RNA G-Quadruplex Structures: Functions in Noncoding RNAs and Disease

Emily Benner

Follow this and additional works at: https://dsc.duq.edu/etd

Part of the Biochemistry Commons, and the Biophysics Commons

Recommended Citation

This One-year Embargo is brought to you for free and open access by Duquesne Scholarship Collection. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Duquesne Scholarship Collection.
RNA G-QUADRUPLE STRUCTURES:
FUNCTIONS IN NONCODING RNAs AND DISEASE

A Dissertation
Submitted to the Bayer School of Natural and Environmental Sciences

Duquesne University

In partial fulfillment of the requirements for
the degree of Doctor of Philosophy

By
Emily M. Benner

May 2021
Paraspeckles are a relatively newly identified subnuclear body whose function is still largely unknown. They are membraneless organelles found within mammalian nuclei originally identified as being comprised of different paraspeckle proteins (PSPs), and later proven to also contain RNA. This study investigates the long non-coding RNA nuclear enriched abundant transcript 1 (NEAT1), which has been found to be indispensable in the formation of paraspeckles, with each containing up to 50 NEAT1 molecules. NEAT1 has two isoforms, the shorter NEAT1_1, and the longer NEAT1_2. The focus of this work will be on the NEAT1_2, which acts as a scaffold in the formation of the paraspeckles. Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disorder affecting motor neurons and several studies indicate NEAT1 upregulation in ALS, presumably due to an increase in paraspeckle formation in motor neurons of ALS patients.
We identified five regions within NEAT1_2 that have the potential to form secondary structure G quadruplexes (GQs), with one of the GQs being localized 20 nucleotides away from a triple helix (TH) secondary structure previously identified at its 3’ terminus, formed by the binding of an adenine (A)-rich tail to two uracil (U)-rich motifs. The proximity of the TH and GQ at the 3’ end of the RNA leads to the hypothesis that these two structures will affect one another, potentially affecting the stability of NEAT1_2.

Here, we used various biophysical techniques to characterize the GQ structure present near the 3’ end of NEAT1_2 and determine how the presence or absence of the GQ affects the formation of the terminal TH. We and others have shown that the potential GQ region nearest the TH terminus, “GQ5” does indeed form a GQ. We have also shown that this GQ is parallel and intramolecular. Various binding studies reveal that when the GQ is unable to form, the TH is also unable to form, supporting our hypothesis that the formation of the TH is dependent upon the GQ structure. Understanding the secondary structure of NEAT1_2 can help in the overall understanding of paraspeckle structure and function, how NEAT1 and paraspeckles may work together in the progression of some diseases, and lead to potential treatment of neurodegenerative diseases like ALS.
DEDICATION

I would like to dedicate this dissertation to the late Dr. Jeffry D. Madura. Jeff was my original advisor at Duquesne, and he was a great advisor, mentor, and friend. He was the department dad, “dad-dura”, and he was one of the smartest, kindest, and warmest people I have ever had the pleasure of knowing. Without Jeff, there is no way I would be where I am now. I will never take anything I learned from Jeff for granted, but if I had to take any one thing away from my time with him, it would be the way he treated people. Jeff had such a kind and open heart, and his door was always open to talk about research, life, or anything else you wanted to discuss at that moment. We had plenty of hockey talks, plenty of scientific discussions, and when you needed a suggestion for a good place to eat, Jeff was the person to go to. He was taken too soon and although I didn’t finish out the project he and I started, I know he would be extremely proud and happy to see that I finished my PhD in Dr. Rita’s lab. So, thank you, Jeff, for pushing me to do what I never thought I could, for teaching me to keep going even when I felt stuck, and for providing me with the best lab environment I could have had for five years. This one’s for you.
ACKNOWLEDGEMENT

They say it takes a village, and earning a PhD is no exception to that. I would not have been able to accomplish what I have without the help of so many people. I would like to start by acknowledging my advisor, Dr. Mihaela-Rita Mihăilescu, for her continued and unwavering support throughout my time at Duquesne. Dr. Rita was always there to encourage me and to help me through times when I felt stressed. Even when I did not believe in myself or in my ability to finish this degree, Dr. Rita always believed in me. Without her continued support, encouragement, and help, I truly don’t know if I would have completed this work. Affectionately referred to as the “mom” of the Chemistry and Biochemistry department, Dr. Rita is one of the kindest people I have met, and I hope to be as good of an advisor to others (in any aspect) throughout my life.

I’d like to thank my other committee members, Dr. Jeffrey Evanseck, Dr. Michael Cascio, and Dr. Danith Ly. Dr. Evanseck, thank you for always making me think outside the box – from asking questions in joint group meetings when I first started at Duquesne, to continually challenging me to perfect ways of presenting and tackling research questions. Thank you for taking me under your wing when Dr. Madura passed. That was an act of true kindness and friendship in such a difficult time. Dr. Cascio, thank you for always being there to lend a helping hand. You and your research group were always a fantastic resource for my work in the Madura Lab and beyond. Thank you for being the source of “adventure” on many biophysics trips – I don’t think any conference would have been the same without those adventures! Thank you, Dr. Ly for agreeing to be my external reader – even on short notice. I appreciate the time and effort you took to help me finalize my dissertation.
There are countless former and current lab members I would like to mention. Thank you all for helping me with research, always providing laughs, and being there for each other when we needed it the most. I’d especially like to thank the members of the Madura Lab – Dr. Riley Workman, Dr. Matt Srnec, Kendy Pellegrene-Guarinoni, Shiv Upadhyay, Amanda Dumi, and Dr. Mike Wasko. The time spent working with all of you and getting to know you all like family was a truly special time during my graduate career. Thank you to my fellow members of the Rita Lab – Dr. Josh Imperatore, Caylee Cunningham, and Caleb Frye. Thank you, Josh, for answering all of my questions and not getting too annoyed with me when I’d take up all your time asking you to explain things. Caylee, your expertise in Disney was the perfect distraction when it was often much needed. Let’s go on a Disney trip soon?! Caleb, you’ve been a great addition to the lab, and I know you’ll do great things at Duquesne. Also, I want to extend a special thank you to all of the undergraduate students who have assisted on my projects, who have helped me learn and become a better researcher, and who have also become great friends. It’s been a privilege watching you all grow and move on to bigger and better things.

I would like to thank Dan Bodnar, Lance Crosby, and Chris Lawson from the instrument shop. Thank you for always troubleshooting instrument problems and being very prompt in responding to requests and concerns. Your help was always greatly appreciated. I’d also like to extend a special thank you to Scott Boesch. Scott, you were instrumental in any and all computer issues I had during my time in the Madura Lab. Thanks for always being available to help, to listen, and to talk about science and life.

A very special thank you goes out to all of my family and friends. Thank you to my mom and Tom for putting up with me through all of this. I know I can be frustrating
and (frankly) annoying at times, but I know you’ve always got my back. Thanks for always believing in me, even when I wasn’t always the most pleasant. Gram, thanks for always checking in and supporting me. And I hope you’re happy that I “made it in 2020, ok?!” I know Pap would be saying the same! Thanks to all my friends, both in Pittsburgh and outside. I knew I always had people to count on when I needed a friend to talk to, complain to, and celebrate with. 2020 is certainly a unique year to wrap up graduate school, but it only seems fitting to end on an unconventional note, considering everything that’s happened the past several years. Thanks, again, to everyone who has been there for me. It means so much!
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td></td>
<td>iv</td>
</tr>
<tr>
<td>DEDICATION</td>
<td></td>
<td>vi</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENT</td>
<td></td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td></td>
<td>xii</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td></td>
<td>xiv</td>
</tr>
<tr>
<td>CHAPTER 1: Introduction</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1.1</td>
<td>Central Dogma of Biology and Non-Coding RNAs</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>Paraspeckles</td>
<td>5</td>
</tr>
<tr>
<td>1.3</td>
<td>Nuclear Enriched Abundant Transcript 1</td>
<td>8</td>
</tr>
<tr>
<td>1.4</td>
<td>G Quadruplexes</td>
<td>11</td>
</tr>
<tr>
<td>1.5</td>
<td>Fused in Sarcoma and Neurodegeneration</td>
<td>14</td>
</tr>
<tr>
<td>1.6</td>
<td>Motivation</td>
<td>16</td>
</tr>
<tr>
<td>CHAPTER 2: Materials and Methods</td>
<td></td>
<td>18</td>
</tr>
<tr>
<td>2.1</td>
<td>RNA Synthesis</td>
<td>18</td>
</tr>
<tr>
<td>2.2</td>
<td>Protein Synthesis</td>
<td>18</td>
</tr>
<tr>
<td>2.3</td>
<td>Polyacrylamide Gel Electrophoresis</td>
<td>18</td>
</tr>
<tr>
<td>2.4</td>
<td>Circular Dichroism Spectroscopy</td>
<td>20</td>
</tr>
<tr>
<td>2.5</td>
<td>$^1$H NMR Spectroscopy</td>
<td>21</td>
</tr>
<tr>
<td>2.6</td>
<td>Ultraviolet Spectroscopy Thermal Denaturation</td>
<td>21</td>
</tr>
<tr>
<td>CHAPTER 3: Results and Discussion</td>
<td></td>
<td>23</td>
</tr>
</tbody>
</table>
3.1 NEAT1_2 is proposed to form G-quadruplex structures in five guanine-rich regions .......................................................................................................................... 23
3.2 Biophysical Characterization of NEAT1 GQ5.......................................................... 24
3.3 Characterization of GQ and TH in the NEAT1 Triple Helix long sequence .......... 31
3.4 Determining the influence of GQ on TH formation ............................................... 37
3.5 Interaction of NEAT1_2 GQ5 with FUS................................................................... 46

CHAPTER 4: Conclusions and Future Work................................................................. 49

CHAPTER 5: References................................................................................................. 54

CHAPTER 6: Appendix.................................................................................................... 71
6.1 Introduction ............................................................................................................. 71
6.2 Materials and Methods .......................................................................................... 73
6.2.1 Cell Transformation for (G4C2)45....................................................................... 73
6.2.2 Agarose Gel Preparation .................................................................................. 73
6.2.3 Phenol Chloroform Extraction........................................................................... 74
6.3 Results and Discussion .......................................................................................... 74
6.3.1 Transcription of (G4C2)45 Expansion................................................................. 74
6.3.2 Determining G Quadruplex Formation in the (G4C2)45 Sequence................. 76
6.3.3 Binding of Peptide Nucleic Acid to the (G4C2)45 Sequence ......................... 78
6.4 Conclusions and Future Work............................................................................... 79
6.5 References ............................................................................................................. 80
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Central dogma of biology</td>
<td>1</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Roles of IncRNAs</td>
<td>4</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Structure of paraspeckles</td>
<td>7</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Two isoforms of NEAT1</td>
<td>9</td>
</tr>
<tr>
<td>Figure 5</td>
<td>MALAT1 Triple helix sequence and crystal structure</td>
<td>10</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Planar and 3D G quadruplex</td>
<td>12</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Topology of G quadruplexes</td>
<td>13</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Schematic representation of FUS</td>
<td>15</td>
</tr>
<tr>
<td>Figure 9</td>
<td>Sequence and predicted structure of 3’ end of NEAT1_2</td>
<td>17</td>
</tr>
<tr>
<td>Figure 10</td>
<td>KCl dependence PAGE for NEAT1 GQ5 WT</td>
<td>25</td>
</tr>
<tr>
<td>Figure 11</td>
<td>1H NMR spectra of NEAT1 GQ5 WT</td>
<td>26</td>
</tr>
<tr>
<td>Figure 12</td>
<td>CD spectra of NEAT1 GQ5 WT and TmPyP4</td>
<td>27</td>
</tr>
<tr>
<td>Figure 13</td>
<td>Structure of TmPyP4</td>
<td>28</td>
</tr>
<tr>
<td>Figure 14</td>
<td>UV spectra of NEAT1 GQ5 WT</td>
<td>29</td>
</tr>
<tr>
<td>Figure 15</td>
<td>KCl dependence PAGE for NEAT1 TH long</td>
<td>31</td>
</tr>
<tr>
<td>Figure 16</td>
<td>CD spectra for NEAT1 TH long</td>
<td>32</td>
</tr>
<tr>
<td>Figure 17</td>
<td>1H NMR spectra for NEAT1 TH long</td>
<td>34</td>
</tr>
<tr>
<td>Figure 18</td>
<td>Proposed structures of NEAT1 TH long</td>
<td>36</td>
</tr>
<tr>
<td>Figure 19</td>
<td>UV spectra for NEAT1 TH long</td>
<td>37</td>
</tr>
<tr>
<td>Figure 20</td>
<td>Proposed structure of NEAT1 TH long with mutated GQ5 region</td>
<td>38</td>
</tr>
<tr>
<td>Figure 21</td>
<td>KCl dependence PAGE for NEAT1 TH short WT GQ5 and mut GQ5</td>
<td>39</td>
</tr>
</tbody>
</table>
Figure 22. 1H NMR spectra of NEAT1 TH short WT GQ5 .............................................. 40
Figure 23. 1H NMR overlay at 150 mM KCl ...................................................................... 41
Figure 24. UV spectra of all NEAT1 TH short sequences ................................................ 42
Figure 25. PAGE with fluorescent A-tail binding to NEAT1 TH short ............................... 44
Figure 26. PAGE with fluorescent A-tail binding to NEAT1 TH short WT GQ5 and NEAT1 TH short GQ5 mut ................................................................................................... 46
Figure 27. FUS RGG binding gels with NEAT1 GQ5 WT ................................................ 47
Figure 28. Structure of PNAs ....................................................................................... 72
Figure 29. Transcription gel for (G₄C₂)₄₅ ........................................................................... 75
Figure 30. Results of phenol chloroform extraction of (G₄C₂)₄₅ ........................................ 76
Figure 31. NMM stained agarose gel of (G₄C₂)₄₅ ............................................................. 77
Figure 32. Control NMM GQ agarose gel ........................................................................ 77
Figure 33. PNA binding to (G₄C₂)₄₅ .............................................................................. 78
LIST OF ABBREVIATIONS

$^1$H NMR  proton nuclear magnetic resonance spectroscopy

ALS  amyotrophic lateral sclerosis

CD  circular dichroism spectroscopy

FUS  fused in sarcoma

GQ  G-quadruplex

KCl  potassium chloride

IncRNA  long non-coding RNA

MALAT1  metastasis associated lung adenocarcinoma transcript 1

miRNA  microRNA

mut  mutant

ncRNA  non-coding RNA

NEAT1  nuclear enriched abundant transcript

NMM  n-methyl mesoporphyrin IX

PAGE  polyacrylamide gel electrophoresis

PSP  paraspeckle protein

RBP  RNA binding protein

RGG  arginine-glycine-glycine

$T_m$  melting temperature

TH  triple helix

UV  ultraviolet spectroscopy

WT  Wild type
1.1 Central Dogma of Biology and Non-Coding RNAs

The central dogma of biology in its most basic form states that DNA makes RNA and RNA makes protein. It is this basic framework that allows for the understanding of how genetic information is carried through living organisms. Within this framework, DNA is replicated, RNA is transcribed from sections of DNA, and amino acid sequences which fold into proteins are translated from RNA (Figure 1). The processes occur via an intricate framework of machinery in the cell. Though the central dogma is considered the “golden standard”, not all RNA is translated into proteins. Coding RNA, known as messenger RNA (mRNA) is that which will be translated into protein, whereas non-coding RNA (ncRNA) is that which is not translated into protein (Figure 1). These ncRNAs serve different functions and play different roles than their coding counterparts.

Figure 1. Central dogma of biology – DNA (red block) is replicated to form identical DNA, DNA is transcribed to form complimentary RNA (gold block), RNA is translated into amino acid sequences which form protein (green block). RNA can either be coding (light blue block), like messenger RNA, or non-coding (light orange block).
Non-coding RNAs were first characterized in 1965 when researchers discovered an alanine transfer RNA (tRNA) in baker’s yeast (Holley et al. 1965). The three-dimensional structure of the tRNA was established years later when two independent research groups confirmed the proposed “cloverleaf” structure known today (Kim et al. 1973, Ladner et al. 1975). Another RNA within this non-coding family is ribosomal RNA (rRNA). rRNA makes up ribosomes and is an essential component of the cell. The main role of these rRNAs which assemble into the ribosomal two-subunits is protein synthesis, carried out by reading the coding mRNA. Once ncRNAs were established as functionally important entities, the discovery of different types of ncRNAs has continued, including various short and long ncRNAs.

One type of short ncRNA is micro RNA (miRNA). The first miRNA was discovered in 1993 (Lee et al. 1993) and since then, miRNAs have been studied extensively to determine their roles and biological functions. More than a thousand miRNAs have been characterized and have been found to be highly conserved among humans, other animals, and plants. Most miRNAs can interact with human mRNAs, suggesting that they play important roles in developmental processes (reviewed, Bartel 2018). These 22-24 nucleotide molecules function in RNA silencing – base-pairing with complementary mRNA – and posttranscriptional regulation of gene expression (Bartel 2009). miRNAs are found abundantly in many mammalian cell types (Lagos-Quintana et al. 2002) and they recognize their target mRNAs with a “seed region”, usually 6-8 nucleotides in length (Lewis et al. 2003, Lewis et al. 2005, Friedman et al. 2009). miRNAs resemble another type of short ncRNA, the small interfering RNA (siRNA). These siRNAs are double stranded RNAs that are found in the RNA interference pathway which work by degrading mRNA after transcription, thus preventing translation (Lagana et al. 2014). The implications of miRNAs in disease is another reason they are a valuable area of study. Cancer, heart disease, kidney disease, and obesity are
among the diseases miRNAs have been linked to, and a publicly available comprehensive database of miRNAs in disease was curated to help document the dysregulation of miRNAs in disease (Jiang et al. 2008).

Due to the diversity of the single stranded miRNAs, they are able to bind targets other than the mRNAs mentioned previously. In fact, miRNAs have been shown to interact with other ncRNAs like long non-coding RNAs (lncRNA) (Jalali et al. 2013). Long non-coding RNAs are classified as ncRNAs greater than 200 nucleotides in length. Because of transcriptional and structural similarities to mRNA, it is no surprise that miRNAs will target lncRNAs as well. lncRNAs can even act as miRNA “sponges”, that is, a molecule by which miRNA binding causes a loss of function of the miRNA (Qu et al. 2015, Chen et al. 2018). The roles of lncRNA do not stop at miRNA regulation, and are extremely varied, as lncRNA have become more widely studied because of their roles in translation and post-translation processes and implications in disease (Liu and Ding 2017, Chen et al. 2018, Liu et al. 2018).

Other lncRNA functions are depicted in Figure 2 (Balas and Johnson 2018, Wang and Chang 2011). Unlike miRNAs, lncRNAs have been found to be mostly (78%) tissue specific (Cabili et al. 2011). The function of lncRNAs in certain disease states is still being investigated, though function seems to be dependent on location – that is, depending on where the lncRNA is in the cell, its functions will be different. Around 30% of lncRNAs are found exclusively in the nucleus, around 15% are found exclusively in the cytoplasm, and the remaining 55% can be shuttled back and forth between the nucleus and the cytoplasm (Kapranov et al. 2007). Nuclear lncRNAs are essential for transcriptional modulation, while cytoplasmic lncRNAs can actually target mRNAs, affecting their function and stability (Mercer and Mattick 2013).
Considerable research has been done in the area of lncRNAs and their roles in disease (Dhanoa et al. 2018). Specifically, certain lncRNAs have been shown to be either upregulated or downregulated in cancers. Metastasis associated lung adenocarcinoma transcript 1 (MALAT1), a highly characterized lncRNA, is overexpressed in various types of cancer (Prensner and Chinnaiyan 2011). MALAT1 has been shown to act as a miRNA sponge in ovarian cancer, where miRNA-200c acts as a tumor suppressor (Pa et al. 2017). In their study, they found that miRNA-200c binds directly with a complementary sequence on lncRNA MALAT1. Here, MALAT1 actually reversed the tumor suppressive function of miRNA-200c by regulating its activity, and thus MALAT1 was able to achieve its function to promote metastasis of ovarian cancer cells (Pa et al. 2017). This study and others like it indicate the importance of lncRNAs biologically, as well as their potential for therapeutic targets in disease. These findings also stress the importance of interactions of ncRNAs with one another.

The specific functions that lncRNA can serve stem from four general categories of activity (Figure 2). Those categories include signaling lncRNAs, decoy lncRNAs, guide lncRNAs, and scaffold lncRNAs. Signal lncRNAs are believed to be such because of their cell type specific expression. As mentioned previously, 78% of lncRNAs are tissue
specific; this suggests they are under tremendous transcriptional control. This allows them to serve as molecular signals, recruiting molecules to specific tissues, because transcription of individual lncRNAs occurs at a very specific place (Wang and Chang 2011, Li et al. 2014). Some signal lncRNAs have regulatory roles, the advantage of which indicates that some regulatory functions do not rely on protein translation. A second function of lncRNAs is that they act as decoys. The decoys seek to limit the availability of regulatory factors by acting as a “molecular sink” (Balas and Johnson 2018). This works because the lncRNAs bind and titrate away protein targets, but do not perform any other functions. It is believed that the lncRNAs that fit into the decoy category act by negatively regulating effectors. Thirdly, lncRNAs can act as guides, where they bind to proteins and direct the ribonucleoprotein complex to a specific target. Guides are also necessary for the localization of factors specific for the regulation of the genome (Balas and Johnson 2018).

One last function of lncRNAs is that they are necessary in the formation of dynamic scaffolds. Here, lncRNAs can serve as central platforms for relevant molecular components to assemble, such as RNA and protein structures within the cell. Proteins were originally thought to be the main component of these diverse scaffolds, but more recently however, it has been found that lncRNAs might play a similar role (Li et al. 2014). One of these ribonucleoprotein scaffolds, located in the inter-chromatin space of mammalian cells, is known as the paraspeckle.

1.2 Paraspeckles

The nucleus of a cell is highly organized because of the presence of many different nuclear bodies. Many of these nuclear bodies are becoming more and more promising areas of interest in the study of different neurodegenerative diseases. One of the nuclear bodies of particular interest is the paraspeckle. Paraspeckles are relatively newly identified subnuclear bodies that were accidentally discovered by a postdoctoral student studying fusion proteins (Fox et al. 2002) which
get their name because of their proximity to nuclear splicing speckles (they are parallel to the nuclear splicing speckles). They are only present in mammalian nuclei, and the number varies by species and cell type (Fox et al. 2002, Bond and Fox 2009). It was discovered that a novel protein, termed paraspeckle protein 1 (PSP1), localized to specific areas within the interchromatin space. PSP1 has become the standard marker used to identify these membraneless organelles (Bond and Fox 2009). Between 2012 and 2013, 53 different proteins had been identified as localizing to paraspeckles (Naganuma et.al 2012, Fong et al. 2013).

Paraspeckles have been described as dynamic structures, since they are comprised of proteins that cycle to and from the nucleoli (Fox et al. 2002). Along with PSP1, several other proteins that belong to the same family were also found to be abundant within paraspeckles. These newly identified PSPs are not exclusive to paraspeckles, however, so studying the function of paraspeckles is difficult. It has been established that paraspeckles act as a molecular sponge, sequestering PSPs and modulating their functions outside of the paraspeckle (Chen et al. 2009, Imamura et al. 2014). At the organism level, paraspeckles have been shown to increase or suppress tumor function, depending on the cell lines studied (Adriaens et al. 2016, Mello et al. 2017).

A few years after they were first described, it was suggested that paraspeckles were RNase sensitive structures when a study showed that RNA binding proteins localized to these paraspeckles (Fox et al. 2005). Upon further study, the PSPs were found to each contain RNA binding domains. It follows, then, that paraspeckles would be comprised of essential PSPs, as well as various RNA that would bind to the PSPs, creating the scaffold of the paraspeckle that allows it to function (Figure 3).
The function of paraspeckles is not fully known; however, it is suggested that they localize PSPs that may be involved in transcription, directing the activity of the process in the nucleus (Fox et al. 2018). Therefore, it is thought paraspeckles provide a higher level of organization in the nucleus for both PSPs and certain lncRNA. One of the lncRNAs that was found to be essential in the formation and structure of paraspeckles is called nuclear enriched abundant transcript 1 (NEAT1/MENβ). When studies of NEAT1 knockout cells were performed, there was a decrease in the number of structured paraspeckles formed (Clemson et al. 2009, Sasaki et al. 2009, Sunwoo et al. 2009). Using high resolution microscopy, West et al. identified the 5’ and 3’ termini of a NEAT1 isoform were present in paraspeckle shells, suggesting bundles of the RNA within the organelle.

Figure 3. Paraspeckle structure. PSPs that bind to NEAT1, like FUS (green ball), accumulate on the NEAT1 scaffold (blue ribbon) and recruit other proteins, forming the mature paraspeckle structure, around 360 nm in diameter. Figure adapted from Fox 2019 (website).
1.3 Nuclear Enriched Abundant Transcript 1

Nuclear enriched abundant transcript 1 (NEAT1) is also called nuclear paraspeckle assembly transcript 1 because of its role in the formation of paraspeckles. Discovered in 2007, it has been found to be prevalent in a number of tissues, including ovary, pancreas, colon, and prostate (Hutchinson et al. 2007). It is transcribed from a gene on chromosome 11, specifically from the multiple endocrine neoplasia locus (MEN1), which gives NEAT1 its other name, MENβ (Guru et al. 1997, Prinz et al. 2019). There is no evidence of NEAT1 in embryonic stem cells, but it appears upon cell differentiation (Chan and Carmichael 2009). NEAT1, as indicated by its name, is enriched in the nucleus, but can be found in the cytoplasm as well (van Heesch et al. 2014). Overexpression of NEAT1 has been implicated in several cancerous tumors, including those for non-small cell lung cancer, ovarian cancer, cervical cancer, pancreatic cancer, breast cancer, and more (Dong, et al. 2018 and references therein). NEAT1 acts in cancer by preventing cell apoptosis and promoting metastasis (Dong et al. 2018), making it a potential candidate for therapeutics of specific types of cancers. In addition to cancer, NEAT1 has been found to be of importance in certain viral infections, namely HIV (Zhang et al. 2013). When NEAT1 was knocked down, viral production increased. Furthermore, studies have shown that NEAT1 plays an important role in corpus luteum formation in mice and is essential for mammary gland development and lactation in mice (Nakagawa et al. 2014, Standaert et al. 2014). With a countless number of regulatory and biological applications, NEAT1 is an interesting RNA to study, and its role in the formation and structure of paraspeckles makes it especially interesting to our research.

NEAT1 has been investigated for its ability to bind paraspeckle proteins, namely NONO, an essential PSP, at its 5’ and 3’ ends (Murthy and Rangarajan 2010). The location of the PSP binding sites at the ends of the RNA support the study of West et al. that the arrangement of
NEAT1 in paraspeckles is that like which is seen in Figure 3 – the 5’ and 3’ terminals are near the outer portion of paraspeckles, and the middle section is located in the core, indicating the structure of NEAT1 is similar to a ‘V’ shape.

NEAT1 has two different isoforms, the NEAT1_1 isoform is the shorter of the two, around 3,700 nucleotides, and the NEAT1_2 isoform is much longer, containing around 22,700 nucleotides (Figure 4). NEAT1 has been identified as an integral component of paraspeckles (Clemson et al. 2009, Sasaki et al. 2009, Sunwoo et al. 2009). Specifically, the NEAT1_2 isoform is essential in the formation of paraspeckles, and mouse cells that only express NEAT1_1 do not form paraspeckles (Nakagawa et al. 2011, Isobe et al. 2020). The two isoforms are identical to the end of NEAT1_1, indicating that the sequence and/or structure that is present within the longer region of the NEAT1_2 isoform is indispensable in the formation of paraspeckles. Fused in sarcoma (FUS), a PSP that will be discussed in further detail below, is responsible for regulating the assembly of NEAT1 into paraspeckles (West et al. 2016).

At the 3’ terminus of NEAT1_2, there exists a triple helix (TH) structure (Wilusz et al. 2012) that is thought to be integral for protection of that 3’ end. TH structures form when two nucleotides interact via Watson-Crick hydrogen bonding, and a third nucleotide interacts with one of the two via Hoogsteen hydrogen bonding, forming a base triple like that shown in Figure 5A. The TH structure is not exclusive to NEAT1_2 and has been described in other lncRNAs as well. For example, another lncRNA, MALAT1, has been identified as containing a TH structure at its
3’ end as well (Brown et al. 2014). MALAT1 plays a role in the progression of some cancers and its TH has been highly characterized (Ageeli et al. 2018, Brown et al. 2012, Brown et al. 2014). To date, most TH structures that have been studied show no more than six consecutive base triples. Recently, a crystal structure for MALAT1 containing 11 base triples was solved to a resolution of 3.10 Å (Ruszkowska et al. 2020). Figure 5B shows the 11 base triple MALAT1 TH sequence, and Figure 5C shows the 3-dimensional representation of the solved crystal structure. The crystal structure clearly shows how the 3’ end of the molecule folds upon itself to form the base triples involved in the TH structure, protecting it from degradation by exonucleases.

The terminal TH of NEAT1_2, along with the fact that it is the isoform that is indispensable for paraspeckle formation, suggests that secondary structure of NEAT1 plays an important role in its function and necessity in paraspeckle formation. In addition to TH, another non-canonical secondary structure shown to facilitate RNA function is the G quadruplex (GQ). Thus, we used
the QGRS Mapper software (http://bioinformatics.ramapo.edu/QGRS/analyze.php), to predict that five regions within NEAT1_2 have the potential to form GQ structures. These regions are listed in Table 1. Of these structures, four are common to both NEAT1_1 and NEAT1_2, whereas the last identified region, is unique to NEAT1_2, being located 20 nucleotides away from its terminal TH. The influence these secondary structures, TH and GQ, have on one another may be integral to understanding NEAT1 stability and function.

<table>
<thead>
<tr>
<th>Nucleotide position</th>
<th># bases</th>
<th>Sequence</th>
<th>GQ potential</th>
</tr>
</thead>
<tbody>
<tr>
<td>425</td>
<td>30</td>
<td>GGGTCAGGGCTGCAGGAGGGCGGGAGGG</td>
<td>38</td>
</tr>
<tr>
<td>2070</td>
<td>28</td>
<td>GGGAGTTGTGCGGGGGACGGAGAAGGGG</td>
<td>41</td>
</tr>
<tr>
<td>2171</td>
<td>27</td>
<td>GGGATGAGGGTTGAGAAGGGGAGAGGG</td>
<td>39</td>
</tr>
<tr>
<td>3226</td>
<td>23</td>
<td>GGGAAGGGGATGGGGATTGTGGG</td>
<td>40</td>
</tr>
<tr>
<td>22619</td>
<td>15</td>
<td>GGAGGGAGGGGAGGG</td>
<td>42</td>
</tr>
</tbody>
</table>

Table 1. Potential GQs in NEAT1

1.4 G Quadruplexes

Traditionally, DNA adopts a double helix structure like that which was first described in 1953 by Watson and Crick (Watson and Crick 1953). While this is the canonical structure most commonly adopted by DNA, there are several noncanonical structures that DNA can adopt as well. In addition to the TH structure discussed above, another non-canonical structure formed by guanine-rich sequences of both DNA and RNA is the GQ. The GQ (review, Bochman et al. 2012) – a structure that was first discovered in 1963 (Gellert et al. 1962) is formed when four guanine residues within a guanine-rich region of DNA/RNA form planar tetrads that stack upon one another. Referencing Stegle et al., the standard nucleotide formula for GQs is: $G_XN_{1-7}G_XN_{1-7}G_XN_{1-7}G_X$, where $X (>2)$ represents the number of guanines in the G-rich sequence and N is the number of nucleotides in the loop sequence between tetrads. The length of N has been shown to
have an effect on the stability of the GQ, and the shorter the loop sequence, the more stable the structure (Zhang et al. 2011).

GQs are thought to serve an important role in the human genome and have several biological functions. Many DNA sequences that have been characterized to have GQ structures are intramolecular telomeric sequences (Wang and Patel 1993, Mills et al. 2002, Neidle et al. 2003). Because of their prevalence in human telomeres, GQs are thought to be regulators of telomere processing, and they can be potential targets for cancer therapeutics (Neidle 2010). The prevalence of GQs in telomeric sequences, as well as their presence in promoter regions of DNA throughout the human genome, lead to the assumption that GQs play an important role in transcription regulation (Huppert and Balasubramanian 2006).

Typically, base pairs share a Watson-Crick hydrogen bonding, however the guanine residues of GQs share a Hoogsteen hydrogen bonding, allowing this square planar orientation to manifest, as shown in Figure 6A. The tetrads are stabilized by intercalating potassium (K+) ions, leading to the stacked structure seen in Figure 6B (Hud et al. 1996, Sen and Gilbert 1990). GQs

![Figure 6](image)

**Figure 6.** GQs form as four guanine residues come together via Hoogsteen hydrogen bonding to form planar G tetrads (A). The tetrads stack on top of one another (B) to form the GQ structure, stabilized by K⁺ ions.
can form as intermolecular (between different strands) or intramolecular (within the same strand) species and can adopt a parallel or antiparallel orientation seen in Figure 7. Parallel GQs are named such because the loops between the tetrads are oriented in the same direction, that is, they are all in the 5’ – 3’ direction. Antiparallel GQs on the other hand have loops that are oriented in opposite directions. For instance, one loop sequence may be oriented 5’ – 3’, while the next is oriented 3’ – 5’. The structures can also exist as a hybrid.

![Figure 7. Topology of GQ structure. GQs can be parallel (A), antiparallel (B), or a hybrid (C and D). Figure adapted from Moraca et al. 2016.](image)

Though originally described in DNA, GQs have been reported in RNA as well (Kumari et al. 2007). The typically single-stranded nature of RNA is favorable for GQ formation, since there is no second strand contributing to structural hindrance, and there are less bases involved with Watson-Crick base pairing (Agarwala et al. 2015, Fay et al. 2017). Unlike their DNA counterparts, RNA GQs do not require K^+ in order to fold, however, the absence of K^+ means the resulting GQ structure will have lower stability (Joachimi et al. 2009). Studies have shown that GQ structures in RNA 5’-UTR regions act as regulators, usually repressors, of translation (Beaudoin and Perreault 2010, Bugaut and Balasubramanian 2012). Another role of GQ structure lies within the miRNA pathway. The short non-coding miRNAs also regulate transcription and translation and their dysregulation has been implicated in the progression of different cancers (reviewed in Lorio and Croce 2012, Chen et al. 2013). Understanding the role of GQs in the progression of disease
is key to understanding how to better treat patients, as well as provide potential targets for therapeutics.

A role of GQs in RNA that will be focused on here is the ability for them to bind to the RNA binding sites of RNA binding proteins (RBPs). Specifically, GQs have shown to bind selectively to arginine-glycine-glycine (RGG) repeat regions. These are regions that are rich in arginine and glycine residues but are not exclusively RGG repeat expansions. Binding of GQs to these regions is influenced by arrangement of RGG repeats, as well as by the gap amino acids between repeats (Huang et al. 2018). The proteins that bind to GQs have been identified as being involved in several important biological roles and processes, including alternative splicing, translation, and repeat expansion diseases (Cooper et al. 2009, Ji et al. 2011). FUS, mentioned previously, is a PSP and RBP that has been shown to bind RNA GQs within its RGG repeat region (Takahama et al. 2013, Imperatore et al. 2020).

1.5 **Fused in Sarcoma and Neurodegeneration**

FUS is also known as heterogeneous nuclear ribonuclear protein P2. It belongs to the FET family of proteins, a family that also includes Ewing’s Sarcoma and TATA-binding protein associated factor 15. The FET family of proteins can be found expressed in all examined tissue types, is found predominantly in the nucleus, and also has some functions in the cytoplasm of the cell (Zinsner et al. 1994, Morohoshi et al. 1996, Andersson et al. 2008). The main roles of FUS lie within the nucleus guiding transcription through interactions with RNA Polymerase II and the TATA binding protein, and it also functions in mRNA processing (Aman et al. 1996, Powers et al. 1998, Tan and Manly 2009, Lagier-Tourenne, et al. 2010). FUS can be shuttled outside of the nucleus, a process that is facilitated by its C-terminal proline-tyrosine nuclear localization signal (Lee et al. 2006, Dormann et al. 2012). Outside of the nucleus, FUS can be found within neurophil
granules and within the post-synaptic space of rat hippocampus cells (Aoki et al. 2012, Schoen et al. 2016). The mechanisms of function of FUS outside of the nucleus, however, are not well understood.

The FET proteins contain several conserved regions, including an RNA recognition motif (RRM), arginine-glycine-glycine (RGG) repeat regions, and a Cys2-Cys2 zinc finger region. Specifically, FUS contains an RRM, a low-complexity prion-like domain, zinc finger, three RGG domains, as well as a N-terminus region with transcriptional activating properties, and a C-terminus region that is capable of binding DNA and RNA (Iko et al. 2004, Law et al. 2006, Tan and Manly 2009). A schematic representation of FUS can be found in Figure 8. FUS can bind secondary structures of DNA and RNA, specifically hairpin structures with U-U or U-C base pairing and GQ structures (Hoell et al. 2011, Takahama et al. 2013). Literature has reported that the C-terminal RGG domain (RGG3) of FUS specifically binds to GQ secondary structure (Takahama and Oyoshi 2013, Ozdilek et al. 2017, Yagi et al. 2018, Imperatore, et al. 2020).

![Figure 8. Schematic representation of FUS. Diagram shows the low-complexity prion-like domain (LC Domain), RNA recognition motif (RRM), zinc finger region (ZnF), as well as all three RGG domains.](image)

Mutations in FUS have been associated with the onset of certain neurological diseases (Lai et al. 2011). Namely, amyotrophic lateral sclerosis (ALS) has been linked to specific mutations of the FUS gene on chromosome 16 (Kwiatkowski et al. 2009, Vance et al. 2009, DeJesus-Hernandez 2010). Most of the FUS mutations associated with ALS are located within its C-terminal RGG3 domain and there is an increase in the amount of cytoplasmic FUS with these mutations (Hewitt et al. 2010). ALS is the most common motor neuron disease in adults and is
always fatal, usually within 1-5 years of onset (Cleveland and Rothstein 2001). Upwards of 90% of ALS cases are sporadic (sALS), while around 10% of cases are familial (fALS). The disease is characterized by a progressive loss of motor function and results in the atrophy of muscles, leading to paralysis and eventual death by respiratory failure. There currently exists no cure and only few treatments to extend life after diagnosis. Interestingly, there is a link between ALS progression and FUS with NEAT1 and the formation of paraspeckles.

FUS has been recognized as an essential paraspeckle protein, located within the core of the paraspeckle (Shelkovnikova et al. 2013, Hennig et al. 2015). Its role in paraspeckle formation is the regulation of assembly of NEAT1_2 RNA into the scaffold structure (West et.al 2016). In healthy individuals, paraspeckle formation and NEAT1_2 recruitment by FUS is normal; however, individuals with ALS exhibit FUS mutations mentioned previously and abnormal paraspeckle levels. As reviewed in An et al. 2018, literature reports that NEAT1_2 is upregulated in early stages of ALS. Particularly, an abundance of paraspeckles were observed in spinal motor neurons of patients (Nishimoto et al. 2013). Further, studies have shown that FUS mutations linked to ALS leads to an abundance of structurally and functionally compromised paraspeckles (An et al. 2019).

1.6 Motivation

NEAT1_2 has been proven to be indispensable in the formation of paraspeckles. The link between ALS and the increased number of paraspeckles in affected individuals leaves a window of opportunity to study the role of this lncRNA on disease onset, progression, and potential therapeutic development. The 3’ terminal TH region, along with the potential for GQ formation in several regions of the RNA, makes studying the secondary structure of NEAT1_2 particularly interesting.
Of the five potential GQ forming regions of NEAT1 that were analyzed in-house, four are common to both isoforms, while one is unique to the NEAT1_2 isoform. The GQ region that is unique to NEAT1_2, “GQ5”, is located at nucleotide position 22,619, just 20 bases away from the beginning of the TH structure discussed previously (Figure 9). To date, there has not been a study done that includes the GQ5 region when characterizing the TH of NEAT1_2. The structural motifs proposed by Wilusz et al. 2012 and Yamazaki et al. 2018 contain only the region that forms the TH; however, the sequence used in this study also considers the region proposed to form GQ5, as depicted in Figure 9B.

We hypothesize first that the GQ5 region will form a stable GQ which we will characterize using several biophysical techniques. Secondly, we propose that this GQ region will bind to PSP FUS at its RGG3 domain, which could lead to insight for paraspeckle development and NEAT1_2 recruitment by FUS. Lastly, we hypothesize that the presence of the GQ at region GQ5 will influence the formation of the TH at the 3’ terminus of NEAT1_2, thereby influencing the stability of the NEAT1_2 molecule.

Figure 9. Predicted structure of the triple helix region of NEAT1_2 including the GQ5 region (A) and the sequence of the region with the proposed triple GQ indicated in the box (B).
Chapter 2: Materials and Methods

2.1 RNA Synthesis

All RNA sequences (Table 1), wild-type NEAT1 GQ5, its NEAT1 GQ5 mutant, extended
NEAT1 GQ5 TH, NEAT1 TH short, NEAT1 TH short WT GQ5, and NEAT1 TH short GQ5
mutant sequences were synthesized \textit{in vivo} from synthetic DNA templates (Integrated DNA
Technologies) using T7 RNA polymerase that was produced in-house (Milligan \textit{et al}. 1987). The
RNAs were purified using 20% denaturing urea polyacrylamide gel electrophoresis (PAGE)
followed by electroelution and extensive dialysis against 10 mM cacodylic acid at pH 6.5. The A-
rich tail sequences (Table 1) were purchased from Dharmacon, Inc., including the A-rich tail
sequence tagged with the fluorophore DY547.

2.2 Peptide Synthesis

FUS RGG peptide regions were chemically synthesized by the Peptide Synthesis Unit at
the University of Pittsburgh Center for Biotechnology and Engineering.

2.3 Polyacrylamide Gel Electrophoresis

\textit{KCl dependence}

RNA samples were prepared at various concentrations in the presence of different
concentrations of KCl and annealed by boiling for five minutes, followed by cooling for 10 minutes
at room temperature on the bench top. Non-denaturing loading dye (1 µL) was added to each
sample prior to loading onto a 20% acrylamide:bis-acrylamide TBE gel and running at 88 V for
approximately 3.5 hours at 4°C. For longer sequences (>100 nt), samples were loaded onto a 15%
acrylamide:bis-acrylamide TBE gel and run at 88 V for approximately 3 hours at 4°C. Upon
completion, gels were visualized with UV-shadowing at 254 nm (Hendry and Hannan, 1996) using
an Alphalmager (AlphaInnotech). Gels were subsequently stained with GQ-specific N-methyl
mesoporphyrin 9 (NMM) dye (Arthanari et al 1998) and visualized with the Alphaimager. BDNF and miR-122 were used as positive and negative controls, respectively. Each of the experiments were performed at least in duplicate.

**RNA concentration**

Samples were prepared with varying concentrations of NEAT1 GQ5 WT RNA (1 nM – 10 µM). Samples were annealed by boiling for five minutes and cooled for 10 minutes at room temperature on the bench top. Once cool, non-denaturing loading dye (1 µL) was added to each sample prior to loading onto a 20% acrylamide:bis-acrylamide TBE gel and running at 88 V for approximately 3.5 hours at 4°C. Completed gels were stained in ethidium bromide (EtBr) and visualized using the Alphaimager as described previously. Subsequently, the same gel was stained in SYBR gold dye, a dye more sensitive to lower concentrations and again visualized with the Alphaimager.

**FUS RGG3 binding**

For FUS RGG3 binding studies, samples were prepared at appropriate RNA concentrations, and the KCl concentration was held constant at 5 mM. Samples were annealed by boiling for five minutes and then cooled for 10 minutes at room temperature on the bench top. Once cool, FUS RGG3 peptide was added in varying concentrations and allowed to equilibrate for 15 minutes at room temperature. Samples were run onto a 20% acrylamide:bis-acrylamide TBE gel at 88 V for approximately 3.5 hours at 4°C. For longer sequences (>100 nt), samples were run onto a 15% acrylamide:bis-acrylamide TBE gel at 88 V for approximately 3 hours at 4°C. Upon completion, gels were visualized with UV-shadowing at 254 nm (Hendry and Hannan, 1996) using an Alphaimager. Gels were also stained with NMM or SYBR Gold dyes, depending on conditions,
and visualized with the AlphaImager. Each of the experiments were performed at least in duplicate.

**A-rich tail binding gels**

A-rich tail binding gels were performed with the NEAT1 TH short, NEAT1 TH short WT GQ5, and NEAT1 TH short GQ5 mut sequences. Samples were prepared with 1 µM RNA, 50 mM KCl, 10% glycerol (vol/vol), and 1 mM MgCl₂. These conditions were modeled after the methods in (Kunkler, et al. 2019). Samples were annealed for 5 minutes by boiling and then immediately snap-cooled on dry ice for 10 minutes and equilibrated to room temperature on the bench top. The A-rich tail sequence was added in increasing concentrations from 0.25:1 – 3:1 A-tail:NEAT1. Separate experiments were performed for native A-rich tail and A-rich tail that was tagged with the fluorophore DY547. Samples were run on a 20% acrylamide:bis-acrylamide TBM gel. Gels containing the TH short sequence were run at 150 V for 4 hours at room temperature. Gels containing the TH short WT GQ5 and TH short GQ5 mut sequences were run at 150 V for 6 hours at room temperature. Once complete, gels were stained in SYBR gold dye for 5 minutes and visualized using the AlphaImager. Gels including the fluorescently tagged A-tail were first visualized using a LI-COR Odyssey Fc on which fluorescent channels 500 and 600 were used to excite the DY547 fluorophore. After fluorescent visualization, gels were stained in SYBR Gold dye for 5 minutes and visualized using the AlphaImager as described above.

**2.4 Circular Dichroism (CD) Spectroscopy**

All CD experiments were performed using a Jasco J-810 spectropolarimeter at 25°C, using a 200 µL quartz cuvette (Starna Cells) with a 1 mm path length, with scans from 180 to 300 nm. Each sample was scanned seven times with a 1 s response time and 2 nm bandwidth. For potassium dependent GQ formation experiments, KCl was titrated into the sample and allowed to equilibrate
for 5 minutes between spectra acquisition. After the final titration, the sample was annealed by boiling for 5 minutes and cooled for 10 minutes at room temperature before spectra acquisition. For GQ destabilization studies, TmPyP4 was titrated into the sample in varying ratios up to 10:1 TmPyP4:RNA. Samples were allowed to equilibrate for 5 minutes before spectra acquisition after each addition of TmPyP4.

2.5 1H Nuclear Magnetic Resonance (NMR) Spectroscopy

1-dimensional proton (1D 1H) NMR spectra were acquired at 25°C using a 500 MHz Bruker AVANCE spectrometer. Water suppression was performed using the Watergate pulse sequence (Piotto et al., 1992). Each acquisition averaged 2048 scans and for longer sequences (>100 nt), 8192 scans. RNA samples were prepared in 10 mM cacodylic acid at pH 6.5, in a 90% H2O/10% D2O ratio to a final volume of 250 μL. Potassium dependent GQ formation was observed by titrating KCl into samples up to 150 mM monitoring the imino proton resonances in the 10-15 ppm region of the spectrum.

2.6 Ultraviolet Spectroscopy Thermal Denaturation

All UV/Vis experiments were performed using a Varian Cary 3E UV-visible spectrophotometer with a Peltier temperature control cell holder. For the isolated NEAT1 GQ5 sequence, 200 μL samples were prepared in with 5 mM KCl and 10 mM cacodylic acid with a pH of 6.5. Samples were annealed for 5 minutes by boiling and then thermally denatured by increasing the temperature of the instrument from 25°C – 95°C at a rate of 0.2°C/minute, holding at each temperature for 3 minutes to allow for equilibration. To prevent evaporation, 200 μL of mineral oil was carefully added to the top of the sample cuvette. Absorbance changes were monitored at 295 nm, a wavelength that is sensitive to GQ denaturation (Mergny et al. 1998). RNA concentrations were varied from 5 μM – 50 μM to determine if the GQ structure that forms was
intramolecular or intermolecular. For intermolecular structures, the RNA concentration is dependent on the melting temperature \(T_m\), as shown by equation (1), where \(R\) is the gas constant and \(\Delta H^{\circ}_{\text{AH}}\) and \(\Delta S^{\circ}_{\text{AH}}\) are the Van’t Hoff thermodynamic parameters.

\[
\frac{1}{T_m} = \frac{R(n-1)}{\Delta H^{\circ}_{\text{vH}}} \ln C_T + \frac{\Delta S^{\circ}_{\text{vH}} - (n-1)R \ln 2 + R \ln n}{\Delta H^{\circ}_{\text{vH}}} \tag{1}
\]

For intramolecular GQ structures, the RNA concentration is independent of the \(T_m\), shown by equation (2).

\[
\frac{1}{T_m} = \frac{\Delta S^{\circ}_{\text{vH}}}{\Delta H^{\circ}_{\text{vH}}} \tag{2}
\]

Thermodynamic parameters were obtained by fitting the thermal denaturation curves to equation (3), where \(R\) is the gas constant and \(A_U\) and \(A_F\) represent the absorbance of the unfolded (denatured) and folded (native) GQ structures, respectively.

\[
A(T) = \frac{A_U + A_F e^{-\Delta H / RT} e^{\Delta S / R}}{e^{-\Delta H / RT} e^{\Delta S / R+1}} \tag{3}
\]

For triple helix denaturing studies with the NEAT1 TH short, NEAT1 TH short WT GQ, and NEAT1 TH short GQmut sequences, samples were prepared to 200 \(\mu\)L final volume with 1 \(\mu\)M RNA, 50 mM KCl, 10% glycerol (vol/vol), and 1 mM MgCl\(_2\) to match the conditions of the PAGE binding gels. The same experimental conditions were followed as before, however, absorbance changes were monitored at 265 nm, a wavelength more sensitive to TH and double strand denaturation.
Chapter 3: Results and Discussion

3.1 NEAT1_2 is proposed to form G-quadruplex structures in five guanine-rich regions

NEAT1_2 is just under 23,000 nucleotides in length, compared to its isoform, NEAT1_1, which contains just under 4,000 nucleotides. It has been shown to form a TH structure at its 3’ end (Wilusz et al. 2012), and there are several guanine-rich (G-rich) regions within this lncRNA, leading to the potential for GQ formation. Using the QGRS Mapper software (http://bioinformatics.ramapo.edu/QGRS/analyze.php), we have predicted five G-rich sequences to be potential sites for GQ formation (listed in Table 1). Our focus for this study will be on region 5, the only region that is exclusive to NEAT1_2. Beginning at nucleotide 22,619, this region has the highest potential of the five predicted regions to form a GQ as defined by its G score of 42. While this work was in progress, in 2020, Simko, et al. published work that focused on these same five regions of interest. Using circular dichroism (CD) spectroscopy, they determined that each of the five proposed regions formed a GQ structure (Simko et al. 2020).

| NEAT1 GQ5 WT | 5’ GGG AGG GAG GGA GGG AGG CGG 3’ |
| NEAT1 GQ5 mut | 5’ GGG AGG GAA GGA GAG AGG CGG 3’ |
| NEAT1 TH long | 5’ UCC AGC CCA UGG CGC CAC UGA GCC GGG UCA GCC GCU UUU ACU AGG GAG GAG AAA AAG CAAA 3’ |
| NEAT1 TH short | 5’ UCC AGC CCA UGG CGC CAC UGA GCC GGG UCA GCC GCU UUU ACU AGG GAG GAG AAA AAG CAAA 3’ |
| NEAT1 TH short WT GQ5 | 5’ UCC AGC CCA UGG CGC CAC UGA GCC GGG UCA GCC GCU UUU ACU AGG GAG GAG AAA AAG CAAA 3’ |
| NEAT1 TH short GQ5 mut | 5’ UCC AGC CCA UGG CGC CAC UGA GCC GGG UCA GCC GCU UUU ACU AGG GAG GAG AAA AAG CAAA 3’ |
| A-rich tail | 5’ GGAAAAAGCAAAA 3’ |
| A-rich tail DYSG7 | 5’ DYSG7-GGAAAAAGCAAAA 3’ |

Table 2. Various NEAT1 RNA sequences used for these studies.
Our hypothesis states that the GQ5 and TH regions affect one another, such that the TH will not form unless the GQ forms. In order to characterize the GQ structure, TH structure, and the influence of one on the other, several RNA sequences were studied (Table 2).

### 3.2 Biophysical Characterization of NEAT1_2 GQ5

_In vitro_ T7 RNA polymerase driven transcription was utilized to produce the NEAT1 RNA GQ5 sequences (Table 2) from synthetic DNA templates. To first determine if a GQ forms in this region, three different RNA sequences were studied, including two wild type (WT) sequences, and one mutant (mut) sequence (first three rows of Table 2). Several biophysical characterization techniques were utilized to help us determine if the GQ5 region was, indeed, favorable to forming a GQ. First, KCl dependence non-denaturing PAGE was performed for the isolated NEAT1 GQ5 WT sequence because potassium (K⁺) is a stabilizing ion for GQ structure (Hud et al. 1996, Sen and Gilbert 1990). The gel was visualized by UV shadowing where all RNAs, regardless of secondary structure, will be visible, even in the absence of KCl. Then the gel was stained by n-methyl mesoporphyrin IX (NMM) dye which will stain only the GQ structures (Arthanari _et al._ 1998). As seen in Figure 10A, under UV shadowing, NEAT1 GQ5 WT shows a main upper band with a lower faint band that becomes more apparent as the KCl concentration is increased. When the gel was stained in NMM (Figure 10B), both bands stained, indicating that they contain GQ structures. Two control samples were also used in this gel as a GQ positive (GQ+) RNA and a GQ negative (GQ-) RNA. Both RNA controls are present in the UV shadowed gel (Figure 10A) but only the GQ+ control RNA is present in the NMM stained gel (Figure 10B). When the same experiments were performed with the NEAT1 GQ5 mutant, in which the GQ sequence was disrupted, UV shadowing revealed RNA bands in each lane under all KCl conditions (Figure 10C), but when stained in NMM dye, no bands, other than the band in the GQ+ control lane, were present.
Taken together, these results support the hypothesis that the NEAT1 GQ5 WT sequence is a favorable region for GQ formation.

To further validate the presence of a GQ structure, $^1$H NMR spectroscopy experiments were performed for NEAT1 GQ5 WT, monitoring the imino proton resonance region changes with increasing concentrations of KCl from 0-150 mM (Figure 11). Focusing first on Figure 11A, there are prominent resonances in the region between 10-12 ppm at all concentrations of KCl investigated. Only slight changes were seen between spectra at higher concentrations of KCl: the resonance centered around 11.1 ppm appears to broaden, while smaller resonances between 11.3-11.7 ppm disappear. Changes in resonances can be seen more clearly in Figure 11B, showcasing the overlay of the spectra.

The region between 10-12 ppm of the NMR spectrum corresponds to the imino protons involved in Hoogsteen base pairs, the type of bonding that occurs between guanine residues in a
G-quartet stabilizing GQ structure (Sen and Gilbert 1990). In contrast, imino protons found in traditional Watson-Crick hydrogen bonding appear between 12-14.5 ppm. The spectra in Figure 11 show strong resonances between 10-12 ppm, the Hoogsteen region, even in the absence of KCl, indicating the formation of a GQ structure that does not require the presence of K⁺ ions. The disappearance of the resonances between 11.3-11.7 ppm could indicate that alternate GQ structures are converted to a more stable conformation in the presence of K⁺ ions. The observed broadening of all resonances upon addition of higher KCl concentration could indicate the presence of stacked

**Figure 11.** ¹H NMR spectroscopy of NEAT1 GQ5 WT sequence. (A) KCl was titrated into samples from 0 mM (blue line) to 150 mM (lime green line). (B) Overlay of spectra at differing KCl concentrations
GQ structures promoted by the high RNA concentration used in the NMR experiments. Absent from the spectra, however, are resonances within the Watson-Crick region between 12-14.5 ppm, indicating that there is no hairpin structure formed by this sequence. Based on the nucleotide sequence, this is not surprising, as the GQ is the predominant structure that could be formed by this RNA. The $^1$H NMR spectroscopy experiments confirm the results of the KCl dependence PAGE, indicating the formation of a GQ structure by the NEAT1 GQ5 WT sequence.

CD spectroscopy can be used to study the folding of the RNA to determine if a GQ is parallel or antiparallel. CD spectroscopy works by exposing a sample to circularly polarized light; depending on the extent to which right or left circularly polarized light is absorbed, there will be different signatures on the resulting spectrum (reviewed, Ranjibar and Gill, 2009). For the NEAT1 GQ5 WT sequence, there is a strong positive band observed around 265 nm, and a negative band observed around 240 nm (Figure 12A). Similar to the NMR spectroscopy experiments, KCl was titrated into the sample at increasing concentrations. There is no distinct difference between spectra upon the addition of KCl, and only when the sample was boiled (orange line, Figure 12A).

**Figure 12.** CD spectroscopy results for A) NEAT1 GQ5 WT sequence in the presence of KCl and B) NEAT1 GQ5 WT sequence with TmPyP4
was there any small change in intensity of the positive band at 265 nm. This indicates that there is a parallel GQ present within the GQ5 region of NEAT1_2, as CD spectra with positive bands around 265 nm and negative bands around 240 nm are signatures of parallel structure (Williamson 1994). Consistent with the NMR spectroscopy results, the addition of KCl does not change the intensity of the bands observed, as seen in Figure 12A.

Another CD spectroscopy experiment was performed with the KCl concentration held constant at 5 mM, this time in the presence of varying concentrations of TmPyP4 (Figure 13) up to 10:1 concentration ratio. TmPyP4 is a porphyrin compound that has been shown to stabilize DNA GQs but destabilize RNA GQs (Fujiwara, et al. 2015, Morris, et al. 2012, Mikami-Terao, et al. 2008). The graph in Figure 12B shows CD spectra for the NEAT1 GQ5 WT region as TmPyP4 was added to the sample, showing that the band at 265 nm diminishes in intensity. Similarly, there is a decrease in intensity of the negative band at 240 nm. These results agree with literature (Fujiwara, et al. 2015) that TmPyP4 acts as a destabilizing agent for RNA GQs. The red spectrum in Figure 12B, recorded in the absence of TmPyP4, is the most intense and as TmPyP4 is added to the sample, we see a definite diminishing of intensity of the bands at 265 nm. Once a 10:1 concentration ratio of TmPyP4 was reached (hot pink spectrum in Figure 12B), the intensity decreased by almost half, indicating that the GQ structure had been destabilized.
After confidently determining that the GQ5 region of NEAT1_2 formed a parallel GQ, UV spectroscopy was utilized to determine if the GQ was intramolecular or intermolecular. A series of UV thermal denaturation experiments were performed with varying concentrations of NEAT1 GQ5 WT RNA. The temperature was increased from 25°C to 95°C at a rate of 0.2°C/minute recording the absorbance changes at 295 nm and the data was normalized. The results show a hypochromic transition between 70°C - 90°C, attributed to the denaturation of the GQ (Figure 14).

**Figure 14.** UV thermal denaturation spectra for NEAT1 GQ5 WT. (A) Overlay of spectra at various concentrations of RNA (B) Plot of T_m calculated from hypochromic transitions (C) 20% non-denaturing PAGE at various concentrations of NEAT1 GQ5 WT stained in ethidium bromide (top image) and SYBR gold (bottom image)
14A), present at all concentrations of RNA examined. Using Origin (OriginLab Corporation), the data was fit to Equation 3 to determine the melting temperature ($T_m$) for each thermal denaturation experiment. $T_m$ for each experiment can be found in Table 3. The plot in Figure 14B shows the $T_m$ for each RNA concentration, with a red line representing the average, 79.1°C. For intramolecular GQs, $T_m$ may be independent of the concentration of RNA (Equation 2). Figure 14B reveals that the $T_m$ does not change when RNA concentration is changed, indicating that the GQ formed may be intramolecular. Given that the range of concentrations investigated is narrow (5-50 µM), we ran a PAGE experiment with a wider range of RNA concentrations. Figure 14C shows the results of the experiment with a concentration range of 1 nM – 10 µM. The gel was first stained in ethidium bromide (EtBr) to get a clear image of bands at higher concentrations. Then, the same gel was stained in SYBR gold, a dye more sensitive to lower concentrations, in order to see bands in the nM range. This experiment allows us to observe a wider range of concentrations while the number of bands in each lane remains the same. This further supports that there may be an intramolecular GQ present.

<table>
<thead>
<tr>
<th>[NEAT1] (µM)</th>
<th>$T_m$ (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>79.50 ± 0.04</td>
</tr>
<tr>
<td>10</td>
<td>79.30 ± 0.01</td>
</tr>
<tr>
<td>20</td>
<td>78.40 ± 0.02</td>
</tr>
<tr>
<td>30</td>
<td>78.60 ± 0.01</td>
</tr>
<tr>
<td>50</td>
<td>79.90 ± 0.01</td>
</tr>
<tr>
<td><strong>Average $T_m$</strong></td>
<td><strong>79.10 ± 0.02</strong></td>
</tr>
</tbody>
</table>

**Table 3.** Melting temperatures ($T_m$) determined from UV thermal denaturation experiments with different concentrations of RNA.
3.3 Characterization of GQ and TH in the NEAT1 Triple Helix long sequence

After successfully confirming the presence of a GQ in the NEAT1 GQ5 WT sequence, we set out to determine if the presence of a TH at the 3’ end of NEAT1_2 would affect the formation of the GQ. Using in vitro T7 RNA polymerase driven transcription, the RNA sequence NEAT1 TH long (Table 2), was produced. This sequence contains the GQ5 region of NEAT1_2 (bold Gs in Table 2) as well as the terminal 3’ triple helix region (highlighted blue and pink in Table 2).

Similar to the NEAT1 GQ5 WT sequence, non-denaturing PAGE was performed in the presence of varying amounts of KCl for the NEAT1 TH long sequence. Figure 15A shows the results of this experiment when the gel was visualized by UV shadowing and Figure 15B shows the same gel visualized upon staining in NMM. Under UV shadowing, bands are present in each lane, whereas upon NMM staining the same bands remain visible in each NEAT1 TH long lane, at each KCl concentration and even without KCl present. As expected, the positive GQ+ control stains in NMM, whereas the negative GQ- control is observable only by UV shadowing and not by NMM staining. Similar to what was observed with the NEAT1 GQ5 WT sequence, a very stable GQ forms in NEAT1 TH long sequence, apparent even in the absence of KCl.

![Figure 15](image.png)

**Figure 15.** 15% non-denaturing KCl dependence PAGE for NEAT1 TH long sequence visualized under UV shadowing (A) and stained in NMM dye (B)
CD spectroscopy was also performed to determine if the GQ present was parallel or antiparallel. Figure 16 shows the overlaid CD spectra obtained for the NEAT1 TH long sequence at varying concentrations of KCl. The spectra show intense positive bands around 265 nm and negative bands around 240 nm, similar to data obtained for the NEAT1 GQ5 WT sequence, which represent bands for a parallel GQ structure, further validating that there is a parallel GQ forming in the NEAT1 TH long sequence. However, for this sequence we can also determine that KCl is a stabilizing factor for the GQ structure, as there is a noticeable increase in the intensity of the band at 265 nm upon the addition of KCl. Also present in these spectra is a strong negative band around 210 nm, indicative of the presence of a hairpin structure in the sequence. This is unsurprising, given the expected structure of the NEAT1 TH long sequence (Figure 18B). These data indicate the presence of both GQ and hairpin structures in the NEAT1 TH sequence.

![Figure 16](image_url)

**Figure 16.** CD spectroscopy results for the NEAT1 TH long sequence with various concentrations of KCl.
$^1$H NMR spectroscopy experiments were performed next, monitoring the imino proton resonance region changes upon the addition of increasing KCl concentrations. As seen in Figure 17A, there are resonances present in the 10-12 ppm region, but unlike the resonances observed in the same region for the isolated NEAT1 GQ5 WT (Figure 11), here we observe both a subset of broad resonances between 10-10.9 ppm and a subset of sharp resonances between 10.9 -11.9 ppm at 0 mM KCl. Sharp imino proton resonances in the region 10-12 ppm have been previously assigned to G imino protons involved in G-U base pairs (Figure 17C) and the presence of such resonances in this region is consistent with the predicted formation of multiple G-U base pairs by the NEAT1 TH long sequence (Figure 18A, Us shown in blue) (Glemarec et al. 1996). Additionally, for this sequence there are also intense resonances in the Watson-Crick hydrogen bonding region between 12-14.5 ppm, indicating the presence of hairpin structures, results of which are consistent with the CD experiments. The region 12-13 ppm also contains signals from the U imino protons involved in G-U base pairs, since a G-U base pair gives rise to two imino proton resonances, one from the G imino and one from the U imino (Glemarec et al. 1996). These resonances could originate from stem-loop structures predicted in this sequence (Figure 18).

We noted that the sharp resonances in the 10.9-11.9 ppm region decrease in intensity with the addition of KCl, with the concomitant apparition of broad resonances in the same region. Additionally, we observed the disappearance of sharper resonances in the region 12-13 ppm with the addition of KCl. Thus, we attribute these spectral changes to a structural rearrangement in the TH long sequence, where upon with the addition of KCl, the G-U base pairs are disrupted and the G bases become involved in the formation of the GQ structure which is stabilized by the K$^+$ ions. This structural rearrangement does not affect the conserved stem-loop shown in purple in Figures 18 A and B, which is predicted to remain the same in the presence or absence of the GQ structure.
Figure 17. (A) $^1$H NMR spectra for NEAT1 TH long with increasing concentrations of KCl. (B) Overlay of spectra in A at different KCl concentrations. (C) G-U noncanonical base pair.
We did not observe clear resonances originating from the formation of additional A-U base pairs predicted to be engaged in the formation of the triple helix structure (pink and blue in Figure 18B). Being at the end of the sequence, this region is likely very dynamic experiencing “fraying” and giving rise to very broad resonances. We note the presence of very weak and broad resonances between 14-14.5 ppm, almost at the noise level for this sample whose concentration was not optimal for NMR studies (74 µM), which could originate from the A-U base pairs of the triple helix.

We hypothesize that the GQ structure promotes the formation of the TH structure; if the Gs upstream of the TH forming sequence are not engaged in GQ formation, they are predicted by RNA Structure (Reuter and Mathews 2010) to form base pairs with the Us typically involved in the formation of the TH structure. Figure 18B shows a prediction of the structure of the NEAT1 TH long sequence if both GQ and TH were able to form. If the GQ is able to form, there are no free guanine residues available to pair with uracil residues of the U-rich motif, and this allows the TH formation. To the same end, if the U-rich motifs involved in TH stabilization can pair with the A-rich tail, then there are no available uracil residues to pair with guanine residues in the GQ region. From just the \(^1\)H NMR spectra, however, we cannot ascertain whether or not a TH structure is present. To determine the presence of TH, UV thermal denaturation experiments were utilized.

Literature has reported that triple helix sequences will denature in a double transition manner, with the first (lower T\(_m\)) transition representing the dissociation of the third strand from the helix, and the second (higher T\(_m\)) transition representing the dissociation of the rest of the double-stranded RNA (Gondeau et al. 1998). Because the triple helix long sequence contains a characterized GQ (as proven by NMM stained gels, CD and NMR spectroscopy data) and potential TH, UV thermal denaturation experiments were expected to yield curves with two
hyperchromic transitions and one hypochromic transition. The two hyperchromic transitions are predicted to correspond to the dissociation of the TH third strand followed by the dissociation of the remaining duplex whereas the hypochromic transition is predicted to correspond to the dissociation of the GQ structure.

UV thermal denaturation experiments were performed at two wavelengths; for GQ detection, absorbance was recorded at 295 nm (Mergny et al. 1998) and for TH detection, absorbance was recorded at 275 nm. At 275 nm, we observe a double hyperchromic transition indicative of triple helix dissociation (Figure 19), with the $T_m$ for the lower transition 46°C, and for the higher transition, 74°C. At 295 nm, there is a single hypochromic transition indicative of GQ denaturation with a $T_m$ of 80°C. This agrees with the $T_m$ from the previous experiment with NEAT1 GQ5 WT of 79°C. These data confirm the presence of both GQ and TH structures in the NEAT1 TH long sequence.
3.4 Determining the influence of GQ on TH formation

Following the experiments with the NEAT1 GQ5 WT sequence and the NEAT1 TH long sequence, we decided to further explore the influence of the GQ structure on the formation of the TH by performing various binding experiments. First, using in-house transcription protocols, three separate RNA sequences were transcribed (Table 2). The first sequence, called NEAT1 TH short, contained only the U-rich motif regions involved in forming the TH structure and did not include the GQ5 sequence or the A-rich tail. Because this sequence excludes the upstream Gs which would compete for binding with the Us involved in TH, we believe it is able to easily form the TH structure upon addition of the short A-rich tail sequence. The second RNA, called NEAT1 TH

Figure 19. UV thermal denaturation of NEAT1 TH long at 275 nm (orange) and 295 nm (blue) at 5 mM KCl.
short WT GQ5, contains the GQ5 region, but not the A-rich tail, thus this sample will be able to form a TH only upon the addition of the short A-tail (Wilusz, et al. 2012, Kunkler et al. 2019). Finally, the third, called NEAT1 TH short GQ5 mut, contained a mutated GQ5 region (using the same mutations that we proved previously disrupt the GQ structure formation, Table 2, row 2), but not the A-rich tail. This sample will be used to determine if indeed the GQ structure formation is required for the TH formation. NEAT1 TH short GQ5 mut has the same U-rich motif sequence as the NEAT1 TH short, however, it contains an upstream mutated sequence that cannot form a GQ structure, but retains the ability to base pair with the Us normally involved in TH formation (Figure 20). This sequence contains a competing upstream sequence that cannot form a GQ, revealing directly if the absence of the GQ impacts TH formation. The A-rich tail RNA sequences were purchased from Dharmacon.

**Figure 20.** Predicted structure of NEAT1 TH short GQ5 mut sequence. The mutated GQ5 region (red outlined circles, mutated bases are red letters) cannot form a GQ and therefore allows G-U base pairs to form within the U-rich motifs (blue circles). The A-rich tail (pink circles) do not interact with the U-rich motifs to form the TH. Conserved stem-loop region highlighted in purple. The G-U base pairs are indicated by stars.
Before performing binding experiments, we first needed to characterize the presence or absence of GQ structure formation in the sequences. Similar to the previous characterization methods used, here, KCl dependence PAGE stained in NMM elucidates the presence of GQ in the NEAT1 TH short WT GQ5 sequence, while, as expected, there is no GQ present in the NEAT1 TH short GQ5 mut sequence. The results of the KCl dependence PAGE experiments (Figure 21) show that RNA bands are present for both sequences when visualized by UV shadowing (Figure 21A and C), but once stained in NMM, the only bands visible are those corresponding to the NEAT1 TH short WT GQ5 gel and to the positive control GQ+ sequence (Figure 21B and D). These data support that GQ is forming in the NEAT1 TH short WT GQ5 sequence but not in TH short GQ5 mut when the sequence is mutated.

**Figure 21.** 15% non-denaturing KCl dependence PAGE for NEAT1 TH short WT GQ5 and NEAT1 TH short GQ5 mut with UV shadowing (A) and (B) and stained in NMM (C) and (D)

Further characterization of the GQ was completed with \(^1\)H NMR spectroscopy experiments. KCl was titrated into samples from 0 mM to 150 mM and the resulting spectra are shown in Figure 22 A and B. Similar to the NEAT1 TH long sequence, the resonances in the region 10-12 ppm are divided into broad resonances from 10-10.9 ppm and sharper resonances
from 10.9-11.9 ppm at 0 mM KCl. Also, similarly to what was observed for the NEAT1 TH long sequence, the sharper resonances in the region 10.9-11.9 ppm of NEAT1 TH short WT GQ5 sequence disappear with the concomitant appearance of broad resonances as KCl is titrated in the sample. This indicates the disruption of the competing G-U base pairs as the Gs become engaged in GQ structure formation which is stabilized by K$^+$ ions. A direct comparison of the imino proton resonance spectra of NEAT TH long, NEAT TH short WT GQ5 and of the isolated NEAT1 GQ5 in 150 mM KCl is shown in Figure 23. Most of the resonances in the Watson-Crick region appear at the same chemical shift and also some of the broad resonances corresponding to the GQ structure overlap, indicating that the stem-loop structures are similar in NEAT TH long, NEAT TH short WT GQ5, and also that the GQ structure formed by the isolated sequence is similar to that formed

![Figure 22. $^1$H NMR for NEAT1 TH short WT GQ5 sequence in the presence of various concentrations of KCl (A) and the overlay of those spectra (B).](image)
in NEAT TH long, NEAT TH short WT GQ5. While is possible that the very broad resonances (indicated by stars in Figure 23) seen in NEAT TH long and absent in TH short WT GQ5 (which lacks the A tail) originate from the A-U base pairs formed in the TH, they are almost at the noise level and we are not confident in these assignments.

Thus, in order to unambiguously determine whether or not a TH forms when A-rich tail is added to this sequence, various binding experiments were performed and the formation of the TH was observed by UV thermal denaturation experiments, monitoring the absorbance changes at 260 nm. Each sample was prepared in the presence of 50 mM KCl, 1 mM MgCl₂ (Kunkler, et al. 2019), and a 1:1 ratio of A-tail:TH sequence was added to each sample. The TH short sequence (pink line on the graph in Figure 24) is expected to act as a positive control, forming a TH sequence, and indeed it reveals a dual hyperchromic transition. The lower transition has a $T_m$ of 58°C, while the higher transition has a $T_m$ of 87°C. The TH short WT GQ5 sequence (orange line on the graph in Figure 24) reveals a dual hyperchromic transition as well. This NEAT1 sequence contains the
GQ5 region previously characterized, as well as the U-rich motifs necessary for the formation of the TH. Upon addition of the A-tail sequence, we see from the graph in Figure 24 that a triple helix forms, as the lower transition with $T_m$ of 57°C indicates that the third strand of the triple helix dissociates at around the same temperature as in the control NEAT1 TH short sequence. The $T_m$ of the higher transition is also around the same temperature as the $T_m$ for the higher transition of the TH short sequence. These results indicate that a TH forms when the GQ5 region is included at the 3’ end of NEAT1 TH short, as in the presence of KCl, the GQ structure is stabilized and the Gs will not be available for competitive base pairing with the Us involved in TH formation. Lastly, the TH short GQ5 mut sequence (blue line on the graph in Figure 24), which has a mutated GQ sequence but retains the competing upstream sequence that could bind the Us involved in TH formation, shows a single transition with a $T_m$ of 88°C. The absence of the lower $T_m$ transition in this sample clearly indicates that the TH structure does not form in the TH short GQ5 mut

**Figure 24.** UV thermal denaturation for NEAT1 TH short (pink), NEAT1 TH short WT GQ5 (orange), and NEAT1 TH short GQ5 mut (blue) at 260 nm.
sequence. Taken together, the results in the graph in Figure 24 indicate that when a competing sequence is present upstream, the TH structure formation is dependent on the GQ structure formation.

After determining the presence of the TH in the sequence via binding experiments monitored by UV spectroscopy, next, the TH formation was monitored by non-denaturing PAGE. The A-rich tail terminus sequence was fluorescently labeled with Dy547 at the 5’ end. Because the A-rich tail terminus sequence is very short, we are unable to capture the free RNA bands on the gel, while simultaneously capturing its complex formation with the TH forming sequences. Figure 25A shows the RNA sequences used in this experiment. First, the binding was assessed between the NEAT1 TH short sequence and the terminus Dy547-A-rich tail region that has been fluorescently labeled. The U-rich motifs of the TH short sequence is highlighted in blue, and the A-rich region of the terminus sequence is highlighted in pink in Figure 25A. These are the regions that are predicted to bind to one another to form the TH.

The non-denaturing PAGE results for this experiment are shown in Figure 25B and C. As seen in Figure 25B which shows the results of this gel after staining in SYBR Gold, as the A-rich tail is added to the samples, an upper band corresponding to the complex formed by NEAT1 TH short and the A-rich tail becomes more intense, while the bottom band corresponding to the free NEAT1 TH short becomes fainter and almost disappears. Figure 25C depicts the same gel, but the image was taken with a LICOR Odyssey Fc fluorescent imager, visualizing the fluorescence of the A-rich tail. This image shows one fluorescent band that increases in intensity from left to right as the concentration of A-rich tail increases, with no fluorescence signal in the lane containing free TH short sequence. The fluorescent bands in Figure 25C correspond perfectly to the upper
The information gained from this positive control experiment helps in determining how the TH is formed. If the A-rich tail is binding to the NEAT1 TH short sequence, we expect to see a shift in the position of the band on the gel. Seeing the formation of this complex band indicates that binding between the two is occurring, and this is further verified with the fluorescent gel. This experiment was replicated with both the NEAT1 TH short WT GQ5 and NEAT1 TH short GQ5
mut sequences and the results from these binding experiments are shown in Figure 26. The gel in Figure 26A shows the results of the NEAT1 TH short WT GQ5 and A-rich tail binding. We observe in each lane a distinct band, but unlike the gel in Figure 25B, there is not a second visible band that appears upon the addition of the A-rich tail. Since the NEAT1 TH short WT GQ5 sequence is much larger than the NEAT1 TH short sequence, the formation of a complex upon the addition of the A-rich tail cannot be monitored by a clear shift of the complex band. To determine whether the A-rich tail binds to the sequence, it was visualized using a LICOR Odyssey Fc imager, visualizing the A tail fluorescence, the results of which are in Figure 26B. Although a distinct band showing the complex between NEAT1 TH short WT GQ5 and A-rich tail is not seen with SYBR Gold dye, when visualized using fluorescence, bands appear across the gel in the lanes containing the A-rich tail, increasing in intensity as A-rich tail was added in higher concentrations. This gel also contained a positive control sample in which the A tail was added in a 1:1 ratio with the TH short sequence, and this lane clearly shows the formation of the TH structure due to its fluorescence.

Conversely, the binding experiment results with the NEAT1 TH short GQ5 mut sequence show a different result. The gel stained in SYBR Gold (Figure 26C) shows a distinct band in each lane across the gel as before; however, when this gel was visualized using fluorescent imaging, the only band that appears is that in the control lane containing the TH short-A-tail complex (Figure 26D). This result clearly shows that when the GQ sequence is mutated and the GQ cannot form, the presence of the upstream competing sequence prevents the formation of the TH structure. This result confirms the UV thermal denaturation result (Figure 20, blue line) which showed that a TH does not form upon the addition of the A tail to the NEAT1 TH short GQ5 mut sequence.
3.5 Interaction of NEAT1_2 GQ5 with FUS

One final part of our hypothesis deals with the interaction of NEAT1_2 GQ5 with the FUS RGG3 domain. To investigate the binding, we performed native PAGE experiments with each of the three RGG domains of FUS. FUS RGG3 has been shown to bind GQs, but not FUS RGG1 or RGG2 (Takahama et al. 2013). Figure 27 shows the results of the PAGE binding experiments with FUS. Figure 27A shows results for the binding experiments with FUS RGG1 and RGG2, where there is one distinct band visible in each lane. We see no band shifting or smearing of the bands which would typically imply binding. On the contrary, in Figure 27B, we see binding
between NEAT1 GQ5 and FUS RGG 3. This is made evident because of the change in position of bands, as well as a smearing of the bands in each lane. The first lane in Figure 27B is the free RNA NEAT1 GQ5 WT and as FUS RGG3 is added in increasing concentrations, we see a definitive shift in the bands towards the top of the gel. This implicates a complex that is forming between the two, as the larger complex will not travel as far on the gel. At very high RGG3 concentrations, the complex is large enough that it does not even enter the gel, which is why the bands in the 3:1 and 4:1 lanes, for instance, appear much lighter than the lower concentrations.

This result confirms that the GQ forming in NEAT1 GQ5 WT is binding with FUS RGG3 which is in agreement with literature showing that FUS can bind to GQ structures (Takahama et al. 2013, Imperatore et al. 2020). This finding is particularly important in understanding how the

**Figure 27.** 20% non-denaturing PAGE binding gels stained in NMM with NEAT1 GQ5 WT and FUS RGG domains. (A) NEAT1 GQ5 shows no binding to FUS RGG1 or RGG2. (B) NEAT1 GQ5 binds to FUS RGG3 as concentration increases. Experiments were performed at 5 mM KCl.
two interact with respect to paraspeckle formation and structure, since both NEAT1_2 and FUS are essential components of paraspeckles.
Chapter 4: Conclusions and Future Work

The results presented in this dissertation elucidate novel information about IncRNA NEAT1 and the importance of secondary structure for the formation of paraspeckles and the onset and/or progression of neurodegenerative diseases like ALS. We predicted five potential GQ regions in the NEAT1_2 isoform, four of which were common to both NEAT1_1 and NEAT1_2, and one that was exclusive to NEAT1_2 near the 3’ terminus, “GQ5”. This region was of particular interest because of the indispensable nature of NEAT1_2 to the formation of paraspeckles (Nakagawa et al. 2011). Using several biophysical techniques, such as KCl dependent PAGE, we successfully characterized the formation of a stable GQ at the GQ5 region within the NEAT1 GQ5 WT sequence, supporting our original hypothesis. We also verified this result using 1H NMR spectroscopy, where we observed resonances within imino proton region corresponding to guanines involved in Hoogsteen hydrogen bonding. In 2020, Simko et al. published work that characterized the same five regions we proposed; their work also supports the formation of GQ in the GQ5 region.

After successfully showing the formation of a stable GQ, we were able to determine the structure and topology of this GQ. Using CD spectroscopy, we were able to determine that the GQ has parallel topology, indicated by signature bands at 265 nm (positive) and 240 nm (negative) on the spectra (Ranjibar 2009). CD spectroscopy also confirmed that the GQ is very stable, as intensity of the bands changed only slightly as K+, a GQ stabilizing ion, was added. Further, we determined that the GQ was intramolecular using UV spectroscopy thermal denaturation techniques. By varying the concentration of RNA in each experiment while keeping the other conditions the same, we were able to determine several melting temperatures (Tm) for the sequence
(Equation 1). For intramolecular GQs, $T_m$ is independent of RNA concentration (Equation 2), and we can determine from our experiments that the GQ5 region forms an intramolecular structure.

We then investigated the formation of GQ in the presence of the 3’ TH region using the NEAT1 TH long sequence. First, we were able to characterize the formation of GQ in this sequence similarly to the NEAT1 GQ5 WT sequence. Both KCl dependent PAGE and CD spectroscopy showed that a stable parallel GQ was forming, even in the absence of KCl. This also supports our original prediction that GQ will form in this region, but we also needed to account for the presence of the TH. $^1$H NMR data for this sequence shows resonances within both the imino proton region corresponding to Hoogsteen base pairs, as well as in the imino proton region corresponding to Watson-Crick base pairs. To determine if the resonances were a result of solely stem-loop interactions, or if they were also a result of TH interactions, we utilized UV spectroscopy. Literature has shown that TH structures denature in a biphasic manner (Gondeau et al. 1998). Our study shows that the NEAT1 TH long sequence denatures in a biphasic manner, indicating the presence of TH.

Future studies for this project include investigation into the characterization for the other four proposed GQ regions of NEAT1. Although these regions were shown to form GQ by Simko et al., extensive characterization can be done to determine topology of the GQs, as well as inter- or intramolecular nature. Although only NEAT1_2 is essential for paraspeckle formation, NEAT1_1 is still found within paraspeckles. By characterizing the other potential GQ regions of NEAT1, we can better understand interactions between NEAT1 and PSPs like FUS. These characterization studies are being carried out by other graduate and undergraduate members of the lab.
Next we investigated binding of the GQ5 region to FUS, an essential PSP (Shelkovnikova et al. 2013, Hennig et al. 2015). Previous work has shown that FUS binds to GQ structures (Takahama and Oyoshi 2013, Ozdilek et al. 2017, Yagi et al. 2018), like that which we characterized in our NEAT1 sequences. With PAGE experiments, we were able to determine that FUS binding does occur at the GQ5 region of NEAT1_2. Specifically, we identified FUS binding at its RGG3 region, a region previously shown to bind DNA and RNA secondary structures (Takahama et al. 2013). Conversely, PAGE experiments performed with FUS RGG1 and RGG2 showed no binding. This finding supports the literature documenting FUS RGG3 binding to GQs. Future work characterizing the GQs in NEAT1 will correspond with future work involving FUS binding studies. Investigating the interaction of an essential PSP with an essential paraspeckle RNA can lead to the better understanding of how these two interact with one another to allow the formation of paraspeckles and will contribute greatly to the field of understanding the pathology of neurodegenerative diseases like ALS. Additionally, further studies need to be performed to elucidate the mechanism of FUS RGG3 binding to the GQ regions of FUS.

The final facet to this research involved examining the effect of GQ5 on the 3’ terminal TH. Our work had already shown the formation of a stable GQ in both the NEAT1 GQ5 WT and NEAT1 TH long sequence. What we sought to discover next was how these two secondary structures affected one another; that is, was the TH able to form without the GQ or vice-versa. We proposed that, without GQ formation, the guanine residues would be free to form noncanonical G-U base pairs with the uracil residues of the U-rich motifs typically involved with the TH structure, thus interfering with TH formation. With a series of shortened NEAT1 sequences, we were able to show for the first time that the GQ structure plays a role in modulating the 3’ TH formation of
NEAT1_2. Since the triple helix is related to stability of NEAT1_2, it follows that the GQ forming in the GQ5 region also plays a role in stability.

Each shortened sequence excluded the A-rich tail region that is responsible for creating the third base in the base triples of the TH, while one excluded both the A-rich tail and the GQ5 regions. We showed using PAGE that even without the A-rich tail region, GQ was still able to form in the NEAT1 TH short WT GQ5 sequence, but not in its mutated counterpart, NEAT1 TH short GQ5 mut. This confirms that the GQ formation is not dependent on presence of the TH. However, when the GQ is disrupted, we did not see formation of the TH. We showed using UV spectroscopy and PAGE with a fluorescently labeled A-rich tail that when GQ was intact, the TH was able to form, but when the GQ region was mutated, TH was not able to form. These findings support our original hypothesis that the TH and GQ regions at the 3’ end of NEAT1_2 affect one another, specifically, that TH formation is dependent upon the presence of the GQ.

The work presented here provides interesting insight into NEAT1_2 and how it can be used as a potential drug target for ALS. Literature has shown that paraspeckles and NEAT1 are overexpressed in individuals who are affected by ALS (Nishimoto et al. 2013, An et al. 2019). Since the TH at the 3’ end of NEAT1_2 is thought to affect the stability of the molecule and disrupting the GQ adjacent to it can disrupt the formation of the TH, we can use the GQ as a potential drug target. PNAs have been shown to bind GQ structures (Panyutin 2012), making them attractive tools for potential therapeutics. Disrupting the GQ will prevent TH formation, thus allowing the 3’ end of NEAT1_2 to be degraded which can counteract the overexpression of the molecule in ALS. Figure 20 shows the proposed structure when the GQ region is mutated. The same effect can be achieved if a PNA targets the first 11 nts of the sequence. The binding of the
PNA in this region would be enough to disrupt the GQ, but still allow the G-U base pairs to form like those shown with stars in Figure 20, thus preventing the formation of the 3’ TH.

Further work in this area includes investigating the role of the TH in the stability and protection of the 3’ end of NEAT1_2. Degradation studies using exonucleases, like those performed in Donlic et al. 2020 on the TH region of MALAT1, would allow better understanding of how the TH and GQ5 work together to provide structural integrity for NEAT1_2 at the 3’ terminus. Forming base triples effectively protects the A-rich tail from being easily degraded. However, should the TH be degraded, the GQ located in the GQ5 region may act as another hindrance to exonucleases.

Overall the work presented here shows novel importance of the TH and GQ5 regions at the 3’ end of NEAT1_2 and that the TH may rely on the presence of GQ5 to form. This work contributes to the greater field of study because it allows better understanding of the lncRNA that forms the scaffold of paraspeckles. Both NEAT1 and paraspeckles are implicated in the onset of ALS, a progressive neurodegenerative disorder. Understanding the structure and interactions between NEAT1 and FUS can lead to the potential development of therapeutics and possibly PNAs to treat ALS and other neurodegenerative disorders.
Chapter 5: References


*BMC Cell Biol.* **2008**, *9*, 37.


Bugaut, A. Balasubramanian, S. 5’-UTR RNA GQs: translation, regulation, and targeting.


Cabili, M. N.; Trapnell, C.; Goff, L.; Koziol, M.; Tazon-Vega, B.; Regev, A.; Rinn, J. L.


Boguski, M. S.; Roe, B. A.; Collins, F. S.; Marx, S. J.; Burns, L.; Chandrasekharappa, S.
C. A transcript map for the 2.8-Mb region containing the multiple endocrine neoplasia

Hendry, P.; Hannan, G. Detection and quantitation of unlabeled nucleic acids in polyacrylamide

binding proteins are essential for building subnuclear paraspeckles. *J. Cell Biol.* **2015**, 
*210*, 529-539.

Ince, P. G.; McDermott, C. J.; Shaw, P. J. Novel FUS/TLS mutations and pathology in

Borkhardt, A.; Sander, C.; Tuschl, T. RNA targets of wild type and mutant FET family

Holley, R. W.; Apgar, J.; Everett, G. A.; Madison, J. T.; Marquisee, M.; Merrill, S. H.;

Identification of G-quadruplex-binding protein from the exploration of RGG motif/G-


Chapter 6: Appendix

6.1 Introduction

Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are neurodegenerative disorders that belong to the same spectrum consisting of similar symptoms and pathology. ALS is a fatal neurodegenerative disorder characterized by loss of motor neurons in the brain and spinal cord (Haverkamp, et al. 1995, Cleveland and Rothstein 2011). FTD is the second most common form of dementia next to Alzheimer’s disease, leading to degeneration of frontal and anterior lobes of the brain (Neary et al. 1998). These disorders are caused by large expansions of the hexanucleotide repeat GGGGCC (G₄C₂) in the 5’ region of intron 1 on the C9ORF72 gene (DeJesus-Hernandez et al. 2011, Renton et al. 2011, Fratta et al. 2012).

Healthy, unaffected individuals generally have less than 30 of these repeats, while those affected by disease have more than 250 repeats and can even have upwards of 700-1600 repeats (Gomez-Tortosa et al. 2013). A hallmark of a diseased brain is the aggregation of mRNA transcripts that are expressed from the expanded repeats (DeJesus-Hernandez et al. 2011, Lee et al. 2013). About 90% of ALS cases are sporadic (sALS) and at least 8% of sALS patients exhibit these G₄C₂ repeats; conversely, about 10% of ALS cases are familial (fALS) and at least 40% of fALS patients exhibit these G₄C₂ repeats (DeJesus-Hernandez et al. 2011, Renton et al. 2011, Majounie et al. 2012), making these hexanucleotide repeat regions exciting targets for potential treatment.

One potential approach to target this RNA is to use peptide nucleic acids (PNAs). PNAs are manufactured biomolecular tools that can be used as probes and antigene agents (Pellestor and Paulasova 2004). PNAs mimic a protein backbone but have a nucleic acid base in place of an amino acid sidechain (Figure 28) (Nielsen 1999). Because PNAs have properties of both protein
and nucleic acid, they are protected from proteases and nucleases and can bind to DNA or RNA via traditional Watson-Crick base pairing (Figure 28). They have neutral backbones which allow them to penetrate DNA and/or RNA duplexes with no electrostatic repulsion (Neilson 2001). Another potential target of PNAs is the G quadruplex (GQ) structure (Panyutin 2012). Due to the highly G-rich tracts of the G₄C₂ hexanucleotide expansion regions, these regions are favorable for GQ formation (Cammas and Millevoi 2016).

*We hypothesize that the G₄C₂ hexanucleotide expansion region will form a GQ structure.* We will investigate this with a long (G₄C₂)₄₅ sequence and characterizing it using biophysical techniques. We chose this length because 45 repeats are relevant disease length. We also hypothesize that PNA binding will occur at the GQ region with a complementary (C₄G₂) sequence. Binding studies with a synthesized PNA and (G₄C₂)₄₅ RNA will reveal if binding is possible, lending insight into the viability of PNAs as a potential therapeutic for diseased individuals.
6.2 Materials and Methods

6.2.1 Cell Transformation for (G₄C₂)₄₅

Plasmids encoding for the (G₄C₂) hexanucleotide expansion, 45 repeat region were provided by our collaborators. Cell transformation was performed using Stbl3 cells. All media used for cell growth was Luria-Bertani (LB; Fisher Scientific) with 200 µg/mL ampicillin (AMP). Cultures were incubated overnight at 37°C. A single colony from the growth plates was placed in 250 mL of LB broth and 1 mL AMP and incubated overnight at 37°C with shaking at 225 RPM.

Plasmids were subsequently extracted using E. Z. N. A. Endo-free Plasmid Extraction MiniKit (Omega) and the protocol therein. After extraction, plasmids were digested using Buffer 3.1 and restriction enzyme (EcoRV). In an Eppendorf tube, 5 µL Buffer 3.1, 39 µL dH₂O, 3 µL DNA plasmid, and 3 µL EcoRV were incubated at 37°C for 2.5 hours. Transcription was carried out using the mMessage mMachine T7 Megascript Kit and the protocol therein. An Eppendorf tube containing 8 µL of Nucleotide mix, 2 µL of GTP, 2 µL of 10X Reaction Buffer, 5 µL of Digested DNA template, 2 µL of T7, and 3 µL of nuclease free H₂O was incubated for 6 hours at 37°C. After incubation, 1 µL of TurboDNAse was added, followed by an additional 15 minutes of incubation at 37°C. Samples were stored at -20°C.

6.2.2 Agarose Gel Preparation

For the (G₄C₂)₄₅ sequences, 1% agarose gels were utilized for separation. Gels were prepared using agar dissolved upon heating in a 1/2X TBE buffer solution. RNA samples were prepared at appropriate concentrations and loaded onto the gel against a 1kb+ ladder (Invitrogen) for reference. Agarose gels were run at room temperature for approximately 1.5 hours at 80 V. Gels were initially visualized with UV-shadowing at 254 nm using an AlphaImager.
(AlphaInnotech) and subsequently stained in ethidium bromide solution for 30 minutes. For PNA binding studies, all conditions remained the same, except 0.8% agarose gel was used.

6.2.3 Phenol Chloroform Extraction

Agarose gel experiments were run as described and upon completion, the bands of interest were carefully cut from the gel with little excess agarose. Gel pieces were crushed inside of an Eppendorf tube and placed in a 65°C water bath until the agarose was melted. Phenol was added in equal volume to that of the melted gel. The same was centrifuged for 10 minutes at 13,000 RPM. After centrifuging, the top (aqueous) layer was carefully pipetted off and placed in a new Eppendorf tube. To this layer was added an equal volume of 1:1 phenol:chloroform solution. The sample was vortexed to emulsion and then centrifuged for 10 minutes at 13,000 RPM. Again, the top (aqueous) layer was pipetted off and placed in a new Eppendorf tube. Sodium acetate (pH 5.5) was added to a volume of 1/10, followed by 2 volumes of 100% ethanol. Invert to mix well for several seconds before centrifuging for 10 minutes at 13,000 RPM. Remove the ethanol layer and place in a new Eppendorf tube. RNA should be in this layer.

6.3 Results and Discussion

6.3.1 Transcription of (G₄C₂)₄₅ Expansion

After performing cell transformation with Stbl3 cells, circular plasmids were extracted and digested, leaving us with a sample for transcription. Upon completion of transcription, we are left with the (G₄C₂)₄₅ sample of interest. In order to determine whether each step of the process was successful, Agarose gel experiments were performed. Using a 1kB+ ladder as a reference, we are able to determine that the sample prepared is the (G₄C₂)₄₅ of interest – the sequence is 494 nucleotides in length, which corresponds well to the 500 nucleotide (nt) marker of the ladder (Figure 29). The transcription product of the cell transformation protocol, however, does not
produce a product that is pure (red circle, Figure 29). Lanes 4 and 7 of the agarose gel have two distinct RNA bands, one that corresponds closely to the 500 nt marker (that would be our band of interest), and one that is higher up the gel. In order to fully characterize the (G₄C₂)₄₅ sequence, we need a pure sample.

To obtain a pure product, phenol chloroform purification techniques were utilized. In doing so, the RNA can be extracted from crushed gel pieces, effectively isolating the band of interest. Through much trial and error, the phenol chloroform extraction process was able to yield a single band, pure product, circled in red in Figure 30. This band corresponds closely with the 500 nt marker in the 1 kB+ ladder and with the lower band of the impure product (Figure 30).
Although this result is promising as it indicates an effective way to purify the sample, the resultant band is faint, signifying a poor yield from the extraction procedure, and we have not been able to duplicate the result.

6.3.2 Determining G Quadruplex Formation in the (G₄C₂)₄₅ Sequence

The first goal of this project is to determine if the (G₄C₂)₄₅ sequence forms a GQ structure. Similar to techniques used for PAGE, agarose gels were stained in NMM to elucidate the presence of GQ structure. Figure 31A shows the results of staining a 1\% agarose gel in NMM – there are no bands present. While we would not expect to see staining of the ladder or plasmids in NMM (NMM is specific to GQ structures), if the transcribed sample forms a GQ, we should see staining. When the same gel was stained in ethidium bromide (Figure 31B), bands appear on the gel as would be expected. This begs the question, then, does the NMM fail to stain because there is no GQ, or does the NMM fail to stain because it cannot penetrate the thick agarose gel. Figure 32 shows two control gels, where the only samples used were ladder, a GQ+ control (BDNF), and a
GQ- control (pre-miR 149). If NMM can penetrate the agarose gel, the GQ+ band will be visible.

The gel in Figure 32A was stained in NMM and no bands were visible upon visualization, suggesting that NMM cannot penetrate the gel to stain the RNA. When the same gel was stained
in EtBr, bands in each lane were visible (Figure 32B). Due to the inability to use NMM dye to determine presence of GQ in gels, the results for this part of the experiment remain inconclusive.

### 6.3.3 Binding of Peptide Nucleic Acid to the (G₄C₂)₄₅ Sequence

The second goal of this project is to determine whether PNAs will bind to the sequence. PNAs can bind to DNA or RNA via traditional Watson-Crick base pairing and they also have the ability to bind to GQ structures with the potential to either disrupt or stabilize the structure. To determine if PNAs would bind to the (G₄C₂)₄₅ sequence, binding studies were performed and visualized using agarose gels. The PNA used in the study, Lys(C₄G₂)₂, is complementary to the G₄C₂ hexanucleotide repeat. The results of the binding experiments are shown in Figure 33.

![Figure 33](image)

**Figure 33.** 0.8% agarose gel showing results from PNA binding experiments. The free RNA band (red circle) corresponds closely to the 500 nt marker in the ladder. Upon addition of PNA, a significant shift in bands upwards on the gel is seen (yellow box), indicating binding. The same experiment with the addition of KCl produces faint, streaked bands (blue box).

PNA was added to samples from 10:1 – 30:1, and in each case, binding was observed. Figure 33 shows the free RNA circled in red; when compared to the lanes with PNA added, a significant shift in the bands is observed (yellow box, Figure 33). When KCl was added to
samples, binding was also observed, but in this case, the resulting bands were faint and streaked on the gel. A potential reason for the bands being faint or non-visible is that, once binding occurs, the complex is too big to enter the gel. The findings from this experiment indicate that PNA binding does occur at the G₄C₂ hexanucleotide repeat region and that they can be used as potential therapeutics for disease.

6.4 Conclusions and Future Work

The findings reported in this research indicate that a long (G₄C₂)₄₅ hexanucleotide repeat region can be successfully transcribed in house. Though the initial transcription product is impure, using a phenol chloroform extraction technique, we were able to purify the sample, although to a low yield. Results for the formation of GQ are inconclusive from gel studies alone. Because NMM cannot penetrate the thick agarose gels, it cannot be used as a reliable tool for determining the presence of GQ in the (G₄C₂)₄₅ sequence. Finally, PNA binding studies show promising results indicating these manufactured biomolecules can be used to bind potential GQ regions in the diseased states of individuals with many G₄C₂ repeats. Future work for this project involves characterizing the potential GQ formation using atomic force microscopy (AFM). Purified samples will be sent to collaborators to use AFM, which will hopefully be able to ‘see’ the GQ structures due to the size of the (G₄C₂)₄₅ repeat. In house, we hope to perform 5% PAGE experiments which will allow visualization with NMM dye. Overall, the preliminary results discussed here show promise for PNAs that target G₄C₂ sequences and their use for potential therapeutics for disease.
6.5 References


