TARGETING MYOTONIC DYSTROPHY WITH SMALL MOLECULES

by

LESLIE A. COONROD

A DISSERTATION

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

June 2012

DISSERTATION APPROVAL PAGE

Student: Leslie A. Coonrod

Title: Targeting Myotonic Dystrophy with Small Molecules

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

Dr. Alice Barkan	Chairperson
Dr. J. Andrew Berglund	Advisor
Dr. Kenneth Prehoda	Member
Dr. Eric Johnson	Member
Dr. Victoria DeRose	Outside Member

and

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

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

Degree awarded June 2012

DISSERTATION ABSTRACT

Leslie Adrianne Coonrod

Doctor of Philosophy

Department of Biology

June 2012

Title: Targeting Myotonic Dystrophy with Small Molecules

Myotonic dystrophy (DM) is one of the most common forms of muscular dystrophy, characterized by its hallmark symptom myotonia. DM is an autosomal dominant disease caused by a toxic gain of function RNA. The toxic RNA is produced from expanded non-coding CTG/CCTG repeats, and these CUG/CCUG repeats sequester a family of RNA binding proteins. The Muscleblind-like (MBNL) family of RNA binding proteins are sequestered to the expanded CUG/CCUG repeats. The MBNL proteins are regulators of alternative splicing, and their sequestration to the toxic RNA leads to mis-splicing events, which are believed to cause the symptoms observed in DM patients.

A previously reported screen for small molecules used to identify compounds that could disrupt MBNL from binding the toxic CUG repeats found that pentamidine was able to rescue splicing defects associated with DM. Herein, we present a new class of molecules (phenolsulphonphthaleins) that inhibited MBNL1/CUG repeat complex formation in a competitive electrophoretic mobility shift assay (EMSA). Additionally, one of these molecules, bromophenol blue (BPB), acted in a synergistic manner with the previously described inhibitor pentamidine. We also demonstrated that the halogenation of the phenolsulphonphthalein dyes is an important factor for activity. Moreover, we present

an analysis of a series of methylene linker variants of pentamidine that revealed heptamidine (an analog of pentamidine) could reverse splicing defects in a DM1 tissue culture model and rescue myotonia in a DM1 mouse model.

Finally, we report on a new crystal structure of CUG repeats, crystallized in the context of a GAAA tetraloop/receptor which facilitated ordered packing within the crystal. This structure was consistent with previous structures showing that the repeats are essentially A-form RNA, despite having a U-U mismatch every third base pair. We also identified six types of U-U mismatch in the context of the 5'CUG/3'GUC motif, suggesting that the interactions between the uridines are dynamic. This structure also contains the highest resolution GAAA tetraloop/receptor structure (1.95 Å) reported to date.

This dissertation includes previously unpublished co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Leslie A. Coonrod

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, OR
Illinois College, Jacksonville, IL

DEGREES AWARDED:

Doctor of Philosophy, Biology, 2012, University of Oregon
Bachelor of Science, Chemistry, Mathematics, 2006, Illinois College

AREAS OF SPECIAL INTEREST:

Biochemistry
Molecular Biology

PROFESSIONAL EXPERIENCE:

Graduate Student, Dr. J. Andrew Berglund, 2006-2012

Undergraduate Research Assistant, Juanita B. Leonhard, 2005-2006

GRANTS, AWARDS, AND HONORS:

Howard Hughes Medical Institute Science Literacy Program Fellow, 2012

GK-12 Fellow, University of Oregon, 2008-2010

Trustee Scholarship, Illinois College, 2002-2006

Leo F. and Lorraine O. Heneghan Chemistry Prize, Illinois College, 2006

Earle B. Miller Award in Mathematics, Illinois College, 2004

PUBLICATIONS:

Gates, D. P., Coonrod, L. A., & Berglund, J. A. (2011). Autoregulated Splicing of muscleblind-like 1 (MBNL1) Pre-mRNA. *J. Biol Chem*, *286*(39), 34224–34233.

Coonrod, L.A., Lohman, J.R., and Berglund, J.A. Utilizing the GAAA tetraloop/receptor to facilitate crystal packing and determine the structure of a CUG RNA helix. (*Submitted to Biochemistry*)

ACKNOWLEDGMENTS

Firstly, I want to thank my advisor, Dr. Andy Berglund, for his advice and support throughout the years. I will always be grateful for your guidance during my time as a graduate student. I would also like to thank my committee for their invaluable input and guidance throughout my research. I am also appreciative of the funding from the National Science Foundation during my time as a GK-12 fellow, and the NIH grant to the lab.

My sincerest thanks to all the members of the Berglund lab, past and present. Our scientific discussions helped me to grow as a scientist, and our friendships helped me to grow as a person. I will never forget our time together. Equally important in helping me maintain my sanity were the friends I made outside of the lab who were always ready to take a walk and grab a coffee with me when I needed a break from an experiment that wasn't going well. Special thanks to Jen, Devika, Jamie, and Stacey. I don't know what I would have done without you.

Thanks to all the support staff here at the University of Oregon, from the secretaries who made sure I had all my paperwork filed, to the kitchen, who saved me untold amounts of time by washing dishes and making buffers. Thanks to Dan and Nigel, who kept equipment and computers running against all odds, and thanks to Ron for making sure I meant to order all that RNA. All of you helped smooth my progress through my work.

Finally, I want to thank my family. I would not be here without your support. Thanks to my parents, who always encouraged me, no matter what I wanted to do. Thanks to my brothers for teaching me patience and for keeping me entertained. Thanks to my in-laws, who have truly welcomed me into their family as a daughter. And of

course, my husband Evan, without whom I could not have done this. You have always believed in me and supported me, even when I told you I wanted to move 2000 miles across the country to go to graduate school. Thank you for all the times you listened when my experiments were not going well, and for all those late night trips into lab so I could change the media on my cells. I am so lucky to have married my best friend.

This dissertation is dedicated to my parents, who have always supported me in my quest for knowledge, and to my wonderful husband, who has always believed in me.

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

CHAPTER I

INTRODUCTION

Alternative splicing

In order to create a proteome of millions of different proteins from a genome made up of only 20,000-25,000 genes (*1*), humans employ differential processing of mRNAs. After transcription from DNA, but before translation to protein, a pre-mRNA must be processed. This process includes splicing, a molecular process that results in the removal of non-coding introns and fusion of exons to create mature mRNAs. In order for splicing to occur, the splicing machinery, known as the spliceosome, must recognize key motifs in the pre-mRNA that indicate a sequence is intronic, and thus removed, or exonic, and retained (*2*). Of course, not every pre-mRNA is processed in the same manner every time. If this were true, then a single gene would code for only a single protein, which we know is not the case. Most exons are constitutively spliced, that is, they are always included in the mature mRNA. However, there are many regulated exons that are not constitutively spliced, but rather alternatively spliced (*3*), resulting in multiple mRNA isoforms being derived from a single gene. In the extreme example of the *Drosophila* gene Dscam, there are 38,016 potential isoforms (*4*). Although this is an extreme example, it is estimated that more than 90% of human transcripts are alternatively spliced (*5*).

There are many different forms of alternative splicing (Figure 1). The most common type of alternative splicing event is the cassette exon, where an exon is either included or excluded from the final mRNA. There are also mutually exclusive exons, in which a choice between one of two exons is made. Exon length can also change by the use of

different 5’ or 3’ splice sites. Sometimes an intron will fail to be removed, leading to intron retention (*3*).

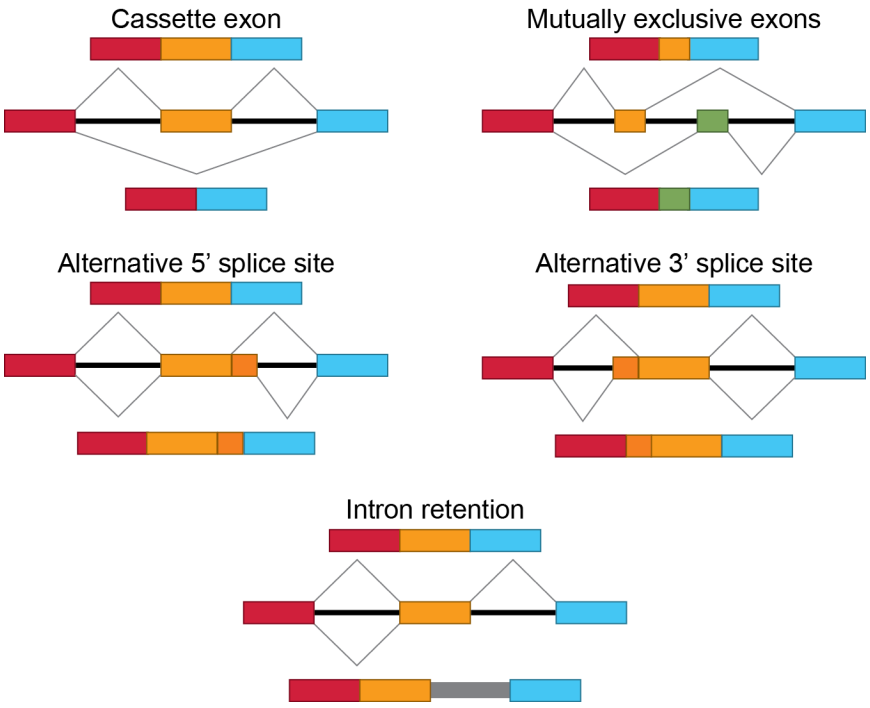


Figure 1. Types of alternative splicing. A cassette exon event will either include or exclude an exon, while in the case of mutually exclusive exons, either one or the other exon will be included. Alternative 5’ and 3’ splice sites will utilize an alternative 5’ or 3’ splice site of an exon, respectively. In some cases, introns can fail to be spliced out, leading to intron retention.

Alternative splicing is, necessarily, a highly regulated process involving a vast array of regulatory proteins and RNA motifs. Since different splice isoforms can result in proteins that have different biological activities, or that are targeted by nonsense-mediated decay due to the introduction of a premature stop codon, it is easy to envision

how errors in alternative splicing can result in disease. Some alternative splicing diseases are caused by cis-factors, mutations to the RNA motifs that function as splicing signals, while others are due to trans-factors, changes in the proteins, or the levels of proteins, that regulate splicing (*6*). One such disease, which affects levels of proteins involved in the regulation of alternative splicing, is myotonic dystrophy.

Myotonic dystrophy

Myotonic dystrophy (DM) is the most common form of adult onset muscular dystrophy, characterized by many different symptoms including myotonia, muscle wasting, cataracts, cardiac conduction defects, and cognitive dysfunction, among others (*7*). There are two types of DM: type 1 (DM1) and type 2 (DM2). DM1 is caused by an aberrant expansion of CTG repeats in the 3' untranslated region of the *DMPK* gene. Unaffected individuals have between 5 and 37 CTG repeats, while affected individuals have more than 50 (*8*). When transcribed, these expanded CUG repeats sequester RNA binding proteins in nuclear foci (Figure 2) (*9*, *10*). The DM2 disease mechanism is very similar, except instead of a trinucleotide repeat, it is a tetranucleotide repeat (CCTG) expansion in the first intron of the *ZNF9* gene (*11*).

One of the families of proteins sequestered by CUG/CCUG repeats is the Muscleblind-like (MBNL) family of proteins (*10*). In humans, there are three paralogs: MBNL1, MBNL2, and MBNL3. Within this body of work, I discuss studies focused primarily on MBNL1, since an MBNL1 knockout mouse has been shown to phenocopy DM (*12*). The MBNL proteins are zinc-finger RNA binding proteins that regulate

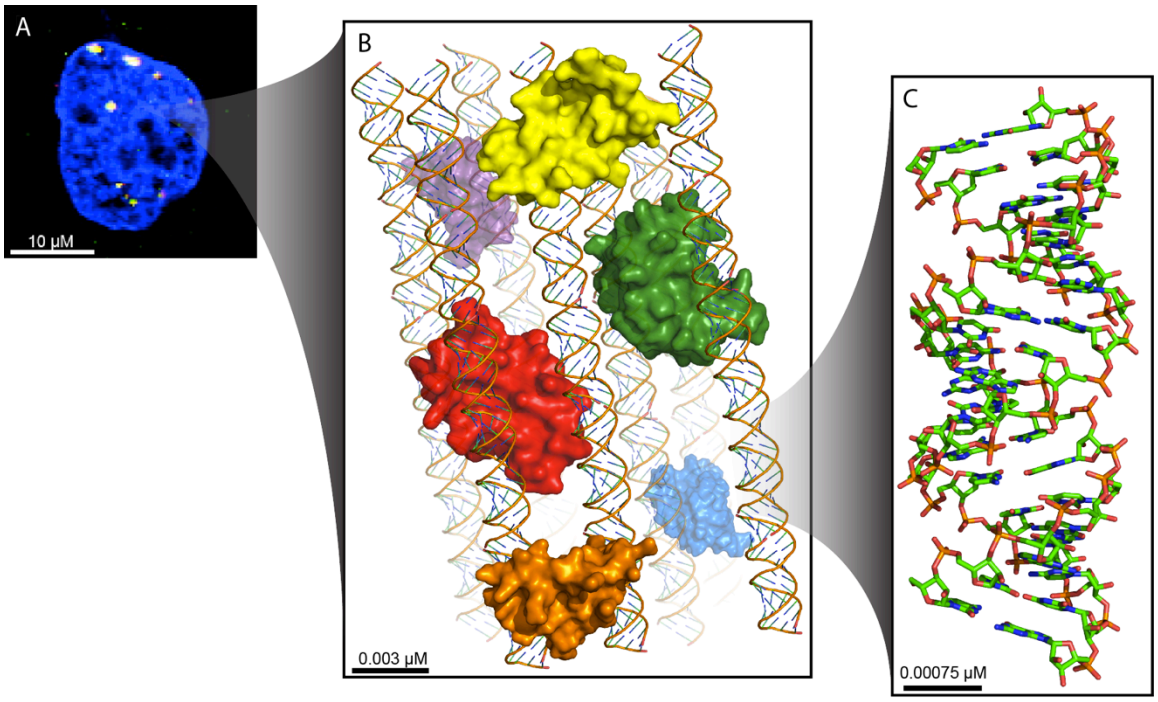
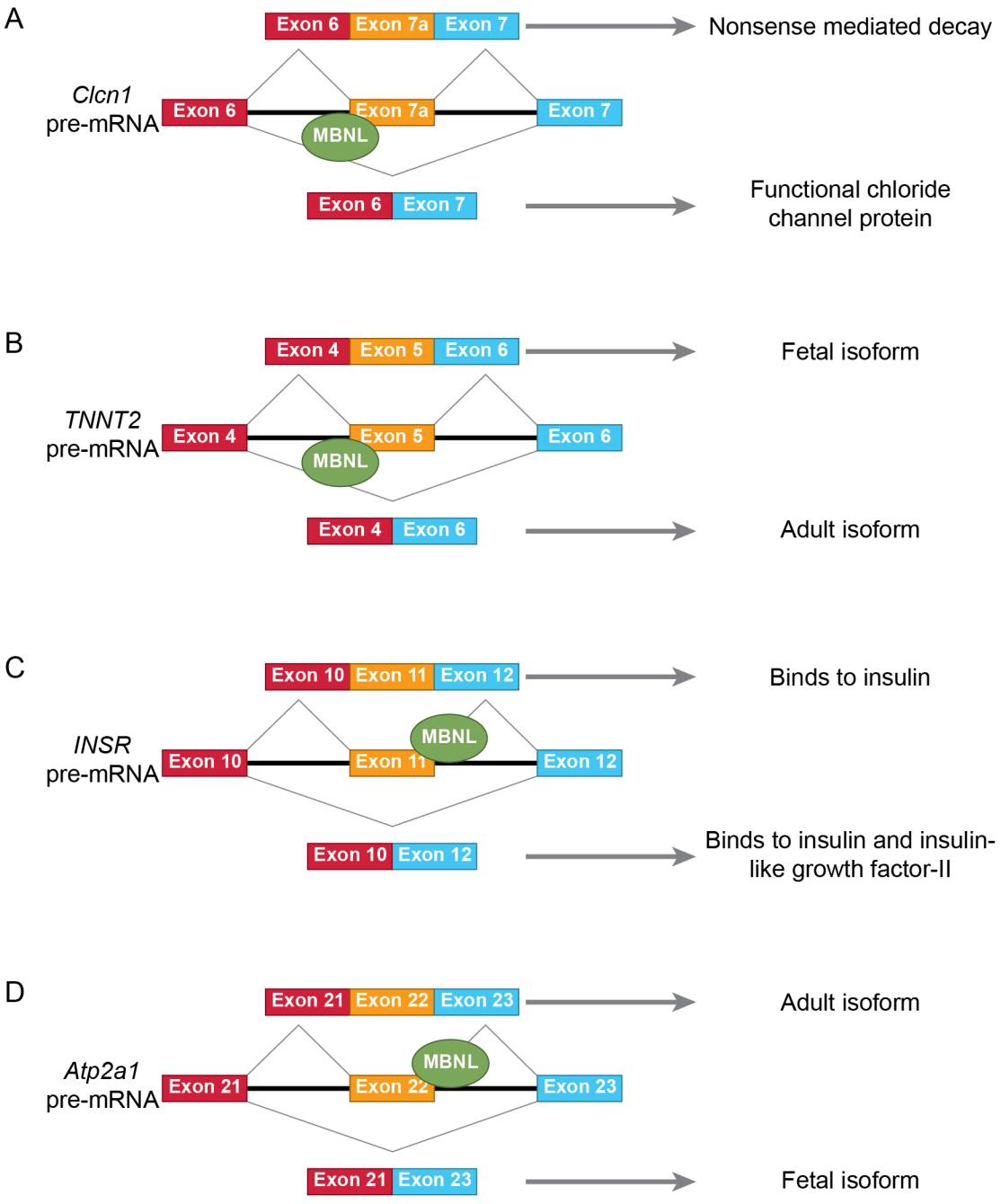


Figure 2. Characteristic nuclear foci observed in DM composed of RNA and proteins. (A) Nucleus of a HeLa cell expressing CUG repeats and MBNL1. Blue DAPI stains the DNA, and the foci are white. (B) Foci are composed of CUG repeat RNA and various RNA binding proteins, including MBNL. (C) Molecular structure of a duplex of six CUG repeats. The CUG repeats are essentially A-form in solution, despite having a U-U mismatch every third base pair.

alternative splicing (*13*). When sequestered to CUG/CCUG repeats, they are no longer able to properly regulate their target transcripts (*14-18*). Many of the mis-splicing events in DM have been found to be MBNL dependent, and some of these events have been correlated with disease symptoms. One target of the MBNL proteins is the *Clcn1* gene, which encodes a chloride ion channel. The MBNL proteins are negative regulators of exon 7a in *Clcn1* (when MBNL proteins are present, exon 7a is excluded). In the presence of CUG/CCUG repeats, the levels of free MBNL are not sufficient to properly regulate splicing of exon 7a (Figure 3A). This leads to chloride channelopathy and the characteristic DM symptom myotonia (*19*). Several other MBNL regulated mis-splicing events in DM have been well characterized, including *TNNT2* (*14*, *16*), *INSR* (*17*), and *Atp2a1* (*15*). MBNL promotes the exclusion of exon 5 in *TNNT2*, resulting in the expression of the adult isoform (Figure 3B). MBNL can also act as a splicing enhancer as it does in the cases of *INSR* (Figure 3C) and *Atp2a1* (Figure 3D).

Figure 3 (next page). Schematic of alternative splicing events in the presence and absence of MBNL. (A) MBNL mediated alternative splicing of *Clcn1*. When MBNL is present, exon 7a is skipped, resulting in a functional chloride channel protein. When MBNL is not present, perhaps due to sequestration by CUG/CCUG repeats, exon 7a is included. Exon 7a contains a premature stop codon, resulting in degradation of the transcript by the nonsense mediated decay pathway. (B) MBNL mediated alternative splicing of *TNNT2*. In the presence of MBNL, exon 5 is excluded, leading to the expression of the adult isoform of *TNNT2*; however, when MBNL is absent, exon 5 is skipped resulting in the fetal isoform. (C) MBNL mediated alternative splicing of *INSR*. MBNL binding promotes exon 11 inclusion, resulting in an insulin receptor that only binds to insulin, while exon 11 exclusion (in the absence of MBNL) produces an insulin receptor that can bind to both insulin and insulin-like growth factor-II. This contributes to the insulin resistance phenotype observed in DM patients. (D) MBNL mediated alternative splicing of *Atp2a1*. MBNL facilitates the inclusion of exon 22, leading to the adult isoform.


A
Exon 6
Exon 7a
Exon 7
Nonsense mediated decay
Clcn1
pre-mRNA
Exon 6
Exon 7a
Exon 7
MBNL
Exon 6
Exon 7
Functional chloride
channel protein
B
Exon 4
Exon 5
Exon 6
Fetal isoform
TNNT2
pre-mRNA
Exon 4
Exon 5
Exon 6
MBNL
Exon 4
Exon 6
Adult isoform
C
Exon 10
Exon 11
Exon 12
Binds to insulin
INSR
pre-mRNA
MBNL
Exon 10
Exon 11
Exon 12
Exon 10
Exon 12
Binds to insulin and insulin-
like growth factor-II
D
Exon 21
Exon 22
Exon 23
Adult isoform
Atp2a1
pre-mRNA
MBNL
Exon 21
Exon 22
Exon 23
Exon 21
Exon 23
Fetal isoform

Since many mis-splicing events have been shown to be MBNL dependent, one potential method of alleviating disease symptoms would be to release MBNL from the CUG/CCUG repeat foci, so that it could be free within the cell to perform its proper role in splicing regulation. There are many potential methods of displacing MBNL including utilizing small molecules (*20*-*23*), multivalent compounds (*24*-*26*), small peptides (*27*, *28*), or antisense oligonucleotide analogs (*29*). Because most FDA approved drugs are small molecules (*30*), and because there are many small molecule drugs that bind RNA (*31*), we decided to take a small molecule approach to relieving MBNL sequestration.

Pentamidine as a potential DM1 therapeutic

In order to identify small molecules that have the potential to disrupt MBNL/CUG complexes, we undertook a screen of ~30 molecules that are known to interact with nucleic acid. Using a competitive electrophoretic mobility shift assay (EMSA), we discovered a small molecule, pentamidine (a currently FDA approved drug for the treatment of *Pneumocystis carinii* infections), had the ability to disrupt the MBNL1/$(CUG)_4$ complex with an IC_{50} of 58 ± 5 μM (*20*) (Figure 4A). In addition to disrupting the complex *in vitro*, pentamidine was also able to reverse splicing defects (Figure 4B) and reduce the number the nuclear foci in a HeLa cell model of DM1 (Figure 4C). Furthermore, when mice that express 220 CUG repeats under a human skeletal actin promoter were treated with pentamidine, partial splicing rescue of two endogenous mRNAs, including *Clcn1* exon 7a, were observed (Figure 4D).

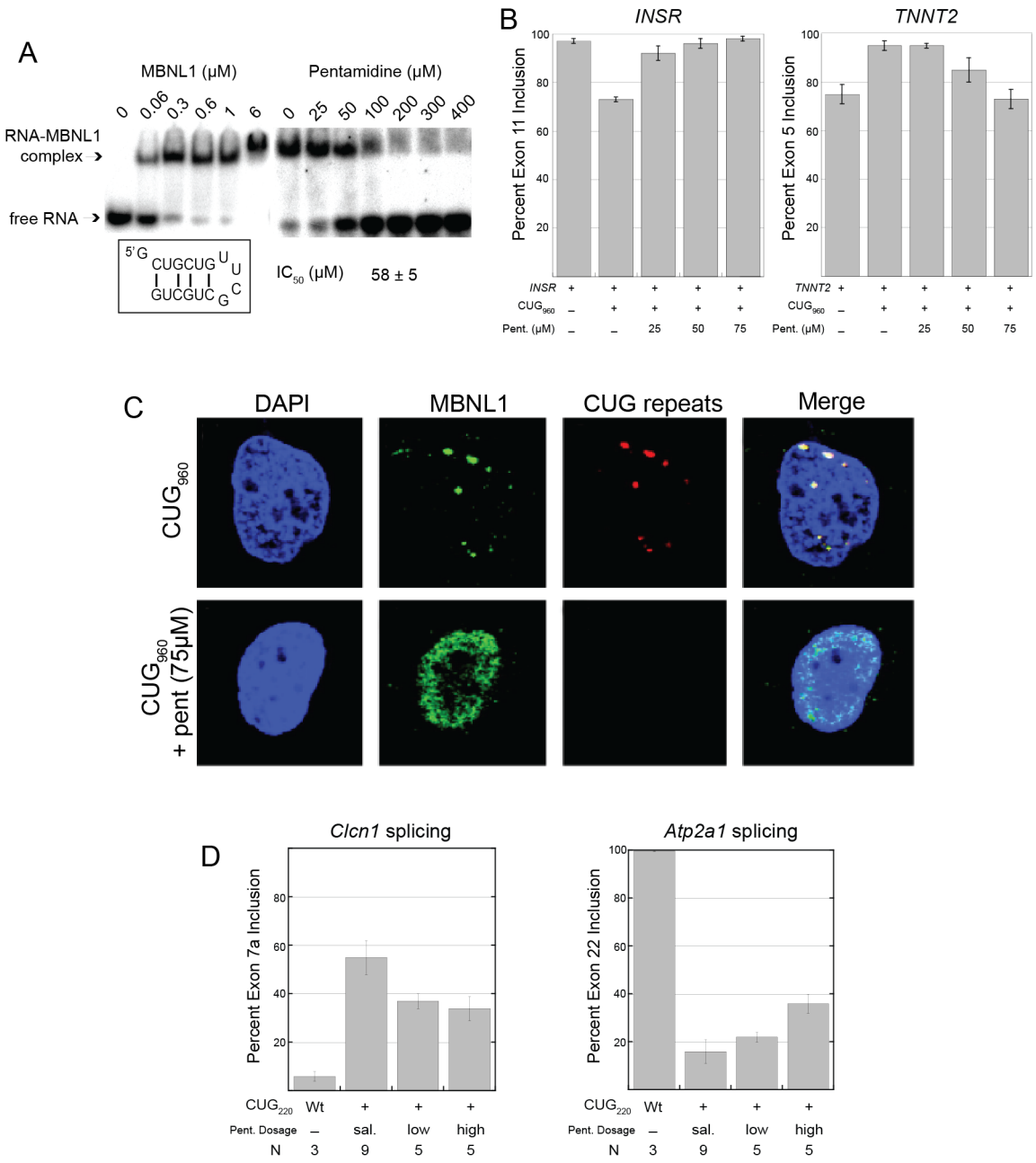


Figure 4. Pentamidine disrupts the MBNL1/$(CUG)_4$ complex *in vitro* and rescues mis-splicing *in vivo*. (A) EMSA showing MBNL1 binds to $(CUG)_4$, a short hairpin RNA. When pentamidine is added, it disrupts the MBNL1/$(CUG)_4$ complex with an IC_{50} of 58 μM. (B) Pentamidine rescues the mis-splicing of two transiently transfected minigenes in a DM1 HeLa cell model. (C) Pentamidine reduces the number of nuclear foci in cells expressing 960 CUG repeats, freeing MBNL1 within the nucleus. (D) DM1 mice treated with pentamidine show mis-splicing rescue of two endogenous transcripts.

Myotonic dystrophy and RNA structure

Since the expanded CUG repeats are the causative agent of DM1, their structure has been the object of much interest since the discovery of their toxic nature. The first studies on CUG structure were probing experiments that found that extended CUG repeats formed hairpin structures (*32*). The first crystal structure of the CUG repeats showed that, despite having a U-U mismatch every third base pair, the structure was essentially A-form (Figure 2C) (*33*). This structure also found that the U-U mismatches formed one hydrogen bond, and did not distort the backbone of the RNA, as other U-U mismatches had been observed to do (*34*). Since the first crystal structure, several other crystal structures of CUG repeat RNA have been released (*35*, *36*), as well as an NMR structure (*37*), confirming the A-form nature of CUG repeats. These structures also found that the U-U mismatches in the 5'CUG/3'GUC motif were not strictly locked in to the single hydrogen bond conformation.

Understanding the structure of CUG repeats may help us understand how MBNL proteins recognize the YGCY motif (*38*) through the Watson-Crick face of the GC dinucleotide (*39*), but in the context of double-stranded RNA these faces of the nucleotides are not available for binding by MBNL proteins. Additionally, knowing the structure of CUG repeats and how they interact with small molecules, can help guide rational drug design.

In order to improve upon our initial lead compound pentamidine, we undertook a structure activity relationship (SAR) study. In Chapter II, I discuss complications surrounding the original screening assay and demonstrate that pentamidine needs a "helper" compound in order to efficiently disrupt the MBNL1/$(CUG)_4$ complex in the

competitive EMSA. Chapter III and Appendix A discuss the SAR study performed on pentamidine methylene linker analogs using the original competitive EMSA and an *in vivo* splicing assay. We show that as the methylene linker length increases, compounds are more effective at disrupting the MBNL1/$(CUG)_4$ complex *in vitro*, and also rescue the mis-splicing of two different MBNL targets in a HeLa cell model of DM1. Additionally, we observed that heptamidine (a derivative of pentamidine containing a seven carbon methylene linker) was able to reverse myotonia in a DM1 mouse model. In Chapter IV, I crystallized CUG repeats utilizing a crystallization module in order to facilitate intermolecular contacts and analyze the flexibility of the U-U mismatch within the context of the 5'CUG/3'GUC motif. Chapters II, III, and IV contain unpublished, co-authored material.

CHAPTER II

A NEW CLASS OF SMALL MOLECULES DISRUPTS MBNL1/CUG REPEAT BINDING AND WORKS SYNERGISTICALLY WITH PENTAMIDINE

I was the primary contributor to the experiments described in the following chapter and did all of the writing. Dr. J. Andrew Berglund was the principal investigator of this work.

INTRODUCTION

Myotonic dystrophy (DM) is an autosomal dominant disorder caused by a toxic, gain-of-function RNA. In myotonic dystrophy type 1 (DM1), there is an expansion of a CTG trinucleotide repeat in the 3' UTR of the DMPK gene. Unaffected individuals have between 5 and 37 CTG repeats, whereas DM1 patients have 50 or more. Upon transcription, these CTG repeats form stable, structured CUG duplex RNA which aberrantly interacts with many different nuclear proteins. One of the proteins that binds the CUG repeats is Muscleblind-like 1 (MBNL1), a regulator of alternative splicing (reviewed in ref (*40*)). MBNL1 has been shown to co-localize with the CUG repeats in the nuclear foci that are a hallmark of DM (*41*). Additionally, MBNL1 has been shown to directly interact with CUG repeats *in vitro* (*16*, *42*, *43*). Thus, MBNL1 is sequestered to this expanded RNA and less available to perform its function of regulating alternative splicing events (*14-18*). It is this sequestration of MBNL1 that leads to mis-splicing events observed in DM1, a number of which have been correlated with disease symptoms (reviewed in refs (*40*, *44*)).

We previously described a small molecule screen used to identify small molecule inhibitors of the CUG/MBNL1 interaction (*20*). The screen used a competitive electrophoretic mobility shift assay (EMSA) wherein the amount of MBNL1 and $(CUG)_4$ (a small hairpin RNA composed of two CUG repeats, a UUCG cap, and two more CUG repeats) were held constant. As molecules were added, those effective at disrupting the MBNL1/$(CUG)_4$ complex resulted in an increase in free RNA. We discovered that pentamidine (a currently FDA approved drug for the treatment of Pneumocysitis carinii infections) is able to disrupt MBNL1/$(CUG)_4$ binding with an IC_{50} of 58 ± 5 μM. We further showed that in an *in vivo* HeLa cell splicing system, pentamidine is able to rescue mis-splicing events caused by the presence of 960 CUG repeats and to disrupt the characteristic nuclear foci observed in DM1. Finally, when pentamidine is administered to mice that express 250 CUG repeats, partial reversal of DM-related splicing defects is observed. From this study, it was evident that a small molecule approach to treating DM1 has the potential to be successful, since pentamidine has the ability to disrupt the MBNL1/CUG complex *in vitro* and to presumably do the same in both HeLa cells and mouse tissue to reverse splicing defects.

During subsequent studies of the pentamidine/CUG repeat interaction, we discovered that the concentration of dye in the EMSA dramatically altered the effectiveness of pentamidine. This observation lead to the discovery that bromophenol blue (BPB) is able to compete with the MBNL1/$(CUG)_4$ complex on its own, as well as synergistically with pentamidine. An SAR study of BPB showed that iodophenol blue (IPB) demonstrates an approximately 7.5-fold increase in efficacy over BPB. Other methods can disrupt the MBNL1/CUG repeat complex, including oligonucleotide analogs and several different

small molecules (reviewed in ref (*45-47*)), but phenolsulphonphthaleins are unique because it appears this class of compounds function through a different mechanism than the previously described molecules.

RESULTS and DISCUSSION

BPB disrupts the MBNL1/$(CUG)_4$ complex. While optimizing conditions for the competitive EMSA, we discovered that BPB, the dye used to monitor reaction mixtures and progress through the polyacrylamide matrix, acted to disrupt the MBNL1/$(CUG)_4$ complex with an IC_{50} of 945 ± 30 μM (Figure 5).

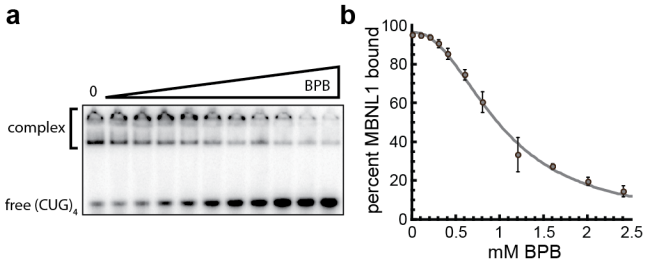


Figure 5. Titration of BPB. (a) EMSA of the competition of BPB with the MBNL1/$(CUG)_4$ complex at 0, 100, 200, 300, 400, 600, 800, 1200, 1600, 2000, and 2400 μM BPB. (b) IC_{50} curve of BPB competition with MBNL1.

Since BPB is an anionic molecule at the pH at which the competitive EMSA is performed, it is almost certainly disrupting the complex differently than pentamidine. We suspect that BPB is interacting with the protein rather than the negatively charged RNA. Additionally, BPB is known to interact with other proteins (*48-51*), so it is likely BPB is interacting with MBNL and disrupting its RNA binding ability.

BPB and pentamidine work synergistically. Since BPB and pentamidine probably act through different mechanisms to disrupt the MBNL1/$(CUG)_4$ complex (pentamidine is suspected of interacting with the RNA), we decided to test them together, to observe if pentamidine and BPB could act simultaneously. Interestingly, when BPB and pentamidine are both present in the *in vitro* assay, they act synergistically: in the presence of 38 μM pentamidine, the IC_{50} of BPB drops to 376 μM (Figure 6, Table 1), compared to an IC_{50} 945 ± 30 μM with BPB alone. The same is true of pentamidine: with increasing concentrations of BPB, pentamidine appears to be more effective at disrupting the complex (Figure 6a). In fact, under these assay conditions, ≥200 μM BPB is required for pentamidine dependent complex disruption (Figure 6b). With increasing BPB concentrations, the effect of pentamidine is magnified, from complete loss of efficacy without BPB, to an IC_{50} of 15 μM at 1200 μM BPB (Figure 6b, Table 1). Thus it is clear that pentamidine and BPB are working synergistically in the EMSA assay to disrupt the MBNL1/$(CUG)_4$ complex.

Table 1. IC_{50} values of pentamidine with varying concentrations of BPB and BPB with varying concentrations of pentamidine.

Pentamidine titrations		Bromophenol blue titrations	
[BPB]	IC_{50} pent	[pent]	IC_{50} BPB
0 μM	–	0 μM	945 ± 30 μM
100 μM	–	6.6 μM	925 ± 56 μM
200 μM	84 ± 3 μM	11 μM	892 ± 60 μM
300 μM	46 ± 2 μM	17 μM	823 ± 40 μM
400 μM	36 ± 3 μM	22 μM	663 ± 77 μM
600 μM	27 ± 1 μM	26 μM	516 ± 47 μM
800 μM	23 ± 2 μM	32 μM	426 ± 37 μM
1200 μM	15 ± 3 μM	38 μM	347 ± 17 μM

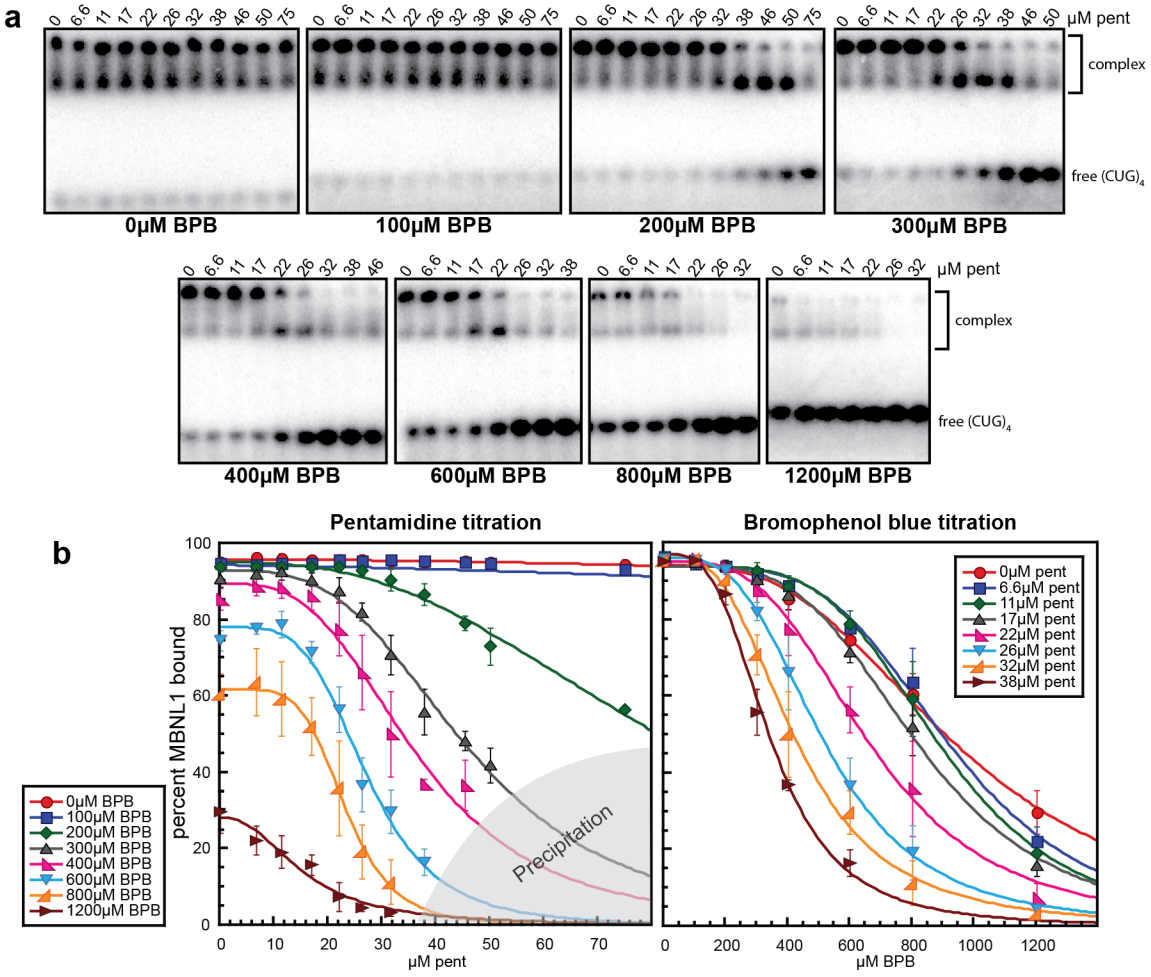


Figure 6. Pentamidine (pent) and bromophenol blue (BPB) titrations. (a) EMSA of the competition of pentamidine and BPB with the MBNL1/$(CUG)_4$ complex. In the absence of BPB, it is clear that pentamidine has no effect on disrupting the protein/RNA complex. As the BPB concentration increases, pentamidine is better able to free the $(CUG)_4$ RNA. At high concentrations of both BPB and pentamidine, precipitation was observed, both as an increase in the number of counts in the wells and by a blue precipitate in the wells. These points were excluded. (b) IC_{50} curves for each concentration of BPB (left graph) and pentamidine (right graph).

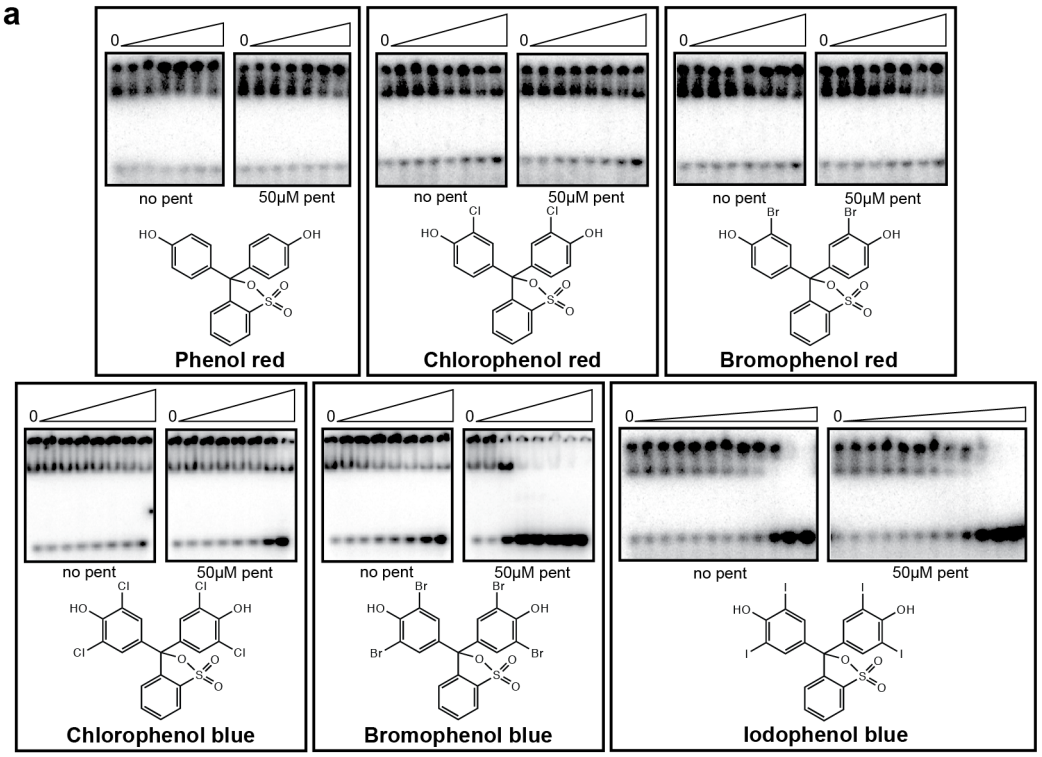
There are several potential mechanisms to account for the synergistic effect of pentamidine and BPB on the MBNL1/CUG repeat interaction. It is possible that BPB decreases the affinity of MBNL1 for the CUG RNA enough so that the effect of pentamidine is more easily observed. This could take place through binding interactions with MBNL1. BPB is known to interact with other proteins, so it is plausible that BPB is able to interact with MBNL1 and lessen its affinity for RNA. Another possibility is that BPB and pentamidine form an ion pair, since, at physiological pH, BPB carries a negative charge on the arylsulfonate group, and pentamidine has a positive charge on each of the amidine groups. If formed, this ion pair could target the RNA better than either pentamidine or BPB alone. In an effort to identify important structural features of BPB, we decided to test dyes similar to BPB to determine if we could identify more potent inhibitors and gain insight into the mechanism of inhibition.

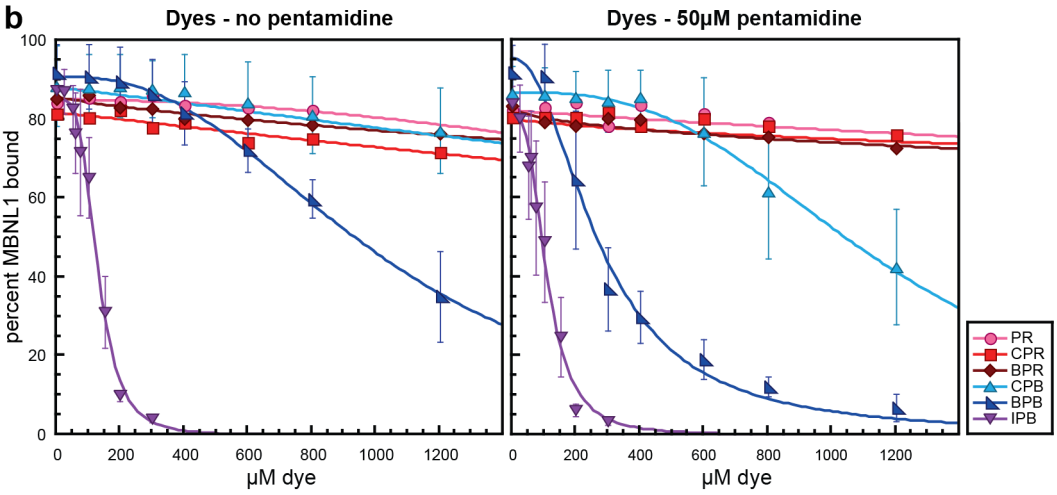
Dye efficacy depends on ring halogenation. BPB is a pH indicator composed of two brominated phenolic rings and a benzene sulfonyl ring joined by a single carbon. We found many other commercially available indicators that had substitutions at the position(s) of the bromines. These indicators substituted other halogens, were mono-halogenated, or had no halogens. Each of these dyes was tested for its ability to disrupt the MBNL1/$(CUG)_4$ complex alone and with 50 μM pentamidine. Phenol red (PR), which has no halogens, failed to compete both on its own and with pentamidine (Figure 7a). The same was true of bromophenol red (BPR) and chlorophenol red (CPR), which both have only one halogen on each of the phenol groups (Figure 7a). Chlorophenol blue (CPB), with two chlorines on each phenol group, was not able to compete with the MBNL1/$(CUG)_4$ complex on its own (Figure 7); however, with the addition of 50 μM

pentamidine, it was able to disrupt the complex with an IC_{50} of 1166 ± 255 μM. As expected, BPB was able to compete on its own (IC_{50} = 945 ± 30 μM) and with in-creased efficacy in the presence of pentamidine (IC_{50} = 276 ± 30 μM). Of all the BPB derivatives, iodophenol blue (IPB), which has two iodines on each phenol group, performed the best. It had an IC_{50} of 127 ± 17 μM alone and an IC_{50} of 107 ± 26 μM with pentamidine (Figure 7a), demonstrating only a modest, if any, increase in efficacy in the presence of pentamidine (Figure 7b).

The increase in complex disruption efficacy observed when pentamidine and a phenolsulphonphthalein dye are together is clearly affected by the halogenation of the dye. There seems to be an inverse relationship between electronegativity and ability to compete; the chlorinated dye (CPB), which is the most electronegative, is the least effective, while the iodinated (IPB) dye is the most effective and the least electronegative.

Figure 7 (next page). Titrations of dyes with and without pentamidine. (a) EMSA of the competition with the MBNL1/$(CUG)_4$ complex of the six dyes alone and with 50μM pentamidine. All dyes, with the exception of PR and IPB were tested at the concentrations 0, 100, 200, 300, 400, 600, 800, and 1200μM. PR was tested at 0, 100, 200, 300, 400, 600, and 800μM due to solubility issues and IPB was tested at 0, 25, 50, 60, 75, 100, 150, 200, and 300μM since it was much more effective than any of the other dyes. PR clearly has no effect, alone or with pentamidine, while CPB and BPR may be beginning to compete at their highest concentrations. CPB, while no observable competition is taking place with the dye alone, when added to pentamidine, together they are able to compete quite well. IPB is the best competitor of the group, and shows only modest increases in efficacy in the presence of pentamidine. (b) IC_{50} curves of for each dye alone (left graph) and with 50μM pentamidine (right graph). The most drastic changes are observed with CPB and BPB.


a
0
no pent
50μM pent
Phenol red
Chlorophenol red
Bromophenol red
Chlorophenol blue
Bromophenol blue
Iodophenol blue


b
Dyes - no pentamidine
Dyes - 50μM pentamidine
percent MBNL1 bound
μM dye
PR
CPR
BPR
CPB
BPB
IPB

We suspect that CPB, BPB, and IPB are somehow interacting with MBNL1, sufficiently weakening the interaction between MBNL1 and $(CUG)_4$ so that pentamidine's effect can be observed *in vitro*. Although pentamidine does not effectively inhibit the MBNL1/$(CUG)_4$ complex in the EMSA in the absence of dyes, pentamidine works quite will in both HeLa cells and a DM1 mouse model (*20*). Initially, we considered the possibility that PR and pentamidine were working together in cell culture, since PR is used as a pH indicator in the growth media, but that appears unlikely based on the EMSA results. No dyes were added in the mouse studies with pentamidine, suggesting in the cell/animal environment, pentamidine is either more "potent" or an endogenous molecule(s) replaces the dyes.

Interestingly, we have discovered a new class of compounds able to disrupt MBNL1/CUG complexes. Although there is no existing treatment for DM1, many different approaches are currently being developed. Many focus on displacing MBNL from the CUG repeat structures, either through utilizing small molecules (*20*, *21*), multivalent compounds (*25*), small peptides (*27*), or antisense oligonucleotide analogs (*29*). Other techniques include cleaving the CUG repeats (*52-54*) or triggering their degradation (*55-57*). The phenolsulphonphthalein class of compounds is unique because, based on literature precedence, (*48*, *49*, *51*, *58*) we believe it is acting on the protein rather than the nucleic acid. We have already been able to make significant improvements over the lead compound BPB. By simply replacing the bromines of BPB for the iodines of IPB, we have observed 7.5 fold decrease in the IC_{50} of IPB as compared to BPB. It may well be worth further optimizing BPB-like compounds, or investigating other halogenated ring systems. More importantly, we are now aware of the *in vitro* dye effect,

and can properly control for it in future drug comparison studies searching for pentamidine analogs with better efficacy. Determining if dye-like molecules are working in the cell to aid pentamidine (and potentially other small molecule inhibitors), and identifying these compounds are important questions to address in the future.

METHODS

MBNL1 purification. The MBNL1 1-260 protein was purified as previously described (*20*) with the following changes. After purification, the protein was dialyzed into 500 mM NaCl, 20 mM Tris pH 7.5, 50% glycerol, and 5 mM β-mercaptoethanol. Aliquots were stored at –80 °C until just before use.

RNA labeling. The $(CUG)_4$ RNA construct used in gel shifts (5'–GCUGCUGUUCGCUGCUG) was ordered from IDT and 5' labeled using [γ–32P]ATP. After kinasing, the RNA was run over a spin column containing Bio-Gel P2 Gel, brought to final stock concentration of 50 nM using low TE, and stored at –20 °C until use.

Gel shift assay. Kinased $(CUG)_4$ RNA was snap annealed in 250 mM NaCl, 25 mM $MgCl_2$, and 75 mM Tris pH 7.5 by heating at 95 °C for 3 minutes. The reaction was placed on ice for 5 minutes, BSA and heparin were added, and the reaction placed back on ice for another 5 minutes. MBNL1 was then added to the RNA, and incubated at room temperature (RT) for 10 minutes. Pentamidine or water was added, followed by another 5 minute RT incubation, followed by the addition of dye. The final reaction volume of 10 μL was incubated at RT for 25 minutes. Final reaction conditions were 0.5 nM $(CUG)_4$, 175 mM NaCl, 20 mM Tris pH 7.5, 5 mM $MgCl_2$, 2 mg mL^{-1} BSA, 0.1 mg mL^{-1} heparin, 1.25 mM β-mercaptoethanol, 1.25% glycerol, and 250 nM MBNL. Then 3 μL of the reaction was run on a 5% 80:1 pre-chilled polyacrylamide gel with 0.5x TB. Gels were

run at 4 °C at 170V for 35 minutes, dried and autoradiographed. IC_{50} values were calculated as previously described (Warf, 2009). In brief, gels were quantified using ImageQuant (Molecular Dynamics). The percent of RNA bound was determined by taking the ratio of the RNA:protein complex (including any well shifting) to the total RNA, per lane. To determine the IC_{50}, the equation

$$Y = \frac{m_3}{1 + \left(\frac{m_0}{m_1}\right)^{m_2}}$$

was used, where m_0 = small molecule concentration, $m_1 = IC_{50}$, m_2 = Hill coefficient, and m_3 = fraction of MBNL bound without small molecule present. When present, errors were determined by calculating the standard deviation of triplicate data.

Pentamidine was purchased from Sigma-Aldrich as the isethionate salt, converted to the chloride salt, and dissolved in water. PR, CPR, BPR and IPB (3', 3", 5', 5"-tetraiodophenolsulfonphthalein) were purchased from TCI, CPB was purchased from City Chemical, and BPB was purchased from Sigma. Dyes were dissolved in water with a small amount of NaOH to help with solubility and bring to a neutral pH.

CHAPTER III

STRUCTURE OPTIMIZATION REVEALS A SMALL MOLECULE THAT REVERSES THE SPLICING DEFECTS AND MYOTONIA IN A MOUSE MODEL OF MYOTONIC DYSTROPHY

I was the primary contributor to the experiments described in the following chapter and did all of the writing. Dr. Masayuki Nakamori performed all the experiments involving mice. Dr. Cameron Hilton and Dr. Micah Bodner synthesized the small molecules. Dr. Michael Haley and Dr. Charles Thornton helped design experiments, and Dr. J. Andrew Berglund was the principal investigator of this work.

INTRODUCTION

Myotonic dystrophy (DM) is characterized by many different symptoms, including, but not limited to, cataracts, cardiac conduction defects, cognitive dysfunction, muscle wasting, and, its hallmark symptom, myotonia, where muscles have delayed relaxation. There are two types of DM: type 1 (DM1) and type 2 (DM2). DM1 is caused by the aberrant expansion of CTG repeats in the 3' UTR of the *DMPK* gene. Unaffected individuals have between 5 and 37 repeats, whereas DM1 patients have more than 50 repeats. Upon transcription, these repeats form a CUG repeat stem loop RNA, which is capable of sequestering RNA binding proteins in nuclear foci. Similar circumstances cause DM2, where CCTG repeats in the first intron of the *ZNF9* gene expand and lead to production of a CCUG repeat stem loop (reviewed in refs (*40*, *59*)).

One of the protein families capable of binding these CUG/CCUG repeats is the Muscleblind-like (MBNL) family of proteins (*10*). In humans, there are three paralogs:

MBNL1, MBNL2, and MBNL3. The MBNL proteins are zinc-finger RNA binding proteins that regulate alternative splicing events (*13*). When extended CUG/CCUG repeats are present, MBNL proteins are sequestered to the repeat RNA and are no longer able to properly regulate alternative splicing of their target transcripts (*14-18*). Many of the mis-splicing events observed in DM are MBNL dependent, and some of these events have been correlated with disease symptoms. One target of the MBNL proteins is the *Clcn1* gene, which encodes a chloride ion channel. The MBNL proteins are negative regulators of exon 7a in *Clcn1* (when MBNL proteins are present, exon 7a is excluded). In the presence of CUG/CCUG repeats, the levels of free MBNL are not sufficient to properly regulate splicing of exon 7a. This leads to chloride channelopathy and the characteristic DM symptom myotonia (*19*).

Although there are currently no treatments available to patients with DM, many different approaches are under development. Several methods focus on displacing MBNL proteins from the CUG/CCUG repeat structures, either through utilizing small molecules (*20-22*, *22*, *23*), multivalent compounds (*24-26*), small peptides (*27*, *28*), or antisense oligonucleotide analogs (*29*). Other techniques include cleaving the repeats (*52-54*) or triggering their degradation (*55-57*, *60*). We have recently discovered that pentamidine, an FDA approved drug for treatment of *Pneumocystis carinii* infections, disrupts MBNL1/$(CUG)_4$ interactions in a competitive electrophoretic mobility shift assay (EMSA) (*20*). Pentamidine also rescues mis-splicing of transiently transfected minigene reporters in a HeLa DM1 cell model expressing 960 interrupted CUG repeats. Most encouragingly, pentamidine was used to improve mis-splicing of endogenous transcripts in a DM1 mouse model expressing 220 CUG repeats. From this study, it was evident that

a small molecule approach for treating DM1 has the potential to be successful since pentamidine has the ability to rescue splicing defects in both HeLa cells and mouse tissue.

In an effort to develop a pentamidine analog with improved physiochemical properties, we undertook a structure activity relationship (SAR) study. Analogs containing between three to nine methylene carbons were synthesized and tested using the *in vitro* EMSA assay for their ability to rescue the mis-splicing of MBNL regulated minigenes in our HeLa cell splicing assay. In general, analogs with more methylenes more efficiently rescued splicing. The most promising compound, heptamidine, was tested for its ability to rescue mis-splicing and myotonia in a transgenic mouse model of DM1.

RESULTS and DISCUSSION

Pentamidine analogs rescue two mis-spliced minigenes in tissue culture. A common tactic to improve the physiochemical properties of a drug candidate is to vary the hydrophobicity of the molecule, and we utilized this strategy by varying the length of pentamidine's methylene linker. The most polar analog was propamidine, with a three carbon linker, and the least polar was nonamidine, with a linker of nine carbons (Figure 8). During our studies with pentamidine, we discovered that a dye, bromophenol blue (BPB), used to aid in gel loading/tracking, was necessary for pentamidine to robustly inhibit the MBNL1/$(CUG)_4$ interaction. In fact, BPB was able to dissociate the MBNL1/$(CUG)_4$ complex in the absence of pentamidine (Figure 9a, b). Additionally, we observed that BPB and pentamidine were able to function synergistically in the electrophoretic mobility shift assay (EMSA) (Figure 9c). That is, with increasing

concentrations of BPB, pentamidine appeared to be more effective at disrupting the MBNL1/$(CUG)_4$ complex. Due to this complication in the EMSA, we decided to use our DM1 cell model for screening the activity of the methylene linker analogs.

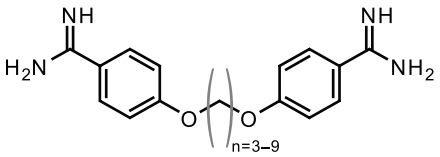


Figure 8. Structure of the pentamidine analogs. Methylene carbon linker highlighted by gray parenthesis.

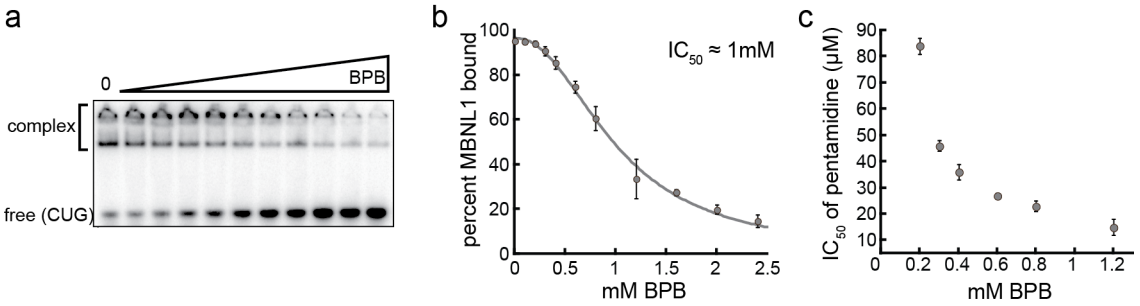


Figure 9. Titration of BPB and synergy with pentamidine. (a) EMSA of the competition of BPB with the MBNL1/$(CUG)_4$ complex at 0, 0.1, 0.2, 0.3, 0.4, 0.6, 0.8, 1.2, 1.6, 2, and 2.4 mM BPB. (b) IC_{50} curve of BPB competition with MBNL1. (c) Plot of the IC_{50} of pentamidine at 0.2, 0.3, 0.4, 0.6, 0.8, and 1.2 mM BPB (at 0 and 0.1 mM BPB, pentamidine did not disrupt the MBNL1/$(CUG)_4$ complex in the EMSA). As BPB concentration increases, the IC_{50} of pentamidine decreases. Error bars denote the standard deviation of at least three observations.

Each analog was tested for the ability to rescue splicing of transiently transfected minigenes in a HeLa cell model of DM1. Two different minigenes containing exons that

are mis-spliced in DM were tested: *TNNT2* (also known as cTNT), containing the alternatively spliced exon 5, and *INSR*, containing the alternatively spliced exon 11. MBNL proteins facilitate exclusion (negative regulation) of exon 5 of *TNNT2*: when MBNL proteins are present, exon 5 is excluded. The typical level of exon 5 inclusion when the *TNNT2* minigene is expressed in HeLa cells is 64 ± 2%. The inclusion level increased to 82 ± 3% when HeLa cells co-expressed a DMPK plasmid containing 960 CUG repeats (Figure 10a). Presumably, this change in *TNNT2* exon 5 inclusion is due to the sequestration of endogenous MBNL proteins to the CUG repeats. Propamidine did not have the desired effect on *TNNT2* mis-splicing at concentrations up to 130 μM (higher concentrations were toxic). Also unable to rescue splicing defects before causing significant cell death (at 4 μM) was nonamidine. The remaining linker analogs rescued *TNNT2* mis-splicing to varying degrees. The concentration needed to observe 50% rescue (EC_{50}) for butamidine was 23 ± 5 μM, which was similar to pentamidine (EC_{50} = 20 ± 4 μM). Interestingly, at higher concentrations, butamidine and pentamidine lowered exon 5 inclusion levels below that of wild type. Hexamidine and heptamidine showed improvements over pentamidine with EC_{50} values of 12 ± 3 μM and 15 ± 6 μM, respectively. Finally, octamidine showed a slight rescue of *TNNT2* mis-splicing; however, because of toxicity issues, it was not possible to treat the cells with high enough concentrations to obtain an accurate EC_{50} value (Figure 10b). Thus, hexamidine is able to rescue *TNNT2* mis-splicing at the lowest dose compared to other linker analogs.

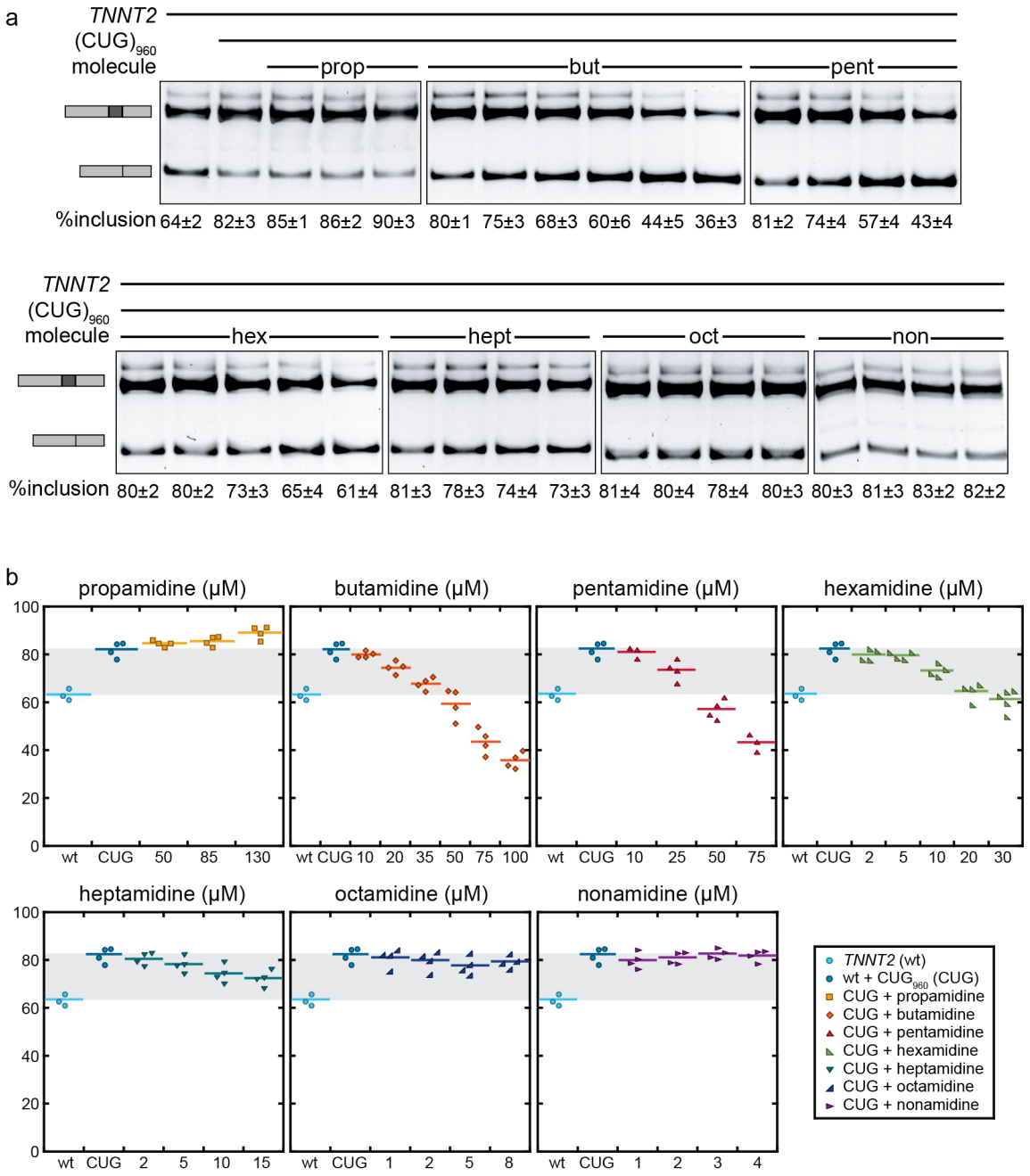


Figure 10. Pentamidine analogs rescue mis-splicing of *TNNT2* minigene reporter in a HeLa cell DM1 model. (a) RT-PCR data of *TNNT2* splicing. Errors are standard deviation. (b) Jitter plot representation of *TNNT2* splicing. Each point is one experiment and the line represents the average of all experiments for that condition (at least three for each concentration). Butamidine, pentamidine, hexamidine, and heptamidine rescue *TNNT2* mis-splicing in the presence of 960 CUG repeats. Octamidine may have a slight effect. Gray area denotes range between typical splicing and DM mis-splicing.

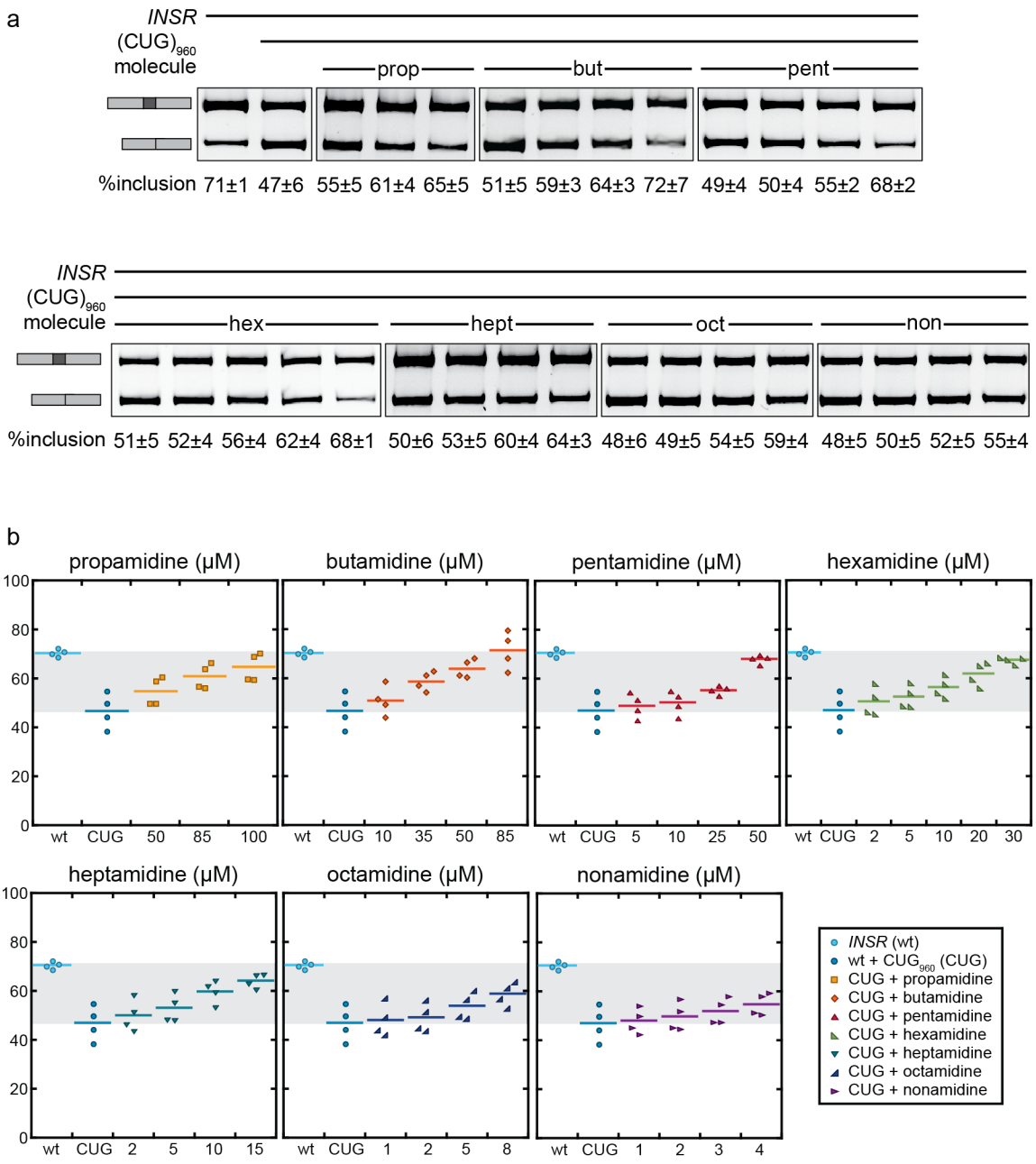


Figure 11. Pentamidine analogs rescue mis-splicing of *INSR* minigene reporter in a HeLa cell DM1 model. (a) RT-PCR data of *INSR* splicing. Errors are standard deviation. (b) Plots of *INSR* splicing. All pentamidine linker analogs are able to fully or partially rescue *INSR* mis-splicing in the presence of 960 CUG repeats. Gray area denotes range between typical splicing and DM mis-splicing.

We also tested the ability of the linker analogs to rescue an exon that is positively regulated by MBNL proteins: exon 11 of the *INSR* gene. When wild type HeLa cells expressed the *INSR* reporter minigene, exon 11 inclusion was 71 ± 1%. When 960 CUG repeats were co-expressed with the *INSR* reporter, exon 11 inclusion dropped to 48 ± 7% (Figure 11a). Unlike the *TNNT2* study, all linker analogs were able to partially or fully rescue the mis-splicing of exon 11 of *INSR* when CUG repeats were expressed (Figure 11b). Propamidine rescued mis-splicing with an EC_{50} of 64 ± 10 μM. Butamidine and pentamidine were similar with EC_{50} values of 34 ± 7 μM and 31 ± 2 μM, respectively. Hexamidine also rescued *INSR* mis-splicing (EC_{50} = 15 ± 1 μM), as did heptamidine (EC_{50} = 9 ± 1 μM). Unlike with the *TNNT2* minigene, both octamidine and nonamidine rescued *INSR* mis-splicing with EC_{50} values of 7 ± 1 μM and 6 ± 1 μM.

All the compounds were able to partially or fully rescue *INSR* mis-splicing, while only butamidine, pentamidine, hexamidine, and heptamidine rescued *TNNT2*. One possible explanation for this observation is that different MBNL targets require different concentrations of MBNL proteins in the cell to be properly regulated. For example, nonamidine is toxic to HeLa cells at concentrations above 4 μM. *TNNT2* is not rescued; however, *INSR* is rescued by 30%. Presumably, the same amount of MBNL proteins are freed inside the cell, so the difference in rescue observed between the two minigenes is likely due to the amount of MBNL required for proper splicing regulation. Interestingly, when rescue was observed with *TNNT2*, those compounds had slightly lower EC_{50} values as compared to *INSR* (Figure 12). Additionally, *TNNT2* exon 5 inclusion was often decreased to inclusion levels that are lower than wild type (Figure 10b), suggesting that these compounds may function to block exon 5 inclusion of the *TNNT2* pre-mRNA,

possibly by interacting with the *TNNT2* pre-mRNA (*20*). The effect of pentamidine and its analogs on the splicing of the *TNNT2* pre-mRNA does not appear to be a general phenomenon, because splicing of *INSR* and other pre-mRNAs tested were not altered by pentamidine in the absence of CUG repeats (*61*).

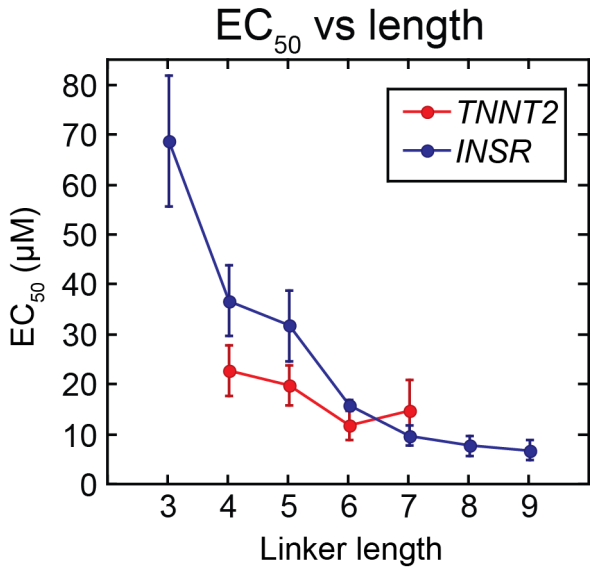


Figure 12. Plot of EC_{50} values versus methylene linker length. When *INSR* splicing is monitored, the inverse relationship between methylene linker length and EC_{50} holds. The trend is less clear with *TNNT2*. Error bars are standard deviation.

As linker length increases, we observe an increase in efficacy, but we also observe an increase in toxicity to HeLa cells. This could be due to the increased lipophilicity of the molecule as the carbon chain becomes longer. While pentamidine satisfies some of Lipinski's "rule of 5" requirements for orally available drug-like compounds (*62*), its constitutively charged amidines hinder its ability to diffuse through membranes. As linker length increases, the hydrophobicity of the molecule also increases. Thus, the compounds

with longer linkers may cross the membrane more effectively, resulting in increased intracellular concentrations, leading to increased efficacy and toxicity observed in the HeLa cell splicing assay.

Heptamidine rescues mis-splicing in a DM1 mouse model and reduces myotonia symptoms. Because heptamidine rescued mis-splicing of both minigenes in HeLa cells while retaining water solubility, it was tested in the HSA^{LR} transgenic DM1 mouse model. The HSA^{LR} DM1 mouse model expresses 220 CUG repeats under the skeletal promoter (*63*). Two different endogenous pre-mRNAs were observed: *Clcn1*, the mis-splicing of which has been shown to cause myotonia (*19*), and *Atp2a1* (also called Serca1), which is a robust marker of mis-splicing in DM (*64*). MBNL proteins have been shown to promote exclusion of exon 7a of the *Clcn1* gene. Wild type adult mice included exon 7a of the *Clcn1* pre-mRNA at a level of 4 ± 1% while HSA^{LR} mice included exon 7a at steady state levels of 47 ± 1% (Figure 13, note that exon 7a inclusion isoforms are subject to nonsense mediated decay). Treatment with heptamidine caused a dose-dependent reduction of exon 7a inclusion in HSA^{LR} mice, returning to wild type levels (6 ± 1%) at the dose of 20 mg kg^{-1} heptamidine (Figure 13).

The MBNL regulated exon of *Atp2a1* (exon 22) is included in the presence of MBNL proteins (positively regulated). Wild type adult mice included this exon at 100 ± 1% (Figure 13). When MBNL proteins are sequestered by the 220 CUG repeats present in HSA^{LR} mice, exon 22 inclusion dropped to 23 ± 3%. Although full rescue with heptamidine could not be reached, under a treatment regimen of 30 mg kg^{-1} per day for 7 d, exon 22 inclusion levels returned to 62 ± 4% (Figure 13). Additionally, after being treated with 30 mg kg^{-1} heptamidine for 7 d, mice went untreated for 10 d, and splicing of

Clcn1 and *Atp2a1* was examined. In both mRNAs, exon inclusion levels returned to control HSA^{LA} levels (Figure 13).

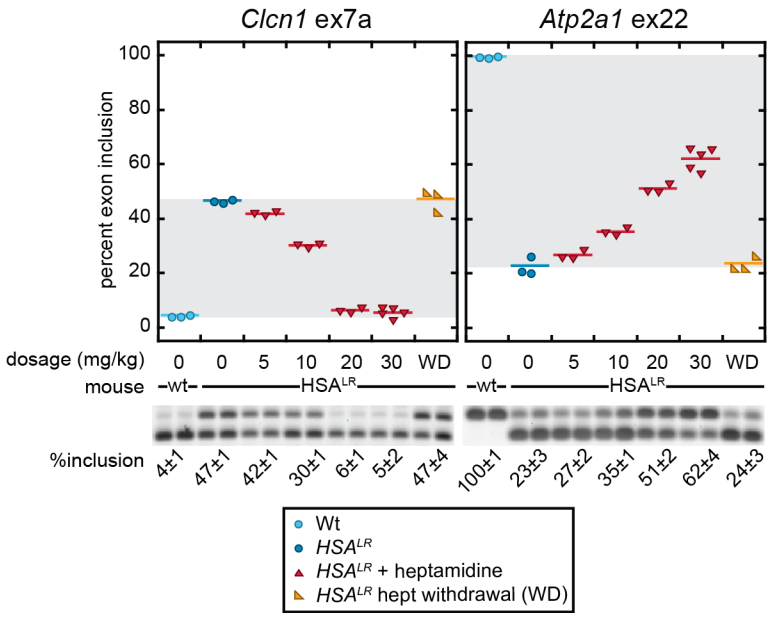


Figure 13. Heptamidine rescue of endogenous mis-splicing events in HSA^{LR} mice. *Clcn1* showed a complete rescue by 20 mg/kg for 7 d. Each symbol represents the splicing outcome for vastus muscle of a single mouse. After treatment, withdrawal mice (WD) were maintained for 10 d with no additional heptamidine injections. These mice showed a complete return to disease state splicing. *Atp2a1* showed partial rescue of mis-splicing. At 30 mg/kg heptamidine, exon 22 inclusion is approximately 50% rescued. WD mice again reverted to the pre-treatment splicing levels. Errors are standard deviation. Gray area denotes range between typical splicing and DM mis-splicing.

In addition to splicing defects, HSA^{LR} mice exhibit myotonia, manifested by runs of repetitive action potentials. We graded the severity of myotonia by insertion of extracellular recording electrodes into muscle tissue (electromyography) under general

anesthesia. In glucose-treated controls, we observed grade 3 myotonia in vastus muscle, which indicates abundant repetitive discharges with nearly all electrode insertions. When treated with 20 or 30 mg kg^{-1} heptamidine, the myotonia was reduced from grade 3 to grade 1 (occasional myotonic discharge) or grade 0 (no myotonia) (Figure 14). These results are consistent with *Clcn1* splicing rescue observed under high heptamidine dosages, and show that by correcting DM1 mis-splicing events with a small molecule, myotonia symptoms can be alleviated.

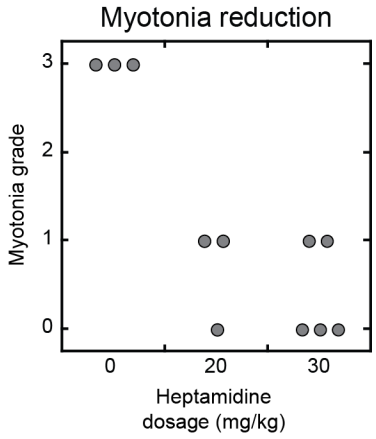


Figure 14. Myotonia rescue in *HSA*^{LR} mice treated with 0, 20, or 30 mg/kg heptamidine for 7 d. Mice not treated with heptamidine showed myotonic discharge with nearly all electrode insertions (grade 3), whereas those treated with 20 or 30 mg/kg heptamidine had occasional myotonic discharges with less than 50% of insertions (grade 1) or no myotonia (grade 0).

In the *HSA*^{LR} mouse model, heptamidine completely reverses *Clcn1* mis-splicing and partially reverses *Atp2a1* mis-splicing at lower concentrations than pentamidine (*20*). Heptamidine also was able to rapidly (one week) and significantly reduce the myotonia

phenotype that has been associated with *Clcn1* mis-splicing. It is encouraging that varying the linker length resulted in significant improvements over our lead compound pentamidine, but additional optimization is required. There is still toxicity in both the HeLa cell and mouse models (some mice treated with a higher dose of 40 mg kg^{-1} heptamidine daily did not survive for one week). We are currently working on analogs that will aid in identifying the pharmacophore of pentamidine-like compounds in order to guide our design of next generation drugs.

From our lead compound, pentamidine, we wanted to synthesize analogs that would have increased efficacy at disrupting the MBNL1/$(CUG)_4$ complex, lower toxicity, and increased cellular permeability. To identify these compounds, we conducted a SAR study varying the length of the methylene linker between the phenyl amidines from 3 to 9 carbons. It seemed likely that increasing the hydrophobic character of the molecule would lead to more favorable pharmacodynamic properties by increasing membrane permeability.

In our HeLa cell model of DM1, we observed an increase in efficacy as linker length increased (Figure 10-11). However, we are still unsure of the mechanism of pentamidine and its analogs in cells and mice. It is possible that these compounds bind the expanded CUG repeat RNA and displace MBNL proteins as we originally proposed (*20*); however, due to the BPB complication, and difficulty demonstrating a robust interaction between pentamidine and CUG repeat RNA, this mode of action appears increasingly unlikely. Alternative models include mechanisms wherein pentamidine does not displace MBNL, but rather eliminates the CUG repeats, either through affecting transcription or RNA stability. Pentamidine is known to bind double-stranded DNA (*65*), so it is possible to

envision pentamidine interacting with the CTG•CAG repeats present in both the HeLa cell and mouse models and down-regulating transcription. This would be consistent with our previous observation that pentamidine significantly reduced nuclear foci formation in HeLa cells (*20*). However, interactions of pentamidine with CUG repeat RNA that either displace MBNL directly or decrease the RNA stability are also consistent with observations thus far, and cannot be ruled out. It is also conceivable that pentamidine is acting through more than a single mechanism. It is clear that more work needs to be done to determine pentamidine's mode (or modes) of action in order to guide further small molecule drug development for DM1.

METHODS

MBNL purification. The MBNL1 medium construct (1-260) was purified as previously described (*20*) with the following changes: after purification, the protein was dialyzed into 500 mM NaCl, 20 mM Tris pH 7.5, 50% (v/v) glycerol, and 5 mM β-mercaptoethanol. Aliquots were stored at –80 °C until just before use.

RNA labeling. The $(CUG)_4$ RNA construct used in the *in vitro* EMSA (5'-GCUGCUGUUCGCUGCUG) was ordered from IDT and 5' end labeled using [γ–^{32}P]ATP. After phosphorylation, the RNA was purified using a spin column containing Bio-Gel P2 Gel (Bio-Rad), brought to final stock concentration of 50 nM using low TE, and stored at –20 °C until use.

Synthesis of pentamidine analogs. Pentamidine isethionate was purchased from Sigma-Aldrich and the HCl salt was accessed by recrystallization from hot 10% (w/v) aqueous HCl. Propamidine, butamidine and hexamidine were synthesized as described by

Tidwell and coworkers (*66*). Heptamidine, octamidine and nonamidine were prepared in an analogous fashion.

propamidine•HCl: White powder. ^{1}H NMR (600 MHz, $(CD_3)_2SO$) δ 2.34 (pent, J = 6.0 Hz, 2H), 4.27 (t, J = 6.0 Hz, 4H), 7.18 (d, J = 9.0 Hz, 4H), 7.86 (d, J = 9.0 Hz, 4H), 9.08 (br s, 4H), 9.28 (br s, 4H); ^{13}C NMR (150 MHz, $(CD_3)_2SO$) δ 28.22, 64.80, 114.80, 119.52, 130.22, 162.81, 164.67.

butamidine•HCl: White powder. ^{1}H NMR (600 MHz, $(CD_3)_2SO$) δ 1.91 (pent, J = 6.0 Hz, 4H), 4.17 (t, J = 6.0 Hz, 4H), 7.15 (d, J = 9.0 Hz, 4H), 7.88 (d, J = 9.0 Hz, 4H), 9.12 (br s, 4H), 9.31 (br s, 4H); ^{13}C NMR (150 MHz, $(CD_3)_2SO$) δ 25.12, 67.73, 114.77, 119.31, 130.20, 162.99, 164.70.

pentamidine•HCl: White powder. ^{1}H NMR (600 MHz, $(CD_3)_2SO$) δ 1.58 (pent, J = 7.2 Hz, 2H), 1.80 (pent, J = 6.0, 7.2 Hz, 4H), 4.15 (t, J = 6.0 Hz, 4H), 7.14 (d, J = 9.0 Hz, 4H), 7.86 (d, J = 9.0 Hz, 4H), 9.11 (br s, 4H), 9.29 (br s, 4H); ^{13}C NMR (150 MHz, $(CD_3)_2SO$) δ 22.50, 28.60, 68.46, 115.19, 119.69, 130.63, 163.49, 165.14.

hexamidine•HCl: White powder. ^{1}H NMR (600 MHz, $(CD_3)_2SO$) δ 1.48 (m, 4H), 1.76 (pent, J = 6.0 Hz, 4H), 4.09 (t, J = 6.0, 4H), 7.13 (d, J = 9.0 Hz, 4H), 7.87 (d, J = 9.0 Hz, 4H), 9.11 (br s, 4H), 9.30 (br s, 4H); ^{13}C NMR (150 MHz, $(CD_3)_2SO$) δ 25.60, 28.83, 68.47, 115.18, 119.66, 130.63, 163.49, 165.14.

heptamidine•HCl: White powder. ^{1}H NMR (600 MHz, $(CD_3)_2SO$) δ 1.43 (m, 6H), 1.74 (m, 4H), 4.08 (t, J = 6.6 Hz, 4H), 7.13 (d, J = 8.4 Hz, 4H), 7.88 (d, J = 8.4 Hz, 4H), 9.14 (br s, 4H), 9.31 (br s, 4H); ^{13}C NMR (150 MHz, $(CD_3)_2SO$) δ 25.79, 28.81, 28.87, 68.50, 115.17, 119.62, 130.63, 163.50, 165.15.

octamidine•HCl: White powder. ^{1}H NMR (600 MHz, $(CD_3)_2SO$) δ 1.35 (m, 4H), 1.42 (m, 4H), 1.73 (m, 4H), 4.07 (t, J = 6.6 Hz, 4H), 7.13 (d, J = 8.4 Hz, 4H), 7.87 (d, J = 8.4 Hz, 4H), 9.13 (br s, 4H), 9.31 (br s, 4H); ^{13}C NMR (150 MHz, $(CD_3)_2SO$) δ 25.79, 28.87, 29.10, 68.52, 115.16, 119.62, 130.62, 163.51, 165.15.

nonamidine•HCl: White powder. ^{1}H NMR (600 MHz, $(CD_3)_2SO$) δ 1.32 (m, 6H), 1.40 (m, 4H), 1.73 (m, Hz, 4H), 4.07 (t, J = 6.6 Hz, 4H), 7.13 (d, J = 9.0 Hz, 4H), 7.86 (d, J = 9.0 Hz, 4H), 9.09 (br s, 4H), 9.28 (br s, 4H); ^{13}C NMR (150 MHz, $(CD_3)_2SO$) δ 25.83, 28.88, 29.11, 29.36, 68.53, 115.17, 119.65, 130.62, 163.51, 165.14.

Gel shift assay. The 5' end labeled $(CUG)_4$ RNA was annealed in 250 mM NaCl, 25 mM $MgCl_2$, and 75 mM Tris pH 7.5 by heating at 95 °C for 3 min and was cooled on ice for 5 min, BSA and heparin were added, and the reaction cooled another 5 min. MBNL1 was then added to the RNA and incubated at room temperature (RT) for 10 min. Pentamidine was added, followed by another 5 min RT incubation, followed by the addition of BPB. The final reaction volume of 10 μL was incubated at RT for 25 min. Final reaction conditions were 0.5 nM $(CUG)_4$, 175 mM NaCl, 20 mM Tris pH 7.5, 5 mM $MgCl_2$, 2 mg mL^{-1} BSA, 0.1 mg mL^{-1} heparin, 1.25 mM β-mercaptoethanol, 1.25% (v/v) glycerol, and 250 nM MBNL1. Then 3 μL of the reaction was run on a 5% (w/v) 80:1 pre-chilled polyacrylamide gel with 0.5x TB. Gels were run at 4 °C at 170 V for 35 min, dried, and autoradiographed. IC_{50} values were calculated as previously described (Warf, 2009). In brief, gels were quantified using ImageQuant (Molecular Dynamics). The percent of RNA bound was determined by taking the ratio of the RNA:protein

complex (including any well shifting) to the total RNA, per lane. IC_{50} values were determined with KaleidaGraph (Synergy) software using the equation

$$Y = \frac{m_3}{1 + \left(\frac{m_0}{m_1}\right)^{m_2}}$$

where m_0 = small molecule concentration, $m_1 = IC_{50}$, m_2 = Hill coefficient, and m_3 = fraction of MBNL1 bound without small molecule present. Errors were determined by calculating the standard deviation of triplicate data.

Splicing analysis in cell culture. Splicing was performed as described previously (*20*) with the following changes. Approximately 2 x 10^5 cells were plated in 6 well plates and transfected 24–36 h later. After transfection, cells were incubated in OPTI-mem (GIBCO) for 6 h and then washed with 1x PBS and placed in DMEM with GLUTAMAX (GIBCO) supplemented with 10% FBS (v/v) (GIBCO) along with pentamidine analogs. Cells were harvested after 16–18 h and RNA was isolated immediately using an RNeasy kit (QIAGEN). After DNase treatment, RNA was reverse transcribed with SuperScript II and a plasmid specific reverse primer. This cDNA was then subjected to PCR (22 rounds for *INSR* or 24 rounds for *TNNT2*). Resulting PCR products were then run on a 6% (w/v) 19:1 native polyacrylamide gel containing 0.5x TB at 300 V for 90 min. The gel was then stained with 1x SYBR I dye (Applied Biosystems) in 0.5xTB for 15 min. Quantification of bands was performed using the Alpha Imager HP software from Alpha Innotech. EC_{50} values were determined by performing linear regression analysis of the data using Excel (Microsoft), and determining the concentration would result in 50% splicing rescue, calculated as halfway between wild type and disease state levels of exon inclusion.

The following primers were used. For the *TNNT2* minigene, the RT primer was 5' AGC ATT TAG GTG ACA CTA TAG AAT AGG G. The forward primer for PCR was 5' GTT CAC AAC CAT CTA AAG CAA GAT G and the reverse primer was 5' GTT GCA TGG CTG GTG CAG G. For the *INSR* minigene, the RT primer was 5' GCT GCA ATA AAC AAG TTC TGC. The forward primer for PCR was 5' CGA ATT CGA ATG CTG CTC CTG TCC AAA GAC AG, and the reverse primer was 5' TCG TGG GCA CGC TGG TCG AG. The *TNNT2* minigene was a gift from the lab of Thomas Cooper and the *INSR* minigene was a gift from Nicholas Webster.

Heptamidine treatment of mice. Homozygous HSA^{LR} transgenic mice in line 20b (FVB inbred background) were previously described (*63*). Gender-matched mice of 10–14 wks age were treated with heptamidine at the indicated dose by daily intraperitoneal injection for 7 d. Control mice received 5% glucose injections. Mice were sacrificed 1 d after the final injection and vastus (quadriceps) muscle was obtained for splicing analysis. Mice in the withdrawal (WD) group were treated with 30 mg kg^{-1} heptamidine for 7 d and then left untreated for 10 d before sacrifice. Mice were sacrificed 1 d after final treatment and vastus muscle was obtained for splicing analysis. RNA was isolated, reverse transcribed, and amplified by PCR, and analyzed on agarose gels using a fluorimager as previously described (*20*).

Electromyography. Electromyography was performed under general anesthesia as described previously (*12*). Briefly, at least 10 needle insertions were performed in vastus muscle and myotonic discharges were graded on a four point scale: 0, no myotonia; 1, occasional myotonic discharge in less than 50% of needle insertions; 2, myotonic discharge with more than 50% of insertions; and 3, myotonic discharge with nearly all insertions.

CHAPTER IV

UTILIZING THE GAAA TETRALOOP/RECEPTOR TO FACILITATE CRYSTAL PACKING AND DETERMINE THE STRUCTURE OF A CUG RNA HELIX

I was the primary contributor to the experiments described in the following chapter and did all of the writing. Dr. Jeremy Lohman aiding in determining the structure trCUG-3, and Dr. J. Andrew Berglund was the principal investigator of this work.

INTRODUCTION

Microsatellite expansion diseases are the result of aberrant expansion of short repeating segments of DNA (between 1–10 bases) (*67*). These expansions can occur in either in coding regions where they result in abnormal proteins (*68*), or in noncoding regions of the genome where upon transcription, act through a toxic RNA gain-of-function mechanism (*40*, *69*). Myotonic dystrophy (DM), which is the most common form of adult onset muscular dystrophy, affecting ~1 in 8000 individuals, is the result of microsatellite expansions (reviewed in (*40*, *59*, *67*)). DM type 1 (DM1) is caused by a CTG expansion in the 3' untranslated region in the *DMPK* gene (*8*). When transcribed to RNA, these CUG repeats form stable stem loops, which are capable of sequestering RNA binding proteins within nuclear foci (*9*, *10*). One family of sequestered proteins that co-localizes with expanded CUG repeats, the Muscleblind-like (MBNL) proteins, are regulators of alternative splicing, resulting in the mis-splicing of target transcripts, leading to disease symptoms (*14-18*).

It is the expanded CUG RNA which is toxic in DM1, and a handful of CUG-repeat structures have been published showing that, despite containing a U-U mismatch every third base pair, the repeats formed an essentially A-form structure, with the U-U mismatches forming only one hydrogen bond without distorting the backbone (*33*, *35*, *36*). However, two of these structures contained some degree of disorder, and all of the structures had very tight crystal packing, leaving no potential room for co-crystallization with ligands (either small molecules or short peptides). In order to overcome these issues, we decided to use an RNA crystallization motif in order to design specific intermolecular contacts and minimize close packing, and introduce "free spaces" within the crystal.

We chose the GAAA tetraloop and its conserved 11-nucleotide receptor as our engineered crystal contacts. This tetraloop/receptor was initially identified as a tertiary motif in Group I and II self-splicing introns (*70-72*). The three adenosines in the loop flip outward and form specific hydrogen bonds with the 11-nucleotide receptor (*72*), making it an ideal candidate for engineering crystal contacts. This motif has been utilized to aid in the crystallization of larger RNA constructs, such as the hepatitis delta virus, among others (*73*, *74*). Here, we utilized the tetraloop/receptor to crystallize a short RNA helix composed of CUG repeats allowing us to obtain a strongly diffracting crystal. Additionally, we were able to use the previously solved structure of the tetraloop/receptor as our search model, allowing us to quickly and easily solve the structure using molecular replacement.

MATERIALS and METHODS

Purification of RNA. RNA was purchased from Dharmacon and desalted per their instructions. RNA was then resuspended in 1x denaturing dye (25% w/w formamide, 1x

TBE, 0.1% bromophenol blue, 0.1% xylene cyanol) and run on a denaturing 8% polyacrylamide (19:1) gel with 7 M urea. RNA was located using UV shadowing, excised and eluted overnight using the Elutrap Electroelution System (Whatman) in 1x TBE. RNA was then ethanol precipitated and resuspended in ddH_2O and desalted using a Micro Bio-Spin column with Bio-Gel P2 Gel (Bio-Rad). RNA was 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 $MgCl_2$.

RNA crystallization. The RNAs were annealed by heating to 70 ºC for 5 min and rapidly cooled to 4 ºC. RNAs were screened for crystallization using the Mosquito nano-liter high throughput robot (TTP labtech) and the Natrix screen (Hampton Research) by the hanging drop vapor diffusion method at room temperature. Promising hits were screened in 4 μL hanging drops. The best crystals of trCUG-3 were grown from a mixture of 2 μL RNA solution and 2 μL well solution containing 4 mM $MgSO_4$, 50 mM Tris pH 8.5, and 30% (w/v) 1,6-hexanediol. Crystals appeared in approximately one week.

Data collection. Crystals of at least 0.02 mm in the smallest dimension were mounted in rayon loops and flash frozen in liquid nitrogen. Experimental data were collected at Advanced Light Source BL 5.0.1 under a cryostream. The x-ray data were integrated, merged, and scaled using the HKL-2000 program suite (*75*), and converted to structure factors using the CCP4i GUI (*76*) for the CCP4 program suite (*77*). Data collection statistics are listed in Table 2.

Table 2. Summary of data collection and refinement statistics

Measurement		**Value**		
Spacegroup		R3		
Unit cell dimensions	**a, b, c (Å)**	70.34	70.34	68.24
	α, β, γ (°)	90	90	120
Resolution range, Å		22.75 – 1.95		
Total number of reflections		98787		
Number of unique reflections		8621		
Average redundancy		5.5		
% completeness		99.3 (98.9)		
I/σI		37.0 (2.2)		
$\boldsymbol{R_{merge}}$ [a]		0.068 (0.60)		
Average B-factors [atoms]				
	nucleotides	54.58 [744]		
	solvent	60.53 [44]		
$\boldsymbol{R_{free}}$ [b]		0.266		
$\boldsymbol{R_{work}}$ [b]		0.208		

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

Structure determination. Structure was solved using the molecular replacement method by the program molrep (*78*), a part of the CCP4 program suite (*77*), with the tetraloop/receptor from 1GID, the structure of group I ribozyme domain (*72*), as the search model. The model was manually rebuilt using Coot (*79*), and refined using refmac (*80*). Coot and PyMOL (*81*) were used to generate figures. Refinement statistics are listed in Table 2. Structural parameters were calculated using 3DNA (*82*).

RESULTS

Design of RNA constructs for crystallization. In order to overcome crystallographic disorder present in previous CUG repeat structures (*33*, *35*), we designed ten different constructs utilizing the GAAA tetraloop/receptor to facilitate intermolecular contacts (Figure 15). Each construct contained either four or six CUG repeats, one or two stabilizing base pairs, the receptor, a short upper helix, and the GAAA tetraloop. The upper helix was kept short in order to favor intermolecular rather than intramolecular interactions. Only trCUG-8 failed to form any crystals, possibly due to the 3' overhanging nucleotide. Of the remaining constructs, only trCUG-2, 3, 6, and 10 grew crystals of sufficient size to mount. Neither trCUG-2 nor trCUG-10 diffracted to better than 10Å. Both trCUG-3 and trCUG-6 diffracted to better than 3Å resolution; however, the 3' overhanging G of trCUG-6 was disordered, and both structures looked essentially the same.

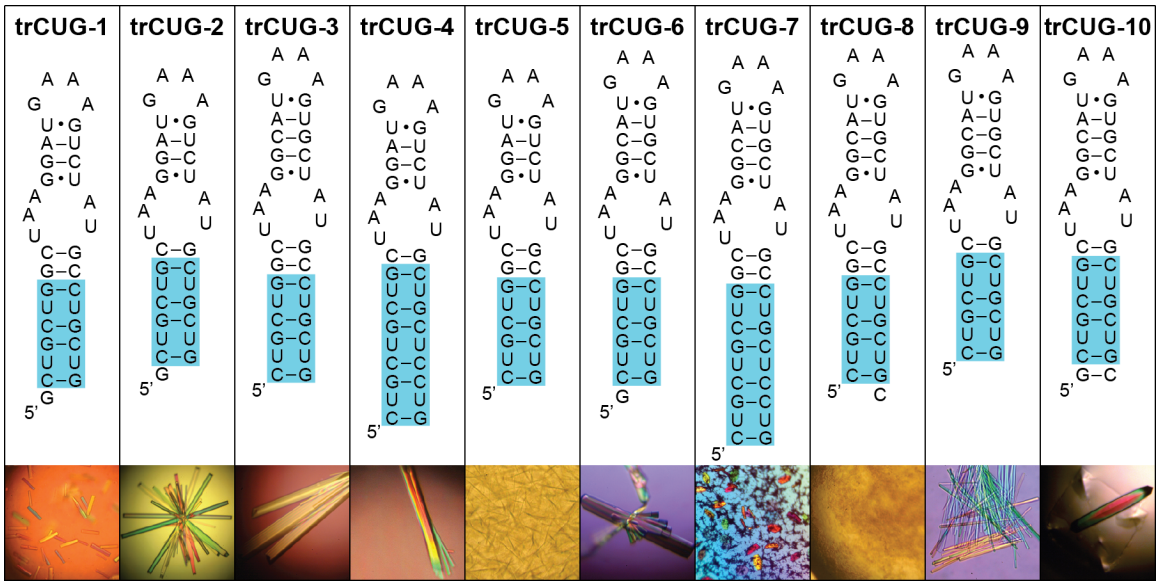


Figure 15. Sequence of RNAs used to identify CUG repeat containing crystals that diffracted to high resolution using the tetraloop/receptor crystallization module. Each RNA construct consists of four or six CUG repeats (highlighted in blue), the 11-nucleotide receptor, a short upper stem, and the GAAA tetraloop. Below each sequence is a representative picture of how well each RNA crystallized. The only sizable trCUG-6 crystal grew off a fiber.

Structure determination using tetraloop/receptor as a model for molecular replacement. Since the sequence, and presumably the structure, of the tetraloop/receptor in trCUG-3 was the same as the tetraloop/receptor in the group I ribozyme domain structure determined by Cate *et al*. (*72*) (with the exception of the first base pair, which was reversed), this portion of their structure was used for molecular replacement. The 11-nucleotide receptor and the 4-nucleotide loop (plus the two adjacent base pairs) was successfully used as our search model. The un-modeled base pairs were clearly visible in the maps and the nucleotides were easily modeled into the density (Figure 16A).

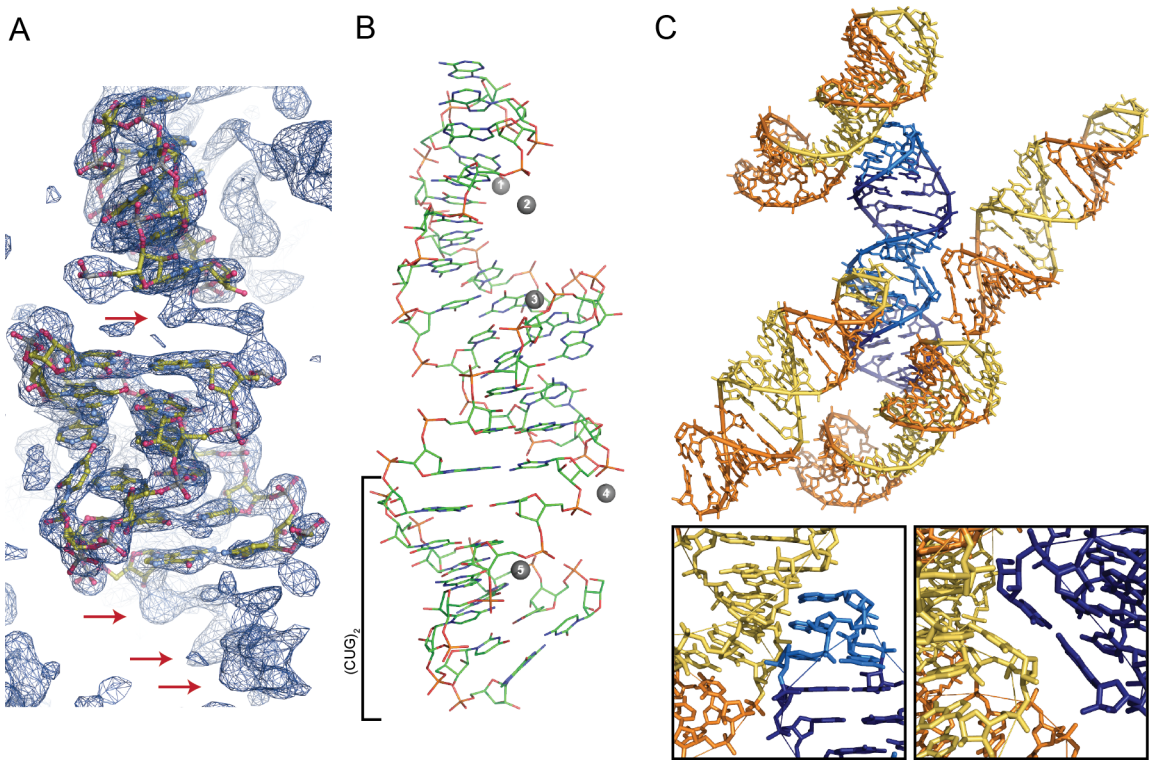
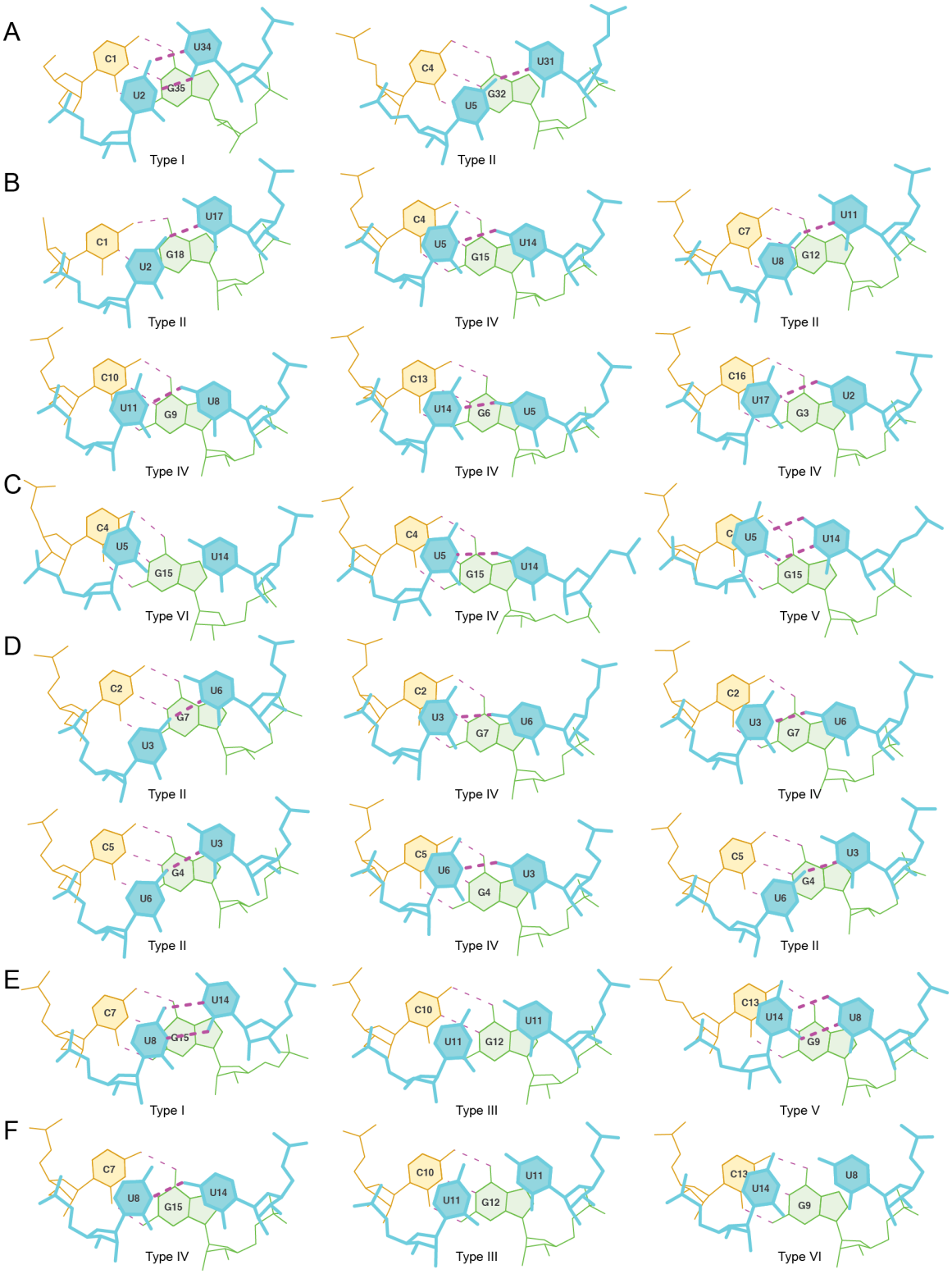


Figure 16. Structure and crystal packing of trCUG-3. (A) Molecular replacement solution and 2Fo-Fc map (contoured at 1.5 σ) after one round of refinement. The density of the base pairs not yet modeled, both between the tetraloop/receptor and directly below the receptor, are clearly observable in the density (red arrows). (B) Secondary structure of trCUG-3. The $(CUG)_2$ region, containing four CUG repeats at the base of the hairpin, is denoted by a bracket. Magnesium ions are represented by numbered gray spheres. Waters not shown. (C) Crystal packing of trCUG-3. Each RNA interacts with four other RNAs in the structure: at the loop, at the receptor, on the opposite side of the receptor and at the bottom of the helix. Tetraloop and receptor are denoted by lighter colors. Inset gives a closer view of each interaction.

Five peaks were observed in the maps that were unlikely to be waters due to geometric considerations, and based on previous structures of the tetraloop/receptor (*72*, *83*, *84*) and their proximity to the phosphate backbone, we modeled the peaks as magnesium ions (Figure 16B). The final model has 35 nucleotides, 44 water molecules, and five magnesium ions with an R_{work} and R_{free} of 0.208 and 0.266, respectively.

Structure of trCUG-3. The 35-nucleotide trCUG-3 RNA crystallized in space group *R3*. As designed, the RNA formed a hairpin with a GAAA tetraloop, a short five base pair stem, the tetraloop/receptor, and an eight base pair CUG-containing stem (Figure 16B). The upper stem begins and ends with a G•U wobble base pair. The lower stem contains two different U-U mismatches, while the G-C base pairs formed Watson-Crick interactions. Each RNA makes specific crystal contacts with four other RNAs: two of these interactions were predicted tetraloop/receptor interactions, and two were the unanticipated stacking of the base of the helix against the opposite side of the receptor of a symmetry mate (Figure 16C).

Figure 17 (next page). Comparison of U-U mismatches in the context of CUG repeats. (A) Lower and upper U-U mismatch of trCUG-3. The upper U5-U31 mismatch forms the more typical (in the context of CUG repeats) one hydrogen bond, while the lower U2-U34 mismatch forms two hydrogen bonds, shortening the C1'-C1' distance. (B) U-U mismatches of the de-twinned $(CUG)_6$ structure. All U-U mismatches in this structure form one hydrogen bond, maintaining the C1'-C1' distance typical of a Watson-Crick interaction. (C) U-U mismatches from the NMR structure of CCGCUGCGG, forming zero, one, or two hydrogen bonds. (D) U-U mismatches from $G(CUG)_2C$ crystal structure. The first column contains the U-U mismatches in duplex A+B, the second contains C+D and the third contains E+E*. Every mismatch within the $G(CUG)_2C$ crystal structure forms one hydrogen bond. (E-F) U-U mismatches from $UUGGGC(CUG)_3CUCC$ pdb entry 3SYW (E) and 3SZX (F).


A
C1
U34
U2
G35
Type I
C4
U31
U5
G32
Type II
B
C1
U17
U2
G18
Type II
C4
U5
G15
U14
Type IV
C7
U11
U8
G12
Type II
C10
U11
G9
U8
Type IV
C13
U14
G6
U5
Type IV
C16
U17
G3
U2
Type IV
C
C4
U5
G15
U14
Type VI
C4
U5
G15
U14
Type IV
C
U5
G15
U14
Type V
D
C2
U6
G7
U3
Type II
C2
U3
G7
U6
Type IV
C2
U3
G7
U6
Type IV
C5
U3
G4
U6
Type II
C5
U6
G4
U3
Type IV
C5
U3
G4
U6
Type II
E
C7
U14
U8
G15
Type I
C10
U11
U11
G12
Type III
C13
U14
U8
G9
Type V
F
C7
U8
G15
U14
Type IV
C10
U11
U11
G12
Type III
C13
U14
U8
G9
Type VI

Within the CUG-containing stem, two different conformations of the U-U mismatch were observed (Figure 17A). The lower U2-U34 mismatch formed two hydrogen bonds, which resulted in a shorter C1'-C1' distance (8.6 Å) as compared to the upper U5-U31 mismatch, which formed one hydrogen bond and had a C1'-C1' distance of 10.6 Å (Table 3). Both mismatches were inclined towards the minor groove ($\lambda 1 < 54°$), and as a result, do not appear to stack with the preceding C. The incline was more severe in the U5-U31 mismatch, which is most likely a result of the single hydrogen bond.

Examining the overall structure of the CUG portion of the lower helix (hereafter referred to as $(CUG)_2$) reveals that it is primarily an A-form helix. The average helical rise of $(CUG)_2$ is 2.39 Å (Table 4), which is closer to an A-form conformation (2.83 Å) than a B-form conformation (3.29 Å) (*85*). The average helical twist is 35.9°, which is greater than the expected 32.5° for A-form, and not much less than the expected 36.5° for B-form. Base pairs in $(CUG)_2$ are also steeply inclined (18.6°) as is expected for A-form helices. Local dimer step parameters, including roll, twist, slide, and most importantly z_P and $z_P(h)$, also match well with average A-form values (Table 4). Interestingly, the structure did not seem to be affected, beyond the first CU/UG step, by the crystal packing interaction that occurred at the base of the helix (Figure 16C). This base pair step was strongly tipped (20.6°), tilted (-11.3°), and rolled (19.4°) as compared to the other base pairs in the structure.

Table 3. Comparison of U-U mismatches

		C1'-C1' (Å)	1st hbond (Å)	2nd hbond (Å)	λ I (°)	λ II (°)	Incline[1]	Shear (°)	Stretch (Å)	Stagger (Å)	Buckle (°)	Propeller (Å)	Opening (°)	Type
trCUG-3	**U2-U34**	8.6	2.8	2.8	44.8	75.9	minor	-2.47	-1.89	0.26	-7.12	-16.4	11.6	Type I
	U5-U31	10.6	2.7	-	28.2	55.8	minor	-2.63	-1.33	-0.1	5.62	-7.12	-25.98	Type II
$(CUG)_6$ ***(35)***	**U2-U17**	10.3	2.7	-	34.1	56.7	minor	-2.32	-1.38	-0.18	0.89	-9.98	-21.23	Type II
	U5-U14	10.6	3	-	53.8	33.1	major	2.27	-1.17	-0.36	-6.03	-15.03	-22.06	Type IV
	U-8-U11	10.3	2.7	-	32.4	56.9	minor	-2.5	-1.48	0.45	2.36	-11.27	-21.02	Type II
	U11-U8	10	2.9	-	57.8	40.4	major	1.71	-1.29	-0.44	-1.86	-12.2	-11.93	Type IV
	U14-U5	10.5	2.7	-	55.9	32.4	major	2.65	-1.29	-0.23	6.92	-11.44	-28.01	Type IV
	U17-U2	10.3	3	-	58.2	39.2	major	2.04	-1.11	-0.15	-4.69	-14.81	-14.8	Type IV
Disney NMR ***(37)***	**U5-U14 (0)**	10.7	-	-	71.2	32	major	3.57	-0.95	-0.05	-13.48	-9.09	-5.62	Type VI
	U5-U14 (1)	10.6	2.9	-	61.1	28.9	major	3.03	-1.14	-0.09	-22.07	-6.46	-24.3	Type IV
	U5-U14 (2)	8.9	2.9	2.9	73.3	48.3	major	2.24	-1.74	0.02	12.56	-17.25	4.58	Type V
Kiliszek A-B ***(35)***	**U3-U6**	10.7	2.8	-	26.2	57.9	minor	-2.82	-1.19	-0.47	12.51	-8.17	-31.5	Type II
	U6-U3	10.3	2.9	-	31.6	67.7	minor	-3.03	-1.31	-0.13	14.01	-8.36	-16.37	Type II
Kiliszek C-D ***(35)***	**U3-U6**	10.9	2.8	-	56.1	22.6	major	2.98	-1.22	-0.44	-4.51	-8.69	-34.42	Type IV
	U6-U3	10.2	2.6	-	59.2	32.4	major	2.59	-1.51	-0.28	0.57	-15.81	-21.16	Type IV
Kiliszek E-E* ***(35)***	**U3-U6**	10.6	2.6	-	52.7	29.1	major	2.38	-1.31	-0.48	-4.07	-11.84	-30.56	Type IV
	U6-U3	10.6	2.6	-	29.1	52.7	minor	-2.38	-1.31	-0.48	4.07	-11.84	-30.56	Type II
Disney 3SYW ***(36)***	**U8-U14**	8.8	3	3	46.3	72.4	minor	-2.2	-1.7	0.21	0.27	-14.34	10.06	Type I
	U11-U11	9.9	-	-	50.8	50.8	none	0.01	-1.17	0.07	0.11	-5.73	-8.25	Type III
	U14-U8	8.8	2.9	3	72.5	46.3	major	2.2	-1.71	0.21	-0.76	-14.31	10	Type V
Disney 3SZX ***(36)***	**U8-U14**	10.5	2.7	-	53.9	35.3	major	1.85	-1.16	-0.11	-5.24	-8.68	-29.19	Type IV
	U11-U11	10	-	-	46	53	none	0.14	-0.99	0.18	1.98	-10.74	-0.21	Type III
	U14-U8	10.5	-	-	58.8	48.1	major	1.3	-0.7	-0.3	-2.99	-11.12	-8.97	Type VI

Classification	# of hbonds	Incline	Amount
Type I	2	minor	2
Type II	1	minor	6
Type III	0	none	2
Type IV	1	major	9
Type V	2	major	2
Type VI	0	major	2

[1]Direction of incline was determined by considering the difference between λI and λII. If (λI – λII) < -8, then the pair was considered to be inclined toward the minor groove; however, if the difference was > 8, the pair was considered to be inclined toward the major groove. Any values between -8 and 8 were considered to be not inclined.

Table 4. Comparison of CUG repeat helical parameters

	Roll	Twist	Slide	Rise	Inclination	Helical twist	x-displacement	Helical rise	Zp	Zp(h)
B Form[1]	0.60	36.00	0.23	3.32	2.10	36.50	0.05	3.29	-0.36	-0.02
A Form[1]	8.00	31.10	-1.53	3.31	14.70	32.50	-4.17	2.83	2.24	4.19
(CUG)2	9.17	33.48	-1.34	3.10	18.57	35.89	-4.70	2.39	2.35	4.62
(CUG)6	9.16	32.77	-1.90	3.25	16.98	34.54	-5.08	2.49	2.69	4.81
A-B	6.55	31.70	-1.80	3.32	12.41	33.25	-4.55	2.96	2.52	4.10
C-D	8.56	33.11	-1.59	3.30	6.93	34.78	-4.58	2.66	2.40	4.57
E-E*	6.78	32.77	-1.90	3.32	14.12	34.05	-5.03	2.69	2.52	4.32
3SYW	8.51	28.68	-1.73	3.16	18.76	30.62	-5.27	2.39	2.27	4.64
3SZX	8.63	32.33	-1.64	3.21	14.93	33.55	-4.26	2.71	2.52	4.46

[1]A and B-form structural parameters from reference *(82)*

Within the structure, we identified five suspected magnesium ions. Two were near the GAAA tetraloop, two more flanked the 11-nucleotide receptor, and the final magnesium ion was near the center of the CUG portion of the lower helix (Figure 16B). Only the third and fourth magnesium ions were close enough to the RNA to be making direct contacts: the rest were making interactions via water molecules. Although none of the suspected magnesium ions appeared to be fully coordinated, the first and second magnesium ions near the loop interacted with four water molecules.

DISCUSSION

Comparison to other CUG helices. To date, there have been four other crystal structures of CUG repeat RNA crystal structures published (*33*, *35*, *36*). The first, and longest, to be crystallized was a $(CUG)_6$ duplex (*33*). Unfortunately, this construct crystallized with twofold translational disorder, complicating analysis. The next structure to be released was another duplex: $G(CUG)_2C$ (*35*). This structure contained three different duplexes within the same unit cell: the A+B duplex, the C+D duplex, and the E+E* duplex, which were related to one another by crystallographic symmetry. This publication also de-twinned the data from Mooers *et al*. The final two structures are different conformations of a third duplex: $UUGGGC(CUG)_3UCC$ (*36*). Using 3DNA, we compared the helical parameters of $(CUG)_2$ to the $(5'CUG/3'GUC)_x$ portions of these six structures. (For simplicity, the de-twinned $(CUG)_6$ data were used.)

Overall, all seven structures appeared to be more similar to an A-form helix than a B-form helix, especially when parameters that demonstrate the greatest distinction between A-form and B-form are considered (roll, slide, inclination, x-displacement, z_P, and $z_P(h)$) (Table 4). Although $(CUG)_2$ from trCUG-3 has the greatest helical twist of all the

published CUG helices (35.9°), it also has the smallest helical rise (2.39 Å). In general, it appears that the $(CUG)_2$ model is consistent with existing models.

U-U mismatches are dynamic. In addition to comparing helical parameters of other CUG structures, we also examined all the U-U mismatches, in the context of CUG repeats that were available (Figure 17, Table 3). U-U mismatches from all the CUG crystal structures (*33*, *35*, *36*) and an NMR structure (*37*) were compared. By examining these mismatches, it is clear that the U-U mismatch in the 5'CUG/3'GUC motif is dynamic, and at least six unique conformations have been captured, considering just potential number of hydrogen bonds and direction of inclination (towards the major or minor groove, or not inclined). There is one other potential type of U-U mismatch that has not been observed in the currently available data, which is a mismatch with no hydrogen bonds inclined towards the minor groove. It is likely this type of mismatch occurs (as a counterpart to type VI, which also forms no hydrogen bonds, but is inclined toward the major groove) and has simply not been observed/reported at this time.

Within the trCUG-3 structure, we find two of the six types: type I and II (Figure 17A, Table 3). The type I mismatch forms two hydrogen bonds, shortening the C1'-C1' distance, and is inclined towards the minor groove. One other type I mismatch has been observed, in $UUGGGC(CUG)_3UCC$ (Figure 17E) (*36*), suggesting this is an actual conformation and not simply due to potential distortion of the helix of $(CUG)_2$ due to the crystal packing at the base of the trCUG-3 helix. The other mismatch we observed was a type II mismatch, the second most frequent type of U-U mismatch. The type II mismatch is also inclined towards the minor groove but forms only one hydrogen bond. It is observed twice in $(CUG)_6$ (Figure 17B), twice in the $G(CUG)_2C$ A+B duplex, and once in

the E+E* duplex (Figure 17D). The most commonly observed mismatch, type IV, is the counterpart to type II, and forms one hydrogen bond while being inclined toward the major groove. It is found four times in $(CUG)_6$ (Figure 17B), in the NMR structures (Figure 17C), twice in the $G(CUG)_2C$ C+D duplex, once in the E+E* duplex (Figure 17D), and once in the $UUGGGC(CUG)_3UCC$ duplex (Figure 17F). Of course, in solution, it is probable that the U-U mismatches found in the 5'CUG/3'GUC motif sample all these conformations, and that these "types" of U-U mismatches are not discrete, but rather found in a continuum of different states. These data are consistent with the flexibility observed in the NMR structure of the CCGCUGCGG duplex (Figure 17C) (*37*).

Metal ions in structure. Within the trCUG-3 crystal structure, we identified five potential magnesium ions. It is unsurprising that magnesium ions would be identified within the structure, especially the tetraloop/receptor region, since it has previously been shown that tetraloop/receptor interaction is dependent on magnesium (*86*) or other divalent metal ions (*87*). Several other structures of the GAAA tetraloop/receptor have been published, so we compared our structure and metal ion sites with several previously published structures (Figure 18). Comparing trCUG-3 and 1GID (*72*), we observe our first three magnesium ions surrounding the cobalt hexamine ion found near the GAAA loop (Figure 18B). Our fourth magnesium ion near the receptor has a nearby counterpart in 1GID. When compared to 1HR2 (*83*) (Figure 18C), we see a very similar picture: our first three magnesium ions surround one magnesium ion near the loop. However, the second, third, and fourth magnesium ions all have nearby counterparts in the 1HR2

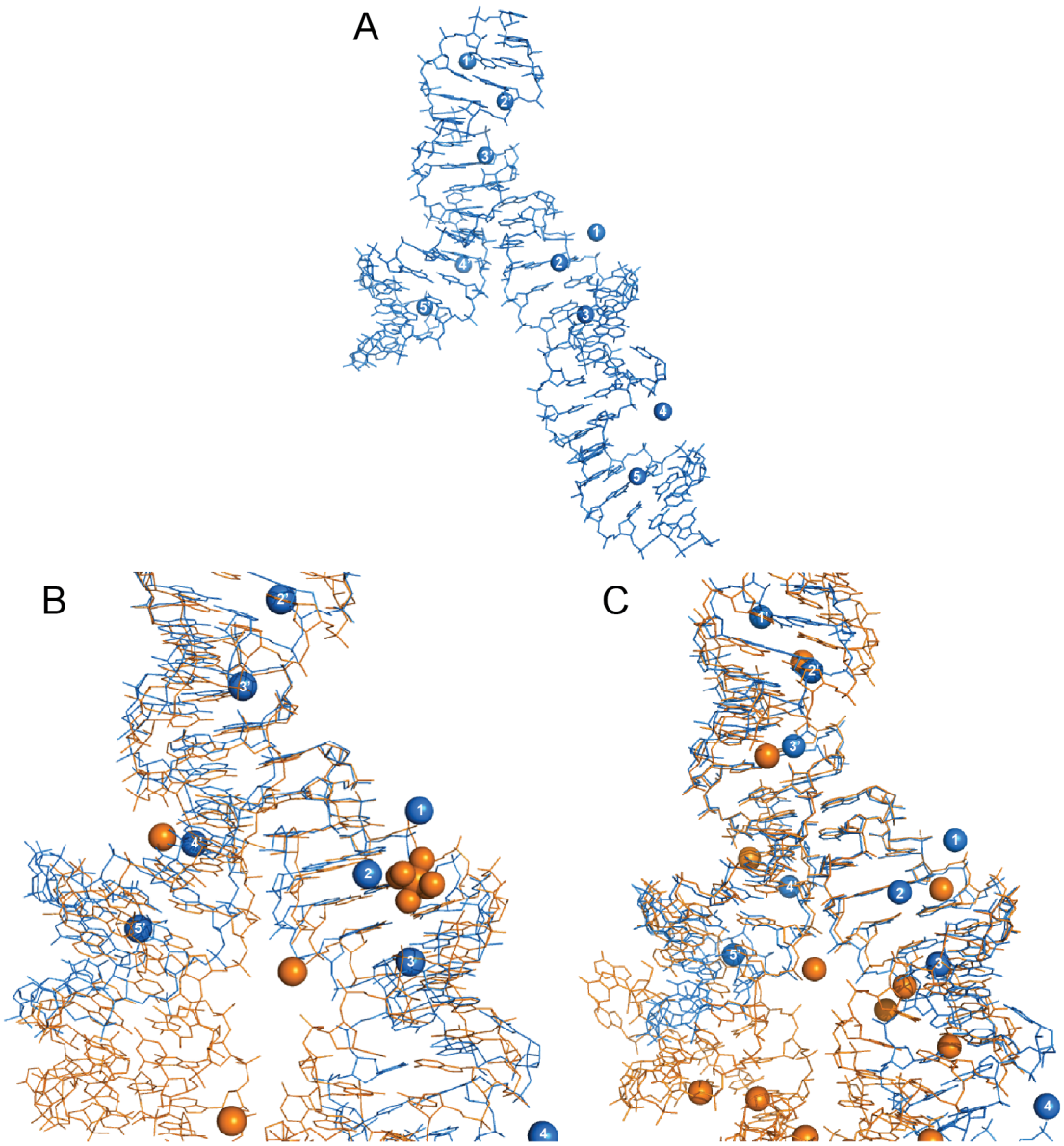


Figure 18. Alignment of trCUG-3 with two different GAAA tetraloop/receptor structures. (A) Structure of trCUG-3 and a symmetry mate used for comparison to other tetraloop/receptor structures. Magnesium ions are numbered. (B) Alignment of the tetraloop/receptor of trCUG-3 (blue) with that of 1GID (orange). The first three magnesium ions surround the cobalt hexamine ion found in 1GID. The fourth magnesium ion near the receptor is near a corresponding magnesium ion in 1GID. (C) Alignment of the tetraloop/receptor of trCUG-3 (blue) with that of 1HR2 (orange). Again, the first three magnesium ions in trCUG-3 surround a single magnesium ion found in 1HR2. The second, third and fourth magnesium ions also have counterparts near the receptor.

structure. The additional ions observed in our structure may be due to the close proximity of our tetraloop and receptor as compared to either 1GID or 1HR2, or it could simply be that the magnesium ions are not important for making specific contacts with the RNA, but rather important for allowing close contact between RNA strands.

It is clear that utilizing the GAAA tetraloop/receptor was a successful strategy for crystallizing a short helix and avoids the pseudosymmetry problems associated with duplex RNA. Additionally, examining the crystal packing of this particular construct, we observe a large "open" space in the crystal lattice (Figure 19), which could potentially allow co-crystallization of small molecules, short peptoids, and possibly small zinc-finger domains in complex with the CUG stem. This GAAA tetraloop/receptor framework can also be utilized to crystallize other short repeat RNA stems, such as the CCUG repeats responsible for DM2 (*11*), CAG repeats responsible for many spinocerebellar ataxias, and CGG repeats responsible for fragile X-associated tremor ataxia syndrome, among many others (reviewed in (*88*)). We have already crystallized a CCUG helix utilizing the tetraloop/receptor construct, which diffracted to a resolution of better than 3 Å (data not shown).

We have crystallized a CUG containing RNA construct that has captured two different orientations of the clearly mobile U-U mismatch. By examining other crystal and NMR structures, it is clear that U-U mismatches, in the context of CUG repeats, can flex between many different conformations, from fully paired with one another, to only sharing a single hydrogen bond, to being too distant to interact with one another (Figure 3). Additionally, these U-U mismatches can go from being inclined towards the

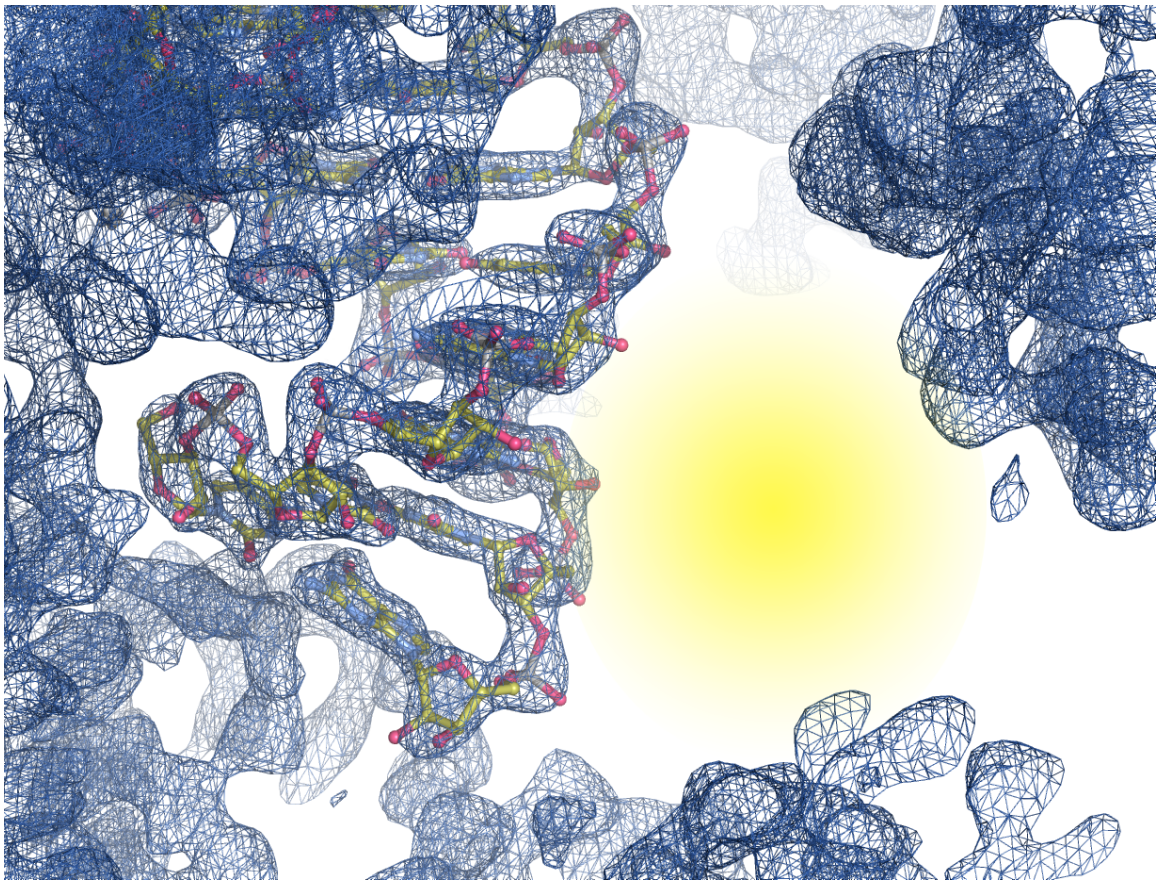


Figure 19. Structure and 2Fo-Fc map contoured at 1.5 σ near the CUG repeat stem. Yellow area highlights large open space near the minor groove of the $(CUG)_2$ stem.

minor groove, through no inclination, all the way to being inclined towards the major groove. This conformational heterogeneity suggests a weak interaction that may play an important role in MBNL proteins gaining access to single stranded CUG repeats in order to bind the Watson-Crick face of the GC dinucleotide as found in the MBNL-RNA crystal structure (3D2S) (*39*).

CHAPTER V

CONCLUSIONS AND FUTURE DIRECTIONS

Pentamidine requires the dye bromophenol blue (BPB) for efficient disruption of the MBNL1/$(CUG)_4$ complex *in vitro*

The initial small scale screen for molecules that could disrupt the MBNL1/$(CUG)_4$ complex was performed using BPB, a dye use to monitor reaction mixtures and progress through the polyacrylamide matrix. What we did not realize during these experiments was that BPB itself was able to compete with the complex with an IC_{50} of approximately 1 mM (Chapter II, Figure 5). When I tested the ability of pentamidine to disrupt the MBNL1/$(CUG)_4$ complex in the absence of BPB, I was not able to observe any competition in the concentration range in which pentamidine and the RNA remained soluble. However, when pentamidine and BPB are used in combination, they function in a synergistic manner (Chapter II, Figure 6). I found that at least 200 μM BPB was required to observe competition under the conditions used for the EMSA.

I was also interested if other phenolsulphonphthalein dyes would be able to function similarly to BPB in disrupting the MBNL1/$(CUG)_4$ complex alone and in conjunction with pentamidine. Of the six dyes tested, only BPB and IPB were able to disrupt the complex on their own. In combination with 50 μM pentamidine, CPB, BPB, and IPB all showed increased efficacy (Chapter II, Figure 7). Overall, it appeared that the activity of the phenolsulphonphthalein dyes was greatly affected by the halogenation, and that the more electronegative the halogen, the less effective the complex disruption.

We suspect that the phenolsulphonphthalein dyes disrupt the protein/RNA complex through interactions with MBNL1 rather than $(CUG)_4$ since BPB and other phenolsulphonphthalein dyes are known to interact with other proteins (*48-51*, *58*, *89*). Indeed, we may be able to reduce the necessary concentration of BPB required to observe efficient MBNL1/$(CUG)_4$ complex disruption by reducing the amount of BSA present in the reaction, which would effectively raise the concentration of "free" BPB. It is clear that pentamidine's mechanism of action is not as simple as we first predicted, since pentamidine requires BPB efficiently disrupt the MBNL1/$(CUG)_4$ complex. Additionally, something else must be occurring in cells, since no BPB is added to the HeLa cell assay, and yet pentamidine is able to reverse mis-splicing events. We decided to further investigate pentamidine's mode of action in the DM1 HeLa cell model.

Pentamidine reduces the amount of $(CUG)_{960}$ transcript within DM1 HeLa model

Due to the BPB complications with the competitive EMSA, and our inability to measure a high affinity interaction between CUG repeats and pentamidine, we began to question if pentamidine was acting via a mechanism other than (or in addition to) direct binding to extended CUG repeats, displacing MBNL. In order to examine other potential mechanisms through which pentamidine could act, I decided to observe $(CUG)_{960}$ levels with and without pentamidine treatment (Figure 20). Pentamidine clearly reduces CUG repeat transcript levels while not affecting GAPDH levels. This suggests that pentamidine is acting specifically on the transfected DNA or its transcript, rather than generally affecting transcription. In order to further examine pentamidine's effect on transcript levels, we are planning to perform an RNA-seq experiment, using both a HeLa cell DM1 model as well as myoblasts from DM1 patients treated with several different

concentrations of pentamidine. Through this transcriptional profiling, we will not only be able to observe changes in transcript levels, but also changes in alternative splicing, alternative promoter usage, and alternative polyadenylation sites.

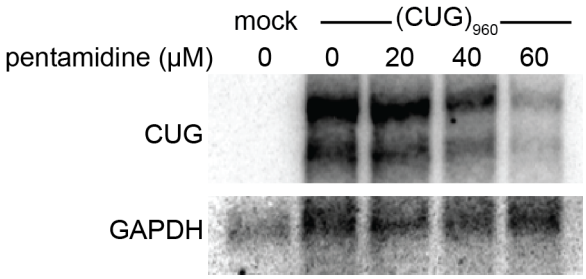


Figure 20. Northern of $(CUG)_{960}$ levels. The first lane is mock transfected HeLa cells which are not expressing CUG repeats. In the cells expressing 960 CUG repeats, levels of CUG transcripts drop as cells are treated with increasing concentrations of pentamidine (0, 20, 40, and 60 μM) while GAPDH levels remain the same.

These data suggest that pentamidine can act through some other mechanism, either instead of or in addition to, binding to CUG repeat RNA and displacing MBNL. Exactly what that mechanism is we have not yet determined. Pentamidine could be binding to CTG repeat DNA and inhibiting transcription, it could be affecting stability of the CUG repeats, it could be interacting with a DNA/RNA hybrid, or some combination of mechanisms of action. More work needs to be done in order to elucidate pentamidine's mode of action *in vivo*. We also want to determine if pentamidine can reduce CUG transcripts in patient myoblasts as well as in the DM1 mouse model.

As methylene linker length increases, pentamidine analogs become more effective

Due to the BPB complications with the competitive EMSA, and the fact that pentamidine reduces $(CUG)_{960}$ transcript levels, we decided to screen our compounds using the HeLa cell splicing rescue assay (competitive EMSA characterization can be found in Appendix A). We observed that analogs with longer methylene linker lengths had better efficacy than those with shorter linkers (Chapter III, Figure 10-12). Interestingly, these results correlate with the EMSA characterizations: as linker length increased, analogs were better at disrupting the MBNL1/$(CUG)_4$ complex. There are several possible reasons for the correlation. It is possible the correlation is simply coincidental: longer linkers may bind $(CUG)_4$ with higher affinity, while longer linkers are more hydrophobic, leading to better ability to cross the cellular membrane, effectively raising the intracellular concentration. There is also a possibility that while optimizing the compounds for $(CUG)_4$, we also accidentally optimized them for CTG repeat DNA, leading to better inhibition of transcription. On the other hand, it may be that pentamidine and its analogs do bind to the CUG repeat RNA and trigger its degradation, so optimization in the EMSA would be observed as better efficacy in the splicing assay. Finally, we have not ruled out that pentamidine is able to bind to expanded CUG repeats in the cell and displace MBNL as was originally hypothesized (*20*) or that there could be a combination of these different modes of action happening simultaneously.

After screening seven pentamidine analogs, we tested our best compound that remained water soluble (heptamidine) in a mouse model of DM1. We observed that heptamidine was able to completely reverse the mis-splicing of the *Clcn1* gene and partially rescue the *Apt2a1* gene in a dose dependent manner. After seven days of

treatment, mice that went untreated for an additional ten days showed a complete reversion to disease levels of splicing, demonstrating that heptamidine treatment was responsible for the change. Additionally, mice that were treated with the highest doses of heptamidine saw significant reduction in the myotonia phenotype. This demonstrates that we can make chemical modifications to our lead drug compounds to identify other potential small molecules that may be more effective and/or less toxic than pentamidine. It also demonstrates that a small molecule can be successful, not only at reversing DM related mis-splicing, but also at alleviating symptoms as well. However, we still need to identify pentamidine's mode of action in order to better guide our future drug design.

Crystallization of a CUG repeat stem by utilizing a crystallization motif

If pentamidine or other small molecule drug targets are interacting with CUG repeat RNA, it would be helpful to crystallize the two together, in order to design molecules with higher affinity and specificity. The currently crystallized constructs contained either disorder (*33*) or were packed too tightly to allow for co-crystallization with a small molecule (*35*, *36*). We successfully overcame both these issues by utilizing the GAAA tetraloop/receptor attached to a CUG repeat stem to engineer crystal contacts. The tetraloop/receptor locked the stem into a specific orientation, eliminating disorder, while also leaving free space along the side of the helix, which could allow for co-crystallization with small molecules or short peptides.

I also compared the trCUG-3 structure with all other available CUG repeat structures and found that although all the structures appear to be essentially A-form, that the U-U mismatches have a wide range of conformations (Chapter IV, Figure 17, Table 3). I identified six different types of U-U mismatches in the context of the 5'CUG/3'GUC

motif based on direction of inclination and number of hydrogen bonds formed (Chapter IV, Figure 17, Table 3). I also predict a seventh type of U-U mismatch that has not yet been reported. I have also used this GAAA tetraloop/receptor motif to facilitate the crystallization of other short duplex RNAs including the CCUG repeats of DM2. It will be interesting to see if the C-U mismatches are similarly dynamic.

APPENDIX

CHARACTERIZATION OF PENTAMIDINE METHYLENE LINKER ANALOGS USING COMPETITIVE EMSA

Introduction

In addition to the *in vivo* splicing assay used to characterize pentamidine analogs discussed in Chapter III, all seven analogs were also characterized with the competitive EMSA, which was used in the small scale screen that discovered pentamidine as a potential drug candidate for DM1. Due to the requirement of BPB in order to observe efficient MBNL1/$(CUG)_4$ complex disruption (discussed in Chapter II), and because pentamidine reduced $(CUG)_{960}$ transcript levels (discussed in Chapter V), these results were not included in the manuscript that composes Chapter III.

In an effort to develop a pentamidine analog with improved physiochemical properties, we undertook a structure activity relationship (SAR) study. Analogs containing three to nine methylene carbons were synthesized and tested using the *in vitro* EMSA assay for their ability to compete with the MBNL1/$(CUG)_4$ complex, and *in vivo* for their ability to rescue the mis-splicing of MBNL regulated minigenes in the HeLa cell splicing assay (discussed in Chapter IV). In general, analogs containing more methylenes had lower IC_{50} values in the EMSA and more efficiently rescued splicing. The most promising compound, heptamidine, was tested for its ability to rescue mis-splicing and myotonia in a transgenic mouse model of DM1.

RESULTS

In vitro efficacy increases with linker length

A common tactic to improve the physiochemical properties of a drug candidate is to vary the hydrophobicity of the molecule, and we utilized this strategy by varying the length of pentamidine's methylene linker. The most polar analog being propamidine, with a three carbon linker, and the least polar being nonamidine, with a linker of nine carbons (Figure 21a). The activities of these compounds were tested using the competitive EMSA in which MBNL1 and $(CUG)_4$ (a small hairpin RNA composed of two CUG repeats, a UUCG cap, and two more CUG repeats) were held constant. As molecules were added, those effective at disrupting the MBNL1/$(CUG)_4$ complex caused an increase in free RNA. We found that propamidine had the lowest activity, with an IC_{50} of 250 ± 10 μM, followed by butamidine (IC_{50} = 85 ± 17 μM), both of which disrupted complexes less effectively than our lead compound pentamidine (IC_{50} = 39 ± 2 μM) (Figure 21b,c Table 5). Analogs with longer linkers showed an improvement over pentamidine. Hexamidine disrupted the MBNL1/$(CUG)_4$ complex with an IC_{50} of 26 ± 2 μM. Heptamidine and octamidine were comparable with IC_{50} values of 20 ± 2 μM and 20 ± 1 μM, respectively. Nonamidine was the most effective in our competitive EMSA with an IC_{50} of 14 ± 1 μM (Figure 21b,c, Table 5). These results showed that analogs containing longer oligomethylene linkers are more effective at disrupting the MBNL1/$(CUG)_4$ complex.

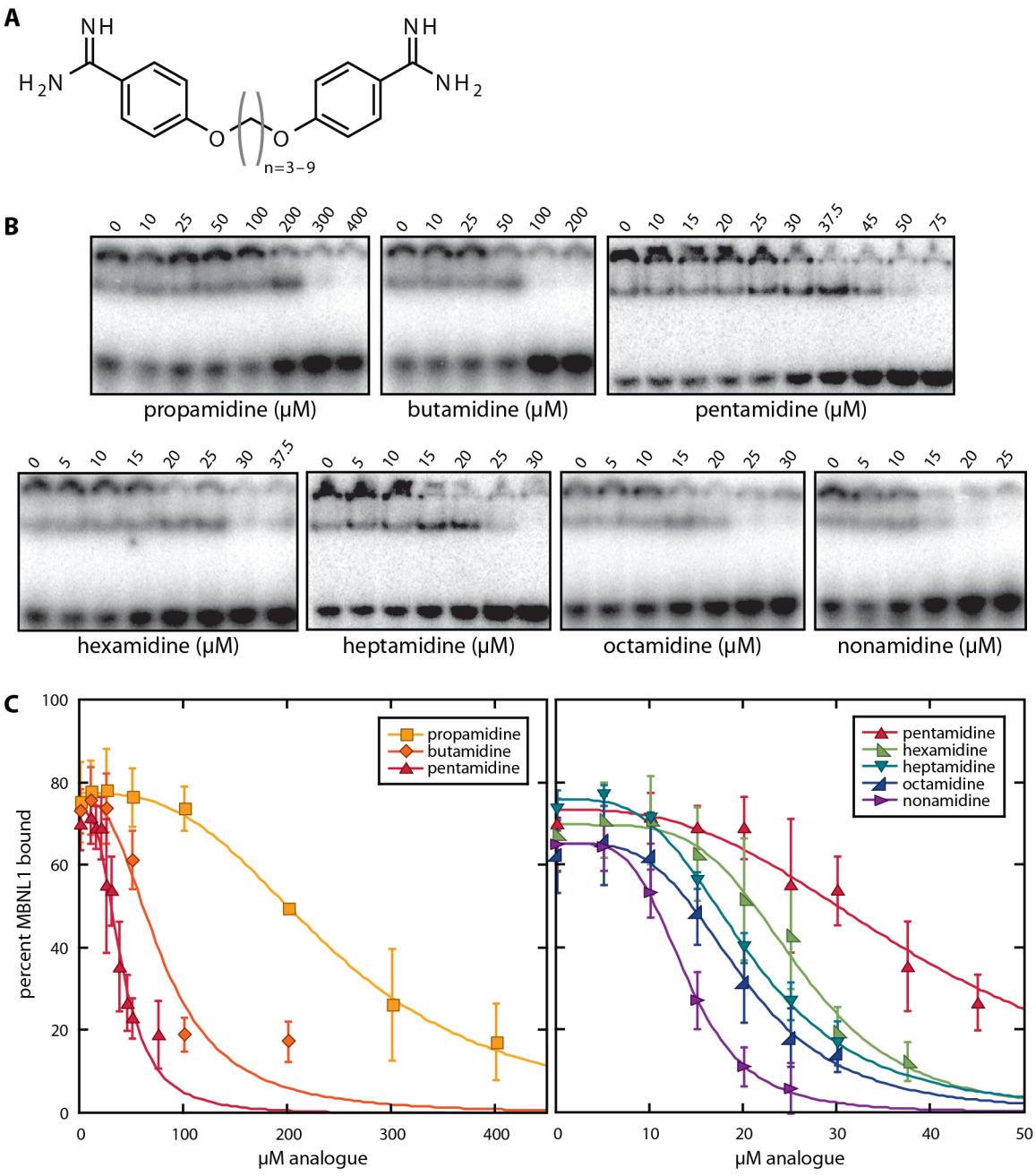
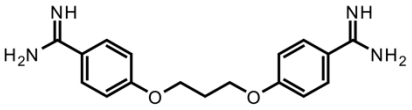
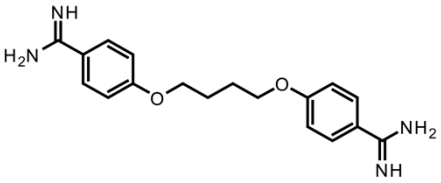
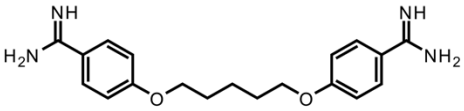
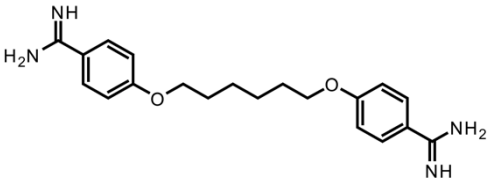
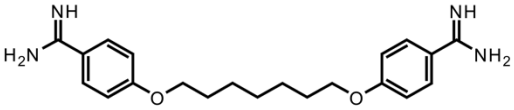
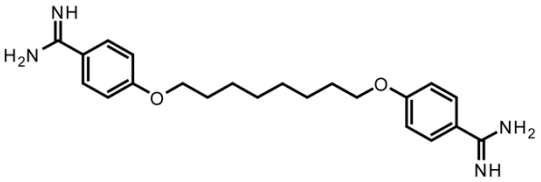
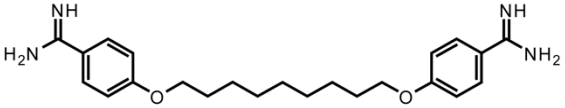


Figure 21. Pentamidine analogs containing longer linkers disrupt the MBNL1/$(CUG)_4$ complex more effectively. (a) Structure of the pentamidine analogs. Methylene carbon linker highlighted by gray parenthesis. (b) EMSA of the competition with the MBNL1/$(CUG)_4$ complex. Each compound can compete with the MBNL1/$(CUG)_4$ complex. (c) IC_{50} curves for each linker analog. Left graph shows propamidine, butamidine, and pentamidine on a scale from 0–450 μM, and the right graph shows pentamidine, hexamidine, heptamidine, octamidine, and nonamidine on a scale from 0–50 μM since they are more effective. Error bars are standard deviation for each point.

Table 5. Structures and IC_{50} of each linker analog.

Small molecule	Name	Linker length	IC_{50}
	propamidine	3	250 ± 10 μM
	butamidine	4	85 ± 17 μM
	pentamidine	5	39 ± 2 μM
	hexamidine	6	26 ± 2 μM
	heptamidine	7	20 ± 2 μM
	octamidine	8	20 ± 1 μM
	nonamidine	9	14 ± 1 μM

DISCUSSION

From our lead compound, pentamidine, we wanted to synthesize analogs that would have increased efficacy at disrupting the MBNL1/$(CUG)_4$ complex, lower toxicity, and increased ability to be absorbed by cells (further discussed in Chapter III). To identify these compounds, we conducted a SAR study varying the length of the methylene linker between the phenyl amidines from 3 to 9 carbons. We suspected that by varying the chain length, we could fine-tune the distance between the positively charged amidine groups on either side of the molecule. We hypothesize that pentamidine binds the stretched U-U wobble pairs present in the CUG helix (*33*, *35*) because bulges are often sites of small molecule/RNA interactions. We suspected varying the number of methylenes would either allow the charged amidines to adopt a more favorable conformation for hydrogen bonding or strengthen hydrophobic interactions between the molecule and RNA (a key component of pentamidine's selective interaction with A/T rich DNA (*65*)). Finally, it seemed likely that increasing the hydrophobic character of the molecule would lead to more favorable pharmacodynamic properties by increasing membrane permeability.

In our competitive EMSA, we observed an increase in efficacy as linker length increased (Figure 22). However, since we did not identify an optimal linker length, we cannot be sure that the observed increase in efficacy was strictly due to an ideal distance between the amidine ends of the molecules. It is possible that the hydrophobic linker contributes significantly to the analog's activity.

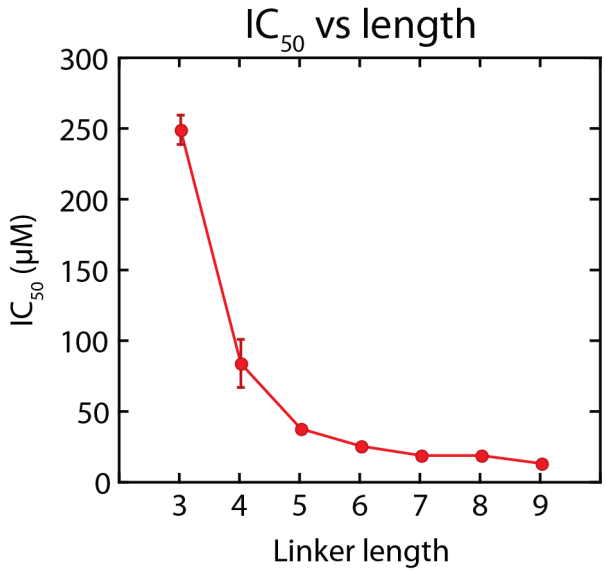


Figure 22. Plot of IC_{50} values versus methylene linker length. As length increases, IC_{50} decreases. Error bars are standard deviation.

METHODS

Protein purification, RNA labeling, analog synthesis, and competitive EMSA were conducted as described in Chapter III.

REFERENCES CITED

1. International Human Genome Sequencing Consortium. (2004) Finishing the euchromatic sequence of the human genome., *Nature 431*, 931–945.

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

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

4. 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.

5. 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.

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

7. Larkin, K., and Fardaei, M. (2001) Myotonic dystrophy--a multigene disorder., *Brain Res. Bull. 56*, 389–395.

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

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

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

11. Liquori, C. L., Ricker, K., Moseley, M. L., Jacobsen, J. F., Kress, W., Naylor, S. L., Day, J. W., and Ranum, L. P. (2001) Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9., *Science 293*, 864–867.

12. Kanadia, R. N. (2003) A Muscleblind Knockout Model for Myotonic Dystrophy, *Science 302*, 1978–1980.

13. 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.

14. 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.

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

16. 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.

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

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

19. 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., *Molecular Cell 10*, 35–44.

20. 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., *Proc. Natl. Acad. Sci. U.S.A. 106*, 18551–18556.

21. Arambula, J. F., Ramisetty, S. R., Baranger, A. M., and Zimmerman, S. C. (2009) A simple ligand that selectively targets CUG trinucleotide repeats and inhibits MBNL protein binding., *Proc. Natl. Acad. Sci. U.S.A. 106*, 16068–16073.

22. Wong, C.-H., Fu, Y., Ramisetty, S. R., Baranger, A. M., and Zimmerman, S. C. (2011) Selective inhibition of MBNL1–CCUG interaction by small molecules toward potential therapeutic agents for myotonic dystrophy type 2 (DM2), *Nucleic Acids Res 39*, 8881–8890.

23. Parkesh, R., Childs-Disney, J. L., Nakamori, M., Kumar, A., Wang, E., Wang, T., Hoskins, J., Tran, T., Housman, D., Thornton, C. A., and Disney, M. D. (2012) Design of a Bioactive Small Molecule That Targets the Myotonic Dystrophy Type 1 RNA via an RNA Motif–Ligand Database and Chemical Similarity Searching, *Journal of the American Chemical Society 134*, 4731–4742.

24. Lee, M. M., Pushechnikov, A., and Disney, M. D. (2009) Rational and Modular Design of Potent Ligands Targeting the RNA That Causes Myotonic Dystrophy 2, *ACS Chem. Biol. 4*, 345–355.

25. Pushechnikov, A., Lee, M. M., Childs-Disney, J. L., Sobczak, K., French, J. M., Thornton, C. A., and Disney, M. D. (2009) Rational design of ligands targeting triplet repeating transcripts that cause RNA dominant disease: application to myotonic muscular dystrophy type 1 and spinocerebellar ataxia type 3., *Journal of the American Chemical Society 131*, 9767–9779.

26. 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.*

27. Gareiss, P. C., Sobczak, K., McNaughton, B. R., Palde, P. B., Thornton, C. A., and Miller, B. L. (2008) Dynamic combinatorial selection of molecules capable of inhibiting the (CUG) repeat RNA-MBNL1 interaction in vitro: discovery of lead compounds targeting myotonic dystrophy (DM1)., *Journal of the American Chemical Society 130*, 16254–16261.

28. Garcia-Lopez, A., Llamusí, B., Orzáez, M., Pérez-Payá, E., and Artero, R. D. (2011) In vivo discovery of a peptide that prevents CUG-RNA hairpin formation and reverses RNA toxicity in myotonic dystrophy models., *Proc. Natl. Acad. Sci. U.S.A. 108*, 11866–11871.

29. Wheeler, T. M., Sobczak, K., Lueck, J. D., Osborne, R. J., Lin, X., Dirksen, R. T., and Thornton, C. A. (2009) Reversal of RNA dominance by displacement of protein sequestered on triplet repeat RNA., *Science 325*, 336–339.

30. Overington, J. P. J., Al-Lazikani, B. B., and Hopkins, A. L. A. (2006) How many drug targets are there?, *Nat Rev Drug Discov 5*, 993–996.

31. Guan, L., and Disney, M. D. (2012) Recent advances in developing small molecules targeting RNA., *ACS Chem. Biol. 7*, 73–86.

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

33. 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.

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. 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 Res 37*, 4149–4156.

36. Kumar, A., Park, H., Fang, P., Parkesh, R., Guo, M., Nettles, K. W., and Disney, M. D. (2011) Myotonic dystrophy type 1 RNA crystal structures reveal heterogeneous 1 × 1 nucleotide UU internal loop conformations., *Biochemistry 50*, 9928–9935.

37. Parkesh, R., Fountain, M., and Disney, M. D. (2011) NMR spectroscopy and molecular dynamics simulation of r(CCGCUGCGG)$_2$ reveal a dynamic UU internal loop found in myotonic dystrophy type 1., *Biochemistry 50*, 599–601.

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

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

40. Ranum, L. P., and Cooper, T. A. (2006) RNA-mediated neuromuscular disorders, *Annu Rev Neurosci 29*, 259–277.

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

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

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

44. de León, M. B., and Cisneros, B. (2008) Myotonic dystrophy 1 in the nervous system: from the clinic to molecular mechanisms., *J. Neurosci. Res. 86*, 18–26.

45. Wheeler, T. M. (2008) Myotonic dystrophy: therapeutic strategies for the future., *Neurotherapeutics 5*, 592–600.

46. Mulders, S. A. M., van Engelen, B. G. M., Wieringa, B., and Wansink, D. G. (2010) Molecular therapy in myotonic dystrophy: focus on RNA gain-of-function., *Hum. Mol. Genet. 19*, R90–97.

47. Foff, E. P., and Mahadevan, M. S. (2011) Therapeutics development in myotonic dystrophy type 1., *Muscle Nerve 44*, 160–169.

48. Hiebert, M., Gauldie, J., and Hillcoat, B. L. (1972) Multiple enzyme forms from protein-bromphenol blue interaction during gel electrophoresis., *Analytical Biochemistry 46*, 433–437.

49. Murakami, K. (1981) Kinetic studies of the interaction of bromophenol blue with bovine serum albumin by pressure-jump method, *Bull. Chem. Soc. Jpn. 54*, 862–868.

50. Subramanian, M., Sheshadri, B. S., and Venkatappa, M. P. (1984) Interaction of lysozyme with dyes. II. Binding of bromophenol blue., *J. Biochem. 96*, 245–252.

51. Kim, B. B., Abdul Kadir, H., and Tayyab, S. (2008) Bromophenol blue binding to mammalian albumins and displacement of albumin-bound bilirubin., *Pak. J. Biol. Sci. 11*, 2418–2422.

52. Phylactou, L. A., Darrah, C., and Wood, M. J. (1998) Ribozyme-mediated trans-splicing of a trinucleotide repeat., *Nat. Genet. 18*, 378–381.

53. Langlois, M.-A., Lee, N. S., Rossi, J. J., and Puymirat, J. (2003) Hammerhead ribozyme-mediated destruction of nuclear foci in myotonic dystrophy myoblasts., *Mol. Ther. 7*, 670–680.

54. Chen, H. Y., Kathirvel, P., Yee, W. C., and Lai, P. S. (2009) Correction of dystrophia myotonica type 1 pre-mRNA transcripts by artificial trans-splicing., *Gene Ther. 16*, 211–217.

55. Furling, D., Doucet, G., Langlois, M.-A., Timchenko, L., Belanger, E., Cossette, L., and Puymirat, J. (2003) Viral vector producing antisense RNA restores myotonic dystrophy myoblast functions., *Gene Ther. 10*, 795–802.

56. Langlois, M.-A., Boniface, C., Wang, G., Alluin, J., Salvaterra, P. M., Puymirat, J., Rossi, J. J., and Lee, N. S. (2005) Cytoplasmic and nuclear retained DMPK mRNAs are targets for RNA interference in myotonic dystrophy cells., *J Biol Chem 280*, 16949–16954.

57. Mulders, S. A. M., van den Broek, W. J. A. A., Wheeler, T. M., Croes, H. J. E., van Kuik-Romeijn, P., de Kimpe, S. J., Furling, D., Platenburg, G. J., Gourdon, G., Thornton, C. A., Wieringa, B., and Wansink, D. G. (2009) Triplet-repeat oligonucleotide-mediated reversal of RNA toxicity in myotonic dystrophy., *Proc. Natl. Acad. Sci. U.S.A. 106*, 13915–13920.

58. Mayburd, A. L., Tan, Y., and Kassner, R. J. (2000) Complex formation between Chromatium vinosum ferric cytochrome c' and bromophenol blue., *Arch. Biochem. Biophys. 378*, 40–44.

59. Cho, D. H., and Tapscott, S. J. (2007) Myotonic dystrophy: Emerging mechanisms for DM1 and DM2, *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease 1772*, 195–204.

60. Lee, J. E., Bennett, C. F., and Cooper, T. A. (2012) RNase H-mediated degradation of toxic RNA in myotonic dystrophy type 1., *Proc. Natl. Acad. Sci. U.S.A. 109*, 4221–4226.

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

62. Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings., *Adv. Drug Deliv. Rev. 46*, 3–26.

63. Mankodi, A., Logigian, E., Callahan, L., McClain, C., White, R., Henderson, D., Krym, M., and Thornton, C. A. (2000) Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat., *Science 289*, 1769–1773.

64. Kimura, T. (2005) Altered mRNA splicing of the skeletal muscle ryanodine receptor and sarcoplasmic/endoplasmic reticulum Ca2+-ATPase in myotonic dystrophy type 1, *Hum. Mol. Genet. 14*, 2189–2200.

65. Edwards, K. J., Jenkins, T. C., and Neidle, S. (1992) Crystal structure of a pentamidine-oligonucleotide complex: implications for DNA-binding properties., *Biochemistry 31*, 7104–7109.

66. Tidwell, R. R., Jones, S. K., Geratz, J. D., Ohemeng, K. A., Cory, M., and Hall, J. E. (1990) Analogues of 1,5-bis(4-amidinophenoxy)pentane (pentamidine) in the treatment of experimental Pneumocystis carinii pneumonia., *J. Med. Chem. 33*, 1252–1257.

67. Cooper, T. A., Wan, L., and Dreyfuss, G. (2009) RNA and Disease, *Cell 136*, 777–793.

68. Li, Y.-C., Korol, A. B., Fahima, T., Beiles, A., and Nevo, E. (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review., *Mol. Ecol. 11*, 2453–2465.

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

70. Murphy, F. L., and Cech, T. R. (1994) GAAA tetraloop and conserved bulge stabilize tertiary structure of a group I intron domain., *J. Mol. Biol. 236*, 49–63.

71. Costa, M., and Michel, F. (1995) Frequent use of the same tertiary motif by self-folding RNAs., *EMBO J. 14*, 1276–1285.

72. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R., and Doudna, J. A. (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing., *Science 273*, 1678–1685.

73. Ferré-D'Amaré, A. R., Zhou, K., and Doudna, J. A. (1998) A general module for RNA crystallization., *J. Mol. Biol. 279*, 621–631.

74. Reiter, N. J., Osterman, A., Torres-Larios, A., Swinger, K. K., Pan, T., and Mondragón, A. (2010) Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA., *Nature 468*, 784–789.

75. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, *Method Enzymol 276*, 307–326.

76. Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003) A graphical user interface to the CCP4 program suite., *Acta Crystallogr. D Biol. Crystallogr. 59*, 1131–1137.

77. 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 Crystallogr. D Biol. Crystallogr. 67*, 235–242.

78. Vagin, A., and Teplyakov, A. (1997) MOLREP: an Automated Program for Molecular Replacement, *J Appl Crystallogr 30*, 1022–1025.

79. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot., *Acta Crystallogr. D Biol. Crystallogr. 66*, 486–501.

80. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood Method, *Acta Crystallogr. D Biol. Crystallogr. 53*, 240–255.

81. Schrodinger, LLC. (2010, May) The PyMOL Molecular Graphics System, Version 1.5.0.1.

82. 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., *Nat Protoc 3*, 1213–1227.

83. Juneau, K., Podell, E., Harrington, D. J., and Cech, T. R. (2001) Structural basis of the enhanced stability of a mutant ribozyme domain and a detailed view of RNA--solvent interactions., *Structure 9*, 221–231.

84. Davis, J. H., Foster, T. R., Tonelli, M., and Butcher, S. E. (2006) Role of metal ions in the tetraloop-receptor complex as analyzed by NMR, *RNA 13*, 76–86.

85. Olson, W. K., Bansal, M., Burley, S. K., Dickerson, R. E., Gerstein, M., Harvey, S. C., Heinemann, U., Lu, X. J., Neidle, S., Shakked, Z., Sklenar, H., Suzuki, M., Tung, C. S., Westhof, E., Wolberger, C., and Berman, H. M. (2001) A standard reference frame for the description of nucleic acid base-pair geometry., *J. Mol. Biol. 313*, 229–237.

86. Qin, P. Z., Butcher, S. E., Feigon, J., and Hubbell, W. L. (2001) Quantitative analysis of the isolated GAAA tetraloop/receptor interaction in solution: a site-directed spin labeling study., *Biochemistry 40*, 6929–6936.

87. Downey, C. D., Fiore, J. L., Stoddard, C. D., Hodak, J. H., Nesbitt, D. J., and Pardi, A. (2006) Metal ion dependence, thermodynamics, and kinetics for intramolecular docking of a GAAA tetraloop and receptor connected by a flexible linker., *Biochemistry 45*, 3664–3673.

88. Wojciechowska, M., and Krzyzosiak, W. J. (2011) Cellular toxicity of expanded RNA repeats: focus on RNA foci., *Hum. Mol. Genet. 20*, 3811–3821.

89. Sheikh, M. I., and Gerdes, U. (1978) Interaction of phenolsulphonphthalein dyes with rabbit plasma and rabbit serum albumin., *Arch. Int. Physiol. Biochim. 86*, 613–623.