STRUCTURE-STABILIZING RNA MODIFICATIONS PREVENT MBNL BINDING TO TOXIC CUG AND CCUG REPEAT RNA IN MYOTONIC DYSTROPHY

by

ELAINE DELORIMIER

A DISSERTATION

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

June 2015

DISSERTATION APPROVAL PAGE

Student: Elaine deLorimier

Title: Structure-Stabilizing RNA Modifications Prevent MBNL Binding to Toxic CUG and CCUG Repeat RNA in 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 and Biochemistry by:

Dr. Victoria DeRose	Chair
Dr. J. Andrew Berglund	Advisor
Dr. Brad Nolen	Core Member
Dr. Diane Hawley	Core Member
Dr. Alice Barkan	Institutional Representative

and

Dr. Scott Pratt	Dean of the Graduate School

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

Degree awarded June 2015

DISSERTATION ABSTRACT

Elaine deLorimier

Doctor of Philosophy

Department of Chemistry and Biochemistry

June 2015

Title: Structure-Stabilizing RNA Modifications Prevent MBNL Binding to Toxic CUG and CCUG Repeat RNA in Myotonic Dystrophy

Myotonic dystrophy is a genetic neurodegenerative disease caused by repeat expansion mutations. Myotonic dystrophy type 1 (DM1) is caused by a CTG repeat expansion in the 3' UTR of the dystrophia myotonia protein kinase (*DMPK*) gene, while myotonic dystrophy type 2 (DM2) is caused by a CCTG repeat expansion in intron 1 of the zinc finger protein nine (*Znf9*) gene. When expressed, these genes produce long CUG/CCUG repeat RNAs that bind and sequester a family of RNA-binding proteins known as muscleblind-like 1, 2 and 3 (MBNL1, MBNL2, MBNL3). Sequestration of these proteins plays a prominent role in pathogenicity in myotonic dystrophy. MBNL proteins regulate alternative splicing, and myotonic dystrophy symptoms are a result of mis-spliced transcripts that MBNL proteins regulate. MBNL proteins bind to a consensus sequence YGCY (Y = pyrimidine), which is found in CUG and CCUG repeats, and cellular RNA substrates that MBNL proteins bind and regulate. CUG and CCUG repeats can form A-form helices, however it is hypothesized that MBNL proteins bind to the helices when they are open and the YGCY binding site is single-stranded in nature. To evaluate this hypothesis, we used structure-stabilizing RNA modifications pseudouridine

(Ψ) and 2'-O-methylation to determine if stabilization of CUG and CCUG repeat helices affected MBNL1 binding and toxicity. We also used Ψ to determine if the structure-stabilizing modification affected MBNL binding to single-stranded YGCY RNA. CUG repeats modified with Ψ or 2'-O-methyl groups exhibited enhanced structural stability and reduced affinity for MBNL1. Ψ also stabilized the structure of CCUG repeats and rigidified single-stranded YGCY RNA and inhibited MBNL1 binding to both of these RNAs. Binding data from CCUG repeats and single-stranded YGCY RNA suggest that both pyrimidines in the YGCY motif must be modified for significant inhibition. Molecular dynamics and X-ray crystallography suggest a potential water-bridging mechanism for Ψ-mediated CUG repeat stabilization. Molecular dynamics simulations suggest that Ψ increases base-stacking interactions, and reducing the flexibility of single-stranded RNA leads to reduced MBNL1 binding. Ψ modification rescued mis-splicing in a cellular DM1 model and prevented CUG repeat toxicity in zebrafish embryos.

This dissertation includes previously published and unpublished coauthored material.

CURRICULUM VITAE

NAME OF AUTHOR: Elaine deLorimier

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, OR
University of California at Santa Barbara, Santa Barbara, CA

DEGREES AWARDED:

Doctor of Philosophy, Chemistry and Biochemistry, 2015, University of Oregon
Bachelor of Science, Chemistry and Biochemistry, 2007, University of California at Santa Barbara

AREAS OF SPECIAL INTEREST:

Biochemistry
Molecular Biology
Gene Regulation

PROFESSIONAL EXPERIENCE:

Graduate Student, Dr. J. Andrew Berglund, 2009-2015

Chemical Analyst, Cranmer Engineering Inc., 2007-2009

Undergraduate Research Assistant, Dr. Martin Sagermann, 2006-2007

GRANTS, AWARDS AND HONORS:

American Heart Association Pre-doctoral Fellowship, *MBNL1 Interaction with Modified CUG Repeat RNA.,* 2013-2015

University of Oregon CMB NIH Training Grant Fellow, 2010-2013

Merck Index Award for Excellence in Organic Chemistry, 2007

PUBLICATIONS:

deLorimier E, Copperman J, Guenza MG, Berglund JA. 2015. Pseudouridine modification inhibits muscleblind-like 1 (MBNL1) binding to single-stranded RNA and CCUG repeats through reduced RNA flexibility. (submitted for publication, Journal of Biological Chemistry)

deLorimier E, Coonrod LA, Copperman J, Taber A, Reister EE, Sharma K, Todd PK, Guenza MG, Berglund JA. 2014. Modifications to toxic CUG RNAs induce structural stability, rescue mis-splicing in a myotonic dystrophy cell model and reduce toxicity in a myotonic dystrophy zebrafish model. *Nucleic Acids Research* **42**: 12768–12778. PMID: 25303993

Sagermann M, Chapleau RR, deLorimier E, Lei M. 2008. Using affinity chromatography to engineer and characterize pH-dependent protein switches. *Protein Science.* **18**: 217-228. PMID: 19177365

ACKNOWLEDGEMENTS

I would like to acknowledge my advisor, Dr. J. Andrew Berglund, for his advise and support throughout my work towards my doctoral degree. Andy consistently provided valuable advise in an amicable manner whenever I had a problem or needed guidance. I would also like to thank my committee members for their advice throughout the process. The amount of information and ideas that came out of every committee meeting astonished and impressed me. I left each committee meeting full of new ideas and interpretations for my project. I would also like to acknowledge my funding sources, first the NIH Training grant, and second the American Heart Association Pre-doctoral fellowship.

The members of the lab also deserve my sincerest thanks. Input from your co-workers is invaluable when working in a Biochemistry/Molecular Biology lab. In particular, I'd like to thank previous members of the lab who helped me most during the first few years of graduate school, when I needed the most guidance. These people helped me learn techniques, and provided valuable discussion on the particulars of the subjects we study. I also had a number of very valuable friends who helped me understand the process of obtaining a Ph.D. in our field from a perspective outside of my own lab. Not only did they provide support and advice from the perspective of more senior graduate students, they also provided a source of release when work was overwhelming, and more importantly, we had a lot of fun. Special thanks to Alex Taber, Michelle Lu and Marisa Connell.

I'd also like to acknowledge the staff in the Institute of Molecular Biology. Dan Graham does an amazing job at keeping all of the equipment working. I don't think the

institute would be able to run without him. I'd also like to thank the media kitchen staff for saving us all time washing our glassware and preparing media and other buffers. The rest of the staff of the institute were also very helpful with ordering, contracts and grant submission.

Finally, I'd like to thank my family for all of the support they provided. My older brother Michael finished his doctorate while I was working on my own. Looking up to him and observing how he went through his education has always inspired me, and it continues to do so to this day. The rest of my family provided invaluable emotional support, and asked very interesting questions regarding my project. It is very important to be able to explain the in-depth biochemical and molecular biological projects to people who have expertise in areas such as Art, Design, Library Science and Foreign Language and Culture. Also, thanks to my boyfriend Justin for all the love and laughs. Finally, thanks to the furry creatures in my life. They are such wonderful and entertaining company.

This dissertation is dedicated to Rory, a faithful and loving companion.

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

CHAPTER I

INTRODUCTION

Alternative splicing

The human genome consists of roughly 20,000 genes, while the human proteome is comprised of millions of different proteins[1,2]. The central dogma of molecular biology states that genomic DNA is transcribed into messenger RNA (mRNA), which is then translated into protein. If the process were this simple, each gene would encode for only a single protein, which, due to the difference in genome size versus proteome size, cannot be the case. RNA that is transcribed directly off of genomic DNA is called pre-mRNA, and contains many sequence elements that do not encode the protein sequence. Sequence elements such as introns, and 5' and 3' untranslated regions are removed from pre-mRNA, leaving mature mRNA that is made up of exons and other mRNA features. Diverse cellular mechanisms exist to modify pre-mRNA so that a number of different protein isoforms can be expressed off of a single gene. These include alternative pre-mRNA splicing, post-transcriptional modifications and alternative polyadenylation sites[2]. Intronic sequences are noncoding elements of pre-mRNA, while exonic sequences encode protein sequences. The spliceosome is the complex molecular machine that removes introns and ligates exons together through transesterification reactions, one of the process that is required for mature mRNA production[3]. Most exons in any certain gene are constitutively included in mature mRNA, while some exons are only included in the mRNA in certain tissues or developmental environments. Certain exons are either included or excluded in mRNA, and this process is known as alternative splicing. Alternative splicing and occurs in more than 90% of human transcripts[3,4]. The

Drosophila Dscam gene can express over 30,000 potential isoforms, but this is an extreme example[5].

There are a number of different types of alternative splicing, including cassette exon inclusion or exclusion, alternative 5' or 3' splice sites, mutually exclusive exons and intron retention (Figure 1). Cassette exon alternative splicing, in which exons are either

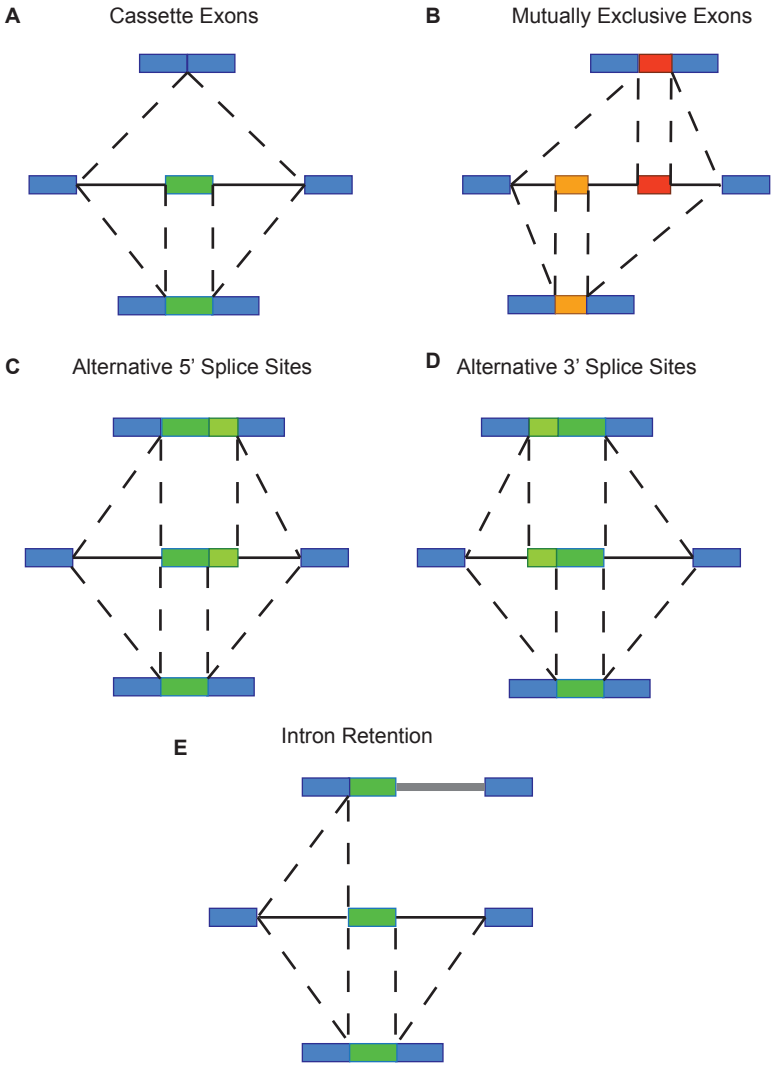


Figure 1: Types of alternative splicing. (A) A cassette exon event will either include or exclude a certain exon. (B) In the case of mutually exclusive exons, the inclusion of one exon prevents the inclusion of another. (C and D) The spliceosome can use alternative 5' or 3' splice sites to express different isoforms. (E) Sometimes intronic sequences are included in protein-coding mRNAs.

simply included or excluded, is the most common form of alternative splicing[6]. During pre-mRNA processing, the spliceosome must choose which exons to include or exclude based on regulatory signals and splice-site strength. Cis-acting elements in the RNA itself influence splice-site choice. Some examples of these are the strength of the 5' and 3' splice sites, the polypyrimidine tract, and the branch point sequence. Certain cis-acting elements are considered enhancers or suppressors, which either encourage or discourage the spliceosome to choose a splice-site. Point mutations in any of these sequences can result in mis-splicing, producing transcripts that could express nonfunctional proteins, or be prematurely decayed[7]. Trans-acting elements are also critically important for appropriate splice-site choice, which can differ depending on tissue type or developmental phase. These trans-acting elements are regulatory molecules, including RNA-binding proteins and non-coding RNAs and they influence exon inclusion or exclusion[8,9].

Because alternative splicing is so highly regulated by proteins and non-coding RNAs, errors in this process can be detrimental to cellular function. Mis-spliced mRNA could result in biologically inactive proteins, or cause nonsense mediated decay due to introduction of premature stop codons. Therefore, depletion or disruption of certain isoforms due to mis-regulated alternative splicing is the basis for many diseases[10]. Diseases can result from errors in cis-factors, as well as depletion or disruption of trans-acting splicing regulatory factors. Myotonic dystrophy is a disease that is largely attributed to disruption of proteins that regulate alternative splicing, resulting in the expression of mis-spliced isoforms[11].

Myotonic dystrophy

Myotonic dystrophy is a neurodegenerative disease that is the most common adult-onset form of muscular dystrophy. Myotonic dystrophy patients experience a wide variety of symptoms including cataracts, heart defects, insulin resistance, cognitive defects, muscle wasting, and myotonia. The namesake symptom of the disease, myotonia, is a condition in which muscles will contract, but have trouble relaxing. The disease is a genetic, autosomal dominant disorder that affects approximately 1 in 8000 people. There are two types of myotonic dystrophy. Myotonic dystrophy type one (DM1) is caused by a mutation that expands CTG repeats in the 3' UTR of the dystrophia myotonica protein kinase (*DMPK*) gene. Unaffected individuals have 5-38 CTG repeats in this gene, while DM1 patients have up to thousands of CTG repeats[12]. Myotonic dystrophy type two (DM2) is caused by an expanded CCTG repeat mutation in intron one of the zinc finger protein nine (*ZNF9*) gene. DM2 patients can have very long expanded CCTG repeats (~75 to 11,000), however DM2 patients tend to have more mild symptoms than DM1 patients. In both cases, the mutation is in a non-coding region of the affected gene, which results in the expression of long CUG or CCUG repeat RNAs in DM1 and DM2, respectively[13]. When these long non-coding RNAs are expressed, they aggregate and form distinct nuclear foci, and RNA-binding proteins associate with these foci[14].

Long CUG and CCUG repeats contain an abundance of YGCY motifs (Y = pyrimidine), which is the binding site for a family of RNA-binding proteins that regulate alternative splicing[15]. The muscleblind family of proteins (MBNL1, MBNL2 and MBNL3) are RNA-binding proteins that regulate alternative splicing of hundreds of transcripts[16,17]. These proteins have been shown to bind to CUG and CCUG repeats and

are sequestered to the nuclear foci[18-20]. When sequestered to nuclear foci, MBNL proteins are no longer able to regulate alternative splicing, resulting in a number of mis-spliced transcripts expressed in myotonic dystrophy patients. Furthermore, a number of symptoms in DM1 and DM2 can be attributed to the expression of fetal isoforms in adult tissue[11]. MBNL proteins are expressed in adult tissues, and myotonic dystrophy is commonly an adult-onset disease, thus the lack of regulation of alternative splicing in myotonic dystrophy patients results in fetal splicing patterns. Some specific examples of symptom-causing fetal isoforms expressed in myotonic dystrophy patients are the chloride ion channel (*Clcn1*), cardiac troponin T (*TNNT2*), the insulin receptor (*INSR*) and sarcoplasmic/endoplasmic reticulum Ca^{2+}-ATPase (*Atp2a1*). Myotonia can be caused by the lack of function of chloride ion channels in skeletal muscles, and MBNL proteins regulate alternative splicing of *Clcn1*. Patients with myotonic dystrophy have reduced levels of this protein in skeletal muscle, due to aberrant alternative splicing. Specifically, exon 7a of *Clcn1* is excluded when MBNL proteins are present, and inclusion of exon 7a results in expression of a premature stop codon, which leads to nonsense-mediated decay (Figure 2A)[21,22]. Myotonic dystrophy patients can also have heart defects, caused by cardiac muscle dysfunction. Heart defects are the second most common cause of death in myotonic dystrophy patients[23,24]. MBNL proteins also regulate alternative splicing of *TNNT2*, and this transcript is mis-spliced in myotonic dystrophy patients' cardiac muscles, resulting in the expression of the fetal isoform. The fetal isoform of *TNNT2* contains exon 5, while exon 5 is absent in the MBNL-regulated adult version (Figure 2B)[20,25]. Another transcript that is found to be mis-spliced in myotonic dystrophy patients is *INSR*, which results in a resistance to insulin. When MBNL proteins are present, they

promote inclusion of *INSR* exon 11, which allows for expression of a functioning insulin receptor and healthy glucose metabolism (Figure 2C). The isoform lacking exon 11 has a

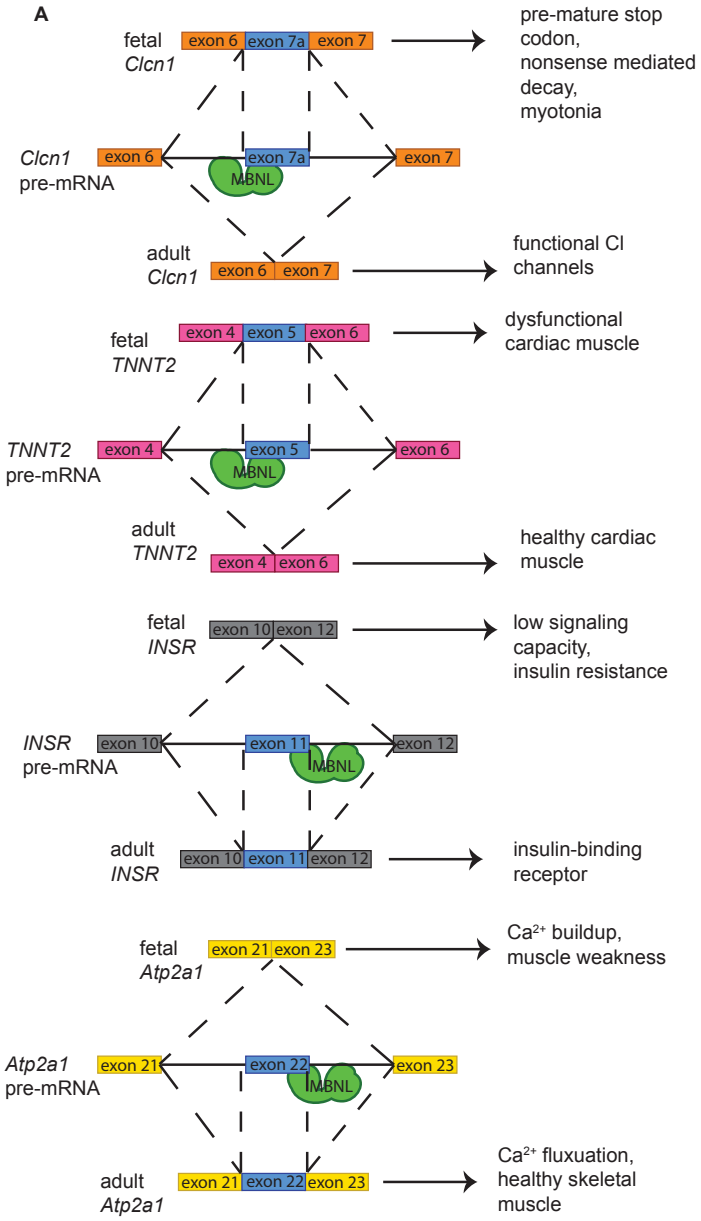


Figure 2: Examples of MBNL-regulated alternative splicing. (A) MBNL promotes exclusion of the *Clcn1* exon 7a. Inclusion of exon 7a results in a premature stop codon and nonsense mediated decay. (B) MBNL promotes exclusion of the *TNNT2* exon 5. Inclusion of exon 5 leads to cardiac muscle defects. (C) MBNL promotes inclusion of the *INSR* exon 11. Exclusion of exon 11 leads to insulin resistance. (D) MBNL promotes inclusion of the *Atp2a1* exon 22. Exclusion of exon 22 leads to Ca^{2+} buildup in skeletal muscle and muscle

lower signaling capacity, and an increase in expression of the receptor leads to insulin resistance[12,26,27]. Myotonic dystrophy patients also express higher levels of the fetal isoform of *Atp2a1*, which codes for a protein that regualtes Ca^{2+} fluxuation in adult muscle. MBNL proteins promote inclusion of exon 22, which is the isoform that is present at decreased levels in myotonic dystrophy patients (Figure 2D). Increased Ca^{2+} levels and muscle weakness in myotonic dystrophy patients has been attributed to this mis-splicing event[28,29]. All of these MBNL-regulated transcripts contain its consensus binding sequence (YGCY) and MBNL binds directly to them[15,30]. However, when MBNL proteins are bound and sequestered to CUG and CCUG repeats, they are unable to access their target transcripts and mis-regulation of the transcripts leads to myotonic dystrophy symptoms. It is therefore important to understand exactly how MBNL proteins interact with their YGCY binding site, in order to develop therapeutic strategies that would prevent MBNL binding to CUG and CCUG repeats.

RNA structure and MBNL1 interaction

Efforts to understand the molecular basis for myotonic dystrophy began shortly after CUG repeats and CCUG repeats were identified as the toxic molecules myotonic dystrophy patients. Specifically, studies were planned to determine the structure of CUG repeats in order to understand how they function as toxic molecules and provide a basis for structure-based drug design. An initial study used chemical and enzymatic probing to determine that CUG repeats form “slippery” hairpins[31]. This finding was further validated with NMR and crystal structures, establishing that CUG repeats form A-form helices with G-C base pairs and U-U mismatch base pairs (Figure 3A, C)[32,33]. The crystal structure of CUG repeats indicated that the U-U mismatches do not distort the backbone,

as other structures have been shown to do[34].

By comparing the conformation of U-U mismatches from a number of structures, it was determined that the U-U mismatches in CUG repeat helices are rather dynamic[33]. Because of the number of different possible U-U mismatches that form in CUG repeats, there is no established conformation of the U-U mismatch. A later study also found that CCUG repeats form A-form helices (Figure 3B)[35]. With the knowledge that CUG and CCUG repeats are helical, a model was developed in which MBNL associates with CUG and CCUG repeats by interacting with the major or minor groove of helical RNA. However, efforts to co-crystallize MBNL1 with CUG repeats in order to determine MBNL1's mode of binding to CUG repeats were unsuccessful, possibly due to unstructured regions of the protein.

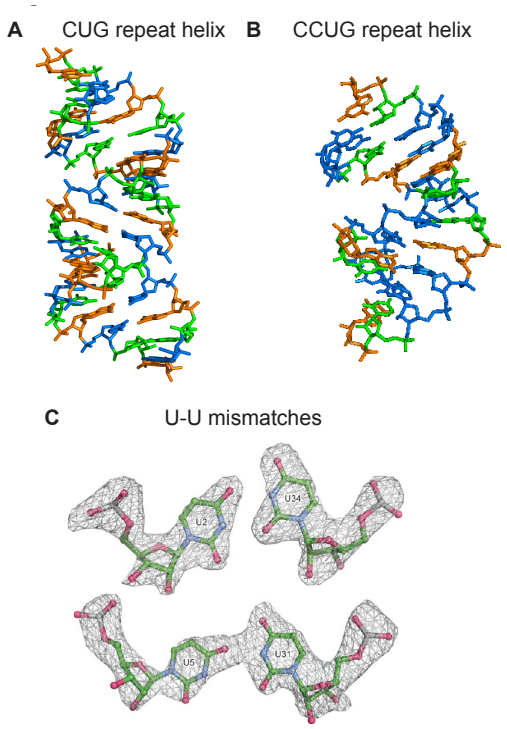


Figure 3: Structures of CUG and CCUG repeats and examples of U-U mismatches. (A) Crystal structure of CUG repeats in an A-form helix. (B) Crystal structure of CCUG repeats in an A-form helix. (C) Two examples of the conformations of U-U mismatch base pairs found in different crystal structures.

MBNL proteins contain four highly conserved zinc finger domains that fold into two distinct and similar motifs (ZnF1/2 and ZnF3/4)[36,37]. Teplova et al in 2008 co-crystallized the isolated ZnF3/4 with a short YGCY-containing RNA (CGCUGU). This structure refuted the previous hypotheses that MBNL proteins bind the outer faces of

CUG and CCUG repeat helices. The RNA associated with ZnF3/4 makes contacts with the protein via the Watson-Crick face of the GC dinucleotide (Figure 4)[38]. In helical CUG and CCUG repeats, the Watson-Crick face of the GC dinucleotide forms base pairs with the opposing strand, rendering it unable to interact with MBNL proteins. This discovery led to the development of a new hypothesis on how MBNL interacts with CUG and CCUG repeats. It is proposed that mismatch base pairs destabilize the helices, and they exist in equilibrium between helical RNA and an open, single-stranded state. Furthermore, we propose that when CUG and CCUG repeats are open and single-stranded, MBNL proteins bind to the accessible Watson-Crick face of their YGCY binding sites. Following this model, stabilization of CUG and CCUG repeat helices should prevent MBNL binding to the toxic RNA and allow it to function normally, relieving myotonic dystrophy symptoms.

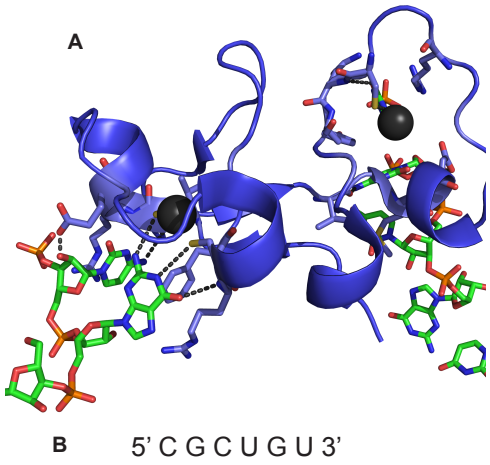


Figure 4: Structure of a zinc finger of MBNL interacting with YGCY RNA. (A) A depiction of MBNL1's ZnF3/4 interacting with YGCY RNA. The protein is shown as a purple cartoon, RNA is shown as sticks and zinc atoms are shown as black spheres. (B) The sequence of the RNA that co-crystallized with the zinc finger in this study.

RNA modification

Post-transcriptional RNA modifications are commonly employed in cells to alter the structure and function of RNA. Like alternative splicing and other post-transcriptional RNA processing events, RNA modifications can also function as regulatory elements[39]. There are more than 90 different types of RNA modifications that have various functions. These functions include stabilizing

secondary structure, acting as sensors for a cell's metabolic state, increasing translation efficiency and participating in other cellular processes[40,41]. A common and well-established function of RNA modifications is to stabilize the structure of non-coding functional RNAs such as tRNA, rRNA, snRNA and RNA components of the spliceosome[41]. Pseudouridine (Ψ) and 2'-O-methylation are common RNA modifications found in the RNAs mentioned above. Ψ is an isomerization product of uridine, with a C-C bond connecting the ribose to the base, and an extra N-H group at nucleoside position three (Figure 6A). Ψ modification in tRNA, rRNA and spliceosomal RNA is position-specific and highly conserved, suggesting an important functional role for this modification[42]. 2'-O-methylation is similarly conserved and has been shown to affect RNA folding to promote hairpin formation rather than duplexing, implying it is important for structural integrity[43]. 2'-O-methylation of the U2 snRNA is fundamental for spliceosome assembly[44]. 2'-O-methylation is also commonly found in small RNAs in order to protect against 3' to 5' degradation [45]. Both modifications have been shown to increase the thermal stability of RNA via different mechanisms[46-48]. Ψ, with its extra N-H group, has the capability of forming unique hydrogen bonds with neighboring atoms and it has been shown to form water bridges to induce stability[46]. Other studies have attributed an increase in structural stability with Ψ modifications to enhanced base-stacking interactions with neighboring bases[46,49]. 2'-O-methyl modifications appear to induce structural stability by pushing the ribose to maintain a 3'endo conformation. Specifically, the methyl group induces C3'endo conformation through steric repulsion between the 2-carbonyl group and the 2' methoxy group[48].

Based on the model for CUG and CCUG repeat toxicity, we envisioned that if

structure-stabilizing modifications are incorporated in these repeats, MBNL proteins will be less likely to bind to them and become sequestered. In chapters II and III, I address the structural effects Ψ and 2'-O-methylation have on CUG repeats, CCUG repeats and single-stranded YGCY RNA, and MBNL's ability to bind to modified RNA. Chapters II and III contain data that indicates that Ψ and 2'-O-methyl modifications do increase the structural stability of these RNAs and weaken MBNL1's binding affinity *in vitro*. Chapter II provides evidence that MBNL is not sequestered to modified CUG repeats *in vivo*, and splicing patterns are normal in cells transfected with modified CUG repeats. In order to move forward with this information to a therapeutic strategy, we must address how modifications are carried out *in vivo*.

Pseudouridylation is carried out *in vivo* by an enzyme that catalyzes the breaking of the bond between N1 and the ribose, rotation of the base and formation of a bond between C5 and the ribose. This enzyme is part of pseuouridine-synthase, which is an enzyme-RNA complex designed for site-specific modification, called an H/ACA box RNP. The RNA in these complexes is small nucleolar RNA (snoRNA), and acts as a guide to target the specific site for modification. H/ACA box RNAs contain binding sites for pseudouridine synthases (Figure 5)[50]. 2'-O-methylation is carried out by C/D box RNPs, which have binding sites for methylases instead of pseudouridine-synthases[51]. H/ACA box and C/D box RNPs are otherwise quite similar. Both are composed of two interrupted stem-loops separated by a protein-binding site. Within the stem loops are single-stranded target recognition sequences that base pair with the RNA targeted for modification. During the modification process, the RNPs assemble around the targeted RNA, and the enzymes carry out the modification. Previous studies have shown that it is

possible to make designer H/ACA box and C/D box RNAs by modifying the target sequence such that it recognizes a different RNA, presumably one under investigation[52,53]. Thus, it is possible to envision that an H/ACA box or C/D box RNA designed to target CUG or CCUG repeats could be used as a therapeutic for myotonic

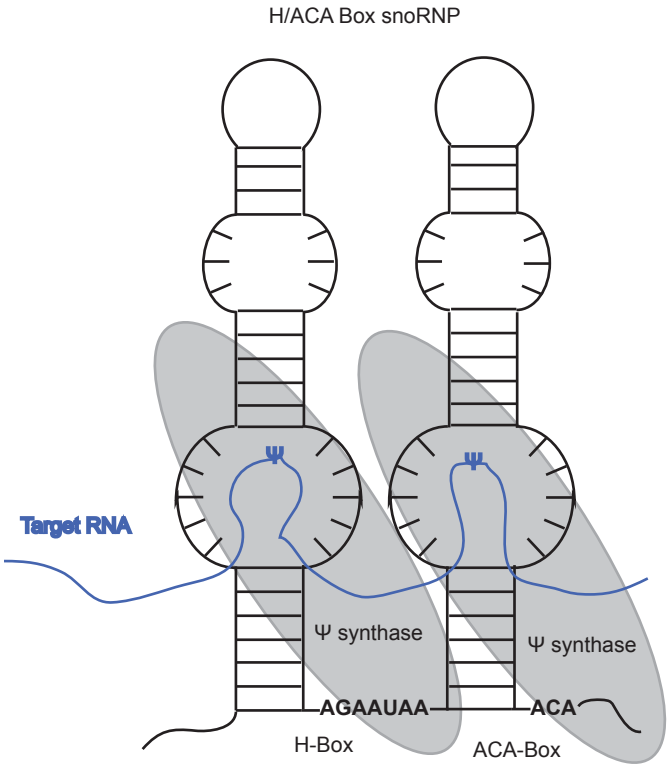


Figure 5: Schematic of H/ACA box snoRNP containing pseudouridine synthase. RNA targeted for pseudouridylation base-pair with targeting sequences in the stem loops of the snoRNA. Pseudouridine synthases bind to the H and ACA boxes and catalyze pseuouridylation of the target RNA.

dystrophy patients. RNAs introduced to patients would assemble with enzymes on the toxic repeats, which would then be modified. Modification would induce structural stability of the RNA, preventing MBNL binding, and allowing it to function normally.

This dissertation includes previously published and unpublished coauthored material.

CHAPTER II

MODIFICATIONS TO TOXIC CUG RNAS INDUCE STRUCTURAL STABILITY, RESCUE MIS-SPLICING IN A MYOTONIC DYSTROPHY CELL MODEL AND REDUCE TOXICITY IN A MYOTONIC DYSTROPHY ZEBRAFISH MODEL

This work was published in Volume 42, Issue 18 of the journal Nucleic Acids Research in October 2014. I was the primary author on this manuscript. I performed the thermal melt assays, the binding assays, the *in vitro* transcription and the *in vivo* splicing assays. Dr. Leslie Coonrod and Emily Reister performed the crystallography. Jeremy Copperman performed the molecular dynamics simulations. Dr. Peter Todd and Kush Sharma performed the zebrafish viability and spontaneous coiling exeriments. Dr. J. Andrew Berglund was the principle investigator of this work.

INTRODUCTION

Myotonic dystrophy is a genetic disorder that is the most common adult-onset form of muscular dystrophy, affecting ~1 in 8000 people. Myotonic dystrophy type one (DM1) is caused by a CTG repeat expansion mutation in the 3' non-coding region of the dystrophia myotonica protein kinase (*DMPK*) gene. Unaffected individuals have up to 40 CTG repeats at this location, while DM1 patients have >40 and up to thousands of CTG repeats[29,54]. Our current understanding of the DM1 disease mechanism is that the CTG repeats are transcribed into toxic CUG-repeat RNA, which sequesters RNA-binding proteins that are important for gene regulation[12,37].

Muscleblind-like proteins (MBNL1, MBNL2 and MBNL3) are the primary RNA-

binding proteins that are sequestered to the expanded CUG repeats[18]. MBNL proteins regulate alternative splicing and other RNA processing events[55]. During pre-mRNA splicing, the spliceosome must choose which exons to include based on developmental or environmental signals. This is a common regulatory event that vastly diversifies the proteome by allowing multiple protein isoforms to be expressed from a single gene. MBNL proteins function by interacting with their RNA binding sites within pre-mRNA downstream or upstream of alternatively spliced exons and direct the spliceosome to either include or exclude certain exons[15,56]. Many symptoms of DM1 are caused by expression of fetal isoforms due to mis-regulated alternative splicing[11]. For example, when MBNL genes are expressed in adults, they regulate the inclusion of the insulin receptor's (*INSR*) exon 11[27]. One symptom of DM1 is insulin resistance that is likely partly due to the mis-regulation of this transcript[11]. Another MBNL target that is mis-regulated in DM1 is cardiac troponin T (*TNNT2*). In this case MBNL promotes the exclusion of exon 5[57]. Accordingly, one of the leading causes of death in DM1 is cardiac dysfunction, in part possibly due to mis-splicing of *TNNT2*[16,23].

MBNL proteins contain four Zn fingers that fold into two similar domains (each domain contains two Zn fingers) that bind RNA on opposing faces of the domain[58,59]. The consensus MBNL binding site is YGCY (Y represents pyrimidines), which is found in its pre-mRNA targets, and several copies of this motif are usually necessary for splicing regulation by MBNL proteins[15,20]. The CUG repeats are toxic because they contain hundreds or thousands of YGCY motifs repeated in the expanded CUG repeats[60]. Presumably, the many MBNL proteins binding the expanded CUG repeats leads to formation of aggregated MBNL-CUG repeat foci in patient cells and model systems[61,62].

Structure-probing, crystal and NMR structures have shown that CUG repeats can form double stranded A-form helices[31-33,63,64]. A crystal structure by Teplova and colleagues showed how a single MBNL1 zinc finger domain interacts with an RNA containing a YGCY binding site primarily via the Watson-Crick face of the nucletoides[38]. This result suggested that MBNL proteins will not bind or bind weakly to CUG repeats when they are in double-stranded or helical conformation, as the Watson-Crick faces of the bases form hydrogen bonds with the opposite strand, leaving them unavailable for interaction with MBNL. The resulting prediction is that if CUG repeats are stabilized in a double-stranded or helical structure, these repeats will have reduced or no toxicity because MBNL proteins will not be sequestered.

Structure-stabilizing modifications are commonly found in RNAs such as tRNA, rRNA and spliceosomal snRNAs[42,65,66]. Pseudouridine (Ψ) is an isomerization product of uridine and is the most common RNA modification (Figure 6A). It is known to play many important roles in many cellular processes[42]. Ψ has been shown to stabilize RNA structural components via base-stacking interactions and water-bridging interactions[49,67]. Another common RNA modification found in many cellular RNAs is the replacement of the hydrogen on the 2'-hydroxyl group with a methyl group and is known as 2'-O-methylation (Figure 6B)[45]. Due to steric hindrance, 2'-O-methylation induces A-form helices in RNA and is also know to function as a stabilizing element in tRNA and rRNA[48].

Ψ and 2'-O-methyl modifications were introduced into CUG repeats to determine if these nucleotides stabilize the CUG repeats in a helical conformation. Both modifications do indeed increase the thermal stability of model CUG repeat RNAs and

reduce MBNL1's affinity for model CUG repeats. A crystal structure of CUG repeats with a single Ψ substitution showed a water molecule bridging the Ψ-U non-canonical base pair. Molecular dynamics simulations support the model that this water bridge reduces the dynamic nature of the Ψ-U pair in helical CUG repeats. HeLa cells treated with *in vitro*-transcribed CUG repeat RNA caused mis-splicing of MBNL1 targets.

Replacing uridine with Ψ in the CUG repeats rescued MBNL1-mediated splicing events. Furthermore, we found that incorporation of Ψ into a toxic CUG repeat RNA in a zebrafish model of DM1 ameliorated defects in motor function and enhanced viability.

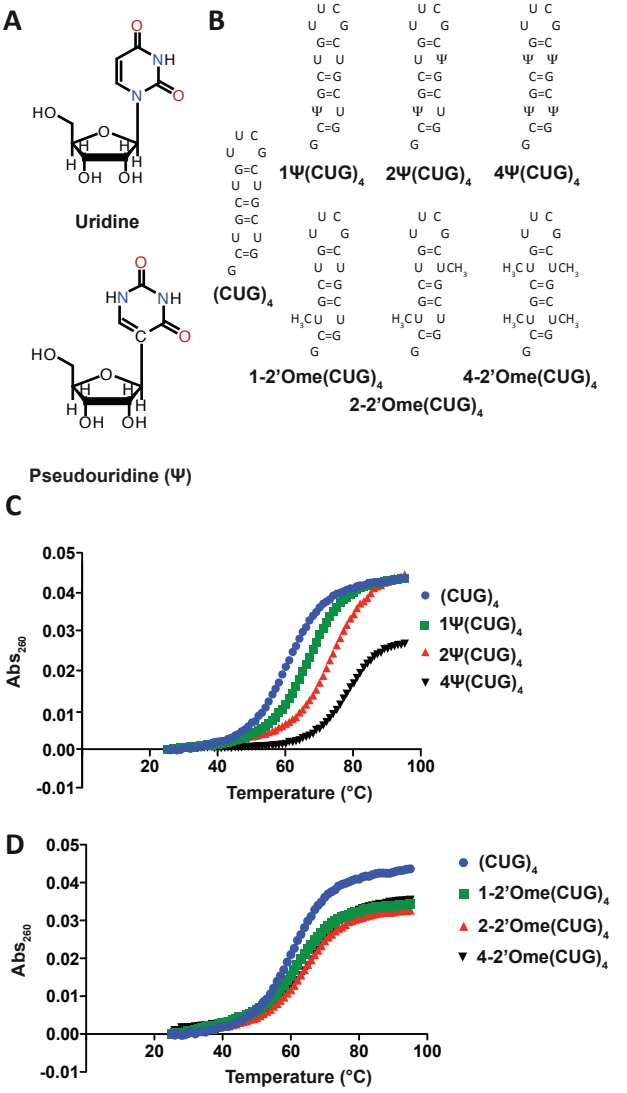


Figure 6: Pseudouridine and 2'-O-methyl modifications increase the structural stability of $(CUG)_4$ RNA. (A) Uridine and Ψ. Ψ an isomerization product of uridine. (B) Depictions of the structures of unmodified $(CUG)_4$ RNA along with $(CUG)_4$ with Ψ and 2'-O-methyl modifications. (C) Melting curves for $(CUG)_4$ RNAs modified with Ψ. The total change in melting temperature between unmodified $(CUG)_4$ and $4Ψ(CUG)_4$ is ~20 °C. (D) Melting curves for $(CUG)_4$ RNAs modified with 2'O methyl groups. The total change in melting temperature between unmodified $(CUG)_4$ and 4-2'Ome$(CUG)_4$ is ~7 °C.

RESULTS

Pseudouridine and 2'-O-methyl modifications increase the thermal stability of short $(CUG)_4$ stem-loops

A thermal melting assay was used to determine if Ψ and 2'-O-methyl modifications stabilize a short, helical $(CUG)_4$ RNA (Figure 6B). The modifications were targeted to the uridines because these were proposed as dynamic mismatches[33]. $(CUG)_4$ was synthesized with one, two and four Ψs in place of uridines (Figure 6B). Native $(CUG)_4$ melted at 59 °C, while the analogs with one, two and four Ψs melted at 67, 72 and 78 °C, respectively (Figure 6C, Table 1). $(CUG)_4$ RNA containing one, two and four 2'-O-methyl modification were studied as well (Figure 6B). The T_m increased to 62, 62 and 66°C with one, two and four 2'-O-methyl modifications. These data indicate that Ψ has a more pronounced effect on stabilization of CUG repeats in comparison to 2'-O-methylation.

Table 1: Melting temperatures and MBNL1-affinities for $(CUG)_4$ RNAs

RNA	T_m (°C)	MBNL K_D(μM)
CUG4	59+/-1	0.30 +/- 0.1
1ΨCUG4	67+/-2	1.0 +/- 0.3
2ΨCUG4	72+/-1	6.0 +/- 2.0
4ΨCUG4	78+/-1	>5
1-2'OmeCUG4	62+/-2	0.65 +/- 0.2
2-2'OmeCUG4	62+/-1	0.89 +/- 0.2
4-2'OmeCUG4	66+/-1	1.0 +/- 0.2

MBNL1 has a reduced affinity for pseudouridylated and 2'-O-methylated $(CUG)_4$

The RNAs shown in figure 6B were used in gel-shift assays to determine how Ψ and 2'-O-methyl modifications affected MBNL1's binding affinity for the $(CUG)_4$ RNA. MBNL1 binds to $(CUG)_4$ with a relatively strong affinity (K_d = 0.3 μM). The substitutions of one and two Ψs decreased MBNL1's affinity for $(CUG)_4$, resulting in K_ds for these RNAs of 1.0 and 6.0 μM. Four Ψ replacements in $(CUG)_4$ reduced MBNL1's affinity to such an extent that the K_d was immeasurable in the assay used (Figure 7A, C,

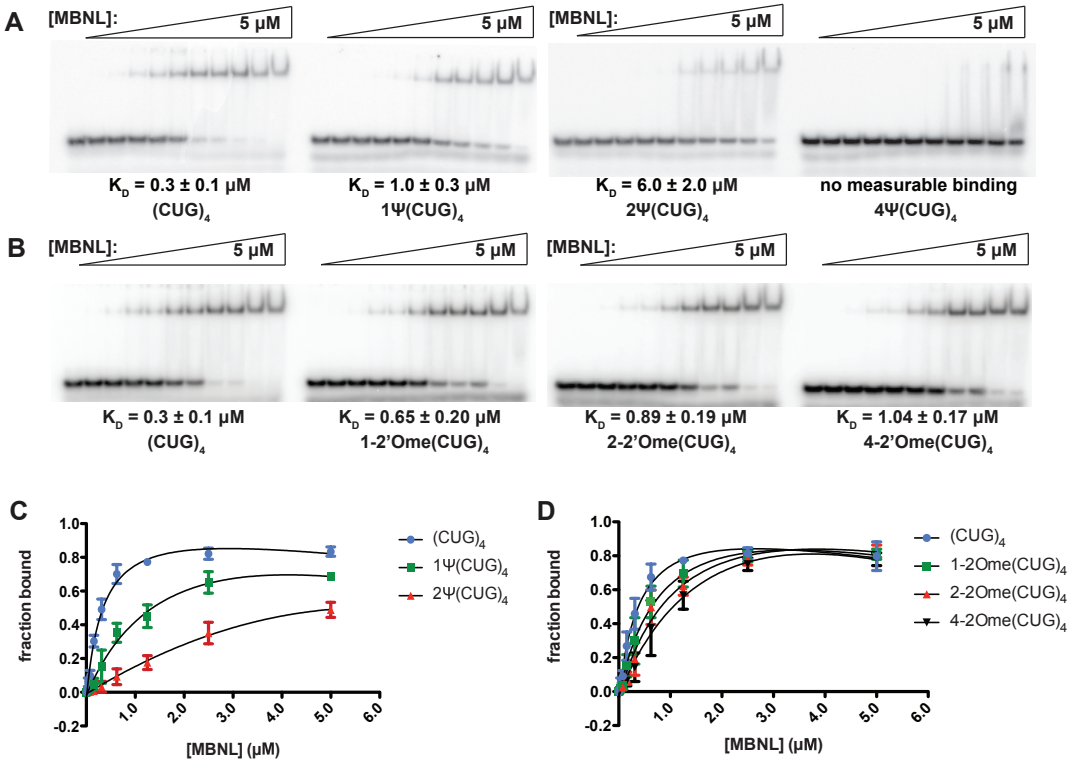


Figure 7: Binding gels and curves for MBNL1-$(CUG)_4$ RNA interactions. (A) Binding gels from MBNL1-$(CUG)_4$ affinity assays for RNAs modified with Ψ. RNAs were radiolabelled with α-P^{32} and combined with different concentrations of a consensus MBNL1 sequence (amino acids 2-260). RNA-protein complex formation induces a band-shift due to an increase in size. K_Ds were calculated based on the signal from the radiolabeled RNA and values are approximately the [MBNL1] when 50% of RNA is bound. (B) Binding gels from MBNL1-$(CUG)_4$ affinity assays for RNAs modified with 2-'O-methylation. (C) Binding curves for $(CUG)_4$ RNAs modified with Ψ. Data were fit to the following equation to determine K_D values: $f_{bound} = f_{max}([MBNL1]/([MBNL1 + K_D))$. (D) Binding curves for $(CUG)_4$ RNAs modified with 2'-O-methylation. K_D values were calculated using the above

Table 1).

Modifying $(CUG)_4$ with 2'-O-methylation also reduced MBNL1 binding such that 1, 2 and 4 methyl modifications increased the K_d to 0.7, 0.9 and 1.0 μM, respectively (Figure 7B, D, Table 1). While 2'-O-methylation did decrease MBNL1's affinity for CUG repeats, the effect was weaker than that of Ψ. These data are consistent with our model that increasing the stability of the double-stranded, helical conformation of the CUG repeats reduces or eliminates MBNL binding to CUG repeats in this conformation.

A water molecule bridges the Ψ-U mismatch in a crystal structure and molecular dynamics simulations

We previously utilized the GAAA tetraloop and its conserved 11 nucleotide receptor in order to facilitate the crystallization of CUG repeats (trCUG-3, Figure 8 A)[33]. In order to observe how Ψ affected the structure of CUG repeats, we substituted a Ψ in place of the U at position five in the trCUG-3 construct to create trCUG-3(Ψ5) (Figure 8A). The trCUG-3(Ψ5) construct readily crystallized and the structure was successfully solved using the trCUG-3 structure for molecular replacement (see Table 2 for data collection and refinement statistics).

Figure 8 (next page): A structure of CUG repeats with an incorporated pseudouridine and water occupancy through a molecular dynamics simulation. (A) Sequence and predicted secondary structure of constructs used in crystallography. CUG repeats (highlighted in blue) are attached to a GAAA tetraloop/receptor and the Ψ substitution is highlighted in red. trCUG-3 is the native structure and tr-CUG-3(Ψ5) has the Ψ substitution. (B) Contacts made in the U5-U31 base pair and the Ψ5-U31 base pair. A bridging water coordinates the amine group of Ψ5 and the carbonyl group of U31. (C) Simulated annealing omit map (Fo − Fc) of Ψ-U pair and bridging water. (D) $(CUG)_5$ RNA used in MD simulations. (E) Average degree of water occupancy throughout an MD simulation in U-U, Ψ-U and Ψ-Ψ base pairs. (F) Free energy required to open central base pairs.

This new structure of trCUG-3(Ψ5) at 1.90 angstroms resolution is nearly identical to the native trCUG-3 structure. Comparison of helical parameters of the CUG portion of

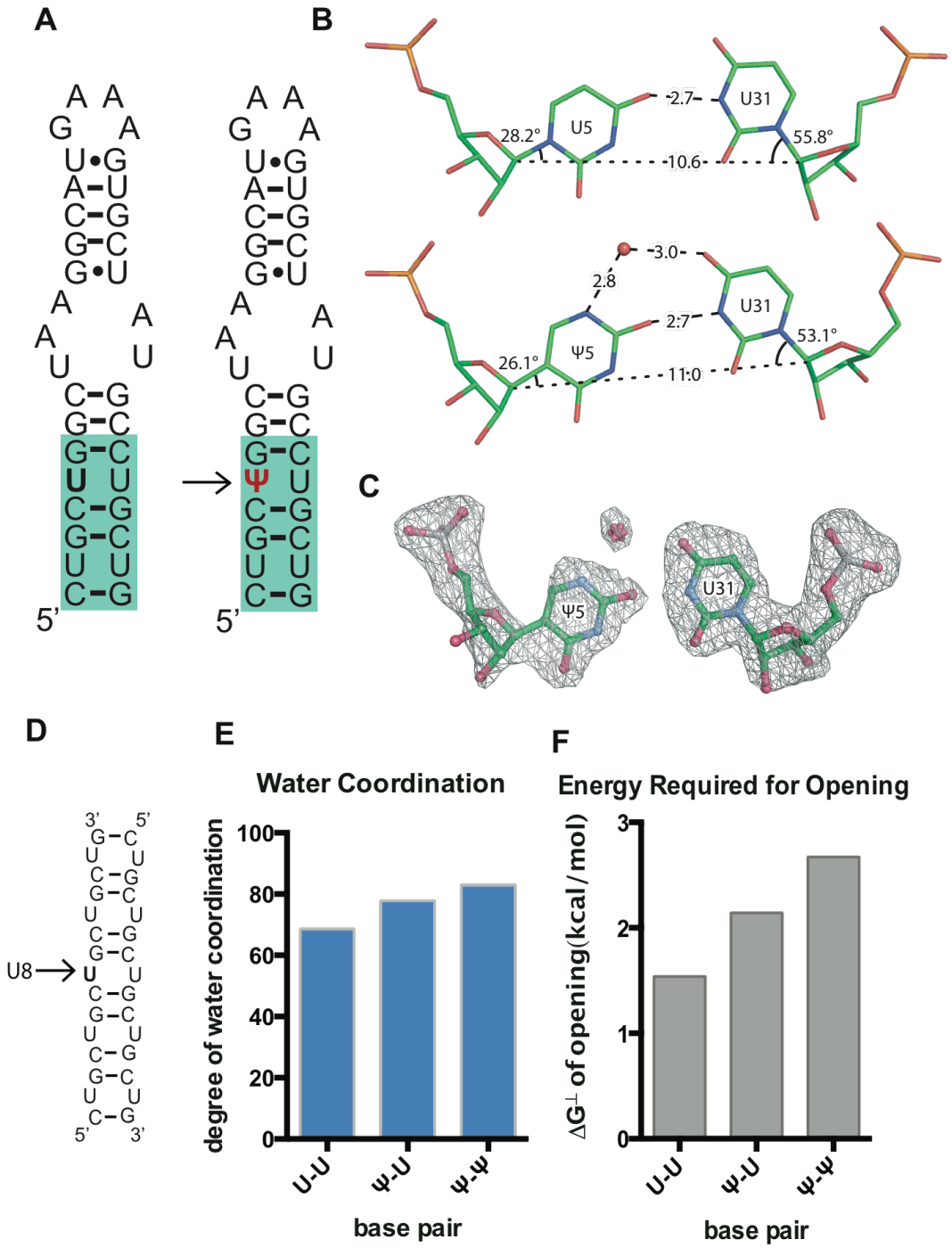


Table 2: Summary of crystallography data collection and refinement statistics

Measurement		**Value**		
Space group		R3		
Unit cell dimensions	**a, b, c (Å)**	69.70	69.70	67.80
	α, β, γ (°)	90	90	120
Resolution range, Å		22.60 – 1.90		
Total number of reflections		46959		
Number of unique reflections		9697		
Average redundancy		5.0		
% completeness		96.9 (89.6)		
I/σI		22.7 (2.4)		
R_{merge} [a]		0.051 (0.44)		
Average B-factors [no. of atoms]				
	nucleotides	52.32 [744]		
	solvent	55.783 [65]		
R_{free} [b]		0.267		
R_{work} [b]		0.196		

Values in parentheses represent highest resolution shell

[a] $R_{merge} = \frac{\sum |I - \langle I \rangle|}{\sum \langle I \rangle}$ where I is the observed intensity and $\langle I \rangle$ is the average of intensities obtained from multiple observations of symmetry-related reflections.

[b] $R_{factor} = \frac{\sum ||F_0| - |F_c||}{\sum |F_0|}$ where F_o and F_c are the observed and calculated structure amplitudes, respectively.

the lower helices showed two analogous, primary A-form helices (see Table 3 for helical parameters). However, there were key differences when the U5-U31 base pair from the native structure was compared to the Ψ5-U31 base pair (Figure 8B). Strikingly, there was a water molecule coordinating the amine group of Ψ5 to the carbonyl group of U31. Positions of the water and the non-canonical Ψ-U pair were confirmed by creating a simulated annealing omit map (Figure 8C). In order to form these two additional

hydrogen bonds, the Ψ5-U31 base pair has a greater opening angle compared to U5-U31 (Table 4) as well as a greater C1'-C1' distance (Figure 8B). This bridging water adds two additional hydrogen bonds between uridine and Ψ and may contribute to Ψ's structure-stabilizing capabilities in the context of CUG repeats.

To assess the potential importance of this bridging water molecule in the Ψ-U base pair we used molecular dynamics (MD) simulations to study the structure of CUG repeats containing Ψ-substitutions. MD simulations allowed us to investigate how Ψ substitutions affected water coordination and stability of an internal U-U base pair in a simulated solution. The systems investigated *in silico* were hybridized $(CUG)_5$ duplex RNA with zero, one, and two Ψ-substitutions at the middle (U8-U8) base pair (Figure 8D). Simulations clearly displayed the presence of bridging water molecules, consistent with the crystal structure. RNA with Ψ is much more likely to adopt structures with bridging water molecules. It also showed how the likelihood for RNA of adopting secondary structures mediated by bridging water molecules increases with one and two Ψ substitutions. Bridging water molecules were counted by looking for any identifying solvent molecules that made contact with both pairing bases simultaneously. In a typical 10 ns run of a MD simulation, the average degree of water coordination in a U-U mismatch throughout an MD simulation was 68.5% with respect to the water bridge. This value increased to 77.8% and 82.9% in Ψ-U and Ψ-Ψ mismatches, respectively (Figure 8E).

We also used *in silico* approaches to study the energy required to break open U-U, Ψ-U and Ψ-Ψ base pairs in MD simulations. As expected, pseudouridylated base pairs

required more energy to open. The Ψ-Ψ required an additional 1.5 kcal/mol more energy to open compared to the U-U base pair (Figure 8F). Here we've shown that CUG repeats containing Ψ are not only able and likely to form a water bridge, the presence of the water appears to contribute to structural stability.

Table 3: Comparison of CUG repeat helical parameters

	B DNA[1]	A DNA[1]	(CUG)2	(CUG)2(Ψ5)
Roll	**0.6**	**8**	**9.17**	**10.09**
Twist	36	31.1	33.48	33.68
Slide	0.23	-1.53	-1.34	-1.34
Rise	3.32	3.31	3.1	3.17
Inclination	2.1	14.7	18.57	20.05
Helical twist	36.5	32.5	35.89	36.09
x-displacement	0.05	-4.17	-4.7	-4.53
Helical rise	3.29	2.83	2.39	2.44
Zp	-0.36	2.24	2.35	2.23
Zp(h)	-0.02	4.19	4.62	4.74

Table 4: Comparison of U-U and Ψ-U non-canonical base pairs.

	trCUG-3		trCUG-3(Ψ5)	
	U2-U34	**U5-U31**	**U2-U34**	**Ψ5-U31**
C1'-C1'	8.6	10.6	8.7	11
1st hbond	2.8	2.7	2.8	2.7
2nd hbond	2.8	-	2.9	-
λ I	44.8	28.2	44.5	26.1
λ II	75.9	55.8	73	53.1
Incline[1]	minor	minor	minor	minor
Shear	-2.47	-2.63	-2.26	-3
Stretch	-1.89	-1.33	-1.87	-1.17
Stagger	0.26	-0.1	0.15	-0.34
Buckle	-7.12	5.62	-0.62	7.48
Propeller	-16.4	-7.12	-18.37	-6.69
Opening	11.6	-25.98	10.22	-38.77
Type	Type I	Type II	Type I	Type II

***In vitro*-transcribed CUG repeats disrupt MBNL-mediated alternative splicing and the effects are rescued by Ψ incorporation**

After determining that Ψ substitutions stabilized the formation of secondary structure in the CUG repeat RNA and prevented MBNL1 binding, we next asked how pseudouridylation of CUG repeats would affect MBNL-mediated splicing events in a cell model. The significant effect on MBNL1 binding to CUG repeats suggested that pseudouridylated CUG repeats would not sequester MBNL proteins and the proteins would be free to regulate alternative splicing. To test this hypothesis, *in vitro*-transcribed CUG repeats with and without Ψ modifications were transiently transfected into HeLa cells along with MBNL-regulated splicing targets.

We used the T7 transcription system to generate $(CUG)_{54}$ RNA with 0%, 50% and 100% Ψ content. The T7 polymerase produced equivalent quantities of RNAs with UTP and ΨTP, suggesting that this enzyme efficiently incorporated the Ψ base. After cleanup, these RNAs were co-transfected into HeLa cells with two different mini-genes containing MBNL1-regulated exons. The *TNNT2* and *INSR* mini-genes were used to monitor the effect of pseudouridylation on the ability of $(CUG)_{54}$ RNA to alter splicing. A plasmid [DMPK$(CTG)_{960}$] that had previously been shown to significantly alter MBNL-regulated splicing was used as a positive control[15], as well as a plasmid [DMPK$(CTG)_{40}$] with fewer repeats for comparison to the T7-produced $(CUG)_{54}$ RNA.

The basal level of exon 5 inclusion for *TNNT2* splicing reporter alone (mock) was 50% due to the activity of endogenous MBNL proteins in the HeLa cells (Figure 9A,C). Transfecting the *TNNT2* reporter as well as DMPK$(CTG)_{40}$ and DMPK$(CTG)_{960}$ resulted in 60% and 76% exon inclusion due to sequestration of MBNL proteins and a loss of

activity. MBNL proteins promote exclusion of exon 5 of *TNNT2*[15,57], so cellular expression of CUG repeats resulted in an increased level of exon inclusion as expected. Cells transfected with the *in vitro*-transcribed $(CUG)_{54}$ with 0%, 50% and 100% Ψ altered the *TNNT2* exon inclusion levels to 62%, 54% and 48%, respectively (Figure 9A,C). The 62% level of exon 5 inclusion is comparable to 60% exon 5 inclusion

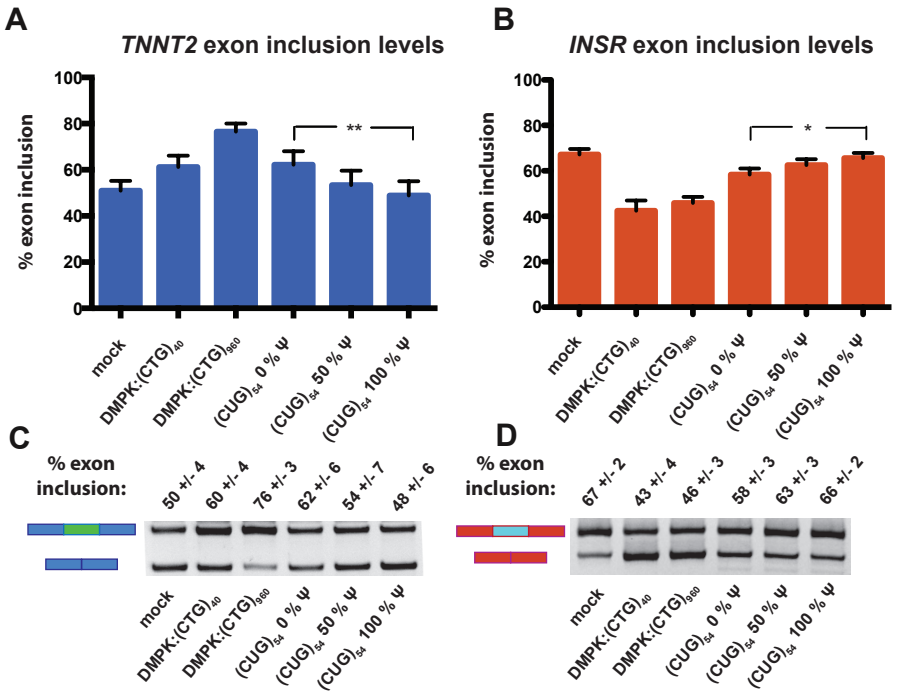


Figure 9: CUG repeats with increasing levels of pseudouridine are unable to induce mis-splicing of MBNL1-regulated targets *TNNT2* and *INSR*. (A) *TNNT2* exon inclusion levels from a HeLa cell-splicing assay. The difference in exon inclusion between the 0% Ψ samples and the 100% Ψ samples was statistically significant via a student's t test with a p=0.0021. (B) *INSR* exon inclusion levels from a HeLa cell-splicing assay. p=0.0239. (C) A representative gel from the *TNNT2* HeLa cell-splicing assay. Upper bands represent spliced transcripts with exons included and lower bands represent spliced transcripts with exons excluded. Quantification of these bands was used to obtain percent exon inclusion. (D) A representative gel from the *INSR* HeLa cell-splicing assay.

observed with the DMPK$(CTG)_{40}$ plasmid, suggesting these two approaches resulted in similar levels of CUG repeats, and thus MBNL sequestration, in the cells. Ψ incorporation into $(CUG)_{54}$ resulted in reversion to the basal level of exon inclusion (50%), presumably because MBNL proteins were no longer sequestered to the CΨG repeats and they were free to regulate *TNNT2* exon 5 inclusion.

To determine if pseudouridylation was able to reduce or eliminate the change in splicing of another MBNL-mediated event, we measured exon inclusion levels of *INSR*. For this splicing event, MBNL1 promotes inclusion of the regulated exon, exon 11[27]. HeLa cells transfected with only the *INSR* reporter had an exon inclusion level of 67%, and co-transfection with DMPK$(CTG)_{40}$ and DMPK$(CTG)_{960}$ resulted in 43% and 46% exon 11 inclusion (Figure 9B,D). Transfection of $(CUG)_{54}$ with 0% Ψ resulted in 58% exon 11 inclusion, an effect which is partially rescued by 50% Ψ (63% exon 11 inclusion) and nearly fully rescued by 100% Ψ (66% exon 11 inclusion) (Figure 9B,D). These results are consistent with pseudouridylation of the CUG repeats reducing or eliminating the ability of the modified RNA to sequester MBNL proteins in this DM1 cell model.

Pseudouridylated CUG repeats form reduced-sized foci

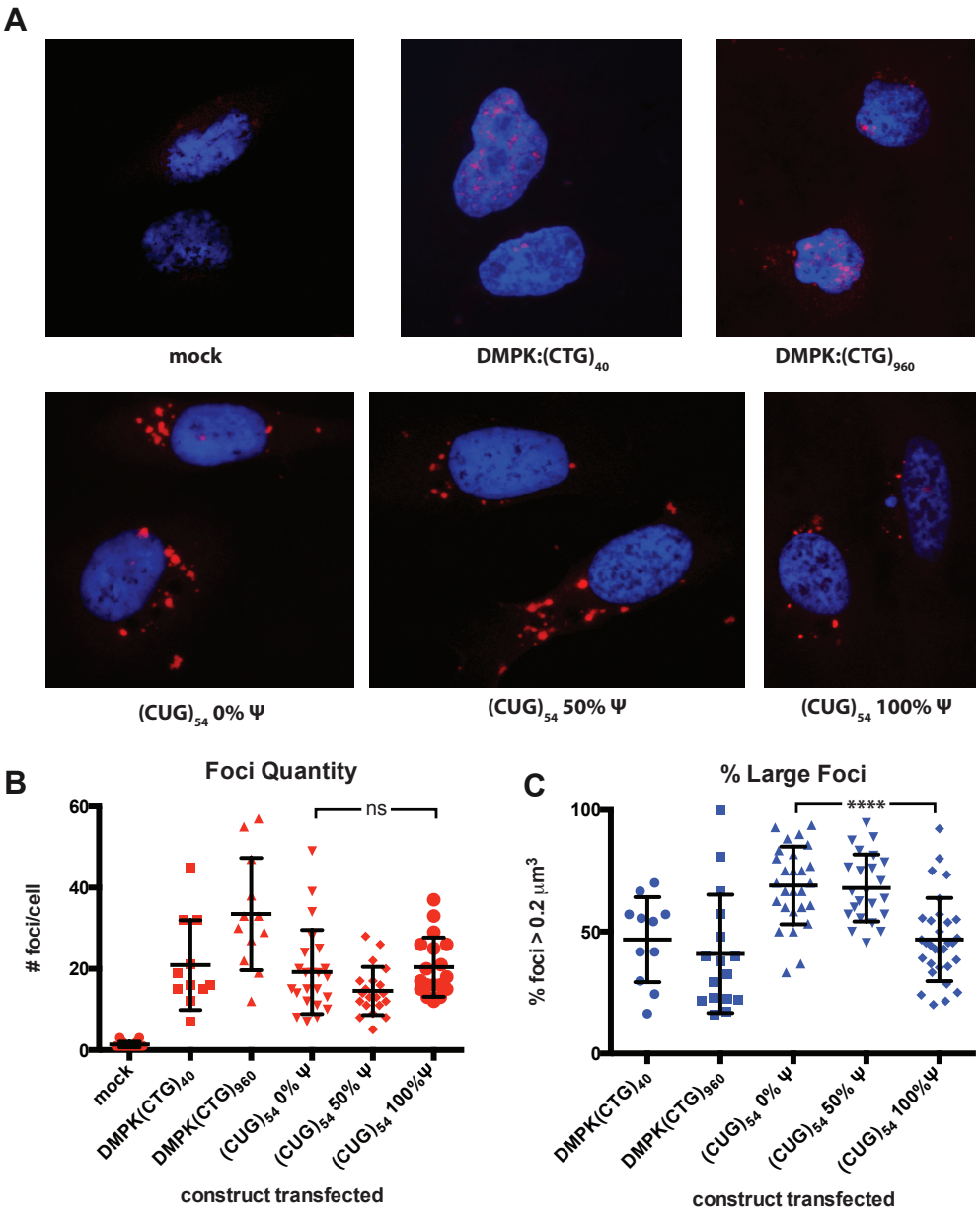
Previous studies have show that extended CUG repeats form nuclear foci in DM1 tissues, animal and cell culture models[14]. To determine whether pseudouridylated RNA is capable of forming RNA foci, we performed Fluorescence In-Situ Hybridization, probing for CUG repeats with a Cy-3 labeled $(CAG)_{10}$ probe. HeLa cells were transfected with $(CUG)_{54}$ RNA transcribed with 0%, 50% and 100% Ψ, as well as plasmids containing the DMPK 3'UTR with 40 and 960 CTG repeats as described above for the splicing assay.

Cells transfected with an unrelated plasmid (mock) formed 1.3 ± 0.6 foci/cell, while those transfected with DMPK$(CTG)_{40}$ and DMPK$(CTG)_{960}$ formed 21 ± 11 and 34± 12 foci/cell, respectively. When we transfected cells with $(CUG)_{54}$ RNA with 0%, 50% and 100% Ψ they formed 19 ± 10, 15 ± 6 and 20 ± 7 foci/cell, respectively and there was no significant difference between unmodified and fully modified transfected RNA (Figure 10, Table 5). However, cells transfected with fully modified CUG repeats tended to form smaller foci than those transfected with unmodified RNA. Large foci were deemed those with a volume larger than 0.2 μm^3. The cells transfected with unmodified CUG repeats had 69 ± 17% large foci, while those transfected with fully modified CUG repeats had 47 ± 17% large foci. The difference in percent-large-foci between unmodified and modified RNA is significant with a p value less than 0.0001 (Figure 10, Table 5).

Table 5: Quantification of CUG repeat foci

construct	# foci	% lg foci (0.2 um^3)
mock	1.3 ± 0.6 N=16	N/A
DMPK:(CTG)40	21 ± 11 N=11	47 ± 17 N=11
DMPK:(CTG)960	34 ± 12 N=12	41 ± 24 N=16
(CUG)54 0%	19 ± 10 N=23	69 ± 17 N=27
(CUG)54 50%	15 ± 6 N=20	68 ± 14 N=23
(CUG)54 100%	20 ± 7 N=18	47 ± 17 N=31

Figure 10 (next page): Pseudouridylated CUG repeats form reduced-size foci in HeLa cells. (A) Images of HeLa cells transfected with mock plasmid, plasmids containing 40 and 960 CTG repeats in the DMPK gene as controls, as well as cells transfected with (CUG)54 RNA with 0%, 50% and 100% Ψ . (B) Cells transfected with 960 CTG repeats tended to form more foci, while those transfected with 40 CTG repeats and modified and unmodified RNA formed similar numbers of foci. (C) Modified RNA did induce formation of foci that tended to be smaller than 0.2 μm3.



Pseudouridylation suppresses CUG repeat-associated toxicity in a zebrafish model of myotonic dystrophy

To assess the *in vivo* effects of Ψ incorporation on CUG repeat-associated toxicity, we utilized an established zebrafish model of DM1[68]. This model utilizes *in vitro*-transcribed RNA encoding GFP fused with the 3'UTR of DM1, which contains 91

CUG repeats. When injected into single cell embryos, this CUG RNA triggers alterations in morphology, viability, and motor behaviors within the first 24 hours of embryo development (Figure 11B,C)[68]**.** To assess the impact of pseudouridylation on CUG repeat toxicity *in vivo,* transcribed GFP-$(CUG)_{91}$ RNAs with Ψ-NTP were injected into zebrafish embryos. We used GFP alone-encoding mRNA lacking CUG expansions with uridine or Ψ as a control (Figure 11A). RNA levels post-injection were measured using

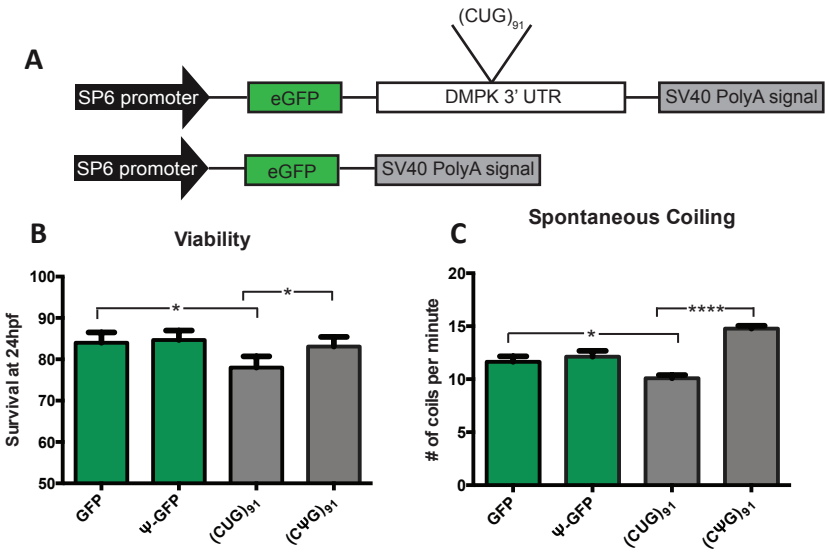


Figure 11: Viability and motoric function of zebrafish embryos injected with CUG repeat RNA is modified by pseudouridylation. (A) DMPK and GFP control constructs injected into zebrafish. (B) Viability of zebrafish embryos at 24 hours post fertilization after injection with *in vitro*-transcribed GFP or GFP:$(CUG)_{91}$ There was no significant difference in the viability of embryos injected with GFP with and without Ψ. N>400/>500 zebrafish per genotype over at least three independent experiments. For GFP vs $(CUG)_{91}$ the p value was 0.014. For RNA ($\Psi(CUG)_{91}$ vs $(CUG)_{91}$ the p value was 0.030. $P<0.05$ by Fishers exact test. (C) Spontaneous coiling movements in 24 hpf zebrafish embryos injected with the indicated *in vitro*-transcribed RNAs. $\Psi(CUG)_{91}$ RNA vs GFP ($p<0.0001$). N>100/ zebrafish per genotype over 3 independent experiments were analyzed. All groups were significantly different by a Kruskal Wallis ANOVA. * indicates $p<0.0001$ by Post-hoc Mann-Whitney U test.

qRT-PCR to determine whether pseudouridylated RNA has a similar stability in the embryos and there was no significant difference between levels (Figure 12). As previously described[68], injection of GFP-$(CUG)_{91}$ mRNA into embryos led to a significant decrease in viability at 24 hours post-fertilization (hpf) compared to embryos injected with an equimolar amount of GFP RNA [GFP vs $(CUG)_{91}$] (Figure 11B)[68]. In contrast, when embryos were injected with GFP-$(CUG)_{91}$ RNA transcribed with 100% Ψ-NTP, survival at 24 hpf was significantly enhanced compared to 0% Ψ-NTP GFP-$(CUG)_{91}$ and was similar to embryos injected with GFP RNA (Figure 11B).

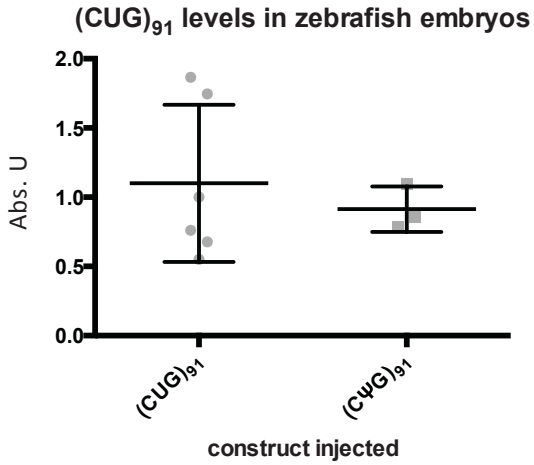


Figure 12: CUG repeat levels in zebrafish are independent of modification. CUG repeat levels were measured in zebrafish embryos injected with unmodified and fully pseudouridylated (CUG)91 . The method used to measure RNA levels was qRT-PCR using primers specific for the injected transcript. Modified and unmodified levels of RNA were insignificantly different (p=0.61).

Zebrafish models of neuromuscular disorders often exhibit abnormalities in basic motor behaviors during early development [68,69]. The first observable indication of skeletal muscle activity is spontaneous coiling, the alternating contraction of trunk and tail that begins at 17 hpf, peaks at 19 hpf and then decreases over the next 8 hours [70]. At 24 hpf, zebrafish embryos injected with CUG repeats displayed a defect in spontaneous coiling

[GFP: 12 ± 0.52 coils/min, $(CUG)_{91}$:10 ± 0.30 coils/min, p=0.022] (Figure 11C) [68]. However, pseudouridylation of the CUG RNA completely abolished this toxicity, resulting in an increase in coiling activity [Ψ $(CUG)_{91}$:15 ± 0.26 coils/min, p<0.0001] (Figure 11C). These results indicate that substituting Ψ for uridine in CUG repeats reduces the toxicity of the RNA as demonstrated by the rescue in viability and rescue (enhanced activity) of coiling activity.

DISCUSSION

Structural and biochemical studies on MBNL1 with RNA substrates indicated that this family of proteins prefers to bind single-stranded or partially structured RNA[38]. This lead us to hypothesize that stabilizing the CUG repeats in a helical conformation with RNA modifications would decrease or eliminate MBNL binding to this toxic RNA. Of two RNA modifications tested, we found Ψ significantly stabilized the CUG repeats in a helical (A-form) conformation, and, as hypothesized, this stabilization of CUG repeats inhibited MBNL1 binding (Figures 6 and 7). Furthermore, in a DM1 cell model, pseudouridylation of CUG repeats reduced and eliminated the ability of the CUG repeats to cause mis-splicing of two MBNL1 splicing reporters (Figure 9). Significantly, pseudouridylation of a CUG repeat-containing mRNA protected zebrafish embryos from increased morbidity and decreased spontaneous coiling compared to mRNA without the Ψ modification (Figure 11). These studies are consistent with the model shown in figure 13 that depicts pseudouridylation stabilizing CUG repeats in a helical conformation and not sequestering MBNL proteins (Figure 13C,D) while unmodified CUG repeats are less stable (or more prone to partial unfolding) and MBNL proteins are sequestered by the

repeats, and mis-splicing results (Figure 13A,B). While CUG repeats have been shown to form stable stem-loop structures via structure-probing assays, we have shown that Ψ induces further stabilization, preventing MBNL1 binding[31].

Past efforts to understand the mechanism through which Ψ stabilizes RNA structure indicated that this modification leads to stronger base-stacking interactions with surrounding nucleotides and increased hydrogen bonding contacts through coordination

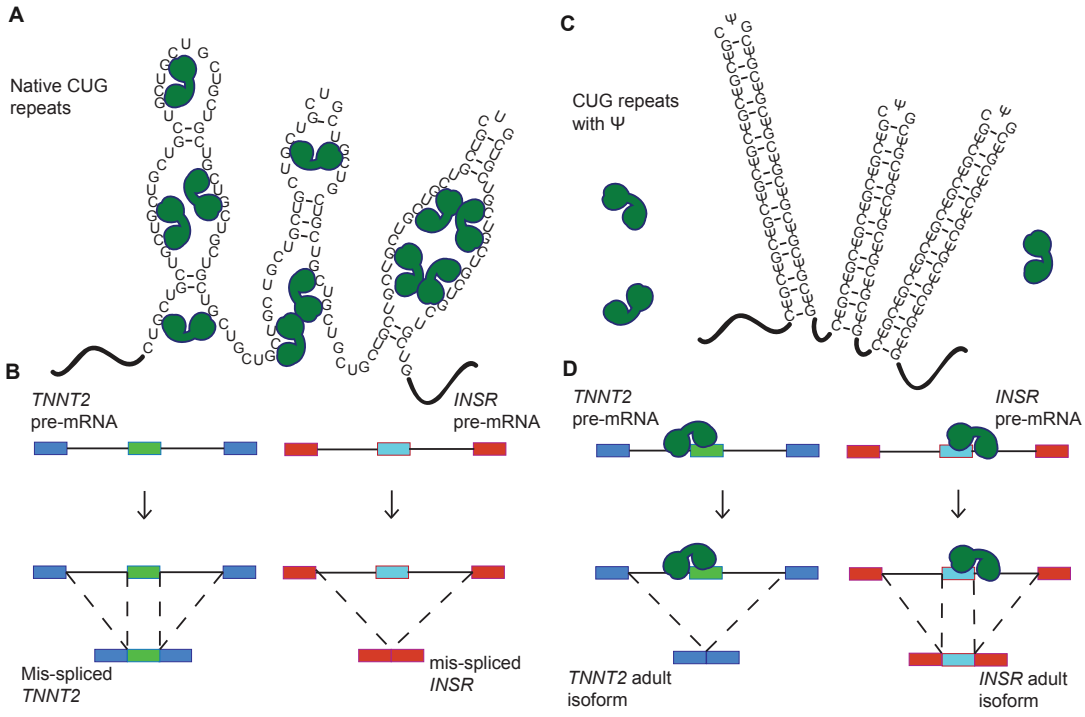


Figure 13: A model for CUG repeat detoxification via pseudouridylation. (A) CUG repeats form unstable helices that adopt multiple conformations, including partially unstructured RNA. MBNL1 binds to the CUG repeats and is sequestered to their location in the nucleus. MBNL1 is represented by green shapes. (B) When MBNL1 is sequestered to CUG repeats, it is unavailable to bind to the pre-mRNAs that it regulates. Here CUG repeats increased exon inclusion in the *TNNT2* transcript. CUG repeats result in decreased exon inclusion in the *INSR* mRNA in the presence of CUG repeats. (C) CUG repeats with Ψ (CΨG repeats) form stable stem-loop structures. In these structures, MBNL1's binding site is unavailable for interaction and MBNL1 is not sequestered. (D) MBNL1 is free to bind to its target transcripts when CΨG repeats are present.

with water molecules[49]. According to our analyses, the coordination of the water molecule induced formation of two extra H-bonds in the Ψ-U base pair compared to the U-U base pair (Figure 8B,C). Although this water in the crystal structure could be due to crystal packing, the MD studies with a CUG RNA containing Ψ showed a significant increase in the degree of water coordination in this same position compared to the unmodified CUG RNA (Figure 8D-F). It is likely that Ψ stabilizes RNA structure through different mechanisms and this new mechanism adds to the repertoire[49].

Our results are consistent with CAG antisense and small molecule approaches that stabilize the CUG repeats in helical structures (no mismatches in CAG-CUG antisense interactions) or through binding of molecules to the CUG repeats in helical conformation[71-73]. However, CAG antisense and small molecule approaches do not address the secondary structure of RNA. Because modifying CUG repeats to stabilize structure prevents MBNL binding, we've shown that the secondary structure of the RNA is important in the DM1 disease mechanism. Furthermore, these data are congruent with approaches that stabilize CUG repeats in helical conformations and specific targeting of the U-U mismatches. SnoRNAs have been used to target RNAs with some promising results for altering RNA structure and function in cells[52,53]. To significantly pseudouridylate CUG repeats *in vivo* with snoRNAs would be quite challenging because of the abundance of sites and lack of specificity for specific uridines within the repeats, therefore this avenue is not currently being pursued.

More broadly, our results suggest that the conformation of toxic RNAs should be considered for disease mechanisms and therapeutic strategies. For example, the CCUG repeats that cause myotonic dystrophy type 2 can adopt multiple RNA structures

including possibly unfolded when protein-bound, partially unfolded, and two different helical structures with different mismatches[20,35]. The expanded CGG repeats in the 5' UTR of the fragile X mental retardation 1 gene results in FTAXS, a disease characterized by ataxia, tremors, and neurodegeneration. Extended expansion of these repeats results in FXS, a common form of mental retardation[74]. RNA binding proteins associate with helical CGG repeats and are sequestered to nuclear inclusions[63]. Spinocerebellar ataxia 8 (SCA8) is caused by expansion of CTG repeats in the *ataxin* 8 gene, and the resulting RNA gain of function produces cerebellar atrophy and coordination defects[75,76]. Expanded CUG repeats in SCA8 could exhibit the same pathogenic mechanism as that in DM1 and DM2, and RNA modifications could be used to help guide therapeutic strategies. SCA10 and ALS/FTD are autosomal dominant neurodegenerative diseases that are caused by expanded penta- and hexanucleotide repeats, and in both cases RNA-binding proteins are sequestered to repeats, resulting in mis-regulation of splicing[77-79]. The RNA modification strategy described here is a powerful approach to address the conformation of RNA repeats associated with human diseases that should be considered for disease mechanism and therapeutic targeting.

MATERIALS AND METHODS

Thermal melts were carried out in 20 mM PIPES buffer at a pH of 7.0 and 150 mM NaCl. The buffer was de-gassed before diluting RNA to a concentration of 2 μM. The temperature was raised by 1 °C per minute over a range of 25-95 °C, and absorbance at 260 nm was monitored by a Cary UV/Vis spectrophotometer. Absorbance was normalized by subtracting the absorbance at 25 °C.

GST-tagged MBNL1 (amino acids 2-260) was expressed in Arctic Express *E. coli* cells (Agilent Technologies™) grown to an OD of 0.4, induced with 25 mM IPTG and shaken for 20 hours at 16 °C. GST-MBNL1 was initially purified over a Hi-TrapGST column on an FPLC. MBNL1 was then cleaved from GST by *PreScission Protease* and purified further by affinity chromatography over a Hi-TrapHeparin column. Both columns and the protease were made by GE Healthcare Life Sciences™. MBNL1 concentration was measured by a Bradford assay and the protein was stored in a buffer of 500 mM NaCl, 50% glycerol, 25 mM Tris at pH of 7.5 and 5 mM BME at –80 °C until needed.

RNA synthesized by Dharmacon was radiolabeled with α-^{32}P phosphate with a polynucleotide kinase and stored at -20 °C in 20 mM Tris, pH 7.5. Prior to the binding reaction, CUG helices were diluted and folded by incubation at 95 °C in 20 mM Tris, pH 7.5, 150 mM NaCl, and 5 mM $MgCl_2$ for 2 min, followed by incubation on ice for 5 min. The RNA was then incubated with varying MBNL1 concentrations in 175 mM NaCl, 20 mM Tris, pH 7.5, 1 mM BME, 10% glycerol, 5 mM $MgCl_2$, 0.1 mg/mL heparin, 2 mg/mL BSA, 0.02% xylene cyanol, and 0.05% bromophenol blue. The incubation was 20 minutes at room temperature. RNA-protein complexes were separated from free RNA on a 6% native acrylamide gel. Complex formation was quantified by exposure on a phosphorimager screen followed by analysis with *ImageQuant* from GE Healthcare Life Sciences™. Affinity constants were then calculated with the following equation: $f_{bound} = f_{max}([\text{MBNL1}]/([\text{MBNL1}+K_d))$. The images of the gels were enhanced equally across each gel using Adobe Photoshop. No specific feature was enhanced.

The trCUG-3(Ψ5) construct was purchased from Dharmacon and deprotected per the manufacturer's instructions. The RNA was then purified using HPLC and brought to a final concentration of 0.5 mM in a solution of 15 mM NaCl, 5 mM Tris (pH 7.5), and 5 mM $MgCl_2$. The RNA was annealed by heating to 70°C for 5 min and rapidly cooled to 4°C. The best crystals grew from a mixture of 2 μL of RNA solution and 2 μL of well solution containing 4 mM $MgSO_4$, 50 mM Tris (pH 8.5), and 30% (w/v) 1,6-hexanediol. Crystals appeared in approximately 1 week.

Crystals were mounted in rayon loops and flash frozen in liquid nitrogen. Experimental data were collected at the Structural Biology Center at Argonne National Laboratory on the 19ID beamline under a cryostream. The X-ray data were integrated, merged, and scaled using the HKL-2000 program suite[80] and converted to structure factors using the CCP4i GUI[81] for the CCP4 program suite[82]. Data collection statistics are listed in Table 2.

The structure was determined using molrep[83], a part of the CCP4 program suite[82], using the Protein Data Bank (PDB) entry 4FNJ, the structure of trCUG-3 [33], as the search model. The model was rebuilt using Coot[84] and refined with refmac[85]. Refinement statistics are listed in Table 2. Coot and PyMOL [Schrodinger, LLC (2010) The PyMOL Molecular Graphics System, Version 1.5.0.1.] were used to generate figures. Simulated annealing omit maps were created using CNS[86] by deleting the Ψ5-U31 base pair as well as the coordinating water while all other nucleotides remained fixed during annealing. The structure was deposited in the PDB as entry 4PCJ. Structural parameters were computed using 3DNA[87].

Double-stranded $(CUG)_5$ starting structures with 0, 1, and 2 Ψ modifications at

the central base pair were constructed from helical crystal structures of CUG repeats (PDB ID: 4FNJ). All simulations were performed utilizing the GROMACS[88-91] molecular dynamics package on the ACISS cluster at the University of Oregon, using an AMBER[92] force-field optimized for stem-loop RNA structures with modified force-field parameters for the Ψ base[93]. All simulations were performed in explicit solvent utilizing the spc/e water model and 150 mM NaCl with excess sodium ions to neutralize the system, Ewald summation for the electrostatics, standard force-field cutoffs and parameters, and in the NVT ensemble. A 1 fs timestep was utilized, along with standard energy minimization and equilibration techniques, with over 5 nanoseconds of equilibration before collecting production data.

After equilibration, water-bridge occupancy was measured using the water-bridge collective variable of the PLUMED software package[94]. This collective variable is defined by a numerical function which assigns a continuous measure to the number of solvent atoms which are within 2 angstroms of an atom belonging to the A strand U8/Ψ8 base and also within 2 angstroms of an atom belonging to the pairing B strand U8/ Ψ8 base. This numerical function is rather lengthy and we refer the reader to the PLUMED documentation for further details. Statistics were collected over 10 ns of simulation.

To obtain the free-energy in the center-of-mass base separation distance, first an approximate opening pathway was obtained by using the PLUMED[94] implementation of Parrinello's metadynamics algorithm[95] to jointly bias the number of base-pairing contacts and the center-of-mass base separation. Twenty images were chosen along this pathway, anchored using a harmonic restraining potential, from which umbrella sampling runs of 1 ns each were performed. Final calculation of the free energy along the distance

coordinate was performed using the weighted historgram analysis method[96] (Grosfield, Alan, "WHAM: the weighted histogram analysis method", version 2.09, http://membrane.urmc.rochester.edu/content/wham).

$(CUG)_{54}$ was transcribed off of 300 ng of a linearized plasmid template containing a T7 promoter upstream of 54 CTG repeats. The reaction conditions were as follows: 0.1 M DTT, 40 mM Tris, pH 8.0, 8 mM $MgCl_2$, 50 mM NaCl, 2 mM spermidine 1 mM ATP, 1 mM CTP, 1 mM GTP, 1 mM UTP/ΨTP, 1 µL RNasin (Ambion) and 1 µL T7 polymerase in a final volume of 30 µL. Reactions were incubated at 37 °C for 2 hrs followed by addition of 3.3 uL of RQ1 DNase Buffer and 1 µL RQ1 DNase (Promega) and a 1 hr incubation at 37 °C. Transcribed RNA was purified by the addition of 80 µL of a 3.75 M LiCl, 25 mM EDTA solution and an overnight incubation at -20 °C. The RNA was then pelleted by centrifugation at 13,000 RPM and the liquid was removed. The pellets were washed in 1 mL of 70% ethanol, followed by re-pelleting and re-suspension in 30 µL of ddH_2O. RNA concentrations were measured on a NanoDrop UV/vis spectrophotometer.

Reporter minigenes, $DMPK(CTG)_{960}$, $DMPK(CTG)_{40}$, and *in vitro*-transcribed RNA were transfected using Lipofectamine 2000 (Life Technologies). Control plasmids were added to transfections containing only reporter minigenes such that each set of cells received 1000 ng of nucleic acid. HeLa cells were routinely cultured as a monolayer in Dulbecco's modified Eagle's medium (DMEM)-GlutaMax medium (Invitrogen) supplemented with 10% fetal bovine serum (Gibco) and 1x antibiotic/antimycotic at 37 °C under 5% CO_2. Prior to transfection, cells were plated in six-well plates at a density of 2.0×10^6 cells per well. Cells were transfected 24 hrs later at approximately 90%

confluence using 500 μL of Opti-MEM (Life Technologies) and 5 μL of Lipofectamine 2000. After 4 hours, Opti-MEM was replaced with DMEM. Cells were harvested in-plate after 24 hrs using the RLT buffer from QIAgen's RNeasy kit.

Total RNA was isolated with the RNeasy kit per the manufacturer's instructions. RNA samples were treated with RQ1 DNase and exon inclusion levels were analyzed by reverse transcription and PCR amplification as performed previously in this group[97]. Images of the gels were inverted and enhanced equally across each gel using Adobe Photoshop. No specific feature was enhanced.

Zebrafish experiments were conducted as described[68]. Briefly, embryos were isolated after paired mating of AB zebrafish (zFIN, Eugene, OR) and injected at the 1–2-cell stage using a Drummond Nanoject. Diluted *in vitro*-transcribed capped and polyadenylated mRNA (4.6 nl each) was injected at a concentration of 100 ng/μL for all constructs unless otherwise specified (approximate total amount of RNA injected/embryo = 0.46 ng). Survival was determined by measuring the number of intact embryos after injection for each genotype and then determining the number of viable embryos at 24 hpf. For each injected RNA, as least 3 independent experiments were conducted, with at least two separate injection batches conducted per injected RNA at each experiment. The survival % was pooled across all studies and the 95% confidence interval was calculated using the method of Newcombe[98]. Fisher exact test was used to compare groups.

Spontaneous coiling was measured as previously described[69,99]. Briefly, sets of 5 embryos injected with the indicated RNAs were collected at 24 hpf and observed over four different 15-second periods. The number of coiling movements per minute per embryo was calculated. N>100/genotype from 3 independent experiments. A one-way

Kruskal Wallis ANOVA across genotypes was performed with post-hoc unpaired mann Whitney U tests conducted on pre-specified comparisons (GFP vs. CUG91 and CUG91 vs. Ψ CUG91).

Flourescence In-Situ Hybridization was carried out on HeLa cells transfected using the same procedure described above. Sixteen hours after transfection, cells were fixed in 4% paraformaldehyde for 15 minutes. Cells were then permeabilized with 0.5% triton X-100, in 1X PBS at RT for 5 min. Cells were prewashed with 30% formamide, 2X SSC for 10 min at RT. Cells were then probed overnight at 37 °C, with a 1 ng/μL of Cy3 $(CAG)_{10}$ probe (IDT, IA) in 30% formamide, 2X SSC, 2 μg/mL BSA, 66 μg/mL yeast tRNA, 2 mM vanadyl complex. Cells were then washed for 30 minutes in 30% formamide, 2X SSC at 42 °C, and then with 1X SSC for 30 min at RT. Cells were mounted onto glass slides using the hard-set mounting media that contains DAPI (Vectashield). Cells were imaged on an Olympus Fluoview FV1000 with a Bx61 scope. Unpaired, two-tailed t-tests were performed using Graphpad's Prism 6. Images were prepared using ImageJ and Adobe Photoshop and the number of foci were quantified using ImageJ's object counter. Large foci were quantified with ImageJ as well.

CHAPTER III

PSEUDOURIDINE MODIFICATION INHIBITS MUSCLEBLIND-LIKE 1 (MBNL1) BINDING TO CCUG REPEATS AND SINGLE-STRANDED RNA THROUGH REDUCED RNA FLEXIBILITY

This manuscript has been submitted and reviewed by the Journal of Biological Chemistry. I am the primary author on this manuscript. I performed the thermal melt assays and the binding experiments. Jeremy Copperman performed the molecular dynamics simulations. Dr. J. Andrew Berglund was the primary investigator for this work.

INTRODUCTION

Myotonic dystrophy type 1 (DM1) is a genetic neurodegenerative disease caused by an expansion of CTG repeats in the 3'UTR of the dystrophia myotonia protein kinase (*DMPK*) gene. Similar to DM1, myotonic dystrophy type 2 (DM2) is caused by expanded CCTG repeats in the zinc finger protein 9 (*ZNF9*) gene. DM1 and DM2 occur when the CTG/CCTG repeats are expanded beyond 40 repeats and patients can have up to thousands of CTG/CCTG repeats[29,54]. The currently accepted DM1 and DM2 disease mechanism is that the expanded repeats sequester RNA binding proteins (primarily the Muscleblind-like family), which prevents these proteins from performing their functions in cells[12,37].

The Muscleblind-like family of proteins (MBNL1, MBNL2 and MBNL3) bind RNA and regulate several RNA processing pathways including alternative splicing, pre-

miRNA biogenesis, mRNA localization and circular RNA generation[16,17,100,101]. MBNL proteins bind the consensus YGCY RNA sequence[15,16]. CUG and CCUG repeats are composed of YGCY motifs creating hundreds or thousands of perfect MBNL binding sites resulting in large numbers of MBNL proteins binding the repeats and forming nuclear foci[18]. When MBNL proteins are sequestered they are unable to regulate RNA processing events and consequently, many DM1 and DM2 symptoms are caused by mis-regulated alternative splicing and potentially the loss of other MBNL activities[11]. It is therefore important to understand how MBNL proteins bind their toxic and cellular RNA substrates, in order to develop mechanisms to alleviate MBNL sequestration in DM1 and DM2.

MBNL proteins have two sets of Zinc finger domains that are proposed to bind RNA on opposing faces of the domain[58,59]. Teplova *et al.* published a crystal structure of one of MBNL1's Zinc finger domains in complex with a YGCY-containing RNA, showing that MBNL1 interacts with the Watson-Crick face of the GC dinucleotide[38]. This structure indicates that at least this portion of the RNA must be single-stranded in order to interact with MBNL1, as the Watson-Crick face interacts with the opposing strand in double-stranded RNA.

Pseudouridine (Ψ) is the most abundant RNA modification, often referred to as the fifth RNA base. Ψ is found in tRNA, rRNA and many spliceosomal RNAs and functions to stabilize the structure of these RNAs[42]. Ψ is a 5-ribosyl isomer of uridine, with a C-C bond connecting the ribose to the base instead of a C-N bond and an extra H-bonding donor on the N1 (Fig. 14A). Nuclear magnetic resonance, X-ray crystallography and MD simulations have shown that Ψ induces rigidity in both single and double-

stranded RNA via base-stacking and hydrogen bonding interactions[42,46,49]. There is evidence that in some cases, base-stacking interactions are the primary stabilizing force in RNA containing Ψ[47].

Replacing uridines with Ψ in the CUG repeats induced structural stabilization, prevented MBNL1 binding and rescued mis-splicing. X-ray crystallography data suggested that Ψ stabilized the CUG repeats through water-bridging via hydrogen bonding of Ψ's N1-H with the uridine on the opposing strand [25]. In this work we have extended our studies to determine that Ψ can stabilize CCUG repeat RNA and single-stranded RNA containing YGCY motifs and this modification inhibits MBNL1 binding. Molecular dynamics simulations indicate that Ψ increases base-stacking interactions in single-stranded YGCY RNAs.

RESULTS

Replacing uridine with Ψ in CCUG repeat RNA increased structural stability

Previous studies showed that Ψ increased the thermal stability of CUG repeat RNA in a helical conformation[25]. Here we determined if Ψ had a similar effect on stabilization of CCUG repeats. The additional cytosine in CCUG repeats renders them different from CUG repeats in two ways. First, there is only uridine adjacent to the GC dinucleotide, so Ψ substitutions can only change the 5' side of the YGCY binding site unlike the CUG repeats. Second, due to the double cytosine in CCUG repeats, it is possible for CCUG repeats to form helices in two different registers. CCUG repeat helices in register one have C-U and U-C mismatched base pairs, and adjacent G-C base pairs, helices in register two have U-U and C-C mismatches and spaced G-C base pairs

(Figure 14B). These RNAs were designed to determine if the different CCUG registers have the same degree of structural stability with Ψ modification. Unmodified $(CCUG)_6$ in register one melted at 45 ± 2°C, while $(CCUG)_6$ in register one with Ψ substitutions melted at 59 ±1°C. Similarly, unmodified $(CCUG)_6$ in register two melted at 46 ± 2°C and Ψ substitutions increased the melting temperature to 60 ± 1°C (Figure 14C, 14D and Table 6). These data showed that Ψ increased the thermal stability in CCUG repeats and the register didn't appear to make a difference.

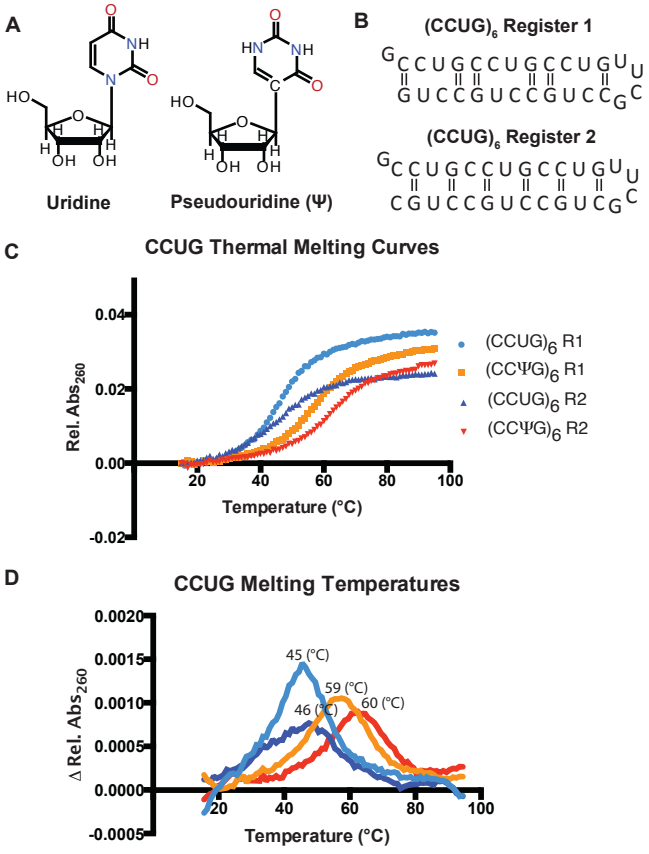


Figure 14: Pseudouridine (Ψ) increases thermal stability of CCUG repeat RNA. (A) representative images of uridine and Ψ. (B) Sequence and predicted secondary structures of the two different registers of the $(CCUG)_6$ repeat helices used in thermal melt and gel-shift assays.. (C) Thermal melt curves of $(CCUG)_6$ in both registers either unmodified or fully pseudouridylated. (D) Derivatives of thermal melt curves give the melting temperature of each construct. The largest derivative, or the steepest point on the curve, is the melting temperature.

MBNL has a reduced affinity for CCUG repeats modified with Ψ

Gel shift assays were used to determine if the Ψ stabilization observed in the $(CCUG)_6$ RNAs inhibited MBNL1 binding. The dissociation constant (K_d) for MBNL1 binding to unmodified $(CCUG)_6$ in register one is 0.16 ± 0.03 μM, modification of uridines to Ψ in $(CC\Psi G)_6$ resulted in an increased K_d of 0.39 ± 0.08 μM. The measured dissociation constants for $(CCUG)_6$ and $(CC\Psi G)_6$ in register two were similar to the register one RNAs, with K_ds of 0.12 ± 0.03 μM and 0.47 ± 0.09 μM, respectively (Fig 15A,B and Table 6). These results showed that Ψ had a modest inhibitory effect on MBNL1 binding to the YGCY binding sites in the CCUG repeats. This effect is significantly less than that observed with Ψ modification of CUG repeats[20,25].

MBNL has reduced affinity for a single-stranded RNA modified with Ψ

The model MBNL1 RNA substrate $[U_4(GC)U_{11}(GC)U_4]$ with little RNA structure as previously demonstrated[59], was used to determine if Ψ modification would inhibit MBNL1's ability to bind ssRNA. The Ψ modification was placed 5' or 3' of the GC dinucleotide and flanking both GC dinucleotides in the $U_4(GC)U_{11}(GC)U_4$ RNA (Figure 16). The dissociation constant of MBNL1 binding to $U_4(GC)U_{11}(GC)U_4$ (unmodified RNA) is 0.14 ± 0.09 μM. The addition of a Ψ modification 5' to the GC increased the K_d to 0.83 ± 0.15 μM, while Ψ placement 3' to the GC dinucleotide increased the K_d to 3.2 ± 0.66 μM, respectively (Fig. 16A,B and Table 7). When four pseudouridines flank the two GC dinucleotides, the fraction of RNA bound does not reach 0.5 in these assays, so we can only

Table 6: Melthing temperatures and MBNL affinities of $(CCUG)_6$ RNAs

RNA	T_M (°C)	K_D (μM)
$(CCUG)_6$ R1	45 ± 2	0.16 ± 0.03
$(CC\Psi G)_6$ R1	59 ± 1	0.39 ± 0.08
$(CCUG)_6$ R2	46 ± 2	0.12 ± 0.03
$(CC\Psi G)_6$ R2	60 ± 1	0.47 ± 0.09

state that the dissociation constant is > 3.2 µM (Figure 16A). Binding curves used to calculate these values are shown in Fig. 16B. These results show that MBNL1 was generally inhibited by Ψ incorporation; the nucleotide 3' to the GC dinucleotide had a more pronounced affect on MBNL1 binding than the 5' position as the K_d for the RNA with the 3' Ψ modifications is ~4 fold greater than for the RNA with 5' modifications

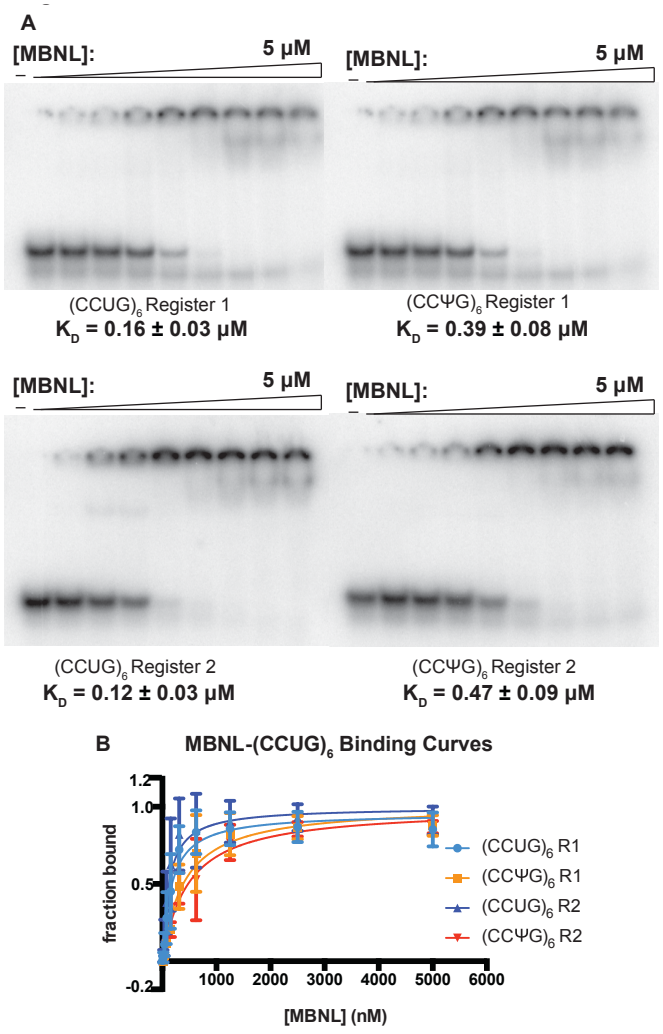


Figure 15: MBNL1 has reduced affinity for pseudouridylated $(CCUG)_6$. (A) Representative binding gels of MBNL1-$(CCUG)_6$ RNAs with Ψ modification and without modification. MBNL1 concentration increased in 2-fold steps from 20 nM to 5 µM. K_d values for RNAs are reported below the representative gels. (B) Binding curves for unmodified and Ψ modified $(CCUG)_6$ RNAs

(Figure 16).

Ψ increases the thermal stability of single-stranded YGCY RNAs

Melting studies were performed to determine if the Ψ modifications would increase the thermal stability of the MBNL1 model ssRNA substrate. Previous reports have shown that Ψ induced base-stacking interactions in single-stranded RNAs predicting that increased base-stacking might be observed [49]. $U_4(GC)U_{11}(GC)U_4$ melted at a

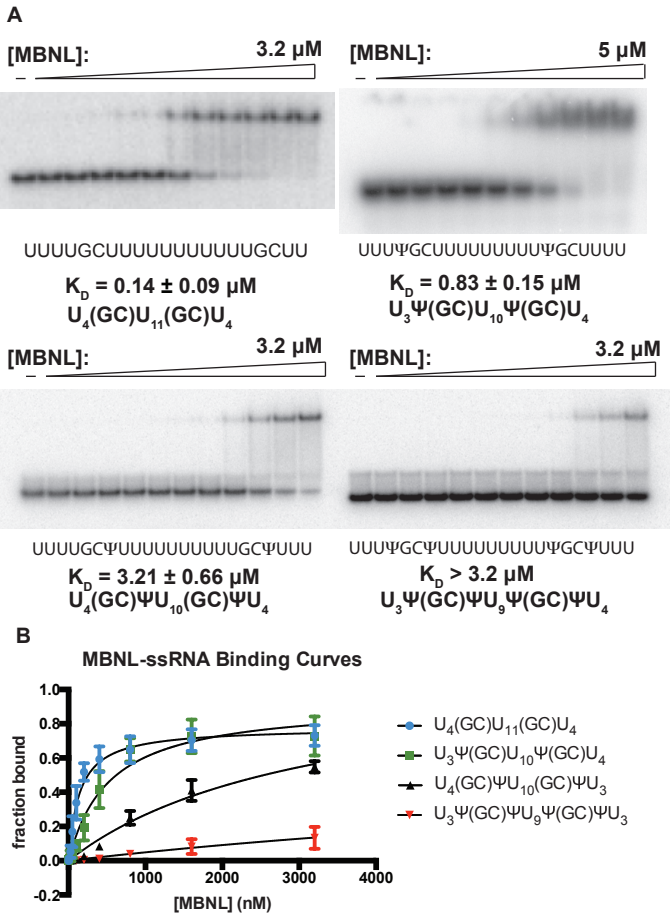


Figure 16: MBNL1 has a reduced affinity for single-stranded RNA containing pseudouridines. (A) Binding gels of MBNL1-$U_4(GC)U_{11}(GC)U_4$ RNAs with increasing number of pseudouridines. MBNL1 concentration increased in 2-fold steps from 6.25 nM to 3.2 µM except for the gel in the upper right panel where MBNL1 concentration ranged from 9.8 nM to 5 µM. K_d values for RNAs are reported below the representative gels. (B) Binding curves for $U_4(GC)U_{11}(GC)U_4$ RNA and Ψ modified RNAs.

temperature of 29 ± 2°C, which is presumably the un-stacking of the bases in a primarily single-stranded structure because there is little possibility of secondary structure with only two guanosines in a long string of pryrimidines. The RNA with the two pseudouridines 5' of the GC dinucleotides melted at 42 ± 4°C. The RNA containing pseudouridines 3' to the dinucleotide melted at 40 ± 2°C. The RNA containing Ψ substitutions flanking the GC dinucleotides melted at 51 ± 1°C (Figure 17A,B and Table 7). As shown in figure 17B, melting temperatures are calculated by taking the derivative of the melting curve, and the steepest (largest derivative) point is reported. These data indicate that Ψ stabilized this single-stranded RNAs by increasing the base stacking in the single-stranded RNA. Consistent with this interpretation is that additional Ψ substitutions increased the melting temperature (Figure 17B). An increase in thermal stability of these ssRNAs is likely due to strand rigidifying via base-stacking interactions.

Table 7: Melting temperatures and MBNL affinities of ssRNA

RNA	T_M (°C)	K_D (μM)
$U_4(GC)U_{11}(GC)U_4$	29 ± 2	0.14 ± 0.09
$U_3Ψ(GC)U_{10}Ψ(GC)U_4$	42 ± 4	0.83 ± 0.15
$U_4(GC)ΨU_{10}(GC)ΨU_3$	40 ± 2	3.21 ± 0.66
$U_3Ψ(GC)ΨU_9Ψ(GC)ΨU_3$	51 ± 1	N/A

Molecular dynamics simulations indicate Ψ increases base-stacking in single-stranded YGCY RNAs

We conducted MD simulations to study whether Ψ induces increased base-stacking interactions in a fully atomistic explicit solvent model single-stranded YGCY RNA. MD simulations allowed us to computationally investigate the dynamics of single-

stranded $U_5(GC)U_5$, $U_4\Psi(GC)U_5$, $U_5(GC)\Psi U_4$, and $U_4\Psi(GC)\Psi U_4$ RNA. There were many differences in the conformational ensemble of the native U versus the pseudouridylated constructs, overall the pseudouridylated structures were much more compact and stable than the ensemble of native U structures. In order to quantify base-stacking interactions, we calculated the relative free energy of the distance between the U/Ψ5-G and the G-C6 bases throughout the simulation (Figure 18A,B). We define the distance between the bases to be the distance between the C2 atom on the G base and the C2 (C4) atom on the U (Ψ) base

U5-G bases had a global minimum in the stacked configuration but a very small

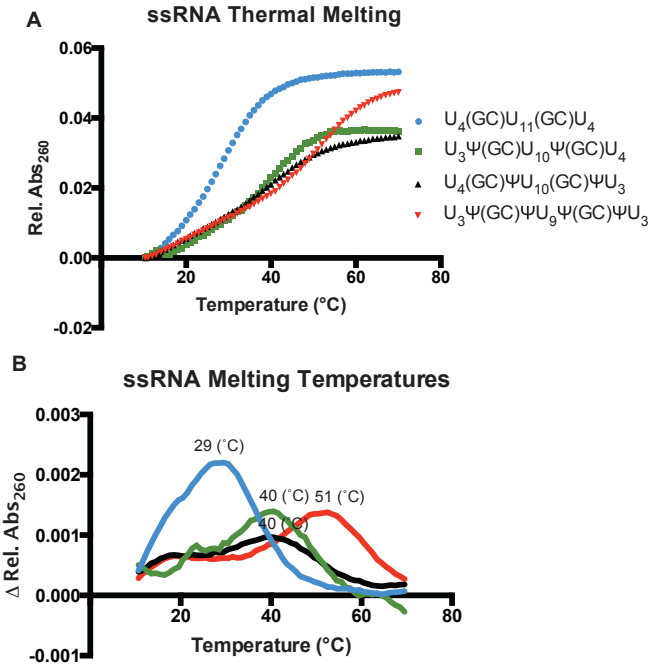


Figure 17: Pseudouridine increases the thermal stability of single-stranded YGCY RNAs. (A) Melting curves for $U_4(GC)U_{11}(GC)U_4$ RNA and modified RNA with Ψ as depicted in the legend. Absorbance at 260 nm was measured over the temperature range of 10-70°C. (B) Derivatives of the thermal melt curves give the melting temperature of

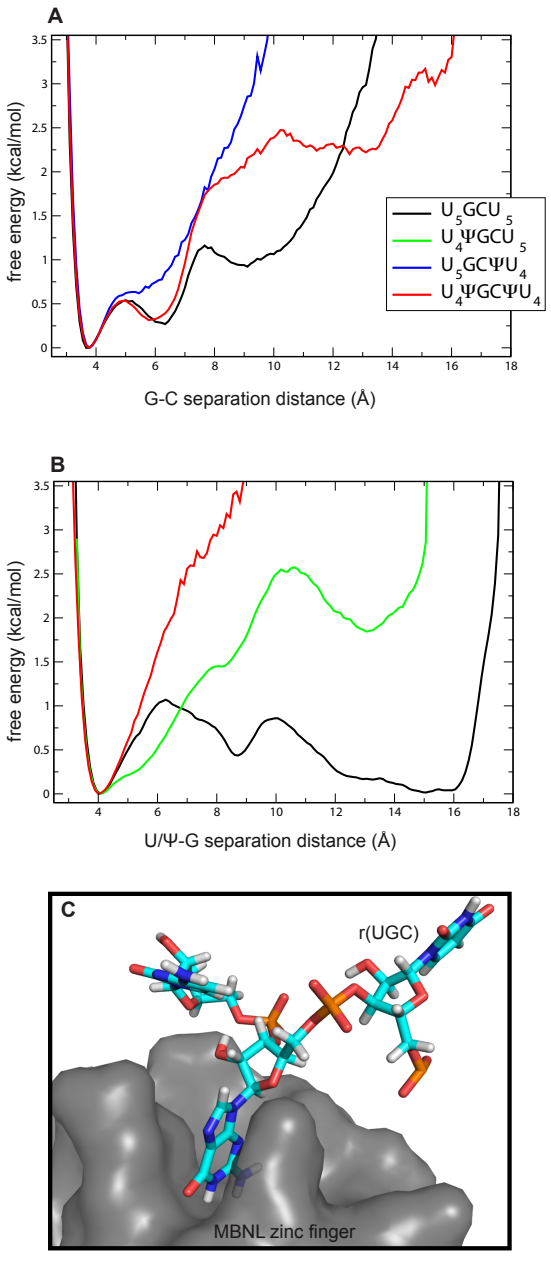


Figure 18: Molecular dynamics simulations indicate Ψ increases base-stacking in single-stranded YGCY RNAs. (A) Free energy of constructs at different G-C separation distances in $U_5(GC)U_5$ modified with Ψ as depicted in the legend. Unmodified RNA has energy minima at larger G-C separation distances, while pseudouridylated RNA only has energy minima at small G-C separation distances. (B) Free energy of constructs at different U/Ψ-G separation distances. Unmodified RNA has energy minima over a large range of U/Ψ-G separation distances, while the pseudouridylated constructs only have favorable conformations at short U/Ψ-G separation distances. (C) A depiction of the G base stacking with MBNL1's amino acids when bound to the protein.

barrier of 1.0 kcal/mol separating a range of open states of nearly equal free energy to the stacked state. Overall the range of open conformations, defined as structures with a U/Ψ5-G base separation greater than 7 angstroms were about four times more probable than closed conformations (Figure 18B). The widespread distribution of base separation distances in the U_5GCU_5 simulation indicates that the bases are dynamic, without a strong preference for base-stacking. Furthermore, analysis of the ensemble of open structures of the U_5GCU_5 from the simulation show that G6 is commonly unstacked from both neighboring bases.

While some pseudouridylated constructs displayed slightly increased flexibility at small separations, open configurations with a stacking distance to the guanine base of greater than 7 angstroms were highly disfavored in comparison to the U_5GCU_5 construct. Pseudouridylation of the 5' U to G6 resulted in increased base-stacking with G6, while pseudouridylation of the 3' U7 alone led to increased base-stacking between G6 and C7. Pseudouridylation of both U5 and U7 led to increased base stacking on both sides of G6. These data suggest that Ψ increases base-stacking interactions in these single-stranded YGCY RNAs.

DISCUSSION

Replacing the uridines in CCUG repeats with pseudouridines increased the thermal stability of this RNA in two different registers (Figure 14) by 15°C. Surprisingly, this stabilization only resulted in a modest 3- or 4-fold inhibition of MBNL1 binding. Stabilization of CUG repeats with Ψ resulted in similar thermal stability (13°C or 19°C for two and four Ψ modifications)[25] but had a more dramatic effect on MBNL1 binding (20-fold and loss of binding)[25] compared to CCUG repeat binding. A possible explanation for the difference in MBNL1 binding observed between CUG and CCUG repeats with Ψ modification is that CCUG repeats are less stable compared to CUG repeats and the increased stability provided by the Ψ modification to the CCUG repeats wasn't sufficient to strongly inhibit MBNL1 binding. Another possibility is that modifying both pyrimidines of the YGCY motif is necessary for robust inhibition of MBNL1 binding.

The model ssRNA was used to probe the difference between the two pyrimidine positions in the YGCY motif. The modification studies with the ssRNA revealed that the Ψ substitution 3' to the GC dinucleotide had a much more pronounced effect on MBNL1 binding affinity compared to the 5' substitution (Figure 16). This difference in binding affinity doesn't appear to be due to thermal stability as the 5' and 3' modifications were similarly stabilized by Ψ modification (T_ms of 40°C, Figure 17 and Table 7). The difference in MBNL1 inhibition at the 5' and 3' pyrimidine positions may be due to contacts made by the Zinc fingers of MBNL1 with the 3' pyrimidine observed in the crystal structure[38]. The 3' Ψ modification likely adopts conformations with bases in stacked structures that are less compatible with MBNL1 binding. The double Ψ modification that nearly eliminated MBNL1 binding (Figure 16) again favors RNA conformations with single strand base stacking. These studies are consistent with the crystal structure that showed the zinc fingers of MBNL1 interact primarily with the bases in an un-stacked conformation.

Consistent with the melting studies, molecular dynamics simulations showed that Ψ increased the base-stacking interactions in single-stranded YGCY RNAs, particularly at the pyrimidine position 5' to the GC dinucleotide. While the structure published by Teplova et al in 2008 shows that the 3' uridine in the YGCY motif makes contact with the protein, this structure also shows that while the RNA is bound to the protein, the G of the GC dinucleotide stacks with the protein's amino acids (Figure 18C). These results impose a model in which MBNL1 is able to form interactions with YGCY RNA when it is flexible enough for the G to un-stack with its neighboring pyrimidine. At the U-G base separation distance of 10.7 angstroms as measured in the crystal structure of bound

MBNL1 and RNA, the difference in free energy between the native U and pseudouridylated constructs studied in the MD simulations was 1.9 kcal/mol. This is significant in a conformational selection model of binding especially when you consider this free energy penalty is for a single binding site, while the binding of MBNL1 $U_4(GC)U_{11}(GC)U_4$ interaction involves binding to two of the YGCY sites. The MD simulations indicate that Ψ causes the RNA to prefer conformations in which the G is stacked with its neighbor. Based on the distribution of base separation distances, unmodified RNA also tends to form stacking interactions, but the population of those conformations is much less abundant than the range of open configurations. These data suggest that Ψ stabilizes YGCY RNA in repeats and single-stranded form likely via a combination of base-stacking and hydrogen bonding interactions.

MATERIALS AND METHODS

Thermal melt assays

Thermal melts were carried out in 20 mM PIPES buffer at a pH of 7.0 and 150 mM NaCl. The buffer was de-gassed before diluting RNA to a concentration of 2 μ M. The temperature was raised by 1°C per minute over a range of 10-75°C for ssRNAs and 15-95°C for CCUG RNAs, and absorbance at 260 nm was monitored by a Cary UV/Vis spectrophotometer. Absorbance was normalized by subtracting the absorbance at the lowest temperature measured.

Protein expression and purification

GST-tagged MBNL1 (amino acids 2-260) was expressed in Arctic Express *E. coli* cells (Agilent Technologies™) grown to an OD of 0.4, induced with 25 mM IPTG and

shaken for 20 hours at 16 °C. GST-MBNL1 was initially purified over a Hi-TrapGST column on an FPLC. MBNL1 was then cleaved from GST by *PreScission Protease* and purified further by affinity chromatography over a Hi-TrapHeparin column. Both columns and the protease were made by GE Healthcare Life Sciences™. MBNL1 concentration was measured by a Bradford assay and the protein was stored in a buffer of 500 mM NaCl, 50% glycerol, 25 mM Tris at pH of 7.5 and 5 mM BME at –80 °C until needed.

Gel-shift binding assays

RNA synthesized by Dharmacon was radiolabeled with α-^{32}P phosphate with a polynucleotide kinase and stored at -20 °C in 20 mM Tris, pH 7.5. Prior to the binding reaction, CUG helices were diluted and folded by incubation at 95 °C in 20 mM Tris, pH 7.5, 150 mM NaCl, and 5 mM $MgCl_2$ for 2 min, followed by incubation on ice for 5 min. The RNA was then incubated with varying MBNL1 concentrations in 175 mM NaCl, 20 mM Tris, pH 7.5, 1 mM BME, 10% glycerol, 5 mM $MgCl_2$, 0.1 mg/mL heparin, 2 mg/mL BSA, 0.02% xylene cyanol, and 0.05% bromophenol blue. The incubation was 20 minutes at room temperature. RNA-protein complexes were separated from free RNA on a 6% native acrylamide gel. Complex formation was quantified by exposure on a phosphorimager screen followed by analysis with *ImageQuant* from GE Healthcare Life Sciences™. Affinity constants were then calculated with the following equation: $f_{bound} = f_{max}([MBNL1]/([MBNL1+K_d))$. The images of the gels were enhanced equally across each gel using Adobe Photoshop. No specific feature was enhanced.

Molecular dynamics simulations

Atomistic explicit solvent simulations of single-stranded $U_5(GC)U_5$, $U_5\Psi(GC)U_5$, $U_5(GC)\Psi U_5$, and $U_5\Psi(GC)\Psi U_5$ RNA constructs were performed using the GROMACS [88-91] molecular dynamics package on the local ACISS cluster at the University of Oregon, using an AMBER [92] force-field optimized for stem-loop RNA structures with modified force-field parameters for the Ψ base [93]. All simulations were performed in explicit solvent utilizing the spc/e water model and 150 mM NaCl with excess sodium ions to neutralize the system, Ewald summation for the electrostatics, standard force-field cutoffs and parameters. The RNA was initially in an A-form structure which was energy minimized after solvation. A 1 fs timestep was utilized in the simulation, which was equilibrated over the course of a 5 nanoseconds period. An additional 10ns simulation was performed in the production NVT ensemble at 300K, from which 10 random snapshots were taken to seed 10 production runs of 10ns each from which to collect statistics.

CHAPTER IV

CONCLUSIONS AND FUTURE DIRECTIONS

MBNL has a decreased affinity for CUG repeats that are structurally stabilized by pseudouridine and 2'-O-methyl groups

We used structure-stabilizing RNA modifications to address our model of the mode of MBNL binding to its targets brought about by the structure published by Teplova et al in 2008. While there is abundant evidence that CUG and CCUG repeats form helices[31-33], this structure shows that MBNL accesses its YGCY binding site through interaction with the Watson-Crick face of the GC dinucleotide, which is taken up in intra-strand interactions in helices[38]. Thus, RNA modifications that stabilize helical structure should inhibit MBNL binding by preventing it access to its YGCY binding sites in CUG and CCUG repeats. As described in chapter II, Ψ and 2'-O-methylation have been shown to stabilize RNA secondary structure through various mechanisms[46-48].

To determine whether these modifications have stabilizing effects on CUG repeats, we used thermal-melt assays on $(CUG)_4$ RNAs with one, two and four of each different modification (Figure 6B). As expected, based on previous structural data regarding these modifications[42,45], our thermal melt assays showed an elevation of melting temperature of the RNAs with modifications, indicating an increase in structural stability (Figure 6C). Consistent with the conclusion that these modifications increase structural stability, fewer modifications had less of an effect on melting temperature, while fully modified RNAs had the most elevated melting temperatures (Figure 6C, Table 1)[25].

After verifying that 2'-O-methylation and pseudouridylation stabilize CUG repeat helices, we asked whether this helical stabilization prevents MBNL binding as we predicted in our model. We used gel-shift assays to measure MBNL1's affinity for unmodified $(CUG)_4$ and the RNA with one, two and four of each modification. Pseudouridylation and 2'-O-methylation decreased MBNL1's affinity for the described $(CUG)_4$ RNAs (Figure 7)[25]. 2'-O-methylation has a more modest effect on both structural stability and MBNL1 affinity than pseudouridylation, indicating that the extent of stability corresponds with MBNL1's affinity, further supporting our model. Because 2'-O-methylation has only modest effects on CUG repeats structure and MBNL1 affinity, we focused primarily on Ψ modifications in subsequent studies.

MBNL has a decreased affinity for CCUG repeats and single-stranded YGCY RNA stabilized by pseudouridine

In chapter III, I describe how we continued our studies on how Ψ affects RNA structural stability and MBNL association, this time addressing CCUG repeats and single-stranded YGCY RNAs. While the bulk of research on myotonic dystrophy addresses CUG repeats that are the pathogenic molecule in myotonic dystrophy type 1, it is also important to gather information on CCUG repeats, which are pathogenic in myotonic dystrophy type 2[12]. Based on the model established by the study in chapter II, MBNL proteins interact with YGCY when it is single-stranded. Therefore, we expanded our investigation to include Ψ's affects on single-stranded YGCY RNAs.

Because Ψ has a stabilizing effect on CUG repeats, we predicted that it would also stabilize CCUG repeats[25]. The extra C in the tetranucleotide CCUG repeat sequence inflicts a few important features that make them different from CUG repeats. First, they

are capable of forming helices in two different registers (Figure 14B). Register one has adjacent G-C base pairs and C-U mismatches. Register two has separated G-C base pairs and U-U and C-C mismatches. Secondly, all YGCY motifs in CCUG repeats are flanked by a 5' U and a 3' C, so Ψ substitution can only be 5' to the GC dinucleotide (CC**U**G**C**CUG). We used $(CCUG)_6$ RNA in both registers to determine how Ψ affects its thermal stability. As expected, Ψ substitution did increase the thermal stability of CCUG repeats (Figure 14C,D, Table 6). Ψ's effect on CCUG repeat stability was not as robust as it was on the thermal stability of CUG repeats documented in chapter II. Following the establishment that Ψ stabilizes CCUG repeats in both registers, we measured MBNL1's affinity for pseudouridylated $(CCUG)_6$ in both registers. Here we saw that Ψ reduces MBNL1's affinity for CCUG repeats, as it does for CUG repeats described in chapter II (Figure 15 A,B, Table 6). However, MBNL1 has a higher affinity for fully pseudouridylated CCUG repeats than it does for fully pseudouridylated CUG repeats. This could be because CCUG repeat helices are less stable than CUG repeats, and fully pseudouridylated CCUG repeats still have enough single-stranded nature for MBNL1 to bind. Alternatively, modifying only the pyrimidine 5' to the GC dinucleotide might not be sufficient to prevent MBNL1 binding.

We also investigated Ψ's effects on MBNL1's ability to bind to single-stranded YGCY RNA ($U_4(GC)U_{11}(GC)U_4$) and its thermal stability. This RNA is predicted to be single-stranded, and was previously shown to be a very good MBNL1 substrate[59]. Replacing the pyrimidines both 5' and 3' to each GC dinucleotide severely hampered MBNL1 binding. When we measured MBNL1's affinity for this RNA with Ψ substitutions at only the 3' bases, we also saw a strong decrease in affinity (Figure 16

A,B, Table 7). However, when the Ψs are substituted only 5' to the GC dinucleotide MBNL1's affinity was 4 fold stronger than the former case. This could be due to interactions MBNL makes with the 3' pyrimidine. The structure produced by Teplova et al shows that the uridine 3' to the GC dinucleotide forms hydrogen bonds with MBNL1's zinc finger[38]. Ψ in this position in place of uridine could alter the conformation of the bases such that it is unable to form those hydrogen bonds.

To determine if MBNL1's lower affinity for single-stranded RNA with Ψ is due to increased thermal stability, we performed thermal melt assays on these RNAs and saw that Ψ increased the thermal stability of RNA. In these experiments, however, the position of the Ψ does not affect the change in melting temperature. Ψ substitutions elevate the melting temperature to the same extent whether 5' or 3' to the GC dinucleotide. Four substitutions had an even greater effect on the melting temperature of this single-stranded RNA (Figure 17 A,B, Table 7). In chapter II we showed how Ψ stabilizes secondary structure of CUG repeat helices, but here we see that RNA that doesn't form helices is also stabilized by Ψ and that also blocks MBNL1 binding. This implies that not only does MBNL1 prefer to bind to single-stranded RNA, the binding site must also be flexible for optimum MBNL1 binding.

Pseudouridine stabilizes CUG repeats, CCUG repeats and single-stranded YGCY RNA through a combination of mechanisms

In light of the structure-stabilizing capabilities exhibited by Ψ, we were curious as to how it stabilizes helical and single-stranded RNA. Ψ has previously been shown to stabilize RNA through base-stacking interactions, which would explain how it can induce thermal stability of single-stranded RNA[49]. It is also known to stabilize RNA via

hydrogen-bonding interactions by incorporating water bridges between base pairs[46]. We investigated the structure and dynamics of RNAs containing Ψ to address its mechanism of stabilization.

To determine how Ψ affects the structure of CUG repeat helices, we crystallized a CUG repeat RNA containing a Ψ substitution (Figure 8A). We used an RNA with four CUG repeats attached to a GAAA tetraloop and its receptor to facilitate crystallization, which we had previously used to obtain the structure of native CUG repeats[33]. Initially it appeared that the Ψ substitution had very little affect on the helical structure of CUG repeats, however further inspection indicated a water molecule bridging the Ψ-U mismatch in the crystal structure (Figure 8 B,C, Tables 3,4). Molecular dynamics simulations on CUG repeat helices with Ψ substitutions allowed us to investigate water coordination and stability of U-U mismatches in a simulated solution. These simulations showed that Ψ substitutions in the U-U mismatch induced a higher degree of water coordination, which also increased the energy required to break the base pair (Figure 8 D-F). We determined that the extra hydrogen bonds present in the Ψ-U mismatch with a water bridge could account for Ψ's stabilizing effects.

Since Ψ is able to stabilize single-stranded RNA, and single-stranded RNA has no base pairs, Ψ must act through an additional stabilizing mechanism. Because Ψ has been shown to induce base-stacking interactions, we developed a method of measuring base-stacking through molecular dynamics simulations. To do this, we calculated the relative free energy of the distance between the G-C and U/Ψ5-G throughout simulations on U_5GCU_5 with 5' and 3' Ψ substitutions. Molecules with Ψ substitutions tended to stay in a state in which the Ψ stacks with the G, while unmodified molecules adopted many

conformations with varying distances between the U and the G. Unmodified YGCY RNA was also able to adopt more conformations with larger G-C separation distances than RNAs with Ψ (Figure 18 A,B). These data indicate that Ψ induces base-stacking interactions in single-stranded YGCY RNAs. This provides evidence that Ψ stabilizes CUG and CCUG repeat helices via a combination of increased base-stacking and water molecule coordination.

CUG repeats modified with pseudouridine lack the ability to abrogate MBNL-regulated alternative splicing and myotonic dystrophy phenotypes

Evidence supporting our model for how MBNL binds CUG repeat helices was established *in vitro*, however I was curious as to whether stabilization of CUG repeat helices would affect MBNL's function *in vivo*. To address this, I used our established splicing assay to determine if MBNL is free to regulate alternative splicing in the presence of pseudouridylated repeats. When HeLa cells are transfected with plasmids containing long CTG repeats, the cells express long CUG repeats, which form foci and sequester MBNL (Figure 10A, upper right panel). We then extract cellular RNA and measure exon inclusion levels to determine MBNL's extent of regulation. In order to test whether pseudouridylated CUG repeats affect MBNL-regulated alternative splicing, I *in vitro* transcribed CUG repeats with varying levels of Ψ and transfected them into HeLa cells. Unmodified *in vitro* transcribed CUG repeats disrupted MBNL-regulated alternative splicing of its targets *TNNT2* and *INSR*, producing exon inclusion levels similar to the CTG plasmid control. Fully pseudouridylated *in vitro*-transcribed CUG repeats did not disrupt alternative splicing of these transcripts, and exon inclusion levels

were similar to wild-type (Figure 9). This data provided evidence that the model established *in vitro* functions *in vivo* as well.

Our collaborator Dr. Peter Todd recently developed a myotonic dystrophy model zebrafish, which provided us with an opportunity to further test this model *in vivo.* When zebrafish embryos are injected with long CUG repeat RNA, they develop muscle function phenotypes and are less viable. Consistent with our observation that pseudouridylated CUG repeats do not obstruct MBNL regulated alternative splicing, zebrafish embryos injected with pseudouridylated CUG repeats did not develop the muscle function phenotype and were as viable as embryos injected with control RNA (Figure 11). These results, coupled with the alternative splicing results further establish that our model is relevant *in vivo*. Therefore pseudouridylated CUG repeats would likely be less toxic to myotonic dystrophy patients.

Possible therapeutic approaches to myotonic dystrophy through CUG and CCUG repeat structural stabilization

Efforts to develop therapeutic molecules to treat myotonic dystrophy patients have been approached from different angles. For example, we have identified small molecules that reduce levels of CUG repeats *in vivo*[102]. Molecules designed to disrupt MBNL binding to CUG and CCUG repeats have also been introduced as potential drug target molecules[35,56,72]. Another therapeutic approach that has been investigated is the development of antisense oligonucleotides that base pair with CUG repeats and block MBNL binding[71]. The results from my work present another possible therapeutic strategy for myotonic dystrophy patients. Introducing Ψ modifications to CUG and CCUG repeats in myotonic dystrophy patients would prevent MBNL sequestration and likely relieve

disease symptoms. In order to introduce Ψ modifications to CUG and CCUG repeats in patient tissues, we would design H/ACA box snoRNAs with target sequences that anneal to CUG and CCUG repeats. Once introduced to cells, the snoRNPs containing the designer snoRNA and pseudouridine synthase would assemble on the CUG or CCUG repeats and convert uridines to Ψ. Advances in delivery systems has made RNA's use as a therapeutic a plausible treatment mechanism, so a snoRNA-based therapeutic is conceivable[103].

REFERENCES CITED

(1) Ezkurdia, I., Juan, D., Rodriguez, J. M., Frankish, A., Diekhans, M., Harrow, J., Vazquez, J., Valencia, A., and Tress, M. L. (2014) Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. *Human Molecular Genetics 23*, 5866–5878.

(2) Rieder, L. E., and Reenan, R. A. (2012) The intricate relationship between RNA structure, editing, and splicing. *Semin. Cell Dev. Biol. 23*, 281–288.

(3) Jurica, M. S., and Moore, M. J. (2003) Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell 12*, 5–14.

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

(5) Schmucker, D., Clemens, J. C., Shu, H., Worby, C. A., Xiao, J., Muda, M., Dixon, J. E., and Zipursky, S. L. (2000) Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell 101*, 671–684.

(6) Black, D. L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem. 72*, 291–336.

(7) Gates, D. P., Coonrod, L. A., and Berglund, J. A. (2011) Autoregulated splicing of muscleblind-like 1 (MBNL1) Pre-mRNA. *Journal of Biological Chemistry 286*, 34224–34233.

(8) Zhang, C., Lee, K.-Y., Swanson, M. S., and Darnell, R. B. (2013) Prediction of clustered RNA-binding protein motif sites in the mammalian genome. *Nucleic Acids Research 41*, 6793–6807.

(9) Rinn, J. L., and Chang, H. Y. (2012) Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem. 81*, 145–166.

(10) Tazi, J., Bakkour, N., and Stamm, S. (2009) Alternative splicing and disease. *Biochim. Biophys. Acta 1792*, 14–26.

(11) Machuca-Tzili, L., Brook, D., and Hilton-Jones, D. (2005) Clinical and molecular aspects of the myotonic dystrophies: A review. *Muscle Nerve 32*, 1–18.

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

(13) Ranum, L. P. W., and Day, J. W. (2004) Myotonic dystrophy: RNA pathogenesis comes into focus. *Am. J. Hum. Genet. 74*, 793–804.

(14) Taneja, K. L., McCurrach, M., Schalling, M., Housman, D., and Singer, R. H. (1995) Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. *J. Cell Biol. 128*, 995–1002.

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

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

(17) Wang, E. T., Cody, N. A. L., Jog, S., Biancolella, M., Wang, T. T., Treacy, D. J., Luo, S., Schroth, G. P., Housman, D. E., Reddy, S., Lécuyer, E., and Burge, C. B. (2012) Transcriptome-wide Regulation of Pre-mRNA Splicing and mRNA Localization by Muscleblind Proteins. *Cell 150*, 710–724.

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

(19) Kino, Y. (2004) Muscleblind protein, MBNL1/EXP, binds specifically to CHHG repeats. *Human Molecular Genetics 13*, 495–507.

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

(21) Mankodi, A., Takahashi, M. P., Jiang, H., Beck, C. L., Bowers, W. J., Moxley, R. T., Cannon, S. C., and Thornton, C. A. (2002) Expanded CUG repeats trigger aberrant splicing of ClC-1 chloride channel pre-mRNA and hyperexcitability of skeletal muscle in myotonic dystrophy. *Mol. Cell 10*, 35–44.

(22) Charlet-B, N., Savkur, R. S., Singh, G., Philips, A. V., Grice, E. A., and Cooper, T. A. (2002) Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing. *Mol. Cell 10*, 45–53.

(23) de Die-Smulders, C. E., Höweler, C. J., Thijs, C., Mirandolle, J. F., Anten, H. B., Smeets, H. J., Chandler, K. E., and Geraedts, J. P. (1998) Age and causes of death in adult-onset myotonic dystrophy. *Brain 121 (Pt 8)*, 1557–1563.

(24) Groh, W. J., Groh, M. R., Saha, C., Kincaid, J. C., Simmons, Z., Ciafaloni, E., Pourmand, R., Otten, R. F., Bhakta, D., Nair, G. V., Marashdeh, M. M., Zipes, D. P., and Pascuzzi, R. M. (2008) Electrocardiographic abnormalities and sudden death in myotonic dystrophy type 1. *N. Engl. J. Med. 358*, 2688–2697.

(25) deLorimier, E., Coonrod, L. A., Copperman, J., Taber, A., Reister, E. E., Sharma, K., Todd, P. K., Guenza, M. G., and Berglund, J. A. (2014) Modifications to toxic CUG RNAs induce structural stability, rescue mis-splicing in a myotonic dystrophy cell model and reduce toxicity in a myotonic dystrophy zebrafish model. *Nucleic Acids Research 42*, 12768–12778.

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

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

(28) Kimura, T., Nakamori, M., Lueck, J. D., Pouliquin, P., Aoike, F., Fujimura, H., Dirksen, R. T., Takahashi, M. P., Dulhunty, A. F., and Sakoda, S. (2005) Altered mRNA splicing of the skeletal muscle ryanodine receptor and sarcoplasmic/endoplasmic reticulum Ca2+-ATPase in myotonic dystrophy type 1. *Human Molecular Genetics 14*, 2189–2200.

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

(30) Kino, Y., Washizu, C., Oma, Y., Onishi, H., Nezu, Y., Sasagawa, N., Nukina, N., and Ishiura, S. (2009) MBNL and CELF proteins regulate alternative splicing of the skeletal muscle chloride channel CLCN1. *Nucleic Acids Research 37*, 6477–6490.

(31) Napierała, M., and Krzyzosiak, W. J. (1997) CUG repeats present in myotonin kinase RNA form metastable “slippery” hairpins. *J. Biol. Chem. 272*, 31079–31085.

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

(33) Coonrod, L. A., Lohman, J. R., and Berglund, J. A. (2012) Utilizing the GAAA Tetraloop/Receptor To Facilitate Crystal Packing and Determination of the Structure of a CUG RNA Helix. *Biochemistry 51*, 8330–8337.

(34) Lietzke, S. E., Barnes, C. L., Berglund, J. A., and Kundrot, C. E. (1996) The structure of an RNA dodecamer shows how tandem U-U base pairs increase the range of stable RNA structures and the diversity of recognition sites. *Structure 4*, 917–930.

(35) Childs-Disney, J. L., Yildirim, I., Park, H., Lohman, J. R., Guan, L., Tran, T., Sarkar, P., Schatz, G. C., and Disney, M. D. (2014) Structure of the myotonic dystrophy type 2 RNA and designed small molecules that reduce toxicity. *ACS Chem. Biol. 9*, 538–550.

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

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

(38) Teplova, M., and Patel, D. J. (2008) Structural insights into RNA recognition by the alternative-splicing regulator muscleblind-like MBNL1. *Nature Structural & Molecular Biology 15*, 1343–1351.

(39) Helm, M., and Alfonzo, J. D. (2014) Posttranscriptional RNA Modifications: playing metabolic games in a cell's chemical Legoland. *Chem. Biol. 21*, 174–185.

(40) Rezgui, V. A. N., Tyagi, K., Ranjan, N., Konevega, A. L., Mittelstaet, J., Rodnina, M. V., Peter, M., and Pedrioli, P. G. A. (2013) tRNA tKUUU, tQUUG, and tEUUC wobble position modifications fine-tune protein translation by promoting ribosome A-site binding. *Proceedings of the National Academy of Sciences 110*, 12289–12294.

(41) Limbach, P. A., Crain, P. F., and McCloskey, J. A. (1994) Summary: the modified nucleosides of RNA. *Nucleic Acids Research 22*, 2183–2196.

(42) Charette, M., and Gray, M. W. (2000) Pseudouridine in RNA: what, where, how, and why. *IUBMB life 49*, 341–351.

(43) Micura, R., Pils, W., Höbartner, C., Grubmayr, K., Ebert, M. O., and Jaun, B. (2001) Methylation of the nucleobases in RNA oligonucleotides mediates duplex-hairpin conversion. *Nucleic Acids Research 29*, 3997–4005.

(44) Yu, Y. T., Shu, M. D., and Steitz, J. A. (1998) Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing. *EMBO J. 17*, 5783–5795.

(45) Ji, L., and Chen, X. (2012) Regulation of small RNA stability: methylation and beyond. *Cell Res. 22*, 624–636.

(46) Durant, P. C., and Davis, D. R. (1999) Stabilization of the anticodon stem-loop of tRNALys,3 by an A+-C base-pair and by pseudouridine. *J. Mol. Biol. 285*, 115–131.

(47) Yarian, C. S., Basti, M. M., Cain, R. J., Ansari, G., Guenther, R. H., Sochacka, E., Czerwinska, G., Malkiewicz, A., and Agris, P. F. (1999) Structural and functional roles of the N1-and N3-protons of Ψ at tRNA's position 39. *Nucleic Acids Research 27*, 3543–3549.

(48) Kawai, G., Yamamoto, Y., Kamimura, T., Masegi, T., Sekine, M., Hata, T., Iimori, T., Watanabe, T., Miyazawa, T., and Yokoyama, S. (1992) Conformational rigidity of specific pyrimidine residues in tRNA arises from posttranscriptional modifications that enhance steric interaction between the base and the 2'-hydroxyl group. *Biochemistry 31*, 1040–1046.

(49) Davis, D. R. (1995) Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Research 23*, 5020–5026.

(50) Lui, L., and Lowe, T. (2013) Small nucleolar RNAs and RNA-guided post-transcriptional modification. *Essays Biochem. 54*, 53–77.

(51) Darzacq, X., Jády, B. E., Verheggen, C., Kiss, A. M., Bertrand, E., and Kiss, T. (2002) Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. *EMBO J. 21*, 2746–2756.

(52) Stepanov, G. A., Semenov, D. V., Kuligina, E. V., Rabinov, O. K. I., Kit, Y. Y., and Richter, V. A. (2012) Analogues of Artificial Human Box C/D Small Nucleolar RNA As Regulators of Alternative Splicing of a pre-mRNA Target. *Acta Naturae 4*, 32.

(53) Zhao, X., and Yu, Y.-T. (2007) Targeted pre-mRNA modification for gene silencing and regulation. *Nat Meth 5*, 95–100.

(54) O'Rourke, J. R., and Swanson, M. S. (2009) Mechanisms of RNA-mediated disease. *J. Biol. Chem. 284*, 7419–7423.

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

(56) Warf, M. B., Nakamori, M., Matthys, C. M., Thornton, C. A., and Berglund, J. A. (2009) Pentamidine reverses the splicing defects associated with myotonic dystrophy. *PNAS 106*, 18551–18556.

(57) Philips, A. V., Timchenko, L. T., and Cooper, T. A. (1998) Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. *Science 280*, 737–741.

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

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

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

(61) Kanadia, R. N., Shin, J., Yuan, Y., Beattie, S. G., Wheeler, T. M., Thornton, C. A., and Swanson, M. S. (2006) Reversal of RNA missplicing and myotonia after muscleblind overexpression in a mouse poly (CUG) model for myotonic dystrophy. *Proc. Natl. Acad. Sci. U.S.A. 103*, 11748–11753.

(62) Du, H., Cline, M. S., Osborne, R. J., Tuttle, D. L., Clark, T. A., Donohue, J. P., Hall, M. P., Shiue, L., Swanson, M. S., Thornton, C. A., and Ares, M. (2010) Aberrant alternative splicing and extracellular matrix gene expression in mouse models of myotonic dystrophy. *Nature Structural & Molecular Biology 17*, 187–193.

(63) Kiliszek, A., Kierzek, R., Krzyzosiak, W. J., and Rypniewski, W. (2011) Crystal structures of CGG RNA repeats with implications for fragile X-associated tremor ataxia syndrome. *Nucleic Acids Research 39*, 7308–7315.

(64) Kiliszek, A., Kierzek, R., Krzyzosiak, W. J., and Rypniewski, W. (2009) Structural insights into CUG repeats containing the “stretched U-U wobble”: implications for myotonic dystrophy. *Nucleic Acids Research 37*, 4149–4156.

(65) Newby, M. I., and Greenbaum, N. L. (2001) A conserved pseudouridine modification in eukaryotic U2 snRNA induces a change in branch-site architecture. *RNA 7*, 833–845.

(66) Ofengand, J., Bakin, A., Wrzesinski, J., Nurse, K., and Lane, B. G. (1995) The pseudouridine residues of ribosomal RNA. *Biochem. Cell Biol. 73*, 915–924.

(67) Newby, M. I., and Greenbaum, N. L. (2002) Investigation of Overhauser effects between pseudouridine and water protons in RNA helices. *Proc. Natl. Acad. Sci. U.S.A. 99*, 12697–12702.

(68) Todd, P. K., Ackall, F. Y., Hur, J., Sharma, K., Paulson, H. L., and Dowling, J. J. (2014) Transcriptional changes and developmental abnormalities in a zebrafish model of myotonic dystrophy type 1. *Dis Model Mech 7*, 143–155.

(69) Telfer, W. R., Busta, A. S., Bonnemann, C. G., Feldman, E. L., and Dowling, J. J. (2010) Zebrafish models of collagen VI-related myopathies. *Human Molecular Genetics 19*, 2433–2444.

(70) Brustein, E., Saint-Amant, L., Buss, R. R., Chong, M., McDearmid, J. R., and Drapeau, P. (2003) Steps during the development of the zebrafish locomotor network. *J. Physiol. Paris 97*, 77–86.

(71) Wheeler, T. M., Leger, A. J., Pandey, S. K., MacLeod, A. R., Nakamori, M., Cheng, S. H., Wentworth, B. M., Bennett, C. F., and Thornton, C. A. (2012) Targeting nuclear RNA for in vivo correction of myotonic dystrophy. *Nature 488*, 111–115.

(72) Childs-Disney, J. L., Hoskins, J., Rzuczek, S. G., Thornton, C. A., and Disney, M. D. (2012) Rationally designed small molecules targeting the RNA that causes myotonic dystrophy type 1 are potently bioactive. *ACS Chem. Biol. 7*, 856–862.

(73) Jahromi, A. H., Nguyen, L., Fu, Y., Miller, K. A., Baranger, A. M., and Zimmerman, S. C. (2013) A Novel CUGexp· MBNL1 Inhibitor with Therapeutic Potential for Myotonic Dystrophy Type 1. *ACS Chem. Biol. 8*, 1037–1043.

(74) Jacquemont, S., Hagerman, R. J., Leehey, M., Grigsby, J., Zhang, L., Brunberg, J. A., Greco, C., Portes, Des, V., Jardini, T., Levine, R., Berry-Kravis, E., Brown, W. T., Schaeffer, S., Kissel, J., Tassone, F., and Hagerman, P. J. (2003) Fragile X premutation tremor/ataxia syndrome: molecular, clinical, and neuroimaging correlates. *Am. J. Hum. Genet. 72*, 869–878.

(75) Day, J. W., Schut, L. J., Moseley, M. L., Durand, A. C., and Ranum, L. P. (2000) Spinocerebellar ataxia type 8: clinical features in a large family. *Neurology 55*, 649–657.

(76) Daughters, R. S., Tuttle, D. L., Gao, W., Ikeda, Y., Moseley, M. L., Ebner, T. J., Swanson, M. S., and Ranum, L. P. W. (2009) RNA gain-of-function in spinocerebellar ataxia type 8. *PLoS Genet. 5*, e1000600.

(77) Matsuura, T., Yamagata, T., Burgess, D. L., Rasmussen, A., Grewal, R. P., Watase, K., Khajavi, M., McCall, A. E., Davis, C. F., Zu, L., Achari, M., Pulst, S. M., Alonso, E., Noebels, J. L., Nelson, D. L., Zoghbi, H. Y., and Ashizawa, T. (2000) Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. *Nat Genet 26*, 191–194.

(78) Wakamiya, M., Matsuura, T., Liu, Y., Schuster, G. C., Gao, R., Xu, W., Sarkar, P. S., Lin, X., and Ashizawa, T. (2006) The role of ataxin 10 in the pathogenesis of spinocerebellar ataxia type 10. *Neurology 67*, 607–613.

(79) DeJesus-Hernandez, M., Mackenzie, I. R., Boeve, B. F., Boxer, A. L., Baker, M., Rutherford, N. J., Nicholson, A. M., Finch, N. A., Flynn, H., and Adamson, J. (2011) Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS. *Neuron 72*, 245–256.

(80) Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Macromolecular Crystallography, Pt A 276*, 307–326.

(81) Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003) A graphical user interface to the CCP4 program suite. *Acta crystallographica. Section D, Biological crystallography 59*, 1131–1137.

(82) Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta crystallographica. Section D, Biological crystallography 67*, 235–242.

(83) Vagin, A., and Teplyakov, A. (1997) MOLREP: an Automated Program for Molecular Replacement. *Journal of Applied Crystallography 30*, 1022–1025.

(84) Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography 66*, 486–501.

(85) Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta crystallographica. Section D, Biological crystallography 53*, 240–255.

(86) Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta crystallographica. Section D, Biological crystallography 54*, 905–921.

(87) Lu, X.-J., and Olson, W. K. (2008) 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nature Protocols 3*, 1213–1227.

(88) Berendsen, H. J., van der Spoel, D., and van Drunen, R. (1995) GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications 91*, 43–56.

(89) Lindahl, E., Hess, B., and van der Spoel, D. (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Molecular modeling annual 7*, 306–317.

(90) van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. C. (2005) GROMACS: fast, flexible, and free. *J. Comput. Chem. 26*, 1701–1718.

(91) Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput 4*, 435–447.

(92) Zgarbová, M., Otyepka, M., Sponer, J., Mládek, A., Banáš, P., Cheatham, T. E., and Jurečka, P. (2011) Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J Chem Theory Comput 7*, 2886–2902.

(93) Aduri, R., Psciuk, B. T., Saro, P., Taniga, H., Schlegel, H. B., and SantaLucia, J. (2007) AMBER force field parameters for the naturally occurring modified nucleosides in RNA. *J Chem Theory Comput 3*, 1464–1475.

(94) Bonomi, M., Branduardi, D., Bussi, G., Camilloni, C., Provasi, D., Raiteri, P., Donadio, D., Marinelli, F., Pietrucci, F., and Broglia, R. A. (2009) PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Computer Physics Communications 180*, 1961–1972.

(95) Laio, A., and Parrinello, M. (2002) Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A. 99*, 12562–12566.

(96) Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H., and Kollman, P. A. (1992) THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem. 13*, 1011–1021.

(97) Purcell, J., Oddo, J. C., Wang, E. T., and Berglund, J. A. (2012) Combinatorial mutagenesis of MBNL1 zinc fingers elucidates distinct classes of regulatory events. *Molecular and Cellular Biology 32*, 4155–4167.

(98) Newcombe, R. G. (1998) Two-sided confidence intervals for the single proportion: comparison of seven methods. *Stat Med 17*, 857–872.

(99) Dowling, J. J., Low, S. E., Busta, A. S., and Feldman, E. L. (2010) Zebrafish MTMR14 is required for excitation-contraction coupling, developmental motor function and the regulation of autophagy. *Human Molecular Genetics 19*, 2668–2681.

(100) Rau, F., Freyermuth, F., Fugier, C., Villemin, J.-P., Fischer, M.-C., Jost, B., Dembele, D., Gourdon, G., Nicole, A., Duboc, D., Wahbi, K., Day, J. W., Fujimura, H., Takahashi, M. P., Auboeuf, D., Dreumont, N., Furling, D., and Charlet-Berguerand, N. (2011) Misregulation of miR-1 processing is associated with heart defects in myotonic dystrophy. *Nature Structural & Molecular Biology 18*, 840–845.

(101) Ashwal-Fluss, R., Meyer, M., Pamudurti, N. R., Ivanov, A., Bartok, O., Hanan, M., Evantal, N., Memczak, S., Rajewsky, N., and Kadener, S. (2014) circRNA biogenesis competes with pre-mRNA splicing. *Mol. Cell 56*, 55–66.

(102) Coonrod, L. A., Nakamori, M., Wang, W., Carrell, S., Hilton, C. L., Bodner, M. J., Siboni, R. B., Docter, A. G., Haley, M. M., Thornton, C. A., and Berglund, J. A. (2013) Reducing levels of toxic RNA with small molecules. *ACS Chem. Biol. 8*, 2528–2537.

(103) Leng, Q., Woodle, M. C., Lu, P. Y., and Mixson, A. J. (2009) Advances in Systemic siRNA Delivery. *Drugs Future 34*, 721.