

**INVESTIGATING THE CODING REGION DETERMINANT BINDING
PROTEIN-RNA INTERACTION *IN VITRO*, IN CELLS AND *IN SILICO***

by

Gerrit van Rensburg

B.Sc., University of Northern British Columbia, 2012

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
MATHEMATICAL, COMPUTATIONAL, PHYSICAL AND MOLECULAR
SCIENCES
(BIOCHEMISTRY)

UNIVERSITY OF NORTHERN BRITISH COLUMBIA

JULY 2014

© Gerrit van Rensburg, 2014

UMI Number: 1526511

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI 1526511

Published by ProQuest LLC 2015. Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

The Coding Region-Determinant Binding-Protein (CRDBP) belongs to a family of RNA-binding proteins involved in embryogenesis and oncogenesis. Accumulative evidence suggests that the physical interaction with target mRNAs is a critical determinant for CRDBP function. The goal of this study is to better understand the CRDBP-RNA interaction. The role of the glycine-X-X-glycine motif in each of the KH domains of CRDBP for RNA binding *in vitro*, in cells and *in vivo* was investigated. Smaller RNA fragments of c-myc, MITF, MAPK4 and β -TrCP1 bound by CRDBP were determined and used for *in silico* analysis. Overall, the results showed that the G-X-X-G motif in the KH domains of CRDBP are critical for RNA binding, and that the KH domains play differential roles in binding different RNA molecules. The mapped smaller RNA regions aid in discovering inhibitory molecules and *in silico* analysis leading to improved approach for understanding the CRDBP RNA interaction.

TABLE OF CONTENTS

Abstract	ii
Table of Contents	iii
List of Tables	vi
List of Figures	viii
List of Abbreviations	xiv
Acknowledgements	xv
Research Contributions	xvi
Appendix	xvii
References	xxiii

Chapter 1- Introduction

1.1 Gene Expression and Post Transcriptional Regulation	1
1.2 mRNA Stability and mRNA Degradation Pathways	2
1.2.1 mRNA Processing, Half-Life, and Stability	2
1.2.2 mRNA Degradation Pathways	3
1.2.3 cis-Acting Elements	5
1.2.4 trans-Acting Factors	6
1.3 RNA-Binding Proteins	7
1.3.1 Well-characterized mammalian mRNA-binding proteins	7
1.3.2 <i>VICKZ</i> , CRDBP Family of RNA-Binding Protein Orthologs and RNA Partners	9
1.3.3 The Molecular Basis of RNA Binding Proteins: Well-Characterized RNA- Binding Domains	10
1.4 The Coding Region Determinant Binding Protein (CRDBP)	16
1.4.1 Proposed Model for CRDBP RNA Recognition and Binding	18
1.4.2 CRDBP in Cancer	20
1.5 Microphthalmia-associated transcription factor (MITF)	21
1.5.1 Function of MITF	21
1.5.2 Expression of MITF	22
1.6 RNA Folding: Pros and Cons	23
1.7 Objectives of this Research	30
1.7.1 Characterizing the structure of RNA targets of CRDBP	30
1.7.2 Characterizing the RNA-binding ability of the KH domains of CRDBP	30

Chapter 2- Characterizing the MITF, c-myc, MAPK4, and β -TrCP1 RNAs that bind CRDBP

2.1 Methodology	31
2.1.1 Protein Purification of Recombinant CRDBP	31
2.1.1.1 Transformation	31
2.1.1.2 Induction	32
2.1.1.3 Purification	32
2.1.1.4 12% SDS-PAGE of Purified CRDBP Protein Fractions	33
2.1.2 Dialysis of Isolated CRDBP Protein	34

2.1.3 Generation of Internally Radiolabelled RNA	34
2.1.4 Electrophoretic Mobility Shift Assays (EMSA).....	35
2.2 Results.....	36
2.2.1 MITF.....	39
2.2.2 c-myc	44
2.2.3 MAPK4.....	50
2.2.4 β -TrCP1	52
2.3 Discussion	57

Chapter 3- Analysis of CRDBP-RNA Interaction

Part A Characterizing the CRDBP-RNA interaction *in vitro* using Electrophoretic Mobility Shift Assay, in cells and *in vivo*

Part B Analysis of RNA Species using Circular Dichroism Spectroscopy

3.1 Methodology	60
Part A	
3.1.1 Purification and Dialysis of WT-CRDBP and Variants	60
3.1.2 Radiolabeling of RNA and Electrophoretic Mobility Shift Assay	61
3.1.3 Quantification and Representation of EMSA	61
3.1.4 Immuno-precipitation of CRDBP Variants from HeLa cells	61
3.1.5 RT-qPCR of RNA Associated with FLAG-CRDBP Variants	64
3.1.6 <i>D. rerio</i> studies	65
3.1.6.1 Preparation of pSp64TNE-Plasmid	65
Part B	
3.1.7 Circular Dichroism Conditions and Settings	66
3.2 Results	
Part A	
Part I: <i>In vitro</i>	67
3.2.1 EMSAs with c-myc RNA and CRDBP Variants	67
3.2.2 EMSA with MITF RNA and CRDBP Variants	70
Part II: In cells.....	75
Part III: <i>In vivo</i>	76
Part B	78
3.2.7 Analysis of RNA Species using Circular Dichroism Spectroscopy	78
3.3 Discussion	85
Part A	
Part I: <i>In vitro</i>	85
Part II: In cells.....	87
Part III: <i>In vivo</i>	88
Part B: Analysis of RNA species using Circular Dichroism spectroscopy	88

Chapter 4- Using *in silico* analytical methods to compare and contrast RNAs that bind CRDBP

4.1 Methodology	92
4.2 Results	96
4.3 Discussion	126

Chapter 5- General Discussion

5.1 General Overview	133
5.2 Mapping RNA	134
5.3 Investigating the Role of KH Domains of CRDBP in Binding RNA Targets <i>In Vitro</i> and in Cells	135
5.4 The Attempt to Use Zebrafish Model to Understand the Significance of KH domains of CRDBP in Conferring an <i>In Vivo</i> Phenotype	139
5.5 An Attempt to Use Circular Dichroism Spectroscopy to Understand CRDBP-RNA Interaction	140
5.6 <i>In Silico</i> Analyses of the Structure of RNAs that Bind CRDBP	140
5.7 Future Directions	142
5.8 Concluding Remarks	144

List of Tables

Table 1. Comparing computational algorithms for RNA structure prediction	28
Table 2. Reagents used in the purification, analysis, dialysis and EMSA protocols of 6xHIS CRDBP.....	35
Table 3. List of primers used for PCR to generate MITF DNA templates for <i>in vitro</i> transcription.....	37
Table 4. List of primers used for PCR to generate c-myc DNA templates for <i>in vitro</i> transcription.....	37
Table 5. List of primers used for PCR to generate MAPK DNA templates for <i>in vitro</i> transcription.....	38
Table 6. List of primers used for PCR to generate β -TrCP1 DNA templates for <i>in vitro</i> transcription.....	39
Table 7. MITF RNA fragments assessed for binding with recombinant WT-CRDBP	41
Table 8. c-myc RNA fragments examined by EMSA for interaction with CRDBP	45
Table 9. MAPK4 RNA fragments examined by EMSA for interaction with CRDBP	51
Table 10. β -TrCP1 RNA fragments examined by EMSA for interaction with CRDBP	53
Table 11. Primer sets used in real-time RT-PCR experiments	64
Table 12. Primers used in the gene amplification of WT-CRDBP and Variants	66
Table 13. RNA Molecules analyzed by Circular Dichroism	67
Table 14. Summary table of the different CRDBP variants and determined K_d values	74
Table 15. The progression of RNA primary sequence associated with the VICKZ family of RNA Binding proteins.....	92
Table 16. Oncogenic RNAs with high affinity for CRDBP	93
Table 17. Structures generated with Mfold of the six oncogenic RNAs	97
Table 18. RNAfold generated structures of the six oncogenic RNAs.	101
Table 19. Sfold generated secondary structures of the six oncogenic RNAs	104
Table 20. MC-FOLD Top 1 from 100 non-pseudoknot and Top 1 from 20 pseudoknot and Top Pseudoknot structure folded in MC-SYM.....	112
Table 21. MC-SYM Tertiary Structure Information with possible H-type pseudoknots.....	115
Table 22. LocARNA results from the six oncogenic RNAs.....	117
Table 23. LocARNA results from 2000 bps surrounding the smaller regions of interest of the six oncogenic RNAs. The table includes the alignment and shared secondary structure	118
Table 24. LocARNA results from the six oncogenic RNAs.....	118
Table 25. Predicted shared stem regions of RNA alongside the predicted secondary structure of RNA without GC Threshold Enabled in CARNAC.....	120
Table 26. Predicted shared stem regions of RNA alongside the predicted secondary structure of RNA with GC Threshold Enabled in CARNAC.....	123

Table 27. A Summary of the Dissociation Constant K_d of Four RNAs that Bind CRDBP	138
Table S1. List of all the base-paired secondary structure features.....	xviii
Table S2. List of the single stranded bases	xix
Table S3. Summary table of the calculated average percent bound and standard deviation for WT-CRDBP with c-myc 1705-1886 (182 nts) RNA	xxi

List of Figures

Figure 1. The VICKZ Phylogenetic tree indicating each orthologs accession number and the number of amino acid substitutions. Adaptation from Bell <i>et al.</i> , 2013 (1).....	9
Figure 2. Illustration of the Type I and Type II KH Domains. Indicated in black is the minimal KH motif requirement with β -strands labeled as arrows and α -helices as cylinders. This figure is Fig. 1 from Valverde <i>et al</i> 2008 (2)	14
Figure 3. The Annotation of <i>Mus musculus</i> CRDBP Coding Region from N-terminus to C-terminus. The entire sequence of <i>M. musculus</i> CRDBP coding region is 577 amino acids. KH1 amino acid 199 to 248. KH2 amino acid 280 to 330. KH3 amino acid 409 to 458. KH4 amino acid 491to 541	16
Figure 4. Cartoon of MITF dimer in cyan and green with DNA in grey. Missing protein loops are shown by dotted lines. Adaptation from Pogenberg <i>et al.</i> 2012 (3).....	21
Figure 5. RNA Molecules and their predicted tertiary structures. Superimposed <i>in silico</i> predicted structures (blue) are superimposed onto experimental structures (gold). (a) a hairpin with internal loop (28SP), (b) a duplex with two bulges (1I9X), (c) the pseudoknot 1KPZ, (d) Hammerhead ribozyme RNA 1NYI, (e) tRNA 1J1U and (f) SRP RNA 1Z43. Adaptation from Zhao <i>et al.</i> 2012 (4).....	24
Figure 6. Electrophoretic Mobility Shift Assay of recombinant CRDBP with MITF RNA fragments (1330-1550 and 1550-1740) and c-myc 1705-1886. [³² P] RNA (20,000 cpm/reaction) incubated with varying concentrations of CRDBP as shown. Samples were electrophoresed in a 4% native polyacrylamide followed by autoradiography	40
Figure 7. <i>In vitro</i> transcription of ³²P-UTP labeled MITF RNAs. PCR-amplified DNAs were used as templates for <i>in vitro</i> transcription to generate MITF RNA fragments. (A) nts 1330-1550 and 1550-1740. (B) 1550-1740, 1581-1740, and 1671-1740. (C) 1550-1619, 1550-1589, 1671-1740, and 1701-1740. (D) 1550-1740, 1550-1709, and 1550-1669. RNA samples were then electrophoresed in a 6% denaturing polyacrylamide gel followed by autoradiography	41
Figure 8. Electrophoretic Mobility Shift Assay of recombinant CRDBP with MITF RNA fragments. [³² P] MITF RNAs (20,000 cpm) nts 1550-1740, 1550-1709, and 1550-1669 were incubated with varying concentrations of CRDBP as shown. Samples were electrophoresed in a 4% native polyacrylamide followed by autoradiography	42
Figure 9. Electrophoretic Mobility Shift Assay of recombinant CRDBP with MITF RNA fragments. [³² P] MITF RNAs (20,000 cpm) nts 1550-1740, 1550-1619, and 1550-1589 were incubated with varying concentrations of CRDBP as shown. Samples were electrophoresed in a 4% native polyacrylamide followed by autoradiography	42
Figure 10. Electrophoretic Mobility Shift Assay of recombinant CRDBP with MITF RNA fragments. [³² P] MITF RNAs (20,000 cpm) nts 1550-1740 and 1581-1740 were incubated with varying concentrations of CRDBP as shown.	

Samples were electrophoresed in a 4% native polyacrylamide followed by autoradiography	43
Figure 11. Electrophoretic Mobility Shift Assay of recombinant CRDBP with MITF RNA fragments. [³² P] MITF RNAs (20,000 cpm) nts 1621-1740, 1701-1740, and 1671-1740 were incubated with varying concentrations of CRDBP as shown. Samples were electrophoresed in a 4% native polyacrylamide followed by autoradiography	43
Figure 12. Summary of MITF RNA fragments and relative binding affinity for CRDBP. The region of interest is shown in blue and the smaller fragments are shown in black. (++++), (+++), (++) , and (+) indicate ~90%, ~70%, ~40% and ~10% binding respectively. (+/-) and (-) indicate weak and no protein-RNA complex respectively	44
Figure 13. <i>In vitro</i> transcription of ³²P-UTP labeled c-myc RNA Fragments. PCR-amplified DNAs were used as templates for in vitro transcription to generate c-myc RNA fragments nts 1705-1886, 1705-1854, 1705-1814, 1735-1792, and 1705-1744. RNA samples were then electrophoresed in a 6% denaturing polyacrylamide gel followed by autoradiography	45
Figure 14. <i>In vitro</i> transcription of ³²P-UTP labeled c-myc RNA Fragments, PCR-amplified DNAs were used as templates for in vitro transcription to generate c-myc RNA fragments nts 1705-1886, 1735-1886, 1775-1886, 1815-1886, and 1845-1886. RNA samples were then electrophoresed in a 6% denaturing polyacrylamide gel followed by autoradiography	46
Figure 15. Electrophoretic Mobility Shift Assay of recombinant CRDBP with c-myc RNA fragments. [³² P] c-myc RNAs (20,000 cpm) nts 1705-1886, 1705-1854, and 1705-1814 were incubated with varying concentrations of CRDBP as shown. Samples were electrophoresed in a 4% native polyacrylamide followed by autoradiography	47
Figure 16. Electrophoretic Mobility Shift Assay of recombinant CRDBP with c-myc RNA fragments. [³² P] c-myc RNAs (20,000 cpm) nts 1705-1886, 1705-1792, and 1705-1744 were incubated with varying concentrations of CRDBP as shown. Samples were electrophoresed in a 4% native polyacrylamide followed by autoradiography	47
Figure 17. Electrophoretic Mobility Shift Assay of recombinant CRDBP with c-myc RNA fragments. [³² P] c-myc RNAs (20,000 cpm) nts 1705-1886, 1735-1886, 1775-1886, and 1845-1886 were incubated with varying concentrations of CRDBP as shown. Samples were electrophoresed in a 4% native polyacrylamide followed by autoradiography	48
Figure 18. Electrophoretic Mobility Shift Assay of recombinant CRDBP with c-myc RNA fragments. [³² P] c-myc RNAs (20,000 cpm) nts 1735-1886, 1775-1886, 1735-1792, 1815-1886, and 1845-1886 were incubated with varying concentrations of CRDBP as shown. Samples were electrophoresed in a 4% native polyacrylamide followed by autoradiography	48
Figure 19. <i>In vitro</i> transcription of ³²P-UTP labeled c-myc RNA Fragment and Electrophoretic Mobility Shift Assay of recombinant CRDBP. (A) PCR-amplified DNA used as templates for in vitro transcription to generate c-myc RNA fragment nts 1735-1792. RNA samples were then electrophoresed in a 6%	

denaturing polyacrylamide gel followed by autoradiography. (B) [³² P] c-myc RNAs (20,000 cpm) nts 1735-1886, 1775-1886, and 1735-1792 were incubated with varying concentrations of CRDBP as shown. Samples were electrophoresed in a 4% native polyacrylamide followed by autoradiography	49
Figure 20. Summary of c-myc RNA fragments and their relative binding affinity for CRDBP. The region is interest is shown in blue and the smaller fragments are shown in black. (++++), (+++), (++) , and (+) Indicate ~90%, ~70%, ~40% and ~10% binding respectively. (+/-) and (-) Indicate weak and no protein-RNA complex was visualized	50
Figure 21. <i>In vitro</i> transcription of ³²P-UTP labeled MAPK RNA Fragments. PCR-amplified DNAs were used as templates for in vitro transcription to generate MAPK RNA fragments nts 2764-3089, 3071-3424, 3403-3754, 3723-4063, 4036-4386 and 4386-4713. RNA samples were then electrophoresed in a 6% denaturing polyacrylamide gel followed by autoradiography	51
Figure 22. Electrophoretic Mobility Shift Assay of recombinant CRDBP with MAPK RNA fragments. [³² P] MAPK RNAs (20,000 cpm) nts 2764-3089, 3723-4063, 4036-4386, 4368-4713, and β-TrCP1 1021-1219 were incubated with varying concentrations of CRDBP as shown. Samples were electrophoresed in a 4% native polyacrylamide followed by autoradiography	52
Figure 23. Summary of MAPK RNA fragments and their relative binding affinity for CRDBP. The region is interest is shown in blue and the smaller fragments are shown in black. (++++), (+++), (++) , and (+) Indicate ~90%, ~70%, ~40% and ~10% binding respectively. N/A indicates the binding affinity could not be assessed	52
Figure 24. <i>In vitro</i> transcription of ³²P-UTP labeled β-TrCP1 RNA Fragments PCR-amplified DNAs were used as templates for in vitro transcription to generate β-TrCP1 RNA fragments nts 661-862, 853-1040, and 1021-1219 RNA samples were then electrophoresed in a 6% denaturing polyacrylamide gel followed by autoradiography	54
Figure 25. Summary of β-TrCP1 RNA fragments and their relative binding affinity for CRDBP. The region of interest is shown in blue and the smaller fragments are shown in black. (++++), (+++), (++) , and (+) Indicate ~90%, ~70%, ~40% and ~10% binding respectively	54
Figure 26. Electrophoretic Mobility Shift Assay of recombinant CRDBP with β-TrCP1 RNA fragments. [³² P] β-TrCP1 RNAs (20,000 cpm) nts 510-678, and 1021-1219 were incubated with varying concentrations of CRDBP as shown. Samples were electrophoresed in a 4% native polyacrylamide followed by autoradiography	54
Figure 27. Electrophoretic Mobility Shift Assay of recombinant CRDBP with β-TrCP1 RNA fragments. [³² P] β-TrCP1 RNAs (20,000 cpm) nts 853-1040, 1021-1219, and 1203-1372 were incubated with varying concentrations of CRDBP as shown. Samples were electrophoresed in a 4% native polyacrylamide followed by autoradiography	55
Figure 28. <i>In vitro</i> transcription of ³²P-UTP labeled β-TrCP1 RNA Fragments. PCR-amplified DNAs were used as templates for in vitro transcription to generate β-TrCP1 RNA fragments nts 1051-1219, 1091-1219,	

1131-1219, 1021-1190, 1021-1150 and 1021-1110 RNA samples were then electrophoresed in a 6% denaturing polyacrylamide gel followed by autoradiography	55
Figure 29. Electrophoretic Mobility Shift Assay of recombinant CRDBP with β-TrCP1 RNA fragments. [32 P] β -TrCP1 RNAs (20,000 cpm) nts 1021-1219, 1051-1219, 1091-1219 and 1131-1219 were incubated with varying concentrations of CRDBP as shown. Samples were electrophoresed in a 4% native polyacrylamide followed by autoradiography	56
Figure 30. Electrophoretic Mobility Shift Assay of recombinant CRDBP with β-TrCP1 RNA fragments. Autoradiograph of β -TrCP1 RNA (20,000 cpm) incubated with varying concentrations of CRDBP after electrophoresis at 25mA for 55mins in a 4% polyacrylamide gel exposed for 8 hours. [32 P] β -TrCP1 RNAs (20,000 cpm) nts 1021-1219, 1021-1190, 1021-1150 and 1021-1110 were incubated with varying concentrations of CRDBP as shown. Samples were electrophoresed in a 4% native polyacrylamide followed by autoradiography	56
Figure 31. Summary of β-TrCP1 RNA fragments and their relative binding affinity for CRDBP. The region of interest is shown in blue and the smaller fragments are shown in black. (++++), (+++), (++) and (+) Indicate ~90%, ~70%, ~40% and ~10% binding respectively	57
Figure 32. The equation for calculating molar ellipticity per residue from measured mdeg	67
Figure 33. Electrophoretic mobility shift assay of the recombinant wild-type CRDBP and CRDBP variants. [32 P] c-myc 1705-1886 (182 nts) RNA substrate (20,000 cpm/reaction) was incubated with varying concentrations of CRDBP variants as shown. Samples were electrophoresed in a 4% native PAGE gel followed by autoradiography. Data shown are representatives from at least four experiments using proteins from at least two separate preparations.....	68
Figure 34. CRDBP variants binding curves as generated by the Hill Equation. Binding curves were generated using densitometry and the data from Figure 3A-B with at least two biological replicates as well as three separate experiments were pooled. CRDBP variants were incubated with [32 P] c-myc RNA 1705-1886 (182 nts) and the binding profiles were analyzed using densitometry. Percent bound vs Percent unbound values were used to generate the binding curves and calculate the dissociation constant.....	69
Figure 35. Electrophoretic mobility shift assay of recombinant wild-type and mutant CRDBP. [32 P] MITF RNA 1550-1740 (191 nts) RNA substrate (20,000 cpm/reaction) was incubated with varying concentrations of CRDBP variants as shown. Samples were electrophoresed in a 4% native PAGE gel followed by autoradiography. Data shown are representatives from at least three experiments using at least two separately prepared recombinant proteins.....	71
Figure 36. CRDBP variants binding curves as generated by the Hill Equation. Binding curves were generated using densitometry from Figure 35A-H. The data were pooled from at least two biological replicates as well as four separate experiments. CRDBP variants were incubated with [32 P] MITF RNA 1550-1740 (191 nts) and the binding profiles were analyzed using densitometry.	

Percent bound vs Percent unbound values were used to generate the binding curves and calculate the dissociation constant.....	73
Figure 37. Analysis of specific mRNAs physically associated with the WT CRDBP and its variants. pcDNA-FLAG-CRDBP plasmids were transfected into HeLa cells as described in the Methodology section. FLAG-CRDBP-RNA complexes in cell lysates were immune-precipitated with anti-FLAG antibody. (A) Immuno-precipitated lysate was used in Western blot analysis to detect and confirm comparable exogenous FLAG-CRDBP protein expression using anti-FLAG antibody. (B-D) β -actin, CD44 and c-myc mRNAs isolated from the complexes were measured using qRT-PCR. mRNA levels were expressed over the pcDNA-FLAG control and compared to WT-CRDBP. Statistical significance between the expression of target genes in WT-CRDBP and CRDBP variants-transfected cells was assessed by the unpaired Student's t-test (n=4). The asterisk indicates that the p-value is less than 0.05	76
Figure 38. Circular Dichroism Spectra of Kras 240-332 (93 nts) RNA at different concentrations. (A) Kras 240-332 (93 nts) at 435nM. (B) Kras 240-332 (93 nts) at 869nM. (C) Kras 240-332 (93 nts) at 1738nM. (D) Kras 240-332 (93 nts) at 3314nM. Kras RNA was resuspended in nuclease-free water.....	79
Figure 39. Circular Dichroism Spectra of different RNA molecules in water. The c-myc 1705-1792 (88 nts) RNA at 3514 nM is in blue. The MITF 1621-1709 (89 nts) RNA at 3491 nM is in purple. The Kras 240-332 (93 nts) RNA at 3337 nM is in green. The Kras 148-239 (92 nts) RNA at 3314 nM is in orange.....	80
Figure 40. Circular Dichroism Spectra of different RNA molecules in dialysis buffer. The c-myc 1705-1792 (88 nts) RNA at 3514 nM is in blue. The MITF 1621-1709 (89 nts) at 3491 nM is in purple. The Kras 240-332 (93 nts) RNA at 3337 nM is in green. The Kras 148-239 (92 nts) RNA at 3314 nM is in orange	81
Figure 41. Comparison between water and dialysis buffer on the Circular Dichroism Spectra of different RNA molecules. The solid colour line represents the sample diluted in water and the dotted black line represents the sample diluted in dialysis buffer. (A) c-myc 1705-1792 (88 nts) at 3514 nM. (B) MITF 1621-1709 (89 nts) at 3491nM. (C) Kras 240-332 (93 nts) at 3337nM. (D) Kras 148-239 (92 nts) at 3314 nM.....	82
Figure 42. Circular Dichroism Spectra of Kras 148-239 (92 nts) RNA in the presence of WT-CRDBP. (A) Kras 148-239 (92 nts) 1807 nM RNA in dialysis buffer. (B) Kras 148-239 (92 nts) 1807 nM RNA in the presence of WT-CRDBP 1807 nM. (C) Overlay of CD Spectra A and B.....	83
Figure 43. Circular Dichroism Spectra of different Kras RNA molecules in the presence of WT-CRDBP. (A) Kras 240-332 (93nts) 1824nM RNA in the presence of WT-CRDBP 1824nM. (B) Kras 148-239 (92nts) 1807nM RNA in the presence of WT-CRDBP 1807nM. (C) Overlay of CD Spectra A and B.....	84
Figure 44. Circular Dichroism Spectra of Kras 240-332 (93nts) RNA at different concentrations in the presence of WT-CRDBP. (A) Kras 240-332 (93nts) 1824nM RNA in the presence of WT-CRDBP 1824nM. (B) Kras 240-332 (93nts) 912nM RNA in the presence of WT-CRDBP 1824nM. (C) Overlay of CD Spectra A and B	85

Figure 45. Predicted Dot-Bracket Structures from CARNAC without GC Threshold Enabled. These were input into Mfold and displayed in Table 25.....	119
Figure 46. Clustal Alignment by CARNAC with GC Threshold Enable, displaying the nucleotide composition above and dot-bracket secondary structure information below of the six oncogenic RNAs.....	122
Figure 47. Predicted Dot-Bracket Structures from CARNAC with GC Threshold Enabled. These were input into Mfold and displayed in Table 26.....	123
Figure S1. Ribonuclease secondary structure probing of human <i>c-myc</i> CRD RNA nts 1705-1792 (88nts), indicating the locations of bases accessible for enzyme probes. Adapted from Tafech <i>et al</i> , 2007 (5)	xvii
Figure S2. Ribonuclease secondary structure probing of human CD44 RNA nts 2861-2958 (98nts). Adapted from Kim <i>et al</i> , 2010 (6)	xviii
Figure S3. Autoradiograph to calculate the dissociation constant (K_d). Each autoradiograph that was generated was then used to calculate the K_d for CRDBP. Total signal from each lane (red lines) was calculated by drawing quantification boxes (blue boxes) around the signal generated from the [32 P] <i>c-myc</i> CRD 1705-1886 RNA substrate. K_d values were then calculated using the Hill equation	xxi
Figure S4. WT-CRDBP binding curves as generated by the Hill Equation. WT-CRDBP was incubated with [32 P] <i>c-myc</i> CRD 1705-1886 from Table S3.....	xxii
Figure S5. The Hill Equation for calculating dissociation constant (K_d).....	xxii

List of Abbreviations

AONs	Anti-sense oligonucleotides
Bps	Base pairs
CD	Circular Dichroism
Cpm	counts per minute
CRDBP	Coding Region Determinant-Binding Protein
DNA	deoxyribose nucleic acid
ds	double stranded
EMSA	Electrophoretic Mobility Shift Assay
FMR1	Fragile-X Mental Retardation protein-1
IP	Immuno-precipitation
IVT	<i>In vitro</i> transcription
KH	K homology
M	Molar
MFE	minimum free energy
miRNA	microRNA
N/A	Not available
Nts	nucleotides
PCR	Polymerase Chain Reaction
RBP	RNA Binding Protein
RNA	Ribose nucleic acid
siRNA	small interfering RNA
ss	single stranded
UD	undetermined
μ L	Microliter
UTR	Untranslated Region
VICKZ	Family of RNA binding proteins: Vera, IMPs, CRDBP, KOC, and ZBP-1
WT	Wild Type
ZBP-1	β -actin zipcode-binding protein 1

Acknowledgements

Thank you Joseph and Eunice Kim for my introduction to Dr. Chow Lee, which led to the beginning of my research. I am very thankful for Dr. Chow Lee and Dr. Maggie Li, they have provided me with endless learning experiences and excellent mentorship. A special thank you is due for my committee members Dr. Andrea Gorrell and Dr. Dezene Huber for their generous support throughout this entire process. My journey as a graduate student has been filled with challenges that have taught me valuable lessons in working efficiently while under pressure. How I approach obstacles in the future has been shaped by these challenges faced. My research was conducted alongside three great individuals: Mark Barnes, Kashif Mehmood and Sebastian Mackedenski, all three of them were always encouraging, gregarious and intellectually driven persons. Thank you to the Rader, Gorrell and Gray lab members as they also helped me, and their company was always enjoyable. My parents and sister have always been there for me from the start and their continual support is always cherished. Most importantly I thank Veronica Thurlborn, for her unwavering encouragement and financial support while completing graduate school.

Research Contributions

Publication

Barnes M, **van Rensburg G**, Li WM, Mehmood K, Mackedenski S, King DT, Miller AL, and Lee CH. (2014) Insights into the CRD-BP-RNA Interaction through Site-Directed Mutagenesis at the GXXG Motif in KH Domains. (Submitted to the Journal of Biological Chemistry)

Abstracts

van Rensburg G, Barnes M, and Lee CH. (2014) Functional Investigation of the CRD-BP hnRNP K Homology Domains for *c-myc* Interaction. AACR 105th Annual Meeting 2014, San Diego, California.

Barnes M, **van Rensburg G**, and Lee CH. (2014) Investigating a Novel Approach to Interrupt the Cancer Cell Invasion. Realities of Northern Oncology Conference 2014.

van Rensburg G, Barnes M, and Lee CH. (2013) Structural Importance of CRD-BP hnRNP K Homology Variants *in-vitro* and in cells. RiboWest 2013, UNBC, Prince George.

van Rensburg G, Barnes M, and Lee CH. (2013) Determining the Functional Domains of the RNA-Binding Protein CRD-BP. 8th Annual UNBC Graduate Conference, UNBC, Prince George.

Chapter 1- Introduction

1.1 Gene Expression and Post Transcriptional Regulation of Gene Expression

One of the critical molecular processes in life is gene expression, the transfer of genetic information within the cell: from the nucleus, stored as deoxyribonucleic acid (DNA) to ribonucleic acid (RNA) and then to protein. At each of these steps, from DNA to RNA (transcription), RNA to protein (translation), at the protein level, and in transition between these steps is where regulation can occur. Six major points of control are, (i) control of how often a gene is transcribed (transcriptional control), (ii) control of the splicing and processing of pre-RNA transcripts (RNA processing control), (iii) selection of which mature mRNA are exported from the nucleus and where they are localized in the cytosol (RNA export and localization control), (iv) selection of which mRNA are translated in the cytosol (translation control), (v) selectively destabilizing specific mRNA in the cytosol (mRNA degradation control), and (vi) selectively activating, deactivating, degrading, or localizing specific protein after translation (protein activity control) (7).

For most genes, the major regulation point is at transcription, as this ensures exorbitant gene expression is prevented. However, once a gene is expressed, tight regulation of the mRNA is required to prevent cellular deregulation and disease. Post-transcriptional regulation, specifically at the level of translation and mRNA degradation, is how a loss of regulation can lead to pathogenesis. During stages of embryonic development, up until and including apoptosis, translational regulation influences cellular processes.

1.2 mRNA Stability and mRNA Degradation Pathways

1.2.1 mRNA Processing, Half-Life, and Stability

How is mRNA half-life regulated and how does the process of mRNA degradation influence gene expression? The current knowledge of mRNA stability and its relationship to the control of gene expression aims to answer these questions.

The half-life of eukaryotic mRNA ranges from 10-15 minutes for some genes (e.g. *c-fos* proto-oncogene) to several hours (>10hrs) (e.g. house-keeping genes such as β -actin) (8). The longevity of mature mRNA in the cytoplasm is influenced by cis-elements (sequences at 5'-UTR, 3'-UTR and coding region) and trans-acting factors (ribonucleases, RNA-binding proteins, miRNAs, non-coding RNAs).

The 5' cap of mRNA acts as a regulation point, required for export through the nuclear pore after recognition by the cap binding complex. Capping of the 5' end of mRNA with a guanosine triphosphate (GTP) in a 5' to 5' phosphate linkage, and methylation at the 7-nitrogen (m^7G) protects the 5' end from exonuclease degradation (9). Messenger mRNA stability is enhanced by the 5' cap in conjunction with eIF-4E/eIF4G and the 3'-polyadenylated tail forming a loop structure, which promotes translation by recruiting the 40S ribosomal subunit (9).

The 3'-end of the mRNA is polyadenylated before export from the nucleus. The polyadenylated tail is preceded by an AAUAAA signal sequence and the polyadenylation ranges in length from 80 to 250 nucleotides, serving to protect mRNA from 3'-end degradation. The poly(A) tail also aids in the exportation of the transcript from the nucleus, as well as in promoting translation (10).

1.2.2 mRNA Degradation Pathways

An important part of regulating mRNA degradation is breakdown of the transcript via exonucleases and endonucleases through the standard or specialized mRNA decay pathways. Having the ability to degrade transcripts, allows the cell to abolish any further protein production, controlling translation levels, protein levels and how the cell develops and responds to environmental factors.

Exonucleolytic degradation is the destruction of the RNA transcript from its 5' or 3'-end. These two events can occur independently of each other to halt translation, or deadenylation at the 3' end can occur followed by decapping at the 5' end before transcripts are degraded exonucleolytically (11).

The 5'-to-3' degradation pathway begins with excision of the m⁷G cap, which protects the 5' end of the RNA (11, 12). This decapping action is enzymatically facilitated by Dcp1p and Dcp2p, which exposes the 5' end of the transcript to Xrn1p, a major 5'-to-3' exonuclease (12). For 3'-to-5' exonuclease activity, the poly(A) tail is first deadenylated in mammalian cells by PARN to allow the exosome complex to degrade the transcript in a 3'-to-5' fashion (11).

Endonuclease cleavage within the body of RNA exposes both a 5'-end and 3'-end site of mRNA that can be easily degraded by Xrn1p and the exosome as mentioned above. However, this mRNA degradation pathway is still not fully defined. This is mainly due to the difficulty in detecting endonucleolytic mRNA decay products and hence the responsible endoribonucleases (11).

RNA interference (RNAi) is an example of a well-characterized two-step endonucleolytic pathway. First, dsRNA is formed by micro RNA (miRNA) or small

interfering RNA (siRNA) with the mRNA. The mRNA is then cleaved by the enzyme Dicer, into 20-25 base pair fragments known as siRNA. Second, these siRNA are unraveled into single stranded RNA (ssRNA) and bound by the RNA-induced slicing complex (RISC). RISC then uses this ssRNA, as a guide strand to bind other mRNA and cleave them with its catalytic Argonaute protein unit (13). Similarly, miRNA, which binds to mRNA, result in the mRNA being endonucleolytic cleaved by RISC.

Specialized mRNA degradation pathways act as a mode for controlling the quality of mRNA being translated. For example, in nonsense-mediated decay, mRNA containing a premature stop codon are decapped and destroyed before truncated proteins with deleterious properties are made (14). In the case of non-stop mRNA decay, a ribosome becomes stalled due to lack of a stop codon, Ski7p adapter protein binds to the stalled ribosome and recruits the exosome to mediate the mRNA decay (15). In some cases, the ribosome stalls mid-translation, then no-go mRNA decay occurs involving Dom34p and Hbs1p (16). In AU-rich element (ARE) mediated decay, adenine (A) and uracil (U) rich regions, or elements containing a 5'-AUUUA-3' sequence, promote the binding of ARE-binding proteins, resulting in increased deadenylation and endonuclease activity of the transcript (17). Other mRNA degradation pathways exist, aside from the mRNA degradation pathways mentioned above.

Amongst all of these pathways, one with marked relevance is the regulation of specific mRNA by microRNAs (miRNA). How RNA-binding proteins that recognize sequences in the untranslated regions (UTR) of the RNA and interrupt the miRNA mediated mRNA decay is of importance.

1.2.3 cis-Acting Elements

Sequences of RNA, called cis-elements, provide mRNA an additional mode of post-transcriptional or translational control by interacting with specific trans-acting factors.

Early examples of cis-acting elements, which allow for alternate gene expression and regulation by initiation of translation, was originally identified in virus RNA genomes, called internal ribosome entry segments or sites (IRESs) (18). These IRESs initiated translations via recruitment of the ribosome in the coding region, far downstream of the initial start codon and without the presence of a 5'-cap structure (cap-independent translation) (19-21). Although there is no strong evidence for either of the primary or secondary structures of IRES regions being responsible for this recruitment, most IRESs are GC-rich and therefore considered to have a complex secondary structure, suggesting secondary structure could play a key role in the function of IRESs (18).

Adenylate/uridylylate-rich elements (AU-rich elements or AREs) are a cis-regulatory element located in the 3'-UTR of several proto-oncogenic mRNA (22). These AREs are destabilizing determinant regions that down-regulate mRNA half-life (22, 23).

Iron response elements (IREs) are located at the 5'-UTR and 3'-UTR of Ferritin mRNA and Transferrin mRNA respectively (24). These IREs control the translation and degradation of Ferritin mRNA and Transferrin mRNA respectively (7).

The coding region instability determinant (CRD) is another cis-acting element found in *c-myc*, *c-fos*, interleukin-2 and other mRNAs (25). The CRD is defined as a pause site, due to a string of rare codons, which forces the ribosome to wait until a scarce

aminoacyl-tRNA arrives thus pausing the translation (25). This pause correlates with mRNA instability, allowing endoribonucleases to target the mRNA at the CRD site (25). The CRD acting as a destabilizing element can be protected from CRD-binding protein (CRDBP) by binding to the CRD site (25, 26). Once the CRD is bound by CRDBP, the mRNA half-life increases due to the protection from endonucleolytic degradation at that site (26).

From the above examples, it is clear that cis-acting elements play an important role in the control of mRNA degradation and translation. It also pointed out how the interaction between mRNA and protein can dictate the fate of an mRNA

1.2.4 trans-Acting Factors

These proteins, which interact with cis-acting elements, are called trans-acting factors (TAFs), or internal initiation trans-acting factors (ITAFs). ITAFs are thought to be responsible for initiation of translation when cap-dependent initiation is decreased, but also regulate mRNA decay through ribonucleases (RNases), microRNAs (miRNAs) and RNA-binding proteins (RBPs) (27).

As discussed in section 1.2.2, ribonucleases are protein enzymes important in cell metabolism. They control mRNA levels by catalyzing the degradation of mRNA into its component pieces. Two major groups are the endo and exo RNases. Exo RNases degrade RNA from either the 5'-end or 3'-end in a 5'-to-3' or 3'-to-5' direction respectively. These proteins function as TAFs by recognizing specific regions of ssRNA, dsRNA and RNA-DNA hybrids.

A class of short (~22nt) noncoding RNA, known as microRNAs (miRNAs) play an important role in controlling post-transcriptional gene expression by mRNA

degradation (28). In processing from pre-miRNA to mature miRNA, the mature miRNA associate with proteins, forming the RNA Induced Silencing Complex (RISC). This RISC complex that acts as a TAF to locate and bind to specific mRNA, regulating their half-life, as part of normal cell metabolism and gene regulation at the mRNA transcript level.

There are thousands of genes being regulated by miRNA at the transcriptional level, examples of specific interest are miR-183 with coding region of β -TrCP1 and miR-340 with 3'-UTR of MITF (29, 30). Both of these transcripts, which are down regulated by miRNA at targeted sites, are also protected by CRDBP, in the same regions preventing miRNA degradation of those transcripts (29, 30).

In conclusion, trans-acting factors can regulate gene expression at the post-transcriptional level, via interaction with mRNA.

1.3 RNA-Binding Proteins

1.3.1 Well-characterized mammalian mRNA-binding proteins

It is clear from the above discussion that RNA-binding proteins (RBPs) play a critical role in the post-transcriptional regulation of mRNA. In turn, the mRNAs bound by RBPs are known to play significant roles in many cellular processes, which include cell signaling, division, growth, development and cancer biology. Despite this, there is still a vast lack of studies on mRNA-binding proteins (31, 32).

Three well-characterized mRNA-binding proteins are HuR, CUG-BP, and FMR1. HuR, (also known as ELAVL1) is known to bind to several mRNAs including cytokines, interleukins and the 3'-UTR of CAT1 mRNA, which codes for a cationic amino acid

transporter in the liver (33, 34). In human hepatocarcinoma cell under cellular stress, HuR loses the ability to bind to the 3'-UTR of CAT1 mRNA and results in the release of CAT1 mRNA from P-bodies and recruitment to polysomes (31, 32, 35). This 3'-UTR region of CAT1 is the site for miRNA-122 binding, which end in translational repression and accumulation in P-bodies (32, 35).

CUG-BP, also known as Elav like family of RNA binding protein-1 (CELF-1) regulates several steps of RNA processing, and found in early embryonic development, as well as heart and skeletal muscle function (36). CELF-1 functions as a mRNA stability factor, and plays a role in myotonic dystrophy (DM), a form of muscle dystrophy affecting the skeletal and heart muscle (36). The onset of DM is seen as related to overexpression of CLEF1 due to CLEF1's importance in striated muscle function and RNAi knockdown of *C. elegans* homolog (ETR-1) results in defects in muscle attachment (36).

Fragile X syndrome is an inheritable disease that presents as constrictions on two arms of the X chromosome resulting in learning disability (37). This constriction in the arms of the X chromosome causes a loss of FMR-1 RBP, altering brain development, and in severe cases the protein is expressed, but carries a point mutation in the second K homology domain (I304N) (38).

Additional characterized family of RBPs is the VICKZ family of RBPs that includes coding region determinant binding protein (CRDBP). CRDBP is implicated in protecting mRNA during embryonic development and is also highly associated with cancer. CRDBP is the focus of this thesis and is discussed in greater detail in the following sections.

1.3.2 VICKZ, CRDBP Family of RNA-Binding Protein Orthologs and RNA Partners

There are a number of different RBPs with most RBPs consist of a few RBDs. One highly conserved family of RBPs responsible for recognizing specific *cis*-acting elements in variety of different RNAs, the VICKZ family (Figure 1) (1, 39, 40).

The VICKZ family of regulatory RBPs consist of the proteins Vg1 RBP/Vera, IMP-1,2,3, CRDBP, KOC, and ZBP-1, all homologs consisting of two RRM at the N-terminus of the protein and four KH domains at the C-terminus (40). Originating from a wide range of different organisms, all these proteins interact with mRNA and result in aiding of localization, stabilization and translocation of their target mRNA which are implicated in cell polarity, cell migration, cell proliferation and cancer (39). These post-transcriptional changes experienced by the mRNA in the presence of VICKZ proteins is seen as a source for the invasive, mobile and uncontrolled cell proliferation characteristics associated with cancer.

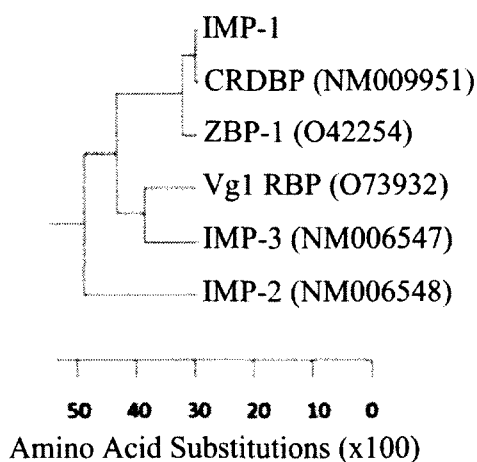


Figure 1. The VICKZ Phylogenetic tree indicating each orthologs accession number and the number of amino acid substitutions. Adaptation from Bell *et al.*, 2013 (1).

Vg1 RBP and ZBP-1 were recognized for RNA trafficking mechanism with specific RNA targets in *Xenopus laevis* (frog) oocytes and *Gallus gallus* (chick) embryos, respectively. Based on Vg1 RBP's high affinity binding to the 3'-UTR vegetal localization element (VLE) of *Vg1* RNA, and ability to localize the RNA to the vegetal cortex in *X. laevis* oocytes, Vg1 RBP is considered to be part of a larger complex of several RNA binding and motor proteins. The ZBP-1 present in *G. gallus* embryos is responsible for the leading edge of migrating neurons and fibroblasts (41, 42).

Nielsen and colleagues researched the human insulin-like growth factor II mRNA-binding proteins IMP-1, IMP-2 and IMP-3 on the basis of their ability to regulate translation of the *IGF-II* mRNA. It has also been shown in HeLa cells that IMP-1 and IMP-3 help orchestrate the formation of invadopodia and regulate cellular adhesion through binding with *CD44* mRNA (43). IMP-1 binds to the *IGF-II* mRNA at the 5' UTR, and in transgenic mice an increase in IGF-II protein can be measured after over-expression of IMP-1, leading to the induction of mammary tumors (44). Other oncogenes regulated by IMP-1 are *H19* mRNA at the 3'-UTR, *Tau* mRNA at the 3'-UTR and *c-myc* mRNA at the coding region determinant (45-47).

From the study on pancreatic carcinomas and hepatocellular carcinomas, Mueller-Pillasch found the KH-domain-containing protein is over-expressed in cancer (KOC) (48). KOC is a synonym for IMP-3.

1.3.3 The Molecular Basis of RNA Binding Proteins: Well-Characterized RNA-Binding Domains

When a protein interacts with nucleic acid, many non-covalent interactions are formed, with four main definitions of interaction (49). The first manner is the protein

side chain and the RNA nucleobases and is the major form of sequence specific interactions. Specificity arises due to the complexity and complementary of hydrogen bonding, van der Waals forces, hydrophobic attractions and global electrostatic forces presented by both protein and RNA (49).

The next possible mode of interaction (between protein and nucleic acid) can occur at the protein side chain with the phosphate backbone of nucleic acid. This commonly takes place as positively charged amino acids such as lysine and arginine attracted to the negatively charged phosphate backbone (49). Hydrogen bonding can also take place between the backbone of nucleic acid and side chain of protein, or amine groups of the protein backbone (49).

The last two ways in which protein can interact with nucleic acid is through the amine groups in the protein backbone with either the nucleic acids nucleobases or phosphate backbone (49). Often backbone contacts are considered a means of non-specific binding, however a protein can recognize and interrupt specific tertiary structure of nucleic acid, particular single stranded RNA (ssRNA), of which is dependent on particular base sequences (49). This then allows for such contacts to be used for sequence specific interactions. This basic understanding of how protein and nucleic acid could interact provides insight for how different protein domains recognize and interrupt RNA, and how ensembles of RNA-Binding domains (RBDs) in combination give rise to specificity and diversity of function.

RNA localization, stabilization, processing, rate of translational control, and nuclear export all naturally take place in the presence of proteins, by formation of protein-RNA complexes known as ribonucleoproteins (RNPs) (50). This diversity of

RNA-binding proteins (RBPs) outnumber the handful of RNA-binding domains (RBDs) they are made of and therefore the diversity of RBPs stems rather from the presence of multiple domains and their arrangement, as opposed to a wide variety of RBDs (50). The repetition of multiple similar domains is thought to provide the protein with increased affinity and specificity compared to individual domains, where individual domains tend to bind with a weak affinity for shorter stretches of RNA. By having several domains together allow increased lengths of RNA to be recognized and for the recognition of certain stretches of RNA to be separated by some lengths of nucleotides, or sequences that belong to two different RNAs. These individual domains can then evolve independently to better suit the proteins function. The six RBDs that will be discussed, are RNA Recognition Motifs (RRMs), heterogeneous nuclear ribonucleoprotein K-Homology domains (hnRNP KH domains), Double-stranded RNA Binding Domains (dsRBDs), S1 Domain, PAZ/PIWI Domains, and Zinc-Fingers.

RRMs or Ribonucleoprotein (RNP) domain are the best-characterized RNA-binding motif that were first found in yeast polyadenylated-binding protein and then in 250 additional proteins (50). Consisting of 80 to 90 amino acids in length, to form a 4-stranded anti-parallel β -sheet packed against two α -helices giving a split $\alpha\beta$ topology of β - α - β - α - β (49, 50). From crystalized RRM-RNA complexes, most show the β -sheet surface acts as the point of RNA recognition, and binding by three conserved residues of Arg or Lys, which forms a salt bridge with a phosphodiester backbone and two aromatic residues, making stacking interactions with nucleobases (49, 50). This is located within the domain on the central β -strands of the four-strand β -sheet, as two conserved motifs, RNP1 and RNP2 (49, 50). These RNP motifs are responsible for RNA binding

specificity of two nucleotides. An additional two nucleotides are recognized on either side of this central region, and by utilizing the other secondary structure and exposed loops not found in the canonical structure of the RRM, a total of four to eight nucleotides can be recognized. A classic example is the U1a spliceosomal protein containing two such RRM domains that binds to the hairpin of U1 snRNA (51). This structure is solved, and shows that the β -sheets form a binding platform allowing nonspecific interactions between the U1 snRNA hairpin and a highly negatively charged loop of residues, which join the two β -strands in the sheet (51). Here the RNA conformation is stabilized in order for the RRM to have specific interactions at the RNP1 and 2 motifs. Due to the small number of nucleotides recognized at one time (by a single RRM domain), multiple RRM domains are needed in order to obtain specificity and unique binding sequences. Aside from its original finding of being a RNA Recognition Motif, RRM, RRM domains do not all interact with RNA. In fact, some interact with protein (52).

K-homology (KH) domains, more formally hnRNP-KH domains, were first identified in the heterogeneous nuclear RNA-binding protein, hnRNP K, which contains 3 copies of this domain. The VICKZ family of proteins and FMR1 protein are other examples of KH domain containing proteins (53). The FMR1 has been established as a clear clinical target, where a mutation near a key signature sequence causes fragile-X mental retardation syndrome (53, 54). The KH domain is noted to span the eukaryotes, eubacteria and archaea domains of life. This protein domain is close to 70 amino acids in length and is composed of 3 β -sheets packed against 3 α -helices, and two alternate topologies exist giving rise to two types of KH domains (49, 50). Type I is β - α - α - β - β - α and type II is α - β - β - α - α - β (Figure 2) (2, 49, 50). In both types there is a signature

sequence near its center of (I/L/V)IGXXGXX(I/L/V) and four nucleotides are recognized in a cleft, due to a G-X-X-G loop motif (50). This G-X-X-G motif flanks the helices of the β -strand from α -2 (Type I) or α -3 (Type II) and a variable loop between β -2 and β -3 (Type I) or between α -2 and β -2 (Type II) (2, 49). Unlike the RRM, the binding platform has no aromatic amino acids, and recognition of RNA is through hydrogen bonding, electrostatic interactions and shape complementarity (50).

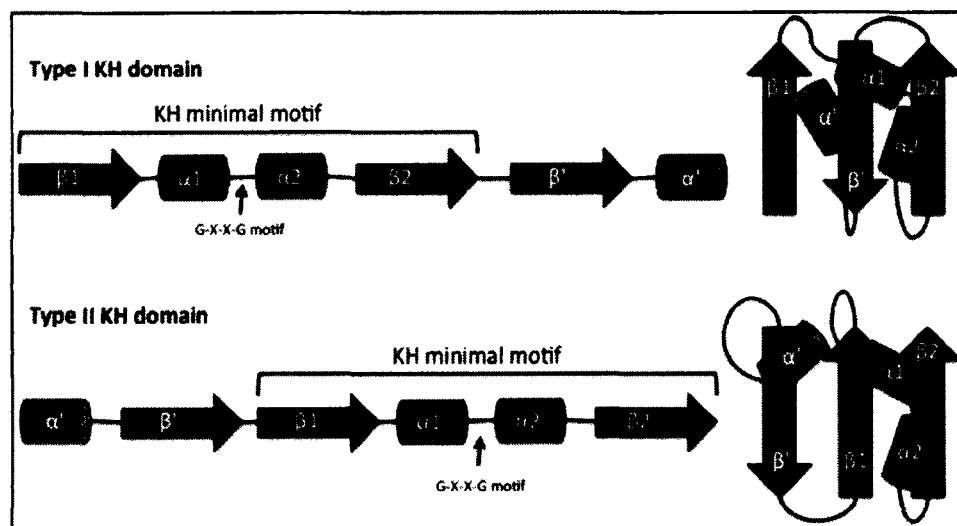


Figure 2. Illustration of the Type I and Type II KH Domains. Indicated in black is the minimal KH motif requirement with β -strands labeled as arrows and α -helices as cylinders. Adaptation from Fig. 1 of Valverde *et al* 2008 (2).

Double-stranded RBDs can be as short as 65 residues, but tend to range from 70 to 90 and are found in several proteins which bind dsRNA in bacteria and eukaryotes (49, 50). Like the KH domain, the dsRBD is centered on a three-strand β -sheet, but there are two α -helices (α - β - β - α), which is also similar to the RRM domain(49). When binding to dsRNA, the dsRBD binds to the face of the dsRNA helix, across two successive minor grooves and the major groove between, with no specific nucleobase interactions (50). This binding interaction occurs at a cleft in the domain between the second α -helix and the β -sheet, with most of the contacts dependent on the 2'-hydroxyl (-OH) groups and

phosphate backbone, which is unlike the RRM or KH domains (50). Although the dsRBD does not recognize specific nucleobase interactions, multiple dsRBDs can confer structure specificity as they could recognize different arrangements of helices. There are also some dsRBD having an N-terminal helix, which binds to bulges, stem-loops and irregular A-form RNA helices adding to the binding diversity of dsRBDs (50).

The identification of the ribosomal protein S1 led to the initial finding of the S1 domain, and since then other RBPs containing the S1 domain such as exo-nucleases have been identified (55). The S1 domain has approximately 70 residues, which form a 5-stranded anti-parallel β -barrel capped with a 3_{10} helix (50). This capped β -barrel is similar to the oligonucleotide/oligosaccharide binding (OB)-fold superfamily, which also contains the RNA-binding cold-shock domain (50). In common with the OB-fold binding surface, the S1 domain recognizes nucleic acids via two β -strands surrounded by several loops and is similar to the RRM domain, with the two-stranded β -sheet at the core and involving aromatic residues for base stacking with nucleic acids (50).

The PAZ and Piwi domains are an S1 or OB-fold like β -barrel linked to a small $\alpha\beta$ domain, forming a clamp like structure and are 110 amino acids in length (50). Dicer and Argonaute proteins have PAZ and Piwi domains for positioning miRNA and aid in cleavage of the target strand in the RISC (50).

Zinc Finger motifs (ZFM) are classically a DNA binding motif, but can also bind to RNA. The ZFM is classified on the residues responsible to coordinating a zinc ion, (e.g., Cys2His2, CCHH. Or CCCH, CCHC) and is usually present as multiple repeats in a protein. First identified in TFIIIA, with nine CCHH motifs, of which 3 interact with 5S RNA as a form of protein synthesis regulation (49). The TIS11d zinc-finger protein is an

example of CCCH zinc finger motifs bound to AU-rich RNA elements (ARE) (49). The ZMF interact with RNA at two loops, using the same recognition helices found in motifs for recognizing DNA via hydrogen bonding directly to the major groove or Watson-Crick base pairs (49). The major groove of RNA is too narrow to accept the conventional recognition helix presented, so it is possible that distinct residues make DNA and RNA contacts, or RNA is severely distorted at the recognition site to allow access of the major groove (49).

From these different RBDs it is important to note that they are vastly outnumbered by the total number of individual RBPs due to the combination and arrangement of RBDs. One must also then keep in mind that many different nucleic acid-binding structures are also occasionally involved in protein-protein interactions.

1.4 The Coding Region Determinant Binding Protein (CRDBP)

The c-myc coding region determinant binding protein (CRDBP) was initially identified in mice based on its ability to stabilize c-myc mRNA upon binding to a specific coding region of the transcript named the coding region determinant (56).

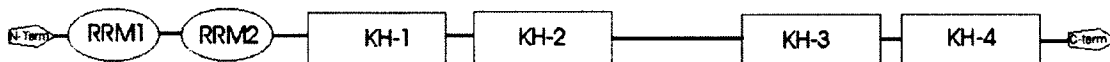


Figure 3. The Annotation of *Mus musculus* CRDBP Coding Region from N-terminus to C-terminus. The entire sequence of *M. musculus* CRDBP coding region is 577 amino acids. KH1 amino acid 199 to 248. KH2 amino acid 280 to 330. KH3 amino acid 409 to 458. KH4 amino acid 491 to 541.

1.4.1 The Cellular Role of CRDBP as an mRNA-Binding Protein

CRDBP binds to the following mRNAs whose gene products have been implicated in cancer: *c-myc*, *CD44*, *MITF*, *MAPK4*, *Gli-1*, β -*TrCP1*, *IGF-II*, *K-Ras*, *MDR-1* (30, 57-61). The working model is that CRDBP binds and shields these mRNAs

from endonucleolytic attack by endoribonuclease or miRNA, thereby prolonging half-life of the transcripts (29, 41, 62-64).

As a proto-oncogene, *c-myc*, codes for the protein MYC, which acts as a transcription factor regulating cell proliferation and differentiation (65). The stabilization of *c-myc* mRNA by CRDBP ultimately results in increased *c-myc* mRNA translation in cancer cells (57, 61, 63, 66). One study shows that c-Myc promotes the transcription of CRDBP and with this stabilization action and promotion relationship, a positive feedback loop is an easy conclusion (67). Similarly, stabilization of *CD44* mRNA by CRDBP leads to increased invadopodia and cell adhesion, as *CD44* is a cell surface protein for cellular attachment to extracellular matrix components and cell-cell adhesion (43, 58, 65, 68).

The overexpression of CRDBP also results in suppression of apoptosis of cancer cells that prolongs the turnover rate of β -*TrCP1* (61). β -*TrCP1* mRNA has a coding region which is bound by miR-183 and Argonaute-2 to facilitate the degradation of β -*TrCP1* mRNA and down regulate the translation for Skp1-Cullin1-F-box protein (SCF) ^{β TrCP1} E3 ligase which is a regulatory part of the cell cycle. This miR-183 binding site can be disrupted and/or occupied by CRDBP causing oncogenesis and the suppression of apoptosis (29).

Insulin-like growth factor II (IGF-II) mRNA plays an important role in cell proliferation, and is well known in regulating the development of breast cancer (69). Pacher *et al* found increased levels of IGF-II in breast cancer, and CRDBP interaction with IGF-II mRNA is a likely cause (69, 70). Using ELISA, increases in serum IGF-II was measured in MCF7 cells overexpressed with IGF-II indicating CRDBP regulates bioavailability and half-life of IGF-II circulation (70). Serum CRDBP in early-stage

breast cancer in women was detected using ELISA (69). Multi-drug resistance cancer cells commonly occur due to the production of a membrane protein that functions as an efflux pump with protective properties (71). This membrane protein, P-glycoprotein is coded for by MDR-1 mRNA, which has similar affinity as *c-myc* RNA for CRDBP, with a K_d of 500 nM (61, 71). Such affinity between CRDBP and MDR-1 mRNA was proposed to be the mechanism for the over expression of MDR1 and the multidrug resistance phenotype observed in cancer cells (61).

1.4.1 Proposed Model for CRDBP RNA Recognition and Binding

CRDBP's implication in cancer is based on its binding/stabilizing interactions with various mRNAs and its measured levels within a wide range of cancers. Binding sites for CRDBP-mRNA are found in the coding regions of *c-myc*, *β -TrCP1*, and *MDR-1* mRNAs and at the 3'-UTR of *CD44* mRNA. Segments of the ZBP-1 structure, a homologous protein to CRDBP, have been determined using X-ray crystallography. The study illustrated the truncated KH3-4 domains of ZBP-1 as the key domains for binding *β -actin* RNA as opposed to truncated RRM1-2 and KH1-2 domains (72).

The domains of many RNA binding proteins have been revealed through cDNA production of many hnRNPs indicating the K homology domain which shares identity with the VICKZ protein family and CRDBP (73). The common characteristic of most KH domains in RNA-binding proteins and this motif is the mechanism for RNA interaction is the Gly-X-X-Gly motif (74). The KH domain of NS1, an influenza virus protein, has a highly flexible loop comprising the same conserved G-X-X-G motif, which stiffens and stabilizes upon binding to RNA (74). The G-X-X-G motif, located between

α -helices 1 and 2 acts a C-terminus and N-terminus cap for the helices providing flexible stability, which allows interaction with RNAs (75). The α -helices 1 and 2 provided by this cap has been visualized under X-ray crystallography {Lewis, 2000 #23}. The α -helix 1 and β -sheet 2 of Nova-2 KH3 domain has been crystallized showing interaction with mRNA (54, 74). More importantly the interaction between the KH3 domain of Nova 2 and mRNA is due to the G-X-X-G sequence, producing Watson-Crick type hydrogen bonding with 5'-CA-3' after recognizing a 5'-UCAC-3' sequence of mRNA (54). By acting as a 5'-UG-3' sequence, the G-X-X-G motif binds to mRNA with a variant loop between β sheet 1 and the β' sheet that clamps the mRNA in place (54). However, the binding of each KH domain is very individualistic based on the location of the G-X-X-G sequence relative to the β sheets and α helices and the properties associated with the β sheets and α helices. The two RRM's did not demonstrate any contribution to the binding of the ZBP-1 to 3'-UTR of *β -actin*, but that the anti-parallel pseudodimer conformation of KH3 and KH4 together are responsible for the binding based on EMSA assays of truncated ZBP-1 at each of the RRMs and KH domains (individually and in combination) (72).

Researchers demonstrated the importance of the KH34 didomain of Vg1 RBP, the homologue of CRDBP in frogs, as a key component for binding and stabilizing interactions between Vg1 RBP and *Vg1* mRNA {Git, 2002 #24}. By examining point mutations introduced at the KH3 and KH4 domains, separately and simultaneously, only simultaneous point mutations interrupted the binding interaction (76). Mutants of the KH1, KH2 and KH1-2 domains showed change in Vg1 RBP abilities to bind RNA, but the single mutants (KH1 and KH2), and double mutant (KH1-2) did not have a large

difference in binding reduction, suggesting these KH domains are not as required for RNA binding (76). Although it was shown that KH34 didomains have a better interaction with the *β-actin* (for ZBP-1) and *Vg1* (for Vg1RBP) RNAs than the other domains together, and individually, these interactions might be mRNA specific and have different results with different mRNA (72, 76). This also suggests that no individual domain is absolutely necessary which is also seen with the conserved sequences in all of the KH domains. Producing truncated forms of the CRDBP homologues also greatly affects their tertiary structure and how the RRM and KH domain interact as a whole, with high affinity interactions possibly requiring all of the KH domains.

1.4.2 CRDBP in Cancer

CRDBP is also known as IMP1 (Insulin-like growth factor II mRNA-binding protein) identified as an oncofetal protein (62, 63, 77, 78). CRDBP is expressed during embryogenesis in *Mus musculus* and as IMP-1 in *Homo sapiens* in the developing epithelial, muscle and placental tissue. As this tightly regulated protein becomes unregulated, it is defined as an oncofetal protein, overexpressed *de novo* in various cancers and associated with aggression and invasiveness (62, 63, 67, 78).

One of the cancers implicated with CRDBP is colorectal carcinoma, with 46 of 78 cases (59%) measuring high levels of CRDBP (25). Expression of CRDBP has been measured by Ioannidis *et al* in a wide variety of cancers, such as non-small cell lung cancer, breast carcinomas, chronic myelogenous leukemia, sarcomas, and mesenchymal solid tumors (67, 78, 79). Previous research has shown that CRDBP expression is present in 73% of 33 malignant tumors tested and 40% of 10 benign mesenchymal tumors (64). In comparison to adjacent normal tissues, tissue from brain tumors, non-small cell lung

carcinomas, malignant, and benign neuroepithelial tumors all had high expression of CRDBP (64). By understanding how CRDBP's binds with various mRNA uncovers a biochemical mechanism of how those cancer overexpressing CRDBP can be targeted.

1.5 Microphthalmia-associated transcription factor (MITF)

1.5.1 Function of MITF

Microphthalmia-associated transcription factor (MITF) plays a role in the development, survival and function of melanocytes, retinal pigment epithelial cells, osteoclasts, mast cells as well as stem cells of the hair bulge. MITF is comprised of three main structural regions, (1): a Helix-Loop-Helix Motif, (2): a Leucine Zipper Motif, and (3): a Basic Motif. The Helix-Loop-Helix and Leucine Zipper motifs (bHLHzip) (Regions 1 and 2) are critical for protein interactions and key for MITF to form homodimers or heterodimers with other transcription proteins. Once dimerized, the Basic Motif is the functional region of MITF, which recognizes and binds deoxyribonucleic acid (DNA), facilitating transcription (Figure 4).

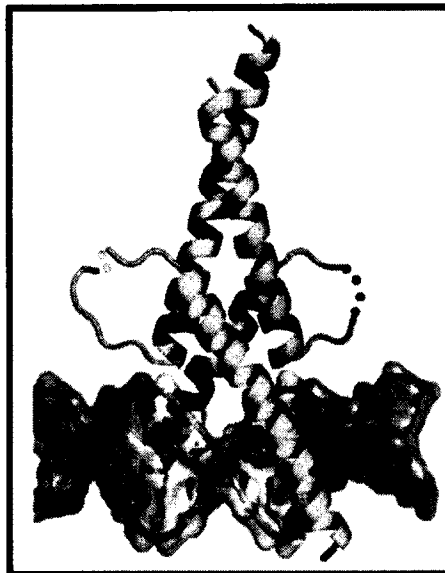


Figure 4. Cartoon of MITF dimer in cyan and green with DNA in grey. Missing protein loops are shown by dotted lines. Adaptation from Pogenberg *et al.* 2012 (3).

1.5.2 Expression of MITF

Hallmarks of melanoma are rapid and continuous cell division, increased cell pigmentation and cell migration. As a transcription factor, MITF is a master regulator of melanocyte development and an important oncogene in melanoma. Gene targets of MITF responsible for increased cell proliferation include T-box transcription factor 2 (TBX-2), cyclin-dependent kinase-2 (CDK2) and B-cell lymphoma-2 (BCL-2).

TBX-2 is a transcription factor, which down regulates p21, resulting in a suppression of cell senescence. CDK-2 modulates cell cycle progression at the G-1/S transition and S-phase and correlates with melanoma growth. BCL-2's role as an anti-apoptotic factor leads to sustained melanoma cell survival, by inhibiting apoptosis of mitochondria (80). The down regulation of BCL-2 with siRNA showed sensitization of melanoma cells resistant to apoptotic effects (81). MITF also regulates the transcription of tyrosinase (TYR), tyrosinase-related protein-1 (TYRP-1) and tyrosinase-related protein-2 (TYRP-2/DCT), which promotes the induction of melanin (82). The continual expression of MITF in melanoma cells results in melanocytes becoming melanoma, which exhibit uncontrolled cell proliferation, resistance to apoptosis and increase in cell pigmentation (melanin) (83).

The regulation of MITF protein production is considered to come from MITF 3'-UTR mRNA sequences which show high levels of conservation within vertebrate and longer in length than most mRNAs (84). The down regulation of *MITF* mRNA by miR-148 and miR-137 binding in the 3'-UTR has been demonstrated which further adds to the importance of the 3'-UTR in the regulation of MITF gene expression (84, 85). A mutation, loss of function or inability of miRNA to interact at the 3'-UTR of *MITF*

mRNA is a likely cause of melanomagenesis. Cell line and developmental studies have shown that miRNA-340 has the ability to regulate MITF at the 3'-UTR and this regulation can be interrupted and blocked in the presence of CRDBP (30).

Aside from melanoma, MITF has not been linked to any other cancers. MITF is part of a larger family of bHLHzip transcription factors with TFE3, TFEB and TFEC (86). Like Myc, as dimerizing transcription factors, TFE3 and TFEB have been linked to renal cell carcinoma, but MITF has only been shown to be associated with melanoma (87).

1.6 RNA Folding: Pros and Cons

The secondary structure of RNA is a result of base pairing of the nucleotides in canonical (A:U, U:A, G:C, C:G) and non-canonical (A:A, G:U) manner. The base pairings result in features such as stems, loops, bulges, and in combination, stem-loops, pseudoknots, and internal bulges.

The organization of RNA secondary structure into three-dimensional space gives rise to tertiary structure (Figure 5). Examples include double-helix formation in stems, coaxial base stacking between helices, and GAAA tetraloop interactions all of which can be divided into structural motifs with specific biological function.

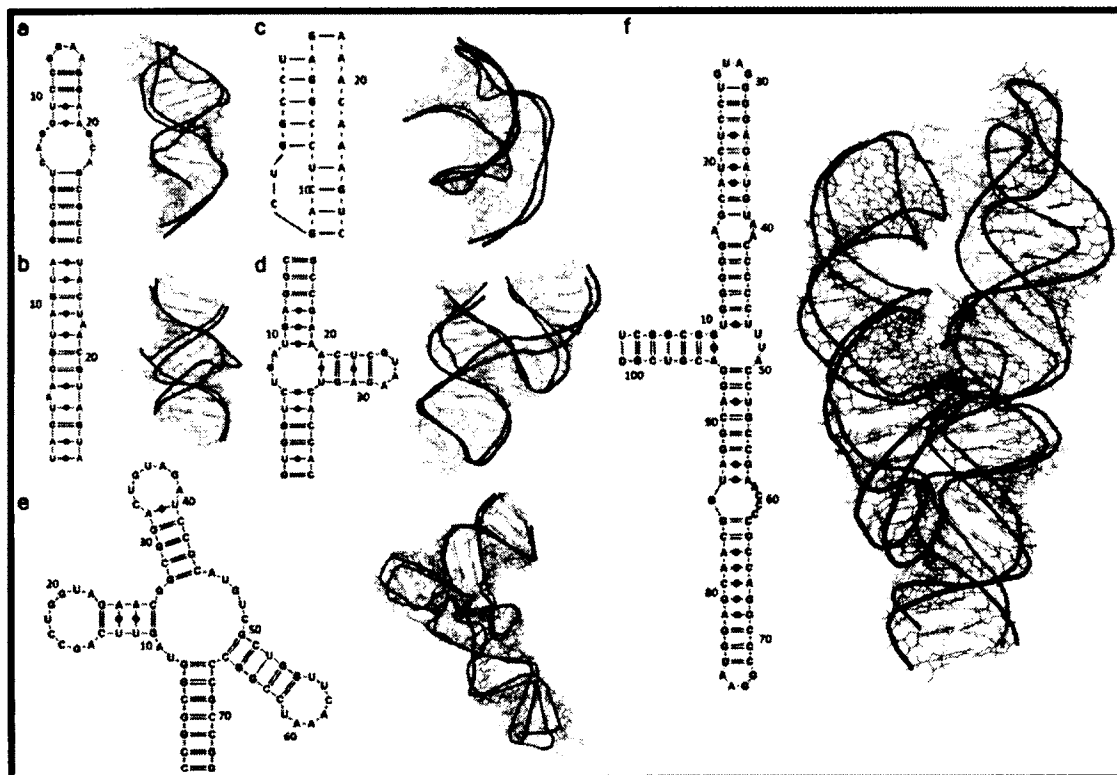


Figure 5. RNA Molecules and their predicted tertiary structures. Superimposed *in silico* predicted structures (blue) are superimposed onto experimental structures (gold). (a) a hairpin with internal loop (28SP), (b) a duplex with two bulges (119X), (c) the pseudoknot 1KPZ, (d) Hammerhead ribozyme RNA 1NYI, (e) tRNA 1J1U and (f) SRP RNA 1Z43. Adaptation from Zhao *et al.* 2012 (4).

The ribosome and spliceosome are examples of functional quaternary RNA structures, where several RNAs come together to perform a specific task. Investigating secondary and tertiary structure will be most beneficial for identifying common structural features, which may associate with protein. Determining quaternary structure will not be necessary, as quaternary structure is important for active RNA function such as ribozymes as opposed to structural binding features.

How RNA functions and is able to be bound by protein is dependent on its structure, and by gaining knowledge of RNA structure, function can be assigned. The binding of distantly related RNAs by protein suggests that it is not the primary structure

that defines the RNAs ability to be bound by protein, but rather their secondary and tertiary structure (88).

Determining RNA secondary and tertiary structure can be divided into the categories of computational analysis and experimental probing. To increase the accuracy of RNA secondary structure by computational prediction to identify motifs for binding to protein domains, it is important to assess different RNAs that demonstrate similar function experimentally. This eliminates the possibility of primary structure as the feature for binding, and reinforces the notion that secondary and tertiary structure as the mechanism for binding. It is also worth noting that the single stranded regions of the RNA secondary structure is where specific RNA-protein interactions are more likely to occur, because those regions are free to interact with the protein. Three most common ways to predict RNA secondary structure *in silico* are, (i) “single sequence secondary structure prediction”, (ii) “align-then-fold” and (iii) “simultaneously folding and alignment” (89).

The “single sequence secondary prediction” model relies on the thermodynamic stability of the overall secondary structure by maximizing the number of base pairs and scoring each structure by the minimum free energy (MFE) associated (90). This method uses the Zuker-Steigler algorithm for generating the MFE and is used in programs such as MFOLD/UNAFold and RNAfold (91-94). Possible improvements in secondary structure prediction can come from a dynamic algorithm, *contraFOLD*, which uses experimentally determine parameters to determine the MFE in combination with probabilistic models (95). *Sfold* uses a Boltzmann distribution to assign probabilities of representative secondary structures, and then clusters similar motifs into ensembles (96).

The MFE might not always be included, but *Sfold* allows manual addition of RNA structures (89). *RNAkinetics* provides ensembles of structures with assigned probabilities by taking into account how RNA comes into existence as an elongating strand (97). The program calculates secondary structure as the RNA sequence is growing to demonstrate how RNA is dynamic (89). *Vienna+P* calculates RNA secondary structure with a model that simulates the presence of general single stranded RNA binding proteins (93, 98). This model tries to account for *in vivo* protein interactions, which could alter the secondary structure of the RNA. These secondary structure prediction models, from a single sequence are a good starting point as an initial indicator of possible homologous RNA secondary structure (89).

Being able to compare multiple RNA sequences binding to the same protein can be accomplished with the “align-then-fold” and “simultaneously folding and alignment” methods. Sequences aligned with higher than 55% sequence identity produce consistent results with the “align-then-fold” method compared to the “simultaneously folding and alignment” method which performs better with sequence similarity of 50-60% or lower (89). When performing “simultaneously folding and alignment” algorithms on a set of sequences aligning below 50-60%, it is recommended that additional structural information is included when running the program to achieve accurate alignments (89).

QRNA is an “align-then-fold” algorithm using a hidden Markov based model to assess non-coding gene segments (99). By aligning two sequences with synonymous substitutions, the model identifies a coding region, and compares regions to identify conserved RNA secondary structure such as the presence of stems (99). *RNAz* utilizes an MFE algorithm in conjunction with known RNA motifs in sets of aligned sequences

(100). A support vector machine (pattern recognition algorithm) is used to assign a conservation of secondary structure index, and allows for genome-wide searches of similar sequences (89, 100). The *QRNA* and *RNAz* “align-then-fold” methods assumes the aligned sequences are evolutionary related or homologous sequences, but this will not be the case when assessing RNA sequences of different genes which possibly possess similar secondary structure (89). *CMFinder* is the most advantageous model, finding motifs in sets of unrelated RNA sequences using covariance models (101).

Simultaneous folding and alignment methods are applicable for set of sequences below 55-60% similar identity using the Sankoff algorithm (102). *FOLDALIGN* scans multiple sequences for local and global similarities in secondary structure after simultaneous alignment and prediction of secondary structure (103). Using pairwise connection, *CARNAC*, attempts to predict conserved secondary structure of homologous non-coding RNAs in three steps (104). Firstly *CARNAC* predicts all the possible stem structures, then compares those in a pairwise fashion to extract the most related structures and compiled them into stem graphs (104). Two unaligned RNA sequences secondary structure can also be aligned, predicted in a MFE manner, and then additional experimentally characterized constraints included using *Dynalign* (105).

Prediction of RNA tertiary structure would provide a wealth of information for determining how the binding of protein is associated. Generating three-dimensional structure requires the incorporation of primary, secondary and tertiary information. Being able to distinguish similar features between various RNAs would provide a vast amount of insight into understanding RNA-protein interaction. *MC-FOLD/MC-SYM* pipeline predicts tertiary structure of RNA from primary sequence (89, 106). The *MCFOLD*

algorithm predicts secondary structure with all possible nucleotide cyclic motifs, which are then organized into building blocks (106). The building blocks are matched to a database of determined three-dimensional structures, and *MC-SYM* then compiles all of the blocks to generate three-dimensional model (106).

Table 1. Comparing computational algorithms for RNA structure prediction.

Mfold/UNAFold	MFE	Simple/fast Web browser	Accuracy 1 sequence
RNAfold	MFE	Simple/fast Web browser	Accuracy 1 sequence
contraFOLD	Probabilistic/Experimentally determined parameters	Dynamic model with increased accuracy	Accuracy 1 sequence
Sfold	Boltzmann statistical ensemble/MFE structures can be manually added	Dynamic model provides groupings of structures which would help identify protein binding sites	Accuracy 1 sequence
RNAkinetics	MFE of local sequences	Provides ensembles of structures with probabilities Accounts for growing RNA	Accuracy: RNA would fold as it is made, then introns are cut 1 sequence
Vienna+P	MFE in the presence of generalized ssRNA binding proteins	Calculates probability for each nucleotide to bind protein in the entire RNA	Does not account biological data: Known Protein domains or # of nucleotides to bind protein -2 sequence limit.
QRNA	Compares synonymous vs. compensatory mutations		-Assumes sequences are evolutionarily related.
RNAz	Averaged MFE structures vs. individual MFE structures	Provides structural conservation index	-Assumes sequences are evolutionarily related.
CMfinder	Covariance	Applicable to unrelated	

Dynalign	Model Simultaneous alignment and prediction	sequences Predictions can be constrained to experimentally determined base pairings Good for sequence below 55-60% similarity	2 sequence limit
FOLDALIGNM	Simultaneous alignment and prediction Folds RNA based on local and global alignments	Scans for common local motifs after folding >2 sequences Good for sequence below 55-60% similarity	
CARNAC	Simultaneous alignment and prediction	Good for sequence below 55-60% similarity	Assumes sequences are homologous
MC-FOLD/MC-SYM	Folds secondary structure into building blocks then compares blocks to determined 3D structures	Uses primary information to generate secondary and 3D structure with experimentally determined structural motifs	Database of 3D structures possibly limited

Folding biases due to the nature of the folding program exists. Which parameters are considered or not, and how these different parameters are weighted, all contribute to certain biased outcomes. These biases are then further exemplified when RNA sequences are truncated, as the full sequence has a potential to form different *in vivo* than a mere segment of it. Lastly, it is not known the exact dynamics of RNA and, some RNA sequences could be very fluid and easily manipulated to take on multiple different forms, all of them present at the same time.

1.7 Objectives of this Research

The goals of this thesis is to gain a better understanding of the molecular interaction between CRDBP and its target RNAs, *MITF* and *c-myc*, by examining: (i) the structure of target RNAs, and (ii) assessing the RNA-binding ability of KH variants of CRDBP *in vitro* and *in vivo*.

1.7.1 Characterizing the structure of RNA targets of CRDBP

The first aspect for understanding CRDBP's interaction with RNA is to gain knowledge of the RNA structure that facilitates binding to CRDBP. By mapping out smaller regions of RNA bound by CRDBP allows for a starting point for finding commonality in how CRDBP selects RNA. Using computer software to model secondary and tertiary structure of the mapped RNA regions, I am aiming to provide more accurate information as to how CRDBP interacts with its targets RNA. As an additional tool, I also used Circular Dichroism Spectroscopy to study *MITF* and *c-myc* RNA to examine RNA structure and their dynamics upon association with CRDBP.

1.7.2 Characterizing the RNA-binding ability of the KH domains of CRDBP

The structure of CRDBP will be characterized *in vitro* and in cells by assessing the RNA-binding profile of site-directed KH variants of CRDBP. *In vitro* characterization of CRDBP will consist of electromobility shift assays to assess the binding of various CRDBP KH variants to *c-myc* and *MITF* RNAs. The cell studies will be conducted by the expression and immunoprecipitation of CRDBP KH variants in HeLa cells, followed by quantification of *c-myc*, *CD44* and β -*actin* mRNAs by q-PCR. In addition, I attempted to use zebra fish model as an *in vivo* model to determine if defect in CRDBP-*MITF* RNA interaction can lead to defect in pigmentation.

Chapter 2- Characterizing the MITF, c-myc, MAPK4, and β -TrCP1 RNAs that bind CRDBP

The goal of this Chapter is to further characterize the RNAs that bind CRDBP. Specifically, the aim is to determine the smallest regions of several RNAs that still have affinity for CRDBP. Mapping the smaller regions of RNA fragments that bind CRDBP is important for several reasons. Firstly, the smaller regions of RNA molecules that bind CRDBP are of interest because they provide an opportunity to be analyzed *in silico*. It is important to point out that most *in silico* methods that predict the secondary and tertiary structures of RNA molecules require input of smaller sized RNA molecules. Secondly, the knowledge of smaller RNA binders of CRDBP permits development of high-throughput fluorescent anisotropy method to screen for inhibitors of CRDBP-RNA interactions. Thirdly, the knowledge of smaller RNA molecules that bind CRDBP have led to the discovery of specific oligonucleotides that inhibit CRDBP-RNA interaction *in vitro* and suppress specific gene expression in cells as demonstrated for CD44 and GLI1 RNAs (107, 108).

2.1 Methodology

2.1.1 Protein Purification of Recombinant CRDBP

2.1.1.1 Transformation

The appropriate pET28b plasmid (~100 ng) is transformed into BL21 competent *E. coli* (100 μ L). The BL21 competent *E. coli* 100 μ L aliquot is first thawed on ice from -80°C, and then incubated on ice with approximately 100 ng of corresponding plasmid for 30 minutes. The cells are then heat shocked for 90 seconds at 42°C and then incubated on ice for 1 minute. The cells are then incubated for 30 minutes at 37°C in 500

μ L LB broth and plated on LB-kanamycin agar 10 mL dishes. The LB-kanamycin agar plates are then incubated at 37°C for 16 hours.

2.1.1.2 Induction

From the LB-kanamycin plate, 20 colonies are picked, inoculated into 100 mL of LB-Kanamycin broth in a 250 mL flask and placed in an incubator shaker at 37°C with 200rpm for 3 hours. The 100 mL culture is then transfers into a 3 L flask, with 900 mL of LB-kanamycin and grown to an optical density of 0.5 (~2.5 hours). Upon reaching an optical density of 0.5, induction of protein was initiated with the addition of 0.3mM isopropyl- β -D-thiogalactopyranoside (IPTG) (Invitrogen), and incubated on a shaker for 6 hours, at 37°C and 200 rpm. The cells are then spun at 3000 rpm for 15 minutes at 4°C, with the supernatant removed and the pellets stored at -80°C overnight.

2.1.1.3 Purification

The cell pellets stored at -80°C were retrieved, thawed on ice, and resuspended in 12 mL of buffer B (Table 2). The solution was kept on ice and placed on a shaker for 1 hour of gentle mixing, followed with centrifugation at 13,200rpm, 4°C for 30 min. The cell lysate was then syringed filtered (0.45 μ m filter) and mixed with 1mL of nickel-NTA resin that had been washed with 5mL of H₂O and 5mL of buffer B (Table 2). The filtered-lysate and Ni-NTA mixture is then placed on a bench-top plate shaker with ice for 1 hour. The resin and lysate suspension was pipetted back into the column, washed with 5 mL of buffer B, C, and D, while collecting 1 mL fractions (Table 2). The protein was then eluted with 12 mL of buffer E and F, collected in 0.5mL fractions. All the collected fractions were then stored at -80°C, and the column was washed with 6 mL of 0.5M NaOH, 10 mL of H₂O, capped with 30% ethanol and stored at 4°C.

2.1.1.4 12% SDS-PAGE of Purified CRDBP Protein Fractions

A BioRad Mini-Protean System was used to prepare 0.75 mm 12% SDS-PAGE gels, and have them electrophoresed at 240 V in 1X Tris-Glycine running buffer (Table 2). Of selected fractions from each of the CRDBP variants, 16 μ L were boiled for 5 minutes with 4 μ L of 5X sample buffer and allowed to cool for 1 minute prior to being loaded (Table 2). Following electrophoresis, the gels were stained with Coomassie blue dye for 1 hour, destained overnight and imaged (Table 2).

2.1.2 Dialysis of Isolated CRDBP Protein

The isolated fractions of purified CRDBP protein contained 8 M urea, 0.1 M NaH_2PO_4 , 0.01 M Tris-Cl; pH 4.5/5.9 was dialyzed in a two-step gradient to remove urea, normalize pH and alter buffer concentrations (Table 2). Slide-A-Lyzer® MINI dialysis units (3,500 MWCO, from Thermo Scientific) were loaded with a maximum of 100 μ L (~60 ng) of CRDBP variant and placed in a beaker with 50 mL of dialysis buffer A for 26 hours at 4°C. The mini dialysis units are then transferred into dialysis buffer B for 2 hours at 4°C, and then transferred again into a new beaker of 200 mL dialysis buffer B for 2 hours at 4°C. The dialyzed CRDBP protein was then pipetted into a 1.5mL microcentrifuge tube, centrifuged at 13,200 rpm for 30 minutes at 4°C, and supernatant transferred to a new 1.5 mL microcentrifuge tube and stored at 4°C (centrifugation was to remove any large aggregates). The concentration of protein was determined using BCA™ Protein Assay Kit (Pierce, IL, U.S.A.).

2.1.3 Generation of Internally Radiolabelled RNA

Radiolabeled RNA was prepared using ^{32}P -UTP and corresponding cDNA templates. The template DNA is incubated with ^{32}P -UTP, 1X transcription buffer, 20

μmol DDT, 20U rRNasin, 1 μmol each of ATP, GTP, CTP, 10 μmol UTP and 30 U T7-RNA polymerase to a final volume of 20 μL for 30 minutes at 37°C. Next 10 μL of RQ1 (1 U/ μL , RNase free DNase) was added to each sample and incubated at 37°C for an additional 10 minutes. Then 10 μL of loading dye is added and electrophoresed at 25 mA for 1 hour on a 6% denaturing polyacrylamide gel (7 M Urea). The gel was imaged with a cyclone storage phosphor screen system, printed off and placed underneath the gel to excise the RNA transcripts. The excised transcripts are crushed with an eppendorf pestle and eluted in 400 μL of DEPC water for 10 minutes at 70°C. The elution is followed with removal of the gel using Performa® DTR Gel Filtration Cartridges (EdgeBio, MD, U.S.A) a standard phenol/chloroform extraction and an ethanol precipitation. The precipitated RNA is resuspended in 30 μL of ddH₂O and 2 μL is placed in 1 mL of scintillation fluid and the radioactivity is measured on a Hidex liquid scintillation analyzer in counts per minute (cpm), then stored at -20°C.

2.1.4 Electrophoretic Mobility Shift Assays (EMSA)

Prior to each shift assay, the required EMSA binding buffer was prepared on ice, and the EMSA reaction mixture was pipetted. Internally radiolabeled RNA samples were refolded by heating each sample to 50°C for 5 minutes and cooling at room temperature for 7 minutes, and then 2 μL of RNA was added to each of the reaction tubes. The dialyzed CRDBP protein was then incubated in the EMSA binding buffer containing the internally radiolabeled RNA alternating twice between 35°C for 10 minutes, followed incubation on ice for 5 minutes. The binding reaction had 2 μL of loading dye added and 15 μL loaded onto a 4% non-denaturing polyacrylamide gel electrophoresed at 25 mA for 55 minutes (Table 2). The gel was then transferred onto chromatography paper, wrapped

in plastic-wrap, and exposed overnight on a phosphor imager screen. Cyclone Storage Phosphor Screen System and Optiquant software was used to image the exposed phosphor screen and generate the autoradiographs.

Table 2. Reagents used in the purification, analysis, dialysis and EMSA protocols of 6xHIS CRDBP.

Reagent Identity	Reagent Composition
Buffer B (Wash)	8 M Urea, 0.1 M NaH ₂ PO ₄ , 0.01 M Tris-Cl; pH 6.5.
Buffer C (Wash)	8 M Urea, 0.1 M NaH ₂ PO ₄ , 0.01 M Tris-Cl; pH 6.5.
Buffer D (Wash)	8 M Urea, 0.1 M NaH ₂ PO ₄ , 0.01 M Tris-Cl; pH 6.5.
Buffer E (Elution)	8 M Urea, 0.1 M NaH ₂ PO ₄ , 0.01 M Tris-Cl; pH 5.9.
12% SDS-PAGE Lower gel	12% Acrylamide, 0.3% Bis-acrylamide, 1X Lower gel buffer, 0.04% Ammonium per sulfate, 0.06% TEMED.
Lower gel buffer	18% (w/v) Tris-base, 0.4% (w/v) Sodium Dodecyl Sulfate, pH 8.8.
Tris-Glycine Running Buffer	0.3% (w/v) Tris-base, 1.44% (w/v) Glycine, 0.1% Sodium Dodecyl Sulfate.
Coomassie Blue dye	0.1% (w/v) Coomassie Blue.
Dialysis Buffer A*	2 M Urea, 1 mM L-glutathione reduced, 0.1 mM L-glutathione oxidized.
Dialysis Buffer B	2 M Urea, 1 mM L-glutathione reduced, 0.1 mM L-glutathione oxidized.
Loading Dye	9 M Urea, 0.01% Bromophenol Blue, 0.01% Xylene cyanole FF and 0.01% Phenol.
Elution Buffer	8 M Urea, 0.1 M NaH ₂ PO ₄ , 0.01 M Tris-Cl; pH 5.9.

10x Glycogen	0.2 mg/ml Glycogen and 3M Sodium Acetate (pH 5.2).
EMSA Binding Buffer	50 mM Tris-Cl (pH 7.4), 12.5 mM EDTA (pH 8.0), 10 mM DDT, 25% Glycerol, and 0.01% Triton X-100.
EMSA Loading Dye	250 mM Tris-Cl (pH 7.4), 0.2% Bromophenol Blue, 0.2% Xylene cyanol, and 40% Sucrose.

* Each Dialysis buffer contains a solution of 200 mM sodium chloride (NaCl), 20 mM Tris (pH 8.0), 10% (v/v) glycerol, 0.01% (v/v) Triton X-100, pH 7.4

2.2 Results

All four RNA fragments, MITF, c-myc, MAPK, and β -TrCP1 and their regions of interest were chosen based on findings reported in the literature. The goal was then to find the smallest fragments of each RNA that could still bind to CRDBP. However, the first step was to generate DNA fragments corresponding to the desired RNA fragments using PCR. The primers used in the PCR encoded the T7 promoter region are shown in Tables 7 to 10. The amplified DNA templates were then used for ^{32}P -UTP *in vitro* transcription reaction to generate the radiolabelled RNA for use in the EMSAs.

Table 3. List of primers used for PCR to generate MITF DNA templates for *in vitro* transcription.

Primer	Primer Sequence
MITF Forward 1338	5' <u>GGA TCC TAA TAC GAC TCA CTA TAG G</u> ACG CAA ATA AGT GTG 3'
MITF Forward 1550	5' <u>GGA TCC TAA TAC GAC TCA CTA TAG G</u> ATT GTA CAA ATA AGT GTG 3'
MITF Reverse 1580	5' TAC TTG CCA GCC ATT TGT 3'
MITF Reverse 1740	5' TGG TTG CTT TAC ATT CTT 3'
MITF Forward 1671	5' <u>GGA TCC TAA TAC GAC TCA CTA TAG G</u> TTT GCG TCA GAG AAA TGT 3'
MITF Forward 1621	5' <u>GGA TCC TAA TAC GAC TCA CTA TAG G</u> TTT GCG TCA GAG AAA TGT 3'
MITF Forward 1701	5' <u>GGA TCC TAA TAC GAC TCA CTA TAG G</u> ATC AAA TGT AAA GTT TAA 3'
MITF Reverse 1619	5' TCC TCT TGC TCA GAA AGC 3'
MITF Reverse 1709	5' ACA TTT GAT TTC CAA GGA 3'

Underlined is the SP7 promoter sequence.

Table 4. List of primers used for PCR to generate c-myc DNA templates for *in vitro* transcription.

Primer	Primer Sequence
c-myc Forward 1735	5' <u>GGA TCC TAA TAC GAC TCA CTA TAG G</u> AGG CCC CCA AGG TAG TTA 3'
c-myc Forward 1815	5' <u>GGA TCC TAA TAC GAC TCA CTA TAG G</u> GAA GAG GAC TTG TTG CGG 3'
c-myc Reverse 1744	5' TTG GGG GCC TTT TCA TTG 3'
c-myc Reverse 1814	5' AGA AAT GAG CTT TTG CTC 3'
c-myc Reverse 1886	5' CGC ACA AGA GTT CCG TAG 3'

Underlined is the SP7 promoter sequence.

Table 5. List of primers used for PCR to generate MAPK DNA templates for *in vitro* transcription.

Primer	Primer Sequence
MAPK Forward 3071	5' <u>GGA TCC TAA TAC GAC TCA CTA TAG G</u> TCT TGC AGA GGT AGC TCC 3'
MAPK Forward 4368	5' <u>GGA TCC TAA TAC GAC TCA CTA TAG G</u> GCA AAC TAT AAT TTG TAC 3'
MAPK Reverse 3754	5' GTC TGG AGA GGC GCT GTG 3'
MAPK Reverse 3424	5' ATG CTT GGT GAC TAG GAG 3'
MAPK Reverse 4386	5' TGA TTA GAT TTA AGG TGC 3'

Underlined is the SP7 promoter sequence.

Table 6. List of primers used for PCR to generate β -TrCP1 DNA templates for *in vitro* transcription.

Primer	Primer Sequence
β -TrCP1 Forward 661	5' <u>GGA TCC TAA TAC GAC TCA CTA TAG G</u> ACC TGA AGC CCA TGT TGC 3'
β -TrCP1 Forward 1021	5' <u>GGA TCC TAA TAC GAC TCA CTA TAG G</u> TGC AGA GGA TTC AGT GCC 3'
β -TrCP1 Reverse 678	5' GCA ACA TGG GCT TCA GGT 3'
β -TrCP1 Reverse 1040	5' GCG GCA CTG AAT CCT CTG 3'
β -TrCP1 Reverse 1372	5' ATG TCC CAC ACA GCA ATG 3'
β -TrCP1 Forward 1091	5' <u>GGA TCC TAA TAC GAC TCA CTA TAG G</u> TAT CAG TGG CCT ACG AGA 3'
β -TrCP1 Reverse 1190	5' CAG ACA GAG GAC AGA GCC 3'
β -TrCP1 Reverse 1110	5' TAT CTC GTA GGC CAC TGA 3'

Underlined is the SP7 promoter sequence.

2.2.1 MITF

Determining the smallest region of MITF RNA best bound by CRDBP began with a literature search which indicated that CRDBP binds MITF mRNA within the 3'-UTR, which spans nucleotides 1261-1822 (30). Nucleotide (nt) 1330 was chosen as the start site because Goswami *et al* has noted miRNA-340 first binding site, which spans nts 1444 to 1464 (marked in bold in Figure 2D of Goswami *et al*), and hence the start of nt 1330 would encompass the miRNA binding sites (30). The 3'-UTR region of MITF RNA from nts 1330 to 1740 were selected for RNA mapping studies. The two miRNA-340 and miR-548c-3p sites located close to the 3'-end of the 3'-UTR are critical as these are

the only sites likely to be present in both melanocytes and melanoma. The miRNAs are expected to be blocked from binding to the MITF RNA due to binding of CRDBP (30). Goswami *et al* noted that the shorter RNA isoforms are present in melanoma cells, hence nt 1740 was chosen as the end point of the MITF RNA in the RNA mapping studies (30).

The MITF region of 1330-1740 was initially divided into two regions, 1330-1550 (211nts) and 1550-1740 (191nts). The two portions of MITF's 3'-UTR (1330-1550 and 1550-1740) were compared alongside with c-myc RNA nts 1705-1886 which served as a positive control. From Figure 6, it is clear that both MITF RNA fragments bind CRDBP. It is also clear that MITF nts 1550-1740 had a greater affinity (Figure 6).

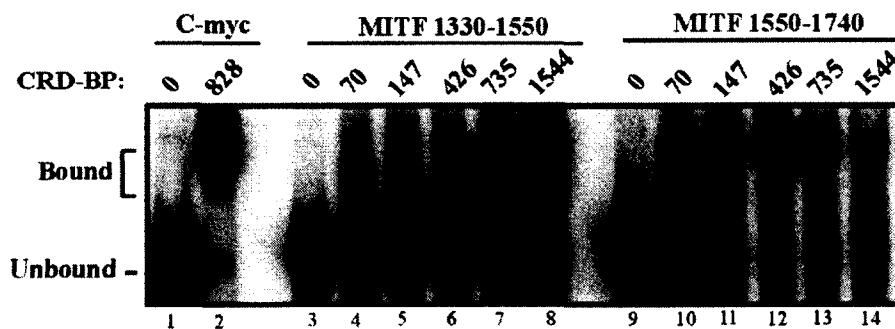


Figure 6. Electrophoretic Mobility Shift Assay of recombinant CRDBP with MITF RNA fragments (1330-1550 and 1550-1740) and c-myc 1705-1886. [³²P] RNA (20,000 cpm/reaction) incubated with varying concentrations of CRDBP as shown. Samples were electrophoresed in a 4% native polyacrylamide followed by autoradiography.

The higher affinity 1550-1740 MITF fragment was then further divided into eight fragments of decreasing length from the 3' and 5'-end (Table 7) using the PCR amplification. The PCR-amplified DNA templates were then used in the *in vitro* transcription reaction to produce the radiolabelled RNA fragments for use in EMSA binding assays (Figure 7).

Table 7. MITF RNA fragments assessed for binding with recombinant WT-CRDBP

RNA Fragment Identity	RNA Fragment Length	RNA Fragment Nucleotide Composition				RNA Fragment's affinity for CRDBP ^a
		A	C	G	U	
MITF 1330-1550	221nts	62	43	28	88	+++
MITF 1550-1709	160nts	45	25	34	56	++++
MITF 1550-1619	70nts	23	12	15	20	-
MITF 1581-1740	152nts	50	25	32	53	++++
MITF 1671-1740	72nts	26	9	13	22	+/-

^aData is from Figures 8 to 11.

(++++), (+++), (++) and (+) Indicate ~90%, ~70%, ~40% and ~10% binding respectively. (+/-) and (-) Indicate weak and no protein-RNA complex was visualized.

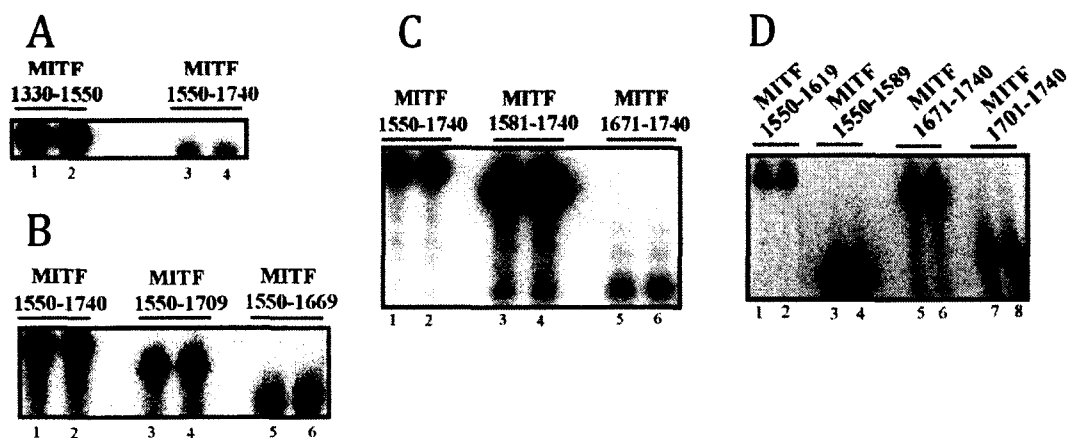


Figure 7. *In vitro* transcription of ³²P-UTP labeled MITF RNAs. PCR-amplified DNAs were used as templates for *in vitro* transcription to generate MITF RNA fragments. (A) nts 1330-1550 and 1550-1740. (B) 1550-1740, 1581-1740, and 1671-1740. (C) 1550-1619, 1550-1589, 1671-1740, and 1701-1740. (D) 1550-1740, 1550-1709, and 1550-1669. RNA samples were then electrophoresed in a 6% denaturing polyacrylamide gel followed by autoradiography.

From EMSAs of these MITF RNA fragments, nts 1550-1709, 1550-1669, 1581-1740, and 1621-1740 all showed binding affinity for CRDBP (Figure 8, 10 and 11). In contrast, MITF RNA fragments 1550-1619, 1550-1589, and 1701-1740 did not exhibit

binding affinity for CRDBP (Figures 9 and 11). MITF fragment 1671-1740 showed weak affinity for CRDBP (Figure 11).

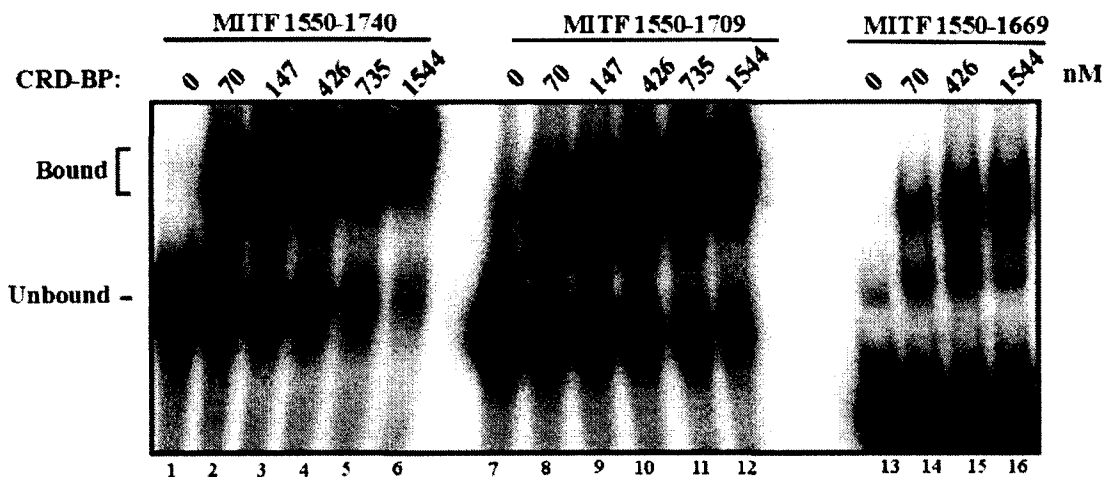


Figure 8. Electrophoretic Mobility Shift Assay of recombinant CRDBP with MITF RNA fragments. [32 P] MITF RNAs (20,000 cpm) nts 1550-1740, 1550-1709, and 1550-1669 were incubated with varying concentrations of CRDBP as shown. Samples were electrophoresed in a 4% native polyacrylamide followed by autoradiography.

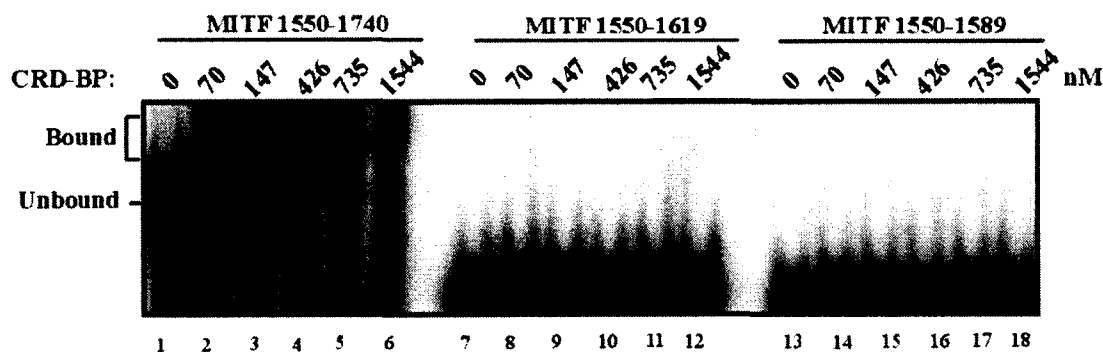


Figure 9. Electrophoretic Mobility Shift Assay of recombinant CRDBP with MITF RNA fragments. [32 P] MITF RNAs (20,000 cpm) nts 1550-1740, 1550-1619, and 1550-1589 were incubated with varying concentrations of CRDBP as shown. Samples were electrophoresed in a 4% native polyacrylamide followed by autoradiography.

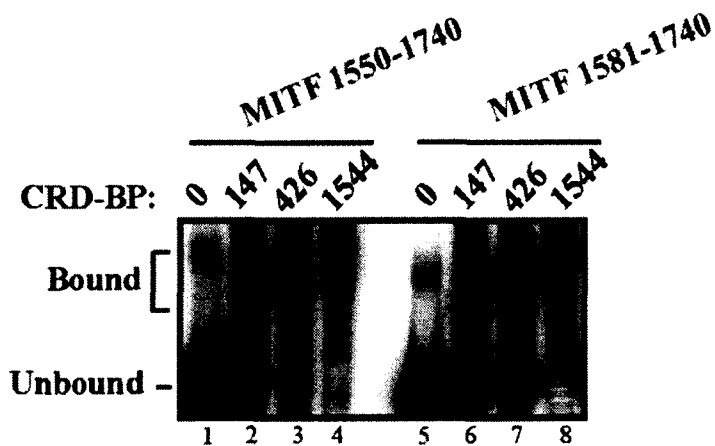


Figure 10. Electrophoretic Mobility Shift Assay of recombinant CRDBP with MITF RNA fragments. [32 P] MITF RNAs (20,000 cpm) nts 1550-1740 and 1581-1740 were incubated with varying concentrations of CRDBP as shown. Samples were electrophoresed in a 4% native polyacrylamide followed by autoradiography.

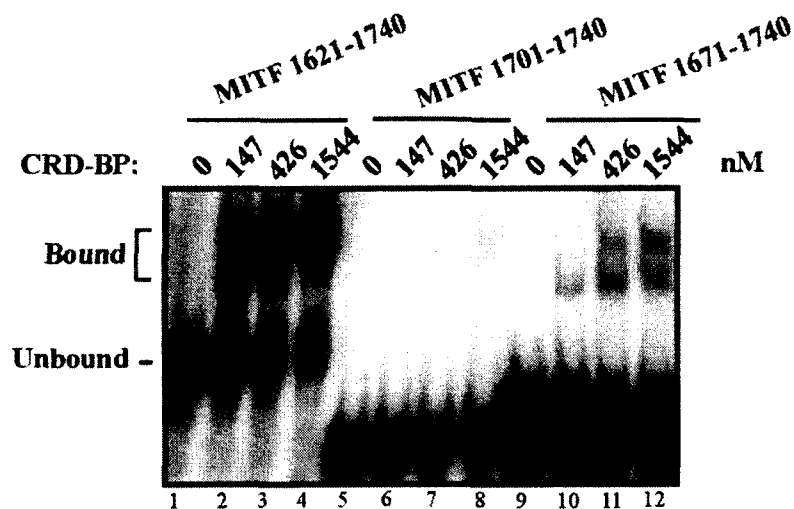


Figure 11. Electrophoretic Mobility Shift Assay of recombinant CRDBP with MITF RNA fragments. [32 P] MITF RNAs (20,000 cpm) nts 1621-1740, 1701-1740, and 1671-1740 were incubated with varying concentrations of CRDBP as shown. Samples were electrophoresed in a 4% native polyacrylamide followed by autoradiography.

			CRD-BP Binding
1550	————— 1740	191nts	++++
1550	————— 1709	160nts	++++
1550	————— 1669	120nts	++
1550	————— 1619	70nts	-
1550	——— 1589	40nts	-
1581	————— 1740	152nts	++++
	1621 ————— 1740	112nts	++++
	1671 ————— 1740	72nts	+/-
	1701 ————— 1740	42nts	-

Figure 12. Summary of MITF RNA fragments and relative binding affinity for CRDBP. The region of interest is shown in blue and the smaller fragments are shown in black. (++++), (+++), (++) and (+) indicate ~90%, ~70%, ~40% and ~10% binding respectively. (+/-) and (-) indicate weak and no protein-RNA complex respectively.

2.2.2 c-myc

The 182 nts c-myc fragment of nts 1705-1886 was originally shown by Doyle *et al* to be bound with high affinity to CRDBP (56). This region was even further narrowed down to 1705-1792, an 88 nt fragment (56). Additional research by Mr. Dan Sparanese in Dr. Lee's Lab also indicated that c-myc nts 1705-1886 as an excellent binding partner of CRDBP (61). This 1705-1886 CRD region of c-myc was therefore further truncated from the 5' and 3' ends into eight DNA fragments by PCR method. The PCR-amplified DNA templates were then *in-vitro* transcribed using radiolabelled UTP to produce the RNA fragments (Table 8, Figures 13 and 14).

Table 8. c-myc RNA fragments examined by EMSA for interaction with CRDBP

RNA Fragment Identity	RNA Fragment Length	RNA Fragment Nucleotide Composition				RNA Fragment's affinity for CRDBP ^a
		A	C	G	U	
c-myc 1705-1886	182nts	64	43	42	33	++++
c-myc 1705-1814	110nts	39	29	22	20	+++
c-myc 1705-1744	40nts	16	11	9	4	-
c-myc 1775-1886	112nts	36	26	28	22	+/-
c-myc 1845-1886	42nts	14	10	9	9	-

^aData is from Figures 15 to 19.

(++++), (+++), (++) and (+) Indicate ~90%, ~70%, ~40% and ~10% binding respectively. (+/-) and (-) Indicate weak and no protein-RNA complex was visualized.

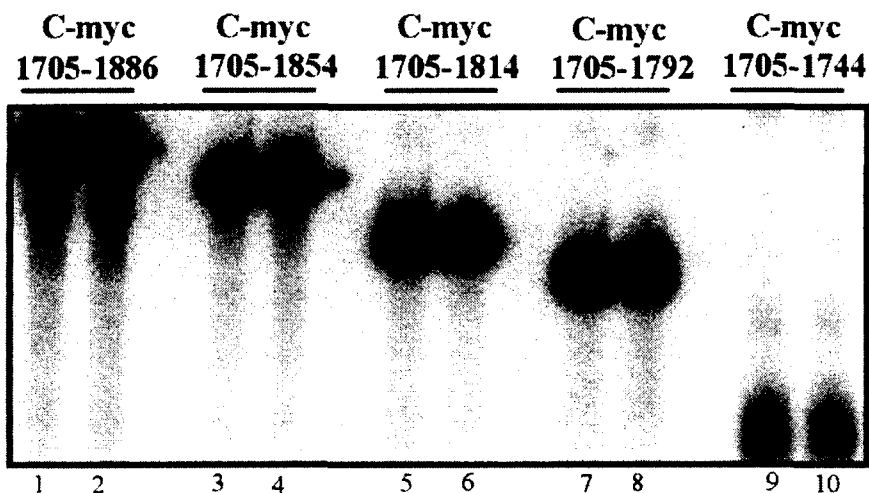


Figure 13. *In vitro* transcription of ³²P-UTP labeled c-myc RNA Fragments. PCR-amplified DNAs were used as templates for *in vitro* transcription to generate c-myc RNA fragments nts 1705-1886, 1705-1854, 1705-1814, 1735-1792, and 1705-1744. RNA samples were then electrophoresed in a 6% denaturing polyacrylamide gel followed by autoradiography.

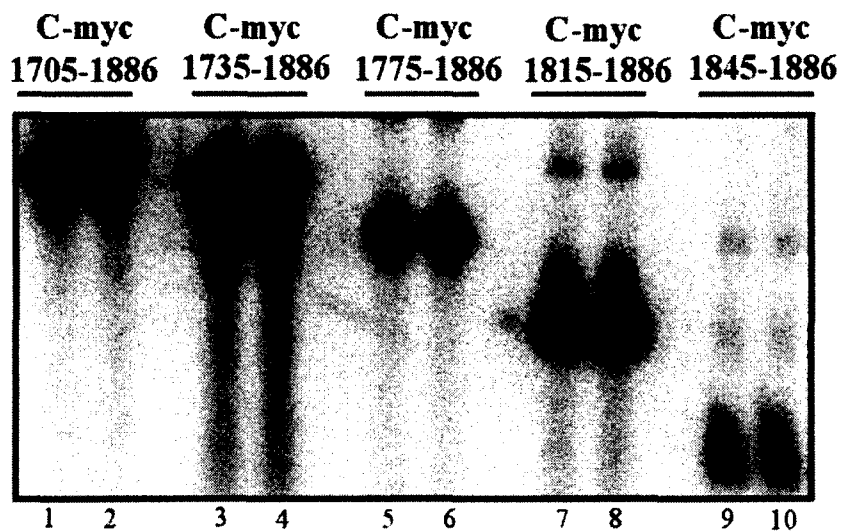


Figure 14. *In vitro* transcription of ^{32}P -UTP labeled c-myc RNA Fragments, PCR-amplified DNAs were used as templates for *in vitro* transcription to generate c-myc RNA fragments nts 1705-1886, 1735-1886, 1775-1886, 1815-1886, and 1845-1886. RNA samples were then electrophoresed in a 6% denaturing polyacrylamide gel followed by autoradiography.

The c-myc fragments of 1705-1854, 1705-1814, 1705-1792, and 1735-1886, all demonstrated high affinity binding with CRDBP in EMSA (Figures 15, 16, and 17). The c-myc RNA fragments 1735-1792, 1775-1886, 1815-1886, also exhibited binding with CRDBP although with much reduced affinity (Figures 16, 17 and 18). In contrast, c-myc RNA fragment 1705-1744 and 1845-1886 did not show any interaction with CRDBP (Figures 16 and 17).

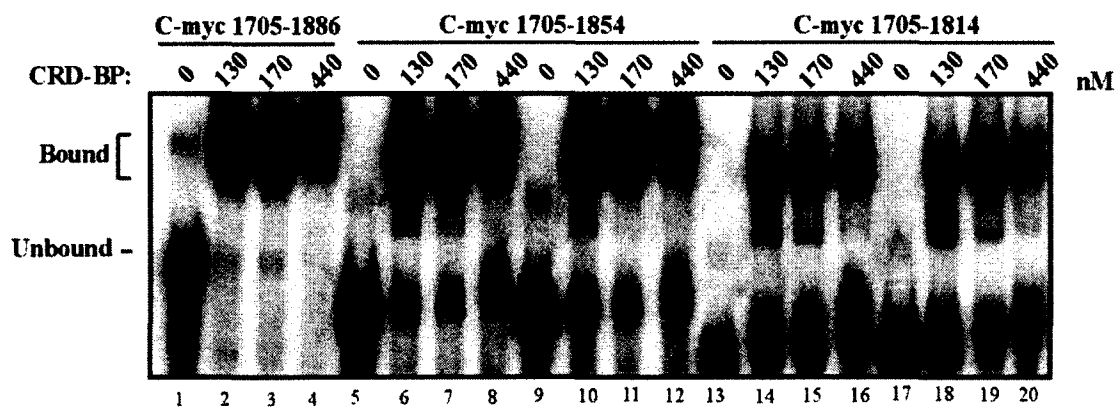


Figure 15. Electrophoretic Mobility Shift Assay of recombinant CRDBP with c-myc RNA fragments. [32 P] c-myc RNAs (20,000 cpm) nts 1705-1886, 1705-1854, and 1705-1814 were incubated with varying concentrations of CRDBP as shown. Samples were electrophoresed in a 4% native polyacrylamide followed by autoradiography.

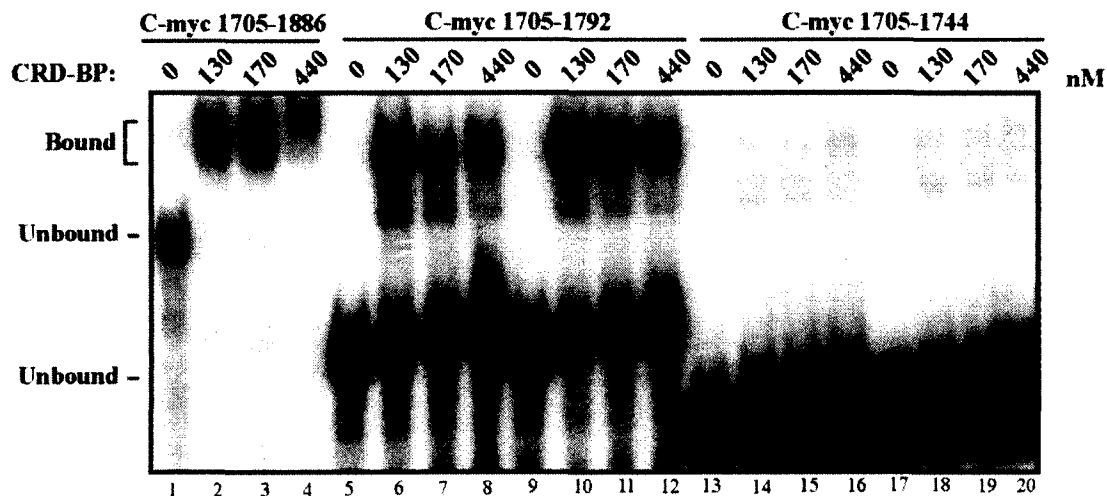


Figure 16. Electrophoretic Mobility Shift Assay of recombinant CRDBP with c-myc RNA fragments. [32 P] c-myc RNAs (20,000 cpm) nts 1705-1886, 1705-1792, and 1705-1744 were incubated with varying concentrations of CRDBP as shown. Samples were electrophoresed in a 4% native polyacrylamide followed by autoradiography.

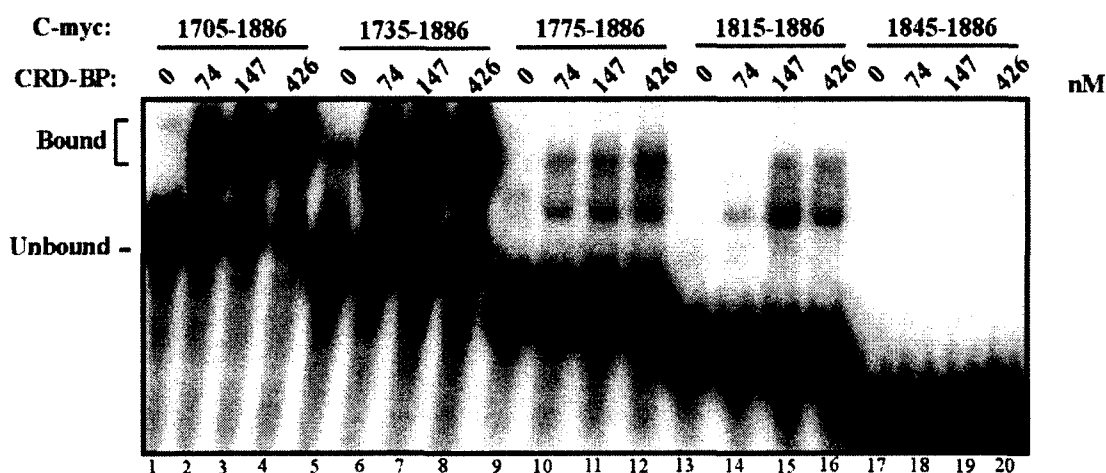


Figure 17. Electrophoretic Mobility Shift Assay of recombinant CRDBP with c-myc RNA fragments. [32 P] c-myc RNAs (20,000 cpm) nts 1705-1886, 1735-1886, 1775-1886, and 1845-1886 were incubated with varying concentrations of CRDBP as shown. Samples were electrophoresed in a 4% native polyacrylamide followed by autoradiography.

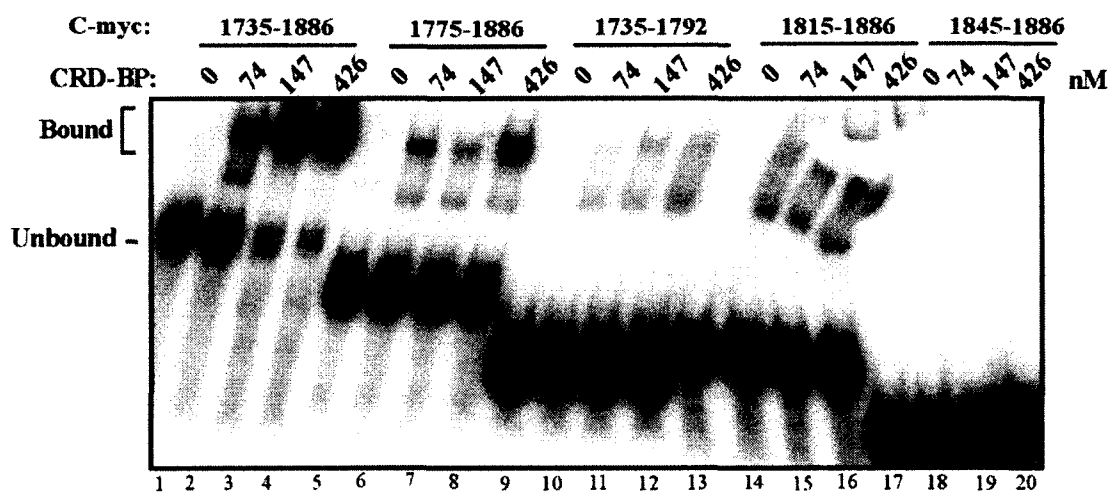


Figure 18. Electrophoretic Mobility Shift Assay of recombinant CRDBP with c-myc RNA fragments. [32 P] c-myc RNAs (20,000 cpm) nts 1735-1886, 1775-1886, 1735-1792, 1815-1886, and 1845-1886 were incubated with varying concentrations of CRDBP as shown. Samples were electrophoresed in a 4% native polyacrylamide followed by autoradiography.

An additional smaller c-myc RNA fragment nts 1735-1792 (58 nts) could also be synthesized. This fragment was truncated from both the 5'-end and 3'-end and has the possibility of interacting with CRDBP. Furthermore, analysis of this fragment would

help in narrowing down the smallest region of c-myc RNA that interacts with CRDBP.

EMSA of c-myc RNA fragment 1735-1792 indicated that this fragment was indeed capable of binding to CRDBP (Figure 19). Another RNA fragment smaller than the 182 nts c-myc fragment 1705-1886 that indicated binding to CRDBP was 1815-1886 (72 nts) (Figures 17 and 18).

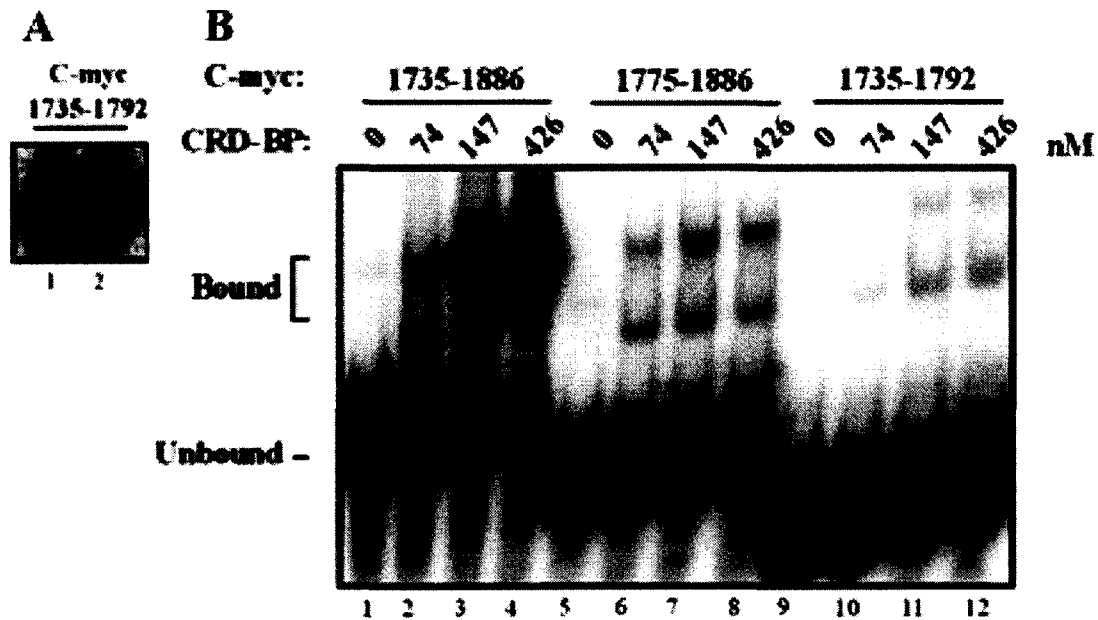


Figure 19. *In vitro* transcription of ^{32}P -UTP labeled c-myc RNA Fragment and Electrophoretic Mobility Shift Assay of recombinant CRDBP. (A) PCR-amplified DNA used as templates for *in vitro* transcription to generate c-myc RNA fragment nts 1735-1792. RNA samples were then electrophoresed in a 6% denaturing polyacrylamide gel followed by autoradiography. (B) [^{32}P] c-myc RNAs (20,000 cpm) nts 1735-1886, 1775-1886, and 1735-1792 were incubated with varying concentrations of CRDBP as shown. Samples were electrophoresed in a 4% native polyacrylamide followed by autoradiography.

			<u>CRD-BP Binding</u>
1705	————— 1886	182nts	++++
1705	————— 1854	150nts	++++
1705	————— 1814	110nts	+++
1705	————— 1792	88nts	+++
1705	——— 1744	40nts	-
1735	————— 1886	152nts	+++
	1775 ————— 1886	112nts	+/-
	1815 ————— 1886	72nts	+/-
	1845 ————— 1886	42nts	-
	1735 ————— 1792	58nts	+/-

Figure 20. Summary of c-myc RNA fragments and their relative binding affinity for CRDBP. The region of interest is shown in blue and the smaller fragments are shown in black. (++++), (+++), (++) and (+) indicate ~90%, ~70%, ~40% and ~10% binding respectively. (+/-) and (-) indicate weak and no protein-RNA complex was visualized.

2.2.3 MAPK4

Stöhr *et al* 2012 were the first to discover that IGF2BP1 (IMP-1) inhibited MAPK4 expression by interfering with mRNA translation upon associating with the MAPK4 3'-UTR (59). The MAPK4 nucleotide sequence was then identified on NCBI as Homo sapiens mitogen-activated protein kinase 4 (MAPK4) mRNA (NCBI Reference Sequence: NM_002747.3) and the 3'-UTR region spanning nucleotide 2765 to 4736, with a poly A signal starting at 4713. The 3'-UTR region for mapping MAPK4 RNA was then selected from nucleotide 2764 to 4713 (1949 nts). The MAPK 2764-4713 (1949 nts)

region was divided into six fragments of ~340 nts in length. RNA Fragments 3071-3424 (354 nts) and 3403-3754 (352 nts) could not be generated. This reason is not clear for either 3071-3424 (354 nts) or 3403-3754 (352 nts), but 3403-3754 (352 nts) does have a higher number of cytosine at 141 than the rest (Table 9). The remaining four RNA fragments, 2764-3089 (326 nts), 3723-4063 (341 nts), 4036-4386 (351 nts), and 4368-4713 (346 nts) were successfully transcribed (Table 9 and Figure 21).

Table 9. MAPK4 RNA fragments examined by EMSA for interaction with CRDBP.

RNA Fragment Identity	RNA Fragment Length	RNA Fragment Nucleotide Composition				RNA Fragment's affinity for CRDBP ^a
		A	C	G	U	
MAPK4 2764-3089	326nts	74	87	109	56	++++
MAPK4 3403-3754	352nts	73	141	59	79	N/A
MAPK4 4036-4386	351nts	88	102	85	76	++++

^aData is from Figure 22.

(++++), (+++), (++) and (+) Indicate ~90%, ~70%, ~40% and ~10% binding respectively. N/A indicates the binding affinity could not be assessed because the RNA fragments could not be generated.

MAPK 2764-3089 MAPK 3071-3424 MAPK 3403-3754 MAPK 3723-4063 MAPK 4036-4386 MAPK 4368-4713



Figure 21. *In vitro* transcription of ³²P-UTP labeled MAPK RNA Fragments. PCR-amplified DNAs were used as templates for *in vitro* transcription to generate MAPK RNA fragments nts 2764-3089, 3071-3424, 3403-3754, 3723-4063, 4036-4386 and 4386-4713. RNA samples were then electrophoresed in a 6% denaturing polyacrylamide gel followed by autoradiography.

The four successfully transcribed shorter RNA fragments of MAPK (2764-3089, 3724-4063, 4036-4386 and 4368-4713) were found to bind CRDBP effectively (Figure 22).

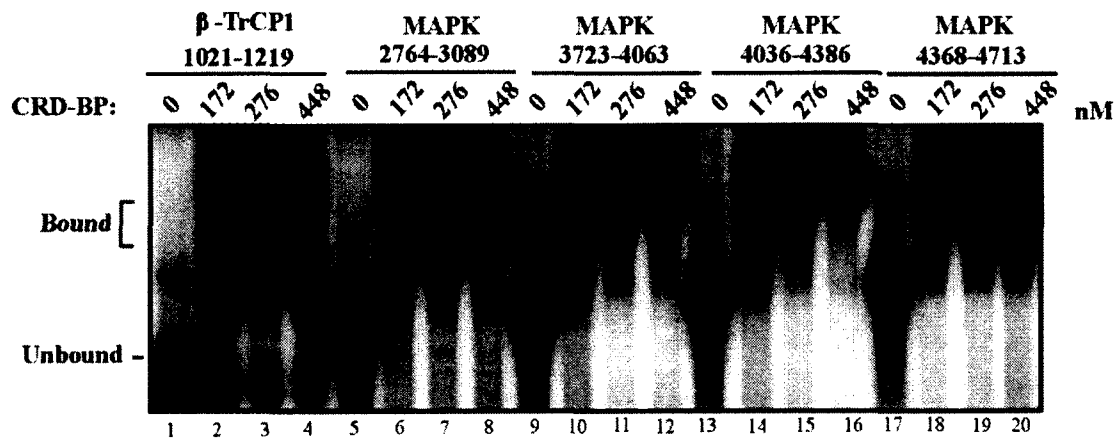


Figure 22. Electrophoretic Mobility Shift Assay of recombinant CRDBP with MAPK RNA fragments. [^{32}P] MAPK RNAs (20,000 cpm) nts 2764-3089, 3723-4063, 4036-4386, 4368-4713, and β -TrCP1 1021-1219 were incubated with varying concentrations of CRDBP as shown. Samples were electrophoresed in a 4% native polyacrylamide followed by autoradiography.

Fragment (nts)	CRDBP Binding
2764 ————— 4713	1950nts
2764 — 3089	326nts ++++
3071 — 3424	354nts N/A
3403 — 3754	352nts N/A
3723 — 4063	341nts ++++
4036 — 4386	351nts ++++
4368 — 4713	346nts ++++

Figure 23. Summary of MAPK RNA fragments and their relative binding affinity for CRDBP. The region of interest is shown in blue and the smaller fragments are shown in black. (++++), (+++), (++) and (+) indicate ~90%, ~70%, ~40% and ~10% binding respectively. N/A indicates the binding affinity could not be assessed.

2.2.4 β -TrCP1

Noubissi *et al* 2006 were the first to describe the ability of CRDBP to stabilize the 577-1182 (606 nts) region of β -TrCP1 RNA (60). Elcheva *et al* 2009 then further described the stabilization and protection of β -TrCP1 RNA (nts 577-1182) by CRDBP, regulating the RNA in miRNA-dependent manner by interrupting Argonaute2 and

miRNA-182 (29). Having established nucleotides 577-1182 as the important region of β -TrCP1 RNA for interaction with CRDBP, the regions of 510-1372 (863 nt) were then selected and split into five smaller regions to ensure full coverage (Table 10).

Table 10. β -TrCP1 RNA fragments examined by EMSA for interaction with CRDBP.

RNA Fragment Identity	RNA Fragment Length	RNA Fragment Nucleotide Composition				RNA Fragment's affinity for CRDBP ^a
		A	C	G	U	
β -TrCP1 510-678	169nts	59	29	33	48	+++
β -TrCP1 853-1040	188nts	61	40	43	44	+++
β -TrCP1 1203-1372	170nts	39	37	46	48	++
β -TrCP1 1021-1150	130nts	44	19	30	37	+++
β -TrCP1 1051-1219	169nts	52	27	41	49	++++
β -TrCP1 1131-1219	89nts	24	18	23	24	++

^aData is from Figures 26, 27, 29 and 30.

β -TrCP1 nts 510-678 and 661-862 were initially analyzed in Dr. Lee's lab (Mackedenski S, unpublished data). (++++), (+++), (++) and (+) Indicate ~90%, ~70%, ~40% and ~10% binding respectively. (+/-) and (-) Indicate weak and no protein-RNA complex was visualized.

β -TrCP1 RNA fragment 510-678 (169 nts) and 661-862 (202 nts) were previously analyzed in Dr. Lee's Lab to show that β -TrCP1 RNA fragment 510-678 (169 nts) bound best to CRDBP (Table 10) (Mackedenski S, unpublished data). The next step was to *in vitro* transcribe and compare the 510-678 (169 nts) β -TrCP1 RNA fragment with 853-1040 (188 nts), 1021-1219 (199 nts) and 1203-1372 (170 nts) (Figure 24).

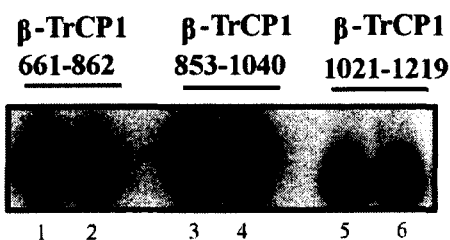


Figure 24. *In vitro* transcription of 32 P-UTP labeled β -TrCP1 RNA Fragments PCR-amplified DNAs were used as templates for *in vitro* transcription to generate β -TrCP1 RNA fragments nts 661-862, 853-1040, and 1021-1219 RNA samples were then electrophoresed in a 6% denaturing polyacrylamide gel followed by autoradiography.

β -TrCP1 RNA fragment 1021-1219 (199 nts) was found to bind better than 510-678 (169 nts), 853-1040 (188 nts) and 1203-1372 (170 nts) (Figures 25, 26 and 27).

			CRD-BP Binding
510	—————	1372	710nts
510	————— 678		169nts +++
	661 ——— 862		202nts ++
	853 ——— 1040		188nts +++
	1021 ——— 1219		199nts ++++
	1203 ——— 1372		170nts ++

Figure 25. Summary of β -TrCP1 RNA fragments and their relative binding affinity for CRDBP. The region of interest is shown in blue and the smaller fragments are shown in black. (++++), (+++), (++) and (+) Indicate ~90%, ~70%, ~40% and ~10% binding respectively.

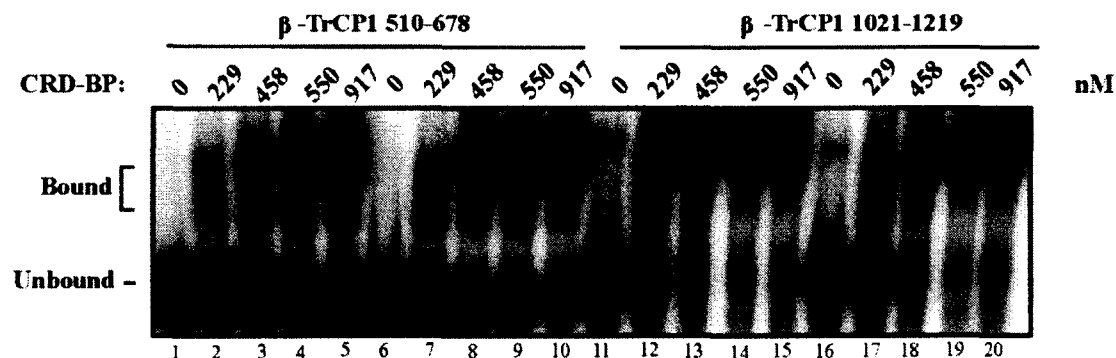


Figure 26. Electrophoretic Mobility Shift Assay of recombinant CRDBP with β -TrCP1 RNA fragments. [32 P] β -TrCP1 RNAs (20,000 cpm) nts 510-678, and 1021-1219 were incubated with varying concentrations of CRDBP as shown. Samples were electrophoresed in a 4% native polyacrylamide followed by autoradiography.

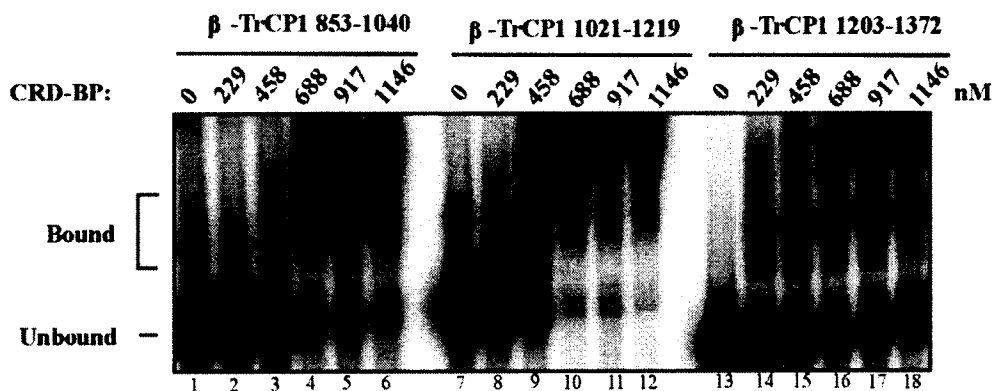


Figure 27. Electrophoretic Mobility Shift Assay of recombinant CRDBP with β -TrCP1 RNA fragments. [32 P] β -TrCP1 RNAs (20,000 cpm) nts 853-1040, 1021-1219, and 1203-1372 were incubated with varying concentrations of CRDBP as shown. Samples were electrophoresed in a 4% native polyacrylamide followed by autoradiography.

Since the 1021-1219 (199 nts) β -TrCP1 RNA fragment bound best, it was further divided and *in vitro* transcribed into six smaller fragments by shortening the RNA at both the 5' and 3' ends (Table 10 and Figure 28).

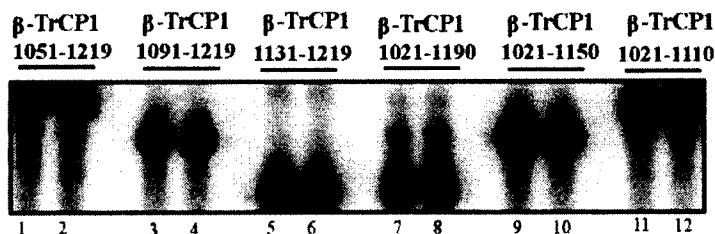


Figure 28. *In vitro* transcription of 32 P-UTP labeled β -TrCP1 RNA Fragments. PCR-amplified DNAs were used as templates for *in vitro* transcription to generate β -TrCP1 RNA fragments nts 1051-1219, 1091-1219, 1131-1219, 1021-1190, 1021-1150 and 1021-1110 RNA samples were then electrophoresed in a 6% denaturing polyacrylamide gel followed by autoradiography.

All the six smaller fragments tested, 1051-1219 (169 nts), 1091-1219 (129 nts), 1131-1219 (89 nts), 1021-1190 (170 nts), 1021-1150 (130 nts) and 1021-1110 (90 nts), were found to bind CRDBP with slightly varying degree of affinity (Figures 29, 30 and 31).

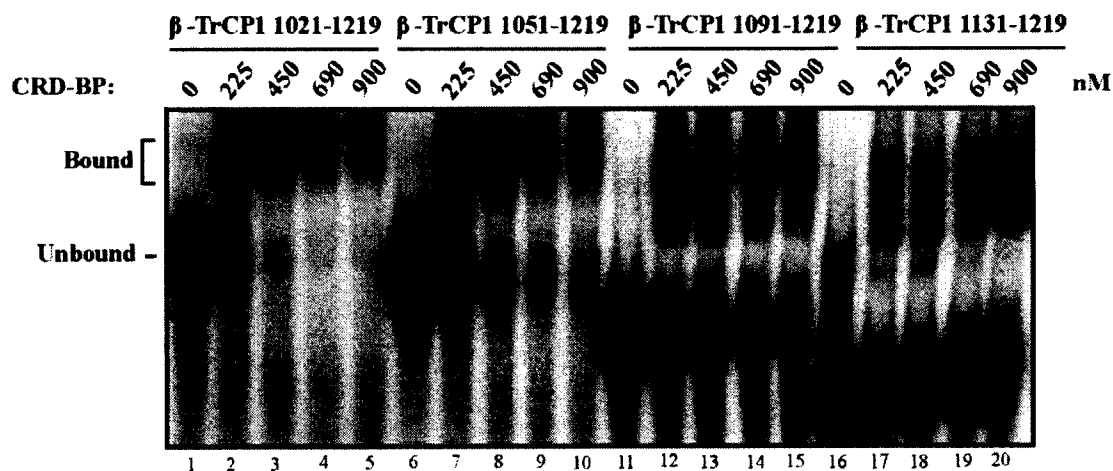


Figure 29. Electrophoretic Mobility Shift Assay of recombinant CRDBP with β -TrCP1 RNA fragments. [32 P] β -TrCP1 RNAs (20,000 cpm) nts 1021-1219, 1051-1219, 1091-1219 and 1131-1219 were incubated with varying concentrations of CRDBP as shown. Samples were electrophoresed in a 4% native polyacrylamide followed by autoradiography.

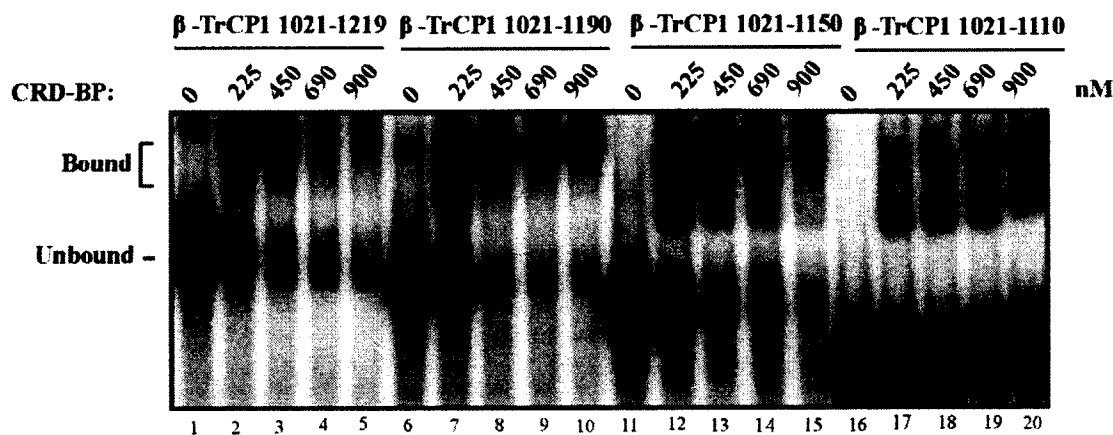


Figure 30. Electrophoretic Mobility Shift Assay of recombinant CRDBP with β -TrCP1 RNA fragments. Autoradiograph of β -TrCP1 RNA (20,000 cpm) incubated with varying concentrations of CRDBP after electrophoresis at 25mA for 55mins in a 4% polyacrylamide gel exposed for 8 hours. [32 P] β -TrCP1 RNAs (20,000 cpm) nts 1021-1219, 1021-1190, 1021-1150 and 1021-1110 were incubated with varying concentrations of CRDBP as shown. Samples were electrophoresed in a 4% native polyacrylamide followed by autoradiography.

			<u>CRD-BP Binding</u>
1021	————— 1219	199nts	++++
1021	————— 1190	170nts	++++
1021	————— 1150	130nts	+++
1021	————— 1110	90nts	+
1051	————— 1219	169nts	++++
	1091 ————— 1219	129nts	+++
	1131 ————— 1219	89nts	++

Figure 31. Summary of β -TrCP1 RNA fragments and their relative binding affinity for CRDBP. The region of interest is shown in blue and the smaller fragments are shown in black. (++++), (+++), (++) and (+) Indicate ~90%, ~70%, ~40% and ~10% binding respectively.

Discussion

In general, the RNA mapping data of MITF, c-myc and β -TrCP1 indicated that as the RNA region is decreased in size, so does the binding affinity. All the regions of MITF that retained binding affinity for CRDBP were 1550-1709, 1550-1669, 1581-1740, and 1621-1740, and this outlined a smaller region of MITF 1621-1709 (89 nts) that could interact with CRDBP. All these regions, including the smaller 1621-1709 (89 nts) region, outline a binding site for miR-548c-3p spanning 1658-1680 (Figure 2D from Goswami *et al*) (30). This suggests that CRDBP could bind at this position, making it a potential point for the regulation of MITF RNA. Two miR-340 binding sites that flank the MITF 1621-1709 region could also be interrupted when CRDBP binds in the MITF region of 1621-1709.

There was initial difficulty in the synthesis of MITF 1550-1740 in comparison to MITF 1330-1550 and other fragments due to the higher than usual number of repeated uracils (U's) in a row. To compensate for this, the normal *in vitro* transcription reaction was spiked with four times the usual amount of cold UTP, from 10 μ mol UTP to 40 μ mol UTP.

The mapping of c-myc RNA fragments indicated all regions of c-myc RNA bound to CRDBP except for 1845-1886 (42 nts) and 1705-1744 (40 nts) (Figures 16 and 17). The other two 40 nt regions of MITF, 1550-1589 and 1701-1740 also did not indicate any binding to CRDBP (Figure 9 and 11). Given that a 39 nt fragment of CD44 RNA had been established to be bound by CRDBP, it is unlikely that the two 40 nt fragments of c-myc and MITF RNAs that did not bind was due to an artifact based on their reduced size (109). The MITF RNA fragment 1550-1619 (70 nts) clearly showed no binding activity, which also encompasses the smaller 40 nt fragment, supporting the notion that the two smaller 40 nt fragments of c-myc and MITF are in fact incapable of binding CRDBP. From this, I conclude that the smallest region of c-myc RNA bound by CRDBP is 1735-1792 (58 nts), which lies within previously the reported region of 1705-1792 (88 nts) and 1705-1886 (182 nts) CRD instability region of c-myc (56, 61). In the neighbouring region 1594-1614 of c-myc, miR-185-3p has shown to bind to c-myc and regulate c-myc protein level (110). Binding of CRDBP to the CRD region (1705-1886 nts) could disrupt the ability of miR-185-3p to regulate c-myc RNA.

The smaller MAPK regions, which were transcribed all showed binding to CRDBP, suggesting that there may be multiple areas where CRDBP is able to bind with MAPK RNA.

The initial β -TrCP1 RNA region of 510-1372 (863 nts) that was split into five smaller fragments all exhibited binding affinity with CRDBP. Although they all indicated some affinity for CRDBP, fragment 1021-1219 (199 nts) had the greatest affinity for CRDBP.

From these mapping experiments, it is clear that there is some specificity for which RNA fragments can bind CRDBP. The mapping of these smaller regions of RNA could share similarities in sequence or structure. Being able to recognize these similarities could pave the way to the design of small molecules that mimic the RNA. The understanding of the structures associated with the RNAs mapped here, as well as the K-Ras, GLI1 and MDR1 RNAs mapped by other members in the Lee lab, will be discussed in Chapter 4.

Chapter 3- Analysis of CRDBP-RNA Interaction

Part A Characterizing the CRDBP-RNA interaction *in vitro* using Electrophoretic Mobility Shift Assay, in cells and *in vivo*

As described in Chapter 1, it is important to understand the function of CRDBP, notably in how it binds to its oncogenic target RNAs. Previous research using deletion studies of CRDBP orthologs have demonstrated that the KH domains are responsible for interacting with RNA (76, 111). In addition, previous work on an RNA binding protein NusA, which interacts through KH domains have demonstrated clearly that mutation at the G-X-X-G motif to D-X-X-G, renders the KH domain inactive for binding with RNA molecules (112). The aim of this section of Chapter 3 was to identify which of the KH domain(s) of CRDBP play a key role in interacting with RNA in the context of the entire protein. To do so, site-directed mutagenesis was executed to generate and purify recombinant CRDBP single and double KH variants (Barnes *et al* submitted). The same mutations were also introduced on pcDNA3-FLAG-CRDBP plasmid to express the CRDBP variants in HeLa cells and to assess their ability to physically interact with specific mRNA species. These variants were also sub-cloned into pSp64TNE vectors for the expression of CRDBP in zebrafish (*Danio rerio*) in an effort to assess their influence on the CRDBP-MITF RNA interaction and hence on pigmentation *in vivo*.

3.1 Methodology

3.1.1 Purification and Dialysis of WT-CRDBP and Variants

Purification of wild type and variant CRDBP proteins were performed in identical fashion as described in Chapter 2. The purified WT-CRDBP and variant proteins were dialyzed as described in Chapter 2. All the proteins were quantified using the BCA

protein assay kit (Catalog # PI23225, Thermo Scientific).

3.1.2 Radiolabeling of RNA and Electrophoretic Mobility Shift Assay (EMSA)

Radiolabeling of c-myc 1705-1886 (182 nts) and MITF 1550-1740 (191 nts) RNA fragments were performed as described in Chapter 2. The manner in which the radiolabeled RNA and dialyzed protein were assessed in EMSA is described in Chapter 2.

3.1.3 Quantification and Representation of EMSA

The autoradiographs of the EMSAs generated by the Cyclone Storage Phosphor Screen System and Optiquant software was used to determine the dissociation constant (K_d) of CRDBP with the [32 P] c-myc CRD 1705-1886 and [32 P] MITF 1550-1740 RNA substrates. Using the Optiquant Software, bound and unbound values were determined by calculating the total signal in each lane and measuring how much of that belonged to either bound or unbound states of the respective [32 P] RNA substrate (Figure 1). The Hill equation was then used to calculate the K_d using these bound and unbound values (Appendix, Figures S3 to S5).

3.1.4 Immuno-precipitation of CRDBP Variants from HeLa cells

Immuno-precipitation was performed using monoclonal ANTI-FLAG[®] M2 antibody (Sigma-Aldrich) to assess mRNA associated with the wild-type and variant CRDBP. Human cervical cancer cells, HeLa, purchased from American Tissue Culture Collection (ATCC) were cultured in Minimum Essential Media (Gibco) containing 10%

Fetal Bovine Serum and plated on 100 mm dishes at 1×10^5 cells/mL. After one day of growth on the 100 mm dishes, transfection was performed using Lipofectamine 2000 reagent (Invitrogen), Opti-MEM[®] (Gibco) and pcDNA3-FLAG plasmid containing, empty vector, wild-type and variant CRDBP. Two days of growth after transfection, the cells were lysed using TCLB-L buffer [TCLB (50 mM Tris, 130 mM NaCl, 625 μ M EDTA, 2.5% Triton x-100), 7 x TCLB-C (TCLB, complete tablet), VRC (800 μ M), DTT (400 μ M), RNasin (40 U/ μ L)]. The cells were incubated on ice for 5 minutes after addition of the TCLB-L buffer. The cells were transferred to an eppendorf tube on ice, after the use of a cell scraper to remove the cells from the plates. The lysed cells were then aspirated using a 26 gauge needle 5 times to break up the nuclei before incubation on ice for an additional 30 minutes.

During the 30-minute incubation on ice, Protein-G agarose beads used for the immuno-precipitation were equilibrated. First the beads are resuspended until homogenous, and 150 μ L of the homogenized beads were aliquot into Eppendorf tubes for each plate of cells. The beads were then centrifuged for 1 minute at 3000 x g, and the supernatant was removed. The beads were then resuspended in 500 μ L of TCLB-L buffer, centrifuged for 1 minute at 3000 x g, and the supernatant removed. This TCLB-L wash step was repeated once more and then the Protein-G agarose beads were resuspended in 75 μ L of TCLB-L buffer.

After the 30-minute incubation on ice, the lysed cells were centrifuged for 10 minutes at 16,100 x g at 4°C. The supernatant was collected in a new eppendorf in order to pre-clear the cell lysate. The cell lysate was pre-cleared by the addition of 50 μ L of the equilibrated Protein-G agarose beads and incubating on a shaker for a minimum of 1 hour

at 4°C before centrifuging the lysate and bead mixture at 3000 x g for 1 minute at 4°C.

Pre-clearing was performed one more time, for a total time of 3 hours of incubation. The pre-cleared cell lysate was then incubated with 5 µL of the monoclonal ANTI-FLAG[®] M2 antibody (Sigma-Aldrich) on shaker overnight at 4°C.

The next morning, the remaining equilibrated Protein-G agarose beads were added to the pre-cleared lysate and antibody mixture, and incubated on a shaker for four hours at 4°C. After the incubation, the lysate and bead mixture was centrifuged at 3000 x g for 1 minute at 4°C. The supernatant was removed, and the beads were washed four times with TCLB-W buffer [TCLB, DTT (400 µM), VRC (800 µM)] for 5 minute at 4°C each. Next the beads were washed five times with TCLB-U [TCLB, DTT (800 µM), VRC (1.60 mM), 0.46 M Urea] for 15 minutes at 4°C each. After each wash the bead and wash mixture was centrifuged at 3000 x g for 1 minute at 4°C. After resuspension of the beads on the fourth wash of the TCLB-U wash step, half of the resin was removed, centrifuged at 3000 x g for 1 minute, and re-suspended in 16 µL of H₂O and stored at -80°C to be used for Western blot analysis. The remaining beads were washed for a fifth time, supernatant removed, and incubated with 100 µL of TCLB-PK [500 µL TCLB-W, Proteinase K (150 µg), 0.1 % SDS] for 30 minutes at 50°C. Phenol/chloroform extraction was performed twice to extract the RNA, which was then precipitated at -80°C with the addition of 4 µL yeast tRNA (5 mg/ml), 24 µL 3 M sodium acetate and 500 µL 100% ethanol in a new eppendorf tube. The precipitated RNA was then resuspended in 8 µL of nuclease-free H₂O and DNase treated using a DNA-free kit (Ambion). The resuspended 8 µL RNA had 1 µL of 10X buffer and 1 µL of DNase1 added and incubated at 37°C for 30 minutes as instructed by the Ambion DNase treatment kit. Next, 2 µL of the

inactivating reagent was added to the mixture and incubated at room temperature for 2 minutes, and then centrifuged at 10,000 x g for 1.5 minutes. The purified RNA was then used as template to produce cDNA for real-time quantitative polymerase chain reaction (RT-qPCR) analysis.

3.1.5 RT-qPCR of RNA Associated with FLAG-CRDBP Variants

The BioRad iScript™ cDNA Synthesis Kit was used for cDNA synthesis. Equivalent amount of RNA were used for each cDNA synthesis reaction. Each reaction consisted of RNA (~1 ug), 4 uL of 5 x iScript mix, 1 uL of reverse transcriptase and H₂O (up to 20 uL) which was placed in the thermocycler and the first strand synthesis of cDNA was done using the following program: 25°C 5 minutes, 42°C 30 minutes, 85°C 5 minutes. The qPCR reaction mixtures (25 uL) then consisted of 2 uL cDNA, 12.5 uL iQ™ SYBR® Green Supermix (Bio-Rad), 1.0 uL forward primer (10 uM), 1.0 uL reverse primer (10 uM) and 9.5 uL nuclease-free water and were added to 96-well PCR plates (Axygen). The 96-well plates were placed in a BioRad iQ5 cycler, where the cDNA was amplified in a two-step PCR reaction. Optimal temperature for each primer set used (β -actin, c-myc and CD44 Table 1) were previously described in Dr. Lee's lab, using melt-curve analysis, prior to all qPCR experiments (109). The efficiency of the qPCR reactions were established using the standard curve method.

Table 11. Primer sets used in real-time RT-PCR experiments.

Target Gene	Primer Sequences
β -actin	Forward, 5'-CTCCTTCTGATGTCCTGGG-3' Reverse, 5'-GGGACGATGAGCCAGTAC-3'
c-myc	Forward, 5'-ACGAAACTTTGCCCATAGCA-3' Reverse, 5'-GCAAGGAGAGCCTTTCAGAG-3'
CD44	Forward, 5'-TATGATGTTGAGGCTGCTGCAATGC-3' Reverse, 5'-TGTGATGTTGAGGCTGCTGCAATGC-3'

3.1.6 *D. rerio* studies

3.1.6.1 Preparation of pSp64TNE-Plasmid

In order to *in vitro* transcribe capped mouse WT-CRDBP and CRDBP variant RNAs for injection into *Danio rerio*, the genes encoding the mouse WT-CRDBP and CRDBP variants needed to be first sub-cloned from pcDNA3 vector into pSp64TNE vector. This required gene amplification with the introduction of restriction sites to allow for ligation into pSp64TNE vector. Polymerase chain reaction was performed using 100 ng of pcDNA3 CRDBP gene of interest, 5X Phusion DNA polymerase buffer, 0.28 mM dNTPs, 2 Units of Phusion DNA Polymerase and 0.5 μ M primers (Table 2). The reactions were then placed in a thermocycler using the following program – 98 °C for 30 seconds, then 98°C-15 seconds, 55°C-30 seconds, 72°C-2 minutes for 25 cycles followed by incubation at 72°C for 10 minutes. The PCR product was analyzed on a 2% agarose gel followed by gel extraction using QIAEX® II Gel Extraction Kit. The gel extracted insert and pSp64TNE plasmid were double digested using ~500 ng of DNA, 10 Units of KpnI, 20 Units of EcoRI-HF, NEBuffer 2.1 and heated to 37°C for 2 hours. The digested vector and inserts were then analyzed on 2% agarose gel, and gel extracted using QIAEX® II Gel Extraction Kit. Next two ligation reactions were prepared using digested pSp64TNE and digested insert in a 1:1 and 2:1 (pSp64TNE:Insert) together with 10X ligation buffer, and 400 Units of Ligase. A control was prepared with 10X ligation buffer, 40 Units of Ligase and digested pSp64TNE. All ligation reactions were left at room temperature for 2 hours. Competent DH5 α *E. coli* cells were thawed from -80°C on ice, added to the ligation reactions and incubated on ice for 30 minutes. The DH5 α *E. coli* cells were then heat shocked at 42°C for 90 seconds, placed on ice for 2 minutes, 500

μ L of LB broth added and then incubated at 37°C for 1 hour. The transformed DH5 α *E. coli* cells were then transferred onto LB-Ampicillin agar plates and incubated overnight at 37°C. Colonies from both the 1:1 and 2:1 (pSp64TNE:Insert) were compared to the control (vector only ligation) and cultured in 3mL LB-Ampicillin broth overnight at 37°C. Liquid colony PCR was performed on the overnight cultures with 10X ThermoPol™ Reaction buffer, 0.28 mM dNTPs, 5 Units of DNA Taq Polymerase and 0.5 μ M primers (Table 1). The liquid colony PCR reactions were then placed in a thermocycler using the following program – 95 °C for 2 minutes, then 95°C-15 seconds, 55°C-15 seconds, 72°C-1 minute for 25 cycles followed by incubation at 72°C for 2 minutes. The PCR product was analyzed on a 1% agarose gel. DNA Mini-prep (QIAGEN) was performed on colonies with appropriate size bands visible on the 1% agarose gel and then sent for DNA sequencing for verification. A total of five pSp64TNE sub-clones were generated: WT-CRDBP, CRDBP-KH1, CRDBP-KH4, CRDBP-KH1-2 and CRDBP-KH3-4.

Table 12. Primers used in the gene amplification of WT-CRDBP and Variants.

Primer	Primer Sequence		
CRDBP EcoRI Reverse	5' ACCA Handle	GAATTC EcoRI Site	TAC CTT CCT CCG AGC CTG 3' Reverse Complement of CRDBP Stop

3.1.7 Circular Dichroism Conditions and Settings

All samples were scanned in triplicate from 200 to 320 nm at a rate of 20 nm/min using the Jasco J-815 Circular Dichroism Spectrometer. The raw scanned data were transferred from the Jasco Spectral Manager software into an excel spreadsheet and

converted from mdeg to molar ellipticity per residue as shown Figure 32. The samples analyzed by circular dichroism (Table 13) were either diluted in nuclease-free water or dialysis buffer as described in Table 2. All samples were blanked with a matching reference sample containing either nuclease-free water or dialysis.

$$\Delta\epsilon = \frac{\theta}{32,980 \times C \times L \times N}$$

θ = Raw circular dichroism amplitude in mdeg

C = Sample concentration in mol/L

L = Cell path length in cm

N = Number of residues in the molecule

Figure 32. The equation for calculating molar ellipticity per residue from measured mdeg.

Table 13. RNA Molecules analyzed by Circular Dichroism.

RNA Sample	Molecular Weight	Number of Residues
Kras 240-332	30,173.1 g/mol	93
MITF 1621-1709	28,643.9 g/mol	89

3.2 Results

Part A: Characterizing the CRDBP-RNA interaction *in vitro*, in cells, and *in vivo*

Part I: *In Vitro*

3.2.1 EMSAs with c-myc RNA and CRDBP Variants

The binding profile of CRDBP KH variants on c-myc RNA was assessed using EMSA. As shown in Figure 33B, both CRDBP-KH3 and CRDBP-KH4 showed comparable binding to [³²P] c-myc RNA 1705-1886 (182 nts) as the WT-CRDBP. In contrast, single KH mutation variants CRDBP-KH1 and CRDBP-KH2 showed a reduction in binding as compared to the WT-CRDBP (Figure 33A).

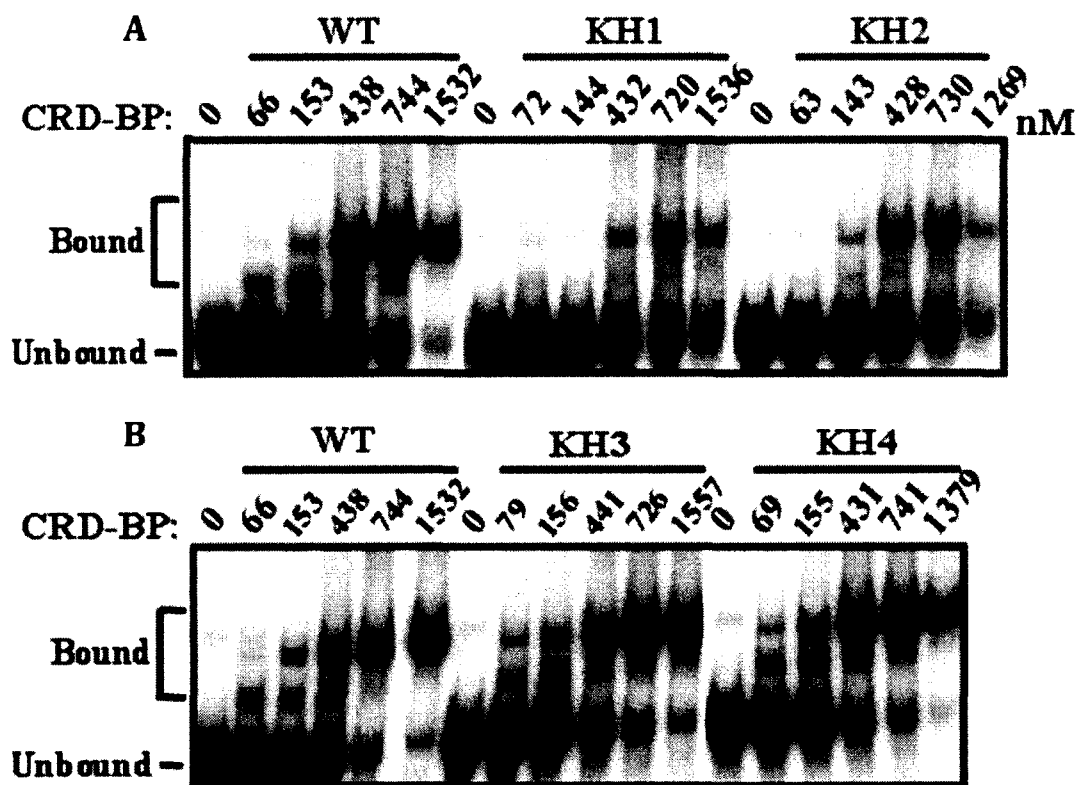


Figure 33. Electrophoretic mobility shift assay of the recombinant wild-type CRDBP and CRDBP variants. [32 P] c-myc 1705-1886 (182 nts) RNA substrate (20,000 cpm/reaction) was incubated with varying concentrations of CRDBP variants as shown. Samples were electrophoresed in a 4% native PAGE gel followed by autoradiography. Data shown are representatives from at least four experiments using proteins from at least two separate preparations.

To obtain quantitative measurements on the affinity of CRDBP for the [32 P] c-myc RNA 1705-1886 (182 nts), the dissociation constant (K_d) was determined using the Hill equation. The binding data for each of the CRDBP variants (Figure 33A and 33B) showed a sigmoidal curve when plotted and were thus fit using the Hill equation (Figure 34A-D).

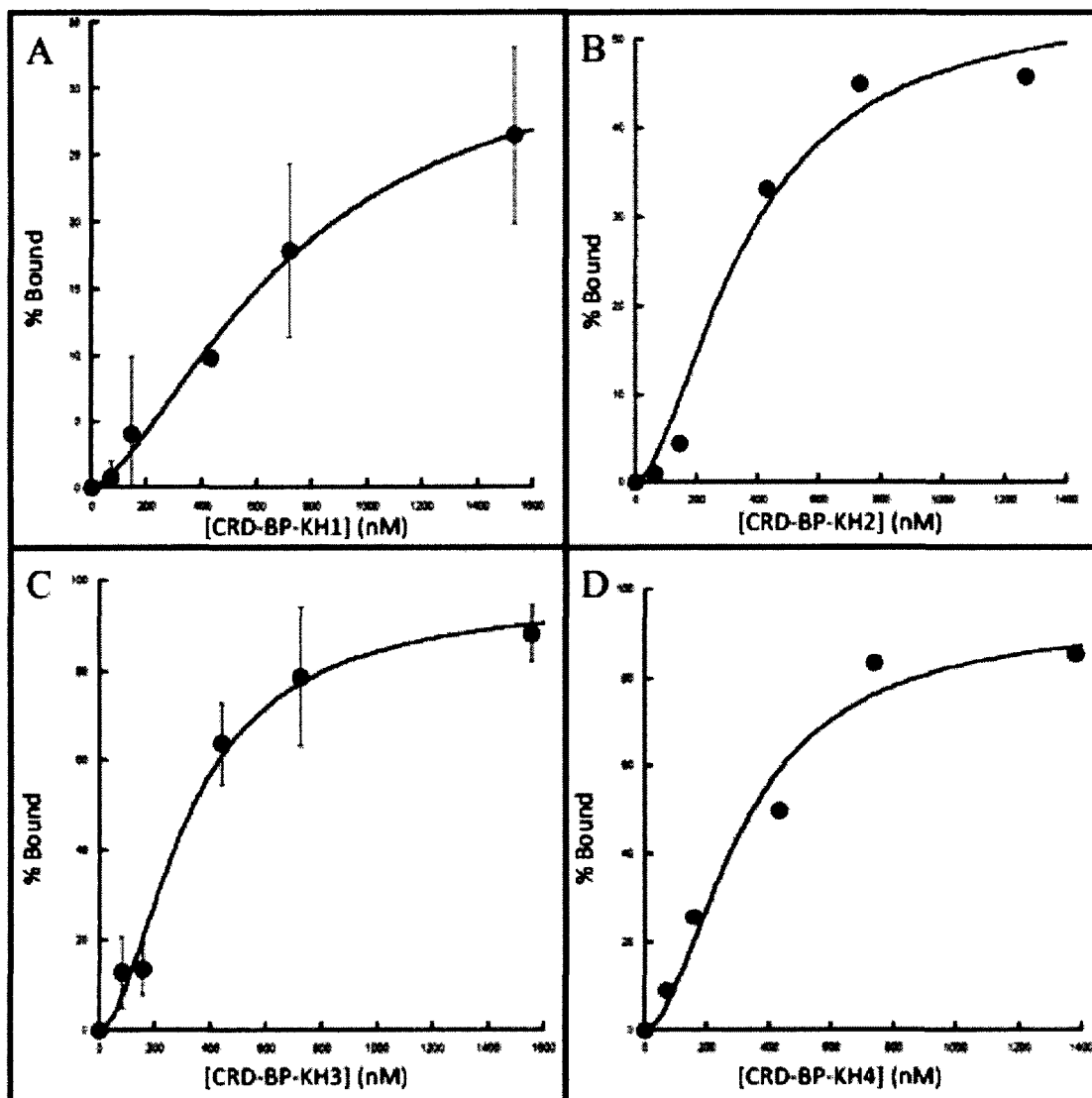


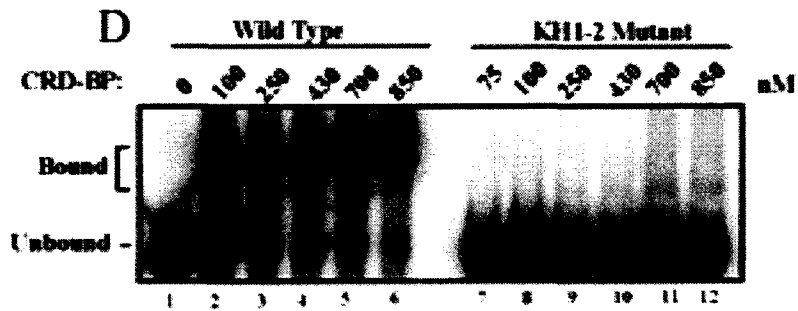
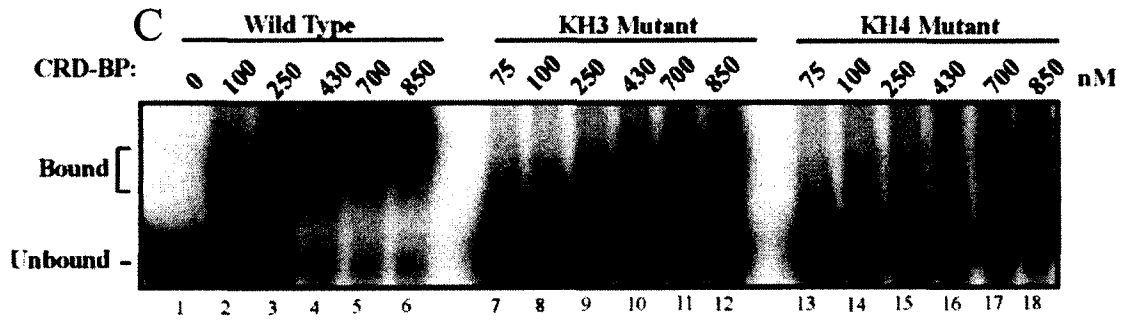
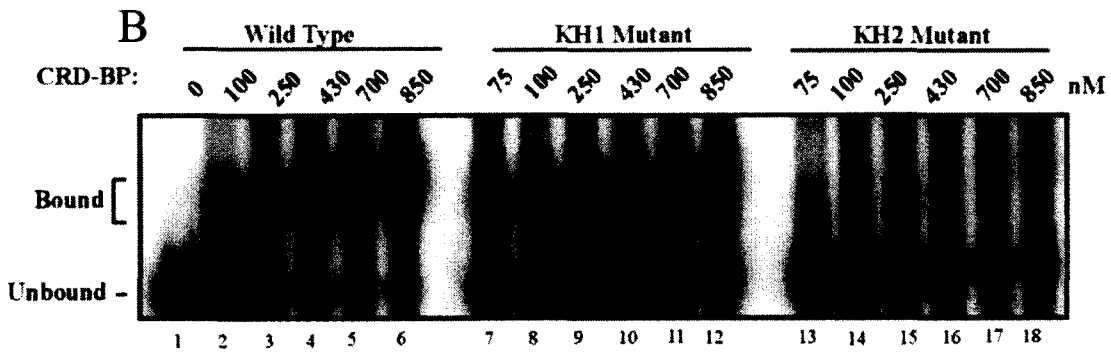
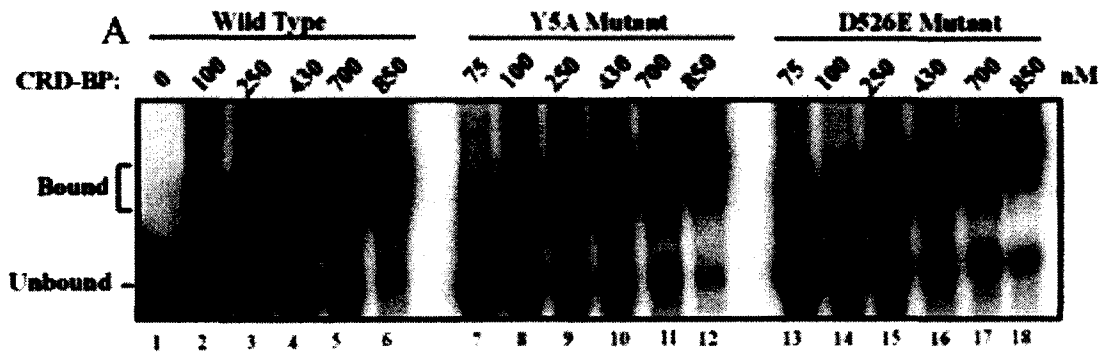
Figure 34. CRDBP variants binding curves as generated by the Hill Equation. Binding curves were generated using densitometry and the data from Figure 3A-B with at least two biological replicates as well as three separate experiments were pooled. CRDBP variants were incubated with [32 P] c-myc RNA 1705-1886 (182 nts) and the binding profiles were analyzed using densitometry. Percent bound vs Percent unbound values were used to generate the binding curves and calculate the dissociation constant. Errors bars indicate the standard error mean for A-D.

Each of the following dissociation constants were determined with [32 P] c-myc RNA 1705-1886 (182 nts) as the substrate and the use of densitometry to fit into the Hill equation. The dissociation constant for WT-CRDBP was determined to be 398 ± 52 nM

(Figure 33A). The dissociation constant for CRDBP-KH1 was determined to be 723 ± 75 nM (Figure 33A). The dissociation constant for CRDBP-KH2 was determined to be 360 ± 70 nM (Figure 33A). The dissociation constant for CRDBP-KH3 was determined to be 320 ± 42 nM (Figure 33B). The dissociation constant for CRDBP-KH4 was determined to be 320 ± 59 nM (Figure 33B). The measured dissociation constants are summarized in Table 14.

3.2.2 EMSA with MITF RNA and CRDBP Variants

Figure 35 shows representative results from EMSA experiments in assessing the WT CRDBP and its KH variants for ability to bind to MITF RNA. CRDBP-Y5A and CRDBP-D526E showed comparable binding to [32 P] MITF RNA 1550-1740 (191 nts) as the WT-CRDBP (Figure 35A). CRDBP-Y5A and CRDBP-D526E variants are negative controls and were expected to have no effect on binding affinity and similar binding profiles as the wild type CRDBP. The Y5A mutation is located in RRM1 of CRDBP and D526E is located within the variable loop of KH4 between β_2 β_3 . Both Y5A and D526E are negative controls because those two regions are not involved in the interaction with RNA. From the single mutant variants, CRDBP-KH1 showed a reduced binding affinity to [32 P] MITF RNA 1550-1740 (191 nts) as compared to the WT-CRDBP (Figure 35A-B). The other single KH variants, CRDBP-KH2, CRDBP-KH3, CRDBP-KH4 exhibited even more reduced binding with [32 P] MITF 1550-1740 (191 nts) in comparison to the WT-CRDBP (Figure 35A-C). Except for CRDBP-KH3-4, all CRDBP variants with double point mutations (CRDBP-KH1-2, KH1-3, KH1-4, KH2-3, KH2-4) showed complete abrogation of binding with [32 P] MITF RNA 1550-1740 (191 nts) (Figure 35A, 35D-H).



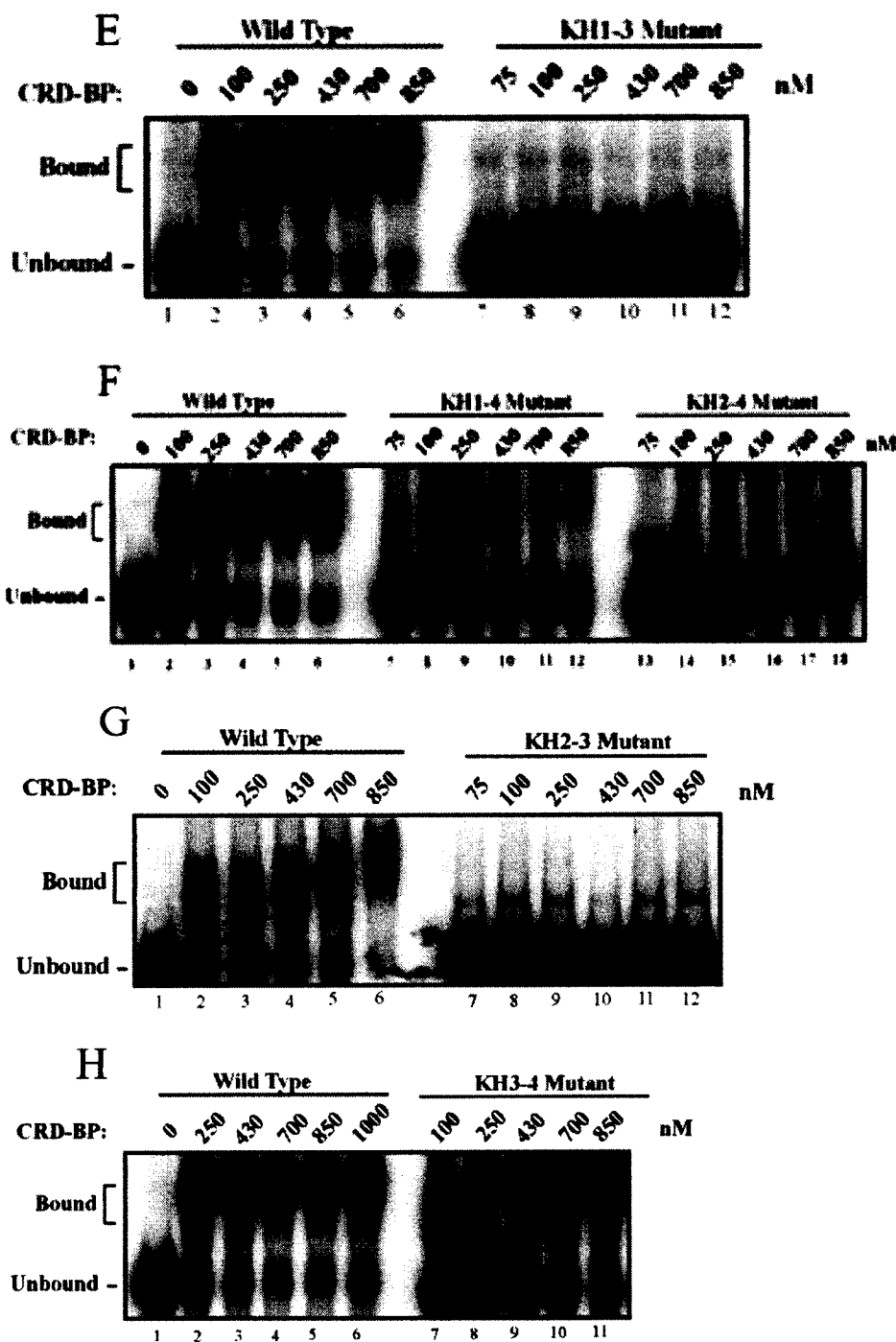


Figure 35. Electrophoretic mobility shift assay of recombinant wild-type and mutant CRDBP. [³²P] MITF RNA 1550-1740 (191 nts) RNA substrate (20,000 cpm/reaction) was incubated with varying concentrations of CRDBP variants as shown. Samples were electrophoresed in a 4% native PAGE gel followed by autoradiography. Data shown are representatives from at least three experiments using at least two separately prepared recombinant proteins.

From the EMSAs of [32 P] MITF RNA 1550-1740 (191 nts) with WT-CRDBP, CRDBP-Y5A, CRDBP-D526E and CRDBP-KH3-4 densitometry was performed to quantify the binding profiles. For CRDBP-KH2, KH3, KH4, KH1-2, KH1-3, KH1-4, KH2-3 and KH2-4 densitometry was not measured as these variants exhibited little to no binding affinity with MITF 1550-1740 (191 nts) (Figure 35B-C). The binding data was then fit into the Hill equation and showed a sigmoidal curve when plotted (Figure 36).

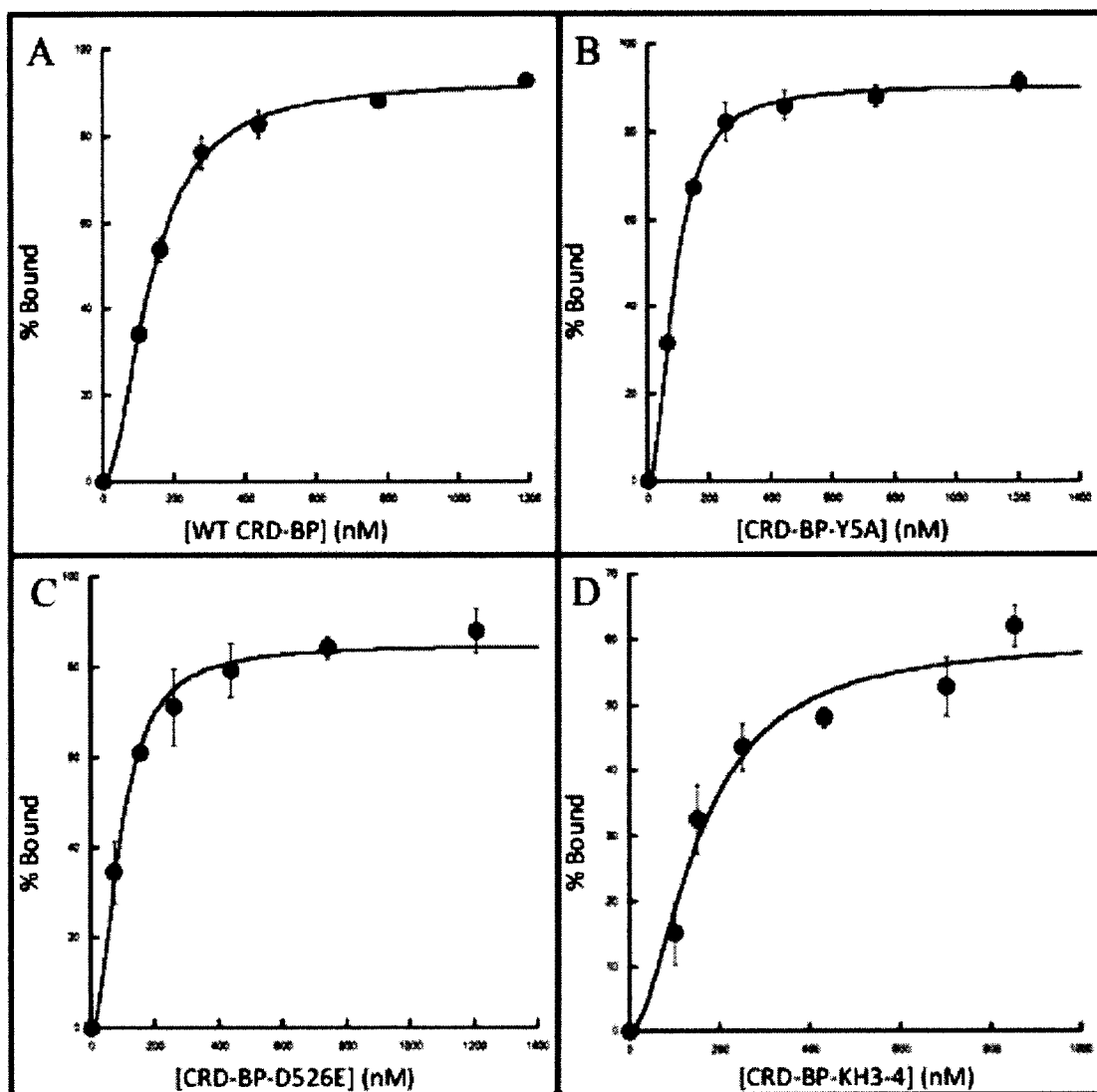


Figure 36. CRDBP variants binding curves as generated by the Hill Equation. Binding curves were generated using densitometry from Figure 35A-H. The data were pooled from at least two biological replicates as well as four separate experiments.

CRDBP variants were incubated with [³²P] MITF RNA 1550-1740 (191 nts) and the binding profiles were analyzed using densitometry. Percent bound vs Percent unbound values were used to generate the binding curves and calculate the dissociation constant.

Each of the following dissociation constants were determined with [³²P] MITF RNA 1550-1740 (191 nts) as the substrate and the use of densitometry fit into the Hill equation. The dissociation constant for WT-CRDBP was determined to be 131 ± 3 nM (Figure 35A, Table 14). The dissociation constant for CRDBP-Y5A was determined to be 86 ± 2 nM (Figure 35A, Table 14). The dissociation constant for CRDBP-D526E was determined to be 91 ± 6 nM (Figure 35A). The dissociation constant for CRDBP-KH3-4 was determined to be 155 ± 19 nM (Figure 35H).

The summary of K_d values for each of the CRDBP variants that bind with c-myc 1705-1886 (182 nts) and MITF 1550-1740 (191 nts) are shown in Table 14. Statistical analysis was performed using student t-test model.

Table 14. Summary table of the different CRDBP variants and determined K_d values.

Protein	c-myc 1705-1886 (182nts)	MITF 1550-1740 (191nts)
Y5A	468 ± 46 nM ^{MB}	85 ± 2 nM
KH1	723 ± 75 nM*	UD
KH3	321 ± 43 nM	No Binding
KH1-2	No Binding ^{MB}	No Binding
KH1-4	No Binding ^{MB}	No Binding
KH2-4	No Binding ^{MB}	No Binding

Student's t-test (n=4, $P^* < 0.05$) was used to determine whether there was a statistically significant difference between the K_d values of the KH single mutants as compared to the WT-CRDBP K_d . The asterisk indicates that the p-value is less than 0.05. The initials MB indicate Mr, Mark Barnes completed the EMSAs and K_d values (109).

Part II: In cells

To determine if mutation in KH domains has any effect on the ability of CRDBP to physically associate with its target mRNA in cells, pcDNA-FLAG-CRDBP plasmids were expressed in HeLa cells. Immuno-precipitation (IP) using anti-FLAG antibody was then performed. IP was followed by western blot analysis to ensure equivalent protein expression, confirming difference in bound RNA levels between wild-type and variants are due to differences in the ability of those immuno-precipitated CRDBP variants to bind with RNAs. As shown in Figure 37A, equal amount of FLAG-CRDBP were present in the cell lysates of all samples. RT-qPCR was then performed on the isolated RNA for measurements of c-myc, CD44 and β -actin mRNA. The measured mRNA levels of c-myc and CD44 for all variants containing a double mutation were found to be reduced in comparison to the wild-type CRDBP as shown in Figure 37B and 37C. For β -actin mRNA levels, the relative amounts between FLAG-WT-CRDBP and the variants did not change as shown in Figure 37D.

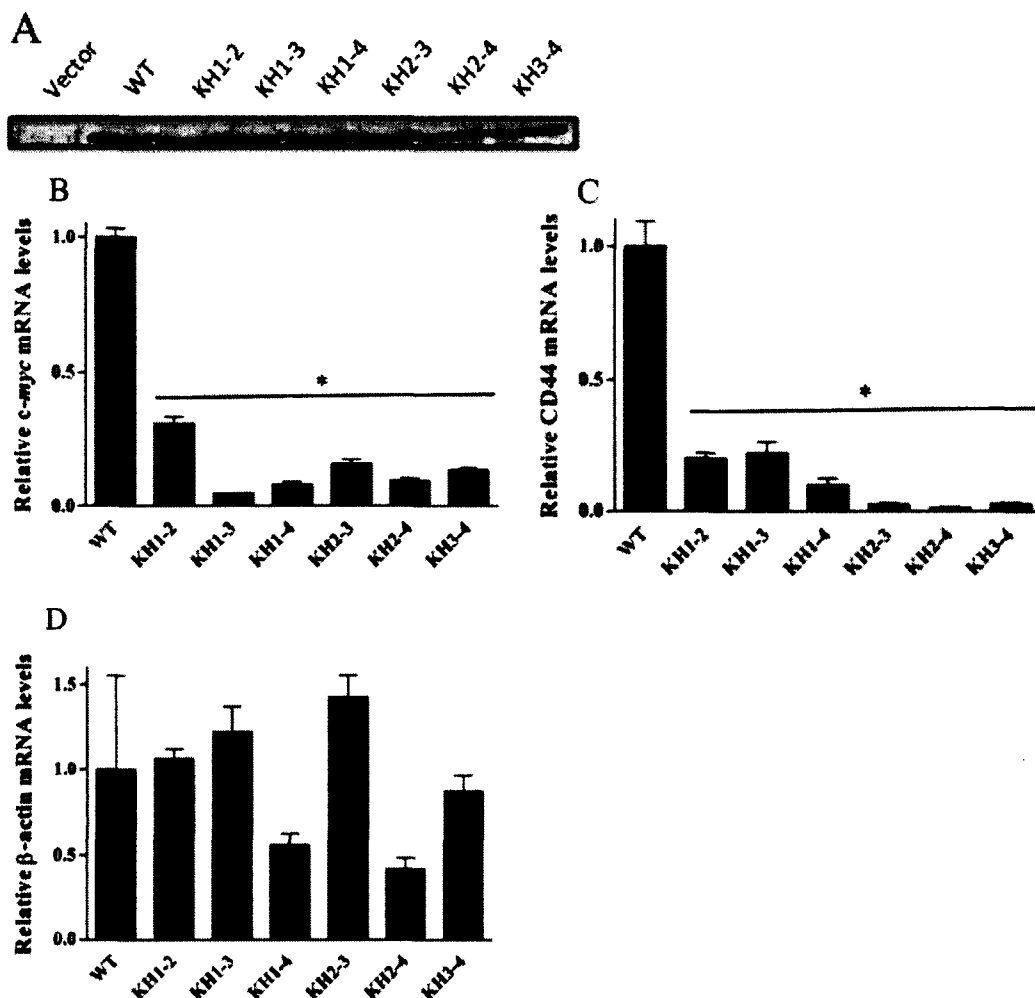


Figure 37. Analysis of specific mRNAs physically associated with the WT CRDBP and its variants. pcDNA-FLAG-CRDBP plasmids were transfected into HeLa cells as described in the Methodology section. FLAG-CRDBP-RNA complexes in cell lysates were immune-precipitated with anti-FLAG antibody. (A) Immuno-precipitated lysate was used in Western blot analysis to detect and confirm comparable exogenous FLAG-CRDBP protein expression using anti-FLAG antibody. (B-D) β -actin, CD44 and c-myc mRNAs isolated from the complexes were measured using qRT-PCR. mRNA levels were expressed over the pcDNA-FLAG control and compared to WT-CRDBP. Statistical significance between the expression of target genes in WT-CRDBP and CRDBP variants-transfected cells was assessed by the unpaired Student's t-test ($n=4$). The asterisk indicates that the p-value is less than 0.05.

Part III: *In vivo*

A previous study (Goswami et al 2010) showed that upon injection of zebrafish embryos with a specific morpholino against zebrafish CRDBP, suppression of *mif1a* (zebrafish's MTF) mRNA level was observed which was followed by reduction in

pigmentation in zebrafish 48 hours after fertilization (30). This study prompted us to investigate the role of KH domains of CRDBP in manifesting an *in vivo* phenotype, which in this case would be cell pigmentation in zebrafish. In collaboration with Dr. Andrew Miller's lab in the Hong Kong University of Science and Technology, Dr. Lee first needed to determine whether the mouse WT CRDBP can in fact rescue the pigmentation in zebrafish which had been previously knocked down with the specific morpholino against zebrafish CRDBP. Unfortunately, the experimental plan turned out to be more complicated than previously anticipated. Firstly, Dr. Lee discovered a significant mistake was made by the previous study. The morpholino used in the previous study was in fact targeting IGFBP1 (Insulin-like growth factor-binding protein 1) mRNA, and not IGF2BP1 (CRDBP) mRNA. Dr. Lee later designed a new morpholino, which is specific for IGF2BP1 (CRDBP) and performed the planned zebrafish experiments. Injections of four concentrations of CRDBP-morpholinos (4, 8, 12 and 16 ng) into at least three zebrafish embryos did not result in any observable changes in phenotype, including pigmentation. The experiments were repeated two more times using zebrafish eggs produced from different zebrafish. But again, no changes in phenotype, including pigmentation, were observed. Also surprisingly, when Dr. Lee used the wrong morpholino previously used by Goswami *et al*, no reduction in pigmentation was observed in zebrafish (30).

Part B: Analysis of RNA Species using Circular Dichroism Spectroscopy

3.2.7 Analysis of RNA Species using Circular Dichroism Spectroscopy

Circular dichroism (CD) spectroscopy has been successfully used to assess secondary structure changes in nucleic acids (DNA and RNA) upon interaction with proteins (113, 114). Here, I explore the possibility of using the CD spectroscopy to measure secondary structural changes in RNA upon interaction with CRDBP.

For the circular dichroism spectroscopy, optimal RNA concentration had to be first established. Kras 240-332 (93 nts) RNA was scanned at four concentrations, 435 nM, 869 nM, 1738 nM and 3314 nM in nuclease-free water (Figure 38A-D). As the concentration of the Kras 240-332 (93 nts) RNA molecule increased, the quality of the spectra was found to increase (Figure 38A-D). At the maximum concentration of 3314 nM of Kras 240-332 (93 nts) RNA molecule used, the spectra was at the best quality. However, regardless of the concentrations of Kras RNA used, a maximal peak was observed at around 265 nm. This ~3000 nM concentration range is where further RNA spectral analyses were conducted.

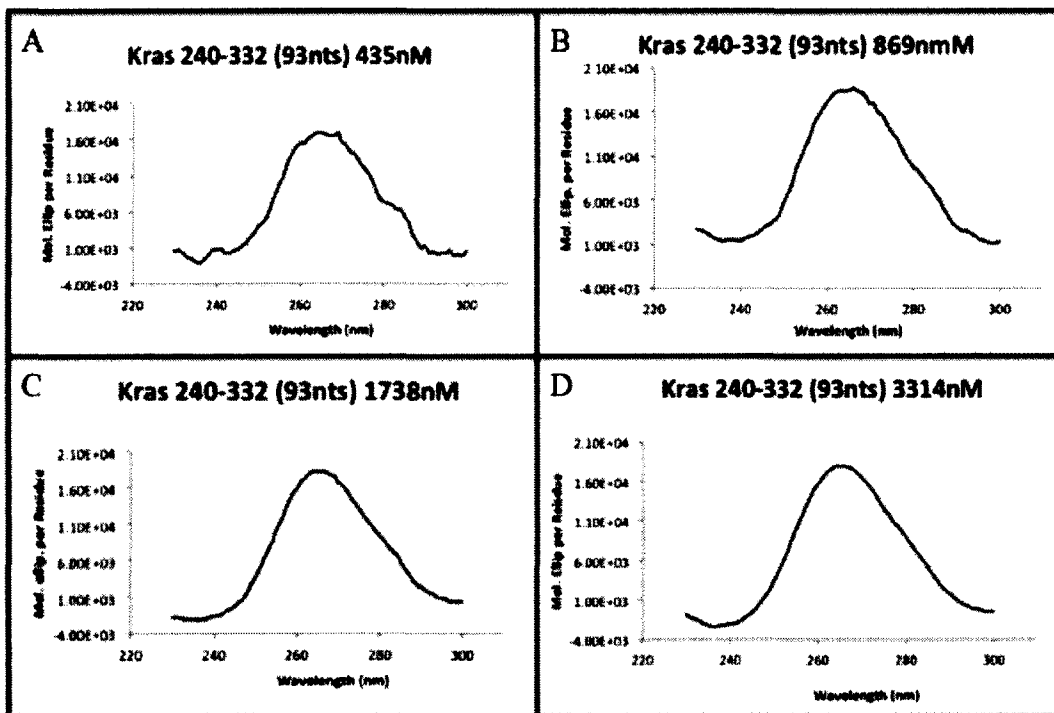


Figure 38. Circular Dichroism Spectra of Kras 240-332 (93 nts) RNA at different concentrations. (A) Kras 240-332 (93 nts) at 435nM. (B) Kras 240-332 (93 nts) at 869nM. (C) Kras 240-332 (93 nts) at 1738nM. (D) Kras 240-332 (93 nts) at 3314nM. Kras RNA was resuspended in nuclease-free water.

Next, in addition to Kras 240-332 (93 nts), c-myc 1705-1792 (88 nts), MITF 1621-1709 (89 nts) and Kras 148-239 (92 nts) were also scanned using the circular dichroism spectrophotometer in nuclease-free water (Figure 39). The Kras 148-239 (92 nts) RNA molecule does not bind with CRDBP, whereas the Kras 240-332 (93 nts) RNA molecule does (Mackedenski S, unpublished observation). The data from all four of these RNA molecules were then plotted onto a single graph in order to determine possible differences between them (Figure 39). Other than the slight differences in the height of the peak close to 263nm, the CD spectra of all the four RNAs look remarkably similar with a single peak (Figure 39).

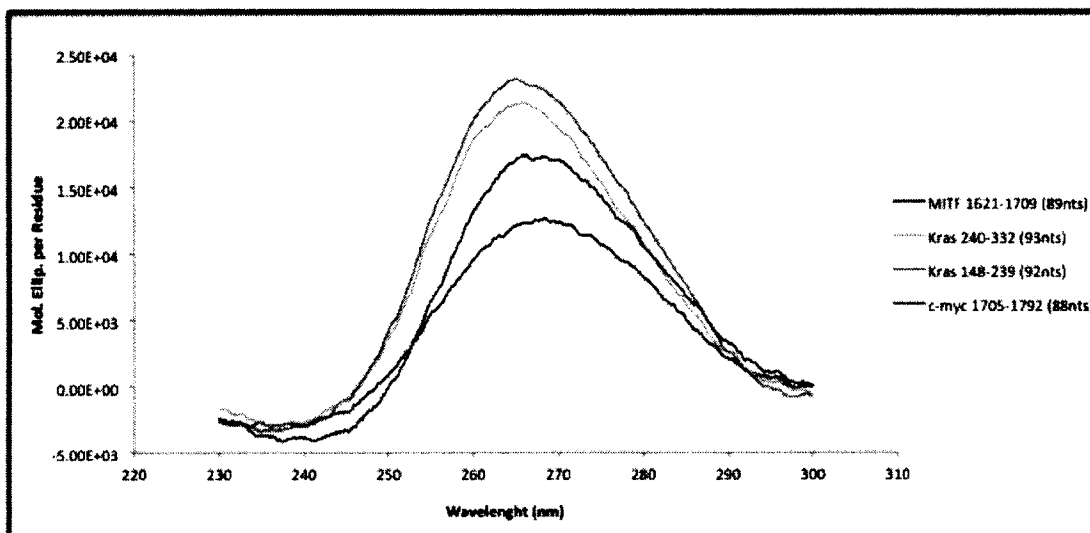


Figure 39. Circular Dichroism Spectra of different RNA molecules in water. The c-myc 1705-1792 (88 nts) RNA at 3514 nM is in blue. The MITF 1621-1709 (89 nts) RNA at 3491 nM is in purple. The Kras 240-332 (93 nts) RNA at 3337 nM is in green. The Kras 148-239 (92 nts) RNA at 3314 nM is in orange.

The same four RNA molecules were also scanned in dialysis buffer, in order to see if the buffer conditions change the spectra of measured RNA (Figure 40). This is important because subsequent experiments in assessing CRDBP-RNA interactions will utilize dialysis buffer. When the data from all four RNA molecules were compared on one graph, they produced a single peak with varying intensities (Figure 40), which were similar to what was previously observed when the RNAs were resuspended in water (Figure 39).

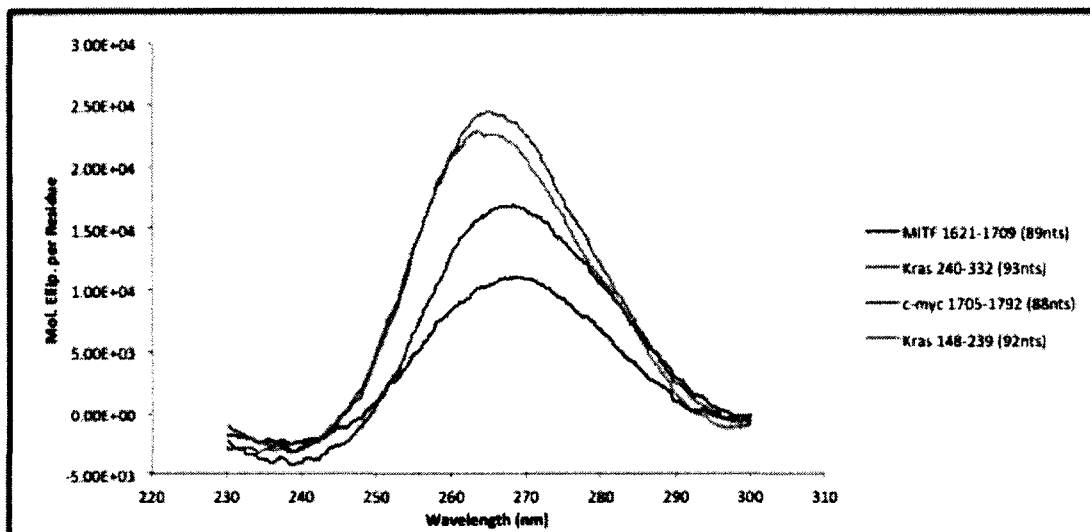


Figure 40. Circular Dichroism Spectra of different RNA molecules in dialysis buffer. The c-myc 1705-1792 (88 nts) RNA at 3514 nM is in blue. The MITF 1621-1709 (89 nts) at 3491 nM is in purple. The Kras 240-332 (93 nts) RNA at 3337 nM is in green. The Kras 148-239 (92 nts) RNA at 3314 nM is in orange.

When the CD spectra of the four RNA molecules in nuclease-free water (Figure 39) are plotted on the same graph as the four RNA molecules in dialysis buffer (Figure 40) as shown in Figures 41A-D the single peaks are similar in shape and intensity. The solid colour line represents the CD spectra from the RNA molecule in nuclease-free water and the black dotted line represents the CD spectra of the RNA molecule in dialysis buffer (Figures 41A-D).

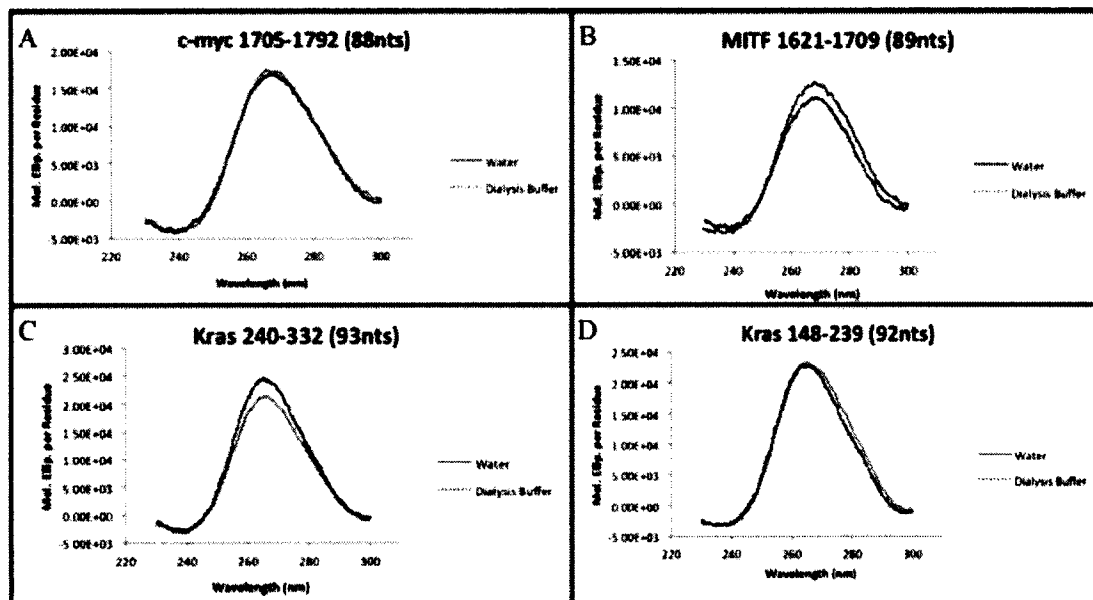


Figure 41. Comparison between water and dialysis buffer on the Circular Dichroism Spectra of different RNA molecules. The solid colour line represents the sample diluted in water and the dotted black line represents the sample diluted in dialysis buffer. (A) c-myc 1705-1792 (88 nts) at 3514 nM. (B) MITF 1621-1709 (89 nts) at 3491nM. (C) Kras 240-332 (93 nts) at 3337nM. (D) Kras 148-239 (92 nts) at 3314 nM.

Next, the CD spectra of RNAs were then determined in the presence of CRDBP (Figures 42A-C, 43A-C and 44A-C). Kras 148-239 (92 nts) 1807 nM RNA molecule was scanned in dialysis buffer in the presence or absence of 1807 nM CRDBP as shown in Figures 42A and 42B respectively. These two scans were then plotted onto the same graph, where the solid colour line represents the Kras 148-239 (92nts) 1807 nM RNA molecule with 1807 nM CRDBP sample, and the black dotted line represents the Kras 148-239 (92nts) 1807 nM RNA alone (Figure 42C). Both the Kras 148-239 (92nts) with or without 1807 nM CRDBP present generated almost identical single peak spectra as shown in Figure 42C.

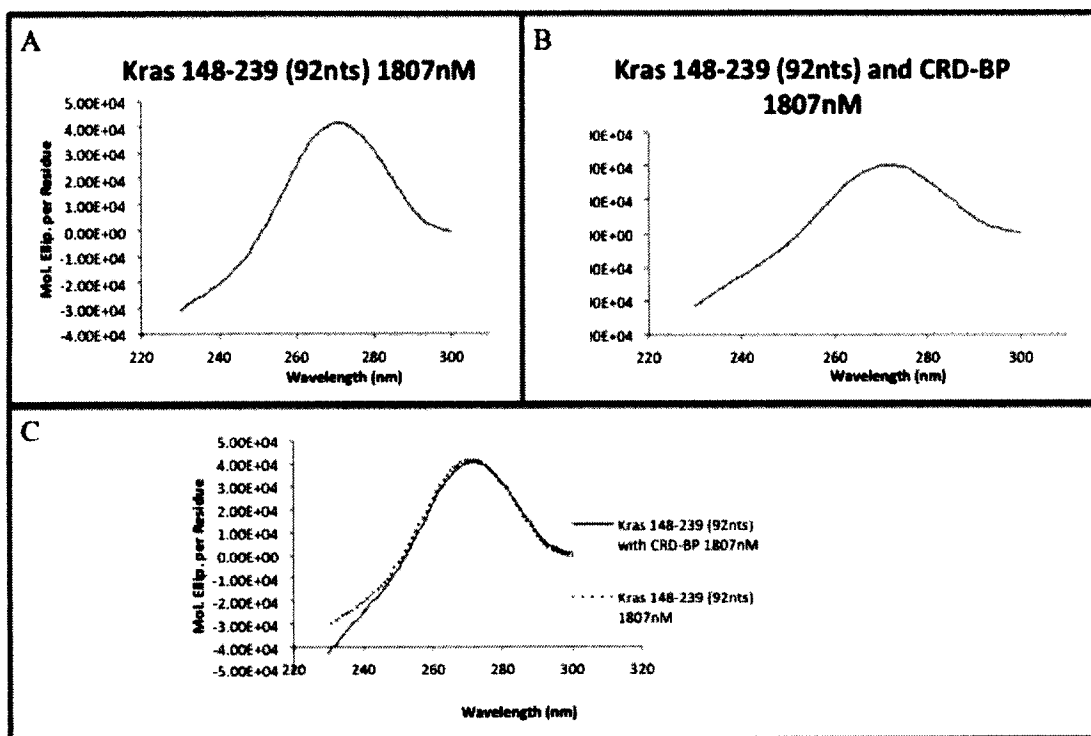


Figure 42. Circular Dichroism Spectra of Kras 148-239 (92 nts) RNA in the presence of WT-CRDBP. (A) Kras 148-239 (92 nts) 1807 nM RNA in dialysis buffer. (B) Kras 148-239 (92 nts) 1807 nM RNA in the presence of WT-CRDBP 1807 nM. (C) Overlay of CD Spectra A and B.

Kras 240-332 (93 nts) 1824 nM in the presence of 1824 nM of CRDBP in dialysis buffer was also scanned (Figure 43A) and compared to Kras 148-239 (92 nts) in the presence of 1807 nM CRDBP (43B) by plotting them onto the same graph as shown in Figure 43C. The two CD spectra scans of two different Kras RNA molecules in solution with CRDBP showed a single peak with similar intensities (Figure 43C). The solid colour line represents Kras 240-332 (93 nts) 1824 nM RNA molecule with 1824 nM CRDBP and the black dotted line represents Kras 148-239 (92 nts) 1807 nM molecule with 1807 nM CRDBP (Figure 43).

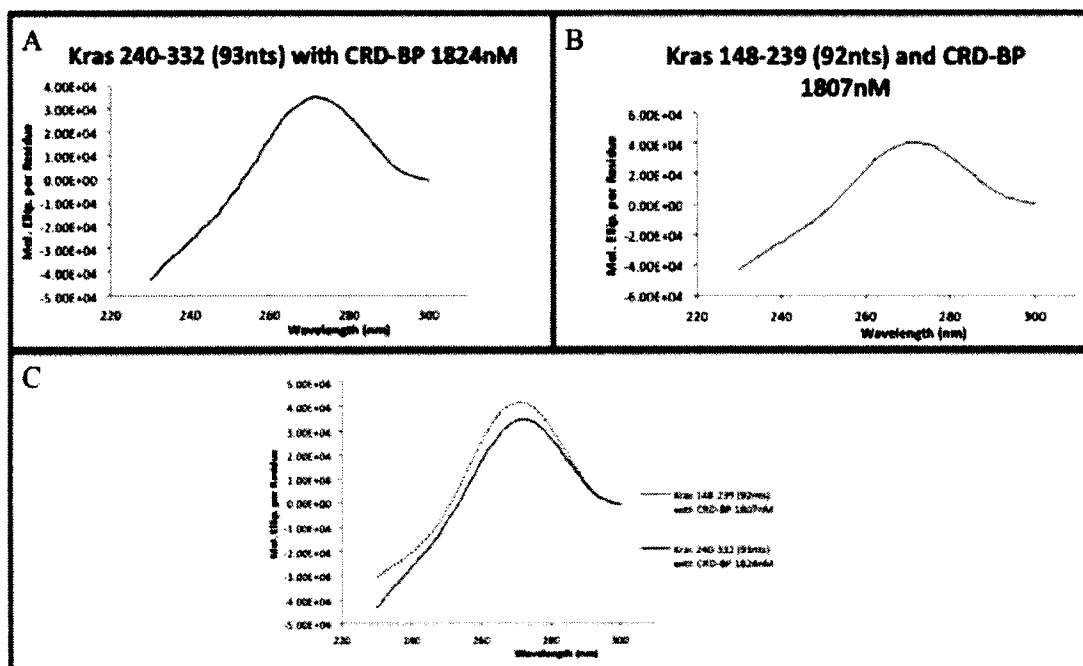


Figure 43. Circular Dichroism Spectra of different Kras RNA molecules in the presence of WT-CRDBP. (A) Kras 240-332 (93nts) 1824nM RNA in the presence of WT-CRDBP 1824nM. (B) Kras 148-239 (92nts) 1807nM RNA in the presence of WT-CRDBP 1807nM. (C) Overlay of CD Spectra A and B.

Reducing the Kras:CRDBP ratio was also conducted to determine if changes in RNA spectra can be observed. The Kras 240-332 (93 nts) 912 nM RNA molecule with 1824 nM CRDBP was scanned (Figure 44B) and compared to Kras 240-332 (93 nts) 1824 nM RNA molecule with 1824 nM CRDBP (Figure 44A). The results were then plotted onto the same graph as shown in Figure 44C. As shown, reducing the RNA concentration of Kras 240-332 (93nts) in half resulted in a single peak with similar intensity (Figure 44C). A higher RNA concentration will give rise to higher peak intensity at 263 nm, but when comparing RNA of similar concentrations, small changes in the peak intensity would be attributed to how RNA molecules absorb the light differently.

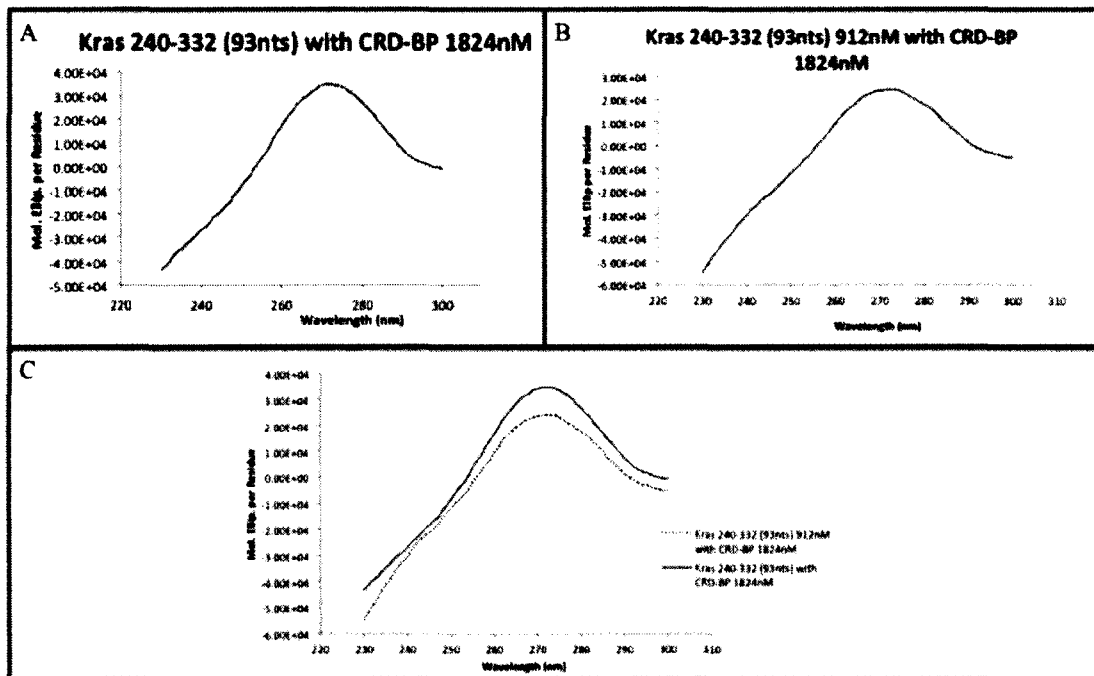


Figure 44. Circular Dichroism Spectra of Kras 240-332 (93nts) RNA at different concentrations in the presence of WT-CRDBP. (A) Kras 240-332 (93nts) 1824nM RNA in the presence of WT-CRDBP 1824nM. (B) Kras 240-332 (93nts) 912nM RNA in the presence of WT-CRDBP 1824nM. (C) Overlay of CD Spectra A and B.

DISCUSSION

Part A:

In vitro

The reason for assessing the different RNA-binding abilities of CRDBP KH variants was to determine which KH domains play an important role in interacting with c-myc 1705-1886 (182 nts) and MITF 1550-1740 (191 nts) RNAs. One concern was whether the point mutations would affect the global structure of the protein, which is not the case as circular dichroism of WT-CRDBP and all the variants indicated similar, if not identical, CD spectra (115).

The CRDBP control variants, Y5A and D526E, showed similar binding affinity as the WT-CRDBP for c-myc 1705-1886 (182 nts) RNA fragment (Table 14). In contrast, KH1 variant exhibited a significant reduction in RNA-binding ability suggesting that the KH1 domain plays an important role in interacting with c-myc 1705-1886 (182 nts) RNA fragment (Table 14, Figure 33A). The importance of the KH1 interaction with c-myc 1705-1886 (182 nts) RNA fragment is exemplified with the fact that the CRDBP variants with mutations at KH1-2, KH1-3, and KH1-4 showed little to no binding (Table 14). It is also interesting, because KH2-3, and KH2-4 showed no binding with c-myc 1705-1886 (182 nts) RNA fragment, and although the KH1 was unaffected, mutations in the other KH domains had an effect (Table 14). The CRDBP-KH3 and KH4 variants, as well as CRDBP-KH3-4 variant showed similar binding affinity as the WT-CRDBP, indicating that the G-X-X-G motif in these two KH domains are less important in interacting with c-myc 1705-1886 (182 nts) RNA fragment (Table 14). This also demonstrates that a single KH domain is not responsible for interacting with c-myc 1705-1886 (182 nts) RNA fragment alone, but at least two.

For MITF 1550-1740 (191 nts) RNA fragment, both CRDBP control variants, Y5A and D526E, as well as KH3-4 variant showed binding affinity similar to that of the WT-CRDBP (Table 14, Figure 35A, 35H). The CRDBP KH2, KH3, KH4, KH1-2, KH1-3, KH2-3, and KH2-4 variants all showed no binding affinity with MITF 1550-1740 (191 nts) RNA fragment (Table 14, Figure 35B-G). In comparison, the CRDBP-KH1 variant maintained some binding affinity with MITF 1550-1740 (191 nts) although the dissociation constant cannot be determined for comparison (Figure 35B, Table 14). Surprisingly the CRDBP-KH3-4 variant showed wild-type like binding affinity for MITF

1550-1740 (191 nts) RNA fragment although single variants CRDBP-KH3 and KH4 did not show binding as seen in Figures 35C and 35H (Table 14). This suggests that regions other than the G-X-X-G motif in KH3 and KH4 domains may have the capacity to cooperate in facilitating RNA interaction. From the single and double CRDBP variant work, the mutation at KH2 showed to have the greatest impact on CRDBP's ability to bind with MITF 1550-1740 (191 nts) RNA fragment, as the mutation at KH1, showed to retain some binding and the double mutation at KH3-4 also retained some binding (Table 14, Figure 35B and 35H).

In comparing the c-myc 1705-1886 (182 nts) with the MITF 1550-1740 (191 nts), the CRDBP-KH1 variant had a greater impact on the c-myc 1705-1886 (182 nts) RNA fragment ability to bind whereas the CRDBP-KH2 impacted MITF 1550-1740 (191 nts) RNA fragment to bind. The CRDBP-KH3-4 variant showed the retention of binding ability for both c-myc 1705-1886 (182 nt) RNA fragment and MITF 1550-1750 (191 nts) RNA fragment (Table 14). It then appears that the G-X-X-G motif in the KH1 and KH2 domains play a greater role in interacting with these two RNA molecules than the G-X-X-G motifs in the KH3 and KH4 domains (Table 14).

In cells

The role of the G-X-X-G motif in the KH domains of CRDBP for binding mRNAs was further investigated in HeLa cells. This work initially began by a previous member in Dr. Lee's lab in assessing the CRDBP single KH variants. It was found that KH1, KH2, KH3 had reduced affinity to c-myc RNA, and KH4 was comparable to the WT-CRDBP (115). The assessment of the CRDBP variants with double KH mutations,

however, showed a decrease in affinity to c-myc and CD44 RNAs, which correlated with the EMSAs of c-myc 1705-1886 (182 nts) and CD44 2862-3055 (194 nts) (Figure 37B-C, Table 14) (115). By comparing the cell data to the *in vitro* data, a clear correlation can be made that the mutations in the KH domains affected CRDBP's RNA binding ability. The mutations at the G-X-X-G motifs in KH domains appeared to be enhanced in the cells as compared to *in vitro*. This is because the CRDBP-KH3-4 variant showed no binding to c-myc and CD44 mRNA from the IP experiments as compared to *in vitro* analysis where the CRDBP-KH3-4 variant demonstrated RNA binding (Figure 37B-C) (115).

In vivo

The investigation into the role of KH domains of CRDBP in conferring *in vivo* phenotype, namely pigmentation in zebrafish, could not be performed. This is because Dr. Lee could not reproduce the results reported in the previous study (Goswami *et al*) using their incorrect morpholino (30). In addition, using a new and specific morpholino against CRDBP in zebrafish embryos, Dr. Lee could not observe any changes in phenotype, including pigmentation, in zebrafish.

Part B: Analysis of RNA species using Circular Dichroism spectroscopy

The optimal concentration for scanning RNA appears to be ~3000 nM. At ~900 nM and ~1500 nM concentrations, the quality of CD spectra generated appeared acceptable, but not as good as the spectra generated when ~3000 nM RNA was used (Figure 38B-D). However, at ~400 nM concentration of RNA, the CD spectra generated were less reproducible and tend to exhibit dimples in the line (Figure 38A). Between c-

myc 1705-1792 (88 nts), MITF 1621-1709 (89 nts), Kras 240-332 (93 nts) and Kras 148-239 (92 nts) RNA molecules, there appears to be only minor differences in peak intensity (Figures 39A-E and 40A-E). Comparing these four RNA molecules in nuclease-free water versus dialysis buffer, again showed only minor differences in peak intensities (Figures 41A-D).

The spectra of Kras 148-239 (92 nts) RNA molecule did not alter much in the presence of equivalent WT-CRDBP protein (Figures 42A-C). At first glance, this is not surprising because Kras 148-239 (92 nts) does not interact with WT-CRDBP in EMSA (Mackedenski S, unpublished data). However, what is surprising is the comparison of Kras 148-239 (92 nts) with WT-CRDBP to Kras 240-332 (93 nts) with WT-CRDBP (Figure 43A) as there was still no change in the generated RNA spectra (Figure 43C), because Kras 240-332 (93 nts) does in fact interacted with WT-CRDBP in EMSA assays (Mackedenski S, unpublished data). One would imagine that the RNA structure of CRDBP-RNA binder would change shape upon interaction with CRDBP and in return have an altered CD spectra. Having compared Kras 148-239 (92 nts) and Kras 240-332 (93 nts) RNA molecules in a 1:1 ratio with WT-CRDBP, which showed no change in CD spectra, a 2:1 Kras 240-332 (93 nts) RNA molecule:WT-CRDBP ratio was also scanned (Figure 44B). Such 2:1, Kras RNA:WT-CRDBP protein ratio was compared to the 1:1 ratio, and still there was no evident of spectral changes (Figure 44C).

From all of these CD spectra scans of different RNA molecules, in the presence or absence of CRDBP, a very similar spectral pattern appears, namely the presence of a single peak at 263nm. This single peak at 263nm represents the absorbance of RNA, and does not provide any detailed structural information of the individual RNAs or how these

RNAs structure has changed in the presence of CRDBP. This being said, perhaps the global RNA structure does not change and hence no changes in CD spectra can be observed. It is possible that the CRDBP-RNA interaction only involves a number of critical nucleotides present in the single-stranded loop RNA regions and such interactions cannot be measured using CD spectroscopy.

Chapter 4. Using *in silico* analytical methods to compare and contrast RNAs that bind CRDBP

Trying to understand the RNA is equally as important as the protein, in the context of how a protein-RNA interaction is formed. Various labs have attempted to decode how an RNA's structure plays a role in being recognized by the corresponding RBP, but most only report on the primary structure, neglecting the secondary and tertiary possibilities. For RBPs in the VICKZ family, this was first described in 1997 by Robert Singer's group studying ZBP-1 which concluded that a 5'-ACACCC-3' sequence of β -actin was the key in the interaction and thereafter translocation (42). In 2003, the Singer group tested truncated ZBP-1 variants to determine the domains that are involved in binding the RNA, and found the truncated variant containing KH3 and KH4 bound as well as the full length ZBP-1 to the established 5'-ACACCC-3' sequence (41). They also performed a SELEX experiment where they found a consensus sequence of 5'-RCACCC-3' (R = pyrimidine), which competed as well as the 5'-ACACCC-3' sequence (41).

The Nielsen group performed IMP-1 granule precipitation, and analyzed the associated mRNA by microarray analysis and found 671 associated transcripts (116). From these six hundred seventy one transcripts, three hundred and seven were selected as positive hits, based on sequences being able to be retrieved from NCBI Entrez, containing less than 1% unknown nucleotides, and sequences less than 5000 nucleotides in length (116). They then searched for these seven nucleotides in length motifs (a "word"), to see if any of the 307 transcript share a common word, and found that these IMP-1 targets enrich a CCYHHCC (Y = C or U and H = C or U or A) motif (116).

Hafner *et al.* performed PAR-CLIP (Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation) to assess RBP recognition elements (RREs) for

FLAG/HA-tagged IMP-1, IMP-2 and IMP-3 (117). After PhyloGibbs analysis, a RRE sequence of CAUH (H = A, U, or C) was found in more than 75% of the top 1000 clusters from 100,000 sequences clusters recognized by the IMPs (117).

As the research for a primary sequence recognized by the VICKZ family has progressed from 1997 by the Singer group, to 2010 by Hafner *et al.*, there has been a decreased in the specificity of the sequence (Table 15) (42, 117). Based on this, it is unlikely that the primary sequence alone is recognized by the VICKZ family of RBPs.

Table 15. The progression of RNA primary sequence associated with the VICKZ family of RNA Binding proteins.

Sequence	Odds of sequence occurring
5'-RCACCC-3' (R = Pyrimidine) (41)	1:2,048
CAUH (H = A, U, or C) (117)	1:85

To date, there has been little or no effort in finding consensus secondary or tertiary structure of RNA recognized by RBP. This is in part due to the complexity associated with potentially sorting through a massive amount of information to find the tentative structure. In addition, further challenges may arise when there is lack of knowledge on determined RNA secondary and tertiary structure. However, investigation into predicted structures of a group of RNAs bound by a common RBP may yield valuable information. To this end, the goal of this investigation is to compare and contrast the predicted structure of a group of oncogenic RNAs known to bind CRDBP.

4.1 Methodology

The sequences for the six RNA fragments that are known to bind CRDBP are: c-myc 1705-1792 (88 nts) (56), MITF 1621-1709 (89 nts) (Chapter 2, this thesis), Kras

240-332 (93 nts) (Mackedenski S, unpublished results), CD44 2861-2958 (98 nts) (107), MDR-1 790-881 (92 nts) and GLI1 307-393 (87 nts) (Table 16) (108).

Table 16. Oncogenic RNAs with high affinity for CRDBP.

Name and Nucleotide Positions of RNA (Size)	RNA Sequence (5'-to-3')
MITF 1621-1709 (89nts)	UUUGCGUCAGAGAAAUGUCUGUCCAUUUUUAUUC AGGGGAAACUUGAUUUGAGAUUUUAUGCCUGUG ACUCCUUGGAAUCAAAUGU
CD44 2861-2958 (98nts)	GAAAUUAGGGCCCAAUUAUAUCAGCAAGAAUU UGAUCGUUCCAGUUCCCACUUGGAGGCCUUUCAU CCCUCGGGUGUGCUAUGGAUGGCUUCUAAC
GLI1 307-393 (87nts)	AUGCCAGCCUGGACCUGCAGACGGUUAUCCGCAC CUCACCCAGCUCCCUCGUAGCUUUCAACUCGC GAUGCACAUCUCCAGGAG

The following programs were used to determine the secondary or tertiary structure of the six RNAs listed in Table 2: Mfold, RNAfold, Sfold, CARNAC, MC-FOLD/MC-SYM and LocARNA. The following are the websites where the *in silico* analysis were performed:

Mfold: <http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form> (118)

RNAfold: <http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi> (119)

(Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure.)

Sfold: <http://sfold.wadsworth.org/cgi-bin/srna.pl> (120)

LocARNA: <http://rna.informatik.uni-freiburg.de/LocARNA/Input.jsp> (121)

CARNAC: <http://bioinfo.lifl.fr/cgi-bin/RNA/carnac/carnac.py> (104)

MC-FOLD/MC-SYM: <http://www.major.ircic.ca/MC-Fold/> (106)

The parameters for each of the RNA folding software were kept at default, unless there were specific options that needed to be specified, in which case the folding was performed for each of the options.

Mfold

The default parameters consist of the sequence being linear, folding temperature remains fixed at 37°C, ionic conditions at 1M NaCl, no divalent ions, 5% sub-optimality structures (considers structures within in 5% of the minimal free energy), an upper bound of 50 total structures, default window parameter (sequence length of 50-119 =default window size of 2, a smaller number represents more computed folding closer to each other, where a larger number computes less structures that are very different from each other), maximum interior/bulge size of 30, maximum asymmetry of an interior/bulge loop of 30, and no limit on the maximum distance between paired bases.

RNAfold

The default minimum free energy function was chosen, with the option to avoid isolated base pairs (prevents undesirable helices of length 1), dangling end energies assigned to free nucleotides on both sides of a helix in any case, RNA energy parameter from Turner model 2004, RNA molecule is assumed to be linear, and the default temperature of 37°C.

Sfold

The default parameters consist of the sequence being linear, folding temperature remains fixed at 37°C, ionic conditions at 1M NaCl, no divalent ions, and no limit on the maximum distance between paired bases.

LocARNA

No RNA prediction folding parameters were provided, and a global standard alignment type and mode format was used.

CARNAC

In the CARNAC program the option of having the GC threshold enabled or disabled was required. Therefore, both options were performed. When GC threshold is enabled, it allows CARNAC to use variable energy thresholds for stems according to the average GC percent of the involved sequence.

MC-FOLD/MC-SYM

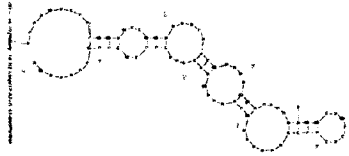
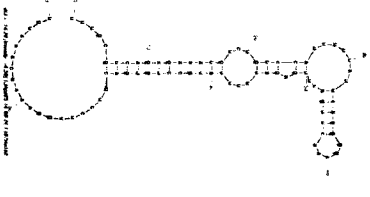
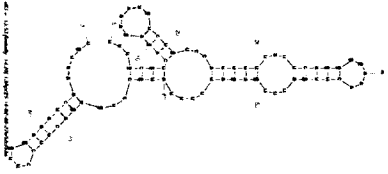
For the MC-FOLD program, the options of specifying to consider pseudoknots enabled and disabled were both performed. The default parameters for MC-FOLD, not considering H-type pseudoknots were to return the best 1000 structures with consideration for the best 15% sub-optimal structures. The default parameters for MC-FOLD when considering H-Type pseudoknots were to return the best 20 structures with consideration for the best 15% sub-optimal structures. The top structures were then directly pipelined into MC-SYM.

4.2 Results

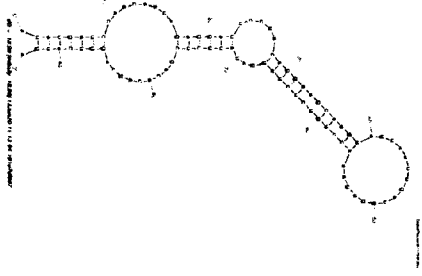
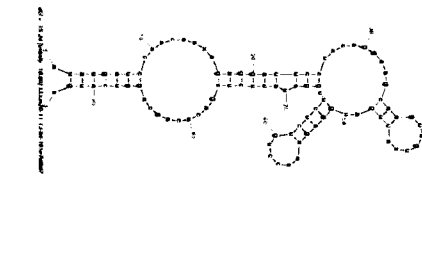
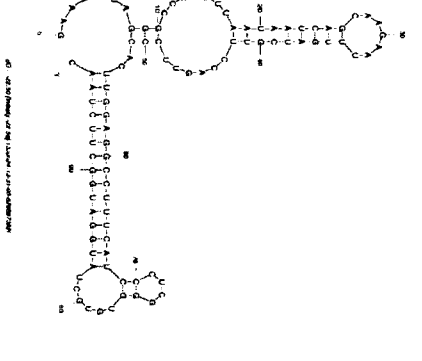
Mfold

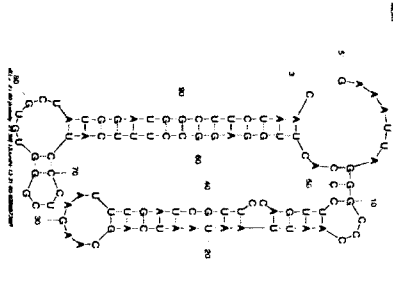
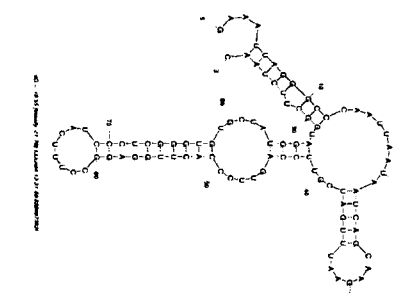
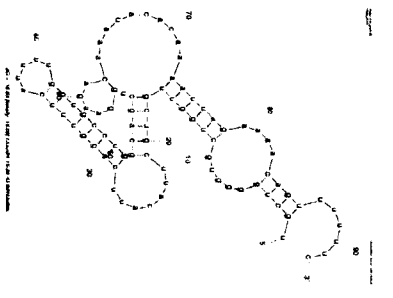
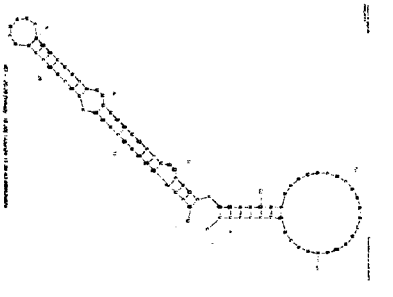
Using Mfold on c-myc 1705-1792 (88 nts) RNA, only one possible structure was generated. As shown in Table 17, the structure has five stem regions, one loop, and eight bulges, and has a ΔG of -14.10kcal/mol. Input of MITF 1621-1709 (89 nts) RNA generated five possible structures, with the most stable structure having a ΔG of -14.20kcal/mol, three stem regions, one loop and seven bulges (Table 17). The Mfold result for Kras 240-332 (93 nts) RNA showed three possible structures. Of these three structures, the second structure was most stable, having three stems, one loop and four bulges with a ΔG of -18.90kcal/mol (Table 17). The CD44 2861-2958 (98 nts) RNAs most stable structure of three had a ΔG of -22.50kcal/mol with four stems, two loops and four bulges (Table 17). There were two structure generated for MDR-1 790-881 (92 nts) RNA and the most stable structure had three stems, one loop and four bulges with a ΔG of -20.20kcal/mol (Table 17). For GLI1 307-393 (87 nts) RNA a single structure was predicted with four stems, two loops, four bulges and a ΔG of -13.95kcal/mol (Table 17).

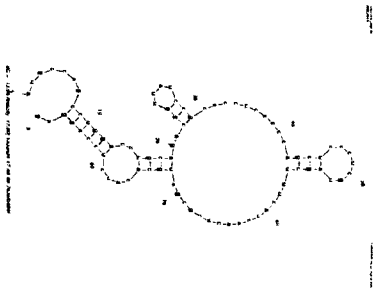
Table 17. Structures generated with Mfold of the six oncogenic RNAs.

Name	Structure	Number of Stems, Loops, Bulges	Total Energy (ΔG) (kcal/mol)
c-myc 1705-1792 (88nts)	(1/1) 	5, 1, 8	-14.10
MITF 1621-1709 (89nts)	(1/5) 	3, 1, 7	-14.20
	(2/5) 	5, 3, 5	-13.53

	<p>(3/5)</p>	5, 2, 6	-11.90
	<p>(4/5)</p>	4, 3, 3	-10.25
	<p>(5/5)</p>	4, 2, 6	-13.20
Kras 240-332 (93nts)	<p>(1/3)</p>	5, 2, 8	-15.80

	<p>(2/3)</p> 	3, 1, 6	-18.90
	<p>(3/3)</p> 	4, 2, 5	-15.24
<p>CD44 2861-2958 (98nts)</p>	<p>(1/3)</p> 	4, 2, 4	-22.50


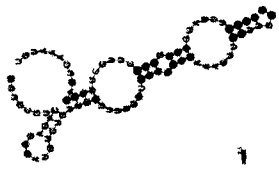
	<p>(2/3)</p> 	<p>4, 2, 4</p>	<p>-21.90</p>
	<p>(3/3)</p> 	<p>4, 2, 5</p>	<p>-19.55</p>
<p>MDR-1 790-881 (92nts)</p>	<p>(1/2)</p> 	<p>4, 2, 4</p>	<p>-16.63</p>
	<p>(2/2)</p> 	<p>3, 1, 4</p>	<p>-20.20</p>


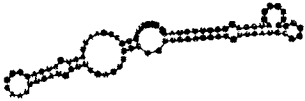

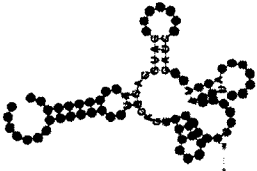
<p>GLI1 307-393 (87nts)</p>	<p>(1/1)</p> 	<p>4, 2, 4</p>	<p>-13.95</p>
---------------------------------	--	----------------	---------------

RNAfold

As shown in Table 18, RNAfold predicted one structure for each of the six sequences. Each of these structures has an associated base-pair probability for each of the nucleotides within the structure, as indicated with colours ranging from blue (zero probability) to red (100% probability).

Table 18. RNAfold generated structures of the six oncogenic RNAs.

Name	MFE Base-pair Probabilities
<p>c-myc 1705-1792 (88nts)</p>	
<p>MITF 1621-1709 (89nts)</p>	

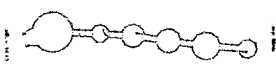
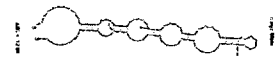


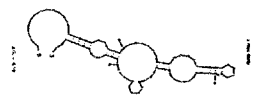
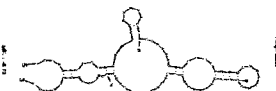
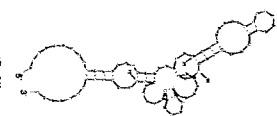
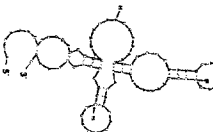

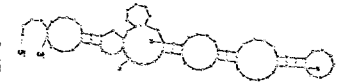
Kras 240-332 (93nts)	
CD44 2861-2958 (98nts)	
MDR-1 790-881 (92nts)	
GLI1 307-393 (87nts)	

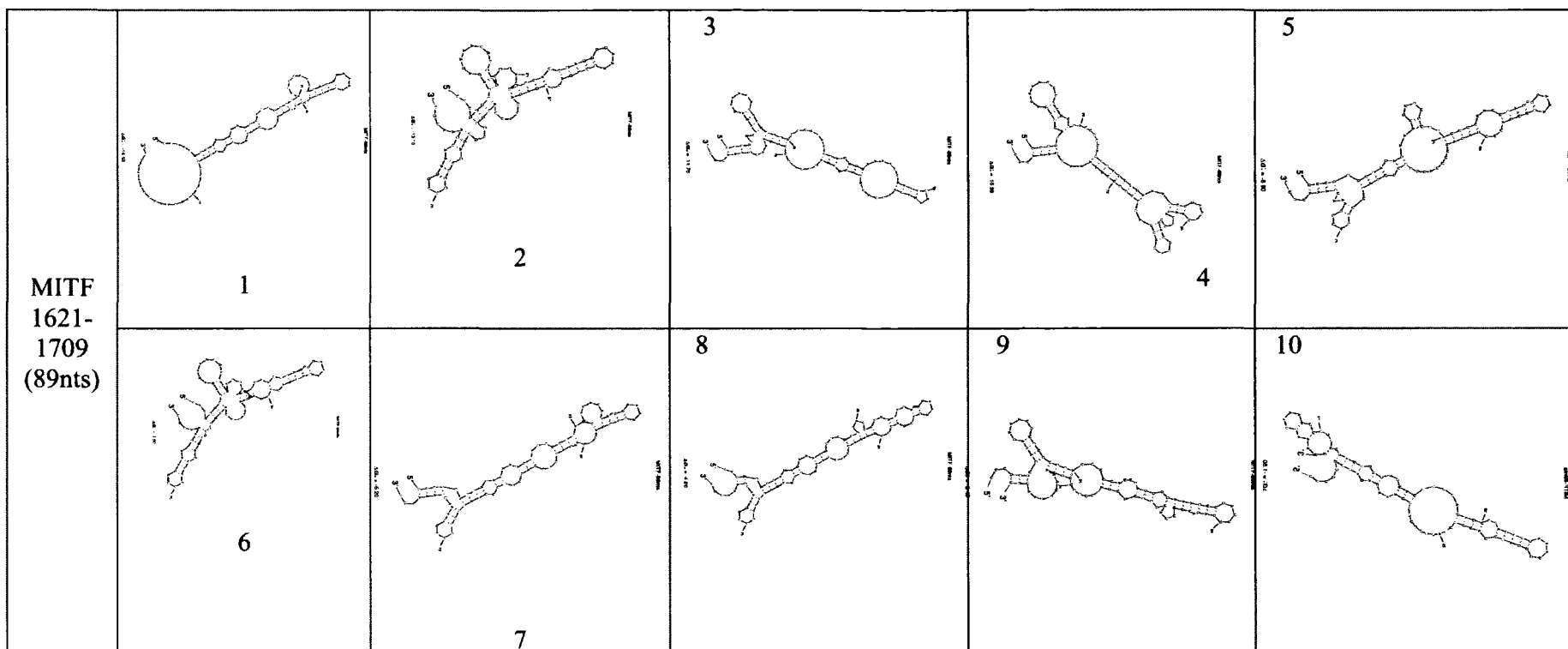
Sfold

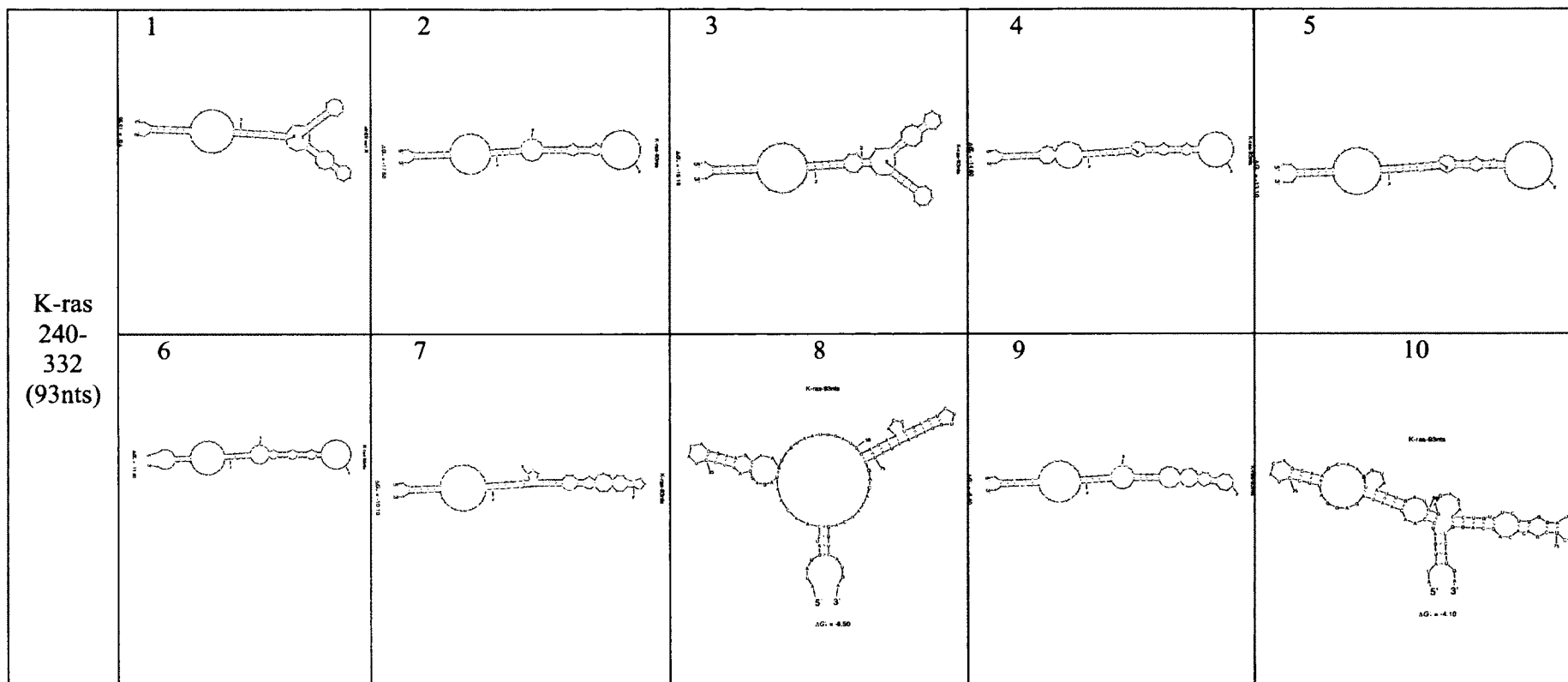
For each sequence input into Sfold statistical RNA folding, ten possible structures are given as outputs with their associated lowest free energy (ΔG) as well as the sample frequency. As shown in Table 19, the first of ten structures for c-myc 1705-1792 (88 nts) RNA had a ΔG of -14.10kcal/mol and 0.499 sample frequency, suggesting this structure as the most likely of the ten predicted. Although the first MITF 1621-1709 (89 nts) RNA structure had the lowest ΔG of -14.50kcal/mol and a sample frequency of 0.192, the

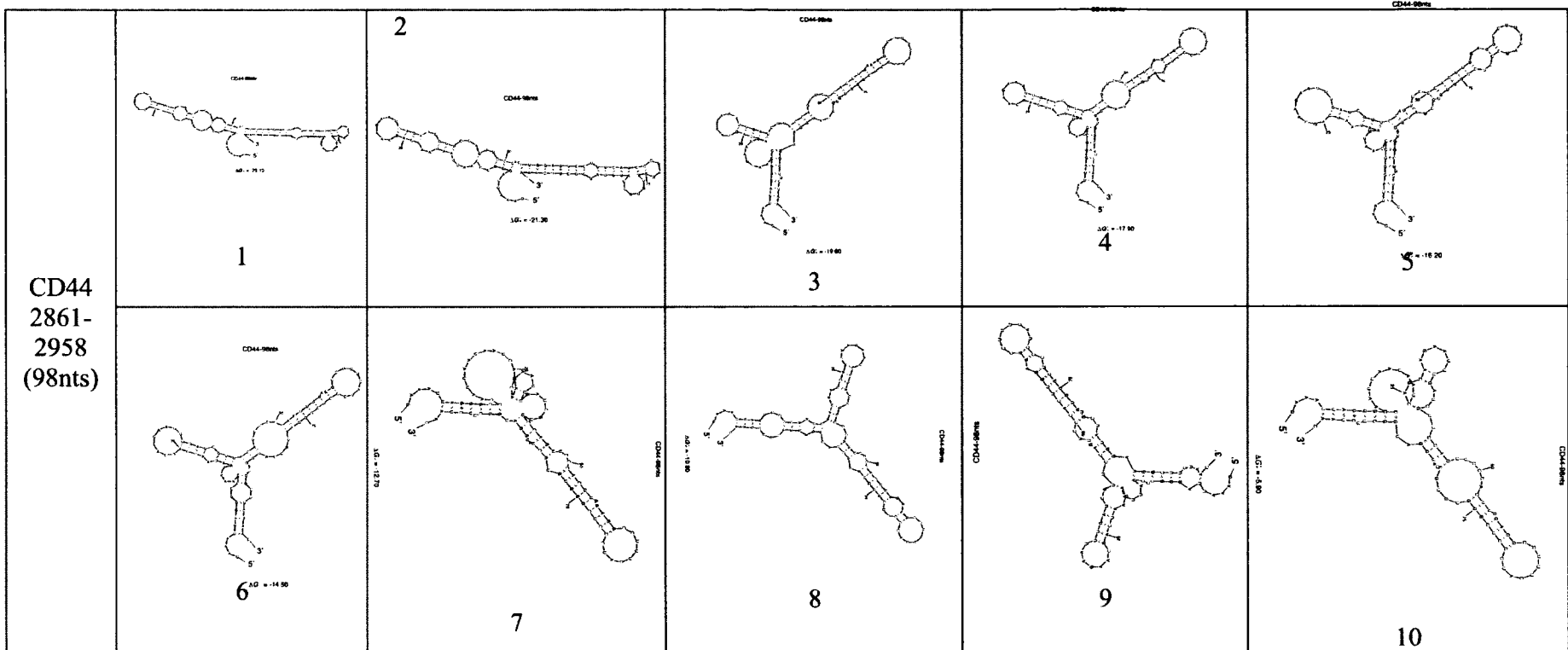
second structure had a greater frequency of 0.226 with a ΔG of -13.10kcal/mol (Table 19). The MITF 1621-1709 (89 nts) RNA structure number three also had a greater sample frequency than that of structure one at 0.212, but had an even lower ΔG of -11.70kcal/mol. From the ten structures predicted for K-ras 240-332 (93 nts) the first three structures had sample frequencies of 0.195, 0.240 and 0.249 and ΔG 's of -19.20kcal/mol, -17.60kcal/mol and -16.10kcal/mol, respectively (Table 19). For CD44 2861-2958 (98 nts) RNA sequence, the second structure had a sample frequency of 0.276 and a ΔG of -21.30kcal/mol in comparison to structure one and three with sample frequencies of 0.118, 0.260 and ΔG 's of -23.10kcal/mol and -19.60kcal/mol respectively. The MDR-1 790-881 (92 nts) RNA structure one had the highest sample frequency of 0.298 and lowest ΔG of -18.80kcal/mol in comparison to structures two and three with sample frequencies of 0.269, 0.212 and free energies of -17.40kcal/mol and -16.10kcal/mol respectively (Table 19). For GLI1 307-393 (87nts) RNA, structure two of ten had the highest sample frequency of 0.301 with a ΔG of -16.50kcal/mol, whereas structures one and three has sample frequencies of 0.207, 0.199 and lowest free energies of -17.80kcal/mol and -15.20kcal/com respectively (Table 19)

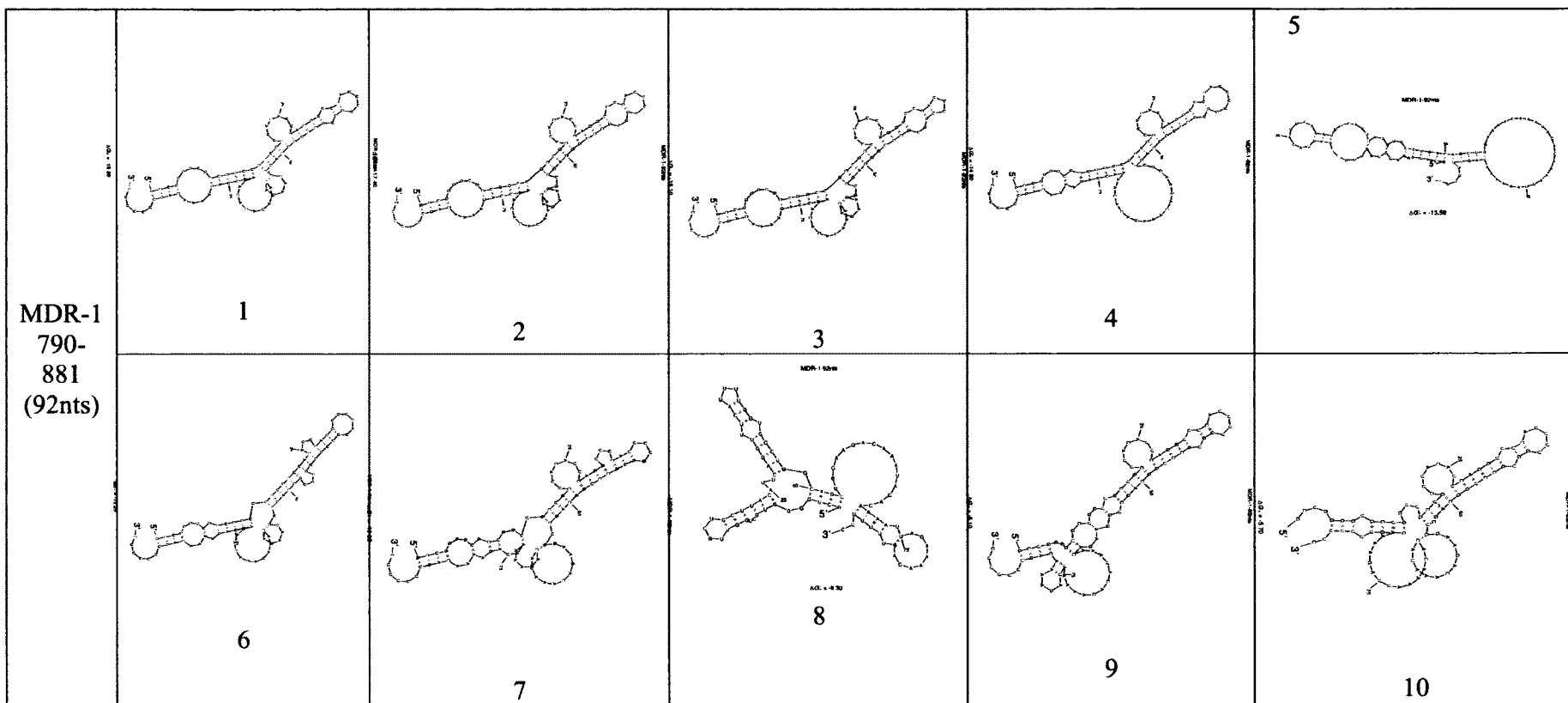
Table 19. Sfold generated secondary structures of the six oncogenic RNAs.

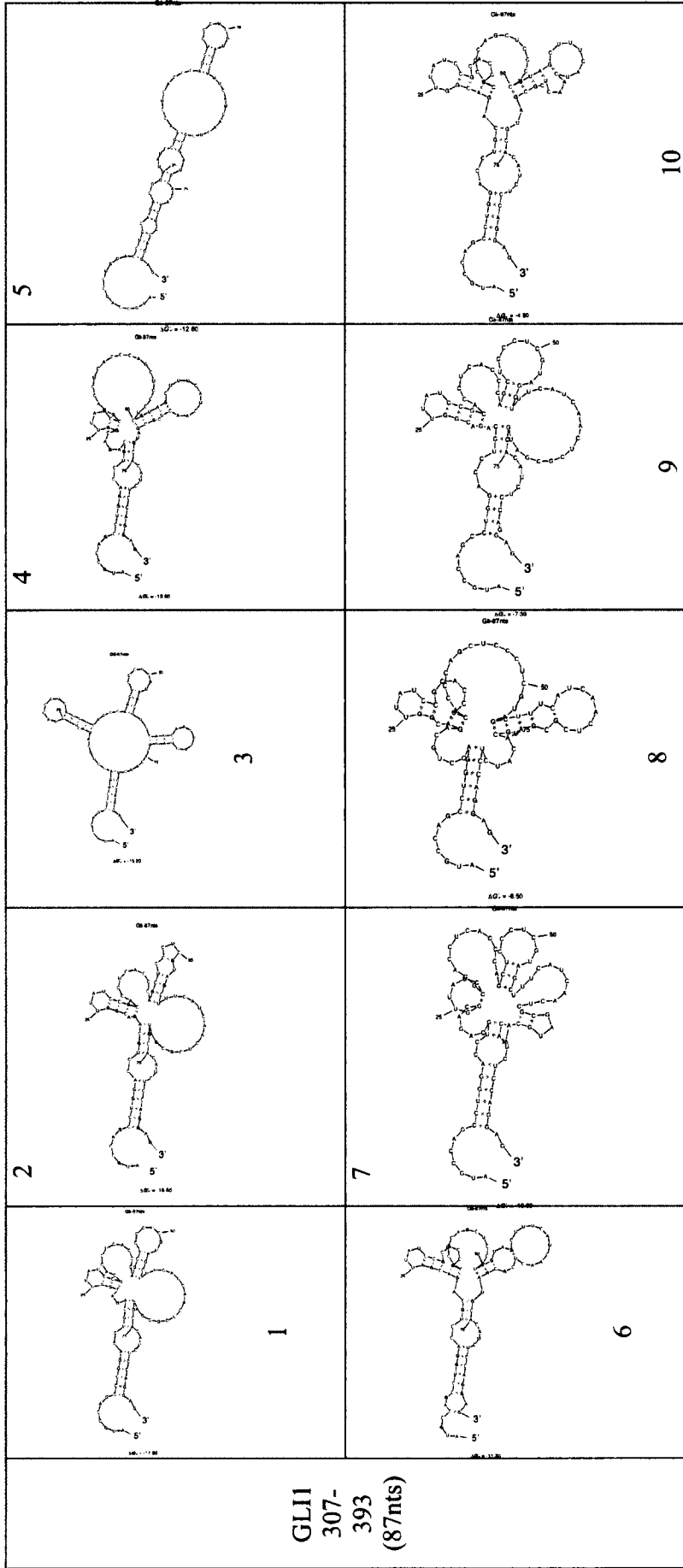
Name	Structure				
	1 	2 	3 	4 	5 
c-myc 1705- 1792 (88nts)	6 	7 	8 	9 	10 











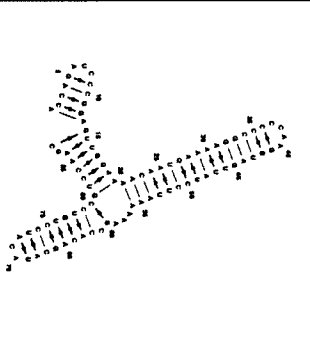
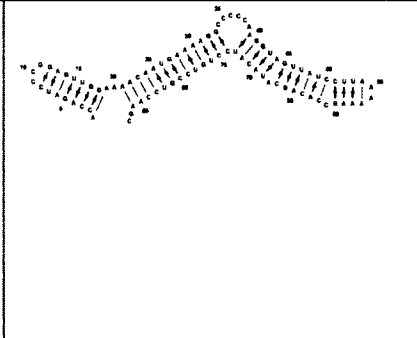
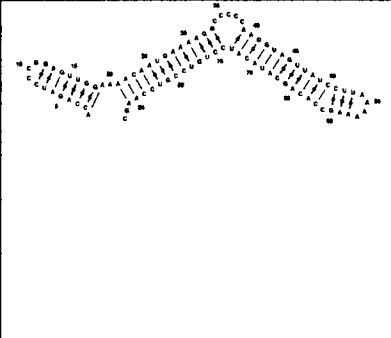

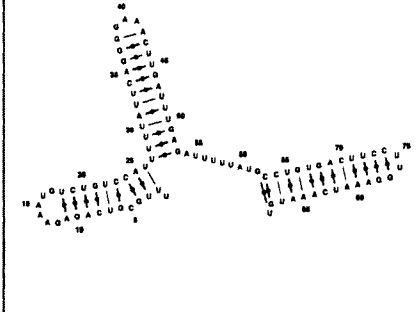
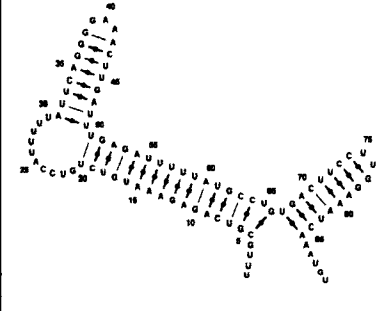
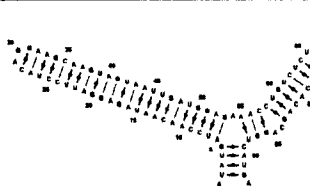
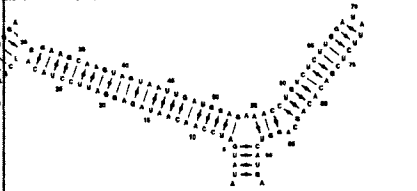
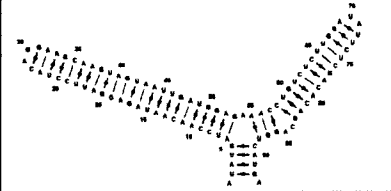
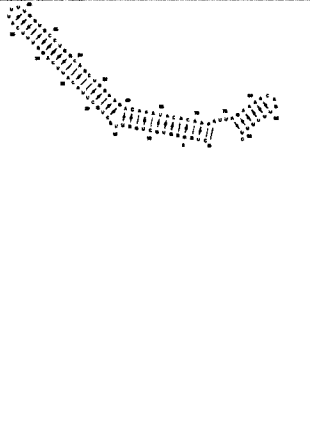
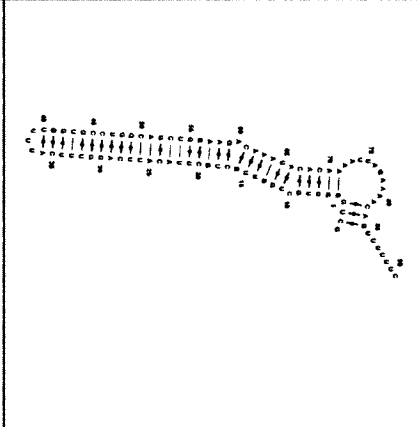
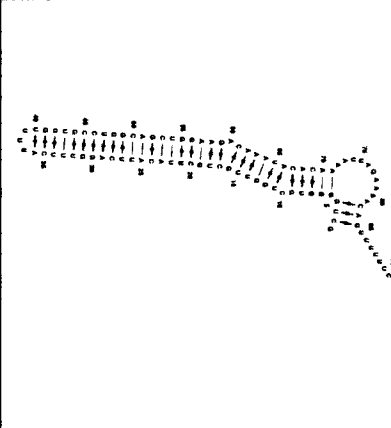
MC-FOLD

From MC-FOLD, the top one hundred structures not containing pseudoknots (no pseudoknots), and top twenty containing pseudoknots were generated. From those, the top one structure of both no pseudoknots and pseudoknots were compiled into a table alongside the top pseudoknot structure, which could be folded into tertiary structure using MC-SYM (Table 20). The non-pseudoknot structure for c-myc 1705-1792 (88 nts) RNA has four stems, three loops and one bulge, with three of the four stems coming together at a three-stem junction (Table 20). The top pseudoknot structure for c-myc 1705-1792 (88 nts) RNA has three stems, two loops and three bulges, but was unable to have a tertiary structure predicted by MY-SYM (Table 20). Pseudoknot structure number four of c-myc 1705-1792 (88 nts) RNA, having one less bulge could be folded by MC-SYM, but predicted MC-FOLD structures numbered one to three could not (Table 20).

The MITF 1621-1709 (89 nts) RNA was predicted to have a three-stem junction with two loops and two bulges (Table 20). The top pseudoknot structure for MITF 1621-1709 (89 nts) RNA does not show a three-stem junction, but retains three stems, and has three loops and one bulge (Table 20). The fifth pseudoknot structure of MITF 1621-1709 (89 nts) RNA with three stems, two loops and one bulge could be folded by MC-SYM into three-dimensional structure, but predicted MC-FOLD structures one to four could not (Table 20). K-ras 240-332 (93 nts) RNA had one consecutive structure as the top structure for both the no pseudoknots and pseudoknots option (Table 20). This conserved structure has three stems, which form a junction, with two loops and one bulge, and also managed to be formed into a tertiary structure by MC-SYM (Table 20). The top non-pseudoknot structure for MDR-1 790-881 (92 nts) RNA has two stems, two loops and

one bulge, in comparison to the top tertiary folded pseudoknot structure with two stems, one loop and three bulges (Table 20). Both CD44 2861-2958 (98 nts) and GLI1 307-393 (87 nts) RNAs were unable to be folded by MC-FOLD, as structures were determined to be too highly branched, and from the available computing power would take too long to resolve (Table 20).

Table 20. MC-FOLD Top 1 from 100 non-pseudoknot and Top 1 from 20 pseudoknot and Top Pseudoknot structure folded in MC-SYM.

Name	Top Non-Pseudoknot Structure	Top Pseudoknot Structure	MC-SYM Computable Structure
c-myc 1705- 1792 (88nts)			
MITF 1621- 1709 (89nts)			
Kras 240-332 (93nts)			
CD44 2861- 2958 (98nts)	Too branched	Too branched	Too branched
MDR-1 790-881 (92nts)			
GLI1 307-393 (87nts)	Too branched	Too branched	Too branched

MC-SYM

The top pseudoknot structures from MC-FOLD were pipelined into MC-SYM to determine possible tertiary structure (Table 20). Once MC-SYM had predicted the possible tertiary structures (not exceeding one thousand structures), each of these structures underwent a four series of minimization, in order to ensure if any discontinuous stretches of RNA are present in the model, that they are placed correctly. The structures were then ranked on Pscore, lowest free energy, van der Waals and electrostatics. Additional analyses such as base-pair interactions and estimated volume was also determined.

The fourth top pseudoknot structure for c-myc 1705-1792 (88 nts) out of twenty could be folded into tertiary structure by MC-SYM (Table 20). MC-SYM provided 127 possible tertiary structures for this one MC-FOLD secondary structure. These 127 structures were then minimized and ranked, highlighting structure number four of 127 because it consistently ranked above the other across the different scoring methods (Table 21). This tertiary structure had a Pscore of -49.50 ranking 1/127, a ΔG of -58.29kcal/mol ranking 2/127, and 6/127 for van der Waals and electrostatic interaction (Table 21).

The fifth top pseudoknot structure of MITF 1621-1709 (89 nts) RNA out of twenty had possible tertiary structure as determined by MC-SYM, with 1000 possibilities (Table 20). Structure three of 1000 ranking consistently highest in the three categories with a Pscore of -59.34 (2/1000), a ΔG of -61.72kcal/mol (4/1000) and 2/1000 for van der Waals and electrostatic interactions (Table 21).


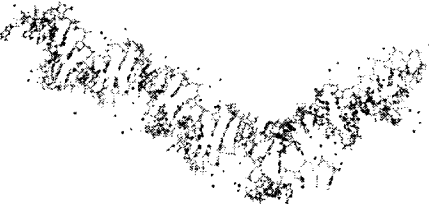

The top MDR-1 790-881 (92 nts) RNA pseudoknot structure of twenty provided three possible tertiary structures (Table 20). Structure one of three was the top structure

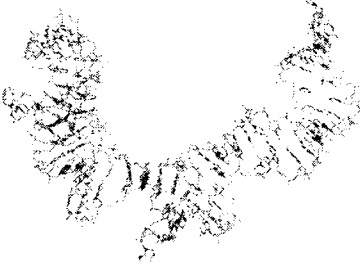
for Pscore (-52.88), lowest free energy (-63.09kcal/mol) and Amber calculation (van der Waals and electrostatic interactions) (Table 21).

MC-SYM predicted 95 structures for the top Kras 240-332 (93 nts) RNA pseudoknot structure (Table 20). The further Kras structure of 95, ranked second for the three analyses with a Pscore of -51.51 and a ΔG of -67.84kcal/mol being consistent in comparison to the structures (Table 21).

MC-SYM tertiary structure determination was not possible for CD44 2861-2958 (98 nts) and GLI1 307-393 (87 nts) RNAs due to the inability of MC-FOLD to provide any secondary structures (Table 20 and 25).

Table 21. MC-SYM Tertiary Structure Information with possible H-type pseudoknots.

Name	Structure	Pscore	Total Energy	Amber Data	Base pair Data	Volume (~ellipsoidal)
c-myc 1705- 1792 (88nts) #4/20		-49.50 (#1/127)	-58.29kcal/mol (#2/127)	-24789.5232 (#6/127)	Tbs = 88 Tprs = 28 Std WC = 20 (71.4%) WWc = 3 (10.7%) HSt = 2 (7.1%)	96,437 Å ³ (#97/127)
MITF 1550- 1750 (89nts) #5/20		-59.34 (#2/1000)	-61.72kcal/mol (#4/1000)	-21062.5423 (#2/1000)	Tbs = 89 Tprs = 31 Std W.C = 20 (64.5%) WWc = 9 (29.0%) HSt = 2 (6.5%)	nan
Kras 240- 332 (93nts) #1/20		-51.51 (#2/95)	-67.84kcal/mol (#2/95)	-22380.3202 (#2/95)	Tbs = 93 Tprs = 39 Std W.C = 24 (61.5%) WWc = 10 (25.6%) HSt = 3 (7.7%)	93,179 Å ³ (#28/95)
CD44 2861- 2958 (98nts)	UD*					

MDR-1 770- 881 (92nts) #1/20		-52.88 (#1/3)	-63.09kcal/mol (#1/3)	-21702.2750 (#1/3)	Tbs = 92 Tprs = 39 Std W.C = 23 (59%) WWc = 11 (28.2%) HSt = 1 (2.6%)	97,157 Å ³ (#1/3)
GLI1 307- 393 (87nts)	UD*					

*UD indicates undetermined due to the inability of MC-FOLD to produce secondary structure that are required to be pipelined into MC-SYM.

LocARNA

From the simultaneous folding and alignment of the six RNA sequences of interest, LocARNA identified a possible shared stem region indicated in the alignment and structure (Table 22). Since LocARNA has a 2000 base-pair limit, 2000 base-pairs surrounding each of the six RNAs regions of interest was also input into LocARNA, and it shows on a larger simultaneous fold and alignment scale that there are conserved regions of secondary structure shared by these RNA (Table 23). In Table 24, a non-binding Kras sequence nts 148-239 (92 nts) was input into LocARNA in addition to the other six RNA sequences, and a shared stem of length 3 was predicted.

Table 22. LocARNA results from the six oncogenic RNAs.


Alignment	Conserved Secondary Structure
<pre> K-ras 93nts AUAUGAUCCAACAAUAGAGGA-UUCCUACAGGAAGCAAG-UAGUAAUUGAUGGA-GAAAC 57 Gli 87nts AUGCCAGCCUGGACCUGCAGA-CGGUU--AUCCG-CACC-UCA-CCCAGCUC---UC 50 MITF 89nts UUUGCUCAGAGAAAUGUCUGUCCAUV--UUUAUUCAGG-GGA-A---ACUUGA---UU 49 C-myc 88nts ACCAGAUCCTGGAGUUGGAAA-ACAAU--GAAAA-GGCC-CCC-A---AGCUAG---UU 47 CD44 98nts GAAAUUAGGGCCAAUUAUAAUACAGC--AAGAAUUUGA-UCG-U---UCCAGUUC--CC 51 HDR-T 92nts GCUG--GGUGCUGGUUGCUGCUACA--UUCAGGUUUCAUUU-U---GGUGCCUG--GC 5010.....20.....30.....40.....50..... K-ras 93nts CUGUCUCUUGGAUA---UUCUC---GACACA-GC-AGGUCAU-G-A 93 Gli 87nts GUAGCUUUAUCA---CUCGC---GAUGCA-CA-UCUCCAG-GAG 87 MITF 89nts UGAGAUUUUUUUGC---CUGUGAC-UUCCUUG-GAAAUCAAU--GU 89 C-myc 88nts AUCCUUAAAAAGC---CACAGCA-UACAUCUUGCCGUCAA--GC 88 CD44 98nts ACUUGGAGGCCUUUCAUCCUCGGUGUGCUAUG-GA-UGGCUUCUAAAC 98 HDR-T 92nts AGCUGGAAGACAAAUACACA--AAAU-U-AGAAA-AC-AGUUUUU--UC 9170.....80.....90.....100..... </pre>	

Table 23. LocARNA results from 2000 bps surrounding the smaller regions of interest of the six oncogenic RNAs. The table includes the alignment and shared secondary structure.

Alignment			Conserved Secondary Structure

Table 24. LocARNA results from the six oncogenic RNAs.

Alignment	Conserved Secondary Structure
<pre> MDR-1_92nts GCUG---GGGUGC-UGGUUGCUGC-UUACA-UUCAGGU-UUCAUUUUGUGCCUGGCAG 52 CD44_98nts GAAAUU--AGGGCC-CAAUUAUAA-UCAGC-AGAUAUU-UGA-UCCGUCCAGUCCAC 53 c-myc_88nts ACCAGA--UCC-CGGAGUU--G-GA-AAACA-AUGAANA-GGGCCCA-MGC-UGUUU 49 Kras_92nts AUGACUGAUA-UAAACUU--GUGG-UAGUU-GGAGCUG-GUGGCUGA-GGC-AGAG-U 51 Gli1_87nts AUGCCA--GCCUGG-ACCUG-CAGA-CGGUUAUCCGCAC-CUCACCCA-GCU-CCCUGU 52 MITF_89nts UUUCCG--UCAGAG-AAAUUCUGU-CUAU-UUUUUC-AGGGAAA-CUU-CAUUU- 50 K-ras_93nts AUUGA--UCCAAC-AAUG-AGGAUUC-CU--A-CAGGAAGCAAGUA-GUA--AUUGAU 4910.....20.....30.....40.....50.....)))..... MDR-1_92nts ---CUGGAAGACAAA--UACACAAA-AU-U---AGAAAACAGUUUUU--UC 91 CD44_98nts ---UUGGAGGCCUUUCAUCCUCGG-GUGU---GCUAUGGAGGCCUUCUAC 98 c-myc_88nts -CCUU--AAAAGGCCA--CAGCAU-AC-A---UCCGUCCUCCAA--GC 88 Kras_92nts -GCCU--UGACGAUACGCUAAUUC-AG-A---AUCAUUUUGGGAC--GA 92 Gli1_87nts -AGCU--UUCAUCAACU--CGC-GAUGC-A---CAUUCUCC--AGG-AG 87 MITF_89nts -GMA-UUUUUAUGCCU--GUGACU-UC-C---U-UGGAAAUCAAU--GU 89 K-ras_93nts GGAGA-AACCUUCUCU---UGGAU-AU-UCUGGACAGCAGGUCAU--GA 9370.....80.....90.....100.....110..... </pre>	

CARNAC

CARNAC has two options to compute the hypothetically conserved RNA secondary structure, without or with a GC threshold, and therefore both options were performed. Without the GC threshold, the clustal alignment did not show any conserved secondary structure, but CARNAC did provide a stem region for c-myc 1705-1792 (88 nts), MDR-1 790-881 (92 nts) and GLI1 307-393 (87 nts) as shown in Table 25. When the GC threshold was enabled, CARNAC found a conserved stem region shared for c-myc 1705-1792 (88 nts) and GLI1 307-393 (87 nts) and a different conserved stem region

for MITF 1621-1709 (89 nts), Kras 240-332 (93 nts), CD44 2861-2958 (98 nts) and MDR-1 790-881 (92 nts) (Table 26). The predicted conserved stems are illustrated alongside the predicted whole structure (converted from Dot-Bracket format using Mfold) (Figures 45 and 47).

```

>Gli_87nts
AUGCCAGCCUGGACCGCAGACGGUUAUCCGCACCCACCCAGCUCGCCUGUAGCUUUCACUACUCGCGAUGCACAUCCACGGAG
.....(((((((.....(((.....)))))).....((((.....)))))).....)))))).. (-17.80)

>MDR-1_92nts
GCUGGGGUGCUGGUUGCUGCUUACAUUCAGGUUUCAUUUUGGUGCCUGGCAGCUGGANGACAAAUACACAAAUFAGAAAACAGUUUUUC
...((.....(((.....(((.....)))))).....))))..(.....)))))).....)))))).. (-16.81)

>K-ras_93nts
AUUUGAUCCACAAUAGAGGUAUUCUACAGGAAGCANGUAGUAAUUGADGGAGAAACCUGUCUCUUGGAAUUCUCGACACAGCAGGUCAUGA
.(((((((.....(((.....(((.....)))))).....))))..((((.....))))..)))))).....)))))).. (-19.2
3)

>C-myc_88nts
ACCGAUCCCGGAGUUGGAAAACAUGAAAAGCCCCCAAGGUAGUUAUCCUUA AAAAAGCCACAGCAUACAUCCUGUCCGUCCAAGC
.....((((.....(((.....(((.....)))))).....))))..)))))).....))))))..... (-14.10)

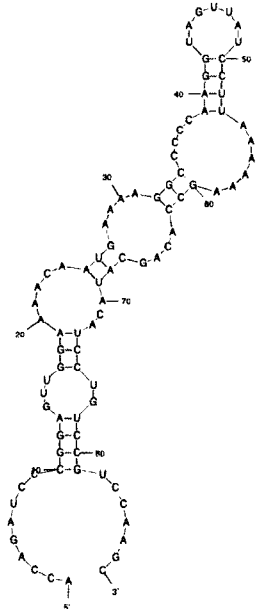
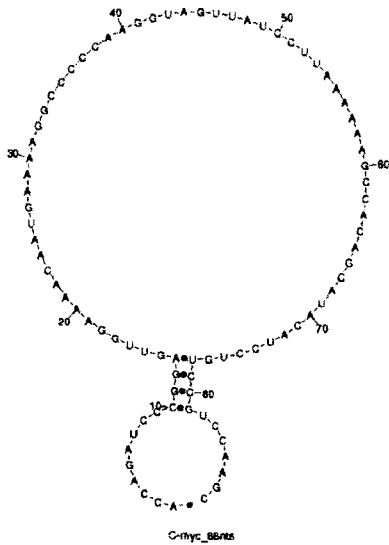
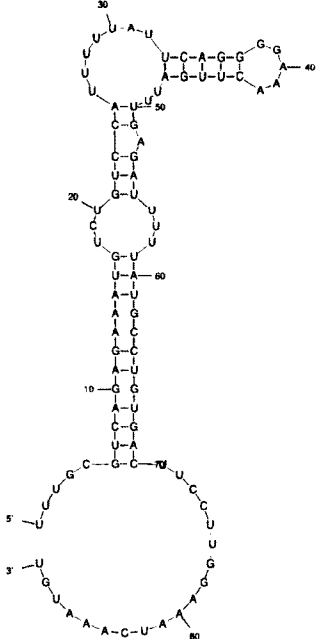
>MITF_89nts
UUUGGUCAGAGAAUUGUCUGUCCAUUUUUAUUCAGGGGAACTUGAUUUGAGAUUUUUAUGCCUGGACUCCUUGGAAAACAAUUCU
.....((((.....(((.....(((.....)))))).....))))..)))))).....))))))..... (-14.50)

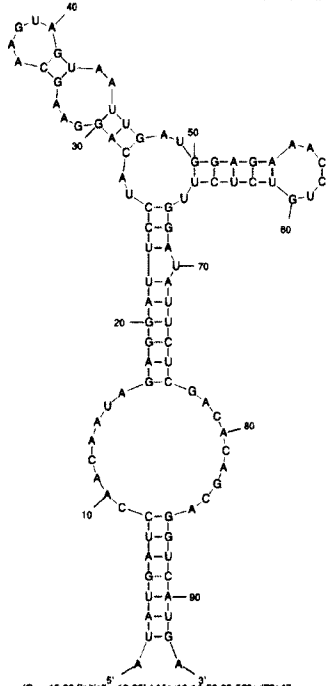
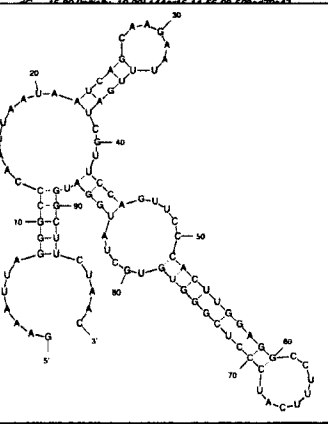
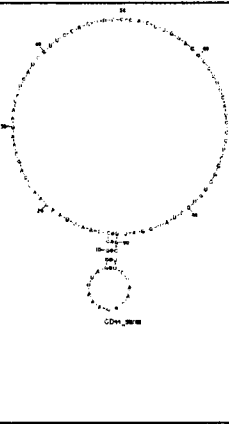
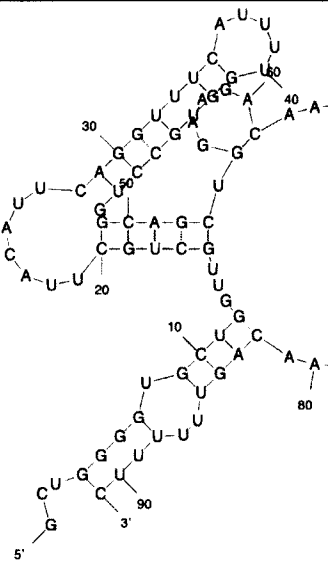
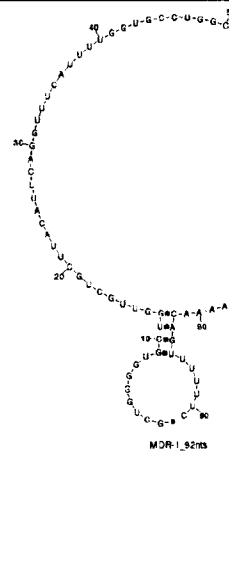
>CD44_98nts
GAAAUUAGGGCCAAUUAUAUAUCAGCAAGAAUUGAUCCGUUCCAGUUCACUUGGAGGCCUUAUCCUUGGGUGGCUAUGGAUUGGCUUUAAC
.....((((.....(((.....(((.....)))))).....))))..((((.....(((.....)))))).....))))))..... (
-19.80)

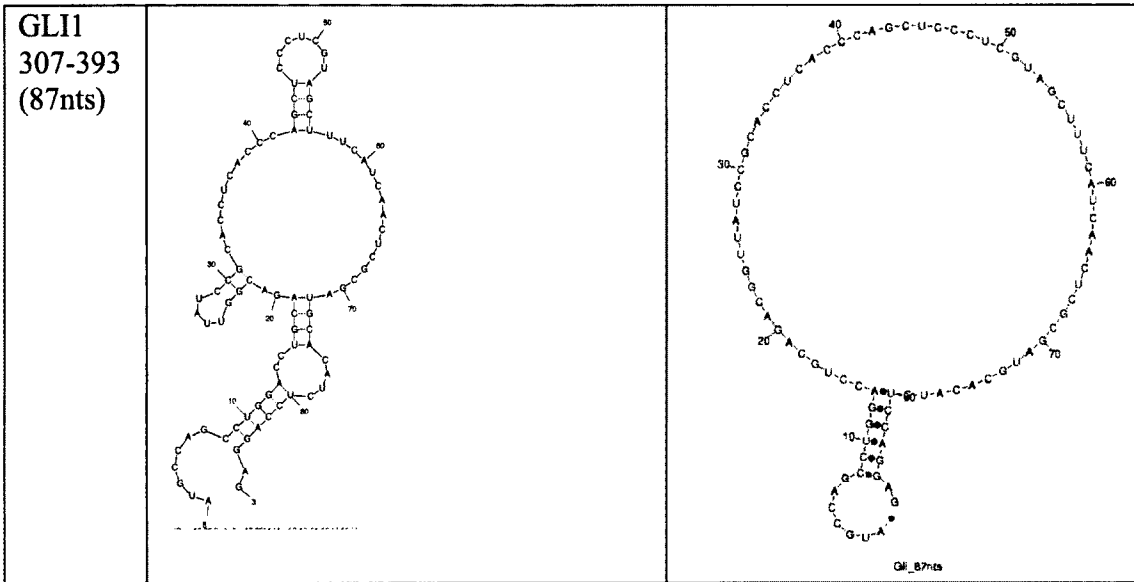
```

Figure 45. Predicted Dot-Bracket Structures from CARNAC without GC Threshold Enabled. These were input into Mfold and displayed in Table 25.

Table 25. Predicted shared stem regions of RNA alongside the predicted secondary structure of RNA without GC Threshold Enabled in CARNAC.

Name	Mfold Structure*	Shared Structure
<p>c-myc 1705- 1792 (88nts)</p>		 <p style="text-align: center;">C-myc_88nts</p>
<p>MITF 1621- 1709 (89nts)</p>		<p>N/A</p>

<p>Kras 240-332 (93nts)</p>		<p>N/A</p>
<p>CD44 2861-2958 (98nts)</p>		
<p>MRD-1 790-881 (92nts)</p>		



* The stem regions and secondary structure was predicted by CARNAC. The structure for the RNA sequences were generated using Mfold from Dot-Bracket Structures.

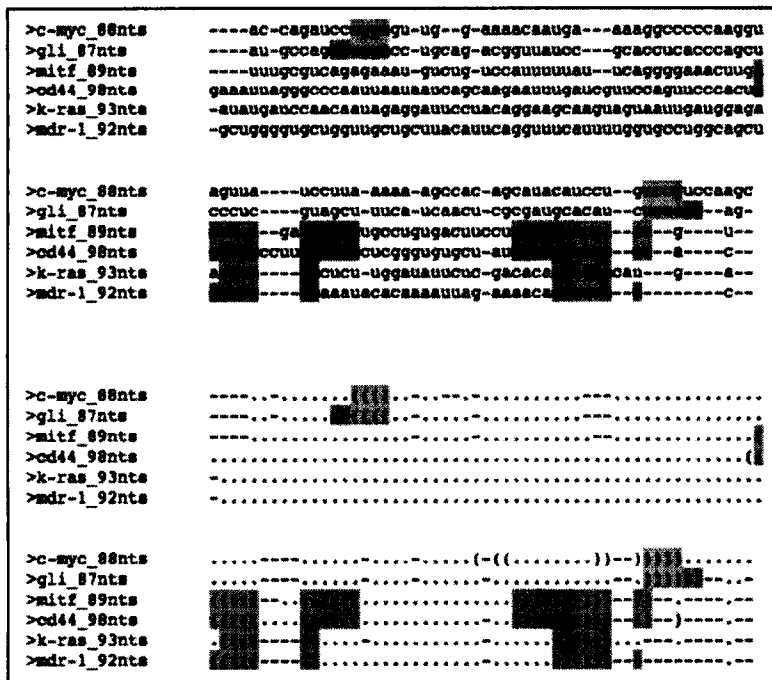
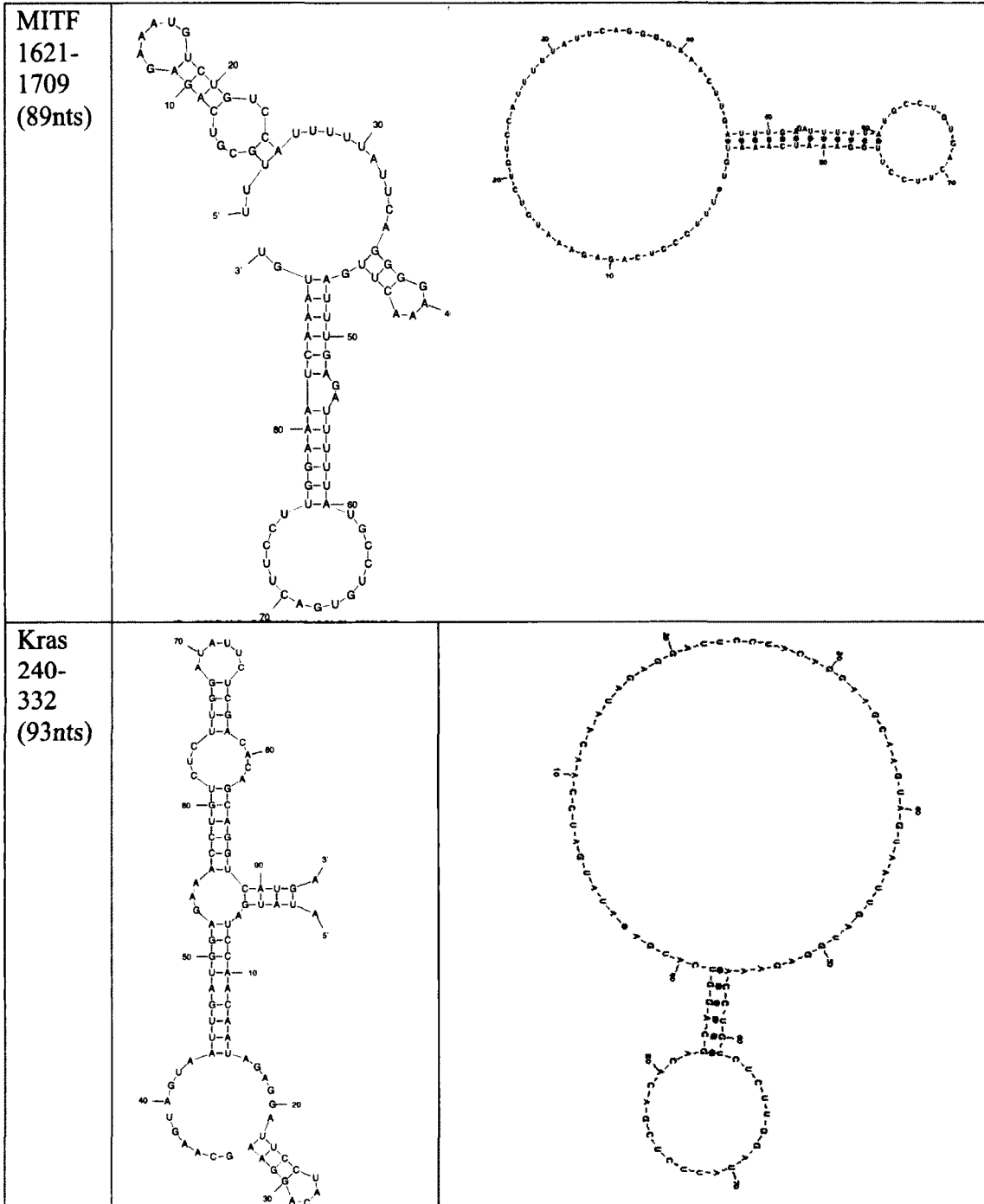
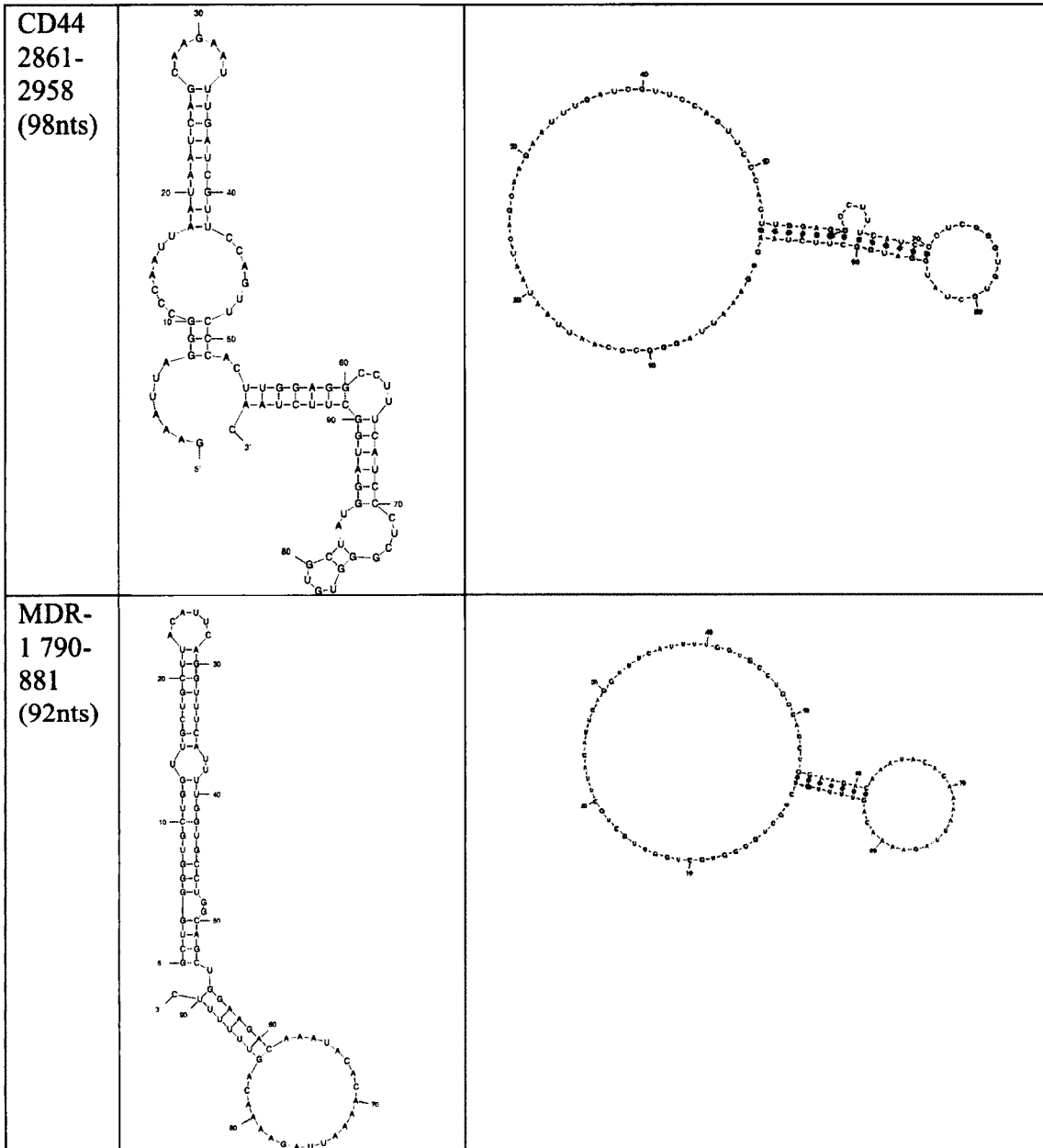
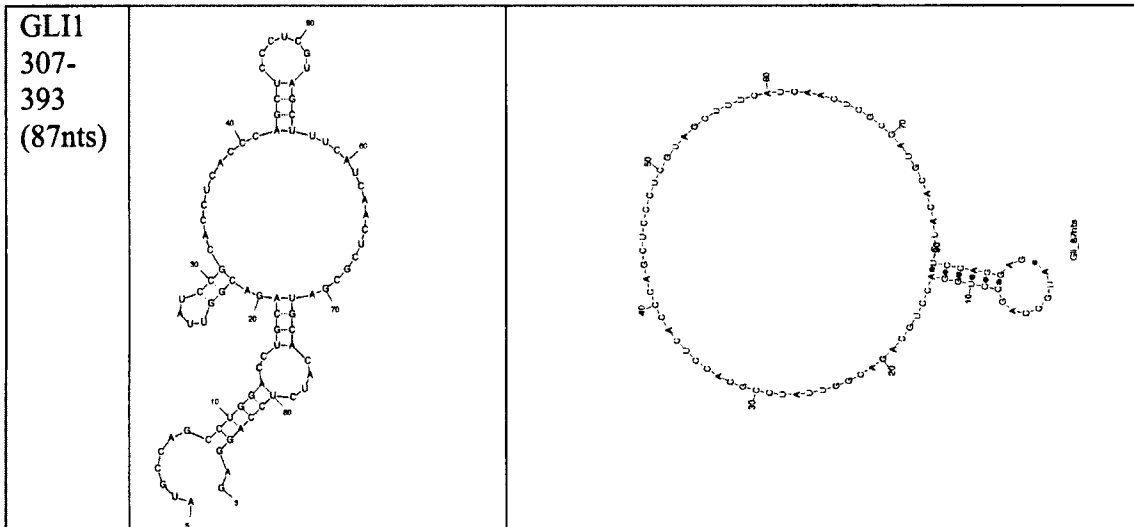


Figure 46. Clustal Alignment by CARNAC with GC Threshold Enable, displaying the nucleotide composition above and dot-bracket secondary structure information below of the six oncogenic RNAs.







*The stem regions and secondary structure was predicted by CARNAC. The structure for the RNA sequences were generated using Mfold from Dot-Bracket Structures.

4.3 Discussion

From the analysis using Mfold and RNAfold, there does not appear to be a single consensus secondary structure for the six RNA sequences being compared. When comparing between the multiple structures generated for a single sequence by Mfold, one can see that the structures with the most stem regions have the lowest free energy, corresponding to highest stability. The results from Sfold suggest that there are multiple structures co-existing and changing between each other, instead of a single structure that each of the RNA sequences occupy. This also indicated that the most stable structure due to energy calculations isn't always the statistically most common structure present either. Between these three folding software, Mfold, RNAfold and Sfold one can also notice the same and very similar structures being produced as representative structures, because these programs are folding the RNA under the same MFE models and biophysical principles.

For c-myc 1705-1792 (88 nts) the exact same structure was produced for Mfold, RNAfold and the first structure in Sfold. The shared stem structures are at nucleotides

10-13 with nucleotides 78-81, nucleotides 17-19 with nucleotides 73-75, nucleotides 25-27 with nucleotides 67-69, nucleotides 32-34 with nucleotides 60-62 and nucleotides 39-42 with nucleotides 50-53 (Tables 21-23, S1 and S2). The shared bulges between Mfold, RNAfold and Sfold of c-myc are located at nucleotides 14-16, 28-31, 35-38, 54-59, 63-66, 70-72 and 76-77 (Tables 21-23, S1 and S2). One shared end loop was located at 43-49 and two free ends at the 5' and 3' ends respectively nucleotides 1-9 and 82-88 (Tables 21-23). The two most conserved stem regions that were maintained in the ten different structures folded in Sfold were at nucleotides 32-34 and 39-42, which consistently maintained the adjacent bulges at nucleotides 35-38, 54-59 and the one end loop at 43-49 (Tables 21-23, S1 and S2). The consensus structure for c-myc 1705-1792 (88 nts) between these three programs is then the first structure generated by all three programs. This structure contains four stem regions, eight bulges, one stem loop and two free ends with a ΔG of -14.10kcal/mol and from Sfold a 0.499 sample frequency (Tables 21-23). Previous research using enzymatic digestion to determine secondary structure of c-myc RNA 1705-1792 validates the structure that has been generated by Mfold, being the same as Sfold and RNAfold (Figure S1) (5).

For MITF 1621-1709 (89 nts) the first Mfold structure of five was identical to the first Sfold structure (Tables 21 and 23). The RNAfold structure was identical to the third Sfold structure with an additional stem in the Sfold structure at nucleotides 1-4 with 83-86 (Tables 22 and 23). Between these two structures, (the first Mfold/Sfold and third Sfold/RNAfold structures) there were no shared stem regions, but a similar stem morphology arose making a symmetrical G-bulge in the middle of a stem region (Table 17 and 23). Shared identical bulges by these two structures were at nucleotides 26-32

and 66 (Table S2). Shared unpaired nucleotides were 1-5, 57-58, 62, 70-75, 80-89 and morphological sequence at two different stems with a sequence of GAAA at nucleotides 39-42 and 12-15 in the Mfold/Sfold structure 1 and RNAfold/Sfold structure 3 respectively (Tables 21-23, S1 and S2). From the additional Sfold information, Sfold structure 1 has a sample frequency of 0.192, being a little lower than 0.212 of Sfold structure 3. There is no consensus between these two structures other than sharing a symmetrical G-bugle in the middle of a stem.

For the Kras 240-332 (93 nts) structures, the second Mfold structure, is almost identical the RNAfold structure with one additional unpaired nucleotides at 34 and 61 (Table 17 and 22). This Mfold structure shares almost all of the same unpaired and paired nucleotides with the second Sfold structure (Table 17 and 22). Two differences are the unpaired nucleotides at 34 and 61, as well as paired nucleotides at 39 with 56 in the Mfold structure, which is not so in the Sfold structure (Table 17 and 22). The stem regions shared by these three very similar structures encompass nucleotides 2-8 with 86-92, 17-22 with 71-76, and 28-39 with 56-67 (Table S1). The shared bulges are nucleotides 9-16, 23-27, 33, 37, 58, 62, 68-70, and 77-85 (Table S2). There was one stem loop, for all three structures, which remained the same, spanning nucleotides 40-55 (Table 17-23 and S2). All three structures also retained a single adenine nucleotide at both the 5' and 3' ends (Tables 17-23). Between these three programs, a consensus structure for Kras 240-332 (93 nts) became apparent, with three stems, six bulges and a single stem loop (Tables 21-23).

For the CD44 2861-2958 (98 nts) the first Mfold structure has a single difference from the RNAfold structure, being unpaired nucleotides 20 with 40 (Tables 17 and 22).

This Mfold structure also has a single difference from the first Sfold structure, being unpaired nucleotides 13 and 46 (Tables 21 and 23). The conserved stems shared by these predicted structure are at nucleotides 8-10 with 49-51, 18-26 with 34-42, and 54-70 with 75-97 (Table S1). The single stranded bulge and stem loop nucleotides shared by all three structures are at nucleotides 1-7, 11-12, 14-17, 21, 27-33, 43-45, 47-48, 52-53, 63, 88, and 98 (Table S2). From Mfold, Sfold and RNAfold, CD44 2861-2956 (98 nts) has a consensus structure composed of two major stems broken up with eight bulges and two stem loops (Tables 21-23). The first Mfold structure of CD44 2861-2956 (98 nts) is also in agreement with experimentally determined structure (Figure S2) (6).

The first Mfold and Sfold structures for MDR-1 790-881 (92 nts) are identical, where as the third Sfold structure shares all the same features as the RNAfold structure with the addition of a base pair at G56 with C62, which is also seen in Mfold and Sfold structure one (Tables 21-23). The shared stem regions generated for these similar structures are nucleotides 2-5 with 83-86, 11-16 with 73-78, and 17-36 with 42 to 54 (Table S1). Conserved bulges for MDR-1 generated structures were at nucleotides 6-10, 22-28, 34, 44, 55, 57-61, 63-72, and 79-82 (Table S2). A shared stem loop region of UUU at nucleotides 38-40 is maintained and two free ends at nucleotides 1, and 87-92.

For GLI1 307-393 (87 nts) the Mfold and first Sfold structures are identical, and the RNAfold structure is similar with the introduction of an additional stem region (Tables 21-23). This extra stem region present in RNAfold comes from pairing of nucleotides 59-62 with 70-73 and pairing 76-77 with 16-17 instead of having 59-62 and 76-77 unpaired while 16-19 is paired with 72-75 as seen in Table 3. Shared stem regions between these two structures are nucleotides 8-17 with 74-85, nucleotides 22-24 with 29-

31, and nucleotides 42-45 with 53 to 56 (Table S1). Shared bulges for GLI1 are nucleotides 14-15, 18-21, 32-41, 57-58, and 78-79 (Table S2). Free nucleotides shared by both structures for GLI1 are 1-7, 63-69, and 86-87 as in Table 17, 22 and 23. Conserved stem loops were 25-28 and 46-52 (Table S2).

When comparing the generated structures of all six RNAs between each other using Mfold, Sfold and RNAfold, all the RNAs formed these long “rod-like” structures with one or two stem loops (Tables 21, 22, 23). GLI1 307-393 (87 nts) and CD44 2861-2956 (98 nts) could be exceptions to this having the possibility of forming three stem loops as seen in the RNAfold structures (Table 18). Another similarity between all these structures is the majority of the free nucleotides (making them available to hydrogen bond with protein) mainly consist of adenine, cytosine, and uracil, which is what has been reported earlier (Table 15, Table S2).

The MC-FOLD software strays away from these MFE models in a way by accounting for non-canonical base pairs, which is not account for by Mfold, RNAfold and Sfold. Instead MC-FOLD uses statistics on nucleotide cyclic motifs from the Protein Data Bank of known crystallized RNA motifs, explaining why the results generated are different by comparison to that of Mfold, Sfold and RNAfold. Although the results are somewhat different, again the overall structures for the RNA maintain these “rod-like” structures (Table 20). For CD44 2861-2958 (98nts) RNA and GLI1 307-393 (87 nts) RNA are determined “too highly branched” for MC-FOLD to output, and this branching can also be seen in the RNAfold structure of GLI1 307-393 (87 nts) (Table 20). The MC-FOLD structures have stem loop regions mostly consisting of two to three nucleotides where as the MFE based programs rarely predict stem loops less than four nucleotides

(Tables 21-24). The MC-FOLD predicted structures also consistently contained less and smaller bulge regions, in comparison to the MFE-based generated structures (Tables 21-24).

Interestingly, from the tertiary data on all four structures generated indicate some form of an L-type structure (Table 21). For both MITF and Kras RNAs, this L-type structure is very evident, and looks very similar to each other (Table 21). This L-type structure can also be identified in the c-myc RNA structure with an additional helical stem loop protruding from one side (Table 21). For MDR-1, this L-type structure is more exemplified as a curved V, but again illustrates a similar overall type of fold (Table 21).

Both LocARNA and CARNAC indicated that these six RNAs sequences unrelated in gene product and evolutionary history, do share elements of conserved secondary structure. LocARNA indicated this with a single stem that all six RNA sequences can form (Table 22). However, the single shared stem generated by LocARNA arise after gaps are inserted into sequences and LocARNA tries to account for all the input RNAs to share some feature (Table 22). The stem predicted by LocARNA was not present in any of the predicted secondary structures. The results of LocARNA are weakened by the additional of gaps within sequences and having the inability to distinguish similar and dissimilar RNAs into distinct groups (Tables 26 and 28). The addition of a non-binding RNA sequence, Kras 148-239 (92 nts), illustrates that LocARNA is unable to distinguish which RNA sequences should be grouped together before folding, but instead will add gaps to the sequence alignment until the primary sequences can account for a shared secondary structure feature.

The CARNAC results indicated that there is a conserved stem shared by c-myc 1705-1792 (88 nts), and GLI1 307-393 (87 nts) with or without the GC threshold present (Table 24). A shared stem for MDR-1 790-881 (92 nts) was also made in the absence of the GC threshold, but from the alignment there was no indication of any conserved structure (Figure 45, Table 25). With the GC threshold enabled, MDR-1 790-881 (92 nts) shares a conserved stem with MITF 1621-1709 (89 nts), Kras 240-332 (93 nts) and CD44 2861-2958 (98 nts) as a separate group from c-myc 1705-1792 (88 nts) and GLI1 307-393 (87 nts) (Figure 46, Table 26). CARNAC also produced dot-bracket format of the entire RNA secondary structure it predicted which was manually input into Mfold (Figure 45 and 47) and illustrated against the predicted shared stem regions (Table 25 and 30). The shared stems that are generated indicate that after an alignment of the primary sequences, stem regions at the same nucleotide numbers form stems. This information is limited to producing useful results only when the RNAs are going to form identical or very similar secondary structures. In the case where only a single stem region is shared at the same nucleotide position, no other useful information is gained, as these stems do not share the same chemical information.

5. General Discussion

5.1 General Overview

To understand the functions of CRDBP during embryonic and cancer development through interaction with RNA via its four KH domains, we aimed at investigating the targeted RNA molecules as well as identifying which of the four KH domains are important. Understanding the RNA molecules that interact with CRDBP began by mapping smaller RNA regions of MITF 1330-1740 (411 nts), c-myc 1705-1886 (182 nts), MAPK 2764-4713 (1949 nts), and β -TrCP1 (863 nts). Knowledge of these smaller regions of RNA was then used *in silico* analysis, and in the design of antisense oligonucleotides (AONs) as inhibitors of CRDBP-MITF RNA interaction. In addition, such knowledge should pave the way for future development of fluorescent polarization method to study CRDBP-RNA interaction and to screen for inhibitors of CRDBP-RNA interaction. The second goal of this investigation was to understand how the four KH domains interact with each other in facilitating CRDBP binding to RNA targets. The significance of the KH domains was investigated by introducing a single point mutation at the first glycine in the G-X-X-G motif to an aspartate (D-X-X-G). In general, this study is expected to expand our current knowledge on how CRDBP and other KH domain-containing RBPs function and provide future researchers with additional tools and insights into the design of compounds (nucleic acid-based and small organic molecules) for the inhibition of protein-RNA interaction.

5.2 Mapping RNA

From the RNA mapping studies, the smallest region of c-myc RNA that still binds CRDBP is 1735-1792 (58 nts). The smallest MITF RNA that retained binding to CRDBP is 1621-1740 (112nts). From the 1949 nt MAPK 3'-UTR region, four regions were identified for having the ability to bind CRDBP, 2764-3089 (326 nts), 3723-4063 (341 nts), 4036-4386 (351 nts) and 4368-4713 (346 nts). As for the β -TrCP1 RNA, the smallest region showing binding with CRDBP is 1131-1219 (89 nts).

The knowledge of mapped small RNA regions from several RNAs that bind CRDBP, including Kras RNA performed in Dr. Lee's lab (Mackedenski S, unpublished observation) has assisted *in silico* analysis in the present study. The *in silico* analysis is discussed in greater detail in section 5.6.

The smaller regions of MITF, c-myc, MAPK, and β -TrCP1 RNAs that have been mapped will further aid in future studies aimed at developing a high-throughput fluorescent polarization method to screen libraries of small molecule for ability to inhibit the CRDBP-RNA interaction. The fluorescent polarization method requires that small RNA be fluorescently-tagged for use as substrate. In fact, a recent study showed that small molecules can be screened for ability to break protein-RNA interaction using fluorescent c-myc RNA as substrate in fluorescent polarization method (122).

Another type of CRDBP-RNA inhibitor, that also takes advantage of the knowledge of smaller regions of RNA, is antisense oligonucleotides (AONs). The design of AON is based on the knowledge of smaller region of RNA that's binds CRDBP. These AONs interrupt the binding of CRDBP by targeting the RNA of interest via the formation of dsRNA segments. These dsRNA regions are expected to hinder the normal

RNA interaction with CRDBP's KH domains. Alternatively, such dsRNA region may be recognized and cleaved by enzymes, as was demonstrated for the siRNA-mediated pathway. Two research groups have indeed demonstrated that AONs are capable of breaking the CRDBP-RNA interaction. One group was targeting the CRDBP-c-myc RNA interaction and the other was targeting the CRDBP-CD44 RNA interaction (107, 123).

5.3 Investigating the Role of KH Domains of CRDBP in Binding RNA Targets *In Vitro* and in Cells

Prior research has shown that VICKZ family members interact with RNA through the KH domains as opposed to the RRM domains (39, 41, 72, 76). The manner in which the KH domains interact with RNA has been determined by studies on other KH-domain containing RBPs (2, 50, 72, 124). The studies illustrated the importance of the flexible G-X-X-G motif, and how the introduction of a negatively charged aspartate either at the first position (D-X-X-G), at the second and third position (G-D-D-G), or two negatively charged glutamate at the second and third position (G-E-E-G) of this motif interrupted the ability of the KH domain to bind RNA (112, 125, 126). Understanding which KH domains of CRDBP that interact with a specific RNA should provide valuable information on how CRDBP interacts with RNA in general. Such knowledge is also useful for future development of drugs that specifically target KH domains known to be important for interacting with specific RNA molecules.

The c-myc 1705-1886 (182 nts) RNA fragment lies within the CRD region and was one of the first RNAs to be shown to have high affinity for CRDBP (56). The ability

of CRDBP to bind c-myc 1705-1886 (182 nts) RNA fragment was re-confirmed using EMSA and was compared to a panel of CRDBP G-X-X-G KH variants. Amongst the KH variants with a single mutation, KH1 showed a modest 2-fold reduction in binding with the 182 nts c-myc RNA as compared to the KH2, KH3, and KH4 variants, as well as the WT-CRDBP (Table 27).

Dr. Lee's lab has previously assessed the variants with double KH mutations and found that CRDBP-KH1-2, KH1-3, KH1-4, KH2-3 and KH2-4 all exhibited no binding to c-myc 1705-1886 (182 nts) RNA, whereas CRDBP-KH3-4 showed binding profile which is similar to that of the WT-CRDBP (Table 27) (109). This loss of binding upon mutation at two KH domains reflects what has been previously reported for how the RBPs ZBP1, PCBP2 and FMR1 KH domains function as a pair. The ZBP1-RNA interaction has been previously described to utilize KH3 and KH4 domains in tandem to bind RNA (72). This was revealed by x-ray crystallography study, where the KH3 and KH4 domains of the truncated ZBP-1-KH-3-4 form an anti-parallel pseudodimer that acts as a binding surface for RNA (72). Chao *et al* also postulated that this pseudodimer would be present in the other VICKZ proteins as they have such a high degree of conserved amino acids (72). Interestingly, both PCBP2 and FMR1 also adopted intramolecular forces between their KH1 and KH2 domains in a similar anti-parallel pseudodimer topology (2, 124).

The MITF 1550-1740 (191 nts) RNA fragment showed binding to the KH1 variant, although the binding coefficient could not be determined due to the relatively weaker binding and hence the inability to fit into a typical binding curve (Table 27). The CRDBP-KH3-4 bound to MITF RNA with similar affinity as the WT-CRDBP (Table

27). The remaining CRDBP KH variants (KH2, KH3, KH4, KH1-2, KH1-3, KH1-4, KH2-3, KH2-4) showed no binding affinity to the MITF 1550-1740 RNA (191 nts) (Table 27). Such results suggested that regions other than the G-X-X-G motif in the KH3 and KH4 domains co-operatively function to interact with the MITF 1550-1740 RNA (191 nts). The data also suggested that all four KH domains play an important role in binding to the MITF RNA, and this appears to contradict the notion that the protein functions with two didomain KH groups (KH1-2 and KH3-4), and that KH1 and KH2 work with each other in binding MITF RNA. An interesting observation made here was that while both CRDBP-KH3 and CRDBP-KH4 variants showed no binding with MITF RNA, the CRDBP-KH3-4 variant surprisingly exhibited binding to the same RNA. It is possible that regions other than the G-X-X-G motif in the KH3 and KH4 domains can somehow cooperate with each other to facilitate binding under situations where the G-X-X-G motif in both KH3 and KH4 domains are mutated. Further experimentation is clearly required to test this hypothesis.

Dr. Lee's lab had previously examined the binding constants for CD44 and GLI1, and showed that these two RNA molecules also bind in a different manner as compared to c-myc and MITF RNAs (Table 27) (108). The comparison of the binding constants from all four RNA fragments shows that CRDBP binds uniquely to each of the individual RNA molecules by utilizing different KH domains (Table 27). CRDBP interacts with different RNA molecules with various combinations of the KH domains, and this is in line with the role RNA structure plays in protein recognition. This is also in congruent with the idea that if primary sequences (recognized by KH domains) of RNA molecules are orientated so two or more KH domains are involved, a binding event would occur.

This would also suggest there are minimal changes that need to place in the structures of both CRDBP and RNA when a binding event is taking place. Such a model is also consistent with the CD spectral data, which showed no noticeable change in RNA structure.

Table 27. A Summary of the Dissociation Constant K_d of Four RNAs that Bind CRDBP.

Variant	c-myc	MITF	CD44 ^a	GLI1 ^a
KH1	723.07 ± 75.13	UD	211.18 ± 34.94	UD
KH3	320.81 ± 42.75	No Binding	219.32 ± 19.35	466.18 ± 136.08
KH1-2	No Binding	No Binding	No Binding	No Binding
KH1-4	No Binding	No Binding	UD	No Binding
KH2-4	No Binding	No Binding	No Binding	No binding

^aAdaptation from Table 9 in the MSc thesis of Mr. Kashif Mehmood (108).

UD denotes undetermined and symbolizes CRDBP variants that retained partial binding ability and could not have a K_d constant determined.

Using IP coupled with qRT-PCR method, Dr. Lee's lab had previously shown that there is a reduction in the binding of CRDBP single KH variants to c-myc and CD44 mRNAs in cells (109). As shown in Chapter 3, the CRDBP double KH variants exhibited even greater decrease in binding to the c-myc and CD44 mRNAs (Figure 37). An explanation for the differences observed between the inability of CRDBP KH variants to bind c-myc and CD44 mRNA in cells as compared to those single KH variants and CRDBP-KH3-4 variant that showed binding *in vitro* is then warranted. One would have to understand the complexity of what is occurring in cells as opposed to the simplicity of an *in vitro* system where only the protein and RNA molecule of interest are interacting. As described by Nielsen *et al*, IMP1 has been shown to cooperatively and sequentially

bind to RNA substrates in dimers in the process of forming ribonucleoproteins (RNPs) complexes; $\text{RNA} + 2\text{Protein} \leftrightarrow \text{RNA-Protein}_1 + \text{Protein} \leftrightarrow \text{RNA-Protein}_2$ (45). This RNP complex aids in the stabilization between IMP1 and RNA substrate for cellular localization (45). If the interaction between CRDBP and the RNA are defective on a small scale, it could then prevent the formation of RNP complexes and result in defective binding required in cells.

5.4 The Attempt to Use Zebrafish Model to Understand the Significance of KH domains of CRDBP in Conferring an *In Vivo* Phenotype

One of the goals of this investigation was to correlate CRDBP function with phenotypic changes in *Danio rerio* (zebrafish). The zebrafish model was selected over other models because: (i) the ease in visualizing the embryos and adult zebrafish; (ii) a shorter time period for fish development that leads to rapid generation of data; (iii) having colleagues with expertise in zebrafish experimentation; and (iv) most importantly, there was a previously observed phenotype, namely a reduction in pigmentation, in zebrafish upon injection of a specific morpholino against CRDBP in *D. rerio* (30). Upon injection of the morpholino, Goswami *et al* reported observations of decreased levels of *mitfa* (zebrafish MITF) mRNA which was followed 48 hours later by a reduction in cell pigmentation in zebrafish (30). For the purpose of this study, I made a number of selected CRDBP variant constructs in anticipation of studying their role in the zebrafish model. But unfortunately, the data published by Goswami *et al* could not be reproduced. It was discovered that the morpholino sequence designed by Goswami *et al* targeted IGFBP1 (Insulin-like growth factor-binding protein 1) mRNA and not the IGF2BP1 (CRDBP) mRNA. Furthermore, when a newly designed specific morpholino

against zebrafish IGF2BP1 was injected into zebrafish embryos, no changes in pigmentation nor any observable phenotypes can be seen in the zebrafish. Hence, the earlier proposed investigation could not be carried out.

5.5 An Attempt to Use Circular Dichroism Spectroscopy to Understand CRDBP-RNA Interaction

The CD Spectroscopy was used to examine the possible CD spectral changes in RNAs upon interaction with CRDBP. Using two ratios of CRDBP to RNA concentration, no observable in CD spectra of RNA was seen. If any observations were made by this technique it would at best indicate that some interaction is taking place between CRDBP and RNA, by having changes present in the generated spectral graphs. Such changes, if any, would not provide detailed information on how these two molecules interact. Rather, it would only complement the *in vitro* EMSA data that already illustrates there is an interaction between the protein and RNA.

5.6 In Silico Analyses of the Structure of RNAs that Bind CRDBP

From the *in silico* analysis using Mfold, RNAfold and Sfold, it is evident that RNA molecules do not form static structures, but instead multiple structures could co-exist and be changing between these different possible structures. The Sfold predictions also indicated that it is not always the most stable structure with the lowest ΔG that is the most prevalent form. These three programs also generated very similar structures for the individual RNA molecules assessed, most likely because all three programs are based on similar biophysical principles of MFE models. In comparing the generated predicted structure of all RNA molecules using the three programs, an overall “rod-like” structure

was consistently modeled. One high note about the data generated from these three programs alone (Mfold, RNAfold, Sfold) was not the structured regions, but instead the “unstructured” free nucleotides. It is common knowledge that guanines and cytosines form the strongest interactions by having three hydrogen bond pairs, and therefore are more likely to make up the majority of the structured stem regions. With most of the structured regions made up of G-C bonds, this left adenines, uracils to make up majority of the free nucleotides. This indeed was the case that adenines and uracils with cytosines and the odd guanine made up most of the free nucleotides for interaction with protein. These are also the nucleotides that make up the predicted primary sequences to interact with the VICKZ family of RNA binding proteins (Table 15).

As with the MFE-based programs, the MC-FOLD-MC-SYM pipeline also generated “rod-like” structures, but with base pairs from solved crystal structure data of other RNA molecules. An L-shaped structure was predicted for MITF and Kras RNA molecules, whereas the MDR-1 RNA structure resembled a more curved V-shape. The c-myc RNA also fit this L-type structure, with an additional stem loop protruding from one side.

In silico algorithms and computing power are not quiet powerful enough to act as a stand-alone prediction of secondary and tertiary structure. Additional secondary/tertiary probing data could be collected, but this kind of information does not always aid with *in silico* analysis. Although the *in silico* data is not conclusive; it did however help establish an idea on how the primary structure is spaced by the secondary structure, and serves as an additional tool to understand how CRDBP and other RNA binding proteins interact with RNA. This notion would require some means of acquiring

RNA structural information, with and without bound protein possibly through x-ray crystallography trials. This could provide a more detailed examination of the protein-RNA interaction and has the potential to bridge the findings of those primary sequences with the importance of the overall RNA structure.

Currently, *in silico* analysis is also limited by the amount of primary sequence that can be computed at a time. When the *in silico* method of modeling RNA molecules can accommodate the entire RNA molecules, the relevance and reliability of this type of method would increase.

5.7 Future Directions

The results generated from this study have paved the way for further basic and applied research based on CRDBP-RNA interaction. Firstly, the design of AONs or sense oligonucleotides against smaller regions of the mapped RNAs (c-myc, MITF, and β -TrCP1) to inhibit specific CRDBP-RNA interaction is now possible.

From this study, it is clear that specific KH domains act together to uniquely bind to specific RNA molecules. Such knowledge should aid in the development of small molecule inhibitors against CRDBP-RNA interaction. One could also envision that these small molecule inhibitors could be designed to be selective for individual KH domains or specific sets of KH domains of CRDBP. This kind of selectivity would inhibit the binding of RNA molecules of interest to CRDBP by blocking the appropriate KH domains.

Currently from *in vitro* studies, the significance of the KH domain has been established as the key player in binding with specific RNA molecules. The use of cell

lines for assessing the role of VICKZ family members to interact with RNA through the KH domains has also been recognized. Assessing CRDBP in cell lines have also been linked to a phenotypic change consisting of the formation of granules in the cytoplasm. In the near future, it would be important to establish an *in vivo* model to further understand CRDBP-RNA interaction in the context of a whole animal. An *in vivo* model using CRDBP variants would translate the previously generated *in vitro* and cell data to a complex organism level, and better link the CRDBP-RNA association to a phenotype and disease. An animal model would also allow for the CRDBP-RNA complex to be assessed in a different context, one that accounts for new biological variables that simulate a biochemical environment closer to that of humans. A working animal model would also provide a more biologically relevant way to further assess small molecules for effectiveness at breaking the CRDBP-RNA interaction and the effects of the small molecules on an *in vivo* phenotype.

The *in silico* analysis should be complemented with validation of the RNA structure using experimental approaches and the RNA-protein interaction by crystallography trials. However, it should be kept in mind that current crystallography techniques also have limitations. Such limitations include the requirement for large amount of RNA sample, and guesswork at optimal crystallization conditions, which also tend to be non-native conditions that can produce structural artifact.

5.8 Concluding Remarks

The knowledge of smaller regions of RNAs, as generated in this study, has provided researchers with additional tools to discover new inhibitory compounds against CRDBP function.

The *in silico* analysis has revealed some commonality in the structure of RNAs that bind CRDBP. It also established the idea that the positioning of the primary structure within the secondary structure could provide a suitable RNA binding surface for interaction with CRDBP.

This study confirmed the importance of the G-X-X-G motif in the KH domains of CRDBP in interacting with RNA. This study also clearly showed that CRDBP binds to different RNA molecules using different KH domains.

Collectively, the studies reported here should aid in future developments for the discovery of new mechanism in disrupting protein-RNA interactions involved in cancer and other diseases.

Appendix

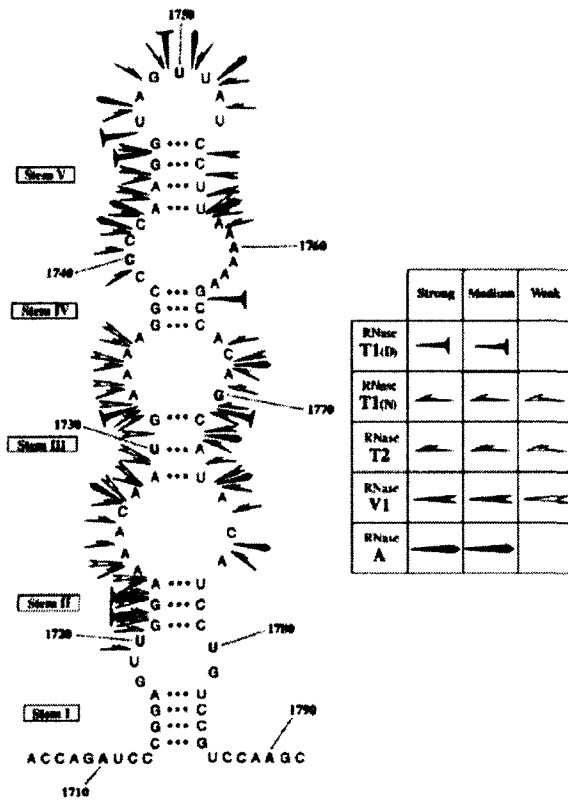


Figure S1. Ribonuclease secondary structure probing of human *c-myc* CRD RNAs 1705-1792 (88nts), indicating the locations of bases accessible for enzyme probes. Adapted from Tafesh *et al*, 2007 (5).

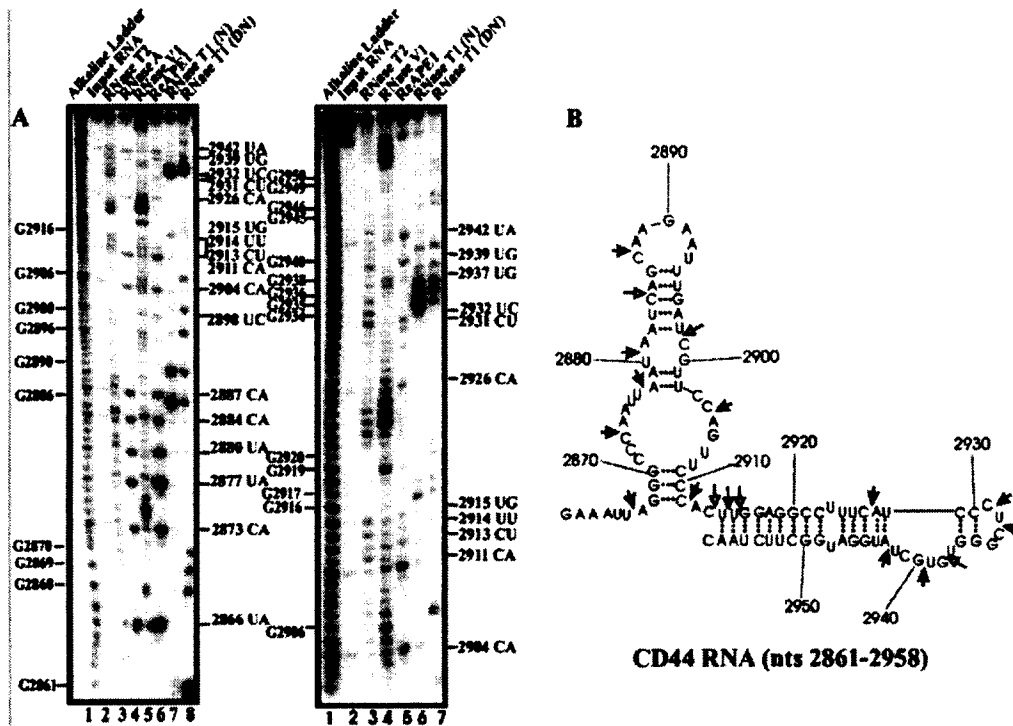


Figure S2. Ribonuclease secondary structure probing of human CD44 RNA nts 2861-2958 (98nts). Adapted from Kim *et al.*, 2010 (6).

Table S1. List of all the base-paired secondary structure features.

RNA Name	Stem Structure of RNA
c-myc stem I	5' CGGA 3' (nts 10-13) 3' GCCU 5' (nts 78-81)
c-myc stem II	5' GGA 3' (nts 17-19) 3' CCU 5' (nts 73-75)
c-myc stem III	5' AUG 3' (nts 25-27) 3' UAC 5' (nts 67-69)
c-myc stem IV	5' GGC 3' (nts 32-34) 3' CCG 5' (nts 60-62)
c-myc stem V	5' AAGG 3' (nts 39-42) 3' UUCC 5' (nts 50-53)
MITF stem I	5' GUCA ^G AG ^{AA} AUG 3' (nts 6-17) 3' CAGU _G UC _{GC} UAU 5' (nts 59-70)
MITF stem II	5' GUC-CA3' (nts 21-25) 3' UAG _A GUS' (nts 50-55)
MITF stem III	5' UCAGG 3' (nts 33-37) 3' AGUUC 5' (nts 43-47)
Kras stem I	5' UAUGAUC 3' (nts 2-8) 3' GUACUGG 5' (nts 86-92)
Kras stem II	5' GAGGAU 3' (nts 17-22) 3' CUCUUA 5' (nts 71-76)

Kras stem III	5' CAGGA ^A GCA ^A GU 3' (nts 28-39) 3' GUUCU _C UGU _C CA 5' (nts 56-67)
CD44 stem I	5' GGG 3' (nts 8-10) 3' CCC 5' (nts 49-51)
CD44 stem II	5' AAU ^A AUCAG 3' (nts 18-26) 3' UUG _C UAGUU 5' (nts 34-42)
CD44 stem III	5' UUGGAGGCC ^U UUCAU-----CC 3' (nts 54-70) 3' AAUCUUCGG _U AGGUA _{UCGUGU} GG 5' (nts 75-97)
MDR-1 stem I	5' GCUG 3' (nts 2-5) 3' UGAC 5' (nts 83-86)
MDR-1 stem II	5' CUGGUU 3' (nts 11-16) 3' GAUAAA 5' (nts 73-78)
MDR-1 stem III	5' GCUGC 3' (nts 17-21) 3' CGACG 5' (nts 50-54)
MDR-1 stem IV	5' CAGGU ^U UC 3' (nts 29-36) 3' GUCCG _U GG 5' (nts 42-49)
GLI1 stem I	5' CCUGGA ^{CC} UG 3' (nts 8-17) 3' GGACCU _{CU} AC 5' (nts 76-85)
GLI1 stem II	5' CGG 3' (nts 22-24) 3' GCC 5' (nts 29-31)
GLI1 stem III	5' AGCU 3' (nts 42-45) 3' UCGA 5' (nts 53-56)
GLI1 stem IV	5' CAUC 3' (nts 59-62) 3' GUAG 5' (nts 70-73)

Table S2. List of the single stranded bases.

RNA Name and Region	Unpaired Nucleotide Position
C-myc	5' GUU 3' (nts 14-16)
C-myc	5' AAACA 3' (nts 20-24)
C-myc	5' AAAA 3' (nts 28-31)
C-myc	5' CCCC 3' (nts 35-38)
C-myc	5' AAAAAA 3' (nts 54-59)
C-myc	5' ACAG 3' (nts 63-66)
C-myc	5' ACA 3' (nts 70-72)
C-myc	5' UG 3' (nts 76-77)
C-myc Stem Loop	5' UAGUUAU 3' (nts 43-49)
MITF	5' UCU 3' (nts 18-20)
MITF	5' UUUUAU 3' (nts 26-32)
MITF	5' UU 3' (nts 48-49)
MITF	5' A 3' (nt 52)
MITF	5' UUU 3' (nts 56-58)
MITF	5' GC 3' (nts 62-63)
MITF *2 positions	5' G 3' (nt 10, 66)

MITF Stem Loop	5' GGAAA 3' (nts 38-42)
Kras	5' CAACAAUA 3' (nts 9-16)
Kras	5' UCCUA 3' (nts 23-27)
Kras *2 positions	5' A 3' (nt 33, 37)
Kras *2 positions	5' C 3' (nt 58, 62)
Kras	5' GAU 3' (nts 68-70)
Kras	5' GACACAGCA 3' (nts 77-85)
Kras Stem Loop	5' AGUAAUUGAUGGAGAA 3' (nts 40-55)
CD44	5' CCCAAUU 3' (nts 11-17)
CD44	5' A 3' (nt 21)
CD44	5' C 3' (nts 39)
CD44	5' CCAGUU 3' (nts 43-48)
CD44	5' AC 3' (nts 52-53)
CD44 *2 positions	5' U 3' (nt 63, 88)
CD44 Stem Loop	5' CAAGAAU 3' (nts 27-33)
CD44 Stem Loop	5' CUCG 3' (nts 71-74)
CD44 Stem Loop	5' UGUGCU 3' (nts 82-77)
MDR-1	5' GGGUG 3' (nts 6-10)
MDR-1	5' UUACAUU 3' (nts 22-28)
MDR-1 *2 positions	5' U 3' (nts 34, 44)
MDR-1	5' UGGAAGACAAUACACAA 3' (nts 55-72)
MDR-1	5' AAAA 3' (nts 79-82)
MDR-1 Stem Loop	5' CAUUUU 3' (nts 38-40)
GLI1	5' CC 3' (nts 14-15)
GLI1	5' CAGA 3' (nts 18-21)
GLI1	5' CACCUCACCC 3' (nts 32-41)
GLI1	5' UU 3' (nts 57-58)
GLI1	5' CA 3' (nts 74-75)
GLI1	5' UC 3' (nts 78-79)
GLI1 Stem Loop	5' UUAU 3' (nts 25-28)
GLI1 Stem Loop	5' CCCUCGU 3' (nts 46-52)
GLI1 Stem Loop	5' AACUCGC 3' (nts 63-69)

Calculation of dissociation constant (K_d):

The percentage of bound RNA was determined using the opti-quant software (Figure S3).

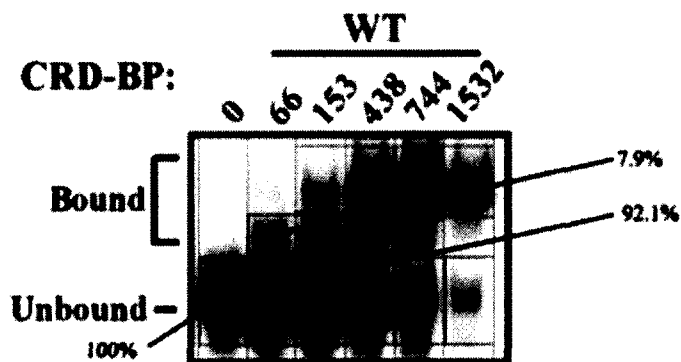


Figure S3. Autoradiograph to calculate the dissociation constant (K_d). Each autoradiograph that was generated was then used to calculate the K_d for CRDBP. Total signal from each lane (red lines) was calculated by drawing quantification boxes (blue boxes) around the signal generated from the [32 P] c-myc CRD 1705-1886 RNA substrate. K_d values were then calculated using the Hill equation.

The average of at least 3 experimental replicates with at least 2 biological replicates had the average and standard deviation calculated (Table S3). The average and standard deviation was then plotted on a graph and the Hill equation was used to calculate the dissociation constant (K_d) (Table S3, Figure S3 and S4).

Table S3. Summary table of the calculated average percent bound and standard deviation for WT-CRDBP with c-myc 1705-1886 (182 nts) RNA.

[Protein] (nm)	Percentage bound for four experimental replicates ^a				Average Percentage Bound	Standard Deviation
0	0	0	0	0	0	0
66	7.9	6.3	8.7	8.9	8.0	1.2
153	49.7	39.7	50.2	53.8	48.4	6.0
438	91.6	93.1	85.5	93.7	91.0	3.8
1532	91.6	93.1	85.5	93.7	91.0	3.8

^a At least two protein preparations.

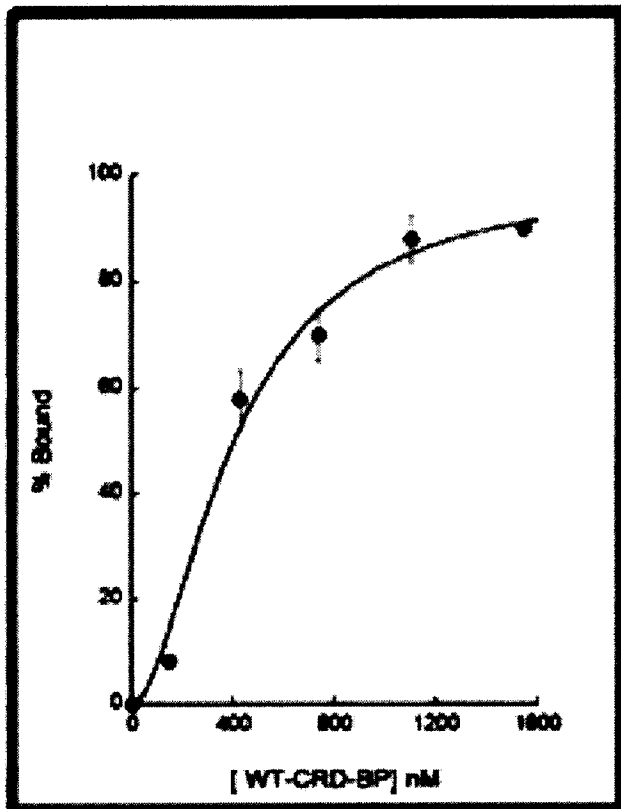


Figure S4. WT-CRDBP binding curves as generated by the Hill Equation. WT-CRDBP was incubated with [32 P] c-myc CRD 1705-1886 from Table S3.

$$\frac{\text{bound}}{\text{Total}} = \frac{\left(\frac{1}{K_d}\right)[L]^n}{1 + \left(\left(\frac{1}{K_d}\right)[L]\right)^n}$$

Hill Equation

K_d = Dissociation constant
 L = CRD-BP concentration
 n = Hill coefficient

Figure S5. The Hill Equation for calculating dissociation constant (K_d).

References

1. Bell JL, Wachter K, Muhleck B, Pazaitis N, Kohn M, Lederer M, Huttelmaier S. 2013. Insulin-like growth factor 2 mRNA-binding proteins (IGF2BPs): post-transcriptional drivers of cancer progression? *Cell Mol Life Sci* 70:2657-2675.
2. Valverde R, Edwards L, Regan L. 2008. Structure and function of KH domains. *FEBS J* 275:2712-2726.
3. Pogenberg V, Ogmundsdottir MH, Bergsteinsdottir K, Schepsky A, Phung B, Deineko V, Milewski M, Steingrimsson E, Wilmanns M. 2012. Restricted leucine zipper dimerization and specificity of DNA recognition of the melanocyte master regulator MITF. *Genes Dev* 26:2647-2658.
4. Zhao Y, Huang Y, Gong Z, Wang Y, Man J, Xiao Y. 2012. Automated and fast building of three-dimensional RNA structures. *Sci Rep* 2:734.
5. Tafech A, Bennett WR, Mills F, Lee CH. 2007. Identification of c-myc coding region determinant RNA sequences and structures cleaved by an RNase1-like endoribonuclease. *Biochim Biophys Acta* 1769:49-60.
6. Kim WC, King D, Lee CH. 2010. RNA-cleaving properties of human apurinic/apryrimidinic endonuclease 1 (APE1). *Int J Biochem Mol Biol* 1:12-25.
7. Alberts B, Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. . 2008. *Molecular Biology of The Cell*, 5th Edition. Garland Science, Taylor & Francis Group, LLC: .
8. Chen CY, Ezzeddine N, Shyu AB. 2008. Messenger RNA half-life measurements in mammalian cells. *Methods Enzymol* 448:335-357.
9. Abaza I, Gebauer F. 2008. Trading translation with RNA-binding proteins. *RNA* 14:404-409.
10. Colgan DF, Manley JL. 1997. Mechanism and regulation of mRNA polyadenylation. *Genes Dev* 11:2755-2766.
11. Collier J, Parker R. 2004. Eukaryotic mRNA decapping. *Annu Rev Biochem* 73:861-890.
12. Dunckley T, Parker R. 1999. The DCP2 protein is required for mRNA decapping in *Saccharomyces cerevisiae* and contains a functional MutT motif. *EMBO J* 18:5411-5422.
13. Tijsterman M, Ketting RF, Plasterk RH. 2002. The genetics of RNA silencing. *Annu Rev Genet* 36:489-519.
14. Cali BM, Anderson P. 1998. mRNA surveillance mitigates genetic dominance in *Caenorhabditis elegans*. *Mol Gen Genet* 260:176-184.
15. Frischmeyer PA, van Hoof A, O'Donnell K, Guerrerio AL, Parker R, Dietz HC. 2002. An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. *Science* 295:2258-2261.
16. Doma MK, Parker R. 2006. Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature* 440:561-564.
17. Barreau C, Paillard L, Osborne HB. 2005. AU-rich elements and associated factors: are there unifying principles? *Nucleic Acids Res* 33:7138-7150.

18. Stoneley M, Willis AE. 2004. Cellular internal ribosome entry segments: structures, trans-acting factors and regulation of gene expression. *Oncogene* 23:3200-3207.
19. Jang SK, Krausslich HG, Nicklin MJ, Duke GM, Palmenberg AC, Wimmer E. 1988. A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *J Virol* 62:2636-2643.
20. Pelletier J, Sonenberg N. 1988. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* 334:320-325.
21. Jang SK, Davies MV, Kaufman RJ, Wimmer E. 1989. Initiation of protein synthesis by internal entry of ribosomes into the 5' nontranslated region of encephalomyocarditis virus RNA in vivo. *J Virol* 63:1651-1660.
22. Chen CY, Shyu AB. 1995. AU-rich elements: characterization and importance in mRNA degradation. *Trends Biochem Sci* 20:465-470.
23. Peng SS, Chen CY, Xu N, Shyu AB. 1998. RNA stabilization by the AU-rich element binding protein, HuR, an ELAV protein. *EMBO J* 17:3461-3470.
24. Nelson DL, DALLALCM. 2008. *Lehninger principles of biochemistry*. W.H. Freeman, New York.
25. Lemm I, Ross J. 2002. Regulation of c-myc mRNA decay by translational pausing in a coding region instability determinant. *Mol Cell Biol* 22:3959-3969.
26. Barnes T, Kim WC, Mantha AK, Kim SE, Izumi T, Mitra S, Lee CH. 2009. Identification of Apurinic/aprimidinic endonuclease 1 (APE1) as the endoribonuclease that cleaves c-myc mRNA. *Nucleic Acids Res* 37:3946-3958.
27. Komar AA, Hatzoglou M. 2011. Cellular IRES-mediated translation: the war of ITAFs in pathophysiological states. *Cell Cycle* 10:229-240.
28. Bartel DP. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116:281-297.
29. Elcheva I, Goswami S, Noubissi FK, Spiegelman VS. 2009. CRD-BP protects the coding region of betaTrCP1 mRNA from miR-183-mediated degradation. *Mol Cell* 35:240-246.
30. Goswami S, Tarapore RS, Teslaa JJ, Grinblat Y, Setaluri V, Spiegelman VS. 2010. MicroRNA-340-mediated degradation of microphthalmia-associated transcription factor mRNA is inhibited by the coding region determinant-binding protein. *J Biol Chem* 285:20532-20540.
31. Agami R. 2010. microRNAs, RNA binding proteins and cancer. *Eur J Clin Invest* 40:370-374.
32. Filipowicz W, Bhattacharyya SN, Sonenberg N. 2008. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet* 9:102-114.
33. Liu J, Hatzoglou M. 1998. Control of expression of the gene for the arginine transporter Cat-1 in rat liver cells by glucocorticoids and insulin. *Amino Acids* 15:321-337.

34. Gonzalez-Feliciano JA, Hernandez-Perez M, Estrella LA, Colon-Lopez DD, Lopez A, Martinez M, Mauras-Rivera KR, Lasalde C, Martinez D, Araujo-Perez F, Gonzalez CI. 2014. The role of HuR in the post-transcriptional regulation of interleukin-3 in T cells. *PLoS One* 9:e92457.
35. Bhattacharyya SN, Habermacher R, Martine U, Closs EI, Filipowicz W. 2006. Relief of microRNA-mediated translational repression in human cells subjected to stress. *Cell* 125:1111-1124.
36. Dasgupta T, Ladd AN. 2012. The importance of CELF control: molecular and biological roles of the CUG-BP, Elav-like family of RNA-binding proteins. *Wiley Interdiscip Rev RNA* 3:104-121.
37. Kaufmann WE. 2002. Neurobiology of Fragile X syndrome: from molecular genetics to neurobehavioral phenotype. *Microsc Res Tech* 57:131-134.
38. Musco G, Stier G, Joseph C, Castiglione Morelli MA, Nilges M, Gibson TJ, Pastore A. 1996. Three-dimensional structure and stability of the KH domain: molecular insights into the fragile X syndrome. *Cell* 85:237-245.
39. Oberman F, Rand K, Maizels Y, Rubinstein AM, Yisraeli JK. 2007. VICKZ proteins mediate cell migration via their RNA binding activity. *RNA* 13:1558-1569.
40. Yisraeli JK. 2005. VICKZ proteins: a multi-talented family of regulatory RNA-binding proteins. *Biol Cell* 97:87-96.
41. Farina KL, Huttelmaier S, Musunuru K, Darnell R, Singer RH. 2003. Two ZBP1 KH domains facilitate beta-actin mRNA localization, granule formation, and cytoskeletal attachment. *J Cell Biol* 160:77-87.
42. Ross AF, Oleynikov Y, Kislauskis EH, Taneja KL, Singer RH. 1997. Characterization of a beta-actin mRNA zipcode-binding protein. *Mol Cell Biol* 17:2158-2165.
43. Vikesaa J, Hansen TV, Jonson L, Borup R, Wewer UM, Christiansen J, Nielsen FC. 2006. RNA-binding IMPs promote cell adhesion and invadopodia formation. *EMBO J* 25:1456-1468.
44. Tessier CR, Doyle GA, Clark BA, Pitot HC, Ross J. 2004. Mammary tumor induction in transgenic mice expressing an RNA-binding protein. *Cancer Res* 64:209-214.
45. Nielsen J, Kristensen MA, Willemoes M, Nielsen FC, Christiansen J. 2004. Sequential dimerization of human zipcode-binding protein IMP1 on RNA: a cooperative mechanism providing RNP stability. *Nucleic Acids Res* 32:4368-4376.
46. Bernstein PL, Herrick DJ, Prokipcak RD, Ross J. 1992. Control of c-myc mRNA half-life in vitro by a protein capable of binding to a coding region stability determinant. *Genes Dev* 6:642-654.
47. Atlas R, Behar L, Elliott E, Ginzburg I. 2004. The insulin-like growth factor mRNA binding-protein IMP-1 and the Ras-regulatory protein G3BP associate with tau mRNA and HuD protein in differentiated P19 neuronal cells. *J Neurochem* 89:613-626.
48. Mueller F, Bommer M, Lacher U, Ruhland C, Stagge V, Adler G, Gress TM, Seufferlein T. 2003. KOC is a novel molecular indicator of malignancy. *Br J Cancer* 88:699-701.

49. Twyman RM, Wisden W. 1999. *Advanced Molecular Biology: A Concise Reference*. Springer Verlag.
50. Lunde BM, Moore C, Varani G. 2007. RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol* 8:479-490.
51. Nagai K, Oubridge C, Ito N, Jessen TH, Avis J, Evans P. 1995. Crystal structure of the U1A spliceosomal protein complexed with its cognate RNA hairpin. *Nucleic Acids Symp Ser*:1-2.
52. Kielkopf CL, Lucke S, Green MR. 2004. U2AF homology motifs: protein recognition in the RRM world. *Genes Dev* 18:1513-1526.
53. Rackham O, Brown CM. 2004. Visualization of RNA-protein interactions in living cells: FMRP and IMP1 interact on mRNAs. *EMBO J* 23:3346-3355.
54. Lewis HA, Musunuru K, Jensen KB, Edo C, Chen H, Darnell RB, Burley SK. 2000. Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome. *Cell* 100:323-332.
55. Bycroft M, Hubbard TJ, Proctor M, Freund SM, Murzin AG. 1997. The solution structure of the S1 RNA binding domain: a member of an ancient nucleic acid-binding fold. *Cell* 88:235-242.
56. Doyle GA, Betz NA, Leeds PF, Fleisig AJ, Prokipcak RD, Ross J. 1998. The c-myc coding region determinant-binding protein: a member of a family of KH domain RNA-binding proteins. *Nucleic Acids Res* 26:5036-5044.
57. Leeds P, Kren BT, Boylan JM, Betz NA, Steer CJ, Gruppuso PA, Ross J. 1997. Developmental regulation of CRD-BP, an RNA-binding protein that stabilizes c-myc mRNA in vitro. *Oncogene* 14:1279-1286.
58. Marhaba R, Zoller M. 2004. CD44 in cancer progression: adhesion, migration and growth regulation. *J Mol Histol* 35:211-231.
59. Stohr N, Kohn M, Lederer M, Glass M, Reinke C, Singer RH, Huttelmaier S. 2012. IGF2BP1 promotes cell migration by regulating MK5 and PTEN signaling. *Genes Dev* 26:176-189.
60. Noubissi FK, Elcheva I, Bhatia N, Shakoory A, Ougolkov A, Liu J, Minamoto T, Ross J, Fuchs SY, Spiegelman VS. 2006. CRD-BP mediates stabilization of betaTrCP1 and c-myc mRNA in response to beta-catenin signalling. *Nature* 441:898-901.
61. Sparanese D, Lee CH. 2007. CRD-BP shields c-myc and MDR-1 RNA from endonucleolytic attack by a mammalian endoribonuclease. *Nucleic Acids Res* 35:1209-1221.
62. Dimitriadis E, Trangas T, Milatos S, Foukas PG, Gioulbasanis I, Courtis N, Nielsen FC, Pandis N, Dafni U, Bardi G, Ioannidis P. 2007. Expression of oncofetal RNA-binding protein CRD-BP/IMP1 predicts clinical outcome in colon cancer. *Int J Cancer* 121:486-494.
63. Ioannidis P, Mahaira L, Papadopoulou A, Teixeira MR, Heim S, Andersen JA, Evangelou E, Dafni U, Pandis N, Trangas T. 2003. 8q24 Copy number gains and expression of the c-myc mRNA stabilizing protein CRD-BP in primary breast carcinomas. *Int J Cancer* 104:54-59.
64. Liao B, Patel M, Hu Y, Charles S, Herrick DJ, Brewer G. 2004. Targeted knockdown of the RNA-binding protein CRD-BP promotes cell proliferation

- via an insulin-like growth factor II-dependent pathway in human K562 leukemia cells. *J Biol Chem* 279:48716-48724.
65. Dang CV. 1999. c-Myc target genes involved in cell growth, apoptosis, and metabolism. *Mol Cell Biol* 19:1-11.
 66. Doyle GA, Bourdeau-Heller JM, Coulthard S, Meisner LF, Ross J. 2000. Amplification in human breast cancer of a gene encoding a c-myc mRNA-binding protein. *Cancer Res* 60:2756-2759.
 67. Noubissi FK, Nikiforov MA, Colburn N, Spiegelman VS. 2010. Transcriptional Regulation of CRD-BP by c-myc: Implications for c-myc Functions. *Genes Cancer* 1:1074-1082.
 68. Lesley J, Hyman R, Kincade PW. 1993. CD44 and its interaction with extracellular matrix. *Adv Immunol* 54:271-335.
 69. Pacher M, Seewald MJ, Mikula M, Oehler S, Mogg M, Vinatzer U, Eger A, Schweifer N, Varecka R, Sommergruber W, Mikulits W, Schreiber M. 2007. Impact of constitutive IGF1/IGF2 stimulation on the transcriptional program of human breast cancer cells. *Carcinogenesis* 28:49-59.
 70. Espelund U, Cold S, Frystyk J, Orskov H, Flyvbjerg A. 2008. Elevated free IGF2 levels in localized, early-stage breast cancer in women. *Eur J Endocrinol* 159:595-601.
 71. Phillips MB, and pinkerton, C.R. 1993. MDR1 mRNA in Resistant Neuroblastoma Bone Marrow. *Clinical Oncology* 5:263.
 72. Chao JA, Patskovsky Y, Patel V, Levy M, Almo SC, Singer RH. 2010. ZBP1 recognition of beta-actin zipcode induces RNA looping. *Genes Dev* 24:148-158.
 73. Siomi H, Matunis MJ, Michael WM, Dreyfuss G. 1993. The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. *Nucleic Acids Res* 21:1193-1198.
 74. Chien CY, Tejero R, Huang Y, Zimmerman DE, Rios CB, Krug RM, Montelione GT. 1997. A novel RNA-binding motif in influenza A virus non-structural protein 1. *Nat Struct Biol* 4:891-895.
 75. Grishin NV. 2001. KH domain: one motif, two folds. *Nucleic Acids Res* 29:638-643.
 76. Git A, Standart N. 2002. The KH domains of Xenopus Vg1RBP mediate RNA binding and self-association. *RNA* 8:1319-1333.
 77. Nielsen J, Christiansen J, Lykke-Andersen J, Johnsen AH, Wewer UM, Nielsen FC. 1999. A family of insulin-like growth factor II mRNA-binding proteins represses translation in late development. *Mol Cell Biol* 19:1262-1270.
 78. Ioannidis P, Mahaira L, Papadopoulou A, Teixeira MR, Heim S, Andersen JA, Evangelou E, Dafni U, Pandis N, Trangas T. 2003. CRD-BP: a c-Myc mRNA stabilizing protein with an oncofetal pattern of expression. *Anticancer Res* 23:2179-2183.
 79. Ioannidis P, Trangas T, Dimitriadis E, Samiotaki M, Kyriazoglou I, Tsiapalis CM, Kittas C, Agnantis N, Nielsen FC, Nielsen J, Christiansen J, Pandis N. 2001. C-MYC and IGF-II mRNA-binding protein (CRD-BP/IMP-1) in benign and malignant mesenchymal tumors. *Int J Cancer* 94:480-484.

80. Eberle J, Hossini AM. 2008. Expression and function of bcl-2 proteins in melanoma. *Curr Genomics* 9:409-419.
81. Chawla-Sarkar M, Bae SI, Reu FJ, Jacobs BS, Lindner DJ, Borden EC. 2004. Downregulation of Bcl-2, FLIP or IAPs (XIAP and survivin) by siRNAs sensitizes resistant melanoma cells to Apo2L/TRAIL-induced apoptosis. *Cell Death Differ* 11:915-923.
82. Yasumoto K, Yokoyama K, Shibata K, Tomita Y, Shibahara S. 1994. Microphthalmia-associated transcription factor as a regulator for melanocyte-specific transcription of the human tyrosinase gene. *Mol Cell Biol* 14:8058-8070.
83. da Cruz AT, Jasiulionis MG. 2012. miRNAs and Melanoma: How Are They Connected? *Dermatol Res Pract* 2012:528345.
84. Haflidadottir BS, Bergsteinsdottir K, Praetorius C, Steingrimsdottir E. 2010. miR-148 regulates Mitf in melanoma cells. *PLoS One* 5:e11574.
85. Bemis LT, Chen R, Amato CM, Classen EH, Robinson SE, Coffey DG, Erickson PF, Shellman YG, Robinson WA. 2008. MicroRNA-137 targets microphthalmia-associated transcription factor in melanoma cell lines. *Cancer Res* 68:1362-1368.
86. Davis IJ, Kim JJ, Ozsolak F, Widlund HR, Rozenblatt-Rosen O, Granter SR, Du J, Fletcher JA, Denny CT, Lessnick SL, Linehan WM, Kung AL, Fisher DE. 2006. Oncogenic MITF dysregulation in clear cell sarcoma: defining the MiT family of human cancers. *Cancer Cell* 9:473-484.
87. Steingrimsdottir E, Copeland NG, Jenkins NA. 2004. Melanocytes and the microphthalmia transcription factor network. *Annu Rev Genet* 38:365-411.
88. Klein DJ, Schmeing TM, Moore PB, Steitz TA. 2001. The kink-turn: a new RNA secondary structure motif. *EMBO J* 20:4214-4221.
89. Gerst JE. 2011. RNA Detection and Visualization Methods and Protocols vol. 714. Humana Press.
90. Eddy SR. 2004. How do RNA folding algorithms work? *Nat Biotechnol* 22:1457-1458.
91. Zuker M. 2000. Calculating nucleic acid secondary structure. *Curr Opin Struct Biol* 10:303-310.
92. Markham NR, Zuker M. 2008. UNAFold: software for nucleic acid folding and hybridization. *Methods Mol Biol* 453:3-31.
93. Hofacker IL. 2003. Vienna RNA secondary structure server. *Nucleic Acids Res* 31:3429-3431.
94. Zuker M, Stiegler P. 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res* 9:133-148.
95. Do CB, Woods DA, Batzoglou S. 2006. CONTRAfold: RNA secondary structure prediction without physics-based models. *Bioinformatics* 22:e90-98.
96. Ding Y, Lawrence CE. 2003. A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Res* 31:7280-7301.
97. Danilova LV, Pervouchine DD, Favorov AV, Mironov AA. 2006. RNAKinetics: a web server that models secondary structure kinetics of an elongating RNA. *J Bioinform Comput Biol* 4:589-596.

98. Forties RA, Bundschuh R. 2010. Modeling the interplay of single-stranded binding proteins and nucleic acid secondary structure. *Bioinformatics* 26:61-67.
99. Rivas E, Eddy SR. 2001. Noncoding RNA gene detection using comparative sequence analysis. *BMC Bioinformatics* 2:8.
100. Washietl S, Hofacker IL, Stadler PF. 2005. Fast and reliable prediction of noncoding RNAs. *Proc Natl Acad Sci U S A* 102:2454-2459.
101. Yao Z, Weinberg Z, Ruzzo WL. 2006. CMfinder--a covariance model based RNA motif finding algorithm. *Bioinformatics* 22:445-452.
102. Sankoff D. 1985. Simultaneous Solution of the RNA Folding, Alignment and Protosequence Problems. *SIAM Journal of Applied Mathematics* 45:810-825.
103. Havgaard JH, Lyngso RB, Gorodkin J. 2005. The FOLDALIGN web server for pairwise structural RNA alignment and mutual motif search. *Nucleic Acids Res* 33:W650-653.
104. Touzet H, Perriquet O. 2004. CARNAC: folding families of related RNAs. *Nucleic Acids Res* 32:W142-145.
105. Mathews DH. 2005. Predicting a set of minimal free energy RNA secondary structures common to two sequences. *Bioinformatics* 21:2246-2253.
106. Parisien M, Major F. 2008. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* 452:51-55.
107. King DT, Barnes M, Thomsen D, Lee CH. 2014. Assessing specific oligonucleotides and small molecule antibiotics for the ability to inhibit the CRD-BP-CD44 RNA interaction. *PLoS One* 9:e91585.
108. Mehmood K. 2014. Biochemical and cellular characterization of CRD-BP-mRNA interaction.
109. Barnes M. 2013. Characterizing the CRD-BP-RNA interaction in-vitro and in cells.
110. Liao JM, Lu H. 2011. Autoregulatory suppression of c-Myc by miR-185-3p. *J Biol Chem* 286:33901-33909.
111. Nielsen FC, Nielsen J, Kristensen MA, Koch G, Christiansen J. 2002. Cytoplasmic trafficking of IGF-II mRNA-binding protein by conserved KH domains. *J Cell Sci* 115:2087-2097.
112. Zhou Y, Mah TF, Greenblatt J, Friedman DI. 2002. Evidence that the KH RNA-binding domains influence the action of the E. coli NusA protein. *J Mol Biol* 318:1175-1188.
113. Ranjbar B, Gill P. 2009. Circular dichroism techniques: biomolecular and nanostructural analyses- a review. *Chem Biol Drug Des* 74:101-120.
114. Spencer KA, Hiscox JA. 2006. Characterisation of the RNA binding properties of the coronavirus infectious bronchitis virus nucleocapsid protein amino-terminal region. *FEBS Lett* 580:5993-5998.
115. Mark Barnes GvR, Wai-Ming Li, Kashif Mehmood, Sebastian Mackedenski, Dustin T. King, Andrew L. Miller, and Chow H. Lee. 2014. Insights into the Coding Region Determinant-Binding Protein-RNA Interaction through Mutagenesis at the Glycine-X-X-Glycine Motif in heterogeneous nuclear Ribonucleoproteins-K-homology Domains. *Journal of Biological Sciences*.

116. Jonson L, Vikesaa J, Krogh A, Nielsen LK, Hansen T, Borup R, Johnsen AH, Christiansen J, Nielsen FC. 2007. Molecular composition of IMP1 ribonucleoprotein granules. *Mol Cell Proteomics* 6:798-811.
117. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jr., Jungkamp AC, Munschauer M, Ulrich A, Wardle GS, Dewell S, Zavolan M, Tuschl T. 2010. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 141:129-141.
118. Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31:3406-3415.
119. Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P. 1994. Fast folding and comparison of RNA secondary structures. *Monatsh Chem* 125:167-188.
120. Ding Y, Chan CY, Lawrence CE. 2004. Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res* 32:W135-141.
121. Smith C, Heyne S, Richter AS, Will S, Backofen R. 2010. Freiburg RNA Tools: a web server integrating INTARNA, EXPARNA and LOCARNA. *Nucleic Acids Res* 38:W373-377.
122. Mahapatra L, Mao C, Andruska N, Zhang C, Shapiro DJ. 2014. High-throughput fluorescence anisotropy screen for inhibitors of the oncogenic mRNA binding protein, IMP-1. *J Biomol Screen* 19:427-436.
123. Coulis CM, Lee C, Nardone V, Prokipcak RD. 2000. Inhibition of c-myc expression in cells by targeting an RNA-protein interaction using antisense oligonucleotides. *Mol Pharmacol* 57:485-494.
124. Du Z, Lee JK, Tjhen R, Li S, Pan H, Stroud RM, James TL. 2005. Crystal structure of the first KH domain of human poly(C)-binding protein-2 in complex with a C-rich strand of human telomeric DNA at 1.7 Å. *J Biol Chem* 280:38823-38830.
125. Hollingworth D, Candel AM, Nicastro G, Martin SR, Briata P, Gherzi R, Ramos A. 2012. KH domains with impaired nucleic acid binding as a tool for functional analysis. *Nucleic Acids Res* 40:6873-6886.
126. Wachter K, Kohn M, Stohr N, Huttelmaier S. 2013. Subcellular localization and RNP formation of IGF2BPs (IGF2 mRNA-binding proteins) is modulated by distinct RNA-binding domains. *Biol Chem* 394:1077-1090.