CONTRIBUTION OF MUSCLEBLIND-LIKE1 (MBNL1) BINDING EVENTS TO DOSE-RESPONSE BEHAVIOR OF *MBNL1*

By RITI GUPTA

A THESIS

Presented to the Department of Chemistry
And The Honors College of the University of Oregon
In partial fulfillment of the requirements
For the degree of
Bachelor of Science

June 2013

An Abstract of the Thesis of

Riti Gupta for the degree of Bachelor of Science

in the Department of Chemistry to be taken May 29, 2013

Title: CONTRIBUTION OF MUSCLEBLIND-LIKE1 (MBNL1) BINDING EVENTS TO DOSE-RESPONSE BEHAVIOR OF *MBNL1*

Approved: _____
Dr. J. Andrew Berglund

Mis-regulation of the alternative splicing factor muscleblind-like 1 (MBNL1) plays a significant role in the disease myotonic dystrophy (DM). MBNL1 regulates alternative splicing of genes involved in development of skeletal muscle, heart and the central nervous system. In myotonic dystrophy the lack of properly localized MBNL1 leads to mis-splicing of many pre-mRNAs. One of the mis-spliced pre-mRNAs is the *MBNL1* pre-mRNA that codes for the MBNL1 protein (an auto-regulated event). Specifically, the mis-splicing is aberrant inclusion of exon 5 in the *MBNL1* pre-mRNA. Previous work has shown that intron 4 of the *MBNL1* gene is highly conserved and contains multiple MBNL1 binding sites. It has been shown that a 90-nucleotide region within intron 4 is necessary for regulation by MBNL1. This study investigates the sensitivity of the auto-regulated *MBNL1* splicing event by generating MBNL1 dose-response curves using HEK293 cells with an inducible MBNL1 expression system. Coupling this expression system with *MBNL1* constructs containing deletions of various MBNL1

binding sites within the important 90 nucleotide regulation region, it was shown that a single central MBNL1 site was fundamentally more important that other sites for splicing regulation.

ACKNOWLEDGEMENTS

I would like express sincere appreciation to both Professor Berglund and Dr. Stacey Wagner for their help in developing this project and for their assistance in preparing this manuscript. In addition, special thanks are due to the entire Berglund lab for their ideas and help throughout the process of working on this project. Finally, I would like to express the utmost love and appreciation to Mukta, Rakesh, and Lipi Gupta for their unstinting support.

TABLE OF CONTENTS

Background: The Central Dogma of Molecular Biology

The central dogma of molecular biology describes how genetic information given by DNA results in protein in biological systems. The ideas of the central dogma of molecular biology are all based on experimental evidence and represent a synthesis of the knowledge gained by many scientists over time. The term dogma was used in order to reflect how fundamental the core idea is to modern scientists.

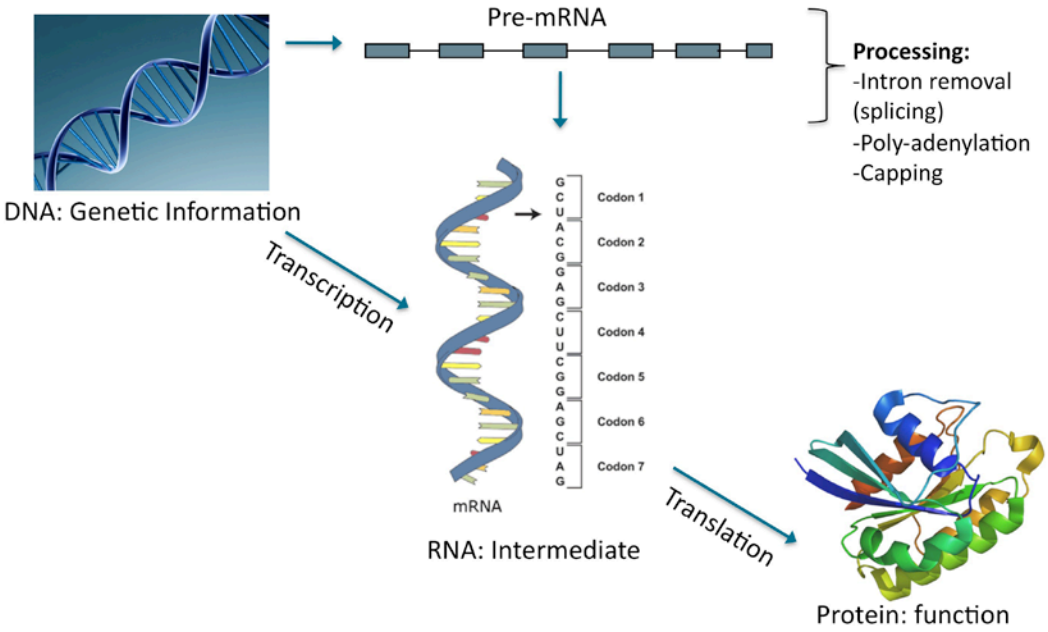

Figure 1. The Central Dogma of Molecular Biology.

DNA is the hereditary material in all organisms, including humans. DNA is packaged in structures called chromosomes, which are found in the nucleus of a cell. The information in DNA is given by a code that is made up on four chemical bases: adenine (A), cytosine (C), guanine (G) and thymine (T). The human genome consists of about 3 billion of these bases. Of these bases about 99 percent are the same in all humans. The order or sequence of these bases is what gives the information necessary for proper growth and function of an organism.

DNA bases pair with one another: A with T and C with G to form what are called base pairs. Each of the chemical bases is attached to a sugar and phosphate molecule. The base, sugar, phosphate unit is called a nucleotide and these nucleotides are arranged in two long anti-parallel strands (meaning that they have opposite polarity) that form a double-helix. The asymmetric ends of DNA are called 5' and 3' because the 5' end has a terminal phosphate group while the 3' has a terminal hydroxyl group.

While DNA is the total hereditary material, its basic physical unit is called a gene. Gene expression results in the formation of a protein that has a specific role in the cell. The first step of gene expression occurs when the DNA of a gene is copied into ribonucleic acid (RNA) through a process called transcription (Figure 1). In this process a complementary anti-parallel strand of RNA is formed which includes the same base pairs as DNA except that instead of thymine (T), RNA includes uracil (U). This transcript of RNA is termed pre-messenger RNA or pre-mRNA. This pre-mRNA undergoes different types of processing within the nucleus and is then moved to the cytoplasm. Here, the mRNA is translated into protein by the ribosome.

At it's most basic level the central dogma relates how a gene becomes a protein. Within this context, however, there are small changes that can effectively tune how the flow of genetic information works. For example, at any one time only a small portion of the DNA contained in any cell is being actively transcribed. Furthermore, the genes that are transcribed contain intervening sequences that provide for further change in the gene product (alternative splicing). This can reflect how different protein isoforms are being expressed either at different levels or in different tissues. At any level of the flow chart shown in Figure 1, expression of protein (and the form of the protein) can be controlled.

This study aims to understanding how protein isoform levels are varied at the alternative splicing level.

Background: Pre-mRNA processing

After the initial transcript is made from a gene (pre-mRNA) it must undergo processing in the nucleus. This includes the addition of a 5' cap, splicing, editing, and polyadenine tail (Poly(A) tail) at the 3' end. A typical mature RNA (mRNA) contains the 5' cap, 5' untranslated region (UTR), coding sequence, 3' UTR, and Poly(A) tail (Figure 2, 1).

Both the 5' UTR and 3' UTR are non-coding (not translated by the ribosome into protein) but are necessary for many reasons. They are important for mRNA localization, stability, and translational efficiency by the ribosome. Proteins can bind to either of the UTRs and either promote or inhibit translation.

The 5' cap is a modified guanine nucleotide and critical for proper attachment of the mRNA to the ribosome. It also functions as a protection from exonucleases. The 3' Poly(A) tail promotes export from the nucleus and promotes translation. Like the 5' cap, it also serves to protect the mRNA from degradation.

The mature coding sequence is formed through process known as splicing. In this process, introns are excised and exons are ligated together to form the mature RNA. This occurs because large parts of any gene are non-coding (introns). Exons are any sequence needed for encoding the final protein product. Alternative splicing allows for one gene to produce more than one type of RNA, which can encode for multiple different proteins, thus increasing proteomic diversity. About 92-94% of human multi-

exon genes are alternatively spliced (2). Of these at least 85% have a minor isoform that is expressed at a frequency of 15% (3).

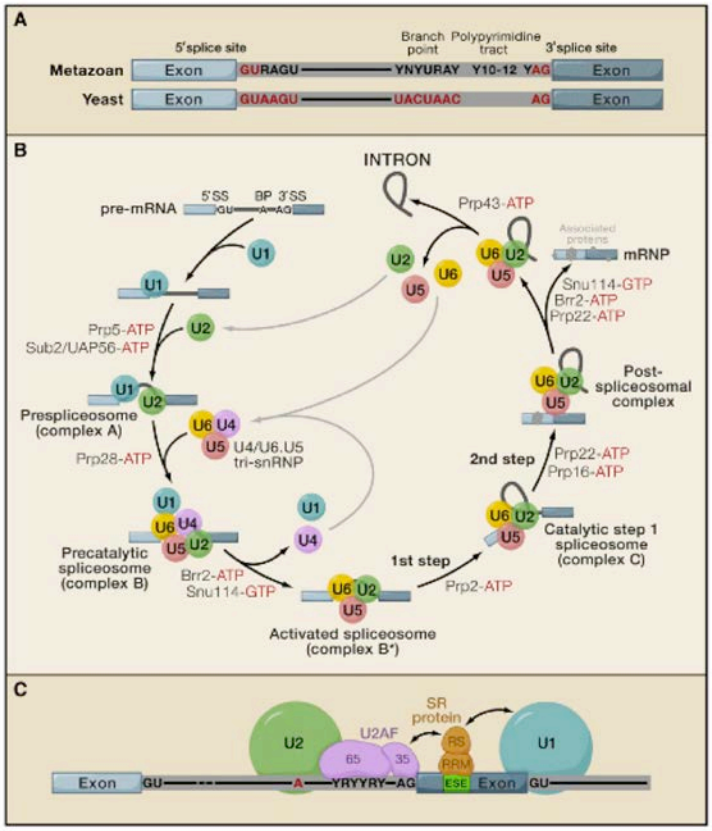

Figure 2. Pre-mRNA Splicing by the Major Spliceosome
(A) Conserved sequence elements of metazoan and yeast pre-mRNAs. Here, two exons (blue) are separated by an intron (gray). The consensus sequences in metazoans and yeast at the 5′ splice site (SS), branch point sequence (BPS), and 3′ splice site (SS) are as indicated, where N is any nucleotide, R is a purine, and Y is a pyrimidine. The polypyrimidine tract is a pyrimidine-rich stretch located between the BPS and 3′SS.
(B) Cross-intron assembly and disassembly cycle of the major spliceosome. The stepwise interaction of the spliceosomal snRNPs (colored circles), but not non-snRNP proteins, in the removal of an intron from a pre-mRNA containing two exons (blue) is depicted. Only the spliceosomal complexes that can be resolved biochemically in mammalian splicing extracts are shown. Eight evolutionarily conserved DExD/H-type RNA-dependent ATPases/helicases act at specific steps of the splicing cycle to catalyze RNA-RNA rearrangements and RNP remodeling events. These enzymes include Sub2 (UAP56 in humans), Prp5, Prp28, Brr2, Prp2, Prp16, Prp22, and Prp43 (with Brr2 and Prp22 acting at more than one step in the cycle). The GTPase Snu114 also functions at several steps during the cycle. In yeast, Prp28 acts at a later stage during spliceosome activation (the B complex to B* complex transition) (Staley and Guthrie, 1998). Several of these proteins, such as Prp5, Prp16, and Prp22, also carry out proofreading functions at the stages where they are shown.
(C) Cross-exon splicing complexes form on long introns during the earliest stage of spliceosome assembly. An SR protein containing an arginine-serine-rich (RS) domain and RRM (RNA recognition motif) is depicted as interacting with an exonic splicing enhancer (ESE). The U1 (blue) and U2 (green) spliceosomal snRNPs and the two subunits of the U2 auxilliary factor (U2AF), U2AF65 and U2AF35, are also shown interacting with the splice sites flanking the exon.

Splicing of pre-mRNA is performed by an RNA and protein complex that is known as the spliceosome (2). Different trans-acting proteins known as repressors and activators, cis-acting regulatory sites (enhancers and silencers) and other RNA features such as secondary structure determine how splicing is regulated and how it will occur based on the cellular environment.

Pre-mRNA splicing is dependent upon the recognition of four canonical motifs (5’ splice site, branch point sequence, polypyrimidine tract (PY tract), and the 3’ splice site). Both the 3’ splice site and 5’ splice site have consensus sequences but these do not

contain enough information to always define which exons must be included or excluded. For this reason non-canonical splicing signals are also important as they act to recruit the spliceosome and other splicing factors (4).

The spliceosome itself is a large, complex, and not entirely understood molecular machine made of both small RNAs and proteins. The human splicing machinery contains many small nuclear RNAs (U1, U2, U4, U5, and U6) and more than 300 proteins (5). Once the intron is recognized by the spliceosome, splicing occurs in two straight forward steps. First the 2' hydroxyl of the branchpoint adenosine attacks the 5' splice junction which results in breakage of that phosphodiester bond. Simultaneously, a 2'-5' phosphodiester linkage between the branch point and the 5' terminal nucleotide of the intron is formed. At the end of the first step the intron is the form of a lariat. In the next step the released 3' hydroxyl of the 5' exon attacks the 3' splice junction breaking the phosphodiester bond while forming a new phosphodiester bond between the 5' exon and the 3' exon. The intron is then released in the lariat form (6).

There are many different types of alternative splicing, which include mutually exclusive exons, alternative 5' splice sites, alternative 3' acceptor sites, intron retention, and cassette exon skipping (Figure 1). Exon skipping is the most common type of splicing in mammals (7). In exon skipping, an exon may either be excised out of the pre-mRNA or retained. Often whether or not an exon is included depends on the stage of development or whether or not certain regulatory factors are available. This is the

type of splicing focused on in the present study.

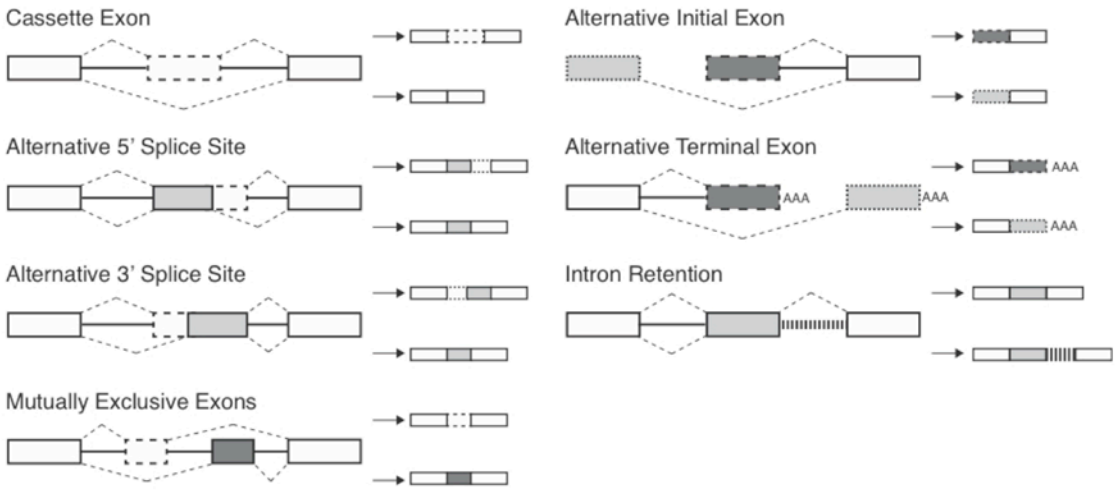

Figure 3. Types of alternative splicing.

Background: Human Genetic Diseases

Genes code for proteins, which serve a many functions in the cell. When a mutation occurs in the DNA of a gene, human genetic diseases arise. The human genome contains between 20,000 to 25,000 genes. Single-gene diseases are the result of a mutation in just one gene.

Humans have 23 pairs of chromosomes. One chromosome of each pair is inherited from each parent. Each member of the homologous pair has the same genetic material and is located at the same physical positions on the chromosome. Each pair of chromosomes carries specific genes and of those genes can be slightly different, which is defined as alleles. If both parents contribute the same allele to the offspring then the offspring is homozygous for that allele. If the parents contribute different alleles then the offspring is a heterozygote.

The interaction between the two alleles can be defined as either recessive or dominant. Which ever parent the heterozygous offspring is phenotypically identical to is the parent that contributed the dominant allele. Thus diseases can be either dominant (inherited from one parent) or recessive (must be inherited from both) and either autosomal (not on a sex chromosome) or sex-linked (on a sex chromosome).

There are many single-gene diseases. Diseases such as phenylketonuria and cystic fibrosis are autosomal recessive, meaning the child inherited an allele from each parent that leads to the disease phenotype. Other diseases such as Huntington's and myotonic dystrophy type 1 are autosomal dominant, meaning that one parent has the disease and it has been passed on to the child. The purpose of this project is to examine the aberrant alternative splicing that plays a key role in causing myotonic dystrophy type 1.

Background: Myotonic Dystrophy Type 1 (DM1) and Muscleblindlike-1 (MBNL1)

Nearly 20 developmental and degenerative diseases caused by the expansion of unstable repeats are known today (8). Among these there are three classes of disorders: those that are caused by expansions of non-coding repeats and results in the loss of protein function, those that are causing by coding repeats and result in altered protein function and those that are caused by expansions of non-coding repeats which results in altered RNA function (Table 1).

Disease	Mutation/ repeat unit	Gene name (protein product)	Putative function	Normal repeat length	Pathogenic repeat length
Diseases that are caused by loss of protein function					
FRDA	$(GAA)_n$	*FRDA* (frataxin)	Mitochondrial iron metabolism	6–32	200–1,700
FRAXA	$(CGG)_n$	*FMR1* (FMRP)	Translational regulation	6–60	>200 (full mutation)
FRAXE	$(CCG)_n$	*FMR2* (FMR2)	Transcription?	4–39	200–900
Diseases that are caused by altered protein function					
SCA1	$(CAG)_n$	*SCA1* (ataxin 1)	Transcription	6–39	40–82
SCA2	$(CAG)_n$	*SCA2* (ataxin 2)	RNA metabolism	15–24	32–200
SCA3 (MJD)	$(CAG)_n$	*SCA3* (ataxin 3)	De-ubiquitylating activity	13–36	61–84
SCA6	$(CAG)_n$	*CACNA1A* ($CACNA1_A$)	P/Q-type α1A calcium channel subunit	4–20	20–29
SCA7	$(CAG)_n$	*SCA7* (ataxin 7)	Transcription	4–35	37–306
SCA17	$(CAG)_n$	*SCA17* (TBP)	Transcription	25–42	47–63
DRPLA	$(CAG)_n$	*DRPLA* (atrophin 1)	Transcription	7–34	49–88
SBMA	$(CAG)_n$	*AR* (androgen receptor)	Steroid-hormone receptor	9–36	38–62
HD	$(CAG)_n$	*HD* (huntingtin)	Signalling, transport, transcription	11–34	40–121
Diseases that are caused by altered RNA function					
DM1	$(CTG)_n$	*DMPK* (DMPK)	RNA-mediated	5–37	50–1,000
DM2	$(CCTG)_n$	*ZNF9* (ZNF9)	RNA-mediated	10–26	75–11,000
FXTAS	$(CGG)_n$	*FMR1* (FMRP)	RNA-mediated	6–60	60–200 (premutation)
Diseases of unknown pathogenic mechanism(s)					
SCA8	$(CTG)_n$	*SCA8* (transcribed/ untranslated)	Unknown	16–34	>74
SCA10	$(ATTCT)_n$	Unknown	Unknown	10–20	500–4,500
SCA12	$(CAG)_n$	*PPP2R2B* (PPP2R2B)	Phosphatase regulation	7–45	55–78
HDL2	$(CTG)_n$	*JPH3* (junctophilin 3)	PM/ER junction protein	7–28	66–78

Table 1. Repeat expansion diseases (8)

Myotonic dystrophy type 1 is associated with a toxic gain-of-function by a CUG repeat expansion in the 3’ untranslated region (UTR) of the dystrophin myotonin protein kinase (*DMPK*) gene on chromosome 19 (9,10). Unaffected individuals have between 5 and 37 CTG repeats. Individuals with 50 repeats can start to exhibit symptoms. The severity of the disease is correlated with repeat length. With more than 1500 repeats the result is often a severe congenital form. DM affects many different organ systems and has many major symptoms that include progressive skeletal wasting, impaired muscle relaxation, cardiac conduction defects resulting in arrhythmias, iridescent cataracts, insulin insensitivity, and others (11).

The CUG repeats are proposed to be toxic primarily through their action sequestering the RNA binding protein, muscleblind-like 1 (MBNL1). This protein is important because it regulates alternative splicing of genes involved in development of skeletal muscle, the heart and the central nervous system. In DM1 the expanded CUG repeats co-localize with MBNL1 to form foci that make the active protein unavailable (12). Furthermore, with increased CUG repeats more MBNL1 is sequestered. A knockout of *MBNL1* in mice produces a myotonic myopathy similar to DM1 along with DM1-like cataracts and cardiac disease, showing that MBNL1 sequestration plays a key role in DM1 (3). For example, the reduced concentrations of MBNL1 lead to the mis-splicing of the chloride challen (*CLCN1),* which leads to the characteristic myotonia

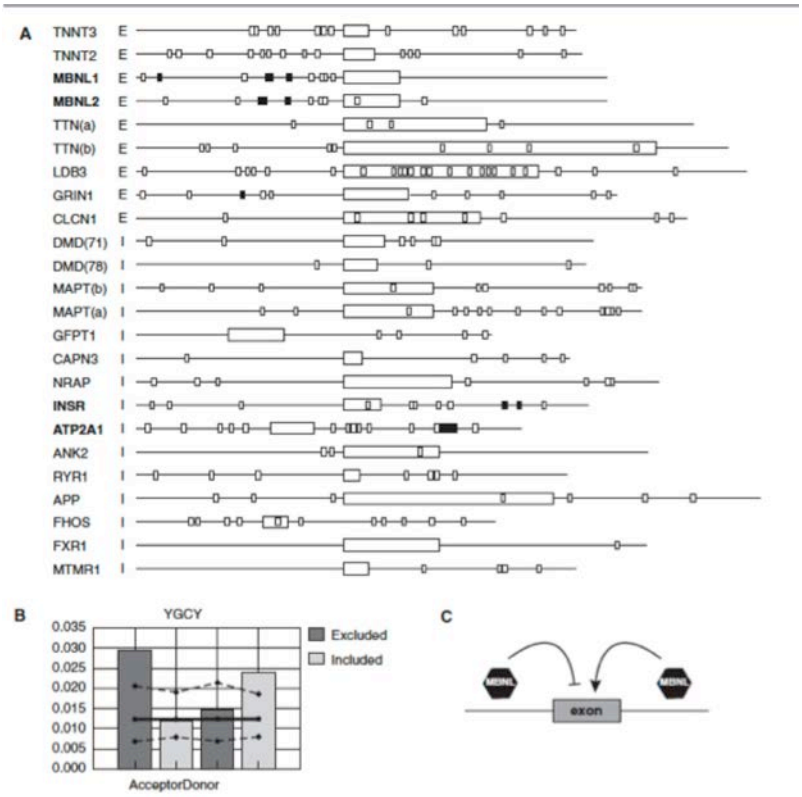

Figure 4. A. Schematic representation of exons and flanking regions that are mis-spliced in DM1 with instances of YGCY motifs that have been tested for MBNL1 binding in black. 'I' or 'E' indicated whether that exon is included or excluded in normal individuals. B. YGCY motifs are enriched in the upstream acceptor flank of exons that are normally excluded and are enriched in the downstream donor flanks. C. Model of how MBNL1 binding correlates to exon regulation. (9)

(12). Other CUG binding proteins, CUG triplet repeat RNA-binding protein (CUGBP1) and ETR3-like factors (CELFs) have also been implicated in DM1 pathogenesis (8).

MBNL1 regulated splicing is dictated by the presence of RNA binding sites for this protein within the regulated pre-mRNAs. It was found that YGCY (where Y represents either U or C) is the minimal RNA binding site for MBNL1 (13). Most of the transcripts that are mis-spliced in DM1 include these motifs (Figure 4). The location of the YGCY motifs corresponds to whether MBNL1 causes inclusion or exclusion of the exon. If the site is located upstream of the exon then MBNL1 binding will generally lead to exon inclusion whereas when the YGCY motifs are located downstream of the exon, MBNL1 binding generally leads to exon inclusion (Figure 4, C).

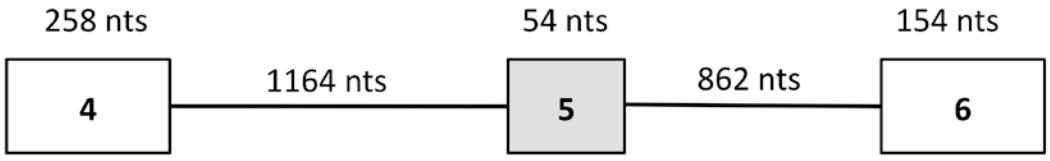

Figure 5. Schematic of the wild-type *MBNL1* minigene containing exon 4, intron 4, exon 5, intron 5, and exon 6.

Exon 5 of the *MBNL1* pre-mRNA is mis-spliced in DM (14). Inclusion of the exon causes MBNL1 to be localized in the nucleus whereas isoforms of MBNL1 lacking exon 5 are found both in the nucleus and the cytoplasm. The mis-splicing of *MBNL1* was studied using a minigene that contained exons 4, 5 and 6 and introns 4 and 5 of the *MBNL1* gene (Figure 5). It was found that a 90-nucleotide region upstream of exon 5 containing a cluster of 10 MBNL1 binding sites was necessary in order to regulate inclusion of exon 5 by MBNL1 (Figure 6). Small deletions within the MBNL1 response element were also made (Figure 6) but did not eradicate the ability of MBNL1 to regulate exon 5 exclusion (Figure 6, B) (14).

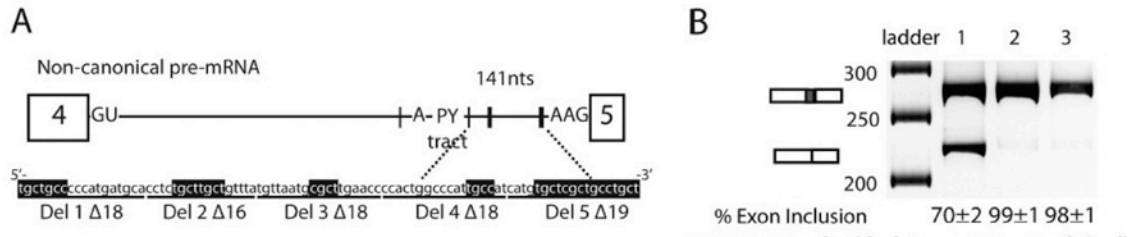

Figure 6. A. Schematic of non-canonical intron four and the 90-nucleotide regulatory response element. B. Splicing of 90 nucleotide deletion minigene. Lane 1 shows the wild type splicing of the *MBNL1* minigene. Lanes 2 and 3 show splicing of the 90 nucleotide deletion in HeLa cells in the absence and presence of overexpressed MBNL1-eGFP, respectively. (11)

In the absence of overexpressed MBNL1, percent exon 5 inclusion increased in all deletions as shown in Figure 7 (e.g. compare lane 1 to lane 2). When MBNL1 was overexpressed, inclusion of the exon decreased significantly and differences in activity between deletion constructs became evident (e.g. compare lane 5 to lane 7). Differences seen at high MBNL1 levels suggest that simply deleting one of these regions does not eliminate regulation, but that the degree of splicing regulation might be different.

To fully understand how MBNL1 binding sites affect splicing, each deletion construct (Figure 7), must be examined over a range of MBNL1 concentrations in comparison with the splicing activity of the wild type *MBNL1* minigene. Each deletion construct varies in the number of sites deleted and the location of the deletion or the distance from other sites that are important for splicing (i.e. splice sites). Examining each deletion construct at varying levels of MBNL1 would elucidate how these sites affect splicing decisions in the MBNL1 autoregulation splicing event. The possession of multiple sites could increase the probability of MBNL1 binding or these pre-mRNAs could require multiple MBNL1 binding events in order to block spliceosome regulation of exon 5.

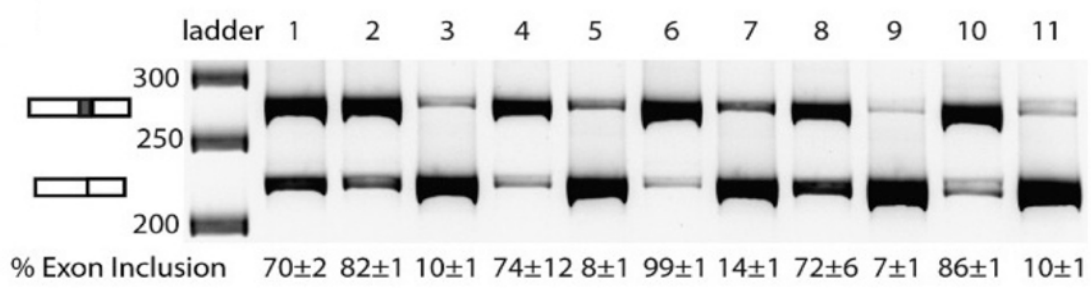

Figure 7. Splicing of del1-5 in HeLa cells in the absence (first of set) and presence (second of set) of overexpressed MBNL1-eGFP. Lane 1 shows the wild type splicing of the *MBNL1* minigene and lanes 2-11 are sets of the following deletions: del1 (lanes 2 and 3), del2 (lanes 4 and 5), del3 (lanes 6 and 7), del4 (lanes 8 and 9), del5 (lanes 10 and 11). (14)

Methodology: Overview

In order to study how MBNL1 affects splicing in myotonic dystrophy, a cell based splicing assay was used. Human embryonic kidney (HEK) 293 cells with the stably integrated MBNL1 gene were transfected with reporter minigenes of the events being studied. After transfection, a titration of doxycycline (dox) was added in order to induce expression of MBNL1. The titration of dox induced different amounts of MBNL1 protein to be made which mimicks how much free MBNL1 is available with different CUG length repeats. To measure splicing the RNA was extracted, reverse transcribed to DNA, PCR amplified, and quantified.

Methodology: Stable MBNL1 expressing HEK293 cell line

As discussed earlier, previous work that was done in order to understand MBNL1's role in has used knockout models and over-expression models of MBNL1. These studies fail to examine the relationship between observed splicing and MBNL1 concentration (which is affected by CUG repeat length). In order to directly investigate the effects of MBNL1, I will employ an inducible MBNL1 expression system to control MBNL1 protein levels to explore the dose-response effect on splicing of different *MBNL1* deletion constructs.

The TRex Flp-In system (Invitrogen) was used to create a stable HEK293 cell line expressing a HA-tagged MBNL1. Levels of MBNL1 can be varied by titrating different concentrations of dox (Figure 8).

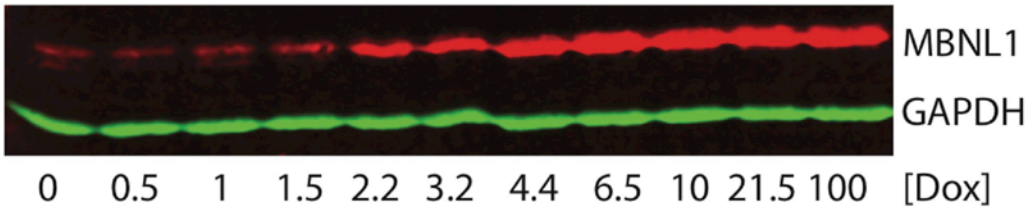

Figure 8. Induction of MBNL1 by doxycycline. Western blot using α-HA to detect HA-MBNL1 expressed as a result of doxycycline titrations into a stable, inducible HA-MBNL HEK293 cell line. GAPDH is used as a loading control. (Stacey Wagner, unpublished)

Methodology: Construction of Splicing Reporter Constructs (14)

The MBNL-eGFP construct was obtained from the laboratory of Maury Swanson (University of Florida), and the DMPK CUG_{960} plasmid was obtained from the laboratory of Thomas Cooper (Baylor College of Medicine). The wild type MBNL1 minigene was made by amplifying regions of the MBNL1 gene containing 51 nucleotides from the 3' -end of intron 3, exon 4, intron 4, exon 5, intron 5, exon 6, and 33 nucleotides of the 5' -end of intron 6 from HeLa genomic DNA using PCR primers with unique restriction sites. The forward primer (5'-CCACAGGATCCGCTTCTT CTTCTTCATGTTGACTAAACCTCATG-3') contained a BamHI site, and the reverse primer (5'ATTCTTATGCGGCCGCCAGATTCATTTATTAAGAAACCCCACCCC-3') contained a NotI site. The amplified genomic DNA was cut with BamHI and NotI, inserted into a pcDNA3 plasmid, and sequenced.

The Δ90 minigene was made in two segments. The first segment was made using the forward primer 5'CCACAGGATCCGCTTCTTCTTCTTCATGTTGACT AAACCTCATG-3' and the reverse primer 5'-GGCTTTCAATTGGTGCATTTTTG GGTAGGTGAGAAAAAACA-3'. The second segment was made using the forward primer 5' –GGCTTTCAATTGAATTAAGACTCAGTCGGCTGTCAAATCAC-3' and the reverse primer 5'-ATTCTTATGCGGCCGCCAGATTCATTTATTAAGAA

ACCCCACCCC-3’. Segment 1 was cut with MfeI and BamHI, and segment 2 was cut with MfeI and NotI. Segments 1 and 2 were then ligated into a pcDNA3 plasmid and sequenced.

The del1 minigene was made in two segments. The first segment was made using the forward primer 5’-CCACAGGATCCGCTTCTTCTTCTTCATGTTGACTA AACCTCATG-3’ and the reverse primer 5’-CATTAACATAAACAGCAAGCAG AGGGTGCATTTTTGGGTAGG-3’. The second segment was made using the forward primer 5’-CCTCTGCTTGCTGTTTATGTTAATGCGCTTGAACC-3’ and the reverse primer 5’ –ATTCTTATGCGGCCGCCAGATTCATTTATTAAGAAACCCCACCCC-3’. The two segments were ligated using standard PCR techniques, inserted into a pcDNA3 plasmid, and sequenced.

The del2, del3, del4, and del 5 minigenes were made using the PCR techniques described for the del1 minigene. All del minigenes used the same forward primer for the first segment and the same reverse primer for the second segment. The del2 minigene used the reverse primer 5’ –GGTTCAAGCGCATTAACATGCATCATGGGGCAGC-3’ for the first segment and the forward primer 5’ –TGTTAATGCGCTTGAACCCCAC TGGCCATTGC-3’ for the second segment. The first segment of the del3 minigene was made using the reverse primer 5’-CATGATGGCAATGGGCCAGTGGTAAACAGCA AGCAGAGG- 3’, and the second segment was made using the forward primer 5’-CCACTGGCCCATTGCCATCATGTGCTCGC-3’. The first segment of the del4 minigene was made using the reverseprimer 5’ –GCAGGCAGCGAGCACATGGGTTCAAGCGCATTAAC-3’. The second segment of the del4 minigene was made using the forward primer 5’ –CATGTGCTCGCTGC

CTGCTAATTAAGACTCAGTCGGC-3’ . The first segment of the del5 minigene was made using the reverse primer 5’ –GACAGCCGACTGAGTCTTAATTATGGC AATGGGCCAGTGG-3’. The second segment of the del5 minigene was made using the forward primer 5’-AATTAAGACTCAGTCGGCTGTCAAATCACTGAAGC GACCCC-3’.

The del3M minigene was made by site-directed mutagenesis of the wild type MBNL1 minigene. It was made using the forward primer 5'-AGCAAGCAGAGGT GCATCATG-3' and reverse primer 5'-GTTTATGTTAATCCCCTTGAACCCCAC-3'. This was amplified using standard PCR techniques, inserted into a pcDNA3 plasmid and sequenced.

Methodology: Cell Culture and Transfection

HEK293 cells were routinely cultured as a monolayer in DMEM media (Invitrogen) supplemented with 10% Fetal Bovine Serum (Gibco) at 37C under 5% CO_2. Prior to transfection, cells were plated in 24-well plates at a density of 1.5 **x** 10^5 cells/well. Cells were transfected 18–24 h later at ~ 80% confluency. In double transfection experiments, 250 ng of each plasmid was transfected into a single well; however, in single transfection experiments 250 ng of empty pcDNA3 vector was used to normalize plasmid concentration between wells. Dosing with dox was done 4-5 hours after transfection. Cells were harvested 18–24 h after transfection. RNA was isolated from the cell pellets using an RNeasy kit (QIAGEN) according to the manufacturer’s protocol.

Methodology: *In-vivo* splicing assay

All reporters were reverse transcribed using a pcDNA3 plasmid-specific

antisense primer, 5’ -AGCATTTAGGTGACACTATAGAATAGGG-3’. The -RT reactions were treated the same as the +RT reactions except that no SuperScript II was added to the -RT reactions. The cDNA from the RT reaction (2 µl) was subjected to 26 rounds of PCR (within linear range) in a 20 µl reaction. PCR amplification for all splice products was done using the sense primer 5’ -GATCAAGGCTGCCCAATACCAG-3’ and the antisense primer 5’ –ATTCTTATGCGGCCGCCAGATTCATTTATTAAGAA ACCCCACCCC-3’. The PCR products were resolved on a 6% native polyacrylamide gel (40% 19:1 acrylamide:bisacrylamide) using SYBR Green (Applied Biosystems). The SYBR Green was diluted 10000X. Quantification of bands was done using the Alpha Imager HP Software from Alpha Innotech. Percent exon inclusion was calculated by dividing the amount of the band indicating inclusion by the total amount of splice product (bands indicating inclusion and exclusion). Background was taken from the space below the two bands. The splicing experiments shown in Figure 11 to 17 were done in triplicate. Averages with standard deviation are shown below gels in the figures.

Results: Splicing assay of *MBNL1* deletions constructs in HeLa and HEK293 cells

Previous work on these deletion constructs was performed in HeLa cells. The present study utilizes an inducible MBNL1 expression system in the HEK293 cell line. In order to understand how the previous studies compare to results from the new

system, splicing assays using both cell types were carried out. Differences in splicing activity helped to elucidate fundamental differences between the two cell lines.

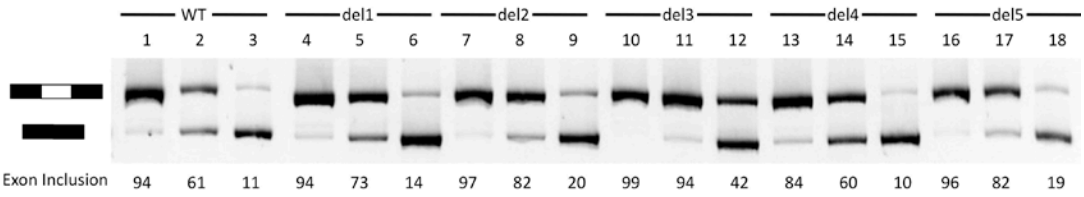

Figure 9. Splicing of *MBNL1* constructs in HeLa cells. Lanes 1-3 show splicing of wild-type *MBNL1*, Lanes 4-6 show splicing of Del 1, Lanes 7-9 show splicing of Del 2, Lanes 10-12 show splicing of Del 3, Lanes 13-15 show splicing of Del 4 and Lanes 16-18 show splicing of Del 5. The first of each set is a cotransfection with DMPK CUG_{960} minigene, the second is without treatment and the third is in the presence of overexpressed MBNL1-eGFP.

Figure 9 shows the splicing of wild type *MBNL1* and each of the five deletion constructs in HeLa cells. The first lane of each set includes a co-transfection with the CUG_{960} plasmid that mimics the disease state. The second of each set is without treatment and represent splicing under endogenous MBNL1 protein levels. The third lane shows splicing in the presence of overexpressed MBNL1 protein. The previous study did not assay splicing under disease state conditions. This study included a co-transfection with CUG_{960} repeat plasmid in order to ascertain if differences existed at this state. As seen in Figure 9 exon 5 inclusion ranges from 94-99% indicating that the all constructs respond similarly in disease conditions. Under endogenous MBNL1 levels splicing activity of the constructs ranges from 60-94% exon inclusion indicating that the constructs behave differently when MBNL1 is at low levels. Thus we can see that deletion of binding sites (or systematic deletion of 16-19 nucleotide between the polypyrimidine tract and 3'-splice site) is having an effect on MBNL1's ability to regulate exon 5 splicing. This becomes yet more apparent in the activity levels of splicing in the presence of overexpressed MBNL1. While exon 5 inclusion can be

driven down significantly, each construct responds differently and cannot be driven down to the same degree (ranging from 42% exon inclusion to 10%).

The same set of experiments was repeated using the inducible MBNL1 expression system in the HEK293 cell line (Figure 10). In addition to the points that were assayed above, a high concentration of dox (20 ng/ml) was also included in order to find further differences between varying concentrations of MBNL1 protein. Just as in HeLa cells, the disease state does not vary much between different constructs indicating that at high levels of MBNL1 sequestration there is limited MBNL1 available for splicing. Differences between the wild type and deletion constructs become apparent at endogenous and high levels of MBNL1 just as they do in HeLa cells. Interestingly, for al six minigenes (wild type and the five mutants), overexpression of MBNL1-eGFP lead to a decrease in exon 5 inclusion compared to 20 ng/ml dox suggesting that overexpression via plasmid resulted in more MBNL1 protein.

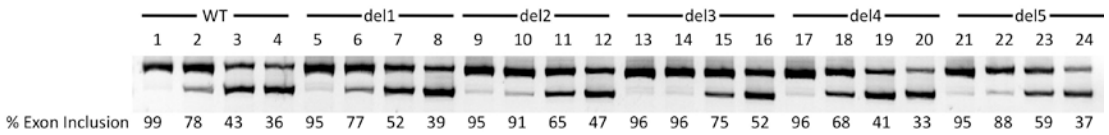

Figure 10. Splicing of MBNL1 constructs in HEK cells. Lanes 1-4 show splicing of wild-type MBNL1, Lanes 5-8 show splicing of Del 1, Lanes 9-12 show splicing of Del 2, Lanes 13-16 show splicing of Del 3, Lanes 17-20 show splicing of Del 4 and Lanes 21-24 show splicing of Del 5. The first of each set is a co-transfection with CUG_{960} minigene, the second is without treatment and the third is with a 20 ng/mL treatment of dox and the fourth is in the presence of overexpressed MBNL1-eGFP.

It is important to note that though the level of exon 5 inclusion is significantly different in HeLa and HEK293 cells. In HeLa cells percent exon 5 inclusion can be pushed to as low as 10 % in the presence of overexpressed MBNL1, in HEK293 cells the level only drops to 36 %. If equal transfection efficiency occurred in both cell lines, the same amount of MBNL-eGFP was expressed, these results suggest that MBNL1 is less able to regulate exon 5 in HEK293 cells compared to HeLa cells. This is seen in all

constructs and under both endogenous and induced levels of MBNL1. Presumably these differences are functions of the inherent differences between HeLa and HEK293 cells. HeLa cells are a mature, adult cell line while HEK293 cells are embryonic. MBNL1, along with other splicing factors, are developmentally regulated. MBNL1 levels are low at early stages and higher in adult cells. It is also possible that other factors work with MBNL1 to regulate this event and are at low levels or missing the HEK293 cells compared to the HeLa cells.

Results: Dose-response behavior of the wild type *MBNL1* minigene

In the *MBNL1* autoregulation splicing event MBNL1 negatively regulates exon 5 inclusion by binding upstream of the cassette exon where there is a clustering of 10 YGCY motifs. The percent exon 5 inclusion decreases as MBNL1 protein levels increase (Figure 11). A significant increase in splicing activity is seen when comparing the disease state (Lane 1, Figure 11, C), and zero treatment or endogenous MBNL1 levels (Lane 2, Figure 11, C). With 960 CUG repeats it can be assumed that there is no free MBNL1 in the nucleus. Endogenous levels, however, are enough to see a response from the *MBNL1* minigene (80 % inclusion versus 98 % under the disease model). The response ranges from 39 percent exon 5 inclusion at high levels of dox / MBNL1 to 75 percent exon 5 inclusion at the lowest level of MBNL1 for a 46 % change over the concentration range of dox tested. By 2.2 ng/ml dox the level of exon 5 inclusion had leveled off at the low 40 % range.

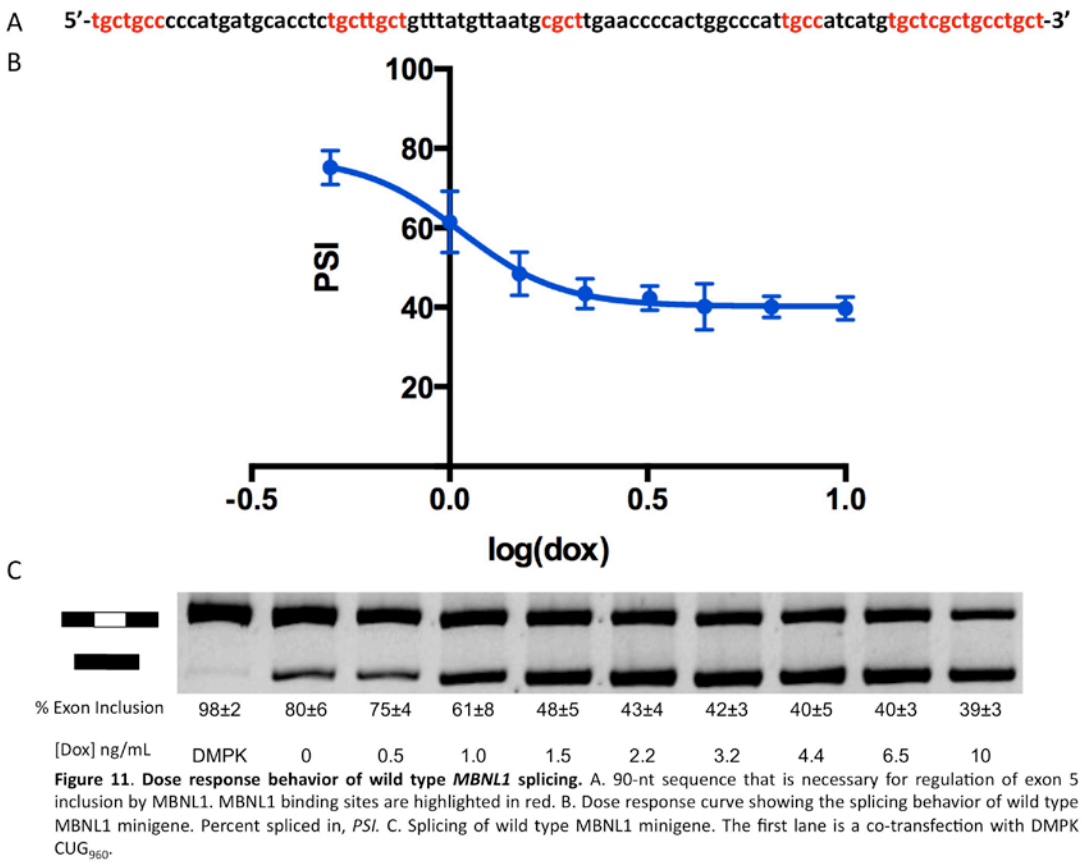

Figure 11. Dose response behavior of wild type *MBNL1* splicing. A. 90-nt sequence that is necessary for regulation of exon 5 inclusion by MBNL1. MBNL1 binding sites are highlighted in red. B. Dose response curve showing the splicing behavior of wild type MBNL1 minigene. Percent spliced in, *PSI*. C. Splicing of wild type MBNL1 minigene. The first lane is a co-transfection with DMPK CUG_{960}.

Results: Dose-response behavior of del1

The del1 mutant (Figure 12) has 18 nucleotides removed from the 5' end of the 90 nucleotide MBNL1 response element, which contains two MBNL1 binding sites. When treated with a titration of MBNL1 protein, del1 showed a different response than that of the wild type *MBNL1* minigene. Endogenous levels of MBNL1 produce a change from the disease state in which 960 CUG repeats are co-transfected. The splicing activity of del1 ranges from 54 percent exon 5 inclusion at high concentrations of dox / MBNL1 to 82 percent exon 5 inclusion at the lowest concentration of dox / MBNL1. This range is reduced (28 % change) compared to wild type that had a 46 % change in exon 5 inclusion. In addition, more dox (3.2 ng/ml) is required to reach the

point at which the del1 mutant levels off (~55% exon 5 inclusion) compared to 2.2 ng/ml for the wild type event.

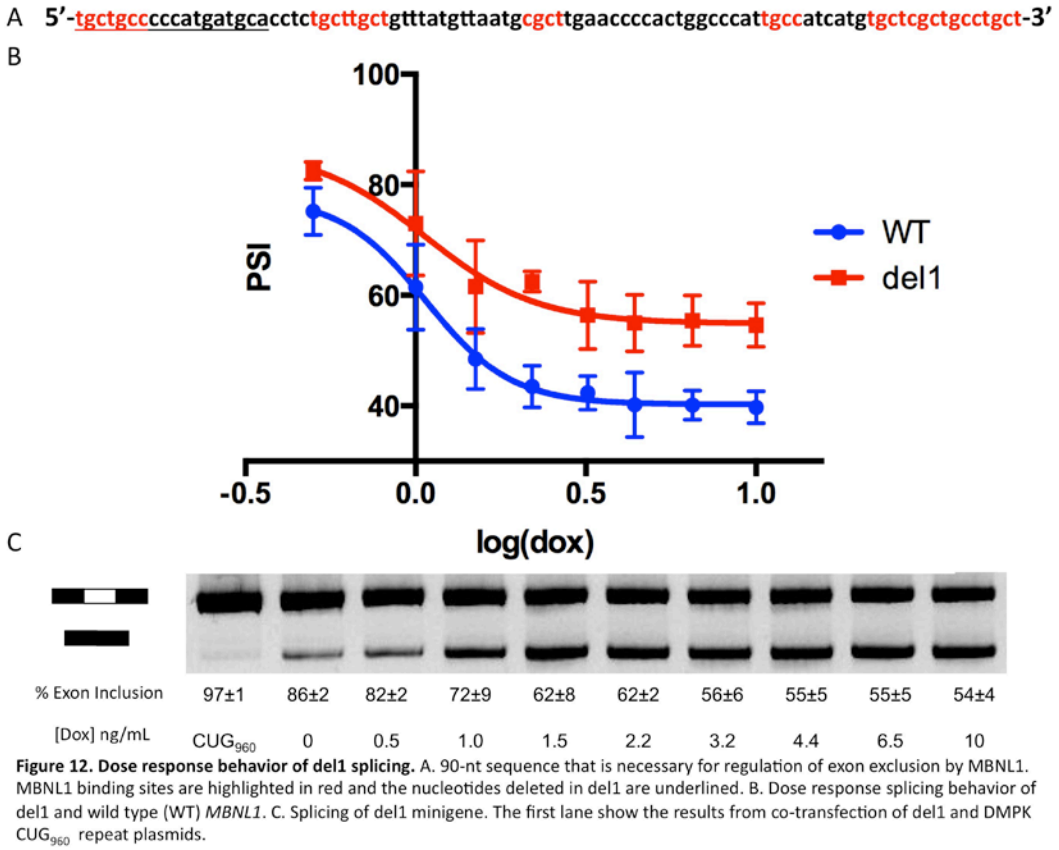

Figure 12. Dose response behavior of del1 splicing. A. 90-nt sequence that is necessary for regulation of exon exclusion by MBNL1. MBNL1 binding sites are highlighted in red and the nucleotides deleted in del1 are underlined. B. Dose response splicing behavior of del1 and wild type (WT) *MBNL1*. C. Splicing of del1 minigene. The first lane show the results from co-transfection of del1 and DMPK CUG_{960} repeat plasmids.

Results: Dose-response behavior of del2

Sixteen nucleotides were deleted in the del2 mutant construct. These 16 nucleotides include two MBNL1 binding sites (Figure 13, A). Though the regulation of exon 5 inclusion in del2 responded to a titration of MBNL1, the response was different from that of the wild type *MBNL1* minigene. Activity ranged from 61 percent exon 5 inclusion at high levels of dox / MBNL1 to 82 percent at low levels of dox / MBNL1, which is significantly higher compared to wild type and modestly higher compared to del1. The slope of the dose-response curve also decreased along with the decrease in range of activity compared to wild type.

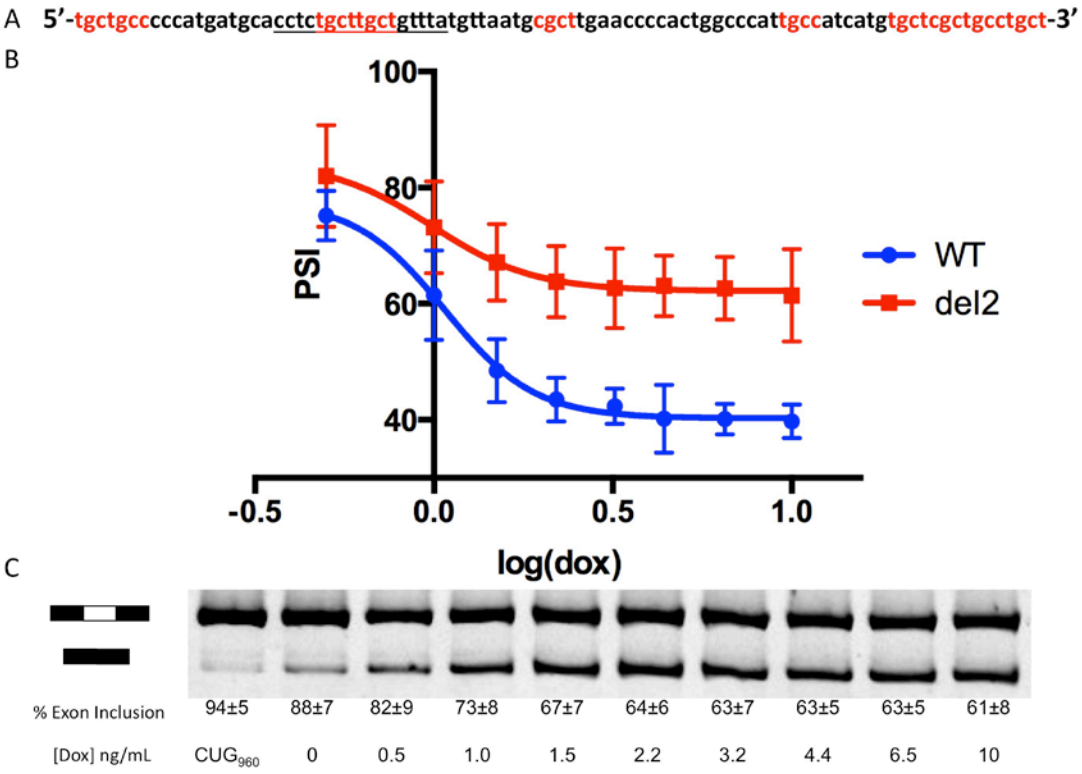

Figure 13. Dose response behavior of del2 splicing. A. 90-nt sequence that is necessary for regulation of exon exclusion by MBNL1. MBNL1 binding sites are highlighted in red and the nucleotides deleted in del2 are underlined. B. Dose response splicing behavior of del2 and wild type (WT) *MBNL1*. C. Splicing of del1 minigene. The first lane show the results from co-transfection of del2 and DMPK CUG_{960} repeat plasmids.

Results: Dose-response behavior of del3

In the del3 construct, 18 nucleotides have been deleted from the wild type *MBNL1* minigene (Figure 14, A). These 18 nucleotides contain only one YGCY MBNL1 binding motif. Deletion of this site greatly decreases splicing activity compared to wild type and the other deletion mutants. Compared to the range of activity seen in splicing of the wild type splicing event (46 %), the range of del3 is significantly decreased (19 %). The percent exon 5 inclusion is shifted compared to wild type such that at the highest dox / MBNL1 concentration exon 5 is still included at 74 %. The slope in the linear portion of the curve is less steep compared to wild type as well. There is no change in splicing between the disease state (CUG_{960} plasmid co-transfection,

Lane 1 Figure 14, C.) and endogenous MBNL1 levels (zero treatment, Lane 2, Figure 14, C), suggesting that the low level endogenous MBNL1 was not able to regulate del3 at all.

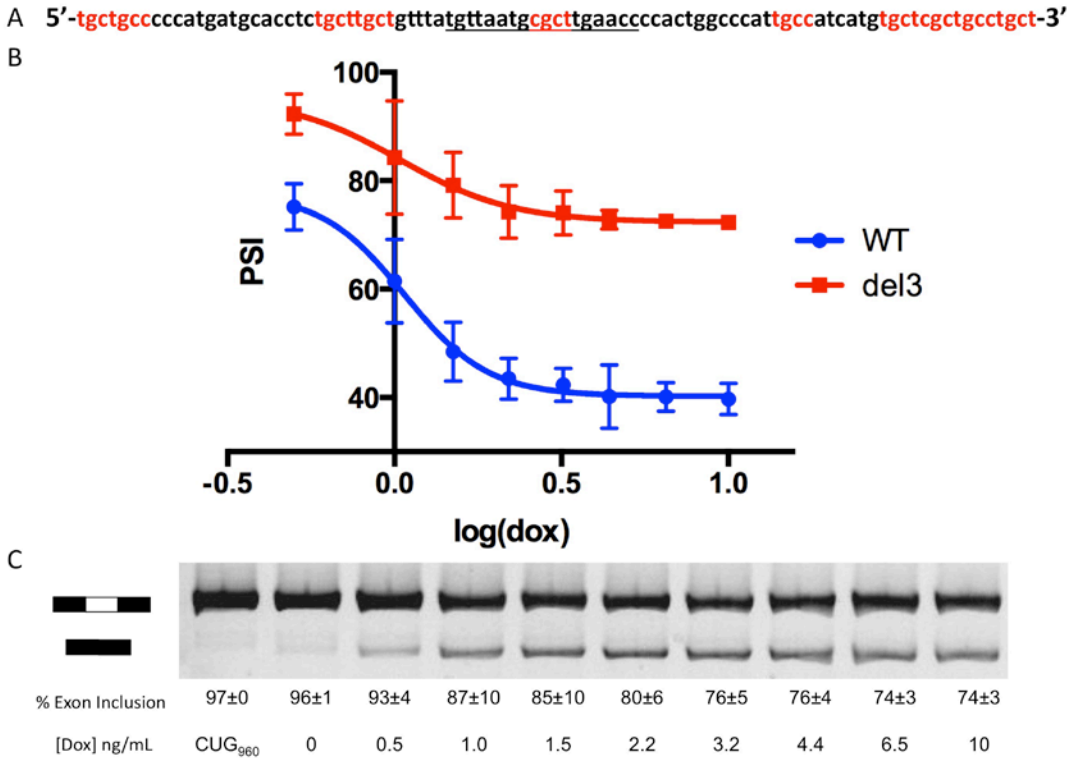

Figure 14. Dose response behavior of del3 splicing. A. 90-nt sequence that is necessary for regulation of exon exclusion by MBNL1. MBNL1 binding sites are highlighted in red and the nucleotides deleted in del3 are underlined. B. Dose response splicing behavior of del3 and wild type (WT) *MBNL1*. C. Splicing of del1 minigene. The first lane show the results from co-transfection of del3 and DMPK CUG_{960} repeat plasmids.

Results: Dose-response behavior of del4

Like del3, del4 also has one MBNL1 binding site in the 18 nucleotide deleted portion of the construct (Figure 15, A). Similar to the wild type *MBNL1* minigene, del4 shows an immediate difference in splicing activity between disease state (Lane 1, Figure 15, C) and endogenous MBNL1 levels (Lane 2, Figure 15, C) indicating del 4 is sensitive to endogenous MBNL1 protein. Interestingly, del4 had the lowest level of exon 5 inclusion at 32 % (wild type 39%). The range of activity for del4 (33%) is lower compared to wild type (46%) meaning the activity curve for del4 is shifted below wild

type. The slope for del4 is comparable to that of the wild type dose response curve. Although both del3 and del4 results in the removal of one MBNL1 binding site, only del3 had a profound effect on splicing activity while del4 had the most modest effect on MBNL1 activity.

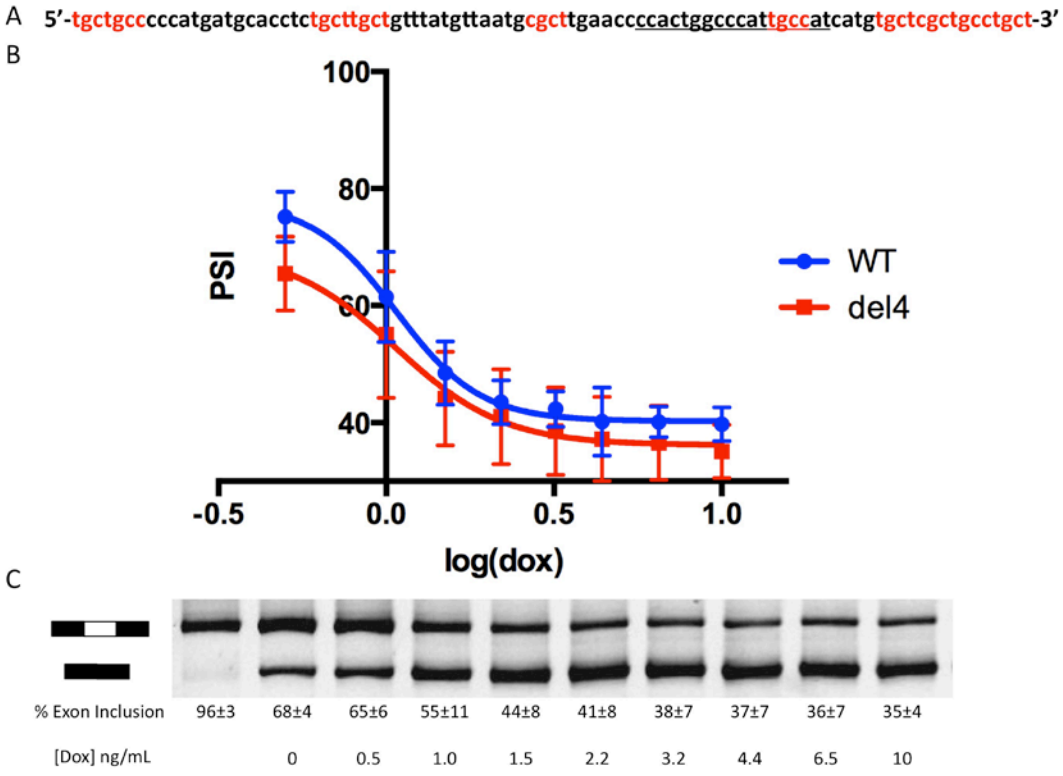

Figure 15. Dose response behavior of del4 splicing. A. 90-nt sequence that is necessary for regulation of exon exclusion by MBNL1. MBNL1 binding sites are highlighted in red and the nucleotides deleted in del4 are underlined. B. Dose response splicing behavior of del4 and wild type (WT) *MBNL1*. C. Splicing of del4 minigene. The first lane show the results from co-transfection of del4 and DMPK CUG_{960} repeat plasmids.

Results: Dose-response behavior of del5

Nineteen nucleotides were deleted in the del5 construct. This 19 nucleotide region contained four MBNL1 binding sites. Deletion of these four sites did not result in the loss of regulation of exon 5 inclusion by MBNL1 but did reduce the range of activity. Del5 has a similar reduction in range of activity (26 %) compared to the other mutants although not a reduced compared to del3 (Figure 16).

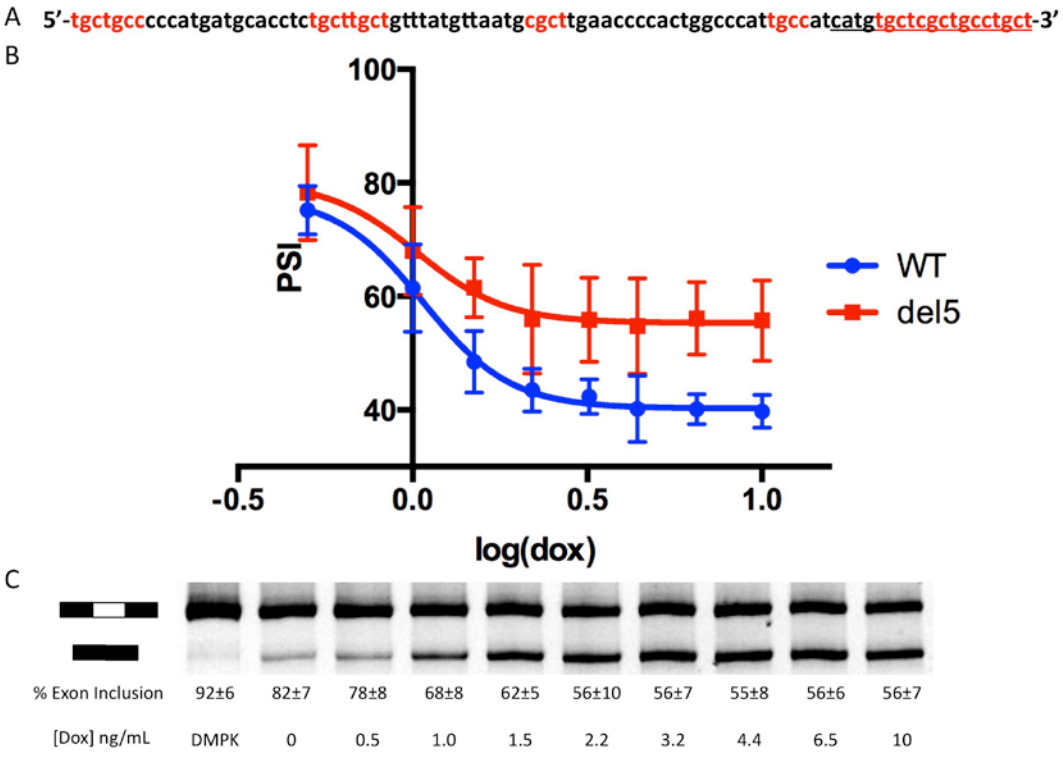

Figure 16. Dose response behavior of del5 splicing. A. 90-nt sequence that is necessary for regulation of exon exclusion by MBNL1. MBNL1 binding sites are highlighted in red and the nucleotides deleted in del5 are underlined. B. Dose response splicing behavior of del5 and wild type (WT) *MBNL1*. C. Splicing of del5 minigene. The first lane show the results from co-transfection of del5 and DMPK CUG_{960} repeat plasmids.

Results: Dose-response behavior of 3M

The most significant difference in splicing activity compared to wild type was seen in the del3 mutant as both the sensitivity (no regulation by endogenous MBNL1) and range of response were reduced. To determine if the single MBNL1 site or other aspects of del3 were important a site mutant of the del3 mutant was made (Figure 17, A). No nucleotides were deleted. Instead, the MBNL1 binding site within the 18 nucleotide region of del3 was mutated (two guanosines were changed to cytosines). A dose response curve for this site mutant, 3M, was generated using the same splicing assay as described previously. The change in splicing activity for 3M ranged from 25 % (lowest dox / MBNL1) to 85 % exon 5 inclusion (highest dox / MBNL1). Although the 3M splicing reporter was similar to del 3 (Figure 17, C), the range of splicing activity

for del 3 was less at 19 % compared to 25 % for 3M. Also, del3 only reduced exon 5 inclusion to 74 % while 3M reduced exon 5 inclusion to 60 %. This result suggests that the YGCY site within the del3 region is major element of regulation and that additional aspects of this region (spacing of YCGY motifs and/or other sequence elements) likely modulate MBNL1 splicing regulation.

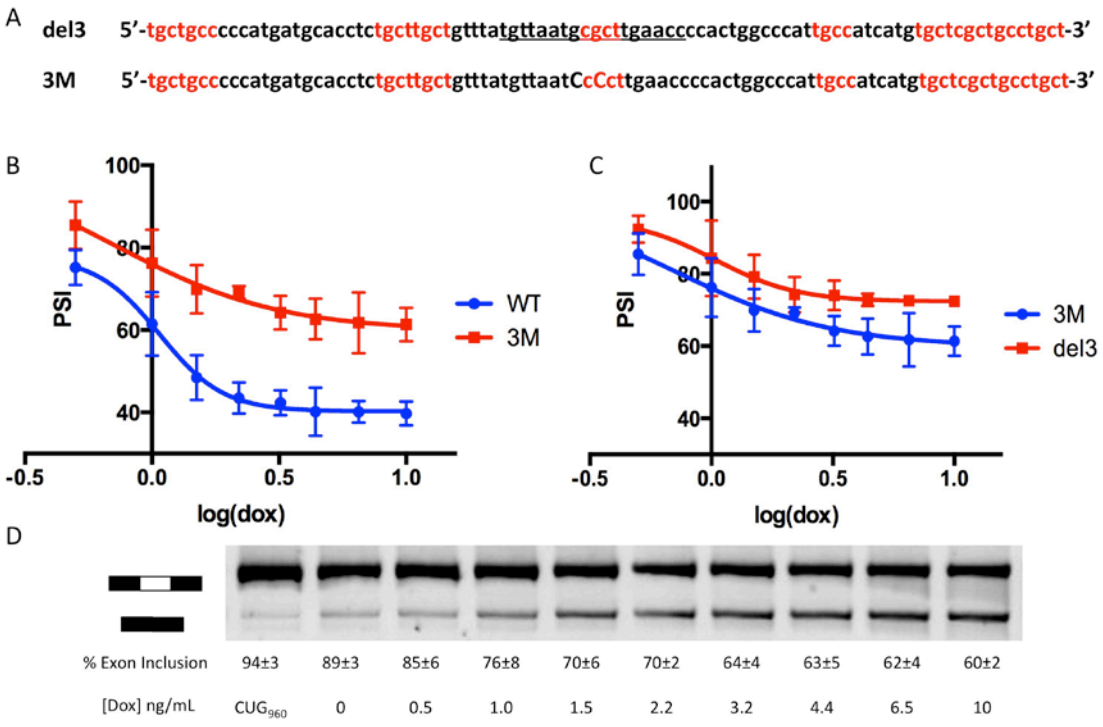

Figure 17 Dose response behavior of 3M splicing. A. 90-nt sequence that is necessary for regulation of exon exclusion by MBNL1. MBNL1 binding sites are highlighted in red and the nucleotides mutated in the 3M construct are uppercase. B. Dose response splicing behavior of 3M and wild type (WT) *MBNL1*. C. Dose response splicing behavior of del3 and 3M. D. Splicing of 3M minigene. The first lane show the results from co-transfection of 3M and DMPK CUG_{960} repeat plasmids.

Discussion: Overview

This study aimed to examine how the alternative splicing decisions in the MBNL1 pre-mRNA are dependent upon the number of MBNL1 binding sites. Inclusion of exon 5 is negatively regulated by MBNL1. The 3'-end of the flanking intron (intron

4) contains an ultraconserved region that is longer than 200 nucleotides within which a 90 nucleotide region contains 10 YGCY MBNL1 binding sites that are necessary for exon 5 regulation by MBNL1 (14). Deletion constructs with binding sites deleted were examined over a range of MBNL1 protein concentrations. Though each deletion construct was still able to regulate exon 5 inclusion, this activity varied greatly showing that each region of the 90 nucleotide region was important and that the binding sites are not redundant. Del4 showed splicing activity similar to that of the wild type *MBNL1* minigene, while del3 exhibited splicing activity most different from wild type.

Minigene	ΔSC	ΔPSI	Slope	[Dox] (ng/ml)
WT MBNL1	18 %	36 %	-3.5	2.2
del1	11 %	28 %	-2.6	3.2
del2	6 %	21 %	-3.5	2.2
del3	1 %	19 %	-2.0	3.2
del4	28 %	30 %	-2.7	3.2
del5	10 %	22 %	-3.4	2.2
3M	5 %	25 %	-2.7	3.2

Table 2. ΔSC: change in splicing CUG_{960} co-transfection versus endogenous MBNL1 (zero dox). ΔPSI: change in splicing from lowest dox / MBNL1 to highest dox / MBNL1. slope: Slope: slope of linear range of curve from Figures 11-17. [Dox] (ng/ml): Concentration of dox at which response to MBNL1 leveled off.

Discussion: Sensitivities

Each deletion construct exhibited a different dose response to MBNL1 as seen in Figures (11-17). The sensitivity of each construct to MBNL1 was defined as the change in splicing (ΔSC) under disease conditions (co-transfection with CUG_{960}) to endogenous MBNL1 levels (zero dox). Table 2 shows the ΔSC for the wild type

minigene and the six mutants. While ΔSC for wild type was 18 %, this was not the most sensitive construct. The ΔSC for del 4 was 28% indicating that at endogenous MBNL1 levels, splicing decisions are being shifted farther away from inclusion than in wild type. Del3 was the least sensitive to MBNL1 at endogenous levels. The 3M mutation, modestly increased the sensitivity (5%). Since there was only a small change in the sensitivity of 3M, this indicates that the MBNL1 binding site in this region is important for regulation but that other aspects of this region are also important when regulating splicing decisions.

Discussion: Dose-response behavior of constructs to dox titration

Table 1 shows the range of response of each minigene to dox / MBNL1 titration (ΔPSI) and the slope of each curve. While the ΔSC and ΔPSI provide information about sensitivity and range of response respectively, the slope provides information about the different in the mechanism by which each deletion construct is regulating splicing.

ΔPSI for wild type was 46 % which was the largest range among the seven minigenes and has a slope of -3.5. Though del4 exhibited higher sensitivity to MBNL1 than the wild type minigene, the ΔPSI for del4 is 33 % which is smaller than the wild type range. This is because at 0.5 ng/ml of dox, exon 5 inclusion is ~65% whereas for wild type exon 5 inclusion is ~75%. Thus del4 is already closer to the final percentage of exon 5 inclusion than wild type is. Also, the range being capture is smaller because the curve is shifted to the left (Figure 15). Del3 had the smallest range (ΔPSI = 19 %) compared to wild type and the most different slope (-2.0). The 3M site mutant had an slightly increased range (ΔPSI = 25 %) indicating again that the YGCY motif, while

being an important regulatory event, there are other aspects of the del3 region that are important for splicing regulation.

Discussion: Binding site number and distance from 3'-splice site

Del3, which showed the largest deviation from wild type *MBNL1* behavior (Table 2), contained the deletion of a single YGCY binding motif and was in the center of the 90 nucleotide regulatory element. In order to understand whether this deletion has such a profound effect on splicing activity due to the deletion of the site or deletion of the 18 nucleotides which affected the distance of binding sites from the 3'-splice site, the 3M mutant was assayed. It was found to have an activity similar to that of the del3 mutant though there were some differences in activity (both the sensitivity and range of response increased slightly). This indicates that both the MBNL1 binding site contained within the region of del3 and the spacing of YGCY motifs in is important for regulation.

Furthermore, the number of missing potential MBNL1 binding sites does not correlate with reduced range of activity because if this simple correlation existed we would have expected del5 (four for del 5) to display the largest reduction in range of activity. The deletion of one site in one specific location (del3) affects splicing more than the deletion of multiple sites (del5) (Table 2). Del5 includes a cluster of four binding sites that is closest in the 90 nucleotide regulatory element to the 3'-splice site. Yet their removal does not have as dramatic of an effect as del3, which includes only one site and is farther from the 3'-splice site. If the binding of MBNL1 at the sites upstream of the 3'-splice site was necessary in order to block spliceosome activity, del5 would presumably effect splicing activity more. In the same way if multiple sites were needed in order to promote MBNL1 binding then the deletion constructs with more than

just one MBNL1 binding site would show decreased splicing activity. Neither hypothesized mechanism fully explains how MBNL1 is affecting splicing since the deletion of a single site in del3 greatly increases exon 5 inclusion. Del4, however, which is also a deletion that contains only one MBNL1 binding site, behaves as the wild type *MBNL1* minigene does. This suggests that the mechanism of MBNL1's negative regulation of exon 5 inclusion is more complicated than multiple MBNL1 binding events upstream of exon 5. The data suggest that the different MBNL1 binding sites are neither equivalent nor redundant, but that each one contributes to splicing activity to some degree.

It is possible that other splicing factors have binding sites within this regulatory regions and the deletion of their binding sites is what is affecting splicing activity levels in the deletion constructs. Further mutational and knockdown studies could ascertain what other splicing factors are affecting the splicing events that are regulated by MBNL1.

Future Directions

This assay has provided much more information about each construct and how the different regions contribute to splicing than a simple overexpression assay. In order to further study the mechanism behind MBNL1 promotion of exon 5 exclusion in the wild type *MBNL1* pre-mRNA, more mutational analysis should be performed using this assay. It would be informative to know which of the 10 binding sites are sufficient for regulation. This could be done via further mutational analysis. The 90 nucleotide region

could also be replaced with repeating YGCY repeats in order to see if splicing activity can be increased beyond that of the wild type *MBNL1* minigene.

In order to further characterize what makes del3 (and 3M) so different from the rest of the deletions, it would be useful to shift the location of the YGCY site within the 18 nucleotides of del3. The region of del3 could also be placed in a different part of the 90 nucleotide regulatory region in order to determine how distance from the 3'splice site of this site affects splicing.

Del4 is another interesting construct in that it is more sensitive to endogenous MBNL1 than the wild type minigene. Moving the del4 region to a different location within the 90 nucleotide regulatory element would help determine how the spacing of this YGCY motif affects splicing.

Furthermore, mutations in which a YGCY binding region from a different minigene regulated by MBNL1 is inserted upstream of the MBNL1 exon 5 regulation event could show if the change in splicing can be rescued.

Knocking down other splicing factors could also show what splicing factors are necessary besides MBNL1 and define what are available in HEK293 cells.

References

1. Wahl, M.C., Will, C.L., Luhrmann, R. (2009) *Cell* 136, 701-718.

2. Hallegger M, Llorian M, and Smith CWJ. (2010) *FEBS Journal* 277:856-886.

3. Wang, et. al. (2008) *Nature* 456:470-476.

4. Murray, J.I., Voelker, R.B., Henschied, K.L., Warf, M.B., Berglund, J.A. (2008) *Genome*

Biol. 9(6), R97.

5. Hoskins, A.A., Moore, M.J. *Trends Biochem Sci.* (2012) 37(5):179-188.

6. Nilsen, T.W. (2003) *Bioessays* 25:1147-1149.

7. Sammeth M, Fossaic S, Guigo R. (2008) *PLoS Comput Bio* 4:8.

8. Gatchel, J.R., Zoghbi, H.Y. (2005). *Nature Reviews*. 6, 743-755.

9. Ranum, L. P., and Day, J. W. (2004) *Am. J. Hum. Genet*. 74, 793– 804

10. Ranum, L. P., and Cooper, T. A. (2006) *Annu. Rev. Neurosci*. 29, 259 –277

11. Osborne, R. J., and Thornton, C. A. (2006) *Hum. Mol. Genet.* 15, R162–R169 J

12. Fardaei, M., Larkin, K., Brook, J. D., and Hamshere, M. G. (2001) *Nucleic Acids Res.*

29, 2766 –27711

13. Charlet-B, N., Savkur, R. S., Singh, G., Philips, A. V., Grice, E. A., and Cooper, T. A.

(2002) *Mol. Cell* 10, 45–53

14. Goers, E. S., Purcell, J., Voelker, R. B., Gates, D. P., and Berglund, J. A. (2010) *Nucleic Acids Res.* 38, 2467–2484

15. Terenzi, F., and Ladd, A. N. (2010) *RNA Biol.* 7, 43–55

16. Gates D.P., Coonrod L.A., and Berglund J.A. (2011). *J Biol Chem* 286: 34224-34233

17. Berejano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., Haussley, D. (2004) *Science*. 304, 1321-1325.