Secondary Structure Analysis of Intra- and Intermolecular Guanine Quadruplexes and Targeting through Complementary and Homologous Binding using Peptide Nucleic Acid

April Berlyoung

To my grandparents, Jerry & Esther Berlyoung and John & Mary Szakal.

Without their love and support none of this would have been possible.

Acknowledgments

There are countless people to thank for helping me through this journey. Here, I attempt to show a small amount of gratitude for the unwavering support shown to me.

This thesis would not be possible without my advisor, Dr. Bruce Armitage. Thank you for your support over the past five years. Thank you for meeting with me at 8:30am in February of 2014. That introductory meeting not only convinced me to enroll at CMU, but also of my perfect fit with the Armitage Lab. Your ability to mentor young scientists is inspiring. I would also like to thank my committee members: Dr. Marcel Bruchez, Dr. Danith Ly, and Dr. Rita Mihailescu. Dr. Bruchez and Dr. Ly provided helpful feedback and thorough discussions regarding my research progress report and my original proposal.

I must also thank Dr. Cheryl Telmer, who provided materials for the work in Chapter 2. Other professors at CMU who have assisted me in this journey include Dr. Gloria Silva, Dr. Susan Graul, and Carolyn Neiderlander. They made TA assignments feel less like work and more like an enjoyable, collaborative environment. I am appreciative of their mentorship and support.

Former Armitage Lab members who were instrumental in my training include Dr. Lisa Rastede, Dr. Stanley Oyaghire, Dr. Munira Fouz, Dr. Karen Kormuth, and Dr. Taylor Canady. These colleagues and friends welcomed me into the lab and graciously taught me the ways of the group. Rotation in a lab is critical, and the Armitage lab has gained three new students during my tenure. There's never a dull conversation with Dima Kolodieznyi. Thank you for always teaching me something, and for your dedication to Mellon FIT. Thank you Nisha Hasija for helping me with patience and understanding. Thank you Srijani Sarkar for working through the many frustrations of SPR and PNA synthesis with me. I also had the pleasure of working with several undergraduate researchers; Gunnar Gillespie, Justin Wei, Sarah Simon, and Melissa Greco were

all pleasures to work with. I am grateful for the time I was able to spend with each of them, and the ability to watch their skills grow exponentially over the course of the semester/year.

I have been privileged to work with a variety of organizations while at CMU. The Graduate Student Association (GSA) opened my eyes to the number of students dedicated to improving graduate student life through a student-run governing body. The executive board continues to impress me every month with the changes they are able to implement. Mellon FIT has also been an amazing part of my life at CMU. I am so thankful to be a part of an organization that cares so deeply about the wellbeing of graduate students in the Mellon Institute and beyond. Thank you Dr. Olivia Molinar for your dedication and initiating the club that would become Mellon FIT.

I have to thank the unsung heroes of the Mellon Institute: Sara Wainer, Lorna Williams, Valerie Bridge, and Brenda Chambers. These people are essential to the function of the department and have assisted me with a variety of activities during my time at CMU.

I would not be here without the friends I made along the way. Dr. Anne Arnold, thank you for your unwavering belief in me. Thank you for welcoming me into your home and allowing me to spend time with Izzy. Zoe Wright, thank you for your patience, encouragement, and writing help. Thank you both for being there every step of this Ph.D. I know you will both do great things.

My family has been my largest support system throughout my graduate career. My grandparents, Jerry and Esther Berlyoung, are my biggest supporters. A conversation with Esther is not complete unless she tells me how proud of me she is. Jerry & Esther have lit my path through some of my most difficult times in graduate school. Having them two hours away allowed me to spend weekends with them, which has been amazing since I lived nearly 700 miles away for the majority of my younger life. Jerry has taught and continues to teach me patience. I am very grateful to have known all four of my grandparents. John Szakal, my maternal grandfather, passed away in

Table of Contents

List of Figures

List of Schemes

List of Tables

List of Supporting Figures

Abstract

Many of the projects discussed in this thesis utilize one or more of the binding modes shown in Figure 1.12. As a result of this work, we expand our understanding of GQ secondary structures and show the ability to target these GQs using multiple binding modes.

Chapter 2 analyzes variations on a short peptide nucleic acid (PNA) probe that homologously binds to guanine-rich telomeric DNA. We compare the PNA probe with and without backbone modifications and utilize a GQ binding small molecule fluorophore; a combination of two binding modes. This work identifies that TO-labeled probes show significantly increased thermal stability compared to their unlabeled counterparts. Modifications inducing a right handed helix in the PNA show no difference in thermal stability. This short probe does not possess selectivity as evidenced by in vitro translation experiments performed in cell lysate.

In Chapter 3, we expand the current literature on RNA:DNA heteroquadruplexes (RDQ). This GQ contains both RNA and DNA guanine tracts and has recently been identified as a regulator between transcription and translation. We show the formation and stabilities of three suspected RDQ sequences using an optimized scaffold duplex. The formation of the RDQs was also monitored with a small molecule fluorophore, Thioflavin T.

Chapter 4 utilizes the RDQ systems created in Chapter 3 and compares them to DNA and RNA homoquadruplexes. With this chapter, we make DNA:DNA and RNA:RNA homoquadruplexes using the scaffold duplex created in chapter 3 and compare their biophysical properties to our RDQs. A modified PNA oligomer is then used to duplex invade each heteroquadruplex. We identify that RDQs possess characteristics of both RNA and DNA heteroquadruplexes. Our PNA probe is also able to sufficiently target each heteroquadruplex with varying IC_{50} values.

Chapter 5 describes the use of dual gammaPNA (γPNA) probes that hybridize to alternating sites along a telomere. Each probe contains a fluorophore, and can undergo Förster resonance energy transfer (FRET). We study the biophysical characteristics of these probes and observe bright staining of telomeres in

nuclei. Finally, we show the incorporation of additional fluorophores does not significantly improve

FRET efficiency

1 Chapter 1. Introduction to DNA and RNA G-Quadruplexes: Targeting through Complementary and Homologous Binding

1.1 Nucleic Acid Secondary Structures

A variety of secondary structures have been discovered since the identification of the DNA double helix (B-DNA)[1]. Some variations reside within the helix family, such as A-DNA and Z-DNA[2,3]. Both use Watson-Crick base pairing (adenine to thymine and guanine to cytosine), but differ in overall structure. For example, A-DNA is broader and has more base pairs per helical turn, while Z-DNA runs left handed instead of right handed (Figure 1.1). A variety of structures outside of the double helix have also been identified and have shown to impact biological systems.
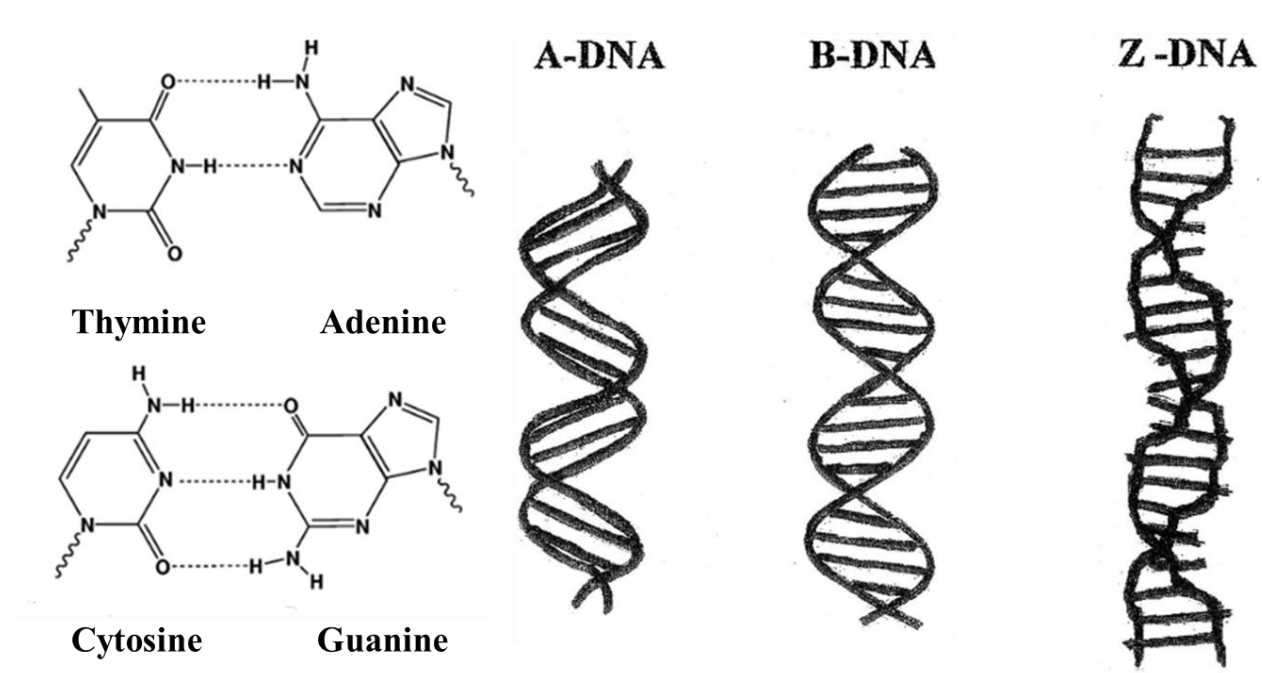
Figure 1.1 Watson-Crick base pairing of cytosine-guanine and thymine-adenine (left) and variations of the double helix, A- B- and Z- DNA (right). Adapted from Reference 4.

1.1.1 Non Canonical Nucleic Acid Structures

Other nucleic acid secondary structures exhibit hydrogen bonds outside of the typical Watson-Crick. One such form is triplex DNA, where a purine-rich duplex and a third nucleic acid strand bind by Hoogsteen hydrogen bonding[5–7]. Triplex-forming DNA has been implicated in regulating DNA metabolism and gene function[6]. Advancements in the study of triplex-forming DNA have led to the development of nucleic acid therapeutics and improvements in DNA nanotechnology[6,8].

The final two nucleic acid secondary structures are intercalated motifs (i-motifs) and guanine quadruplexes (GQs). These structures are often associated, since i-motifs arise from cytosine-rich sequences and GQs arise from guanine-rich sequences. When an i-motif forms in duplex DNA, the complementary guanine bases are unbound and can form a GQ, and vice versa. i-motifs consist of two intercalated duplexes held by hydrogen bonds of hemi-protonated cytosine (shown as green bars in Figure 1.2)[9]. This protonated cytosine is favored by slightly acidic conditions for stabilization; however, stable i-motifs have been observed at neutral pH under certain environmental conditions, such as in the presence of molecular crowding and negative superhelicity[9]. Bioinformatics indicate i-motifs are located at key points of interest in the genome in addition to the growing body of *in vitro* assays implicating i-motifs as gene expression regulators and telomerase regulation[9,10]. GQs form in guanine-rich sequences and utilize both the Watson-Crick and Hoogsteen faces of the guanine in order to form tetrads (shown as red squares in Figure 1.2)[11]. High throughput sequencing experiments suggest there are over 700,000 potential GQ forming sequences[12]. Furthermore, GQs are particularly interesting because of their potential involvement in DNA replication, telomere maintenance, and gene expression[13,14].
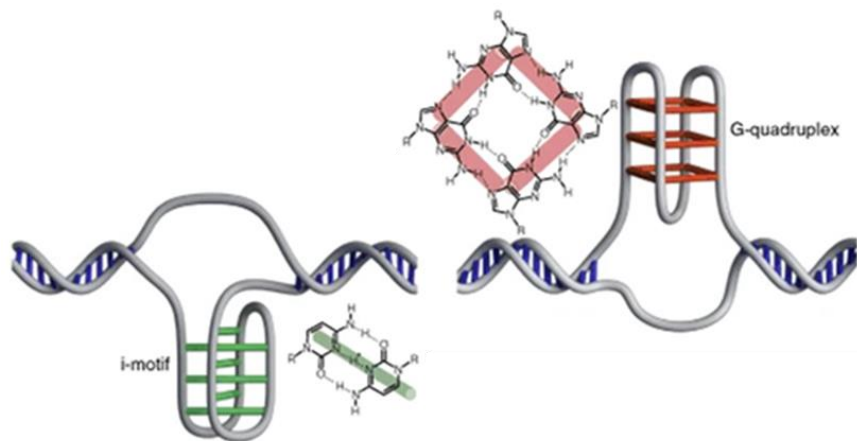
Figure 1.2 i-motif and guanine quadruplex structures. Adapted from Reference 15.

My thesis projects focus on the understanding of GQ secondary structures and targeting these sequences with the nucleic acid analogue peptide nucleic acid (PNA). The remainder of this introduction provides a background for GQ topology, biological implications, and targeting strategies.

1.1.2 Guanine Quadruplex Structures: Sequence Requirements, Topology, and Stability

Guanine rich sequences can form GQs by assembling into tetrads, where four guanine bases interact by hydrogen bonds from both Watson-Crick and Hoogsteen faces[16]. This allows for eight total hydrogen bonds as well as electrostatic stabilization between O6 and a typically monovalent cation (shown as X+ in Figure 1.3)[16]. Tetrads then align atop one another by π-π stacking. Typically, GQs form within the following sequences: $G_AX_BG_AX_BG_AX_BG_AX_B$[16]. Initial GQ literature identified sequences where A was 2-3 guanines and B was between 1 and 7 nucleotides. More recently, studies show the formation of GQs where B is much greater than 7 and in some cases up to 70 nucleotides[17,18].

Figure 1.3 Structure of a guanine tetrad.

Additionally, bulges within guanine tracts, up to five nucleotides, show

3

stable GQ formation[19]. This allows for GQ sequences to vary greatly in their primary sequence, while maintaining the overall integrity of the GQ.

Sequences that allow for three or four tetrads typically form more stable GQs than those with two tetrads, although all three versions have been formed *in vitro*. An additional sequence component affecting GQ stability is the loop sequence, or the nucleic acid between each G-tract of the quadruplex. The spatial location of the loop varies with the topology of the GQ, which can impact the affinity of GQ binding ligands. Loop sequence has also been shown to play a role in the topology and folding of GQs[20]. Intramolecular GQs, where the GQ consists of only one nucleic acid strand, are typically more stable because the forced proximity of a single strand allows easier association of the GQ. This is in contrast to intermolecular GQs (i.e. consisting of two or more nucleic acid strands). Intermolecular GQs show a broad range of thermal stabilities but typically possess slower association kinetics since two or more molecules need to gather[21–23]. This kinetics issue can be circumvented if the nucleic acids are already in close proximity, such as multiple RNA transcripts immediately after transcription.

Both inter- and intramolecular GQs can orient themselves in three different ways. First, when all guanine tracts are in the same 5' - 3' direction, the GQ is in a *parallel* conformation[16]. Second is *antiparallel*, where half the tracts run 5' - 3' and the other half run 3' - 5'[16]. The third *hybrid* structure is a combination of parallel and antiparallel strand orientation[16]. Strand orientation impacts the glycosidic bond and whether the base resides in the *syn* or *anti* configuration while in the GQ. GQs, which can accommodate both *syn* and *anti* configurations, have the potential to possess different loop structures and could be useful in specific GQ recognition. Additionally, DNA GQs can transition between topologies upon a change in cation concentration and identity[24,25]. Due to the 2'-OH in the ribose sugar, RNA guanines involved in GQ formation assume the *anti* position and are therefore parallel GQs[26]. The existence of an antiparallel RNA has only occurred by force[27,28].
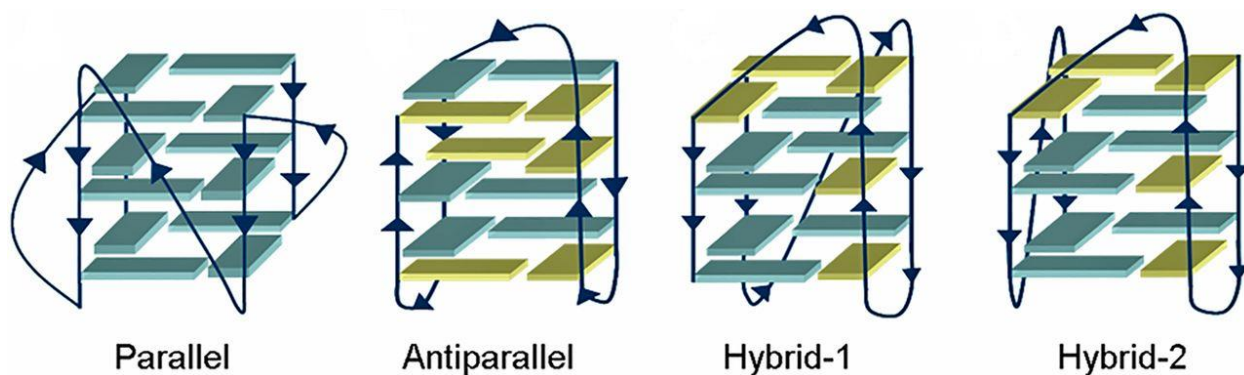
Figure 1.4 Schematic representation of different GQ topologies. Guanines with *syn* and *anti* glycosidic bonds are shown in yellow and cyan. Adapted from Reference 29.

A variety of factors impact the stability of both DNA and RNA GQs. The monovalent cation stabilizing O6 plays a key role, as its size and charge density are crucial for GQ stability. Potassium (K^+) is one of the best GQ stabilizing ions, followed by sodium (Na^+)[30]. With an ionic radius of 1.33Å, K^+ is too large to fit within the tetrad[31]. Therefore, K^+ coordinates between two tetrads. Sodium is small enough (ionic radius 0.95Å) to coordinate within the tetrad itself[31]. Additionally, both have relatively low hydration energies. Hydration energy must be overcome to displace water molecules from the ions so that they are available to stabilize the GQ[30]. K^+ and Na^+ possess relatively small hydration energies, with K^+ having the smaller of the two, and therefore the cost of ion-dehydration does not prevent GQ formation. Furthermore, lithium ions (Li^+) do not support the formation of GQs. The lack of coordination of Li^+ is due to a small ionic radius (0.76Å) and a significantly higher hydration energy[31,32]. Throughout this thesis, K^+ and Li^+ buffers will be used to compare the formation of GQs.

1.2   Impact of GQ Structures on Biological Functions

While much knowledge has been collected about the overall structure and formation of GQs, it is also critically important to identify biologically relevant GQs. These structures exist in both DNA and RNA sequences, and they must be understood at a biophysical level as well as a biochemical level. Once these properties are understood, the knowledge can be used to impart some change in the biological function by targeting the GQ.

5

1.2.1 Cellular Evidence for GQs

Whole genome sequencing has allowed in-depth computational studies on the potential prevalence of GQs. Over 700,000 potential GQ sequences have been identified using updated GQ models by incorporating longer loop lengths, as well as observing possible bulges within guanine tracts[12]. However, the presence of a potential GQ sequence does not prove a GQ will form at that particular location. This challenge has been addressed by two independent research studies. One of the most notable DNA GQ sequences, identified in both sequence and cellular analyses, is telomeric DNA. A second common location of DNA GQs is near transcription start sites, especially surrounding the promoter region of genes.

Cellular evidence for GQs in the human genome was sought after the completion of sequence analysis. Two independent laboratories, Balasubramanian and Lansdorp, developed different GQ-specific antibodies which were used alongside secondary and fluorophore-labeled tertiary antibodies to label GQs in fixed cell nuclei[33,34]. These experiments provided two very interesting results. First, they note the antibodies localized on previously identified GQs, such as the chromosomal ends called telomeres and other unidentified sequences throughout the chromosome[33]. Second, there is a cell cycle dependence on GQ presence, with the greatest number of GQ sites observed during the S phase[33]. This is thought to be due to the unwinding of DNA during replication, which would promote the formation of GQs. However, one must also consider the interaction between antibody and GQ. GQs are often stabilized by binding partners, and these results could be skewed because of any stabilization effects provided by the antibodies. There is still ample research which supports the regulatory power of GQs. Now that GQs have been established, in depth studies can be performed on specific regions to better understand their function.

1.2.2 Regulatory Roles of DNA GQs

One of the first GQs identified in cells is located in telomeric DNA. This 6mer repeating sequence (TTAGGG in vertebrates) caps each chromosome with thousands of repeats and slowly degrades over the

course of an individual cell's lifetime[35]. The discovery of how these telomeres provide genome stability and the enzyme that elongates telomeres was awarded the Nobel Prize in 2009[36]. Telomeres consist of both double stranded DNA as well as a 5' overhang which decreases in length after each replication. This is known as the end-replication problem. One method the cell has to alleviate this problem is to use telomerase, an enzyme capable of extended telomere length that is typically inactive in somatic cells. However, telomeres are used as a regulatory mechanism for cellular age, and telomeres which are too long are considered a cancer hallmark. GQs can form on this 5' overhang with four or more TTAGGG repeats and have been shown to decrease telomerase activity. However, there are a variety of binding

Figure 1.5. Regulatory roles of DNA GQs in transcription.

factors and proteins that play a role in telomere maintenance and telomerase regulation[37]. Research is still needed to fully understand the role of telomeric DNA GQs within telomere biology.

High-throughput sequencing studies noted that DNA GQs often occur in the promoter region of a particular gene[38]. With such a prevalence of GQs at promoters across different species, it was speculated that these sequences had some sort of biological significance. Well-studied promoters containing GQs include c-MYC[39], RAS(h and k)[40], and VEGF[41]. kRAS has been shown to have at least three GQs by sequence, identified in 2016 by the Brooks lab. Mutations in kRAS have been identified in 98% of pancreatic adenocarcinomas and the kRAS protein has been notoriously difficult to selectively target, given it has two other isoforms[42]. Therefore, studying and developing therapeutics to target the kRAS gene and its promoter has been at the forefront of the field. Brooks and colleagues were able to show two of the three GQs had strong CD melting profiles and DMS footprinting[40]. The group also found that a GQ binding small molecule (TMPyP4), which bound to both kRAS GQs, knocked down luciferase expression in a dual luciferase-renilla expression assay[40]. Individual controls with only one of the two GQs present showed that

7

both GQs are involved in expression suppression, however, the longer more G-rich mid sequence showed slightly better suppression[40]. The cytosine-rich strand in kRAS has also been shown to form secondary structures and can modulate between i-motif and hairpin structures[43]. All three i-motifs negatively regulate transcription when stabilized by small molecule binders. Targeting both guanine and cytosine rich sequences might allow enhanced transcription regulation in future experiments. However, small molecules are not the only kRAS binding ligands with regulatory roles. Recently, HMGB1 (a protein involved in a plethora of functions including cell growth, immunity, and DNA chaperoning) was identified to bind one of the previously identified kRAS GQs[44]. Binding of HMGB1 to kRAS was shown by surface plasmon resonance (SPR) and fluorescence assays[44]. mRNA expression assays where HMGB1 siRNA was transfected into pancreatic cancer cells showed an increase in kRAS expression[44]. The interaction between HMGB1 and kRAS remains unknown, as well as how selective HMGB1 is to kRAS. These questions will need to be addressed in order to utilize kRAS DNA as a therapeutic target.

1.2.3   Regulatory Roles of RNA GQs

Biologically relevant GQs are not solely restricted to the genome, as RNA GQs can result from RNA transcripts. RNA GQs have shown significant regulatory capacity and are considered a strong target for regulating protein expression and controlling protein concentrations within cells where full elimination of a protein is not necessary[45]. There are a few distinct differences between DNA and RNA GQs, and they are important to understanding GQ dynamics. First, since RNA is typically single stranded, the formation of a GQ is more favorable. Single stranded RNA GQs do not have to compete with a duplex strand like genomic DNA GQs[46]. Second, RNA GQs always possess parallel folding due to the 2' hydroxyl. This –OH group allows for more intramolecular hydrogen bonds to form, which decreases the number of bound water molecules and therefore decreases the entropic penalty during GQ folding[26]. For these reasons, RNA GQs typically have higher thermal stabilities than DNA GQs, which can easily be monitored by UV melting.

Telomeric repeats transcribed to RNA (TERRA), are involved in telomere extension and are critical for telomerase function[47]. Telomerase, the enzyme responsible for telomere extension, is a ribonucleoprotein that requires a TERRA sequence in order to function[48]. For this reason, TERRA molecules are found in high concentrations near telomeres and have been shown to have higher concentrations in cancer cells where telomerase is typically overactive[48]. In systems lacking telomerase, TERRA molecules can assist in telomere length regulation in telomerase-negative ALT systems[49]. The formation of RNA GQs formed by TERRA sequences has been well studied *in vitro*, however, the interaction between TERRA GQs and the wide array of interactions with telomerase, as well as telomeric DNA, is a topic in need of further research.

mRNA transcripts, and more specifically the 5′ untranslated region (UTR), show a strong localization of GQ forming sequences across species. The presence of these possible GQ sequences concentrated around these regions indicates a possible evolutionary conservation and regulatory role[45]. RNA GQs have been shown to regulate translation by preventing the 5′ cap from binding to translation initiation factors, referred to as cap-dependent translation[46]. RNA GQs were subsequently shown to inhibit translation by cap-independent translation, where GQ folding modifies the internal ribosome entry sites (IRES) and impedes the ribosome from binding[46]. These findings are promising for researchers looking to use RNA GQs as therapeutic targets. Our lab has participated in this subject, utilizing peptide nucleic acid to bind and prevent RNA GQ formation[50]. By doing so, we have shown transcription inhibition in *in vitro* luciferase assays[50]. But, some research suggests targeting 5′ UTR GQs leads to an enhancement in translation[51]. This discrepancy between different sequences is not well understood and needs to be further studied in order to effectively target with the preferred outcome (either enhancement or inhibition of translation).

1.2.4 Existence and Implications of RNA:DNA Heteroquadruplexes

One area of GQ research that has not been studied as thoroughly as DNA and RNA GQs is the formation of RNA:DNA Heteroquadruplexes (RDQs). RDQs can occur when guanine-rich DNA and RNA are in close proximity and a GQ forms with both DNA and RNA contributing to the GQ tetrads. Although DNA and RNA typically exist in different locations within a cell (DNA in the nucleus and RNA in the cytosol), there are important instances in which they have the potential to interact. The two following examples illustrate these instances and the potential functional roles.

One example of this interaction is between telomeric DNA and TERRA. TERRA has been speculated to Watson-Crick bind with the C-rich strand of telomeric DNA, allowing a DNA GQ to form

and impact how telomeric binding proteins, such as the shelterin complex, interact with telomeres[48]. This RNA:DNA duplex, called an R-loop, is thought to strongly impact the telomere length and homeostasis in cancer cells. However, with guanine-rich DNA and RNA in close proximity, one could also imagine the formation of an RDQ.
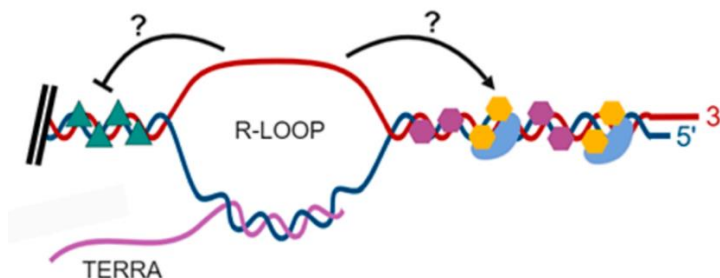
Figure 1.6. Formation of RNA-DNA hybrid duplexes (R-loops) at telomeres. R-loops are thought to play a role in DNA methylation (left) and protein recruitment (right). Adapted from Reference 52.

Telomeric RDQs have been studied *in vitro*, utilizing oligonucleotide models that create a light up probe when bound. The DNA was labeled with an azidocoumarin moiety and RNA was labeled with an alkyne moiety[53]. This system underwent a click reaction when the two strands are in close proximity, in the presence of Cu catalyst, and form a fluorophore tag[53]. These oligonucleotides were then inserted into HeLa cells to observe RDQ biocompatibility in cells[53]. These results showed localization of the RDQs in the nucleus and indicated RDQs could form *in vivo*[53]. Though these results are promising, the development of an RDQ specific assay will be needed to identify their formation in cells and confirm their genomic locations.

A second interaction between DNA and RNA where a heteroquadruplex can form is during transcription. Under the condition where the DNA template strand is cytosine-rich, the RNA transcript and DNA non-template strand will both be guanine-rich. Therefore, recently transcribed RNA has the potential to interact with the unwound non-template DNA and can form an RDQ. The most promising sequence to date is at the conserved sequence block II (CSB II) region in human mitochondrial DNA. Previous literature indicated that this region was responsible for transcription termination and
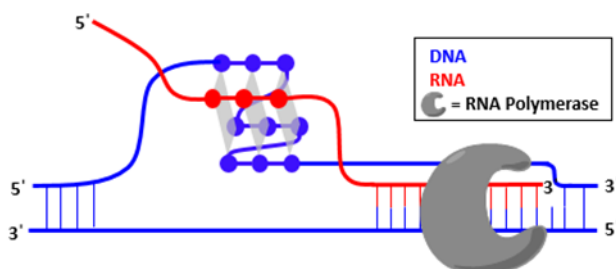
Figure 1.7. Formation of RNA:DNA Heteroquadruplexes (RDQs) during transcription.

was thought to serve as a pseudo-switch between DNA replication and transcription[54]. The idea of an RDQ serving this purpose was introduced in the late 1990s, however, it was thought that the RNA in the RDQ served as the transcription primer and was bound to the replication origin H[54]. This work has significantly expanded since 2010, when the Gustafsson lab reported the formation of an RDQ at the mitochondrial CSB II region and demonstrated that the R-loop formed during transcription was extended and more stable[55]. Since this work, the Tan laboratory has been the primary lab studying CSB II RDQs. They were able to replicate previous literature and confirm the formation of an RDQ[56,57]. They were also able to establish the sequence of events that led to the formation of the CSB II RDQ[58]. *In vitro* transcription assays were conducted using a T7 promoter with an inserted CSB II sequence[58]. A GQ binding ligand (Zn-TTAPc) selectively cleaves the adjacent guanine bases upon radiation with ultraviolet light and was visible in polyacrylamide gel in order to identify existing GQs and RDQs[58]. Using this method, as well as DMS footprinting and control treatments with RNase and DNase, it was established that the transcribed RNA remains duplex-bound to the DNA, forming an R-loop[58]. During the next round of transcription, the bound RNA is displaced by the polymerase and the newly formed transcript[58]. The displaced RNA is then free to form an RDQ with the non-template strand. In two different experiments, the formation of the RDQ was directly related to the formation of the R-loop[58 57]. Since these RDQs are dependent on transcription events,

one might assume they are relatively short lived. However, this proved to not be the case using a FRET-based transcription model[59]. This model utilized DNA where the guanine-rich motif is flanked by a donor and acceptor dye. This system was used with sequences that could form RDQ, DNA GQ, and no GQ. The formation of both RDQs and DNA GQs was fast (over 50% quadruplex formation within minutes) and the rate of collapse was slow when left undisturbed for at least an hour after transcription stopped[59]. Upon the introduction of RNase, the RDQ was disrupted by approximately 67%, however, the remaining 33% was still persistent throughout the one hour test[59]. The persistence of the RDQs and DNA GQs has the potential to significantly affect genome transcription, however, further sequence analyses are needed to identify specific locations for future study.

Given the likelihood that RDQs play a regulatory role in transcription, it is critical to identify genomic regions which can form RDQs, as well as identify if these sequences are evolutionarily persistent. Protein coding genes and their upstream flanking regions, including the TSS, were studied through sequence analysis for a statistical prediction of whether RDQ formation was possible[60]. Of the 60 species tested, RDQs near the TSS appeared in higher order animals (i.e. Amphibia, birds, and mammals)[60]. The use of the non-template strand to form the RDQ is specific to mammals and persists throughout the mammalian class[60]. However, this is not to say that bacteria and other eukaryotes cannot be affected by RDQ. Plasmid transfection into E. coli yielded similar results to the previous *in vitro* experiments, indicating these RDQs can form under *in vivo* conditions[61]. These statistical analyses provide a starting point for scientists to identify pertinent sequences for further analyses, with a focus on therapeutic targeting.

These examples of RDQ formation are particularly exciting for scientists looking to use RDQ as therapeutic targets. Since RDQs are comprised of two nucleic acid strands, there are two unique sequences that can be targeted. Complementary probes can be used to target either the DNA or the RNA, selectively. However, because RDQs are a relatively understudied area of GQ literature, much research is needed at both a biophysical and biochemical level. Biophysically, research needs to address the formation and stability of RDQs. With that information, scientists can begin to research specific locations where predicted RDQs form in order to better understand their biological role and target ability.

1.3 Strategies for Targeting GQs

Due to the biological relevance of GQs, scientists aim to manipulate GQ formation by designing probes. There are a variety of factors to consider when creating probes. The following section discusses some of the options in probe design.

1.3.1 A Delicate Balance between Selectivity and Affinity

The terms selectivity and affinity are consistently discussed when targeting GQs. Selectivity (or specificity) describes how effectively a probe binds to its target over off-target binding. In the case of a GQ target, the probe should only bind the structure of interest. Affinity refers to the strength of the binding interaction between target and probe. In an ideal scenario, a probe would bind to its target with high selectivity and affinity. This can be difficult, since selectivity often decreases as binding affinity increases. Targeting an individual GQ can be difficult due to similarities in overall sequence (high G content) and secondary structure compared to off-target GQs. Nonetheless, scientists have developed GQ binding ligands based on two different recognition modes: shape and sequence.

Shape based recognition of GQ sequences utilizes two primary interactions: π-π stacking and stabilizing electrostatic interactions. In order to increase affinity, one could increase the overall number of these interactions in order to stabilize the ligand. Typical small molecule shape binders contain planar regions that interact with the planar top or bottom tetrad. While this may lead to a GQ-selective ligand with high affinity, these ligands are not selective to a specific GQ. Small molecule ligands that are selective to a particular GQ often contain modifications that expand the stacking interactions beyond the tetrad, such as groove binding.

Sequence-based recognition of GQ sequences uses Watson-Crick or Hoogsteen base pairing between the nucleobases of the GQ target and oligomer probe. In this recognition method, the length and sequence are critically important to the probes' affinity and selectivity. Probes must contain a minimum number of bases in order to possess sufficient binding affinity. Additionally, short probes used in cellular

or plasmid studies run the risk of having multiple perfect match sequences in a genome. In contrast, long probes can permit mismatches in the target sequence, leading to off-target effects. This leads researchers to design probe modifications and use multiple sequence-based binding modes to modulate affinity and selectivity.

The following two sections will explore shape and sequence recognition modes, discussing small molecule and nucleic acid analogues.

1.3.2 Small Molecules

A variety of small molecules have been shown to bind and stabilize GQs. Many possess fluorescence properties, which allow for GQ detection. Small molecule GQ binders include: cyanine dye derivatives, thioflavin t, porphyrins, and a few metal complexes; all of which bind with affinity in the micromolar to nanomolar range[62,63]. Often, these fluorophores are light-up probes, meaning they only fluoresce upon binding. One challenge to the field has been identifying ligands that are selective only to GQs and not to other nucleic acid secondary structures. More recently, scientists have identified ligands that are selective to individual GQ sequences and topologies. Thioflavin T (ThT) was identified as a GQ-specific fluorophore and is widely used as a sensor for GQ formation[64,65]. ThT-E contains an ethyl modification and has improved affinity over ThT, but remains selective to the generic GQ structure[66].
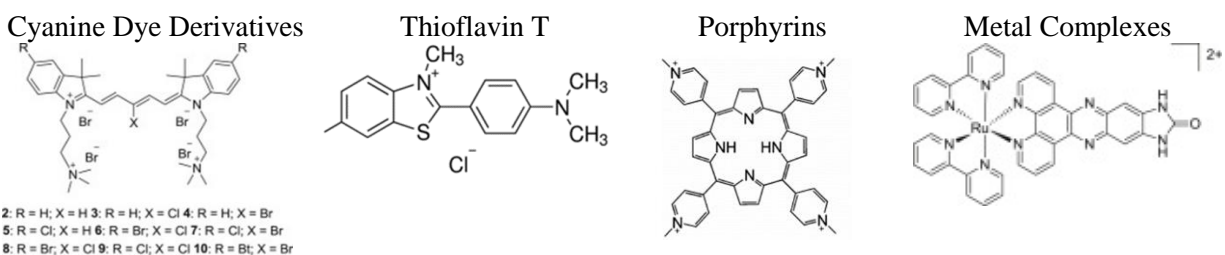
Figure 1.8. Fluorescent GQ binding ligands. Selected from References 61 and 62.

Non-fluorescent small molecules have also been used to target GQs. A small molecule microarray technique developed by the Schneekloth Lab has identified chlorohexidine to be specific to kRAS GQs[67].

The use of a small molecule microarray allows for the screening of a wide variety of compounds, and works well for identifying potential binders from pre-existing compounds. An alternative approach from the Wei Lab utilized molecular modeling and structure-based screening to create a compound (9CI) selective to the c-MYC Pu22 GQ[68]. This approach allows researchers to design a probe with a specific target in mind. It also requires a strong understanding of the secondary structure of the GQ of interest.

Chlorohexidine   9CI
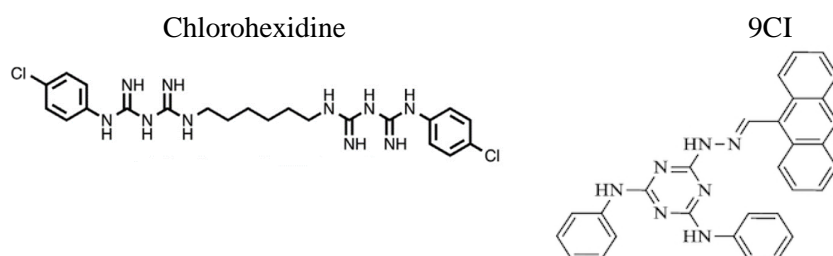
Figure 1.9. Nonfluorescent GQ binding ligands. Selected from References 66 and 67.

Structural and molecular modeling studies suggest small molecules primarily bind GQs by end-stacking via π-π interactions to the terminal tetrads[69]. This is reasonable considering most GQ ligands are planar molecules. However, there are examples of GQ ligands which bind the groove of the GQ[69,70]. Both the π-π interactions and groove binding moieties have the potential to impact the overall stability of the GQ. For this reason, it is critical for researchers to understand the impact of their small molecules on GQ stability and secondary structure.

1.3.3   Nucleic Acid Analogues

The use of nucleic acid molecules and analogues as binding probes provides an alternate binding mode to the previously discussed small molecules. While small molecules require π-π stacking and hydrophobic interactions, nucleic acid molecules utilize hydrogen bonding and π-π stacking between base pairs. Various nucleic acid analogues exist outside of the natural DNA and RNA; these include 2'O-Me

RNA, locked nucleic acid (LNA), and peptide nucleic acid (PNA) shown in Figure 1.10. These structures differ in their backbones, leaving the nucleobase unchanged.
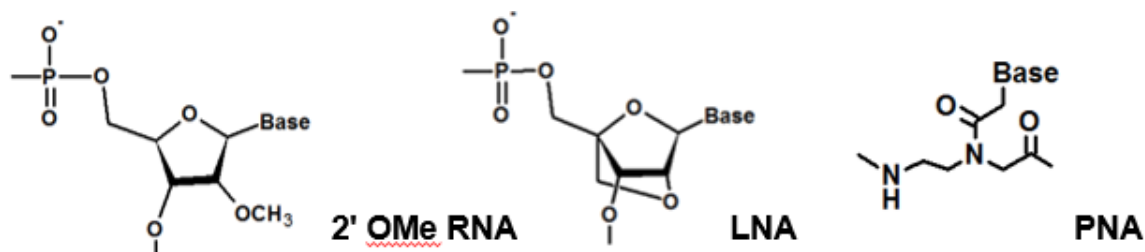
Figure 1.10 Nucleic acid analogs including 2'-OMe RNA, locked nucleic acid (LNA), and peptide nucleic acid (PNA)

Scientists can design probes to bind GQs in two distinct modes: complementary or homologous binding. 2'OMe RNA and LNA have primarily been used as complementary binders. PNA is uniquely capable of both, and can be optimized by modifications.

1.3.3.1 Complementary Binders

Complementary binding occurs when two sequences hybridize using Watson-Crick hydrogen bonding. In order to design a complementary probe, knowledge regarding the sequence of interest is required. The resulting secondary structure, upon binding, will be a hybrid duplex between the target and the probe. Therefore, this approach is typically used when disruption of the GQ is favored[50].

A variety of nucleic acid analogues have been used in this manner. 2'-OMe modified oligomers were used to target the 5' untranslated region (UTR) of a histone gene involved in several cancers[71]. Utilization of this probe was promising, however, if the GQ was prefolded the oligomer was unable to invade and disrupt formation[71]. A second analogue uses locked nucleic acid (LNA). LNA is a modified RNA nucleotide where the ribose contains an extra carbon connecting the 2' oxygen and 4' carbon, and restricts the ribose to the endo conformation (Figure 1.10). This modification was able to significantly improve binding affinity and further decrease gene expression compared to probes with fewer or no

modifications[72]. Peptide nucleic acid (PNA) contains N-(2-aminoethyl)glycine subunits instead of the sugar phosphate backbone[73]. The neutral PNA backbone has distinct properties including resistance to degradation by nucleases and proteases, as well as high binding affinity[74,75]. Our lab has extensively used PNA, and a modified γPNA, to target GQ structures. The resulting DNA-PNA and RNA-PNA hybrid duplexes have high stability and are formed with relatively short probes[76]. Using complementary binding, we can target the loop regions of the GQ, which leads to increased selectivity[77].

1.3.3.2   Homologous Binders

Guanine-rich oligomers are able to bind GQs homologously and thus substitute guanines in the GQ. Homologous binding allows invasion of the GQ by the probe, but maintains the presence of a heteroquadruplex. This can be beneficial for systems where the presence of a GQ is desired, and in many cases the heteroquadruplex has higher stability than the native GQ.
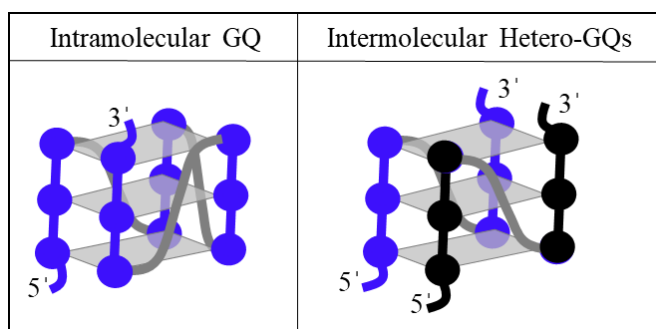
Figure 1.11 Homologous binding probes (shown in black) create an intermolecular hetero-GQ.

The Armitage Lab initially identified the ability of a PNA to form a hetero-GQ with *Oxytricha nova* telomeric DNA[78]. This complex bound 2:1 PNA:DNA, so that two hetero-GQs were formed from a single GQ target. A significant increase in thermal stability was also observed, compared to the DNA alone[78]. Further use of homologous probes was conducted on Myc-19 DNA. This short, 8mer, PNA again formed a heteroquadruplex in a 2:1 stoichiometry and provided a significant increase in thermal stability, which could be in part due to the stacking of the heteroGQs[79]. Modified PNA has further shown to eliminate off

target complementary binding, which can be a concern for genomic sequences. Incorporation of mini-poly-ethylene glycol (mini-PEG) monomers in place of the loop structures showed similar stabilities[80]. Left handed modified PNA molecules further destabilized off-target complementary binding, and the combination of mini-PEG abasic sites and left handed gamma modifications nearly eliminated complementary binding[80]. More recently, these probes have been utilized to target a GQ possessing a 17 nucleotide long loop[81]. In this GQ, the PNA probe bound with 6:1 stoichiometry, indicated by SPR and ESI-MS[81].

In vitro experiments found that the 8-mer PNA was able to inhibit polymerase extension past the GQ forming sequence[82]. This inhibition via hetero-GQ formation was found to be stronger than an already established GQ-stabilizing ligand (BRACO-19) and complementary PNA[82]. Homologous probes have also been used for the regulation of biological processes. G-rich RNA oligomers were used to inhibit EGFP expression *in vivo*, using plasmids transfected into HeLa cells[83]. These experiments provide a strong background for targeting GQ structures using homologous binding.

1.3.4   Combining Binding Methods to Optimize Selectivity and Affinity

While the binding modes above are discussed independently, they can be utilized simultaneously in order to optimize selectivity and affinity for GQ sequences (Figure 1.12). One example uses a nucleic acid probe to bind homologously to a GQ region and complementarily to the sequence proximal to the GQ[84]. This dual recognition probe was able to decrease gene expression in *in vitro* assays and in cancer cells[84]. A second example combines a hairpin DNA minor-groove binder [hairpin pyrrole-imidazole polyamides (hPIP)] and a GQ binding cyclic polyamide. This probe was able to effectively bind, and further improved the thermal stability of the GQ[85]. Each of these dual recognition probes aim to optimize the selectivity and affinity of their probes. In these examples, the homologous binder aims to stabilize the GQ region while the duplex binder serves to increase selectivity and affinity to the specific GQ of interest. The use of multiple binding modes will likely continue to be used since no single binding probe has ideal selectivity and affinity.
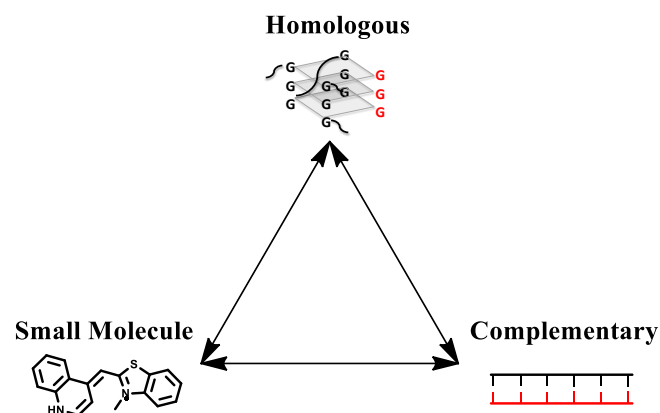
Figure 1.12 Combinations of binding modes (small molecule, homologous and complementary) for probe design.

1.4 Chapter Overviews and Contributions to the GQ Field

Many of the projects discussed in this thesis utilize one or more of the binding modes shown in Figure 1.12. As a result of this work, we expand our understanding of GQ secondary structures and show the ability to target these GQs using multiple binding modes.

Chapter 2 analyzes variations on a short peptide nucleic acid (PNA) probe that homologously binds to guanine-rich telomeric DNA. We compare the PNA probe with and without backbone modifications and utilize a GQ binding small molecule fluorophore; a combination of two binding modes. This work identifies that TO-labeled probes show significantly increased thermal stability compared to their unlabeled counterparts. Modifications inducing a right handed helix in the PNA show no difference in thermal stability. This short probe does not possess selectivity as evidenced by *in vitro* translation experiments performed in cell lysate.

In Chapter 3, we expand the current literature on RNA:DNA heteroquadruplexes (RDQ). This GQ contains both RNA and DNA guanine tracts and has recently been identified as a regulator between transcription and translation. We show the formation and stabilities of three suspected RDQ sequences using an optimized scaffold duplex. The formation of the RDQs was also monitored with a small molecule fluorophore, Thioflavin T.

Chapter 4 utilizes the RDQ systems created in Chapter 3 and compares them to DNA and RNA homoquadruplexes. With this chapter, we make DNA:DNA and RNA:RNA homoquadruplexes using the scaffold duplex created in chapter 3 and compare their biophysical properties to our RDQs. A modified PNA oligomer is then used to duplex invade each heteroquadruplex. We identify that RDQs possess characteristics of both RNA and DNA heteroquadruplexes. Our PNA probe is also able to sufficiently target each heteroquadruplex with varying IC_{50} values.

Chapter 5 describes the use of dual gammaPNA (γPNA) probes that hybridize to alternating sites along a telomere. Each probe contains a fluorophore, and can undergo Förster resonance energy transfer (FRET). We study the biophysical characteristics of these probes and observe bright staining of telomeres in nuclei. Finally, we show the incorporation of additional fluorophores does not significantly improve FRET efficiency.

1.5   References

(1)   Watson, J. D.; Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. In *50 Years of DNA*; Nature Publishing Group, 2016; Vol. 171, pp 83–84.

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

(3)   Franklin, R. E.; Gosling, R. G. The Structure of Sodium Thymonucleate Fibres. I. The Influence of Water Content. *Acta Crystallogr.* **1953**, *6* (8), 673–677.

(4)   Gagna, C. .; Lambert, W. . The Halting Arrival of Left-Handed Z-DNA. *Med. Hypotheses* **2003**, *60* (3), 418–423.

(5)   Felsenfeld, G.; Davies, D. R.; Rich, A. Formation of a Three-Stranded Polynucleotide Molecule. *Journal of the American Chemical Society*. American Chemical Society April 1957, pp 2023–2024.

(6)   Jain, A.; Wang, G.; Vasquez, K. M. DNA Triple Helices: Biological Consequences and Therapeutic Potential. *Biochimie* **2008**, *90* (8), 1117–1130.

(7)   Hoogsteen, K.; IUCr. The Structure of Crystals Containing a Hydrogen-Bonded Complex of 1-Methylthymine and 9-Methyladenine. *Acta Crystallogr.* **1959**, *12* (10), 822–823.

(8)   Chandrasekaran, A. R.; Rusling, D. A. Triplex-Forming Oligonucleotides: A Third Strand for DNA Nanotechnology. *Nucleic Acids Res.* **2018**, *46* (3), 1021–1037.

(9)   Abou Assi, H.; Garavís, M.; González, C.; Damha, M. J. I-Motif DNA: Structural Features and Significance to Cell Biology. *Nucleic Acids Res.* **2018**, *46* (16), 8038–8056.

(10)   Fleming, A. M.; Ding, Y.; Rogers, R. A.; Zhu, J.; Zhu, J.; Burton, A. D.; Carlisle, C. B.; Burrows, C. J. 4 *n* –1 Is a "Sweet Spot" in DNA i-Motif Folding of 2′-Deoxycytidine Homopolymers. *J. Am. Chem. Soc.* **2017**, *139* (13), 4682–4689.

(11)   Kwok, C. K.; Merrick, C. J. G-Quadruplexes: Prediction, Characterization, and Biological Application. *Trends Biotechnol.* **2017**, *35* (10), 997–1013.

(12) Chambers, V. S.; Marsico, G.; Boutell, J. M.; Di Antonio, M.; Smith, G. P.; Balasubramanian, S. High-Throughput Sequencing of DNA G-Quadruplex Structures in the Human Genome. *Nat. Biotechnol.* **2015**, *33* (8), 877–881.

(13) Lipps, H. J.; Rhodes, D. G-Quadruplex Structures: In Vivo Evidence and Function. *Trends Cell Biol.* **2009**, *19* (8), 414–422.

(14) Rhodes, D.; Lipps, H. J. G-Quadruplexes and Their Regulatory Roles in Biology. *Nucleic Acids Res.* **2015**, *43* (18), 8627–8637.

(15) Zeraati, M.; Langley, D. B.; Schofield, P.; Moye, A. L.; Rouet, R.; Hughes, W. E.; Bryan, T. M.; Dinger, M. E.; Christ, D. I-Motif DNA Structures Are Formed in the Nuclei of Human Cells. *Nat. Chem.* **2018**, *10* (6), 631–637.

(16) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. Quadruplex DNA: Sequence, Topology and Structure. *Nucleic Acids Res.* **2006**, *34* (19), 5402–5415.

(17) Guédin, A.; Gros, J.; Alberti, P.; Mergny, J.-L. How Long Is Too Long? Effects of Loop Size on G-Quadruplex Stability. *Nucleic Acids Res.* **2010**, *38* (21), 7858–7868.

(18) Kwok, C. K.; Sherlock, M. E.; Bevilacqua, P. C. Effect of Loop Sequence and Loop Length on the Intrinsic Fluorescence of G-Quadruplexes. *Biochemistry* **2013**, *52* (18), 3019–3021.

(19) Mukundan, V. T.; Phan, A. T. Bulges in G-Quadruplexes: Broadening the Definition of G-Quadruplex-Forming Sequences. *J. Am. Chem. Soc.* **2013**, *135* (13), 5017–5028.

(20) Tippana, R.; Xiao, W.; Myong, S. G-Quadruplex Conformation and Dynamics Are Determined by Loop Length and Sequence. *Nucleic Acids Res.* **2014**, *42* (12), 8106–8114.

(21) Merkina, E. E.; Fox, K. R. Kinetic Stability of Intermolecular DNA Quadruplexes. *Biophys. J.* **2005**, *89* (1), 365–373.

(22) Mergny, J. L.; Phan, A. T.; Lacroix, L. Following G-Quartet Formation by UV-Spectroscopy. *FEBS Lett.* **1998**, *435* (1), 74–78.

(23) Víglaský, V.; Bauer, L.; Tlučková, K. Structural Features of Intra- and Intermolecular G-Quadruplexes Derived from Telomeric Repeats. *Biochemistry* **2010**, *49* (10), 2110–2120.

(24) Kettani, A.; Bouaziz, S.; Gorin, A.; Zhao, H.; Jones, R. A.; Patel, D. J. Solution Structure of a Na Cation Stabilized DNA Quadruplex Containing G·G·G·G and G·C·G·C Tetrads Formed by G-G-G-C Repeats Observed in Adeno-Associated Viral DNA. *J. Mol. Biol.* **1998**, *282* (3), 619–636.

(25) Bouaziz, S.; Kettani, A.; Patel, D. J. A K Cation-Induced Conformational Switch within a Loop Spanning Segment of a DNA Quadruplex Containing G-G-G-C Repeats. *J. Mol. Biol.* **1998**, *282* (3), 637–652.

(26) Joachimi, A.; Benz, A.; Hartig, J. S. A Comparison of DNA and RNA Quadruplex Structures and Stabilities. *Bioorg. Med. Chem.* **2009**, *17* (19), 6811–6815.

(27) Bonnat, L.; Dejeu, J.; Bonnet, H.; Génnaro, B.; Jarjayes, O.; Thomas, F.; Lavergne, T.; Defrancq, E. Templated Formation of Discrete RNA and DNA:RNA Hybrid G-Quadruplexes and Their Interactions with Targeting Ligands. *Chem. - A Eur. J.* **2016**, *22* (9), 3139–3147.

(28) Xiao, C.-D.; Ishizuka, T.; Xu, Y. Antiparallel RNA G-Quadruplex Formed by Human Telomere RNA Containing 8-Bromoguanosine. *Sci. Rep.* **2017**, *7* (1), 6695.

(29) Moraca, F.; Amato, J.; Ortuso, F.; Artese, A.; Pagano, B.; Novellino, E.; Alcaro, S.; Parrinello, M.; Limongelli, V. Ligand Binding to Telomeric G-Quadruplex DNA Investigated by Funnel-Metadynamics Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (11), E2136–E2145.

(30) Bhattacharyya, D.; Mirihana Arachchilage, G.; Basu, S. Metal Cations in G-Quadruplex Folding and Stability. *Front. Chem.* **2016**, *4*, 38.

(31) and, A. W.; Wu*, G. Selective Binding of Monovalent Cations to the Stacking G-Quartet Structure Formed by Guanosine 5'-Monophosphate: A Solid-State NMR Study. **2003**.

(32) Hud, N.; Plavec, J. Quadruplex Nucleic Acids. In *Quadruplex Nucleic Acids*; Neidle, S., Balasubramanian, S., Eds.; Royal Society of Chemistry, 2006; pp 100–130.

(33) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative Visualization of DNA G-Quadruplex Structures in Human Cells. *Nat. Chem.* **2013**, *5* (3), 182–186.

(34) Henderson, A.; Wu, Y.; Huang, Y. C.; Chavez, E. A.; Platt, J.; Johnson, F. B.; Brosh, R. M.; Sen,

D.; Lansdorp, P. M. Detection of G-Quadruplex DNA in Mammalian Cells. *Nucleic Acids Res.* **2014**, *42* (2), 860–869.

(35)   Bochman, M. L.; Paeschke, K.; Zakian, V. A. DNA Secondary Structures: Stability and Function of G-Quadruplex Structures. *Nat. Rev. Genet.* **2012**, *13* (11), 770–780.

(36)   Andersen, P. [The Nobel Prize in Physiology or Medicine 1967]. *Tidsskr. Nor. Laegeforen.* **1967**, *87* (22), 1926–1927.

(37)   Shay, J. W.; Wright, W. E. Telomeres and Telomerase: Three Decades of Progress. *Nat. Rev. Genet.* **2019**, *20* (5), 299–309.

(38)   Armas, P.; David, A.; Calcaterra, N. B. Transcriptional Control by G-Quadruplexes: In Vivo Roles and Perspectives for Specific Intervention. *Transcription* **2017**, *8* (1), 21–25.

(39)   Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. Direct Evidence for a G-Quadruplex in a Promoter Region and Its Targeting with a Small Molecule to Repress c-MYC Transcription. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (18), 11593–11598.

(40)   Morgan, R. K.; Batra, H.; Gaerig, V. C.; Hockings, J.; Brooks, T. A. Identification and Characterization of a New G-Quadruplex Forming Region within the KRAS Promoter as a Transcriptional Regulator. *Biochim. Biophys. Acta - Gene Regul. Mech.* **2016**, *1859* (2), 235–245.

(41)   Kim, B. G.; Long, J.; Dubins, D. N.; Chalikian, T. V. Ionic Effects on VEGF G-Quadruplex Stability. *J. Phys. Chem. B* **2016**, *120* (22), 4963–4971.

(42)   Eser, S.; Schnieke, A.; Schneider, G.; Saur, D. Oncogenic KRAS Signalling in Pancreatic Cancer. *Br. J. Cancer* **2014**, *111* (5), 817–822.

(43)   Kaiser, C. E.; Van Ert, N. A.; Agrawal, P.; Chawla, R.; Yang, D.; Hurley, L. H. Insight into the Complexity of the I-Motif and G-Quadruplex DNA Structures Formed in the *KRAS* Promoter and Subsequent Drug-Induced Gene Repression. *J. Am. Chem. Soc.* **2017**, *139* (25), 8522–8536.

(44)   Amato, J.; Madanayake, T. W.; Iaccarino, N.; Novellino, E.; Randazzo, A.; Hurley, L. H.; Pagano, B. HMGB1 Binds to the *KRAS* Promoter G-Quadruplex: A New Player in Oncogene Transcriptional Regulation? *Chem. Commun.* **2018**, *54* (68), 9442–9445.

(45)   Millevoi, S.; Moine, H.; Vagner, S. G-Quadruplexes in RNA Biology. *Wiley Interdiscip. Rev. RNA* **2012**, *3* (4), 495–507.

(46)   Bugaut, A.; Balasubramanian, S. 5'-UTR RNA G-Quadruplexes: Translation Regulation and Targeting. *Nucleic Acids Res.* **2012**, *40* (11), 4727–4741.

(47)   Wang, C.; Zhao, L.; Lu, S. Role of TERRA in the Regulation of Telomere Length. *Int. J. Biol. Sci.* **2015**, *11* (3), 316–323.

(48)   Rippe, K.; Luke, B. TERRA and the State of the Telomere. *Nat. Struct. Mol. Biol.* **2015**, *22* (11), 853–858.

(49)   Dilley, R. L.; Greenberg, R. A. ALTernative Telomere Maintenance and Cancer. *Trends in cancer* **2015**, *1* (2), 145–156.

(50)   Oyaghire, S. N.; Cherubim, C. J.; Telmer, C. A.; Martinez, J. A.; Bruchez, M. P.; Armitage, B. A. RNA G-Quadruplex Invasion and Translation Inhibition by Antisense γ-Peptide Nucleic Acid Oligomers. *Biochemistry* **2016**, *55* (13), 1977–1988.

(51)   Morris, M. J.; Negishi, Y.; Pazsint, C.; Schonhoft, J. D.; Basu, S. An RNA G-Quadruplex Is Essential for Cap-Independent Translation Initiation in Human VEGF IRES. *J. Am. Chem. Soc.* **2010**, *132* (50), 17831–17839.

(52)   Bettin, N.; Oss Pegorar, C.; Cusanelli, E. The Emerging Roles of TERRA in Telomere Maintenance and Genome Stability. *Cells* **2019**, *8* (3).

(53)   Xu, Y.; Ishizuka, T.; Yang, J.; Ito, K.; Katada, H.; Komiyama, M.; Hayashi, T. Oligonucleotide Models of Telomeric DNA and RNA Form a Hybrid G-Quadruplex Structure as a Potential Component of Telomeres. *J. Biol. Chem.* **2012**, *287* (50), 41787–41796.

(54)   Lee, D. Y.; Clayton, D. A. Properties of a Primer RNA-DNA Hybrid at the Mouse Mitochondrial DNA Leading-Strand Origin of Replication. *J. Biol. Chem.* **1996**, *271* (39), 24262–24269.

(55)   Wanrooij, P. H.; Uhler, J. P.; Shi, Y.; Westerlund, F.; Falkenberg, M.; Gustafsson, C. M. A Hybrid G-Quadruplex Structure Formed between RNA and DNA Explains the Extraordinary Stability of

the Mitochondrial R-Loop. *Nucleic Acids Res.* **2012**, *40* (20), 10334–10344.

(56)   Zheng, K.; Xiao, S.; Liu, J.; Zhang, J.; Hao, Y.; Tan, Z. Co-Transcriptional Formation of DNA:RNA Hybrid G-Quadruplex and Potential Function as Constitutional Cis Element for Transcription Control. *Nucleic Acids Res.* **2013**, *41* (10), 5533–5541.

(57)   Zheng, K.; Wu, R.; He, Y.; Xiao, S.; Zhang, J.; Liu, J.; Hao, Y.; Tan, Z. A Competitive Formation of DNA:RNA Hybrid G-Quadruplex Is Responsible to the Mitochondrial Transcription Termination at the DNA Replication Priming Site. *Nucleic Acids Res.* **2014**, *42* (16), 10832–10844.

(58)   Zhang, J.; Zheng, K.; Xiao, S.; Hao, Y.; Tan, Z. Mechanism and Manipulation of DNA:RNA Hybrid G-Quadruplex Formation in Transcription of G-Rich DNA. *J. Am. Chem. Soc.* **2014**, *136* (4), 1381–1390.

(59)   Zhao, Y.; Zhang, J.; Zhang, Z.; Tong, T.; Hao, Y.; Tan, Z. Real-Time Detection Reveals Responsive Cotranscriptional Formation of Persistent Intramolecular DNA and Intermolecular DNA:RNA Hybrid G-Quadruplexes Stabilized by R-Loop. *Anal. Chem.* **2017**, *89* (11), 6036–6042.

(60)   Xiao, S.; Zhang, J.-Y.; Zheng, K.-W.; Hao, Y.-H.; Tan, Z. Bioinformatic Analysis Reveals an Evolutional Selection for DNA:RNA Hybrid G-Quadruplex Structures as Putative Transcription Regulatory Elements in Warm-Blooded Animals.

(61)   Wu, R.; Zheng, K.; Zhang, J.; Hao, Y.; Tan, Z. Formation of DNA:RNA Hybrid G-Quadruplex in Bacterial Cells and Its Dominance over the Intramolecular DNA G-Quadruplex in Mediating Transcription Termination. *Angew. Chemie Int. Ed.* **2015**, *54* (8), 2447–2451.

(62)   Ma, D.-L.; Zhang, Z.; Wang, M.; Lu, L.; Zhong, H.-J.; Leung, C.-H. Recent Developments in G-Quadruplex Probes. *Chem. Biol.* **2015**, *22* (7), 812–828.

(63)   Monchaud, D.; Teulade-Fichou, M.-P. A Hitchhiker's Guide to G-Quadruplex Ligands. *Org. Biomol. Chem.* **2008**, *6* (4), 627–636.

(64)   Mohanty, J.; Barooah, N.; Dhamodharan, V.; Harikrishna, S.; Pradeepkumar, P. I.; Bhasikuttan, A. C. Thioflavin T as an Efficient Inducer and Selective Fluorescent Sensor for the Human Telomeric G-Quadruplex DNA. *J. Am. Chem. Soc.* **2013**, *135* (1), 367–376.

(65)   Xu, S.; Li, Q.; Xiang, J.; Yang, Q.; Sun, H.; Guan, A.; Wang, L.; Liu, Y.; Yu, L.; Shi, Y.; et al. Thioflavin T as an Efficient Fluorescence Sensor for Selective Recognition of RNA G-Quadruplexes. *Sci. Rep.* **2016**, *6*.

(66)   Guan, A.; Zhang, X.-F.; Sun, X.; Li, Q.; Xiang, J.-F.; Wang, L.-X.; Lan, L.; Yang, F.-M.; Xu, S.-J.; Guo, X.-M.; et al. Ethyl-Substitutive Thioflavin T as a Highly-Specific Fluorescence Probe for Detecting G-Quadruplex Structure. *Sci. Rep.* **2018**, *8* (1), 2666.

(67)   Calabrese, D. R.; Zlotkowski, K.; Alden, S.; Hewitt, W. M.; Connelly, C. M.; Wilson, R. M.; Gaikwad, S.; Chen, L.; Guha, R.; Thomas, C. J.; et al. Characterization of Clinically Used Oral Antiseptics as Quadruplex-Binding Ligands. *Nucleic Acids Res.* **2018**, *46* (6), 2722–2732.

(68)   Zhai, Q.; Gao, C.; Ding, J.; Zhang, Y.; Islam, B.; Lan, W.; Hou, H.; Deng, H.; Li, J.; Hu, Z.; et al. Selective Recognition of C-MYC Pu22 G-Quadruplex by a Fluorescent Probe. *Nucleic Acids Res.* **2019**, *47* (5), 2190–2204.

(69)   O'Hagan, M. P.; Morales, J. C.; Galan, M. C. Binding and Beyond: What Else Can G-Quadruplex Ligands Do? *European J. Org. Chem.* **2019**.

(70)   Luigi Martino, †; Ada Virno, ‡; Bruno Pagano, §; Antonella Virgilio, ‡; Simone Di Micco, §; Aldo Galeone, ‡; Concetta Giancola, †; Giuseppe Bifulco, §; Luciano Mayol, ‡ and; Antonio Randazzo*, ‡. Structural and Thermodynamic Studies of the Interaction of Distamycin A with the Parallel Quadruplex Structure [d(TGGGGT)]4. **2007**.

(71)   Rouleau, S. G.; Beaudoin, J. D.; Bisaillon, M.; Perreault, J. P. Small Antisense Oligonucleotides against G-Quadruplexes: Specific MRNA Translational Switches. *Nucleic Acids Res.* **2015**, *43* (1), 595–606.

(72)   Kumar, N.; Patowary, A.; Sivasubbu, S.; Petersen, M.; Maiti, S. Silencing C-*MYC* Expression by Targeting Quadruplex in P1 Promoter Using Locked Nucleic Acid Trap [†]. *Biochemistry* **2008**, *47*

(50), 13179–13188.

(73) Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. Sequence-Selective Recognition of DNA by Strand Displacement with a Thymine. *Source Sci. New Ser.* **1991**, *254* (5037), 1497–1500.

(74) Krishnan, Y.; Simmel, F. C. Nucleic Acid Based Molecular Devices. *Angewandte Chemie - International Edition*. John Wiley & Sons, Ltd March 28, 2011, pp 3124–3156.

(75) Panyutin, I. G.; Onyshchenko, M. I.; Englund, E. A.; Appella, D. H.; Neumann, R. D. Targeting DNA G-Quadruplex Structures with Peptide Nucleic Acids. *Curr. Pharm. Des.* **2012**, *18* (14), 1984–1991.

(76) Datta, B.; Armitage, B. A. Hybridization of PNA to Structured DNA Targets: Quadruplex Invasion and the Overhang Effect. *J. Am. Chem. Soc.* **2001**, *123* (39), 9612–9619.

(77) Gupta, A.; Lee, L. L.; Roy, S.; Tanious, F. A.; Wilson, W. D.; Ly, D. H.; Armitage, B. A. Strand Invasion of DNA Quadruplexes by PNA: Comparison of Homologous and Complementary Hybridization. *ChemBioChem* **2013**, *14* (12), 1476–1484.

(78) Datta, B.; Schmitt, C.; Armitage, B. A. Formation of a PNA2-DNA2 Hybrid Quadruplex. *J. Am. Chem. Soc.* **2003**, *125* (14), 4111–4118.

(79) Roy, S.; Tanious, F. A.; Wilson, W. D.; Ly, D. H.; Armitage, B. A. High-Affinity Homologous Peptide Nucleic Acid Probes for Targeting a Quadruplex-Forming Sequence from a MYC Promoter Element. *Biochemistry* **2007**, *46* (37), 10433–10443.

(80) Lusvarghi, S.; Murphy, C. T.; Roy, S.; Tanious, F. A.; Sacui, I.; Wilson, W. D.; Ly, D. H.; Armitage, B. A. Loop and Backbone Modifications of Peptide Nucleic Acid Improve G-Quadruplex Binding Selectivity. *J. Am. Chem. Soc.* **2009**, *131* (51), 18415–18424.

(81) Kormuth, K. A.; Woolford, J. L.; Armitage, B. A. Homologous PNA Hybridization to Noncanonical DNA G-Quadruplexes. *Biochemistry* **2016**, *55* (12), 1749–1757.

(82) Murphy, C. T.; Gupta, A.; Armitage, B. A.; Opresko, P. L. Hybridization of G-Quadruplex-Forming Peptide Nucleic Acids to Guanine-Rich Dna Templates Inhibits DNA Polymerase η Extension. *Biochemistry* **2014**, *53* (32), 5315–5322.

(83) Ito, K.; Go, S.; Komiyama, M.; Xu, Y. Inhibition of Translation by Small RNA-Stabilized MRNA Structures in Human Cells. *J. Am. Chem. Soc.* **2011**, *133* (47), 19153–19159.

(84) Bhattacharyya, D.; Nguyen, K.; Basu, S. Rationally Induced RNA:DNA G-Quadruplex Structures Elicit an Anticancer Effect by Inhibiting Endogenous EIF-4E Expression. *Biochemistry* **2014**, *53* (33), 5461–5470.

(85) Asamitsu, S.; Obata, S.; Phan, A. T.; Hashiya, K.; Bando, T.; Sugiyama, H. Simultaneous Binding of Hybrid Molecules Constructed with Dual DNA-Binding Components to a G-Quadruplex and Its Proximal Duplex. *Chem. - A Eur. J.* **2018**, *24* (17), 4428–4435.

2 Chapter 2. Comparative Analysis of PNA Modifications on Heteroquadruplex Formation.

2.1 Introduction

2.1.1 Brief Introduction to Guanine Quadruplexes

Guanine-quadruplexes (GQ) are secondary structures of nucleic acid that form in guanine-rich sequences[1]. These sequences form tetrads (Figure 2.1) which utilize Hoogsteen and Watson-Crick faces for hydrogen bonding and rely on cations from the surrounding buffer for stabilization[1]. Tetrads stack via π-π interactions and form a GQ[1]. The cation (X^+ in Figure 2.1) stabilizes the GQ by coordinating O6 in each guanine residue by both size and charge[2]. Potassium is one of the best GQ stabilizing cations due to its optimal atomic radius[3]. Additional mono- and divalent ions that are known to stabilize GQs are sodium and magnesium[3]. Lithium, which possesses a smaller ionic radius, does not coordinate the tetrad, and therefore does not stabilize GQ formation[3].
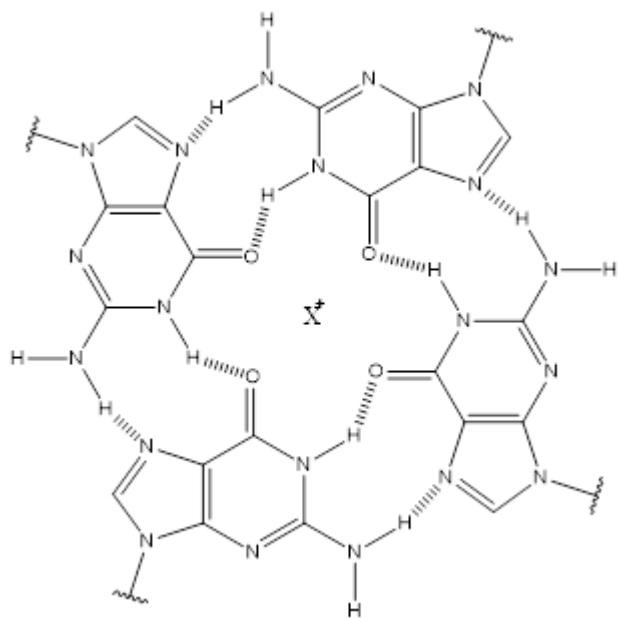
Figure 2.1 Basic structure of a G-tetrad.

GQs can be classified according to the direction of the nucleic acid strand in the tetrad. This can be determined by identifying the glycosidic bond in deoxyguanosine(dG) as syn- or *anti*[4]. When all strands in a GQ are in the same 5' to 3' orientation, they are labeled as parallel[4]. *Anti*-parallel GQs form when two strands are opposite in orientation, labeled as 2+2 in Figure 2.2[4]. Parallel GQs possess *anti*-dG, while *anti*-parallel GQs contain both syn- and *anti*-dG[4]. In unimolecular GQs, antiparallel structure results from alternating orientation between each continuous strand (2+2)[4]. Additionally, GQs can form hybrid structures where one strand runs opposite the

other three (3+1) resulting in an uneven number of syn- and *anti*-dG[4]. Some GQ structures have been characterized as displaying specific conformations, while other researchers have found that GQs can be forced from one conformation to another via external factors[4].
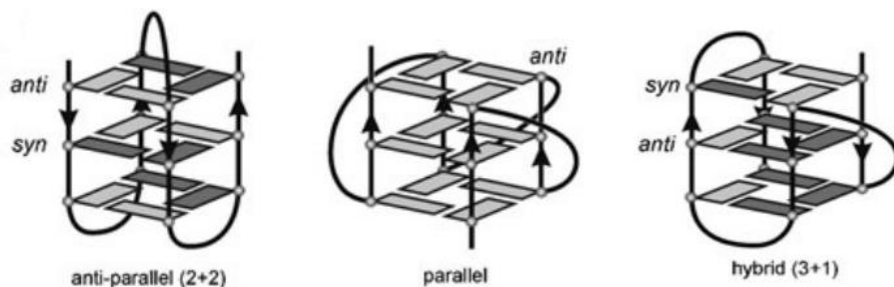
Figure 2.2 GQ structure as anti-parallel, parallel, and hybrid. Adapted from Reference 5.

2.1.2 Biological Implications of Guanine Quadruplexes

Recent research shows *in vivo* GQs play a critical role in many cellular processes, like transcription[6] and post-transcriptional gene regulation[7]. The highest concentration of G-quadruplex forming sequences occurs in repetitive DNA sequences like telomeres[8]. Telomeres are located at chromosomal ends, consist of both double and single stranded portions, and have the repeating sequence TTAGGG in humans[9]. Cells use telomere length as an indicator of the number of replications a cell has undergone[9]. However, cells possessing telomere extension mechanisms cannot utilize such an indicator and do not undergo cellular senescence, leading to DNA mutation[10]. As a result, telomere length has been associated with cellular aging, tumorigenesis, and cancer, making telomeric GQs interesting targets for disease treatment[11,12].

Telomeric repeats transcribed to RNA(TERRA) are involved in telomere extension and are critical for telomerase function[13]. Telomerase, the enzyme responsible for telomere extension, is a ribonucleoprotein that requires a TERRA sequence in order to function[14]. For this reason, TERRA molecules are found in high concentrations near telomeres and have been shown to have higher concentrations in cancer cells where telomerase is typically overactive[14]. In systems lacking telomerase, TERRA molecules can assist in telomere length regulation in telomerase-negative ALT systems[15]. The

formation of RNA GQs formed by TERRA sequences has been well studied *in vitro*, but the interaction between TERRA GQs and the wide array of interactions with telomerase (as well as telomeric DNA) is a topic in need of further research.

mRNA transcripts, and more specifically the 5' untranslated region (UTR), show a strong localization of GQ forming sequences across species. The presence of these possible GQ sequences concentrated around these regions indicates a possible evolutionary conservation and regulatory role[16]. RNA GQs have been shown to regulate translation by preventing the 5' cap from binding to translation initiation factors, referred to as cap-dependent translation[17]. RNA GQs were subsequently shown to inhibit translation by cap-independent translation, where GQ folding modifies the internal ribosome entry sites (IRES) and impedes the ribosome from binding[17]. These findings are promising for researchers looking to use RNA GQs as therapeutic targets. Our lab has participated in this subject by utilizing peptide nucleic acid to bind and prevent RNA GQ formation[18,19]. By doing so, we have shown transcription inhibition in *in vitro* luciferase assays[18]. But, some research suggests manipulating 5' UTR GQs leads to an enhancement in translation[20]. This discrepancy between different sequences is not well understood and needs to be considered on a case by case basis in order to effectively target with the preferred outcome (either enhancement or inhibition of translation).

2.1.3   Methods for Targeting Guanine Quadruplexes

Previous GQ targeting agents include small molecules; however, these do not target specific bases in GQ structure[11]. Nucleic acid probes offer the ability to target both primary and secondary structures[21]. Peptide nucleic acid (PNA) is one such type of probe that offers more specific targeting due to its backbone design. PNA possesses the base structures of DNA, but the sugar phosphate backbone is replaced with N-(2-aminoethyl)glycine subunits (Figure 2.3)[22]. The neutral PNA backbone has
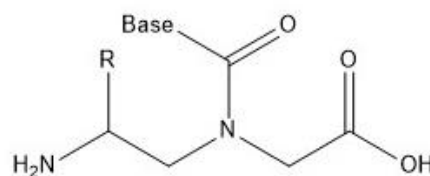
Figure 2.3 The basic structure of peptide nucleic acid.

distinct properties including resistance to degradation by nucleases and proteases as well as high binding affinity[21,23]. Modifications of the PNA structure have been shown to provide additional benefits; particularly, a miniPEG (polyethylene glycol) modification to the gamma carbon on the PNA backbone (R in Figure 2.3)[24]. This modification gives the PNA a helical motif and can significantly increase affinity and selectivity in the PNA's ability to bind single and double stranded DNA[24].

Fluorescent dyes are often used to label, monitor, and further stabilize PNA bound structures; one example is thiazole orange (TO). TO-labeled PNA possesses high fluorescence when hybridized to complementary DNA or RNA and low fluorescence when free in solution[25]. This change in fluorescence is due to the rotational restriction of the carbon-carbon bond between the two aromatic portions, seen in Figure 2.4[26]. TO is an intercalator dye, meaning it is able to intercalate (or insert) between base pairs in a duplex nucleic acid sequence[27]. In a GQ structure, TO likely π-stacks on either the top or the bottom tetrad, leading to an increase in fluorescence.

Figure 2.4 Thiazole orange (TO) structure.

The Armitage lab studies GQ structures and the ability to invade GQs to form hetero-GQs. Initially, the lab identified the ability of a PNA to form a hetero-GQ with *Oxytricha nova* telomeric DNA[28]. This complex, called PNA$_2$-DNA$_2$, showed a significant increase in thermal stability compared to the DNA alone[28]. This novel recognition motif opened doors for a new type of nucleic acid probe.

Modifications of PNA probes were also studied in order to identify an optimal probe for the Myc19 DNA sequence[29]. These modifications sought to enhance probe selectivity by decreasing the affinity for complementary binding without interfering with homologous binding[29]. Modifications included the incorporation of abasic sites and backbone modifications[29]. The incorporation of abasic sites favored the formation of a hetero-GQ over binding a complementary strand based on thermal melting temperatures[29]. Backbone modifications at the gamma position (to induce a right or left handed helix in the PNA molecule) showed that right handed PNA had no effect on the thermal stability of complementary binding[29]. However,

left handed PNA molecules destabilized complementary binding and the combination of abasic sites and left handed gamma modifications nearly eliminated complementary binding[29].

Our lab also showed the first example of kinetic variation between GQ sequences. Using surface plasmon resonance (SPR), kinetics experiments were performed using five different GQ forming sequences and the G-rich 8-mer PNA from previous experiments[30]. Interestingly, PNA displayed the fastest association with parallel GQs[30]. I*n vitro* experiments were done to test the ability of a PNA-DNA hetero-GQ to prevent extension by a translation polymerase[31]. Analysis of the polymerase extension mechanism found that the 8-mer PNA was able to inhibit polymerase extension past the GQ forming sequence[31]. Inhibition via hetero-GQ formation was found to be stronger than an already established GQ-stabilizing ligand (BRACO-19) and complementary PNA[31].

PNA can also be used to form a GQ in a sequence that would not otherwise permit GQ formation. Work from Sylvain Ladame's lab utilizes a dye-labeled short peptide nucleic acid (PNA) to target a truncated form of human telomeric (hTelo) DNA[32]. The truncated hTelo DNA contains three guanine-rich tracts and requires the short guanine-rich PNA to form the final tract of the GQ, i.e. a 3+1 binding mode[32]. Data from this work demonstrates acridone-labeled PNA is able to form a bimolecular GQ with truncated hTelo DNA[32]. UV-Vis, CD and fluorescence experiments were used to identify PNA-DNA interactions[32]. An acridone dye moiety on the N-terminus of the PNA allows fluorescent information to be obtained, as well as providing additional stabilization to the complex by base stacking[32]. The group showed higher thermal stability in the PNA-DNA GQ than the DNA alone using melting temperatures[32]. CD spectra of the complexes showed that DNA alone possesses *anti*-parallel-favoring hybrid GQ, and DNA + PNA generally shows a hybrid GQ that is more parallel in nature[32]. This work employs a novel probe which uses Hoogsteen and Watson-Crick interactions, as well as dye association by π-stacking. We wanted to utilize this 3+1 binding mode and test it using a variety of PNA modifications, and identify if the 3+1 binding mode would invade a full length GQ sequence (as opposed to the previously used truncated version).

2.1.4   Project Goals

The project discussed here uses a dual-recognition approach to target GQs. Similarly to Ladame's approach discussed above, we use two binding domains, a homologous GQ binder and a stacking dye, in order to modulate selectivity and affinity. Four different PNAs with the homologous binding sequence GGG were used (Table 2.1) to accomplish the following aims: 1. Extend Ladame's findings with the interactions of PNA on hTelo DNA by using both truncated and full length sequences. 2. Identify the effects of different PNAs on the truncated and full length nucleic acid using the following techniques: UV-Vis, CD, and surface plasmon resonance (SPR). 3. Characterize the PNA's ability to inhibit translation in a cell lysate system.

The following table will be used in order to explicitly identify the comparisons being made throughout this project. The specific components of discussion will be shaded in each respective section.

Length of Nucleic Acid	Type of Nucleic Acid	PNA Modification
		No modification
3G3	DNA	γ backbone
4G3	RNA	+TO

2.2   Materials and Methods

Table 2.1 Peptide nucleic acids used in *in vitro* studies. Gamma residues are underlined.

Name	Sequence
PNA	GGG
PNA-TO	GGG-TO
γPNA	GGG
γPNA-TO	GGG-TO

Table 2.2 Nucleic acids used in *in vitro* studies.

Name	Sequence
3G3 D/RNA	5'-(AGGGTT)$_3$-3'
4G3 D/RNA	5'-(AGGGTT)$_4$-3'

PNA Sample Preparation PNA and PNA-TO were synthesized by Dr. Lisa Rastede using solid phase synthesis[33] by Boc- protected monomers, purified using HPLC, and confirmed using MALDI-TOF mass spectrometry with α-cyano-4-hydroxycinnamic acid as the matrix dissolved in 50:50 H_2O/ACN with 0.1% TFA.

The final product was lyophilized and dissolved in nanopure water to make a 200 μM stock solution. γPNA and γPNA-TO were purchased from PNA Innovations (http://pnainnovations.com/).

DNA and RNA Sample Preparation All oligonucleotides were purchased from Integrated DNA Technologies (https://www.idtdna.com) and dissolved in nanopure water. All samples were prepared in 100 mM KCl T_m buffer (10 mM Tris Base, 0.1 mM salt free EDTA, and 100 mM KCl pH=7.00).

UV-Vis Spectroscopy was conducted using a Cary 300 UV-Vis Spectrophotometer with a thermoelectrically controlled Peltier cell holder. Concentration checks were performed by scanning absorbance from 240-600 nm above 90 °C. Melting curves measured absorbance at 295 nm from 95 °C to 15 °C (cooling at 1° per min). Quartz spectrophotometer cuvettes (10 mM path lengths) from Starna Cells Inc. (http://www.starnacells.com/) were used for all experiments. The melting curve was normalized and the maximum of the first derivative plot was identified as the melting point (T_m).

Fluorescence Spectroscopy was conducted using a Tecan Infinite M1000 Plate Spectrometer. Job plot analysis utilized varying concentrations of GQ and (γ)PNA-TO, such that the overall concentration remained constant but the ratio varied from 0-1. Fluorescence was measured using an excitation wavelength of 495 nm and an emission wavelength of 520 nm.

Circular Dichroism (CD) Spectroscopy was conducted using a JASCO 715 CD spectropolarimeter with a Peltier water-circulating temperature controller. Specific parameters included scan rate (100 nm/min), data interval (0.5 nm), number of scans (6), and temperature control (37 °C).

SPR Experiments

Table 2.3 SPR DNA Sequences.

Name	Sequence
3G3 DNA (SPR)	5'-biotin-TTTTTAGGGTTAGGGTTAGGGTTA-3'
4G3 DNA (SPR)	5'-biotin-TTTTTAGGGTTAGGGTTAGGGTTAGGG-3'

1X SPR buffer consisted of 100 mM KCl (or LiCl), 10 mM Tris, 1 mM salt free EDTA, and 0.0025% P-20 (pH=7.4). SPR buffer was prepared at 5X and diluted to 1X with nanopure water.

SPR Chip Preparation Streptavidin and each DNA were immobilized on the CM5 sensor chip. The chip surface was activated by injecting a mixture of N-hydroxysuccinimide and N-ethyl-N'-(dimethylaminopropyl) carbodiimide over the chip for 10 minutes (rate=5μL/min). Then streptavidin was injected until approximately 7000 response units (RU) were immobilized on all four flow cells (rate=5μL/min). Ethanolamine was then injected for 10 minutes to cap the chip (rate=5μL/min), followed by seven injections of regeneration solution (10 mM NaOH and 1M NaCl) (rate=50μL/min). After overnight stabilization in the SPR instrument, biotinylated DNA was immobilized on the chip surface. 3G3 DNA (SPR) and 4G3 DNA (SPR) were injected into separate flow cells until approximately 100RU were immobilized (rate=2μL/min). One flow cell was left with streptavidin as a control. Chip preparation was conducted using lithium 1X SPR buffer.

SPR Concentration Series and Direct Binding Experiments Concentration series and direct binding experiments utilized a series of injection cycles: regeneration, buffer, and sample. These experiments were performed in both potassium and lithium buffer, which is specified for each experiment. All runs were background subtracted using the streptavidin flow cell and buffer subtracted. In direct binding experiments where the type of PNA varied, the response units were corrected for the difference in molecular weight of each PNA.

Luciferase Assay

Table 2.4 Plasmid 5'-UTR G-quadruplex sequences.

Plasmid	5'-UTR G-quadruplex Sequence
3G3	GGGTTAGGGTTAGGG
4G3	GGGTTAGGGGTTAGGGGTTAGGG
noG3	TCAGATCCGCTAGCGCTACCG

Plasmid Amplification DNA plasmids were provided by Dr. Cheryl Telmer. Each plasmid (3G3, 4G3, and noG3 shown in Table 2.4 and Supporting Figure 2.4-Supporting Figure 2.6) was mini-prepped to amplify the DNA concentration. This was done by heat shocking the plasmid into MACH1/T1 *Escherichia coli*. A

single colony from the plated bacteria was added to 50 mL of lysogeny broth (LB) media and grown in a shaker overnight. Plasmid DNA was then extracted and purified using a ThermoFisher GeneJET Purification kit.

Plasmid Linearization A restriction digest was performed using the APAL1 restriction site. Samples were incubated at 37 °C for two hours. The linearized plasmid was extracted using a ThermoFisher Gel Extract Kit (https://www.thermofisher.com/).

Transcription Linearized DNA (1µg) was transcribed *in vitro* with T7 RNA polymerase (30 units), Ribolock RNase Inhibitor (50 units), 2 mM of each rNTP (total volume 50 µL) at 37 °C for two hours. The mRNA was then purified using the ThermoFisher GeneJET RNA clean-up kit (https://www.thermofisher.com/).

Annealing and Translation mRNA samples were annealed with PNA probes at 37 °C for two hours under translation salt conditions (79 mM KCl and 7.9 mM Tris-HCl). The annealed samples were then mixed with nuclease-treated rabbit reticulocyte lysate, amino acid mixtures minus leucine, amino acid mixtures minus methionine, and RNase inhibitor (all purchased from Promega http://www.promega.com/). Translation samples (35 µL total volume) were incubated at 30 °C for 90 minutes.

Luciferase Assay The luciferase assay was conducted by mixing 10µL of the translation samples with 20µL of a luciferin solution (D-luciferin, Mg^{2+}, and ATP) (Promega http://www.promega.com/) and measured on a Tecan Infinite M1000 Plate Spectrometer as Relative Light Units (RLU).

2.3 Results

2.3.1 *3G3 DNA forms a weak bimolecular GQ with itself and a strong bimolecular GQ with unmodified PNA.*

Length of Nucleic Acid	Type of Nucleic Acid	PNA Modification
		No modification
3G3	DNA	γ backbone
4G3	RNA	+TO

3G3 DNA is a human telomere based sequence with three repeating tracks of guanines. This sequence is unable to form an intramolecular GQ. However, two strands of 3G3 DNA could form an intermolecular GQ. This is observed by thermal melting spectra, where an unstable GQ transition is observed at 35 °C. Thermal melting curves monitor absorbance at a particular wavelength over a temperature range. The formation of GQs are unique because they exhibit a hypochromic shift at 295 nm[34]. Overall absorbance was converted to fraction bound for these experiments in order to compare melting curves. The transition observed for 3G3 DNA is quite unstable (Figure 2.5). A ΔT_m of approximately 11 °C was observed between the 3G3 DNA and the 3G3 DNA-PNA, indicating the formation of a DNA-PNA GQ and an increase in stability when PNA is present.
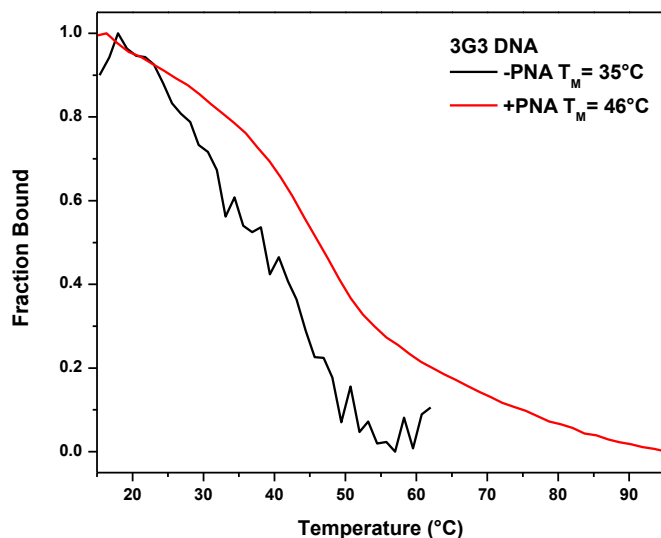
Figure 2.5 Thermal Melting of 3G3 DNA with and without PNA in 100 mM KCl Buffer. [3G3 DNA] = 1 µM [PNA] = 1 µM. Melting temperatures without PNA = 35 °C and with PNA = 46 °C.

Circular dichroism (CD) measures differences in the way chiral molecules absorb right handed and left handed circularly polarized light. These differences have specific profiles for unique secondary structures, allowing scientists to gather information about the secondary structure of these molecules. For GQs, there are three commonly observed secondary structures: parallel, *anti*-parallel, and hybrid. The three structures indicate the directionality of the guanine tracts in relation to the 5′ to 3′ run of the DNA. Parallel GQs show a maximum at 260 nm and a minimum at 240 nm, while *anti*-parallel GQs show maxima at 295 nm and 245 nm and a minimum at 260 nm[4]. Hybrid GQs are a combination of parallel and *anti*-parallel elements and show a combination of both CD spectra[5]. 3G3 DNA shows hybrid GQ CD signal, indicating the formation of an intermolecular GQ. However, this signal is very weak, which would be expected considering the T_M for this GQ is 35 °C. Upon the addition of PNA, the peak trends remain the same but the overall signal intensity increases. Due to the low signal intensity, these experiments were repeated at 5μM DNA and PNA and confirmed overall curve structure (Supporting Figure 2.1).
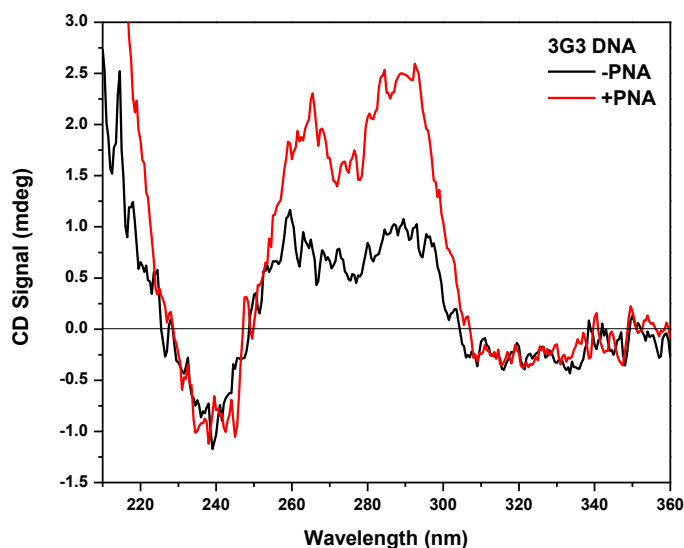
Figure 2.6 Circular dichroism spectra for 3G3 DNA with and without PNA. [3G3 DNA] = 1 µM [PNA] = 1 µM. Antiparallel secondary structure is indicated for both GQs.

2.3.2 *TO modified PNA increases stability in 3G3 DNA but only slightly affects the overall secondary structure.*

Length of Nucleic Acid	Type of Nucleic Acid	PNA Modification
		No modification
3G3	DNA	γ backbone
4G3	RNA	+TO

Selective fluorophores have been used as light up probes for the past few decades. Specific to PNA, acridone was previously tethered to a PNA molecule in order to modulate binding and stabilize the DNA-PNA GQ[32]. Light up fluorophores generally π-π stack on the top and/or bottom tetrad of GQs, and release absorbed energy in the form of fluorescence through such stacking interactions[25,27]. Due to these stacking interactions, GQ binding fluorophores are also generally GQ stabilizing. This is observed in the T_M, where PNA-TO increases the melting temperature by 22 °C compared to unlabeled PNA (Figure 2.7 A). Furthermore, the 3G3 DNA-PNA-TO transition occurs over a narrower temperature range compared to 3G3 DNA-PNA, indicating the stabilizing effects of the TO.
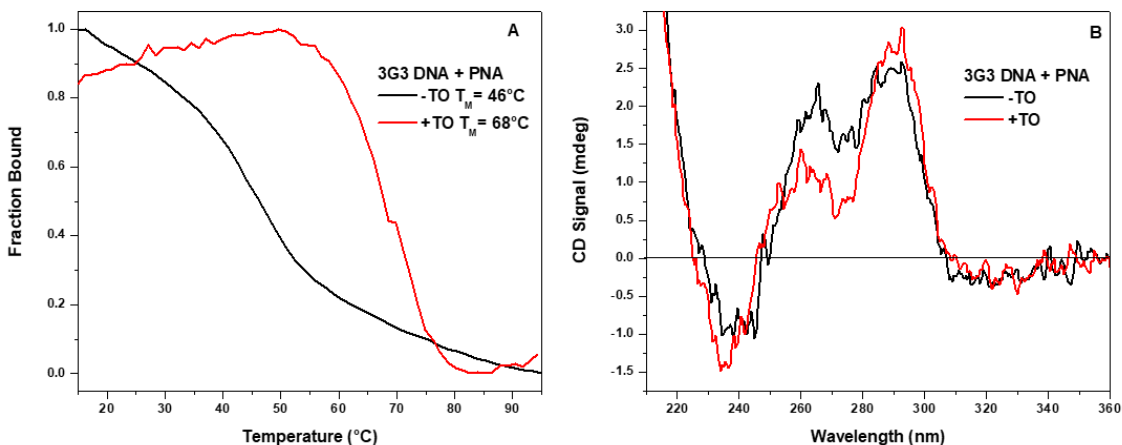
Figure 2.7 (A) Thermal Melting and (B) circular dichroism of 3G3 DNA + PNA with and without TO in 100 mM KCl Buffer. [3G3 DNA] = 1 µM [PNA(-TO)] = 1 µM. Melting temperatures without TO = 46 °C and with TO = 68 °C. CD suggests a shift to antiparallel secondary structure in the presence of TO.

The CD spectra aid in identifying any changes in secondary structure in the presence of TO. Without TO, as discussed in 2.3.1, we observe hybrid CD signal. While TO was expected to interact with the GQ via π-π stacking, we would not expect the overall secondary structure to change. PNA with TO shows hybrid GQ structures (Figure 2.7 B). There is a decrease in parallel GQ signal (decrease in peak intensity at 260 nm) indicating a possible change in the orientation of the GQ formed in these two samples. However, this decrease is not significant enough to warrant a change in overall topology.
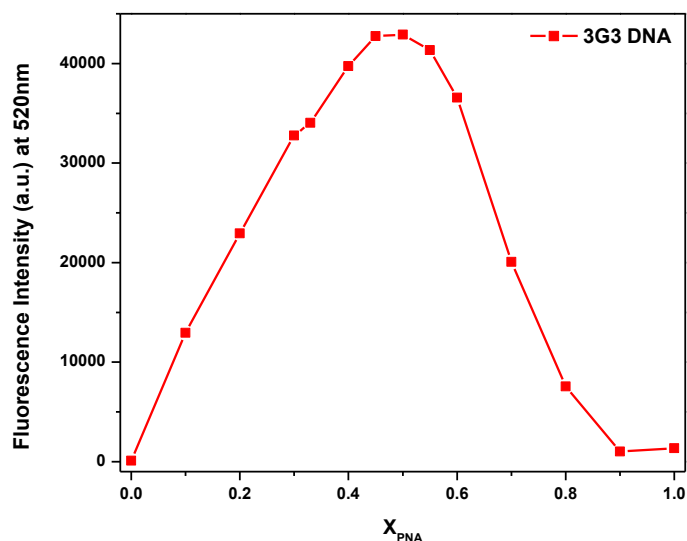
Figure 2.8 Continuous variation experiments (i.e. Job Plot) using PNA-TO with 3G3 DNA. 1:1 binding stoichiometry is indicated by a peak maxima at 0.5 X_{PNA}.

Continuous variation experiments further identified a 1:1 binding stoichiometry (Figure 2.8) and confirmed the 3+1 motif investigated by the Ladame lab.
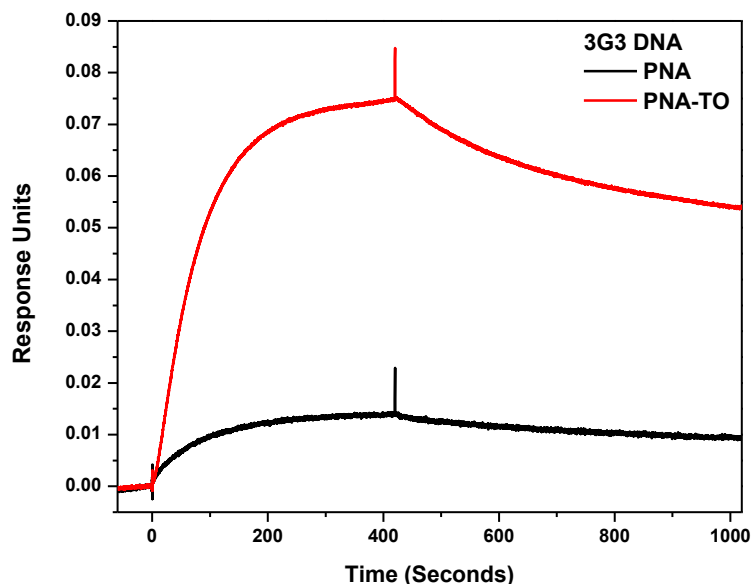
Figure 2.9 SPR sensorgrams of direct binding experiments with 20 nm PNA in 100 mM KCl SPR buffer. The increase in signal and on-rate in the presence of TO indicates an increase in kinetics.

Surface plasmon resonance (SPR) direct binding experiments were used to confirm the UV-Vis and CD results. SPR monitors the binding of PNA to DNA by passing the PNA probe over a DNA bound SPR chip. These experiments allow further evaluation of PNA-DNA GQ formation, as well as an initial on rate approximation. In addition to confirming binding between the DNA and PNA, we can compare the kinetic interaction between different PNAs. This could not be determined by our previous methods because thermal melting curves can only provide thermodynamic information, given that samples are preannealed.

Direct binding studies were conducted using 20 nm samples of each PNA in 1X SPR buffer (100 mM KCl). Response units are corrected to account for differences in molecular weight. Both unmodified PNAs (without and with TO) show an increase in response during the 420 second injection (time 0-420 seconds). Additionally, during the 600 second wash with buffer some PNA dissociated (time 420-1020 seconds). This association and dissociation indicates PNA binding to DNA-labeled surface. TO-labeled PNA had a significantly higher on-rate than its non-labeled counterpart (Figure 2.9). This indicates the TO has an impact on the kinetics in addition to the increased thermal stability observed in the previous thermal melting data.

2.3.3 *γ modified PNA does not significantly change stability in 3G3 DNA.*

Length of Nucleic Acid	Type of Nucleic Acid	PNA Modification
		No modification
3G3	DNA	γ backbone
4G3	RNA	+TO

Modifications to PNA have allowed scientists to fine tune their properties. Specifically, modifying the γ position on the backbone of PNA (γPNA) induces a right handed helical turn[24]. γPNA has significantly higher affinity to complementary nucleic acid because the pre-structured γPNA reduces the enthalpic penalty incurred by PNA. However, the helical pitch required for homologous quadruplex binding is not identical to that required for complementary binding, thus negating the benefit of γPNA.
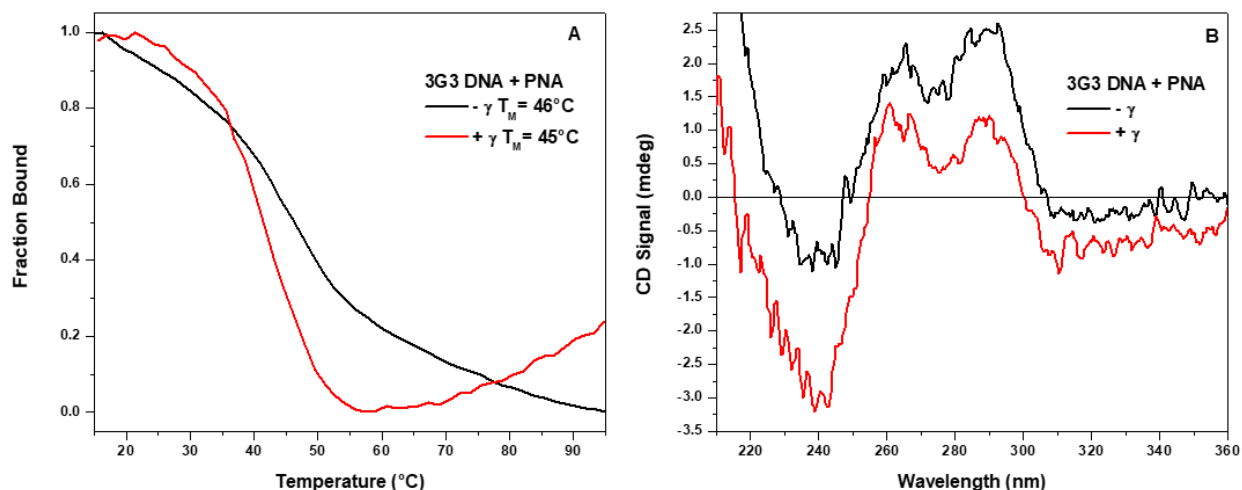
Figure 2.10 (A) Thermal Melting and (B) circular dichroism of 3G3 DNA + PNA with and without gamma modifications in 100 mM KCl Buffer. [3G3 DNA] = 1 µM [(γ)PNA] = 1 µM. Melting temperatures without gamma modification = 46 °C and with gamma modification = 45 °C. CD suggests no change in secondary structure with the gamma modification.

γPNA did not have a significant impact on the melting temperature, where γPNA showed a lower melting temperature than unmodified PNA by one degree (Figure 2.10 A). This indicated some enthalpic penalty due to the structural difference in helical pitch between duplexes and quadruplexes. Little difference is observed in the overall shape of CD spectra, indicating similar binding modes (Figure 2.10 B). However,

the overall signal intensity is lower which could be possibly be a result of the poor baseline (i.e. the difference in 320-340 nm). This difference was not observed in the background scan and therefore could not be removed from the spectra.
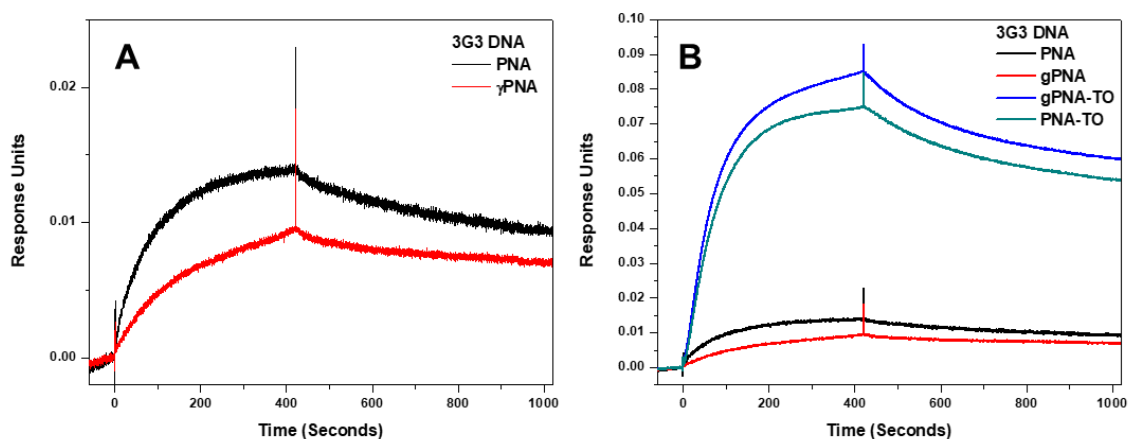
Figure 2.11 SPR sensorgrams of direct binding experiments with 20 nm PNA in 100 mM KCl SPR buffer. The presence of a gamma modification seems to hinder binding of PNA (A) and does not significantly change TO labeled PNAs (B).

Direct binding SPR studies were used to further understand the impact of γPNA on homologous GQ formation with 3G3 DNA. The SPR sensorgram shows γPNA has a slower on rate compared to unmodified PNA (Figure 2.11 A). γPNA must alter its preordered right handed helix prior to homologous binding, since the final formed GQ does not have the same helical pitch as duplex DNA. Therefore, the slower on rate could be attributed to the enthalpic penalty when γPNA unwinds to bind 3G3 DNA. However, this trend was not upheld comparing the PNA-TO to γPNA-TO (Figure 2.11 B). Both PNA-TO and γPNA-TO have much faster on rates than their unlabeled counterparts. This is due to the stabilizing effects of the TO, and is especially important in the case of γPNA since the presence of the TO negates the ethalpic penalty.

2.3.4 *PNA can invade GQ formed by 4G3 DNA. Trends with modifications stay constant.*

Length of Nucleic Acid	Type of Nucleic Acid	PNA Modification
		No modification
3G3	DNA	γ backbone
4G3	RNA	+TO
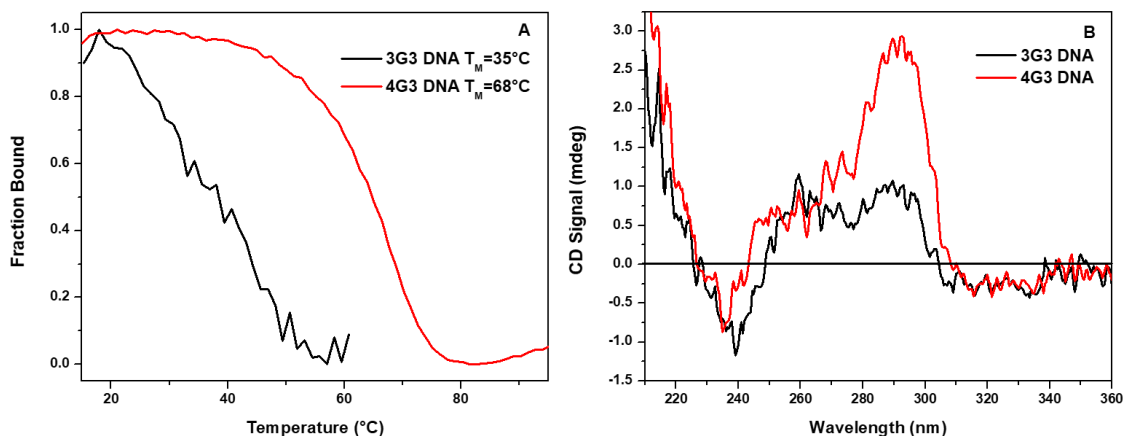
Figure 2.12 Thermal melting curves (A) and circular dichroism spectra (B) comparing the truncated 3G3 DNA to the full length 4G3 DNA. [3G3 DNA] = 1 μM [4G3 DNA] = 1 μM. Melting temperature of 3G3 DNA = 35 °C and 4G3 DNA = 68 °C. CD suggests a shift to antiparallel secondary structure from 3G3 DNA to 4G3 DNA.

4G3 DNA contains four repeating tracts of guanines and is able to form an intramolecular GQ without the addition of PNA and allows us to determine if the PNA probes can displace one of the GQ tracts to form a 3+1 GQ. 4G3 DNA alone shows a strong thermal melting transition at 68 °C (Figure 2.12 A) and *anti*-parallel GQ structure (Figure 2.12 B) by CD indicating the formation of an intramolecular GQ. This secondary structure is different from 3G3 DNA, which had primarily hybrid structure (Figure 2.12 B).
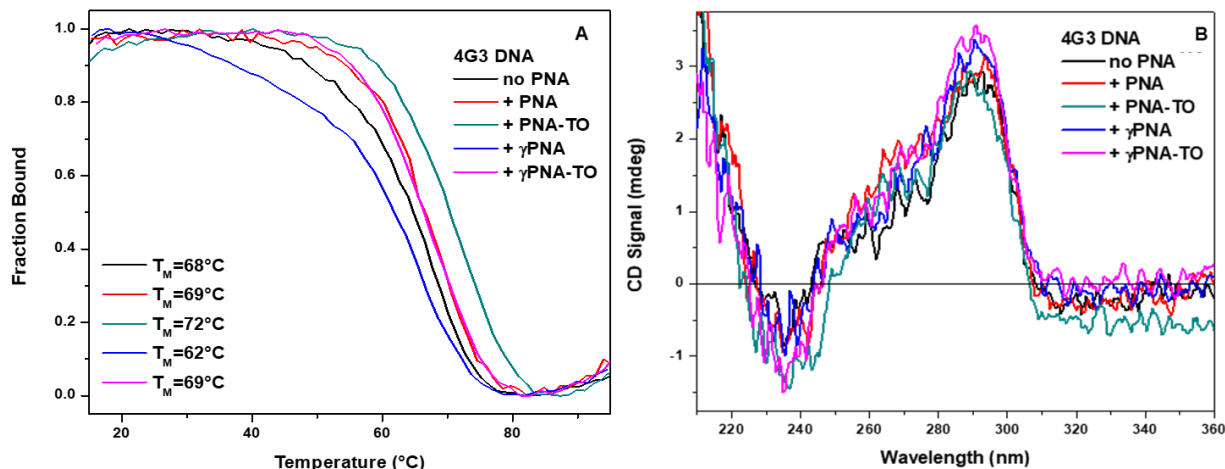
Figure 2.13 Thermal melting curves (A) and circular dichroism spectra (B) with 4G3 DNA and each binding PNA. [4G3 DNA] = 1 µM [(γ)PNA(-TO)] = 1 µM

Upon the addition of PNA to 4G3 DNA, the melting temperature changed by less than a degree (Figure 2.13 A). Again, primarily *anti*-parallel GQ structure is observed; however, there is an increase in the shoulder around 260 nm (indicative of parallel GQ). Unlike 3G3 DNA, no increase in CD signal was observed upon addition of PNA (Figure 2.13 B). This is likely due to the stability of the intramolecular GQ formed by 4G3 DNA.

PNA-TO consistently had higher melting temperatures when compared with the unlabeled PNA. 4G3 DNA + PNA-TO showed a ΔT_m of 6 °C compared to the addition of unlabeled PNA (while the 3G3 DNA showed a 22 °C ΔT_m) and confirmed binding of PNA to 4G3 DNA (Figure 2.13 A). The overall T_m increase, compared with the unlabeled PNA, is due to the stabilizing effect of TO. When TO dye is added to a sample of 4G3 DNA, a 5 °C increase in melting temperature is observed (Supporting Figure 2.2). The difference between the ΔT_m between the 4G3 and 3G3 DNA suggests that the TO better stabilizes the PNA complex with 3G3 DNA than 4G3 DNA. This could be because the overhanging GGG tract from the 4G3 DNA could interfere with TO stacking on the tetrad, allowing TO to bind more tightly to the 3G3 DNA.
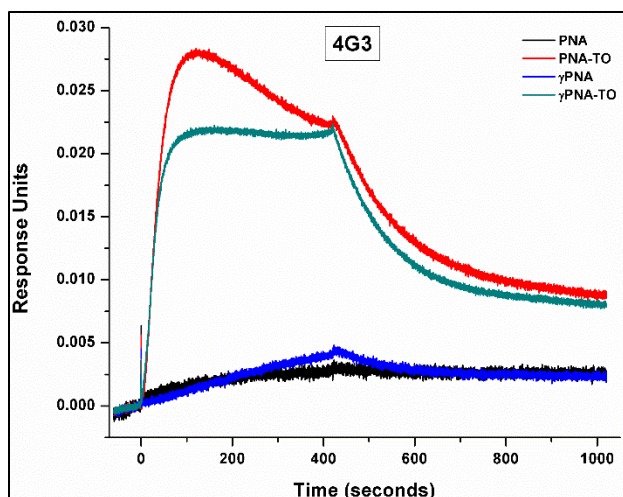
Figure 2.14 SPR sensorgrams of direct binding experiments with 20 nm PNA in 100 mM KCl SPR buffer.

SPR shows a much lower signal intensity in the 4G3 DNA than the 3G3 DNA (Figure 2.14). This was expected, since SPR experiments are performed at 25 °C in 100 mM KCl. Under these conditions, the 4G3 DNA likely forms a GQ and the PNA must invade the preformed GQ. Therefore, the kinetics of GQ breathing and PNA invasion play a more significant role than in the case of 3G3 DNA. However, the general trends observed for 3G3 DNA hold true. TO-labeled PNAs have the fastest binding, and there does not appear to be a difference in the initial rate between the gamma modified and unmodified PNA. There is a strange maxima in signal noted in the PNA-TO (and to a lesser extent in γPNA-TO) when bound to 4G3 DNA (Figure 2.14). The peak is primarily seen when a TO-labeled PNA is passed over 4G3 DNA. This increase in signal could be due to nonspecific binding of the TO to the DNA, which is then washed away over time, leading to the decrease in signal. Additionally, PNA-TO is able to aggregate in the system to a greater extent than γPNA-TO, since PNA-TO is more hydrophobic. This causes the maxima to be lower.

2.3.5   *PNA can form GQs with 3G3 and 4G3 RNA.*

Length of Nucleic Acid	Type of Nucleic Acid	PNA Modification
		No modification
3G3	DNA	γ backbone
4G3	RNA	+TO

Since we intend to use our PNA probes to target biological guanine-rich nucleic acids, we wanted to confirm that our PNAs are able to target both truncated (3G3) and full length (4G3) RNA GQs. RNA GQs are often the primary target sequence for therapeutic probes. This is because RNA is typically single stranded, in the cytoplasm, and there is a larger number of copies allowing for enhanced targeting. This contrasts with the double helix DNA, where the GQ competes with the double helix. Here, we observe the differences between the DNA and RNA targets, and targeting by PNA.
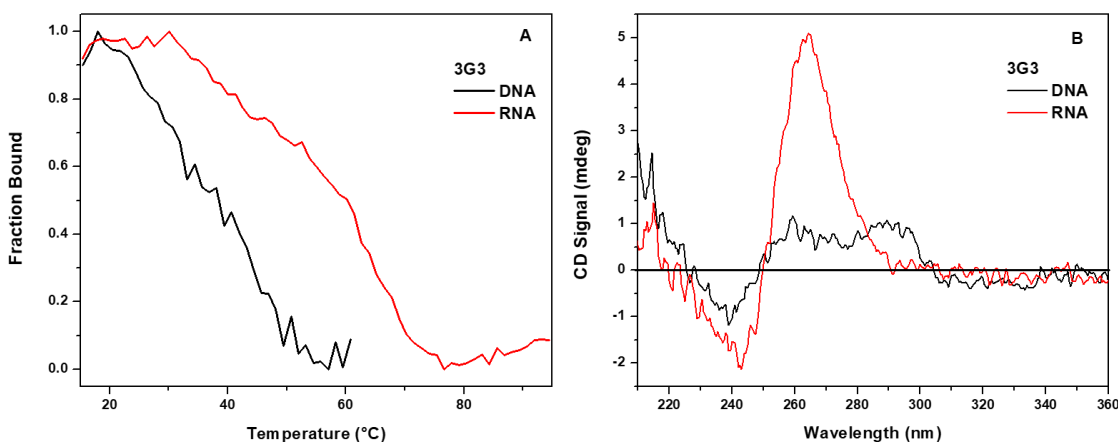
Figure 2.15 Thermal melting curves (A) and circular dichroism spectra (B) comparing 3G3 DNA and RNA. [3G3 DNA] = 1 µM [3G3 RNA] = 1 µM. Melting temperature of 3G3 DNA = 35 °C and 3G3 RNA = 62 °C. CD suggests a shift to parallel secondary structure in the presence of 3G3 RNA compared to 3G3 DNA.

Thermal melting curves of 3G3 RNA (Figure 2.15 A) showed a T_M of 62 °C, whereas the DNA showed a very weak melting transition around 35 °C. Even though the melting temperature increases by 27 °C using RNA, the melting transition is quite broad. This indicates that while the thermal stability increased, there is still some difficulty in forming the intermolecular GQ. The formation of the GQ is also supported

by CD (Figure 2.15 B). Parallel secondary structure is observed, which is expected due to the steric imposed by the 2ʹ OH[35].
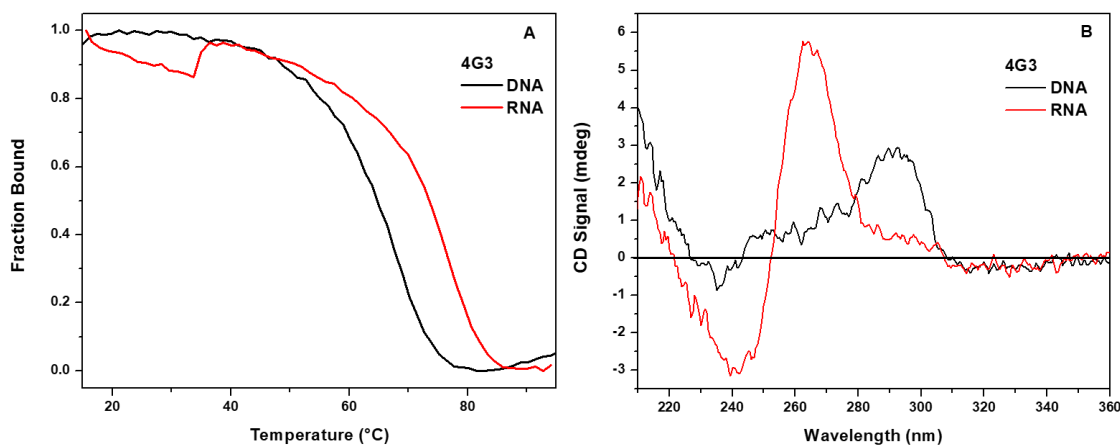
Figure 2.16 Thermal melting curves (A) and circular dichroism spectra (B) comparing 4G3 DNA and RNA. [4G3 DNA] = 1 µM [4G3 RNA] = 1 µM. Melting temperature of 4G3 DNA = 68 °C and 4G3 RNA = 76 °C. CD suggests a shift to parallel secondary structure in the presence of 4G3 RNA compared to 4G3 DNA.

Utilizing the full length telomeric RNA (4G3) showed an 8 °C increase in melting temperature over DNA (68 °C for 4G3 DNA and 76 °C for 4G3 RNA) (Figure 2.16 A). These thermal melting curves are generally smoother and have narrower melting transitions than their truncated counterparts, observed previously. Again, a shift from antiparallel secondary structure in the DNA to a parallel secondary structure in the RNA is observed (Figure 2.16 B).
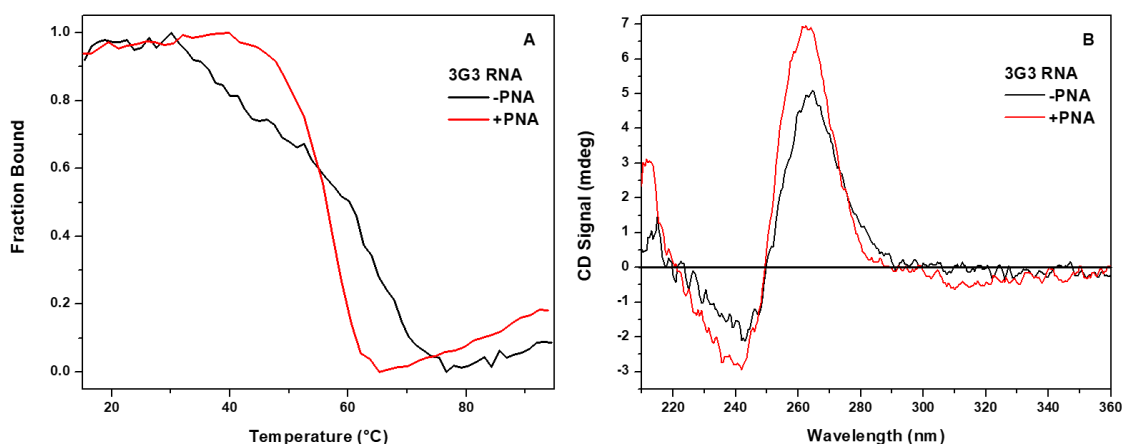
Figure 2.17 Thermal melting curves (A) and circular dichroism spectra (B) comparing 3G3 RNA with and without PNA. [3G3 RNA] = 1 µM [PNA] = 1 µM. Melting temperature without PNA = 61 °C and with PNA = 58 °C. CD suggests parallel secondary structure for both GQs.

In the presence of PNA, 3G3 RNA shows a narrow transition indicative of GQ formation (Figure 2.17 A). Although this transition has a lower melting temperature than 3G3 RNA alone (58 °C vs 61 °C), the narrowing of the transition indicates binding between the PNA and RNA. CD also supports PNA binding by the slight increase in parallel GQ signal (Figure 2.17 B).
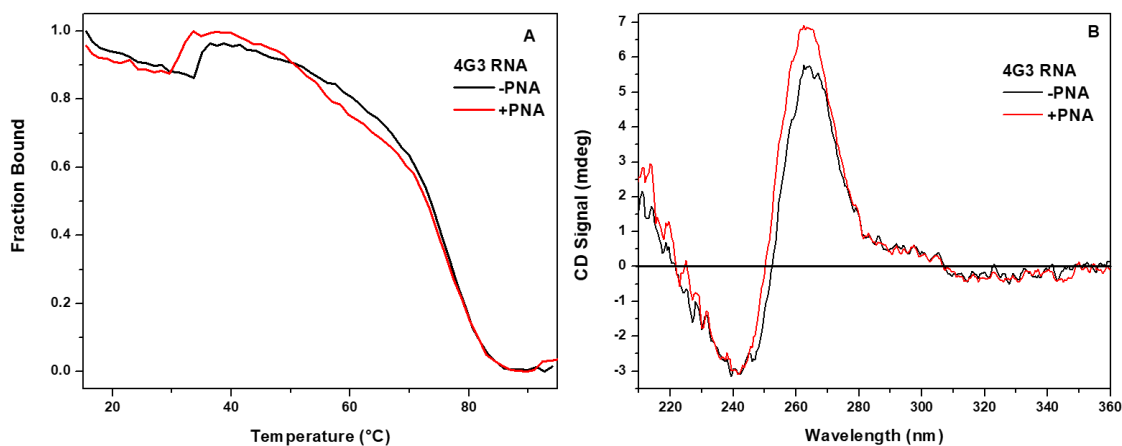
Figure 2.18 Thermal melting curves (A) and circular dichroism spectra (B) comparing 4G3 RNA with and without PNA. [4G3 RNA] = 1 µM [PNA] = 1 µM. Melting temperature without PNA = 76 °C and with PNA = 76 °C. CD suggests parallel secondary structure for both GQs.

46

4G3 RNA showed no difference in melting temperature upon addition of PNA (76 °C) (Figure 2.18 A). This neither confirmed nor denied the binding between PNA and 4G3 RNA. Furthermore, only a slight increase was observed in the CD spectra (Figure 2.18 B), which could indicate binding; however, we wanted further evidence of a binding event. Job plot experiments using PNA-TO showed 1:1 binding interaction (Supporting Figure 2.3), confirming the ability of our TO labeled PNAs to invade 4G3 RNA.

2.3.6 *Cell Lysate experiments indicate PNA binding; however, low selectivity is a concern.*

We then examined the implications of our short PNA probes in the context of a luciferase reporter assay. This assay has been used previously to evaluate the PNA's ability to enhance or inhibit protein expression[25, 26]. We use a puc19 DNA plasmid, which contains the GQ sequence in the 5′-UTR of the firefly luciferase gene. The linearized plasmid is transcribed into mRNA. After incubating the target mRNA with a specific concentration of PNA, the mRNA is translated. The amount of firefly luciferase is determined by the production of light when luciferin is added.

The luciferase assay experiment was run with three different DNA plasmids. They are 3G3, 4G3, and noG3, and the sequences that are inserted into the 5'-UTR of the firefly luciferase gene can be seen in Table 2.4. Full sequences are shown in Supporting Figure 2.4 - Supporting Figure 2.6.

Dose response curves were run with PNA concentrations varying from 0-500 nm PNA. Each plasmid was run with PNA at 0, 100, 250, 300, 450, and 500 nm PNA concentrations. Luciferase activity is converted to percent relative light units (%RLU) where the 0 nm PNA sample is set to 100%.
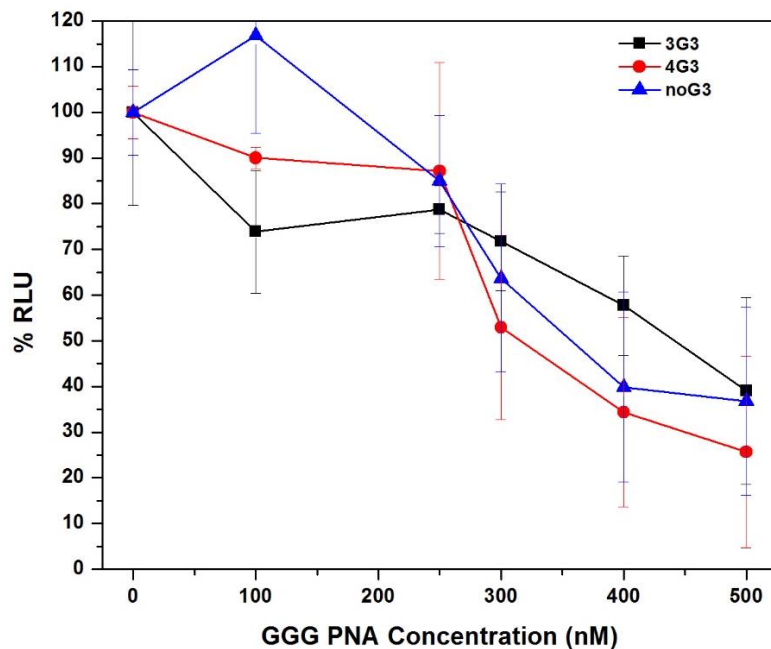
Figure 2.19 Dose response curve for 3G3 (black), 4G3 (red), and noG3 (blue) with GGG PNA (0, 100, 250, 300, 450, and 500 nm PNA). (n=3)

Results from the dose response curve (Figure 2.19) indicate no significant difference in knockdown between the different plasmids (3G3, 4G3, and noG3) with unmodified PNA. 3G3 shows significant knockdown (approximately 60%) however the 4G3 and noG3 also show approximately 60-75% knockdown. This indicates knockdown is potentially due to off-target effects and not to GQ formation. This result is not surprising, as short probes have significantly reduced selectivity. In order to selectively target the GQ sequence in the 5'UTR of the firefly luciferase gene, the probe will need to be altered.

Using our modified PNAs (PNA-TO and γPNA) did not improve our selectivity issue (Figure 2.20). In Figure 2.20, we show a single dose (500 nm) of each PNA for all three plasmids. PNA-TO and γPNA showed 0-20% knockdown in the 3G3 and 4G3 plasmid. These probes also had 50-60% knockdown in the noG3 plasmid. This discrepancy indicates both PNA-TO and γPNA have off target binding sites. Overall, PNA had the most significant knockdown, with over 90% knockdown in the 3G3 plasmid and 75% knockdown in the 4G3 plasmid. However, the significant knockdown observed in the noG3 plasmid indicates this is due to off target binding. Upon inspection of the luciferase gene, over 10 different cytosine

rich binding sites were identified as potential off target binding sites (shown as light blue in Supporting Figure 2.4 - Supporting Figure 2.6).
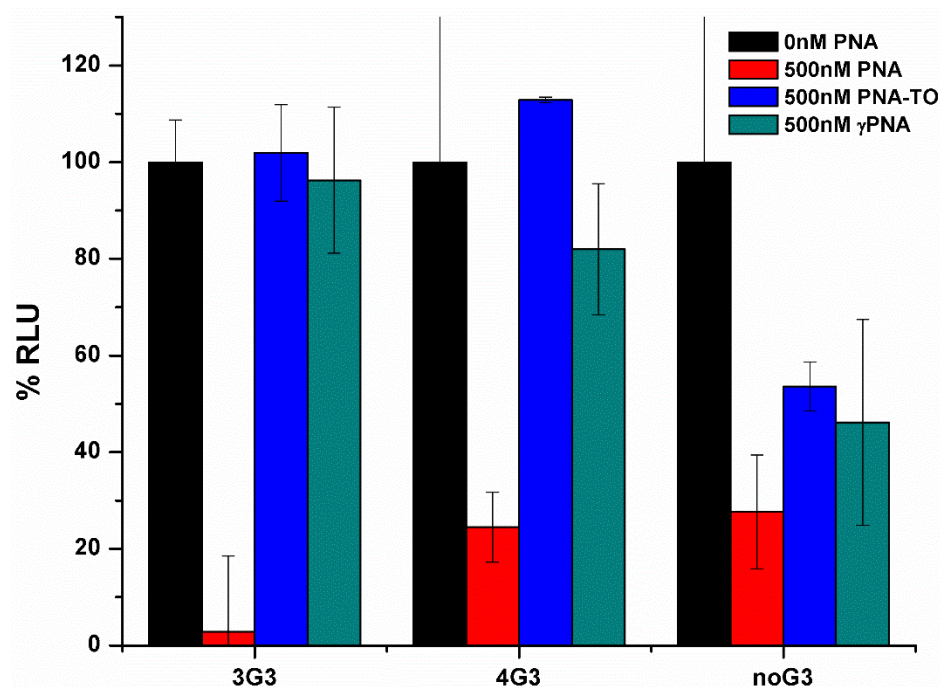
Figure 2.20 Dose response for 3G3 (left), 4G3 (middle), and noG3 (right) with no PNA (black), 500 nm PNA (red), 500 nm PNA-TO (blue), 500 nm γPNA (green). (n=3)

2.4   Discussion

This project uses short guanine-rich PNA probes to bind and target GQs. We target two telomere-based sequences: The first is a truncated sequence, called 3G3, which only contains three guanine tracks and requires the addition of PNA in order to form a stable GQ. This sequence is similar to that used in Ladame's approach[32]. The second is a full length sequence (called 4G3) which contains four guanine tracks, and upon the addition of PNA must displace a guanine track in order to form a PNA-DNA GQ. Our PNA probes also vary in design. We combined our homologous binding PNA with a small molecule ligand by linking thiazole orange (TO) to our PNA. We further incorporate backbone modifications to the gamma position in order to preorganize the PNA. The biophysical characterization of these probes will assist in the

design of probes for targeting telomeric and other GQ systems *in vivo* to treat a wide variety of diseases where GQs play key roles.

We first characterized the addition of unmodified PNA to each of our DNA telomere targets (3G3 DNA and 4G3 DNA). Thermal melting curves showed 3G3 DNA alone was able to form a weak bimolecular GQ (Figure 2.5) and the PNA-3G3 DNA GQ showed a significantly increased melting temperature. Circular dichroism also confirmed the incorporation of PNA, showing antiparallel secondary structure with an overall increase in signal upon the addition of PNA (Figure 2.6). Using a target nucleic acid that has the ability to form a GQ by itself (i.e. 4G3 DNA) was important for this project as both a comparison to the 3G3 DNA and to observe the PNAs ability to invade an existing GQ. This made determining if PNA bound more difficult, since the difference in melting temperature and secondary structure changed insignificantly (Figure 2.13). The SPR sensorgrams were crucial in identifying the binding of each PNA to 4G3 DNA (Figure 2.14). Even though the SPR signal was significantly lower than 3G3 DNA due to the required unwinding of a preexisting GQ, the binding of each PNA is observed with similar trends to that of 3G3 DNA.

Thiazole orange was selected as a light up probe for our PNAs. The presence of TO, compared to an unmodified PNA, significantly increased the thermal stability of the GQ (Figure 2.7 A and Figure 2.13 A) and greatly improved the overall structure of the transition. Although TO was expected to alter the thermal stability of the GQ[36], there was not any precedent for whether the secondary structure would change. We do observe a slight decrease in the parallel signal in the case of the 3G3 DNA target, indicating a possible change in the overall secondary structure (Figure 2.7 B). This secondary structure change does not significantly impact the formation of the PNA-DNA GQ, but it is worth noting the change in secondary structure if there are concerns in future *in vivo* experiments regarding altering native secondary structure. Direct binding surface plasmon resonance (SPR) experiments further confirmed/added to the thermal stability and secondary structure data. These experiments indicated TO plays a significant role in increasing the initial on rate of the PNA to 3G3 and 4G3 DNA.

Gamma modifications to PNA (γPNA) have significantly increased binding affinity to complementary nucleic acid targets by preorganizing the PNA into a right handed structure. We wanted to identify what role the gamma modification would play in targeting telomeric nucleic acid. Overall, γPNA did not contribute any significant changes to the thermal stability or secondary structure (Figure 2.10). SPR sensorgrams indicated that the gamma modification decreases the initial on rate in systems without TO, but no change in on rate was observed in systems with TO. These results indicate that while the gamma modification improves stability for complementary binding probes, it does not significantly modify stability of homologous binding probes. This is in agreement with a previously studied 2+2 PNA:GQ where right handed gamma PNA was shown to have little impact on the stability of heteroquadruplexes[29]. Further, left handed gamma modified PNA was shown to improve selectivity for homologous binding[29]. Utilizing these types of probes slightly destabilizes the PNA:DNA heteroquadruplex, but the benefit of enhanced selectivity could improve *in vivo* work where there are significantly more off-target binding sites for a short homologous PNA.

In addition to targeting telomeric DNA, we tested our PNA probes' ability to target RNA GQs. We utilized telomeric RNA, called TERRA, using the same truncated (3G3) and full length (4G3) system. RNA GQs, both 3G3 and 4G3, showed melting temperatures that were significantly higher than their DNA counterparts (Figure 2.15 and Figure 2.16). Nearly all RNA GQs show parallel secondary structure, so the shift in hybrid secondary structure in DNA to parallel secondary structure in RNA was expected. The addition of PNA was more difficult to observe for the RNA, since little change in melting temperature and no change in secondary structure were observed (Figure 2.17 and Figure 2.18). Melting hypochromicity changes were able to confirm binding of PNA to 3G3 RNA (Figure 2.17). However, Job plots were able to confirm TO-labeled PNA binding to each RNA (Supporting Figure 2.3).

We utilize cell lysate luciferase assays to evaluate the PNA probes' ability to impact translation. The sequences of interest, in this case 3G3 and 4G3, were incorporated into the 5′-UTR of the luciferase gene on a puc19 plasmid. Following linearization and transcription to RNA, we incubated our PNA in cell lysate. We can then observe the impact of the PNA on gene expression by measuring the light production

51

of luciferase in the presence of luciferin. Unmodified PNA successfully knocked down gene expression in 3G3 and 4G3; however, there was also significant knockdown in the noG3 control (Figure 2.19). Contrary to our previous biophysical results, PNA-TO (and γPNA) did not improve the activity in these luciferase experiments (Figure 2.20). Further analysis of the luciferase promoter and gene sequence indicated our short PNA probes are likely not selective for our 3G3 and 4G3 target. This selectivity concern can hopefully be alleviated by the incorporation of a second binding mode to the PNA probe.

2.5   Future Directions

The project discussed here characterized short homologous binding PNAs. These PNAs contained a variety of modifications, including a combination of homologous and small molecule binding modes (Scheme 2.1) in order to improve binding. While it appears these probes are very effective (i.e. binding to the target), they are not selective enough to target a single GQ region on a gene sequence. This is not surprising since these probes are relatively short in length. Shorter probes have a higher probability for off-target effects, since there is an increased likelihood for a matching sequence elsewhere in the gene or genome. These off-target effects are seen in the firefly luciferase experiments. Longer probes have increased affinity; however, mismatches are more easily tolerated, reducing selectivity.

Improvement in probe selectivity could be obtained by incorporating an additional recognition domain (complementary binding in Scheme 2.1) to the probe. This probe could utilize two parts: a duplex-forming sequence and a GQ sequence (Scheme 2.2). This combination of recognition domains has been studied using a DNA probe to target an mRNA transcript[37]. Basu and colleagues designed a 22mer DNA probe containing 15 bases complementary to bases adjacent to the GQ sequence and 7
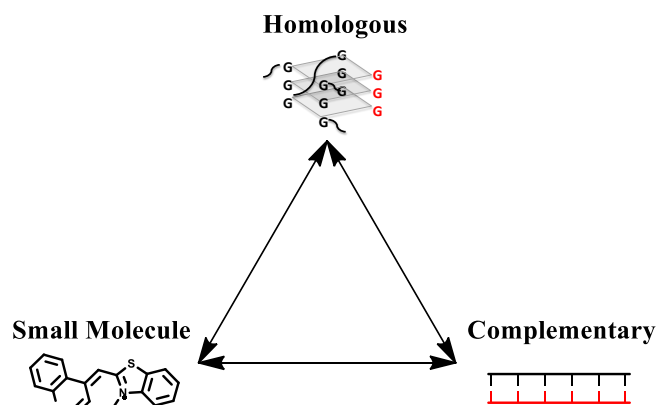
Scheme 2.1 Binding method utilized in probe design.

bases capable of forming a heteroquadruplex[37]. Further, DNA modifications made the DNA probe resistant to nucleases and was used to study protein expression in cancer cells[37]. Using PNA would allow inherent nuclease resistance, permit shorter sequences due to enhanced affinity, and improve selectivity by incorporating backbone modifications in one or both domains.
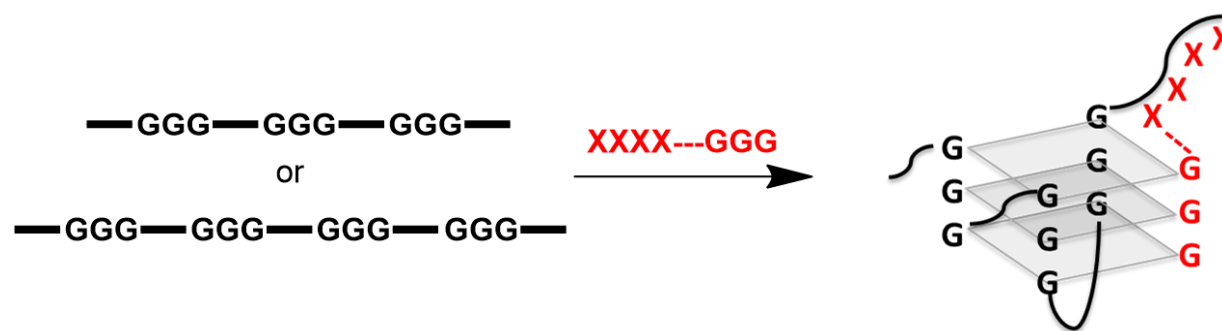
Scheme 2.2 Potential probe design using complementary and homologous binding.

2.6 References

(1)   Parkinson, G. N. Quadruplex Nucleic Acids. In *Quadruplex Nucleic Acids*; Neidle, S., Balasubramanian, S., Eds.; Royal Society of Chemistry: Cambridge, 2006; pp 1–30.

(2)   Phillips, K.; Dauter, Z.; Murchie, A. I. H.; Lilley, D. M. J.; Luisi, B. The Crystal Structure of a Parallel-Stranded Guanine Tetraplex at 0.95Å Resolution. *J. Mol. Biol.* **1997**, *273* (1), 171–182.

(3)   Hardin, C. C.; Watson, T.; Corregan, M.; Bailey, C. Cation-Dependent Transition between the Quadruplex and Watson-Crick Hairpin Forms of d(CGCG3GCG). *Biochemistry* **1992**, *31* (3), 833–841.

(4)   Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. Quadruplex DNA: Sequence, Topology and Structure. *Nucleic Acids Res.* **2006**, *34* (19), 5402–5415.

(5)   Roe, S.; Ritson, D. J.; Garner, T.; Searle, M.; Moses, J. E. Tuneable DNA-Based Asymmetric Catalysis Using a G-Quadruplex Supramolecular Assembly. *Chem. Commun.* **2010**, *46* (24), 4309.

(6)   Huppert, J. L.; Balasubramanian, S. G-Quadruplexes in Promoters throughout the Human Genome. *Nucleic Acids Res.* **2007**, *35* (6), 2105–2105.

(7)   Huppert, J. L.; Bugaut, A.; Kumari, S.; Balasubramanian, S. G-Quadruplexes: The Beginning and End of UTRs. *Nucleic Acids Res.* **2008**, *36* (19), 6260–6268.

(8)   Lipps, H. J.; Rhodes, D. G-Quadruplex Structures: In Vivo Evidence and Function. *Trends Cell Biol.* **2009**, *19* (8), 414–422.

(9)   Phan, A. T. Human Telomeric G-Quadruplex: Structures of DNA and RNA Sequences. *FEBS Journal*. 2010, pp 1107–1117.

(10) Dahse, R.; Fiedler, W.; Ernst, G. Telomere Und Telomerase. *Pathologe* **1997**, *18* (6), 425–429.

(11) Neidle, S. Quadruplex Nucleic Acids as Novel Therapeutic Targets. *J. Med. Chem.* **2016**, *59* (13), 5987–6011.

(12) Oganesian, L.; Bryan, T. M. Physiological Relevance of Telomeric G-Quadruplex Formation: A Potential Drug Target. *BioEssays* **2007**, *29* (2), 155–165.

(13) Wang, C.; Zhao, L.; Lu, S. Role of TERRA in the Regulation of Telomere Length. *Int. J. Biol. Sci.* **2015**, *11* (3), 316–323.

(14) Rippe, K.; Luke, B. TERRA and the State of the Telomere. *Nat. Struct. Mol. Biol.* **2015**, *22* (11), 853–858.

(15) Dilley, R. L.; Greenberg, R. A. ALTernative Telomere Maintenance and Cancer. *Trends in cancer* **2015**, *1* (2), 145–156.

(16) Millevoi, S.; Moine, H.; Vagner, S. G-Quadruplexes in RNA Biology. *Wiley Interdiscip. Rev. RNA* **2012**, *3* (4), 495–507.

(17) Bugaut, A.; Balasubramanian, S. 5'-UTR RNA G-Quadruplexes: Translation Regulation and Targeting. *Nucleic Acids Res.* **2012**, *40* (11), 4727–4741.

(18) Oyaghire, S. N.; Cherubim, C. J.; Telmer, C. A.; Martinez, J. A.; Bruchez, M. P.; Armitage, B. A. RNA G-Quadruplex Invasion and Translation Inhibition by Antisense γ-Peptide Nucleic Acid Oligomers. *Biochemistry* **2016**, *55* (13), 1977–1988.

(19) Canady, T. D.; Telmer, C. A.; Oyaghire, S. N.; Armitage, B. A.; Bruchez, M. P. In Vitro Reversible Translation Control Using ΓpNA Probes. *J. Am. Chem. Soc.* **2015**, *137* (32), 10268–10275.

(20) Morris, M. J.; Negishi, Y.; Pazsint, C.; Schonhoft, J. D.; Basu, S. An RNA G-Quadruplex Is Essential for Cap-Independent Translation Initiation in Human VEGF IRES. *J. Am. Chem. Soc.* **2010**, *132* (50), 17831–17839.

(21) Krishnan, Y.; Simmel, F. C. Nucleic Acid Based Molecular Devices. *Angewandte Chemie - International Edition*. John Wiley & Sons, Ltd March 28, 2011, pp 3124–3156.

(22) Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. Sequence-Selective Recognition of DNA by Strand Displacement with a Thymine. *Source Sci. New Ser.* **1991**, *254* (5037), 1497–1500.

(23) Panyutin, I. G.; Onyshchenko, M. I.; Englund, E. A.; Appella, D. H.; Neumann, R. D. Targeting DNA G-Quadruplex Structures with Peptide Nucleic Acids. *Curr. Pharm. Des.* **2012**, *18* (14), 1984–1991.

(24) Dragulescu-Andrasi, A.; Rapireddy, S.; Frezza, B. M.; Gayathri, C.; Gil, R. R.; Ly, D. H. A Simple γ-Backbone Modification Preorganizes Peptide Nucleic Acid into a Helical Structure. *J. Am. Chem. Soc.* **2006**, *128* (31), 10258–10267.

(25) Svanvik, N.; Westman, G.; Wang, D.; Kubista, M. Light-Up Probes: Thiazole Orange-Conjugated Peptide Nucleic Acid for Detection of Target Nucleic Acid in Homogeneous Solution. *Anal. Biochem.* **2000**, *281* (1), 26–35.

(26) Silva, G. L.; Ediz, V.; Yaron, D.; Armitage, B. A. Experimental and Computational Investigation of Unsymmetrical Cyanine Dyes: Understanding Torsionally Responsive Fluorogenic Dyes. *J. Am. Chem. Soc.* **2007**, *129* (17), 5710–5718.

(27) Nygren, J.; Svanvik, N.; Kubista, M. The Interactions between the Fluorescent Dye Thiazole Orange and DNA. *Biopolymers* **1998**, *46* (1), 39–51.

(28) Datta, B.; Schmitt, C.; Armitage, B. A. Formation of a PNA2-DNA2 Hybrid Quadruplex. *J. Am. Chem. Soc.* **2003**, *125* (14), 4111–4118.

(29) Lusvarghi, S.; Murphy, C. T.; Roy, S.; Tanious, F. A.; Sacui, I.; Wilson, W. D.; Ly, D. H.; Armitage, B. A. Loop and Backbone Modifications of Peptide Nucleic Acid Improve G-Quadruplex Binding Selectivity. *J. Am. Chem. Soc.* **2009**, *131* (51), 18415–18424.

(30) Roy, S.; Zanotti, K. J.; Murphy, C. T.; Tanious, F. A.; Wilson, W. D.; Ly, D. H.; Armitage, B. A. Kinetic Discrimination in Recognition of DNA Quadruplex Targets by Guanine-Rich Heteroquadruplex-Forming PNA Probes. *Chem. Commun.* **2011**, *47* (30), 8524–8526.

(31) Murphy, C. T.; Gupta, A.; Armitage, B. A.; Opresko, P. L. Hybridization of G-Quadruplex-Forming Peptide Nucleic Acids to Guanine-Rich Dna Templates Inhibits DNA Polymerase η Extension. *Biochemistry* **2014**, *53* (32), 5315–5322.

(32) Paul, A.; Sengupta, P.; Krishnan, Y.; Ladame, S. Combining G-Quadruplex Targeting Motifs on a Single Peptide Nucleic Acid Scaffold: A Hybrid (3 + 1) PNA-DNA Bimolecular Quadruplex. *Chem. - A Eur. J.* **2008**, *14* (28), 8682–8689.

(33) Braasch, D. A.; Nulf, C. J.; Corey, D. R. Synthesis and Purification of Peptide Nucleic Acids. *Curr. Protoc. Nucleic Acid Chem.* **2002**, *9* (1), 4.11.1-4.11.18.

(34) Mergny, J. L.; Phan, A. T.; Lacroix, L. Following G-Quartet Formation by UV-Spectroscopy. *FEBS Lett.* **1998**, *435* (1), 74–78.

(35) Fay, M. M.; Lyons, S. M.; Ivanov, P. RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms. *J. Mol. Biol.* **2017**, *429* (14), 2127–2147.

(36) Chilka, P.; Desai, N.; Datta, B. Small Molecule Fluorescent Probes for G- Quadruplex Visualization as Potential Cancer Theranostic Agents. *Molecules* **2019**, *24* (4).

(37) Bhattacharyya, D.; Nguyen, K.; Basu, S. Rationally Induced RNA:DNA G-Quadruplex Structures Elicit an Anticancer Effect by Inhibiting Endogenous EIF-4E Expression. *Biochemistry* **2014**, *53* (33), 5461–5470.

2.7   Supporting Figures
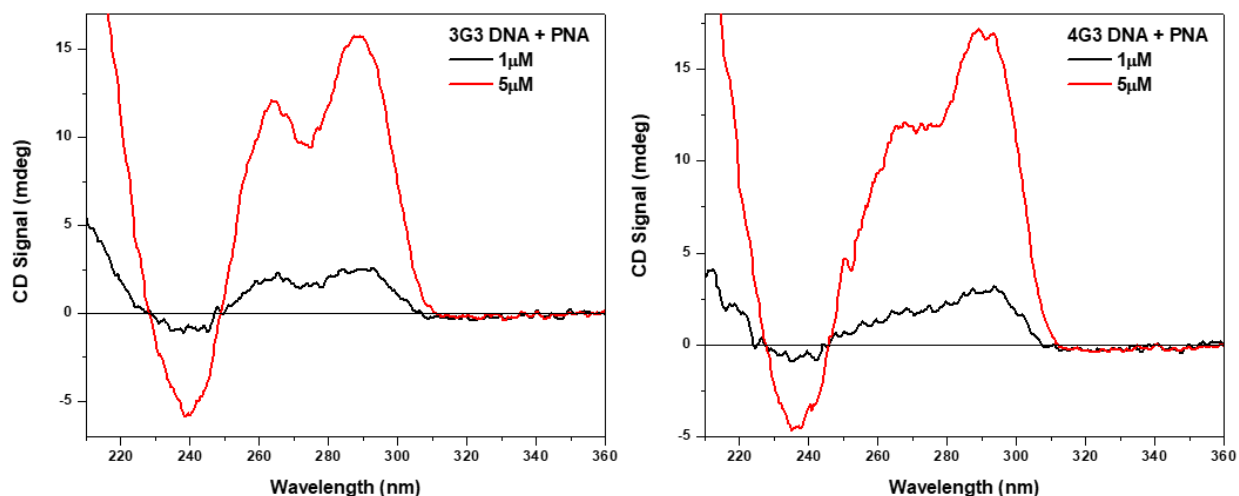
Supporting Figure 2.1 CD Spectra comparing 1µM and 5µM DNA-PNA. These spectra indicate hybrid secondary structure for both concentrations in 3G3 DNA (left) and antiparallel secondary structure for both concentrations in 4G3 DNA (right).
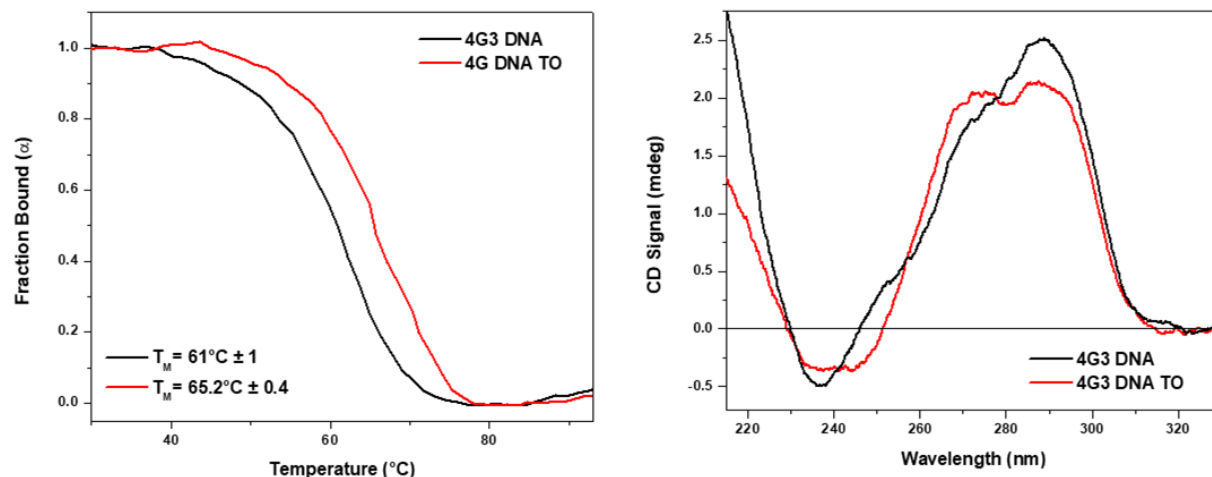
Supporting Figure 2.2 Thermal melting curves (left) and circular dichroism (right) of 4G3 DNA with (red) and without (black) added thiazole orange. Experiments contained 1μM 4G3 DNA and 1μM TO.
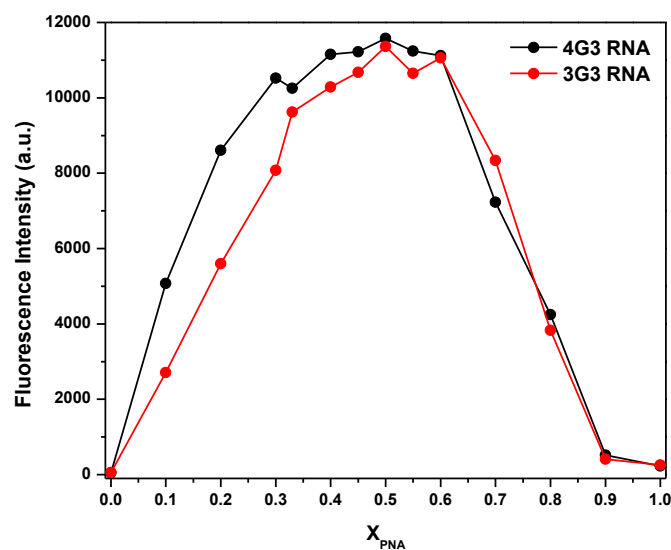
Supporting Figure 2.3 Continuous variation experiments (i.e. Job Plot) using PNA-TO with 3G3 and 4G3 RNA. 1:1 binding stoichiometry is indicated.

Supporting Figure 2.4 3G3 DNA linearized sequence. Shown in light blue and green are potential C-rich and G-rich binding sites, respectively. Other colored regions: base 1 (red) is the last base in the T7 promoter, base 8 (dark blue) is the HindIII restriction cut site, base 28 (lavender) is the 3G3 site, base 56 (yellow) is the firefly luciferase gene.

Supporting Figure 2.5 4G3 DNA linearized sequence. Shown in light blue and green are potential C-rich and G-rich binding sites, respectively. Other colored regions: base 1 (red) is the last base in the T7 promoter, base 8 (dark blue) is the HindIII restriction cut site, base 28 (light green) is the 4G3 site, and base 62 (yellow) is the firefly luciferase gene.

Supporting Figure 2.6 noG3 DNA linearized sequence. Shown in light blue (with the exception of bases 14-21 and 28-34) and green are potential C-rich and G-rich binding sites, respectively. Other colored regions: base 1 (red) is the last base in the T7 promoter, base 8 (dark blue) is the HindIII restriction cut site, bases 14-34 make up the CMV transcript (light blue) and NheI (pink), and base 48 (yellow) is the firefly luciferase gene.

3   Chapter 3. RNA:DNA Heteroquadruplexes (RDQs) Formed Using a Duplex Scaffold: Analyzing Biophysical Differences Between Sequences and Targeting Using Peptide Nucleic Acid (PNA)

3.1   Introduction

3.1.1   Brief Introduction to RNA:DNA Heteroquadruplexes (RDQs)

This thesis focuses on the non-canonical secondary structure formed in guanine-rich nucleic acid, called guanine quadruplexes (GQs). One newly identified variation on the GQ motif known as an RNA-DNA heteroquadruplex (RDQ) utilizes both RNA and DNA in order to form the quadruplex. RDQs form in a similar manner to DNA and RNA bimolecular GQs (shown as blue and red in tetrad in Scheme 3.1), where two nucleic acid strands assemble to form a GQ. These guanine-rich sequences assemble into tetrads (Scheme 3.1), where four guanine bases interact by hydrogen bonds from both Watson-Crick and Hoogsteen faces[1]. Multiple tetrads can align atop one another by π-π stacking to form the quadruplex. Little is known about RDQ secondary structure compared to the vast body of knowledge surrounding other GQ motifs[2,3]. In this chapter, we aim to further understand the structural and biophysical properties of RDQs.
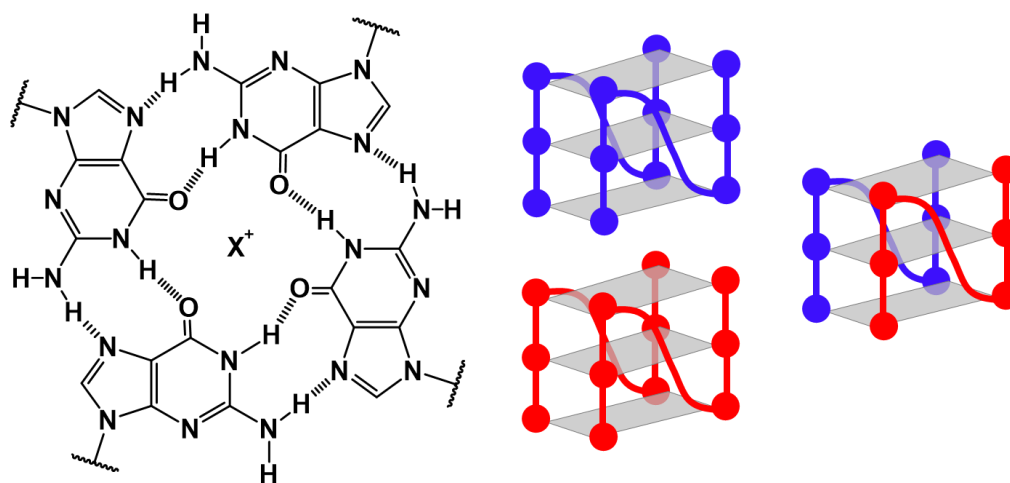
Scheme 3.1 Basic structure of a G-tetrad (left), DNA and RNA bimolecular GQs (center), and RNA:DNA HeteroGQ (right).

G quadruplexes (GQs) arise from guanine-rich DNA or RNA sequences, which assemble/fold in order to form stacking G-tetrads (Scheme 3.1, left). GQ sequences require four tracts of guanines in order to form. However, these tracts need not be present on a single nucleic acid strand. Bimolecular GQs made of RNA or DNA can form (Scheme 3.1, middle) with two strands of R/DNA which collectively contain four tracts of guanine. Additionally, bimolecular hetero-GQs made of both RNA and DNA can form (Scheme 3.1, right). There are three different possible RNA:DNA combinations: 1:3, 2:2, and 3:1 where the digits refer to the number of G-tracts provided by the RNA and DNA strands, respectively. The growing biological interest in RDQs motivated us to develop model systems with which to study their structural and biophysical properties.

3.1.2   Proposed Biological Functions of RDQs

RDQs can occur when guanine-rich DNA and RNA are in close proximity. In order for DNA and RNA to interact in the cell, specific spatial and temporal conditions must be met. One such example which meets these conditions occurs during transcription. In particular, when the DNA template strand is cytosine-rich, the RNA transcript and DNA non-template strand will both be guanine-rich. Therefore, recently transcribed RNA has the potential to interact with the unwound non-template DNA to form an RDQ. Furthermore, while the guanine-rich nature of the template and non-template strands is comparable throughout transcription start sites, the non-template strand is guanine-rich downstream of the start site[4–6]. Bioinformatic analysis suggests RDQs forming at transcription start sites could be significantly more common than RNA and DNA GQs, in part because of the relaxed sequence requirements (only needing two guanine tracts instead of four)[7]. Recently, the conserved sequence block II (CSB II) region in human mitochondrial DNA and its associated RNA transcript have been proposed to form an RDQ that serves as a switch between transcription and DNA replication[8–11].

A second proposed biological RDQ uses telomeric repeat-containing RNA (TERRA). Telomeres comprise the ends of chromosomes and serve to protect coding regions from degradation[12]. TERRA is the

non-coding G-rich RNA transcript of telomeres, and it is proposed to perform a variety of functions including telomere length regulation[13]. The interaction between TERRA and telomeric DNA has long been speculated, however this interaction has primarily focused on the formation of R-loops[14,15]. The study of RDQs using telomeres and TERRA is considered an additional protective measure for telomeres[16–18].

3.1.3   Methods for Forming RDQs in vitro

In order to analyze RDQs *in vitro*, scientists have developed scaffolds to assist formation and mimic the forced proximity of naturally formed RDQs. One example of a scaffold which enables specific control for the formation of RDQs utilizes a cyclic peptidic scaffold, called a template-assembled synthetic g-quadruplex (TASQ). This design, developed by the Defrancq lab, allows for precise control over the secondary structure of the quadruplex of interest and has been used to study a wide variety of quadruplexes including bimolecular DNA GQs as well as tetramolecular DNA, RNA, and RDQs[19–21]. TASQs were initially developed to study the interaction between small molecule ligands and constrained GQ topologies. TASQs were recently used to study a constrained parallel RDQ based on telomeric DNA and TERRA[21]. This scaffold allowed for in depth biophysical characterization, and their results support the hypothesis that RDQs can form stable structures and can be selectively targeted. However, their analysis utilized tetramolecular RDQs, which did not possess loop structures. TASQ enabled the formation and analysis of an RDQ, but there is still much to be learned about the biophysical properties of biologically relevant RDQs.

A second scaffold for the creation of RDQs utilizes a scaffolding duplex to study enzymatic stability and secondary structure[8,10]. Introduced by the Tan lab, the scaffold duplex provides an anchor to hold the guanine-rich RNA and DNA in close proximity, resulting in the formation of an RDQ. Two different duplex scaffolds were used. The first utilizes extended RNA and DNA strands that bind a third complementary DNA (creating 32 base pairs) (cDNA in Scheme 3.2)[8]. The second contains complementary RNA and DNA, which form 18 base pairs prior to the RDQ analogous to our approach (Scaffold Duplex in Scheme 3.2)[8].

These studies demonstrated the formation of RDQs using a scaffold duplex by RNAse digestion and subsequent gel electrophoresis, as well as secondary structure analyzed by circular dichroism[8,10].
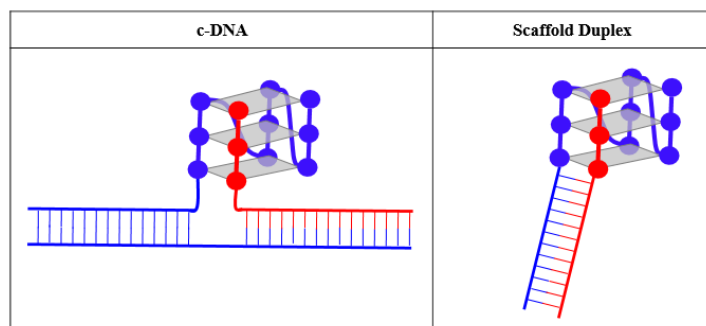
Scheme 3.2 Different methods for forming RDQs using a complementary DNA strand (left) and scaffold duplex (right).

In this chapter, we identify that a scaffold system is required for the formation of *in vitro* RDQs. Further, we utilize a scaffold duplex system to study telomeric and mitochondrial RDQs. Although there is a wide array of literature on intramolecular DNA and RNA GQs ranging from biophysical characterization to biological targeting, our understanding of RDQ systems is not as robust. Despite the many potential RDQ sequences, particularly those arising from G-rich RNA transcripts and non-coding RNAs that remain in the nucleus[7], little is known about their secondary structure and stability. The secondary structure analysis is of particular interest to us because of the unique differences between DNA and RNA GQs. Intermolecular DNA GQs have significant variation in structural topology (forming parallel, antiparallel and hybrid structures) and loop morphology (forming propeller, lateral and diagonal loops)[1,22]. However, RNA GQs almost always contain parallel topology and typically form propeller loops[3]. Since our constructs contain both DNA and RNA, this analysis will help identify if RDQ structures favor a specific structure, similar to RNA, or if they are more flexible in secondary structure, similar to DNA.

3.2 Materials and Methods

Table 3.1 Sequences of each RDQ.

Name	GQ Sequence
R2D2-Telo	GGGTTAGGG (D) GGGUUAGGG (R)
R1D3-Telo	GGGTTAGGGTTAGGG (D) GGG (R)
R1D3-CSB	GGGAGGGGGGG (D) GGG (R)

Table 3.2 Initial and Extended Duplex Sequences. Upper case letters indicate DNA and lower case letters indicate RNA.

Name	Duplex Sequence
Initial Duplex Sequence	5'TGTACGTCACAGCTA 3'acaugcagugucgau
Extended Duplex Sequence	5'GATCTCGTGATCGCTAGCTGTACGTCACAGCTA 3'CTAGAGCACTAGCGATCGACATGCAGTGTCGAT

UV-Vis Spectroscopy was conducted using a Cary 300 UV-Vis Spectrophotometer with a thermoelectrically controlled Peltier cell holder. Concentration checks were performed by scanning absorbance from 240-600 nm above 90 °C. Melting curves measured absorbance at 295 nm from 95 °C to 15 °C (cooling at 1° per min). Quartz spectrophotometer cuvettes (10 mM path lengths) from Starna Cells Inc. (http://www.starnacells.com/) were used for all experiments. The melting curve was normalized and the maximum of the first derivative plot was identified as the melting point (T_m).

Thermal Difference Spectra (TDS) were collected in triplicate from samples following a high temperature anneal (samples were held at 95 °C for five minutes and then slowly cooled to room temperature). UV cuvettes containing the samples were scanned (240-600 nm) individually at 25 °C. Samples were left in the heat block while the temperature was adjusted to 95 °C, and individual scans were repeated. TDS were

produced by subtracting the absorbance spectrum at 25 °C from the absorbance spectrum at 95 °C for each sample.

Circular Dichroism (CD) Spectroscopy was conducted using a JASCO 715 CD spectropolarimeter with a Peltier water-circulating temperature controller. Specific parameters included scan rate (100 nm/min), number of scans (6), and temperature control (25 °C).

Fluorescence Spectroscopy was conducted using a Tecan Infinite M1000 Plate Spectrometer for Job Plot analysis and a CARY Eclipse fluorescence spectrophotometer for all other fluorescence scans. Job plot analysis utilized varying concentrations of RDQ and ThT, such that the overall concentration stayed constant but the ratio varied from 0-1. Fluorescence was measured using an excitation wavelength of 430 nm and an emission wavelength of 482 nm. All other fluorescence scans using the CARY instrument were conducted using 1 µM RDQ and 10 µM ThT. Fluorescence titration utilized 200 nm ThT and RDQ concentration varied between 0-6.4 µM.

Fluorescence titrations with γPNA were performed on the CARY Eclipse Fluorescence Spectrophotometer to determine the degree of RDQ disruption. Each RDQ was prepared at 0.5 µM in triplicate and pre-annealed with γPNA ranging from 0 to 3.5 µM in 100 mM KCl Tris buffer. ThT was added after a 95 °C anneal up to 0.5 µM and then allowed to incubate for 5 minutes before the scan. Samples were excited at 430 nm and emission was monitored from 440 nm to 750 nm. In cases when the titration had been carried out more than once, the data was normalized to the maximum fluorescence at 482 nm in the titration for that specific RDQ.

3.3 Results

3.3.1 *A scaffold system is required in order to form an RDQ.*

In order to identify if a scaffold is required for the formation of RDQs *in vitro*, we combined DNA and RNA that contained two G-tracts of the telomeric sequence (i.e. GGGTTAGGG and GGGUUAGGG). In isolation, these strands form homo-GQs based on hypochromic transitions in UV thermal melting scans

(Figure 3.1, A and B) with T_M values of 37 °C and 71 °C. When both the RNA and DNA strands are present in the same sample (Figure 3.1 C), the result is two well-resolved melting transitions that match the melting temperatures and relative hypochromicities of the RNA and DNA alone. This indicates that the homologous RNA and DNA strands segregate to form bimolecular homoquadruplexes rather than combining to form heteroquadruplexes (i.e. RDQs). The significant difference in melting temperature between the RNA and DNA likely increases the preference for the formation of homoquadruplexes over RDQs. These experiments indicate an alternative method is required in order to promote RDQ formation.
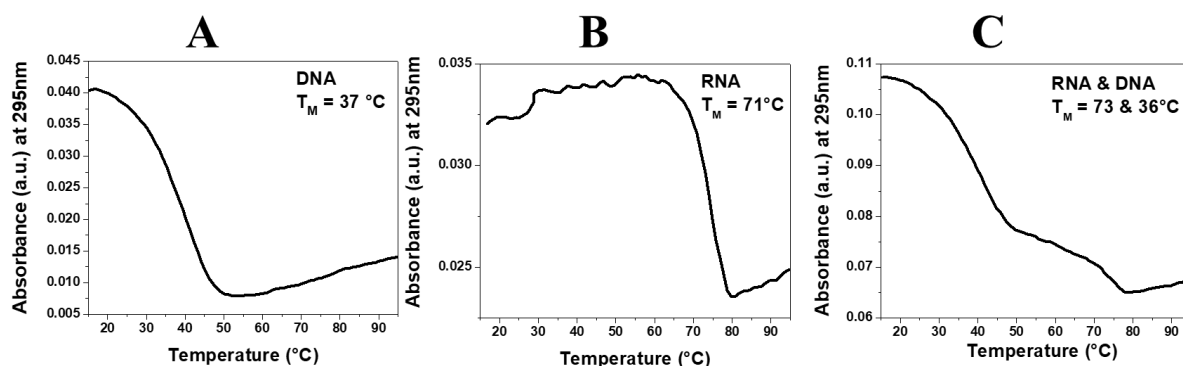
Figure 3.1 UV melting curves at 295 nm monitoring GQ formation with DNA T_M = 37 °C (A), RNA T_M = 71 °C (B), and RNA & DNA T_Ms = 73 & 36 °C (C).

3.3.2 *Using a scaffold duplex allows for the formation of an RDQ, and a linker between the scaffold and RDQ region is not required.*

To promote the formation of RDQs, we extended the DNA and RNA strands with complementary overhangs (sequences in Table 3.2) in order to form a heteroduplex that would force the respective G-tracts into close proximity. This method mimics one way in which an RDQ could form in a biological system. Transcription holds the unwound DNA and RNA transcript in close proximity, because both are locally tethered to the RNA polymerase. The high local concentration of G-rich RNA and DNA can then lead to intermolecular heteroquadruplex formation. We were uncertain about whether the duplex and quadruplex structures would form properly if positioned directly adjacent to one another due to differences in helical pitch[23], so we also tested sequences that contained a 5mer thymine (T_5) linker on one or both strands in

65

between the duplex- and quadruplex-forming sequences (Scheme 3.3). These sequences contain 1 guanine track of RNA and three tracts of DNA, shown in Scheme 3.3.

RNA+/DNA+
T_M (295nm)= 51.0 ±0.6

RNA-/DNA-
T_M (295nm)= 53.0 ±1.7

RNA-/DNA+
T_M (295nm)= 46.1 ±1.5

RNA+/DNA-
T_M (295nm)= 48.8 ±0.7
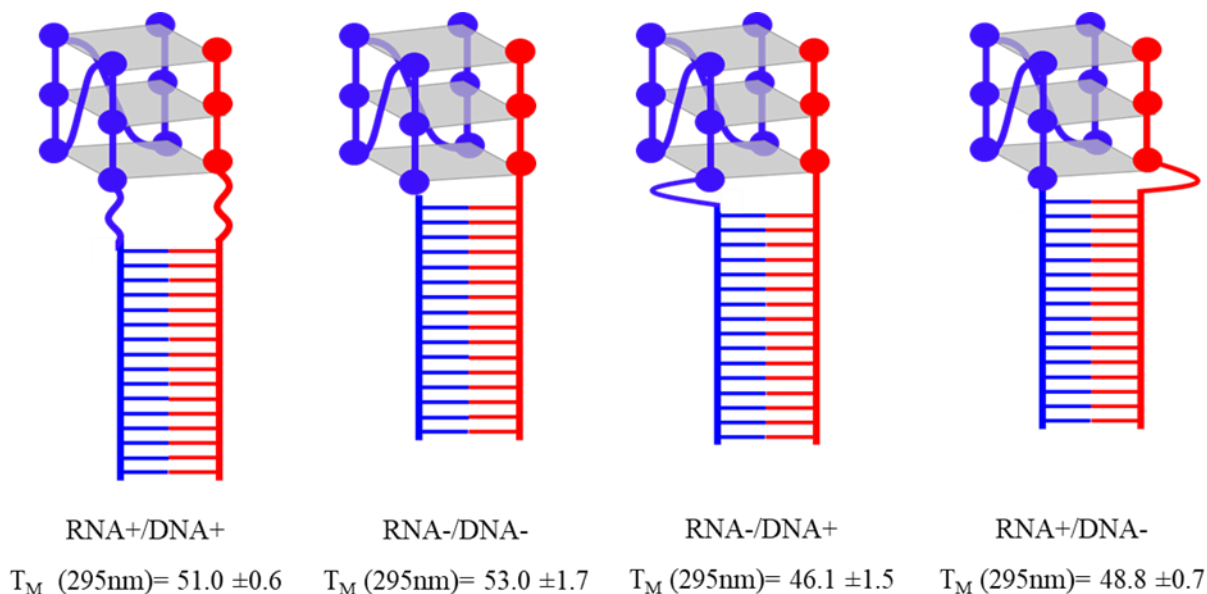
Scheme 3.3 RDQ formation and melting temperature using combinations (+/-) of the T_5 linker.

Scheme 3.3 illustrates our RDQ construct, with (+) or (-) indicating the presence or absence of the T_5 linker on the corresponding strand and the melting temperature (UV melting curves in Figure 3.2). The duplex (260 nm) and RDQ (295 nm) domains showed hyperchromic and hypochromic transitions, respectively. All RDQ domains show a single melting transition, indicating RDQ formation instead of RNA and DNA GQs. The highest temperature transitions for both the duplex and quadruplex domains were observed when neither strand had a T_5 linker. Additionally, the RDQ melting temperature decreased by 4-7 °C when only one strand (either RNA or DNA) contained the T_5 linker, indicating the importance of proximity between the two domains.
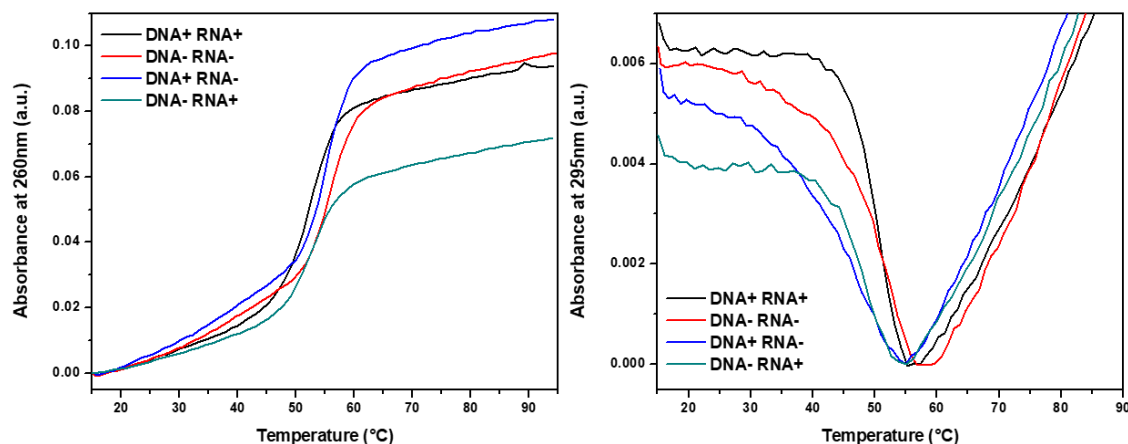
Figure 3.2 UV melting curves recording duplex (left) and RDQ (right) formation with and without a T5 linker on either DNA or RNA. n=3 DNA=1µM, RNA=1µM, 100mM KCl.

Circular dichroism (CD) is a technique used to characterize the secondary structure of nucleic acids[24]. Duplexes typically show a positive peak at 260 nm and a minimum at 240 nm[24]. Spectra of GQ structures are more variable and depend on the GQ topology. Parallel GQs, where the guanine tracts are oriented in the same direction, show maxima at 260 nm and minima at 210 nm[24]. Antiparallel GQs, containing guanine tracts in alternating directions, have maxima at 280-290 nm and minima at 240 nm[24]. Hybrid GQs are a mix of parallel and antiparallel G-tract orientations and exhibit spectral characteristics of both parallel and antiparallel structures. We utilized CD to identify any structural changes in the presence of the T_5 linker (Figure 3.3). Since the 15mer duplex scaffold in our RDQ structures contributes to the CD signal, we recorded the CD spectrum of the duplex alone and subtracted it from the overall CD to obtain the contribution from the heteroquadruplex (Figure 3.3, right). Since there is no difference in either the duplex subtracted or the raw CD signal, we conclude that the duplex and quadruplex domains do not require a linker in order to optimally stack on one another. The linker was omitted from all further experiments.
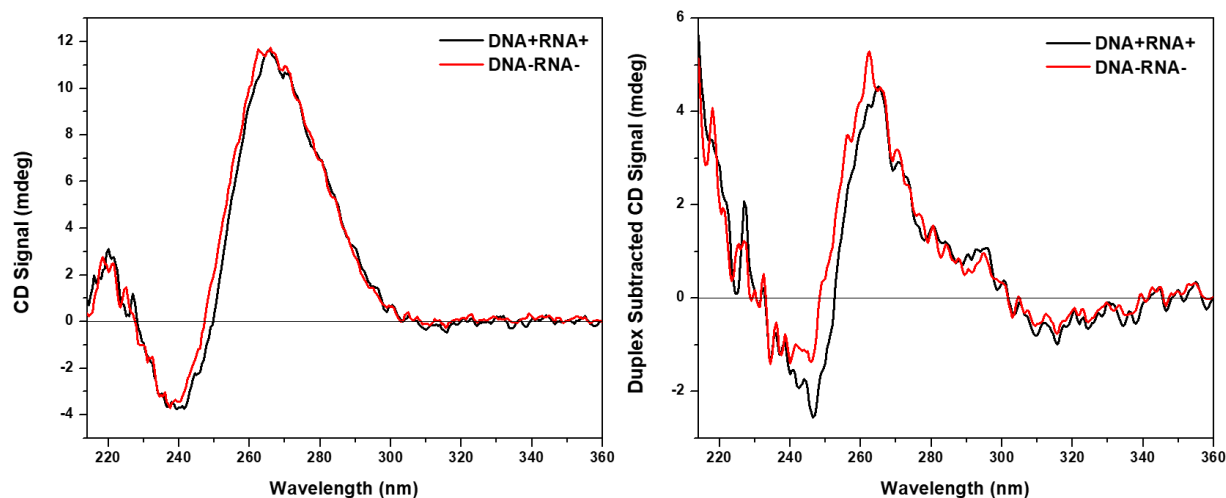
Figure 3.3 Circular dichroism spectra of RDQ constructs with (+) and without (-) T5 linkers. Spectra on the left include both duplex and RDQ, while the right spectra have the duplex region subtracted to only visualize the RDQ structure. DNA=1 µM, RNA=1 µM, 100 mM KCl.

3.3.3 *Different RDQ sequences can be made using the duplex scaffold and have differing thermal stabilities.*

Having demonstrated the ability to construct a stable RDQ using a duplex scaffold, we next sought to expand the scope of our design by varying the composition of the RDQ. In particular, we tested an additional permutation of the telomeric sequence (namely a 1:3 motif) to compare with the 2:2 motif described above, as well as a different 1:3 RDQ derived from the CSB-II sequence region of the mitochondrial genome (Scheme 3.4, RDQ sequences in Table 3.1).
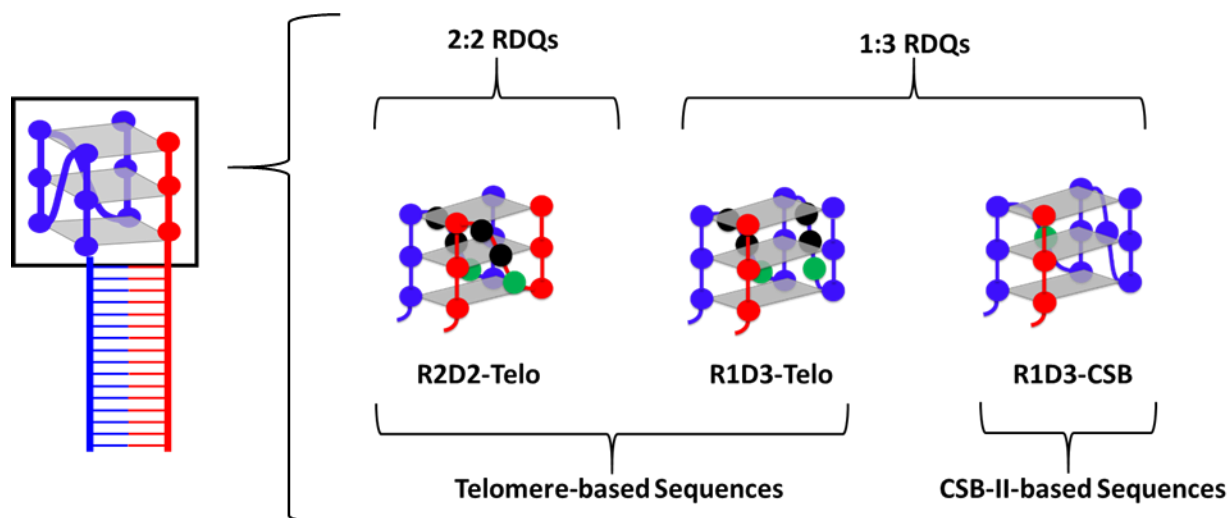
Scheme 3.4 RDQ variants based on the number of RNA and DNA tracts (2:2 and 1:3) as well as telomeric and mitochondrial genomic sequences.

Thermal melting curves for each RDQ are shown in Figure 3.4. R2D2-Telo and R1D3-Telo both exhibit clear hypochromic transitions in KCl buffer, with T_{MS} of 55.5 ± 0.4 and 52.3 ± 0.5 °C, respectively. These transitions are not observed in LiCl, consistent with formation of RDQs in KCl. The CSB-II RDQ (R1D3-CSB) shows a melting transition in KCl ($T_M = 57.5 \pm 0.4$), although the baselines of the curve have significant slopes.
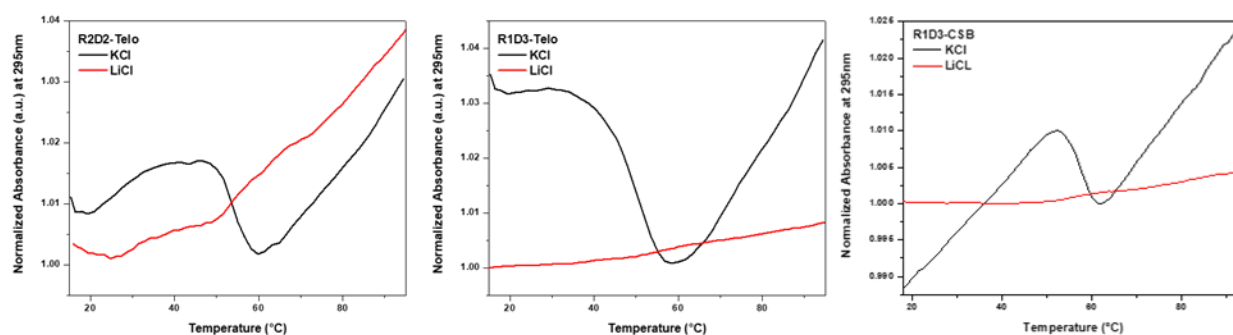
Figure 3.4 Thermal melting curves recorded for the 3 different RDQ sequences in the presence of 100 mM KCl and LiCl. GQs are indicated by hypochromic transitions (observed in KCl, black curves).

Thermal difference spectroscopy (TDS) is an alternate method to determine nucleic acid secondary structure. This analysis utilizes UV-Vis scans during bound and unbound states (i.e. low temperature and high temperature)[25]. The difference between these two spectra gives structure-specific characteristics and can differentiate between various DNA orientations, triplexes, i-motifs, and G-quadruplexes[25]. GQ TDS typically show positive peaks at 243 and 273 nm and a negative peak at 295 nm[25]. While these positive peaks overlap with antiparallel DNA, the addition of the negative 295 nm peak is distinct to GQs[25]. TDS with our telomere-based RDQ constructs have negative peaks in the 295 nm range (Figure 3.5). Comparing each RDQ in the presence of K^+ and Li^+ (Figure 3.5) confirmed RDQ formation for R2D2-Telo and R1D3-Telo. While R1D3-CSB did not show a negative peak in K^+, the signal was distinctly lower at 295 nm in the presence of K^+ compared to Li^+, indicating the possible presence of a GQ. The low intensity of the negative peak at 295 nm is likely due to the large number of Watson-Crick base pairs. While TDS is a useful tool for determining secondary structure, we could only use it in conjunction with our other analyses to confirm RDQ formation (UV melting, CD, and fluorescence).
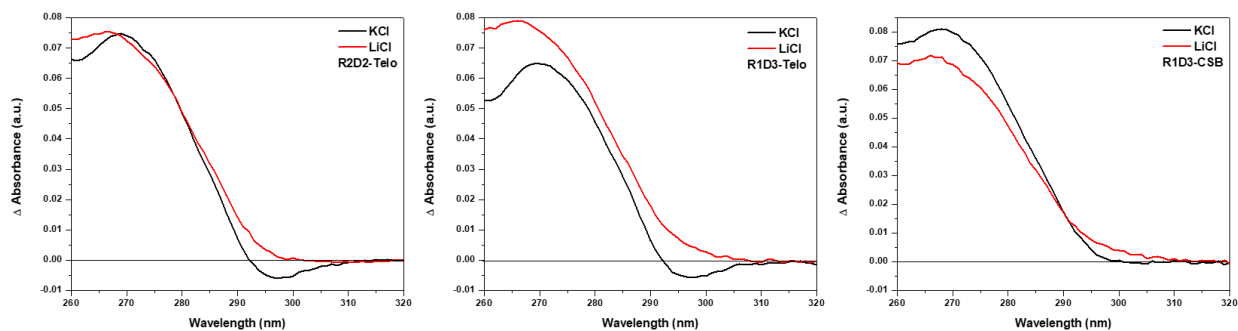
Figure 3.5 Thermal Difference Spectra with each RDQ sequence in 100 mM KCl or LiCl buffer.

3.3.4 *Extension of scaffold duplex can change the melting temperature of RDQs.*

Melting temperatures for the initial duplex of each RDQ ranged from 53.4 °C to 54.7 °C in the presence of potassium, which also supports the presence of RDQs. The overlap of the duplex and RDQ melting temperature raises questions about the actual melting temperature of the RDQs. In particular, since the duplex is required in order for the RDQ to form, the RDQ could actually melt considerably higher than

is observed. In order to address this, we increased the melting temperature of the duplex by more than doubling its length (Table 3.2). By using a 33 base pair (bp) DNA:DNA duplex, the melting temperature of the duplex increased by 19 °C in R2D2-Telo (Table 3.3, Figure 3.6). Furthermore, a 5.7 °C increase was observed in the RDQ melting temperature, relative to the shorter duplex used in the earlier system, indicating this RDQ was moderately more stable than previously identified. Meanwhile, R1D3-Telo showed a 20 °C increase in its duplex transition but only a 1.7 °C increase in RDQ melting temperature. These results demonstrate that the inherent stabilities of the telomeric RDQs were only slightly underestimated by the shorter constructs used in the earlier experiments. However, since all future experiments utilize high temperature annealed RDQ, we chose to utilize the shorter 15mer RNA:DNA duplex for subsequent experiments.

Table 3.3 Melting Temperatures (°C) obtained from UV melting curves recording RDQ formation in 100 mM KCl of each 15bp and 33bp duplex RDQ sequence.

T_M (°C)	15bp DNA:RNA duplex		33bp DNA:RNA duplex	
	260nm	**295nm**	**260nm**	**295nm**
R2D2-Telo	53.6 ± 0.8	55.5 ± 0.42	73.0 ± 0.3	61.2 ± 0.7
R1D3-Telo	53.4 ± 0.9	52.3 ± 0.46	73.3 ± 0.1	54 ± 1.3
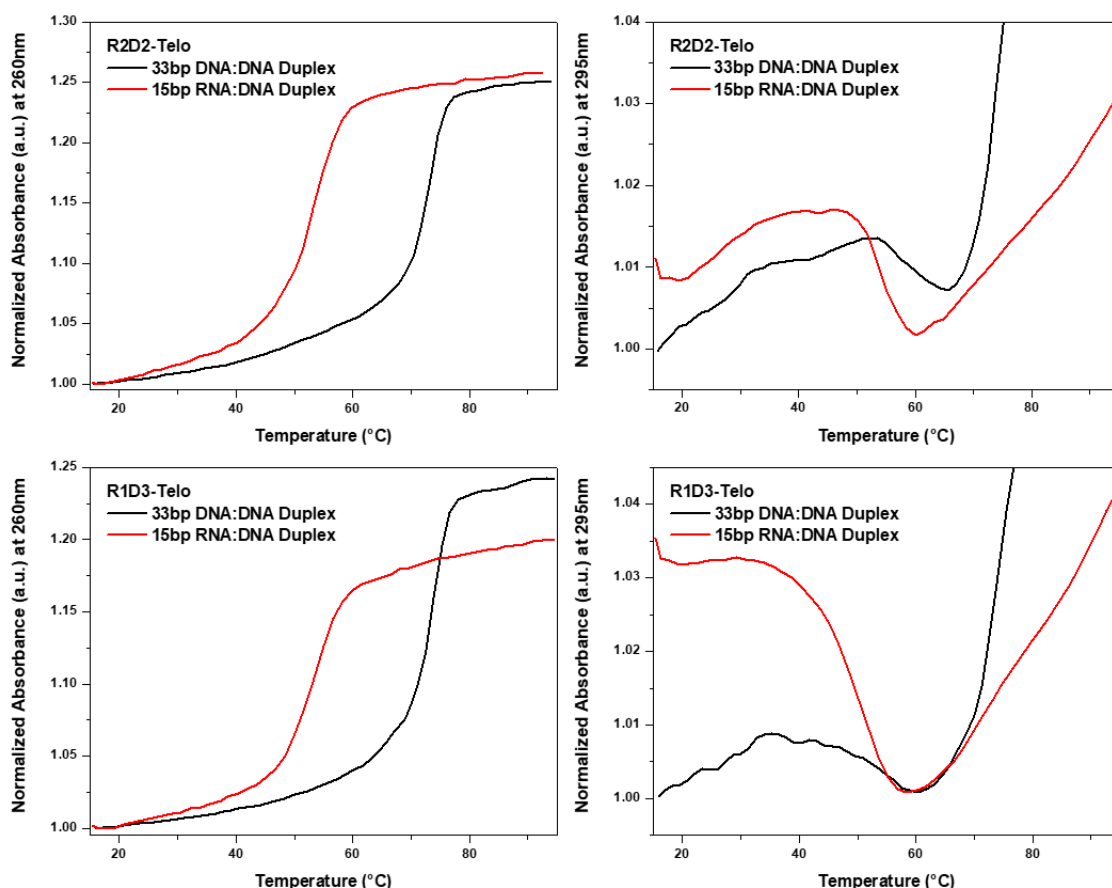
Figure 3.6 Thermal melting curves recorded at 260 nm (left) and 295 nm (right) for an extended 33bp DNA:DNA duplex fused to 2:2 (top) and 1:3 (bottom) telomeric RDQs.

3.3.5   *Telomere and CSBII sequences differ in secondary structure.*

Circular dichroism (CD) was again used to determine the secondary structure of each RDQ construct. Briefly, duplexes typically show a positive peak at 260 nm and a minimum at 240 nm[24]. GQ CD signal depends on the GQ topology: parallel GQs show maxima at 260 nm and minima at 210 nm[24], antiparallel GQs have maxima at 280-290 nm and minima at 240 nm[24], and hybrid GQs are a mix of parallel and antiparallel G-tract orientations and exhibit spectral characteristics of both parallel and antiparallel structures. Since the 15mer duplex scaffold in our RDQ structures contributes to the CD signal, we recorded the CD spectrum of the duplex alone (Supporting Figure 3.1, sample duRduD) and subtracted it from the

overall CD to obtain the contribution from the heteroquadruplex (Figure 3.7). Comparing potassium and lithium buffer, a quadruplex secondary structure is clearly observed in each RDQ. The telomeric RDQs showed primarily parallel structure based on maxima and minima at 260 and 240 nm, respectively. However, a shoulder at 295 nm is evident in R2D2-Telo and more so in R1D3-Telo, suggesting hybrid secondary structure. R1D3-Telo likely has a more prominent hybrid structure because of the increased number of more flexible DNA G-tracts. RNA G-tracts typically have parallel secondary structure due to structural constraints resulting from the 2'-OH in the ribose sugar[26]. GQs formed from telomeric DNA can range in secondary structure but typically have antiparallel structure[27]. Therefore, the increase in antiparallel nature of a telo-based sequence with an increased number of DNA tracts contributing to the RDQ is not surprising. The CSB-II based RDQ showed exclusively parallel secondary structure. While this sequence has the same number of DNA G-tracts contributing to the RDQ as in R1D3-Telo, previous reports also suggest CSB-II sequences possess parallel secondary structure[8]. These results indicate RDQs likely feature a combination of DNA and RNA GQ characteristics and are capable of variation in structure across sequences.
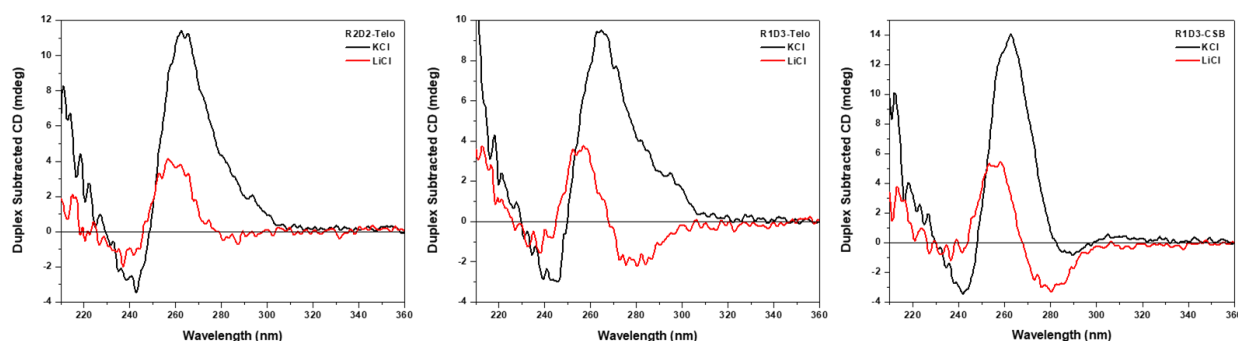
Figure 3.7 Duplex subtracted circular dichroism spectra for 1 µM RDQ in 100 mM KCl and LiCl. Spectra have the duplex region subtracted to only visualize the RDQ structure. Spectra were recorded at 25 °C.

3.3.6 *GQ selective fluorophore, Thioflavin T (ThT), can be used to indicate RDQ formation.*

Thioflavin T (ThT) is a known GQ binding dye that has been used to study both RNA and DNA GQs[28,29]. Specifically, ThT is a light-up probe, where enhanced fluorescence occurs when the benzylamine and benzothiazole rings of ThT are immobilized upon binding to a GQ (Figure 3.8). This immobilization reduces nonradiative decay of the excited state through conformational motions, thereby enhancing the fluorescence quantum yield of the dye.
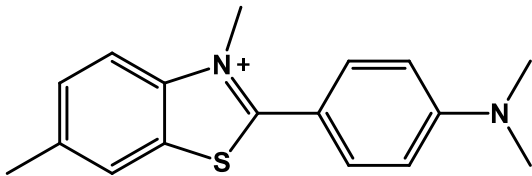
Figure 3.8 Thioflavin T (ThT) Structure

Initially, we wanted to confirm previous literature indicating 1:1 binding stoichiometry. Continuous variation experiments (i.e. Job plot) provide stoichiometry information by monitoring fluorescence intensity across a ratio of RDQ:ThT from 0 to 1. These experiments indicated 1:1 stoichiometry for the Telo based RDQs and between 1:1 and 2:1 stoichiometry for the CSB-based RDQ (Figure 3.9). This will play a larger role when fitting the fluorescence data later in section 3.3.7. Additionally, no fluorescence enhancement was observed for the duplex alone (duDduR).
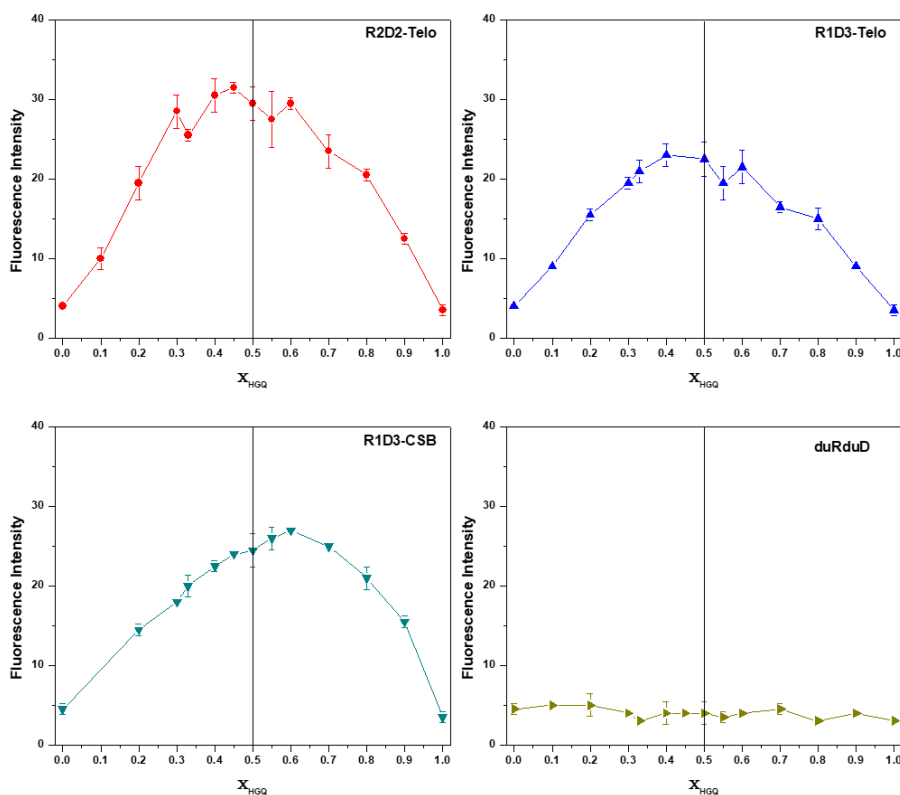
Figure 3.9 Job Plots with each RDQ sequence and Thioflavin T (ThT) in 100 mM KCl. A maxima at 0.5 X_{HGQ} indicates 1:1 binding stoichiometry. Excitation wavelength of 430 nm and emission wavelength of 482 nm.

We then utilized ThT to indicate the formation of each RDQ (Figure 3.10). Significant fluorescence enhancement is observed for each of the RDQs compared to the duplex scaffold alone (Figure 3.10, sample duRduD). This confirms RDQ formation in each model, including R1D3-CSB which gave somewhat conflicting results in UV melting and CD experiments.
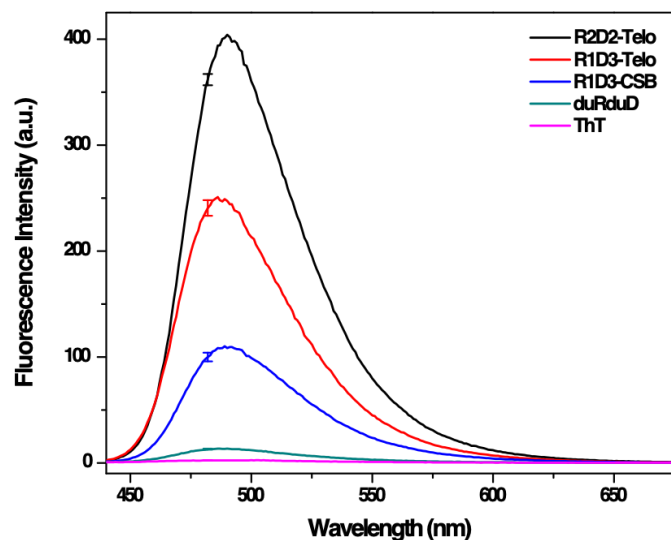
Figure 3.10 Fluorescence spectra of 1 µM RDQ with 10 µM ThT in 100 mM KCl. Samples were excited at 430 nm.

3.3.7   *Fluorescence can be used to calculate dissociation constants of ThT for RDQs.*

Florescence titrations allow the determination of dissociation binding constants, shown in Figure 3.11 and Table 3.4. From our previous Job plots, we determined R2D2-Telo and R1D3-Telo best fit a 1:1 binding model, while R1D3-CSB binds ThT in either 1:1 or 2:1 stoichiometry (Figure 3.9). R1D3-CSB titration data fit well using a 1:1 binding model (a 2:1 binding model fit is shown in Supporting Figure 3.2 and Supporting Table 3.1). K_d values ranged from 3.3 to 4.1 µM (Table 3.4). These values are comparable to other RNA and DNA intramolecular G quadruplexes, which showed K_ds ranging from 2.7-5.8 µM in similar ionic strengths (80-100 mM)[28,29].
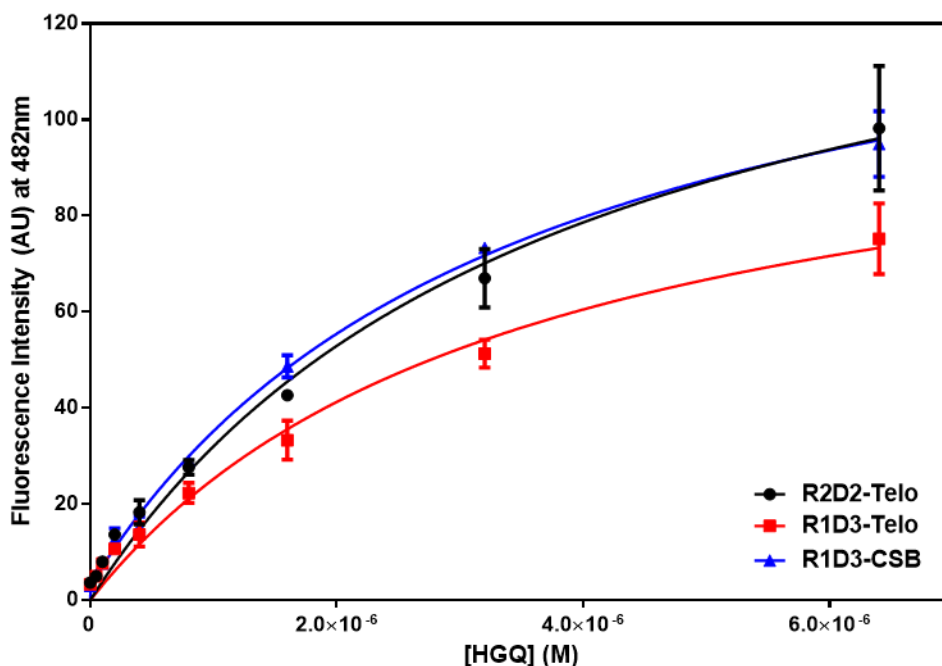
Figure 3.11 Fluorescence binding assay between each RDQ and ThT in 100 mM KCl. [ThT]=200 nm [RDQ]= 0-6.4 µM. ThT was excited at 430 nm and emission was recorded at 482 nm. Kd values were calculated using a one site binding model.

Table 3.4 K_d values calculated from fluorescence binding assay between each RDQ and ThT in 100 mM KCl.

	K_d (µM)
R2D2-Telo	3.5 ± 0.6
R1D3-Telo	3.3 ± 0.5
R1D3-CSB	4.1 ± 0.5

3.4 Discussion

DNA and RNA can interact in several different ways, which have the potential for different biological effects. Advancements in biochemistry have enabled the discovery of nuanced pathways for RNA-DNA interaction, allowing for a deeper understanding of biology. The first identified interaction is a Watson-Crick duplex formed between recently transcribed RNA and the coding DNA strand. The transcribed RNA:DNA hybrid leaves the non-template DNA as a single strand, forming an r-loop[30,31]. R-loops have been implicated in transcription as well as cellular differentiation[32–34]. A second interaction

between RNA and DNA is the formation of RNA:DNA triplexes, which usually form via a DNA duplex and RNA transcript. RNA:DNA triplexes have thoroughly analyzed structures but their biological implications are not well understood[35,36].

A third and more recently identified RNA:DNA interaction is the formation of RNA:DNA heteroquadruplexes (RDQs). RDQs were first identified in the conserved sequence block II (CSBII) sequence in human mitochondrial DNA[11,37]. CSBII RDQs were further implicated in mediating transcription termination and replication initiation[8]. In parallel, bioinformatics studies show conserved G-rich sequences in the non-template strands at and surrounding transcription start sites, providing further support for the hypothesis that RDQs regulate transcription and replication[7]. Since these findings, research has focused on the validation of *in vitro* and *in vivo* RDQs formed during transcription and on chromosomal ends (telomeres)[8–10,18,38,39].

While the biological relevance of RDQs has been examined, biophysical characterization has been somewhat less developed. This is likely due to the inherent challenges with making RDQs, since mixing guanine-rich RNA and DNA in solution predominantly forms RNA:RNA and DNA:DNA homoquadruplexes (Scheme 3.1, Figure 3.1). Forced proximity is vital to the formation of RDQs. Methods for covalently linking DNA and RNA allow for unique control over RDQ formation *in vitro* and *in vivo*[8,17,40,41]. Another example of a non-native RDQ uses a DNA oligonucleotide containing a duplex and quadruplex region[42]. The DNA oligo targeted the 5-UTR and coding sequence of RNA, where the duplex region allowed for increased selectivity to the location of interest and the quadruplex region formed an RDQ with the guanine-rich RNA[42]. This therapeutic successfully inhibited protein translation *in vitro* and *in vivo*[42]. While these systems provide a method for the visualization of RDQs and can be very useful in nucleic acid therapeutics, they do not allow an in depth biophysical characterization of naturally occurring RDQs.

We chose to utilize a duplex scaffold for a number of reasons, including ease of design and to resemble proposed biological structures. The experiments described above established biophysical properties of biologically relevant RNA:DNA heteroquadruplex (RDQ) sequences. In order to form an

RDQ, we utilize a duplex which hybridizes the 5' end of the DNA and 3' end of the RNA, preventing the formation of homo-GQs. Additionally, we observed no benefit in the presence of a T5 linker, indicating the closeness of the scaffold duplex does not hinder RDQ formation and actually favors it, based on T_M measurements.
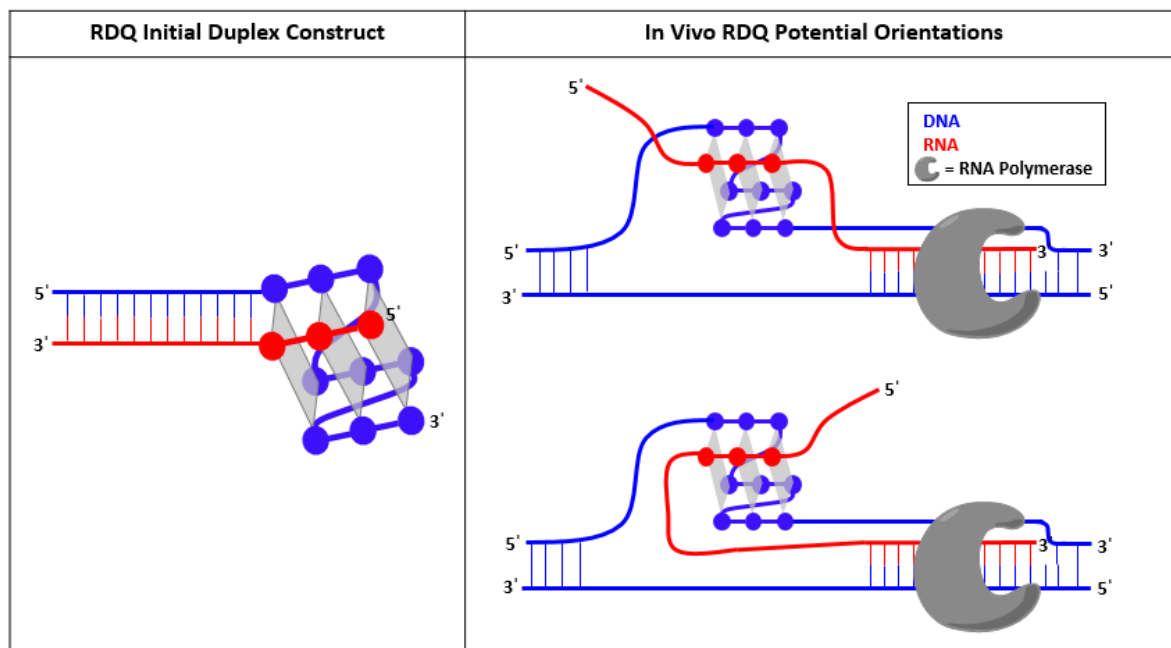
Scheme 3.5 Comparison of RDQ construct and *in vivo* RDQ strand orientations highlighting the differing orientations of RDQs.

While the scaffold duplex does not hinder RDQ formation, the RDQ orientation in our construct might deviate from that of an *in vivo* RDQ formed during transcription (Scheme 3.5). RDQs made up of an RNA transcript and the non-template DNA strand could run parallel to one another (5' – 3' and 5' – 3'), although it should be noted that the overall orientation of *in vivo* RDQs is not well understood. It would be possible for RDQ strands to run antiparallel, given sufficient single stranded DNA and RNA (Scheme 3.5). The construct shown in this report requires antiparallel stands (5' – 3' and 3' – 5') because of the duplex scaffold. However, when a T5 linker was incorporated between the RDQ and duplex no orientation changes were observed by circular dichroism (Figure 3.3). This indicates no negative impact on the RDQ using this antiparallel construct. Although the orientation used in this report has the potential to impact the overall

topology compared to native RDQs, it does not appear to do so in this case. Furthermore, the use of a duplex scaffold provides a convenient strategy for assembling RDQs for biophysical and structural characterization.

UV melting curves showed RDQ formation for each system (R2D2-Telo, R1D3-Telo, and R1D3-CSB) via clear hypochromic transitions. Furthermore, differences were observed between potassium (a GQ supporting ligand) and lithium (an unsupportive ligand) indicate RDQ formation.

The extension of the initial duplex provided the true melting temperature for each RDQ. This is because the duplex must be present in order for the RDQ to form. Therefore, if the RDQ had a T_M higher than ~55 °C (the highest duplex T_M), the observed T_M would not report the inherent T_M for the quadruplex. Instead, the RDQ would melt at a temperature similar to the duplex. This issue was addressed by extending the duplex sequence from 15 to 33 base pairs, making the T_M of the duplex 73 °C (about 20 °C higher). R2D2-Telo RDQ T_M increased from 55.5 to 61.2 °C. This 5.7 °C increase indicates the RDQ is more stable than initially identified. R1D3-Telo showed little change (1.7 °C) in T_M.

Circular dichroism (CD) was used to show variation in secondary structure for each RDQ, specifically parallel and hybrid structures. The use of a duplex control allowed for separation of GQ and duplex CD signal, since these signals overlap in wavelength. CD spectra showed two distinct secondary structures for the RDQs: parallel and hybrid. These structures differ in the strand orientation of the GQ's guanine tracts. R2D2-Telo and R1D3-Telo showed some degree of hybrid structure indicated by a shoulder at 295 nm, with R1D3-Telo having the more prominent shoulder. We hypothesize that this is because R1D3-Telo has an additional DNA G-tract in the RDQ, since the antiparallel tracts likely come from DNA based on literature reports. RNA is typically found in a parallel orientation due to constraints imposed by the 2'-OH[26]. Interestingly, this suggests structural variation in different constructs containing the same overall sequence. The CSB-II RDQ showed explicitly parallel structure, consistent with prior reports[8].

The selective fluorescence enhancement of Thioflavin T (ThT) by guanine quadruplexes allows structural confirmation for each of the studied RDQs. Here, ThT is likely stabilized by π-π interactions with one of the surfaces of the RDQ, immobilizing the benzylamine and benzothiazole rings of ThT.

Quadruplexes have two theoretical binding sites for ThT; one on each outside tetrad of the GQ. However, literature typically reports 1:1 GQ:ThT binding[28,29]. We confirmed 1:1 RDQ:ThT binding with Job plots, although the R1D3-CSB RDQ displayed an intermediate maximum in the Job plot between 1:1 and 2:1 RDQ:ThT binding. However, this did not significantly modify the ability to visualize RDQ formation by ThT. Fluorescence scans indicate RDQ formation where fluorescence enhancement was observed over a duplex control. Dissociation constants obtained from fluorescence assays indicate the binding of ThT to the RDQs is of similar affinity to the binding events between ThT and DNA or RNA GQs.

3.5 Future Directions

Using an initial duplex, we are able to form RDQ structures which vary in RNA:DNA content and loop sequence. Both telomere and CSB-II sequences form stable RDQs. Thermally stable RDQs were observed in the telomere sequences, while CSB-II RDQs were confirmed using CD and fluorescence. Structural differences were also observed between the three different RDQs. Specifically, telomere based RDQs show some hybrid structure, while the CSB-based RDQ exhibits primarily parallel secondary structures. This speaks to the possible variation in other RDQ sequences. Additionally, the dissociation constants between ThT and our RDQs fall within previously identified dissociation constants for DNA and RNA GQs.

3.6 References

(1)   Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. Quadruplex DNA: Sequence, Topology and Structure. *Nucleic Acids Res.* **2006**, *34* (19), 5402–5415.
(2)   Hänsel-Hertsch, R.; Di Antonio, M.; Balasubramanian, S. *DNA G-Quadruplexes in the Human Genome: Detection, Functions and Therapeutic Potential*; 2017.
(3)   Millevoi, S.; Moine, H.; Vagner, S. G-Quadruplexes in RNA Biology. *Wiley Interdiscip. Rev. RNA* **2012**, *3* (4), 495–507.
(4)   Maizels, N. G4 Motifs in Human Genes. *Ann. N. Y. Acad. Sci.* **2012**, *1267*, 53–60.
(5)   Eddy, J.; Maizels, N. Conserved Elements with Potential to Form Polymorphic G-Quadruplex Structures in the First Intron of Human Genes. *Nucleic Acids Res.* **2008**, *36* (4), 1321–1333.
(6)   Zhao, Y.; Du, Z.; Li, N. Extensive Selection for the Enrichment of G4 DNA Motifs in Transcriptional Regulatory Regions of Warm Blooded Animals. *FEBS Lett.* **2007**, *581* (10), 1951–1956.
(7)   Xiao, S.; Zhang, J.-Y.; Zheng, K.-W.; Hao, Y.-H.; Tan, Z. Bioinformatic Analysis Reveals an Evolutional Selection for DNA:RNA Hybrid G-Quadruplex Structures as Putative Transcription

Regulatory Elements in Warm-Blooded Animals.

(8)   Zheng, K.; Wu, R.; He, Y.; Xiao, S.; Zhang, J.; Liu, J.; Hao, Y.; Tan, Z. A Competitive Formation of DNA:RNA Hybrid G-Quadruplex Is Responsible to the Mitochondrial Transcription Termination at the DNA Replication Priming Site. *Nucleic Acids Res.* **2014**, *42* (16), 10832–10844.

(9)   Zheng, K.; Xiao, S.; Liu, J.; Zhang, J.; Hao, Y.; Tan, Z. Co-Transcriptional Formation of DNA:RNA Hybrid G-Quadruplex and Potential Function as Constitutional Cis Element for Transcription Control. *Nucleic Acids Res.* **2013**, *41* (10), 5533–5541.

(10)   Zhang, J.; Zheng, K.; Xiao, S.; Hao, Y.; Tan, Z. Mechanism and Manipulation of DNA:RNA Hybrid G-Quadruplex Formation in Transcription of G-Rich DNA. *J. Am. Chem. Soc.* **2014**, *136* (4), 1381–1390.

(11)   Wanrooij, P. H.; Uhler, J. P.; Simonsson, T.; Falkenberg, M.; Gustafsson, C. M. G-Quadruplex Structures in RNA Stimulate Mitochondrial Transcription Termination and Primer Formation. *Proc. Natl. Acad. Sci.* **2010**, *107* (37), 16072–16077.

(12)   Rippe, K.; Luke, B. TERRA and the State of the Telomere. *Nat. Struct. Mol. Biol.* **2015**, *22* (11), 853–858.

(13)   Cusanelli, E.; Chartrand, P. Telomeric Repeat-Containing RNA TERRA: A Noncoding RNA Connecting Telomere Biology to Genome Integrity. *Front. Genet.* **2015**, *6*, 143.

(14)   Balk, B.; Maicher, A.; Dees, M.; Klermund, J.; Luke-Glaser, S.; Bender, K.; Luke, B. Telomeric RNA-DNA Hybrids Affect Telomere-Length Dynamics and Senescence. *Nat. Struct. Mol. Biol.* **2013**, *20* (10), 1199–1205.

(15)   Toubiana, S.; Selig, S. DNA:RNA Hybrids at Telomeres - When It Is Better to Be out of the (R) Loop. *FEBS J.* **2018**, *285* (14), 2552–2566.

(16)   Luke, B.; Lingner, J. TERRA: Telomeric Repeat-Containing RNA. *EMBO J.* **2009**, *28* (17), 2503–2510.

(17)   Xu, Y.; Ishizuka, T.; Yang, J.; Ito, K.; Katada, H.; Komiyama, M.; Hayashi, T. Oligonucleotide Models of Telomeric DNA and RNA Form a Hybrid G-Quadruplex Structure as a Potential Component of Telomeres. *J. Biol. Chem.* **2012**, *287* (50), 41787–41796.

(18)   Xu, Y.; Suzuki, Y.; Ishizuka, T.; Xiao, C.-D.; Liu, X.; Hayashi, T.; Komiyama, M. Finding a Human Telomere DNA–RNA Hybrid G-Quadruplex Formed by Human Telomeric 6-Mer RNA and 16-Mer DNA Using Click Chemistry: A Protective Structure for Telomere End. *Bioorg. Med. Chem.* **2014**, *22* (16), 4419–4421.

(19)   Murat, P.; Bonnet, R.; Van der Heyden, A.; Spinelli, N.; Labbé, P.; Monchaud, D.; Teulade-Fichou, M.-P.; Dumy, P.; Defrancq, E. Template-Assembled Synthetic G-Quadruplex (TASQ): A Useful System for Investigating the Interactions of Ligands with Constrained Quadruplex Topologies. *Chem. - A Eur. J.* **2010**, *16* (20), 6106–6114.

(20)   Bonnat, L.; Bar, L.; Génnaro, B.; Bonnet, H.; Jarjayes, O.; Thomas, F.; Dejeu, J.; Defrancq, E.; Lavergne, T. Template-Mediated Stabilization of a DNA G-Quadruplex Formed in the HIV-1 Promoter and Comparative Binding Studies. *Chem. - A Eur. J.* **2017**, *23* (23), 5602–5613.

(21)   Bonnat, L.; Dejeu, J.; Bonnet, H.; Génnaro, B.; Jarjayes, O.; Thomas, F.; Lavergne, T.; Defrancq, E. Templated Formation of Discrete RNA and DNA:RNA Hybrid G-Quadruplexes and Their Interactions with Targeting Ligands. *Chem. - A Eur. J.* **2016**, *22* (9), 3139–3147.

(22)   Joachimi, A.; Benz, A.; Hartig, J. S. A Comparison of DNA and RNA Quadruplex Structures and Stabilities. *Bioorg. Med. Chem.* **2009**, *17* (19), 6811–6815.

(23)   Egli, M.; Pallan, P. S. The Many Twists and Turns of DNA: Template, Telomere, Tool, and Target. *Curr. Opin. Struct. Biol.* **2010**, *20* (3), 262–275.

(24)   Kypr, J.; Kejnovská, I.; Bednářová, K.; Vorlíčková, M. Circular Dichroism Spectroscopy of Nucleic Acids. In *Comprehensive Chiroptical Spectroscopy*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2012; pp 575–586.

(25)   Mergny, J.-L.; Li, J.; Lacroix, L.; Amrane, S.; Chaires, J. B. Thermal Difference Spectra: A Specific Signature for Nucleic Acid Structures. *Nucleic Acids Res.* **2005**, *33* (16), e138.

(26) Fay, M. M.; Lyons, S. M.; Ivanov, P. RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms. *J. Mol. Biol.* **2017**, *429* (14), 2127–2147.

(27) Neidle, S.; Parkinson, G. N. The Structure of Telomeric DNA. *Curr. Opin. Struct. Biol.* **2003**, *13* (3), 275–283.

(28) Gabelica, V.; Maeda, R.; Fujimoto, T.; Yaku, H.; Murashima, T.; Sugimoto, N.; Miyoshi, D. Multiple and Cooperative Binding of Fluorescence Light-up Probe Thioflavin T with Human Telomere DNA G-Quadruplex. *Biochemistry* **2013**, *52* (33), 5620–5628.

(29) Xu, S.; Li, Q.; Xiang, J.; Yang, Q.; Sun, H.; Guan, A.; Wang, L.; Liu, Y.; Yu, L.; Shi, Y.; et al. Thioflavin T as an Efficient Fluorescence Sensor for Selective Recognition of RNA G-Quadruplexes. *Sci. Rep.* **2016**, *6*.

(30) Lee, D. Y.; Clayton, D. A. Properties of a Primer RNA-DNA Hybrid at the Mouse Mitochondrial DNA Leading-Strand Origin of Replication. *J. Biol. Chem.* **1996**, *271* (39), 24262–24269.

(31) Xu, B.; Clayton, D. A. RNA-DNA Hybrid Formation at the Human Mitochondrial Heavy-Strand Origin Ceases at Replication Start Sites: An Implication for RNA-DNA Hybrids Serving as Primers. *EMBO J.* **1996**, *15* (12), 3135–3143.

(32) Santos-Pereira, J. M.; Aguilera, A. R Loops: New Modulators of Genome Dynamics and Function. *Nat. Rev. Genet.* **2015**, *16* (10), 583–597.

(33) Chiang, H.-C.; Zhang, X.; Li, J.; Zhao, X.; Chen, J.; Wang, H. T.-H.; Jatoi, I.; Brenner, A.; Hu, Y.; Li, R. BRCA1-Associated R-Loop Affects Transcription and Differentiation in Breast Luminal Epithelial Cells. *Nucleic Acids Res.* **2019**.

(34) Al-Hadid, Q.; Yang, Y. R-Loop: An Emerging Regulator of Chromatin Dynamics. *Acta Biochim. Biophys. Sin. (Shanghai).* **2016**, *48* (7), 623–631.

(35) Sentürk Cetin, N.; Kuo, C.-C.; Ribarska, T.; Li, R.; Costa, I. G.; Grummt, I. Isolation and Genome-Wide Characterization of Cellular DNA:RNA Triplex Structures. *Nucleic Acids Res.* **2019**, *47* (5), 2306–2321.

(36) Li, Y.; Syed, J.; Sugiyama, H. Review RNA-DNA Triplex Formation by Long Noncoding RNAs. *Cell Chem. Biol.* **2016**.

(37) Wanrooij, P. H.; Uhler, J. P.; Shi, Y.; Westerlund, F.; Falkenberg, M.; Gustafsson, C. M. A Hybrid G-Quadruplex Structure Formed between RNA and DNA Explains the Extraordinary Stability of the Mitochondrial R-Loop. *Nucleic Acids Res.* **2012**, *40* (20), 10334–10344.

(38) Wu, R.; Zheng, K.; Zhang, J.; Hao, Y.; Tan, Z. Formation of DNA:RNA Hybrid G-Quadruplex in Bacterial Cells and Its Dominance over the Intramolecular DNA G-Quadruplex in Mediating Transcription Termination. *Angew. Chemie Int. Ed.* **2015**, *54* (8), 2447–2451.

(39) Zhao, Y.; Zhang, J.; Zhang, Z.; Tong, T.; Hao, Y.; Tan, Z. Real-Time Detection Reveals Responsive Cotranscriptional Formation of Persistent Intramolecular DNA and Intermolecular DNA:RNA Hybrid G-Quadruplexes Stabilized by R-Loop. *Anal. Chem.* **2017**, *89* (11), 6036–6042.

(40) Haase, L.; Dickerhoff, J.; Weisz, K. DNA-RNA Hybrid Quadruplexes Reveal Interactions That Favor RNA Parallel Topologies. *Chem. - A Eur. J.* **2018**, *24* (57), 15365–15371.

(41) Qi, J.; Shafer, R. H. Human Telomere Quadruplex: Refolding and Selection of Individual Conformers via RNA/DNA Chimeric Editing †. *Biochemistry* **2007**, *46*, 7599–7606.

(42) Bhattacharyya, D.; Nguyen, K.; Basu, S. Rationally Induced RNA:DNA G-Quadruplex Structures Elicit an Anticancer Effect by Inhibiting Endogenous EIF-4E Expression. *Biochemistry* **2014**, *53* (33), 5461–5470.

3.7   Supporting Figures
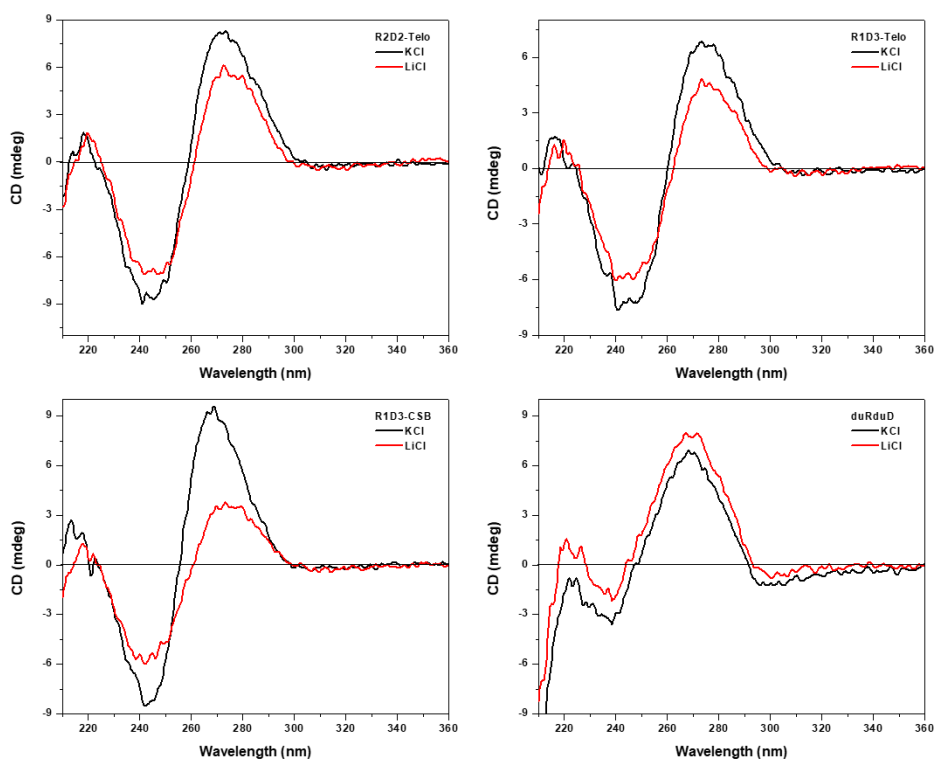
Supporting Figure 3.1. Circular dichroism spectra for 1 µM RDQ or duplex control in 100 mM KCl or LiCl. Spectra were recorded at 25 °C. Duplex control spectrum were subtracted from RDQ spectra to give results shown in Figure 3.7.
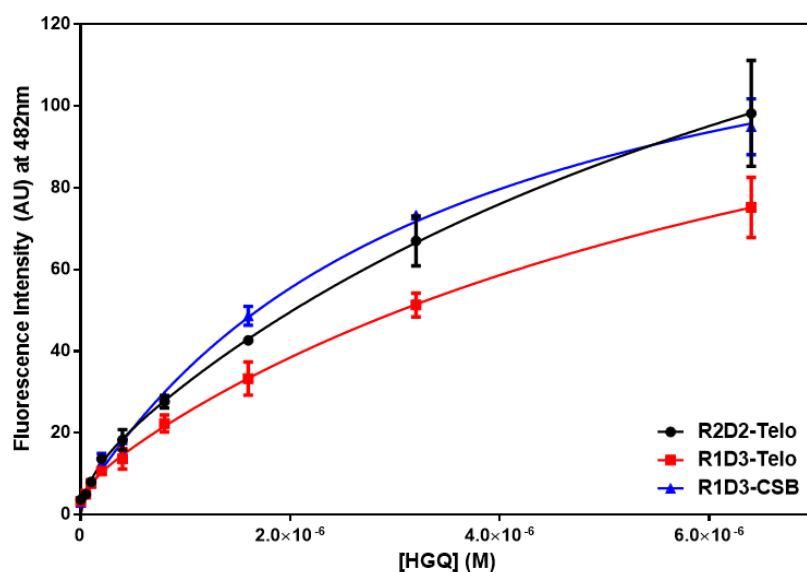
Supporting Figure 3.2 Fluorescence binding assay between each RDQ and ThT in 100 mM KCl. [ThT]=200 nm [RDQ]= 0-6.4 µM. ThT was excited at 430 nm and emission was recorded at 482 nm. Kds were calculated using a two site-specific binding model.

Supporting Table 3.1 Kd and Fmax values calculated from fluorescence binding assay between RDQ and ThT in 100 mM KCl using a 2-site binding model.

	Kd (μM)	**Fmax**
R1D3-CSB 2site	Site 1: X (not within confidence thresholds) Site 2: 3.4 ± 0.7	Site 1: 3 ± 3 Site 2: 143 ± 9

An additional set of experiments was conducted to compare different sequence orders using

R2D2-Telo. Underlined in Supporting Table 3.2, we studied the difference in loop sequence of the RNA

component of the RDQ.

Supporting Table 3.2 RDQ sequences used. RNA shown in bold and the sequence differences are underlined.

Name	RDQ Sequences RNA shown in bold)
R2D2-Telo	5' GGGTTAGGG 3' **GGGUUAGGG**
R2D2-Alt-Telo	5' GGGTTAGGG 3' **GGGAUUGGG**
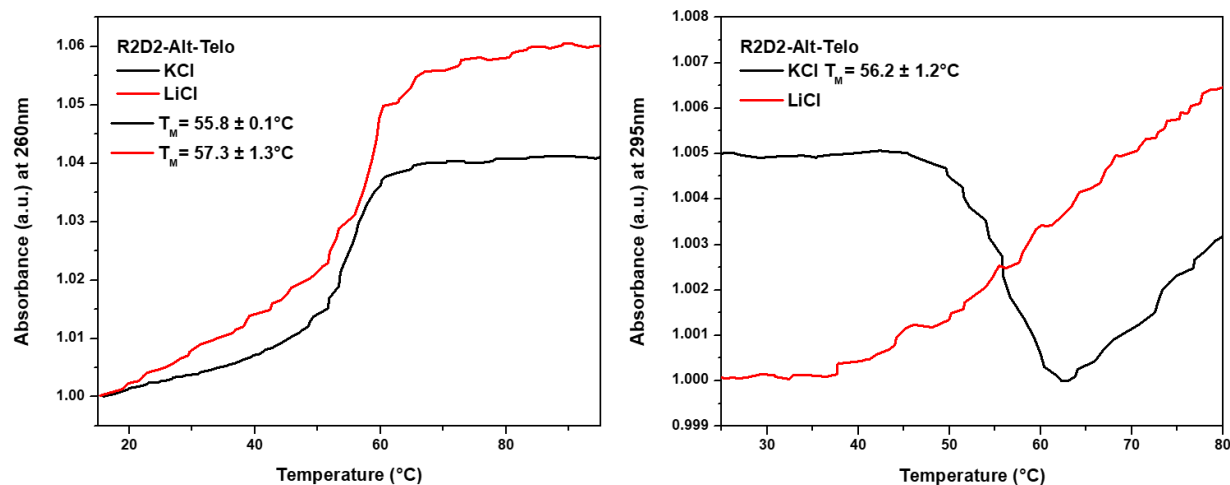
Supporting Figure 3.3 Thermal melting curves at 260 nm (left) and 295 nm (right) of R2D2-Alt-Telo in KCl and LiCl Buffer.
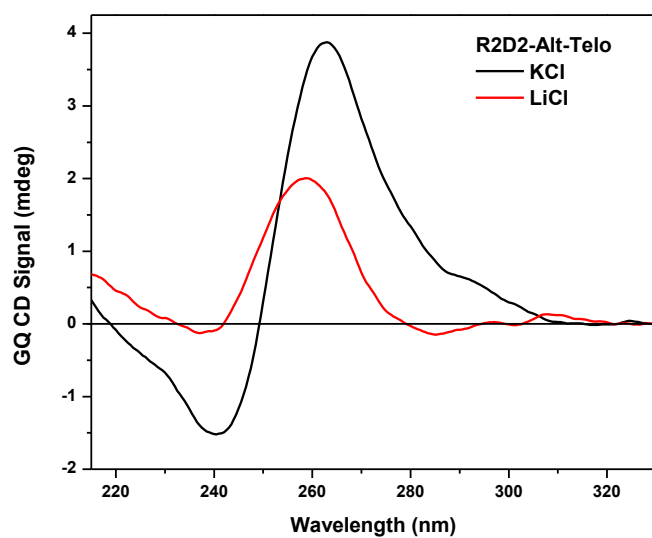
Supporting Figure 3.4 GQ circular dichroism signal of R2D2-Alt-Telo in KCl and LiCl Buffer.
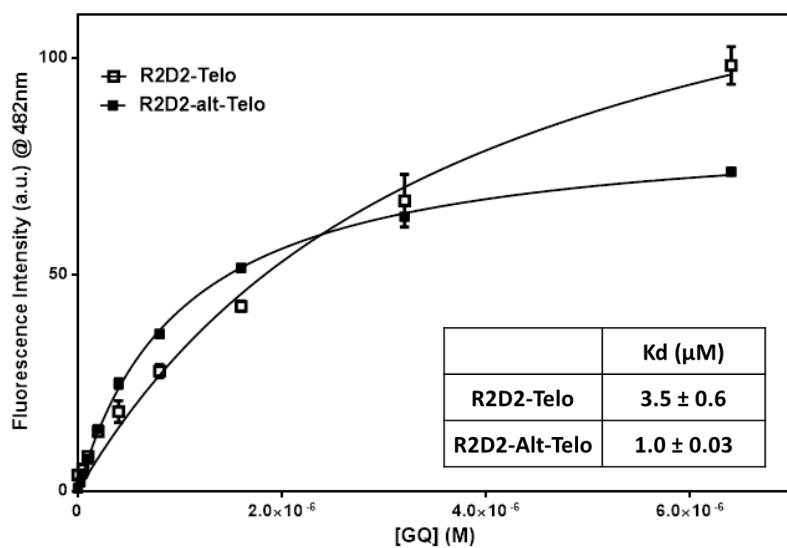
Supporting Figure 3.5 Fluorescence binding assay between each RDQ and ThT in 100 mM KCl. [ThT]=200 nm [RDQ]= 0-6.4 µM. ThT was excited at 430 nm and emission was recorded at 482 nm. Kds were calculated using a one site-specific binding model.

Thermal melting curves, recorded at 295 nm show a melting transition at $56.2 \pm 1.2°C$ for the R2D2-Alt-Telo sequence (Supporting Figure 3.3). This is very close to the $55.5 \pm 0.4°C$ melting temperature recorded for R2D2-Telo, indicating no differences in stability between the two sequences. Additionally, no change was observed in the secondary structure of the RDQ. Both show primarily parallel secondary structure with a slight hybrid shoulder (Supporting Figure 3.4). The most interesting result with R2D2-Alt-Telo was in the ThT dissociation constants, where R2D2-Telo showed a K_d of $3.5 \pm 0.6 \, \mu M$ whereas R2D2-Alt-Telo was 3X lower at $1.0 \pm 0.03 \, \mu M$ (Supporting Figure 3.5). This difference is likely a result of different stabilizing interactions by the loop. Overall, these results suggest that the difference in loop sequence does not significantly impact the stability or secondary structure, it can impact the binding interactions with small molecules.

4   Chapter 4. An Analysis of Monomeric GQs, Dimeric HomoGQs, and Dimeric HeteroGQs and Targeting by γPNA.

4.1   Introduction

4.1.1   Variations in GQ Molecularity (monomeric and dimeric GQs)

Chapter 4 will compare three different quadruplex motifs: monomeric guanine quadruplexes (GQs) (labeled DNA and RNA in Scheme 4.1), dimeric homoGQs (DNA:DNA and RNA:RNA in Scheme 4.1), and dimeric heteroGQs (RNA:DNA in Scheme 4.1). Each motif requires guanine-rich nucleic acid to assemble into tetrads, followed by tetrad stacking to form the GQ. Dimeric heteroGQs are discussed in detail in Chapter 3, where we analyze three different RNA:DNA heteroquadruplexes (RDQs). In this chapter, we first utilize a telomeric sequence to analyze differences in structure and stability that arise from variations between monomeric and dimeric GQs. Then, we target our dimeric homoGQs and heteroGQs with gamma modified peptide nucleic acid (γPNA).
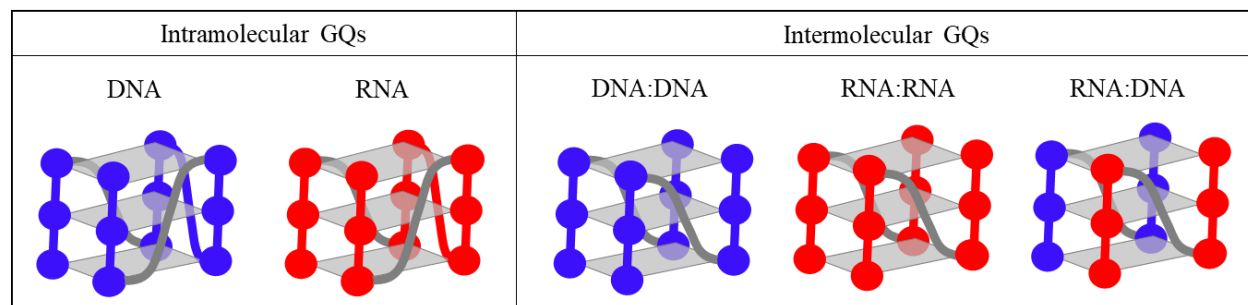
Scheme 4.1 Intra- and Intermolecular GQs shown with DNA (blue) and RNA (red).

This thesis has discussed a variety of GQ constructs; here we will focus on intermolecular GQs and differences between other intermolecular GQs and single molecule GQs. Single molecule GQs, also called intramolecular GQs, contain enough guanines to create a GQ in a single nucleic acid strand; this is typically shown as $G_A X_B G_A X_B G_A X_B G_A X_B$ where A is 2-3 guanines and B between 1 and up to 70 nucleotides[1–3]. In contrast, dimeric GQs (intermolecular GQs) utilize two nucleic acid strands to create a GQ (Scheme 4.1)[4].

Dimeric GQs can be classified in two different ways: homoGQs, where identical strands come together to create the GQ or heteroGQs, where two different strands comprise the GQ.
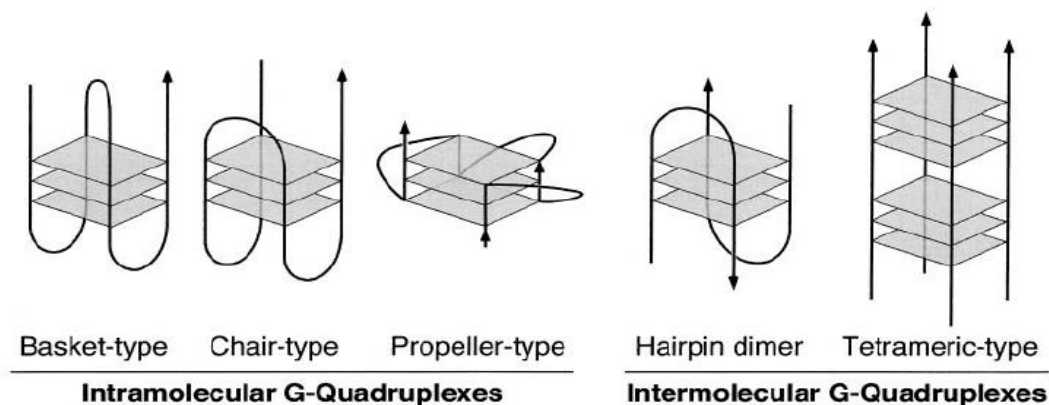
Scheme 4.2 Structural variations within intra- and intermolecular GQs. Adapted from Reference 5.

Intramolecular and intermolecular GQs can vary in topology, observed in Scheme 4.2. This variation is based on the directionality of the guanine tracts. Furthermore, the orientation of the glycosidic bond can also be used to determine directionality. The glycosidic bond is the covalent bond linking the C1 of the ribose sugar to the N9 of the guanine base[1]. This nucleoside can be found in the syn- or the *anti*-conformation; the difference between the two is a 180° rotation of the glycosidic bond[1]. When every strand in a GQ shares the same 5ʹ to 3ʹ orientation, all glycosidic bonds are in the *anti*- conformation. These GQs are labeled as parallel. *Anti*-parallel GQs form when two strands are oppositely oriented and contain equal numbers of syn- and *anti*- glycosidic bonds. In single molecular GQs, antiparallel structure results from alternating orientation between each continuous guanine tract. In contrast, intermolecular GQs can form an antiparallel GQ by alternating strand directionality. Additionally, GQs can form hybrid structures where one strand runs opposite the other three, resulting in an uneven number of syn- and *anti*-dG. Some GQ structures have been characterized as displaying specific conformations, while others have found that GQs can be forced from one conformation to another via external factors[6]. One GQ that has been observed to have a variety of secondary structures is the telomeric DNA GQ.

4.1.2   Rationale for Forming Dimeric GQs Using a Scaffold

Chapter 3 illustrated that while RNA:DNA heteroquadruplexes (RDQs) require a scaffold system, dimeric DNA:DNA (DDQ) and RNA:RNA (RRQ) GQs do not need a scaffold. However, one of the goals of this chapter is to compare dimeric homoGQs (DDQ and RRQ) to the RDQs shown in Chapter 3. In order to make an accurate comparison to our RDQs, we utilize the duplex scaffold from our RDQ construct to form DDQ-Scaffold and RRQ-Scaffold. This allows us to not only compare the RNA:DNA heteroquadruplexes to dimeric homoquadruplexes, but also to compare dimeric homoquadruplexes with and without a scaffold.

4.1.3   Dimeric G-Quadruplex Targeting by PNA

Synthetic nucleic acid analogs, such as peptide nucleic acid (PNA), are well established probes for targeting GQ structures. Our lab has extensively analyzed PNA-GQ interactions, by both homologous and complementary binding[7–11]. These experiments can be designed to fully disrupt GQ formation or maintain GQ structure, with the overall purpose of impacting some biological function.

We selected a complementary gamma modified PNA (γPNA) to target our RDQ constructs. Using a complementary probe allows us to selectively disrupt the presence of an individual RDQ, whereas a homologous binding PNA would not be as selective. Further, the presence of the RDQ has been shown to regulate transcription/replication, making the disruption of RDQs an interesting therapeutic target[12]. Since the RDQ constructs designed here utilize a scaffold duplex, the γPNA molecules can be designed to target both the RDQ as well as the scaffold duplex. This ensures sufficient binding and assists in determining specificity. Previous work has utilized short, 6mer, PNA probes to target bimolecular GQs. These sequences were effective, however the short cytosine-rich sequences formed triplexes with the GQ sequences[13].

Gamma modified PNA (γPNA) was selected for its high binding affinity. Incorporating a mini-PEG on the gamma position pre-organizes the PNA into a right handed helix, decreasing the entropic penalty paid during binding[14]. The 15mer γPNA was designed so that the first three bases of the guanine

tract and the subsequent 12 bases of the scaffold duplex are invaded. This allows disruption of both the GQ and the duplex, ensuring full separation.

4.1.4   Project Goals

This project aims to develop our understanding of dimeric GQs; specifically, how they relate to intramolecular GQs and how different dimeric GQs relate to each other. We first compare an intramolecular GQ to an intermolecular homoGQ. Then we compare intermolecular homoGQs formed with and without duplex scaffolds. Next, we observe differences between intermolecular homo- and heteroGQs. Finally, we utilize γPNA probes to understand binding affinities between different GQ systems.

4.2   Materials and Methods

Table 4.1 Sequences of each GQ. RNA is shown in bold.

Name	Sequence (RNA is bolded)
Intra-GQ	5'GGGTTAGGGTTAGGGTTAGGG
DDQ	5'GGGTTAGGG
RRQ	5'**GGGUUAGGG**
RDQ-Scaffold (also R2D2-Telo)	5'TGTACGTCACAGCTAGGGTTAGGG 3'ACATGCAGTGTCGAT**GGGUUAGGG**
RRQ-Scaffold	5'TGTACGTCACAGCTA**GGGUUAGGG** 3'ACATGCAGTGTCGAT**GGGUUAGGG**
DDQ-Scaffold	5'TGTACGTCACAGCTAGGGTTAGGG 3'ACATGCAGTGTCGATGGGTTAGGG
R1D3-Telo	5'TGTACGTCACAGCTAGGGTTAGGGTTAGGG 3'ACATGCAGTGTCGAT**GGG**
R1D3-CSB	5'TGTACGTCACAGCTAGGGAGGGGGGG 3'ACATGCAGTGTCGAT**GGG**

Table 4.2 γPNA Sequence.

Name	Sequence (C-N)
γPNA	NH_2- CCC ATC GAC ACT GCA –H

DNA and RNA Sample Preparation All oligonucleotides were purchased from Integrated DNA Technologies (https://www.idtdna.com) and dissolved in nanopure water. All samples were prepared in either 100 mM KCl buffer (10 mM Tris Base, 0.1 mM salt free EDTA, and 100 mM KCl pH=7.00) or 100 mM LiCl buffer (10 mM Tris Base, 0.1 mM salt free EDTA, and 100 mM LiCl pH=7.00).

Peptide Nucleic Acid Synthesis PNA oligomers were synthesized by solid phase synthesis using Fmoc/Bhoc protected γPNA monomers where the gamma modification is a mini-PEG. Oligomers were purified with high-performance liquid chromatography (HPLC). The PNA was injected into the HPLC and run through the semi-prep column at a flow rate of 10 mL/min with gradient of 95-40% water/acetonitrile with 0.1% TFA for 40 minutes (Supporting Figure 4.1). The column was held at 50 °C. Molecular weight (Supporting Figure 4.2) was confirmed with matrix assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF) using α-cyano-4-hydroxycinnamic acid as the matrix dissolved in 50:50 H_2O/ACN with 0.1% TFA. (Expected mass: 5761g/mol Observed mass: 5755g/mol)

UV-Vis Spectroscopy was conducted using a Cary 300 UV-Vis Spectrophotometer with a thermoelectrically controlled Peltier cell holder. Concentration checks were performed by scanning absorbance from 240-600 nm above 90 °C. Melting curves measured absorbance at 295 nm from 95 °C to 15 °C (cooling at 1° per min). Quartz spectrophotometer cuvettes (10 mM path lengths) from Starna Cells Inc. (http://www.starnacells.com/) were used for all experiments. The heating curve was normalized and the melting point (T_m) was determined by alpha plot. Alpha plots normalize the melting curve by identifying upper and lower base lines in order to determine the melting point, where alpha = 0.5.

Circular Dichroism (CD) Spectroscopy was conducted using a JASCO 715 CD spectropolarimeter with a Peltier water-circulating temperature controller. Specific parameters included scan rate (100 nm/min), number of scans (6), and temperature control (25 °C).

Fluorescence Spectroscopy was conducted using a CARY Eclipse fluorescence spectrophotometer. Fluorescence titration utilized 200 nm ThT and RDQ concentration varied between 0-6.4 µM. For γPNA titrations, each RDQ was prepared at 0.5 µM in triplicate and pre-annealed with γPNA ranging from 0 to 3.5 µM in 100 mM KCl Tris buffer. ThT was added after a 95 °C anneal up to 0.5 µM and then allowed to

incubate for 5 minutes before the scan. Samples were excited at 430 nm and emission was monitored from 440 nm to 750 nm. Data was normalized using an RDQ control sample to account for day to day variation in the fluorimeter.

4.3   Results

4.3.1   *Intra- vs Intermolecular GQs have differing melting temps and differing secondary structures*
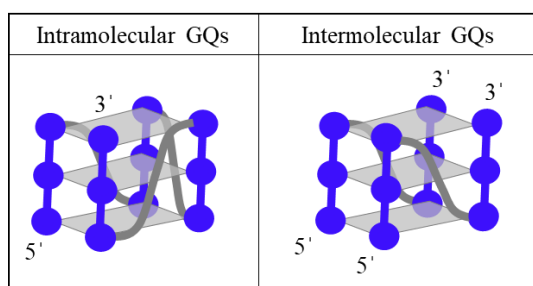
Scheme 4.3 Schematic representation of intramolecular and intermolecular DNA GQs.

Telomeres are chromosomal end caps that consist of the repeating sequence TTAGGG in humans. The number of repeats is highly variable, depending on genetics and lifestyle factors[15]. Telomeres average between 5-15kb with a 3′ overhang between 35-600 nucleotides in length[15]. This 3′ overhang is the result of DNA replication, since the replication machinery runs unidirectionally. The result is a G-rich overhang that is able to form GQs in both an intra- and intermolecular fashion. Intramolecular GQs contain a single nucleic acid strand while intermolecular GQs are comprised of at least two nucleic acid strands (Scheme 4.3). Here we highlight the differences between our intramolecular, GGG(TTAGGG)$_3$, and intermolecular, GGGTTAGGG, GQs.
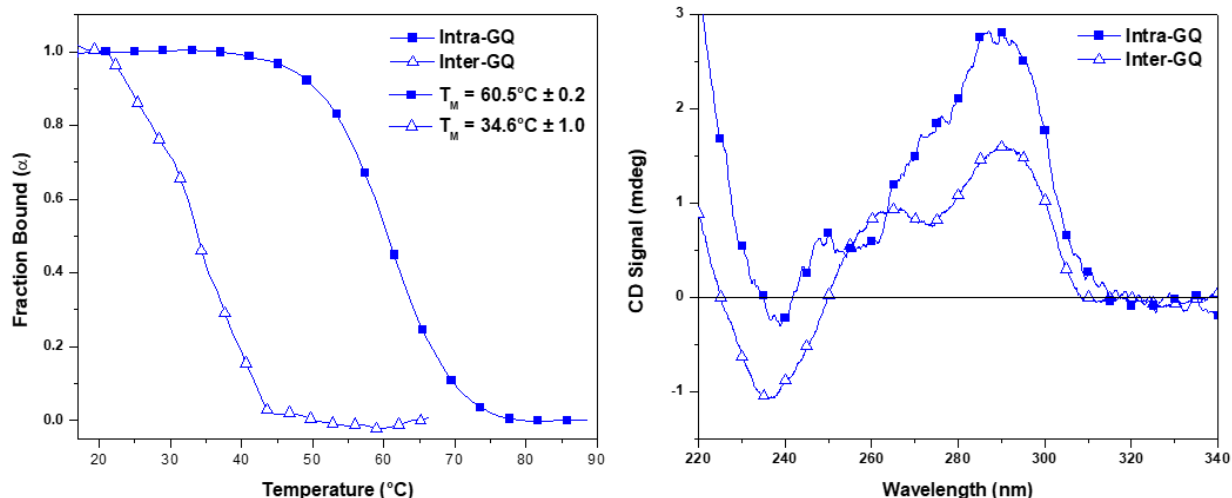
Figure 4.1 Thermal UV melting curves and circular dichroism (CD) of intramolecular (Intra-GQ, GGG(TTAGGG)$_3$) and intermolecular DNA (Inter-GQ, GGGTTAGGG). [GQ]=1 µM n=3. Melting temperature for intramolecular GQ = 60.5 °C ± 0.2 with antiparallel secondary structure and intermolecular GQ = 34.6 °C ± 1.0 with hybrid secondary structure.

Thermal melting curves indicate intramolecular GQs have significantly higher thermal stability than intermolecular GQs (ΔT_M= 26.5 °C). This is likely due to the entropic penalty paid by having two individual strands. Figure 4.1 shows the heating curve for each sample, however comparing the heating and cooling curve for each GQ is shown in Supporting Figure 4.3. These curves show a slight hysteresis in the Inter-GQ, which is typical for intermolecular structures. No hysteresis is observed for Intra-GQ. These DNA strands have vastly different secondary structures, indicated by circular dichroism (CD). Intra-GQ shows primarily antiparallel secondary structure, whereas inter-GQ shows hybrid secondary structure. This change in secondary structure indicates the number of strands is critical to the overall secondary structure of the DNA.

Intramolecular GQs using the human telomeric sequence are often reported in literature. Sequences with no overhangs GGG(TTAGGG)$_3$ have melting temperatures between 65-69 °C in 100 mM KCl determined by UV-Vis and CD[16–18]. The T_M identified here is slightly lower. Intermolecular, specifically bimolecular GQs show significantly lower melting temperatures than intramolecular GQs with similar sequences (39 °C for a 7 µM sample of GGGTTAGGG in 100 mM KCl)[18]. Again, the T_M identified here

is slightly lower. However, the melting temperatures of intermolecular GQs are concentration dependent[19]. Therefore, we would expect differences in the reported literature values.

Intramolecular telomeric GQs have received significant attention in the literature. Using CD, telomeric GQs with no overhangs (GGG(TTAGG)₃) generally possess antiparallel structure; sometimes with a slight shoulder at 260 nm, in 100 mM KCl[17,20]. However, in the presence of NaCl, the secondary structure shifts to a primarily parallel secondary structure[21]. Additional methods have been used to elucidate the secondary structure and loop structure of intramolecular GQs. Again, antiparallel secondary structure, with potential parallel shoulder present, is implicated for intramolecular telomeric GQs using structural NMR and ^{125}I radiolabeling[22–24]. It is worth noting that while these methods agree on the overall topology, they disagree over loop orientation[22–24]. In contrast, crystal structures of telomeric intramolecular GQs indicate parallel topology, although the authors address that this difference could be due to the crystallization conditions[25,26]. Intermolecular GQs have been less scrutinized. These dimers show hybrid structure, similar to that observed in our system; however, this indicates two unique structures. The first is a true hybrid, where three guanine tracks run up while one track runs down. The second is a mixing of parallel and antiparallel structure. Using CD and gel electrophoresis, these intermolecular GQs possess both parallel and antiparallel topologies, with both species occurring simultaneously[21,27].

4.3.2 *Dimeric GQs can be formed with or without a scaffold, impacting both thermal stability and secondary structure.*

In order to compare the dimeric GQs to the RNA:DNA heteroquadruplexes (RDQs) discussed previously, we must utilize a scaffold duplex to create the dimeric DNA:DNA (DDQ) and RNA:RNA (RRQ) heteroquaduplexes. Although we have previously identified the scaffold duplex is not required to form DDQ and RRQ, it is required in order to make an adequate comparison to the RDQ systems. Additionally, we can compare the scaffold DDQ and RRQ with the non-scaffold systems in order to

understand how the duplex scaffold impacts the formation of DDQ and RRQ in a manner that was not possible with the RDQ systems.
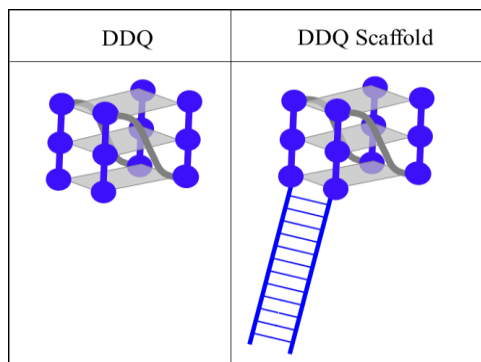
Scheme 4.4 Schematic representation of DNA:DNA dimeric homoGQs without (DDQ) and with a duplex scaffold (DDQ Scaffold).

The DDQ system was analyzed first (Scheme 4.4). Thermal melting curves converted to fraction bound show a significant increase in DDQ thermal stability in the presence of a duplex scaffold (ΔT_M=22 °C, Figure 4.2). The melting temperature of the duplex scaffold was calculated to be 56.3 ± 0.1 °C (black solid line, Figure 4.2), which accounts for the significant increase in the T_M of DDQ. Circular dichroism (CD) indicated a change in the secondary structure of DDQ.

DDQ possesses hybrid secondary structure with a slight preference for an antiparallel structure. DDQ-Scaffold shows more antiparallel nature, which concurs with the antiparallel runs imposed by the duplex scaffold (Figure 4.2). Overall, these results indicate the duplex scaffold increases thermal stability and limits secondary structure to a primarily antiparallel configuration.
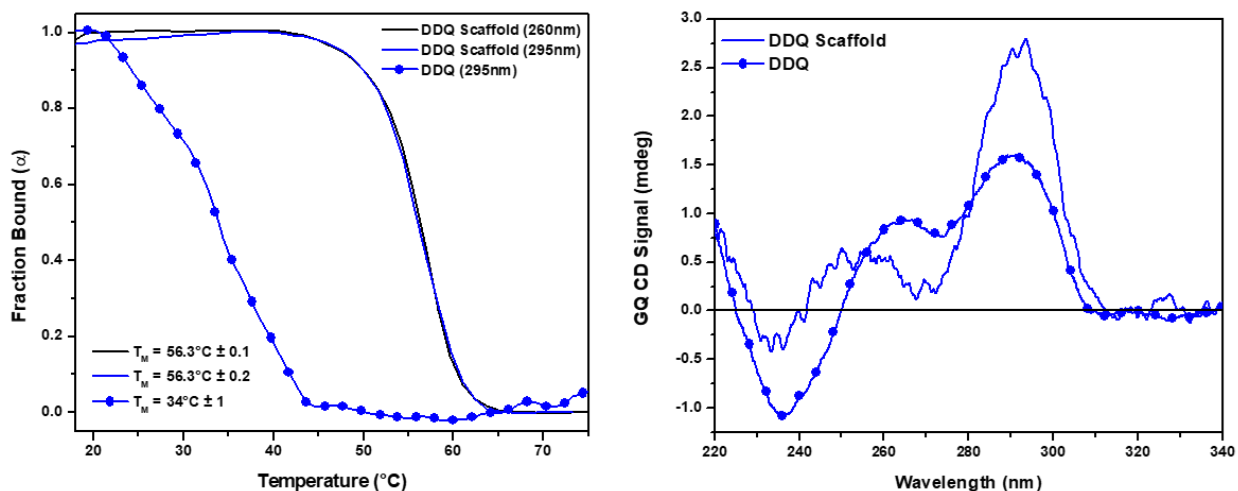
Figure 4.2 Thermal melting curves (left) for DDQ Scaffold (blue and black solid lines for 295 nm and 260 nm respectively) and DDQ (blue circles for 295 nm). Circular dichroism (CD) (right) of DDQ-Scaffold (blue solid line) and DDQ (blue circles). [GQ]=1 µM n=3
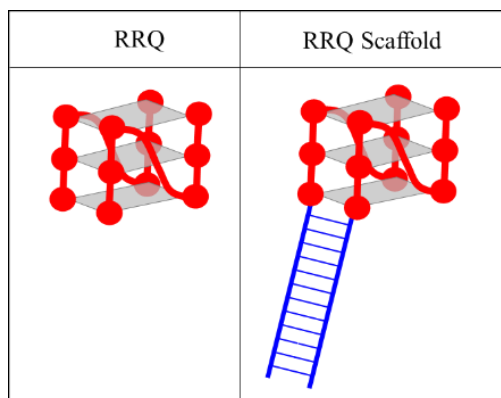
Scheme 4.5 Schematic representation of RNA:RNA dimeric homoGQs without (RRQ) and with a duplex scaffold (RRQ Scaffold).

RRQ is the second analyzed system; shown in Scheme 4.5. This system utilizes the RNA homologue of the previously discussed DDQ. Telomeric DNA is transcribed to RNA (TERRA) which is further responsible for the regulation of telomerase and heterochromatin formation[28]. Here we compare the RNA:RNA heteroquadruplex (RRQ) to its scaffold counterpart (RRQ-Scaffold). RRQ contains the homologue of the DDQ sequence (GGGUUAGGG), and the scaffold duplex is identical to that of the DDQ-Scaffold.
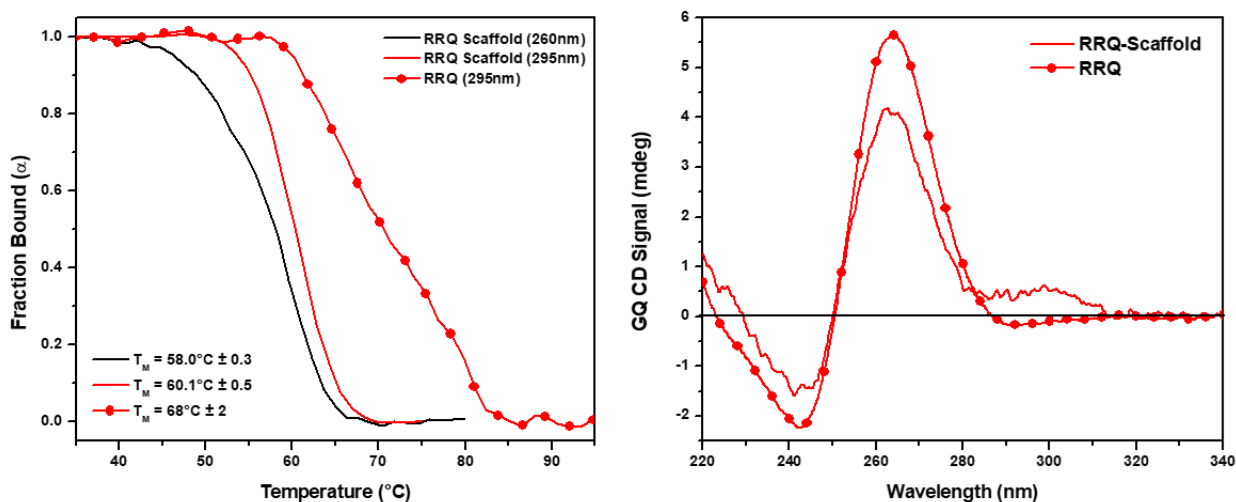
Figure 4.3 Thermal melting curves (left) for RRQ Scaffold (red and black solid lines for 295 nm and 260 nm respectively) and RRQ (red circles for 295 nm). Circular dichroism (CD) (right) of RRQ-Scaffold (red solid line) and RRQ (red circles). [GQ]=1 μM n=3

The melting temperature of RRQ was 68 °C. While the transition shown in Figure 4.3 is rather broad, the melting temperature corresponds well with previously reported literature[29]. The addition of a scaffold duplex decreased the melting temperature of RRQ by 8 °C. The scaffold duplex (T_M= 58 °C, black line Figure 4.3) is likely the cause of this decrease. The scaffold duplex destabilizes the RRQ because of

97

the stand orientation. This is indicated by the CD spectra. While the RRQ shows a fully parallel secondary structure, the RRQ-Scaffold shows decreased parallel signal at 260 nm and the emergence of a slight peak at 295 nm, indicative of antiparallel structure. This indicates the presence of the scaffold prevents RRQ-Scaffold from forming a preferred fully parallel structure. Furthermore, RNA GQs typically possess parallel secondary structure, due to the sterics imposed by the 2ʹOH[30].

4.3.3   *Using a duplex scaffold enables comparisons between RDQs and the dimeric homoGQs*

Although the scaffold duplex seems to play a significant role in the formation of dimeric homoGQs,
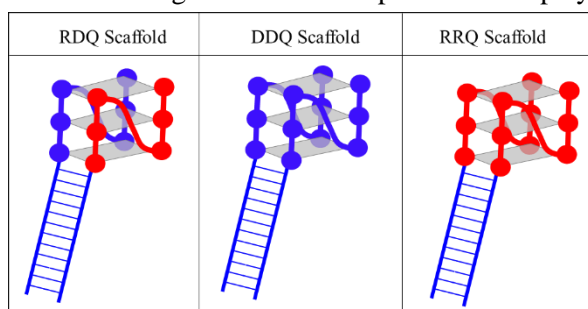
Scheme 4.6 Schematic representation of RNA:DNA heteroquadruplexes (RDQ-Scaffold), DNA:DNA dimeric homoGQs (DDQ-Scaffold), and RNA:RNA dimeric homoGQs (RRQ-Scaffold).

it is necessary to make comparisons between the dimeric homo- and heteroGQs (Scheme 4.6). Here we observe a telomere-based RDQ, which utilizes one DNA strand (GGGTTAGGG) and one RNA strand (GGGUUAGGG) and compare it to RRQ-Scaffold and DDQ-Scaffold.
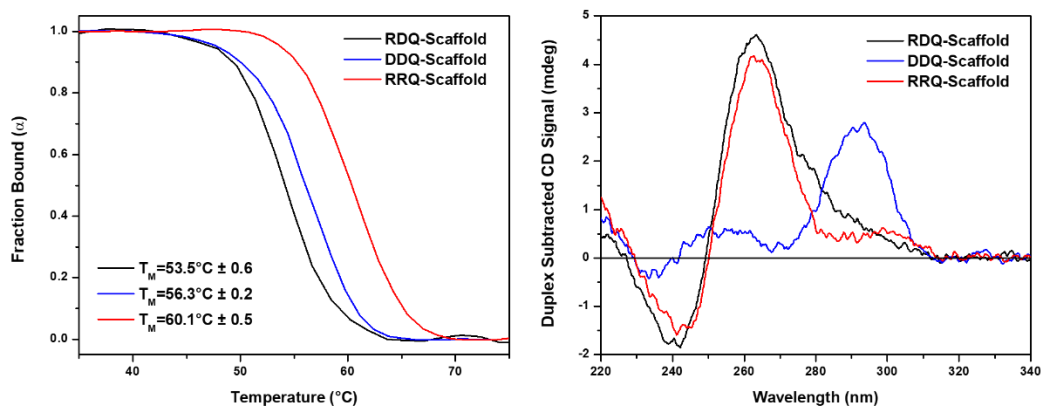
Figure 4.4 Thermal melting curves measured at 295 nm (left) and circular dichroism (CD) (right) of RDQ-Scaffold (black), DDQ-Scaffold (blue), and RRQ-Scaffold (red). [GQ]=1 µM. n=3

Interestingly, both DDQ-Scaffold and RRQ-Scaffold have increased melting temperatures compared to the RDQ-Scaffold. While RDQ-Scaffold had a T_M (at 295 nm) of 53.5 °C, DDQ-Scaffold and RRQ-Scaffold had T_{MS} of 56.3 °C and 60.1 °C respectively. The 6.6 °C difference in melting temperature between RDQ-Scaffold and RRQ-Scaffold was expected, since RNA GQs are generally known to have higher stability than GQs containing DNA. In contrast, we would have expected the DDQ-Scaffold GQ to have a decreased melting temperature compared to RDQ-Scaffold (instead of a 2.8 °C increase). The orientation of these GQs likely plays a role in their stability. CD spectra indicate our RDQ-Scaffold construct closely matches the RRQ-Scaffold system, with primarily parallel secondary structure. However, the RDQ-Scaffold possesses a slight shoulder around 290 nm that is not observed in RRQ-Scaffold. This hybrid structure likely arises from the DNA tracts of RDQ-Scaffold. RNA bases typically form parallel structures and the orientation of the guanine tracts adjacent to the duplex in the RRQ-Scaffold are likely forced into an antiparallel orientation. This could destabilize the formation of RRQ-Scaffold. Similar structural constraints exist for RDQ-Scaffold; however, the parallel secondary structure conflicts with these constraints. This strained parallel orientation likely leads to a decrease in melting temperature.

The melting temperature of the scaffold duplex also likely plays a role in the stability of the quadruplex, since the T_M values of the duplex (260 nm) are in the same order as the quadruplex (Figure 4.5). Even though the scaffold duplex is identical across each system, there is a 6° difference between them. The RDQ-Scaffold duplex melts first at 52.8 °C, followed by DDQ-Scaffold at 56 °C, and then the RRQ-Scaffold at 58.0 °C. The reason for the difference in duplex stability is unclear, however, it does provide some explanation for the inflated melting temperature of the DDQ-Scaffold GQ.
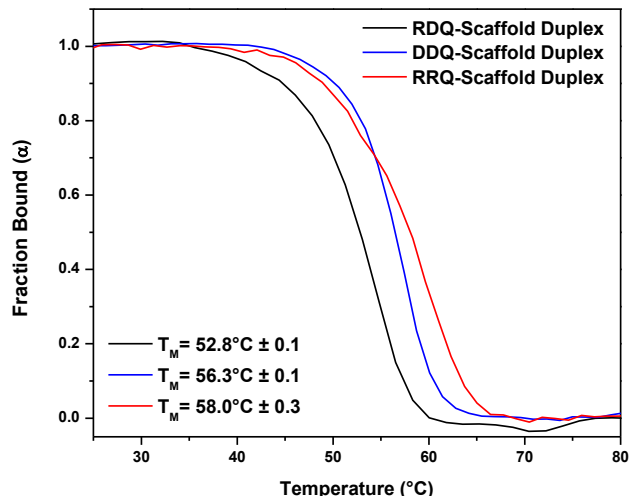
Figure 4.5 Thermal melting curves monitored at 260 nm of RDQ-Scaffold (black), DDQ-Scaffold (blue), and RRQ-Scaffold (red). [GQ]=1 µM n=3

4.3.4   *Fluorescence titrations suggest scaffold systems have similar binding with Thioflavin T compared to intramolecular GQs.*

Thioflavin T (ThT) is a known GQ binding dye that has been used to study both RNA and DNA GQs[31,32]. Specifically, ThT is a light-up probe, where enhanced fluorescence occurs when the benzylamine and benzothiazole rings of ThT are immobilized upon binding to a GQ. ThT has shown a slight increase in binding antiparallel DNA GQs compared to parallel DNA GQs, implying that ThT might have a tighter binding to DNA compared to RNA GQs[31,33]. However, calculated dissociation constants do not show distinct differences between DNA and RNA[32]. Our study is the first direct comparison of homologous DNA and RNA sequences.

Fluorescence titrations with Thioflavin T (ThT) allow the determination of dissociation constants (Figure 4.6 and Table 4.3). Here, we compare our intramolecular GQ and the three scaffold dimeric GQs. DDQ-Scaffold had the lowest dissociation constant (0.77 µM). RDQ-Scaffold, Intra-GQ, and RRQ-Scaffold followed with dissociation constants less than 10 µM (3.8, 4.4, and 6.4 µM respectively).

100
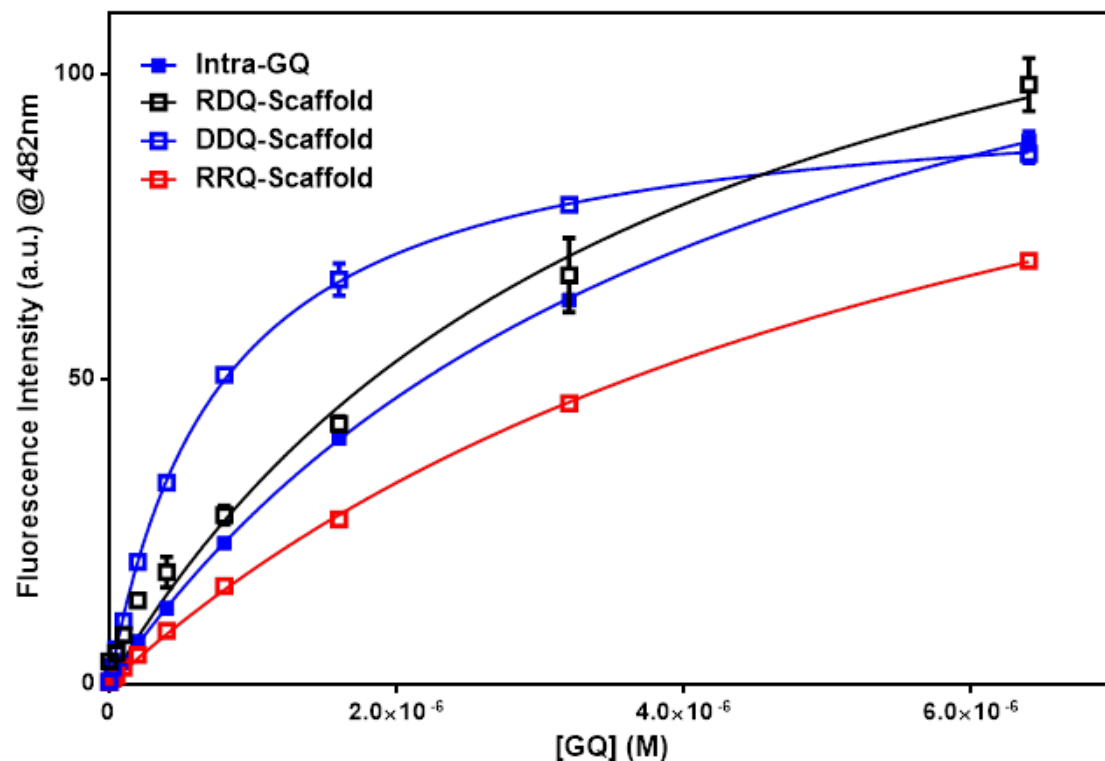
Figure 4.6 Fluorescence binding assay between each GQ and ThT in 100 mM KCl. [ThT]=200 nm [RDQ]= 0-6.4 µM. ThT was excited at 430 nm and emission was recorded at 482 nm. Kd values were calculated using a one site binding model.

Table 4.3 K_d values calculated from fluorescence binding assay between each GQ and ThT in 100 mM KCl.

	K_d (µM)
Intra-GQ	4.4 ± 0.8
RDQ-Scaffold	3.8 ± 0.8
DDQ-Scaffold	0.77 ± 0.02
RRQ-Scaffold	6.4 ± 0.3

4.3.5 *γPNA successfully invades RDQ-Scaffold, DDQ-Scaffold, and RRQ-Scaffold.*

Targeting GQ structures is crucial to modulating their function. The high affinity and ability to invade GQs makes complementary γPNA an excellent choice as an RDQ probe. γPNA can be synthesized to target either RNA or DNA distinguished by the scaffold duplex segments (for this system, RNA is

targeted as shown in Scheme 4.7). Here, we utilize complementary γPNA to target and disrupt the above hetero- and homoGQs (RDQ-Scaffold, DDQ-Scaffold, and RRQ-Scaffold).
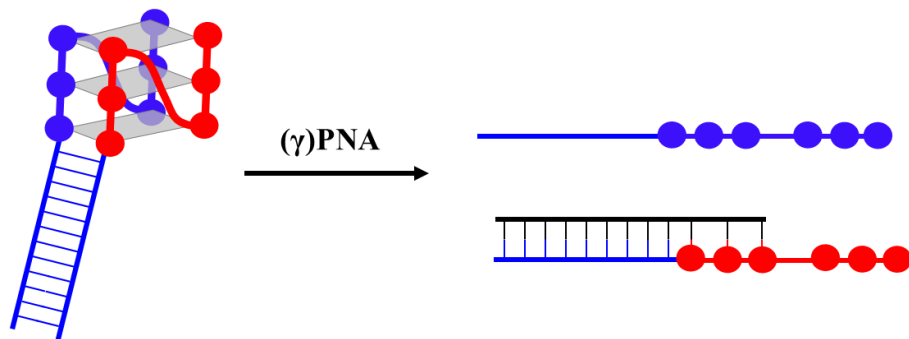
Scheme 4.7 Scheme showing (γ)PNA disruption of RDQs. (γ)PNA, shown in black, hybridizes to the RNA strand and invades one guanine tract and the scaffold duplex.

The γPNA used here is a 15mer that is complementary to the first three guanines of the quadruplex and the subsequent 12 bases of the scaffold duplex. The invasion of the first three guanines is sufficient for disruption of the quadruplex. The invasion of the scaffold duplex ensures separation between the bimolecular GQ. Further, this design allows us to target all of our bimolecular GQs using the same γPNA sequence. Thioflavin T, ThT, can be used to monitor the invasion of γPNA. Prior to adding γPNA, ThT binds to a stable GQ and fluoresces. As γPNA is added to the system, it disrupts the GQ and fluorescence should decrease. In this section, we will compare the RDQ-Scaffold, DDQ-Scaffold, and RRQ-Scaffold systems.
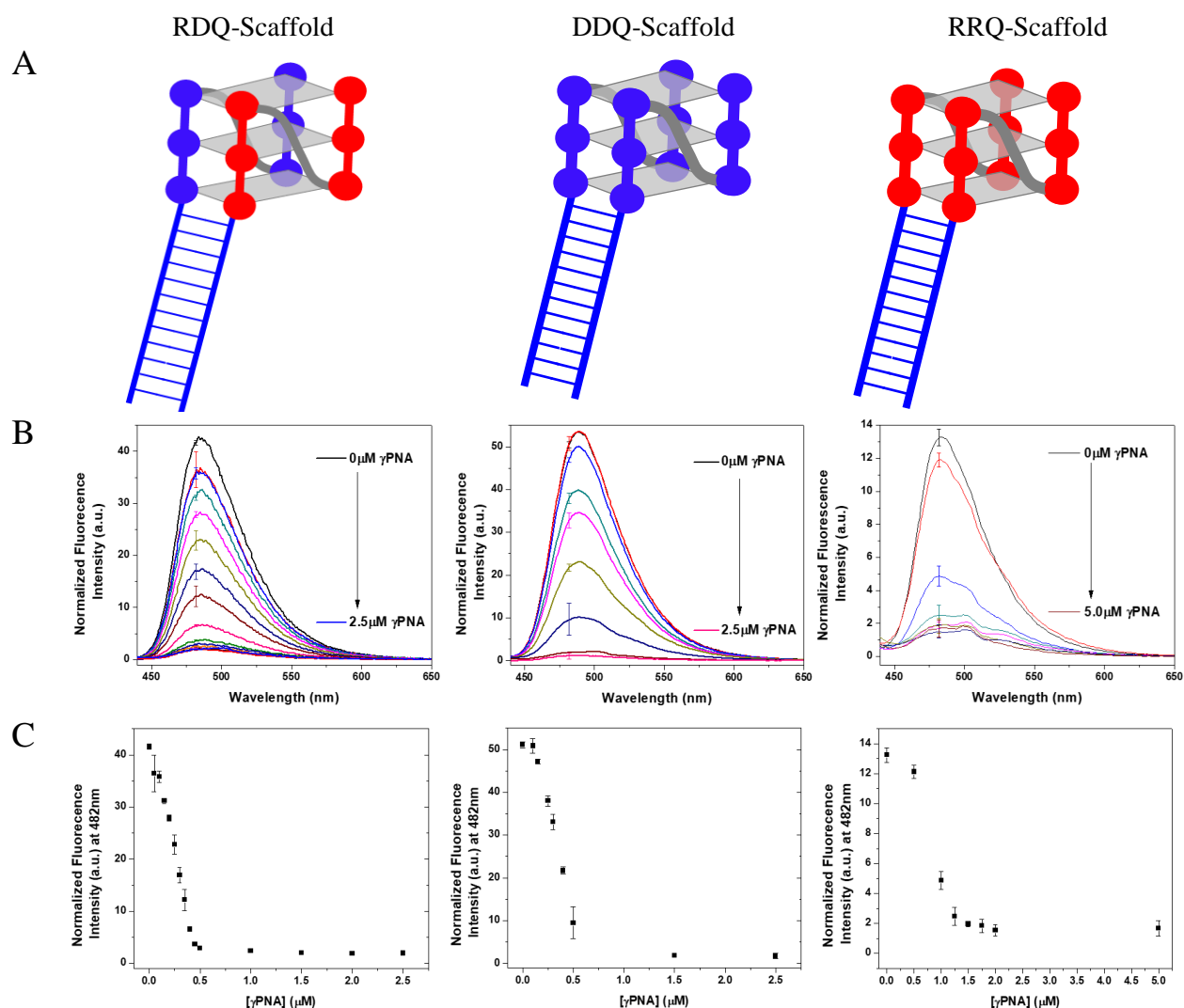
Figure 4.7 (A) RDQ-Scaffold, DDQ-Scaffold, and RRQ-Scaffold systems, (B) Fluorescence γPNA Titration scans 440-750nm, and (C) Fluorescence intensity at 482 nm vs. [γPNA] in 100mM KCl. [ThT]= 0.5µM [RDQ]= 0.5µM [γPNA]= 0-2.5µM. n = 3. ThT was excited at 430nm and emission maxima were recorded at 482nm.

Table 4.4 IC_{50} values of RDQ-Scaffold, DDQ-Scaffold, and RRQ-Scaffold calculated from fluorescence titrations by converting to Fraction Bound (shown in Supporting Figure 4.4) in 100 mM KCl.

	IC_{50} (µM)
RDQ-Scaffold	0.27
DDQ-Scaffold	0.35
RRQ-Scaffold	0.82

γPNA successfully disrupted RDQ-Scaffold with near baseline fluorescence, up to 2.5 µM γPNA, and an IC_{50} value of 0.27 µM. RDQ-Scaffold showed a strong decrease, even in the presence of 50 nm

γPNA. DDQ-Scaffold also showed strong disruption with an IC_{50} value of 0.35 µM, reaching baseline fluorescence at 2.5 µM γPNA. RRQ-Scaffold had a significantly lower starting fluorescence than the other two systems. A decrease in fluorescence for the RRQ-Scaffold system was expected based on the fluorescence titrations and K_d values determined in Section 4.3.4. Baseline fluorescence was achieved with an increase in γPNA, up to 5 µM, and an IC50 value of 0.82 µM was calculated.

4.3.6   *γPNA successfully invades RDQ-Scaffold constructs (R2D2-Telo, R1D3-Telo, and R1D3-CSB).*

To further our understanding of γPNA's ability to disrupt our scaffold GQs, we conducted fluorescence titrations using the RDQ constructs explored in Chapter 3. These RDQs contain differing sequences and utilize different numbers of DNA and RNA guanine tracts to make up the GQ (Figure 4.8 A). R2D2-Telo, also called RDQ-Scaffold earlier in this chapter, is a telomere-based 2:2 RDQ (meaning two guanine tracts are from RNA and two tracts are from DNA). R1D3-Telo is a second telomere-based RDQ where one guanine tract is from RNA and three tracts from DNA. The last RDQ is R1D3-CSB. This RDQ has one RNA guanine tract and three DNA guanine tracts; the sequence represents a biologically implicated RDQ found in the CSB-II region of human mitochondrial DNA. Again, we are able to use ThT to monitor RDQ disruption by γPNA.
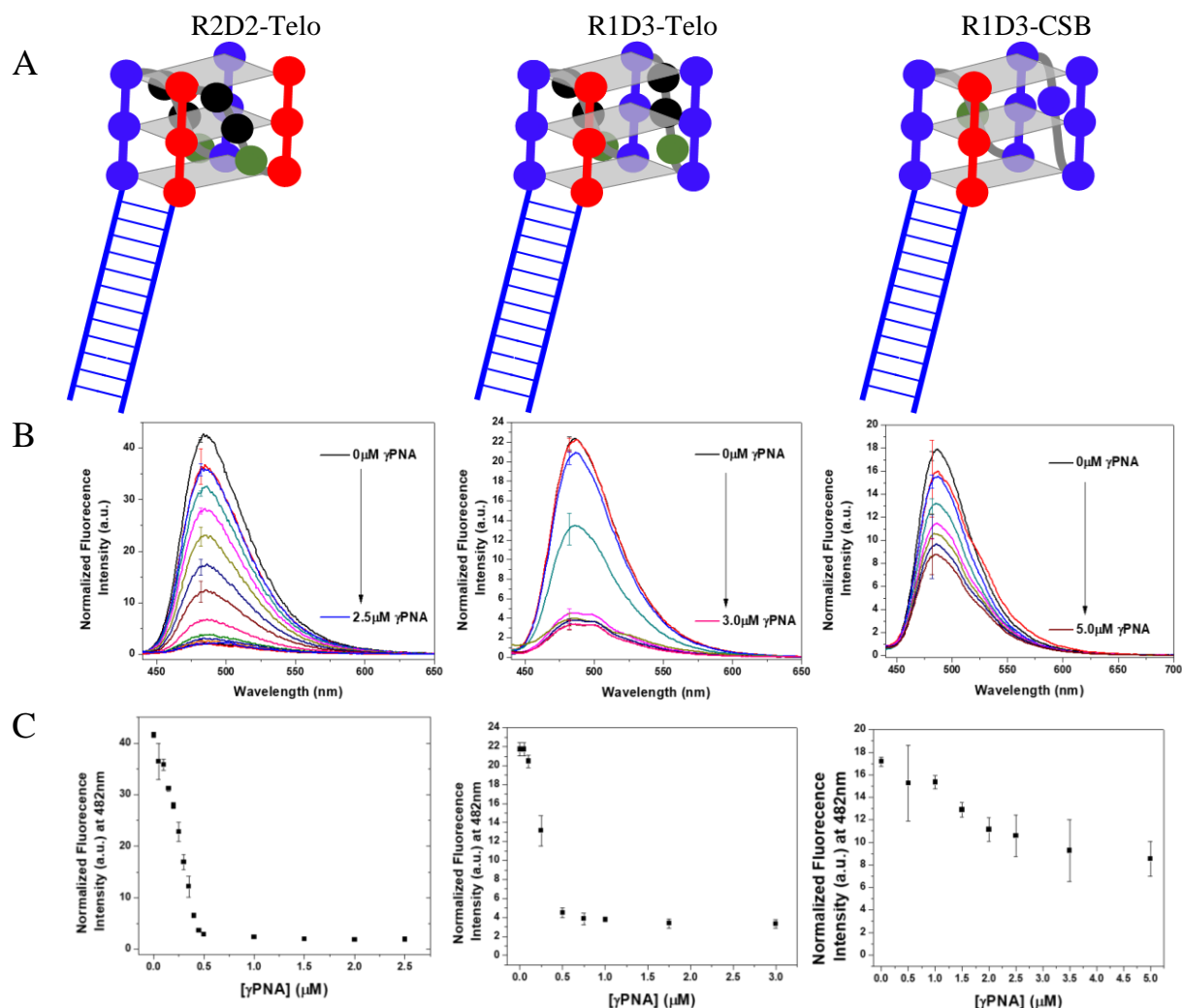
Figure 4.8 (A) R2D2-Telo, R1D3-Telo, and R1D3-CSB systems, (B) Fluorescence γPNA Titration scans 440-750nm, and (C) Fluorescence intensity at 482 nm vs. [γPNA] in 100mM KCl. [ThT]= 0.5µM [RDQ]= 0.5µM [γPNA]= 0-2.5µM. n = 3. ThT was excited at 430nm and emission maxima were recorded at 482nm.

Table 4.5 IC$_{50}$ values of R2D2-Telo, R1D3-Telo, and R1D3-CSB calculated from RDQ fluorescence titrations by converting to Fraction Bound (shown in Supporting Figure 4.4) in 100 mM KCl.

	IC$_{50}$ (µM)
RDQ-Scaffold (R2D2-Telo)	0.27
R1D3-Telo	0.26
R1D3-CSB	1.5

γPNA disrupted formation of each RDQ. R2D2-Telo was fully disrupted at 2.5 µM γPNA in the presence of 0.5 µM RDQ and ThT (IC$_{50}$ = 0.27 µM). R1D3-Telo also showed significant disruption, with an IC$_{50}$ value 0.26 µM; however, a slightly elevated baseline remained. R1D3-CSB showed some disruption

(approximately 50%) in the presence of increased γPNA concentration, with a significantly elevated baseline (IC_{50} = 1.5 µM). The remaining ThT signal indicated the presence of a quadruplex even at 10X γPNA.

In order to understand the fluorescence signal observed at high PNA concentrations in R1D3-CSB RDQ, we measured fluorescence of the DNA alone for each of our RDQs. The DNA portion of each construct contains at least two guanine tracks. One could imagine when γPNA and RNA bind, the guanine regions of the released DNA could form tetrads and therefore inflate the fluorescence observed by ThT. To test this, we preannealed (95 °C for 5mins and then slowly cooled to room temperature) the DNA from each construct alone, added ThT, and observed the fluorescence. R1D3-CSB DNA had significantly increased fluorescence over R2D2-Telo and R1D3-Telo DNA (Figure 4.9 left). This is likely a result of the sequence of this DNA (GGGAGGGGGGG). Additionally, R1D3-Telo DNA shows a slight increase in fluorescence intensity that could explain the slightly elevated baseline observed in the titration. These results impact how we interpret the fluorescence titration results, since fluorescence from the DNA alone contributes background signal.
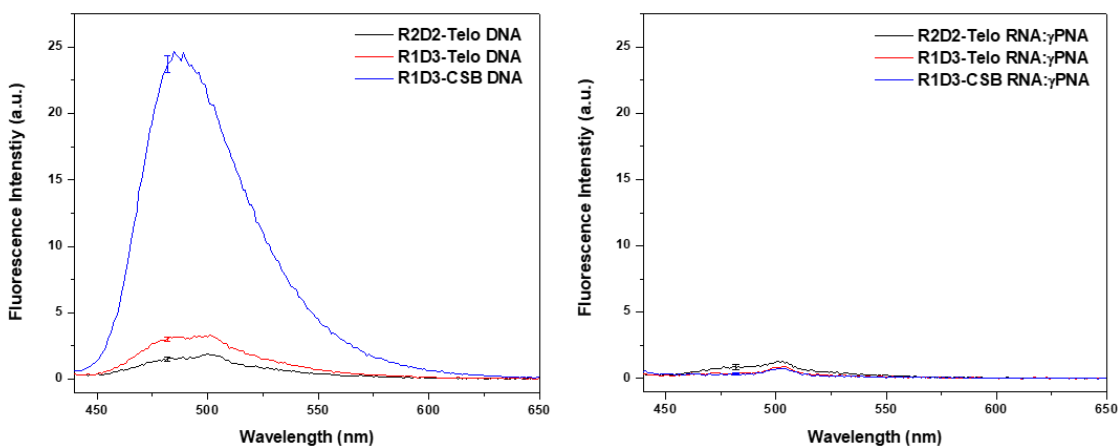
Figure 4.9 Fluorescence of DNA alone (left) and RNA:γPNA duplex (right) in 100 mM KCl. [ThT]= 0.5 µM [RDQ DNA]= 0.5 µM/ [RNA:γPNA duplex]= 0.5 µM. ThT was excited at 430 nm and emission maxima were recorded at 482 nm.

We then confirmed no additional ThT fluorescence was originating from the RNA:γPNA duplex. In order to accomplish this, each RNA and γPNA were preannealed followed by the addition of ThT. These results show little fluorescence enhancement (seen in Figure 4.9 scaled to the DNA signal). A zoomed version, Figure 4.10, shows no signal enhancement from R1D3-Telo and R1D3-CSB compared to ThT alone. R2D2-Telo shows a very slight increase in fluorescence; however, this is considered nominal.
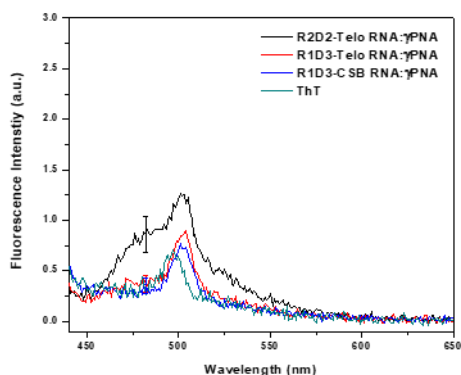
Figure 4.10 Fluorescence of DNA alone in 100mM KCl. [ThT]= 0.5µM [RDQ DNA]= 0.5µM. ThT was excited at 430nm and emission maxima were recorded at 482nm.

4.4 Discussion

This chapter analyzes differences in monomeric, dimeric hetero-, and dimeric homoquadruplexes. We first look at the differences between intramolecular and intermolecular (or dimeric) quadruplexes. There were significant differences in thermal stability and secondary structure. Our intramolecular DNA GQ had a 26.5 °C higher melting temperature than the intermolecular GQ. Additionally, the intramolecular GQ showed primarily antiparallel secondary structure, whereas the intermolecular GQ possessed hybrid secondary structure.

Our intermolecular DNA:DNA homoquadruplex was further studied using a duplex scaffold. This duplex scaffold confined the secondary structure to primarily antiparallel, instead of hybrid secondary structure for DDQ, and increased the melting temperature by 22 °C compared to DDQ. In contrast, the presence of a duplex scaffold decreased the melting temperature of the RNA:RNA homoquadruplex by 8 °C. The secondary structure of RRQ-Scaffold showed decreased parallel structure and a slight increase in antiparallel structure compared to RRQ. This increase in antiparallel structure is likely because the construct requires antiparallel stands (5' – 3' and 3' – 5') due to the duplex scaffold. RDQs composed of an RNA transcript and the non-template DNA strand could run parallel to one another (5' – 3' and 5' – 3'), although

107

it should be noted that the overall orientation of *in vivo* RDQs is not well understood. It would be possible for RDQ strands to run antiparallel, given sufficient single-stranded DNA and RNA. Even though the duplex scaffold affected the stability of DDQ-Scaffold and RRQ-Scaffold, it was necessary for comparing to the previously studied RNA:DNA heteroquadruplexes (RDQ-Scaffold).

Comparing scaffold systems allowed us to draw conclusions about the nature of RDQs compared to DDQs and RRQs. Surprisingly, RDQ-Scaffold had the lowest melting temperature, 53.5 °C, with DDQ-Scaffold following at 56.3 °C. RRQ-Scaffold had the highest melting temperature of 60.1 °C. These melting temperatures correspond well with the melting temperatures of the scaffold duplex, indicating the scaffold duplex plays a significant role in forming these GQs. Since the systems are high temperature annealed in all subsequent fluorescence titrations, this role was not analyzed further. Secondary structure analysis indicated that DDQ-Scaffold formed a primarily antiparallel GQ while RRQ-Scaffold and RDQ-Scaffold formed primarily parallel GQs. This was expected, since GQs containing RNA are typically parallel.

Thioflavin T (ThT) is a GQ-selective fluorophore that has been utilized to study both DNA and RNA intramolecular GQs. The results shown here are the first comparison of intra- and intermolecular GQs, as well as the first comparison of homo- and hetero- GQs. The calculated dissociation constants show intramolecular GQs have similar binding affinities to ThT as the RDQ-Scaffold and RRQ-Scaffold systems. DDQ-Scaffold had an exceptionally low dissociation constant. The tight binding could be a result of the DDQ-Scaffold orientation. ThT has shown a slight preference for antiparallel GQs[34]. Since DDQ-Scaffold is primarily antiparallel, it could lead to increased binding affinity between the ThT and DDQ-Scaffold.

We then utilize ThT in order to monitor γPNA invasion of the GQ. This requires two distinct steps. First, the GQ system with or without γPNA is preannealed, heated to 95 °C for five minutes and then slowly cooled to room temperature. This allows complete binding between γPNA and its complement. Next, ThT is added to the γPNA:GQ mixture and allowed to equilibrate for five minutes prior to a fluorescence scan. The incremental increase of γPNA to the system shows decreases in fluorescence because ThT is only fluorescent in the presence of GQs. All of the telomere-based sequences have IC_{50} values in the sub-micromolar range. R1D3-Telo and R2D2-Telo had the lowest IC_{50} values; 0.26 and 0.27 µM respectively.

DDQ-Scaffold and RRQ-Scaffold, the intermolecular homoGQs, had higher IC_{50} values of 0.35 and 0.82 µM. R1D3-CSQ showed the highest IC_{50} value; however, there is likely some error in this value since the DNA alone is able to form a GQ structure, as evidenced by increased fluorescence in the DNA alone. This sequence could be mutated for future studies, or a secondary probe could be used to prevent GQ formation from the DNA.

4.5   Future Directions

Intermolecular GQ structures can have vastly different secondary structures and thermal stabilities, even if they arise from identical repeating sequences. Even though dimeric DNA and RNA GQs (DDQ and RRQ) can form without assistance from a scaffold, we utilize a scaffold to compare these GQs to the RDQs studied in Chapter 3. We are able to target these scaffold systems using γPNA. Future work aims to continue developing these γPNA probes. All of the targeting data shown here is conducted using high temperature anneals. In order to understand how these probes would interact with RDQ systems *in vivo*, further experiments expanding upon the kinetic properties of the probes will be helpful. γPNA probes will hopefully lead to the selective targeting of RDQs, which could impact telomere maintenance and transcription using the sequences studied here.

4.6   References

(1)   Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. Quadruplex DNA: Sequence, Topology and Structure. *Nucleic Acids Res.* **2006**, *34* (19), 5402–5415.
(2)   Guédin, A.; Gros, J.; Alberti, P.; Mergny, J.-L. How Long Is Too Long? Effects of Loop Size on G-Quadruplex Stability. *Nucleic Acids Res.* **2010**, *38* (21), 7858–7868.
(3)   Kwok, C. K.; Sherlock, M. E.; Bevilacqua, P. C. Effect of Loop Sequence and Loop Length on the Intrinsic Fluorescence of G-Quadruplexes. *Biochemistry* **2013**, *52* (18), 3019–3021.
(4)   Vorlíčková, M.; Bednářová, K.; Kejnovská, I.; Kypr, J. Intramolecular and Intermolecular Guanine Quadruplexes of DNA in Aqueous Salt and Ethanol Solutions. *Biopolymers* **2007**, *86* (1), 1–10.
(5)   Kim, M.-Y.; Gleason-Guzman, M.; Izbicka, E.; Nishioka, D.; Hurley, L. H. The Different Biological Effects of Telomestatin and TMPyP4 Can Be Attributed to Their Selectivity for Interaction with Intramolecular or Intermolecular G-Quadruplex Structures. *Cancer Res.* **2003**, *63* (12), 3247–3256.
(6)   Cang, X.; Šponer, J.; Cheatham, T. E.; III. Explaining the Varied Glycosidic Conformational, G-

Tract Length and Sequence Preferences for Anti-Parallel G-Quadruplexes. *Nucleic Acids Res.* **2011**, *39* (10), 4499–4512.

(7)   Datta, B.; Armitage, B. A. Hybridization of PNA to Structured DNA Targets: Quadruplex Invasion and the Overhang Effect. *J. Am. Chem. Soc.* **2001**, *123* (39), 9612–9619.

(8)   Marin, V. L.; Armitage, B. A. RNA Guanine Quadruplex Invasion by Complementary and Homologous PNA Probes. *J. Am. Chem. Soc.* **2005**, *127* (22), 8032–8033.

(9)   Marin, V. L.; Armitage, B. A. Hybridization of Complementary and Homologous Peptide Nucleic Acid Oligomers to a Guanine Quadraplex-Forming RNA. *Biochemistry* **2006**, *45* (6), 1745–1754.

(10)   Gupta, A.; Lee, L. L.; Roy, S.; Tanious, F. A.; Wilson, W. D.; Ly, D. H.; Armitage, B. A. Strand Invasion of DNA Quadruplexes by PNA: Comparison of Homologous and Complementary Hybridization. *ChemBioChem* **2013**, *14* (12), 1476–1484.

(11)   Oyaghire, S. N.; Cherubim, C. J.; Telmer, C. A.; Martinez, J. A.; Bruchez, M. P.; Armitage, B. A. RNA G-Quadruplex Invasion and Translation Inhibition by Antisense γ-Peptide Nucleic Acid Oligomers. *Biochemistry* **2016**, *55* (13), 1977–1988.

(12)   Zheng, K.; Wu, R.; He, Y.; Xiao, S.; Zhang, J.; Liu, J.; Hao, Y.; Tan, Z. A Competitive Formation of DNA:RNA Hybrid G-Quadruplex Is Responsible to the Mitochondrial Transcription Termination at the DNA Replication Priming Site. *Nucleic Acids Res.* **2014**, *42* (16), 10832–10844.

(13)   Amato, J.; Oliviero, G.; De Pauw, E.; Gabelica, V. Hybridization of Short Complementary PNAs to G-Quadruplex Forming Oligonucleotides: An Electrospray Mass Spectrometry Study. *Biopolymers* **2009**, *91* (4), 244–255.

(14)   Sahu, B.; Sacui, I.; Rapireddy, S.; Zanotti, K. J.; Bahal, R.; Armitage, B. A.; Ly, D. H. Synthesis and Characterization of Conformationally Preorganized, (R)-Diethylene Glycol-Containing γ-Peptide Nucleic Acids with Superior Hybridization Properties and Water Solubility. *J. Org. Chem.* **2011**, *76* (14), 5614–5627.

(15)   Samassekou, O.; Gadji, M.; Drouin, R.; Yan, J. Sizing the Ends: Normal Length of Human Telomeres. *Ann. Anat. - Anat. Anzeiger* **2010**, *192* (5), 284–291.

(16)   Chaires, J. B. Human Telomeric G-Quadruplex: Thermodynamic and Kinetic Studies of Telomeric Quadruplex Stability. *FEBS J.* **2010**, *277* (5), 1098–1106.

(17)   Li, W.; Wu, P.; Ohmichi, T.; Sugimoto, N. Characterization and Thermodynamic Properties of Quadruplex/Duplex Competition. *FEBS Lett.* **2002**, *526* (1–3), 77–81.

(18)   Mergny, J. L.; Phan, A. T.; Lacroix, L. Following G-Quartet Formation by UV-Spectroscopy. *FEBS Lett.* **1998**, *435* (1), 74–78.

(19)   Mergny, J.-L.; Lacroix, L. *Analysis of Thermal Melting Curves*; 2003; Vol. 13.

(20)   Tran, P. L. T.; Mergny, J.-L.; Alberti, P. Stability of Telomeric G-Quadruplexes. *Nucleic Acids Res.* **2011**, *39* (8), 3282–3294.

(21)   Vorlickova, M.; Chládková, J.; Kejnovská, I.; Fialová, M.; Kypr, J. Guanine Tetraplex Topology of Human Telomere DNA Is Governed by the Number of (TTAGGG) Repeats. *Nucleic Acids Res.* **2005**, *33* (18), 5851–5860.

(22)   Wang, Y.; Patel, D. J. Solution Structure of the Human Telomeric Repeat d[AG3(T2AG3)3] G-Tetraplex. *Structure* **1993**, *1* (4), 263–282.

(23)   He, Y.; Neumann, R. D.; Panyutin, I. G. Intramolecular Quadruplex Conformation of Human Telomeric DNA Assessed with 125I-Radioprobing. *Nucleic Acids Res.* **2004**, *32* (18), 5359–5367.

(24)   Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D. Human Telomeric Sequence Forms a Hybrid-Type Intramolecular G-Quadruplex Structure with Mixed Parallel/Antiparallel Strands in Potassium Solution. *Nucleic Acids Res.* **2006**, *34* (9), 2723–2735.

(25)   Parkinson, G. N.; Lee, M. P. H.; Neidle, S. Crystal Structure of Parallel Quadruplexes from Human Telomeric DNA. *Nature* **2002**, *417* (6891), 876–880.

(26)   Li, J.; Correia, J. J.; Wang, L.; Trent, J. O.; Chaires, J. B. Not so Crystal Clear: The Structure of the Human Telomere G-Quadruplex in Solution Differs from That Present in a Crystal. *Nucleic Acids Res.* **2005**, *33* (14), 4649–4659.

(27) Pedroso, I. M.; Duarte, L. F.; Yanez, G.; Burkewitz, K.; Fletcher, T. M. Sequence Specificity of Inter- and Intramolecular G-Quadruplex Formation by Human Telomeric DNA. *Biopolymers* **2007**, *87* (1), 74–84.

(28) Cusanelli, E.; Chartrand, P. Telomeric Repeat-Containing RNA TERRA: A Noncoding RNA Connecting Telomere Biology to Genome Integrity. *Front. Genet.* **2015**, *6*, 143.

(29) Bao, H.-L.; Ishizuka, T.; Sakamoto, T.; Fujimoto, K.; Uechi, T.; Kenmochi, N.; Xu, Y. Characterization of Human Telomere RNA G-Quadruplex Structures in Vitro and in Living Cells Using 19F NMR Spectroscopy. *Nucleic Acids Res.* **2017**, *45* (9), 5501–5511.

(30) Fay, M. M.; Lyons, S. M.; Ivanov, P. RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms. *J. Mol. Biol.* **2017**, *429* (14), 2127–2147.

(31) Gabelica, V.; Maeda, R.; Fujimoto, T.; Yaku, H.; Murashima, T.; Sugimoto, N.; Miyoshi, D. Multiple and Cooperative Binding of Fluorescence Light-up Probe Thioflavin T with Human Telomere DNA G-Quadruplex. *Biochemistry* **2013**, *52* (33), 5620–5628.

(32) Xu, S.; Li, Q.; Xiang, J.; Yang, Q.; Sun, H.; Guan, A.; Wang, L.; Liu, Y.; Yu, L.; Shi, Y.; et al. Thioflavin T as an Efficient Fluorescence Sensor for Selective Recognition of RNA G-Quadruplexes. *Sci. Rep.* **2016**, *6*.

(33) Verma, S.; Ghuge, S. A.; Ravichandiran, V.; Ranjan, N. Spectroscopic Studies of Thioflavin-T Binding to c-Myc G-Quadruplex DNA. *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* **2019**, *212*, 388–395.

(34) Mohanty, J.; Barooah, N.; Dhamodharan, V.; Harikrishna, S.; Pradeepkumar, P. I.; Bhasikuttan, A. C. Thioflavin T as an Efficient Inducer and Selective Fluorescent Sensor for the Human Telomeric G-Quadruplex DNA. *J. Am. Chem. Soc.* **2013**, *135* (1), 367–376.

(35) Jin, R.; Gaffney, B. L.; Wang, C.; Jones, R. A.; Breslauer, K. J. Thermodynamics and Structure of a DNA Tetraplex: A Spectroscopic and Calorimetric Study of the Tetramolecular Complexes of d(TG3T) and d(TG3T2G3T). *Proc. Natl. Acad. Sci.* **2006**, *89* (18), 8832–8836.

(36) Viglasky, V.; Bauer, L.; Tluckova, K.; Javorsky, P. Evaluation of Human Telomeric G-Quadruplexes: The Influence of Overhanging Sequences on Quadruplex Stability and Folding. *J. Nucleic Acids* **2010**, *2010*.

4.7   Supporting Figures
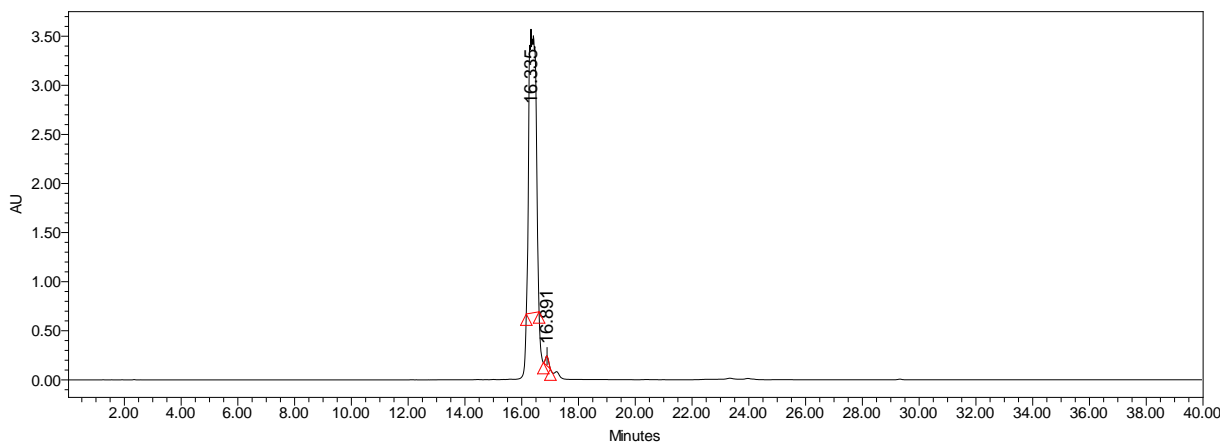
Supporting Figure 4.1. γPNA after purification via high-performance liquid chromatography (HPLC). γPNA was injected into the HPLC and run through the semi-prep column at a flow rate of 10 mL/min with gradient of 95-40% water/acetonitrile with 0.1% TFA for 40 minutes. Column temperature: 50 °C.
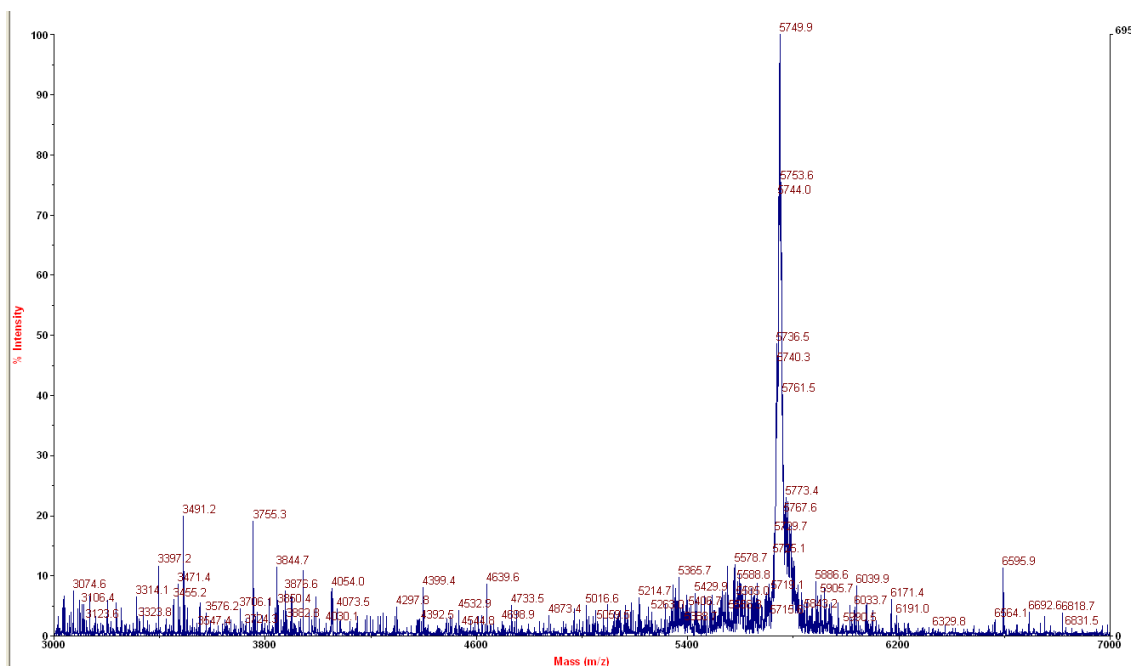
Supporting Figure 4.2 Mass spectra of γPNA using matrix assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF). γPNA was mixed with matrix (α-cyano-4-hydroxycinnamic acid as the matrix dissolved in 50:50 H_2O/ACN with 0.1% TFA). Expected mass: 5761g/mol observed mass: 5749g/mol adjusted to 5755g/mol via an external standard.
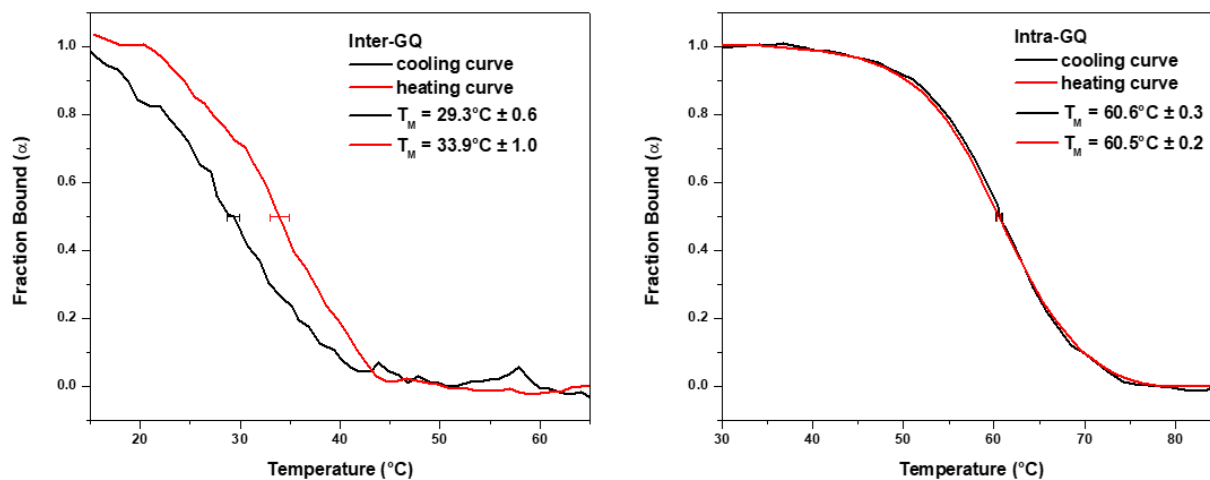
Supporting Figure 4.3 Thermal UV melting heating and cooling curves for Inter-GQ on left and Intra-GQ on right. [GQ]=1 μM n=3
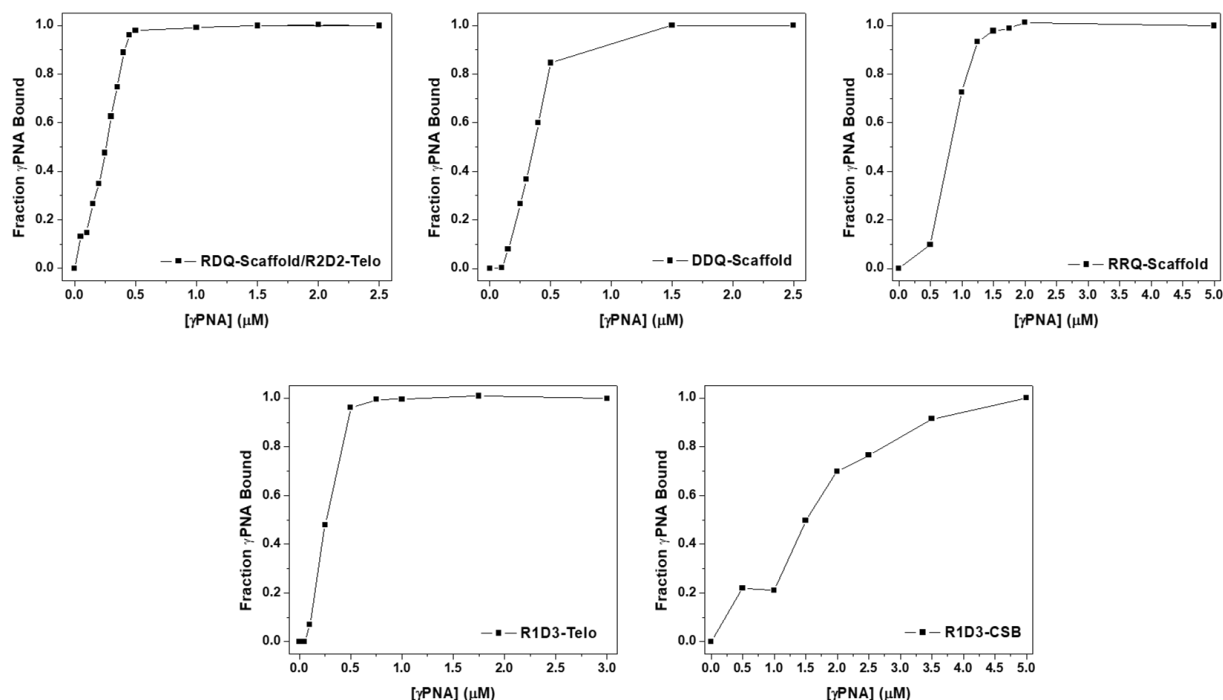
Supporting Figure 4.4 GQ-γPNA titration data converted to Fraction γPNA bound (where maximum fluorescence was normalized to 0% bound and minimum fluorescence was set to 100% bound. IC$_{50}$ values were calculated at the concentration where 50% γPNA is bound.

Supporting Table 4.1 DNA and RNA sequences of 9 and 10mer intermolecular homoGQs.

Name	Sequence (RNA is bolded)
DDQ 10mer	5'GGGTTAGGGT
RRQ 10mer	5'**GGGUUAGGGU**
DDQ 9mer	5'GGGTTAGGG
RRQ 9mer	5'**GGGUUAGGG**

An additional analysis made during this chapter that did not fit within the ultimate goal was the presence of T and U base overhangs on the intermolecular homoGQs. Overhangs have been well studied for intramolecular DNA and RNA GQs and studied to a lesser degree for intermolecular GQs. The overall consensus is that overhangs further stabilize the GQ by base stacking on the outermost tetrad; however, the amount of increase is not well established[35,36]. Here, we observe little difference in our overhang constructs. For DDQ, the 9mer melting temperature is about 2 °C lower than that with a T overhang (Supporting Figure 4.5 left). Additionally, the T overhang does not significantly change the secondary structure of the GQ (Supporting Figure 4.5 right). The RRQ construct saw no change in melting temperature with the U

overhang, however, the 9mer showed a much broader transition than the 10mer (Supporting Figure 4.6, left). No secondary structure differences were observed (Supporting Figure 4.6, right). These results indicate the effect of a single base addition to the GQ is not sufficient to significantly improve thermal stability or change the secondary structure. However, the addition of more overhangs could provide compounded stability.
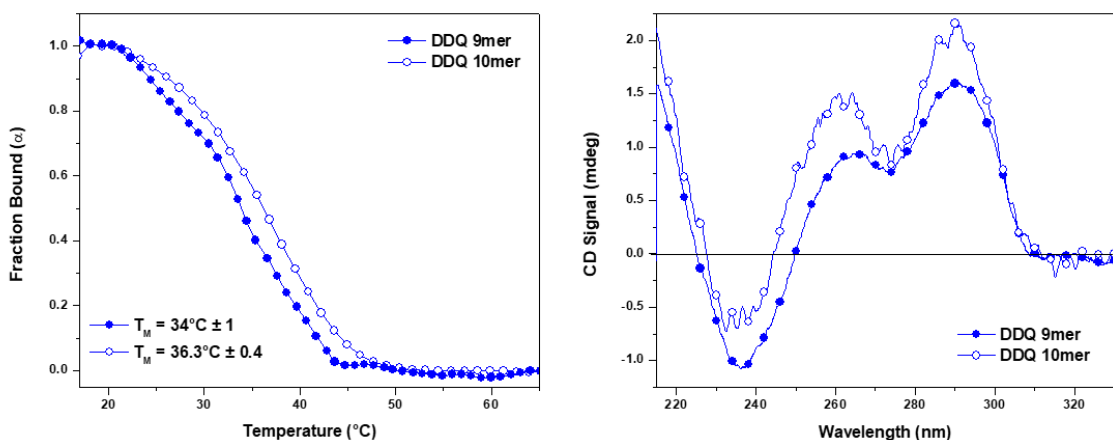
Supporting Figure 4.5 Thermal melting curves and circular dichroism (CD) of homoGQs DDQ 9mer (GGGTTAGGG) and DDQ 10mer (GGGTTAGGGT). [GQ]=1 µM n=3.
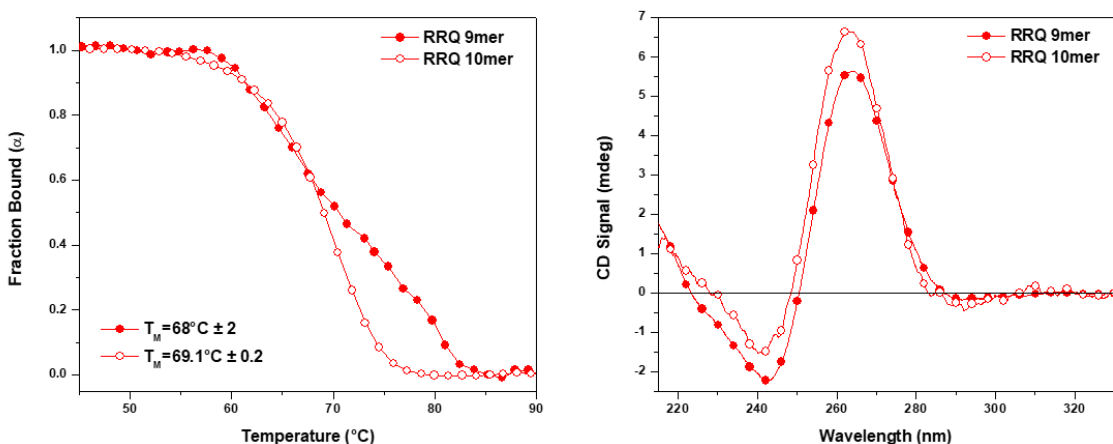
Supporting Figure 4.6 Thermal melting curves and circular dichroism (CD) of homoGQs DDQ 9mer (GGGTTAGGG) and DDQ 10mer (GGGTTAGGGT). [GQ]=1 µM n=3.

5 Chapter 5. γPNA FRET pair miniprobes for quantitative fluorescent in situ hybridization to telomeric DNA.

5.1 Introduction

5.1.1 Using Fluorescent-Labeled Probes to Target Nucleic Acid.

Fluorescent-labeled probes are a broad class of molecules used in biomedical research for identifying, tracking, or quantifying biomolecules. In order to target native biomolecules, probes (such as small molecules, peptides, or nucleic acids) can be covalently linked to a fluorophore. In this chapter, we target telomeric nucleic acid, a repeating sequence located at chromosomal ends. In order to target nucleic acid, a complementary oligomer is often used. A variety of complementary probe designs using fluorescence have been applied for targeting nucleic acids[1].

Traditional fluorophore-labeled oligomers, shown in Scheme 5.1 A, hybridize their target and are useful for target detection. These types of probes have been used in a variety of biological applications such as the imaging of RNA transcripts, detection of DNA mutations, and fluorescence in situ hybridization (FISH)[2–5]. Since these probes are continuously fluorescent, high background fluorescence is generally observed. For this reason, washing steps are required in order to eliminate any unbound probe[6]. Additional probe designs have been implemented in order to decrease background signal. One example is the molecular beacon, shown in Scheme 5.1 B. Molecular beacons utilize a single strand of nucleic acid containing two molecules at the 5'/3' end[7,8]. This pair, shown as a fluorophore and quencher in Scheme 5.1 B, is held in close proximity by a stem loop in the probe and is thus not fluorescent[7,8]. However, when the probe binds its target, the quencher is no longer close enough to the fluorophore to quench and fluorescence is observed. This system has shown promise, with decreased background signal compared to traditional probes. However, nonspecific opening of the hairpin has been observed[9,10].
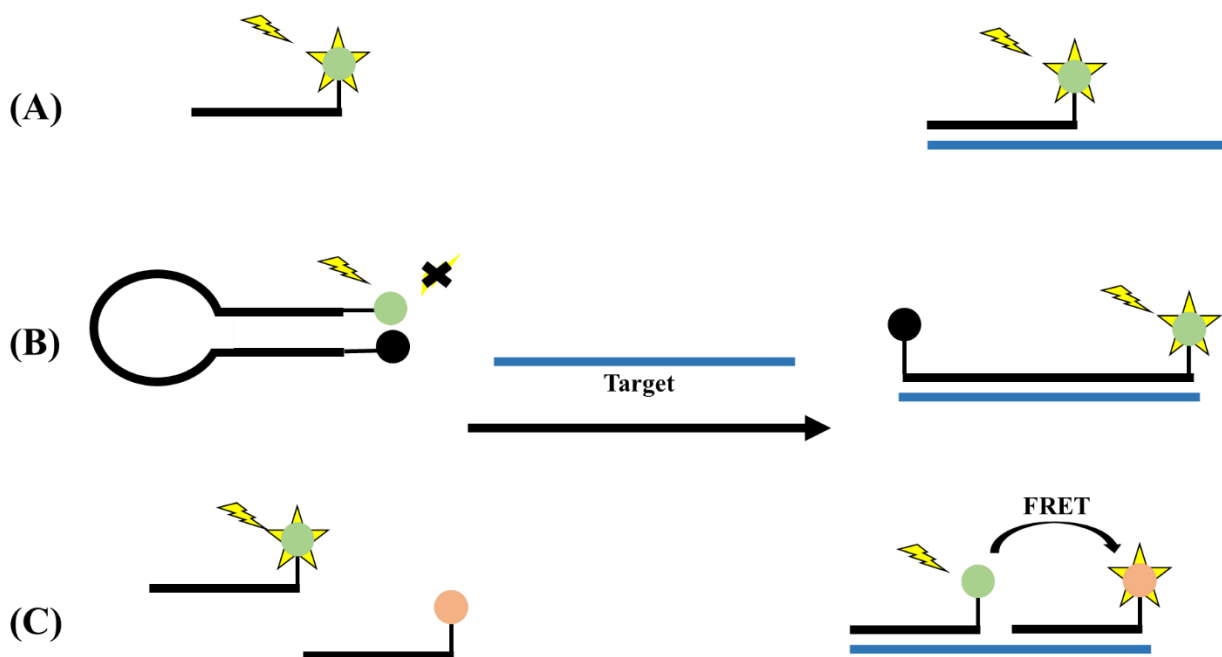
Scheme 5.1 Strategies for targeting nucleic acid using fluorescent-labeled nucleic acid probes. (A) Standard fluorescent-labeled probe (B) Molecular beacon using fluorophore and quencher (C) Two FRET pair probes.

An alternative probe design uses a highly sensitive fluorescence phenomenon: Förster Resonance Energy Transfer (FRET). FRET utilizes two fluorophore molecules, called a donor and acceptor. These dyes must have spectral overlap; meaning the emission spectrum of the donor must overlap with the excitation spectrum of the acceptor[11]. When the donor molecule is in the excited state, it is able to transfer its energy and excite the acceptor molecule[11]. This allows visualization of fluorescence of the acceptor molecule via excitation of the donor molecule[11]. In order for FRET to occur, the donor and acceptor probes must be in close proximity, leading to decreased background signal for unbound probes. The proximity requirements allow conformational changes to be observed, and has been used for a variety of projects including protein interactions and enzyme kinetics, in addition to nucleic acid targeting[12,13].

5.1.2   Telomeric DNA as a Target for FRET Probes.

Telomeres are chromosomal end caps which consist of repeating sequences (TTAGGG in vertebrates) with thousands of repeats that slowly degrade over the course of an individual cell's lifetime[14].

The discovery of how these telomeres provide genome stability and the enzyme that elongates telomeres was awarded the Nobel Prize in 2009[15]. Telomeres consist of both double stranded DNA as well as a 5' overhang, which decreases in length after each replication[16]. This is known as the end-replication problem. Cells can use telomerase, an enzyme capable of extended telomere length, to alleviate this problem[17,18].

Additionally, the lengths of telomeres have received significant attention because of their potential involvement in aging and disease[19]. Interestingly, a variety of lifestyle factors have been shown to play a role in the length of telomeres, including: smoking, exposure to toxins, exercise, and diet[20]. While the shortening of telomeres can lead to damaged genomic DNA, long telomeres are not always desirable. Upregulation of telomerase, leading to increased telomere length, has been shown to prevent the onset of degenerative diseases but it is also strongly associated with cancer[21].

Recent work has developed methods for determining telomere length. One method uses telomere restriction fragment analysis by southern blot; however, this method only provides an average telomere length for a population of cells and requires a significant amount of genomic material ($3\mu g$)[22]. Quantitative PCR is a second method available for quantifying telomere length, but it only provides an average telomere length[23]. A third method for determining telomere length is quantitative fluorescence in situ hybridization (Q-FISH). This method utilizes fluorophore-labeled nucleic acid probes that hybridize to telomere repeats[24]. Q-FISH has been further improved using peptide nucleic acid (PNA, Figure 5.1), where complementary PNA probes quantified the shortest telomeres in cells and tissue[25].
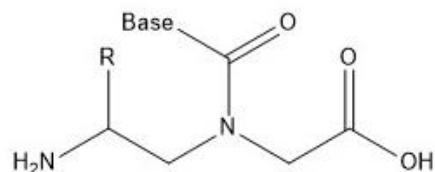
Figure 5.1 Structure of peptide nucleic acid. R indicates the gamma position, which can be modified to preorganize the secondary structure of PNA.

Previous PNA probes have required relatively long (>12) sequences for effective binding[5,26]. In the context of a FRET pair (Scheme 5.1 C), this requires the FRET pair to be attached on the 5' end of the first probe and the 3' end of the second probe in order to be close enough for efficient FRET. If the FRET pair is too close, fluorophore quenching can occur, so the probes must be separated by a few nucleotides[8,27]. Strategies to reduce the donor-acceptor distance while targeting a continuous nucleic acid sequence utilize

internally tethered dyes[28,29]. An alternative method utilizes relatively short probes with fluorophore labels on the 5ʹ ends. This not only allows for continuous hybridization, but also increases signal brightness because a larger number of probes can bind to a single nucleic acid sequence. Shorter probes require high binding affinity, which can be achieved using gamma modified PNA (γPNA). Incorporating modifications to the gamma position preorganizes the PNA to a right handed helix and significantly increases binding affinity[30,31]. Our group has used 12mer and 6mer γPNA miniprobes to fluorescently label telomeres[32]. These miniprobes showed brighter fluorescence in the staining of cells with short telomeres compared to the 18mer probes[32].

5.1.3   PNA Rationale and Design.

We designed two 9mer γPNA miniprobes that are complementary to 1.5 telomeric repeats. Use of these sequences allows the probes to hybridize in an alternating fashion, ensuring that each donor dye is adjacent to an acceptor dye (Figure 5.2).
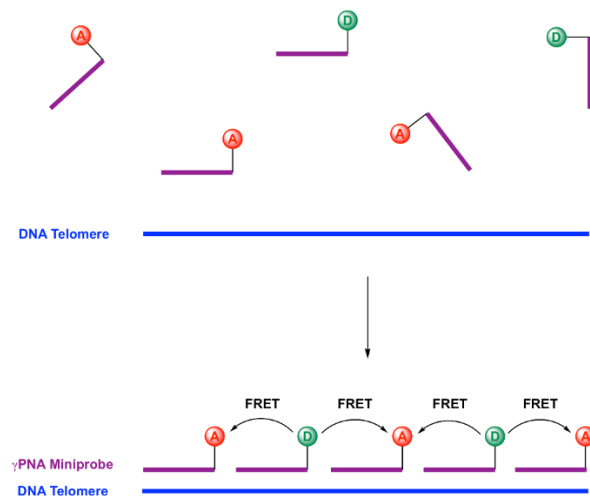
Figure 5.2 Alternating γPNA miniprobes on a DNA telomere for effective FRET signal. Adapted from Reference 33.

Previously, work from our lab utilized cyanine-3 (Cy3) in fluorescent miniprobes, so the Cy3/Cy5 FRET pair was initially chosen. These probes were analyzed by Dr. Lisa Rastede, who characterized the FRET pair on a variety of telomere lengths. In this chapter, we utilize a sulfonated Cy3 (SCy3) to improve solubility. We also utilize an Alexa647 (X647) fluorophore as the acceptor dye to improve brightness.

The biophysical characterization of these probes analyze their ability to bind telomere targets consisting of three or 4.5 telomeric repeats (Figure 5.3). This allows us to understand any cooperativity in the presence of one or two acceptor dyes.
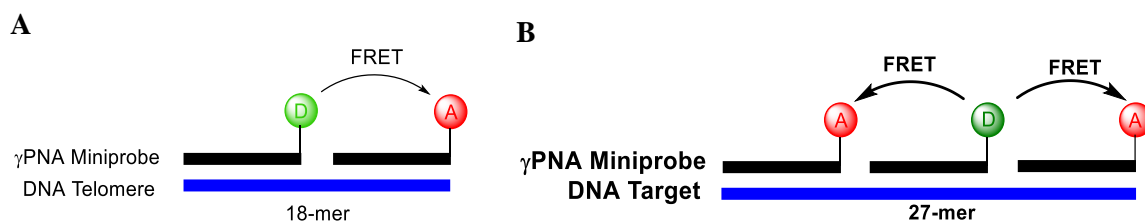
Figure 5.3 Donor/acceptor placement on DNA targets of 18 (A) and 27 (B) nucleotides in length Adapted from Reference 33.

We then collaborated with Dr. Patricia Opresko in order to perform FISH analysis on telomeres in cells. FISH methods generally require washing steps in order to reduce background noise, however the high affinity γPNA FRET pair could lead to the elimination of wash steps, since unhybridized probes will not exhibit FRET.

Finally, we tested longer 12mer γPNA miniprobes to identify if the incorporation of additional donor dyes would improve FRET efficiency. These miniprobes utilized SCy3 dyes that were covalently bound to the terminal ends of the γPNA or internally bound dyes through covalent binding to a uracil base. We conducted a biophysical characterization to determine any impact on the stability of the probe hybridization, and determined if the brightness of the probes increased.

5.2   Materials and Methods

Table 5.1 Probe, DNA, and RNA sequences utilized.

Name	Sequence
SCy3-γPNA-A	H_2N- Lys -TCC CAA TCC- SCy3
Cy5-γPNA-B	H_2N- Lys -CAA TCC CAA- Cy5
X647-γPNA-B	H_2N- Lys -CAA TCC CAA- X647
γPNA-B	H_2N-Lys-CAA TCC CAA- H
Telo-3 DNA	5'-(AGG GTT)$_3$-3'
Telo-4.5 DNA	5'-(AGG GTT)$_4$AGG-3'
compA DNA	5'-AGG GTT AGG-3'
compB DNA	5'-GTT AGG GTT-3'

Table 5.2 γPNA and DNA sequences used to study dual dye-labeled probes.

Name	Sequence
Term_PNA	H_2N-Lys-AATCCCAATCCC-**SCy3**
Mid_PNA	H_2N-Lys-AATCCCAAu**SCy3**CCC-Ac
Term&Mid_PNA	H_2N-Lys-AATCCCAAu**SCy3**CCC-**SCy3**
Telo-2 DNA	5′-TTAGGGTTAGGG-3′

Materials DNA oligonucleotides were purchased from Integrated DNA Technologies (idtdna.com) with standard desalting purification. γPNA miniprobes were provided by PNAInnovations.

Thermal Melting Analysis Thermal melting analysis was performed using a Varian Cary 3 UV-vis spectrophotometer with a temperature controlled multicell holder. Samples were buffered in 10 mM Tris, 0.1 mM EDTA, 100 mM KCl, pH 7.00 and strand concentrations were 1 µM for each DNA and γPNA. Samples were heated to 95 °C and annealed by slowly cooling to 15 °C (rate = 1 °C/min). The samples were then slowly heated to 95 °C (rate = 1 °C/min). Absorbance was measured at 260 nm. The data are reported as α, or the fraction of DNA/PNA in the duplexed state[34]. Alpha values were calculated by first, identifying the upper and lower baselines, corresponding to the fully single-stranded and fully duplexed states, respectively. Upper baselines in Figure 5.6 utilized absorbance values from 15 °C to 30 °C and lower baselines utilized 85 °C to 90 °C. Next, alpha values were calculated for each temperature using a ratio of the change in absorbance between the upper baseline and the experimental point, and the change in absorbance between the upper and lower baselines for each temperature on the melting curve. Melting temperatures were determined by calculating the temperature where there is 50% duplex and 50% single stranded nucleic acid (i.e. temperature where α = 0.5).

Circular Dichroism Spectroscopy CD data were collected on a Jasco 715 spectrophotometer at 25 °C using samples previously annealed in UV melting experiments. Each spectrum was obtained by averaging six scans (200-330 nm) collected at 100 nm/min.

Fluorescence Spectroscopy Fluorescence data were obtained using a Cary Eclipse Fluorescence spectrophotometer. Samples were buffered in 10 mM Tris, 0.1 mM EDTA, 100 mM KCl, pH 7.00 and strand concentrations for samples using Telo-3 DNA were 200 nm for Telo-3 DNA, SCy3-γPNA, and X647-γPNA. Samples using Telo-4.5 DNA contained 200 nm Telo-4.5 DNA, 200 nm SCy3-γPNA, and 400 nm X647-γPNA. (Telo-4.5 DNA has two binding sites for X647-γPNA.) Scans were collected at room temperature after samples were annealed (heated at 95 °C for five minutes and then slowly cooled to room temperature). SCy3 was excited at 520 nm and emission was recorded from 530-800 nm in order to monitor both SCy3 and FRET emission. Control scans of X647 were excited at 600 nm and monitored from 610-800 nm.

Cell Culture Human telomerase expressing U2OS osteosarcoma cells were purchased from ATCC. The HeLa VST was provided by Dr. Roderick O'Sullivan (University of Pittsburgh). Cells were cultured in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 50 units/mL penicillin, and 50 units/mL streptomycin (Gibco) at 37 °C in humidified chambers with 5% CO_2 and 20% O_2. Cells were cultured with 10% FBS from Hyclone and at 5% O_2, which improves proliferation of this cell line.

Fluorescent In Situ Hybridization of Interphase Cells Staining of interphase cell nuclei for telomere foci was based on previous protocols[35] with slight modification. Briefly, 50,000 – 150,000 cells were seeded in 35 mM glass bottom dishes (Mattek) and allowed to attach overnight. Cells were subsequently fixed in 3.8% paraformaldehyde, washed three times in 1xPBS, then successively dehydrated with 70%, 90%, then 100% ethanol. γPNA miniprobes (final concentration 1 μg/mL each) were added to a hybridization solution (70% formamide, 5% 20x $MgCl_2$ buffer, 0.5% blocking reagent, 10 mM Tris HCl pH = 7.5). 20x $MgCl_2$ buffer is 20 mM $MgCl_2$, 82 mM NaH_2PO_4, 9 mM citric acid, pH = 7.4. 10% blocking reagent (Roche 11096176001) is dissolved in maleic acid buffer 100 mM maleic acid, 150 mM NaCl, pH = 7.5. The solution was incubated for 5 min at 95 °C, iced for 5 min, then applied to samples and allowed to hybridize in a dark humid chamber at 70 °C for 10 minutes, followed by room temperature for 2-3 hours. After hybridization, the samples were rinsed with PBS once, then counterstained with DAPI, rinsed again in PBS, then allowed

to dry. Samples were mounted with ProLong Diamond Antifade (Invitrogen) and allowed to cure 24 hours before imaging.

Fluorescent In Situ Hybridization of Metaphase Chromosomes Metaphase chromosome spreads were prepared and stained as described previously with some modifications[36]. Briefly, 2x10^6 cells were seeded into flasks and allowed to grow over night. Cells were then incubated in media containing 0.05 µg/mL colcemid for three hours, then incubated in 75 mM KCl for 7 minutes, and fixed in 3:1 methanol: glacial acetic acid solution. Samples were then dropped onto slides, soaked in PBS for 5 minutes, fixed again in 3.8% paraformaldehyde for 10 minutes, and washed three times in PBS. Samples were treated with 250 µg/mL RNaseA for 15 minutes at 37 °C, 1 mg/mL pepsin for 15 minutes, 3.8% paraformaldehyde fixative for 2 minutes, and then three rinses with PBS. This was followed by successive dehydration in 70%, 90%, and 100% ethanol. SCy3 and X647 FRET miniprobes were mixed in ddH$_2$O in a 10:1 ratio. The mixture was added to the hybridization solution as described above for a final concentration of 1µg/mL SCy3-γPNA-A and 0.1 µg/mL X647-γPNA-B. The solution was heated for 5 minutes at 95 °C, iced for 5 minutes, and then applied to samples. Hybridization was at 70 °C for 10 minutes in a dark humid chamber, followed by overnight at 4 °C. Samples were rinsed three times in PBS with 0.5 µg/mL DAPI in the last wash. Samples were rinsed twice with ddH$_2$O, and then successively dehydrated with 70%, 90% and 100% ethanol. Slides were air dried, mounted with ProLong Diamond Antifade (Invitrogen), and allowed to cure 24 hours before imaging.

Imaging Cells were imaged using a Nikon Ti90 epi-fluorescence microscope (Nikon Inc., NY) equipped with a PlanApo 60×/1.40 oil immersion objective. Samples were imaged under three filter combinations: 1) Cy3 excitation and visualization of Cy3 emission, 2) Cy5 excitation and visualization of Cy5 emission and 3) Cy3 excitation and visualization of Cy5 emission (FRET channel), where the Cy3 and Cy5 channels are appropriate for SCy3 and Alexa647, the dyes used on our γPNA miniprobes. The Nikon NIS element advanced software was used to acquire the images. Exposure times and look up table settings were identical for all imaging conditions. For the interphase nuclei and metaphase chromosomes a series of z- stacked images (0.20 micron steps) were acquired for each and processed by deconvolution using Nikon NIS

elements software to remove out-of-focus light. A maximum intensity projection of each image was obtained using the Nikon NIS elements ND processing software, and was analyzed for fluorescence signal intensity.

γPNA Fluorescence Intensity Measurements The fluorescence intensity of γPNA stained telomere foci in interphase nuclei and metaphase chromosomes were measured by detecting objects and quantitating fluorescence intensity units in Nikon NIS Elements. A consistent minimum object threshold was established across each experiment. The total sum of intensity per object was measured.

Statistical Analyses Statistical significance was calculated using Prism 6 (GraphPad Software Inc.) and two-tailed unpaired Student's t-test at 95% confidence level.

5.3   Results

5.3.1   *Optimizing the Cy3/Cy5 FRET pair to SCy3/X647.*

Previous work conducted by Dr. Lisa Rastede utilized 9mer γPNA FRET miniprobes where the donor dye was cyanine-3 (Cy3) and the acceptor dye was cyanine-5 (Cy5). These probes had a FRET efficency of 34% when bound 1:1:1 with Telo-3. Biological samples tested by Dr. Patricia Opresko's lab provided promising results using the miniprobes to image telomeres. However, concerns regarding solubity of the miniprobes, specifically Cy3, lead us to utilize a sulfonated Cy3 (SCy3) on PNA-A. Adding the charged sulfonate groups to the cyanine dye increases water solubility and decreases dye aggregation.
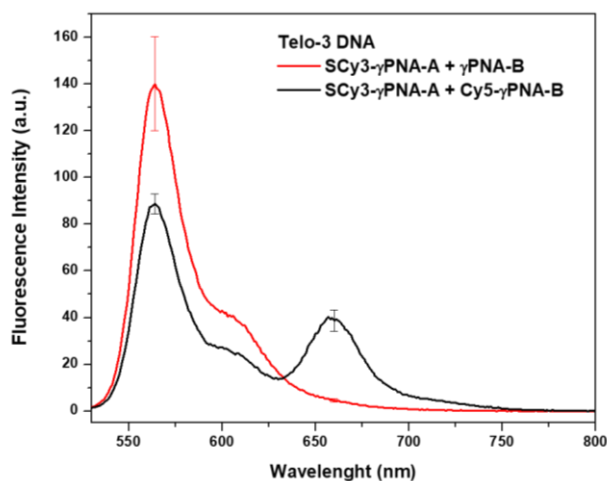
Figure 5.4 Fluorescence spectra of FRET miniprobes with Telo-3 DNA. Red curve is Telo-3 with SCy3-γPNA-A and γPNA-B. Black curve is Telo-3, SCy3-γPNA-A and Cy5-γPNA-B, and SCy3-γPNA-A. Samples were excited at 530 nm.

The 9mer PNA probes were preannealed to DNA, where Telo-3 complementary binds a single PNA-A and PNA-B strand. A non-labeled PNA-B was annealed to account for any differences in quantum yield when the probes are fully hybrid (black in Figure 5.4). When both PNA probes are labeled with donor and acceptor dyes and excited at 520 nm (within the excitation range of SCy3) we observe a decrease in the signal from SCy3 and an increase in signal from Cy5 (red in Figure 5.4), indicating FRET was successful. FRET efficiency was calculated to be 41%, which is slightly higher than observed with the non-sulfonated Cy3 probes.
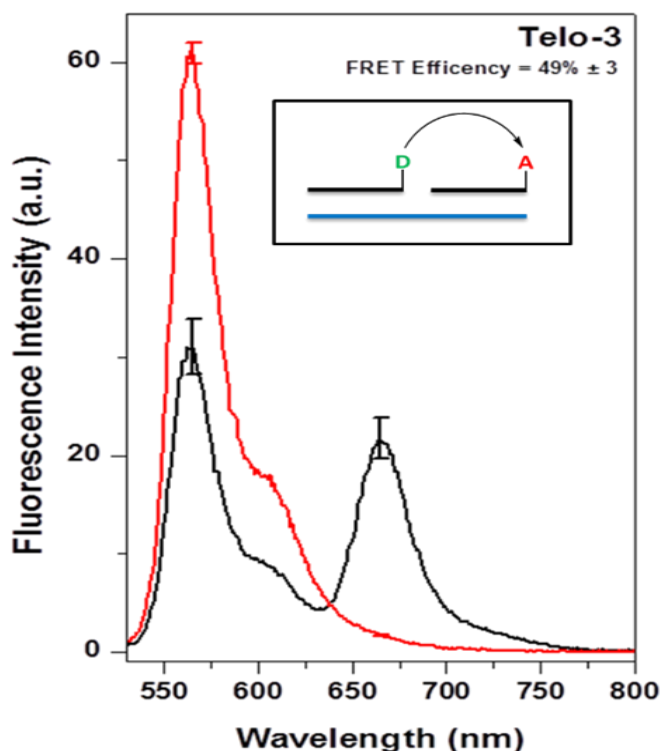
Figure 5.5 Fluorescence spectra of FRET miniprobes with Telo-3 DNA. Red curve is Telo-3 with SCy3-γPNA-A and γPNA-B. Black curve is Telo-3, SCy3-γPNA-A and X647-γPNA-B, and SCy3-γPNA-A. Telo-3 DNA and γPNA concentration was 200 nm. Samples were excited at 530 nm .Inset is a schematic representation of each DNA (blue) with donor (green) and acceptor (red) γPNAs.

We continued by modifying the acceptor to an Alexa647 (X647). Again, SCy3-γPNA-A and X647-γPNA-B were preannealed to Telo-3, excited at 520 nm, and emission spectra were compared to the system that did not contain an acceptor dye. Using an X647 fluorophore, we saw an increase in FRET efficiency (49%) (Figure 5.5).

5.3.2 *Biophysical characterization of SCy3/X647 indicates high stability.*

Each miniprobe forms a stable duplex with its 9-mer DNA complement ($T_m = 56$-57 ºC; Figure 5.6, pink and blue curves). The melting temperature improves by approximately 5ºC when both miniprobes are

combined with DNA Telo-3, an 18-mer on which the miniprobes can hybridize at adjacent positions. We postulate the enhanced stability results from coaxial stacking between the adjacent γPNA miniprobes, as we observed previously for a 12-mer γPNA[32]. The duplex secondary structure was additionally confirmed by circular dichroism (Supporting Figure 5.1). This showed a strong duplex signal with a maximum at 260 nm and a minimum at 240 nm. The DNA alone showed weak peaks at 260 nm and 290 nm, indicative of single stranded nucleic acid and potentially some bimolecular GQ formation.
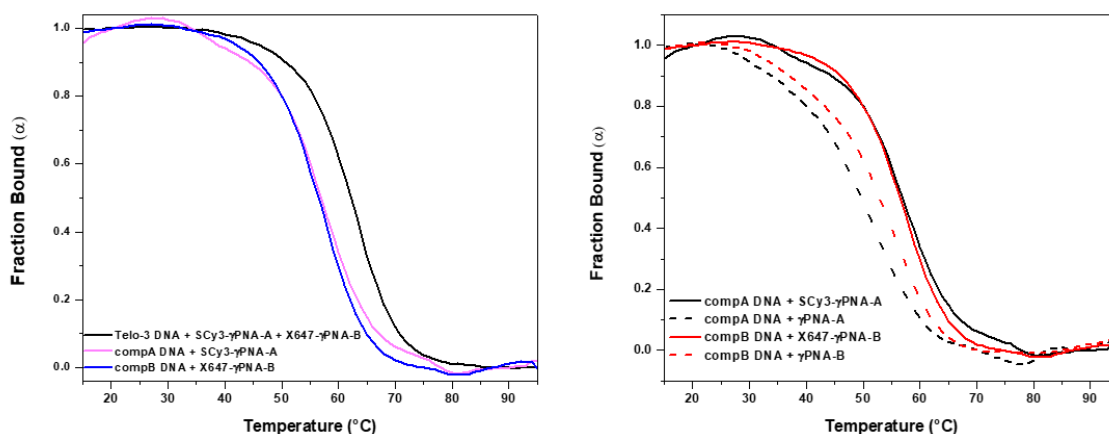
Figure 5.6 Alpha plots of thermal melting curves where α is the fraction in the duplex state. Left shows Telo-3 DNA with both labeled probes (SCy3-γPNA-A and X647-γPNA-B) (black) compared to both probes bound to their complementary 9mer (compA DNA + SCy3-γPNA-A, pink) and (compB DNA + X647-γPNA-B, blue). Right shows the labeled (solid) and unlabeled (dashed) γPNAs with their 9mer complements (γPNA-A in red and γPNA-B in black). DNA and γPNA concentration was 1 µM.

We further analyzed the effects of the fluorophores on each miniprobe with their complement DNA. Melting temperatures showed a 7.5 °C increase, from 49.5 °C to 57.0 °C, in the presence of SCy3 on γPNA-A, black dashed and solid lines in Figure 5.6 right. A 4 °C increase was observed for X647; where γPNA-B with its complementary DNA had a melting temperature of 52.4 °C and X647-γPNA-B with its complementary DNA melted at 56.5 °C (red dashed and solid lines in Figure 5.6 right). This slight increase in melting temperature is likely due to stabilizing stacking interactions from the fluorophores[37].

5.3.3 *Increasing acceptor/donor ratio increases FRET efficiency in DNA.*

We wanted to further understand how the γPNA miniprobes would interact with longer DNA sequences. We increased the number of telomeric repeats from 3 (18nt) to 4.5 (27nt). This positions a donor in between two acceptors, which was expected to increase the FRET efficiency (schematic shown as insert in Figure 5.7). FRET is evident based on the decrease in the donor SCy3 peak at 560 nm, and the new X647 emission peak at 660 nm (Figure 5.7, red and black curves). FRET efficiency was calculated as 68% using Telo-4.5. This is a significant increase from the 49% observed using Telo-3 in section 5.3.1.

Figure 5.7 Fluorescence spectra of FRET miniprobes with Telo-4.5. Red curve is Telo-3 with SCy3-γPNA-A and γPNA-B. Black curve is Telo-3, SCy3-γPNA-A and X647-γPNA-B, and SCy3-γPNA-A. Telo-3 DNA and γPNA concentration was 200 nm. Samples were excited at 530 nm. Inset is a schematic representation of each DNA (blue) with donor (green) and acceptor (red) γPNAs.

5.3.4 *FRET γPNA miniprobes effectively detect telomeric repeats in cells.*

In order to demonstrate the utility of the γPNA miniprobes in biological samples, Dr. Patricia Opresko (University of Pittsburgh) imaged fixed human U2OS osteosarcoma cells. Nuclei from interphase human U2OS osteosarcoma cells were stained with each γPNA probe separately in the absence of wash steps. When stained individually, SCy3-γPNA-A and X647-γPNA-B show emission in their respective channels but no fluorescence in the FRET channel (Figure 5.8 a, first and second row). This indicates

binding of each probe to its respective target and minimal cross talk between the channels. Next, a 1:1

mixture of SCy3-γPNA-A and X647-γPNA-B was used to stain the U2OS cells (Figure 5.8 a, third row).

Distinct FRET signals are observed in the SC3-excitation/X647-emission (FRET channel). The use of

traditional washing steps, which are typically required to remove unhybridized probe(s), significantly

decreases the overall brightness (Figure 5.8 a, fourth row). Quantifying the number of foci (Figure 5.8 b

and c) shows the wash steps reduce both the telomere foci intensity and the number of telomeric foci

detected per nucleus. This suggests the FRET γPNA miniprobes would be able to improve detection of

critically short telomeres that might otherwise be missed using wash conditions.
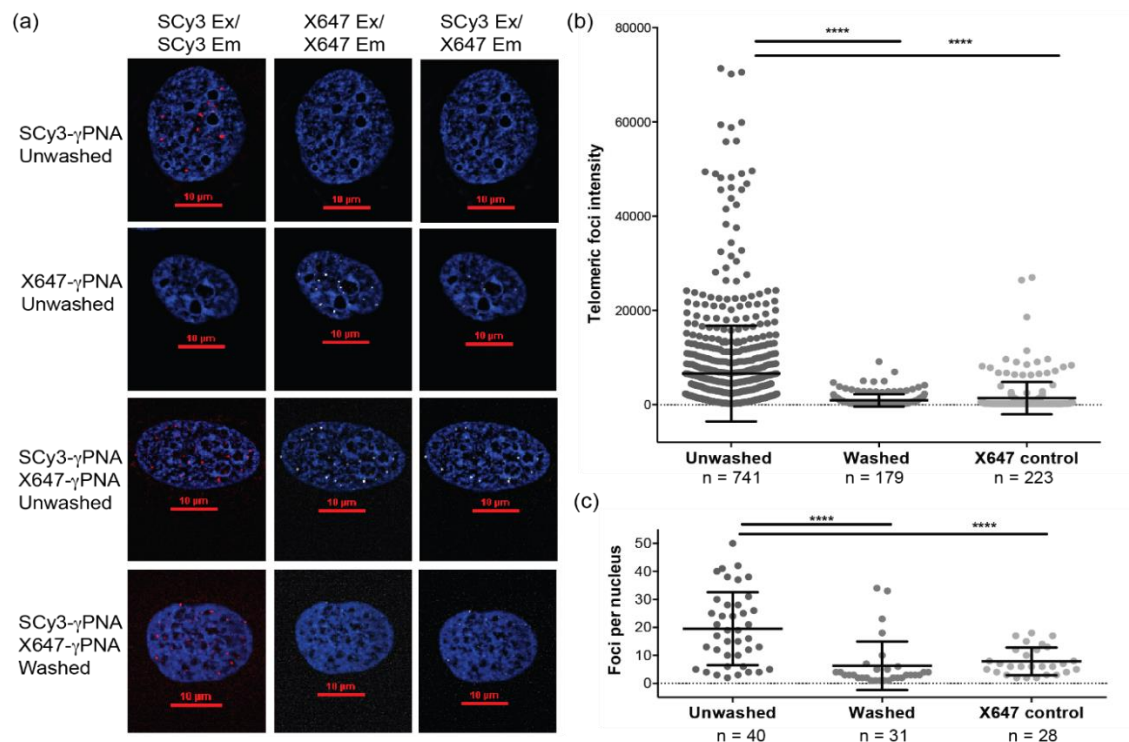
Figure 5.8 (**a**) Stained interphase U2OS cell nuclei using SCy3-γPNA-A, X647-γPNA-B, or both by FISH with or without wash steps. SCy3 excitation and emission (red), X647 excitation and emission (white) or SCy3 excitation and X647 emission (white, FRET channel). Nuclei were stained with DAPI. Fluorescence signal intensity for each telomeric foci (**b**) and the number of telomeric foci per nucleus (**c**) in the FRET channel were measured. Data from unwashed SCy3-γPNA + X647-γPNA (n = 40 nuclei), washed SCy3-γPNA + X647-γPNA (n = 31 nuclei) and unwashed X647-γPNA alone (n = 28 nuclei). Mean and SEM from total foci (**b**) or nuclei (**c**) from three independent experiments. **** $p < 0.0001$, unpaired t-test. Images and analysis by Dr. Patricia Opresko, Dr. Elise Fourquerel, and Alexander Orenstein.

5.3.5 *FRET miniprobes can be used on cells with critically short telomeres.*

Cells possessing critically short telomeres are difficult to stain[35]. We wanted to analyze the γPNA miniprobe's ability to test short telomeres. For these experiments, HeLa VST cells were chosen for staining because they possess very short telomeres (approximately 3.7kb) compared to the previously used U2OS cells which have highly variable telomere lengths[38,39]. Using a 1:1 mixture of SCy3-γPNA-A and X647-γPNA-B, HeLa VST cells were stained without washing (Figure 5.9). Strong FRET signal (SCy3 Ex/X647 Em) was observed, indicating the γPNA miniprobes efficiently stain critically short telomeres.
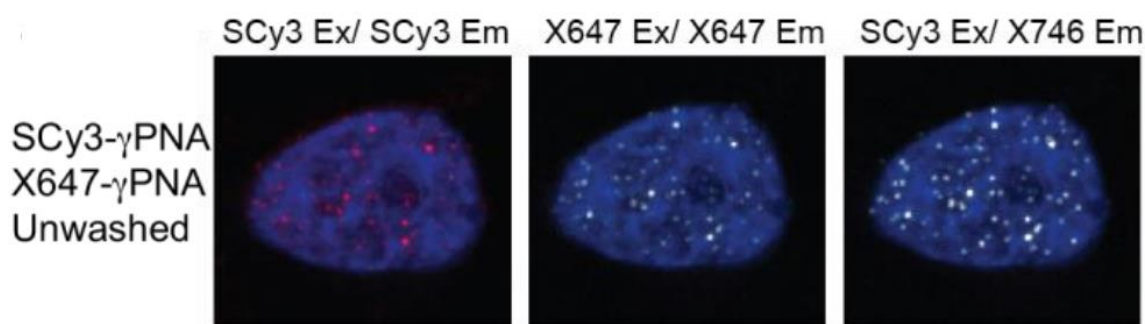
Figure 5.9 Interphase HeLa VST cell nuclei were stained with a mix of SCy3-γPNA and X647-γPNA probes (1:1 ratio) by FISH in the absence of washings. Images were captured using filters for SCy3 excitation and emission (red) , X647 excitation and emission (white) or SCy3 excitation and X647 emission (white, FRET channel). Images and analysis by Dr. Patricia Opresko, Dr. Elise Fourquerel, and Alexander Orenstein.

5.3.6 *Additional donor dyes neither affect PNA binding nor improves brightness.*

After experiencing success staining telomeres in cells, we wanted to continue improving the overall brightness of our telomere miniprobes by incorporating additional fluorophores. For these experiments, we utilized a 12mer γPNA with a SCy3 fluorophore (A) on the N-terminus, (B) in the middle of the probe (via a covalent linkage to a uracil base), or (C) both (Scheme 5.2 and Supporting Figure 5.2).
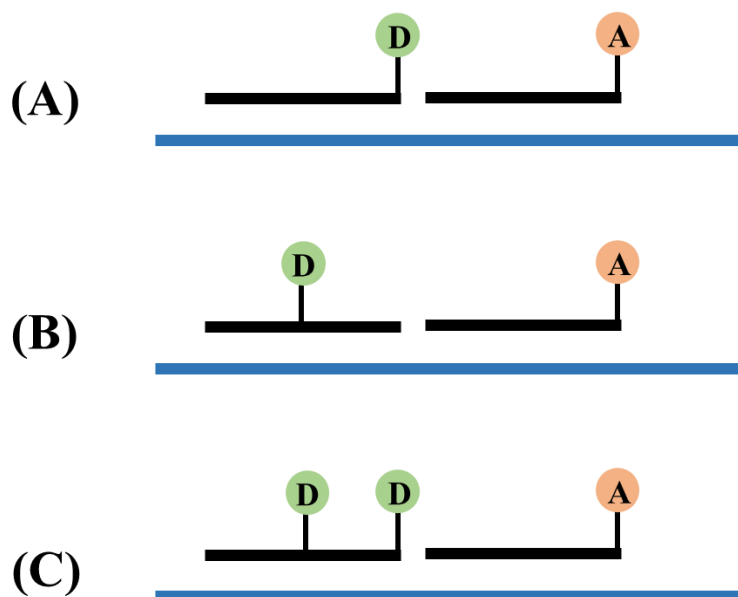
Scheme 5.2 Schematic design of dual dye-labeled γPNA probes.

We first confirmed effective duplex binding of the γPNA via thermal melting curves. Additionally, we were interested in identifying any thermodynamic penalties that the middle dye would impose on the binding event. As expected, each γPNA formed a duplex in the presence of 12mer complementary DNA as shown in Figure 5.10. All γPNA-DNA duplexes formed stable duplexes with melting temperatures between 72-75 °C (Term_PNA + DNA= 73.6 °C, Mid_PNA + DNA= 72.3 °C, Term&Mid_PNA = 72.9 °C, and Unlabeled_PNA = 75.3 °C). The melting temperature of the unlabeled probe is slightly higher than the three fluorescent-labeled probes (ΔT_M between 1.7 and 3 °C). This indicates including a fluorescent label decreases the thermal stability. However, the loss in thermal
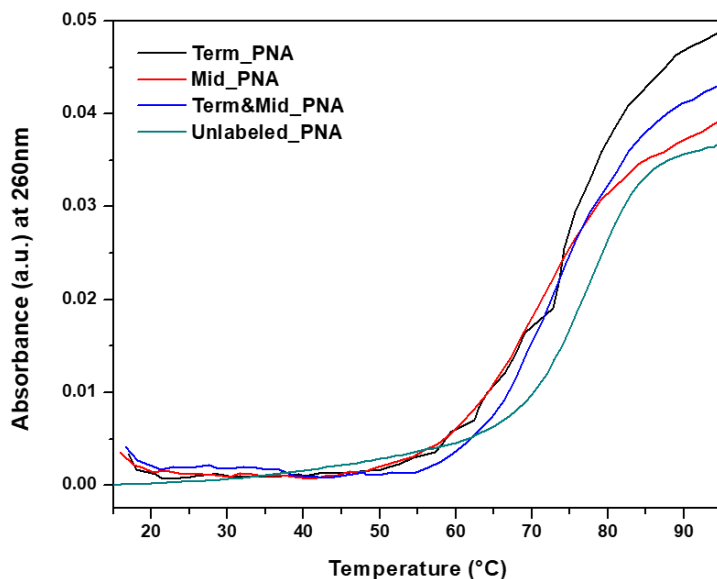
Figure 5.10 Thermal melting curves using single and dual dye-labeled probes. Term_PNA (black) Mid_PNA (red), Term&Mid_PNA (blue), and unlabeled_PNA (green) with complementary DNA. γPNA and DNA concentration was 1 µM.

stability appears to be a one-time penalty since there is no significant difference in the melting temperatures between the mono- and dual-labeled γPNAs.

Fluorescence experiments were performed in order to determine if a dual dye-labeled probe would be preferential to a mono dye-labeled probe. Initially, fluorescence experiments were conducted at 200 nm, shown in Figure 5.11.

Surprisingly, the mono (terminal)-labeled γPNA had the highest fluorescence. Additionally, the mono (middle)-labeled γPNA exhibited the lowest fluorescence, indicating that the
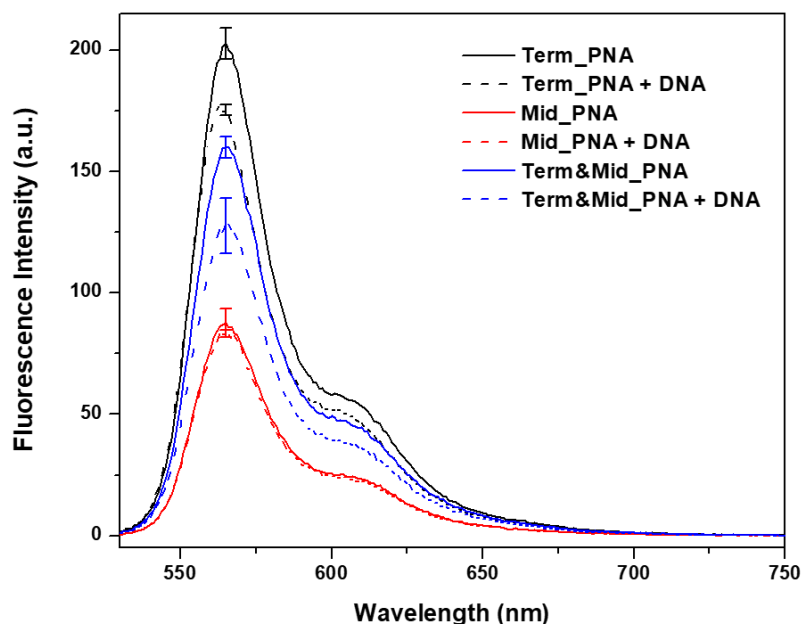
Figure 5.11 Fluorescence scans (right panel) with each γPNA (with and without complementary DNA) using lower concentrations (200 nm γPNA and 200 nm DNA). Excitation wavelength 520 nm. (n=3)

location of the dye is not ideal. The dual-labeled dye had an average fluorescence between the terminal and middle-labeled γPNA. This is strange since one could expect an additive effect, causing the overall fluorescence to be higher. However, with two dyes in close proximity, it is possible for the dyes to self-quench. The signal trends remained the same at 1 μM (with the exception of Term_PNA + DNA which had a significant drop in fluorescence) shown in Figure 5.12, right panel. UV-Vis scans (Figure 5.12, left panel) indicate the formation of a dimer via an increase in the 520 nm peak. This peak is most pronounced in the dual dye-labeled probe, which could also cause the overall fluorescence to decrease. Additionally, Mid_PNA shows significantly lower absorbance compared to the other two PNAs, Figure 5.12. This decrease in absorbance leads to a decrease in fluorescence of the probe, observed in Figure 5.11 and Figure 5.12, and would decrease the brightness of the probe.
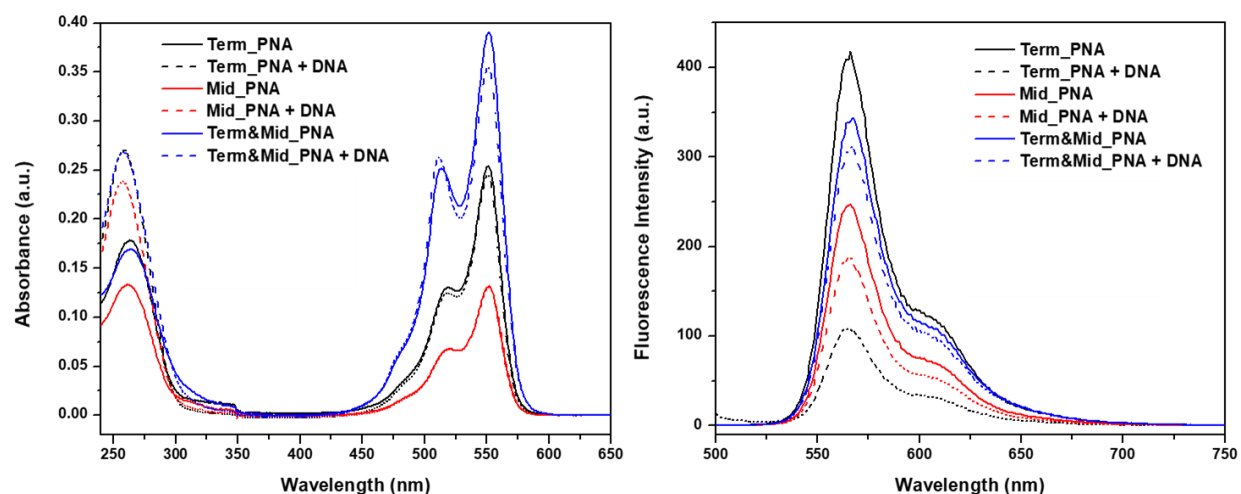
Figure 5.12 UV-Vis scans (left panel) and fluorescence scans (right panel) with each γPNA (with and without complementary DNA) (n=1). Excitation wavelength 490 nm. Concentration of γPNA and DNA was 1 μM.

5.4 Discussion

The ability to determine telomere length has become an important goal for scientists. Critically short telomeres can be caused by a variety of factors including disease and lifestyle[20]. Additionally, critically short telomeres can be difficult to identify since current methods (including southern blot and quantitative PCR) provide an average telomere length[22,23]. Identifying telomere length in individual chromosomes or cells with critically short telomeres would be more useful to researchers. To this end, we utilize γPNA miniprobes to analyze telomeres.

We first modified the FRET pair from the probes previously used by Dr. Lisa Rastede to a SCy3/Cy5 fluorophore pair. The sulfonated Cy3 improves water solubility, which would benefit future cellular experiments. With these probes, we observed a 41% FRET efficiency with a three repeat telomere sequence, containing one donor probe and one acceptor probe. We next modified the Cy5 dye to an Alexa fluorophore (X647). Alexa fluorophores have shown increased brightness compared to dyes with similar excitation and emission profiles. Probes using SCy3/X647 dyes showed a 49% FRET efficiency; slightly higher than that observed with the Cy5-labeled probe. Biophysical characterization indicates the probes have high thermal stability with their perfect complements. The SCy3 and X647 fluorophores were shown

131

to increase thermal stability by 7.5 °C and 4 °C, respectively. This stability is likely a result of coaxial stacking between γPNA probes[32].

Increased FRET efficiency was observed using a telomeric DNA containing 4.5 repeats that is able to bind one donor-containing probe and two acceptor-containing probes. This efficiency indicates improved energy transfer across a longer telomere sequence, and offers promising results for cellular studies. Cellular studies using human osteosarcoma cells where metaphase chromosomes in fixed nuclei showed effective FRET. We observed reduced background signal, which was sufficient to remove washing steps typically required in FISH experiments. The removal of wash steps improved signal intensity and allowed for a greater number of telomeric foci to be observed per nucleus. To explore the robustness of our FRET γPNA miniprobes, we monitored FRET using cells with critically short telomeres, HeLa VST cells. These results show strong promise for the use of γPNA miniprobes in quantitative telomere measurements.

We attempted to improve the efficiency of our FRET miniprobes by increasing the number of donor dyes per probe. SCy3 was used to label a 12mer γPNA probe on the N-terminus and internally, covalently linked to a uracil base. The probes showed a slight decrease in melting temperature in the presence of fluorophore, compared to an unlabeled probe, in the presence of a perfect complement. However, this appears to be a single penalty and the presence of an internal dye label does not further decrease stability. Fluorescence and absorbance analysis revealed that a dual-labeled donor probe may not improve our γPNA probes. The dual-labeled probe showed lower fluorescence signal than the terminally-labeled probe. Absorbance scans indicated this is likely due to the formation of a dimer, leading to self-quenching and to lower overall absorbance of the internal dye.

5.5 Future Directions

Our short γPNA probes offer a method for detecting and quantifying telomeres using FISH methods, without requiring washing steps. These probes could be extremely beneficial in a clinical setting as a potential diagnostic tool. Future work could focus on improving the FRET efficiency between

fluorophores in order to improve brightness and detect lower concentrations of telomere repeats. This could

aid in identifying diseased cells where critically short telomeres are implicated[19,40].

Although the incorporation of internal fluorophores was not successful, their placement was

determined by the location of the uracil base. A report from the Appella lab has utilized the gamma position

of PNA probes in order to internally label any position on the probe[41]. This could be useful in order to

increase separation between the two donor dyes, and prevent dimerization.

5.6   References

(1)   Su, X.; Xiao, X.; Zhang, C.; Zhao, M. Nucleic Acid Fluorescent Probes for Biological Sensing. *Applied Spectroscopy*. 2012, pp 1249–1261.

(2)   Raj, A.; van den Bogaard, P.; Rifkin, S. A.; van Oudenaarden, A.; Tyagi, S. Imaging Individual MRNA Molecules Using Multiple Singly Labeled Probes. *Nat. Methods* **2008**, *5* (10), 877–879.

(3)   Femino, A. M.; Kay, F.; Fogarty, K.; Singer, R. Visualization of Single RNA Transcripts in Situ. *Science (80-.).* **1998**, *280* (5363), 585–590.

(4)   Ranasinghe, R. T.; Brown, T. Fluorescence Based Strategies for Genetic Analysis. *Chemical Communications*. November 28, 2005, pp 5487–5502.

(5)   Lansdorp, P. M.; Verwoerd, N. P.; Van De Rijke, F. M.; Dragowska, V.; Little, M. T.; Dirks, R. W.; Raap, A. K.; Tanke, H. J. Heterogeneity in Telomere Length of Human Chromosomes. *Hum. Mol. Genet.* **1996**, *5* (5), 685–691.

(6)   Baerlocher, G. M.; Lansdorp, P. M. Telomere Length Measurements Using Fluorescence In Situ Hybridization and Flow Cytometry. In *Methods in Cell Biology*; Academic Press, 2006; Vol. 75, pp 719–750.

(7)   Tyagi, S.; Kramer, F. R. Molecular Beacons: Probes That Fluoresce Upon Hybridization. *Nat. Biotechnol.* **1996**, *14* (3), 303–308.

(8)   Guo, J.; Ju, J.; Turro, N. J. Fluorescent Hybridization Probes for Nucleic Acid Detection. *Analytical and Bioanalytical Chemistry*. April 17, 2012, pp 3115–3125.

(9)   Molenaar, C. Linear 2' O-Methyl RNA Probes for the Visualization of RNA in Living Cells. *Nucleic Acids Res.* **2001**, *29* (17), 89e – 89.

(10)   Santangelo, P. J.; Nix, B.; Tsourkas, A.; Bao, G. Dual FRET Molecular Beacons for MRNA Detection in Living Cells. *Nucleic Acids Res.* **2004**, *32* (6), e57–e57.

(11)   Didenko, V. V. Dna Probes Using Fluorescence Resonance Energy Transfer (Fret): Designs and Applications. *Biotechniques* **2001**, *31* (5), 1106–1121.

(12)   Kalinin, S.; Peulen, T.; Sindbert, S.; Rothwell, P. J.; Berger, S.; Restle, T.; Goody, R. S.; Gohlke, H.; Seidel, C. A. M. A Toolkit and Benchmark Study for FRET-Restrained High-Precision Structural Modeling. *Nat. Methods* **2012**, *9* (12), 1218–1225.

(13)   Sapsford, K. E.; Berti, L.; Medintz, I. L. Materials for Fluorescence Resonance Energy Transfer Analysis: Beyond Traditional Donor-Acceptor Combinations. *Angewandte Chemie - International Edition*. John Wiley & Sons, Ltd July 10, 2006, pp 4562–4588.

(14)   Neidle, S.; Parkinson, G. N. The Structure of Telomeric DNA. *Curr. Opin. Struct. Biol.* **2003**, *13* (3), 275–283.

(15)   Andersen, P. [The Nobel Prize in Physiology or Medicine 1967]. *Tidsskr. Nor. Laegeforen.* **1967**, *87* (22), 1926–1927.

(16)   Harley, C. B.; Futcher, A. B.; Greider, C. W. Telomeres Shorten during Ageing of Human Fibroblasts. *Nature* **1990**, *345* (6274), 458–460.

(17)   Shay, J. W.; Wright, W. E. Telomeres and Telomerase: Three Decades of Progress. *Nat. Rev. Genet.* **2019**, *20* (5), 299–309.

(18)   Bodnar, A. G.; Ouellette, M.; Frolkis, M.; Holt, S. E.; Chiu, C. P.; Morin, G. B.; Harley, C. B.; Shay, J. W.; Lichtsteiner, S.; Wright, W. E. Extension of Life-Span by Introduction of Telomerase into Normal Human Cells. *Science (80-.).* **1998**, *279* (5349), 349–352.

(19)   Calado, R. T.; Young, N. S. Telomere Diseases. *N. Engl. J. Med.* **2009**, *361* (24), 2353–2365.

(20)   Shammas, M. A. Telomeres, Lifestyle, Cancer, and Aging. *Curr. Opin. Clin. Nutr. Metab. Care* **2011**, *14* (1), 28–34.

(21)   Jafri, M. A.; Ansari, S. A.; Alqahtani, M. H.; Shay, J. W. Roles of Telomeres and Telomerase in Cancer, and Advances in Telomerase-Targeted Therapies. *Genome Medicine*. BioMed Central 2016, p 69.

(22)   Kimura, M.; Stone, R. C.; Hunt, S. C.; Skurnick, J.; Lu, X.; Cao, X.; Harley, C. B.; Aviv, A. Measurement of Telomere Length by the Southern Blot Analysis of Terminal Restriction Fragment Lengths. *Nat. Protoc.* **2010**, *5* (9), 1596–1607.

(23)   Cawthon, R. M. Telomere Measurement by Quantitative PCR. *Nucleic Acids Res.* **2002**, *30* (10), 47e – 47.

(24)   Ourliac-Garnier, I.; Londoño-Vallejo, A. Telomere Length Analysis by Quantitative Fluorescent in Situ Hybridization (Q-FISH). In *Methods in Molecular Biology*; Humana Press, 2017; Vol. 1587, pp 29–39.

(25)   Vera, E.; Blasco, M. A. Beyond Average: Potential for Measurement of Short Telomeres. *Aging*. June 8, 2012, pp 379–392.

(26)   Karadag, A.; Riminucci, M.; Bianco, P.; Cherman, N.; Kuznetsov, S. A.; Nguyen, N.; Collins, M. T.; Robey, P. G.; Fisher, L. W. A Novel Technique Based on a PNA Hybridization Probe and FRET Principle for Quantification of Mutant Genotype in Fibrous Dysplasia/McCune-Albright Syndrome. *Nucleic Acids Res.* **2004**, *32* (7), e63.

(27)   Robertson, K. L.; Yu, L.; Armitage, B. A.; Lopez, A. J.; Peteanu, L. A. Fluorescent PNA Probes as Hybridization Labels for Biological RNA. *Biochemistry* **2006**, *45* (19), 6066–6074.

(28)   Sei-Iida, Y. Real-Time Monitoring of in Vitro Transcriptional RNA Synthesis Using Fluorescence Resonance Energy Transfer. *Nucleic Acids Res.* **2000**, *28* (12), 59e – 59.

(29)   Tsuji, A.; Sato, Y.; Hirano, M.; Suga, T.; Koshimoto, H.; Taguchi, T.; Ohsuka, S. Development of a Time-Resolved Fluorometric Method for Observing Hybridization in Living Cells Using Fluorescence Resonance Energy Transfer. *Biophys. J.* **2001**, *81* (1), 501–515.

(30)   Dragulescu-Andrasi, A.; Rapireddy, S.; Frezza, B. M.; Gayathri, C.; Gil, R. R.; Ly, D. H. A Simple γ-Backbone Modification Preorganizes Peptide Nucleic Acid into a Helical Structure. *J. Am. Chem. Soc.* **2006**, *128* (31), 10258–10267.

(31)   Sahu, B.; Sacui, I.; Rapireddy, S.; Zanotti, K. J.; Bahal, R.; Armitage, B. A.; Ly, D. H. Synthesis and Characterization of Conformationally Preorganized, (R)-Diethylene Glycol-Containing γ-Peptide Nucleic Acids with Superior Hybridization Properties and Water Solubility. *J. Org. Chem.* **2011**, *76* (14), 5614–5627.

(32)   Pham, H. H.; Murphy, C. T.; Sureshkumar, G.; Ly, D. H.; Opresko, P. L.; Armitage, B. A. Cooperative Hybridization of ΓPNA Miniprobes to a Repeating Sequence Motif and Application to Telomere Analysis. *Org. Biomol. Chem.* **2014**, *12* (37), 7345–7354.

(33)   Orenstein, A.; Berlyoung, A. S.; Rastede, E. E.; Pham, H. H.; Fouquerel, E.; Murphy, C. T.; Leibowitz, B. J.; Yu, J.; Srivastava, T.; Armitage, B. A.; et al. ΓPNA FRET Pair Miniprobes for Quantitative Fluorescent in Situ Hybridization to Telomeric DNA in Cells and Tissue. *Molecules* **2017**, *22* (12).

(34)   Marky, L. A.; Breslauer, K. J. Calculating Thermodynamic Data for Transitions of Any Molecularity from Equilibrium Melting Curves. *Biopolymers* **1987**, *26* (9), 1601–1620.

(35)   Fouquerel, E.; Lormand, J.; Bose, A.; Lee, H.-T.; Kim, G. S.; Li, J.; Sobol, R. W.; Freudenthal, B. D.; Myong, S.; Opresko, P. L. Oxidative Guanine Base Damage Regulates Human Telomerase Activity. *Nat. Struct. Mol. Biol.* **2016**, *23* (12), 1092–1100.

(36)   Pope-Varsalona, H.; Liu, F. J.; Guzik, L.; Opresko, P. L. Polymerase η Suppresses Telomere Defects Induced by DNA Damaging Agents. *Nucleic Acids Res.* **2014**, *42* (21), 13096–13109.

(37)   Moreira, B. G.; You, Y.; Owczarzy, R. Cy3 and Cy5 Dyes Attached to Oligonucleotide Terminus Stabilize DNA Duplexes: Predictive Thermodynamic Model. *Biophys. Chem.* **2015**, *198*, 36–44.

(38)   O'Sullivan, R. J.; Arnoult, N.; Lackner, D. H.; Oganesian, L.; Haggblom, C.; Corpet, A.; Almouzni, G.; Karlseder, J. Rapid Induction of Alternative Lengthening of Telomeres by Depletion of the Histone Chaperone ASF1. *Nat. Struct. Mol. Biol.* **2014**, *21* (2), 167–174.

(39)   Pickett, H. A.; Reddel, R. R. Molecular Mechanisms of Activity and Derepression of Alternative Lengthening of Telomeres. *Nat. Struct. Mol. Biol.* **2015**, *22* (11), 875–880.

(40)   O'Sullivan, J. N.; Bronner, M. P.; Brentnall, T. A.; Finley, J. C.; Shen, W. T.; Emerson, S.; Emond, M. J.; Gollahon, K. A.; Moskovitz, A. H.; Crispin, D. A.; et al. Chromosomal Instability in Ulcerative Colitis Is Related to Telomere Shortening. *Nat. Genet.* **2002**, *32* (2), 280–284.

(41)   Englund, E. A.; Appella, D. H. Synthesis of γ-Substituted Peptide Nucleic Acids: A New Place to Attach Fluorophores without Affecting DNA Binding. *Org. Lett.* **2005**, *7* (16), 3465–3467.

5.7   Supporting Figures
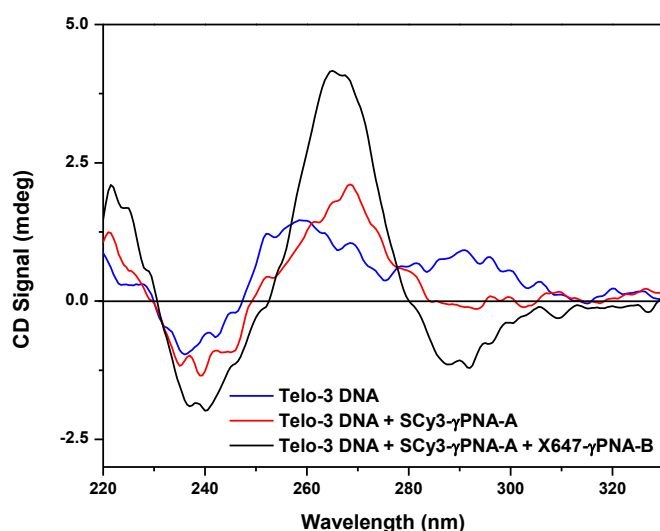
Supporting Figure 5.1 Circular dichroism spectra of Telo-3 DNA (black), Telo-3 DNA + SCy3-γPNA-A (red) and Telo-3 DNA + SCy3-γPNA-A + X647-γPNA-B (blue). DNA and γPNA concentration was 1 μM.
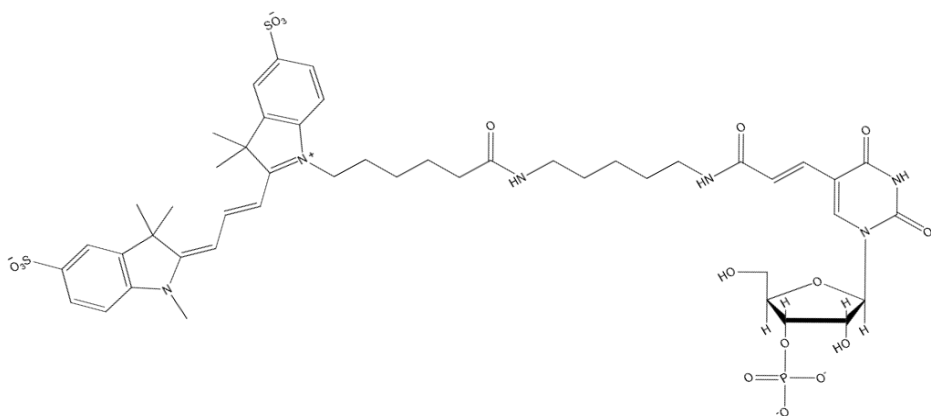
Supporting Figure 5.2 Structure of internal SCy3 fluorophore.

6   Chapter 6. Conclusions and Contributions to the G-Quadruplex Field

The projects discussed in this thesis highlight my work on guanine quadruplexes (GQs). I am interested in the biophysical characteristics of GQs and utilizing various binding modalities to target and disrupt GQs. Broadly, I have shown GQ targeting using three different binding modes: small molecule binding, complementary binding, and homologous binding (where the latter two utilize nucleic acid analogs as the probe). As a result of this work, I have expanded the understanding of GQ secondary structures and shown the ability to target these GQs using multiple binding modes. Here, I will discuss the future for the field as it relates to each chapter and highlight opportunities for further development.

Chapter 2 analyzes variations on a short peptide nucleic acid (PNA) probe that homologously binds to guanine-rich telomeric DNA. We compare the PNA probe with and without backbone modifications and utilize a GQ binding small molecule fluorophore; a combination of homologous and small molecule binding. This work identifies that TO-labeled probes show significantly increased thermal stability compared to their unlabeled counterparts. Modifications inducing a right handed helix in the PNA show no difference in thermal stability. This short probe does not possess selectivity as evidenced by in vitro translation experiments performed in cell lysate. Although these probes were effective in buffer systems, they require improvement in selectivity to be effective in more complex, biological systems. Incorporating an additional binding mode, such as a complementary binding sequence, would increase the selectivity. This was more recently analyzed by Basu and colleagues using a DNA oligonucleotide probe containing a duplex and quadruplex region[1]. The DNA probe targeted the 5'-UTR and coding sequence of RNA, where the duplex region allowed for increased selectivity to the location of interest, and the quadruplex region formed a GQ with a guanine-rich segment of the RNA[1]. This probe successfully inhibited protein translation in vitro and in vivo, demonstrating the potential impact of this type of probe[1].

In Chapter 3, we expand the current literature on RNA:DNA heteroquadruplexes (RDQ). This GQ contains both RNA and DNA guanine tracts. Its formation has been suggested at telomeres, and was

more recently identified at the CSB-II region of human mitochondrial DNA where the RDQ acts as a regulator between transcription and translations[2,3]. We show the formation and stabilities of three suspected RDQ sequences using an optimized scaffold duplex. This is the first study comparing RDQ sequences, showing the biophysical similarities between RDQs. We also show how RDQs have a combination of GQ characteristics, reflecting their hybrid, DNA+RNA makeup. For example, melting temperatures are between those observed for DNA:DNA and RNA:RNA intermolecular GQs. Additionally, dissociation constants with a GQ selective fluorophore fit within those calculated for intramolecular DNA and RNA GQs. This chapter provides a better understanding of RDQs, however, there is significantly less literature regarding RDQs than for other types of GQs. It is important to further our understanding of RDQs with additional structural analysis. These include NMR or crystal structures in order to elucidate the potential roles of loops and overhangs in RDQ structure and function.

Chapter 4 utilizes the RDQ systems created in Chapter 3 and compares them to DNA and RNA homoquadruplexes. We make DNA:DNA and RNA:RNA homoquadruplexes using the scaffold duplex created in Chapter 3 and is the first work to compare homoquadruplex biophysical properties to RDQs. A modified PNA oligomer is then used to duplex-invade each quadruplex. To understand the potential of the PNA probes, polymerase assays developed by the Tan lab could be used to study PNA invasion of the CSB-II RDQ[4]. This RDQ has been shown to serve as a regulator between transcription and translation. Using polymerase assays, we could further understand this regulatory role and be able to modulate transcription and translation with the addition of PNA.

Chapter 5 describes the use of dual gammaPNA (γPNA) probes that hybridize to alternating sites along a telomere. Each probe contains a fluorophore and can undergo Förster resonance energy transfer (FRET). We study the biophysical characteristics of these probes and optimize FRET efficiency by modifying the fluorophores. These miniprobes showed bright staining of telomeres in nuclei. The use of FRET for monitoring probe binding eliminated the need for wash steps, allowing the visualization of critically short telomeres. We also show the incorporation of additional donor fluorophores does not significantly improve FRET efficiency. Future work to improve brightness could utilize a hybridization

chain reaction in order to increase the overall signal and increase visualization of critically short telomeres[5,6].

Much of my thesis work merges the *characterization* of GQs with *targeting* of GQs. Our lab often utilizes multiple binding modes to target GQs, to tailor selectivity and affinity. Types of binders include small molecules, which bind via base stacking, and nucleic acid analogues, which interact with the target via homologous and complementary binding. Small molecules are useful as light-up fluorophores; however, their binding domain is often confined to the tetrad surface and grooves. This leads to poor selectivity. To improve the selectivity, I incorporate a nucleic acid analogue probe (γPNA), utilizing homologous binding probes in Chapter 2, and complementary probes in Chapters 4 and 5. These experiments highlight that guanine-rich homologous probes are not always selective, especially in lysate systems where the probe can bind to many genomic regions. The use of a complementary probe can provide additional selectivity. Probes that utilize dual binding modes have a variety of therapeutic applications, including GQs located within promoter regions such as KRAS, since they can be tailored to optimize selectivity and affinity.

The advancements made by the scientific community in the past few decades regarding GQ research have been illuminating. They span from the development of antibodies for GQ detection in vivo to manipulating GQ structures via molecular probes to modulate their function[7,8]. While there are still ongoing discussions regarding the effect of binding moieties on GQ stability and how they may preferentially form GQs, these methods are crucial for the identification of in vivo GQs. As a field, creating selective GQ methods will be crucial for identifying different types of GQs, including the RDQs and heteroquadruplexes discussed in this thesis. With these goals in mind, the GQ field can continue to grow through identification, characterization, and function modulation.

References

(1)   Bhattacharyya, D.; Nguyen, K.; Basu, S. Rationally Induced RNA:DNA G-Quadruplex Structures Elicit an Anticancer Effect by Inhibiting Endogenous EIF-4E Expression. *Biochemistry* **2014**, *53* (33), 5461–5470.

(2)   Xu, Y.; Ishizuka, T.; Yang, J.; Ito, K.; Katada, H.; Komiyama, M.; Hayashi, T. Oligonucleotide

Models of Telomeric DNA and RNA Form a Hybrid G-Quadruplex Structure as a Potential Component of Telomeres. *J. Biol. Chem.* **2012**, *287* (50), 41787–41796.

(3) Zheng, K.; Wu, R.; He, Y.; Xiao, S.; Zhang, J.; Liu, J.; Hao, Y.; Tan, Z. A Competitive Formation of DNA:RNA Hybrid G-Quadruplex Is Responsible to the Mitochondrial Transcription Termination at the DNA Replication Priming Site. *Nucleic Acids Res.* **2014**, *42* (16), 10832–10844.

(4) Zhang, J.; Zheng, K.; Xiao, S.; Hao, Y.; Tan, Z. Mechanism and Manipulation of DNA:RNA Hybrid G-Quadruplex Formation in Transcription of G-Rich DNA. *J. Am. Chem. Soc.* **2014**, *136* (4), 1381–1390.

(5) Huang, J.; Wang, H.; Yang, X.; Quan, K.; Yang, Y.; Ying, L.; Xie, N.; Ou, M.; Wang, K. Fluorescence Resonance Energy Transfer-Based Hybridization Chain Reaction for: In Situ Visualization of Tumor-Related MRNA. *Chem. Sci.* **2016**, *7* (6), 3829–3835.

(6) Choi, H. M. T.; Chang, J. Y.; Trinh, L. A.; Padilla, J. E.; Fraser, S. E.; Pierce, N. A. Programmable in Situ Amplification for Multiplexed Imaging of MRNA Expression. *Nat. Biotechnol.* **2010**, *28* (11), 1208–1212.

(7) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative Visualization of DNA G-Quadruplex Structures in Human Cells. *Nat. Chem.* **2013**, *5* (3), 182–186.

(8) Ma, D.-L.; Zhang, Z.; Wang, M.; Lu, L.; Zhong, H.-J.; Leung, C.-H. Recent Developments in G-Quadruplex Probes. *Chem. Biol.* **2015**, *22* (7), 812–828.