

Targeting and Biological Implications of Guanine Quadruplexes

Submitted in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy
The Department of Biological Sciences

Karen A. Kormuth

Carnegie Mellon University
Pittsburgh, PA

April 4, 2016

Abstract

The multiplicity of nucleic acid structural conformations allows these molecules to drive intermolecular interactions and support essential processes within biological systems. Of these structures, the guanine (G) quadruplex has been demonstrated to function in the majority of cellular processes involving nucleic acids. Driven by a goal of developing a better understanding of the role of G-quadruplexes in ribosomal genes, we have identified new, noncanonical G-quadruplex motifs. Characterization of these structures has allowed us to learn about how they interact with synthetic nucleic acid analogs *in vitro* and has provided clues to the *in vivo* functions of these quadruplexes in yeast.

Chapter 2 is a summary of our analysis of homologous, G-rich, peptide nucleic acid (PNA) hybridization to noncanonical, long-looped G-quadruplexes. We have confirmed that these PNAs, initially developed in the Armitage group, are compatible with long-looped quadruplexes. In recent years, the acceptance of noncanonical quadruplex motifs, including those with long loops, has been established within the field, enhancing the relevance of our studies. Surprisingly, we discovered that long-looped quadruplexes accommodate higher order heteroquadruplex formation with homologous PNA, wherein >2 PNA strands hybridize to 1 DNA strand. In the future, we intend to explore strategies to exploit the unique features of long-looped quadruplexes to allow for the design of more selective and efficacious PNA-based probes.

Chapter 3 is an expansion of previous studies in our group focused on analysis of γ -modified PNA hybridization to DNA G-quadruplex targets. We have

demonstrated the compatibility of homologous, diethylene glycol-modified γ PNAs with DNA G-quadruplex invasion. We have also explored PNA/DNA interactions under *in vitro* conditions that mimic the crowded intracellular environment, determining that molecular crowding enhances thermodynamics of PNA/DNA hybridization. These results help to confirm the applicability of similar PNAs to targets within biological systems.

In Chapter 4, we initiated studies of our noncanonical G-quadruplex motifs in yeast ribosomal (r) DNA and rRNA. Our efforts also included the development of better systems with which to introduce and assay rDNA quadruplex mutations. Our results support the importance of quadruplex motifs in ribosome biogenesis, as well as the unprecedented hypothesis that specific regions of the rRNA may take on a quadruplex fold at certain points during assembly or translation.

Taken together, this thesis contributes to our understanding of quadruplex biology and the interactions between these structures and synthetic PNA probes. Future development of selective and biologically compatible PNAs will be required to assist verification of quadruplex folding and function in cells. The human genome contains >700,000 G-quadruplexes, which hints at the potential for exciting discoveries in G-quadruplex biology yet to come.

Table of Contents

Title Page.....	i
Abstract.....	ii
Table of Contents.....	iv
List of Figures.....	vii
List of Tables.....	xiv
List of Schemes.....	xvi
Acknowledgements.....	xvii

1. Chapter 1: G-quadruplex Biology and Strategies for Synthetic Regulation of G-quadruplex function

1.1 G-quadruplex Structure and Biology.....	1
1.1.1 G-quadruplex Structure.....	2
1.1.2 Conventional vs. Updated G-quadruplex Motifs.....	5
1.1.3 G-quadruplexes in DNA.....	8
1.1.4 G-quadruplexes in RNA.....	14
1.1.5 G-quadruplexes in Eukaryotic Ribosomal Genes.....	16
1.2 Artificial Modulation of G-quadruplex function.....	21
1.2.1 Small Molecules.....	21
1.2.2 Antisense Oligonucleotides.....	22
1.2.3 Complementary Peptide Nucleic Acids.....	25
1.2.4 Homologous Peptide Nucleic Acids.....	31
1.3 Thesis Overview.....	34
1.4 References.....	34

2. Chapter 2: Homologous PNA Hybridization to Noncanonical DNA G-quadruplexes

2.1 Chapter Summary.....	39
2.2 Introduction.....	39
2.3 Results.....	44
2.4 Discussion.....	56
2.5 Conclusion.....	62
2.6 Future Directions.....	62
2.7 Materials and Methods.....	65
2.7.1 Materials.....	65

2.7.2	Equipment	66
2.7.3	UV Melting	66
2.7.4	Thermal Difference Spectroscopy	67
2.7.5	Circular Dichroism.....	67
2.7.6	Surface Plasmon Resonance	67
2.7.7	Electrospray Ionization Mass Spectrometry (ESI-MS)	68
2.8	Acknowledgement.....	69
2.9	Appendix	70
2.10	References	78
3.	Chapter 3: Molecular Crowding Enhances γ-modified and Unmodified Homologous PNA Hybridization to Discrete DNA G-quadruplex Targets	
3.1	Chapter Summary	83
3.2	Introduction	84
3.3	Results	89
3.4	Discussion	95
3.5	Conclusion and Future Perspective	98
3.6	Materials and Methods.....	100
3.6.1	Materials	100
3.6.2	Equipment	100
3.6.3	UV Melting	100
3.6.4	Circular Dichroism.....	101
3.7	Acknowledgement.....	101
3.8	References	101
4.	Chapter 4: Evidence for the Biological Significance of Ribosomal DNA G-quadruplex Motifs in Ribosomal RNA Biogenesis	
4.1	Chapter Summary	104
4.2	Introduction	105
4.3	Results	110

4.4 Discussion	120
4.5 Conclusion	125
4.6 Future Directions	126
4.7 Materials and Methods	129
4.7.1 Materials	129
4.7.2 Equipment	130
4.7.3 UV Melting	130
4.7.4 Circular Dichroism	130
4.7.5 Plasmid Mutagenesis	130
4.7.6 Yeast Strains for rDNA Mutant Analysis	131
4.7.7 Standard Spotting Assay for Cell Growth	132
4.7.8 Polysome Profiles	132
4.7.9 Northern Blotting	132
4.7.10 <i>In vivo</i> SHAPE	133
4.8 Acknowledgements	134
4.9 Appendix	135
4.10 References	139

List of Figures

Chapter 1

- Figure 1.1.** **2**
G-quadruplex structural features. Four guanines interact via Hoogsteen H-bonding to form a planar G-tetrad (A; adapted from Bochman *et al.*, 2012). The canonical quadruplex motif contains 4 G-tracts that interact to form the body of the quadruplex, with stacked G-tetrads and the intervening sequences forced out as loops (B).
- Figure 1.2.** **3**
Variations in G-quadruplex topology (adapted from Chen and Yang, 2012). Parallel G-quadruplexes achieve uniform strand polarity using propeller loops (left), while antiparallel quadruplexes can accommodate both diagonal (center) and lateral loops (right).
- Figure 1.3.** **12**
DNA G-quadruplex regulation of gene expression during transcription. G-quadruplexes near promoters tend to prevent transcription initiation, especially when further stabilized through bound repressors (green). Depending on the DNA strand on which they fold, stable G-quadruplexes can either enhance or inhibit transcription elongation by stalling the polymerase (blue) or holding open the transcription bubble.
- Figure 1.4.** **18**
Eukaryotic rDNA repeating unit from the yeast, *S. cerevisiae*. 5S rRNA is transcribed by RNA polymerase III, while the remaining rRNAs and spacers are transcribed by RNA polymerase I.
- Figure 1.5.** **25**
Comparison of DNA (left) and PNA (right) backbone chemical structures.
- Figure 1.6.** **31**
Homologous vs. complementary PNA hybridization to G-quadruplexes (adapted from Gupta *et al.*, 2013). Complementary probes hybridize by Watson-Crick basepairing, while homologous probes form G-tetrads via Hoogsteen H-bonding. The complementary probe unwinds the quadruplex to form a duplex (left), while the homologous probe invades the target forming an intermolecular heteroquadruplex (right).
- Figure 1.7.** **33**
PNA backbone and base modifications (adapted from Lusvarghi *et al.*, 2009). Modification at the γ carbon introduces chirality to the monomer (center). G-tracts can be separated with an abasic, miniPEG monomer (right), which provides flexibility and reduces affinity for potential duplex targets.

Chapter 2

Figure 2.1. 48

Long loop-containing DNAs fold into G-quadruplexes *in vitro*. A) CD spectra of QFS3, QFS3_{L2}, and QFS3_{L3} in 100 mM K⁺. B) TDS of QFS3, QFS3_{L2}, QFS3_{L3}, and hTERT in 100 mM K⁺. The inset graph highlights the region around 295 nm, which includes a quadruplex-specific inversion for each oligomer.

Figure 2.2. 49

P_{eg2} forms DNA/PNA heteroquadruplexes with long loop-containing DNA G-quadruplexes. A) CD spectra of each **P_{eg2}**/DNA complex in 100 mM K⁺ suggests overall more parallel topology (min. at 240 nm, max. at 260 nm) than those of the DNA alone (Figure 1A). B) TD spectroscopy in 100 mM K⁺ detects G-quadruplex formation for each complex as evidenced by the inversion near 295 nm. The inset graph highlights the region around 295 nm.

Figure 2.3. 50

P_{eg2} hybridization to QFS3 results in PNA/DNA heteroquadruplex formation. SPR sensorgrams of **P_{eg2}** binding to immobilized QFS3 in 100 mM KCl (left). In 100 mM LiCl, the response is lost, indicating the molecules do not interact in quadruplex-destabilizing conditions (right). [**P_{eg2}**] = 5, 10, 25, 50, 75 and 100 nM.

Figure 2.4. 51

P_{eg2} readily associates with long-looped G-quadruplexes. (A) SPR sensorgrams of 50 nM **P_{eg2}** binding to various immobilized DNA quadruplexes in 100 mM K⁺. (B) Sensorgrams are normalized to the maximum RU_{max} value for each quadruplex to allow for visual comparison of on/off rates of **P_{eg2}** to each target.

Figure 2.5. 52

P_{eg2} forms a higher order heteroquadruplex with the long-looped DNA G-quadruplex, QFS3. SPR sensorgrams of **P_{eg2}** binding to immobilized QFS3 in 100 mM K⁺. [**P_{eg2}**] = 100, 250, 500, 750, and 1000 nM. The red dashed lines indicate theoretical RU_{max} for 2:1, 5:1, and 6:1 PNA:DNA.

Figure 2.6. 54

Detection of T-loop DNAs G-quadruplexes and DNA/PNA heteroquadruplexes *in vitro*. A) CD spectra and TDS (B) of T2, T4, and T6 with and without **P_{eg2}** in 10 mM K⁺.

Figure 2.7. 56

Higher order complex formation between **P_{eg2}** PNA and three T-loop G-quadruplexes in 100 mM KCl detected by SPR. Signal saturation of PNA binding to immobilized T2 ([**P_{eg2}**] = 100, 200, 400, and 500 nM) occurs close to the theoretical RU_{max} corresponding to 4:1 stoichiometry. Signal saturation of PNA binding to immobilized T4 ([**P_{eg2}**] = 100, 200, 400, and 500 nM) occurs close to the theoretical RU_{max} corresponding to 5:1 stoichiometry. Signal saturation of

PNA binding to immobilized T6 ($[P_{eg2}] = 100, 200, 400, \text{ and } 500 \text{ nM}$) occurs close to the theoretical RU_{max} corresponding to 5:1 stoichiometry.

Figure 2.8. **65**
Small molecule quadruplex ligands. NMM (www.molbase.com), TmPyP4, and quindoline (adapted from Chen and Yang 2012).

Figure 2.S1. **70**
TDS of Myc19 DNA and Myc19 in complex with P_{eg2} PNA. Pre-annealed samples ($1 \mu\text{M}$ DNA and $1 \mu\text{M}$ DNA + $2 \mu\text{M}$ PNA) were buffered in $100 \text{ mM } K^+$, 10mM Tris pH 7, and $0.1 \text{ mM } Na_2EDTA$. The TD spectrum of Myc19 alone (black) shares similar spectral features with that of QFS3_{L3}. Although these features are less pronounced in the TDS of the DNA/PNA complex (black), both spectra contain the G-quadruplex-specific inversion around 295 nm .

Figure 2.S2. **71**
DNA G-quadruplexes containing long loops fold into intramolecular structures *in vitro*. UV melting analyses (295 nm) indicate that the T_m of each quadruplex [(A) QFS3, (B) QFS3_{L2}, (C) QFS3_{L3}, (D) hTERT] is not dependent on DNA concentration. DNA samples were prepared in buffer containing $100 \text{ mM } K^+$, except for hTERT, which required a lower ionic strength ($10 \text{ mM } K^+$) to allow melting.

Figure 2.S3. **71**
UV melting curves recorded for QFS3 in the presence of either 100 mM KCl (black) or LiCl (red). The lower T_m and hypochromicity for LiCl is consistent with the known quadruplex-destabilizing effect of Li^+ relative to K^+ .

Figure 2.S4. **72**
 P_{eg2} PNA enhances the thermal stability of each long-looped DNA G-quadruplex. UV melting analyses of G-quadruplex DNA with (red) and without (black) P_{eg2} at 295 nm . $1 \mu\text{M}$ DNA [QFS3 (A), QFS3_{L2} (B), QFS3_{L3} (C), and hTERT (D)] was melted alone or in the presence of $2 \mu\text{M } P_{eg2}$. Samples were buffered in 10mM (hTERT) or $100 \text{ mM } K^+$, 10 mM Tris (pH 7), and $0.1 \text{ mM } Na_2EDTA$.

Figure 2.S5. **73**
Comparison of DNA G-quadruplex and DNA/PNA heteroquadruplex structural topology by CD at $30 \text{ }^\circ\text{C}$. Pre-annealed samples ($1 \mu\text{M}$ DNA [black] or $1 \mu\text{M}$ DNA + $2 \mu\text{M}$ PNA [red]) were buffered in $100 \text{ mM } K^+$, 10 mM Tris pH 7, and $0.1 \text{ mM } Na_2EDTA$. The presence of the PNA intensifies the parallel features (min. at 240 nm , max. at 260 nm) of the CD spectra for each DNA target: (A) QFS3, (B) QFS3_{L2}, (C) QFS3_{L3}, and (D) hTERT.

Figure 2.S6. **74**
 P_{eg2} interaction with long-looped DNA G-quadruplexes is greatly enhanced under quadruplex-stabilizing conditions. SPR sensorgrams of P_{eg2} ($[PNA] = 5, 10, 25,$

50, 75, and 100 nM) binding to immobilized DNA in 100 mM K⁺ (left) or 100 mM Li⁺ (right).

Figure 2.S7.

75

P_{eg2} forms stable intermolecular heteroquadruplexes with QFS3 under conditions used in ESI-MS. UV melting (A), CD (B; 30 °C), and TDS (C) detect quadruplex formation *in vitro*. Samples containing only QFS3 DNA (black), 5:1 **P_{eg2}**:QFS3 (red), and 6:1 **P_{eg2}**:QFS3 (blue) were buffered in ESI-MS buffer containing 150 mM NH₄OAc.

Figure 2.S8.

76

ESI-MS detects peaks corresponding to 6:1 (left) and 5:1 (right) PNA-DNA.

Figure 2.S9.

76

Higher order complex formation between **P_{eg2}** PNA and three long-looped G-quadruplexes in 100 mM KCl detected by SPR. A) Binding to immobilized QFS3_{L2}. [**P_{eg2}**] = 100, 200, 400, 500, 550, and 600 nM. Signal saturation occurs close to the theoretical RU_{max} corresponding to 4:1 stoichiometry. B) Binding to immobilized QFS3_{L3} in 100 mM K⁺. [**P_{eg2}**] = 100, 200, 400, and 600 nM. Signal saturation occurs close to the theoretical RU_{max} corresponding to 4:1 stoichiometry. C) Binding to immobilized hTERT. [**P_{eg2}**] = 100, 200, 400, 500, 550, and 600 nM. Signal saturation occurs close to the theoretical RU_{max} corresponding to 5:1 stoichiometry.

Figure 2.S10.

77

2:1 **P_{eg2}**:Myc19 heteroquadruplex formation detected by SPR. SPR sensorgrams of **P_{eg2}** binding to immobilized QFS3 in 100 mM K⁺. [**P_{eg2}**] = 5, 10, 20, 25, 30, 40, 50, 60, 70, 75, 80, 90, 100, 150, 200, and 250 nM. The red dashed lines indicate theoretical RU_{max} for 1:1 and 2:1 PNA:DNA.

Figure 2.S11.

77

P_{eg2} PNA enhances the thermal stability of each T-loop DNA G-quadruplex. UV melting analyses of G-quadruplex DNA with and without **P_{eg2}** at 295 nm. 1 μM of each DNA was melted alone or in the presence of 2 μM **P_{eg2}**. Samples were buffered in 10mM (hTERT) or 10 mM K⁺, 10 mM Tris (pH 7), and 0.1 mM Na₂EDTA.

Figure 2.S12.

78

P_{eg2} interaction with T-loop DNA G-quadruplexes is greatly enhanced under quadruplex-stabilizing conditions. SPR sensorgrams of **P_{eg2}** ([PNA] = 5, 10, 25, 50, 75, and 100 nM) binding to immobilized DNA in 100 mM K⁺ (left) or 100 mM Li⁺ (right).

Chapter 3

- Figure 3.1.** **85**
Chemical structure of the **P_{eg}2** PNA probe.
- Figure 3.2.** **92**
Effect of molecular crowding on Myc19, γ PNA heteroquadruplex formation. Molecular crowding with PEG 200, but not glycerol, enhances the thermal stability of the heteroquadruplex as measured by UV melting with both **γ 6P_{eg}2** (A) and **γ 2P_{eg}2** (B). Molecular crowding of heteroquadruplexes, including fully (C) or partially modified (D) PNA, with PEG 200 amplifies the parallel CD signal (min. at 240 nm, max. at 260 nm).
- Figure 3.3.** **94**
Effect of molecular crowding on hTelo22, γ PNA heteroquadruplex formation. Molecular crowding with PEG 200, but not glycerol, enhances the thermal stability of the heteroquadruplex as measured by UV melting for both the fully modified (A) and partially modified (B) PNA probe. Molecular crowding with PEG 200 amplifies the CD signal of the hTelo22/ γ **6P_{eg}2** heteroquadruplex, although the spectrum does not become fully parallel (min. at 240 nm, max. at 260 nm) until maximal molecular crowding at 40% PEG 200 (C). Molecular crowding with PEG 200 amplifies the CD signal of the hTelo22/ γ **2P_{eg}2** heteroquadruplex, although it retains a broad peak from 280-300 nm, which suggests hybrid/mixed quadruplex topology (D).
- Figure 3.4.** **95**
In vitro molecular crowding of QFS3, QFS3/**P_{eg}2** G-quadruplexes. (A) The addition of PEG 200 increases the thermal stability of both QFS3 and the QFS3/**P_{eg}2** heteroquadruplex, as determined by UV melting. (B) PEG 200 enhances the parallel character (min. at 240 nm, max. at 260 nm) of both QFS3 and the QFS3/**P_{eg}2** heteroquadruplex. The structure displaying the most parallel character is the crowded heteroquadruplex.

Chapter 4

- Figure 4.1.** **108**
rDNA mutagenesis system to assay mutant rRNA phenotypes. Mutations are introduced into an RNA Polymerase I-driven rDNA vector having neutral sequence tags with the 18S and 25S genes. Tagged, plasmid-derived rRNA can be detected separately from chromosome-derived rRNA, allowing for analysis of potentially lethal recessive mutations.
- Figure 4.2.** **108**
rDNA mutagenesis system to assay mutant rRNA growth phenotypes. Mutations are introduced into an RNA Polymerase I-driven rDNA vector (*TRP1*), which is transformed into cells having all chromosomal rDNA repeats deleted and

containing a wildtype rDNA vector (*URA3*). Transformants are grown on 5-FOA to select for loss of the *URA3* vector, resulting in the restriction of all cellular rRNA production to the mutant plasmid.

Figure 4.3. **111**

Positions of QFS1-4 in the yeast repeating rDNA unit. QFS2 and QFS3 are overlapping G-quadruplex motifs derived from a series of G-tracts close together in the 25S region of the repeat.

Figure 4.4. **112**

CD spectra of QFS1-4 in dilute (A) and crowded (B) conditions. Molecular crowding with PEG 200 drives the topology of each quadruplex to a more parallel structure (min. at 240 nm, max. at 260 nm).

Figure 4.5. **114**

QFS2/3 mutations designed to modulate G-quadruplex stability. G-tract mutations are shown in red and compensatory mutations, (designed to maintain the predicted stem-loop secondary structure) are shown in blue. QFS2/3 secondary structure is adapted from <http://www.rna.icmb.utexas.edu>. G-tracts predicted to participate in G-tetrad formation are indicated in red, dashed boxes. QFS2/3 mutant 1 is designed to destabilize G-quadruplex structure by mutation of the central G's in each G-tract. QFS2/3 mutant 2 is designed to stabilize G-quadruplex structure through the introduction of an additional G-tract in the long 3' loop of QFS3. QFS2/3 mutant 3 is designed to stabilize G-quadruplex structure by conversion of G₃-tracts to G₄-tracts.

Figure 4.6. **115**

QFS2/3 mutants 1 and 2 growth phenotypes. Serial dilutions of cell cultures were spotted on C-trp and incubated at 30 °C (top) or 18 °C (bottom). The top row of spots in each image is the wildtype rDNA control. QFS2/3 mutant 2 grows similarly to wildtype, while QFS2/3 mutant 1 is cold sensitive.

Figure 4.7. **116**

Polysome profiles of QFS2/3 mutants at 30 °C (top) and 16 °C (bottom). Compared to wildtype, QFS2/3 mutants 1 and 2 both produce a reduced 60S peak phenotype at 30 °C, which is further exaggerated for QFS2/3 mutant 1 at 16 °C.

Figure 4.8. **117**

rRNA phenotypes of QFS2/3 mutants as determined by northern blotting. RNA was extracted from cells containing wildtype (WT), empty vector (EV), or mutant plasmids, cultured at 30 °C or 16 °C. Plasmid-derived 18S and 25S rRNAs were detected using probes complementary to the neutral sequence tags inserted within the genes. Northern blots (top) were quantified (bottom) using ImageJ. The quantified band intensities represent the ratio of each rRNA to the total 18S + 25S for each sample (n = 2). U2 snRNA is included as a loading control. The

25S:18S ratios for mutants 2 and 3 (both designed to stabilize QFS2/3 quadruplexes) decrease slightly compared to the wildtype.

Figure 4.9. **120**
QFS2/3 mutant SHAPE reactivity profiles. Reverse transcription products of wildtype and QFS2/3 mutant rRNAs were run on a sequencing gel (left) following treatment with a SHAPE reagent (NAI) or vehicle control (DMSO). Nucleotides that react differently in the mutants compared to the wildtype are marked with red (more reactive) or blue (less reactive) spheres. SHAPE reactivity profiles for QFS2/3 mutants 1 and 2 are plotted on the predicted wildtype secondary structure [red (more reactive), blue (less reactive)]. Wildtype G-tracts are marked with dashed red boxes. QFS2/3 mutant 1 (destabilizes G-quadruplexes) changes the SHAPE reactivity profile to more directly reflect the predicted stem-loop secondary structure in the mature 25S rRNA.

Figure 4.S1. **135**
QFS1-4 fold into intramolecular G-quadruplexes *in vitro*. Absorbance spectra at 295 nm were measured by UV melting for 1 μ M and 10 μ M DNA : (A) QFS1, (B) QFS2, (C) QFS3, and (D) QFS4.

Figure 4.S2. **136**
Effect of molecular crowding on QFS1-4 thermal stability. Molecular crowding with PEG 200, but not glycerol, enhances the thermal stability as measured by UV melting of (A) QFS1, (B) QFS2, (C) QFS3, and (D) QFS4.

Figure 4.S3. **137**
QFS1 mutant design and northern blot analysis. (A) QFS1 mutant 1 includes quadruplex-destabilizing mutations (red; 5'G₃ to U₃), which are superimposed on the wildtype secondary structure of this region of the 18S rRNA. QFS1 secondary structure is adapted from <http://www.rna.icmb.utexas.edu>. Compensatory mutations (mutants 2 and 3) designed to reestablish the disrupted helix are listed in the blue dashed box. (B) Northern blots detect almost complete loss of 18S rRNA in mutant 1, which is restored in both compensatory mutants, while 25S rRNA is unaffected. U2 snRNA is included as a loading control.

List of Tables

Chapter 2

- Table 2.1.** **45**
Sequences of PNA probes and DNA G-quadruplex targets. (G-tracts predicted or reported to participate in tetrad formation shown in bold and underlined.)
- Table 2.2.** **46**
Concentration dependence of DNA quadruplex melting temperature (T_m , °C) determined in 100 mM K^+ .
- Table 2.3.** **48**
PNA-DNA heteroquadruplex melting temperature (T_m , °C) and stabilization relative to DNA homoquadruplex (ΔT_m , °C) determined in 100 mM K^+ .
- Table 2.4.** **51**
PNA association (on) and dissociation (off) rates in $RU \cdot s^{-1}$ and ratio (Off/On) determined at 50 nM P_{eg2} .
- Table 2.5.** **53**
Sequences of T-loop DNA G-quadruplex targets. (G-tracts predicted to participate in tetrad formation shown in bold and underlined.)
- Table 2.6.** **54**
DNA quadruplex and DNA/PNA heteroquadruplex melting temperatures (T_m , °C) determined by UV melting in 10 mM K^+ .
- Table 2.S1.** **75**
 T_m values (°C) of P_{eg2} /QFS3 Complexes in 150 mM NH_4^+ .

Chapter 3

- Table 3.1.** **90**
Homologous PNA probes and DNA G-quadruplex targets. (G-tracts predicted or reported to participate in tetrad formation are shown in bold and underlined.)
- Table 3.2.** **91**
Thermal stability (T_m , °C) of Myc19/ γ PNA heteroquadruplexes under molecular crowding conditions.
- Table 3.3.** **93**
Thermal stability (T_m , °C) of hTelo22/ γ PNA heteroquadruplexes under molecular crowding conditions (10 mM K^+).

Table 3.4.	95
Enhanced Thermal Stability (T_m , °C) of QFS3 upon P_{eg2} hybridization, molecular crowding (100 mM K ⁺).	
Chapter 4	
Table 4.1.	110
Yeast rDNA G-quadruplex motifs. (G-tracts predicted to participate in tetrad formation are shown in bold and underlined.)	
Table 4.2.	111
QFS1-4 T_m (°C) dependence on DNA concentration (left) and molecular crowding (right) determined by UV melting.	
Table 4.3.	115
QFS2/3 mutant growth phenotypes as determined by standard spotting assays. Plates were incubated at 30 °C and 18 °C.	
Table 4.S1.	137
<i>S. cerevisiae</i> strain list.	
Table 4.S2.	139
<i>S. cerevisiae</i> rDNA G-quadruplex motif mutations.	

List of Schemes

Chapter 2

Scheme 2.1. **43**
A homologous targeting strategy for PNA hybridization to G-quadruplex DNA. A short, homologous PNA can invade a DNA G-quadruplex forming a 2PNA:1DNA heteroquadruplex.

Scheme 2.2. **60**
(A) Model 5:1 and 6:1 PNA:DNA heteroquadruplex structures. (B) Q1-Q4 represented in an end-to-end stacked conformation.

Scheme 2.3. **64**
Strategies for complementary probe hybridization to QFS3. A complementary PNA (red) targeted to the long 3' loop should stabilize the quadruplex (top left), while a PNA targeted to one of the G-tracts and part of the loop (top right) should form a partial duplex and destabilize the quadruplex. Conjugation of a small molecule quadruplex ligand (red box; bottom left) to a complementary PNA targeted to the loop should even further stabilize the quadruplex than the PNA alone.

Chapter 4

Scheme 4.1. **125**
A model for structural interconversion of ES27L rRNA

Scheme 4.2. **127**
Proposed sequence tag placement in ITS2 pre-rRNA. The sequence tag (red) is inserted into helix IV of ITS2 (black), which is depicted in its hairpin conformation. Regions of ITS2 required for rRNA processing are marked with dashed blue boxes.

Scheme 4.3. **129**
Chimeric γ PNA-based affinity purification of pre-60S ribosomal subunits. The right-handed domain (red) of the capture strand, which is conjugated to a bead (gray sphere), hybridizes to the sequence tag in ITS2 (red). The particle is eluted by introduction of a chiral γ PNA complementary to both the right- and left-handed (gray) domains of the capture strand. Right- and left-handed domains are separated by an abasic spacer (blue sphere).

Acknowledgements

Reflection upon the past six years has let me to realize how deeply grateful I am for the people who have carried me through graduate school and made the completion of this thesis possible. First, I am so thankful to have experienced the bulk of my graduate study as a co-advised student under Dr. John Woolford and Dr. Bruce Armitage, who welcomed me into their groups when I feared I would have no place to go. Working within these two groups has allowed me to explore science on my own terms, blurring the lines between fields. Despite the challenges I have encountered with my projects, my advisors have provided me with a unique and special experience for which I will always be grateful.

I am also indebted to the Department of Biological Sciences and CNASt, as well as all of their members for providing a collaborative and welcoming research environment. I am especially grateful for the people who eased my transition into the Armitage and Woolford labs, Dr. Nathan Urban and Emily Stark. And I thank Dr. Shoba Subramanian for giving me the chance to productively pursue my passion for writing and contribute my talents to the department.

I am also very thankful to Dr. Mark Macbeth for providing me with my first home in the Mellon Institute. My time with Dr. Macbeth and my labmates, Dr. Any Kehr, Androo Markham, and especially Liz Ransey, allowed me build my confidence as a researcher and develop my love for RNA, which I have carried with me throughout graduate school.

My Armitage labmates have given me a home in the Chemistry Department and have guided me through obstacles as much as the development of new ideas – and taught me how to properly serve cake. I am so grateful to have worked with Dr. Anisha Gupta, Dr. Connor Murphy, Dr. Junriz Delos Santos, Dr. Ha Pham, Dr. Wendy Brotherton, Dr. Tumul Srivastava, Dr. Xiaohong Tan, Dr. Lauren Regula, Dr. Gloria Silva, Stan Oyaghire, Munira Fouz, April Berlyoung, and Dima Kolodieznyi. Dr. Lisa Rastede and Taylor Canady, especially, have pushed me to reach beyond my own, self-imposed limitations and become a better scientist and a better person. My Woolford labmates have equally supported me and my unconventional project. I would not have been able to pursue study of rDNA G-quadruplex biology without Jelena Jakovljevic, Stephanie Biedka, Dr. Michael Gamalinda, and Dr. Jason Talkish. Along with Beril Tutuncuoglu, Dr. Madhu Ramesh, and Dr. Sal Konikkat, who have become dear friends and always provided a soft place to land, all my labmates have supported me throughout the past three years.

Beyond science, graduate school has given me the opportunity to create friendships, which I hope to carry forward into the future. I have been so lucky to know Dr. Shanna Bowman and Dr. Idil Ulengin-Talkish, who are both fierce scientists and great friends. I am also very grateful for my cohort, Dr. Shanna Bowman, Liz Ransey, Dr. Olivia Molinar, Dr. John Pettersson, Dr. Yi Wang, Dr. Jigar Desai, Dr. Gino Gallo, Dr. Sal Konikkat, Dr. Madhu Ramesh, Sahil Sangani, Zhongling Sun, Han Lai, and Ting Liu, who have from the beginning been a warm and supportive group. I am also very thankful to have had my friends outside of

CMU, especially Rachel Stamateris, Katie O'Connor, and Dr. Katie Ryan, to rely on throughout this experience.

Finally, I owe the most gratitude to my two greatest supporters, my mom and dad, Zorica Jarnjevic and Joe Kormuth. Words are not sufficient to express all they have done to positively impact my time here, and I hope that if I have accomplished nothing else, I have at least been able to make them proud.

1. Chapter 1: G-quadruplex Biology and Strategies for Synthetic Regulation of G-quadruplex function

Beyond the canonical double helix, nucleic acids can take on a huge variety of alternate secondary structures that serve to regulate biological processes at both the DNA and RNA levels. In recent years, G-quadruplexes, uniquely-structured nucleic acid elements formed from G-rich motifs, have been emerging as biologically relevant structures with the potential to regulate many biological processes. As new quadruplex functions have been uncovered, simultaneously, new motifs that include a range of noncanonical features have also been defined, greatly expanding the number of potential biological G-quadruplexes. Together, these findings present two major challenges to quadruplex research: 1) to identify which motifs fold into functional G-quadruplexes within the cell, and 2) to understand the function of individual G-quadruplexes. The answer to these questions will require development of selective and effective quadruplex-targeting probes that can aid in the identification of G-quadruplex function and enhance future therapeutic strategies.

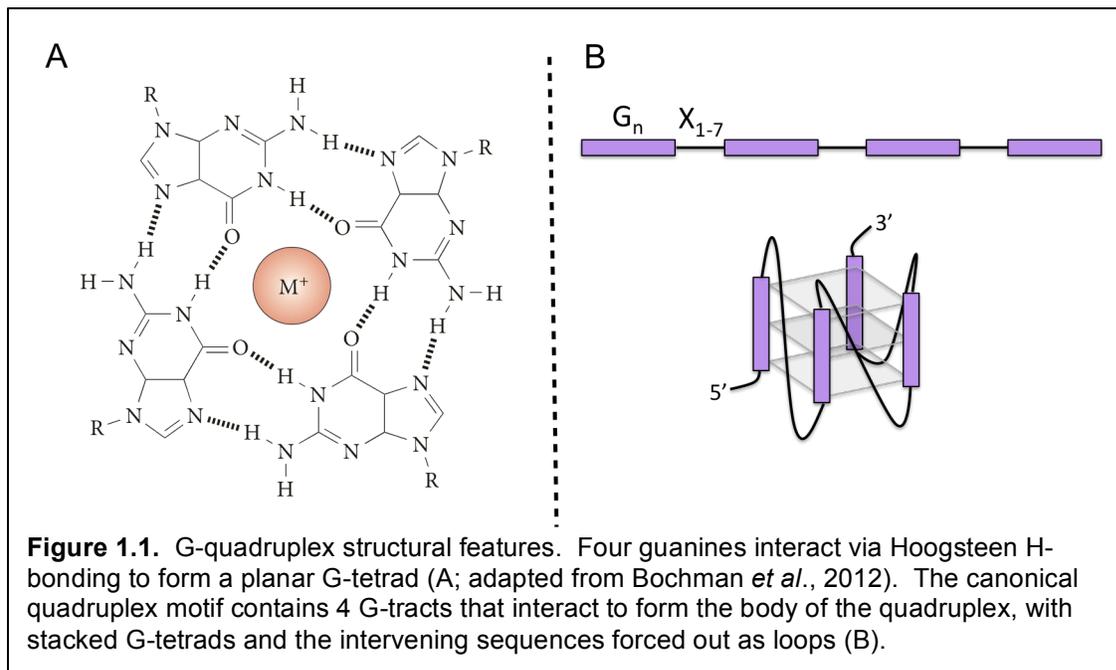
1.1 G-quadruplex Structure and Biology

G-quadruplexes as a whole are an unusual class of structures in that they can be formed from nucleic acids that differ greatly in primary sequence, creating structures with varying topologies. This potential for diversity in quadruplex structure confers these elements with the ability to regulate biological processes by a variety of mechanisms and through interaction with many cellular factors. A broad goal in G-quadruplex biology is to understand the complex relationship

between sequence, structure, and function to allow for efficacious and selective artificial regulation.

1.1.1 G-quadruplex Structure

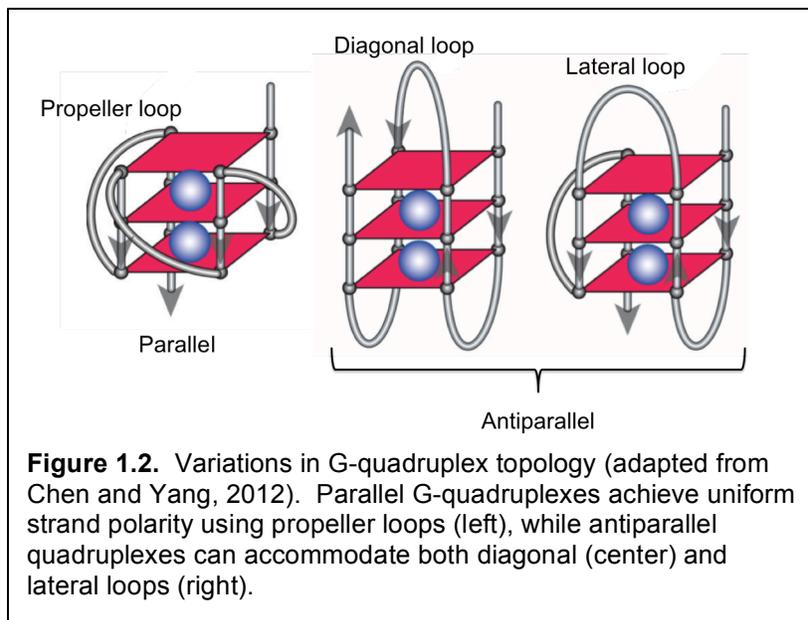
The structural unit around which G-quadruplex stability is based is the G-tetrad, wherein four guanines interact via Hoogsteen H-bonding in a planar arrangement (Figure 1.1A)[1]. In biological systems, G-quadruplexes can form



from one or more G-rich DNA or RNA molecules. Motifs that have been historically defined as being able to support intramolecular G-quadruplex formation typically conform to the following sequence arrangement: $G_m X_n G_m X_o G_m X_p G_m$ [2]. According to this definition, a putative quadruplex motif must contain four G-tracts (each having at least two guanines), with the length of the shortest G-tract determining the number of G-tetrads that can form in the folded structure[2]. The sequences separating the G-tracts ($X_{n,o,p}$), hereafter referred to as loops, can vary in both length and composition, but each range in

length between 1 and 7 nts[3]. In the quadruplex secondary structure, the G-tetrads stack on top of each other and the central cavity is occupied by monovalent cations, which also serve to stabilize the structure (Figure 1.1B). The ionic radius of K^+ is the most ideal for quadruplex stabilization by occupation of the central cavity and minimization of phosphate charge repulsion, although other cations (including Na^+) can also stabilize quadruplexes to varying degrees[4]. Hydration energy also plays a role in quadruplex stability, as water molecules are released upon folding of Hoogsteen-based structures, resulting in enhanced stability of these structures in conditions where hydration is unfavorable, such as the crowded intracellular environment[5].

G-quadruplex topology is considerably more variable than other types of nucleic acid secondary structures. Although G-quadruplexes can be composed



of one or more molecules, for the purpose of clarity, this section will focus on intramolecular, single-stranded structures. G-quadruplexes can differ in strand polarity and loop placement.

Two broad classes of quadruplex structures are parallel (Figure 1.1B; all four G-tracts have the same polarity) and antiparallel (at least one G-tract has an

opposing polarity from the others)[2]. Parallel quadruplexes include guanines having glycosidic angles only in the *anti* conformation and require propeller loops, which connect the bottom G-tetrad directly to the top G-tetrad (Figure 1.2)[2]. Antiparallel quadruplexes include both *syn* and *anti* guanines and can accommodate both lateral loops (connect adjacent G-tracts) and diagonal loops (connect opposite G-tracts; Figure 1.2)[2]. Very recently, Mergny and colleagues identified several quadruplex motifs able to switch conformations following changes in cation identity and concentration[6]. Specifically, they detected a shift from antiparallel to fully parallel topology after addition of small amounts of K^+ to a Na^+ -stabilized structure[6]. This quadruplex was also responsive to changes in K^+ concentration, transitioning from a 3-tetrad (two bound cations) to a 2-tetrad (one bound cation) structure at very low K^+ concentrations[6]. These findings have implications for biological quadruplexes, and may provide a platform for the development of new quadruplex-based K^+ sensors.

Loop length is another factor that directly affects G-quadruplex folding topology. Studies done by Hazel and colleagues on a series of 3-tetrad quadruplexes having thymine (T)-loops of various lengths began to clarify the relationship between loop length and topology through the identification of several rules: 1) motifs having T_1 -loops support only parallel structure, 2) motifs with three T_2 -loops OR at least one T_1 -loop can fold into either conformation, with a preference for parallel topology, 3) motifs having three T-loops between 2 and 6 nts long have a preference for antiparallel topology[7]. Further studies confirmed that the presence of two single nucleotide loops is sufficient to

constrain the quadruplex to a parallel structure[8]. The tendency of more flexible systems (long-looped quadruplexes) to fold into antiparallel conformations can be explained by the contribution of stacking interactions between the bases within the loop, which stabilize the structure[7]. In addition to length, loop sequence identity can also contribute to the conformation and folding dynamics of the quadruplex[9]. In the case of more flexible motifs that can achieve multiple folds, more than one conformation can exist together in solution, adding to the complexity of the system[7].

1.1.2 Conventional vs. Updated G-quadruplex Motifs

Although all G-quadruplexes are defined by the characteristic stacking of multiple G-tetrads, the true scope of variation in both sequence and structural topology is only beginning to be uncovered. In general, G-quadruplexes having shorter loops tend to be the most stable. Melting analyses of T-loop quadruplexes suggested that folding of quadruplexes with shorter loop lengths was more energetically favorable than those with longer loops[7]. Additional detailed analyses of quadruplex-forming DNA oligonucleotide libraries, designed to average the effect of loop sequence identity on quadruplex stability, revealed that the influence of added loop length is most pronounced for motifs having a total loop length of <5 nts. Specifically, each nucleotide added results in a significant decrease in the thermal stability of the structure[8]. Motifs having >5 loop nucleotides do not experience the same decrease in stability as total loop length is increased[8]. These findings have likely contributed to the bias towards shorter looped motifs in the majority of quadruplex studies.

More recent work has explored G-quadruplexes with much longer loops, which has been a focal point of the work described in this thesis. An investigation of artificial 3-tetrad DNA quadruplex motifs having T-loops of varying lengths (1-30 nt) was consistent with previous results in the finding of an inverse correlation between T_m and loop length (2 °C decrease in T_m with each added base), specifically in K^+ solution[3]. However, despite having longer loops, many of the motifs tested were able to fold into G-quadruplexes having T_m 's above physiological temperature (37 °C), suggesting that similar, long-looped motifs may have more biological relevance than previously thought[3]. Beyond artificially designed sequences, many long-looped quadruplex motifs have been discovered in biological systems. An interesting example is the tandem quadruplex found within a G-tract rich region of the hTERT core promoter[10]. *Taq* polymerase stop assays and DMS (dimethyl sulfate) footprinting revealed an unusual configuration of two stacked quadruplexes, with the second quadruplex having a 26 nt central loop that forms its own hairpin[10]. More recent work identified nearly 1500 potential G-quadruplex motifs containing first and third loops each having 1 nt and very long central loops (up to 90 bases) in the 5'-UTRs of human mRNAs[11]. Quadruplex formation was characterized for many of these sequences with maximum central loop length up to 70 nts, and several were able to modulate luciferase gene expression *in cellulo*[11]. Additionally, G-quadruplex motifs can fold into multiple conformations, especially in regions of DNA/RNA that have more than four G-tracts clustered close enough together in the primary sequence. Studies of human telomeric DNA (repeating units of

TTAGGG) using optical tweezers identified a complex, heterogeneous population of structures, including some that had one or more G-rich repeats within the loops[12]. In combination, these findings suggest the potential biological relevance of long-looped quadruplexes in the regulation of gene expression and genome maintenance, as well as confirming that the canonical 7 nt maximal loop length limit is far too conservative.

In addition to increasing loop length, G-quadruplexes can also vary in the number of loops. Specifically, a class of 3-tetrad motifs including G-tracts interrupted by one or more T nucleotides, termed bulges, has been shown to fold into stable G-quadruplexes *in vitro*[13]. Further, bulges can be located at many positions within the quadruplex, provided that the quadruplex does not contain more than two bulges[13]. Consistent with canonical loops, the stability of the structure is inversely proportional to bulge length and is dependent on bulge location, sequence and quadruplex topology[13].

Now that the G-quadruplex motif definition has been amended to include long loops and bulges, the number of potentially biological G-quadruplexes has drastically increased. The most recent high-throughput sequencing analysis of G-quadruplexes within the human genome, accounting for long-looped and bulged motifs, has identified more than 700,000 potential quadruplex motifs, more than half of which were previously unrecognized[14]. Of these motifs, sequences including long loops and bulges each accounted for approximately 20% of the identified quadruplexes[14]. The monumental task ahead will be to

characterize these motifs, identifying which ones fold into stable G-quadruplexes *in vivo* and what functions they perform for the cell.

The continuous identification of new features within stable G-quadruplexes expands quadruplex motif definition and enhances the complexity of possible structures. These findings also highlight the need for more complex computational (and structure prediction) tools that are able to predict all types of possible quadruplex motifs. Further, such features could serve to differentiate individual quadruplexes from each other in biological systems by acting as recognition elements for intracellular factors. The studies described here do not take into account the variety of intracellular factors and interacting partners that can influence quadruplex stability *in vivo*, which is important to consider as part of detailed analyses of individual biological motifs. More complex quadruplexes could also be ideal targets for selective synthetic regulation of quadruplex function through the development of molecules that can interact with specific long loops or bulges, in addition to the tetrad core of the quadruplex.

1.1.3 G-quadruplexes in DNA

Many G-rich motifs found throughout the genome have the potential to fold into G-quadruplex structures under conditions that result in unwinding of the double helix. Although DNA G-quadruplexes have been implicated in a variety of biological processes, the precise locations and functional roles of these G-quadruplexes are continuing to be uncovered.

The landmark breakthrough resulting in quantitative visualization of DNA G-quadruplexes on a genome-wide scale was executed by the Balasubramanian

group in 2013 through their engineering of a DNA G-quadruplex specific antibody, BG4[15]. In cultured human cells, BG4 localized not only to the telomeres, regions of the DNA having well-characterized quadruplex-forming capacity, but also dispersed throughout the chromosomes[15]. Specifically, only 25% of BG4 foci were localized to telomeres, with the other 75% found in other regions of the DNA[15]. Nuclease assays confirmed the specificity of the antibody for G-quadruplex DNA, as BG4 localization was sensitive to DNase treatment, but not RNase treatment[15]. Further, the number of BG4 foci was dependent on the cell cycle. While BG4 foci increased approximately 2.5-fold just prior to the start of DNA replication (at the G1/S checkpoint), blocking cellular DNA synthesis with aphidicolin resulted in a 2-fold reduction in BG4 foci[15]. Replication leads to more of the DNA being unwound from the double helix, presumably allowing for the formation of additional G-quadruplexes, thus increasing the number of BG4 binding sites. A caveat of this study is that the BG4 binding may artificially induce the formation of stable quadruplexes that did not exist prior to the introduction of the antibody. Also, the specificity of the antibody is likely to be limited to certain classes of quadruplexes and, therefore, it may not recognize every possible DNA G-quadruplex. However, this work stands as the most compelling evidence for genome-wide G-quadruplex formation within human cells.

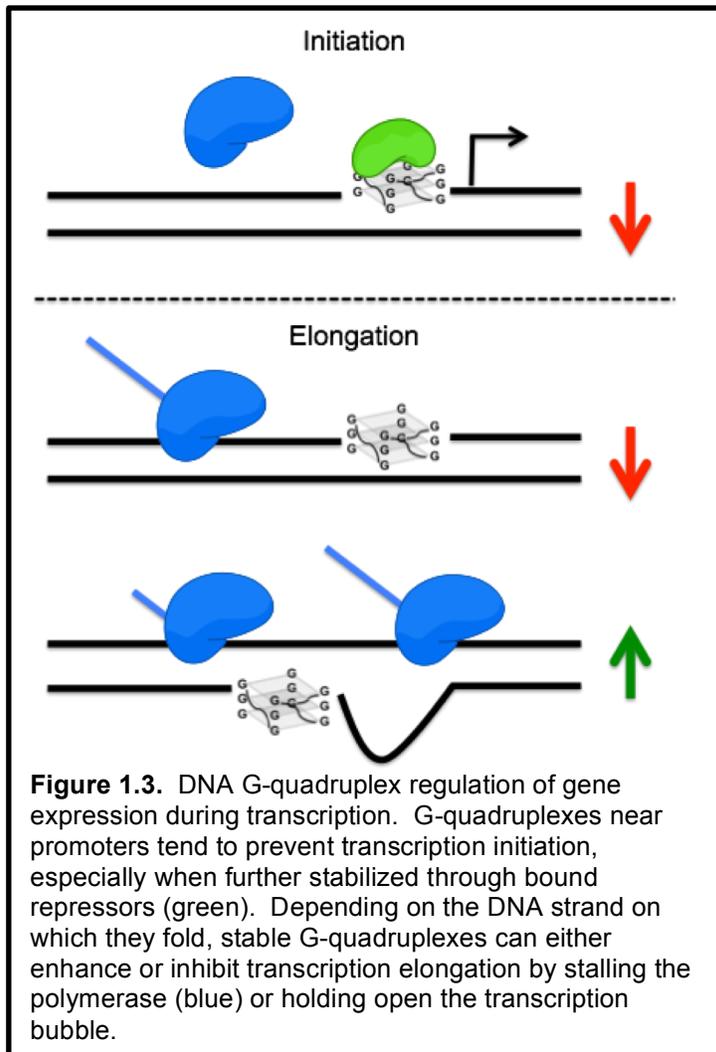
Although the BG4 antibody is a powerful tool for the recognition of putative DNA G-quadruplexes, computational and sequencing-based methods can provide more information about the specific location of G-quadruplex motifs.

Recently, the Balsubramanian group developed G4-Seq, a technique combining polymerase stop assays and Illumina next-generation sequencing to identify stable G-quadruplex structures within human DNA[14]. Using this method, sequencing data from human B lymphocytes were compared from conditions that either enhanced or reduced G-quadruplex stability[14]. Detection of scrambled sequences resulting from polymerase pausing in quadruplex-stabilizing conditions was used to identify G-quadruplex sites[14]. The identified G-quadruplexes were not randomly dispersed throughout the genome, but rather were localized to specific regions of the DNA. Sites with high G-quadruplex density included predicted 5'-UTRs and splice sites of pre-mRNAs or mRNAs, and genes associated with cancer development, including oncogenes and tumor suppressors[14]. Of particular interest were genes that had previously not been known to contain G-quadruplexes, including *BRCA2*[14]. This work provides strong support for the biological relevance of DNA G-quadruplexes, as well as identification of a series of potential new targets for cancer intervention. Apart from the identification of DNA G-quadruplexes, another major challenge in quadruplex biology is the classification of the roles of specific G-quadruplexes, including interactions with intracellular factors that modulate stability and propagate function. Although there is much more work to be done, a number of DNA G-quadruplexes have been already identified to have a role in a variety of biological processes.

Eukaryotic telomeres, the protective ends of linear chromosomes, are especially predisposed to G-quadruplex folding due to the asymmetric

distribution of G-rich sequences between DNA strands[16]. Specifically, vertebrate telomeres are composed of repeating units of 5'-TTAGGG-3', which has been a widely studied motif. The most conclusive experimental evidence for G-quadruplexes in the telomeres of eukaryotic cells resulted from the development of a scFv (single-chain antibody fragment) against the ciliate telomeric repeating unit[17]. *Stylonychia lemnae* telomeres are capped with a 3' overhang of four repeating units of T₄G₄, which can fold into either parallel or antiparallel quadruplexes *in vitro*[17]. These organisms have unusual DNA storage mechanisms, with a high density of minichromosome DNA structures found within macronuclei, making for an ideal system to visualize telomeric G-quadruplexes[17]. The DNA in these macronuclei is highly organized and interconnected into long fibrous structures, which may be linked partly by interchromosomal G-quadruplex formation[17]. Although the existence of telomeric G-quadruplexes is now widely accepted, the roles of these structures are still being uncovered. Putative functions of telomeric G-quadruplexes include protection of the chromosome ends from nucleases, binding sites for scaffolding proteins, and binding sites for ligands involved in telomerase regulation[16].

G-quadruplexes also have the potential to regulate gene expression at the transcriptional level (Figure 1.3). During elongation, a quadruplex folded on the



template strand may serve to inhibit transcription, while a quadruplex on the non-template strand may prevent re-annealing of the two strands, enhancing the transcriptional efficiency of the affected gene[16]. Additionally, DNA G-quadruplexes close to promoters have the ability to regulate transcription initiation. G-quadruplex motifs tend to be overrepresented in the

promoters of oncogenes, and such quadruplexes have been characterized in genes involved in the six hallmarks of cancer, highlighting the clinical significance of these structures[18]. Perhaps the most notable example is the quadruplex-forming sequence found in the nuclease hypersensitive element III₁ of the *c-MYC* promoter, which is responsible for >80% of transcriptional control of this oncogene[19]. In this case, the *c-MYC* quadruplex represses transcription

following further stabilization by interaction with the quadruplex binding protein and transcription factor, nucleolin[18]. Transcription of this gene can also be synthetically repressed through interaction with quadruplex binding small molecules, further suggesting the role of this G-quadruplex as a transcriptional regulatory element[19].

Apart from transcription, DNA G-quadruplexes have also been implicated in other biological processes, including DNA recombination. A 16 nt quadruplex-forming motif was identified as a specialized recombination initiation sequence within the antigenically variable pilin locus in *Neisseria gonorrhoeae*[20]. Mutagenesis of this site or introduction of the quadruplex-binding drug, NMM (N-methyl mesoporphyrin ix), inhibited pilin antigenic variation, a feature of this pathogen that has obstructed development of a vaccine against it[20]. Under certain conditions DNA G-quadruplexes can also be deleterious, requiring their destabilization by intracellular factors. Loss of the *Caenorhabditis elegans* DEXH-box helicase, for example, leads to genomic instability as it is needed to resolve stable structures in G-rich DNA during lagging strand synthesis in DNA replication[21]. Additionally, DNA G-quadruplexes may be involved in the origin of DNA replication, meiosis, and epigenetic regulation[16]. G-quadruplexes can also act as regulatory elements in viruses. The long terminal repeat (LTR) promoter of the HIV-1 provirus, for example, contains G-quadruplex motifs that, when folded, repress promoter activity and produce antiviral effects[22]. The growing number of putative quadruplex sequences throughout the DNA of many organisms will likely add to this list long into the future.

1.1.4 G-quadruplexes in RNA

Functional G-quadruplexes are not restricted to the genome, and similar structures are able to fold from the motifs that are transcribed into RNA. RNA G-quadruplexes comprise another class of regulatory elements involved in co- and post-transcriptional gene expression. Additionally, the discovery of RNA G-quadruplex regulation of pathogen proliferation offers new opportunities for targeted intervention of infectious disease. Coupled with the broad range of cellular RNA functions, RNA G-quadruplexes can be considered paramount to the regulation of gene expression throughout life.

A distinguishing property of RNA G-quadruplexes is superior stability when compared to their DNA counterparts. A comparative study of known DNA quadruplex-forming oligomers and their RNA analogs by CD (circular dichroism) spectroscopy and CD thermal melting established that the RNA oligomers formed more stable structures, always having parallel folding topology[23]. This enhanced stability is due to the presence of C2' hydroxyl groups, which allow for more intramolecular H-bonds, reducing the entropic cost of folding by decreasing the number of bound water molecules and augmenting the enthalpic contribution to the free energy of folding[24]. Additionally, RNA G-quadruplexes experience less structural constraints than motifs within dsDNA, due to the lack of competitive interactions with a complementary, duplex-forming strand[24]. The identification of RNA G-quadruplex stability substantiated the role of these structures in RNA biology.

Biological applications of RNA G-quadruplexes include functions in many cellular processes such as pre-mRNA splicing, pre-mRNA polyadenylation, transcriptional termination, mRNA translation, mRNA targeting, and telomere homeostasis within telomeric repeat-containing RNAs (TERRAs; reviewed in Millevoi *et al.*, 2012)[25]. 5'-UTR G-quadruplexes, in particular, have been explored as regulators of translation initiation. Specifically, quadruplexes within the 5'-UTR generally inhibit cap-dependent translation by interfering with 5' cap binding to translation initiation factors, or preventing proper scanning of the small ribosomal subunit[24, 25]. However, quadruplex folding may augment the function of internal ribosome entry sites, enhancing cap-independent translation[24]. The identification of long-looped 5'-UTR G-quadruplex motifs has expanded the number of potential regulatory elements, which will require further studies to pinpoint their specific functions[11]. RNA G-quadruplexes have also been identified in viral mRNAs. For example, a cluster of G-quadruplexes within the open reading frame of the Epstein-Barr virus-encoded nuclear antigen 1 (EBNA1) have been found to restrict self-translation, maintaining a minimal concentration of viral antigens and preventing detection by the host immune system[26].

A relatively new aspect of quadruplex-mediated regulation of transcription is the formation of DNA/RNA heteroquadruplexes during elongation[27]. Specifically, as transcription proceeds, the nascent RNA binds to the complementary DNA template forming an R-loop, which is subsequently melted as the next polymerase elongates along the same template. G-rich ssRNA is

then able to interact with the G-rich, homologous DNA in the non-template strand, forming a DNA/RNA heteroquadruplex[27]. This model was created following biochemical analyses of putative quadruplex motifs in the conserved sequence block II of human mitochondrial DNA[27]. Based on electrophoretic separation of RNase cleavage products of annealed, dye-labeled oligomers, heteroquadruplex formation was found to lag behind R-loop formation, and could be inhibited by the presence of an additional complementary DNA oligonucleotide[27]. Although DNA/RNA hybrids are just beginning to be explored, the prevalence of these structures is expected to be quite high due to their need for only two consecutive G-tracts within the DNA, as opposed to four G-tracts required for intramolecular structures.

Collectively, the pervasiveness of G-quadruplexes throughout the genomes and transcriptomes of organisms highlights the need for further study to identify the biological/clinical roles of these structures and develop selective and efficacious methods for synthetic regulation of their function.

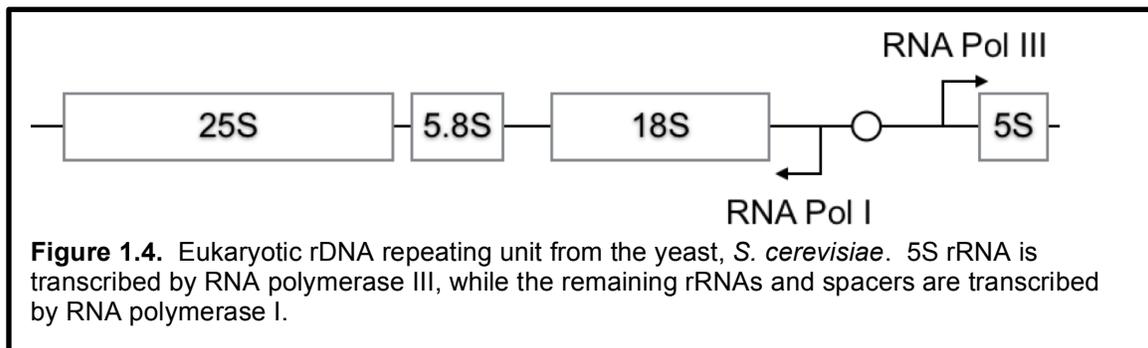
1.1.5 G-quadruplexes in Eukaryotic Ribosomal Genes

Eukaryotic ribosomal DNA (rDNA) is an area of the genome predicted to have a particularly high capacity for G-quadruplex formation[28]. 'rDNA' refers to the genes from which the ribosomal RNA (rRNA) is transcribed and subsequently processed, folded, and assembled to become the structural/functional core of the ribosome. Ribosomes are responsible for decoding mRNA and catalyzing polymerization of amino acids into proteins, which become a significant component of the cell's infrastructural and functional units. Control of the

biosynthesis of ribosomes is important because it is directly proportional to protein synthesis and growth rates[29]. Therefore, G-quadruplexes within these genes may have a significant impact on the regulation of cell growth, and may be especially important under conditions resulting in aberrant proliferation, such as cancer. Despite the potential biological significance, the role of rDNA G-quadruplexes has not yet been thoroughly studied. Investigation of G-quadruplexes in the yeast, *Saccharomyces cerevisiae*, rDNA and rRNA has been a major focus of the work described in this thesis.

Although the specific function(s) of eukaryotic rDNA G-quadruplexes are not yet known, the modest body of literature relating to this topic suggests that they do play an important role, and may be good targets for intervention in diseases, including cancer. The potential significance of rDNA G-quadruplexes is not only related to the importance of the genes in which they are located, but also the unusual structure of this part of the genome. Specifically, the rDNA locus is composed of a series of repeating units containing genes for the 5S rRNA and a polycistronic unit that templates the remaining rRNAs and transcribed spacer elements (Figure 1.4; reviewed in Woolford and Baserga, 2013)[30]. While the 5S rRNA is transcribed by RNA polymerase III, the remaining rRNAs are transcribed by RNA polymerase I as a single unit, which is subsequently processed to form mature rRNAs[30]. Adding to the complexity is the sheer number of repeats within the chromosome. Yeast, for example, carry approximately 150-200 rDNA repeats on chromosome XII, while humans have >400 copies across several chromosomes[30, 31]. Despite having so many

rDNA copies, only a subset of these are expressed at any one time, as having high concentrations of heavily transcribed genes would be toxic to cells [32]. Eukaryotes, specifically, have evolved to retain these extra copies of untranscribed rDNA. Yeast lacking sufficient rDNA repeats are more sensitive to mutagens, as these gene copies are needed for efficient replication-coupled recombinatorial repair[32]. However, the amount of rDNA gene copies results in cells having the monumental task of organizing proper replication, recombination, and transcription. G-quadruplexes may be involved in one or more of these processes.



rDNA is especially G-rich across the entire repeat, and many of these guanines are housed within tracts of at least 3 G's, which create potential quadruplex motifs[28]. Further, these G-tracts are predominantly restricted to the non-template strand, and so are also present in the rRNA, adding another level of complexity to potential G-quadruplex function. In humans and other metazoans, a clue to the role of rDNA G-quadruplexes lies in their interaction with the nucleolar protein, nucleolin. Two separate domains of nucleolin are able to interact with G-quadruplex DNA with very high affinity ($K_D = 1$ nM), specifically, the RNA binding domains and the C-terminal Arg-Gly-Gly repeats[28]. This

binding affinity coupled with the localization of nucleolin to the region of the nucleolus where rDNA is actively transcribed, suggests that this protein may effect quadruplex-mediated regulation of replication, recombination, and/or transcription of rDNA[28]. The role of nucleolin in RNA polymerase I-mediated transcription has been established. Chromatin immunoprecipitation assays demonstrated localization of nucleolin to the rDNA genes transcribed by RNA polymerase I, but not to genes transcribed by RNA polymerase II or RNA polymerase III[33]. Also, RNA interference knockdown of nucleolin resulted in inhibition of transcription by RNA polymerase I[33].

The connection between rDNA G-quadruplexes and nucleolin has also been implicated in cancer cell biology. Specifically, a fluoroquinolone-derived small molecule, CX-3543 (Quarfloxin), was found to disrupt nucleolin/rDNA G-quadruplex interactions, preventing RNA polymerase I transcription and causing apoptosis in cancer cells[31]. This drug eventually reached phase II clinical trials, and was the first quadruplex-targeting agent to enter human clinical trials[31]. Although this evidence supports the role of G-quadruplexes in rDNA transcription, it does not necessarily preclude their role in other processes such as recombination and replication. Apart from CX-3543, many other classes of small molecules, including natural alkaloids, antibiotics kinase inhibitors, DNA intercalators, and metal conjugates, have been explored as anticancer agents through their abilities to interact with or modulate the function of the nucleolus (reviewed in Pickard and Bierbach, 2013)[34]. These molecules can be more selective than other chemotherapeutics by taking advantage of the structural and

functional properties unique to the nucleoli of cancer cells[34]. Conjugate molecules able to both specifically target rDNA quadruplexes and modulate nucleolar function may be useful in future chemotherapeutic strategies.

Similarly, yeast have putative G-quadruplex motifs in the rDNA, (and rRNA) as well as within the rest of the genome[35, 36]. Computational searches for 3-tetrad quadruplex motifs using a 100 nt maximum sequence length, allowing for the detection of long-looped motifs, identified 21 potential quadruplex sequences all on the non-template strand of the rDNA[35]. When compared to a genome-wide search for quadruplex motifs, the rDNA displayed a 10-fold higher density for these sequences, with the exception of telomeric quadruplex-forming sequences[35]. Upon treatment of yeast cells with the quadruplex-binding drug, NMM, genes related to nucleolar function, including those involved in rRNA processing and ribosome assembly, were dramatically downregulated, suggesting the stabilization of G-quadruplexes at these sites[35]. Together, these results suggest that rDNA/rRNA G-quadruplexes may also have important functions in yeast, although these roles may differ slightly from humans, as yeast do not contain a nucleolin homolog. However, the availability of genetic tools that allow manipulation of rDNA sequence in yeast makes this an attractive system in which to approach this problem[37].

The body of literature regarding natural DNA and RNA G-quadruplexes, including those within ribosomal genes, continues to grow and support both their prevalence and biological function. The major challenge ahead will be to

selectively target and modulate the function of these structures to expand our understanding of G-quadruplex biology in both gene expression and disease.

1.2 Artificial Modulation of G-quadruplex function

G-quadruplexes function as unusually structured *cis*-acting elements in the genomes and transcriptomes of biological systems. Artificial manipulation of these functions could lead to the development of better tools to control gene expression and study quadruplex biology. Broadly, this aim can be achieved through the design of probes that specifically target either the sequence or unique structural features of G-quadruplexes.

1.2.1 Small Molecules

Small molecules have been frequently used as G-quadruplex targeting agents, even in some cases demonstrating therapeutic potential[31]. Molecules that selectively bind G-quadruplexes generally interact with their targets via shape recognition, wherein the surface area of the molecule is similar to that of the G-quartet. A well-characterized class of such molecules includes porphyrin dyes, such as TMPyP4 and NMM, which also fluoresce upon binding to quadruplex DNA and allow detection of these structures apart from other secondary structures[38]. A report, now several years old, cataloged >800 known quadruplex ligands, which suggests the potential utility of these molecules in technological or therapeutic applications[39]. Stabilization of endogenous G-quadruplexes *in vivo* by small molecules was verified using a DNA G-quadruplex-specific antibody, BG4, in human cells that had been treated with the quadruplex-stabilizing drug, pyridostatin[15]. Specifically, treated cells were found to have a

2.9-fold increase in BG4 foci when compared to untreated cells[15]. However, the major limitation with shape recognition-based strategies is the inability to selectively interact with one quadruplex structure over another. Therefore, an alternative strategy is the use of natural or artificial nucleic acid oligomer-based probes that can hybridize to G-quadruplex motifs using sequence selectivity.

1.2.2 Antisense Oligonucleotides

The antisense strategy takes advantage of the barcode-like nature of nucleic acids, relying on sequence specificity to minimize off-target effects. In the case of G-quadruplexes, the C-rich antisense oligonucleotide will typically act to unravel the quadruplex, resulting in the formation of a Watson-Crick base-paired intermolecular duplex at that site[40, 41]. The potential utility of G-quadruplex-targeting antisense oligonucleotides as an alternative therapeutic strategy for the treatment of latent viral infections has recently been explored. Stabilization of EBNA1 G-quadruplexes by the small molecule, pyridostatin, reduced *in vitro* and *in vivo* translation of EBNA1 mRNA and limited recognition of infected cells by T-cells[26]. However, pyridostatin, like other small molecules quadruplex ligands, cannot be specifically targeted to these viral quadruplexes, limiting complete understanding of its *in vivo* mechanism. In contrast, destabilization of these G-quadruplexes through hybridization with duplex-forming antisense oligonucleotides enhanced translation of the mRNA, limiting antigen presentation on virally infected cells[26]. However, the G-quadruplex-forming sequence (5'-GGGGCAGGAGCAGGAGGA-3') investigated in this study has a relatively moderate thermal stability *in vitro* ($T_m = 54.1 \pm 1.1$ °C in 100 mM KCl), which likely

made destabilization by the antisense probes possible[26]. Further, potential G-quadruplex-forming sequences have been identified in the open reading frames of additional gammaherpesviral maintenance proteins[26]. This finding suggests that G-quadruplex-mediated changes to structural population dynamics of viral mRNAs may be a more prevalent translational regulatory mechanism than previously realized.

The antisense strategy is not limited to use with only intramolecular G-quadruplexes. Recently, transcription-induced DNA/RNA heteroquadruplex formation was inhibited by the addition of a C-rich DNA oligomer designed to hybridize with the G-rich portion (CSB II of the human mtDNA; 5'-GGGGGAGGGGG-3') of the nascent transcript[27]. This antisense probe not only served to inhibit the heteroquadruplex, but also produced a slight negative effect on R-loop formation, likely due to increased competition from re-annealing of the template and non-template DNA strands[27]. The existence of similar potential heteroquadruplex-forming sequences in greater than 97% of human protein encoding genes hints at the magnitude of untapped potential for the development of novel methods of transcriptional regulation[42]. The use of similar probes will likely aid in the continued revelation of the biological significance of these structures. Additionally, these sites must now also be considered as potential unintended recipients of probes designed to target other intracellular G-quadruplexes.

Antisense oligonucleotides may also be used to promote G-quadruplex formation by sequestering nucleotides that would otherwise base pair with the

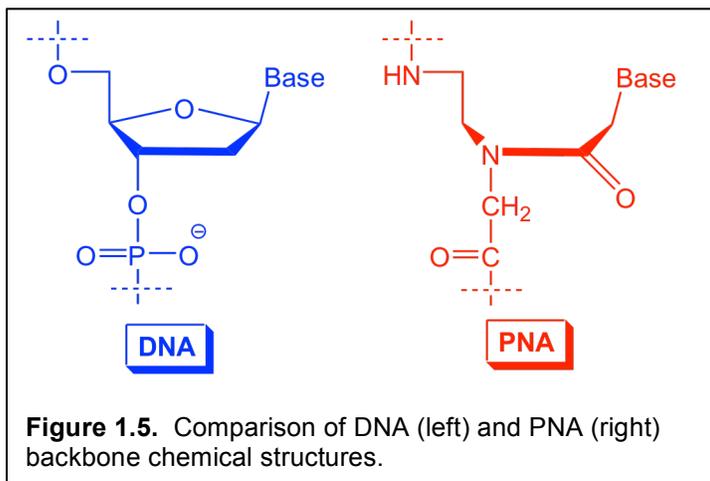
quadruplex motif. This strategy, in addition to unfolding of the quadruplex, has been employed to regulate translation in artificial and endogenous mRNAs[43]. In this work, specificity was maximized by targeting G-quadruplexes containing particularly long central loops, in addition to the adjacent G-tracts, including one found within the 5' UTR of the H2AFY (human histone variant; 5'-GGGCGGGAAGCGAAGAGGCGGG-3') gene[43]. In this case, unfolding of the G-quadruplex led to a modest increase in translation, while stabilization of the G-quadruplex produced the opposite effect[43]. However, this strategy cannot be applied universally. Similar experiments with another G-quadruplex, (the 5' UTR of Akirin2 mRNA) did not result in increased translation, due to the inability of the antisense oligonucleotide to invade that structure[43]. Nevertheless, this demonstration of switchable regulation of translation provides new insights into the biological significance and experimental utility of targeting 5'UTR mRNA G-quadruplexes. However, the limited efficacy and versatility of antisense oligonucleotides in this context alludes to the need for the development of even better probes.

Together, these studies cement G-quadruplex-mediated antisense regulation of translation as an effective experimental, and potentially therapeutic tool. The variety of intracellular G-quadruplex motifs is vast and growing. Therefore, special care must be taken to select an appropriate probe design depending on the G-quadruplex of interest. In the case of very stable structures, synthetic nucleic acid mimics may be more useful for the manipulation of G-quadruplex structures.

1.2.3 Complementary Peptide Nucleic Acids

As an alternative to DNA and RNA-based probes, synthetic nucleic acid-mimicking oligomers have also been widely explored as G-quadruplex targeting probes. Peptide nucleic

acids (PNAs), are nucleic acid oligomers having the nucleobases of DNA/RNA conjugated to an uncharged, amide backbone (Figure 1.5)[44]. Due to the lack of



backbone charge repulsion, PNAs form very stable complexes with nucleic acid targets[44, 45]. This ability to achieve high affinity hybridization with nucleic acids, in combination with chemical modifications, have allowed PNAs to be utilized in a variety of applications, including fluorescent labeling, anti-gene technology, and nucleic acid purification (reviewed in Achim *et al.*, 2008)[46].

An obvious concern when targeting G-quadruplex nucleic acids is that the probe will be unable to overcome thermodynamic and kinetic barriers required for the structural conversion to an intermolecular duplex. Synthetic nucleic acid mimicking oligomers, such as PNA and LNA (locked nucleic acid), are being established as valid alternatives for targeting G-quadruplexes. LNAs are RNA nucleotides modified via connection of the 2'O to the 4'C[47]. Invasion of the *c-MYC* promoter G-quadruplex using a complementary LNA modified probe resulted in duplex formation at that site and reduced promoter activity[48]. PNAs

offer numerous additional advantages when used as probes for nucleic acid hybridization. Covalent modification of PNAs can result in improved target recognition and expanded probe functionality. For example, incorporation of *trans*-cyclopentane residues within PNA oligomers has been shown to improve PNA/DNA complex stability in duplex, triplex, and quadruplex conformations[49]. More recently, the introduction of a hydrophilic R-group at the γ carbon of the PNA backbone results in the oligomer having a pre-organized helical conformation, ideal for duplex formation[50]. Variations of these modifications are continuing to enhance the applicability of PNA based probes.

In 2001, the Armitage group first demonstrated G-quadruplex invasion by a short, complementary PNA probe at physiological temperature, which could not be accomplished using a DNA oligonucleotide of the same sequence[40]. The quadruplex-forming sequence targeted in these experiments was the thrombin binding aptamer (TBA; 5'-GGTTGGTGTGGTTGG-3'), an antiparallel, two-tetrad, DNA G-quadruplex. Several important insights were gained into DNA/PNA hybridization by this work. In addition to confirming the ability of PNA to invade DNA G-quadruplex structures, the dependence of this interaction on ionic strength was also identified. Specifically, at low ionic strength (10 mM KCl), the thermal stability of the PNA/TBA heteroduplex was higher than that of the G-quadruplex (55.2 and 46.7 °C, respectively)[40]. However, in 250 mM KCl, the T_m of the G-quadruplex increased to 55.5 °C, allowing TBA to re-fold into a quadruplex upon thermal denaturation of the heteroduplex[40]. This result, by taking into account the effect of salt concentration on target structure, contrasts

with the previous finding that PNA/DNA hybridization is indifferent to physiologically-relevant variations in ionic strength[51]. Additionally, overhanging DNA bases beyond those involved in duplex formation further promoted stability of the complex, an effect that was not seen with a DNA/TBA duplex, suggesting additional benefits of targeting G-quadruplexes found within the context of a longer, functioning nucleic acid unit[40].

The invasion of complementary PNAs is not limited to two-tetrad DNA G-quadruplexes. Subsequent reports have produced additional insights into the hybridization of these complexes. FRET experiments with the human telomeric repeat DNA (5'-GGGTTAGGGTTAGGGTTAGGG-3') found that opening of the quadruplex was independent of PNA concentration, suggesting that the rate-limiting step of the interaction to be a slow, internal rearrangement of the quadruplex structure, followed by fast hybridization of the PNA[41]. This mechanism of hybridization can be explained by a model where the quadruplex structure exists in equilibrium with a partially folded intermediate, which can hybridize to the complementary strand[41]. Additionally, complementary PNA invasion to both RDQ, a G-quadruplex-containing RNA aptamer of the Fragile X mental retardation protein and RQ, the truncated quadruplex-forming region of this sequence, has been reported[52, 53]. Fluorescent Job plot results, in addition to UV melting experiments, confirmed the formation of the expected stable 1PNA:1RNA duplexes[52, 53]. Additionally, CD data provided evidence of parallel topology for both RDQ and RQ, expanding the variety of G-quadruplex structures able to be invaded by complementary PNA[53]. Similar parallel

character was later found to be a common feature of RNA G-quadruplexes[23]. However, an important caveat to bear in mind is that the structure of each individual G-quadruplex target may dictate the potency of the chosen recognition mode. More recently, our group compared the hybridization of complementary and homologous (G-rich; to be discussed in the next section) PNAs targeted to both a parallel (Myc19; 5'-AGGGTGGGGAGGGTGGGGA-3') and a hybrid/mixed (hTelo22; 5'-AGGGTTAGGGTTAGGGTTAGGG-3') DNA G-quadruplex[54]. All PNAs were designed to interact with the central two G-tracts and loop of each DNA. Surprisingly, the complementary PNA, even after thermal annealing, was unable to hybridize to Myc19 under physiologically relevant salt concentrations (100 mM KCl)[54]. However, this type of interaction is dependent on both the G-quadruplex in question and the conditions of the experiment. A separate report confirmed duplex formation between a four-tetrad DNA G-quadruplex [(5'-TGGGGT-3')₄] and its hexamer PNA complement only after thermal annealing in ammonium acetate[55]. This work reinforces the need to carefully design and optimize probes for each specific target of interest.

In addition to being able to invade and hybridize with potential target sequences, PNAs must also be compatible with biological systems. When the goal is G-quadruplex-mediated regulation of transcription or translation, for example, an obvious concern is that the PNA will only serve to replace one stable structure with another, preventing proper function of the polymerase or ribosome. In cases where unfolding of the G-quadruplex is intended to upregulate gene/protein expression, inhibition of the relevant biological

machinery may not be the desired result[26]. However, recent work has shown that under certain circumstances, DNA/PNA duplexes may be less inhibitory than G-quadruplexes[56]. Specifically, the presence of a complementary PNA resulted in less inhibition of DNA polymerase η than a G-rich, heteroquadruplex-forming PNA or small molecule-stabilized G-quadruplex. Polymerase η is a *trans*-lesion synthesis polymerase, previously found to be able to read through unusual structural motifs[57]. Polymerase extension assays confirmed a moderate increase of polymerase η activity in the presence of a complementary PNA on a template containing a well-characterized G-quadruplex-forming sequence (NHE-III₁ of the *c-MYC* promoter)[56]. Interestingly, the reverse experiment, hybridization of a G-rich PNA to a C-rich template, resulted in slight inhibition of the polymerase[56]. Although the implications of these results are complex, they indicate that G-quadruplex destabilization via interaction with a complementary PNA is sufficient to promote the activity of this DNA polymerase. Further investigation will be needed to confirm the specificity of these PNA/DNA interactions, as well as how such complexes will behave in other biological contexts. Nevertheless, this work expands the potential utility of complementary PNAs as G-quadruplex-destabilizing agents, which could, in principle, be further improved by modification of probe design to optimize affinity and specificity.

Alternatively, complementary PNAs have been utilized as G-quadruplex-stabilizing ligands. Such probes perform similarly to G-quadruplex-stabilizing small molecules, but with the added benefit of sequence-specificity. The DNA G-quadruplex targeted in these experiments, cKit87up (5'-

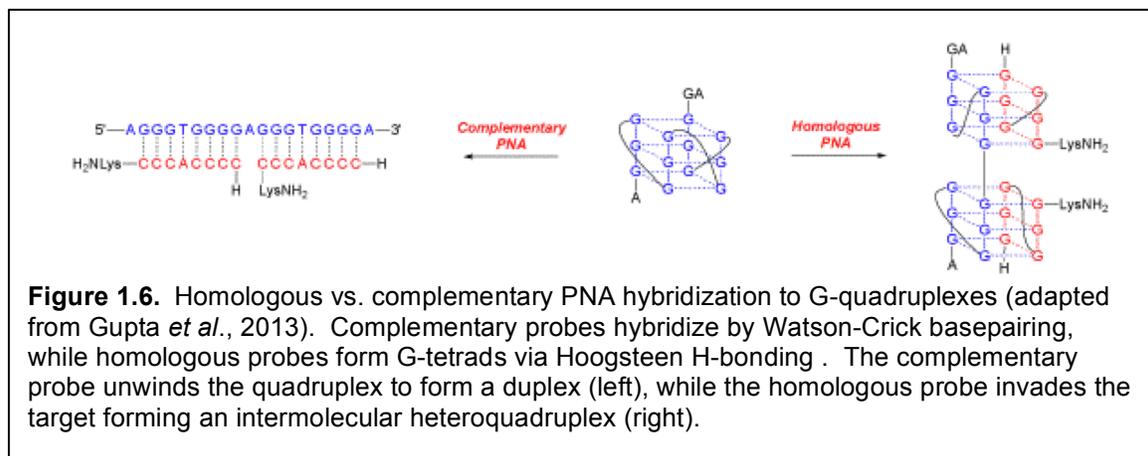
AGGGAGGGCGCTGGGAGGAGGG-3'), adopts an unusual structure with four loops [58]. A complementary PNA targeted to the fourth loop (A₁₆-G₂₀) acted as a quadruplex stabilizer in solutions containing K⁺ or NH₄⁺[59]. Additional complementary PNAs extending beyond the loop into the tetrad core behave similarly in K⁺ solution, but unfold the quadruplex to form a heteroduplex in NH₄⁺ solution[59]. In the latter case, duplex formation is favored due to the slight destabilizing effect of NH₄⁺ on the G-quadruplex compared to K⁺ solution. Stabilization of this G-quadruplex by probe hybridization to the long loop is reminiscent of another long loop-containing G-quadruplex-forming sequence. Similarly, the hTERT core promoter contains a quadruplex motif including a 26 base central loop that folds into its own hairpin within the G-quadruplex, greatly stabilizing the structure[10]. Therefore, in the case of cKit87up, the PNA likely functions to confer both structure to the long loop and stability to the overall G-quadruplex structure. This work produced a G-quadruplex targeting system capable of switching binding modes from G-quadruplex stabilizer to the more conventional G-quadruplex "opener", depending on both probe sequence and chemical environment. Of course, this specific strategy was made possible by the unique topology of cKit87up. As the definition of G-quadruplex motifs continues to expand, similar approaches to other targets become more feasible.

G-quadruplexes are typically considered as uniquely structured regions of nucleic acids. However, the use of complementary, sequence-specific probes has been demonstrated as a viable method of artificial recognition and functional regulation at these sites. Among synthetic nucleic acid mimics, PNAs have been

established as versatile probes able to de-/stabilize G-quadruplexes and up/down-regulate function depending on the biological and chemical environment. Collectively, these studies bring to light the important reality that a one-size-fits-all protocol for sequence-based G-quadruplex targeting does not exist. Probe design and experimental conditions must be optimized depending on the structure of the target, environment, and intended outcome of the interaction. However, our growing understanding of G-quadruplex biology and widening definition of the G-quadruplex motif will expand the application of this and similar strategies in the future.

1.2.4 Homologous Peptide Nucleic Acids

As an alternative to complementary probe hybridization, G-quadruplexes are unique in that they are amenable to interactions with homologous, G-rich probes

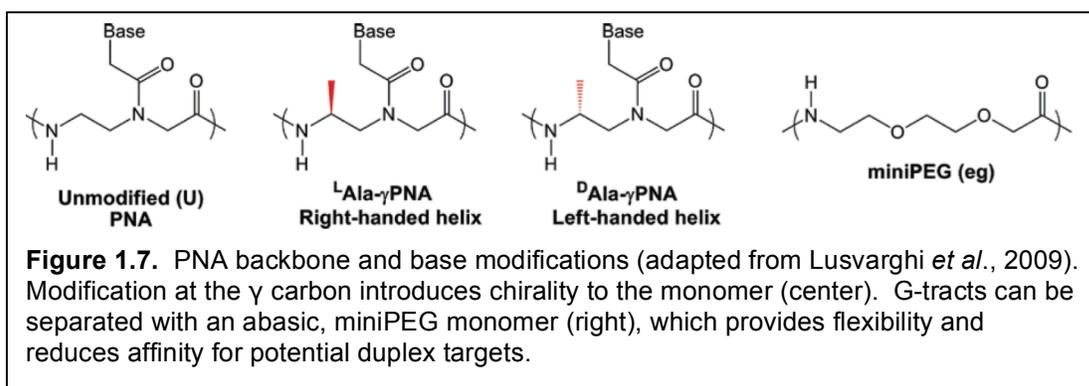


(Figure 1.6). The utility of homologous PNAs as G-quadruplex binders was first demonstrated in the Armitage group with the *Oxytrichia nova* telomeric sequence, $d(G_4T_4G_4)[60]$. With this method, the PNA interacts with its targets via intermolecular G-tetrad formation, creating a DNA/PNA heteroquadruplex. FRET (Förster resonance energy transfer) assays demonstrated that the N-terminus of

the PNA preferentially aligns with the 5' end of the DNA strand[60]. Subsequent studies demonstrated the compatibility of homologous PNAs with other DNA and RNA G-quadruplex targets. Detailed analyses of these interactions with an RNA quadruplex derived from the RDQ Fragile X protein aptamer revealed that while the complementary probe interacts with the target in a 1:1 fashion, two homologous PNA strands hybridize to the quadruplex forming a 2:1 PNA:DNA intermolecular heteroquadruplex[52, 53]. Similar findings were determined with the DNA quadruplex derived from the *MYC* gene promoter, (Myc19: 5'-AGGGTGGGGAGGGTGGGGA-3'), where a homologous 8-mer probe (P_{myc} : H₂NLys-GGGGAGGG-H) bound to the target with low nanomolar affinity ($K_d = 5$ nM), forming a 2:1 PNA:DNA complex[61]. Although this work transformed the utility of PNA as a quadruplex-targeting agent, it also presented the challenge of developing a mechanism to allow the PNA to distinguish between the G-quadruplex target and the complementary, opposing DNA strand.

The problem of targeting homologous PNAs specifically to G-quadruplexes has been approached through chemical modification of the probes intended to promote quadruplex over duplex formation. Modification at the γ position of the monomer was first utilized as a mechanism to incorporate fluorophores and did not negatively affect the thermal stability of the PNA/DNA complex[62]. γ -modifications also introduce chirality into the molecule, allowing the oligomer to be pre-organized into a right- or left-handed helix, depending on the stereochemistry of the modification[63]. Early iterations of this strategy incorporated ^LAla- or ^DAla- γ -modified oligomers into homologous probes (Figure

1.7). ^DAla- γ PNAs in combination with the inclusion of abasic, miniPEG (polyethylene glycol) linkers designed to destabilize duplex binding were found to enhance probe selectivity for quadruplex vs. duplex targets. More recently, the alanine-based γ substituent has been replaced with an (*R*)-diethylene glycol unit, which also serves to enhance the water solubility of the probe, making the PNA more suitable for biological applications[50].



An important consideration that must be addressed when designing homologous quadruplex targeting probes is the preferred mode of hybridization for the target of interest. Homologous PNAs, for example, are known to display kinetic discrimination among a pool of DNA quadruplex targets, with a tendency to hybridize most quickly to targets already having a parallel topology[64]. In the case of the antiparallel DNA quadruplex, hTelo22, SPR experiments indicated a preference for the complementary over the homologous PNA[54].

The homologous PNA strategy has been demonstrated as an effective method for use with a broad range of quadruplex targets. However, a significant challenge that still remains is the selective hybridization to specific quadruplex targets, with minimal off-target quadruplex interactions. Further advances in PNA chemistry to allow for combination of hybridization strategies (e.g. both

homologous and complementary) into a single probe will be required to help overcome this obstacle. Additionally, the development of reliable techniques to identify all PNA binding sites in a complex sample, such as a cell, will be needed to assay the efficacy and selectivity of new probes. An orthogonal approach is the consideration of quadruplex targets that are more amenable to selective probe hybridization. The growing library of motifs known to support stable G-quadruplex formation offers a large pool of yet untested targets, creating new opportunities for sequence-specific G-quadruplex regulation and identification of biological function.

1.3 Thesis Overview

Chapter 2, much of which was summarized in a recent publication[65], summarizes a novel mode of homologous PNA hybridization to noncanonical, long loop-containing G-quadruplexes, wherein the long loops allow >2 PNAs to hybridize to the target forming higher order heteroquadruplex structures. Chapter 3 is a continuation of previous studies in our group exploring the effects of simulated molecular crowding on the thermodynamics of homologous PNA hybridization to several DNA G-quadruplex targets. In this study, we also investigated the interaction of partially and fully γ -modified homologous PNAs to these quadruplexes. Chapter 4 is a summary of the evidence for G-quadruplex mediated regulation of ribosome biogenesis in the *Saccharomyces cerevisiae* model system.

1.4 References

1. Gellert, M., M.N. Lipsett, and D.R. Davies, *Helix formation by guanylic acid*. Proc Natl Acad Sci U S A, 1962. **48**: p. 2013-8.

2. Burge, S., et al., *Quadruplex DNA: sequence, topology and structure*. Nucleic Acids Res, 2006. **34**(19): p. 5402-15.
3. Guedin, A., et al., *How long is too long? Effects of loop size on G-quadruplex stability*. Nucleic Acids Res, 2010. **38**(21): p. 7858-68.
4. Hardin, C.C., et al., *Cation-Dependent Transition between the Quadruplex and Watson-Crick Hairpin Forms of d(CGCG3GCG*. Biochemistry, 1992. **31**: p. 833-841.
5. Miyoshi, D., H. Karimata, and N. Sugimoto, *Hydration regulates thermodynamics of G-quadruplex formation under molecular crowding conditions*. J. Am. Chem. Soc., 2006. **128**: p. 7957-7963.
6. Largy, E., et al., *Quadruplex Turncoats: Cation-Dependent Folding and Stability of Quadruplex-DNA Double Switches*. J. Am. Chem. Soc., 2016. **138**(8): p. 2780-92.
7. Hazel, P., et al., *Loop-length-dependent folding of G-quadruplexes*. J. Am. Chem. Soc., 2004. **126**: p. 16405-16415.
8. Bugaut, A. and S. Balasubramanian, *A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes*. Biochemistry, 2008. **47**: p. 689-697.
9. Tippana, R., W. Xiao, and S. Myong, *G-quadruplex conformation and dynamics are determined by loop length and sequence*. Nucleic Acids Res, 2014. **42**(12): p. 8106-14.
10. Palumbo, S.L., S.W. Ebbinghaus, and L.H. Hurley, *Formation of a unique end-to-end stacked pair of G-quadruplexes in the hTERT core promoter with implications for inhibition of telomerase by G-quadruplex-interactive ligands*. J. Am. Chem. Soc., 2009. **131**: p. 10878-10891.
11. Jodoin, R., et al., *The folding of 5'-UTR human G-quadruplexes possessing a long central loop*. RNA, 2014. **20**(7): p. 1129-41.
12. Koirala, D., et al., *Long-loop G-quadruplexes are misfolded population minorities with fast transition kinetics in human telomeric sequences*. J. Am. Chem. Soc., 2013. **135**(6): p. 2235-41.
13. Mukundan, V.T. and A.T. Phan, *Bulges in G-quadruplexes: broadening the definition of G-quadruplex-forming sequences*. J. Am. Chem. Soc., 2013. **135**(13): p. 5017-28.
14. Chambers, V.S., et al., *High-throughput sequencing of DNA G-quadruplex structures in the human genome*. Nat Biotechnol, 2015.
15. Biffi, G., et al., *Quantitative visualization of DNA G-quadruplex structures in human cells*. Nat Chem, 2013. **5**(3): p. 182-6.
16. Bochman, M.L., K. Paeschke, and V.A. Zakian, *DNA secondary structures: stability and function of G-quadruplex structures*. Nat Rev Genet, 2012. **13**(11): p. 770-80.
17. Schaffitzel, C., et al., *In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei*. Proc Natl Acad Sci U S A, 2001. **98**: p. 8572-8577.
18. Brooks, T.A., S. Kendrick, and L. Hurley, *Making sense of G-quadruplex and i-motif functions in oncogene promoters*. FEBS J, 2010. **277**(17): p. 3459-69.

19. Siddiqui-Jain, A., et al., *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): p. 11593-8.
20. Cahoon, L.A. and H.S. Seifert, *An alternative DNA structure is necessary for pilin antigenic variation in Neisseria gonorrhoeae*. Science, 2009. **325**(5941): p. 764-7.
21. Cheung, I., et al., *Disruption of dog-1 in Caenorhabditis elegans triggers deletions upstream of guanine-rich DNA*. Nat Genet, 2002. **31**(4): p. 405-9.
22. Perrone, R., et al., *A dynamic G-quadruplex region regulates the HIV-1 long terminal repeat promoter*. J Med Chem, 2013. **56**(16): p. 6521-30.
23. Joachimi, A., A. Benz, and J.S. Hartig, *A comparison of DNA and RNA quadruplex structures and stabilities*. Bioorg Med Chem, 2009. **17**(19): p. 6811-5.
24. Bugaut, A. and S. Balasubramanian, *5'-UTR RNA G-quadruplexes: translation regulation and targeting*. Nucleic Acids Res, 2012. **40**(11): p. 4727-41.
25. Millevoi, S., H. Moine, and S. Vagner, *G-quadruplexes in RNA biology*. Wiley Interdiscip Rev RNA, 2012. **3**(4): p. 495-507.
26. Murat, P., et al., *G-quadruplexes regulate Epstein-Barr virus-encoded nuclear antigen 1 mRNA translation*. Nat Chem Biol, 2014. **10**(5): p. 358-64.
27. Zhang, J.Y., et al., *Mechanism and manipulation of DNA:RNA hybrid G-quadruplex formation in transcription of G-rich DNA*. J. Am. Chem. Soc., 2014. **136**(4): p. 1381-90.
28. Hanakahi, L.A., H. Sun, and N. Maizels, *High affinity interactions of nucleolin with G-G-paired rDNA*. J. Biol. Chem., 1999. **274**(22): p. 15908-15912.
29. Kief, D.R. and J.R. Warner, *Hierarchy of elements regulating synthesis of ribosomal proteins in Saccharomyces cerevisiae*. Mol Cell Biol, 1981. **1**: p. 1016-1023.
30. Woolford, J.L., Jr. and S.J. Baserga, *Ribosome biogenesis in the yeast Saccharomyces cerevisiae*. Genetics, 2013. **195**(3): p. 643-81.
31. Drygin, D., et al., *Anticancer activity of CX-3543: a direct inhibitor of rRNA biogenesis*. Cancer Res, 2009. **69**(19): p. 7653-61.
32. Ide, S., et al., *Abundance of ribosomal RNA gene copies maintains genome integrity*. Science, 2010. **327**(5966): p. 693-6.
33. Rickards, B., et al., *Nucleolin is required for RNA polymerase I transcription in vivo*. Mol Cell Biol, 2007. **27**(3): p. 937-48.
34. Pickard, A.J. and U. Bierbach, *The cell's nucleolus: an emerging target for chemotherapeutic intervention*. ChemMedChem, 2013. **8**(9): p. 1441-9.
35. Hershman, S.G., et al., *Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in Saccharomyces cerevisiae*. Nucleic Acids Res, 2008. **36**(1): p. 144-56.

36. Capra, J.A., et al., *G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in Saccharomyces cerevisiae*. PLoS Comput Biol, 2010. **6**(7): p. e1000861.
37. Cole, S.E. and F.J. LaRiviere, *Analysis of nonfunctional ribosomal RNA decay in Saccharomyces cerevisiae*. Methods Enzymol, 2008. **449**.
38. Arthanari, H., et al., *Fluorescent dyes specific for quadruplex DNA*. Nucleic Acids Res, 1998. **26**.
39. Li, Q., et al., *G4LDB: a database for discovering and studying G-quadruplex ligands*. Nucleic Acids Res, 2013. **41**(Database issue): p. D1115-23.
40. Datta, B. and B.A. Armitage, *Hybridization of PNA to structured DNA targets: quadruplex invasion and the overhang effect*. J. Am. Chem. Soc., 2001. **123**: p. 9612-9619.
41. Green, J.J., et al., *Kinetics of unfolding the human telomeric DNA quadruplex using a PNA trap*. J. Am. Chem. Soc., 2003. **125**: p. 3763-3767.
42. Xiao, S., et al., *Bioinformatic analysis reveals an evolutionary selection for DNA:RNA hybrid G-quadruplex structures as putative transcription regulatory elements in warm-blooded animals*. Nucleic Acids Res, 2013. **41**(22): p. 10379-90.
43. Rouleau, S.G., et al., *Small antisense oligonucleotides against G-quadruplexes: specific mRNA translational switches*. Nucleic Acids Res, 2015. **43**(1): p. 595-606.
44. Nielsen, P.E., et al., *Sequence-Selective Recognition of DNA by Strand Displacement with a Thymine-Substituted Polyamide*. Science, 1991. **254**(5037): p. 1497-1500.
45. Egholm, M., et al., *PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules*. Nature, 1993. **365**: p. 566-568.
46. Achim, C., et al., *Peptide Nucleic Acids (PNAs)*. Wiley Encyclopedia of Chemical Biology, 2008.
47. Singh, S.K., et al., *LNA (locked nucleic acids): synthesis and high-affinity nucleic acid recognition*. Chem Commun, 1998(4): p. 455-456.
48. Kumar, N., et al., *Silencing c-MYC Expression by Targeting Quadruplex in P1 Promoter Using Locked Nucleic Acid Trap*. Biochemistry, 2008. **47**(50): p. 13179-13188.
49. Englund, E.A., et al., *PNA-DNA duplexes, triplexes, and quadruplexes are stabilized with trans-cyclopentane units*. J. Am. Chem. Soc., 2006. **128**: p. 16456-16457.
50. Sahu, B., et al., *Synthesis and characterization of conformationally preorganized, (R)-diethylene glycol-containing gamma-peptide nucleic acids with superior hybridization properties and water solubility*. J Org Chem, 2011. **76**(14): p. 5614-27.
51. Tomac, S., et al., *Ionic effects on the stability and conformation of peptide nucleic acid complexes*. J. Am. Chem. Soc., 1996. **118**: p. 5544-5552.

52. Marin, V.L. and B.A. Armitage, *RNA guanine quadruplex invasion by complementary and homologous PNA probes*. J. Am. Chem. Soc., 2005. **127**: p. 8032-8033.
53. Marin, V.L. and B.A. Armitage, *Hybridization of complementary and homologous peptide nucleic acid oligomers to a guanine quadruplex-forming RNA*. Biochemistry, 2006. **45**(6): p. 1745-1754.
54. Gupta, A., et al., *Strand invasion of DNA quadruplexes by PNA: comparison of homologous and complementary hybridization*. ChemBiochem, 2013. **14**(12): p. 1476-84.
55. Amato, J., et al., *Hybridization of short complementary PNAs to G-quadruplex forming oligonucleotides: An electrospray mass spectrometry study*. Biopolymers, 2009. **91**(4): p. 244-55.
56. Murphy, C.T., et al., *Hybridization of G-Quadruplex-Forming Peptide Nucleic Acids to Guanine-Rich DNA Templates Inhibits DNA Polymerase eta Extension*. Biochemistry, 2014.
57. Bergoglio, V., et al., *DNA synthesis by Pol eta promotes fragile site stability by preventing under-replicated DNA in mitosis*. J Cell Biol, 2013. **201**(3): p. 395-408.
58. Phan, A.T., et al., *Structure of an Unprecedented G-Quadruplex Scaffold in the Human c-kit Promoter*. J. Am. Chem. Soc., 2007. **129**: p. 4386-4392.
59. Amato, J., et al., *Targeting G-quadruplex structure in the human c-Kit promoter with short PNA sequences*. Bioconjug Chem, 2011. **22**(4): p. 654-63.
60. Datta, B., C. Schmitt, and B.A. Armitage, *Formation of a PNA2-DNA2 Hybrid Quadruplex*. J. Am. Chem. Soc., 2003. **125**: p. 4111-4118.
61. Roy, S., et al., *High-affinity homologous peptide nucleic acid probes for targeting a quadruplex-forming sequence from a MYC promoter element*. Biochemistry, 2007. **46**: p. 10433-10443.
62. Englund, E.A. and D.H. Appella, *Synthesis of γ -substituted peptide nucleic acids: a new place to attach fluorophores without affecting DNA binding*. Organic Letters, 2005. **7**(16): p. 3465-3467.
63. Yeh, J.I., et al., *Crystal Structure of Chiral γ PNA with Complementary DNA Strand: Insights into the Stability and Specificity of Recognition and Conformational Preorganization*. J. Am. Chem. Soc., 2010. **132**: p. 10717-10727.
64. Roy, S., et al., *Kinetic discrimination in recognition of DNA quadruplex targets by guanine-rich heteroquadruplex-forming PNA probes*. Chem Commun, 2011. **47**(30): p. 8524-6.
65. Kormuth, K.A., J.L. Woolford, and B.A. Armitage, *Homologous PNA Hybridization to Noncanonical DNA G-Quadruplexes*. Biochemistry, 2016.

2. Chapter 2: Homologous PNA Hybridization to Noncanonical DNA G-quadruplexes

2.1 Chapter Summary

Potential guanine (G) quadruplex-forming sequences (QFSs) found throughout the genomes and transcriptomes of organisms have emerged as biologically relevant structures. These G-quadruplexes represent novel opportunities for gene regulation at the DNA and RNA levels. Recently, the definition of functional QFSs has been expanding to include a variety of unconventional motifs, including relatively long loop sequences (i.e. > 7 nucleotides) separating adjacent G-tracts. We have identified a QFS within the 25S rDNA gene from *S. cerevisiae* that features a long loop separating the two 3'-most G-tracts. An oligonucleotide based on this sequence, QFS3, folds into a stable G-quadruplex *in vitro*. We have studied the interaction between QFS3 and several loop mutants with a small, homologous (G-rich) peptide nucleic acid (PNA) oligomer that is designed to form a DNA/PNA heteroquadruplex. The PNA successfully invades the DNA quadruplex target to form a stable heteroquadruplex, but with surprisingly high PNA:DNA ratios based on surface plasmon resonance and mass spectrometric results. A model for high stoichiometry PNA-DNA heteroquadruplexes is proposed and the implications for quadruplex targeting by G-rich PNA is discussed.

2.2 Introduction

Certain sequence motifs within nucleic acids have the ability to adopt distinct secondary structural topologies, such as guanine (G) quadruplexes[1]. *In vivo*,

G-quadruplexes can act as structured regulatory elements that have the ability to regulate biological processes at the DNA and RNA levels through interaction with intracellular factors[2-5]. The quadruplex-forming capability of regions across the human genome has been experimentally demonstrated by immunofluorescence studies relying on a G-quadruplex binding antibody, supporting the biological relevance of these structures[6]. The prevailing, sequencing-informed estimate places the number of G-quadruplexes within the human genome at >700,000[7]. The breadth of potential quadruplex-forming sequences (QFS) across DNA alone highlights the monumental challenge of identifying the functions of individual G-quadruplexes within the cell.

G-quadruplexes are formed from the mutual interaction of four sets of G-tracts via Hoogsteen H-bonding to form a series of stacked, planar G-tetrads (Chapter 1, Figure 1.1). In an intramolecular G-quadruplex, the nucleotides separating the G-tracts become loops, which are external to the stacked G-tetrads that form the core of the structure. The canonical intramolecular G-quadruplex motif was originally defined as: $G_{2-4}N_{1-7}G_{2-4}N_{1-7}G_{2-4}N_{1-7}G_{2-4}$ [8, 9]. This formula was designed to identify motifs having features that are most likely to support *in vivo* quadruplex folding. For example, short loops confer the quadruplex with greater structural stability than longer loops[10, 11]. However, in recent years, the quadruplex motif definition has been expanded to include a variety of non-canonical sequences. For example, certain motifs having the backbone G-tracts broken up by one or more bulge nucleotides are able to support G-quadruplex folding[12]. It has also become clear that the loops

separating the G-tracts need not be limited to only 7 nucleotides in length. A striking example of a long-looped quadruplex was identified within the hTERT (human telomerase reverse transcriptase) core promoter. This region of the gene promoter contains 12 individual G-tracts that can fold into a pair of tandem G-quadruplexes, including one with a 26 nt loop that forms its own internal hairpin, conferring enhanced stability to the structure[13]. More recently, a long-looped quadruplex-forming sequence was identified in the P1 promoter element of the *BCL2* gene[14]. This sequence was identified to fold into two separate similar conformations having either 11 or 12 nt central hairpin loops, with its folding and structure likely related to the negative superhelicity of the DNA associated with transcription levels[14]. Long-looped G-quadruplexes may, in fact, represent a significant portion of the intracellular quadruplex population. Specifically, potential QFSs with very long (up to 90 nt) central loops and single nucleotide flanking loops have been identified within the 5' UTRs of human mRNAs[7]. Further, the most updated high-throughput search for potential QFSs in the human genome, which accounts for structures having bulges and long loops, suggests that unconventional G-quadruplexes are much more prevalent than previously thought[7]. This broader diversity in G-quadruplex sequence and structure enhances the possibility for developing synthetic ligands that can target and regulate specific quadruplexes within the cell.

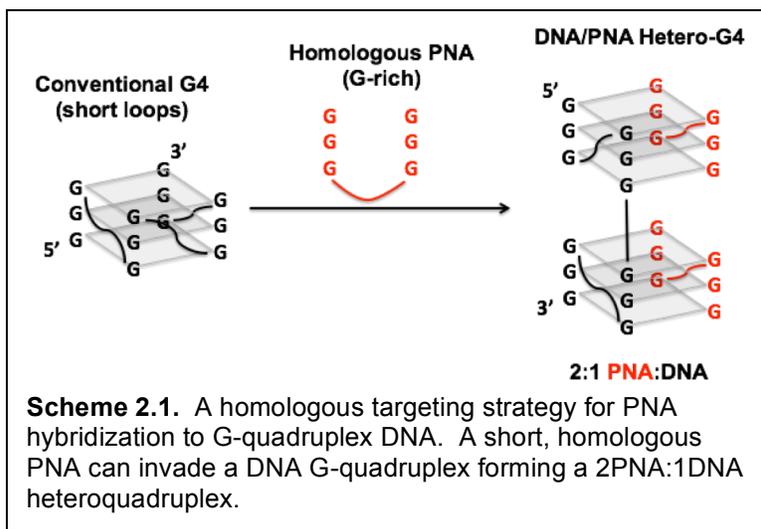
Synthetic nucleic acid analogues, such as peptide nucleic acid (PNA)[15, 16], are useful for targeting specific DNA and RNA sequences. G-quadruplexes are unusual targets in that they can interact with such probes not only via Watson-

Crick pairing to form heteroduplexes, but also via Hoogsteen pairing to form heteroquadruplexes, which are stabilized by hetero-G-tetrads. Although it is not the focus of this report, quadruplex-targeting by heteroduplex-forming complementary probes is well established for PNA[17-24], as well as other classes of oligonucleotides[25-28]. Interestingly, the resulting heteroduplexes appear to be less inhibitory than the native G-quadruplex to a processing ribosome[27, 28] or DNA polymerase enzyme[24].

The alternative option for sequence-based targeting of G-quadruplexes relies on G-rich probes that disrupt the native quadruplex in order to form a heteroquadruplex. We refer to G-rich oligomers that bind in this manner as *homologous* probes, even if the linker joining the G-tracts differs from the loops within the target because, to a first approximation, only the guanines from the probe and the target interact. We first reported this hybridization mode for targeting G-quadruplexes in 2001[29], and there have been numerous additional examples provided by our lab[21-23, 30-32] and others[33-36]. Although detailed comparisons of the two binding modes have not been made in most cases, we recently demonstrated that targeting a G-quadruplex with a heteroquadruplex-forming G-rich PNA inhibited DNA polymerase η , whereas targeting the same site with a heteroduplex-forming C-rich PNA enhanced polymerase activity[24]. The imposition of opposing effects on quadruplex stability and function through the use of different targeting mechanisms motivates further study.

This study focuses on recognition of a long-loop-containing DNA quadruplex by a G-rich PNA. We have previously found homologous PNAs having two G-

tracts invade DNA quadruplexes to form 2:1 heteroquadruplexes (Scheme 2.1).



Although we have demonstrated this binding mode for several different G-quadruplexes, the rate of hybridization can vary by more than an order of magnitude, depending

on the particular fold of the quadruplex target[31]. Given the expanding definition of what constitutes a quadruplex motif, studying the effects of long loops on heteroquadruplex formation by homologous PNA probes became important to ultimately develop the most selective and potent quadruplex-targeting PNAs.

With this work we have identified a new, 3-tetrad quadruplex motif, containing a moderately long central loop in addition to a very long 3' loop, that folds into a stable, intramolecular G-quadruplex *in vitro*. Further, we have found that a short, homologous PNA can spontaneously invade this long-looped quadruplex via intermolecular heteroquadruplex formation. However, in contrast to previous results with short-looped quadruplexes, the long loops allow for hybridization of additional PNA strands to the DNA, leading to formation of higher order (5:1 and 6:1 PNA:DNA) complexes. Based on testing of additional DNA quadruplexes, this new binding mode appears to be applicable to long-looped quadruplexes in general. We have also initiated studies that indicate loop sequence identity, in

addition to loop length, similarly can impact DNA/PNA heteroquadruplex stoichiometry.

2.3 Results

Our lab has previously shown that short PNAs consisting of two G_3 tracts and various linkers hybridize to conventional G-quadruplexes forming 2:1 heteroquadruplexes[21, 22, 30]. As the definition of a quadruplex-forming sequence expands, conventional, short-looped motifs are coming to be understood as the minority within biological systems, at least based on bioinformatic analysis[7]. Motivated in part by recent reports of quadruplexes having long (>7 nts) loops between the core G-tracts, we have become interested in learning how PNA heteroquadruplex formation may change with G-quadruplexes having very different topologies[7, 37].

Identification and design of long loop-containing G-quadruplex motifs. This study was conducted using a series of G-quadruplex-forming DNA oligomers to assess the effects of very long loops on *in vitro* G-quadruplex formation and hybridization to a small, homologous (G-rich) PNA, P_{eg2} (Table 2.1). P_{eg2} is designed to invade and hybridize to DNA G-quadruplexes forming intermolecular, DNA/PNA heteroquadruplexes. Specifically, the two G_3 tracts allow for formation of a 3-tetrad heteroquadruplex in combination with two additional G_3 tracts donated by the target DNA. The G-tracts are separated by two abasic diethylene glycol “miniPEG” units; the abasic linker was originally designed to suppress hybridization to complementary sequences, but does not impact the affinity or hybridization kinetics of heteroquadruplex formation[32]. Meanwhile, the target

QFS3 was originally identified in the 25S rDNA of the yeast *Saccharomyces cerevisiae* using the algorithm $G_3X_nG_3X_nG_3X_nG_3$, with a maximum total sequence length of 50 nucleotides, i.e. the three X_n loops in total could comprise up to 38 nucleotides[38]. QFS3 is particularly interesting because it contains both a moderately long (6 nt) central loop and a very long (17 nt) 3' loop. Myc19 is derived from a quadruplex-forming sequence found within the nuclease hypersensitivity element III₁ of the *c-MYC* promoter[39, 40]. In contrast to QFS3, Myc19 has three very short (1, 2, and 1 nt) loops. Our lab has utilized Myc19 as a model G-quadruplex target for PNA probes[30]. In order to study the effects of varied loop length and placement on G-quadruplex formation, two artificial potential G-quadruplex-forming sequences were designed. QFS3_{L2} and QFS3_{L3} are both derived from QFS3 but each has a loop replaced with the corresponding short loop of Myc19. hTERT was included as an example of a G-quadruplex having an especially long central loop (26 nt) that folds into its own hairpin, greatly stabilizing the overall structure[13].

Table 2.1. Sequences of PNA probes and DNA G-quadruplex targets. (G-tracts predicted or reported to participate in tetrad formation shown in bold and underlined.)

PNA (N-to-C)	
P _{eg2}	H- GGG -(miniPEG) ₂ - GGG -Lys-NH ₂
DNA (5'-to-3')	
Myc19[39]	AGGGTGGGGAGGGTGGGGA
QFS3	GGGTCGGGTAGTGAGGGCCTTGGTCAGACGCAGCGGG
QFS3 _{L2}	GGGTCGGGGAGGGCCTTGGTCAGACGCAGCGGG
QFS3 _{L3}	GGGTCGGGTAGTGAGGGTGGG
hTERT[13]	GGGGCTGGGCCGGGGACCCGGGAGGGGTCGGGACGGGGCGGGG

G-rich DNA oligomers having long 3' loops fold into stable, intramolecular G-quadruplexes in vitro. G-quadruplex formation and stability in vitro can be

detected using biophysical and spectroscopic techniques. We have confirmed that QFS3, QFS3_{L2}, and QFS3_{L3} DNA oligomers fold into G-quadruplexes using three optical spectroscopic methods. *In vitro* G-quadruplex formation was detected by UV melting at 295 nm, with each oligomer producing a quadruplex-specific hypochromic transition at this wavelength[41]. From these data we extracted a melting temperature (T_m) for each quadruplex (Table 2.2). The T_m values determined in 100 mM K^+ are significantly lower than what would be observed for a quadruplex having only short loops, consistent with the destabilizing effect that long loops are reported to have on quadruplex structures[42, 43]. (Note that the T_m value for the hTERT quadruplex was determined in the presence of 10 mM KCl due to its higher thermal stability relative to the other quadruplexes.) In addition, a minimal effect is observed on the T_m values recorded at 5-fold higher DNA concentration. This result confirms that each QFS forms intramolecular G-quadruplexes *in vitro*. Finally, replacing KCl with LiCl in the buffer significantly destabilized the quadruplexes, as shown in Supplemental Figure 2.S3 for QFS3, consistent with the known quadruplex-destabilizing effect of Li^+ [44].

Table 2.2. Concentration dependence of DNA quadruplex melting temperature (T_m , °C) determined in 100 mM K^+ .

DNA	1 μ M DNA	5 μ M DNA
QFS3	54.9 \pm 3.0	57.0 \pm 0.2
QFS3 _{L2}	56.7 \pm 1.3	56.7 \pm 0.3
QFS3 _{L3}	68.7 \pm 0.2	68.3 \pm 0.6
hTERT [†]	64.8 \pm 1.6	61.8 \pm 0.8

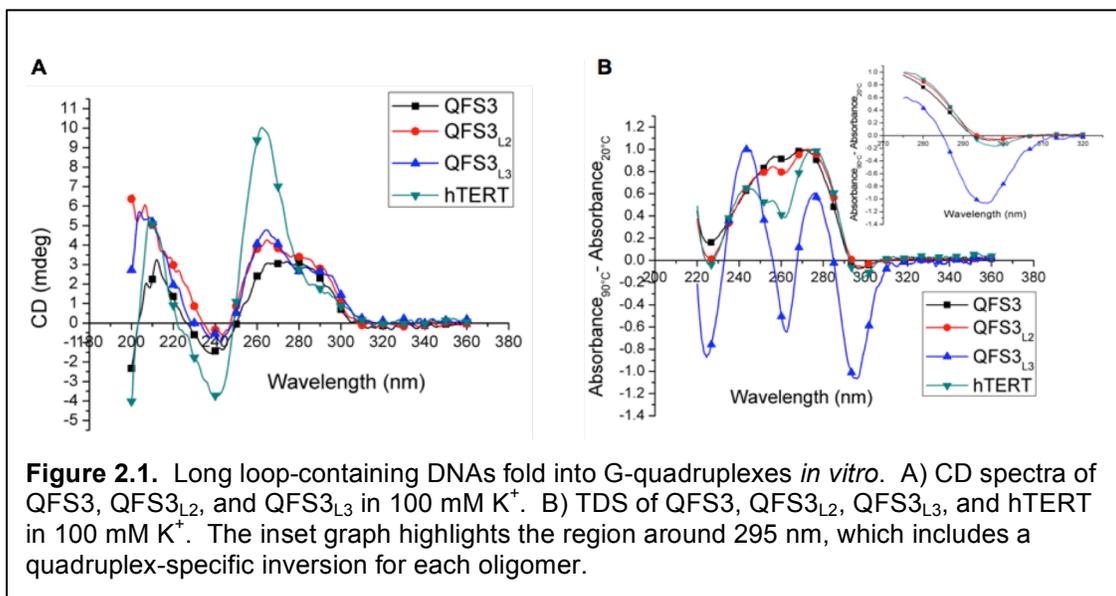
[†] Measured in 10 mM K^+

We also used circular dichroism (CD) spectropolarimetry to confirm the presence of G-quadruplexes in each sample and gain insight into the overall

topology of each structure. G-quadruplexes are often categorized based on the orientation of the G-tracts relative to each other. For example, in a parallel quadruplex, all four G-tracts run in the same direction, while an antiparallel quadruplex has two pairs of G-tracts running in opposite directions; these structural differences give rise to unique CD signatures[45]. The CD spectra of QFS3, QFS3_{L2}, and QFS3_{L3} contain characteristic G-quadruplex spectral features. The broad peak from 260-300 nm produced by QFS3, QFS3_{L2}, and QFS3_{L3} suggests a more complex topology (e.g. hybrid structure) than a standard parallel (min. at 240 nm, max. at 260 nm) structure (Figure 2.1A). Additionally, the unpaired bases in the loop should contribute to the shape of the spectrum and this contribution would be greater for longer-looped quadruplexes. The stronger signal from hTERT likely arises from the internal hairpin, present within loop 2, since B-form DNA also exhibits a positive peak at 260 nm[13]. The long loops of the other three quadruplexes are not predicted to form a base-paired secondary structure.

Thermal difference spectroscopy (TDS), provides additional insight into nucleic acid secondary structures based on features of the difference spectrum that is produced by subtracting the UV absorption spectrum of the folded nucleic acid from that of the unfolded (i.e. thermally denatured) form[46]. Each DNA oligomer in Table 2.1 produces a TDS containing spectral features characteristic of G-quadruplexes, in particular the distinctive inversion near 295 nm (Figure 2.1B, see inset). Although this feature is weak for QFS3 and QFS3_{L2}, the intensity is similar to that of the known long-loop quadruplex hTERT. The

spectrum of QFS3_{L3} differs considerably from those of the other quadruplexes in Figure 2.1B, but is similar to that of Myc19 (Supplemental Figure 2.S1). While the origin of the differences among the spectra is unclear, variation in thermal difference spectra produced by different DNA G-quadruplexes has been reported for other systems[46].



PNA invasion of long loop-containing G-quadruplexes. The G-rich PNA **P_{eg2}** interacts favorably with each long-looped G-quadruplex, evidenced by the increase in T_m upon the addition of two equivalents of PNA to each DNA (Table 2.3; Supplemental Figure 2.S4).

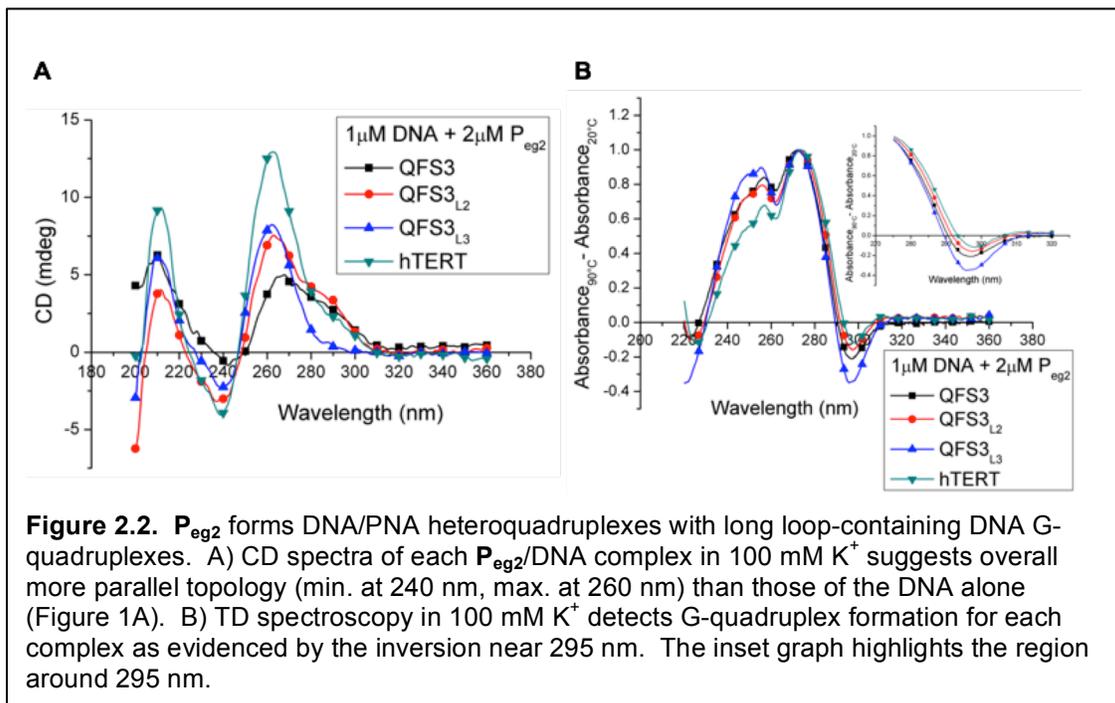
Table 2.3. PNA-DNA heteroquadruplex melting temperature (T_m , °C) and stabilization relative to DNA homoquadruplex (ΔT_m , °C) determined in 100 mM K⁺.

DNA	T_m	ΔT_m
QFS3	61.6 ± 0.4	6.7
QFS3 _{L2}	75.6 ± 3.1	18.9
QFS3 _{L3}	77.2 ± 3.9	8.5
hTERT [†]	> 95	> 30

[DNA] = 1 μM, [P_{eg2}] = 2 μM

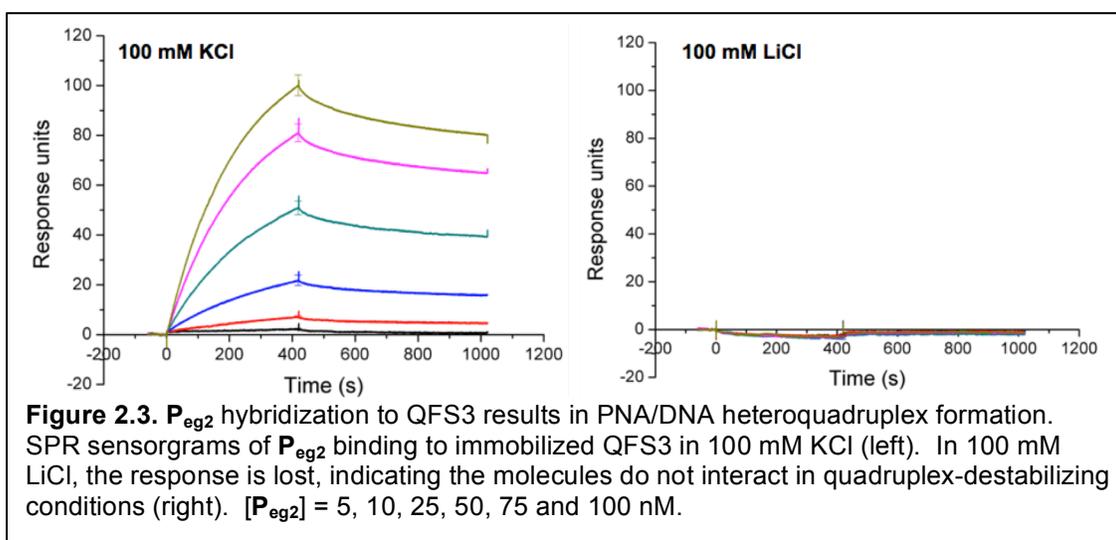
[†] Measured in 10 mM K⁺.

CD and TDS experiments are consistent with hetero-G-quadruplex formation in pre-annealed samples of 2:1 PNA:DNA (Figure 2.2). In the presence of PNA, the CD signal for each QFS is intensified and appears more parallel (i.e. increased positive intensity at 260 nm) than that of the DNA alone (Figure 2.2A; overlaid CD spectra for each quadruplex with and without PNA are shown in Supplemental Figure 2.S5). Similar results were observed for heteroquadruplex formation between G-rich PNA and a DNA QFS based on the human telomeric repeat[31]. Analogous results are obtained from TDS, where the spectral shapes for each QFS become much more similar in the presence of PNA than for the DNA alone (compare Figures 2.1B and 2.2B).



To confirm that the PNA interacts with these DNAs via intermolecular G-quadruplex formation, we carried out SPR experiments where PNA was flowed

past immobilized QFS3 (Figure 2.3). An obvious increase in response units was detected upon injection of increasing concentrations of PNA in buffer containing 100 mM K^+ . When the experiment was repeated in buffer containing 100 mM Li^+ , nearly all the response signal was lost, consistent with PNA-DNA heteroquadruplex formation. We obtained similar results for additional SPR experiments conducted with QFS3_{L2}, QFS3_{L3}, and hTERT (Supplemental Figure 2.S6).



Sensorgrams obtained for P_{eg2} hybridization to five different G-quadruplexes at 50 nM PNA concentration are shown in Figure 2.4, with initial association and dissociation rates shown in Table 2.4. Heteroquadruplex formation for the long-looped G-quadruplexes is noticeably slower than to the short-looped quadruplex, Myc19. This result is consistent with our prior findings that heteroquadruplex formation is slower for DNA quadruplexes that adopt hybrid structures than for parallel morphologies[31].

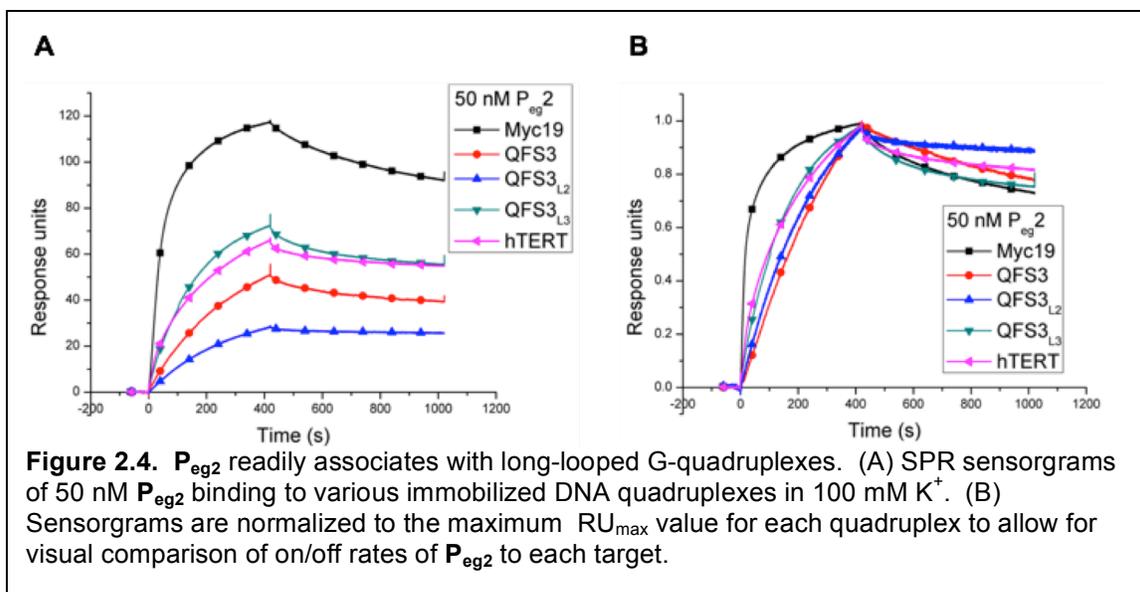
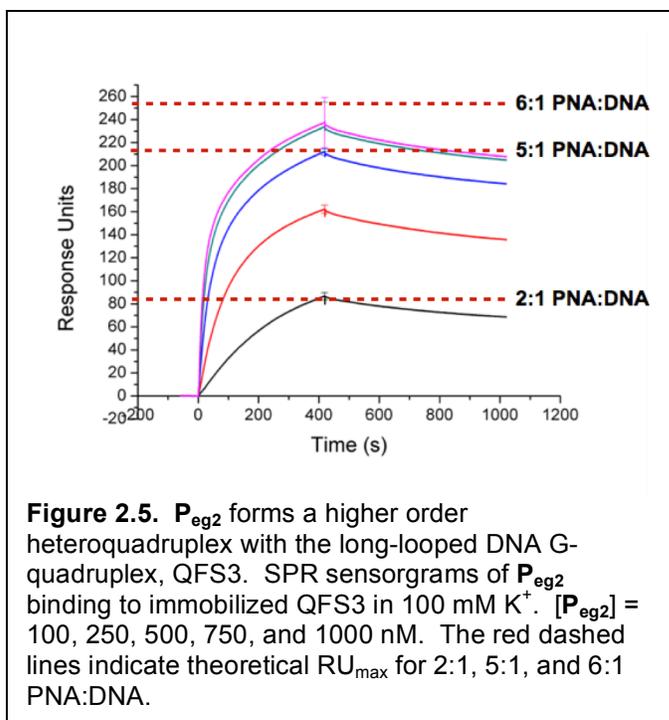


Table 2.4. PNA association (on) and dissociation (off) rates in $RU \cdot s^{-1}$ and ratio (Off/On) determined at 50 nM P_{eg2} .

DNA	On Rate	Off Rate	Ratio
Myc19	0.501 ± 0.008	0.086 ± 0.003	0.17
QFS3	0.171 ± 0.007	0.338 ± 0.001	0.18
QFS3_{L2}	0.115 ± 0.006	0.014 ± 0.004	0.12
QFS3_{L3}	0.338 ± 0.017	0.077 ± 0.004	0.20
hTERT	0.353 ± 0.004	0.048 ± 0.004	0.14



Detection of higher order complex formation between P_{eg2} and long-looped G-quadruplexes. A curious result from the SPR experiments was that considerably more PNA was required to saturate the signal than expected for a 2:1 stoichiometry. For example, Figure 2.5 shows sensorgrams

for binding of P_{eg2} to QFS3. The red dashed lines indicate the calculated saturating response units (RU_{max}) for 2:1, 5:1 and 6:1 stoichiometries, based on the amount of DNA immobilized on the SPR chip (determined as described in the Experimental section). Based on these calculations, the PNA:DNA stoichiometry appears to be either 5:1 or 6:1. (Uncertainty arises from the fact that a small amount of DNA is released from the SPR chip during the regeneration phase after each injection of PNA, i.e. the RU_{max} value for a given stoichiometry decreases over the duration of the experiment.) Electrospray ionization mass spectrometry experiments also supported formation of both 5:1 and 6:1 heteroquadruplexes, as described in supporting information (Figures 2.S7 and 2.S8). High-stoichiometry (e.g. 4:1 and 5:1) complexes were also indicated by SPR for P_{eg2} binding to the other long-loop quadruplexes (Supplemental Figure 2.S9) indicating that this phenomenon is not specific to QFS3. However, higher

order complex formation does not appear to require exceptionally long loops, as the longest loop of QFS3_{L3} has only 6 nucleotides. In contrast, binding of **P_{eg2}** to the short-looped quadruplex Myc19 saturates at 2:1 stoichiometry (Supplemental Figure 2.S10), consistent with our previous results for a slightly different quadruplex-forming PNA[30].

The target loop sequence affects DNA/PNA complex stoichiometry.

Although initial studies focused on the effect of G-quadruplex loop length on the stoichiometry of the DNA/PNA heteroquadruplex, loop sequence identity may also be an important factor in PNA hybridization. To assess this question, we designed a series of simplified, ‘T-loop’ 3-tetrad DNA G-quadruplex targets having central loops of varied lengths (Table 2.5). Despite the maximum loop length of QFS3_{L3} being only 6 nt, this motif still supports higher order heteroquadruplex formation with **P_{eg2}**. Therefore, we selected 6 nt as the maximum central loop length for our T-loop quadruplexes.

Table 2.5. Sequences of T-loop DNA G-quadruplex targets. (G-tracts predicted to participate in tetrad formation shown in bold and underlined.)

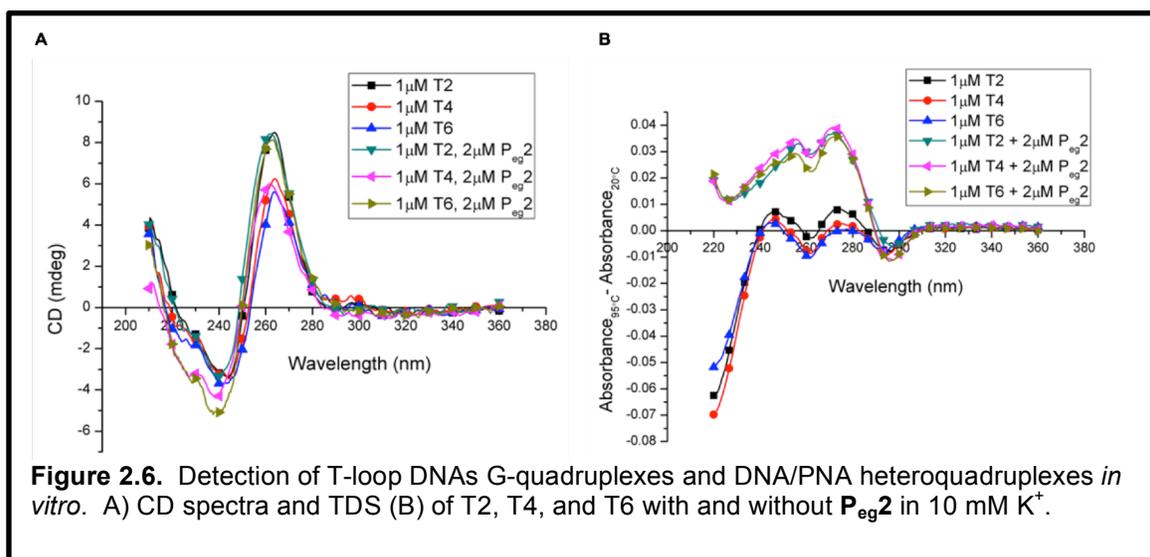
DNA (5'-to-3')	
T2	<u>GGTGGG</u>TT<u>GGGTGGG</u>
T4	<u>GGTGGG</u>TTTT<u>GGGTGGG</u>
T6	<u>GGTGGG</u>TTTTTT<u>GGGTGGG</u>

UV melting (Table 2.6; Supplemental Figure 2.S11), CD, and TDS data confirmed that each of the T-loop DNAs folded into stable G-quadruplexes *in vitro*, which were even further stabilized upon invasion of **P_{eg2}**, forming DNA/PNA heteroquadruplexes (Figure 2.6). In both cases, loop length is inversely proportional to quadruplex stability, with T6 producing the lowest T_m 's, even in

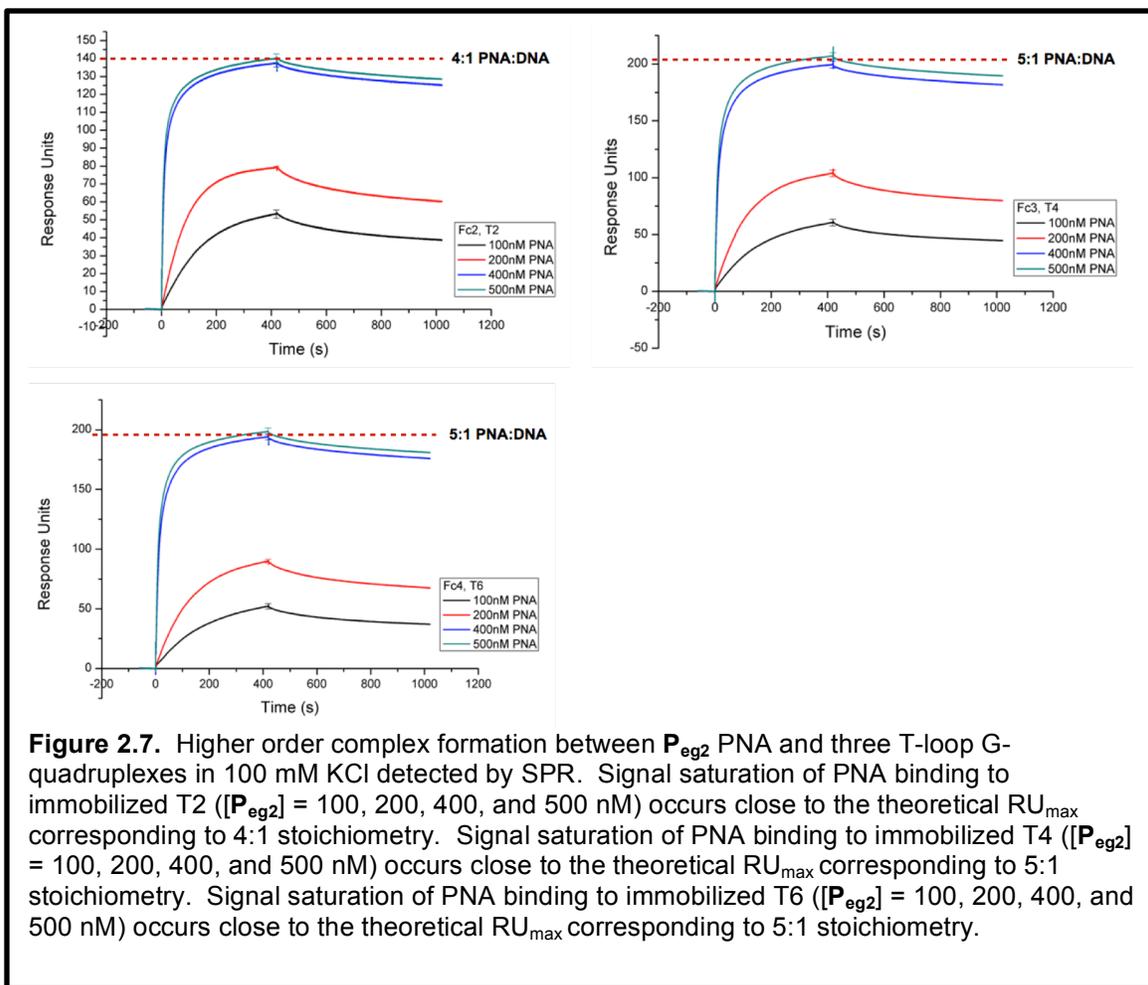
the presence of the PNA. Each of the T-loop quadruplexes can be characterized as parallel by CD. As with the long-looped quadruplex targets, introduction of the PNA enhances this parallel signal, particularly at the 260 nm wavelength (Figure 2.6A). The quadruplex-specific inversion at 295nm is present in all the TD spectra, confirming quadruplex formation for each of the DNAs, as well as the DNA/PNA complexes. TD spectra of the T-loop motifs is similar to the spectra of QFS3_{L3} (Figure 2.1B) and Myc19 (Supplemental Figure 2.S1), which all do not exceed a 6 nt maximum loop length. Each of the heteroquadruplexes produce similar TD spectra with $\Delta_{\text{Abs}} > 0$ at wavelengths below 290 nm. This result matches the spectra for the long-looped quadruplexes in complex with the PNA (Figure 2.2B), suggesting that PNA hybridization contributes directly to this change in signal.

Table 2.6. DNA quadruplex and DNA/PNA heteroquadruplex melting temperatures (T_m , °C) determined by UV melting in 10 mM K^+ .

DNA	1 μM DNA	1 μM DNA, 2 μM PNA
T2	77.2 \pm 1.2	ND
T4	64.1 \pm 0.6	78.3 \pm 0.7
T6	57.9 \pm 1.2	73.1 \pm 1.3



SPR analyses confirmed that PNA interaction with each of these DNAs resulted in heteroquadruplex formation, as the response signal was lost in buffer containing 100 mM Li⁺ (Supplemental Figure 2.S12). However, each of the T-loop quadruplexes, including the shortest quadruplex-forming sequence (T₂) were able form higher order complexes with **P_{eg}2** (Figure 2.7), suggesting that higher order heteroquadruplex formation is also dependent on loop sequence identity in addition to length. This result, however, contrasts our findings with Myc19, which is a similar, short-looped quadruplex motif. The central loop of Myc19 contains two purines (GA), while the central G-tracts are separated by two pyrimidines (T₂), which may contribute to this difference in the upper limit of complex stoichiometry.



2.4 Discussion

G-rich motifs with varied long loop placement support stable G-quadruplex formation. Early bioinformatics studies determined that >300,000 G-quadruplex forming sequences are present in the human genome. While this is a large number, recent research indicates that it substantially under-estimates the actual number of QFSs due to the fact that quadruplex formation is supported by a considerably broader range of sequences than defined by the original search algorithm. The ability of QFSs to tolerate extra bases (bulges) within G-tracts, to adopt non-canonical folds wherein corners of the quadruplex include Gs from non-contiguous tracts, and the successful folding of quadruplexes even when

long (> 7 nt) loops are present[7, 13, 37], all indicate that the definition of a QFS is much broader than was originally anticipated. The results presented above represent our first effort to investigate the impact of long loops on G quadruplex recognition by a heteroquadruplex-forming PNA.

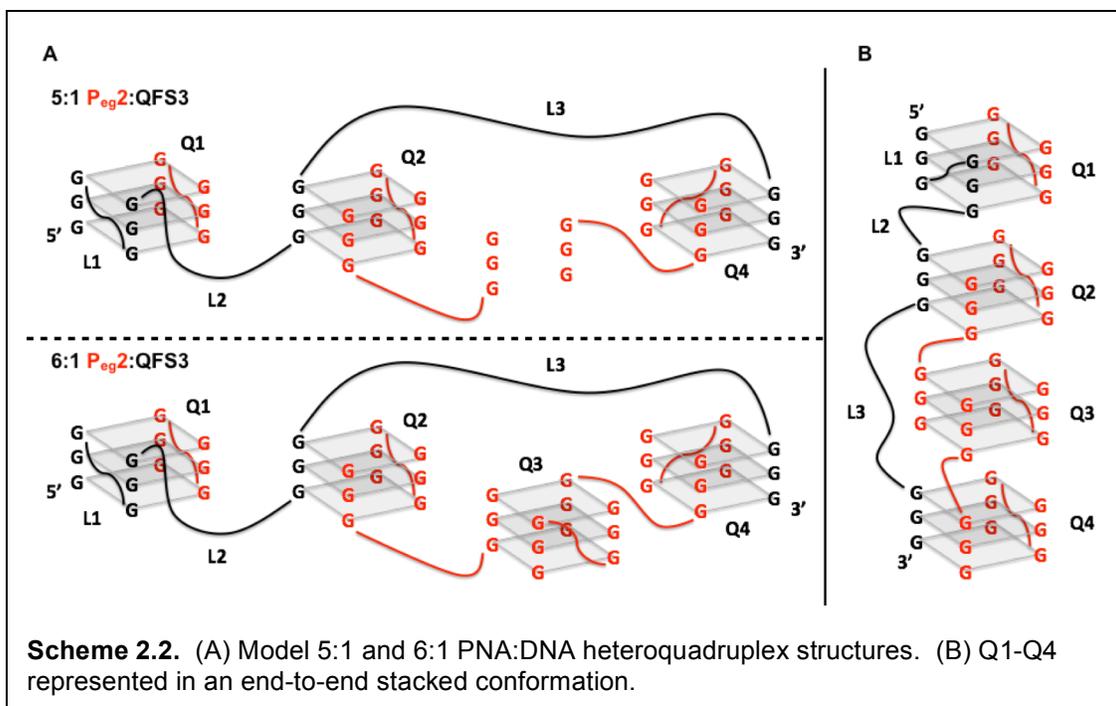
To the best of our knowledge, QFS3 represents a new quadruplex motif having a moderately long central loop in addition to an especially long 3' loop (in contrast to a long central loop with single nucleotide flanking loops). We have shown that QFS3 and similar DNA oligomers can fold into intramolecular G-quadruplexes *in vitro* with melting temperatures well above physiological temperature (30 °C for yeast), suggesting that these structures have the potential to form *in vivo*. Furthermore, a G-quadruplex of moderate stability may be useful to the cell as a regulatory switch especially in the presence of other molecular factors or changing intracellular conditions. QFS3 is also a potentially interesting biological quadruplex, due to its location within yeast 25S rDNA. The presence of sequences with quadruplex-forming potential in eukaryotic rDNA has been noted previously, although the biological role of such quadruplexes is yet to be determined[38, 47]. In addition, the fact that QFS3 is on the non-template strand means that it will also be present in the 25S rRNA. While there is no evidence that it folds into a quadruplex within the mature ribosome[48], it is possible that a quadruplex forms transiently during rDNA replication/recombination, rRNA processing, or ribosome assembly. Alternatively, a DNA-RNA heteroquadruplex could form during transcription in order to regulate ribosome levels in the cell[49].

Long loops allow for hybridization of >2 homologous PNAs to a DNA quadruplex. Our group has previously established a homologous targeting strategy using G-rich PNA oligomers to invade G-quadruplexes forming stable, intermolecular heteroquadruplexes[21-23, 29-32]. We have now shown that long-looped G-quadruplexes are also amenable to homologous PNA hybridization. However, the long loops provide these quadruplexes with remarkable plasticity in regards to the number of PNA strands in complex with each DNA.

Scheme 2.2A illustrates possible structures for 5:1 and 6:1 PNA:DNA heteroquadruplexes. After opening of the quadruplex, the first PNA molecule hybridizes analogous to the proposed binding mode for short-loop quadruplexes[30] to form heteroquadruplex Q1. However, the long third loop hinders binding of another PNA molecule to both the third and fourth G-tracts. (The extended loop likely provides the molecule with sufficient flexibility to accommodate binding of additional PNA strands.) Rather, two PNA molecules combine with the third G-tract to form a termolecular heteroquadruplex (Q2). Two PNAs also combine with the fourth G-tract in this manner (Q4), giving rise to the 5:1 complex. This model leaves a dangling G-tract for each of the two PNAs that are bound within Q2 and Q4, which can be crosslinked into an all-PNA quadruplex using the sixth PNA molecule (Q3). Further, these tandem quadruplexes can potentially stack end-to-end (Scheme 2.2B), enhancing the stability of the overall structure, which could explain the drive toward higher-stoichiometry complexes[13, 50]. While this model stands in contrast to the

'beads-on-a-string' conformation where the individual quadruplexes do not stack, as has been proposed for long telomeric DNAs[51-53], the differences in linker length and composition for the PNA-DNA heteroquadruplexes reported here could give rise to different preferred structures[53]. Regardless of the structural details, the model for a 6:1 complex is appealing because the total number of G-tracts is 16, with four contributed by QFS3 and two contributed by each of the 6 PNA molecules. If the total number of G-tracts is a multiple of 4, a structure can form in which each of the G-tracts participates in a homo- or heteroquadruplex.

SPR also indicated the formation of higher-order heteroquadruplexes between P_{eg2} and DNA quadruplexes QFS3_{L2}, QFS3_{L3}, and hTERT. The results for QFS3_{L3}, which has a maximum loop length of only 6 nt, suggests that the quadruplex target need not have an exceptionally long loop before higher order PNA hybridization stoichiometries become possible. Motifs with very short loop lengths, such as T2, are also capable of higher order heteroquadruplex formation, although this is not the case with Myc19. G-quadruplex loop sequence, as well as length, influences the confirmation of the structure[54], and loop nucleotides can favorably contribute to the enthalpy of quadruplex folding via stacking with the core G-tetrads[55]. Myc19 could be restricted to 2:1 heteroquadruplex formation due to its central GA loop, which may participate in stacking interactions that rigidify the structure. The poly-T loops, conversely, may be more flexible, thus allowing for hybridization of more than two PNA strands.



Evidence for formation of higher stoichiometry PNA-DNA heteroquadruplexes was recently reported by Panyutin and coworkers, through investigation of PNA (G₄T₄G₄) hybridization to the homologous target sequence in a DNA oligonucleotide, as well as in the context of a DNA hairpin loop and a plasmid[33]. Observation of multiple gel-shifted bands indicated formation of at least two different heteroquadruplexes of differing mobility, while optical spectroscopy detected at least one of these heteroquadruplexes with a 3:1 stoichiometry (PNA:DNA). Interestingly, the target only contained two G-tracts, precluding formation of its own intramolecular G-quadruplex. In our QFS3 system, which has four G-tracts, we propose a stoichiometry of 5:1 or 6:1, with the latter being double of what Panyutin *et al.* found for a truncated target with half as many G-tracts.

Implications for biological G-quadruplex targeting. We have shown that long loop-containing G-quadruplexes can be invaded by homologous, G-rich PNA. However, the tendency of these long loops to accommodate higher order complex formation *in vitro* is not ideal for biological applications of this targeting method, since the higher PNA concentrations required to achieve stable binding would increase the likelihood of unwanted, off-target interactions. Additionally, our T-loop data suggest that sequence identity similarly influences PNA hybridization, possibly due to enhanced flexibility of the poly-T loops when compared to similar oligonucleotides, such as Myc19. Further studies will be required to more completely understand these effects. Therefore, future probe designs for targeting similar quadruplexes should focus on decreasing the stoichiometry of these complexes. A potential solution, particularly for long-looped quadruplexes, is the combination of homologous and complementary domains, allowing for bimodal quadruplex hybridization, i.e. hetero-duplex and – quadruplex formation by the same probe, as reported recently by Basu and coworkers[56]. The complementary domain could be targeted either to an external flanking region or within the long loop, while the homologous domain then associates with two nearby G-tracts to form the heteroquadruplex part of the final structure. Optimization of the relative affinities of the two domains should then allow selective targeting of one quadruplex (or quadruplex-forming sequence) among the millions of QFSs present in the genome and transcriptome.

2.5 Conclusion

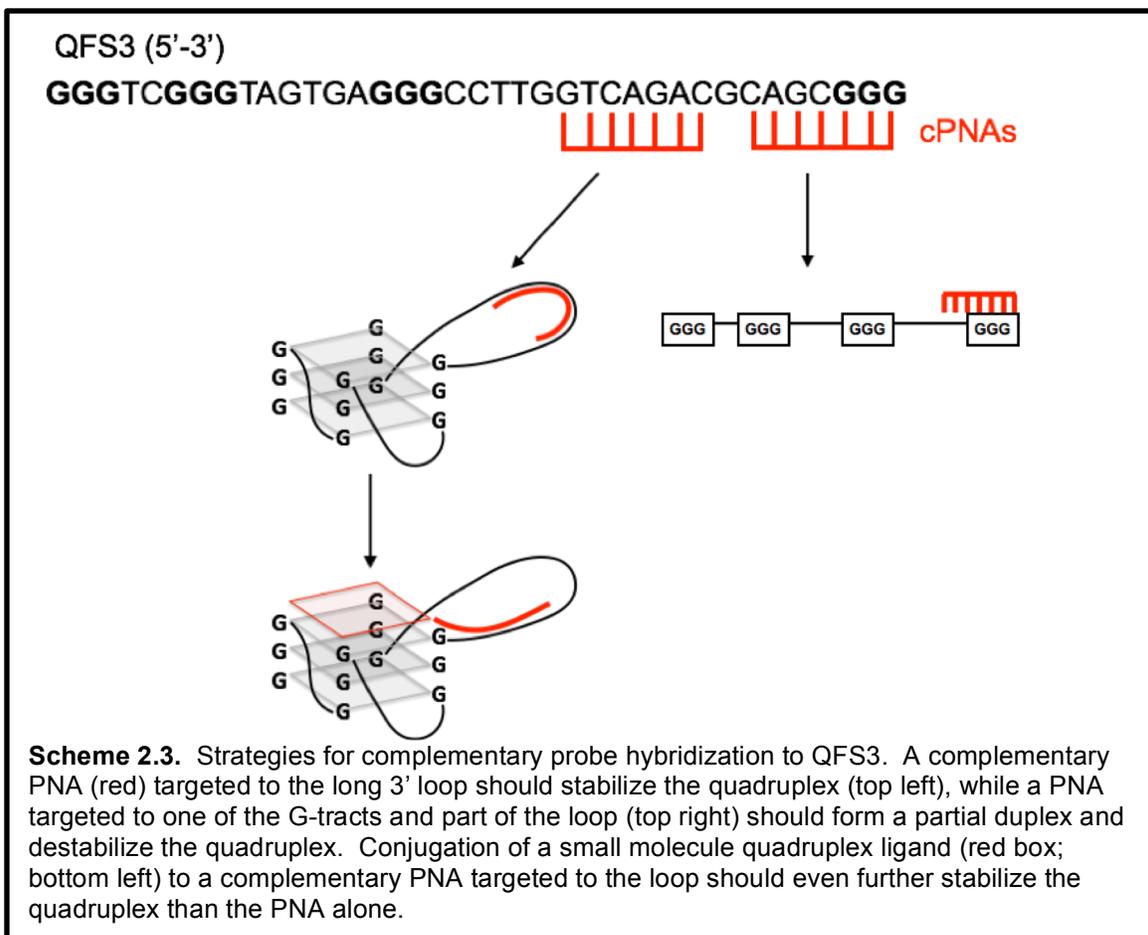
This work provides new insight into potential targeting of a specific class of G-quadruplexes. Additional studies will be required to develop a better understanding of how loop length, structure, and sequence affect PNA hybridization. Nevertheless, we expect that the prevalence of unusual quadruplex motifs in biology will require the development of new targeting strategies to allow identification and modulation of G-quadruplex functions in the future.

2.6 Future Directions

The work discussed here has expanded our understanding of homologous PNA interactions with unconventional DNA G-quadruplexes. However, we have not yet addressed the problem of selectivity that frequently limits the application of synthetic quadruplex ligands. Future efforts should be directed toward the development of technologies that exploit the unique features of noncanonical G-quadruplexes, including long loop regions. Long loops can potentially act as handles for hybridization of complementary PNA probes. Inspired by previous efforts with the cKit87up promoter quadruplex[57] (discussed in detail in Chapter 1), incorporation of structure within the long loop using a complementary PNA may serve to enhance overall quadruplex stability. Specifically, a complementary PNA designed to hybridize to only loop nucleotides, but not interact with the G-tetrad core, could aid in structuring of disordered loops, and enhance overall structural stability. In the case of long-looped quadruplexes, such as QFS3, such a PNA could be designed for selective hybridization entirely within the

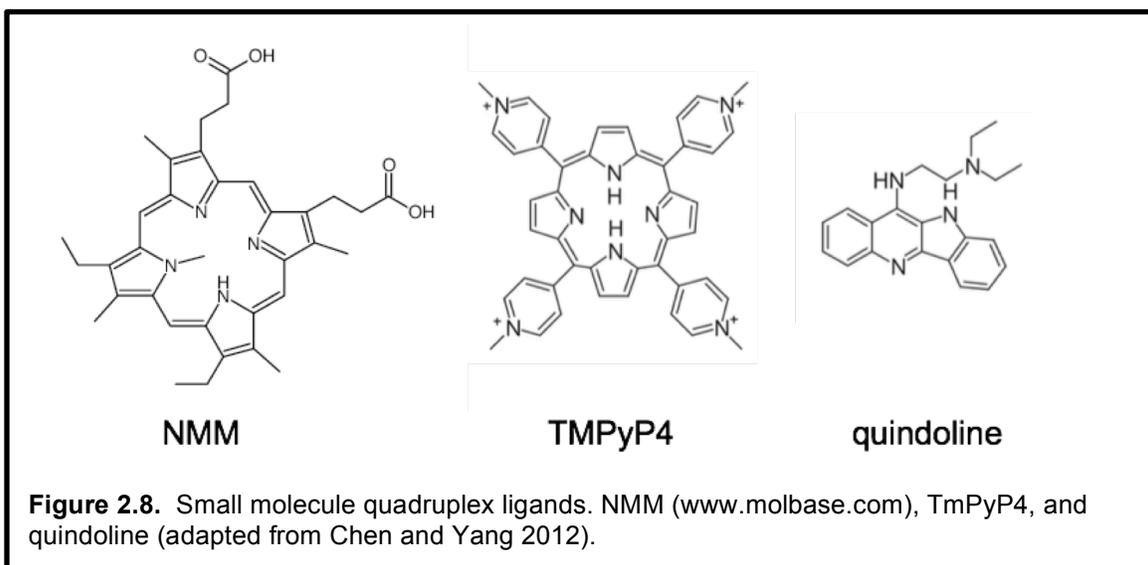
quadruplex motif, with the added benefit of minimizing potential interactions between the probe and *cis*-elements in adjacent nucleotides. This strategy also mimics the natural G-quadruplex topology of the hTERT[13] and, more recently, Bcl2[14] promoter quadruplexes, which include hairpin loops that enhance the stability of the structure.

cKit87up is an especially unusual G-quadruplex, having four loops instead of three[57]. Therefore, QFS3 may be a more desirable candidate to explore complementary PNA/long loop interactions. I propose to expand applications of complementary PNA targeting to long-looped G-quadruplexes, which can involve probes designed to either stabilize or destabilize the quadruplex structure, depending on the desired experimental output. Specifically, I expect that a complementary PNA targeted to the center of the long 3' loop of QFS3 will enhance the stability of the structure. Shifting the PNA targeting site to include the 3' G-tract should result in conversion to a partial duplex structure. A similar strategy with the cKit87up quadruplex was not sufficient to unfold the structure until the quadruplex was slightly destabilized by substitution of NH_4^+ for K^+ in the buffer[57]. Based on the relatively low thermal stability of QFS3, I anticipate that the PNA will be able to unfold this quadruplex, especially if the probe is designed as a right-handed helix through the inclusion of γ -modified monomers[58, 59] (discussed in detail in Chapter 3). The results of this work should expand the applications of PNA probe hybridization to long-looped G-quadruplexes, adding complementary PNAs that de-/stabilize these structures, along with heteroquadruplex-forming PNAs discussed in this chapter (Scheme 2.3).



Second generations of the quadruplex-stabilizing probe should seek to apply small molecule quadruplex ligands in the context of a selective complementary probe (Scheme 2.3). Conjugation of a small molecule to a PNA, producing a probe capable of bi-modal hybridization to a quadruplex target, should even further stabilize the quadruplex to which it binds. Small molecule ligands, such as porphyrins, can enhance quadruplex stability via π - π stacking and electrostatic interactions (reviewed in Monchaud and Teulade-Fichou, 2007)[60]. A candidate molecule, TmPyP4 (Figure 2.8), is compatible with several binding modes, including external stacking or intercalation between the G-tetrads[60]. N-methyl mesoporphyrin IX (NMM) could also be explored as an alternative ligand

(Figure 2.8). NMM is selective for binding to quadruplex over duplex DNA, but with low affinity[60]. Conjugation to a high affinity PNA probe could allow for targeted hybridization to specific quadruplex structures. These or similar molecules could be linked to the ends of the PNA probe, or potentially incorporated into PNA monomers that could allow for variation of the placement of these molecules within probes[61]. However, the symmetrical structure of these porphyrins may make chemical synthesis challenging. To avoid this problem, crescent-shaped quadruplex ligands, such as quindoline (Figure 2.8) could also be explored[62]. The application of this type of strategy should result in significant advancement in selective quadruplex targeting, especially in the context of long-looped G-quadruplex structures.



2.7 Materials and Methods

2.7.1 Materials

DNA oligonucleotides were purchased from Integrated DNA Technologies (www.idtdna.com) and Fisher Scientific (www.fishersci.com). Boc/Z-protected

PNA monomers were purchased from Poly Org (www.polyorginc.com). PNA oligomers were synthesized manually on a lysine-substituted MBHA resin by standard solid-phase peptide synthesis techniques[63, 64]. Monomers were activated by HBTU for 1 minute in the presence of DIEA and DMF. The PNA oligomer was cleaved from the solid support using TFMSA:TFA:thioanisol:m-cresol (1:1:2:6), purified by reverse-phase high-performance liquid chromatography (HPLC) and characterized by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry (calculated mass: 2184, observed mass: 2183).

2.7.2 Equipment

Absorbance data were collected using a Varian Cary 3 spectrophotometer with a temperature controlled multicell holder. Circular dichroism experiments were completed in a Jasco J-715 spectropolarimeter having a single cell holder with a water circulating temperature controller.

2.7.3 UV Melting

Samples (varied concentrations of DNA and/or PNA) were buffered in 100 mM KCl, 10 mM Tris pH 7, and 0.1 mM Na₂EDTA. The experiments were completed in two phases, first where the samples were heated to 95 °C and then annealed by slowly cooling to 15 °C (1 °C/min). The samples were then slowly heated to 95 °C. Absorbance was measured at 295 nm. Melting temperatures were determined by averaging the minimum of the first derivative of each melting curve, in triplicate (OriginPro 8). To allow for comparison with ESI-MS

conditions, similar experiments were completed in buffer solution containing 150 mM NH₄OAc (pH 6.7), 10% MeOH, and 0.1% HCOOH.

2.7.4 Thermal Difference Spectroscopy

Thermal difference (TD) data were collected in triplicate from samples (DNA or DNA+PNA) previously annealed in UV melting experiments. UV cuvettes containing the samples were scanned individually (200 nm/min) at 20 °C. Samples were left in the heat block while the temperature was adjusted to 90 °C and individual scans were repeated. TD spectra were produced by subtracting the absorbance spectrum at 20 °C from the absorbance spectrum at 90 °C for each sample.

2.7.5 Circular Dichroism

CD data were collected at 30 °C using samples (DNA or DNA+PNA) previously annealed in UV melting experiments. Each spectrum was obtained by averaging 6 scans collected at 100 nm/min.

2.7.6 Surface Plasmon Resonance

SPR analyses were performed on a Biacore T100 instrument (GE Healthcare) using four flow cell sensor chips. CM5 sensor chips were purchased from GE Healthcare and functionalized with streptavidin (approx. 5000 RU's) using NHS-EDC coupling. 5' biotinylated DNA ligands were individually immobilized (approx. 150 RU's) onto the streptavidin surface of separate flow cells to complete preparation of the sensor chips. Chips were prepared for PNA injection by priming 5x in the appropriate buffer. Direct binding experiments were completed in buffer containing 10 mM HEPES pH 7.4, 100 mM KCl or LiCl, 3 mM

EDTA, and 0.005% v/v polysorbate 20. PNA samples ($[P_{eg2}] = 5, 10, 25, 50, 75,$ and 100 nM) were injected randomly and in triplicate with a total surface contact time of 420 s. In order to determine complex stoichiometry, similar direct binding experiments were completed with higher concentrations of PNA ($[P_{eg2}] = 250, 500, 750,$ and 1000 nM) in the 100 mM KCl SPR buffer to saturate the response signal. Approximate PNA:DNA stoichiometries were then calculated using the following equation, where RU DNA_{immob} corresponds to the response units of immobilized DNA, RU_{max} PNA corresponds to the saturating response units obtained during PNA titration, MW is molecular weight and x is the number of PNA strands hybridized to each DNA strand:

$$\frac{\text{RU DNA}_{\text{immob}}}{\text{MW DNA}_{\text{immob}}} = \frac{\text{RU}_{\text{max PNA}}}{x(\text{MW PNA})}$$

2.7.7 Electrospray Ionization Mass Spectrometry (ESI-MS)

A buffer solution of 150 mM NH₄OAc (pH 6.7) was used for the ESI-MS analyses and prepared with doubly distilled H₂O and filtered with a 0.22 μM filter. PNA and DNA sequences were dissolved to concentrations of 100 μM each in 150 mM NH₄OAc buffer. The QFS3 DNA sequence was ordered through Integrated DNA Technologies (Coralville, IA) and HPLC-purified. Both QFS3 DNA and P_{eg2} PNA were heated to 65 °C, slowly cooled to 25 °C, and stored at 4 °C until use. PNA and DNA complexes were mixed at their respective concentration ratios and allowed to set for up to 24 hours at 4 °C. Up to 10% MeOH (v/v) was added per sample to promote solution-to-gas phase transition. Within the MeOH, a fixed percentage of the appropriate acid/base was also

included to help improve signal with 0.1% formic acid (v/v) for positive ion mode analyses and 0.5% NH₃ for negative ion mode analyses.

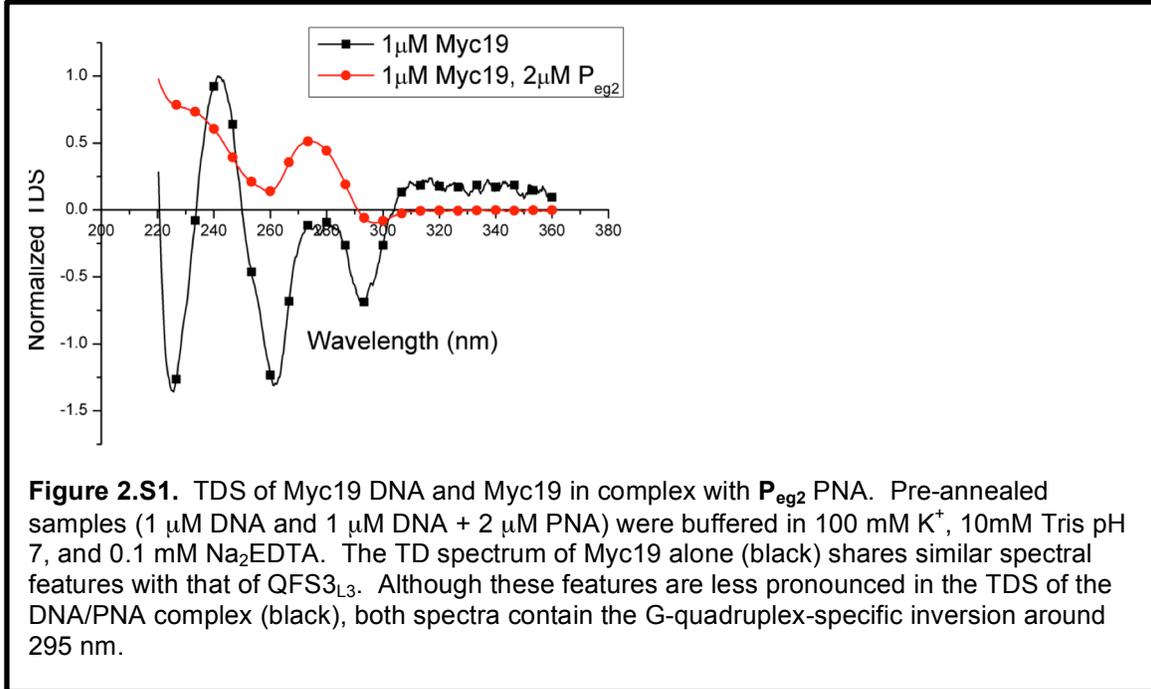
Samples were run using a Waters Micromass ESI-Q-TOF Mass Spectrometer (Milford, MA). DNA samples were scanned within a range of *m/z* 300 – 3,000 while PNA samples were scanned between *m/z* 100 – 2,000. PNA was initially tested by ESI-MS positive ion mode at a concentration of 20 µM and 8% MeOH (0.1% HCOOH). QFS3 DNA was tested by both ESI-MS (-) and (+) ion modes at a concentration of 5 µM with 10% MeOH (0.5% NH₃) and 10% MeOH (0.1% HCOOH), respectively. PNA-QFS3 was prepared at a ratio of [6:1] where concentration of PNA was 30 µM and QFS3 concentration equaled 5 µM in 10% MeOH (0.1% HCOOH). Optimal parameters used:

capillary voltage, 2400 V; sample cone voltage, 32 V; extraction voltage, 1.5 V; source temperature, 80 °C; desolvation temperature, 110 °C; collision energy, 2.0 V; cone voltage, 30 L/hr; desolvation gas, 450 L/hr.

2.8 Acknowledgement

The authors thank Prof. W. David Wilson and Sarah Laughlin-Toth of Georgia State University for acquiring mass spectrometry data and for helpful discussions and the David Scaife Family Charitable Foundation for financial support of this research (Award 141RA01).

2.9 Appendix



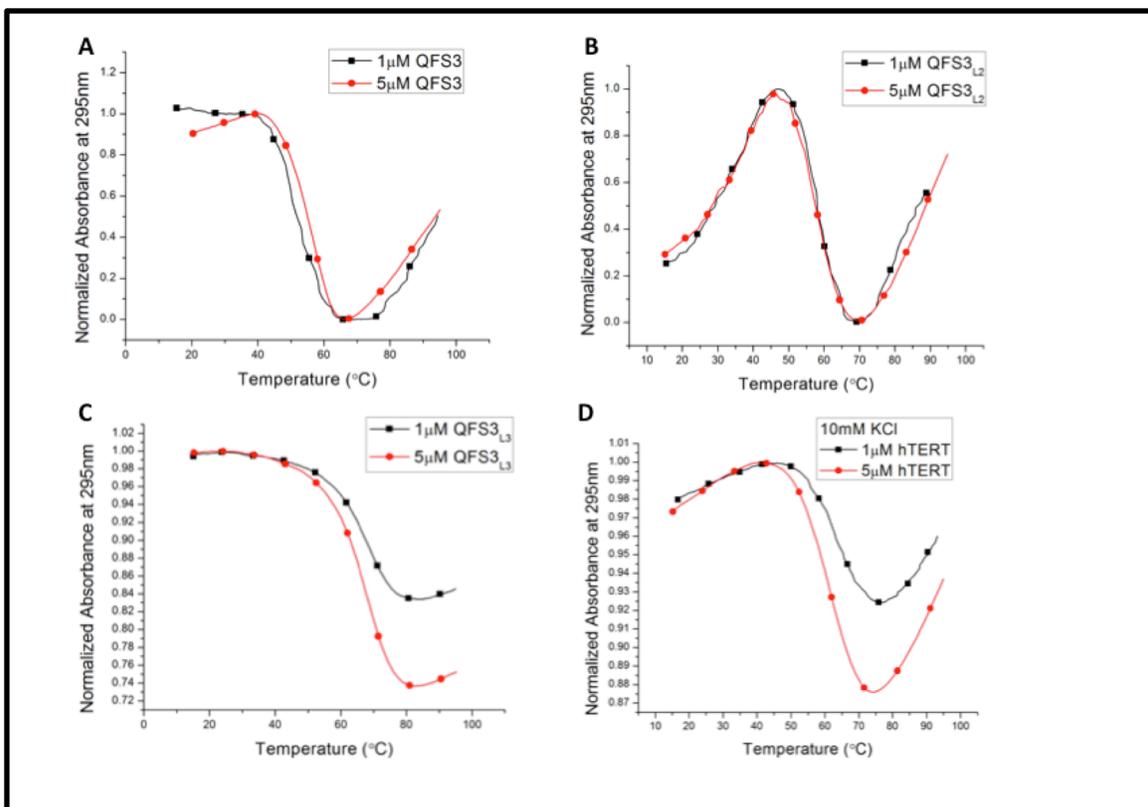


Figure 2.S2. DNA G-quadruplexes containing long loops fold into intramolecular structures *in vitro*. UV melting analyses (295 nm) indicate that the T_m of each quadruplex [(A) QFS3, (B) QFS3_{L2}, (C) QFS3_{L3}, (D) hTERT] is not dependent on DNA concentration. DNA samples were prepared in buffer containing 100 mM K^+ , except for hTERT, which required a lower ionic strength (10 mM K^+) to allow melting.

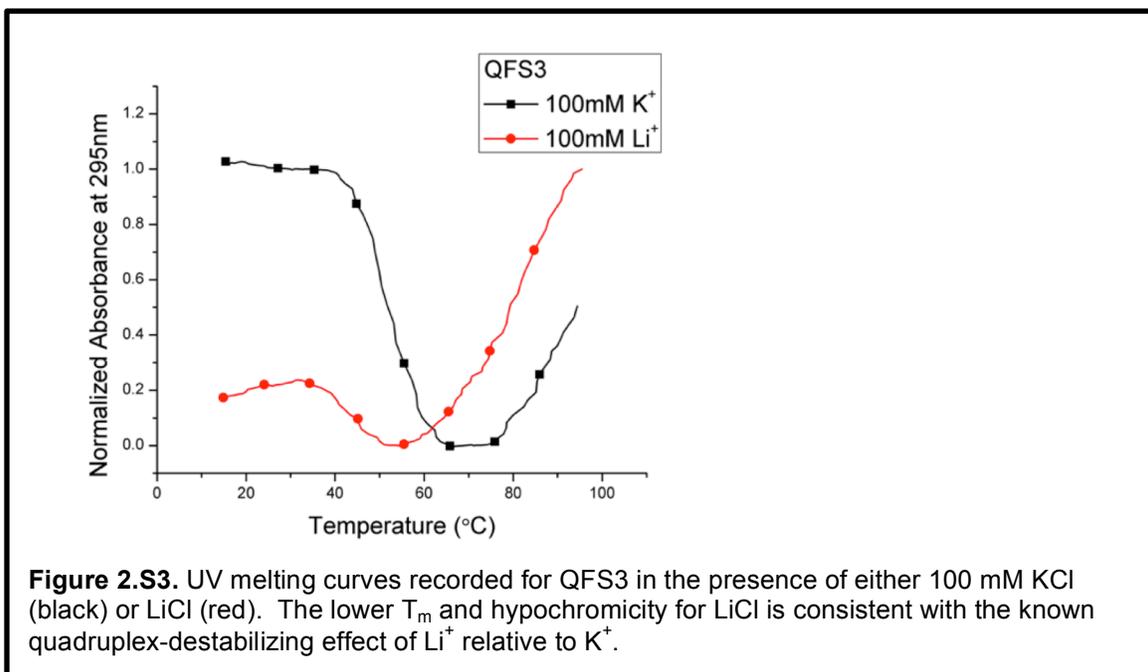


Figure 2.S3. UV melting curves recorded for QFS3 in the presence of either 100 mM KCl (black) or LiCl (red). The lower T_m and hypochromicity for LiCl is consistent with the known quadruplex-destabilizing effect of Li^+ relative to K^+ .

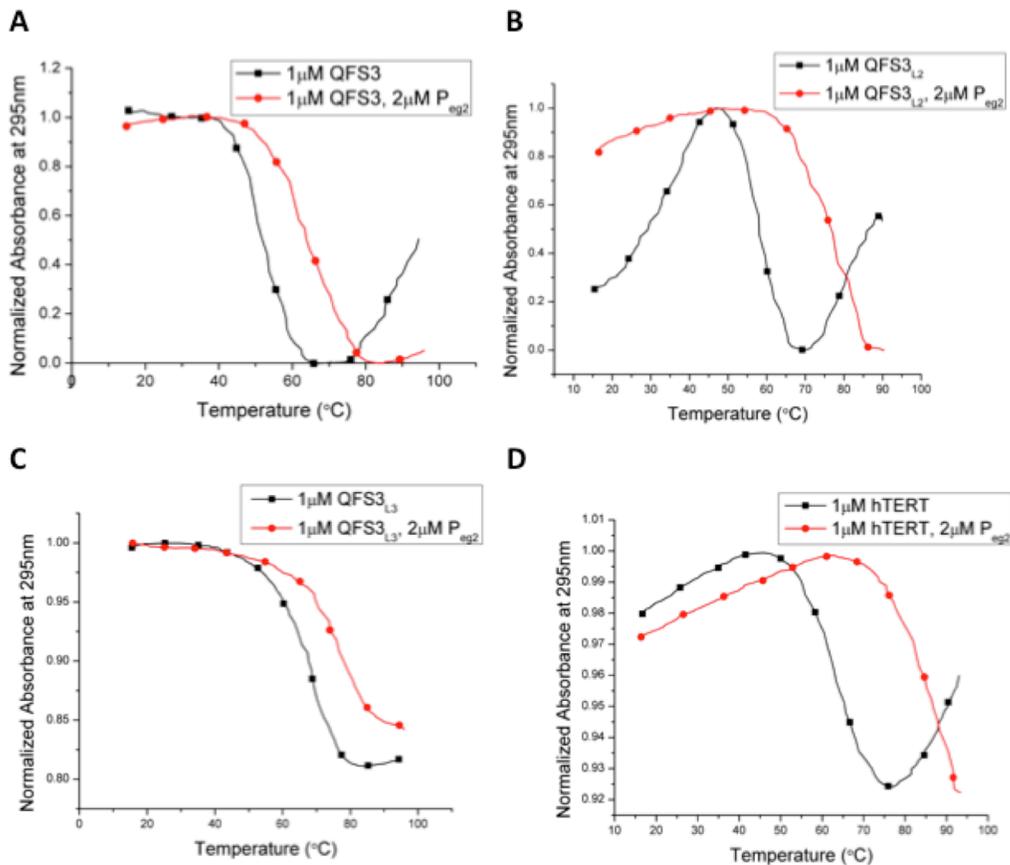


Figure 2.S4. P_{eg2} PNA enhances the thermal stability of each long-looped DNA G-quadruplex. UV melting analyses of G-quadruplex DNA with (red) and without (black) P_{eg2} at 295 nm. 1 μ M DNA [QFS3 (A), QFS3_{L2} (B), QFS3_{L3} (C), and hTERT (D)] was melted alone or in the presence of 2 μ M P_{eg2} . Samples were buffered in 10mM (hTERT) or 100 mM K^+ , 10 mM Tris (pH 7), and 0.1 mM Na_2EDTA .

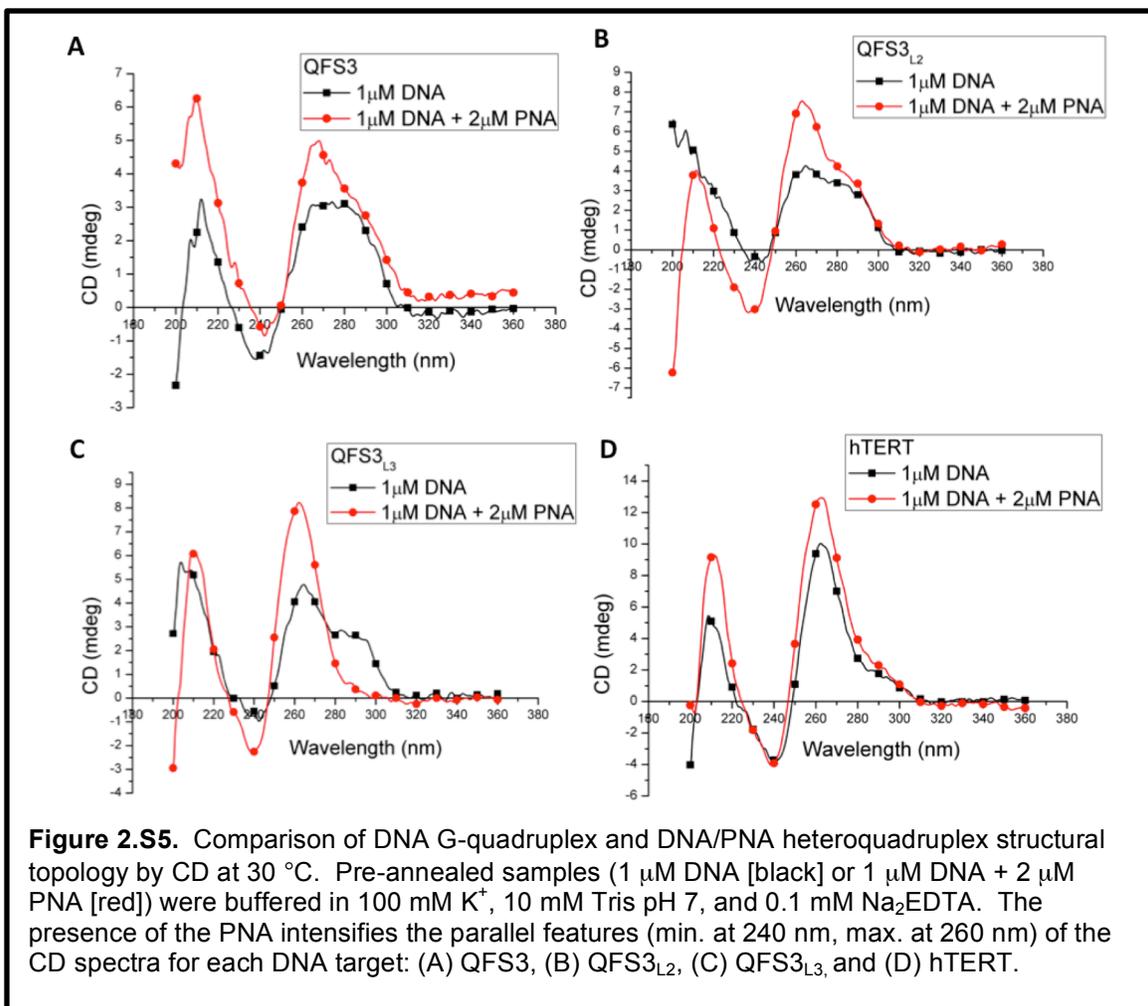
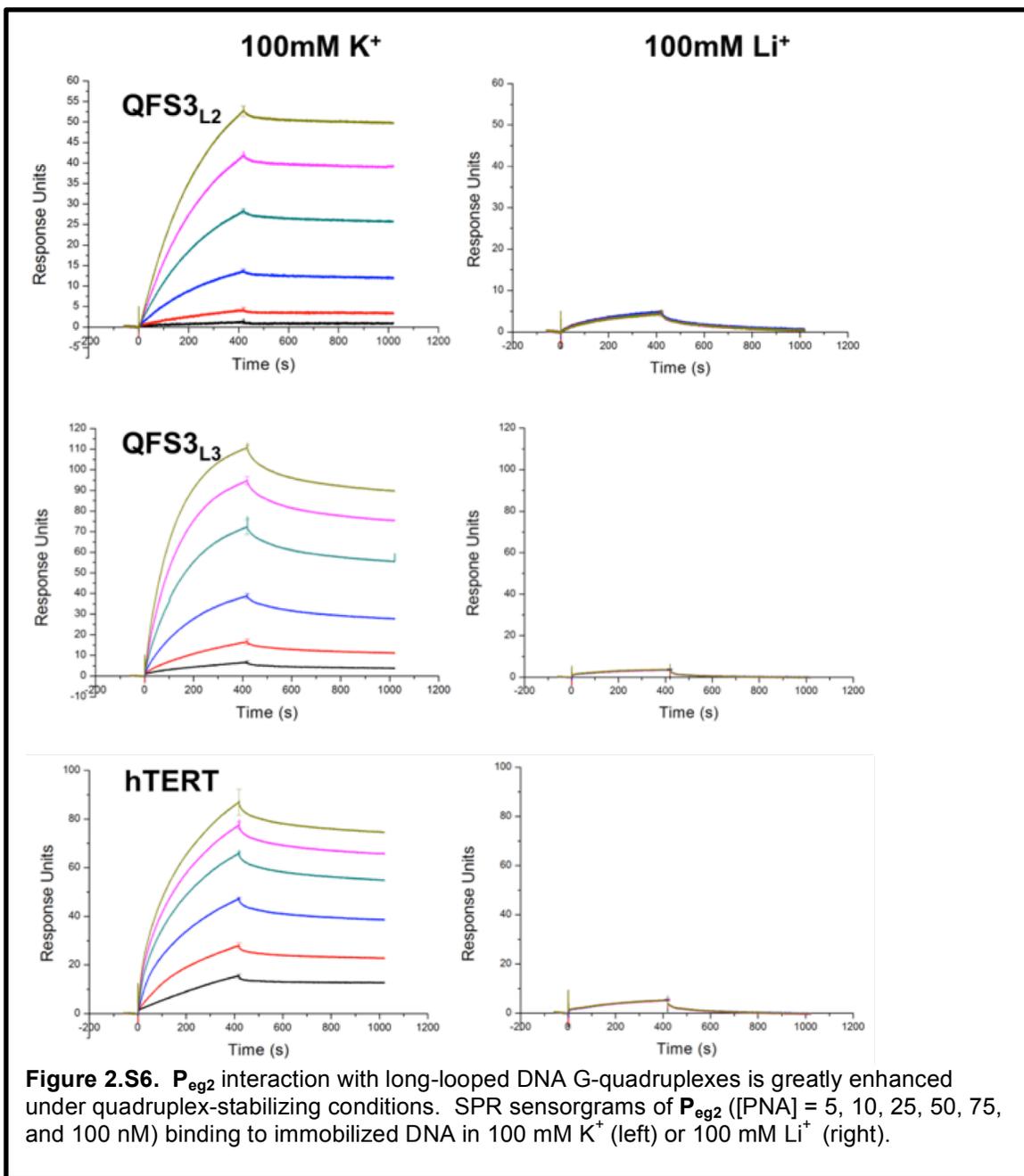


Figure 2.S5. Comparison of DNA G-quadruplex and DNA/PNA heteroquadruplex structural topology by CD at 30 °C. Pre-annealed samples (1 μM DNA [black] or 1 μM DNA + 2 μM PNA [red]) were buffered in 100 mM K⁺, 10 mM Tris pH 7, and 0.1 mM Na₂EDTA. The presence of the PNA intensifies the parallel features (min. at 240 nm, max. at 260 nm) of the CD spectra for each DNA target: (A) QFS3, (B) QFS3_{L2}, (C) QFS3_{L3}, and (D) hTERT.



To gain further information about these complexes, we used electrospray ionization mass spectrometry (ESI MS). High KCl concentrations are not compatible with ESI, so the buffer was changed to ammonium acetate with no added salt. UV melting, CD spectroscopy, and TDS experiments demonstrated that (a) QFS3 still folds into a quadruplex and (b) P_{eg2} forms a stable

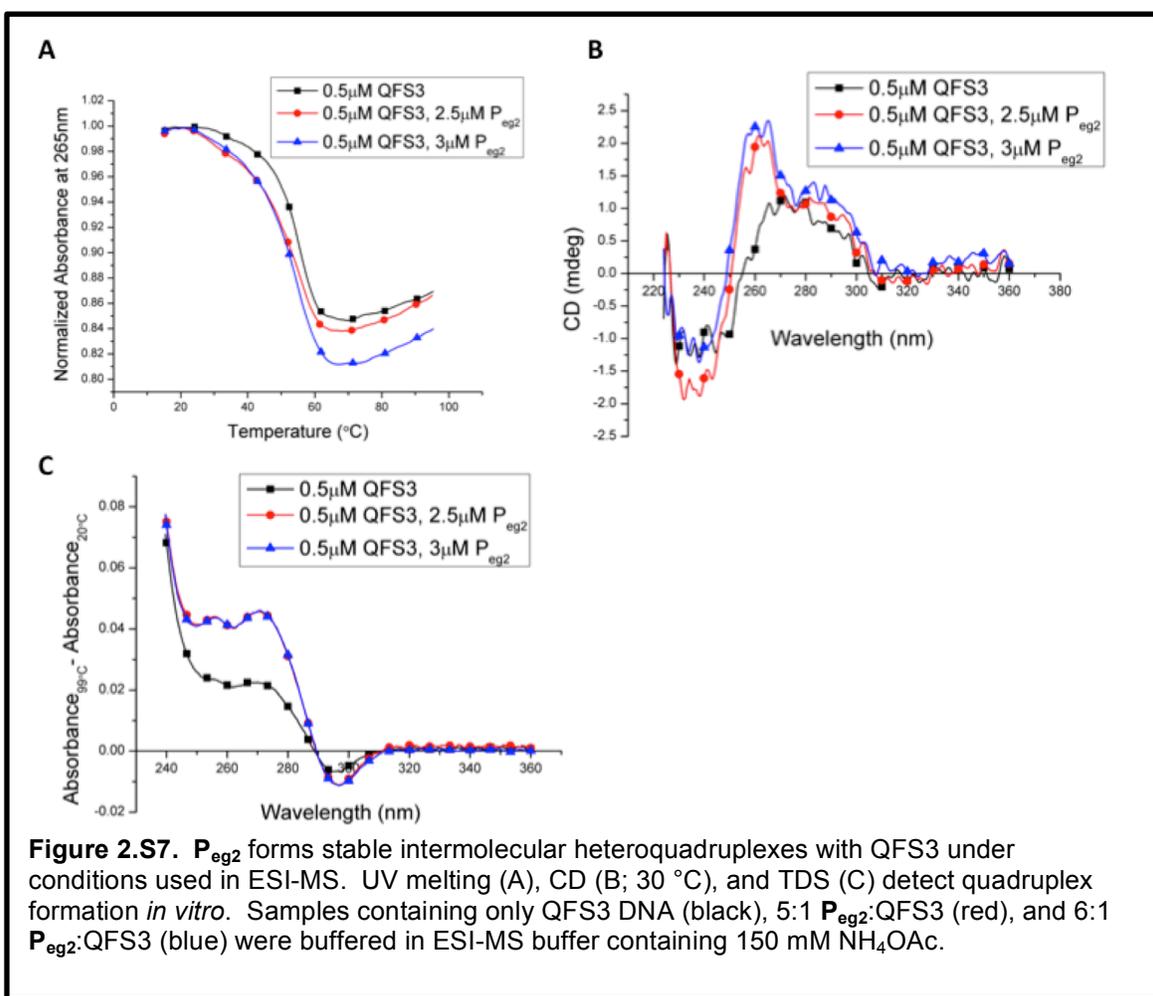
heteroquadruplex with QFS3 under these conditions (Table 4.S1; Figure 4.S7).

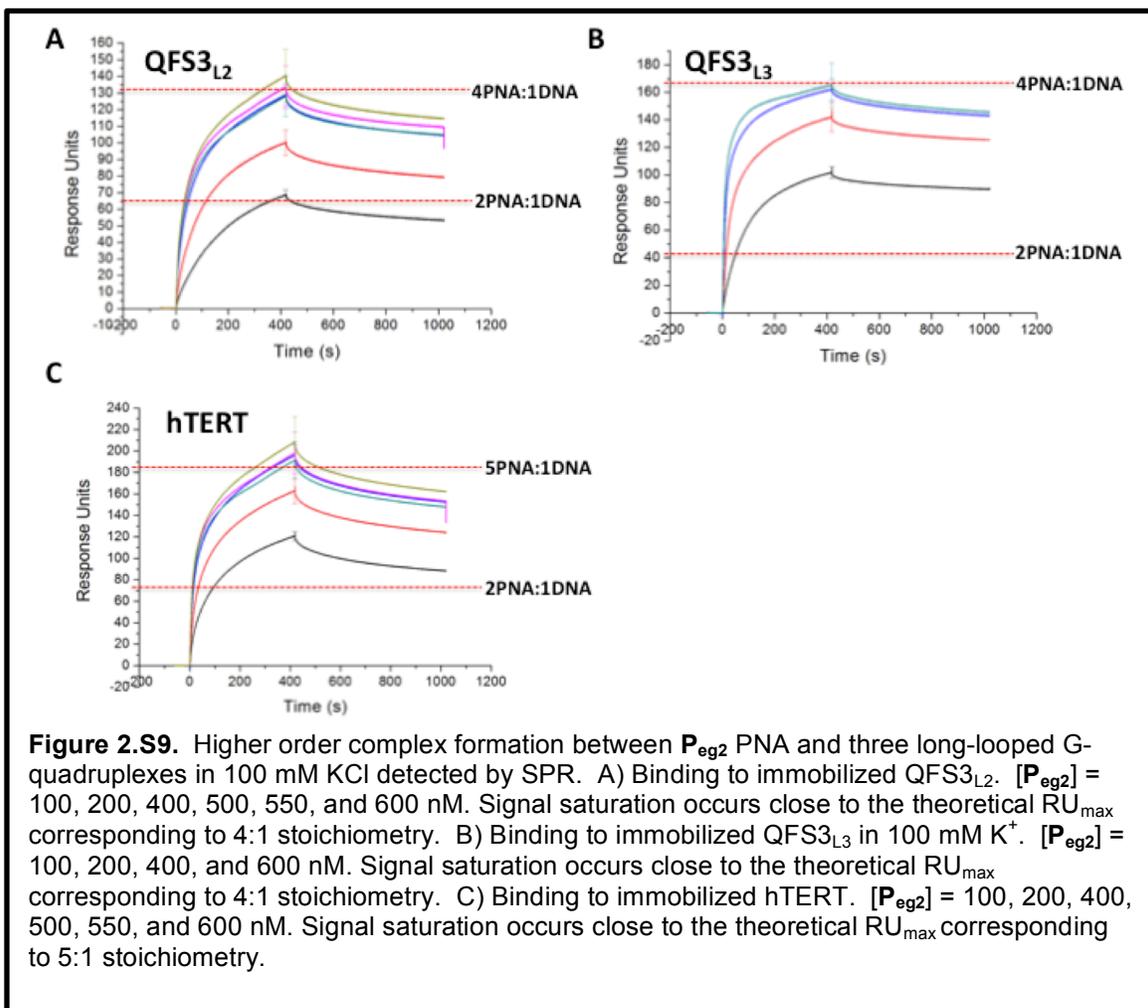
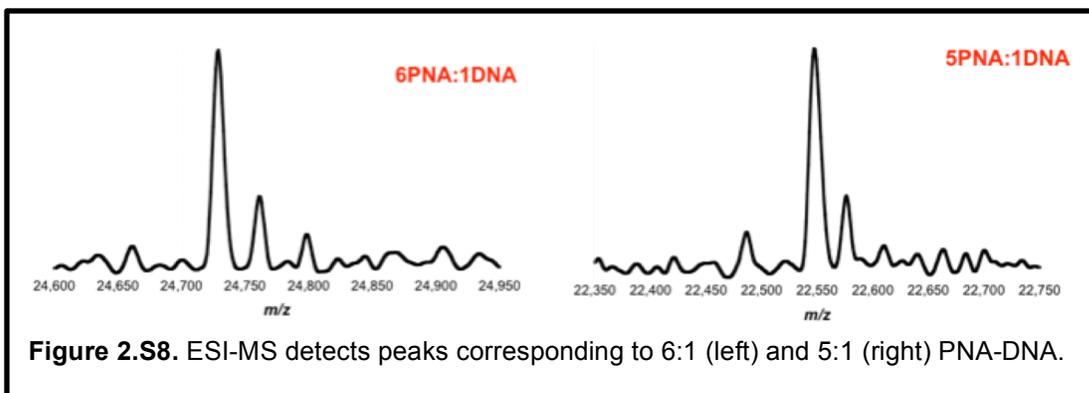
As shown in Figure 4.S8, both 5:1 and 6:1 P_{eg2} :QFS3 complexes are observed in the mass spectra.

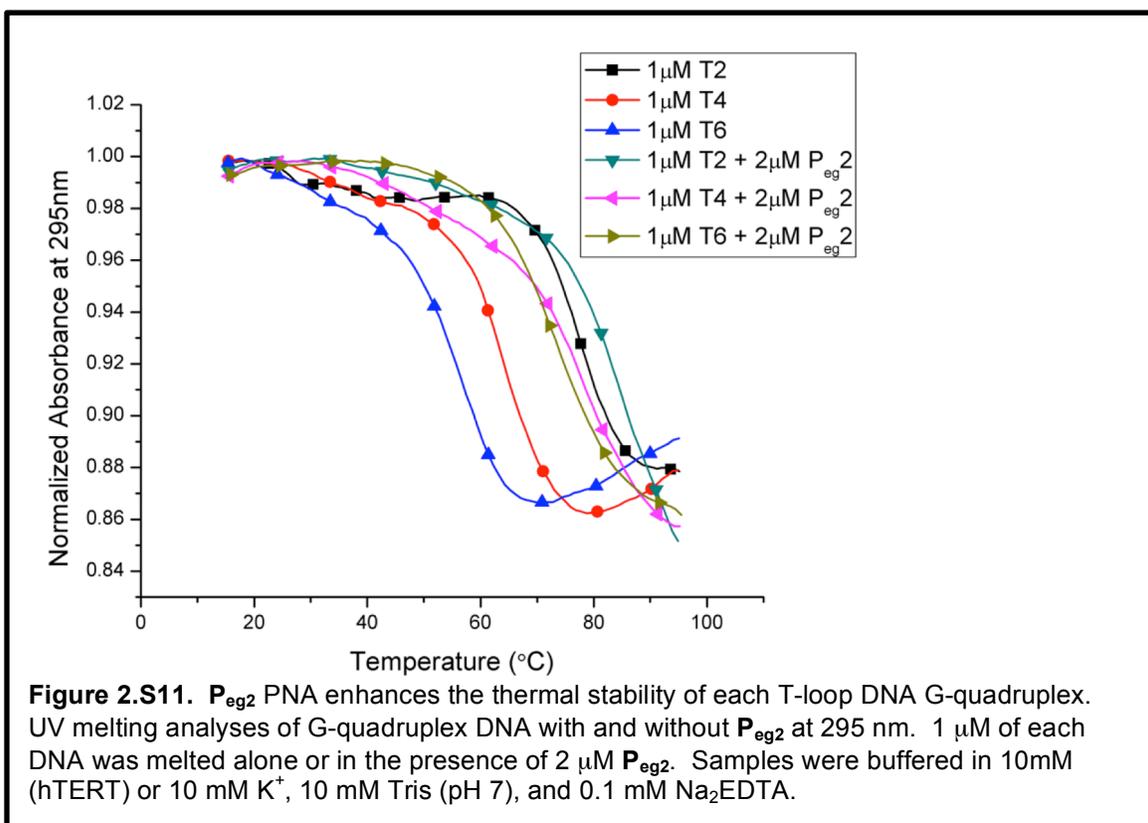
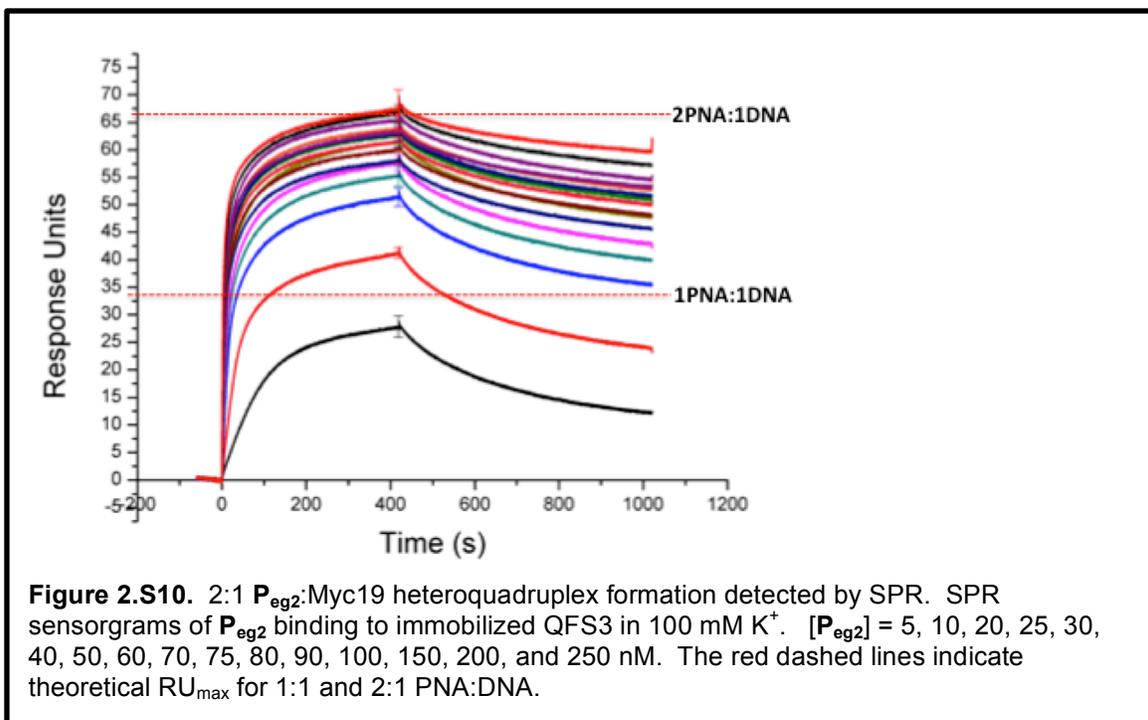
Table 2.S1. T_m values ($^{\circ}C$) of P_{eg2} /QFS3 Complexes in 150 mM NH_4^+ .

[DNA] μM	[PNA] μM	T_m
0.5	0	55.5 ± 0.4
0.5	2.5	$55.3 \pm 0.3^*$
0.5	3	55.1 ± 0.3

*n=2







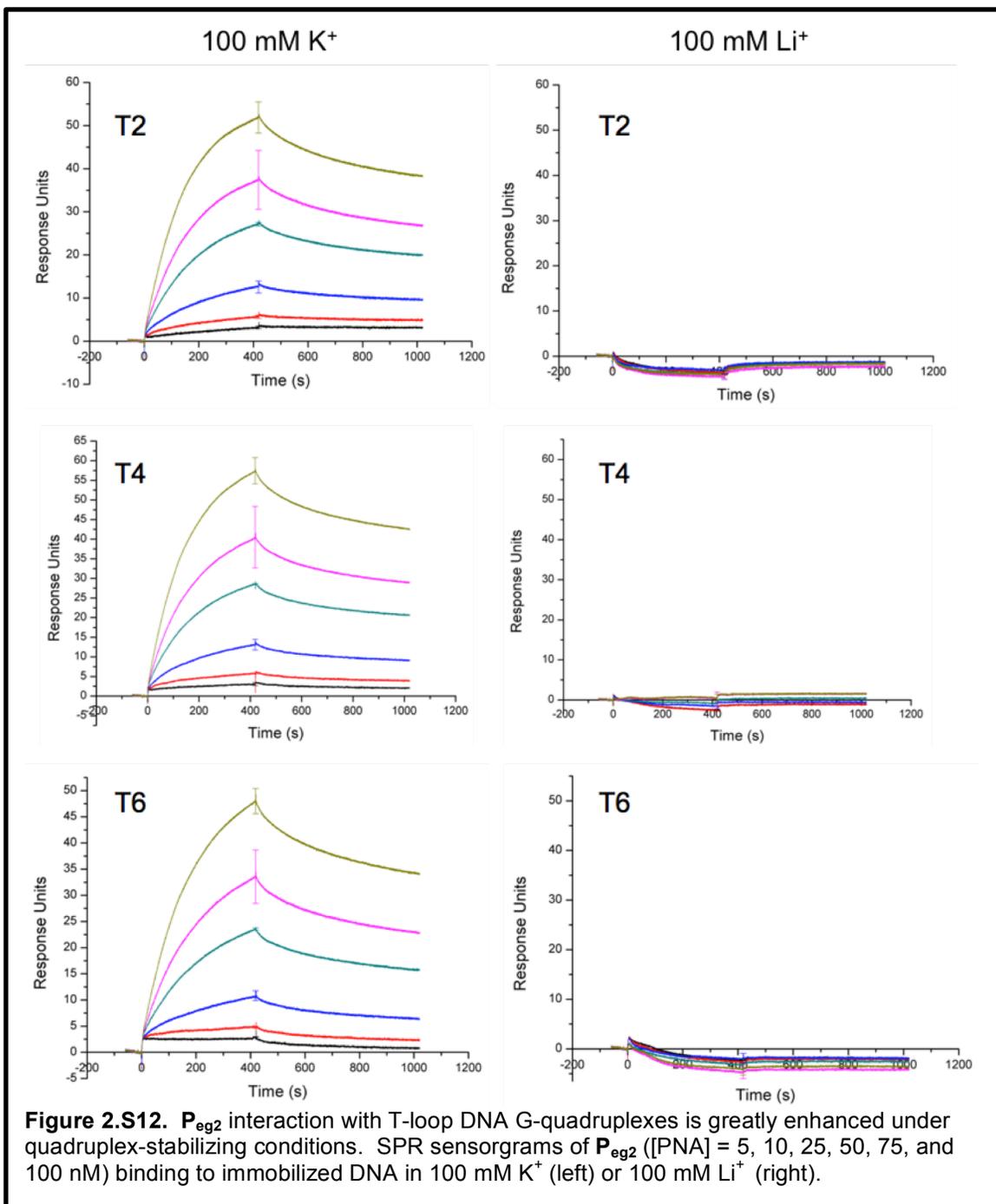


Figure 2.S12. P_{egg2} interaction with T-loop DNA G-quadruplexes is greatly enhanced under quadruplex-stabilizing conditions. SPR sensorgrams of P_{egg2} ([PNA] = 5, 10, 25, 50, 75, and 100 nM) binding to immobilized DNA in 100 mM K⁺ (left) or 100 mM Li⁺ (right).

2.10 References

1. Burge, S., et al., *Quadruplex DNA: sequence, topology and structure*. *Nucleic Acids Res*, 2006. **34**(19): p. 5402-15.
2. Patel, D.J., A.T. Phan, and V. Kuryavyi, *Human Telomere, Oncogenic Promoter and 5'-UTR G-Quadruplexes: Diverse Higher Order DNA and RNA Targets for Cancer Therapeutics*. *Nucleic Acids Research*, 2008. **35**: p. 7429-7455.

3. Balasubramanian, S. and S. Neidle, *G-Quadruplex Nucleic Acids as Therapeutic Targets*. *Current Opinion in Chemical Biology*, 2009. **13**: p. 345-353.
4. Lipps, H.J. and D. Rhodes, *G-Quadruplex Structures: In Vivo Evidence and Function*. *TRENDS in Cell Biology*, 2009. **19**: p. 414-422.
5. Patel, D.J., A.T. Phan, and V. Kuryavyi, *Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics*. *Nucleic Acids Res*, 2007. **35**(22): p. 7429-55.
6. Biffi, G., et al., *Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells*. *Nat Chem*, 2014. **6**(1): p. 75-80.
7. Chambers, V.S., et al., *High-throughput sequencing of DNA G-quadruplex structures in the human genome*. *Nat Biotechnol*, 2015.
8. Huppert, J.L. and S. Balasubramanian, *Prevalence of quadruplexes in the human genome*. *Nucleic Acids Res*, 2005. **33**(9): p. 2908-16.
9. Todd, A.K., M. Johnston, and S. Neidle, *Highly prevalent putative quadruplex sequence motifs in human DNA*. *Nucleic Acids Res*, 2005. **33**(9): p. 2901-7.
10. Hazel, P., et al., *Loop-length-dependent folding of G-quadruplexes*. *J Am Chem Soc*, 2004. **126**(50): p. 16405-15.
11. Guedin, A., et al., *How long is too long? Effects of loop size on G-quadruplex stability*. *Nucleic Acids Res*, 2010. **38**(21): p. 7858-68.
12. Mukundan, V.T. and A.T. Phan, *Bulges in G-quadruplexes: broadening the definition of G-quadruplex-forming sequences*. *J Am Chem Soc*, 2013. **135**(13): p. 5017-28.
13. Palumbo, S.L., S.W. Ebbinghaus, and L.H. Hurley, *Formation of a unique end-to-end stacked pair of G-quadruplexes in the hTERT core promoter with implications for inhibition of telomerase by G-quadruplex-interactive ligands*. *J. Am. Chem. Soc.*, 2009. **131**: p. 10878-10891.
14. Onel, B., et al., *A New G-Quadruplex with Hairpin Loop Immediately Upstream of the Human BCL2 P1 Promoter Modulates Transcription*. *J Am Chem Soc*, 2016. **138**(8): p. 2563-70.
15. Nielsen, P.E., et al., *Sequence-Selective Recognition of DNA by Strand Displacement with a Thymine-Substituted Polyamide*. *Science*, 1991. **254**(5037): p. 1497-1500.
16. Egholm, M., et al., *PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules*. *Nature*, 1993. **365**: p. 566-568.
17. Datta, B. and B.A. Armitage, *Hybridization of PNA to structured DNA targets: quadruplex invasion and the overhang effect*. *J. Am. Chem. Soc.*, 2001. **123**: p. 9612-9619.
18. Green, J.J., et al., *Kinetics of unfolding the human telomeric DNA quadruplex using a PNA trap*. *J. Am. Chem. Soc.*, 2003. **125**: p. 3763-3767.

19. Amato, J., et al., *Hybridization of short complementary PNAs to G-quadruplex forming oligonucleotides: An electrospray mass spectrometry study*. Biopolymers, 2009. **91**(4): p. 244-55.
20. Amato, J., et al., *Targeting G-quadruplex structure in the human c-Kit promoter with short PNA sequences*. Bioconjug Chem, 2011. **22**(4): p. 654-63.
21. Marin, V.L. and B.A. Armitage, *RNA guanine quadruplex invasion by complementary and homologous PNA probes*. J. Am. Chem. Soc., 2005. **127**: p. 8032-8033.
22. Marin, V.L. and B.A. Armitage, *Hybridization of complementary and homologous peptide nucleic acid oligomers to a guanine quadruplex-forming RNA*. Biochemistry, 2006. **45**(6): p. 1745-1754.
23. Gupta, A., et al., *Strand invasion of DNA quadruplexes by PNA: comparison of homologous and complementary hybridization*. ChemBiochem, 2013. **14**(12): p. 1476-84.
24. Murphy, C.T., et al., *Hybridization of G-quadruplex-forming peptide nucleic acids to guanine-rich DNA templates inhibits DNA polymerase ϵ extension*. Biochemistry, 2014. **53**(32): p. 5315-22.
25. Kumar, N., et al., *Silencing c-MYC Expression by Targeting Quadruplex in P1 Promoter Using Locked Nucleic Acid Trap*. Biochemistry, 2008. **47**(50): p. 13179-13188.
26. Kumar, N. and S. Maiti, *Role of locked nucleic acid modified complementary strand in quadruplex/Watson-Crick duplex equilibrium*. J. Phys. Chem. B, 2007. **111**: p. 12328-12337.
27. Rouleau, S.G., et al., *Small antisense oligonucleotides against G-quadruplexes: specific mRNA translational switches*. Nucleic Acids Res, 2015. **43**(1): p. 595-606.
28. Murat, P., et al., *G-quadruplexes regulate Epstein-Barr virus-encoded nuclear antigen 1 mRNA translation*. Nat Chem Biol, 2014. **10**(5): p. 358-64.
29. Datta, B., C. Schmitt, and B.A. Armitage, *Formation of a PNA2-DNA2 Hybrid Quadruplex*. J. Am. Chem. Soc., 2003. **125**: p. 4111-4118.
30. Roy, S., et al., *High-affinity homologous peptide nucleic acid probes for targeting a quadruplex-forming sequence from a MYC promoter element*. Biochemistry, 2007. **46**: p. 10433-10443.
31. Roy, S., et al., *Kinetic discrimination in recognition of DNA quadruplex targets by guanine-rich heteroquadruplex-forming PNA probes*. Chem Commun, 2011. **47**(30): p. 8524-6.
32. Lusvarghi, S., et al., *Loop and backbone modifications of peptide nucleic acid improve g-quadruplex binding selectivity*. J Am Chem Soc, 2009. **131**(51): p. 18415-24.
33. Gaynutdinov, T.I., et al., *G-quadruplex formation between G-rich PNA and homologous sequences in oligonucleotides and supercoiled plasmid DNA*. Nucleic Acid Ther, 2015. **25**(2): p. 78-84.

34. Usui, K., et al., *Control of guanine-rich DNA secondary structures depending on the protease activity using a designed PNA peptide*. *Org Biomol Chem*, 2015. **13**(7): p. 2022-5.
35. Ishizuka, T., et al., *G-rich sequence-specific recognition and scission of human genome by PNA/DNA hybrid G-quadruplex formation*. *Angew Chem Int Ed Engl*, 2012. **51**(29): p. 7198-202.
36. Englund, E.A., et al., *PNA-DNA duplexes, triplexes, and quadruplexes are stabilized with trans-cyclopentane units*. *J. Am. Chem. Soc.*, 2006. **128**: p. 16456-16457.
37. Jodoin, R., et al., *The folding of 5'-UTR human G-quadruplexes possessing a long central loop*. *RNA*, 2014. **20**(7): p. 1129-41.
38. Hershman, S.G., et al., *Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in *Saccharomyces cerevisiae**. *Nucleic Acids Res*, 2008. **36**(1): p. 144-56.
39. Siddiqui-Jain, A., et al., *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): p. 11593-8.
40. Simonsson, T., P. Pecinka, and M. Kubista, *DNA tetraplex formation in the control region of c-myc*. *Nucleic Acids Res*, 1998. **26**(5): p. 1167-72.
41. Mergny, J.-L., A.-T. Phan, and L. Lacroix, *Following G-Quartet Formation by UV-Spectroscopy*. *FEBS Letters*, 1998. **435**: p. 74-78.
42. Guédin, A., et al., *How Long is Too Long? Effects of Loop Size on G-Quadruplex Stability*. *Nucleic Acids Research*, 2010. **38**: p. 7858-7868.
43. Kwok, C.K., M.E. Sherlock, and P.C. Bevilacqua, *Effect of Loop Sequence and Loop Length on the Intrinsic Fluorescence of G-Quadruplexes*. *Biochemistry*, 2013. **52**: p. 3019-3021.
44. Williamson, J.R., M.K. Raghuraman, and T.R. Cech, *Monovalent Cation-Induced Structure of Telomeric DNA: The G-Quartet Model*. *Cell*, 1989. **59**: p. 871-880.
45. Vorlícková, M., et al., *Circular Dichroism and Guanine Quadruplexes*. *Methods*, 2012. **57**: p. 64-75.
46. Mergny, J.L., et al., *Thermal difference spectra: a specific signature for nucleic acid structures*. *Nucleic Acids Res*, 2005. **33**(16): p. e138.
47. Hanakahi, L.A., H. Sun, and N. Maizels, *High affinity interactions of nucleolin with G-G-paired rDNA*. *J Biol Chem*, 1999. **274**(22): p. 15908-12.
48. Ben-Shem, A., et al., *Crystal structure of the eukaryotic ribosome*. *Science*, 2010. **330**(6008): p. 1203-9.
49. Wanrooij, P.H., et al., *G-quadruplex structures in RNA stimulate mitochondrial transcription termination and primer formation*. *Proc Natl Acad Sci U S A*, 2010. **107**(37): p. 16072-7.
50. Tothova, P., P. Krafcikova, and V. Viglasky, *Formation of highly ordered multimers in G-quadruplexes*. *Biochemistry*, 2014. **53**(45): p. 7013-27.
51. Petraccone, L., et al., *Structure and stability of higher-order human telomeric quadruplexes*. *J Am Chem Soc*, 2011. **133**(51): p. 20951-61.

52. Yu, H.Q., D. Miyoshi, and N. Sugimoto, *Characterization of structure and stability of long telomeric DNA G-quadruplexes*. J Am Chem Soc, 2006. **128**(48): p. 15461-8.
53. Yu, H., et al., *Beads-on-a-string structure of long telomeric DNAs under molecular crowding conditions*. J Am Chem Soc, 2012. **134**(49): p. 20060-9.
54. Tippana, R., W. Xiao, and S. Myong, *G-quadruplex conformation and dynamics are determined by loop length and sequence*. Nucleic Acids Res, 2014. **42**(12): p. 8106-14.
55. Fujimoto, T., et al., *Thermodynamics-hydration relationships within loops that affect G-quadruplexes under molecular crowding conditions*. J Phys Chem B, 2013. **117**(4): p. 963-72.
56. Bhattacharyya, D., K. Nguyen, and S. Basu, *Rationally Induced RNA:DNA G-Quadruplex Structures Elicit an Anticancer Effect by Inhibiting Endogenous eIF-4E Expression*. Biochemistry, 2014. **53**: p. 5461-5470.
57. Phan, A.T., et al., *Structure of an Unprecedented G-Quadruplex Scaffold in the Human c-kit Promoter*. J. Am. Chem. Soc., 2007. **129**: p. 4386-4392.
58. Sahu, B., et al., *Synthesis and characterization of conformationally preorganized, (R)-diethylene glycol-containing gamma-peptide nucleic acids with superior hybridization properties and water solubility*. J Org Chem, 2011. **76**(14): p. 5614-27.
59. Yeh, J.I., et al., *Crystal Structure of Chiral γ PNA with Complementary DNA Strand: Insights into the Stability and Specificity of Recognition and Conformational Preorganization*. J. Am. Chem. Soc., 2010. **132**: p. 10717-10727.
60. Monchaud, D. and M.P. Teulade-Fichou, *A hitchhiker's guide to G-quadruplex ligands*. Org Biomol Chem, 2008. **6**(4): p. 627-36.
61. Englund, E.A. and D.H. Appella, *Gamma-substituted peptide nucleic acids constructed from L-lysine are a versatile scaffold for multifunctional display*. Angew Chem Int Ed Engl, 2007. **46**(9): p. 1414-8.
62. Chen, Y. and D. Yang, *Sequence, stability, and structure of G-quadruplexes and their interactions with drugs*. Curr Protoc Nucleic Acid Chem, 2012. **Chapter 17**: p. Unit17 5.
63. Christensen, L., et al., *Solid-Phase Synthesis of Peptide Nucleic Acids*. Journal of Peptide Science, 1995. **3**: p. 175-183.
64. Koch, T., *PNA Synthesis by Boc Chemistry*, in *Peptide Nucleic Acids: Protocols and Applications*, P.E. Nielsen, Editor. 2004, Horizon Bioscience: Norfolk. p. 37-60.

3. Chapter 3: Molecular Crowding Enhances γ -modified and Unmodified Homologous PNA Hybridization to Discrete DNA G-quadruplex Targets

3.1 Chapter Summary

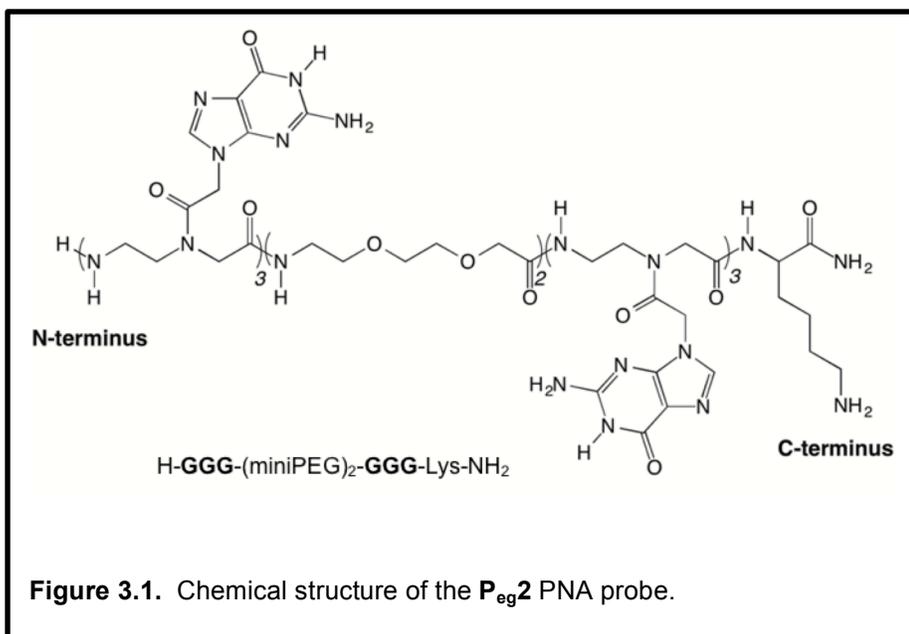
The growth rate of identified G-quadruplex functions in biological systems suggests that these structures are nearly omnipotent with respect to their ability to regulate cellular processes involving nucleic acids. The vast number and variety of biological G-quadruplexes within DNA and RNA presents a huge challenge to investigation of these structures, as well as synthetic regulation, requiring the development of selective and potent quadruplex-binding ligands. The ability of G-rich, homologous peptide nucleic acids to invade and hybridize to DNA and RNA G-quadruplexes has been well established *in vitro*. Further advances in PNA chemistry have expanded the variety of molecules that can be applied to these functions. However, the transition from *in vitro* to *in vivo* studies may not always be direct partly due to the complexity of the intracellular environment. As an expansion of previous studies in our group, we have explored the interaction of homologous γ -modified and unmodified PNAs with several DNA quadruplex targets under molecular crowding conditions. Our results suggest that partially and fully (*R*)-diethylene glycol γ -modified PNAs are compatible with DNA/PNA heteroquadruplex formation, and expand upon our previous understanding of molecular crowding-induced enhancement of heteroquadruplex stability. Together, these findings reinforce the applicability of homologous PNA-based probes in biological systems.

3.2 Introduction

The concept of homologous PNA hybridization to G-quadruplex DNA was initiated in the Armitage lab 15 years ago[1], and has since continued to be developed with a number of biologically interesting DNA and RNA quadruplexes[2-4]. Simultaneous advances in PNA chemistry have expanded the toolkit of synthetic molecules available for use with this strategy. A broad goal of the G-quadruplex field is the development of mechanisms for selective and effective hybridization to modulate G-quadruplex function. Therefore, exploration of G-quadruplex/probe interactions in controlled settings designed to reflect biological systems is also necessary to advance the development of these tools.

A goal of this work has been to investigate thermodynamic properties of quadruplex invasion using a G-rich, homologous PNA introduced in Chapter 2, **P_{eg}2** (Figure 3.1), and its derivatives (Table 3.1). The simplicity of this PNA oligomer makes it ideal for the study of Hoogsteen-driven DNA/PNA interactions. Specifically, **P_{eg}2** is designed to form 3-tetrad quadruplexes using its two G₃-tracts. Additionally, two diethylene glycol (miniPEG) linker units have been incorporated between the G-tracts to add flexibility to the molecule and reduce the chance of duplex formation with its targets[5]. Modifications at the γ -positioned carbon of the PNA monomer have been developed that confer specific properties when incorporated into an oligomer, which have the potential to affect quadruplex binding. Initial γ PNA iterations incorporated alanine-like monomers having a methyl group conjugated to the γ carbon[5]. Importantly, γ modifications

introduce chirality to the molecule, which, depending on their stereochemistry, can pre-organize the γ PNA oligomer into a right- or left-handed helix[5, 6]. Although a right-handed helix is ideal for duplex formation with natural nucleic acids, these γ PNAs are also sufficiently flexible for G-quadruplex invasion[5]. More recently, alanine-based modifications have been replaced with hydrophilic (*R*)-diethylene glycol units, which similarly pre-organize the PNA into a right-handed helix but also have the added benefit of enhanced water solubility and reduced propensity for aggregation[7]. In this study, we have expanded upon our understanding of P_{eg2} hybridization to G-quadruplex DNA using both fully ($\gamma 6P_{eg2}$) and partially ($\gamma 2P_{eg2}$) modified derivatives of this probe (Table 3.1).



The interior of a cell is a highly complex and fluid environment, and contributing factors such as pH, salt content, and molecular crowding have the potential to influence intra- and intermolecular interactions. Therefore, molecules designed for nucleic acid hybridization in biological contexts must be studied in

conditions that mimic the intracellular environment. Living cells are extraordinarily crowded, with 30-40% of the intracellular volume occupied with biomolecules[8, 9]. This biological molecular crowding greatly influences intracellular components, including the structural conformation and stability of DNA (reviewed in Myoshi and Sugimoto, 2008)[10].

G-quadruplex conformational and thermodynamic properties are particularly susceptible to molecular crowding, which has been studied in great detail *in vitro*. The human telomeric DNA quadruplex (from which the hTelo22 oligonucleotide is derived) switches from a hybrid conformation to a parallel structure in the presence of K^+ and a polyethylene glycol (PEG 200) crowding agent, suggesting that the *in vivo* conformation of this quadruplex may be different from the structure determined in dilute solution[11]. Similarly, the $G_4T_4G_4$ quadruplex-forming sequence undergoes a conversion from antiparallel to parallel topology in the presence of PEG 200[12]. However, some controversy surrounds the effect of PEG 200, specifically in the context of human telomeric quadruplex DNA. Imposition of osmotic stress conditions has been found to lead to structural conversion to the parallel conformation in a manner similar to that detected previously with PEG 200[13]. In this study, the authors suggested that PEG is not sufficient to affect quadruplex structure by osmotic stress, but rather likely interacts with the propeller loops present in the parallel conformation[13]. Further, they proposed that the parallel, propeller conformation of the quadruplex should not be supported by molecular crowding because it was determined to be less compact than the hybrid structure by sedimentation velocity

experiments[13]. This work calls into question the physiological telomeric DNA quadruplex structure, as well as the utility of PEG 200 as a molecular crowding agent. However, the documented influence of PEG 200 on DNA G-quadruplex structure and stability suggests that this molecule can still be useful for approximation of intracellular conditions.

G-quadruplexes tend to also be stabilized by molecular crowding. The reduction of water volume caused by molecular crowding agents enhances the effect of dehydration on G-quadruplex folding[10]. Viscous co-solutes having vicinal hydroxyl groups, such as glycerol, are able to mimic hydration of the DNA and, therefore, do not influence structure and stability to the same degree as molecular crowding agents like PEG 200[10, 14]. *In silico* molecular dynamics simulations have provided explanations for this phenomenon using ethanol as a crowding agent[15]. Ethanol alone cannot support quadruplex formation without G-tetrad stabilizing cations but the presence of this co-solute can increase G-quadruplex stability by enhancing stacking interactions between tetrads, as well as H-bonds within the G-tetrads through the exclusion of water volume and decreased water activity[15]. The reduction of water volume also can increase quadruplex affinity for stabilizing cations present in solution[15].

In vitro studies of crowding effects on isolated quadruplexes are important, but biological quadruplex motifs are generally housed within the context of longer nucleic acid polymers, and, in the case of DNA, have to compete with duplex interactions with the complementary strand. Work by Tan and colleagues concerning the effects of PEG 200 on previously identified quadruplexes within

long double-stranded DNAs has set the precedent for crowding-induced stabilization of these structures within a more biologically relevant context[16]. Specifically, structure probing and gel electrophoresis experiments confirmed that the presence of PEG 200-promoted stabilization of G-quadruplexes formed after melting of the DNA duplex by either physical (heat) or biological (RNA transcription) mechanisms, even in the presence of the competing, complementary DNA strand[16]. These results support the hypothesis that G-quadruplex structures can be maintained *in vivo* following slight melting of dsDNA.

Apart from the effects on individual G-quadruplexes, molecular crowding can also influence the interactions between these structures and quadruplex-targeting probes. An investigation of ligand binding to human telomeric DNA quadruplexes in the presence of PEG 200 demonstrated that stabilization of these structures by a series of small molecules (TmPyP4, BMVC, and Hoechst 33258) was reduced in the molecular crowding condition[17]. The ligands were also found to be much less effective inhibitors of telomerase activity in the crowding condition[17]. The authors suggested that the decreased affinity between the quadruplex and ligands was due to the reduction of water activity and increase in the viscosity of the medium[17]. This work demonstrates that the effects of quadruplex-binding ligands determined in dilute solutions can be misleading. Therefore, exploration of probe/target interactions is important to consider prior to beginning assays in more complex, and sometimes costly *in vivo* environments.

Preliminary studies previously conducted in our lab have indicated that homologous PNAs do not experience the same decrease in affinity to DNA G-quadruplexes under molecular crowding conditions[18]. Specifically, crowding studies with a series of homologous PNAs (including **P_{eg2}**) in complex with either Myc19 or hTelo22 confirmed that DNA/PNA heteroquadruplex thermal stability increases with increasing PEG 200 content (up to 40% w/v)[18]. However, analogous experiments with the viscous co-solvent, glycerol, did not produce the same degree of heteroquadruplex stabilization, suggesting that molecular crowding-induced stability is due to the increase in PNA and DNA concentrations due to the exclusion of water volume[18]. These initial studies were important to confirm that DNA/PNA interactions are compatible with molecular crowding and therefore, should not be adversely affected by the crowded intracellular environment. We have expanded upon these earlier studies with additional PNA and DNA oligomers.

Here we describe the effect of molecular crowding on hybridization of both γ -modified and unmodified homologous PNAs to several DNA G-quadruplex targets that vary in both primary and secondary structure. Consistent with previous work in our lab, *in vitro* molecular crowding enhances the thermal stability of these DNA/PNA heteroquadruplexes, providing support for the potential utility of similar PNA probes in biological systems.

3.3 Results

Heteroquadruplex formation between homologous PNAs and nucleic acids is an established mechanism for quadruplex-based probe/target hybridization. The

availability of modifications afford PNAs with various physical and chemical properties that can impact these interactions. Further, to truly exploit the full potential of PNAs in biological systems, we must understand how these molecules interact in conditions that mimic the intracellular environment. Here we have explored the effect of molecular crowding on homologous PNA/DNA hybridization for three DNA quadruplex targets: Myc19, hTelo22, and QFS3 (Table 3.1). Previous work in the Armitage group has explored the thermodynamic and kinetic components of Myc19 and hTelo22 interaction with homologous PNA probes in great detail[2, 5, 19-21]. We have expanded upon these studies using diethylene glycol γ -modified PNAs in conditions that mimic the crowded intracellular environment (up to 40% PEG 200), in addition to investigating the effects of molecular crowding on the QFS3/**P_{eg}2** heteroquadruplex introduced in Chapter 2[22].

Table 3.1. Homologous PNA probes and DNA G-quadruplex targets. (G-tracts predicted or reported to participate in tetrad formation are shown in bold and underlined.)

PNA (N-to-C)	
P_{eg}2	H- <u>GGG</u> -miniPEG-miniPEG- <u>GGG</u> -Lys-NH ₂
γ6P_{eg}2	H- γ <u>GγGγG</u> -miniPEG-miniPEG- γ <u>GγGγG</u> -Lys-NH ₂
γ2P_{eg}2	H- <u>GγGG</u> -miniPEG-miniPEG- <u>GγGG</u> -Lys-NH ₂
DNA (5'-to-3')	
Myc19	<u>AGGGTCGGGGAGGGTGGGGA</u>
hTelo22	<u>AGGGTTAGGGTTAGGGTTAGGG</u>
QFS3	<u>GGGTCGGGTAGTGAGGGCCTTGGTCAGACGCAGCGGG</u>

Molecular crowding γ -modified PNAs in complex with a parallel DNA G-quadruplex target. Myc19 is a canonical 3-tetrad quadruplex-forming oligomer that exists as a regulatory element in the *c-MYC* gene promoter (discussed in

Chapter 2)[23, 24]. UV melting analyses at 295 nm confirm stable heteroquadruplex formation between the fully modified $\gamma 6P_{eg}2$ and Myc19, which is enhanced with increasing PEG 200 concentration (Table 3.2; Figure 3.2A). To allow for complete melting of the heteroquadruplexes, these experiments were done in reduced salt conditions (1 or 10 mM KCl), which suggests that these complexes should be significantly more stable at physiological K^+ concentrations (approx. 100 mM K^+). Further support that the additional stability conferred by PEG 200 is due to molecular crowding is that inclusion of the viscous co-solvent, glycerol, does not produce a similar effect (Table 3.2; Figure 3.2A). Similarly, the partially modified $\gamma 2P_{eg}2$ invades Myc19 to form a heteroquadruplex with even higher stability than that with $\gamma 6P_{eg}2$, which is strengthened in the presence of PEG 200, but not glycerol (Table 3.2, Figure 3.2B).

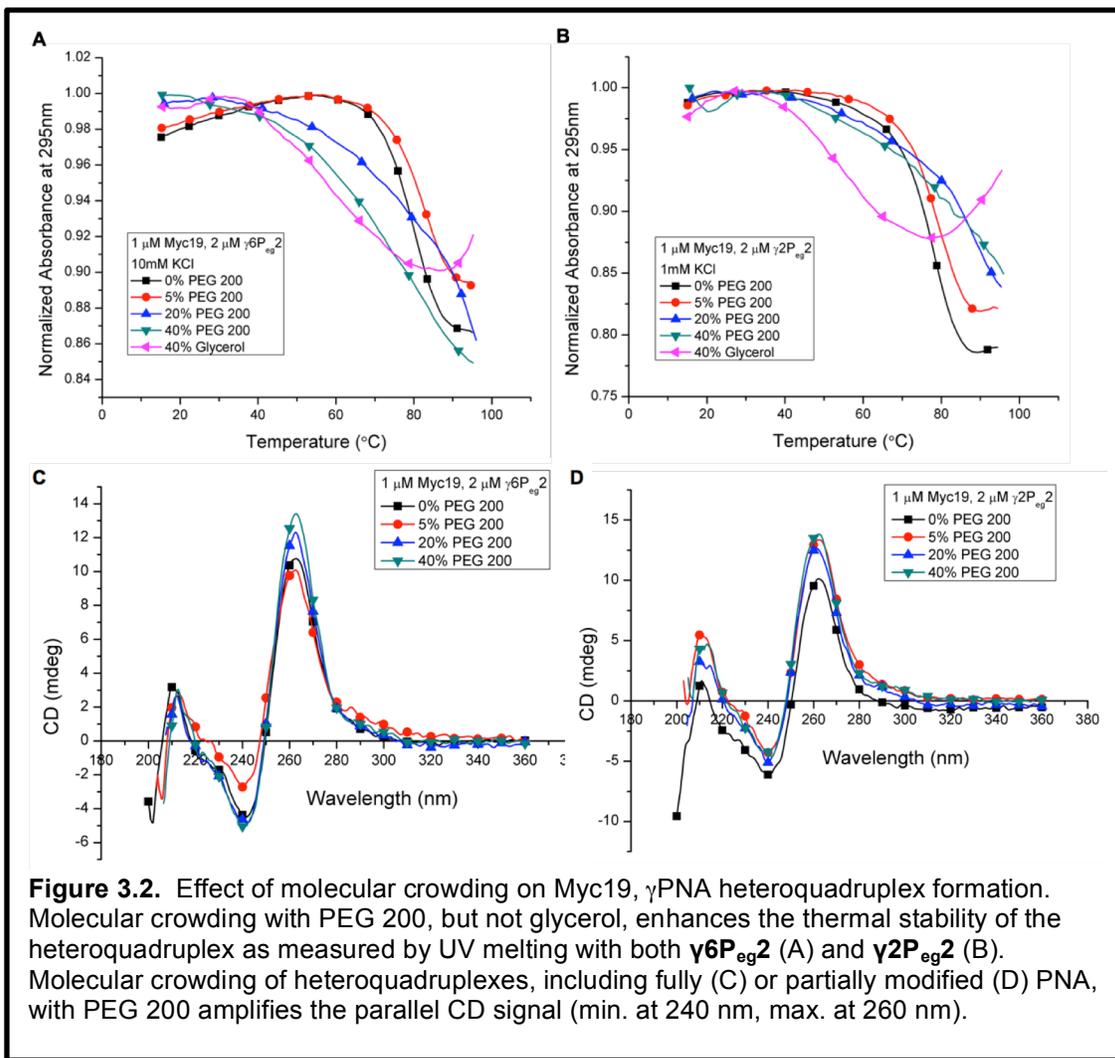
Table 3.2: Thermal stability (T_m , °C) of Myc19/ γ PNA heteroquadruplexes under molecular crowding conditions.

γ PNA	0% PEG 200	5% PEG 200	20% PEG 200	40% PEG 200	40% Glycerol (10 mM K^+)
$\gamma 6P_{eg}2$ (10 mM K^+)	71.7±0.9	73.9±0.8	83.7±0.4	ND*	55.4±0.7
$\gamma 2P_{eg}2$ (1 mM K^+)	78.7±1.1	78.6±0.3	ND*	ND*	58.1±1.9

* T_m could not be determined due to incomplete transition.

CD spectropolarimetry indicates parallel folding topology (min. at 240 nm, max. at 260 nm) for both PNA/DNA heteroquadruplexes, with slight amplification

of this signal in the presence of PEG 200 (Figure 3.2C, D).



Molecular crowding of γ -modified PNAs in complex with a hybrid DNA G-quadruplex target. hTelo22 is a quadruplex-forming oligomer derived from the human repeating telomeric unit. hTelo22 is not vastly different from Myc19 in terms of primary structure, however, it produces a CD signature consistent with a hybrid/mixed antiparallel quadruplex, at least in dilute *in vitro* conditions[2, 20]. As with Myc19, UV melting at 295 nm detects heteroquadruplex formation between both of the γ PNAs and hTelo22, which is further stabilized by PEG 200, but not glycerol (Table 3.3; Figure 3.3A, B). However, the stabilities of the

heteroquadruplexes formed with hTelo22 do not match those with Myc19, especially considering that these measurements were taken at ten times the K^+ concentration (10 mM) in the case of $\gamma 2P_{eg}2$. Additionally, the level of thermodynamic discrimination between $\gamma 6P_{eg}2$ and $\gamma 2P_{eg}2$ is not as pronounced with hTelo22 as with Myc19. Together these data are suggestive of the preference of these probes for parallel quadruplex targets.

Table 3.3. Thermal stability (T_m , °C) of hTelo22/ γ PNA heteroquadruplexes under molecular crowding conditions (10 mM K^+).

γ PNA	0% PEG 200	5% PEG 200	20% PEG 200	40% PEG 200	40% Glycerol
$\gamma 6P_{eg}2$	63.9±2.0	63.8±1.1	64.1±1.6	ND*	60.5±6.9
$\gamma 2P_{eg}2$	59.3±3.1	62.7±2.7	69.0±0.2	ND*	62.8±2.2

* T_m could not be determined due to incomplete transition.

The CD spectra of hTelo22 in complex with both $\gamma 6P_{eg}2$ and $\gamma 2P_{eg}2$ display parallel features, but also retain features indicative of a hybrid/mixed antiparallel structure (broad shoulder from 280-300 nm; Figure 3.3C, D). While increasing the concentration of PEG 200 enhances the parallel signal, the hTelo22/ $\gamma 2P_{eg}2$ heteroquadruplex is resistant to complete transition to parallel topology even at 40% PEG 200, which may reflect the greater conformational flexibility expected with the partially modified PNA compared to the fully modified probe. Similarly, the CD spectrum of hTelo22 in complex with an unmodified PNA in 40% PEG 200 also retains some antiparallel/hybrid signal[18].

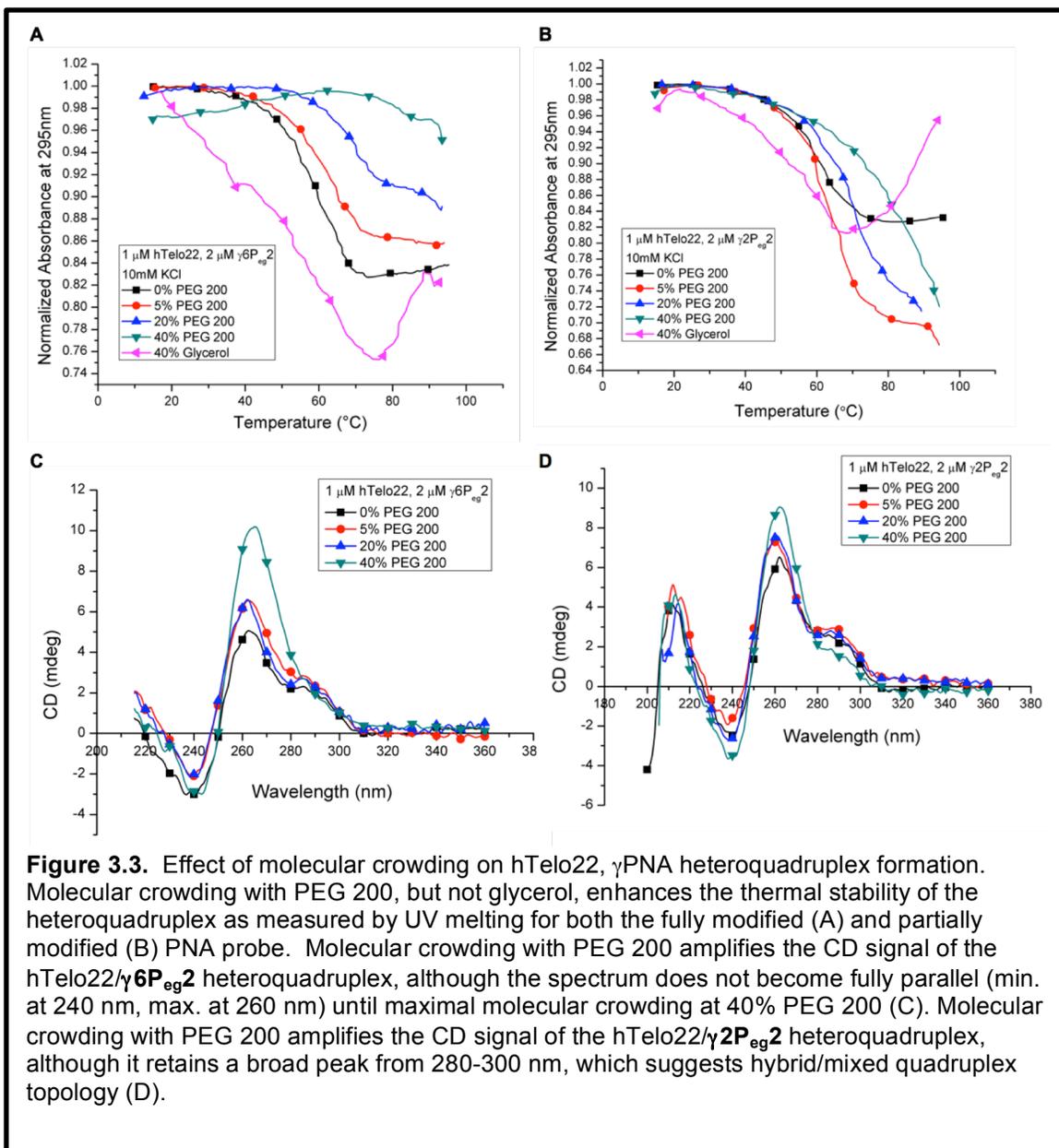


Figure 3.3. Effect of molecular crowding on hTelo22, γ PNA heteroquadruplex formation. Molecular crowding with PEG 200, but not glycerol, enhances the thermal stability of the heteroquadruplex as measured by UV melting for both the fully modified (A) and partially modified (B) PNA probe. Molecular crowding with PEG 200 amplifies the CD signal of the hTelo22/ γ 6P_{eg}2 heteroquadruplex, although the spectrum does not become fully parallel (min. at 240 nm, max. at 260 nm) until maximal molecular crowding at 40% PEG 200 (C). Molecular crowding with PEG 200 amplifies the CD signal of the hTelo22/ γ 2P_{eg}2 heteroquadruplex, although it retains a broad peak from 280-300 nm, which suggests hybrid/mixed quadruplex topology (D).

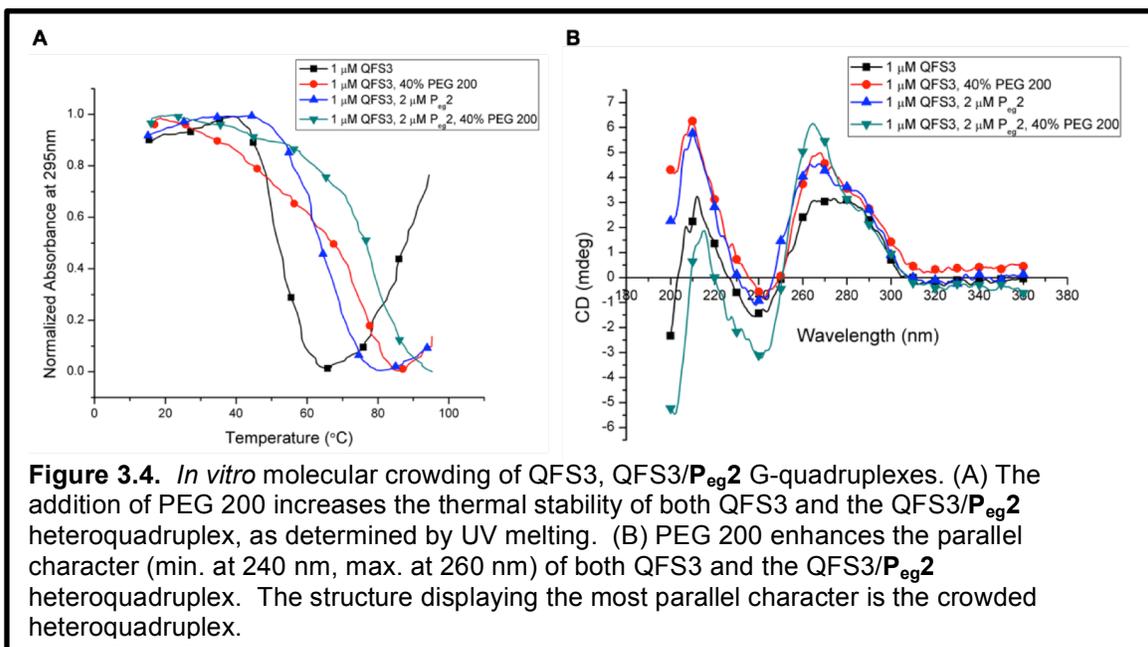
Molecular crowding of the QFS3/P_{eg}2 heteroquadruplex. QFS3 is a long loop-containing quadruplex-forming sequence found within the *S. cerevisiae* 25S rDNA[22]. We have characterized the folding of QFS3 and its interaction with the homologous PNA, P_{eg}2, in Chapter 2[22]. Here we have investigated homologous PNA interaction with a long-looped quadruplex in conditions that mimic biological molecular crowding. As with the canonical quadruplex targets,

the stability of the QFS3/**P_{eg}2** heteroquadruplex is intensified in 40% PEG 200 (Table 3.4, Figure 3.4A).

Table 3.4. Enhanced Thermal Stability (T_m , °C) of QFS3 upon **P_{eg}2** hybridization, molecular crowding (100 mM K^+).

DNA only	54.9±3.0
+ 40% PEG 200	77.1±2.0
+ 2 μM PNA	61.6±0.4
+ 40% PEG 200 + 2 μM PNA	82.3±5.3

QFS3 alone produces a CD signal indicating a more complex topology than a simple parallel or antiparallel structure, likely due to contribution from the nucleotides in the long loop. While the addition of **P_{eg}2** begins to shift the CD spectrum towards that of a parallel quadruplex, this effect is magnified in the presence of 40% PEG 200 (Figure 3.4B).



3.4 Discussion

(R)-diethylene glycol γ PNAs support heteroquadruplex formation with Myc19 DNA. We have demonstrated stable heteroquadruplex formation between

homologous γ PNAs and two DNA G-quadruplex targets. Both $\gamma 6P_{eg}2$ and $\gamma 2P_{eg}2$ interact favorably with the parallel quadruplex, Myc19, which is consistent with previous studies of similar PNAs with this target. Initial studies of unmodified homologous PNA hybridization to Myc19 detected 2:1 PNA:DNA heteroquadruplex formation with low nanomolar binding affinity[19, 21]. Introduction of abasic linkers separating the G-tracts in the $P_{eg}2$ PNA produced even more stable heteroquadruplexes, while also eliminating loop bases that could potentially contribute to off-target duplex interactions[5]. Chiral γ modifications confer helical character to the PNA oligomer[6]. Both right- and left-handed versions of homologous PNAs have been shown to interact favorably with Myc19, although the right-handed probes also hybridize very well to complementary targets[5]. As an addendum to these studies, we have explored the interactions of PNAs combining G-quadruplex-homology, abasic linkers, and diethylene glycol γ substituents with Myc19. As with earlier homologous PNAs, our probes invade Myc19 to form heteroquadruplexes that produce parallel CD signatures very similar to the CD spectrum of Myc19 alone[5, 19]. These similarities in CD signal suggest that our γ PNAs interact with Myc19 in a comparable manner to the previously reported probes. Interestingly, the partially modified $\gamma 2P_{eg}2$ probe forms more stable complexes with Myc19 than its fully modified counterpart. The Myc19/ $\gamma 2P_{eg}2$ heteroquadruplex has comparable stability to the previously reported alanyl- γ -modified homologous probe in complex with this target[5], confirming that this diethylene glycol moiety does not negatively influence thermodynamics of hybridization much more than the

alanine-based modifications. This result suggests that each position of the PNA does not need to be γ -modified for effective quadruplex hybridization, which leaves opportunities for the incorporation of other useful chemical substituents, such as fluorophores.

Homologous (R)-diethylene glycol γ PNAs favor parallel DNA G-quadruplexes.

We have shown that γ 6P_{eg}2 and γ 2P_{eg}2 interact favorably with Myc19 and hTelo22. Additionally, hybridization of our γ PNA probes to hTelo22 intensifies the CD signal at wavelengths characteristic of a parallel quadruplex structure compared to that of hTelo22 alone, which is similar to previous reports of homologous PNA interaction with this DNA[18, 20]. However, both γ PNAs display thermodynamic discrimination favoring the parallel target, Myc19, even at a 10-fold lower K⁺ concentration. These results are consistent with previous findings demonstrating thermodynamic, as well as kinetic preferences of unmodified homologous PNA for Myc19 over hTelo22[20]. The kinetic selectivity, specifically faster association rates, of homologous PNAs for parallel quadruplexes has been suggested to be related to the tendency of G's within DNA/PNA heteroquadruplexes to be in the *anti* conformation[20]. However, the difference among quadruplexes can be mitigated by slower off-rates for antiparallel targets[20]. Although we did not examine kinetic parameters of hybridization of our γ PNAs to Myc19 and hTelo22, we anticipate that the γ modifications will reduce the conformational flexibility of the probes, which may result in slower quadruplex invasion when compared to unmodified PNAs. Previous studies have indicated that both right- and left-handed alanyl- γ -modified

homologous probes hybridize to Myc19 more slowly than an unmodified PNA[5]. Further studies will be required to confirm this hypothesis with the probes studied here.

Molecular crowding enhances DNA/PNA heteroquadruplex hybridization thermodynamics. In this study we have demonstrated that *in vitro* molecular crowding with PEG 200 increases the thermal stability of unmodified and γ PNA/DNA heteroquadruplexes. This effect is not seen with the viscous co-solvent, glycerol. Our CD data also suggest that molecular crowding drives the folding topology of these heteroquadruplexes towards the parallel conformation, which is especially evident for hTelo22 and QFS3. We are limited to thermodynamic analysis of crowded heteroquadruplexes due to the incompatibility of viscous solutions with the SPR apparatus, impeding collection of kinetic data. Similar effects of molecular crowding on thermal stability and folding topology have been observed previously in the Armitage group upon hybridization of unmodified homologous PNAs to Myc19 and hTelo22[18]. We have expanded upon these principles using both fully and partially modified γ PNAs, in addition to the noncanonical target, QFS3. Together, these data suggest that hybridization of several classes of homologous PNAs to DNA G-quadruplexes should be favorable in the crowded intracellular environment, and provide information about how these molecules may interact in such conditions.

3.5 Conclusion and Future Perspective

With this work, we have demonstrated that partially and fully ditheylene glycol γ -modified homologous PNAs are compatible with DNA/PNA heteroquadruplex

formation, specifically with Myc19 and hTelo22. These γ PNAs drive parallel heteroquadruplex folding topology, forming stable complexes that are further stabilized by molecular crowding. We have also expanded our understanding of the effects of molecular crowding on homologous PNA hybridization to three distinct DNA G-quadruplexes. Our results suggest that *in vitro* studies of DNA/PNA interactions should include analysis of these complexes under molecular crowding conditions, which may produce vastly different results than the same assays in dilute solution. Additionally, crowding agents (e.g. bovine serum albumin, acetonitrile, or ethanol) should be explored as alternatives to PEG. While this work has expanded upon knowledge of PNA/DNA G-quadruplex interactions, future studies should be directed toward the identification of an ideal probe design for specific quadruplex targets. For example, we know that hTelo22 preferentially hybridizes to complementary PNAs[2]. γ PNAs can be conformationally pre-organized into right-handed helical polymers, which are ideal for duplex formation with nucleic acids[6]. Therefore, complementary γ PNA probes should also be tested with this G-quadruplex.

The most recent, high-throughput experimental and computational approaches have estimated the number of putative G-quadruplexes throughout the human genome at nearly three quarters of a million, including motifs having unconventional features such as long loops and bulges[25]. Simultaneously, the library of synthetic molecules able to interact with these quadruplexes has continued to expand. Previous work has confirmed that the ideal mechanism of probe hybridization depends on the quadruplex being targeted. Specifically, our

group has shown that hTelo22 is more amenable to invasion by a complementary, duplex-forming PNA than a homologous probe, at least in dilute solution[2]. In contrast, Myc19 hybridizes more efficiently to a homologous PNA than a complementary probe under conditions mimicking physiological K^+ concentration[2]. Therefore, detailed analyses of the interactions among varying classes of G-quadruplexes and quadruplex-targeting probes must be completed to facilitate selection of the best probe design for hybridization for a G-quadruplex of interest.

3.6 Materials and Methods

3.6.1 Materials

DNA oligonucleotides were purchased from Integrated DNA Technologies (www.idtdna.com) and Fisher Scientific (www.fishersci.com). (*R*)-diethylene glycol γ PNA oligomers were purchased from PNA Innovations (www.pnainnovations.com).

3.6.2 Equipment

UV melting data were collected with a Varian Cary 3 spectrophotometer combined with a temperature controlled multicell holder. CD data were collected on a Jasco J-715 spectropolarimeter with a single cell holder and water circulating temperature controller.

3.6.3 UV Melting

Samples (1 μ M DNA + 2 μ M PNA) were buffered in 100 mM KCl, 10 mM Tris pH 7, and 0.1 mM Na_2EDTA , unless otherwise noted. Molecular crowding was done through the addition of increasing concentrations of the crowding agent (0,

5, 20, and 40% w/v PEG 200). As a control, 40% of the sample volume was replaced with glycerol. Samples were heated to 95 °C and then annealed by slow cooling to 15 °C (1 °C/min). The samples were then slowly melted by heating to 95 °C while absorbance was measured at 295 nm. Melting temperatures were calculated by averaging the minimum of the first derivative of each melting curve, in triplicate (OriginPro8).

3.6.4 Circular Dichroism

CD data were collected at the physiological temperature corresponding to each quadruplex-forming sequence: 37 °C for Myc19, hTelo22 and 30 °C for QFS3. Each spectrum was obtained by averaging 6 scans collected at 100 nm/min.

3.7 Acknowledgement

We thank the David Scaife Family Charitable Foundation for financial support of this research (Award 141RA01).

3.8 References

1. Datta, B. and B.A. Armitage, *Hybridization of PNA to structured DNA targets: quadruplex invasion and the overhang effect*. J. Am. Chem. Soc., 2001. **123**: p. 9612-9619.
2. Gupta, A., et al., *Strand invasion of DNA quadruplexes by PNA: comparison of homologous and complementary hybridization*. Chembiochem, 2013. **14**(12): p. 1476-84.
3. Murphy, C.T., et al., *Hybridization of G-quadruplex-forming peptide nucleic acids to guanine-rich DNA templates inhibits DNA polymerase ϵ extension*. Biochemistry, 2014. **53**(32): p. 5315-22.
4. Usui, K., et al., *Control of guanine-rich DNA secondary structures depending on the protease activity using a designed PNA peptide*. Org Biomol Chem, 2015. **13**(7): p. 2022-5.
5. Lusvardi, S., et al., *Loop and backbone modifications of peptide nucleic acid improve g-quadruplex binding selectivity*. J Am Chem Soc, 2009. **131**(51): p. 18415-24.

6. Yeh, J.I., et al., *Crystal Structure of Chiral γ PNA with Complementary DNA Strand: Insights into the Stability and Specificity of Recognition and Conformational Preorganization*. J. Am. Chem. Soc., 2010. **132**: p. 10717-10727.
7. Sahu, B., et al., *Synthesis and characterization of conformationally preorganized, (R)-diethylene glycol-containing gamma-peptide nucleic acids with superior hybridization properties and water solubility*. J Org Chem, 2011. **76**(14): p. 5614-27.
8. Zimmerman, S.B. and S.O. Trach, *Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli*. J. Mol. Biol., 1991. **222**: p. 599-620.
9. Zimmerman, S.B. and A.P. Minton, *Macromolecular crowding: biochemical, biophysical, and physiological consequences*. Annu. Rev. Biophys. Biomol. Struc., 1993. **22**: p. 27-65.
10. Miyoshi, D. and N. Sugimoto, *Molecular crowding effects on structure and stability of DNA*. Biochimie, 2008. **90**(7): p. 1040-51.
11. Xue, Y., et al., *Human telomeric DNA forms parallel-stranded intramolecular G-quadruplex in K⁺ solution under molecular crowding condition*. J Am Chem Soc, 2007. **129**(36): p. 11185-91.
12. Miyoshi, D., A. Nakao, and N. Sugimoto, *Molecular crowding regulates the structural switch of the DNA G-quadruplex*. Biochemistry, 2002. **41**: p. 15017-15024.
13. Buscaglia, R., et al., *Polyethylene glycol binding alters human telomere G-quadruplex structure by conformational selection*. Nucleic Acids Res, 2013. **41**(16): p. 7934-46.
14. Miyoshi, D., H. Karimata, and N. Sugimoto, *Hydration regulates thermodynamics of G-quadruplex formation under molecular crowding conditions*. J. Am. Chem. Soc., 2006. **128**: p. 7957-7963.
15. Verdian Doghaei, A., M.R. Housaindokht, and M.R. Bozorgmehr, *Molecular crowding effects on conformation and stability of G-quadruplex DNA structure: insights from molecular dynamics simulation*. J Theor Biol, 2015. **364**: p. 103-12.
16. Zheng, K.W., et al., *Molecular crowding creates an essential environment for the formation of stable G-quadruplexes in long double-stranded DNA*. Nucleic Acids Res, 2010. **38**(1): p. 327-38.
17. Chen, Z., et al., *Reduced or diminished stabilization of the telomere G-quadruplex and inhibition of telomerase by small chemical ligands under molecular crowding condition*. J Am Chem Soc, 2009. **131**: p. 10430-10438.
18. Mohammed, H.S., *Ph.D. Dissertation, Carnegie Mellon University*. 2012.
19. Roy, S., et al., *High-affinity homologous peptide nucleic acid probes for targeting a quadruplex-forming sequence from a MYC promoter element*. Biochemistry, 2007. **46**: p. 10433-10443.
20. Roy, S., et al., *Kinetic discrimination in recognition of DNA quadruplex targets by guanine-rich heteroquadruplex-forming PNA probes*. Chem Commun, 2011. **47**(30): p. 8524-6.

21. Mohammed, H.S., J.O. Delos Santos, and B.A. Armitage, *Noncovalent binding and fluorogenic response of cyanine dyes to DNA homoquadruplex and PNA-DNA heteroquadruplex structures*. *Artif DNA PNA XNA*, 2011. **2**(2): p. 43-49.
22. Kormuth, K.A., J.L. Woolford, and B.A. Armitage, *Homologous PNA Hybridization to Noncanonical DNA G-Quadruplexes*. *Biochemistry*, 2016.
23. Simonsson, T., P. Pecinka, and M. Kubsita, *DNA tetraplex formation in the control region of c-myc*. *Nucleic Acids Res*, 1998. **26**(5): p. 1167-1172.
24. Siddiqui-Jain, A., et al., *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): p. 11593-8.
25. Chambers, V.S., et al., *High-throughput sequencing of DNA G-quadruplex structures in the human genome*. *Nat Biotechnol*, 2015.

4. Chapter 4: Evidence for the Biological Significance of G-quadruplex Motifs in Ribosomal RNA Biogenesis

4.1 Chapter Summary

Although G-quadruplex motifs are found throughout the genome, the potential regulatory power of these structures may be unmatched in ribosomal genes due to their localization in the non-template rDNA strand and also in rRNA. Ribosome biosynthesis is intimately connected to cell growth. Therefore, regulation of all processes involved in this task, ranging from maintenance of ribosomal genes to processing and assembly of rRNA, must be tightly regulated to ensure the health of the cell. DNA or RNA G-quadruplexes can conceivably assist regulation in one or more of these processes, although a quadruplex structure has not been detected in the mature ribosome. Certain disease states, such as cancer, are driven partly by nucleolar dysfunction, which may be related to aberrant rDNA quadruplex interactions with intracellular factors. Despite evidence supporting the formation and function of rDNA G-quadruplexes, the specific roles of these structures in cell biology are not yet understood. We have approached this problem using the yeast, *Saccharomyces cerevisiae*, model system. We have identified at least four long-looped G-quadruplex motifs within regions of the gene that template the mature 18S and 25S rRNAs, and confirmed that these all fold into stable, intramolecular quadruplexes *in vitro*. To overcome the challenges associated with genomic rDNA repeat copy number and the limitations of previously existing tools for rDNA mutagenesis, we have developed a two-step protocol to introduce and study quadruplex mutants *in vivo*. Our

results suggest that several of these quadruplex motifs are important for assembly and/or stability of mature rRNA and may even exist as transitional conformations, distinct from predicted stem-loop structures, within flexible regions of the mature ribosome. Future efforts will be required to confirm that these motifs do fold into G-quadruplexes *in vivo* and to identify their specific functions, providing a more complete understanding of ribosome biogenesis.

4.2 Introduction

The rDNA houses the genes that are among those most vital for life, as they are responsible for templating the rRNA that becomes the infrastructure and functional center of the ribosome (reviewed in Woolford and Baserga, 2013)[1]. The rDNA locus contains two genes, one of which is transcribed by RNA polymerase III to produce 5S rRNA. The other gene, which is transcribed by RNA polymerase I, is a polycistron that produces the 18S, 5.8S, and 25S rRNAs, initially transcribed as a single rRNA unit. Together the 25S, 5S, and 5.8S RNAs become the rRNA components of the 60S large ribosomal subunit, while 18S rRNA is incorporated into the 40S small subunit. These rRNAs undergo complicated and interconnected processing and assembly steps with assembly factors and ribosomal proteins to form the mature ribosome[1]. Eukaryotic rDNA quadruplex motifs are uniquely localized to the non-template DNA strand, but unlike other genes, are found throughout the spacers and rRNA regions of the gene[2, 3]. This unusual localization suggests that these quadruplexes, at the DNA level, may be involved in processes wherein the DNA duplex is melted, such as transcription, replication, or recombination, thus allowing formation of

intramolecular G-quadruplex structures[2]. Based on this perception, our initial expectation was that these quadruplexes could serve to support rRNA transcription via stabilization of the transcription bubble, allowing for the passage of additional polymerases through this highly transcribed region of the genome[4].

We elected to explore this problem using specialized *S. cerevisiae* strains designed for introduction and analysis of rDNA mutations. As in higher eukaryotes, G-quadruplex motifs have been identified across the yeast rDNA repeat[3, 5]. Apart from the telomeres, the rDNA contains approximately a 10-fold higher density of quadruplex motifs than the rest of the genome[3]. These loci also respond to conditions that modulate quadruplex stability. Specifically, treatment with the quadruplex-stabilizing drug, NMM (n-methyl mesoporphyrin IX), downregulates loci associated with rDNA function[3], although the biological implications of this finding are not completely clear. Despite the prevalence of yeast rDNA quadruplex motifs, the functions of these quadruplexes may not be completely analogous to similar loci in humans. Specifically, the yeast genome does not contain many canonical (short-looped) quadruplex motifs when compared to higher eukaryotes[6]. However, the growing acceptance in the field for alternative motifs reinforces the need to expand investigations of these quadruplexes. Additionally, yeast rDNA quadruplexes are evolutionarily conserved among related fungi, which provides evidence for *in vivo* functions of these structures[5]. We anticipate that, even if yeast rDNA quadruplex functions

do not exactly mirror those in human cells, investigation of these motifs could help to define new paradigms in G-quadruplex biology.

A practical advantage of the yeast model system is the accessibility of genetic tools that allow for rDNA mutagenesis. rDNA exists as a series of repeated copies, rendering introduction of mutations within the genome impossible[7]. Although a variety of systems have been designed to overcome this problem, these frequently utilize artificial promoters or result in high levels of contamination from wildtype, chromosome-derived ribosomes[7]. Therefore, we designed and implemented our own two-step protocol for analysis of rDNA mutants. Initially, we inserted neutral sequence tags[7] into the 25S and 18S regions of an rDNA plasmid[8] containing the entire rDNA repeat under control of the endogenous RNA polymerase I promoter. This design is similar to previous strategies[9], although we selected sequence tags different from those previously employed in other rDNA mutagenesis vectors[7]. To study the rRNA phenotypes of rDNA mutants, we simply transformed the tagged, mutant vectors into wildtype cells and probed specifically for the plasmid-derived rRNAs using northern blots (Figure 4.1). While this method allows for detection of mutant rRNA, the cells do not rely on this rRNA for survival, allowing for analysis of potentially lethal mutations. However, the caveat of this protocol is that there is no mechanism to separate and detect plasmid-derived pre-rRNAs separately

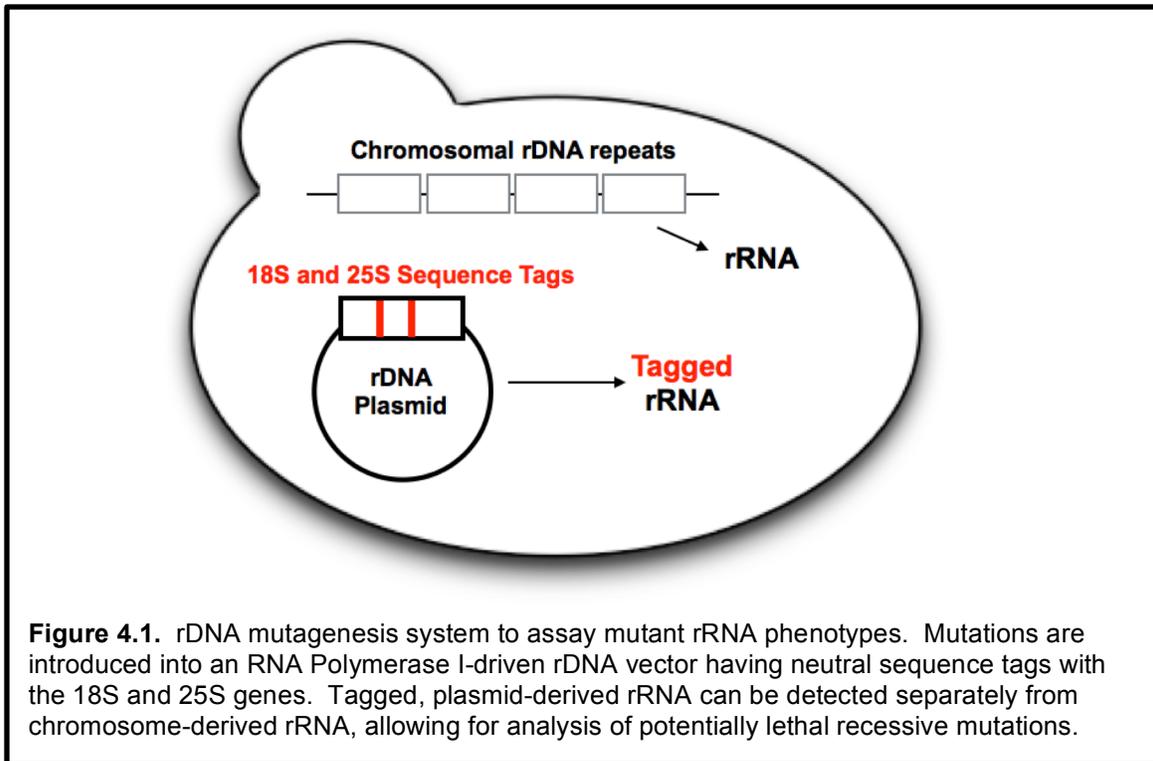


Figure 4.1. rDNA mutagenesis system to assay mutant rRNA phenotypes. Mutations are introduced into an RNA Polymerase I-driven rDNA vector having neutral sequence tags with the 18S and 25S genes. Tagged, plasmid-derived rRNA can be detected separately from chromosome-derived rRNA, allowing for analysis of potentially lethal recessive mutations.

from endogenous copies, limiting our analysis of the effects of these mutations on rRNA processing and ribosome assembly.

To explore the effects of sublethal mutations on cell growth, ribosome stability/function, and rRNA structure, we have employed a strain of yeast lacking

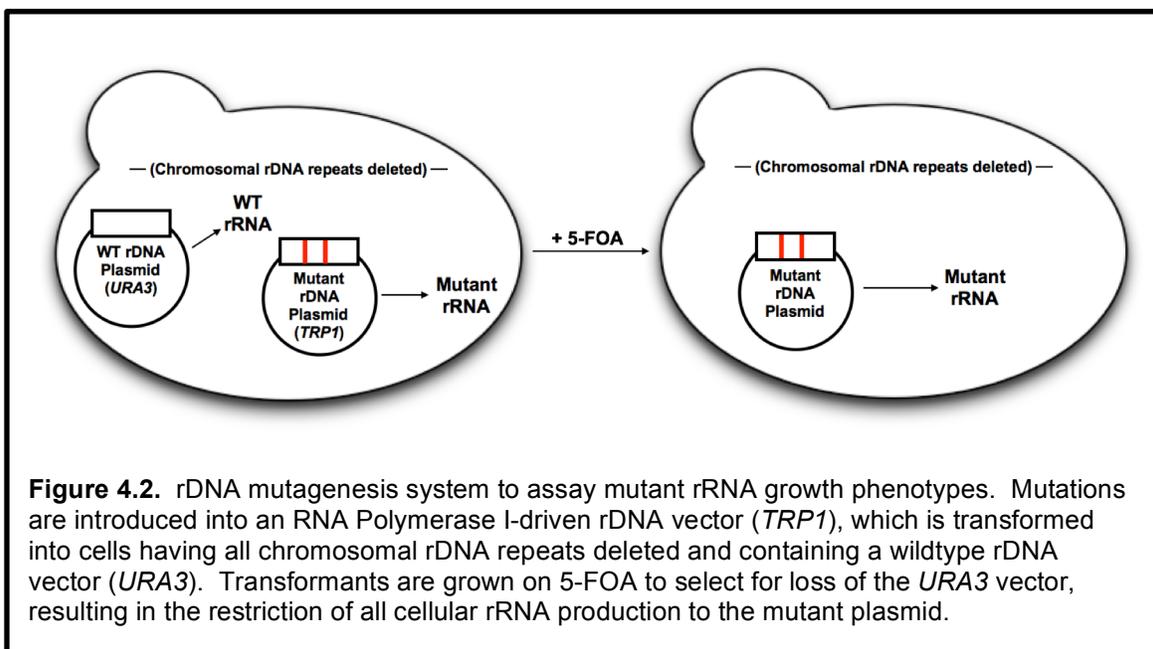


Figure 4.2. rDNA mutagenesis system to assay mutant rRNA growth phenotypes. Mutations are introduced into an RNA Polymerase I-driven rDNA vector (*TRP1*), which is transformed into cells having all chromosomal rDNA repeats deleted and containing a wildtype rDNA vector (*URA3*). Transformants are grown on 5-FOA to select for loss of the *URA3* vector, resulting in the restriction of all cellular rRNA production to the mutant plasmid.

all chromosomal repeats of the rDNA (*rdn1ΔΔ*)[7]. This strain produces rRNA from a high-copy rDNA plasmid containing a *URA3* selectable marker. To this strain we introduce our mutant vector (*TRP1* selection) and then incubate transformants with 5-FOA (5-fluoroorotic acid), which promotes loss of the *URA3* vector in favor of the *TRP1* plasmid[8]. This system (Figure 4.2) allows all the cellular rDNA to be produced by the mutant vector, expanding our capacity for analysis of rDNA mutant phenotypes.

With this study, we have expanded upon our *in vitro* analyses of noncanonical G-quadruplex motifs in the *S. cerevisiae* model system. We identified and characterized the folding of four yeast rDNA quadruplex motifs (QFS1-4) *in vitro*. To approach analysis of the biological functions of these sequences, we designed a two-step strategy to overcome problems associated with rDNA mutagenesis and implemented our own protocols to study G-quadruplex motif mutants *in vivo*. Our data support the functional importance of QFS2 and QFS3 within the rRNA. Specifically, our results indicate that integrity of the QFS2/3 region is important for large ribosomal subunit stability and cell growth. QFS2/3 is located within a particularly flexible portion of the 25S rRNA, which may adopt a quadruplex fold as a conformational intermediate. Our work provides evidence for the biological relevance of rDNA/rRNA G-quadruplexes in yeast, although future studies will be required to identify the exact cellular functions of these structures.

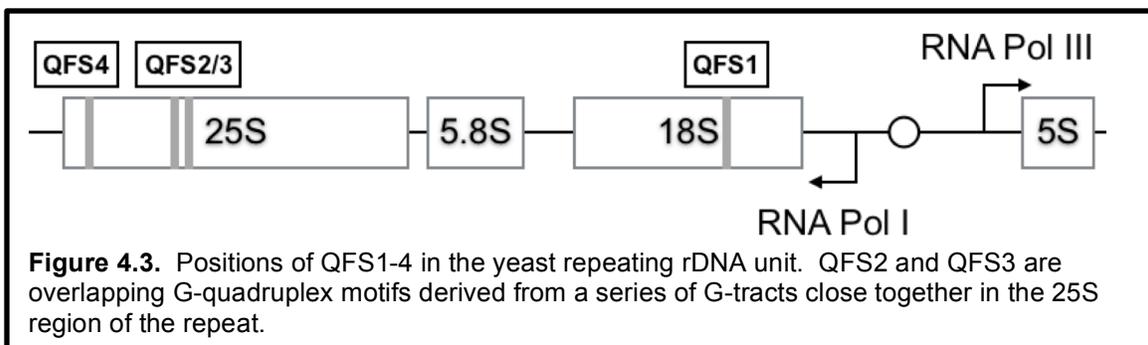
4.3 Results

Noncanonical G-quadruplex motifs identified in the yeast rDNA fold into stable quadruplexes in vitro. We have identified four potential long-looped G-quadruplex motifs (QFS1-4) in the *S. cerevisiae* rDNA repeat using an algorithm[3] to search for sequences following the pattern, $G_3X_nG_3X_nG_3X_nG_3$, with a maximum total length of 50 nucleotides (Table 4.1). All loci of putative G-quadruplex motifs in the *S. cerevisiae* genome were previously annotated by Capra and colleagues[5]. However, we identified QFS1-4 specifically from the sequence of a wildtype yeast rDNA vector (pWL160[7]) and, importantly, determined that these quadruplex motifs were invariant in all the wildtype rDNA vectors used in this study (Supplemental Table 4.S1).

Table 4.1. Yeast rDNA G-quadruplex motifs. (G-tracts predicted to participate in tetrad formation are shown in bold and underlined.)

DNA (5'-to-3')	
QFS1	<u>GGG</u>TAAC<u>GGG</u>GAATA<u>GGG</u>TTTCGATTCCGGAG<u>GGG</u>
QFS2	<u>GGG</u> ATAAGGATTGGCTCTA <u>GGG</u> T <u>GGG</u> TAGTG <u>GGG</u>
QFS3	<u>GGG</u> T <u>GGG</u> TAGTG <u>GGG</u> CCTTGGTCAGACGCAGC <u>GGG</u>
QFSS4	<u>GGG</u> GCTGATCC <u>GGG</u> TTGAAGACATTGTCAGGT <u>GGG</u> GAGTTTGGCT <u>GGG</u>

While all four QFS's are located on the nontemplate DNA strand (and in the rRNA), QFS1 is found within the 18S region of the gene and QFS2-4 are within the 25S region (Figure 4.3). In both cases G-quadruplex formation is in competition with alternative secondary structure formation, specifically, duplex binding to the complementary strand in the DNA and stem-loop formation of the rRNA in the mature ribosome.

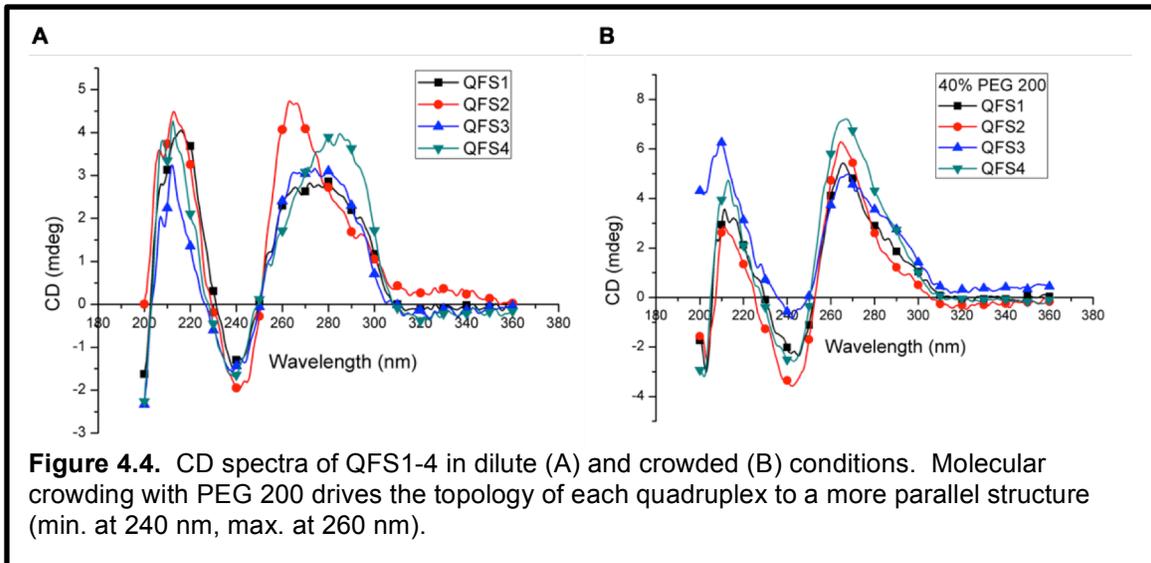


As we have previously demonstrated with QFS3, the other QFSs all fold into intramolecular G-quadruplex structures *in vitro*, which we have confirmed by UV melting (Table 4.2 and Supplemental Figure 4.S1) and circular dichroism (Figure 4.1). Molecular crowding with PEG 200 further enhances the stability of these quadruplexes (Table 4.2 and Supplemental Figure 4.S1). Substitution of 40% glycerol in place of PEG 200 does not produce the same effect, suggesting that the enhancement in stability is due to excluded volume effects (Table 4.2 and Supplemental Figure 4.S2).

Table 4.2. QFS1-4 T_m ($^{\circ}\text{C}$) dependence on DNA concentration (left) and molecular crowding (right) determined by UV melting.

	1 μM DNA	10 μM DNA	1 μM DNA	
			+ 40% PEG 200	+ 40% Glycerol
QFS1	46.7 \pm 1.3	47.6 \pm 2.9	71.2 \pm 1.6	41.7 \pm 5.3
QFS2	48.1 \pm 5.1	50.8 \pm 4.9	77.5 \pm 2.6	50.0 \pm 1.9
QFS3	54.9 \pm 3.0	56.0 \pm 0.4	77.1 \pm 2.0	57.4 \pm 1.1
QFS4	45.9 \pm 2.9	42.9 \pm 2.6	52.8 \pm 0.4	49.8 \pm 2.0

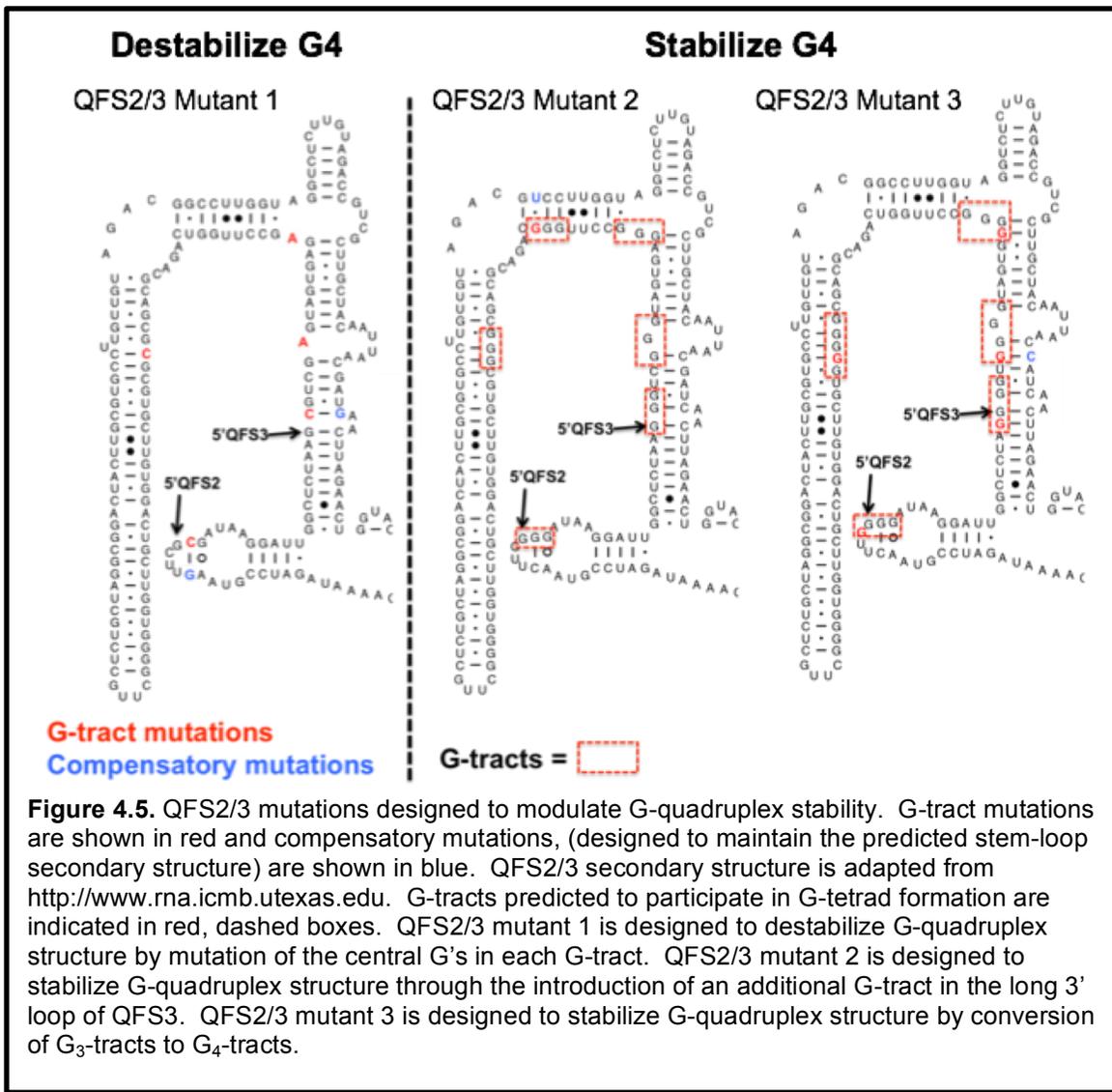
Similar to results discussed in detail Chapter 3, PEG 200 also induces a more parallel quadruplex topology for each QFS than the structure that is detected in dilute solution (Figure 4.4).



Design of QFS2/3 mutations. Mutations introduced into the QFS2/3 region of the 25S rDNA were designed to modulate G-quadruplex structural stability at this site. QFS2/3 is located within a highly dynamic region of the rRNA, eukaryote-specific expansion segment 27 (ES27L), for which we do not yet have a high-resolution structure in yeast [10, 11], although the commonly accepted secondary structure of this region is an extended stem-loop[12]. The mutations introduced into QFS2/3 are designed to either destabilize (mutant 1) or stabilize (mutants 2 and 3) G-quadruplex formation, while maintaining the integrity of the mature, stem-loop secondary structure (Figure 4.5). Specifically, QFS2/3 mutant 1 is composed of G-tract mutations, wherein the central guanine of each G-tract is changed to an A or C. If the mutated nucleotide is predicted to be involved in a base-pairing interaction, the corresponding partner is also mutated to allow duplex interactions, (which is generally the case for all three QFS2/3 mutants). QFS2/3 mutant 2 introduces an additional G₃-tract into the 3' long loop region of QFS3 via a simple G-U wobble base pair flip. QFS2/3 mutant

3 is more complex, extending the length of each G₃-tract to at least four guanines.

We also assayed several 18S rDNA mutants intended to reduce the stability of the QFS1 putative quadruplex and compensate for disruptions to the canonical, stem-loop secondary structure of this region of the rRNA. While we detected a decrease in 18S rRNA from northern blots with this mutant, further analysis suggested that defects in 18S rRNA synthesis indicated this phenotype was compensated by mutations designed to mimic the wildtype, mature secondary structure (Supplemental Figure 4.S3). In this case, our data indicate that the rRNA phenotype resulted from disruption of a specific helix, masking any potential quadruplex-derived phenotypes in this QFS.



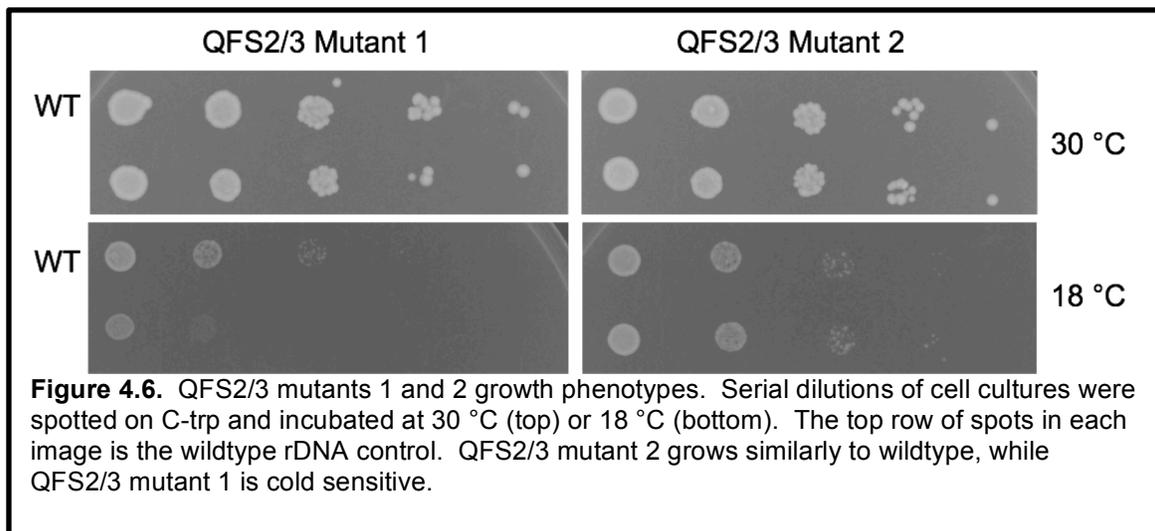
QFS2/3 mutant cell growth and polysome profile defects. Cell growth phenotypes of the QFS2/3 mutants were assayed following transformation of mutant plasmids (*TRP1*) into cells relying entirely on a wildtype rDNA vector (*URA3*) for rRNA transcription. Incubation of transformants on C-trp + 5-FOA resulted in loss of the wildtype vector, allowing total rRNA production to be derived from the mutant plasmid, providing that the mutations were sublethal. Each of the QFS2/3 mutants produced different cell growth phenotypes (Table

4.3 and Figure 4.6). We consider QFS2/3 mutant 3 to be lethal because these transformants did not produce viable colonies after replica plating on C-trp + 5-FOA. Following subsequent transformations, we detected a number of background colonies with this mutant, likely due to recombination between the mutant and wildtype vectors (data not shown). As a result of this problem, we chose not to pursue further analysis of this mutant in the chromosomal rDNA deletion strain. The other quadruplex-stabilizing mutant, QFS2/3 mutant 2, did not produce any detectable growth defects. However, QFS2/3 mutant 1 was reproducibly cold sensitive during incubation at 18 °C.

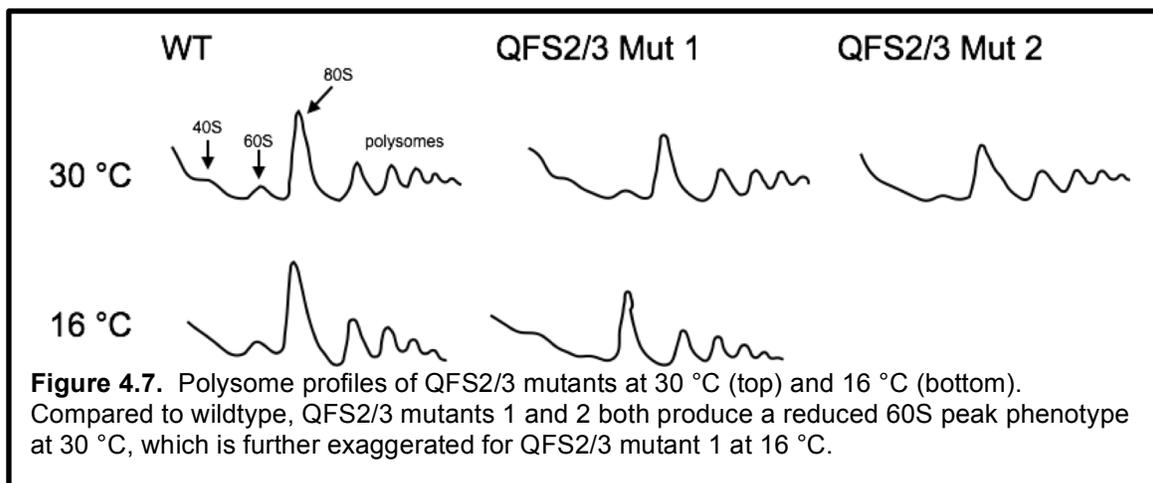
Table 4.3. QFS2/3 mutant growth phenotypes as determined by standard spotting assays. Plates were incubated at 30 °C and 18 °C.

Mutant	Growth Phenotype
QFS2/3 mut 1	cold sensitive
QFS2/3 mut 2	like wildtype
QFS2/3 mut 3	lethal*

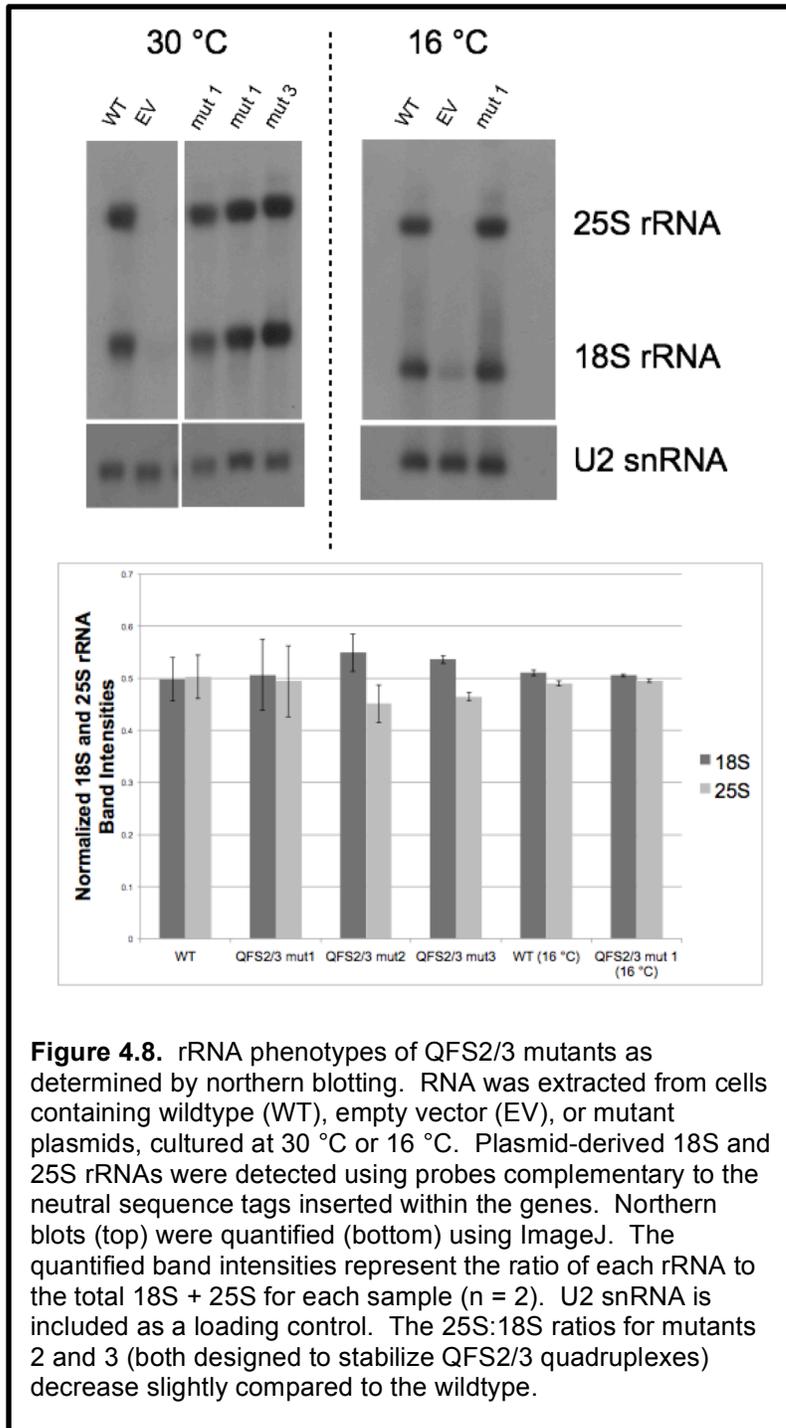
*No viable colonies after replica-plating on C-trp + 5-FOA



Although mutant growth phenotypes were quite varied, subsequent polysome analysis revealed similarities in 60S subunit defects between QFS1 mutants 1 and 2 (Figure 4.7). Polysome profiles were generated following sucrose gradient fractionation of cell extracts. At 30 °C, both mutants have visibly less robust 60S peaks upon qualitative comparison to wildtype, which is consistent with the location of the mutations in the 25S rRNA (primary component of the 60S large ribosomal subunit). The presence of 80S monosome and polysome (which represent translating ribosomes) peaks at 30 °C aligns with the growth phenotypes, confirming that sufficient functional ribosomes are made to support cell growth. At 16 °C, the QFS2/3 mutant 1 60S defect is exaggerated such that the peak is almost completely lost, which is consistent with the cold-sensitivity of this mutant. The 80S and polysome peaks present in this mutant are likely the result of residual functional ribosomes produced prior to the shift to the colder temperature, suggesting that these mutants are alive in the cold but are not able to grow or proliferate due to defects in 60S biogenesis and/or stability.



Mutations at QFS2/3 designed to stabilize quadruplexes decrease the 25S:18S rRNA ratio. To assay the effects of the QFS2/3 mutations on rRNA

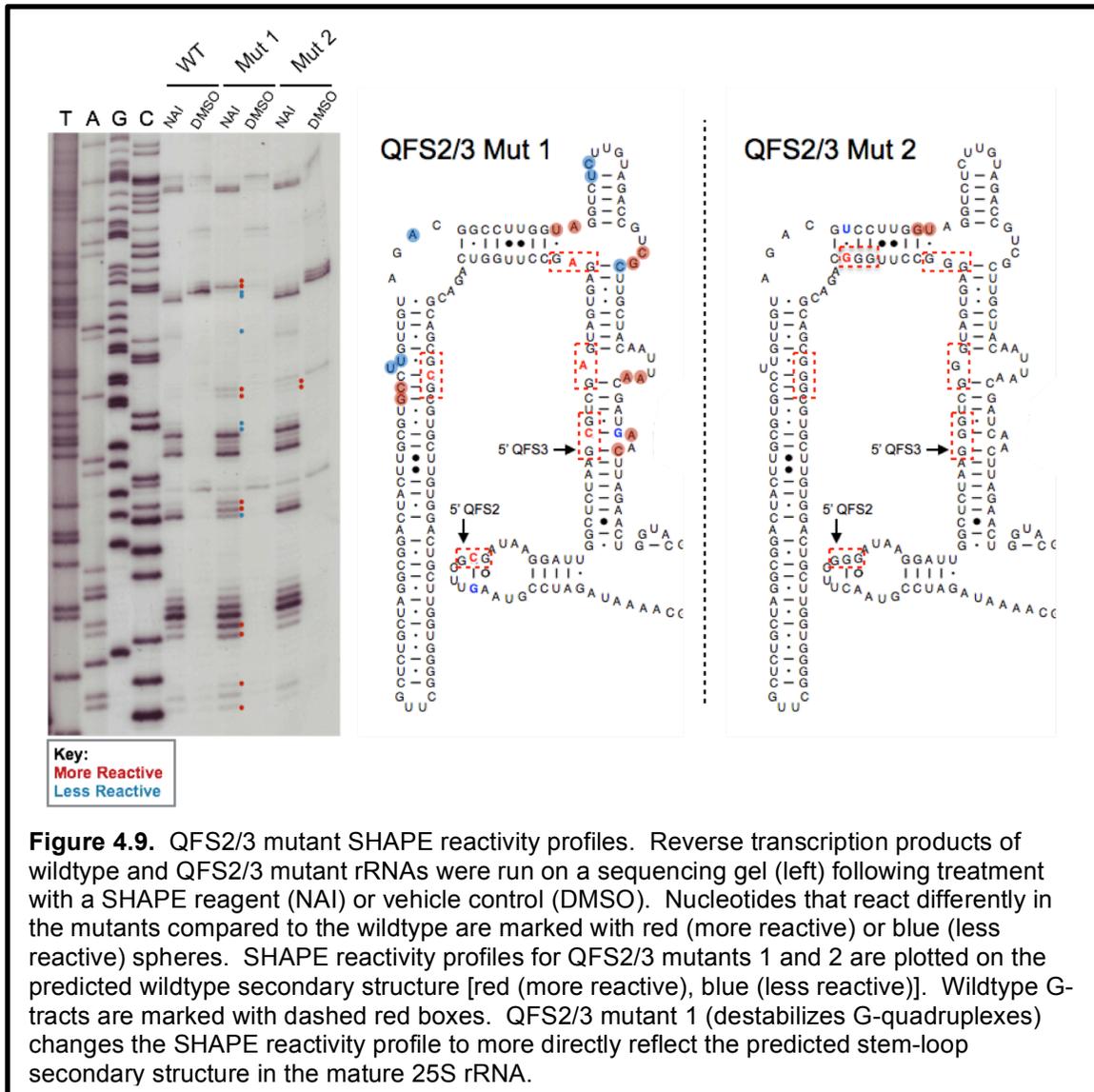


biogenesis/stability, we transformed mutant vectors into wildtype cells and detected the plasmid-derived rRNAs by northern blotting using oligonucleotides complementary to neutral sequence tags in the 18S and 25S rRNAs (Figure 4.8). This protocol allowed for analysis of all three mutants, including the lethal QFS2/3 mutant 3. Although none of the mutants produced drastic rRNA

phenotypes at 30 °C, both mutants designed to stabilize the QFS2/3 quadruplex resulted in a decrease in the 25S:18S ratio, suggesting a slight defect in 60S biogenesis/stability. We could not detect an rRNA phenotype for QFS2/3 mutant 1, suggesting that the level of functional rRNA production is close to that of wildtype cells. A similar phenotype was observed for QFS2/3 mutant 1 at 16 °C, likely due to the presence of residual functional ribosomes made prior to the shift to the colder temperature. Additional *in vitro* studies of oligonucleotides corresponding to these mutants will be required to ascertain whether they actually form more stable G-quadruplex structures than the wildtype motif.

QFS2/3 mutants 1 and 2 influence the structure of ES27L. Thus far, our results suggest that the QFS2/3 mutants cause defects in 60S subunit biogenesis and/or stability, hinting at the role of this region of the rRNA in the structural and functional integrity of the ribosome. In order to gain a more detailed understanding of the influence of our mutations on the structure of the QFS2/3 region of the rRNA, we completed SHAPE (selective 2'-hydroxyl acylation analyzed by primer extension) analysis of QFS2/3 mutants 1 and 2 in the chromosomal rDNA deletion strain. Specifically, whole-cell RNA was extracted from cells treated with NAI (2-methylnicotinic acid imidazolide), which acylates 2' hydroxyl groups of unconstrained or unpaired nucleotides[13]. SHAPE reactivity profiles were produced by electrophoretic separation of reverse transcription products of the modified rRNA (Figure 4.9). Originally, we intended to assay the full length of the stem-loops that house QFS2/3, but the long, central stem was not amenable to RT primer annealing (data not shown), limiting our

analysis to nucleotides downstream of QFS2/3. Importantly, the SHAPE reactivity profiles of the mutants are not drastically different from that of the wildtype, confirming that the mutations did not completely distort the predicted stem-loop secondary structure of this region of the rRNA. However, both mutants produced nucleotides with differential reactivity compared to wildtype. Specifically, we detected only two nucleotides with enhanced reactivity in the QFS2/3 mutant 2, which are localized close to the mutation site in the predicted secondary structure. Conversely, SHAPE reactivity in QFS2/3 mutant 1 increased specifically in the loop regions, suggesting that this mutant may be more predisposed to adopting this structure than even the wildtype. Several nucleotides in this mutant were less modified than the corresponding wildtype nucleotides, which may have resulted from changes in base-pairing or stacking interactions caused by the G-tract or compensatory mutations.



4.4 Discussion

Although G-quadruplex motifs are prevalent in the ribosomal genes and rRNA of eukaryotes[2], the functions of these quadruplexes are not yet understood. With this study, we have initiated *in vivo* analyses of the effects of several quadruplex motif mutants on cell growth, ribosome/rRNA biogenesis, and rRNA structure. Together, our results suggest that these motifs are needed for large

ribosomal subunit biogenesis and or stability, and may serve a functional role at the rRNA levels.

Noncanonical quadruplex motifs in the yeast rDNA fold into stable G-quadruplex structures in vitro. While yeast rDNA is rich in G-tracts, we were unable to find any canonical, 3-tetrad quadruplex motifs within the rDNA repeating unit. After modifying our search algorithm to increase the maximum sequence length to 50 nt (as described in Chapter 2), we identified four long-looped quadruplex motifs in the 18S and 25S genes. As we have demonstrated and discussed in Chapters 2 and 3, these noncanonical motifs fold into intramolecular quadruplex structures *in vitro*[14]. The thermal stability of each quadruplex is greatly enhanced by molecular crowding with PEG 200, providing evidence that the folding of these structures may be even more favorable in the crowded intracellular environment than in dilute solution. Further, the reduced stability of these structures compared to canonical quadruplex motifs could support their potential function(s) as regulatory switches, which may require transitioning between two or more secondary structure conformations to function properly at the rDNA and/or rRNA levels.

QFS2/3 quadruplex-stabilizing mutants suppress 25S rRNA biogenesis or stability. We have shown that QFS2/3 mutant 2 does not produce any apparent growth defects under idealized laboratory conditions, but we did detect an approximately 10% decrease in the amount of 25S rRNA (normalized to total 18S + 25S rRNAs) for each of the quadruplex-stabilizing mutants, as well as a reduced 60S peak in the polysome profile of QFS2/3 mutant 2. Of the three

mutants, QFS2/3 mutant 2 contains the fewest changes to the primary rRNA sequence. Our SHAPE data suggest that these mutations can influence, albeit mildly, the secondary structure of ES27L in which they are located. An important point to remember is that the hydrolysis rate of the SHAPE reagent, NAI is quite slow ($t_{1/2} = 33$ min)[13], which means that the SHAPE reactivity profile results from the averaged SHAPE reactivities of a heterogeneous population of structures. In *S. cerevisiae*, complete deletion of ES27L is known to be lethal and analysis of mutations within this region has identified structural features important for normal biogenesis and 25S rRNA stability[15]. Specifically, ES27L includes structural features required for processing of the ITS2 spacer sequence (separates 5.8S and 25S rRNAs in the unprocessed rRNA transcript) and, therefore, disruption of these structures impedes biogenesis and stability of these rRNAs[15]. Further studies will be required to identify whether QFS2/3 mutant 2 causes similar defects in ribosome assembly, which could be confirmed by detection of accumulating 5.8S/25S rRNA precursors. QFS2/3 mutant 3 appeared to follow a similar pattern as mutant 2 in regards to rRNA phenotype, but its apparent lethality precluded further study of this mutant.

QFS2/3 mutant 1 compromises ES27L structure and 60S subunit biogenesis/stability. QFS2/3 mutant 1 is designed to destabilize G-quadruplex formation in this region of the rRNA. Northern blot analysis of 18S and 25S rRNAs from this mutant did not detect any apparent defects in rRNA biosynthesis. Cells expressing this mutant grow normally at physiological temperature, confirming that it supports synthesis of sufficient ribosomes for

normal growth. However, polysome profiles of this mutant are abnormal (reduced 60S peak), especially at 16 °C, which is consistent with its cold sensitivity growth phenotype. Cold sensitivity may result from folding rRNA becoming stuck in kinetic traps, creating nonproductive 60S subunits. Our SHAPE data clearly support the hypothesis that QFS2/3 mutant 1 affects the structure of ES27L, likely preventing proper function of this region of the rRNA.

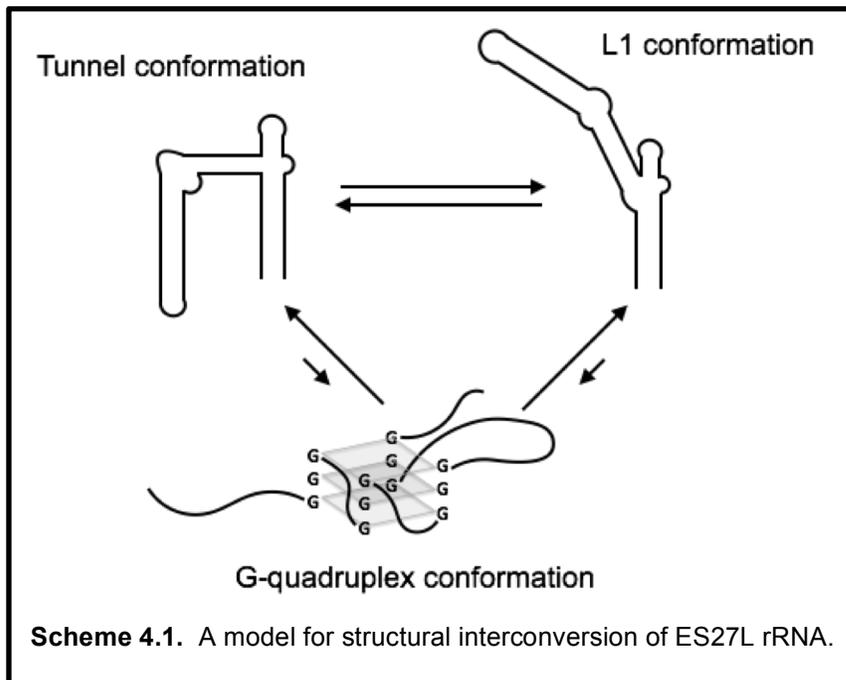
ES27L is an important functional component of the pre-60S particle and mature ribosome. ES27L is an essential[15] and highly dynamic[10, 16, 17] component of the 25S rRNA, with multi-layered functionality in ribosome assembly and function. While ES27L is required for rRNA processing[15], its dynamic nature allows it to be involved in other processes as well. Specifically, the 60S-specific assembly factor, Arx1 associates, in part, with ES27L in the immature, pre-60S particle[11, 18, 19]. Interaction of Arx1 (in combination with assembly factor, Alb1) reduces the flexibility of ES27L, holding it in a conformation that is positioned toward the polypeptide exit tunnel[18, 19]. This 'tunnel conformation' has been suggested to be important for prevention of the premature association of factors that process the nascent polypeptide with the immature particle[18]. Cryo-EM studies have also detected a population of pre-60S particles with Arx1, but lacking Alb1, in which ES27L is much less ordered and appears to be directed towards the L1 stalk structural component of the pre-ribosome (L1 conformation)[19].

Interplay between the tunnel and L1 conformations of ES27L has been previously described as important for mature ribosome function[16, 17]. In yeast,

these two conformations have been suggested to play a role in the coordination with extra-ribosomal factors that interact with the polypeptide exit tunnel and nascent peptide chain, such as chaperones, enzymes, signal recognition particle, and the endoplasmic reticulum translocon[16]. Weak density of this region further evidenced the existence of a series of different conformational states, rather than two distinct structures[16, 17]. A higher resolution (5.5Å) cryo-EM structure of a translating plant (*Triticum aestivum*) ribosome has provided additional insight into the interconversion of ES27L structures[17]. In this structure, ES27L was found to be associated with several ribosomal proteins and occupied a different, intermediate conformation from those described in yeast, which nevertheless allowed for molecular modeling of the tunnel and L1 conformations[17].

A model G-quadruplex structural intermediate within ES27L. Taken together, these studies demonstrate that ES27L participates in a variety of biological processes and that the conformational flexibility of the rRNA region is required for the proper execution of these functions. In yeast ES27L, we have found an especially high density of G-tracts, potentially capable of *in vivo* G-quadruplex formation. Although the predicted stem-loop structure is likely the primary conformation of this region, the sheer number of ES27L structural conformations and interacting partners during both assembly and translation could allow for the formation of a quadruplex-based structural intermediate during one or more of these processes. Formation of a quadruplex may even be required to allow proper release of assembly factors or extra-ribosomal factors involved in nascent

chain interactions. We propose a model wherein a specific subpopulation of rRNAs may take on a G-quadruplex fold within ES27L, and that this structure may be supported by interactions with intercellular factors (Scheme 4.1). QFS2/3 mutant 1 should preclude quadruplex formation, resulting in an increase in the population of particles having ES27L in a stem-loop conformation. Our SHAPE data support this conclusion, as we see increased reactivity in nucleotides predicted to be unpaired in the secondary structure. Our quadruplex-stabilizing mutant does not produce a similar SHAPE reactivity profile. Further studies will be required to ascertain the validity of this model and identify candidate interacting partners that may facilitate G-quadruplex formation at this site.



4.5 Conclusion

Eukaryote specific expansion segments, such as ES27L are notoriously variable in both length and sequence across organisms (reviewed in Wilson and

Doudna Cate, 2012)[20]. For example, in yeast and plants ES27L is approximately 150 nts long, while human ES27L is closer to 700 nts in length[12]. This variability suggests different functional roles of this expansion segment in more complex, higher eukaryotes. Nevertheless, the QFS2/3 component of ES27L may yet provide interesting insights into the applicability of G-quadruplex structures to biological systems. Further studies of the other quadruplex motifs identified in yeast rDNA will also be needed to gain a complete understanding of these structures in gene expression and ribosome biosynthesis.

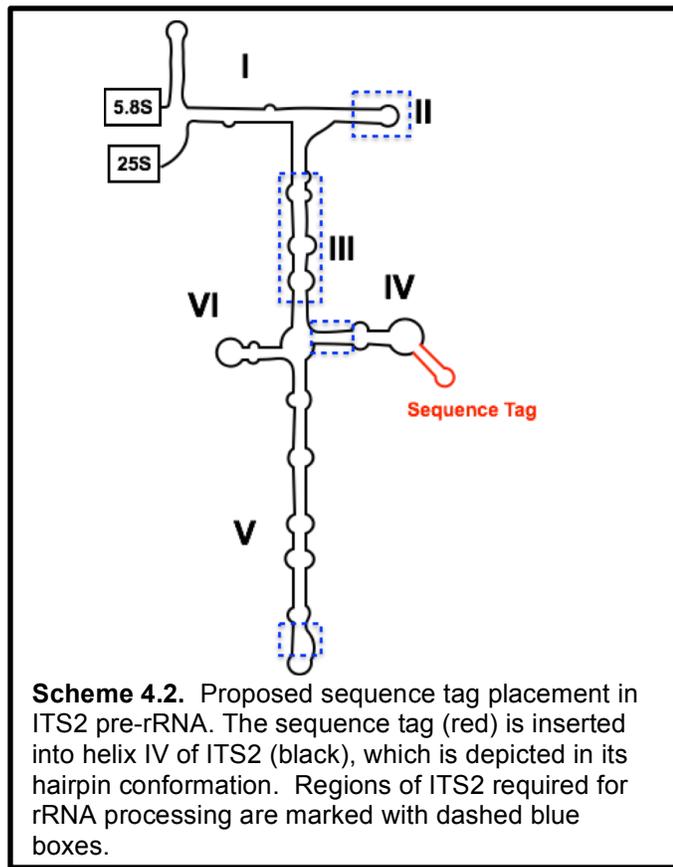
4.6 Future Directions

rDNA mutants are included in studies of ribosome biogenesis much less frequently than r protein or assembly factor mutants partly due to the technical challenges associated with introducing and assaying these mutations. However, rRNAs are the primary structural and catalytic components of the ribosome, which necessitates the development of better technologies that can be used to address this problem. The rDNA mutagenesis strategies discussed in this chapter are limited in their capacity to address lethal mutants, which can arguably provide the most useful information regarding the roles of rRNAs in ribosome biogenesis and function. Therefore, future efforts should be focused on creating a mechanism for PNA-based affinity purification of plasmid-derived pre-ribosomes.

We have demonstrated that transformation of tagged, mutant rDNA vectors into wildtype cells (Figure 4.1) is an effective way to study the rRNA phenotypes of mutant rDNAs, without perturbation of endogenous rRNA production.

However, this protocol only allows for analysis of mature 18S and 25S rRNAs, because there is no mechanism to separate the plasmid-derived and wildtype rRNAs. This problem can be overcome by insertion of a neutral sequence tag into the ITS2 (internal transcribed spacer 2) rRNA that separates the 5.8S and 25S large subunit rRNAs in the primary transcript, and is present during the nucleolar stages of rRNA processing[1]. The most abundant pre-ribosomal particles with ITS2 contain the large subunit pre-rRNAs (5.8S and 25S) separated from the small subunit pre-rRNA (18S) and, therefore, can be used to directly study early steps in large ribosomal subunit assembly. Insertion of a neutral sequence tag into this site could then facilitate oligomer-based affinity purification and analysis of pre-60S subunits.

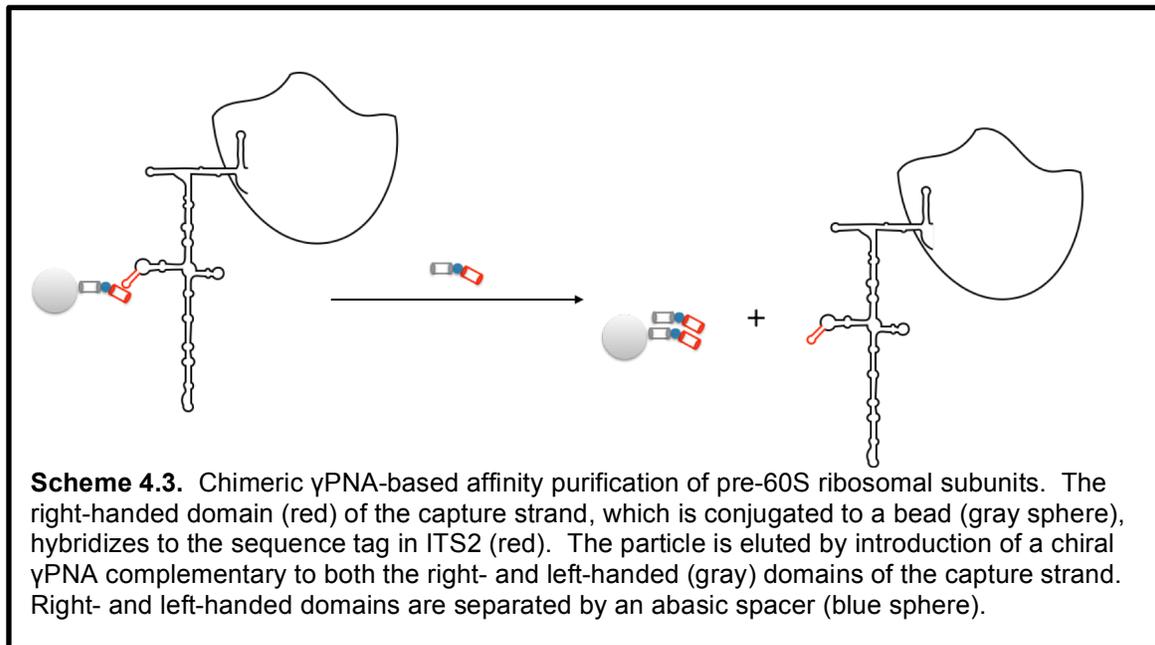
ITS2 contains secondary and tertiary structure elements that are required for proper processing and 60S subunit biogenesis[21]. Generally, these *cis*-elements correlate with the most evolutionarily conserved nucleotides of the spacer[21]. Therefore, location of a sequence tag must be carefully selected so as not to perturb function of ITS2. ITS2 has been described as



having two possible secondary structures, a ring conformation and a hairpin conformation (reviewed in Fernández-Pevida *et al.*, 2015)[22]. For simplicity, I will refer to the hairpin conformation of ITS2. The best candidate location for insertion of a neutral sequence tag is the terminal stem-loop in helix IV of ITS2 (approx. 80 nts downstream of the 5' end; Scheme 4.2). Careful analysis will be required to determine the best nucleotide sequence for such a tag, but a partial (or complete) insertion of the tag we have previously utilized in the 18S gene of our rDNA mutagenesis vector, which is effective for hybridization in northern blots, may work well for this purpose.

Helical, γ PNA-based probes[23] (discussed in detail in Chapter 3) complementary to the ITS2 sequence tag can be used for affinity purification of pre-ribosomes. A similar mechanism for purification of pre-30S bacterial ribosomes has been previously developed using oligonucleotide-based affinity purification[24]. However, the eukaryotic rRNA is much longer and more complex than its bacterial counterpart, which may increase the chance of off-target interactions with high affinity γ PNAs. The solution to this problem lies in the chimeric γ PNA probes currently under development in the Armitage group (T. Canady, personal communication). Specifically, a γ PNA probe can be designed having a right-handed domain (complementary to the sequence tag) and a left-handed domain separated by an abasic (diethylene glycol) spacer. The PNA can be conjugated to a bead and the right-handed domain will serve to trap particles containing the complementary spacer sequence. These particles can be eluted off the column using a chimeric γ PNA containing right- and left-handed domains

both complementary to the capturing PNA probe (Scheme 4.3). Inclusion of left-handed regions in these probes eliminates potential off-target interactions between these domains and natural nucleic acids, which should greatly enhance the purity of any particles extracted using this method.



The ability to effectively purify pre-ribosomal particles containing mutant rRNA will greatly enhance our ability to study the functions of rRNAs in assembly and aid in developing a more complete model of the ribosome biogenesis pathway. Apart from ribosomes, similar methods could be employed to study biogenesis and function of other cellular ribonucleoproteins, enhancing our understanding of the complex roles of RNAs within biological systems.

4.7 Materials and Methods

4.7.1 Materials

DNA oligonucleotides were purchased from Integrated DNA Technologies (www.idtdna.com) and Fisher Scientific (www.fishersci.com).

4.7.2 Equipment

UV melting data were collected on a Varian Cary 3 spectrophotometer combined with a temperature controlled multicell holder. A Jasco J-715 spectropolarimeter with a single cell holder and water circulating temperature controller was used to collect CD data.

4.7.3 UV Melting

Samples (1 μ M DNA) were buffered in 100 mM KCl, 10 mM Tris pH 7, and 0.1 mM Na₂EDTA, unless otherwise noted. Molecular crowding was approximated through the inclusion of 40% w/v PEG 200, or 40% v/v glycerol control. Samples were heated to 95 °C and then annealed by slow cooling to 15 °C (1 °C/min), followed by slow melting by heating to 95 °C while absorbance was measured at 295 nm. Melting temperatures were calculated by averaging the minimum of the first derivative of each melting curve, in triplicate (OriginPro8).

4.7.4 Circular Dichroism

CD data were collected at yeast physiological temperature, 30 °C. Each spectrum was obtained by averaging 6 scans collected at 100 nm/min.

4.7.5 Plasmid Mutagenesis

Mutations were introduced by PCR-based site-directed mutagenesis using the QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent Technologies). Dpn1-digested PCR products were transformed into XL10-Gold ultracompetent

E. coli. Plasmid DNA was extracted from the *E. coli* by miniprepping and mutations were confirmed by DNA sequencing (Genewiz; www.genewiz.com).

4.7.6 Yeast Strains for rDNA Mutant Analysis

All strains used in this study are listed in Supplemental Table 4.S1. The rDNA mutagenesis vector was derived from the pJD180.TRP plasmid, kindly provided by J. Dinman. pJD180.TRP is a 2 μ m plasmid containing a complete copy of the rDNA repeat[8]. Neutral sequence tags were introduced by serial site-directed mutagenesis into the 18S (5'-AGCCCCCTAGTTGGATCCTCGGCG-3') and 25S (5'-CTGGTAGGAAGCTGCAGCCTGCCC-3')-templating regions of the gene, as has been previously described for the pWL160 tagged rDNA vector[7], to produce the pJD180T rDNA vector. Quadruplex motif mutations were introduced directly into pJD180T plasmid. All mutations are listed in Supplemental Table 4.S2.

Growth, polysome, and SHAPE analyses were completed using *rdn1* $\Delta\Delta$ cells in combination with a wildtype or mutant pJD180T plasmid. *rdn1* $\Delta\Delta$ lack all chromosomal rDNA repeats and carry a high copy rDNA plasmid (pNOY290; *URA3*)[7]. *rdn1* $\Delta\Delta$ was transformed with wildtype or mutant pJD180T; loss of the *URA3* vector was promoted by two rounds of replica-plating the transformants on C-trp + 5-FOA (5-fluoroorotic acid), followed by streaking on C-trp and C-ura to confirm the plasmid substitution.

Northern blot analysis of quadruplex mutants was completed using wildtype cells (BY4742 *trp1* Δ) in combination with a wildtype or mutant pJD180T plasmid. The empty vector, pRS314 (*TRP1*), was included a negative control. 20 mL

cultures in C-trp were incubated overnight and diluted to $OD_{600} = 0.2$ the following morning. Cells were incubated in an air shaker or water bath at the desired temperature until they reached an approximate $OD_{600} = 0.6$. Cells were harvested by centrifugation and stored at $-80\text{ }^{\circ}\text{C}$ until RNA extraction.

4.7.7 Standard Spotting Assay for Cell Growth

2 mL cultures in C-trp were incubated overnight at $30\text{ }^{\circ}\text{C}$. Cultures were normalized to the lowest OD_{600} and 5 serial dilutions of each culture were prepared in C-trp. $8\text{ }\mu\text{L}$ of each dilution were plated on C-trp and C-ura and incubated at the desired temperature.

4.7.8 Polysome Profiles

Large volume (100 mL) cell cultures in C-trp were incubated at the desired temperature until reaching $OD_{600} = 0.6-0.7$. Cell cultures were treated with 5mg cycloheximide to prevent ribosome dissociation from mRNA transcripts. Cells were harvested by centrifugation and lysed with acid washed glass beads. Cell extracts were clarified by centrifugation and loaded onto sucrose gradients (7 and 47% sucrose) for fractionation by ultracentrifugation.

4.7.9 Northern Blotting

Whole cell RNA was prepared by phenol/chloroform/isoamyl alcohol extraction and ethanol precipitation. RNA pellets were resuspended in Nanopure water, and normalized to approximately $1\text{ }\mu\text{g}/\mu\text{L}$. $5\text{ }\mu\text{g}$ RNA samples were separated on a large, 1.2% agarose gel. The RNA was placed on a nitrocellulose membrane by capillary transfer, which was stabilized by UV crosslinking. Northern blotting was done using $\gamma^{32}\text{P}$ -ATP-labeled DNA

oligonucleotides complementary to the 18S and 25S sequence tags or U2 snRNA control. Membranes were incubated with labeled oligonucleotides and washed, followed by exposure to film for visualization of RNA bands. Band intensities were quantified using ImageJ 1.48v.

4.7.10 *In vivo* SHAPE

NAI was prepared by drop-wise addition of 927 mg 1,1'-carbonyldiimidazole (dissolved in 3 mL DMSO) to 822 mg 2-methylnicotinic acid (dissolved in 3 mL DMSO). The reaction was incubated, stirring at room temperature for 1 hour. NAI was stored at -80 °C until use in SHAPE.

Cell cultures were grown to $OD_{600} = 0.6$ and harvested by centrifugation. Cells were resuspended in 5 mL 1x PBS and 555 μ L DMSO or NAI was added directly to the cells. Cells were incubated rocking at 30 °C for 20 minutes. After the SHAPE treatment, cells were harvested by centrifugation and washed with 5 mL cold Nanopure water. RNA was prepared by phenol/choloroform/isoamyl alcohol extraction and ethanol precipitation.

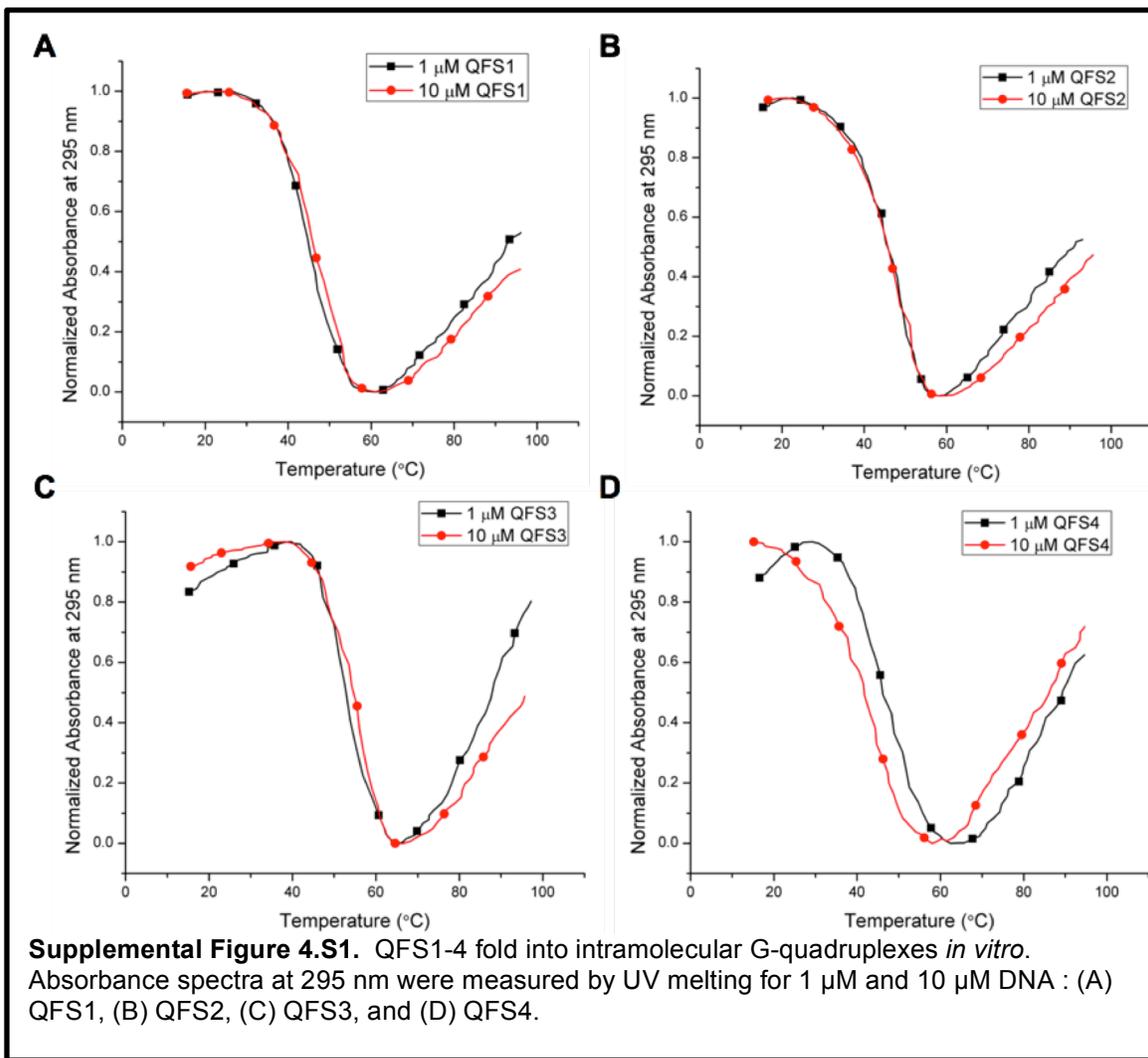
The RT (reverse transcription) primer (5'-CAGTCAGATTCCCCTTGTCCGTACC-3') was labeled using γ 32P-ATP and T4 Polynucleotide Kinase (New England Biolabs). RT reactions were completed for each sample using Transcriptor Reverse Transcriptase (Roche). Sequence reactions were completed using untreated RNA samples extracted from a wildtype cell line (W303). Following the RT, RNA was hydrolyzed using NaOH and the samples were ethanol precipitated. RT products were resuspended in

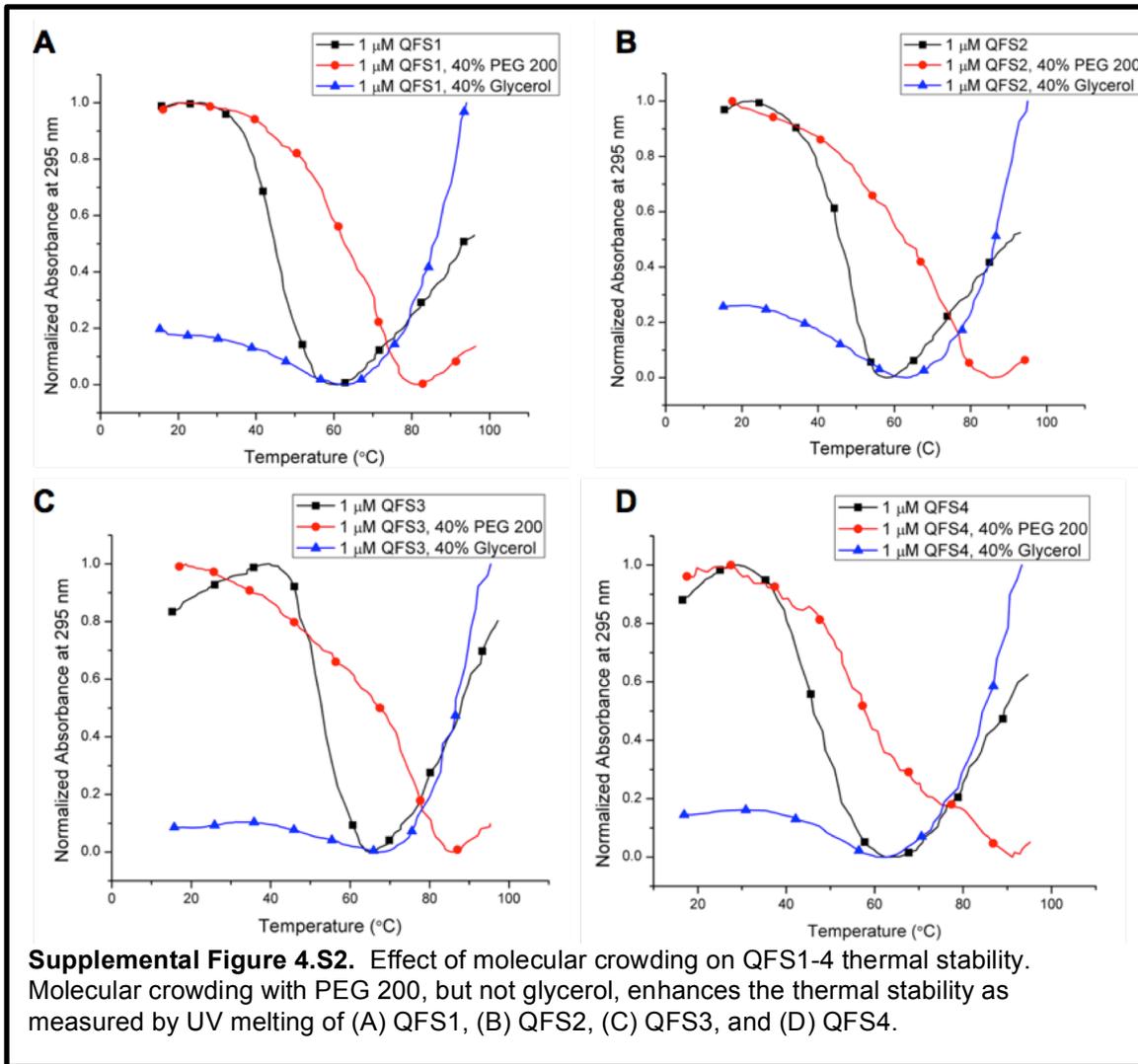
loading dye and analyzed by electrophoretic separation, followed by exposure to a film for visualization of RT product bands.

4.8 Acknowledgements

We are grateful to the F. Bradley Johnson lab at the University of Pennsylvania for the G-quadruplex prediction software and the Jonathan Dinman lab at the University of Maryland for generously providing us with yeast strains and plasmids. We also thank the David Scaife Family Charitable Foundation for financial support of this research (Award 141RA01).

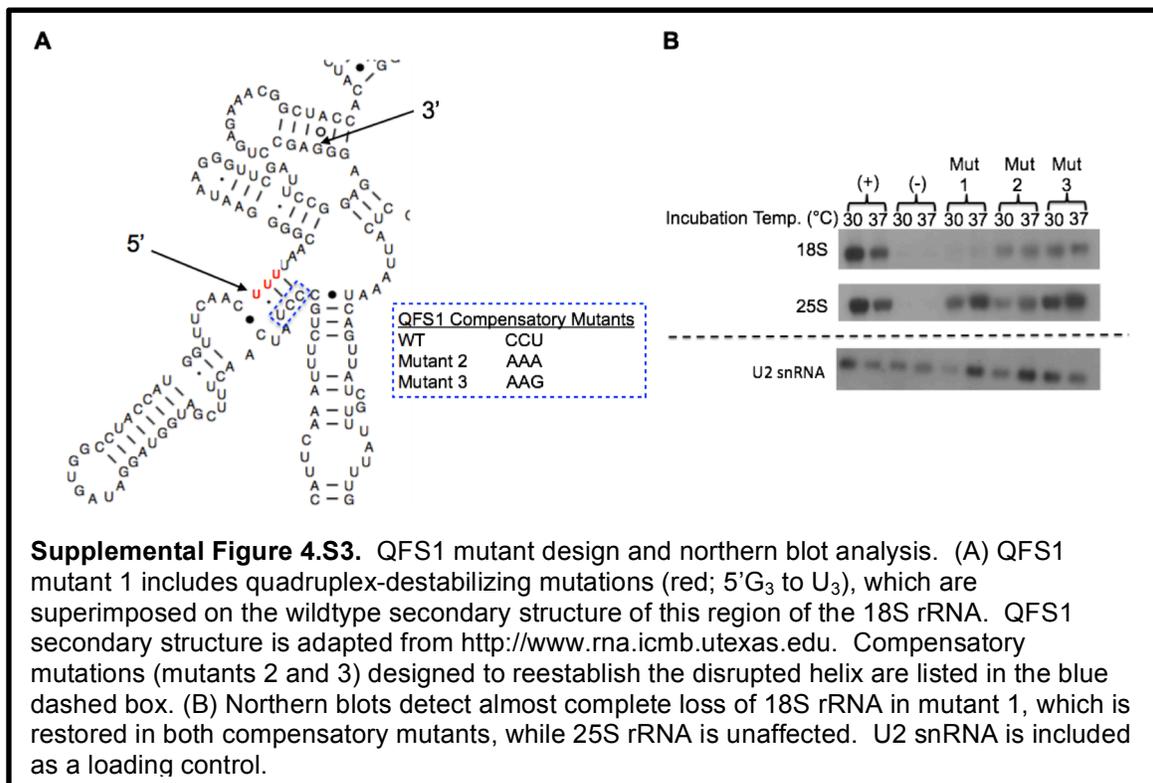
4.9 Appendix





QFS1 mutants were introduced and assayed in an alternative rDNA mutagenesis system combining a tagged rDNA vector with an RNA polymerase I temperature sensitive (inactive at 37 $^{\circ}$ C) mutant yeast strain[7]. In this system, the rDNA vector is under the control of a *GAL* promoter. Therefore, when cells are grown in galactose and shifted to the higher temperature, only plasmid-derived rRNA is produced (although residual wildtype ribosomes are retained within cells). QFS1 mutant 1 produces an rRNA phenotype characterized by a severe depletion in 18S rRNA. This phenotype is not likely the result of a defect

in rRNA transcription, as 18S and 25S rRNAs are derived from the same transcriptional unit, and we do not detect a similar depletion of 25S rRNA. Further, both mutants that were designed to reestablish the helix in the mature ribosome that is destabilized by QFS1 mutant 1 compensate for this phenotype, suggesting that it results from disruption of the mature rRNA, rather than loss of a quadruplex. We chose not to pursue further analysis of these mutants for this reason.



Supplemental Table 4.S1. *S. cerevisiae* strain list.

Number	Date Frozen	Strain	Plasmid	Media
JWY 11001	29-Aug-13	NOY504	(none)	YEPD
JWY 11002	29-Aug-13	NOY504	pRS314	C -trp Gal
JWY 11003	29-Aug-13	NOY504	pWL160 QFS1 G3 1 to T3	C -trp Gal
JWY 11004	6-Nov-13	NOY504 <i>leu2Δ</i> , <i>trp1Δ</i>	(none)	YEPD
JWY 11005	19-Nov-13	NOY504 <i>leu2Δ</i> , <i>trp1Δ</i>	pWL160	C -trp Gal

JWY 11006	19-Nov-13	NOY504 <i>leu2Δ</i> , <i>trp1Δ</i>	pRS314	C -trp Gal
JWY 11007	19-Nov-13	NOY504 <i>leu2Δ</i> , <i>trp1Δ</i>	pWL160 QFS1 G3 1 to T3	C -trp Gal
JWY 11008	3-Mar-14	NOY504 <i>leu2Δ</i> , <i>trp1Δ</i>	pWL160 QFS1 comp AAA	C -trp
JWY 11009	3-Mar-14	NOY504 <i>leu2Δ</i> , <i>trp1Δ</i>	pWL160 QFS1 comp AAG	C -trp
JWY11010	21-Mar-14	NOY504 <i>leu2Δ</i>	(none)	YEPD
JWY11011	3-Apr-14	NOY504 <i>leu2Δ</i>	pRS315	C -leu
JWY11012	3-Apr-14	NOY504 <i>leu2Δ</i>	pRS314	C -trp
JWY11013	3-Apr-14	NOY504 <i>leu2Δ</i>	pWL160	C -trp
JWY11014	3-Apr-14	NOY504 <i>leu2Δ</i>	pWL160 QFS1 G3 1 to T3	C -trp
JWY11015	10-Apr-14	NOY504	pWL160 QFS1 comp AAA	C -trp
JWY11016	10-Apr-14	NOY504	pWL160 QFS1 comp AAG	C -trp
JWY11017	11-Apr-14	NOY504 <i>leu2Δ</i>	pWL160 QFS1 comp AAA	C -trp
JWY11018	11-Apr-14	NOY504 <i>leu2Δ</i>	pWL160 QFS1 comp AAG	C -trp
JWY11019	29-Apr-14	NOY504	pWL160	C -trp
JWY11020	29-Jul-14	W303	(none)	YEPD
JWY11021	29-Aug-14	<i>sgs1Δ</i>	(none)	YEPD
JWY11022	21-Oct-14	BY4742	(none)	YEPD
JWY11023	18-Nov-14	W303	pRS314	C -trp
JWY11024	18-Nov-14	W303	pJD180.trpT (pJD180T)	C -trp
JWY11025	25-Nov-14	BY4742 <i>trp1Δ</i>	(none)	C-His
JWY11026	13-Dec-14	BY4742 <i>trp1Δ</i>	pJD180T	C -trp
JWY11027	13-Dec-14	BY4742 <i>trp1Δ</i>	pRS314	C -trp
JWY11028	6-Jan-15	<i>sgs1Δ trp1Δ</i>	pJD180T	C -trp
JWY11029	6-Jan-15	<i>sgs1Δ trp1Δ</i>	pRS314	C -trp
JWY11030	27-Jan-15	<i>sgs1Δ trp1Δ</i>	(none)	YEPD
JWY11031	2-Feb-15	BY4742 <i>trp1Δ</i>	pJD180T QFS1 mut	C -trp
JWY11032	2-Feb-15	BY4742 <i>trp1Δ</i>	pJD180T QFS1 comp AAA	C -trp
JWY11033	2-Feb-15	BY4742 <i>trp1Δ</i>	pJD180T QFS1 comp AAG	C -trp
JWY11034	24-Feb-15	<i>rdn1ΔΔ</i> (yJD1045)	pNOY353 (<i>GAL</i> 35S rDNA; <i>TRP1</i>)	YEPD
JWY11035 (x2)	14-Aug-15	<i>rdn1ΔΔ</i> (yJD1044)	pNOY290 (<i>URA3 leu2Δ</i> rDNA hygr 2μ)	YEPD
JWY11036	4-Mar-15	BY4742 <i>trp1Δ</i>	pJD180T QFS1 comp HS	C -trp
JWY11037	11-Mar-15	<i>rdn1ΔΔ</i>	pJD180T	C -trp
JWY11038	27-Mar-15	<i>rdn1ΔΔ</i>	pJD180T QFS1 mut	C -trp
JWY11039	31-Mar-15	<i>rdn1ΔΔ</i>	pJD180T QFS1 comp AAA	C -trp
JWY11040	2-Apr-15	<i>rdn1ΔΔ</i>	pJD180T QFS1 comp AAG	C -trp
JWY11041	2-Apr-15	<i>rdn1ΔΔ</i>	pJD180T QFS1 comp HS	C -trp
JWY11042	14-Aug-15	BY4742 <i>trp1Δ</i>	pJD180T QFS2/3 mut 1	C -trp
JWY11043	14-Aug-15	BY4742 <i>trp1Δ</i>	pJD180T QFS2/3 mut 2	C -trp
JWY11044	14-Aug-15	BY4742 <i>trp1Δ</i>	pJD180T QFS2/3 mut 3	C -trp
JWY11045	9-Oct-15	<i>rdn1ΔΔ</i>	pJD180T	C -trp

JWY11046	16-Oct-15	<i>rdn1ΔΔ</i>	pJD180T QFS2/3 mut 1	C -trp
JWY11047	16-Oct-15	<i>rdn1ΔΔ</i>	pJD180T QFS2/3 mut 2	C -trp

Supplemental Table 4.S2. *S. cerevisiae* rDNA G-quadruplex motif mutations.

Name	Description	Location of Mutation
QFS1 mutant 1	Destabilize quadruplex by replacing 5' G ₃ with T ₃ (U ₃)	18S rDNA/rRNA
QFS1 mutant 2	Mature rRNA 2° structure compensatory mutant (CCU to AAA)	18S rDNA/rRNA
QFS1 mutant 3	Mature rRNA 2° structure compensatory mutant (CCU to AAG)	18S rDNA/rRNA
QFS2/3 mutant 1	Designed to destabilize quadruplexes (eliminates central G of each G ₃ -tract)	25S rDNA/rRNA
QFS2/3 mutant 2	Designed to stabilize quadruplexes (additional G ₃ in long loop of QFS3)	25S rDNA/rRNA
QFS2/3 mutant 3	Designed to stabilize quadruplexes (G ₃ -tracts to G ₄)	25S rDNA/rRNA

4.10 References

1. Woolford, J.L., Jr. and S.J. Baserga, *Ribosome biogenesis in the yeast Saccharomyces cerevisiae*. Genetics, 2013. **195**(3): p. 643-81.
2. Hanakahi, L.A., H. Sun, and N. Maizels, *High affinity interactions of nucleolin with G-G-paired rDNA*. J. Biol. Chem., 1999. **274**(22): p. 15908-15912.
3. Hershman, S.G., et al., *Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in Saccharomyces cerevisiae*. Nucleic Acids Res, 2008. **36**(1): p. 144-56.
4. Bochman, M.L., K. Paeschke, and V.A. Zakian, *DNA secondary structures: stability and function of G-quadruplex structures*. Nat Rev Genet, 2012. **13**(11): p. 770-80.
5. Capra, J.A., et al., *G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in Saccharomyces cerevisiae*. PLoS Comput Biol, 2010. **6**(7): p. e1000861.
6. Maizels, N., *G4-associated human diseases*. EMBO Rep, 2015. **16**(8): p. 910-22.
7. Cole, S.E. and F.J. LaRiviere, *Analysis of nonfunctional ribosomal RNA decay in Saccharomyces cerevisiae*. Methods Enzymol, 2008. **449**.
8. Smith, M.W., et al., *Saturation mutagenesis of 5S rRNA in Saccharomyces cerevisiae*. Mol Cell Biol, 2001. **21**(24): p. 8264-75.

9. Musters, W., et al., *A system for the analysis of yeast ribosomal DNA mutations*. Mol Cell Biol, 1989. **9**(2): p. 551-9.
10. Ben-Shem, A., et al., *Crystal structure of the eukaryotic ribosome*. Science, 2010. **330**(6008): p. 1203-9.
11. Leidig, C., et al., *60S ribosome biogenesis requires rotation of the 5S ribonucleoprotein particle*. Nat Commun, 2014. **5**: p. 3491.
12. Cannone, J.J., et al., *The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs*. BMC Bioinformatics, 2002. **3**: p. 2.
13. Spitale, R.C., et al., *RNA SHAPE analysis in living cells*. Nat Chem Biol, 2013. **9**(1): p. 18-20.
14. Kormuth, K.A., J.L. Woolford, and B.A. Armitage, *Homologous PNA Hybridization to Noncanonical DNA G-Quadruplexes*. Biochemistry, 2016.
15. Jeeninga, R.E., et al., *Variable regions V13 and V3 of Saccharomyces cerevisiae contain structural features essential for normal biogenesis and stability of 5.8S and 25S rRNA*. RNA, 1997. **3**(5): p. 476-88.
16. Beckmann, R., et al., *Architecture of the protein-conducting channel associated with the translating 80S ribosome*. Cell, 2001. **107**(3): p. 361-72.
17. Armache, J.P., et al., *Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-Å resolution*. Proc Natl Acad Sci U S A, 2010. **107**(46): p. 19748-53.
18. Greber, B.J., et al., *Cryo-EM structures of Arx1 and maturation factors Rei1 and Jjj1 bound to the 60S ribosomal subunit*. Nat Struct Mol Biol, 2012. **19**(12): p. 1228-33.
19. Greber, B.J., et al., *Insertion of the Biogenesis Factor Rei1 Probes the Ribosomal Tunnel during 60S Maturation*. Cell, 2016. **164**(1-2): p. 91-102.
20. Wilson, D.N. and J.H. Doudna Cate, *The structure and function of the eukaryotic ribosome*. Cold Spring Harb Perspect Biol, 2012. **4**(5).
21. van Nues, R.W., et al., *Evolutionarily conserved structural elements are critical for processing of Internal Transcribed Spacer 2 from Saccharomyces cerevisiae precursor ribosomal RNA*. J Mol Biol, 1995. **250**(1): p. 24-36.
22. Fernandez-Pevida, A., D. Kressler, and J. de la Cruz, *Processing of preribosomal RNA in Saccharomyces cerevisiae*. Wiley Interdiscip Rev RNA, 2015. **6**(2): p. 191-209.
23. Sahu, B., et al., *Synthesis and characterization of conformationally preorganized, (R)-diethylene glycol-containing gamma-peptide nucleic acids with superior hybridization properties and water solubility*. J Org Chem, 2011. **76**(14): p. 5614-27.
24. Clatterbuck Soper, S.F., et al., *In vivo X-ray footprinting of pre-30S ribosomes reveals chaperone-dependent remodeling of late assembly intermediates*. Mol Cell, 2013. **52**(4): p. 506-16.