INVESTIGATING THE RNA BINDING DOMAINS OF MBNL1 AND THE ALTERNATIVE SPLICING MOTIFS THEY RECOGNIZE

by

JAMIE PURCELL

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

March 2012

DISSERTATION APPROVAL PAGE

Student: Jamie Purcell

Title: Investigating the RNA Binding Domains of MBNL1 and the Alternative Splicing Motifs They Recognize

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:

Diane Hawley	Chairperson
J. Andrew Berglund	Advisor
Bradley Nolen	Member
Tom Stevens	Member
Alice Barkan	Outside Member

and

Kimberly Andrews Espy	Vice President for Research & Innovation/Dean of the Graduate School

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

Degree awarded March 2012

DISSERTATION ABSTRACT

Jamie Purcell

Doctor of Philosophy

Department of Chemistry

March 2012

Title: Investigating the RNA Binding Domains of MBNL1 and the Alternative Splicing Motifs They Recognize

Muscleblind-like 1 (MBNL1) is a ubiquitously expressed RNA binding protein that regulates the alternative splicing of a variety of transcripts. In Myotonic Dystrophy (DM) aberrant cellular localization of MBNL1 results in disease-associated mis-splicing of several MBNL1 target pre-mRNAs. Due to its role in DM pathogenesis, MBNL1 has been a topic of intense study for the last decade, however many open mechanistic questions remain regarding how MBNL1 recognizes RNA substrates to mediate splicing.

The RNA recognition motif for MBNL1, 5'-YGCY-3', was defined herein. This motif was used to identify novel MBNL1 binding sites within regulated transcripts and create synthetic MBNL1-regulated splicing reporters. MBNL1 contains four zinc finger (ZF) RNA binding domains arranged into two pairs of two ZFs. A comprehensive, combinatorial mutagenic study of MBNL1 was conducted to determine the role of each ZF in RNA binding and splicing activity. Functional analysis of the mutant proteins in cellular splicing assays and assessment of RNA binding activity demonstrated that the ZF pairs (*i.e.* ZF1-2 or ZF3-4) do not have equivalent activity. The ZF1-2 pair is responsible for MBNL1's high affinity RNA binding and splicing activity, whereas the ZF3-4 pair has reduced affinity for RNA and impaired ability to regulate splicing of some transcripts. Hierarchical clustering analysis revealed that two distinct classes of MBNL1-

regulated splicing events exist within the small set of splicing events examined. For Class II splicing events the binding and splicing activity for the ZF mutants correlated well. However, for Class I events there was no significant correlation between RNA binding and splicing activity. For pre-mRNAs in the latter class it appears that MBNL1 exerts surprisingly robust splicing activity in the absence of strong RNA binding, suggesting that MBNL1 may be recruited to some pre-mRNA substrates through protein-protein interactions. This study provides the first demonstration that functionally distinct classes of MBNL1-mediated splicing events exist in terms of requirements for different ZFs and the importance of RNA binding.

This dissertation includes previously published and unpublished co-authored material as well as recently co-authored material that has been submitted for publication.

CURRICULUM VITAE

NAME OF AUTHOR: Jamie Purcell

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, OR
Emory University, Atlanta, GA
Utah State University, Logan, UT

DEGREES AWARDED:

Doctor of Philosophy, Chemistry, 2012, University of Oregon
Master of Science, Chemistry, 2005, Utah State University
Bachelor of Science, Chemistry, 2004, Utah State University

AREAS OF SPECIAL INTEREST:

Biochemistry
Molecular Biology
Alternative Splicing

PROFESSIONAL EXPERIENCE:

Graduate Research Fellow, Department of Chemistry, Dr. J. Andrew Berglund, University of Oregon, 2008-2012

Graduate Research Fellow, Graduate Division of Biological and Biomedical Sciences, Dr. Dennis Liotta, Emory University, 2005-2008

Graduate Research Fellow, Department of Chemistry and Biochemistry, Dr. Alvan Hengge, Utah State University, 2004-2005

Undergraduate Research Assistant, Department of Chemistry and Biochemistry, Dr. Alvan Hengge, Utah State University, 2001-2004

GRANTS, AWARDS, AND HONORS:

NIH Training Grant, Institute of Molecular Biology (Molecular Biology and Biophysics), Department of Chemistry, University or Oregon, 2008-2011

NIH Training Grant, Biochemistry, Cell and Developmental Biology Program, Graduate Division of Biological and Biomedical Sciences, Emory University, 2005-2007

American Institute of Chemists Foundation Undergraduate Award, Utah State University Department of Chemistry and Biochemistry, 2004

Undergraduate Research and Creative Opportunities Grant, Utah State University, Vice Provost, 2003

Undergraduate Research Summer Internship Grant, Utah State University, Department of Chemistry and Biochemistry, 2002

Presidential Academic Scholarship, Utah State University, 2000-2004

Society of Women in Engineering Scholarship, 2000

PUBLICATIONS:

Purcell, J., Oddo, J.C., Wang, E.T. and Berglund, J.A. Combinatorial Mutagenesis of MBNL1 Zinc Fingers Elucidate Distinct Classes of Regulatory Events. 2012 (*Manuscript in Preparation*)

Thomas, S.L., Zhao J., Li Z., Lou B., Du Y., Purcell J., Snyder J.P., Khuri F.R., Liotta D., Fu, H. Activation of the p38 pathway by a novel monoketone curcumin analog, EF24, suggests a potential combination strategy. *Biochem. Pharmacol*. **2010** 80(9): 1309-1316.

*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): 2467-2484. (* Authors contributed equally to this work)

*Catrina, I., *O'Brien, P.J., *Purcell, J., Nikolic-Hughes, I., Zalatan, J.G., Hengge, A.C., and Herschlag, D. Probing the origin of the compromised catalysis of *E. coli* alkaline phosphatase in its promiscuous sulfatase reaction. *J. Am. Chem. Soc.* **2007** 129(17): 5760-5765. (* Authors contributed equally to this work)

Purcell, J. and Hengge, A.C. The thermodynamics of phosphate versus phosphorothioate ester hydrolysis. *J. Org. Chem.* **2005** 70(21): 8437-8442.

Onyido, I., Swierczek, K., Purcell, J. and Hengge, A.C. A concerted mechanism for the transfer of the thiophosphinoyl group from aryl dimethylphosphinothioate esters to oxyanionic nucleophiles in aqueous solution. *J. Am. Chem. Soc.* **2005** 127(21): 7703-7711.

Grzyska, P.K., Czyryca, P.G., Purcell, J. and Hengge, A. C. Transition state differences in hydrolysis reactions of alkyl versus aryl phosphate monoester monoanions.*J. Am. Chem. Soc.* **2003** 125(43): 13106-13111.

ACKNOWLEDGMENTS

I would like to thank Andy Berglund and members of the Berglund lab, past and present, for their help, guidance, and scientific discussion over the years. I also want to express gratitude to my friends and family for their support. I am especially grateful for the support and encouragement that my husband has provided throughout my graduate career. I am also appreciative of funding provided by the Institute of Molecular Biology training grant (GM-07759), the Institute of Molecular Biology support staff, and for guidance provided by my graduate committee.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I
INTRODUCTION

CONSTITUTIVE PRE-MRNA SPLICING

The central dogma of molecular biology dictates that genetic information is passed from DNA to transient RNA messages that are translated into proteins. In eukaryotic cells, the transmission of genetic data from gene to protein is highly regulated, particularly at the RNA level. Removal of non-coding, intronic content from the pre-messenger RNA via splicing is an essential process for creating mature messages that are competent for translation.

Introns are removed from pre-mRNA by the spliceosome, which catalyzes intron removal via two sequential transesterification reactions to create mature mRNA. The major spliceosome core consists of five small nuclear ribonucleoproteins (snRNPs) named U1, U2, U4, U5, and U6 (1), (2). A mature spliceosome contains the core splicing machinery along with more than a 150 accessory proteins (3), (4), making it one of the most complex macromolecular machines in the cell. Each snRNP contains an RNA component (snRNA) that is necessary for base pairing with specific sequences within the pre-mRNA to direct step-wise spliceosome assembly at exon/intron junctions. In vertebrates, pre-mRNA splicing is based on the recognition of four canonical intron motifs: the 5'-splice site (5'ss), branchpoint sequence (BPS), poly-pyrimidine tract (py-tract), and 3'-splice site (3'ss) (Figure 1A). The 5'-splice site is defined by a 6-nt consensus motif with the sequence 5'-GURAGU-3' (where R is either A or G), which marks the 5'-end of the intron (5), (6). The consensus BPS for mammals is 5'-CURAY-3' (7), (6). The adenosine at position 4 of the BPS initiates the first catalytic step of intron removal by attacking the 5'-splice site (8). The 3'-end of the intron is defined by a run of pyrimidine rich sequence and terminates with 5'-YAG-3' at the next intron/exon boundary.

Spliceosomal formation proceeds through four distinct complexes (Figure 1B) (9). Spliceosome assembly is initiated by U1 snRNP binding to the 5'-splice site via base pairing interactions between the splice site and the U1 snRNA component. At the 3'-end of the intron, branchpoint binding protein, or SF1, binds to the BPS and the large (65

kDa) and small (35 kDa) subunits of the dimeric auxiliary factor, U2AF65/35, bind the py-tract (10), resulting in formation of the early splicing complex (E-complex) (Figure 1B). Following E-complex, U2AF65 recruits the U2 snRNP to the branchpoint in an ATP-dependent process (11), (12) to form A-complex (Figure 1B). Next, the U4/U5/U6 tri-snRNP displaces U1 from the 5'ss to form B-complex (13). Finally, B-complex undergoes a complicated rearrangement and U4 is lost to form the catalytic, C-complex (14). It is within C-complex that the two chemical steps of splicing occur resulting in intron removal as a lariat and ligation of the remaining exons.

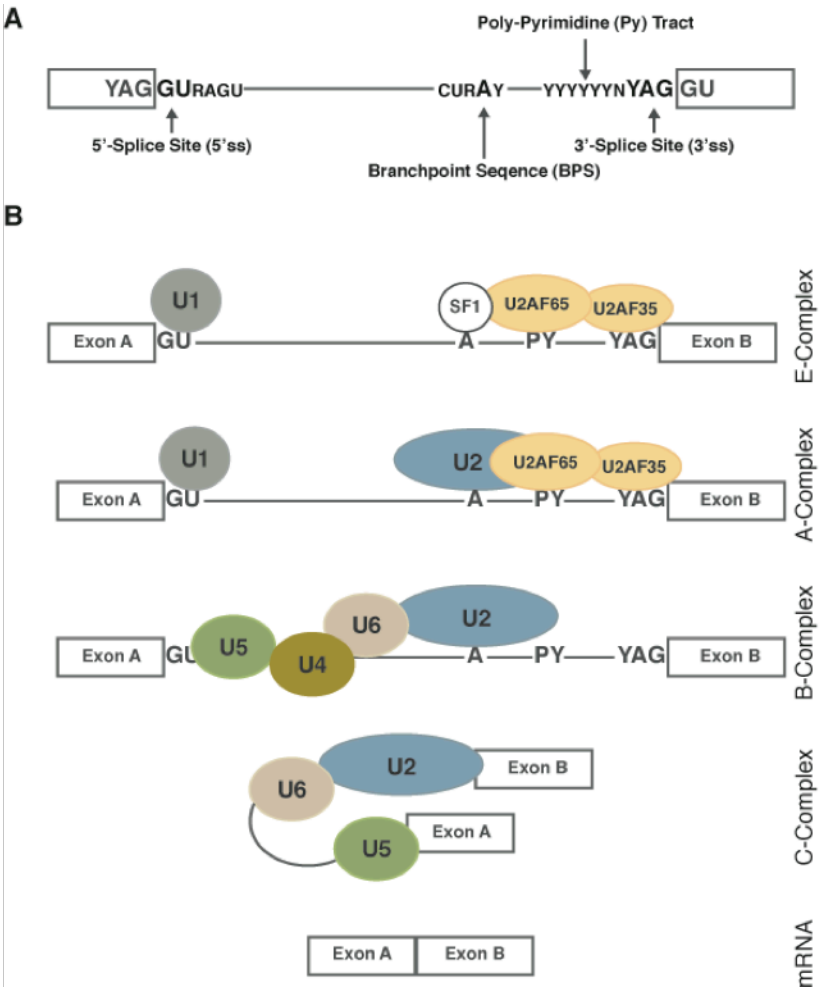


Figure 1. Model of spliceosome assembly. (A) Diagram depicting consensus motifs for the four, main intronic sequence elements. "Y" denotes C or U, "R" denotes A or G, and "N" represents any nucleotide. (B) Schematic depicting sequential spliceosome assembly on a pre-mRNA from E-complex (top) to the final mature mRNA (bottom). Spliceosomal components associated with the transcript throughout spliceosome assembly and catalysis, and their corresponding binding positions, are labeled. In panels A and B exons are depicted with boxes and introns with lines.

ALTERNATIVE SPLICING

Alternative splicing allows for one, multi-exon pre-mRNA to give rise to several mature mRNAs depending on which exons are included or excluded in the final transcript. Nature has utilized alternative splicing to enrich the proteomic diversity of higher order eukaryotes by creating millions of possible gene products from less than 30,000 genes. Recent approximations suggest that more than 90% of human genes are alternatively spliced (15). Alternative splicing plays an essential role in modulating numerous cellular processes, including: development, sex determination, tissue specific protein expression, apoptosis, cell excitation and contraction, and cellular differentiation (2).

There are several different ways that alternative splicing of an exon can occur (Figure 2) and most transcripts contain multiple alternative exon events. Cassette exons, which are the most common type of alternative splicing event, occur when a particular exon is either included or excluded in the final splice isoform. Alternative 5'- and 3'-splice site usage events are also common and result in lengthening or shortening of exons. In pre-mRNAs containing mutually exclusive exons, several coding exons are present but only one is included in the final transcripts. The proteins encoded by various splice isoforms may have small peptide changes that affect enzymatic properties, such as ligand binding or allostery, or protein localization or large changes such as entire domain deletions or insertions (2).

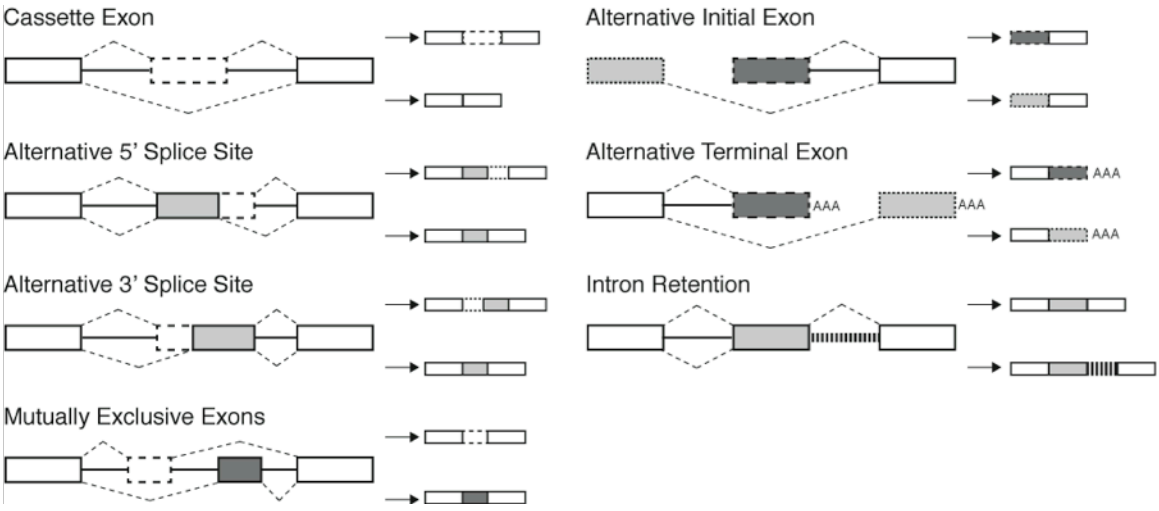


Figure 2. Common alternative splicing events.

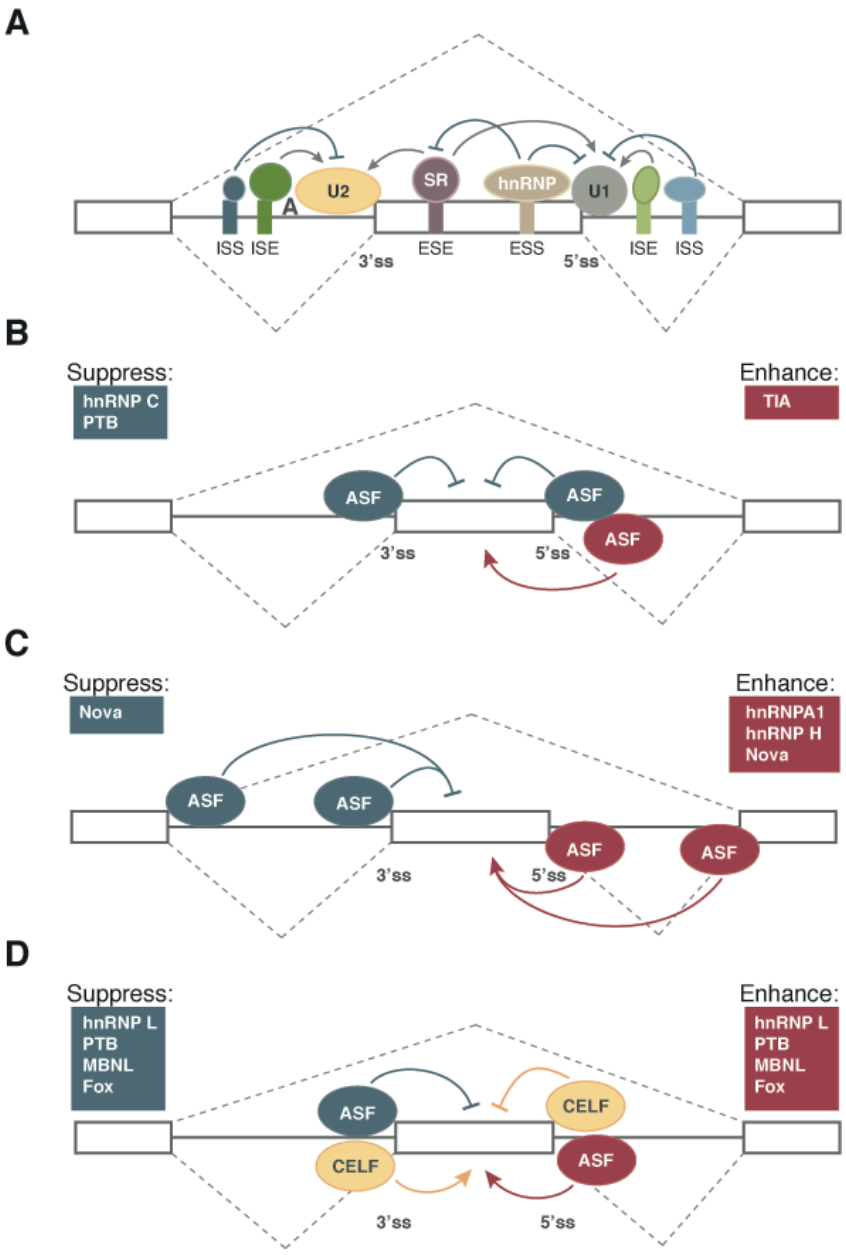
Alternative splicing requires a very complex, but beautifully orchestrated, series of protein-protein and protein-RNA interactions. However, very little is known about the exact cellular mechanisms that govern regulation of alternative splicing. Regulation of a single transcript draws on multiple themes of biological regulation at both the RNA and protein levels. Once E-complex is formed, all of the necessary RNA: protein interactions required to define the exon/intron boundary have been made and the intron is committed to being spliced. Therefore regulating E-complex formation is the most common mechanistic theme in splicing regulation. Splice site strength is defined by the sequence complementarily between the U1 snRNP and the 5'-splice site, and the extent of U2AF35/65 pairing to the py-tract and 3'-splice site. Extensive bioinformatic efforts have been made to score the strength of splice sites based on complementarity to the spliceosome (16)-(18). However the splice site consensus sequences in eukaryotes are somewhat degenerate and are generally not sufficient for spliceosome assembly. Consequently, auxiliary splicing sequences are found throughout introns and exons and have evolved to recruit auxiliary splicing factors (ASF) to aid in definition of the exon/intron boundary and to recruit spliceosomal components to the appropriate regions of the pre-mRNA.

AUXILIARY SPLICING FACTORS

Exonic Regulatory Elements. In higher order eukaryotes, splicing is not only complicated by discrepancies in splice site strength, but these organisms also have extremely small exons (typically 50-250 base pairs in length) in comparison to large intronic sequences (typically 100s to 1,000s of bps in length), and contain a number of "decoy" splice sites and pseudoexons (19), (20). Therefore auxiliary splicing sequences are vital for proper exon definition and for splicing to proceed. Exons often contain exonic splicing enhancer (ESE) or exonic splicing suppressor (ESS) signals that help the splicing machinery discern exonic sequence from the sea of surrounding intronic sequence (Figure 3A). Most ESEs are recognized by members of the serine/arginine (SR)-rich protein family and activate splicing at nearby splice sites (21). SR proteins contain a characteristic C'-terminal extended arginine-serine (RS) dipeptide repeat and at least one, N'-terminal RNA recognition motif (RRM) that binds RNA (22), (23). Once

bound to ESEs, the RS domains of the SR proteins interact with RS domains on other splicing factors, specifically U1 and U2AF35/65, and recruit them to splice sites (24). Proteins of the heterogenous nuclear RNP, or hnRNP, class most often recognize ESSs. Most hnRNPs contain one or more RRM and generally harbor glycine-rich, acidic, or proline-rich domains (23). Once bound, hnRNPs suppress splicing using a variety of mechanism involving protein: protein interactions, including: blocking recruitment or interaction of snRNPs, repressing spliceosome assembly by coating exons through multimerization, or looping out entire exons (25). SR and hnRNP proteins have been shown to influence both constitutively and alternatively spliced exons (26), (22) suggesting that these proteins may participate in most, if not all, splicing events (Figure 3A).

Figure 3 (next page)**. Regulation of splicing via auxiliary signals.** (A) The spliceosome recognizes the intron via interactions of U1 snRNP with the 5'-splice site (5'ss) and U2 snRNP with the 3'splice site (3'ss). Exonic ESEs bind SR proteins and function with ISEs to promote exon definition (arrowheads). ESSs generally bind hnRNP proteins and work with ISSs to block exon definition. (B) – (D) Splicing maps depicting how binding site position relative to a regulated exon relates to splicing activity for several alternative splicing factors (ASF). Binding events that activate inclusion are shown in red and those that suppress exon splicing are shown in blue. (B) ASFs with binding sites that occur only on one side of the regulated exon or flanking the exon. (C) ASFs that function to enhance or suppress splicing via interaction with both ends of the regulated intron. (D) ASFs that suppress splicing when bound upstream of a regulated exon and enhance inclusion when bound near 5'ss. CELF proteins show the opposite splicing trend. And are shown in yellow. In all panels exons are depicted with boxes and lines represent exons.



Intronic Regulatory Elements. Many of the sequences responsible for the regulation of alternative splicing are found within introns and bind alternative splicing factors. Intronic splicing enhancer (ISE) and intronic splicing suppressor (ISS) signals are found near splice junctions and facilitate exon identification. Many proteins are known to function through ISEs and ISSs, including some SR and hnRNP proteins, however intronic elements are less well characterized than their exonic counterparts (Figure 3A).

Defining ISE and ISSs is difficult because most alternative splicing factors act to either enhance or suppress splicing based on the location of their binding site relative to the alternatively the regulated exon. Therefore the same motif can act as an ISE or ESE depending upon where it occurs within the pre-mRNA. Additionally, expression of most alternative splicing factors is tissue-specific and developmentally controlled. This phenomenon has inspired the use of genome-wide approaches to create "splicing maps" for several well-studied alternative splicing factors.

UNDERSTANDING SPLICING REGULATION THROUGH SPLICING MAPS

Splicing maps correlate binding site locations for particular RNA binding proteins to regulated splicing effects on a genome-wide scale, allowing for a panoramic view of protein: RNA interactions within regulated transcripts. These maps reveal splicing regulation patterns unique to each splicing factor and have helped to predict if particular binding motifs acts as ISEs or ISSs based on their position relative to regulated exons (27), (28).

TIA-1, is a factor that only binds U-rich ISEs downstream of 5'-splice sites to stimulate U1 binding (29). hnRNP C binds runs of uridines (*i.e.* UUUUU), primarily found in py-tracts, and silences exon inclusion when bound near either splice site of an alternative exon (Figure 3B) (30). Poly-pyrimidine tract binding protein, or PTB, is a potent splicing suppressor that binds CU-rich ISSs. When PTB is simultaneously bound to position within py-tracts and near 5'-splice sites it also suppress exon inclusion (Figure 3B) (31).

hnRNP A1 and hnRNP H bind GA-rich and G-triple (GGG) containing sequences, respectively, which are found near the ends of most introns. Both factors can function as potent splicing suppressors when bound near 5'-splice sites by antagonizing SR proteins to block splice site usage by U1. However, more recent studies have shown that both hnRNP A1 and hnRNP H also promote efficient intron removal when they are bound at both the 5'- and 3'- ends of an intron simultaneously (Figure 3C), and that the two factors can collaborate on a single intron to promote splicing (32), (33). Nova is a neuronal-specific splicing factor that binds YCAY clusters (34), (35). The splicing map obtained for Nova indicates that Nova enhances exon inclusion by binding at both ends of

a downstream intron and suppresses exon inclusion by binding at both ends of and upstream intron (Figure 3C) (36), (37). The bimodal binding and well-defined homodimerization properties of hnRNPs have led to the proposal of an intron looping model to explain splicing regulation by factors containing this type of activity pattern. The intron looping model proposes that splicing factors with this binding pattern use protein: protein interactions to bring the two ends of the intron in close proximity to one another, looping out the intervening intron, to help define the nearby exons (30), (33), (27).

Ataxin 2-binding proteins (A2BP1/2; also known as Fox-1/2) bind the hexanucleotide sequence 5'-UGCAUG-3' and strongly enhances exon splicing when the motif is located downstream of an exon, near a 5' splice site. The same protein facilitates exon skipping when bound to the same motif when the motif is located upstream of the regulated exon, near a 3' splice site (38), (39). Muscleblind-like (MBNL) proteins, which bind YGCY clusters (40), (41), hnRNP L, and PTB have splicing patterns similar to Fox-1/2 in that they suppress exons when bounds near 3'-splice sites and activate splicing when bound near 5'-splice sites (Figure 3D). Members of the CUGBP and ETR-like factors (CELF) family, including the namesake CUGBP1, activate splicing with the opposite pattern of Fox-1/2 through binding to UG-rich sequences near 3'-splice sites and suppress splicing when binding motifs occur near 5'-splice sites (Figure 3D) (42). MBNL, Fox-1/2, and CELF binding sites occur near each other in a subset of regulated exon events, suggesting that MBNL and Fox proteins may cooperate in splicing activity while CELF proteins can antagonize function of either splicing factor.

Exonic and intronic regulatory elements are usually grouped together (Figure 3A) and alternative splicing factors can be expressed in tissue- or developmental stage-specific manner or ubiquitously, making the cellular concentration and ratio between alternative factors very important in the regulation of alternative splicing. Subtle differences in affinity or local concentration of multiple alternative splicing factors can dramatically affect a pre-mRNA's splicing fate, suggesting that unlocking subtle nuances in location of splicing factor binding sites or affinities for various targets may be key to understanding how large networks of splicing factors cooperate to regulate alternative splicing. Upsetting the cellular balance of alternative splicing factors can cause improper

splicing regulation that results in expression of aberrant transcripts and ultimately disease. The future of the alternative splicing field lies in combining the wealth of biochemical knowledge with recent global approaches to define splicing networks for each alternative splicing factor. Once combined these networks will help to create a "splicing code" that can be used as a predictive tool for understanding how alternative splicing patterns changes in response to disease and advance the discovery of treatments to correct improper splicing.

ALTERNATIVE SPLICING AND DISEASE

Myotonic Dystrophy (DM) is a degenerative muscle disease which effects 1 in 8,500 adults and is characterized by muscle hyperexcitability (myotonia), cataracts, cardiac defects, and insulin resistance (43). DM type 1 (DM1) is caused by a trinucleotide (CTG) repeat expansion in the 3'-untranslated region of the *DMPK* gene (44), (45) and DM type 2 (DM2) results from expansion of a CCTG repeat found within the first intron of the *ZNF9* gene (46). The repeats are not detrimental at the genomic level. However, once transcribed into RNA the CUG or CCUG repeats form ultra-stable hairpin structures sequester RNA binding proteins, such as Muscleblind-like 1(MBNL1), away from their endogenous targets, and into ribonuclear foci (47)-(49). Expression of CUGBP1 is also up-regulated in DM (50). Sequestration of MBNL1, and increased splicing antagonism by CUGBP1, results in mis-splicing of MBNL1 target transcripts and disease symptoms (51), (52), (43).

Due to its role in DM pathology, MBNL1 has been the focus of intense study for over a decade. Global approaches have been used to identify hundreds of DM-associated, MBNL1-mediated, mis-splicing events (53), (40). However, at the point that my research began only a handful of functional MBNL1 binding sites had been identified and there was no defined consensus binding motif for this factor. There is also very little mechanistic information currently available to explain how MBNL1 regulates splicing. Particularly, there has been no mechanistic exploration of how MBNL1 activates inclusion of upstream exons when bound near 5'-splice junctions. Therefore, the studies conducted herein were designed to further our mechanistic understanding of how this important splicing factor functions in alternative splicing.

In Chapter II we define the consensus binding motif of MBNL1 as 5'-YGCY-3' and were the first group to demonstrate the correlation between binding site position relative to a regulated exon and splicing activity described above (41). In Chapter II and Appendix A, I demonstrated that clusters of 5'-YGCY-3' motifs were sufficient to induce MBNL1-mediated regulation of a synthetic reporter and characterized a novel MBNL1-binding site in the ATP2A1 transcript. In Chapter III and Appendix B, I conducted an extensive combinatorial mutagenesis of MBNL1 to determine how the protein's multiple RNA binding domains contribute to RNA binding and splicing function. I was able to demonstrate that at least two, functionally distinct classes of MBNL1-regulated splicing events exist. Based on the experiment in Chapter III I proposed a model to discriminate between the two classes. The model suggests that one class of MBNL1-mediated splicing event may require protein: protein interactions. In Chapter IV, I used an artificial tether system to investigate potential protein: protein interaction domains within MBNL1. I also created chimeric reporters in Chapter IV to determine if sequences found near MBNL1 binding sites in endogenous transcripts contribute to splicing activity. Chapter II contains previously published, co-authored material. Chapter III contains unpublished, co-authored material.

CHAPTER II

MBNL1 RECOGNIZES YGCY MOTIFS EMBEDDED IN PYRIMIDINE RICH RNA SITES

This work was published in Volume 38 Issue 7 of the Nucleic Acids Research journal in January 2010. Emily S. Goers and I were co-first authors on this publication. Emily Goers performed the SELEX experiments, characterized SELEX RNAs and identified potential endogenous binding sites; I performed all the cloning and *in vivo* splicing assays; Rodger B. Voelker performed all the bioinformatics analysis; Devika P. Gates 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

Alternative splicing is essential for creating a diverse and functional proteome as well as for establishing tissue and developmentally specific repertoires of mRNAs. It has been shown that approximately 90% of human pre-mRNAs are alternatively spliced (22). Several proteins (*e.g.* NOVA1, CUGBP1, MBNL1, A2BP1 (also known as Fox-1) and their related paralogues) have been shown to play important roles in the regulation of alternative splicing (53), (35), (38). NOVA1 is a neuron-specific regulator of alternative splicing that binds YCAY clusters in or near alternatively spliced exons and promotes exon inclusion or exclusion, depending on the location of the binding site (35), (36). A2BP1 is expressed in brain, heart and skeletal muscle and binds the 5'-UGCAUG-3' RNA motif (54). Based on hundreds of predicted A2BP1 binding sites, A2BP1 binding upstream of a regulated exon promotes exclusion while binding downstream promotes inclusion (38). Although much less is known about MBNL1 binding sites, a similar model of alternative splicing regulation has been proposed for the MBNL1 proteins (53), (55).

The original member of the muscleblind family of proteins, muscleblind (Mbl) was identified in *Drosophila* and found to be important in photoreceptor and muscle differentiation (56), (57). The orthologous proteins, muscleblind-like 1-3 (MBNL1, MBNL2 and MBNL3) were discovered in humans as the proteins sequestered to the toxic CUG and CCUG repeats that cause Myotonic Dystrophies 1 and 2 (DM1 and DM2),

respectively (47)-(49). The muscleblind proteins are generally highly conserved, especially in the zinc finger domains, which bind RNA in a specific fashion (55), (58)-(61). The sequestration of MBNL1 results in its lack of binding to normal pre-mRNA targets. This lack of binding by MBNL1 causes important developmentally controlled transcripts to become mis-spliced and leads to symptoms of DM (for reviews see (51), (52), (62)). For example, insulin receptor (INSR) and chloride ion channel (CLCN1) pre-mRNAs are mis-spliced in DM1 patients leading to inappropriate expression of fetal isoforms and/or degradation of the transcript (63)-(65). The lack of appropriate INSR and CLCN1 splice isoforms in DM1 patients is thought to lead to the symptoms of insulin resistance and myotonia, respectively.

MBNL1 promotes the exclusion of exon 5 in the TNNT2 (also known as cTNT) pre-mRNA, which produces a splice product found in adult tissue. However, in DM1, exon 5 is included aberrantly, thus producing a splice product normally found in fetal tissue (50). It has now been shown that the sequestration of MBNL1 and MBNL2 is responsible for this mis-splicing (66). MBNL1 binds a 32-nucleotide region upstream of exon 5 and regulates splicing through this site (55), (66), (67). In addition to TNNT2, several other pre-mRNA transcripts are regulated by MBNL1, including ATP2A1 (also known as SERCA1), and auto-regulation of MBNL1 and MBNL2 pre-mRNAs (for reviews see (51), (62)). The only previously characterized MBNL1 binding site in a human pre-mRNA is the 32- nucleotide TNNT2 site. The identification of additional MBNL1 RNA binding sites would allow for a deeper understanding of MBNL1's RNA binding specificity. This will help determine pre-mRNA targets regulated by MBNL1 and to predict MBNL1 binding sites within these targets.

We performed a doped SELEX (Systematic Evolution of Ligands by Exponential Enrichment) experiment, using the TNNT2 binding site as a template, to identify sequences that MBNL1 binds with high affinity. The doping incorporated the endogenous binding site nucleotides at a rate of 51% for each of the 32 residue positions. The majority of sequences recovered after 5 rounds of selection bound MBNL1 with an affinity similar to MBNL1's affinity for the native TNNT2 site. We observed that there was a general selection for pyrimidines and a specific selection for the motif YGCY. We showed that placing multiple YGCY motifs in a splicing reporter minigene, which was

not previously regulated by MBNL1, was sufficient for MBNL1-dependent splicing regulation. We analyzed the sequences surrounding cassette exons in 24 pre-mRNAs that are mis-regulated in DM1 patients and identified over 100 YGCY potential MBNL1 binding sites. We then demonstrated that MBNL1 binds with high affinity to RNA oligomers derived from several of these endogenous YGCY motifs. We also showed that multiple YGCY instances significantly contribute to MBNL1 regulated splicing of ATP2A1, an exon known to be mis-spliced in DM1. And finally, we observed that YGCY motifs are enriched in the intronic regions upstream of cassette exons that are normally excluded (aberrantly included in DM1), and, conversely, are enriched downstream of exons that are normally included (aberrantly excluded in DM1). Together, these results aid in defining MBNL1's RNA binding specificity, which will ultimately help lead to the identification of more MBNL1 endogenous binding sites and provide a better understanding of the mechanisms by which MBNL1 regulates splicing.

RESULTS

Doped SELEX with an MBNL1 RNA Binding Site

MBNL1 binds just upstream of exon 5 in the TNNT2 pre-mRNA (55), (66). It has been previously shown that a 32-nucleotide RNA corresponding to this region binds MBNL1 with high affinity and that mutations that disrupt MBNL1 binding also disrupt MBNL1 regulated exclusion of the downstream exon (55). In order to identify the nucleotides that contribute most to recognition by MBNL1, we performed a doped SELEX experiment (68). Unlike traditional SELEX, which starts with a pool of uniformly random RNAs, doped SELEX begins with a population of RNAs synthesized such that they are biased towards a specific starting sequence, but each position is still allowed to vary. For this particular experiment we began with a population of 32-nucleotide RNAs synthesized such that each position was 51% likely to be the equivalent base found in the 32-nucleotide TNNT2 MBNL1 binding site. Each of the other nucleotides had a 16.3% probability of being used. For example, the first position in the 32-nucleotide region was cytidine, so in 51% of the RNAs it remained a cytidine, while the other 49% of the time the other three nucleotides were equally incorporated at that position (Figure 1A). This allowed MBNL1 to sample other possible residues at each

position with a bias toward the endogenous residue. The 32 endogenous nucleotides were flanked on each side by constant regions, 26 nucleotides on the 5' side and 23 nucleotides on the 3' side, that are necessary for PCR and transcription. The MBNL1 construct included residues 1-260 and an N-terminal GST tag. This protein, termed MBNL1 throughout the paper, bound RNA as tightly as the full-length version of MBNL1 (55).

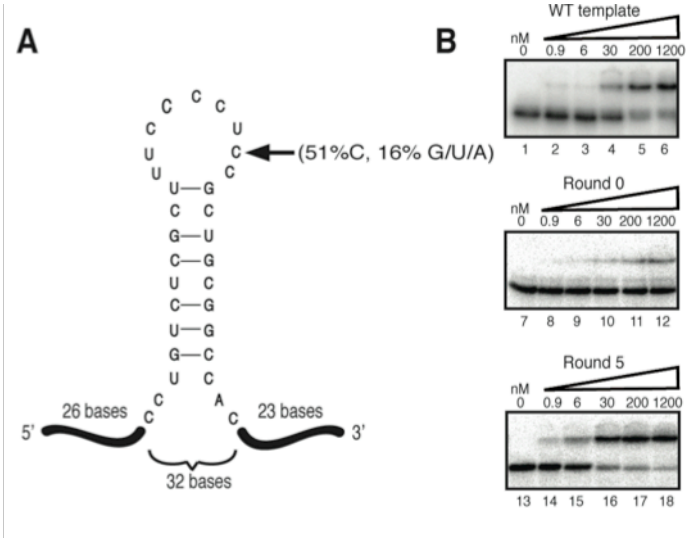


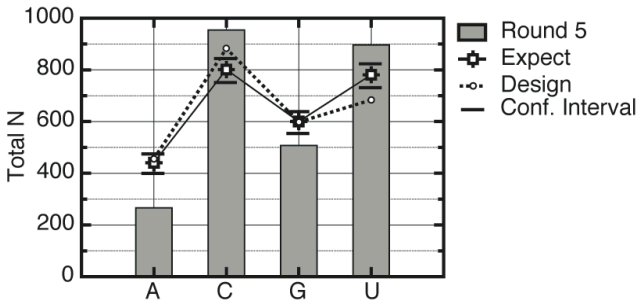
Figure 1. **Doped SELEX.** (A) The TNNT2 template used to create the pool 0 RNA for SELEX. The dark lines indicate the constant regions, while the central 32 bases denote the doped, variable region. The arrow indicates an example of the doping. For that particular position, the C was maintained in 51% of the sequences, while G, U and A were incorporated equally in 49% of the other sequences. (B) Gel mobility shift assays of MBNL1 with the 81-nucleotide wild type TNNT2 template (top), Round 0 RNAs (middle), and Round 5 RNAs (bottom).

After five rounds of selection, the binding affinities of the recovered pools of RNAs were tested using a gel mobility shift assay. The binding affinity of MBNL1 was low for pool zero (K_d of greater than 1.2 uM) and high for pool five (Figure 1B). Pool 5 RNA bound MBNL1 with slightly higher affinity compared to the endogenous TNNT2 binding site template (Figure 1B, compare lanes 1-6 and lanes 13-18), which demonstrated that we were successful in selecting RNAs with high affinity towards MBNL1. 82 sequences were obtained from Round 5, and there were no duplicates,

indicating that over-selection did not take place. The sequence most closely related to the endogenous TNNT2 template contained six bases that differed from the original TNNT2 site. This result was not surprising, as pool 0 started out with only a 1 in $2x10^9$ chance that a sequence would be identical to the endogenous template. Pool zero was sequenced (96 individual sequences) to insure that there was no bias in the initial pool, other than the designed doping. Analysis of these sequences demonstrated that pool 0 sequences had compositions similar to that expected according to the doping design. However, the pool 0 RNAs did show a slight enrichment in the number of uridines and a corresponding slight decrease in the number of cytidines expected according to the doping design (Figure 2A).

Figure 2 (next page)**. Nucleotide composition, K-mer analysis and the frequency of nucleotides flanking GC dimers confirm a high occurrence of pyrimidines and YGCY motifs within SELEX sequences.** (A) Single nucleotide counts for the randomized regions derived from Round 5 RNAs are compared to those expected (labeled 'Expect') given the frequencies observed in the starting pool (derived from the counts observed in Round 0). Upper and lower confidence intervals (α=0.05) are indicated above and below each of the expected values (see Materials and Methods for the calculation). Also shown are the values expected according to the original doping design (labeled 'Design'). (B) The K-mer analysis reveals highly enriched 4-, 5- and 6-mers ($p < 0.01$) found within the SELEX sequences from Round 5. The fifteen K-mers listed have the highest Z-scores out of each K-mer size category (4-6). Observed is the total number of occurrences of the K-mer within the sample of 82 SELEX sequences. The expected is the mean number of the K-mer within 1000 sets of 82 random sequences, taking the doping into account. (C) The nucleotide frequencies flanking all GC dimers found within the SELEX Round 5 randomized regions were determined. The frequencies are represented as sequence-logos with the size of the nucleotide corresponding to frequency (numbers to the right of each nucleotide in percentages). The nucleotides are ordered top to bottom, highest frequency to lowest frequency.

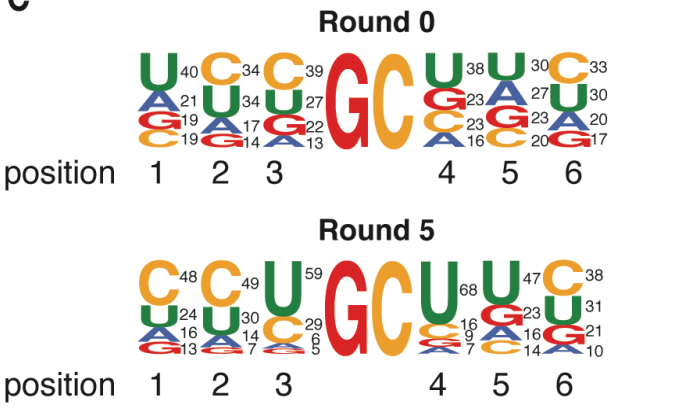
A



B

K-mer	Observed	Expected	Z-score
UGCU	107	12	27
GCUU	116	15	27
UUGC	69	14	15
CUGC	75	19	13
CGCU	78	23	12
UGCUU	66	3	36
CUGCU	54	4	24
GCUUU	51	5	21
UUGCU	39	3	19
CGCUU	44	6	15
CUGCUU	37	1	35
UGCUUU	27	1	27
CCUGCU	28	1	24
UGCUUG	18	1	23
UUGCUU	22	1	22

C



The Majority of SELEX-Derived RNAs Bind MBNL1 with High Affinity and Are Highly Enriched in YGCY Motifs

MBNL1's endogenous TNNT2 binding site is pyrimidine rich (28% U, 50% C) and purine poor (19% G, 3% A). In order to determine if there was an overall selection for pyrimidines over purines during the SELEX process, we analyzed the single

nucleotide frequencies for the randomized region from the population of sequences recovered during Round 5 and compared these to the sequences from Round 0 (Figure 2A). This analysis demonstrated that there was a statistically significant enrichment in the overall uridine and cytidine content in the Round 5 RNAs when compared to Round 0 RNAs, suggesting that MBNL1 has a preference for pyrimidines over purines.

An analysis of the K-mer (4-6 nucleotide) composition of the Round 5 RNAs relative to the composition expected by chance (based upon the starting template and doping regime, see Materials and Methods) revealed that the Round 5 RNAs are enriched in specific K-mers (Figure 2B). Meanwhile, the same analysis identified no significantly enriched K-mers within the Round 0 RNA sequences (data not shown). All of the most highly enriched K-mers contain a GC-dimer. Of these, the most significantly enriched contain UGCU (107 occurrences compared to 12 expected by chance, $Z= 27$). Closer inspection revealed that CGCU ($Z= 12$) and UGCC ($Z= 3$) are also enriched, but at lower levels (Figure 2B).

We aligned all GC-dimer instances in the Round 5 sequences, plus three nucleotides on either flank, and compared these to GC-dimers found in Round 0 (Figure 2C). This analysis revealed a selection toward pyrimidines in the three positions flanking either side of the central GC-dimer. This, and the previous analysis, suggests that MBNL1 has a higher affinity towards UGCU; however, since cytidines are also tolerated we chose to represent the general MBNL1 binding site as YGCY.

The majority of SELEX sequences contain two to six YGCY motifs (Class I, Figure 3). Over half of the sequences contain YGCY in the same locations as the wild type TNNT2 CGCU motifs. However, many of the sequences show that MBNL1 selected YGCY motifs that are shifted with respect to the positions of the YGCY motifs in the wild type TNNT2. In some cases, this change is a slight shift of the YGCY position in comparison to the location of the wild type motifs, whereas in other sequences a whole new second, third or fourth additional motif appeared nearby or in a different region.

Binding studies were performed on 12 of the SELEX sequences to determine the affinities of these sequences to MBNL1. Almost all of the sequences tested have binding affinities roughly equal to or higher than the endogenous TNNT2 binding-site (Table 1).

Table 1. Binding affinities of MBNL1 for selected SELEX sequences

Sequence	K_d (nM) 81mer	K_d (nM) 32mer
WT cTNT	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	
H01	10 ± 5	
H01 CA1[b]	10 ± 3	
H01 CA2[b]	2.3 ± 0.6	
H01 CA3[b]	3.9 ± 1.2	
H01 CA1-2[b]	120 ± 35	
H01 CA1-3[b]	21 ± 9	
H01 CA2-3[b]	8.3 ± 2.4	
H01 CA1-2-3[b]	>1.2 μM	
D12	25 ± 7	
D12 CA1[b]	11 ± 6	
D12 CA2[b]	24 ± 12	
D12 CA3[b]	4.9 ± 1.2	
D12 CA4[b]	8.4 ± 2.8	
D12 CA1-2[b]	20 ± 11	
D12 CA1-4[b]	3.9 ± 1.8	
D12 CA 2-4[b]	3.8 ± 0.7	
D12 CA1-2-4[b]	20 ± 3	
D12 CA1-2-3-4[b]	>1.2 μM	

[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.

The SELEX sequences were grouped into two classes. Class I contains sequences with two or more YGCY motifs and K_ds between 1 nM and 25 nM (of the RNAs tested). Class II sequences contain zero or one YGCY motif and have K_ds greater than 1.2 uM (of the RNAs tested). Interestingly, two sequences in Class II that contain only one motif (C12 and B08), don't bind MBNL1, suggesting an important role for multiple YGCY motifs for MBNL1 binding (Figure 3, Table 1).

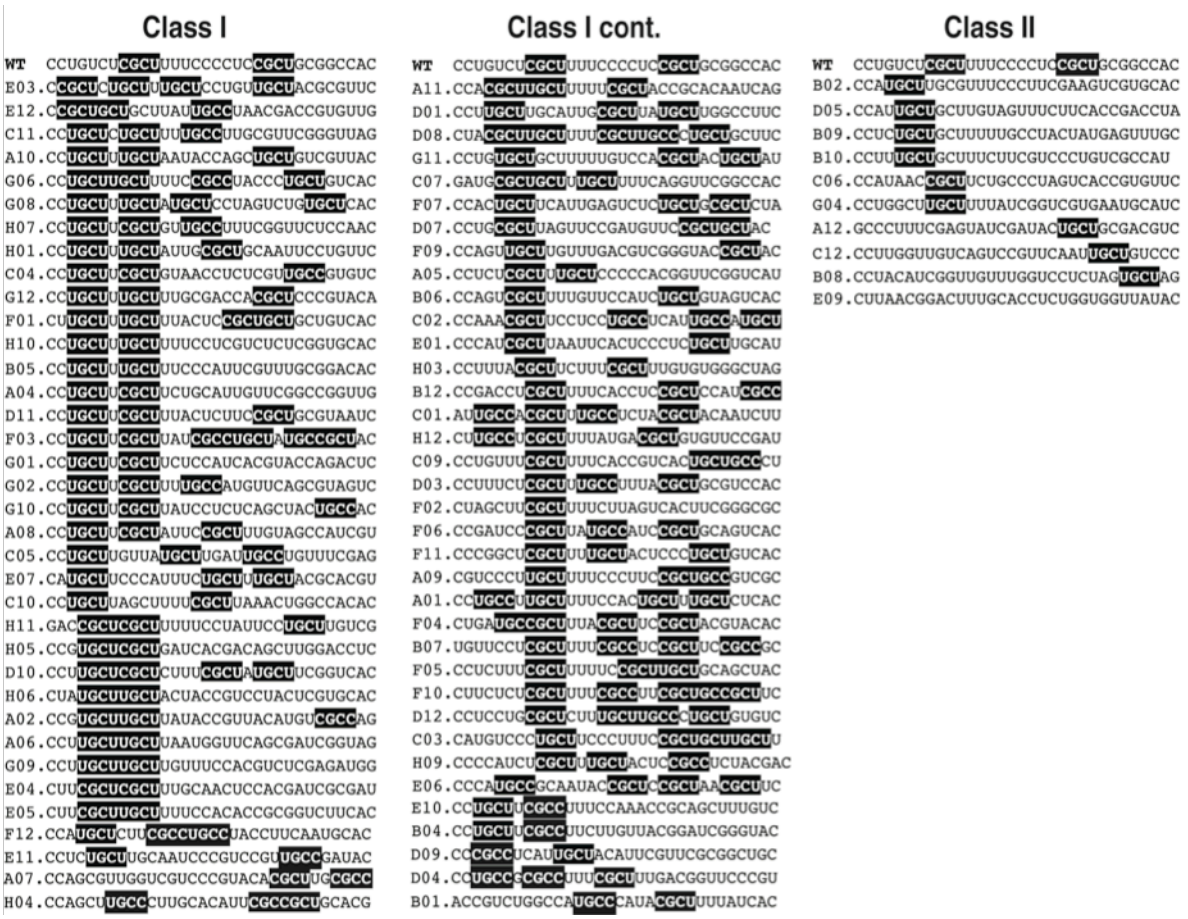


Figure 3. SELEX sequences are highly enriched in YGCY motifs. The YGCY motifs highlighted in all 82 Round 5 SELEX sequences. Class I RNAs contain two or more YGCY motifs and bind MBNL1 with high affinity while Class II contain zero or one motif and do not bind MBNL1 with high affinity. The wild type TNNT2 sequence contains two YGCY motifs and is shown at the top of each column for reference.

To determine if the constant (non-randomized) regions were playing a role in the binding of MBNL1 to the SELEX sequences, the constant regions were removed from several SELEX sequences and binding studies were performed. The addition of the constant regions to the TNNT2 sequence only slightly weakened binding (K_d of 30 nM), compared to a K_d of 20 nM for the 32-mer endogenous TNNT2 site. Similar studies were performed on three of the SELEX sequences (F06, B07, A04) to determine the effects of removing the constant regions for these RNAs. The F06 32-mer showed similar binding to the equivalent 81-mer (Table 1). The B07 32-mer bound MBNL1 with high affinity (K_d of 0.66 nM), which was nearly 10-fold tighter than the B07 81-mer The A04 32-mer

bound much weaker to MBNL1 with a K_d of 180 nM compared to 6.0 nM for the 81-mer RNA. Although these studies show that the constant regions can influence MBNL1 binding, in general the randomized regions appear to contain the necessary elements for MBNL1 high affinity binding.

Mutations in YGCY Motifs in SELEX Sequences Decrease MBNL1's Binding Affinity

To determine the importance of the YGCY motif for MBNL1 recognition, this motif was mutated in three different 81-mer SELEX sequences (F06, H01 and D12). The F06 sequence contains three YGCY motifs (Figure 4A). Motif #1 in F06 was mutated from CGCU to UCCA, which caused a 140-fold decrease in binding affinity to MBNL1. The H01 sequence contains three YGCY motifs, which were mutated from GC to CA individually (H01 CA1, HO1 CA2, and HO1 CA3), in pairs (H01 CA1-2, HO1 CA1-3, HO1 CA2-3), or all three at once (H01 CA1-2-3) (Figure 4B, Table 1). The H01 mutants with one or two motifs mutated bound with similar affinity to H01 (K_d's of 2-10 nM), with the exception of H01 CA1-2, which bound with 12-fold lower affinity to MBNL1 compared to H01. When all three motifs were mutated (H01 CA1-2-3), binding was greatly weakened, with a K_d greater than 1.2 uM (Figure 4B, lanes 19-24). Together, these results demonstrate that, for this particular RNA, only one motif is required for MBNL1 binding. However, in the case of two SELEX sequences from Class II, C12 and B08, MBNL1 was unable to bind them although each contains one YGCY motif (Figure 3 and Table 1). This indicates that MBNL1 binding is context dependent and in some cases, requires more than one YGCY motif.

The D12 RNA contains four YGCY motifs (Figure 4C). These were mutated from GC dimers to CA dimers one at a time, two at a time, three at a time or all four at once (Figure 4C, Table 1). All single, double, or triple mutant combinations of D12 RNAs have an affinity within a 10-fold range of the native sequence (K_d of 25 nM). In contrast, when all four motifs were mutated (D12 CA1-2-3-4, Table 1) binding was drastically reduced (K_d greater than 1.2 uM). Certain motifs within these RNAs may play a more significant role in MBNL1 binding, such as the H01 motifs #1 and #2 compared to HO1 motif #3, but in general, MBNL1 does not appear to have a preference for specific

YGCY motifs in these RNAs (Table 1). In conclusion, mutations in these SELEX sequences indicate that MBNL1 requires at least one YGCY for high affinity binding.

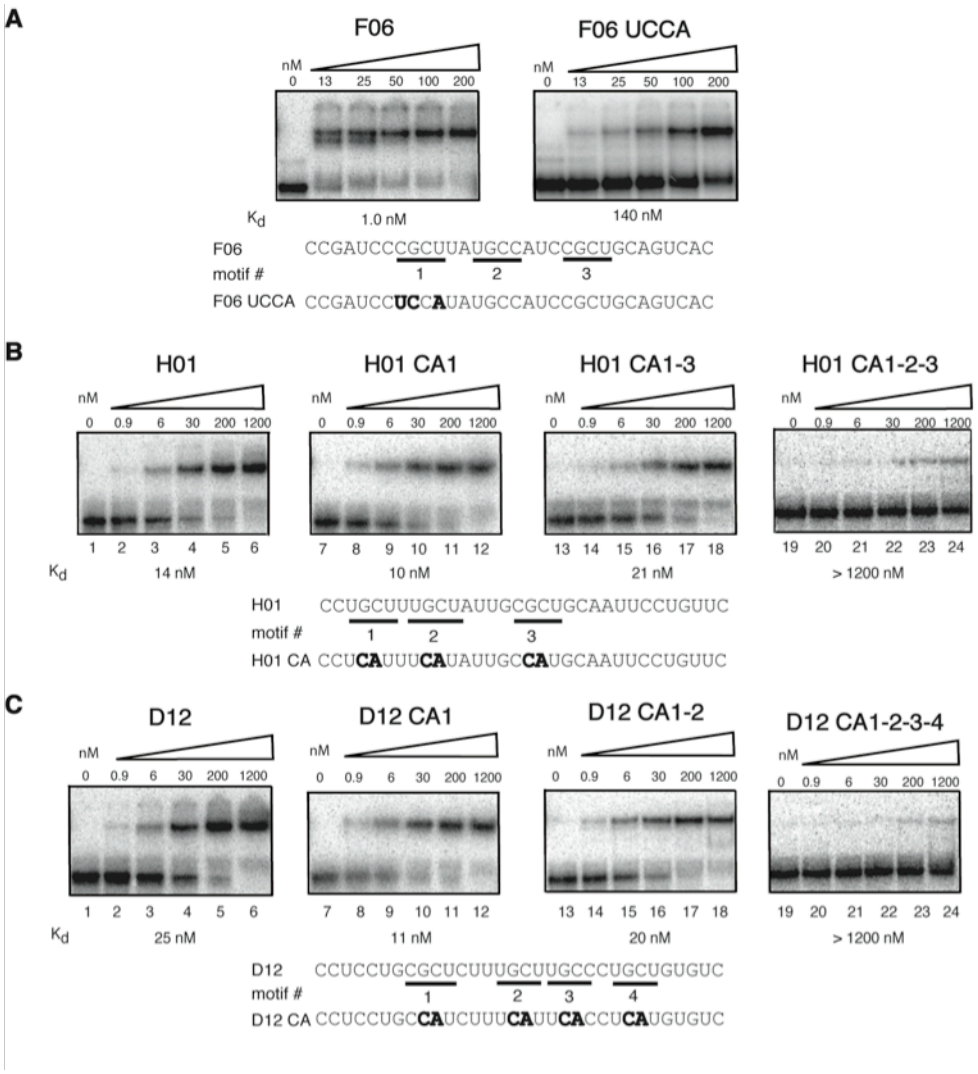


Figure 4. Mutations in YGCY motifs decrease binding to SELEX RNAs. (A) SELEX RNA F06 has three YGCY motifs, indicated by motif #1, #2 and #3. Pictured are gel mobility shift assays to F06 and mutated F06 (F06 UCCA), with the K_ds listed below. The residues in bold indicate mutated residues. (B) RNA H01 contains three YGCY motifs and binds MBNL1 with a K_d of 14 nM. Representative gel mobility shift assays and K_ds of MBNL1 binding to one of each type of mutation are shown. Mutations relative to H01 are shown in bold. H01 CA1 indicates the GC in motif #1 was mutated to CA, H01 CA1-3 indicates motifs #1 and #3 were mutated (but not motif #2). H01 CA1-2-3 has all three motifs mutated. All RNAs are 81 nucleotides long. (C) D12 contains four YGCY motifs and binds MBNL1 with a K_d of 25 nM, indicated by the gel mobility shift assay labeled D12. D12 CA1, D12 CA1-2 and D12 CA1-2-3-4 refer to mutations in the corresponding motifs shown in the lower portion of the figure. Representative gel mobility shift assays and K_ds are shown for D12 mutant RNAs bound to MBNL1.

YGCY Motifs Are Sufficient for MBNL1 to Regulate Splicing in a PLEKHH2 Minigene Reporter

In order to determine if YGCY is sufficient for MBNL1 to regulate splicing, we created an artificial binding site using the SELEX CUGCUU K-mer (SELEX motif 1, SM1) because it had the highest Z-score in the K-mer analysis of 6-mers enriched in the SELEX sequences (Figure 2B). In addition, CUGCUU is found in several SELEX sequences that MBNL1 binds with high affinity (E01, A04, H01). MBNL1 binds $(CUGCUU)_6$ RNA with high affinity in a gel mobility shift assay (Figure 5A).

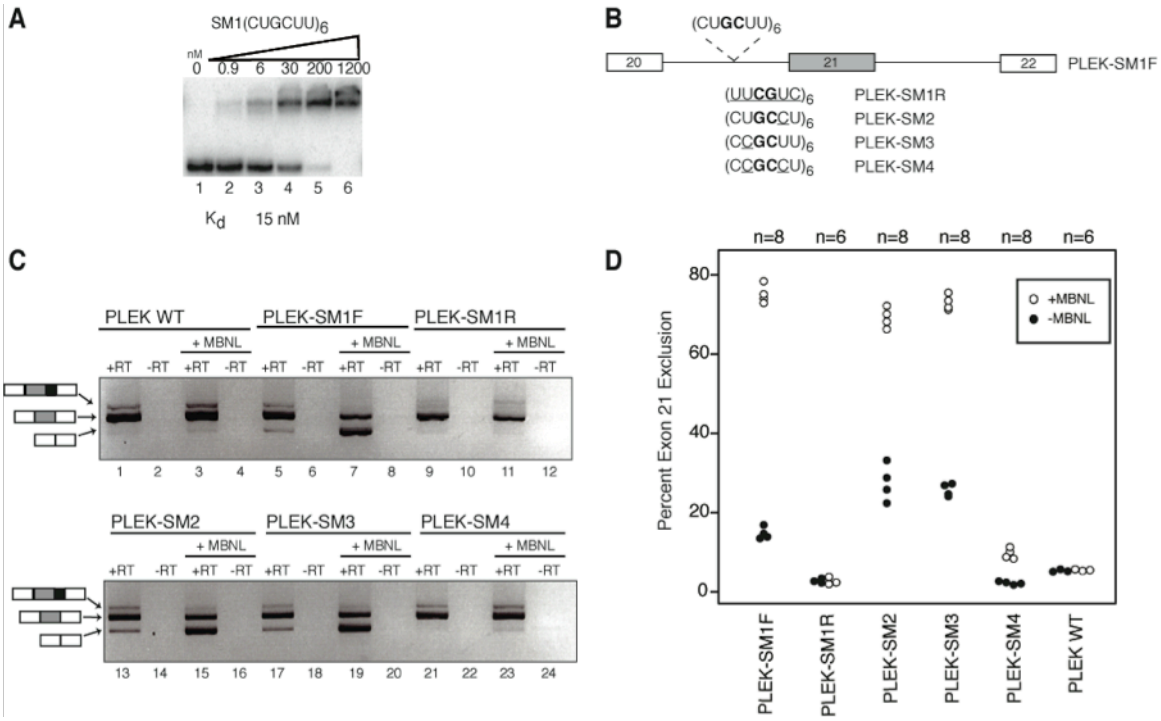


Figure 5. Multiple YGCY motifs embedded in pyrimidines are sufficient for MBNL1 regulated splicing of a PLEKHH2 reporter minigene. (A) Gel mobility shift assay of MBNL1 to SELEX motif 1$(CUGCUU)_6$ RNA. (B) A diagram of the PLEKHH2 minigene reporter, PLEKHH2-SM1F, containing 6 copies of the SELEX motif 1 (CUGCUU). Additional reporters were constructed by varying the sequence of the inserted motif. Nucleotide variations in the motif around the GC core (bold) are underlined. (C) HeLa cells were transfected with either wild type or SELEX-motif containing splicing reporters with or without MBNL1 over-expression (+MBNL). Representative agarose gels of RT-PCR products are shown. A diagram of splice products including exon 20 or 22 (white boxes), exon 21 (gray boxes), the alternative 5' splice site (downstream of exon 21) (black boxes) are shown to the left of their corresponding bands in the gel. (D) Dot plot depicting splicing of the PLEK-SM reporters with (open circles) and without (solid circles) MBNL co-expression.

We designed several splicing reporter minigenes based on the PLEKHH2 gene transcript. This transcript was chosen because MBNL1 does not normally regulate the splicing of this minigene. We designed several minigenes that include exons 20-22 and the intervening introns (Figure 5B). We replaced 36 nucleotides (positions -76 to -40) with a 50-nucleotide sequence containing six repeats of either CUGCUU (PLEKHH2-SM1F) or, as a negative control, UUCGUC (PLEKHH2-SM1R), the reverse of SM1. We also wanted to explore the effect of C versus U in the positions immediately flanking the GC core of the motif. To this end, we created three additional splicing reporter constructs in the same manner as PLEKHH2-SM1F using variations of SELEX motif 1; PLEKHH2-SM2, PLEKHH2-SM3, and PLEKHH2-SM4. PLEKHH2-SM2 contains six repeats of CUGC<u>C</u>U (the changed nucleotide is underlined). This reporter was designed to assess the effects of a C, instead of a U, immediately 3' of the GC. Similarly, PLEKHH2-SM3 contains $(C\underline{C}GCUU)_6$, to evaluate the effects of a C at the position immediately 5' of the GC. PLEKHH2-SM4 contains a C at both positions flanking the GC core to give $(C\underline{C}GC\underline{C}U)_6$ (Figure 5B). In all cases the motifs are positioned approximately 10 nucleotides upstream of the putative branch site and of the poly-pyrimidine tract. It is important to note that the reversed sequence (PLEKHH2-SM1R) contains the same overall pyrimidine content and the same number of guanosines as in SELEX motif 1; however, in the reversed sequence the UGCU motif is converted to UCGU. All three constructs were transfected into HeLa cells with or without an MBNL1 protein expression vector (Figure 5C). Splice products were observed via harvesting of HeLa RNA and subsequent RT-PCR using primers to PLEKHH2.

Several splicing products were observed for the PLEKHH2 WT construct (Figure 5C, lane 1). The major splice product (Figure 5C, lane 1, middle band) corresponds to inclusion of exon 21. We also observed a smaller amount of product that arose from usage of an alternative 5' splice site lying downstream of the normal 5' splice site (Figure 5C, lane 1, top band). Very little product (5%) corresponding to the skipping of exon 21 (Figure 5C, lane 1, bottom band) was observed in the WT construct. The splicing pattern for PLEKHH2 WT was completely unaffected by expression of MBNL1 (Figure 5C, lane 3 versus lane 1 and 5D), which demonstrates that the WT splicing pattern is MBNL1 independent.

The overall splicing pattern observed for PLEKHH2-SM1F, without expression of MBNL1, was very similar to that observed for the PLEKHH2 WT construct (Figure 5C, lane 5), however exon 21 exclusion is increased to 15% for this reporter (Figure 5D). In sharp contrast to the effect seen for PLEKHH2 WT, when MBNL1 was co-expressed with the PLEKHH2-SM1 reporter, 75% of the splicing product observed corresponded to the skipped isoform (Figure 5C, lane 7 and 5D) demonstrating that MBNL1 robustly suppresses inclusion of exon 21 in this synthetic construct. HeLa cells contain a low level of endogenous MBNL1; this could explain the low level of the exon 21 skipped isoform seen when expressing PLEKHH2-SM1without co-expression of MBNL1 (Figure 5C, lane 5 versus lane 1).

The splicing pattern for PLEKHH2-SM1R (containing SELEX motif 1 reversed), like the other constructs, consists primarily of the exon 21 inclusion isoform (Figure 5C, lane 9), with only 2% exon 21 exclusion. However, in addition to the small amount of transcript derived from usage of the downstream alternative 5' splice site, there is a minimal amount of higher molecular weight product corresponding to a cryptic 3' splice site. Importantly, and in sharp contrast to PLEKHH2-SM1F, when MBNL1 is co-expressed with PLEKHH2-SM1R there is no corresponding increase in skipping of exon 21 observed and the overall percentage of exon 21 exclusion remains unchanged at 2% (Figure 5C, lane 11 versus lane 9 and 5D).

The splicing patterns for PLEKHH2-SM2 and PLEKHH2-SM3 were very similar to PLEKHH2-SM1F, with exon 21 exclusion percentages of 28% and 26%, respectively (Figure 5C lanes 13 and 17 and 5D). Also, similar splicing changes were observed in both reporters with MBNL1 co-expression leading to strong MBNL1-induced increases in exon 21 exclusion to 70% and 73%, respectively (Figure 5C, lanes 15 and 19 and 5D). These findings suggest that MBNL1 recognizes a C upstream or downstream of the GC core equally efficiently when the other Y position in the YGCY motif is a U.

Splicing of the PLEKHH2-SM4 construct gave 2% exon 21 exclusion (Figure 5C lane 21). Surprisingly, however, only a slight increase in exon 21 exclusion to 10% was observed upon co-expression of MBNL1 (Figure 5C, lane 23 and 5D). This observation suggests that MBNL1 recognizes CGCC less efficiently than UGCU, UGCC or CGCU,

which suggests that MBNL1 requires at least one U flanking the GC core for high affinity recognition.

Together, these results demonstrate that, at least in the case of PLEKHH2, insertion of multiple YGCY motifs is sufficient for conferring MBNL1 dependent alternative splicing upon an exon that is normally constitutively included and does not normally require MBNL1 for splicing regulation. The observation that the reversed sequence is incapable of conferring MBNL1 dependent splicing demonstrates that a GC step (rather than a CG or GU step) is necessary for MBNL1 specific regulation. This distinction differentiates the MBNL1 binding site from the compositionally similar ETR-3 binding site: UGUU (69). Additionally, the observation that the exon 21 exclusion ratio in PLEKHH2-SM1F, PLEKHH2-SM2, and PLEKHH2-SM3 reporters are all robustly increased with MBNL1 co-expression suggests that MBNL1 recognizes a C or U flanking the GC core of the motif (i.e. CGCU or UGCC) with similar efficiency. Importantly however, MBNL1 co-expression with the PLEKHH2-SM4 reporter was much less efficient at inducing exon 21 exclusion in that reporter. This observation suggests that MBNL1 tolerates a C at either Y position only as long as the other position remains a U, therefore MBNL1 does not appear to recognize CGCC as well as it recognizes UGCU, UGCC, or CGCU.

Novel Putative MBNL1 Binding Sites Identified in Pre-mRNAs Mis-spliced in DM1

We predicted potential endogenous MBNL1 binding sites by identifying YGCY motifs clustered within introns flanking five exons mis-spliced in DM1. We evaluated these sites *in vitro* to determine if MBNL1 can bind these potential sites with similar affinity to that of the known TNNT2 site. Two sites in the MBNL1 pre-mRNA upstream of exon 7, one site downstream of ATP2A1 exon 22, one site in the MBNL2 pre-mRNA upstream of exon 7, one site downstream of exon 11 in INSR, and one site upstream in GRIN1 were tested for MBNL1 binding (Figure 6).

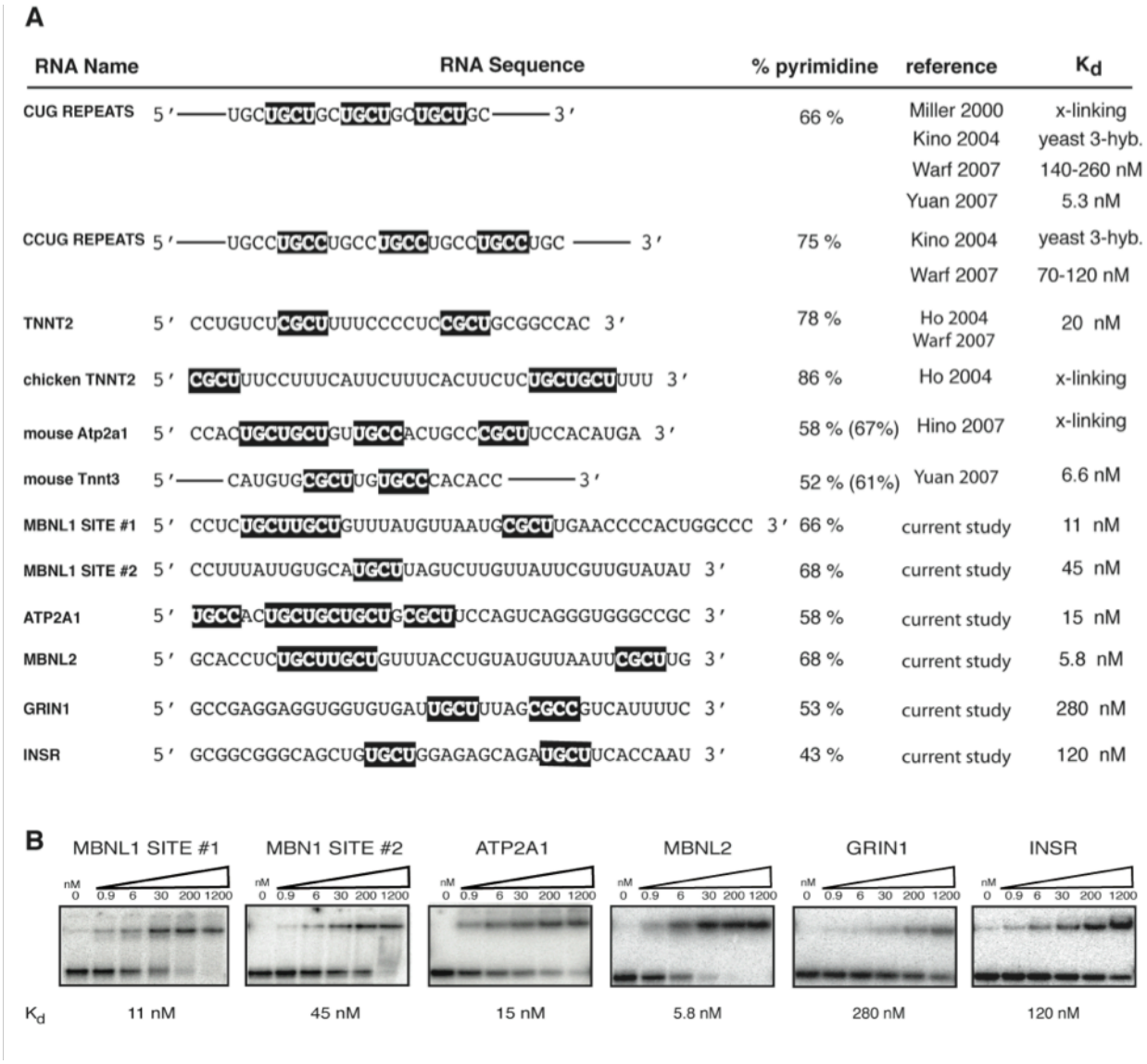


Figure 6. YGCY motifs are found in multiple MBNL1 binding sites. (A) Shown are twelve RNAs that MBNL1 binds. The K_ds on the right are reported in the literature (reference) or in the current study. YGCY motifs are highlighted in black. The percent pyrimidine was calculated based on the sequences that the various binding studies were performed with. The percent pyrimidine calculations in parenthesis were calculated based on the sequences shown in the figure (binding studies were performed on larger regions of RNA in some cases). (B) Representative gel mobility shift assays and K_ds to MBNL1 are shown for all binding sites identified in this study. The concentration (nM) of MBNL1 is indicated above each gel.

MBNL1 site #1 contains upstream residues 35-79 (45-mer) and has three YGCY motifs. MBNL1 site #2 contains upstream residues 154-193 (40-mer) and has one YGCY motif. The potential MBNL1 binding site found in ATP2A1 pre-mRNA has five motifs

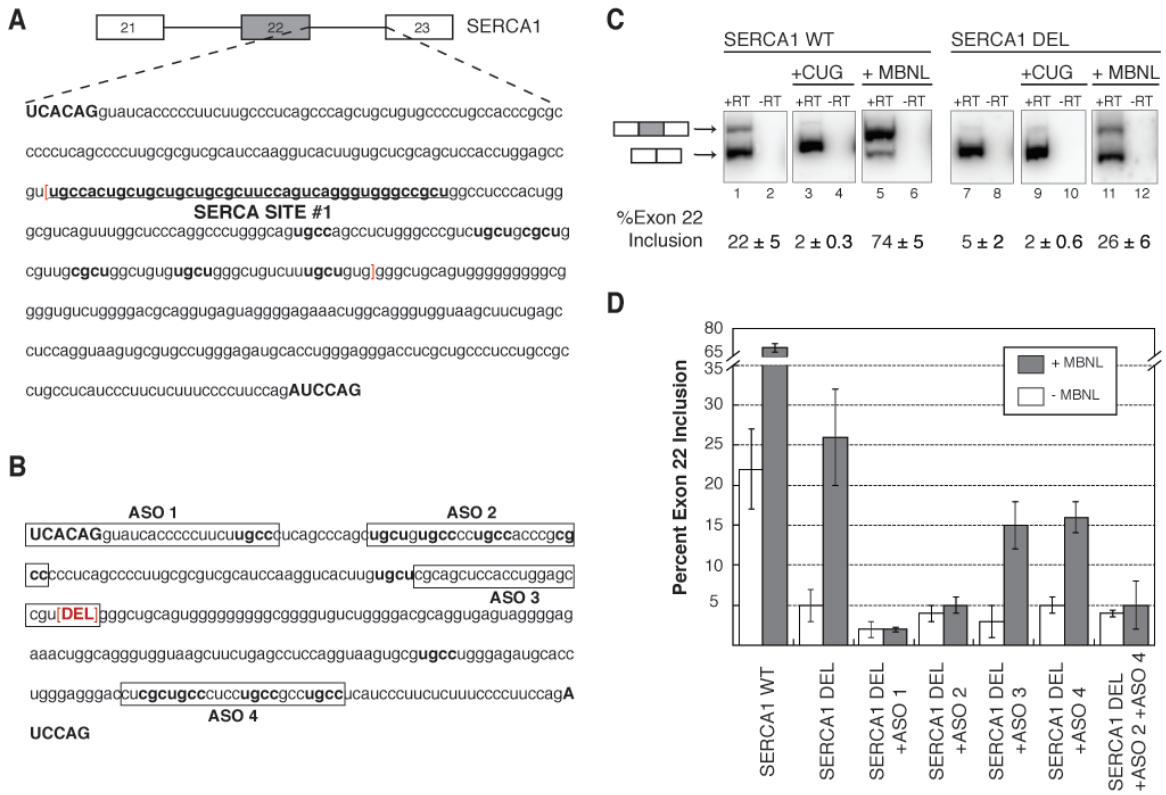
and is located downstream and contains residues 117-156 (40-mer). The site in the MBNL2 pre-mRNA is between upstream residues 50-89 (40-mer) and contains three motifs. The K_ds of MBNL1 to the potential RNA binding sites; MBNL1 site #1, MBNL1 site #2, ATP2A1 and MBNL2 sites were measured as 11 nM, 45 nM, 15nM and 5.8 nM, respectively (Figure 6B), well within the range of MBNL1's RNA binding affinity to TNNT2 (K_d of 20 nM). MBNL1 bound the 40 nucleotide GRIN1 and INSR RNAs, each containing two YGCY motifs, but did so with lower affinity (K_ds of 280 nM and120 nM, respectively). This observation suggests that aspects other than the YGCY motifs, such as pyrimidine content of adjacent regions of RNA, number of motifs, and RNA structure may also play a role in MBNL1's RNA binding affinity. Although these other aspects almost certainly play a role in MBNL1 binding, simply searching for YGCY occurrences has proven a fruitful method for predicting potential MBNL1 binding sites.

Multiple YGCY Motifs Contribute to MBNL1 Regulated Splicing of the ATP2A1 Exon

In order to validate the functionality of YGCY as the MBNL1 recognition motif in an endogenous target we created a splicing reporter minigene of the ATP2A1 transcript. Our minigene contains exons 21, 22, and 23 of the ATP2A1 transcript and the intervening introns (Figure 7A).

Figure 7 (next page)**. MBNL1 recognizes multiple YGCY motifs in an endogenous target, ATP2A1.** (A) Diagram of the ATP2A1 minigene, with white boxes indicating exons 21 and 23, lines to represent intervening introns, and a grey box representing the alternatively spliced exon 22, the inclusion of which is regulated by MBNL1. Sequence of intron 22 is shown with high affinity MBNL1 binding site (ATP2A1 binding site) underlined. The intronic sequence contained within the red brackets indicates the portion of the intron (151 nucleotides) that was deleted to create the ATP2A1 deletion (ATP2A1 DEL) minigene in panel B. Capitalized nucleotides represent exonic sequence; lower case nucleotides represent intronic sequence. YGCY motifs contained within the deletion region are in bold. (B) HeLa cells were transfected with either ATP2A1 WT or ATP2A1 DEL splicing reporter minigenes alone, or with either MBNL (+MBNL) or DMPK-CUG_{960} (+CUG) co-expression. Splice products are diagramed. The top band corresponds to the splice product containing exon 22 (grey box), whereas the bottom band corresponds to the splice product wherein exon 22 is excluded. The resulting splicing products were quantified using radiolabeled RT-PCR and percent exon 22

inclusion is reported below representative acrylamide gels. The values reported are an average of at least 4 independent experiments. (C) Sequence of intron 22 after the 151 nucleotide deletion (red bracketed DEL). Target sequences of antisense oligonucleotides (ASO) are boxed and YGCY motifs contained within the ATP2A1 DEL minigene are indicated in bold. (D) HeLa cells were transfected with the ATP2A1 DEL minigene in combination with the ASOs diagrammed in panel B, both with (grey columns) or without (white columns) MBNL co-expression, and the percentage of exon 22 was quantified as in panel C. Inclusion values reported are the result of at least 3 independent experiments.



In DM1 patients, this exon is aberrantly excluded and thus is potentially regulated by MBNL1, as shown in mouse Atp2a1 (70), (71). As expected, our ATP2A1 reporter gave two major splicing product bands, corresponding to exon 22 inclusion (Figure 7B, lane 1, top band) and exon 22 exclusion (Figure 7B, lane 1, bottom band). For the ATP2A1 minigene we observed 22% exon 22 inclusion when spliced in HeLa cells (Figure 7B, lane 1 and 7D). As previously mentioned HeLa cells express some endogenous MBNL1, so the exon inclusion observed for the reporter is most likely due to exon inclusion induced by the endogenous MBNL1. In DM1, it is thought that MBNL1

is sequestered away from the ATP2A1 transcript and unable to regulate exon 22 inclusion, resulting in very low inclusion and an aberrant transcript (70). As expected, when our minigene is spliced in the presence of 960 CUG repeats, the inclusion of exon 22 drops to 2% (Figure 7B, lane 3). Importantly, splicing of the ATP2A1 reporter in combination with over-expression of MBNL1 lead to an increase in exon 22 inclusion to 74% (Figure 7B, lane 5 and 7D). In combination, these observations suggest that our ATP2A1 minigene is a good reporter and is amenable to studying MBNL1-mediated splicing effects in a DM1-related transcript.

Multiple YGCY motifs are found in the intron downstream of ATP2A1 exon 22. We demonstrated that MBNL1 binds to an RNA oligonucleotide derived from this region (Figure 6B), making this site a candidate for an endogenous MBNL1 binding site responsible for regulation of exon 22. To test the importance of this site and the surrounding YGCY motifs on splicing regulation we deleted 151 nucleotides of intron 22 in the splicing minigene to create an ATP2A1 deletion minigene (ATP2A1 DEL) that lacks both the high affinity ATP2A1 binding site and six YGCY motifs located downstream of the ATP2A1 site (Figure 7C). When the ATP2A1 DEL minigene was tested under the same splicing conditions as the WT ATP2A1 minigene we observed that deletion of the high affinity site lead to a decrease in exon 22 inclusion from 22% to 5% (Figure 7C, compare lane 7 to lane 1). This observation could imply that removal of the high affinity site created a minigene that endogenous MBNL1 recognizes much less efficiently. Inclusion of exon 22 of the ATP2A1 DEL minigene was also reduced to 2% upon co-expression of CUG repeats (Figure 7B, lane 9). This is an important observation because it implies that although MBNL1-mediated regulation of the ATP2A1 DEL minigene is impaired significantly, it isn't completely eliminated because MBNL1 sequestration is still capable of reducing the inclusion ratio slightly. Following this line of reasoning, splicing of the ATP2A1 DEL minigene in combination with over-expression of MBNL1 caused exon 22 inclusion to increase to 26% (Figure 7B, lane 11). This observation also suggests that the deletion was successful in reducing the MBNL1-mediated splicing effects of the ATP2A1 minigene, but did not eliminate MBNL1's ability to regulate exon 22 inclusion. Taken together, these findings imply that deletion of

the high affinity ATP2A1 site from intron 22 of the minigene is not sufficient to eliminate all of the functional MBNL1 binding sites in the ATP2A1 transcript.

Upon further examination of the ATP2A1 DEL minigene, we observed two remaining regions of intron 22 containing clusters of YGCY motifs (Figure 7C). If these potential MBNL1 binding sites are functional in regulation of the ATP2A1 transcript, then they may be responsible for maintaining MBNL1-mediated regulation of exon 22 in the context of the deletion. Next, we used antisense oligonucleotides (ASO) (Figure 7C) to block the remaining motifs (see Materials and Methods) in combination with the deletion to determine if removing additional YCGY motifs is sufficient to completely eliminate MBNL1-mediated splicing effects on ATP2A1 exon 22. ASOs have previously been shown to affect splicing *in vitro* and in tissue culture (72), (73). Splicing of the ATP2A1 DEL minigene in the presence of each ASO was evaluated both with and without co-expression of MBNL1 and percentage of exon 22 inclusion was quantified (Figure 7D). ASO1 was designed as a control to show that blocking the 5'splice site of the intron with an ASO directed against that region is sufficient to block the splicing machinery from recognizing the 5' splice site, leading to a robust change in the splicing pattern of the minigene that is independent of MBNL1 over-expression. As expected, blocking the 5'splice site with ASO1 resulted in a reduction in exon 22 inclusion to 2%, an effect that was MBNL1 independent (Figure 7D). ASO3 was designed as a control to show that not all ASO binding events alter splicing regulation. In the context of ATP2A1 DEL minigene and MBNL1 over-expression, ASO3 had little effect on MBNL1's ability to regulate exon 22 inclusion (Figure 7D).

ASO2 targets a section of the ATP2A1 DEL that contains four YGCY motifs (Figure 7C) and is a logical candidate for an endogenous MBNL1-recognition site because it is located within the 5' end of the intron, near the alternatively regulated 5'-splice junction. Blocking the potential MBNL1 binding site of the ATP2A1 DEL minigene with ASO2 caused a decrease in exon 22 inclusion to 4% (Figure 7D). Interestingly, over-expression of MBNL1 in combination with ASO2 treatment showed no significant increase in exon 22 inclusion. This observation suggests that blocking the four YCGY motifs immediately downstream of the 5'splice junction via ASO2 is

sufficient to eliminate MBNL1-mediated regulation of ATP2A1 exon 22 in the context of the deletion.

Finally, ASO4 was designed to a region at the 3' end of the intron that also contains four YGCY motifs and is therefore another potential MBNL1 binding site. ASO4 treatment in combination with ATP2A1 DEL reduced exon 22 inclusion to 3%, similar to the effect observed upon ASO2 treatment. Although ASO4 treatment alone is capable of suppressing the ATP2A1 DEL minigene further than the deletion alone, in combination with over-expression of MBNL1, the inclusion was increased to 15% (Figure 7D). This observation suggests that blocking these four YGCY motifs found in ASO4 is not sufficient to eliminate MBNL1's ability to regulate the inclusion of this exon. Not surprisingly however, when the SERCA DEL minigene is spliced with ASO2 in combination with ASO4 the amount of exon 22 inclusion was decreased to 4%. When this minigene is subjected to splicing with ASO2, ASO4, and MBNL1 over-expression no significant change in exon inclusion is observed. Therefore, the decrease in exon 22 inclusion with the combination of ASO2 and ASO4 is MBNL1-independent, suggesting that the combination of ASO2 and ASO4 is also sufficient to completely eliminate MBNL1-mediated regulation of ATP2A1 exon 22 in the context of the deletion.

These experiments indicate that the ATP2A1 transcript, an endogenous target of MBNL1, contains multiple functional and, potentially high affinity, MBNL1 binding sites. It is important to note that in this particular case, simply eliminating one YGCY motif or high affinity binding site is not sufficient to eliminate MBNL1-mediated splicing effects. Rather, multiple YGCY motifs, spanning much of the regulated intron, must be removed or blocked to fully remove MBNL1 regulation.

YGCY Motifs Are Enriched in the Intronic Regions Flanking DM1 Mis-spliced Exons

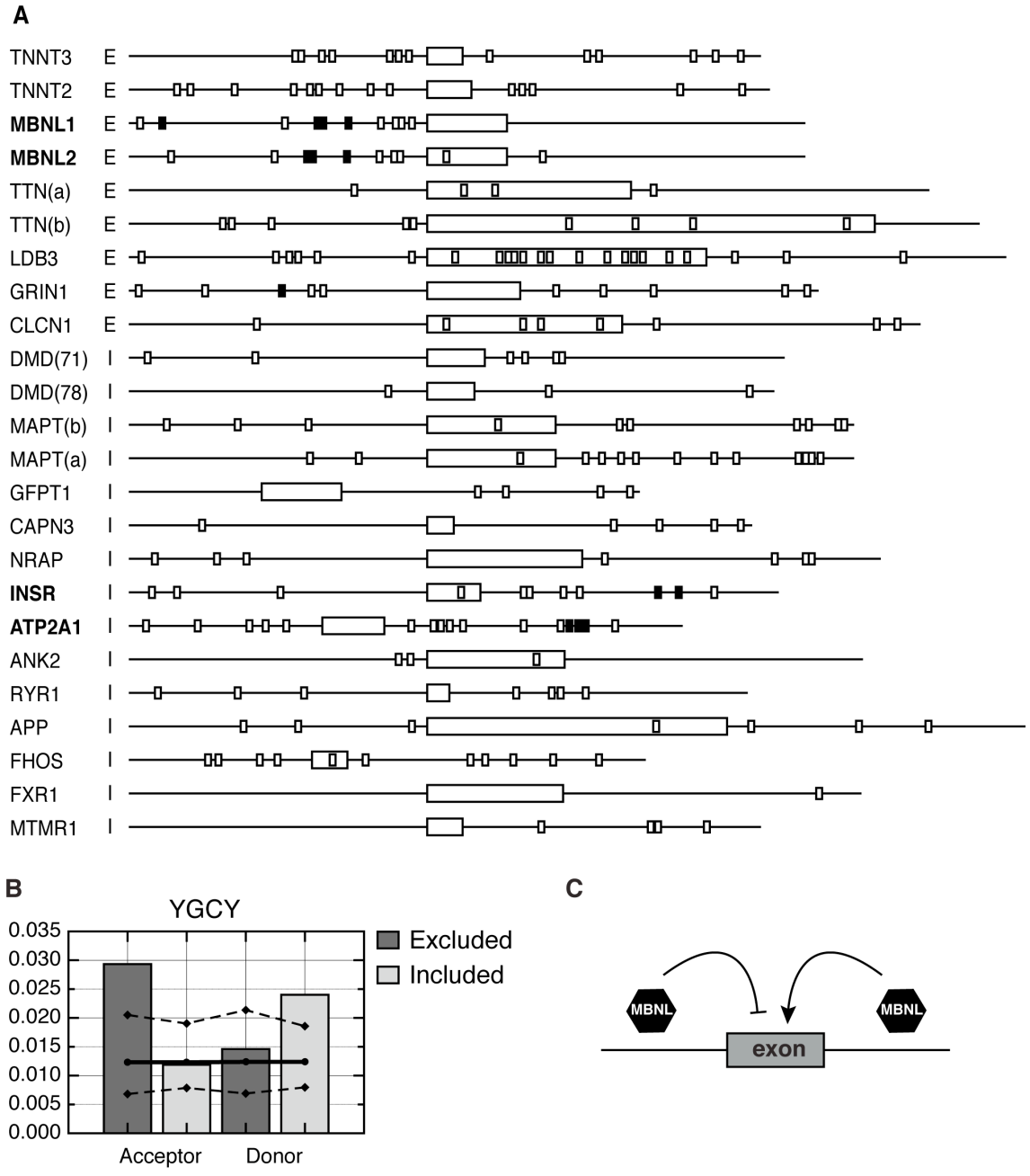
To identify new MBNL1 sites, we analyzed the intronic regions encompassing 24 human exons known to be mis-spliced in DM1. We identified the locations of all instances of YGCY within the last 200 nucleotides of the upstream intron and the first 200 nucleotides of the downstream intron (if either intron was less than 400 nucleotides in length only half of the intron was included) (Figure 8A). Although we don't know if

MBNL1 directly regulates the splicing of all of these exons, all 24 of these exons have one or more instances of YGCY in the adjacent intronic sequences and several contain instances within the exon as well. Considering both the small size and the degeneracy of this motif this observation by itself is not surprising. We carried out several statistical tests to determine if there is a significant correlation between the occurrence of the YGCY motif within these regions and the mis-regulation of splicing in DM1.

A correlation between binding site location and enhancer or silencer activity has been observed for several alternative splicing regulators such as NOVA1 (36) and A2BP1 (38). In order to determine if a similar correlation is observed for instances of YGCY and DM1 related mis-regulation, we evaluated the association between the frequency of YGCY in the upstream (acceptor) and downstream (donor) intronic flanks versus the type of mis-splicing seen in DM1. We observed that, if the exon is included (represented by an I in Figure 8A) in normal individuals (excluded in DM1), occurrences of YGCY are biased towards the donor region (Figure 8B). Conversely, if the exon is excluded (represented by an E in Figure 8A) in normal individuals, the motifs are biased towards the acceptor side (Figure 8B). The chi-squared test for association confirmed the significance of this bias within this dataset ($P = 7.4 \times 10^{-7}$).

Figure 8 (next page)**. Position of YGCY motifs in transcripts that are mis-spliced in DM1.** (A) Schematic representation of exons and flanking regions that are mis-spliced in DM1 with 200 bases (unless the intron is < 400 bases, then only half of the intron was used) upstream and downstream of the mis-spliced exon are represented for each gene. All regions are drawn to scale. Exons are indicated as boxes and introns as lines. The gene names are those that appear in the UCSC genome browser (http://genome.ucsc.edu/). The 'I' or 'E' after each gene indicates that the exon is generally included or excluded in normal individuals. The smaller open boxes denote YGCY motifs. The filled boxes indicate YGCY motifs that are within regions tested for binding to MBNL1 in Figure 6. (B) YGCY motifs are enriched in the upstream acceptor flank of exons that are normally excluded and are enriched in the downstream donor flanks of exons that are normally included. The bars indicate the frequency of YGCY motifs in the upstream (acceptor) and downstream (donor) intronic flanks for the exons shown in panel A. For each bar, the background frequency for YGCY is indicated as a solid line. The upper and lower dashed lines represent the upper and lower 99% confidence intervals that define the range of frequencies expected if the same number of intronic flanks were randomly drawn from the population of all intronic flanks. (C) Model of how MBNL1 binding correlates to exon regulation. When MBNL1 binds

upstream of an alternatively spliced exon, MBNL1 suppresses inclusion of that exon. Conversely, when MBNL1 binds downstream of an alternatively spliced exon, MBNL1 enhances exon inclusion.



In order to determine whether or not YGCY motifs are enriched within these regions relative to human intronic flanks in general, we determined the background frequency of YGCY within the corresponding regions of all human intronic flanks (solid

line, Figure 8B) and then calculated a 99% confidence interval for the range of frequencies of YGCY expected within sample sizes equivalent to those used in this analysis (dashed lines, Figure 8B, and Materials and Methods). This test revealed that, despite the small size and degeneracy of the putative MBNL1 binding site and the small sample size of putative MBNL1 regulated exons currently available, YGCY is significantly enriched above background levels in the acceptor intronic flanks of exons that are excluded in normal individuals and in the donor intronic flanks of exons that are normally included (Figure 8B). Meanwhile the frequency of YGCY in the alternate flank is essentially the level expected by chance. To further demonstrate that this association is a phenomenon of YGCY, we performed identical analysis using several motifs similar to YGCY but which contain specific substitutions: YGUY, YCGY, and RGCR. We observed that, of the motifs tested, only YGCY shows a significant and biased enrichment correlated with the inclusion/exclusion of the adjacent exon. Although YGCY motifs are often located both upstream and downstream of regulated exons (Figure 8A) it is likely that only subsets of these are functionally relevant MBNL1 binding sites. Additional features that we haven't yet fully explored are likely to play important roles in creating physiologically functional sites, such as pyrimidine content and structure. However, the data presented here supports the model that MBNL1 binding upstream of an exon is likely to cause silencing of the downstream splice site while MBNL1 binding downstream of an exon is likely to enhance usage of the upstream splice site (Figure 8C).

DISCUSSION

The data presented here have provided insights into MBNL1's RNA binding specificity. The SELEX experiment recovered many RNAs that bind MBNL1 with high affinity. Analysis of the selected RNAs suggests that MBNL1 has an overall preference for pyrimidines and requires the motif YGCY for high affinity binding. This motif is quite similar to the previously reported motif of $YGCU^{U}/_{G}Y$, which was identified by comparing four MBNL1 binding sites, two in chicken TNNT2 and two in human TNNT2 (66). All nine RNAs that MBNL1 has been shown to bind, including CUG repeats, chicken and human TNNT2, mouse Atp2a1, mouse Tnnt3, and the four new putative sites identified in this study (those with K_ds below 50 nM are being considered putative sites),

have the common motif of YGCY (Figure 6). The analysis of nucleotide frequencies of regions flanking GC dimers suggests the Y positions in YGCY have a strong bias toward uridine (Figure 2C). Interestingly, the four known MBNL1 binding sites (human and chicken TNNT2, mouse Atp2a1 and mouse Tnnt3, Figure 7A) all contain multiple CGCU, UGCU or UGCU motifs and don't contain any CGCC motifs. GRIN1 is the only RNA in Figure 6A that contains CGCC and it binds MBNL1 with lower affinity (Figure 6A). In addition, our work on MBNL1-dependent PLEKHH2 minigene reporter splicing indicates that MBNL1 requires at least one U in either Y position in the YGCY motif to regulate splicing (Figure 5). The CCUG repeats have also been shown to be high affinity binding sites for MBNL1 (55), (59) and the common motif in these repeats is UGCC not CGCC (Figure 6A). Taken together, these results strongly suggest that MBNL1 requires multiple GC steps flanked by at least one U and a pyrimidine on the opposite flank.

Several lines of evidence suggest that local pyrimidine content influences MBNL1 binding: 1) This SELEX experiment demonstrated overall selection for uridines, 2) the native TNNT2 binding site and most of the high-affinity sites adjacent to regulated exons (this study) are generally pyrimidine rich, and 3) CUG and CCUG repeat RNAs are highly pyrimidine rich, and 4) The sequences inserted in PLEKHH2-SM1F, PLEKHH2-SM2 and PLEKHH2-SM3 that confer MBNL1 splicing regulation are 83% pyrimidine. The mouse Tnnt3 and the mouse Atp2a1 sites contain lower levels of pyrimidines in general compared to the other known MBNL1 binding sites (Figure 6). However, the regions of these transcripts proposed to bind MBNL1 (an 18 nucleotide stem-loop and two YGCY motifs within 12 nucleotides of each other, respectively) contain higher levels of pyrimidines (71), (74). In other words, for these specific transcripts the regions of the transcripts proposed to contain an MBNL1 binding site have higher, local pyrimidine content than the overall transcript. Therefore, when only these regions of the transcripts are considered the mouse Tnnt3 and the mouse Atp2a1 transcripts are similar in pyrimidine composition to the other sites. This supports the model that MBNL1 prefers to bind YGCY motifs when they are embedded within pyrimidines. The GRIN1 and INSR RNAs bind more weakly to MBNL1 compared to the other RNAs, and, interestingly, both of these RNAs contain a lower pyrimidine content compared to the other RNAs (Figure 6). This observation also supports the model that

high affinity MBNL1 binding sites will be found within pyrimidine rich sequences. Although MBNL1 bound the GRIN1 and INSR sites more weakly, it is still possible that MBNL1 interacts with these sites and regulates splicing through them, especially if other protein factors enhance the binding of MBNL1 to these sites.

Another commonality among the RNAs that MBNL1 binds with high affinity is that they contain multiple YGCY motifs. A significant fraction of the SELEX RNAs contains two or more YGCY motifs (Figure 3). The CUG repeats, CCUG repeats, TNNT2, ATP2A1 and several potential MBNL1 sites identified in this study also contain multiple copies of the YGCY motif. These results suggest that high affinity MBNL1 sites will be both pyrimidine rich and contain multiple copies of the YGCY motif. It will be important to follow up the identification of these putative new MBNL1 sites with functional studies to determine if the YGCY motifs regulate the splicing of these exons and if their effect on regulating the inclusion and exclusion of these exons is MBNL1 dependent. We showed that in the case of ATP2A1 exon 22 regulation, MBNL1 utilizes multiple YGCY motifs, spanning a large portion of intron 22 for proper splicing regulation. This finding is in agreement with the model that MBNL1 requires multiple YGCY motifs for binding and splicing regulation.

MBNL1 has been shown to bind to several different types of RNA targets. The first type is the toxic CUG and CCUG repeat RNA (47), (59). Specifically the expanded CUG and CCUG repeats which form long, A-form stem-loops containing pyrimidine-pyrimidine mismatches in between G-C and C-G base pairs (75)-(78). Similarly, MBNL1's binding sites in TNNT2 and mouse Tnnt3 pre-mRNAs are stem-loops containing pyrimidine-pyrimidine mismatches (55), (74), suggesting MBNL1 is recognizing all of these sites through a common mode of recognition. A recent crystal structure of Zn fingers 3-4 from human MBNL1 in complex with a 6-mer RNA (CGCUGU) shows that the GC core of the YGCY motif is recognized via its Watson-Crick face, indicating that the GC core is not able to simultaneously bind MBNL1 and form Watson-Crick base-pairs in a stem-loop structure (61). This structure, in combination with the data presented here showing that MBNL1 prefers YGCY motifs embedded in pyrimidine rich RNAs, suggests MBNL1 binds this motif when it is partially or fully unfolded. It is possible that RNA structure in the remainder of the RNA

enhances binding to MBNL1. A similar combination of primary sequence and secondary structure has been shown to be important for NOVA1 binding in which its site can be presented in the loop portion of a stem-loop (34).

We used Sfold (website: http://sfold.wadsworth.org) to predict structures of the six potential MBNL1 binding sites identified in this study (40-45 nucleotides) (data not shown) (79), (80). Interestingly, there is a mix of strong stem-loops and weaker stem-loops and no strong correlation between MBNL1's binding affinity and the stability of the predicted structure. The analysis of the RNAs from the SELEX using Sfold also did not indicate a correlation between predicted RNA structure and MBNL1 binding. It does not appear that within this set of RNAs, a specific RNA structure is required for MBNL1 binding. Structure may be one aspect influencing how MBNL1 regulates splicing and another influence could be binding site location.

The creation of the first "synthetic" MBNL1 regulated exon (PLEKHH2-SM1F) demonstrates that six copies of the motif (CUGCUU) are all that is necessary and sufficient for MBNL1 to block inclusion of an exon. The minigene containing the reverse of the motif in PLEKHH2-SMIF (PLEKHH2-SM1R) shows that the GC core is critical for this control by MBNL1 (Figure 5). We chose to place the MBNL1 motifs 10 nucleotides upstream of the putative branch site and poly-pyrimidine tract in intron 20 so as to avoid direct steric competition with the splicing factors recognizing these elements. This region upstream of the 3' splice site has been shown by us and others to frequently contain splicing enhancers and repressors (81), (82). Therefore, we felt this was an appropriate location for placement of the six CUGCUU motifs. It is likely that MBNL1 is repressing the use of the PLEKHH2-SM1F 3' splice site through a different mechanism than the blocking of U2AF65 binding at the poly-pyrimidine tract, which is how MBNL1 negatively regulates exon 5 of the TNNT2 pre-mRNA (67). In the future it will be interesting to determine the multiple mechanisms through which MBNL1 acts to negatively and positively regulate splicing.

Multiple alternative splicing regulators have been shown to have positional biases with respect to where their binding site is relative to the regulated exon. For example, A2BP1 has been shown to function as a repressor when bound upstream of a regulated exon and as an enhancer when bound downstream (38). NOVA and the CELF family of

proteins have also been shown to have a position-dependent mode of splicing regulation (53), (54). Although there has not been a comprehensive analysis of MBNL1 binding sites, several pieces of evidence suggest MBNL may share common splicing mechanisms with these other factors. The characterization of MBNL1's binding site in TNNT2 confirms that MBNL1 acts as a repressor when binding upstream of exon 5 (55), (66). In addition, the presence of a branch point sequence and a poly-pyrimidine tract near the MBNL1 site suggests that MBNL1 blocks recognition of those sequences by inhibiting the binding of the splicing factor U2AF65 (67). In the case of the PLEKHH2 transcript, placing MBNL1 binding sites upstream of exon 21 also causes MBNL1-dependent repression of exon 21, similar to how MBNL1 regulates TNNT2 exon 5 splicing (Figure 5). In regards to positive regulation, YGCY sites downstream of exon 22 in the human ATP2A1 pre-mRNA function as positive regulators of exon inclusion through MBNL1 binding (Figure 7). And finally, statistical analysis of introns flanking DM1 mis-spliced exons revealed that YGCY sites are enriched upstream of exons that are normally excluded (relative to DM levels) and are enriched downstream of exons that are normally included (relative to DM levels). These results support the model (Figure 8C) that MBNL1 acts as a negative regulator when bound upstream of a regulated exon and as a positive regulator when bound downstream.

Identification of additional MBNL1 binding sites in pre-mRNA targets will be very useful towards understanding MBNL1's alternative splicing mechanisms and to identify the full list of pre-mRNA targets MBNL1 regulates. However, with a potential structural aspect as well as the sequence aspects, MBNL1's binding sites may be degenerate and challenging to recognize in a sea of intronic sequence. This doped SELEX has provided a useful set of criteria for identifying potential MBNL1 binding sites; YGCY motifs embedded in pyrimidines. We have shown that sites created (in the case of PLEKHH2 reporters) or identified (in the case of ATP2A1) using these criteria have proven to be sufficient for MBNL1-dependent splicing regulation. In the future these criteria will aid in the identification of additional MBNL1 targets and currently has furthered the understanding of MBNL1's mode of splicing regulation.

MATERIALS AND METHODS

SELEX. The MBNL1 expression construct included residues 1-260 and contains an N-terminal GST tag. This construct, termed MBNL1 throughout the paper, binds RNA as tightly as the full-length version (55). MBNL1 was expressed and purified as previously described (55). The DNA template used for the SELEX experiment was the following 81-mer oligonucleotide: 5'-GGGAATGGATCCACATCTACGAATTC **(**CCT GTCTCGCTTTTCCCCTCCGCTGCGGCCAC**)** AAGACTCGATACGTGACGAACCT -3'. The oligonucleotide depicted here contains the 32-nucleotide MBNL1 binding site in TNNT2 (in parenthesis) flanked by two constant regions. To create SELEX pool 0, the nucleotides in parenthesis were randomized; 51% of the time the original nucleotide was kept at each position and the other 49% of the time each position was varied to incorporate an equal mix of the other three nucleotides. The forward primer for the SELEX was 5'-GATAATACGACTCACTATAGGGAATGGAT CCACATCTACGA-3' (including a T7 RNA polymerase site for in vitro transcription) and the reverse was 5'-AGGTTCGTCACGTATCGAGTCTT -3'. The initial PCR reaction was amplified by 8 rounds of PCR and $1x10^{14}$ DNA molecules from the PCR reaction were used for the initial transcription reaction. The transcription reactions were done under the following conditions: 500 ng/uL template, 40 mM Tris pH 7.9, 26 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl, 5 mM of each nucleotide, 0.1 ug/100 uL yeast pyrophosphatase, ~2 mg/mL T7 polymerase, and 40 mM DTT. The transcription reactions were performed with trace amounts of [α-P^{32}] CTP for monitoring purposes. After transcription, reactions were treated with DNase for 1 hr at 37°C, gel purified, and ethanol precipitated.

For each round, the appropriate concentration of MBNL1 (see below) was bound to 20 uL of glutathione-agarose beads at 4°C for 15 min, then washed once with 200 uL of SELEX binding buffer (100 mM NaCl, 20 mM Tris pH 7.5, 5 mM $MgCl_2$, 0.02% triton x-100 and 5 mM DTT added the day of use). All washes were done in a similar manner. The RNA was heated in SELEX binding buffer at 95°C for 3 min and placed immediately on ice for 10 min. The MBNL1-bound beads were incubated with the annealed RNA at 25°C for 20 min. The beads were washed once with 200 uL of SELEX binding buffer at 25°C and RNA was released and collected from the beads by a phenol-chloroform extraction and subsequent ethanol precipitation. The binding conditions were

chosen such that 5-10% of the RNA was retained after each round. The collected RNA was reverse transcribed using AMV reverse transcriptase, 1X AMV buffer, 8.8 uM reverse primer, and 2/3rds of the RNA isolated after the round for 1 hr at 42°C. After the reverse transcription was complete, the DNA was amplified by 11 PCR cycles. The DNA was transcribed and the SELEX cycle repeated a total of five times with the concentration of RNA and MBNL1 varying. The concentrations of RNA and MBNL1 used were as follows: Round 1 (20 uM RNA, 20 uM MBNL1), Round 2 (9.8 uM RNA, 4.9 uM MBNL1), Round 3 (8 uM RNA, 2 uM MBNL1), Round 4 (2 uM RNA, 0.33 uM MBNL1), Round 5 (1 uM RNA, 0.13 uM MBNL1). After five rounds were completed, individual clones were isolated by TOPO cloning (Invitrogen) and sequenced.

Labeling of RNAs for gel mobility shift assays. SELEX RNA oligonucleotides were transcribed from DNA templates using T7 RNA polymerase and [α-P^{32}] CTP. Commercially purchased RNA oligonucleotides were kinased using Polynucleotide Kinase and [γ-P^{32}] ATP.

Gel mobility shift assay. Gel mobility shift assays and K_d calculations as described previously (55), with the following modifications. RNA (8 uL) was snap annealed in (75 mM NaCl, 5 mM $MgCl_2$, 15 mM Tris pH 7.5, 0.25 mM β-ME), then mixed with 2uL of protein for final reaction conditions of 175 mM NaCl, 5 mM MgCl2, 20 mM Tris pH 7.5, 1.25 mM β-ME, 10% glycerol, 2 mg/mL BSA, 0.1 mg/mL heparin and trace amounts of bromophenol blue. The binding reactions were incubated for 10-25 minutes at room temperature and 2-4 uL were loaded onto a 6% acrylamide gel (6% 37.5:1 acrylamide: bisacrylamide, 0.5X TB). Gels were run for 35 min at 170 V at 4°C. The K_d values were calculated based on a minimum of three gel mobility shift assays for each RNA, with the exception of F06 UCCA and $(CUGCUU)_6$ RNAs, which were calculated based on two gel mobility shift assays.

Single nucleotide composition of selected RNAs. In order to examine biases in the single nucleotide composition that resulted during selection, we compared the total count for each nucleotide within the randomized region for Round 5 versus those observed in Round 0. To evaluate the statistical significance of the differences observed, we calculated a confidence interval (α=0.05) for the binomial distribution using the previously described method (83) for a population that is the size of the sequences

recovered (82 oligos x 32 nt length = 2624). The expected probabilities used in this estimate were those derived from the frequencies observed within the randomized regions of the Round 0 sequences.

K-mer enrichment. All K-mers from 4- to 6-nt were counted in the final 82 SELEX sequences, including the random sequence plus 5 nt of the constant flanking sequence. The inclusion of the flank allows one to analyze K-mers that overlap with the edges of the constant sequence. To calculate a Z-score for the observed vs. expected occurrences we used Monte-Carlo simulation to estimate the mean (μ) and standard deviation (σ) for the occurrence of all K-mers in 1000 independent populations of 82 sequences constructed using the same random biases used to synthesize the oligos used as input in the SELEX experiment. The Z-score (Z) for each K-mer was calculated according to Equation 1 where μ is the mean observed over the random samples, σ is the standard deviation for the K-mer occurrence within the random samples, and x is the observed occurrence within the population derived from SELEX.

$$Z = \frac{x - \mu}{\sigma} \qquad \text{Equation 1}$$

Construction of splicing reporter constructs. The construction of the MBNL-eGFP plasmid was previously described and obtained from the laboratory of Maury Swanson (84). The DMPK-CUG_{960} plasmid was obtained from the laboratory of Thomas Cooper (66). The wild type PLEKHH2 minigene (PLEKHH2 WT) was constructed by amplifying regions of the PLEKHH2 gene from HeLa genomic DNA using PCR primers containing unique restriction sites. Introns 20 and 21 are 2.4 Kb and 1.7 Kb, respectively, so both were truncated using primers directed against the introns. The construct was built in three segments. The front segment contains exon 20 and the first 922 nucleotides of intron 20 flanked by KpnI and BamHI sites (sense primer: 5'-CGGGGTACCAAATGC TGCAGTTGACTCTCC-3' and anti-sense primer: 5'-CGCGGATCCCTGTGGCTAAC AGGCAGTCA-3'). The middle segment is flanked by BamHI and NotI sites and contains the last 194 nucleotides of intron 20, exon 21, and the first 297 nucleotides of intron 21(sense primer: 5'- CGCGGATCCGGATCATAGATCTGACCCAATG-3' and

anti-sense primer: 5'-AGGAACATAGCGGCCGCTTGAATGAACACCCACTAAT GC-3'). The final segment is flanked by NotI and XhoI sites and contains the last 958 nucleotides of intron 21 and exon 22 (sense primer: 5'-AGGAACATAGCGGCCGCAT CTGCCTACAGGGCACTTG-3' and anti-sense primer: 5'-CCGCTCGAGCCATTCAT GAAGTGCACAGG-3'). The ligated segments were inserted between KpnI and XhoI sites of the pcDNA3 plasmid (Invitrogen), which contains the hCMV promoter/enhancer and a bovine growth hormone (BGH) poly (A) signal after the multiple cloning site.

The PLEKHH2-SM1F, PLEKHH2-SM1R, PLEKHH2-SM2, PLEKHH2-SM3, and PLEKHH2-SM4 constructs were all created from the PLEKHH2 WT template. To create PLEKHH2-SM1F, six, tandem copies of the SELEX motif 5'-CUGCUU-3' were inserted into intron 20 of the PLEKHH2 minigene using PCR and standard cloning techniques. The 36-nucleotide SELEX motif repeat, flanked by ClaI and PacI unique restriction sites (a total of 50 nucleotides), was inserted 40 nucleotides upstream of the 3' splice site of intron 20, and replaced 36-nucleotides (positions -76 to -41 relative to the 3' splice site) of the intron to ensure that the relative size of the intron remains fixed. PLEKHH2-SM1R was created in the same manner except the six SELEX motifs were inserted into the intron backwards, 5'-UUCGUC-3', as a negative control. PLEKHH2-SM2, PLEKHH2-SM3, and PLEKHH2-SM4 were also created in the same manner, but contain six tandem copies of 5'-CUGCCU-3', 5'-CCGCUU-3', and 5'-CCGCCU-3' motifs, respectively.

Construction of ATP2A1 splicing minigenes. The wild type ATP2A1 minigene was constructed by amplifying the region of the ATP2A1 gene from HeLa genomic DNA containing exon 21, intron 21, exon 22, intron 22, and exon 23 using PCR primers (sense primer: 5'-GGGGTACCACCTCACCCAGTGGCTCATG-3' and antisense primer: 5'-CGGGATCCCACAGCT CTGCCTGAAGATG-3') containing unique KpnI and BamHI restriction sites. The PCR product was purified, digested, and ligated into the pcDNA3 plasmid.

The ATP2A1 deletion minigene was created, using standard PCR techniques, by deleting 151 residues of intron 22 (positions -117 to -267). The deletion minigene was created in two segments; the front portion (antisense primer: 5'-TGCTTACAATTG ACGGCTCCAGGTGGAGCTGCGAGCACAAGTG-3') contains exon 21, intron 21,

exon 22, and the first 116 nucleotides of intron 22 flanked by KpnI and MfeI restriction sites. The second section (sense primer: 5'-TGCTTACAATTGGGGCTGCAGTGGGG GGGGGCGGG-3') contains the last 171 nucleotides of intron 22 and exon 23 flanked by MfeI and BamHI sites. The digested fragments were ligated in the KpnI and BamHI sites of pcDNA3.

Cell culture and transfection. HeLa cells were routinely cultured as a monolayer in DMEM+GLUTAMAX media (Invitrogen) supplemented with 10% Fetal Bovine Serum (Gibco) at 37° C under 5% CO_2. Prior to transfection, cells were plated in 6-well plates at a density of $1.8x10^6$ cells/well. Cells were transfected 18-24 hours later at approximately 80% confluency. Plasmid (1 ug/well) and when applicable ASO (100 pmol/well) were transfected into each well with 5 mL Lipofectamine2000 (Invitrogen) following manufacturer's protocol. All ASOs contained a 2'-*O*-Methyl modification at every base and phosphorothioate backbones. The sequences of ASOs used herein to target intron 22 of the ATP2A1 minigene are as follows: ASO1 (5'-GGGCAAGAAGGG GGTGATACCTGTG-3'), ASO2 (5'-GGCGCGGGTGGCAGGGGCACAGCA-3'), ASO3 (5'-TTGACGGCTCCAGGTGGAGCTGCG-3'), and ASO4 (5'-GGCAGGCGGC AGGAGGGCAGCGAG-3'). In double transfection experiments, 500 ng of each plasmid was transfected into a single well; however, in single transfection experiments 500 ng of empty pcDNA3 vector was used to normalize plasmid concentration between wells. Cells were harvested 18-24 hours after transfection using TripLE (GIBCO) and then pelleted by centrifugation. RNA was isolated from the cell pellets using an RNeasy kit (QIAGEN).

In vivo splicing: RT-PCR. Splicing assays were conducted as previously described (82). Briefly, Isolated RNA (500 ng) was incubated with 1 unit of RQI DNase (Promega) in a 10 uL reaction for 1 hour at 37° C. DNased RNA (2 uL; 100 ng) was reverse transcribed in a 10 uL reaction (1:5 dilution) using Superscript II (Invitrogen), according to manufacturer's protocols, with the exception that we used half the recommended amount of Superscript II. All PLEKHH2 reporters and the ATP2A1 minigenes were reverse transcribed using an antisense primer (5'-AGCATTTAGGTGA CACTATAGAATAGGG-3') designed to the Sp6 promoter site of the pcDNA3 plasmid. The minus reverse transcription (-RT) reactions were treated identically to plus reverse

transcription (+RT) reactions except the Superscript II was replaced with water. All reverse transcription reactions (2 uL) were subjected to 22 rounds of PCR amplification, which was found to be within the linear range for all primers used (data not shown), in a 20 uL reaction (1:10 dilution). PCR to analyze the splice products of all PLEKHH2 reporters was conducted using the sense cloning primer designed to exon 20 and the antisense cloning primer designed to exon 22 described above.

The ATP2A1 WT sense and antisense cloning primers designed to exon 21 and 23, respectively, were used for PCR of the ATP2A1 minigene. For quantification of the PLEKHH2 reporters, PCR was conducted using radiolabeled sense cloning primer. For the ATP2A1 minigene splicing, radiolabeled antisense primer was used in the PCR reaction for quantification. The resulting PCR products were resolved by electrophoresis on 8% (19:1) polyacrylamide native gels. The gels were dried and exposed overnight on a phosphorimager screen. Quantification of the radioactive bands corresponding to splice products was performed using ImageQuant software (Molecular Dynamics). The percent of exon inclusion was calculated by dividing the amount of the band corresponding to inclusion splice product by the total amount of splice product (calculated by adding the inclusion splice product band to the exclusion splice product band). Percentage of exon exclusion was computed in the same manner, except the exclusion product was divided by the total amount.

Calculation of the enrichment of YGCY motifs in the regions flanking exons mis-spliced in DM1. The identities of 24 exons that are mis-spliced in human DM1 tissues were gleaned from a search of literature. The sequences for the exons and up to 200 nt of the flanking introns were recovered from public databases. If the intron was less than 400 nt in length only the adjacent half was evaluated. Each exon was categorized as 'excluded' (E) or 'included' (I) if it was generally excluded or included, respectively, in normal tissues relative to DM1 tissues. The frequency of YGCY motifs in the flanking intronic regions was calculated according to equation 2 where n equals the number of YGCY motifs observed, l equals the length of each intronic sequence evaluated, and k indicates the number of introns evaluated. The denominator effectively represents the total number of positions that could contain the motif within the set of sequences evaluated.

$$f = \frac{n}{\sum_{k}(l-3)}$$ Equation 2

The background frequency of YGCY in the upstream and downstream intronic flanks for all human introns (according to the UCSC hg18 annotation) was also calculated. The upper and lower confidence intervals (α=0.01) for the frequencies expected in a randomly drawn sample that is the size of the regions of interest was calculated using the background frequencies, the sample sizes, and the method of Agresti et al. (83).

CHAPTER III

COMBINATORIAL MUTAGENESIS OF MBNL1 ZINC FINGERS ELUCIDATES DISTINCT CLASSES OF REGULATORY EVENTS

This manuscript is currently in preparation. I designed and cloned all mutants used in this study, designed and conducted all splicing experiments and data analysis, purified mutant proteins, and prepared the manuscript; Julia Oddo and I performed binding experiments; Eric Wang contributed statistical correlation analyses. Dr. J. Andrew Berglund was the principle investigator of this work and contributed to experimental design and manuscript preparation.

INTRODUCTION

Alternative splicing is a cellular mechanism that is used to create proteomic diversity from a limited number of genes. Although the mechanisms that govern the regulation of alternative splicing are vast, the overarching mechanistic theme is that alternative splicing results from a careful balance between positive and negative splicing signals in the pre-mRNA and the relative concentrations of the many proteins involved in recognizing these signals (reviewed in (2), (85), (22), (25) and (27)).

Muscleblind-Like 1, or MBNL1, is an alternative splicing factor that has been the focus of intense study over the last decade due to its involvement in Myotonic Dystrophy (DM) pathogenesis (43), (86). DM is a debilitating, multi-systemic disease that is caused by the expansion of certain non-coding, CTG- and CCTG-containing repeats within the genome (44)-(46). Once transcribed into RNA, the CUG- or CCUG-containing expansions form stable structures that are capable of aberrantly sequestering RNA binding proteins, including MBNL1, in structures referred to as foci (48), (49), (87). Once sequestered to the toxic RNA, MBNL1 is no longer able to perform its normal cellular role in the regulation of key splicing events, leading to mis-splicing and ultimately disease symptoms (62), (51). Many disease-associated and MBNL1-dependent splicing events have been defined, however mechanistic insights into how MBNL1 regulates splicing are limited. An important step towards a comprehensive understanding of the regulatory mechanisms governed by MBNL1 is a thorough understanding of how MBNL1 recognizes its cellular targets. To this end, the following comprehensive study

was conducted to evaluate the consequences of MBNL1 mutagenesis on splicing function and RNA binding.

The architecture of MBNL1 is seemingly very simple. MBNL1 contains four zinc fingers (ZFs) of the $CX_7CX_{4-6}CX_3H$-type (57), (47). The four ZF domains are the only known RNA binding domains of MBNL1. The ZFs are commonly referred to as ZF1, ZF2, ZF3, and ZF4 and are arranged such that the first two ZFs fold into one domain and the second pair of ZFs folds into another domain, thus resulting in two RNA binding domains. Crystal structures have been obtained of just the ZF1-2 and ZF3-4 domains of MBNL1 (61). Sequence alignments and structural overlays of the two ZF domains (*i.e.* ZF1-2 and ZF3-4) show that ZF1 and ZF3 have high sequence similarity, as do ZF2 and ZF4 and that the two domains have nearly identical structures. The major structural difference between the two domains is an extended a-helix that extends off of the C'-terminus of the ZF1-2 domain. The conservation and structural similarity has led some to believe that the two pairs arose from a tandem duplication event (60). The nature of this protein's architecture suggests that the pairs of ZFs are capable of functioning as independent domains, with the ZF1-2 pair and ZF3-4 pair having redundant function. The idea of functional redundancy between the ZF pairs arises from studies performed on the *Drosophila Melanogaster* ortholog of MBNL1, *muscleblind* (*mbl*). It was shown that an isoform of Mbl, containing only one ZF pair, was able to regulate splicing of some MBNL1 target transcripts in mammalian cell culture systems (88).

Like many splicing factors, such as CUGBP1 (89), FOX 1/2 (38), (39) , Nova 1/2 (36), (37) and PTB (31), MBNL1 can act as either a splicing suppressor or activator in a transcript-specific manner. Also similar to these splicing factors, there appears to be a positional dependence between MBNL1 binding site position, relative to the regulated exon, that determines if MBNL1 enhances or inhibits splicing of that exon (40), (41). For MBNL1 the pattern is that binding upstream of a regulated exon results in suppression, whereas when MBNL1 binds downstream of an exon it enhances inclusion. Various studies from multiple groups have defined the minimal consensus binding site of MBNL1 as 5'-YGCY-3' (66), (40), (41).

A handful of functional binding sites within MBNL1-regulated transcripts have been defined, such as: cardiac troponin T type 2 (TNNT2 or cTNT) (66), (55),

sarcoplasmic/endoplasmic reticulum Ca^{2+}-ATPase 1 (ATP2A1 or SERCA1) (71), (41), insulin receptor (INSR) (90), (91), mouse nuclear factor I/X (Nfix) (40), mouse very low density lipoprotein receptor (Vldlr) (40), and within the MBNL1 pre-mRNA (*MBNL1*) (92). To eliminate any confusion the *MBNL1* transcript will be italicized throughout the text. No obvious rules for MBNL1 function have been gleamed from the architecture of the MBNL1 binding sites within these transcripts. For some binding events, such as TNNT2, INSR, and Vldlr, the functional MBNL1 binding sites appear to contain only a few YGCY motifs occurring within 50 nucleotides of the regulated splice site. However, in other transcripts, such as ATP2A1, Nfix and *MBNL1*, it appears that MBNL1 requires multiple clusters of YGCY motifs spaced throughout the intronic and exonic sequence, with no single YGCY motif being sufficient for function. Although great strides have been made in defining the architecture of MBNL1 binding sites within endogenous targets, questions remain to be answered, such as: 1) Are ZF1-2 and ZF3-4 domains functionally equivalent? 2) Are all transcripts recognized by MBNL1 in the same fashion? 3) Does MBNL1's splicing activity correlate with its RNA binding affinity? In order to address these questions and advance the mechanistic understanding of MBNL1-regulated splicing we have conducted a combinatorial mutagenic study of the ZFs of MBNL1 to refine our understanding of the functional RNA:MBNL1 interaction.

RESULTS

Ablation of Individual Zinc Fingers Has Minimal Effect on MBNL1 Function

To determine the functional importance of each individual ZF, the terminal Cys and His residue of each finger was mutated to alanine, resulting in ZFs with a $CX_7CX_{4-6}AX_3A$ motif (Figure 1A). By mutating two of the four zinc ligands we expected that the mutated ZF would no longer be able to chelate zinc, which would therefore most likely ablate the entire ZF domain. All of the mutants were cloned and expressed as N'-terminal eGFP fusions. Proteins levels for each chelation mutant (CM) were assessed to ensure that the mutations didn't significantly alter expression of the protein (Figure 1B).

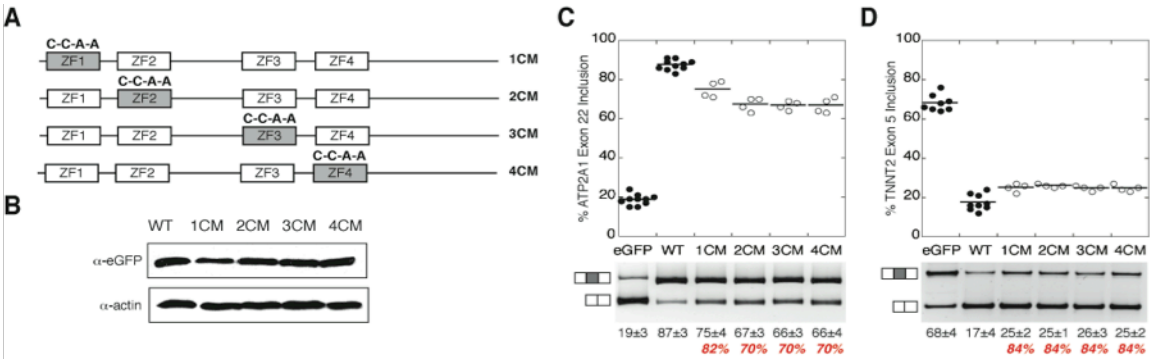


Figure 1. Single ZF Chelation Mutants (sCMs). (A) Schematic depicting individual ZF mutations of MBNL1. ZF motif mutations are shown above shaded boxes. Open boxes represent wild-type ZFs. (B) Western blot showing expression levels of single CMs in comparison to wild-type MBNL1 in HeLa cells. (C) and (D) Dot plots of ATP2A1 and TNNT2 splicing with single CMs. Values for percent exon inclusion and percent splicing activity (red) are listed below representative splicing gels in Panels C and D. Values were obtained from at least 4, independent splicing experiments.

Each chelation mutant was then tested in a cell-based splicing assay to assess if mutational ablation of any individual ZF would have an effect on MBNL1's ability to regulate splicing. The splicing of two minigene reporters, TNNT2 and ATP2A1, was analyzed with these mutants. These particular minigenes were chosen because: 1) of their robust splicing response in the cell-based assay, 2) splicing of both minigenes has been well characterized, and 3) they represent both major classes of MBNL1 regulated event, exon exclusion and exon inclusion. Exon 22 of ATP2A1 is an example of a cassette exon event wherein MBNL1 activates exon inclusion. When this reporter was spliced in the absence of MBNL1, exon 22 inclusion was 19±3%. When wild-type MBNL1 protein was co-expressed with the reporter the inclusion increased to 87±3%. Of the 4 CMs tested, the mutation of ZF1 (1CM) had the weakest affect on splicing of the ATP2A1 minigene with this mutant retaining 82% of splicing activity in comparison to wild-type. In comparison to 1CM, the other three mutants (*i.e.* 2CM, 3CM, and 4CM) suffered a greater reduction in function, but are still able to regulate exon 22 inclusion with 70% of wild-type activity (Figure 1C). Under cellular conditions where MBNL1 is not over-expressed, exon 5 of the TNNT2 transcript is included at 68±4%. However, when MBNL1 is co-expressed with the TNNT2 exon inclusion is reduced to 17±4%. All

of the CMs were able to regulate splicing of TNNT2 exon 5 with nearly identical activity to each other, and activity was reduced only 16% compared to wild-type protein (Figure 1D). These results suggest that no individual ZF is absolutely required for MBNL1 to function as either a splicing repressor or activator for these splicing events.

MBNL1 Requires One Zinc Finger Pair for Splicing Regulation

A combinatorial approach was employed wherein the individual ZF chelation mutations from Figure 1 were combined in all possible combinations to create double ZF chelation mutants. In these mutants, two of the four ZFs were simultaneously mutated from C-C-C-H to C-C-A-A domains as depicted in schematic form in Figure 2A. Expression of each double CM was verified in HeLa cells via Western blot (Figure 2B). Splicing activity of each of the six double CMs was assayed using ATP2A1, TNNT2, and *MBNL1* minigene reporters to determine what affect ablating two of the four ZFs had on splicing activity. MBNL1 has been shown to auto-regulate the inclusion of exon 5 of the *MBNL1* pre-mRNA transcript (92). Splicing of *MBNL1* in HeLa cells resulted in 72±2% exon 5 inclusion. When MBNL1 is over-expresses in the presence of this reporter the inclusion of exon 5 is suppressed to 8±2%. For all three minigene reporters tested, we observed that any of the double mutants that resulted in one CM in each pair of ZFs (*i.e.* 1,3CM, 1,4CM, 2,3CM and 2,4CM) resulted in protein with less than 10% of wild-type splicing activity (Figure 2, compare panels C-E). For the ATP2A1 and TNNT2 minigenes the 1,2CM and 3,4CM proteins were able to regulate splicing with approximately 80% wild-type activity, which is nearly identical to the values observed with the single ZF CMs (compare Figure 1C to 2C and Figure 1D to 2D). These observations support the idea that MBNL1 only requires one pair of ZFs to function at near wild-type levels for these reporters.

Contrary to this model, was the observation that the 1,2CM retains only 34% of wild-type spicing activity with the *MBNL1* minigene reporter (Figure 2E). However in stark contrast, the 3,4CM, which retains a wild-type ZF1-2 pair, regulated splicing of this minigene with 93% of wild-type activity (Figure 2E). Taken together these observations provide the first evidence of a differential splicing activity between the ZF1-2 and ZF3-4 pairs, indicating that the ZF pairs are not functionally equivalent for all splicing events.

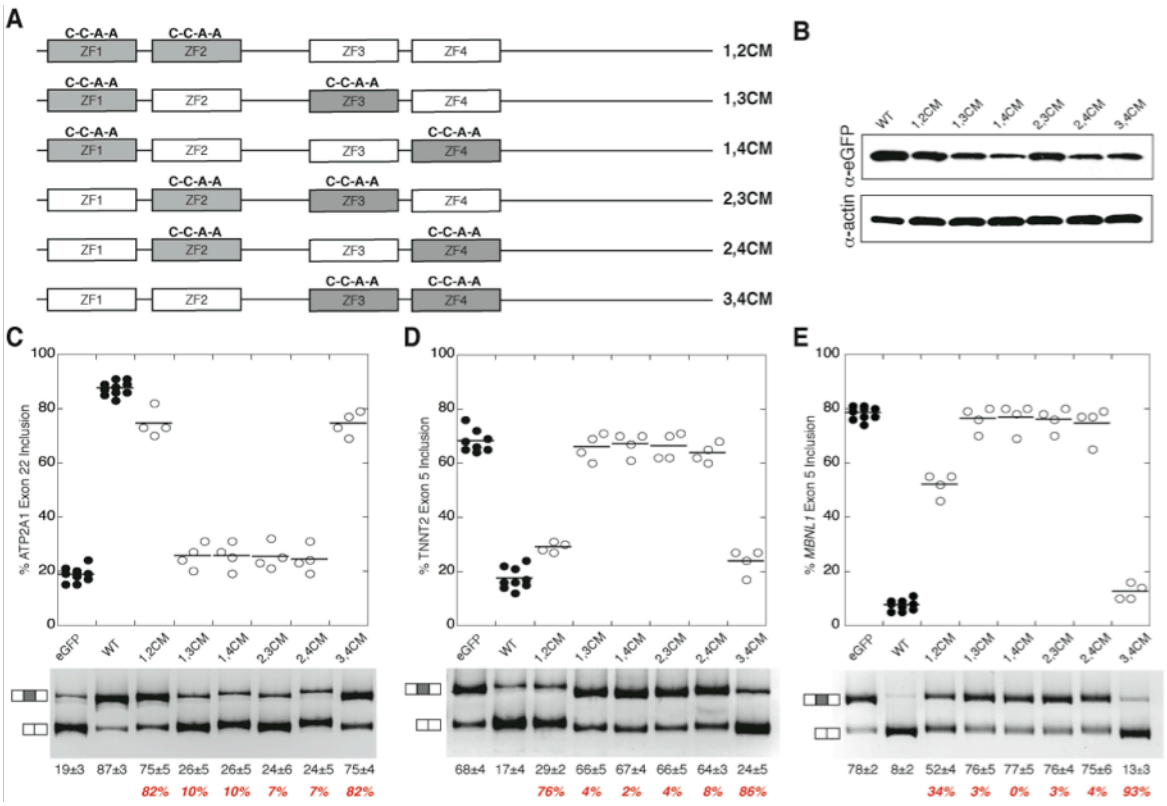


Figure 2. Double ZF Chelation Mutants (dCMs). (A) Schematic depicting double ZF mutations of MBNL1. ZF motif resultant from mutation is shown above shaded boxes. Open boxes represent wild-type ZFs. (B) Western blot showing expression levels of double CMs in HeLa cells in comparison to wild-type MBNL1. (C), (D), and (E) Dot plots of ATP2A1, TNNT2, and *MBNL1* splicing with double CMs, respectively. Values for percent exon inclusion and percent splicing activity (red) are listed below representative splicing gels in Panels C, D, and E. Values were obtained from at least 4, independent splicing experiments.

The nearly complete lack of splicing activity for 1,3CM, 1,4CM, 2,3CM, and 2,4CM on all three minigenes tested suggest that the mutations used to ablate the ZFs are too severe to leave the other half of the ZF-ZF domain functional. One plausible explanation for the lack of activity observed for these mutants is that both ZF domains are partially or fully unfolded, resulting in a protein with essentially no functional RNA binding domains. Unfortunately with only 4 ZFs, MBNL1 has very little secondary structure, which makes testing this hypothesis with classical biophysical techniques impossible. However, the contrast in activity that we observed between the ZF1-2 and ZF3-4 pairs in regards to splicing of the *MBNL1* minigene warranted examining the

functional proficiency of non-traditional ZF pairings, such as ZF2-3 and ZF1-4, in a context other than chelation mutants. To this end a more subtle mutagenic approach was devised to evaluate each ZF's contribution to overall splicing activity of MBNL1.

Mutations that Disrupt Specific MBNL1-RNA Interactions Reveal Two Distinct Classes of Splicing Events

We embarked upon creating mutations in the ZFs of MBNL1 designed to inhibit the individual ZF's ability to interact with RNA without seriously perturbing the overall structure of the ZF. Using guidance from the crystal structure of MBNL1's ZF3-4 domain bound to a 6-nucleotide (nt) RNA ligand (61), we mutated residues in each ZF that were shown to be important for either hydrogen bonding or base stacking interactions with the 5'-GC-3' dinucleotide of the RNA ligand. Particular interest was paid to targeting the interactions with the GC dinucleotide because the wealth of biochemical data suggests that the majority of the MBNL1:RNA interaction can be explained by interactions with the GC dinucleotide. Glu187 of ZF3 and Asp223 of ZF4 are key acidic residues that form hydrogen bonding interactions with the RNA backbone and cytosine of the GC dinucleotide. Therefore these amino acids were chosen as mutagenic targets in order to reduce the hydrogen bonding potential of the ZFs with the GC dinucleotide. No crystal structure is available for MBNL1's ZF1-2 domain with a ligand bound, however the structures of the ZF1-2 domain and ZF3-4 domain superimpose on one another, enabling us to identify the corresponding acidic residues in ZF1 (Glu21) and ZF2 (Asp55) through alignment.

Initially we created a subset of single ZF and double ZF mutants with only these acidic residues mutated to Alanine and tested their splicing. We observed no measurable effects on splicing regulation until three out of four zinc fingers had the targeted acidic residues mutated. Therefore, we added a second point mutation to each ZF to further reduce the ZF:RNA interaction. The second mutation was designed to target the py-stacking interaction that occurs between Phe202 of ZF3 or Tyr236 of ZF4 with the guanidine of the RNA substrate. As previously mentioned, alignments were used to identify the paralagous resides in ZF1 (Phe36) and ZF2 (Tyr68). All residues were mutated to Alanine (see schematic in Figure 3A). The resulting MBNL1 RNA Interaction

Mutants (RIMs) had no significant change in protein expression level compared to wild-type MBNL1 (Figure 3B).

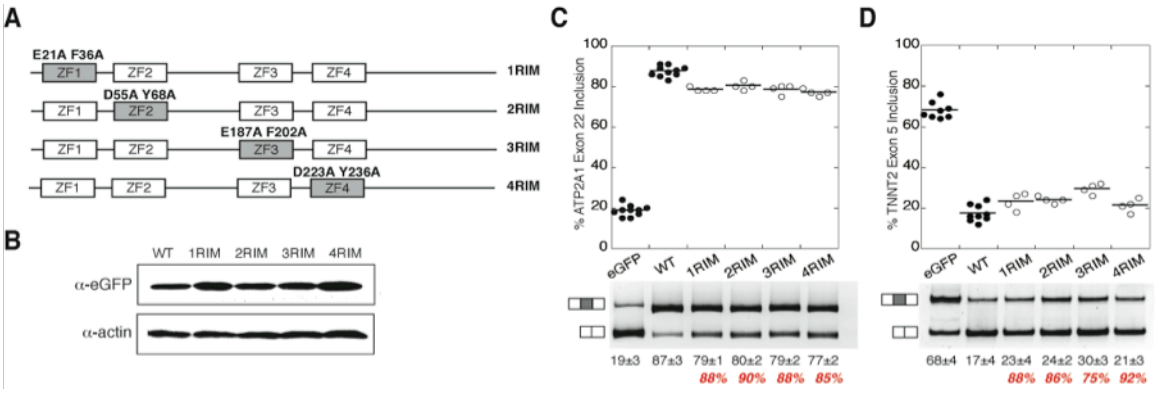


Figure 3. Single ZF RNA Interaction Mutants (sRIMs). (A) Schematic of single ZF RIMs. Open boxes represent wild-type ZFs. Shaded boxes represent mutated ZFs. Mutated residues are denoted above shaded boxes. (B) Western blot comparing expression levels of single RIMs to wild-type MBNL1 in HeLa cells. (C) and (D) Dot plot of ATP2A1 and TNNT2 splicing with single RIMs. Values for percent exon inclusion and percent splicing activity (red) are listed below representative splicing gels in Panels C and D. Values were obtained from at least 4, independent splicing experiments.

Splicing of two reporter minigenes, ATP2A1 and TNNT2, was conducted to assess the functional consequence of these mutations. Splicing of the ATP2A1 minigene with single ZF RIMs resulted in a 10-15% reduction in splicing (Figure 3C). Similarly, splicing activity of the RIMs was reduced 8-25% for the TNNT2 splicing reporter (Figure 3D). In comparison to the single CM mutations (Figure 1), in which the ZFs are ablated, it appears as though the RIMs generally have a weaker effect on splicing (compare Figure 1C to 3C and Figure 1D to 3D), indicating that the RIMs are more subtle mutations.

The single ZF RIMs, which contain two point mutations per ZF, were combined in all possible combinations to create a series of MBNL1 double ZF RIMs. The double RIMs contain a total of four point mutations per construct. The position of each mutation is depicted in schematic form in Figure 4A. Mutations that affect two of the four ZFs had a moderate effect on the expression of some mutants in HeLa cells. In comparison to wild-type protein expression levels, 1,2RIM, 1,4RIM, 2,4RIM and 3,4RIM had slightly decreased expression when analyzed via Western blot (Figure 4B). However

the decreases in expression did not correlate with a decrease in the splicing function, and these protein levels are still over-expressed in comparison to endogenous MBNL1 protein levels.

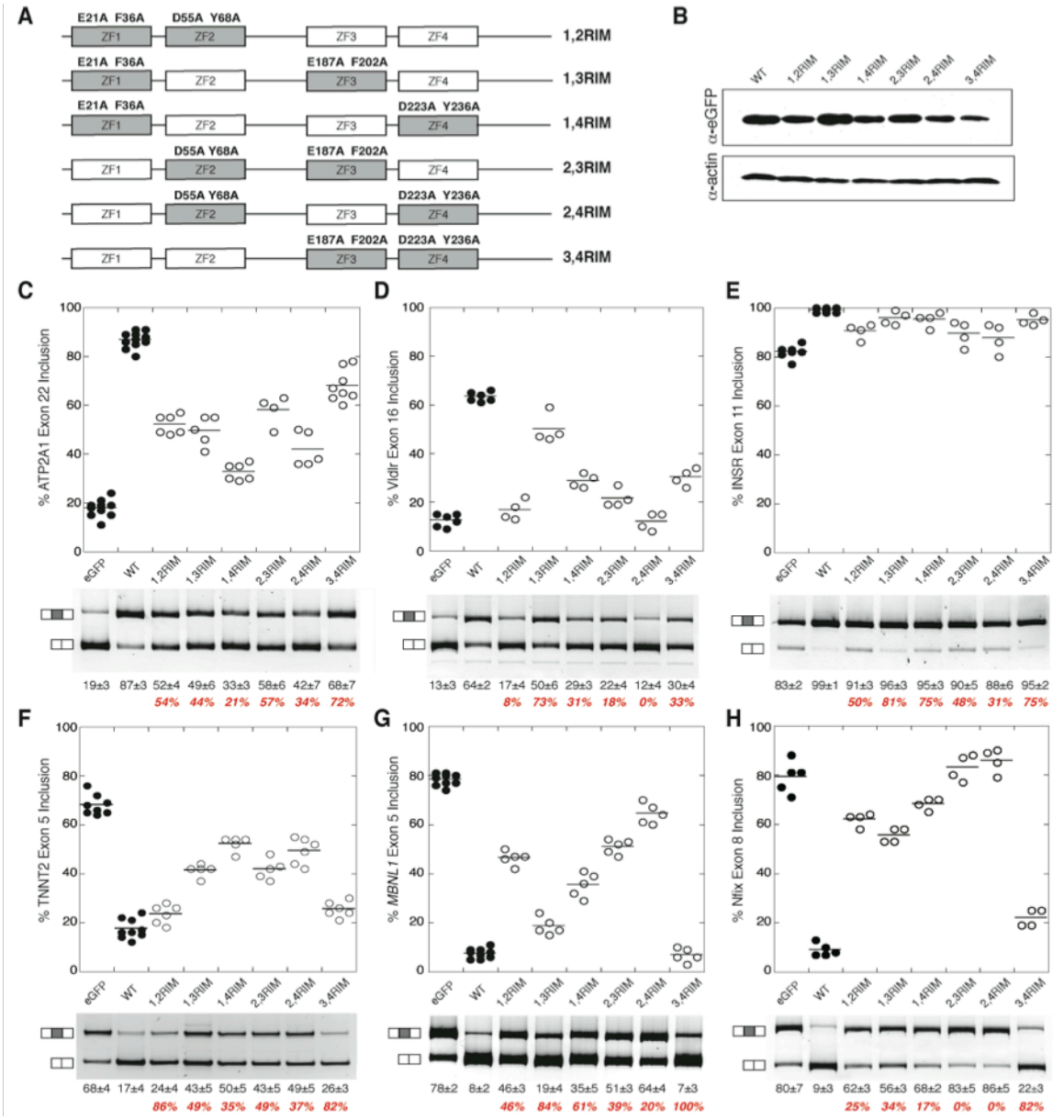


Figure 4. Double ZF RNA Interaction Mutants (dRIMs). (A) Schematic of double RIMs. Open boxes represent wild-type ZFs. Shaded boxes represent mutated ZFs. Mutated residues are denoted above shaded boxes. (B) Western blot comparing expression levels of double RIMs to wild-type MBNL1 in HeLa cells. (C) - (E) Dot plots of ATP2A1, Vldlr, and INSR splicing with double RIMs. (F) – (H) Dot plots of TNNT2, *MBNL1*, and Nfix splicing with double RIMs. Exon inclusion (%) and splicing activity (%; red) values are listed below representative splicing gels in Panels C-H. Splicing values were obtained from at least 4, independent experiments.

Initially the splicing of ATP2A1 and TNNT2 minigenes was assayed. Contrary to what was observed with the double CMs, differences in splicing activities amongst the different combinations of ZF pairs became apparent with the RIMs. In the case of ATP2A1 splicing, 1,2RIM, 1,3RIM, and 2,3RIM all had similar splicing activities. These three mutants were able to regulate splicing at about 50% of wild-type activity. The 1,4RIM and 2,4RIMs were the least active mutants with regards to ATP2A1 splicing with only 21% and 34% activity, respectively. Also notable is the fact that the 3,4RIM, which retains a wild-type ZF1-2 pair, had the highest activity (72%) of all RIMs for the ATP2A1 reporter (Figure 4C). The splicing pattern observed for ATP2A1 with the RIMs deviated substantially from the pattern of ATP2A1 splicing observed with the double CMs (compare Figure 2C to Figure 4C).

In contrast to the ATP2A1 minigene, splicing of the TNNT2 minigene with the double RIMs yielded a splicing pattern similar to that observed for the double CM splicing. For the TNNT2 minigene, all of the RIMs that contained a mutated ZF in each pair (*i.e.* 1,3RIM, 1,4RIM, 2,3RIM, and 2,4RIM) had similarly reduced splicing activities, only regulating splicing with 35%-49% activity. However, mutations that retain an intact wild-type ZF pair (*i.e.* 1,2RIM and 3,4RIM) retain close to wild-type activity (Figure 4F). This was the same splicing trend observed for splicing of the TNNT2 minigene with the double CMs (compare Figure 2B to Figure 4F), with the major exception that the 1,3RIM, 1,4RIM, 2,3RIM, and 2,4RIMs retain significantly more splicing function than their CM counterparts. It is important to note that both the 1,2RIM and 3,4RIMs have nearly identical splicing activity (86% and 82%, respectively) in regards to the TNNT2 minigene, suggesting that for this splicing event the ZF pairs have nearly identical function and in this case could be considered functionally redundant or interchangeable.

The differential splicing patterns observed for the double RIMs between TNNT2 and ATP2A1 minigenes, combined with the observation that the *MBNL1* minigene showed an interesting difference in splicing activity between the 1,2CM and 3,4CM (Figure 2E), prompted us to broaden our analysis to include three splicing events from each type (negative versus positive) of MBNL1 regulated event. We were interested in investigating how various transcripts might result in differential splicing

activities with the double RIMs. As previously mentioned ATP2A1 is a specific event where MBNL1 regulates exon inclusion by binding downstream of a regulated exon. To further characterize the behavior of the double RIMs on substrates of this type two additional minigenes were chosen to represent MBNL1-mediated exon activation events in which the MBNL1 binding sites have been characterized: human INSR and mouse Vldlr.

Vldlr showed a large difference in splicing activity amongst the double RIMs (Figure 4D). 1,3RIM was the most active RIM with 73% wild-type activity. It is surprising that 1,3 RIM, which retains wild-type ZFs 2 and 4, retains more splicing activity than either 1,2RIM or 3,4RIM, both of which have a wild-type pair of ZFs preserved. With approximately 30% function, 1,4RIM and 3,4RIM function nearly identically in regards to Vldlr splicing. 1,2RIM and 2,3RIM have very little function, only 8% and 18%, respectively. 2,4RIM was completely inactive and unable to regulate splicing of the Vldlr minigene to any measurable degree (Figure 4D). Interestingly, the mutant that retained wild-type ZF2-4 was the most active and the mutant with a mutated ZF2-4 was completely inactive, suggesting that ZFs 2 and 4 are quite important for MBNL1's ability to regulate the splicing of Vldlr.

The INSR minigene has a very small splicing range with exon 11 inclusion only ranging from 80 to 100%, which makes differences between the double RIMs less obvious than those observed with Vldlr, however still significant. Much like the splicing profiles observed for Vldlr, 1,3RIM retains the highest splicing function on INSR with 81% activity. 1,2RIM and 2,3RIM have nearly identical activities, retaining approximately 50% of wild-type function. 1,4RIM and 3,4RIM also have identical function with 75% activity. The 2,4RIM had the lowest level of splicing function on the INSR minigene of all of the double RIMs with 31% activity (Figure 4E).

Exons 5 of both TNNT2 and *MBNL1* represent MBNL1 regulated events wherein binding upstream of a regulated exon induces exon skipping. Splicing of mouse Nfix was chosen as the third minigene of this type. The MBNL1 binding sites within these transcripts have been characterized. The differences in splicing activity observed amongst the RIMs on the *MBNL1* minigene substrate yielded a splicing pattern that is quite interesting (Figure 4G). There is a very significant difference in the splicing activity

of the two ZF pairs for this minigene. The 1,2RIM, which retains a wild-type ZF3-4 pair, has only 46% activity on this substrate while the 3,4RIM retains 100% wild-type activity. This observation strongly correlates with the double CM data wherein 3,4CM has close to wild-type activity and splicing function of 1,2CM was drastically reduced (compare Figure 4G to 2E). It is important to note the significant similarity in the splicing activity profile between *MBNL1* and Vldlr minigenes, with the largest exception being the activity of 3,4RIM. For *MBNL1* the 3,4RIM has 100% activity, but for the Vldlr minigene, the 3,4RIM only retains 33% activity. 1,4RIM and 2,3RIM have intermediate splicing activities for *MBNL1* of 61% and 39%, respectively. As seen in both Vldlr and INSR splicing, 1,3RIM has robust function in regards to *MBNL1* splicing at 84%, second in activity only to 3,4RIM. Also similar to the splicing trends observed for other substrates, 2,4RIM is clearly the least active of all mutants for *MBNL1* with only 20% activity (Figure 4G).

Although there are some similarities in splicing activities between Nfix and the other minigenes analyzed herein, Nfix has a few differences that make it unique (Figure 4H). With a 57% difference in the splicing activity between 1,2RIM and 3,4RIM, Nfix arguably has the largest discrimination between the ZF pairs. 1,2RIM and 1,4RIM have similarly reduced function at 25% and 17% activity, respectively. In contrast to Vldlr, INSR, and *MBNL1* the 1,3RIM retains only slightly higher activity at 34%, than the other mutants. On the other hand, the trend holds that 2,4RIM has no measurable splicing activity for the Nfix substrate. Uniquely however, 2,3RIM also has no splicing activity. In all other minigenes examined the 2,3RIM retained at least 20% activity. Similar to TNNT2 and *MBNL1*, the 3,4RIM retains the most function in regards to Nfix splicing with 82% of wild-type activity.

The trend that 1,2RIM has less splicing activity than 3,4RIM on most reporters strongly suggests that maintaining a wild-type ZF1-2 pair is more important for splicing activity than the ZF3-4 pair. The ZF3-4 pair seems to be completely dispensable for splicing of *MBNL1*, TNNT2, and Nfix. These splicing profiles also illuminated an observation unique to this study that ZFs 2 and 4 appear to have a significant role in splicing regulation for most substrates.

We conducted pair-wise correlation analysis of the double RIM splicing

activities for all 6 splicing events (Figure 5). Of the 15 possible iterations of paired, double RIM splicing events, 8 events have splicing profiles that correlate very well (R ≥0.80). For 3 of the paired-events (TNNT2 vs. Vldlr, TNNT2 vs. INSR, and INSR vs. ATP2A1) there was no significant correlation between splicing activity (R value less that 0.50). The 4 remaining events had less significant correlations with R-values ranging from 0.50 to 0.70.

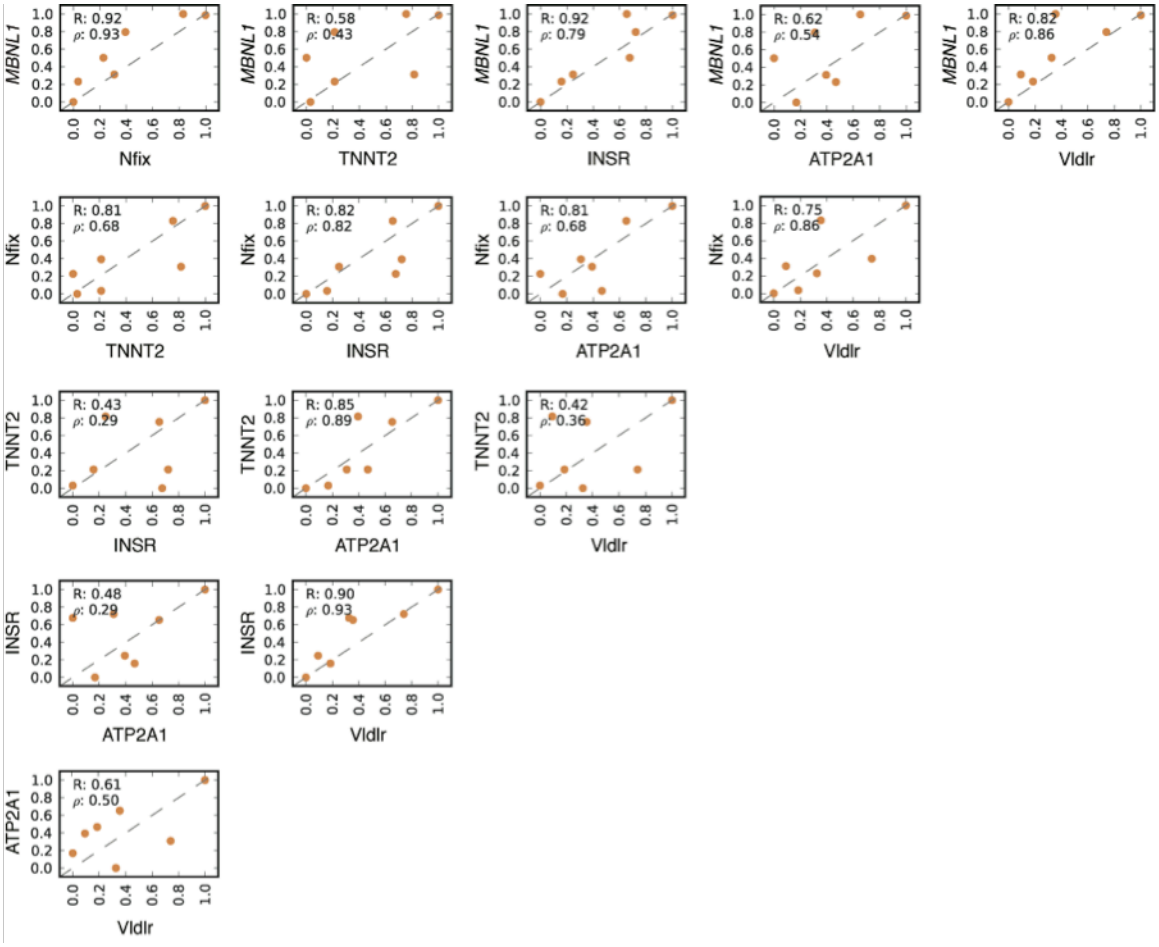


Figure 5. Pair-wise correlations comparing dRIM splicing activity for all six minigenes. Correlations were performed using splicing activity (%) values from Figure 4. Pearson's (R) and Spearman's rank (ρ) correlation values are provided for each pair.

We found the pair-wise correlations encouraging and clustered all of the splicing data by both minigene and RIM. The clustering revealed that of the 6 minigenes

tested, two major classes of splicing events exist (Figure 6). The first class (Class I) contains only ATP2A1 and TNNT2. The second class (Class II) contains Vldlr, *MBNL1*, INSR , and Nfix. These classes are evident in Figure 6, where it is clearly apparent that the splicing patterns of TNNT2 and ATP2A1 are very similar to each other and distinct from the other four minigenes, which have very similar splicing patterns.

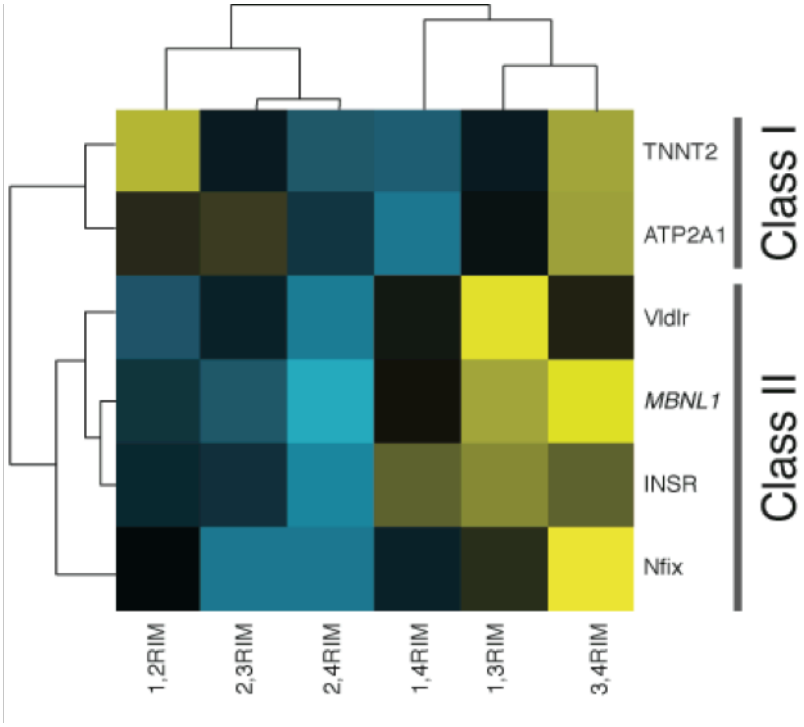


Figure 6. Double RIMs elucidate two classes of MBNL1-mediated splicing event. Heat map depicting two-dimensional correlation clustering of splicing activities (red values in Figure 4) obtained for each double RIM (x-axis) with the six minigenes tested (y-axis) in Figure 4. Splicing activities for each minigene were mean centered prior to analysis. Yellow represents mutants with high splicing activity, black represents activity near the mean, and blue represents mutants with low splicing activity.

The other important observation gained from the clustering heat map was the observation that the differences in splicing profile that define the two classes seem to be due to mutation of ZF2. 1,2RIM, 2,3RIM, and 2,4RIM clustered together in the Class II exon events and in all cases have the least activity of all the RIMs for the four minigenes

in this class. Therefore, for Class II splicing events MBNL1 function is strongly dependent on ZF2. However in Class I events, 1,2RIM and 2,3RIM have high levels of activity while the lowest activity cluster contains 1,4RIM and 2,4RIM. This observation suggests that for Class I events, wild-type function of ZF4 might be important for splicing regulation.

RNA Binding Analysis of Double RIMs with RNA Substrates

To determine if the RNA binding of the double RIMs correlated with splicing activity, we expressed N'-terminal GST fusions of truncated versions (amino acids 2-260) of each double RIM and measured the binding affinity for RNA substrates derived from each minigene. Sequences of all RNA substrates tested are provided in Table 1. Binding to the 32- nt region of the TNNT2 minigene that lies just upstream of the 3'ss in intron 4 has been extensively characterized (55; cTNT 32mer), (41). Therefore, we chose to use that RNA to represent the TNNT2 binding site. The ATP2A1 minigene has several high affinity MBNL1 binding sites, consisting of clusters of YGCY motifs downstream of exon 22. Functional analysis has shown that all of these sites must be deleted to lose MBNL1 function; therefore we believe that most if not all of these sites are functional (41). We chose one of these previously characterized, high affinity sites to represent an ATP2A1 binding site. Similar to ATP2A1, the *MBNL1* minigene contains multiple high affinity MBNL1 binding sites, containing clusters of YGCY motifs, located upstream of exon 5. Deletion analysis of a 141-nt region of *MBNL1* intron 4 showed that no single motif within this region is necessary for MBNL1 activity (92). As with ATP2A1, it is believed that all of the MBNL1 binding sites within this intronic region may be functional. Taking these observations into consideration, we designed a 40-nt *MBNL1* RNA that contains multiple YGCY clusters for binding to the double RIMs. MBNL1 was shown to associate with a 37-nt region of INSR downstream of exon 11 via co-immunoprecipitation (90). Grammatikakis *et al.* demonstrated direct binding to a truncated (31-nt) version of this RNA (91). We chose to use the larger substrate for the binding studies conducted herein. Mutation of YGCY motifs in the Nfix and Vldlr minigenes followed by splicing characterization identified the required MBNL1 binding sites in both minigenes (40). We derived 40-nt RNA substrates from Vldlr and Nfix that

contain the domains required for MBNL1 function. In the case of Nfix we derived two substrates, the first (Nfix_Site #1) includes the YGCY motifs on the most 5' end of intron 7. The second RNA encompasses the YGCY motifs in exon 8 of the Nfix minigene (Nfix_Site #2).

Of all the RNAs tested, the double RIMs bound the *MBNL1* RNA with the highest affinity (Table 1). In general, the RNA binding trends observed for the *MBNL1* RNA substrate corresponded to splicing activity. Wild-type protein binds *MBNL1* with an apparent K_d of 5±1 nM, followed by 3,4RIM with a 3-fold decrease in affinity (17±5 nM). 1,4RIM, 1,2RIM, and 1,3RIM, which have intermediate splicing activities, have K_d values of 124±20 nM, 138±14 nM, and 204±66 nM, respectively. 2,3RIM and 2,4RIM, the mutants with the lowest splicing activity for the *MBNL1* minigene, have the lowest affinities for this substrate at 496±206 nM and greater than 4uM, respectively. Representative binding gels for the *MBNL1* substrate are shown in Figure 7A. For this RNA we were able to resolve multiple RNA-protein complexes that appear to vary between mutants (Figure 7A).

Table I. Dissociation constants for double RIMs.

		Apparent K_d (nM)						
RNA	**Sequence**	**WT**	**1,2RIM**	**1,3RIM**	**1,4RIM**	**2,3RIM**	**2,4RIM**	**3,4RIM**
MBNL1	5'-AUGA**UGC**ACCUC**UGCUUGCU**GUUUAUGUUAA**UGCGCU**UGA-3'	5±1	138±14	204±66	124±20	496±206	> 4000	17±5
[a,b]INSR (ST1)	5'-G**GCU**UUU**UGC**A**UGC**GGCGGGCA**GCU**G**UGCU**GGAGAGC-3' (↓↓↓ AGG)	18±5	352±30	1620±155	2136±840	2802±639	> 4000	150±11
[c]Nfix_Site #1 (int.)	5'-**GCUGCU**GGG**UGCU**GAGAAGAGAAGGAGCGG**GCUGCU**UAGC-3' (*)	10±4	352±69	1105±559	910±315	1131±334	> 4000	145±38
[c]Nfix_Site #2 (ex.)	5'- G**GCU**UUCACCCACCACC**CGCUGCC**UG**UGCUUGCU**GGAGUC-3' (**, *, *)	55±6	1378±347	1171±154	372±36	1114±689	> 4000	118±11
[d]ATP2A1	5'-**UGC**CAC**UGCUGCUGCUGCGCU**UCCAGUCAGGGUGG**GCC**GC-3'	12±2	333±52	1937±317	1362±176	1768±154	> 4000	157±19
[d,e]TNNT2	5'-CCUGUCU**CGCU**UUUCCCCUC**CGCUGC**G**GCC**AC-3'	15±3	> 4000	> 4000	> 4000	> 4000	> 4000	43±7
[c]Vldlr	5'-CUGUUAAUCUCAACUAAC**UGCCGCU**UAAAUAAUUAG**UGC**A-3' (* * *)	35±7	> 4000	> 4000	> 4000	> 4000	> 4000	73±5

BOLD: YGC or GCY motif
Underlined: YGCY motif

[a] Sen 2010; ↓ Mutations that disrupted co-immunoprecipitation with MBNL1
[b] Grammatikakis 2011; Sequence used to demonstrate direct binding to MBNL1
[c] Du 2010; * G to C mutation that disrupt MBNL1 regulation
[d] Goers 2010
[e] Warf 2007

More specifically, although all 3 mutants have similar binding affinities, 1,3RIM and 1,4RIM form multiple, distinct complexes with the *MBNL1* RNA, however 1,2RIM only forms a single, slowly migrating complex. Additionally, 2,3RIM forms different

complexes than 1,3RIM or 1,4RIM at similar concentrations (compare panels in Figure 7A). Also, 3,4RIM forms multiple complexes at high protein concentration, however wild-type protein does not.

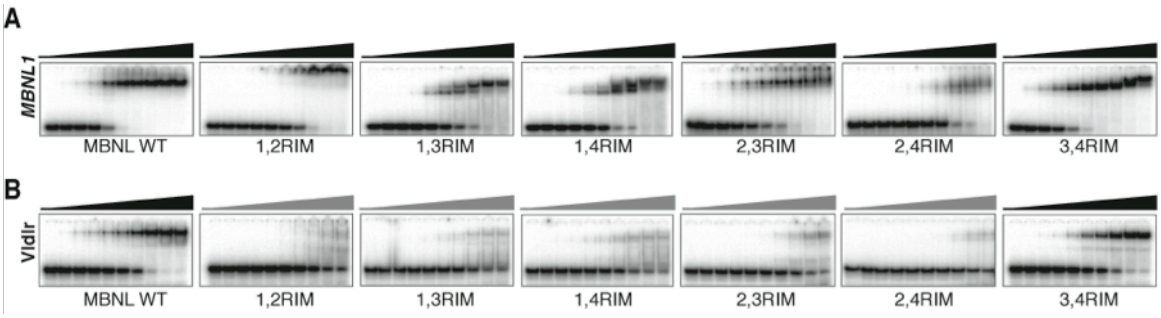


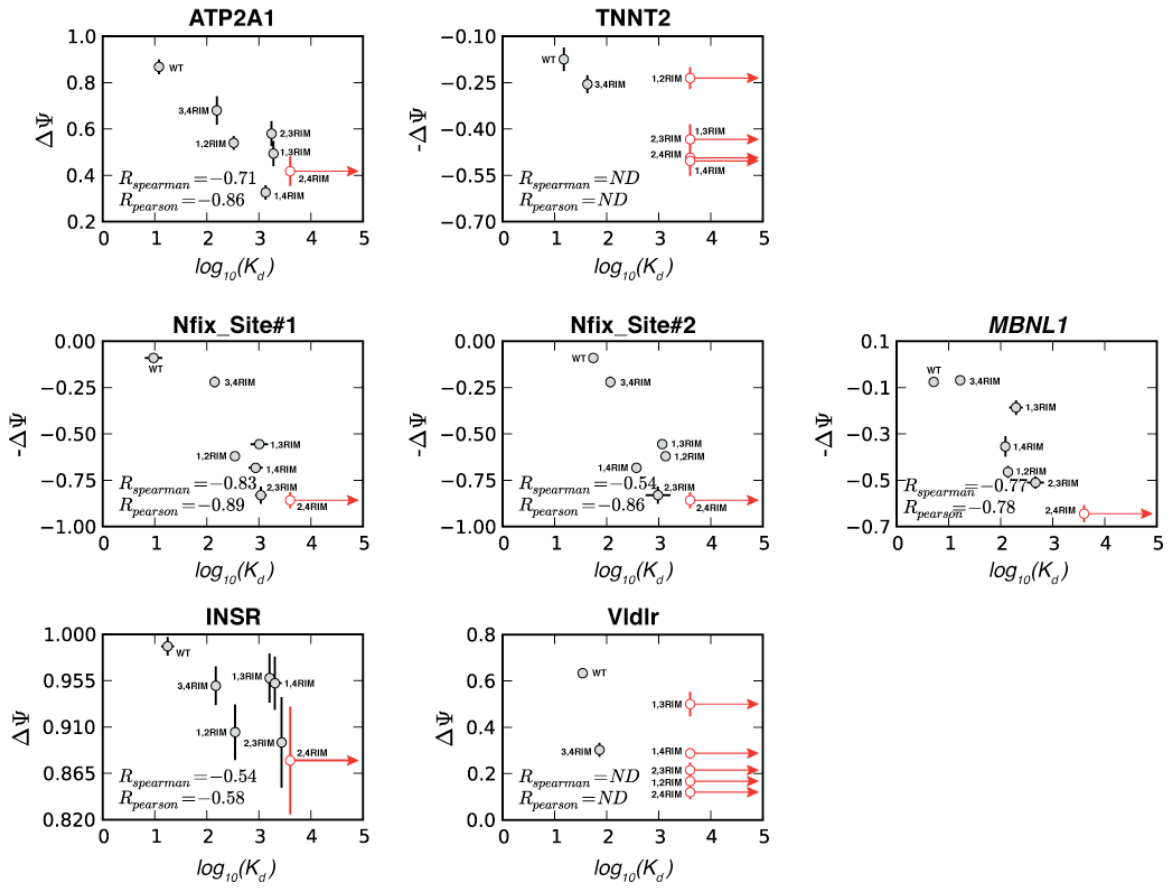
Figure 7. Double RIM binding to RNA substrates. (A) Representative EMSA gels of dRIMs binding to *MBNL1* RNA substrate in comparison to wild-type protein. (B) Representative EMSA gels depicting double RIMs lack of binding to Vldlr RNA substrate in comparison of wild-type protein. Sequences of RNA substrates can be found in Table 1. Triangles indicate increase in protein concentration. Grey triangles denote binding events where the percentage of RNA bound at highest concentration of protein was less than 20%. Protein concentrations used were: 0, 1, 3, 9, 27, 81, 243, 729, 2187, 4000 nM.

Pair-wise analysis between the *MBNL1*, Nfix and INSR minigenes yielded some of the highest correlation values obtained in this study (Figure 5). Additionally, clustering analysis placed all three of these events into Class II, because the splicing pattern obtained with the RIMs was very similar amongst these minigenes (Figure 6). Therefore, it was not surprising to find that the corresponding *MBNL1*, Nfix and INSR RNA substrates showed similar binding trends in regards to RIMs. Wild-type MBNL1 bound to INSR, Nfix_Site #1, and Nfix_Site#2 substrates with K_ds of 18±5 nM, 10±4 nM and 55±6 nM, respectively. Similar to *MBNL1*, 3,4RIM bound the most tightly of all mutants to these RNAs and with similar apparent K_d values of 150±11 nM, 145±38 nM, and 118±11 nM, respectively. With approximately 20-30-fold lower affinity that wild-type protein, 1,2RIM had identical binding affinities for INSR (352±30 nM) and Nfix_Site#1(352±69 nM), but much weaker binding to Nfix_Site#2 (1378±347 nM). 1,3RIM, 1,4RIM, and 2,3RIMs bound the INSR and Nfix subset of RNAs with K_d values ranging from 900 to 2800 nM (refer to Table 1 for exact values). A notable exception to

this generality is 1,4RIM binding to Nfix_Site#2 with a K_d of 372±36 nM, which is 4-fold lower than 1,2RIM and 2-fold greater than 3,4RIM. This observation is in stark contrast to the splicing data, which shows that 1,4RIM has the lowest, measurable splicing activity for Nfix. 2,4RIM had dissociation constants greater than 4uM for INSR and Nfix RNAs. Similar to *MBNL1*, although less distinct, we observed multiple complexes with the Nfix and INSR substrates that were mutant dependent.

Contrary to what the pair-wise correlation and clustering data would predict, binding trends for ATP2A1 amongst the RIMs was very similar to binding seen for Nfix and INSR. Wild-type MBNL1 bound to ATP2A1 with an apparent K_d of 12±2 nM. Binding affinities measured for the 3,4RIM (157±19 nM) and 1,2RIM (333±52 nM) binding to ATP2A1 were nearly identical to the values obtained for the INSR and Nfix substrates. 1,3RIM, 1,4RIM, and 2,3RIM also bound ATP2A1 with values between 1400-2000 nM, also similar to INSR and Nfix. Once again, 2,4RIM bound with a K_d greater than 4 uM. These binding trends are surprising because ATP2A1 has a very different splicing activity profile from Nfix, INSR, and *MBNL1*, and therefore clustered into a Class I. Correlation analysis was conducted to compare the splicing activities of the double RIMs to the binding activities measured for the RNA substrates derived from each minigene (Figure 8). This analysis confirmed that *MBNL1*, INSR, Nfix, and ATP2A1 all have significant correlations between splicing and RNA binding (Figure 8).

Figure 8 (next page)**. Statistical analysis of correlation between dRIM binding and splicing profile**. The dissociation constants (K_d) for each dRIM mutant obtained for each RNA substrate evaluated in Figure 5 (x-axis) is plotted against to the change in percent activity (Y) measured for each mutant in Figure 3 (y-axis). Spearman (r) and Pearson (R) correlations are provided. MBNL1-mediated exclusion events are represented as negative changes in splicing activity (-Y). Dissociation constants for both Nfix RNAs were correlated to the single Nfix splicing activities. Red arrows represent binding events with dissociation constants greater than 4000 nM.



The remaining RNA substrates, TNNT2 and Vldlr, had nearly identical binding activities. This was surprising because these minigenes behaved very differently from one another in the splicing assays and clustered into separate classes. In fact, with an R-value of 0.42, the correlation between TNNT2 and Vldlr was the lowest observed, suggesting that these minigenes had the least similar RIM splicing profiles. Wild-type MBNL1 displayed an apparent K_d of 15±3 nM for the TNNT2 RNA and 35±7 nM for Vldlr. 3,4RIM bound both substrates with a K_d only 2-3 fold lower than wild-type. However in stark contrast to the other 5 RNA substrates, all of the other RIMs bound to TNNT2 and Vldlr with apparent dissociation constants greater that 4 uM. This was a shocking observation in both cases. In the case of TNNT2, all of the double RIMs were able to regulate splicing of the TNNT2 minigene with at least 35% activity. In fact, 1,2RIM retained slightly more activity, in regards to TNNT2 splicing (86% vs. 82%), than 3,4RIM, but has no significant binding affinity for the 32-nt RNA. In the case of Vldlr

the splicing activity ranged from 0% to 73%, but the difference in splicing activity in no way correlated to binding to the RNA substrate. Representative Vldlr binding gels can be seen in Figure 7B. The observation that only a 2-fold difference in binding activity between wild-type MBNL1 and 3,4RIM corresponds to almost a 70% reduction in splicing activity is intriguing (compare WT and 3,4RIM in Figure 4D). Also noteworthy is the disparity in binding affinities between 1,2RIM and 3,4RIM for Vldlr, which seems to be much more drastic than the difference in splicing activity (8% vs. 33%, respectively) would suggest. Another observation unique to Vldlr is that 1,3RIM, which is the most active of all double RIMs for this minigene, has no significant binding although it regulates splicing with 40% more activity than 3,4RIM (Figure 4D). The lack of correlation between splicing and binding activities for TNNT2 and Vldlr are echoed in the correlation analysis, wherein both substrates show very poor correlations similar to ATP2A1 (Figure 8).

As a whole these binding experiments provided us with several clear observations. The first of which is that 2,4RIM undeniably has the lowest affinity binding for every substrate. The 2,3RIM also has severely reduced RNA binding ability. These are very meaningful observations considering that the clustering analysis revealed that these mutants were clearly the least active. Therefore there is a strong correlation between activity and binding for the 2,3RIM and 2,4RIM and it is clear that ZFs 2 and/or 4 are required for high affinity binding. The other strong conclusion is that in every case 1,2RIM, with a wild-type ZF3-4 pair, binds all RNAs much weaker than 3,4RIM, with its wild-type ZF1-2 pair. This binding information indicates that the ZF1-2 pair has inherently higher RNA binding affinity than the ZF3-4 pair, providing a satisfying explanation for the discrepancy in splicing activity between the two ZF pairs for some substrates. What the binding data does not provide is a clear explanation of how mutants with essentially no affinity for RNA substrates (such as TNNT2) can still regulate splicing of these minigenes. The formal possibility exists that the RNA substrates that we used are not appropriate representations of the actual endogenous binding sites. With that being said, it seems apparent that in some instances one must invoke that MBNL1's splicing activity is due to more than just direct MBNL1:RNA interactions.

High Affinity MBNL1 Binding Is Required to Regulate a Synthetic Minigene

The insights gained evaluating the splicing and binding activity of the double RIMs on endogenous MBNL1 targets, led us to believe that evaluation of a target that has not evolved to be regulated by MBNL1 could be informative. The analysis of such a reporter could remove some of the complications that are inherent to evaluating endogenous targets; such as the existence of multiple MBNL1 binding sites with varying affinity in the introns and exons of regulated targets and recruitment of MBNL1 via protein-protein interactions through factors bound at adjacent sites in the pre-mRNA.

We previously created a synthetic splicing reporter from the PLEKHH2 minigene (41) by inserting 6, tandem 5'-CUGCUU-3' motifs upstream of exon 21, within intron 20, in a PLEKHH2 minigene to create a high affinity MBNL1 binding site. The resulting synthetic reporter, PLEKHH2-SM1F (Int 20), is depicted in schematic form in Figure 9A and is robustly regulated by MBNL1. Using the same high affinity MBNL1 binding site, we also created a reporter wherein the motifs were inserted downstream of exon 21, within intron 21. The resulting reporter, PLEKHH2-SM1F (Int 21) (see schematic in Figure 9B), is also regulated by MBNL1. The basal exon inclusion levels for the PLEKHH2-SM1F (Int 20) and PLEKHH2-SM1F (Int 21) reporters are similar at 90±5% and 99±2%, respectively. When MBNL1 is co-expressed with either reporter, a decrease in inclusion level to 41±6% or 35±5%, respectively, is observed.

Splicing of the double RIMs was evaluated with both the PLEKHH2-SM1F (Int 20) (Figure 9C) and PLEKHH2-SM1F (Int 21) (Figure 9D) reporters. Interestingly, 1,2 RIM and 3,4RIM were the only RIMs that were capable of regulating splicing of the PLEKHH2-SM1F (Int 20) synthetic reporter (Figure 9C). 3,4RIM retains 76% of wild-type activity, which is well within the range of activities observed for this mutant in regards to the splicing of the endogenous minigenes. 1,2RIM only has 12% of wild-type activity for this synthetic reporter. Although this value seems low at face value, it is important to remember that values of 8% and 25% were observed for 1,2RIM splicing of Vldlr (Figure 4D) and Nfix (Figure 4H), respectively. Hence, 12% activity for this synthetic reporter is not dramatically different from values that we observed for the endogenous minigenes with this mutant. Interestingly however, only 3,4RIM was able to regulate splicing of the PLEKHH2-SM1F (Int 21) reporter and does so with 94% wild-

type activity (Figure 9D).

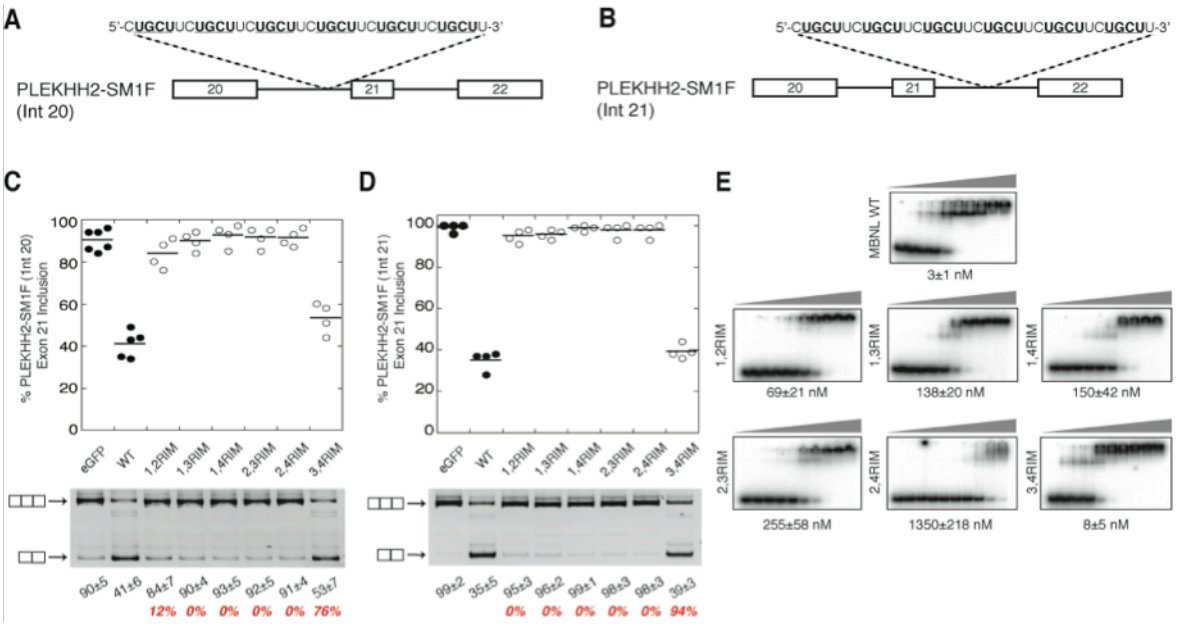


Figure 9. Evaluation of splicing and binding activity of double RIMs with synthetic splicing reporters. (A) and (B) Schematic representations of the synthetic reporters, PLEKHH2-SM1F(Int 20) and PLEKHH2-SM1F(Int 21), respectively. Open boxes represent exons and lines represent introns. Reporters were created by inserting a 36-nt motif, containing six YGCY motifs (bold and underlined), upstream of Exon 21 into the PLEKHH2 minigene. The sequence of the motif (SM1F) is provided above the schematic in Panels A and B. (C) Dot plot of PLEKHH2-SM1F (Int 20) splicing with double RIMs. (D) Dot plot of PLEKHH2-SM1F (Int 21) splicing with double RIMs. Values for percent exon inclusion and percent splicing activity (red) are listed below representative splicing gels in Panels C and D. (E) Representative EMSA gels showing double RIM binding to the SM1F motif in comparison to wild-type MBNL1. Apparent binding constants (K_d) for each mutant are listed below the corresponding gel. Triangles indicate increase in protein concentration. Protein concentrations used were: 0, 1, 3, 9, 27, 81, 243, 729, 2187, 4000 nM.

Binding affinity was measured for the double RIMs with the 36-nt RNA, SM1F, used to create the reporter (Figure 9E). The sequence of SM1F is provided in the schematics in Figures 9A and 9B. The SM1F RNA, with an apparent K_d of 3±1 nM, has the highest affinity for wild-type MBNL1 of all the RNA substrates used in this study. Echoing the previous binding observations, 3,4RIM binds SM1F with an approximate 2-fold decrease in affinity (8±5 nM), in comparison to wild-type MBNL1, whereas 1,2RIM

binding is 23-fold weaker (69±21 nM). One surprise in these results resides in the fact that 3,4RIM binds SM1F with only a slight decrease in affinity from wild-type protein, and regulates splicing of the PLEKHH2-SM1F (Int 21) reporter with near wild-type levels, however 3,4RIM splicing suffers a 24% decrease when this motif is placed upstream of the exon. Another noteworthy observation is that 1,2RIM, which has significantly decreased affinity for the SM1F substrate, retains modest activity when placed upstream of the exon (*i.e.* PLEKHH2-SM1F (Int 20)), but is inactive when placed downstream of the exon (*i.e.* PLEKHH2-SM1F (Int 21)) (compare Figure 9C to 9D).

Close examination of the gel shifts shows that although wild-type MBNL1 and 3,4RIM have similar affinities for the SM1F substrate, the two proteins form distinctly different complexes with the RNA at identical protein concentrations (Figure 9E). In fact, all of the mutant proteins, form complexes that correspond in shift to the wild-type MBNL1:SM1F complex at low protein concentration, but then form slower migrating complexes at higher concentrations. 1,3RIM, 1,4RIM and 2,3RIM all have similar affinity for the SM1F substrate with K_d values of 138±20 nM, 150±42 nM, and 255±58 nM, respectively. Although weak, SM1F was the only substrate with which we were able to measure 2,4RIM binding (1350±218 nM), indicating that this mutant is able to weakly bind very high affinity RNA substrates.

DISCUSSION

The ZF1-2 Pair of MBNL1 Has Higher Levels of Activity than the ZF3-4 Pair

By individually mutating the ability of each ZF to either chelate zinc or to interact with RNA, we have shown that MBNL1 does not require 4 functional ZFs to regulate splicing (Figures 1 and 3). By conducting pair-wise ZF mutations, we have clearly demonstrated that the ZF1-2 and ZF3-4 pairs of MBNL1 do not have equivalent function (Figures 2 and 4) nor do they have equivalent affinities for RNA (Figure 7 and Table 1). As previously reported, in some MBNL1-regulated splicing events, such as TNNT2, the pairs of ZFs are redundant and MBNL1 can function with either pair (91). For these events it appears as though MBNL1 does not require an intact ZF-ZF domain as mutations that disrupt one ZF in each pair (*i.e.* 1,3RIM, 1,4RIM, 2,3RIM, and 2,4RIM) retain moderate levels of splicing activity. However in a second subset of MBNL1-

regulated events, such as Nfix, MBNL1 requires a functional ZF1-2 pair to regulate splicing and cannot robustly regulate splicing with the ZF3-4 pair alone. We were only able to definitively demonstrate the difference in splicing activity between the two ZF pairs because our analysis was extended to contain multiple regulatory events. The difference in splicing activity between the ZF1-2 and ZF3-4 pairs could largely be due to their inherent differences in RNA binding affinity. For all of the RNA substrates we tested, the ZF1-2 pair had a significantly higher binding affinity than the ZF3-4 pair (Table 1). The disparity in binding affinity that we observed between ZF1-2 and ZF3-4 domains echoes recent binding studies conducted with structured RNAs (93).

RNA Interaction Mutants Elucidate at Least Two Distinct Classes of MBNL1-Regulated Splicing Event

Using a combinatorial approach to study the function of the ZFs we created six, MBNL1 double RIMs wherein two of the four ZFs were mutated to have reduced ability to interact with RNA. We tested the splicing activity of these RIMs on six MBNL1-regulated transcripts and observed splicing activities ranging from 0% to 100% (Figure 4). Clustering of the RIM splicing activity profiles revealed that two distinct classes of splicing events exist within the subset of splicing events we tested, which we've denoted Class I and Class II (Figure 6).

An interesting outcome from the clustering is that splicing events did not cluster based on similarities in the mode of MBNL1 regulation (*i.e.* activation or repression) or by similarities in MBNL1-binding site architecture. In other words, examples of MBNL1-mediated inclusion and exclusion events can be found in both classes and examples of binding sites with only a few YGCY motifs or extensive clusters of YGCY motifs can also be found in both classes. The Class I events, which include ATP2A1 and TNNT2, showed high activity for 1,2RIM and 3,4RIMs and decreased but similar activities for the remaining mutants. The Class II minigenes, which contain INSR, *MBNL1*, Nfix and Vldlr, shared similar splicing profiles due to a gradient in splicing activity amongst the six RIMs. Events in Class II all share characteristically low activity for 1,2RIM, 2,3RIM, and 2,4 RIM. We also measured RNA binding affinities for each RIM with RNA substrates derived from each minigene (Table 1 and Figure 7). We found

that 2,4RIM had the lowest affinity for RNA of all mutants, and that 2,3RIM only weakly binds RNA. Taken together these finding suggests that a functional ZF2, or a moderately high affinity interaction with the ZF1-2 pair, is required for splicing activity of transcripts found within Class II.

MBNL1 May Require Higher Order Intramolecular Interactions for Splicing Function

Splicing of the synthetic MBNL1 reporters, containing six YGCY motifs artificially inserted into the PLEKHH2 minigene, allowed us to measure the splicing activity of MBNL1 without the presence of additional splicing signals that may exist in endogenous transcripts (Figure 9). Our initial observation that placing the motifs either upstream or downstream of PLEK exon 21 elicited the same effect, exon skipping, is intriguing because it challenges the simple theory that position of the MBNL1 binding motifs with regards to an exon is sufficient to elicit the positional regulatory effects characteristic of MBNL1. We found that only two mutants, 1,2RIM and 3,4RIM, were able to regulate splicing of either synthetic reporter, resulting in a splicing profile unique to these minigenes (compare Figure 4C-4H to Figure 9C and 9D).

We assessed binding of each RIM to the MBNL1 binding motif (*i.e.* SM1F) that we inserted to create the reporter and found that all of the RIMs were able to bind the motif with relatively high affinity compared to endogenous RNA substrates (compare Figure 9E to Table 1). Interestingly however, we observed that very little of the complexes formed in the RIM binding gels correspond to the wild-type MBNL1: RNA complex, but rather form more slowly migrating complexes. Hence the RIM proteins can bind RNA with somewhat high affinity, but may not bind in the same confirmation as wild-type protein. The poor correlation between splicing and binding activity for the synthetic reporters puts them in a class of their own. We believe that, unlike the Class I and Class II events, the synthetic reporters require confirmation dependent, high affinity MBNL1 interactions for function. This requirement could partially be explained by the lack of additional protein-protein interactions in the reporters that we previously hypothesized to be important for recruiting the splicing factor. Although we don't completely understand the mechanisms by which MBNL1 is able to regulate splicing of

these synthetic reporters, we think mechanistic studies of these reporters are worthy of future investigation.

Binding experiments conducted with RIMs also resulted in multiple protein: RNA complexes for some endogenous substrates. The existence of multiple complexes appeared to be specific to each mutant and concentration dependent. Specific examples of this phenomenon can be seen by comparing the 1,2RIM binding gels to the 3,4RIM binding gels for all of the endogenous substrates. In every case, 1,2RIM forms complexes that are more slowly migrating than those formed by wild-type MBNL1, whereas 3,4RIM mostly forms complexes that correspond to those of wild-type (Figure 7). Additionally, we find it worth noting that 1,3RIM and 1,4RIMs, both of which retain a wild-type ZF2, mostly form complexes that correspond to those formed with wild-type protein. Taken together the trends in complex formation and binding affinity indicate that ZF2 is important for facilitating high affinity protein: RNA interactions of the appropriate confirmation, which help to explain its importance in MBNL1 function. We also observed that the RIMs eluted as multiple peaks from a Heparin column, but that the peaks had similar affinities for RNA (see Materials and Methods). We hypothesize that these observations could be due to the presence of higher order MBNL1 structures that contribute to RNA binding which are perturbed in the RIMs. Although beyond the scope of this study, investigating the functional contribution of MBNL1 confirmations or potential ZF intramolecular interactions may warrant further investigation.

Definition of Functional Classes Provides Novel Mechanistic Insights into How MBNL1 Regulates Alternative Splicing

With the exception of Vldlr, the splicing activities for Class II events correlated very well with binding affinities (Figure 8), suggesting that Class II-type splicing events are primarily dependent on MBNL1 binding to RNA. A very poor correlation between RIM binding and splicing activity for the Vldlr substrate, as well as both Class I events, was observed. However the splicing activity of Vldlr differs significantly from the Class I splicing activities observed with TNNT2 and ATP2A1. For both the Vldlr and TNNT2 substrate we were only able to measure binding to one mutant, 3,4RIM, however nearly all of the mutants were able to regulate splicing of both minigenes to some extent.

Although the lack of correlation between binding and splicing activity for Vldlr suggests that it might be a unique Class II splicing event (Figure 8), we are still grouping Vldlr with the other three Class II events due to the strong correlation in splicing activity that we observed in Figure 6.

The lack of correlation between splicing and binding for both Class I events suggests that MBNL1 functions independently of a strong RNA interaction within this class. An intriguing explanation for this observation is that splicing events within Class I are partially mediated by protein: protein interactions that are important for recruiting MBNL1 to the transcript. Regions outside of the ZFs that may be important for mediating such interactions have been recently defined (91). If this is the case, then it is worth noting that no dominant negative effects were observed with the most deleterious mutant, 2,4RIM. The lack of dominant negative effects suggests that MBNL1 is not interacting with a core component of the splicing machinery, but rather with other RNA binding proteins with binding sites that co-occur in MBNL1-regulated introns, such as Fox1/2, CUGBP1, or hnRNP factors (94), (95) that regulate splicing. The proposed mechanistic differences between the two classes have been summarized in a model in Figure 10.

Reinterpretation of Current Models for How MBNL1 Binds to RNA

Models have been proposed to explain how MBNL1 binds to RNAs containing multiple YGCY motifs, based on either the crystal structure of MBNL1 alone (61), or in combination with biochemical studies which assessed MBNL1 binding to simple RNA model substrates (96). Both of these models make the assumption that the four ZF domains of MBNL1 have equal affinity, and therefore that the ZF1-2 and ZF3-4 pairs also have equivalent affinities for RNA. Therefore both models predict that conformational distortions of the protein or RNA substrate can result in binding confirmations wherein ZFs in both pairs are simultaneously contacting the RNA, allowing MBNL1 to bind YGCY motifs that are very close together or spaced further apart in target transcripts based on which combination of ZFs bind. For example the model put forth by Cass *et al.* proposes that an MBNL1 mutant containing only wild-type ZF1 and ZF3 would bind RNA substrates with closely spaced YGCY motifs as well as, or better than, a mutant containing only the wild-type ZF1-2 pair (96). The first mutant

would be able to wrap around the RNA, due to inherent flexibility related to the long linker between the pairs of ZFs, and access both motifs. However, the latter mutant would have reduced ability to bind the same substrate because it is limited by the proximity of the binding sites within the RNA, making the RNA unable to wrap around the ZF1-2 pair and therefore unable to simultaneously interact with the closely spaced motifs.

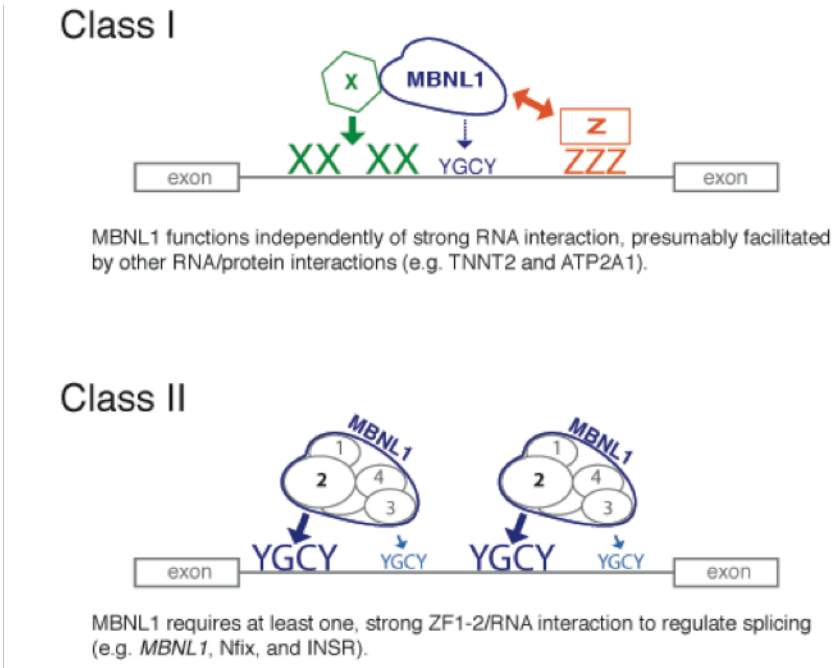


Figure 10. Model depicting differences between MBNL1's role in Class I and Class II splicing events. For Class I events (top panel) there is no correlation between MBNL1 binding and splicing activity, therefore other RNA:protein (depicted with protein Z) or MBNL1:protein (depicted with protein X) interactions may mediate MBNL1's ability to bind RNA and regulate splicing. For Class II events (bottom panel) the correlation between binding affinity and splicing activity suggest that MBNL1's activity is largely due to direct RNA recognition. However, in Class II events, ZF2 of MBNL1 seems to be important for high affinity RNA interactions.

Although we have presented evidence to support a model wherein MBNL1 utilizes variable protein: RNA confirmations for splicing activity, the binding and splicing data presented herein are directly inconsistent with this simple model. We saw that such a mutant (*i.e.* 2,4RIM) was completely incapable of regulating splicing or

binding to endogenous RNA substrates, regardless of the spacing of YGCY motifs. Whereas the mutant containing only the ZF1-2 pair (*i.e.* 3,4RIM) had the highest binding affinity of all the mutants for every substrate tested and retained high levels of splicing activity. We have good evidence that MBNL1: RNA interactions are somehow facilitated by a functional ZF2 or 4 because losing both strongly inhibits MBNL1's ability to bind RNA and consequently regulate splicing. We propose that MBNL1 binding to endogenous transcripts is most likely transcript specific, depends upon the presence of a ZF2 interaction, and in some cases might be facilitated by other RNA binding proteins. Hence, we believe that the mode in which MBNL1 binds to endogenous transcripts cannot be explained as simply as the previous models, which were derived from small, synthetic RNAs, might suggest.

In summary, we have conducted the first comprehensive study, specifically designed to dissect the functional contribution of each ZF to MBNL1's overall splicing activity. We have learned that although no single ZF is absolutely required for function that either ZF2 or ZF4 seems to be necessary for high affinity interactions with RNA. We have also concisely demonstrated that the two pairs of ZFs within MBNL1 do not have equivalent RNA binding properties and consequently do not have redundant function. Our studies lead us to identify multiple classes of MBNL1-regulated splicing events. There is still much to learn about how MBNL1 uses binding site composition and architecture within endogenous targets to facilitate multiple modes of splicing regulation. Our hope is that future studies can use the information provided herein to further classify other MBNL1 binding events as well as to discern what additional proteins may be involved in the mechanisms governing MBNL1-mediated splicing.

MATERIALS AND METHODS

Protein Mutagenesis Cloning and Splicing Reporters. The construction MBNL1-eGFP plasmid was previously described (84) and obtained from the laboratory of Maury Swanson. The wild-type MBNL1-eGFP protein (1-382; splice isoform a; NCBI accession number NP_066368) was used as the wild-type control in all splicing experiments and as the cloning template for all mutants. Mutations were installed using PCR via standard mutagenic cloning techniques. Primers used to install the point

mutations are provided in Table 2. All mutants were cloned into the pEGFP-C1 (Clontech) plasmid using XhoI and BamHI sites, resulting in mammalian expression of all mutant proteins as N'-terminal eGFP fusions.

Table 2. Mutagenic Cloning Primers

Mutant	Mutation	Sense Primer Sequence
1CM	C34A/H38A	5'-CGGCCAGACACGGAAGCTAAATTTGCAGCTCCTTCGAAAA GCTGC-3'
2CM	C66A/H70A	5'-GCTCCAGGGAGAACGCCAAATATCTTGCTCCACCCCCACAT TTAAAAACGC-3'
3CM	C200A/H204A	5'-CGAGGAGAAAATGATGCTCGGTTTGCTGCTCCTGCTGACAG CAC-3'
4CM	C234A/H238A	5'-GCTCTCGGGAAAAGGCCAAATACTTTGCTCCCCCTGCACATT TAC-3'
1RIM	E21A	5'-CTGGAAGTATGTAGAGCGTTCCAGAGGGGG-3'
1RIM	F36A	5'-CACGGAATGTAAAGCTGCACATCCTTC-3'
2RIM	D55A	5'-GTAATCGCCTGCTTTGCTTCATTGAAAGGCCG-3'
2RIM	Y68A	5'-GAACTGCAAAGCTCTTCATCCACCC-3'
3RIM	E187A	5'-CTTGAGGTATGTCGAGCGTACCAACGTGGCAATTGC-3'
3RIM	F202A	5'-GAAAATGATTGTCGGGCTGCTCATCCTG-3'
4RIM	D223A	5'-GTCACTGTGTGTATGGCTTACATCAAAGGGAG-3'
4RIM	Y236A	5'-GGAAAAGTGCAAAGCCTTTCATCCCC-3'
pEGFP-C1 to pGEX-6P1		Sense: 5'- CGCGGATCCATGGCTGTTAGTGTCACACCAATT -3'
pEGFP-C1 to pGEX-6P1		Anti-Sense: 5'- GCGGCCGCTGCAGCTGCAGCCTGGTTGACCT -3'

For bacterial expression, the double RIMs were sub-cloned from pEGFP-C1 plasmids into pGEX-6P1 (Amersham) via PCR and subsequent digest and ligation with BamHI and NotI restriction enzyme sites. A previous study demonstrated that the C'-terminal region (amino acids 261-382) of MBNL1 is not required for high affinity binding (55), therefore all double RIMs were expressed as truncations at amino acid 260 for binding experiments.

Construction of TNNT2 (also known as hcTNT) (50), (97), PLEKHH2-SM1F (41), ATP2A1 (41), *MBNL1* (92), INSR (98), Nfix (40), and Vldlr (40) minigenes were previously described. The cTNT minigene was a gift from the laboratory of Thomas Cooper. Vldr and Nfix minigenes were a gift from Manuel Ares Jr. The INSR minigene was a gift from Nicholas Webster.

Bacterial Protein Expression. All proteins used for binding experiments were expressed as N'-terminal GST fusions. Protein expression of all double RIMs (2-260) and

wild-type MBNL1 (2-260) was conducted using ArcticExpress cells (Agilent) due to the limited solubility of double RIMs. Cells were induced following manufacturer's protocol for 24 hr at 10°C with 1 mM IPTG. Following induction the cell pellets were lysed with B-Per Reagent (Pierce) supplemented with DNAseI (10 U/mL) and lysozyme (0.5 mg/mL). The resulting lysate was diluted with 1 volume of 1XPBS and incubated on ice for 30 minutes prior to centrifugation at 17,000 RPM.

The resulting supernatant was filtered and loaded onto a GST affinity column (GSTrap FF; GE Healthcare). The column was washed with 5 column volumes of 1M NaCl followed by elution of the GST-tagged proteins with reduced glutathione (40 mM Bicine pH 8.3, 50 mM NaCl, and 10 mM reduced glutathione). The resulting eluant was diluted with 1 volume of water, to reduce the salt concentration to 25 mM, and subjected to cation exchange purification using a HiTrap Heparin HP column (GE Healthcare). Each double RIM protein eluted off of the heparin column as multiple peaks (Figure 11).

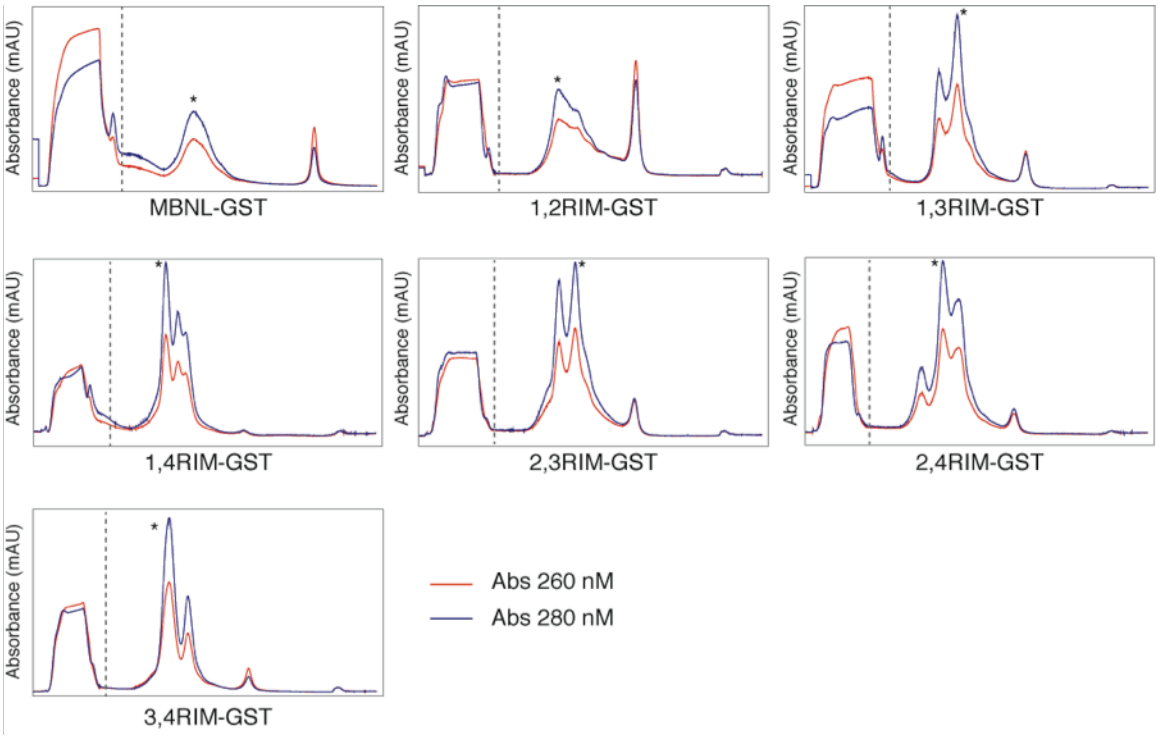


Figure 11. Double RIMs elute as multiple peaks off of heparin column.
Absorbance traces for elution of MBNL1-GST and dRIM-GST proteins off of heparin column. Dotted grey lines indicate the beginning of fraction collection. The peak that was found to have the highest RNA-binding activity via EMSA and was used for measurement of binding constants is denoted with an asterisk.

Fractions corresponding to the individual peaks for each protein were separately pooled, concentrated, and dialyzed into storage buffer (25 mM Tris pH 7.5, 500 mM NaCl, 5 mM β-ME, 50% glycerol) for storage at -80°C. For each mutant protein evaluated in this study we independently measured the RNA binding activity of each peak via EMSA and found that in every case the largest peak was the most active peak. Consequently, the protein corresponding to the major peaks (see asterisks in Figure 11) for each mutant were used in binding assays.

Cell Culture and Transfection. HeLa cells were routinely cultured as a monolayer in DMEM+GLUTAMAX media (Invitrogen) supplemented with 10% Fetal Bovine Serum (Gibco) at 37°C under 5% CO_2. Prior to transfection, cells were plated in 6-well plates at a density of $1.8x10^5$ cells/well. Cells were transfected 18-24 hours later at approximately 80% confluency. Plasmid (1 ug/well) was transfected into each well with 5 uL Lipofectamine2000 (Invitrogen) following manufacturer's protocol. For all splicing assays 500 ng of each mutant protein plasmid, wt plasmid, or empty pEGFP-C1 plasmid was co-transfected into a single well with 500 ng of minigene reporter. Cells were harvested with TrypLE (Invitrogen) 18-24 hrs after transfection and pelleted via centrifugation.

Cell-Based In Vivo Splicing Assay. RNA was isolated from HeLa cell pellets using an RNeasy kit (QIAGEN). Isolated RNA (500 ng) was incubated with 1 unit of RQI DNase (Promega) in a 10 uL reaction for 1 hour at 37°C. After DNase treatment, RNA (2 uL; 100 ng) was reverse transcribed in a 10 uL reaction using Superscript II (Invitrogen), according to manufacturer's protocols, with the exception that we used half the recommended amount of Superscript II.

TNNT2, PLEK-SM1, ATP2A1 and *MBNL1* minigene reporters were reverse transcribed using an antisense primer designed to the Sp6 promoter site of the pcDNA3 plasmid. The Nfix and Vldlr minigenes were reverse transcribed with a previously described DUP33 antisense primer, DUP1 (99), which anneals to the downstream DUP exon 3 (globin exon 2). The INSR minigene was reverse transcribed with a previously described pSG5 plasmid specific antisense primer (98). All reverse transcription reactions (2 uL) were subjected to PCR amplification in a 20 uL reaction (1:10 dilution) using flanking exon specific primers. The number of amplification cycles was determined

to be within the linear range for all primers used. Primer sequences used for RT-PCR amplification are provided in Table 3. The resulting PCR products were resolved by gel electrophoresis on 1.5 mm 6% (19:1) native polyacrylamide gels, ran at 300 V for 90 min. Splice products were visualized and quantified using SYBR Green I nucleic acid stain (Invitrogen) in combination with AlphaImager HP system (Alpha Innotech). The percent of exon inclusion was calculated as the fraction of signal corresponding to inclusion splice product divided by the total splice product signal (calculated by adding the volume of the splice product band to that of the exclusion product band).

Table 3. RT-PCR Primers

Minigene	RT Primer	PCR Primer (Sense)	PCR Primer (Anti-Sense)	#	ET* (°C)
TNNT2	5'-AGCATTTAGG TGACACTATAGA ATAGGG-3'	5'-GTTCACAACCATCT AAAGCAAGATG-3'	5'-GTTGCATGGCTGGTG CAGG-3'	24	55
MBNL1	5'-AGCATTTAGGT GACACTATAGAA TAGGG-3'	5'-GATCAAGGCTGCCC AATACCAG-3'	5'-CAGATTCATTTATTAA GAAACCCCACCCCTTAC-3'	26	55
ATP2A1	5'-AGCATTTAGGT GACACTATAGAA TAGGG-3'	5'-GTCCTCAAGATCTC ACTGCCAGT-3'	5'-GCCACAGCTCTGCCT GAAGATG-3'	20	55
INSR	5'-GCTGCAATAAA CAAGTTCTGC-3'	5'-CGAATTCGAATGCT GCTCCTGTCCAAAGA CAG-3'	5'-TCGTGGGCACGCTGG TCGAG-3'	22	55
Nfix/Vldlr	5'-GCAGCTCACTC AGTGTGGCA-3'	5'-GACACCATGCATGG TGCACC-3'	5'-GCAGCTCACTCAGTG TGGCA-3'	24	58
PLEKHH2-SM1F	5'-AGCATTTAGGT GACACTATAGAA TAGGG-3'	5'-GCTGCAGAATGAAA TTTGCTGTCAG-3'	5'-GTCACCATTTTGTTG CGTTCTTTCTAC-3'	23	58

*PCR elongation temperature
Rounds of PCR used for amplification

Western Blot Analysis. HeLa cells were plated, transfected, and harvested for Western blot analysis in the same fashion as for splicing assays. Cell pellets were lysed in RIPA buffer (50 mM Tris pH 7.4, 150 mM NaCl, 1% NP-40, 0.25% sodium deoxycholate, 0.1% SDS, 1 mM PMSF) supplemented with 1X protease inhibitor cocktail (SIGMA*FAST*; Sigma) via 3 freeze-thaw cycles. Concentration of protein was quantified and normalized using BCA reagent (Peirce) and 5-10 mg of total protein was loaded on 12% SDS-PAGE gels. Resolution of proteins and transfer onto nitrocellulose was conducted using standard techniques. MBNL mutant proteins were probed with 1:1,000 eGFP (FL) primary antibody (Santa Cruz; sc-8334) followed by 1:2,000 goat α-

rabbit secondary IgG-HRP (Santa Cruz; sc-2004). The actin loading control was probed using actin (I-19) (Santa Cruz; sc-1616) primary (1:1,000) antibody and the same secondary as eGFP. Chemilluminescence was visualized using a SuperSignal (Pierce) kit.

Radiolabeling and Purification of RNA Substrates. All of the RNA substrates used for gel mobility shift assays were commercially synthesized by Integrated DNA Technologies. The oligonucleotides were 5' end labeled with [γ-P32] ATP, followed by gel purification on 8% denaturing polyacrylamide gel (19:1) and elution into 0.3 M sodium acetate. RNAs were then ethanol precipitated with glycogen and resuspended into TE buffer (10 mM Tris pH 7.5, 1 mM EDTA).

Gel Mobility Shift Assays (EMSA). RNA was snap annealed at 95°C for 1 min and then placed on ice for 2 minutes. Once cooled the RNA was mixed with protein to yield final reaction conditions of 100 mM NaCl, 5 mM $MgCl_2$, 20 mM Tris pH 7.5, 1 mM β-ME, 5 mM DTT, 0.02% Triton X-100, 10% glycerol, 2 mg/mL BSA, 0.1 mg/mL heparin and 0.02% bromophenol blue. After mixing, the protein and RNA binding reaction was subjected room temperature incubation for 25 minutes prior to electrophoresis. Binding samples (3 uL) were loaded onto pre-chilled 1.5mm 6% acrylamide (37.5:1) gels. Gels were run for 45 min at 170 V at 4°C. Gels were dried for overnight exposure on phosphorus plates. Binding curves were quantified using ImageQuant (Molecular Dynamics). The fraction of RNA bound was calculated as the ratio of all RNA/protein complexes over total RNA per lane. The apparent K_d values were calculated using a previously defined equation (96) based on a minimum of at least three, independent gel shift assays per RNA.

Statistical Analysis, Correlation Plots and Clustering Analysis. For each mutant protein evaluated, splicing activity was calculated as a percentage of wild-type activity based off of the percentage of exon inclusion measured. Percent activity was defined as the change in exon inclusion caused by expression of the mutant relative to the mock splicing levels (*i.e.* eGFP) divided by the maximum change in exon inclusion, resultant from MBNL1-eGFP expression relative to mock. Conversion of the exon inclusion values to percent activity was the only means of normalization applied to the data set prior to clustering analysis. The dataset was mean centered only by minigene (row) prior to clustering. Hierarchical clustering (correlation (uncentered), centroid linkage) was

performed on both minigenes (rows) and dRIMs (columns) using Cluster 3 (100). The resulting Cluster tree and heatmap were visualized in JavaTreeView Version 1.1.6 (101).

CHAPTER IV
UTILIZING MS2 AND CHIMERIC MINIGENES TO INVESTIGATE MBNL1-MEDIATED SPLICING REGULATION

INTRODUCTION

Aberrant splicing of a number of MBNL1 regulated transcripts has been linked to the multifaceted disease symptoms of Myotonic Dystrophy, DM1 (51). One of the disease hallmarks in DM is the formation of ribonuclear foci, which contain the pathogenic CUG repeats and MBNL1 (48), (87). In DM1 the normal mRNA splice variants are produced, however the developmental splicing pattern is disrupted, resulting in temporally inappropriate expression of fetal protein isoforms that are not usually present in adult tissues (102). MBNL1-specific mis-splicing of the insulin receptor (INSR) (65), chloride channel (CLCN1) (63), ryanodine receptor 1 (RYR1) and sarcoplasmic/endoplasmic reticulum Ca^{2+}-ATPase 1(SERCA1 or ATP2A1) (70) genes results in insulin resistance, myotonia, and impaired calcium homeostasis, respectively, all of which are characteristic DM1 symptoms. Although these mis-splicing examples cleanly support the disease model, there are some well-characterized disease symptoms that have not been tied to specific mis-splicing events. Nor does every known mis-splicing event correlate precisely with a specific disease symptom.

Investigations into the mis-localization of MBNL1 to ribonuclear foci in DM1 cells concluded that foci formation and disruption of MBNL1-regulated splicing by CUG repeats are separable events (84). Analysis of DM1 cells identified other RNA binding factors, besides MBNL, that colocalize with CUG repeats in the nuclear foci (103). However, siRNA mediated knockdown of MBNL1 expression in DM1 cells causes disruption of the nuclear foci (104), implying that MBNL1 must be present for the RNA:protein foci in the nucleus to be stable. Therefore, this evidence suggests that MBNL1 is capable of mediating the formation of RNA:protein complexes. These findings challenge the simplicity of the current disease model, and suggest that sequestration of MBNL1 alone may not explain all of the disease-associated mis-splicing events and that MBNL1's cellular functions may be very complex. When sequestered to toxic RNA repeats in the nucleus, MBNL1 might be capable of recruiting other proteins

to and stabilizing the ribonuclear foci. If other important splicing factors are co-localizing to nuclear foci with MBNL1 and CUG repeats in DM1, identification of protein interaction domains within MBNL1 that mediate protein:protein interactions may help to interrogate the disease-associated mis-splicing events currently unaccounted for in the MBNL1 sequestration model. One of the largest motivations for understanding MBNL1's function as a splicing factor in respect to DM1 is to move towards the design of effective DM1 treatments. Therefore a thorough understanding of the mechanism(s) by which MBNL1 regulates alternative splicing is vital to truly understanding the disease. Especially those subtleties related to how MBNL1 directs the splicing machinery's decision to include or exclude alternatively spliced exons within its target transcripts.

Most of the splicing factors that have been examined contain unconserved regions outside of their RNA binding domains that are open to bind to other proteins and splicing factors. Current thinking dictates that these protein interaction domains are open to form either "splicing activation" or "splicing suppression" complexes based on the proteins that they interact with, which allows for control and fine tuning of alternative splicing decisions. Therefore it is easy to suggest that when bound to an exon that these domains act as protein/protein domains and recruit activation or suppression complexes. This has been demonstrated extensively with the SR family of proteins (105), (24), (106) which have RNA binding domains that are separable from protein interactions domains. Based on these observations, one could argue that once bound to a target transcripts that MBNL1 mediated the splicing decision via interactions with other splicing factors to form either activation or suppression complexes at the RNA transcripts, and that those complexes subsequently regulate splicing

In humans there are 3 paralogues of the splicing factor MBNL1; denoted MBNL1, MBNL2, and MBNL3. These genes gives rise to at least nine, three, and six protein isoforms, respectively (60). Of the 18 protein isoforms expressed, 10 of them differ in regions outside of the zinc finger (ZF) RNA binding domains (60), suggesting that regions outside of the ZFs may be important for MBNL1 activity. The full-length isoform of MBNL1, denoted MBNL1_v3, contains 388 amino acids and is the most abundant and ubiquitously expressed MBNL1 protein (60). This is the isoform of MBNL1 whose improper localization and function has been implicated as the causative

factor in DM1. Therefore, the majority of MBNL1 isoforms have potentially relevant domains with currently undefined functions. Recent MBNL1 deletion analysis followed by functional characterization narrowed down the potential domains to two 80 amino acid regions of MBNL1, lying just outside of the ZF domains (91).

The MS2 coat protein is derived from the MS2 bacteriophage (R17) and specifically binds a well-characterized 19-nt RNA aptamer that with very high affinity (107). If fused to the MS2 coat protein, any protein can be artificially recruited to any RNA target by inserting the MS2 aptamer into the target sequence. The MS2 system has been used in various "tether" assays as a way of recruiting a protein to an RNA substrate independent of a direct protein: RNA interaction. The splicing field adopted the MS2 system as a means of classifying binding sites of splicing factors within transcripts.

Using MS2 tethers, researchers demonstrated that the SR family of splicing factors, proteins known to stimulate splicing, has protein: protein interaction motifs that are separable from RNA binding domains (108), (106). MS2 has also been used to not only define the silencing function of hnRNP A1 (109) and to show that TIA-1 enhances splicing via direct interactions with U1 (110), but has been used to directly study the relative strengths of these splicing factor with respect to one another on a single transcript (111). Most recently, the MS2 system has been used to develop a synthetic genetic system that senses cellular protein concentrations to effect alternative splicing decisions, which subsequently reprogram gene expression events (112). I previously demonstrated in Chapter III that potential protein:protein interaction may governs MBNL1-mediated splicing decisions for Class I-type splicing events. Using an MS2 tethered function assay, MBNL1's function as an RNA binding protein can be separated from its ability to regulate enhancement or suppression of alternative splicing. Through systematic analysis of MBNL1 domains via MS2, residues within the splicing factor that are essential for regulation of splicing, but independent of RNA binding can be studied. The MS2 system also provides a powerful means for identifying and evaluating proteins that may be interacting with MBNL1 near splice sites.

Like other well studied splicing factors, including Fox-1/2 (113), (39), (38), Nova (36), (37), PTB (31), and CUGBP1 (89), MBNL1 has been shown to act as either a splicing enhancer or suppressor based on the location of its binding site within target

transcripts (40), (41). Therefore, the positional dependence of splicing factor binding in relation to alternatively spliced exons is a regulatory theme of alternative splicing that has recently begun to take shape. However, using synthetic reporters, I demonstrated that simply repositioning a high affinity MBNL1 binding motif relative to an exon (See Appendix A) is not sufficient to recapitulate the positional effects of MBNL1 binding on splicing regulation. I was only able to create reporters for MBNL1-mediated suppression.

I hypothesize that additional RNA signals, sequences and/or structures, within MBNL1-target transcripts but outside of the MBNL1 binding site are important for regulation. Every human transcript is riddled with splicing signals of varying strengths and overlapping protein-binding sites also of varying strength. Thus, it is likely that MBNL1 binding sites and have co-evolved with other splicing motifs or binding sites within endogenous targets and that these signals work synergistically with MBNL1 to mediate some splicing events. Following this line of reasoning it seems likely that I was unable to create a synthetic reporter for MBNL1-mediated exon inclusion with only MBNL1 binding sites due to the lack of surrounding RNA signals. Therefore I designed a series of chimeric reporters using two well-characterized MBNL1-regulated exon inclusion events, INSR and ATP2A1. The chimeras are designed to combine minimal intronic and exonic sequences from the endogenous targets with flanking PLEK sequence to investigate the minimal requirements for MBNL1-mediated exon inclusion. The goal in designing these reporters was to create a minimal reporter for MBNL1-mediated exon inclusion that could be used to investigate and potentially define the mechanism(s) governing MBNL1's splicing enhancer activity.

RESULTS

MBNL1-MS2 Regulates Splicing of PLEK-MS2 Reporters

Three, synthetic MS2 splicing reporters were cloned by inserting MS2 aptamers into the PLEKHH2 minigene. Studies using MS2 aptamers in synthetic splicing constructs have demonstrated that having multiple MS2 sites near each other increases the efficiency of splicing (108), therefore all splicing reporters contain multiple MS2 sites. Two of the splicing reporters contained MS2 sites downstream of PLEK exon 21 and one reporter was made with MS2 sites upstream of exon 21 to evaluate the positional

effects of MBNL1 recruitment to splice junctions. Two different insert sequences, both containing two MS2 aptamers, were used to create the various reporters. We tested two different sequences to ensure an MS2-mediated splicing effect would be observed. See Figure 1A for a schematic of the three reporters. The different MS2 inserts are denoted with different colors in Figure 1A. Details regarding the sequence of the inserts can be found in the Material and Methods. The MS2 coat protein (14 kDa) was fused onto the C'-terminus of MBNL1 (1-283; 31kDa). The MBNL1-MS2 (48 kDa) fusion protein had slightly increased expression level in HeLa cells in comparison to wild-type MBNL1 (1-382; 41 kDa) (Figure 1B).

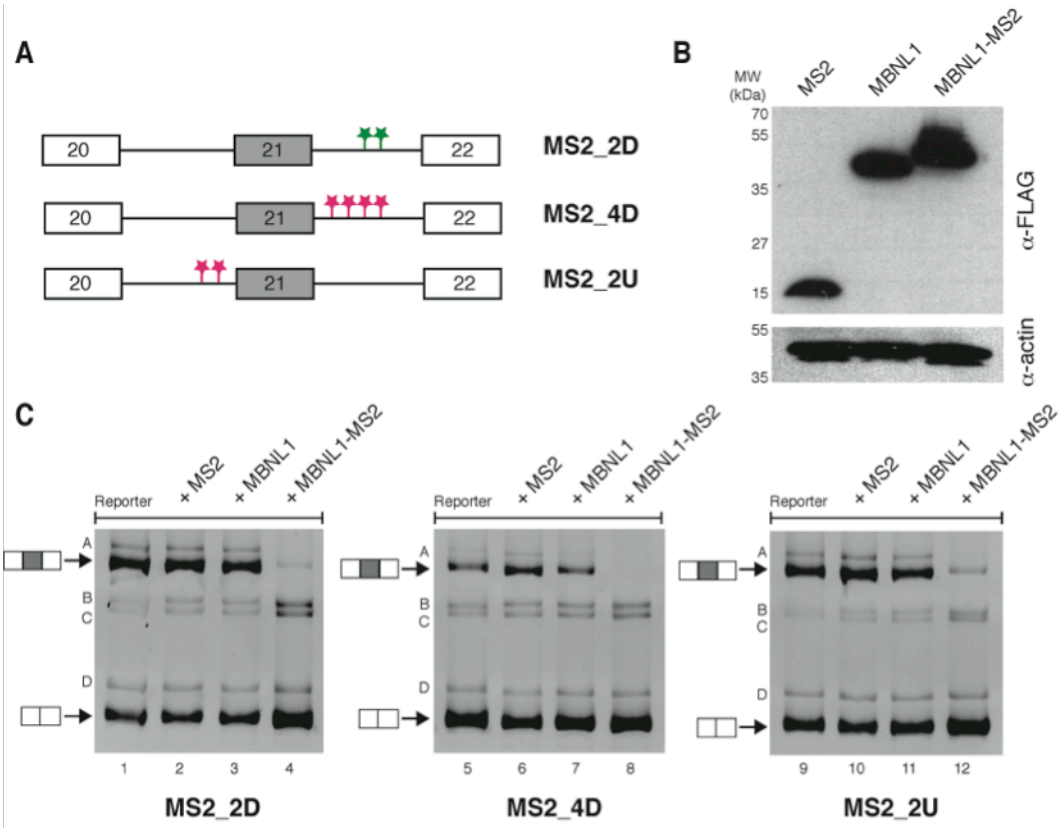


Figure 1. Splicing of PLEK-MS2 reporters. (A) Schematic representation of PLEK-MS2 reporters. Star lollipops represent position and quantity of MS2 sites inserted into PLEKHH2 minigene. MS2 inserts denoted in green contain motifs derived from SELEX experiments, while the pink motifs contain the endogenous aptamer. See Material and Methods for exact sequences. (B) Western blot showing expression of N'-terminally FLAG-tagged MS2, MBNL1, and MBNL1-MS2 in HeLa cells. (C) Representative splicing gels showing PLEK-MS2 reporters splicing. Exon inclusion ans exclusion products are denoted. Bands A-D are unidentified splice products.

Splicing of all three MS2 splicing reporters in HeLa cells resulted in multiple bands (Figure 1C, lanes 1, 5, and 6). Several attempts were made to sequence these bands but were unsuccessful. For that reason the bands corresponding to uncharacterized splice products are labeled A-D in Figure 1C. Splice products A-D are created via activation of cryptic splice sites that are found throughout the PLEKHH2 minigene. However, the specific location and identity of the splice sites that are activated in these reporters is currently unknown. Based on the size of the bands, splice product A contains exon 21, and is therefore an exon inclusion product. Bands B-D could be the result of activation of cryptic splice sites within exon 21, and are therefore also inclusion products. Or, bands B-D could be the result of activation of distal splice sites found near the flanking exons, which would make them exclusion products. It is important to note that MS2_2D and MS2_2U, which contain MS2 inserts that differ slightly in sequence, spliced identically (Figure 1C, lanes 1 and 9).

In comparison to the reporter that only has two MS2 sites (*i.e.* MS2_2D), the presence of four MS2 aptamers in MS2_4D reduced basal exon 21 inclusion levels (Figure 1C, compare exon 21 inclusion band and band A in lane 1 to the same bands in lane 5). This observation is not surprising because it is well known that when RNA structures are found near splice sites they reduce usage of the splice site (114). It is also apparent that increasing the number of MS2 sites from two to four in these reporters does not increase the splicing efficiency.

All three splicing reporters were non-responsive to either MS2 or MBNL1 over-expression (Figure 1C, compare lanes 1-3, 5-7 and 9-11). However, all three reporters were responsive to over-expression of the MBNL1-MS2 fusion protein (Figure 1C, compare lanes 1 to 4, 5 to 8 and 9 to 12). For all three MS2 reporters, over-expression of the fusion protein caused a drastic decrease in exon 21 inclusion, resulting in increased exon skipping and activation of cryptic splice products B and C. Splice product A is decreased in response to MBNL1-MS2 over-expression for all three reporters. This response supports the size-based prediction that it is a second exon 21 inclusion band. The opposite trend was observed for splice products B-D, which all increased in response to MBNL1-MS2 over-expression. Therefore these splice products are most likely due to

activation of 3'- or 5'-splice sites near the flanking exons 20 or 22 because they respond like exon exclusion products.

MBNL1 Activates Exon Inclusion in Chimeric Reporters

MBNL1 regulates exon 11 inclusion in the insulin receptor (INSR) transcript (65) and exon 22 in the sarcoplasmic/endoplasmic reticulum Ca^{2+}-ATPase 1(SERCA1 or ATP2A1) transcript (70). Functional MBNL1 binding sites have been identified in both reporters, however the architecture of the binding site differs substantially between the two transcripts. In the case of INSR, the MBNL1 binding site seems to be very simple and contains a single cluster of YGCY motifs located within 50-nucleotides of exon 11(90), (91). The MBNL1 binding site within the ATP2A1 contains vast clusters of YGCY motifs that span more than 200-nucleotides on intronic sequence downstream of exon 22. As discussed in detail in Chapter II, using a series of deletions and ASOs to block these extensive YGCY clusters, I found that no single YGCY motif, or cluster of YGCY motifs was necessary for MBNL1 function but rather that all motifs contributed to function (41). I also demonstrated in Chapter III that upon analysis with MBNL1 mutants, INSR and ATP2A1 minigenes fell into different splicing classes based on MBNL1-function. Thus is it likely that these exon events are both positively regulated by MBNL1 but via different mechanisms. Therefore in designing the chimeric reporters, sequences from both INSR and AT2A1, were used.

For all chimeric splicing reporters the PLEKHH2 minigene was used as the splicing minigene scaffold. To create the PLEK/INSR (PI) reporter, exon 21 and intron 21 of PLEK were swapped for exon 11 and intron 11 of INSR. The resulting chimera is diagrammed in Figure 2A. Splicing of the PI reporter in HeLa cells resulted in approximately 100% exon skipping (Figure 2B, lane 1). Co-expression of MBNL1 with the PI reporter resulted in a robust activation of splicing at cryptic splice sites (Figure 2B, lane 3, bands A and B) that lie upstream of the INSR insert within PLEK intron 20. This result was promising because it suggested that: 1) MBNL1 was recognizing and regulating splicing of the chimera, and 2) that MBNL1 was activating splicing at upstream splice sites. However MBNL1 was not activating splicing at the splice sites that correspond to INSR exon 11. In order to eliminate usage of the cryptic splice sites, I

mutated five cytosines to thymine to strengthen the poly-pyrimidine (py) tract in PLEK intron 20 and increase the splicing potential of the INSR exon (Figure 2A). The resulting reporter, PI-PY, spliced with 100% exon skipping (Figure 2C, lane 1), identical to the PI reporter (compare lanes 1 in Figures 2B and 2C). Co-expression of MBNL1 with the PI-PY reporter resulted in robust exon 11 inclusion and modest activation of splice product C (Figure 2C, lane 3). MBNL1-mediated splicing of the PI-PY reporter is the first observation of MBNL1 mediated exon inclusion in a synthetic reporter system.

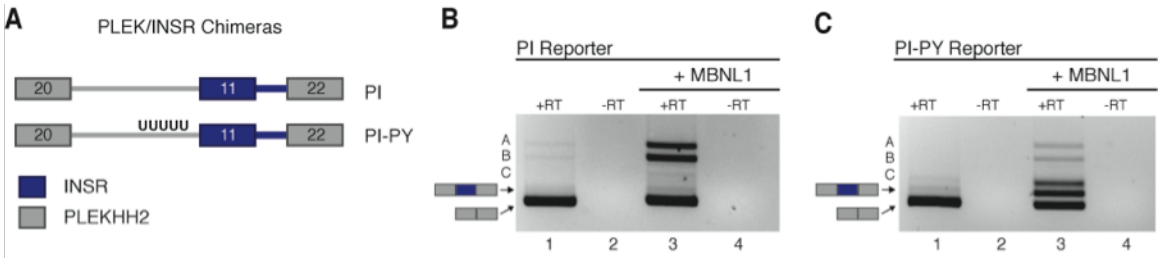


Figure 2. Splicing of PLEK/INSR Chimeras. (A) Schematic of PLEKHH2/INSR chimeric splicing reporters, PI and PI-PY. The py-tract was mutated in the PI reporter to create the PI-PY reporter (designated with Us). PLEK sequence is shown in grey and INSR sequence in blue. Boxes represent exons and lines represent introns. (B) and (C) Splicing of PI and PI-PY reporters, respectively, with (+ MBNL1) and without MBNL1 co-expression. Exon 11 inclusion and exclusion product are designated. Bands A-C are due to unidentified splice products.

The PLEK/ATP2A1 (PA) chimeric reporter was created by replacing exon 21 and intron 21 of PLEK with exon 22 and intron 22 of ATP2A1 (Figure 3A). Similar to the PI reporter, 100% exclusion of exon 22 was observed when the PA reporter was spliced in HeLa cells (Figure 3B, lane 1) and co-expression of MBNL1 with the PA reporter resulted in robust activation of upstream splice sites (Figure 3B, lane 3). In contrast to PI however, MBNL1 robustly activated inclusion of ATP2A1 exon 22 in addition to the upstream cryptic sites (Figure 3B, lane 3, bands A and B) in the PA reporter. This observation may be explained by the presence of a large number of MBNL1 binding sites in the PA reporter, but only a single site in the PI reporter, hence the local concentration of MBNL1 is higher near splice sites in the PA reporter than in PI. Strengthening the

py-tract of PLEK intron 20 in the PA reporter (*i.e.* PA-PY) resulted in a complete ablation of cryptic splice usage and a robust increase in basal exon 22 inclusion levels (compare lanes 1 in Figure 3B and 3C). Over-expression of MBNL1 with the PA-PY reporter resulted in almost complete exon 22 inclusion (Figure 3C, lane 3).

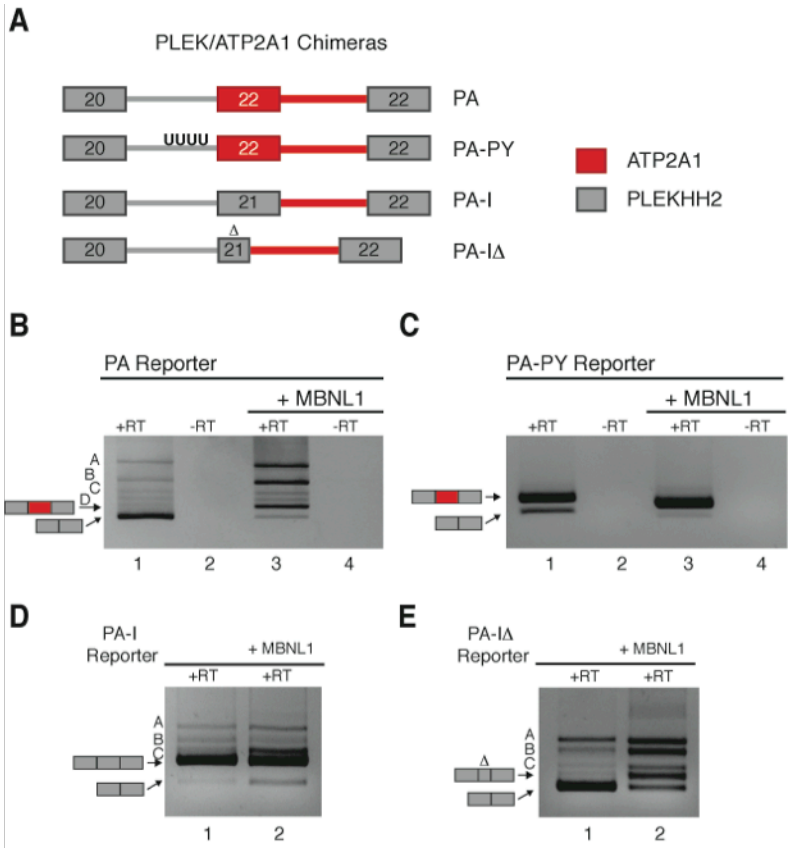


Figure 3. Splicing of PLEK/ATP2A1 Chimeras. (A) Schematic of PLEKHH2/ATP2A1 chimeric splicing reporters: PA, PI-PY, PA-I, and PA-IΔ. PLEK sequence is shown in grey and ATP2A1 sequence in red. Boxes represent exonic sequence and lines represent introns. PLEK exon 21 was truncated to create the PA-IΔ reporter. The truncation is indicated above the exon (Δ). The py-tract was mutated in the PA reporter to create the PA-PY reporter (designated with Us). (B) and (C) Splicing of PA and PA-PY reporters, respectively, with (+ MBNL1) and without MBNL1 co-expression. (D) and (E) Splicing of the PA-I and PA-IΔ reporters, respectively, which contain only intronic sequence from ATP2A1 with (+ MBNL1) and without MBNL1 co-expression. Exon 11 inclusion and exclusion product are designated. Bands A-C are unidentified splice products.

Due to the robust splicing activity of the PA and PA-PY reporters I wanted to use this scaffold to investigate exon 22's contribution to MBNL1-mediated exon activation.

To this end I created an additional reporter that simply replaced PLEK intron 21 with ATP2A1 intron 22 (PA-I, Figure 3A) to test if intron 22 was sufficient for positive regulation by MBNL1. In stark contrast to the PA reporter, PA-I spliced with nearly 100% exon inclusion (Figure 3D, lane1). PA-I contains PLEK exon 21, which is nearly twice the size of ATP2A1 exon 22, explaining the size difference between inclusion products relative to exclusion products in lanes 1 of Figures 3B and 3D. Although the basal exon 21 inclusion level for PA-I was 100%, co-expression of MBNL1 with this reporter still resulted in activation of upstream cryptic splice sites (Figure 3D, lane 2, band C), suggesting that intron 22 of ATP2A1 contains all the required splicing signald for MBNL1-mediated regulation. The robust inclusion of exon 21 observed in the PA-I reporter is not surprising because this exon is constitutively spliced and contains several potent exonic splicing enhancers (ESEs) in its 3' end. To reduce the strength of exon 21 within the PA-I reporter, I truncated the 98-nucleotide exon by deleting 42-nucleotides that contains predicted SR protein binding sites near the 3' end (PA-ID, Figure 3A). The last 12-nucleotides (*i.e.* 3' end) of the exon were retained to ensure that the PLEK exon 21-ATP2A1 intron 22 junction was unchanged. In dramatic contrast to the PA-I reporter, exon exclusion was the major splice product observed for PA-ID (Figure 3E, lane 1) with minor usage of cryptic sites within PLEK intron 20 (Figure 3E, lane 1, bands A and B). When MBNL1 was co-expressed with PA-ID, it not only activated splicing of exon 22 but also activated every splice site upstream of exon 22 (Figure 3E, lane 2), suggesting that removal of the ESEs was critical for MBNL1 to mediate exon inclusion and that ATP2A1 intron 22 contains all the necessary signals for MBNL1 to facilitate splice site activation.

DISCUSSION AND FUTURE DIRECTIONS

By utilizing the well-characterized MS2 tether system, I was able to artificially recruit MBNL1 to splice sites within synthetic reporters and evaluate its ability to regulate splice site activation independent of a direct RNA interaction (Figure 1). Recruiting MBNL1 to positions upstream or downstream of exon 21 in various PLEK-based reporters all resulted in MBNL1-dependent exon suppression. Although preliminary, these observations may suggest that the splicing effects observed

independent of a direct RNA interaction are the consequence of MBNL1's ability to interact with other splicing factors or components of the core splicing machinery.

Splicing patterns observed with the PLEK-MS2 reporters were similar to those observed for synthetic PLEK reporters containing high affinity MBNL1 binding sites ((41)and Appendix I), suggesting that MBNL1 regulates splicing of all of the synthetic PLEK-based reporters in a similar fashion. I have used the PLEKHH2 minigene as the template for these synthetic reporters because it a constitutively spliced event that is not regulated by MBNL1. However, MBNL1-mediated exon exclusion is consistently observed with all synthetic PLEK reporters. This trend might suggest that once recruited to PLEK, either via artificial insertion of high-affinity binding sites or through MS2, MBNL1 "finds" YGCY motifs within the minigene and affects splicing either through displacing splicing enhancers, interfering with the spliceosome's ability to recognize appropriate splice sites, or by inducing inhibitory structure near splice sites. If this is true, then the splicing observed in the MS2 system is not due to MBNL1 recruiting other protein factors to the RNA substrate, but rather is caused by MBNL1 itself interacting with the substrate after being localized to the substrate via MS2. I describe a series of MBNL1 mutants in Chapter III and Appendix B (*i.e.* triple RIMs) that are unable to bind to RNA with high affinity. A logical control that remains to be conducted is to fuse an MBNL1 mutant, with severely reduced RNA binding activity (such as any triple RIM), to MS2 and evaluate if the inherent RNA binding activity of MBNL1 is responsible for the splicing effects observed with the MS2 reporters or if MBNL1-MS2 splicing activity in independent of MBNL1's ability to bind RNA. If the mutant-MS2 is able to regulate splicing as well as wild-type MBNL1-MS2, then the next logical step would be to utilize the deletion mutants described by Grammatikakis et al. and narrow down the regions of MBNL1 required for splicing activity to identify what factors might be interacting with those domains (91).

MBNL1-Mediated Exon Inclusion Requires Additional Splicing Signals Found Within Regulated Introns

When given intronic sequences from endogenous targets in a chimeric context, MBNL1 is able to enhance exon inclusion (Figures 2 and 3). This is a very significant

observation because neither insertion of high affinity binding sites alone, nor recruitment of MBNL1 to synthetic reporters with MS2 resulted in MBNL1-mediated exon inclusion. Taken together these results argue that there are other sequences and/or structures within MBNL1-regulated introns that are required for MBNL1 mediated inclusion. One possible explanation is that there are other protein binding sites within these intronic regions that co-occur with MBNL1 binding sites and cooperate with MBNL1 to recruit factors that enhance splicing and/or recruit the splicing machinery.

As a whole the results from the chimeric reporters indicate that MBNL1 can mediate exon inclusion if the context of the reporter is appropriate. For example, strengthening the py-tract in PLEK intron 20 helped reduce activation of cryptic splicing signals when endogenous exons were inserted in place of PLEK exon 21 (Figure 2B and 3C). In both cases the endogenous exon doesn't contain the strong splicing signals contained within PLEK exon 21, therefore strengthening the py-tract helps to define the endogenous exon.

Exon 21 of PLEK contains some strong splicing enhancers that over-powered splicing signals found within the MBNL1 target introns. However, truncation of PLEK exon 21 resulted in a drastic decrease in the strength of the exon, resulting in a splicing readout that was amenable to visualizing splicing enhancement effects. Future experimentation should be done to strengthen the Py-tract in the PA-ID reporter and to mutate the cryptic splice site responsible for band C in the PI-PY reporter. I predict that these changes would reduce activation of cryptic splice sites. The resulting reporters would be a robust splicing constructs containing minimal MBNL1 sequence and would be ideal for investigating the mechanisms governing MBNL1-mediated exon inclusion and identifying additional factors involved in this type of splicing regulation.

MATERIALS AND METHODS

Cloning of PLEK-MS2 splicing reporters. PLEK-MS2 reporters were cloned using standard, overlapping PCR techniques. The PLEK-D mutant minigene was used as the cloning template for all MS2 reporters. The wild-type PLEK mutant minigene has a potent intronic enhancer element (ISE) downstream of exon 21. In the PLEK-D reporter this ISE was mutated so that the mutant reporter has a reduced basal exon 21 inclusion

ratio in comparison to wild-type PLEK, which was necessary in creating the reporters for this study so that they have an exon 21 inclusion ratio near 50%.

To create the MS2_4D and MS2_2U reporters a 56-nucleotide MS2 insert was inserted either 33-nucleotides upstream or downstream of PLEK exon 21. The MS2_4D reporter contains 2, tandem copies of the 56-nucleotide insert. The 56-nucleotide insert contains two, tandem 19-nucletotide MS2 aptamers (5'-ACATGAGGATCACCCATGT-3') (111) proceeded by an 8-nucletide spacer sequence (5'-CCATTCAA-3') (107). This spacing was provided to allow neighboring MS2 sites to fold properly. To create the PLEK-MS2 reporter with two MS2 sites downstream of PLEK exon 21 (MS2_2D) the following primers were used: 5'- CCATTCAAACATGAGGATCACCCATGTCCATTC AAACATGAGGATCACCCATGTCGTTTCAGAAGAATGTATGAAACAAATATTG GTACAGG -3' (sense) and 5'- CGACATGGGTGATCCTCATGTTTGAATGGACATG GGTGATCCTCATGTTTGAATGGATAAACAGTTCAGCTGGCTAGTATTCTGC -3' (antisense). To create the PLEK-MS2 reporter with four (MS2_4D) MS2 sites downstream of PLEK exon 21 the following primers were used, twice: 5'- CCATTCAA ACATGAGGATCACCCATGTCCATTCAAACATGAGGATCACCCATGTCGTTTC AGAAGAATGTATGAAACAAATATTGGTACAGG 3' (sense) and 5'- CGACATGGG TGATCCTCATGTTTGAATGGACATGGGTGATCCTCATGTTTGAATGGATAAA CAGTTCAGCTGGCTAGTATTCTGC -3' (antisense).

The MS2_2D reporter was created by inserting a 68-nucleotide motif containing two, tandem 34-nucleotide MS2 binding sites 66-nucleotides downstream of PLEK exon 21. The 34-nucleotide sequence (5'-CGCGTACACCATCAGGGTACGCTGCAGTCG ACTT-3') was previously used as a minimal MS2 substrate in *in vitro* splicing assays with HeLa nuclear extract (105) and is based off of the consensus MS2 binding sequence derived from SELEX experiments (115). This reporter was cloned using: 5'- CGCGTAC ACCATCAGGGTACGCTGCAGTCGACTTCGCGTACACCATCAGGGTACGCTGC AGTCGACTTAAGAATGTATGAAACAAATATTGGTACAGGAA -3' (sense) and 5'- AAGTCGACTGCAGCGTACCCTGATGGTGTACGCGAAGTCGACTGCAGCGTA CCCTGATGGTGTACGCGACTTATGATTTATCGTAGACTTTCAGAGGCCA -3' (antisense) primers.

For all three MS2 reporters, flanking PLEK exon 20 sense (5'-CGGGGTACCAA ATGCTGCAGTTGACTCTCC -3') and PLEK exon 22 antisense (5'- CCGCTCGAGCC ATTCATGAAGTGCACAGG -3') primers, containing KpnI and XhoI restriction enzyme sites, respectively, were used in combination with the mutagenic primers to created full-length minigene inserts via PCR. The final PCR products were ligated into pcDNA3 using KpnI and XhoI sites.

Cloning of MBNL1-MS2 protein. The pCI-MS2-NLS-FLAG construct was a generous gift from Dr. Richard Breathnach (109). This construct contains a single, unique StuI restriction enzyme site between the N'-terminal FLAG epitope and the C'-terminal SV40NLS-MS2 sequence. MBNL1 contains a single StuI restriction enzyme site after amino acid 283. Therefore a truncated version of MBNL1 (1-283) was sub-cloned from the pEGFP-C1 vector and inserted into the pCI-MS2-NLS-FLAG vector using flanking StuI sites. The sense primer used for sub-cloning MBNL1 from the eGFP-C1 vector added a StuI site to the 5'end of the PCR product (5'-AGGCCTATGGCTGT TAGTGTCACACCAATTCGG -3'). This primer was used in combination with an eGFP-C1 plasmid specific antisense primer (5'-GGGAGGTGTGG GAGGTTTT-3'). The resulting PCR product was digested with StuI and ligated into the pCI-MS2 vector. The final MBNL1-MS2 fusion vector contains an N'-terminal FLAG epitope (MDYKDDDDK), followed by MBNL1 (1-283), the nuclear localization signal from simian virus 40 T antigen (PPKKKRKVD), and finally the MS2 (1-130) coat-protein. Please note that several attempts were made to express MBNL1-MS2 fusion proteins in human cell culture from pcDNA3 (Invitrogen), pcDNA3.1 (Invitrogen), or pTRIEX (Novagen) vectors and in all cases expression of fusion protein was never observed via Western blot.

Cloning of PLEK/INSR and PLEK/ATP2A1 chimeras. The PI, PA, and PA-I reporters were created using overlapping PCR techniques in two segments. For the PI and PA reporters the portion of the reporter containing PLEK exon 20 and PLEK intron 20 was amplified off of the full length PLEKHH2 minigene using the following primers: 5'-GCACCAGTGCCTGAAGAGGTTTTTCTGGGCAGGGAGAAGCAC -3' (antisense for PI) and 5'- GATGGAGGAGGGGGAACAGTTATCCTGGGCAGGGAGAAGCAC -3' (antisense for PA). For the PA-I reporter, which contains PLEK exon 20, intron 20, and

exon 21, the following antisense primer was used: 5'- CTGGAATCTGCATTCCTCTTT AGGTG -3'. All three antisense primers were used in combination with the PLEK exon 20 sense primer described above to create the 5' portion of the chimera. The INSR portion of the PI reporter and the ATP2A1 portions of the PA and PA-I reporters were amplified from the INSR (98) and ATP2A1 (41) minigenes using a sense primers to exon 11 (5'- GTGCTTCTCCCTGCCCAGAAAAACCTCTTCAGGCACTGGTGC -3'), exon 22 (5'- GTGCTTCTCCCTGCCCAGAAAAACCTCTTCAGGCACTGGTGC -3'), or intron 22 (5'- CACCTAAAGAGGAATGCAGATTCCAGGTATCACCCCCTTCTT GCCC -3'), respectively. All of these sense primers contain a 5'-overhang that is complementary to PLEK intron 20 or exon 21 and were used with the PLEK exon 22 antisense primer described above. Once PCR product for both halves of each reporter was obtained, overlapping PCR was conducted with the flanking PLEK exon 20 and 22 primers to combine the two portions into a final PCR product corresponding to the full-length chimeric reporter.

The PI and PA reporters were used as templates to create the PI-PY and PA-PY reporters. Five using cytosine residues were mutated to thymidine using standard mutagenic techniques. The primers used to install py-tract point mutation into PI were: 5' CTAACCTGT GGTGCTTCTTTTTGTTCAGAAAAACC 3'(sense) and 5'- GGTTT TTCTGAACAA AAAGAAGCACCACAGGTTAG -3'(antisense). The primers used to install py-tract point mutation into PA were: 5' GTGGTGCTTCTTTTTGTTCAGGATA ACTGTTCC 3'(sense) and 5'- GGAACAGTTATCCTGAACAAAAAGAAGCA CCAC -3'(antisense). The PA-ID reporter was cloned using the PA-I reporter as a template with the following primers: 5'-CTTGGCACTCTGCGTTGGGTGCAGATTCCAGGTATCA CCCCC-3' (sense) and 5'- CCCAACGCAGAGTGCCAAG -3' (antisense). All full-length PCR products were digested with KpnI and XhoI and ligated into pcDNA3.

Cell culture and transfection. HeLa cells were routinely cultured as a monolayer in DMEM+GLUTAMAX media (Invitrogen) supplemented with 10% Fetal Bovine Serum (Gibco) at 37°C under 5% CO_2. Prior to transfection, cells were plated in 6-well plates at a density of $1.8x10^5$ cells/well. Cells were transfected 18-24 hours later at approximately 80% confluency. Plasmid (1 ug/well) was transfected into each well with 5 uL Lipofectamine2000 (Invitrogen) following manufacturer's protocol. For all splicing

assays 500 ng of each protein plasmid or empty pcDNA3 plasmid was co-transfected into a single well with 500 ng of reporter. Cells were harvested with TrypLE (Invitrogen) 18-24 hrs after transfection and pelleted via centrifugation.

Cell-based splicing assay and RT-PCR. Total RNA was isolated from HeLa pellets using an RNeasy kit (Qiagen). Isolated RNA (500 ng) was incubated with 1 unit of RQ1 DNase (Promega) in a 10 uL reaction for 1 hour at 37° C. DNased RNA (2 uL (100 ng)) was reverse transcribed in a 10 uL reaction (1:5 dilution) using Superscript II (Invitrogen), according to manufacturer's protocols, with the exception that we used half the recommended amount of Superscript II. All PLEKHH2 reporters were reverse transcribed using an antisense primer (5'-AGCATTTAGGTGACACTATAGAATAG GG-3') designed to the Sp6 promoter site of the pcDNA3 plasmid. The minus reverse transcription (-RT) reactions were treated identically to plus reverse transcription (+RT) reactions except the Superscript II was replaced with water. All reverse transcription reactions (2 uL) were subjected to 22 rounds of PCR amplification, which was found to be within the linear range for the primers used (data not shown), in a 20 uL reaction (1:10 dilution). PCR to analyze the splice products for all PLEKHH2 reporters was conducted using the PLEK exon 20 and exon 22 primers described above**.** The resulting PCR products were resolved on 3% agarose gels and the splice products were visualized using SYBR Green I nucleic acid stain (Invitrogen) in combination with AlphaImager HP system (Alpha Innotech).

Western blot analysis. HeLa cells were plated, transfected, and harvested for Western blot analysis in the same fashion as for splicing assays. Cell pellets were lysed in RIPA buffer (50 mM Tris pH 7.4, 150 mM NaCl, 1% NP-40, 0.25% sodium deoxycholate, 0.1% SDS, 1 mM PMSF) supplemented with 1X protease inhibitor cocktail (SIGMA*FAST*; Sigma) via 3 freeze-thaw cycles. Concentration of protein was quantified and normalized using BCA reagent (Peirce) and 5-10 mg of total protein was loaded on 12% SDS-PAGE gels. Resolution of proteins and transfer onto nitrocellulose was conducted using standard techniques. All proteins were N'-terminally FLAG tagged amd were probed 1:1,000 FLAG (M2) primary antibody (Invitrogen) followed by 1:2,000 bovine α-mouse secondary IgG-HRP (Santa Cruz; sc-2371). The actin loading control was probed using actin (I-19) (Santa Cruz; sc-1616) primary (1:1,000) antibody and

donkey α-rabbit secondary IgG-HRP (Santa Cruz; sc-2313). Chemilluminescence was visualized using a SuperSignal (Pierce) kit.

APPENDIX A

FUNCTIONAL CHARACTERIZATION OF ADDITIONAL PLEK-BASED SYNTHETIC SPLICING REPORTERS

INTRODUCTION

A subset of synthetic PLEK splicing reporters were cloned and assayed for this study in addition to the reporters described in Figure 5 of Chapter II and the corresponding text. Due to spatial limitation these results were not included in the published manuscript.

The crystal structure (61) of MBNL1 and biochemical data (96) defined the core RNA recognition motif for MBNL1 as 5'-YGC-3' or 5'-GCY-3'. More specifically, each zinc finger (ZF) of MBNL1 specifically binds one GC dinucleotide. MBNL1 is able to bind synthetic substrates containing only two GC dinucleotides with similarly high affinities whether the dinucleotides are closely spaced or separated by several nucleotides (96). Mbl, the *Drosophila* orthologue of MBNL1, only has two ZFs. There is unpublished biochemical evidence from our lab indicating that, unlike MBNL1, the binding affinity and specificity of Mbl is dependent upon spacing of GC dinucleotides within RNA substrates. It appears as though Mbl cannot bind synthetic RNA substrates containing only two, closely spaced GC dinucleotides (*i.e*. no nucleotides between GCs) with high affinity. However, Mbl does bind RNA substrates with high affinity if the two GC nucleotides are separated by nine to thirteen nucleotides.

Models that invoke RNA and protein looping have been proposed to explain these observations (96), (61). These models propose that MBNL1 is able to interact with multiple GCs that are either separated by multiple nucleotides or closely spaced because MBNL1 has a long, flexible linker between the two ZF pairs. This linker allows the two ZF pairs to wrap around single stranded RNA substrates and interact with multiple, closely spaced GC dinucleotides. In contrast, for the single pair of ZFs in Mbl to interact with multiple GC dinucleotides in a single RNA substrate, the RNA must wrap around the pair of ZFs. Therefore, Mbl can only interact simultaneously with appropriately spaced GCs.

I was interested in testing this phenomenon in a biological system using synthetic reporters. The first subset of synthetic reporters was based off of the original PLEK-SM1F (Int 20) reporter that is described in detail in Chapter II. This reporter was created by inserting a 36-nt, high affinity MBNL1-binding site, into intron 20 of the PLEK minigene. Additional reporters were constructed by altering the 36-nt motif to include a reduced number of GC dinucleotides with varied spacing between the YGCY motifs. I then tested the ability of MBNL1 to regulate splicing of these reporters in cell-based splicing assays.

A correlation between the position of MBNL1 binding sites within endogenous targets, relative to regulated exons, and MBNL1's mode of splicing regulation was described in Chapter II (41). In general, for MBNL1-mediated exclusion events the MBNL1 binding sites are located within the acceptor side of the intron. Conversely, binding sites for MBNL1-mediated inclusion events are generally found near intronic donor sites. The final subset of synthetic reporters created for this study was based on the prediction that recruitment of MBNL1 to intronic regions downstream of an exon would result in increased inclusion of that exon. Therefore, I created an additional synthetic reporter wherein the high affinity MBNL1 binding sites was placed downstream of PLEK exon 21. Splicing of this reporters was evaluated to determine if changing the position of the MBNL1 binding sites, relative to exon 21, could result in a synthetic reporter wherein MBNL1-mediates exon inclusion.

RESULTS

MBNL1 Regulates Splicing of PLEK Reporters with a Reduced Number of GC Dinucleotides

I created three PLEK reporters where the number of GC dinucleotides in the SM1F motif in intron 20 of PLEK was reduced from six GCs to only two GCs, and one reporter that contained 4GCs. The sequence of the resulting 36-nt motifs are provided in Figure 1A. For the PLEK-2GC1 reporter the four remaining GCs were mutated to CC. Two of the uridines were also mutated to cytosine to control for the unintentional creation of PTB binding sites (*i.e.* CUCU) (116) and to avoid runs of uridines, which could create a pseudo py-tract and activate improper splicing. For the PLEK-2GC2 and PLEK-2GC3

reporters, four of the GC motifs were mutated to GG. In the PLEK-2GC1 and PLEK-2GC2 reporters the remaining GC motifs are separated by 16 nucleotides. In the PLEK-2GC3 reporter the remaining GC dinucleotides are only separated by 4 nucleotides. The PLEK-4GC reporter, has two GC motifs mutated to GG, and contains two GC motifs that separated by 4 nucleotides flanked by a second set of GC dinucleotides with a 10-nucleotide spacer on each side.

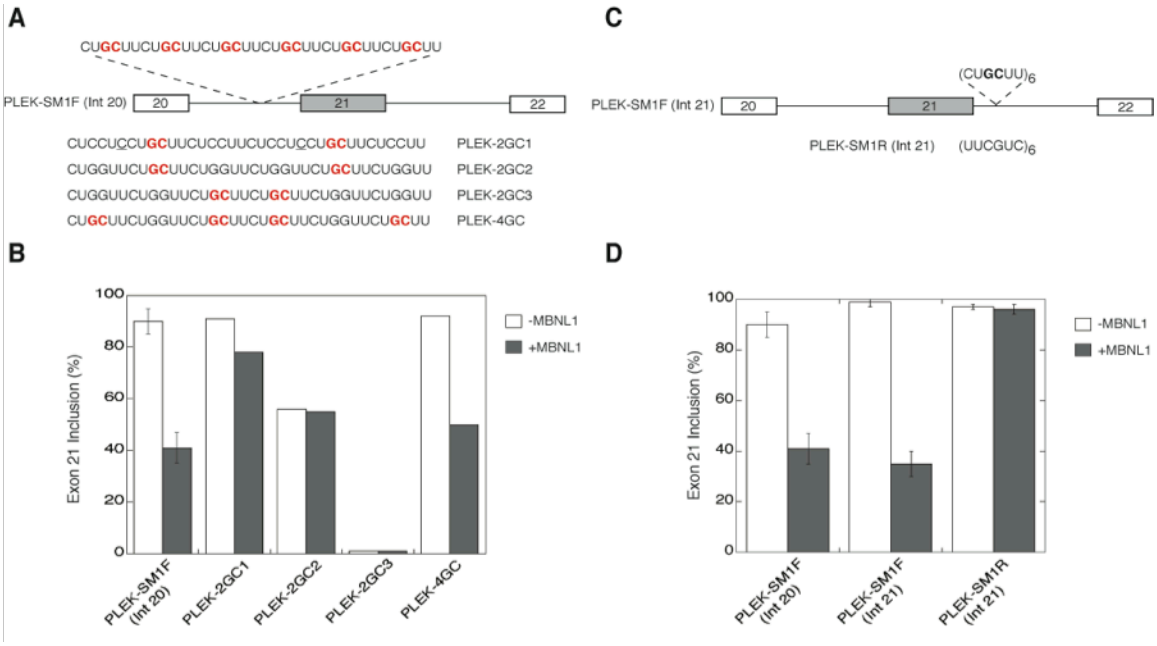


Figure 1. Splicing of synthetic PLEK Reporters. (A) Schematic of PLEK-2GC and PLEK-4GC reporters in comparison to reference reporter, PLEK-SM1F (Int 20). Sequences for the inserted, 36-nucleotide motif are provided under schematic. Positions of GC motifs are highlighted in red. (B) Splicing of PLEK-2GC and PLEK-4GC reporters with (grey bars) and without (white bars) MBNL1 over-expression as percent inclusions relative to the reference reporter. (C) Schematic of the PLEK-SM1F (Int 21) reporter wherein SM1 motif is inserted in PLEK intron 21. Motif inserted to create the PLEK-SM1R (Int 21) reporter is provided under schematic. (D) Splicing of PLEK Int21 reporters with (grey bars) and without (white bars) MBNL1 over-expression as percent inclusion relative to the Int 20 reference reporter.

With greater than 90% exon 21 inclusion, the PLEK-2GC1 and PLEK-4GC reporters both spliced with basal inclusion levels similar to the template PLEK-SM1F (Int 20) reporter. However the PLEK-2GC2 and PLEK-2GC3 reporters had significantly

reduced basal exon 21 inclusion levels in comparison to the PLEK-SM1F reporter (Int 20) (open bars in Figure 1B). Over-expression of MBNL1 caused a significant decrease in exon 21 inclusion in both the PLEK-2GC1 and PLEK-4GC reporters, however MBNL1 was unable to regulate splicing of the PLEK-2GC2 and PLEK-2GC3 reporters (shaded bars in Figure 1B).

MBNL1 Induces Exon Skipping when Binding Sites are Placed Downstream of Exon 21 in a Synthetic Reporter

To create the PLEK-SM1F (Int 21) reporter the 36-nt, SM1F motif (*i.e.* $(CUGCUU)_6$) was placed downstream of PLEK exon 21 in intron 21(Figure 1C). As demonstrated in Figure 5 of Chapter II, I wanted to ensure the necessity of YGCY motifs for MBNL1 activity, so I also inserted the SM1F motif with the orientation reversed so that the motif reads $(UUCGUC)_6$ (Figure 1C). Creation of the PLEK-SM1F (Int 21) reporter was complicated by the presence of a well-characterized intronic splicing enhancer (ISE) sequence within intron 21 of PLEK. This ISE is a potent splicing activation signal. Mutation of this ISE was previously shown (unpublished data) to reduce PLEK exon 21 inclusion. The goal in creating the PLEK-SM1F (Int 21) reporter was to create a splicing reporter that would be amenable to MBNL1-mediated inclusion. Therefore, I predicted that replacing the ISE with the SM1 motif would reduced the basal exon 21 inclusion levels of this reporter enough that we would be able to observe any MBNL1- mediated increases in exon inclusion that might result from placing binding site downstream of an exon. Unfortunately, basal splicing of the PLEK-SM1F (Int 21) reporter was 100%, which was actually slightly higher than the basal level observed for the PLEK-SM1F (Int 20) reporter (compare open bars in Figure 1D). Hence, replacing the ISE with the SM1 motif was not sufficient to reduce the exon inclusion level to a level amenable for assessing MBNL1-mediated inclusion. However, contrary to our prediction, we observed a decrease in exon 21 inclusion of PLEK-SM1F (Int 21) upon MBNL1 over-expression (shaded bars in Figure 1D) that is nearly identical to the decrease in exon inclusion measured for the upstream reporter.

DISCUSSION AND FUTURE DIRECTIONS

In comparison to the PLEK-SM1F (Int 20) reporter, which contains 6 GCs, decreasing the number of GC dinucleotides in the synthetic reporters to either 4GCs (i.e. PLEK-4GC) or 2GCs (i.e. PLEK-2GC1) resulted in a corresponding decrease in MBNL1-mediated splicing. MBNL1 regulated splicing of the PLEK-4GC reporter almost as well as the PLEK-SM1F (Int 20) reporter, which contains 6 GC dinucleotides. This result may suggest that not all 6 GCs are simultaneously occupied with MBNL1 in the original reporter. Based on biochemical characterization of synthetic oligos containing varying GC motifs, Cass et al. demonstrated that MBNL1 bound synthetic oligos with 2 GCs, 3GCs, or 4 GCs, surrounded by pyrimidines, with similar affinity (96). Therefore for these reporters it is somewhat surprising to see a correlation between the decrease in MBNL1-mediated exon suppression as the number of GC binding motifs decreases. Although contrary to biochemical observations, the decrease in MBNL1 activity might results from a decrease in MBNL1 affinity for the intronic motif as GCs are removed. Hence, reducing the number of MBNL1 recognition motifs in the transcript may result in less MBNL1 occupying a single transcript and subsequently lower activity.

Another interesting observation from these reporters is that the PLEK-2GC2 and PLEK-2GC3 reporters are not regulated by MBNL1 (Figure 1B). PLEK-2GC1 and PLEK-2GC2 are identical in length and have identically spaced GC dinucleotides, however they vary in sequence at the positions of GC mutation. In the case of the PLEK-2GC1 reporter CC motifs were substituted for GC motifs, however in PLEK-2GC2 reporter GG motifs replaced the GC motifs. It appears as though substitution of GG motifs in the PLEK-2GC2 reporter suppressed basal exon 21 inclusion levels (compare white bar for PLEK-2GC1 to that of PLEK-2GC2 in Figure 1B) and removed MBNL1's ability to regulate splicing of this reporter (compare white bar to grey bar for PLEK-2GC2 in Figure 1B). The most surprising splicing result from these reporters came from the PLEK-2GC3 reporter, which contains a pair of GCs separated by only 4 nucleotides. For this reporter exon 21 inclusion was drastically reduced (less than 5%) and regulation by MBNL1 was lost. One possible explanation for these interesting observations is that in the PLEK-2GC2 and PLEK-2GC3 reporters the GG motifs are spaced closely enough to interact with each other resulting in structures (such as G-quadruplexes (117)) that

inhibitory to splicing because they reduce the ability for the core splicing machinery to associate with the 3'-splice site and block MBNL1 from binding the GC dinucleotides.

In designing these preliminary reporters the spacing between GC dinucleotides was somewhat arbitrary. It seems apparent from this data that the spacing of GCs could be the focus of an entire study dedicated to understanding the spacing requirements for regulation via MBNL1. Such studies would also be useful in determining if specificity for Mbl versus MBNL1 could be obtained by simply changing the orientation of GC dinucleotides relative to one another in the appropriate synthetic system.

Splicing of the PLEK-SM1F (Int 21) reporter demonstrated that placing MBNL1 binding sites downstream of an exon in a synthetic system if not sufficient to induce MBNL1-mediated exon inclusion. The observation that insertion of high affinity MBNL1 binding sites either upstream or downstream of PLEK exon 21 results in nearly identical splicing effects begs an interesting mechanistic questions: How does MBNL1 regulates splicing of these synthetic reporters? Several groups have demonstrated that inserting structured RNA sequence near splice sites is sufficient to inhibit splicing and cause exon skipping (reviewed in (114)). However inhibition of splicing is not observed without MBNL1 over-expression, so one would have to argue that MBNL1 binding to the SM1 motif near the 3'ss (in the case of PLEK-SM1F (Int 20)) or 5'ss (in the case of PLEK-SM1F (Int 21)) and induces the inhibitory RNA structure. Another simple explanation for MBNL1-mediated skipping in the context of both upstream and downstream reporter is that MBNL1 binding to the motif sterically blocks binding of core spliceosomal components. Therefore, exon skipping could be the result of simple steric competition between MBNL1 and the splicing machinery at splice sites leading to a reduction in exon 21 definition and subsequent skipping.

Although the mechanisms governing and RNA components necessary for MBNL1-mediated exon inclusion remain a mystery, these experiments indicate that binding site position is one of many components that are currently poorly understood. Enhancer sequences within MBNL1- regulated exons, or the lack of splicing signals within the exon or intron, or presence of other RNA-binding proteins sites within target introns all might contribute to the splicing enhancement activity of MBNL1.

MATERIALS AND METHODS

Cell culture methods, *in vivo* splicing assays, and RT-PCR experiments were conducted as described in Chapter II.

Construction of Synthetic Reporters. Mutant reporters were created using modified QuickChange (Agilent) mutagenic PCR techniques. All reporters were cloned into pcDNA3 using KpnI and XhoI sites flanking exons 20 and 22 as described in Chapter II. The sequences of the mutagenic primers are provided in Table 1.

Table 1. Primers Used to Clone Synthetic PLEK reporters.

Reporter	Sense Primer	Antisense Primer
PLEK-2GC1	5'-ATCGATCTCCTCCTGCTTCTCCTTCTCCTCCTGCTTCTCCTTTTAATTAAGGCACGTGGTATCTAACCTGTGGTG-3'	5'-TTAATTAAAAGGAGAAGCAGGAGGAGAAGGAGAAGCAGGAGGAGATCGATGATGACAGTCTCTCTTAACCCAAAAACG-3'
PLEK-2GC2	5'-CTGGTTCTGGTTCTGCTTCTGCTTCTGGTTCTGGTTGGCACGTGGTATCTAACCTGTGGT-3'	5'-AACCAGAACCAGAAGCAGAAGCAGAACCAGAACCAGGATGACAGTCTCTCTTAACCCAAAAACGTTTTGC-3'
PLEK-2GC3	5'-CTGGTTCTGCTTCTGGTTCTGGTTCTGCTTCTGGTTGGCACGTGGTATCTAACCTGTGGT-3'	5'-AACCAGAAGCAGAACCAGAACCAGAAGCAGAACCAGGATGACAGTCTCTCTTAACCCAAAAACGTTTTGC-3'
PLEK-4GC	5'-CTGCTTCTGGTTCTGCTTCTGCTTCTGGTTCTGCTTGGCACGTGGTATCTAACCTGTGGT-3'	5'-AAGCAGAACCAGAAGCAGAAGCAGAACCAGAAGCAGGATGACAGTCTCTCTTAACCCAAAAACGTTTTGC-3'

APPENDIX B

FUNCTIONAL CHARACTERIZATION OF ADDITIONAL MBNL1 MUTANTS

This appendix contains experiments regarding a subset of unique, MBNL1 mutants. These mutants were created for the study described in Chapter III, but were not included in the final manuscripts due to spatial limitations.

As previously mentioned, MBNL1 has four ZFs of the $CX_7CX_{4-6}CX_3H$-type arranged into two tandem pairs of two ZFs. In the first pair (*i.e.* ZF1-2), ZF1 is separated from ZF2 by a 14 amino acid linker. In the second pair (*i.e.* ZF3-4), ZF3 is separated from ZF4 by a 16 amino acid linker (60). The C-C-C-H type of ZF was originally structurally characterized within the TIS11 protein family (118). TIS11 proteins play a critical role in the inflammatory response pathway by controlling the lifetime of certain genes. These proteins function by binding to AU-rich elements in the 3'-untranslated regions of some response genes and promoting deadenylation and subsequent messenger RNA turnover via degradation (119)-(122).

Tis11d is a specific member of this family and contains two, tandem $CX_8CX_5CX_3H$ ZFs separated by an 18 amino acid linker. Hudson et al. structurally characterized the ZFs of Tis11d, bound to a 9-nt AU-rich RNA via NMR (123). This structure demonstrated that the two ZFs of Tis11d bind to adjacent sites on the single stranded RNA substrate in a symmetrical fashion and that the linker between the two ZFs is largely unstructured (Figure 1A) (123). Classically, the structure of ZF proteins is defined as having multiple, small, individually folded ZF domains separated by amino acid linkers such that there is no significant higher order structure outside of the ZF domains. This structural organization has been coined the beads-on-a-string configuration and is typical of ZF proteins (124), including Tis11d.

Although the overall Tis11d protein structure is characteristic of most ZF proteins, the Tis11d structure provided a novel, sequence specific mechanism for ZF association with RNA via hydrophobic intercalation of aromatic residues between dinucleotide bases (Figure 1A) (125). The crystal structure obtained with the ZF3-4 domain of MBNL1 bound to a 6-nt RNA ligand demonstrated that MBNL1 interacts with

the Watson-Crick face of the 5'-GC-3' dinucleotide (Figure 1B) (61). This was a substantial observation because MBNL1 interacts with single stranded RNA in the same manner that Tis11d interacts with the 5'-AU-3' dinucleotide (123). Therefore, MBNL1 and Tis11d recognize RNA substrates through nearly identical ZF interactions.

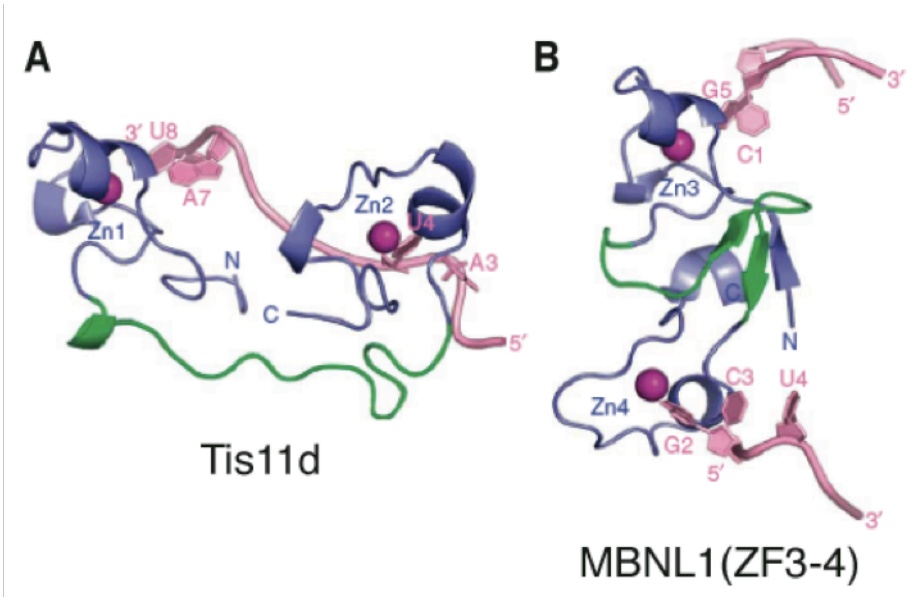


Figure 1. Structural comparison of MBNL1 and Tis11d linkers (61). (A) Ribbon structure of Tis11d (blue) with RNA substrate (pink) bound. Unstructured inker between ZF1 and ZF2 is represented in green. (B) Ribbon structure of ZF3-4 of MBNL1 (blue) with RNA substrate bound (pink). Structured**,** b-sheet linker between ZF3 and ZF4 is also represented in green.

Structure based alignments between the two pairs of ZFs in MBNL1 and the single ZF pair of Tis11d show a large degree of structural similarity between the ZF domains (61). However, MBNL1 has a structural attribute that makes it unique from Tis11d and other ZF proteins. The 14 and 16 amino acid linkers between ZF1 and ZF2 and ZF3 and ZF4, respectively, are highly structured in MBNL1 and adopt a previously uncharacterized symmetrical fold (61). The structured linkers consist of β-sheets, which orient the ZFs anti-parallel to one another creating a higher ordered, ZF-β-sheet-ZF, domain (Figure 1B). Teplova and Patel hypothesized that the anti-parallel orientation suggests that the ZF domains of MBNL1 do not bind adjacent dinucleotide motifs, such

as is the case with Tis11d, but rather bind motifs that are spaced far apart or on two distinct strands of RNA(61). I was interested in investigating the functional role of these structured linkers in regards to splicing regulation. To this end I used domain-swapping techniques to create a series of MBNL1 mutants wherein either the ZF1-2 linker, or ZF3-4 linker, or both was replaced with the unstructured 18 amino acid Tis11d linker.

In Chapter III I described and functionally characterized MBNL1 RNA Interaction Mutants (RIMs) wherein one of the four (*i.e.* single RIMs) or two of the four (*i.e.* double RIMs) ZFs was mutated to reduce RNA binding abilities. I also created a subset of MBNL1 mutant wherein three of the four ZFs were mutated to reduce the RNA interaction. These mutants are referred to as triple RIMs. The triple RIMs were evaluated using *in vitro* splicing assays to assess their ability to regulate splicing. As a result of the extensive mutation within this subset of mutants, each triple RIM only retains one wild-type ZF however the spacing and orientation of the ZFs with regards to one another is maintained. Therefore, I was interested in using these mutants to dissect the individual contribution of each ZF in regards to splicing function.

RESULTS

Substitution of Structured ZF Linkers Affects MBNL1 Activity

To create the MBNL1-Tis mutants, the 14 and 16 amino acid linkers between ZF1 and ZF2 and between ZF3 and ZF4, respectively, of MBNL1 were replaced with the 18 amino acid sequence of the Tis11d ZF linker. For clarity, Tis11d only has one pair of zinc fingers and therefore one linker, so for all mutants the MBNL1 linkers were replaced by the same Tis11d sequence. A schematic depicting the Tis-linker mutants can be found in Figure 2A. The MBNL1-Tis linker mutants were expressed in HeLa cells and Western blots were used to evaluate expression level (Figure 2B).

The splicing activity of these mutants was evaluated using the ATP2A1, TNNT2, and *MBNL1* minigenes previously described in Chapter III. Both the Tis1-2 and Tis3-4 mutants had near wild-type activity (92% and 94%, respectively) for the TNNT2 minigene (Figure 2D). This observation is completely consistent with splicing trends that were observed with other MBNL1 mutants on this reporter. However, the Tis1-2/3-4 mutant, which has both linkers mutated, had significantly reduced function at only 39%

of wild-type activity. ATP2A1 splicing resulted in a different splicing pattern. For this minigene, the Tis1-2 mutant, which retains a wild-type ZF3-4 domain, had the most activity at 72% and the Tis 3-4 mutant retained 60% activity (Figure 2C). Similar to what was observes with TNNT2, the Tis1/2-3/4 double mutant has the least activity for ATP2A1 at only 18%. The splicing trends for the *MBNL1* minigene also echoed those that seen with other mutants. Specifically the Tis 3-4 mutant, which retains a wild-type ZF1-2, has nearly wild-type activity (96%), whereas the Tis 1-2 mutant, with a wild-type ZF3-4, has only 63% activity (Figure 2E). Similar to both ATP2A1 and TNNT2, the Tis1-2/3-4 mutant had the least activity for *MBNL*1 at only 16%.

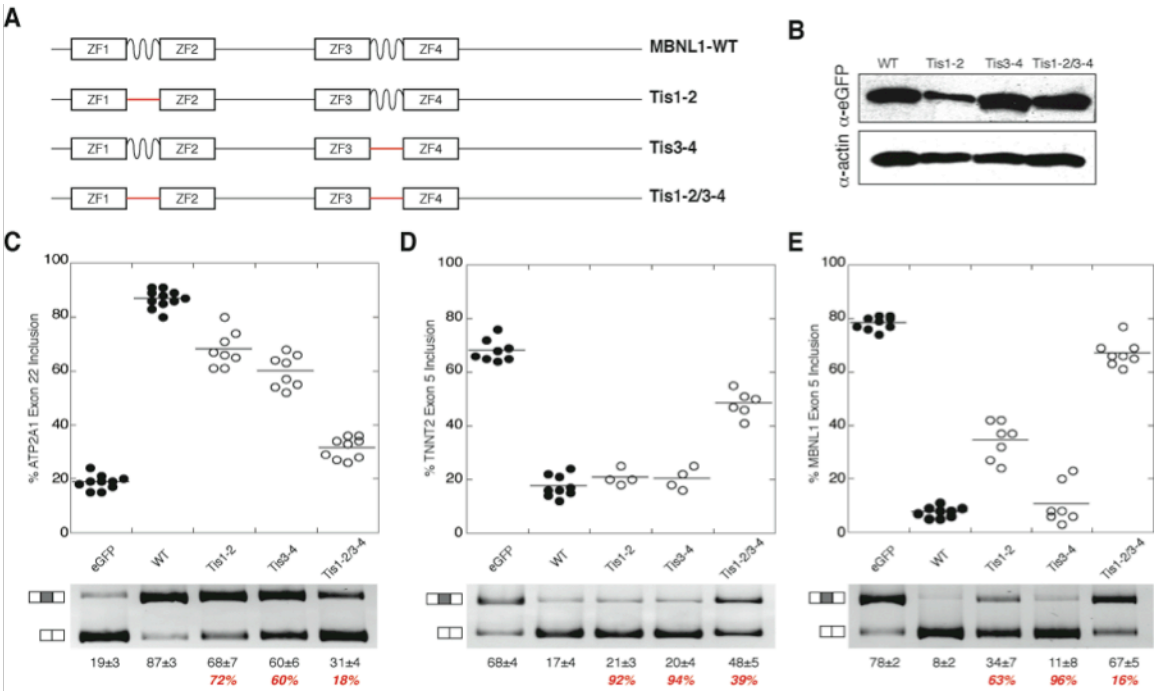


Figure 2. Tis Linker Mutants. (A) Schematic depiction of MBNL1-Tis mutants in comparison to wild-type protein. Solid red lines indicate positions where unstructured Tis11d linker domain was substituted for structured MBNL1 linker (black, wavy line). (B) Western blot comparing expression levels of Tis mutants to wild-type MBNL1 in HeLa cells. (C), (D), and (E) Dot plots of ATP2A1, TNNT2, and *MBNL1* splicing with Tis mutants, respectively. Values for percent exon inclusion and percent activity (red) are provided below representative gels in Panels C-E. All values were obtained from at least 4, independent experiments.

Triple RIM Mutants Cannot Regulate Splicing of Class I Minigenes.

For this subset of MBNL1 mutants, each of the three mutated ZF contains two point mutations, for a total of six amino acid substitutions per protein. The specific substitutions are provided in schematic form in Figure 3A. Western blot analysis of triple RIMs expressed in HeLa cells showed that the mutations did not have significant effect on protein expression levels (Figure 3B). Splicing of both the ATP2A1 (Figure 3C) and TNNT2 (Figure 3D) minigene reporters was measured with all four triple RIMs. For both reporters the splicing activity of the triple RIMs was greatly reduced in comparison to wild-type protein (Figure 3C and 3D). The 1,2,3RIM, which retains only a wild-type ZF4, was able to very modestly regulate splicing of both ATP2A1 and TNNT2 with 34% and 12% activity, respectively. With less than 12% activity 1,2,4RIM, 1,3,4RIM, and 2,3,4RIM are essentially unable to regulate splicing.

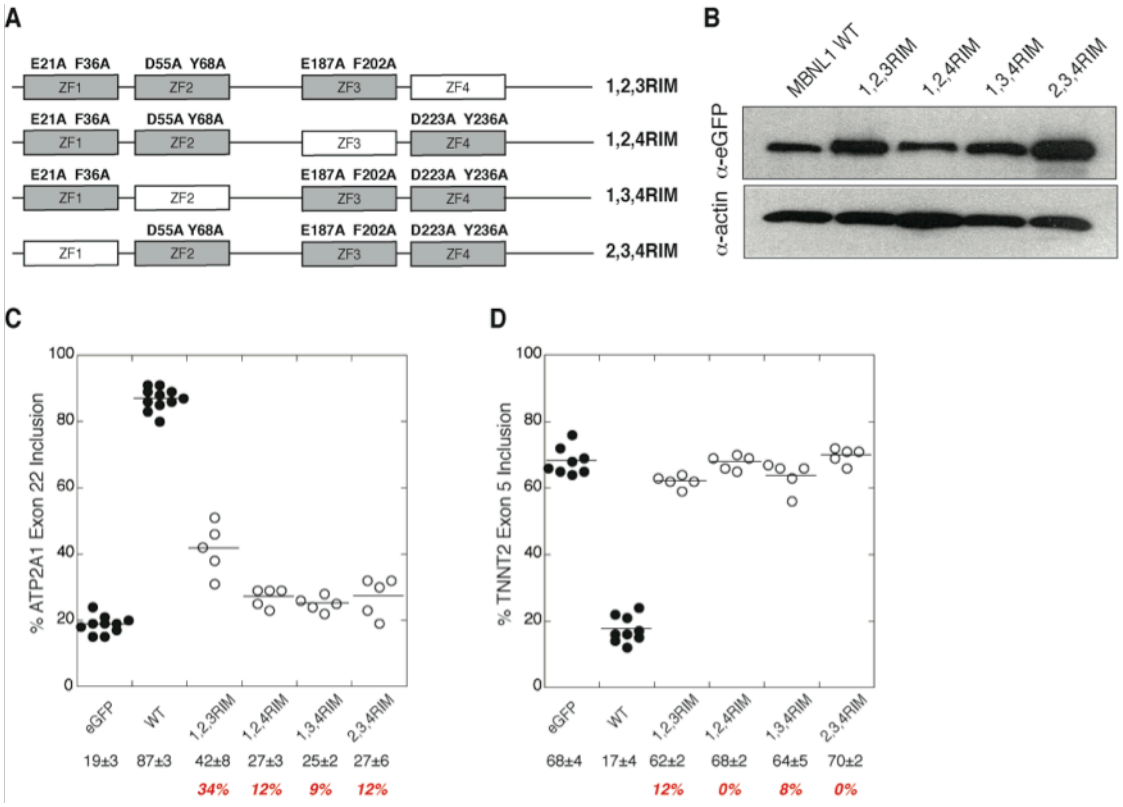


Figure 3. Triple ZF RNA Interaction Mutants (tRIMs). (A) Schematic of triple ZF RIMs. Mutated residues are denoted above shaded boxes. (B) Western blot comparing expression levels of triple RIMs to wild-type MBNL1 in HeLa cells. (C) and (D) Dot plots of ATP2A1 and TNNT2 splicing with single RIMs. Exon inclusion (%) and percent splicing activity (red) values are listed below representative splicing gels in Panels C and D. Values were obtained from at least 5, independent splicing experiments.

DISCUSSION

Tis Mutants Function Similarly to CMs and RIMs

In general the splicing activities that we measured for the Tis linker mutants echoed the splicing results that we obtained with other mutants. More specifically, for the ATP2A1 and TNNT2 minigenes the splicing activities for all three Tis linker mutants were similar to what was observed for the corresponding CM and RIM mutants. For these Class I minigenes, mutations in either ZF1-2 or ZF3-4 resulted in similarly reduced function. The only Class II minigene examined was *MBNL1*. We observed a similar splicing trend for the Tis mutants with this minigene in that mutation in the ZF1-2 pair have larger affects on splicing activity than mutations in the ZF3-4 pair. Interestingly however, with 63% splicing activity the Tis 1-2 mutant has significantly higher splicing activity than either the 1,2CM (34%) or 1,2RIM (46%) for this minigene. This observation may suggest that ZF1 and ZF2 are still able to weakly interact with RNA in this mutant to facilitate a stronger ZF3-4 interaction.

At first glance, analysis of the Tis1-2/3-4 double mutant suggests that turning MBNL1 into a "beads on a string" type ZF protein inhibited its splicing activity, suggesting that the linkers are important for orienting the ZFs with respect to one another and are therefore vital for recognizing RNA substrates. It is possible that swapping the structured linker for an unstructured linker resulted in a global unfolding of the ZF pairs, similar to what we hypothesized to occur with a subset of the double ZF chelation mutants. However, in every case the double Tis linker mutant (*i.e.* 1-2/3-4) retained more splicing activity than the corresponding CMs (see Chapter II) wherein one ZF in each pair was mutated. The most drastic example is in the case of TNNT2 splicing, wherein the Tis 1-2/3-4 mutant retained 39% splicing activity (Figure 2D) however CMs that targeted one ZF in both pair (*i.e.* 1,3CM or 2,3CM) had less than 5% activity. This observation may suggest that the presence of Tis linkers does not affect the folding of the ZF domains as severely as the chelation mutations do, and that the orientation of the ZFs with respect to one another is necessary for high affinity interactions. Future work is needed to characterize the binding affinity of these Tis mutants for RNA substrates. Further examination of the biophysical and binding properties of these mutants may lead to understanding the role of the structured linker within MBNL1 ZF pairs.

Triple RNA Interactions Mutants (RIMs) Have Negligible Splicing Activity

The lack of triple RIM splicing activity observed for the ATP2A1 and TNNT2 reporters is an interesting finding because both of these are Class I-type splicing events. I demonstrated with other MBNL1 mutants in Chapter III that there is no correlation between splicing activity and binding affinity for minigenes of this class. Based on splicing activities of double RIM mutants we anticipated that although the triple RIM mutants would have greatly reduced RNA binding affinity they should be able to facilitate modest splicing activity, potentially via protein:protein interactions. It is possible that we were only able to observe significant splicing effects with some of the double RIM mutants on Class I minigenes due to over-expression of the mutants, which compensates for the low affinity binding. If this is the case then mutation of the additional ZF to create the triple RIMs most likely resulted in binding affinities that are so low that over-expression of the mutants no longer compensates for poor binding.

The double RIM splicing and binding data also suggested that ZF2 and ZF4 seemed to play a larger role in splicing activity than ZF1 or ZF3. The observation that 1,2,3RIM has moderate splicing activity on the ATP2A1 minigene supports the previous findings, however the triple RIM splicing provides no additional evidence that maintaining a wild-type ZF2 (*i.e.* 1,3,4RIM) is important for MBNL1 function. Based on these observations and what was previously demonstrated with the double RIMs, it would be very interesting to see if the triple RIMs could regulate splicing of any of the Class II minigenes (*i.e.* Vldlr, Nfix, *MBNL1* and INSR).

MATERIALS AND METHODS

Cell culture, *in vivo* splicing assays, RT-PCR, and Western blot methods were conducted as described in Chapter III.

Mutagenic cloning. Triple MBNL1 RIMs were created using the mutagenic PCR primers listed in Table 2 of Chapter III with double RIM proteins as templates. Tis11d mutant were cloned using wild-type MBNL1 as template. Mutations were installed using overlapping oligo PCR and standard techniques. The primers used to install the Tis11d linkers are: Tis1-2 (Fwd) 5'- CGCGACATCCTAAGTACAAAACAGAACTGTGCT TTGATTCATTGAAAGGCCGTTGCTCC- 3'; Tis1-2 (Rev) 5'- CAGTTCTGTTTTGTA

CTTAGGATGTCGCGTAAGTGACCGCAGCTCATGAAAAGGATGTGCAAATTTA CATTCCGTGTC-3'; Tis3-4 (Fwd) 5'- CTTACGCGACATCCTAAGTACAAAACAGA ACTGTGTATGGATTACATCAAAGGGAGATCG-3'; Tis3-4 (Rev) 5'- CAGTTCTGT TTTGTACTTAGGATGTCGCGTAAGTGACCGCAGCTCATGAAAAGGATGAGC AAACCGACAATCATTTTCTCC-3'. All resulting PCR products were digested and ligated into pEGFP-C1 as described in Chapter III.

REFERENCES CITED

1. Staley, J.P. and Guthrie, C. (1998) Mechanical Devices of the Spliceosome: Motors, Clocks, Springs, and Things. *Cell*, **92**, 315-326.

2. Black, D.L. (2003) Mechanisms of Alternative Pre-Messenger RNA Splicing. *Ann. Rev. Biochem.*, **72**, 291-336.

3. Zhou, Z., Licklider, L.J., Gygi, S.P. and Reed, R. (2002) Comprehensive proteomic analysis of the human spliceosome. *Nature*, **419**, 182-185.

4. Jurica, M.S. and Moore, M.J. (2003) Pre-mRNA Splicing: Awash in a Sea of Proteins. *Mol. Cell*, **12**, 5-14.

5. Burge, C.B., Tuschl, T. and Sharp, P.A. (1999) *The RNA World*. 2nd ed. Cold Spring Harb. Lab., Cold Spring Harbor, NY.

6. Iwata, H. and Gotoh, O. (2011) Comparative analysis of information contents relevant to recognition of introns in many species. *BMC Genomics*, **12**, 45.

7. Schwartz, S., Silva, J.o., Burstein, D., Pupko, T., Eyras, E. and Ast, G. (2008) Large-scale comparative analysis of splicing signals and their corresponding splicing factors in eukaryotes. *Gen. Res.*, **18**, 88-103.

8. Williams, K.P., Fujimoto, D.N. and Inoue, T. (1994) Two Universally Conserved Adenosines of the Group I Intron That Are Important for Self-Splicing but Not for Core Catalytic Activity. *J. Biochem.*, **115**, 126-130.

9. Konarska, M.M. and Sharp, P.A. (1986) Electrophoretic separation of complexes involved in the splicing of precursors to mRNAs. *Cell*, **46**, 845-855.

10. Das, R., Zhou, Z. and Reed, R. (2000) Functional Association of U2 snRNP with the ATP-Independent Spliceosomal Complex E. *Mol. Cell*, **5**, 779-787.

11. Michaud, S. and Reed, R. (1991) An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. *Genes Dev.*, **5**, 2534-2546.

12. Valcárcel, J., Gaur, R.K., Singh, R. and Green, M.R. (1996) Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA. *Science*, **273**, 1706-1709.

13. Konarska, M.M. and Sharp, P.A. (1988) Association of U2, U4, U5, and U6 small nuclear ribonucleoproteins in a spliceosome-type complex in absence of precursor RNA. *Proc. Natl. Acad. Sci.*, **85**, 5459-5462.

14. Lamond, A.I., Konarska, M.M., Grabowski, P.J. and Sharp, P.A. (1988) Spliceosome assembly involves the binding and release of U4 small nuclear ribonucleoprotein. *Proc. Natl. Acad. Sci.*, **85**, 411-415.

15. Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P. and Burge, C.B. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature*, **456**, 470 - 476.

16. Harris, N.L. and Senapathy, P. (1990) Distribution and consenus of branch point signals in eukaryotic genes: a computerized statistical analysis. *Nucleic Acids Res.*, **18**, 3015.

17. Zhang, X.H.F. and Chasin, L.A. (2004) Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev.*, **18**, 1241-1250.

18. Yeo, G. and Burge, C.B. (2004) Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals. *J. Comput. Biol.*, **11**, 377-394.

19. Sun, H. and Chasin, L.A. (2000) Multiple Splicing Defects in an Intronic False Exon. *Mol. Cell. Biol.*, **20**, 6414-6425.

20. Zhang, X.H.F., Kangsamaksin, T., Chao, M.S.P., Banerjee, J.K. and Chasin, L.A. (2005) Exon Inclusion Is Dependent on Predictable Exonic Splicing Enhancers. *Mol. Cell. Biol.*, **25**, 7323-7332.

21. Lam, B.J. and Hertel, K.J. (2002) A general role for splicing enhancers in exon definition. *RNA*, **8**, 1233-1241.

22. Wang, Z. and Burge, C.B. (2008) Splicing regulation: From a parts list of regulatory elements to an integrated splicing code. *RNA*, **14**, 802-813.

23. Busch, A. and Hertel, K.J. (2012) Evolution of SR protein and hnRNP splicing regulatory factors. *WIREs RNA*, **3**, 1-12, doi: 10.1002/wrna.100.

24. Graveley, B.R. (2000) Sorting out the complexity of SR protein functions. *RNA*, **6**, 1197-1211.

25. Hertel, K.J. (2008) Combinatorial Control of Exon Recognition. *J. Biol. Chem.*, **283**, 1211-1215.

26. Schaal, T.D. and Maniatis, T. (1999) Multiple Distinct Splicing Enhancers in the Protein-Coding Sequences of a Constitutively Spliced Pre-mRNA. *Mol. Cell. Biol.*, **19**, 261-273.

27. Witten, J.T. and Ule, J. (2011) Understanding splicing regulation through RNA splicing maps. *Trends Genet.*, **27**, 89-97.

28. Darnell, R.B. (2010) HITS-CLIP: panoramic views of protein–RNA regulation in living cells. *WIREs RNA*, **1**, 266-286, doi: 10.1002/wrna.3.

29. Forch, P., Puig, O., Kedersha, N., Martinez, C., Granneman, S., Seraphin, B., Anderson, P. and Valcarcel, J. (2000) The Apoptosis-Promoting Factor TIA-1 Is a Regulator of Alternative Pre-mRNA Splicing. *Mol. Cell*, **6**, 1089-1098.

30. Konig, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M. and Ule, J. (2010) iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.*, **17**, 909-915.

31. Xue, Y., Zhou, Y., Wu, T., Zhu, T., Ji, X., Kwon, Y.-S., Zhang, C., Yeo, G., Black, D.L., Sun, H. *et al.* (2009) Genome-wide Analysis of PTB-RNA Interactions Reveals a Strategy Used by the General Splicing Repressor to Modulate Exon Inclusion or Skipping. *Mol. Cell*, **36**, 996-1006.

32. Martinez-Contreras, R., Fisette, J.-F., Nasim, F.-u.H., Madden, R., Cordeau, M. and Chabot, B. (2006) Intronic Binding Sites for hnRNP A/B and hnRNP F/H Proteins Stimulate Pre-mRNA Splicing. *PLoS Biol.*, **4**, e21.

33. Fisette, J.-F.o., Toutant, J., Dugre-Brisson, S., Desgroseillers, L. and Chabot, B. (2010) hnRNP A1 and hnRNP H can collaborate to modulate 5' splice site selection. *RNA*, **16**, 228-238.

34. Jensen, K.B., Musunuru, K., Lewis, H.A., Burley, S.K. and Darnell, R.B. (2000) The tetranucleotide UCAY directs the specific recognition of RNA by the Nova K-homology 3 domain. *Proc. Natl. Acad. Sci.*, **97**, 5740-5745.

35. Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A. and Darnell, R.B. (2003) CLIP Identifies Nova-Regulated RNA Networks in the Brain. *Science*, **302**, 1212-1215.

36. Ule, J., Stefani, G., Mele, A., Ruggiu, M., Wang, X., Taneri, B., Gaasterland, T., Blencowe, B.J. and Darnell, R.B. (2006) An RNA map predicting Nova-dependent splicing regulation. *Nature*, **444**, 580-586.

37. Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X. *et al.* (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature*, **456**, 464-469.

38. Zhang, C., Zhang, Z., Castle, J., Sun, S., Johnson, J., Krainer, A.R. and Zhang, M.Q. (2008) Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. *Genes Dev.*, **22**, 2550-2563.

39. Yeo, G.W., Coufal, N.G., Liang, T.Y., Peng, G.E., Fu, X.-D. and Gage, F.H. (2009) An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat. Struct. Mol. Biol.*, **16**, 130-137.

40. Du, H., Cline, M.S., Osborne, R.J., Tuttle, D.L., Clark, T.A., Donohue, J.P., Hall, M.P., Shiue, L., Swanson, M.S., Thornton, C.A. *et al.* (2010) Aberrant alternative splicing and extracellular matrix gene expression in mouse models of myotonic dystrophy. *Nat. Struct. Mol. Biol.*, **17**, 187-193.

41. Goers, E.S., Purcell, J., Voelker, R.B., Gates, D.P. and Berglund, J.A. (2010) MBNL1 binds GC motifs embedded in pyrimidines to regulate alternative splicing. *Nucleic Acids Res.*, **38**, 2467-2484.

42. Barreau, C., Paillard, L., Mereau, A. and Osborne, H.B. (2006) Mammalian CELF/Bruno-like RNA-binding proteins: molecular characteristics and biological functions. *Biochimie.*, **88**, 515-525.

43. Lee, J.E. and Cooper, T.A. (2009) Pathogenic mechanisms of myotonic dystrophy. *Biochem. Soc. Trans.*, **37**, 1281 - 1286.

44. Brook, J.D., McCurrach, M.E., Harley, H.G., Buckler, A.J., Church, D., Aburatani, H., Hunter, K., Stanton, V.P., Thirion, J.P. and Hudson, T. (1992) Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell*, **69**, 385.

45. Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barcelo, J., O'Hoy, K. *et al.* (1992) Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. *Science*, **255**, 1253-1255.

46. Liquori, C.L., Ricker, K., Moseley, M.L., Jacobsen, J.F., Kress, W., Naylor, S.L., Day, J.W. and Ranum, L.P.W. (2001) Myotonic Dystrophy Type 2 Caused by a CCTG Expansion in Intron 1 of ZNF9. *Science*, **293**, 864-867.

47. Miller, J.W., Urbinati, C.R., Teng-umnuay, P., Stenberg, M.G., Byrne, B.J., Thornton, C.A. and Swanson, M.S. (2000) Recruitment of human muscleblind proteins to (CUG)n expansions associated with myotonic dystrophy. *EMBO J.*, **19**, 4439-4448.

48. Mankodi, A., Urbinati, C.R., Yuan, Q.-P., Moxley, R.T., Sansone, V., Krym, M., Henderson, D., Schalling, M., Swanson, M.S. and Thornton, C.A. (2001) Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2. *Hum. Mol. Genet.*, **10**, 2165-2170.

49. Fardaei, M., Larkin, K., Brook, J.D. and Hamshere, M.G. (2001) In vivo co-localisation of MBNL protein with DMPK expanded-repeat transcripts. *Nucleic Acids Res.*, **29**, 2766-2771.

50. Philips, A.V., Timchenko, L.T. and Cooper, T.A. (1998) Disruption of Splicing Regulated by a CUG-Binding Protein in Myotonic Dystrophy. *Science*, **280**, 737-741.

51. Osborne, R.J. and Thornton, C.A. (2006) RNA-dominant diseases. *Hum. Mol. Genet.*, **15**, R162-R169.

52. Cho, D.H. and Tapscott, S.J. (2007) Myotonic dystrophy: Emerging mechanisms for DM1 and DM2. *Biochim. Biophys. Acta.*, **1772**, 195-204.

53. Kalsotra, A., Xiao, X., Ward, A.J., Castle, J.C., Johnson, J.M., Burge, C.B. and Cooper, T.A. (2008) A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart. *Proc. Natl. Acad. Sci*, **105**, 20333-20338.

54. Jin, Y., Suzuki, H., Maegawa, S., Endo, H., Sugano, S., Hashimoto, K., Yasuda, K. and Inoue, K. (2003) A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG. *EMBO J.*, **22**, 905-912.

55. Warf, M.B. and Berglund, J.A. (2007) MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T. *RNA*, **13**, 2238 - 2251.

56. Artero, R., Prokop, A., Paricio, N., Begemann, G., Pueyo, I., Mlodzik, M., Perez-Alonso, M. and Baylies, M.K. (1998) The muscleblind Gene Participates in the Organization of Z-Bands and Epidermal Attachments of Drosophila Muscles and Is Regulated by Dmef2. *Dev. Biol.*, **195**, 131-143.

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

58. Goers, E.S., Voelker, R.B., Gates, D.P. and Berglund, J.A. (2008) RNA Binding Specificity of Drosophila Muscleblind. *Biochemistry*, **47**, 7284-7294.

59. Kino, Y., Mori, D., Oma, Y., Takeshita, Y., Sasagawa, N. and Ishiura, S. (2004) Muscleblind protein, MBNL1/EXP, binds specifically to CHHG repeats. *Hum. Mol. Genet.*, **13**, 495-507.

60. Pascual, M., Vicente, M., Monferrer, L. and Artero, R. (2006) The Muscleblind family of proteins: an emerging class of regulators of developmentally programmed alternative splicing. *Differentiation*, **74**, 65-80.

61. Teplova, M. and Patel, D.J. (2008) Structural insights into RNA recognition by the alternative-splicing regulator muscleblind-like MBNL1. *Nat. Struct. Mol. Biol.*, **15**, 1343 - 1351.

62. Ranum, L.P.W. and Cooper, T.A. (2006) RNA-Mediated Neuromuscular Disorders. *Annu. Rev. Neurosci.*, **29**, 259-277.

63. Charlet-B, N., Savkur, R.S., Singh, G., Philips, A.V., Grice, E.A. and Cooper, T.A. (2002) Loss of the Muscle-Specific Chloride Channel in Type 1 Myotonic Dystrophy Due to Misregulated Alternative Splicing. *Mol. Cell*, **10**, 45-53.

64. Mankodi, A., Takahashi, M.P., Jiang, H., Beck, C.L., Bowers, W.J., Moxley, R.T., Cannon, S.C. and Thornton, C.A. (2002) Expanded CUG Repeats Trigger Aberrant Splicing of ClC-1 Chloride Channel Pre-mRNA and Hyperexcitability of Skeletal Muscle in Myotonic Dystrophy. *Mol. Cell*, **10**, 35-44.

65. Savkur, R.S., Philips, A.V. and Cooper, T.A. (2001) Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy. *Nat. Genet.*, **29**, 40-47.

66. Ho, T.H., Charlet-B, N., Poulos, M.G., Singh, G., Swanson, M.S. and Cooper, T.A. (2004) Muscleblind proteins regulate alternative splicing. *EMBO J.*, **23**, 3103-3112.

67. Warf, M.B., Diegel, J.V., von Hippel, P.H. and Berglund, J.A. (2009) The protein factors MBNL1 and U2AF65 bind alternative RNA structures to regulate splicing. *Proc. Natl. Acad. Sci*, **106**, 9203-9208.

68. Bartel, D.P., Zapp, M.L., Green, M.R. and Szostak, J.W. (1991) HIV-1 rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA. *Cell*, **67**, 529-536.

69. Faustino, N.A. and Cooper, T.A. (2005) Identification of Putative New Splicing Targets for ETR-3 Using Sequences Identified by Systematic Evolution of Ligands by Exponential Enrichment. *Mol. Cell. Biol.,* **25**, 879-887.

70. Kimura, T., Nakamori, M., Lueck, J.D., Pouliquin, P., Aoike, F., Fujimura, H., Dirksen, R.T., Takahashi, M.P., Dulhunty, A.F. and Sakoda, S. (2005) Altered mRNA splicing of the skeletal muscle ryanodine receptor and sarcoplasmic/endoplasmic reticulum Ca^{2+}-ATPase in myotonic dystrophy type 1. *Hum. Mol. Genet.*, **14**, 2189-2200.

71. Hino, S.-i., Kondo, S., Sekiya, H., Saito, A., Kanemoto, S., Murakami, T., Chihara, K., Aoki, Y., Nakamori, M., Takahashi, M.P. *et al.* (2007) Molecular mechanisms responsible for aberrant splicing of SERCA1 in myotonic dystrophy type 1. *Hum. Mol. Genet.*, **16**, 2834-2843.

72. Dominski, Z. and Kole, R. (1993) Restoration of correct splicing in thalassemic pre-mRNA by antisense oligonucleotides. *Proc. Natl. Acad. Sci.*, **90**, 8673-8677.

73. Sierakowska, H., Sambade, M.J., Agrawal, S. and Kole, R. (1996) Repair of thalassemic human beta-globin mRNA in mammalian cells by antisense oligonucleotides. *Proc. Natl. Acad. Sci.*, **93**, 12840-12844.

74. Yuan, Y., Compton, S.A., Sobczak, K., Stenberg, M.G., Thornton, C.A., Griffith, J.D. and Swanson, M.S. (2007) Muscleblind-like 1 interacts with RNA hairpins in splicing target and pathogenic RNAs. *Nucleic Acids Res*, **35**, 5474 - 5486.

75. Michalowski, S., Miller, J.W., Urbinati, C.R., Paliouras, M., Swanson, M.S. and Griffith, J. (1999) Visualization of double-stranded RNAs from the myotonic dystrophy protein kinase gene and interactions with CUG-binding protein. *Nucleic Acids Res.*, **27**, 3534-3542.

76. Mooers, B.H.M., Logue, J.S. and Berglund, J.A. (2005) The structural basis of myotonic dystrophy from the crystal structure of CUG repeats. *Proc. Natl. Acad. Sci.*, **102**, 16626-16631.

77. Napierala, M. and Krzyzosiak, W.J. (1997) CUG Repeats Present in Myotonin Kinase RNA Form Metastable "Slippery" Hairpins. *J. Biol. Chem.*, **272**, 31079-31085.

78. Sobczak, K., de Mezer, M., Michlewski, G., Krol, J. and Krzyzosiak, W.J. (2003) RNA structure of trinucleotide repeats associated with human neurological diseases. *Nucleic Acids Res.*, **31**, 5469-5482.

79. Ding, Y. and Lawrence, C.E. (2003) A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Res.*, **31**, 7280-7301.

80. Ding, Y.E., Chan, C.Y. and Lawrence, C.E. (2005) RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble. *RNA*, **11**, 1157-1166.

81. Ladd, A. and Cooper, T. (2002) Finding signals that regulate alternative splicing in the post-genomic era. *Genome Biol.*, **3**, reviews0008.0001 - reviews0008.0016.

82. Murray, J.I., Voelker, R.B., Henscheid, K.L., Warf, M.B. and Berglund, J.A. (2008) Identification of motifs that function in the splicing of non-canonical introns. *Genome Biol.*, **9**, R97.

83. Agresti, A. and Coull, B.A. (1998) Approximate Is Better than "Exact" for Interval Estimation of Binomial Proportions. *The American Statistician*, **52**, 119-126.

84. Ho, T.H., Savkur, R.S., Poulos, M.G., Mancini, M.A., Swanson, M.S. and Cooper, T.A. (2005) Colocalization of muscleblind with RNA foci is separable from mis-regulation of alternative splicing in myotonic dystrophy. *J. Cell Sci.*, **118**, 2923-2933.

85. House, A.E. and Lynch, K.W. (2008) Regulation of alternative splicing: more than just the ABCs. *J. Biol. Chem.*, **283**, 1217 - 1221.

86. Fernandez-Costa, J.M., Llamusi, M.B., Garcia-Lopez, A. and Artero, R. (2011) Alternative splicing regulation by Muscleblind proteins: from development to disease. *Biol. Rev. Camb. Philos. Soc.*, **86**, 947-958.

87. Fardaei, M., Rogers, M.T., Thorpe, H.M., Larkin, K., Hamshere, M.G., Harper, P.S. and Brook, J.D. (2002) Three proteins, MBNL, MBLL and MBXL, co-localize in vivo with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. *Hum. Mol. Genet.*, **11**, 805-814.

88. Vicente-Crespo, M., Pascual, M., Fernandez-Costa, J.M., Garcia-Lopez, A., Monferrer, L.n., Miranda, M.E., Zhou, L. and Artero, R.D. (2008) *Drosophila* Muscleblind Is Involved in *troponin T* Alternative Splicing and Apoptosis. *PLoS ONE*, **3**, e1613.

89. Han, J. and Cooper, T.A. (2005) Identification of CELF splicing activation and repression domains in vivo. *Nucleic Acids Res.*, **33**, 2769-2780.

90. Sen, S., Talukdar, I., Liu, Y., Tam, J., Reddy, S. and Webster, N.J. (2010) Muscleblind-like 1 (Mbnl1) promotes insulin receptor exon 11 inclusion via binding to a downstream evolutionarily conserved intronic enhancer. *J. Biol. Chem.*, **285**, 25426 - 25437.

91. Grammatikakis, I., Goo, Y.-H., Echeverria, G.V. and Cooper, T.A. (2011) Identification of MBNL1 and MBNL3 domains required for splicing activation and repression. *Nucleic Acids Res.*, **39**, 2769-2780.

92. Gates, D.P., Coonrod, L.A. and Berglund, J.A. (2011) Autoregulated Splicing of muscleblind-like 1 (MBNL1) Pre-mRNA. *J. Biol. Chem.,* **286**, 34224-34233.

93. Fu, Y., Ramisetty, S.R., Hussain, N. and Baranger, A.M. (2011) MBNL1–RNA Recognition: Contributions of MBNL1 Sequence and RNA Conformation. *Chembiochem.*, **13**, 112-119.

94. Paul, S., Dansithong, W., Kim, D., Rossi, J., Webster, N.J.G., Comai, L. and Reddy, S. (2006) Interaction of musleblind, CUG-BP1 and hnRNP H proteins in DM1-associated aberrant IR splicing. *EMBO J.*, **25**, 4271-4283.

95. Suyama, M., Harrington, E.D., Vinokourova, S., von Knebel Doeberitz, M., Ohara, O. and Bork, P. (2010) A network of conserved co-occurring motifs for the regulation of alternative splicing. *Nucleic Acids Res.*, **38**, 7916-7926.

96. Cass, D., Hotchko, R., Barber, P., Jones, K., Gates, D.P. and Berglund, J.A. (2011) The four Zn fingers of MBNL1 provide a flexible platform for recognition of its RNA binding elements. *BMC Mol. Biol.*, **12**, 20.

97. Ladd, A.N., Charlet-B, N. and Cooper, T.A. (2001) The CELF Family of RNA Binding Proteins Is Implicated in Cell-Specific and Developmentally Regulated Alternative Splicing. *Mol. Cell. Biol.*, **21**, 1285-1296.

98. Kosaki, A., Nelson, J. and Webster, N.J.G. (1998) Identification of Intron and Exon Sequences Involved in Alternative Splicing of Insulin Receptor Pre-mRNA. *J. Biol. Chem.*, **273**, 10331-10337.

99. Modafferi, E.F. and Black, D.L. (1997) A complex intronic splicing enhancer from the c-src pre-mRNA activates inclusion of a heterologous exon. *Mol. Cell. Biol.*, **17**, 6537-6545.

100. de Hoon, M.J.L., Imoto, S., Nolan, J. and Miyano, S. (2004) Open source clustering software. *Bioinformatics*, **20**, 1453-1454.

101. Saldanha, A.J. (2004) Java Treeview--extensible visualization of microarray data. *Bioinformatics*, **20**, 3246-3248.

102. Lin, X., Miller, J.W., Mankodi, A., Kanadia, R.N., Yuan, Y., Moxley, R.T., Swanson, M.S. and Thornton, C.A. (2006) Failure of MBNL1-dependent post-natal splicing transitions in myotonic dystrophy. *Hum. Mol. Genet.*, **15**, 2087-2097.

103. Kim, D.-H., Langlois, M.-A., Lee, K.-B., Riggs, A.D., Puymirat, J. and Rossi, J.J. (2005) HnRNP H inhibits nuclear export of mRNA containing expanded CUG repeats and a distal branch point sequence. *Nucleic Acids Res.*, **33**, 3866-3874.

104. Dansithong, W., Paul, S., Comai, L. and Reddy, S. (2005) MBNL1 is the primary determinant of focus formation and aberrant insulin receptor splicing in DM1. *J. Biol. Chem.*, **280**, 5773 - 5780.

105. Graveley, B.R. and Maniatis, T. (1998) Arginine/Serine-Rich Domains of SR Proteins Can Function as Activators of Pre-mRNA Splicing. *Mol. Cell*, **1**, 765-771.

106. Philipps, D., Celotto, A.M., Wang, Q.Ä., Tarng, R.S. and Graveley, B.R. (2003) Arginine/serine repeats are sufficient to constitute a splicing activation domain. *Nucleic Acids Res.*, **31**, 6502-6508.

107. Carey, J., Cameron, V., De Haseth, P.L. and Uhlenbeck, O.C. (1983) Sequence-specific interaction of R17 coat protein with its ribonucleic acid binding site. *Biochemistry*, **22**, 2601-2610.

108. Graveley, B.R., Hertel, K.J. and Maniatis, T. (1998) A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. *EMBO J.*, **17**, 6747-6756.

109. Del Gatto-Konczak, F., Olive, M., Gesnel, M.-C. and Breathnach, R. (1999) hnRNP A1 Recruited to an Exon In Vivo Can Function as an Exon Splicing Silencer. *Mol. Cell. Biol.* , **19**, 251-260.

110. Gesnel, M.-C., Theoleyre, S., Del Gatto-Konczak, F. and Breathnach, R. (2007) Cooperative binding of TIA-1 and U1 snRNP in K-SAM exon splicing activation. *Biochem. Biophys. Res. Commun.*, **358**, 1065-1070.

111. Gesnel, M.-C., Del Gatto-Konczak, F. and Breathnach, R. (2009) Combined Use of MS2 and PP7 Coat Fusions Shows that TIA-1 Dominates hnRNP A1 for K-SAM Exon Splicing Control. *J. Biomed. Biotechnol.*, **2009**, Article ID 104853, doi:10.1155/2009/104853.

112. Culler, S.J., Hoff, K.G. and Smolke, C.D. (2010) Reprogramming Cellular Behavior with RNA Controllers Responsive to Endogenous Proteins. *Science*, **330**, 1251-1255.

113. Fukumura, K., Kato, A., Jin, Y., Ideue, T., Hirose, T., Kataoka, N., Fujiwara, T., Sakamoto, H. and Inoue, K. (2007) Tissue-specific splicing regulator Fox-1 induces exon skipping by interfering E complex formation on the downstream intron of human F1gamma gene. *Nucleic Acids Res.*, **35**, 5303-5311.

114. Warf, M.B. and Berglund, J.A. (2010) Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem. Sci.*, **35**, 169-178.

115. Schneider, D., Tuerk, C. and Gold, L. (1992) Selection of high affinity RNA ligands to the bacteriophage R17 coat protein. *J. Mol. Biol.*, **228**, 862-869.

116. Spellman, R. and Smith, C.W.J. (2006) Novel modes of splicing repression by PTB. *Trends Biochem. Sci.*, **31**, 73-76.

117. Huppert, J.L. (2008) Hunting G-quadruplexes. *Biochimie.*, **90**, 1140-1148.

118. Gomperts, M., Pascall, J.C. and Brown, K.D. (1990) The nucleotide sequence of a cDNA encoding an EGF-inducible gene indicates the existence of a new family of mitogen-induced genes. *Oncogene*, **5**, 1081-1083.

119. Caput, D., Beutler, B., Hartog, K., Thayer, R., Brown-Shimer, S. and Cerami, A. (1986) Identification of a common nucleotide sequence in the 3'-untranslated region of mRNA molecules specifying inflammatory mediators. *Proc. Natl. Acad. Sci.*, **83**, 1670-1674.

120. Shaw, G. and Kamen, R. (1986) A conserved AU sequence from the 3'-untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell*, **46**, 659-667.

121. Lai, W.S., Carballo, E., Thorn, J.M., Kennington, E.A. and Blackshear, P.J. (2000) Interactions of CCCH zinc finger proteins with mRNA. Binding of tristetraprolin-related zinc finger proteins to Au-rich elements and destabilization of mRNA. *J. Biol. Chem.* , **275**, 17827-17837.

122. Blackshear, P.J. (2002) Tristetraprolin and other CCCH tandem zinc-finger proteins in the regulation of mRNA turnover. *Biochem. Soc. Trans.*, **30**, 945-952.

123. Hudson, B.P., Martinez-Yamout, M.A., Dyson, H.J. and Wright, P.E. (2004) Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d. *Nat. Struct. Mol. Biol.*, **11**, 257-264.

124. Miller, J., McLachlan, A.D. and Klug, A. (1985) Repetitive zinc-binding domains in the protein transcriptional factor IIIA from Xenopus oocytes. *EMBO J.*, **4**, 1609-1615.

125. Brown, R.S. (2005) Zinc finger proteins: getting a grip on RNA. *Curr. Opin. Struct. Biol.*, **15**, 94-98.