

THE PENNSYLVANIA STATE UNIVERSITY  
SCHREYER HONORS COLLEGE

DEPARTMENT OF CHEMISTRY

INVESTIGATING PROPERTIES AND CHARACTERISTICS OF G-QUADRUPLEX  
STRUCTURE, FOLDING, AND FLUORESCENCE

MADELINE E. SHERLOCK  
SPRING 2014

A thesis  
submitted in partial fulfillment  
of the requirements  
for baccalaureate degrees in Chemistry and German  
with honors in Chemistry

Reviewed and approved\* by the following:

Philip C. Bevilacqua  
Professor of Chemistry  
Thesis Supervisor

Raymond L. Funk  
Professor of Chemistry  
Honors Adviser

Scott A. Showalter  
Assistant Professor of Chemistry  
Faculty Reader in Chemistry

\* Signatures are on file in the Schreyer Honors College.

## ABSTRACT

G-Quadruplex Structures (GQS) have been of recent interest due to their unique folding and spectroscopic properties, as well as their potential for target binding, modification for uses in sensing, and possible regulatory roles *in vivo*. This thesis explores all of these categories in both RNA and DNA G-Quadruplexes. GQSs with longer G-stretch length were developed to explore intermediates along the folding pathway and their effect on folding cooperativity to create a broadened response range for possible application in detection of analytes with highly variable concentrations. The recently-discovered intrinsic fluorescence of DNA GQSs was proven to also exist in RNA GQSs, which can be used as an alternative readout signal and could have potential for fluorescent imaging. This fluorescence was further characterized in DNA GQSs by systematic variation of the G-stretch length, loop nucleotides, and loop length. Effects on emission wavelength and signal intensity were observed and compared to high resolution structural data to better understand the nature of the GQS as a fluorophore. In order to further this understanding, the effects of dangling nucleotide ends and GQS concentration as well as varied conditions, such as  $Mg^{2+}$ , temperature, quenchers, and crowding, were explored; the aim of these studies was increasing fluorescence signal toward the goal of developing an intrinsically fluorescent RNA for live cell imaging. Lastly, interactions of RNA GQSs with divalent metal ions were studied in the context of gene regulation by GQSs for heavy metal response with the finding of reversible unfolding by  $Cu^{2+}$  ions.

## TABLE OF CONTENTS

List of Figures .....	iii
List of Tables .....	iv
List of Abbreviations .....	v
Acknowledgements .....	vii
Dedication .....	ix
Chapter 1 Introduction to Nucleic Acid Folding and G-Quadruplexes .....	1
1.1 Nucleic Acid Structure .....	2
1.2 Nucleic Acids as a Biochemical Tool .....	4
1.3 G-Quadruplex Structure .....	5
1.4 GQS Roles in Biology .....	7
1.5 GQs as a Biochemical Tool .....	8
1.6 References .....	8
Chapter 2 Decrease of RNA Folding Cooperativity by Deliberate Population of Intermediates in RNA G-Quadruplexes .....	13
2.1 Abstract .....	13
2.2 Introduction .....	13
2.3 Results and Discussion .....	16
2.4 Materials and Methods .....	23
2.4.1 RNA Preparation .....	23
2.4.2 Spectroscopic Titration Experiments .....	24
2.4.3 Data Fitting .....	25
2.4.4 UV-Detected Thermal Denaturation .....	26
2.4.5 G2 GQS Mixing Simulation .....	26
2.5 Author Contributions .....	27
2.6 Acknowledgements .....	27
2.7 References .....	27
Chapter 3 The Effect of Loop Sequence and Loop Length on the Intrinsic Fluorescence of G-Quadruplexes .....	32
3.1 Abstract .....	32
3.2 Introduction .....	32
3.3 Result and Discussion .....	34
3.4 Materials and Methods .....	40
3.4.1 DNA Preparation .....	40
3.4.2 Spectroscopic Titration Experiments .....	40
3.4.3 Data Fitting .....	41
3.4.4 UV-Detected Thermal Denaturation .....	41

3.5 Author Contributions .....	42
3.6 Acknowledgements .....	42
3.7 References .....	42
Chapter 4 Investigations into the Origin of DNA G-Quadruplex Fluorescence .....	46
4.1 Abstract .....	46
4.2 Introduction .....	46
4.3 Results and Discussion .....	48
4.4 Future Directions .....	53
4.5 Materials and Methods .....	55
4.5.1 DNA Preparation .....	55
4.5.2 Spectroscopic Titration Experiments .....	55
4.6 Acknowledgements .....	56
4.7 References .....	56
Chapter 5 Heavy Metal Effects on RNA G-Quadruplex Folding and Unfolding .....	58
5.1 Abstract .....	58
5.2 Introduction .....	58
5.3 Results and Discussion: .....	60
5.3.1 Selection of Putative Quadruplex Sequences with Heavy Metal Function .....	60
5.3.2 RNA G-Quadruplex Titrations with $K^+$ ions .....	61
5.3.3 RNA G-Quadruplex titrations with Divalent and Heavy Metal ions .....	62
5.3.4 Reversible G-Quadruplex Unfolding with EDTA .....	65
5.3.5 $Cu^{2+}$ Titrations in Varying Background $K^+$ and pH .....	65
5.4 Conclusions .....	66
5.5 Future Directions .....	67
5.6 Materials and Methods .....	67
5.6.1 RNA Preparation .....	67
5.6.2 CD Spectroscopic Titration Experiments .....	68
5.6.3 Data Fitting .....	68
5.6.4 Bioinformatic Data Sets and Programs .....	70
5.7 Acknowledgements .....	70
5.8 References .....	70
Appendix A Supplementary Information for Chapter 2 .....	74
References .....	83
Appendix B Supplementary Information for Chapter 3 .....	84
References .....	92
Appendix C Supplementary Information for Chapter 4 .....	93
Appendix D Supplementary Information for Chapter 5 .....	97
ACADEMIC VITA .....	104

## LIST OF FIGURES

Figure 1.1 The Central Dogma of Molecular Biology .....	1
Figure 1.2 The Five Nucleobases of RNA and DNA .....	2
Figure 1.3 Nucleic Acid Backbone .....	3
Figure 1.4 Watson Crick Base Pairing in DNA .....	4
Figure 1.5 A Top View of G-Quadruplex Structure .....	5
Figure 1.6 Parallel GQS formation with a $K^+$ ion .....	6
Figure 1.7 Gene Regulation Controlled by a GQS in the Promoter Region .....	7
Figure 2.1 Two Approaches to Achieve Broad Response Sensing .....	15
Figure 2.2 Fraction Folded Plots for all of the G2 GQS, Displaying Different $K^+$ Binding Strengths.....	18
Figure 2.3 Fraction Folded Plot of G2m1 and its Longer G-stretch Counterparts at 262nm by CD Spectroscopy.....	20
Figure 2.4 Comparison of CD- and Fluorescence-Monitored GQS $K^+$ Titrations.....	21
Figure 3.1 Experimental Design and Fluorescence Titration Results .....	36
Figure 3.2 Biophysical Characterization of dG3T. ....	37
Figure 4.1 DNA Concentration Dependence of Fluorescent Signal Intensity. ....	50
Figure 4.2 Fluorescence Enhancement with a PEG 200 Crowding Agent. ....	52
Figure 5.1 F19G14 GQS CD Metal Ion Unfolding Comparison. ....	63
Figure 5.2 NRAMP6 GQS CD $Cu^{2+}$ titration curve plot with EDTA signal rescue. ....	65
Figure A.1 G2w1 GQS CD $K^+$ Titration and Fraction Folded plot. ....	75
Figure A.2 G2m1 GQS CD $K^+$ Titration and Fraction Folded plot. ....	75
Figure A.3 G2s1 GQS CD $K^+$ Titration and Fraction Folded plot.....	76
Figure A.4 G2w2 GQS CD $K^+$ Titration and Fraction Folded plot. ....	76
Figure A.5 G2m2 GQS CD $K^+$ Titration and Fraction Folded plot. ....	77
Figure A.6 G2s2 GQS CD $K^+$ Titration and Fraction Folded plot.....	77

Figure A.7 G3m1 GQS CD K <sup>+</sup> Titration and Fraction Folded plot. ....	78
Figure A.8 A.8 G3m1 GQS CD 2 and 3 State Fraction Folded Plot with Residuals.....	78
Figure A.9 G4m1 GQS CD K <sup>+</sup> Titration and Fraction Folded plot. ....	79
Figure A.10 A.10 G4m1 GQS CD 2 and 3 State Fraction Folded Plot with Residuals.....	79
Figure A.11 G5m1 GQS CD K <sup>+</sup> Titration and Fraction Folded plot. ....	80
Figure A.12 G6m1 GQS CD K <sup>+</sup> Titration and Fraction Folded plot. ....	80
Figure A.13 G2m1 GQS Fluorescence K <sup>+</sup> Titration and Fraction Folded plot.....	81
Figure A.14 G6m1 GQS Fluorescence K <sup>+</sup> Titration and Fraction Folded plot.....	81
Figure A.15 Dependence of Melting Temperature on RNA Concentrations for G2m1 and G6m1 GQS.....	82
Figure A.16 Hill Coefficients (n) of GQS Sequences Studied.....	83
Figure B.1 K <sup>+</sup> Titration Result of dG2A using CD and Fluorescence Spectroscopy. ....	84
Figure B.2 K <sup>+</sup> Titration Result of dG2A2 using CD and Fluorescence Spectroscopy. ....	85
Figure B.3 K <sup>+</sup> Titration Result of dG2A3 using CD and Fluorescence Spectroscopy. ....	85
Figure B.4 K <sup>+</sup> Titration Result of dG2T using CD and Fluorescence Spectroscopy. ....	86
Figure B.5 K <sup>+</sup> Titration Result of dG2T2 using CD and Fluorescence Spectroscopy. ....	86
Figure B.6 K <sup>+</sup> Titration Result of dG2T3 using CD and Fluorescence Spectroscopy. ....	87
Figure B.7 K <sup>+</sup> Titration Result of dG3A using CD and Fluorescence Spectroscopy. ....	87
Figure B.8 K <sup>+</sup> Titration Result of dG3A2 using CD and Fluorescence Spectroscopy. ....	88
Figure B.9 K <sup>+</sup> Titration Result of dG3A3 using CD and Fluorescence Spectroscopy. ....	88
Figure B.10 K <sup>+</sup> Titration Result of dG3T using CD and Fluorescence Spectroscopy. ....	89
Figure B.11 K <sup>+</sup> Titration Result of dG3T2 using CD and Fluorescence Spectroscopy. ....	89
Figure B.12 K <sup>+</sup> Titration Result of dG3T3 using CD and Fluorescence Spectroscopy. ....	90
Figure B.13 K <sup>+</sup> Titration Result of dG3C using CD and Fluorescence Spectroscopy.....	90
Figure B.14 Overlay of Fluorescence Intensity versus K <sup>+</sup> Concentration for dG3T, dG3T2, and dG3T3. ....	91

Figure B.15 Overlay of Fluorescence Spectra of GQS at 1M K <sup>+</sup> .....	91
Figure B.16 NMR Structure of d(GGA) <sub>4</sub> .....	92
Figure C.1 Representative K <sup>+</sup> Titration of G3T and G3T 5'T using Fluorescence Spectroscopy.....	93
Figure C.2 Representative K <sup>+</sup> Titration of G3T and G3T 5'T using Fluorescence Spectroscopy.....	93
Figure C.3 Representative K <sup>+</sup> Titration of G3T and G3T 5'T using Fluorescence Spectroscopy.....	94
Figure C.4 Temperature Dependence of Fluorescence Intensity.....	94
Figure C.5 Counter-anion Effects on Fluorescence Emission Intensity for the G3T Sequence.....	95
Figure C.6 Quenching Effect of Molecular Oxygen.....	95
Figure C.7 Mg <sup>2+</sup> Effects on Fluorescence Emission Intensity.....	96
Figure D.1 GGT3 GQS CD K <sup>+</sup> Titration and λ <sub>max</sub> Curve Plot.....	97
Figure D.2 F19G14 GQS CD K <sup>+</sup> Titration and λ <sub>max</sub> Curve Plot.....	98
Figure D.3 NRAMP6 GQS CD K <sup>+</sup> Titration and λ <sub>max</sub> Curve Plot.....	98
Figure D.4 HMA5 GQS CD K <sup>+</sup> Titration and λ <sub>max</sub> Curve Plot.....	99
Figure D.5 COBL10 GQS CD K <sup>+</sup> Titration and λ <sub>max</sub> Curve Plot.....	99
Figure D.6 NRAMP6 GQS CD-Monitored Cd <sup>2+</sup> Slow Kinetics.....	100
Figure D.7 GGT3 GQS CD Cu <sup>2+</sup> and Cd <sup>2+</sup> Titration Curve Plot.....	100
Figure D.8 F19G14 GQS CD Cu <sup>2+</sup> and Cd <sup>2+</sup> Titration Curve Plot.....	101
Figure D.9 NRAMP6 GQS CD Cu <sup>2+</sup> and Cd <sup>2+</sup> Titration Curve Plot.....	101
Figure D.10 HMA5 GQS CD Cu <sup>2+</sup> and Cd <sup>2+</sup> Titration Curve Plot.....	102
Figure D.11 COBL10 GQS CD Cu <sup>2+</sup> and Cd <sup>2+</sup> Titration Curve Plot.....	102
Figure D.12 NRAMP GQS CD Cu <sup>2+</sup> Titration Curve Plot with Varied K <sup>+</sup> .....	103
Figure D.13 GGT3 GQS CD Cu <sup>2+</sup> Titration Curve Plot at Varied pH.....	103

**LIST OF TABLES**

Table 2.1 GQS Folding Parameters .....	17
Table 3.1 DNA GQS Sequence, Topology, and Emission Wavelength Maximum.....	35
Table 4.1 DNA GQS Names, Sequences, and Relative Fluorescence Enhancement .....	48
Table 5.1 RNA GQS Names, Sequences, Accession Numbers and GO Terms. ....	61
Table 5.2 RNA GQS K <sup>+</sup> Folding Results.....	62
Table 5.3 RNA GQS Cu <sup>2+</sup> Unfolding Results .....	64
Table A.1 RNA Oligonucleotide Sequences.....	74

**LIST OF ABBREVIATIONS**

$\epsilon$	Ellipticity
$\lambda$	Wavelength
A	Adenine
C	Cytosine
CD	Circular Dichroism
DNA	Deoxyribonucleic Acid
ds	Double-stranded
EDTA	Ethylenediaminetetraacetic acid
Fl	Fluorescence
G	Guanine
G2	G-quadruplex with two G-tracts
G3	G-quadruplex with three G-tracts
G4	G-quadruplex with four G-tracts
G5	G-quadruplex with five G-tracts
G6	G-quadruplex with six G-tracts
GFP	Green Fluorescent Protein
GO	Gene Ontology
GQS	G-Quadruplex Structure
L	Loop
LiCac	Lithium cacodylate
K <sup>+</sup>	Potassium ion
n	Hill coefficient / cooperativity
NMR	Nuclear Magnetic Resonance

nt	Nucleotide
OAc	Acetate
PDB	Protein Data Bank
PEG	Polyethylene glycol
PNA	Peptide Nucleic Acid
RNA	Ribonucleic Acid
SELEX	Systematic Evolution of Ligands by EXponential Enrichment
ss	single-stranded
T	Thymine
TAIR	The Arabidopsis Information Resource

## ACKNOWLEDGEMENTS

First and foremost, I would like to thank my research mentor, Dr. Phil Bevilacqua, for giving me the opportunity for an amazing undergraduate research experience. Phil, I didn't tell you this at the time, but I had written a draft of an email to ask you if I could join your lab while I was taking Chem 110H, but I was too nervous to send it! Luckily, you asked me a few weeks later if I was interested in working for you, so I never needed to send it. I can't imagine my four years at Penn State without working in the lab. You taught me how to think as a scientist and ask important questions, and I will take what I've learned from you and apply it for the rest of my career.

My time in lab also would not have been the same without my graduate mentor, Dr. Chun Kit Kwok. Kit, you were patient with me and pushed me to work hard, even when I was in Germany or you were in Hong Kong. We accomplished more together than I ever would have imagined. I wish you the best of luck in your postdoc and your future career. Maybe we will even be collaborators again someday! Additionally I would like to thank the rest of the Bevilacqua Lab members over the years. You all provided a wonderful work environment and treated me as an equal, which always meant a lot to me as an undergraduate.

I want to take this chance to thank my two additional thesis committee members, Dr. Raymond Funk and Dr. Scott Showalter. Thank you to Dr. Eric Meggers, who provided me with the opportunity to continue doing research during my year abroad in Germany. Thank you to Dr. Bortiatynski for your advice and support, as well as giving me the opportunity to be involved in SEECoS, which was a truly enriching experience and helped to support my summer research. I would also like to thank Dr. Bratoljub Milosavljevic, Dr. Christine Keating, and Dr. Mark Maroncelli for the use of their fluorimeters over the course of my many fluorescence

experiments. A special thanks to Dr. Sarah Assmann and Dr. Melissa Mullen Davis for helpful discussions in plant subgroup meetings.

Last but not least, I want to thank my family and friends. To my parents Susie and Phil, you are incredibly supportive and have never let me forget how proud you are. You always make me feel loved and empowered to accomplish whatever I set my mind to. Thanks to my brothers, Nevin and Carl, and sister-in-law Melissa for being awesome siblings and friends. To all my friends, thanks for being there for me in stressful times and making my time at Penn State the craziest and most memorable four years of my life.

## DEDICATION

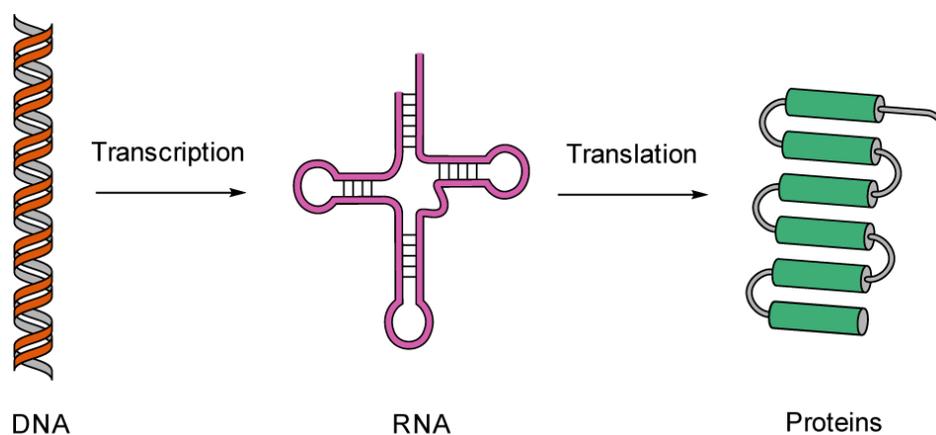
I would like to dedicate this thesis in the honor of Peter C. Breen (1991-2014).

Peter, you were a wonderful classmate, labmate, co-worker, and friend. You were always helpful and gave great insight on difficult problems. You would always help me at the last minute on homework if I forgot to do it. You made me feel better if I was struggling and never let me get overconfident if I was excelling. We shared so much and I will always cherish those memories. Your contributions at Penn State did not go unnoticed and you had an incredibly bright future ahead of you. Although you didn't have the chance to finish your honors thesis, everyone in the Bevilacqua Lab knows how hard you worked and you completely deserved your Schreyer honors medal. Peter, you were loved by all who knew you and we all miss you dearly.

## Chapter 1

### Introduction to Nucleic Acid Folding and G-Quadruplexes

The molecular basis of life lies in three biopolymers: deoxyribonucleic acid (DNA), ribonucleic acid (RNA) and proteins. The Central Dogma of molecular biology describes the information channel for genetic instructions in which genes are stored in DNA, transcribed into RNA, and finally translated into proteins that form the structures and execute the majority of the processes that govern life (Figure 1.1). It has now become clear that these processes are much more sophisticated and that RNA is much more than an intermediate template for the production of proteins. RNA has been found to have both gene regulatory function in the cases of riboswitches<sup>2,3</sup>, microRNAs<sup>4,5</sup>, and RNA G-Quadruplexes<sup>6-8</sup>, as well as catalytic function in the case of a ribozyme undergoing self-cleavage<sup>9-11</sup>. These discoveries have led to the RNA World hypothesis, which postulates RNA as the original genetic material due to its versatility<sup>12</sup>.

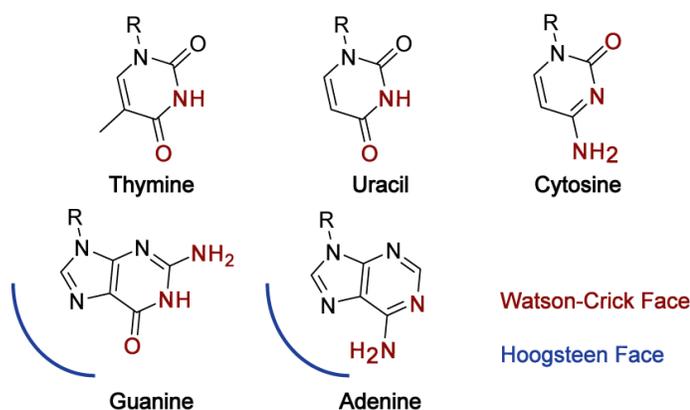


**Figure 1.1 The Central Dogma of Molecular Biology**

DNA is transcribed to produce RNA, which is then translated into proteins.

## 1.1 Nucleic Acid Structure

The unique structures of RNA and DNA are crucial for their respective biological roles. At the primary level of structure, nucleic acids consist of three major components: a nucleobase, a sugar, and a phosphate group. DNA and RNA each consist of four nucleobases: Adenine, Guanine, Cytosine and either Thymine (DNA) or Uracil (RNA); Thymine and Uracil differ by a single methyl group (Figure 1.2).



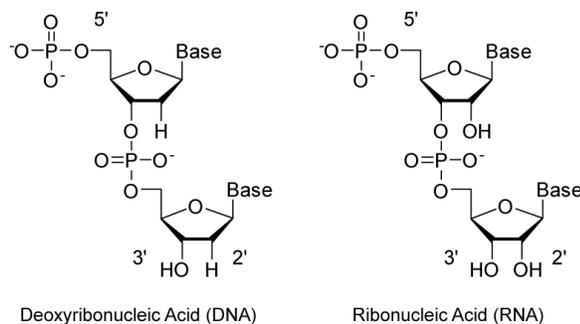
**Figure 1.2 The Five Nucleobases of RNA and DNA**

The red atoms denote the hydrogen bond donors and acceptors on the Watson-Crick face. The blue curve represents the Hoogsteen face, which is accessible on the purines during base-pairing.

Adenine (A) and Guanine (G) are purines and Cytosine (C), Thymine (T), and Uracil (U) are pyrimidines. G and C have three complementary hydrogen bonding sites on the Watson-Crick face, while A and T (or U) only have two complementary sites.

The nucleobase is connected to the sugar by the glycosidic bond and these two components combined are a nucleoside. A nucleotide is comprised of the nucleobase, sugar, and phosphate group. RNA and DNA are able to polymerize by linkage through phosphodiester bond on the phosphate sugar backbone, which goes from the 3' oxygen to the adjacent 5' phosphate group (Figure 1.3). The phosphate groups are unprotonated at physiological pH, leading to the negatively-charged backbone. The highly charged backbone is stabilized by positively-charged ions, of which  $\text{Mg}^{2+}$  is the most common<sup>13</sup>. The difference between the

backbone of RNA and DNA is that RNA contains a 2'OH group on the sugar whereas DNA contains a 2' H.



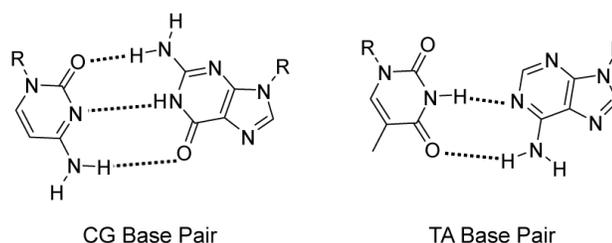
**Figure 1.3 Nucleic Acid Backbone**

The 5' to 3' linkage on the backbone of DNA and RNA.

The seemingly minor differences between ribose and deoxyribose actually leads to major dissimilarities in the conformational flexibility of RNA compared to DNA<sup>14,15</sup>. The pentose ring of the sugar adopts a pucker formation to achieve the torsion angle with minimal overlap and maximal stability; for ribose (RNA), the most stabilizing conformation is C2' endo and for deoxyribose (DNA), the most stabilizing conformation is C3' endo<sup>16</sup>.

At the secondary level of structure, the bases form hydrogen bonds with one another, leading to a compact and stable structure. DNA typically exists as two self complementary strands involved in Watson-Crick base pairing, in which G and C have three hydrogen bonds and A and T have two hydrogen bonds<sup>17</sup> (Figure 1.4). While DNA is almost always in this B-form double-stranded (ds) helix it can also be single-stranded (ss), especially during replication and in the telomere regions.

In contrast to DNA, RNA exists mostly as a single strand, which then undergoes base pairing and base stacking for added stability. RNA also forms Watson Crick (GC and AU) base pairs as well as non-Watson Crick pairing, such as the GU wobble, GA shear, and GG interactions.



**Figure 1.4 Watson Crick Base Pairing in DNA**

Cytosine and Guanine (CG) share three hydrogen bonds and Thymine and Adenine (TA) share two hydrogen bonds.

Vertical stacking of the heterocyclic nucleobases, which is a non-covalent interaction, is also stabilizing, and purines (G and A) stack more efficiently than pyrimidines (A and T/U). For example, adenosine stacking has a favorable -1 kcal/mole stabilization energy at 25 °C as separate nucleosides<sup>18</sup>. Unpaired bases, particularly dangling ends off of either side of a structure (helix, G-quadruplex, etc.), tend to stack to minimize energy. For RNA, a 3'-dangling end stacks better while a 5'-dangling end stacks better in DNA<sup>19</sup>.

Tertiary interactions between regions involved base pairs are also possible and common. Examples of tertiary RNA structures are kissing loops, pseudoknots, and helical junctions<sup>16,20</sup>. These structures lend additional stabilization and compaction to the overall structure. RNA and DNA also regularly interact with one another, as well as proteins and small molecules. Small molecule binding to RNA has recently become of particular interest in drug development and discovery, especially in aptamer development<sup>21</sup>.

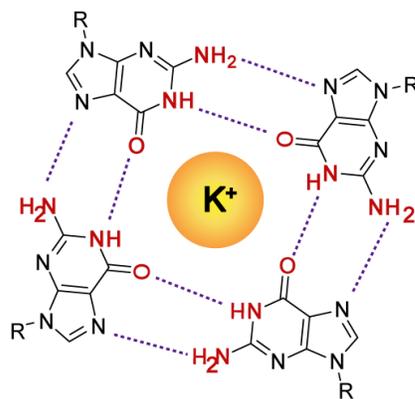
## 1.2 Nucleic Acids as a Biochemical Tool

In order to achieve specific and strong binding between a nucleic acid and a target molecule, a great deal of effort has recently gone into designing and engineering modified sequences for aptamer development using SELEX, and even modified nucleic acids. Peptide

nucleic acids (PNA), which combine the neutral backbone of polypeptides with the sidechain nucleobases of RNA and DNA, show potential in pharmaceuticals and personalized medicine<sup>22</sup>. Recently, the Jaffrey Lab has created RNA mimics of Green Fluorescent Protein (GFP), which they have named Spinach and Spinach2<sup>23-25</sup>. These RNA aptamers bind to the fluorophore component identified in GFP and has been shown in principle to be used in live-cell imaging with a fluorescent readout. These developments show great promise for new techniques in following RNA expression, especially for non-coding RNAs, which cannot be followed by protein expression using GFP and whose functions remain largely unidentified.

### 1.3 G-Quadruplex Structure

A less studied folding motif that can form in ssDNA or ssRNA is the Guanine-Quadruplex Structure (GQS). The first report of this structure was in the context of self-assembling guanosine monophosphates in solution followed by their discovery in G-rich telomere regions, which are located at the end of chromosomes<sup>26-28</sup>. The basis of the quadruplex is a set of four guanines associated by hydrogen bonding as modeled in Figure 1.5.

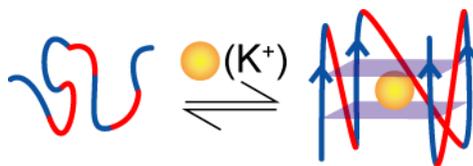


**Figure 1.5 A Top View of G-Quadruplex Structure**

Four guanines assemble in a quartet structure via hydrogen bonding. Red denotes the Watson Crick face of the nucleobase and R denotes the sugar. A desolvated  $K^+$  ion resides in the middle to stabilize the structure.

These four guanines form a quartet plane and the quadruplex structure arises from quartet planes stacked on top of one another. At least two quartet planes are necessary for quadruplex formation, leading to the requirements of four repeats of at least two guanines with at least one loop nucleotide in between. The general form of the consensus sequence for a GQS is  $G_xL_aG_xL_bG_xL_cG_x$ , where  $x$  is  $\geq 2$  nt and loops (L) a, b, and c are  $\geq 1$  nt. The nomenclature used in this thesis refers to a GQS with 2 quartet planes ( $x=2$ ) as a ‘G2’, 3 planes as a ‘G3’, etc.

Because the structure requires four highly negative backbones into close proximity, GQSs require monovalent cations to stabilize folding (Figure 1.6). While  $Li^+$  is too small to provide adequate stabilization, both  $Na^+$  and  $K^+$  induce GQS formation.  $Na^+$  resides directly in the quartet plane with the four guanines while  $K^+$ , which has a larger atomic radius, sits between between quartet planes; however,  $K^+$  is more stabilizing due to a lower enthalpy of dehydration<sup>29</sup>.



**Figure 1.6 Parallel GQS formation with a  $K^+$  ion**

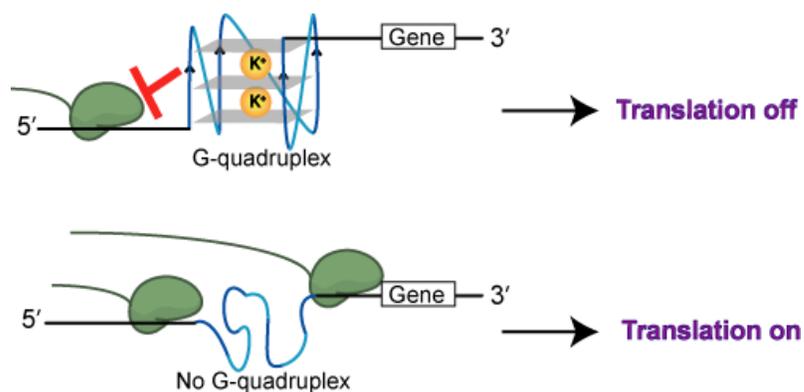
A randomly arranged nucleic acid forms a GQS upon addition of  $K^+$  ions. Blue regions represent the G-stretches and red represents loop nucleotides. Strand directionality is parallel, as is typical for RNA.

A GQS is a secondary folding structure and therefore must compete with other secondary structures. For this reason, GQSs typically form in RNA and telomere regions of DNA because these particular regions are single-stranded. One difference between RNA and DNA GQSs is the preferred topology of the quadruplex. Due to the C2' endo sugar pucker of ribose and the associated orientation of the glycosidic bond, RNA nucleobases adopt the *anti* conformation, which limits them to a parallel strand topology, as shown in Figure 1.6. DNA, on the other hand, can adopt both parallel and anti-parallel topologies as well as a mixture of the two.

## 1.4 GQS Roles in Biology

Although GQS formation has been studied *in vitro* for decades, their presence *in vivo* has been heavily debated. Earlier studies were inconclusive, due to the possibility that antibodies bound to GQSs affect their folding and induce formation. Recently, there has been conclusive proof of *in vivo* GQS formation in human cells using both a fluorescent dye and a specially engineered antibody reported<sup>8,30</sup>. G-Quadruplexes form both in telomere, promoter, and oncogene regulator regions, making them a prime target for drug development<sup>31-34</sup>. Telomere stabilization is of particular interest because these overhang regions at the end of the chromosome protect genetic coding material from degradation.

Additionally, GQSs have functions in regulation of biological processes, such as transcription and translation. For example, a GQS in a promoter region can halt transcription when folded<sup>6,35</sup> (Figure 1.7). Due to the stability of the GQS, RNA polymerase as well as the ribosome are thought to stall or fall off due to the structure physically blocking its path.



**Figure 1.7 Gene Regulation Controlled by a GQS in the Promoter Region**

In the presence of  $K^+$  ions, the GQSs will fold, physically blocking the ribosome, halting translation and downregulating expression. In the absence of GQS folding, translation proceeds as normal<sup>6,35</sup>

## 1.5 GQSs as a Biochemical Tool

Due to their unique folding structure and spectroscopic properties, GQSs have also been modified and engineered to bind specific targets. Targets of interest include fluorophores, small molecules, peptides<sup>36</sup>, and even ions. A GQS sensor was developed for ATP detection by combining it with a GQS-binding fluorophore system with analyte-dependent fluorescence<sup>37</sup>. The thrombin aptamer is a DNA GQS-forming sequence that binds thrombin and inhibits thrombin-catalyzed fibrin coagulation<sup>38,39</sup>. An intermolecular GQS was modified to include a copper-binding component, which can reversibly bind and report Cu<sup>2+</sup> ions in a system of interest<sup>40</sup>.

In the studies described herein, the potential for using GQSs is explored with the goal to better understand the physical characteristics of its folding and binding properties.

## 1.6 References

- 1 Crick, F. H. On protein synthesis. *Symp Soc Exp Biol* **12**, 138-163 (1958).
- 2 Breaker, R. R. Riboswitches and the RNA world. *Cold Spring Harb. Perspect. Biol.* **4**, (2012).
- 3 Serganov, A. & Nudler, E. A Decade of Riboswitches. *Cell* **152**, 17–24 (2013).
- 4 Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
- 5 Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215–233 (2009).
- 6 Kumari, S., Bugaut, A., Huppert, J. L. & Balasubramanian, S. An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol.* **3**, 218–221 (2007).

- 7 Bugaut, A. & Balasubramanian, S. A Sequence-Independent Study of the Influence of Short Loop Lengths on the Stability and Topology of Intramolecular DNA G-Quadruplexes. *Biochemistry* **47**, 689–697 (2008).
- 8 Biffi, G., Di Antonio, M., Tannahill, D. & Balasubramanian, S. Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat. Chem.* **6**, 75–80 (2014).
- 9 Stark, B. C., Kole, R., Bowman, E. J. & Altman, S. Ribonuclease P: an enzyme with an essential RNA component. *Proc. Natl. Acad. Sci.* **75**, 3717–3721 (1978).
- 10 Cech, T. R., Zaug, A. J. & Grabowski, P. J. In vitro splicing of the ribosomal RNA precursor of tetrahymena: Involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell* **27**, 487–496 (1981).
- 11 Lilley, D. M. J. Catalysis by the nucleolytic ribozymes. *Biochem. Soc. Trans.* **39**, 641–646 (2011).
- 12 Gilbert, W. Origin of life: The RNA world. *Nature* **319**, 618–618 (1986).
- 13 Hud, N. V.. *Nucleic acid-metal ion interactions*. (RSC Publishing, 2009).
- 14 Saenger, W. *Principles of nucleic acid structure*. Springer-Verlag, 1984).
- 15 Auffinger, P. & Westhof, E. RNA solvation: a molecular dynamics simulation perspective. *Biopolymers* **56**, 266–274 (2000).
- 16 Bloomfield, V. A., Crothers, D. M. & Tinoco, I. *Nucleic acids: structures, properties, and functions*. University Science Books, 2000.
- 17 Crick, F. H. C. & Watson, J. D. The Complementary Structure of Deoxyribonucleic Acid. *Proc. R. Soc. Lond. Ser. Math. Phys. Sci.* **223**, 80–96 (1954).
- 18 Alderfer, J. L., Tazawa, I., Tazawa, S. & Ts'o, P. O. P. 2'-O-Alkyl polynucleotides. 2. Comparative studies on homopolymers of adenylic acid possessing different C-2' substituents of the furanose. Poly(deoxyriboadenylic acid), poly(riboadenylic acid),

- poly(2'-O-methyladenylic acid), and poly(2'-O-ethyladenylic acid). *Biochemistry* **13**, 1615–1622 (1974).
- 19 Bommarito, S., Peyret, N. & Jr, J. S. Thermodynamic parameters for DNA sequences with dangling ends. *Nucleic Acids Res.* **28**, 1929–1934 (2000).
- 20 Chang, K. Y. & Tinoco, I. Characterization of a 'kissing' hairpin complex derived from the human immunodeficiency virus genome. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 8705–8709 (1994).
- 21 Huang, Z. & Szostak, J. W. Evolution of aptamers with a new specificity and new secondary structures from an ATP aptamer. *RNA* **9**, 1456–1463 (2003).
- 22 Ray, A. & Nordén, B. Peptide nucleic acid (PNA): its medical and biotechnical applications and promise for the future. *FASEB* **14**, 1041–1060 (2000).
- 23 Paige, J. S., Wu, K. Y. & Jaffrey, S. R. RNA Mimics of Green Fluorescent Protein. *Science* **333**, 642–646 (2011).
- 24 Song, W., Strack, R. L., Svensen, N. & Jaffrey, S. R. Plug-and-Play Fluorophores Extend the Spectral Properties of Spinach. *J. Am. Chem. Soc.* **136**, 1198–1201 (2014).
- 25 Strack, R. L., Disney, M. D. & Jaffrey, S. R. A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat-containing RNA. *Nat. Methods* **10**, 1219–1224 (2013).
- 26 Gellert, M., Lipsett, M. N. & Davies, D. R. Helix Formation By Guanylic Acid. *Proc. Natl. Acad. Sci. U. S. A.* **48**, 2013–2018 (1962).
- 27 Henderson, E., Hardin, C. C., Walk, S. K., Tinoco, I., Jr & Blackburn, E. H. Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs. *Cell* **51**, 899–908 (1987).
- 28 Williamson, J. R., Raghuraman, M. K. & Cech, T. R. Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell* **59**, 871–880 (1989).

- 29 Hud, N. V., Smith, F. W., Anet, F. A. & Feigon, J. The selectivity for K<sup>+</sup> versus Na<sup>+</sup> in DNA quadruplexes is dominated by relative free energies of hydration: a thermodynamic analysis by <sup>1</sup>H NMR. *Biochemistry* **35**, 15383–15390 (1996).
- 30 Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.* **5**, 182–186 (2013).
- 31 Phan, A. T. & Mergny, J.-L. Human telomeric DNA: G-quadruplex, i-motif and Watson–Crick double helix. *Nucleic Acids Res.* **30**, 4618–4625 (2002).
- 32 Azzalin, C. M., Reichenbach, P., Khoriauli, L., Giulotto, E. & Lingner, J. Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* **318**, 798–801 (2007).
- 33 Williamson, J. R. G-quartet structures in telomeric DNA. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 703–730 (1994).
- 34 Phan, A. T. Human telomeric G-quadruplex: structures of DNA and RNA sequences. *FEBS J.* **277**, 1107–1117 (2010).
- 35 Shahid, R., Bugaut, A. & Balasubramanian, S. The BCL-2 5' Untranslated Region Contains an RNA G-Quadruplex-Forming Motif That Modulates Protein Expression. *Biochemistry* **49**, 8300–8306 (2010).
- 36 Kobayashi, K., Matsui, N. & Usui, K. Use of a Designed Peptide Library to Screen for Binders to a Particular DNA G-Quadruplex Sequence. *J. Nucleic Acids* **2011**, (2011).
- 37 Zhang, Z., Sharon, E., Freeman, R., Liu, X. & Willner, I. Fluorescence Detection of DNA, Adenosine-5'-Triphosphate (ATP), and Telomerase Activity by Zinc(II)-Protoporphyrin IX/G-Quadruplex Labels. *Anal. Chem.* **84**, 4789–4797 (2012).
- 38 Macaya, R. F., Schultze, P., Smith, F. W., Roe, J. A. & Feigon, J. Thrombin-binding DNA aptamer forms a unimolecular quadruplex structure in solution. *Proc. Natl. Acad. Sci.* **90**, 3745–3749 (1993).

- 39 Kankia, B. I. & Marky, L. A. Folding of the Thrombin Aptamer into a G-Quadruplex with Sr<sup>2+</sup>: Stability, Heat, and Hydration. *J. Am. Chem. Soc.* **123**, 10799–10804 (2001).
- 40 Engelhard, D. M., Pievo, R. & Clever, G. H. Reversible Stabilization of Transition-Metal-Binding DNA G-Quadruplexes. *Angew. Chem. Int. Ed.* **52**, 12843–12847 (2013).

## Chapter 2

### **Decrease of RNA Folding Cooperativity by Deliberate Population of Intermediates in RNA G-Quadruplexes**

[Published as a paper entitled “Decrease of RNA Folding Cooperativity by Deliberate Population of Intermediates in RNA G-Quadruplexes” by Chun Kit Kwok, Madeline E. Sherlock, and Philip C. Bevilacqua in *Angewandte Chemie International Edition* **52**, 683-686 (2013). This paper was recommended by F1000 Faculty. <http://f1000.com/prime/717970147>] Author contributions are listed in Section 2.5.

#### **2.1 Abstract**

Conventional biochemical detection systems only have a 100-fold linear response range. In this study, we broaden the range of potassium concentrations detected by an RNA G-quadruplex-containing sequence by intentionally populating multiple binding and folding intermediate states. The broadened RNA response can be monitored by both circular dichroism and intrinsic fluorescence spectroscopy.

#### **2.2 Introduction**

RNA folding is central to diverse biological and chemical processes, including gene regulation and biosensing. Cooperativity is a key feature of RNA folding, and critical to the folding of small and large RNAs<sup>1-4</sup> alike and to the response of naturally occurring riboswitches<sup>5</sup>. Some biological RNAs fold with little or no cooperativity, whereas others exhibit high

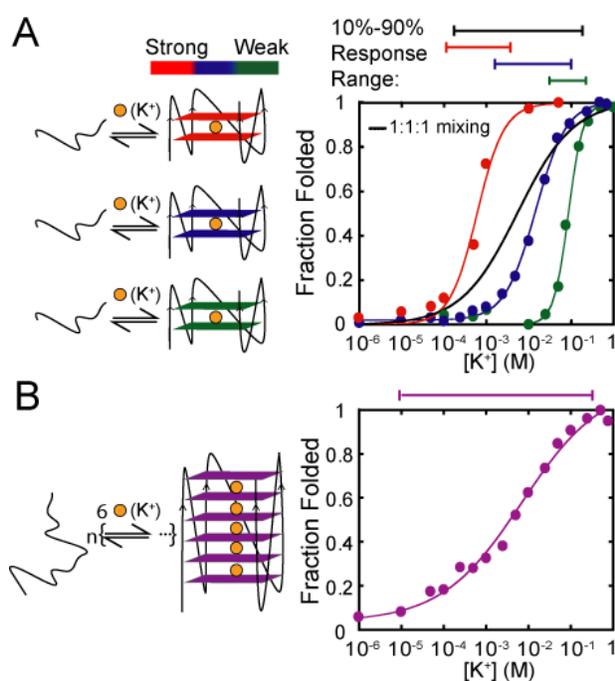
cooperativity<sup>5-8</sup>. Concerning biosensing, conventional or “classical” biosensors without cooperative behavior bind an analyte with a responsive range of ~100-fold in analyte concentration, using 10-90% of maximal signal. When the analyte concentration is above or below this range, however, “dead zones” are found wherein the sensor will not work<sup>9,10</sup>; as such, biosensors with extended dynamic ranges are needed. Tuning negative (or “anti-”) cooperativity into RNA folding provides one potential means of achieving these desirable outcomes. The goals of our study are to gain new fundamental insights into the folding of biological RNAs and to aid the development of biosensors with diverse response ranges through the development of a series of RNA scaffolds with tunable cooperativity.

Several types of broad ligand response biopolymers have been reported previously. Broad-response detection of pH and ionic strength were attained by coupling signal output to a one-state downhill folding protein scaffold.<sup>11</sup> Broad-response sensing was also achieved by mixing together three (or more) biosensors that differ in analyte  $K_d$  by ~1-2 logs<sup>12-14</sup>. Using such an approach, broad-response sensors were prepared that detect changes in pH<sup>13</sup> and in DNA<sup>12</sup> and protein<sup>14</sup> concentration.

In the present study, we develop another, simpler approach to broad-range folding and sensing, which we link to anticooperative RNA folding. We use just a *single*, rationally designed RNA sequence as a scaffold for ligand detection. Our strategy exploits deliberate population of intermediates along both the RNA folding and analyte binding pathways (Figure 5.1). We make use of a G-quadruplex sequence (GQS) to explore the relationship between RNA sequence and anticooperative RNA folding. Proof-of-principle of analyte sensing and riboswitching are demonstrated by monitoring the concentration of  $K^+$ , with fraction RNA folded read out using either circular dichroism (CD) or fluorescence spectroscopy to monitor GQS folding.

A GQS is a nucleic acid motif formed by guanine-rich RNA or DNA sequences of the pattern  $G_xL_aG_xL_bG_xL_cG_x$ , where  $x$  is the length of G-stretch that is involved in G-quartet

interaction, and  $L_a$ ,  $L_b$ , and  $L_c$  are the three loops between the four G-strands that can be of the same or different length and sequence<sup>15</sup>; such a sequence is referred to herein as a “GX” GQS. A GQS generally consists of two or more stacked quartet planes, with a dehydrated  $K^+$  between planes, which stabilizes the quadruplex<sup>16,17</sup>. Extensive studies on DNA and RNA GQS in different sequence contexts have revealed that GQS bind  $K^+$  selectively over other types of metal ions<sup>18-21</sup>. In addition, GQS has been used as a biosensor to determine  $K^+$  concentration, in cells<sup>22</sup> and other samples.



**Figure 2.1 Two Approaches to Achieve Broad Response Sensing**

(A) Mixing of different sensors, as previously reported.<sup>12</sup> Mixing together three different G2 GQS sequences with  $\Delta K_d$  of  $\sim 200$  fold (G2w2, G2m1, and G2s2). The  $K^+$  binding affinity towards individual GQS is represented by the color-coded bars above the plot, in which red, blue, and green represent the 10-90% response range of the strong, medium, and weak binders respectively. Mixing the three binders should in principle allow broad-response sensing, as depicted by the black curve, which simulates a 1:1:1 mixture of the three G2 GQS. (B) Extending the G-stretch length of a GQS developed herein. The color-coded bar above the graph depicts the 10-90% response range. Longer G-stretch lengths allow more G-quartet planes to form, and thus more intermediate states in the RNA folding and potassium binding pathway to populate, which results in a broadened response to  $K^+$ . Shown here is the G6 GQS.

RNA GQS predominately form a parallel topology upon  $K^+$  titration<sup>23-25</sup>. This topology leads to an increase in ellipticity at  $\sim 260$  nm. Circular Dichroism (CD) spectroscopy has been used extensively to study folding of other G-quadruplexes<sup>19,26-28</sup>. More recently, it has been reported that DNA G-quadruplexes are intrinsically fluorescent<sup>29,30</sup>. To assay for GQS formation herein, we used both CD and fluorescence spectroscopy to monitor this change upon  $K^+$  addition. In addition, we show for the first time that RNA GQS are intrinsically fluorescent and that fluorescence signal responds to  $K^+$  concentration in a fashion similar to CD signal.

## 2.3 Results and Discussion

First, we investigate the folding of three G2 GQS sequences termed G2w1, G2m1, and G2s1, where w, m, and s represent weak, medium, and strong binding of  $K^+$  using CD signal<sup>31</sup>. These sequences have positive or neutral cooperativity and are related to those from our recent report<sup>27</sup>. Binding strength was controlled by the sequence and length of the loops.  $K^+_{1/2}$  values were determined as  $410 \pm 70$ ,  $14 \pm 1$ , and  $2.2 \pm 0.5$  mM respectively (Table 2.1). We then identified a second set of G2 GQS—G2w2, G2m2, and G2s2—having more A-content in the loops, which led to similar  $K^+$  binding profiles, and somewhat higher Hill coefficients (Table 2.1)<sup>27</sup>. Folding transitions for all six G2 GQS are provided in Figure 5.2. The G2 GQS were then classified as either classical response ( $\sim 2$  logs) or digital response ( $< 2$  logs) depending on whether their n values were  $\sim 1$  or  $\sim 2$  respectively<sup>32</sup>.

To determine the effective dynamic range of response possible by using the previously published approach of *mixing* probes,<sup>12-14</sup> we selected G2w2, G2m1 and G2s2 GQS, which vary by  $\sim 2$  logs in  $K^+_{1/2}$ . We simulated a linear broad-response signal (Eq. 2.4) and fit the resulting response with Eq. 5.1 (Figure 5.1A, black line). This yielded an apparent  $K^+_{1/2}$  value of 12 mM and an apparent Hill coefficient of 0.6. The mixing approach was thus expected to widen the

10%-90% response range to  $\sim 3$  logs in  $K^+$  concentration, as represented by the black bar above Figure 5.1A.

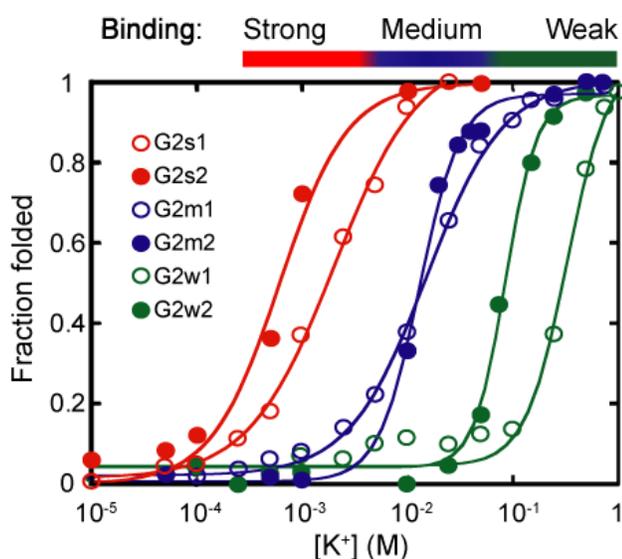
**Table 2.1 GQS Folding Parameters**

GQS motif <sup>[d]</sup>	Fitting <sup>[a]</sup>	$n^{[a,c]}$	$K^+_{1/2}$ (mM) <sup>[a,c]</sup>	Reference
G2w1	2-state	$2.2 \pm 0.3$	$410 \pm 70$	This study
G2m1	2-state	$1.2 \pm 0.1$	$14 \pm 1$	This study
G2s1	2-state	$1.2 \pm 0.1$	$2.2 \pm 0.5$	This study
G2w2	2-state	$2.7 \pm 0.1$	$117 \pm 28$	<sup>27</sup>
G2m2	2-state	$2.5 \pm 0.4$	$14 \pm 1$	<sup>27</sup>
G2s2	2-state	$1.7 \pm 0.4$	$0.8 \pm 0.2$	<sup>27</sup>
G3m1	2-state	$0.5 \pm 0.1$	$0.7 \pm 0.2$	This study
G3m1 (tr1) <sup>[b]</sup>	3-state	$1.5 \pm 0.1$	$0.06 \pm 0.01$	This study
G3m1 (tr2) <sup>[b]</sup>	3-state	$1.3 \pm 0.2$	$6.4 \pm 2.8$	This study
G4m1	2-state	$0.5 \pm 0.1$	$2.7 \pm 2.2$	This study
G4m1 (tr1) <sup>[b]</sup>	3-state	$1.2 \pm 0.2$	$0.06 \pm 0.06$	This study
G4m1 (tr2) <sup>[b]</sup>	3-state	$1.2 \pm 0.1$	$12 \pm 1$	This study
G5m1	2-state	$0.7 \pm 0.1$	$3.7 \pm 0.7$	This study
G6m1	2-state	$0.5 \pm 0.1$	$12 \pm 2$	This study

**Table Footnote:** [a] GQS folding parameters were obtained from fitting CD titration experiments using a two-state model (Eq. 2.1). [b] In several cases (“tr1” and “tr2” for transitions 1 and 2), two distinct transitions were observed, and three-state fitting was performed (Eq. 2.2). [c] Values reported in the table are average  $\pm$  standard deviation, obtained from three separate experiments. [d] Full sequences are provided in Table A.1.

We next sought to attain broad-response sensing of  $K^+$  by our proposed method of deliberately populating intermediates. To do so, we took advantage of our previously reported observation that a G2 GQS sequence generally has a higher Hill coefficient and thus a steeper

(i.e. more “digital”) response to  $K^+$  concentration than an otherwise similar G3 GQS<sup>27</sup>. We previously attributed this trend to G2 sequences having poorly populated folding and binding intermediates—a phenomenon that generally enhances folding cooperativity<sup>33,34</sup>. Fewer populated intermediates in G2 sequences was attributed to fewer ways to assemble the quartets incorrectly, and to the presence of just one ion binding site, which precludes ion-ion interactions<sup>27</sup>. We then reasoned that the inverse may hold: *i.e. increasing* the length of the G-stretches in a single sequence might enhance the number and population of folding and binding intermediates, thereby lessening folding and binding cooperativity and broadening response to analyte concentration (Figure 5.1B).



**Figure 2.2 Fraction Folded Plots for all of the G2 GQS, Displaying Different  $K^+$  Binding Strengths**

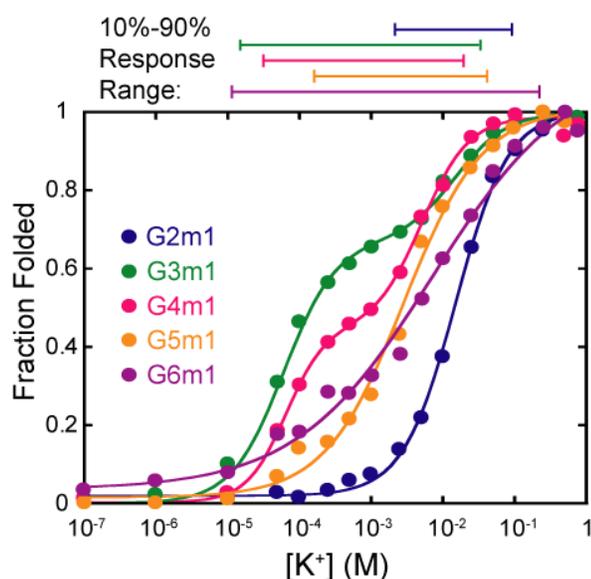
Data at 262nm were well fit by a two-state Hill equation (Eq. 2.1). Red, blue, and green colors represent strong, medium, and weak binders, respectively.  $n$  and  $K^+_{1/2}$  values are provided in Table 2.1. See Figures A.1-A.6 for full CD spectra and individual fits.

To test this idea, we selected G2m1—which has the sequence  $G_2UAG_2UAG_2CG_2$  and a conventional 2-log response—as our model GQS and systematically extended the length of the G-stretch, affording G3m1–G6m1, where G6m1 has the sequence  $G_6UAG_6UAG_6CG_6$  (all sequences in Table A.1). Fitting of CD-detected  $K^+$  titrations for G3m1 and G4m1 revealed

distinctly three-state folding behavior (Figure 2.3, green and magenta lines)—a conclusion that was further supported by plots of the residuals (Figures A.8 and A.10). Notably, each of the two transitions for G3m1 and G4m1 has a Hill coefficient of  $\sim 1$  (Table 2.1), supporting well-separated transitions and a well-populated folding intermediate and usage of the three-state model (Eq. 2.2). This intermediate may represent slipped quartets (i.e. with X-1 or X-2 quartets), and/or weakened binding of the “second” or “third” ion in G3m1 and G4m1, respectively, owing to ion-ion repulsion. On the one hand, population of the folding intermediate broadened the response, as anticipated; but, on the other hand, the response was no longer linear and instead had a prominent dead zone in the center of the graph (Figure A.3, green and magenta lines)<sup>35</sup>. Nonetheless, intermediates clearly populate as the number of quartets increases, which motivated us to examine even longer GQS.

Potassium ion titrations of G5m1 and G6m1 sequences are provided in Figure 2.3 (gold and purple lines, respectively) and are largely linear in the 10-90% transition region (i.e. no significant dead zone like in G3m1 and G4m1) and broad in their response. Given such behavior, these traces could be fit to Eq. 2.1 for apparent two-state transition (Figures A.11 and A.12), which gave apparent Hill coefficients of just  $0.7 \pm 0.1$  and  $0.5 \pm 0.1$ , respectively (Table 2.1). Reversion to apparent two-state behavior for G5m1 and G6m1 is due to the presence and population of *many* folding and binding intermediates, so many so that the multiple transitions blur into one very broad transition. This is supported by the above trends and the fractional Hill coefficients. Such intermediates may represent an ensemble of states containing slipped quartet registers, and/or weakened binding of the “fourth” and “fifth”  $K^+$  ions owing to repulsion from the other bound  $K^+$  ions. Overall, G5m1 and G6m1 have astonishingly wide linear responses to  $K^+$  concentration, spanning  $\sim 3$  to 4.5 logs as compared to just 2 logs for G2m1 (Figure 2.3). Indeed, these response ranges are similar to, or exceed, those extrapolated from the above G2 sequence mixtures (Figure 2.1A).

Next, we explored fluorescence as an alternative method of monitoring GQS folding. As mentioned above, it was recently shown that the intrinsic fluorescence of DNA G-quadruplexes increases upon quadruplex formation<sup>29,30</sup>; moreover, fluorescence is often used in developing biosensors. The same potassium titrations described above for CD detection were performed on unlabeled G2m1 and G6m1 using fluorescence detection.



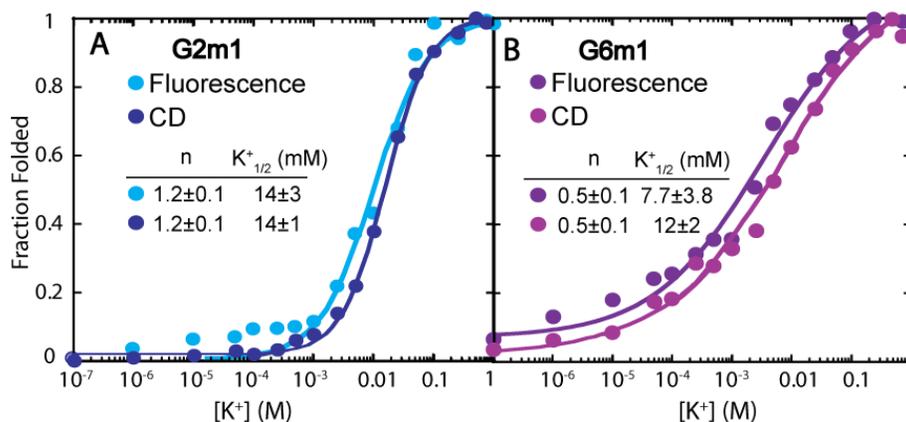
**Figure 2.3 Fraction Folded Plot of G2m1 and its Longer G-stretch Counterparts at 262nm by CD Spectroscopy.**

G3m1 and G4m1 were fit to the three-state equation (Eq. 2.2), while G2m1, G5m1, and G6m1 were fit to the two-state equation (Eq. 2.1). The conventional G2m1 has a response range of <2 logs, whereas G5m1 and G6m1 have 3- and >4-log responsive ranges. The response range for each GQS is defined from 10 to 90% folded and is depicted with bars of matching color above the panel. The  $K^{+}_{1/2}$  and  $n$  values are provided in Table 2.1. See Figures A.2 and A.7-A.12 for full CD spectra and individual fits.

We found that these RNA G-quadruplexes, like their DNA counterparts, exhibit an increase in fluorescence intensity upon folding when excited at 270 nm (Figures A.13 and A.14). The emission intensity increased upon potassium ion addition, with an emission maximum of ~360 nm. The fraction folded titration curves produced using the data at 360 nm are provided in Figure 2.4, along with a comparison to the data produced by CD-monitored experiments. The G6m1 data were again fit to the apparent 2-state model, owing to the presence of multiple

intermediate states. The  $K_{1/2}^+$  values and Hill coefficients between the two methods are within standard deviation of each other (Figure 2.4), making both methods equally useful in GQS detection. The intrinsic fluorescence of the quadruplexes is valuable not only because fluorescence is a convenient way to monitor folding and binding, but also because it can be used on unmodified oligonucleotides and so avoids potentially perturbing the interaction.

In order to ensure the signal changes were due to intramolecular and not intermolecular GQS folding, concentration-dependent UV-detected (295nm) thermal denaturations were conducted, focusing on the limiting cases of G2m1 and G6m1. The melting temperature ( $T_m$ ) was found to be independent of RNA concentration over the RNA concentration range used in CD and fluorescence spectroscopy (Figure A.15), supporting that intramolecular GQS folding was being monitored during the  $K^+$  titrations.



**Figure 2.4 Comparison of CD- and Fluorescence-Monitored GQS  $K^+$  Titrations.**

CD data points were taken at 262nm and fluorescence emission data points were taken at 360 nm. The  $K_{1/2}^+$  and Hill coefficient (n) are provided in the figure for each method. (A) Fraction-folded plots of G2m1, detected by both fluorescence and CD spectroscopy and fit to a two-state equation (Eq. 2.1). (B) Fraction-folded plots of G6m1, detected by both fluorescence and CD spectroscopy and fit to the apparent two-state equation (Eq. 2.1). See Figure A.13-A.14 for full fluorescence spectra and individual fits.

In this study we controlled RNA folding cooperativity by deliberately populating intermediates. Broad-range response was achieved with G5 and G6 GQS RNAs, which populate folding and binding intermediates, whereas narrow-range sensing was accomplished with G2

GQS RNAs, which do not populate such intermediates, as described earlier<sup>27</sup>. The binding event was read out by both CD and fluorescence spectroscopy, using the fluorescence intrinsic to the GQS. Although potassium sensing has been achieved previously, the steepness of the response has not been controlled and fluorophores had to be attached. Moreover, this is the first time that broad response has been achieved by a *single* sequence rather than a mixture of sequences. The previously established biosensor mixing approach would provide a broadened response range, of  $\sim 3\text{-}4$  logs<sup>12-14</sup>. This approach, however, has severe limitations: multiple biosensors with well-controlled  $K_d$ 's must be designed and prepared, and accurate concentration determination and mixing of each probe is needed to ensure consistency in the response profile. Moreover, such sensors could interfere by base pairing with each other or by ion-induced aggregation.

As an alternative, we introduced the concept of intentionally populating folding and binding intermediates in a single RNA sequence as a means to broadening the linear response range. The G-stretch extension approach is one such simple and robust strategy and should be applicable in DNA. It requires no mixing of multiple probes and achieves a similar or even better dynamic range of response to  $K^+$  concentration. Moreover, it is much simpler to introduce single nucleic acid probes *in vivo* than multiple probes. We clearly detected population of a folding intermediate for G3 and G4 GQS herein. Such an intermediate could include “slipped” GQS registers—absent in the simpler and steeper folding G2-sequences—and/or through-space electrostatic repulsion of the adjacent  $K^+$  ions located between quartets. Similar phenomena of weak association of “later-binding ions” has been reported for  $K^+$  binding in membrane ion channels<sup>36</sup>. The longer G-stretches reported in the G5 and G6 sequences allow even more G-quartet planes to form, and therefore more folding and binding intermediates to populate. Evidently, the transition region of G5 and G6 GQS has enough overlapping states that they revert back to apparent two-state behavior, as compared to G3 and G4 GQS which displayed distinctly

three-state folding behavior. Broadening of response with number of quartets is most evident in a graph of apparent Hill coefficient versus number of quartets (Figure A.16).

One unique feature of this study is the intrinsic fluorescence of an RNA GQS. Although the fluorescence is somewhat weak, we find that it can be enhanced by altering loop size, loop sequence, RNA topology, and ionic conditions (manuscript in preparation, see Chapter 4). In addition, tandem GQS could be introduced to multiply the fluorescence signal, as shown for related systems<sup>37</sup>. Coupling of GQS folding with ligand detection would then allow a unique readout of ligand binding, which circumvents the need for introduction of synthetic fluorophores, such as FRET pairs and fluorophore-quencher pairs.

In summary, we have shown that the cooperativity of RNA folding can be tuned by adjusting the number of G-quartets present. Although the particular GQS biosensor developed in this study senses only  $K^+$ , our approach to broad-response sensing is potentially applicable to other RNA or DNA aptamer systems<sup>38,39</sup> in which folding intermediates can be engineered into the system, for example through alternative folds or overlapping binding sites.

## **2.4 Materials and Methods**

### ***2.4.1 RNA Preparation***

RNA oligonucleotides were purchased from Dharmacon, Inc. Sequences and abbreviations are provided in Table D.1. All RNAs were dialyzed in an eight-well microdialysis apparatus (Gibco-BRL Life Technologies) at a flow rate of 25 mL/min. The RNA samples were first dialyzed against 100 mM LiCl for 6 hours to replace RNA backbone cation, then ddH<sub>2</sub>O for 6 hrs to remove excess LiCl, and finally 10 mM LiCacodylate (pH 7.0) for >12 hours. The

concentration of the dialyzed RNAs was then determined by UV-spectroscopy and stored in -20°C before and after use.

#### ***2.4.2 Spectroscopic Titration Experiments***

Circular Dichroism (CD): CD spectroscopy was performed using a Jasco CD J810 Spectropolarimeter, and data were analyzed with KaleidaGraph v.3.5 (Synergy software). RNA oligonucleotides were prepared to a concentration of 2.5-5  $\mu\text{M}$  in 10 mM LiCacodylate (pH 7.0) buffer. RNA was renatured at 95°C for 2 minutes and allowed to cool at room temperature for 15 minutes. Spectra were acquired at each wavelength from 230-300 nm at 25°C. For replicate experiments, spectra were acquired at each wavelength from 252-272 nm. Each reported spectrum is an average of 2 scans with a response time of 2s/nm. Data were buffer subtracted, normalized to provide molar residue ellipticity values, and smoothed over 5 nm<sup>40</sup>.

Fluorescence: Fluorescence spectroscopy experiments were performed using a Horiba Scientific Fluorolog Model FL3-1/HR320 IR spectrofluorimeter, and data were analyzed with Kaleidagraph v.3.5 (Synergy software). RNA oligonucleotides were prepared to a concentration of 5  $\mu\text{M}$  in 10 mM LiCacodylate (pH 7.0) buffer. RNA was renatured at 95°C for 2 minutes and allowed to cool at room temperature for 15 minutes. Emission spectra were acquired in triplicate from separate experiment using excitation at 270nm and scanning from 280-450 nm at 25°C. The entrance and exit slit widths were both 5 nm and a 0.5 second integration time was used. Data were buffer subtracted, normalized and smoothed over 5 nm.

### 2.4.3 Data Fitting

Titration experiments were performed with KCl to determine the amount of potassium ion ( $K^+$ ) required to drive G-quadruplex formation. Both CD and fluorescence data were processed similarly. Using CD as an example here, to determine  $K^+_{1/2}$  values, ellipticity data at  $\lambda_{max}$  (262nm) for each sequence were fit with KaleidaGraph v. 3.5 (Synergy software) according to the apparent two-state Hill equation (Eq. 2.1) or three-state Hill equation (Eq. 2.2):

$$\varepsilon = \varepsilon_F + \frac{\varepsilon_U - \varepsilon_F}{1 + \left(\frac{[K^+]}{[K^+]_{1/2}}\right)^n} \quad (\text{Eq. 2.1})$$

where  $\varepsilon$  is the molar ellipticity,  $\varepsilon_F$  is the normalized CD signal corresponding to the fully folded GQS;  $\varepsilon_U$  is signal for the unfolded GQS;  $[K^+]$  is the potassium ion concentration,  $[K^+]_{1/2}$  is the potassium ion concentration needed to fold 50% of the RNA, and  $n$  is the Hill coefficient. Choice of Eq. 2.1 or Eq. 2.2 to fit the data was made on the basis of acceptable residuals (see Figures A.8 and A.10).

$$\varepsilon = \frac{\varepsilon_U + \varepsilon_I \left(\frac{[K^+]}{[K^+]_{1/2}1}\right)^{n_1} + \varepsilon_F \left(\frac{[K^+]}{[K^+]_{1/2}1}\right)^{n_1} \left(\frac{[K^+]}{[K^+]_{1/2}2}\right)^{n_2}}{1 + \left(\frac{[K^+]}{[K^+]_{1/2}1}\right)^{n_1} + \left(\frac{[K^+]}{[K^+]_{1/2}1}\right)^{n_1} \left(\frac{[K^+]}{[K^+]_{1/2}2}\right)^{n_2}} \quad (\text{Eq. 2.2})$$

where  $\varepsilon$  is the molar ellipticity,  $\varepsilon_U$  is the normalized CD signal for fully unfolded RNA,  $\varepsilon_I$  is the normalized CD signal for the intermediate RNA, and  $\varepsilon_F$  is the normalized CD signal for the fully folded RNA GQS.  $[K^+]_{1/2}1$  and  $n_1$  are the  $K^+_{1/2}$  and Hill coefficient values for the U-to-I transition, and  $[K^+]_{1/2}2$  and  $n_2$  are the values for the I-to-F transition.

Data shown in Table 2.1 were fit using Eq. 2.1 or Eq. 2.2, and  $K^+_{1/2}$  and  $n$  values reported are the averages of 3 separate experiments; the errors listed are the standard deviations. In addition, data were plotted as fraction folded plots in main text and were fit using the following equation:

$$f_F = \frac{([K^+]/[K_{1/2}^+])^n}{([K^+]/[K_{1/2}^+])^n + 1} \quad (\text{Eq. 2.3})$$

#### 2.4.4 UV-Detected Thermal Denaturation

RNA was prepared to a concentration of 1-25  $\mu\text{M}$  in 10 mM LiCacodylate (pH 7.0). Samples were renatured at 95°C for 1.5 min then allowed to cool to room temperature for 15mins. KCl was added for a final concentration of 5mM or 100 mM (See fig. legend for detail) and samples were incubated at room temperature for 45 min to allow quartets to form. Thermal denaturation was performed on a Gilford Response II spectrophotometer. Data were collected every 0.5°C while heating over the temperature range of 5-95° C. The unfolding transitions were monitored at 295 nm. Data were fit with KaleidaGraph v. 3.5 (Synergy Software) using a Marquadt algorithm for non-linear curve fitting.<sup>41</sup>

#### 2.4.5 G2 GQS Mixing Simulation

Broad response signal from mixing G2s2, G2m1, and G2w2 was generated from the equation previously described.<sup>12</sup> The equation used was as follow

$$\varepsilon = \sum_{i=1}^3 \frac{[K^+]\varepsilon_i}{[K^+] + [K_{1/2}^+]_i} \quad (\text{Eq. 2.4})$$

where  $\varepsilon_i$  is the CD ellipticity amplitude for probe i, and  $K_{1/2,i}^+$  is the dissociation constant for probe i.

## 2.5 Author Contributions

C.K.K and M.E.S performed the CD experiments. C.K.K performed the UV experiments. M.E.S performed the fluorescence experiments. C.K.K and M.E.S performed the data analysis. C.K.K and M.E.S contributed equally to this work. All authors contributed ideas, discussed the data, and wrote the paper.

## 2.6 Acknowledgements

This study was supported by Human Frontier Science Program (HFSP) Grant RGP0002/2009-C to P.C.B and Penn State Summer Discovery Grant to M.E.S. We would like to thank Prof. Sarah Assmann, Prof. Bratoljub Milosavljevic and Dr. Melissa Mullen for helpful discussion.

## 2.7 References

- 1 Moody, E. M. & Bevilacqua, P. C. Folding of a stable DNA motif involves a highly cooperative network of interactions. *J. Am. Chem. Soc.* **125**, 16285-16293 (2003).
- 2 Moody, E. M., Feerrar, J. C. & Bevilacqua, P. C. Evidence that folding of an RNA tetraloop hairpin is less cooperative than its DNA counterpart. *Biochemistry* **43**, 7992-7998 (2004).
- 3 Sattin, B. D., Zhao, W., Travers, K., Chu, S. & Herschlag, D. Direct measurement of tertiary contact cooperativity in RNA folding. *J. Am. Chem. Soc.* **130**, 6085-6087 (2008).
- 4 Solomatin, S. V., Greenfeld, M. & Herschlag, D. Implications of molecular heterogeneity for the cooperativity of biological macromolecules. *Nat. Struct. Mol. Biol.* **18**, 732-734 (2011).

- 5 Mandal, M. *et al.* A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* **306**, 275-279 (2004).
- 6 Erion, T. V. & Strobel, S. A. Identification of a tertiary interaction important for cooperative ligand binding by the glycine riboswitch. *RNA* **17**, 74-84 (2011).
- 7 Butler, Ethan B., Xiong, Y., Wang, J. & Strobel, Scott A. Structural basis of cooperative ligand binding by the glycine riboswitch. *Chemistry & Biology* **18**, 293-298 (2011).
- 8 Lipfert, J., Sim, A. Y. L., Herschlag, D. & Doniach, S. Dissecting electrostatic screening, specific ion binding, and ligand binding in an energetic model for glycine riboswitch folding. *RNA* **16**, 708-719 (2010).
- 9 Koshland, D., Goldbeter, A. & Stock, J. Amplification and adaptation in regulatory and sensory systems. *Science* **217**, 220-225 (1982).
- 10 Vallee-Belisle, A., Ricci, F. & Plaxco, K. W. Thermodynamic basis for the optimization of binding-induced biomolecular switches and structure-switching biosensors. *Proc. Natl. Acad. Sci. U S A* **106**, 13802-13807 (2009).
- 11 Cerminara, M., Desai, T. M., Sadqi, M. & Munoz, V. Downhill protein folding modules as scaffolds for broad-range ultrafast biosensors. *J. Am. Chem. Soc.* **134**, 8010-8013 (2012).
- 12 Vallee-Belisle, A., Ricci, F. & Plaxco, K. W. Engineering biosensors with extended, narrowed, or arbitrarily edited dynamic range. *J. Am. Chem. Soc.* **134**, 2876-2879 (2012).
- 13 Lin, J. & Liu, D. An optical pH sensor with a linear response over a broad range. *Analytica Chimica Acta* **408**, 49-55 (2000).
- 14 Drabovich, A. P., Okhonin, V., Berezovski, M. & Krylov, S. N. Smart aptamers facilitate multi-probe affinity analysis of proteins with ultra-wide dynamic range of measured concentrations. *J. Am. Chem. Soc.* **129**, 7260-7261 (2007).
- 15 Huppert, J. L. Hunting G-quadruplexes. *Biochimie* **90**, 1140-1148 (2008).

- 16 Sannohe, Y. & Sugiyama, H. Overview of formation of G-quadruplex structures. *Curr. Protoc. in Nucleic Acid Chem.* **Chapter 17, Unit 17.2**, pp1-17, Wiley, New York (2010).
- 17 Halder, K. & Hartig, J. S. RNA quadruplexes. *Met. Ions Life Sci.* **9**, 125-139 (2011).
- 18 Hud, N. V., Smith, F. W., Anet, F. A. L. & Feigon, J. The selectivity for K<sup>+</sup> versus Na<sup>+</sup> in DNA quadruplexes is dominated by relative free energies of hydration: a thermodynamic analysis by <sup>1</sup>H NMR *Biochemistry* **35**, 15383-15390 (1996).
- 19 Mullen, M. A. *et al.* RNA G-quadruplexes in the model plant species *Arabidopsis thaliana*: prevalence and possible functional roles. *Nucleic Acids Res.* **38**, 8149-8163 (2010).
- 20 Zhang, A. Y. Q., Bugaut, A. & Balasubramanian, S. A sequence-independent analysis of the loop length dependence of intramolecular RNA G-quadruplex stability and topology. *Biochemistry* **50**, 7251-7258 (2011).
- 21 Bugaut, A. & Balasubramanian, S. A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes. *Biochemistry* **47**, 689-697 (2008).
- 22 Takenaka, S. & Juskowiak, B. Fluorescence detection of potassium ion using the G-quadruplex structure. *Anal Sci* **27**, 1167-1167 (2011).
- 23 Halder, K., Wieland, M. & Hartig, J. S. Predictable suppression of gene expression by 5'-UTR-based RNA quadruplexes. *Nucleic Acids Res.* **37**, 6811-6817 (2009).
- 24 Phan, A. T. Human telomeric G-quadruplex: structures of DNA and RNA sequences. *FEBS J.* **277**, 1107-1117 (2009).
- 25 Lane, A. N., Chaires, J. B., Gray, R. D. & Trent, J. O. Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.* **36**, 5482-5515 (2008).

- 26 Viglasky, V., Tluckova, K. & Bauer, L. The first derivative of a function of circular dichroism spectra: biophysical study of human telomeric G-quadruplex. *Eur Biophys J* **40**, 29-37 (2011).
- 27 Mullen, M. A., Assmann, S. M. & Bevilacqua, P. C. Toward a digital gene response: RNA G-quadruplexes with fewer quartets fold with higher cooperativity. *J. Am. Chem. Soc.* **134**, 812-815 (2012).
- 28 Serganov, A., Polonskaia, A., Phan, A. T. n., Breaker, R. R. & Patel, D. J. Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature* **441**, 1167-1171 (2006).
- 29 Mendez, M. A. & Szalai, V. A. Fluorescence of unmodified oligonucleotides: A tool to probe G-quadruplex DNA structure. *Biopolymers* **91**, 841-850 (2009).
- 30 Dao, N. T., Haselsberger, R., Michel-Beyerle, M.-E. & Phan, A. T. Following G-quadruplex formation by its intrinsic fluorescence. *FEBS Lett.* **585**, 3969-3977 (2011).
- 31 (The sequences chosen here are either directly from a known gene, such as in the case of G2w2, G2m2, G2s2, and G4m1, or are closely related to these and representative of classes known to exist in genes).
- 32 (UA-rich loops led to Hill constants approaching unity, which may be because of the formation of intermediates involving base pairing of the G's with the loops.).
- 33 Woodson, S. A. Compact intermediates in RNA folding. *Annu Rev Biophys* **39**, 61-77 (2010).
- 34 Weikl, T. R., Palassini, M. & Dill, K. A. Cooperativity in two-state protein folding kinetics. *Protein Sci* **13**, 822-829 (2004).
- 35 (Notably, we previously reported folding intermediates in related G3 sequences, observed through three-state CD-detected titrations, presence of intermediates on native gels, and RNase T1 protection.).

- 36 Doyle, D. A. *et al.* The structure of the potassium channel: molecular basis of K<sup>+</sup> conduction and selectivity. *Science* **280**, 69-77 (1998).
- 37 Vargas, Diana Y. *et al.* Single-molecule imaging of transcriptionally coupled and uncoupled splicing. *Cell* **147**, 1054-1065 (2011).
- 38 Liu, J., Cao, Z. & Lu, Y. Functional nucleic acid sensors. *Chem. Rev.* **109**, 1948-1998 (2009).
- 39 Mok, W. & Li, Y. Recent progress in nucleic acid aptamer-based biosensors and bioassays. *Sensors* **8**, 7050-7084 (2008).
- 40 Sosnick, T. R. Characterization of tertiary folding of RNA by circular dichroism and urea. *Curr. Protoc. in Nucleic Acid Chem.*, 11.15. 11-11.15. 10 (2001).
- 41 Siegfried, N. A. & Bevilacqua, P. C. Thinking inside the box: designing, implementing, and interpreting thermodynamic cycles to dissect cooperativity in RNA and DNA folding. *Methods Enzymol.* **455**, 365-393 (2009).

## Chapter 3

### **The Effect of Loop Sequence and Loop Length on the Intrinsic Fluorescence of G-Quadruplexes**

[Published as a paper entitled “The Effect of Loop Sequence and Loop Length on the Intrinsic Fluorescence of G-Quadruplexes” by Chun Kit Kwok, Madeline E. Sherlock, and Philip C. Bevilacqua in *Biochemistry* **52**, 3019-3021 (2013).] Author contributions are listed in Section 3.4.

#### **3.1 Abstract**

Guanine quadruplex structures (GQS) exhibit unique spectroscopic features, including an inverse melting profile at 295 nm, distinctive circular dichroism features, and intrinsic fluorescence. Herein, we investigate effects of loop sequence and loop length on the intrinsic fluorescence of 13 DNA GQSs. We report label-free fluorescence enhancements upon intramolecular GQS formation of up to 16-fold and a shift in fluorescence emission maximum to the visible light portion of the spectrum. Effects can be understood in the context of available NMR GQS structures. Intrinsic fluorescence of GQS may be useful for nucleic acid studies and for the development of label-free detection methods.

#### **3.2 Introduction**

Guanine quadruplex structures (GQS) have been of great interest for decades. They are found in DNA telomere repeats and the promoter region of certain oncogenes, such as *c-myc*, *c-kit*, and *bcl-2*<sup>1-3</sup>. Guanine-rich sequences in DNA or RNA can potentially form a GQS if they

contain the pattern  $G_xL_aG_xL_bG_xL_cG_x$ , where  $x \geq 2$ , and loops a, b, and c are  $\geq 1$  and can be of the same or different length and sequence. The four stretches of G's interact with one another to form 2 or more quartets that stack upon each other, with the backbone in a parallel or antiparallel geometry. Formation of the GQS requires dehydrated  $K^+$  or  $Na^+$  to be present between or within the stacked quartet planes, which stabilizes charge accumulation from having four adjacent strands<sup>4</sup>. Detailed ion binding studies have been performed on diverse GQS contexts under various ionic conditions, and the general consensus is that they prefer  $K^+$  over other ions<sup>5-7</sup>.

The unique structure and folding of GQS lead to distinctive spectroscopic features as compared to other nucleic acid motifs. For example, GQS has a characteristic “inverse” or hypochromic melting profile when monitored at 295nm, in which absorbance decreases as the structure unfolds<sup>8</sup>. Shifting of the wavelength from the standard 260 nm to 295 nm is due to the extended conjugation from four bases interacting, while absorbance decrease is likely due to loss of partial additivity of the guanine dipoles within a quartet. The GQS also has characteristic patterns in circular dichroism (CD) experiments: for a parallel topology there is a positive peak near 260-265 nm with a minimum near 245 nm, whereas for an antiparallel topology there are positive peaks at 245 and 295 nm with a minimum near 260-265 nm<sup>9</sup>. The intrinsic fluorescence of DNA GQS was recently reported<sup>10</sup> and is also likely due to the extended conjugation of the four-base quartet. Fluorescence studies of DNA GQS to date have been conducted primarily on the telomeric GQS sequence<sup>11-13</sup>. Our lab recently reported that intramolecular RNA GQS exhibit intrinsic fluorescence as well. We demonstrated that the folding cooperativity of a GQS, which can be tuned by adjusting the loop sequence or number of guanines in each stretch<sup>14,15</sup>, can be assessed by monitoring GQS fluorescence as a function of  $K^+$  concentration<sup>15</sup>.

There is, however, little knowledge of what parameters influence the intrinsic fluorescence of GQS, including the effects of the DNA loop sequence and length. It seemed likely that loop sequence and length would play important roles for fluorescence given the key roles that

these structural features play in GQS folding and stability<sup>6,16,17</sup>. In this study, we conduct fluorescence measurements to investigate the effect of loop sequence and length on the intrinsic fluorescence of dG2 and dG3 GQS upon  $K^+$  titration. We then characterize the molecularity and folding cooperativity of a highly fluorescent sequence, dG3T, compare it to its A-loop counterpart and available structures, and discuss advantages and potential applications of intrinsic GQS fluorescence.

### 3.3 Result and Discussion

To systematically investigate the effects of loop sequence and length on GQS fluorescence, we designed 13 DNA GQSs with various single-base loops (A, C, or T)<sup>18</sup>, loop lengths (1-3 nt), and G-stretch lengths (2-3). To simplify the study, the loops within each GQS were chosen to comprise a single base but be of variable length, following the pattern “dG<sub>x</sub>L<sub>y</sub>”, where x and y represent the number of quartets and loop nucleotides, respectively. The sequence of the DNA GQS motifs used in this study and their abbreviations are provided in Table 3.1<sup>19</sup>.

The GQS topology was determined by monitoring the ellipticity at 5  $\mu$ M DNA and 10 mM Li-Cacodylate (pH 7.0) at 1M  $K^+$ . The CD and fluorescence spectra for all 13 sequences are provided in Figures B.1-B.13, and the GQS topologies are summarized in Table 3.1. The CD results reveal that GQS with loops of 1 nucleotide (i.e. “dG<sub>2</sub>L” and “dG<sub>3</sub>L”) have parallel topologies as expected<sup>6,20</sup>. GQS with longer loops were found by comparison to CD to be parallel, antiparallel, or a combination of the two topologies (Table 3.1).

In order to compare fluorescence intensity at unfolded, physiological  $K^+$ , and saturated  $K^+$  states, we collected fluorescence data on each sequence under 0, 100 and 1000 mM  $K^+$  conditions (Figure 3.1A). We found that the wavelength of maximum emission varies by ~80 nm among the GQS, ranging from 340-420 nm (Table 3.1 and Figures B.1-B.13). In general, dG<sub>3</sub>L<sub>y</sub>

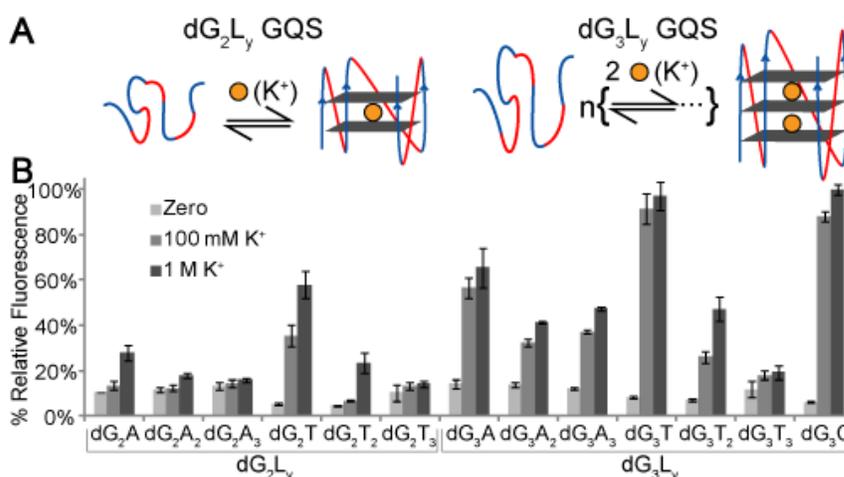
GQSs are more fluorescent than their dG<sub>2</sub>L<sub>y</sub> counterpart (Figure 3.1B). In addition, GQSs with loops of 1 nucleotide have significantly higher fluorescence emission at saturating K<sup>+</sup> condition than loops of 2 or 3 nucleotides. Indeed, for T loops, fluorescence decreases markedly with increasing loop size, for both dG<sub>2</sub> and dG<sub>3</sub> sequences, to the point where there is almost no fluorescence for the T<sub>3</sub>-loop sequences (Figure B.15).

**Table 3.1 DNA GQS Sequence, Topology, and Emission Wavelength Maximum**

dG <sub>x</sub> L <sub>y</sub>	Sequence	Topology <sup>[a]</sup>	λ <sub>em</sub> (nm) <sup>[b]</sup>
dG <sub>2</sub> A	GGAGGAGGAGG	Parallel	400
dG <sub>2</sub> A <sub>2</sub>	GGAAGGAAGGAAGG	Parallel	410
dG <sub>2</sub> A <sub>3</sub>	GGAAAGGAAAGGAAAGG	Parallel	420
dG <sub>2</sub> T	GGTGGTGGTGG	Parallel	385
dG <sub>2</sub> T <sub>2</sub>	GGTTGGTTGGTTGG	Antiparallel	385
dG <sub>2</sub> T <sub>3</sub>	GGTTTGGTTTGGTTTGG	Antiparallel	NA <sup>[c]</sup>
dG <sub>3</sub> A	GGGAGGGAGGGAGGG	Parallel	385
dG <sub>3</sub> A <sub>2</sub>	GGGAAGGGAAGGGAAGGG	Antiparallel	340,420
dG <sub>3</sub> A <sub>3</sub>	GGGAAAGGGAAAGGGAAAGGG	Antiparallel	390
dG <sub>3</sub> T	GGGTGGGTGGGTGGG	Parallel	390
dG <sub>3</sub> T <sub>2</sub>	GGGTTGGGTTGGGTTGGG	Parallel	380
dG <sub>3</sub> T <sub>3</sub>	GGGTTTGGGTTTGGGTTTGGG	Mix <sup>[d]</sup>	340
dG <sub>3</sub> C	GGGCGGGCGGGCGGG	Parallel	390

**Table Footnote:** [a] Topology judged by CD spectroscopy at 1 M K<sup>+</sup>. [b] Values determined by peak maximum in emission spectrum at 1 M K<sup>+</sup> with excitation at 260 nm. [c] Fluorescence was weak and peak wavelength was not available (NA). [d] Assigned as “mix” because CD has positive ellipticity at both 260 and 295 nm.

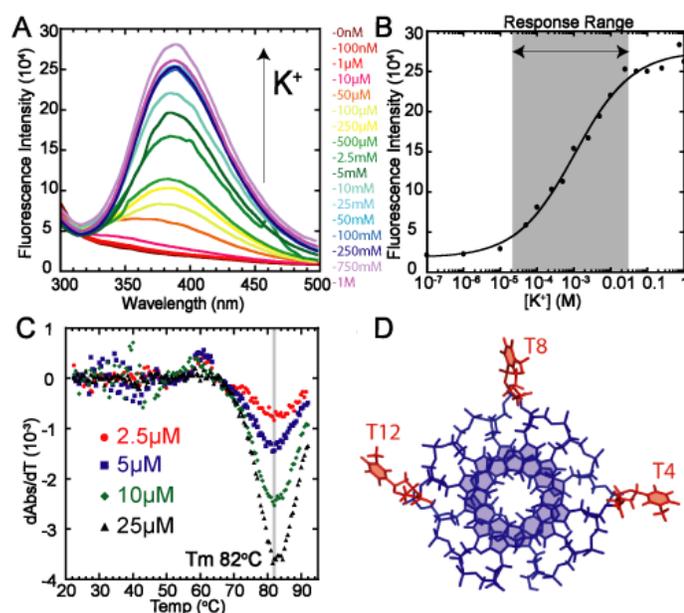
Out of the dG<sub>2</sub>L<sub>y</sub> GQSs tested, dG<sub>2</sub>T is the most fluorescent, being ~2 to 6 times more fluorescence than the other GQS-forming dG<sub>2</sub>L<sub>y</sub> sequences. Among *all* DNA GQSs tested, dG<sub>3</sub>T and dG<sub>3</sub>C achieve the highest overall fluorescence in 1M K<sup>+</sup>, as well as the largest enhancement in fluorescence, ~12- and ~16-fold respectively, upon GQS formation (Figure 3.1B).



**Figure 3.1 Experimental Design and Fluorescence Titration Results**

(A) GQS folding model for  $dG_2L_y$  GQS (top) and  $dG_3L_y$  GQS (bottom). The guanine-stretches ( $G_x$ ) are depicted in blue, and the loops ( $L_y$ ) are in red.  $dG_2L_y$  and  $dG_3L_y$  GQS with parallel topologies are depicted in panel A, although antiparallel quartets are found too (Table 3.1). (B) Relative fluorescence intensity of individual GQS upon  $K^+$  titration. Fluorescence intensity for each GQS was obtained at  $\lambda_{em}$  shown in Table 3.1; for  $dG_2T_3$ , we chose the fluorescence intensity at 340 nm (see Fig. B.6B). For direct comparison, experiments were performed at 5  $\mu$ M GQS concentration for all candidates, with identical instrument settings, and the highest fluorescence intensity was normalized as 100%.

We selected  $dG_3T$  for further spectroscopic and thermodynamic evaluation because NMR structural information is available for it<sup>21</sup>. Full  $K^+$  titrations were performed, which led to a  $K^+_{1/2}$  value of  $1.2 \pm 0.4$  mM and a Hill coefficient ( $n$ ) of  $0.57 \pm 0.02$  (Figures 3.2A and 3.2B). The lower than unity value of  $n$  indicates that GQS folding is anticooperative, which is also reflected in the wide, ~1000-fold analyte response range. Such anticooperativity is consistent with our recent GQS folding studies on long RNA GQS<sup>14,15</sup>. In addition, UV-melting was monitored at 295nm at  $dG_3T$  concentrations ranging from 2.5-25  $\mu$ M, and the melting temperature ( $T_m$ ) was found to be independent of concentration (Figure 3.2C), indicating that the fluorescence signal observed at 5  $\mu$ M  $dG_3T$   $K^+$  titration is due to intramolecular GQS folding.



**Figure 3.2 Biophysical Characterization of dG<sub>3</sub>T.**

(A)-(B) Emission spectra and K<sup>+</sup> titration plot monitored at 390 nm for dG<sub>3</sub>T. K<sup>+</sup><sub>1/2</sub> and the Hill coefficient (n) are  $1.2 \pm 0.4$  mM and  $0.57 \pm 0.02$  using Eq. 3.1. The average was obtained from three separate experiments, and the error is standard deviation. Response range (10 % - 90 % signal) of dG<sub>3</sub>T is shaded in gray in (B). (C) UV-melting of dG<sub>3</sub>T monitored at 295 nm at various concentrations. The vertical bar shows the position of the T<sub>m</sub>, which is invariant with DNA concentration. (D) NMR structure of d(GGGT)<sub>4</sub> (PDB ID 2LE6)<sup>21</sup>. The 3' terminal T (T16) is not shown. The loops (T4, T8, T12) and G-quartets are colored in red and blue, respectively.

Our above results indicate that the increase in fluorescence observed upon K<sup>+</sup> titration was due to formation of GQS, as GQS-specific signals were observed in the CD under these conditions (Figure B.1-B.13). Likewise, non GQS-forming sequences did not exhibit an increase in fluorescence upon K<sup>+</sup> titration<sup>10</sup>. We also tested different anions (fluoride and acetate) and found that fluorescence intensity changes were similar to those with chloride, associating the observed signal change with K<sup>+</sup>-induced GQS formation. In addition, comparison of fluorescence intensity among GQS candidates at 1M K<sup>+</sup> is justified, as plateaus were observed at higher K<sup>+</sup> concentration, indicating that the GQS is at its fully folded states (Figure B.14). The origin of GQS intrinsic fluorescence is proposed to be contributed by <sup>1</sup>G\*G excimer<sup>10</sup>. It is likely that the

loop sequence and length alter the GQS structure, thereby affecting the excited states and energy transfer properties of GQS.

It is therefore of interest to correlate the spectroscopic observations herein with GQS structures. An NMR structure is available for 5'-d(GGGT)<sub>4</sub><sup>21</sup> which has the same CD profile and T<sub>m</sub> within 2 C of dG<sub>3</sub>T<sup>22</sup>. In the structure, the T-loops are extruded and do not interact with the G-quartets (Figure 3.2D). The flipped-out nature of the loop bases is likely responsible for the stronger fluorescence of this GQS, as T's are known to quench fluorescence<sup>23</sup>. As mentioned, dG<sub>3</sub>T and dG<sub>3</sub>C achieved a maximum of ~12 and 16-fold increase in fluorescence upon GQS folding. Since dG<sub>3</sub>C had similar relative fluorescence intensity and emission wavelength maximum as dG<sub>3</sub>T, and the same topology as dG<sub>3</sub>T (both parallel) (Table 3.1), it is likely that the one-pyrimidine loops of both dG<sub>3</sub>C and dG<sub>3</sub>T are extruded. Moreover, the strong decrease in fluorescence with T-loop length (Figure B.15) likely arises because lengthening the extruded loops leads to enhanced collisional quenching by the thymines.

As compared to dG<sub>3</sub>C and dG<sub>3</sub>T, dG<sub>3</sub>A has significantly weaker fluorescence. Moreover, among the dG<sub>2</sub>L sequences, dG<sub>2</sub>A has significantly lower fluorescence than dG<sub>2</sub>T. Interestingly, an NMR structure of 5'-d(GGA)<sub>4</sub>, similar in sequence to dG<sub>2</sub>A, reveals that the A's interact with G-quartet to form a heptad structure (Figure B.16)<sup>24</sup>. It is likely that such direct GQS-interactions quench fluorescence in dG<sub>2</sub>A and dG<sub>3</sub>A. Moreover, the greater dependence of fluorescence emission intensity loss on loop length for T-loops than A-loops likely reflects the greater flexibility of the extra-helical T's (Figure B.15); indeed whereas dG<sub>3</sub>T is ~2x more fluorescent than dG<sub>3</sub>A, dG<sub>3</sub>A<sub>3</sub> is ~2x more fluorescent than dG<sub>3</sub>T<sub>3</sub>.

The dependence of maximum emission wavelength on loop sequence can also be correlated with GQS structures. In particular, dG<sub>2</sub>A<sub>1-3</sub> emit at the longest wavelengths with emission maxima of 400-420 nm, whereas dG<sub>2</sub>T<sub>1-2</sub> have emission maxima of ~385 nm (Table 3.1 and Figure B.15). Apparently, the heptads present in the A-loop GQS extend the conjugation of

the G-quartet leading to the fluorescence shift into the visible light portion of the spectrum. Notably, two peaks were observed in the fluorescence emission of dG<sub>3</sub>A<sub>2</sub>, one at 340 nm and the other at 420 nm (Figure B.8), which suggests that the quartets in dG<sub>3</sub>A<sub>2</sub> may experience different environments—one with a heptad and one without.

The high fluorescence signal enhancement and wide dynamic range of K<sup>+</sup> detection of dG<sub>3</sub>T upon GQS folding demonstrates its merit for sensor development and broad response range K<sup>+</sup> detection. The intrinsic fluorescence of GQS can be used in fluorescence melting, kinetics or quadruplex primer amplification (QPA) readouts<sup>22</sup> without the incorporation of unnatural bases such as 2-aminopurine (2-AP), which have been found to affect the T<sub>m</sub> of the dG<sub>3</sub>T GQS<sup>25</sup> and to interfere with certain biological studies. Moreover, the intrinsic fluorescence of dG<sub>3</sub>T could be utilized as a real-time readout to replace GQS ligand binding reporter such as ZnPPIX in G-quadruplex integrated hybridization chain reaction (GQ-HCR) to probe target DNA or RNA sequences<sup>26</sup>.

In summary, we have investigated the intrinsic fluorescence of DNA GQS with various loops sequences and lengths. We observe a wide dynamic range of emission intensity and maxima, and these can be understood in light of available NMR structures. Together with our work on RNA GQS intrinsic fluorescence<sup>15</sup>, the label-free fluorescence of GQS motif described provides a tool for nucleic acid studies and sensor development where fluorescence output is desired.

### 3.4 Materials and Methods

#### 3.4.1 DNA Preparation

DNA oligonucleotides were purchased from Integrated DNA Technologies, Inc (Coralville, Iowa). Sequences and abbreviations are provided in Table 6.1. All DNAs were dialyzed in an eight-well microdialysis apparatus (Gibco-BRL Life Technologies) at a flow rate of 25 mL/min. The DNA samples were dialyzed in three steps: first against 100 mM LiCl for 6 h to replace DNA backbone cation, then double distilled H<sub>2</sub>O for 6 h to remove excess LiCl, and finally 10 mM LiCacodylate (pH 7.0) for >12 h. (Lithium was chosen as the cation since it does not stabilize GQS.) The concentration of the dialyzed DNAs was determined by UV-spectroscopy and stored in -20°C.

#### 3.4.2 Spectroscopic Titration Experiments

Circular Dichroism (CD): CD spectroscopy was performed using a Jasco CD J810 Spectropolarimeter and analyzed with KaleidaGraph v.3.5 (Synergy software). DNA oligonucleotides were prepared to a concentration of 5  $\mu$ M in 10 mM LiCacodylate (pH 7.0). Prior to CD, DNA was renatured at 95°C for 2 min and allowed to cool at room temperature for 15 min. Spectra were acquired at each wavelength from 230-310 nm at 25 °C. Each reported spectrum is an average of 2 scans with a response time of 2 s/nm. Data were normalized to provide molar residue ellipticity values and smoothed over 5 nm<sup>27</sup>.

Fluorescence: Fluorescence spectroscopy experiments were performed using a Horiba Jobin Yvon Fluorolog 3-21 spectrofluorimeter and analyzed with KaleidaGraph v.3.5 (Synergy software). DNA oligonucleotides were prepared to a concentration of 5  $\mu$ M in 10 mM

LiCacodylate (pH 7.0). Prior to fluorescence, DNA was renatured at 95°C for 2 min and allowed to cool at room temperature for 15 min. Emission spectra were acquired in triplicate from separate experiments using excitation at 260 nm and scanning from 300-500 nm at 25 °C. The entrance and exit slit widths were 5 nm, and a 0.4 s integration time was used. Instrumental settings were kept constant for all spectra. Data were normalized and smoothed over 5 nm.

### 3.4.3 Data Fitting

Titration experiments were performed with KCl to determine the amount of potassium ion ( $K^+$ ) required to drive G-quadruplex formation. To determine  $K^+_{1/2}$  values, emission intensity data at  $\lambda_{max}$  for each sequence as a function of  $K^+$  concentration were fit with KaleidaGraph v. 3.5 (Synergy software) according to the apparent two-state Hill equation (Eq.3.1)

$$\varepsilon = \varepsilon_F + \frac{\varepsilon_U - \varepsilon_F}{1 + \left(\frac{[K^+]}{[K^+]_{1/2}}\right)^n} \quad (\text{Eq. 3.1})$$

where  $\varepsilon$  is the fluorescence signal,  $\varepsilon_F$  is the fluorescence signal corresponding to the fully folded GQS;  $\varepsilon_U$  is signal for the unfolded GQS;  $[K^+]$  is the potassium ion concentration,  $[K^+]_{1/2}$  is the potassium ion concentration needed to fold 50% of the DNA, and  $n$  is the Hill coefficient.

### 3.4.4 UV-Detected Thermal Denaturation

DNA was prepared to a concentration of 2.5-25  $\mu\text{M}$  in 10 mM LiCacodylate (pH 7.0). Samples were renatured at 95°C for 2 min then allowed to cool to room temperature for 15 min. KCl was added for a final concentration of 5 mM, similar to reported elsewhere<sup>7</sup>, and samples were incubated at room temperature for 15 min to allow quartets to form. Thermal denaturation

was performed on a Gilford Response II spectrophotometer. Data were collected every 0.5°C while heating over the temperature range of 20-95°C. The unfolding transitions were monitored at 295 nm. The melting temperature (T<sub>m</sub>) was obtained from the first derivative plot.

### 3.5 Author Contributions

M.E.S and C.K.K performed the CD and fluorescence experiments. C.K.K performed the UV experiments. M.E.S and C.K.K performed the data analysis. M.E.S and C.K.K contributed equally to this work. All authors contributed ideas, discussed the data, and wrote the paper.

### 3.6 Acknowledgements

This study was supported by Human Frontier Science Program (HFSP) Grant RGP0002/2009-C to P.C.B and Penn State Summer Discovery Grant to M.E.S. We thank Prof. Christine Keating for use of fluorimeter.

### 3.7 References

- 1 Balasubramanian, S. & Neidle, S. G-quadruplex nucleic acids as therapeutic targets. *Curr. Opin. Chem. Biol.* **13**, 345-353 (2009).
- 2 Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.* **5**, 182-186 (2013).
- 3 Sannohe, Y. & Sugiyama, H. Overview of formation of G-quadruplex structures. *Curr. Protoc. Nucleic Acid Chem.* **Chapter 17**, Unit 17 12 11-17 (2010).
- 4 Halder, K. & Hartig, J. S. RNA quadruplexes. *Met. Ions Life Sci.* **9**, 125-139 (2011).

- 5 Hud, N. V., Smith, F. W., Anet, F. A. L. & Feigon, J. The selectivity for K<sup>+</sup> versus Na<sup>+</sup> in DNA quadruplexes is dominated by relative free energies of hydration: a thermodynamic analysis by <sup>1</sup>H NMR *Biochemistry* **35**, 15383-15390 (1996).
- 6 Bugaut, A. & Balasubramanian, S. A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes. *Biochemistry* **47**, 689-697 (2008).
- 7 Zhang, A. Y. Q., Bugaut, A. & Balasubramanian, S. A sequence-independent analysis of the loop length dependence of intramolecular RNA G-quadruplex stability and topology. *Biochemistry* **50**, 7251-7258 (2011).
- 8 Mergny, J.-L., Phan, A.-T. & Lacroix, L. Following G-quartet formation by UV-spectroscopy. *FEBS Lett.* **435**, 74-78 (1998).
- 9 Vorlíčková, M. *et al.* Circular dichroism and guanine quadruplexes. *Methods* **57**, 64-75 (2012).
- 10 Mendez, M. A. & Szalai, V. A. Fluorescence of unmodified oligonucleotides: A tool to probe G-quadruplex DNA structure. *Biopolymers* **91**, 841-850 (2009).
- 11 Miannay, F.-A., Banyasz, A., Gustavsson, T. & Markovitsi, D. Excited states and energy transfer in G-quadruplexes. *J. Phys. Chem. C* **113**, 11760-11765 (2009).
- 12 Dao, N. T., Haselsberger, R., Michel-Beyerle, M.-E. & Phan, A. T. Following G-quadruplex formation by its intrinsic fluorescence. *FEBS Lett.* **585**, 3969-3977 (2011).
- 13 Hua, Y. *et al.* Cation effect on the electronic excited states of guanine nanostructures studied by time-resolved fluorescence spectroscopy. *J. Phys. Chem. C* **116**, 14682-14689 (2012).
- 14 Mullen, M. A., Assmann, S. M. & Bevilacqua, P. C. Toward a digital gene response: RNA G-quadruplexes with fewer quartets fold with higher cooperativity. *J. Am. Chem. Soc.* **134**, 812-815 (2012).

- 15 Kwok, C. K., Sherlock, M. E. & Bevilacqua, P. C. Decrease in RNA folding cooperativity by deliberate population of intermediates in RNA G-quadruplexes. *Angew. Chem. Int. Ed.* **52**, 683-686 (2013).
- 16 Guédin, A., De Cian, A., Gros, J., Lacroix, L. & Mergny, J.-L. Sequence effects in single-base loops for quadruplexes. *Biochimie* **90**, 686-696 (2008).
- 17 Guédin, A., Gros, J., Alberti, P. & Mergny, J.-L. How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Res.* **38**, 7858-7868 (2010).
- 18 (We did not study loops of G as these would lead to heterogeneous mixtures of states.).
- 19 (According to CD, dG2C, dG3C2 and dG3C3 were found to not form GQS presumably due to competitive Watson-Crick structures (data not shown), and dG2C2 and dG2C3 were not attempted. It appears that for dGxCy to form a GQS, x-y should be  $\geq 2$ .).
- 20 Hazel, P., Huppert, J., Balasubramanian, S. & Neidle, S. Loop-length-dependent folding of G-quadruplexes. *J. Am. Chem. Soc.* **126**, 16405-16415 (2004).
- 21 Do, N. Q., Lim, K. W., Teo, M. H., Heddi, B. & Phan, A. T. Stacking of G-quadruplexes: NMR structure of a G-rich oligonucleotide with potential anti-HIV and anticancer activity. *Nucleic Acids Res.* **39**, 9448-9457 (2011).
- 22 Kankia, B. I. Self-dissociative primers for nucleic acid amplification and detection based on DNA quadruplexes with intrinsic fluorescence. *Anal. Biochem.* **409**, 59-65 (2011).
- 23 Kierzek, R., Li, Y., Turner, D. H. & Bevilacqua, P. C. 5'-Amino pyrene provides a sensitive, nonperturbing fluorescent probe of RNA secondary and tertiary structure formation. *J. Am. Chem. Soc.* **115**, 4985-4992 (1993).
- 24 Matsugami, A. *et al.* An intramolecular quadruplex of (GGA)<sub>4</sub> triplet repeat DNA with a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad, and its dimeric interaction. *J. Mol. Biol.* **313**, 255-269 (2001).

- 25 Johnson, J., Okyere, R., Joseph, A., Musier-Forsyth, K. & Kankia, B. Quadruplex formation as a molecular switch to turn on intrinsically fluorescent nucleotide analogs. *Nucleic Acids Res.* **41**, 220-228 (2013).
- 26 Dong, J., Cui, X., Deng, Y. & Tang, Z. Amplified detection of nucleic acid by G-quadruplex based hybridization chain reaction. *Biosens. Bioelectron.* **38**, 258-263 (2012).
- 27 Sosnick, T. R. Characterization of tertiary folding of RNA by circular dichroism and urea. *Curr. Protoc. in Nucleic Acid Chem.*, 11.15. 11-11.15. 10 (2001).

## Chapter 4

### Investigations into the Origin of DNA G-Quadruplex Fluorescence

#### 4.1 Abstract

The fluorescence properties of DNA G-quadruplexes reported in Chapter 3 are studied more extensively herein. The addition of tailing nucleotides decreases fluorescence emission intensity when located on the 5' end, but the same is not true of the 3' end. G-Quadruplexes have higher fluorescence at low temperature. The presence of  $O_2$  (g) and  $Cl^-$  (aq), which are both well-characterized quenchers of fluorescence, do not lead to additional quenching of G-Quadruplex fluorescence. While the addition of  $Mg^{2+}$  ions has no effect on the fluorescence of GQSs, the crowding agent PEG 200 has a significant impact and increases emission intensity. These results and links to GQS structure are discussed as well as future experiments.

#### 4.2 Introduction

As seen in Chapters 2 and 3, G-Quadruplex structures exhibit fluorescent properties, which can be used in biophysical studies and have been systematically characterized<sup>1-3</sup>. In this chapter, the origin and nature of the fluorescence are further explored. By themselves, the nucleobases have exceptionally low fluorescence quantum yields between  $3.0 \times 10^{-5}$  and  $1.2 \times 10^{-4}$  and a picoseconds lifetime<sup>4</sup>. In duplex DNA, fluorescence increases slightly with a threefold quantum yield increased<sup>5</sup>. Given the exceptionally large extinction coefficients, it is possible for bases to have observable fluorescence even with modest quantum yield, as observable fluorescence emission brightness is the total light emitted and is a product of the extinction

coefficient and quantum yield. On the other hand, DNA fluorescence is significantly enhanced when the quadruplex structure folds, changing from a signal indistinguishable from a buffer signal when no  $K^+$  is present to large signal peaks when approaching the  $K^+_{1/2}$  concentration<sup>2</sup>. Results from Chapter 3 suggest the stacking overlap between the quartets creates conjugated stacking, which we showed can be diminished by increasing the number of loop nucleotides.

As an extension of the systematic loop study, this chapter investigates the effect of adding dangling nucleotides to both the 3' and 5' end, as these could have similar or very different effects from the loop nucleotides. Another goal of this chapter was to test a hypothesis from Phan and co-workers who, through NMR studies as well as fluorescence, proposed that the fluorescence is due to a 5' to 5' end quadruplex dimer<sup>6</sup>. We hypothesized that the addition of tailing nucleotides would create steric clashes and obstruct dimerization, but only with 5' dangling ends; if fluorescence were still observed, it would help disprove their hypothesis. Another approach taken to test the dimer theory was measuring fluorescence at a constant  $K^+$  concentration with different DNA concentrations, as a linear or decreasing relationship between fluorescence intensity and quadruplex concentration would suggest no relationship between dimer formation and the fluorescence property of the quadruplex.

Aside from the dimerization experiments, additional studies in this chapter focus on creating conditions that alter (quench or enhance) the quadruplex fluorescent signal without changing the sequence through variation of experimental parameters. The reason for pursuing this direction is that the fluorescent signal of the quadruplexes is weak, considering these experiments are performed in the  $\mu\text{M}$  range, and most fluorescent dyes for nucleic acid labeling are effective at nM concentrations. The recent development of the Spinach aptamers shows the potential for RNA fluorescent probes in live-cell imaging, but these depend on a fluorescent co-factor to produce signal<sup>7-9</sup>. GQS fluorescence only depends on  $K^+$ , which is abundant *in vivo*, but the current limits in quantum yield and emission range make it difficult to apply. For GQs to

become a useful tool in sensing and nucleic acid design, the readout must be much stronger and easily identifiable at reasonably low concentrations. It is therefore extremely desirable to find conditions where the quantum yield is significantly improved. Identification of the type of quenching that prevents the quadruplex from being highly fluorescent, and the effects of other known chemical quenchers are also useful tools in understanding the fundamental nature of quadruplex fluorescence. These aims were pursued by performing experiments in which the temperature was varied above and below 25°C, and by adding fluorescence quenchers, and divalent metal ions and polymer crowding agents, each of which stabilizes the quartet.

### 4.3 Results and Discussion

**Table 4.1 DNA GQS Names, Sequences, and Relative Fluorescence Enhancement**

Name	Sequence	Fl Enhancement (0.1M K <sup>+</sup> ) <sup>[a]</sup>	Fl Enhancement (1M K <sup>+</sup> ) <sup>[b]</sup>
G3T	<u>GGGTGGGTGGGTGGG</u>	12.7	12.9
G3T 5'T	T <u>GGGTGGGTGGGTGGG</u>	17.7	16.9
G3T 5'TT	TT <u>GGGTGGGTGGGTGGG</u>	3.0	7.9
G3T 5'AA	AA <u>GGGTGGGTGGGTGGG</u>	2.0	3.1
G3T 5'CC	CC <u>GGGTGGGTGGGTGGG</u>	2.4	3.9
G3T 3'TT	<u>GGGTGGGTGGGTGGG</u> TT	19.4	18.6

**Table Footnote:** [a] The fluorescence enhancement is reported as a ratio of the fluorescent signal intensity at the maximum  $\lambda_{em}$  in 100mM KCl to the signal at 0M KCl. [b] The relative fluorescence reported is a ratio of the fluorescent signal intensity at the maximum  $\lambda_{em}$  in 1M KCl to the signal at 0M KCl.

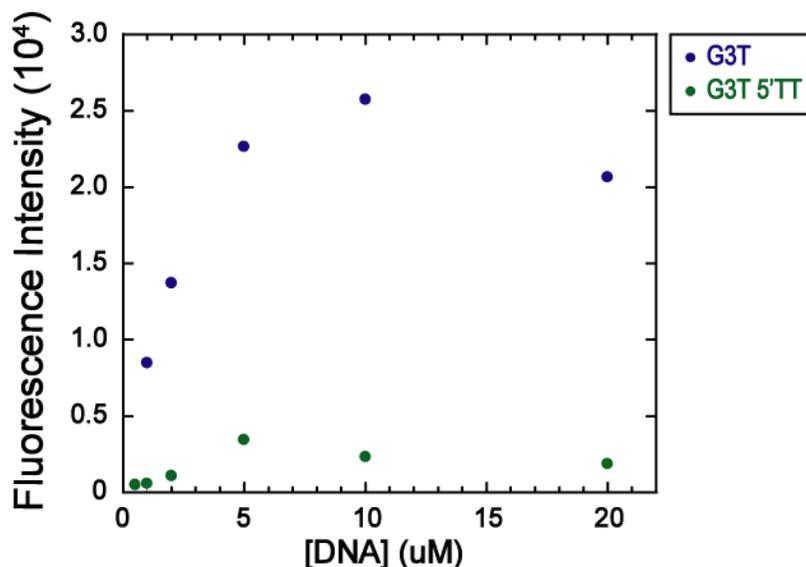
The main GQS chosen for this study was the sequence dG3T (provided in Table 4.1), a parallel GQS which had the highest signal and was thoroughly characterized in Chapter 3 (Figure 3.2). Using this as a core sequence, modifications were made by adding 5' or 3' nucleotides, and the fluorescence was measured at 0, 0.1, and 1M KCl (Figures C.1-C.3). The relative fluorescence intensity enhancement, a ratio of either the 0.1M KCl or 1M KCl signals to the 0 M

KCl signal at the maximum emission wavelength, for each sequence is reported in Table 4.1. From these results, there appears to be an increase in fluorescence from two tailing Ts on the 3' end of the GQS. The addition of one 5' T causes an increased fluorescence intensity but two 5' nucleotides, regardless of identity, caused a decrease in signal, especially at 100mM KCl.

The difference between 3' and 5' results could be related to the difference in stabilization from dangling end stacking; in DNA a 5' dangling end stack is more stabilizing<sup>10</sup>. That the addition of two 5' nucleotides does not completely destroy signal suggests either that dimerization is hindered by dangling ends and is not necessary for fluorescence, or that dimerization is not completely hindered by this addition and the decrease in signal could be due to an equilibrium between dimer and protomer states as suggested by Phan and co-workers. In the first case, the decrease observed could be from collisional quenching or static quenching due to stacking of the dangling ends.

One method to assess the type of quenching was a series of temperature dependent experiments. In the case of 5' nucleotide stacking (static quenching), a higher temperature should melt these stacks, allowing the unstacked quadruplex form to dominate and fluorescence signal should increase. Alternatively, collisional quenching should increase as the temperature is raised due to more movement and more frequent collisions<sup>11</sup>. I observed that fluorescence decreases in the G3T 5'TT sequence by about 33% when the temperature was increased from 25 to 45 °C (data noisy, not shown). Keep in mind that the  $T_m$  of this GQS is 82 °C, making quadruplex unfolding only a minimal contribution to the signal decrease (see Figure 3.2). Interestingly, signal also decreased for the G3T sequence with no dangling ends, which had a 33% decrease when the temperature was only raised to 37 °C, while a 30% signal *enhancement* was observed at 10 °C (Figure C.4). This suggests that more components of the quadruplex structure are contributing to collisional quenching than static quenching from just the dangling ends. These results could also

be interpreted such that the dimer proposed by Phan and co-workers is melting out at higher temperatures, which would also lead to the observed trend.



**Figure 4.1 DNA Concentration Dependence of Fluorescent Signal Intensity.**

Fluorescence Intensity is reported as the signal at the maximum wavelength of emission in 1 M KCl. Each point represents a separate experiment and is plotted at the corresponding DNA concentration.

Another approach taken to investigate the dimerization was varying the DNA concentration used for the fluorescence experiments. These experiments were performed on the G3T and G3T 5'TT sequences using DNA concentrations between 1 and 20  $\mu\text{M}$  (Figure 4.1). Between 1 and 5  $\mu\text{M}$ , the data are linear, but above this concentration the signal begins to level off and drops off before the 25  $\mu\text{M}$  data point. This was not an anticipated outcome of either hypothesis, as dimerization would cause a parabolic increase in signal, and unfortunately seems to be an artifact of the fluorescence. The signal decrease comes from an inner filter effect, which is observed as a decrease in signal when high concentrations of a fluorophore are used or due to the fluorescence coming from the monomer<sup>12</sup>. This most often occurs either when the excitation and emission spectra overlap, or when the fluorophore is very weak, as in this case, and it is

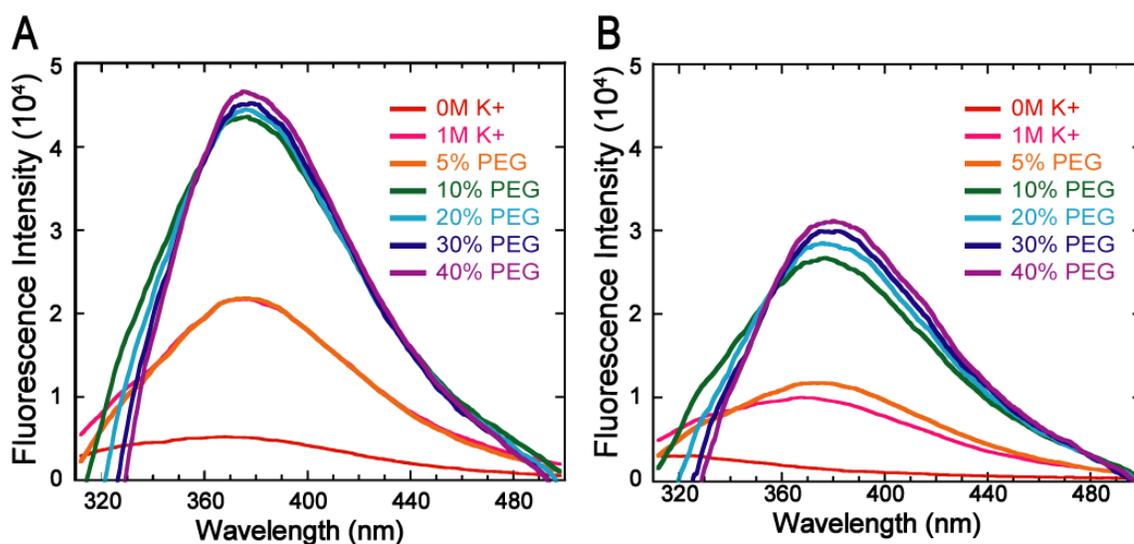
difficult to correct for<sup>13</sup>. Because of this trend, the DNA concentration-dependent experiment does not give great insight into the fluorescent mode of the quadruplex.

To gain deeper insight into the mechanism of the G-Quadruplex as a fluorophore, experiments were performed to investigate if known fluorescence quenchers were responsible for a lower signal than the maximal possible output. The first concern was that the counter anion for  $K^+$  experiments was always  $Cl^-$ , a fluorescence quencher<sup>11</sup>. Given that higher temperature decreased fluorescence, we hypothesized that collisional quenching was important and that  $Cl^-$  would quench fluorescence. To test whether or not the signal was affected by the  $Cl^-$  ion, experiments were repeated using potassium salts with other counterions including KF and KOAc (Figure C.5). Fluorescence at the maximum wavelength was unchanged between all three salts, indicating that chloride has no quenching effect for this fluorophore. An additional compound that might have been present and unknowingly quenching fluorescence was molecular oxygen. This was explored by bubbling  $N_2$  into the cuvette for 20 minutes prior to collecting the emission spectrum; this too showed no difference in signal intensity (Figure C.6).

The absence of a quenching effect may be due to the inaccessibility of the quadruplex to the quencher, or it may be that the dynamic motions of the structure itself are already causing significant quenching. The former hypothesis is difficult to believe given that  $O_2$  is uncharged and small. To explore the effects of self-quenching, additives that are known to compact RNA folding were tested to see if reduced dynamics or increased stabilization of the structure would result in higher signal intensity.

The first additive tested was  $Mg^{2+}$ , which is ubiquitously recognized as an RNA tertiary structure stabilizer<sup>14</sup>. It seemed plausible that in the background of  $K^+$  present in significant quantities to stabilize the hydrogen bonding of the quadruplex,  $Mg^{2+}$  would help stabilize the structure by binding with the phosphate groups. By relieving the negative charge repulsions of the backbone, the structure might become more compact, which would reduce dynamic self-

quenching and result in a higher quantum yield; however the addition of 10 mM  $Mg^{2+}$  in a background of 1M  $K^+$  did not result in an increase in emission intensity (Figure C.7). While the  $Mg^{2+}$  ions are likely binding with the phosphate backbone, this binding event was not significant enough to reduce dynamic quenching.



**Figure 4.2 Fluorescence Enhancement with a PEG 200 Crowding Agent.**

Fluorescence Intensity is reported as a function of emission wavelength. PEG200 was added to a cuvette already containing 1 M KCl (pink trace) and was added up to the final concentration percentage reported in the legend. Signal is normalized by a background of 10 mM LiCac, 1M KCl, and PEG 200 at the concentration corresponding to the concentration of each trace. It should be noted that the PEG 200 solution used exhibited signal that tailed off before the max  $\lambda_{em}$ , but its subtraction caused the lack of overlap at lower wavelength. (A) Titration of G3T sequence. (B) Titration of G3T 5'TT sequence.

The other additive tested was polyethylene glycol with a molecular weight of 200 (PEG 200). PEG is a polymer commonly used as a crowding agent to simulate cellular conditions *in vitro*<sup>15,16</sup>. By adding this viscous polymer, the Quadruplex should become more compact due to size exclusion. At 10% PEG 200, the fluorescent signal increase of the G3T sequence was 90% and the G3T 5'TT was 173% (Figure 4.2). This was a more significant enhancement in intensity than the temperature experiments and suggests that the compaction of the structure locks the quadruplex into the fluorescent mode and decreases random motions. Additionally, PEG has been shown to dehydrate the backbone, which could be rigidifying the structure or  $H_2O$  has a

quenching effect<sup>17</sup>. From results of CD-monitored  $K^+$  titrations, the quadruplex should be nearly 100% folded at 1M  $K^+$ , but even at this extremely high salt concentration, the PEG had a significant effect. This experiment leads to the interesting conclusion that a quadruplex can be completely folded without being fully stable and compact.

#### 4.4 Future Directions

While these studies have resulted in the furthered understanding of G-quadruplex fluorescence, some questions remain unanswered and many additional questions have arisen. There are a number of experiments, which are described below, that were planned but not finished by the thesis submission deadline.

Two sequences of interest are the G3T with three T nucleotides added on the 5' or 3' end to test if the trend of quenching and non-quenching, respectively, continues as the number of dangling nucleotides increases. The dangling end experiments should also be repeated in RNA to see if the trend is reversed due to the reversal of 5' and 3' tailing nucleotide stacking stability. The same experiments should also be performed on an antiparallel GQS, because all of these experiments used the parallel dG3T.

Another sequence is an oligonucleotide with the same four G3 stretches as the sequence used throughout the chapter, but without bases in the loops. This would be accomplished by replacing the T site with an abasic spacer. This would be a follow-up to studies in Chapter 3 to see if the trend of increased signal as loop length decreases still holds when the number of nucleotides in the loops becomes zero.

To follow up on the inner filter effect observed in the DNA concentration dependent studies, non-GQS forming DNA will be added to test if the same inner filter effect is achieved. If

this is indeed an artifact of the inner filter effect, then the same results should be observed when the non-interacting DNA is added at the same concentrations as in Figure 4.1.

Since the two quenchers that were explored in this study had no effect, it is likely that most known fluorescence quenchers would have no effect. More known quenchers will be tested, such as  $\Gamma$ ,  $\text{Cu}^{2+}$ ,  $\text{Pb}^{3+}$ , and acrylamide to confirm this hypothesis<sup>11</sup>. If one quencher does have quenching activity, this may lead to insight about the quadruplex by learning about the mode of action of that quencher.

Although the addition of  $\text{Mg}^{2+}$  had no effect in the background of 1M  $\text{K}^+$ , it might have an effect at lower and more physiologically relevant (10 or 100 mM)  $\text{K}^+$  concentrations. At this lower concentration, the backbone will not already be saturated with monovalent ions and might rigidify with a divalent ion. Additionally, polyvalent amines, such as spermine and spermidine, might also neutralize and rigidify the backbone

With the preliminary result of crowding agents being effective, it would be of interest to test other crowding agents of various sizes, such as PEG 8000, dextran, or ficoll, to see if any of them can compact the quadruplex more effectively than PEG 200. A follow-up to the fluorescence studies using crowding agents, it would be desirable to know if quadruplex folding as observed by other methods is also affected by the presence of PEG 200. This would be assessed by performing CD-monitored  $\text{K}^+$  titrations and UV thermal melting in the presence of a PEG 200 background. This would differentiate whether crowding affects the folding equilibrium and  $\text{K}^+_{1/2}$  or only the dynamics of a fully folded quadruplex.

## 4.5 Materials and Methods

### 4.5.1 DNA Preparation

DNA oligonucleotides were purchased from Integrated DNA Technologies, Inc (Coralville, Iowa). Sequences and abbreviations are provided in Table 4.1. All DNAs were dialyzed in an eight-well microdialysis apparatus (Gibco-BRL Life Technologies) at a flow rate of 25 mL/min. The DNA samples were dialyzed in three steps: first against 100 mM LiCl for 6 h to replace DNA backbone cation, then double distilled H<sub>2</sub>O for 6 h to remove excess LiCl, and finally 10 mM LiCacodylate (pH 7.0) for >12 h. The concentration of the dialyzed DNAs was determined by UV-spectroscopy and samples were stored in -20°C.

### 4.5.2 Spectroscopic Titration Experiments

Fluorescence spectroscopy experiments were measured using a Horiba FluoroLog FL3-11. Data were analyzed with KaleidaGraph v.3.5 (Synergy software). DNA oligonucleotides were prepared to a concentration of 5 μM in [DNA] independent studies, and from 0.5-25 μM for [DNA] dependent studies, in 10 mM LiCacodylate (pH 7.0). Prior to fluorescence, DNA was renatured at 95°C for 3 min and allowed to cool at room temperature for 15 min. Emission spectra were acquired using excitation at 260 nm and scanning from 310-500 nm at 25 °C, when no temperature changes are specifically specified. Temperature was controlled using a water chiller circulating fluid through the sample holder. The entrance slit width was 5 nm, exit slit width was 2nm, and a 1 second integration time was used. Instrumental settings were kept constant for all spectra. Data were normalized and smoothed over 5 nm.

## 4.6 Acknowledgements

Financial support was provided by Undergraduate Research Funding from the Penn State Chemistry Department to M.E.S. Thank you to Dr. Mark Maroncelli for the use of the fluorimeter and to Jens Breffke for his help in designing and setting up fluorescence experiments.

## 4.7 References

- 1 Kwok, C. K., Sherlock, M. E. & Bevilacqua, P. C. Decrease in RNA Folding Cooperativity by Deliberate Population of Intermediates in RNA G-Quadruplexes. *Angew. Chem. Int. Ed.* **52**, 683–686 (2013).
- 2 Kwok, C. K., Sherlock, M. E. & Bevilacqua, P. C. Effect of Loop Sequence and Loop Length on the Intrinsic Fluorescence of G-Quadruplexes. *Biochemistry* **52**, 3019–3021 (2013).
- 3 Dao, N. T., Haselsberger, R., Michel-Beyerle, M.-E. & Phan, A. T. Following G-quadruplex formation by its intrinsic fluorescence. *FEBS Lett.* **585**, 3969–3977 (2011).
- 4 Bloomfield, V. A., Crothers, D. M. & Tinoco, I. *Nucleic acids: structures, properties, and functions*. University Science Books, (2000).
- 5 Vayá, I., Gustavsson, T., Miannay, F.-A., Douki, T. & Markovitsi, D. Fluorescence of Natural DNA: From the Femtosecond to the Nanosecond Time Scales. *J. Am. Chem. Soc.* **132**, 11834–11835 (2010).
- 6 Dao, N. T., Haselsberger, R., Michel-Beyerle, M.-E. & Phan, A. T. Excimer Formation by Stacking G-Quadruplex Blocks. *ChemPhysChem* **14**, 2667–2671 (2013).

- 7 Strack, R. L., Disney, M. D. & Jaffrey, S. R. A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat-containing RNA. *Nat. Methods* **10**, 1219–1224 (2013).
- 8 Paige, J. S., Wu, K. Y. & Jaffrey, S. R. RNA Mimics of Green Fluorescent Protein. *Science* **333**, 642–646 (2011).
- 9 Song, W., Strack, R. L., Svensen, N. & Jaffrey, S. R. Plug-and-Play Fluorophores Extend the Spectral Properties of Spinach. *J. Am. Chem. Soc.* **136**, 1198–1201 (2014).
- 10 Bommarito, S., Peyret, N. & Jr, J. S. Thermodynamic parameters for DNA sequences with dangling ends. *Nucleic Acids Res.* **28**, 1929–1934 (2000).
- 11 Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*. (Springer, 2007).
- 12 Miller, J. N. *Stand. Fluorescence Spectrom.* **27–43** Springer Netherlands (1981).
- 13 Parker, C. A. & Rees, W. T. Correction of fluorescence spectra and measurement of fluorescence quantum efficiency. *Analyst* **85**, 587–600 (1960).
- 14 Hud, N. V. *Nucleic acid-metal ion interactions*. RSC Publishing, (2009).
- 15 Zhou, H.-X., Rivas, G. & Minton, A. P. Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.* **37**, 375–397 (2008).
- 16 Strulson, C. A., Boyer, J. A., Whitman, E. E. & Bevilacqua, P. C. Molecular crowders and cosolutes promote folding cooperativity of RNA under physiological ionic conditions. *RNA* (2014).
- 17 Nakano, S., Miyoshi, D. & Sugimoto, N. Effects of Molecular Crowding on the Structures, Interactions, and Functions of Nucleic Acids. *Chem. Rev.* **114**, 2733–2758 (2014).
- .

## Chapter 5

### Heavy Metal Effects on RNA G-Quadruplex Folding and Unfolding

#### 5.1 Abstract

Although GQS interactions with alkali metals have been thoroughly studied, interactions with other metal ions are still largely uncharacterized. In this chapter, RNA GQS interactions with divalent and heavy metal ions were explored, both in terms of ability to unfold the structure and for possible biological roles. Ions studied include  $Mg^{2+}$ ,  $Ca^{2+}$ ,  $Zn^{2+}$ ,  $Cu^{2+}$ ,  $Cd^{2+}$ , and  $Pb^{2+}$ ; of these,  $Cu^{2+}$  was by far the most active metal ion in GQS unfolding, although at concentrations too high for biological relevance, and this process is reversible. Additional effects, such as pH and competition between  $Cu^{2+}$  and  $K^+$  ions were also explored.

#### 5.2 Introduction

RNA and DNA G-Quadruplex folding as a function of monovalent metal ions has been well documented with the stabilization trend of  $K^+ > Na^+ > Li^{+1-4}$ . Studies characterizing alternative metal ion interactions with G-Quadruplexes are much less abundant. At the nucleotide level, metal ions generally bind more strongly to guanine than the other three nucleobases with the strongest binding sites at the N7 and O6<sup>5,6</sup>. A number of transition metals, such as  $Zn^{2+}$ ,  $Cu^{2+}$ ,  $Mn^{2+}$ ,  $Cr^{2+}$ ,  $Ag^+$ ,  $Cd^{2+}$ ,  $Pt^{2+}$  and  $Ni^{2+}$  have the highest affinity for these two sites<sup>5-7</sup>. Additional divalent metal ions such as  $Mg^{2+}$  and, to a lesser extent  $Ca^{2+}$ , bind preferentially to the phosphate groups and not the nucleobase themselves, which stabilizes RNA tertiary structure<sup>8</sup>. It therefore

seems likely that metal ions other than the typically classified monovalent cations will interact with G-Quadruplexes, possibly at positions that would induce GQS unfolding.

Most studies that have been conducted on G-Quadruplex interactions with divalent metal ions either develop engineered GQSs that include a metal ion-binding component or domain outside of the GQS itself, or GQS folding with other metal ions in the absence of stabilizing monovalent ions ( $K^+$  or  $Na^+$ ). A few studies have characterized  $Cd^{2+}$  and  $Pb^{2+}$ -stabilized GQSs and even developed biosensors by exploiting these characteristics<sup>9-11</sup>. Other studies have used additional modifications or introduced compounds to make  $Cu^{2+}$  binding sites on the quadruplex or on a quadruplex-stabilizing ligand, which can also be used for biosensing application<sup>12-14</sup>. Unfortunately, these results do not reflect GQS activity *in vivo* where  $K^+$  and  $Na^+$  are readily available. This led to our interest in heavy metal interactions with RNA G-Quadruplex structures.

While trace amount of certain heavy metal ions are necessary to maintain the normal function in many organisms, they can be extremely toxic if present in too high of a concentration<sup>15</sup>. Organisms of special interest with regards to heavy metals are plants. Heavy metal ions are introduced into the soil in high quantities due to mining, smelting, and other industrial processes, therefore heavy metal ion contamination and other stresses can have a significant impact on crop production worldwide<sup>15</sup>. Plants are sessile and therefore need defense mechanisms and molecular responses to deal with an influx of heavy metal ions to prevent cell death. In *Arabidopsis thaliana*, a model plant organism, many genes have been identified having function in transportation of or response to heavy metal ions.

Genes that respond to this type of stress need to be strictly regulated. When the concentration of heavy metal ions is at a certain minimal level, the proteins that are supposed to remove ions from the cell should not be present. This response needs to be very selective and specific to the exact concentration range where that metal ion starts to build up and could start damaging vital processes. The idea arose in our lab that since G-quadruplexes can have

regulatory function since they can slow or stall enzymes involved in replication<sup>16</sup>, the G-quadruplex itself could have regulatory function of the heavy metal responders at the RNA level. In this model, the quadruplexes would fold and block translation in the presence of potassium and other metal ions under physiological conditions; upon the introduction of heavy metal ions, the quadruplex structure would be unfolded, allow translation, and begin the process of dealing with high levels of that metal ion. This model requires specificity of G-Quadruplex destruction, meaning that a gene coding for response to one specific metal ion should be denatured by that heavy metal with significantly higher affinity than all other heavy metal ions. This hypothesis, as well as a more general biophysical analysis of divalent metal-GQS interactions, was explored in the following experiments.

### **5.3 Results and Discussion:**

#### ***5.3.1 Selection of Putative Quadruplex Sequences with Heavy Metal Function***

In studies that were begun with the work of Dr. Melissa Mullen Davis under the additional guidance of Dr. Sarah Assmann, multiple bioinformatics approaches were used to identify relevant genes that both contain G-Quadruplex sequences and code for something involved in heavy metal ion response. These and many other functions have been identified and characterized using gene ontology (GO) terms. Using amiGO, a bioinformatics searching tool, all of the heavy metal ion-related genes in the coding regions of the *A. thaliana* genome were searched and identified as well as their sequence and location<sup>17</sup>. Another program called Quadparser was used to search the set of sequences output by amigo for sequences that fit the G-Quadruplex motif<sup>18</sup>. For the purposes of the project, a G-Quadruplex sequence was defined as having: 1) at least four repeats of 2) at least two guanines with 3) loops of at least one but no

more than 4) four nucleotides in length. Using these four parameters, the set of sequences with GO terms in heavy metal function were searched and many had quadruplex motifs. From this set, five representative RNA sequences were selected for experiments. These sequences have gene annotations for either Cu<sup>2+</sup> or Cd<sup>2+</sup> transport or response. Only the portion of the sequence with the Quadruplex-containing sequence was used for the following *in vitro* experiments. The five sequences, accession numbers, and gene annotations can be found in Table 5.1.

**Table 5.1 RNA GQS Names, Sequences, Accession Numbers and GO Terms.**

Gene	Sequence	Accession # <sup>[a]</sup>	GO term <sup>[b]</sup>
GGT3	GGUGGCAGGUGGG	AT1G69820	Cd <sup>2+</sup> response
F19G14	GGGGUUGGGGUUGGGGCGGGG	AT3G15780	Cu <sup>2+</sup> response
NRAMP6	GGCAGGAGAUGGCGG	AT1G15960	Cu <sup>2+</sup> /Fe <sup>3+</sup> transport
HMA5	GGCGGCGGAGG	AT1G20580	Cu <sup>2+</sup> ion response
COBL10	GGAAUGCCGGGCGAGGAUGGAGGAGGAGG	AT3G20580	Phytochelatin (Cd <sup>2+</sup> )

**Table Footnote:** [a] Accession Numbers and gene names are reported as assigned by TAIR9. 'AT' is for *Arabidopsis thaliana*, and the following number is which of the 5 chromosomes the gene is found on. [b] GO terms reported as the output from amiGO.

### 5.3.2 RNA G-Quadruplex Titrations with K<sup>+</sup> ions

CD spectroscopy-monitored K<sup>+</sup> titrations were conducted on these five RNA sequences to determine their folding characteristics, such as the K<sup>+</sup> <sub>1/2</sub> value, folding cooperativity, and the conformation (Table 2). All sequences were found to form a parallel GQS as evidence by the characteristic positive peak at 262nm, which is the expected conformation for RNA GQSs<sup>19-21</sup> (Figures D.1-D.5). F19G14 is a G4 GQS and has a three-state response and it should be noted that it is the same as the G4m1 sequence from Chapter 2. This information was used to fix upper and lower baselines of signal response in later experiments to fit Cu<sup>2+</sup> unfolding titration curves. The next step was to choose an appropriate background K<sup>+</sup> concentration, in which to study divalent

and other heavy metal ion effects. I found that four of the five sequences are >95% folded at 100 mM K<sup>+</sup>, which is around the physiological concentration in *A. thaliana*, and therefore this concentration was selected for further metal ions competition studies.

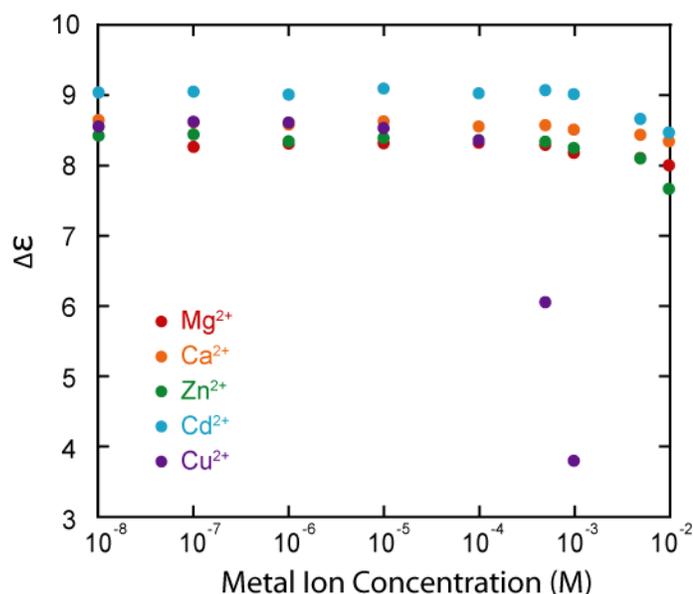
**Table 5.2 RNA GQS K<sup>+</sup> Folding Results.**

Gene	K <sup>+</sup> 1/2 (mM) <sup>[a]</sup>	n <sup>[a]</sup>
GGT3	0.11	1.3
F19G14 (tr1) <sup>[b]</sup>	0.091	0.97
F19G14 (tr2) <sup>[b]</sup>	13	1.2
NRAMP6	250	1.4
HMA5	20	0.90
COBL10	1.5	1.0

**Table Footnote:** [a] GQS folding parameters were obtained from fitting CD titration experiments using a two-state model (Eq. 5.1). [b] In this case (“tr1” and “tr2” for transitions 1 and 2), two distinct transitions were observed, and three-state fitting was performed (Eq. 5.2).

### 5.3.3 RNA G-Quadruplex titrations with Divalent and Heavy Metal ions

I performed divalent metal ion (Mg<sup>2+</sup>, Ca<sup>2+</sup>, Zn<sup>2+</sup>, Cu<sup>2+</sup>, Cd<sup>2+</sup>, Pb<sup>2+</sup>) titration in the background of 100 mM KCl and 10mM LiCac, and found that the copper ion is a more effective quadruplex destabilize for all five sequences studied. A representative overlay of all metal ion effects is shown in Figure 1. Main group divalent metal ions such as Ca<sup>2+</sup> and Mg<sup>2+</sup> have no distinguishable stabilizing or destabilizing effect at these K<sup>+</sup> concentrations. This is consistent with their known property of interacting with the phosphate backbone rather than the bases. Pb<sup>2+</sup> was shown to induce a conformational switch in DNA G-quadruplexes in a potassium ion background and can fold quadruplexes even in the absence of K<sup>+</sup><sup>11</sup>. We found only a slight unfolding effect at 1mM lead, similar to cadmium, and saw no evidence of reorganization or conformational changes.



**Figure 5.1 F19G14 GQS CD Metal Ion Unfolding Comparison.**

CD titrations of the F19G14 sequence at a RNA concentration of 5 $\mu$ M. Ellipticity is plotted vs. metal ion concentration at the  $\lambda_{\text{max}}$  in a 100mM KCl background.

Although only at high concentrations, Cd<sup>2+</sup> and, to a lesser extent, Zn<sup>2+</sup> are also able to partially destabilize one of the GQSs (Cd<sup>2+</sup> titration results in Figures D.7B-D.11B). It became of interest whether more time is needed for GQS to unfold to achieve equilibrium, i.e. slow kinetics. Previous studies have shown quadruplex unfolding in the presence of K<sup>+</sup> ions to be on the order of tens of minutes<sup>22</sup>. In a typical titration, a spectrum is finished measuring about three minutes after each addition. In the experiments with longer time points, the Cd<sup>2+</sup> was added and data was collected at the maximum wavelength for 30-60 minutes after addition (Figure D.6). These results were fit with Eq. 5.3 and show a ~25% decrease in CD signal after the additional 3 hours of experiment time, and most the decrease occurred in the first 10 minutes. Even after the additional time, the overall unfolding effect was still not as much as in the Cu<sup>2+</sup> titrations.

All five sequences undergo unfolding in the  $\sim 10^{-4}$  M concentration range of Cu<sup>2+</sup> ion (Table 5.3, Figures D.7A-D.11A). The steepness of the transition region is much greater than that of the K<sup>+</sup> folding transition, suggesting that Cu<sup>2+</sup>-induced unfolding is more cooperative than K<sup>+</sup>-

induced folding. A possible molecular explanation is that once the  $\text{Cu}^{2+}$  ion gets into a stable binding site and interrupts a necessary bond for GQS stability, unfolding occurs cooperatively afterwards.

**Table 5.3 RNA GQS  $\text{Cu}^{2+}$  Unfolding Results**

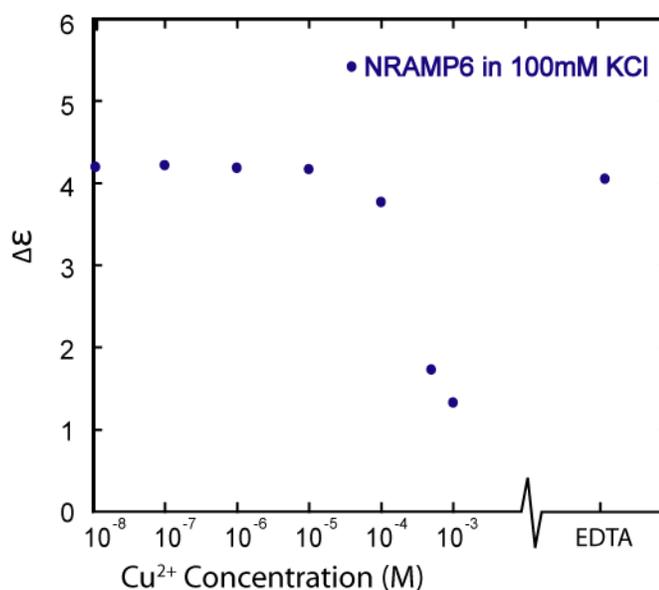
Gene	$[\text{K}^+]$ (M) <sup>[a]</sup>	pH	$\text{Cu}^{2+}$ $_{1/2}$ ( $\mu\text{M}$ ) <sup>[b][c]</sup>	n <sup>[a]</sup>
GGT3	0.1	7	120 $\pm$ 30	2.0 $\pm$ 1.5
GGT3	0.1	5.5	7800 $\pm$ 4000 <sup>[d]</sup>	
F19G14	0.1	7	760 $\pm$ 30	3.0 $\pm$ 0.3
NRAMP6	0.1	7	130 $\pm$ 30	2.3 $\pm$ 1.4
NRAMP6	0.3	7	400 $\pm$ 180 <sup>[d]</sup>	1.3 $\pm$ 0.4
HMA5	0.1	7	670 $\pm$ 30	1.2 $\pm$ 0.5
COBL10	0.1	7	390 $\pm$ 20	1.8 $\pm$ 0.1

**Table Footnote:**[a] This refers to the concentration of  $\text{K}^+$  ions present in solution prior to the titration of  $\text{Cu}^{2+}$  ions. [b] GQS folding parameters were obtained from fitting CD titration experiments using a two-state model (Eq. 5.1). [c] Values reported in the table are average  $\pm$  standard deviation. [d] Large error is due to a poorly defined transition region and the complete lack of lower baseline.

As was stated in the introduction, we selected these RNA G-Quadruplexes based on their annotated function. If this trend had been found, the  $\text{Cu}^{2+}$   $_{1/2}$  would have been much lower for the F19G14, NRAMP6, and HMA5 genes than for GGT3 and COBL1. There also should have been appreciable decrease in signal for the  $\text{Cd}^{2+}$  titrations of GGT3 and COBL10; I found no such correlation with each sequence's  $\text{Cu}^{2+}$  vs.  $\text{Cd}^{2+}$  ion unfolding with  $\text{Cu}^{2+}$  unanimously having more unfolding activity, suggesting that the presence of an RNA GQS, at least in vitro, is not responsible for turning on these genes when too much of that metal ion is present in the cells.

### 5.3.4 Reversible G-Quadruplex Unfolding with EDTA

The concern arose that the change in signal may be due to irreversible cleavage or damage to the RNA and not simply unfolding. To address this issue, 10mM EDTA was added following a copper unfolding titration. , where in the final  $\text{Cu}^{2+}$  concentration was 1 mM. Upon addition of EDTA, which binds  $\text{Cu}^{2+}$  very tightly and with a 1:1 stoichiometry, CD signal was completely restored to its initial upper baseline value. This indicates a reversible folding and unfolding process by  $\text{Cu}^{2+}$  ions and no noticeable cleavage (Figure 5.2). These results are consistent with the findings of Monchaud et.al, which showed  $\text{Cu}^{2+}$  can be reversibly used as a ligand on a modified GQS-forming oligonucleotide<sup>12</sup>.



**Figure 5.2 NRAMP6 GQS CD  $\text{Cu}^{2+}$  titration curve plot with EDTA signal rescue.**

CD titration of the NRAMP6 sequence at a RNA concentration of  $5\mu\text{M}$ . Ellipticity is plotted vs.  $\text{Cu}^{2+}$  concentration at the  $\lambda_{\text{max}}$ , followed by the addition of 10mM EDTA.

### 5.3.5 $\text{Cu}^{2+}$ Titrations in Varying Background $\text{K}^+$ and pH

As previously mentioned, one of the selected sequences is only partially folded at 100mM  $\text{K}^+$ . Due to the smaller available range for signal variation, this sequence was also tested

in a 300mM  $K^+$  background (Figure C.12). When a larger fraction of GQS was folded, the  $Cu^{1/2}$  value subsequently increased, indicating competition and a negative heterotropic linkage between  $K^+$  and  $Cu^{2+}$  ions (Table 5.3).

To further investigate the destabilization mechanism, the potassium and copper titrations were also carried out at pH 5.5. The binding constants did not change for folding with  $K^+$ , but the unfolding by copper was much less effective at this lower pH (Figure C.13).  $Cu^{2+}$  ions have been shown to interact favorably with the N7 and O6 of guanine. This suggests that the unfolding mechanism involves  $Cu^{2+}$  ions coordinating on the Hoogsteen face and replacing the hydrogen bonds necessary for quadruplex formation.  $Cu^{2+}$  ion binding has been shown to have the highest affinity at this site to any other site on the four nucleobases in the gas phase, making this interaction likely to occur<sup>8</sup>. Copper has a smaller atomic radius and charge density than cadmium, which may allow copper to penetrate into the quadruplex during regular vibrations and base pair opening better than cadmium and the other heavy metal ions tested.  $Cu^{2+}$  must additionally overcome a higher enthalpy of hydration<sup>23</sup>, which is significant because the relative stabilization between  $K^+$  and  $Na^+$  is decided by this same principle<sup>1</sup>.

## 5.4 Conclusions

Even though the metal ion binding specificity based on gene function was not found as was hypothesized, this study gave insight into the nature of metal ion-quadruplex interactions outside of the monovalent cations that are normally studied in association with quadruplexes. The biophysical characterization of these interactions and the binding could be useful in other studies that are more focused on designing sensors for heavy metal ions, in addition to testing for sensing of physiological divalent metal ion level through RNA G-quadruplexes *in vivo*.

## 5.5 Future Directions

A major aim in the future should be to focus on more biological conditions, as described by the following proposed experiments

Only the slow kinetics of  $\text{Cd}^{2+}$  ions were explored, therefore it would be of interest to test whether the same decrease occurs with  $\text{Cu}^{2+}$  on the hour timescale. This experiment should also be performed with the heavy metal ions added before  $\text{K}^+$  ions to allow them better access to the structure. The effect of pH, which was only tested at 5.5 and 7, should be extended to 7.5, which is around the physiological value to test if the unfolding effect is even larger as the concentration of  $\text{H}^+$  decreases further.

The unfolding experiments should be repeated at higher temperatures, as this will increase dynamics and potentially allow more metal ions access to binding sites that are disruptive to GQS folding. The addition of PEG 200, or other molecular crowding agents, may also increase the unfolding effect by dehydrating the backbone and creating tighter binding between the GQS and the metal ion.

Another experiment that could be pursued is to add flanking sequences, because these RNA transcripts would not exist as just the GQS portion. The addition of flanking sequences may introduce additional dynamics, which could destabilize the GQS folded state and aid in the unfolding from metal ions.

## 5.6 Materials and Methods

### 5.6.1 RNA Preparation

RNA oligonucleotides were purchased from Dharmacon, Inc. All RNAs were deprotected and desalted, then dialyzed in an eight-well microdialysis apparatus (Gibco-BRL Life

Technologies) at a flow rate of 25 mL/min. The RNA samples were first dialyzed against 100 mM LiCl for 6 hours to replace RNA backbone cation, followed by ddH<sub>2</sub>O for 6 hrs to remove excess LiCl, and finally 10 mM LiCacodylate (pH 7.0) for >12 hours. In the low pH experiments, the final dialysis buffer was 10 mM LiOAc (pH 5.5) The concentration of the dialyzed RNAs was then determined by UV-spectroscopy and stored in -20°C before and after use.

### ***5.6.2 CD Spectroscopic Titration Experiments***

Circular Dichroism (CD): CD spectroscopy was performed using a Jasco CD J810 Spectropolarimeter, and data were analyzed with KaleidaGraph v.3.5 (Synergy software). RNA oligonucleotides were prepared to a concentration of 5 μM in 10 mM LiCacodylate (pH 7.0) buffer. For K<sup>+</sup> titration experiments, RNA was renatured at 95°C for 2 minutes and allowed to cool at room temperature for 15 minutes. For heavy metal ion titration experiments, RNA in 100mM KCl and buffer was renatured at 95°C for 2 minutes and allowed to cool at room temperature for 15 minutes. Spectra were acquired at each wavelength from 220-310 nm at 25°C. Each reported spectrum is an average of 3 scans with a response time of 2s/nm. Data were buffer subtracted, normalized to provide molar residue ellipticity values, and smoothed over 5 nm<sup>24</sup>.

### ***5.6.3 Data Fitting***

Titration were performed with KCl to determine the amount of potassium ion (K<sup>+</sup>) required to drive G-quadruplex formation. Both CD and fluorescence data were processed similarly. Using CD as an example here, to determine K<sup>+</sup><sub>1/2</sub> values, ellipticity data at λ<sub>max</sub> (262nm) for each sequence were fit with KaleidaGraph v. 3.5 (Synergy software) according to the apparent two-state Hill equation (Eq. 5.1) or three-state Hill equation (Eq. 5.2):

$$\varepsilon = \varepsilon_F + \frac{\varepsilon_U - \varepsilon_F}{1 + \left(\frac{[K^+]}{[K^+]_{1/2}}\right)^n} \quad (\text{Eq. 5.1})$$

where  $\varepsilon$  is the molar ellipticity,  $\varepsilon_F$  is the normalized CD signal corresponding to the fully folded GQS;  $\varepsilon_U$  is signal for the unfolded GQS;  $[K^+]$  is the potassium ion concentration,  $[K^+]_{1/2}$  is the potassium ion concentration needed to fold 50% of the RNA, and  $n$  is the Hill coefficient. Choice of Eq. 5.1 or Eq. 5.2 to fit the data was made on the basis of an observed intermediate dead region in the titration.

$$\varepsilon = \frac{\varepsilon_U + \varepsilon_I \left(\frac{[K^+]}{[K^+]_{1/2}]} \right)^{n_1} + \varepsilon_F \left(\frac{[K^+]}{[K^+]_{1/2}]} \right)^{n_1} \left(\frac{[K^+]}{[K^+]_{1/2}]} \right)^{n_2}}{1 + \left(\frac{[K^+]}{[K^+]_{1/2}]} \right)^{n_1} + \left(\frac{[K^+]}{[K^+]_{1/2}]} \right)^{n_1} \left(\frac{[K^+]}{[K^+]_{1/2}]} \right)^{n_2}} \quad (\text{Eq. 5.2})$$

where  $\varepsilon$  is the molar ellipticity,  $\varepsilon_U$  is the normalized CD signal for fully unfolded RNA,  $\varepsilon_I$  is the normalized CD signal for the intermediate RNA, and  $\varepsilon_F$  is the normalized CD signal for the fully folded RNA GQS.  $[K^+]_{1/2}]$  and  $n_1$  are the  $K^+$  and Hill coefficient values for the U-to-I transition, and  $[K^+]_{1/2}]$  and  $n_2$  are the values for the I-to-F transition.

In addition, Eq 5.1 was used to fit the  $\text{Cu}^{2+}$  unfolding experiments. In cases where the lower baseline was ill-defined,  $\varepsilon_U$  was replaced with the experimentally obtained lower baseline value from the  $K^+$  titration experiments.

The slow kinetics experiment was fit as an exponential and linear decay function

$$\varepsilon = Ae^{-kt} + mt + b \quad (\text{Eq. 5.3})$$

where  $\varepsilon$  is the molar ellipticity,  $t$  is time,  $k$  is the exponential decay constant and  $m$  is the linear decay constant.

#### **5.6.4 Bioinformatic Data Sets and Programs**

Sequences searched were obtained from TAIR9 (The Arabidopsis Information Resource)<sup>25</sup>. Only coding sequences were used as an input; the 5' UTR, 3'UTR, intron, and intergenic regions were not included. The program Quadparser was used to search sequences for G-Quadruplex motifs within the coding sequence region<sup>18</sup>. The parameters used in the search were “2414” and “3417” which will search for G<sub>2</sub>L<sub>1-4</sub> and G<sub>3</sub>L<sub>1-7</sub> sequences, respectively. Genes were identified using the keyword searches “Copper ion binding” and “metal ion binding” using amiGO<sup>17</sup>. These were obtained by Dr. Melissa Mullen Davis.

#### **5.7 Acknowledgements**

Financial support was provided by the Penn State Summer Discovery Grant to M.E.S. Thanks to Dr. Sarah Assmann for helpful discussion, and to Dr. Melissa Mullen Davis for her input and help with initial data collection with Quadparser as well as experimental training.

#### **5.8 References**

- 1 Hud, N. V., Smith, F. W., Anet, F. A. & Feigon, J. The selectivity for K<sup>+</sup> versus Na<sup>+</sup> in DNA quadruplexes is dominated by relative free energies of hydration: a thermodynamic analysis by <sup>1</sup>H NMR. *Biochemistry* **35**, 15383–15390 (1996).
- 2 Mullen, M. A. *et al.* RNA G-Quadruplexes in the model plant species *Arabidopsis thaliana*: prevalence and possible functional roles. *Nucleic Acids Res.* **38**, 8149–8163 (2010).

- 3 Zhang, A. Y. Q., Bugaut, A. & Balasubramanian, S. A Sequence-Independent Analysis of the Loop Length Dependence of Intramolecular RNA G-Quadruplex Stability and Topology. *Biochemistry* **50**, 7251–7258 (2011).
- 4 Bugaut, A. & Balasubramanian, S. A Sequence-Independent Study of the Influence of Short Loop Lengths on the Stability and Topology of Intramolecular DNA G-Quadruplexes. *Biochemistry* **47**, 689–697 (2008).
- 5 Kothekar, V. & Dutta, S. Molecular orbital calculations of metal ion interaction with nucleic acid bases. I. Binding of Cu(II) with adenine, guanine, uracil, and cytosine. *Int. J. Quantum Chem.* **12**, 505–514 (1977).
- 6 Robertazzi, A. & Platts, J. Binding of transition metal complexes to guanine and guanine–cytosine: hydrogen bonding and covalent effects. *J. Biol. Inorg. Chem.* **10**, 854–866 (2005).
- 7 Marino, T., Mazzuca, D., Toscano, M., Russo, N. & Grand, A. Gas phase interaction of zinc ion with purine and pyrimidine DNA and RNA bases. *Int. J. Quantum Chem.* **107**, 311–317 (2007).
- 8 Hud, N. V. *Nucleic acid-metal ion interactions*. RSC Publishing, (2009).
- 9 Olsen, C. M., Gmeiner, W. H. & Marky, L. A. Interaction of Cd<sup>2+</sup> with G-Quadruplexes Containing K<sup>+</sup> or Sr<sup>2+</sup>. *J. Biomed. Nanotechnol.* **2**, 62–70 (2006).
- 10 Guo, L. *et al.* A G-quadruplex based label-free fluorescent biosensor for lead ion. *Biosens. Bioelectron.* **35**, 123–127 (2012).
- 11 Liu, W. *et al.* Kinetics and mechanism of conformational changes in a G-quadruplex of thrombin-binding aptamer induced by Pb<sup>2+</sup>. *J. Phys. Chem. B* **115**, 13051–13056 (2011).
- 12 Monchaud, D., Yang, P., Lacroix, L., Teulade-Fichou, M.-P. & Mergny, J.-L. A Metal-Mediated Conformational Switch Controls G-Quadruplex Binding Affinity. *Angew. Chem. Int. Ed.* **47**, 4858–4861 (2008).

- 13 Engelhard, D. M., Pievo, R. & Clever, G. H. Reversible Stabilization of Transition-Metal-Binding DNA G-Quadruplexes. *Angew. Chem. Int. Ed.* **52**, 12843–12847 (2013).
- 14 Chu, J.-F., Wang, Z.-F., Tseng, T.-Y. & Chang, T.-C. A Novel Method for Screening G-quadruplex Stabilizers to Human Telomeres. *J. Chin. Chem. Soc.* **58**, 296–300 (2011).
- 15 Pilon-Smits, E. Phytoremediation. *Annu. Rev. Plant Biol.* **56**, 15–39 (2005).
- 16 Bochman, M. L., Paeschke, K. & Zakian, V. A. DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Genet.* **13**, 770–780 (2012).
- 17 Moore, D. & Meškauskas, A. A comprehensive comparative analysis of the occurrence of developmental sequences in fungal, plant and animal genomes. *Mycol. Res.* **110**, 251–256 (2006).
- 18 Huppert, J. L. & Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **33**, 2908–2916 (2005).
- 19 Halder, K., Wieland, M. & Hartig, J. S. Predictable suppression of gene expression by 5'-UTR-based RNA quadruplexes. *Nucleic Acids Res.* **37**, 6811–6817 (2009).
- 20 Phan, A. T. Human telomeric G-quadruplex: structures of DNA and RNA sequences. *FEBS J.* **277**, 1107–1117 (2010).
- 21 Lane, A. N., Chaires, J. B., Gray, R. D. & Trent, J. O. Stability and Kinetics of G-Quadruplex Structures. *Nucleic Acids Res.* **36**, 5482–5515 (2008).
- 22 Green, J. J., Ying, L., Klenerman, D. & Balasubramanian, S. Kinetics of Unfolding the Human Telomeric DNA Quadruplex Using a PNA Trap. *J. Am. Chem. Soc.* **125**, 3763–3767 (2003).
- 23 Smith, D. W. Ionic hydration enthalpies. *J. Chem. Educ.* **54**, 540 (1977).
- 24 Sosnick, T. R. Characterization of tertiary folding of RNA by circular dichroism and urea. *Curr. Protoc. Nucleic Acid Chem. Ed. Serge Beaucage Al* **11.5** (2001).

- 25 Lamesch, P. *et al.* The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* **40**, D1202–D1210 (2012).

## Appendix A

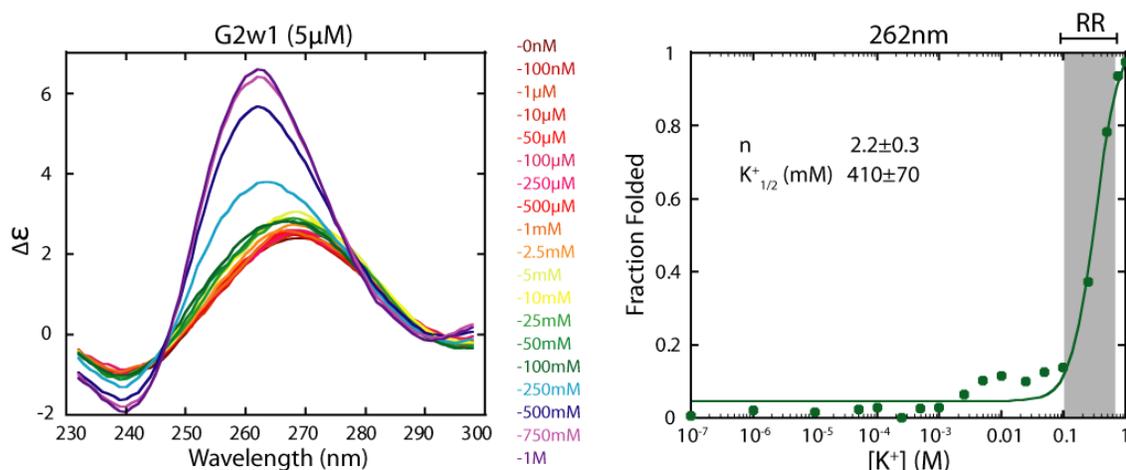
### Supplementary Information for Chapter 2

[Published as Supplementary Online Material for a paper entitled “Decrease of RNA Folding Cooperativity by Deliberate Population of Intermediates in RNA G-Quadruplexes” by Chun Kit Kwok, Madeline E. Sherlock, and Philip C. Bevilacqua in *Angewandte Chemie International Edition* **52**, 683-686 (2013). This paper was recommended by F1000 Faculty. <http://f1000.com/prime/717970147>]

**Table A.1 RNA Oligonucleotide Sequences**

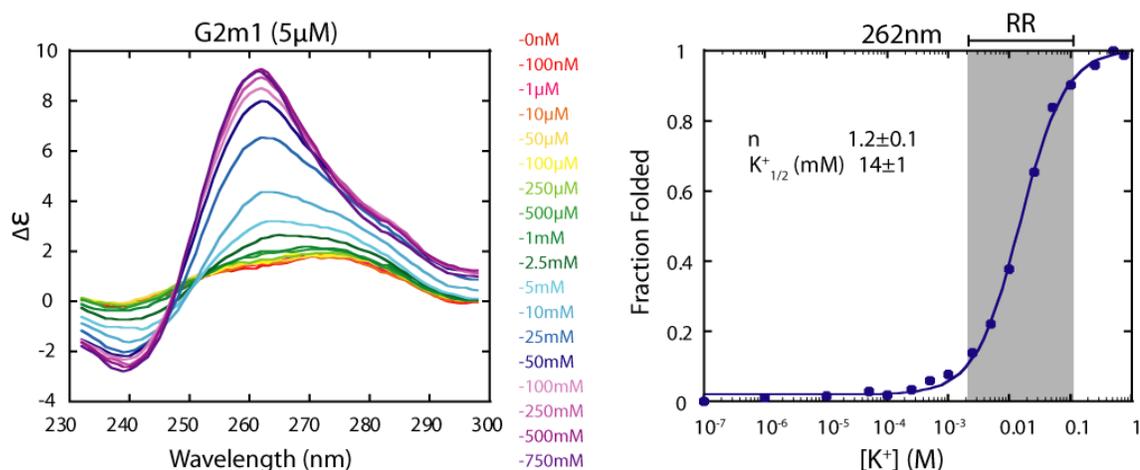
GQS name	Sequence
G2w1	<u>GG</u> UAU <u>A</u> <u>GG</u> UAU <u>A</u> <u>GG</u> ACA <u>GG</u>
G2m1	<u>GG</u> UA <u>GG</u> UA <u>GG</u> C <u>GG</u>
G2s1	<u>GG</u> UA <u>GGU</u> UAG <u>GU</u> AG <u>G</u>
G2w2	<u>GG</u> AUA <u>GG</u> AUA <u>GG</u> AUA <u>GG</u>
G2m2	<u>GG</u> AA <u>GG</u> AA <u>GG</u> AA <u>GG</u>
G2s2	<u>GG</u> A <u>GG</u> A <u>GG</u> A <u>GG</u>
G3m1	<u>GGG</u> UA <u>GGG</u> UA <u>GGG</u> C <u>GGG</u>
G4m1	<u>GGGG</u> UA <u>GGGG</u> UA <u>GGGG</u> C <u>GGGG</u>
G5m1	<u>GGGGG</u> UA <u>GGGGG</u> UA <u>GGGGG</u> C <u>GGGGG</u>
G6m1	<u>GGGGGG</u> UA <u>GGGGGG</u> UA <u>GGGGGG</u> C <u>GGGGGG</u>

This table provides RNA nucleotide sequences corresponding to all GQS used in these experiments. The names of all of the GQS were given based on the length of the G-stretches and whether the sequence is a “strong”, “medium” or “weak” potassium binder. The last number is used to distinguish the two sets of G2’s; one set was designed for this study and the second set was from Mullen *et al.*<sup>1</sup>



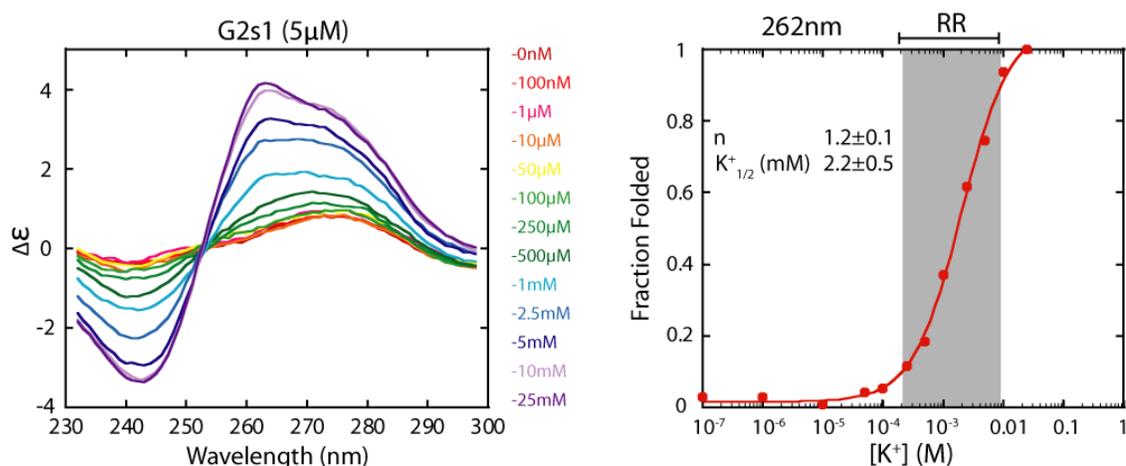
**Figure A.1 G2w1 GQS CD  $K^+$  Titration and Fraction Folded plot.**

CD titrations of the G2w1 sequence at a RNA concentration of 5 $\mu$ M. The left-hand plot is normalized ellipticity versus wavelength, where each trace represents a different  $K^+$  concentration. From this data, the  $\lambda_{\max}$  was determined to be 262nm and the data points at that wavelength were extracted and converted to fraction folded values using the maximum and minimum ellipticity values. The right-hand plot is fraction folded vs.  $K^+$  concentration, which appears as a two-state titration curve that was fit with Eq. 2.1. The shaded gray region highlights the response range (RR), which is defined from 10%-90% folded. (Note that the range of the x-axis is 7 logs in all Supporting Figures to facilitate comparison.) The  $K^+_{1/2}$  and  $n$  values also appear on the right-hand plot, which are the average of three trials  $\pm$  standard deviation. The color of the curve in the right-hand panel matches the color in the main text.



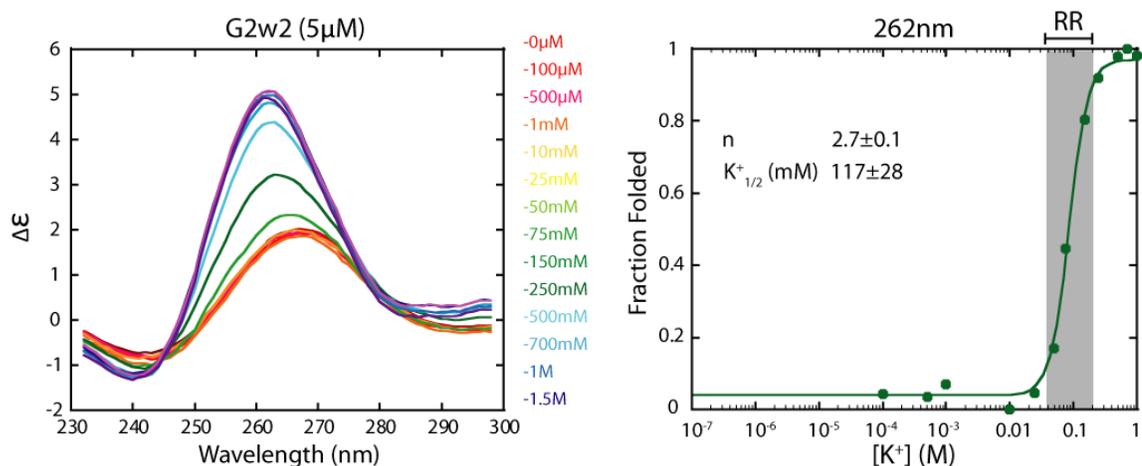
**Figure A.2 G2m1 GQS CD  $K^+$  Titration and Fraction Folded plot.**

CD titrations of the G2m1 sequence at a RNA concentration of 5 $\mu$ M. The left-hand plot is normalized ellipticity versus wavelength, where each trace represents a different  $K^+$  concentration. From this data, the  $\lambda_{\max}$  was determined to be 262nm and the data points at that wavelength were extracted and converted to fraction folded values using the maximum and minimum ellipticity values. The right-hand plot is fraction folded vs.  $K^+$  concentration, which appears as a two-state titration curve that was fit with Eq. 2.1. The shaded gray region highlights the response range (RR), which is defined from 10%-90% folded. The  $K^+_{1/2}$  and  $n$  values also appear on the right-hand plot, which are the average of three trials  $\pm$  standard deviation. The color of the curve in the right-hand panel matches the color in the main text.



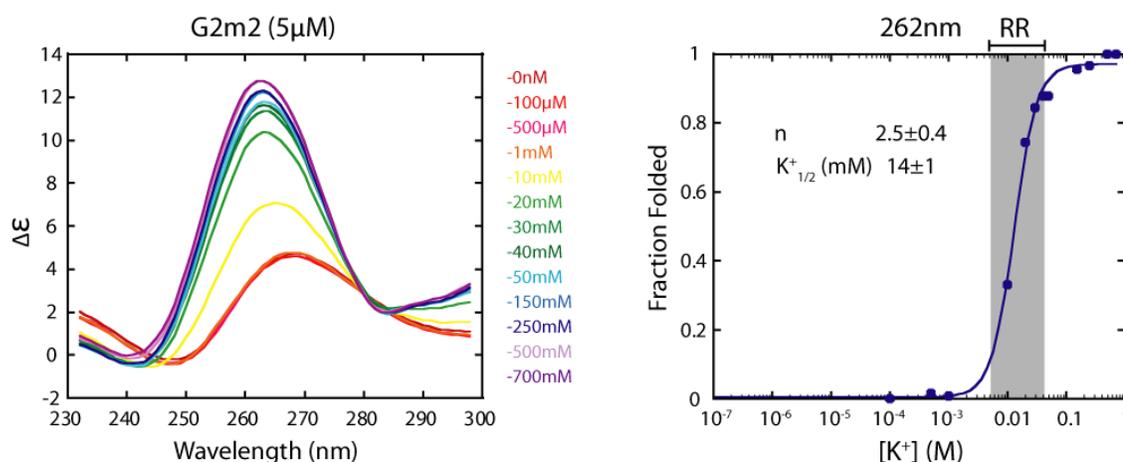
**Figure A.3 G2s1 GQS CD  $K^+$  Titration and Fraction Folded plot.**

CD titrations of the G2s1 sequence at a RNA concentration of 5 $\mu$ M. The left-hand plot is normalized ellipticity versus wavelength, where each trace represents a different  $K^+$  concentration. From this data, the  $\lambda_{\max}$  was determined to be 262nm and the data points at that wavelength were extracted and converted to fraction folded values using the maximum and minimum ellipticity values. The right-hand plot is fraction folded vs.  $K^+$  concentration, which appears as a two-state titration curve that was fit with Eq. 2.1. The shaded gray region highlights the response range (RR), which is defined from 10%-90% folded. The  $K^+_{1/2}$  and  $n$  values also appear on the right-hand plot, which are the average of three trials  $\pm$  standard deviation. The color of the curve in the right-hand panel matches the color in the main text.



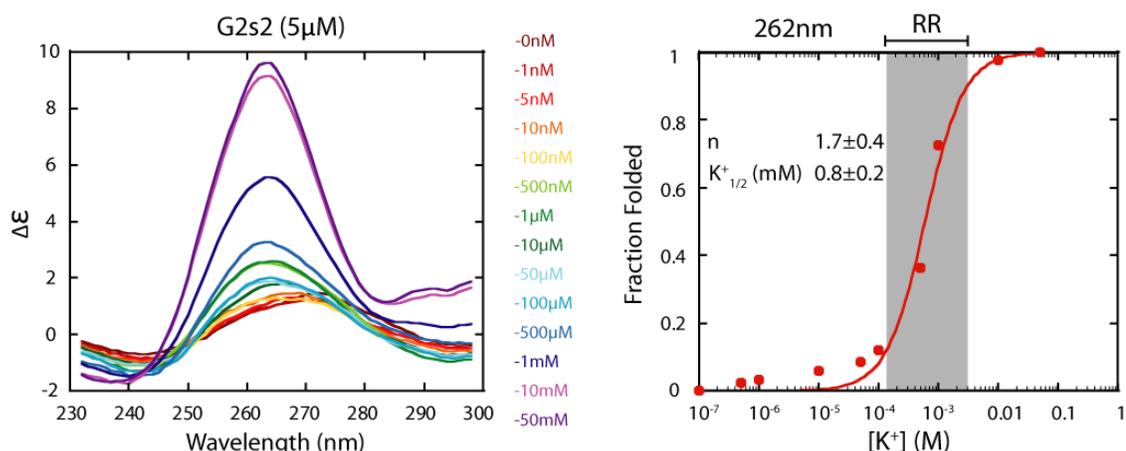
**Figure A.4 G2w2 GQS CD  $K^+$  Titration and Fraction Folded plot.**

CD titrations of the G2w2 sequence at a RNA concentration of 5 $\mu$ M. The left-hand plot is normalized ellipticity versus wavelength, where each trace represents a different  $K^+$  concentration. From this data, the  $\lambda_{\max}$  was determined to be 262nm and the data points at that wavelength were extracted and converted to fraction folded values using the maximum and minimum ellipticity values. The right-hand plot is fraction folded vs.  $K^+$  concentration, which appears as a two-state titration curve that was fit with Eq. 2.1. The shaded gray region highlights the response range (RR), which is defined from 10%-90% folded. The  $K^+_{1/2}$  and  $n$  values also appear on the right-hand plot, which are the average of three trials  $\pm$  standard deviation. The color of the curve in the right-hand panel matches the color in the main text.



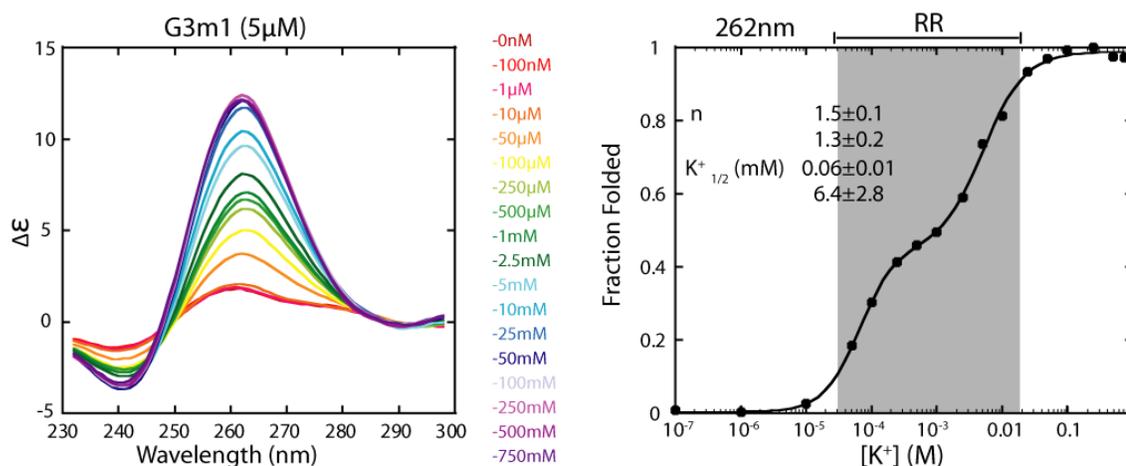
**Figure A.5 G2m2 QQS CD  $K^+$  Titration and Fraction Folded plot.**

CD titrations of the G2m2 sequence at a RNA concentration of  $5\mu\text{M}$ . The left-hand plot is normalized ellipticity versus wavelength, where each trace represents a different  $K^+$  concentration. From this data, the  $\lambda_{\text{max}}$  was determined to be 262nm and the data points at that wavelength were extracted and converted to fraction folded values using the maximum and minimum ellipticity values. The right-hand plot is fraction folded vs.  $K^+$  concentration, which appears as a two-state titration curve that was fit with Eq. 2.1. The shaded gray region highlights the response range (RR), which is defined from 10%-90% folded. The  $K^+_{1/2}$  and  $n$  values also appear on the right-hand plot, which are the average of three trials  $\pm$  standard deviation. The color of the curve in the right-hand panel matches the color in the main text.



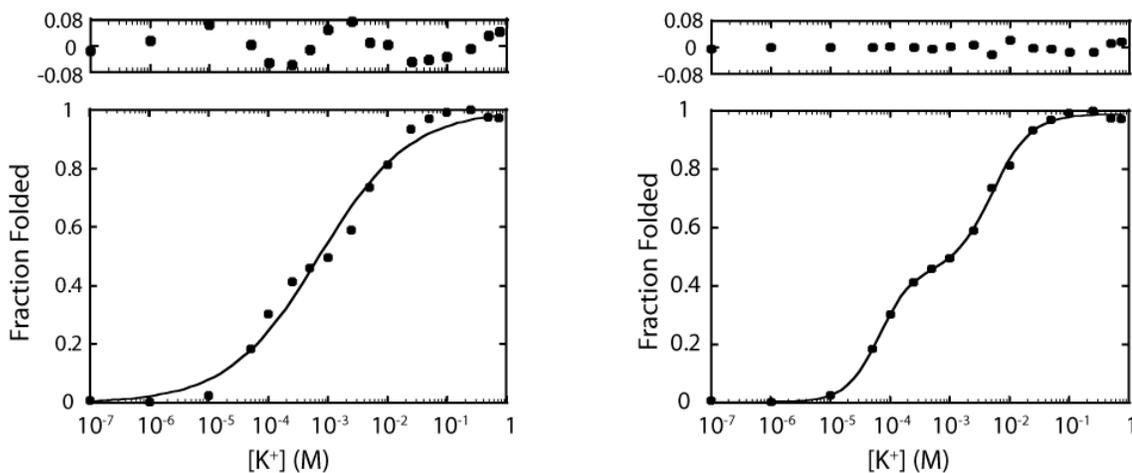
**Figure A.6 G2s2 QQS CD  $K^+$  Titration and Fraction Folded plot.**

CD titrations of the G2s2 sequence at a RNA concentration of  $5\mu\text{M}$ . The left-hand plot is normalized ellipticity versus wavelength, where each trace represents a different  $K^+$  concentration. From this data, the  $\lambda_{\text{max}}$  was determined to be 262nm and the data points at that wavelength were extracted and converted to fraction folded values using the maximum and minimum ellipticity values. The right-hand plot is fraction folded vs.  $K^+$  concentration, which appears as a two-state titration curve that was fit with Eq. 2.1. The shaded gray region highlights the response range (RR), which is defined from 10%-90% folded. The  $K^+_{1/2}$  and  $n$  values also appear on the right-hand plot, which are the average of three trials  $\pm$  standard deviation. The color of the curve in the right-hand panel matches the color in the main text.



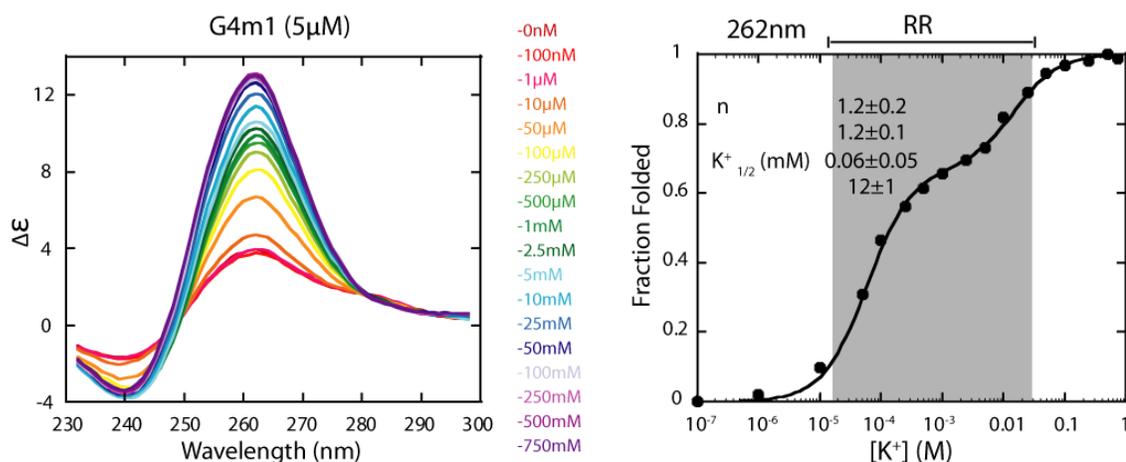
**Figure A.7 G3m1 QQS CD  $K^+$  Titration and Fraction Folded plot.**

CD titrations of the G3m1 sequence at a RNA concentration of  $5\mu\text{M}$ . The left-hand plot is normalized ellipticity vs. wavelength, where each trace represents a different  $K^+$  concentration. From this data, the  $\lambda_{\text{max}}$  was determined to be 262nm and the data points at that wavelength were converted to fraction folded values using the maximum and minimum ellipticity values. The right-hand plot is fraction folded vs.  $K^+$  concentration, which appears as a three-state titration curve that was fit with Eq. 2.2. The shaded gray region highlights the response range (RR), which is defined from 10%-90% folded. The  $K^+_{1/2}$  and  $n$  values for both transitions also appear on the right-hand plot, which are the average of three trials  $\pm$  standard deviation.



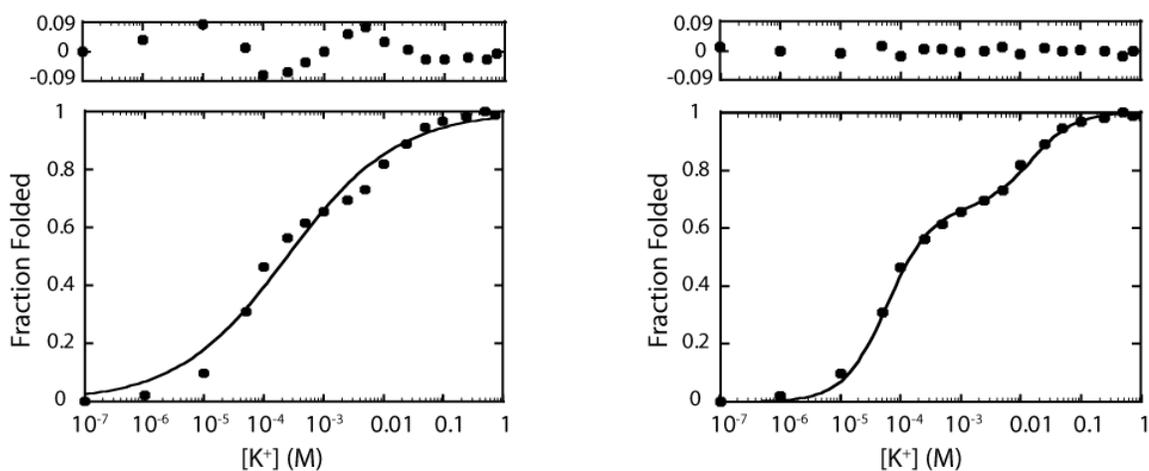
**Figure A.8 G3m1 QQS CD 2 and 3 State Fraction Folded Plot with Residuals.**

CD titration curves at 262nm of the G3m1 sequence. The left-hand plot is fit to a 2-state model with Eq. 2.1, and the right-hand plot is fit to a 3-state model with Eq. 2.2. Above both plots are the residuals, which were calculated by taking the difference between the value calculated by the curve fit equation and the experimentally determined ellipticity value at each  $K^+$  concentration. The sinusoidal pattern in the residuals of the left-hand plot and the linear pattern in the residuals of the right-hand plot confirm that the 3-state model is more appropriate for this sequence.



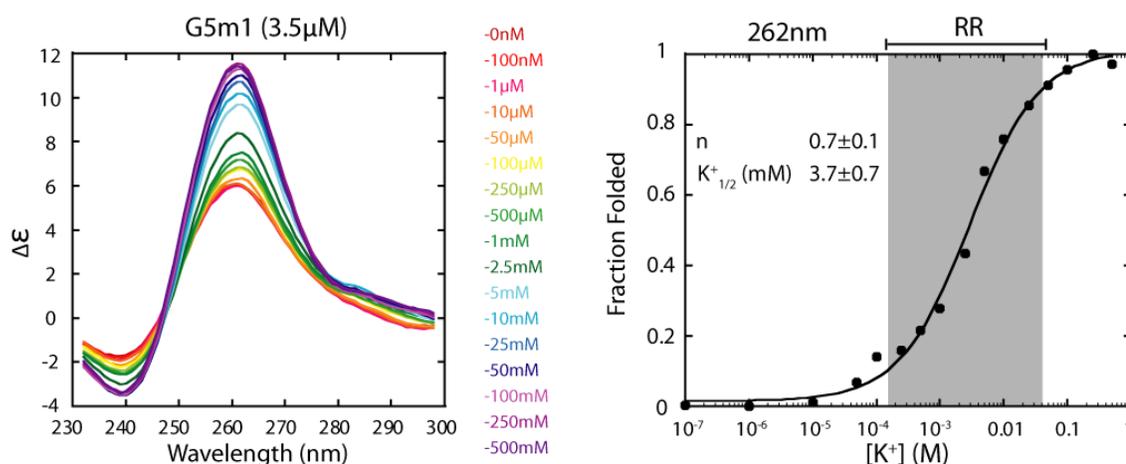
**Figure A.9 G4m1 GQS CD  $K^+$  Titration and Fraction Folded plot.**

CD titrations of the G3m1 sequence at a RNA concentration of  $5\mu\text{M}$ . The left-hand plot is normalized ellipticity vs. wavelength, where each trace represents a different  $K^+$  concentration. From this data, the  $\lambda_{\text{max}}$  was determined to be 262nm and the data points at that wavelength were converted to fraction folded values using the maximum and minimum ellipticity values. The right-hand plot is fraction folded vs.  $K^+$  concentration, which appears as a three-state titration curve that was fit with Eq. 2.2. The shaded gray region highlights the response range (RR), which is defined from 10%-90% folded. The  $K^+_{1/2}$  and  $n$  values for both transitions also appear on the right-hand plot, which are the average of three trials  $\pm$  standard deviation.



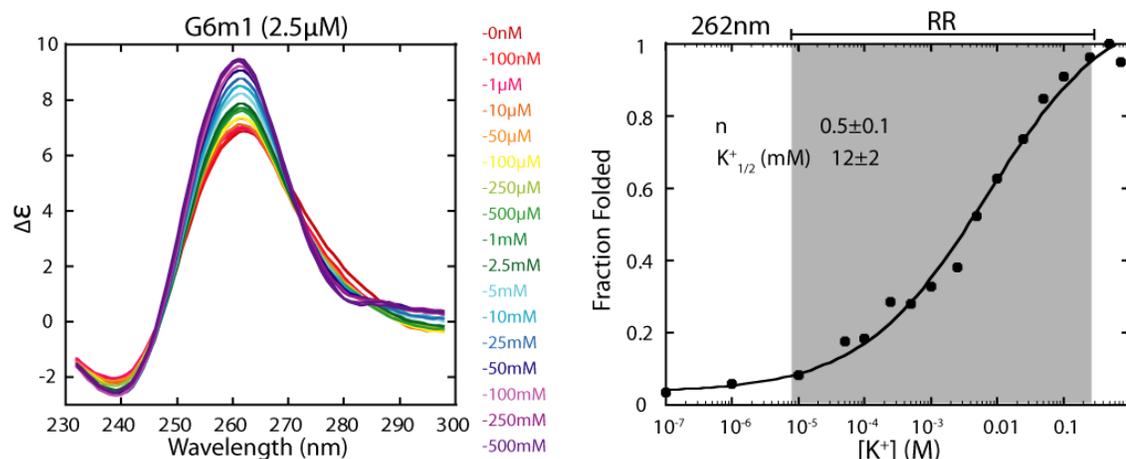
**Figure A.10 G4m1 GQS CD 2 and 3 State Fraction Folded Plot with Residuals.**

CD titration curves at 262nm of the G4m1 sequence. The left-hand plot is fit to a 2-state model with Eq. 2.1, and the right-hand plot is fit to a 3-state model with Eq. 2.2. Above both plots are the residuals, which were calculated by taking the difference between the value calculated by the curve fit equation and the experimentally determined ellipticity value at each  $K^+$  concentration. The sinusoidal pattern in the residuals of the left-hand plot and the linear pattern in the residuals of the right-hand plot confirm that the 3-state model is more appropriate for this sequence.



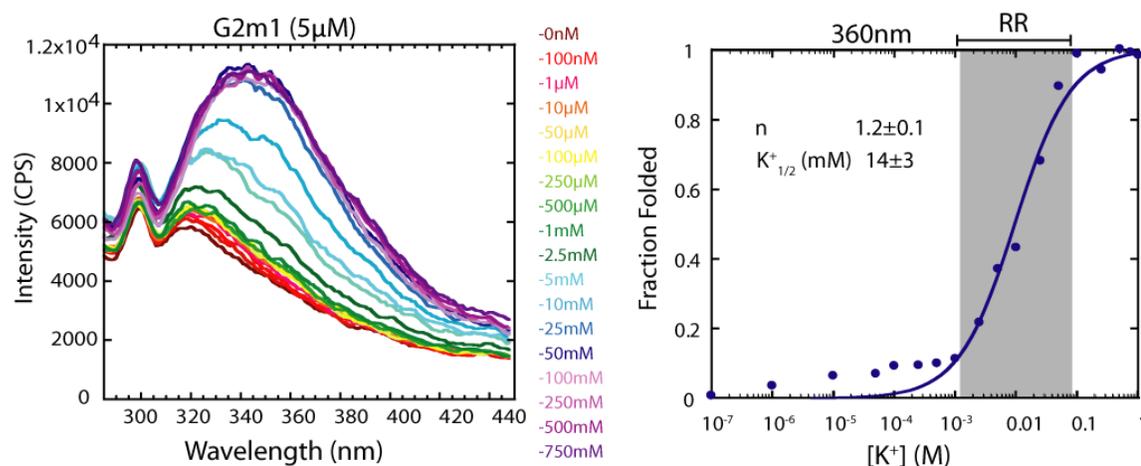
**Figure A.11 G5m1 GQS CD  $K^+$  Titration and Fraction Folded plot.**

CD titrations of the G5m1 sequence at a RNA concentration of  $3.5\mu\text{M}$ . The left-hand plot is normalized ellipticity versus wavelength, where each trace represents a different  $K^+$  concentration. From this data, the  $\lambda_{\text{max}}$  was determined to be 262nm and the data points at that wavelength were extracted and converted to fraction folded values using the maximum and minimum ellipticity values. The right-hand plot is fraction folded vs.  $K^+$  concentration, which appears as a two-state titration curve that was fit with Eq. 2.1. The shaded gray region highlights the response range (RR), which is defined from 10%-90% folded. The  $K^+_{1/2}$  and  $n$  values also appear on the right-hand plot, which are the average of three trials  $\pm$  standard deviation.



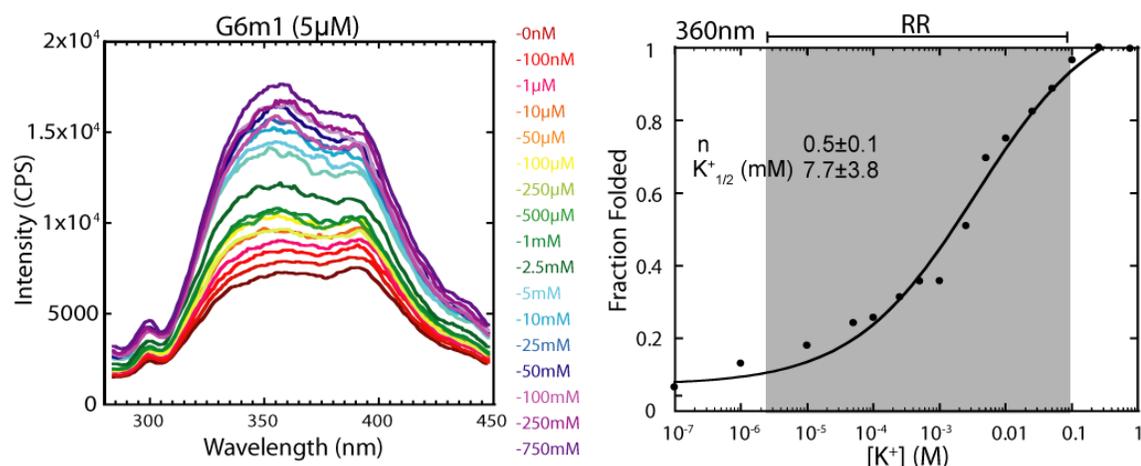
**Figure A.12 G6m1 GQS CD  $K^+$  Titration and Fraction Folded plot.**

CD titrations of the G6m1 sequence at a RNA concentration of  $2.5\mu\text{M}$ . The left-hand plot is normalized ellipticity versus wavelength, where each trace represents a different  $K^+$  concentration. From this data, the  $\lambda_{\text{max}}$  was determined to be 262nm and the data points at that wavelength were extracted and converted to fraction folded values using the maximum and minimum ellipticity values. The right-hand plot is fraction folded vs.  $K^+$  concentration, which appears as a two-state titration curve that was fit with Eq. 2.1. The shaded gray region highlights the response range (RR), which is defined from 10%-90% folded. The  $K^+_{1/2}$  and  $n$  values also appear on the right-hand plot, which are the average of three trials  $\pm$  standard deviation.



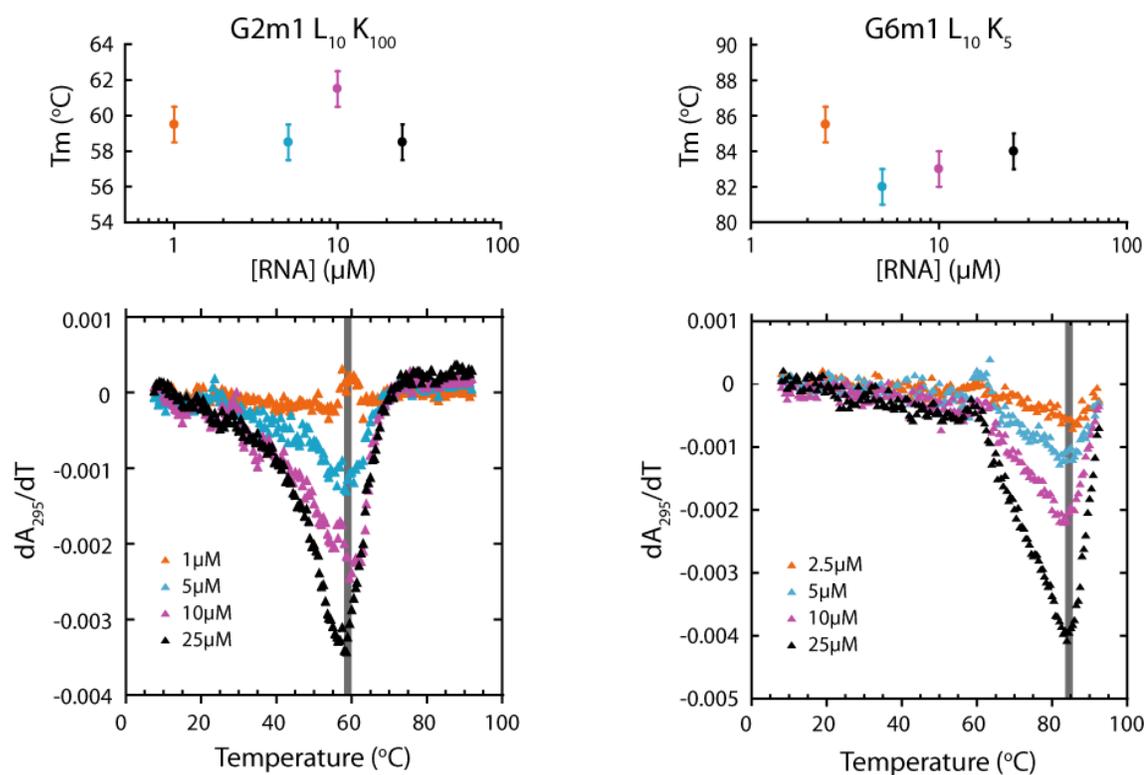
**Figure A.13 G2m1 GQS Fluorescence  $K^+$  Titration and Fraction Folded plot.**

Fluorescence titrations of the G2m1 sequence at a RNA concentration of  $5\mu\text{M}$ . The left-hand plot is normalized intensity in counts per second versus wavelength, where each trace represents a different  $K^+$  concentration. From this data, the  $\lambda_{\text{max}}$  was determined to be 360nm and the data points at that wavelength were extracted and converted to fraction folded values using the maximum and minimum intensity values. The right-hand plot is fraction folded vs.  $K^+$  concentration, which appears as a two-state titration curve that was fit with Eq. 2.1. The  $K^+_{1/2}$  and  $n$  values also appear on the right-hand plot, which are the average of three trials  $\pm$  standard deviation.



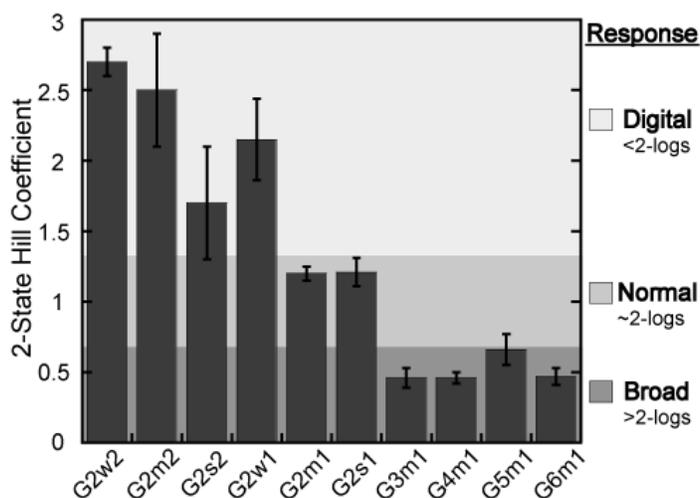
**Figure A.14 G6m1 GQS Fluorescence  $K^+$  Titration and Fraction Folded plot.**

Fluorescence titrations of the G6m1 sequence at a RNA concentration of  $5\mu\text{M}$ . The left-hand plot is normalized intensity in counts per second versus wavelength, where each trace represents a different  $K^+$  concentration. From this data, the  $\lambda_{\text{max}}$  was determined to be 360nm and the data points at that wavelength were extracted and converted to fraction folded values using the maximum and minimum intensity values. The right-hand plot is fraction folded vs.  $K^+$  concentration, which appears as a two-state titration curve that was fit with Eq. 2.1. The  $K^+_{1/2}$  and  $n$  values also appear on the right-hand plot, which are the average of three trials  $\pm$  standard deviation.



**Figure A.15** Dependence of Melting Temperature on RNA Concentrations for G2m1 and G6m1 GQS.

Upper Row: UV melting experiments were conducted, and  $T_m$  was monitored at 295nm, a characteristic feature for GQS. The  $T_m$  values at increasing RNA concentration are plotted to show the concentration independence for both sequences. The  $T_m$  values were extracted from the first derivative graph below with error bars of  $\pm 1$  °C, as reported previously.<sup>2</sup> Lower Row: The first derivative plots are shown to aid in visual inspection of the melting temperature ( $T_m$ ) at four different RNA concentrations. The vertical, gray-shaded bars show the narrow variation in the  $T_m$  values between the four concentrations. The notation ‘L<sub>10</sub>K<sub>100</sub>’ represents buffer and salt conditions of 10 mM LiCacodylate (pH 7.0) and 100 mM KCl; the notation ‘L<sub>10</sub>K<sub>5</sub>’ represents 10 mM LiCacodylate (pH 7.0) and 5 mM KCl. For G2m1, the KCl concentration was chosen such that ~90% of the population of G-quartets were folded according to Eq. 5.1. For G6m1, the transition temperature was above observable temperature range using 100mM KCl, so 5mM KCl was used instead, as per Balasubramanian and colleagues.<sup>3</sup>



**Figure A.16 Hill Coefficients (n) of GQS Sequences Studied**

All GQSs were fit using the apparent two-state model (Eq. 2.1). Results were binned into three classes: digital, normal, and broad. n values are apparent, provided in Table 2.1. Error bars are the standard deviation of three trials. For this figure, G3m1 and G4m1 were ‘forced’ to fit with Eq. 2.1.

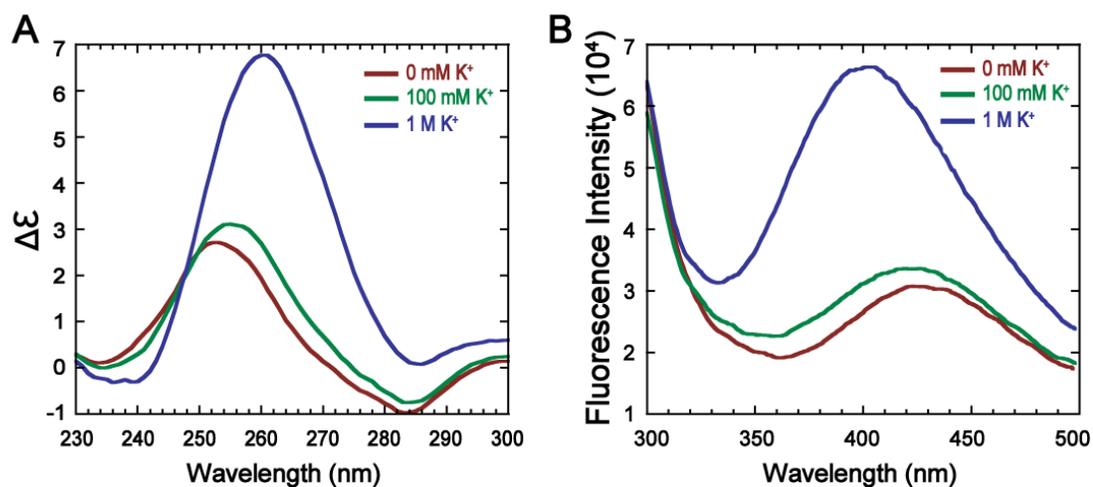
## References

- 1 Mullen, M. A., Assmann, S. M. & Bevilacqua, P. C. Toward a digital gene response: RNA G-quadruplexes with fewer quartets fold with higher cooperativity. *J. Am. Chem. Soc.* **134**, 812-815 (2012).
- 2 Moody, E. M. & Bevilacqua, P. C. Folding of a stable DNA motif involves a highly cooperative network of interactions. *J. Am. Chem. Soc.* **125**, 16285-16293 (2003).
- 3 Zhang, A. Y. Q., Bugaut, A. & Balasubramanian, S. A sequence-independent analysis of the loop length dependence of intramolecular RNA G-quadruplex stability and topology. *Biochemistry* **50**, 7251-7258 (2011).

## Appendix B

## Supplementary Information for Chapter 3

[Published as Supplementary Online Material for a paper entitled “The Effect of Loop Sequence and Loop Length on the Intrinsic Fluorescence of G-Quadruplexes” by Chun Kit Kwok, Madeline E. Sherlock, and Philip C. Bevilacqua in *Biochemistry* **52**, 3019-3021, 2013).]



**Figure B.1** K<sup>+</sup> Titration Result of dG2A using CD and Fluorescence Spectroscopy.

(A) K<sup>+</sup> titration spectra detected by CD. (B) K<sup>+</sup> titration spectra detected by fluorescence.

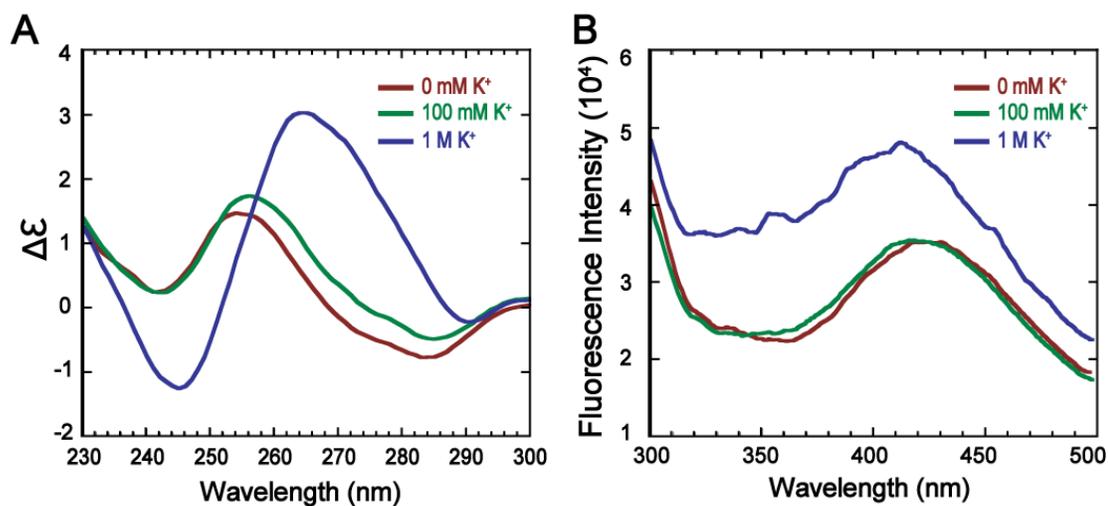


Figure B.2  $K^+$  Titration Result of dG2A2 using CD and Fluorescence Spectroscopy.

(A)  $K^+$  titration spectra detected by CD. (B)  $K^+$  titration spectra detected by fluorescence.

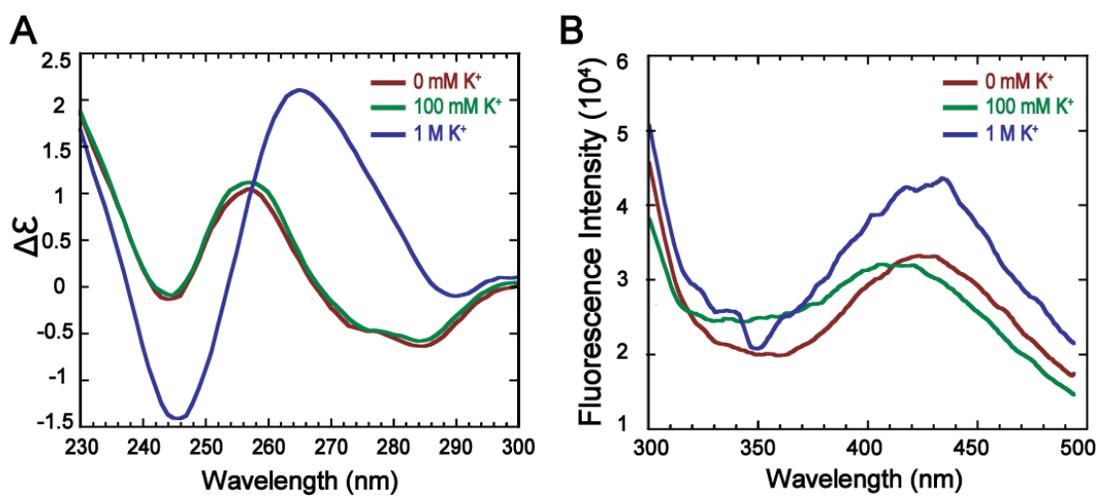


Figure B.3  $K^+$  Titration Result of dG2A3 using CD and Fluorescence Spectroscopy.

(A)  $K^+$  titration spectra detected by CD. (B)  $K^+$  titration spectra detected by fluorescence.

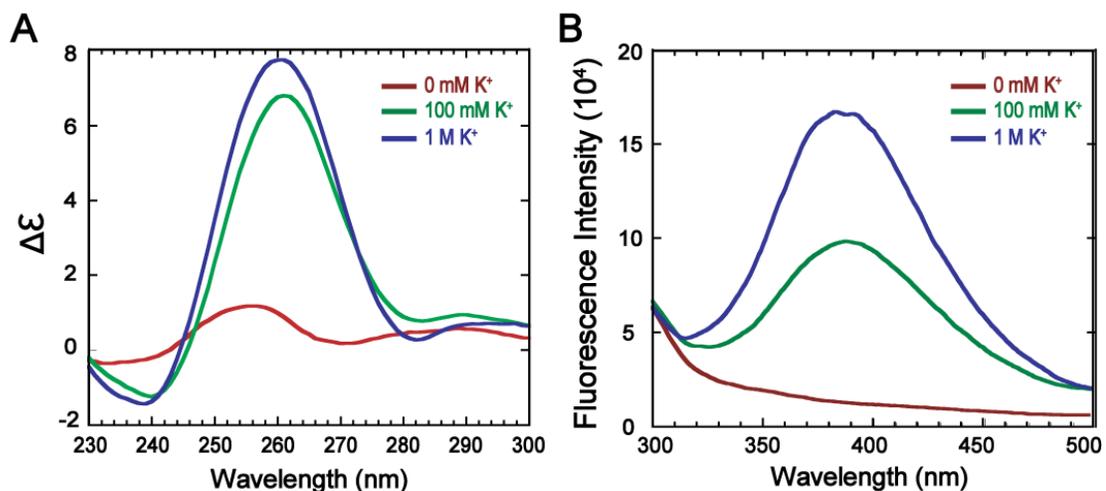


Figure B.4 K<sup>+</sup> Titration Result of dG2T using CD and Fluorescence Spectroscopy.

(A) K<sup>+</sup> titration spectra detected by CD. (B) K<sup>+</sup> titration spectra detected by fluorescence.

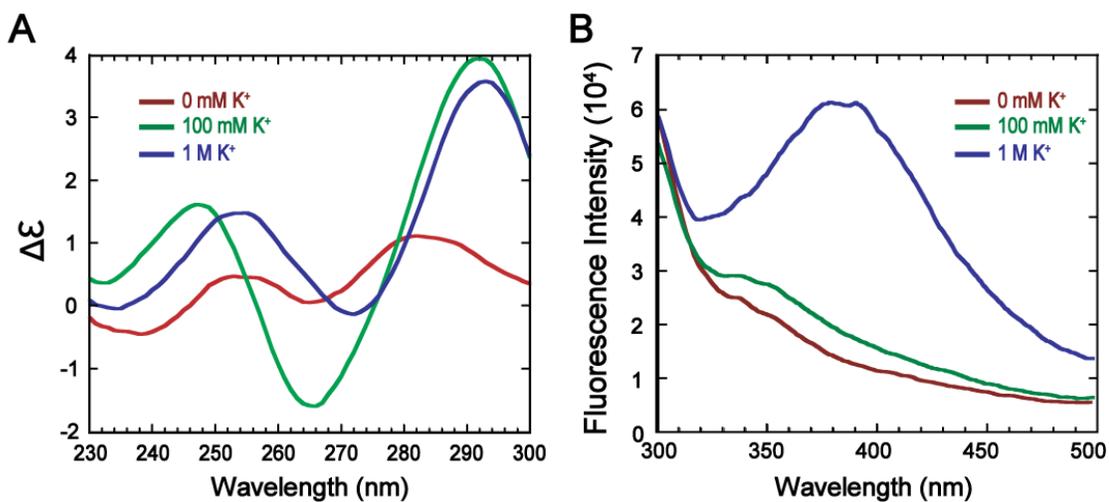


Figure B.5 K<sup>+</sup> Titration Result of dG2T2 using CD and Fluorescence Spectroscopy.

(A) K<sup>+</sup> titration spectra detected by CD. (B) K<sup>+</sup> titration spectra detected by fluorescence.

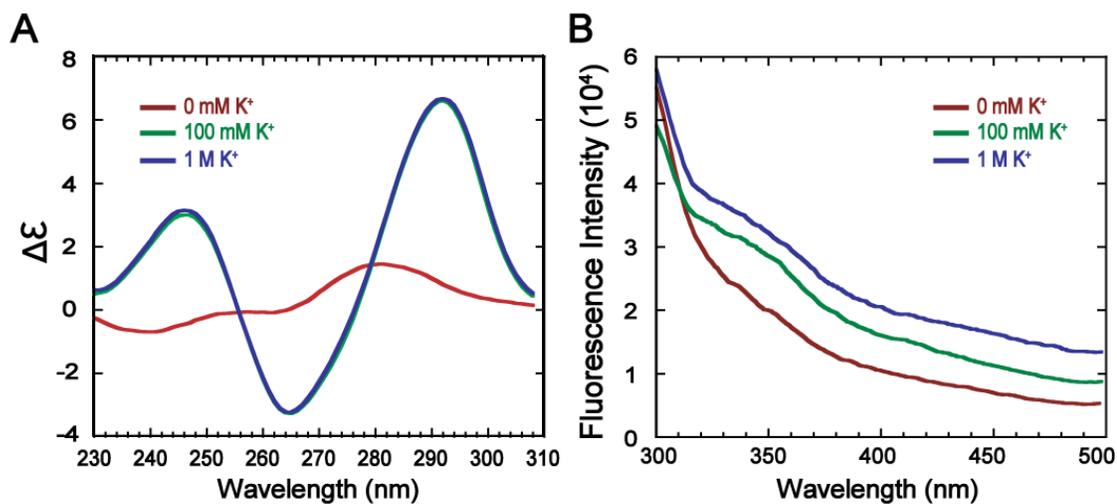


Figure B.6 K<sup>+</sup> Titration Result of dG2T3 using CD and Fluorescence Spectroscopy.

(A) K<sup>+</sup> titration spectra detected by CD. (B) K<sup>+</sup> titration spectra detected by fluorescence.

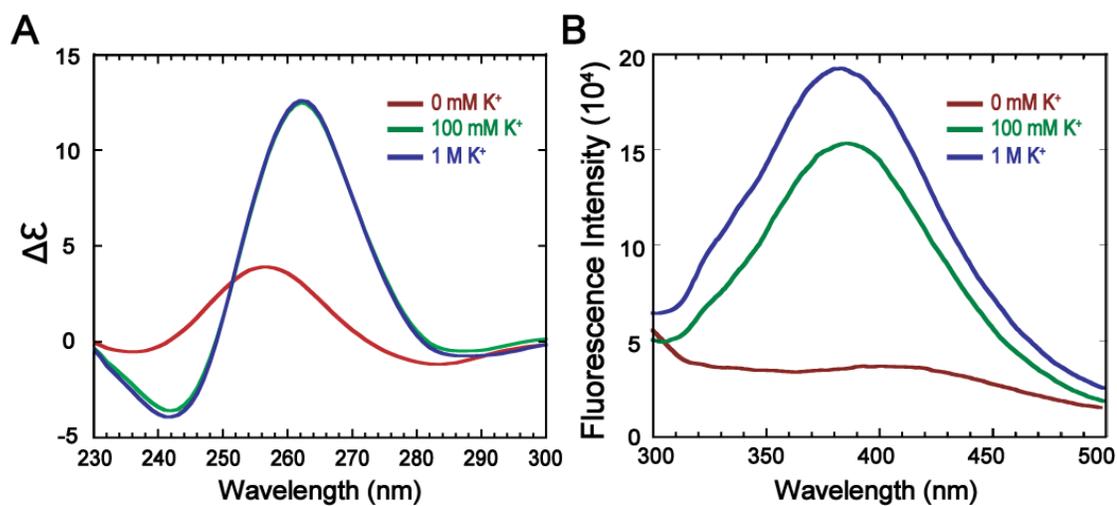


Figure B.7 K<sup>+</sup> Titration Result of dG3A using CD and Fluorescence Spectroscopy.

(A) K<sup>+</sup> titration spectra detected by CD. (B) K<sup>+</sup> titration spectra detected by fluorescence.

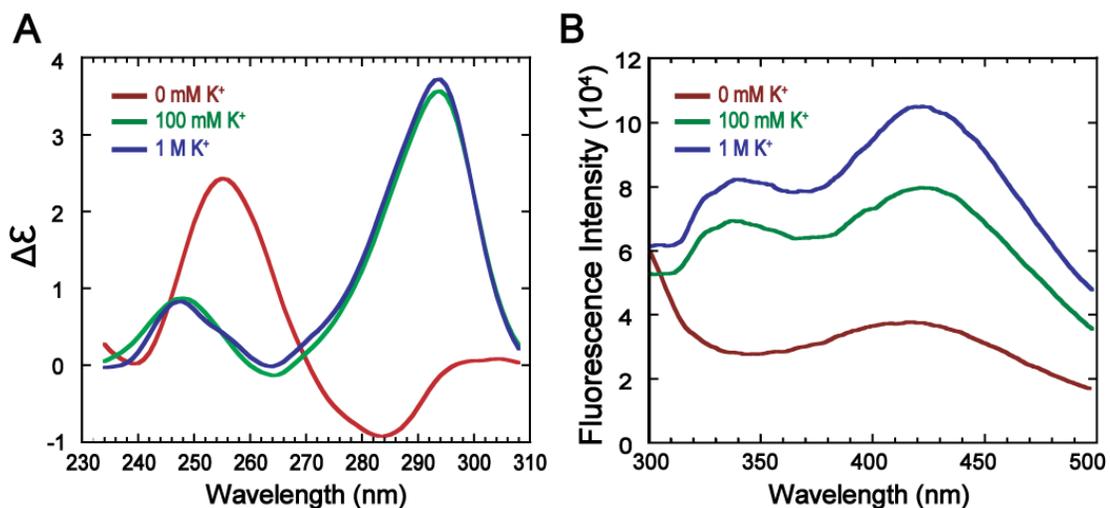


Figure B.8 K<sup>+</sup> Titration Result of dG3A2 using CD and Fluorescence Spectroscopy.

(A) K<sup>+</sup> titration spectra detected by CD. (B) K<sup>+</sup> titration spectra detected by fluorescence.

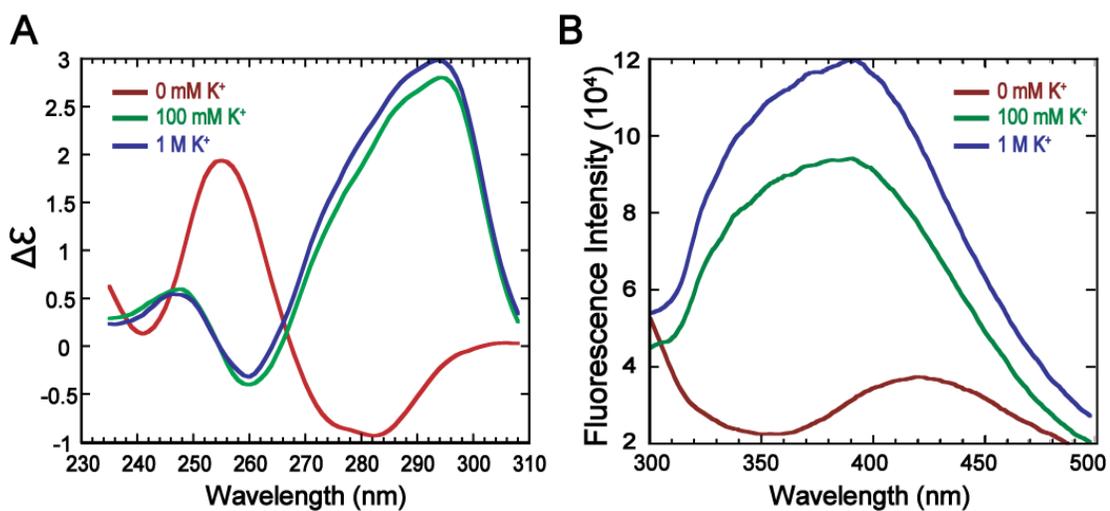


Figure B.9 K<sup>+</sup> Titration Result of dG3A3 using CD and Fluorescence Spectroscopy.

(A) K<sup>+</sup> titration spectra detected by CD. (B) K<sup>+</sup> titration spectra detected by fluorescence.

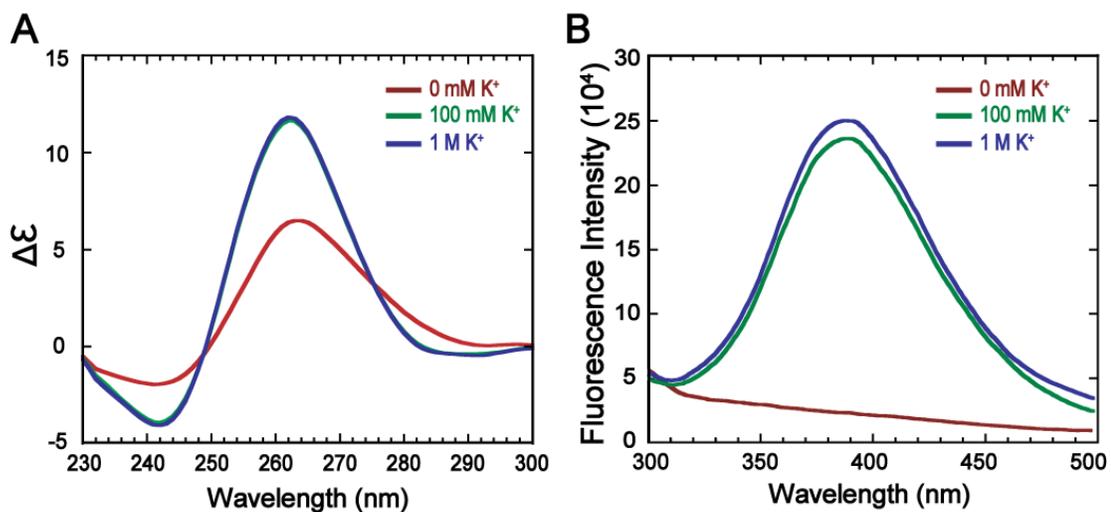


Figure B.10  $K^+$  Titration Result of dG3T using CD and Fluorescence Spectroscopy.

(A)  $K^+$  titration spectra detected by CD. (B)  $K^+$  titration spectra detected by fluorescence.

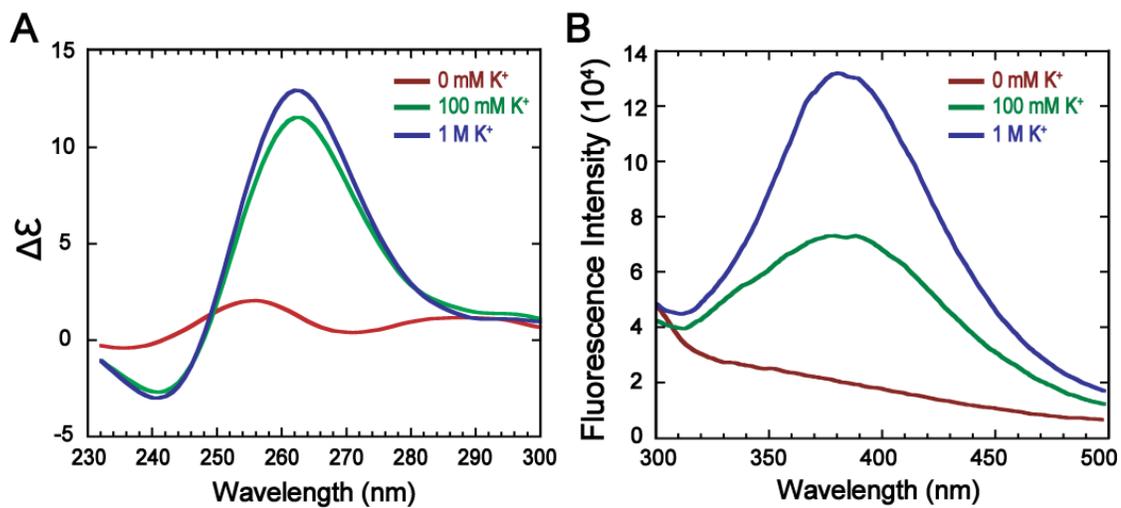


Figure B.11  $K^+$  Titration Result of dG3T2 using CD and Fluorescence Spectroscopy.

(A)  $K^+$  titration spectra detected by CD. (B)  $K^+$  titration spectra detected by fluorescence.

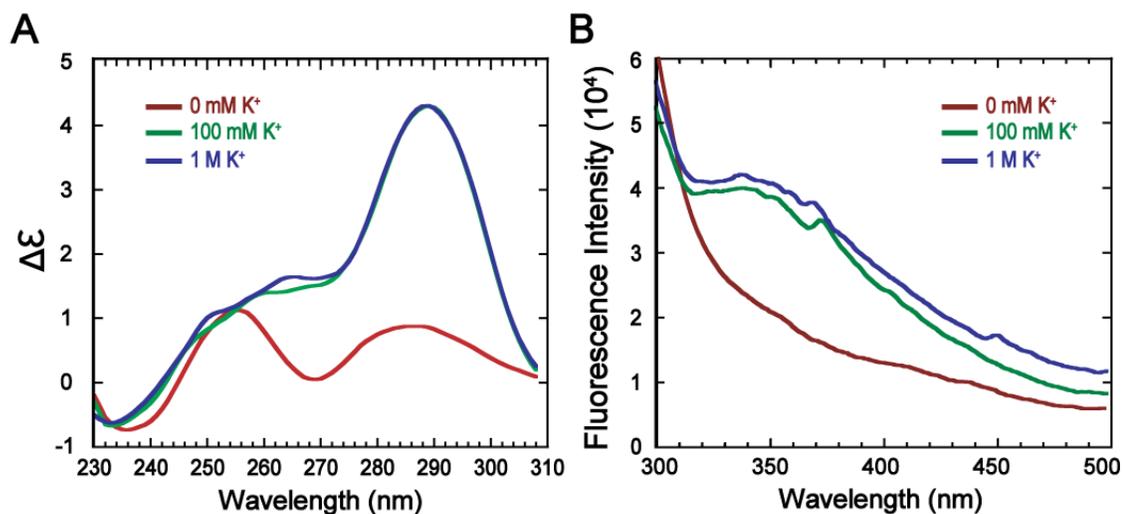


Figure B.12 K<sup>+</sup> Titration Result of dG3T3 using CD and Fluorescence Spectroscopy.

(A) K<sup>+</sup> titration spectra detected by CD. (B) K<sup>+</sup> titration spectra detected by fluorescence.

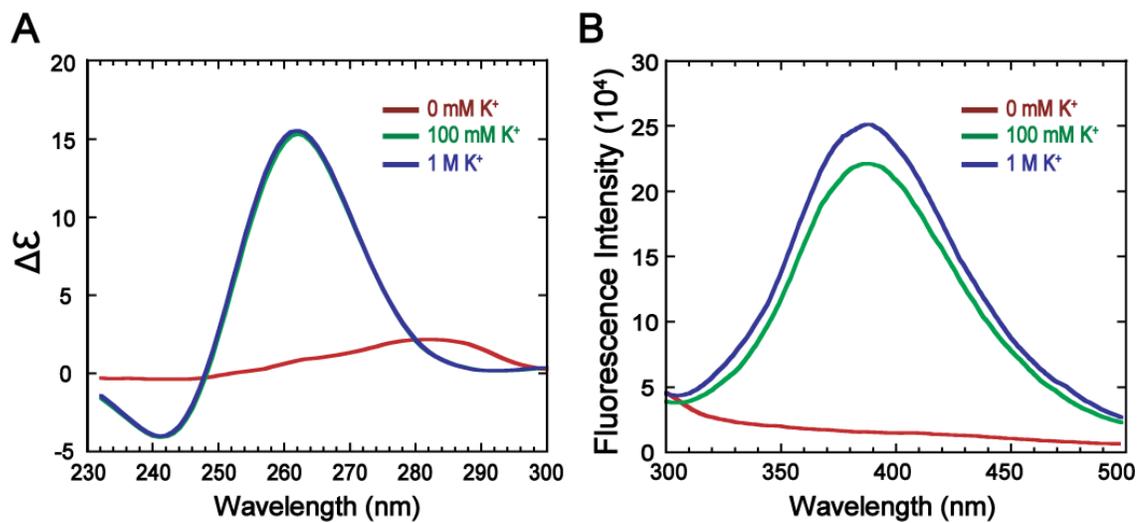
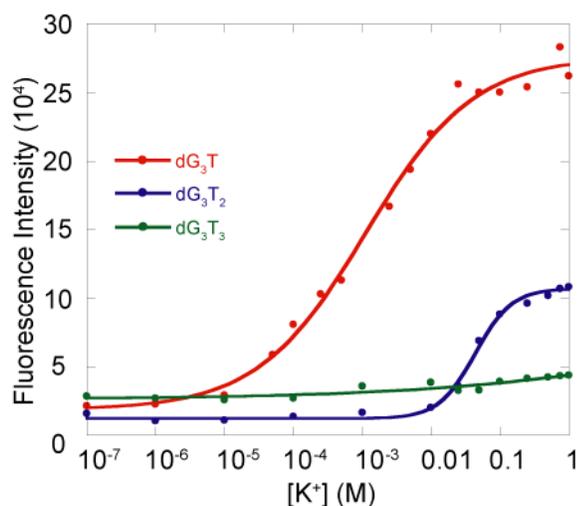


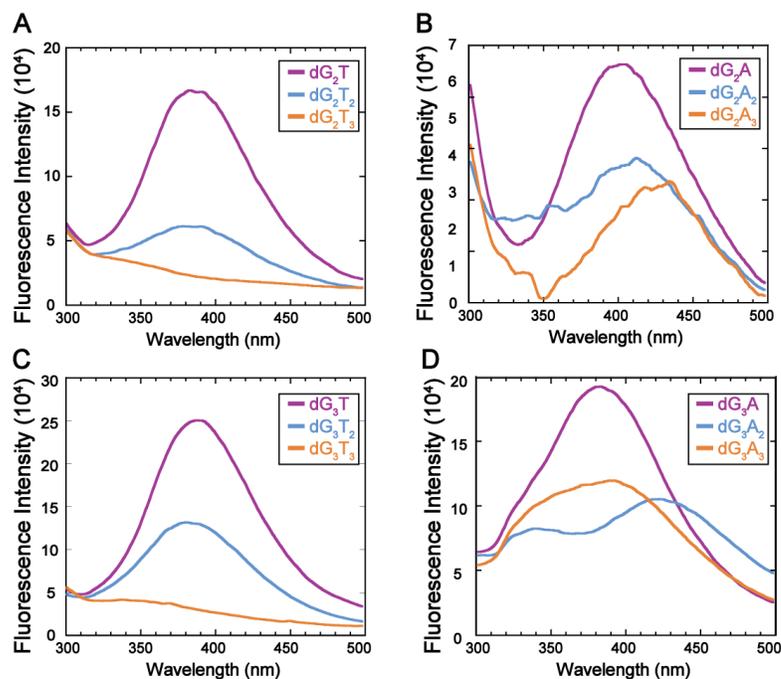
Figure B.13 K<sup>+</sup> Titration Result of dG3C using CD and Fluorescence Spectroscopy.

(A) K<sup>+</sup> titration spectra detected by CD. (B) K<sup>+</sup> titration spectra detected by fluorescence.



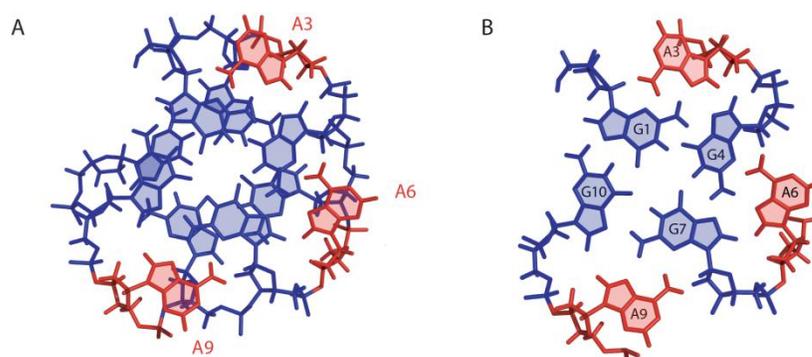
**FigureB.14** Overlay of Fluorescence Intensity versus  $K^+$  Concentration for dG3T, dG3T2, and dG3T3.

The fluorescence intensity of each GQS candidates was collected at the corresponding  $\lambda_{em}$  maximum as provided in Table 6.1. The presence of a plateau observed in fluorescence intensity at higher  $K^+$  concentrations for dG<sub>3</sub>T and dG<sub>3</sub>T<sub>2</sub> sequences clearly indicates that these are in their saturated folded state (i.e. 100% quadruplex) when compared at 1 M  $K^+$ . Although the above fluorescence data is too weak to allow such an assessment for dG<sub>3</sub>T<sub>3</sub>, analysis of CD data collected at 100 mM and 1 M  $K^+$  (Figure B.12) indicates that dG<sub>3</sub>T<sub>3</sub> too is fully in the quadruplex state at 1 M  $K^+$ .



**FigureB.15** Overlay of Fluorescence Spectra of GQS at 1M  $K^+$ .

(A) Fluorescence spectra of dG<sub>2</sub>T<sub>1-3</sub>. (B) Fluorescence spectra of dG<sub>2</sub>A<sub>1-3</sub>. (C) Fluorescence spectra of dG<sub>3</sub>T<sub>1-3</sub>. (D) Fluorescence spectra of dG<sub>3</sub>A<sub>1-3</sub>.



**Figure B.16 NMR Structure of d(GGA)4.**

(A) Overall structure viewed along the dyad axis. The A loops in red (A3, A6, A9) are intra-helical and interact with the G-quartets in blue. The 3'terminal A (A12) is not shown. (B) G's and A's in the heptad. The structure was obtained from PDB ID 1MYQ<sup>1</sup>.

## References

- 1 Matsugami, A. et al. An intramolecular quadruplex of (GGA)4 triplet repeat DNA with a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad, and its dimeric interaction. *J. Mol. Biol.* **313**, 255-269 (2001).

## Appendix C

## Supplementary Information for Chapter 4

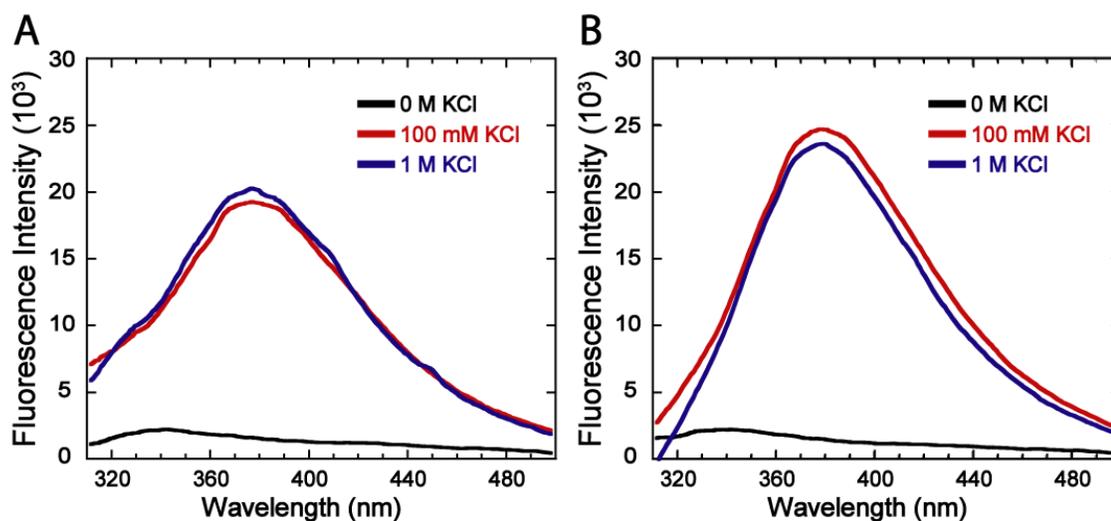


Figure C.1 Representative K<sup>+</sup> Titration of G3T and G3T 5'T using Fluorescence Spectroscopy.

(A) G3T. (B) G3T 5'T

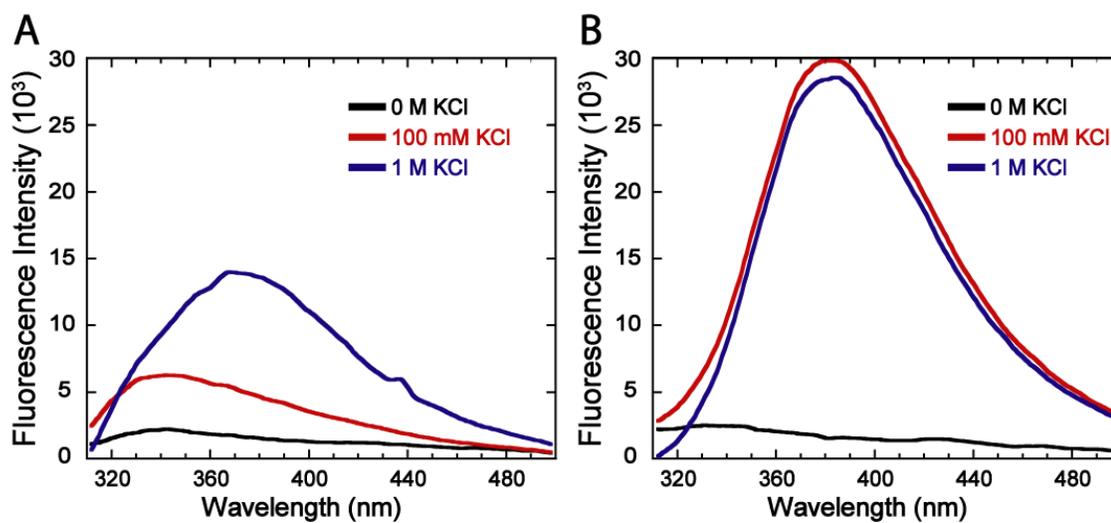
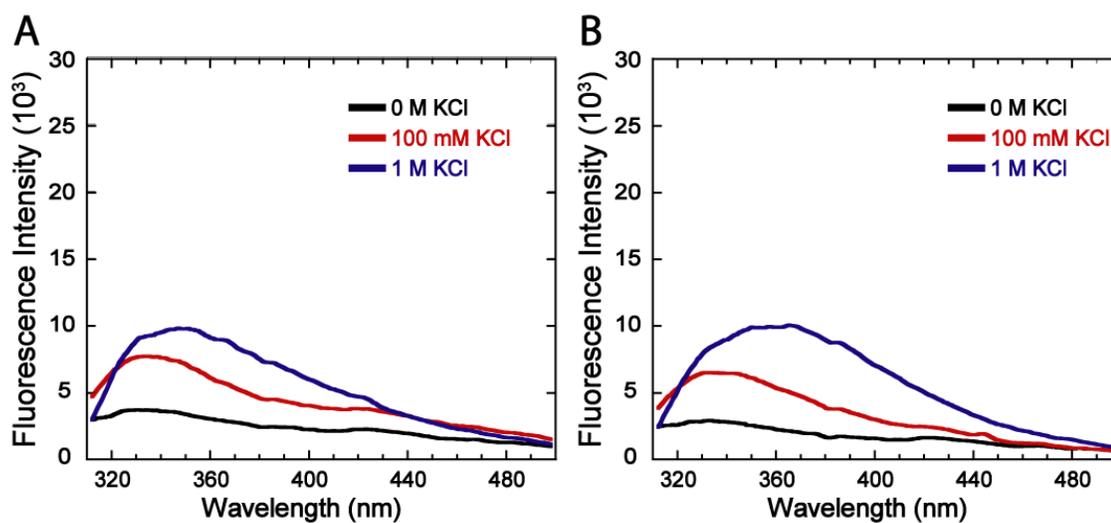


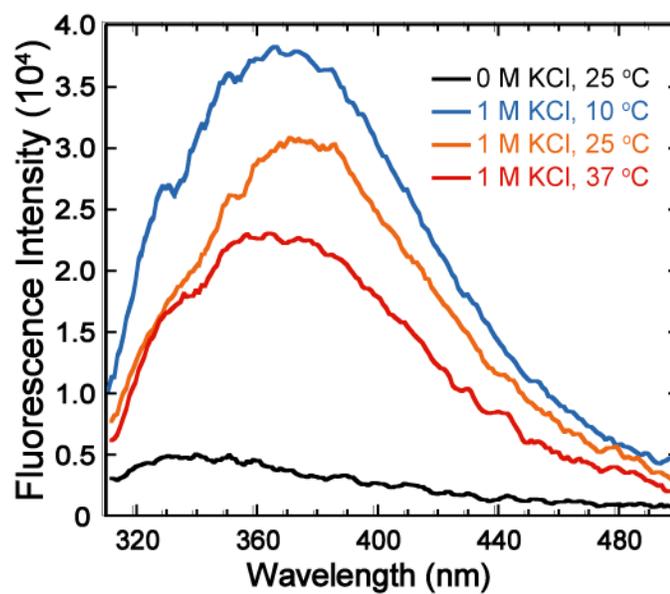
Figure C.2 Representative K<sup>+</sup> Titration of G3T and G3T 5'T using Fluorescence Spectroscopy.

(A) G3T 5'TT. (B) G3T 3'TT



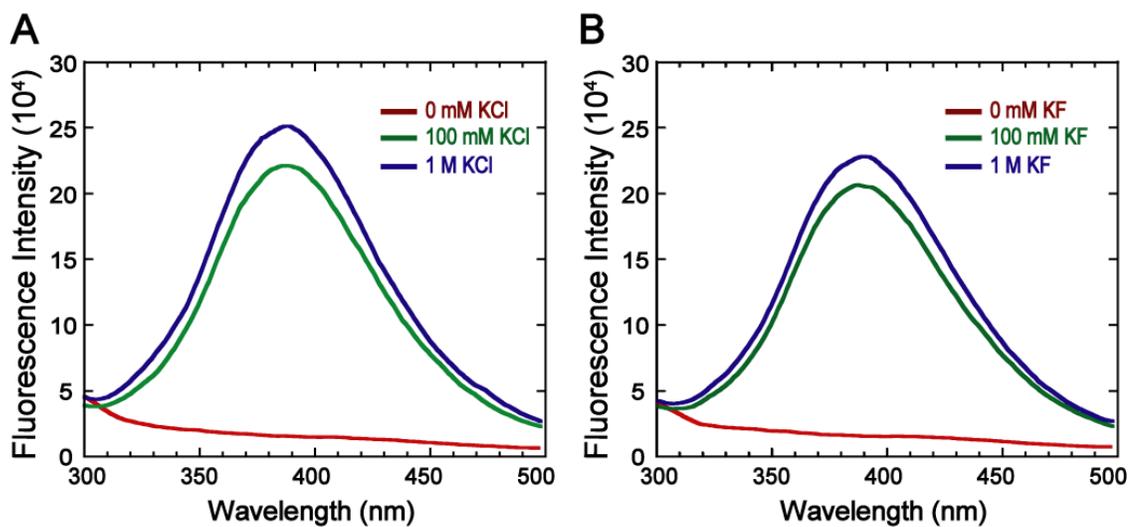
**Figure C.3 Representative K<sup>+</sup> Titration of G3T and G3T 5'T using Fluorescence Spectroscopy.**

(A) G3T 5'AA. (B) G3T 5'CC



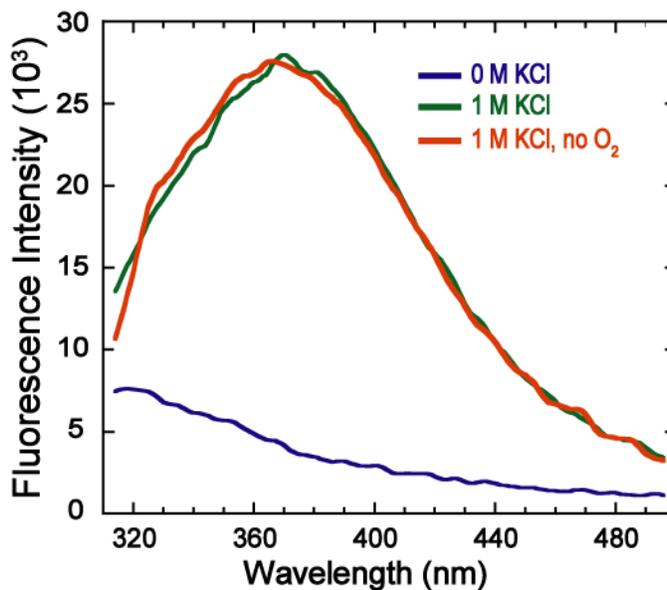
**Figure C.4 Temperature Dependence of Fluorescence Intensity.**

An overlay of fluorescence emission traces of the G3T sequence in 1 M KCl at increasing temperature.



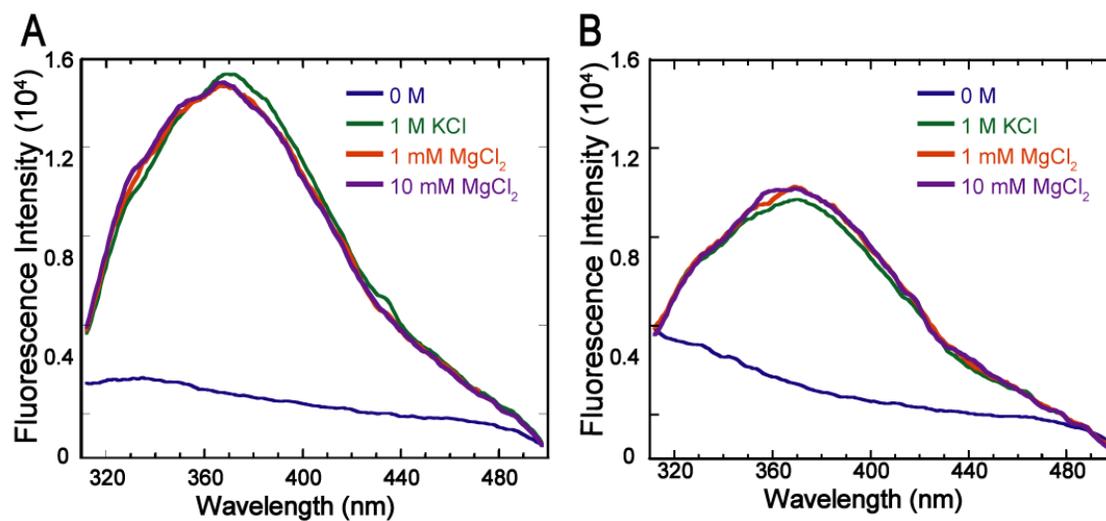
**Figure C.5 Counter-anion Effects on Fluorescence Emission Intensity for the G3T Sequence.**

(A) Representative  $K^+$  titration using KCl salt. (B) Representative  $K^+$  titration using KF salt.



**Figure C.6 Quenching Effect of Molecular Oxygen.**

An overlay of fluorescence emission traces of the G3T sequence in 1 M KCl before and after degassing to remove  $O_2$ .

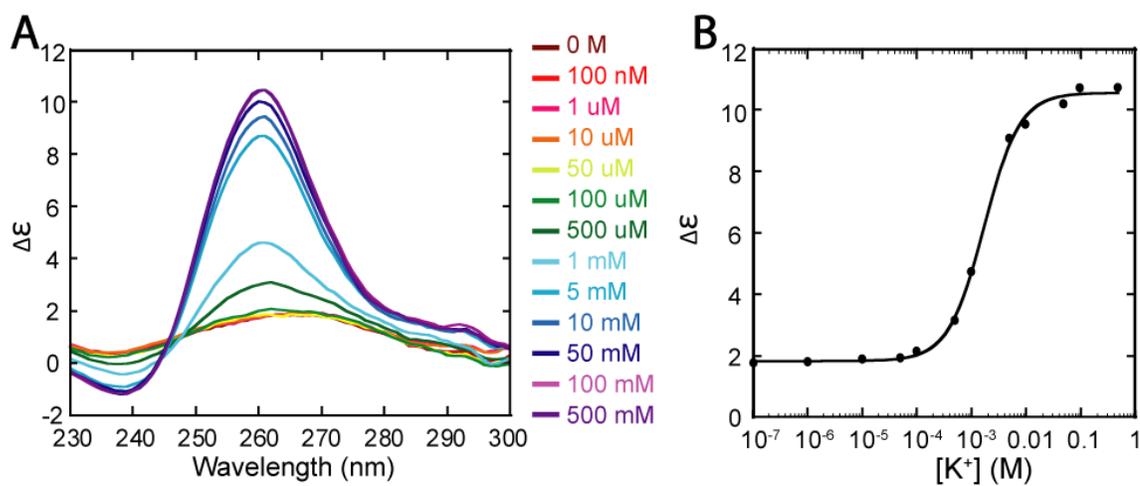


**Figure C.7 Mg<sup>2+</sup> Effects on Fluorescence Emission Intensity.**

An overlay of traces at 1 M KCl, 1 mM MgCl<sub>2</sub> in a 1M KCl background, and 10 mM MgCl<sub>2</sub> in a 1M KCl background. (A) Emission Spectrum for the G3T sequence. (B) Emission Spectrum for the G3T 5'TT sequence.

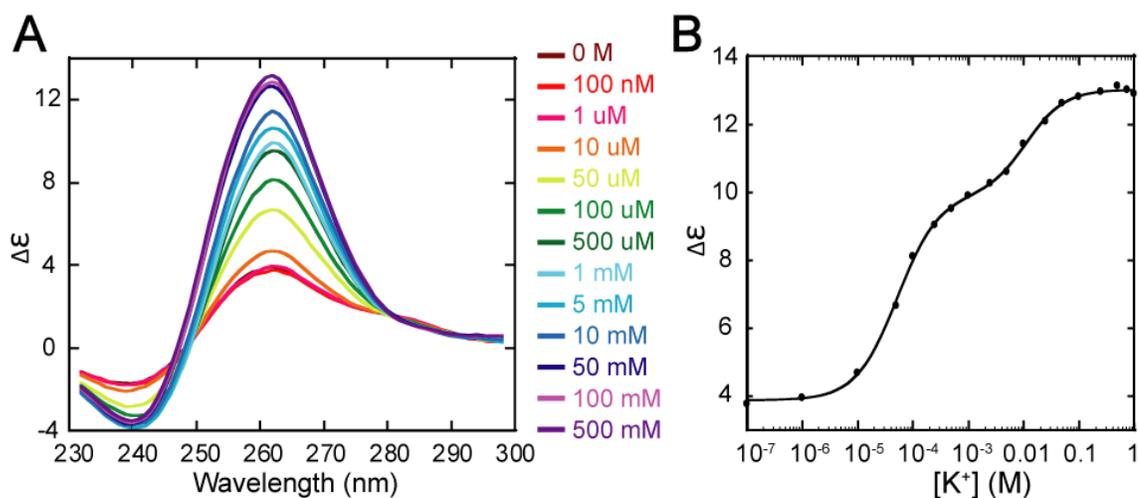
## Appendix D

## Supplementary Information for Chapter 5



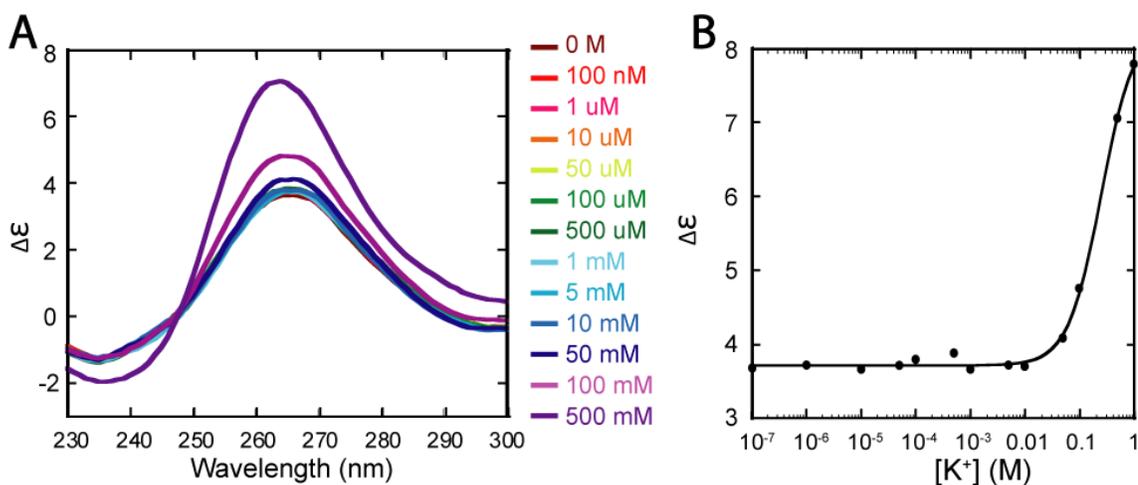
**Figure D.1 GGT3 QGS CD  $K^+$  Titration and  $\lambda_{max}$  Curve Plot.**

CD titrations of the GGT3 sequence at a RNA concentration of  $5\mu\text{M}$ . (A) Plot of normalized ellipticity versus wavelength, where each trace represents a different  $K^+$  concentration. From this data, the  $\lambda_{max}$  was determined to be 262nm and the data points at that wavelength were extracted. (B) Plot of ellipticity vs.  $K^+$  concentration at the  $\lambda_{max}$ , which appears as a two-state titration curve that was fit with Eq. 5.1. The  $K^+_{1/2}$  and n values are provided in Table 5.2.



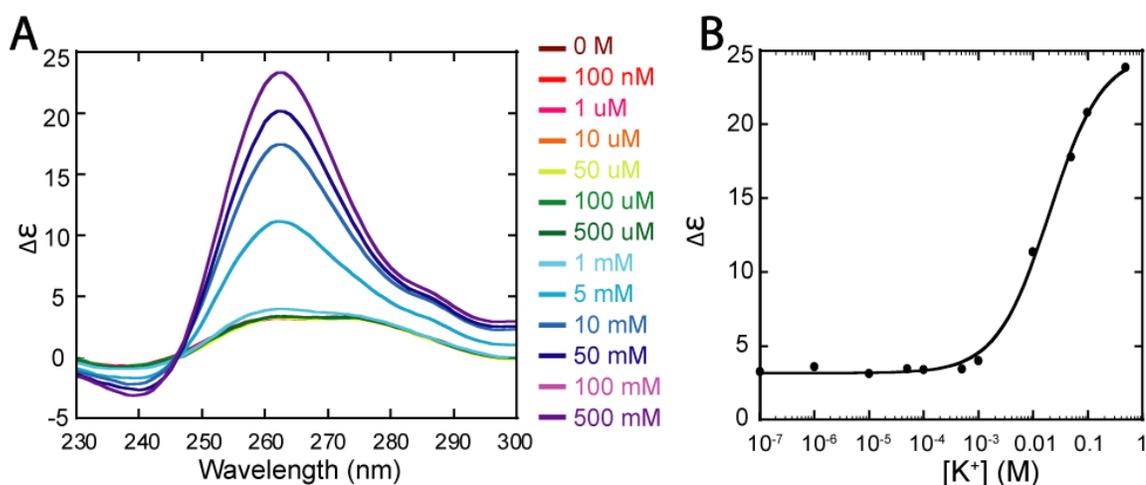
**Figure D.2 F19G14 GQS CD  $K^+$  Titration and  $\lambda_{\max}$  Curve Plot.**

CD titrations of the F19G14 sequence at a RNA concentration of  $5\mu\text{M}$ . (A) Plot of normalized ellipticity versus wavelength, where each trace represents a different  $K^+$  concentration. From this data, the  $\lambda_{\max}$  was determined to be 262nm and the data points at that wavelength were extracted. (B) Plot of ellipticity vs.  $K^+$  concentration at the  $\lambda_{\max}$ , which appears as a two-state titration curve that was fit with Eq. 5.1. The  $K^+_{1/2}$  and  $n$  values are provided in Table 5.2.



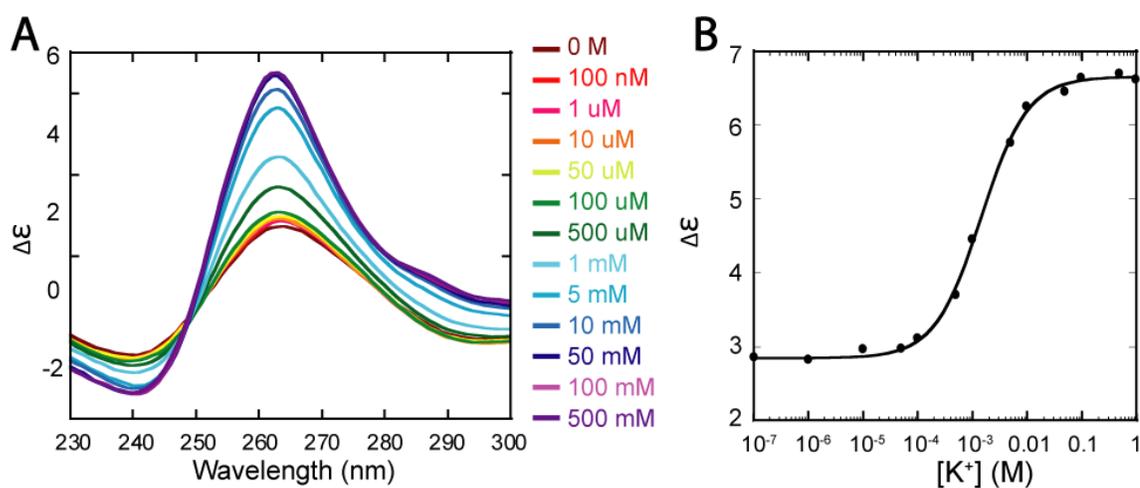
**Figure D.3 NRAMP6 GQS CD  $K^+$  Titration and  $\lambda_{\max}$  Curve Plot.**

CD titrations of the NRAMP6 sequence at a RNA concentration of  $5\mu\text{M}$ . (A) Plot of normalized ellipticity versus wavelength, where each trace represents a different  $K^+$  concentration. From this data, the  $\lambda_{\max}$  was determined to be 262nm and the data points at that wavelength were extracted. (B) Plot of ellipticity vs.  $K^+$  concentration at the  $\lambda_{\max}$ , which appears as a two-state titration curve that was fit with Eq. 5.1. The  $K^+_{1/2}$  and  $n$  values are provided in Table 5.2.



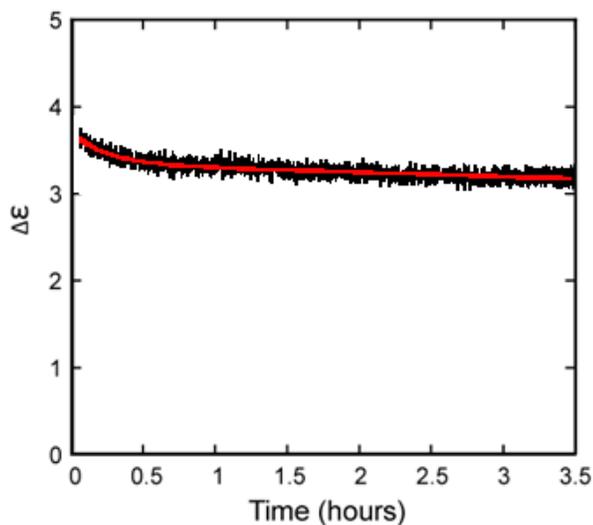
**Figure D.4 HMA5 GQS CD  $K^+$  Titration and  $\lambda_{max}$  Curve Plot.**

CD titrations of the GGT3 sequence at a RNA concentration of 5  $\mu$ M. (A) Plot of normalized ellipticity versus wavelength, where each trace represents a different  $K^+$  concentration. From this data, the  $\lambda_{max}$  was determined to be 262nm and the data points at that wavelength were extracted. (B) Plot of ellipticity vs.  $K^+$  concentration at the  $\lambda_{max}$ , which appears as a two-state titration curve that was fit with Eq. 5.1. The  $K^+_{1/2}$  and  $n$  values are provided in Table 5.2.



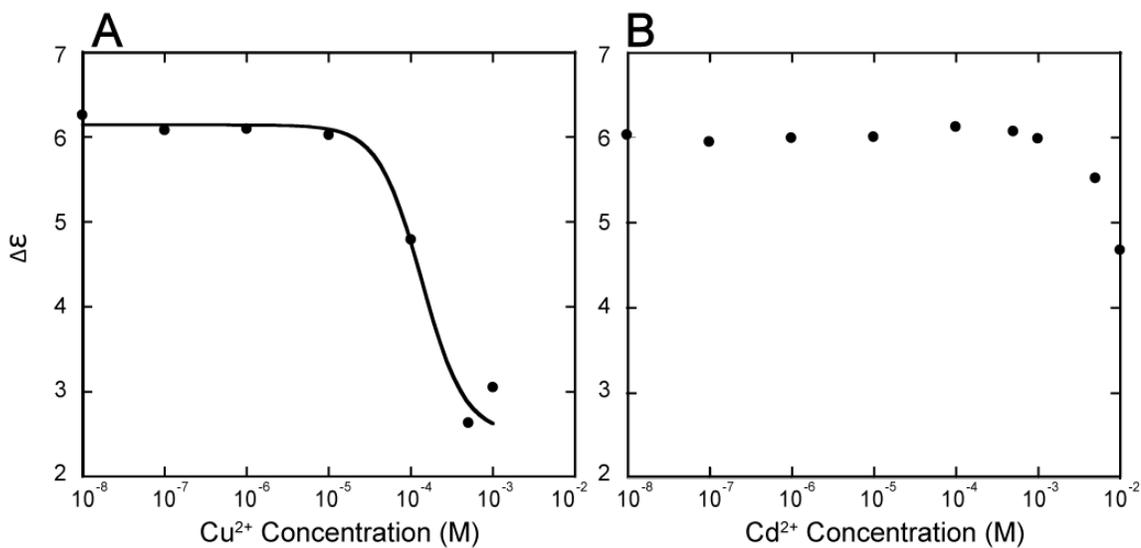
**Figure D.5 COBL10 GQS CD  $K^+$  Titration and  $\lambda_{max}$  Curve Plot.**

CD titrations of the COBL10 sequence at a RNA concentration of 5  $\mu$ M. (A) Plot of normalized ellipticity versus wavelength, where each trace represents a different  $K^+$  concentration. From this data, the  $\lambda_{max}$  was determined to be 262nm and the data points at that wavelength were extracted. (B) Plot of ellipticity vs.  $K^+$  concentration at the  $\lambda_{max}$ , which appears as a two-state titration curve that was fit with Eq. 5.1. The  $K^+_{1/2}$  and  $n$  values are provided in Table 5.2.



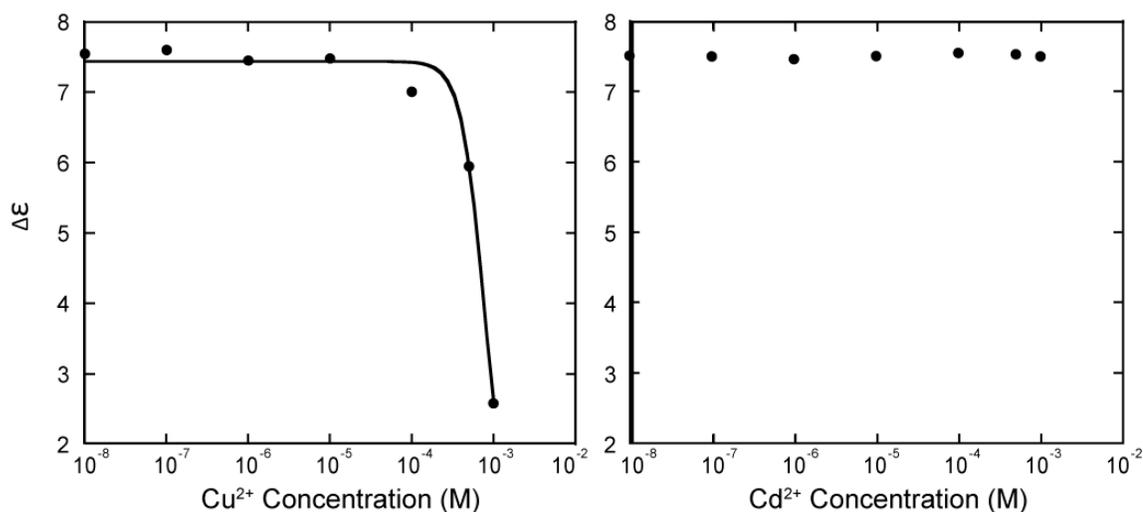
**Figure D.6 NRAMP6 GQS CD-Monitored  $\text{Cd}^{2+}$  Slow Kinetics.**

Normalized ellipticity measured for longer timepoints of the NRAMP6 sequence at a RNA concentration of  $5\mu\text{M}$ . The concentration of  $\text{Cd}^{2+}$  ions was  $100\mu\text{M}$  in a background of  $10\text{mM}$  LiCac and  $100\text{mM}$  KCl. The trace was fit with Eq. 5.3 and the exponential decay constant was determined to be 3.8.



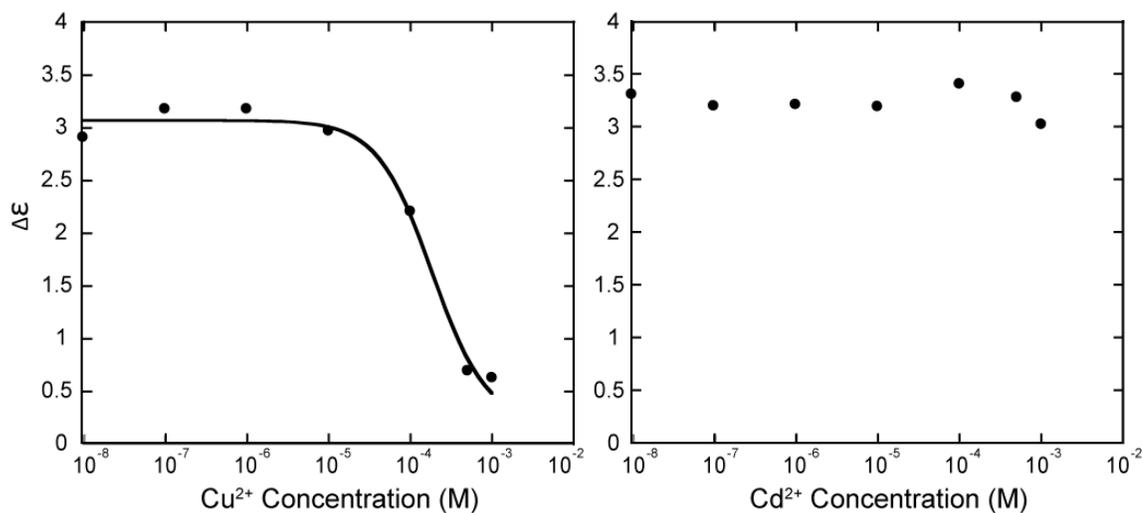
**Figure D.7 GGT3 GQS CD  $\text{Cu}^{2+}$  and  $\text{Cd}^{2+}$  Titration Curve Plot.**

CD titrations of the GGT3 sequence at a RNA concentration of  $5\mu\text{M}$ . (A) Plot of ellipticity vs.  $\text{Cu}^{2+}$  concentration at the  $\lambda_{\text{max}}$  appears as a two-state titration curve that was fit with Eq. 5.1. The  $\text{Cu}^{2+}_{1/2}$  and  $n$  values are provided in Table 5.3. (B) Plot of ellipticity vs.  $\text{Cd}^{2+}$  concentration at the  $\lambda_{\text{max}}$ .



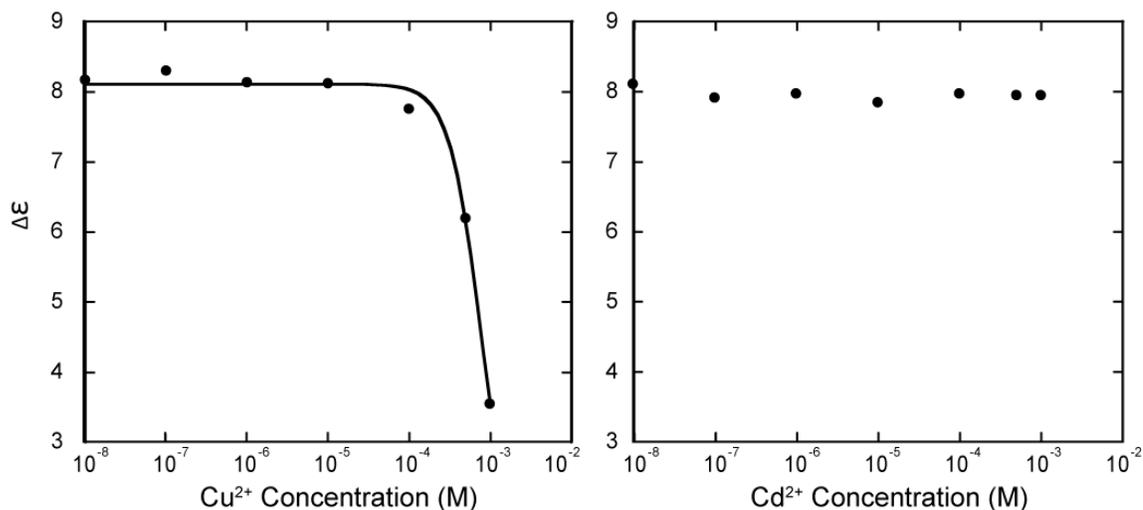
**Figure D.8 19G14 GQS CD  $\text{Cu}^{2+}$  and  $\text{Cd}^{2+}$  Titration Curve Plot.**

CD titrations of the F19G14 sequence at a RNA concentration of  $5\mu\text{M}$ . Plot of ellipticity vs.  $\text{Cu}^{2+}$  concentration at the  $\lambda_{\text{max}}$  appears as a two-state titration curve that was fit with Eq. 5.1. The  $\text{Cu}^{2+}_{1/2}$  and  $n$  values are provided in Table 5.3. The lower baseline was fit using the  $\text{K}^{+}$  titration data in Figure D.2. (B) Plot of ellipticity vs.  $\text{Cd}^{2+}$  concentration at the  $\lambda_{\text{max}}$ .



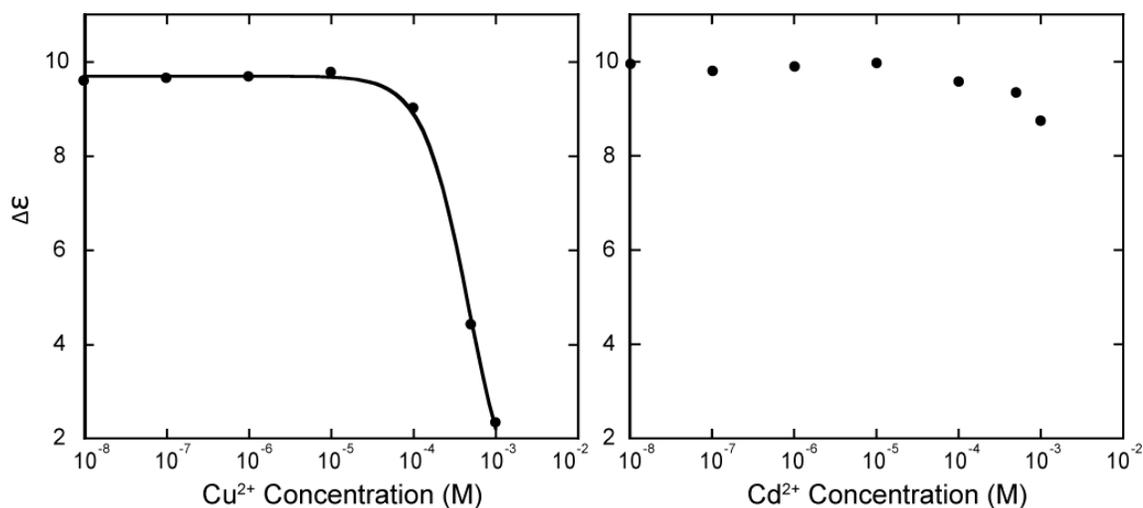
**Figure D.9 NRAMP6 GQS CD  $\text{Cu}^{2+}$  and  $\text{Cd}^{2+}$  Titration Curve Plot.**

CD titrations of the NRAMP6 sequence at a RNA concentration of  $5\mu\text{M}$ . Plot of ellipticity vs.  $\text{Cu}^{2+}$  concentration at the  $\lambda_{\text{max}}$  appears as a two-state titration curve that was fit with Eq. 5.1. The  $\text{Cu}^{2+}_{1/2}$  and  $n$  values are provided in Table 5.3. (B) Plot of ellipticity vs.  $\text{Cd}^{2+}$  concentration at the  $\lambda_{\text{max}}$ .



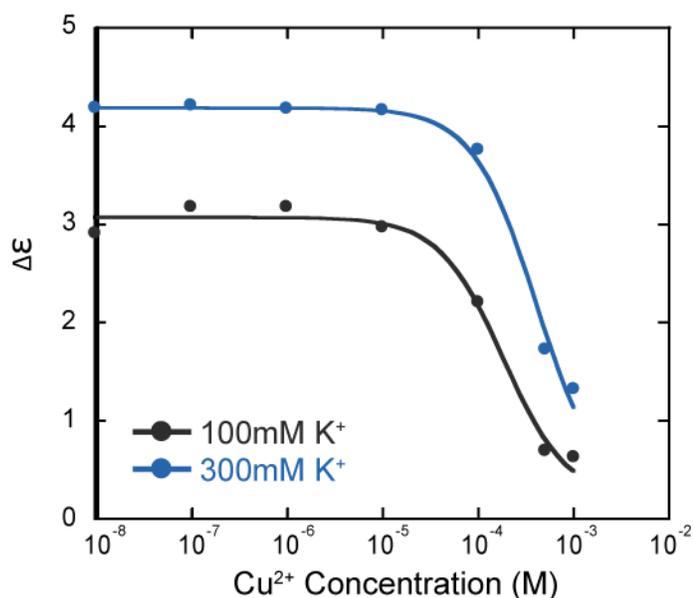
**Figure D.10 HMA5 GQS CD  $\text{Cu}^{2+}$  and  $\text{Cd}^{2+}$  Titration Curve Plot.**

CD titrations of the HMA5 sequence at a RNA concentration of  $5\mu\text{M}$ . Plot of ellipticity vs.  $\text{Cu}^{2+}$  concentration at the  $\lambda_{\text{max}}$  appears as a two-state titration curve that was fit with Eq. 5.1. The  $\text{Cu}^{2+}_{1/2}$  and  $n$  values are provided in Table 5.3. The lower baseline was fit using the  $\text{K}^+$  titration data in Figure D.4. (B) Plot of ellipticity vs.  $\text{Cd}^{2+}$  concentration at the  $\lambda_{\text{max}}$ .



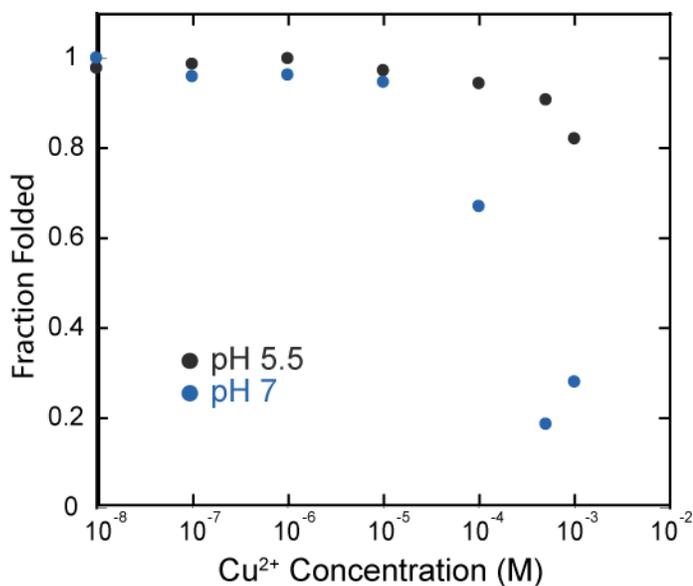
**Figure D.11 COBL10 GQS CD  $\text{Cu}^{2+}$  and  $\text{Cd}^{2+}$  Titration Curve Plot.**

CD titrations of the COBL10 sequence at a RNA concentration of  $5\mu\text{M}$ . Plot of ellipticity vs.  $\text{Cu}^{2+}$  concentration at the  $\lambda_{\text{max}}$  appears as a two-state titration curve that was fit with Eq. 5.1. The  $\text{Cu}^{2+}_{1/2}$  and  $n$  values are provided in Table 5.3. The lower baseline was fit using the  $\text{K}^+$  titration data in Figure D.5. (B) Plot of ellipticity vs.  $\text{Cd}^{2+}$  concentration at the  $\lambda_{\text{max}}$ .



**Figure D.12 NRAMP GQS CD Cu<sup>2+</sup> Titration Curve Plot with Varied K<sup>+</sup>.**

CD titrations of the NRAMP6 sequence at a RNA concentration of 5 $\mu$ M. Ellipticity in either a 100mM or 300mM KCl background is plotted vs. Cu<sup>2+</sup> concentration at the  $\lambda_{\max}$ , which appears as a two-state titration curve that was fit with Eq. 5.1. The Cu<sup>2+</sup><sub>1/2</sub> and n values are provided in Table 5.3.



**Figure D.13 GGT3 GQS CD Cu<sup>2+</sup> Titration Curve Plot at Varied pH.**

CD titrations of the GGT3 sequence at a RNA concentration of 5 $\mu$ M. Ellipticity at either pH 5.5 or pH 7 is plotted vs. Cu<sup>2+</sup> concentration at the  $\lambda_{\max}$  and appears as a two-state titration curve that was fit with Eq. 5.1. Cu<sup>2+</sup><sub>1/2</sub> values are reported in Table 5.3.

## ACADEMIC VITA

Madeline E. Sherlock  
3884 Buffalo Run Road, Port Matilda, PA 16870 --- mes5567@psu.edu

---

### Education

The Pennsylvania State University, University Park, PA

Chemistry B.S., May 2014; German B.S., May 2014

Honors in Chemistry

Minor: International Studies

### Research and Professional Experience

PSU Chemistry Department Teaching Assistant:

- Fall 2013: Honors chemistry course (Chem 110H) for Dr. P. Bevilacqua
- Spring 2014: Fundamentals in Organic Chemistry Lab (Chem 203) for Dr. J. Bortiatynski
- Tutor: undergraduate chemistry courses and labs

Bevilacqua Chemistry Lab (2011-2014) - Chemistry Department, Penn State University

- Individual project involving RNA and DNA folding using Circular Dichroism, UV-vis, and fluorescence spectroscopy methods

Meggers Chemistry Lab(2012-2013) – Chemistry Department, Philipps-Universität Marburg

- One-year visiting research opportunity during exchange program
- Project: Synthesis of a ruthenium-based kinase inhibitor to be used for in vivo testing

Summer Experience in the Eberly College of Science (SEECoS) (2011-2012)

- Research mentor: supervised 3 high school students to conduct experiments in the lab
- The 6-week program (SEECoS) is part of the Upward Bound Initiative at Penn State

Ordway Neuroscience Lab(2010) - Life Sciences Department, Penn State University

- Group project studying neurological diseases of *Drosophila melanogaster*
- Used microbiology techniques – PCR, Gel Electrophoresis, DNA digestion

### **Awards and Activities**

- DAAD Undergraduate Scholarship (2012-2013) – Study abroad scholarship in Germany
- Teas Scholarship (2011-present) – Presented by the Penn State Chemistry Department
- PSU Undergraduate Discovery Summer Grant recipient (2011)
- Penn State Club Softball Team (2010-2012) – Penn State Dance Marathon chairperson

### **Conferences and Presentations**

- Presented a research poster “Structural Studies of G-Quadruplexes and their Intrinsic Fluorescence” at the 2013 Rustbelt RNA Meeting in Cleveland, OH
- Attended the 63rd Lindau Nobel Laureate Meeting in Chemistry, July 2013
- Presented a research poster “Heavy Metal Effects on RNA G-Quadruplex Formation in *Arabidopsis thaliana*” at the 2011 local ACS Meeting at Penn State

### **Publications**

Kwok, C.K.\*, Sherlock, M.E.\*, Bevilacqua, P.C. "Decrease in RNA folding cooperativity by deliberate population of intermediates in RNA G-quadruplexes." *Angew. Chem. Int. Ed.* **2013**, *52*, 683-686.

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

Kwok, C.K., Ding, Y., Sherlock, M.E., Assmann, S.M., Bevilacqua, P.C. "A hybridization-based approach for quantitative and low-bias single-stranded DNA ligation." *Analytical Biochemistry* **2013**, *435*, 181-186.

\*These authors contributed equally to this work