THE REGULATION OF ALTERNATIVE SPLICING ASSOCIATED WITH MYOTONIC DYSTROPHY

by

MICHAEL BRYAN WARF

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

September 2009

University of Oregon Graduate School

Confirmation of Approval and Acceptance of Dissertation prepared by:

Michael Warf

Title:

"The regulation of alternative splicing associated with Myotonic Dystrophy"

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

Kenneth Prehoda, Chairperson, Chemistry
J. Andrew Berglund, Advisor, Chemistry
Victoria DeRose, Member, Chemistry
Peter von Hippel, Member, Chemistry
Alice Barkan, Outside Member, Biology

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

September 5, 2009

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

An Abstract of the Dissertation of

Michael Bryan Warf for the degree of Doctor of Philosophy

in the Department of Chemistry to be taken September 2009

Title: THE REGULATION OF ALTERNATIVE SPLICING ASSOCIATED WITH MYOTONIC DYSTROPHY

Approved: _____

Dr. J. Andrew Berglund

Myotonic Dystrophy (DM) is a genetic disorder with multisystemic symptoms that is caused by expression (as RNA) of expanded repeats of CTG or CCTG in the genome. It is hypothesized that the protein MBNL1 (Muscleblind-like-1) is sequestered to the expanded CUG or CCUG RNAs. MBNL1 regulates the alternative splicing of a variety of pre-mRNAs and its mis-localization results in mis-splicing of a subset of pre-mRNAs that are linked to the symptoms found in DM patients.

I initially demonstrated that MBNL1 can bind short structured CUG and CCUG repeats with high affinity and specificity *in vitro*. Next, I was able to determine and articulate the first structure of a binding site of MBNL1 in an endogenous pre-mRNA that

it regulates. I found that MBNL1 binds a stem-loop in the cardiac troponin T (cTNT) pre-mRNA. The stem-loop contains two mismatches and resembles both CUG and CCUG repeats. I determined that MBNL1 regulated exon 5 by directly competing with the essential splicing factor U2AF65 for binding upstream of exon 5. When U2AF65 is prevented from binding, factors in the spliceosome can no longer be recruited and the following exon is skipped. Furthermore, I found that MBNL1 and U2AF65 compete by binding mutually exclusive RNA structures.

I also characterized a potential therapeutic approach for DM. Current data suggest that if MBNL1 is released from sequestration, disease symptoms may be alleviated. Using a targeted screen of small molecules known to bind structured nucleic acids, I identified the small molecule pentamidine as a compound that disrupted MBNL1 binding to CUG repeats *in vitro*. I showed in cell culture that pentamidine was able to reverse the mis-splicing of two pre-mRNAs affected in DM. Pentamidine also significantly reduced the formation of RNA foci in tissue culture cells, which are characteristic of DM. MBNL1 was released from the foci in the treated cells. Furthermore, pentamidine partially rescued splicing defects of two pre-mRNAs in mice expressing expanded CUG repeats.

This dissertation includes three previously published co-authored publications.

CURRICULUM VITAE

NAME OF AUTHOR: Michael Bryan Warf

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene OR

Guilford College, Greensboro NC

DEGREES AWARDED:

Doctor of Philosophy in Chemistry, 2009, University of Oregon

Bachelor of Science in Chemistry, 2004, Guilford College

Bachelor of Science in Biology, 2004, Guilford College

AREAS OF SPECIAL INTEREST:

Pre-mRNA splicing and processing

Gene Regulation

PROFESSIONAL EXPERIENCE:

Graduate Research Fellow, Department of Chemistry, University of Oregon, Eugene, 2006-2009

Graduate Teaching Fellow, Department of Chemistry, University of Oregon, Eugene, 2005-2006

Research Assistant, Dept of Chemistry, Wake Forest University, Winston-Salem, 2003-2004

Research Experience for Undergraduates (REU), University of Utah, Salt Lake City, 2002

GRANTS, AWARDS AND HONORS:

American Heart Association Graduate Fellowship, Department of Chemistry, University of Oregon, 2008-2009

Clarence and Lucille Dunbar Award, Department of Chemistry, University of Oregon, 2008

Myotonic Dystrophy Foundation Award for Excellence in Research, 2007

NIH Training Grant, Department of Chemistry, University of Oregon, 2005-2008

Ted Benfy Senior Award in Chemistry, Department of Chemistry, Guilford College, 2004.

Graduate with Departmental Honors, Chemistry and Biology, Guilford College, 2004.

Graduate with High Honors, Guilford College, 2004

Dean's Award for Writing in the Sciences, Guilford College, 2003 and 2004

PUBLICATIONS:

Warf, MB*. Nakamori, M*. Matthys, C. Thornton, C and JA Berglund. (2009) "Pentamidine reverses the splicing defects associated with Myotonic Dystrophy." *PNAS*, under revision.

* These authors contributed equally

Banerjee, P. Warf, MB. and R Alexander. (2009) "Effect of a domain-spanning disulfide on aminoacyl-tRNA synthetase activity." *Biochemistry*, under review.

Warf, MB. Diegel, JV. Von Hippel, PH. and JA Berglund. (2009). "The protein factors MBNL1 and U2AF65 bind alternative RNA structures to regulate splicing." *PNAS*, 106 (23): 9203-8.

Murray, J. Voelker, R. Henscheid, K. Warf, MB and JA Berglund. (2008). "Identification of motifs that function in the splicing of non-canonical introns." *Genome Biology*, 9 (6) R97.

Warf MB and Berglund JA. (2007). MBNL binds similar RNA structures in the CUG repeats of Myotonic Dystrophy and its pre-mRNA substrate cardiac troponin T. *RNA*, 13 (12): 2238-51.

ACKNOWLEDGEMENTS

I thank all the past and present members of the Berglund and von Hippel labs. I also want to thank members of the Barkan, Prehoda and Zong labs for experimental advice, reagents, and various other forms of help. I especially want to thank Andy Berglund, Peter von Hippel, Alice Barkan, Ken Prehoda, Vickie DeRose, Rodger Voelker, Sandra Grieves, Kausikie Datta, Emily Goers, Dylan Farnsworth, Amy Mahady, Kristy Henscheid, Jana Marcette, Jamie Purcell, Brit Ventura, Marisa Connell, Stephen Garrey, Meredith Price, Amy Mahady, Kenneth Watkins and Danielle Cass. I thank Charles Thornton, Tom Cooper, Maury Swanson and Nicholas Webster for plasmids. I also thank Manny Ares, Ralf Krahe and Charles Thornton for advice on manuscripts.

This work was supported by grants from the NIH (AR053903), and the MDA (grant #93113) to J.A.B, and the NIH (GM-15792) to Peter H. von Hippel, who serves as co-advisor (with J.A.B.) to M.B.W. M.B.W. was supported by NIH training grant GM-07759 to the Institute of Molecular Biology, and an American Heart Association Graduate Fellowship.

TABLE OF CONTENTS

Chapter Page

Chapter Page

Chapter Page

LIST OF FIGURES

Figure Page

Figure Page

Chapter IV

LIST OF TABLES

Table Page

CHAPTER I

INTRODUCTION

Constitutive pre-mRNA splicing

The canonical model for gene expression in cells is for DNA to be transcribed into RNA, and the RNA used as a template for translation into a protein. However, in eukaryotic cells RNA must undergo significant processing to mature into a message that can be used for translation. One key processing event is splicing.

During the process of splicing, non-coding introns are removed from pre-mRNAs, and coding exons are ligated together. The spliceosome is the machinery that recognizes and splices out introns (see Figure 1 for a model of the steps of spliceosomal recruitment, and for a review see (*1, 2*)). The spliceosome is composed of many different proteins and protein-RNA complexes, that are used to recognize different sequence elements in the pre-mRNA, as well as interact with other parts of the spliceosome. It is thought that certain proteins and protein-RNA complexes recognize the intron in a sequential manner and discreet complexes form on the pre-mRNA during these different stages. The major spliceosome, which splices the majority of introns, has 5 major protein-RNA complexes, with a host of other protein factors. These large protein-RNA complexes are called snRNPs (small nuclear ribonuclear proteins), and are labeled U1, U2, U4, U5 and U6 (*3, 4*).

Early complex, or E complex (*5*), is first to form on the pre-mRNA, and is generally defined as the point in which the U1 snRNP has been recruited to the 5' splice site and the RNA portion of the U1 snRNP base-pairs with the 5' splice site; at the 3' splice site only the U2AF65/35 heterodimer and SF1 have been recruited (*6*). This is an ATP independent step, where many key nucleotide sequences in the intron are recognized for the first time, such as the 3' and 5' splice sites and the branch-point adenosine. Following E complex, U2AF65 recruits the U2 snRNP to the branch-point sequence to form A complex, which is an ATP dependent process (*7, 8*). The RNA portion of the U2 snRNP base-pairs with the branch-point sequence, with the

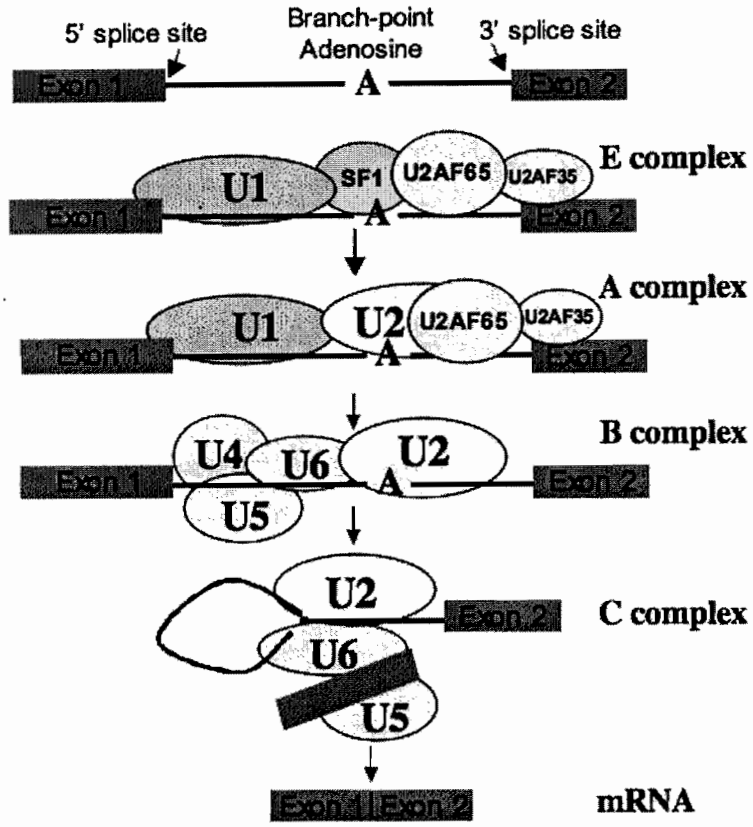

Figure 1. ***Model of spliceosomal assembly.*** Exons are represented by boxes, the intron by a line. The branch-point adenosine and 5' and 3' splice sites are noted. Spliceosomal assembly is seen to be a sequential process, at least *in vitro*, as different proteins and protein-RNA complexes (known as snRNPs, or small nuclear ribonuclear proteins) load onto the pre-mRNA. For future reference note that, in E complex, the proteins SF1, U2AF65 and U2AF35 recognize sequences in the 3' end of the intron and help recruit the U2 snRNP to the intron in A complex.

branch-point base (nearly always an adenosine) excluded from base-pairing and flipped out. Next, the U4/U5/U6 tri-snRNP is recruited to the 5' splice site to form B complex (*9, 10*). Finally, structural rearrangements in the snRNPs and the dissociation of U4 lead to the catalytic C complex, where the intron is spliced out and the exons ligated together (*9, 10*).

There are many key sequences within the pre-mRNA that help define the 3' and 5' splice sites, as well as the location of the branch-point. Currently, four canonical sequences are considered to be important for splicing (Figure 2), with a host of other auxiliary sequences in adjacent regions of the pre-mRNA also being important for splicing in certain pre-mRNAs (*11*). Both the 5' and 3' splice site have strong consensus sequences and do not vary often. The 5' splice site consensus sequence AGGTRAG base-pairs with part of the RNA portion of the U1

snRNP, while the 3' splice site YAG is bound by the protein factor U2AF35 (*11*). The branch-point sequence, which is bound by the RNA portion of the U2 snRNP, is not well conserved in higher eukaryotes. The "ideal" branch-point for base pairing with the U2 snRNP is UACUAAC, but in human, the consensus branch-point is the more degenerate sequence CURAY (*11*). The branch-point is ideally located within 50 bases of the 3' splice site, but there are many exceptions to this in higher eukaryotes. Finally, the poly-pyrimidine tract (py-tract), is the least conserved of all the sequences. It preferably consists of a run of at least 8 pyrimidine residues, and is located somewhere between the branch-point and the 3' splice site. The py-tract serves as a binding site for the protein U2AF65, and aids in E complex formation, when U2AF65 first binds the pre-mRNA (*12, 13*). Only ~40% of introns have ideal py-tracts, suggesting that this is not always a required sequence for splicing.

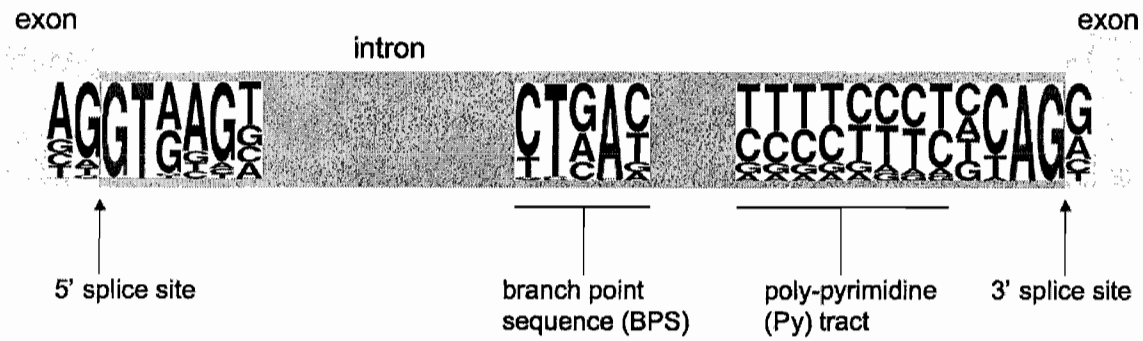

Figure 2. ***Canonical sequence elements within an intron.*** The four main sequence elements within an intron are noted. The 5' and 3' splice sites have relatively good consensus sequences, while the branch-point sequence and py-tract are much more degenerate in their sequence and location. Adapted from (*11*).

Alternative pre-mRNA splicing

During splicing, many different mRNAs can be made from a single pre-mRNA, depending on which exons are included in the final mRNA. This is called alternative splicing (for review see (*14*)). A current study of 10,000 human genes found that 74% were alternatively spliced (*15*), while other studies have found that potentially 95% of genes are alternatively spliced, to some degree (*16*). There are many different ways in which a pre-mRNA can be alternatively spliced (Figure 3). The most common type of alternative splicing is when a single exon is either chosen to be included or excluded, and is referred to as a cassette exon (Figure 3A). Another common type is for an alternate 3' or 5' splice site to be chosen and is referred to as a 3'

or 5' isoform, depending on which splice site is alternatively spliced (Figure 3B-C). The alternative splice site that is chosen can be either further into the intron, or the exon. Finally, a third common form of alternative splicing is intron retention, when an intron is simply not spliced at all (Figure 3D). This can often introduce a premature termination codon into the mRNA, leading to mRNA degradation through nonsense-mediated decay.

In many cases, sequences other than the four canonical sequences have been shown to affect alternative splicing. In some cases, specific proteins which bind these sequences. Sequences that increase splicing efficiency of a certain splice variant are known as enhancers, and more specifically as an Intronic Splicing Enhancer (ISE) if the sequence is in the intron, or Exonic Splicing Enhancers (ESE) if it is in the exon. Conversely, sequences that inhibit splicing efficiency of a certain splice variant are known as splicing suppressors. If the sequence is in the intron it is called an Intronic Splicing Suppressor (ISS), and an Exonic Splicing Suppressor (ESS) if it is in the exon (for review on auxiliary splicing signals see (*17*)).

It is assumed that these sequences are bound by alternative splicing factors to regulate which splice variant is made. In many cases, these alternative splicing factors interact with the pre-mRNA during the initial recognition and definition of the intron. At these early stages of intron recognition, these alternative splicing factors can aid or inhibit in the recruitment of spliceosomal proteins in their binding of their cognate sequence elements within the pre-mRNA.

Myotonic Dystrophy (DM) and alternative splicing

Many diseases are known to arise from the mis-regulation of alternative splicing (*18*). Some of these diseases occur because mutations alter binding sites for essential spliceosomal proteins, which leads to mis-splicing events and therefore functional changes in proteins (*19*). One of the first documented examples is from the aberrant splicing of the β-globin pre-mRNA. In patients with the disease β+ thalassemia, mutations were found that activated an abnormal 3' splice site in the pre-mRNA, which changes the β-globin protein and leads to anemia and sometimes death (*20*).

Splicing may also be mis-regulated due to changes in the function of alternative splicing factors. One of the best studied models of such a disease is Myotonic Dystrophy (DM, for a review see (*21, 22*)). In this complicated disease, there are a range of seemingly unrelated symptoms, such as myotonia, myopathy, cardiac arrhythmia, insulin resistance and cataracts. Many heart defects are also seen and are generally the most lethal symptom. This disease is

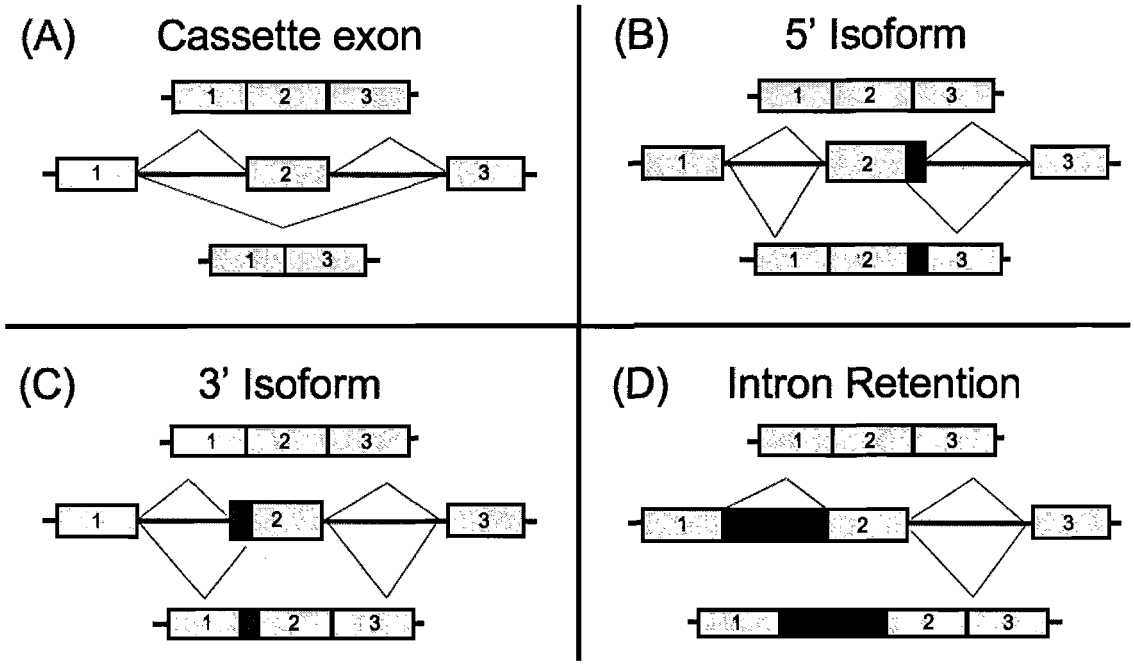

Figure 3. ***Common forms of alternative splicing.*** **(A)** Cassette exon splicing is where a single exon is either included or excluded. This is the most common form of alternative splicing. **(B-C)** 5' or 3' isoform is another relatively common form of alternative splicing where another splice site nearby is chosen. Note that this new splice site may be further into the intron as well, though it is depicted as further into the exon. **(D)** Intron retention is a form of alternative splicing where the intron is not spliced out at all.

hypothesized to occur mainly through the mis-localization of the alternative splicing factor Muscleblind-like-1 (MBNL1) (see Figure 4 for model of DM). MBNL1 is mis-localized due to the expression of aberrant CTG or CCTG repeat expansions that, on the RNA level (as CUG and CCUG repeats), form stable stem-loop structures. MBNL1 binds these stem-loops and is sequestered from its normal role of regulating alternative splicing (*23*). The pre-mRNAs that MBNL1 regulates are subsequently mis-spliced, which is hypothesized to lead to subsequent disease symptoms. As dozens of pre-mRNAs are known to be mis-spliced in DM, the sequestration of MBNL1 may have a substantial impact of DM patients. Presently, two MBNL1 regulated mis-splicing events have been strongly correlate to certain symptoms of DM. Mis-splicing of the chloride channel pre-mRNA is thought to give rise to myotonia (the inability to relax a muscle) (*24*), and the mis-splicing of the insulin receptor pre-mRNA may lead to insulin resistance (*25*). While other splicing factors have been implicated in DM (such as CUG-BP (*21*)), the mis-localization of MBNL1 is thought to be the primary cause for mis-splicing seen in DM (*26*).

Currently a few dozen pre-mRNAs are known to be mis-spliced in DM (*22, 27*), though it is likely that global studies will identify many more pre-mRNAs that are also mis-regulated. It is unclear if MBNL1 directly regulates all these mis-spliced pre-mRNAs, but MBNL1 is thought to at least directly regulated a sub-set of them. At the point when my research began, MBNL1 was only known to cross-link with one pre-mRNA target, the cardiac troponin T (cTNT) pre-mRNA. MBNL1 was seen to cross-link to an ISS directly upstream of exon 5, and cause that exon to be excluded (*28*). While sequences required for MBNL1 binding were determined for the cTNT pre-mRNA, an articulated binding site for MBNL1 was not determined in that pre-mRNA. Furthermore, a mechanistic understanding of how MBNL1 regulated the cTNT pre-mRNA (or any other pre-mRNA) also remained unclear. The second chapter of my dissertation describes my articulation of MBNL1's binding site in the cTNT pre-mRNA, and initial studies on how MBNL1 may regulate exon 5 and contains previously co-authored published material. The third chapter describes an in-depth study into the mechanism of MBNL1's regulation of exon 5. Both chapters contain previously published co-authored material.

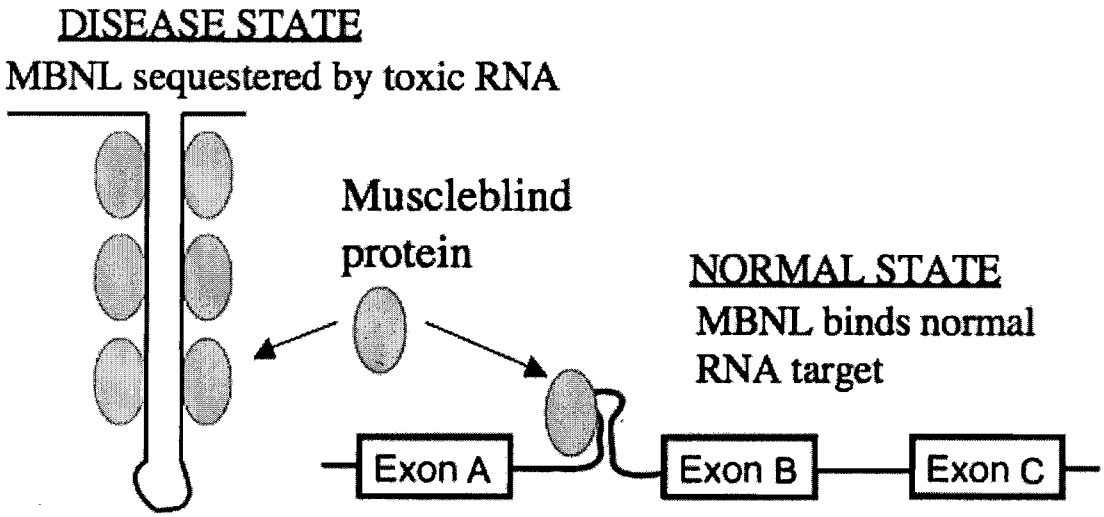

Figure 4. ***Model of Myotonic Dystrophy.*** The current model of DM is that the toxic RNA repeats (either CUG or CCUG) form stable stem-loop structures. These stem-loops aberrantly bind the protein Muscleblind (MBNL), which is an alternative splicing factor. When MBNL is sequestered, it does not bind and regulate its pre-mRNA targets, which leads to mis-splicing and subsequent disease symptoms.

Relevance to cardiovascular disease and heart development

The heart defects seen in DM are variable and are usually not defects in the heart muscle itself, but with its ability to conduct the electrical signal that keeps it beating regularly. Specific defects include atrioventricular and intraventricular conduction abnormalities, atrial fibrillation and ventricular arrhythmias. Sudden and lethal arrhythmias sometimes necessitate pacemakers or implantable defibrillators. Conduction defects do not always precede lethal arrhythmias, making sudden deaths hard to prevent (*29*).

Determining how the alternative splicing of the cTNT pre-mRNA is regulated is likely to be important for understanding a potential cause for the heart defects seen in DM. The troponin protein family consists of three types of troponin proteins and is required for muscle function (for review see (*30*)). The three types of troponin are troponin T, troponin I, and troponin C. All three types are found in all three muscle types (fast, slow and cardiac). The three proteins form a complex in muscle sarcomeres that binds both actomyosin (actin associated to myofibrils) and tropomyosin. Their function is unclear, but they all seem to jointly regulate the ATPase activity and Ca^{2+} sensitivity of actomyosin (*31, 32*). Each troponin protein has a different gene for each muscle type (fast, slow and cardiac), as well as multiple protein isoforms within each muscle type. This leads to a complex expression pattern, where various isoforms for each troponin gene are differentially expressed in each muscle type in a developmental fashion.

Cardiac troponin T (cTNT) is one of the three main types of muscle troponin T's. It has at least 4 protein isoforms in humans, which are generated through alternative splicing. At least one of the exon skipping events is developmentally controlled in human. Exon 5 is included in fetal cardiac cells, but excluded in adult cells (*33*). Exon 5 of cTNT encodes a highly acidic 10 amino acid sequence that is conserved in nearly all mammals (*32*). The retention of this sequence increases Ca^{2+} sensitivity, increases ATPase activity, increases force development of the muscle and slows relaxation of myocardial muscle (*31, 32*). The aberrant retention of exon 5 in adult cardiac cells can have severe consequences for heart health, as exon 5 retention (or the analogous exon in other organisms) has been linked to cardiomyopathy and spontaneous heart failure in human, cat, dog, guinea pig and turkey (*32*).

While it is unclear how aberrant retention of this exon in adult heart cells may cause cardiomyopathy, it is hypothesized that retention of this isoform leads to arrhythmia by desynchronization of heart muscle contraction, which is an important factor in cardiomyopathy (*32*). Conduction abnormalities and arrhythmia of the heart are arguably the most lethal symptom

in DM, while the heart muscle itself seems to be normal (which is unanticipated in a muscular dystrophy with heart defects) (*29*). As exon 5 of the cTNT pre-mRNA is drastically increased in DM, it seems likely that the mis-splicing of this one pre-mRNA could solely account for the specific and unusual arrhythmia heart defects in the disease. In animal models where increased exon 5 had high retention, similar heart defects were seen when compared with DM patients (*29, 31, 32*). By understanding how this process is regulated, we would gain insight into one of the largest causes of death in DM, as well as the role of alternative splicing in heart development. The third chapter of this dissertation details my studies on how MBNL1 mechanistically regulates exon 5 exclusion by its competition with the essential splicing factor U2AF65 and contains previously published co-authored research.

Potential therapeutic avenues for treating Myotonic Dystrophy

Understanding how MBNL1 mechanistically regulates alternative splicing is a vital step towards understanding Dm and creating a cure for the disease. Now that we have a better understanding of the disease state and the role of MBNL1's sequestration in causing DM, it is possible to hypothesize how the disease may be treated. A cure for DM would be for the CTG or CCTG expansions to be deleted from the patients genome, but this is infeasible with the current genetic tools. Other recent therapeutic approaches for DM1 have ranged from over-expression of MBNL1, RNA interference against the CUG repeats, to targeted degradation of the mutant DMPK transcript with an RNA ribozyme (*34-37*). Aside from the over-expression of MBNL1, another possible approach to overcoming the sequestration of MBNL1 is to identify small molecules that specifically bind the CUG repeats to competitively release the sequestered MBNL1. The fourth chapter of this dissertation focuses on determining if a small molecule could compete with MBNL1 for binding of CUG repeats and alleviate mis-splicing seen in various DM model systems and contains previously published co-authored material.

CHAPTER II

MBNL1 BINDS SIMILAR RNA STRUCTURES IN THE CUG REPEATS OF MYOTONIC DYSTROPHY AND ITS PRE-MRNA SUBSTRATE CARDIAC TROPONIN T

Contribution note: this chapter is previously published (Warf and Berglund, 2007) (84). The other contributor, J. Andrew Berglund, helped with experimental design, data analysis and manuscript preparation.

Introduction

Myotonic Dystrophy (DM) is a genetic disorder with multisystemic symptoms that include myotonia, cardiac arrhythmia, insulin resistance and muscular weakness. There are two subtypes of Myotonic Dystrophy: DM1 and DM2. DM1 has been linked to a $(CTG)_n$ repeat expansion in the 3' untranslated region (3' UTR) of the *DMPK* gene. DM2 has been linked to $(CCTG)_n$ repeats expansion in intron 1 of the *ZNF9* gene. The genetic mutations in each subtype are in two unrelated genes on different chromosomes. The symptoms observed in the two subtypes are remarkably similar, with the only molecular commonality being the repeat expansions. The similarity in symptoms and sequence motifs as well as the fact that both repeats are non-coding indicate a common mechanism for both subtypes.

CUG expanded repeats fold into extended stem-loop structures, with guanosines and cytosines forming base pairs, while the uridines form mismatches (*38-40*). Biochemical and structural studies have shown that extended helical regions of the stem-loops are primarily A-form in structure and are thermodynamically stable (*40, 41*). The CCUG repeats of DM2 also fold into an extended stem-loop structure. It is currently thought this stem-loop consists of two adjacent guanosine-cytosine base pairs and two adjacent cytosine-uracil mismatches (*42, 43*). However, it is possible for the RNA to anneal in another structure in which single guanosine-cytosine base pairs are interspersed with uracil-uracil and cytosine-cytosine mismatches (see

Figure 3A for schematic) and it is unclear if one structure predominates. Thermodynamically, the first structure is the only one that is predicted (*44*).

One proposed mechanism for the disease is that, upon transcription, the CUG and CCUG repeats sequester RNA binding proteins from their normal cellular functions. It is hypothesized that the specific sequestration of the RNA binding protein MBNL1 (Muscleblind-like) primarily leads to DM symptoms. Supporting this hypothesis is a model in which expression of 250 non-coding CUG repeats causes symptoms similar to patients with DM (*45*). The link between MBNL11 and DM was strengthened with a mouse knockout model, where the MBNL11 gene was inactivated through deletion and the mice developed many key symptoms of DM (*46*).

The muscleblind family of proteins was originally identified in *Drosophila melanogaster* as a gene required for muscle development and eye differentiation (*47*). For a review of muscleblind see (*48*). There are three muscleblind paralogues in human, named MBNL11-3. Of the three human MBNL1 proteins, MBNL11 and MBNL12 are more abundant and have been shown to co-localize with CUG and CCUG repeats in the nucleus, forming nuclear foci in both DM1 and DM2 (*23, 49-55*). All three muscleblind proteins are similar in sequence, and appear to have similar functions, as they can regulate alternative splicing in tissue culture (*26, 28, 56*). However, MBNL12 has also been shown to function in RNA localization in the cytoplasm (*57*). Less is known about MBNL13, but it is possible that MBNL11 and MBNL13 may act antagonistically to each other when regulating gene expression that is involved in muscle differentiation (*58*).

The RNA binding protein, CUG-BP also has an important role in DM pathogenesis. CUG-BP and hnRNP H are over-expressed in the presence of expanded CUG repeats and are involved in controlling the alternative spicing of many of the same genes regulated by MBNL1 (*25, 56, 59, 60*). An antagonistic relationship exists between MBNL1 and CUG-BP for several regulated exons; where one protein acts as a positive regulator while the other protein acts as a negative regulator. However, the actual role of each protein appears to depend on the specific pre-mRNA and splice-junction in question. For example, for the fifth exon of the cardiac troponin T (cTNT) pre-mRNA, CUG-BP is a positive regulator of exon five inclusion while MBNL1 is a negative regulator of exon five inclusion (*28, 60, 61*). Conversely, for exon 11 of the insulin receptor pre-mRNA, the roles of CUG-BP and MBNL1 have been reversed and CUG-BP is a negative regulator while MBNL1 is a positive regulator (*25, 26, 56, 62*).

It is currently thought that the relative levels of these two splicing factors and other splicing factors leads to specific pre-mRNA splice patterns (*63*). As these two factors compete in

their regulation of splicing, the mis-splicing seen in DM has been hypothesized to be due to both decreased levels of MBNL1 and/or increased levels of CUB-BP. Supporting both models, mice in which MBNL1 is knocked out show DM symptoms, as do mice in which CUG-BP is over-expressed (*46, 61, 64*). However, it has recently been shown in a tissue culture model that loss of MBNL1 has a more drastic effect on mis-splicing, while the increased levels of CUG-BP had only a secondary effect (*26*). This suggests that sequestration of MBNL1 to the CUG and CCUG repeats is the primary cause for the mis-splicing observed in DM.

To understand the function of MBNL11 in the disease state and during pre-mRNA splicing, we characterized the RNA binding activity of purified recombinant MBNL11. We tested binding to CUG and CCUG repeats, to a series of RNAs where portions of the CUG repeat are mutated, and to a fragment from the cTNT pre-mRNA. We found that MBNL11 preferentially binds short helical A-form RNA regions in both the CUG and CCUG repeats and the cTNT pre-mRNA structure. In all cases the helical structures contain similar pyrimidine-pyrimidine nucleotide mismatches. This suggests that MBNL11 may recognize and bind similar structures in both the pathogenic repeats and in its pre-mRNA targets.

Results

The elongated zinc finger domains of MBNL1 are sufficient for RNA binding

To characterize the RNA binding of MBNL11, we used a truncated version of MBNL11 (containing amino acids 1-260). Recombinant protein was expressed in *e. coli* BL21* cells, and purified using and GST affinity tag and ion exchange chromatography. When compared to full-length MBNL1(1-382), the truncated version of MBNL11(1-260) bound three different RNAs with similar affinity: an RNA with 90 CUG repeats, a shortened CUG_4 construct, and to a region of the cardiac troponin T pre-mRNA (Figure 1 and Figure 5). Kino and colleagues also previously found that two truncated versions of MBNL11 (a.a. 1-248 and 1-269) bound RNA substrates in a three-hybrid assay better than a full length version of MBNL11 (*65*). MBNL11(1-260) contains all four zinc fingers of the full length protein (Figure 1A), and was used for all of the studies presented here and will be referred to as MBNL1 throughout the remainder of the text.

Biophysical methods were used to characterize MBNL1. Analytical ultracentrifugation (AUC) with purified recombinant MBNL1 (Figure 1B) was used to measure a molecular weight of 28.2 kDa (predicted molecular weight is 28.5 kDa). This indicates that MBNL1 is a monomer

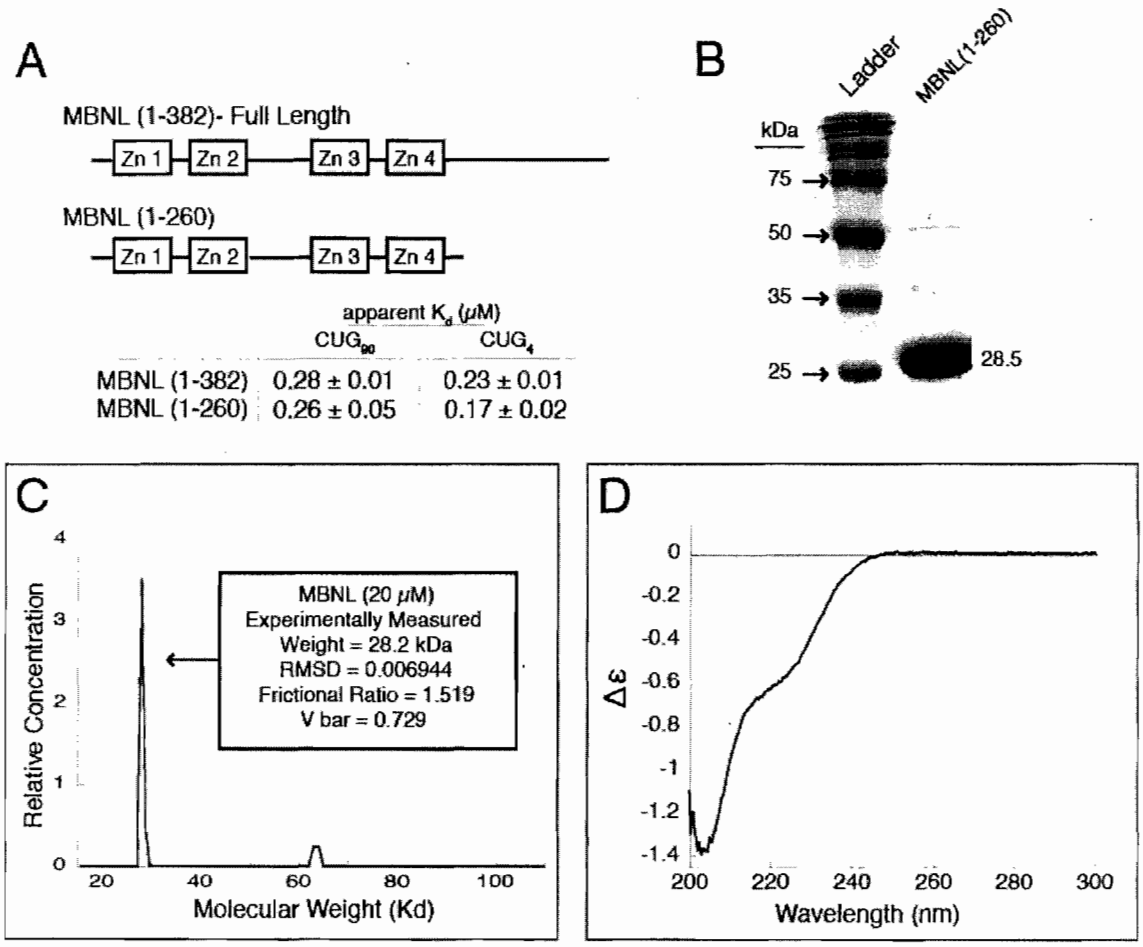

Figure 1. ***Structural characterization of MBNL1.*** (**A**) A schematic of MBNL1 showing the truncated form MBNL1(1-260) used in these studies. (**B**) Recombinant expressed MBNL1 has a molecular weight of 28.5 kDa on a 10% SDS-PAGE gel. Note there is a small amount of co-purifying contaminant with a molecular weight of approximately 55-60 kDa. (**C**) Analytical ultracentrifugation of MBNL1 shows that it sediments with the weight of a monomer. MBNL1 has a slightly elongated shape, as its frictional ratio is 1.519. (**D**) Circular dichroism spectra of MBNL1, showing a major peak at 203 and a minor peak at 220 nm, indicating that a portion of MBNL1 is disordered and MBNL1 lacks significant a-helical content in its structure (units Δε are molar circular-dichroic absorption).

in solution (Figure 1C), as the zinc finger domains do not seem to mediate any oligomerization. It may also be possible that the C terminal domain (amino acids 261-382) may mediate oligomerization. The frictional ratio of MBNL1 was 1.5189. Globular proteins have a ratio of 1.2 (*66*), and an increased ratio is consistent with a protein that is elongated in solution. A co-purifying contaminant (seen in Figure 1B), of approximately 65 kDa was also seen to sediment.

However, the main peak at 28.2 kDa accounted for over 95% of the sedimented protein signal (Figure 1C).

Circular dichroism (CD) was used to probe the secondary structure of MBNL1 (Figure 1D). The weak α-helical signal at 220 nm and the strong signal at 203 nm (indicative of disorder in protein structures) suggests that MBNL1 does not have significant α-helical structure, although it is possible that MBNL1 contains some or significant β-sheet structure, as the signal for this structure was undetectable due to interference with the buffer. The addition of a known RNA substrate did not alter the signal at 220 nm, suggesting no new α-helices were forming upon RNA binding (data not shown). Mutations of individual cysteines within the zinc finger domains reduce the stability of MBNL1 and reduce its RNA binding affinity (data not shown), indicating that zinc fingers competent to bind zinc are required for RNA binding. The addition of 5 μM zinc chloride did not alter the 203 nm to 220 nm ratio, indicating that the added Zn^{++} caused no structural changes in the protein and therefore that the four zinc finger domains were already saturated with zinc ions (data not shown).

MBNL1 binds moderate and short CUG expansions with similar affinity

To identify a minimal CUG repeat construct that could be used for characterizing the binding specificity of MBNL1, we started with an RNA substrate containing 90 CUG repeats (CUG_{90}); this RNA was then truncated until a minimal RNA substrate capable of binding MBNL1 was identified. In all assays, the protein was always in at least a ten fold excess of the total number of CUG triplets, to ensure that multiple binding sites on the longer CUG expansions would not artificially enhance the apparent affinity of MBNL1 for the longer repeats. To stabilize the shorter CUG repeats that contained 8 repeats or fewer, an ultra stable UUCG tetraloop was used to cap the short stem-loops (*67*). Stabilizing the short CUG repeats was likely necessary because Miller and colleagues didn't observe binding with short CUG repeats (*51*), possibly because in their experiments the shorter repeats were not forming stem-loops.

A stem-loop containing two pairs of CUG repeats separated by the tetraloop (CUG_4) bound MBNL1 with similar affinity compared to CUG_{90} (Figure 2A versus 2B). The slightly reduced affinity of MBNL1 for CUG_{90} compared to CUG_4 may be because multiple MBNL1 proteins need to bind to CUG_{90} to cause a shift suggesting the apparent K_d for this RNA is likely to be lower than 260 nM. Our truncation studies of the CUG repeats indicate that the minimal binding site for MBNL1 is six base pairs or less.

Thermal melts were performed to verify that the small RNA constructs formed the predicted stem-loop structures. Discrete thermal transitions were seen for all RNAs constructs, indicating that a stable stem-loop structure was indeed formed (see Table 1 for a listing of estimated T_m values). In addition, the T_m of CUG_4 was tested and found to be concentration independent up to 2 μM, indicating that only an intramolecular structure formed at these concentrations at the RNA concentrations used during our experiments (as opposed to intermolecular structures which may start to form at higher concentrations and would likely have a different melting curve).

MBNL1 recognizes both the mismatch and Watson-Crick base-pairs within the CUG repeat stem-loop

To determine the importance of the U-U mismatch in the CUG stem-loop RNA, we replaced the U-U mismatch with other mismatches or placed Watson-Crick base pairs in this position (see Figure 2). MBNL1 binds to sequences containing any pyrimidine-pyrimidine mismatch in any combination with similar affinity (apparent K_d values range from 140 to 170 nM) (Figure 2B, 2C and Table 1). The substitution of purine-purine mismatches (A-A, G-G and A-G) for U-U mismatches all moderately inhibit the binding of MBNL1 by 10 to 20-fold (Figure 2D, 2E and Table 1). Replacement of the U-U mismatch with C-A mismatches only reduces binding approximately 4-fold while a G-U wobble base pair reduces binding more than 20-fold (Figure 2F, 2G and Table 1). Replacing the U-U mismatches with Watson-Crick base pairs almost completely abolishes MBNL1 binding (Figure 2F), indicating the mismatch plays an essential role in MBNL1's binding.

The identity and location of the C-G and G-C base pairs in the CUG repeats are also important for MBNL1 binding. When the G-C base pairs in CUG_4 were changed such that cytosines were all one strand and the guanosines were all one strand, MBNL1 binding was reduced 15-fold (Figure 2H, lanes 1-6). Changing the polarity of the RNA from CUG to GUC completely eliminated binding (Figure 2H), demonstrating that the polarity and order of the base pairs in the sequence is important. Finally, an RNA with UCA repeats that form all Watson-Crick base pairs capped with a UUCG loop does not interact with MBNL1 (Figure 2H).

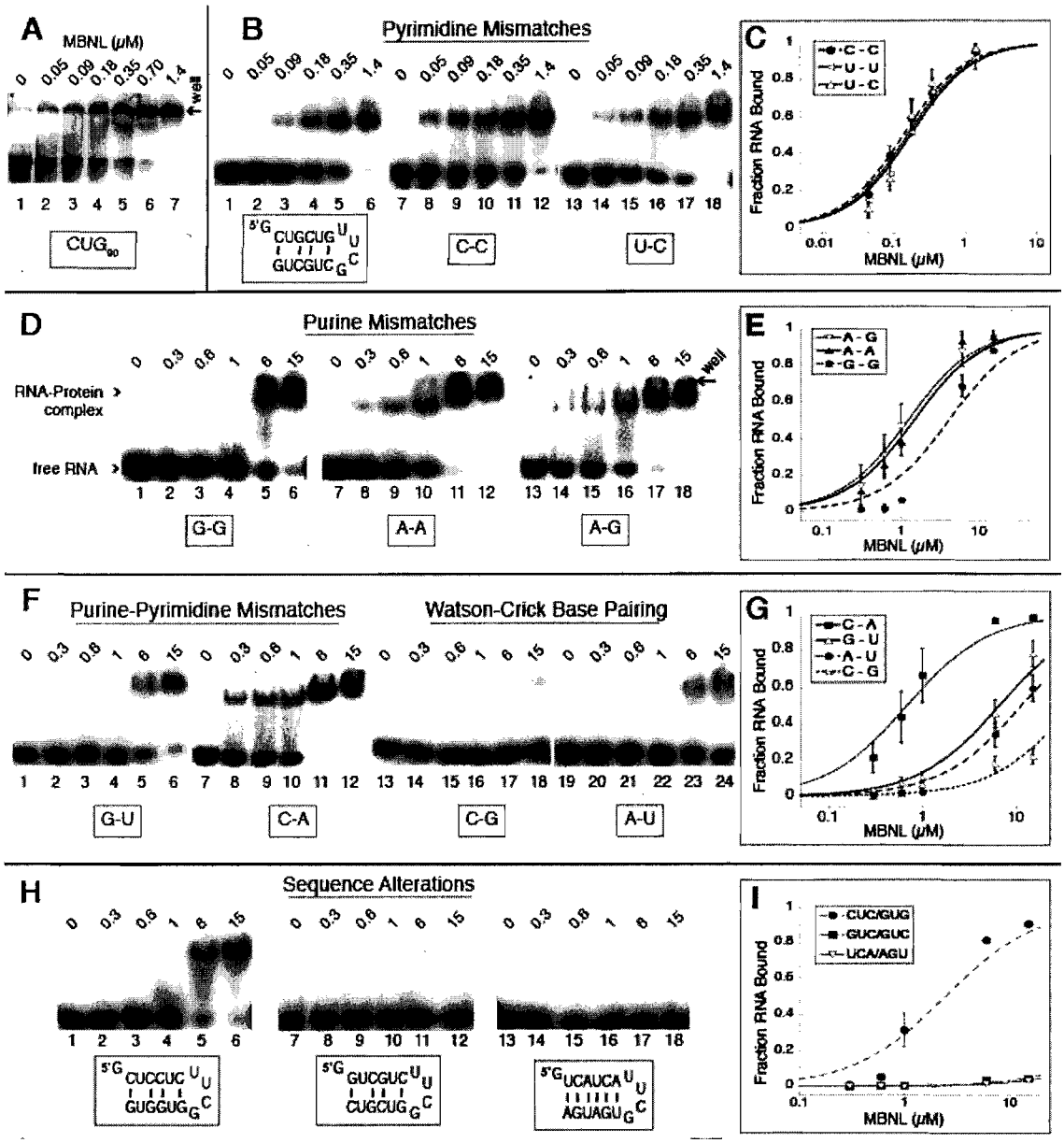

Figure 2. ***MBNL1 binds short CUG repeats with similar affinity to longer repeats and the presence of mismatches is necessary for binding.*** (**A**) MBNL1 binding CUG_{90} repeats (lanes 1-6) with an apparent K_d of 230 nm. The concentration of MBNL1 is shown above each lane in μM in the gel shift assay. **(B)** MBNL1 binding to pyrimidine-pyrimidine mismatches. The sequence of the $(CUG)_4$ RNA is shown below lanes 1-6. The boxed C-C (lanes 7-12) and U-C (lanes 13-18) represent the replacement of the U-U mismatches with these mismatches. **(D)** MBNL1 binding to purine-purine mismatches in place of the U-U mismatch, boxed sequence indicates mismatch replacement base pairs. (**F**) MBNL1 binding to G-U (lanes 1-6) and C-A (lanes 7-12) mismatches and Watson-Crick base pairs in place of the U-U mismatches (**H**) Binding of MBNL1 to RNAs containing sequence alterations to the Cytosine and Guanine positions in CUG_4 and a control RNA with no sequence similarity to CUG repeats. (**C, E, G, I**) Binding curves.

TABLE 1. Apparent K_d and T_m values for all tested RNA constructs

RNA construct	Apparent K_d (μM)	T_m (°C)
CUG_{90}	0.26 ± 0.05	
CUG_8	0.14 ± 0.04	60
CUG_6	0.14 ± 0.04	61
CUG_4	0.17 ± 0.02	58
$CCUG_8$	0.07 ± 0.01	41
$CCUG_6$	0.09 ± 0.01	41
$CCUG_4$	0.12 ± 0.02	42
$CCUG_{6-2}$	0.06 ± 0.01	42
CXG_4 mismatch screen		
Pyrimidine mismatch		
U-U	0.17 ± 0.02	58
C-C	0.14 ± 0.03	53
U-C	0.15 ± 0.03	53
Purine mismatch		
A-G	1.4 ± 0.2	65
G-G	4.2 ± 0.4	72
A-G	1.4 ± 0.4	67
Purine–pyrimidine mismatch		
C-A	0.7 ± 0.1	69
G-U	6.5 ± 0.6	79
Watson–Crick base pair		
C-G	>15[a]	>95
A-U	11.5 ± 0.3	83
CUG_4 sequence alteration		
CUC_2/GUG_2	2.5 ± 0.3	53
GUC_2/GUC_2	>15[a]	54
UCA_2/AGU_2	>15[a]	63
Intronic target		
cTNT 50mer	0.022 ± 0.004	52
cTNT 50mer mutant	2.1 ± 0.3	34
cTNT 32mer	0.05 ± 0.01	58
cTNT 32mer G-U flip mutant	2.1 ± 0.3	61
Potential stem #1	1.00 ± 0.09	47, 83[b]
Potential stem #2	0.22 ± 0.02	61
Potential stem #3	1.20 ± 0.07	54

[a] K_d exceeds the highest protein concentration tested.
[b] This RNA had two discreet thermal transitions.

Thermal melts were again performed on all RNA constructs, to verify secondary structures formed (Table 1). All RNA structures were relatively stable, with T_ms greater than 50 °C. Analysis of the range of the transitions indicates that all RNA structures should be fully intact at 25 °C, the temperature at which all binding studies were performed.

MBNL1 binds short CCUG stem-loops

A series of short CCUG RNA constructs with 4, 6 and 8 repeats were designed and named $CCUG_4$, $CCUG_6$ and $CCUG_8$, respectively (all which also contained a tetraloop to act as a cap). MBNL1 bound all three RNA constructs with similar affinity and on average with 2-fold stronger affinity compared to CUG repeats (Figure 3 and Table 1). CCUG expansions can anneal in two possible structures (Figure 3A). Presently, CCUG expansions are thought to be in the first structure (*42, 43*), where two C-G base pairs are interspaced by two U-C mismatches. CCUG repeats are thermodynamically predicted by Mfold to be in this structure (*44*).

However, another structure is possible in which G-C base pairs are flanked by alternating U-U and C-C mismatches. A $CCUG_6$ construct was designed that forced the stem-loop to anneal in this alternate structure ($CCUG_{6\text{-}2}$, Figure 3A). MBNL1 bound this construct with slightly enhanced affinity compared to $CCUG_6$, indicating that MBNL1 binds CCUG expansions in either configuration but may have a slight preference for the conformation containing the alternating U-U and C-C mismatches. Melting studies performed on the CCUG stem-loops showed lower T_ms than those for the CUG stem-loops (Table 1), which reflects the decreased stability of these structures due to the increased proportion of mismatches in the structure.

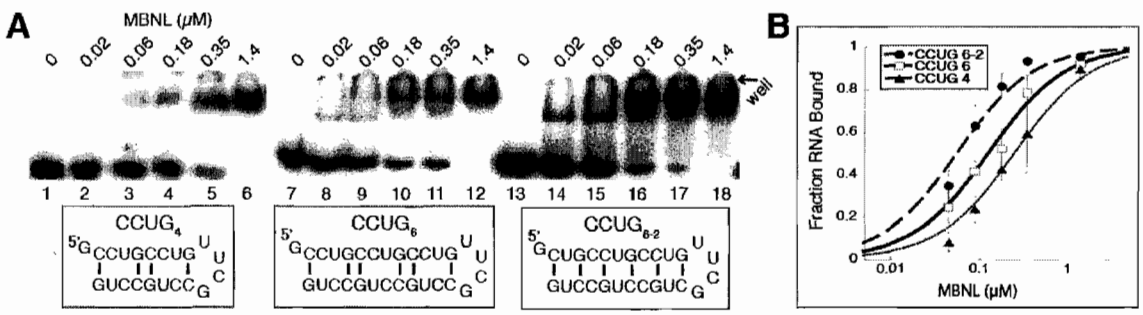

Figure 3. ***MBNL1 binds CCUG expansions with high affinity, in two possible registers.*** **(A)** MBNL1 binding to $CCUG_4$ (lanes 1-6), $CCUG_6$ (lanes 7-12) and CCUG repeats in an alternate register (labeled $CCUG_{6\text{-}2}$) in lanes 13-18. Concentration of MBNL1 is in μM labeled above each lane. **(B)** Binding curve for CCUG constructs.

MBNL1 binds a helical A-form structure within the cTNT pre-mRNA

It has been previously shown that MBNL1 can be cross-linked to the 3' end of the fourth intron of the cardiac troponin T (cTNT) pre-mRNA *in vitro* (*28*). However, quantitative binding to this substrate has not been performed. We found that MBNL1 binds a 50 nucleotide fragment from the 3' end of the intron (cTNT 50mer) with the highest affinity of any of the RNAs we tested, with an apparent K_d of 22 nM (Figure 4C and D). This 50mer spans nucleotide residues 8 through 58 directly upstream of exon 5 (Figure 4A). The full length MBNL11 (1-382) bound the cTNT 50mer with similar affinity to MBNL1(1-260) (Figure 5).

Because MBNL1 was found to bind short stem-loop RNAs, we hypothesized that the cTNT 50mer might also contain a short stem-loop recognized by MBNL1 (Figure 4B). UV melting shows that the cTNT 50mer unfolded with a single transition and a T_m of 52 °C (Figure 4E), indicating at least one structural element within the RNA. As CUG repeats are known to be A-form (*40*), we analyzed the circular dichroism (CD) spectrum of the cTNT 50mer and the CUG_4 stem-loop RNA to determine if this structure was also A-form in nature. A-form duplex RNA has a characteristic peak between 260-270 nm in a CD spectrum, while single stranded RNA has a peak at approximately 275 nm (*68, 69*). The cTNT 50mer (1 μM) had a peak at 271 nm and CUG_4 (1 μM) had a strong peak at 269 nm (Figure 4F). These values are in the higher range of the duplex A-form range, but currently only the CD spectrum of purely duplex or single stranded RNA has been characterized, and the mismatches in the CUG_4 stem and potential mismatches in the cTNT stem structures may shift the peak. Because the cTNT 50mer has a peak at a similar wavelength, it indicates that its stem structure is similar to the CUG_4 stem.

Addition of MBNL1 (1 μM) to the cTNT 50mer decreases the signal by 23% with the peak shifting to 272.5 nm. Both the reduction in signal and shift suggest that MBNL1 binding alters the RNA structure in some way that may reduce the base stacking of the RNA. MBNL1 has no signal at 270 nm (Figure 4F), meaning the change in signal is due solely to structural changes in the RNA. Approximately 85% of the RNA is predicted to be bound at these RNA and protein concentrations, indicating that the peak shift and the 23% decrease in signal is not stoichiometric, but only a partial decrease in signal by the majority of the RNA population.

The addition of 1 μM MBNL1 to CUG_4 similarly decreased the CD signal by 20% and caused a peak shift of 2.5 nm, from 269 to 271.5 nm (Figure 4G). The GUC_4 (1 μM) construct was tested and had a signal at 270 nm, which was unchanged in the presence of 1 μM MBNL1

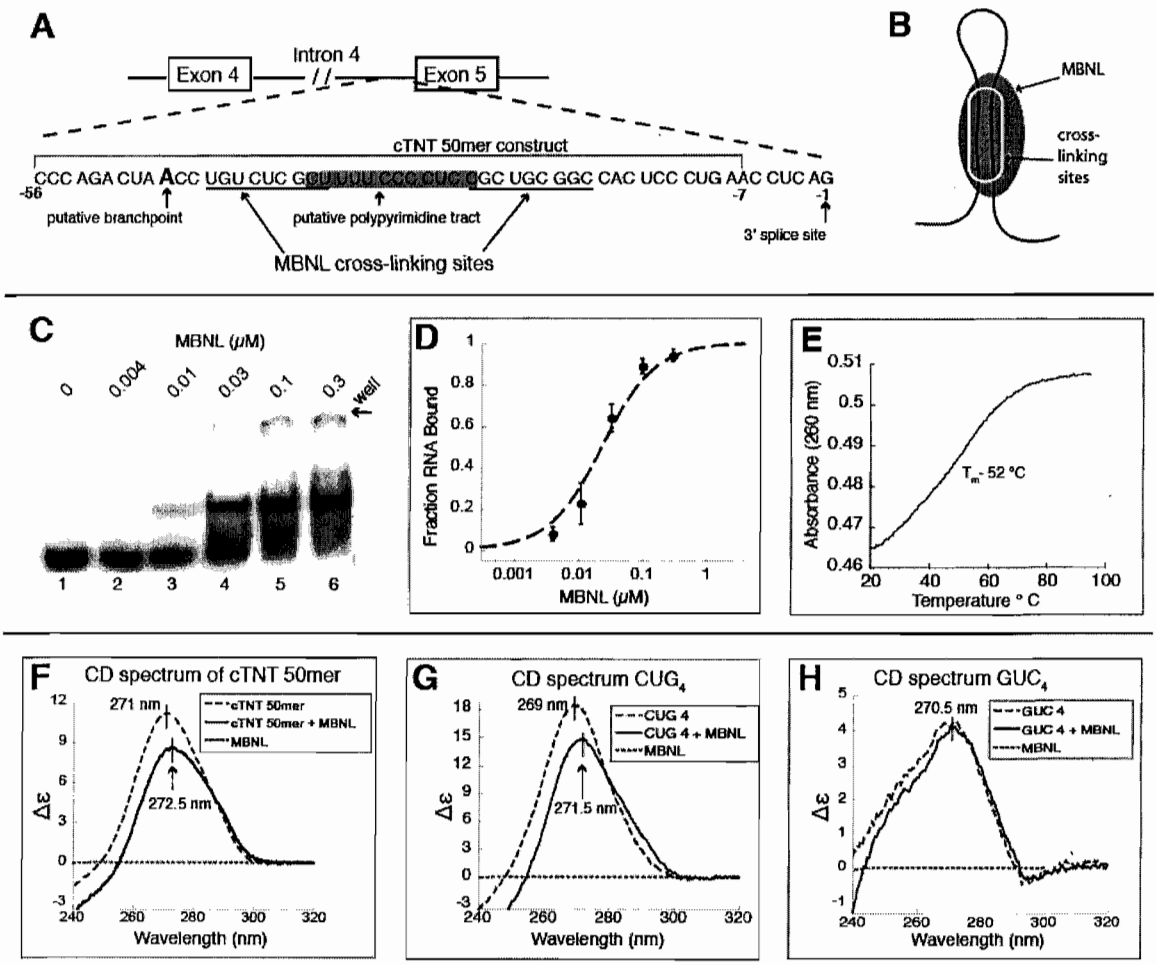

Figure 4. ***MBNL1 binds a structured region in the 3' end of intron 4 in the human cardiac troponin T (cTNT) pre-mRNA.*** (**A**) A schematic of intron 4 of the cTNT pre-mRNA. The cTNT 50mer used for binding studies is indicated. (**B**) A schematic model of MBNL1 binding this RNA as stem-loop. (**C**) Binding of MBNL1 to cTNT 50mer with the concentration of protein labeled above each lane. (**D**) Binding curve of cTNT 50mer. (**E**) Thermal melt of the cTNT 50mer, with a T_m of 52 °C. (**F**) Circular dichroism (CD) spectra of the cTNT 50mer with and without MBNL1 present (units $\Delta\varepsilon$ are molar circular-dicroic absorption). (**G**) CD spectra of CUG_4 with and without MBNL1 present. (**H**) CD spectra of GUC_4 with and without MBNL1 present.

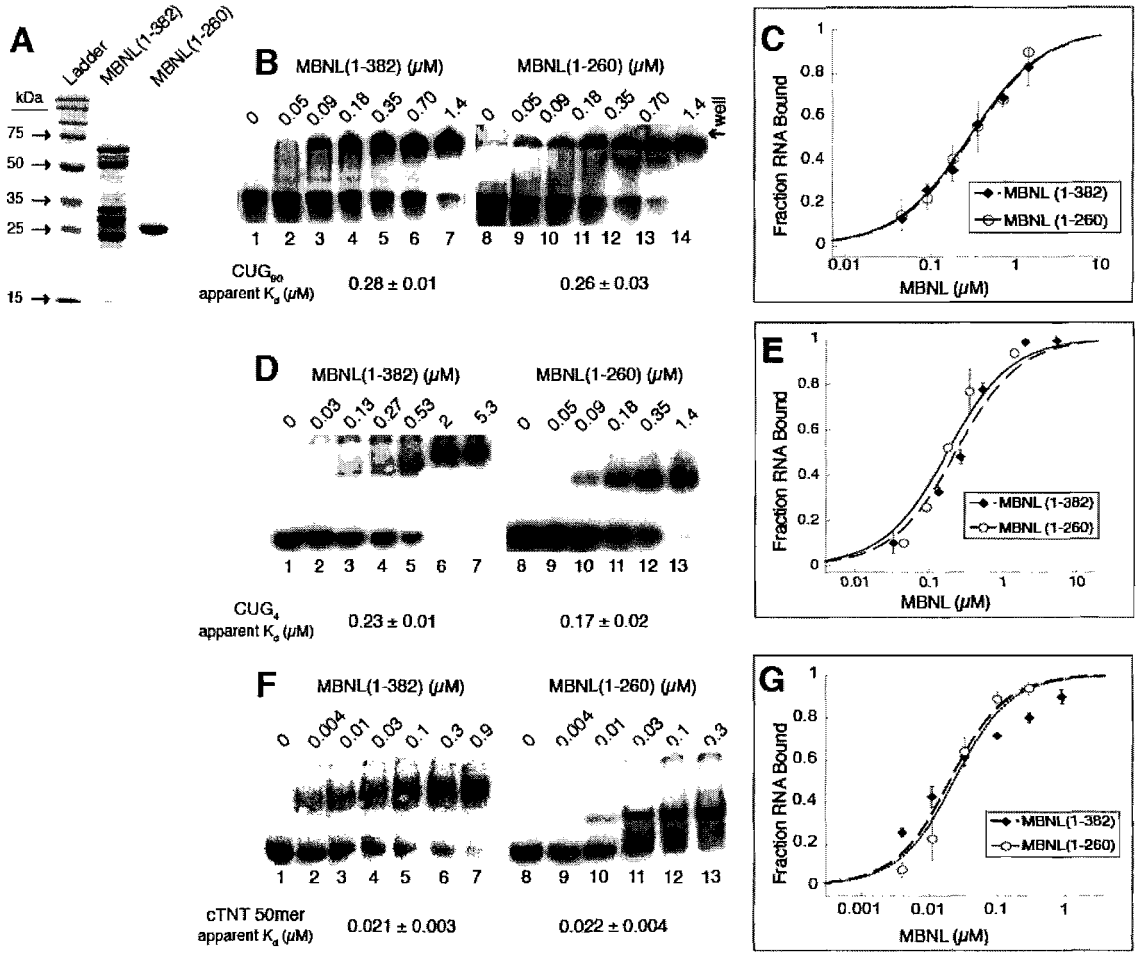

Figure 5. ***A truncated version of MBNL1 has similar RNA binding activity compared to the full length protein.*** **(A)** Full length recombinantly expressed MBNL1(1-382) with the GST affinity tag has a molecular weight of 67.5 kDa and MBNL1(1-260) without the affinity tag has a weight of 28.5 kDa on a 10% SDS-PAGE gel. Note, in the MBNL1(1-382) sample there are moderate amounts of 4 smaller degradation products with molecular weights of approximately 55, 50, 30 and 27 kDa, as well as small amounts of contaminating GST with a molecular weight of 26 kDa. The tag was kept on MBNL1(1-382), as it was more stable and degraded less, which was not a problem with MBNL1(1-260), where the affinity tag was removed. The concentration of MBNL1(1-382) could only be approximated, as the sample is not completely pure. The approximation made was that equal concentrations of MBNL1(1-382) and MBNL1(1-260) were loaded on the gel shown. **(B)** Binding assay with CUG_{90} for MBNL1(1-382) and MBNL1(1-260). **(C)** Quantitation of CUG_{90} gels. **(D)** Binding assay with CUG_4. **(E)** Quantitation of CUG_4 gels. **(F)** Binding assay with 50 base region of the cTNT pre-mRNA (cTNT 50mer). **(G)** Quantitation of cTNT 50mer gels.

(Figure 4H). The relative amplitude differences between cTNT 50mer, CUG_4 and GUC_4 are likely due to differences in base stacking due to sequence differences.

MBNL1 binds a stem-loop containing mismatches at the 3' end of the 4th cTNT intron

To determine the structure of the putative stem-loop in intron 4 of the cTNT pre-mRNA, mung bean nuclease was used to probe the secondary structure. A shortened version of the cTNT RNA (cTNT 32mer) was used in the structure probing assay because the cTNT 50mer was less stable in these assays and degraded more easily. MBNL1 binds the cTNT 32mer with high affinity, with an apparent K_d of 50 nM (Table 1).

Mung bean nuclease is a non-specific single-stranded cutter that cleaved residues 12 to 22 and 30-31, while most of the other residues were not. This suggests that residues 12-22 are likely single-stranded and located within a loop, with the final three residues being in a 3' tail (Figure 6A). Residues 4-11 and 22-29 are protected except position 8 is cleaved at a low level (lane 3), indicating these two regions form the stem (Figure 6B).

This cleavage pattern suggests the two MBNL1 sites identified through cross-linking (*28*) come together to form the stem. Therefore, it appears that these two separate sites, underlined in Figure 5B, are actually one individual site (Figure 6B). Three different base pairing configurations are possible for this stem containing slightly different base pairing and mismatches possibilities (Figure 6C-D). To determine which (if any) of the three stems are favored by MBNL1, we created three stems with the different base pairing configurations capped with the UUCG tetraloop (Figure 6D). MBNL1 bound all three RNA structures, but it clearly prefers stem #2. MBNL1 bound this RNA with 5-6 fold greater affinity compared to stems #1 and #3, with an apparent K_d of 0.22 μM for stem #2. This K_d is in the same range as the K_d for the CUG_4 RNA, although it is 4-fold weaker than the endogenous cTNT 32mer. This suggests that Stem #2 is the preferred structure that MBNL1 binds in the endogenous pre-mRNA target but that loop or tail regions may make additional contacts with MBNL1. Alternatively, the UUCG cap may perturb the structure of RNA in a way that negatively effects MBNL1 binding.

Thermal melts were performed on these stem-loops, and stem #2 was found to have the highest T_m, suggesting it is the most stable stem and likely the biologically relevant structure. However, the stabilities of these stems were likely altered by the presence of the UUCG cap. For instance, stem #1 had a second discreet transition at 83 °C (Table 1). It is likely that stem #1 has two separable structural elements, separated by the C bulge and U-C mismatch (Figure 6C). The

UUCG cap is known to form a stable fold with just two adjacent G-C base pairs (*67*), making it likely that the second structural element of stem #1 is strongly stabilized by this cap, which is not part of the wild-type sequence.

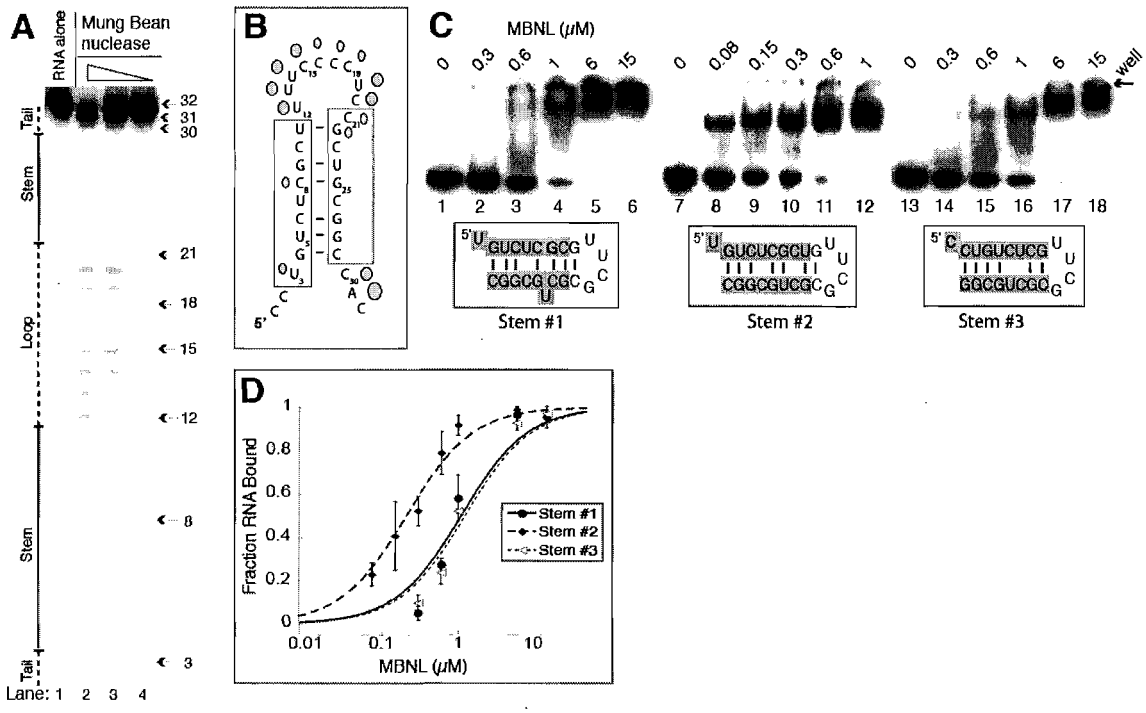

Figure 6. ***MBNL1 binds the cTNT 32mer as a stem-loop.*** (**A**) Mung bean nuclease cleavage pattern of cTNT 32mer. The concentration of mung bean nuclease is 100, 10, and 1 units per μL, in lanes 2-4 respectively. (**B**) Schematic of the likely stem-loop within cTNT intron 4, showing cleavage locations of mung bean nuclease. Larger shapes indicate strong cleavage events while smaller shapes represent weak cleavage events. The cross-linking sites determined previously are boxed (*28*). (**C**) Binding curve of the three potential stem structures of the cTNT intron 4, showing MBNL1 prefers structure #2. (**D**) Gel shifts showing MBNL1 binding to the three potential stems of cTNT intron 4. Concentration of MBNL1 is labeled above each lane, note the lower concentrations in lanes 8-12 compared to lanes 2-6 and lanes 14-18. Sequences highlighted in grey are sequences from the cTNT intron in the three different potential base pair and mismatch configurations. The non-grey sequence is the tetraloop cap and an additional base pair to stabilize the structure if necessary.

Mutations that destabilize the stem in the cTNT 4th intron also significantly reduce binding of MBNL1

We hypothesized that the stem loop structure in the cTNT RNA was critical for MBNL1 binding and that mutations that destabilize the stem loop would abolish MBNL1 binding. Previously, Ho and colleagues found that four simultaneous Guanosine point mutations (which will be referred to as the 4G construct) in this region of the cTNT 4th intron reduce MBNL1's ability to cross-link to this RNA and the ability of MBNL1 to negatively regulate the inclusion of the downstream exon is eliminated (*28*). These four point mutations change four of the six guanosines that participate in base pairing or wobble base pairing in the stem loop we identified (Figure 7A).

The 4 mutations in the cTNT 50mer reduced binding of MBNL1 approximately 100-fold, the K_d changed from 23 nM for the wild-type cTNT RNA to 2.1 μM for the mutant RNA (Figure 7B-C, Table 1). This result is consistent with the nearly complete reduction of MBNL1 cross-linking to this mutant RNA (*28*). A UV melt showed that these mutations significantly reduced the stability of the RNA structure within the cTNT RNA, as expected. The T_m was shifted from 52°C to 34°C for the mutant RNA (Figure 7D). The CD spectra of the cTNT 50mer mutant was also different from the wild-type cTNT RNA, with the mutant having a reduced signal and the peak shifted to 275 nm, into the known single stranded RNA wavelength region (Figure 7E). Addition of MBNL1 (1 μM) reduced the peak only 4% (compared to a 23% reduction of the wild-type sequence) and did not shift the peak wavelength. These results suggest that the stem-loop structure within the cTNT pre-mRNA is the recognition site for MBNL1.

Regulated splicing by MBNL1 requires a stem-loop containing a MBNL1 binding site

To test the role of the stem-loop *in vivo*, a cTNT minigene that includes exons 4-6 (*28*) was used to make mutations in the MBNL1 binding site and monitored for changes in the splicing of the cTNT minigene in HeLa cells containing MBNL1 (*51*). To determine if splicing of the mutated MBNL1 binding sites changed upon MBNL1 sequestration, a plasmid expressing 950 CUG repeats was co-transfected into the HeLa cells. We reproduced the results of Ho and colleagues showing that MBNL1 regulates the splicing of the wild-type cTNT minigene (Figure 8B and 8C), finding that exon 5 inclusion shifts from 66% to 89% when MBNL1 is sequestered by the CUG repeats. When the four guanosines are mutated to cytosines (4G mutations) MBNL1 no longer regulates the splicing of this cTNT minigene.

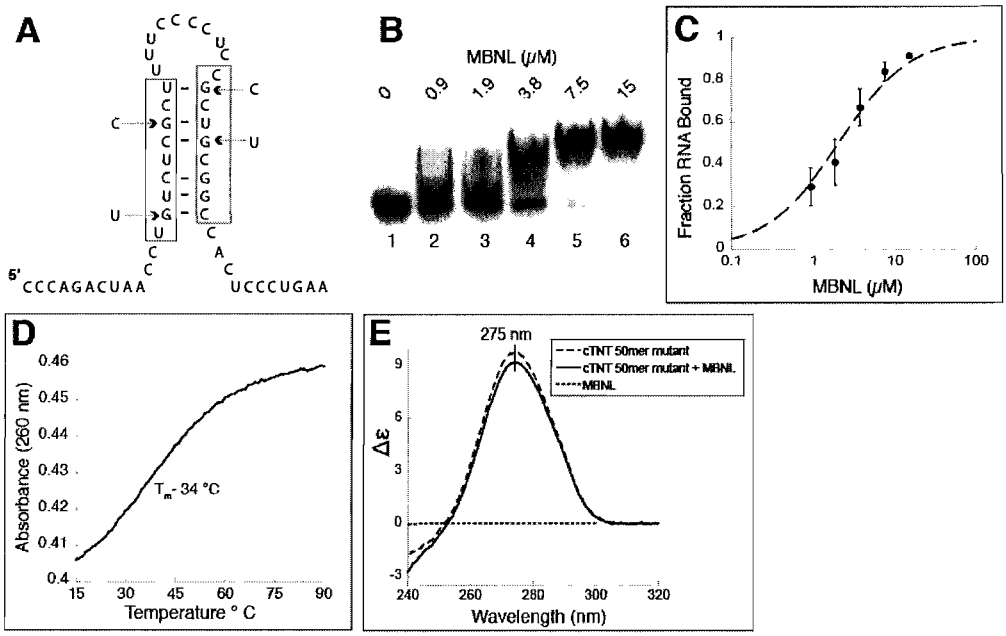

Figure 7. ***Four point mutations destabilize the stem in the cTNT 50mer and reduce MBNL1 binding.*** **(A)** Schematic of the stem-loop with the four point mutations indicated by arrows. **(B)** Gel shift assay shows MBNL1 binding to the mutated 50mer. **(C)** Binding curve of MBNL1 to the mutant cTNT 50mer. **(D)** UV melt of the cTNT 50mer mutant, with a T_m of 34 °C. **(E)** Circular dichroism spectra of the cTNT 50mer mutant alone (1 μM) and with MBNL1 (1 μM).

The 4G mutant had slightly reduced exon 5 inclusion, 57% compared to 66% for the wild-type sequence (Figure 8B). This reduction in exon 5 inclusion is an unexpected result, as the point mutations both disrupt the stem-loop and MBNL1 binding, which should cause an increase in exon 5 inclusion. However, these point mutations create a large new single stranded element in the pre-mRNA, and this structural change might have unanticipated effects, such as the recruitment of other protein factors, or the creation of an unknown silencing element. Upon co-transfection with the DMPK-CUG_{950} minigene, no change was seen in exon 5 inclusion with this construct (Figure 8C-D), indicating that once MBNL1's binding site is abolished, the splicing of this construct is independent of MBNL1. Ho and colleagues also observed this surprising result with the 4G mutant minigene (*28*).

To determine if a stem-loop that contains CUG repeats similar to the CUG_4 construct would function to recruit MBNL1, the upper portion of the stem was replaced with two CUG repeats on each side (CUG stem - CUGS, Figure 8A). This CUGS minigene was regulated by MBNL1. In the presence of MBNL1 only 29% inclusion of exon 5 was observed (Figure 8B)

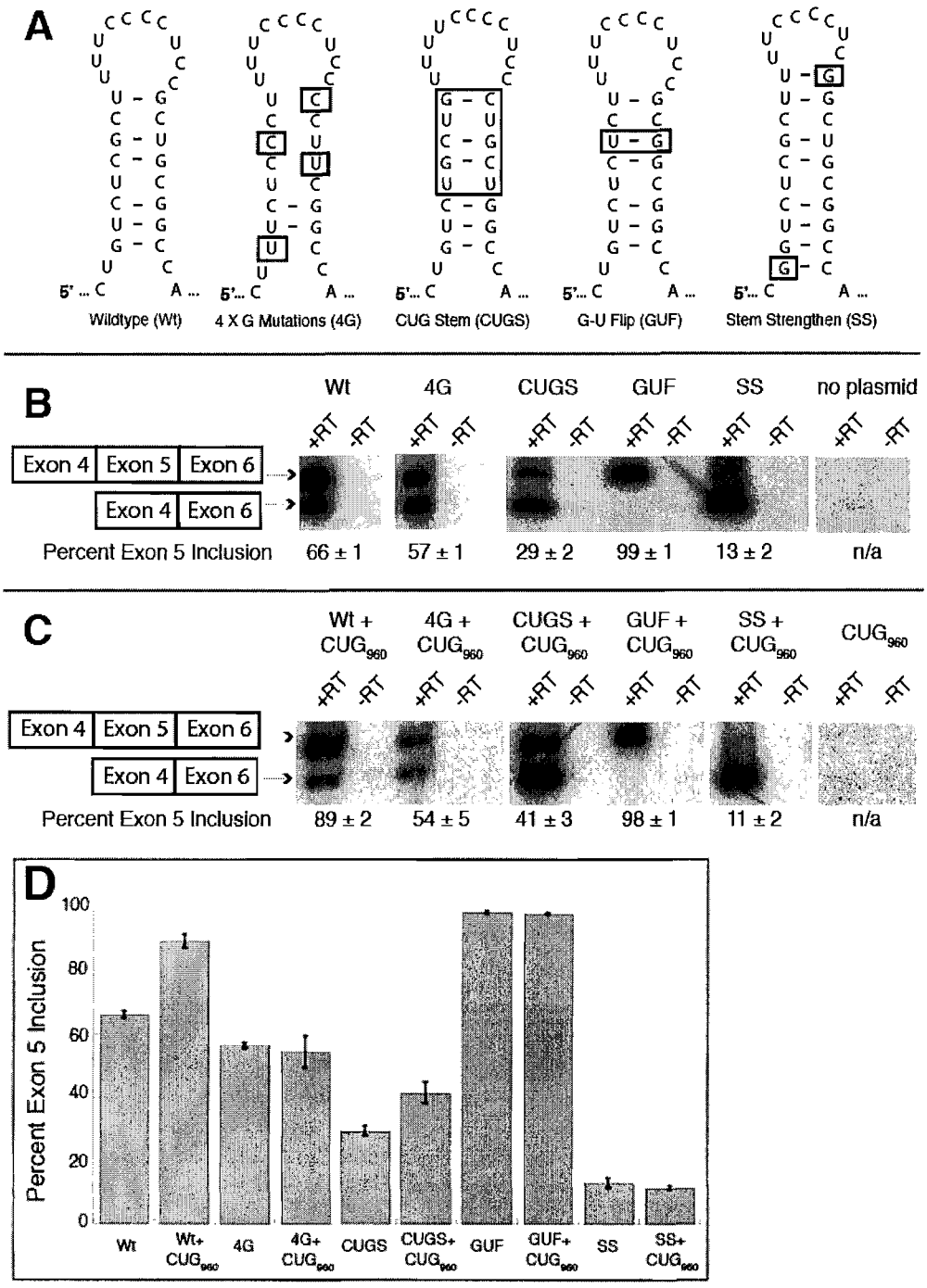

Figure 8. ***The sequence and structure of the cTNT stem-loop upstream of exon 5 is important for regulated splicing by MBNL1.*** **(A)** Schematic of mutations made to the stem-loop in intron 4 of the cTNT minigene. **(B)** RT-PCR results of the wild-type and mutant cTNT minigenes. **(C)** Co-transfection with a second minigene that contains 950 CUG repeats to test the role of MBNL1 sequestration on the splicing of the cTNT minigenes. **(D)** Graphical representation of exon 5 inclusion for the different constructs with and without the overexpression of the 950 CUG repeats.

while sequestration of MBNL1 by the CUG_{950} RNA leads to 41% of the exon 5 inclusion. The lower percentage of exon 5 inclusion compared to wild-type and the weaker effect of MBNL1 regulated is likely due to the stability of the stem-loop being different between these two RNAs.

Further strengthening the stem-loop with two additional base pairs, one at each end of the helix (Figure 8A, Stem Strengthen – SS), almost completely inhibited the use of exon 5 both in the presence of MBNL1 and when it was sequestered (Figure 8B-D). This result indicates that a strong stem-loop inhibits the use of the 3' splice site and the binding of MBNL1 probably does not enhance the effect because the stem is sufficient on its own, while the weaker stems (wild-type and CUGS) can be melted by the splicing factors binding at the 3' end of the intron.

The *in vitro* binding experiments demonstrated that MBNL1 prefers to have purines on both sides of the helix and therefore we predicted that altering the cTNT stem-loop in a manner that shifted all the purines to one side and the pyrimidines on the other side would result in an MBNL1 unregulated exon. This was done by flipping a G-U base pair (G-U flip - GUF, Figure 8A). This change results in complete inclusion of exon 5 (99%) in the presence or sequestration of MBNL1 (Figure 8B-D). As expected this RNA is bound weakly by MBNL1 (Table 1, a 42-fold decrease in binding compared to cTNT 32mer). The complete inclusion of exon 5 is likely the result of strengthening the poly-pyrimidine tract, which is recognized by the constitutive splicing factor U2AF65.

Discussion

MBNL1's structure and its lack of cooperative RNA binding

The AUC and CD experiments with MBNL1 suggest the protein adopts a slightly elongated structure. When monitored by CD, the α-helical structure of MBNL1 does not appear to significantly alter upon the addition of CUG_4, although other structural components have yet to be measured upon RNA binding. We propose MBNL1, like other zinc finger proteins, is organized into domains around each zinc ion, and that each domain contacts the RNA in a relatively independent manner (*70-73*).

The lack of significantly higher affinity binding to longer CUG repeats suggests that MBNL1 does not bind in a highly cooperative manner to CUG repeats. An analysis of our binding data supports this conclusion. Scatchard and Hill plots (data not shown) reveal MBNL1 binds CUG_{90} with only minimal cooperativity (Hill constant of 1.4) under the binding conditions used in these studies. Analysis of the other repeat RNA substrates indicates only very weak

cooperativity, and for cTNT 50mer there is no evidence for cooperative binding at all. These results suggest this version of MBNL1 recognizes the many binding sites on long CUG repeats as independent binding sites.

It should be noted that the apparent K_ds reported here are dependent on our binding conditions: 175 mM NaCl, 5 mM $MgCl_2$, 20 mM Tris (pH 7.5), 1.25 mM BME, 12.5% glycerol, 2 mg/ml BSA and 0.1 mg/ml Heparin. These conditions are quite stringent and the apparent K_ds are lowered if NaCl, Heparin concentrations or temperature are decreased in the binding conditions. If we perform binding studies under these less stringent conditions our apparent K_ds for CUG repeats are similar to those measured by Yuan and colleagues (*74*).

The RNA binding specificity of MBNL1 for CUG and CCUG repeats

Analysis of RNA substrates in which the CUG repeat tract sequences are modified in various ways (Figure 2) clearly shows the preference of MBNL1 for pyrimidine-pyrimidine mismatches as well as the Watson-Crick base pairs in their particular positions. The requirement of mismatches for MBNL1 binding could be due to two different reasons; either the mismatches are directly recognized by MBNL1 or the mismatches allow the helix to be easily distorted so MBNL1 can gain access to the C-G and G-G base pairs.

Both the specific recognition and structural distortion are likely playing a role in MBNL1 binding the CUG repeats. The distortion is supported by the decrease in CD signal and the peak shift upon MBNL1 binding to the CUG repeats and the cTNT 50mer (Figure 4). Furthermore, there also appears to be a general inverse trend between MBNL1 binding affinity and the stability of the RNA structure. All the RNA substrates to which MBNL1 binds with strongest affinity generally have lower T_m values (Table 1). The CCUG substrates, as a group, have the strongest affinity for MBNL1 and also have the lowest T_m values. The pyrimidine-pyrimidine mismatches also, as a group (Table 1), have the lowest T_m values for RNAs containing a mismatch, as well as the strongest binding affinity for MBNL1. This trend further suggests that MBNL1 alters the structures of the RNAs when it binds. If more energy is required to distort or disrupt the base pairing interactions of the RNA, it might result in weaker binding by MBNL1.

MBNL1 binds CCUG repeats with approximately 2-fold stronger affinity in both structural conformations (Figure 3) compared to the CUG repeats. Previously, Kino and colleagues also qualitatively found that MBNL1 preferred CCUG repeats over CUG repeats. These results are surprising because patients with DM2 tend to have more CCUG repeats

compared to patients with CUG repeats (*75*). Additionally, the *ZNF9* pre-mRNA containing the CCUG repeats appears to be expressed at similar or even higher levels than the *DMPK* pre-mRNA (*50, 76*). Yet, the symptoms of DM2 patients are less serious than those of DM1 patients (*77*). This supports the model that the expanded CUG repeats, unlike the CCUG repeats, are affecting transcription or other processes in the cell and create another layer of mis-regulation in DM1 compared to DM2 (*21*).

Recognition of a helical element within the cTNT pre-mRNA by MBNL1

Our observation that MBNL1 binds short stem-loops prompted us to consider that MBNL1 might recognize a short stem-loop within the cTNT pre-mRNAs as well. We chose to study the cTNT intron 4 because it was the only available pre-mRNA substrate with a well identified binding site (*28*). The combination of UV melting of the cTNT 50mer RNA (Figure 4E), CD of this RNA in the absence and presence of MBNL1 (Figure 4F) and structure probing of the cTNT 32mer (Figure 5A-B) show that this RNA folds into a stem-loop structure which appears to be partially A-form, which was also found for the CUG repeats. This proposed helix has some similarities to the CUG and CCUG helices in that one of the pyrimidine-pyrimidine mismatches is bracketed by G-C base pairs (though the other mismatch is flanked by G-U wobble base pairs). The similarities between the CUG, CCUG and cTNT stems suggest MBNL1 recognizes both its pathogenic and natural RNA targets through an analogous mode of recognition.

The stem-loop structure in the 4^{th} intron of the cTNT pre-mRNA is not predicted by m-fold, perhaps due to its multiple mismatches and the minimal number of consecutive base pairs. It was therefore previously predicted to be single stranded. As stated above, our data strongly support the formation of this stem-loop and that MBNL1 recognizes this RNA structure. The lack of binding to the cTNT 50mer containing the four mutations due to the disruption of the stem-loop and potential removal of key nucleotides for MBNL1 recognition (Figure 7), further support the model that MBNL1 binds this stem-loop. The lack of cross-linking by MBNL1 to a cTNT RNA containing these four mutations as well as the elimination of regulated splicing of the cTNT exon 5 when these mutations are introduced into a cTNT minigene (*28*); suggest the presence of this stem-loop within intron 4 of the cTNT pre-mRNA is important for the regulated splicing of exon 5 by MBNL1.

Clearly MBNL1 can bind a range of RNA stems, but the common theme of pyrimidine-pyrimidine mismatches and presence of G-C and C-G base pairs indicates these are requirements for binding by MBNL1. Further studies are necessary to fully define the specificity of MBNL1 before predictions can be made to identify binding sites in other MBNL1 regulated pre-mRNAs. Another challenge is that these potential regulatory stem-loops will not necessarily be predicted by folding programs due to the lack of consecutive base pairs and presence of mismatches as observed for the cTNT intron 4 stem-loop.

MBNL1's binding to a helical element in its role as a splicing regulator

A possible model for the mechanism through which MBNL1 regulates the exclusion of exon 5 in the cTNT mRNA is that MBNL1 competes for binding of the intron with other splicing factors. When MBNL1 binds the stem-loop at the 3' end of the intron, the presence of MBNL1 or the stabilization of the stem-loop may inhibit recognition of this 3' splice site by the splicing machinery causing exon skipping to occur. If the stem-loop does not form and MBNL1 is not present, the splicing machinery recognizes this site and exon 5 is included (Figure 9).

The mutations made to the cTNT splicing minigene support the model that recruitment of MBNL1 to a stem-loop causes repression of exon 5. The replacement of the endogenous binding site with another sequence that MBNL1 binds causes MBNL1 dependent repression of exon 5, while minor mutations that significantly reduce MBNL1's binding abolishes MBNL1 ability to repress inclusion of exon 5. These results indicate that direct binding of MBNL1 to this stem-loop is required for the repression of exon 5. Mutations that add additional base pairs that stabilize the stem-loop also strongly cause repression of exon 5, showing that increased stability of the stem correlates with repression of exon 5. It appears that if the stem is strengthened enough, MBNL1 is no longer needed for repression, as sequestering MBNL1 function does not affect splicing of this construct.

This mechanism may have two levels of regulation. First, the stem-loop alone might inhibit the use of this 3' splice site; second, the addition of MBNL1 might further stabilize this stem-loop and increase the block at this 3' splice site. Previously, stem-loop formation has been shown to reduce the use of 5' and 3' splice sites (*78, 79*), but this is the first example to our knowledge in which the stem-loop is specifically recognized by a factor that regulates splicing. Therefore MBNL1 may be a member of a new second class of splicing regulators utilizing

secondary structure, while other splicing regulators have previously been shown to recognize single-stranded motifs, such as the SR splicing factors (for review see (*80, 81*)).

Like other factors regulating splicing, MBNL1 functions in one context to exclude an exon as described for the cTNT exon 5, while for other exons the presence of MBNL1 enhances the inclusion of a particular exon. One possible mechanism through which MBNL1 could enhance exon inclusion would be to sequester exonic or intronic splicing silencers in a stem-loop. Alternatively the location of MBNL1 binding (upstream or downstream of the exon or within the exon) may determine if it acts as splicing enhancer or repressor as has been found for other splicing factors, such as NOVA (*82*).

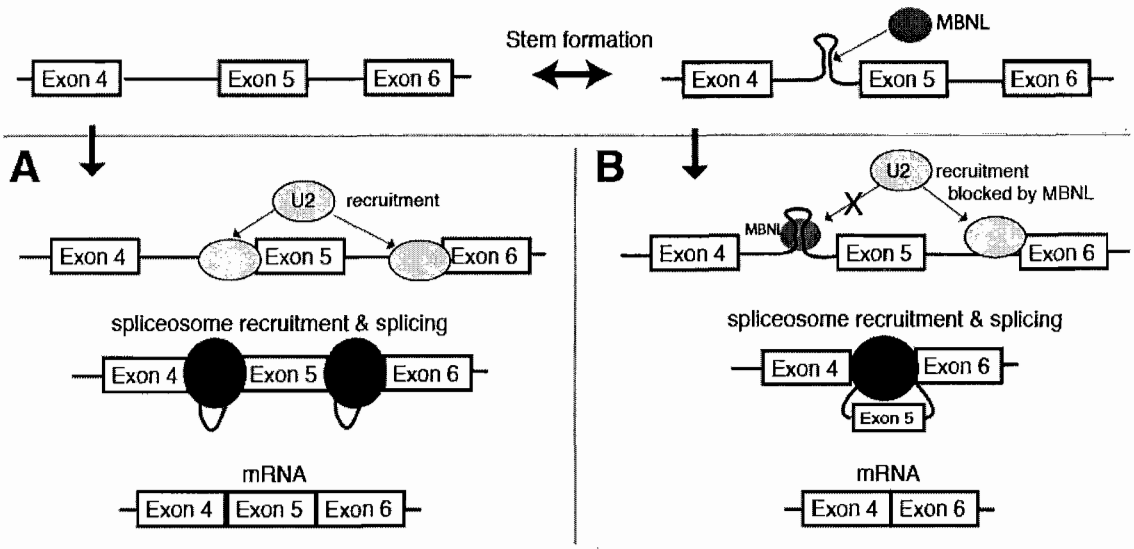

Figure 9. ***Model of MBNL1's regulation of the cTNT pre-mRNA splicing.*** (**A**) If the stem-loop doesn't form the splicing machinery represented by U2 (U2 snRNP) recognizes this 3' splice site and exon 5 is included. (**B**) When the stem-loop forms and MBNL1 binds, this 3' splice site is not recognized by the splicing machinery and exon 5 is skipped.

Materials and methods

Cloning and protein purification. MBNL11 was PCR amplified and was cloned into GST fusion vector pGEX-6P-1 (Amersham), using DNA (MBNL1 isoform with amino acids 1-382) provided by Maury Swanson (*28*). Both the full length MBNL11 and MBNL11(1-260) constructs were cloned using BamHI and NotI restriction sites.

Using BL21-Star expression cells (Invitrogen), protein expression was induced with 0.25 mM IPTG at an OD_{600} ~ 0.5-1, for 3-4 hours at 37°C. Cells were lysed in 30 mL of buffer (500 mM NaCl, 25 mM Tris pH 7.5, 10 mM β-Mercaptoethanol (BME) and 5% glycerol) using 1

mg/ml of lysozyme followed by sonication (3 x 30 seconds). Cell extract was centrifuged for 15 minutes at 17,000 rpm, and lysate which contained GST-MBNL1 was collected. GST-MBNL1 was bound to GST affinity beads for 30 minutes at 4°C. Beads were washed 5 times with buffer (1M NaCl, 25 mM Tris pH 7.5 and 5mM BME); MBNL1 was cleaved from the affinity tag with Precision Protease (Amersham) and collected from the beads. The protein was then run over an anion exchange (Q) column. Co-purifying contaminants bind the column, but MBNL1 does not. MBNL1 was collected in the column flow through, concentrated and dialyzed into storage buffer (50% glycerol, 500 mM NaCl, 20 mM Tris pH 7.5, 5 mM BME) and stored at -20 °C.

RNA synthesis, labeling and purification. The RNA substrates CUG_{90} and CUG_{54} were transcribed with T7 polymerase off the pCTG54 and pCTG90 plasmids, respectively, provided by Maury Swanson (*51*). During transcription, RNAs were radiolabeled using [α-^{32}P]CTP. All other RNA substrates were ordered from IDT DNA, and 5' end-labeled using [γ-^{32}P]ATP. All RNAs were purified on 8% poly-acrylamide denaturing gels.

Gel shift assay. Solutions for the protein-RNA binding experiments contained 175 mM NaCl, 5 mM $MgCl_2$, 20 mM Tris (pH 7.5), 1.25 mM BME, 12.5% glycerol, 2 mg/ml BSA and 0.1 mg/ml Heparin. Prior to incubation, RNA substrates were annealed by incubation at 95 °C for 2 minutes and then placed directly on ice for 20 minutes in 66 mM NaCl, 6.7 mM $MgCl_2$, and 27 mM Tris (pH 7.5). Protein was then added to the RNA. The binding reaction was 10 µL volume and was incubated for 20 minutes at room temperature before 3-5µL were loaded on a pre-chilled 4 °C gel. RNA and RNA-protein complexes were separated on 3% acrylamide (37.5:1, or 80:1), 0.3% agarose, 0.5x Tris-Borate (TB) gels, run for approximately 30 minutes at 4 °C, 175 volts. Gels were dried and autoradiographed.

For binding curves, gels were quantified using ImageQuant (Molecular Dynamics). The percent RNA bound was determined by taking the ratio of RNA:Protein complex to total RNA, per lane. Binding curves were graphed and apparent K_d values were determined with KaleidaGraph (Synergy) software using the following equation: $y=((m2 + m1 + m0)-((-m2 - m1 - m0)^2-(4 \times m1 \times m0))^{0.5}/(2 \times m1)$, where y=% bound, m2= K_d, m1=total RNA concentration and m0=protein concentration. This equation assumes a 1:1 interaction between the RNA and protein, which allows only an apparent K_d to be determined for CUG_{90} and RNAs containing more than one binding site. To determine the standard error of the apparent dissociation constants, 3-5

binding titrations were performed with each substrate and the apparent K_d values determined for each titration separately, prior to averaging. The error bars on the binding curve were obtained by averaging the individual titration points and calculating the standard deviation. Data points greater than two standard deviations from the average were discarded, with at least three data points remaining for standard deviation analysis.

Structure probing assay. End-labeled RNA was incubated in the presence of protein or buffer for 20 minutes at room temperature, using the same binding conditions used for the gel shift assay (except BSA was excluded and 0.1 mg/ml tRNA was used instead of heparin, and the volume of the binding reaction was 9 μL). RNases (1 μL) were added, and the RNA digested for 2 minutes at room temperature. The reaction was quenched with phenol; RNA was collected through ethanol precipitation, re-suspended in denaturing dye and heated at 95 °C for 10 minutes. The sample was then run on a 15% poly-acrylamide (19:1), 8M Urea, 1x TBE gel.

Analytical ultracentrifugation (AUC). AUC runs with MBNL1 (20 μM) were made in solutions containing 170 mM NaCl, 20 mM Tris (pH 7.5), BME 0.2 mM and 1% glycerol (v/v). Data was collected on an Optima XL-I (Beckman), for ~10 hours at 60,000 rpm, 4 °C. Data was analyzed using SedFit 97 (NIH).

Circular dichroism (CD) and thermal melts. The CD spectra of MBNL1 (2 μM) were measured in 50 mM NaCl and 10 mM sodium phosphate buffer (pH 7.5). Data was collected on a J-720 spectropolarimeter (Jasco) for 0.5-1 hour at room temperature at a rate of 5 nm/min. For the CD spectra of RNA, the samples were annealed by snap-cooling (95 °C for 2 minutes, then directly on ice for 15 minutes) in 50 mM NaCl, 5 mM $MgCl_2$ and 10 mM phosphate (pH 7.5); lower salt and phosphate were used to reduce interference at lower wavelengths. CD measurements were made at room temperature over a period of 2-3 hours, at spectral scanning rates of 5 or 10 nm/min. For measurements on protein-RNA complexes the components were mixed and incubated for 20 minutes prior to data collection.

For the CD thermal melt, RNA was prepared as normal and heated at a rate of 2°C per minute. For UV thermal melts, RNAs (1 μM) were snap-annealed and melted at a rate of 2 °C per minute, monitored at 260 nm. T_m values were calculated by determining the inflection point for each thermal transition, or the mid-point of the transition if the inflection point was not apparent.

In vivo *splicing*. Wild-type cTNT, the 4G cTNT mutant, and the DMPK-CUG_{950} minigenes were obtained from the lab of Thomas Cooper (*28*). All additional mutations were made using PCR from the WT sequence. Sal1 and Spe1 restriction sites were used to sub-clone all mutants after PCR.

HeLa cells were grown in monolayers in DMEM with GLUTAMAX (GIBCO) and supplemented with 10% fetal bovine serum (GIBCO). Approximately 2.0 (± 0.2) $x10^5$ cells were plated in 6 well plates and transfected 18-20 hours later at 70-90% confluency. 1 ug of plasmid was transfected into each well of cells, using 5 uL of Lipofectin2000 (Invtitrogen) according to the manufacturer's protocols. For co-transfection, 1 ug of total plasmid was transfected, at 500 ng of each construct. Cells were harvested 20-24 hours after transfection using triplE reagent (GIBCO). Immediately following harvesting, RNA was isolated from the cell pellets using an RNeasy kit (QIAGEN). 500 ng of isolated RNA was DNased with RQI DNase (Promega) according to manufacture's protocol. 100 ng of DNased RNA was reverse transcribed with Superscript II and a cTNT-specific reverse primer according to manufacturer's protocols. 30 ng of the RT reaction was subjected to 22-25 rounds of PCR amplification using cTNT specific primers spiked with a kinased cTNT forward primer. The linear range for PCR was determined for the WT and 4G construct and found to be between 20-26 cycles. The resulting PCR products were run on a 6% (19:1) poly-acrylamide denaturing gel at 6W for 2 hours. The gel was subsequently autoradiographed and quantitation of the radioactive bands was performed using ImageQuant software. T-Tests were performed in Excel, assuming two distribution tails using a homoscedatic model.

CHAPTER III

THE PROTEIN FACTORS MBNL1 AND U2AF65 BIND ALTERNATIVE RNA STRUCTURES TO REGULATE SPLICING

Contribution note: this chapter is previously published (Warf et al., 2009) (132). The other contributors were Julien V. Diegel, Peter H. von Hippel and J. Andrew Berglund. Julien V. Diegel performed a few experiments, and Peter H. von Hippel and J. Andrew Berglund helped with experimental design, data analysis and manuscript preparation.

Introduction

Alternative splicing is a fundamentally important process that many organisms use to increase proteomic diversity. Many diseases are known to arise from the mis-regulation of alternative splicing (*18*). Myotonic Dystrophy (DM) is an example of a disease where the alternative splicing of many pre-mRNAs is mis-regulated. The mis-splicing seen in DM is thought to be caused, at least in part, by the mis-localization of the splicing factor MBNL1 (*21, 22, 84*). MBNL1 is known to interact directly with three pre-mRNAs that are mis-regulated in DM, the cardiac troponin T (cTNT) (*28, 85*), fast troponin T (*74*), and SERCA1 (*86*). It is unclear how many more of the mis-regulated pre-mRNAs observed in DM may also be directly bound and regulated by MBNL1.

We have recently shown that MBNL1 binds a stem-loop within intron 4 of the cardiac troponin T (cTNT) pre-mRNA. This stem-loop is located directly upstream of exon 5, which is mis-regulated in this disease (*28, 85*). The mechanism used by MBNL1 to repress exon 5 remains undetermined. One mechanism might entail a direct binding competition between MBNL1 and other splicing factors. One of the best articulated models of the regulation of alternative splicing is sex determination in *Drosophila melanogaster*, where the protein Sex Lethal (Sxl) competes with the splicing factor U2AF65 at the 3' end of certain introns (reviewed in (*87*)). U2AF65 is thought to be one of the first splicing factors to bind an intron and thus helps define the 3' end of the intron. In concert with other proteins that bind near the 3' end, U2AF65 is responsible for helping to recruit the U2 snRNP to the branch-point sequence (*7, 8, 88*). When Sxl inhibits

U2AF65 binding and the subsequent recruitment of the U2 snRNP, the spliceosome selects another 3' splice site, or the intron is retained if another 3' splice site cannot be defined (*88, 89*).

We hypothesize that MBNL1 may act through a similar mechanism to compete with U2AF65. U2AF65 is a potential competitive target of MBNL1 because a putative U2AF65 binding site appears to be in the loop portion of the stem-loop which MBNL1 binds (*84*). We found that MBNL1 does compete with U2AF65 for binding of a region within intron 4. This competition with U2AF65 is functionally important, as recruitment of the U2 snRNP is reduced by MBNL1. Furthermore we found that MBNL1 and U2AF65 compete by binding mutually exclusive RNA structures.

Results

U2AF65 recognizes a canonical sequence within intron 4 of the cTNT pre-mRNA

We have previously shown that MBNL1 binds a stem-loop located at the 3' end of intron 4 (Figure 1A-B) (*84*). A putative poly-pyrimidine tract (py-tract) for intron 4, the likely U2AF65 binding site, exists primarily within the loop portion of this stem-loop (Figure 1A-B). This loop contains an uninterrupted sequence of five uracil and seven cytosine residues and is 25 residues upstream of the 3' splice site and nine residues downstream of a consensus branch-point.

The affinities of U2AF65 for regions of this intron were determined using an electrophoretic mobility shift (gel shift) assay. U2AF65 bound to a 50 nucleotide sequence (cTNT 50mer) containing the putative py-tract with an affinity (K_d) of 310 ± 30 nM (Figure 1C, see S. Figure 1 for representative gel shifts). Truncated versions of this RNA that retained the putative py-tract, but removed another run of pyrimidines nearer to the 3' splice site, showed an unchanged binding affinity for U2AF65 (cTNT 32mer, Figure 1B-C). Mutations that replaced the putative py-tract nearly abolished U2AF65 binding (UUCG Loop, Figure 1B-C). Based on its location, the sequence of the region, and the binding data and the truncations, it is likely that this region is indeed the py-tract and comprises the binding site of U2AF65 within this intron.

MBNL1 and U2AF65 bind competitively to the 3' end of intron 4

To determine whether MBNL1 competes with U2AF65 for binding at this intronic site, a UV cross-linking assay was used. While holding U2AF65 concentration constant, increasing the concentration of MBNL1 strongly reduced U2AF65 cross-linking (Figure 2). If the two proteins could bind to the RNA at the same time, it was expected that a ternary complex of higher

molecular weight would be observed. However, since U2AF65 cross-linking was significantly decreased and little or no apparent ternary complex was observed, it appears that MBNL1 and U2AF65 bind competitively to the cTNT 50mer-Wt sequence.

A mutated RNA (cTNT 50mer-4G, Figure 2) was used to determine if non-specific binding of MBNL1 affected U2AF65 binding. In this RNA, four guanosine residues in the stem were mutated that are required for MBNL1 binding (*84*) and cross-linking (*28*), while U2AF65 binding and cross-linking remain intact. As expected, cross-linking of this RNA to U2AF65 was not affected by MBNL1, showing that MBNL1 specifically competes with U2AF65.

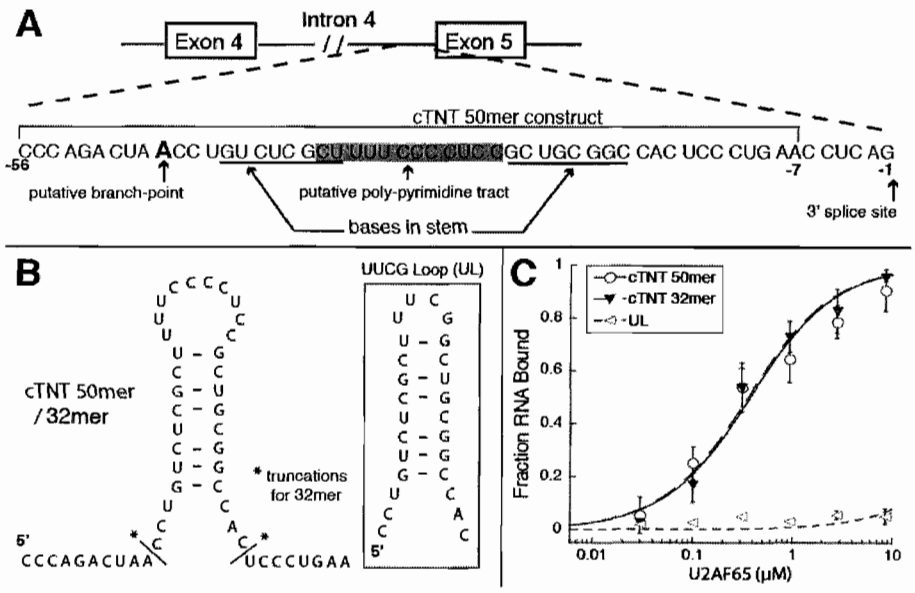

Figure 1. ***U2AF65 binds the putative poly-pyrimidine tract in the cTNT intron 4.*** **(A)** Schematic of the 3' end of intron 4 of the cTNT pre-mRNA. The cTNT 50mer used for binding studies are indicated, as well as the bases in the stem-loop structure. The 3' splice site, putative poly-pyrimidine tract and putative branch-point are also indicated. **(B)** Structures of tested RNAs. **(C)** Binding curves of tested RNAs.

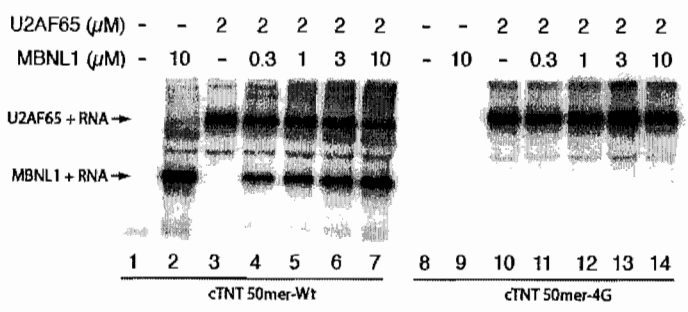

Figure 2. ***MBNL1 and U2AF65 competitively bind intron 4 of the cTNT pre-mRNA.*** MBNL1 and U2AF65 were UV cross-linked to the cTNT 50mer-Wt RNA and run on a denaturing gel. MBNL1 competes with U2AF65 for the cTNT-Wt sequence, but not for the cTNT-4G mutant, which MBNL1 does not bind.

MBNL1 inhibits "A complex" formation on intron 4

The role of U2AF65 is to define the 3' end of an intron and help recruit the U2 snRNP to the branch-point (*7, 8, 87*), forming "A complex" (*3, 4*). If U2AF65 does not bind, recruitment of the U2 snRNP to the branch-point is compromised (*7, 8*). We predicted that competition between MBNL1 and U2AF65 would inhibit formation of the A complex on this intron.

An RNA construct was made that contained human cTNT exon 5, together with 100 nucleotides of the 3' end of the RNA upstream of intron 4. Sequences from a constitutively spliced intron from chicken troponin I (TNI) were placed upstream of that sequence (see Figure 3A for diagram). Radiolabeled RNA was incubated with HeLa nuclear extract, allowing splicing complexes to form on the RNA substrate. A time course shows that this cTNT pre-mRNA forms A complex well (Figure 3B). The ADML pre-mRNA substrate served as a positive control because it forms A complex robustly (Figure 3B). In the absence of ATP, no complex band is formed on the cTNT-Wt pre-mRNA, indicating that this is A complex (Figure 3C). At later time points, larger complexes (B and C) are seen on the ADML pre-mRNA, but the cTNT pre-mRNAs are not robust enough to form these complexes *in vitro* prior to RNA degradation.

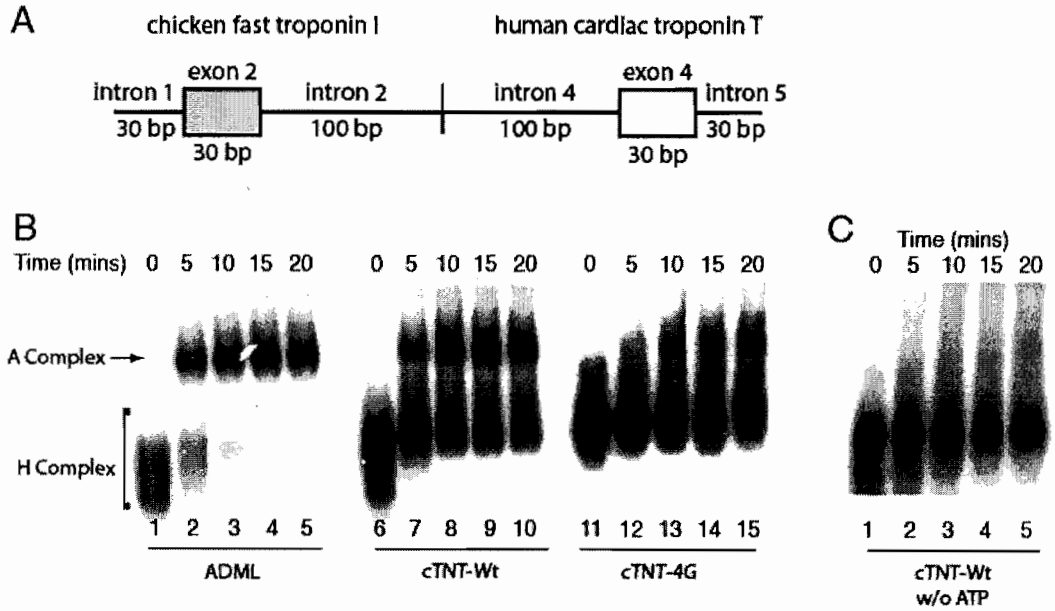

Figure 3. ***Time course and ATP dependence of complex formation assay.*** **(A)** Schematic of cTNT pre-mRNA used in complex formation assay. **(B)** Time course of A complex formation on cTNT-Wt, ADML and cTNT-4G pre-mRNAs. **(C)** Time course of A complex formation on cTNT-Wt in the absence of ATP.

The 10 minute time point was chosen to assay the affect of MBNL1 on A complex formation because RNA degradation was minimal and complex formation was optimal on the cTNT pre-mRNA (Figure 3B). Complex formation on the cTNT-Wt pre-mRNA was reduced by nearly 2 fold with increasing concentrations of MBNL1, while little effect was seen on ADML pre-mRNA or on the mutant cTNT-4G pre-mRNA, (Figure 4A-B). This indicates that MBNL1 reduces the formation of A complex only on its specific pre-mRNA target, and does not perturb A complex formation in general.

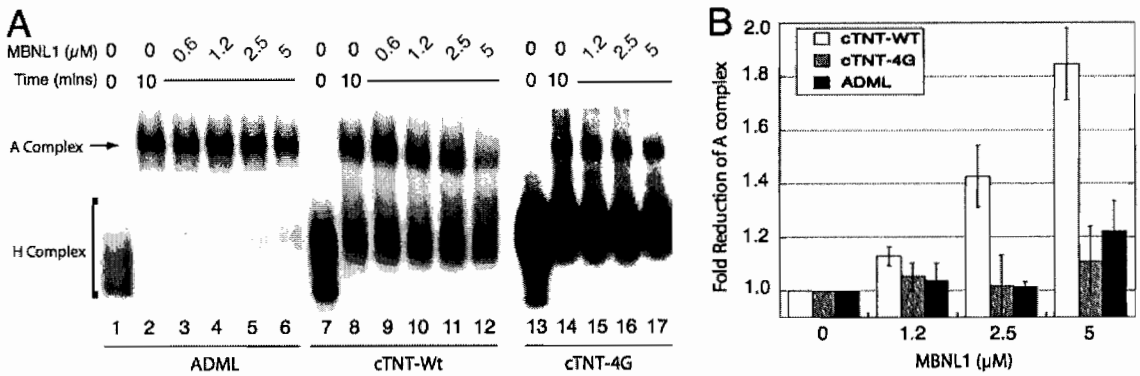

Figure 4. ***MBNL1 reduces "A complex" formation specifically on intron 4 of the cTNT pre-mRNA.*** **(A)** MBNL1 reduces A complex formation strongly on the cTNT-Wt pre-mRNA, but not on an ADML or cTNT-4G pre-mRNA. **(B)** Quantitation showing the affect of MBNL1 on A complex formation on cTNT-Wt, cTNT-4G and ADML pre-mRNAs.

Stabilization of the stem-loop reduces U2AF65 binding affinity and represses exon 5 inclusion in vivo

We previously observed that mutations that increased the number of base-pairs in the stem-loop repressed exon 5 inclusion, independently of MBNL1 (*84*). This led us to hypothesize that these mutations might repress the inclusion of exon 5 by reducing the affinity of U2AF65 for the py-tract, in a manner similar to the repression mechanism used by MBNL1. To investigate this possibility we tested the affinity of U2AF65 for the cTNT intron 4 py-tract using stem-loops with increased stability.

A correlation was seen between the stability of the stem-loop and the binding affinity of U2AF65, with stronger stems nearly abolishing U2AF65 binding (Figure 5A-B). The Stem Strengthen (SS) mutation stabilized the stem by adding one base pair to each end of the stem, increasing the T_m of the stem-loop by 5 °C, reduced U2AF65 binding ~10-fold (Figure 5A-B).

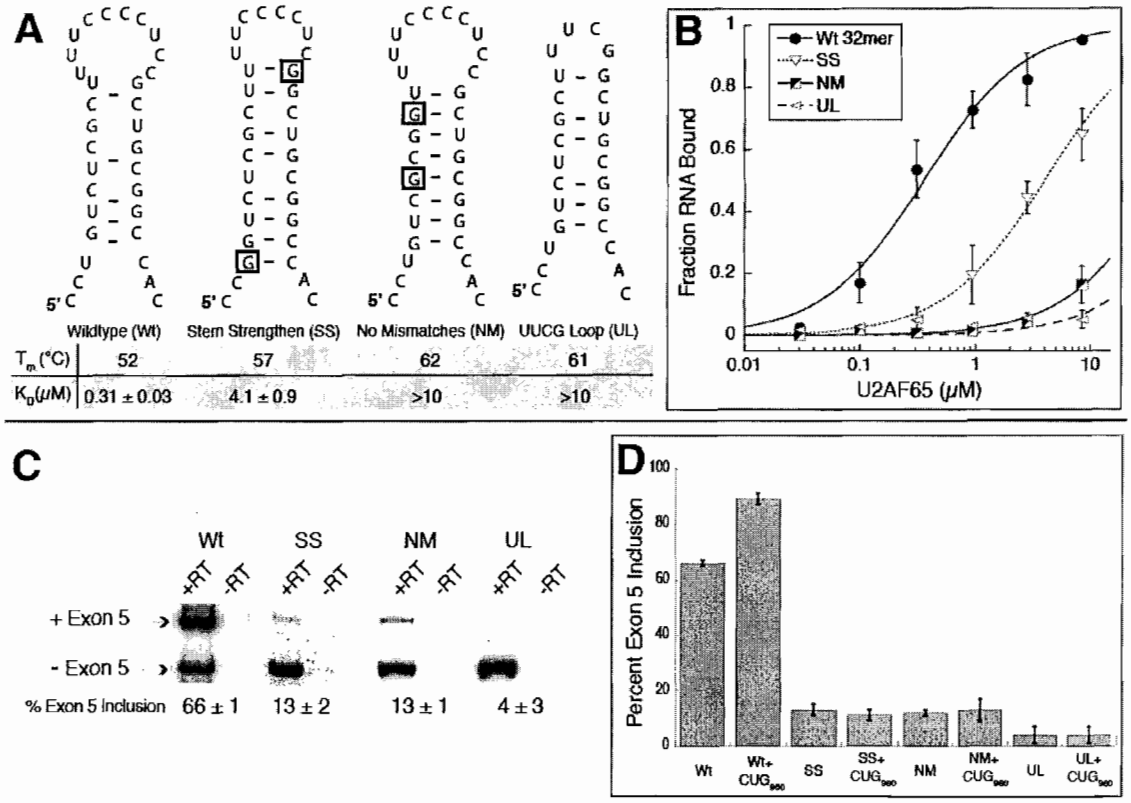

Figure 5. ***Mutations that stabilize the stem-loop structure reduce U2AF65 binding affinity and repress exon 5 inclusion*** **in vivo. (A)** Schematic of different mutations, the thermal stability of each mutation and the binding affinity of U2AF65 for each RNA. **(B)** Binding curve of tested mutations. **(C)** Representative RT-PCR data from *in vivo* splicing of cTNT mutant series. **(D)** Quantitation of RT-PCR data, as well as data from CUG_{960} co-transfection experiments.

The No Mismatch (NM) mutation increased the stability of the stem by 10 °C and reduced U2AF65 binding to the point where a K_d value could not be determined (Figure 5A-B). It is interesting that the mutation that more strongly stabilized the stem structure had a larger effect on binding. This result suggests that U2AF65 binds the py-tract in a single-stranded region of this RNA. As a negative control, a mutation (UUCG loop, UL) was made that removed the py-tract from the loop to fully abolish U2AF65 binding (Figure 5A-B).

Furthermore, we found a strong correlation between U2AF65 binding affinity and exon 5 inclusion *in vivo* (Figure 5C-D). The NM, SS, UL mutations all showed minimal exon 5 inclusion using an *in vivo* splicing assay. This repression was independent of MBNL1, as co-expression of expanded CUG repeats (which would sequester MBNL1) had no affect on exon 5 inclusion for

these mutations (Figure 5D). These results indicate that inclusion of exon 5 can be regulated either by competition of MBNL1 with U2AF65 at the 3' end of the intron, or by the formation of RNA secondary structures that reduce the affinity of U2AF65 for its binding site.

U2AF65 and MBNL1 bind distinct RNA structures

To determine the effect of MBNL1 or U2AF65 binding on the stem-loop structure, a fluorescence-based assay was used. The fluorescent adenosine analogue 2-amino purine (2-AP) was used to replace a guanosine residue that forms a "wobble base pair" with a uracil residue near the base of the stem (Figure 6A). 2-AP was chosen as a spectroscopic probe for this purpose because its fluorescence is strongly quenched when it is in a base-paired structure (here with uracil), while its fluorescence strongly increases when it is in a single-stranded environment. The guanosine residue near the base of the stem was selected for replacement as no adenosine residues were available in the stem and replacement of this position had no affect MBNL1 or U2AF65 binding or stem formation (data not shown).

Titration of U2AF65 to the RNA increased the fluorescence of the probe nearly two-fold (Figure 6B). In general, the fluorescence of a 2-AP probe has been shown to increase 2- to 3-fold when 2-AP switches from a base-paired to a single-stranded environment (*90*). Therefore this result suggests that U2AF65 binds this RNA in a single-stranded structure. U2AF65 was titrated to a point where the RNA should be ~85% bound, based on the K_d determined from the gel shift assay. A U2AF65 protein that contains four mutations which inhibit RNA binding (U2AF65–Quad Mutant or QM (*91*)) had no effect on the fluorescence of the probe (Figure 6B).

In contrast to U2AF65, titration of MBNL1 (to a concentration at which 85% of the RNA should be bound) had no significant effect on the fluorescence of the 2-AP probe (Figure 6C). This result indicates that the stem structure remains intact upon MBNL1 binding. However, a recent crystal structure of two zinc fingers from MBNL1 with a six nucleotide RNA suggests that at least two of MBNL1's zinc fingers bind RNA sequences in a single-stranded structure (*92*). This crystal structure was only obtained with a minimal RNA sequence that was not long enough to form a double-stranded structure, so it is unclear if a longer RNA sequence would be in a different structure. It might be possible that MBNL1 does bind the cTNT stem as single-stranded RNA, but the protein quenches the 2-AP probe so that the fluorescence does not increase. However, this is unlikely, as 2-AP probes have not been seen to be strongly quenched by direct protein binding when compared to base stacking and base pairing in a stem (*90, 93*).

The recent crystal structure argues that it is probable that MBNL1 opens at least one or two bases in the stem. However, this fluorescence data suggests that other portions of the stem remain structured. We therefore conclude that at least a portion of the stem remains intact upon MBNL1 binding, while U2AF65 binds to a completely single-stranded structure.

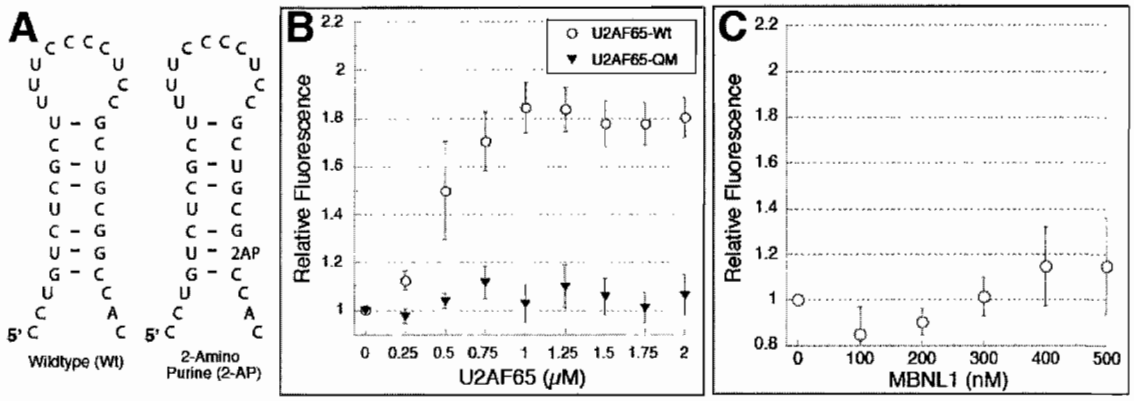

Figure 6. ***U2AF65 destabilizes the stem-loop upon binding, while MBNL1 binds the stem-loop structure.*** **(A)** Schematic of fluorescent RNA used to probe cTNT structure. **(B)** Effects of wild-type or Quad Mutant (QM) U2AFA65 on cTNT fluorescent probe. **(C)** Effect of MBNL1 on the fluorescent cTNT probe.

Discussion

MBNL1 regulates the cTNT exon 5 through competition with U2AF65

To our knowledge, the competition of MBNL1 with U2AF65 represents the first example of an alternative splicing factor that regulates splicing primarily through modulation of an RNA structural element and not through direct occlusion of another splicing factor's binding site. The competition of MBNL1 with U2AF65 does not constitute a novel mechanism, as other alternative splicing factors have previously been shown to compete directly with U2AF65. The alternative splicing factor Sex Lethal (Sxl) has been shown to compete with U2AF65 in the *tra* and *msl-2* pre-mRNAs (*88, 89*), while the Poly-Pyrimidine Tract Binding Protein (PTB) has been observed to compete with U2AF65 for sequences within the α- or β-tropomyosin pre-mRNAs (*94, 95*).

Sxl and PTB both have binding specificities similar to that of U2AF65, as all three proteins primarily prefer long runs of single-stranded uracil residues (*94*). When competitively binding with U2AF65, both Sxl and PTB are considered to occlude the U2AF65 binding site, and sterically inhibit U2AF65 from accessing its binding site. On the other hand, MBNL1 has a

different binding specificity from U2AF65, and binds adjacent sequences primarily outside of the py-tract. We cannot rule out steric contributions to the competition between MBNL1 and U2AF65, but the fluorescence assay strongly suggests that these proteins bind this RNA in different conformations (Figure 6). Both the differential binding sites of MBNL1 and U2AF65, and the role that RNA secondary structure can play in modulating U2AF65 binding, suggest that these splicing factors compete for recognition of the 3' end of intron 4 largely through binding mutually exclusive RNA structures (see Figure 7 for model).

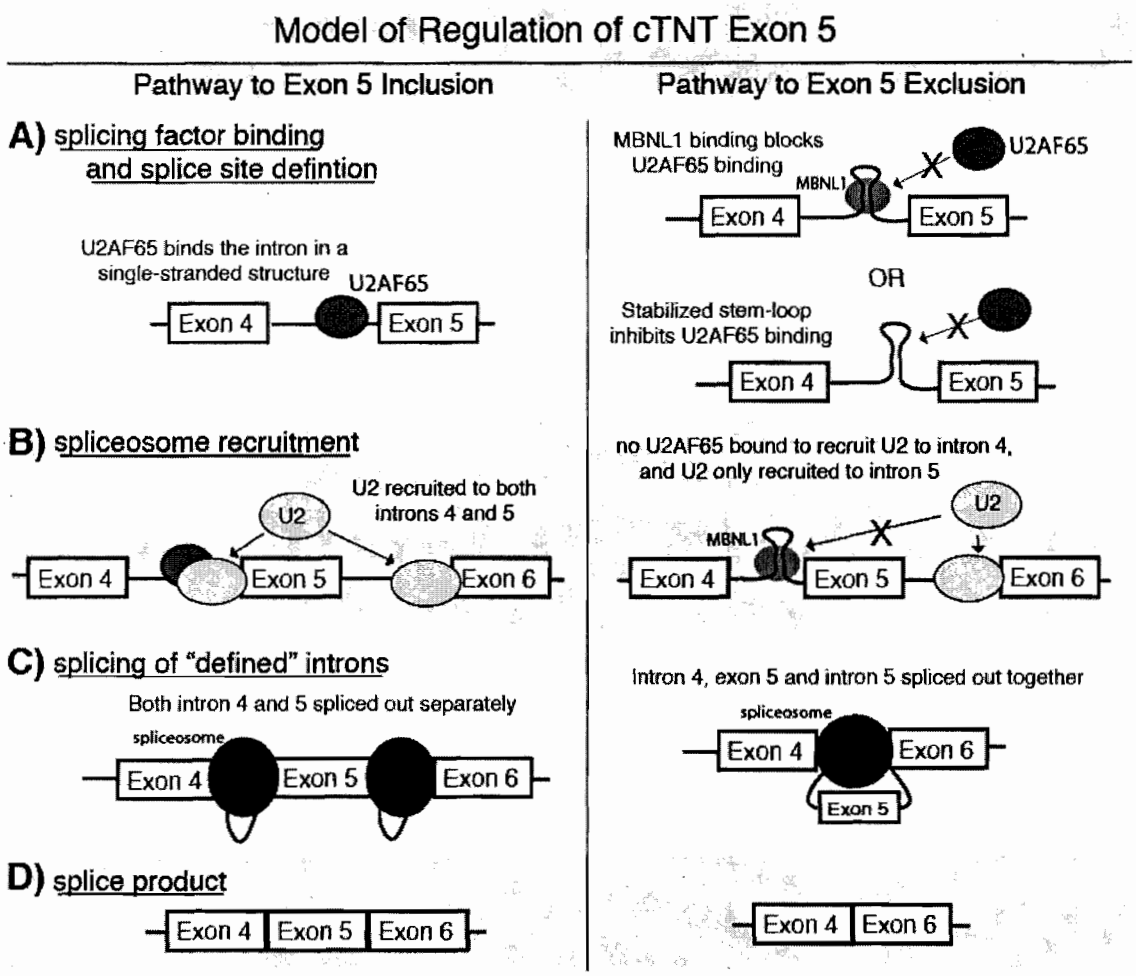

Figure 7. ***Model of regulation of cTNT exon 5 by MBNL1 and RNA structure.*** **(A)** Initial recognition of intron 4 by splicing factors. U2AF65 binds the intron in a single-stranded structure, but U2AF65 is inhibited if it cannot destabilize the stem due to MBNL1 binding or mutations that stabilize the stem-loop. **(B)** U2 snRNP recruitment to all 3' splice sites were U2AF65 is present. **(C)** Spliceosomal recruitment and splicing. **(D)** mRNA splice products.

The novel role of MBNL1 in regulating alternative splicing by modulating RNA secondary structure

It is well documented that RNA structural elements alone can regulate alternative splicing. For instance, stable structures have been shown to inhibit U1 snRNP binding to the 5'-splice site following exon 7 of both the *SMN1* and *SMN2* pre-mRNAs (*96*). Other structural elements that encompass exon 6B in chicken β-tropomysin have been shown to inhibit binding of all the U snRNPs, promoting the skipping of that exon (*97*). At present, it is unclear whether these examples involve alternative splicing factors that regulate these RNA structures, as MBNL1 appears to regulate secondary structures within the cTNT pre-mRNA.

Other alternative splicing factors have been postulated to regulate RNA structural elements, but there is no clear evidence to show that these factors modulate RNA structure in a way that directly regulates alternative splicing. For example, the splicing factor hnRNP A1 has been shown to inhibit binding of splicing factors ASF/SF2 and SC35 when they regulate splicing of the *tat* pre-mRNA from HIV-1 (*98*). Binding of hnRNP A1 is thought to alter the secondary structure of the pre-mRNA in a way that affects splicing (*99*), but this has not been demonstrated. Similarly, ribosomal protein L32 from *S. cerevisiae* binds a structured RNA element near the 5'-splice site of an intron within its own pre-mRNA, leading to intron retention (*100*). However, this structured binding site allows for recruitment of the U1 snRNP, showing that RNA structures do not always inhibit binding of the U snRNPs (*101*). It is still unclear how L32 causes intron retention if it does not inhibit binding of U1.

In regards to the competition between MBNL1 and U2AF65, it has been demonstrated with a crystal structure that U2AF65 binds the py-tract in a single-stranded structure (*102*). In this structure, seven uridine residues bind along two RNA recognition motifs of U2AF65, leaving the 3' and 5' ends spatially far from each other. This crystal structure sheds light on why U2AF65 must bind the cTNT intron in a single-stranded form, as it is unlikely that the stem could form with the loop in such an extended structure. At present, MBNL1 is the first alternative splicing factor that has been shown to regulate alternative splicing on the level of RNA structure. It is likely that other examples will follow as it is becoming increasingly apparent that RNA structure plays important roles in controlling pre-mRNA splicing.

Materials and methods

Cloning and protein purification. MBNL1 and U2AF65 were purified as previously described (*84, 91*).

RNA synthesis, labeling and purification. The RNA substrates used for the Complex formation assay were transcribed with T7 polymerase, off linearized Puc19 plasmid. During transcription, RNAs were radiolabeled using [α-^{32}P]CTP. All other RNA substrates were ordered from IDT DNA, and 5' end-labeled using [γ-^{32}P]ATP.

Cross-linking assay. RNA and proteins were incubated at room temperature for 10 minutes, then on ice for 10 minutes, under the same conditions as used in the gel shift assay. Samples were placed on a pre-chilled block and exposed to UV for 1 minute 20 seconds (1 Joule / cm^2), 3.5 cm from the light source using a FB-UVXL-1000 lamp (Fisher Scientific). Samples were then incubated in protein denaturing buffer for 2 minutes at 95 °C and were resolved on a 10% denaturing SDS-PAGE gel. Gels were then dried and autoradiographed.

Complex formation assay. HeLa nuclear extract was prepared (*103*) and complex formation was performed (*5*) as previously described.

Thermal melting curves. T_m values were obtained as previously described (*84*).

Gel shift assay. Final conditions for the RNA-U2AF65 binding experiments were 175 mM NaCl, 5 mM $MgCl_2$, 20 mM Tris (pH 7.5), 1.25 mM BME, 12.5% glycerol, 2 mg/ml BSA and 0.1 mg/ml Heparin. Prior to incubation, RNA substrates were snap annealed in 66 mM NaCl, 6.7 mM $MgCl_2$, and 27 mM Tris (pH 7.5). Protein was then added to the RNA, in a 10 µL volume and incubated for 10 minutes at room temp before 3-5µL were loaded on a pre-chilled 5% acrylamide (37.5:1), 0.5x TB gel, and run for 1 hour at 175 volts, at 4 °C. Gels were dried and autoradiographed.

For binding curves, gels were quantified using ImageQuant (Molecular Dynamics). The percent RNA bound was determined by taking the ratio of RNA:Protein complex to total RNA, per lane. Binding curves were graphed and apparent K_d values were determined with KaleidaGraph (Synergy) software using the following equation: y=((m2 + m1 + m0)-((-m2 - m1 -

$m0)^2$-(4 x m1 x m0))$^{0.5}$/(2 x m1), where y=% RNA bound, m2= K_d, m1=total RNA concentration and m0=protein concentration. This equation assumes a 1:1 interaction between the RNA and protein, which allows only an apparent K_d to be determined for RNAs containing more than one binding site. To determine the standard error of the apparent K_ds, 3-5 binding titrations were performed with each substrate and the apparent K_d values determined for each titration separately.

Complex formation assay. HeLa nuclear extract was prepared (*103*) and complex formation was performed (*5*) as previously described. In short, HeLa extract was added to a solution with RNA and buffer and incubated at 30 °C for the appropriate time point. The sample volume was 12 μL, and contained 25% HeLa extract, 55 mM KOAc, 30 mM KCl, 14.5 mM Hepes (pH 8.0), 10 % glycerol, 0.12 mM EDTA (pH 8.0), 0.5 mM DTT, 3.2 mM $MgCl_2$, 0.5 mM ATP, 20 mM creatine phosphate, 0.7 units / μL of RNase Inhibitor (Ambion). After the appropriate time point, samples were incubated with 4 μL of heparin (4 mg/mL), and incubated at room temp for 2 minutes. ADML and cTNT-wt samples were then loaded onto a pre-chilled 1.8% agarose gel (1x TG), and run at 75 volts for 1.5 hours, at 4 °C, in a 1x TG solution. To better resolve A complex, cTNT-4G samples had 80 mM KOAc, and were run on a 1.2% agarose gel (1x TG), for 4.5 hours at 40 volts, 4 °C. The gels were then fixed in 10% acetic acid, 10% methanol solution for 10 minutes. Gels were dried, autoradiographed, and quantified using ImageQuant (Molecular Dynamics). To determine the fold reduction, the percent of complex formed in the absence of MBNL1 was divided by the percent complex formed in the presence of MBNL1.

2-Amino purine fluorescence. RNA (Dharmacon RNA Technologies) was diluted to 1 μM concentration in the binding buffer (175 mM NaCl, 20 mM NaCl, 5 mM $MgCl_2$) and snap annealed. All reactions were performed at room temperature, in the binding buffer, using an L-formate Jobin-Yvon Horiba Fluoromax fluorimeter. A 3 mm wide Spectrosil microcell cuvette (Starna Cells, Inc) was used. The 2-amino purine was excited at 312 nm, and spectra were collected from 310 – 420 nm. The fluorimeter slits were 3 nm , with an integration time of 0.2 seconds. Spectra collected with buffer and protein were used to subtract out background. Three spectra were collected and averaged for every titration point. To calculate relative fluorescence, the value at 375 nm was used.

In vivo *splicing.* HeLa cell transfections, and RT-PCR was done as previously described (*84*).

CHAPTER IV

PENTAMIDINE REVERSES THE SPLICING DEFECTS ASSOCIATED WITH MYOTONIC DYSTROPHY

Contribution note: this chapter is under review for publication at PNAS. The other contributors were Masayuki Nakamori, Catherine M. Matthys, Charles A. Thornton and J. Andrew Berglund. Masayuki Nakamori was co-first author with me on this publication and helped with experimental design, performed many experiments and did a large amount of data analysis. Catherine M. Matthys performed many initial experiments. Charles A. Thornton helped with experimental design and data analysis. J. Andrew Berglund helped with experimental design, data analysis and manuscript preparation.

Introduction

Myotonic Dystrophy (DM) is an autosomal dominant genetic disorder that is characterized by a variety of symptoms. There are two types of Myotonic Dystrophy, type 1 (DM1) and type 2 (DM2). DM1 is linked to a $(CTG)_n$ repeat expansion in the 3' untranslated region of the *DMPK* gene, while DM2 is linked to a $(CCTG)_n$ repeat expansion in intron 1 of the *ZNF9* gene. The current model is that the expanded repeats are toxic on the RNA level, where either repeat can form a stable structured RNA that aberrantly interacts with proteins in the nucleus (for review see (*22, 85*)).

One proposed molecular mechanism that may account for the disease symptoms is that, upon transcriptions of the expansions, either the CUG or CCUG repeats sequester RNA binding proteins from their normal cellular functions. The protein MBNL1 (Muscleblind-like 1) has been shown to bind both expanded CUG and CCUG repeats *in vitro* (*38, 66, 75*), and co-localize with these expanded repeats *in vivo* (*23, 51-54*). MBNL1 is an alternative splicing factor and its sequestration leads to the mis-splicing of multiple pre-mRNAs in DM, which is thought to give rise to many of the symptoms of the disease. In support of this model, it has been shown that the disruption of the MBNL1 gene or expression of CUG repeats in mice causes symptoms and mis-splicing similar to those seen in DM patients (*46, 47, 62, 65*). Furthermore, disease symptoms can

be rescued and mis-splicing of many pre-mRNAs can be reversed in mice expressing CUG repeats by over-expression of MBNL1 (*34*).

Aside from the over-expression of MBNL1, another possible approach to overcoming the sequestration of MBNL1 is to identify small molecules that specifically bind the CUG repeats and would competitively release the sequestered MBNL1. As a step towards identifying a small molecule therapeutic for DM, a small library of molecules known to bind structured nucleic acid were screened for their ability to disrupt an MBNL1-CUG repeat interaction *in vitro*. Two molecules were identified that strongly disrupted the complex *in vitro*. Further testing showed that the molecule pentamidine was able to rescue mis-splicing of four tested pre-mRNAs in both cell culture and a mouse model of DM1.

Results

Identification of small molecules that disrupt an MBNL1–CUG repeat complex in vitro

Twenty-six small molecules that are known to bind structured nucleic acid (Table 1) were screened to identify compounds that would disrupt a protein-RNA complex using a previously characterized MBNL1-CUG repeat gel shift assay (Figure 1A) (*84*). The CUG RNA (CUG_4) used in this screen contains four repeats, which were stabilized into a stem-loop structure by using the ultra stable UUCG loop (Figure 1A). The molecule that most effectively disrupted the complex was pentamidine, which competed with MBNL1 and disrupted the protein-RNA complex with an IC_{50} of 58 ± 5 μM (Figures 1B-D). Neomycin B was the next best molecule with an IC_{50} of 280 (Figures 1B-D). The small molecules ethidium bromide and thiazole orange also disrupted the MBNL1-CUG_4 complex with an IC_{50} in the 200-300 μM range (Table 1). However due to toxicity issues with ethidium bromide and thiazole orange, only pentamidine and neomycin B were chosen for further study to determine their ability to disrupt the MBNL1-CUG repeat interaction *in vivo*.

Pentamidine rescues mis-splicing in HeLa cells of two pre-mRNAs mis-regulated in DM

To determine if either small molecule could rescue mis-splicing in cells, we investigated the alternative splicing of the insulin receptor (IR) and cardiac troponin T (cTNT) minigenes in HeLa cells, in the presence or absence of expanded CUG repeats. Both of these pre-mRNAs are mis-

Table 1. ***List of 26 screened small molecules.*** Molecules are divided into groups based on structure. Listed are the providers and product order number.

Molecule	EC_{50} (μM)	Prov-ider	Product No.	Known Nucleic Acid Binding Activity	Ref.
Aminoglycosides					
Actinomycin D	/	Sigma	A4262	Strongly Binds DNA CTG repeats, with the T:T mismatch as the main recognition site	(*104, 105*)
Apramycin Sulfate	/	Sigma	A2024	Binds rRNA in eukaryotes and prokaryotes	(*106, 107*)
Geneticin Disulfate	/	Sigma	G5013	Binds A site in bacterial ribosome	(*108*)
Genistein	/	Sigma	91955	Binds RNA and DNA	(*109*)
Gentamicin Sulfate	~2,000	Sigma	G3632	Binds A site in bacterial ribosome	(*110, 111*)
Kanamycin Sulfate	/	IBI Sci	IB02120	Binds A site in bacterial ribosome	(*110, 111*)
Lividomycin A Sulfate	/	Sigma	L4518	Binds A site in bacterial ribosome	(*106, 110*)
Neomycin B	280 ± 40	Sigma	N1876	Binds A site in bacterial ribosome	(*110, 111*)
Paromomycin Sulfate	/	Sigma	P9297	Binds A site in bacterial ribosome	(*106, 111*)
Proflavine Hydrochloride	~1,600	MP Bio	210360010	Binds DNA with alternating purines and pyrimidines	(*112*)
Ribostamycin Sulfate	/	Sigma	R2255	Binds A site in bacterial ribosome	(*106, 110*)
Streptomycin Sulfate	/	Sigma	S6501	Binds bacterial 16 S rRNA	(*113*)
Tetracyclin Hydrochloride	/	IBI Sci	IB02200	Binds A and P site in bacterial ribosome	(*114*)
Tobramycin Sulfate	/	Sigma	T1783	Binds A site in bacterial ribosome	(*111*)
Dyes					
Bisbenzimidine Hoechst No. 33342	~1,000	Sigma	B2261	Binds minor groove of DNA, in AT rich regions	(*115*)
Bisbenzimidine Hoechst No. 33258	~1,000	Sigma	B2883	Binds and inhibits splicing of Group 1 introns, binds TAR RNA from HIV	(*115, 116*)
DAPI	/	Sigma	D9542	Intercalates DNA and A-form RNA	(*117, 118*)
Ethidium Bromide	210 ± 40	Biorad	161-0433	Intercalates A-form RNA	(*119, 120*)
Thiazole Orange	300 ± 60	Sigma	17237	Binds RNA	(*121*)
Amines					
9-Aminoacridine Hydrochloride Hydrate	/	Sigma	A38401	Intercalates DNA with high affinity	(*122*)
L-Arginiamide Dihydrochloride	/	Sigma	A3913	Binds non-canonically structured DNA	(*123*)
Pentamidine Isethionate	58 ± 5	Sigma	P0547	Binds structured RNA. Binds AT rich DNA in the minor groove. Inhibits splicing of group I and II introns. Modifies Ubiquitin.	(*124-127*)
Spermidine	/	Sigma	S2626	Binds tRNA.	(*128*)
Spermine	/	Sigma	S3256	Binds tRNA, and A-form DNA.	(*129, 130*)
Misc.					
Ampicillin	/	IBI Sci	IB02040		
Berenil	~2,000	Sigma	D7770	Binds Tar RNA from HIV and structured RNA stems.	(*117*)

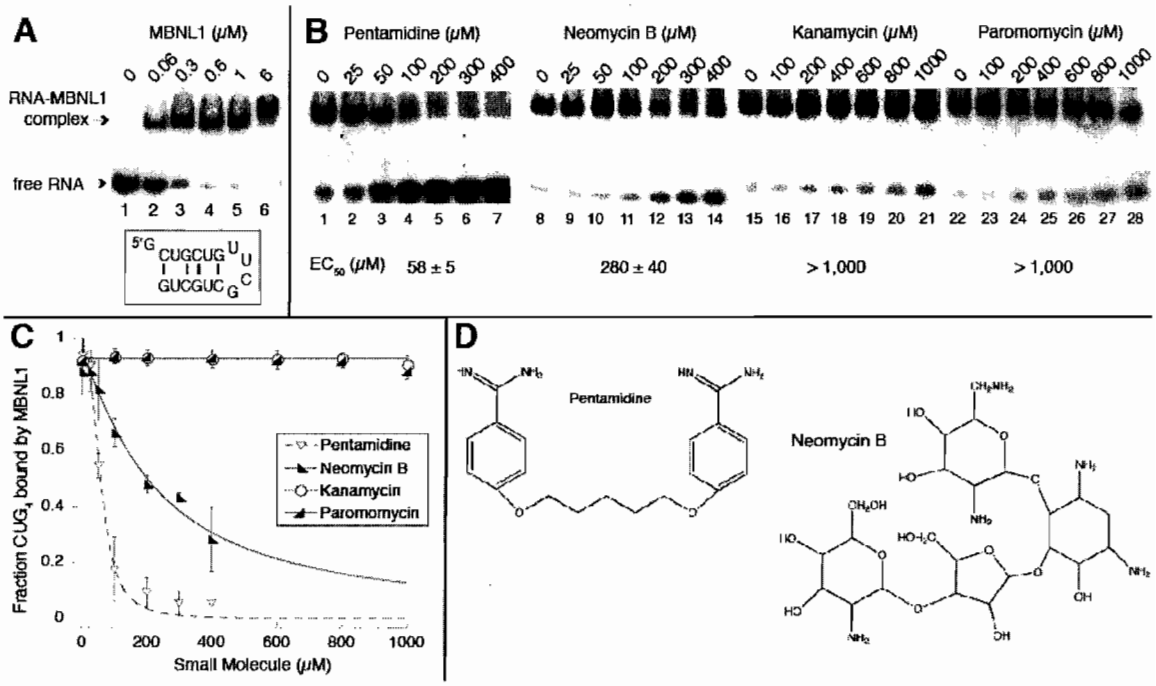

Figure 1. ***Screen for small molecules that disrupt an MBNL1-CUG complex*** **in vitro.**
(A) Representative gel of MBNL1 complexed with the CUG_4 construct. **(B)** Representative gels of pentamidine and neomycin B disrupting the MBNL1-CUG_4 complex, and two small molecules that had no effect, kanamycin and paromomycin. The MBNL1 concentration in each lane is 500 nM. **(C)** Graph depicting competition of pentamidine, neomycin B, kanamycin and paromomycin with MBNL1. **(D)** Molecular structure of pentamidine and neomycin B.

spliced in DM patients (*25, 26, 28, 56, 60, 62, 84, 131*). MBNL1 represses inclusion of exon 5 in the cTNT pre-mRNA and binds directly upstream of exon 5 (*28, 84*). MBNL1 positively regulates the inclusion of exon 11 of the IR pre-mRNA, but it is unknown if MBNL1 directly binds this pre-mRNA (*25, 26, 28, 49, 56*). It has been shown that the over-expression of 960 interrupted CUG repeats sequesters MBNL1 to foci containing the CUG repeats, and ablates the regulation of multiple pre-mRNAs by MBNL1, including the regulation of these two minigene. pre-mRNAs (*25, 28, 49, 84*). A plasmid containing these 960 repeats can be transfected into HeLa cells, thus providing a cell based model for studying the splicing defects of DM.

We observed that the addition of pentamidine was able to rescue the mis-splicing of both the cTNT and IR minigene pre-mRNAs caused by the expression of the expanded CUG repeats (Figure 2A-C). For the IR pre-mRNA, the basal level of inclusion of exon 11 in HeLa cells is 97 ± 1%, while co-transfection with the CUG repeat plasmid reduces inclusion to 73 ± 1%. The mis-splicing was mostly rescued at 25 μM pentamidine, returning exon 11 inclusion to 92 ± 3% (Figure 2A-B). Exon 11 levels were fully rescued to normal levels by 50 μM pentamidine.

We next tested if mis-splicing of the cTNT pre-mRNA could be rescued by pentamidine (Figure 2C). Exon 5 inclusion was seen to have a basal level of inclusion of 75 ± 4% in HeLa cells, while co-transfection of CUG repeats raised it to 95 ± 2%. Pentamidine also rescued this mis-splicing, though higher concentrations were required for rescue compared to the IR substrate. Inclusion of the cTNT exon 5 to normal levels was observed at 75 μM pentamidine, while 50 μM pentamidine showed a partial rescue, and 25 μM pentamidine had no effect (Figure 2C). Unexpectedly, in the absence of CUG repeats, pentamidine was observed to reduce the inclusion of exon 5. We recently reported that MBNL1 competes with the essential splicing factor U2AF65 at the 3' end of intron 4 to block inclusion of the cTNT exon 5 (*132*). MBNL1 binds a stem-loop in this area and this interaction blocks U2AF65 binding in this region of the intron because U2AF65's single-stranded binding site is sequestered in the stem-loop. We hypothesized that pentamidine was binding this stem-loop and blocking U2AF65 binding, through a mechanism similar to MBNL1 inhibition. Supporting this model we found that pentamidine competed with U2AF65 for binding to the 3' end of the intron (Figure 3). This result suggests that for the cTNT pre-mRNA both MBNL1 and pentamidine bind the same site to additively repress exon 5 inclusion in the absence of the CUG repeats, thus explaining the low levels of exon 5 inclusion in the absence of CUG repeats (Figure 2C). However, as the effect is less penetrant in the presence of CUG repeats, it appears that pentamidine is titrated to these repeats when they are present.

Since pentamidine was found to regulate cTNT exon 5 levels of inclusion in the absence of CUG repeats, we wanted to investigate if pentamidine might generally affect splicing of other RNAs. We first tested its affect on the splicing of a highly expressed pre-mRNA from another minigene, as the previous pre-mRNAs tested were also from minigenes. The splicing of the PLEKHH2 pre-mRNA (that contains exons 20-22, all of which are constitutively included when endogenously expressed) was tested and found to be unaffected by the addition of either pentamidine or the expression of CUG repeats (Figure 2D). We next determined if pentamidine affected the alternative splicing of endogenous pre-mRNAs in HeLa cells. None of the four alternatively spliced exons tested were affected by either the expression of the CUG repeats or the addition of 75 μM pentamidine (Figure 2E). These results suggest that pentamidine does not globally affect either constitutive or alternative splicing, but only a subset of pre-mRNAs, which are MBNL1 dependent. Furthermore, it suggests that pentamidine does not affect either highly expressed pre-mRNAs (such as the PLEKHH2, which was expressed from a plasmid with a strong promoter), or the four endogenously expressed gene, which are likely to be expressed at much lower levels. Finally, it should be noted that transcript levels appeared relatively unchanged upon the addition of pentamidine for any of these pre-mRNAs. This indicates that pentamidine does not globally affect transcription or RNA turnover at the concentrations tested.

To determine if neomycin B was able to rescue the mis-splicing induced by CUG repeat expression, it was also tested in the splicing assay. Neomycin B did not alter the splicing of either the cTNT or IR pre-mRNAs (Fig 4). For the cTNT pre-mRNA, concentrations up to 200 μM neomycin B were tested (Fig 4A), for the IR pre-mRNA 500 μM was tested (Fig 4B). These results indicate that neomycin B does not specifically bind the CUG repeats in this DM1 cell model, suggesting that neomycin B is likely binding other substrates in the cell.

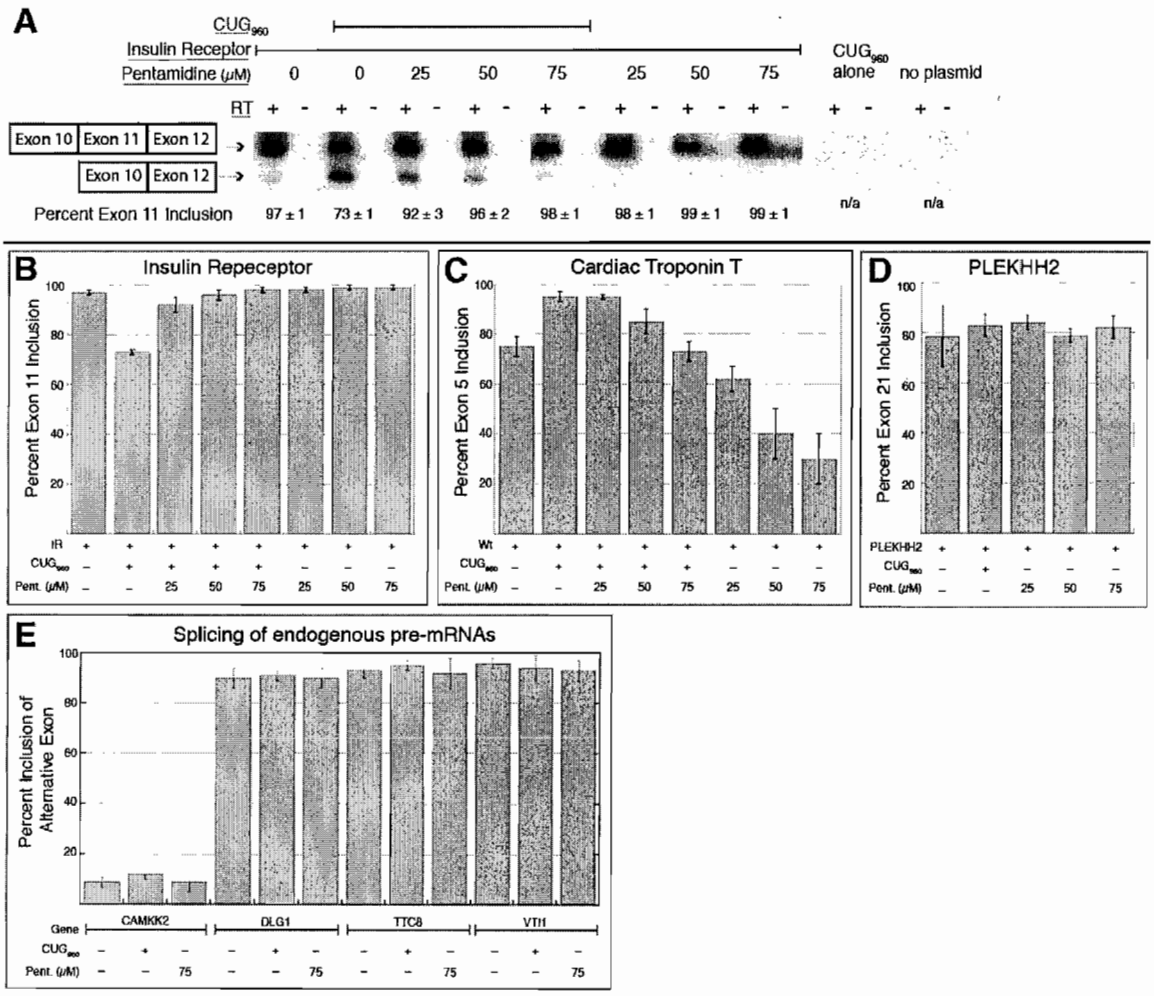

Figure 2. ***Pentamidine rescues mis-splicing of MBNL1 dependent pre-mRNAs in HeLa cells.*** **(A)** Representative RT-PCR data for the IR minigene pre-mRNA (exons 10-12) under pentamidine treatment. **(B)** Bar graph representation of pentamidine rescuing mis-splicing of IR exon 11. **(C)** Bar graph representation of pentamidine rescuing mis-splicing of cTNT exon 5. **(D)** Bar graph representation of splicing of the PLEKHH2 minigene (exons 20-22) in the presence of either CUG repeats or pentamidine. **(E)** Bar graph representation of splicing of four endogenously expressed genes in HeLa that are alternatively spliced. Splicing of the alternative exon was tested in the presence of either CUG repeats or 75 μM pentamidine.

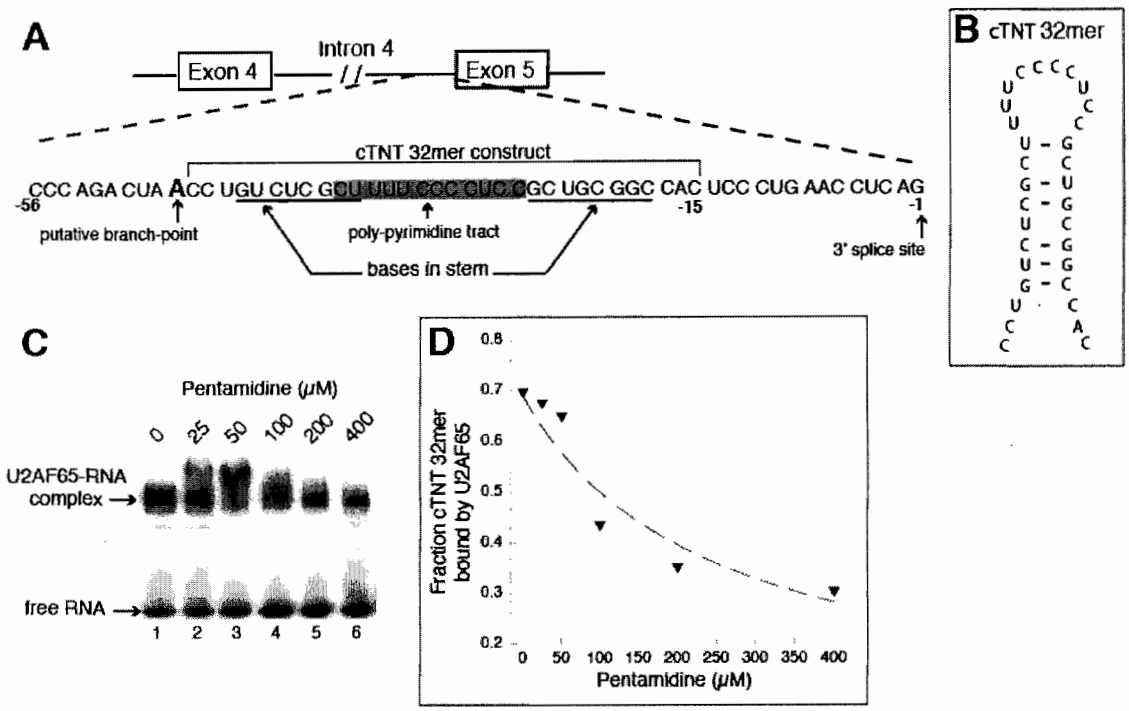

Figure 3. ***Pentamidine competes with U2AF65 binding upstream of cTNT exon 5.*** **(A)** Schematic of the 3' end of intron 4 in the cTNT pre-mRNA. Noted are the 3' splice site, the poly-pyrimidine tract (which is the U2AF65 binding site), the putative branch-point and the bases in the stem-loop which is MBNL1's binding site. **(B)** Structure of the stem-loop which MBNL1 binds. **(C)** U2AF^5 was pre-bound to the cTNT 32mer at a concentration of 750 nM, and a titration of pentamidine disrupted the protein-RNA complex. **(D)** Graph of pentamidine's competition with U2AF65.

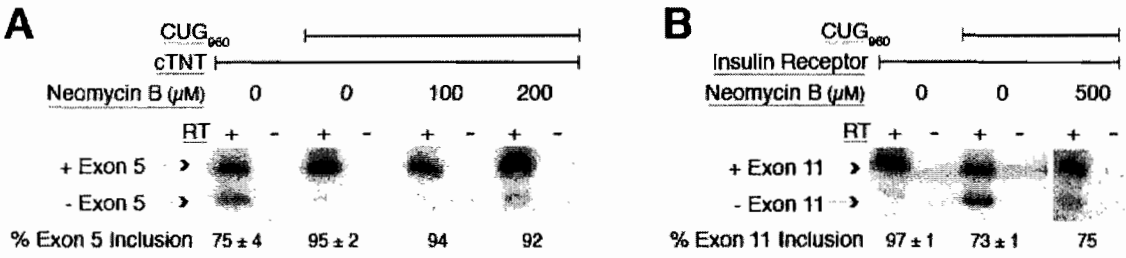

Figure 4. ***Neomycin B fails to rescue mis-splicing of MBNL1 dependent pre-mRNAs in HeLa cells.*** **(A)** RT-PCR data showing that neomycin B has no affect on mis-splicing of the cTNT exon 5. **(B)** RT-PCR data showing that neomycin B, at even higher concentrations, has no affect on mis-splicing of the insulin receptor exon 11.

Pentamidine reduces CUG repeat foci formation and relieves MBNL1 sequestration in HeLa cells

To directly visualize pentamidine's effect on MBNL1 sequestration, CUG foci formation was investigated in the presence and absence of the small molecule. Using an anti-MBNL1 antibody, MBNL1 localizes to the nucleus of HeLa cells, without CUG repeats or pentamidine present (Figure 5A-D). Upon transfection with the CUG_{960} plasmid, characteristic RNA ribonuclear foci were formed, which sequestered nearly all of the MBNL1 (Figure 5E-H). Upon the addition of 75 μM pentamidine, we observed that foci formation was decreased. We scored CUG transfected cells, with or without pentamidine, for foci formation. Cells were scored as either having foci or having no foci. Foci formation was reduced by 21% in the pentamidine treated cells. While this is not a full reduction, a chi test gave a P value of 0.004, showing the difference between treated and untreated cells was significant. In cells treated with pentamidine where foci were eliminated, MBNL1 was found to be diffuse throughout the nucleus (Figure 5I-L). This assay was also performed in HEK293 cells, which contain higher levels of endogenous MBNL1. In this cell line, foci formation was reduced 28% by treatment with 50 μM pentamidine. A chi test gave a P value of 0.5×10^{-40}, showing the difference between treated and untreated cells was also significant in this cell line.

Pentamidine partially rescues mis-splicing of two pre-mRNAs in a DM1 mouse model

We tested pentamidine's ability to rescue mis-splicing in a mouse model (HSA^{LR}) for DM1 in which 250 CUG repeats are expressed under an actin promoter (*45*). Pentamidine was administered by intraperitoneal injection in two dosage regimens, and mis-splicing of the chloride-1 (Clc-1) and Serca1 pre-mRNAs was assayed. Both of these pre-mRNAs are mis-splicing in DM; MBNL1 represses inclusion of exon 7a in the Clc-1 mRNA and enhances inclusion of exon 22 in the Serca1 mRNA (*24, 85, 133*).

Adult mice normally have a Clc-1 exon 7a inclusion of 6 ± 1%, while the HSA^{LR} line has a level of 54 ± 6%. Pentamidine treatment at a lower dosage (25 mg/kg twice at day) partially rescued inclusion to 37 ± 3%, while a higher dosage regimen (40 mg/kg once a day) rescued splicing to 34 ± 4% exclusion (Figure 6A, C). When comparing the low dosage treatment to injection of only saline, the P value was <0.003 when using a Mann-Whitney U-test. The P value for the high dosage treatment was <0.001. The number of mice used for each trial is noted in Figure 6C-D.

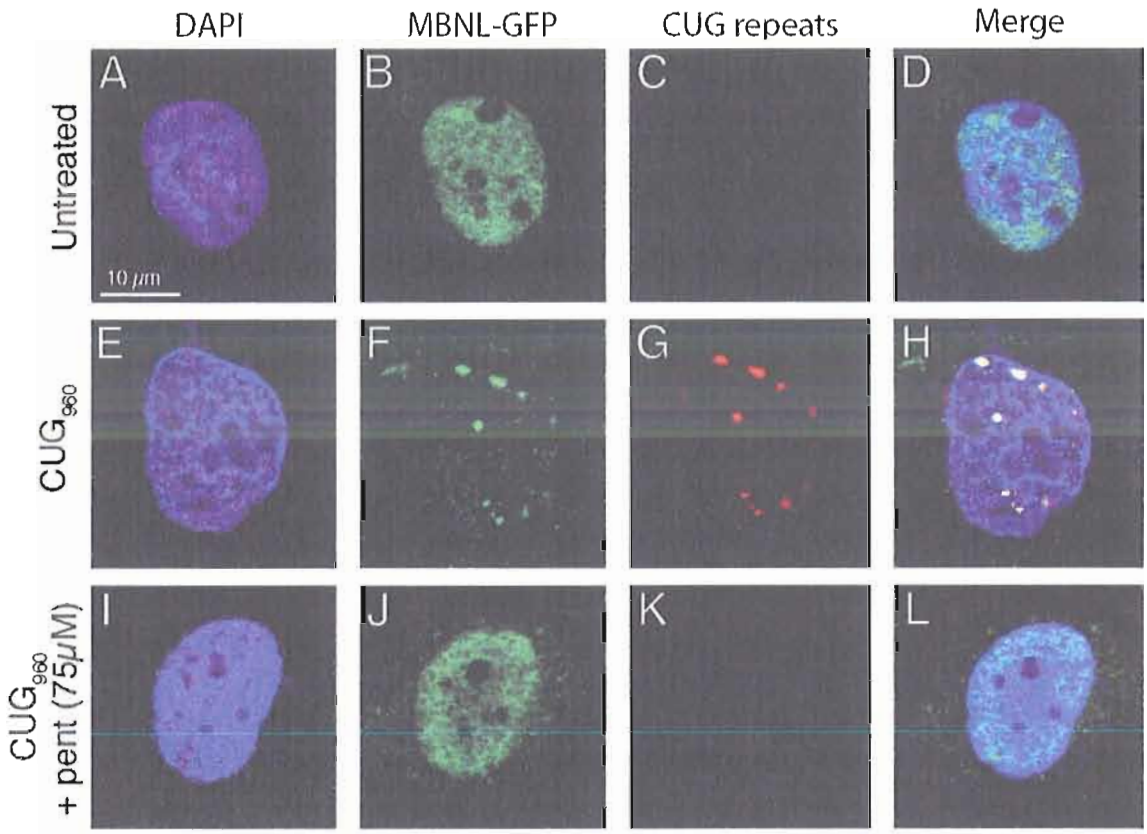

Figure 5. ***Pentamidine reduces the formation of CUG ribonuclear foci and relieves MBNL1 sequestration.*** **(A-D)** Untreated HeLa cells, with MBNL1 localizing to the nucleus. **(E-H)** Transfection with CUG_{960} gives rise to ribonuclear foci, characteristic of DM, with MBNL1 co-localizes with the CUG repeats at the foci. **(I-L)** In cells where pentamidine (75 μM) reduced formation of the foci, MBNL1 sequestration is fully relieved.

Inclusion of the Serca1 exon 22 is normally 99.7 ± 0.1% in adult mice, while HSA^{LR} mice have drastically lower levels of 17 ± 4%. Lower dosage pentamidine treatment minimally rescued exon 22 inclusion to 22 ± 2%, while the higher dosage treatment more significantly rescued exon 22 inclusion to 40 ± 4% (Figure 6B, D). The P value for the lower dosage treatment regimen was < 0.01, while the higher dosage regimen had a P value that was < 0.001.

Further increases to the dosage regimens had substantial toxicity in the mice. An increase to 30 mg/kg twice daily was lethal, while dosages at or less than 25 mg/kg twice a day, or 40mg/kg once a day were not. It is likely that higher levels of pentamidine could indeed further reverse the mis-splicing of these transcripts. However, this full rescue may not be achievable with pentamidine in HSA^{LR} mice, due to the toxicity. Nonetheless, the partial rescue that was observed was statistically significant and strongly indicates that a small molecule can indeed work in a mammalian DM model to reverse mis-splicing.

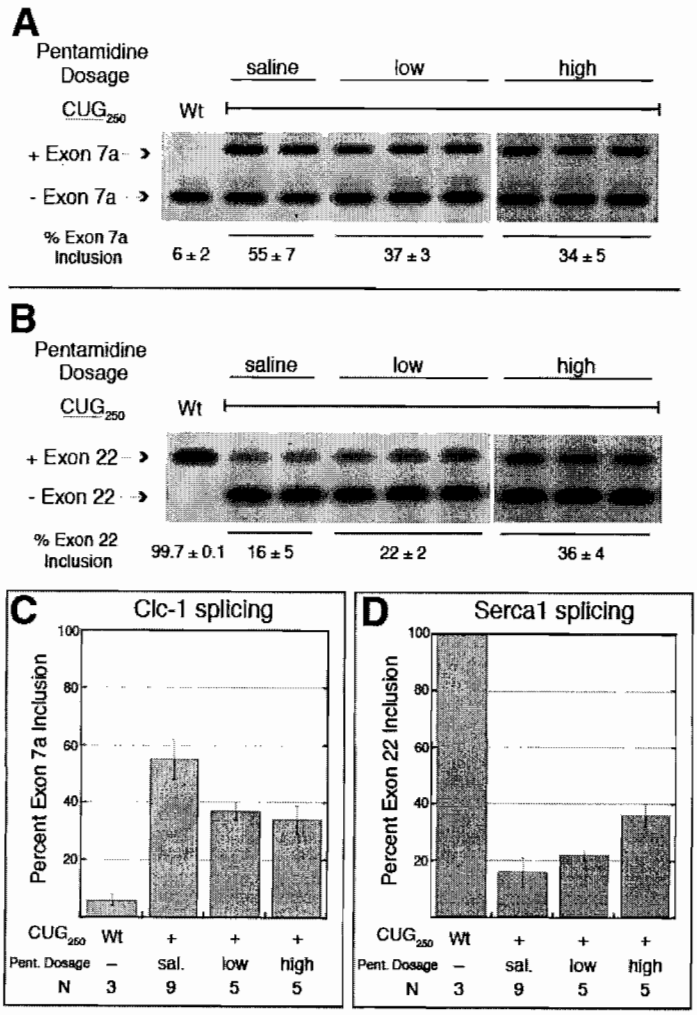

Figure 6. ***Pentamidine partially rescues mis-splicing of two pre-mRNAs in a DM1 mouse model.***
(A) RT-PCR analysis showing a dose dependent partial rescue of mis-splicing of Clc-1 exon 7a. **(B)** RT-PCR of partial dose dependent rescue of the mis-splicing of Serca1 exon 22. **(C)** Bar graph representation of partial rescue of Clc-1 exon 7a splicing. The number of mice used for each sample is denoted as N. Using a Mann-Whitney U-test, the significance of rescue for both the low compared to the saline treatment had a P value < 0.003, while the high dose treatment had a P value < 0.001. **(D)** Bar graph representation of partial rescue of Serca1 exon 21 splicing. The low dosage had a P value < 0.01, while the high dosage had a P value < 0.001.

Discussion

Small molecules that bind nucleic acids as therapeutic agents for disease treatment

Many small molecules have been shown to bind and affect nucleic acid structure and function. Examples include aminoglycosides binding to specific sites in ribosomal RNA to inhibit various steps of translation or cisplatin cross-linking DNA to inhibit transcription and replication (*134-136*). Dervan and colleagues have been quite successful in designing molecules that target specific sequences in DNA (*137*), but that technology doesn't exist for targeting specific RNA structures and sequences. Efforts are currently underway to understand the rules of small molecule-RNA interactions (*138, 139*). To determine if CUG repeats could be targeted by small molecules we screened a small library of known nucleic acid binding molecules and found that pentamidine and neomycin B bound CUG repeats and could displace MBNL1 *in vitro*. This limited screen of small molecules, plus recent screens by Gareiss and colleagues and Pushechnikov and colleagues with short peptides and peptoids (*140, 141*), clearly demonstrates that CUG repeats can be targeted and that high throughput screens with large libraries of small molecules will almost certainly produce many additional lead compounds other than pentamidine for further studies.

Pentamidine is currently used to treat *Pneumocystis carinii* infections (pneumonia) in AIDS (Acquired Immuno-Deficiency Syndrome) patients and to treat patients with Trypanosomiasis and Leishmaniasis infections (*142, 143*). Although pentamidine has been shown to bind to both DNA and RNA, the mechanism of its antifungal activity is unknown. Pentamidine may inhibit *P. carinii* growth by inhibiting the splicing of essential group I introns in this organism (*124, 126*), although more recently pentamidine has also been shown to inhibit translation (*144*). It has also been proposed that pentamidine blocks DNA replication (*143*), as a structure of pentamidine with DNA shows that it binds DNA in the minor groove (*127*). Therefore it is likely that pentamidine interacts with many different nucleic acid targets within cells. Our results suggest pentamidine can, at least, partially target CUG repeats and release MBNL1. It is interesting to note that we previously showed that CUG repeats form an A-form structure, despite the U-U mismatches (*40*), although the minor grooves of A-form and B-form nucleic acid differ in architecture it is tempting to speculate that pentamidine binds the minor groove of the A-form CUG repeats.

The results with pentamidine are in contrast to neomycin B which also disrupted the MBNL1-CUG repeat interaction *in vitro* (Figure 1), but did not rescue mis-splicing of any of the

tested targets (Figure 4). Both pentamidine and neomycin B are known to bind many different nucleic targets, but only pentamidine displayed the ability to reverse the mis-splicing of four different pre-mRNAs affected by CUG repeats. These results suggest that pentamidine sufficient specificity for CUG repeats to have an effect in these assays, specificity that neomycin B apparently lacks.

The efficacy of pentamidine as a potential therapeutic for DM

The window of efficacy for pentamidine treatment in the DM1 cell culture and mouse model is narrow. Substantial cell death is observed in HeLa cells with pentamidine treatments of 100 to 125 μM. Raising the concentration of pentamidine for the mice treatments is also problematic because pentamidine dosages of 30 mg/kg twice a day are lethal to the mice, while 40 mg/kg twice once a day only partially relieves mis-splicing of the two tested pre-mRNAs (Figure 6). To overcome this toxicity, modifications to pentamidine will be performed to enhance the affinity and specificity of pentamidine for CUG repeats, thus allowing for a lower dosage treatment. The length of the carbon linker between the aromatic groups will be optimized for binding to the CUG repeats and functional groups containing H-bond donors and acceptors will be added to the molecule in an attempt to create new specific interactions with the CUG repeats. Another approach to improve pentamidine's affinity and specificity for CUG repeats will be to oligomerize pentamidine to obtain multivalent binding to the repeating CUGs. A multivalent strategy was recently used to obtain ligands that bind CCUG repeats with nanomolar affinity (*145*), supporting the idea that covalent linkage of multiple pentamidines should result in compounds with significantly better affinities to CUG repeats.

Previous potential therapeutic approaches for DM1 have ranged from over-expression of MBNL1, RNA interference against the CUG repeats, to targeted degradation of the mutant DMPK transcript with an RNA ribozyme (*34-37*). We have now demonstrated that releasing MBNL1 from the CUG repeats with a small molecule is another valid approach that should be considered for treating DM1.

Material and methods

MBNL1 cloning and purification. MBNL1 was PCR amplified and was cloned into GST fusion vector pGEX-6P-1 (Amersham), using DNA (MBNL1 isoform with amino acids 1-382) provided by Maury Swanson (*28*). The MBNL1(1-260) construct used in this study was cloned using BamHI and NotI restriction sites. Using BL21-Star expression cells (Invitrogen), protein expression was induced with 0.25 mM IPTG at an $OD_{600} \sim 0.5$-1, for 3-4 hours at 37°C. Cells were lysed in 30 mL of buffer (500 mM NaCl, 25 mM Tris pH 7.5, 10 mM β-Mercaptoethanol (BME) and 5% glycerol) using 1 mg/ml of lysozyme followed by sonication (3 x 30 seconds). Cell extract was centrifuged for 15 minutes at 17,000 rpm, and lysate which contained GST-MBNL1 was collected. GST-MBNL1 was bound to GST affinity beads for 30 minutes at 4°C. Beads were washed 5 times with buffer (1M NaCl, 25 mM Tris pH 7.5 and 5mM BME); MBNL1 was cleaved from the affinity tag with Precision Protease (Amersham) and collected from the beads. The protein was then run over an anion exchange (Q) column. Co-purifying contaminants bind the column, but MBNL1 does not. MBNL1 was collected in the column flow through, concentrated and dialyzed into storage buffer (30% glycerol, 300 mM NaCl, 20 mM Tris pH 7.5, 5 mM BME) and stored at -80 °C. Aliquots were removed from -80 °C and temporarily placed in -20 °C until used for an experiment.

RNA labeling and purification. All RNA substrates that were radiolabeled were ordered from IDT DNA, and 5' end-labeled using [γ-^{32}P]ATP and all RNAs were purified on an 8% denaturing acrylamide gel.

Gel shift assay. The final concentration in the reaction was 125 mM NaCl, 5 mM $MgCl_2$, 20 mM Tris (pH 7.5), 7.5% glycerol, 2 mg/ml BSA and 0.1 mg/ml Heparin. Prior to mixture with protein, RNA substrates were snap annealed (95 °C for 1 min then directly on ice for 20 min) in 66 mM NaCl, 6.7 mM $MgCl_2$, and 27 mM Tris (pH 7.5). Protein and the small molecule were then added to the RNA. The binding reaction was in 10 μL volume and was incubated for 10 min at room temperature (RT) before 2-4μL were loaded on a pre-chilled gel. RNA and RNA-protein complexes were separated on 3% acrylamide (37.5:1), 0.3% agarose (low melting point), 0.5x Tris-Borate (TB) gels that were run for 2-3 hours at 4 °C, 25-50 volts. For U2AF65, a 6% acrylamide (37.5:1), 0.5x TB was used and run at 175 volts, 4 °C for 2 hours. Gels were dried and

autoradiographed. All small molecules were dissolved in water, ethanol, or ethanol/DMSO (50:50) according to manufacture's guidance for solubility.

For binding curves, the apparent K_d values were determined as previously described (*84*). In brief, gels were quantified using ImageQuant (Molecular Dynamics). The percent RNA bound was determined by taking the ratio of RNA:Protein complex to total RNA, per lane. Binding curves were graphed and apparent K_d values were determined with KaleidaGraph (Synergy) software using the following Equation 1, where y=% bound, m2= K_d, m1=total RNA concentration and m0=protein concentration.

$$\text{Equation 1:} \quad y = \left((m0 + m1 + m2) - \sqrt{(-m0 - m1 - m2)^2 - (4 \times m0 \times m1)}\right) \Big/ (2 \times m1)$$

This equation assumes a 1:1 interaction between the RNA and protein, which allows only an apparent K_d to be determined if the RNA contains more than one binding site. To determine the standard error of the apparent dissociation constants, 3-5 binding titrations were performed with each substrate and the apparent K_d values determined for each titration separately, prior to averaging. The error bars on the binding curve were obtained by averaging the individual titration points and calculating the standard deviation. Data points greater than two standard deviations from the average were discarded, with at least three data points remaining for standard deviation analysis.

To determine the IC_{50}, Equation 2 was used, where m0= small molecule concentration, m1= IC_{50}, m2= Hill coefficient, and m3= fraction MBNL1 bound without small molecule present.

$$\text{Equation 2:} \quad Y = m3 / \left[1 + (m0/m1)^{m2}\right]$$

In vivo *splicing*. Wild-type cTNT and DMPK-CUG_{960} minigenes were obtained from the lab of Thomas Cooper (Baylor College of Medicine). The insulin receptor minigene was obtained from the lab of Nicholas Webster (UCSD). The PLEKHH2 minigene was cloned from HeLa genomic DNA. Exon naming was determined using release 18 of the UCSC genome browser.

HeLa cells were grown in monolayers in DMEM with GLUTAMAX (GIBCO) and supplemented with 10% fetal bovine serum (GIBCO). Approximately 2.0 (± 0.2) $x10^5$ cells were plated in 6 well plates and transfected 18-20 hours later at 70-90% confluency. 1 ug of plasmid was transfected into each well of cells, using 5 uL of Lipofectin2000 (Invtitrogen) according to the manufacturer's protocols. For co-transfection, 1 ug of total plasmid was transfected, at 500 ng of each construct. The remainder of total plasmid was empty vector, if needed. Cells were

incubated in low growth OPTI-mem media (GIBCO) for four hours, then cells were washed with 1x PBS, and placed in normal DMEM growth media. Pentamidine was added at this point, if being assayed.

After 20-24 hours, cells were washed with 1x PBS and harvested using triplE reagent (GIBCO). Following harvesting, cells were either stored temporarily at -80 °C, or RNA was immediately isolated using an RNeasy kit (QIAGEN). Isolated RNA (500 ng) was DNased with RQI DNase (Promega) according to manufacture's protocol. DNased RNA (100 ng) was reverse transcribed with Superscript II and a gene specific reverse primer according to manufacturer's protocols. DNA from the RT reaction (30 ng) was subjected to 22-25 rounds of PCR amplification using specific primers spiked with a kinased forward primer. The linear range for PCR was determined for the cTNT and the IR construct and found to be between 20-26 cycles. The linear range for PCR for PLEKH2 was determined to be 25-27 cycles. The resulting PCR products were run on a 6% (19:1) poly-acrylamide denaturing gel at 6W for 2 hours. The gel was subsequently autoradiographed and quantitation of the radioactive bands was performed using ImageQuant software.

For testing of the endogenously expressed genes, cells were transfected with an empty pcDNA3 vector, with or without the construct containing the CUG expansions. A total of 1 μg of plasmid was transfected. Cells were then treated with or without pentamidine. Cells and RNA were harvested as normal. All the isolated RNA was DNased, using RQ1 DNase. Following DNase, RNA was ethanol precipitated and re-suspended in approximately 10-20 μL of water. RNA was standardized to 300 ng/μL and cDNA was made. Both random hexamers and oligoDT were used for the RT reaction. Random hexamers and oligoDT were used at a concentration of 95 ng and 600ng per 3,600 ng of RNA, respectively. For the RT reaction, DNased RNA, primers and 0.8 mM dNTPs were incubated at 70 °C for 5 minutes, then placed on ice for 5 minutes. RNA was then reverse transcribed with Super Script II, according to the manufacture's protocol. DNA (600 ng of RNA, or 2μL) was then used as a template for PCR, with gene specific primers, with a T_m of 55 °C, and 30 cycles. Samples were then run on a 2-3% agarose gel and imaged with ethidium bromide. ImageQuant was used to quantify the gel.

The following primer pairs were used. For the cTNT minigene, the forward primer was 5'GTT CAC AAC CAT CTA AAG CAA GAT G, and the reverse primer was 5'GTT GCA TGG CTG GTG CAG G. For the IR minigene, the forward primer was 5' GTA CAA GCT TGA ATG CTG CTC CTG TCC AAG ACA G, and the reverse primer was 5' GCC CTC GAG CGT GGG

CAC GCT GGT C. The underlined sections are for cloning purposes. For the PLEKHH2 minigene, the forward primer was 5' CGG GGT ACC AAA TGC TGC AGT TGA CTC TCC, and the reverse primer was 5' CCG CTC GAG CCA TTC ATG AAG TGC ACA GG. For the endogenous genes, the following primer pairs were used. CAMKK2 forward primer was 5' CCTGGTGAAGACCATGATACG, and the reverse primer was 5' GGCCCAGCAACTTTCCAC, with PCR products sizes 247 and 204 bases (for inclusion of exclusion of alternative exon). For DLG1, the forward primer was 5' AGCCCGATTAAAAACAGTGAAA, and the reverse primer was 5' CGTATTCTTCTTGACCACGGTA, with possible PCR product sizes of 209, 179 and 143 bases (with the largest product size being denoted as inclusion, and the smallest as exclusion). For TTC8 the forward primer was 5' AGCTATTTTAGGCGCAGGAAGT, and the reverse primer was 5' TTTTCATCCAGCATCATTTCTG, with PCR products of 209 and 179 bases. For VTI1, the forward primer was 5' GATCGCCTACAGTGACGAAGTA, and the reverse primer was 5' TCCACTGCTATTTGGTATCCAG, with PCR products sizes of 168 and 147 bases.

Microscopy. HeLa or HEK294 cells were plated in 6 well plates onto coverslips. Cells were transfected with the normal protocol. For transfection, 500 ng of CUG_{960} plasmid were transfected for each experiment. A total of 1 μg of plasmid were transfected, with the additional amount being empty pcDNA3 vector. After transfection, pentamidine was added and the cells were fixed 16 hours later.

Cells were fixed for 15 min at RT with 4% PFA and washed 5 times for 10 min in 1x PBS at RT. Cells were stored in 4° C, if not probed immediately. For the FISH (fluorescence *in situ* hybridization) procedure, cells were permeabilized with 0.5% triton x-100, in 1x PBS at RT for 5 min. Cells were pre-washed with 30% formamide, 2x SSC for 10 min at RT. Cells were then probed for 2 hours at 37 °C, with a 1 ng/μL of Cy3 CAG_{10} probe (IDT, IA) in 30% formamide, 2x SSC, 2μg/mL BSA, 66 μg/mL yeast tRNA, 2mM vanadyl complex. Cells were then washed for 30 min in 30% formamide, 2x SSC at 42° C, then with 1x SSC for 30 min at RT. Cells were next washed twice in 1x PBS, 10 min at RT, and then probed overnight, at 4 ° C with anti-MBNL1 antibody (1:5000 dilution, A2764 antibody) in 1x PBS. Cells were washed two times, for 10 mins at RT with 1x PBS. Next, cells were incubated with goat anti-rabbit Alexa 488 (1:500 dilution) for two hours at RT. Cells were washed two times, for 10 mins at RT with 1x PBS and then mounted onto glass slides using hardest mounting media that contains DAPI

(Vectashield, Peterborough England). Cells were imaged on an Olympus Fluoview FV1000 with a Bx61 scope (Center Valley, PA). The number of cells scored was 254 for the CUG_{960} transfection, and 193 cells were scored when cells were treated with pentamidine (75 μM). A chi test with two degrees of freedom was performed using Excel (Microsoft).

Treatment in mice. Mouse handling and experimental procedures were conducted in accordance with the Association for Assessment and Accreditation of Laboratory Animal Care. HSA^{LR} transgenic mice in line 20b were previously described (*45*). These mice express human skeletal actin mRNA with ~250 CUG repeats in the 3′ UTR. Age- and gender-matched HSA^{LR} mice were injected intraperitoneally with pentamidine or saline alone. For a low-dosage regimen, 25 mg/kg of pentamidine were administrated twice daily for 5 days. For a higher-dosage regimen, 40 mg/kg of pentamidine was injected once daily for 7 days. Mice were sacrificed one day after the last injection, and vastus muscle was obtained for splicing analysis. RNA extraction and cDNA preparation were performed as described previously (*55*). PCR amplification was carried out for 22–24 cycles with the following primer pairs:
Clc1 forward: 5′-TGAAGGAATACCTCACACTCAAGG-3′,
and reverse: 5′-CACGGAACACAAAGGCACTG-3′.
Serca1 forward: 5’-GCTCATGGTCCTCAAGATCTCAC-3’,
and reverse: 5’-GGGTCAGTGCCTCAGCTTTG-3’. The PCR products were analyzed on agarose gels and scanned with a laser fluorimager (Typhoon, GE Healthcare). Differences between two groups were evaluated by the Mann-Whitney U-test.

CHAPTER V

CONCLUSIONS

The RNA binding specificity of MBNL1 for CUG and CCUG repeats

Prior to my dissertation research, very little was known of MBNL1's binding specificity. It was clear that MBNL1 co-localized with CUG and CCUG repeats, and could be pulled down from cellular extract by CUG repeats. Cross-linking data also indicated that it might directly bind to the cTNT pre-mRNA to directly regulate pre-mRNA splicing. However, it was unclear what MBNL1's RNA binding specificity and affinity was for various targets, and how strongly it bound the CUG/CCUG expansions in comparison to any endogenous pre-mRNAs that it might directly regulate.

One of the first steps that I made towards dissecting MBNL1's binding specificity was the mutational study of CUG repeats. Analysis of RNA substrates in which the CUG repeat tract sequences are modified in various ways (Chapter II, Figure 2) clearly shows the preference of MBNL1 for pyrimidine-pyrimidine mismatches as well as the Watson-Crick base pairs in their particular positions. The lack of significantly higher affinity binding to longer CUG repeats suggests that MBNL1 does not bind in a highly cooperative manner to CUG repeats. These results suggest this version of MBNL1 recognizes the many binding sites on long CUG repeats as independent binding sites. Both the specific recognition and structural distortion of CUG repeats are likely playing a role in MBNL1 binding. The distortion of CUG repeats is supported by the decrease in CD signal and the peak shift upon MBNL1 binding to the CUG repeats and the cTNT 50mer (Chapter II, Figure 4).

I next found that MBNL1 binds CCUG repeats with approximately 2-fold stronger affinity in both structural conformations (Chapter II, Figure 2-3) compared to the CUG repeats. These results are surprising because patients with DM2 tend to have more CCUG repeats compared to DM1 patients with CUG repeats (*75*). Additionally, the *ZNF9* pre-mRNA containing the CCUG repeats appears to be expressed at similar or even higher levels than the *DMPK* pre-

mRNA (*50, 76*). Yet, the symptoms of DM2 patients are less serious than those of DM1 patients (*77*). This supports the model that the expanded CUG repeats, unlike the CCUG repeats, are affecting transcription or other processes in the cell and create another layer of mis-regulation in DM1 compared to DM2 (*21*). Further studies are required to understand how DM1 and DM2 differ, as my results suggest it is not as simple as just the sequestration of MBNL1 to the RNA repeat expansions.

The RNA binding specificity of MBNL1 for the cTNT pre-mRNA

My observation that MBNL1 binds short CUG and CCUG stem-loops prompted me to consider that MBNL1 might recognize a short stem-loop within the cTNT pre-mRNAs as well. I chose to study the cTNT intron 4 because it was the only available pre-mRNA substrate with a putative binding site (*28*). The combination of UV melting of the cTNT 50mer RNA (Chapter II, Figure 4E), circular dichroism information in the absence and presence of MBNL1 (Chapter II, Figure 4F), structure probing of the cTNT 32mer (Chapter II, Figure 6A-B) and fluorescence studies (Chapter III, Figure 6C) show that MBNL1 binds this RNA in a stem-loop structure which appears to be partially A-form. This proposed helix has some similarities to the CUG and CCUG helices in that one of the pyrimidine-pyrimidine mismatches is bracketed by G-C base pairs (though the other mismatch is flanked by G-U wobble base pairs). The similarities between the CUG, CCUG and cTNT stems suggest MBNL1 recognizes both its pathogenic and natural RNA targets through an analogous mode of recognition. However, the differences between the RNAs suggest that MBNL1 may have complex binding specificity.

Clearly MBNL1 can bind a range of RNA stems, but the common theme of pyrimidine-pyrimidine mismatches and presence of G-C and C-G base pairs indicates these are requirements for binding by MBNL1. Further studies are necessary to fully define the specificity of MBNL1 before predictions can be made to identify binding sites in other MBNL1 regulated pre-mRNAs. Another challenge is that these potential regulatory stem-loops will not necessarily be predicted by folding programs due to the lack of consecutive base pairs and presence of mismatches as observed for the cTNT intron 4 stem-loop.

MBNL1's role as a splicing regulator for the cTNT pre-mRNA

At the beginning of my dissertation work, I hypothesized that a possible mechanism through which MBNL1 might regulate the exclusion of exon 5 in the cTNT pre-mRNA is that

MBNL1 could compete for binding of the intron with other splicing factors. My initial findings supported this model, as mutations made to the cTNT splicing minigene showed that recruitment of MBNL1 to a stem-loop with CUG repeats (instead of the endogenous sequence) also causes repression of exon 5. The replacement of the endogenous binding site with another sequence that MBNL1 binds causes MBNL1 dependent repression of exon 5, while minor mutations that significantly reduce MBNL1's binding abolishes MBNL1 ability to repress inclusion of exon 5. These results indicate that direct binding of MBNL1 to this stem-loop is required for the repression of exon 5.

Analysis of the sequences adjacent to MBNL1's binding site suggested that U2AF65 binding might be affected by MBNL1. I found that MBNL1 and U2AF65 do indeed directly compete for binding and that this competition inhibits U2AF65 from recruiting the U2 snRNP to the branch-point (Chapter III). The competition of MBNL1 with U2AF65 does not itself constitute a novel mechanism, as other alternative splicing factors have previously been shown to compete directly with U2AF65. The alternative splicing factor Sex Lethal (Sxl) has been shown to compete with U2AF65 in the *tra* and *msl-2* pre-mRNAs (*88, 89*), while the poly-pyrimidine Tract Binding Protein (PTB) has been observed to compete with U2AF65 for sequences within the α- or β-tropomyosin pre-mRNAs (*94, 95*). Sxl and PTB both have binding specificities similar to that of U2AF65, as all three proteins primarily prefer long runs of single-stranded uracil residues (*94*). When competitively binding with U2AF65, both Sxl and PTB are considered to occlude the U2AF65 binding site, and sterically inhibit U2AF65 from accessing its binding site. On the other hand, MBNL1 has a different binding specificity from U2AF65, and binds adjacent sequences outside of the py-tract. I cannot rule out steric contributions to the competition between MBNL1 and U2AF65, but the fluorescence assay strongly suggests that these proteins bind this RNA in different conformations (Chapter III, Figure 6). Both the differential binding sites of MBNL1 and U2AF65, and the role that RNA secondary structure can play in modulating U2AF65 binding, suggest that these splicing factors compete for recognition of the 3' end of intron 4 largely through binding mutually exclusive RNA structures.

Like other factors regulating splicing, MBNL1 functions in one context to exclude an exon as described for the cTNT exon 5, while for other pre-mRNA the presence of MBNL1 enhances the inclusion of a particular exon. One general mechanism for MBNL1's regulation of alternative splicing could be that the location of MBNL1 binding dictates whether it represses or enhances inclusion of an exon. One simply model would be that MBNL1 represses exon

inclusion if it binds upstream of the exon, or enhances exon inclusion if it binds downstream of the exon. This has been found for other splicing factors, such as NOVA (*82*). Supporting this model for MBNL1, I found that MBNL1 represses cTNT exon 5 and bound directly upstream of the exon 5 (Chapter II and III), and Yuan and colleagues found that MBNL1 represses exon F in the Tnnt3 pre-mRNA and binds directly upstream of that exon as well (*74*). Furthermore, MBNL1 enhances inclusion of exon 22 of the Serca1 pre-mRNA and was found to cross-link downstream of the exon (*85*). While there are only three identified binding sites, they are consistent with the model. More binding sites are required to fully articulate the affect of MBNL1's binding site location on its regulatory function.

Beyond the location of the binding, the general mechanisms by which MBNL1 either represses or enhances inclusion of an exon are relatively unclear. When binding upstream of exon 5 in the cTNT pre-mRNA, I found that MBNL1 competed with the splicing factor U2AF65 to repress exon 5 (Chapter II and III). While the mechanism for repression of exon F in the Tnnt3 pre-mRNA is unknown, MBNL1's binding site is also adjacent to a potential U2AF65 binding site in that intron (*74*). This is highly suggestive that MBNL1 may repress inclusion of exon F in that pre-mRNA in a similar mechanism as it does in the cTNT pre-mRNA. As all the protein and protein-RNA complexes that recognize the 3' end of introns are thought to bind the intron in a single-stranded fashion, it is possible that MBNL1 may repress exon inclusion through a general mechanism of sequestering important sequences in stem-loops. Further binding sites need to be articulated and mechanistically studied to see if this is a general mechanism by which MBNL1 represses exons.

Less data is available to propose a mechanism through which MBNL1 enhances inclusion of exons, as only one cross-linking site has been determined in the Serca1 pre-mRNA. As MBNL1 can inhibit recognition of U2AF65 by sequestering its binding site in a loop in the cTNT pre-mRNA, a potential mechanism through which MBNL1 could enhance exon inclusion would be to sequester exonic or intronic splicing silencers in a stem-loop. This would be an elegant model, in that MBNL1 would always regulate splicing by sequestering binding sites of other splicing regulators. However, MBNL1 may have a completely different mechanism for enhancing exonic inclusion. Another possibility is that MBNL1 could recruit other proteins to the 5' splice site, which was recently shown for the splicing regulator CUG-BP2 (*146*).

The novel role of MBNL1 in regulating alternative splicing by modulating RNA secondary structure

To our knowledge, the competition between MBNL1 and U2AF65 represents the first example of an alternative splicing factor that regulates splicing primarily through modulation of an RNA structural element and not through direct occlusion of another splicing factor's binding site. It is well documented that RNA structural elements alone can regulate alternative splicing. For instance, stable structures have been shown to inhibit U1 snRNP binding to the 5'-splice site following exon 7 of both the *SMN1* and *SMN2* pre-mRNAs (*96*). Other structural elements that encompass exon 6B in chicken β-tropomysin have been shown to inhibit binding of all the U snRNPs, promoting the skipping of that exon (*97*). At present, it is unclear whether these examples involve alternative splicing factors that regulate these RNA structures, as MBNL1 appears to regulate secondary structures within the cTNT pre-mRNA.

Other alternative splicing factors have indeed been postulated to regulate RNA structural elements, but there is no clear evidence to show that these factors actually modulate RNA structure in a way that directly regulates alternative splicing. For example, the splicing factor hnRNP A1 has been shown to inhibit binding of splicing factors ASF/SF2 and SC35 when they regulate splicing of the *tat* pre-mRNA from HIV-1 (*98*). Binding of hnRNP A1 is thought to alter the secondary structure of the pre-mRNA in a way that affects splicing (*99*), but this has not been demonstrated. Similarly, ribosomal protein L32 from *S. cerevisiae* binds a structured RNA element near the 5'-splice site of an intron within its own pre-mRNA, leading to intron retention (*100*). However, this structured binding site allows for recruitment of the U1 snRNP, showing that RNA structures do not always inhibit binding of the U snRNPs (*101*). It is still unclear how L32 causes intron retention if it does not inhibit binding of U1. In regards to the competition between MBNL1 and U2AF65, it has been demonstrated with a crystal structure that U2AF65 binds the py-tract in a single-stranded structure (*102*). In this structure, seven uridine residues bind along two RNA recognition motifs of U2AF65, leaving the 3' and 5' ends spatially far from each other. This crystal structure sheds light on why U2AF65 must bind the cTNT intron in a single-stranded form, as it is unlikely that the stem could form with the loop in such an extended structure.

At present, I have shown that MBNL1 is the first alternative splicing factor to be identified that regulates alternative splicing on the level of RNA structure. It is likely that other

examples will follow as it is becoming increasingly apparent that RNA structure plays important roles in controlling pre-mRNA splicing.

The small molecule pentamidine as a potential therapeutic for Myotonic Dystrophy

Many small molecules have been shown to bind and affect nucleic acid structure and function, which were reviewed in the Discussion of Chapter IV. My results indicate that small molecules can also be used to target CUG repeats and alleviate MBNL1 sequestration. The small molecule pentamidine shows strong promise as a lead compound as a therapeutic for DM as it was effective altering MBNL1 sequestration in an *in vitro* assay and three *in vivo* based assays. My results suggest pentamidine can, at least, partially target CUG repeats and release MBNL1. As pentamidine is known to bind B-form DNA in the minor groove (*127*), it is interesting to note that Mooers and colleagues previously showed that CUG repeats form an A-form structure, despite the U-U mismatches (*40*). Although the minor grooves of A-form and B-form nucleic acid differ in architecture it is tempting to speculate that pentamidine binds the minor groove of the A-form CUG repeats.

The results with pentamidine are in contrast to neomycin B which also disrupted the MBNL1-CUG repeat interaction *in vitro* (Chapter IV, Figure 1), but did not rescue mis-splicing of any of the tested targets (Chapter IV, Figure 4). Both pentamidine and neomycin B are known to bind many different nucleic targets, but only pentamidine displayed the ability to reverse the mis-splicing of four different pre-mRNAs affected by CUG repeats. These results suggest that pentamidine has sufficient specificity for CUG repeats to have an effect in these assays, specificity which neomycin B apparently lacks.

The window of efficacy for pentamidine treatment in the DM1 cell culture and the mouse model is narrow. Substantial cell death is observed in HeLa cells with pentamidine treatments of 100 to 125 μM. Raising the concentration of pentamidine for the mice treatments is also problematic because pentamidine dosages of 30 mg/kg twice a day are lethal to the mice, while 40 mg/kg twice once a day only partially relieves mis-splicing of the two tested pre-mRNAs (Chapter IV, Figure 6). To overcome this toxicity, modifications to pentamidine will be performed to enhance the affinity and specificity of pentamidine for CUG repeats, thus allowing for a lower dosage treatment. Possible modifications are reviewed in the Discussion of Chapter IV.

Previous potential therapeutic approaches for DM1 have ranged from over-expression of MBNL1, RNA interference against the CUG repeats, to targeted degradation of the mutant DMPK transcript with an RNA ribozyme (*34-37*). We have now demonstrated that releasing MBNL1 from the CUG repeats with a small molecule is another valid approach that should be considered for treating DM1.

BIBLIOGRAPHY

1. J. P. Staley, C. Guthrie, *Cell* **92**, 315 (Feb 6, 1998).
2. M. C. Wahl, C. L. Will, R. Luhrmann, *Cell* **136**, 701 (Feb 20, 2009).
3. R. Reed, *Proc Natl Acad Sci U S A* **87**, 8031 (Oct, 1990).
4. R. Reed, J. Griffith, T. Maniatis, *Cell* **53**, 949 (Jun 17, 1988).
5. R. Das, R. Reed, *RNA* **5**, 1504 (Nov, 1999).
6. R. Reed, *Genes Dev* **3**, 2113 (Dec, 1989).
7. J. Valcarcel, R. K. Gaur, R. Singh, M. R. Green, *Science* **273**, 1706 (Sep 20, 1996).
8. B. Ruskin, P. D. Zamore, M. R. Green, *Cell* **52**, 207 (Jan 29, 1988).
9. M. M. Konarska, P. J. Grabowski, R. A. Padgett, P. A. Sharp, *Nature* **313**, 552 (Feb 14-20, 1985).
10. M. M. Konarska, P. A. Sharp, *Cell* **46**, 845 (Sep 12, 1986).
11. *The RNA World*. E. R. G. a. J. Atkins, Ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1993), pp.
12. P. D. Zamore, J. G. Patton, M. R. Green, *Nature* **355**, 609 (Feb 13, 1992).
13. P. D. Zamore, M. R. Green, *Proc Natl Acad Sci U S A* **86**, 9243 (Dec, 1989).
14. C. W. Smith, J. Valcarcel, *Trends Biochem Sci* **25**, 381 (Aug, 2000).
15. J. M. Johnson *et al.*, *Science* **302**, 2141 (Dec 19, 2003).
16. Q. Pan, O. Shai, L. J. Lee, B. J. Frey, B. J. Blencowe, *Nat Genet* **40**, 1413 (Dec, 2008).
17. K. J. Hertel, *J Biol Chem* **283**, 1211 (Jan 18, 2008).

18. M. A. Garcia-Blanco, A. P. Baraniak, E. L. Lasda, *Nat Biotechnol* **22**, 535 (May, 2004).

19. J. F. Caceres, A. R. Kornblihtt, *Trends Genet* **18**, 186 (Apr, 2002).

20. M. Busslinger, N. Moschonas, R. A. Flavell, *Cell* **27**, 289 (Dec, 1981).

21. D. H. Cho, S. J. Tapscott, *Biochim Biophys Acta* **1772**, 195 (Feb, 2007).

22. L. P. Ranum, T. A. Cooper, *Annu Rev Neurosci* **29**, 259 (2006).

23. M. Fardaei *et al.*, *Hum Mol Genet* **11**, 805 (Apr 1, 2002).

24. A. Mankodi *et al.*, *Mol Cell* **10**, 35 (Jul, 2002).

25. R. S. Savkur, A. V. Philips, T. A. Cooper, *Nat Genet* **29**, 40 (Sep, 2001).

26. W. Dansithong, S. Paul, L. Comai, S. Reddy, *J Biol Chem* **280**, 5773 (Feb 18, 2005).

27. R. J. Osborne, C. A. Thornton, *Hum Mol Genet* **15 Spec No 2**, R162 (Oct 15, 2006).

28. T. H. Ho *et al.*, *EMBO J* **23**, 3103 (Aug 4, 2004).

29. J. W. Day, L. P. Ranum, *Neuromuscul Disord* **15**, 5 (Jan, 2005).

30. S. V. Perry, *J Muscle Res Cell Motil* **19**, 575 (Aug, 1998).

31. A. V. Gomes, G. Guzman, J. Zhao, J. D. Potter, *J Biol Chem* **277**, 35341 (Sep 20, 2002).

32. B. J. Biesiadecki, B. D. Elder, Z. B. Yu, J. P. Jin, *J Biol Chem* **277**, 50275 (Dec 27, 2002).

33. T. A. Cooper, C. P. Ordahl, *J Biol Chem* **260**, 11140 (Sep 15, 1985).

34. R. N. Kanadia *et al.*, *Proc Natl Acad Sci U S A* **103**, 11748 (Aug 1, 2006).

35. L. A. Phylactou, C. Darrah, M. J. Wood, *Nat Genet* **18**, 378 (Apr, 1998).

36. M. A. Langlois *et al.*, *J Biol Chem* **280**, 16949 (Apr 29, 2005).

37. M. A. Langlois, N. S. Lee, J. J. Rossi, J. Puymirat, *Mol Ther* **7**, 670 (May, 2003).

38. B. Tian *et al.*, *RNA* **6**, 79 (Jan, 2000).

39. S. Michalowski *et al.*, *Nucleic Acids Res* **27**, 3534 (Sep 1, 1999).

40. B. H. Mooers, J. S. Logue, J. A. Berglund, *Proc Natl Acad Sci U S A* **102**, 16626 (Nov 15, 2005).

41. M. Napierala, W. J. Krzyzosiak, *J Biol Chem* **272**, 31079 (Dec 5, 1997).

42. K. Sobczak, M. de Mezer, G. Michlewski, J. Krol, W. J. Krzyzosiak, *Nucleic Acids Res* **31**, 5469 (Oct 1, 2003).

43. R. Dere, M. Napierala, L. P. Ranum, R. D. Wells, *J Biol Chem* **279**, 41715 (Oct 1, 2004).

44. M. Zuker, *Nucleic Acids Res* **31**, 3406 (Jul 1, 2003).

45. A. Mankodi *et al.*, *Science* **289**, 1769 (Sep 8, 2000).

46. R. N. Kanadia *et al.*, *Science* **302**, 1978 (Dec 12, 2003).

47. G. Begemann *et al.*, *Development* **124**, 4321 (Nov, 1997).

48. M. Pascual, M. Vicente, L. Monferrer, R. Artero, *Differentiation* **74**, 65 (Mar, 2006).

49. T. H. Ho *et al.*, *J Cell Sci* **118**, 2923 (Jul 1, 2005).

50. R. N. Kanadia *et al.*, *Gene Expr Patterns* **3**, 459 (Aug, 2003).

51. J. W. Miller *et al.*, *EMBO J* **19**, 4439 (Sep 1, 2000).

52. A. Mankodi *et al.*, *Hum Mol Genet* **10**, 2165 (Sep 15, 2001).

53. A. Mankodi *et al.*, *Ann Neurol* **54**, 760 (Dec, 2003).

54. H. Jiang, A. Mankodi, M. S. Swanson, R. T. Moxley, C. A. Thornton, *Hum Mol Genet* **13**, 3079 (Dec 15, 2004).

55. X. Lin *et al.*, *Hum Mol Genet* **15**, 2087 (Jul 1, 2006).

56. S. Paul *et al.*, *EMBO J* **25**, 4271 (Sep 20, 2006).

57. Y. Adereth, V. Dammai, N. Kose, R. Li, T. Hsu, *Nat Cell Biol* **7**, 1240 (Dec, 2005).

58. R. M. Squillace, D. M. Chenault, E. H. Wang, *Dev Biol* **250**, 218 (Oct 1, 2002).

59. B. N. Charlet *et al.*, *Mol Cell* **10**, 45 (Jul, 2002).

60. A. V. Philips, L. T. Timchenko, T. A. Cooper, *Science* **280**, 737 (May 1, 1998).

61. T. H. Ho, D. Bundman, D. L. Armstrong, T. A. Cooper, *Hum Mol Genet* **14**, 1539 (Jun 1, 2005).

62. R. S. Savkur *et al.*, *Am J Hum Genet* **74**, 1309 (Jun, 2004).

63. A. N. Ladd, M. G. Stenberg, M. S. Swanson, T. A. Cooper, *Dev Dyn* **233**, 783 (Jul, 2005).

64. N. A. Timchenko, P. Iakova, Z. J. Cai, J. R. Smith, L. T. Timchenko, *Mol Cell Biol* **21**, 6927 (Oct, 2001).

65. Y. Kino *et al.*, *Hum Mol Genet* **13**, 495 (Mar 1, 2004).

66. J. Dam, P. Schuck, *Methods Enzymol* **384**, 185 (2004).

67. M. Molinaro, I. Tinoco, Jr., *Nucleic Acids Res* **23**, 3056 (Aug 11, 1995).

68. S. Ivanov, Y. Alekseev, J. R. Bertrand, C. Malvy, M. B. Gottikh, *Nucleic Acids Res* **31**, 4256 (Jul 15, 2003).

69. S. H. Hung, Q. Yu, D. M. Gray, R. L. Ratliff, *Nucleic Acids Res* **22**, 4326 (Oct 11, 1994).

70. T. M. Hall, *Curr Opin Struct Biol* **15**, 367 (Jun, 2005).

71. R. S. Brown, *Curr Opin Struct Biol* **15**, 94 (Feb, 2005).

72. S. D. Auweter, F. C. Oberstrass, F. H. Allain, *Nucleic Acids Res* **34**, 4943 (2006).

73. B. P. Hudson, M. A. Martinez-Yamout, H. J. Dyson, P. E. Wright, *Nat Struct Mol Biol* **11**, 257 (Mar, 2004).

74. Y. Yuan *et al.*, *Nucleic Acids Res* **35**, 5474 (2007).

75. C. L. Liquori *et al.*, *Science* **293**, 864 (Aug 3, 2001).

76. K. Shimizu, W. Chen, A. M. Ashique, R. Moroi, Y. P. Li, *Gene* **307**, 51 (Mar 27, 2003).

77. G. Meola, R. T. Moxley, 3rd, *J Neurol* **251**, 1173 (Oct, 2004).

78. D. Libri, A. Piseri, M. Y. Fiszman, *Science* **252**, 1842 (Jun 28, 1991).

79. A. Watakabe, K. Inoue, H. Sakamoto, Y. Shimura, *Nucleic Acids Res* **17**, 8159 (Oct 25, 1989).

80. R. Singh, J. Valcarcel, *Nat Struct Mol Biol* **12**, 645 (Aug, 2005).

81. B. J. Blencowe, J. A. Bowman, S. McCracken, E. Rosonina, *Biochem Cell Biol* **77**, 277 (1999).

82. J. Ule *et al.*, *Nature* **444**, 580 (Nov 30, 2006).

83. K. R. Nykamp, M. S. Swanson, *Prog Mol Subcell Biol* **35**, 57 (2004).

84. M. B. Warf, J. A. Berglund, *RNA* **13**, 2238 (Dec, 2007).

85. S. Hino *et al.*, *Hum Mol Genet* **16**, 2834 (Dec 1, 2007).

86. L. O. Penalva, L. Sanchez, *Microbiol Mol Biol Rev* **67**, 343 (Sep, 2003).

87. R. K. Gaur, J. Valcarcel, M. R. Green, *RNA* **1**, 407 (Jun, 1995).

88. L. Merendino, S. Guth, D. Bilbao, C. Martinez, J. Valcarcel, *Nature* **402**, 838 (Dec 16, 1999).

89. J. Valcarcel, R. Singh, P. D. Zamore, M. R. Green, *Nature* **362**, 171 (Mar 11, 1993).

90. K. Datta, N. P. Johnson, P. H. von Hippel, *J Mol Biol* **360**, 800 (Jul 21, 2006).

91. K. L. Henscheid, R. B. Voelker, J. A. Berglund, *Biochemistry* **47**, 449 (Jan 8, 2008).

92. M. Teplova, D. J. Patel, *Nat Struct Mol Biol* **15**, 1343 (Dec, 2008).

93. H. Y. Yi-Brunozzi, O. M. Stephens, P. A. Beal, *J Biol Chem* **276**, 37827 (Oct 12, 2001).

94. R. Singh, J. Valcarcel, M. R. Green, *Science* **268**, 1173 (May 26, 1995).

95. C. H. Lin, J. G. Patton, *RNA* **1**, 234 (May, 1995).

96. N. N. Singh, R. N. Singh, E. J. Androphy, *Nucleic Acids Res* **35**, 371 (2007).

97. P. Sirand-Pugnet, P. Durosay, B. C. Clouet d'Orval, E. Brody, J. Marie, *J Mol Biol* **251**, 591 (Sep 1, 1995).

98. J. Zhu, A. Mayeda, A. R. Krainer, *Mol Cell* **8**, 1351 (Dec, 2001).

99. C. K. Damgaard, T. O. Tange, J. Kjems, *RNA* **8**, 1401 (Nov, 2002).

100. F. J. Eng, J. R. Warner, *Cell* **65**, 797 (May 31, 1991).

101. J. Vilardell, J. R. Warner, *Genes Dev* **8**, 211 (Jan, 1994).

102. E. A. Sickmier *et al.*, *Mol Cell* **23**, 49 (Jul 7, 2006).

103. A. R. Krainer, T. Maniatis, B. Ruskin, M. R. Green, *Cell* **36**, 993 (Apr, 1984).

104. M. H. Hou, H. Robinson, Y. G. Gao, A. H. Wang, *Nucleic Acids Res* **30**, 4910 (Nov 15, 2002).

105. F. M. Chen, *Biochemistry* **37**, 3955 (Mar 17, 1998).

106. R. H. Griffey, S. A. Hofstadler, K. A. Sannes-Lowery, D. J. Ecker, S. T. Crooke, *Proc Natl Acad Sci U S A* **96**, 10129 (Aug 31, 1999).

107. T. Hermann, V. Tereshko, E. Skripkin, D. J. Patel, *Blood Cells Mol Dis* **38**, 193 (May-Jun, 2007).

108. Q. Vicens, E. Westhof, *J Mol Biol* **326**, 1175 (Feb 28, 2003).

109. S. Usha, I. M. Johnson, R. Malathi, *J Biochem Mol Biol* **38**, 198 (Mar 31, 2005).

110. B. Francois *et al.*, *Nucleic Acids Res* **33**, 5677 (2005).

111. R. Schroeder, C. Waldsich, H. Wank, *EMBO J* **19**, 1 (Jan 4, 2000).

112. C. Bailly, J. P. Henichart, P. Colson, C. Houssier, *J Mol Recognit* **5**, 155 (Dec, 1992).

113. C. Spickler, M. N. Brunelle, L. Brakier-Gingras, *J Mol Biol* **273**, 586 (Oct 31, 1997).

114. B. Epe, P. Woolley, H. Hornig, *FEBS Lett* **213**, 443 (Mar 23, 1987).

115. X. Zhang, F. L. Kiechle, *Biochem Biophys Res Commun* **248**, 18 (Jul 9, 1998).

116. M. D. Disney, J. L. Childs, D. H. Turner, *Chembiochem* **5**, 1647 (Dec 3, 2004).

117. T. E. Edwards, S. T. Sigurdsson, *Biochemistry* **41**, 14843 (Dec 17, 2002).

118. R. Sinha, M. Hossain, G. S. Kumar, *Biochim Biophys Acta* **1770**, 1636 (Dec, 2007).

119. R. Sinha *et al.*, *Bioorg Med Chem* **14**, 800 (Feb 1, 2006).

120. V. Peytou *et al.*, *J Med Chem* **42**, 4042 (Oct 7, 1999).

121. M. T. Makler, L. G. Lee, D. Recktenwald, *Cytometry* **8**, 568 (Nov, 1987).

122. L. D. Van Vliet *et al.*, *J Med Chem* **50**, 2326 (May 17, 2007).

123. X. Guo, Z. Liu, S. Liu, C. M. Bentzley, M. F. Bruist, *Anal Chem* **78**, 7259 (Oct 15, 2006).

124. Y. Zhang, Z. Li, D. S. Pilch, M. J. Leibowitz, *Nucleic Acids Res* **30**, 2961 (Jul 1, 2002).

125. P. A. Nguewa *et al.*, *Chem Biodivers* **2**, 1387 (Oct, 2005).

126. Y. Zhang, A. Bell, P. S. Perlman, M. J. Leibowitz, *RNA* **6**, 937 (Jul, 2000).

127. K. J. Edwards, T. C. Jenkins, S. Neidle, *Biochemistry* **31**, 7104 (Aug 11, 1992).

128. M. C. Lee, G. Knapp, *J Biol Chem* **260**, 3108 (Mar 10, 1985).

129. L. Frydman *et al.*, *Proc Natl Acad Sci U S A* **89**, 9186 (Oct 1, 1992).

130. A. Garcia, R. Giege, J. P. Behr, *Nucleic Acids Res* **18**, 89 (Jan 11, 1990).

131. Y. Nezu, Y. Kino, N. Sasagawa, I. Nishino, S. Ishiura, *Neuromuscul Disord* **17**, 306 (Apr, 2007).

132. M. B. Warf, J. V. Diegel, P. H. von Hippel, J. A. Berglund, *Proc Natl Acad Sci U S A* **106**, 9203 (Jun 9, 2009).

133. T. Kimura *et al.*, *Hum Mol Genet* **14**, 2189 (Aug 1, 2005).

134. F. Franceschi, E. M. Duffy, *Biochem Pharmacol* **71**, 1016 (Mar 30, 2006).

135. M. A. Fuertesa, J. Castillab, C. Alonsoa, J. M. Perez, *Curr Med Chem* **10**, 257 (Feb, 2003).

136. S. G. Chaney, S. L. Campbell, E. Bassett, Y. Wu, *Crit Rev Oncol Hematol* **53**, 3 (Jan, 2005).

137. C. F. Hsu *et al.*, *Tetrahedron* **63**, 6146 (Jul 2, 2007).

138. E. S. DeJong, B. Luy, J. P. Marino, *Curr Top Med Chem* **2**, 289 (Mar, 2002).

139. J. R. Thomas, P. J. Hergenrother, *Chem Rev* **108**, 1171 (Apr, 2008).

140. A. Pushechnikov *et al.*, *J Am Chem Soc* (Jun 24, 2009).

141. P. C. Gareiss *et al.*, *J Am Chem Soc* **130**, 16254 (Dec 3, 2008).

142. R. Paredes, F. Laguna, B. Clotet, *J Int Assoc Physicians AIDS Care* **3**, 22 (Jun, 1997).

143. G. J. Frayha, J. D. Smyth, J. G. Gobert, J. Savel, *Gen Pharmacol* **28**, 273 (Feb, 1997).

144. T. Sun, Y. Zhang, *Nucleic Acids Res* **36**, 1654 (Mar, 2008).

145. M. M. Lee, A. Pushechnikov, M. D. Disney, *ACS Chem Biol* **4**, 345 (May 15, 2009).

146. Y. H. Goo, T. A. Cooper, *Nucleic Acids Res* (May 14, 2009).