

THE MUSCLEBLIND PROTEIN FAMILY'S RNA SEQUENCE ELEMENTS,
STRUCTURAL ELEMENTS AND NOVEL BINDING SITES
DEFINED THROUGH SELEX

by

EMILY SARAH MARIE GOERS

A DISSERTATION

Presented to the Department of Chemistry
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

December 2008

University of Oregon Graduate School

Confirmation of Approval and Acceptance of Dissertation prepared by:

Emily Goers

Title:

"The Muscleblind Protein Family's RNA Sequence Elements, Structural Elements and Novel Binding Sites Defined Through SELEX"

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Chemistry by:

Kenneth Prehoda, Chairperson, Chemistry
J. Andrew Berglund, Advisor, Chemistry
Victoria DeRose, Member, Chemistry
Alice Barkan, Member, Biology
Eric Johnson, Outside Member, Biology

and Richard Linton, Vice President for Research and Graduate Studies/Dean of the Graduate School for the University of Oregon.

December 13, 2008

Original approval signatures are on file with the Graduate School and the University of Oregon Libraries.

An Abstract of the Dissertation of

Emily Sarah Marie Goers for the degree of Doctor of Philosophy
in the Department of Chemistry to be taken December 2008

Title: THE MUSCLEBLIND PROTEIN FAMILY'S RNA SEQUENCE ELEMENTS,
STRUCTURAL ELEMENTS AND NOVEL BINDING SITES DEFINED
THROUGH SELEX

Approved: _____
Dr. J. Andrew Berglund

Myotonic Dystrophy type I (DM1) is caused by muscleblind protein sequestration to aberrantly expanded CUG repeats. When muscleblind is sequestered it can no longer fulfill its role as an alternative splicing regulator, leading to mis-splicing events in both humans and *Drosophila*. The muscleblind protein family's RNA binding specificity has been minimally characterized. Only one pre-mRNA target in humans, cardiac troponin T (cTNT), has a known MBNL1 binding site. In order to understand muscleblind's RNA binding specificity and identify a consensus binding motif, systematic evolution of ligands by exponential enrichment (SELEX) was performed on both the *Drosophila* muscleblind protein, Mbl, and the human ortholog, MBNL1.

Drosophila has provided a useful model for studying the disease mechanism of DM1. Studies of Mbl's RNA binding specificity to CUG repeats concluded that replacing the U-U mismatches with different pyrimidine-pyrimidine mismatches was

tolerated, but no other mutations were. To understand Mbl's RNA binding specificity, SELEX was performed. After 6 rounds, several sequences were identified that bound with high affinity, all containing the 5'-AGUCU-3' consensus motif. One sequence, SELEX RNA 20 was analyzed further. In addition to the guanosine in the consensus motif of SELEX RNA 20, two other guanosines were shown to be protected by Mbl in a footprinting assay, indicating that Mbl has a strong preference for binding guanosine. Also, two "tail" regions of SELEX RNA 20 were shown to be single stranded and required for binding by Mbl. These results indicate that Mbl is a highly specific RNA binding protein with preference for both single and double stranded guanosine-rich regions.

A doped SELEX was performed on MBNL1's binding site from the cTNT pre-mRNA to determine which sequences and structural aspects were important for recognition by MBNL1. Pool 5 RNA sequences bound with high affinity, and the motif 5'-YGCUU-3' was selected. This motif was then used to identify new MBNL1 binding sites in pre-mRNAs regulated by MBNL1, SERCA1 and MBNL1. The identification of this motif and two new MBNL1 sites provide insight into MBNL1-mediated alternative splicing.

This dissertation includes both my previously published co-authored material and my unpublished co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Emily Sarah Marie Goers

PLACE OF BIRTH: San Luis Obispo, California

DATE OF BIRTH: July 16, 1981

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon

University of California, Santa Cruz

DEGREES AWARDED:

Doctor of Philosophy, Chemistry, 2008, University of Oregon

Bachelor of Science, Biochemistry and Molecular Biology, 2003, University of California, Santa Cruz

AREAS OF SPECIAL INTEREST:

Protein-RNA interactions

Alternative Splicing

Myotonic dystrophy

PROFESSIONAL EXPERIENCE:

Doctoral Research, Under Dr. J. Andrew Berglund, University of Oregon, 2004-present

Graduate Teaching Fellow, Department of Chemistry, University of Oregon, 2004-2005

Research Technician, Under Dr. R. Scott Lokey, University of California, Santa Cruz, 2003-2004

GRANTS, AWARDS AND HONORS:

NIH Training Grant, University of Oregon, 2005-2008

6th International Myotonic Dystrophy Consortium Meeting: Young Scientists
Travel and Registration fellowship, 2007

Roche Bioscience Scholarship, University of California, Santa Cruz, 2002

PUBLICATIONS:

Goers, E. S., Voelker, R. B., Gates, D. P., and Berglund, J. A. (2008) RNA binding specificity of *Drosophila* muscleblind, *Biochemistry* 47, 7284-7294.

Simon, R. A., Schuresko, L., Dendukuri, N., Goers, E., Murphy, B., and Lokey, R. S. (2005) One-bead-one-compound library of end-capped dipeptides and deconvolution by microflow NMR, *J Comb Chem* 7, 697-702.

ACKNOWLEDGEMENTS

I would like to express sincere appreciation to Dr. Andy Berglund for being an amazing advisor. He provided the perfect level of guidance; he gently pushed me when I needed it and held back at times to let me develop my independence. He let me stray from the lab's expertise to dabble in *in vivo* studies and let me travel to Italy for an amazing conference. I don't think I could have picked a better advisor, thank you.

I would like to thank the Berglund lab members, past and present, for all the support and scientific discussion. We had a great time inside and outside of the lab. I'll never forget the swimming hole trips, the Halloween parties and throwing handfuls of hail at each other from the windowsills. I could go on and on, but just know I looked forward to coming in to work everyday to be with you folks. I would especially like to thank Rodger Voelker and Devika Gates for working so hard on our papers, I couldn't have done them without you. I would also like to thank all the students from "my year" for helping maintain my work and play balance, you did a great job.

I would like to thank my committee members for direction and support inside and outside of annual committee meetings. When I did not have the experience, you gave me much needed perspective.

I would like to thank all the staff in the Institute of Molecular Biology for keeping me sane in the moments of panic when my computer stopped working, or when I just

needed that one piece of glassware to finish the experiment, or when I needed to ship something immediately and couldn't figure out the FedEx packaging system.

I would like to thank my family for constant support during each decision I made in graduate school. Every care package, call and visit got me a little closer to graduation. Mom, I'm sorry that Dad won out in the war of scientist versus artist, but just remember that the minute amount of artistic skill I express comes from you. I would like to thank my husband, Scott, for helping me develop into the scientist and wife that I am. I don't think I could have found a better person to spend the rest of my life with, I love you.

Last but not least, I would like to thank the creatures that supported me through it all. The chickens, Scratchy and Feisty, for sacrificing their eggs and blood for antibodies, without a choice in the matter. The two rabbits that sacrificed their lives so I could attain a deeper understanding of the workings of cells. Nash, Wyla and Parsley reminded me to stay a little lighter even when things seemed dark.

In memory of Feisty, Scratchy, and all the animals who were sacrificed for the research
herein, I will never forget you.

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION	1
Myotonic Dystrophy	2
Alternative Splicing Regulation Mechanisms.....	7
The Muscleblind Family of Proteins	11
The Myotonic Dystrophy Type I <i>Drosophila</i> Model	14
RNA Binding Specificity of MBNL to CUG Repeats and Other Targets	16
Systematic Evolution of Ligands by Exponential Enrichment	19
Dissertation Overview	23
II. RNA BINDING SPECIFICITY OF <i>DROSOPHILA</i> MUSCLEBLIND	25
Introduction	25
Materials and Methods	28
Protein Construct Design and Protein Purification	28
Transcription and Kinase Reactions	29
Gel Mobility Shift Assays	29
SELEX	30
Structure Probing and Footprinting	31
Determination of <i>n</i> -mer Bias in SELEX Sequences	32
Results.....	33
A Minimal Mbl Protein Binds Specifically to Expanded CUG and Short CUG Repeats	33
Identification of RNA Sequences that Bind with High Affinity to Mbl	35
The Secondary Structure and Mbl Binding Site on a SELEX Group I RNA	38
Truncation and Mutational Analysis Identifies Features of SELEX RNA 20 that are Essential for Binding by Mbl	41
Mbl Recognizes an MBNL1 Binding Site in the cTNT pre-mRNA with High Affinity	44

Chapter	Page
Discussion	47
III. DOPED SELEX WITH MBNL1 REVEALS A MOTIF USED TO IDENTIFY NOVEL MBNL1 BINDING SITES	54
Introduction	54
Materials and Methods	57
SELEX	57
Labeling of RNA Oligomers for Gel Mobility Shift Assays	59
Gel Mobility Shift Assay	59
Positional Selection	60
K-mer Enrichment	60
Results	61
MBNL1 Binding Site in cTNT Intron Provides a Template for a Doped RNA SELEX	61
The Majority of SELEX Sequences Bind MBNL1 with High Affinity and Provide a Consensus Motif of 5'-YGCUU-3'	64
Mutations in the 5'-YGCUU-3' Consensus Motif Decrease MBNL1's Binding Affinity	71
Novel Binding Sites Identified in pre-mRNAs Regulated by MBNL1	72
Discussion	73
IV. CONCLUSIONS AND FUTURE DIRECTIONS	79
Conclusions from Chapter II: RNA Binding Specificity of <i>Drosophila</i> Muscleblind	80
Future Directions for Chapter II: RNA Binding Specificity of <i>Drosophila</i> Muscleblind	83
Conclusions from Chapter III: Doped SELEX with MBNL1 Reveals a Motif Used to Identify Novel MBNL1 Binding Sites.....	84
Future Directions for Chapter III: Doped SELEX with MBNL1 Reveals a Motif Used to Identify Novel MBNL1 Binding Sites	87

Chapter	Page
APPENDIX: DOPED SELEX ON MBNL1 REVEALS ROUND 7 SEQUENCES HAVE MULTIPLE 5'-YGCUU-3' MOTIFS	89
REFERENCES	93

LIST OF FIGURES

Figure	Page
Chapter I	
1. A Disease Model for Myotonic Dystrophy Type 1	5
2. The Muscleblind Proteins are Highly Conserved, Especially in the Zinc Fingers	11
3. Systematic Evolution of Ligands by Exponential Enrichment	21
Chapter II	
1. Mbl Binds CUG Repeats with a High Degree of Specificity	34
2. High-Affinity RNA Ligands for Mbl Identified by SELEX	36
3. RNA Sequences From Group I and II Bind Mbl with High Affinity	40
4. Secondary Structure Determination and Identification of the Mbl Binding Site on RNA 20	42
5. Mbl Requires 5'- and 3'-Single-Stranded Tails but not Full-Length Stem-Loop 1 for High-Affinity RNA Binding	45
6. Mbl Binds Specific Structural Elements and Sequences	46
7. Mbl Binds a Human MBNL1 Binding Site in cTNT pre-mRNA with High Affinity and Specificity	48
8. Comparison of Three Stem-Loop Binding Sites for the Muscleblind Family of Proteins	50
Chapter III	
1. Round 5 and Round 7 of SELEX RNA Oligomers Bind Significantly Tighter than Round 0 Oligomers	62
2. The 82 SELEX Sequences From Round 5 Fell Into One Main Group	64
3. Regions of Positive and Negative Selection Reveal a Consensus Binding Sequence on the Left Region of the cTNT Stem	67
4. Most SELEX Sequences Contain at Least One 5'-YGCUU-3' Motif and Bind with High Affinity	69
5. SELEX Sequences Bind MBNL1 with High Affinity With or Without the Constant Regions	73

Figure	Page
6. Mutations to the 5'-YGCUU-3' Motif in SELEX sequences Decrease Binding	74
7. Novel MBNL1 Binding Sites Identified Near MBNL1 Regulated Exons in MBNL1 and SERCA1 pre-mRNAs	76
 Appendix	
1. The 76 SELEX sequences from round 7 fall into two main groups.....	90
2. Novel MBNL1 Binding Sites Identified Near MBNL1 Regulated Exons in MBNL1 and SERCA1 pre-mRNAs	92

LIST OF TABLES

Table	Page
Chapter II	
1. Concentration of RNA and Protein used in SELEX	30
2. Highly Enriched n-mers Found in SELEX Sequences	39
Chapter III	
1. Concentration of RNA and Protein Used for Rounds of SELEX	59
2. Highly Enriched k-mers Found in SELEX Sequences	68
3. Binding Affinities of Selected SELEX Sequences and the Wild Type cTNT Binding Site	70

CHAPTER I

INTRODUCTION

According to the central dogma of biology, DNA codes for the functioning units of the cell, proteins. RNA acts as the intermediate, transferring information from the genes encoded in the DNA to create the diverse set of proteins that carry out the majority of the work needed by the cell. RNA is an unstable molecule and is heavily protected by a host of proteins that prevent its degradation. Although typically regarded as an intermediate, RNA is as powerful as protein and DNA in that it can compose a viral genome, regulate the translation of protein and can even be a catalyst, such as catalyzing the reactions necessary for splicing.

Splicing occurs co-transcriptionally, with a specifically orchestrated series of events transforming pre-messenger RNA, pre-mRNA, into mRNA. In most cases, the intron is initially recognized and bound by the U1 small nuclear RiboNucleoProtein (U1 snRNP), splicing factor 1 (SF1), and U2 auxiliary factor (U2AF), together termed E complex. U2 snRNP then replaces SF1 bound to the branch point sequence, termed A complex. The transition from E complex to A complex is ATP dependent. U1 snRNP is then replaced by the U4/U5/U6 tri-snRNP at the 5' splice site, termed B complex. C complex occurs next and includes the two transesterification reactions necessary for ligation of the two exons. The 2' hydroxyl group from the branch point adenosine attacks

the phosphate in the guanosine at the 5' splice site. The 3' hydroxyl from the recently freed exon attacks the 3' splice site of the second exon and ligates the two exons together. This continues until all exons are ligated to form the mRNA (reviewed in refs 1 and 2).

In addition to the spliceosomal complexes, other regulatory factors, termed alternative splicing factors, play a role in choosing which exons to include and which to exclude from the final mRNA. There are several ways to alternatively splice a pre-mRNA: cassette exons, mutually exclusive exons, alternative 5' and 3' splice sites, alternative promoters, alternative poly(A) sites, and intron retention (reviewed in ref 3). The job of an alternative splicing factor is to aid in splice site choice. Specific binding sites in the pre-mRNA, in addition to other protein and structure factors, communicate with alternative splicing factors to help with splice site decisions. The complexity of alternative splicing regulation has made it difficult to identify many alternative splicing regulators, much less understand the underlying mechanisms through which they operate. This dissertation addresses this exciting, active field of research by analyzing muscleblind, an alternative splicing regulator involved in myotonic dystrophy.

MYOTONIC DYSTROPHY

Myotonic Dystrophy falls under the category of muscular dystrophy, a general term for over forty diseases characterized by muscle wasting. There are two types of myotonic dystrophy, type I (DM1) and type 2 (DM2). They have many overlapping symptoms with only subtle differences, such as the age of onset and group of muscles affected. DM1 has a rare, congenital form and is also the most common form of adult-

onset muscular dystrophy while DM2 has no congenital form and is much less common (4). Also, DM1 symptoms include distal skeletal muscle degeneration at onset while DM2 include symptoms of proximal skeletal muscle degeneration (5). Both diseases have seemingly unrelated symptoms including myotonia, muscle wasting, insulin resistance, cardiac abnormalities and iridescent cataracts. In 1992, DM1 patients were identified as having aberrant CTG expansions in the dystrophin myotonia protein kinase gene (DMPK) and in 2001 DM2 patients were identified as having abnormal CCTG expansions in the first intron of the zinc finger 9 gene (ZNF9) (6, 7). Normal individuals have a smaller number of CTG and CCTG repeats, 5-37 and less than 75, respectively (6, 7). In DM1, the age of onset and severity of symptoms are linked to the number of repeats. Patients that develop DM1 as adults have around 100 repeats or less, patients that develop DM1 as young adults or children have up to 500 repeats while congenital cases are the most severe and have approximately 1000 repeats (8). DM2 patients usually have an average of 5000 repeats even though the disease is less severe than DM1 (9). The commonalities between the two diseases most likely stem from an RNA gain of function disease mechanism. Both repeats occur in non-coding regions, ruling out mutant protein expression as causing the diseases, which is the case with many triplet repeat diseases. Although there are some alterations in DMPK expression post-transcriptionally and the expression of a nearby gene (SIX5) is affected, these changes do not appear to play a large role in the disease mechanism because knock out mouse models of DMPK and SIX5 display only two of the symptoms of DM1 (10-14). In addition, changes in levels of DMPK and SIX5 expression cannot explain the same set of symptoms in DM2 as the repeats are in different locations.

Although levels of DMPK protein and transcripts are not the main culprit in DM1 patients, DMPK transcripts containing the aberrant CUG repeats accumulate in the nucleus of DM1 cells (15-17). Many proteins have been identified to co-localize or bind to the aberrant DMPK transcript, however, muscleblind is the most prevalent and prefers long CUG repeats (>20) over short (≤ 11) (18-20). Interestingly, muscleblind is a zinc finger protein that binds RNA and regulates alternative splicing (21). The proposed disease mechanism is that the expanded CUG repeats of the mutant DMPK transcript act as muscleblind binding sites and compete with muscleblind's normal RNA binding targets and sequester muscleblind to the expanded CUG repeats, which are observed as nuclear foci (22, 23). The DMPK transcript/MBNL foci are transient, with a portion of mobile MBNL and a portion of immobile MBNL, however, enough MBNL is sequestered to cause abnormal splicing events (21). These aberrant splicing events lead to symptoms of DM1 (Figure 1). This model is supported by a mouse knockout of MBNL1 in which the mice display symptoms of myotonia, cataracts and splicing defects, all of which are seen in DM1 patients (24).

Many transcripts have shown a dependence on MBNL1 for proper alternative splicing (reviewed in (4)). Alternative splicing accounts for the high level of protein diversity in humans, with over half the genes predicted to be alternatively spliced (3). Alternatively spliced transcripts require an elevated level of regulation because expression at the incorrect time or tissue may result in deleterious effects, including disease. Muscleblind regulates its pre-mRNA targets in a developmentally specific and tissue specific fashion, with its expression and nuclear localization corresponding

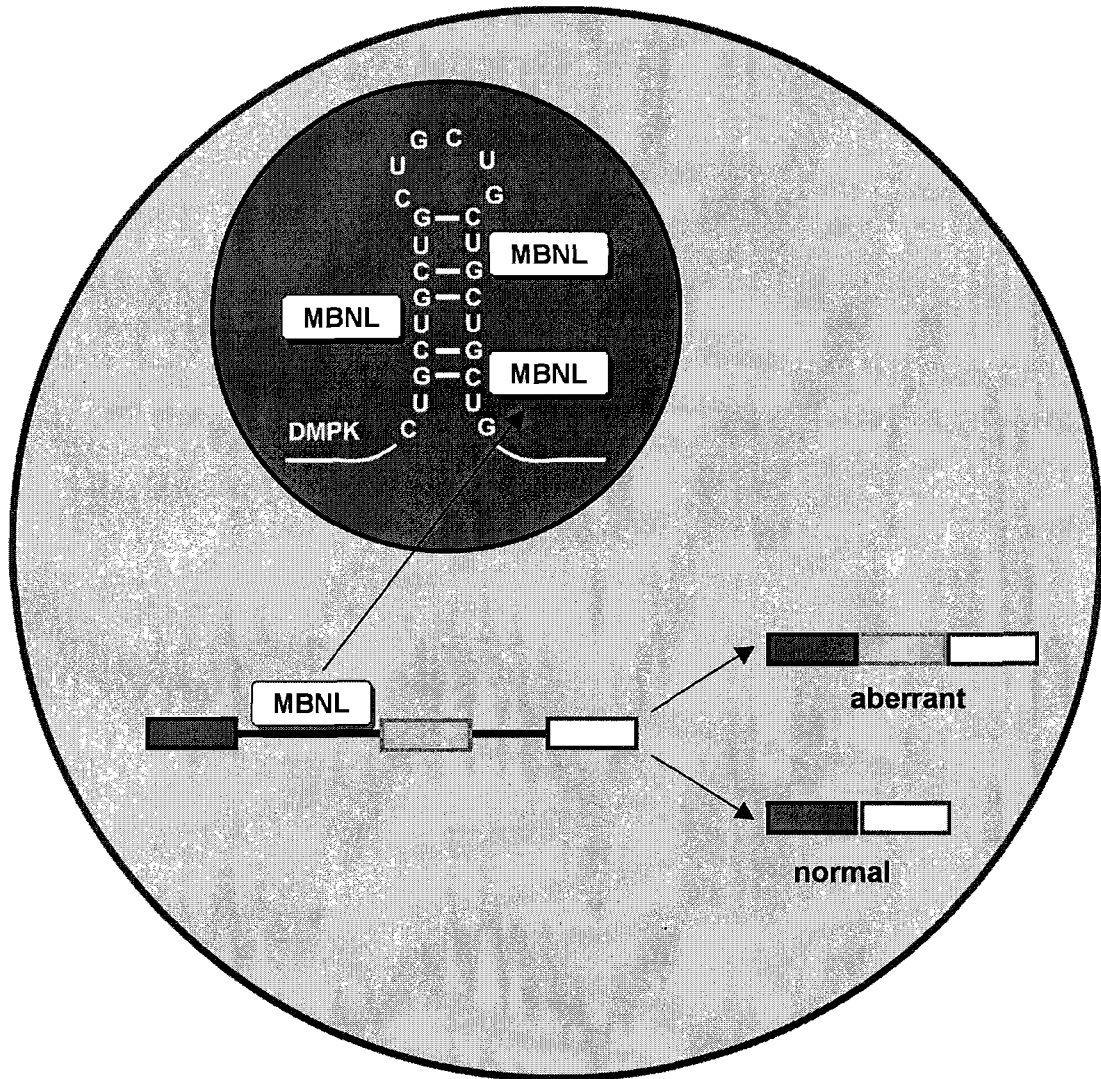


Figure 1. A disease model for myotonic dystrophy type 1. The nucleus of a DM1 patient cell is shown as the dark gray circle and the cytoplasm as the light gray circle. The stem structure in the nucleus represents aberrant CUG repeats that form stem-loops and sequester muscleblind. In the cytoplasm, MBNL binds a pre-mRNA transcript; the boxes represent exons and the black line, introns. In this example, MBNL binds upstream of the central exon and promotes exclusion. When MBNL is sequestered, it cannot promote exon exclusion and an aberrant transcript, labeled “aberrant”, would be translated. MBNL can also promote exonic inclusion as well.

to the post-natal development stage (25). MBNL1 regulates the alternative splicing of insulin receptor (IR) and chloride ion channel (CIC-1) transcripts, both of which are mis-regulated in DM1 patients (26-28). One isoform of IR, IR-B, is expressed in skeletal muscle and is responsible for binding insulin and aiding in glucose homeostasis. In DM1 patients, isoform IR-A is expressed which has a higher affinity for insulin, but a decreased signaling ability. The expression of IR-A in adults instead of IR-B leads to the symptom of insulin resistance in DM1 (28). CIC-1 is similarly mis-spliced with a truncated or non-functional form of CIC-1 expressed in DM1 patients. CIC-1 is responsible for maintaining chloride ion channels and regulating action potentials in muscle tissue (26). The main cause of myotonia, the inability to relax a contracted muscle, in DM1 patients is thought to arise from the hyperexcitability of muscle fibers leading to involuntary contractions. This is most likely a direct result of MBNL1 no longer regulating the proper splicing of CIC-1. Some evidence indicates that muscleblind in *Drosophila* also regulates splicing in a similar development and tissue-specific manner as MBNL1, as discussed in “the myotonic dystrophy type 1 *Drosophila* model” section.

A second alternative splicing regulator, CUG binding protein, CUG-BP, has been identified as regulating many of the same pre-mRNAs as MBNL1, including CIC-1 and IR (26, 28). MBNL1 and CUG-BP act antagonistically on many of the transcripts they regulate; when MBNL1 promotes exon inclusion, CUG-BP promotes exon exclusion and vice versa (see section on “alternative splicing regulation mechanisms”). CUG-BP has been implicated in the disease mechanism of DM1 as levels have shown to be increased in DM1 cells. This increase in CUG-BP could account for some of the mis-splicing

events (18, 28, 29). Evidence suggests that MBNL1 is the dominant player in DM1 because over-expression of MBNL1 in tissue culture is able to rescue many of the splicing defects seen in DM1. However, siRNA of CUG-BP has a milder rescue effect (18). Plus, it is controversial whether CUG-BP levels are in fact increased in DM1 patients.

Mis-splicing events are thought to be the main cause of DM1, with the partial list at 24 known mis-spliced transcripts (30). Symptoms of myotonia and insulin resistance have been accounted for by mis-splicing events, however, many symptoms remain unaccounted for. Other mis-splicing events will likely lead to symptoms of muscle wasting, iridescent cataracts, cardiac abnormalities and testicular atrophy. The full list of MBNL1-regulated transcripts is yet to be completed, but is likely to be extensive.

ALTERNATIVE SPLICING REGULATION MECHANISMS

Alternative splicing is involved in a host of cellular processes, including development, cell specificity and even sex determination. There are multiple mechanisms in which an alternative splicing factor can regulate these processes. In one mechanism, alternative splicing factors act by blocking splicing factors from binding to regions of the intron that are important for splicing. The polypyrimidine tract, the branch point sequence and the splice sites are all key binding sites for spliceosomal proteins. Also, recognition of intronic/exonic splicing enhancers (ISE's, ESE's, respectively) and intron/exonic splicing silencers (ISS's and ESS's, respectively) are essential for proper alternative splicing of many pre-mRNAs. For example, Sex lethal, Sxl, in *Drosophila*

binds an ISS in the polypyrimidine tract and the 5' splice site in an intron of the Mxl2 transcript. By binding in those regions, Sxl blocks both the binding of U2AF and the binding of TIA-1, a regulatory factor, to the 5' splice site. These simultaneous blocking events inhibit splicing of the first intron and eventually halt translation of the Mxl2 mRNA (31). The downstream effect in males, where Sxl is not expressed, is expression of Msl2 which regulates X chromosome dosage compensation. It is also possible for an alternative splicing regulator to act positively and recruit other splicing factors.

A second mechanism of alternative splicing regulation is for an alternative splicing regulator to act in conjunction with a cis-acting factor, such as secondary structure, within the pre-mRNA target. Little is known about this type of mechanism, but there is speculation that splicing factors can act to alleviate or promote secondary structure and thus promote or prevent, respectively, other splicing factors to bind. For example, the β -tropomyosin pre-mRNA has been shown to have a pair of stem-loops in the exon and the polypyrimine tract that prevent inclusion of the exon. However, when mutations abolish the stems, the exon can be recognized and included in the mRNA (32). It is possible that alternative splicing regulators are involved in destabilizing the stem in order to promote exon inclusion.

An example of secondary structures working with trans-acting factors to regulate splicing is the RBMY protein. The RBMY protein in humans is an active alternative splicing regulator in the testis and has been shown to bind with high affinity to a specific stem-loop sequence and stabilize it *in vitro*. It is possible that RBMY can bind a stem-loop in or near a regulated exon and stabilize the structure and prohibit binding by

splicing factors that bind single stranded RNA (33). This type of regulation has also been seen for the yeast protein RPL32 which regulates its own splicing. One of the alternatively spliced exons contains a stem-loop structure that includes the 5' splice site. L32 binds the stem-loop and most likely prohibits the binding of U1 RNA (34). The muscleblind family of proteins may act in a similar way by binding stem-loops near or within the branch point sequence and/or the polypyrimidine tract and blocking binding of splicing factors.

The binding site for MBNL1 in cTNT pre-mRNA is a stem-loop about 35 bases upstream of exon 5. MBNL1 binds this site and promotes the exclusion of exon 5. The stem forms on either side of the polypyrimidine tract, with the polypyrimidine tract in the loop region. Mutations to abolish or strengthen the stem result in decreased binding affinity of MBNL1 to the site *in vitro*. Interestingly, the same mutations in a cTNT minigene prevent MBNL1 from regulating the splicing of exon 5 in tissue culture (21, 35). One set of mutations extends the length of the polypyrimidine tract and potentially promotes the binding of U2AF, which would result in inclusion, instead of exclusion, of exon 5. In contrast, when the wild type stem is in place and MBNL1 is present, MBNL1 may bind and stabilize the stem-loop structure and prevent U2AF from binding. This would promote exon 5 exclusion. However, more work needs to be conducted before this mechanism can be established.

One proposed mechanism for how the muscleblind family of proteins may regulate splicing is by binding upstream of the regulated exon to promote exon exclusion, as seen in cTNT, and bind downstream to promote exon inclusion (21, 35). This type of

regulation has recently been shown for FOX-1 and FOX-2 (FOX-1/2), another family of tissue specific splicing regulators (36). Zhang et al. showed a clear bias in brain, skeletal muscle and heart tissues for FOX-1/2 to repress exon inclusion when binding sites were located upstream of the regulated exon, while sites downstream enhanced exon inclusion. Nova is a neural-specific splicing regulator and acts very similarly to FOX-1/2. Nova also displays this same pattern of regulation, when it binds downstream of the regulated exon Nova acts as an enhancer and when bound upstream functions as a repressor (37). This suggests that this class of regulatory splicing factors (FOX, Nova and muscleblind) share a common mechanism through which to regulate alternative splicing.

Alternative splicing regulators also can act on the same exon but one is a positive regulator while the other is a negative regulator. For example, MBNL1 and CUG-BP regulate several transcripts involved in DM1 pathogenesis in this manner, including cTNT exon 5. CUG-BP has been shown to bind several CUG motifs downstream of and promote the inclusion of exon 5 (38). MBNL1, on the other hand, binds upstream of exon 5 and promotes the exclusion of exon 5 (21). The two proteins act antagonistically, when MBNL1 levels are down, as in the case of DM1, exon 5 is included and the fetal isoform of cTNT is aberrantly expressed (4). In DM1 patients, exon 5 is aberrantly included, consistent with MBNL1 sequestration to CUG repeats. Levels of CUG-BP may be increased in DM1, which could also lead to expression of the cTNT fetal isoform of cTNT (38). Most likely, both proteins play a role in the pathogenesis of DM1, with MBNL1 being the dominant factor.

THE MUSCLEBLIND FAMILY OF PROTEINS

The muscleblind family of proteins are highly conserved and found in many, if not all, metazoans but not in bacteria, fungi or plants. This indicates that muscleblind is a

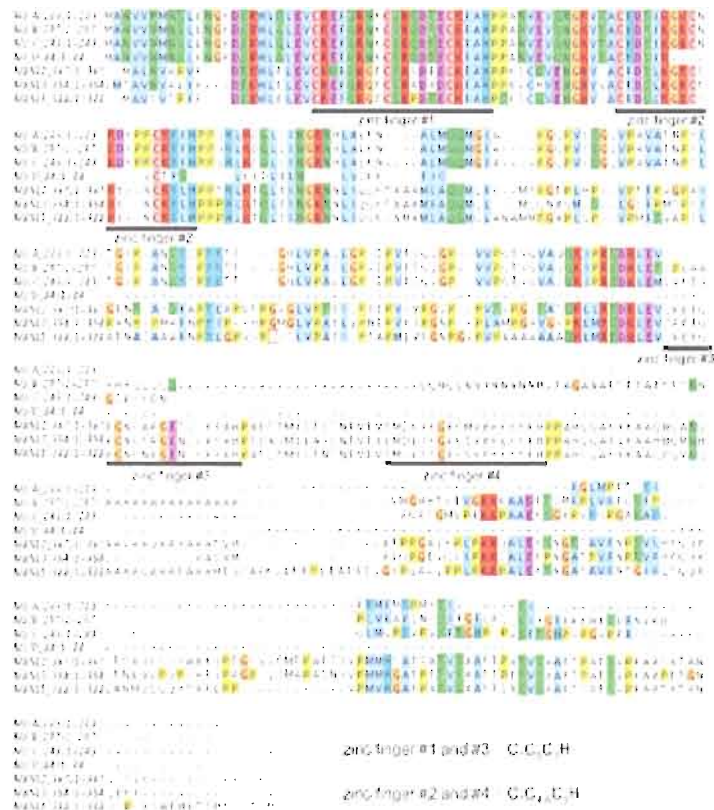


Figure 2. The muscleblind proteins are highly conserved, especially in the zinc finger, RNA binding domains. Shown is an alignment of all four *Drosophila* muscleblind splice isoforms (Mbl A-D) and the longest splice isoforms of the three human muscleblinds, MBNL1, 2 and 3. The number of amino acids in each isoform is displayed after the name of each protein. The zinc fingers are labeled below the amino acids that compose them. The shading indicates regions of similar amino acid identity. All seven proteins shown high sequence conservation for the first two zinc fingers.

relatively modern protein (39). Most invertebrates have only one muscleblind gene while most vertebrates have three muscleblind genes. Muscleblind proteins are composed of two to four zinc finger domains of the CCCH type. The spacing between the cysteines and histidines varies from domain to domain, but in general, is either $CX_7CX_6CX_3H$ or $CX_7CX_4CX_3H$, X being any amino acid. Less complex species like *Drosophila*, contain two zinc fingers, while more complex species, such as humans, have four (Figure 2). It appears that a duplication event of the first pair of zinc fingers occurred to produce genes with four zinc fingers. This is noticeable due to the similarity of the 1st and 3rd zinc finger and the 2nd and 4th in MBNL1. Other domains such as the LEV box, the NGR box, alanine-rich regions, phenylalanine rich regions and proline-rich regions are moderately conserved but have unknown functions. There are multiple, predicted phosphorylation sites throughout the muscleblind family of proteins, including some conserved in the zinc finger domains. When phosphorylated, they could potentially regulate the RNA binding function via steric hindrance (39).

The *Drosophila* muscleblind gene, Mbl, is expressed and spliced into 4 splice isoforms, A-D. A-C splice isoforms have identical 179 N-terminal residues and contain both the zinc fingers and are 203 amino acids, 316 amino acids, and 243 amino acids, respectively. Splice isoform D is much shorter, only 84 amino acids, and contains 1.5 zinc fingers. Mbl C is likely the most ancestral splice isoform as homologs are found in many other organisms (39). Mbl C is the most highly expressed throughout embryo, larval and pupal stages, with Mbl B and A at slightly lower expression levels. Mbl D is only expressed at the end of the larval stage and the beginning of the pupal stage. Mbl A,

B and D are not expressed in adult *Drosophila*, and C is moderately expressed in adults (40). This expression profile is a sign that Mbl is involved in developmental processes as discussed in “the myotonic dystrophy type I *Drosophila* model” section.

The human muscleblind-like proteins have multiple splice isoforms, MBNL1 has 9, MBNL2 has 3 and MBNL3 has 6 (20, 41, 42). The majority of splice isoforms of MBNL1 contain all 4 zinc fingers, indicating that they are probably capable of binding RNA. MBNL1 splice isoforms that contain only 0-3 zinc fingers are not capable of binding RNA, according to a yeast 3 hybrid study (19). Similar to the expression pattern of Mbl in *Drosophila*, mouse muscleblind-like proteins 1-3 are expressed throughout development from 4.5 dpc to 18.5 dpc (42). Human muscleblind-like proteins 1 and 2 are expressed in skeletal muscle, liver, placenta, brain, and heart tissues, with MBNL1 most highly expressed in skeletal muscle. MBNL3 is only expressed in placental tissue, not the other tissues tested (41). The differential expression patterns between MBNL1, 2 and 3 may indicate different roles, in addition to overlapping roles, for each of them. For example, MBNL2 shuttles integrin α_3 -mRNA to the required location in the cell membrane for proper translation (43).

Drosophila have only one muscleblind gene, not three, so it is possible that the four isoforms have some unique functions and some overlapping functions, similar to the varied functions of MBNL1, 2 and 3. In a Mbl *Drosophila* knockout of Mbl, Mbl C is able to rescue the embryonic lethal phenotype while Mbl B and A can only partially rescue the Mbl knockout. In addition, over-expression of Mbl C has the strongest effect on α -actinin alternative splicing. This indicates that the three isoforms are only partially

redundant, with Mbl C fulfilling the main roles. Mbl A, B and C are expressed in both the nucleus and cytoplasm of tissue culture cells with Mbl A preferentially expressed in the cytoplasm and Mbl B and C preferentially expressed in the nucleus. Mbl A, B and C all co-localize with CUG repeat RNA and form nuclear foci (40). It is possible that Mbl C fills a similar role as MBNL1 in humans, regulating a developmental switch. Evidence for this includes the ability of MBNL1 to rescue the lethal phenotype in Mbl knockout in flies (44). Mbl A, on the other hand, may perform a similar role to MBNL2, as it is primarily located in the cytoplasm and MBNL2 regulates the localization of integrin α_3 -mRNA in the cytoplasm (43). Overall, the *Drosophila* muscleblind splice isoforms and the human muscleblind splice isoforms have very similar functions, if not identical.

THE MYOTONIC DYSTROPHY TYPE I *DROSOPHILA* MODEL

Muscleblind was first identified in a *Drosophila* genetic screen looking for genes that are required for photoreceptor development and/or differentiation. In Mbl null flies, deficient cells are recruited into developing ommatidia, begin early stages of differentiation but are unable to fully differentiate into terminal photoreceptors. This indicates that Mbl is required for photoreceptor differentiation (45). Artero et al. showed that muscleblind was also important for terminal muscle cell differentiation (46). Mutant Mbl *Drosophila* are paralyzed, have contracted abdomen, lack striation of muscles and die as embryos or in the early larval stage (46). This coincides with ubiquitous expression of Mbl in the later stages of terminal muscle differentiation. Mbl was originally hypothesized to be involved in transcriptional regulation and perhaps bind nucleic acids,

but the function for its zinc finger domains was not known. Currently, Mbl has been shown to bind RNA and regulate alternative splicing (8, 47, 48).

In DM1, the expanded CUG repeat RNA gains a toxic function by recruiting MBNL1 and causing symptoms of the disease. To investigate RNA sequestration of Mbl in *Drosophila*, de Haro et al. created a *Drosophila* line that expresses 480 interrupted CUG repeats (8). In this DM1 *Drosophila* model, Mbl was sequestered to CUG repeats in nuclear foci, there was muscle wasting and degeneration of indirect flight muscles, disorganization and fusion of ommatidia and duplication or loss of bristles in the eyes. These phenotypes are similar to DM1 symptoms in humans of muscle wasting and cataracts. In addition, over-expression of MBNL1 relieved the eye phenotype while over-expression of CUG-BP enhanced this phenotype. This is similar to cTNT splicing events in DM1 tissue and cell culture. CUG-BP over-expression causes expression of the developmentally incorrect isoform of cTNT while over-expression of MBNL allows for expression of the proper cTNT isoform (21, 38).

The *Drosophila* DM1 model suggests that Mbl plays a similar functional role as MBNL1, an alternative splicing factor. In fact, Mbl regulates the alternative splicing of two transcripts shown to be regulated by MBNL1, ZASP and TNNT3 (47, 48). In addition, Mbl is highly expressed throughout all pre-adult life stages, and expressed at lower levels in adult *Drosophila*, indicating a role as alternative splicing regulator during critical stages of development (40). Both of these transcripts, ZASP and TNNT3, are mis-spliced in DM1 patients, indicating that the mis-splicing may give rise to disease symptoms. It is not surprising that Mbl and MBNL1 play similar roles in the cell, as they

are highly conserved, especially in the RNA binding domains which are involved in the binding of the RNA and the regulation of splicing.

RNA BINDING SPECIFICITY OF MBNL TO CUG REPEATS AND OTHER TARGETS

Muscleblind has only recently been discovered to be an RNA binding protein. The first in-depth look at MBNL1's binding to RNA and DNA was carried out by Kino et al. (19). They show that MBNL1 has a preference for binding RNA over DNA and that binding is salt dependent, with NaCl concentrations of 250 mM or greater causing disruption of the MBNL1/RNA complex. They used a yeast 3-hybrid assay and confirmed that MBNL1 prefers longer CUG repeats (16+), over short repeats, as was also seen in crosslinking experiments of MBNL1 to CUG repeats (20). Interestingly, MBNL1 is capable of binding a CUG repeat of only 4, however this was achieved by forcing a stem-loop conformation (35). This coincides with MBNL1 having a preference for binding longer CUG repeats because they are able to form stem-loops, whereas shorter repeats are not (49).

A yeast 3 hybrid study showed that MBNL1 binds CCUG, CCCG, CUUG and CAAG repeats, with a preference for CCUG and CCCG repeats. These repeat RNAs are capable of forming stems with mismatches. However, MBNL doesn't bind repeats, like CUG₁₆/CAG₁₆, that do not form mismatches (19). This was also seen in a mutational screen of CUG repeats where various positions along the CUG repeat stem were mutated. When the U-U mismatch was replaced with anything other than a pyrimidine-pyrimidine

mismatch, such as a purine-purine mismatch, purine-pyrimidine mismatch, or Watson-Crick base pairing, binding was significantly decreased. Changes in 5' to 3' orientation and flipping the C-G base pair to a G-C base pair decreased MBNL1's binding affinity as well (35). MBNL1 is a specific RNA binding protein, with a preference for stem-forming CUG repeats and pyrimidine-pyrimidine mismatches. Because Mbl contains the same RNA binding domains as MBNL1, analysis of the RNA binding specificity of Mbl could give insight into endogenous binding sites for both MBNL1 and Mbl.

DMPK transcripts with expanded CUG repeats are retained in the nucleus of DM1 patients and co-localize with MBNL1 (22, 23). MBNL1 binds double stranded RNA and thus prefers to bind longer CUG repeats as they can form stem-loops easily. The CUG repeat stem-loops are "slippery", with the sides of the stem sliding and creating different registers. Also, the longer the CUG repeat, the more stable it becomes (49). It is possible that MBNL1 can bind and stabilize the stem. The CUG repeat stem consists of two C-G base pairs interspersed with U-U mismatches in a relaxed, A-form state (49, 50). This has been shown in the crystal structure of CUG repeat RNA and surprisingly the presence of multiple U-U mismatches doesn't significantly alter the CUG repeat RNA structure compared to the canonical A-form RNA structure. The six CUG repeat structure reveals an accessible minor groove where MBNL1 most likely interacts directly with the U-U mismatches. In electron microscopy studies, MBNL binds to the middle portion of the CUG repeat stem, not the loop or the base (51). On the other hand, CUG-BP has been shown to bind near the base of CUG repeat stems most likely in the tail region. This coincides with CUG-BP binding single stranded RNA (52). This supports

the model that MBNL1 co-localizes with long CUG repeats because they form double stranded structures, while CUG-BP, which binds single stranded RNA, does not.

The sequestration of MBNL1 to the CUG repeats in DM1 patients results in the decrease or absence of MBNL1 at its endogenous RNA binding sites. These endogenous targets are hypothesized to be pre-mRNA transcripts of which MBNL1 regulates the alternative splicing. One such target is cTNT, where it has been shown to bind upstream of and promote the exclusion of exon 5 (21, 35). The binding site is within a 32 nucleotide region that folds into a stem-loop, similar to a short CUG repeat stem-loop. The cTNT stem-loop contains several C-G and U-G base pairs and two pyrimidine-pyrimidine mismatches, but lacks adenosines. Strengthening the cTNT stem by adding two Watson-Crick base pairs or weakening the stem by mutating 4 guanosines, decreased binding by MBNL1 *in vitro* (35). This indicates that MBNL1 prefers weaker stems over stronger, purely Watson-Crick stems. A potential binding site motif for MBNL1 of 5'-YGCU(U/G)Y-3' was postulated by Ho et al., which is different than the 5'-CHHG-3' motif put forth by Kino et al. (19, 21). The 5'-YGCU(U/G)Y-3' motif was identified by looking at 4 potential binding sites in cTNT introns, 2 in a human cTNT intron and 2 in a chicken cTNT intron. The 5'-CHHG-3' motif was identified by a yeast 3 hybrid assay that used various RNA 3-mer and 4-mer repeats as bait. It is possible that MBNL1's binding site is degenerate and that a combination or portions of both motifs are correct. It is also possible that MBNL1 recognizes different binding sites in endogenous targets than it does in CUG or CCUG repeats. Evidence for this comes from the binding affinity for the endogenous cTNT binding site being 5-10 fold higher affinity than the affinity for

CUG repeats *in vitro*. More information about MBNL1's endogenous binding sites will help to determine the correct binding motif.

SYSTEMATIC EVOLUTION OF LIGANDS BY EXPONENTIAL ENRICHMENT

MBNL1 and Mbl regulate alternative splicing, however only a handful of pre-mRNA targets have been identified. Within those targets, only one MBNL1 and no Mbl binding sites have been recognized. The understanding of both DM1 symptoms and muscleblind's functions in regulating splicing would be helped greatly if binding sites were identified. This would allow for the creation of a complete list of muscleblind targets and binding locations within each target. Systematic evolution of ligands by exponential enrichment (SELEX) is a technique used in Chapter II and III to discover RNA sequence elements and structural elements important for binding by Mbl and MBNL1.

SELEX was established in 1990 as a technique to identify high affinity and/or high specificity RNA aptamers (53, 54). SELEX applications cover a broad range, the most common include identification of binding site motifs and selection for more efficient or specific ribozymes (55). Examples include identification of RNA aptamers that bind specifically to organic dyes, cleave single stranded DNA substrates more efficiently, bind secondary structures with high affinity, and bind specifically to a protein and block protease activity (53, 54, 56, 57). SELEX has even aided in understanding what sequence and structural element the HIV REV protein binds in the HIV RNA

genome (58). With this information, drugs can be designed to disrupt the REV/HIV genome complex.

SELEX can be performed with many permutations in technique; however, the general scheme includes the design of a large starting pool of randomized DNA. The initial DNA pool is then transcribed to create a large random pool of RNA. The protein or target is then immobilized on beads or a surface. The pool of RNA is incubated with the target to allow binding to occur. Washes are performed to remove the RNA that is not bound. The bound RNA is harvested by chemical or heat denaturation of the target to release the RNA. The RNA is then reverse-transcribed and PCR is performed to create DNA. The DNA is then transcribed into a new pool of RNA termed round one. This round one pool of RNA is enriched in RNA molecules that bind the target. This cycle continues until enrichment is no longer taking place, anywhere from one cycle to tens of cycles are performed (Figure 3).

SELEX is highly variable and can be suited to many different RNA-target combinations. Each step of the SELEX process can be catered to the identification of particular aptamers. The template usually includes two constant regions that play a dual role; they serve as priming sites for the RT-PCR step and as an RNA polymerase T7 binding site. Usually the size of the constant regions are minimized to limit interaction of the target to areas other than the random region. The random region can vary in length according to the target, if a smaller binding site or a smaller active site is required, then the region may be around 10 to 40 residues. However, if one is looking for a larger binding site, one may use up to 100 random residues or more. A highly diverse initial

pool is key to ensure that the best aptamers are available. Commonly $10^{14} - 10^{15}$ unique RNA molecules are used in the first round of SELEX (59).

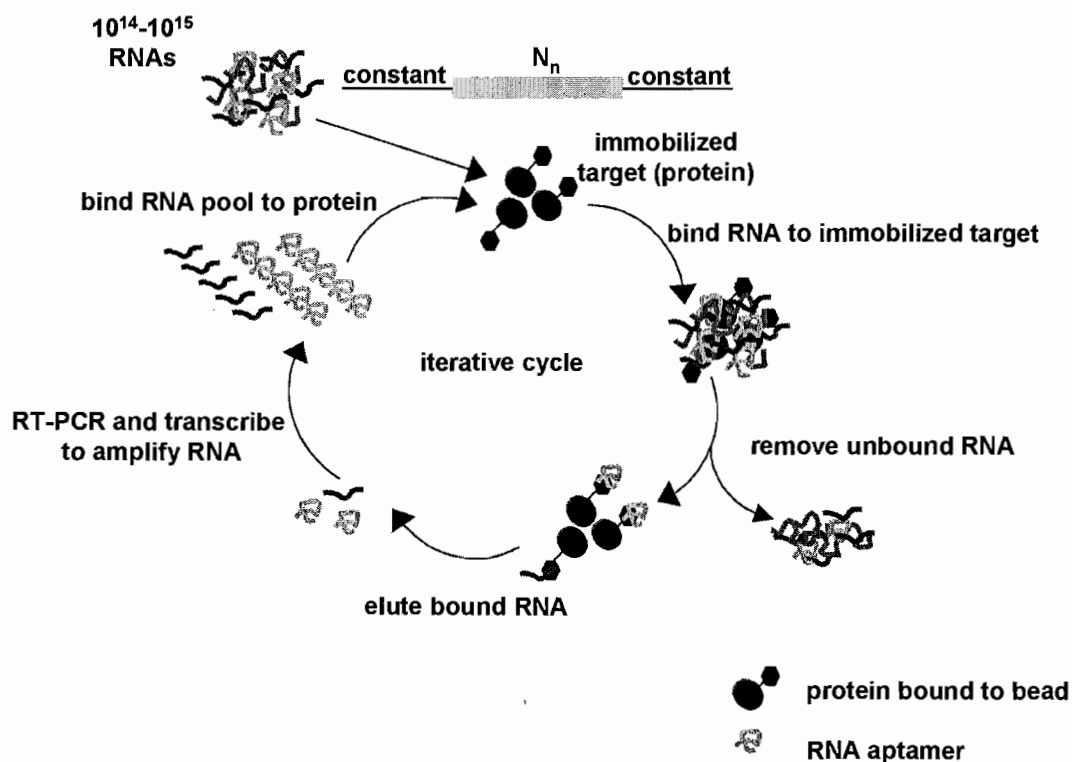


Figure 3. Systematic evolution of ligands by exponential enrichment (SELEX). A random pool of RNA aptamers composed of two, flanking constant regions and a randomized region is bound to an immobilized target. The immobilized target and pool of RNA are incubated. The unbound RNA is removed and the bound RNA aptamers are eluted. The RNA aptamers are reverse transcribed and PCR is performed to amplify a pool of DNA. The DNA is transcribed to create an enriched pool of RNA aptamers. The process is repeated until enrichment of RNA aptamers is no longer observed.

The binding conditions should take into account the interaction between the RNA aptamer and the target so that some selectivity can take place. It is important to have a portion of the RNA bind, but not all of it. The stringency of the binding conditions can

be increased as the rounds increase to ensure removal of low affinity aptamers. Stringency can be changed by increasing the ratio of RNA to protein or increasing the concentration of salt and/or a nonspecific inhibitor. This allows the target to sample large numbers of RNA molecules and bind only the highest affinity aptamers. To monitor the effectiveness of each round, assays are performed to determine how many rounds are required. In the case of identifying RNAs that bind with high affinity to a protein of interest, gel mobility shift assays are performed to check the change in binding affinity over the rounds. Once there is no longer an increase in binding affinity, the first round where the highest affinity is achieved is cloned and sequenced. It is important to not over-select as the aptamer diversity or biological relevance may be compromised.

A variation on the common SELEX technique is doped SELEX. Instead of a completely randomized region for the target to bind, the “randomized” region is doped with varying amounts of a known, wild type binding site. The doping can be as extensive as the majority of wild type positions being randomized (partially or fully) or as low as only a few positions being partially randomized with the rest remaining wild type. This technique is useful if a large region of RNA is known to have a protein binding site in it, but the exact nucleotides it interacts with or the nucleotides required to set up a specific structure are unknown. In the case of the HIV-1 REV protein, which binds the HIV-1 genome, the binding region had been narrowed down to 66 nucleotides. A doped SELEX narrowed down REV’s binding specificity to an essential G-G bulge in a stem-bulge-stem region (58). Doped SELEX is a powerful tool to identify specific structural or sequence elements that are required for binding by a protein of interest.

DISSERTATION OVERVIEW

The previous sections have outlined the role that muscleblind proteins play in both the disease mechanism of myotonic dystrophy type 1 and as regulators of alternative splicing. Muscleblind regulates the splicing of several important tissue and developmentally specific transcripts. When MBNL1 becomes sequestered to aberrant CUG repeats in DM1 patients, MBNL1 can no longer regulate these transcripts. This results in expression or degradation of inappropriately spliced mRNA. The lack of appropriately spliced products leads to symptoms of DM1, such as myotonia and insulin resistance. However, many more symptoms are unaccounted for and only a handful of MBNL1-regulated pre-mRNAs have been identified. In addition, only one endogenous binding site has been identified for MBNL1.

The *Drosophila* DM1 model has already proven helpful in understanding the role of muscleblind in development and CUG/Mbl foci formation. There is only one muscleblind gene in *Drosophila*, not three like in humans, and Mbl only contains two zinc fingers, not four like humans. This simplicity makes Mbl an ideal protein to characterize muscleblind's RNA binding specificity. In Chapter II, binding of Mbl to structured CUG repeat RNA is analyzed. In addition, SELEX is performed in order to determine key sequence and structural elements required for binding by Mbl. The results of these studies indicate that Mbl is a highly specific RNA binding protein that prefers both single stranded and double stranded regions of RNA as well as guanosines. Chapter II is currently published in the *Biochemistry* (Vol. 47, pp. 7284-7294, 2008) and titled "RNA Binding Specificity of *Drosophila* Muscleblind". I am the first author and

performed the majority of the experimental work with the help of Devika P. Gates. Rodger B. Voelker, the third author, performed the bioinformatics. My advisor, J. Andrew Berglund, is last author and helped in authorship, editing and experimental design.

The identification of additional MBNL1 binding sites in endogenous targets is useful for linking mis-splicing events to DM1 and DM2 symptoms and characterizing MBNL1's mode of alternative splicing regulation. Little is known about the mechanisms by which tissue specific alternative splicing regulators work. Even less is known about splicing regulators that bind structured RNA, like MBNL1. Chapter III discusses the results from a doped SELEX performed on the only known MBNL1 binding site in cTNT pre-mRNA. This doped SELEX aids in understanding MBNL1's RNA sequence preferences. Chapter III work is currently being prepared as a manuscript with the help of Devika P. Gates , Rodger B. Voelker and J. Andrew Berglund.

CHAPTER II

RNA BINDING SPECIFICITY OF *DROSOPHILA* MUSCLEBLIND

Reproduced with permission from Goers, E. S., Voelker, R. B., Gates, D. P., and Berglund, J. A. (2008) RNA binding specificity of *Drosophila* muscleblind, *Biochemistry* 47, 7284-7294.

INTRODUCTION

Drosophila muscleblind (Mbl) is an RNA binding protein important in differentiation and development. Mbl is known to have roles in photoreceptor differentiation as well as terminal muscle differentiation (1, 2). In addition to roles in differentiation, Mbl is necessary for early development as Mbl null flies die at late embryo stages or early larval stages (2). Mbl has recently been shown to regulate the alternative splicing of two genes involved in the cytoskeleton, α -actinin and *tun* (3, 4). The human orthologues of Mbl are the muscleblind-like proteins (MBNL1-3) and these proteins also regulate alternative splicing of transcripts important in development (5, 6). Alternative splicing of the Mbl transcript leads to the production of at least four different Mbl protein isoforms ranging in size from 84 to 316 amino acids. Three Mbl isoforms presumably confer RNA binding through two CX₇CX₆CX₃H zinc fingers; the shortest Mbl isoform contains only one zinc finger.

The MBNL proteins have been shown to play an important role in myotonic dystrophy (DM) in humans. There are two subtypes of DM, both involving repeat expansions that result in toxic RNAs. Myotonic dystrophy type 1 (DM1) is caused by an expansion of CTG repeats in the 3'-UTR of the *DMPK* gene (7-9), while myotonic dystrophy type 2 (DM2) is caused by an expansion of CCTG repeats in the first intron of the *ZNF9* gene (10). These two types of expansions appear to function through a common disease mechanism: upon transcription, both the CUG and CCUG repeats fold into double-stranded stem-loops with the helices adopting a primarily A-form structure (11-13). MBNL proteins bind these stem-loop structures leading to the sequestration of these important splicing factors in nuclear foci (14-16). MBNL sequestration results in mis-splicing of various transcripts, several of which have been linked to symptoms observed in DM patients (reviewed in refs 5 and 6).

Approximately 20 different human pre-mRNAs have been shown to be mis-spliced in DM (6). The most well-characterized example of mis-splicing in DM is the cardiac troponin T (cTNT) pre-mRNA in which MBNL1 has been shown to promote exclusion of an alternatively spliced exon (17). cTNT is developmentally regulated, and the presence or absence of this exon determines if the expressed protein is the fetal or adult isoform, respectively (18-21). Incorrect cTNT splicing in DM1 patients may be linked to cardiac abnormalities, one of the symptoms of DM1 (5). The binding site for MBNL1 in the cTNT pre-mRNA has been mapped to a region upstream of the skipped exon (17). Recently, we have shown that this section of cTNT RNA folds into a stem-loop with mismatches and adopts a structure similar to the CUG repeat RNA which

contains U-U mismatches (22). Yuan and colleagues have also shown that MBNL1 binds a stem-loop containing a pyrimidine-pyrimidine mismatch in an intron near the *Tnnt3* fetal exon (23), suggesting MBNL recognizes structured RNA elements in its pre-mRNA targets.

Drosophila has been used as a model to investigate several aspects of DM1 pathology. *Drosophila* overexpressing 162 CUG repeats have nuclear foci that contain localized Mbl and CUG repeats but do not display any abnormal phenotypes (24). Longer, interrupted CUG repeats (480 repeats) result in nuclear foci and cause the flies to have muscle, eye, and other degeneration phenotypes, consistent with the Mbl null flies (1, 25). Overexpression of human MBNL1 can rescue these mutant phenotypes (26), indicating MBNL1 and Mbl are functionally conserved. Since *Drosophila* has one *mbl* gene with four protein isoforms (while humans have three muscleblind-like genes, each with multiple isoforms) experimental design and data analysis can be simplified when using *Drosophila* as a model system. Characterization of the RNA binding properties of *Drosophila* Mbl will lead to a better understanding of its role in alternative splicing and the role of MBNL in DM. In this study we show that recombinant Mbl (amino acids 1-105) is a highly specific RNA binding protein with a strong preference for structured RNA containing specifically placed guanosines.

MATERIALS AND METHODS

Protein Construct Design and Protein Purification

Mbl cDNA (amino acids 1-105) was amplified by PCR from the *Drosophila* cDNA library DGC release 2 using primers: 5'-GCGAATTCGCCAACGTTGTCAATATGAACAGC-3' and 5'-ATAGTTATGCGGCCGCTCAGCCCATCTGTTGCATCAGTGC-3' and restriction digested with EcoR1 and Not1. The PCR product was ligated into pGEX6P-1 (Amersham) to create the Mbl (1-105) construct, which has a N-terminal glutathione S-transferase (GST) tag. This construct was expressed in *Escherichia coli* Rosetta cells and grown in LB medium with 50 mM ZnCl₂, 50 mg/mL Amp, 25 mg/mL chloramphenicol, and 2% glucose at 37°C. Cells were induced at an OD₆₀₀ of 0.5-1.0 with 0.4 mM IPTG at 30°C for 3 h. Cells were then pelleted and stored at -80°C. The pellet was dissolved in lysis buffer (500 mM NaCl, 25 mM Tris pH 8.0, 10 mM BME and 5% glycerol), sonicated, and centrifuged at 31,000 x g for 15 min. The supernatant was bound to glutathione-agarose beads for 30 min to 1h at 4°C. The protein was eluted with 10 mM reduced glutathione, diluted to 60 mM NaCl, and run over a Source 30Q anion-exchange column (Amersham). Protein impurities bound to the column, but Mbl did not; therefore, flow-through was collected, concentrated, and dialyzed into protein storage buffer (50% glycerol, 25 mM Tris pH 7.5, 5 mM BME, 0.1 % triton x-100 and 500 mM NaCl) overnight at 4°C. The presence of the GST tag does not appear to alter RNA binding because removing GST did not alter the RNA binding of Mbl (data not shown).

Transcription and Kinase Reactions

RNA oligonucleotides were transcribed with T7 RNA polymerase and (α - ^{32}P) CTP from linearized plasmids, PCR products or fully complementary DNA oligos. The transcription reaction mixtures were incubated at 37°C for 3 h or 25°C overnight. RNAs were kinased using T4 polynucleotide kinase and [γ - ^{32}P] ATP. Both transcribed and kinased RNAs were gel purified with 8% denaturing polyacrylamide gels.

Gel Mobility Shift Assay

Radiolabeled RNA was heated at 95°C for 3 min and then placed directly on ice for 10 min in binding buffer (20 mM Tris pH 7.5, 5 mM MgCl₂ and 100 mM NaCl). RNA was equilibrated to 25°C for 5 min. Heparin (final concentration of 0.5 mg/mL) and loading dye were added to the RNA. Protein was serially diluted into protein storage buffer on ice and aliquoted into 2 mL per binding reaction mixture at 25°C. RNA in binding buffer (8 mL) was added to the aliquoted protein and allowed to bind for 15-30 min at 25°C. The RNA concentration was equal to or less than the dissociation constant (K_d) value for each reaction. RNA-protein complexes (2 ml per binding reaction mixture) were resolved with a native polyacrylamide gel (6% 37.5:1 acrylamide/bisacrylamide mixture and 0.5X TB) run at 4°C for 30 min at ~150 V. The gels were dried and exposed to a phosphoimager screen (Molecular Dynamics) overnight at 25°C. K_d values were determined on the basis of quantitation of the bound and unbound RNA fractions using ImagQuant (Molecular Dynamics). K_d values were calculated as described by Warf and Berglund (22).

SELEX

The template for the SELEX (systematic evolution of ligands by exponential enrichment) (27) experiment was the 80-mer DNA oligonucleotide 5'-

GGGAATGGATCCACATCTACGAATTCN₃₀TTCAGTGCAGACTTGACGAAGCTT-

Table 1: Concentration of RNA and Protein used in SELEX

round	RNA (μ M)	Protein (μ M)
1	20	5
2	1.2	0.12
3	4.3	0.43
4	2.4	0.24
5	5.6	0.56
6	0.56	0.056

3'. The forward primer for SELEX was 5'-

GATAATACGACTCACTATAGGGAATGGATCCACATCTACGA-3' and the reverse

primer was 5'-AAGCTTCGTCAAGTCTGCAGTGAA-3'. The initial PCRs were

amplified by eight rounds of PCR and calculated to a concentration of 1×10^{14} molecules.

The transcription reactions were conducted under the following conditions: 100-2000 ng

of template, 40 mM Tris pH7.9, 26 mM MgCl₂, 2 mM spermidine, 10 mM NaCl,

nucleotides (5 mM each), 0.1 μ g of yeast pyrophosphatase/100 μ L, \sim 2 mg/mL T7

polymerase, and 40 mM DTT. The transcription reactions were performed with small

amounts of [α -³²P] CTP for monitoring purposes. After transcription, reactions mixtures

were treated with DNase for 1 hr at 37°C, gel purified, ethanol precipitated and run over a

Bio-Spin-6 size exclusion column (Bio-Rad). For each round, the appropriate

concentration of Mbl (1-105) was bound to 20 mL of glutathione-agarose beads at 4°C for 15 min, then washed three times with 200 mL of SELEX binding buffer (160 mM NaCl, 25 mM Tris pH 7.5, 5 mM MgCl₂, 0.002% triton x-100 and 1 mM BME added the day of use). This amount of beads should be enough to completely bind all of the added Mbl. All washes were conducted in a similar manner. The RNA was heated in SELEX binding buffer at 95°C for 3 min and placed immediately on ice for 10 min. tRNA (0.1 mg/mL) was added, as a nonspecific inhibitor, to the RNA (in all rounds except round 3). Negative selection was carried out on glutathione-agarose beads prior to binding to Mbl-bound beads by prebinding the RNA to the glutathione-agarose beads. The Mbl-bound beads were incubated with the RNA at 25°C for 20 min. The beads were washed three times in 200 mL of SELEX binding buffer and RNA was released and collected from the beads by a phenol/chloroform extraction and subsequent ethanol precipitation. The collected RNA was reverse transcribed using AMV reverse transcriptase, AMV buffer, 8.8 μM reverse primer, and two-thirds of the RNA isolated after the round for 1 hr at 42°C. After the reverse transcription was complete, the DNA was amplified by 11 PCR cycles. The DNA was transcribed and the SELEX cycle repeated six times with the concentration of RNA and Mbl varying according to Table 1. After the rounds were completed, individual clones were isolated by TOPO cloning and then sequenced.

Structure Probing and Footprinting

RNA was transcribed, and the terminal triphosphate was removed using shrimp alkaline phosphatase (USB) and then kinased using T4 polynucleotide kinase (NEB) and [³²P]ATP to 5'-end label the RNA. The RNA was gel purified and

ethanol precipitated. Purified RNA was incubated with either RNase T1, RNase V1, or RNase 1 or under alkaline hydrolysis conditions (Ambion) in footprinting buffer and protein storage buffer (20% glycerol, 0.04% triton X-100, 2 mM BME, 20 mM Tris pH 7.5, 200 mM NaCl, 100 mM KCl, 0.5 mg/mL tRNA and 10 mM MgCl₂) to a final volume of 10 μ L for 3 min. Reactions were quenched with phenol-chloroform and extracted. In the reaction mixtures containing Mbl, Mbl was added in protein buffer to a final concentration of 4 or 12 μ M and incubated with RNA for 10 min at 25°C before RNases were added. After the phenol extraction, the RNA was ethanol precipitated and cleavage products were resolved with a denaturing 15% polyacrylamide gel for ~2 h at 35 W. The gel was dried and exposed to a phosphorimager screen overnight at 25°C.

Determination of n-mer Bias in the SELEX Sequences

SELEX was initiated with a population of RNAs containing a uniformly random sequence; however, the random sequence was flanked by two constant regions that are necessary for amplification. It is possible that the nonrandom sequences can inadvertently influence the selection process (e.g., by directly acting as a binding site or by interacting with the nonrandom region to form secondary structures that influence binding).

Therefore, we developed a method for calculating enrichment that attempts to account for the constant regions. To calculate the degree to which *n*-mers were enriched during the SELEX, we first performed a sliding-window count of all *n*-mers (3 to 5 nt) within the non-redundant sample of 25 selected sequences. This count included the entire constant and random regions. To calculate the maximum likelihood probabilities for expected occurrences we also performed a similar sliding-window count on a set of 500,000

computer generated sequences containing a uniformly random 30 nt region flanked by the same constant regions used in the experiment. For both of these data sets the counts were transformed to probabilities and the enrichment was determined according to the binomial confidence interval method previously described (28).

RESULTS

A Minimal Mbl Protein Binds Specifically to Expanded CUG and Short CUG Repeats

To characterize the RNA binding specificity and affinity of Mbl, we have designed a minimal protein – RNA system. Initially we began our RNA binding studies with a full-length version of Mbl (amino acids 1-243), isoform C of the four different alternative splice forms of Mbl. However, due to its low level of solubility and instability, Mbl (1-243) proved difficult for use in binding studies. The first 100 amino acids are highly conserved and contain the zinc finger domains. We therefore truncated Mbl by removing the C-terminus and fused the N-terminal 105 amino acids to an N-terminal GST tag, creating a more stable protein. This construct will be referred to as Mbl and all experiments were performed with Mbl unless otherwise indicated.

Mbl bound to fifty four CUG repeats (CUG_{54}) with a K_d of 500 nM (Figure 1A) while a minimal RNA containing only four CUG repeats with a UUCG cap (CUG_4) bound Mbl with a K_d of 430 nM. (Figure 1A). The UUCG cap is an ultra-stable loop that forces stems to form and maintain stability (29). Normally 10 or more CUG repeats are necessary for duplex formation and human MBNL1 binding (16). This minimal (CUG_4) RNA facilitates the dissection of the RNA binding specificity of Mbl by replacing

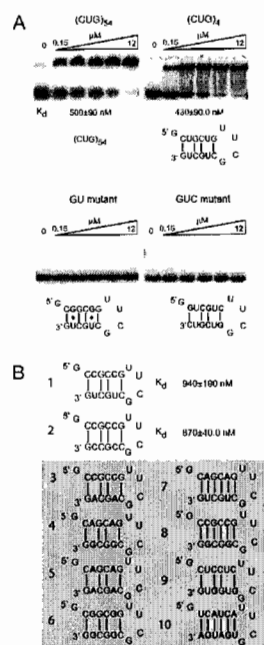


Figure 1. Mbl binds CUG repeats with a high degree of specificity. (A) The four gels are gel mobility shift assays with the name of the RNA above each gel, the structure of the RNA below each gel, and increasing concentrations of Mbl along the top. The GU mutant gel contains (CUG)₄ with the U-U mismatches replaced with G-U base pairs. The GUC mutant has reversed 5' to 3' orientation compared to (CUG)₄. The nucleotides mutated relative to (CUG)₄ are shown in bold. The Mbl concentrations in panel A are 0.15 μM, 0.45 μM, 1.4 μM, 4.1 μM, and 12 μM. The K_d values refer to the RNA oligonucleotides shown in the gel above. (B) The secondary structures of 10 other CUG mutant RNA oligonucleotides that were tested for binding by Mbl. Only the C-U and C-C mutants (1 and 2) bound to Mbl, while the other eight RNA oligonucleotides (numbers 3-10, shaded in gray) showed no binding with up to 12 μM Mbl. The nucleotides mutated relative to (CUG)₄ are shown in bold. The K_d values were determined on the basis of a minimum of three independent gel mobility shift experiments.

the U-U mismatches and/or C-G base pairs with other bases. Using this strategy we measured the binding affinities of twelve (CUG)₄ mutants. We found that when the U-U mismatches were replaced with C-U or C-C mismatches, Mbl bound with only a 2-fold decrease in affinity compared to (CUG)₄ (Figure 1B). However, mutating the U-U mismatches to G-U wobble base pairs or purine-purine mismatches such as A-G, A-A or G-G abolished binding (Figure 1B). Replacing the U-U mismatches with Watson-Crick base pairs also disrupted binding by Mbl (RNA #7 and #8, Figure 1B). Furthermore,

changes to the C-G and G-C Watson-Crick base pairs including reversing the orientation of the RNA (5'-GUC-3') and making one strand all pyrimidines and the other all purines eliminated binding by Mbl (Figure 1B). Finally, an RNA with different Watson-Crick base pairs at all of the positions didn't bind Mbl (RNA #10, Figure 1B). These results indicate that all of the positions of the CUG helix are important for binding, and that the pyrimidine-pyrimidine mismatch is essential for binding.

Identification of RNA Sequences That Bind with High Affinity to Mbl

SELEX was performed using Mbl and a random pool of 10^{14} RNA oligonucleotides. Each RNA consisted of a 30-nucleotide randomized region flanked by shorter constant regions, creating an 80-nucleotide RNA. Twenty-five unique RNA sequences were identified (five sequences were found multiple times) after six rounds of SELEX. The sequences were classified into three groups according to the presence of enriched motifs, binding affinities to Mbl and predicted secondary structures (Table 2 and Figure 2).

An n -mer analysis comparing the expected occurrences in a sample of uniformly random RNA sequences to the 25 SELEX sequences identified several highly enriched n -mers (Table 2). The n -mers were found predominately in groups I and II and rarely in group III with AGUCU identified as the most highly enriched 5-mer. Most of the high scoring n -mers were located within two main regions of the SELEX sequences; UGUG, GUGCG and CGGUA were 5' of the AGUCU sequence in Group I while AGUC

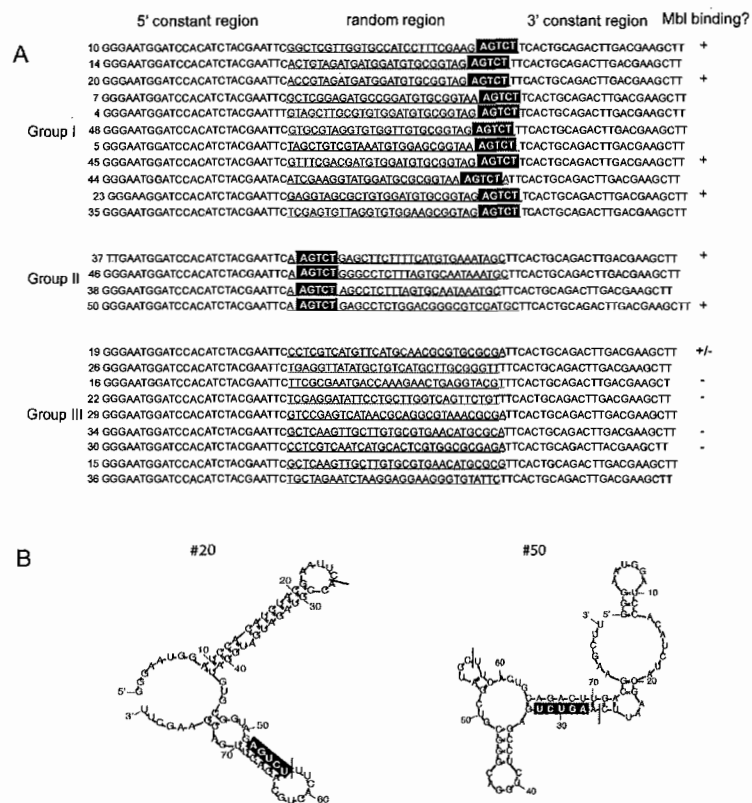


Figure 2. High-affinity RNA ligands for Mbl identified by SELEX. (A) The 25 SELEX RNA sequences were separated into three groups on the basis of whether they contained 5'-AGUCU-3' and then further categorized by the location of the 5'-AGUCU-3' sequence within the random region. The location of the 5'-AGUCU-3' sequence dictated the fold of the RNA, which also contributed to the grouping. Group I RNA sequences contain 5'-AGUCU-3' (black boxes) near the 3'-end of the random region while Group II RNA sequences contain 5'-AGUCU-3' near the 5'-end of the random region. Both Group I and II RNA sequences bound Mbl with K_d values of ≤ 50 nM. The + indicates a K_d value of < 50 nM, the +/- indicates a K_d value of ~ 300 nM and the - indicates a K_d value of > 300 nM. Group III RNA sequences did not contain 5'-AGUCU-3', did not fold into a common structure, and had K_d values of > 300 nM. (B) The 5'-AGUCU-3' sequences of both group I and II (black boxes) form a double-stranded region with 5'-AGACU-3' found in the 3'-constant region, giving rise to characteristic, predicted secondary structures within each group. RNA 20 represents the common fold of group I RNA sequences, and RNA 50 represents the common fold of group II RNA sequences. The hatch marks indicate the divisions between the constant regions and the randomized region.

was found within the AGUCU sequence. Almost all group I SELEX sequences fold into a common structure (similar to RNA 20, figure 2B) of two stems connected by a linker of one to three nucleotides. The location within the predicted secondary structures for the *n*-mers found 5' of AGUCU was generally in the linker between stem 1 and stem 2 or the base of stem 2. It is possible that these sequences were selected during SELEX to promote specific secondary structures.

Group I and II RNA sequences bound Mbl with high affinity (K_d values ranged from 0.050 to 1.0 nM, Figure 3), while group III RNA sequences bound Mbl weakly or not at all (Figure 3). Group I and II RNAs both contained the highly enriched sequence, 5'-AGUCU-3', in the random region that base pairs to a region in the 3'-constant region. In group I, this consensus sequence was located at the 3'-terminus of the randomized region, while in group II the sequence was located at the 5'-terminus of the randomized region (Figure 2A). Occasionally, the consensus sequence spanned the junction of the randomized region and the 3'-constant region (for the group I sequences). For group I sequences, the reverse complement of the consensus sequence (5'-AGUCU-3') was found six to nine nucleotides downstream from the consensus sequence. For sequences in Group II, the reverse complement of the consensus sequence was found 32 nucleotides downstream in the constant region. Regardless of the distance between the consensus sequence and its reverse complement, the minimal free energy structures (according to mfold) predicted that the 5'-AGUCU-3'/5'-AGACU-3' sequences were paired. The double-stranded consensus region (5'-AGUCU-3'/5'-AGACU-3') was found in all SELEX RNA sequences that Mbl bound tightly and corresponded to regions of common

structural aspects across Group I and II (Figure 3A,B), suggesting this motif is a core component for Mbl binding.

The Secondary Structure and Mbl Binding Site on a SELEX Group I RNA

Structure probing and footprinting were performed to determine if the secondary structure of the group I RNA sequences (predicted by mfold 30) was correct and to identify the Mbl binding site. RNA 20 was chosen for this characterization because, of the group I RNAs tested, it bound the tightest (Figures 2A and 3A). The structure probing with RNase T1, RNase V1 and RNase 1 revealed that the mfold structure prediction was consistent with the experimentally-determined structure (compare Figures 2B and 4C). The results of the RNase T1 cleavage, which is specific for single-stranded guanosines, are shown in panels A and B of Figure 4 (lane 3). Cleavage sites were identified at G1, G2, G8, G21, G39, G40, G43, G45, G47, G48, G51, G53, G63, G66, G71, G74, and G77. Lane 6 (Figure 4A,B) shows the results of RNase V1, which preferentially cleaves in double-stranded regions or regions adjacent to double-stranded RNA. The cleavages by RNase V1 were seen in stem 1A (nucleotides 11-22), stem 1B (nucleotides 31, 36, 39, 40, and 42), the base of stem 2A (nucleotides 46 and 47), and stems 2A and 2B (nucleotides between positions 53 and 78 display cleavage; however, the resolution of the gel does not allow discrete nucleotides to be identified). Different levels of cleavage by RNase V1 in stem 1A and stem 1B were observed; it is possible that the tertiary structure of RNA 20 allows some regions of stem 1 to be more accessible to cleavage than other regions.

The addition of Mbl significantly altered the cleavage pattern at three positions. All three nucleotides are guanosines, which were strongly protected from RNase T1, RNase V1, and RNase 1 activity. G53 was protected from RNase V1 by Mbl at both

Table 2: Highly Enriched *n* -mers Found in SELEX Sequences

<i>n</i> -mer	observed ^a	expected ^b	<i>P</i> value
AGUCU	15	0.740438	1.03E-61
GUGCG	11	0.625438	2.54E-39
CGGUA	10	0.717688	5.82E-28
UGUG	17	2.63369	7.81E-19
UGCG	16	2.61488	1.06E-16
AGUC	16	2.65144	2.07E-16
GUG	32	10.9052	1.33E-10
GCG	29	10.9132	3.68E-08
GUC	23	10.9851	2.56E-04

^a Observed is the total occurrences of the *n* -mer within the sample of 25 SELEX sequences.

^b Expected is the number of occurrences according to the maximum likelihood estimate in a random pool of RNA of the same size as the 25 SELEX sequences.

concentrations of Mbl, 4 μ M and 12 μ M (Figure 4B, lanes 7 and 8). G63 was protected from RNase T1 cleavage at both Mbl concentrations as well. G77 was protected from RNase T1 cleavage but only slightly protected from RNase V1 and RNase 1 by Mbl. This protection indicates Mbl binds RNA 20 within stem-loop 2 and the 3'-tail (Figure 4C).

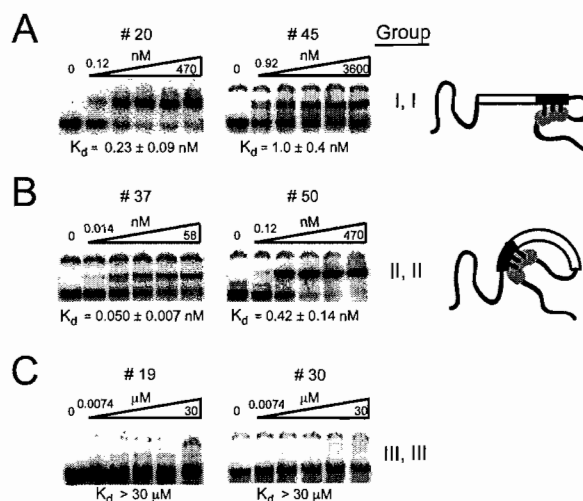


Figure 3. RNA sequences from Group I and II bind Mbl with high affinity. (A) Gel mobility shift assays with RNAs 20 and 45 from Group I are shown. A schematic representation of their predicted secondary structures is shown to the right of the gels. The box represents the random region, and the black portion represents the 5'-AGUCU-3' sequence and the gray section the 5'-AGACU-3' sequence in the 3'-constant region. Mbl concentrations for the RNA 20 gel were 0.12 nM, 0.92 nM, 7.3 nM, 58 nM, and 470 nM. Mbl concentrations for the RNA 45 gel were 0.92 nM, 7.3 nM, 58 nM, 470 nM and 3600 nM. (B) Gel mobility shift assays with RNAs 37 and 50 from Group II and a schematic representation of their predicted secondary structures. Mbl concentrations for the RNA 37 gel were 0.014 nM, 0.12 nM, 0.92 nM, 7.3 nM and 58 nM. Mbl concentrations for the RNA 50 gel were 0.12 nM, 0.92 nM, 7.3 nM, 58 nM, and 470 nM. (C) RNAs 19 and 30 from Group III bind Mbl with a K_d value >30 μ M. Mbl concentrations for RNA 19 and 30 gels were 0.0073 μ M, 0.058 μ M, 0.47 μ M, 3.6 μ M, and 30 μ M. The K_d values were determined on the basis of a minimum of three independent gel mobility shift experiments.

No protection was observed in stem 1 or the 5'-tail, which is consistent with the binding information used to group the SELEX RNA sequences. G53 is in the 5'-AGUCU-3' consensus motif, implying Mbl directly binds this particular guanosine in stem 2.

In the SELEX sequences, guanosines made up 33% of all bases while adenosine, uridine, and cytosine made up 22, 27, and 18%, respectively, of the base composition of the randomized region after six rounds of SELEX. These results suggest Mbl selected for

guanosine-rich sequences and against cytosine-rich sequences while adenosines and uridines were neutral. The selection against cytosines is surprising because Mbl and MBNL bind CUG and CCUG sequences. The bias toward guanosines in the SELEX sequences could be the result of selecting for structured, stable RNAs within a relatively short randomized region due to the fact that guanosine can bind both cytosine and uridine.

Truncation and Mutational Analysis Identifies Features of SELEX RNA 20 That Are Essential for Binding by Mbl

To determine a minimal binding site within SELEX RNA 20, binding studies were conducted with the 5'- and 3'-halves of the RNA, dividing it between nucleotides 44 and 45 to maintain stem 1 (stem-loop 1 half) and stem 2 (stem-loop 2 half). The stem-loop 2 half RNA contained an additional three guanosines at the 5'-end that was necessary for transcription using T7 RNA polymerase. No binding was observed for stem-loop 1 half RNA and weaker binding was observed for stem-loop 2 half RNA (~1000-fold lower affinity than that of full length RNA 20) (Figure 5). This is consistent with the Mbl footprinting to stem 2, but the significant loss of binding with just stem-loop 2 half compared to the full-length RNA indicates that at least a portion of stem-loop 1 half aids Mbl binding. Many of the highly enriched *n*-mers (Table 2) are located at the base of stem 1 and the linker between stem 1 and stem 2 which were disrupted in the stem-loop 1 half and stem-loop 2 half truncations. To identify the 5'-region that aids in Mbl binding, we began deleting sections of the RNA and found that removing the

majority of the eight nucleotide 5'-tail abolished Mbl binding, indicating that the 5'-tail is required for binding (truncated 5'-tail, Figure 5). When a different RNA tail (six of eight nucleotides in the 5'-tail were mutated) was added back, binding was partially restored (K_d value of 140 nM for mutated 5'-tail RNA compared to 0.23 nM for full-

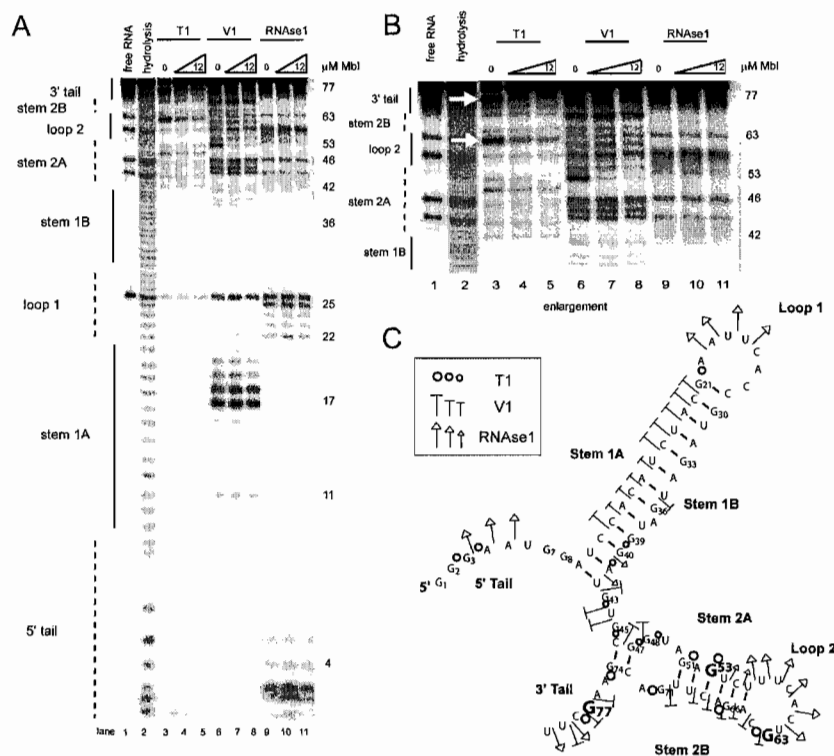


Figure 4. Secondary structure determination and identification of the Mbl binding site on RNA 20. (A) Structure probing was performed on RNA 20 with RNase T1, V1, and RNase1 in the absence and presence of Mbl. Lane 1, free RNA, shows some degradation that was considered as background cleavage. Alkaline hydrolysis of RNA 20 was performed and run in lane 2. The concentration of Mbl ranged from 4 to 12 μ M. The solid lines to the left of the gels indicate regions of double-stranded RNA; dashed lines indicate regions of single-stranded RNA. The numbers on the right of the gel correspond to residue number. (B) This is an enlargement of the upper region of the gel shown in panel A. The white arrows highlight protection of three guanosines by Mbl: G53, G63 and G77. (C) The predicted secondary structure of RNA 20 from mfold with cleavages from the three RNases (30). The three sizes of symbols indicate the intensity of the cleavage. Pins with circles indicate RNase T1 cleavages, T symbols indicate RNA V1 cleavages and triangles indicate RNase 1 cleavages. The large, bolded guanosines indicate bases protected by Mbl.

length RNA 20, Figure 5), indicating that both the presence of the tail and its sequence are important for Mbl binding.

Not surprisingly, deleting the six nucleotide 3'-tail abolished Mbl binding; one of the guanosines that Mbl protected is within this region (Figure 5). To determine if the 5'- and 3'-tails from the SELEX RNA 20 could generally enhance Mbl binding, we added these tails to the (CUG)₄ RNA and tested Mbl binding. Apparently, the tails are specific to the SELEX RNA because the addition of the tails to the CUG repeats did not increase the level of binding (data not shown). Next we reduced the length of stem 1, maintaining the first four base pairs at the base of this stem and capping it with an ultrastable UUCG sequence. Tight binding was maintained with no loss in affinity compared to the full-length RNA 20 (shortened stem-loop 1 RNA, Figure 5). These binding studies indicate that both the 3' and 5' tails are required for binding by Mbl, but that a minimal stem can replace the extended stem 1 and maintain high-affinity binding. This suggests that the 5'-tail, 3'-tail and stem 1 are important for the overall structure of the RNA and Mbl binding.

To test the importance of the guanosines protected by Mbl, each was mutated individually to adenosine. Each G to A mutation severely inhibited or eliminated Mbl binding (Figure 6). G63 and G77 were in single-stranded regions that, when mutated, did not affect the predicted structure. G53 and C68 were both mutated to form a new U-A base pair to maintain the secondary structure of RNA 20; however mfold predicted a change in structure. To determine if nucleotides near these guanosines are important for binding, we introduced two additional single mutations. C59 was mutated from a cytosine

to adenosine (data not shown) and A75 from an adenosine to a guanosine (Figure 6, bottom right panel). Neither mutation affected Mbl binding. Our results suggest that G53, G63, and G77 are positioned within the secondary structure in RNA 20 in such a way as to facilitate high-affinity binding by Mbl.

Although the G53A mutation suggested this specific nucleotide was important for Mbl binding, we wanted to determine if Mbl would still bind with high affinity if the 5'-C₅₁CACAC-3' / 5'-GUGUGG-3' to maintain the structure but not the sequence (2A/2B-6, Figure 6). The affinity did not decrease compared to that of RNA 20, indicating that the stem structure was selected for and not necessarily the sequence, although Mbl could be contacting a different guanosine in this new sequence. The lack of obvious pyrimidine-pyrimidine mismatches in SELEX RNA 20 was initially a puzzling result. We hypothesized that loop 2 might contain one or more pyrimidine-pyrimidine mismatches and function in a manner analogous to that of the mismatches found in the MBNL1 binding sites. To determine if the U57-C64 mismatch was important for Mbl binding, U57 was mutated to a guanosine (U57G) to form a G-C base pair (G57-C64). The level of binding to U57G was reduced 30-fold compared to that of RNA 20 suggesting that this mismatch plays a role in Mbl binding (Figure 6).

Mbl Recognizes an MBNL1 Binding Site in the cTNT Pre-mRNA with High Affinity

MBNL1 binds a stem-loop in the cTNT pre-mRNA and regulates the splicing of the adjacent exon through this interaction (17, 22). To determine if Mbl binds this

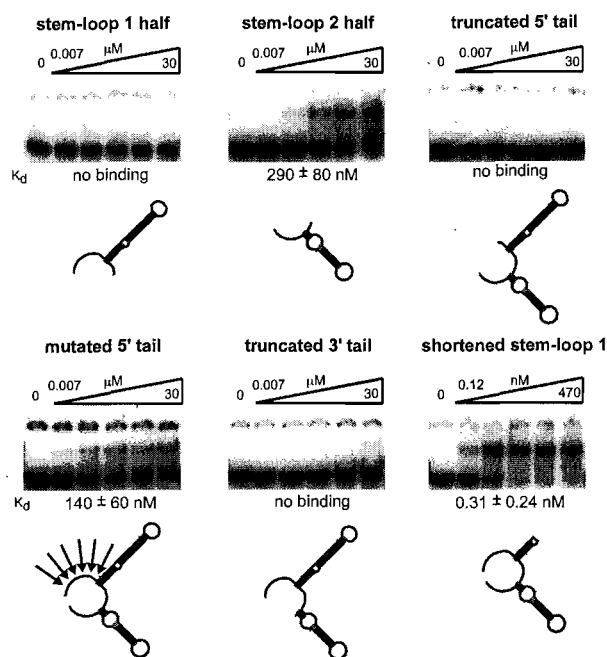


Figure 5. Mbl requires 5'- and 3'-single-stranded tails but not full-length stem-loop 1 for high-affinity RNA binding. Various truncations and mutations of RNA 20 were bound to Mbl and analyzed using gel mobility shift assays. Stem-loop 1 half and stem-loop 2 half RNAs were cleaved between residues 44 and 45 with three guanosines added 5' to G45 in stem-loop 2 half RNA. Truncated 5' tail RNA was deleted up to residue 6, and U6 was replaced with a guanosine. Six residues were changed in mutated 5' tail RNA as indicated by the arrows. The residues were mutated from 5'-G₃AAUGG-3' to 5'-A₃CUCCA-3'. Truncated 3'-tail RNA was deleted up to residue 74. The shortened stem-loop 1 RNA was truncated up to the 12-39 base pair and replaced with a stable UUCG cap. Mbl concentrations were 0.0073 μ M, 0.059 μ M, 0.47 μ M, 3.8 μ M, and 30 μ M for stem-loop 1 half, stem-loop 2 half, truncated 5'-tail, mutated 5'-tail and truncated 3'-tail RNAs. Mbl concentrations for shortened stem-loop 1 were 0.12 nM, 0.93 nM, 7.3 nM, 59 nM and 470 nM. The secondary structures shown below each gel mark the location of the truncation or mutation. The K_d values were determined on the basis of a minimum of three independent gel mobility shift experiments.

MBNL1 pre-mRNA target, we performed binding studies with this cTNT RNA. A 50-nucleotide region from within intron 4 of the cTNT pre-mRNA was found to form a stem-loop that contains several pyrimidine-pyrimidine mismatches bordered by G-C and

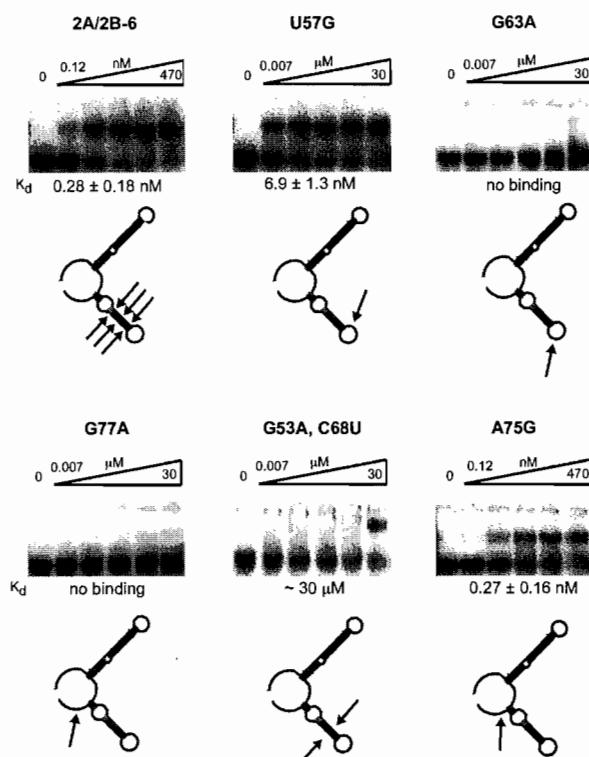


Figure 6. Mbl binds specific structural elements and sequences. In 2A/2B-6 mutant RNA, six base pairs in stem 2 (containing the AGUCU consensus region) were mutated to maintain the structure but change the sequence. In U57G mutant RNA, a potential U-C pyrimidine-pyrimidine mismatch was mutated to a G-C base pair. The three guanosines (G53, G63, and G77) that exhibited strong protection in the footprinting assay were individually mutated to adenosines. The arrows indicate the residue or group of residues that was mutated. A75G mutant RNA was tested for binding to Mbl as a control. Mbl concentrations for U57G, G63A, G77A, and G53A/C68U mutant gels were 0.0073 μM , 0.059 μM , .47 μM , 3.8 μM , and 30 μM . Mbl concentrations for 2A/2B-6 and A75G mutant gels were 0.12 nM, 0.92 nM, 7.3 nM, 59 nM, and 470 nM. The K_d values were determined on the basis of a minimum of three independent gel mobility shift experiments.

G-U base pairs (Figure 7A). Mbl binds this RNA with a K_d value of 0.67 nM, an affinity similar to that of the Mbl SELEX RNA sequences (Figure 7B). Mutating four of the guanosines to pyrimidines eliminates binding at the Mbl concentrations that were tested (Figure 7B). These mutations reduce the level of MBNL1 binding by 100-fold (22),

while Mbl appears to display even more specificity (>500-fold). This result shows that Mbl is capable of recognizing human MBNL1 sites and suggests Mbl may recognize similar RNA motifs during splicing regulation in *Drosophila*.

DISCUSSION

This first characterization of the RNA binding properties of Mbl shows that the protein is a very specific and high-affinity RNA binding protein. Mbl binds CUG repeats as expected (Figure 1) on the basis of the findings that, like MBNL1, Mbl colocalizes with CUG repeats in *Drosophila* (24). The high specificity of Mbl for the CUG repeat RNA is surprising. We previously found that MBNL1 tolerated many different mutations to the CUG repeats (22), while Mbl only tolerated replacement of the U-U mismatch with other pyrimidine-pyrimidine mismatches (Figure 1). This indicates that Mbl is a more specific RNA binding protein than MBNL1.

Using SELEX, we identified RNAs that bound Mbl with K_d values ranging from the picomolar to the low nanomolar range. All of these RNAs contain the consensus motif of 5'-AGUCU-3' (Figure 2). The footprinting studies suggest Mbl interacts directly with the guanosine in this sequence, and mutating this nucleotide eliminated Mbl binding (Figures 4 and 6); however, the importance of this sequence is partly or fully due to the fact that it base pairs with a region (5'-AGACU-3') in the constant region. This was demonstrated by changing the sequence of the stem 2 and finding that the affinity of this RNA for Mbl was the same as that for RNA 20 (Figure 6). It is possible that Mbl

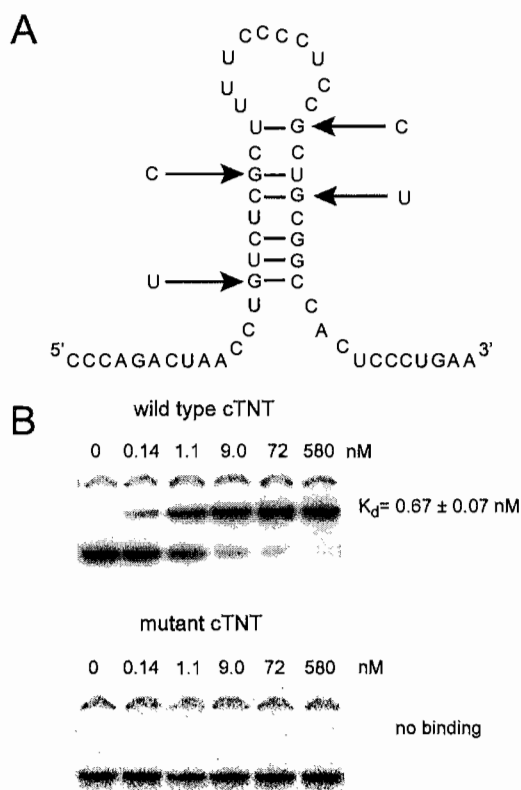


Figure 7. Mbl binds a human MBNL1 binding site in cTNT pre-mRNA with high affinity and specificity. (A) Secondary structure of the cTNT RNA based on structure probing by Warf and Berglund (22). The arrows indicate the four guanosine to pyrimidine mutations (mutant cTNT) that abolish the stem structure. (B) Gel mobility shift assays of the 50 nucleotide region of cTNT RNA as the wild-type (top gel) or mutated (bottom gel). The K_d values were determined on the basis of a minimum of three independent gel mobility shift experiments.

interacts with a different guanosine in this new sequence, although it would be in a different location in the helix. Interestingly, the 5'-AGUCU-3' sequence and the putative MBNL1 consensus sequence (YGCUUY) (17) are similar in that both contain a guanosine followed by a run of pyrimidines. The major difference between the motifs is the presence of the adenosine in the SELEX motif. Also importantly, the SELEX 5'-AGUCU-3' motif is involved in all Watson-Crick base pairs, while the YGCUUY motifs in cTNT, when base paired, contain both Watson-Crick base pairs and pyrimidine-

pyrimidine mismatches (22). However, in the SELEX RNA 20, a pyrimidine-pyrimidine mismatch 3' to the 5'-AGUCU-3' motif (U57-C64) appears to be important for binding because if the mismatch is replaced with a G-C base pair, the level of binding is reduced 30-fold (Figures 6 and 8).

Although stem-loop 2 is clearly important for Mbl binding, the 5'- and 3'-single-stranded tails are also important for binding (Figure 5). These tails could be involved in tertiary interactions with other regions of the RNA that are difficult to monitor with gel mobility shift assays and footprinting. It is possible that the G-rich regions of the 5'-tail interact with loop 1 or the 3'-tail and aid in loading of Mbl onto the RNA or proper folding of the RNA, and upon removal, these interactions are lost. Apparently, the tails are specific to the SELEX RNA because the addition of the tails to the CUG repeats did not enhance binding (data not shown). The 50-nucleotide cTNT RNA also contains 5'- and 3'-tails that could be aiding in the formation of the correct RNA structure recognized by Mbl.

A comparison of the SELEX RNA 20 with the human cTNT RNA reveals a site that the two RNAs may share. As shown in Figure 8, the loop and upper portion of stem-loop 2 of RNA 20 and the upper portion of the cTNT RNA are quite similar in sequence and structure, assuming RNA 20 and cTNT adopt similar three dimensional structures. Both of these RNAs contain two pyrimidine-pyrimidine mismatches (positions 1 and 3) with a potential U-G wobble base pair between them (position 2). Below the position 3 mismatches are two additional base pairs, with position 5 being a C-G base pair in both RNA 20 and the cTNT structures. MBNL1 has also been shown to bind a structured

RNA site within the mouse Tnnt3 pre-mRNA and regulate splicing through this site

(23). Mutational analysis revealed the recognition site for MBNLI as being a stem-loop

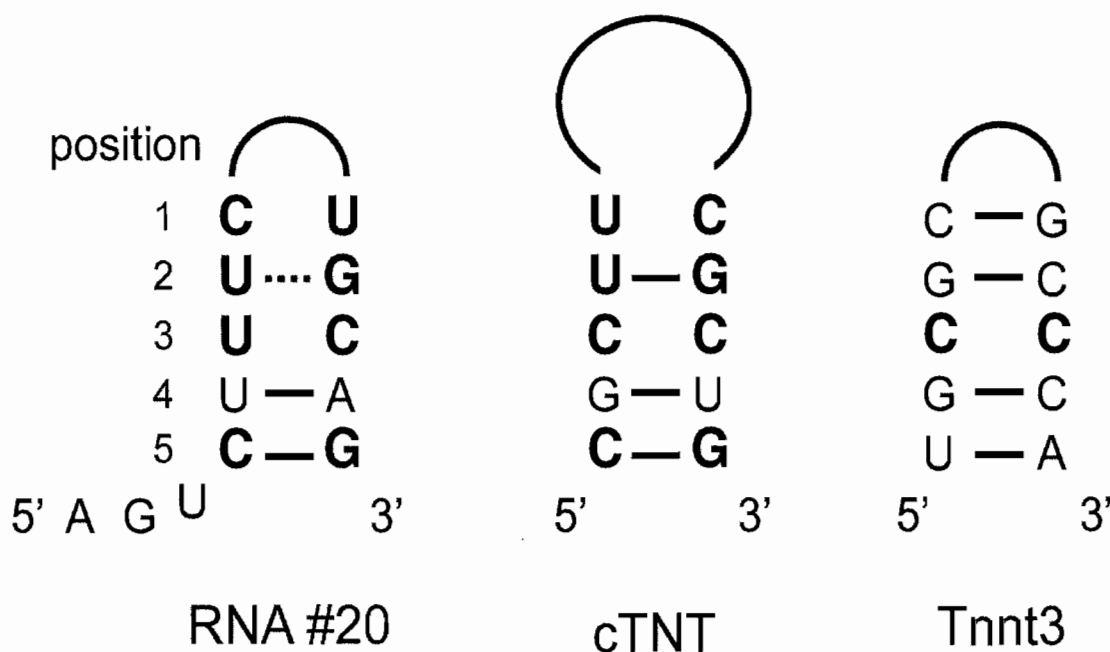


Figure 8. Comparison of three stem-loop binding sites for the muscleblind family of proteins. From the SELEX RNA 20, a portion of stem-loop 2 is shown beginning with the highly enriched 5'-AGUCU-3' sequence (RNA 20). The upper portion of the cTNT stem-loop structure (cTNT) is shown. The upper portion of the mouse Tnnt3 stem-loop structure (Tnnt3) is shown (23). The position and numbers refer to base pairs or mismatches in the structures to the right. The bold nucleotides highlight the sequence similarity between the structures. The dashed line indicates a potential G-U wobble base pair in RNA 20.

structure within a larger RNA structure. This stem-loop also contains a pyrimidine-pyrimidine mismatch but lacks the sequence similarity to SELEX RNA 20 and the cTNT RNA (Figure 8). The comparison of these three RNAs indicates Mbl and MBNL1 bind structured RNA stem-loops and that there is significant flexibility in the composition of

the sequences, which makes it challenging to identify a consensus sequence within the context of a stem or stem-loop.

We speculate that stem-loop structures containing correctly positioned guanosines and pyrimidine-pyrimidine mismatches will be found as the binding sites in Mbl's endogenous targets such as *tun* and α -*actinin* pre-mRNA. These transcripts were recently shown to be mis-spliced in Mbl mutant flies (3). Vicente and colleagues (4) proposed a CUG-rich region downstream of a regulated exon in the α -*actinin* pre-mRNA as a potential binding site for Mbl. This region can potentially fold into a stem-loop containing pyrimidine-pyrimidine mismatches (4). In preliminary studies, we have found a 115-nucleotide RNA, containing this CUG-rich region, binds Mbl with weaker affinity compared to CUG repeats, SELEX RNAs, and the cTNT RNA (data not shown). Although our study has improved our understanding of how Mbl binds RNA, it is clear that additional investigations of Mbl- and MBNL-RNA interactions are required to reach a point at which binding sites can be predicted.

Zinc finger proteins, other than muscleblind proteins, are capable of recognizing complex RNA structures. There are several classes of zinc finger proteins, including C₂H₂, CCCH, and CCHC, that bind RNA (31). Some zinc finger proteins are promiscuous and bind double-stranded DNA and single- and double-stranded RNA (32), while others, like Mbl, bind specifically to double-stranded RNA. An example of a zinc finger protein that binds RNA is TIS11d. TIS11d binds single-stranded RNA with a 5'-UAUU-3' motif by making specific stacking interactions between aromatic side chains and the RNA bases (32). This mode of binding by TIS11d is probably different than

Mbl's mode of binding due to its double-stranded RNA targets. However, the HIV nucleocapsid protein (NC) contains two zinc fingers of the CX₂CX₄HX₄C type and has been found to recognize a variety of stem-loop structures from the HIV RNA (33, 34). There are three exposed guanosines in the loop portion of stem-loop 3 from the HIV ψ -RNA recognition element that directly interact with multiple amino acids positioned by the zinc fingers of the HIV NC protein (35). The zinc fingers of Mbl may be performing a role similar to the role of the HIV NC zinc fingers in that the amino acids are positioned for specific recognition of the guanosines in the SELEX RNA sequences and endogenous RNA targets of Mbl.

The mutagenesis with the CUG repeats and the cTNT RNA revealed that Mbl is a specific RNA binding protein and is apparently more specific than MBNL1 since it displays a greater degree of specificity for both the CUG repeats and the cTNT RNA substrates. The primary difference between these two muscleblind proteins is that Mbl contains only two zinc fingers while MBNL1 contains four zinc fingers. One possibility is that the two additional zinc fingers of MBNL1 are nonspecific RNA binding domains, resulting in a less discriminate RNA binding protein. The third and fourth zinc fingers in MBNL1 appear to be related to the first two of Mbl and MBNL1 in sequence and could be the result of a duplication event (36). In future studies, it will be interesting to determine if muscleblind proteins with two zinc fingers are generally more specific RNA binding proteins compared to muscleblind proteins with four zinc fingers. This minimal Mbl protein construct should prove useful for structural studies and in further dissecting

the RNA binding abilities of this important protein as well as the identification of its binding sites in *Drosophila* RNAs.

CHAPTER III

DOPED SELEX WITH MBNL1 REVEALS A MOTIF USED TO IDENTIFY NOVEL MBNL1 BINDING SITES

This chapter is currently being prepared as a manuscript with the help of Devika P. Gates and Rodger B. Voelker. I have performed the majority of the experiments thus far, with the help of Devika P. Gates. Rodger B. Voelker performed the bioinformatics (highly enriched k -mer analysis and positional selection analysis). J. Andrew Berglund aided in experimental design and will be last author on the manuscript.

INTRODUCTION

Alternative splicing is essential for creating a diverse, functional proteome. Approximately 74% of human genes are predicted to be alternatively spliced (1). The proteins accountable for regulating alternative splicing, alternative splicing regulators, are active in tissue- and developmentally-specific fashions. A small group of alternative splicing regulators have been identified, they include NOVA, the CELF family, the FOX family and the muscleblind-like (MBNL) family of proteins. NOVA is a neuron-specific alternative splicing regulator that binds to YCAY clusters in or near regulated exons and promotes exon inclusion or exclusion (2, 3). The FOX family of proteins are expressed in brain, heart and skeletal muscle and bind the UGCAUG RNA motif. The location of

binding, by the FOX family of proteins, CUG-BP (in the CELF family) and NOVA, relative to the regulated exon determines whether the exon is excluded or included in the mRNA(3, 4). Based on hundreds of predicted FOX-1/2 binding sites, FOX-1/2 binding upstream of the regulated exon promotes exclusion while binding downstream promotes inclusion (4). CUG-BP regulates many of the same transcripts that MBNL1 does, however, it is in an antagonistic fashion. For example, MBNL1 binds upstream of exon 5 in the cardiac troponin T (cTNT) pre-mRNA and promotes exon exclusion while CUG-BP binds downstream of exon 5 and promotes inclusion (5, 6).

The muscleblind family of proteins are highly conserved across metazoans, especially in the zinc finger domains, and bind RNA in a specific fashion (6-8). Muscleblind is implicated in myotonic dystrophy as it becomes sequestered to aberrant RNA transcripts in the nucleus of patients. In myotonic dystrophy type 1 (DM1), MBNL1 becomes sequestered to CUG repeats and in myotonic dystrophy type 2 (DM2), to CCUG repeats (9, 10). This sequestration is sufficient to cause a deficiency of MBNL1 binding to pre-mRNA targets. This lack of binding by MBNL1 causes important developmentally-specific transcripts to become mis-spliced and leads to symptoms of DM. For example, insulin receptor (IR) and chloride ion channel (CIC-1) pre-mRNAs are mis-spliced in DM1 patients leading to inappropriate expression of fetal isoforms or degradation of the transcript (11-13). The lack of appropriate IR and CIC-1 splice isoforms in DM1 patients is thought to lead to the symptoms of insulin resistance and myotonia, respectively.

CUG repeats form stem-loop structures with pyrimidine-pyrimidine mismatches (14, 15). MBNL1 has a preference for weak stem-loops containing G-C base pairs, G-U wobble base pairs and pyrimidine-pyrimidine mismatches (6, 7). This preference is apparent because MBNL1 binds CUG repeats, CCUG repeats and a MBNL1 binding site in cTNT pre-mRNA, all of which are capable of forming stem-loops (6, 15). MBNL1's binding site in cTNT is a 32 base region (32mer) about 35 bases upstream of exon 5 (6, 16). MBNL1 promotes the exclusion of exon 5 which produces a splice product found in adult tissue. However, in DM1, MBNL1 is sequestered and causes cTNT exon 5 to be included aberrantly and produces a splice product normally found in fetal tissue (17). Several other pre-mRNA transcripts are regulated by MBNL1, including SERCA1, MBNL1 and IR (for review see(10, 18)). However, the only known human MBNL1 binding site is the one in cTNT. The identification of additional MBNL1 RNA binding sites would allow for a deeper understanding of MBNL1's RNA binding specificity. This will help identify pre-mRNA targets regulated by MBNL1 and to predict MBNL1 binding sites within these targets.

To identify the important residues in the cTNT binding site, doped SELEX was performed on this RNA. The doping incorporated the endogenous cTNT binding site sequence at a rate of 51% for each of the 32 residue positions. After 5 rounds, the majority of SELEX sequences tested bind MBNL1 with high affinity, similar to MBNL1's affinity for the cTNT site. The SELEX sequences revealed selection throughout the doped region. The most pronounced selection occurred on the left side of the cTNT binding site stem; specifically, the selection lead to an over-enrichment of

UGCUU 5-mers in the 82 sequences. We used this motif to identify three potential MBNL1 sites in MBNL1 and SERCA1 pre-mRNAs. The potential MBNL1 and SERCA1 sites are within 200 bases of MBNL1-regulated exons and MBNL1 binds these 40 nucleotide sites with high affinity *in vitro*.

MATERIALS AND METHODS

SELEX

The MBNL1 construct used is truncated and includes residues 1-260 and an N-terminal GST tag. This construct, termed MBNL1 throughout the paper, binds RNA as tightly as the full length version (6). For purification of MBNL1 see Warf and Berglund '07. The DNA template used for the SELEX (Systematic Evolution of Ligands by Exponential Enrichment) (19) experiment was the following 81-mer oligonucleotide: 5'-GGGAATGGATCCACATCTACGAATTC(**CCTGTCTCGCTTTTCCCCTCCGCTGCGGCCAC**)AAGACTCGATACGTGACGAACCT -3'. The oligonucleotide depicted here contains the 32 nucleotide MBNL1 binding site in cTNT (in bold) flanked by two constant regions. To create the SELEX pool 0, the nucleotides in parenthesis were randomized; 51% of the time the original nucleotide was kept at each position and the other 49% of the time each position was varied to incorporate an equal mix of the other three nucleotides. The forward primer for the SELEX was 5'-GATAATACGACTCACTATAGGGAATGGATCCACATCTACGA-3' and the reverse was 5'-AGGTTCGTCACGTATCGAGTCTT -3'. The initial PCR reactions were amplified by 8 rounds of PCR and calculated to approximately 1×10^{14} molecules. The

transcription reactions were done under the following conditions: 500 ng/ μ L template, 40 mM Tris pH 7.9, 26 mM MgCl₂, 2 mM spermidine, 10 mM NaCl, 5 mM of each nucleotide, 0.1 μ g/100 μ L yeast pyrophosphatase, \sim 2 mg/mL T7 polymerase, and 40 mM DTT . The transcription reactions were performed with trace amounts of [α -P³²] CTP for monitoring purposes. After transcription, reactions were treated with DNase for 1 hr at 37°C, gel purified, ethanol precipitated and run over a Bio-Spin-6 size exclusion column (Bio-Rad). For each round, the appropriate concentration of MBNL1 was bound to 20 μ L of glutathione-agarose beads at 4°C for 15 min, then washed 1 time with 200 μ L SELEX binding buffer (100 mM NaCl, 20 mM Tris pH 7.5, 5 mM MgCl₂, 0.02% triton x-100 and 5 mM DTT added the day of use). All washes were done in a similar manner. Round 6 was performed in the same way but with a higher NaCl concentration (150 mM NaCl instead of 100 mM NaCl, the other reagents remained constant). The RNA was heated in SELEX binding buffer at 95°C for 3 min and placed immediately on ice for 10 min. The protein used throughout the paper was a truncated version of MBNL1 with amino acids 1-260 and an N-terminal GST tag as in Warf and Berglund (6). The MBNL1-bound beads were incubated with the RNA at 25°C for 20 min. The beads were washed 1 time with 200 μ L of SELEX binding buffer at 25°C and RNA was released and collected from the beads by a phenol-chloroform extraction and subsequent ethanol precipitation. The collected RNA was reverse transcribed using AMV reverse transcriptase, AMV buffer, 8.8 μ M reverse primer, and 2/3 of the RNA isolated after the round for 1 hr at 42°C. After the reverse transcription was complete, the DNA was amplified by 11 PCR cycles. The DNA was transcribed and the SELEX cycle repeated seven times with the

concentration of RNA and MBNL varying according to Table 1. After the rounds were completed, individual clones were isolated by TOPO cloning (Invitrogen) and sequenced.

Table 1: Concentration of RNA and Protein used for rounds of SELEX

round	RNA (μM)	Protein (μM)
1	20	20
2	9.8	4.9
3	8	2
4	2	0.33
5	1	0.13
6	0.5	0.05
7	0.25	0.025

Labeling of RNA Oligomers for Gel Mobility Shift Assays

RNA oligomers were transcribed from templates using T7 RNA polymerase and [$\alpha\text{-P}^{32}$] ATP. RNA oligomers were kinased using Polynucleotide Kinase and [$\gamma\text{-P}^{32}$] ATP.

Gel Mobility Shift Assay

Gel mobility shift assays done as in Warf and Berglund 2008, with modifications as follows (6). RNA (8 μL) was snap annealed in (75 mM NaCl, 5 mM MgCl₂, 15 mM Tris pH 7.5, 0.25 mM β -ME), mixed with 2 μL of protein for final reaction conditions of 175 mM NaCl, 5 mM MgCl₂, 20 mM Tris pH 7.5, 1.25 mM β -ME, 10% glycerol, 2 mg/mL BSA, 0.1 mg/mL heparin and trace amounts of bromophenol blue. The binding reactions were incubated for 10-25 minutes at room temperature and 2-4 μL were loaded

onto a 6% acrylamide gel (6% 37.5:1 acrylamide:bisacrylamide, 0.5X TB). Gels were run for 35 min at 170 V.

Positional Selection

The selective pressures on individual positions was determined by calculating the Z-score for the observed nucleotide frequencies in each position versus the expected frequencies if there was no positional selective pressures. For each position the expected mean was calculated using the maximum likelihood estimate for the mean (based upon the theoretical probabilities for each position) and the standard deviation was calculated using the standard deviation for the binomial distribution, $\sqrt{np(1-p)}$ where n is the number of observations (which in this case equals 82), and p is the probability of the nucleotide occurring in that position. These means and standard deviations were used to calculate a Z-score using Equation 1. This equation is allowed since the expected numbers are always greater than 10, in this sized sample, which means that the underlying distribution approaches the Gaussian distribution.

K-mer Enrichment

All *k*-mers from 3 to 5 nt were counted in the final 82 SELEX sequences, including the random sequence plus 5 nt of the constant flanking sequence. The inclusion of the flank allows one to analyze *k*-mers that overlap with the edges of the constant sequence. To calculate a Z-score for the observed vs. expected occurrences we used Monte-Carlo simulation to estimate the mean (μ) and standard deviation (σ) for the occurrence of all *k*-mers in 1000 independent populations of 82 sequences constructed

using the same random biases used to synthesize the oligos used as input in the SELEX experiment. The Z-score (z) for each k -mer was calculated according to equation 1 where μ is the mean observed over the random samples, σ is the standard deviation for the k -mer occurrence within the random samples, and x is the observed occurrence within the population derived from SELEX.

$$z = \frac{x - \mu}{\sigma} \quad (\text{Equation 1})$$

RESULTS

MBNL1 Binding Site in cTNT Intron Provides a Template for a Doped RNA SELEX

The 32 nucleotide MBNL1 binding site upstream of exon 5 in cTNT pre-mRNA was used as the template for generating SELEX pool 0 (6, 16). Each position contains the wildtype residue 51% of the time and 48% of the time contains the other 3 residues equally (16% each). For example, the first position in the 32 nucleotide region is C, so in 51% of the RNA sequences, it will remain a C, while the other 48% of the time A, G or U is incorporated at that position. This allows MBNL1 to sample other possible residues at each position with a bias toward the endogenous residue. The 32 endogenous nucleotides are flanked on each side by constant regions, 26 nucleotides on the 5' side and 23 nucleotides on the 3' side for priming and transcription purposes (Figure 1A). The MBNL1 construct used is truncated and includes residues 1-260 and an N-terminal GST tag. This construct, termed MBNL1 throughout the chapter, binds RNA as tightly as the full length version (6).

The initial rounds of selection were carried out under standard conditions (100 mM NaCl, 5 mM MgCl₂, 0.02% triton x-100, 1 mM DTT, 20 mM Tris pH 7.5) with a slightly lower sodium chloride concentration in order to retain roughly 5-10% RNA bound after each round. Binding affinity of MBNL1 was weak for pool zero with a K_d

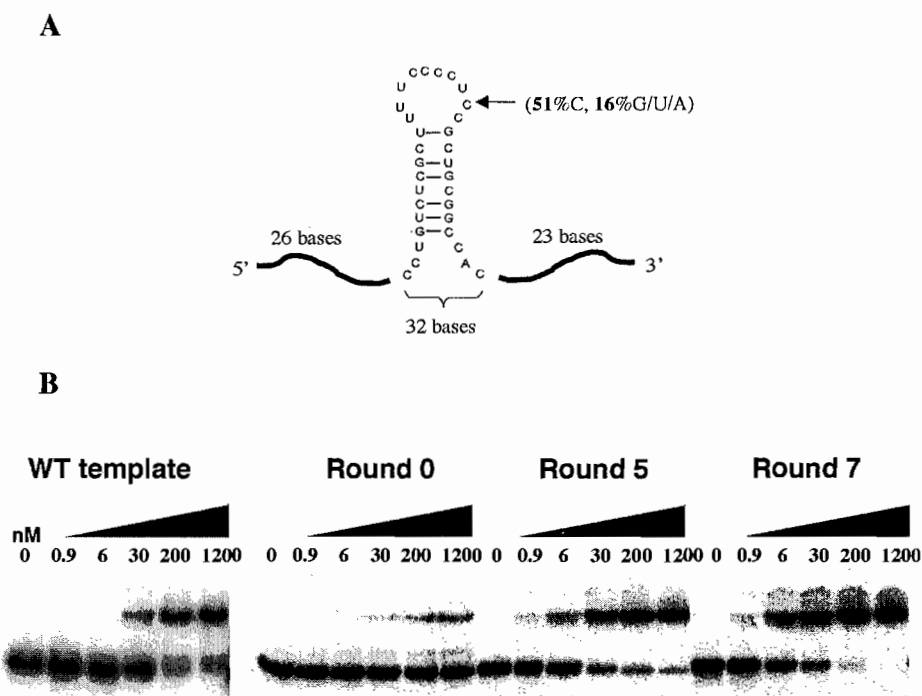


Figure 1. Round 5 and round 7 of SELEX RNA oligomers bind significantly tighter than round 0 RNA oligomers. (A) The template used for the SELEX included MBNL1's cTNT intronic binding site (the 32mer stem-loop) in the center of the template and flanking constant regions. The central 32 bases were doped at each position with 51% of the endogenous base and 49% of the other three nucleotides equally (16% each). (B) Gel mobility shift assays showing MBNL1 binding to the endogenous cTNT template, round 0, round 5 and round 7 RNA sequences. The concentration of MBNL1 is shown at the top of each lane in nM.

greater than 1.2 μ M (Figure 1B). The binding buffer was changed at round seven to higher salt (from 100 mM NaCl to 150 mM NaCl) to make the conditions more stringent. Binding affinities increased after each round until round seven (Figure 1B). 96 sequences

were obtained for round zero, 82 sequences for round five and 76 sequences for round seven in order to analyze the evolution of sequence selection over the subsequent rounds. Pool zero was sequenced to insure that there was no bias in the initial pool other than the designed doping. A usage frequency analysis at each position revealed that pool zero was very close to the doping design (data not shown). Pool five RNA sequences were analyzed more in depth as they showed promise for providing a consensus motif. Analysis of pool seven is discussed in the appendix.

After five rounds, the approximate affinity of the pool was similar to the endogenous cTNT binding site template (Figure 1B). Eighty-two clones from round five were isolated and sequenced. The sequences were grouped according to Jalview dendrogram analysis based on percent identity (20). Such analysis showed only one obvious outlier, with the majority containing relatively close sequence similarities (Figure 2). From the dendrogram, blocks of high sequence similarity are apparent. There are some regions of moderate sequence identity in the 3' region, however, the region of highest identity similarity is the 5' half of the sequences (Figure 2B). There were no duplicates in this round indicating that over-selection did not take place. Also, the sequence most closely related to the endogenous cTNT template (DO3) contained six bases that differed (Figure 4). This was not surprising, as pool zero started out with a 1 in 2,279,098,388 chance that a sequence would be identical to the endogenous template.

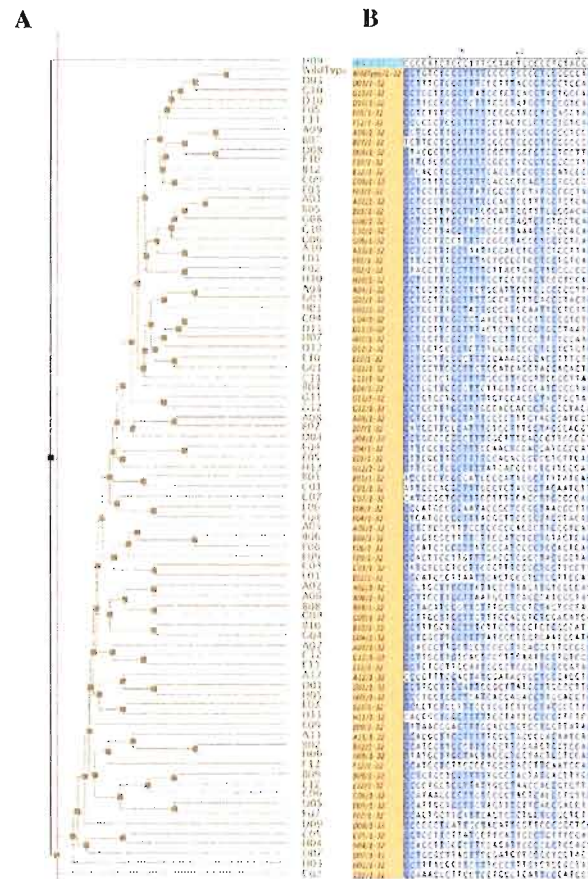


Figure 2. The 82 SELEX sequences from round 5 fell into one main group. (A) Dendrogram showing the 32 nucleotide randomized region of the SELEX sequences aligned based on sequence identity. The wild type cTNT site is included. (B) The sequences ordered based on the dendrogram in (A), with darker blue representing higher sequence identity.

The Majority of SELEX Sequences Bind MBN1 With High Affinity and Provide a Consensus Motif of 5'-YGCUU-3'

To understand the selection for and against certain nucleotides at specific locations along the stem, a Z-score analysis was performed at each residue position, taking into account the doping bias (Figure 3A). A positive Z-score indicates a positive

selection for that particular nucleotide, while a negative Z-score indicates a negative selection for that nucleotide. Most positions have a positive selection for T (U) and negative selection against A, which follows with the fact that the cTNT binding site has very few adenosines and is uracil rich. This indicates that MBNL1 prefers uracil over adenosine in its binding sites. The majority of positions with selection are located in the 5' half of the sequence, which can also be thought of as the left side of the stem (Figure 3B). The first two positions (C_1 and C_2) show strong selection to remain Cs. From analysis of the predicted folds of many of the SELEX sequences, they often form base pairs with the constant region, which is most likely why they were selected to remain as Cs, not necessarily because MBNL1 selected them as part its recognition site. However, they may play a role in setting up the proper structure for MBNL1 binding. There is very little selection throughout the loop region (U_{13} - C_{21}), presumably because MBNL1 binds regions of structured RNA. It is intriguing that there is also less selection in the 3' half of the sequence. G_{27} was selected to become a U, which would no longer allow it to base pair with C_6 . This is consistent with the finding by Warf and Berglund that MBNL1 binds with higher affinity to RNAs that contain a mix of mismatches and base pairs, not complete base pairing (6). With such positive selection in the left side of the cTNT stem and less selection in the right side of the stem, it is possible that MBNL1 prefers very little structure.

Reasoning that the four bases from the endogenous site (G_9 - U_{12}) that were selected to remain the same were probably important for MBNL1's binding, they were included in the consensus sequence of 5'-CCNNCUNUG₉CUU-3'. Using a Z-score cut

off of 4 to define positive selection, several bases 5' to G₉ were also selected for, and are subsequently included in the motif (Figure 3A). U₅, C₆ and U₈ were all selected to switch from the pyrimidine in the endogenous sequence to the other pyrimidine. For example C₈ was selected to become a U and U₅ selected to become a C. It is possible that MBNL1 prefers a short run of pyrimidines 5' of the 5'-G₉CUU-3'. However, MBNL1 bound weakly to a 40mer containing the aforementioned motif flanked on the 5' end by 12 uracils and 12 cytosines on the 3' end (Figure 3D). This indicates that some structure is required for binding by MBNL1. Assuming that MBNL1 interacts with approximately five bases when binding RNA, like other RNA binding proteins such as Nova and FOX1, a consensus motif of 5'-YGCUU-3' is useful (21, 22).

In order to determine if 5'-YGCUU-3' was over-enriched in the 82 round 5 SELEX sequences in comparison to pool 0, an analysis of different 3-mers, 4-mers and 5-mers was performed (materials and methods section, *k*-mer analysis). The analysis included the number of expected occurrences for each *k*-mer (based on 1000 randomly generated sets of 82 sequences) and the observed number of occurrences in the 82 SELEX sequences, taking into account the doping bias (Table 2). The Z-scores range from 14.8 to 33.1, indicating high statistical significance. The UGCUU 5-mer is clearly over-enriched with 61 occurrences in the SELEX sequences compared to only 3.1 expected occurrences. In addition, many shorter permutations of UGCUU, such as GCU and GCUU, were highly enriched. This is strong evidence that 5'-YGCUU-3' was

selected for by MBNL1, with a preference for a U in the Y position. This is interesting as the endogenous cTNT binding site contains 5'-YGCUU-3' with a C, not a U, in the Y position. This further supports the argument that MBNL1 most likely does not have a strong preference for either pyrimidine in that position of the motif.

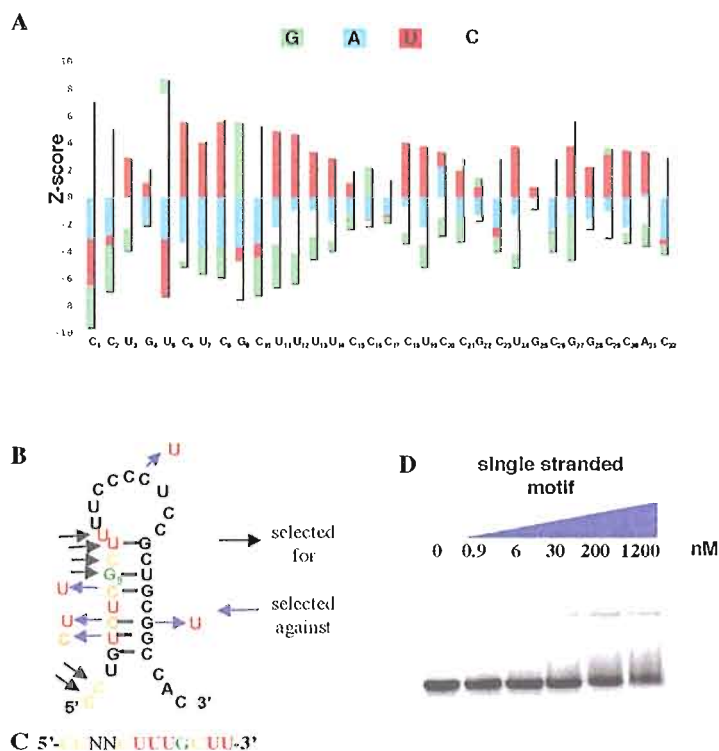


Figure 3. Regions of positive and negative selection reveal a consensus binding sequence on the left region of the cTNT stem. (A) Z-score analysis was performed at each position within the doped region. A negative z-score indicates negative selection and a positive z-score indicates positive selection. The bars representing different nucleotides are stacked on top of each other. The endogenous cTNT sequence is shown below. (B) The stem-loop of the cTNT binding site is shown with colors matching to the consensus sequence in (C). The arrows pointing away from the residue indicate that negative selection occurred for that base to change from its original residue to the residue at the head of the arrow. (C) The original consensus sequence identified using the z-score graph in (A) and taking 1 as the cutoff for statistical significance. (D) Gel mobility shift assay of a single stranded version similar to the motif shown in (C). The RNA sequence is: 5'-CCUCCUCGCUU-3' with 12 uracils flanking on the 5' side and 12 cytosines flanking on the 3' side to ensure single stranded RNA.

The majority of SELEX sequences (74 out of 82) contain one to four 5'-YGCUU-3' motifs (Figure 4). The motifs are mainly located in the 5' half of the

<i>k</i> -mer	expected	observed	Z-score
UGCUU	3.1	61	33.1
GCUU	14.6	116	27.6
UGCU	12	100	26.1
GCUUU	4.8	51	22.1
CUGCU	4.5	50	22
GCU	56.1	197	19.33
UGC	51.9	169	17.9
UUGC	13.8	69	15
CUU	58.3	160	14.8

Observed is the total occurrences of the *k*-mer within the sample of 82 SELEX sequences. The expected number is the mean number observed for 1000 sets of 82 random sequences that are equivalent in size to the SELEX sequences, and, for which, the background probabilities for each nucleotide position are the same as used during synthesis of the starting SELEX population. The Z-score was calculated using the mean and standard deviation derived from the 1000 sets of random sequences. The 9 *k*-mers in the table had the top 3 highest Z-scores within the respective length categories (3-mer, 4-mer, 5-mer) The *k*-mers are ordered by decreasing Z-score.

randomized region with around half of them in the same location as the wild type cTNT 5'-CGCUU-3' motif (Figure 4, column 1 and 2). This is consistent with the Z-score results which demonstrate that a stronger selection took place in the 5' half of the sequence. However, many of the sequences show that MBNL1 selected for the 5'-YGCUU-3' motif to change position with respect to the 5'-YGCUU-3' motif in the wild type cTNT. In some cases, this change is a slight shift of the position of the motif in comparison to the location of the wild type motif, whereas in others a whole new second, third or fourth additional motif appears nearby or in a different region. There are roughly

equal numbers of 5'-UGCUU-3' and 5'-CGCUU-3' motifs, again indicating that MBNL1 simply prefers a pyrimidine in the position just 5' to the G.

Binding studies were performed on 16 of the 82 sequences to determine the affinities of these sequences to MBNL1. Almost all sequences have affinities roughly

column 1		K_d	column 2		K_d
>WT	CGTGTCTGCTTTCCCTTCGGCTGGGGCCAC	31	>WT	CCTGTCTGCTTTCCCTTCGGCTGGGGCCAC	31
>A10	CCGCTGCTGCTTTCCCTTCGGCTGGGGCCAC	~1	>B12	CCGAGCTGCTTTCCCTTCGGCTGGGGCCAC	~1
>E10	CCGCTGCTGCTTTCCCTTCGGCTGGGGCCAC	~1	>C01	ATTGCGAGCTTTCCCTTCGGCTGGGGCCAC	~1
>B04	CCGCTGCTGCTTTCCCTTCGGCTGGGGCCAC	~1	>H12	CTTGCTGCTTTCCCTTCGGCTGGGGCCAC	~1
>G06	CCGCTGCTGCTTTCCCTTCGGCTGGGGCCAC	~1	>C06	CCATAAGCTGCTTTCCCTTCGGCTGGGGCCAC	~1
>G08	CCGCTGCTGCTTTCCCTTCGGCTGGGGCCAC	~1	>C09	CCTGTGCTTTCCCTTCGGCTGGGGCCAC	~1
>H07	CCGCTGCTGCTTTCCCTTCGGCTGGGGCCAC	~1	>D03	CCFTTCTGCTTTCCCTTCGGCTGGGGCCAC	~1
>H01	CCGCTGCTGCTTTCCCTTCGGCTGGGGCCAC	~1	>F02	CTAGCTGCTTTCCCTTCGGCTGGGGCCAC	~1
>C04	CCGCTGCTGCTTTCCCTTCGGCTGGGGCCAC	~1	>F06	CCGATCCGCTTTCCCTTCGGCTGGGGCCAC	1.0
>G12	CCGCTGCTGCTTTCCCTTCGGCTGGGGCCAC	~1	>F11	CCCGGCTGCTTTCCCTTCGGCTGGGGCCAC	~1
>F01	CTGCTGCTTTCCCTTCGGCTGGGGCCAC	~1	>E04	CTTCGCTGCTTTCCCTTCGGCTGGGGCCAC	~1
>H10	CCGCTGCTGCTTTCCCTTCGGCTGGGGCCAC	~1	>E12	CCGCTGCTTTCCCTTCGGCTGGGGCCAC	17
>B05	CCGCTGCTGCTTTCCCTTCGGCTGGGGCCAC	~1	>D05	CCATTGCTGCTTTCCCTTCGGCTGGGGCCAC	~1
>A04	CCGCTGCTGCTTTCCCTTCGGCTGGGGCCAC	6	>G04	CCGCTGCTTTCCCTTCGGCTGGGGCCAC	~1
>B11	CCGCTGCTGCTTTCCCTTCGGCTGGGGCCAC	~1	>A09	CCGCTGCTTTCCCTTCGGCTGGGGCCAC	~1
>F03	CCGCTGCTGCTTTCCCTTCGGCTGGGGCCAC	~1	>G11	CCGCTGCTTTCCCTTCGGCTGGGGCCAC	~1
>G01	CCGCTGCTGCTTTCCCTTCGGCTGGGGCCAC	~1	>B09	CCCTGCTTTCCCTTCGGCTGGGGCCAC	~1
>G02	CCGCTGCTGCTTTCCCTTCGGCTGGGGCCAC	~1	>B10	CCCTGCTTTCCCTTCGGCTGGGGCCAC	~1
>G10	CCGCTGCTGCTTTCCCTTCGGCTGGGGCCAC	~1	>C11	CCGCTGCTTTCCCTTCGGCTGGGGCCAC	~1
>C05	CCGCTGCTGCTTTCCCTTCGGCTGGGGCCAC	~1	>C07	GATGCGCTTTCCCTTCGGCTGGGGCCAC	~1
>E07	CAAGCTGCTTTCCCTTCGGCTGGGGCCAC	~1	>A01	CCTGCTGCTTTCCCTTCGGCTGGGGCCAC	~1
>H06	CTAGCTGCTTTCCCTTCGGCTGGGGCCAC	4	>H11	GACCGCTGCTTTCCCTTCGGCTGGGGCCAC	~1
>A02	CCGCTGCTGCTTTCCCTTCGGCTGGGGCCAC	~6	>F04	CTGATGCTGCTTTCCCTTCGGCTGGGGCCAC	~2.7
>A06	CCGCTGCTGCTTTCCCTTCGGCTGGGGCCAC	~1	>D07	TGTTGCTGCTTTCCCTTCGGCTGGGGCCAC	~1
>G09	CCGCTGCTGCTTTCCCTTCGGCTGGGGCCAC	~1	>F05	CCTGCTGCTTTCCCTTCGGCTGGGGCCAC	~1
>B02	CCGCTGCTGCTTTCCCTTCGGCTGGGGCCAC	~37	>F10	CTTCTGCTTTCCCTTCGGCTGGGGCCAC	~0.9
>E05	CTTCTGCTTTCCCTTCGGCTGGGGCCAC	~1	>C03	CATGCTGCTTTCCCTTCGGCTGGGGCCAC	~0.9
>A11	CCAGCTGCTTTCCCTTCGGCTGGGGCCAC	~1	>H09	CCCATCTGCTTTCCCTTCGGCTGGGGCCAC	~1
>D01	CCGCTGCTGCTTTCCCTTCGGCTGGGGCCAC	25	>D04	CCTGCTGCTTTCCCTTCGGCTGGGGCCAC	~1
>D08	CTAGCTGCTTTCCCTTCGGCTGGGGCCAC	~1	>D12	CTTCTGCTTTCCCTTCGGCTGGGGCCAC	~1
>E11	CTTCTGCTTTCCCTTCGGCTGGGGCCAC	~1	>H01	ACCGTGGCTTTCCCTTCGGCTGGGGCCAC	~37
>F07	CCAGCTGCTTTCCCTTCGGCTGGGGCCAC	~1	>D10	CCGCTGCTTTCCCTTCGGCTGGGGCCAC	~1
>D07	CCGCTGCTGCTTTCCCTTCGGCTGGGGCCAC	~1	>A07	CCAGGCTGCTTTCCCTTCGGCTGGGGCCAC	~1
>F09	CCAGCTGCTTTCCCTTCGGCTGGGGCCAC	~1	>E06	CCCATGCTGCTTTCCCTTCGGCTGGGGCCAC	~1
>A05	CCTGCTGCTTTCCCTTCGGCTGGGGCCAC	~1			
>B06	CCAGCTGCTTTCCCTTCGGCTGGGGCCAC	~1	column 3		
>C02	CCAGCTGCTTTCCCTTCGGCTGGGGCCAC	~1	>WT	CCGTGTCTGCTTTCCCTTCGGCTGGGGCCAC	31
>E01	CCCATGCTGCTTTCCCTTCGGCTGGGGCCAC	~3.3	>A12	GCCCTTTGGAGTATCGATACGCTGCGGAGTC	~1.2 μ M
>E03	CCGCTGCTGCTTTCCCTTCGGCTGGGGCCAC	~1	>B08	CCTACATCGGTTGTTGGTCCCTGATGCTAG	~1.2 μ M
>H03	CCTTTAGCTTTTCCCTTCGGCTGGGGCCAC	~1	>D09	CCCGCTCATTCGATACATTCGCTGCGGGTGC	~1
			>E09	CTTAAGGAACTTTCGACCTCTGCTGCTTATAC	~1
			>F12	CCATGCTTTCCGCTGCTTACCTTCAATGAC	~1
			>H04	CCAGCTGCTTTCCCTTCGGCTGGGGCCAC	~1
			>H05	CCGCTGCTGCTTTCCCTTCGGCTGGGGCCAC	31
			>C12	CCTTGGTTCAGTCCGCTTCATTCGCTGCC	>1.2 μ M

Figure 4. Most SELEX sequences contain at least one 5'-YCGUU-3' motif and bind with high affinity to MBNL1. The 5'-UGCUU-3' motifs are highlighted in green and the 5'-CGCUU-3' motifs in pink. The numbers to the right of the sequences indicate K_d values (nM) of MBNL1 binding. The dissociation constants displayed are based on binding to sequences containing the constant regions (81mers). The ~ indicates only one or two gels were used to calculate the binding affinity, otherwise at least three gels were analyzed. The top of each column contains the endogenous sequence (labeled WT and highlighted in yellow) for comparison. Note- T and U are used interchangeably, both represent uracil.

equal to or higher than the endogenous cTNT binding site (Figure 4 and 5). Only two did not bind MBNL1 at all, showing that the selection was successful. The binding affinities to SELEX sequences that contain 5'-YGCUU-3' are between 0.9 nM and 37 nM, similar to the wild type cTNT site which has a dissociation constant (K_d) of 31 nM. Interestingly, two of the three sequences that lack the motif, don't bind MBNL1, indicating the importance of the motif to MBNL1 binding (Figure 4 and 5). The one sequence (H05) that doesn't have 5'-YGCUU-3', but still binds MBNL1, contains 5'-YGCUG-3' which has a G in the position of the last U. This motif (5'-YGCUGY-3') is also in the 3' region of the wild type cTNT binding site (the right side of the stem) and is postulated to be an MBNL1 binding site by Ho et al. (16).

sequence	K_d 81mer (nM)	K_d 32mer (nM)
D01	25 ± 6	
C03	~ 0.9	
E01	~ 3	
B07	~ 3	~0.5
A04	6 ± 5	~84
B08	>1.2 μM	
H05	~31	
C12	>1.2 μM	
C01	~ 1	
F06	1 ± 0.1	~2
B01	~ 37	
H06	4 ± 4	
A02	~ 6	
B02	~ 37	
E12	17 ± 11	
WT cTNT binding site	31 ± 22	21 ± 11

It is possible that the constant regions are playing a role in the binding of MBNL1 to the SELEX sequences. There is a slight decrease in affinity for the endogenous cTNT SELEX template (81mer, $K_d \sim 31$ nM) compared to the 32mer endogenous cTNT site (32mer, $K_d \sim 21$ nM), indicating that the flanking constant regions have a slight affect on MBNL1 binding (Table 3). Similar studies were performed on several of the SELEX sequences by removing the flanking regions and determining binding affinities to the equivalent 32mer (Figure 5B). The F06 32mer shows similar binding to the equivalent 81mer while A04 and B07 shows decreased binding and increased binding to MBNL1, respectively. These studies show that the conserved flanking regions can either positively or negatively affect MBNL1 binding in a subset of SELEX sequences. However, the binding effects caused by the flanking regions are minor, except for A04 in which binding is reduced approximately 10-fold for the 32mer compared to the affinity of the 81mer (Table 3).

Mutations in the 5'-YGCUU-3' Consensus Motif Decrease MBNL1's Binding Affinity

To determine if MBNL1 is recognizing 5'-YGCUU-3' as a binding site, mutations were made in the motif of two SELEX sequences, F06 and E12. The 5'-CGCUU-3' motif in F06 was mutated to 5'-**UCCA**U-3', with the 3 bases mutated in bold. The 5'-UGCUU-3' motif in E12 was mutated to 5'-**ACCA**U-3', with the 3 bases mutated in bold (Figure 6). The mutations significantly decreased binding, indicating that MBNL1 requires 5'-YGCUU-3' to bind tightly. This could be because MBNL1 recognizes and binds the motif as a sequence, or the motif is necessary for setting up a structure (Figure 6).

Novel Binding Sites Identified in pre-mRNAs Regulated by MBNL1

With the cTNT intronic site the only known binding site of MBNL1 in a human endogenous transcript, we sought to find additional sites that contain the 5'-YGCUU-3' consensus motif. Exons known to be regulated by MBNL1 were selected and 5'-YGCUU-3' was searched for in the 200 bases of intronic sequence immediately upstream and downstream of the regulated exon. Two sites in MBNL1 intron 6 and one site in SERCA1 intron 22 have been identified. Site #1 in MBNL1 is located 75 bases upstream of the alternatively regulated exon 7 (counted from the 3' splice site to the G in 5'-YCGUU-3') and the second site is found 179 bases upstream within the same intron. Interestingly, these two sites flank a potential polypyrimidine tract and branch point sequence that could influence how MBNL1 regulates the splicing of the downstream exon. The potential MBNL1 site found in SERCA pre-mRNA is located 135 nucleotides downstream of exon 22 and may promote exon 22 inclusion (Figure 7B). All three potential sites include bases additional to the 5'-YCGUU-3' motif that are found in the longer motif of 5'-CCNNCUNUG₉CUU-3' identified from the Z-score analysis (Figure 3A and Figure 7A). The MBNL1 site #1 and the SERCA1 site both have runs of pyrimidines 5' to the 5'-YCGUU-3' motif that could be playing a role in MBNL1 binding. Gel mobility shift assays to 40mers containing the three potential sites (MBNL site #1 and #2 and SERCA1 site) were performed (Figure 7C). MBNL1 binds MBNL1 site #1 40mer with a K_d of 2 ± 1 nM and binds the SERCA1 40mer with a K_d of 15 ± 6

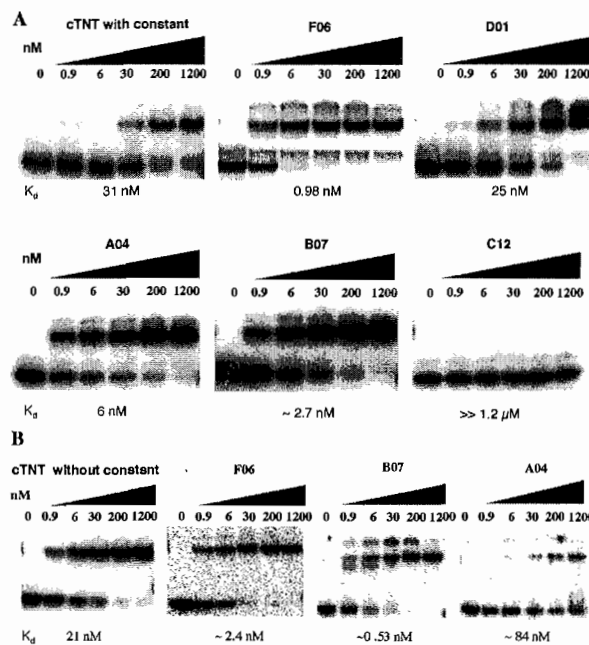


Figure 5. SELEX sequences bind MBNL1 with high affinity with or without the constant regions. (A) Gel mobility shift assays of the endogenous cTNT site flanked by the constant regions (cTNT with constant; 81mer) and five SELEX sequences. The MBNL1 concentrations (nM) are labeled at the top of each lane. (B) The flanking constant regions have variable effects on the binding affinities of MBNL1 to SELEX sequences. The flanking constant regions have been removed from three SELEX sequences and the wild type cTNT site (cTNT without constant; 32mer) and gel mobility shift assays performed to determine the K_d (labeled below each gel). The MBNL1 concentrations (nM) are labeled above each lane.

nM. MBNL1 binds both with high affinity, well within the range of the known MBNL1 binding site in cTNT, which has a K_d of 21 ± 11 nM.

DISCUSSION

The identification of key residues in MBNL1's binding site in cTNT provides a better understanding of MBNL1's RNA binding specificity. A doped SELEX was performed using the 32 base cTNT binding site. By round 5, the pool of RNA sequences bound as tight as the wild type cTNT binding site (Figure 1B). The 82 sequences

identified from round 5 group together based on similar sequence identity, with one outlier and no duplicates, indicating that over-selection did not occur (Figure 2). In

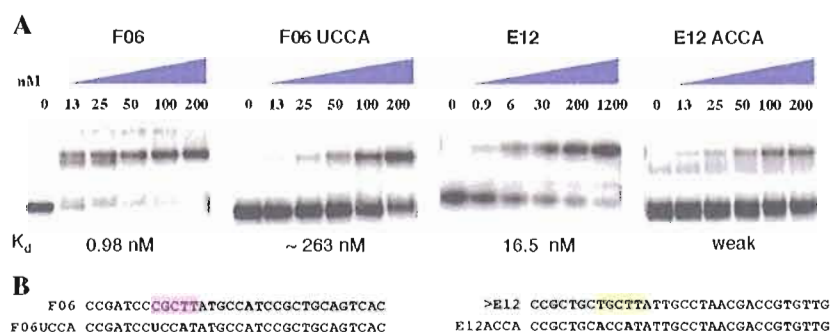


Figure 6. Mutations to the 5'-YGCUU-3' motif in SELEX sequences decrease binding. (A) Gel mobility shift assays of SELEX sequences F06 and E12 and corresponding mutants. The MBNLI concentrations (nM) are labeled at the top of each lane. (B) F06 and E12 each contain one occurrence of 5'-YGCUU-3' highlighted in pink (F06) or yellow (E12). Mutations are shown in the sequence below the original sequence in bold. Binding studies in (A) were performed with F06 and E12 that contain the constant regions (81mers) although only the 32mers are depicted in (B).

addition, 96 sequences from pool 0 were sequenced and showed no bias additional to the doping bias (data not shown). Strong blocks of similar sequence identity are apparent in the 5' region of most of the SELEX sequences, indicating a selective pressure in that region (Figure 3B).

This selective pressure is also apparent when the Z-scores for each position are analyzed. MBNLI generally selects uracil for binding, while adenosine is generally selected against in the SELEX sequences (Figure 3A). Also, over half of the residues in the 5' half of the randomized region show positive selective pressure, which corresponds to selective pressure on the left side of the stem-loop of the cTNT binding site (Figure 3B). This pressure is clear when each 5'-YGCUU-3' motif position is compared to the

position of 5'-CGCUU-3' in the cTNT wild type sequence. Most motifs remain in the same location as the wild type motif, or shift slightly to the 5' side (Figure 4). Using 4 as a Z-score cut off, a motif of 5'-CCNNCUNUG₉CUU-3' is apparent in the 5' half of the SELEX sequences (Figure 3C). With a bias in selection in the left side of the stem, and less selection in the right side, based on these findings it is possible that MBNL1 prefers to bind to weakly structured or single stranded RNA. A single stranded 40mer RNA containing the aforementioned motif with a 5' uracil tail and a 3' cytosine tail, was tested for binding to MBNL1. This single stranded motif RNA bound weakly to MBNL1 (Figure 3D). However, a small portion of the RNA bound at a low MBNL1 concentration, 0.9 nM, suggesting a portion of the RNA adopts a conformation that is favorable for binding by MBNL1. The 5'-YGCUU-3' motif is important for binding by MBNL1 as 3 mutations in the motif decrease MBNL1 binding affinities in F06 and E12 SELEX sequences. This indicates MBNL1 requires 5'-YGCUU-3' for high affinity binding to these sequences.

A *k*-mer analysis comparing the number of expected specific *k*-mers to the number of observed was performed using the 82 SELEX sequences. UGCUU and shorter derivatives were clearly enriched, with UGCUU having a Z-score of 33.1 (Table 2). This indicates that MBNL1 most likely selected to bind to this motif. Thirteen out of sixteen SELEX sequences tested bind MBNL1 with high affinity, and all contain one to four 5'-YGCUU-3' motifs. The two that don't bind MBNL1, don't contain the motif, indicating the importance of the motif. Only one, H05, binds MBNL1 and doesn't contain 5'-YGCUU-3' (Figure 4). However, it does contain 5'-YGCUGY-3', a

previously-identified MBNL1 binding motif, which may account for the high affinity (16). The 5-mer UGCUG occurred 17 times and was only expected to occur 3.2 times in the k -mer analysis, indicating enrichment of the previously identified motif 5'-UGCUG-3' (data not shown).

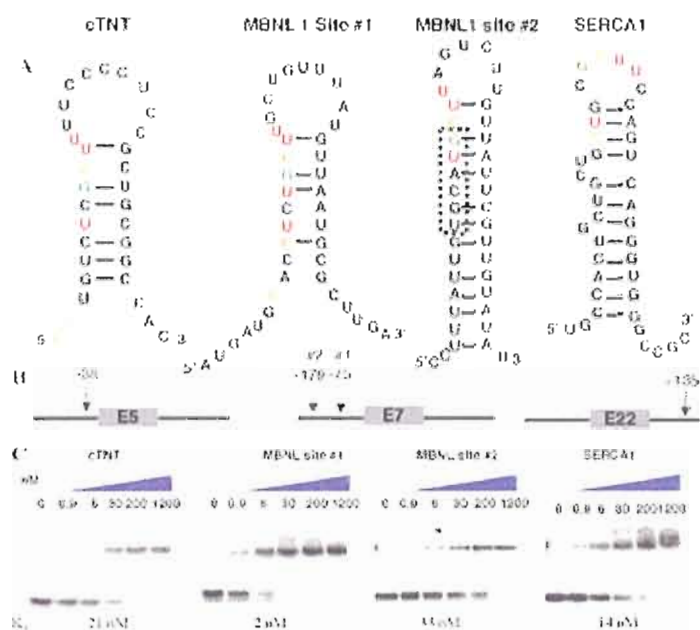


Figure 7. Novel MBNL1 binding sites identified near MBNL1-regulated exons in MBNL1 and SERCA1 pre-mRNA. (A) Shown is the potential secondary structures of the 40mer binding sites in MBNL1 and SERCA1 transcripts. On the left is the secondary structure of MBNL1's binding site (32mer) upstream of exon 5 in cTNT pre-mRNA (Warf and Berglund '07). The colors indicate residues found in the original consensus motif (Figure 3C). The dashed box in MBNL1 site #2 indicates a potential FOX-1/2 binding site. (B) The schematics indicate where the potential binding site is located relative to the MBNL1-regulated exon. (C) Gel mobility shift assays of the 32mer RNA, in the case of cTNT, or 40mer RNA in the case of MBNL1 site #1 and #2 and SERCA1 site, bound to MBNL1.

The constant flanking regions compose slightly more than half of the residues in total SELEX RNA sequences. In SELEX RNA A04, the constant regions decrease binding compared to an equivalent RNA without the constant regions. However, in

SELEX RNA B07, there is roughly a ten-fold increase in binding when the constant regions are present compared to when they are absent (Figure 5 and Table 2). In the case of SELEX RNA F06, the constant regions only marginally affect the binding affinity to MBNL1. The constant region may be setting up a preferred secondary structure for binding by MBNL1 in some cases, or blocking the binding site in other cases.

Reasoning that MBNL1 binds a motif shorter than twelve nucleotides (the length of the long motif) and the fact that 5'-YGCUU-3' is highly enriched in the SELEX sequences, 5'-YGCUU-3' was used to search for potential MBNL1 binding sites in pre-mRNA targets. Three potential binding sites were identified, two in MBNL1 and one in SERCA1, each within 200 bases of intronic sequence flanking exons regulated by MBNL1. One proposed mechanism for how MBNL1 regulates alternative splicing is by binding upstream of the regulated exon to promote exclusion of an alternatively spliced exon and binding downstream to promote inclusion. In total, 6 pre-mRNA MBNL1 targets were searched and 5 contained 5'-YGCUU-3' in the immediate 200 bases of flanking intronic sequence (data not shown). All pre-mRNAs but one followed the rule that the potential MBNL1 site is located downstream when MBNL1 promotes exon inclusion, and upstream when MBNL1 promotes exclusion. The SERCA1 site potentially folds into a similar weak stem-loop as the cTNT binding site, both containing at least one pyrimidine-pyrimidine mismatch (Figure 7A). The MBNL1 site in SERCA1 is 135 bases downstream of the regulated exon and MBNL1 promotes the inclusion of that exon, which follows the rule (23, 24).

MBNL1 binds two possible sites in MBNL1 pre-mRNA, one 75 bases upstream of the regulated exon 7 (site #1) and one 179 bases upstream of exon 7 (site #2). These two sites flank a predicted polypyrimidine tract and branch point sequence. It is possible that binding by MBNL1 to the two sites blocks binding of splicing factors SF1 and/or U2AF, which would promote skipping of exon 7. Alternatively, bound MBNL1 doesn't allow bound SF1 and bound 65 to recruit other splicing factors because the sites overlap. In addition, a FOX-1/2 binding site (5'-UGCAUG₆-3') overlaps with the potential MBNL1 site of 5'-UG₉CUU-3' in the left side of the stem (Figure 7A). FOX-1/2 and MBNL1 may work together to regulate the splicing of MBNL1's exon 7, as FOX-1/2 has been generally shown to promote exon exclusion by binding upstream of regulated exons (4). In addition, the yeast protein RPL32 binds a bulged stem-loops structure in its own transcript that is thought to block binding of U1 RNA. This blocking event inhibits inclusion of the downstream, regulated exon (25). It is possible that MBNL1 behaves similarly to FOX-1/2 and RPL32 and blocks binding of essential splicing factors. *In vivo* binding studies are under way to determine if the three, novel MBNL1 binding sites are required for MBNL1 alternative splicing regulation of MBNL1 and SERCA1 pre-mRNAs.

Identification of additional MBNL1 binding sites in pre-mRNA targets will be very useful for understanding how MBNL1 regulates alternative splicing. Our results should help in the search for novel MBNL1 sites; however, with both a structural aspect and a sequence aspect, MBNL1's binding sites may be degenerate and challenging to recognize in a sea of intronic sequence.

CHAPTER IV

CONCLUSIONS AND FUTURE DIRECTIONS

The muscleblind proteins function as alternative splicing regulators and are involved in the disease mechanism of myotonic dystrophy (DM). They likely regulate alternative splicing by directly binding pre-mRNA targets, such as cTNT (1, 2). However, little is known about their RNA binding specificity. The human protein, MBNL1, co-localizes with aberrant CUG repeat and CCUG repeat RNA in DM type 1 (DM1) and DM type 2 (DM2) patients, respectively. Muscleblind becomes sequestered to these aberrant repeat RNAs in the nuclei of DM patients, forming nuclear foci (3-5). This sequestration event removes muscleblind from regulating alternative splicing, leading to mis-splicing events and disease. Mis-spliced transcripts include cTNT, IR, CIC-1 and a handful of others (reviewed in ref 5). Aberrant fetal splice products of IR and CIC-1 are thought to lead to the DM1 symptoms of insulin resistance and myotonia, respectively (6, 7). DM patients experience a multitude of symptoms, however, only two symptoms have been accounted for from mis-splicing events. There are likely many more undiscovered links between aberrant alternative splicing events and DM symptoms.

Less complex organisms, such as *Drosophila*, have a single muscleblind gene that contains two zinc fingers, while more complex organisms, such as humans, have several muscleblind genes with up to four zinc fingers (8). The zinc finger domains bind RNA in

a specific fashion (2, 9). Muscleblind-like 1 in humans (MBNL1), binds CUG and CCUG repeats in their double stranded form, not as single stranded RNA (10, 11). In addition, mutational studies on CUG repeats show that MBNL1 prefers weak double-stranded RNA with pyrimidine-pyrimidine mismatches (2). MBNL1 binds a weak stem structure in cTNT intron 4 and promotes the exclusion of downstream exon 5. The stem has characteristics similar to CUG repeat stem-loops; both stems include pyrimidine-pyrimidine mismatches interspersed with G-C and/or G-U base pairs (2). Despite the identification of a handful of pre-mRNAs of which MBNL1 regulates the splicing, no other binding sites have been identified. A deeper understanding of muscleblind's RNA binding specificity will allow a complete list of muscleblind regulated transcripts to be identified and provide important information for understanding muscleblind's alternative splicing mechanism.

CONCLUSIONS FROM CHAPTER II: RNA BINDING SPECIFICITY OF *DROSOPHILA* MUSCLEBLIND

As mentioned previously, *Drosophila* have one muscleblind gene (Mbl) with splice isoforms that contain two zinc fingers, the RNA binding domains (8). Mbl regulates alternative splicing, like MBNL1, and Mbl regulates at least two of the same MBNL1-regulated transcripts in *Drosophila*, *Tnnt3* and *Tun* (ZASP in humans) (12, 13). *Drosophila* has provided a useful DM1 model that has phenotypes representative of DM1 symptoms (14). In addition, MBNL1 and Mbl are functionally conserved as MBNL1 can rescue Mbl knockout phenotypes (15). *Drosophila* muscleblind simplifies muscleblind RNA binding specificity studies due to the lack of complicated gene families and high

levels of splice isoforms. Therefore, SELEX was performed on Mbl to determine Mbl's RNA binding specificity. Binding studies were also performed with Mbl on CUG repeat RNA, the toxic RNA in DM1.

Co-localization of Mbl with CUG repeats is observed in the nuclei of the *Drosophila* DM1 line expressing 480 interrupted CUG repeats (14). To confirm that Mbl binds CUG repeats, and to investigate which nucleotides it interacts with, *in-vitro* binding studies were performed on a screen of CUG repeat mutant RNAs. From this screen, it was identified that Mbl is a highly specific RNA binding protein that requires a pyrimidine-pyrimidine mismatch in the U-U mismatch position and does not tolerate changes in 5' to 3' direction, flipping of C-G base pairs to G-C or complete Watson-Crick base pairing. This specificity is similar to MBNL1's specificity; however, MBNL1 has decreased binding to some of the aforementioned CUG repeat mutants while binding by Mbl is completely abolished. This indicates that Mbl follows a similar binding trend to MBNL1, but is more specific. In agreement, Mbl binds an MBNL1 binding site in cTNT pre-mRNA (32 nucleotides) with high affinity. The cTNT binding site forms a weak stem loop resembling CUG repeats in some aspects (2). When the base pairs forming the stem are mutated such that the stem can no longer form, Mbl binding is abolished, while binding by MBNL1 decreases by 100 fold (2).

SELEX was performed on Mbl to determine structural and sequence elements that Mbl recognizes in order to bind. After six rounds of SELEX, 25 sequences were identified and grouped into two main groups, sequences containing 5'-AGUCU-3' and sequences that don't. Of the SELEX sequences tested, the group that contains 5'-

AGUCU-3' bind Mbl with high affinity and the group that doesn't contain 5'-AGUCU-3' bind Mbl weakly or not at all. Structure probing was performed on a SELEX sequence, SELEX RNA 20, that contains 5'-AGUCU-3' and was determined to fold into a two stem structure with two single stranded tail regions. Footprinting analysis showed that Mbl binds three guanosines (Gs), one in the 5'-AGUCU-3' motif, one in a loop region and one in a single stranded tail region. Mutations to any of these three Gs abolished binding by Mbl, indicating that the Gs directly interact with Mbl and/or set up a secondary/tertiary structure necessary for binding by Mbl.

Truncation and mutation studies performed on SELEX RNA 20 reveal regions that aid in binding by Mbl. The two single stranded tails are necessary for high affinity binding by Mbl, and when removed or mutated, binding is decreased significantly. Removal of either of the stems or truncation of one of the stems significantly decreases binding by Mbl. The only truncation that does not affect binding by Mbl was shortening the 13 base pair stem to a 4 base pair stem. Taken together, this information indicates that structural elements and sequence elements are both required for tight binding by Mbl to SELEX RNA 20. Interestingly, there is only one obvious pyrimidine-pyrimidine mismatch in SELEX RNA 20, and when mutated to form a Watson-Crick base pair, binding was decreased. This preference by MBNL1 and Mbl to bind pyrimidine-pyrimidine mismatches in a toxic RNA target (CUG repeats), a SELEX identified RNA target (SELEX RNA 20) and a pre-mRNA target (cTNT), indicates that MBNL1 and Mbl binding sites in pre-mRNA targets may be similar.

FUTURE DIRECTIONS FOR CHAPTER II: RNA BINDING SPECIFICITY OF
DROSOPHILA MUSCLEBLIND

Both MBNL1 and Mbl bind RNA in a specific fashion; however, it is unknown which amino acids confer this specificity. The first two zinc fingers of Mbl and MBNL1 are highly similar in amino acid composition, with some key differences. These could account for the difference in specificity, as Mbl seems to be more specific than MBNL1. It would be interesting to change amino acids in MBNL1 to make it more like Mbl and identify residues that are responsible for Mbl's highly specific RNA binding properties. For example, one could replace the asparagine in MBNL1's first zinc finger to the arginine found in Mbl in that position. This would be replacing a small, non-polar amino acid with a large, polar amino acid. In addition, it would be interesting to see if changes in spacing of the CCCH in the zinc fingers play a role in specificity. Mbl's zinc finger one is two amino acids shorter than MBNL1's zinc finger one. Instead of making changes in the RNA to identify muscleblind's RNA binding specificity, this project would be looking at it from the perspective of the protein.

MBNL1 has been challenging to crystallize, most likely because of MBNL1's unstructured nature (2). Mbl is a great candidate with which to attempt crystallization studies as the portion that binds RNA is small, only 105 amino acids and contains only two zinc fingers, not four. Obtaining a crystal structure of muscleblind bound to RNA would greatly help the DM1 field and the alternative splicing field. A crystal structure would provide information on which nucleic acid residues Mbl interacts with when it binds. This knowledge would give us insight into Mbl's preferred RNA binding site

motif. For example, does Mbl directly interact with pyrimidine-pyrimidine mismatches or the backbone of the RNA? With Mbl and MBNL1 having such similar RNA binding specificities, this information would help to identify both MBNL1's and Mbl's binding sites in pre-mRNA targets. This would help to complete the list of MBNL1's pre-mRNA targets and broaden our understanding of how muscleblind regulates alternative splicing events.

CONCLUSIONS FROM CHAPTER III: DOPED SELEX WITH MBNL1 REVEALS A MOTIF USED TO IDENTIFY NOVEL MBNL1 BINDING SITES

MBNL1 regulates the alternative splicing of cTNT pre-mRNA, along with several other pre-mRNAs (5). cTNT is misregulated in DM1 patients due to aberrant sequestration of MBNL1. MBNL1 binds a weak stem loop structure upstream of cTNT exon 5 and prevents the inclusion of exon 5. The weak stem is composed of G-U and C-G base pairs and two pyrimidine-pyrimidine mismatches. The loop is composed entirely of pyrimidines and is proposed to be the polypyrimidine tract that the splicing factor U2AF binds (M. Bryan Warf, unpublished results). MBNL1 interacts with the stem, however, it is not known specifically which aspects of the stem it is recognizing in order to bind. Doped SELEX was performed on MBNL1, using the cTNT binding site (32 nucleotides) as a template, to determine what structural and sequence elements MBNL1 recognizes to bind.

Five rounds of doped SELEX were carried out and 82 clones were sequenced. In general, there is selective pressure on most of the 32 positions toward uracil and against

adenosine. In addition, there is strong positive selection in the 5' randomized region of the SELEX RNAs toward a 5'-CCNNCUNUG₉CUU-3' motif. This region correlates to the left side of the stem of the cTNT binding site. Four positions out of the aforementioned 12 nucleotide motif, 5'-G₉CUU-3', were selected to remain the same as cTNT wild type sequence, while the other positions selected to change away from the original pyrimidine in the wild type sequence toward the other pyrimidine. From this information, the 5'-YGCUU-3' motif was identified to be a potential binding site for MBNL1. An analysis was performed to identify highly enriched *k*-mers in the SELEX sequences and UGCUU was the highest-ranking *k*-mer, based on Z-scores.

The majority (78 out of 82) of round 5 SELEX sequences contain one to several 5'-YGCUU-3' motifs. The cTNT wild type binding site contains one 5'-CGCUU-3' motif in the 5' half of the randomized region. Over half of the SELEX sequences contained the motif in the same location as the cTNT binding site, in the rest, the location changed. Of the 16 SELEX sequences tested for binding to MBNL1, 14 bind with high affinity (K_d 's ~ 37-0.9 nM) and two bind weakly, with a dissociation constant of > 1.2 μ M. The two that bind weakly do not contain a 5'-YGCUU-3' motif, while 13 out of the 14 that bind tightly do contains 5'-YGCUU-3'. This indicates that the motif is most likely playing an important role in binding by MBNL1. To determine if the 5'-YGCUU-3' motifs are playing a role in binding by MBNL1, mutations in the motif were made to two SELEX sequences, F06 and E12. In both cases, the mutations reduced binding by MBNL1 significantly (25 fold decrease for F06 and 5 fold decrease for E12). This

indicates that the motif is promoting binding by MBNL1 by providing a necessary sequence or by setting up an important secondary or tertiary structure.

To determine if 5'-YGCUU-3' is an endogenous MBNL1 binding site, pre-mRNA targets regulated by MBNL1 were searched for the motif. Three potential MBNL1 binding sites containing 5'-YGCUU-3' were identified, two in MBNL1 and one in SERCA1. The two MBNL1 sites are 75 and 179 bases upstream of the MNBL1 regulated exon, exon 7. MBNL1 promotes exclusion of MBNL1 exon 7 (16). The location of the two bindings sites is consistent with the model that MBNL1 binds upstream of a regulated exon to promotes exon exclusion, and downstream of a regulated exon to promotes inclusion of that exon. Also in agreement with the model, the potential MBNL1 SERCA1 site is 135 bases downstream of the exon of which it promotes the inclusion (17). MBNL1 binds all three sites (40 nucleotides each) with high affinity *in vitro*, comparable to the affinity of MBNL1 for the cTNT binding site. The sites have the potential to fold into stems containing pyrimidine-pyrimidine mismatches, similar to the cTNT binding site, but experiments should be performed to determine their structures.

The two potential MBNL1 binding sites identified upstream of MBNL1 exon 7 flank a putative polypyrimidine tract and a branch point sequence. It is possible that MBNL1 binds the sites and blocks the binding of U2AF and/or SF1. This would inhibit recognition of exon 7 and cause skipping (exon exclusion). In the case of SERCA1, MBNL1 may bind the potential binding site downstream of exon 22 and recruit important splicing factors. This would promote the recognition of exon 22 and include exon 22 in the mRNA. Determining the potential mechanisms for how MBNL1 regulates the

alternative splicing of MBNL1 and SERCA1 pre-mRNAs would be of high interest to the field of alternative splicing.

FUTURE DIRECTIONS FOR CHAPTER III: DOPED SELEX WITH MBNL1 REVEALS A MOTIF USED TO IDENTIFY NOVEL MBNL1 BINDING SITES

The identification of three potential binding sites of MBNL1 in MBNL1-regulated pre-mRNAs is an exciting start to a long journey. First off, other potential MBNL1 binding sites containing 5'-YGCUU-3' need to be tested for binding by MBNL1. Further experimental exploration is needed to determine if MBNL1 regulates alternative splicing via binding at those locations. Also, it is possible that the location of the binding site does not dictate if the nearby exon is included or excluded; this should be analyzed. Assuming MBNL1 does bind and regulate alternative splicing through these newly identified binding sites, this would provide valuable information regarding MBNL1's RNA binding specificity. Structure probing to determine the structure of these sites and footprinting analysis should be performed to analyze which residues MBNL1 interacts with for tight binding. Additional information from round 7 SELEX sequences may also aid in identifying important residues for binding by MBNL1. The small handful of MBNL1 binding sites and high affinity SELEX sequences could be compared and a more accurate consensus motif identified that takes in to account structural elements. With this new consensus motif, the entire human genome could be searched and used to complete the list of MBNL1 binding sites. This will open up the possibility of linking MBNL1 regulated splicing events to mis-splicing events in DM1 patients. This would potentially draw additional connections between symptoms and mis-splicing events.

Information defining MBNL1 binding sites in pre-mRNA splicing targets would allow for more in depth drug design. Currently, the Berglund lab is setting up a screen to screen for drugs that inhibit the binding of MBNL1 to CUG repeats. The idea is to release MBNL1 from binding the toxic CUG repeats in DM1 patients and allow it to properly regulate alternative splicing. However, one major caveat is that if the drug also binds MBNL1's pre-mRNA targets, then it would prevent MBNL1 from regulating splicing properly. However, with more information on the specifics of MBNL1 binding sites, the subtle differences between MBNL1's binding sites on CUG repeats and endogenous targets could improve drug design. Drugs could be designed to bind only CUG repeats and not endogenous targets. However, without more known MBNL1 binding sites, this is a challenge.

MBNL1's alternative splicing regulation of MBNL1 and SERCA1 pre-mRNA should be analyzed in depth. Does the location dictate if MBNL1 exon 7 is excluded or whether SERCA1 exon 22 is included? Does MBNL1 promote exon inclusion by recruiting splicing factors or prevent exon inclusion by blocking splicing factors? These questions need to be answered to begin to understand MBNL1's mode of alternative splicing regulation. Very few tissue specific alternative splicing regulators have been fully characterized, understanding MBNL1's mechanisms would significantly help to advance the field.

APPENDIX

DOPED SELEX ON MBNL1 REVEALS ROUND 7 SEQUENCES HAVE MULTIPLE 5'-YGCUU-3' MOTIFS

Doped SELEX was performed on MBNL1 using a known MBNL1 binding site (32 nucleotides) in cTNT pre-mRNA as a template (Chapter 3). Seven rounds of SELEX were completed with tight binding of MBNL1 by round 7. 76 clones from round 7 were sequenced and aligned based on sequence identity using the Jalview dendrogram analysis tool (Clamp '04). The alignment reveals 2 main groups (Group 3 and 4) and 6 small groups (Groups 1, 2 and 4-8) (Figure 1). The wild type cTNT binding site sequence groups with Group 3 SELEX sequences the largest group. Group 3 contains two main regions of high sequence identity: 5'-GCUU-3' in the 5' half of the randomized region, and 5'-UCCGCU-3' in the 3' half of the randomized region (Figure 1B). The 5'-GCUU-3' consensus motif is also found in the cTNT binding site, indicating that MBNL1 most likely selected to maintain those four bases of the 32 nucleotide binding site, presumably in order to bind with high affinity. MBNL1 binding studies will be performed in order to determine if 5'-GCUU-3' is important for binding by MBNL1.

Round 5 SELEX sequences showed a strong selection for single or multiple 5'-YGCUU-3' motifs in the majority of sequences. After two additional rounds of SELEX, round 7 SELEX sequences show that every sequence but two sequences contain the same

5'-YGCUU-3 motif. Many contain a shift in register of the location of the motif

(Figure 2). Most Group 4 SELEX sequences contain a second 5'-YGCUU-3' motif just

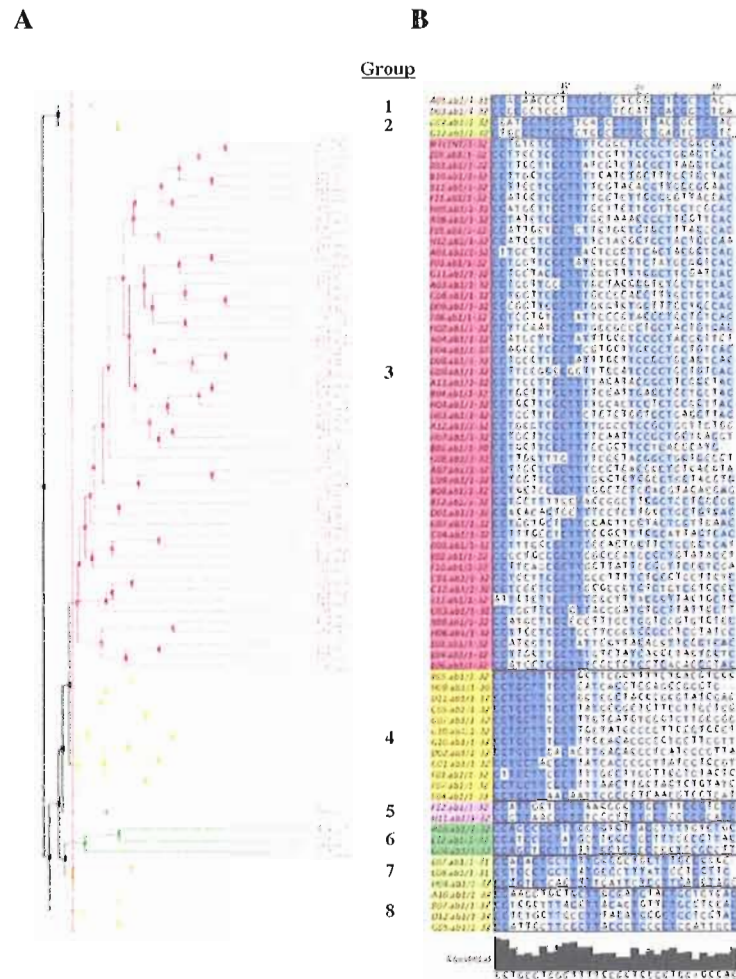


Figure 1. The 76 SELEX sequences from round 7 fall into two main groups. (A) Dendrogram showing the 32 nucleotide randomized region of the SELEX sequences aligned based on sequence identity. The wild type cTNT site is included and groups in Group 3. (B) The sequences ordered based on the dendrogram in (A), with darker blue representing higher sequence identity. The consensus at the bottom indicates the overall consensus sequence. The 4 nucleotides underlined are a potential 5'-GCUU-3' consensus motif.

5' of the location of the original 5'-CGCUU-3' motif in the cTNT binding site. This new register of a highly selected motif may indicate selection of a secondary structure as well. This is yet to be determined and requires binding affinity assays, structure probing assays and/or footprinting assays, however. It appears that MBNL1 selected for similar, if not identical, sequence motifs in round 7 as it did by round 5. It will be interesting to see if MBNL1 selected for the same or different structural elements in round 5 as it may have in round 7.



Figure 2. Round 7 SELEX sequences group based on where 5'-YGCUU-3' motif is located. The groupings are based off the dendrogram analysis in Figure 1. The 5'-UGCUU-3' motifs are highlighted in green and the 5'-CGCUU-3' motifs in pink. The top of each group contains the endogenous cTNT binding site (labeled WT) for comparison. The cTNT binding site contains one 5'-YGCUU-3' motif highlighted in yellow.

REFERENCES

CHAPTER I

1. Burge, C. B., Tuschl, T., and Sharp, P. A. (1999) *The RNA World*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
2. Staley, J. P., and Guthrie, C. (1998) Mechanical devices of the spliceosome: motors, clocks, springs, and things, *Cell* 92, 315-326.
3. Black, D. L. (2003) Mechanisms of alternative pre-messenger RNA splicing, *Annu Rev Biochem* 72, 291-336.
4. Ranum, L. P., and Cooper, T. A. (2006) RNA-mediated neuromuscular disorders, *Annu Rev Neurosci* 29, 259-277.
5. Cho, D. H., and Tapscott, S. J. (2007) Myotonic dystrophy: emerging mechanisms for DM1 and DM2, *Biochim Biophys Acta* 1772, 195-204.
6. Brook, J. D., McCurrach, M. E., Harley, H. G., Buckler, A. J., Church, D., Aburatani, H., Hunter, K., Stanton, V. P., Thirion, J. P., Hudson, T., and et al. (1992) Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member, *Cell* 68, 799-808.
7. Liquori, C. L., Ricker, K., Moseley, M. L., Jacobsen, J. F., Kress, W., Naylor, S. L., Day, J. W., and Ranum, L. P. (2001) Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9, *Science* 293, 864-867.
8. de Haro, M., Al-Ramahi, I., De Gouyon, B., Ukani, L., Rosa, A., Faustino, N. A., Ashizawa, T., Cooper, T. A., and Botas, J. (2006) MBNL1 and CUGBP1 modify expanded CUG-induced toxicity in a *Drosophila* model of myotonic dystrophy type 1, *Hum Mol Genet* 15, 2138-2145.
9. Ranum, L. P., and Day, J. W. (2004) Pathogenic RNA repeats: an expanding role in genetic disease, *Trends Genet* 20, 506-512.

10. Berul, C. I., Maguire, C. T., Aronovitz, M. J., Greenwood, J., Miller, C., Gehrman, J., Housman, D., Mendelsohn, M. E., and Reddy, S. (1999) DMPK dosage alterations result in atrioventricular conduction abnormalities in a mouse myotonic dystrophy model, *J Clin Invest* 103, R1-7.
11. Berul, C. I., Maguire, C. T., Gehrman, J., and Reddy, S. (2000) Progressive atrioventricular conduction block in a mouse myotonic dystrophy model, *J Interv Card Electrophysiol* 4, 351-358.
12. Furling, D., Lemieux, D., Taneja, K., and Puymirat, J. (2001) Decreased levels of myotonic dystrophy protein kinase (DMPK) and delayed differentiation in human myotonic dystrophy myoblasts, *Neuromuscul Disord* 11, 728-735.
13. Klesert, T. R., Cho, D. H., Clark, J. I., Maylie, J., Adelman, J., Snider, L., Yuen, E. C., Soriano, P., and Tapscott, S. J. (2000) Mice deficient in Six5 develop cataracts: implications for myotonic dystrophy, *Nat Genet* 25, 105-109.
14. Sarkar, P. S., Appukuttan, B., Han, J., Ito, Y., Ai, C., Tsai, W., Chai, Y., Stout, J. T., and Reddy, S. (2000) Heterozygous loss of Six5 in mice is sufficient to cause ocular cataracts, *Nat Genet* 25, 110-114.
15. Davis, B. M., McCurrach, M. E., Taneja, K. L., Singer, R. H., and Housman, D. E. (1997) Expansion of a CUG trinucleotide repeat in the 3' untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts, *Proc Natl Acad Sci U S A* 94, 7388-7393.
16. Hamshere, M. G., Newman, E. E., Alwazzan, M., Athwal, B. S., and Brook, J. D. (1997) Transcriptional abnormality in myotonic dystrophy affects DMPK but not neighboring genes, *Proc Natl Acad Sci U S A* 94, 7394-7399.
17. Taneja, K. L., McCurrach, M., Schalling, M., Housman, D., and Singer, R. H. (1995) Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues, *J Cell Biol* 128, 995-1002.
18. Dansithong, W., Paul, S., Comai, L., and Reddy, S. (2005) MBNL1 is the primary determinant of focus formation and aberrant insulin receptor splicing in DM1, *J Biol Chem* 280, 5773-5780.
19. Kino, Y., Mori, D., Oma, Y., Takeshita, Y., Sasagawa, N., and Ishiura, S. (2004) Muscleblind protein, MBNL1/EXP, binds specifically to CHHG repeats, *Hum Mol Genet* 13, 495-507.

20. Miller, J. W., Urbinati, C. R., Teng-Umnuay, P., Stenberg, M. G., Byrne, B. J., Thornton, C. A., and Swanson, M. S. (2000) Recruitment of human muscleblind proteins to (CUG)_n expansions associated with myotonic dystrophy, *Embo J* 19, 4439-4448.
21. Ho, T. H., Charlet, B. N., Poulos, M. G., Singh, G., Swanson, M. S., and Cooper, T. A. (2004) Muscleblind proteins regulate alternative splicing, *Embo J* 23, 3103-3112.
22. Fardaei, M., Larkin, K., Brook, J. D., and Hamshere, M. G. (2001) In vivo co-localisation of MBNL protein with DMPK expanded-repeat transcripts, *Nucleic Acids Res* 29, 2766-2771.
23. Mankodi, A., Urbinati, C. R., Yuan, Q. P., Moxley, R. T., Sansone, V., Krym, M., Henderson, D., Schalling, M., Swanson, M. S., and Thornton, C. A. (2001) Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2, *Hum Mol Genet* 10, 2165-2170.
24. Kanadia, R. N., Johnstone, K. A., Mankodi, A., Lungu, C., Thornton, C. A., Esson, D., Timmers, A. M., Hauswirth, W. W., and Swanson, M. S. (2003) A muscleblind knockout model for myotonic dystrophy, *Science* 302, 1978-1980.
25. Lin, X., Miller, J. W., Mankodi, A., Kanadia, R. N., Yuan, Y., Moxley, R. T., Swanson, M. S., and Thornton, C. A. (2006) Failure of MBNL1-dependent post-natal splicing transitions in myotonic dystrophy, *Hum Mol Genet* 15, 2087-2097.
26. Charlet, B. N., Savkur, R. S., Singh, G., Philips, A. V., Grice, E. A., and Cooper, T. A. (2002) Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing, *Mol Cell* 10, 45-53.
27. Mankodi, A., Takahashi, M. P., Jiang, H., Beck, C. L., Bowers, W. J., Moxley, R. T., Cannon, S. C., and Thornton, C. A. (2002) Expanded CUG repeats trigger aberrant splicing of ClC-1 chloride channel pre-mRNA and hyperexcitability of skeletal muscle in myotonic dystrophy, *Mol Cell* 10, 35-44.
28. Savkur, R. S., Philips, A. V., and Cooper, T. A. (2001) Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy, *Nat Genet* 29, 40-47.
29. Timchenko, N. A., Cai, Z. J., Welm, A. L., Reddy, S., Ashizawa, T., and Timchenko, L. T. (2001) RNA CUG repeats sequester CUGBP1 and alter protein levels and activity of CUGBP1, *J Biol Chem* 276, 7820-7826.

30. Osborne, R. J., and Thornton, C. A. (2006) RNA-dominant diseases, *Hum Mol Genet* 15 Spec No 2, R162-169.
31. Bashaw, G. J., and Baker, B. S. (1997) The regulation of the *Drosophila* msl-2 gene reveals a function for Sex-lethal in translational control, *Cell* 89, 789-798.
32. Libri, D., Piseri, A., and Fiszman, M. Y. (1991) Tissue-specific splicing in vivo of the beta-tropomyosin gene: dependence on an RNA secondary structure, *Science* 252, 1842-1845.
33. Skrisovska, L., Bourgeois, C. F., Stefl, R., Grellscheid, S. N., Kister, L., Wenter, P., Elliott, D. J., Stevenin, J., and Allain, F. H. (2007) The testis-specific human protein RBMY recognizes RNA through a novel mode of interaction, *EMBO Rep* 8, 372-379.
34. Eng, F. J., and Warner, J. R. (1991) Structural basis for the regulation of splicing of a yeast messenger RNA, *Cell* 65, 797-804.
35. Warf, M. B., and Berglund, J. A. (2007) MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T, *RNA* 13, 2238-2251.
36. Zhang, C., Zhang, Z., Castle, J., Sun, S., Johnson, J., Krainer, A. R., and Zhang, M. Q. (2008) Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2, *Genes Dev* 22, 2550-2563.
37. Ule, J., Stefani, G., Mele, A., Ruggiu, M., Wang, X., Taneri, B., Gaasterland, T., Blencowe, B. J., and Darnell, R. B. (2006) An RNA map predicting Nova-dependent splicing regulation, *Nature* 444, 580-586.
38. Philips, A. V., Timchenko, L. T., and Cooper, T. A. (1998) Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy, *Science* 280, 737-741.
39. Pascual, M., Vicente, M., Monferrer, L., and Artero, R. (2006) The Muscleblind family of proteins: an emerging class of regulators of developmentally programmed alternative splicing, *Differentiation* 74, 65-80.
40. Vicente, M., Monferrer, L., Poulos, M. G., Houseley, J., Monckton, D. G., O'Dell K, M., Swanson, M. S., and Artero, R. D. (2007) Muscleblind isoforms are functionally distinct and regulate alpha-actinin splicing, *Differentiation* 75, 427-440.

41. Fardaei, M., Rogers, M. T., Thorpe, H. M., Larkin, K., Hamshere, M. G., Harper, P. S., and Brook, J. D. (2002) Three proteins, MBNL, MBLL and MBXL, co-localize in vivo with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells, *Hum Mol Genet* 11, 805-814.
42. Kanadia, R. N., Urbinati, C. R., Crusselle, V. J., Luo, D., Lee, Y. J., Harrison, J. K., Oh, S. P., and Swanson, M. S. (2003) Developmental expression of mouse muscleblind genes Mbnl1, Mbnl2 and Mbnl3, *Gene Expr Patterns* 3, 459-462.
43. Adereth, Y., Dammai, V., Kose, N., Li, R., and Hsu, T. (2005) RNA-dependent integrin alpha3 protein localization regulated by the Muscleblind-like protein MLP1, *Nat Cell Biol* 7, 1240-1247.
44. Monferrer, L., and Artero, R. (2006) An interspecific functional complementation test in *Drosophila* for introductory genetics laboratory courses, *J Hered* 97, 67-73.
45. Begemann, G., Paricio, N., Artero, R., Kiss, I., Perez-Alonso, M., and Mlodzik, M. (1997) muscleblind, a gene required for photoreceptor differentiation in *Drosophila*, encodes novel nuclear Cys3His-type zinc-finger-containing proteins, *Development* 124, 4321-4331.
46. Artero, R., Prokop, A., Paricio, N., Begemann, G., Pueyo, I., Mlodzik, M., Perez-Alonso, M., and Baylies, M. K. (1998) The muscleblind gene participates in the organization of Z-bands and epidermal attachments of *Drosophila* muscles and is regulated by Dmef2, *Dev Biol* 195, 131-143.
47. Machuca-Tzili, L., Thorpe, H., Robinson, T. E., Sewry, C., and Brook, J. D. (2006) Flies deficient in Muscleblind protein model features of myotonic dystrophy with altered splice forms of Z-band associated transcripts, *Hum Genet* 120, 487-499.
48. Vicente-Crespo, M., Pascual, M., Fernandez-Costa, J. M., Garcia-Lopez, A., Monferrer, L., Miranda, M. E., Zhou, L., and Artero, R. D. (2008) *Drosophila* muscleblind is involved in troponin T alternative splicing and apoptosis, *PLoS ONE* 3, e1613.
49. Napierala, M., and Krzyzosiak, W. J. (1997) CUG repeats present in myotonin kinase RNA form metastable "slippery" hairpins, *J Biol Chem* 272, 31079-31085.
50. Mooers, B. H., Logue, J. S., and Berglund, J. A. (2005) The structural basis of myotonic dystrophy from the crystal structure of CUG repeats, *Proc Natl Acad Sci U S A* 102, 16626-16631.

51. Yuan, Y., Compton, S. A., Sobczak, K., Stenberg, M. G., Thornton, C. A., Griffith, J. D., and Swanson, M. S. (2007) Muscleblind-like 1 interacts with RNA hairpins in splicing target and pathogenic RNAs, *Nucleic Acids Res* 35, 5474-5486.
52. Michalowski, S., Miller, J. W., Urbinati, C. R., Paliouras, M., Swanson, M. S., and Griffith, J. (1999) Visualization of double-stranded RNAs from the myotonic dystrophy protein kinase gene and interactions with CUG-binding protein, *Nucleic Acids Res* 27, 3534-3542.
53. Ellington, A. D., and Szostak, J. W. (1990) In vitro selection of RNA molecules that bind specific ligands, *Nature* 346, 818-822.
54. Robertson, D. L., and Joyce, G. F. (1990) Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA, *Nature* 344, 467-468.
55. Kulbachinskiy, A. V. (2007) Methods for selection of aptamers to protein targets, *Biochemistry (Mosc)* 72, 1505-1518.
56. Garrey, S. M., Voelker, R., and Berglund, J. A. (2006) An extended RNA binding site for the yeast branch point-binding protein and the role of its zinc knuckle domains in RNA binding, *J Biol Chem* 281, 27443-27453.
57. Hirao, I., Harada, Y., Nojima, T., Osawa, Y., Masaki, H., and Yokoyama, S. (2004) In vitro selection of RNA aptamers that bind to colicin E3 and structurally resemble the decoding site of 16S ribosomal RNA, *Biochemistry* 43, 3214-3221.
58. Bartel, D. P., Zapp, M. L., Green, M. R., and Szostak, J. W. (1991) HIV-1 Rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA, *Cell* 67, 529-536.
59. Conrad, R. C., Bruck, F. M., Bell, S., and Ellington, A. D. (1998) *RNA:Protein Interactions*, Oxford University Press, Oxford.

CHAPTER II

1. Begemann, G., Paricio, N., Artero, R., Kiss, I., Perez-Alonso, M., and Mlodzik, M. (1997) Muscleblind, a gene required for photoreceptor differentiation in *Drosophila*, encodes novel nuclear Cys3His-type zinc-finger-containing proteins, *Development* 124, 4321-4331.

2. Artero, R., Prokop, A., Paricio, N., Begemann, G., Pueyo, I., Mlodzik, M., Perez-Alonso, M., and Baylies, M. K. (1998) The muscleblind gene participates in the organization of Z-bands and epidermal attachments of *Drosophila* muscles and is regulated by Dmef2, *Dev. Biol.* 195, 131-143.
3. Machuca-Tzili, L., Thorpe, H., Robinson, T. E., Sewry, C., and Brook, J. D. (2006) Flies deficient in Muscleblind protein model features of myotonic dystrophy with altered splice forms of Z-band associated transcripts, *Hum. Genet.* 120, 487-499.
4. Vicente, M., Monferrer, L., Poulos, M. G., Houseley, J., Monckton, D. G., O'Dell K, M., Swanson, M. S., and Artero, R. D. (2007) Muscleblind isoforms are functionally distinct and regulate alpha-actinin splicing, *Differentiation* 75, 427-440.
5. Ranum, L. P., and Cooper, T. A. (2006) RNA-Mediated Neuromuscular Disorders, *Annu Rev Neurosci.*
6. Osborne, R. J., and Thornton, C. A. (2006) RNA-dominant diseases, *Hum Mol Genet* 15 Spec No 2, R162-169.
7. Brook, J. D., Shaw, D. J., Meredith, L., Bruns, G. A., and Harper, P. S. (1984) Localisation of genetic markers and orientation of the linkage group on chromosome 19, *Hum Genet* 68, 282-285.
8. Fu, Y. H., Pizzuti, A., Fenwick, R. G., Jr., King, J., Rajnarayan, S., Dunne, P. W., Dubel, J., Nasser, G. A., Ashizawa, T., de Jong, P., and et al. (1992) An unstable triplet repeat in a gene related to myotonic muscular dystrophy, *Science* 255, 1256-1258.
9. Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barcelo, J., O'Hoy, K., and et al. (1992) Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene, *Science* 255, 1253-1255.
10. Liquori, C. L., Ricker, K., Moseley, M. L., Jacobsen, J. F., Kress, W., Naylor, S. L., Day, J. W., and Ranum, L. P. (2001) Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9, *Science* 293, 864-867.
11. Napierala, M., and Krzyzosiak, W. J. (1997) CUG repeats present in myotonin kinase RNA form metastable "slippery" hairpins, *J Biol Chem* 272, 31079-31085.

12. Dere, R., Napierala, M., Ranum, L. P., and Wells, R. D. (2004) Hairpin structure-forming propensity of the (CCTG.CAGG) tetranucleotide repeats contributes to the genetic instability associated with myotonic dystrophy type 2, *J Biol Chem* 279, 41715-41726.
13. Mooers, B. H., Logue, J. S., and Berglund, J. A. (2005) The structural basis of myotonic dystrophy from the crystal structure of CUG repeats, *Proc. Natl. Acad. Sci. U S A* 102, 16626-16631.
14. Fardaei, M., Rogers, M. T., Thorpe, H. M., Larkin, K., Hamshere, M. G., Harper, P. S., and Brook, J. D. (2002) Three proteins, MBNL, MBLL and MBXL, co-localize in vivo with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells, *Hum Mol Genet* 11, 805-814.
15. Mankodi, A., Urbinati, C. R., Yuan, Q. P., Moxley, R. T., Sansone, V., Krym, M., Henderson, D., Schalling, M., Swanson, M. S., and Thornton, C. A. (2001) Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2, *Hum. Mol. Genet.* 10, 2165-2170.
16. Miller, J. W., Urbinati, C. R., Teng-Umnuay, P., Stenberg, M. G., Byrne, B. J., Thornton, C. A., and Swanson, M. S. (2000) Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy, *Embo J* 19, 4439-4448.
17. Ho, T. H., Charlet, B. N., Poulos, M. G., Singh, G., Swanson, M. S., and Cooper, T. A. (2004) Muscleblind proteins regulate alternative splicing, *Embo. J.* 23, 3103-3112.
18. Cooper, T. A., and Ordahl, C. P. (1984) A single troponin T gene regulated by different programs in cardiac and skeletal muscle development, *Science* 226, 979-982.
19. Briggs, M. M., and Schachat, F. (1993) Origin of fetal troponin T: developmentally regulated splicing of a new exon in the fast troponin T gene, *Dev Biol* 158, 503-509.
20. Philips, A. V., Timchenko, L. T., and Cooper, T. A. (1998) Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy, *Science* 280, 737-741.
21. Ryan, K. J., and Cooper, T. A. (1996) Muscle-specific splicing enhancers regulate inclusion of the cardiac troponin T alternative exon in embryonic skeletal muscle, *Mol Cell Biol* 16, 4014-4023.

22. Warf, M. B., and Berglund, J. A. (2007) MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T, *RNA Oct 17*; [Epub ahead of print].
23. Yuan, Y., Compton, S. A., Sobczak, K., Stenberg, M. G., Thornton, C. A., Griffith, J. D., and Swanson, M. S. (2007) Muscleblind-like 1 interacts with RNA hairpins in splicing target and pathogenic RNAs, *Nucleic Acids Res* 35, 5474-5486.
24. Houseley, J. M., Wang, Z., Brock, G. J., Soloway, J., Artero, R., Perez-Alonso, M., O'Dell, K. M., and Monckton, D. G. (2005) Myotonic dystrophy associated expanded CUG repeat muscleblind positive ribonuclear foci are not toxic to *Drosophila*, *Hum. Mol. Genet.* 14, 873-883.
25. de Haro, M., Al-Ramahi, I., De Gouyon, B., Ukani, L., Rosa, A., Faustino, N. A., Ashizawa, T., Cooper, T. A., and Botas, J. (2006) MBNL1 and CUGBP1 modify expanded CUG-induced toxicity in a *Drosophila* model of myotonic dystrophy type 1, *Hum. Mol. Genet.* 15, 2138-2145.
26. Monferrer, L., and Artero, R. (2006) An interspecific functional complementation test in *Drosophila* for introductory genetics laboratory courses, *J. Hered.* 97, 67-73.
27. Conrad, R. C., Bruck, F. M., Bell, S., and Ellington, A. D. (1998) In vitro selection of nucleic acid ligands, in *RNA:Protein Interactions* (Smith, C. W. J., Ed.), pp 285-325, Oxford University Press, New York.
28. Voelker, R. B., and Berglund, J. A. (2007) A comprehensive computational characterization of conserved mammalian intronic sequences reveals conserved motifs associated with constitutive and alternative splicing, *Genome Res* 17, 1023-1033.
29. Tuerk, C., Gauss, P., Thermes, C., Groebe, D. R., Gayle, M., Guild, N., Stormo, G., d'Aubenton-Carafa, Y., Uhlenbeck, O. C., Tinoco, I., Jr., and et al. (1988) CUUCGG hairpins: extraordinarily stable RNA secondary structures associated with various biochemical processes, *Proc Natl Acad Sci U S A* 85, 1364-1368.
30. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction, *Nucleic Acids Res* 31, 3406-3415.
31. Hall, T. M. (2005) Multiple modes of RNA recognition by zinc finger proteins, *Curr Opin Struct Biol* 15, 367-373.

32. Brown, R. S. (2005) Zinc finger proteins: getting a grip on RNA, *Curr Opin Struct Biol* 15, 94-98.
33. Hagan, N. A., and Fabris, D. (2007) Dissecting the protein-RNA and RNA-RNA interactions in the nucleocapsid-mediated dimerization and isomerization of HIV-1 stemloop 1, *J Mol Biol* 365, 396-410.
34. Lawrence, D. C., Stover, C. C., Noznitsky, J., Wu, Z., and Summers, M. F. (2003) Structure of the intact stem and bulge of HIV-1 Psi-RNA stem-loop SL1, *J Mol Biol* 326, 529-542.
35. De Guzman, R. N., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N., and Summers, M. F. (1998) Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element, *Science* 279, 384-388.
36. Pascual, M., Vicente, M., Monferrer, L., and Artero, R. (2006) The Muscleblind family of proteins: an emerging class of regulators of developmentally programmed alternative splicing, *Differentiation* 74, 65-80.

CHAPTER III

1. Johnson, J. M., Castle, J., Garrett-Engle, P., Kan, Z., Loerch, P. M., Armour, C. D., Santos, R., Schadt, E. E., Stoughton, R., and Shoemaker, D. D. (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays, *Science* 302, 2141-2144.
2. Ule, J., Jensen, K. B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R. B. (2003) CLIP identifies Nova-regulated RNA networks in the brain, *Science* 302, 1212-1215.
3. Ule, J., Stefani, G., Mele, A., Ruggiu, M., Wang, X., Taneri, B., Gaasterland, T., Blencowe, B. J., and Darnell, R. B. (2006) An RNA map predicting Nova-dependent splicing regulation, *Nature* 444, 580-586.
4. Zhang, C., Zhang, Z., Castle, J., Sun, S., Johnson, J., Krainer, A. R., and Zhang, M. Q. (2008) Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2, *Genes Dev* 22, 2550-2563.
5. Ladd, A. N., Stenberg, M. G., Swanson, M. S., and Cooper, T. A. (2005) Dynamic balance between activation and repression regulates pre-mRNA alternative splicing during heart development, *Dev Dyn* 233, 783-793.
6. Warf, M. B., and Berglund, J. A. (2007) MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T, *RNA* 13, 2238-2251.

7. Kino, Y., Mori, D., Oma, Y., Takeshita, Y., Sasagawa, N., and Ishiura, S. (2004) Muscleblind protein, MBNL1/EXP, binds specifically to CHHG repeats, *Hum Mol Genet* 13, 495-507.
8. Pascual, M., Vicente, M., Monferrer, L., and Artero, R. (2006) The Muscleblind family of proteins: an emerging class of regulators of developmentally programmed alternative splicing, *Differentiation* 74, 65-80.
9. Fardaei, M., Larkin, K., Brook, J. D., and Hamshere, M. G. (2001) In vivo co-localisation of MBNL protein with DMPK expanded-repeat transcripts, *Nucleic Acids Res* 29, 2766-2771.
10. Osborne, R. J., and Thornton, C. A. (2006) RNA-dominant diseases, *Hum Mol Genet* 15 Spec No 2, R162-169.
11. Charlet, B. N., Savkur, R. S., Singh, G., Philips, A. V., Grice, E. A., and Cooper, T. A. (2002) Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing, *Mol Cell* 10, 45-53.
12. Mankodi, A., Takahashi, M. P., Jiang, H., Beck, C. L., Bowers, W. J., Moxley, R. T., Cannon, S. C., and Thornton, C. A. (2002) Expanded CUG repeats trigger aberrant splicing of ClC-1 chloride channel pre-mRNA and hyperexcitability of skeletal muscle in myotonic dystrophy, *Mol Cell* 10, 35-44.
13. Savkur, R. S., Philips, A. V., and Cooper, T. A. (2001) Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy, *Nat Genet* 29, 40-47.
14. Mooers, B. H., Logue, J. S., and Berglund, J. A. (2005) The structural basis of myotonic dystrophy from the crystal structure of CUG repeats, *Proc Natl Acad Sci U S A* 102, 16626-16631.
15. Napierala, M., and Krzyzosiak, W. J. (1997) CUG repeats present in myotonin kinase RNA form metastable "slippery" hairpins, *J Biol Chem* 272, 31079-31085.
16. Ho, T. H., Charlet, B. N., Poulos, M. G., Singh, G., Swanson, M. S., and Cooper, T. A. (2004) Muscleblind proteins regulate alternative splicing, *Embo J* 23, 3103-3112.
17. Philips, A. V., Timchenko, L. T., and Cooper, T. A. (1998) Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy, *Science* 280, 737-741.
18. Ranum, L. P., and Cooper, T. A. (2006) RNA-mediated neuromuscular disorders, *Annu Rev Neurosci* 29, 259-277.

19. Conrad, R. C., Bruck, F. M., Bell, S., and Ellington, A. D. (1998) In vitro selection of nucleic acid ligands, in *RNA:Protein Interactions* (Smith, C. W. J., Ed.), pp 285-325, Oxford University Press, New York.
20. Clamp, M., Cuff, J., Searle, S. M., and Barton, G. J. (2004) The Jalview Java alignment editor, *Bioinformatics* 20, 426-427.
21. Jensen, K. B., Musunuru, K., Lewis, H. A., Burley, S. K., and Darnell, R. B. (2000) The tetranucleotide UCAY directs the specific recognition of RNA by the Nova K-homology 3 domain, *Proc Natl Acad Sci U S A* 97, 5740-5745.
22. Jin, Y., Suzuki, H., Maegawa, S., Endo, H., Sugano, S., Hashimoto, K., Yasuda, K., and Inoue, K. (2003) A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG, *Embo J* 22, 905-912.
23. Kimura, T., Nakamori, M., Lueck, J. D., Pouliquin, P., Aoike, F., Fujimura, H., Dirksen, R. T., Takahashi, M. P., Dulhunty, A. F., and Sakoda, S. (2005) Altered mRNA splicing of the skeletal muscle ryanodine receptor and sarcoplasmic/endoplasmic reticulum Ca²⁺-ATPase in myotonic dystrophy type 1, *Hum Mol Genet* 14, 2189-2200.
24. Lin, X., Miller, J. W., Mankodi, A., Kanadia, R. N., Yuan, Y., Moxley, R. T., Swanson, M. S., and Thornton, C. A. (2006) Failure of MBNL1-dependent post-natal splicing transitions in myotonic dystrophy, *Hum Mol Genet* 15, 2087-2097.
25. Eng, F. J., and Warner, J. R. (1991) Structural basis for the regulation of splicing of a yeast messenger RNA, *Cell* 65, 797-804.

CHAPTER IV

1. Ho, T. H., Charlet, B. N., Poulos, M. G., Singh, G., Swanson, M. S., and Cooper, T. A. (2004) Muscleblind proteins regulate alternative splicing, *Embo J* 23, 3103-3112.
2. Warf, M. B., and Berglund, J. A. (2007) MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T, *RNA* 13, 2238-2251.
3. Fardaei, M., Larkin, K., Brook, J. D., and Hamshere, M. G. (2001) In vivo co-localisation of MBNL protein with DMPK expanded-repeat transcripts, *Nucleic Acids Res* 29, 2766-2771.
4. Mankodi, A., Urbinati, C. R., Yuan, Q. P., Moxley, R. T., Sansone, V., Krym, M., Henderson, D., Schalling, M., Swanson, M. S., and Thornton, C. A. (2001) Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2, *Hum Mol Genet* 10, 2165-2170.

5. Osborne, R. J., and Thornton, C. A. (2006) RNA-dominant diseases, *Hum Mol Genet 15 Spec No 2*, R162-169.
6. Charlet, B. N., Savkur, R. S., Singh, G., Philips, A. V., Grice, E. A., and Cooper, T. A. (2002) Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing, *Mol Cell 10*, 45-53.
7. Savkur, R. S., Philips, A. V., and Cooper, T. A. (2001) Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy, *Nat Genet 29*, 40-47.
8. Pascual, M., Vicente, M., Monferrer, L., and Artero, R. (2006) The Muscleblind family of proteins: an emerging class of regulators of developmentally programmed alternative splicing, *Differentiation 74*, 65-80.
9. Kino, Y., Mori, D., Oma, Y., Takeshita, Y., Sasagawa, N., and Ishiura, S. (2004) Muscleblind protein, MBNL1/EXP, binds specifically to CHHG repeats, *Hum Mol Genet 13*, 495-507.
10. Miller, J. W., Urbinati, C. R., Teng-Umnuay, P., Stenberg, M. G., Byrne, B. J., Thornton, C. A., and Swanson, M. S. (2000) Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy, *Embo J 19*, 4439-4448.
11. Napierala, M., and Krzyzosiak, W. J. (1997) CUG repeats present in myotonin kinase RNA form metastable "slippery" hairpins, *J Biol Chem 272*, 31079-31085.
12. Machuca-Tzili, L., Thorpe, H., Robinson, T. E., Sewry, C., and Brook, J. D. (2006) Flies deficient in Muscleblind protein model features of myotonic dystrophy with altered splice forms of Z-band associated transcripts, *Hum Genet 120*, 487-499.
13. Vicente-Crespo, M., Pascual, M., Fernandez-Costa, J. M., Garcia-Lopez, A., Monferrer, L., Miranda, M. E., Zhou, L., and Artero, R. D. (2008) Drosophila muscleblind is involved in troponin T alternative splicing and apoptosis, *PLoS ONE 3*, e1613.
14. de Haro, M., Al-Ramahi, I., De Gouyon, B., Ukani, L., Rosa, A., Faustino, N. A., Ashizawa, T., Cooper, T. A., and Botas, J. (2006) MBNL1 and CUGBP1 modify expanded CUG-induced toxicity in a Drosophila model of myotonic dystrophy type 1, *Hum Mol Genet 15*, 2138-2145.

15. Vicente, M., Monferrer, L., Poulos, M. G., Houseley, J., Monckton, D. G., O'Dell K, M., Swanson, M. S., and Artero, R. D. (2007) Muscleblind isoforms are functionally distinct and regulate alpha-actinin splicing, *Differentiation* 75, 427-440.
16. Lin, X., Miller, J. W., Mankodi, A., Kanadia, R. N., Yuan, Y., Moxley, R. T., Swanson, M. S., and Thornton, C. A. (2006) Failure of MBNL1-dependent post-natal splicing transitions in myotonic dystrophy, *Hum Mol Genet* 15, 2087-2097.
17. Kimura, T., Nakamori, M., Lueck, J. D., Pouliquin, P., Aoike, F., Fujimura, H., Dirksen, R. T., Takahashi, M. P., Dulhunty, A. F., and Sakoda, S. (2005) Altered mRNA splicing of the skeletal muscle ryanodine receptor and sarcoplasmic/endoplasmic reticulum Ca²⁺-ATPase in myotonic dystrophy type 1, *Hum Mol Genet* 14, 2189-2200.