typical iM is ~ 6.0. However, recent studies (38, 114, 152) have shown that iMs indeed form at neutral pH when molecular crowding agents are present. Since cells are crowded with biopolymers, this finding suggests iMs may be viable in vivo and deserve further study for comparison with the better-characterized G4 structures.

In addition to their possible biological relevance, iMs are currently being used as a biomaterial. Examples include biocompatible pH sensors and drug delivery vehicles. These applications take advantage of their unique topology and proton dependence (6, 153-156). For example, Song et al., (155) reported an iM-based gold nanomaterial that could deliver doxorubicin specifically to human cervix adenocarcinoma cells. The drug is intercalated into duplex DNA containing mismatching pairs that is attached to the gold nanomaterial. Once the nanomaterial reaches the cancer cells, the low pH near the cells causes the DNA to fold into an iM, releasing the intercalated doxorubicin. The properties of such nanomaterials are linked to the properties of the iMs used in their construction, which include the iM’s thermal stability and pKₐ.
Thus, better understanding the properties of iMs will allow better control over the physical properties of iM-based nanomaterials.

Sequences that form iMs are usually derived from the corresponding G4 sequence. To locate potential G4s in any given sequence, Quadfinder (157) and Quadparser (158) were created. These programs find potential G4s within strands by matching the sequence $G_{x_1} N_{y_1} G_{x_2} N_{y_2} G_{x_3} N_{y_3} G_{x_4}$, where $N$ are loop residues, $y_{1,3}$ are the number of bases in each loop, $G$ are guanines, and $x_{1,4}$ are the number of guanines in a row. The default search in both Quadfinder and Quadparser is the sequence of $G_3 N_{1-7} G_3 N_{1-7} G_3$. This default sequence has been based on multiple systematic studies that have shown that lengthening any loop of the G4 will decrease the thermal stability of the G4s, influence the type G4 formed, and increase the potassium dependence of G4 formation (20-23, 159). Although $x$ and $y$ in Quadfinder (157) can be any value, the default sequence is still the most commonly used sequence in the literature (160-163). This search algorithm has also been used to find potential iMs on the opposite strand. However, the most stable G4 and iM formed from the same DNA duplex sequence of the c-Myc gene do not share the same loop lengths with the iM having much longer loops than its G4 counterpart (68). The difference in loop length between structures suggests that a different search algorithm might be necessary to locate the most stable iMs.

To determine the optimal loop length to use in finding potential iMs, we examined the effect of loop length on an iM’s structural stability. We determined the effects of loop length in dilute buffers and in molecular crowded buffers that mimic intracellular conditions. We examined 16 different oligos differing only in loop length and position (Table 4.1; Figure 4.1). We used 40% PEG-300 (w/w) as the crowding agent since low molecular weight PEGs are the most commonly used molecular crowding agents in the literature on both iMs and G4s (30, 38,
Using circular dichroism (CD) spectroscopy and ultraviolet (UV) absorbance spectroscopy, we determined the thermal stability and pKₐ of each oligo under dilute and molecular crowded conditions. We also investigated the effects of PEG size by using PEG-1500 and PEG-3350. Earlier studies reported that a high percentage (> 30%) of other hydrophilic molecules, including PEG, shifted the apparent pH of buffered solutions (168-172). To observe the effect of PEG on apparent pH, we analyzed six different pH indicators and determined their pKₐ values with and without PEG. Increases in the pKₐ of iMs larger than those arising from a change in proton activity (as observed with the indicators) are attributed to PEG’s crowding effect. By obtaining the details about iMs’ physical properties in this report, our results may help guide the creation of novel iM-based nanomaterials. Further, our results allow for a more accurate search method to be created for iMs when the need for complementary G4-forming is unnecessary to be stable.

**Figure 4.1:** Representative folding of a random coil into an iM for oligos differing by loop length and position: (A) A reference structure (T1) where all loops contain one thymine (colored spheres); (B) first (green; Mod1 oligos), (C) central (blue; Mod2 oligos), and (D) third (yellow; Mod3 oligos) loops have 3-20 thymines.
4.3 Experimental

4.3.1 Oligos

All DNA oligos (Table 4.1) were synthesized through standard solid-phase chemistry by Midland Certified Reagent Company, Inc. (Midland, TX, USA). The oligos were stored in 10 mM Tris, 1 mM EDTA Buffer (pH 8.0) until used. The oligo sequence chosen (5’-CCC T CCC T CCC T CCC-3’) is an artificial sequence created to eliminate the possibility of interfering alternate structures that could affect the experiment. The loop length substitutions were thymine repeats of varying length (3, 5, 10, 15, and 20 thymines).

<table>
<thead>
<tr>
<th>Name</th>
<th>5’-Sequence-3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>CCC T CCC T CCC T CCC</td>
</tr>
<tr>
<td>Mod1TX</td>
<td>CCC X CCC T CCC T CCC</td>
</tr>
<tr>
<td>Mod2TX</td>
<td>CCC T CCC X CCC T CCC</td>
</tr>
<tr>
<td>Mod3TX</td>
<td>CCC T CCC T CCC X CCC</td>
</tr>
</tbody>
</table>

Table 4.1: Oligos used and their respective sequences, where X is the placement and number of the thymine repeats (3, 5, 10, 15, 20) in the loop modification.

4.3.2 pKₐ Determination

In keeping with the prior literature, we define the pKₐ for iM formation as the pH at which 50% of the oligo is folded into the iM. To determine pKₐ for iM folding, DNA solutions (4 µM) were made in 30 mM sodium cacodylate buffers with pH ranging from 5.4 to 8.0. Two more sets of DNA solutions (4 µM) were made in these buffers, containing 40% PEG-300 or 40% PEG-3350. All reagents were purchased from Fisher Scientific (Waltham, MA, USA). To eliminate intermolecular structures, prior to the experiments, the solutions were heated to ~ 80 °C for ten minutes and then cooled to room temperature. CD spectra of DNA solutions were collected on an Olis DSM 20 Circular Dichroism (CD) instrument (Olis, Inc. Bogart, GA, USA)
over the wavelength range of 250 to 320 nm at 20 °C with an integration time that adjusted with absorbance. The CD signals observed at 298 nm were then plotted against pH and were fit with Equation 4.1 to obtain $\text{pK}_a$.

\textbf{Equation 4.1:} \[ \text{Signal}_{\text{Total}} = \frac{\text{Signal}_{\text{Folded}} - \text{Signal}_{\text{Unfolded}}}{1 + 10^{(\text{Cooperativity}*(\text{pH}-\text{pK}_a))}} + \text{Signal}_{\text{Unfolded}} \]

4.3.3 Thermal Stability and Denaturation of iMs

Thermal melts of iMs from 20 °C to 80 °C were collected for all oligos (4 μM) at pH 5.4 with and without 40% PEG-300 or 40% PEG-3350 by monitoring CD signals and/or UV absorbance. After verifying iM melting on the CD, thermal melts were also performed on a Cary 100 UV-Visible spectrometer (Agilent Technologies, Santa Clara, CA, USA) at 260 nm. The thermal denaturations were identical with either spectrometer. By assuming a two-state equilibrium model, we determined the melting temperature ($T_m$). All fractions were normalized between 0 and 1 prior to fitting. The change in free energy at 37 °C ($\Delta G^{37 \degree C}$) was found by equation 4.2, where $[N]$ and $[U]$ are the concentrations of the iM and random coil forms of DNA, respectively; $R$ is the gas constant; and $T$ is physiological temperature, 310 K (173). In addition, to ensure the iMs formed were indeed intramolecular, a range of concentrations (0.6 μM, 4 μM, and 20 μM) was analyzed for all oligos. All oligos were found to have the same $T_m$ and $\Delta G^{37 \degree C}$ regardless of concentration, suggesting all oligos form an intramolecular iM structure.

\textbf{Equation 4.2:} \[ K = \frac{[N]}{[U]} = e^{-\frac{\Delta G^{37 \degree C}}{RT}} \]

4.3.4 Relation of Measured pH to Proton Activity

We used the method of Gaboriaud (171) to compare apparent pH measured by a high performance, micro glass-body, Tris compatible, combination pH electrode with a platinum
junction and 3M potassium chloride filling solution (Model pHE-11, GeneMate, BioExpress, Kaysville, UT, USA) with proton activity in PEG solutions. Six pH indicators were used: bromothymol blue; 4-nitrophenol; 3,4-dinitrophenol; methyl red; phenol red; and cytidine monophosphate. The pH titration of each indicator was performed on a Cary 100 UV-Vis spectrometer over a wavelength range of 250 nm – 800 nm. Each indicator was made with a 30 mM sodium cacodylate buffer with 0% PEG, 40% PEG-300, or 40% PEG-3350. The sample size was 3 mL, and 0.05 M NaOH was titrated into the sample at increments of 0.02 to 0.10 mL, depending on the magnitude of the previous pH change. Once the spectral peaks no longer changed with pH, the absorbance values of each peak were plotted against their respective pH after correction for dilution.

4.4 Results and Discussion

4.4.1 pKₐ Determination

The pKₐ of each oligo is given in Table 4.2 with representative structures shown in Figure 4.2. The trend in pKₐ values observed for Mod2 oligos is distinctly different from the other loop modifications (Mod1 and Mod3). Mod2 oligos are able to maintain iM structure at higher pH and longer loop lengths than the other modifications. For example, with ten bases in their respective loops, Mod2T10 has a pKₐ of 6.62 while Mod1T10 and Mod3T10 exhibit a pKₐ of 6.13. The pKₐ remains the same for the Mod2 oligos until more than ten bases have been added to the loop, which suggests that the length of the central loop has less of an effect on the pKₐ than the outer loops. The cooperativity parameter of all oligos examined did not change significantly (~ 3 for all iMs formed). This suggests that extending the loops does not affect how the iM unfolds when shifting from an acidic to a more basic pH. These findings suggest that central loops (e.g., Mod2) shorter than fifteen bases and outer loops (e.g., Mod1 and Mod3)
shorter than ten bases are potentially viable in vivo. This is remarkably different from the results for G4s where all loops are equally limited to seven bases to be potentially viable in vivo.

**Table 4.2:** The pKₐ, Tₘ, and ΔG°₃⁷°C of all oligos in 30 mM sodium cacodylate ± 40% PEG-300. The ΔpKₐ, ΔTₘ, and ΔΔG°₃⁷°C are also shown to emphasize the difference upon adding 40% PEG. All values are reported as mean ± one standard deviation (n = 3), except ΔG°₃⁷°C, which is reported as the mean.

<table>
<thead>
<tr>
<th></th>
<th>pKₐ</th>
<th>pKₐ</th>
<th>ΔpKₐ</th>
<th>Tₘ (°C)</th>
<th>Tₘ (°C)</th>
<th>ΔTₘ (°C)</th>
<th>ΔG°₃⁷°C (kcal/mol)</th>
<th>ΔG°₃⁷°C (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No PEG</td>
<td>40% PEG</td>
<td>No PEG</td>
<td>40% PEG</td>
<td>No PEG</td>
<td>40% PEG</td>
<td>No PEG</td>
<td>40% PEG</td>
</tr>
<tr>
<td>T1</td>
<td>6.67 ± 0.01</td>
<td>7.03 ± 0.05</td>
<td>0.36 ± 0.05</td>
<td>54.9 ± 0.4</td>
<td>58.5 ± 0.5</td>
<td>3.6 ± 0.5</td>
<td>-2.2</td>
<td>-2.9</td>
</tr>
<tr>
<td>Mod1T3</td>
<td>6.66 ± 0.01</td>
<td>7.04 ± 0.05</td>
<td>0.38 ± 0.05</td>
<td>47.7 ± 0.4</td>
<td>53.3 ± 0.5</td>
<td>5.6 ± 0.5</td>
<td>-0.9</td>
<td>-2.0</td>
</tr>
<tr>
<td>Mod1T5</td>
<td>6.25 ± 0.04</td>
<td>6.71 ± 0.12</td>
<td>0.46 ± 0.12</td>
<td>41.9 ± 0.3</td>
<td>48.8 ± 0.4</td>
<td>6.9 ± 0.4</td>
<td>-0.6</td>
<td>-1.3</td>
</tr>
<tr>
<td>Mod1T10</td>
<td>6.14 ± 0.02</td>
<td>6.54 ± 0.05</td>
<td>0.40 ± 0.05</td>
<td>27.8 ± 0.4</td>
<td>34.0 ± 0.6</td>
<td>6.2 ± 0.6</td>
<td>0.7</td>
<td>0.1</td>
</tr>
<tr>
<td>Mod1T15</td>
<td>nd.*</td>
<td>nd.*</td>
<td>nd.*</td>
<td>19.5 ± 0.4</td>
<td>25.8 ± 1.0</td>
<td>6.3 ± 1.0</td>
<td>3.4</td>
<td>1.6</td>
</tr>
<tr>
<td>Mod1T20</td>
<td>nd.*</td>
<td>nd.*</td>
<td>nd.*</td>
<td>15.0 ± 0.5</td>
<td>21.0 ± 0.5</td>
<td>6.0 ± 0.5</td>
<td>4.5</td>
<td>2.3</td>
</tr>
<tr>
<td>Mod2T3</td>
<td>6.66 ± 0.03</td>
<td>7.16 ± 0.05</td>
<td>0.50 ± 0.05</td>
<td>54.7 ± 0.3</td>
<td>66.7 ± 0.7</td>
<td>12.0 ± 0.7</td>
<td>-2.3</td>
<td>-3.3</td>
</tr>
<tr>
<td>Mod2T5</td>
<td>6.68 ± 0.04</td>
<td>7.14 ± 0.05</td>
<td>0.46 ± 0.05</td>
<td>54.5 ± 0.6</td>
<td>64.0 ± 0.6</td>
<td>9.5 ± 0.6</td>
<td>-2.7</td>
<td>-4.0</td>
</tr>
<tr>
<td>Mod2T10</td>
<td>6.62 ± 0.02</td>
<td>6.98 ± 0.07</td>
<td>0.36 ± 0.07</td>
<td>47.1 ± 0.2</td>
<td>55.8 ± 0.8</td>
<td>8.7 ± 0.8</td>
<td>-1.6</td>
<td>-2.8</td>
</tr>
<tr>
<td>Mod2T15</td>
<td>6.33 ± 0.02</td>
<td>6.71 ± 0.06</td>
<td>0.38 ± 0.06</td>
<td>40.6 ± 0.1</td>
<td>49.8 ± 0.7</td>
<td>9.2 ± 0.7</td>
<td>-0.7</td>
<td>-1.7</td>
</tr>
<tr>
<td>Mod2T20</td>
<td>5.98 ± 0.05</td>
<td>6.48 ± 0.06</td>
<td>0.50 ± 0.06</td>
<td>36.3 ± 0.6</td>
<td>46.2 ± 0.8</td>
<td>9.9 ± 0.8</td>
<td>-0.6</td>
<td>-1.5</td>
</tr>
<tr>
<td>Mod3T3</td>
<td>6.70 ± 0.03</td>
<td>7.00 ± 0.06</td>
<td>0.30 ± 0.06</td>
<td>47.0 ± 0.4</td>
<td>55.4 ± 0.3</td>
<td>8.4 ± 0.4</td>
<td>-0.8</td>
<td>-2.0</td>
</tr>
<tr>
<td>Mod3T5</td>
<td>6.27 ± 0.02</td>
<td>6.63 ± 0.08</td>
<td>0.36 ± 0.08</td>
<td>38.9 ± 0.6</td>
<td>45.5 ± 0.9</td>
<td>6.6 ± 0.9</td>
<td>-0.3</td>
<td>-0.9</td>
</tr>
<tr>
<td>Mod3T10</td>
<td>6.12 ± 0.02</td>
<td>6.42 ± 0.05</td>
<td>0.30 ± 0.05</td>
<td>26.9 ± 0.1</td>
<td>34.3 ± 1.3</td>
<td>7.4 ± 1.3</td>
<td>1.1</td>
<td>0.3</td>
</tr>
<tr>
<td>Mod3T15</td>
<td>nd.*</td>
<td>nd.*</td>
<td>nd.*</td>
<td>13.2 ± 0.3</td>
<td>21.2 ± 1.4</td>
<td>8.0 ± 1.4</td>
<td>5.6</td>
<td>3.2</td>
</tr>
<tr>
<td>Mod3T20</td>
<td>nd.*</td>
<td>nd.*</td>
<td>nd.*</td>
<td>10.2 ± 0.2</td>
<td>17.9 ± 0.3</td>
<td>7.7 ± 0.3</td>
<td>5.5</td>
<td>3.2</td>
</tr>
</tbody>
</table>

*Tₘ of oligo ≤ 20 °C, thus pKₐ not determined at 20 °C.
4.4.2 Effects of PEG on Apparent pH

Given our data and other literature, we investigated the difference between the pH determined by electrode measurement and the activity of protons in the PEG solutions. We examined the absorbance changes of six different pH indicators in 30 mM sodium cacodylate buffer. We monitored the change in UV absorbance with change in pH for all indicators with and without 40% PEG-300. We interpret the change in

<table>
<thead>
<tr>
<th>Literature (Reference)</th>
<th>$pK_a$ 4-Nitrophenol</th>
<th>$pK_a$ Bromothymol Blue</th>
<th>$pK_a$ 3,4-Dinitrophenol</th>
<th>$pK_a$ Methyl Red</th>
<th>$pK_a$ Phenol Red</th>
<th>$pK_a$ Cytidine Monophosphate</th>
</tr>
</thead>
<tbody>
<tr>
<td>No PEG</td>
<td>7.10 ± 0.04</td>
<td>7.07 ± 0.03</td>
<td>5.31 ± 0.02</td>
<td>5.04 ± 0.04</td>
<td>7.98 ± 0.05</td>
<td>4.35 ± 0.06</td>
</tr>
<tr>
<td>PEG-300</td>
<td>7.33 ± 0.04</td>
<td>7.35 ± 0.04</td>
<td>5.52 ± 0.04</td>
<td>5.32 ± 0.03</td>
<td>8.27 ± 0.02</td>
<td>4.57 ± 0.04</td>
</tr>
<tr>
<td>PEG-3350</td>
<td>7.35 ± 0.06</td>
<td>7.36 ± 0.04</td>
<td>5.53 ± 0.02</td>
<td>5.28 ± 0.03</td>
<td>8.25 ± 0.02</td>
<td>4.61 ± 0.02</td>
</tr>
<tr>
<td>$\Delta pK_a$</td>
<td>0.24</td>
<td>0.28</td>
<td>0.22</td>
<td>0.26</td>
<td>0.28</td>
<td>0.24</td>
</tr>
<tr>
<td>$\Delta pK_a$ (Avg.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.25 ± 0.03</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.3: $pK_a$ determination for six pH indicators with two different sized PEGs at 40% (w/w) in 30 mM sodium cacodylate buffer. All values are reported as mean ± one standard deviation ($n = 3$).

Figure 4.2: Determination of $pK_a$ of Mod2 oligos by monitoring the CD signal at 298 nm under (A) dilute buffer conditions and (B) molecular crowded conditions (40% PEG-300). The $pK_a$ for these and the Mod1 and Mod3 oligos are recorded in Table 4.2.

4.4.2 Effects of PEG on Apparent pH

Given our (152), and other (38, 114), data on the effect of PEG on iM $pK_a$, we investigated the difference between the pH determined by electrode measurement and the activity of protons in the PEG solutions. We examined the absorbance changes of six different pH indicators in 30 mM sodium cacodylate buffer. We monitored the change in UV absorbance with change in pH for all indicators with and without 40% PEG-300. We interpret the change in
the pK\textsubscript{a} observed for these indicators as primarily due to the change in proton activity, since molecular crowding should not significantly affect the pK\textsubscript{a} of small molecules. As shown in Table 4.3, the observed pK\textsubscript{a} of indicators increases by +0.25 upon the addition of 40% PEG to the buffer, and is consistent for all indicators tested. We also tested the indicators with 40% PEG-3350 to see if the pK\textsubscript{a} is affected by the size of PEG. We observed no difference in pK\textsubscript{a} between PEG-300 and PEG-3350, suggesting the change in activity is linked to the percentage of PEG (w/w) in solution and not on PEG size.

4.4.3 Effects of PEG on iM pK\textsubscript{a}

The pK\textsubscript{a} of all oligos shows an increase upon adding PEG (Table 4.2); however, the increase in pK\textsubscript{a} is not different between oligos of different loop lengths or positions. The lack of a difference in pK\textsubscript{a} between these oligos suggests that the increase in pK\textsubscript{a} caused by molecular crowding conditions is not dependent on loop length. Based on our oligos, the increase in pK\textsubscript{a} is likely associated with the number of C-C\textsuperscript{+} bonds, which all these iMs have in common. In addition, there was no significant change in the cooperativity of the iMs with addition of PEG-300 (~ 3 in both dilute and crowded buffers). The increase in pK\textsubscript{a} with 40% PEG-300 (+0.40 ± 0.07) is larger than the change observe when testing proton activity (+0.25 ± 0.03; Table 4.3), suggesting that a change in proton activity alone does not account for the change of iM pK\textsubscript{a}. We suggest that the additional increase in pK\textsubscript{a} corresponds to the physical effects of molecular crowding, inhibiting the unfolding of the structure.

Next, we examined T1, Mod1T3, and Mod1T10 for any difference in pK\textsubscript{a} when using different size PEGs of 300 and 3350 molecular weight. Mod1 oligos were chosen, as they are symmetric with Mod3 oligos. Mod1 oligos have distinct differences in thermal stability and pK\textsubscript{a} among short (1 base), medium (3-5 bases), and long (10 bases) loops. The pK\textsubscript{a} increased by ~ 0.4
when adding PEG-300 and by ~ 0.7 when adding PEG-3350, showing that PEG size has a direct influence on the pKₐ of iMs. This increase is constant across all loop lengths tested; further suggesting that the pKₐ increase from crowding agents is not dependent on loop length. We can conclude the changes in pKₐ we observed here are from molecular crowding and not an additional change in proton activity since the change in proton activity (as determined from the indicators) is the same for PEG-300 and PEG-3350 (Table 4.3).

4.4.4 Thermal Denaturation of Folded i-Motifs

Thermal denaturations were performed on all oligos from Table 4.1. From the recorded spectra, the melting temperature (Tₘ) of each oligo was determined (Table 4.2; Figure 4.3). The trends observed for the Tₘ’s of Mod1 and Mod3 are similar as loop length increases; however, these Tₘ values are distinctly lower than Mod2 oligos of the same loop length. Mod2 oligos can incorporate loops much longer than Mod1 and Mod3. For example, the Mod2 oligos can have loops up to five bases long and not incur a change in the thermal stability. The trends in Tₘ found for these oligos suggest that longer central loops are more stable in iMs than in G4 oligos.

Figure 4.3: Thermal denaturations of Mod2 oligos in (A) dilute buffer and (B) molecular crowded conditions (40% PEG-300). Both Mod1 and Mod3 showed similar trends. The Tₘ data obtained is recorded in Table 4.2.
In the Mod1 and Mod3 oligos, the loop (Figure 4.1 B & D) may be sterically hindered by the other loop. On the other hand, the central loop (Figure 4.1C) should not be sterically hindered unless 5’ and 3’ overhangs are present. These steric hindrances help to explain the trends observed. All enthalpies of folding determined by Van’t Hoff analysis of $T_m$ data were approximately -5.0 kcal/mol per C-C$^\prime$ bond, which corresponds well with previously published melting enthalpies of iMs (180).

As a better indicator of biological viability, we calculated $\Delta G^\circ$ of folding for each oligo at physiological temperature, 37 °C ($\Delta G^\circ_{37 \, ^\circ C}$ in Table 4.2). Of all iMs examined, Mod2 iMs were most biologically viable, showing a negative $\Delta G^\circ_{37 \, ^\circ C}$ up to a loop length of twenty thymines (Mod2T20). The $\Delta G^\circ_{37 \, ^\circ C}$ for Mod2 oligos start increasing when the central loop is longer than five bases; however, the $\Delta G^\circ_{37 \, ^\circ C}$ values still suggest that long central loops could exist in vivo. In contrast, iMs formed by the Mod1 and Mod3 oligos become unfavorable at 37 °C when there are more than five thymines in the loop, suggesting long outer loops are not biologically viable.

Multiple studies (38, 114, 152) have shown that adding a molecular crowding agent increases the $T_m$ of select iMs. We studied the effects of molecular crowding on the $T_m$ and folding as a function of loop length (Table 4.2; Figure 4.3). The melting temperature of all oligos increases in the presence of PEG-300 as shown in Table 4.2; however, the magnitude of this increase in $T_m$ is dependent on which loop is elongated. For example, in PEG, when the loops of the Mod1 or Mod3 oligos are elongated, the $T_m$ increases by 6 – 8 °C, but if the loops of the Mod2 oligos are elongated, the $T_m$ increases by 9 – 12 °C. These different trends in PEG solutions suggest that longer central loops are more stable in molecular crowded conditions than their outer loop counterparts. Unlike the outer loops, the central loop is not sterically hindered by
nearby bases (Figure 4.1B-D), which allows for Mod2 oligos to fold into and retain the iM structure more readily than their Mod1 or Mod3 counterparts. The enthalpy of folding did not change significantly between the dilute buffer and 40% PEG-300 solutions. The $\Delta G^\circ_{37 \ ^\circ C}$ observed under crowding conditions shows a similar trend to the $T_m$ values, with $\Delta G^\circ_{37 \ ^\circ C}$ of the Mod1 and Mod3 oligos ranging from -2 to 3 kcal/mol and $\Delta G^\circ_{37 \ ^\circ C}$ of the Mod2 oligos ranging from -4 to -1 kcal/mol.

We used the oligos T1, Mod1T3, and Mod1T10 to further examine the changes in melting temperature as the size of PEG increased (Table 4.4). The thermal stability of all oligos showed an increase in $T_m$ in different PEG solutions until the PEG size was equal to or greater than a molecular weight of 600, which is in agreement with a previous study (114). At PEG sizes at or above 600, the $T_m$ remains constant. We surmise that the PEGs < 600 can readily fit into the space the iM needs to fold, preventing the iM from folding. However, the physical crowding of PEG outside of the space the iM folds still forces the iM together, thus the increase in $T_m$ is a combination of these two contrary effects. The PEGs ≥ 600 cannot fit into this space easily, thus does not hinder iM folding; thus, the increase in the $T_m$ comes purely from the physical crowding of PEG.

<table>
<thead>
<tr>
<th>Table 4.4: $T_m$ of three iMs across different sized PEGs at 40% (w/w) in 30 mM sodium cacodylate at pH 5.4. All values are reported as mean ± one standard deviation (n = 3).</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>No PEG</strong></td>
</tr>
<tr>
<td><strong>PEG-300</strong></td>
</tr>
<tr>
<td><strong>PEG-600</strong></td>
</tr>
<tr>
<td><strong>PEG-1500</strong></td>
</tr>
<tr>
<td><strong>PEG-3350</strong></td>
</tr>
</tbody>
</table>
4.5 Chapter Conclusion

We demonstrated that iMs with five or more thymidines in the central loop can form at higher pH than those with loops near the termini of the iM-forming regions. We also established that the increase in pK_a of iMs (+0.40) when PEG is added is independent of loop length and position. From our indicator studies, we found that a portion of this increase in pK_a upon adding PEG is from a change in proton activity, but crowding appears to contribute additional iM stability at higher pH. Our studies also demonstrated that the thermal stabilities of the iMs with three or more bases in the central loop were much higher (~15 °C) in dilute solutions at pH 5.4 than their outer loop counterparts when no 5’ or 3’ overhangs are present. When the iMs are in molecular crowded conditions, we found the outer loops’ T_m increase by 6 – 8 °C while the central loops’ T_m increase by 9 – 12 °C, showing that the increase in thermal stability observed is based on which loop is longer. Together, these trends show that not all loops can be considered equal in iM formation and indicate that iMs with long central loops have a higher melting temperature and pK_a than previously thought. This may allow for better control of the physical properties of iM-based nanomaterials as we demonstrated that small changes to the loop length greatly affect the physical properties of iMs. In addition, we also have shown that iMs with long central loops are potentially viable in vivo at physiological temperature, even with loop lengths of 20 bases. This finding suggests that the optimal search algorithm for locating iMs in genomic databases may be different from those algorithms used for G4 determination.
CHAPTER 5 : SUMMARY

In this dissertation, we explored the effects of various modifications of the iM structure. First, we found that tC° does not perturb iM formation or stability, even when substituted to replace a cytosine in the essential C-C° bond pairs. We also were the first to discover that tC° is much more sensitive to the formation of a tC°-C+ bond pair than a tC°-G bond pair, showing a very large decrease in fluorescence upon formation of an iM structure, allowing for the use of fluorescence to investigate iM formation. Using tC° substitutions in the iM structure, we investigated the hydrodynamics of the iM structure using analytical ultracentrifugation, fluorescence anisotropy, and molecular modeling, and determined solid evidence for a near spherical, compact, rigid iM structure. Next, we determined the mechanism of formation of an iM in solution, being one of the first groups to do so without the aid of NMR. Instead, we implemented more cost-effective techniques, such as fluorescence and circular dichroism. Then, we modified the loop length of all three loops of the iM structure, exploring the changes to formation and stability that each modification caused. We found that the central loop of the iM can accommodate longer loops than the loops near the termini of the iM-forming regions without loss of thermal stability or a lowering of pK_a. We established that the increase in thermal stability observed under crowded conditions was dependant on loop position, but not length, while the increase in pK_a of iMs (+0.40) observed was independent of loop position and length. Through thorough investigation, we became the first to determine that a portion of the increase in pK_a observed for an iM (+0.25 of +0.40) was derived from a change in proton activity while the remaining faction (+0.15 of +0.40) was from physical crowding effects.
Overall, this dissertation expanded the knowledge of iMs and the factors affecting their formation, providing vital information for the utilization of iMs in bio- and nano-material applications. In addition, we have furthered the understanding on iM formation under \textit{in vivo} conditions using PEG, demonstrating that molecular crowding does enhance the possibilities of iM formation \textit{in vivo} by increasing both thermal stability and pK$_a$ of formation. From here on, our laboratory will expand on this work by determining mechanisms for more biologically relevant iM structures, such as VEGF and the unmodified c-Myc; exploring the effects of different loop length modifications, such as adenine, thymine and guanine in the loops together instead of thymine alone; and investigating the effects 5` and 3` overhangs have on loop length dependence.
LIST OF REFERENCES


Figure A.1: Thermal melting of C20T at various concentrations, indicating that the iM formed is unimolecular as the $T_m$ does not change.

Figure A.2: $S_{20,w}$ vs pH of 3 µM iMotif with 10 mM KCl or 200 mM NaCl at 20°C determined from sedimentation coefficient ($S_{app}$) values using Sedanal. The three extra data points in high salt conditions at the lower pH values were included to demonstrate reproducibility at a higher concentration of 6 µM. The best fitted pK is 6.64 at low salt and 6.15 at high salt. Raising the salt concentration increases the $S_{20,w}$ values correcting for non-ideality due to the primary charge (or Donnan) effect. The average sedimentation coefficient (determined by joint Sedanal fitting) at low pH increases by 8.1% (1.827s vs 1.690s) and by 10.7% (1.421s vs 1.331s) at high pH, which is consistent with the decrease in total charge in the folded structure due to protonated cytosines.
Figure A.3: Thermal melting of C20T and all five tC° substituted oligos, indicating that the fluorescent base analog has no effect on the thermal stability of the iM.

Figure A.4: (A) CD spectra and (B) normalized fluorescence spectra of tC°4 (red) and the tC° incorporated into the non-iM forming strand (blue) at pH 5.8 (dashed lines) and pH 8.0 (solid lines), showing that the decrease in fluorescence is attributed to protonation of the cytosines and not to formation of the iM structure.
Table A.1: Fluorescence lifetimes of tC° oligos.

<table>
<thead>
<tr>
<th>Oligo</th>
<th>pH 5.4</th>
<th>pH 8.0</th>
<th>Duplex</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \tau_1 ) (ns)</td>
<td>( f_1 )</td>
<td>( \tau_2 ) (ns)</td>
</tr>
<tr>
<td>tC°4</td>
<td>4.5 ± 0.1</td>
<td>0.59 ± 0.01</td>
<td>1.7 ± 0.1</td>
</tr>
<tr>
<td>tC°9</td>
<td>4.8 ± 0.1</td>
<td>0.49 ± 0.01</td>
<td>1.4 ± 0.1</td>
</tr>
<tr>
<td>tC°10</td>
<td>4.9 ± 0.3</td>
<td>0.49 ± 0.06</td>
<td>1.8 ± 0.4</td>
</tr>
<tr>
<td>tC°12</td>
<td>4.9 ± 0.1</td>
<td>0.60 ± 0.08</td>
<td>1.3 ± 0.2</td>
</tr>
<tr>
<td>tC°15</td>
<td>4.7 ± 0.1</td>
<td>0.80 ± 0.01</td>
<td>1.2 ± 0.1</td>
</tr>
</tbody>
</table>

Table A.2: Fluorescence Rotational Correlation Times (\( \theta \)) of tC°-containing oligos at 25 °C in 150 mM potassium chloride.

<table>
<thead>
<tr>
<th>Oligo</th>
<th>pH 5.4</th>
<th>pH 8.0</th>
<th>Duplex</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \theta ) (ns)</td>
<td>( \theta ) (ns)</td>
<td>( \theta ) (ns)</td>
</tr>
<tr>
<td>tC°4</td>
<td>2.5 ± 0.2</td>
<td>3.1 ± 0.1</td>
<td>8.3 ± 0.3</td>
</tr>
<tr>
<td>tC°9</td>
<td>2.3 ± 0.1</td>
<td>3.2 ± 0.1</td>
<td>8.0 ± 0.1</td>
</tr>
<tr>
<td>tC°10</td>
<td>2.8 ± 0.1</td>
<td>3.3 ± 0.1</td>
<td>8.4 ± 0.1</td>
</tr>
<tr>
<td>tC°12</td>
<td>3.1 ± 0.1</td>
<td>3.5 ± 0.1</td>
<td>8.3 ± 0.3</td>
</tr>
<tr>
<td>tC°15</td>
<td>3.0 ± 0.1</td>
<td>3.4 ± 0.1</td>
<td>8.5 ± 0.1</td>
</tr>
</tbody>
</table>

Table A.3: Dichroism rate constants for C20T at various concentrations.

<table>
<thead>
<tr>
<th>Concentration of C20T (( \mu )M)</th>
<th>CD Rate Constant 1 ( (s^{-1}) )</th>
<th>CD Rate Constant 2 ( (s^{-1}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.61 ± 0.09</td>
<td>0.03 ± 0.02</td>
</tr>
<tr>
<td>8</td>
<td>0.69 ± 0.10</td>
<td>0.06 ± 0.04</td>
</tr>
<tr>
<td>12</td>
<td>0.72 ± 0.11</td>
<td>0.08 ± 0.04</td>
</tr>
</tbody>
</table>
VITA

Samantha Reilly
727B Clearbrook Dr. Oxford, MS 38655
(662) 259-2150 (251) 689-2715 (Mobile) smrb7321@gmail.com

Education

Ph.D., Chemistry, University of Mississippi, University, MS 2014
- Advisor: Randy M. Wadkins
- Dissertation Title: Spectroscopic Analyses of Factors Affecting the Formation and Stability of i-Motif DNA.

B.S., Chemistry, University of South Alabama, Mobile, AL 2009
- Advisor: Alexandra Stenson.
- Research Project: Beryllium (Be$^{+2}$) Complexation Avidity with Hydroxy-keto Heterocycles: an ESI-MS and Fourier Transform Ion-Cyclotron Resonance Mass Spectrometry (FTICR-MS) Investigation.

Experience

Research Assistant, University of Mississippi, University, MS 2014
- Used a variety of spectroscopic instrumentation to analyze fluorescent base analogs incorporated into i-motif DNA
- Collected and evaluated multiple thermal and steady-state spectra to determine the effects of loop modification to the physical properties of i-motif DNA
- Maintained and repaired spectroscopic instrumentation in the laboratory
- Mentored undergraduates in laboratory techniques and research project goals
- Collaborated with other scientists to perform experiments
- Published papers and presented research at conferences

Teaching Assistant, University of Mississippi, University, MS 2009 – 2013
- Created and kept chemical inventory stocked and up-to-date.
- Taught the theories and practical applications of quantitative analysis and instrumental analysis
- Led and mentored students to perform experiment and then write a professional laboratory report based on the JACS Communication Guidelines
- Substituted for corresponding lectures
Supplemental Instructor, University of South Alabama, Mobile, AL  2008 – 2009
  • Helped students understand the basics of analytical chemistry.
  • Reinforced the basics of quantitative chemical analysis.
  • Used and perfected basic communication, teaching, leadership, and organizational skills.

Laboratory Assistant, University of South Alabama, Mobile, AL  2008
  • Kept chemical inventory stocked and up-to-date.
  • Taught the theories and practical applications of physical chemistry

Published Papers


Oral Presentations


Reilly, S. M., Wadkins, R.M. 7th Mississippi Biophysical Consortium, Jackson, MS, June 5-6, 2014.


Reilly, S. M., Stenson, A. University of South Alabama Chemistry Department Seminar, Mobile, AL, April 2009.

**Awards and Accomplishments**

<table>
<thead>
<tr>
<th>Description</th>
<th>Year(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt; Place in University of Mississippi Graduate Student Poster Session</td>
<td>2014</td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt; Place in NSF Mississippi EPSCoR Graduate Poster Session</td>
<td>2014</td>
</tr>
<tr>
<td>Ole Miss Chapter ACS Teaching Assistant Award</td>
<td>2010</td>
</tr>
<tr>
<td>Graduate School Fellowship – Graduate School of the University of Mississippi</td>
<td>2009</td>
</tr>
<tr>
<td>Undergraduate Award in Analytical Chemistry – Awarded by ACS</td>
<td>2008</td>
</tr>
<tr>
<td>Chemistry Scholarship - Awarded by the University of South Alabama</td>
<td>2007</td>
</tr>
<tr>
<td>President’s List – University of South Alabama</td>
<td>2005 – 2009</td>
</tr>
<tr>
<td>Dean’s Honors List – University of South Alabama</td>
<td>2006 – 2008</td>
</tr>
</tbody>
</table>

**Skills and Techniques**: Fluorescence Lifetimes, Fluorescence Anisotropy, UV-Vis spectroscopy, Circular Dichroism (CD), Fluorescence-Detected Circular Dichroism (FDCD), Differential Scanning Calorimetry (DSC), Polyacrylamide Gel Electrophoresis (PAGE), Agarose Gel Electrophoresis, GROMACS (GROningen MAchine for Chemical Simulations), Autodock, Berkeley Madonna, Python, Xcalibur, MALDI-MS, ESI-MS, GC-MS, FTICR-MS