

**TRANSLATION MODULATION OF CELLULAR mRNA  
BY G-QUADRUPLEX STRUCTURES**

A dissertation submitted  
to the Department of Chemistry and Biochemistry  
at Kent State University in partial  
fulfillment of the requirements for  
the degree of Doctor of Philosophy

by

Debmalya Bhattacharyya

August 2016

© Copyright

All rights reserved

Except for previously published materials

**Dissertation written by**

Debmalya Bhattacharyya

B.Sc., Bangalore University, India, 2005

M.Tech., I.I.T Kharagpur, India, 2007

Ph.D., Kent State University, 2016

**Approved by**

Soumitra Basu, Ph.D., \_\_\_\_\_ Chair, Doctoral Dissertation Committee

Roger Gregory, Ph.D., \_\_\_\_\_ Members, Doctoral Dissertation Committee

Paul Sampson, Ph.D. \_\_\_\_\_ Members, Doctoral Dissertation Committee

William C. Merrick Ph.D. \_\_\_\_\_ Members, Doctoral Dissertation Committee

Gail Fraizer Ph.D. \_\_\_\_\_ Members, Doctoral Dissertation Committee

**Accepted by**

Michael Tubergen, Ph.D. \_\_\_\_\_ Chair, Dept. Chemistry and Biochemistry

James L. Blank, Ph.D., \_\_\_\_\_ Dean, College of Arts and Sciences

## TABLE OF CONTENTS

|                                                                       |              |
|-----------------------------------------------------------------------|--------------|
| <b>LIST OF FIGURES.....</b>                                           | <b>IX</b>    |
| <b>LIST OF TABLES.....</b>                                            | <b>XII</b>   |
| <b>LIST OF SCHEMES.....</b>                                           | <b>XIII</b>  |
| <b>LIST OF ABBREVIATIONS.....</b>                                     | <b>XIV</b>   |
| <b>DEDICATION.....</b>                                                | <b>XV</b>    |
| <b>ACKNOWLEDGEMENTS.....</b>                                          | <b>XVI</b>   |
| <b>ABSTRACT.....</b>                                                  | <b>XVIII</b> |
| <b>CHAPTER 1: <i>INTRODUCTION TO G-QUADRUPLEX STRUCTURES</i>.....</b> | <b>1</b>     |
| 1.1 Discovery of G-Quadruplexes.....                                  | 1            |
| 1.2 Structure of G-Quadruplexes.....                                  | 5            |
| 1.2.1 RNA G-Quadruplex structure.....                                 | 8            |
| 1.3 Functions of RNA G-Quadruplex structures.....                     | 11           |
| 1.3.1 Translational regulation.....                                   | 11           |
| 1.3.2 3' end processing and alternative polyadenylation.....          | 13           |
| 1.3.3 Mitochondrial transcription termination.....                    | 13           |
| 1.3.4 mRNA localization.....                                          | 14           |

|                                                                                                                     |                                                                                   |           |
|---------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|-----------|
| 1.3.5                                                                                                               | Alternative Splicing.....                                                         | 14        |
| 1.3.6                                                                                                               | Biogenesis of miRNAs.....                                                         | 15        |
| 1.4                                                                                                                 | Interaction of proteins with RNA G-quadruplex structures.....                     | 16        |
| 1.5                                                                                                                 | Targeting RNA GQ structures for therapeutics.....                                 | 18        |
| 1.6                                                                                                                 | Modulation of translation by GQ structures present in the 5'-UTR.....             | 22        |
| <b>CHAPTER 2: AN INDEPENDENTLY FOLDING RNA G-QUADRUPLEX DOMAIN<br/>DIRECTLY RECRUITS 40S RIBOSOMAL SUBUNIT.....</b> |                                                                                   | <b>27</b> |
| 2.1                                                                                                                 | Role of GQ structures in cap-independent translation initiation.....              | 27        |
| 2.2                                                                                                                 | Materials and methods.....                                                        | 32        |
| 2.2.1                                                                                                               | Plasmid constructions.....                                                        | 32        |
| 2.2.2                                                                                                               | <i>In vitro</i> transcription and radiolabeling of IRES A and mutant<br>RNAs..... | 33        |
| 2.2.3                                                                                                               | Structure mapping analyses.....                                                   | 34        |
| 2.2.4                                                                                                               | Filter binding assay.....                                                         | 35        |
| 2.2.5                                                                                                               | Footprinting of 40S ribosomal subunit interaction with the<br>IRESA.....          | 38        |
| 2.2.6                                                                                                               | Cell culture and transfection of reporter plasmid.....                            | 39        |

|       |                                                                                                           |    |
|-------|-----------------------------------------------------------------------------------------------------------|----|
| 2.3   | Results.....                                                                                              | 39 |
| 2.3.1 | Secondary structure mapping of the IRES A by enzymatic footprinting.....                                  | 39 |
| 2.3.2 | The 40S ribosomal subunit interacts with the 5'-proximal end of IRES A.....                               | 40 |
| 2.3.3 | The independently folding GQ domain in hVEGF IRES A determines the 40S ribosomal subunit recruitment..... | 45 |
| 2.3.4 | The quadruplex forming D2 domain is essential for optimal IRES A mediated translation initiation.....     | 54 |
| 2.4   | Discussion.....                                                                                           | 57 |

**CHAPTER 3: ENGINEERED DOMAIN SWAPPING INDICATES CONTEXT DEPENDENT FUNCTIONAL ROLE OF RNA G-QUADRUPLEXES.....61**

|       |                                                                         |    |
|-------|-------------------------------------------------------------------------|----|
| 3.1   | Domain engineering in biomolecules.....                                 | 61 |
| 3.2   | Materials and Methods.....                                              | 63 |
| 3.2.1 | Plasmid constructions.....                                              | 63 |
| 3.2.2 | In vitro transcription and radiolabeling of IRES A and mutant RNAs..... | 64 |
| 3.2.3 | Structure mapping of the IRES A by RNase T1 footprinting.....           | 65 |
| 3.2.4 | Filter binding assay.....                                               | 66 |

|                                                                  |                                                                            |    |
|------------------------------------------------------------------|----------------------------------------------------------------------------|----|
| 3.2.5                                                            | Reporter gene assay.....                                                   | 67 |
| 3.3                                                              | Results.....                                                               | 68 |
| 3.4                                                              | Discussion.....                                                            | 80 |
| <br>                                                             |                                                                            |    |
| <b>CHAPTER 4: RATIONALLY INDUCED RNA: DNA G-QUADRUPLEX</b>       |                                                                            |    |
| <b><i>STRUCTURES ELICIT ANTI-CANCER EFFECT BY INHIBITING</i></b> |                                                                            |    |
| <b><i>ENDOGENOUS EIF-4E EXPRESSION.....</i></b>                  |                                                                            |    |
| 4.1                                                              | Introduction.....                                                          | 82 |
| 4.2                                                              | Materials and methods.....                                                 | 85 |
| 4.2.1                                                            | In vitro transcription of a segment of 5'-UTR of human eIF4E<br>mRNA.....  | 86 |
| 4.2.2                                                            | DNA Oligonucleotides.....                                                  | 86 |
| 4.2.3                                                            | Folding of the oligonucleotides to form induced GQ.....                    | 87 |
| 4.2.4                                                            | Circular Dichroism (CD) spectroscopy.....                                  | 87 |
| 4.2.4.1                                                          | Analysis of the thermal melting curves obtained by CD<br>spectroscopy..... | 87 |
| 4.2.5                                                            | Isothermal titration calorimetry.....                                      | 90 |
| 4.2.6                                                            | Radiolabeling of RNA and DNA oligonucleotides.....                         | 91 |

|        |                                                                                                                           |     |
|--------|---------------------------------------------------------------------------------------------------------------------------|-----|
| 4.2.7  | Native gel electrophoretic mobility shift assay.....                                                                      | 91  |
| 4.2.8  | RNase T1 structure mapping.....                                                                                           | 92  |
| 4.2.9  | Dimethyl Sulfate (DMS) footprinting.....                                                                                  | 92  |
| 4.2.10 | Plasmid construction.....                                                                                                 | 93  |
| 4.2.11 | Reporter gene assay.....                                                                                                  | 93  |
| 4.2.12 | Quantitative RT-PCR.....                                                                                                  | 94  |
| 4.2.13 | Western blotting.....                                                                                                     | 95  |
| 4.2.14 | Cell viability assay.....                                                                                                 | 95  |
| 4.3    | Results.....                                                                                                              | 95  |
| 4.3.1  | RNA and DNA oligonucleotides associate to form a parallel induced hybrid GQ structure.....                                | 96  |
| 4.3.2  | Enzymatic and chemical footprinting data indicate the formation of an induced GQ and the duplex structures in tandem..... | 99  |
| 4.3.3  | The induced GQ structure is stable at physiological temperature and is resistant to nuclease digestion.....               | 101 |
| 4.3.4  | Induced GQ structures specifically downregulate the expression of a targeted gene in a reporter plasmid.....              | 105 |

|                                    |                                                                                                                        |            |
|------------------------------------|------------------------------------------------------------------------------------------------------------------------|------------|
| 4.3.5                              | Targeted induction of GQ structure downregulates the endogenous human eIF-4E expression by translation repression..... | 108        |
| 4.3.6                              | Targeting endogenous eIF-4E mRNA by GQ inducing sequences inhibits proliferation of human cancer cells.....            | 112        |
| 4.4                                | Discussion.....                                                                                                        | 115        |
| <b>CHAPTER 5: CONCLUSION.....</b>  |                                                                                                                        | <b>120</b> |
| <b>CHAPTER 6: APPENDIX I.....</b>  |                                                                                                                        | <b>123</b> |
| <b>CHAPTER 7: APPENDIX II.....</b> |                                                                                                                        | <b>124</b> |
| <b>REFERENCES.....</b>             |                                                                                                                        | <b>128</b> |

## LIST OF FIGURES

|                                                                                                    |    |
|----------------------------------------------------------------------------------------------------|----|
| <b>Figure 1.1:</b> Structure of the G-quartet.....                                                 | 5  |
| <b>Figure 1.2:</b> Glycosidic bond conformation.....                                               | 6  |
| <b>Figure 1.3:</b> Different topological variants of GQ structures.....                            | 8  |
| <b>Figure 1.4:</b> Functions of RNA GQ structures.....                                             | 12 |
| <b>Figure 1.5:</b> Ligands known to interact with RNA GQ structures.....                           | 20 |
| <b>Figure 1.6:</b> Role of GQ structures in the 5'-UTR.....                                        | 25 |
| <b>Figure 2.1:</b> The 5'-cap structure.....                                                       | 29 |
| <b>Figure 2.2:</b> The classification of viral IRESs.....                                          | 30 |
| <b>Figure 2.3:</b> Imaging of the filter binding assay.....                                        | 36 |
| <b>Figure 2.4:</b> Secondary structure of IRES A.....                                              | 41 |
| <b>Figure 2.5:</b> The IRES A interacts with 40S ribosomal subunit.....                            | 42 |
| <b>Figure 2.6:</b> Nomenclature of the domains in the IRES A.....                                  | 46 |
| <b>Figure 2.7:</b> The 40S ribosomal subunit binds to the GQ in IRES A.....                        | 47 |
| <b>Figure 2.8:</b> Structure mapping and ribosome footprinting of $\Delta$ D1 and $\Delta$ D3..... | 48 |

|                                                                                                       |     |
|-------------------------------------------------------------------------------------------------------|-----|
| <b>Figure 2.9:</b> Structure mapping and ribosome footprinting of $\Delta D2$ and $\Delta D123$ ..... | 51  |
| <b>Figure 2.10:</b> Structure mapping and ribosome binding of $\Delta D4$ and $\Delta D24$ .....      | 53  |
| <b>Figure 2.11:</b> Relative IRES activity of the mutants.....                                        | 56  |
| <b>Figure 3.1:</b> Contexts in which the domain swapping was performed.....                           | 71  |
| <b>Figure 3.2:</b> Structure mapping of IRES A with engineered GQ sequences.....                      | 72  |
| <b>Figure 3.3:</b> The IRES A with engineered GQ sequences binds to the 40S ribosomal<br>subunit..... | 76  |
| <b>Figure 3.4:</b> The IRES A activity with engineered GQ sequences.....                              | 77  |
| <b>Figure 3.5:</b> The engineered GQ sequences in 5'-UTR of MT3-MMP repressed<br>translation.....     | 79  |
| <b>Figure 4.1:</b> <i>In vitro</i> formation of induced GQ.....                                       | 97  |
| <b>Figure 4.2:</b> Enzymatic and chemical structure mapping of induced GQ.....                        | 100 |
| <b>Figure 4.3:</b> Stability of induced GQ.....                                                       | 103 |
| <b>Figure 4.4:</b> Calorimetric data for the titration of the formation of induced GQ.....            | 105 |
| <b>Figure 4.5:</b> The induced GQ is resistant to nucleases.....                                      | 106 |
| <b>Figure 4.6:</b> Induced GQ targeting eIF-4E downregulates reporter gene expression.....            | 109 |
| <b>Figure 4.7:</b> Induced GQ downregulates endogenous eIF-4E protein expression.....                 | 111 |

**Figure 4.8:** Induced GQ structures do not alter mRNA levels.....113

**Figure 4.9:** Anti-proliferative effect of induced GQ structures on cancer cells.....114

**Figure 7.1** The vant't Hoff plots.....124

**Figure 7.2:** Determination of the melting temperature by method 2.....126

## LIST OF TABLES

|                                                                                                                                                                   |     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>Table 3.1:</b> The four GQ forming motifs used in this study for RNA engineering and their sequence details and characteristics.....                           | 70  |
| <b>Table 4.1:</b> DNA oligonucleotide sequences used in the study.....                                                                                            | 86  |
| <b>Table 4.2:</b> Thermodynamic parameters of induced GQ formation.....                                                                                           | 104 |
| <b>Table 4.3:</b> The target sequences for inducing GQ by ON1 and ON2 in the 5'- UTR and the coding region and their position in the human eIF-4E transcript..... | 110 |
| <b>Table 6.1:</b> Comparison of the binding constants derived by two different methods.....                                                                       | 123 |
| <b>Table 7.1:</b> Thermodynamic parameters calculated by Method 1.....                                                                                            | 126 |
| <b>Table 7.2:</b> Thermodynamic parameters calculated by Method 2.....                                                                                            | 127 |
| <b>Table 7.3:</b> Comparison of the thermodynamic parameters.....                                                                                                 | 127 |

## LIST OF SCHEMES

|                                                                                                                                             |    |
|---------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Scheme 2.1:</b> Design of the hVEGFbicis.....                                                                                            | 33 |
| <b>Scheme 2. 2:</b> Arrangement of the membranes in the dot-blot apparatus.....                                                             | 36 |
| <b>Scheme 2.3:</b> Schematic representation of the interaction of GQ motif in IRES A with the<br>40S ribosomal subunit.....                 | 60 |
| <b>Scheme 3.1:</b> Schematic representation of engineered domain swapping.....                                                              | 81 |
| <b>Scheme 4.1:</b> Schematic representation of targeted induction of GQ in eIF-4E mRNA in<br>the 5'- UTR and the protein-coding region..... | 84 |

## **LIST OF ABBREVIATIONS**

**GQ:** G-quadruplex

**IRES:** Internal ribosome entry site

**UTR:** Untranslated region

**40S subunit:** 40S ribosomal subunit

**ITC:** Isothermal titration calorimetry

**PAGE:** Polyacrylamide gel electrophoresis

**CD:** Circular dichroism

**hVEGF:** Human vascular endothelial growth factor

**eIF-4E:** Eukaryotic initiation factor 4E

## **DEDICATION**

This dissertation is dedicated to my parents Nirmalya Kumar Bhattacharya and Swapna Bhattacharya who sacrificed immensely to ensure I get the very best of education and supported me throughout. I would also like to dedicate this to my wife Sudatta and my children Debasmita, Sharanya and Shriyan.

## ACKNOWLEDGEMENTS

I would first like to thank my research advisor Dr. Soumitra Basu, who has always been extremely supportive during my graduate studies. Thanks for providing me the freedom to pursue my own research ideas. I would also like to thank all the faculty members of the Dept. of Chemistry and Biochemistry who always encouraged me, helped me to endure tough times, and immensely contributed to my personal development. I thank you all for dedicating your lives to teach train and develop the future scientists. Special thanks to all my colleagues from the Basu Lab, Mohamed, Gayan, Nate, Prakash, Sumi, Nawar, Brad, Catherine for sharing in both the good times and the frustrating times. I will miss the summer cricket matches and our game of badminton which relieved a lot of stress. I would also like to thank former members of Basu Lab especially Eric, Abby, Mark, and Aparna. Abby and Eric saw me through the initial training in the Basu Lab and have been always there to help. Mark, you set very high standards in the lab and will be forever missed. I would also like to thank all the undergraduate students who have worked with me. Sarah, Justin, Kim and Paige thanks for putting up with my flighty ideas and ways. I loved working with you all.

Special thanks to my lovely wife, Sudatta, who gave up her promising career and have stayed with me in pursuit of my dreams. Thanks for your unending support and love. She never let me worry about anything else other than research and career. To my parents who have always supported me in all my endeavors, have always encouraged and

inspired me to excel in every aspect of my life. Thanks especially my Dad, the man who taught me the values of life and family, you are the humblest, kindest, and hardworking person I have ever met. I miss our debates and arguments over everything. Thank you both for being the best parents that a man could have asked for. Special thanks to my children who made me realize that I have grown up and need to be responsible. I have not always been attentive and spent enough time with them as I would have liked to. I promise I'll be making up for that soon. To all of my family and friends, who I was unable to mention personally, thank you all for your love and support.

Debmalya Bhattacharyya

April 21, 2016, Kent, Ohio

## ABSTRACT

Guanine-rich nucleic acid sequences adopt a four stranded secondary structure referred to as a G-quadruplex (GQ). These structures, when adopted by RNA sequences, have been shown to modulate several biochemical processes, especially translation. The RNA GQ structures are widely perceived as translation repressors although a few recent reports suggest that they are sometimes essential for translation. The presence of GQ structures in IRES elements was known to be essential for cap-independent translation initiation. We investigated the mechanism of how these structures aid IRES-mediated translation initiation wherein we observed that an independently folding RNA GQ domain in the internal ribosome entry site (IRES) of human VEGF mRNA directly recruits the 40S ribosomal subunit thereby controlling non-canonical translation initiation. These findings provide a unique and defined role of an RNA structure in the cap-independent translation initiation by cellular IRESs explaining a hitherto unknown mechanism for the necessity of a GQ structure in IRES function. Furthermore, we employed RNA engineering to demonstrate that RNA GQ function is dependent on the context in which such domains are located. Additionally, we also report a strategy for harnessing the natural ability of RNA GQs to inhibit translation by rationally inducing RNA: DNA hybrid GQ on a targeted mRNA to knockdown endogenous gene expression. Targeted RNA: DNA hybrid GQ structures were induced at the 5'-untranslated region (UTR) and the protein coding region of the eIF-4E mRNA by partially modified extraneous DNA sequences. We observed the formation of a stable induced GQ *in vitro* at a physiologically relevant salt

concentration. The rationally designed GQ inducing sequences resulted in anti-proliferation of HeLa cells in a dose-dependent manner. Also, inducing GQ in the 5'-UTR and the protein-coding region of eIF-4E mRNA in HeLa cells resulted in a 30% and 60% inhibition of the endogenous protein expression respectively. The above strategy opens up a new avenue of anti-cancer RNA therapeutics.

## CHAPTER ONE

### 1. INTRODUCTION

#### 1.1 DISCOVERY OF G-QUADRUPLEXES

Since the discovery of the structure of the DNA double helix by Watson and Crick, DNA has largely been thought to exist as a right-handed double-stranded helix for a considerable period of time (Watson and Crick, 1953). The discovery of the canonical DNA double helix structure comprised of Watson-Crick base pairing has provided the basis for our understanding of the genetic code. However, since then, the discovery of several non-canonical DNA structures was observed to regulate several cellular processes, thereby establishing their importance in the regulation of diverse cellular functions. In 1910, Bang et al. demonstrated that high concentrations of guanylic acid (GMP) can form gels in aqueous solution (Bang, 1910). Almost fifty years later the G-quadruplex (GQ) structure was discovered based on X-ray diffraction studies. It was observed that each guanine residue in the G-quartet acts as both acceptor and donor of two hydrogen bonds by Hoogsteen base pairing (Gellert et al., 1962). Thereafter studies revealed that G-rich repeat sequences in the human telomere were capable of forming GQ structures *in vitro* (Wang and Patel, 1993). It was also established that these GQ structures in the telomere regions inhibit the telomerase activity, which is often overexpressed in cancer cells thus conferring them with immortality (Zahler et al., 1991).

Thereafter bioinformatics reported that G-rich sequences capable of forming GQ

structures were prevalent throughout the entire human genome (Huppert and Balasubramanian, 2005; Todd et al., 2005; Eddy and Maizels, 2008; Mani et al., 2009; Neidle, 2009), especially in some of the key growth regulatory genes and oncogenes (Simonsson et al., 1998; Siddiqui-Jain et al., 2002; Sun et al., 2005; Cogoi and Xodo, 2006; Dexheimer et al., 2006). Bioinformatics studies revealed the existence of 376,000 putative quadruplex-forming sequences (PQS) in the entire human genome and significant enrichment of PQS in the promoter regions (Huppert and Balasubramanian, 2005; 2007). The PQS can be described as G-rich sequences with at least four stretches of G residues where each stretch comprises at least two G residues. Thus, typical PQSs' are defined as  $G_{2+} N_{1-7} G_{2+} N_{1-7} G_{2+} N_{1-7} G_{2+}$ . The numbers of tiers in the GQ are limited by the number of contiguous G residues in the shortest G-stretch. The prevalence of the PQS in the promoter regions of several clinically important genes suggested a role of these structures in transcriptional regulation. Moreover, the high frequency of PQSs in the promoter regions of proto-oncogenes and other clinically significant genes suggested they might be involved in transcriptional regulation of the genes and thereby might play an important role in cancer regulation. The regulation of transcription by GQ structures in the promoter regions of clinically significant genes such as C-MYC, BCL-2, C-KIT, K-RAS has been well established (Siddiqui-Jain et al., 2002; Sun et al., 2003; De Armond et al., 2005; Cogoi and Xodo, 2006; Dai et al., 2006; Fernando et al., 2006) and can be potential targets for chemotherapeutics (Balasubramanian and Neidle, 2009). G-quadruplexes present in the immunoglobulin heavy chain switch regions, mutational hotspots have been implicated in the maintenance of chromosomal integrity and

regulation of replication, transcription and recombination processes (Simonsson, 2001). The functional association between GQ structures in telomeres, oncogenes, and cancer, targeting DNA GQs for therapeutic intervention has been an active area of research (De Cian et al., 2008; Huppert, 2008; Monchaud and Teulade-Fichou, 2008; Nielsen and Ulven, 2010; Zhang et al., 2014b).

Like the DNA GQs, RNAs can also adopt such structures. Apparently, there is no physicochemical barrier towards the formation of RNA GQ structures (Kim et al., 1991; Cheong and Moore, 1992). In fact, the probability of GQ formation by RNA is higher than their DNA counterparts due to the absence of a competing complementary strand and RNA GQ structures were observed to be more stable than their DNA counterparts (Cheong and Moore, 1992; Sacca et al., 2005; Kumari et al., 2007a). Early evidence of GQ formation in RNA was observed in a 19 nucleotide G-rich sequence present in the 3' terminus of 5S rRNA of *E. coli* (Kim et al., 1991). The UG<sub>4</sub>U sequence motif was proposed to form a parallel stranded tetraplex structure containing four stacked G-tetrads and, at least, one U quartet based on NMR and molecular dynamics data (Cheong and Moore, 1992). The RNA GQ structures are extremely stable wherein it was observed that the exchange of the imino protons in the quartets with the solvent occurred over an extensive period at physiological temperature. Thereafter, a GQ structure was reported to be present in the 3'-untranslated region (UTR) of the insulin-like growth factor II (IGF II) mRNA located just downstream of the endonucleolytic cleavage site (Christiansen et al., 1994). A computational survey found GQ forming sequences to be enriched within mRNA processing sites and in the 5'-UTRs of mRNAs of genes related to cancer

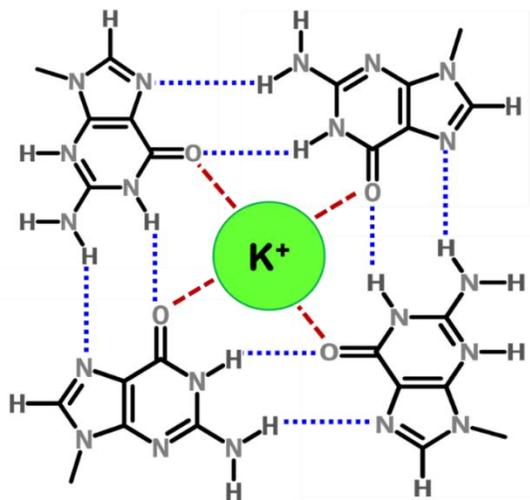
(Kostadinov et al., 2006). Bioinformatics studies suggested that there are approximately 3000 5'-UTRs that had, at least, one RNA PQS (Kumari et al., 2007a; Huppert et al., 2008). Thereafter, several studies further established the role of RNA GQs in the regulation of translation. RNA GQ structures in the 5'-UTR have been shown to repress translation of several clinically important genes such as NRAS, Zic-1, VEGF, TRF2, ERS1, THRA, and BCL-2 (Kumari et al., 2007a; Arora et al., 2008; Balkwill et al., 2009; Morris and Basu, 2009; Gomez et al., 2010; Shahid et al., 2010). These GQ motifs exhibited a dual mode of regulation wherein several naturally occurring RNAs have been shown to have an inhibitory effect on translation as well as to be essential for translation (Bonnal et al., 2003; Morris et al., 2010; Agarwala et al., 2013; Agarwala et al., 2014). Subsequently, several studies reported the presence of RNA GQs and their functional significance in the coding region, micro-RNA biogenesis, long non-coding RNAs and telomeric ends (Arora and Suess, 2011; Ji et al., 2011; Jayaraj et al., 2012; Beaudoin and Perreault, 2013; Endoh et al., 2013a; Martadinata and Phan, 2013; Mirihana Arachchilage et al., 2015).

A lingering question in the field of GQ research until recently whether these structures actually formed *in vivo*. Almost all the studies on GQ structures and their regulatory role presented the evidence of GQ by several biochemical and biophysical methods albeit under *in vitro* conditions. Eventually, the skepticism was put to rest with evidence of intracellular formation of GQ structures by both DNA and RNA. A single chain fragment antibody (BG-4) specific for GQ structures was used for the visualization

of GQs in human cells by immunofluorescence microscopy (Biffi et al., 2013; Biffi et al., 2014).

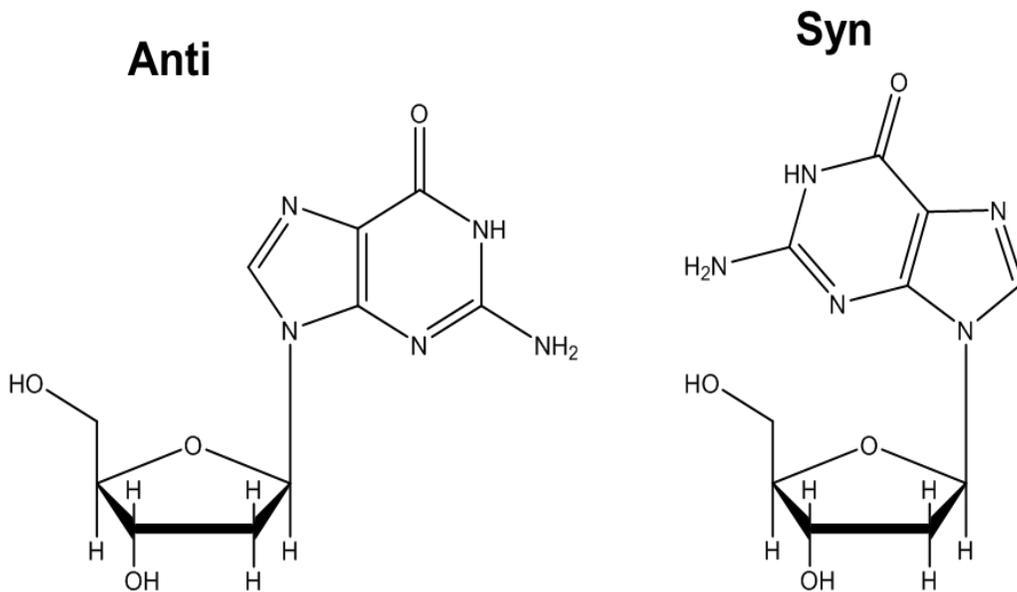
## 1.2 STRUCTURE OF G-QUADRUPLEXES

The GQ structures are non-canonical four stranded secondary structures found in nucleic acid sequences rich in guanine residues. Four guanines can form a square planar tetrad in which each guanine serves as both a hydrogen bond donor and acceptor (Fig. 1.1). The pairing of the N1 on the first guanine with the O6 on the second guanine along with the pairing of N2 on the first guanine with the N7 on the second guanine results in eight hydrogen bonds per G-quartet known as Hoogsteen base pairing. Two or more of these G-quartets stack upon each other to form the GQ structure (Smith and Feigon, 1992). The GQ structures are stabilized by monovalent cations interacting with the carbonyl oxygen atoms located at the center of G-quartet stacks which results in the more stable stacking of tetrads (Neidle and Balasubramanian, 2006). These monovalent cations stabilize the GQ structure in the order  $K^+ > Na^+ > Li^+$  (Burge et al., 2006).



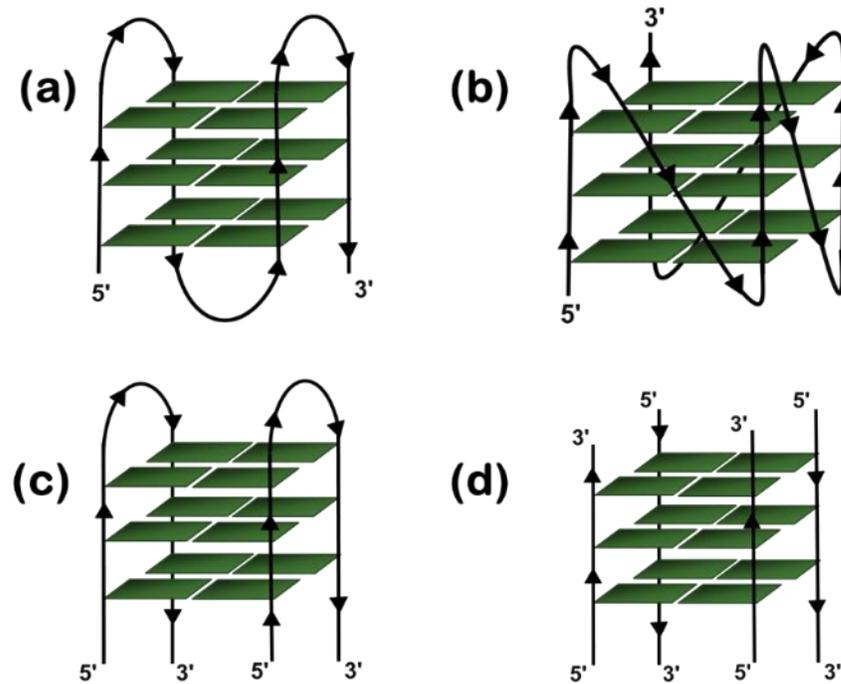
**Figure 1.1- Structure of the G-quartet:** The G-quartet is a square planar orientation of four guanine residues that interact with each other by Hoogsteen bonds (blue dotted). The monovalent ion in the central cavity coordinates with the carbonyl oxygen and stabilizes the quartet.

The GQ structures exhibit diverse topologies depending on whether it is RNA or DNA, the presence of monovalent cation ( $K^+$  or  $Na^+$ ), and glycosidic bond conformation (syn or anti) (Fig. 1.2). The number of nucleic acid molecules involved in the formation of GQ structures results in variants such as, unimolecular (intramolecular) (Fig. 1.3 a & b), bimolecular (Fig. 1.3 c) or tetramolecular (Fig. 1.3 d) structures; relative orientation of the strands leading to parallel (Fig. 1.3 b), antiparallel (Fig. 1.3 a) or a hybrid (Fig. 1.3 d) structures; the number of stacking G-quartets and the nucleotide sequence (Keniry, 2001; Neidle and Balasubramanian, 2006; Patel et al., 2007).



**Figure 1.2- Glycosidic bond conformation:** Conformation of the glycosidic bond between the base and the sugar ribose sugar in the nucleic acid influences the topology of the GQ.

The GQ structures adopt uni-, bi-, or tetra-molecular forms consisting of one, two, or four strands respectively (Neidle and Balasubramanian, 2006). The unimolecular GQ structures are also known as intramolecular structures; however, the intermolecular structure can be more diverse depending upon the molecularity of the strands involved in the formation of the GQ structure. For an intramolecular GQ to form at least four stretches of guanosines, with a minimum of two G residues in tandem in each stretch is required. A typical gap between two consecutive G-stretches can range from one to seven nucleotides (Neidle and Balasubramanian, 2006). Four G-stretches are required for intramolecular GQ formation; three gaps form the loops when the sequence adopts a GQ structure. The length of the loop and the nucleotide composition plays a crucial role in determining the stability of the GQ structures (Risitano and Fox, 2004; Cevc and Plavec, 2005; Rachwal et al., 2007a; b; Rachwal et al., 2007c; Guedin et al., 2010; Pandey et al., 2013). Moreover, the orientation of these strands also determines the topological classification of the GQ structure. A GQ is termed parallel if the polarities of all the strands are oriented in the same direction (Fig. 1.2 b) with respect to one another. In contrast, if each strand has an opposite polarity with respect to the two adjacent strands, the quadruplex is termed anti-parallel (Fig. 1.2 a, c and d) (Neidle and Balasubramanian, 2006).



**Figure 1.3- Different topological variants of GQ structures:** The GQ structures form several conformational variants depending on the number of strands and quartets, the orientation of the strands.

### 1.2.1 RNA G-Quadruplex Structure

The formation of secondary and tertiary structures within the 5'-UTRs of mRNAs have been shown to play important roles in the regulation of gene expression (Gray and Hentze, 1994; Mignone et al., 2002). Several studies on RNA secondary structures with canonical purine/pyrimidine base pairing such as hairpin, loops, bulges, pseudoknots and kissing complexes with functional implications have been reported in the literature (Svoboda and Di Cara, 2006; Jambhekar and Derisi, 2007; Bevilacqua and Blose, 2008; McManus and Graveley, 2011). Guanine-rich sequences of both RNA and DNA can form GQ structures via hydrogen bonding between Watson-Crick and major-groove base edges (Neidle and Balasubramanian, 2006). However, the formation of GQ by RNA is

relatively more favorable compared to its DNA counterpart due to the absence of a competing complementary strand that often restricts the formation of GQ in DNA. RNA GQ structures are more thermodynamically stable, compact and less hydrated compared to their DNA counterpart (Arora and Maiti, 2008; Arora and Maiti, 2009; Joachimi et al., 2009; Zhang et al., 2010; Zhang et al., 2011). In fact, it has been suggested that the 2'-OH of RNA has stabilizing effects on RNA GQs, possibly through extended hydrogen bonding interactions with phosphate and oxygen backbone atoms, O4' sugar oxygen and H-bond acceptors such as the N2 groups of quartet-forming guanines. The higher stability of RNA GQ is attributed to the 2'-OH group of the ribose sugar which acts as a scaffold for an ordered network of water molecules and extensive bonding (Tang and Shafer, 2006). The significance of the 2'-OH group was well established when it was observed that substituting the hydroxyl group by other chemically modified analogs resulted in the destabilization of GQ structure (Sacca et al., 2005). The presence of the 2'-OH group in the RNA GQ exerts conformational constraints on the GQ topology that prevent the syn-conformation. The syn-configuration is a prerequisite in GQ for antiparallel topology. The hydroxyl group restricts orientation of the base about the glycosidic bond to the anti-conformation thereby imparting a C3'-endo pucker and limiting the topology of the RNA GQ to the parallel conformation (Tang and Shafer, 2006). Although the topology of RNA GQ is less dependent on other environmental conditions, the formation of only parallel conformation reduces their topological diversity (Zhang et al., 2010; Zhang et al., 2011).

The substitution of thymine with uracil in the loop of a GQ structure revealed that removal of methyl group stabilizes the stacking interactions and releases the structured water molecules. By instituting the thymine to uracil substitution, the authors also suggested the importance of loop residues in RNA GQ stability (Olsen and Marky, 2009). In addition, various independent studies have demonstrated the effect of varying loop lengths and composition on the stability of the GQ structure (Hazel et al., 2004; Risitano and Fox, 2004; Cevec and Plavec, 2005; Sacca et al., 2005; Hazel et al., 2006; Guedin et al., 2008; Guedin et al., 2010). The stability of all three G-quartets was found to contribute equally to the stability of RNA GQ structures (Agarwala et al., 2015a) as opposed to the more important middle quartet in DNA GQ (Lee and Kim, 2009; Tomasko et al., 2009). Recently, studies on folding dynamics suggested that RNA GQ formation with two G-quartets show positive cooperativity, high dependence on  $K^+$  concentration and very few intermediate states (Mullen et al., 2012).

Although the DNA GQ structures have been thoroughly studied by several NMR and crystallography reports (Kang et al., 1992; Smith and Feigon, 1992; Hud et al., 1996; Neidle and Read, 2001; Parkinson et al., 2002; Matsugami et al., 2003), relatively fewer studies have been conducted on RNA GQs (Collie et al., 2010). The NMR spectrum of an RNA GQ was first reported for sequence R14 (GGAGGUUUUGGAGG) present in mRNAs of immunoglobulin regions (Liu et al., 2000). Crystal structures suggest that 2'-OH group and the rigid C3'-endo sugar pucker in RNA redefine the hydration structure in grooves and hydrogen bonding patterns (Collie et al., 2010). Thereafter, the crystal structure of TERRA GQ complexed with the small molecule acridine also

revealed that the same 2'-OH group of ribose sugar interacts with the acridine (Collie et al., 2010; Collie et al., 2011). Structural studies on RNA GQ using NMR and X-ray crystallography were performed only on intermolecular RNA GQs, however so, detailed structural insight on the intramolecular RNA GQ structures are still lacking.

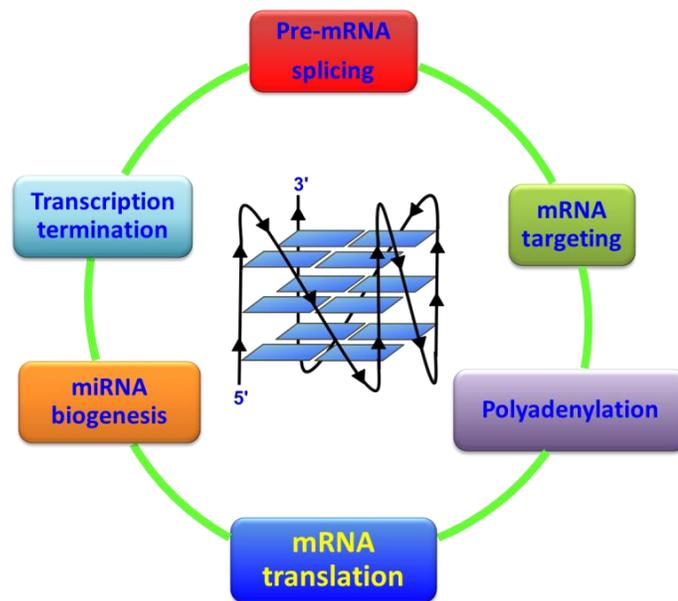
### **1.3 FUNCTIONS OF RNA G-QUADRUPLEX STRUCTURES**

Although the initial focus of research has been on DNA GQs and their role in biology, in recent years the role of RNA GQs as regulatory elements of gene expression has been widely established (Fig. 1.3). The RNA GQs located in the 5'-UTR of mRNAs were found to be involved in translational regulation. However, reports of RNA GQ structures and their involvement in transcriptional regulation, splicing, mRNA transport, miRNA maturation amongst others widened the functional impact of the RNA GQ structures.

#### **1.3.1 Translational regulation**

The first example of translational repression by an RNA GQ was reported in the 5'-UTR of the human NRAS proto-oncogene transcript (Kumari et al., 2007a). Subsequently GQs present in the 5' UTR of the human Zic-1 (Arora et al., 2008), NRAS (Kumari et al., 2007a), exon C of human and bovine ESR (estrogen receptor  $\alpha$ ) (Derecka et al., 2010), MT3-MMP (Membrane-type 3 matrix metalloproteinase) (Morris and Basu, 2009), Bcl2 (B-cell lymphoma 2) (Shahid et al., 2010), TRF-2 (Telomeric repeat binding factor 2) (Gomez et al., 2010), ADAM-10 (Lammich et al., 2011), and cyclin D3 (Weng

et al., 2012) were also observed to repress translation. The first translational upregulation by RNA GQs in the 5'-UTR of mRNA was associated with the cap-independent translation of FGF-2 and VEGF (Bonnal et al., 2003; Morris et al., 2010). Thereafter evidence of GQ structures



**Figure 1.4- Functions of RNA GQ structures:** The RNA GQ structures regulate gene expression by influencing several different mRNA dependent cellular processes.

augmenting the translation in cap-dependent translation initiation was observed in TGF $\beta$ 2 and FOXE3 (Forkhead box E3) (Agarwala et al., 2013; Agarwala et al., 2014). By contrast, the evidence of GQ structures in the 3'-UTR in translation modulation has been sparse. The presence of a GQ structure in the 3'-UTR of proto-oncogene PIM1 transcript inhibits translation, although the mechanism of inhibition is unclear (Arora and Sues, 2011). Nevertheless, the GQ structures in the 3'-UTR are more versatile in their ability to modulate end processing, transcription termination, alternative polyadenylation, and mRNA localization amongst others. RNA GQ structures suppress translation when

present in open reading frame (ORF), that correlates with the stability of the structure (Endoh et al., 2013b). The GQ structure present in the ORF of human estrogen receptor  $\alpha$  was shown to pause translational elongation and thereby influence protein misfolding leading to proteolysis (Endoh et al., 2013a; Endoh et al., 2013c).

### **1.3.2 3' End processing and alternative polyadenylation**

One of the foremost observations of RNA GQ function was initiated with the discovery of a GQ structure immediately downstream of an endonucleolytic cleavage site in the 3'-UTR of IGF2 mRNA (Christiansen et al., 1994; Nielsen and Christiansen, 1995). Thereafter, GQ formation was observed to be critical for maintaining the efficiency of 3'-end processing of p53 following UV irradiation-mediated DNA damage. The GQ interacts with hnRNP H/F protein and maintains the expression of p53, causing apoptosis as a consequence of DNA damage (Decorsiere et al., 2011). Recently it was observed that GQs located in the 3'-UTR of LRP5 and FXR1 increase the efficiency of alternative polyadenylation leading to the generation of several short transcripts. Production of these short transcripts was further implicated in impairing miRNA regulation of FXR1 mRNA (Beaudoin and Perreault, 2013).

### **1.3.3 Mitochondrial transcription termination**

The GQ structures were observed to modulate the termination of transcription of the mitochondrial light strand in a mechanism similar to Rho-independent transcription

termination in prokaryotes, wherein the GQ substitutes for the hairpin loop formed in bacterial mRNA (Wanrooij et al., 2010; Wanrooij et al., 2012).

#### **1.3.4 mRNA localization**

The GQ structures in the 3' UTR of PSD-95 and CaMKIIa are required for their localization from soma to dendrites of cortical neurites. These GQs in the 3'-UTR act as structural “zipcodes” assisting the process of mRNA localization in neurons (Subramanian et al., 2011). It was further demonstrated that depletion of fragile X mental retardation protein (FMRP) was not enough to eliminate the transport of these mRNA thereby refuting the previously accepted role of FMRP in transporting the RNA to the neurites. Bioinformatics analyses revealed the presence of PQS of which 30% are highly conserved in the 3'- UTR of dendritic mRNAs, suggesting that the GQs act as neurite-targeting elements and are important for the establishment and maintenance of cell polarity (Subramanian et al., 2011).

#### **1.3.5 Alternative Splicing**

The GQ structures in the intronic regions of mRNA act as cis-regulatory elements in alternative splicing. The hnRNP F binds to a PQS in the pre-mRNA and is essential in splicing of the scr N1 exon (Min et al., 1995). Genome-wide studies indicated that G-rich tracts in introns of many genes have potential to adopt GQs and thereby can affect the splicing and expression patterns of these genes (Kostadinov et al., 2006). Subsequently, the role of GQs in the regulation of alternative splicing was observed in

several physiologically important genes such as B-tropomyosin, hTERT, FMR1, BACE-1, and PAX9 (Sirand-Pugnet et al., 1995; Gomez et al., 2004; Didiot et al., 2008; Fiset et al., 2012; Ribeiro et al., 2015). The effect of GQ in alternative splicing in p53 is unique wherein the GQ is located in an intron (intron 3) that affects the alternative splicing of another intron (intron 2). Furthermore, the GQ in p53 did not completely alter the splicing pattern but fine-tuned the relative abundance of the splice variants of p53, leading to differential expression of transcripts encoding distinct p53 isoforms (Marcel et al., 2011). The regulation of the cell cycle, apoptosis, and p53 target genes is affected by the relative expression of the various isoforms.

### **1.3.6 Biogenesis of miRNAs**

The pre-miRNA structures are stem-loop structures which are processed by Dicer to form the matured miRNAs. It has been well established that the stem-loop structure is essential for Dicer to act upon pre-miRNAs and cleave them to form mature miRNAs. (MacRae et al., 2007) Recently our laboratory reported the presence of a PQS in pre-miRNA 92b which in the presence of physiological levels of  $K^+$  ions adopts a GQ structure thereby impeding the Dicer-dependent maturation of the miRNA. The study also reported that such GQ forming sequences are prevalent in pre-miRNA (16%) and suggested that the GQ structures play an important role in gene expression by influencing the miRNA maturation (Mirihana Arachchilage et al., 2015). A similar mechanism has been observed in miRNA in global analyses where cells were treated with a GQ destabilizing ligand TmPyP4 (Pandey et al., 2015).

## **1.4 INTERACTION OF PROTEINS WITH RNA G-QUADRUPLEX STRUCTURES**

Reports on proteins binding to RNA GQ's are relatively less compared to examples of DNA GQ binding proteins (Fry, 2007). The binding of FMRP and FMR2P to RNA GQ are probably the most extensively studied cases. These proteins are associated with two forms of mental diseases: the Fragile X Mental Retardation syndrome (FXS) and the FRAXE-associated mental retardation (FRAXE) respectively (Melko and Bardoni, 2010). The FMRP protein interacts using the arginine–glycine–glycine (GGG) box motif which has a high affinity and specificity for RNA GQ motif (Darnell et al., 2001; Blackwell et al., 2010). Studies revealed that the RNA targets of FMRP contain several PQS in their 5'-UTR, coding regions and 3'-UTR (Schaeffer et al., 2001; Todd et al., 2003; Castets et al., 2005; Menon and Mihailescu, 2007; Westmark and Malter, 2007). Microtubules Associated Protein 1b (MAP1B) and the catalytic subunit of Protein Phosphatase 2A (PP2Ac), which are known FMRP targets, revealed the presence of one or more GQ structures in their 5'-UTR (Darnell et al., 2001; Castets et al., 2005). FMRP binds to these motifs in MAP1B mRNA and inhibits its translation during active synaptogenesis in neonatal brain development thereby regulating normal development of dendritic spines (Lu et al., 2004). FMRP also downregulates PP2Ac mRNA translation by binding with its 5'-UTR GQs (Castets et al., 2005). Additionally, a report showed that the longer (CGG)<sub>n</sub> sequences in 5'-UTR of FMR1 gene in FXS form

GQ structures, blocking the migration of the 40S subunit and thereby becoming an impediment of translation of FMR1 mRNA. Indeed, expression of the quadruplex disrupting proteins hnRNP A2 or CBFA were shown to alleviate this translational block (Khateb et al., 2007). RNA targets of FMRP harboring GQ in their 3'-UTR PSD95 and Sem3F mRNAs have been reported (Todd et al., 2003; Menon and Mihailescu, 2007; Melko and Bardoni, 2010; Evans et al., 2012). Interestingly, FMRP was found to regulate its own activity by binding to the GQ present in the coding region of its mRNA and inhibiting its translation by a negative feedback loop (Schaeffer et al., 2001).

Helicase proteins are also known to interact and unwind GQ structures in RNA. RHAU is a DEAH box containing RNA helicase that unwinds both RNA and DNA GQ structures (Vaughn et al., 2005; Creacy et al., 2008; Lattmann et al., 2010; Lattmann et al., 2011; Meier et al., 2013; Booy et al., 2014). This helicase remodels the GQs located in the 5'-end of telomerase RNA and resolves GQs increasing the expression of Yin Yang 1 gene (Booy et al., 2012; Huang et al., 2012). Alternatively, RHAU was reported to bind to the GQ present in the 3'-UTR of PITX1 mRNA and suppress its expression via a unique miRNA-mediated gene regulation (Booy et al., 2014). Another helicase DHX9 was reported to unwind both RNA and DNA GQs (Chakraborty and Grosse, 2011). The increasing repertoire of GQ resolving helicases enables us to understand the mechanism of unfolding and equilibrium between folded and unfolded states of this structure. Recently, transcriptome scale ribosomal profiling of helicase eIF4A-dependent transcripts revealed the enrichment of 5'-UTR sequences with a propensity to form RNA

GQs, indicating the probable role of this factor in the unwinding of this secondary structure during the translation process (Wolfe et al., 2014).

In addition, TLS/FUS- telomere-binding protein was observed to bind telomeric DNA GQ and TERRA RNA GQ via its RGG domain to form a ternary complex. The GQs serve as a scaffold for recruitment of these proteins to the telomere, thereby regulating telomere length by histone modifications of telomere (Takahama and Oyoshi, 2013; Takahama et al., 2013). Protein pull-down followed by MALDI-TOF mass spectrometry identified the proteins interacting with GQs located in 5'-UTR of MMP6 and ARPC2 mRNAs. The identified proteins were predominantly comprised of ribosomal proteins, heterogeneous nuclear ribonucleoproteins, nucleolin and splicing factors (von Hacht et al., 2014). Similarly, by a pull-down strategy, a wide array of proteins interacting with GQ forming TERRA RNA was identified by MALDI-TOF MS (Deng et al., 2009; de Silanes et al., 2010). The shelterin proteins TRF1 and TRF2 were shown to be capable of binding directly to TERRA. Moreover, proteins belonging to the hnRNP class were also found enriched in the pool of binding partners (Deng et al., 2009).

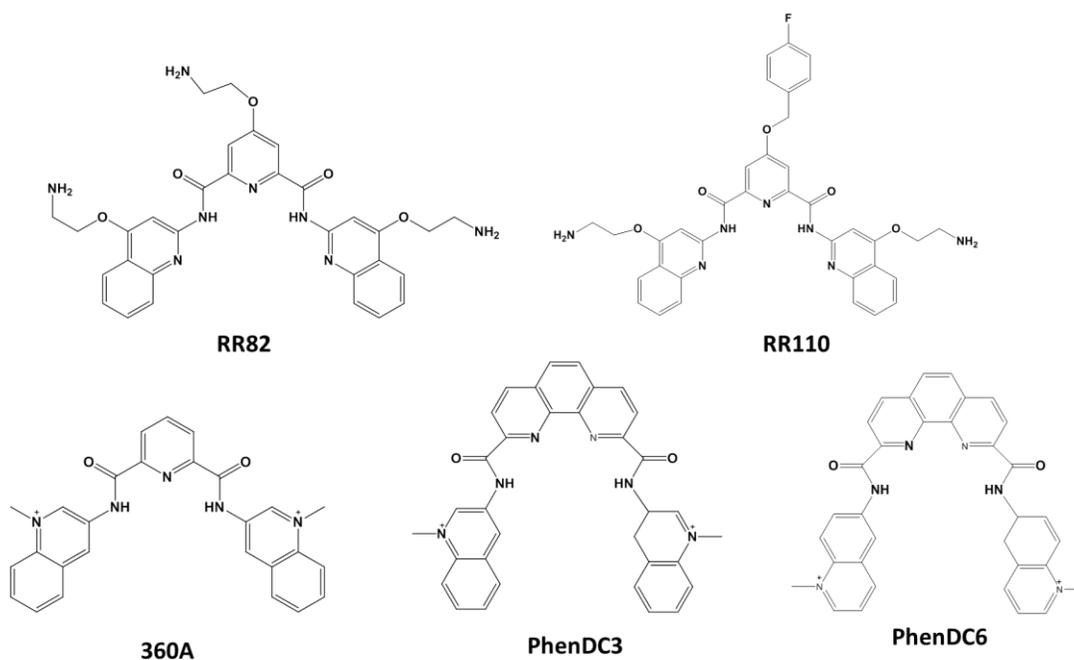
## **1.5 TARGETING RNA GQ STRUCTURES FOR THERAPEUTICS**

Modulation of biochemical processes by targeting nucleic acid secondary structure with small molecules is a promising strategy to regulate gene expression (Collie and Parkinson, 2011; Sissi et al., 2011). Small molecule ligands interacting with GQ can act by either stabilizing or destabilizing the GQ thereby reinforcing or reducing its

function. Furthermore, these ligands can also interfere with the interaction of trans-acting factors with the GQ, thereby modulating the activity of the structure (Bugaut and Balasubramanian, 2012). The GQ structures at the telomeric ends and promoter regions have been extensively studied as targets of small molecule ligands (Ohnmacht and Neidle, 2014). Several small molecule ligands such as porphyrin, acridine, pentacridium, quinacridine, telomestatin, naphthalene diamide, bisquinolium and their derivatives bind selectively and stabilize DNA GQ structures (Phan et al., 2006; Xu, 2011). Several clinically important genes harbor RNA GQ and, therefore, selective targeting of these structures to regulate gene expression, is a promising strategy towards drug discovery for various diseases. Initially, it was observed that a naphthalene diamide derivative was more selective towards RNA GQ compared to its DNA counterpart (Collie et al., 2009). Thereafter, three natural alkaloids namely nitidine, palmatine, and jatrorrhizine were observed to interact with the 5'-UTR RNA GQ of Bcl2 with high affinity (Tan and Yuan, 2013). Pyridostatin is a known molecule to stack over the terminal G-quartets of DNA GQ stabilizing the structure to cause growth arrest in cancer cells by inducing extensive DNA damage (Rodriguez et al., 2012). Another derivative carboxypyridostatin exhibits preferential binding and selectivity towards RNA GQ (Di Antonio et al., 2012).

It was observed that pyridine-2,6-bis-quinolino-dicarboxamide derivative RR82 and the *para*-fluorophenyl substituent at pyridine C4 RR110 (Fig. 1.4) are potent binders and stabilizers of RNA GQ present in the 5'-UTR of NRAS mRNA and foster the inhibition of NRAS gene expression (Bugaut et al., 2010). Moreover, three other

bisquinolium derivatives PhenDC3, PhenDC6 and 360A (Fig. 1.4) with similar binding affinities towards RNA and DNA GQs also have been used for small molecule-mediated gene regulation (Gomez et al., 2010; Shahid et al., 2010; Halder et al., 2011; Marcel et al., 2011; Yu et al., 2014).



**Figure 1.5- Ligands known to interact specifically with RNA GQ structures**

Studies indicate that the same small molecule might elicit varied effect under different contexts. TmPyP4 is a cationic porphyrin that was observed to have no RNA GQ selectivity and thereby, did not inhibit NRAS translation (Bugaut et al., 2010). However, it was conclusively shown to destabilize RNA GQ present in 5'-UTR of MT3-MMP relieving the repressive role of the GQ (Morris et al., 2012). Moreover, TmPyP4 also has been shown to distort RNA GQ structures formed by r(GGGGCC)<sub>n</sub> repeats in

the C9orf72 gene thereby interfering with their interaction with trans-acting factors (Zamiri et al., 2014).

The small molecules observed to bind to GQ structures are predominantly positively charged with a large planar aromatic core, which neutralizes the negatively charged phosphates of the backbone, and maximizes  $\pi$ -stacking interactions with the quartets of GQ respectively. Often the ligands possess positively charged side chains with functional groups to interact with the loops and the grooves of the GQ structure (Collie and Parkinson, 2011). The small molecules binding to RNA GQs in most cases also interact with DNA GQs. Moreover, the limited conformational variation amongst RNA GQ structures also renders selective targeting of a transcript containing GQ much more challenging. Since it has been already established that these structures play an opposing effect on the expression of several genes and their functions are so varied, the non-specific targeting of RNA GQ is not practical for therapeutic purposes. However, contrary to the current notion that the GQ interacting small molecules might bind non-specifically to several GQ structures, it was observed that when a library of RNA GQ forming sequences with varying loop sequence but fixed loop length were screened for their binding to kanamycin A, only a few sequences actually were identified to bind to the ligand (Mirihana Arachchilage et al., 2014).

Present research on ligands has been mostly focused on small molecules with planar aromatic rings to induce  $\pi$ -stacking interactions with the GQ structures. An alternative strategy to target specific GQ structures would be to design and explore

ligands that specifically bind to grooves, loops and other regions of GQ, considering that these are the only variable regions amongst the RNA GQ structures. Exploiting the effects of the 2' hydroxyl group of the ribose on RNA GQ structure can form the basis for ligand design for discerning the differences between DNA and RNA GQs (Xu and Komiyama, 2012).

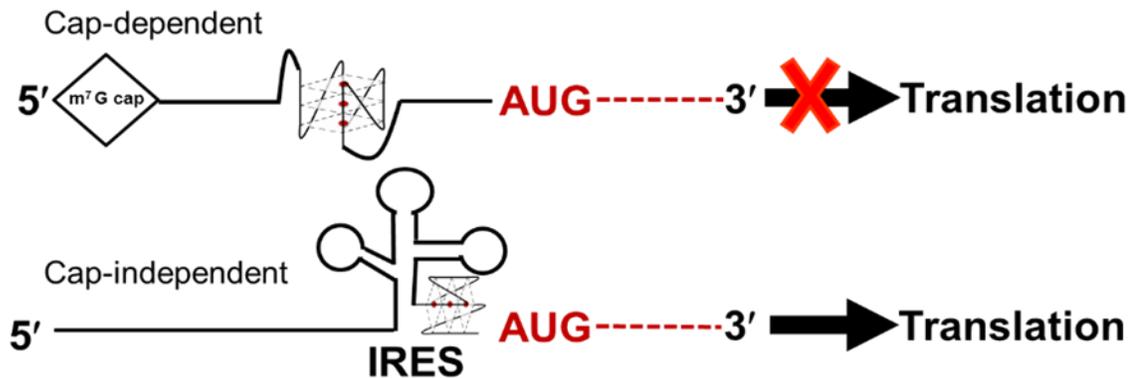
## **1.6 MODULATION OF TRANSLATION BY GQ STRUCTURES PRESENT IN THE 5'-UTR**

The role of RNA GQ structures in the regulation of translation was first highlighted in the case of an RNA GQ located in the 5'-UTR of the proto-oncogene NRAS. The authors discovered an 18 nucleotide (nt.) motif highly conserved both in position (near the 5'-cap) and sequence among various species (Kumari et al., 2007a). The sequence was observed to fold into a very stable GQ structure as identified by *in vitro* analysis via circular dichroism (CD) and UV-melting experiments. *In vitro*, dual luciferase reporter assay showed that there was a 3.7-fold enhancement in activity of both a mutant and the deletion plasmid over the wild type indicating that this particular structure had an inhibitory effect on translation. Bioinformatics analyses indicated that ~3000 transcripts have at least one PQS in their 5'-UTRs (Kumari et al., 2007a; Huppert et al., 2008). Subsequently, in another study, the positional (in the context of the entire 5'-UTR) and stability effect of GQ on the translational repression of the genes was investigated wherein it was demonstrated that the GQ structures repress translation when situated relatively close to the 5'-end of the mRNA (Kumari et al., 2008). The study

suggested that shifting this motif in NRAS 5'-UTR to a position far away from the cap region causes loss of its function, however, it was also observed that the positional shift does not alter the function of RNA GQ structures (Kumari et al., 2008; Halder et al., 2009). The difference in the outcome of shift may be attributed to the different mRNAs or different experimental setups that were employed for the studies. Furthermore, the position specific role of GQ structures at the 5' end of naturally occurring mRNAs was evaluated wherein it was observed that GQ structures inhibited gene expression in AKTIP (AKT interacting protein) and CTSB (Cathepsin B), but augmented protein production in FOXE3. The study concluded that GQ structures exhibit varied functions in different genes and probably is influenced by neighboring sequences, trans-factors, and intracellular conditions (Agarwala et al., 2014). Moreover, the thermodynamic stability of the RNA GQ dictated the repressive effect on translation. It was reported that translational repression is correlated directly to the number of quartets and inversely to the GQ loop length flanking the Shine-Dalgarno (SD) sequence (Wieland and Hartig, 2007). Thereafter, the inhibition of translation in eukaryotic cells by an RNA GQ motif in the 5'-UTR of Zic-1 was reported (Arora et al., 2008). The evolutionarily conserved GQ motif from the 5'-UTR of the mRNA from the zinc-finger protein of the cerebellum 1 (Zic-1) repressed translation up to 80%, without affecting the mRNA expression in HeLa cells (Arora et al., 2008). Subsequently, additional studies reported the presence of GQ forming sequences within the 5'-UTRs of mRNA of several human genes which inhibits translation (Halder et al., 2009; Morris and Basu, 2009; Gomez et al., 2010). The genes studied were all observed to be clinically significant genes that included the extremely

stable GQ motif matrix metalloproteinase MT3-MMP (Morris and Basu, 2009), the estrogen receptor ESR1 (Derecka et al., 2010), the apoptotic regulator BCL2 (Shahid et al., 2010), THRA (Beaudoin and Perreault, 2010) and the telomere shelterin protein TRF2 (Gomez et al., 2010). In most of the above studies and in another report where several GQ forming sequences present in the 5'-UTR were studied, it was observed that the GQ forming sequences when present in 5'-UTRs exhibit translational repression activity (Halder et al., 2009). It has been suggested that the mechanism of translation inhibition by the GQ structures might be similar to that of other stable secondary structures in the 5'-UTR, which interfere with association of trans-acting factors for translation initiation, block the screening mechanism of the 43S preinitiation complex or may recruit some inhibitory factors repressing mRNA translation (Bugaut and Balasubramanian, 2012; Agarwala et al., 2015b).

Interestingly, GQ structures that acted as essential elements, thereby augmenting translation compared to their mutant counterparts, were observed only in case of non-canonical cap-independent translation initiation (Bonnal et al., 2003; Morris et al., 2010). A GQ structure in conjunction with two stem-loop structures located within the FGF-2 IRES was observed to be essential for translation initiation (Bonnal et al., 2003). In our lab, a unique flexible GQ structure required in the IRES A of hVEGF was observed to be essential for cap-independent translation initiation (Morris et al., 2010).



**Figure 1.6- Role of GQ structures in the 5'-UTR:** The existing hypothesis on the regulation of translation by RNA GQ structures was that the GQ structure in 5'-UTR inhibits translation except when they are present in the context of an IRES.

It was widely believed that the GQ structures in the 5'-UTR of mRNAs repress translation except when they are present in an IRES element wherein they augment translation (Fig.1.5) (Bugaut and Balasubramanian, 2012). The presence of multiple G-tracts in the IRES-A suggested the possibility of a flexible GQ switch in response to biological stimuli. Our lab was interested in understanding the mechanism by which the GQ structure in the IRES of hVEGF mRNA modulates translation. On mutation or deletion of the GQ forming sequence, cap-independent translation is significantly repressed. The cellular IRESs are known to recruit the translation factors independent of the 5'-cap (Jackson et al., 2010a; Komar et al., 2012). We hypothesized that the GQ structure might be involved in recruiting some factors of the translation apparatus or the 40S ribosomal subunits. Furthermore, it was also likely that the deletion of the segment could bring about change in the folding of the other regions of the IRES thereby rendering it dysfunctional. We also wanted to investigate the context dependency of the GQ structures in the 5'-UTR of mRNAs. The fact that GQ structures play a different role

in the case of different gene expression can be specific to that gene context or can be that the GQ motif is independent of the context and can play the same role in any gene when inserted in the 5'-UTR. The importance of GQ structures in translation regulation was well established at the beginning of this work. We wanted to apply the translation regulation potential of GQ structures in a targeted fashion to manipulate the regulation of specific genes involved in diseases. We hypothesized that we can specifically target a gene by inducing GQ structures in the 5'-UTR and the protein-coding region of the mRNA, and thereby modulate the expression of the gene. The strategy entailed two contiguous G stretches with more than 3 Gs in each stretch either in the native mRNA sequence at the 5'-UTR region and in the coding sequences.

## CHAPTER TWO

"Adapted with permission from: D. Bhattacharyya, P. Diamond, S. Basu, An Independently folding RNA G-quadruplex domain directly recruits the 40S ribosomal subunit, *Biochemistry*, 54 (2015) 1879-1885. Copyright (2015) American Chemical Society."

### **2. AN INDEPENDENTLY FOLDING RNA G-QUADRUPLEX DOMAIN DIRECTLY RECRUITS THE 40S RIBOSOMAL SUBUNIT**

#### **2.1 ROLE OF GQ STRUCTURES IN CAP-INDEPENDENT TRANSLATION INITIATION**

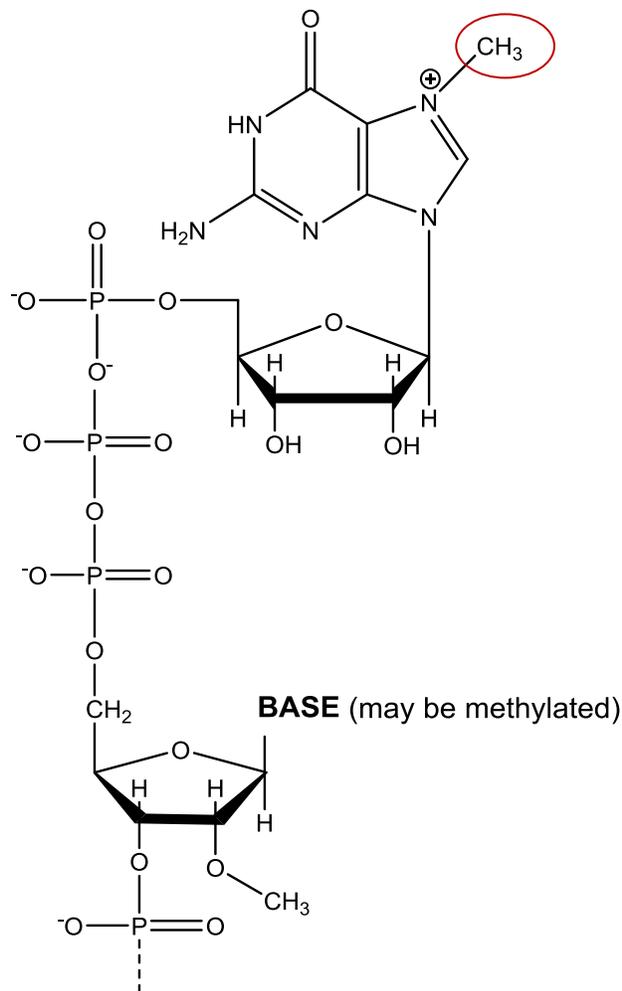
RNA G-quadruplexes (GQ) are secondary structures which when present in the 5'-untranslated (UTR) region of an mRNA can act as both necessary elements and repressors of translation (Bonnal et al., 2003; Kumari et al., 2007a; Morris and Basu, 2009; Morris et al., 2010; Agarwala et al., 2013). The role of GQ structures in translational modulation depends on the relative position of the GQ structure (Bugaut and Balasubramanian, 2012). However, it is well established that in the majority of the cases GQ structures inhibit translation as was observed in several clinically significant mRNAs such as *NRAS*, *ZIC1*, *BCL-2*, *TRf2* and *MT3-MMP* (Kumari et al., 2007a; Arora et al., 2008; Morris and Basu, 2009; Gomez et al., 2010; Shahid et al., 2010). Comprehensive analyses of six different GQ forming sequences in the 5'-UTR region in mRNA amongst many potential GQ forming sequences in the transcriptome suggested that in general, the GQ structures act as translation repressors (Beaudoin and Perreault, 2010). Alternatively, when present in the context of an IRES (Internal Ribosomal Entry Site), for example, in

the cases of *FGF* and *VEGF* the GQ structures act as essential elements for translation initiation (Bonnal et al., 2003; Morris et al., 2010). Nevertheless, the mechanism by which the GQ structures play such context-dependent regulatory role is unknown.

Translation initiation by IRESs involves a cap-independent mechanism wherein the 40S ribosomal subunit or other IRES trans-acting factors (ITAFs) are recruited directly onto the mRNA, thereby rendering the requirement of the 5'-cap and some initiation factors unessential (Komar et al., 2012). The IRES-mediated translation initiation, although initially observed in viral mRNAs, has also been identified in many cellular mRNAs (Hellen and Sarnow, 2001; Jackson et al., 2010b). This non-canonical translation initiation was primarily observed in viral mRNAs and thereafter in several cellular mRNAs. The fundamental difference between these two mechanisms is the requirement for a 5'-cap that is added onto the mRNAs prior to translation. The eukaryotic cytoplasmic mRNAs go through an enzymatic post-transcriptional modification known as the capping of the 5' end of the mRNA. The mRNA cap structure plays a critical role in translation initiation, RNA processing, and nuclear export (Lewis and Izaurflde, 1997). The mitochondrial and the chloroplast mRNAs are generally not capped (Monde et al., 2000; Temperley et al., 2010).

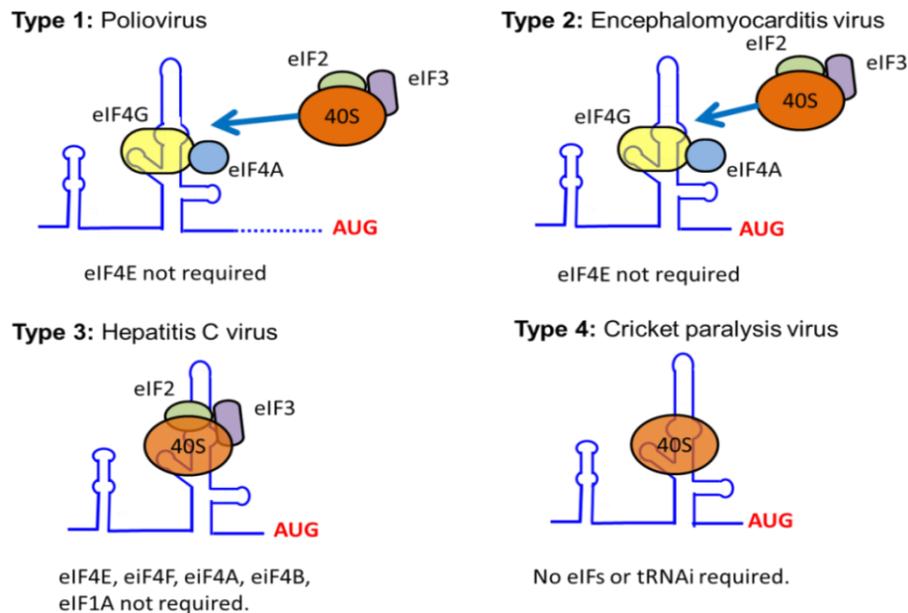
The general structural features of the 5' cap of eukaryotic of mRNAs are the terminal m<sup>7</sup>G and the penultimate nucleotide are joined by an inverted 5'-5' linkage through a triphosphate bridge (Fig. 2.1). The canonical method of translation initiation requires a 5'-cap for orchestrating the assembly of translation initiation machinery. It was

observed that several viral mRNAs were capable of initiating translation even in the absence of the 5'-cap. These RNAs bypassed the canonical initiation process by recruiting the 40S subunit of the ribosomes in a highly structured region in their 5'-UTR. These regions were termed as internal ribosome entry sites (IRESs) and the process was termed IRES-mediated translation initiation or cap-independent translation initiation [163].



**Figure 2.1- The 5'-cap structure:** The base of the first nucleotide in the mRNA is a guanosine methylated at the seventh position and joined to the next nucleotide by a non-canonical 5'-5' triphosphate link.

The occurrences of IRESs were initially observed in two picornaviruses namely the Poliovirus (PV) and Encephalomyocarditis virus (EMCV) (Jang et al., 1988; Pelletier and Sonenberg, 1988). Thereafter IRESs have been discovered in many other viral mRNAs such as Hepatitis C virus (HCV)(Tsukiyama-Kohara et al., 1992), Foot and Mouth Disease virus (FMDV)(Martinez-Salas et al., 2001), classic Swine Fever virus(Rijnbrand et al., 1997), and Human Immunodeficiency virus envelope mRNA (Locker et al., 2011). Although the mechanism by which the viral IRESs initiate translation varies, the common theme amongst all these is their ability to recruit the 40S ribosomal subunit internally regardless of the presence or absence of the 5'-cap. The mode of translation initiation by the viral IRESs can be broadly classified into four types (Fig. 2.2).



**Figure 2.2- The classification of viral IRESs:** Schematic representation of the four different mechanisms of cap-independent translation initiation by the viral IRESs (Jackson et al., 2010b).

Thereafter several IRESs in cellular mRNAs with a range of activities were also discovered which included several growth factors such as fibroblast growth factor (FGF), (Bonnal et al., 2003; Martineau et al., 2004) platelet-derived growth factor B (PDGF-B) (Bernstein et al., 2002), vascular endothelial growth factor (VEGF) (Stein et al., 1991) and oncogenes such as C-MYC (Stoneley et al., 1998; Stoneley et al., 2000), and protein kinase p58 (Tinton et al., 2005). The presence of IRESs in cellular mRNAs allows the cell to modulate its translation and the protein repertoire in response to stress, cell cycle and other stimuli (Plank and Kieft, 2012). It has been suggested that the cellular mRNAs follow the picornavirus initiation mechanism (Jackson et al., 2010b). More detailed investigations of the cellular IRES mechanism are required before we can attribute a common mechanism to all cellular mRNA IRES-mediated translation initiation.

The 5'-UTR of human vascular endothelial growth factor (*hVEGF*) encompasses two IRES elements (Huez et al., 1998). The human VEGF is a key physiological and pathological angiogenic growth factor. An increase in VEGF protein level is not only linked to normal physiological conditions, such as embryonic development, wound repair, adaptation to hypoxia but also in pathological conditions such as proliferative retinopathies, arthritis, psoriasis, and tumor angiogenesis (Ferrara and Davis-Smyth, 1997; Bastide et al., 2008; Arcondeguy et al., 2013). The 5'-UTR of *hVEGF* contains two independently functioning IRESs (A and B) of which the IRES A is a 294-nucleotides (nt)-long fragment (745 to 1038 from 5'-end of the mRNA) immediately upstream of the canonical AUG translation start site (Huez et al., 1998; Miller et al., 1998; Stein et al., 1998). The VEGF IRESs were shown to play a key role in upregulation of VEGF under

hypoxic stress (Stein et al., 1998; Bornes et al., 2007). The presence of a tunable GQ structure in the IRES A was determined to be necessary for optimum cap-independent translation initiation (Morris et al., 2010). Translation initiation by IRESs involves a cap-independent mechanism wherein the 40S subunit and other IRES trans-acting factors (ITAFs) are recruited directly onto the mRNA, rendering the 5'-cap and some of the initiation factors to be non-essential (King et al., 2010; Perard et al., 2010; Komar and Hatzoglou, 2011; Martinez-Salas et al., 2012).

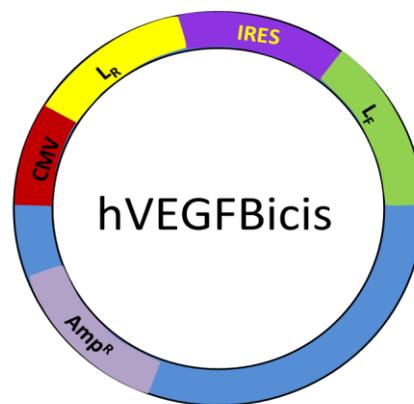
Our investigations on structure mapping of the *hVEGF* IRES A revealed that it forms a very stable and well-defined secondary structure under physiologically relevant salt concentrations. Considering the essential role of the GQ domain in IRES A function, we hypothesized that it plays a key role in IRES A mediated translation initiation by interacting with the 40S subunit.

## **2.2 MATERIALS AND METHODS**

### **2.2.1 Plasmid constructions**

The transcribable plasmid pVFIRESA and bicistronic plasmids hVEGFbicis containing the wild-type IRES A sequence were constructed as previously described (Morris et al., 2010). The schematic representation (Scheme 2.1) of the hVEGFbicis plasmid shows the design of the plasmid wherein the firefly luciferase gene is regulated by the IRES-mediated translation initiation and the *Renilla* luciferase is translated by the canonical translation initiation. The primers for deletion mutation were purchased from

Integrated DNA Technologies, Inc. The deletions of the specific domains were performed by QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent Tech) as per manufacturer's protocol. Plasmids were purified by Pure Yield™ Plasmid Miniprep System (Promega), and successful deletion was confirmed by sequencing at The Plant Microbe Genomic Facility (OSU).



**Scheme 2.1- Design of the hVEGFBicis:** The plasmid was designed to transcribe a single mRNA with both the reporter genes. The *Renilla* luciferase (L<sub>R</sub>) is translated by a canonical cap-dependent translation initiation. The translation of the second reporter gene firefly luciferase (L<sub>F</sub>) can only be initiated by an actively functioning IRES.

### 2.2.2 *In vitro* transcription and radiolabeling of IRES A and mutant RNAs

The 294 nt long IRES A and the mutants of the *hVEGF* mRNA were transcribed using plasmid pVFIRESA and its mutant variants (Morris et al., 2010). The transcribed RNA was purified by 6 % denaturing polyacrylamide gel electrophoresis (PAGE). The RNA bands were harvested by tumbling the gel slices at 4 °C in elution buffer (300 mM NaCl, 10 mM Tris-HCl (pH 7.4), and 0.1 mM EDTA). The eluent was phase

concentrated using 2-butanol and subsequently precipitated from the aqueous phase by ethanol.

The 5'-phosphates of the *in vitro* transcribed RNA were enzymatically removed by Calf intestinal alkaline phosphatase (CIP, NEB). The RNA was extracted by phenol-chloroform and precipitated with 70% ethanol. The CIP treated RNA was 5'-end radiolabeled by T4 polynucleotide kinase (PNK, NEB), [ $\gamma$ - $^{32}$ P] ATP (Perkin Elmer) and incubated for 1 hour at 37 °C. The commercially obtained DNA sequences were directly radiolabeled with T4 PNK as the 5'-phosphate is absent. The reaction was stopped by the addition of an equal volume of stop buffer (7 M urea, 10 mM Tris-HCl, pH 7.5 and 0.1 mM EDTA). The radiolabeled full-length RNA was isolated by 6 % denaturing PAGE. The radiolabeled RNA was then extracted from the gel via the crush and soak method as described previously.

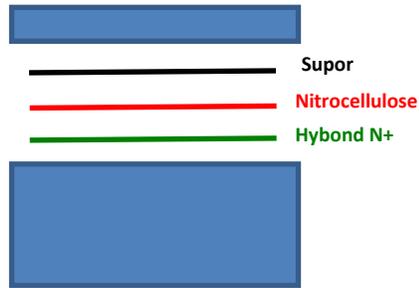
### **2.2.3 Structure mapping analyses**

The 5'-end radiolabeled RNA (60000 cpm) was dissolved in 10 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, 2.5 mM MgCl<sub>2</sub> in the presence of 100 mM KCl or 100 mM LiCl and heated to 70 °C for 10 minutes and slow cooled to room temperature. Once reactions attained the appropriate temperature, the RNA was digested with 0.1 U of Ambion® RNase T1 (Life Technologies), 0.005 µg of Ambion® RNase A for 5 minutes at 37 °C, 1.4×10<sup>-4</sup> U of RNase One™ (Promega) for 5 minutes at 37 °C. The reactions were terminated by using an equal volume of stop buffer (7 M urea, 10 mM Tris-HCl, pH 7.5,

and 0.1 mM EDTA). Treated RNA was electrophoresed on a 6 % denaturing polyacrylamide gel electrophoresis (PAGE, dried on Whatman paper, and exposed to a phosphorimager screen. The gel images were then visualized by scanning the screen on a Typhoon Phosphorimager FLA 9500 (GE Healthcare, Life Sciences).

#### **2.2.4 Filter binding assay**

The filter binding assay for assessing the apparent binding constants of the interaction of 40S ribosomal subunit was performed using a previously published protocol (Kieft et al., 2001). The salt washed 40S ribosomal subunits isolated from rabbit reticulocyte lysate (RRL) were a kind gift from Dr. W. C. Merrick (Case Western Reserve University) and 30S ribosomal subunits (bacterial) were kindly provided by Dr. S. A. Strobel (Yale University). The radiolabeled RNAs (1000 cpm) with an approximate concentration of  $1 \times 10^{-3}$  nM were folded in the presence of the binding buffer (20 mM Tris-HCl, 100 mM CH<sub>3</sub>COOK, 200 mM KCl, 2.5 mM MgCl<sub>2</sub>, 1 mM DTT). The RNAs were then incubated with increasing concentrations of 40S subunit at 37 °C for 20 minutes. The filter binding assay was performed using a Supor® Membrane (0.45 µm) on the top of a Nitrocellulose filter which in turn was placed on top of a *Hybond*<sup>TM</sup>-N+ filter supported by a filter paper. The Supor membrane was added to remove any aggregate, however, none was observed. The filters were presoaked in binding buffer assembled in a Minifold-I Dot-Blot System.

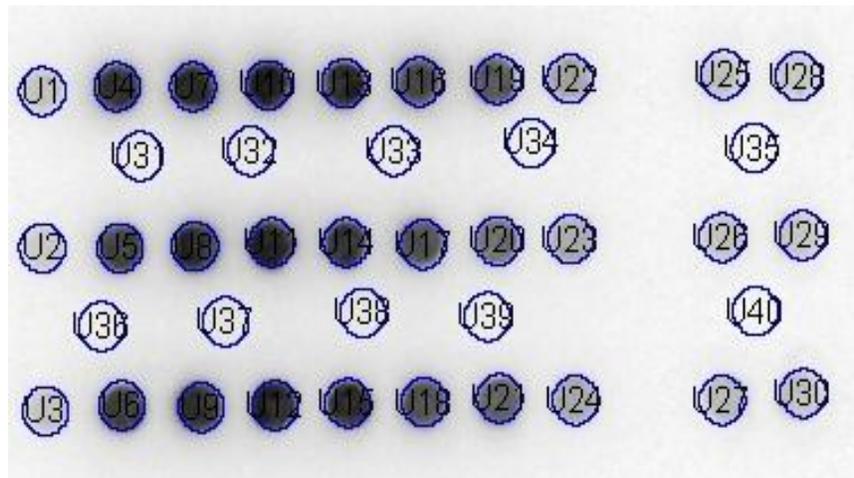


**Scheme 2. 2: Arrangement of the membranes in the dot-blot apparatus**

The reaction mixtures were added to the wells of the Dot Blot apparatus under vacuum. The membranes were air dried and exposed to a phosphorimager screen. The blot images were then visualized by scanning the screen on a Typhoon Phosphorimager FLA 9500 (GE Healthcare, Life Sciences).

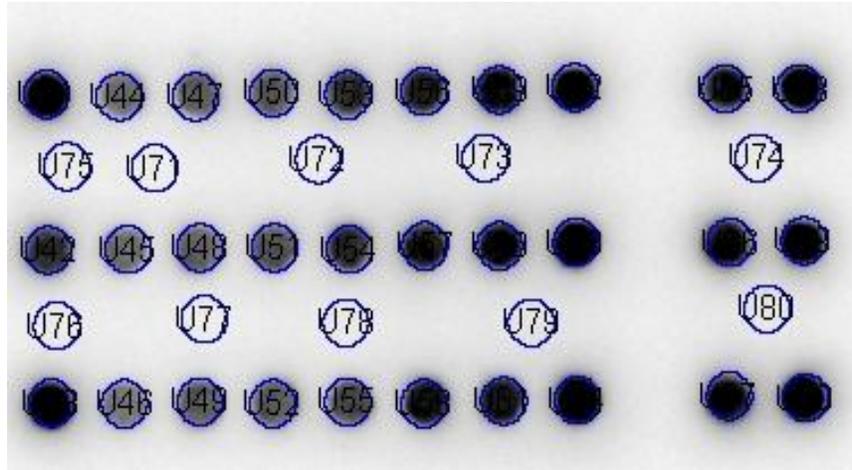
(a)

|   |     |     |    |      |     |      |      |  |       |       |
|---|-----|-----|----|------|-----|------|------|--|-------|-------|
| 0 | 504 | 252 | 99 | 49.5 | 9.9 | 4.95 | 0.99 |  | 0.504 | 0.099 |
|---|-----|-----|----|------|-----|------|------|--|-------|-------|



(b)

|   |     |     |    |      |     |      |      |  |       |       |
|---|-----|-----|----|------|-----|------|------|--|-------|-------|
| 0 | 504 | 252 | 99 | 49.5 | 9.9 | 4.95 | 0.99 |  | 0.504 | 0.099 |
|---|-----|-----|----|------|-----|------|------|--|-------|-------|



**Figure 2.3: Imaging of the filter binding assay:** (a) Image of the Nitrocellulose filter paper. U31-40 were used to calculate the background. The concentration of the 40S in nM in the table above. (b) Image of the Hybond N+ filter paper. U75-80 were used to calculate the background.

The fraction of RNA bound ( $\theta$ ) was calculated as follows:

Fraction bound ( $\theta$ ) = RNA in nitrocellulose/ (RNA in nitrocellulose+ RNA in hybond N+).

We calculated the apparent  $K_d$  by two different equations. Firstly, the data were fit to a Langmuir isotherm described by the equation

$$\theta = [X] / ([X] + K_d) \quad \text{Equation 2.1}$$

in Origin, where  $\theta$  is a fraction of RNA bound to 40S as measured from the fraction retained on the nitrocellulose membrane and  $[X]$  is the 40S subunit concentration. The assumption is that 40S concentration is much greater than the concentration of RNA. The  $K_d$  is given by the concentration of the 40S where the value

fraction bound ( $\theta$ ) is equal to 0.5. The fraction bound did not reach 1 and can be a limitation of the method. But we did observe the plateau suggesting saturation of the RNA. Because of the experimental situation, we have termed the  $K_d$  as apparent  $K_d$ . The results of the  $K_d$  determination by this equation are included in Appendix I along with method 2 and a comparison of the values.

Furthermore, we also determined the  $K_d$  by using the equation

$$\theta = Y_{\max} [X] / ([X] + K_d) \quad \text{Equation 2.2}$$

plotted in Origin assuming the cooperativity ( $n=1$ ), that there was no cooperativity. The  $Y_{\max}$  here gives the filter binding efficiency. We ignored the values of  $Y_{\max}$  less than 0.5 since at the highest experimental concentration of the 40s the fraction bound ( $\theta$ ) was not even 0.5.

### **2.2.5 Footprinting of 40S ribosomal subunit interaction with the IRES A**

The 5'-end radiolabeled RNA (150, 000 cpm) was dissolved in 30 mM HEPES (pH 7.5), 2.5 mM  $MgCl_2$  and 100 mM KCl and heated to 70 °C for 10 minutes and slow cooled to room temperature. The folded RNA was then incubated with 100 nM of 40S ribosomal subunit at 37 °C for 20 minutes. The RNAs were then treated with 0.1 U of RNase T1 for 3 minutes at 37 °C and the reaction was stopped by adding 1% SDS, and then phenol-chloroform extracted followed by ethanol precipitation. The precipitated RNA was then dissolved in formaldehyde loading buffer, counted by a scintillation

counter and normalized for uniform loading in each well, electrophoresed on a 6 % denaturing PAGE, dried on Whatman paper and exposed to a phosphorimager screen. The gel images were then visualized by scanning the screen on a Typhoon Phosphorimager FLA 9500 (GE Healthcare, Life Sciences).

### **2.2.6 Cell culture and transfection of reporter plasmid**

HeLa cells were grown in 96-well plates in Dulbecco's modified Eagle's medium (DMEM) with low glucose supplemented with 10 % fetal bovine serum and 1 % antibiotics at 37 °C in 5% CO<sub>2</sub> in a humidified incubator. The plasmids were transfected with Lipofectamine-2000 for 6 hours. After 24 hours, dual luciferase assay was performed with Dual-Glo® Luciferase Assay System as per manufacturer's protocol and the luminescence signals read by the SpectraMax M4 plate reader (Molecular Devices, LLC). The ratio of the Firefly and *Renilla* luciferase luminescence were calculated and normalized to the ratio of the wild-type plasmid.

## **2.3. RESULTS**

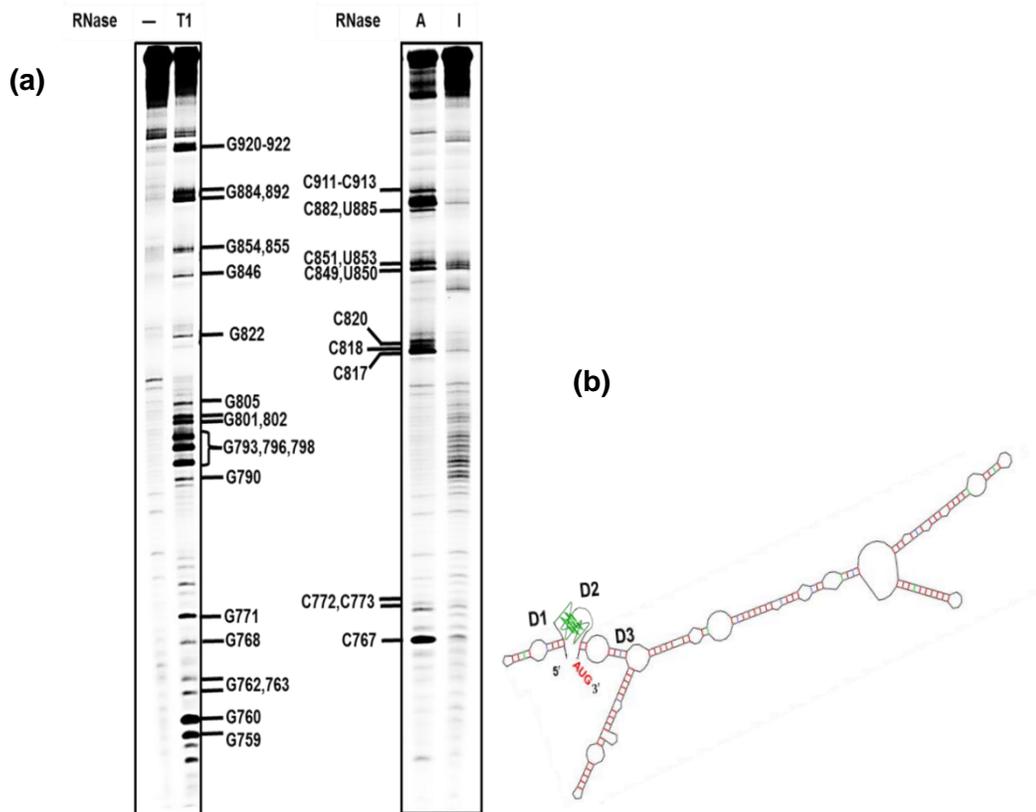
### **2.3.1 Secondary structure mapping of the IRES A by enzymatic footprinting.**

The IRES A forms a very stable secondary structure since is highly enriched with G and C residues (~76 %). A more detailed analysis of the secondary structure of the IRES A was conducted via a series of enzymatic structure mapping (Fig. 2.4 a). The

updated secondary structure of the IRES A that incorporated the constraints derived from structure mapping data was created with the help of Mfold (Zuker, 2003) with the 40S subunit interacting region traced on the folded RNA structure (Fig. 2.4 b). We determined that the GQ is a stable structure formed within the 40S subunit binding region given the strong protection detected in enzymatic (Fig. 2.5 d) and chemical structural mapping (Morris et al., 2010).

### **2.3.2 The 40S ribosomal subunit interacts with the 5'-proximal end of IRES-A.**

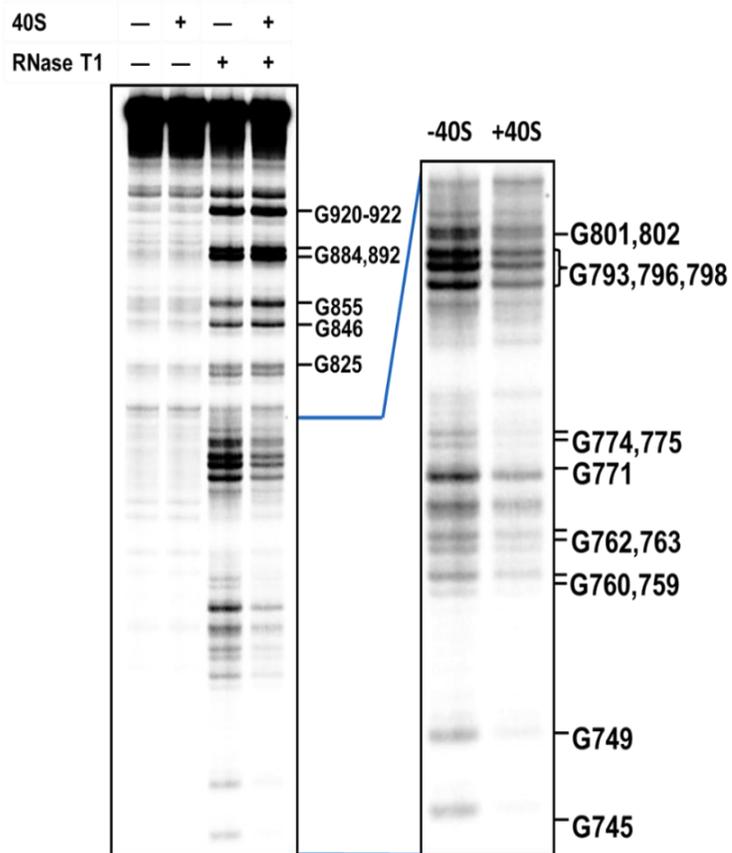
The predominance of guanine (G) residues in the 294 nt long IRES A constituting about 45 % of the total number of residues rendered RNase T1 to be the enzyme of choice for structure mapping and footprinting analyses. RNase T1 cleaves single-stranded G residues in RNA, however, the G residues involved in base pairing to form secondary structures, GQ structures, and tertiary interactions are protected from undergoing cleavage (Takahashi K., 1982). Mapping of IRES A-40S ribosomal subunit interaction by enzymatic footprinting clearly indicates a region of about 60 nt protected from RNase T1 cleavage in the 5'-proximal end of the 294 nt long IRES A (Fig. 2.5 a & b).



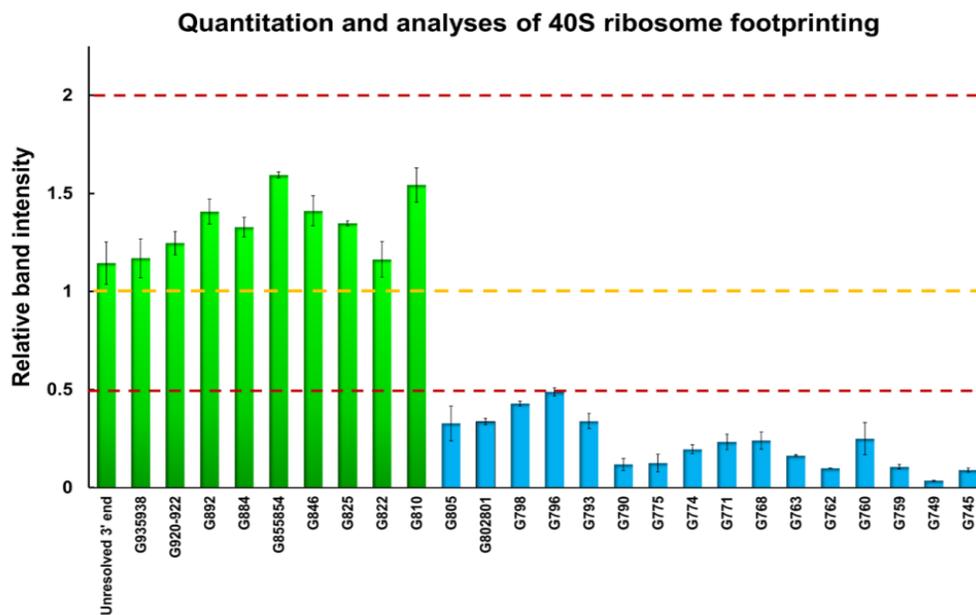
**Figure 2.4- The Secondary structure of IRES A:** (a) A detailed enzymatic structural mapping to deduce the secondary structure of IRES A. No enzyme (lane 1), RNase T1 (lane 2), RNase A (lane 3) and RNase One (lane 4) were used to probe the secondary structure. (b) The information was used to generate the secondary structure of the IRES A in human VEGF with the help of Mfold (RNA folding and hybridization software). The updated secondary structure of the hVEGF IRES A, including the GQ structure in domain 2 (D2).

Interestingly this region harbors the 17 nt long GQ forming region consisting of five stretches of G residues, which are resistant to RNase T1 cleavage in the presence of  $K^+$  (Fig. 2.5 d).

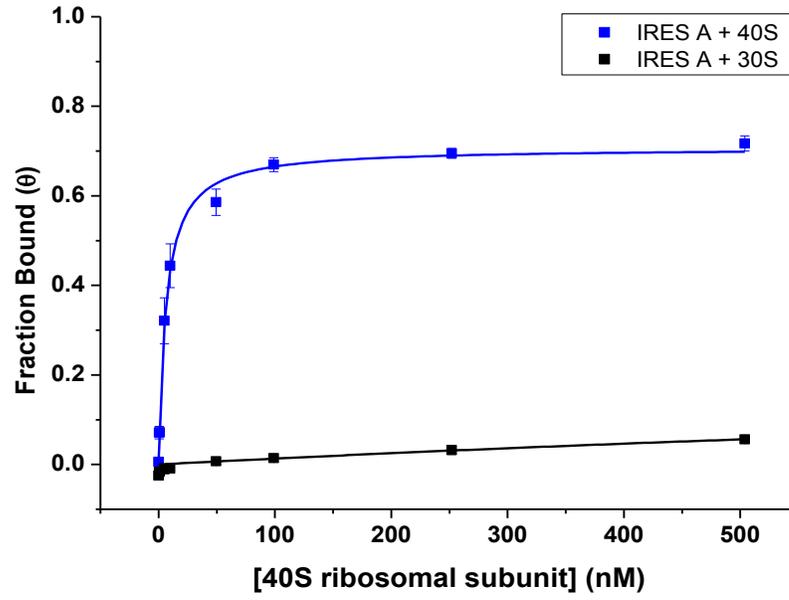
(a)



(b)

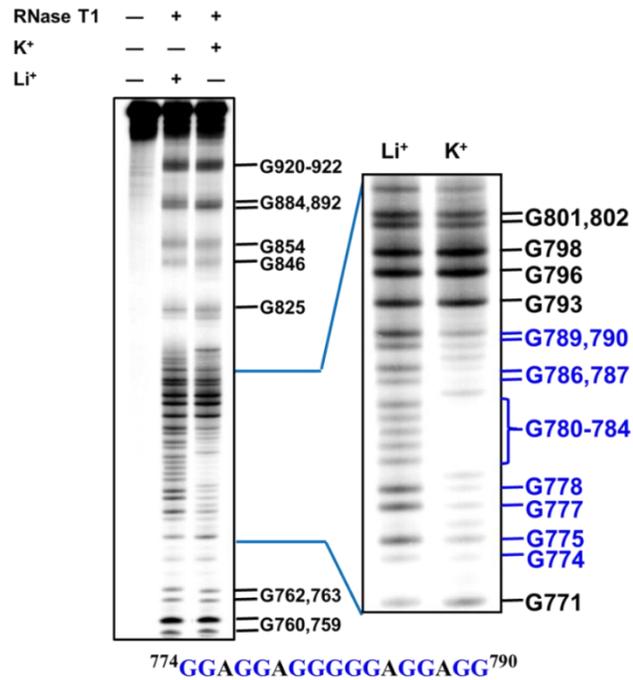


(c)



|                     | $Y_{\max}$  | $K_d$                           |
|---------------------|-------------|---------------------------------|
| <b>IRES A + 40S</b> | <b>0.71</b> | <b><math>6.3 \pm 0.6</math></b> |
| <b>IRES A + 30S</b> | <b>0.30</b> | <b>n. d</b>                     |

(d)



**Figure 2.5- The IRES A interacts with 40S ribosomal subunit:** (a) RNase T1 footprinting of IRES A in the presence of 40S subunit protects a segment of the IRES A from cleavage (745-805). The protected region is enlarged on the sides. (b) The bands were quantified by ImageQuant and normalized to the lane without 40S ribosomal subunit, which indicates the G residues protected from cleavage (blue bars). (c) The 40S ribosomal subunit binds to the IRES A with a strong affinity ( $K_d = 6.3 \pm 0.6$  nM) as observed by filter binding assay whereas there was no binding observed with 30S ribosomal subunit. (d) The 40S interacting region in IRES A comprises of a 17 nt sequence that adopts GQ structure at physiologically relevant  $K^+$  concentrations.

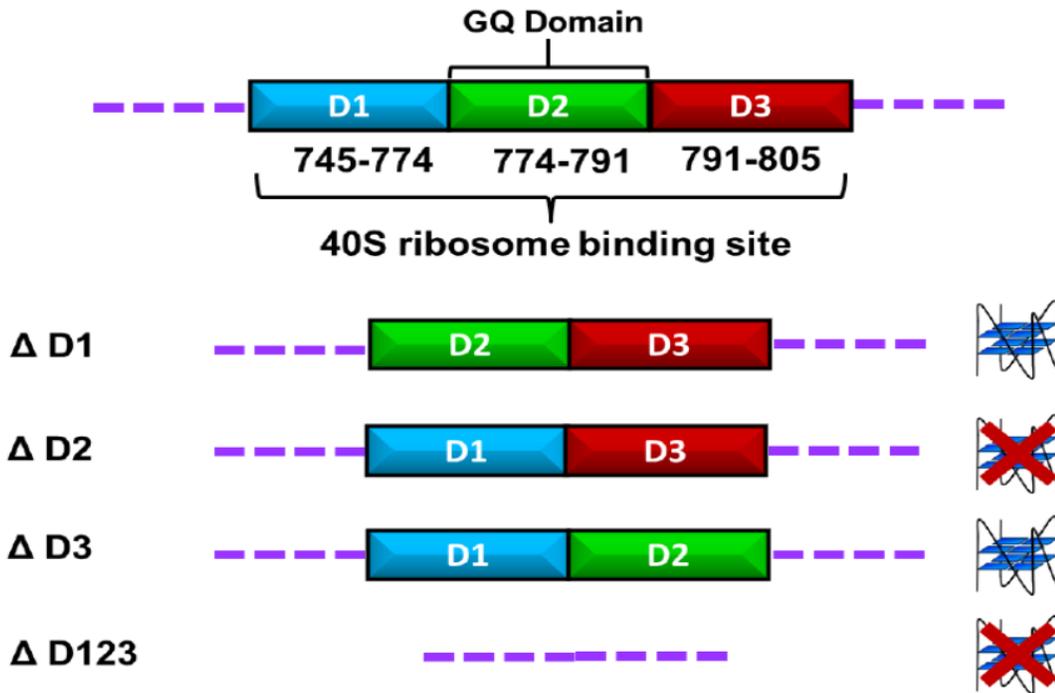
To investigate the binding affinity of the IRES A to the 40S ribosomal subunit we used filter-binding assay to determine the apparent equilibrium dissociation constant ( $K_d$ ) of 40S subunit-IRES A complex and used the bacterial 30S ribosomal subunit as a control. The IRES A was incubated with increasing concentrations (0.1 nM to 500 nM) of either the 40S or 30S salt washed ribosomal subunits. The RNA bound to the 40S or 30S ribosomal subunit stays with the nitrocellulose membrane while the unbound RNA binds to the positively charged membrane underneath. The equilibrium dissociation constant ( $K_d$ ) was calculated by plotting the fraction of bound RNA ( $\theta$ ) to the 40S subunit vs. the 40S subunit concentration and data were fitted to the equation 2.2. The  $K_d$  of the IRES A-40S subunit complex was determined to be  $6.3 \pm 0.6$  nM (Fig. 2.5 c)

Our results precisely trace the region of interaction of a cellular IRES and 40S subunit and their apparent binding affinity in the absence of other ITAFs. The direct association of the 40S subunit to an IRES has been observed only in some type of viral IRESs and our finding suggests that IRES A-mediated translation initiation is similar to that mechanism of initiation (Kieft et al., 2001; Kieft, 2008). Prior to this report, the potential interaction of 40S subunit to a cellular IRES was observed only in case of *c-Src*

kinase although its binding affinity and the precise region important for the interaction are not known (Allam and Ali, 2010). Overall, the cellular VEGF IRES A has the distinct ability to directly recruit the 40S ribosomal subunit with a low nanomolar binding constant.

### **2.3.3. The independently folding GQ domain in hVEGF IRES A determines the 40S ribosomal subunit recruitment**

Thereafter we sought to precisely identify the RNA segment that determines the IRES A-40S subunit interaction. To accomplish the goal, the putative 40S interacting region in the IRES A was divided into three different segments with the central one being the GQ forming region (D2) and its 5' flanking region (D1) and the 3' flanking segment (D3). Deletion mutants were prepared by site-directed mutagenesis of a transcribable plasmid pVFIRESA and used for *in vitro* transcription of the wild-type IRES A and the mutants  $\Delta D1$ ,  $\Delta D2$ ,  $\Delta D3$  and  $\Delta D123$  (Fig. 2.6). The apparent equilibrium dissociation constants ( $K_d$ ) of the 40S ribosomal subunits to the mutants were observed to be  $7.3 \pm 1.0$  nM for  $\Delta D1$  and  $8.4 \pm 1.1$  nM for the  $\Delta D3$  mutant respectively. The GQ forming region was determined to be the key determinant in the binding of the 40S subunit since the deletion of that domain (D2) resulted in extensive weakening of the binding, which precluded us from measuring the apparent  $K_d$  for the mutants  $\Delta D2$  and  $\Delta D123$ , indicating D2's indispensability in 40S subunit recruitment (Fig. 2.7).

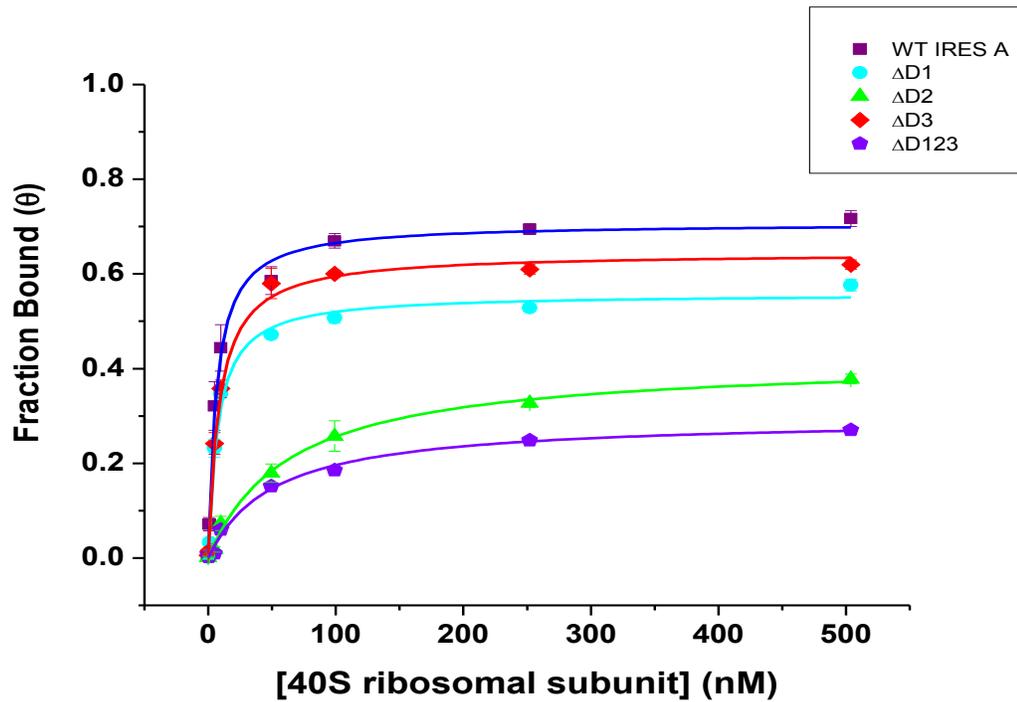


5'- G<sup>745</sup>CUAGCUCGGGCCGGGAGGAGCCGCAGCCG<sup>774</sup>GAGGAGGGGGAGG  
 AGG<sup>790</sup>AAGAAGAGAAGGAAG<sup>805</sup>AGGAGAGGGGGCCGCAGUGGCGACUCGGCGC  
 UCGGAAGCCGGGCUCAUGGACG<sup>858</sup>GGUGAGGCGGCGGUGUGCGCAGACAGUGC  
 UCCAGCCGCGCGCGCUCCCC<sup>907</sup>AGGCCUGGCCCGGGCCUCGGGCCGGGGAGG  
 AAGAGUAGCUCGCCGAGGCGCCGAGGAGAGCGGGCCGCCCCACAGCCCGAGC  
 CGGAGAGGGAGCGCGAGCCGCGCCGGCCCCGGUCGGGCCUCCGAAACCAUG-3'

**Figure 2.6- Nomenclature of the domains in the IRES A:** The 40S subunit interacting region of the IRES A was divided into three regions, wherein the central GQ forming domain was designated D2 (green). The 5'-region flanking the D2 was named D1 (blue) and the 3'- flanking region was named D3 (red). The 50 nt (858-907) long D4 (orange) region was previously reported to be essential for IRES-mediated translation initiation, is not a part of the 40S ribosomal subunit binding.

The filter binding results along with the footprinting confirm the interaction of the 40S ribosomal subunit specifically to a particular region in the IRES A and annuls the

possibility of non-specific interaction. The evidence suggests an unprecedented role of a GQ structure.

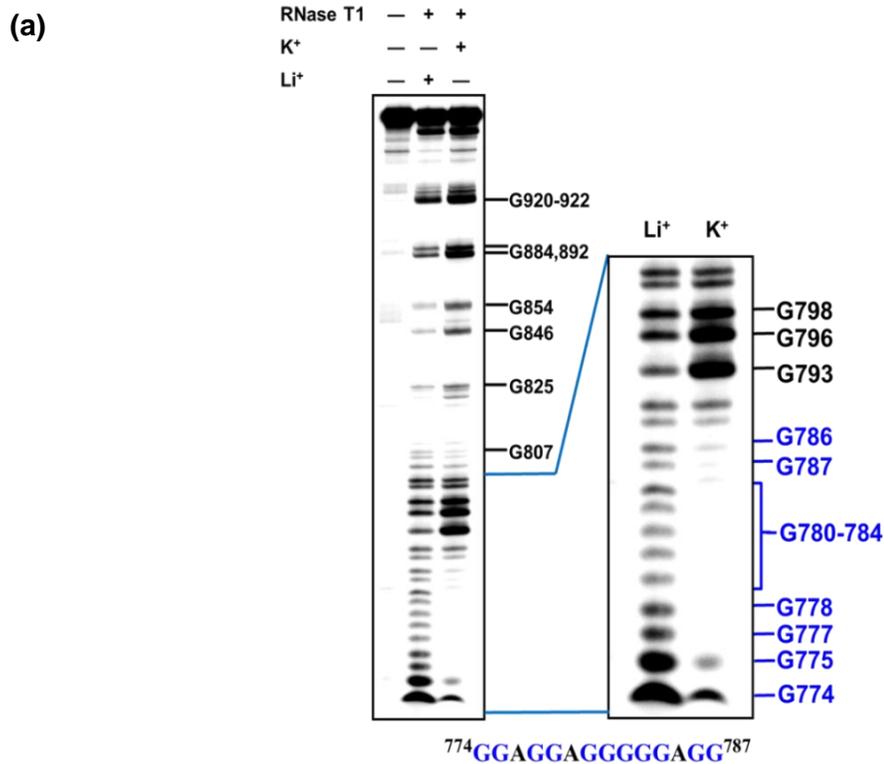


|               | $Y_{max}$ | $K_d$          |
|---------------|-----------|----------------|
| IRES A        | 0.71      | $6.3 \pm 0.6$  |
| $\Delta D1$   | 0.56      | $7.3 \pm 1.0$  |
| $\Delta D3$   | 0.64      | $8.4 \pm 1.1$  |
| $\Delta D2$   | 0.42      | $63.5 \pm 6.3$ |
| $\Delta D123$ | 0.30      | $50.8 \pm 6.6$ |

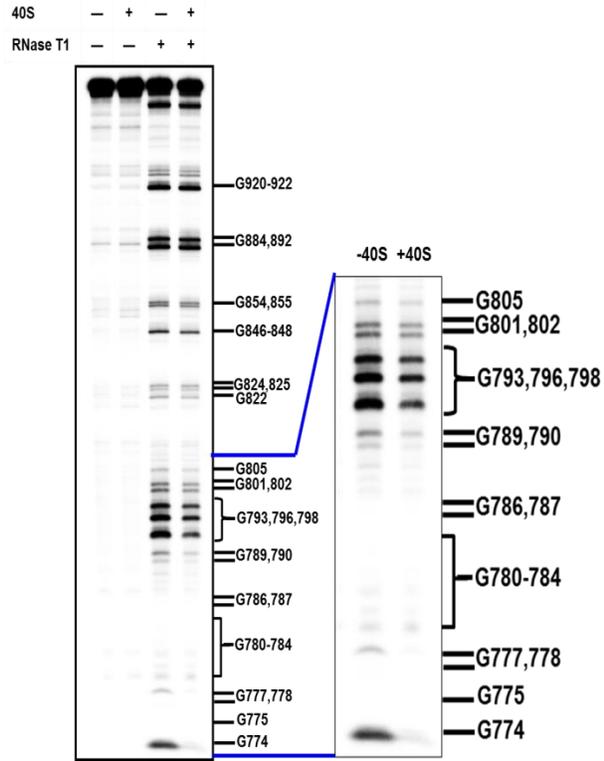
**Figure 2.7- The 40S ribosomal subunit binds to the GQ in IRES A:** The 40S ribosomal subunit dissociation constants ( $K_d$ ) as was calculated from the isotherms of the mutants indicate substantial loss of binding affinity in the absence of GQ i.e. D2 domain. The dissociation constants for  $\Delta D2$  and  $\Delta D123$  could not be determined (n.d) at the

highest 40S concentrations used for the assay. \* n.d indicates  $K_d$  could not be determined at 500 nM 40S ribosomal subunit concentration.

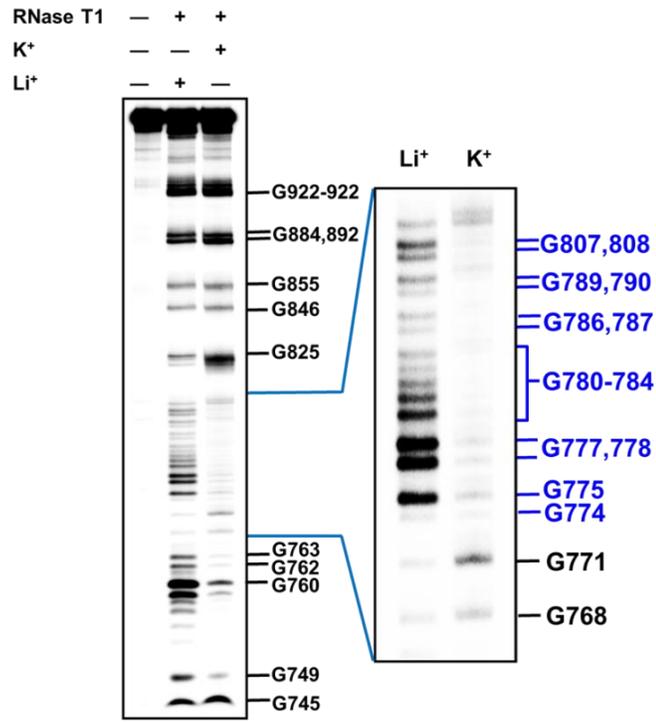
The formation of the GQ structure was confirmed in the mutants  $\Delta D1$  and  $\Delta D3$  by RNase T1 structure mapping in the presence of  $K^+$  and  $Li^+$ . The formation of the GQ structure was observed by its protection from the RNase T1 cleavage in the presence of  $K^+$ . The D2 domain's ability to form the GQ structure in spite of the deletion of one or both of the flanking regions (Fig. 2.8 a & b) established its true independence in terms of adopting the GQ structure.



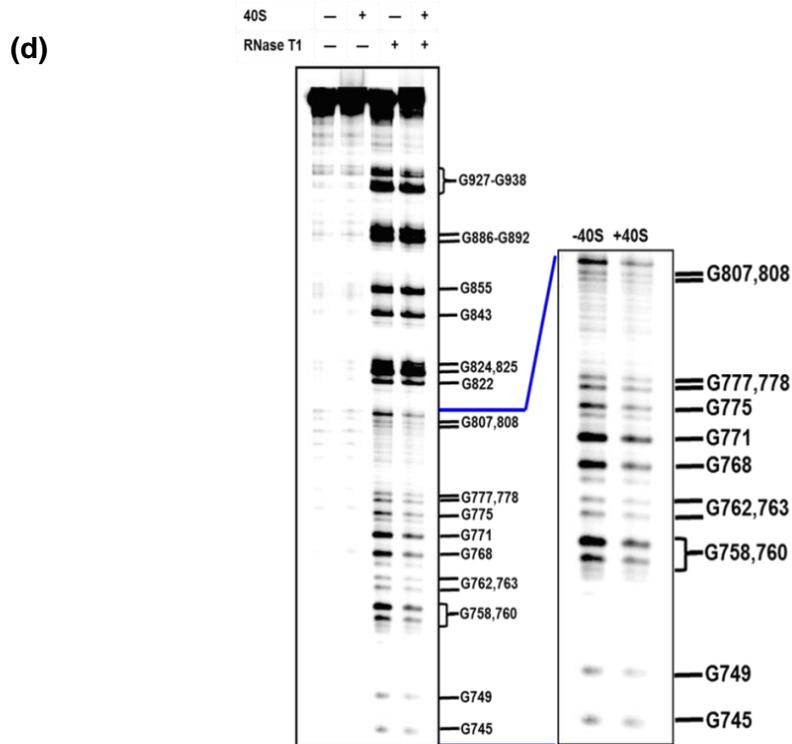
(b)



(c)



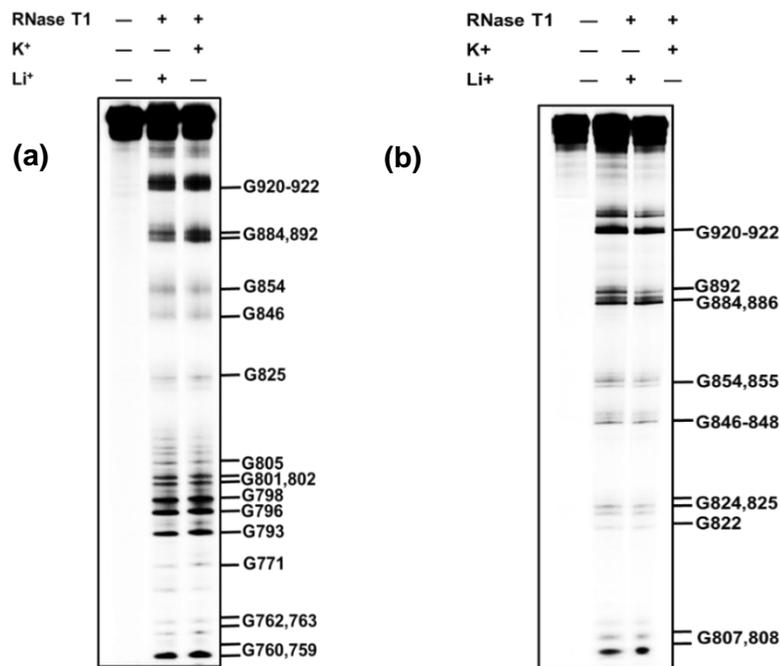
774 GGAGGAGGGGGAGGAGG 790 A 806 GG 808

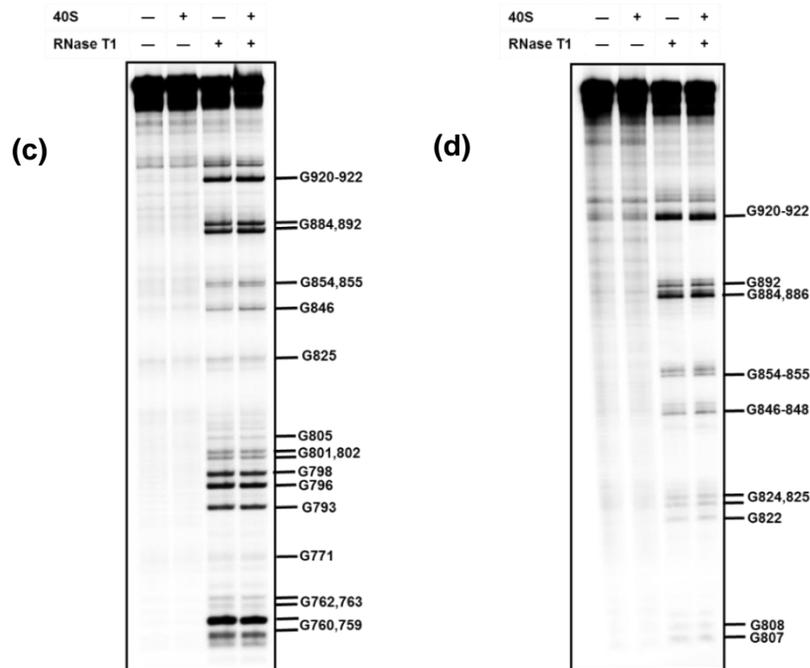


**Figure 2.8- Structure mapping and ribosome footprinting of  $\Delta$ D1 and  $\Delta$ D3:** The D2 domain forms a GQ structure despite the deletion of D1 and D3 domains in  $\Delta$ D1 (a) and  $\Delta$ D3 (c) mutants. Moreover, RNase T1 footprinting of  $\Delta$ D1 and  $\Delta$ D3 mutants in the presence of 40S subunits indicates that it interacts with the mutant IRESs in  $\Delta$ D1 and  $\Delta$ D3 (b & d).

The region of interaction of the 40S subunit with the mutants  $\Delta$ D1 and  $\Delta$ D3 was further confirmed by footprinting (Fig. 2.8 b & d). The region of protection indicated the area of interaction of the mutant IRESs with the 40S subunit and was consistent with the pattern observed for the wild-type IRES A except for the deleted segments. Additionally, mutants  $\Delta$ D2 and  $\Delta$ D123, which lacked the D2 domain, showed no significant structural differences between the  $\text{Li}^+$  versus the  $\text{K}^+$  treated samples (Fig. 2.9 a & b) and also showed no interaction with the 40S subunit (Fig. 2.9 c & d) due to the absence of the key GQ domain. The structure mapping of D2 domain also suggests that there was no

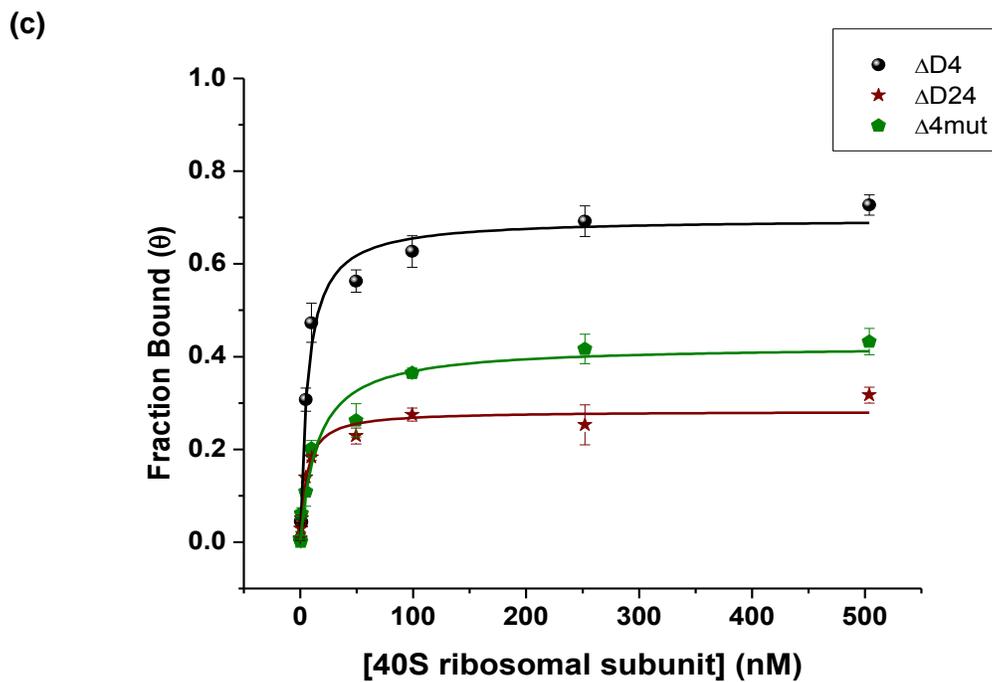
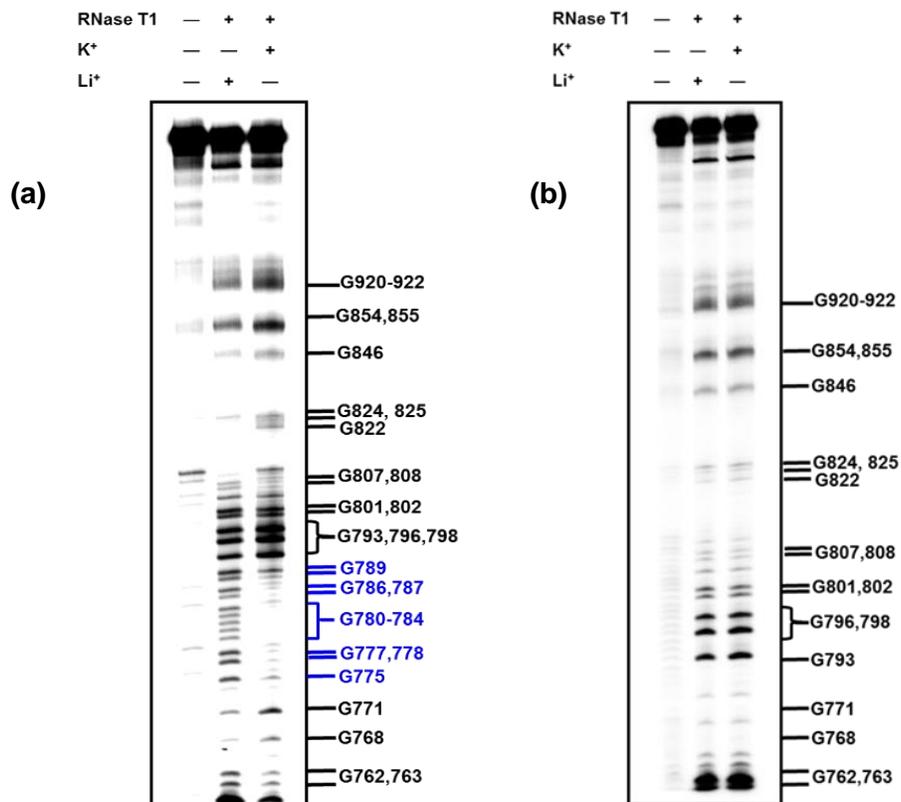
significant change in the overall secondary or tertiary structure of IRES A based upon the relatively unchanged nature in the band pattern. Interestingly, a previously identified 50 nt region (858-907) termed D4, was reported to be important for optimal IRES activity (Bonnal et al., 2003). We ascertained to be not involved in the 40S recruitment (Fig. 2.10 c) as the mutant  $\Delta D4$  interacts with the 40S subunit with an affinity ( $K_d = 8.5 \pm 1.6$  nM) which is similar to the  $K_d$  value of the wild-type IRES A (Huez et al., 1998). However, deletion of the D2 domain in the  $\Delta D4$  mutant to form the  $\Delta D24$  practically eliminated its ability to interact with the 40S subunit (Fig. 2.10 c). Notably, the deletion of the 50 nt D4 domain did not affect the formation of the GQ structure (Fig. 2.10 a). We also verified the 40S ribosome binding affinity of a previously described mutant wherein four G residues in the D2 domain were mutated to uridines in order to disrupt the G-stretches required for adoption





**Figure 2.9- Structure mapping and ribosome footprinting of  $\Delta$ D2 and  $\Delta$ D123:** Deletion of D2 domain resulted in no significant difference in their overall secondary structure (a & b) between the lanes containing  $\text{Li}^+$  and  $\text{K}^+$ . The 40S subunit footprinting also showed no protected region in either  $\Delta$ D2 or  $\Delta$ D123 (c & d).

of a GQ structure. The mutant which was previously established not to adopt a GQ structure was unable to bind to the 40S ribosomal subunits further confirming the requirement of the GQ structure for interaction with the 40S (Fig 2.10 c) (Morris et al., 2010). Importantly, we established that the D2 segment folds: i) independently to form the GQ structure despite the deleted flanking regions, ii) independent of the overall IRES A secondary structure, and iii) is the key determinant of the 40S subunit interaction.



|                                | $Y_{\max}$  | $K_d$                             |
|--------------------------------|-------------|-----------------------------------|
| <b><math>\Delta D4</math></b>  | <b>0.71</b> | <b><math>8.5 \pm 1.6</math></b>   |
| <b><math>\Delta D24</math></b> | <b>0.28</b> | <b><math>4.6 \pm 0.1</math></b>   |
| <b>4 mutant</b>                | <b>0.45</b> | <b><math>24.6 \pm 10.1</math></b> |

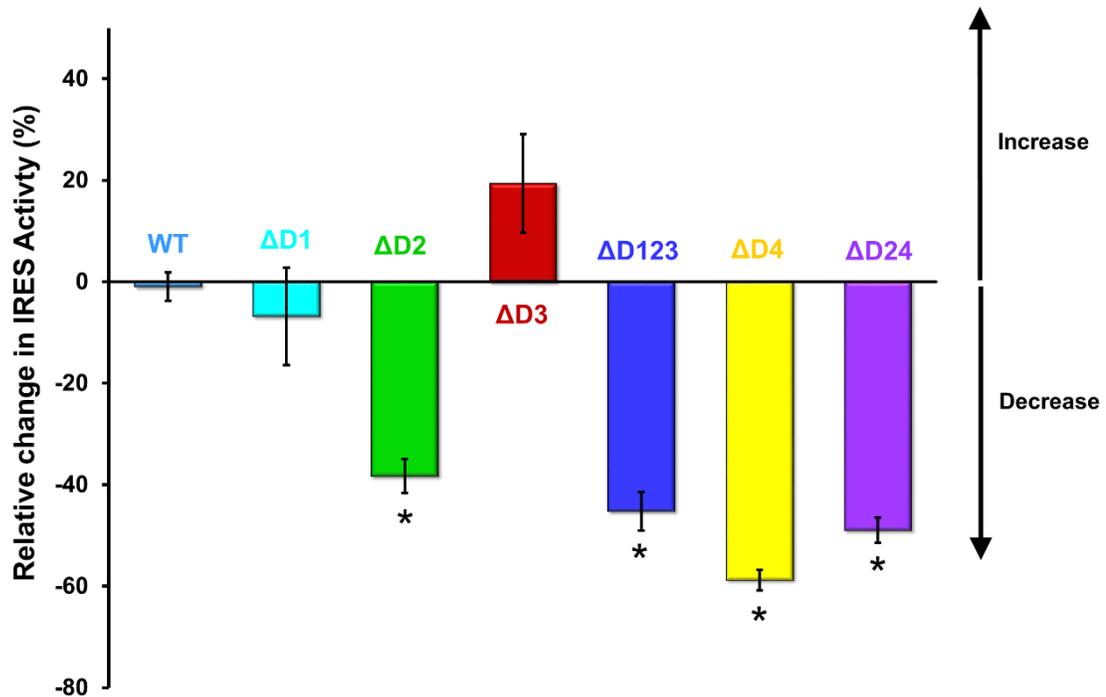
**Figure 2.10- Structure mapping and ribosome binding of  $\Delta D4$  and  $\Delta D24$ :** (a) Indicates the structure mapping of the  $\Delta D4$  mutant and the formation of GQ structure by the G-residues in the D2 domain highlighted by blue color. Simultaneous deletion of both D2 and D4 (b) in  $\Delta D24$  mutant disables GQ formation as observed by structural mapping. (c) The D4 region plays no role in interacting with the 40S subunit since the deletion of the domain did not alter the binding affinity significantly. However, the deletion of both D2 and D4 domains impaired the 40S-IRES binding completely.

### 2.3.4 The D2 domain is essential for optimal IRES A-mediated translation initiation

The next key question that needed to address was the functional relevance of the GQ-40S interaction and the role of different regions of IRES A in the regulation of cap-independent translation. To test that we constructed IRES A mutants similar to the ones described above but in the context of the dual luciferase bicistronic plasmid hVEGFbicis (Morris et al., 2010). The plasmid was designed (Scheme 2.1) in a way such that one reporter gene (*Firefly luciferase*) was translated in a cap-independent manner by IRES A whereas the other reporter gene (*Renilla luciferase*) initiates translation in the canonical pathway serving as the control (Morris et al., 2010; Bhattacharyya et al., 2014). Previously we knocked down eIF-4E, the cap-binding protein essential for canonical translation initiation and when bicistronic plasmid hVEGFbicis activity was measured, a sharp decrease in the *Renilla/Firefly* luciferase luminescence was observed. This

indicates that the cap-dependent translation initiation of the *Renilla* was inhibited specifically on knocking down eIF-4E, which in turn suggests the bicistronic nature of the construct (Bhattacharyya et al., 2014). It turned out that lack of the GQ domain (mutants  $\Delta D2$  and  $\Delta D123$ ) significantly decreased the IRES A activity (Fig. 2.11;  $p < 0.001$ ). The slight decrease in the activity of the  $\Delta D1$  mutant could be explained by its observed peripheral involvement in 40S binding. We also observed an increase in cap-independent translation in the case of the  $\Delta D3$  mutant despite the deletion of the entire region responsible for 40S binding. In previous reports deletion of functional elements has been observed to only partially disrupt the activity and in some cellular IRESs was found to upregulate the activity (Willis and Stoneley, 2004). Furthermore, deletion of the D4 domain yielded significant translation repression compared to the wild-type IRES A as was reported previously, and so did the deletion of both the D2 and D4 domains (Fig. 2.11).

Interestingly the decrease in translation by the mutants  $\Delta D4$  and  $\Delta D24$  were not significantly different from each other. Although the deletion of D2 and D4 resulted in repression of translation initiation by 40% and 60 % respectively. The effect was not observed to be additive when both the regions were deleted. It is possible that although the D2 domain is primarily responsible for 40S ribosomal subunit



**Figure 2.11- Relative IRES activity of the mutants:** The D2 and D4 domains are essential for IRES-mediated translation: HeLa cells were transfected with the bicistronic plasmid and the reporter gene activity measured 24 hours post-transfection. The IRES A activity was significantly ( $p < 0.001$ ) decreased due to deletion of the GQ forming D2 domain from the sequence.

recruitment whereas the D4 domain recruits another ITAFs and hence their effect is, therefore, the IRES-mediated translation initiation is at its maximum with the presence of both the domains and deletion of either of them impairs translation initiation.

Our structure mapping analyses also suggest that variable numbers of G-tracts are utilized to form the GQ in  $\Delta D1$ , WT, and  $\Delta D3$  (Fig 2.5 d, 2.8 a and c). These observations indicate the tunability of the GQ might also be of some importance along with the functional elasticity of the IRES and perhaps the role played by yet to be identified regulatory elements within the hVEGF IRES. Overall, our data indicate that the

independently folding GQ domain is essential for direct interaction of the IRES A with the 40S subunit and for optimal IRES function.

## **2.4 DISCUSSION**

It has been suggested that stable GQ structures in 5'-UTR interfere with the formation of the 43S pre-initiation complex or interrupt the scanning mechanism (Bugaut and Balasubramanian, 2012) and in the coding region, GQs were shown to stall the ribosome movement along the mRNA leading to premature termination of translation and consequently inhibit protein synthesis (Endoh et al., 2013a). Thus, the GQs in mRNAs are almost exclusively translation inhibitors except when present in the context of a cellular IRES element in 5-UTR whereas it plays an essential role. However, the underlying mechanism of how the GQ structures play an essential role in the context of the cellular IRES was not understood. Our findings explain the hitherto undefined mechanistic role of a GQ structure present within a cellular IRES in the cap-independent translation initiation. The presence of about 3000 putative GQ forming sequences in the 5'-UTR poses the question if this interaction is generic for all GQ structures (Kumari et al., 2007a; Huppert et al., 2008). Our findings, when taken in conjunction with the current knowledge on cap-independent translation initiation in cellular IRESs, indicate that the 40S subunit interactions with the GQ structure appear to be of direct functional relevance, only in the context of an IRES function. The association of the 40S to GQ structures alone may not warrant canonical translation initiation due to the requirement of other

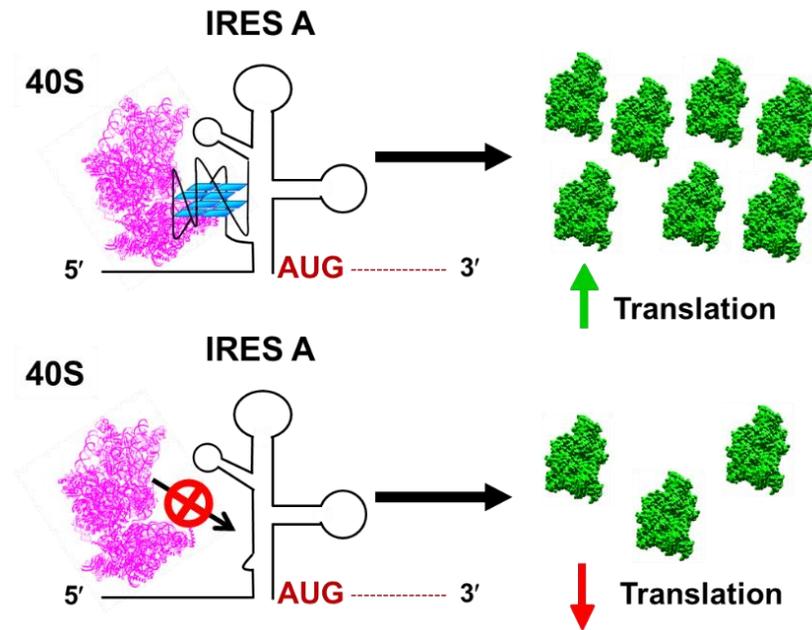
factors for such a process. The discovery of GQ structures in the 5'-UTR, that augments translation of TGF $\beta$ 2 which is not known to harbor any IRES, further complicates the mechanistic paradigm of the GQ structure (Agarwala et al., 2013).

There is yet to be any identification of a common sequence or structural motif that can predict the presence of cellular IRESs. In several instances the canonical stem-loop secondary structures in viral and cellular IRESs have been directly implicated for their role in 40S ribosomal subunit binding, however, a non-canonical structural domain directly playing that role has not been reported previously to the best of our knowledge. The GQ domain in hVEGF IRES A is probably the singular domain identified in any IRES that not only folds into the GQ structure independent of the surrounding native sequences and structures but also appears to 'single-handedly' determine the interaction with the 40S ribosomal subunit. There are only two types of viral IRESs, Type 3 (HCV-like) and Type 4 (Dicistroviridae) that can directly interact with 40S subunit (Hellen, 2009). The structural details of those IRESs indicate that Type 4 IRESs have independently folding regions consisting of stem-loops and pseudoknots that determine the interaction with the 40S ribosomal subunit (Kieft, 2008; Jackson et al., 2010b). Thus, hVEGF IRES A most likely belongs to either Type 3 or Type 4 viral IRESs. It should be noted that although, functionally important, the pseudoknot domain alone is insufficient for recruiting the 40S subunit, unlike the GQ domain in IRES A. However, it is interesting that GQ and pseudoknot, two unrelated non-canonical RNA structures, are crucial for cellular and viral IRES functions respectively. Most of the viral IRESs except for Type 4 IRESs involve other canonical translation initiation factors for cap-

independent translation (Jackson et al., 2010b). Similarly in cellular IRESs of *c-myc* and *N-myc*, canonical translation initiation factors such as eIF-4A and eIF3 were observed to be essential for IRES-mediated translation initiation (King et al., 2010). Furthermore, several ITAFs that are observed to be heterogeneous ribonucleoproteins are known to facilitate cellular IRES-mediated translation (Komar and Hatzoglou, 2011). The requirement of the canonical translation initiation factors along with other ITAFs and their interaction with the GQ or other domains in the IRES A requires further investigation to elucidate the mechanism of IRES A-mediated translation initiation completely.

The post-transcriptional regulation of VEGF is highly complex and consists of almost all known regulatory processes, some of which are involved in cross-talking (Arcondeguy et al., 2013). The VEGF mRNA harbors two IRESs that are capable of initiating translation independent of each other (Huez et al., 1998). The cellular IRESs are known to be active under physiological conditions when the canonical translation initiation is impaired such as mitosis, hypoxia, nutritional stress and cell differentiation (Komar and Hatzoglou, 2011; Komar et al., 2012). The context of the IRES-mediated translation initiation in VEGF is of paramount importance with its established role in wound healing, angiogenesis, and cancer (Ferrara and Davis-Smyth, 1997; Huppert, 2010; Arcondeguy et al., 2013). More importantly, most of the aforementioned conditions that induce IRES-mediated translations are highly relevant to angiogenesis that results in new blood vessels under hypoxic conditions and tumorigenesis. The identification of a GQ structure that plays a central role in the translation regulation of

VEGF expression opens up the possibility of modulation of GQ structures by small molecules that can be an effective strategy to regulate VEGF expression for anticancer therapeutics.



**Scheme 2.3-** Schematic representation of the interaction of GQ motif in IRES A with the 40S ribosomal subunit.

## CHAPTER THREE

### **3. ENGINEERED DOMAIN SWAPPING INDICATES CONTEXT DEPENDENT FUNCTIONAL ROLE OF RNA G-QUADRUPLEXES.**

#### **3.1. DOMAIN ENGINEERING OF BIOMOLECULES**

Rational engineering of biomolecules can be a powerful strategy for understanding rules governing the formation of complex structures and ultimately correlate such knowledge to function. Protein engineering is a far more practiced science compared to RNA engineering, and thus examples of RNA engineering are rather sparse. A key reason for the dearth of RNA engineering might be due to lack of clearly defined interchangeable modules or domains that could maintain their functional topology in different contexts. Some of the best-known examples of RNA engineering are the designer riboswitches, where known RNA structural modules were combined to achieve complex constructs with novel function (Tang and Breaker, 1997; Soukup and Breaker, 1999). Additional examples include RNA domain swapping in the non-template RNA domains of the Telomerase enzyme established the regulatory role played by those sequences (Bhattacharyya and Blackburn, 1997). The majority of the RNA engineering works include riboswitches and ribozymes, which probably are the most well-characterized classes of RNA containing well-defined structural domains with precise functions. Due to the scarcity of well-defined, transplantable and independently folding functional domains in RNA, and the impact of the overall structure on the

functionality limits RNA engineering attempts. RNA domain swapping typically shows conservation of the native function of the domain in a non-native context. In contrast, we employed RNA engineering to demonstrate functional deviation of a G-quadruplex (GQ) that is contingent upon its context-dependent location, which is opposite of their native functional role.

The GQ structures are secondary structures formed by G-rich sequences wherein four G residues interact by Hoogsteen base pairing to form a square planar G-quartet. The quartets stack upon each other to form a G-Q structure (Gilbert and Feigon, 1999; Parkinson et al., 2002; Neidle and Balasubramanian, 2006). The GQ structures present in RNA came into focus with the discovery of a GQ structure in the 5'-untranslated region (5'-UTR) of NRAS was observed to inhibit translation (Kumari et al., 2007a). Thereafter several reports on RNA GQ structures in the 5'-UTR, coding region and the 3'-UTR associated with varied functions translation regulation, transcription termination, pre-mRNA processing, RNA targeting, and RNA turnover were published (Millevoi et al., 2012). The majority of the reports on the functional effect of GQ structures in the 5'-UTR have been associated with translation repression of genes such as, NRAS, ZIC1, MT3-MMP amongst others, thus regulating protein synthesis (Kumari et al., 2007a; Arora et al., 2008; Morris and Basu, 2009). A bioinformatics analysis of putative GQ forming sequence followed by comprehensive *in vitro* and *in vivo* analysis ascribed the GQ structures in 5'-UTR to be translation repressors (Beaudoin and Perreault, 2010). The report of GQ structures acting as essential elements was first observed in an internal ribosomal entry site (IRES) in the cases of FGF and VEGF (Bonnal et al., 2003; Morris et

al., 2010). Initially, it was perceived that the GQ structures generally inhibit translation except when they are present in the context of an IRESs (Bugaut and Balasubramanian, 2012). However, the GQ structures present in non-IRES 5'-UTRs were observed to be necessary for translation of mRNAs such as TGF  $\beta$  2 and FOXE3 (Agarwala et al., 2013; Agarwala et al., 2014). Moreover, it was also observed that inducing GQ structures resulted in inhibition of targeted genes (Ito et al., 2011; Bhattacharyya et al., 2014). In spite of so many examples of GQ mediated translation regulation, the mechanism by which similar structures play a different role is not clear and has perplexed researchers until now. This renders the GQ structures an ideal target for RNA engineering by swapping these motifs to delineate their mechanistic role especially in light of the recently published report of an independently folded GQ functional domain (Bhattacharyya et al., 2015). Here we investigated the context-dependent role of GQ structures by an RNA engineering approach wherein we swapped motifs that have been established biophysically and biochemically to adopt a GQ structure and were functionally relevant due to either acting as an inhibitor or as a necessary element for translation.

### **3.1. MATERIALS AND METHODS**

#### **3.2.1 Plasmid constructions**

The transcribable plasmid pVFIRESA and bicistronic plasmids hVEGFbicis containing the wild-type IRES A sequence were constructed as previously described (Morris et al., 2010). The hVEGFbicis was designed so that the Firefly luciferase expression was by IRES-mediated translation and the *Renilla* luciferase expression was by cap-dependent translation initiation (Scheme 2.1). We have previously observed that knocking down the eIF-4E expression reduces the expression of *Renilla* luciferase specifically which confirms the true bicistronic nature of the plasmid (Bhattacharyya et al., 2014). The primers for deletion mutation were purchased from Integrated DNA Technologies, Inc. The deletion of the specific domains was performed by QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent Tech) as per manufacturer's protocol. Subsequently, the insertion of the NRAS and M3Q sequence in both the plasmids were performed by the same kit using manufacturer's protocol. Simultaneously the dual luciferase plasmid containing the 5'-UTR of MT3-MMP (pwtM3Q) (Morris and Basu, 2009) was also modified to insert the GQ forming sequence of hVEGF and TGF $\beta$ 2. Plasmids were purified by Pure Yield™ Plasmid Miniprep System (Promega), and successful deletion and insertion was confirmed by sequencing at The Plant Microbe Genomic Facility (OSU).

### **3.2.2 *In vitro* transcription and radiolabeling of IRES A and mutant RNAs**

The 294 nt long IRES A and the mutants of the *hVEGF* mRNA were transcribed using plasmid pVFIRESA and its mutant variants (Morris et al., 2010). The transcribed

RNA was purified by 6 % denaturing polyacrylamide gel electrophoresis (PAGE). The RNA bands were harvested by tumbling the gel slices at 4 °C in elution buffer (300 mM NaCl, 10 mM Tris-HCl (pH 7.4), and 0.1 mM EDTA). The eluent was phase concentrated using 2-butanol and subsequently precipitated from the aqueous phase by ethanol.

The 5'-phosphates of the *in vitro* transcribed RNA were enzymatically removed by Calf intestinal alkaline phosphatase (CIP, NEB). The RNA was extracted by phenol-chloroform and precipitated with 70% ethanol. The CIP treated RNA was 5'-end radiolabeled by T4 polynucleotide kinase (PNK, NEB), [ $\gamma$ -<sup>32</sup>P] ATP (Perkin Elmer) and incubated for 1 hour at 37 °C. The commercially obtained DNA sequences were directly radiolabeled with T4 PNK as the 5'-phosphate is absent. The reaction was stopped by the addition of an equal volume of stop buffer (7 M urea, 10 mM Tris-HCl, pH 7.5 and 0.1 mM EDTA). The radiolabeled full-length RNA was isolated by 6 % denaturing PAGE. The radiolabeled RNA was then extracted from the gel via the crush and soak method as described previously.

### **3.2.3 Structure mapping of the IRES A by RNase T1 footprinting**

The 5'-end radiolabelled RNA (60000 cpm) was dissolved in 10 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, 2.5 mM MgCl<sub>2</sub> in the presence of 150 mM KCl or 150 mM LiCl and heated to 70 °C for 10 minutes and slow cooled to room temperature over 30 minutes. Once reactions attained the room temperature, the RNA was digested with 0.1 U

of Ambion® RNase T1 (Life Technologies), 0.005 µg of Ambion® RNase A for 5 minutes at 37 °C,  $1.4 \times 10^{-4}$  U of RNase One™ (Promega) for 5 minutes at 37 °C. The reactions were terminated by using an equal volume of stop buffer (7 M urea, 10 mM Tris-HCl, pH 7.5, and 0.1 mM EDTA).

Treated RNA was electrophoresed on a 6 % denaturing PAGE, dried on Whatman paper, and exposed to a phosphorimager screen. The gel images were then visualized by scanning the screen on a Typhoon Phosphorimager FLA 9500 (GE Healthcare, Life Sciences). The quantitative analyses of the lanes in the gel were done using ImageQuantTL software. The baseline was determined using the rolling-ball method. The peak volumes for each band were then normalized with the total count for each lane. The quantitation was done for Lane 2, 4 and 6 of (Fig. 3.2 b) which were the structure mapping of VF WT, M3Q, and NRAS in 150 mM K<sup>+</sup> respectively. Finally, each peak for the NRAS and M3Q were divided by the corresponding VF WT to determine if there was protection or enhanced cleavage for the band. The quantitative analyses were performed on three separate gel images and the standard error calculated (n=3). Significant protection or enhanced cleavage was considered only when the ratio was either less than 0.5 or 2.0 respectively and the bands are highlighted in the histogram.

#### **3.2.4 Filter binding assay**

The salt washed 40S ribosomal subunits isolated from rabbit reticulocyte lysate (RRL) were a kind gift from Dr. W. C. Merrick (Case Western Reserve University). The

radiolabeled RNAs were folded in the presence of the binding buffer (20 mM Tris-HCl, 100 mM CH<sub>3</sub>COOK, 200 mM KCl, 2.5 mM MgCl<sub>2</sub>, 1 mM DTT). The RNA was then incubated with increasing concentrations of 40S subunits at 37 °C for 20 minutes. The filter binding assay was performed using a Supor® Membrane (0.45 µm) on the top of a Nitrocellulose filter which in turn was placed on top of a *Hybond*<sup>TM</sup>-N+ filter supported by a filter paper. The Supor membrane was added to remove any aggregate, however, none was observed. The filters were presoaked in binding buffer assembled in a Minifold-I Dot-Blot System. The reaction mixtures were added to the wells of the Dot Blot apparatus under vacuum. The membranes were air dried and exposed to a phosphorimager screen. The blot images were then visualized by scanning the screen on a Typhoon Phosphorimager FLA 9500 (GE Healthcare, Life Sciences). The spots were quantified using Quantity One 1-D Analysis Software. To determine the apparent K<sub>d</sub> the data were fit to a Langmuir isotherm described by the equation  $\theta = Y_{\max} [X] / ([X] + K_d)$  where  $\theta$  is a fraction of RNA bound to 40S as measured from the fraction retained on the nitrocellulose membrane and [X] is the 40S subunit concentration. The results of three independent experiments for each labeled RNA are represented in Fig. 1b with the standard error (n=9).

### 3.2.5 Reporter gene assay

HeLa cells were grown in 96-well plates in Dulbecco's modified Eagle's medium (DMEM) with low glucose (HyClone®) supplemented with 10% fetal bovine serum and

1% antibiotics streptomycin and penicillin at 37 °C in 5% CO<sub>2</sub> in a humidified incubator. The plasmids were transfected by Lipofectamine 2000 as per manufacturer's protocol for 8 hours. After 24 hours of the transfection dual luciferase assay was performed with Dual-Glo® Luciferase Assay System as per manufacturer's protocol and the results read by the SpectraMax M4 plate reader (Molecular Devices, LLC).

To estimate the Relative IRES activity of the wild type and the mutants derived from the hVEGFBicis, the ratios of Firefly ( $L_F$ ) / *Renilla* ( $L_R$ ) luciferase expression of the plasmids were calculated and thereafter normalized to the VF deletion mutant to highlight the translation augmentation compared to the no GQ mutant. Alternatively for estimating the Relative gene expression, for the MT3-MMP context, the ratios of *Renilla* ( $L_R$ ) / Firefly( $L_F$ ) luciferase were calculated. To highlight the translation repression, the values were normalized to the GQ deletion mutant to estimate the relative gene expression (Morris and Basu, 2009).

The experiments for each of the relative IRES activity and the relative gene expression were repeated thrice with triplicates for every plasmid transfected in each experiment. The relative gene expression and the relative IRES activity were calculated individually for each experiment. The values of all the experiments were combined to estimate the standard error (n=9) and paired t-Test was used for the statistical analyses to determine if the expression of the engineered GQ domains were significantly different from the deletion mutants.

### **3.3 RESULTS**

The 5'-UTR of the MT3-MMP (MMP16) mRNA harbors a GQ forming sequence (M3Q) which has been shown to repress translational activity of a reporter mRNA in eukaryotic cells by more than fifty percent (Morris and Basu, 2009). The 5'-UTR of the MT3-MMP is 282 nt long (Fig. 3.1) and harbors the M3Q sequence (5'-GGGAGGGAGGGAGAGGGA-3'). It adopts an extremely stable three tier GQ structure that doesn't melt at physiological  $K^+$  concentration (Morris and Basu, 2009). The upregulation of MT3-MMP expression is associated with several cancers and metastasis.

The human vascular endothelial growth factor (hVEGF) has been associated with several pathological conditions such as progressive tumor development, macular degeneration amongst others (Ferrara and Davis-Smyth, 1997; Arcondeguy et al., 2013). A previous report established that the presence of a GQ within IRES-A of the 5'-UTR of hVEGF is essential for the cap-independent initiation of translation (Morris et al., 2010). The IRES A element is a 294 nt long (Fig. 3.1) and is unusually G-rich (45%). The GQ forming sequence (5'-GGAGGAGGGGGAGGAGGA-3') adopts a heterogeneous mix two-tiered GQs using multiple G-stretches. The hVEGF GQ was shown to directly interact with the 40S ribosomal subunit and regulates IRES A mediated translation initiation (Bhattacharyya et al., 2015).

It has been clearly established that these two GQs play opposite roles in translational regulation. Interestingly, both GQ forming sequences are 18 nt in length and all purine in nature. Hence, they were ideal targets for RNA engineering, to decipher the

structural requirement for their activity. Our system provides a unique platform for swapping of the modular quadruplex domains for designed RNA engineering.

The GQ forming NRAS is known to repress translation and the TGF $\beta$ 2 is known to augment translation, were also our motifs of interest (Kumari et al., 2007b; Agarwala et al., 2013). The NRAS GQ is a well-characterized sequence (5'-GGG AGG GGC GGG UCU GGG-3') forms a stable parallel GQ with a melting temperature  $63\pm 11^\circ\text{C}$  in the presence of 1 mM  $\text{K}^+$  and represses translation up to 70% (Kumari et al., 2007a). Alternatively, the TGF $\beta$ 2 sequence (5'-GGGAAAGGGU-GGGAGUCCAAGGG-3') forms a stable parallel GQ with a melting temperature of  $64\pm 1^\circ\text{C}$  in presence 25 mM  $\text{K}^+$  and enhances translation up to 98 % (Agarwala et al., 2013).

| Characteristics of GQ forming motifs used for domain swapping |                         |             |                      |                                                                |                 |                  |                        |               |            |
|---------------------------------------------------------------|-------------------------|-------------|----------------------|----------------------------------------------------------------|-----------------|------------------|------------------------|---------------|------------|
| Motif                                                         | Sequence                | Length (nt) | Function             | $T_m$ (Stability)                                              | UTR length (nt) | GQ position (nt) | Distance from TSS (nt) | GC content    | QGRS score |
| NRAS                                                          | GGGAGGGGCGGGUCUGGG      | 18          | Repress translation  | $63\pm 11^\circ\text{C}$ (1mM $\text{K}^+$ )                   | 254             | 15-32            | 222                    | 65.7 %        | 40         |
| M3Q                                                           | GGGAGGGGAGGAGAGGGA      | 18          | Repress translation  | $72\pm 1^\circ\text{C}$ (1mM $\text{K}^+$ )                    | 282             | 213-230          | 52                     | 61.7%         | 40         |
| VEGF                                                          | GGAGGAGGGGAGGAGGA       | 18          | Enhances translation | $61.5^\circ\text{C}\pm 1^\circ\text{C}$ (100 mM $\text{K}^+$ ) | 1038 (294)      | 774-791          | 247                    | 66.4% (76.2%) | 21         |
| TGF $\beta$ 2                                                 | GGGAAAGGGUGGGAGUCCAAGGG | 23          | Enhances translation | $64\pm 1^\circ\text{C}$ (25mM $\text{K}^+$ )                   | 1368            | 313-335          | 1033                   | 55.0%         | 36         |

**Table 3.1- The four GQ forming motifs used in this study for RNA engineering and their sequence details and characteristics:** The contexts used in this study are that of the M3Q harboring 5'-UTR of the MT3-MMP (MMP16), which is 282 nt in length. This UTR is not known to harbor any IRES. Alternatively, another context used for this study is the 294 nt long IRES A present in the 1038 nt long 5'-UTR of hVEGF mRNA. The IRES A is a highly G-rich sequence with about 45% G-residues and total GC content of 76.2%.

While swapping, we replaced the VEGF GQ domain required for translation with the translation repressors NRAS and M3Q sequences respectively. Our goal was to determine the effect of the otherwise translation repression motifs when they are

exchanged with a well-defined domain essential for translation. The IRES A folds into a very stable secondary structure owing to its highly GC enriched sequence (~75%) and hence the formation of the GQ structure by the engineered sequences in an alien context needed to be confirmed.

**5'- UTR of MT3-MMP (MMP16) context**

5'-UCUUUGAGCCGGAACCGCCGGUGAACUUAGGCGCCACGUUCCCGGUGACUGACCCC  
 GGAGGAUGGUGAACAGGAACUACUGCUGACCCUGUCUUCGCCGCUGCCUCUCGGGGGC  
 UGCCGAGCGCGGGGCCCCUGUCUCCUCCCUUGCCCGGCUGUGGGUGAACCGCAGGC  
 UCCUUACCCACCCGGAGGACUUUUUUUGAAAGGAAACGAG<sup>G<sup>213</sup></sup>GGAGGGAGGGAGAGG  
GA<sup>230</sup>GAGAGGGAGAAAACGAAGGGGAGCUCGUCCAUCAUGAAGCACAGUUCACUAAUG  
 ....3'

Sequences inserted in the GQ forming region:

**VEGF: 5'-GGAGGAGGGGGAGGAGGA-3'**

**TGFβ2: 5'-GGGAAAGGGUGGGAGUCCAAGGG -3'**

**hVEGF IRES A context:**

5'- CUAGCUCGGGCCGGGAGGAGCCGCAGCC<sup>G<sup>774</sup></sup>GAGGAGGGGGAGGAGG<sup>790</sup>AAGAAG  
 AGAAGGAAGAGGAGAGGGGGCCGCAGUGGCGACUCGGCGCUCGGAAGCCGGGCUCAU  
 GGACGGGUGAGGCGGCGGUGUGCAGACAGUGCUCAGCCGCGCGCUCUCCAGG  
 CCCUGGCCCGGGCCUCGGGCCGGGGAGGAAGAGUAGCUCGCCGAGGCGCCGAGGAGA  
 GCGGGCCGCCACAGCCGAGCCGGAGAGGGAGCGCGAGCCGCGCCGGCCCCGGUC  
 GGGCCUCCGAAACCAAUG....3'

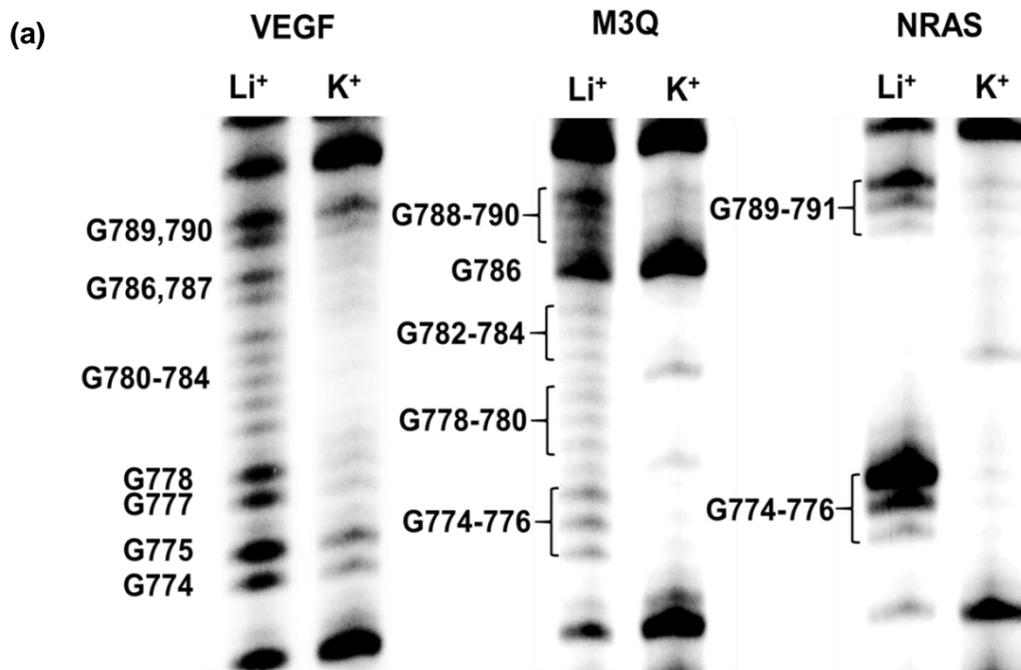
Sequences inserted GQ forming region:

**M3Q: 5'-GGGAGGGAGGGAGAGGGA-3'**

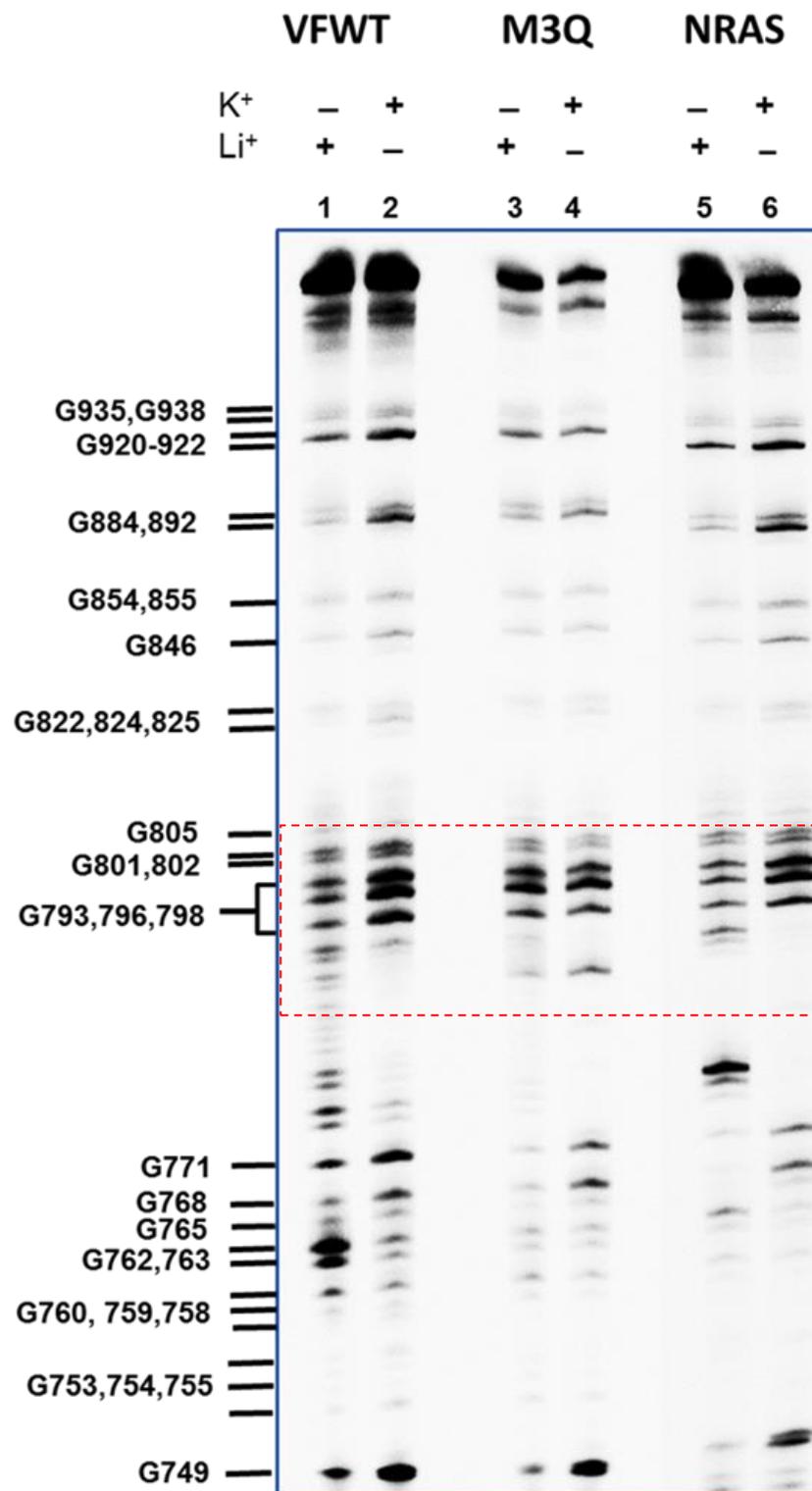
**NRAS: 5'-GGGAGGGGCGGGUCUGGG-3'**

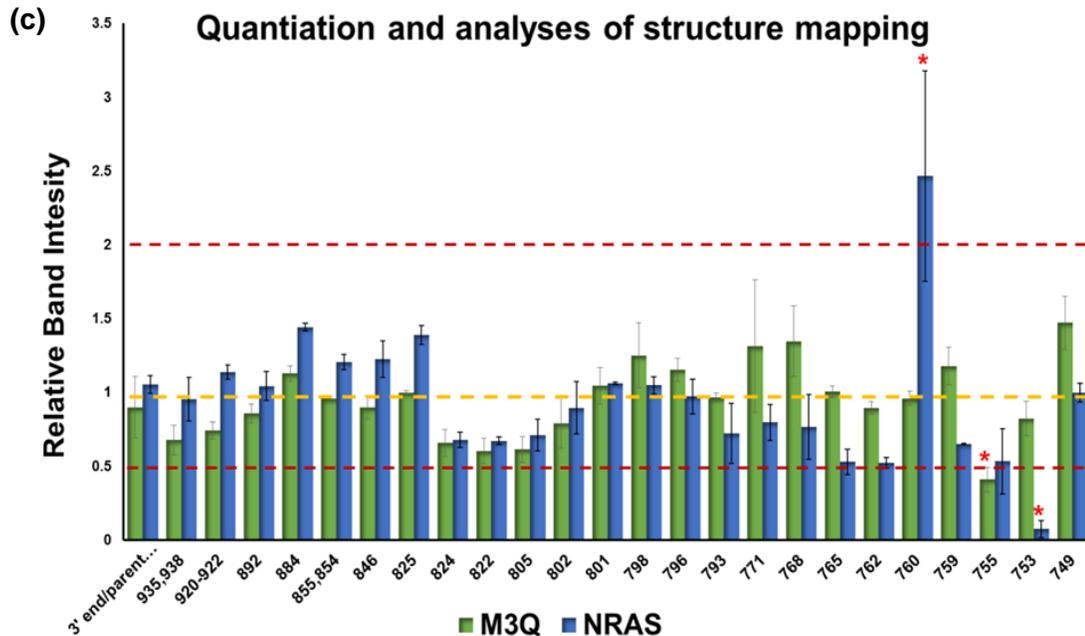
**Figure 3.1- Contexts in which the domain swapping was performed:** The 5'- UTR of MT3-MMP harbors the translation repressing M3Q sequence (red) which was replaced by the translation enhancing **VEGF GQ** or the **TGFβ2** GQ forming sequences. Alternatively, the IRES A of hVEGF harbors a translation enhancing GQ (red) which was swapped with translation repressing **M3Q** and **NRAS GQ** forming sequences.

Structure mapping analyses in the presence of 150 mM  $K^+$  or  $Li^+$  were performed to ensure that the direct replacement of native motif with the NRAS and M3Q did not affect GQ formation (Fig. 3.2 a) and the overall IRES structure (Fig. 3.2 b and c). The structure mapping data confirmed the formation of GQ by the exchanged motifs in their new molecular habitat and no major changes in the cleavage pattern of the G residues by RNase T1. The M3Q sequence is known to show some protection from RNase T1 cleavage even in the absence of  $K^+$ , however, increased protection of the Gs in the presence of  $K^+$  and the susceptibility of the loop to RNase T1 indicate the formation of GQ structure (Morris and Basu, 2009). In the case of NRAS, we detected protection of all the G stretches in the presence of  $K^+$  suggesting the formation of GQ.



(b)



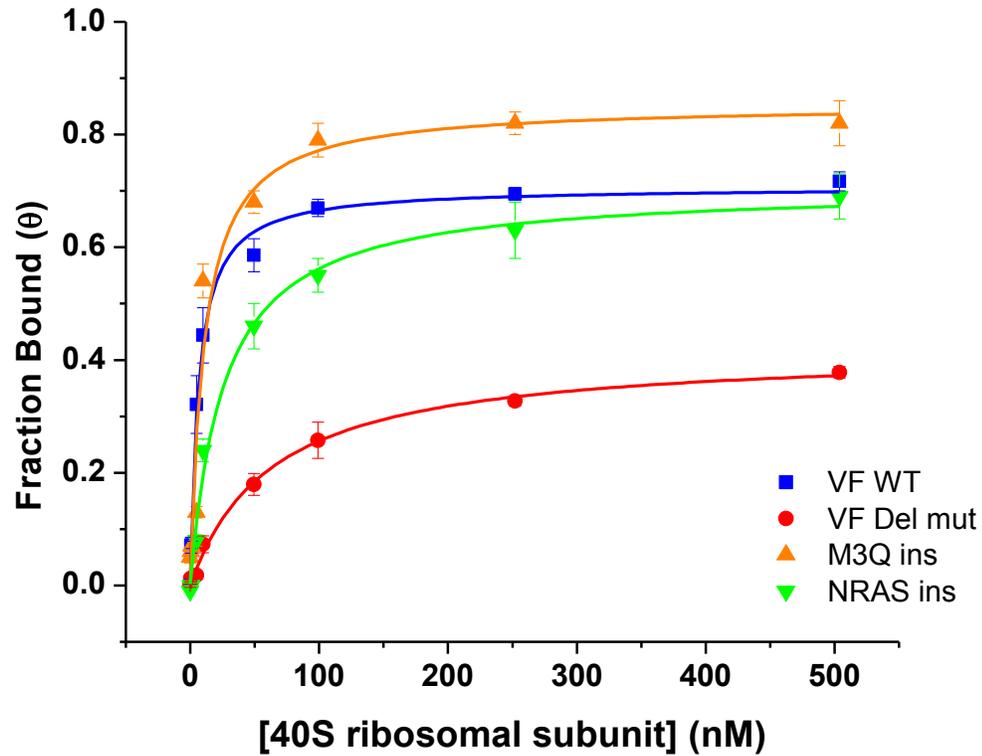


**Figure 3.2- Structure mapping of IRES A with engineered GQ sequences:** (a) RNase T1 footprinting of the GQ forming sequences inserted in IRES A indicates the formation of GQ by the inserted M3Q and NRAS sequence. (b) The structure mapping of the entire IRES A with wildtype VEGF (VFWT), M3Q and NRAS GQ motif in the presence of 150 mM of KCl or LiCl. The quadruplex forming motif (in the box) is enlarged and represented in Figure 1a to show the structural changes with the formation of GQ structure. (c) The histogram represents the quantitation of the cleavage of the G residues in IRES A other than the GQ forming motifs. The results show that there was no significant difference in the overall band pattern when the M3Q and the NRAS sequence were inserted into the IRES A. The overall ratios of most of the bands were between 0.5 to 2.0 and mostly around 1.0. The G residues that had a significant change in protection (<0.5) or enhanced cleavage (>2.0) are indicated by the red asterisk.

Previously, we reported that the endogenous GQ domain within the IRES A folds independently and can directly interact with the 40S ribosomal subunit (Bhattacharyya et al., 2015). This prompted us to test if the IRES A with its engineered GQs still maintains its ability to interact with the 40S. We analyzed the binding of 40S subunit to the engineered IRES A that included either the NRAS or M3Q motifs respectively. Filter-binding assay was performed to determine the nitrocellulose filter bound fraction of RNA

to the 40S ribosomal subunit and the unbound RNA that binds to the positively charged Hybond N<sup>+</sup> membrane underneath. The apparent dissociation constant ( $K_d$ ) was calculated using the equation  $\theta = Y_{\max} [X] / ([X] + K_d)$ , where  $\theta$  is a fraction of RNA bound to 40S as measured by the fraction retained on the nitrocellulose membrane and  $[X]$  is the 40S subunit concentration. We determined that the 40S ribosomal subunit binds to the IRES A with a  $K_d$  of  $6.3 \pm 0.6$  nM. The deletion of the GQ motif abrogated the 40S ribosomal subunit binding. However, when the M3Q and the NRAS GQ motifs were inserted the 40S ribosomal subunit binding was restored. The apparent equilibrium dissociation constant of the IRES A with the inserted M3Q sequence was the lowest indicating the tightest of binding ( $K_d = 10.6 \pm 3.1$  nM) followed by the NRAS sequence ( $K_d = 25.8 \pm 3.6$  nM). Although the NRAS sequence forms the GQ structure, its interaction with the 40S subunit is weaker compared to the other motifs that were tested. It is noteworthy that amongst the three sequences, VEGF and M3Q are all purine sequence, but NRAS has two all pyrimidine loops.

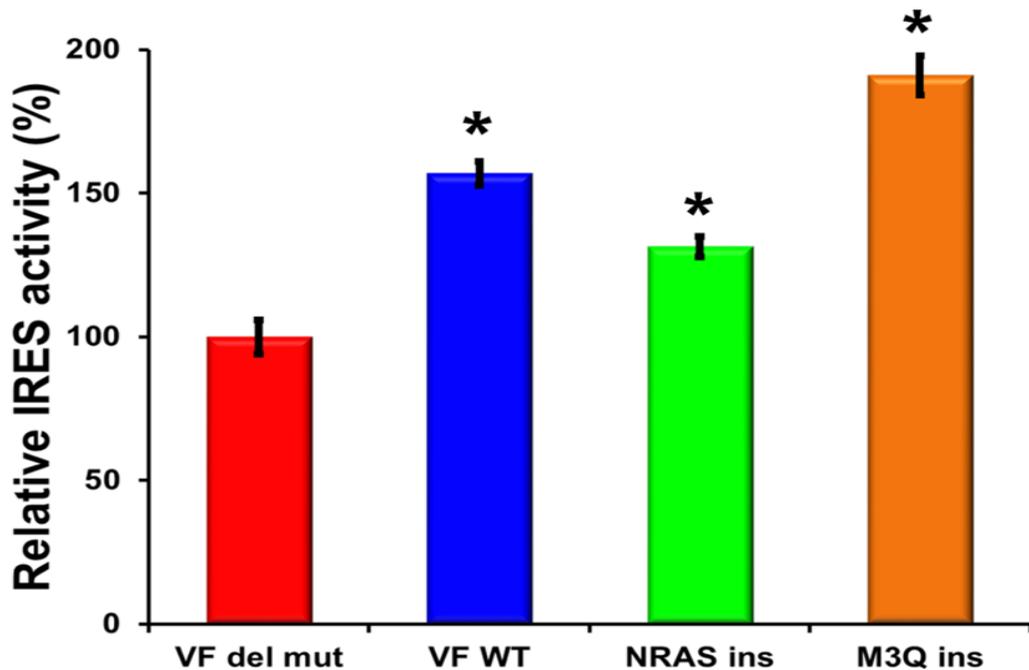
Once we established the formation of the GQ structure by M3Q and NRAS sequences in the context of the entire hVEGF IRES A and determined their interactions with 40S subunits, we proceeded to assess the functional role these structures play in cap-independent translation initiation. The NRAS and M3Q GQ motifs in IRES A were observed to be essential elements compared to that of a deletion mutant wherein the GQ forming motif was deleted. The M3Q motif in the IRES had the most enhancing effect (85%) followed by the native VEGF sequence (60%) and the NRAS sequence (34%).



| Constructs | Method 2  |                |
|------------|-----------|----------------|
|            | $Y_{max}$ | $K_d$          |
| IRES A     | 0.71      | $6.3 \pm 0.6$  |
| Del mut    | 0.42      | $63.5 \pm 6.3$ |
| M3Q ins    | 0.85      | $10.6 \pm 3.1$ |
| NRAS ins   | 0.71      | $25.8 \pm 3.6$ |

**Figure 3.3- The IRES A with engineered GQ sequences bind to the 40S ribosomal subunit:** The inserted GQ forming sequences of M3Q and NRAS in IRES A of hVEGF were capable of interacting with the 40S ribosomal subunit as the deletion of the GQ forming sequence rendered the IRES A incapable of the same interaction.

The probable cause for the high enhancing effect might be the exceptional stability of the M3Q GQ motif that may facilitate interaction with the 40S subunit. It is noteworthy that amongst the three sequences VEGF and M3Q are all purine sequence and the NRAS has two all pyrimidine loops of the total three loop regions. The lower level of cap-independent translation initiation by the NRAS sequence can be a result of the pyrimidines as presumably the stability of the GQ is a not an issue as it is known to form a very stable structure. However, the exact role of the pyrimidines in the loops and how they may interfere with the IRES A mediated translation initiation requires more extensive analyses and is beyond the scope of our present study.

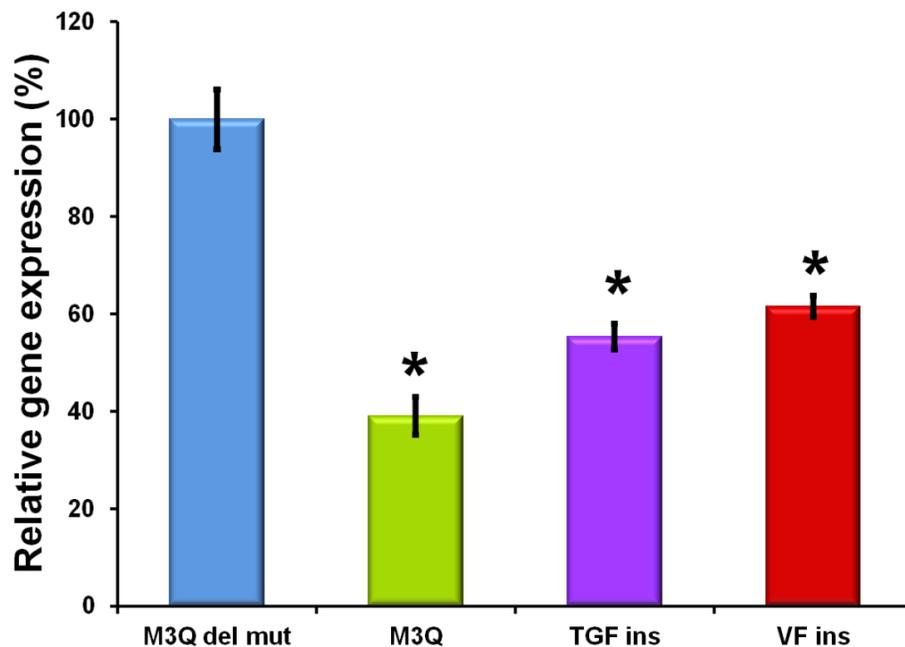


**Figure 3.4- The IRES A activity with engineered GQ sequences:** The relative IRES A activity was dependent on the presence of the GQ forming sequence and was significantly ( $p < 0.05$ ) higher compared to the deletion of the GQ sequences. The relative IRES activity follows the same pattern to that of the 40S ribosomal subunit binding.

Our studies indicate that the GQ in IRES A is responsible for binding to the 40S ribosomal subunit and might be correlated with the flexibility of the motif to adopt multiple GQ structures. The possibility of the requirement of a polypurine stretch along with the GQ cannot be ignored. We observed correspondence of the 40S subunit binding to the activity of the GQ motifs in a bicistronic plasmid. Interestingly during the preparation of this manuscript, another study reported that the stabilization of the GQ in the IRES reduces the cap-independent translation initiation. In the study, the authors replace the VEGF sequence and insert the NRAS sequence in its place and also modified the VEGF sequence to form a three-tiered GQ and observed significant downregulation of the cap-independent translation. This report is in contrast to our observation wherein we observed that the M3Q sequence, which forms an unusually stable GQ, increases the cap-independent translation compared to the wild type sequence. However, the data reported here resulted from experiments performed by at least three researchers at different points in time and consistently suggested the increase in translation with stable GQ sequence.(Morris, 2012) Furthermore the fact that increase in the affinity of the 40S ribosomal subunit is related to the increase in translation initiation strengthens the functional modus operandi of the GQ sequence to modulate cap-independent translation and thus providing a mechanistic validation of our results. In addition, it is consistent with our previous reports on the role of VEGF IRES A GQ in 40S recruitment and translation initiation.

In a separate set of experiments, we inserted the GQ motif of VEGF and TGF $\beta$ 2 in a dual luciferase plasmid containing the entire 5'-UTR of MT3-MMP to test how GQs

that are known to be required for translation act when placed within a molecule where the wild type GQ plays an inhibitory role. The translation repressing M3Q (GQ forming) domain in the sequence was swapped with the GQ forming motifs of the VEGF and TGF $\beta$ 2 required for translation. The VEGF and TGF $\beta$ 2 are well established GQ motifs that are essential for translation initiation (Morris et al., 2010; Agarwala et al., 2013). Although VEGF GQ motif was observed to augment translation in the context of an IRES, the TGF $\beta$ 2 mRNA is not known to harbor any IRES as per our knowledge. In line with the studies reported above, which indicate that the otherwise repressive GQ motifs play the role of essential elements in the context of the VEGF IRES A, the role of essential elements in the context of repressive GQ motif was intended for investigation.



**Figure 3.5- The engineered GQ sequences in 5'-UTR of MT3-MMP repress translation:** The inserted GQ forming sequences of VEGF and TGF $\beta$ 2 in the 5'-UTR of

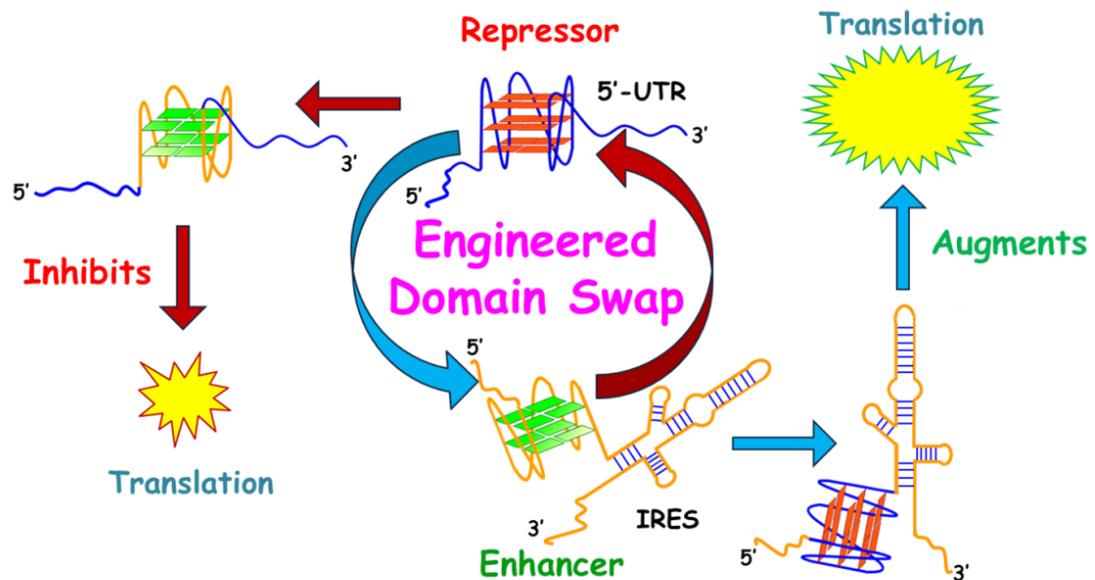
MT3-MMP repressed translation significantly ( $p < 0.05$ ) compared to the deletion of any GQ forming sequence.

The activity of luciferase assay indicates that the M3Q GQ motif repressed the translation by 60%, the TGF $\beta$ 2 GQ motif repressed by 45% and the VEGF motif repressed by 35%. The repressive activities can be correlated to the stability of the GQ structures. The M3Q is an extremely stable GQ structure perhaps the most stable of the GQ motifs followed by TGF $\beta$ 2 and the least stable being the two-tiered VEGF GQ.

### **3.4 DISCUSSION**

Several research groups have widely studied the role of GQs in translation. The ability of a non-canonical structure to substantially modulate translation efficiency relates to considerable effect on gene expression. This led us to induce GQ structures to modulate translation, wherein we observed the similar level of downregulation of the target gene (Bhattacharyya et al., 2014). The presence of GQ structures in translation repression can be explained by the fact that these structures may disrupt the assembly of the initiation factors when present proximal to the 5'-cap. These extremely stable structures can also disrupt the scanning mechanism by the 43S preinitiation complex when present in the 5'-UTR (Bugaut and Balasubramanian, 2012). Thus, the mechanism of repression by the GQ resembles that used by other stable secondary structures within the UTR (Kozak, 1991). In contrast, the mechanism utilized by GQ structures to facilitate translation initiation is intriguing. We have reported previously the GQ when present in the context of an IRES directly recruits 40S ribosomal subunit thus explaining the

mechanistic reason for their existence, for example in human VEGF IRES A (Bhattacharyya et al., 2015). However, it remains to be established that if this is a universal mechanism for all translationally essential GQ elements. Moreover, it is quite possible that these structures can interact with different protein factors, which in turn may upregulate translation.



**Scheme 3.1-Schematic representation of engineered domain swapping:** GQ motif is not associated with a defined inherent function, exhibit considerable structural integrity irrespective of a non-native context but display significant functional adaptability.

## CHAPTER FOUR

"Adapted with permission from: D. Bhattacharyya, K. Nguyen, S. Basu, Rationally induced RNA: DNA G-quadruplex structures elicit an anticancer effect by inhibiting endogenous eIF-4E expression, *Biochemistry*, 53 (2014) 5461-5470. Copyright 2014 American Chemical Society."

### **4. RATIONALLY INDUCED RNA: DNA G-QUADRUPLEX STRUCTURES ELICIT ANTI-CANCER EFFECT BY INHIBITING ENDOGENOUS eIF-4E EXPRESSION**

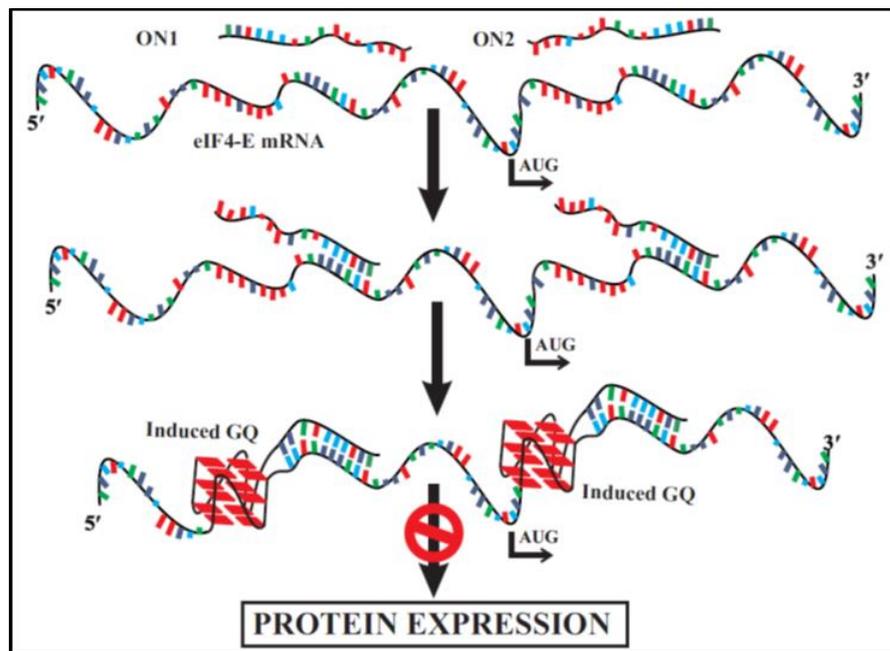
#### **4.1 INTRODUCTION**

RNA GQs have been associated with several functional roles, such as telomere maintenance, transcription termination, the pre-mRNA processing including splicing and polyadenylation, RNA turnover and mRNA targeting (Millevoi et al., 2012). RNA GQ structures in the 5'-UTR are functionally significant in repressing translation of several pathologically important genes such as NRAS, MT3-MMP, TRF2, ERS1, THRA, and BCL-2 (Rogers et al., 1999; Kumari et al., 2007a; Balkwill et al., 2009; Morris and Basu, 2009; Gomez et al., 2010; Shahid et al., 2010). The role of GQ structures in the 5'-UTR of an mRNA in the regulation of translation generally varies with the position and the stability of the structures (Wieland and Hartig, 2007; Kumari et al., 2008; Bugaut and Balasubramanian, 2012). The presence of a GQ in the 5'-UTR has been mostly observed to repress translation with the exception of transforming growth factor  $\beta$ 2 (TGF $\beta$ 2) (Agarwala et al., 2013) and when present in an IRES. (Bonnal et al., 2003; Morris et al., 2010). Recently it was reported that RNA GQ structures within the protein coding sequence of hER $\alpha$  represses translation which result in expression of a truncated protein

that readily undergoes proteolysis (Endoh et al., 2013a). Therefore, the wide range of functional roles of RNA GQ structures makes them attractive therapeutic targets and indeed they have been targeted by several small molecule ligands. For example, small molecules have been used to modulate translation either by disrupting the RNA GQs to revert their repressive effects or stabilize the GQs to enhance repression (Kumari et al., 2007a; Huppert et al., 2008; Balasubramanian and Neidle, 2009; Satyanarayana et al., 2010; Morris et al., 2012). However, the targeting of RNA GQs by small molecules for regulation of specific genes is rather challenging due to the low level of structural variation within RNA GQs, as they are known to almost exclusively adopt parallel conformation (Bugaut and Balasubramanian, 2012). Hence, additional novel and rational design approaches are crucial for targeting specific endogenous RNA GQs.

GQ structures are formed when two or more G-quartets stack upon each other and in the center the oxygen lone pair of the carbonyl groups coordinates with metal cations, typically  $K^+$  (Neidle and Balasubramanian, 2006). A G-quartet is formed by four guanine bases arranged in a square planar pattern with Hoogsteen hydrogen bonding. The formation of an intramolecular G-quadruplex (GQ) requires, at least, four stretches of guanines with each containing two or more contiguous guanine residues. While unimolecular GQs require four stretches of Gs provided by a single strand, bimolecular GQs, need only two stretches of guanines in a strand for interaction with a targeting strand, which would provide in trans two additional guanine stretches. Therefore, a bimolecular GQ structure could be selectively induced in a target mRNA sequence to repress translation if it is strategically located, such as within the 5'-UTR or the protein

coding region of an mRNA (Scheme 4.1) and contains two stretches of Gs. A previous report described the use of a short G-rich RNA (5'-GGGCCCGGG-3' and 5'-GGGUUAGGG-3') to inhibit expression of an engineered reporter gene via formation of an intermolecular RNA GQ in the 5'-UTR and in the coding sequence (Ito et al., 2011). However, such an approach potentially can be extremely nonspecific given the high propensity of occurrence of the sequence  $G_{(3-4)}N_{(1-5)}G_{(3-4)}$  in the transcriptome where N can be any nucleotide. In another report, inducing the formation of a GQ structure in a target RNA sequence with a guanine-tethered oligonucleotide has been observed to block the reverse transcriptase activity *in vitro* (Hagihara et al., 2010), where the mechanism of action is rather unclear.



**Scheme 4.1- Schematic representation of targeted induction of GQ in eIF-4E mRNA in the 5'- UTR and the protein-coding region.**

Human eukaryotic initiation factor-4E (eIF-4E) is an essential factor for translation initiation, which often is designated as an oncogene and is overexpressed in several cancer cell types (Mamane et al., 2004; Hsieh and Ruggero, 2010; Carroll and Borden, 2013). It is a component of the eIF-4F complex and by binding to the 7-methyl guanosine cap structure at the 5'-end of mRNAs, it provides the critical interface between mRNA, recruitment of eIF-4A and eIF-4G, and the 40S ribosomal subunit (Sonenberg et al., 1979). eIF-4E is the least abundant among translation initiation factors and thereby acts as the rate-limiting agent in translation initiation. Additionally, the eIF-4E repression by drugs, antisense oligonucleotides, and siRNA inhibits growth in various cancer cells such as gastric, breast carcinomas and mesothelioma cells which make it a good target for downregulation of its expression for therapeutic application (Jacobson et al., 2013). The eIF-4E mRNA harbors sequences both in the 5'-UTR and in the coding sequence, which are ideal targets for rational induction of GQs. We designed a bifunctional oligonucleotide that contained i) two G-stretches of three contiguous guanines each and ii) a segment that would guide the G-stretches to its target mRNA.

Our studies on inducing GQ structures within the 5'-UTR and the coding region of eIF-4E mRNA showed that targeted induction of GQ could be a powerful strategy for the selective regulation of gene expression for therapeutic application.

## **4.2. MATERIALS AND METHODS**

### **4.2.1 *In vitro* transcription of a segment of 5'-UTR of human eIF4E mRNA**

A 36 nt fragment of the eIF-4E mRNA was *in vitro* transcribed and purified by 15% denaturing polyacrylamide gel electrophoresis (PAGE). The RNA band was harvested via the crush and soak method by tumbling the gel slices at 4 °C in elution buffer (300 mM NaCl, 10 mM Tris-HCl (pH 7.4), and 0.1 mM EDTA). The eluent was phase concentrated using 2-butanol and subsequently precipitated from the aqueous phase by ethanol. The RNA pellet was dissolved in 10 mM Tris-HCl (pH 7.5) and 0.1 mM EDTA unless mentioned otherwise.

#### 4.2.2. DNA Oligonucleotides

Modified and unmodified DNA oligonucleotides for *in vitro* characterization and cell studies were purchased from Integrated DNA Technologies, Inc.

| <b>Modified DNA oligonucleotides used for cell studies</b> |                                               |
|------------------------------------------------------------|-----------------------------------------------|
| <b>ON1</b>                                                 | <b>C*G*TTTAGGCAATCAA GGGAG*G*G</b>            |
| <b>Scrambled Quadruplex 1 (SQ1)</b>                        | <b>C*G*TTTAGGCAATCAA GACCG*A*C</b>            |
| <b>ON2</b>                                                 | <b>A*C*CTTTCCTTGTATA GGGAG*G*G</b>            |
| <b>Scrambled Quadruplex 2 (SQ2)</b>                        | <b>A*C*CTTTCCTTGTATA GACCG*A*C</b>            |
| <b>Scrambled Duplex (SD)</b>                               | <b>N*N*NNNNNNNNNNNNNN GGGAG*G*G</b>           |
| <b>Scrambled Whole (SW)</b>                                | <b>N*N*NNNNNNNNNNNNNNNNNNNNNNNNNNNNNN*N*N</b> |

**Table 4.1- DNA oligonucleotide sequences used in the study.** An asterisk (\*) denotes phosphorothioate modification which renders the oligonucleotides exonuclease resistant.

#### 4.2.3. Folding of the oligonucleotides to form induced GQ

The induced GQ structure was formed by mixing equimolar amounts of RNA and ON1 in 150 mM KCl or LiCl, 10 mM Tris-HCl and 0.1 mM EDTA (pH 7.5). The target RNA and ON1 were also folded individually in the same buffer and salt concentrations. The solutions were heated to 95 °C and cooled to 25 °C with a gradient of 15 °C/10 minutes in a thermocycler.

#### **4.2.4. Circular Dichroism (CD) spectroscopy**

The CD spectra were recorded using a Jasco J-810 spectropolarimeter using a cell with 0.1 cm path length. The oligonucleotides and induced GQ were folded as mentioned above. The spectrum of the only buffer was subtracted from the individual samples for background correction. For the nuclease digestion assay, induced GQ structures formed in 150 mM K<sup>+</sup> were treated with 3 units of DNase 1 and RNase T1 and the CD spectra were observed after every 5 minutes for 15 minutes at 37 °C.

##### **4.2.4.1 Analysis of the thermal melting curves obtained by CD spectroscopy:**

The melting curves indicate a change in molar ellipticity at 263 nm with varying the temperature from 25 °C to 97 °C at a rate of 20 °C per hour in the CD spectropolarimeter. We calculated the thermodynamic parameters by two different methods.

##### **Method 1**

In this method, we calculated the thermodynamic parameters and  $T_m$  values were calculated in a way previously described. (Marky and Breslauer, 1987) In brief, we obtained the melting data from CD machine after smoothing (adaptive). We then calculated the fraction folded ( $\alpha$ ) using the equation

$$\alpha = \frac{\phi - Au(T)}{Af(T) - Au(T)} \quad \text{Equation 4.1}$$

where  $Au(T)$  and  $Af(T)$  corresponds to the upper and lower baseline values, respectively. We assumed that the low temperature baseline represents a completely folded state and the higher temperature baseline represents the unfolded form, the Van't Hoff enthalpy can be calculated using a simple two-state model (Folded  $\leftrightarrow$  Unfolded).

We generated a scattered plot with the alpha value vs the temperature and fit the curve Boltzmann fitting. Thereafter we differentiated the Boltzman fit to generate the 1st derivative plot and get the  $T_m$  value at  $\alpha=0.5$ , using the Origin Software. The equation used for the fit was

$$y = A2 + (A1 - A2)/(1 + \exp^{(x-x_0)/dx}), \quad \text{Equation 4.2}$$

where,

A1 denotes the top asymptote (alpha~ 1); initial value

A2 denotes the bottom asymptote (alpha ~ 0); final value

$x_0$  is the is the midpoint of the fit ( $\sim T_m$ )

$dx$  is the slope/ temperature constant

Thereafter we calculate the  $\Delta H$  using the equation given below:

$$\Delta H_{VH} = (2 + 2n)RT_m^2 \left( \frac{\partial \alpha}{\partial T} \right)_{T=T_m} \quad \text{Equation 4.3}$$

4.3

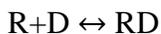
molecular n=1, bimolecular n=2, tetramolecular n=4). The  $(d\alpha/dT)_{T=T_m}$  value is obtained from the Origin at the  $T_m$  value. The  $\Delta S$  is calculated assuming  $\Delta G=0$  and  $T=T_m$  where  $\Delta S=\Delta H/T_m$  and the  $\Delta G_{37}$  is calculated from the above values as  $\Delta G_{37}=\Delta H-T\Delta S$ .

### **Method 2:**

Simultaneously, we calculated the thermodynamic parameters using the van't Hoff plot as described earlier in a previously published method (Mergny and Lacroix, 2003). To derive the  $T_m$ , in this case, we calculated the median of the upper and lower baselines as follows

$$\text{median}(T) = (A_u(T) + A_f(T))/2 \quad \text{Equation 4.4}$$

where  $A_u(T)$  and  $A_f(T)$  corresponds to the upper and lower baseline values, respectively. The fraction folded was calculated similarly given by equation 1. The molecularity of the reaction must be known. In our case for a bimolecular equilibrium, similar to that of the not self-complementary duplex oligonucleotide we determined the  $K_a$  according to the following equation



where the total strand concentration

$$C_{\text{tot}} = [R] + [D] + 2[RD]$$

The fraction of strands folded,  $\alpha = 2[RD] / C$

$$\text{and } (1-\alpha) = ([R] + [D]) / C$$

We define  $K_a = [RD] / [R][D]$

In our case since we used equal total concentrations of R and D,

$$K_a = 2\alpha / C_{\text{tot}} (1-\alpha)^2 \quad \text{Equation 4.5}$$

Thereafter the van't Hoff plot was created by plotting the  $\ln K_a$  vs  $1/T$  in  $K^{-1}$ . The analysis interval was restricted to  $0.15 < \alpha < 0.85$  where  $K_a$  values are most precise. By definition, the free Gibbs enthalpy may be written as:

$$\Delta G = -RT \ln(K_a) = \Delta H - (T \times \Delta S) \quad \text{Equation 4.6}$$

And can, therefore, be derived as

$$\ln(K_a) = [-\Delta H/R] \times [(1/T) + \Delta S/R] \quad \text{Equation 4.7}$$

we assume that the  $\Delta H$  and  $\Delta S$  are independent of temperature and therefore the van't Hoff representation of  $\ln(K_a)$  versus  $1/T$  gives a straight line with a slope of  $-\Delta H/R$  and a Y-axis intercept of  $\Delta S/R$ .

Thereafter we determined the thermodynamic parameters and compared to that determined by method 1. The determination of the thermodynamic parameters and the comparison are included in Appendix II of the dissertation. Herein we discussed the results obtained from Method 1 since it has been regularly used in our laboratory and published articles.

#### **4.2.5. Isothermal titration calorimetry**

The DNA/RNA-binding reactions were conducted using a Microcal, Inc. VP-ITC as previously published (Lang and Schwarz, 2007). The experiment was performed by titrating 7  $\mu\text{L}$  aliquots of 100  $\mu\text{M}$  ON1 into a 10  $\mu\text{M}$  solution of target RNA. The

titrations were 3 minutes apart. The titration run was continued until the addition of the titrant gave no appreciable heat response, where the reaction should be past the saturation point. Heats of dilution of the titrant into just the buffer solution were determined and subsequently subtracted from the binding isotherm. The calorimetric titration data was analyzed and were fit using Origin software from Microcal Inc. that uses a non-linear least squares minimization method. The resulting fit of the data gives observed apparent equilibrium constant,  $K$ , the enthalpy of binding, the binding entropy and the stoichiometric ratio of the reaction,  $N$ . The  $\Delta G$  was calculated using  $\Delta G = \Delta H - T\Delta S$

#### **4.2.6 Radiolabeling of RNA and DNA Oligonucleotides**

The RNA and DNA oligonucleotides were radiolabeled using the same method as described before (Morris and Basu, 2009). In brief, *in vitro* transcribed RNA was enzymatically removed by Calf intestinal alkaline phosphatase (CIP, NEB). The CIP treated RNA and commercially obtained DNA were 5'-end radiolabeled by T4 polynucleotide kinase (PNK, NEB), [ $\gamma$ - $^{32}P$ ] ATP (Perkin Elmer). The radiolabeled full-length RNA and ON1 were isolated by 15% denaturing PAGE and then extracted as described previously.

#### **4.2.7. Native gel electrophoretic mobility shift assay**

The 5'-end radiolabeled RNA, DNA and RNA and DNA (10000 cpm) were mixed with 40 picomoles of unlabeled samples in 150 mM KCl and 150 mM LiCl in 10 mM Tris-HCl pH 7.5 and 0.1 mM EDTA in a total volume of 10  $\mu$ L. The

oligonucleotides were folded and the induced GQ was formed using the same procedure as mentioned above. The complexes were resolved by 13% native polyacrylamide gel electrophoresis at 5-6 °C in Tris-borate-EDTA buffer. The gel was exposed to a phosphorimager screen and then visualized by Typhoon Phosphorimager FLA 9500 (GE Healthcare, Life Sciences).

#### **4.2.8. RNase T1 structure mapping**

The 5'-end radiolabeled RNA was dissolved in 10 mM Tris-HCl (pH 7.5), 0.1 mM EDTA in the presence of 150 mM KCl or 150 mM LiCl and unlabeled ON1 in increasing concentrations (0-5  $\mu$ M) were folded as mentioned above. Once reactions attained the appropriate temperature, the RNA was digested with 0.02 U of Ambion® RNase T1 (Life Technologies) for 1 minute at 37 °C. The reactions were terminated by using an equal volume of stop buffer (7 M urea, 10 mM Tris-HCl, pH 7.5, and 0.1 mM EDTA). Treated RNA was electrophoresed on a 12% denaturing PAGE, dried on Whatman paper, and exposed to a phosphorimager screen and visualized by a Typhoon Phosphorimager FLA 9500 (GE Healthcare, Life Sciences).

#### **4.2.9. Dimethyl Sulfate (DMS) Footprinting**

The 5'-end radiolabeled ON1 was prepared in 10 mM Tris-HCl (pH 7.5), 0.1 mM EDTA in the presence of 150 mM KCl with 5  $\mu$ M of unlabeled DNA. The induced GQ structure was formed as mentioned before with equimolar unlabeled RNA. The samples were treated with 1% DMS at room temperature for 2 minutes. The reactions were

stopped by adding stop buffer (2 M  $\beta$ -mercaptoethanol, 300 mM sodium acetate, 250  $\mu$ g/mL salmon sperm DNA) in an 11:1 ratio and ethanol precipitated. The DNA pellets were then dried and cleaved by the addition of 70  $\mu$ l of 10% piperidine and incubating at 90 °C followed by removal of the piperidine by vacuum drying. The pellets were dissolved in urea loading buffer and the fragments separated on a 15% denaturing polyacrylamide gel. The gel was dried and exposed to a phosphorimager screen and then visualized by scanning the screen on a Typhoon Phosphorimager FLA 9500 (GE Healthcare, Life Sciences).

#### **4.2.10. Plasmid construction**

A 43 nt segment of the 5'-UTR of eIF-4E mRNA was inserted into the psiCHECK-2 plasmid upstream of the *Renilla* luciferase gene (hRluc) by insertion mutagenesis and the plasmid henceforth will be referred to as psiC2-4E-UTR (Fig. 4.6 a). Primers were designed and ordered from Integrated DNA Technologies (IDT). Insertion mutagenesis was performed by QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent Tech) as per manufacturer's protocol and successful insertion into the plasmid was confirmed by sequencing at The Plant Microbe Genomic Facility (Ohio State University). The hVEGF-Bicis plasmid was constructed as previously mentioned (Morris et al., 2010).

#### **4.2.11 Reporter gene assay**

HeLa cells were grown in 96-well plates in Dulbecco's modified Eagle's medium (DMEM) with low glucose (HyClone®) supplemented with 10% fetal bovine serum and 1% antibiotics streptomycin and penicillin at 37 °C in 5% CO<sub>2</sub> in a humidified incubator. The plasmids psiC2-4E-UTR and hVEGFBicis were transfected by Lipofectamine 2000 as per manufacturer's protocol for 8 hours and then treated with 25 μM of ON1 and SW. After 24 hours of treatment of the ON1, dual luciferase assays were performed with Dual-Glo® Luciferase Assay System as per manufacturer's protocol and the results read by the SpectraMax M4 plate reader (Molecular Devices, LLC). In both cases, psiC2-4E-UTR and hVEGFBicis, the ratios of *Renilla* (L<sub>R</sub>)/Firefly(L<sub>F</sub>) luciferase expression were normalized with untreated expression levels.

#### **4.2.12 Quantitative RT-PCR**

HeLa cells were grown in 6-well plates in DMEM as described above. 25 μM of DNA oligonucleotides were added while seeding the cells. Total cellular RNA was extracted from HeLa cells using a TriPure Isolation Reagent (Roche Applied Sciences) as per manufacturer's protocol. The cDNA was synthesized using qScript™ cDNA SuperMix (Quanta Biosciences). The glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and eIF-4E mRNAs were subjected to qRT-PCR using a Perfecta® SYBR® Green Super Mix (Quanta Biosciences) on an Eppendorf Mastercycler® RealPlex2 in the presence of appropriate set of primers. The relative mRNA levels were estimated by the comparative C<sub>t</sub> method (Livak method). (Livak and Schmittgen, 2001; Schmittgen and Livak, 2008)

#### **4.2.13 Western blotting**

HeLa cells were grown as above and treated with 25  $\mu$ M of oligonucleotides for 24 hours. Proteins were extracted from the cells with RIPA buffer (sc-24948, Santa Cruz), 50  $\mu$ g of protein lysate was separated by 12% SDS-PAGE. The proteins were detected by mouse monoclonal eIF-4E antibody (P-2, sc-9976) at 1:200 dilution and GAPDH (G-9, sc-365062) antibody at 1:5000 dilution. Horseradish peroxidase-conjugated goat anti-mouse IgG (sc-2005) was used as secondary antibody at 1:1000 dilutions. Proteins were visualized by Western Blotting Luminol Reagent (sc-2048) in ChemiDoc-It®TS2 Imager (UVP, LLC)

#### **4.2.14 Cell Viability assay**

HeLa cells were seeded in a 96 well plate at 10,000 cells per well and treated with increasing concentrations of the oligonucleotides and incubated for 24 hours. CellTiter 96® Aqueous Non-Radioactive Cell Assay (Promega) was used to assess the cell viability by measuring the absorbance at 490 nm, as per manufacturer's protocol. The cell viability percentages calculated were normalized with untreated controls. The data were fitted using dose-response curve with a variable Hill slope to determine the EC<sub>50</sub> value in Origin 8.

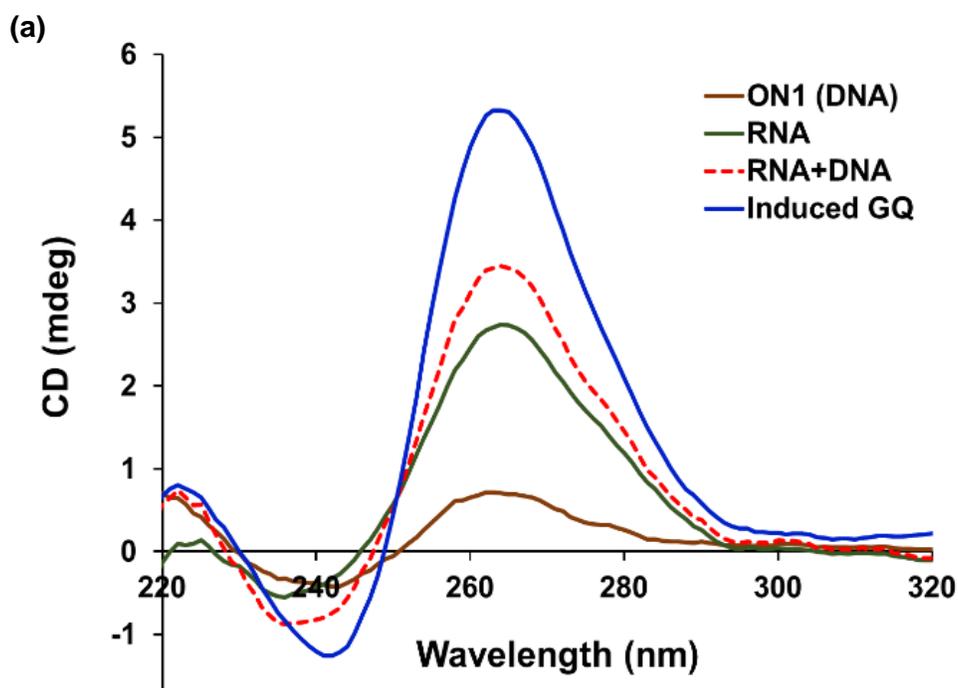
### **4.3 RESULTS**

### **4.3.1 RNA and DNA oligonucleotides associate to form a parallel induced hybrid GQ structure**

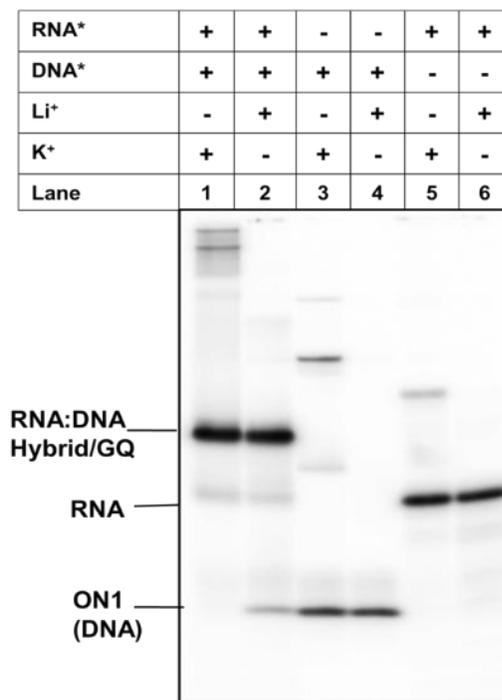
In order to determine the formation of induced GQ structures *in vitro*, we chose a fragment of the 5'-UTR of eIF-4E, which had the requisite characteristics to participate in induced GQ formation. A 36 nucleotide fragment from the human eIF-4E mRNA, which included a GGGAGGG sequence, was used for the initial biophysical and biochemical characterizations. The targeting DNA oligonucleotide ON1 (Table 4.1) and the target RNA were mixed in equimolar amounts (5  $\mu$ M) in the presence of physiologically relevant K<sup>+</sup> ion concentration (150 mM), which was followed by heating and slow cooling to form possible higher order structures.

The sharp increase in molar ellipticity (55%) at 263 nm and a significant decrease (50%) at 240 nm of the CD spectrum (Fig. 4.1 a) in the presence of an equimolar amount of DNA and RNA are indicative of the formation of an induced parallel hybrid (RNA:DNA) GQ. The formation of the induced GQ is further alluded to by the fact that when individual ellipticity values of 5  $\mu$ M of DNA and 5  $\mu$ M RNA were added, it showed a considerably lower intensity than the highest observed intensity of the induced GQ spectrum at 263 nm (Fig. 4.1 a). Parallel RNA:DNA hybrid GQ structures have been observed before (Hagihara et al., 2010) and might be due to the involvement of the RNA strand, as RNA GQs always adopts parallel GQ structures (Wanrooij et al., 2012). The CD spectra also suggest that the target RNA alone forms some higher order structure, which can potentially be due to the formation of an all RNA intermolecular GQ.

The formation of the bimolecular RNA: DNA hybrid GQ should result in a larger complex than either the unstructured RNA or the DNA individually and hence can be observed by native gel electrophoretic mobility shift in a gel. The electrophoretic mobility shift assay was performed with radiolabelled RNA and DNA in the presence of 150 mM of  $\text{Li}^+$  or  $\text{K}^+$ . To maintain the oligonucleotide concentration at 5  $\mu\text{M}$  during the CD experiments, the radiolabeled RNA and ON1 were mixed with their unlabeled versions.



(b)



**Figure 4.1- *In vitro* formation of induced GQ:** (a) CD spectra of induced GQ formation. The spectral pattern of the individual oligonucleotides concentration of 5  $\mu\text{M}$  of DNA (—) and RNA (—) are shown in brown and green solid lines respectively. The induced GQ formed by the 5  $\mu\text{M}$  of RNA and DNA (—) exhibits an increase in the 263 nm absorbance compared to the trace (- -) of the added CD spectrum of individual oligonucleotides. (b) The native gel mobility shift assay shows the association of RNA and DNA in the presence of  $\text{K}^+$  and  $\text{Li}^+$  (Lane 1 and 2).

The mobility shift assay results (Fig. 4.1 b, lanes 1 and 2) clearly indicate the association of RNA and DNA in the presence of  $\text{K}^+$  and  $\text{Li}^+$  respectively. The higher order structure formed by RNA in the presence of  $\text{K}^+$  is probably a bimolecular GQ (lane 5) which corroborates with our CD data. Based on the observation of reduced mobility bands, the ON1 in the presence of  $\text{K}^+$  (Lane 3) very likely forms various higher order species. The CD data in combination with the gel mobility shift data indicate an association of ON1 with the target RNA to form a parallel induced GQ.

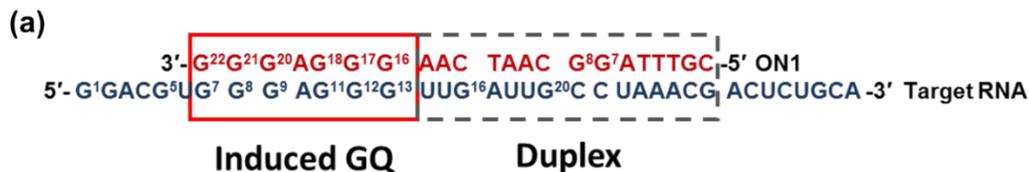
### **4.3.2 Enzymatic and chemical footprinting data indicate the formation of an induced GQ and the duplex structures in tandem**

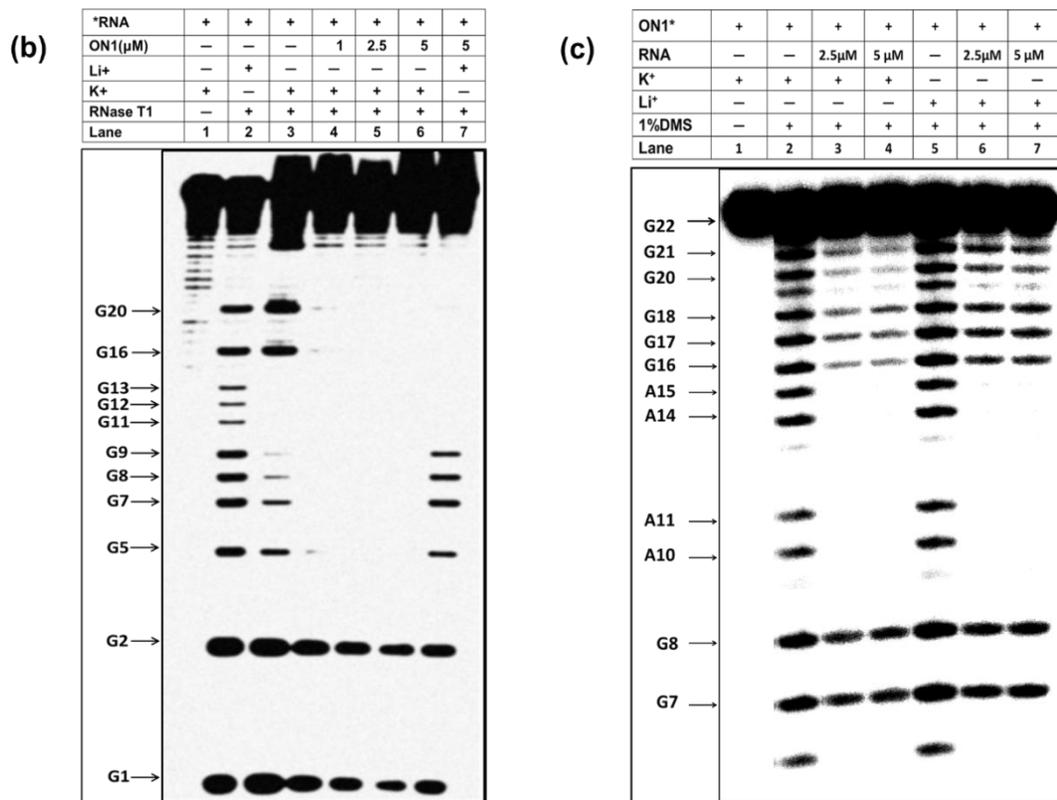
The formation of the induced GQ structure was further confirmed by both enzymatic and chemical footprinting. RNase T1 cleaves single-stranded G-residues in RNA and protection of the G-residues from the cleavage suggests that they are very likely involved in duplex or quadruplex formation. The residues G 7-9 and G 11-13 of the target RNA are designed to be involved in induced GQ structure with the ON1, while the residues G 16 and G 20 would form the duplex (Fig. 4.2 a).

The cleavage (Fig. 4.2 b) of all the G-residues in the 36 nt 5'-radiolabeled RNA was observed in lane 2 in the presence of  $\text{Li}^+$ . The monovalent cation  $\text{Li}^+$  was used as a control as it has been well established that  $\text{Li}^+$  does not stabilize GQ structures. On addition of the DNA oligonucleotide, ON1 (1-5  $\mu\text{M}$ ) in the presence of  $\text{K}^+$ , complete protection was observed both in the putative quadruplex and duplex regions respectively (Fig. 4.3b, lanes 4, 5 and 6). The complete protection pattern persisted in the presence of  $\text{K}^+$  even at the lowest ON1 concentration (1  $\mu\text{M}$ ). However, when the  $\text{K}^+$  was swapped with  $\text{Li}^+$ , band protection in the GQ region (G 7-9) decreased even in the presence of highest used concentration of ON1 (5  $\mu\text{M}$ ). The protection of a subset of residues (G 11-13) persisted which might be due to structural hindrance caused by the proximal duplex region. Thus, the RNase T1 footprinting results clearly indicate the formation of induced RNA: DNA hybrid GQ and the duplex structures in the presence of  $\text{K}^+$ , while in  $\text{Li}^+$  the induced GQ structure is significantly destabilized. We did observe a few G residues in the RNA only lane that showed slight protection in the GQ forming region in presence of

K<sup>+</sup> (Lane 3), which can be explained by the formation of intermolecular GQ structures by the RNA as was previously suggested based on the observations made during CD and gel mobility shift assays.

To further confirm the formation of induced GQ structure, we performed dimethyl sulfate (DMS) footprinting on the radiolabeled ON1 in presence and absence of the unlabeled target RNA. DMS methylates the N7 position of guanine residues and N3 position of adenine (Maxam and Gilbert, 1977). Since the N7 position of the G is involved in the Hoogsteen base-pairing to form the quartets in a GQ, their protection is an indication of the formation of such a structure. The DMS modified all of the G and A residues in the ON1 (Fig. 4.2 c, lane 2). However when the induced GQ structure is formed in the presence of unlabeled target RNA, the A-residues in the ON1 are resistant to methylation, as the N3 in the Watson-Crick face of A is involved in duplex formation. The G-residues in the induced GQ forming region (G 16-18, G 20-21) are strongly protected due to the formation of induced GQ structure in the presence of 150 mM K<sup>+</sup> (Fig. 4.2 c, lane 3 and lane 4) compared to equimolar Li<sup>+</sup> (Fig. 4.2 c, lane 6 and lane 7). Thus, RNase T1 and DMS footprinting confirm the formation of an induced RNA: DNA hybrid GQ structure in the presence of physiologically relevant concentration of K<sup>+</sup> ions.



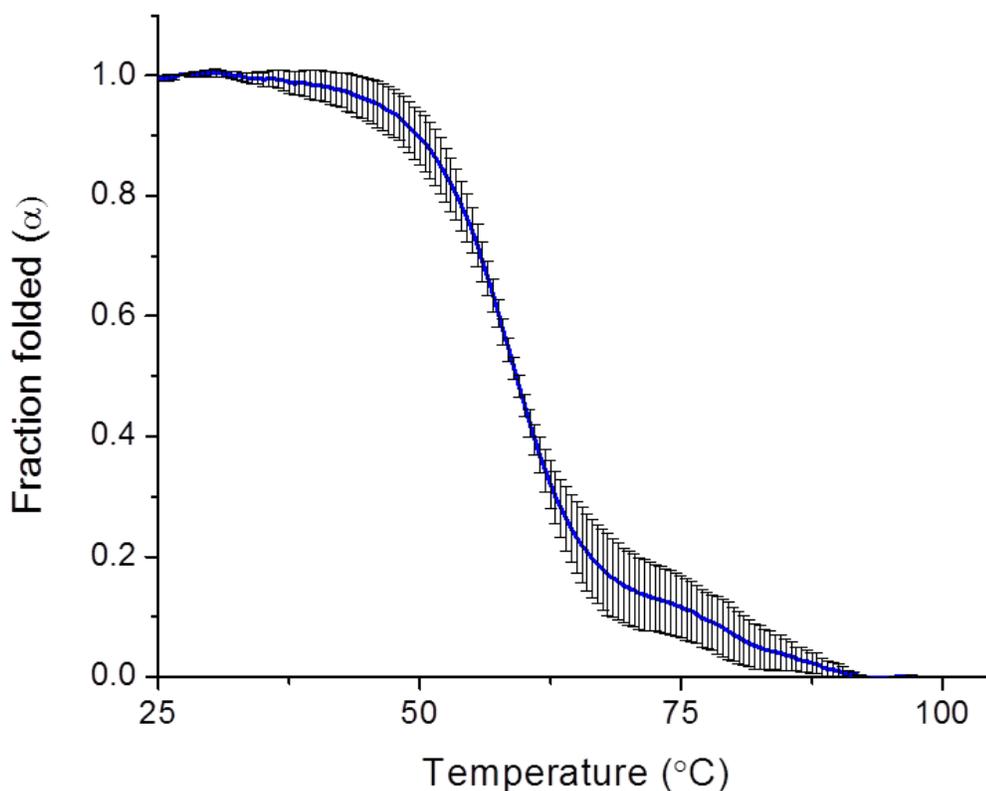


**Figure 4.2- Enzymatic and chemical structure mapping of induced GQ:** (a) Schematic representation of the association of the target RNA sequence with the ON1. The G-residues are numbered for both the RNA and the DNA that are involved in the induced GQ and the duplex formation. The induced GQ and the duplex regions are indicated by separate boxes. (b) Image of a gel showing RNase T1 footprinting of the 5'-radiolabeled target RNA in presence and absence of targeting oligonucleotide ON1 in K<sup>+</sup> and Li<sup>+</sup>. The formation of induced GQ is shown by the protection of the GQ forming residues (G-7, 8, 9, 11, 12, 13) in lanes 4, 5 and 6 in the presence of ON1 and K<sup>+</sup>. In the presence of Li<sup>+</sup> (lane 7), the induced GQ is destabilized as observed by the reappearance (no protection) of the bands. (c) DMS footprinting of the 5'-radiolabeled ON1 in absence and presence of the target RNA and K<sup>+</sup> (lanes 3 and 4) clearly shows the increased protection compared to Li<sup>+</sup> ion (lanes 6 and 7) in the induced GQ forming region in the presence of both RNA and DNA.

### 4.3.3 The induced GQ structure is stable at physiological temperature and is resistant to nuclease digestion

Once the formation of the induced GQ structure was established by enzymatic and chemical footprinting, the stability of the structure was analyzed under the exact same conditions that are amenable to induced GQ formation i.e. 150 mM K<sup>+</sup>, 5 μM of each of RNA and ON1. CD melting experiments were performed by monitoring changes in CD intensity at 263 nm. The T<sub>m</sub> of the structure was observed to be 59.6 ± 0.8 °C that is in agreement with the T<sub>m</sub> of previous studies on similar structures (Hagihara et al., 2010). The annealing and melting curves (Fig. 4.3) showed reversible characteristic, which suggests that the molecules were at thermodynamic equilibrium.

The T<sub>m</sub> values calculated from van't Hoff plots and by first-derivative analyses agreed with each other. Thermodynamic parameters from the melting curves were calculated based on a two-state model that considers only the folded and unfolded states of the molecules. In the presence of 150 mM K<sup>+</sup>, the Gibbs free energy (ΔG°) at 37 °C was calculated to be -19.2 ± 1.4 kJ/mole indicating a stable induced GQ. The calculated ΔH° of -283.6±24.8 kJ/mole and ΔS° of -0.9±0.1 kJ mol<sup>-1</sup> K<sup>-1</sup> values (Table 4.2) compare well with previously reported values for GQs (Rachwal et al., 2007a). Thus, the free energy of formation of induced GQ is solely driven by the favorable change in enthalpy.



**Figure 4.3- Stability of induced GQ:** The normalized CD melting profile of the induced GQ structure formed assuming a two-state model. The melting profile was monitored by the change in molar ellipticity at 263 nm over a temperature range of 25°C to 97°C. The result represents three continuous cycles of melting and cooling and followed by melting respectively. Error bars indicate the standard error of the mean of fraction folded ( $\alpha$ ).

. The stoichiometry ( $N$ ) of the induced GQ formation by Isothermal Titration Calorimetry (ITC) was measured to be  $1.15 \pm 0.026$ . The 1:1 association of the target RNA and ON1 oligonucleotide conforms well to the expected formation of a bimolecular induced GQ structure presumably with a duplex in tandem.

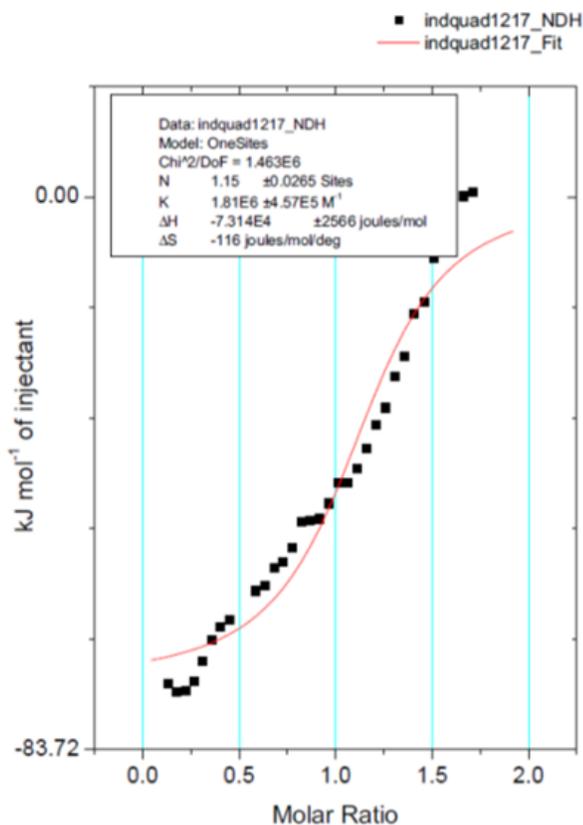
| <b>Thermodynamic parameters of induced GQ formation</b> |                           |                              |                     |                                                     |                                      |
|---------------------------------------------------------|---------------------------|------------------------------|---------------------|-----------------------------------------------------|--------------------------------------|
|                                                         | <b>T<sub>m</sub> (°C)</b> | <b>N<br/>(stoichiometry)</b> | <b>ΔH° (kJ/mol)</b> | <b>ΔS°<br/>(kJ mol<sup>-1</sup> K<sup>-1</sup>)</b> | <b>ΔG°<sub>37</sub><br/>(kJ/mol)</b> |
| <b>CD</b>                                               | <b>59.6 ± 0.8</b>         |                              | <b>-284 ± 25</b>    | <b>-0.85 ± 0.08</b>                                 | <b>-19.2 ± 1.4</b>                   |
| <b>ITC<br/>(at 25 °C)</b>                               |                           | <b>1.15 ± 0.03</b>           | <b>-73 ± 3</b>      | <b>0.12</b>                                         | <b>-38.55</b>                        |

**Table 4.2- Thermodynamic parameters of induced GQ formation:** Thermodynamic parameters calculated from CD melting experiments, first derivative analysis and Isothermal Titration Calorimetry Data represented as mean ± standard error of the mean (SEM).

The 1:1 association of RNA and DNA oligos conforms well to the expected formation of a bimolecular induced GQ structure presumably with a contiguous duplex. The stoichiometry also eliminates the possibility of a tetramolecular induced GQ and intermolecular GQs among RNAs or DNAs. The difference in the enthalpy and free energy values amongst the ones calculated using van't Hoff parameters and the values obtained by ITC may be due to the fact that the CD melting monitors only the GQ structure and assumes a two-state model while the isothermal calorimetry provides a more comprehensive picture of the thermodynamics of induced GQ formation in its entirety comprising of both the duplex and the quadruplex.

We also tested the induced GQ structure for its ability to resist nuclease digestion because nucleases will be less effective in cleaving a compact structure. The induced GQ structure formed in 150 mM of K<sup>+</sup> with 5 μM of RNA and ON1 each was treated successively with DNase 1 and RNase T1. There were no significant quantitative and qualitative changes of the CD spectrum even after 15 minutes of incubation at 37 °C (Fig.

4.5). The results described above suggest that the target RNA and ON1 associate in a 1:1 stoichiometry to form an induced GQ and this duplex is inaccessible to nucleases at physiological temperature.

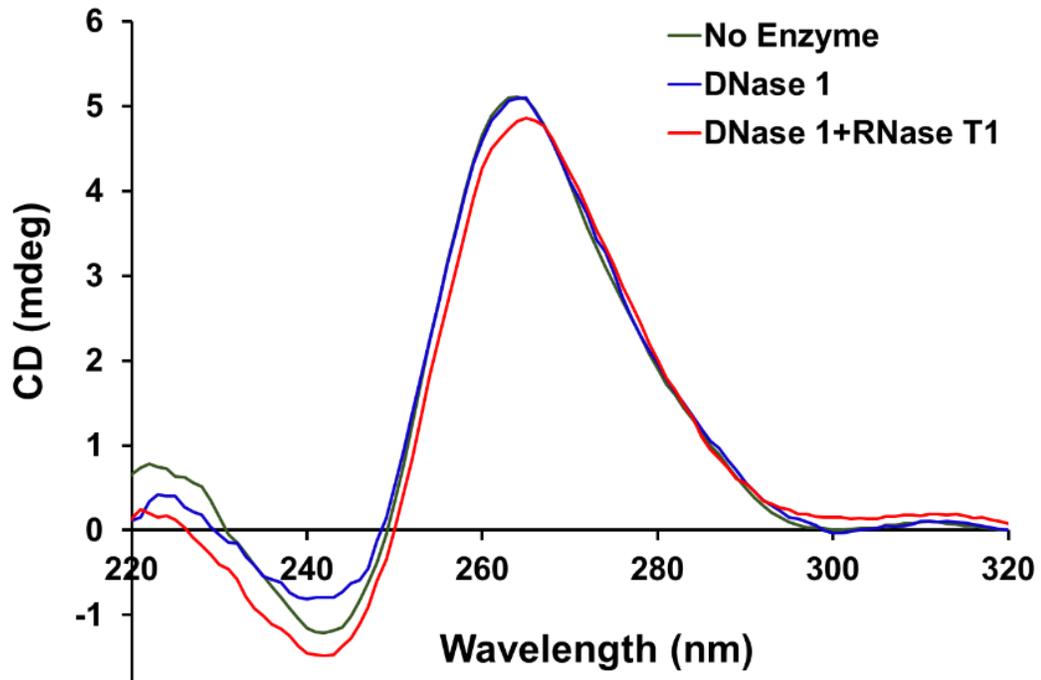


**Figure 4.4:** Calorimetric data for the titration of the formation of induced GQ. Binding isotherms were obtained from the integration of raw data and fitted to a “one-site” model

#### 4.3.4 Induced GQ structures specifically downregulate the expression of a targeted gene in a reporter plasmid

The GQ structures when present within an mRNA modulate its translation. Thus, a GQ structure artificially induced on an mRNA in a targeted fashion might be able to downregulate its translation. The prerequisites for GQ induction on the target mRNA is

the lack of endogenous PQS and an IRES, but the presence of at least two G-stretches. A search for an optimum target resulted in human eIF-4E mRNA as it contains all the needed characteristics for studying induced GQ formation. To test whether induced GQ



**Figure 4.5- The induced GQ is resistant to nucleases:** CD spectra of the induced GQ structure formed by *in vitro* transcribed RNA and ON1 in 150 mM K<sup>+</sup> before and after being treated with either DNase1 or both DNase1 and RNase T1 enzymes.

structures can modulate gene expression in cells, a dual luciferase reporter plasmid was constructed with a 43 nt fragment from the 5'-UTR of human eIF-4E mRNA inserted just upstream of the *Renilla* luciferase gene (Fig. 4.6 a). The plasmid was then transfected into HeLa cells and the cells were treated with DNA oligonucleotides (Table 4.1) for 24 hours following which the dual luciferase activities were measured. The DNA

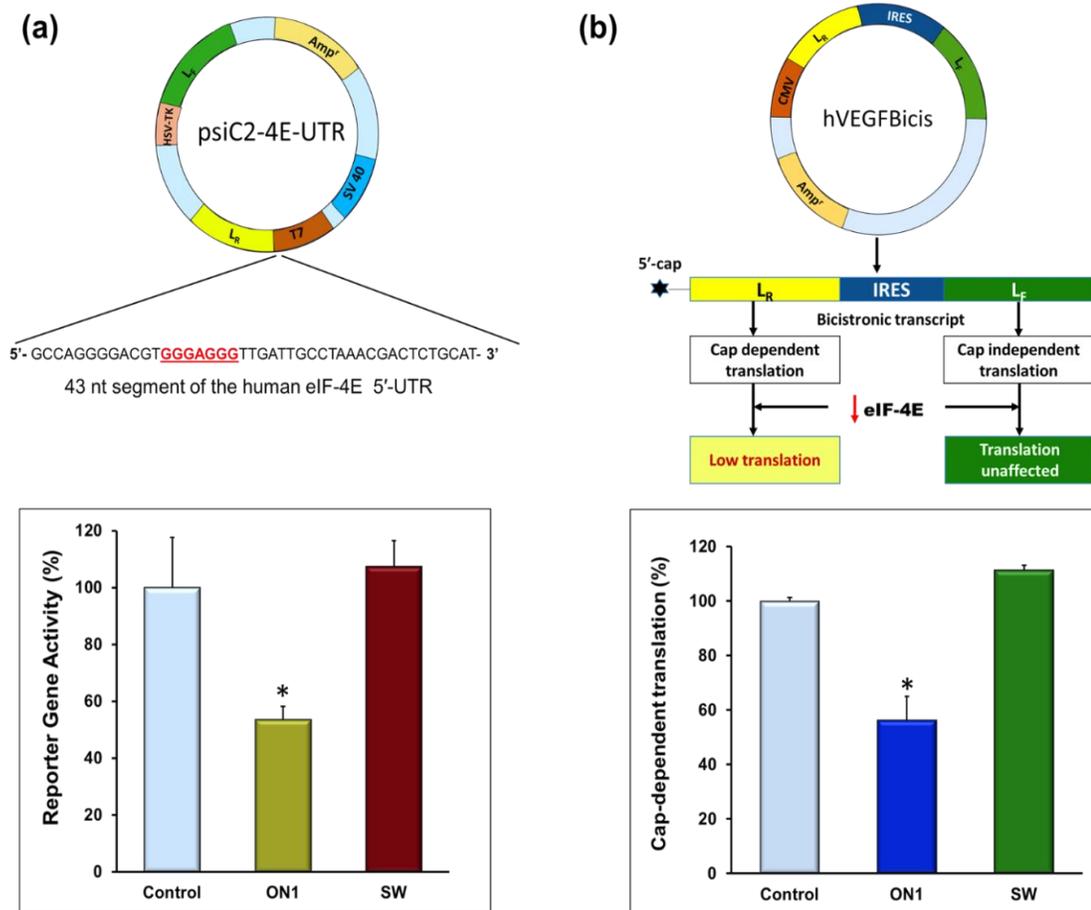
oligonucleotides contained phosphorothioate modification at two nucleotides on both ends to provide resistance against exonucleases (Dias and Stein, 2002). The phosphorothioate modification is not known to interfere with the formation of GQ structures.(Sacca et al., 2005) The ON1 reduced the expression of the reporter gene by ~44 % compared to the untreated cells (Fig. 4.6 a). The whole scrambled control (SW) had no significant change in the expression of the reporter gene. The results indicate that GQ structures repressed translation when induced in the 5'-UTR of reporter plasmids and the repression level is in the range similar to what was observed previously for naturally occurring intramolecular RNA GQ structures (Bugaut and Balasubramanian, 2012).

It is well established that IRES-dependent translation does not require eIF-4E for initiation (Pestova et al., 2001; Jackson et al., 2010a; Perard et al., 2010; Komar et al., 2012; Plank and Kieft, 2012). Thus, the downregulation of eIF-4E that is known to affect the cap-dependent translation initiation should not affect IRES-driven (cap-independent) translation initiation. The bicistronic reporter plasmid hVEGFBicis (Fig. 4.6 b) expresses the *Renilla* luciferase by cap-dependent translation initiation whereas the Firefly luciferase by cap-independent translation under the influence of human vascular endothelial growth factor (hVEGF) IRES A (Morris et al., 2010). The hVEGFBicis was used to observe the effect of the induced GQ structures on endogenous eIF-4E expression with the expectation that the lowering of eIF-4E level would negatively affect cap-dependent translation initiation while the cap-independent translation would remain unaffected (Fig. 4.6 b).

As hypothesized, HeLa cells transfected with the plasmid when treated with ON1 exhibited downregulation of cap-dependent translation as was measured by lowering of the *Renilla*/Firefly activity. However, *Renilla*/Firefly ratio did not show any change when the cells were treated with the control oligonucleotide SW most likely due to its lack of effect on eIF-4E expression. The inhibition of cap-dependent translation by induced GQ formation ON1 established the target specificity of the strategy.

#### **4.3.5 Targeted induction of GQ structure downregulates the endogenous human eIF-4E expression by translation repression**

A reporter construct with an insert derived from the naturally occurring eIF-4E transcript showed reduced reporter gene expression in cancer cells when treated with ON1, presumably via induced GQ formation, which indicates that the region within eIF-4E is a viable target for induction of GQ. In a separate assay, reduction in the expression of the transcript under cap-dependent translation initiation was observed in a dual luciferase reporter gene assay while the cap-independent translation initiation remained unaltered, which provided a strong indirect evidence that endogenous eIF-4E expression was inhibited. This prepared the deck for a direct measurement of the effect of induced GQ formation on endogenous eIF-4E expression level.



**Figure 4.6- Induced GQ targeting eIF-4E downregulates reporter gene expression:** (a) The repression of reporter gene *Renilla* luciferase in psiC2-4EUTR by formation of induced GQ structures by ON1 compared to the untreated and whole scrambled (SW) oligonucleotide (SW) treated cells (b) cells transfected with hVEGFBicis plasmids exhibited repression of cap-dependent translation initiation when treated with ON1. The ratios of *Renilla* (LR)/Firefly (LF) luciferase expression was normalized to untreated expression levels in both assays. Data represented as mean  $\pm$  SEM, and the significance of the data was determined by t-test analysis ( $p < 0.0005$ ).

The human eIF-4E gene produces three transcripts wherein sequence amenable to induced GQ formation in the 5'-UTR region is present in two of the three transcripts whereas a similar sequence was present in the coding region of all three transcripts (Table 4.3). The targeting oligonucleotides ON1 and ON2 consist of two functional segments, a guiding sequence (GS) that is designed to hybridize to the target mRNA forming a duplex

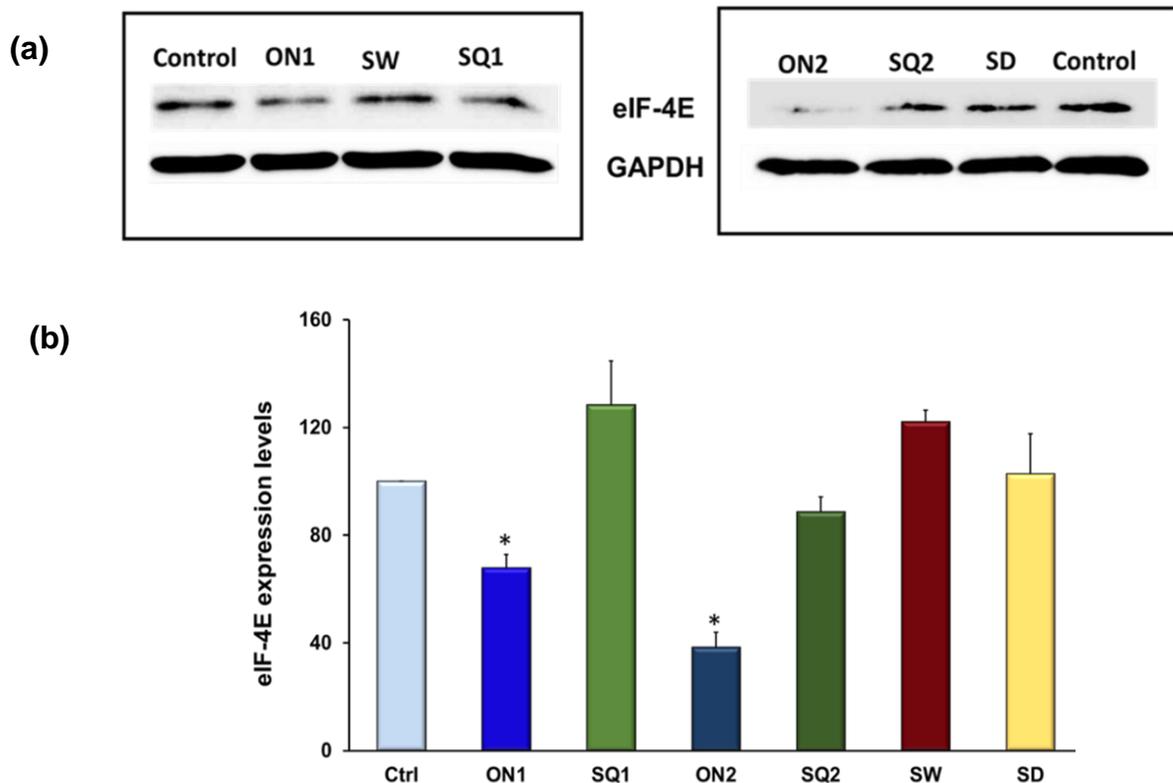
and two G-rich stretches to induce GQ by enlisting two G-stretches from the target RNA. The 15 nt duplex region formed by ON1 and the target RNA can elicit an antisense mechanism potentially causing an inhibition of translation either by an RNase H-mediated degradation of the targeted mRNA or a translation blockade (Flanagan et al., 1996; Dias and Stein, 2002; Chan et al., 2006).

| TARGET SEQUENCE                |     | TARGET REGION         | POSITION OF TARGETS IN THE TRANSCRIPT |                                     |                                  |
|--------------------------------|-----|-----------------------|---------------------------------------|-------------------------------------|----------------------------------|
|                                |     |                       | Transcript 1                          | Transcript 2                        | Transcript 3                     |
| <b>GGGAGGG</b> UUGAUUGCCUAAACG | ON1 | 5'- UTR               | 951-972                               | 951-972                             | Not present                      |
| <b>GGGAGGG</b> UAUACAAGGAAAGGU | ON2 | Protein coding region | 2061-2082<br>( <sup>1524</sup> AUG)   | 2154-2175<br>( <sup>1524</sup> AUG) | 718-739<br>( <sup>121</sup> AUG) |

**Table 4.3: The target sequences for inducing GQ by ON1 and ON2 in the 5'- UTR and the coding region and their position in the human eIF-4E transcript**

Intramolecular GQ structures repress translation via a translation blockade and mostly with unaltered mRNA levels. It was, therefore, necessary to determine the effects of each of the functional regions in isolation and in combination and appropriate controls were designed to address both of those issues (Table 4.1). In one of the controls the quadruplex forming region was scrambled (SQ1) keeping the duplex region unaltered, in the second the duplex forming region was scrambled (SD) keeping the segment that would participate in GQ formation unchanged and finally, one was designed in which the entire sequence was scrambled (SW). HeLa cells were treated with partially modified ON1, ON2, and various control oligonucleotides without any delivery vehicle. The endogenous levels of eIF-4E were measured by Western Blotting (Fig. 4.7 a) using GAPDH as a control and showed a 30% reduction in the eIF-4E levels in ON1 treated

samples and 60% reduction in ON2 treated samples based upon densitometric analysis by Image J software (Fig. 4.7 b). The effects of the duplex forming sequences (SQ1 and SQ2) on eIF-4E levels were not evident from the measured protein levels. The effect of control oligonucleotides SW and SD on protein levels were also insignificant (t-test analyses). The targeted downregulation of the eIF-4E by induced GQ forming oligonucleotides and near absence of any effect in the control samples increased our confidence in the success of such targeted repression strategy.



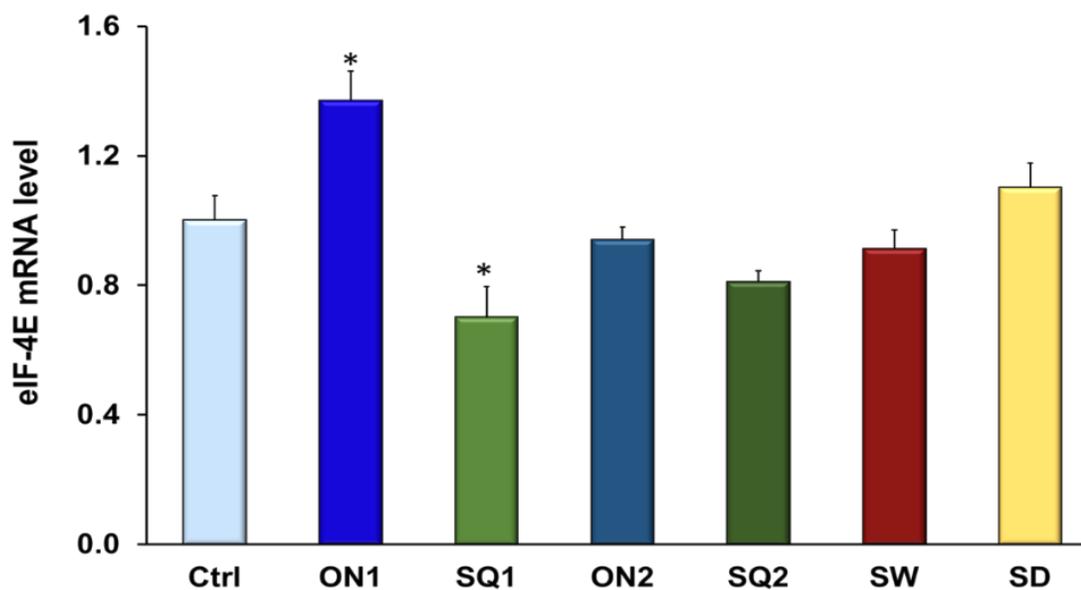
**Figure 4.7- Induced GQ downregulates endogenous eIF-4E protein expression:** eIF4E protein levels in HeLa cells treated with ON1, ON2, and various control oligonucleotides. (a) Western blot of the eIF-4E protein expression levels when treated with various oligonucleotides. (b) Histogram representing densitometry analysis of the western blot images from three independent experiments was performed using Image J software. The eIF-4E bands were normalized with corresponding GAPDH bands to

correct for experimental loading errors. Data represented as mean  $\pm$  SEM and the significance of the data was determined by t-test analysis. ( $p < 0.01$ )

To investigate if the reduction of protein levels was because of translation repression in contrast to RNase H-mediated digestion, quantitative reverse transcription PCR was performed to estimate the mRNA levels when the cells are treated with the targeting and control oligos (Fig. 4.8). Although the protein level of the ON1 treated cells showed about 30% reduction, the mRNA level of ON1 treated cells was not reduced and, in fact, it was 37% higher than the untreated. However, the SQ1 and SQ2 treated cells showed ~30% and ~20% reductions respectively in the mRNA levels. The ON2, SW, and SD treated cells had no significant change in the eIF-4E mRNA levels compared to untreated cells. The ON1 sequence targets the 5'-UTR of eIF-4E mRNA that leads to translation repression via formation of a stable secondary structure on the mRNA. It has been previously reported that formation of a stable secondary structure on the 5'-UTR of some AU-rich mRNAs result in stabilization of the mRNAs (Curatola et al., 1995). Given that eIF-4E mRNA harbors AU-rich elements in its 3'-UTR, it is likely that the GQ formation on its 5'-UTR stabilized the mRNA resulting in its higher expression level. In addition, it is clear from the data that the repression in protein level has negligible if any contribution from the classic RNase H-mediated antisense effect.

#### **4.3.6 Targeting endogenous eIF-4E mRNA by GQ inducing sequences inhibits proliferation of human cancer cells**

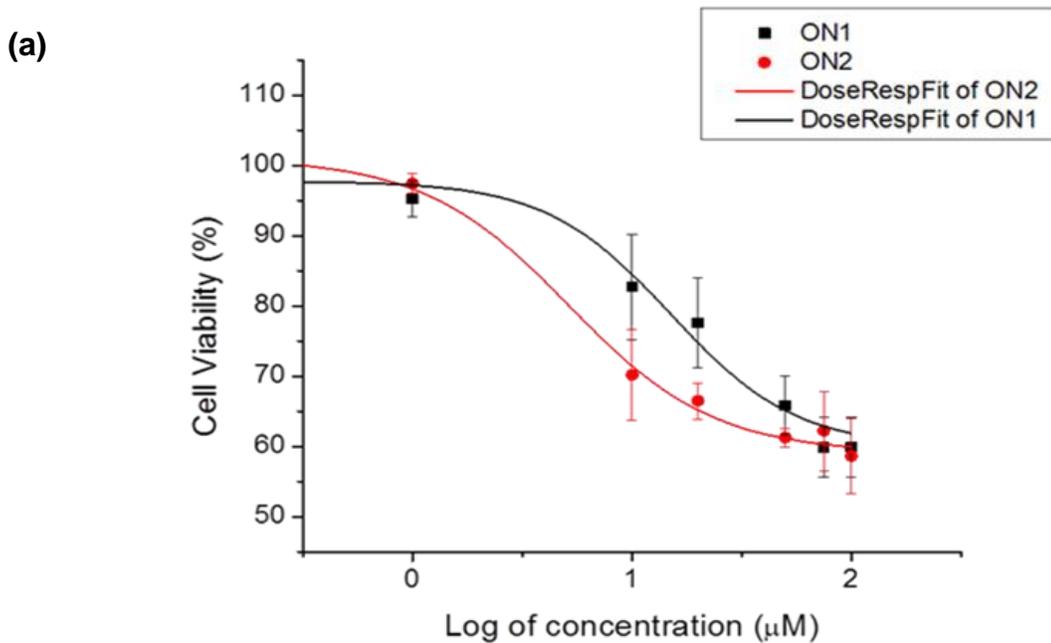
Eukaryotic initiation factor-4E (eIF-4E) is essential for initiation of cap-dependent translation and protein synthesis. Because of its least abundance among translation initiation factors, eIF-4E acts as the rate limiting agent in translation initiation (Mamane et al., 2004; Jacobson et al., 2013). The repression of endogenous eIF-4E levels by induced GQ structures led us to investigate their effects on cell proliferation. We used ON1 and ON2 to determine their anti-proliferative activities in HeLa cells. Cells were grown in the presence of targeting and the control oligonucleotides and cell viability were observed by MTS assay after 24 hours of treatment. Induced GQ forming oligonucleotides ON1 and ON2 had a dose-dependent antiproliferative effect on the cancer cells with  $EC_{50}$  values of 15  $\mu$ M and 5.4  $\mu$ M respectively (Fig. 4.9 a).

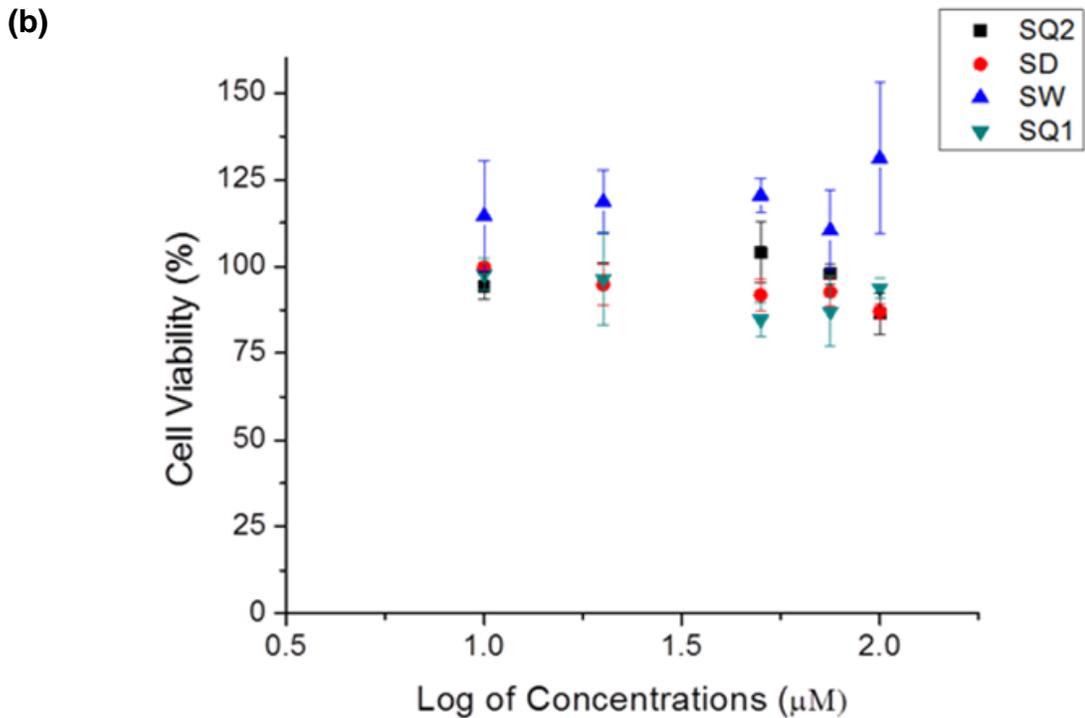


**Figure 4.8- Induced GQ structures do not alter mRNA levels:** To investigate the mechanism of induced GQ mediated effect on protein production that can be either due to mRNA degradation or by translation arrest or a combination of both, eIF-4E mRNA

levels were measured by qRT-PCR in ON1, ON2 and control oligonucleotides treated HeLa cells.

The control oligonucleotides SQ1, SQ2, SW and SD did not exhibit any significant loss of cell viability (Fig. 4.9 b) suggesting a lack of any antisense mechanism in downregulation of eIF-4E expression. The anti-proliferative action of GQ inducing ON1 and ON2 agrees well with their effect on the downregulation of endogenous eIF-4E levels. Interestingly targeted induction of GQ in the coding regions had more robust effect in terms of both repressions of protein expression and inhibition of cell proliferation. Thus, rationally targeted induced GQ formation not only downregulates a gene of choice but also effects antiproliferative activity on human cancer cells.





**Figure 4.9- The Anti-proliferative effect of induced GQ structures on cancer cells:** The dose-response effect of increasing concentrations of ON1, ON2 and control oligonucleotides on HeLa cell viability. (a) The  $EC_{50}$  value of ON1 was observed to be 15  $\mu\text{M}$  and ON2 to be 5.4  $\mu\text{M}$  when the data were fitted using dose-response curve with variable Hill slope. (b) The control oligonucleotides exerted no significant effect on the HeLa cell proliferation. Data represented as mean  $\pm$  SEM

#### 4.4 DISCUSSION

The purpose of this study is to establish that the expression of an endogenous targeted gene can be modulated at the translational level by inducing RNA: DNA hybrid GQ structures. It is suggested that in the case of cap-dependent translation initiation, the formation of an RNA GQ in the 5'-UTR might act as a hindrance to the 5'-3' scanning of the 43S pre-initiation complex which in turn downregulates the translation (Bugaut and Balasubramanian, 2012). Additionally, the formation of a GQ within the coding region

has been shown to result in aborted translation resulting in truncated protein products (Endoh et al., 2013a). At least four G-stretches must be available for a GQ formation and the G-stretches can be provided by one or more strands, which determine the molecularity of the GQ. The stability of the GQ structures depends on the number of quartets among other factors, for example, it is well known the stability of three-tiered GQ structure is more stable than the two-tiered GQs (Pandey et al., 2013) and the majority of the GQs discovered so far are three-tiered. The propensity of occurrence of potential RNA: DNA hybrid GQ enabling sequences are significantly higher compared to potential intramolecular GQ forming sequences and such structures have been known to be formed naturally during transcription (Zheng et al., 2013; Zhang et al., 2014a). Hence, we decided to induce a GQ structure that i) would be stable under normal physiological conditions of temperature and  $K^+$  concentrations and ii) mimic naturally occurring structures and mechanisms in cells.

The formation of a stable induced GQ structures was comprehensibly determined by a diverse set of biochemical and biophysical methods. Induction of the GQ would downregulate the target eIF-4E level that in turn would adversely affect canonical translation initiation. We first used a dual reporter based assay in which we introduced a segment from the 5'-UTR of the eIF-4E mRNA and it harbored two G-stretches of three guanines each. An extraneous DNA sequence (ON1) rationally targeted to the introduced sequence was able to downregulate the reporter gene. Since the introduced reporter plasmid contained sequences inherent to eIF-4E mRNA, when the cells were treated with ON1 sequence, it also should downregulate the endogenous eIF-4E level. We assumed

that under such condition of lower eIF-4E level both *Renilla* luciferase gene and the control Firefly luciferase gene were affected similarly. The downregulation of eIF-4E was further suggested by the observation of a decrease in the ratio of expression of two genes in a bicistronic construct where the translation of the mRNA under the canonical control of translation initiation was normalized to the IRES-controlled one.

The true test for any gene expression inhibition strategy is to be able to downregulate an endogenous gene *in cellulo*. We used human cervical cancer cells to test our strategy and to our knowledge, the effect of repression of eIF-4E in such cell line is not known. The eIF-4E sequence allowed us to target two different regions of the transcripts. The ON1 that targeted the 5'-UTR of eIF-4E showed a little over 30% repression while about 60% inhibition was observed in the presence of ON2 that targeted the coding region of the transcripts. A possible reason for the enhanced repressive effect of ON2 may be due to the fact that there are three eIF-4E transcripts and the target segment for ON1 only occurs in two of them, while the ON2 target exists in all three transcripts. We established that the repression of eIF-4E occurs at the translation level from the qRT-PCR results. We also observed that the levels of eIF-4E mRNA were increased with ON1 treatment. The eIF-4E mRNA harbors AU-rich (ARE) elements in its 3'-UTR (Topisirovic et al., 2009). Previous reports have suggested that translational repression of ARE embedded reporter RNAs containing stable secondary structures increased the half-life of those mRNAs thus increasing the mRNA level (Curatola et al., 1995; Barreau et al., 2005). Thus, the presence of the stably induced GQ structure may have caused translation blockade resulting in increased eIF-4E mRNA level.

The differential effects of ON1 and ON2 on translation repression of eIF-4E may also be due to the unrelated mechanism of inhibition. While the ON1 is expected to cause translation blockade by interfering with ribosome scanning that affects its ability to efficiently start protein synthesis, induced GQ formation in the protein coding region by ON2 might act by early termination of protein synthesis thereby leading to proteolysis of the truncated protein product and decrease in the eIF-4E expression. We are able to draw this inference based upon a previous report that has demonstrated that GQ structures when present in open reading frame slow down or temporarily stalls the translation elongation followed by proteolysis (Endoh et al., 2013a). The data suggest a lack of any classic RNase H-dependent antisense mechanism. For example, the SQ1 and SQ2 control sequences showed a weak downregulation of mRNA which did not translate into lowering of protein expression or cell growth inhibition. The ON2 showed ~60% reduction of protein expression but SQ2 did not show any lowering of protein expression. Thus, the duplex alone failed to elicit any response at the protein level or on cell growth. The oligonucleotide sequences used in the report showed low micromolar EC<sub>50</sub> values. Since our primary goal was to establish a proof of concept of the strategy, we did not employ any external delivery vehicle for the oligonucleotides while treating the cells. We relied solely on the natural ability of the oligonucleotides to enter the cells and elicit their responses. We believe that use of a delivery method would lower the EC<sub>50</sub> values of the oligonucleotides.

The strategy described in this work enables us to utilize multiple targets within an mRNA independently yet simultaneously. In fact, our studies on targeting both the 5'-

UTR and the protein coding region of eIF-4E mRNA simultaneously by ON1 and ON2 resulted in the moderate synergistic effect on anti-proliferation of cancer cells (data not shown). Thus, targeting of multiple sites within transcripts of a gene using this approach may greatly enhance the efficacy of lowering of gene expression.

## CHAPTER FIVE

### 5. CONCLUSION

The non-canonical secondary structures observed in G-rich regions of both DNA and RNA called the GQ structures were observed to be responsible for regulation of several key biochemical processes in the cell. More specifically these structures in the 5'-UTR of mRNA have been shown to modulate translation by several research groups. During the initial days of the dissertation, it was widely believed that the GQ structures in the 5'-UTR of mRNA repress translation except when they are present in internal ribosome entry sites where they augment translation.

During the course of our research, we discovered a novel mechanism of cap-independent translation initiation in a cellular IRES. A GQ, a non-canonical RNA secondary structure, regulated this IRES dependent initiation. The finding is unique by the nature of its interaction and thereby provides us a mechanistic explanation of the enhancing role of RNA GQs in the IRESs. However, it still remains to be established if these GQ structures always augment the IRES-mediated translation by recruiting the 40S ribosomal subunit. This instance of the recruitment of the 40S ribosomal subunit by a GQ domain in cap-independent translation initiation contributes to our nascent understanding of the structural basis of a key aspect of cellular IRES function. RNA GQ structures are known to play diverse roles in several biochemical processes such as pre-mRNA polyadenylation, splicing, mRNA targeting, transcription termination among several

others (Millevoi et al., 2012), and a GQ's direct role in interacting with the 40S subunit adds on to that list of functions. Given the well-established involvement of hVEGF in metastasis and tumor angiogenesis, the GQ located within its 5'-UTR can potentially be a new antitumor target. Because of hVEGF's normal physiological role, potentially it can also be a pro-angiogenic target if a ligand can stabilize the GQ domain. Furthermore, the GQ forming sequence can be a viable target for antisense thereby destabilizing the GQ to reduce hVEGF translation.

During the course of our work, we also observed that the GQ structures escape their native function and adopt an alternate function based on the molecular context in which they are present, making it an apparent general structure-function relationship phenomenon. Furthermore, a GQ motif that is not associated with a defined inherent function, exhibits considerable structural integrity irrespective of a non-native context but displays significant functional adaptability. Although there is considerable structural heterogeneity including subtle differences among GQ structures, it appears that a degree of modularity is sufficient to fulfill their context-dependent functional role in spite of some variation in their primary sequence alluding to the flexibility in the structural requirement.

We were also able to induce GQ structures by specifically targeting the eIF-4E mRNA and thereby repress the translation of the eIF-4E which resulted in anti-proliferation of cancer cells. The anti-proliferative effects of induced GQs provide us with an additional tool for rational drug design against diseases such as cancer. It is widely accepted that there is no magic bullet for several diseases including cancer. Thus,

novel approaches are required to provide alternatives in conjunction with current strategies to fight such diseases. The GQ structures are known to modulate myriad of cellular processes, such as mRNA splicing, editing, transcription termination, the formation of induced GQ structures can potentially modulate such processes in a targeted fashion if appropriate target sites are available. Future studies would include delivery of the oligonucleotides to enhance the efficacy of the oligonucleotides. The above concept opens up a new strategy for targeted modulation of endogenous gene expression. Here we demonstrate our capability to harness and emulate the power of a naturally occurring mechanism for rational control of gene expression.

The work in the dissertation collectively demonstrates the structural plasticity of a non-canonical RNA secondary structure and the adaptive flexibility in its function to modulate translation.

## CHAPTER SIX

### 6. APPENDIX I

#### 6.1 Alternative analyses of 40S ribosome binding to the IRES A

The 40S binding curves were calculated and plotted using two different equations. The plots were created in Origin. The  $Y_{\max}$  here determines the filter binding efficiency and the constructs that did not achieve the  $Y_{\max}$  of at least 0.5 are highlighted. These are the constructs, which are weak binders to the 40S ribosomal subunit. In this section, we report the values obtained by analyzing the curves using the both the equations and comparison of the values.

| Constructs    | Method 2   |                 | Method 1<br>(Kieft et al., 2001) |
|---------------|------------|-----------------|----------------------------------|
|               | $Y_{\max}$ | $K_d$           | $K_d$                            |
| IRES A        | 0.71       | $6.3 \pm 0.6$   | $29.0 \pm 5.7$                   |
| $\Delta$ D1   | 0.56       | $7.3 \pm 1.0$   | 108.40                           |
| $\Delta$ D3   | 0.64       | $8.4 \pm 1.1$   | 55.30                            |
| $\Delta$ D4   | 0.71       | $8.5 \pm 1.6$   | 33.60                            |
| M3Q ins       | 0.85       | $10.6 \pm 3.1$  | $18.0 \pm 4.9$                   |
| NRAS ins      | 0.71       | $25.8 \pm 3.6$  | $88 \pm 18$                      |
| $\Delta$ D123 | 0.30       | $50.8 \pm 6.6$  | n. d                             |
| IRES A + 30S  | 0.30       | n. d            | n. d                             |
| $\Delta$ D24  | 0.28       | $4.6 \pm 0.1$   | n. d                             |
| $\Delta$ D2   | 0.42       | $63.5 \pm 6.3$  | n. d                             |
| 4 mutant      | 0.45       | $24.6 \pm 10.1$ | n. d                             |

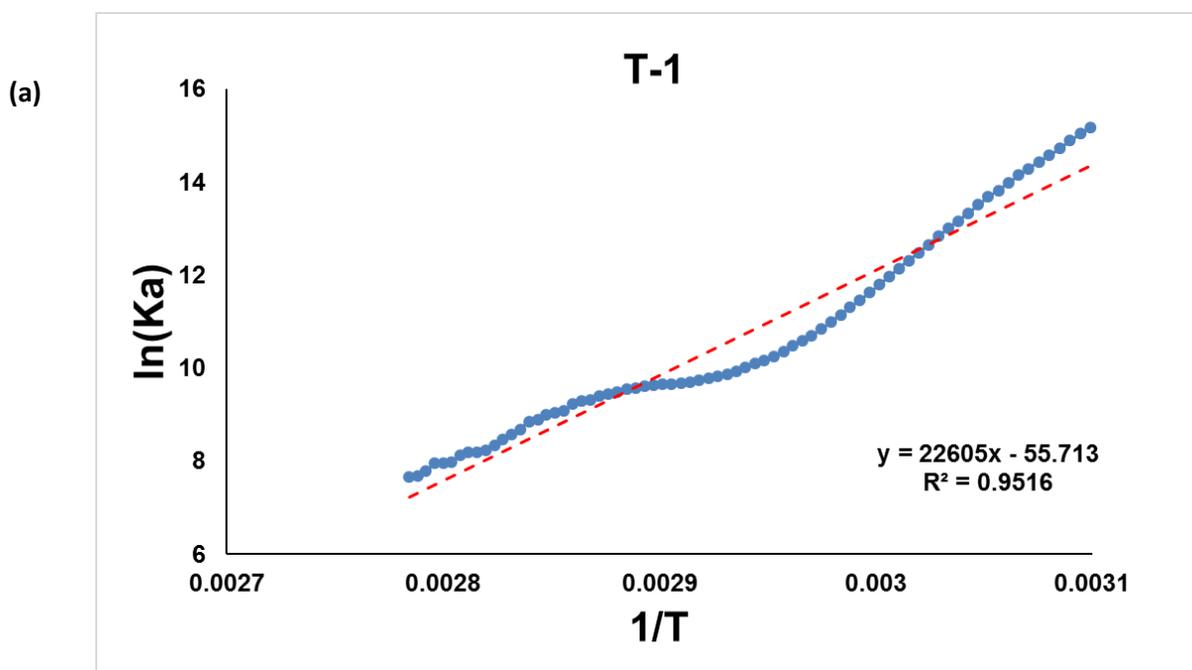
**Table 6.1:** Comparison of the binding constants derived by two different methods.

## CHAPTER SEVEN

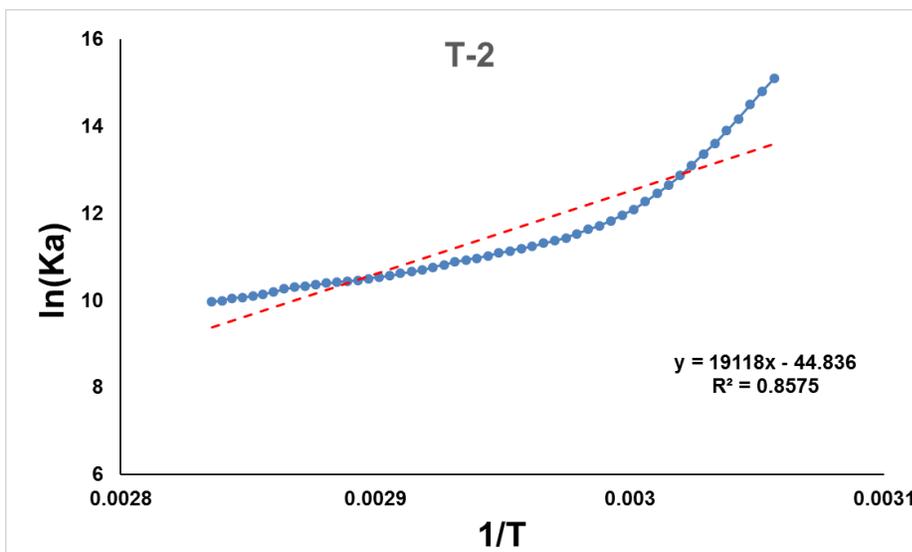
### 7. APPENDIX II

#### 7.1 Analyses of the melting curves of Induced GQ

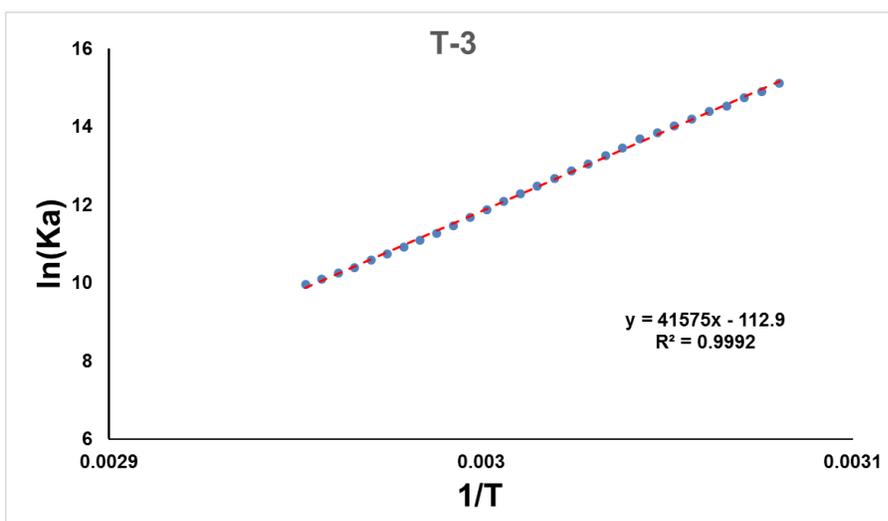
The temperature-dependent CD spectra of three independent trials were analyzed by two methods detailed in Section 4.2.4.1. As discussed before we observed that the formation of the induced GQ was driven by the enthalpy of the induced GQ formation. Analysis by the second method suggests the same. Interestingly the average  $\Delta G_{37}$  calculated by the 2<sup>nd</sup> method was lower than the 1<sup>st</sup> method and much comparable to what was observed by isothermal titration calorimetry.



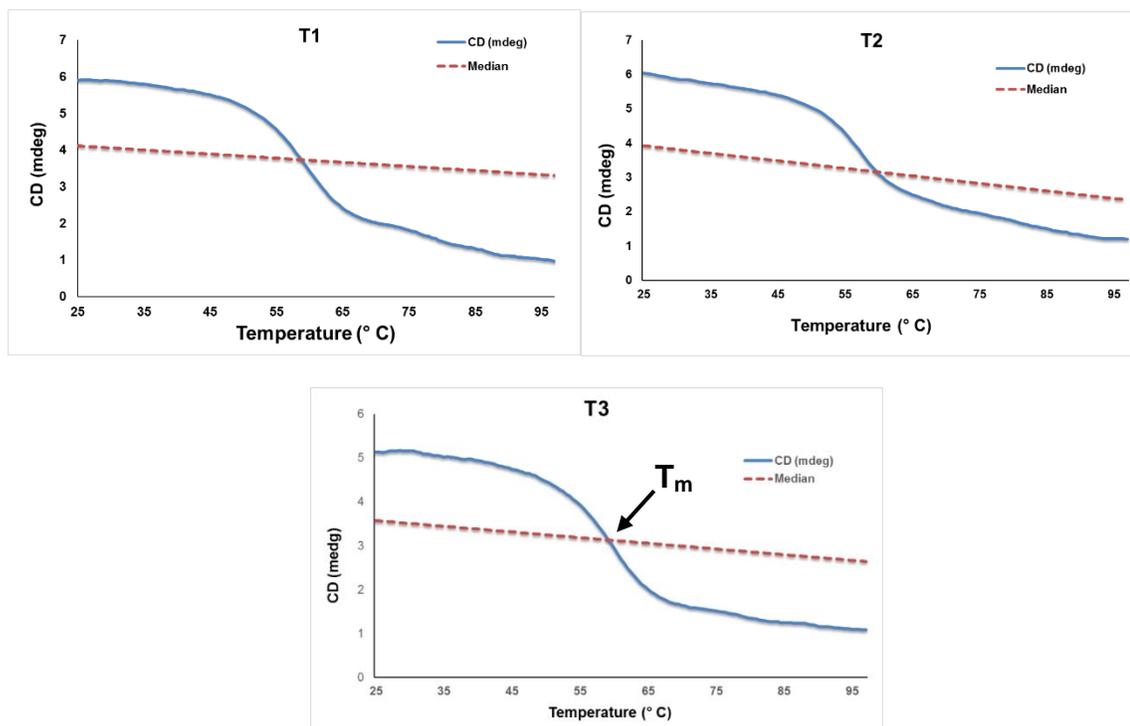
(b)



(c)



**Figure 7.1 The van't Hoff plots:** The van't Hoff plots of the three trials (a, b and) were constructed by plotting the calculated  $\ln(K_a)$  vs the  $1/T$ . The slope represents  $-\Delta H/R$  and the intercept is equal to  $\Delta S/R$ . The  $\Delta H$  is enthalpic and the  $\Delta S$  are the entropic change of the induced GQ formation.



**Figure 7.2: Determination of the melting temperature by method 2:** The temperature where the median of the top and bottom baselines intersect the temperature-dependent CD spectrum was determined to be melting temperature ( $T_m$ ).

| Sample         | $d\alpha/dT$ | $T_m(^{\circ}C)$ | $T_m(K)$      | $\Delta H_{VH} (kJ/mol)$ | $\Delta S (kJ/mol^{\circ}K)$ | $\Delta G_{37} (kJ/mol)$ |
|----------------|--------------|------------------|---------------|--------------------------|------------------------------|--------------------------|
| T1             | -0.05        | 58.68            | 331.83        | -264                     | -0.794                       | -17                      |
| T2             | -0.05        | 61.30            | 334.45        | -254                     | -0.760                       | -18                      |
| T3             | -0.06        | 58.92            | 332.07        | -333                     | -1.003                       | -22                      |
| <b>Average</b> | <b>-0.05</b> | <b>59.63</b>     | <b>332.78</b> | <b>-284</b>              | <b>-0.852</b>                | <b>-19</b>               |
| <b>SEM</b>     | 0.00         | 0.84             | 0.84          | 25                       | 0.076                        | 1                        |

**Table 7.1: Thermodynamic parameters calculated by Method 1:** Thermodynamic parameters calculated by the first method.

| Sample  | T <sub>m</sub> (°C) | ΔH <sub>VH</sub> (kJ/mol) | ΔS (kJ/mol*K) | ΔG <sub>37</sub> (kJ/mol) |
|---------|---------------------|---------------------------|---------------|---------------------------|
| T1      | 59.00               | -188                      | -0.463        | -44                       |
| T2      | 59.50               | -159                      | -0.373        | -43                       |
| T3      | 59.50               | -346                      | -0.939        | -55                       |
| Average | 59.33               | -231                      | -0.592        | -47                       |
| SEM     | 0.20                | 71                        | 0.215         | 4                         |

**Table 7.2 Thermodynamic parameters calculated by Method 2:** Thermodynamic parameters calculated by the second method. The T<sub>m</sub> was determined by the median intersecting the temperature-dependent CD spectra.

| Thermodynamic parameters  | Method 1 |        |        | Method 2 |        |        |
|---------------------------|----------|--------|--------|----------|--------|--------|
|                           | T1       | T2     | T3     | T1       | T2     | T3     |
| T <sub>m</sub> (°C)       | 58.68    | 61.30  | 58.92  | 59.00    | 59.50  | 59.50  |
| ΔH <sub>VH</sub> (KJ/mol) | -264     | -254   | -333   | -188     | -159   | -346   |
| ΔS (KJ/mol*K)             | -0.794   | -0.760 | -1.003 | -0.463   | -0.373 | -0.939 |
| ΔG <sub>37</sub> (KJ/mol) | -17      | -18    | -22    | -44      | -43    | -55    |

**Table 7.3: Comparison of the thermodynamic parameters:** The thermodynamic parameters calculated by two different methods. The calculated Gibbs free energies were lower by the 2<sup>nd</sup> method.

## REFERENCES

- Agarwala, P., Kumar, S., Pandey, S., and Maiti, S. (2015a). Human telomeric RNA G-quadruplex response to point mutation in the G-quartets. *J. Phys. Chem. B* 119(13), 4617-4627.
- Agarwala, P., Pandey, S., and Maiti, S. (2014). Role of G-quadruplex located at 5' end of mRNAs. *Biochim. Biophys. Acta* 12(10), 16.
- Agarwala, P., Pandey, S., and Maiti, S. (2015b). The tale of RNA G-quadruplex. *Org. Biomol. Chem.* 13(20), 5570-5585.
- Agarwala, P., Pandey, S., Mapa, K., and Maiti, S. (2013). The G-quadruplex augments translation in the 5' untranslated region of transforming growth factor beta2. *Biochemistry* 52(9), 1528-1538.
- Allam, H., and Ali, N. (2010). Initiation factor eIF2-independent mode of c-Src mRNA translation occurs via an internal ribosome entry site. *J. Biol. Chem.* 285(8), 5713-5725.
- Arcondeguy, T., Lacazette, E., Millevoi, S., Prats, H., and Touriol, C. (2013). VEGF-A mRNA processing, stability and translation: a paradigm for intricate regulation of gene expression at the post-transcriptional level. *Nucleic Acids Res.* 41(17), 7997-8010.
- Arora, A., Dutkiewicz, M., Scaria, V., Hariharan, M., Maiti, S., and Kurreck, J. (2008). Inhibition of translation in living eukaryotic cells by an RNA G-quadruplex motif. *RNA* 14(7), 1290-1296.
- Arora, A., and Maiti, S. (2008). Effect of loop orientation on quadruplex-TMPyP4 interaction. *J. Phys. Chem. B* 112(27), 8151-8159.
- Arora, A., and Maiti, S. (2009). Stability and Molecular Recognition of Quadruplexes with Different Loop Length in the Absence and Presence of Molecular Crowding Agents. *J. Phys. Chem. B* 113(25), 8784-8792..
- Arora, A., and Suess, B. (2011). An RNA G-quadruplex in the 3' UTR of the proto-oncogene PIM1 represses translation. *RNA Biol.* 8(5), 802-805.
- Balasubramanian, S., and Neidle, S. (2009). G-quadruplex nucleic acids as therapeutic targets. *Curr. Opin. Chem. Biol.* 13(3), 345-353.
- Balkwill, G.D., Derecka, K., Garner, T.P., Hodgman, C., Flint, A.P., and Searle, M.S. (2009). Repression of translation of human estrogen receptor alpha by G-quadruplex formation. *Biochemistry* 48(48), 11487-11495.
- Bang, I. (1910). Untersuchungen u"ber die Guanylsau"re. *Biochemische Zeitschrift* 26, 293-311.
- Barreau, C., Paillard, L., and Osborne, H.B. (2005). AU-rich elements and associated factors: are there unifying principles? *Nucleic Acids Res.* 33(22), 7138-7150.
- Bastide, A., Karaa, Z., Bornes, S., Hieblot, C., Lacazette, E., Prats, H., et al. (2008). An upstream open reading frame within an IRES controls expression of a specific VEGF-A isoform. *Nucleic Acids Res.* 36(7), 2434-2445.
- Beaudoin, J.D., and Perreault, J.P. (2010). 5'-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Res.* 38(20), 7022-7036.
- Beaudoin, J.D., and Perreault, J.P. (2013). Exploring mRNA 3'-UTR G-quadruplexes: evidence of roles in both alternative polyadenylation and mRNA shortening. *Nucleic Acids Res.* 41(11), 5898-5911.

- Bernstein, J.A., Khodursky, A.B., Lin, P.H., Lin-Chao, S., and Cohen, S.N. (2002). Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl. Acad. Sci. U.S.A* 99, 9697 - 9702.
- Bevilacqua, P.C., and Blose, J.M. (2008). "Structures, kinetics, thermodynamics, and biological functions of RNA hairpins," in *Ann. Rev. Phys. Chem.*, 79-103.
- Bhattacharyya, A., and Blackburn, E.H. (1997). A functional telomerase RNA swap in vivo reveals the importance of nontemplate RNA domains. *Proc. Natl. Acad. Sci. U.S.A* 94(7), 2823-2827.
- Bhattacharyya, D., Diamond, P., and Basu, S. (2015). An Independently folding RNA G-quadruplex domain directly recruits the 40S ribosomal subunit. *Biochemistry* 54(10), 1879-1885.
- Bhattacharyya, D., Nguyen, K., and Basu, S. (2014). Rationally induced RNA:DNA G-quadruplex structures elicit an anticancer effect by inhibiting endogenous eIF-4E expression. *Biochemistry* 53(33), 5461-5470.
- Biffi, G., Di Antonio, M., Tannahill, D., and Balasubramanian, S. (2014). Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat. Chem.* 6(1), 75-80.
- Biffi, G., Tannahill, D., McCafferty, J., and Balasubramanian, S. (2013). Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.* 5(3), 182-186.
- Blackwell, E., Zhang, X., and Ceman, S. (2010). Arginines of the RGG box regulate FMRP association with polyribosomes and mRNA. *Hum. Mol. Genet.* 19(7), 1314-1323.
- Bonnal, S., Schaeffer, C., Creancier, L., Clamens, S., Moine, H., Prats, A.C., et al. (2003). A single internal ribosome entry site containing a G quartet RNA structure drives fibroblast growth factor 2 gene expression at four alternative translation initiation codons. *J. Biol. Chem.* 278(41), 39330-39336.
- Booy, E.P., Howard, R., Marushchak, O., Ariyo, E.O., Meier, M., Novakowski, S.K., et al. (2014). The RNA helicase RHAU (DHX36) suppresses expression of the transcription factor PITX1. *Nucleic Acids Res.* 42(5), 3346-3361.
- Booy, E.P., Meier, M., Okun, N., Novakowski, S.K., Xiong, S., Stetefeld, J., et al. (2012). The RNA helicase RHAU (DHX36) unwinds a G4-quadruplex in human telomerase RNA and promotes the formation of the P1 helix template boundary. *Nucleic Acids Res.* 40(9), 4110-4124.
- Bornes, S., Prado-Lourenco, L., Bastide, A., Zanibellato, C., Iacovoni, J.S., Lacazette, E., et al. (2007). Translational induction of VEGF internal ribosome entry site elements during the early response to ischemic stress. *Circ. Res.* 100(3), 305-308.
- Bugaut, A., and Balasubramanian, S. (2012). 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.* 40(11), 4727-4741.
- Bugaut, A., Rodriguez, R., Kumari, S., Hsu, S.T., and Balasubramanian, S. (2010). Small molecule-mediated inhibition of translation by targeting a native RNA G-quadruplex. *Org. Biomol. Chem.* 8(12), 2771-2776.
- Burge, S., Parkinson, G.N., Hazel, P., Todd, A.K., and Neidle, S. (2006). Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* 34(19), 5402-5415.
- Carroll, M., and Borden, K.L. (2013). The oncogene eIF4E: using biochemical insights to target cancer. *J. Interferon Cytokine Res.* 33(5), 227-238.

- Castets, M., Schaeffer, C., Bechara, E., Schenck, A., Khandjian, E.W., Luche, S., et al. (2005). FMRP interferes with the Rac1 pathway and controls actin cytoskeleton dynamics in murine fibroblasts. *Hum. Mol. Genet.* 14(6), 835-844.
- Cevec, M., and Plavec, J. (2005). Role of loop residues and cations on the formation and stability of dimeric DNA G-quadruplexes. *Biochemistry* 44(46), 15238-15246.
- Chakraborty, P., and Grosse, F. (2011). Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA Repair (Amst.)* 10(6), 654-665.
- Chan, J.H., Lim, S., and Wong, W.S. (2006). Antisense oligonucleotides: from design to therapeutic application. *Clin. Exp. Pharmacol. Physiol.* 33(5-6), 533-540.
- Cheong, C., and Moore, P.B. (1992). Solution structure of an unusually stable RNA tetraplex containing G- and U-quartet structures. *Biochemistry* 31(36), 8406-8414.
- Christiansen, J., Kofod, M., and Nielsen, F.C. (1994). A guanosine quadruplex and two stable hairpins flank a major cleavage site in insulin-like growth factor II mRNA. *Nucleic Acids Res.* 22(25), 5709-5716.
- Cogoi, S., and Xodo, L.E. (2006). G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res.* 34(9), 2536-2549.
- Collie, G., Reszka, A.P., Haider, S.M., Gabelica, V., Parkinson, G.N., and Neidle, S. (2009). Selectivity in small molecule binding to human telomeric RNA and DNA quadruplexes. *Chem. Commun.* 28(48), 7482-7484.
- Collie, G.W., Haider, S.M., Neidle, S., and Parkinson, G.N. (2010). A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex. *Nucleic Acids Res.* 38(16), 5569-5580.
- Collie, G.W., and Parkinson, G.N. (2011). The application of DNA and RNA G-quadruplexes to therapeutic medicines. *Chem. Soc. Rev.* 40(12), 5867-5892.
- Collie, G.W., Sparapani, S., Parkinson, G.N., and Neidle, S. (2011). Structural basis of telomeric RNA quadruplex--acridine ligand recognition. *J. Am. Chem. Soc.* 133(8), 2721-2728.
- Creacy, S.D., Routh, E.D., Iwamoto, F., Nagamine, Y., Akman, S.A., and Vaughn, J.P. (2008). G4 Resolvase 1 Binds Both DNA and RNA Tetramolecular Quadruplex with High Affinity and Is the Major Source of Tetramolecular Quadruplex G4-DNA and G4-RNA Resolving Activity in HeLa Cell Lysates. *J. Biol. Chem.* 283(50), 34626-34634.
- Curatola, A.M., Nadal, M.S., and Schneider, R.J. (1995). Rapid degradation of AU-rich element (ARE) mRNAs is activated by ribosome transit and blocked by secondary structure at any position 5'to the ARE. *Mol. Cell. Biol.* 15(11), 6331-6340.
- Dai, J., Chen, D., Jones, R.A., Hurley, L.H., and Yang, D. (2006). NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region. *Nucleic Acids Res.* 34(18), 5133-5144.
- Darnell, J.C., Jensen, K.B., Jin, P., Brown, V., Warren, S.T., and Darnell, R.B. (2001). Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell* 107(4), 489-499.
- De Armond, R., Wood, S., Sun, D., Hurley, L.H., and Ebbinghaus, S.W. (2005). Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1alpha promoter. *Biochemistry* 44(49), 16341-16350.
- De Cian, A., Lacroix, L., Douarre, C., Temime-Smaali, N., Trentesaux, C., Riou, J.F., et al. (2008). Targeting telomeres and telomerase. *Biochimie* 90(1), 131-155.

- de Silanes, I.L., d'Alcontres, M.S., and Blasco, M.A. (2010). TERRA transcripts are bound by a complex array of RNA-binding proteins. *Nat. Commun.* 1, 33.
- Decorsiere, A., Cayrel, A., Vagner, S., and Millevoi, S. (2011). Essential role for the interaction between hnRNP H/F and a G quadruplex in maintaining p53 pre-mRNA 3'-end processing and function during DNA damage. *Genes Dev.* 25(3), 220-225.
- Deng, Z., Norseen, J., Wiedmer, A., Riethman, H., and Lieberman, P.M. (2009). TERRA RNA Binding to TRF2 Facilitates Heterochromatin Formation and ORC Recruitment at Telomeres. *Mol. Cell* 35(4), 403-413.
- Derecka, K., Balkwill, G.D., Garner, T.P., Hodgman, C., Flint, A.P., and Searle, M.S. (2010). Occurrence of a quadruplex motif in a unique insert within exon C of the bovine estrogen receptor alpha gene (ESR1). *Biochemistry* 49(35), 7625-7633.
- Dexheimer, T.S., Sun, D., and Hurley, L.H. (2006). Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter. *J. Am. Chem. Soc.* 128(16), 5404-5415.
- Di Antonio, M., Biffi, G., Mariani, A., Raiber, E.-A., Rodriguez, R., and Balasubramanian, S. (2012). Selective RNA Versus DNA G-Quadruplex Targeting by In Situ Click Chemistry. *Angew. Chem. Int. Ed. Engl.* 51(44), 11073-11078.
- Dias, N., and Stein, C.A. (2002). Antisense oligonucleotides: basic concepts and mechanisms. *Mol. Cancer. Ther.* 1(5), 347-355.
- Didiot, M.C., Tian, Z., Schaeffer, C., Subramanian, M., Mandel, J.L., and Moine, H. (2008). The G-quartet containing FMRP binding site in FMR1 mRNA is a potent exonic splicing enhancer. *Nucleic Acids Res.* 36(15), 4902-4912.
- Eddy, J., and Maizels, N. (2008). Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes. *Nucleic Acids Res.* 36(4), 1321-1333.
- Endoh, T., Kawasaki, Y., and Sugimoto, N. (2013a). Stability of RNA quadruplex in open reading frame determines proteolysis of human estrogen receptor alpha. *Nucleic Acids Res.* 41(12), 6222-6231.
- Endoh, T., Kawasaki, Y., and Sugimoto, N. (2013b). Suppression of Gene Expression by G-Quadruplexes in Open Reading Frames Depends on G-Quadruplex Stability. *Angew. Chem. Int. Ed. Engl.* 52(21), 5522-5526.
- Endoh, T., Kawasaki, Y., and Sugimoto, N. (2013c). Translational halt during elongation caused by G-quadruplex formed by mRNA. *Methods* 64(1), 73-78.
- Evans, T.L., Blice-Baum, A.C., and Mihailescu, M.-R. (2012). Analysis of the Fragile X mental retardation protein isoforms 1, 2 and 3 interactions with the G-quadruplex forming semaphorin 3F mRNA. *Mol. Biosyst.* 8(2), 642-649.
- Fernando, H., Reszka, A.P., Huppert, J., Ladame, S., Rankin, S., Venkitaraman, A.R., et al. (2006). A conserved quadruplex motif located in a transcription activation site of the human c-kit oncogene. *Biochemistry* 45(25), 7854-7860.
- Ferrara, N., and Davis-Smyth, T. (1997). The Biology of Vascular Endothelial Growth Factor. *Endocr. Rev.* 18(1), 4-25.
- Fisette, J.F., Montagna, D.R., Mihailescu, M.R., and Wolfe, M.S. (2012). A G-rich element forms a G-quadruplex and regulates BACE1 mRNA alternative splicing. *J. Neurochem.* 121(5), 763-773.

- Flanagan, W.M., Kothavale, A., and Wagner, R.W. (1996). Effects of oligonucleotide length, mismatches and mRNA levels on C-5 propyne-modified antisense potency. *Nucleic Acids Res.* 24(15), 2936-2941.
- Fry, M. (2007). Tetraplex DNA and its interacting proteins. *Front. Biosci.* 12, 4336-4351.
- Gellert, M., Lipsett, M.N., and Davies, D.R. (1962). Helix formation by guanylic acid. *Proc. Natl. Acad. Sci. U. S.A.* 48, 2013-2018.
- Gilbert, D.E., and Feigon, J. (1999). Multistranded DNA structures. *Curr Opin Struct Biol* 9(3), 305-314.
- Gomez, D., Guedin, A., Mergny, J.L., Salles, B., Riou, J.F., Teulade-Fichou, M.P., et al. (2010). A G-quadruplex structure within the 5'-UTR of TRF2 mRNA represses translation in human cells. *Nucleic Acids Res.* 38(20), 7187-7198.
- Gomez, D., Lemarteleur, T., Lacroix, L., Mailliet, P., Mergny, J.L., and Riou, J.F. (2004). Telomerase downregulation induced by the G-quadruplex ligand 12459 in A549 cells is mediated by hTERT RNA alternative splicing. *Nucleic Acids Res.* 32(1), 371-379.
- Gray, N.K., and Hentze, M.W. (1994). Regulation of protein synthesis by mRNA structure. *Mol. Biol. Rep.* 19(3), 195-200.
- Guedin, A., De Cian, A., Gros, J., Lacroix, L., and Mergny, J.L. (2008). Sequence effects in single-base loops for quadruplexes. *Biochimie* 90(5), 686-696.
- Guedin, A., Gros, J., Alberti, P., and Mergny, J.L. (2010). How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Res.* 38(21), 7858-7868.
- Hagihara, M., Yamauchi, L., Seo, A., Yoneda, K., Senda, M., and Nakatani, K. (2010). Antisense-induced guanine quadruplexes inhibit reverse transcription by HIV-1 reverse transcriptase. *J. Am. Chem. Soc.* 132(32), 11171-11178.
- Halder, K., Largy, E., Benzler, M., Teulade-Fichou, M.P., and Hartig, J.S. (2011). Efficient suppression of gene expression by targeting 5'-UTR-based RNA quadruplexes with bisquinolinium compounds. *Chembiochem* 12(11), 1663-1668.
- Halder, K., Wieland, M., and Hartig, J.S. (2009). Predictable suppression of gene expression by 5'-UTR-based RNA quadruplexes. *Nucleic Acids Res.* 37(20), 6811-6817.
- Hazel, P., Huppert, J., Balasubramanian, S., and Neidle, S. (2004). Loop-length-dependent folding of G-quadruplexes. *J. Am. Chem. Soc.* 126(50), 16405-16415.
- Hazel, P., Parkinson, G.N., and Neidle, S. (2006). Predictive modelling of topology and loop variations in dimeric DNA quadruplex structures. *Nucleic Acids Res.* 34, 2117-2127.
- Hellen, C.U., and Sarnow, P. (2001). Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes Dev.* 15(13), 1593-1612.
- Hellen, C.U.T. (2009). IRES-induced conformational changes in the ribosome and the mechanism of translation initiation by internal ribosomal entry. *Biochim. Biophys. Acta* 1789(9-10), 558-570.
- Hsieh, A.C., and Ruggero, D. (2010). Targeting Eukaryotic Translation Initiation Factor 4E (eIF4E) in Cancer. *Clin. Cancer Res.* 16(20), 4914-4920.
- Huang, W., Smaldino, P.J., Zhang, Q., Miller, L.D., Cao, P., Stadelman, K., et al. (2012). Yin Yang 1 contains G-quadruplex structures in its promoter and 5'-UTR and its expression is modulated by G4 resolvase 1. *Nucleic Acids Res.* 40(3), 1033-1049.
- Hud, N.V., Smith, F.W., Anet, F.A., and Feigon, J. (1996). The selectivity for K<sup>+</sup> versus Na<sup>+</sup> in DNA quadruplexes is dominated by relative free energies of hydration: a thermodynamic analysis by <sup>1</sup>H NMR. *Biochemistry* 35(48), 15383-15390.

- Huez, I., Creancier, L., Audigier, S., Gensac, M.C., Prats, A.C., and Prats, H. (1998). Two independent internal ribosome entry sites are involved in translation initiation of vascular endothelial growth factor mRNA. *Mol. Cell. Biol.* 18(11), 6178-6190.
- Huppert, J.L. (2008). Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes. *Chem. Soc. Rev.* 37(7), 1375-1384.
- Huppert, J.L. (2010). Structure, location and interactions of G-quadruplexes. *FEBS J* 277(17), 3452-3458.
- Huppert, J.L., and Balasubramanian, S. (2005). Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* 33(9), 2908-2916.
- Huppert, J.L., and Balasubramanian, S. (2007). G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* 35(2), 406-413.
- Huppert, J.L., Bugaut, A., Kumari, S., and Balasubramanian, S. (2008). G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res.* 36(19), 6260-6268.
- Ito, K., Go, S., Komiyama, M., and Xu, Y. (2011). Inhibition of translation by small RNA-stabilized mRNA structures in human cells. *J. Am. Chem. Soc.* 133(47), 19153-19159.
- Jackson, R.J., Hellen, C.U., and Pestova, T.V. (2010a). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* 11(2), 113-127.
- Jackson, R.J., Hellen, C.U.T., and Pestova, T.V. (2010b). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* 11(2), 113-127.
- Jacobson, B.A., Thumma, S.C., Jay-Dixon, J., Patel, M.R., Dubear Kroening, K., Kratzke, M.G., et al. (2013). Targeting eukaryotic translation in mesothelioma cells with an eIF4E-specific antisense oligonucleotide. *PLoS One* 8(11), e81669.
- Jambhekar, A., and Derisi, J.L. (2007). Cis-acting determinants of asymmetric, cytoplasmic RNA transport. *RNA* 13(5), 625-642.
- Jang, S.K., Krausslich, H.G., Nicklin, M.J.H., Duke, G.M., Palmenberg, A.C., and Wimmer, E. (1988). A Segment of the 5' Nontranslated Region of Encephalomyocarditis Virus-Rna Directs Internal Entry of Ribosomes during Invitro Translation. *J. Virol.* 62(8), 2636-2643.
- Jayaraj, G.G., Pandey, S., Scaria, V., and Maiti, S. (2012). Potential G-quadruplexes in the human long non-coding transcriptome. *RNA Biol.* 9(1), 81-86.
- Ji, X., Sun, H., Zhou, H., Xiang, J., Tang, Y., and Zhao, C. (2011). Research progress of RNA quadruplex. *Nucleic Acid Ther.* 21(3), 185-200.
- Joachimi, A., Benz, A., and Hartig, J.S. (2009). A comparison of DNA and RNA quadruplex structures and stabilities. *Bioorg. Med. Chem.* 17(19), 6811-6815.
- Kang, C., Zhang, X., Ratliff, R., Moyzis, R., and Rich, A. (1992). Crystal structure of four-stranded Oxytricha telomeric DNA. *Nature* 356(6365), 126-131.
- Keniry, M.A. (2001). Quadruplex structures in nucleic acids. *Biopolymers* 56(3), 123-146.
- Khateb, S., Weisman-Shomer, P., Hershco-Shani, I., Ludwig, A.L., and Fry, M. (2007). The tetraplex (CGG)<sub>n</sub> destabilizing proteins hnRNP A2 and CBF-A enhance the in vivo translation of fragile X premutation mRNA. *Nucleic Acids Res.* 35(17), 5775-5788.
- Kieft, J.S. (2008). Viral IRES RNA structures and ribosome interactions. *Trends Biochem. Sci.* 33(6), 274-283.
- Kieft, J.S., Zhou, K.H., Jubin, R., and Doudna, J.A. (2001). Mechanism of ribosome recruitment by hepatitis C IRES RNA. *RNA* 7(2), 194-206.
- Kim, J., Cheong, C., and Moore, P.B. (1991). Tetramerization of an RNA oligonucleotide containing a GGGG sequence. *Nature* 351(6324), 331-332.

- King, H.A., Cobbold, L.C., and Willis, A.E. (2010). The role of IRES trans-acting factors in regulating translation initiation. *Biochem. Soc. Trans.* 38, 1581-1586.
- Komar, A.A., and Hatzoglou, M. (2011). Cellular IRES-mediated translation The war of ITAFs in pathophysiological states. *Cell Cycle* 10(2), 229-240.
- Komar, A.A., Mazumder, B., and Merrick, W.C. (2012). A new framework for understanding IRES-mediated translation. *Gene* 502(2), 75-86.
- Kostadinov, R., Malhotra, N., Viotti, M., Shine, R., D'Antonio, L., and Bagga, P. (2006). GRSDb: a database of quadruplex forming G-rich sequences in alternatively processed mammalian pre-mRNA sequences. *Nucleic Acids Res.* 34(Database issue), D119-124.
- Kozak, M. (1991). Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J Biol. Chem.* 266(30), 19867-19870.
- Kumari, S., Bugaut, A., and Balasubramanian, S. (2008). Position and stability are determining factors for translation repression by an RNA G-quadruplex-forming sequence within the 5' UTR of the NRAS proto-oncogene. *Biochemistry* 47(48), 12664-12669.
- Kumari, S., Bugaut, A., Huppert, J.L., and Balasubramanian, S. (2007a). An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol.* 3(4), 218-221.
- Kumari, S., Bugaut, A., Huppert, J.L., and Balasubramanian, S. (2007b). An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol.* 3(4), 218-221.
- Lammich, S., Kamp, F., Wagner, J., Nuscher, B., Zilow, S., Ludwig, A.-K., et al. (2011). Translational Repression of the Disintegrin and Metalloprotease ADAM10 by a Stable G-quadruplex Secondary Structure in Its 5'-Untranslated Region. *J. Biol. Chem.* 286(52), 45063-45072.
- Lang, B.E., and Schwarz, F.P. (2007). Thermodynamic dependence of DNA/DNA and DNA/RNA hybridization reactions on temperature and ionic strength. *Biophys. Chem.* 131(1-3), 96-104.
- Lattmann, S., Giri, B., Vaughn, J.P., Akman, S.A., and Nagamine, Y. (2010). Role of the amino terminal RHAU-specific motif in the recognition and resolution of guanine quadruplex-RNA by the DEAH-box RNA helicase RHAU. *Nucleic Acids Res.* 38(18), 6219-6233.
- Lattmann, S., Stadler, M.B., Vaughn, J.P., Akman, S.A., and Nagamine, Y. (2011). The DEAH-box RNA helicase RHAU binds an intramolecular RNA G-quadruplex in TERC and associates with telomerase holoenzyme. *Nucleic Acids Res.* 39(21), 9390-9404.
- Lee, J.Y., and Kim, D.S. (2009). Dramatic effect of single-base mutation on the conformational dynamics of human telomeric G-quadruplex. *Nucleic Acids Res.* 37(11), 3625-3634.
- Lewis, J.D., and Izaurflde, E. (1997). The Role of the Cap Structure in RNA Processing and Nuclear Export. *Eur. J. Biochem.* 247(2), 461-469.
- Liu, H., Kanagawa, M., Matsugami, A., Tanaka, Y., Katahira, M., and Uesugi, S. (2000). NMR study of a novel RNA quadruplex structure. *Nucleic Acids Symp Ser* 44(44), 65-66.
- Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods* 25(4), 402-408.
- Locker, N., Chamond, N., and Sargueil, B. (2011). A conserved structure within the HIV gag open reading frame that controls translation initiation directly recruits the 40S subunit and eIF3. *Nucleic Acids Res.* 39(6), 2367-2377.

- Lu, R., Wang, H.P., Liang, Z., Ku, L., O'Donnell, W.T., Li, W., et al. (2004). The fragile X protein controls microtubule-associated protein 1B translation and microtubule stability in brain neuron development. *Proc. Nat. Acad. Sci U.S.A.* 101(42), 15201-15206.
- MacRae, I.J., Zhou, K., and Doudna, J.A. (2007). Structural determinants of RNA recognition and cleavage by Dicer. *Nat. Struct. Mol. Biol.* 14(10), 934-940.
- Mamane, Y., Petroulakis, E., Rong, L., Yoshida, K., Ler, L.W., and Sonenberg, N. (2004). eIF4E-- from translation to transformation. *Oncogene* 23(18), 3172-3179.
- Mani, P., Yadav, V.K., Das, S.K., and Chowdhury, S. (2009). Genome-wide analyses of recombination prone regions predict role of DNA structural motif in recombination. *PLoS One* 4(2), e4399.
- Marcel, V., Tran, P.L., Sagne, C., Martel-Planche, G., Vaslin, L., Teulade-Fichou, M.P., et al. (2011). G-quadruplex structures in TP53 intron 3: role in alternative splicing and in production of p53 mRNA isoforms. *Carcinogenesis* 32(3), 271-278.
- Marky, L.A., and Breslauer, K.J. (1987). Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers* 26(9), 1601-1620.
- Martadinata, H., and Phan, A.T. (2013). Structure of human telomeric RNA (TERRA): stacking of two G-quadruplex blocks in K(+) solution. *Biochemistry* 52(13), 2176-2183.
- Martineau, Y., Le Bec, C., Monbrun, L., Allo, V., Chiu, I.M., Danos, O., et al. (2004). Internal ribosome entry site structural motifs conserved among mammalian fibroblast growth factor 1 alternatively spliced mRNAs. *Mol. Cell Biol.* 24(17), 7622-7635.
- Martinez-Salas, E., Pineiro, D., and Fernandez, N. (2012). Alternative Mechanisms to Initiate Translation in Eukaryotic mRNAs. *Comp. Funct. Genomics*.
- Martinez-Salas, E., Ramos, R., Lafuente, E., and Lopez de Quinto, S. (2001). Functional interactions in internal translation initiation directed by viral and cellular IRES elements. *J. Gen. Virol.* 82(Pt 5), 973-984.
- Matsugami, A., Okuizumi, T., Uesugi, S., and Katahira, M. (2003). Intramolecular higher order packing of parallel quadruplexes comprising a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad of GGA triplet repeat DNA. *J. Biol. Chem.* 278(30), 28147-28153.
- Maxam, A.M., and Gilbert, W. (1977). A new method for sequencing DNA. *Proc. Nat. Acad. Sci U.S.A.* 74(2), 560-564.
- McManus, C.J., and Graveley, B.R. (2011). RNA structure and the mechanisms of alternative splicing. *Curr. Opin. Genet. Dev.* 21(4), 373-379.
- Meier, M., Patel, T.R., Booy, E.P., Marushchak, O., Okun, N., Deo, S., et al. (2013). Binding of G-quadruplexes to the N-terminal Recognition Domain of the RNA Helicase Associated with AU-rich Element (RHAU). *J. Biol. Chem.* 288(49), 35014-35027.
- Melko, M., and Bardoni, B. (2010). The role of G-quadruplex in RNA metabolism: Involvement of FMRP and FMR2P. *Biochimie* 92(8), 919-926.
- Menon, L., and Mihailescu, M.R. (2007). Interactions of the G quartet forming semaphorin 3F RNA with the RGG box domain of the fragile X protein family. *Nucleic Acids Res.* 35(16), 5379-5392.
- Mergny, J.L., and Lacroix, L. (2003). Analysis of thermal melting curves. *Oligonucleotides* 13(6), 515-537.
- Mignone, F., Gissi, C., Liuni, S., and Pesole, G. (2002). Untranslated regions of mRNAs. *Genome Biol.* 3(3), REVIEWS0004.

- Miller, D.L., Dibbens, J.A., Damert, A., Risau, W., Vadas, M.A., and Goodall, G.J. (1998). The vascular endothelial growth factor mRNA contains an internal ribosome entry site. *FEBS Lett.* 434(3), 417-420.
- Millevoi, S., Moine, H., and Vagner, S. (2012). G-quadruplexes in RNA biology. *Wiley Interdiscip. Rev. RNA* 3(4), 495-507.
- Min, H., Chan, R.C., and Black, D.L. (1995). The generally expressed hnRNP F is involved in a neural-specific pre-mRNA splicing event. *Genes Dev.* 9(21), 2659-2671.
- Mirihana Arachchilage, G., Dassanayake, A.C., and Basu, S. (2015). A potassium ion-dependent RNA structural switch regulates human pre-miRNA 92b maturation. *Chem. Biol.* 22(2), 262-272.
- Mirihana Arachchilage, G., Morris, M.J., and Basu, S. (2014). A library screening approach identifies naturally occurring RNA sequences for a G-quadruplex binding ligand. *Chem. Commun.* 50(10), 1250-1252.
- Monchaud, D., and Teulade-Fichou, M.P. (2008). A hitchhiker's guide to G-quadruplex ligands. *Org. Biomol. Chem.* 6(4), 627-636.
- Monde, R.A., Schuster, G., and Stern, D.B. (2000). Processing and degradation of chloroplast mRNA. *Biochimie* 82(6-7), 573-582.
- Morris, M.J. (2012). *Translational Regulation of mRNA by G-Quadruplex Structures*. Kent State University.
- Morris, M.J., and Basu, S. (2009). An unusually stable G-quadruplex within the 5'-UTR of the MT3 matrix metalloproteinase mRNA represses translation in eukaryotic cells. *Biochemistry* 48(23), 5313-5319.
- Morris, M.J., Negishi, Y., Pazsint, C., Schonhoft, J.D., and Basu, S. (2010). An RNA G-quadruplex is essential for cap-independent translation initiation in human VEGF IRES. *J. Am. Chem. Soc.* 132(50), 17831-17839.
- Morris, M.J., Wingate, K.L., Silwal, J., Leeper, T.C., and Basu, S. (2012). The porphyrin TmPyP4 unfolds the extremely stable G-quadruplex in MT3-MMP mRNA and alleviates its repressive effect to enhance translation in eukaryotic cells. *Nucleic Acids Res.* 40(9), 4137-4145.
- Mullen, M.A., Assmann, S.M., and Bevilacqua, P.C. (2012). Toward a digital gene response: RNA G-quadruplexes with fewer quartets fold with higher cooperativity. *J. Am. Chem. Soc.* 134(2), 812-815.
- Neidle, S. (2009). The structures of quadruplex nucleic acids and their drug complexes. *Curr. Opin. Struct. Biol.* 19(3), 239-250.
- Neidle, S., and Balasubramanian, S. (2006). *Quadruplex Nucleic Acids*. Cambridge, UK: RSC Biomolecular Sciences.
- Neidle, S., and Read, M.A. (2001). G-quadruplexes as therapeutic targets. *Biopolymers* 56(3), 195-208.
- Nielsen, F.C., and Christiansen, J. (1995). Posttranscriptional regulation of insulin-like growth factor II mRNA. *Scand. J. Clin. Lab Invest. Suppl.* 220, 37-46.
- Nielsen, M.C., and Ulven, T. (2010). Macrocyclic G-quadruplex ligands. *Curr. Med. Chem.* 17(29), 3438-3448.
- Ohnmacht, S.A., and Neidle, S. (2014). Small-molecule quadruplex-targeted drug discovery. *Bioorg. Med. Chem. Lett.* 24(12), 2602-2612.

- Olsen, C.M., and Marky, L.A. (2009). Energetic and hydration contributions of the removal of methyl groups from thymine to form uracil in G-quadruplexes. *J. Phys. Chem. B* 113(1), 9-11.
- Pandey, S., Agarwala, P., Jayaraj, G.G., Gargallo, R., and Maiti, S. (2015). The RNA Stem–Loop to G-Quadruplex Equilibrium Controls Mature MicroRNA Production inside the Cell. *Biochemistry* 54(48), 7067-7078.
- Pandey, S., Agarwala, P., and Maiti, S. (2013). Effect of loops and G-quartets on the stability of RNA G-quadruplexes. *J. Phys. Chem. B* 117(23), 6896-6905.
- Parkinson, G.N., Lee, M.P., and Neidle, S. (2002). Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* 417(6891), 876-880.
- Patel, D.J., Phan, A.T., and Kuryavyi, V. (2007). Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res.* 35(22), 7429-7455.
- Pelletier, J., and Sonenberg, N. (1988). Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* 334(6180), 320-325.
- Perard, J., Drouet, E., and Baudin, F. (2010). The role of IRES in translation initiation. *Virologie* 14(4), 241-253.
- Pestova, T.V., Kolupaeva, V.G., Lomakin, I.B., Pilipenko, E.V., Shatsky, I.N., Agol, V.I., et al. (2001). Molecular mechanisms of translation initiation in eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 98(13), 7029-7036.
- Phan, A.T., Kuryavyi, V., and Patel, D.J. (2006). DNA architecture: from G to Z. *Curr. Opin. Struct. Biol.* 16(3), 288-298.
- Plank, T.D., and Kieft, J.S. (2012). The structures of nonprotein-coding RNAs that drive internal ribosome entry site function. *Wiley Interdiscip. Rev. RNA* 3(2), 195-212.
- Rachwal, P.A., Brown, T., and Fox, K.R. (2007a). Effect of G-tract length on the topology and stability of intramolecular DNA quadruplexes. *Biochemistry* 46(11), 3036-3044.
- Rachwal, P.A., Brown, T., and Fox, K.R. (2007b). Sequence effects of single base loops in intramolecular quadruplex DNA. *FEBS Lett.* 581(8), 1657-1660.
- Rachwal, P.A., Findlow, I.S., Werner, J.M., Brown, T., and Fox, K.R. (2007c). Intramolecular DNA quadruplexes with different arrangements of short and long loops. *Nucleic Acids Res.* 35(12), 4214-4222.
- Ribeiro, M.M., Teixeira, G.S., Martins, L., Marques, M.R., de Souza, A.P., and Line, S.R. (2015). G-quadruplex formation enhances splicing efficiency of PAX9 intron 1. *Hum. Genet.* 134(1), 37-44.
- Rijnbrand, R., van der Straaten, T., van Rijn, P.A., Spaan, W.J., and Bredenbeek, P.J. (1997). Internal entry of ribosomes is directed by the 5' noncoding region of classical swine fever virus and is dependent on the presence of an RNA pseudoknot upstream of the initiation codon. *J. Virol.* 71(1), 451-457.
- Risitano, A., and Fox, K.R. (2004). Influence of loop size on the stability of intramolecular DNA quadruplexes. *Nucleic Acids Res.* 32(8), 2598-2606.
- Rodriguez, R., Miller, K.M., Forment, J.V., Bradshaw, C.R., Nikan, M., Britton, S., et al. (2012). Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nat. Chem. Biol.* 8(3), 301-310.

- Rogers, G.W., Richter, N.J., and Merrick, W.C. (1999). Biochemical and kinetic characterization of the RNA helicase activity of eukaryotic initiation factor 4A. *J. Biol. Chem.* 274(18), 12236-12244.
- Sacca, B., Lacroix, L., and Mergny, J.L. (2005). The effect of chemical modifications on the thermal stability of different G-quadruplex-forming oligonucleotides. *Nucleic Acids Res.* 33(4), 1182-1192.
- Satyanarayana, M., Kim, Y.A., Rzuczek, S.G., Pilch, D.S., Liu, A.A., Liu, L.F., et al. (2010). Macrocylic hexaoxazoles: Influence of aminoalkyl substituents on RNA and DNA G-quadruplex stabilization and cytotoxicity. *Bioorg. Med. Chem. Lett.* 20(10), 3150-3154.
- Schaeffer, C., Bardoni, B., Mandel, J.L., Ehresmann, B., Ehresmann, C., and Moine, H. (2001). The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *EMBO J* 20(17), 4803-4813.
- Schmittgen, T.D., and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* 3(6), 1101-1108.
- Shahid, R., Bugaut, A., and Balasubramanian, S. (2010). The BCL-2 5' untranslated region contains an RNA G-quadruplex-forming motif that modulates protein expression. *Biochemistry* 49(38), 8300-8306.
- Siddiqui-Jain, A., Grand, C.L., Bearss, D.J., and Hurley, L.H. (2002). Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. U. S. A.* 99(18), 11593-11598.
- Simonsson, T. (2001). G-quadruplex DNA structures--variations on a theme. *Biol Chem* 382(4), 621-628.
- Simonsson, T., Pecinka, P., and Kubista, M. (1998). DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Res.* 26(5), 1167-1172.
- Sirand-Pugnet, P., Durosay, P., Brody, E., and Marie, J. (1995). An intronic (A/U)GGG repeat enhances the splicing of an alternative intron of the chicken  $\beta$ -tropomyosin pre-mRNA. *Nucleic Acids Res.* 23(17), 3501-3507.
- Sissi, C., Gatto, B., and Palumbo, M. (2011). The evolving world of protein-G-quadruplex recognition: a medicinal chemist's perspective. *Biochimie* 93(8), 1219-1230.
- Smith, F.W., and Feigon, J. (1992). Quadruplex structure of Oxytricha telomeric DNA oligonucleotides. *Nature* 356(6365), 164-168.
- Sonenberg, N., Rupprecht, K.M., Hecht, S.M., and Shatkin, A.J. (1979). Eukaryotic mRNA cap binding protein: purification by affinity chromatography on sepharose-coupled m7GDP. *Proc. Natl. Acad. Sci. U. S. A.* 76(9), 4345-4349.
- Soukup, G.A., and Breaker, R.R. (1999). Engineering precision RNA molecular switches. *Proc. Natl. Acad. Sci. U. S. A.* 96(7), 3584-3589.
- Stein, C.A., Tonkinson, J.L., and Yakubov, L. (1991). Phosphorothioate oligodeoxynucleotides—anti-sense inhibitors of gene expression? *Pharmacol. Ther.* 52(3), 365-384.
- Stein, I., Itin, A., Einat, P., Skaliter, R., Grossman, Z., and Keshet, E. (1998). Translation of vascular endothelial growth factor mRNA by internal ribosome entry: Implications for translation under hypoxia. *Mol. Cell. Biol.* 18(6), 3112-3119.
- Stoneley, M., Paulin, F.E., Le Quesne, J.P., Chappell, S.A., and Willis, A.E. (1998). C-Myc 5' untranslated region contains an internal ribosome entry segment. *Oncogene* 16(3), 423-428.

- Stoneley, M., Subkhankulova, T., Le Quesne, J.P.C., Coldwell, M.J., Jopling, C.L., Belsham, G.J., et al. (2000). Analysis of the c-myc IRES; a potential role for cell-type specific trans-acting factors and the nuclear compartment. *Nucleic Acids Res.* 28(3), 687-694.
- Subramanian, M., Rage, F., Tabet, R., Flatter, E., Mandel, J.L., and Moine, H. (2011). G-quadruplex RNA structure as a signal for neurite mRNA targeting. *EMBO Rep.* 12(7), 697-704.
- Sun, D., Guo, K., Rusche, J.J., and Hurley, L.H. (2005). Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. *Nucleic Acids Res.* 33, 6070–6080.
- Sun, D.Y., Pourpak, A., Beetz, K., and Hurley, L.H. (2003). Direct evidence for the formation of G-quadruplex in the proximal promoter region of the RET protooncogene and its targeting with a small molecule to repress RET protooncogene transcription. *Clin. Cancer Res.* 9(16), 6122S-6123S.
- Svoboda, P., and Di Cara, A. (2006). Hairpin RNA: a secondary structure of primary importance. *Cell Mol. Life Sci.* 63(7-8), 901-908.
- Takahama, K., and Oyoshi, T. (2013). Specific binding of modified RGG domain in TLS/FUS to G-quadruplex RNA: tyrosines in RGG domain recognize 2'-OH of the riboses of loops in G-quadruplex. *J. Am. Chem. Soc.* 135(48), 18016-18019.
- Takahama, K., Takada, A., Tada, S., Shimizu, M., Sayama, K., Kurokawa, R., et al. (2013). Regulation of telomere length by G-quadruplex telomere DNA- and TERRA-binding protein TLS/FUS. *Chem. Biol.* 20(3), 341-350.
- Takahashi K., M.S. (1982). "Ribonuclease T1", in: *The Enzymes V*, (Boyer, P.D, Ed.), Academic Press, New York, the third edition.).
- Tan, W., and Yuan, G. (2013). Electrospray ionization mass spectrometric exploration of the high-affinity binding of three natural alkaloids with the mRNA G-quadruplex in the BCL2 5'-untranslated region. *Rapid Commun. Mass Spectrom.* 27(4), 560-564.
- Tang, C.F., and Shafer, R.H. (2006). Engineering the quadruplex fold: nucleoside conformation determines both folding topology and molecularity in guanine quadruplexes. *J. Am. Chem. Soc.* 128(17), 5966-5973.
- Tang, J., and Breaker, R.R. (1997). Rational design of allosteric ribozymes. *Chem. Biol.* 4(6), 453-459.
- Temperley, R.J., Wydro, M., Lightowlers, R.N., and Chrzanowska-Lightowlers, Z.M. (2010). Human mitochondrial mRNAs—like members of all families, similar but different. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1797(6–7), 1081-1085.
- Tinton, S.A., Schepens, B., Bruynooghe, Y., Beyaert, R., and Cornelis, S. (2005). Regulation of the cell-cycle-dependent internal ribosome entry site of the PITSLRE protein kinase: roles of Unr (upstream of N-ras) protein and phosphorylated translation initiation factor eIF-2alpha. *Biochem J.* 385(Pt 1), 155-163.
- Todd, A.K., Johnston, M., and Neidle, S. (2005). Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.* 33(9), 2901-2907.
- Todd, P.K., Mack, K.J., and Malter, J.S. (2003). The fragile X mental retardation protein is required for type-I metabotropic glutamate receptor-dependent translation of PSD-95. *Proc. Natl. Acad. Sci. U. S. A.* 100(24), 14374-14378.

- Tomasko, M., Vorlickova, M., and Sagi, J. (2009). Substitution of adenine for guanine in the quadruplex-forming human telomere DNA sequence G(3)(T(2)AG(3))(3). *Biochimie* 91(2), 171-179.
- Topisirovic, I., Siddiqui, N., Orolicki, S., Skrabanek, L.A., Tremblay, M., Hoang, T., et al. (2009). Stability of Eukaryotic Translation Initiation Factor 4E mRNA Is Regulated by HuR, and This Activity Is Dysregulated in Cancer. *Mol. Cell. Biol.* 29(5), 1152-1162.
- Tsukiyama-Kohara, K., Iizuka, N., Kohara, M., and Nomoto, A. (1992). Internal ribosome entry site within hepatitis C virus RNA. *J Virol.* 66(3), 1476-1483.
- Vaughn, J.P., Creacy, S.D., Routh, E.D., Joyner-Butt, C., Jenkins, G.S., Pauli, S., et al. (2005). The DEXH protein product of the DHX36 gene is the major source of tetramolecular quadruplex G4-DNA resolving activity in HeLa cell lysates. *J. Biol. Chem.* 280(46), 38117-38120.
- von Hacht, A., Seifert, O., Menger, M., Schutze, T., Arora, A., Konthur, Z., et al. (2014). Identification and characterization of RNA guanine-quadruplex binding proteins. *Nucleic Acids Res.* 42(10), 6630-6644.
- Wang, Y., and Patel, D.J. (1993). Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure* 1(4), 263-282.
- Wanrooij, P.H., Uhler, J.P., Shi, Y., Westerlund, F., Falkenberg, M., and Gustafsson, C.M. (2012). A hybrid G-quadruplex structure formed between RNA and DNA explains the extraordinary stability of the mitochondrial R-loop. *Nucleic Acids Res.* 40(20), 10334-10344.
- Wanrooij, P.H., Uhler, J.P., Simonsson, T., Falkenberg, M., and Gustafsson, C.M. (2010). G-quadruplex structures in RNA stimulate mitochondrial transcription termination and primer formation. *Proc. Natl. Acad. Sci. U. S. A.* 107(37), 16072-16077.
- Watson, J.D., and Crick, F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 171(4356), 737-738.
- Weng, H.-Y., Huang, H.-L., Zhao, P.-P., Zhou, H., and Qu, L.-H. (2012). Translational repression of cyclin D3 by a stable G-quadruplex in its 5' UTR Implications for cell cycle regulation. *RNA Biol.* 9(8), 1099-1109.
- Westmark, C.J., and Malter, J.S. (2007). FMRP mediates mGluR(5)-dependent translation of amyloid precursor protein. *PLoS Biol.* 5(3), 629-639.
- Wieland, M., and Hartig, J.S. (2007). RNA quadruplex-based modulation of gene expression. *Chem. Biol.* 14(7), 757-763.
- Willis, A.E., and Stoneley, M. (2004). Cellular internal ribosome entry segments: structures, trans-acting factors and regulation of gene expression. *Oncogene* 23(18), 3200-3207.
- Wolfe, A.L., Singh, K., Zhong, Y., Drewe, P., Rajasekhar, V.K., Sanghvi, V.R., et al. (2014). RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. *Nature* 513(7516), 65-70.
- Xu, Y. (2011). Chemistry in human telomere biology: structure, function and targeting of telomere DNA/RNA. *Chem. Soc. Rev.* 40(5), 2719-2740.
- Xu, Y., and Komiyama, M. (2012). Structure, function and targeting of human telomere RNA. *Methods* 57(1), 100-105.
- Yu, C.-H., Teulade-Fichou, M.-P., and Olsthoorn, R.C.L. (2014). Stimulation of ribosomal frameshifting by RNA G-quadruplex structures. *Nucleic Acids Res.* 42(3), 1887-1892.

- Zahler, A.M., Williamson, J.R., Cech, T.R., and Prescott, D.M. (1991). Inhibition of telomerase by G-quartet DNA structures. *Nature* 350(6320), 718-720.
- Zamiri, B., Reddy, K., Macgregor, R.B., Jr., and Pearson, C.E. (2014). TMPyP4 Porphyrin Distorts RNA G- quadruplex Structures of the Disease- associated r( GGGGCC) n Repeat of the C9orf72 Gene and Blocks Interaction of RNA-binding Proteins\*. *J. Biol. Chem.* 289(8), 4653-4659.
- Zhang, A.Y., Bugaut, A., and Balasubramanian, S. (2011). A sequence-independent analysis of the loop length dependence of intramolecular RNA G-quadruplex stability and topology. *Biochemistry* 50(33), 7251-7258.
- Zhang, D.H., Fujimoto, T., Saxena, S., Yu, H.Q., Miyoshi, D., and Sugimoto, N. (2010). Monomorphic RNA G-quadruplex and polymorphic DNA G-quadruplex structures responding to cellular environmental factors. *Biochemistry* 49(21), 4554-4563.
- Zhang, J.-y., Zheng, K.-w., Xiao, S., Hao, Y.-h., and Tan, Z. (2014a). Mechanism and Manipulation of DNA:RNA Hybrid G-Quadruplex Formation in Transcription of G-Rich DNA. *J. Am. Chem. Soc.* 136(4), 1381-1390.
- Zhang, S., Wu, Y., and Zhang, W. (2014b). G-quadruplex structures and their interaction diversity with ligands. *ChemMedChem* 9(5), 899-911.
- Zheng, K.W., Xiao, S., Liu, J.Q., Zhang, J.Y., Hao, Y.H., and Tan, Z. (2013). Co-transcriptional formation of DNA:RNA hybrid G-quadruplex and potential function as constitutional cis element for transcription control. *Nucleic Acids Res.* 41(10), 5533-5541.
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31(13), 3406-3415.