AUTO-REGULATION OF THE *MBNL1* PRE-MRNA

by

DEVIKA P. GATES

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

June 2011

DISSERTATION APPROVAL PAGE

Student: Devika P. Gates

Title: Auto-regulation of the *MBNL1* Pre-mRNA

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	Chair
J. Andrew Berglund	Advisor
Victoria J. de Rose	Member
Alice Barkan	Member
Karen Guillemin	Outside Member

and

Richard Linton	Vice President for Research and Graduate Studies/Dean of the Graduate School

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

Degree awarded June 2011

DISSERTATION ABSTRACT

Devika Pooja Rajen Gates

Doctor of Philosophy

Department of Chemistry

June 2011

Title: Auto-Regulation of the *MBNL1* Pre-mRNA

Approved:_____
Dr. J. Andrew Berglund

Muscleblind-like 1 (MBNL1) is a splicing factor whose improper cellular localization is a central component of myotonic dystrophy (DM). In DM, the lack of properly localized MBNL1 leads to mis-splicing of many pre-mRNAs. The mechanism by which MBNL1 regulates it pre-mRNA targets is not well understood. In order to determine the mechanism by which MBNL1 regulates alternative splicing, a consensus RNA binding motif for Mbl (the *Drosophila* ortholog of MBNL1) and MBNL1 were determined using SELEX (Systematic Evolution of Ligands by Exponential Enrichment). These consensus motifs allowed for the identification of high affinity endogenous sites within pre-mRNAs that are regulated by MBNL1. *In vitro* binding studies showed that MBNL1 bound to RNAs that contained the consensus motif surrounded by pyrimidines.

Some of these sites were identified upstream of exon 5 within the *MBNL1* pre-mRNA, and we have shown that MBNL1 auto-regulates the exclusion of exon 5 in HeLa cells. The region of the *MBNL1* gene that includes exon 5 and flanking intronic sequence is highly conserved in vertebrate genomes. The 3' end of intron 4 is non-canonical in that it contains an AAG 3' splice site and a predicted branchpoint that is 141 nucleotides

from the 3' splice site. Using a mini-gene that includes exon 4, intron 4, exon 5, intron 5 and exon 6 of *MBNL1*, we show that MBNL1 regulates inclusion of exon 5. Mapping of the intron 4 branchpoint confirms that branching occurs primarily at the predicted distant branchpoint. Structure probing and footprinting reveal that the highly conserved region between the branchpoint and the 3' splice site is primarily unstructured, and MBNL1 binds within this region of the pre-mRNA, which we have termed the MBNL1 response element. Deletion of the response element eliminates MBNL1 splicing regulation and leads to complete inclusion of exon 5, which is consistent with the suppressive effect of MBNL1 on splicing.

This dissertation includes previously published co-authored material as well as my recent co-authored material that has been submitted for publication.

CURRICULUM VITAE

NAME OF AUTHOR: Devika P. Gates

GRADUATE AND UNDERGRADUATE SCHOOLS:

University of Oregon, Eugene
University of Washington, Seattle

DEGREES AWARDED:

Doctor of Philosophy, Chemistry, 2011, University of Oregon
Bachelor of Science, Biochemistry, 2006, University of Washington

AREAS OF SPECIAL INTEREST:

Chemistry
Biochemistry
Molecular Biology

PROFESSIONAL EXPERIENCE:

General Lab Technician, Dr. Sharon Doty, April 2003-June 2003

Undergraduate Research Assistant, Dr. Gabrielle Varani, 2004-2006

Graduate Student, Dr. J. Andrew Berglund, 2006-2011

Training Internship Research and Development, Pro-Unlimited at Life Technologies, August 2010-December 2010

GRANTS, AWARDS, AND HONORS:

KECK fellow for graduate studies, 2007- 2009

Molecular biology and biophysics training grant appointee, 2009- 2010

KECK fellow for graduate studies, 2011-current

PUBLICATIONS:

Gates, D.P., Coonrod, L.A., Berglund, J.A. *Auto-regulated splicing of the MBNL1 pre-mRNA*. Submitted to JBC. 2011

Cass, D., Hotchko, R., Barber, P., Jones, K., Gates, D.P., Berglund, J.A. *The four Zn fingers of MBNL1 provide a flexible platform for recognition of its RNA binding elements*. BMC Molecular Biology, 2011. **12**: p. 20-26.

Goers, E.S., Purcell, J., Voelker, R.B., Gates, D.P., Berglund, J.A. *MBNL1 binds GC motifs embedded in pyrimidines to regulate alternative splicing.* Nucleic Acids Res, 2010. **38**(7): p. 2467-84.

Goers, E.S., Voelker, R.B., Gates, D.P., Berglund, J.A. *RNA binding specificity of Drosophila muscleblind*. Biochemistry, 2008. **47**(27): p. 7284-94.

ACKNOWLDEDGEMENTS

I would like to thank Dr. Andrew Berglund for all of his advice and mentoring throughout my time as a graduate student in his laboratory. He taught me to think critically and allowed me to develop my independence while providing the perfect amount of support and advice. I have grown a lot scientifically in the last five years and want to thank you for all that you've done to make me the scientist that I am.

I would also like to thank everyone in the Berglund lab (past and present members) who guided me through troubled times when experiments wouldn't work and made me feel like a rock star when experiments did work. The Berglund lab has always functioned like a loving family and it made coming to work every day a wonderful and fun experience.

I would also like to thank all of my committee members, your support and the advice allowed me to get the most out of the graduate school experience. You gently urged me to take certain classes, which aided in the progress of my research and also make me a successful graduate student.

I would like to thank the kitchen staff for doing all the dishes, making all the buffers and LB plates. The time I saved from not washing dishes or making buffers, I was able to spend on my research. Thank you for making my life easier over the last five years.

I would like to thank Dan and Nigel for constantly fixing things in the lab when they were broken. Thank you especially for the time you extracted all the data from my hard drive when my computer stopped working.

I would like to thank my husband for always listening to me talk about my research and providing helpful insight and advice. I will never forget all of those times you came to work with me at 4 am just so I could take something out of the incubator or stop a gel. I will also never forget all those times you would hang out in the lab with me so that we could spend time together in between experiments on weeknights and during the weekend. I am so lucky to have married my best friend.

I would like to thank my family for all of their support over the last five years. I'll always remember how understanding you all were when I had to run into lab because I had an experiment running whenever you guys visited. I want to also thank my Dad who helped me dialyze my protein by pipetting BME into dialysis buffer.

I would also like to thank Anica, Leslie, Jennifer and all of my other friends I made in graduate school. I feel lucky to have spent time with such wonderful, supportive and intelligent people. Thank you for all of your advice, whether it was scientific or not scientific, I'll never forget our time together at graduate school. Anica, I could not have made it through graduate school without your friendship, advice and all of our coffee runs. I know that our friendship will be life long and I can't thank you enough.

This dissertation is dedicated to my loving husband, our daughter who will arrive in one month and my wonderful sister

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Alternative Splicing

Splicing of pre-mRNAs is an important event that contributes to a diverse proteome as well as the regulation of gene expression. It is estimated that more than 90% of human genes undergo alternative splicing [1, 2]. To produce a functional mRNA, introns (non-coding regions) must be accurately removed and the exons (coding regions) must be carefully ligated together to form a functional protein.

Each exon within a pre-mRNA can undergo a different type of splicing. For instance, most exons are constitutive which means they are included in the final mRNA. Cassette exons are regulated such that they are sometimes included or sometimes excluded. In pre-mRNAs with many cassette exons, you can have mutually exclusive splicing where if one exon is excluded, the other exon will be included. Additionally, alternative 5' or 3' splice sites can be used which alter the size of the exon. Differences in splice site choice or skipping of exons commonly results in different protein isoforms, thereby allowing one gene to produce many different proteins.

The removal of introns and ligation of exons relies heavily on special sequences within introns that are recognized by the splicing machinery. Important sequences include the 5' splice site (GURAGU), the branch point sequence (YURAY), the polypyrimidine tract (runs of pyrimidines) and the 3' splice site (YAG). Different proteins bind to these sequences and with the aid of U1, U2, U4, U5 and U6 snRNPs (small nuclear

ribonucleoproteins), splicing occurs (Figure 1). During the splicing process, different complexes form on the pre-mRNA to aid in splicing. In early (E) complex, the U1 snRNP is bound to the 5' splice site, U2AF35 is bound to the 3' splice site, SF1 is bound to the branchpoint sequence and U2AF65 is bound to the polypyrimidine (PY) tract. When U2 snRNP binds, it replaces SF1 and A complex is formed. U2AF65 and U2AF35 leave and the tri snRNP (U4/U5/U6) arrive to form B complex. In B complex the spliceosome undergoes a conformational change to form the C complex where now the U6 snRNP is bound to the 5' splice site (5' ss) and U1 and U4 snRNP are removed. Once C complex has formed, splicing can occur [3].

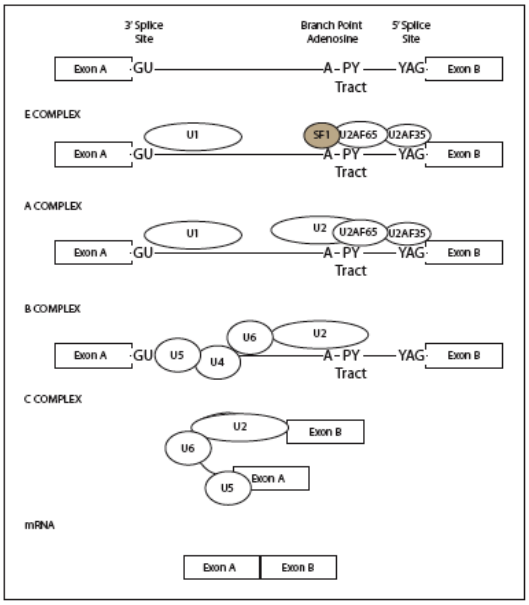

Figure 1. Splicing occurs via two transesterification reactions followed by a ligation. SF1 and U2AF65 bind the branch point and polypyrimidine tract, respectively. U2AF35 binds to the 3' splice site and the U1 snRNA binds the 3' splice site to make the E complex. In complex A, U2 snRNA takes the place of SF1 and ATP is hydrolyzed. In B complex, U4, U5 and U6 bind the intronic region and the U2 snRNA continues to bind the same area. In C complex the U2/U6 snRNA catalyzes the transesterification reactions. Finally the lariat is released and exons are ligated together by ATP hydrolysis.

Splicing occurs by two transesterification reactions followed by ligation of the coding regions [4]. In the first step, the 2'-hydroxyl group of the branchpoint adenosine attacks the phosphate at the 5' ss, thereby cleaving the 5' exon and ligating the 5' end of the intron to the branchpoint adenosine 2'-hydroxyl resulting in a lariat structure. The second transesterification step involves attacking the 3' end of the intron by the 3'-hydroxyl of the removed exon. This allows the two exons to ligate together and form the mature mRNA.

While the proteins involved in the spliceosome complex are very important, there are other proteins that aid in splicing called alternative splicing factors. Alternative splicing factors aid in the regulation of either inclusion or exclusion of an exon by assisting in splice site choice. Alternative splicing factors can do this using many different mechanisms. Muscleblind-like 1 (MBNL1) is one such splicing factor and in this thesis I address how it interacts with RNA and how it regulates the alternative splicing of its own pre-mRNA.

Myotonic Dystrophy and Alternative Splicing

Myotonic dystrophy (DM) is a dominantly inherited neuromuscular disorder and is the most common form of muscular dystrophy in adults. Two genes, and thus two genetic loci are associated with myotonic dystrophy, causing two different diseases with similar symptoms. DM patients show peculiar and seemingly un-related symptoms of which a few are: insulin resistance, cardiac abnormalities, myotonia and progressive muscle weakness. Myotonic dystrophy type 1 (DM1) is caused by a $(CUG)_n$ (where n is a number great than 50) expansion in the 3' untranslated region (UTR) of the dystrophia myotonica protein kinase (*DMPK*) gene [5]. Myotonic dystrophy type 2 (DM2) is caused

by a $(CCUG)_n$ expansion in the first intron of the Zinc Finger 9 (*ZNF9*) gene [6]. Individuals afflicted with DM1 have anywhere from 50 to greater than 4000 CUG repeats in the 3' UTR of the DMPK gene [6]. Larger repeat size correlates with increased severity of the disease as well as younger age of onset.

DM1 is caused by an RNA gain of function, where the CTG expansion is transcribed to CUG repeats. The RNA gain of function hypothesis was first suggested when the following was observed; the loss of function of the *DMPK* gene or surrounding genes didn't cause DM1 [7, 8], long CTG repeats that are transcribed to CUG accumulate in nuclear foci [9] and the observation that the 3' UTR of the *DMPK* gene containing 200 CTG repeats is enough to cause myotonia [10]. Evidence of the RNA gain of function model was provided by a mouse model containing 250 CTG repeats in the human skeletal *a-actin* gene. These mice developed myotonia and muscle histology similar to DM1 [5] suggesting that CUG repeats alone can induce DM1 like symptoms.

Long CUG repeats sequester a family of RNA binding proteins, the muscleblind-like (MBNL) proteins [11, 12]. The human proteins (MBNL1, MBNL2 and MBNL3) derived their name (muscleblind-like) from the founding member of this family, the *Drosophila* muscleblind gene (*Mbl*), which was shown to be important for muscle terminal differentiation and proper eye development [13, 14]. In 2003 Kanadia and colleagues determined that when the mouse *MBNL1* gene is disrupted, it results in muscle, eye and aberrant splicing that is characteristic of DM. In 2004 Ho and colleagues suggested that MBNL1 was an alternative splicing factor that regulates the alternative splicing of the transcripts observed to be mis-spliced in DM and they showed that MBNL1 and MBNL2 regulate the splicing of the *TNNT2* and *IR* pre-mRNAs, two

genes that are mis-spliced in DM1 [15, 16]. This observation provided more evidence for the RNA gain of function model, where the sequestration of MBNL1 to expanded CUG repeats causes the aberrant splicing of several MBNL1 regulated pre-mRNAs. Regulation of alternative splicing often results in the formation of an adult isoform, however, aberrant splicing caused by the sequestration of MBNL1 results in the expression of the neonatal isoform [17]. The pre-mRNA transcripts that are mis-spliced when MBNL1 is sequestered, correlate with disease symptoms. For example, the mis-splicing of the *TNNT2*, *CLCN1* and *SERCA* pre-mRNAs cause cardiac abnormalities, myotonia and the dis-regulation of intracellular skeletal muscle calcium homeostasis. The list of mis-spliced transcripts in DM is partially complete with 24 known transcripts [18].

Another key splicing factor involved in DM is CUG-BP1, which is a conserved heterogeneous nuclear ribonucleoprotein (hnRNP) and binds RNA that contains CUG repeats [19, 20]. MBNL1 and CUG-BP1 can behave in an antagonistic manner for some pre-mRNAs where if MBNL1 regulates the exclusion of an exon, CUG-BP1 regulates inclusion of that same exon [6]. CUG-BP1 protein levels are significantly increased in DM. It has been shown that the inclusion of TNNT2 exon 5 is also regulated by CUG-BP1, therefore the sequestration of MBNL1 and over-expression of CUG-BP1 both lead to the splicing defect observed in this transcript [15]. More recently it has been shown that MBNL1 and CUG-BP1 do not appear to co-regulate very many transcripts [21, 22].

The Muscleblind Family of Proteins

MBNL proteins bind pre-mRNA through conserved CCCH zinc fingers. Human MBNL homologs (MBNL1, MBNL2 and MBNL3) are regulators of alternative splicing

and help promote the inclusion or exclusion of specific exons. MBNL1 and MBNL3 are involved in muscle differentiation and MBNL2 is involved in integrin α3 subcellular localization [23-25]. Alternative splicing of MBNL1, MBNL2 and MBNL3 genes gives rise to nine, three and six protein isoforms [26, 27]. MBNL1 and MBNL2 are expressed in the brain, kidney, liver pancreas and muscle cells, however, MBNL1 is more abundant in the heart and skeletal muscle while MBNL2 levels are similar in all the tissues listed. MBNL3 is expressed at much lower levels in the tissues listed above but is found in high levels in the placenta [24, 26].

Drosophila contains one muscleblind gene (*Mbl*), which shows conservation to the human *MBNL* gene in the zinc fingers. Where MBNL proteins contain four zinc finger domains, Mbl proteins contain two zinc finger domains. It has been hypothesized that a duplication of the first two zinc fingers occurred before the divergence between urochordates and vertebrates, giving rise to the more modern MBNL proteins with four zinc fingers [11]. Mbl contains four splice isoforms (Mbl A, B, C and D) all of which include the two zinc finger domains. Mbl C is highly expressed throughout the embryo, larval and pupal stages, while Mbl B and A are expressed at lower levels. Mbl D is expressed at the end of the larval stage and the beginning of the pupal stage [28]. The expression profile of MBNL and Mbl proteins suggests that they are involved in a developmental process. The binding specificity of Mbl and MBNL are discussed in chapters II and III and contain previously published co-authored material.

MBNL1 Binding and Alternative Splicing of Pre-mRNA Transcripts

Teplova et al. showed that MBNL1 can bind Watson-Crick face of GC dinucleotides through its zinc fingers (Figure 2). The zinc fingers bind RNA by using intermolecular stacking and hydrogen bonding interactions. Zinc finger 3 specifically binds GC dinucleotides while Zinc finger 4 specifically binds GCU trinucleotides. Currently, it is hypothesized that MBNL1 binds RNA in a looped manner, where every zinc finger binds a GC dinucleotide [29].

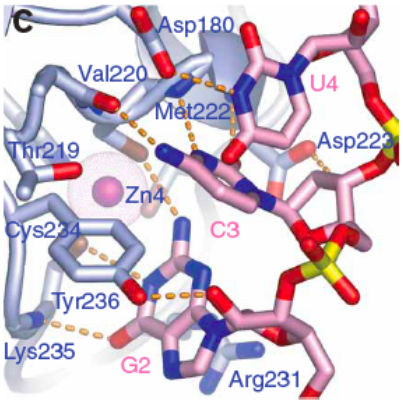

Figure 2. MBNL1 zinc finger 4 bound to single stranded RNA. A crystal structure published by Teplova et al. that shows the protein RNA interface of MBNL1 zinc finger 4 making intermolecular interactions with G2, C3 and U4.

Grammatikakis et al. found that the regions required for splicing regulation in MBNL1 and MBNL3 are separate from the zinc finger RNA binding domains [30]. They identified core regulatory regions for activation and repression 80 amino acids downstream of the N-terminal zinc-finger pair.

MBNL1 has been shown to regulate exon exclusion of exon 5 of the *TNNT2* and exon 7A of the *CLCN1* pre-mRNAs. In both cases MBNL1 directly binds the pre-mRNA. To regulate exon 5 exclusion of the *TNNT2* pre-mRNA, MBNL1 competes with the splicing factor U2AF65 to regulate exon 5 [31]. To regulate exon exclusion of 7A in

the *CLCN1* pre-mRNA MBNL1 binds the 5' end of exon 7A (thereby masking an exonic splicing enhancer) and the upstream intronic region [27].

To determine how MBNL1 identifies and regulates its pre-mRNA targets, a study on the RNA binding specificity of *Drosophila* Mbl and MBNL1 was done. Chapter II discusses the binding specificity of Mbl to SELEX RNAs. The studies from this chapter contains co-authored material by Emily Goers, Rodger Voelker and Devika Gates. The results from the studies discussed in Chapter II indicate that Mbl binds RNAs with secondary structure.

Chapter III discusses the identification of a YGCY consensus motif via a doped SELEX that MBNL1 binds with high affinity. Chapter III contains co-authored material by Emily Goers, Jamie Purcell, Rodger Voelker and Devika Gates. The studies in this chapter allowed for the identification of novel endogenous MBNL1 binding sites.

The studies from Chapter II and Chapter III aided in the identification a 90 nucleotide MBNL1 response element within the *MBNL1* pre-mRNA, which is discussed in Chapter IV. Chapter IV contains co-authored material by Devika Gates and Leslie Coonrod and contains results that provide insight into how MBNL1 regulates the alternative splicing of its own pre-mRNA.

CHAPTER II

RNA BINDING SPECIFICITY OF *DROSOPHILA* MUSCLEBLIND

The work described in this chapter was published in volume 47 of the journal Biochemistry in June 2008. Emily Goers performed the RNA SELEX experiment with *Drosophila* Mbl and performed the characterization of many of the RNA ligands; Rodger Voelker performed bioinformatics analysis; I aided in the characterization of the SELEX RNAs and determining their binding affinity to Mbl. Dr. J. Andrew Berglund was the principle investigator of this work.

Introduction

The human muscleblind protein, MBNL, plays a big role in DM. There are two types of DM, DM1 and DM2. DM1 is caused by a CTG expansion in the 3' UTR region of *DMPK1* gene [32-34] while DM2 is caused by a CCTG repeat expansion in the first intron of *ZNF9* gene [35]. These two different expansions in two different genes result in a disease with a common mechanism. When the repeat expansions are transcribed, they form a double-stranded stem loop structure. MBNL proteins are sequestered to these structures forming nuclear foci, which results in the aberrant splicing of its pre-mRNA targets [23, 26].

Drosophila has been used as a model organism for understanding myotonic dystrophy (DM). It has been shown that *Drosophila* overexpressing 162 CUG repeats result in Mbl sequestered into nuclear foci but no abnormal phenotypes are observed [36]. However, the introduction of 480 interrupted CUG repeats result in not only the nuclear foci but causes symptoms typical of DM in the muscle and eye as well as other degenerative phenotypes [37]. These phenotypes can be rescued with the overexpression

of human MBNL1 indicating that Mbl and MBNL1 are functionally conserved. The *Drosophila* Mbl protein is an excellent model to understand how the muscleblind proteins recognize RNA because Drosophila has only one *Mbl* gene (which has four protein isoforms) and only one zinc finger pair. Here we characterize the RNA binding properties of Mbl.

Results

Characterization of SELEX RNAs- SELEX was done by Emily Goers with Mbl using a library of 10^{14} RNAs, which contained a 30 nucleotide randomized region with flanking constant regions. After six rounds of SELEX, 25 unique RNA sequences were identified and they were classified into three groups depending on their binding affinity to Mbl, their secondary structure and the presence of enriched motifs.

The binding affinity of SELEX RNAs 20, 22, 24, 28, 30, 34 and other RNAs to Mbl was determined using electrophoretic mobility shift assays (EMSA (Figure 3)).

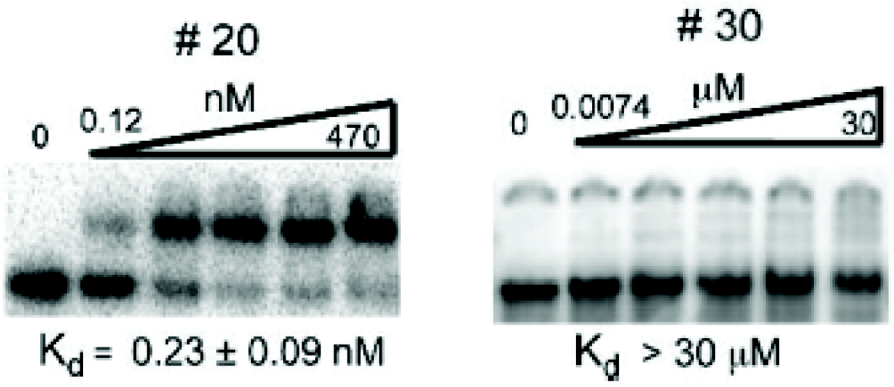

Figure 3. An example of the electrophoretic mobility shift assay (EMSA) used to characterize the SELEX RNAs. In this example, RNA #20 bound Mbl with high affinity while RNA #30 didn't bind Mbl.

Characterization of these and other RNAs allowed us to group the SELEX RNAs. Group I RNAs contained the AGUCU consensus motif near the 3' end of the RNA and

Group II RNAs contained the AGUCU consensus motif near the 5' end of the RNA. Both Group I and Group II RNAs bound Mbl with a $K_d \leq 50nM$. Group 3 RNAs did not contain the AGUCU consensus motif and bound Mbl poorly with Kd values > 300nM (Figure 4). Interestingly, Group I RNAs had a common secondary structure while Group II RNAs had a common secondary structure. This was the result of the AGUCU consensus motif binding a sequence in the 3' constant region. This observation suggests that the SELEX on Mbl may have been biased and the AGUCU consensus motif may not be biologically relevant.

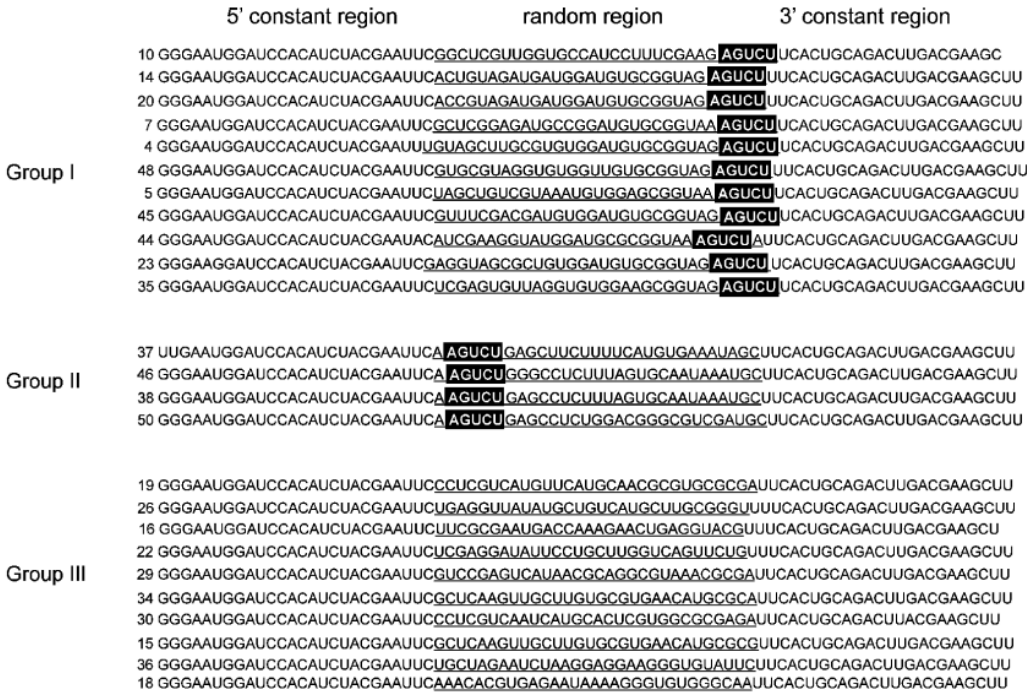

Figure 4. SELEX RNAs classified into three groups. SELEX RNAs were put into separate groups depending on whether they contained a consensus motif (highlighted), where the motif was located and the predicted secondary structure.

Binding studies with SELEX RNA #20 show Mbl binds structure RNA- SELEX RNA #20 was studied in more detail to identify a minimal Mbl binding site and to determine whether structure played a role in Mbl binding of the RNA (Figure 5). RNA

#20 contains two stem-loops separated by a short single stranded linker. In order to determine whether Mbl bound the individual stem loops, gel shift assays were done. We observed, that Mbl did not bind stem-loop 1 and bound stem-loop 2 with a K_d of 290nM, suggesting that Mbl requires both stem-loops to bind. However, if stem-loop 1 was shortened, Mbl still bound with a K_d of 0.31nM. Interestingly, RNAs where the 5' tail and 3' tail were truncated separately did not bind Mbl indicating the importance of these regions for RNA binding.

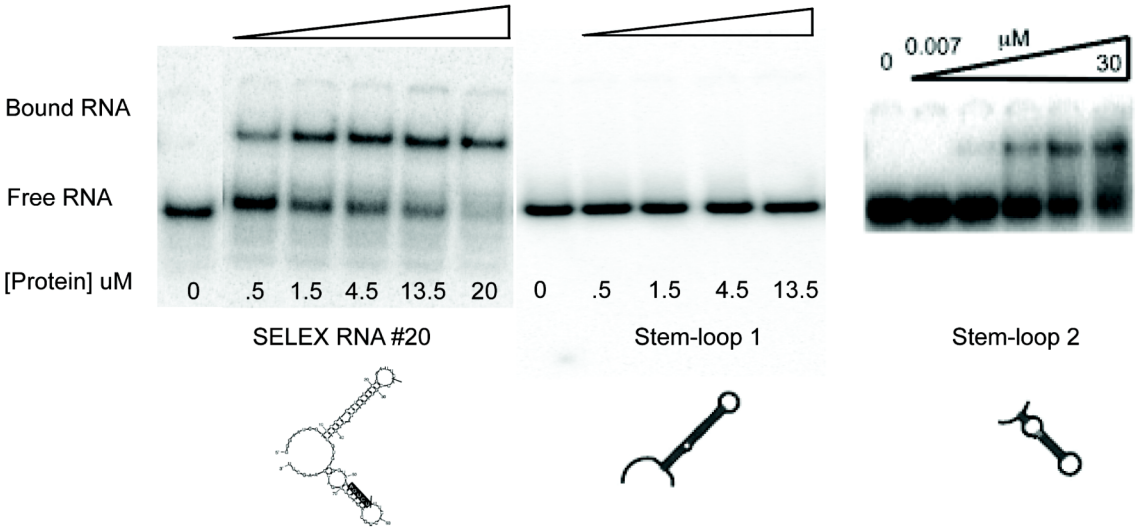

Figure 5. SELEX RNA # 20 binds Mbl with high affinity. To determine whether Mbl binds a particular stem-loop, EMSAs were done with stem-loop 1 and stem-loop 2.

Materials and Methods

Protein Purification. This Mbl construct was expressed in *Escherichia coli* Rosetta cells and grown in LB medium with 50 μM ZnCl2, 50 μg/mL Amp, 25 μg/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 31000*g* for 15 min. The supernatant was

bound to glutathione-agarose beads for 30 min to 1 h 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.

Transcription and Kinase Reactions. RNA oligonucleotides were transcribed with T7 RNA polymerase and [α-32P]CTP from linearized plasmids, PCR products, or fully complementary cDNA oligos. The transcription reaction mixtures were incubated at 37 °C for 3 h or at 25 °C overnight. RNAs were kinased using T4 polynucleotide kinase (New England Biolabs) and [γ-32P]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 μL per binding reaction mixture at 25 °C. RNA in binding buffer (8 μL) 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 μL 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.

Discussion

SELEX allowed for the identification of RNAs that bound Mbl with high affinity, and the consensus motif AGUCU. It is clear from the characterization of several SELEX RNAs that the AGUCU motif binds the constant region and aids in the creation of secondary structure for each RNA. Characterization of the SELEX RNAs using gel shift assays revealed that the RNAs bind in the picomolar to nanomolar range.

Further characterization of SELEX RNA #20 showed that it contained two stem-loops and that both stem-loops are important for Mbl binding. The 5' and 3' tails of the RNA were also shown to be important for binding although it is unclear what role the tails have in binding. It is possible that they could be involved in tertiary interactions with other regions of the RNA.

The AGUCU motif that was enriched in Group I and Group II RNAs did not seem to be a result of Mbl binding but due to the fact that this motif bound the constant region. However, we were able to show that Mbl binds structured RNA with high affinity. A SELEX on MBNL1 will likely yield a different consensus motif.

CHAPTER III

MBNL1 BINDS RNAS THAT CONTAIN A HIGH PERCENTAGE OF PYRIMIDINES AND YGCY MOTIFS

This work was published in volume 38 of the journal Nucleic Acids Research in January 2010. Emily Goers initially identified and performed preliminary expression analysis; Jamie Purcell performed the splicing assays; Rodger Voelker performed bioinformatics analysis; I aided in the characterization of the SELEX RNAs and potential endogenous MBNL1 binding sites. Dr. J. Andrew Berglund was the principle investigator of this work.

Introduction

Myotonic dystrophy (DM) is a dominantly inherited disorder and is the most common form of muscular dystrophy in adults. Two genes, and thus two genetic loci are associated with myotonic dystrophy, causing two different diseases with similar symptoms. DM patients show peculiar and seemingly non-related symptoms of which a few are: insulin resistance, cardiac abnormalities, myotonia and progressive muscle weakness. Myotonic dystrophy type 1 (DM1) is caused by a $(CTG)_n$ (where n is a number great than 50) expansion in the 3' UTR of the *DMPK* gene. Myotonic dystrophy type 2 (DM2) is caused by a $(CCTG)_n$ expansion in the first intron of the *ZNF9* gene. Individuals afflicted with DM1 have anywhere from 50 to greater than 4000 CUG repeats in the 3' UTR of the *DMPK* gene [6]. The toxic CUG and CCUG expansions causes a gain of function disease on the RNA level by forming stable stem loops that sequester RNA binding proteins. The sequestration of these proteins (not the retention of the mutant transcript alone) is thought to cause the diverse symptoms of DM1 and DM2.

MBNL1 is an RNA binding protein sequestered by these toxic CUG and CCUG repeats and has been shown to regulate the splicing of several different transcripts, including the *TNNT2* gene and the *INSR* gene [38]. Symptoms such as insulin resistance and cardiac abnormalities have been attributed to the mis-splicing of the insulin receptor transcript and the cardiac troponin T transcript respectively, however, the cause of many other symptoms have yet to be defined.

24 different transcripts have been shown to be mis-spliced in myotonic dystrophy and of these pre-mRNA transcripts our lab has characterized MBNL1's TNNT2 binding site in detail. Ho et al. demonstrated that MBNL1 regulates the alternative splicing of the *TNNT2* pre-mRNA as well as showing that MBNL1 directly binds upstream of the regulated exon. It was next established that the MBNL1 binding site within the TNNT2 pre-mRNA forms a 32 nucleotide stem loop, which MBNL1 binds to with high affinity [39].

Emily Goers did a doped SELEX (Systematic Evolution of Ligands by Exponential Enrichment) using the 32 nucleotide TNNT2 binding site characterized by Warf et al. [39]. The RNA templates contained constant regions that flanked a randomized region. The randomized region consisted of the wild type TNNT2 template and was generated such that 51% of the time the wild type nucleotide was used and was changed to either of the other three nucleotides 16% of the time. The doped SELEX allowed MBNL1 to sample other residues while having a bias for the endogenous TNNT2 binding site. The RNA sequences that resulted from the five rounds of SELEX bound MBNL1 with similar affinity as the 32 nucleotide TNNT2 site. The fifth round of

SELEX produced RNAs that had a higher percentage of pyrimidines and contained YGCY motifs.

Results

Determining whether the constant regions pay a role in MBNL1 binding- EMSAs were done to determine whether the constant regions of the SELEX RNAs were playing a role in the binding of MBNL1. Gel shifts of RNAs with and without constant regions were done and the K_ds were compared. One of the RNAs that was characterized in such a manner was the F06 RNA (Figure 6). The 32mer F06 RNA (which did not contain the flanking constant regions) bound similarly to the 81mer F06 RNA (which contained the flanking constant regions). In general, this type of analysis on F06 and several other RNAs showed that although the flanking constant regions could slightly influence MBNL1's binding, the randomized 32 nucleotide region contained the primary elements required for MBNL1 binding.

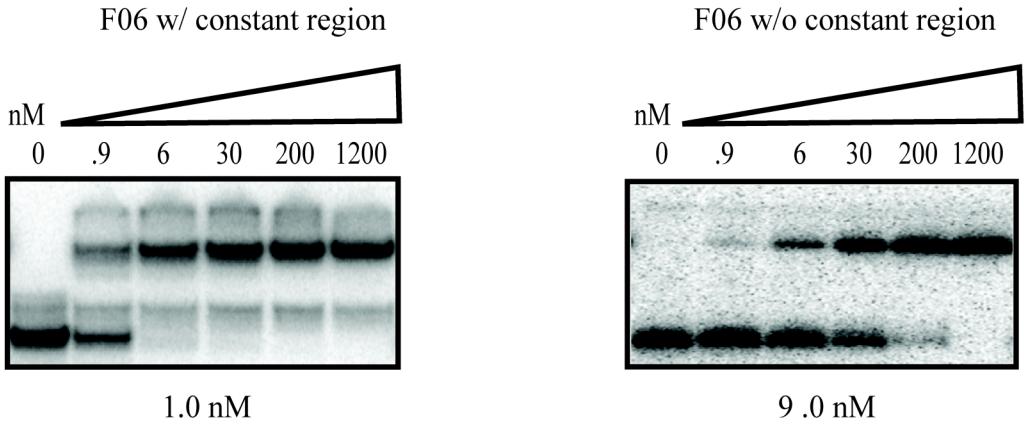

Figure 6. SELEX RNA F06 with and without constant regions. EMSAs showed that MBNL1 bound F06 RNA with constant regions similarly to F06 without constant regions. In general, MBNL1 bound RNAs without constant regions with slight less affinity that RNAs with the constant regions.

MBNL1 binds a consensus motif within the 32 nucleotide TNNT2 binding site- Our lab has characterized in detail MBNL1's 32 nucleotide binding site within the *TNNT2* pre-mRNA. The doped SELEX took advantage of this binding site and allowed for the identification of nucleotides within the *TNNT2* binding site that were important for MBNL1 binding.

RNAs generated for the SELEX were bound to MBNL1, which was bound to GST beads. Unbound RNAs were washed away and RNAs that bound were eluted and were reverse transcribed (RT) and amplified. The RNAs underwent this cycle five times, at which point EMSAs were done to determine whether MBNL1 bound the SELEX RNAs from round 5 with high affinity (Table 1). Round 5 SELEX sequences were sequenced and a common YGCY motif was identified. The sequences were organized into two classes, Class I, which had one or more YGCY motifs and Class II, which has zero to one YGCY motifs (Figure 7). Interestingly, MBNL1 bound Class I RNAs with higher affinity than Class II RNAs.

sequence	K_d (nM) 81mer	K_d (nM) 32mer
WT TNNT2 SELEX	30. ± 20.	20. ± 10.
D01	25 ± 6	
E01	5 ± 2	
H05	25 ± 15	
C12 [a]	>1.2 μM	
B08	> 1.2 μM	
H06	4 ± 4	
E12	13 ± 9	
B07	15 ± 8	0.7 ± 0.5
A04	6 ± 5	180 ± 130
F06	1.0 ± 0.1	9 ± 7
F06 UCCA	~ 140	

[a] gray shading refers to SELEX RNA oligomers in Class II, no shading refers to Class I [b] CA refers to GC to CA mutations in the 5'-YGCY-3' of the SELEX RNA.

Table 1. Binding affinities for selected SELEX sequences: EMSAs were done to determine MBNL1's binding affinity to SELEX RNAs.

Class I

```
WT  CCUGUCUCGCUUUUCCCCUCCGCUGCGGCCAC
E03.CCGCUCUGCUUUGCUCCUGUUGCUACGCGUUC
E12.CCGCUGCUGCUUAUUGCCUAACGACCGUGUUG
C11.CCUGCUCUGCUUUUGCCUUGCGUUCGGGUUAG
A10.CCUGCUUUGCUAAUACCAGCUGCUGUCGUUAC
G06.CCUGCUUGCUUUUCCGCCUACCCUGCUGUCAC
G08.CCUGCUUUGCUAUGCUCCUAGUCUGUGCUCAC
H07.CCUGCUUCGCUGUUGCCUUUCGGUUCUCCAAC
H01.CCUGCUUUGCUAUUGCGCUGCAAUUCCUGUUC
C04.CCUGCUUCGCUGUAACCUCUCGUUGCCGUGUC
G12.CCUGCUUUGCUUUGCGACCACGCUCCCGUACA
F01.CUUGCUUUGCUUUACUCCGCUGCUGCUGUCAC
H10.CCUGCUUUGCUUUUCCUCGUCUCUCGGUGCAC
B05.CCUGCUUUGCUUUCCCAUUCGUUUGCGGACAC
A04.CCUGCUUCGCUUCUGCAUUGUUCGGCCGGUUG
D11.CCUGCUUCGCUUUACUCUUCCGCUGCGUAAUC
F03.CCUGCUUCGCUUAUCGCCUGCUAUGCCGCUAC
G01.CCUGCUUCGCUUCUCCAUCACGUACCAGACUC
G02.CCUGCUUCGCUUUUGCCAUGUUCAGCGUAGUC
G10.CCUGCUUCGCUUAUCCUCUCAGCUACUGCCAC
A08.CCUGCUUCGCUAUUCCGCUUUGUAGCCAUCGU
C05.CCUGCUUGUUAUGCUUGAUUGCCUGUUUCGAG
E07.CAUGCUUCCCAUUUCUGCUUUGCUACGCACGU
C10.CCUGCUUAGCUUUUCGCUUAAACUGGCCACAC
H11.GACCGCUCGCUUUUUCCUAUUCCUGCUUGUCG
H05.CCGUGCUCGCUGAUCACGACAGCUUGGACCUC
D10.CCUUGCUCGCUCUUUCGCUAUGCUUCGGUCAC
H06.CUAUGCUUUGCUACUACCGUCCUACUCGUGCAC
A02.CCGUGCUUGCUUAUACCGUUACAUGUCGCCAG
A06.CCUUGCUUGCUUAAUGGUUCAGCGAUCGGUAG
G09.CCUUGCUUGCUUGUUUCCACGUCUCGAGAUGG
E04.CUUCGCUCGCUUUGCAACUCCACGAUCGCGAU
E05.CUUCGCUUGCUUUCCACACCGCGGUCUUCAC
```

Class I cont.

```
WT  CCUGUCUCGCUUUUCCCCUCCGCUGCGGCCAC
A11.CCACGCUUGCUUUUUCGCUACCGCACAAUCAG
D01.CCUUGCUUGCAUUGCGCUUAUGCUUGGCCUUC
D08.CUACGCUUGCUUUUCGCUUGCCCUGCUGCUUC
G11.CCUGUGCUGCUUUUUGUCCACGCUACUGCUAU
C07.GAUGCGCUGCUUUGCUUUUCAGGUUCGGCCAC
F07.CCACUGCUUCAUUGAGUCUCUGCUGCGCUCUA
D07.CCUGCGCUUAGUUCCGAUGUUCCGCUGCUAC
F09.CCAGUUGCUUGUUUGACGUCGGGUACCGCUAC
A05.CCUCUCGCUUUGCUCCCCCACGGUUCGGUCAU
B06.CCAGUCGCUUUUGUUCCAUCUGCUGUAGUCAC
C02.CCAAACGCUUCCCUCCUGCCUCAUUGCCAUGCU
E01.CCCAUCGCUUAAUUCACUCCCUCUGCUUGCAU
H03.CCUUUACGCUUCUUUCGCUUUGUGUGGGCUAG
B12.CCGACCUCGCUUUUCACCUCCGCUCCAUCGCC
C01.AUUGCCACGCUUUGCCUCUACGCUACAAUCUU
H12.CUUGCCUCGCUUUUAUGACGCUGUGUUCCGAU
C09.CCUGUUUCGCUUUUCACCGUCACUGCUGCCCU
D03.CCUUUCUCGCUUUGCCUUUACGCUGCGUCCAC
F02.CUAGCUUCGCUUUUCUUAGUCACUUCGGGCGC
F06.CCGAUCCCGCUUAUGCCAUCCGCUGCAGUCAC
F11.CCCGGCUCGCUUUUGCUACUCCCUGCUGUCAC
A09.CGUCCCUUGCUUUUCCCUUCCGCUGCCGUCGC
A01.CCUGCCUUGCUUUUCCACUGCUUUGCUCUCAC
F04.CUGAUGCCGCUUUUACGCUUCCGCUACGUACAC
B07.UGUUCCUCGCUUUUCGCCUCCGCUUCCGCCGC
F05.CCUCUUUCGCUUUUUCCGCUUGCUGCAGCUAC
F10.CUUCUCUCGCUUUUUCGCCUUCGCUGCCGCUUC
D12.CCUCCUGCGCUCUUUGCUGCCCUGCUGUGUC
C03.CAUGUCCCUGCUUCCCUUUCCGCUGCUUGCUU
H09.CCCCAUCUCGCUUUGCUACUCCGCCUCUACGAC
E06.CCCAUGCCGCAAUACCGCUCCGCUAACGCUUC
```

Class II

```
WT  CCUGUCUCGCUUUUCCCCUCCGCUGCGGCCAC
E10.CCUGCUUCGCCUUUCCAAACCGCAGCUUUGUC
B04.CCUGCUUCGCCUUCUUGUUACGGAUCGGGUAC
B02.CCAUGCUUGCGUUUCCCUUCGAAGUCGUGCAC
F12.CCAUGCUCUUCGCCUGCCUACCUUCAAUGCAC
E11.CCUCUGCUUGCAAUCCCGUCCGUUGCCGAUAC
D05.CCAUUGCUGCUUGUAGUUUCUUCACCGACCUA
B09.CCUCUGCUGCUUUUUGCCUACUAUGAGUUUGC
B10.CCUUUGCUGCUUUCUUCGUCCCUGUCGCCAU
C06.CCAUAACCGCUUCUGCCCUAGUCACCGUGUUC
G04.CCUGGCUUGCUUUUAUCGGUCGUGAAUGCAUC
D09.CCCGCCUCAUUGCUACAUUCGUUCGCGGCUGC
D04.CCUGCCGCGCCUUUCGCUUUGACGGUUCCCGU
B01.ACCGUCUGGCCAUGCCCAUACGCUUUUAUCAC
A12.GCCCUUUCGAGUAUCGAUACUGCUGCGACGUC
A07.CCAGCGUUGGUCGUCCCGUACACGCUUGCGCC
H04.CCAGCUUGCCCUUGCACAUUCGCCGCUGCACG
C12.CCUUGGUUGUCAGUCCGUUCAAUUGCUGUCCC
B08.CCUACAUCGGUUGUUUGGUCCUCUAGUGCUAG
E09.CUUAACGGACUUUGCACCUCUGGUGGUUAUAC
```

Figure 7. Classification of SELEX RNAs. SELEX RNAs were put into two different classes. Class I contained RNAs that contained two or more YGCY motifs and bound MBNL1 with high affinity. Class II contained RNAs that contained zero to one YGCY motifs and bound MBNL1 with lower affinity.

Mutational analysis of SELEX RNAs reveals YGCY motifs are required for MBNL1 binding- To determine whether MBNL1 recognizes the SELEX RNA sequences through its YGCY motifs, the F06 RNA was mutated. The F06 RNA contains 3 YGCY motifs. Mutation of the first YGCY (CGCU) motif to UCCA significantly reduces MBNL1's ability to bind the F06 RNA by 140 fold (Figure 8).

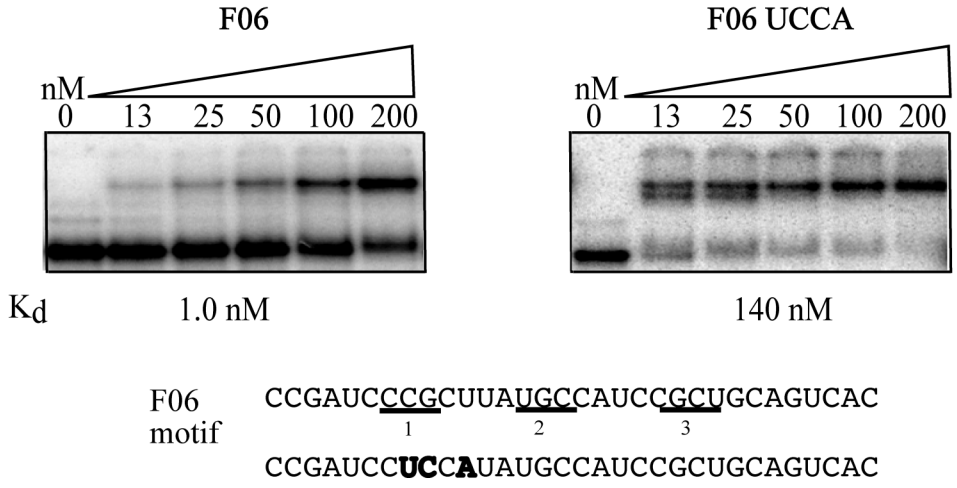

Figure 8. Mutational analysis of the F06 RNA. Mutations in the 1st YGCY motif result in decreased affinity to MBNL1.

Motif identified in the doped SELEX binds MBNL1 in the context of MBNL1's pre-mRNA targets-The YGCY motif was used to determine potential MBNL1 binding sites within the pre-mRNAs that are mis-spliced in myotonic dystrophy. We searched 200 nucleotides upstream, downstream and even within the regulated exon of the pre-mRNA for YGCY motifs in the 24 pre-mRNA transcripts shown to be mis-spliced in myotonic dystrophy. We choose five pre-mRNA transcripts out of the 24 to study in more detail. 40-45 nucleotide regions of the *MBNL1, MBNL2, SERCA1, GRIN1* and

INSR pre-mRNAs that contained YGCY motifs were studied (Figure 9A and 9B). Two potential binding sites within the *MBNL1* pre-mRNA called *MBNL1* Site #1 and *MBNL1* Site #2 bound MBNL1 with a K_d of 11 nM and 45 nM respectively. A potential MBNL1 binding site within the *MBNL2* and *SERCA1* pre-mRNAs each bound MBNL1 with a K_d of 5.8 nM and 15 nM respectively. The RNAs within the *GRIN1* and *INSR* pre-mRNAs that contained YGCY motifs bound with lower affinity. Comparison of the RNA sequences of the potential endogenous MBNL1 binding sites showed that the four RNAs that bound MBNL1 with high affinity not only had one or more YGCY motifs but also contained a high percentage of pyrimidines, which the *GRIN1* and *INSR* RNAs lacked.

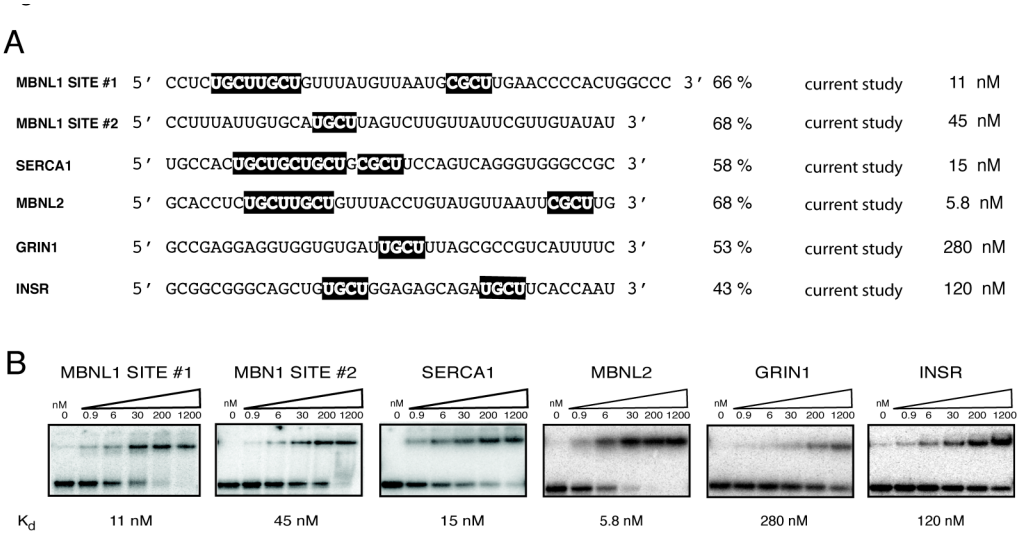

Figure 9. MBNL1 binds 40-45 nucleotide potential endogenous sites. MBNL1 bound potential endogenous sites with YGCY motifs flanked by a high percentage of pyrimidines better than RNAs that did not contain a high percentage of pyrimidines.

Mutational analysis of MBNL1 Site #1- Closer analysis of the five pre-mRNAs we chose to study showed that the *MBNL1* pre-mRNA had an unusually high amount of conservation upstream of the regulated exon (exon 5) as well as within the regulated exon itself. This peaked our interest of this pre-mRNA and we chose to do mutational analysis of MBNL1 Site #1. The predicted secondary structure of MBNL1 Site #1 suggests that

this RNA forms a stable stem loop (Figure 10A). Previous work on MBNL1 has shown that this protein likes to bind to GC dinucleotides in RNAs [29]. We therefore decided to mutate the GCs within the YGCY motifs to CAs as well as add a UUCG cap, which guarantees the RNA will form a stable stem loop so as to not change the secondary structure. This mutant was referred to as MBNL1 CA_CAP (Figure 10B). When EMSA studies were done with MBNL1 CA_CAP we saw abolished binding.

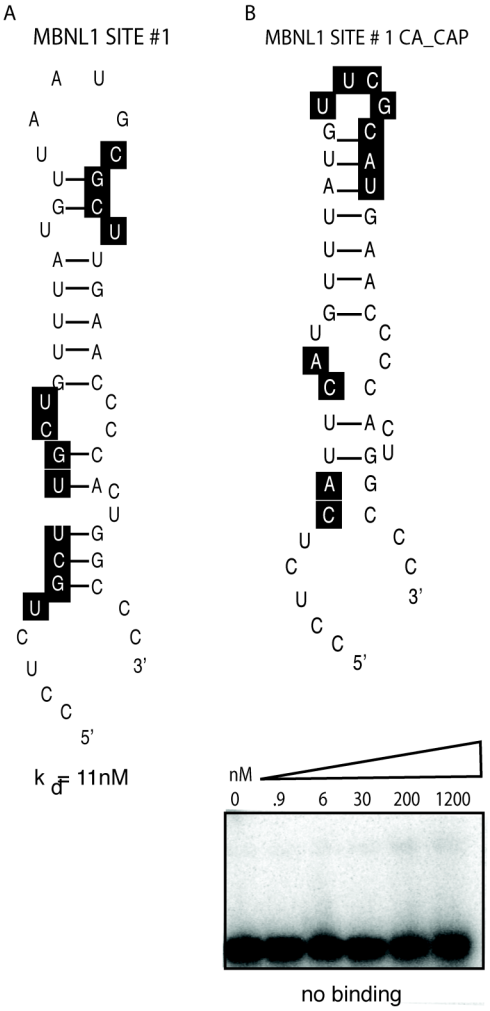

Figure 10. Mutational analysis on MBNL1 Site #1. (A) shows the predicted secondary structure of MBNL1 Site #1 and the K_d at which MBNL1 binds this RNA. Figure (B) shows the mutations that obliterated the UGCU motifs and forced a stem loop structure, this RNA was called MBNL1 Site #1 CA_CAP. MBNL1 did not bind MBNL1 Site #1 CA_CAP.

To determine what the effects of this mutation (on only one of the potential MBNL1 binding sites) within the MBNL1 pre-mRNA are *in vivo*, we made these mutations in a *MBNL1* minigene that was transfected into HeLa cells. MBNL1 regulates exon exclusion of exon 5 within the *MBNL1* pre-mRNA. Transfection of the wild type *MBNL1* minigene resulted in 70% inclusion of exon 5. Transfection of the wild type MBNL1 minigene with the overexpression of MBNL1 resulted in 8% exon inclusion, meaning MBNL1 nearly completely blocks the use of the 3' splice site adjacent to exon 5. Transfection of the mutant minigene resulted in ~90% exon inclusion meaning the mutations enhanced the use of this 3'splice site. Transfection of the mutant minigene with overexpression of MBNL1 resulted in only ~10% exon inclusion, which means MBNL1 is still strongly regulating this exon. These results indicate that MBNL1 either does not regulate the exon through this site or that this site is only one of many MBNL1 sites. It is possible *MBNL1* site #2 and site #1 work together to regulate splicing. Additional sites may also exist.

Materials and Methods

Protein purification. The MBNL1 protein construct, which includes amino acids 1–260 and contains an N-terminal GST tag, was expressed and purified as previously described [39], except for the following differences. For the lysis of bacterial cells, 10μg/mL of DNase was added with the lysozyme (1 mg/mL) and the cell extract was centrifuged for 30 minutes at 15,000 rpm. Supernatant was loaded onto GST beads and washed with PBS-T buffer (1X PBS and 1%Triton X-100) and eluted with elution buffer (10 mM reduced glutathione, 50 mM Tris pH 9.5).

Labeling of RNAs for gel mobility shift assay- The RNA was transcribed from purified PCR product using T7 RNA polymerase and [α-P^{32}] CTP.

Gel mobility shift assay was performed as previously described [40] except for the following changes. Reactions were run on a 5% polyacrylamide gel (.5X TBE, 5% 37.5:1 acrylamide:bisacrylamide and 4% glycerol). The gels were run for 2 hours at 4ºC at 170V.

In vivo splicing. Splicing assays were done as previously described [41], except for the following changes. All reporters were reverse transcribed using a pcDNA3 plasmid specific antisense primer 5'–AGCATTTAGGTGACACTATAGAAT AGGG–3'. The –RT reactions were treated the same as the +RT reactions except that no SuperScriptIII was added to the –RT reactions. The cDNA from the RT reaction (2μL) was subjected to 26 rounds of PCR (within linear range) in a 20μL reaction. PCR amplification for all splice products was done using the sense primer 5'–CCACAGGATC CGCTTCTTCTTCTTCATGTTGACTAAACCTCAT –3' and the antisense primer 5'–A TTCTTATGCGGCCGCCAGATTCATTTATTAAGAAACCCCA CCCC–3'.

Discussion

Most of the SELEX RNAs have in common multiple YGCY motifs. Also, several of the potential endogenous MBNL1 sites contain multiple YGCY motifs flanked by a high percentage of pyrimidines. Much evidence points to MBNL1 requiring high pyrimidine content flanking the YGCY motif(s) since the SELEX RNAs have an overall bias for uridines, the native TNNT2 site contains a high percentage of uridines and CUG and CCUG repeat RNAs are pyrimidine rich. This suggests that high-affinity MBNL1 sites require multiple YGCY motifs in pyrimidine rich RNAs.

Of the potential MBNL1 endogenous sites that were tested via EMSA, we used Sfold to determine the predicted structure of the RNAs. The secondary structures of these 40-45 nucleotide RNAs ranged from strong stem-loops to much weaker stem-loops. There seemed to be no correlation between MBNL1's binding affinity and RNA structure.

The doped SELEX provided us with a YGCY consensus motif as well as a set of criteria to look for potential endogenous MBNL1 binding sites. Further examination of the 5 potential endogenous targets showed that MBNL1 binds RNA that contains YGCY motifs surrounded by pyrimidines. Mutational analysis of MBNL1 site #1 showed that although MBNL1 is no longer able to bind to this RNA *in vitro*, *In vivo* the overexpression of MBNL1 still causes exon 5 exclusion. This suggests that MBNL1 may work through several sites within the *MBNL1* pre-mRNA to regulate exon 5 exclusion. It will be important to determine whether MBNL1 binds and regulates the potential binding sites described above.

CHAPTER IV

AUTO-REGULATION OF MBNL1 PRE-MRNA

I was the primary contributor to the experiments described below and did all of the writing. Leslie Coonrod was helpful in running the splicing gels. J. Andrew Berglund was the principal investigator for this work.

Introduction

Splicing of pre-mRNAs is an important event that contributes to a diverse proteome as well as the regulation of gene expression. It is estimated that more than 90% of human genes undergo alternative splicing [1, 2]. To produce a functional mRNA, non-coding regions must be accurately removed and the coding regions must be ligated together. Splicing occurs by two transesterification reactions that result in removal of the intron and ligation of the exons. This splicing mechanism relies heavily on pre-mRNA sequences, proteins and small nuclear RNAs (snRNA) to aid in the different reactions. Cis-sequences that are important for splicing include the 5' splice site (5' ss), the branchpoint sequence, the polypyrimidine (PY) tract and the 3' splice site (3' ss). These canonical intronic motifs plus additional regulatory splicing motifs found in exons and introns are recognized by splicing factors and snRNPs (U1, U2, U4, U5 and U6) to form the spliceosome, which catalyzes intron removal [42].

There are many splicing factors that are only involved in a sub-set of splicing decisions. These include the human muscleblind-like (MBNL1/2/3) family of RNA binding proteins. The founding member of this family, muscleblind (Mbl), was discovered in *Drosophila* and was shown to be important for photoreceptor differentiation and terminal differentiation of muscles [13, 14]. Subsequently, MBNL

proteins were found to associate with expanded CUG repeats (located in 3' untranslated region of the DMPK gene) that have been shown to act as a toxic RNA and are at least partially responsible for causing myotonic dystrophy type I (DM1) [5]. The expanded CUG repeats sequester MBNL proteins into nuclear foci leading to loss of active protein [11, 12]. Expanded CCUG repeats within the first intron of ZNF9 also sequester MBNL proteins, which is thought to be at least partially responsible for causing DM type 2 (DM2) [6]. The sequestration of MBNL1 leads to mis-splicing of developmental regulated events, and a few of these events have been linked directly to symptoms in DM1 and DM2, such as the mis-splicing of the chloride channel (CLN1) leading to myotonia [43]. Increased levels of CUGBP1 have also been shown to result in mis-splicing of certain pre-mRNA transcripts and are linked to causing heart defects found in DM patients [6].

Exon 5 of both *MBNL1* and *MBNL2* pre-mRNAs are mis-spliced in DM [18]. These paralogous exons are embedded within highly conserved regions of the genome [44, 45]. Figure 11A shows the UCSC genome browser view of the conservation upstream, downstream and within exon 5 of *MBNL1* [The ultra-conserved element described by Bejerano and colleagues [44] is underlined in Figure 11B]. This conserved exon in MBNL1 and MBNL2 encodes 18 amino acids that are C-terminal of the fourth and final Zinc (Zn) finger RNA binding domain. The inclusion of exon 5 causes MBNL1 and MBNL2 to be localized primarily in the nucleus while isoforms of MBNL1 and MBNL2 lacking these amino acids are found both in the nucleus and cytoplasm [27, 46]. MBNL3 differs from MBNL1 and MBNL2 in that it lacks the conserved region upstream of exon 5 as well as the 18 amino acids encoded by exon 5.

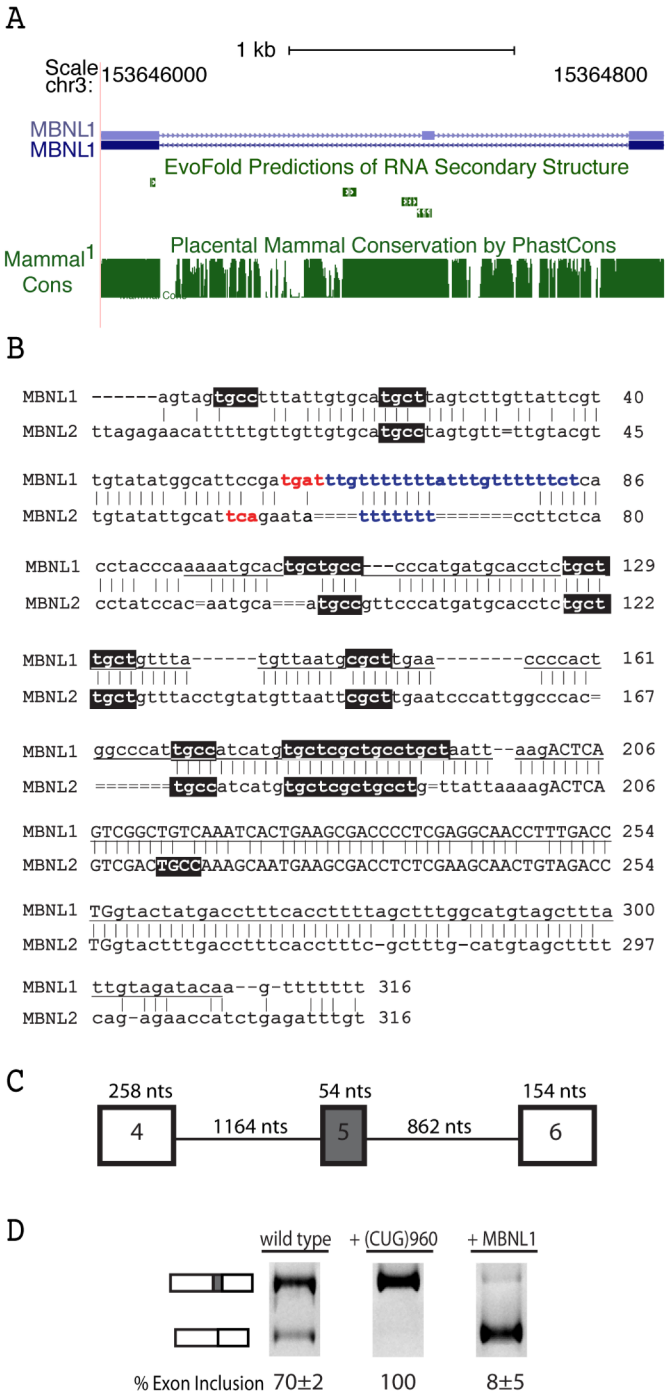

Figure 11. MBNL1 auto-regulated splicing. (A) The *MBNL1* pre-mRNA contains a conserved intronic element upstream of exon 5 (edited from UCSC genome browser). The conserved element is predicted by Evofold [47] to contain several secondary structures (represented by green boxes). (B) Schematic of the wild type *MBNL1* minigene containing exon 4, intron 4, exon 5, intron 5 and exon 6. (C) Exon 5 exclusion of the wild type minigene is MBNL1 dependent. (D) Sequence alignment between *MBNL1* and *MBNL2* that includes the 3' end of intron 4, exon 5 and the highly conserved region of intron 5. Intronic sequence is in lowercase and exon 5 is capitalized. The

MBNL1 pre-mRNA distant branchpoint is shown in red, the PY tract is shown in blue and YGCY (MBNL1 binding sites) motifs are highlighted in black The sequence underlined is the ultra-conserved element described by Bejerano and colleagues. Intron 4 contains a non-canonical AAG 3' ss and intron 5 contains a weak GTACTA 5' ss.

We recently identified YGCY as a minimal MBNL1 RNA binding site [40], and demonstrated that insertion of multiple copies of this motif into an intron adjacent to an exon that is not normally regulated by MBNL1 is sufficient for regulation by MBNL1. In general, the location of the YGCY motifs correlates with the effect (e.g. enhancement or suppression) that MBNL1 has on splicing. When YGCY motifs are located upstream of the exon, MBNL1 binding generally leads to exon exclusion; when YGCY motifs are located downstream of an exon, MBNL1 binding generally leads to exon inclusion [21, 40]. Twelve YGCY motifs occur within the first 200 nucleotides of the upstream acceptor sequence in *MBNL1* but none are found in exon 5 or the donor region (Figure 11B). The intronic sequence upstream of exon 5 in *MBNL2* contains 9 YGCY motifs, while exon 5 and the intronic region downstream (1st 200 nucleotides) only contain one YGCY motif. The location of these YGCY motifs is consistent with MBNL1 proteins acting as repressors of exon 5 in *MBNL1* and *MBNL2*. Interestingly, the location of most of the YGCY motifs upstream of exon 5 in *MBNL1* and *MBNL2* are conserved (Figure 11B). The presence of redundant MBNL1 binding sites (YGCY motifs) in the response element (located between the branchpoint and 3' ss in intron 4) may serve as a way to increase binding at lower cellular concentrations of MBNL1.

Previously, we showed that one mechanism through which MBNL1 acted as a repressor was to compete with the basal splicing factor, U2AF65 for binding at the 3' end of the intron [31]. In this example, the MBNL1 binding site overlaps with the PY tract (U2AF65 binding site) in intron 4 of the *TNNT2* pre-mRNA. We showed that when

MBNL1 bound intron 4 it suppressed U2AF65 binding, resulting in suppression of U2 snRNP binding.

The proposed MBNL1 binding sites within intron 4 for the *MBNL1* pre-mRNA do not appear to overlap with any of the canonical intronic splicing signals. In addition, the architecture of these introns is unique compared to most mammalian introns. For instance, intron 4 of *MBNL1* and intron 4 of *MBNL*2 both have an AAG 3' ss instead of the consensus YAG 3' ss. Intron 4 of *MBNL1* contains a predicted distant branchpoint sequence (TGAT, in red text in Figure 11B) that is 141 nucleotides from the 3' ss. For *MBNL2* the predicted distant branchpoint is also 141 nucleotides away from the 3'ss. Interestingly, the sequence conservation of intron 4 between *MBNL1* and *MBNL2* does not extend to the upstream PY tracts and distant branchpoint sequences. The PY tract found adjacent to the predicted branchpoint sequence within intron 4 of *MBNL1* is a canonical or "strong" site for U2AF65 [41], while the PY tract and branchpoint sequence for intron 4 of *MBNL2* is weaker compared to *MBNL1* (Figure 11B). Introns with distant branchpoint sequences typically lack AG dinucleotides between the branchpoint sequence and 3' ss and this region has been termed an AG Exclusion zone (AGEZ). Introduction of an AG in this zone can lead to its use as a cryptic 3' ss. Introns containing long AGEZs (100 nucleotides or greater) are associated with higher rates of alternative splicing suggesting that these regions contain regulatory elements [48]. Interestingly, many exons with AGEZs of 150 nucleotides or longer are within genes that are either known to be associated with disease or are of biomedical interest [48]. The MBNL1 binding sites are found in the 173 nucleotide AGEZ of intron 4 of *MBNL1* and in the 141 nucleotide AGEZ of intron 4 of *MBNL2*.

To study the auto-regulated MBNL1 splicing we have created a mini-gene that contains exons 4, 5, and 6 of *MBNL1* and its intervening introns (Figure 9C). We show that MBNL1 can regulate a non-canonical intron by binding a mostly unstructured 90 nucleotide response element within the AGEZ, upstream of exon 5. Smaller deletions within the MBNL1 response element did not eliminate MBNL1's ability to regulate exon 5 exclusion. We observed that exon 5 of *MBNL1* is primarily regulated through the usage of a distant branchpoint and deletion of this branchpoint causes exon 5 skipping. These results combined with previous results suggest that MBNL1 does not negatively regulate its target pre-mRNA transcripts through a conserved mechanism.

Results

A mini-gene recapitulates auto-regulation of MBNL1 exon 5- The mini-gene constructed to study splicing of exon 5 contains the full-length sequences of exon 4, intron 4, exon 5, intron 5 and exon 6 (Figure 11C). The mini-gene was transfected into HeLa cells to determine if MBNL1 could regulate splicing of exon 5 in this context. As shown in Figure 11D, when a plasmid that expresses 960 CUG repeats [38] is also transfected into the cells, the inclusion of exon 5 increases to 100% compared to 70% (Figure 11D). This change is presumably the result of the CUG repeats sequestering the endogenous MBNL proteins. The inclusion of exon 5 can be nearly completely blocked (8% inclusion, Figure 11D) by the over-expression of MBNL1 from a co-transfected plasmid. These results show that protein levels of MBNL1 play a significant role in the regulation of *MBNL1* exon 5 and that we can recapitulate MBNL1 regulated splicing in this *in vivo* system.

Characterization of the highly conserved 3' end of MBNL1 intron 4- To determine if the putative distant branchpoint sequence discussed in the introduction was important

for the splicing of intron 4, the sequence ATGAT (proposed branchsite adenosine is underlined) was deleted (Figure 12A, referred to as Δbp). As shown in Figure 12B, the deletion of this motif reduced exon 5 inclusion to 10%. Deletion of the putative branchpoint sequence causes the splicing machinery to skip this 3' ss and select the branchpoint and 3' ss of intron 5 resulting in skipping of exon 5. These results show that this putative branchpoint sequence is necessary for high levels of exon 5 inclusion, presumably because it contains the branchsite adenosine.

In order to directly determine if the adenosine at position –141 within intron 4 functioned as the branchpoint for this intron, branchpoint mapping was performed. The strategy to capture intron 4 lariats is shown in Figure 12C, and previous published protocols were followed [49]. Nested PCR was used to decrease background. Primers C and D were used for the first PCR reaction and primers A and B were used for the second reaction (Figure 12C). From 45 sequences, twenty one mapped to the distant TGAT branchpoint (–141 nts), five mapped to the end of what we have portrayed as the PY tract (–115 nts), one mapped to a region between the PY tract and 3' ss (–34 nts) and one mapped to the first nucleotide of exon 5 (0 nts), as shown in Figure 12C. Dots above and below the nucleotides in Figure 10D are color coded to correlate with the adenosine that may be used as the branchpoint adenosine for each lariat that was sequenced. The remaining 17 sequences either contained multiple templates, sequences that suggested no lariat formation, or sequences that did not map to intron 4. These results are consistent with the 5 nucleotide deletion (Figure 12B) and indicate that the distant branchpoint is the primary branchpoint for intron 4.

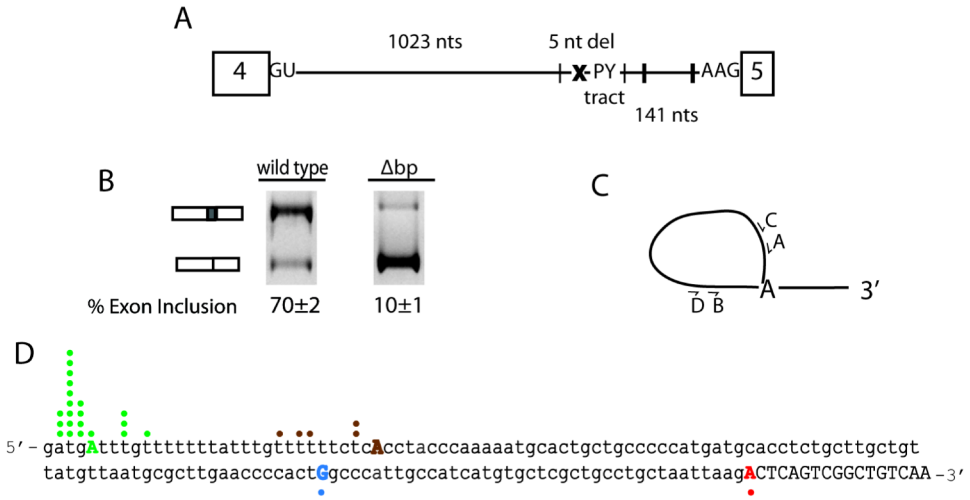

Figure 12. Determination of the branchpoint of the 4th intron in the *MBNL1* pre-mRNA. (A) Schematic of the *MBNL1* pre-mRNA showing the predicted MBNL1 binding sites (black hatch marks indicate instances of YGCY motifs) and the x represents the deletion of the predicted branchpoint (ATGAT) in intron 4. The mini-gene containing the deletion of the predicted branchpoint was termed the Δbp mini-gene. (B) Splicing of the wild type and Δbp mini-gene splicing reporters. (C) RT-PCR of intron 4 lariats of the *MBNL1* pre-mRNA indicates that the 5 nucleotides deleted in the Δbp mini-gene contain the branchpoint sequence. Primer C was used for RT-PCR, primer pairs C and D were used for the initial PCR and primer pairs A and B were used for the final nested PCR. (D) Sequence for the 3' end of intron 4 (in lowercase) and part of exon 5 (capitalized) is shown. Dots correlate with lariats that were sequenced to the particular nucleotide it is placed above or under and are color coded to the branchpoint adenosine that was likely used. The 5 nucleotides that were deleted in the Δbp mini-gene was the most commonly used branchpoint. Other branchpoint sites that weren't used as often were located at the 3' end of the PY tract, between the PY tract and the 3' ss and the first nucleotide of exon 5.

MBNL1 binds a primarily unstructured region between the distant branchpoint and 3' splice site- To determine the structure of the highly conserved intronic region of intron 4 and to identify the binding sites for MBNL1, we performed RNA structure probing and footprinting experiments. This was done with a 491 nucleotide RNA that contains the last 207 nucleotides of intron 4, all 54 nucleotides of exon 5 and the first 227 nucleotides of intron 5. This RNA included the 212 nucleotide ultra-conserved element, which spans the intronic region upstream and downstream of exon 5, as well as exon 5 itself. Additional sequence upstream and downstream of the ultra-conserved element was

used to favor the native secondary structure. Because a larger stretch of sequence upstream of exon 5 is highly conserved and contains proposed MBNL1 binding sites, we focused our structure probing and footprinting studies on this region of the RNA.

Both SHAPE (Selective 2' Hydroxyl Acylation by Primer Extension [50]) and RNases were used to determine the secondary structure of the 3' end of intron 4. RNase T1 cleaves after single-stranded guanosine residues, RNase 1 cleaves after single stranded nucleotides with a bias for single-stranded pyrimidines and SHAPE uses NMIA (N-methylisatoic Anhydride) to form 2'-O-adducts. NMIA reacts with all nucleotides and the extent of the modification is dependent on the flexibility of the nucleotide [50]. Shown in Figure 13A–C are representative gels of all three assays. The quantified (see Materials and Methods) readout from these experiments was fed into RNAStructure (Reuter et al., 2010) to create the secondary structure shown in Figure 13F (only nucleotides 42–210 were quantified). The RNA appears to have a structured 5' end, a primarily unstructured linker region that contains several of the proposed MBNL1 binding sites and then a more structured region that contains the end of intron 4. Secondary structure of exon 5 and intron 5 are predictions from the RNAStructure program. The distant branchpoint sequence is at the junction of two helices with a large bulge that contains the PY tract. The 3' ss is also at the junction of RNA structural elements.

To determine where MBNL1 binds within this RNA, footprinting with T1 and R1 RNases were performed. These experiments revealed that the addition of 5μM MBNL1 changed the cleavage pattern at a number of nucleotides. The differences in cleavage (see lane 3 in Figure 13A–B) were quantified by subtracting the averages of the reactivity

profiles in the presence and absence of MBNL1 and normalizing the data as described in the experimental procedures section (Figure 13D–E). Only nucleotides that showed a difference of more than 0.2 for RNase T1 and RNase 1 are shown in figure 13F (red represents reduced cleavage and blue represents enhanced cleavage in the presence of MBNL1). For example, G144 showed a large reduction in cleavage by RNase T1 suggesting MBNL1 interacts with this nucleotide. G137, which is located within a YGCY motif, and G156 (in UGCG motif) also show significant protection by MBNL1 (Figure 13D and 13F). It is interesting that out of the ten YGCY motifs within the last 141 nucleotides of intron 4, only two YGCY motifs show a footprint (nucleotides G137 and G141). However, nucleotides near YGCY motifs, such as G144–G150, G156, C175 and A196 also show an increased amount of protection suggesting MBNL1 interacts with these nucleotides. Although it appears that there are differences in the RNase T1 cleavage upstream of G86 (Figure 13A, lanes 2 and 3), quantification of 3 different gels showed no significant difference. Interestingly, some nucleotides are cleaved more in the presence of MBNL1, such as nucleotides G150, C167, C172, G173, G176, G198, C190, G194, G197 and G201. Nucleotides with enhanced cleavage near protected sites could be due to MBNL1 affecting the local secondary structure of the RNA.

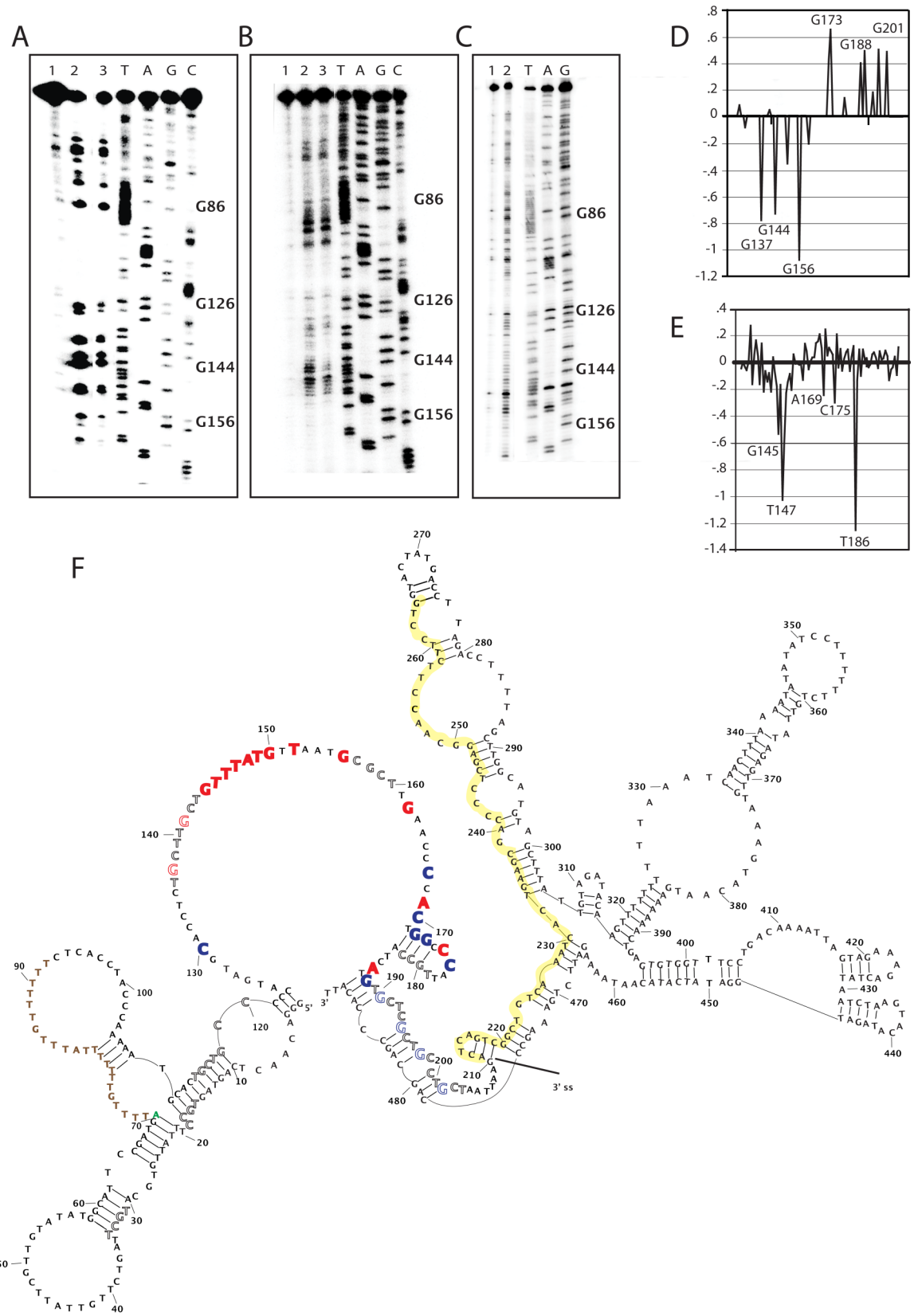
A
1 2 3 T A G C
G86
G126
G144
G156
B
1 2 3 T A G C
G86
G126
G144
G156
C
1 2 T A G
G86
G126
G144
G156
D
G173
G201
G188
G144
G137
G156
E
A169
C175
G145
T147
T186
F
3' ss

Figure 13. Structure probing and footprinting of the 3' end of intron 4 of *MBNL1* pre-mRNA. (A) T1 RNase (cleaves single stranded guanosines) was used for structure probing and footprinting purposes. Lane 1 is the free RNA, lane 2 has RNA and T1 RNase and lane 3 includes RNA, MBNL1 and T1 RNase. Lane 3 shows a footprint between nucleotides G126 and G162. (B) RNase 1 (cleaves single stranded pyrimidines) was used for structure probing and footprinting purposes. Lanes are similar to that described in Figure A except RNase 1 is being used. Lane 3 shows footprinting between nucleotides G126 and G162. (C) SHAPE (used to indicate flexible nucleotides) was used to determine the secondary structure of the intronic region upstream of exon 5. (D) A difference plot of the T1 RNase data was created by subtracting the footprinting data (lane 2) from the structure probing data (lane 1) of the T1 RNase gels. (E) A difference plot of the R1 RNase data was created in the same way as for Figure D. The difference plots in panels D and E focus on the region between nucleotides G126 and G210. (F) Secondary structure of part of intron 4, exon 5 and part of intron 5. Intron 4 contains data from T1, RNase 1 and SHAPE gels. Nucleotides that are highlighted in yellow are in exon 5. Exon 5 and intron 5 structure predictions are based solely on predictions from the RNAStructure program and include no experimental data. Nucleotides in red are nucleotides that are footprinted to in the T1 RNase and R1 RNase assays. Nucleotides in blue are nucleotides that saw greater cleavage in the presence of MBNL1. The adenosine in green is the distant branchpoint and nucleotides in brown are the PY tract. Outlined nucleotides are potential MBNL1 binding site (YGCY motifs).

Identification of a 90 nucleotide MBNL1 regulatory element in the 3' end of intron 4- To determine if the region of intron 4 protected by MBNL1 is required for MBNL1 to negatively regulate exon 5, this section of RNA was deleted from the *MBNL1* mini-gene and replaced by an Mfe1 cut site. Ninety nucleotides between the distant branchpoint and 3' ss were deleted, resulting in an intron 4 lacking all YGCY motifs except two located upstream of the branchpoint. This deletion (Δ90) also resulted in an intron that was more similar to canonical introns in which the PY tract and branchpoint are found closer to the 3' ss (Figure 14A). Splicing of the Δ90 mini-gene results in almost complete inclusion of exon 5 (Figure 14B) and the over-expression of MBNL1 did not alter the splicing pattern (Figure 14B), indicating that this region of the intron (MBNL1 response element) is required for regulation by MBNL1. Interestingly, this element doesn't appear to contain any essential positive splicing signals for exon 5.

To characterize the binding of MBNL1 to this response element, a gel shift assay was performed with this RNA. MBNL1 binds this RNA with high affinity (at 30nM protein, nearly all RNA is bound, lane 6 in Figure 14C). It appears that several MBNL1 proteins bind this RNA because three different complexes can be distinguished (Figure 14C). This result is not surprising given that this 90 nucleotide RNA contains 10 YGCY motifs.

In an effort to determine if the MBNL1 response element could be pared down to a more minimal element, smaller deletions in this region were made. Del 1 Δ18 eliminated the two YGCY motifs closest to the PY tract, Del 2 Δ16 eliminated the next two downstream YGCY motifs, Del 3 Δ18 eliminated one YGCY motif, Del 4 Δ19 eliminated one YGCY motif and Del 5 Δ18 eliminated a string of four YGCY motifs (Figure 14A). Del 1 Δ18, Del 3 Δ18 and Del 5 Δ18 all resulted in an increase of exon 5 inclusion (Figure 14D) where levels ranged from 82% to 99% where as Del 2 Δ16 and Del 4 Δ18 both resulted in wild type levels of exon 5 inclusion. In all deletions MBNL1 was still able to inhibit exon 5 inclusion and the effect of MBNL1 over-expression is strong in all cases (levels range from 7% to 14% inclusion of exon 5). These results indicate that the loss of one to four YGCY motifs in the MBNL1 response element is not sufficient to abrogate MBNL1's ability to regulate exon 5 splicing.

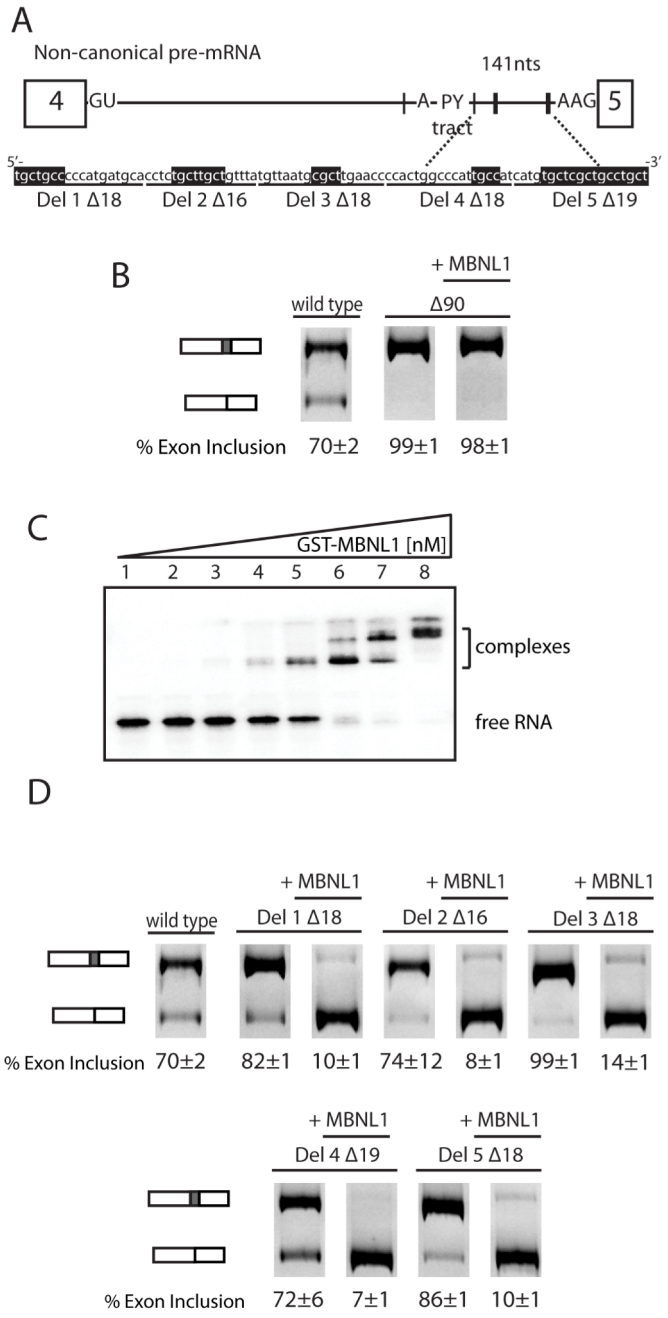

Figure 14. 90 nucleotides between the PY tract and 3' splice site are required for *MBNL1* auto-regulation. (A) A schematic representation of the non-canonical intron 4 of *MBNL*1 and sequence of the 90 nucleotide (sequence shown) MBNL1 response element. Black hatch marks indicate instances of YGCY motifs and the thickness of the hatch mark indicates the frequency of YGCY motifs. These 90 nucleotides (sequence shown) were removed from the mini-gene to create the Δ90 construct. Shorter deletions within the 90 nucleotide region are indicated below the sequence and YGCY motifs are highlighted. (B) Splicing of Δ90 mini-gene in HeLa cells in the absence and presence of overexpressed MBNL1. (C) Gel shift of the 90 nucleotides deleted in the Δ90 minigene. Concentrations of the protein increase from left to right: 0 nM, 0.026 nM, 0.2 nM, 0.9

nM, 6 nM, 30 nM, 200 nM and 1200 nM. (D) Splicing results of the smaller deletions within the 90 nucleotide sequence shown in panel A.

Materials and Methods

Labeling of RNAs for gel mobility shift assay- The RNA was transcribed from purified PCR product using T7 RNA polymerase and [α-P^{32}] CTP. The RNA sequence including the T7 site is 5'– GAUAAUACGACUCACUAUAGGGUGCUGCCCCCAUG AUGCCCUCUGCUUGCUGUUUAUGUUAAUGCGCUUGAACCCCACUGGCCCAU UGCCAUCAUGUGCUCGCUGCCUGCU –3'. The T7 site as well as the three additional guanosines added are underlined.

Gel mobility shift assay- Gel mobility shift assay was performed as previously described [40] except for the following changes. Reactions were run on a 5% polyacrylamide gel (.5X TBE, 5% 37.5:1 acrylamide:bisacrylamide and 4% glycerol). The gels were run for 2 hours at 4ºC at 170V.

Construction of splicing reporter constructs- The MBNL-eGFP construct was obtained from the laboratory of Maury Swanson [51] and the DMPK-CUG_{960} plasmid was obtained from the laboratory of Thomas Cooper [38]. The wild type *MBNL1* minigene was made by amplifying regions of the *MBNL1* gene containing 51 nucleotides from the 3' end of intron 3, exon 4, intron 4, exon 5, intron 5, exon 6 and 33 nucleotides of the 5' end of intron 6 from HeLa genomic DNA using PCR primers with unique restriction sites. The forward primer (5'–CCACAGGATCCGCTTC TTCTTCTTCATGTTGACTAAACCTCATG–3') contained a BamHI site and the reverse primer (5'–ATTCTTATGCGGCCGCCAGATTCATTTATTAAGAAACCCCA CCCC–3') contained a NotI site. The amplified genomic DNA was cut with BamHI and NotI and inserted into a pcDNA3 plasmid and sequenced.

The Δ*bp* minigene was made by deleting 5 nucleotides using standard PCR techniques. The Δ*bp* minigene was created in two segments. The first segment was created by using the forward primer 5'– CCACAGGATCCGCTTCTTCTTCTTCATGT TGACTAAACCT CATG–3' and the reverse primer 5'–GGGTAGGTGAGAAAAAAC AAATAAAAAAACAACGGAATGCCATATACAACGAATAACAAG–3'. The second segment was made using the forward primer 5'–TTGTTTTTTTATTTGTTTTTT CTCACCTACCCAAAAATGCACTGCTGCCCCC–3' and the reverse primer 5'– ATTCTTATGCGGCCGCCAGATTCATTTATTAAGAAACCCCACCCC–3'. Both segments were then used in the same PCR reaction and the forward primer 5'– CCACAGGATCCGCTTCTTCTTCTTC ATGTTGACTAAACCTCATG–3' and reverse primer 5'–ATTCTTATGCGGCCGC CAGATTCATTTATTAAGAAACCCCACCCC– 3' were used to PCR amplify the minigene. The PCR product was cut with BamHI and NotI and ligated into a pcDNA3 plasmid and sequenced.

The Δ*90* minigene was made in two segments. The first segment was made using the forward primer 5'– CCACAGGATCCGCTTCTTCTTCTTCATGTTGACTAAACCT CATG–3' and reverse primer 5'– GGCTTTCAATTGGTGCATTTTTGGGTAGGTGAG AAAAAACA–3'. The second segment was made using the forward primer 5'–GGCTTT CAATTGAATTAAGACTCAGTCGGCTGTCAAATCAC–3' and the reverse primer 5'– ATTCTTATGCGGCCGCCAGATTCATTTATTAAGAAACCCCACCCC–3'. Segment 1 was cut with Mfe1 and BamHI and segment 2 was cut with Mfe1 and NotI. Segment 1 and 2 were then ligated into a pcDNA3 plasmid and sequenced.

The *Del 1 Δ18* minigene was made in two segments. The first part of the segment was made by using the forward primer 5'–CCACAGGATCCGCTTCTTCTTCTTCATG

TTGACTAAACCTCATG–3’ and the reverse primer 5’–CATTAACATAAACAGCAA GCAGAGGGTGCATTTTTGGGTAGG–3’. The second segment was made using the forward primer 5’–CCTCTGCTTGCTGTTTATGTTAATGCGCTTGAACC–3’ and the reverse primer 5’–ATTCTTATGCGGCCGCCAGATTCATTTATTAAGAAACCCCAC CCC–3’. The two segments were ligated using standard PCR techniques and was inserted into a pcDNA3 plasmid and sequenced.

The *Del 2 Δ16, Del 3 Δ18, Del 4 Δ19* and *Del 5 Δ18* minigenes were made using the PCR techniques described for the *Del 1 Δ18* minigene. All *Del* minigenes used the same forward primer for the first segment and the same reverse primer for the second segment. The *Del 2 Δ16* minigene used the reverse primer 5’–GGTTCAAGCGCATTA ACATGCATCATGGGGGCAGC–3’ for the first segment and the forward primer 5’–TG TTAATGCGCTTGAACCCCACTGGCCCATTGC–3’ for the second segment. The first segment of *Del 3 Δ18* minigene was made using the reverse primer 5’–CATGATGGCA ATGGGCCAGTGGTAAACAGCAAGCAGAGG–3’ the second segment was made using the forward primer 5’–CCACTGGCCCATTGCCATCATGTGCTCGC–3’. The first segment of the *Del 4 Δ19* minigene was made using the reverse primer 5’–GCAGGC AGCGAGCACATGGGTTCAAGCGCATTAAC–3’. The second segment of the *Del 4 Δ19* minigene was made using the forward primer 5’–CATGTGCTCGCTGCCTGCTAA TTAAGACTCAGTCGG C–3’. The first segment of the *Del 5 Δ18* minigene was made using the reverse primer 5’–GACAGCCGACTGAGTCTTAATTATGGCAATGGGCCA GTGG–3’. The second segment of the *Del 5 Δ18* minigene was made using the forward primer 5’–AATTAAGACTCAGTCGGCTGTCAAATCACTGAAGCGACCCC–3’.

Cell culture and transfection- HeLa cells were cultured and transfected as previously described [40] except for the following changes. 1X Antibiotic-Antimycotic (Invitrogen) was added to DMEM GLUTAMAX media (Invitrogen). Cells were harvested 16-24 hours after transfection.

In vivo splicing- Splicing assays were done as described in Murray et al., 2008 except for the following changes. All reporters were reverse transcribed using a pcDNA3 plasmid specific antisense primer 5'–AGCATTTAGGTGACACTATAGAATAGGG–3'. The –RT reactions were treated the same as the +RT reactions except that no SuperScriptII was added to the –RT reactions. The cDNA from the RT reaction (2μL) was subjected to 26 rounds of PCR (within linear range) in a 20μL reaction. PCR amplification for all splice products was done using the sense primer 5'– GATCAAGGC TGCCCAATACCAG –3'and the antisense primer 5'–ATTCTTATGCGGCCGCCAGA TTCATTTATTAAGAAACCCCACCCC–3'. The PCR products were resolved on a 6% native polyacrylamide gel (40% 19:1 acrylamide:bisacrylamide) using SYBR Green (Applied Biosystems). The SYBR Green was diluted 1000X in 6X dye. Quantification of bands was done using the Alpha Imager HP Software from Alpha Innotech. Percent exon inclusion was calculated by dividing the amount of the band indicating inclusion by the total amount of splice product (bands indicating inclusion and exclusion). Background was taken from space between the two bands.

Branchpoint mapping- Branchpoint mapping was done as described in Gao et al., 2008 except for the following changes [49]. MBNL1 wt minigene was transfected into HeLa cells, isolated and DNased as described above. Antisense primer C (5'-GAATAGCTTGTAGTCAGATATAGTTGCTC -3') was used to reverse transcribe

using Super Script III. Lariat PCR was done using primers C and sense primer D (5'-GCAGACTCTCTCCTCCTCTCTTCC – 3') and nested PCR was done using antisense primer A (5'- GCTTTTCTGACTGCTAACAAGGAGAGAGC -3') and sense primer B (5'- TAATTAACTACAAAGAGGAGTTATCCTCCC -3'). Nested lariat RT PCR products were purified using PCR Cleanup (Qiagen) and individual lariats were isolated by Topo cloning (Invitrogen) and sequenced.

Protein purification- The MBNL1 protein construct, which includes amino acids 1–260 and contains an N-terminal GST tag, was expressed and purified as described in Warf et al., 2007, except for the following differences [39]. For the lysis of bacterial cells, 10μg/mL of DNase was added with the lysozyme (1mg/mL) and the cell extract was centrifuged for 30 minutes at 15,000 rpm. Supernatant was loaded onto GST beads and washed with PBS-T buffer (1X PBS and 1%Triton X-100) and eluted with elution buffer (10mM reduced glutathione, 50mM Tris pH 9.5).

Transcription of unlabeled RNA- The RNA used for structure probing and footprinting was transcribed from DNA amplified from the MBNL1 pcDNA3 construct used to transfect HeLa cells using sense 5'–GATAATACGACTCACTATAGGGACAA CTCAGTAGTGCCTTTATTGTGCATGCTTAGTCTTGTTATTCGTTGTATATGGCA TTCCG –3' (T7 site and additional guanosines are underlined) and the antisense primer 5–GGGCTGCTGGGCTTTC–3'. To amplify the template, 35 rounds of PCR were done. The template was subjected to PCR cleanup using the PCR cleanup Qiagen kit. To transcribe, 100–500ng of DNA was added to each Transcription Reaction: 40mM Tris pH 8, 10mM $MgCl_2$, 2mM spermidine, 0.01%Triton, 1X rNTP, 10mM DTT, 1 unit of RNasin (Promega), 40% PEG and T7 DNA polymerase at 37°C for 1.5 to 2 hours. 1 unit

of DNase was added to the transcription and incubated at 37°C for 1 hour. The RNA sample underwent phenol chloroform and was run on an 11% denaturing gel (7.5M urea, 11% 29:1 acrylamide:bisacrylamide and 1X TBE). RNA was eluted using an Elutrap and the RNA was EtOH precipitated. The RNA pellet was brought up in low TE and quantified using the Qubit (Life Technologies). RNA was stored at –80°C.

Footprinting and Structure Probing with T1 and RNase 1- RNA, Structure Buffer, and heparin were snap annealed. The mixture was aliquoted and different concentrations of GST tagged MBNL1 were incubated with the RNA sample for 15 minutes at room temperature. Then 0.25U of T1 RNase or 0.3U of RNase 1 was added and the sample was incubated for an additional 15 minutes (structure probing experiments were done without MBNL1). A final concentration of 81nM RNA, 0.48µg of heparin in T1 RNase Structure Buffer (final concentration of 7.7mM Tris pH 7, 77mM KCl, 7.7mM of $MgCl_2$) and R1 RNase Structure Buffer (final concentration of 7.7Mm Tris pH 7, 77mM NaCl, 7.7mM $MgCl_2$) was EtOH precipitated by adding 20µl of inactivation buffer (0.7M NaOAC pH 5.2 in EtOH) and 1µL of 20mg/mL of glycogen. Samples were incubated on ice for 5 minutes, and spun at max speed for 15 minutes at 4°C. Pellets were resuspended in a final concentration of 18nM of radiolabeled reverse primer (5'–CAGGTCAAAGGTTGCCTCG–3') and 0.83mM dNTPs (the reverse primer was kinased using Polynucleotide Kinase and [γ-P32] ATP). Samples were incubated at 65°C for 5 minutes, 35°C for 5 minutes and on ice for one minute. Reverse transcription was carried out by adding appropriate amounts of 5X First Strand buffer, 0.1M DTT, and 200U of SuperScript III. Samples were incubated at 52°C for 1 hour and at 70C for 15 minutes. Samples were then phenol chloroformed, EtOH precipitated and the pellet was

resuspended in 20µL of low TE. An equal volume of 2X Denaturing dye was added to the samples, which were then incubated at 95°C for 2 minutes and run on an 8% denaturing gel (8% 19:1 acrylamide:bisacrylamide, 1X TBE and 7.5M urea)

The SHAPE (Selective 2' Hydroxyl Acylation by Primer Extension) assay was done as described above, except RNA and heparin were snap annealed in HE buffer (10mM HEPES pH 8.0 and 1mM EDTA pH 8.0). 1X Folding mix buffer (Recipe for 3X buffer is 333mM HEPES pH 8.0, 20mM $MgCl_2$, 333mM NaCl) was added and the RNA sample was incubated at 37°C for 20 minutes. RNA samples were aliquoted and 1µL NMIA in neat DMSO were added to samples. Neat DMSO was added to a sample as a control (free RNA). Reactions were incubated at 37°C for 45 minutes. Inactivation buffer was added and reverse transcription was done as described above.

Quantification and normalization of T1, RNase 1 and SHAPE gels- Footprinting and structure probing data was quantified using the SAFA program [52]. Data was normalized for each lane as described in Low et al. 2010 [53]. The sequence and text files were imported to the RNAStructure programs (available from http://rna.urmc.rochester.edu/rnastructure.html) [54]. The slope (m) and intercept (b) chosen were 2.6kcal/mol and –0.8kcal/mol, respectively.

Difference plots were made by subtracting normalized reactivities of the protein lane from the no protein lane. The average of the difference was then added and the values were plotted. The sequence and SHAPE text files were prepared as described in Low et al., 2010.

Discussion

PTB1 and MBNL1 negatively regulate splicing of their pre-mRNAs through introns containing distant branchpoints- PTB1 (polypyrimidine tract binding protein 1), also known as hnRNP I, is an alternative splicing factor that regulates many different splicing events [55]. Like MBNL1 and MBNL2, this factor also auto-regulates the splicing of its own pre-mRNA through the usage of a predicted distant branchpoint [56]. The distant branchpoint is contained within a 351 nucleotide AGEZ in intron 10 of the *PTB1* pre-mRNA [56]. The auto-regulation of *PTB1* leads to exon 11 exclusion, resulting in a premature termination codon that is predicted to induce the NMD (nonsense mediated decay) pathway. This auto-regulation of splicing allows PTB1 to tightly control its own protein levels. MBNL1 and MBNL2 differ in that their auto-regulation leads to different protein isoforms. Presumably, these different isoforms of MBNL1 and MBNL2 have different functions, but currently the major known difference between the isoforms is that those lacking exon 5 are found in both the nucleus and cytoplasm while the isoforms containing exon 5 are primarily nuclear [46].

A non-canonical intron in the *MBNL1* pre-mRNA- The 3' end of MBNL1 intron 4 is different than the 3' end of most other introns. First, the 3' end of intron 4 is contained within an ultra-conserved element longer than 200 nucleotides, and is one hundred percent conserved between human, rat and mouse genomes [44]. Second, this intron is unique because it contains a distant branchpoint and an AAG 3' splice site. Most human introns contain a predicted branchpoint in the last 40 nucleotides of the intron and a YAG 3' ss. In the canonical intron architecture, U2AF35 binds the 3' ss, U2AF65 binds the PY tract and U2 snRNP binds at the branchpoint [57]. It is not clear how these factors

recognize introns containing distant branchpoint sequences. However, it has been suggested that the second step in splicing may involve a mechanism in which the spliceosome performs a linear scan downstream from the distant branchpoint and PY tract until it reaches the first AG dinucleotide [58, 59]. Antibodies to the polypyrimidine tract binding protein were used to analyze components that assemble on alternatively spliced pre-mRNAs that use a distant branchpoint [60]. Out of two pre-mRNAs studied (α and *β-TM*), the group identified a common set of uncharacterized proteins, in addition to PTB1, that assemble on alternatively spliced pre-mRNAs with distant branchpoint sequences. However, it is unclear how these proteins aid in alternative splicing [60].

One possible mechanism is that the 141 nucleotide RNA linker contains binding sites for proteins that interact with U2AF35, U2AF65 and U2 snRNP that facilitate their binding and spliceosome formation. Alternatively, this RNA linker may be sequestered out of the way (likely bound by hnRNPs) in some manner (Figure 15A). It has been shown that the presence of a stem-loop between a distant branchpoint and 3' ss inhibits the second step of splicing [61]. It is possible that like the stem-loop, MBNL1 binding to this region of the intron results in a structure that blocks scanning to the 3' ss (Figure 15B). The inability to splice this pre-mRNA *in vitro* has blocked our efforts to determine at which step MBNL1 is regulating splicing.

A recent bioinformatics study suggests that highly conserved regions are actually less structured compared to other regions of the pre-mRNA [62] and our structure probing data for *MBNL1* is consistent with this hypothesis. It is possible that this lack of structure within highly conserved regions make them more accessible to splicing factors

for regulation, and if these sites are highly regulated by multiple splicing factors, then higher conservation seems plausible.

MBNL1 regulates *MBNL1* exon 5 through a dissimilar mechanism than CLCN1 exon 7A and TNNT2 exon 5- Studies of where MBNl1 binds its pre-mRNA targets to regulate alternative splicing suggests that MBNL1 may not regulate exon exclusion through a conserved mechanism. In intron 4 of the *MBNL1* pre-mRNA, we observed that MBNL1 regulates exon 5 exclusion by binding a response element located within an AGEZ just downstream of a distant branchpoint and PY tract. In the *CLCN1* pre-mRNA, MBNL1 represses exon 7A inclusion by binding the 5' end of exon 7A (which contains an ESE) and flanking intronic regions [27]. Previously, we showed that MBNL1 and U2AF65 compete for binding at the 3' end of intron 4 of the *TNNT2* pre-mRNA to regulate exon 5. MBNL1 regulated *TNNT2* exon 5 exclusion may involve MBNL1 binding the RNA in a looped conformation based on a model from a crystal structure of MBNL1 in complex with a short RNA, while U2AF65 interacts with the RNA in a single-stranded conformation [29, 31].

The binding of multiple MBNL1 proteins to the *MBNL1* intron 4 linker may result in the RNA adopting a conformation that inhibits formation of a functional splicing complex at this 3' ss. Footprinting and structure probing data shows that MBNL1 binds a mostly unstructured part of the RNA (nucleotides 137-162) and causes more cleavage of nucleotides just downstream (nucleotides 167-201), suggesting that upon binding, MBNL1 may cause downstream RNA to become unstructured allowing more MBNL1 proteins to bind (Figure 12C). In the presence of MBNL1, the 3'ss of intron 4 is not accessible to the spliceosome, resulting in the exclusion of exon 5 (Figure 15B). The

MBNL1 response element doesn't overlap with the branchpoint, PY tract or 3' splice site; therefore, direct competition between MBNL1 and constitutive splicing factors doesn't appear to be a likely mechanism for regulation by MBNL1. It is possible that additional MBNL1 proteins bind outside of the response element to compete with constitutive splicing factors or alternatively the presence of MBNL1 blocks the ability of the splicing machinery to locate (via scanning) the 3' ss from the distant branchpoint.

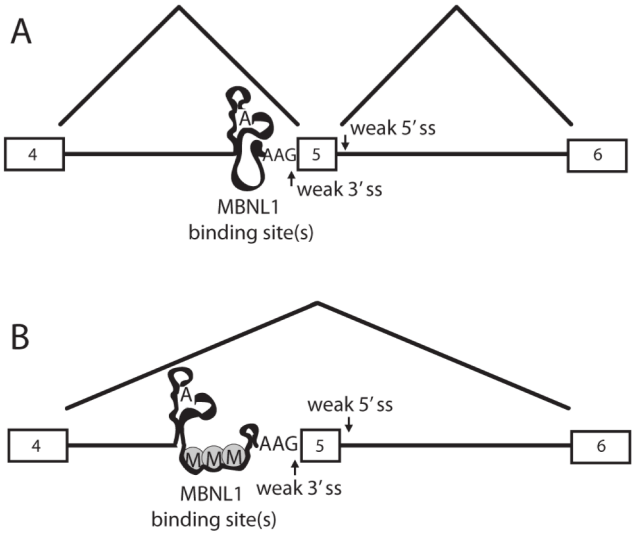

Figure 15. Model for MBNL1 binding and excluding exon 5. (A) Intron 4 of the *MBNL1* pre-mRNA contains a distant branchpoint, PY tract and a 141 nucleotide AGEZ that contains an MBNL1 response element. When MBNL1 is not present to regulate alternative splicing, ISEs (intronic splicing enhancers) and ESEs (exonic splicing enhancers), along with other proteins, regulate exon 5 inclusion. (B) MBNL1 binds the 90 nucleotide regulatory region in the AGEZ to regulate exon 5 exclusion. By binding within the AGEZ, MBNL1 appears to inhibit the ability of the spliceosome to locate the 3' ss.

CHAPTER V

CONCLUSIONS AND FUTURE DIRECTIONS

Conclusions to Chapter II

Results from the Mbl SELEX study discussed in Chapter II showed that the sequence AGUCU was enriched in the SELEX RNAs. Characterization of SELEX RNA #20 showed that Mbl required both stem-loops to bind and truncation of stem-loop 1 could be tolerated. Removal of either the 5' or 3' tail of SELEX RNA #20 resulted in loss of binding by Mbl. However, since the AGUCU motif in Class I and Class II RNAs bound the constant regions the SELEX may have been biased and the AGUCU and additional studies will have to be done to search for a biologically relevant RNA sequence that Mbl binds. However, the binding studies that were done to characterize the different SELEX RNAs show that Mbl can bind structured RNAs. Determining whether MBNL1 can bind the SELEX RNAs (that contain one or more AGUCU motifs) and the RNA #20 mutants would provide insight as to whether MBNL1 and Mbl recognize RNA in a similar manner.

Conclusions to Chapter III

Round 5 SELEX sequences were sequenced and a common YGCY motif was identified. The sequences were further organized into two classes, Class I, which had one or more YGCY motifs and Class II, which had zero to one YGCY motif. Interestingly, MBNL1 bound Class I RNAs with higher affinity than class II RNAs.

The YGCY motif was used to determine potential MBNL1 binding sites within the pre-mRNAs that are mis-spliced in myotonic dystrophy. We searched 200 nucleotides upstream, downstream and even within the regulated exon of the pre-mRNA

for YGCY motifs in the 24 pre-mRNA transcripts shown to be mis-spliced in myotonic dystrophy. We choose 5 pre-mRNA transcripts out of the 24 to study in more detail. 40-45 nucleotide regions of the *MBNL1, MBNL2, SERCA1, GRIN1* and *INSR* pre-mRNAs that contained YGCY motifs were studied.

RNAs that bound MBNL1 with high affinity not only had one or more YGCY motifs but also contained a high percentage of pyrimidines. Closer observation of *MBNL1* site #1 revealed that the it lied within an ultraconserved element upstream of the MBNL1 regulated exon (exon 5). Mutational analysis of *MBNL1* site #1 revealed that there are probably many MBNL1 binding sites upstream of exon 5 within the MBNL1 pre-mRNA. Although simple mutagenesis abolishes MBNL1's ability to bind the 45 nucleotide *MBNL1 CA_CAP in vitro*, the *MBNL1 CA_CAP* mutant does not affect MBNL1's ability to regulate exon 5 exclusion *in vivo*. This suggests that MBNL1 may regulate exon 5 exclusion by binding multiple YGCY motifs within intron 4 outside of *MBNL1* site#1.

Further characterization of the endogenous potential MBNL1 binding sites will provide insight as to where MBNL1 binds its pre-mRNA targets to regulate alternative splicing. Observations from this type of study may lend more support to the hypothesis that MBNL1 and U2AF65 compete for a binding site, or may suggest alternative mechanisms by which MBNL1 may regulate its pre-mRNA targets.

Conclusions to Chapter IV

The identification of two potential binding sites upstream of exon 5 in the *MBNL1* pre-mRNA prompted a series of experiments to determine whether MBNL1 regulates the alternative splicing of it's own pre-mRNA through these sites. We observed that the 3'

end of intron 4 is non-canonical in that it contains an AG exclusion zone and an AAG 3' ss.

To determine whether the ultraconserved region between the potential PY tract and 3' splice site was important for splicing, a mutant *MBNL1* minigene was made where 90 nucleotides were deleted between these two regions. This minigene (*Δ90*) was designed so that the potential PY tract would only be 28 nucleotides away from the 3' splice site. The *Δ90* minigene was transfected into HeLa cells and exon 5 inclusion was seen as it is seen in wild type cells (Figure 12). When the *Δ90* minigene is transfected into HeLa cells and MBNL1 is overexpressed, MBNL1 is no longer able to regulate exclusion of exon 5. Interestingly, the deletion contains the area in which there was a footprint upstream of exon 5. Smaller deletions within the 90 nucleotide MBNL1 response element did not obliterate MBNL1's ability to regulate exon 5 exclusion. This observation is consistent with the *in vivo* slicing of the *MBNL1 CA-CAP* mutant discussed in Chapter III, where the mutation of a handful of YGCY motifs does not affect MBNL1's ability to regulate exon 5 exclusion.

Further characterization by footprinting and structure probing of *MBNL1* intron 4 showed that intron 4 contains a mostly structured distant bp and PY tract and a largely unstructured MBNL1 response element. These results confirm that MBNL1 can bind unstructured, single stranded RNA. Footprinting and branchpoint mapping results have led to the hypothesis that MBNL1 may not regulate exon 5 exclusion by competing with the splicing factor U2AF65 as was suggested for the TNNT2 pre-mRNA (Warf et al., 2009). This is because the YGCY motifs (MBNL1 binding sites) do not overlap with the distant polypyrimidine tract, however U2AF65 binding to *MBNL1* intron 4 has yet to be

studied.

It has been shown that MBNL1 requires a 90 nucleotide sequence upstream of exon 5 in the *MBNL1* pre-mRNA to regulate exon exclusion. It will be interesting to determine if this 90 nucleotide sequence is sufficient for MBNL1 regulation. This could be done by placing the 90 nucleotide element into a heterologous pre-mRNA that is not regulated by MBNL1. If MNBL1 regulates this heterologous pre-mRNA this demonstrates that this elements contains all of the necessary signals for MBNL1 splicing regulation.

Once a minimal RNA sequence required for MBNL1 dependent regulation is found, this construct can be incubated with HeLa nuclear extract to determine whether MBNL1 inhibits a particular complex (E, A, B or C) formation by binding the intron 4.

REFERENCES CITED

1. Wang, E.T., et al., *Alternative isoform regulation in human tissue transcriptomes.* Nature, 2008. **456**(7221): p. 470-6.

2. Pan, Q., et al., *Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing.* Nat Genet, 2008. **40**(12): p. 1413-5.

3. Staley, J.P. and C. Guthrie, *Mechanical devices of the spliceosome: motors, clocks, springs, and things.* Cell, 1998. **92**(3): p. 315-26.

4. Green, M.R., *Pre-mRNA splicing.* Annu Rev Genet, 1986. **20**: p. 671-708.

5. Mankodi, A., et al., *Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat.* Science, 2000. **289**(5485): p. 1769-73.

6. Ranum, L.P. and T.A. Cooper, *RNA-mediated neuromuscular disorders.* Annu Rev Neurosci, 2006. **29**: p. 259-77.

7. Jansen, G., et al., *Abnormal myotonic dystrophy protein kinase levels produce only mild myopathy in mice.* Nat Genet, 1996. **13**(3): p. 316-24.

8. Klesert, T.R., et al., *Mice deficient in Six5 develop cataracts: implications for myotonic dystrophy.* Nat Genet, 2000. **25**(1): p. 105-9.

9. Davis, B.M., et al., *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, 1997. **94**(14): p. 7388-93.

10. Amack, J.D., A.P. Paguio, and M.S. Mahadevan, *Cis and trans effects of the myotonic dystrophy (DM) mutation in a cell culture model.* Hum Mol Genet, 1999. **8**(11): p. 1975-84.

11. Pascual, M., et al., *The Muscleblind family of proteins: an emerging class of regulators of developmentally programmed alternative splicing.* Differentiation, 2006. **74**(2-3): p. 65-80.

12. Cooper, T.A., *Chemical reversal of the RNA gain of function in myotonic dystrophy.* Proc Natl Acad Sci U S A, 2009. **106**(44): p. 18433-4.

13. Begemann, G., et al., *muscleblind, a gene required for photoreceptor differentiation in Drosophila, encodes novel nuclear Cys3His-type zinc-finger-containing proteins.* Development, 1997. **124**(21): p. 4321-31.

14. Artero, R., et al., *The muscleblind gene participates in the organization of Z-bands and epidermal attachments of Drosophila muscles and is regulated by Dmef2.* Dev Biol, 1998. **195**(2): p. 131-43.

15. Philips, A.V., L.T. Timchenko, and T.A. Cooper, *Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy.* Science, 1998. **280**(5364): p. 737-41.

16. Savkur, R.S., A.V. Philips, and T.A. Cooper, *Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy.* Nat Genet, 2001. **29**(1): p. 40-7.

17. Kanadia, R.N., et al., *Reversal of RNA missplicing and myotonia after muscleblind overexpression in a mouse poly(CUG) model for myotonic dystrophy.* Proc Natl Acad Sci U S A, 2006. **103**(31): p. 11748-53.

18. Osborne, R.J. and C.A. Thornton, *RNA-dominant diseases.* Hum Mol Genet, 2006. **15 Spec No 2**: p. R162-9.

19. Timchenko, L.T., et al., *Identification of a (CUG)n triplet repeat RNA-binding protein and its expression in myotonic dystrophy.* Nucleic Acids Res, 1996. **24**(22): p. 4407-14.

20. Timchenko, L.T., et al., *Novel proteins with binding specificity for DNA CTG repeats and RNA CUG repeats: implications for myotonic dystrophy.* Hum Mol Genet, 1996. **5**(1): p. 115-21.

21. Du, H., et al., *Aberrant alternative splicing and extracellular matrix gene expression in mouse models of myotonic dystrophy.* Nat Struct Mol Biol, 2010. **17**(2): p. 187-93.

22. Koshelev, M., et al., *Heart-specific overexpression of CUGBP1 reproduces functional and molecular abnormalities of myotonic dystrophy type 1.* Hum Mol Genet, 2010. **19**(6): p. 1066-75.

23. Miller, J.W., et al., *Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy.* EMBO J, 2000. **19**(17): p. 4439-48.

24. Squillace, R.M., D.M. Chenault, and E.H. Wang, *Inhibition of muscle differentiation by the novel muscleblind-related protein CHCR.* Dev Biol, 2002. **250**(1): p. 218-30.

25. Adereth, Y., et al., *RNA-dependent integrin alpha3 protein localization regulated by the Muscleblind-like protein MLP1.* Nat Cell Biol, 2005. **7**(12): p. 1240-7.

26. Fardaei, M., et al., *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, 2002. **11**(7): p. 805-14.

27. Kino, Y., et al., *MBNL and CELF proteins regulate alternative splicing of the skeletal muscle chloride channel CLCN1.* Nucleic Acids Res, 2009. **37**(19): p. 6477-90.

28. Vicente, M., et al., *Muscleblind isoforms are functionally distinct and regulate alpha-actinin splicing.* Differentiation, 2007. **75**(5): p. 427-40.

29. Teplova, M. and D.J. Patel, *Structural insights into RNA recognition by the alternative-splicing regulator muscleblind-like MBNL1.* Nat Struct Mol Biol, 2008. **15**(12): p. 1343-51.

30. Grammatikakis, I., et al., *Identification of MBNL1 and MBNL3 domains required for splicing activation and repression.* Nucleic Acids Res, 2010.

31. Warf, M.B., et al., *The protein factors MBNL1 and U2AF65 bind alternative RNA structures to regulate splicing.* Proc Natl Acad Sci U S A, 2009. **106**(23): p. 9203-8.

32. Brook, J.D., et al., *Localisation of genetic markers and orientation of the linkage group on chromosome 19.* Hum Genet, 1984. **68**(4): p. 282-5.

33. Fu, Y.H., et al., *An unstable triplet repeat in a gene related to myotonic muscular dystrophy.* Science, 1992. **255**(5049): p. 1256-8.

34. Mahadevan, M., et al., *Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene.* Science, 1992. **255**(5049): p. 1253-5.

35. Liquori, C.L., et al., *Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9.* Science, 2001. **293**(5531): p. 864-7.

36. Houseley, J.M., et al., *Myotonic dystrophy associated expanded CUG repeat muscleblind positive ribonuclear foci are not toxic to Drosophila.* Hum Mol Genet, 2005. **14**(6): p. 873-83.

37. de Haro, M., et al., *MBNL1 and CUGBP1 modify expanded CUG-induced toxicity in a Drosophila model of myotonic dystrophy type 1.* Hum Mol Genet, 2006. **15**(13): p. 2138-45.

38. Ho, T.H., et al., *Muscleblind proteins regulate alternative splicing.* EMBO J, 2004. **23**(15): p. 3103-12.

39. Warf, M.B. and J.A. Berglund, *MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T.* RNA, 2007. **13**(12): p. 2238-51.

40. Goers, E.S., et al., *MBNL1 binds GC motifs embedded in pyrimidines to regulate alternative splicing.* Nucleic Acids Res, 2010. **38**(7): p. 2467-84.

41. Murray, J.I., et al., *Identification of motifs that function in the splicing of non-canonical introns.* Genome Biol, 2008. **9**(6): p. R97.

42. Wahl, M.C., C.L. Will, and R. Luhrmann, *The spliceosome: design principles of a dynamic RNP machine.* Cell, 2009. **136**(4): p. 701-18.

43. Charlet, B.N., et al., *Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing.* Mol Cell, 2002. **10**(1): p. 45-53.

44. Bejerano, G., et al., *Ultraconserved elements in the human genome.* Science, 2004. **304**(5675): p. 1321-5.

45. Kalsotra, A., et al., *A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart.* Proc Natl Acad Sci U S A, 2008. **105**(51): p. 20333-8.

46. Terenzi, F. and A.N. Ladd, *Conserved developmental alternative splicing of muscleblind-like (MBNL) transcripts regulates MBNL localization and activity.* RNA Biol, 2010. **7**(1): p. 43-55.

47. Pedersen, J.S., et al., *Identification and classification of conserved RNA secondary structures in the human genome.* PLoS Comput Biol, 2006. **2**(4): p. e33.

48. Gooding, C., et al., *A class of human exons with predicted distant branch points revealed by analysis of AG dinucleotide exclusion zones.* Genome Biol, 2006. **7**(1): p. R1.

49. Gao, K., et al., *Human branch point consensus sequence is yUnAy.* Nucleic Acids Res, 2008. **36**(7): p. 2257-67.

50. Wilkinson, K.A., E.J. Merino, and K.M. Weeks, *Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution.* Nat Protoc, 2006. **1**(3): p. 1610-6.

51. Ho, T.H., et al., *Colocalization of muscleblind with RNA foci is separable from mis-regulation of alternative splicing in myotonic dystrophy.* J Cell Sci, 2005. **118**(Pt 13): p. 2923-33.

52. Das, R., et al., *SAFA: semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments.* RNA, 2005. **11**(3): p. 344-54.

53. Low, J.T. and K.M. Weeks, *SHAPE-directed RNA secondary structure prediction.* Methods, 2010. **52**(2): p. 150-8.

54. Reuter, J.S. and D.H. Mathews, *RNAstructure: software for RNA secondary structure prediction and analysis.* BMC Bioinformatics, 2010. **11**: p. 129.

55. Sawicka, K., et al., *Polypyrimidine-tract-binding protein: a multifunctional RNA-binding protein.* Biochem Soc Trans, 2008. **36**(Pt 4): p. 641-7.

56. Wollerton, M.C., et al., *Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay.* Mol Cell, 2004. **13**(1): p. 91-100.

57. Black, D.L., *Mechanisms of alternative pre-messenger RNA splicing.* Annu Rev Biochem, 2003. **72**: p. 291-336.

58. Langford, C.J. and D. Gallwitz, *Evidence for an intron-contained sequence required for the splicing of yeast RNA polymerase II transcripts.* Cell, 1983. **33**(2): p. 519-27.

59. Mount, S.M., *A catalogue of splice junction sequences.* Nucleic Acids Res, 1982. **10**(2): p. 459-72.

60. Grossman, J.S., et al., *The use of antibodies to the polypyrimidine tract binding protein (PTB) to analyze the protein components that assemble on alternatively spliced pre-mRNAs that use distant branch points.* RNA, 1998. **4**(6): p. 613-25.

61. Smith, C.W., T.T. Chu, and B. Nadal-Ginard, *Scanning and competition between AGs are involved in 3′ splice site selection in mammalian introns.* Mol Cell Biol, 1993. **13**(8): p. 4939-52.

62. Sathirapongsasuti, J.F., et al., *Ultraconserved cDNA segments in the human transcriptome exhibit resistance to folding and implicate function in translation and alternative splicing.* Nucleic Acids Res, 2010.