UNRAVELING THE INTERACTIONS BETWEEN MBNL1 PROTEIN
AND ITS RNA TARGETS

BY

YUAN FU

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Chemistry
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Doctoral Committee:

Professor Anne M. Baranger, Chair
Professor Steven C. Zimmerman
Professor Paul J. Hergenrother
Professor Susan A. Martinis
ABSTRACT

The muscleblind-like protein 1 (MBNL1) is an RNA binding protein with four conserved zinc fingers that plays an important role in the regulation of alternative splicing. It interacts with pre-mRNA and promotes the inclusion or exclusion of exons to form different mRNA transcripts. Besides its regulatory function in alternative splicing, MBNL1 has also been implicated to interact with CUG and CCUG repeat RNAs that are the causative agents of myotonic dystrophy. As a result of binding to the repeating RNA sequences, MBNL1 is abnormally translocated, which leads to many of the misregulated events in myotonic dystrophy.

MBNL1 can bind various RNA substrates, most of which share a stem-loop structure with two GC base pairs and pyrimidine-pyrimidine mismatches. However, MBNL1 has been shown to have a general binding preference for single-strand RNA containing 5’YGCY3’ motifs. In this work, the secondary structural preference of RNA targets for MBNL1 protein is investigated. With gel shift assays, steady-state fluorescence measurements along with circular dichroism measurements, I revealed that MBNL1 alters the structure of helical RNA targets upon binding, which may explain the selectivity of MBNL1 for less structured RNA sites. The stability of the stem-loop RNA secondary structure has been shown to be inversely correlated with the affinities of RNA substrates for MBNL1. Moreover, both the association and dissociation rates of MBNL1-RNA complexes are affected by stem-loop stability. I propose that the binding of MBNL1 to stem-loop RNA destabilize the RNA secondary structure and lead to partial melting of the stem region in the RNA substrates. Consistent with this hypothesis, RNA structure stabilizers were shown to destabilize the interaction between MBNL1 and CUG repeat RNA.
In addition to characterization of secondary structural preference of MBNL1, the kinetics of association of the MBNL1 – CUG repeat RNA complex has also been examined. Using stopped-flow fluorescence analysis and surface plasmon resonance, I have shown that multiple phases are present during the course of MBNL1 – CUG repeat RNA association. A multiple mode binding model with negative cooperativity has been proposed based on the kinetic data obtained.
ACKNOWLEDGEMENTS

I am very grateful to my research advisor, Professor Anne M. Baranger for her constant support and continuing thought-provoking guidance during the course of my graduate study. I would also like to thank my doctoral committee members Prof. Paul J. Hergenrother, Prof. Steven C. Zimmerman and Prof. Susan A. Martinis for offering insightful suggestions and guidance throughout my graduate career.

I would like to thank all the members of Baranger group for their support and encouragement. I want to express my special thanks to Dr. Sreenivasa Rao Ramisetty for introducing me the MBNL1 research project and for showing me about the gel shift experiments. Thanks to Dr. Divina Anunciado for helping me with circular dichroism experiments and protein expression. I would like to thank Nejmun Hussain, the undergraduate student who helped me with a lot of experiments. Also, I would like to thank Chun-Ho Wong and Amin Haghighat Jahromi in Zimmerman group for the helpful discussion and for their efforts in the success of MBNL1 project. My thanks also go to Professor Robert B Gennis and Dr. Myat Lin for helping me set up the stopped-flow experiments. Thanks to Liping Wang in Immunological Resource Center of University of Illinois for helping me set up the surface plasmon resonance experiments.

Finally, I am grateful to my family for their love and support during the entire course of my study.
# TABLE OF CONTENTS

## CHAPTER 1

**RNA Binding Proteins and their Significance in Myotonic Dystrophy**

1.1 Introduction ........................................................................................................... 1

1.2 Common RNA Binding Domains ........................................................................... 1

1.3 RNA Binding Specificity of RNA Binding Proteins .............................................. 6

1.4 Approaches for Investigation of Protein-RNA Interactions .................................. 9

1.5 RNA Dominant Diseases ...................................................................................... 13

1.6 MBNL1 Protein ...................................................................................................... 17

1.7 Inhibitors for MBNL1-RNA Interactions ................................................................. 20

1.8 Objectives and Significances of My Research ...................................................... 24

1.9 References ............................................................................................................ 25

## CHAPTER 2

**Contribution of MBNL1 Sequence and RNA Conformation to MBNL1-RNA Recognition** ................................................................. 39

2.1 Introduction .......................................................................................................... 39

2.2 Results ................................................................................................................ 40
2.2.1 Analysis of the Requirements of RNA Secondary Structure for MBNL1 Binding ................................................................. 40

2.2.2 Changes of RNA Conformation upon Binding MBNL1 .......................... 44

2.2.3 Evaluation of the RNA Binding Affinities of Truncated and Mutated MBNL1 Proteins ........................................................................................................... 48

2.3 Discussion and Conclusions .................................................................. 51

2.3.1 Truncation of MBNL1 Destabilizes Complexes with RNA ................... 51

2.3.2 MBNL1 Alters Structure of Helical RNA Targets ............................... 52

2.3.3 RNA Secondary Structure May Play a Role in Regulation of Alternative Splicing by MBNL1 .......................................................... 53

2.3.4 New perspectives for Drug Development for Myotonic Dystrophy ......... 55

2.4 Materials and Methods ......................................................................... 55

2.4.1 Plasmid ............................................................................................... 55

2.4.2 Cloning ............................................................................................... 55

2.4.3 Site Directed Mutagenesis ................................................................ 56

2.4.4 Recombinant Protein Preparation and Purification ......................... 58

2.4.5 Radiolabeling and Purification of RNA ............................................. 63

2.4.6 Electrophoresis Mobility Shift Assay ............................................... 63

2.4.7 Thermal Denaturation Study ............................................................ 67

2.4.8 Circular Dichroism Experiments ..................................................... 69

2.4.9 Steady-State Fluorescence Spectroscopy and Anisotropy .................. 69
CHAPTER 3

Correlation between the Binding of RNA to MBNL1 and RNA Secondary Structure Stability

3.1 Introduction

3.2 Results

3.2.1 GST-MBNL1 Induced FRET Changes in RNA Substrate

3.2.2 The Stability of RNA Secondary Structure is Correlated with the Binding Affinity of MBNL1

3.2.3 Both Association and Dissociation Rates are Correlated with Stem-loop Stability

3.2.4 RNA Secondary Structure Stabilizer as a Potential Inhibitor of MBNL1-RNA Interactions

3.2.5 Evaluation of the Binding Affinities and RNA Stabilization Effect of Small Molecule Ligands for CUG Repeat RNA

3.3 Discussion and Conclusions

3.3.1 Target RNA Provides a Flexible Conformation for MBNL1 Recognition

3.3.2 RNA is Partially Melted upon Binding

3.3.3 RNA Melts During the Course of MBNL1 Binding
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3.4 Conclusions</td>
<td>94</td>
</tr>
<tr>
<td>3.4 Materials and Methods</td>
<td>95</td>
</tr>
<tr>
<td>3.4.1 Protein and RNA</td>
<td>95</td>
</tr>
<tr>
<td>3.4.2 Thermal Denaturation Study</td>
<td>96</td>
</tr>
<tr>
<td>3.4.3 Steady State Fluorescence Measurements</td>
<td>97</td>
</tr>
<tr>
<td>3.4.4 Stopped-flow Fluorescence Measurements</td>
<td>99</td>
</tr>
<tr>
<td>3.4.5 Data Analysis</td>
<td>100</td>
</tr>
<tr>
<td>3.5 References</td>
<td>101</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>105</td>
</tr>
<tr>
<td>4.2 Results</td>
<td>105</td>
</tr>
<tr>
<td>4.2.1 Binding Occurs in Multiple Kinetic Phases</td>
<td>105</td>
</tr>
<tr>
<td>4.2.2 Analysis of the MBNL1-CUG repeat RNA Interactions by Surface Plasmon</td>
<td>108</td>
</tr>
<tr>
<td>4.3 Discussion and Conclusions</td>
<td>111</td>
</tr>
<tr>
<td>4.4 Materials and Methods</td>
<td>115</td>
</tr>
<tr>
<td>4.4.1 Plasmid</td>
<td>115</td>
</tr>
</tbody>
</table>
4.4.2 Protein Expression, Electrophoresis Mobility Shift Assay and Stopped-Flow Fluorescence Analysis

4.4.3 in vitro Transcription of CUG Repeat RNA

4.4.4 Biotinylation of RNA Transcript

4.4.5 Surface Plasmon Resonance

4.5 References
CHAPTER 1

RNA Binding Proteins and their Significance in Myotonic Dystrophy

1.1 Introduction

Proteins play an essential role in nearly all biological process. They carry out numerous functions such as catalyzing biological reactions, intercellular communication and small molecule transportation (1-2). One of the main characteristics that allow proteins to carry out numerous functions is their ability to bind and interact with other molecules including DNA, RNA, other proteins, and small molecule substrates. RNA binding proteins (RBPs) are proteins that bind to RNA. They can bind either to single stranded RNA or to double stranded RNA. They can recognize the primary sequence of the RNA or secondary and tertiary structure of the RNA molecule (3). RNA binding proteins play key roles in regulating posttranscriptional process (3-5). Aberrant expression or mis-localization of RNA binding proteins can therefore lead to adverse effects on many biological processes such as alternative splicing and protein translation (6-9). In particular, perturbation of RNA-protein interactions has been implicated to play a role in many human diseases, including myotonic dystrophy (DM) as one of the most important examples. The investigation of protein-RNA interactions in these human diseases will help us to better understand the mechanism of these diseases and to develop methods to target the protein-RNA interactions for therapeutic intervention.

1.2 Common RNA Binding Domains
RNA binding proteins associate with RNA to form ribonucleoprotein (RNP) complexes. Although they all interact with RNA, RNA binding proteins have different affinities and selectivity patterns towards their RNA targets. These differences are mediated by a number of RNA binding scaffolds which consist of a series of conserved modular structures – RNA-binding domains (10). Usually, one RNA binding protein contains more than one RNA-binding domains. The combination of multiple RNA binding domains confers a protein with the ability to interact with its RNA targets with stronger affinities and greater selectivities (10). There are numerous types of RNA binding domains. Some well studied RNA binding domains include the zinc finger domain, RNA recognition motif (RRM), K homology (KH) domain, cold shock domain, double-stranded RNA binding domains and arginine-rich domain (3,5,11-12). Several different RNA-binding domains will be described here, but more focus will be given to RNA-binding proteins with zinc finger domains.

The zinc finger domain is one of the shortest RNA-binding domains. It is stabilized by zinc metal ion, which coordinates with a combination of cysteines or histidine residues in proteins (3,13). The structure of the zinc finger domain is stable and does not undergo conformational changes upon interacting with oligonucleotides. Zinc finger proteins are best known for their ability to bind DNA. However, many zinc finger proteins have also been demonstrated to bind RNA. Zinc finger proteins are usually classified by the order of the zinc chelating residues. Examples of RNA-binding zinc finger proteins can be found in a wide range of organisms. These include the CCHC type HIV-1 nucleocapsid viral protein, the CCCH type
HUA1 nuclear protein, the CCCH type tcZFP1 protein from a parasite, and the C2H2 type mammalian zinc finger protein wig-1 (14).

One of the best-studied RNA binding zinc finger proteins is the transcription factor TFIIIA. It contains nine classical zinc finger domains and binds to the double stranded 5S ribosomal RNA. Zinc fingers 4 to 6 of the protein have been implicated to be important for the RNA binding activity of the protein (15-17). The crystal structure of the protein-RNA complex reveals that zinc fingers 4 and 6 make sequence specific base stacking contacts with nucleobases, while the basic residues in zinc finger 5 non-specifically interact with the phosphate backbones of the 5S ribosomal RNA molecule (17). It is interesting to note that TFIIIA can also interact with DNA using different zinc finger domains. While zinc finger 5 appears to work as a spacer in TFIIIA-RNA complex and makes non-specific contacts with RNA, it makes sequence specific contact with DNA. Zinc fingers 4 and 6 function as spacers in the TFIIIA-DNA complex (18).

The HIV-1 nucleocapsid (NCp7) protein is another example of a zinc finger RNA binding protein. It contains two copies of CCHC type zinc finger motifs with a consensus sequence of C-X$_2$-C-X$_4$-H-X$_4$-C. The two zinc finger domains are linked by a short basic connection sequence “RAPRKKG” (19). The nucleocapsid protein is involved in multiple functions throughout the lifecycle of the virus. It is known to be an RNA chaperone that remodels the RNA structure so that the most stable thermodynamic conformations are gained (19-20). Furthermore, it interacts with the stem loops from the $\psi$ packaging signal at 5’UTR of RNA transcript (19). Besides RNA targets, single stranded DNA has been also identified to be
target of nucleocapsid protein. The structure of nucleocapsid protein-DNA complex reveals that the protein binds to single stranded DNA similarly to how it binds to RNA (21).

**Figure 1.1.** Structures of classical RNA-binding zinc finger proteins. **A.** TFIIIA-RNA complex;(17) **B.** HIV-1 NCp7-RNA complex;(22) **C.** Wilm’s tumor 1 protein;(23) **D.** U11-48K protein;(24) **E.** dsRBP-ZFa protein.(25)

Apart from the above mentioned RNA-binding zinc finger protein, other well studied examples or RNA binding proteins include Wilm’s tumor 1 protein, U11-48K protein and dsRBP-ZFa protein. The structural diversity of the zinc finger proteins is shown in Figure 1.1, in
which the NMR or crystal structures of the proteins or the protein-RNA complexes are presented.

In addition to zinc finger proteins, RNA recognition motifs (RRM), K homology (KH) domain and double-stranded RNA binding domains are also well-studied. The RNA Recognition motif (RRM) is approximately 90 amino acids long with four antiparallel β-sheets and two α-helixes arranged in the order of βαβαβ. There are two high conserved motifs located in β-sheet 1 and β-sheet 3 of each RRM, called RNP1 and RNP2 (26). The amino acid residues in RNP1 and RNP2 are involved in RNA recognition through hydrogen bond interactions and electrostatic interactions (3,27).

K homology (KH) domain was first identified in the human heterogeneous nuclear ribonucleoprotein K (28). It consists of approximately 70 amino acids. There are two types of K homology domain, in which the order and the orientation of the secondary structural elements are different. In type 1, the β-sheets and α-helixes are arranged in the order of βααβα, while in type 2 the secondary structural elements are arranged in the order of αββααβ (29). Both type 1 and type 2 share a βααβ “KH minimal motif” in their linear sequence (29).

The double-stranded RNA binding domain is about 70 -90 amino acids long and folds into an αββα structure in which a three-stranded antiparallel β-sheet is supported by two α-helixes. What is special about the double-stranded RNA binding domain is that it makes non-specific contacts with RNA through the 2’-OH groups and phosphate backbone (30). Therefore, the binding specificity of RNA binding proteins with the double-stranded RNA binding domain is thought to arise from cooperation with additional domains in the proteins.
1.3 RNA Binding Specificity of RNA Binding Proteins

To understand the function of different RNA binding protein, one important question to answer is how RNA binding proteins distinguish their targets from non-targets. Different types of RNA-binding proteins exhibit different selectivity towards their RNA targets. The RNA targets of RNA binding proteins can be either double stranded or single stranded RNA. Table 1.1 lists the names of the identified RNA-binding domains mentioned above with the types or RNA targets they recognize.

<table>
<thead>
<tr>
<th>RNA binding protein</th>
<th>Types of RNA recognized</th>
<th>Examples of RNA binding proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>zinc finger</td>
<td>ssRNA, dsRNA</td>
<td>TFIIIA, HIV-1 NC, MBNL1</td>
</tr>
<tr>
<td>RNA recognition motif</td>
<td>ssRNA</td>
<td>U1A, Sxl, HuD</td>
</tr>
<tr>
<td>K homology</td>
<td>ssRNA</td>
<td>hnRNP K</td>
</tr>
<tr>
<td>double stranded RNA binding domain</td>
<td>dsRNA</td>
<td>PKR, dicer, RNase III</td>
</tr>
<tr>
<td>cold shock</td>
<td>ssRNA</td>
<td>CspB</td>
</tr>
<tr>
<td>Arginine-rich</td>
<td>dsRNA, ssRNA</td>
<td>Tat</td>
</tr>
</tbody>
</table>

There are several important features that affect the RNA-protein recognition. First, RNA helices have an A-form structure. The major grooves in A-form RNA are deep and narrow and sterically restrict protein recognition. Therefore, RNA-binding proteins often recognize dsRNA through minor groove contacts (31). In some cases, the recognition of the helices is solely achieved through contacts with the phosphate-sugar backbone. Second, the dsRNA targets are not always genuinely fully complementary duplex RNA. The RNA targets may include other non-Waston-Crick regions such as loops, bulges, junctions or dangling ends. An RNA-binding protein that interacts with the helices may also recognize these non-Waston-Crick secondary structural elements. For example, the HIV-1 nucleocapsid protein interacts with helices as well.
as the tetraloop structure in stem-loop SL2 of the ψ packaging signal (32). Third, RNA duplexes with mismatches and loops have less favorable free energies of folding. The fact that the RNA helices could be disrupted in protein-RNA complexes should be taken into consideration especially for those targets containing short helices with only two or three base pairs. This aspect will be emphasized in Chapter 2 and Chapter 3 of this thesis.

Apart from the secondary structural elements, the primary sequence of the RNA is another feature that is recognized selectively by RNA-binding proteins. RNA binding proteins achieve sequence specificity through a combination of π-stacking and hydrogen bond interactions with nucleotides.

At the protein-RNA interface, π-stacking interactions can be found between the aromatic rings of nucleobases and the aromatic side-chains of phenylalanine, tyrosine, tryptophan and histidine. They can also be found between positively charged guanidine group of arginine residues and RNA bases. Some π-stacking interactions appear to contribute substantial free energies of binding to the protein-RNA complex. For example, mutation of one phenylalanine residue to alanine in the β-sheet of U1A protein destabilizes the U1A-RNA complex by 10,000 fold (33).

Besides π-stacking interactions, hydrogen bonding is another factor that contributes to the sequence specific recognition. A hydrogen bond is formed between a hydrogen bond donor and a hydrogen bond acceptor. Many amino acid side chains in proteins can act as donors, such as cysteine, tyrosine and asparagine or acceptors such as aspartate, glutamine and serine. A nucleobase can act as either donor or acceptor depending on the sites of interaction. For example,
uracil can act as acceptor, donor, acceptor at position 2, 3, 4 of the ring, respectively. Owing to the requirement of both a donor and an acceptor, the hydrogen bonding interactions are generally more important than π-stackings interaction in establishing sequence specific recognition (34).

For the U1A protein, removing one hydrogen-bond interaction or altering the hydrogen bond pattern by replacing the nucleotide with its analogue can result in approximately 0.9-3.0 kcal/mol change in free energies of protein-RNA interactions. (33,35)

<table>
<thead>
<tr>
<th>RNA-binding protein</th>
<th>RRM domain</th>
<th>target sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sxl</td>
<td>RRM1</td>
<td>UUUUUUUU</td>
</tr>
<tr>
<td>Sxl</td>
<td>RRM2</td>
<td>UGU</td>
</tr>
<tr>
<td>hnRNP A1</td>
<td>RRM1</td>
<td>UAGG</td>
</tr>
<tr>
<td>hnRNP A1</td>
<td>RRM2</td>
<td>UUAGG</td>
</tr>
<tr>
<td>RABP</td>
<td>RRM1</td>
<td>AAAA</td>
</tr>
<tr>
<td>RABP</td>
<td>RRM2</td>
<td>AAAA</td>
</tr>
<tr>
<td>HuD</td>
<td>RRM1</td>
<td>UUAUUU</td>
</tr>
<tr>
<td>HuD</td>
<td>RRM2</td>
<td>UU</td>
</tr>
<tr>
<td>HuD</td>
<td>RRM3</td>
<td>UAU</td>
</tr>
<tr>
<td>PTB</td>
<td>RRM1</td>
<td>UCNU</td>
</tr>
<tr>
<td>PTB</td>
<td>RRM2</td>
<td>CNUNN</td>
</tr>
<tr>
<td>PTB</td>
<td>RRM3</td>
<td>UCNU</td>
</tr>
<tr>
<td>PTB</td>
<td>RRM4</td>
<td>UCNN</td>
</tr>
</tbody>
</table>

By forming networks of π-stacking and hydrogen bonding interactions, RNA-binding proteins recognize many different target RNA sequences. The target RNA sequences vary in sequence and structure even for RNA-binding protein of the same type. Some of the RNA-binding domains are more specific than others and some of the RNA-binding domains bind stronger to their targets than others. Table 1.2 shows examples of different RNA sequences recognized by RRM domains. In an extreme case, the same protein could exhibit sequence-specific binding properties to a number of different RNA sequences. Table 1.3 shows
an example of the many different RNA sequences recognized by one particular protein, HIV-1 NCp7, which is a zinc finger RNA binding protein. In fact, HIV-1 NCp7 is able to bind to nearly all the RNA sequences that are 5-8 nt in length through weak nonspecific interaction with micromolar affinities (36-37), while it binds to specific RNA targets with nanomolar affinities.(19)

Table 1.3 Examples of RNA sequences recognized by HIV-1 NCp7 (19)

<table>
<thead>
<tr>
<th>Structural context</th>
<th>RNA sequencea</th>
<th>Kobs (nM)</th>
<th>Methodb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stem-loop(SL3)</td>
<td>GGACUAGCGAGAGCUAGUCC</td>
<td>21-28</td>
<td>TFQ</td>
</tr>
<tr>
<td>Stem-loop(SL3)</td>
<td>GGACUAGCGAGAGCUAGUCC</td>
<td>100</td>
<td>EMSA</td>
</tr>
<tr>
<td>Stem-loop(SL3)</td>
<td>GGACUAGCGAGAGCUAGUCC</td>
<td>170</td>
<td>ITC</td>
</tr>
<tr>
<td>Stem-loop(SL3)</td>
<td>GGACUAGCGAGGCUAGUCC</td>
<td>551</td>
<td>FA</td>
</tr>
<tr>
<td>Stem-loop(SL2)</td>
<td>GGCGACUGGUGAGUACGCC</td>
<td>110</td>
<td>ITC</td>
</tr>
<tr>
<td>Stem-loop(SL2)</td>
<td>GGCGACUGGUGAGUACGCC</td>
<td>23</td>
<td>TFQ</td>
</tr>
<tr>
<td>Stem-loop(SL2)</td>
<td>GGCGACUGGUGAGUACGCC</td>
<td>400</td>
<td>EMSA</td>
</tr>
<tr>
<td>ssRNA</td>
<td>AAUGCC</td>
<td>285</td>
<td>TFQ</td>
</tr>
<tr>
<td>ssRNA</td>
<td>UGUGCGCGUCU</td>
<td>120</td>
<td>ITC</td>
</tr>
</tbody>
</table>

a, sequence highlighted in bold indicates identified binding site; b, TFQ=tryptophan fluorescence quenching assays, EMSA=gel electrophoresis shift assay, ITC=isothermal titration calorimetry, FA=fluorescence anisotropy.

In summary, RNA-binding proteins recognize both the secondary structural elements and primary RNA sequence to achieve target specificity. Hydrogen bonding interactions and π-stacking interactions contribute to both binding specificity and affinity. In some cases, a generalization of the recognition pattern has become difficult, because the RNA binding proteins recognize multiple sequences.

1.4 Approaches for Investigation of Protein-RNA Interactions

The investigation of protein-RNA interactions is important for understanding many RNA related cellular processes. When quantitatively analyzing protein-RNA interactions, we would
usually like to know the thermodynamic and kinetic properties of the interactions. Furthermore, we would like to obtain structural information on how protein interacts with RNA. We want to know which amino acids residues or nucleotides are involved in specific protein-RNA interactions. Often, it would be useful to first consider traditional approaches before moving onto new approaches. A brief overview of some traditional approaches for characterization of protein-RNA interactions are given in this section.

Electrophoresis mobility shift assay (EMSA) is one of the most widely used approaches for characterizing the binding affinities of a protein to its RNA targets. Free RNA runs further on agarose or acrylamide gel than bound RNA. Therefore, the fraction of RNA bound to protein can be determined by taking an autoradiograph of dried gel. In electrophoresis mobility shift assay, the concentration of RNA is very low and protein is generally used in large excess, so the concentration of free protein will be close to the total protein concentration (38). In general, what EMSA tells us is the binding affinity of the RNA to protein. One drawback of EMSA is that it does not provide us with information about stoichiometry and it assumes that the protein is fully active. Although multiple bands can sometimes be observed when more than one protein molecule bind to RNA, lack of such bands does not necessarily mean 1:1 stoichiometry (39). Another drawback of EMSA is that it is a non-equilibrium technique and the measured dissociation constant does not always agree with dissociation constant determined by other equilibrium methods. Additionally, it is possible to carry out kinetic analysis and competition assay using EMSA. The fraction of bound RNA at varying time can be determined by EMSA, as long as the half-life time is significantly longer than the dead time of the experiment. The
competition assay can be carried out by incubating another oligonucleotide with protein-RNA complex before loading onto gels.

Fluorescence spectroscopy and fluorescence anisotropy are alternative methods for studying the binding affinities of protein with RNA. In both methods, the component with fluorescence can be either the protein or RNA. RNA that has been modified with a fluorophore or protein containing a tryptophan are commonly used in these assays. The measurement of the fluorescence intensity or anisotropy can be carried out either in steady state to obtain the equilibrium values or in real-time to analyze the kinetic process using a stopped flow instrument coupled to a fluorescence spectrometer.

In fluorescence spectroscopy measurements, fluorescence quenching or a shift in the wavelength maximum is monitored when one component is titrated into another. The extent of fluorescence quenching or wavelength shift is proportional to the concentration of protein-RNA complexes (40). The fraction of protein-RNA complex can therefore be determined based on the extent of fluorescence change. In addition, fluorescence resonance energy transfer (FRET) analysis can be used to give quantitative information on distance between sites of interaction.

Fluorescence anisotropy provides another approach to measure the dissociation constant of a protein-RNA complex, if the molecular size of the complex is sufficiently different from that of the free fluorescent component (40). Fluorescence anisotropy describes the extent of rotational diffusion of the emission light, which reflects the size of the complex with fluorescence. In fluorescent anisotropy assays, the anisotropy of fluorescently-labeled RNA is usually monitored
as a function of protein concentration. The increase in molecular weight of the complex will result in a decrease in the extent of depolarization of the emission light.

Surface plasmon resonance can be used to study the real-time kinetics of protein-RNA interactions. The detection principle of SPR relies on the optical properties change of the chemistry modified metal surface of a sensor chip (41). In surface plasmon resonance, one component, either protein or RNA, is immobilized on the metal surface and binding of the other component to the one immobilized on the surface will lead to a change in reflective index. The change will be recorded as resonance unit (RU) by the SPR instrument. Usually, the RNA is immobilized and the protein is allowed to flow over the surface of chip. The increase and decrease in resonance unit as a function of time reflect the association and dissociation process of the protein-RNA interaction.

To obtain structural information about protein-RNA interactions, X-ray crystallography and solution phase NMR can be employed. These two techniques provide precise three dimensional details for specific protein-RNA interactions. Despite the power of the techniques, they are also generally considered to be difficult methods requiring special knowledge in material handling and data interpretation. Although less powerful than crystallography and NMR, circular dichroism spectroscopy and RNA foot printing are commonly used alternative methods for obtaining structural information. Circular dichroism spectroscopy gives information about the secondary structures of protein and RNA, while RNA foot printing shows specific nucleobases that are involved in protein-RNA interactions. For circular dichroism measurements, the far UV spectrum 170-250 nm reflects the change in secondary structure of protein, while the near UV
spectrum 250-300 nm reflects the change in secondary structure of RNA (42). For RNA footprinting, there are two complementary approaches. In one approach, RNA is protected from cleavage reagents when protein binds; in another approach, the bases of RNA are modified and the modifications that prevent binding will result in absent of bands on denaturing gel.

Additional approaches for investigation of protein-RNA interactions include isothermal titration calorimetry (ITC), filter binding assays, RNA pull down assays, yeast three hybrid assay and fluorescent in situ hybridization co-localization (FISH). Overviews of these assays will not be given here due to the focus of this thesis.

1.5 RNA Dominant Diseases

A significant proportion of eukaryotic transcripts are considered to be non-coding RNAs (43). The RNA repeat expansions in non-coding region of the transcriptome are closely related with a few dominantly inherited disorders. At present, myotonic dystrophy (DM) and fragile X tremor ataxia syndrome (FXTAS) are the two clearest examples of RNA dominant diseases. A discussion about the RNA-mediated pathogenesis of DM and FXTAS will be given below respectively.

DM is currently the best-characterized example of an RNA dominant disease. The disease symptoms include cardiac defects, muscular weakness, insulin resistance and other neuromuscular problems. At present, there are no effective treatments for myotonic dystrophy. There are two types of DM, each associated with a repeated sequence of RNA. Patients with myotonic dystrophy type 1 (DM1) have a poly(CTG) repeat, which can be transcribed into poly(CUG) RNA, in the 3′-untranslated region of the dystrophia myotonic protein kinase gene.
Patients with myotonic dystrophy type 2 (DM2) bear a poly(CCTG) repeat in the CCHC-type zinc finger 9 gene (ZNF9) and the poly(CCTG) repeat is transcribed into poly(CCUG) RNA (44-48). The repeat number of CUG repeat expansion or CCUG repeat expansion in unaffected individuals is much shorter than that in patients. The disease severity and age of onset of myotonic dystrophy are correlated with the repeat number. There is considerable support for a toxic RNA gain-of-function of the poly(CUG) or poly(CCUG) transcripts, which tend to accumulate in nuclear foci, as the molecular basis of DM (49-55). One of the most important pieces of evidence for the RNA-gain-of-function model is that the disease symptoms can be reproduced by expression of CUG or CCUG repeat expansions in transgenic mice and flies (53-55).

A main molecular feature of myotonic dystrophy is the mis-regulation of alternative splicing. To explain the molecular basis of mis-regulation of alternative splicing, the RNA gain-of-function model suggests that the CUG or CCUG repeat expansion can exert effects on proteins that regulate alternative splicing. There are a number of pieces of evidence supporting this hypothesis and at least two proteins are known to be involved in this process, MBNL1 and CUG-BP1.

MBNL1 binds to CUG or CCUG repeat RNA. The RNA binding affinities of MBNL1 correlate with the repeat number.(49,56-58) Therefore, it was proposed that MBNL1 was sequestered from its normal functions upon interacting with CUG or CCUG repeat RNA. MBNL1 is an alternative splicing regulator and therefore, loss of MBNL1 can lead to many mis-splicing events. It has been suggested that more than 80% of the mis-splicing pathology
found in a DM mouse model can be explained by sequestration of MBNL1 (59). At least six mis-spliced pre-mRNA have been identified to be RNA targets of MBNL1 protein (57,60-63). Furthermore, studies have shown that an MBNL1 knock-out mouse exhibits similar symptoms in eye, muscle and RNA splicing abnormalities as can be observed in myotonic dystrophy (64).

CUG-BP1 also plays a role in mis-regulation of alternative splicing in myotonic dystrophy. Three mis-regulation events in myotonic dystrophy, cTNT, IR and ClCn1, have been shown to be regulated by CUG-BP1 and the splicing patterns of these pre-mRNAs are consistent with an increase in CUG-BP1 activity (65-67). The expression level, cellular distribution and phosphorylation status of CUG-BP1 are altered in cells from patients with DM and these changes are consistent with an increase in CUG-BP1 activities (67-69). Moreover, transgenic mice with over-expressed CUG-BP1 show a similar splicing pattern as is observed in myotonic dystrophy (70).
Figure 1.2. Schematic illustration of the molecular mechanism of myotonic dystrophy.

Although both MBNL1 and CUG-BP1 contribute to mis-splicing events in myotonic dystrophy, the mechanism by which CUG repeat RNA induces an increase in CUG-BP1 activities is not clear. Experiments have suggested that the loss of MBNL1 appears to play a more important role than the gain of CUG-BP1 activity. However, it is worth mentioning that loss of MBNL1 activities cannot explain all the mis-splicing patterns observed in myotonic dystrophy.
Another example of RNA dominant disease is FXTAS, which is associated with a premature poly(CGG) repeat expansion at the 5’UTR of the FMR gene on the X chromosome. It is interesting to note that the repeat number of poly(CGG) in FXTAS should lie between 70 and 120, because a greater number of repeats will give rise to fragile X syndrome (FXS), another genetic disease, through a loss-of-function mechanism (71). A wealth of evidence supports the RNA gain-of-function model in FXTAS. First, like myotonic dystrophy, nuclear inclusions with CGG repeat RNA are found in nucleus of cells from patients (72). Second, loss of FMR protein does not lead to symptoms in FXTAS, including tremor and ataxia (71). Third, transgenic poly(CGG) knocked-in mice with normal level of FMR protein reproduce symptoms in FXTAS (73). Although FXTAS appears to be caused by RNA gain-of-function mechanism, it is not clear how CGG repeat RNA could lead to the disease symptoms. Further research needs to be carried out to identify the direct targets of CGG repeat RNA.

1.6 MBNL1 Protein

As discussed above, MBNL1 protein plays an important role in myotonic dystrophy, an RNA dominant disease. The focus of my research is MBNL1-RNA interactions. Therefore, an introduction to the functions and RNA binding properties of MBNL1 will be given here.
MBNL1 is a zinc finger RNA binding protein. It contains four CCCH zinc finger domains, positioned as tandem pairs at the N-terminus (ZnF1 and ZnF2) and the middle (ZnF3 and ZnF4) of the protein. ZnF1 and ZnF3 share a similar amino acid sequence, while ZnF2 and ZnF4 share another similar sequence. In ZnF1 and ZnF3 domains, the Cys are found in CX7CX6CX3H sequences, while in ZnF2 and ZnF4 domains, the Cys are found in CX7CX4CX3H sequences. The RNA binding activity of MBNL1 fully resides in the two tandem
zinc finger domains (57,60). The sequence and the crystal structure of the zinc finger domains of MBNL1 are shown in Figure 1.3.

As mentioned in a previous section, MBNL1 regulates alternative splicing of pre-mRNA. It promotes exon exclusion in cTNT pre-mRNA and exon inclusion in IR pre-mRNA (60). It has been shown that MBNL1 regulates cTNT exon exclusion through competition with the splicing factor U2AF2 (75). It binds to the polypyrimidine tract upstream to the exon to be regulated in cTNT gene and blocks spliceosome complex formation (75). It also interacts with IR pre-mRNA via binding to the intronic enhancer downstream to the exon to be regulated (76). It is believed that MBNL1 antagonizes the splicing pattern mediated by CUG-BP1. In an example of alternative splicing regulation of CICN1, MBNL1 inhibits specific exon inclusion, while CUG-BP1 promotes exon inclusion (61). In addition to its splicing regulatory function, MBNL1 has been recently suggested to play a role in mRNA decay through binding to the 3’UTR of mRNA (77).

The RNA binding specificity of MBNL1 has been well characterized. Ishiura and coworkers found that MBNL1 specifically interacts with repeating CHG and CHHG (H=A, U or C) sequences (58). By comparing different binding sites, Cooper and coworkers concluded that MBNL1 recognizes 5’YGCU(U/G)Y3’ (Y=U or C) motifs (62). In another study, SELEX was performed against MBNL1 and revealed that MBNL1 preferably binds to pyrimidine-rich RNAs containing 5’YGCY3’ (Y=U or C) motifs (78). The identification of 5’YGCY3’ (Y=U or C) motifs in pre-mRNA transcripts that are mis-spliced in myotonic dystrophy showed that the
occurrence of 5’YGCY3’ motifs downstream and upstream of the mis-spliced exon is correlated with the missplicing pattern of myotonic dystrophy (78).

Although the binding specificity of MBNL1 has been well-studied, the secondary structural preference of MBNL1 has been a controversial topic. MBNL1 binds to CUG repeats, which share an antiparallel double-stranded helical structure (79-80), but it does not bind to a fully complementary double-stranded RNA comprised of CAG/CUG repeat sequences (58). Berglund and coworkers conclude that MBNL1 recognizes a stem loop structure with mismatches (60,75). However, there was no correlation between MBNL1’s binding affinities and the stem-loop stabilities of the RNA constructs obtained by SELEX (78). The MBNL1 binding sites on IR pre-mRNA transcript identified do not appear to have significant secondary structure (76). Moreover, a crystal structure of zinc fingers of MBNL1 shows that the zinc finger 3 and zinc finger 4 domains of MBNL1 can bind short single stranded RNA (74).

1.7 Inhibitors for MBNL1-RNA Interactions

Strategies for treatment of myotonic dystrophy can be directed at the level of either DNA, RNA or protein. Currently, it is difficult to develop therapeutic strategies aimed at eliminating the DNA repeat expansion and directly correcting the aberrant splicing. By comparison, reduction of RNA-mediated toxicity at the RNA level by selective targeting of the CUG repeat RNA appears to be a promising strategy. A series of recent papers provide proof-of-concept for using small molecule or antisense oligonucleotides to target CUG repeat RNA and prevent MBNL1 from binding on it. An introduction of the inhibitors will be given here with focus on
small molecule inhibitors. Studies of using antisense oligonucleotide for treatment of myotonic dystrophy will not be discussed here due to the focus of my research (81-83).

In identifying small molecule inhibitors, Miller and coworkers reported the first example of using a small organic compound to liberate MBNL1 protein by binding to CUG repeat RNA (84). In their work, a resin-bound dynamic combinatorial library (RBDCL) with a theoretical size of 11,325 members was created for identification of the ligands. Library screening led to identification of four monomers (1-4 in Figure 1.4) allowing for ten possible combinations of homo- or hetero- dimmers for inhibition of MBNL1-CUG repeat RNA interaction. They further evaluated the identified small molecules for RNA binding activities and inhibition activities, as well as their abilities to induce aggregation.

Shortly after this work has been published, Zimmeman and coworkers reported the rational design of a simple ligand that recognizes structured CUG repeat RNA (85). The compound they reported (5 in Figure 1.4) is selective for T-T and U-U mismatches over other mismatches or a fully complementary duplex. In binding experiments, low micromolar concentrations of the ligand destabilizes complexes formed between CUG repeat RNA and MBNL1 even in presence of competitor tRNA. Moreover, two derivatives of the ligand were developed later to selectively target CCUG repeat RNA, which plays a role in myotonic dystrophy type 2 (6 and 7 in Figure 1.4) (86).

Disney and coworkers also used a rational design route to identify inhibitors for MBNL1-CUG repeat RNA interactions (87-91). In their research, kanamycin A or Hoechst 33258 modules were attached to a peptoid backbone to construct multivalent and modularly
assembled ligands for bindings to CUG repeat RNA. (8 in Figure 1.4) A series of compounds were synthesized by changing the number of modules and flexibility of the scaffold and their binding affinities for various RNA targets were determined. An advantage of the modularly assembled compounds they constructed is that these compounds bind to CUG repeat RNA in the low nanomolar range and are generally cell permeable. In addition, these multivalent compounds are promising for treatment of myotonic dystrophy because they can reverse the splicing defects and disrupt nuclei foci formed by CUG repeat RNA in cellular assays (92).

In another report, Berglund and coworkers determined that pentamidine (9 in Figure 1.4), a drug for treatment of Pneumocystis carinii infections in AIDS patients, can disrupt the interaction between MBNL1 and CUG repeat RNA (93). They showed that pentamidine can not only destabilize the complex, but also reverse the splicing defects associated with myotonic dystrophy. They speculated that pentamidine binds to the minor groove of A-form CUG repeat RNA and does not disrupt the stem-loop structure of the RNA.

A summary of the above mentioned inhibitors is given in Table 1.4 and Figure 1.4. Altogether, although much progress has been made, additional basic studies are still required to better understand the molecular basis of the inhibitors in destabilizing the protein-RNA complex and to improve selectivity and lower the toxicity of the inhibitors.
Figure 1.4 Organic compounds developed for inhibitors of MBNL1-CUG repeat RNA interactions
Table 1.4 Binding affinities of the small molecule inhibitors to CUG repeat RNA in literatures

<table>
<thead>
<tr>
<th>Inhibitors$^a$</th>
<th>Target RNA</th>
<th>$K_i$</th>
<th>Experimental Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2,3,4</td>
<td>(CUG)$_6$, (CUG)$<em>8$, (CUG)$</em>{106}$</td>
<td>$1.4-2.6 \mu M$</td>
<td>Fluorescence Titration</td>
</tr>
<tr>
<td>1,2,3,4</td>
<td>(CUG)$_6$, (CUG)$<em>8$, (CUG)$</em>{106}$</td>
<td>$1.9-9.8 \mu M$</td>
<td>Filter Binding Assay</td>
</tr>
<tr>
<td>5</td>
<td>(CUG)$_2$, (CUG)$_4$</td>
<td>$0.43-2.1 \mu M$</td>
<td>ITC$^b$</td>
</tr>
<tr>
<td>5</td>
<td>(CUG)$<em>4$, (CUG)$</em>{12}$</td>
<td>$6-7 \mu M$</td>
<td>Gel Shift Assay</td>
</tr>
<tr>
<td>7</td>
<td>(CCUG)$_6$</td>
<td>$5.2 \mu M$</td>
<td>ITC$^b$</td>
</tr>
<tr>
<td>6,7</td>
<td>(CCUG)$_6$</td>
<td>$2.2 \mu M, 2.5 \mu M$</td>
<td>Gel Shift Assay</td>
</tr>
<tr>
<td>6,7</td>
<td>(CUG)$_{12}$</td>
<td>$6.0 \mu M, 7.0 \mu M$</td>
<td>Gel Shift Assay</td>
</tr>
<tr>
<td>8</td>
<td>(CUG)$_{109}$</td>
<td>$13 \text{nM}^c$</td>
<td>Fluorescence Binding Assay$^d$</td>
</tr>
<tr>
<td>9</td>
<td>(CUG)$_4$</td>
<td>IC$50 = 58 \mu M^e$</td>
<td>Gel Shift Assay</td>
</tr>
</tbody>
</table>

$a$, the numbering of inhibitors is based on the labels under structures in Figure 1.4; $b$, ITC stands for Isothermal Titration Calorimetry; $c$, $K_i$ listed is that of the best ligand. The assembled module of the ligand is Hoechst 33258; $d$, The fluorescence arises from the inhibitor; $e$, The dissociation constant of the small molecule with CUG repeat RNA is not reported.

1.8 Objectives and Significances of My Research

To better understand the molecular basis of myotonic dystrophy, it is important to gain information about MBNL1-RNA recognition. This knowledge can help us predict new targets of MBNL1 as well as develop treatments for myotonic dystrophy. Although the sequence specificity of MBNL1 has been characterized, much remains unknown about the effect of secondary structures of RNA targets on MBNL1-RNA recognition and the kinetic association of MBNL1-RNA complexes. In addition, it is important to investigate the contribution of specific amino acid residues and domains to MBNL1-RNA recognition. Therefore, my research goal is to characterize the MBNL1-RNA interactions with an emphasis on RNA secondary structures, protein sequence and kinetic properties.

In Chapter 2, the interactions between MBNL1 and RNA targets in different secondary structural context have been investigated. Conformational analysis has been carried out with
FRET and circular dichroism measurements. Moreover, I have examined the RNA binding properties of two classes of mutant MBNL1 proteins: MBNL1 with point mutations at conserved amino acid residues and truncated MBNL1 lacking one tandem pair of zinc finger domains. These studies help us understand the contributions of protein sequence and RNA conformation to MBNL1-RNA interactions. In Chapter 3, the correlation between RNA folding energies and MBNL1 binding properties have been investigated from both thermodynamic and kinetic aspects. The study of the correlation leads to a new perspective in using a small molecule RNA stabilizer as a potential drug for treatment of myotonic dystrophy. In addition, melting temperature and binding affinities analysis on a series of rationally designed small molecules inhibitors has been performed to investigate the effect of RNA structure stabilization on disruption of MBNL1-RNA complex. In Chapter 4, kinetic analysis of the interactions between MBNL1 and CUG repeat RNA have been carried out by both SPR and fluorescence stopped-flow measurements to shed insight on the kinetic process of the interaction between MBNL1 and disease-causing RNA interactions.

Taken altogether, these studies add to our knowledge in understanding MBNL1-RNA recognition and are expected to provide insight into predicting new RNA targets of MBNL1 and developing new inhibitors for MBNL1-RNA interactions.

1.9 References


CHAPTER 2

Contribution of MBNL1 Sequence and RNA Conformation to MBNL1-RNA Recognition*

2.1 Introduction

The MBNL1 protein binds to RNA targets, including both helical and single-stranded sequences, containing 5’-YCGY-3’ sequences (1-2). Structured helical targets of MBNL1 include the poly(CUG) repeat sequences of DM1, which form structures that resemble A-form RNA (3-4). Sequences with less stable secondary structures include cardiac troponin T (cTNT), which is targeted by MBNL1 to regulate alternative splicing. MBNL1 can bind to a 32mer segment of the cTNT pre-mRNA and promote the inclusion of exon 5 in cTNT (5-8). The diversity of the structures among MBNL1 targets promoted me to investigate the secondary structural preferences of MBNL1.

MBNL1 contains two tandem pairs of zinc finger domains with conserved sequences (9). An additional question I sought to answer is whether both or only one of the tandem zinc finger domains contribute to the MBNL1-RNA recognition. Furthermore, it would be important to characterize the contributions of individual amino acid residues in zinc finger domains to the RNA binding properties of MBNL1.

Here, I am presenting my efforts in investigating the contributions of MBNL1 sequence and RNA conformation to MBNL1-RNA recognition. I have used in vitro binding assays and steady-state fluorescence measurements with purified recombinant MBNL1 protein to probe the

*Text taken from ChemBioChem Volume 13, Issue 1, pages 112–119
conformational change of RNA targets was investigated by circular dichroim spectroscopy. The data presented here suggest that MBNL1 disrupts helical RNA structure upon binding and that the binding energy is distributed throughout the two pairs of the zinc fingers of MBNL1.

2.2 Results

2.2.1 Analysis of the Requirements of RNA Secondary Structure for MBNL1 Binding

To investigate the binding of MBNL1 to RNA targets with different structures, the oligonucleotides (CUG)_{12} and (CCUG)_{6}, which include a stabilizing GC at the end of the stem, were selected as a mimic of pathogenic, structured RNA in myotonic dystrophy (Figure 2.1). As a mimic of the natural target sites of MBNL1, a segment of cTNT comprised of 18 nucleotides (cTNT18) was selected. Although the most stable secondary structure of cTNT18 that is predicted by UNAfold is shown in Figure 2 (10-11), the T_m was determined to be 26 ± 6 °C. Thus, the structure of this RNA is likely to be an equilibrating mixture of stem loop and single-stranded forms. In contrast, (CUG)_{12} forms a more stable structure with a T_m of 71 ± 2 °C. The homodimers formed by 5’-G(CUG)_{2}C-3’ and 5’-(CUG)_{6}-3’ were found to be A-form RNA helices in crystal structures (3-4). All experiments reported here were carried out with MBNL1, which is comprised of amino acids 1-260 of MBNL, and binds to RNA with similar affinity as the full-length protein (7,12). Apparent dissociation constants for MBNL1-RNA complexes were determined using gel electrophoresis mobility shift assays and fluorescence anisotropy. The apparent dissociation constants for binding to (CUG)_{12} and (CCUG)_{6} are much larger than those for binding to cTNT18, and are similar to previously reported values (2,7). It is worth noting that
the CUG or CCUG repeat RNAs we used in experiments are much shorter than the RNA transcript in the disease state, which may contain up to hundreds of repeats (13-14). It is reasonable to believe that the longer CUG and CCUG repeat RNAs have a better binding affinity due to the availability of multiple binding sites, which could compete with MBNL1’s natural targets to disrupt the proper function of the protein.

**Figure 2.1** Most stable secondary structures predicted by UNAfold (10-11) of (CCUG)_6, (CUG)_12, cTNT32, and cTNT18. The box in the cTNT32 RNA indicates the position of cTNT18 RNA.

To investigate the contribution of each 5’-YGCY-3’ to complex stability in cTNT18, I compared the binding affinities of two cTNT18 variants in which either of the two 5’-GC-3’ sequences was replaced with a 5’-AA-3’ sequence (Figure 2.2). Adenines were chosen to replace GC because substitution with A is unlikely to introduce a new MBNL1 binding site.
Modification at the 5’-GC binding motif (cTNT18_AA1) resulted in an approximately 6000-fold increase in $K_d$, while modification at the second binding motif (cTNT18_AA2) resulted in an approximately 20-fold increase in $K_d$ (Table 1). These results suggest that both 5’-YGCY-3’ motifs contribute to the interaction with MBNL1. The different apparent dissociation constants for the two modified cTNT18 RNAs suggests that the presence of 5’-YGCY-3’ is not the only factor that influences the binding affinity of MBNL1. It is possible that the AA sequence substituted for the more 3’-GC motif can pair with 5’-UU-3’ in the loop region and make the 5’-GC more accessible to binding MBNL1.

![Figure 2.2](image)

**Figure 2.2** Electrophoretic gel mobility shift assays of MBNL1 with cTNT18 analogs, cTNT18_AA1 and cTNT18_AA2. One representative gel is presented for each RNA target. Positions of mutations in 5’-YGCY-3’ motifs in cTNT18_AA1 and cTNT18_AA2 are noted with arrows. In all gels the far right lane contains RNA only.

To probe the ability of MBNL1 to bind 5’-YCGY-3’ target sites in different structural contexts, the cTNT18 RNA was modified so that the two 5’-YCGY-3’ motifs would be separated (cTNT18_m4) or closely associated (cTNT18_m5) in the most stable secondary structures predicted by UNAfold. Similar dissociation constants were obtained for complexes formed with cTNT18 and cTNT18_m4 RNA even though in cTNT18_m4 the two 5’-YCGY-3’ motifs are forced apart by a double-stranded region formed between the original loop region of
cTNT18 and additional RNA sequence at the 3’ end of cTNT18 RNA. In contrast, when a stable stem region is located in the RNA structure to bring the two 5’YGCY3’ motifs into close proximity, as in cTNT18_m5 RNA, the complex is destabilized ($K_d > 1000$ nM). The high $T_m$’s and low folding free energies of the cTNT18_m4 and cTNT18_m5 RNAs indicate that they fold into stable secondary structures at the temperatures used for the gel mobility shift assays.

### Table 2.1 RNA melting temperatures, free energies of folding, and equilibrium dissociations constants of complexes formed with MBNL1$^{a,b}$

<table>
<thead>
<tr>
<th>RNA Name</th>
<th>$T_m$ (°C)</th>
<th>$\Delta G$ (kcal·mol$^{-1}$)</th>
<th>$K_d$ (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(CUG)$_{12}$</td>
<td>71 ± 2</td>
<td>-12.8 ± 0.4</td>
<td>150 ± 20</td>
</tr>
<tr>
<td>cTNT18</td>
<td>26 ± 6</td>
<td>ND$^b$</td>
<td>1.3 ± 0.2</td>
</tr>
<tr>
<td>cTNT18_m1</td>
<td>&gt;89</td>
<td>ND</td>
<td>0.5 ± 0.3</td>
</tr>
<tr>
<td>cTNT18_m2</td>
<td>69 ± 3</td>
<td>-7.8 ± 0.3</td>
<td>&gt;1000</td>
</tr>
<tr>
<td>cTNT18_AA1</td>
<td>ND</td>
<td>ND</td>
<td>&gt;6000</td>
</tr>
<tr>
<td>cTNT18_AA2</td>
<td>ND</td>
<td>ND</td>
<td>26 ± 4</td>
</tr>
<tr>
<td>cTNT21</td>
<td>52 ± 3</td>
<td>-3.9 ± 0.3</td>
<td>14 ± 3</td>
</tr>
</tbody>
</table>

*a*) Errors are the standard deviation of at least three independent measurements. *b*) ND indicates that the free energy of folding could not be unambiguously determined from the melting curve.

As control experiments, binding assays were also carried out with cTNT18_m1, cTNT18_m2 and cTNT18_m3. In cTNT18_m1 and cTNT18_m3, the loop regions in cTNT18 were replaced with poly(A) sequence. In cTNT18_m2, a poly(A) sequence was attached to the 3’ end of cTNT18. The poly(A) sequence was chosen because it is known that MBNL1N binds to C,U,G-rich RNA and mutation into poly(A) can therefore minimize the chance of introducing a new binding site in the RNA analogues. Based on the determined dissociation constants of these
oligonucleotides with MBNL1, changing the sequence in loop region and adding additional sequence at 3’ end of the RNA do not significantly affect the binding properties.

![Electrophoretic gel mobility shift assays of MBNL1 with cTNT18 and analogs](image)

**Figure 2.3** Electrophoretic gel mobility shift assays of MBNL1 with cTNT18 and analogs. One representative gel is presented for each RNA target. The RNA structures shown in the figures are the most stable secondary structures predicted by UNAfold (10-11). Sequences highlighted in the red boxes in the structures are speculated binding sites.

### 2.2.2 Changes of RNA Conformation upon Binding MBNL1

To investigate conformational changes of the RNA upon binding, I designed an RNA target (cTNT21, Figure 2.4 A.) with a fluorophore (FAM) and a quencher (Dabcyl) appended to
the 5’- and 3’- termini, respectively. Analysis of the structure with UNAfold suggests that this RNA forms a duplex structure in which the quencher and fluorophore are in close proximity. Consistent with this structure, the Tm of the labeled RNA (cTNT21) was determined to be 52 ± 3 °C (Figure 2.4). Because all the fluorescence measurements were performed at 20°C, the labeled RNA molecules should form stable stem loop structures in these experiments. The affinity of cTNT21 for MBNL1 was 14 ± 3 nm, which is approximately 10-fold weaker compared that of cTNT18, presumably because of the two additional base pairs, which increase the similarity of this RNA target to (CUG)_{12}.

Because quenching efficiency is directly correlated to the distance between the quencher and the fluorophore, the fluorescence intensity should increase if the two strands are separated (15). We observed a weak signal from the free RNA and a significant increase in fluorescence intensity upon addition of protein (Figure 2.4 D. E.). The increase in fluorescence intensity does not originate from the protein because a mixture of MBNL1 and unlabeled RNA gave no fluorescence signal. As a positive control, antisense DNA was added to the labelled RNA solution in 1:1 or 10:1 ratios, which resulted in a large increase in fluorescence intensity, presumably because the formation of the DNA-RNA duplex disrupts the RNA stem-loop structure. In addition, binding of MBNL1 to an RNA molecule labeled only with the fluorophore FAM resulted in only a small increase in fluorescence signal compared to the RNA labeled with both the fluorophore and quencher. Together, these results suggest that the conformation of the RNA molecule is changed upon protein binding such that the 3’- and 5’- ends of the stem loop are separated.
Figure 2.4 A. Schematic illustration of the structure of the cTNT21 RNA labelled with the fluorophore and quencher, FAM and Dabcyl. B. Melting temperature analysis of the labeled RNA; C. Gel mobility shift assay of MBNL1 and GST-MBNL1 with unlabeled RNA; D. Steady-state emission spectra of 20 nM labeled RNA in the presence and absence of GST-MBNL1. Labels in the box indicate protein concentrations; E. Quantification of fluorescence emission intensity at 520nm under different conditions; cTNT21* is labelled with FAM and Dabcyl; cTNT21f is labelled with FAM only. The concentration of MBNL1 is 1.5 µM and RNA is 20 nM.

The conformational changes of the RNA target upon MBNL1 binding suggested by the fluorescence studies described above are supported by circular dichroism spectroscopy. The CD spectra of (CUG)_{12} cTNT18p, and cTNT18 have maxima at 266, 273, and 275 nm, respectively. The CD spectrum of (CUG)_{12} is similar to those of (CUG)_{4} and (CUG)_{5} spectra that were reported previously (7,16). The (CUG)_{12} spectrum is consistent with A-form RNA, which has a
maximum between 260 and 270 nm (17-18). The maxima of cTNT18 and cTNT18p are at longer wavelength than (CUG)_{12}, which is consistent with these RNA sequences having reduced helical structure compared to (CUG)_{12}.

Upon incubation of the RNA targets with MBNL1, the peak intensity of (CUG)_{12} decreased by 23 %, that of cTNT18p decreased by 14 %, and that of cTNT18 remained nearly unchanged. The decrease in peak intensity implies reduced base stacking in the bound structures of (CUG)_{12} and cTNT18p (19-21). Upon binding MBNL1, the wavelength maximum was red-shifted by 3 nm for (CUG)_{12}, 3 nm for cTNT18, and 5 nm for cTNT18p, which is consistent with changes in RNA structure occurring upon MBNL1 binding. A decrease in peak intensity and shifting towards longer wavelengths upon MBNL1 binding was previously observed with (CUG)_{4} RNA (7). This type of change is also typical when ssDNA binding protein binds DNA and RNA targets (19,22). These data support a model in which the structures of (CUG)_{12} and cTNT18p are altered towards single-stranded upon binding MBNL1, while that of cTNT18 undergoes structural changes. The CD spectra of the bound cTNT18p and cTNT18 RNAs are nearly identical, although those of the free RNA differ, suggesting that the structures of these RNA sequences become similar upon binding MBNL1.
Figure 2.5 Circular dichroism spectrum of A) (CUG)12 RNA free (red) and bound to MBNL1 (blue) and B) Free cTNT18 (green) and cTNT21 (red) free and cTNT18 (blue) and cTNT21 (black) bound to MBNL1N. The RNA concentration is 2.5 µM and the protein concentration is 12 µM.

2.2.3 Evaluation of the RNA Binding Affinities of Truncated and Mutated MBNL1 Proteins

To investigate whether pairs of the zinc finger motifs found in MBNL1 are able to bind RNA, we created ZnF12, which is comprised of the two tandem zinc finger motifs at the N-terminus of the protein, and ZnF34, which is comprised of the more C-terminal tandem zinc finger motifs. Before carrying out binding assays, we first evaluated the zinc inclusion percentage by ICP-MS for both truncated proteins and full length protein, because truncation of protein may likely result in protein misfolding. ICP-MS analysis suggests that the truncated proteins have a similar zinc inclusion rate per zinc finger compared to MBNL1N (Figure 2.6). The complexes formed with both CUG₁₂ and cTNT18 RNAs are destabilized by truncation of either pair of zinc fingers (Table 2.2). No binding of either construct was observed to (CUG)₁₂.
RNA, while binding to cTNT18 RNA of either ZnF12 or ZnF34 was observed only at the highest protein concentrations ($K_d > 3750$ for GST-ZnF12; $K_d > 5000$ for GST-ZnF34; Figure 2.6).

![Diagram](image)

**Figure 2.6** A. Schematic illustration of the relationship between truncated proteins and full length MBNL1N. B. SDS-PAGE analysis of the purified proteins and analysis of zinc inclusion percentage by ICP-MS. C. Representative results of gel mobility shifts of truncated proteins with cTNT18 and (CUG)$_{12}$ RNAs.

To probe the contribution of individual amino acids to the recognition of RNA by MBNL1 I carried out site-directed mutagenesis involving 10 amino acids (Table 2.2). Conserved aromatic amino acid residues in all four motifs were selected. Phe202 and Tyr236 were shown in
the crystal structure of the complex formed between zinc fingers 3 and 4 and CGCUGU to interact with G and C bases in the RNA target (8). Phe36 and Tyr68 are placed at similar positions in zinc fingers 1 and 2. In addition, similarly placed aromatic amino acids in the CCCH motifs of the TIS11d protein intercalate between UU and AU dinucleotides (23). Two arginine residues that interact with RNA in the crystal structure were mutated, Arg195 and Arg201 (8).

Table 2.2. Equilibrium dissociation constants of wild type and mutant MBNL1 proteins with (CUG)\textsubscript{12} and cTNT18 RNA sequences.\textsuperscript{a}

<table>
<thead>
<tr>
<th>Protein</th>
<th>(CUG)\textsubscript{12} (nM)</th>
<th>cTNT18 (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild type</td>
<td>150±20</td>
<td>1.3±0.2</td>
</tr>
<tr>
<td>Phe22Ala</td>
<td>270±50</td>
<td>0.6±0.4</td>
</tr>
<tr>
<td>Phe36Ala</td>
<td>390±20</td>
<td>1.4±0.1</td>
</tr>
<tr>
<td>Phe54Ala</td>
<td>510±50</td>
<td>2±1</td>
</tr>
<tr>
<td>Tyr68Ala</td>
<td>540±50</td>
<td>1.6±0.2</td>
</tr>
<tr>
<td>Tyr188Ala</td>
<td>160±10</td>
<td>1.3±0.4</td>
</tr>
<tr>
<td>Arg195Ala</td>
<td>180±50</td>
<td>1.0±0.4</td>
</tr>
<tr>
<td>Arg201Ala</td>
<td>500±100</td>
<td>0.6±0.2</td>
</tr>
<tr>
<td>Phe202Ala</td>
<td>220±50</td>
<td>1.3±0.3</td>
</tr>
<tr>
<td>Arg231Ala</td>
<td>190±30</td>
<td>1.3±0.2</td>
</tr>
<tr>
<td>Tyr236Ala</td>
<td>70±20</td>
<td>0.9±0.3</td>
</tr>
<tr>
<td>Wild type – GST</td>
<td>160±30</td>
<td>4±2</td>
</tr>
<tr>
<td>ZnF12-GST</td>
<td>NB\textsuperscript{b}</td>
<td>&gt;3750</td>
</tr>
<tr>
<td>ZnF34-GST</td>
<td>NB\textsuperscript{b}</td>
<td>&gt;5000</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Errors are the standard deviation of at least three independent measurements.

\textsuperscript{b} NB indicates that no binding was observed in the gel mobility shift assay.
Each selected amino acid was replaced with alanine to avoid introducing other functional
groups or a bias for a particular peptide structure or solvent exposure. The dissociation constants
of mutant proteins with (CUG)$_{12}$ and cTNT18 RNAs are given in Table 2 (for representative gel
pictures see Figure 2.12). The binding affinities of the mutated proteins for cTNT18 were nearly
identical. The mutations resulted in up to a 4-fold destabilization of the complexes formed with
(CUG)$_{12}$ with the largest destabilization observed for Phe54Ala, Tyr68Ala, and Arg201Ala
mutations. These modest destabilizations show that none of these individual amino acids are
essential to binding.

2.3 Discussion and Conclusions

2.3.1 Truncation of MBNL1 Destabilizes Complexes with RNA

The results presented here suggest that one tandem pair of zinc fingers can bind RNA,
but with much lower affinity than MBL1N. There is conflicting data in the literature about
whether tandem pairs of zinc fingers can bind RNA. These inconsistencies may result from
different buffers being used during binding measurements (8,12). Because our goal is to compare
the binding affinities of truncated and full length MBNL1N, binding measurements were
performed using the same conditions for the truncated and full-length proteins. The data imply
that the two pairs of zinc fingers maybe cooperatively interact with RNA substrates. However,
more experiments need to be carried out in future to confirm the cooperativity between the zinc
fingers. The small complex destabilization caused by substitution of individual amino acids with
alanine may indicate that the binding energy is distributed through large portions of the zinc
finger regions. The similarity of the effects of MBNL1 truncation and mutation on the stability of
complexes formed with (CUG)$_{12}$ and cTNT18 RNAs supports similar recognition of structured and unstructured RNA targets by MBNL1.

### 2.3.2 MBNL1 Alters Structure of Helical RNA Targets

The data presented here suggest that RNA targets with high helicity, such as poly(CUG), are shifted towards single-stranded structures upon binding MBNL1. Thus, when MBNL1 binds to RNA that has a significant secondary structure, the association of the RNA with MBNL1 is in competition with the formation of the RNA secondary structure. As a result the binding affinity of structured RNA targets for MBNL1 is reduced compared to less structured RNA targets.

These observations may explain the affinities of MBNL1 for RNA targets with different secondary structures. For example, in poly(CUG) repeats, although the presence of the U-U mismatch lowers the folding energy of RNAs, the secondary structure is stable. As a result the dissociation constant of the complex formed with (CUG)$_{12}$ RNA is two orders of magnitude greater than that of the complex formed with cTNT18 RNA, a nearly single-stranded target site. The difference in binding free energy of MBNL1 for (CUG)$_{12}$ and cTNT18 RNA is similar to the calculated energetic penalty for disrupting the secondary structure of (CUG)$_{12}$ at an MBNL1 binding site (24-26). Consistent with this hypothesis, the dissociation constant of the complex formed with cTNT21 RNA is in between those of the complexes formed with (CUG)$_{12}$ and cTNT18 RNA. When the stability of the secondary structure of the RNA substrate is increased, for example cTNT18_m5 in Figure 2.3, the protein no longer binds to the RNA because the protein-RNA interaction cannot compete with the formation of the RNA secondary structure. Thus, there is an inverse relationship between the stability of the RNA secondary structure and
that of the complex. With this proposed relationship, we can expect that the more stable a stem loop structure, the lower the binding affinity. This hypothesis would also explain why (CUG)$_{12}$ has a greater apparent dissociation constant comparing with (CCUG)$_6$ (Figure 2.1) even when it has more repeats, because 2X2 bulges destabilize the duplex more than 1X1 bulges do.

### 2.3.3 RNA Secondary Structure May Play a Role in Regulation of Alternative Splicing by MBNL1

The preference of MBNL1 for RNA targets with little secondary structure may be important for MBNL1 function. MBNL1 recognizes the 5’-YGCT-3’ (Y=C/U) motifs in pre-mRNA,(1) but such motifs are expected to have a high occurrence in pre-mRNA sequences. If each of the four bases were to have a 25% chance of occupying at a certain position on the pre-mRNA sequence, the 5’-YGCT-3’ motif will be found every 64 nucleotides on average. Indeed, a sequence search by Berglund and coworkers revealed hundreds of potential MBNL1 binding motifs in pre-mRNAs that are misregulated in myotonic dystrophy (1). However, it may be possible to rule out some of these potential binding sites, by considering the preference of MBNL1 for RNA substrates with low folding energy. We carried out a sliding window analysis of pre-mRNA folding on the two best studied MBNL1 binding sites in the human genome, intron 4 of cTNT pre-mRNA and intron 11 of insulin receptor pre-mRNA (6-7,27). As expected, there is a relatively low mRNA free folding energy at previously identified binding sites, which are highlighted in black in Figure 2.7 (6-7,27).
The data presented here suggest that MBNL1 may be a new example of an RNA chaperone (28-29). It could possibly regulate alternative splicing by helping the pre-mRNAs fold into splicing favored conformations without consumption of ATP. One of the best studied RNA and DNA chaperones is the HIV-1 nucleocapsid protein, which uses CCHC zinc fingers to bind RNA (30-32). Like MBNL1, the NC protein binds with higher affinity to single-stranded than double-stranded nucleic acids. Binding of the NC protein to hairpin DNA sequences shifts the equilibrium towards a partial unwinding of the helix (33). Although it has been proposed that MBNL1 could block the spliceosome assembly to promote the exclusion of exons (7), it is still
unknown how MBNL1 promotes the inclusion of certain exons. The conformational change of RNA induced by MBNL1 may play an important role in these activities.

### 2.3.4 New perspectives for Drug Development for Myotonic Dystrophy

Our results provide new insights for destabilization of the interactions of MBNL1 with poly(CUG) or poly(CCUG) RNA sequences (34-39). First, our data suggest that binding of MBNL1 to cTNT21p results in at least a partial dissociation of the duplex structure. Therefore, to bind to an RNA target, MBNL1 may need to overcome an energy barrier of melting the RNA and any molecule that stabilizes poly(CUG) or poly(CCUG) in a duplex structure could inhibit binding of MBNL1. Second, the results of the experiments performed with truncated MBNL1 proteins suggest that for an inhibitor that targets the protein, blocking the RNA binding of one pair of zinc fingers may be sufficient to destabilize the complex because the protein with only one pair of zinc fingers lose its RNA binding affinity significantly. These findings suggest approaches for drug development and may contribute to an understanding of molecules identified from screening approaches.

### 2.4 Materials and Methods

#### 2.4.1 Plasmid

The expression vector pGEX-6p-1/MBNL1N was obtained from Maurice S. Swanson (University of Florida, College of Medicine, Gainesville, FL, USA). The expressed protein has 272aa with two pairs of zinc fingers. It contains amino acid 1-260 from human MBNL1 and a His₆-tag at its C-terminus.

#### 2.4.2 Cloning
To construct the vector for ZnF12 truncated protein, the zinc finger 1 and 2 region of pGEX-6p-1/MBNL1N was amplified with the primer 5’-CGGCGGGATCCATGGCTGTTAGTGTCA-3’ and the primer 5’-CGGCCGCTCGAGGATTACTCGTCCATT-3’, digested with BamHI and XhoI (Invitrogen), and inserted into pGEX-6p-1 backbone vector containing a His$_6$-tag. The pGEX-6p-1 vector was prepared by digesting the pGEX-6p-1/MBNL1N vector with BamHI and XhoI (Invitrogen), followed by purification of the large fragment with agarose gel using QIAquick Gel Extraction Kit (QIAGEN). To construct the vector for ZnF34 truncated protein, the zinc finger 3 and 4 region of pGEX-6p-1/MBNL1N was amplified using the primer 5’-CGGGATCCATGTTAATGCGAACAGAC-3’ and the primer 5’-CCGCTCGAGCTGTTATTGGG-3’ and subcloned into pGEX-6p-1 following the same procedure as for ZnF12.

Figure 2.8 Schematic illustration of the procedure for subcloning

2.4.3 Site Directed Mutagenesis
For site directed mutagenesis, the pGEX-6p-1/MBNL1N vector was amplified using primers as following:

Phe22Ala
Forward Primer 5’-GTATGTAGAGAGGCCCAGAGGG-3’
Reverse Primer 5’-GTCCCCCCTCTGGGCTCTCTAC-3’

Phe36Ala
Forward Primer 5’-GGAATGTAAGCTGCACATCCTTCG-3’
Reverse Primer 5’-CGAAGGATGTGCAGCTTTACATCC-3’

Phe54Ala
Forward Primer 5’-ATCGCCTGCGCTGATTCATT-3’
Reverse Primer 5’-AATGAATCAGCGCAGGCGAT-3’

Tyr68Ala
Forward Primer 5’-GAACGTGCAAAGCTCTTCATCCA-3’
Reverse Primer 5’-TGGATGAAGAGCTTTGCAGTTC-3’

Tyr188Ala:
Forward Primer 5’-TATGTCGAGAGGCCCAACGT-3’
Reverse Primer 5’-ACGTTGGGCCTCTCGACATA-3’

Arg195Ala:
Forward Primer 5’-AATTGCAACGCGAGAAAATG-3’
Reverse Primer 5’-CATTTTCTCCTGCTTTGCAATT-3’

Arg201Ala:
Forward Primer 5’-GAAAATGATTGTGCGTTTGCTC-3’
Reverse Primer 5’-GAGCAAAACGCACAATCATTTC-3’
Phe202Ala:
Forward Primer 5’-TGATTGTCGGGCTGCTCATCC-3’
Reverse Primer 5’-AGGATGAGCAGCCCGACAATCA-3’
Arg231Ala:
Forward Primer 5’-GGAGATGCTCTGCGGAAAAGTG-3’
Reverse Primer 5’-CACTTTTCCGCAGACATCTCC-3’
Tyr236Ala:
Forward Primer 5’-AAGTGCAAAAGCCTTTCATCCC-3’
Reverse Primer 5’-GGGGATGAAAGGCTTTGCACTT-3’

The PCR product was treated with DpnI (Invitrogen), transformed into X11-blue competent cells (Stratagene) and isolated using QIaquick Spin Miniprep Kit (QIAGEN). The mutations were confirmed by DNA sequencing.

2.4.4 Recombinant Protein Preparation and Purification

Using BL21-CodonPlus(DE3)-RP competent cells (Stratagene), the expression of proteins was induced with 1mM IPTG at OD$_{600}$ 0.6 in LB media with ampicillin for 2 hours at 37 °C. Bacterial cells were collected by centrifugation and were then resuspended in lysis buffer containing 25 mM Tris-Cl (pH=8), 0.5 M NaCl, 10 mM imidazole, 2 mM BME, 5% glycerol, 0.1% Triton X-100, 2 mg/ml lysozyme, 0.1mM PMSF, 1 µM pepstatin, and 1uM leupeptin, and
sonicated six times for 15 s each. The cell pellet was centrifuged, and the supernatant was collected and filtered through a 45 µm Millex Filter (Millipore).

To purify MBNL1N and its related truncated and mutant proteins, Ni-NTA agarose (QIAGEN) was incubated with the lysate for 1 h at 4°C and washed with a washing buffer containing 25 mM Tris-Cl (pH=8), 0.5 M NaCl, 20 mM imidazole and 0.1% Triton X-100, followed by elution with elution buffer of 25 mM Tris-Cl (pH=8), 0.5 M NaCl, 250 mM imidazole and 0.1% Triton X-100.

For fluorescence and anisotropy measurements and ICP-MS analysis, the eluent containing the GST fusion protein was dialyzed against a buffer containing 20mM Tris-Cl (pH=7.5), 100mM NaCl, 5mM MgCl₂, 1mM BME. For electrophoresis shift assay, the eluent was collected and incubated with Glutathione Sepharose 4B (GE Healthcare) for 1 hour at 4°C. After washing with a buffer containing 25 mM Tris-Cl (pH=8), 300 mM NaCl, 5 mM BME and 0.1% TritonX-100, the beads were collected and incubated with PreScission Protease (GE Healthcare) overnight at 4°C. After being cleaved from the beads, the protein was collected in the flow-through of the column, and was concentrated with a Microcon Centrifugal Filter 3000 MWCO (Millipore).

MALDI mass spectrometry was used to confirm the identity of the protein, while SDS-PAGE stained with Coomassies Blue (G-250) was used to evaluate the purity. The concentrations of proteins were determined by Bradford assays (Bio-Rad).
Figure 2.9 Representative coomassie stained SDS gel picture for analyzing the purity of MBNL1 and the mutants
Figure 2.10 Representative MALDI-MS for MBNL1 and mutant proteins. (m/z=30kDa for z=1; m/z=15kDa for z=2; m/z=60kDa for MBNL1 dimer)
Figure 2.10 (cont.)
2.4.5 Radiolabeling and Purification of RNA

All of the RNA substrates were purchased HPLC purified from IDT DNA. RNA was labeled with 50 µCi[γ-32P] ATP using T4 Kinase (Invitrogen). After incubation for 2 hours at 37°C, the reaction mixture was extracted with (25:24:1) phenol/chloroform/isoamyl alcohol and (24:1) chloroform/isoamyl alcohol. Radiolabeled RNA was then precipitated with 75% ethanol. After precipitation, trace ethanol was removed under reduced pressure, and the RNA was diluted into water and stored at -20°C.

2.4.6 Electrophoresis Mobility Shift Assay

Radiolabeled RNAs were heated to 95 °C for 2 mins and then placed on ice for 20 mins. RNAs were diluted to 0.2 nM with buffer containing 27 mM Tris-Cl (pH=7.5), 66 mM NaCl, and 6.7 mM MgCl₂. Proteins were serially diluted in buffer containing 175 mM NaCl, 5 mM MgCl₂, 20 mM Tris-Cl (pH=7.5), 1.25 mM BME, 12.5% glycerol, 2 mg/ml BSA, 0.2% TritonX-100 and 0.1 mg/ml heparin. Protein and RNA solutions were mixed in a volume ratio of 1:1 and incubated at room temperature for 20 mins. After loading on a 6% acrylamide gel, the gel was run at 360V in 0.5X Tris-Borate buffer for 30 mins at 4°C, and dried and visualized on a Molecular Dynamics Storm Phosphorimag. The apparent $K_d$ values were obtained by fitting using the following equation: fraction RNA bound=1/(1+$K_d$/[P]), where [P] is the concentration of protein.
Figure 2.11 Representative results of gel mobility shift assays for MBNL1 with (CCUG)$_6$, (CUG)$_{12}$ and cTNT RNAs. The protein concentration is noted on the top of the gel. The RNA concentration is 0.1 nM. The gel was run at 360V in 0.5X TB buffer.
Figure 2.12 Representative results of gel mobility shift assays of mutant proteins with (CUG)\textsubscript{12} RNA. The protein concentration is noted on top of each gel. The RNA concentration is 0.1 nM. The gels were run at 360V in 0.5X TB buffer.
Figure 2.12 (cont.)
2.4.7 Thermal Denaturation Study

The melting temperatures of the RNAs were measured on a Shimadzu UV2450 spectrophotometer equipped with a temperature-controller. The absorbance of each RNA solution (~3 μM RNA in 27 mM Tris-Cl (pH=7.5), 66 mM NaCl, and 6.7 mM MgCl₂) was monitored at 260nm from 10°C to 90°C at a ramp rate of 0.5 °C/min. The melting temperature ($T_m$) of each sample was determined from the maximum point of the first derivative of the melting curve with LabSolutions – Tm Analysis version 1.0. Free folding energies were determined from the melting curve using Meltwin 3.5. (http://www.meltwin.com/)
Figure 2.13 Representative melting curves of RNAs. The absorbance of each sample was monitored at 260nm from 10°C to 90°C.
2.4.8 Circular Dichroism Experiments

CD measurements were conducted on a Jasco J-700 spectropolarimeter. Protein or RNA samples were dissolved in 250 µl buffer containing 27 mM Tris-Cl (pH=7.5), 66 mM NaCl, and 6.7 mM MgCl₂. The final concentration for protein is 12 µM and the final concentration for RNA is 2.5 µM. Before taking measurements, RNA samples were heat shocked by incubation at 95°C for 5 min and then cooled on ice for 20 min. The spectropolarimeter was scanned from 300 to 250 nm at a rate of 200 nm/min. Data points were acquired every 0.5 nm. The spectrum obtained were averages of 50 accumulations.

2.4.9 Steady-State Fluorescence Spectroscopy and Anisotropy

All of the fluorescein labeled RNAs were purchased HPLC purified from IDT DNA. For fluorescence anisotropy measurements, the fluorescein was always labeled at 3’ end of the RNA.
Figure 2.14 Curve fitting of anisotropy data for determining dissociation constants of GST-MBNL1N with different RNA substrates. Fluorescein labeled RNA (20nM) were incubated with GST-MBNL1 of different concentration.

Steady-state fluorescence spectra were recorded on a FluoroMax-3 spectrometer equipped with a 150W xenon lamp and modified Czerny-Turner spectrometers in both the excitation and emission position utilizing Datamax Spectroscopy Software. Fluorescence emission was measured at a 90° angle. The fluorometer was interfaced with a Neslab RTE-111 temperature controller with a remote sensor. The excitation and emission slits were both set at 4nm and the excitation wavelength was set at 485nm. All fluorescence scans were recorded at
20°C using 1 nm wavelength increments. Samples prepared for fluorescence measurement were all dissolved in a buffer containing 100 mM NaCl, 5 mM MgCl₂, 20 mM Tris-Cl (pH=7.5), 1 mM BME.

Fluorescence anisotropy measurements were carried out on the same instrument with the excitation wavelength at 485 nm and emission wavelength at 520nm. Before measurements, the G value was determined to be 0.93 for fluorescein labeled RNAs at 4 °C. Samples were prepared by mixing 40 nM fluorescein labeled RNA solution with protein solution of different concentrations in a 1:1 ratio in the same buffer as steady state fluorescence measurements. All the measurements were carried out at 4°C. The excitation and emission slits were both set to be 4 nm.

2.4.10 Sliding Window Analysis of mRNA Folding Energy

The RNA folding energy was calculated using the software RNAstructure 4.6, using the default setting: folding occurred at 37°C; the concentrations of RNAs were 1µM; window length was set to 29nt.

2.5 References


transcripts that cause RNA dominant disease: application to myotonic muscular

CHAPTER 3
Correlation between the Binding of RNA to MBNL1 and RNA Secondary Structure Stability

3.1 Introduction

As was discussed in Chapter 1.6, there are discrepancies in the published literature about the RNA secondary structure preference of MBNL1 (1-7). The work presented in Chapter 2 suggests that MBNL1 prefers to bind to ssRNA compared to dsRNA and the two pairs of 5’YGCY3’ motifs in MBNL1 do not need to be base paired for MBNL1 to bind. The fluorescence analysis and circular dichroism spectrum analysis reveal that MBNL1 induces conformational changes in RNA upon binding. These results led me to perform an in depth characterization of the correlation between RNA secondary structure stabilities and MBNL1 binding properties. I reasoned that if RNA targets are melted upon binding, then inverse correlations between the secondary structure stabilities of RNA targets and MBNL1 binding affinities or rates should be observed.

I used an electrophoresis mobility shift assay along with fluorescence stopped-flow spectroscopy to study the correlation between RNA stability and both thermodynamics and kinetics of MBNL1 binding. In addition, FRET analysis was performed to determine the distance change in RNA structure upon binding to MBNL1. Based on these results, we propose that the binding of MBNL1 to stem-loop RNA may destabilize the RNA secondary structure and lead to partial melting of the stem region in the RNA substrates. Consistent with this hypothesis, an
RNA structure stabilizer was shown to destabilize the interaction between MBNL1 and CUG repeat RNA.

### 3.2 Results

#### 3.2.1 GST-MBNL1 Induced FRET Changes in RNA Substrate

In chapter 2, I showed that the stem loop RNA may be melted upon binding to MBNL1 protein (8). To investigate the change in distance between the two ends of a stem-loop RNA upon binding to MBNL1, we designed an artificial RNA substrate containing the MBNL1 binding site in cTNT pre-mRNA with two additional GC base pairs at the end of the stem to stabilize the secondary structure. Fluorescein and TAMRA were covalently attached to the 3’- and 5’- end of the RNA substrate, respectively. The $T_m$ of the RNA was determined to be 58±2°C by UV melting analysis. The most stable secondary structure of the RNA predicted by UNAfold is shown in Figure 3.1 (9). The binding of this dual labeled RNA to GST-MBNL1 was confirmed by steady state fluorescence measurements that monitor the dose-dependent changes in fluorescence intensity upon addition of GST-MBNL1 (Figure 3.1 C) and the $K_d$ was determined to be 1.0±0.1 μM.

The fluorescence emission spectrum of both free and bound RNA were obtained at 20 °C (Figure 3.1 D). Approximately 1-fold increase in donor fluorescence and 2-fold increase in acceptor fluorescence were observed upon incubation of the dual labeled RNA with the protein. The change in emission spectrum suggests that in the free RNA, the donor and acceptor are in close proximity and contact quenching dominates, in which the emission of both donor and acceptor is diminished. FRET quenching becomes the dominant mechanism when the RNA is
bound by MBNL1 and the donor fluorophore has moved away from the acceptor fluorophore (10). To probe the quenching modes and their relation to donor-acceptor distance, we designed an antisense oligonucleotide that could separate the donor from the acceptor by formation of a DNA-RNA duplex. Upon incubation of the dual labeled RNA with antisense DNA in a 1:10 ratio, we observed an approximately 1.5-fold increase in donor fluorescence and a 2-fold decrease in acceptor fluorescence compared to bound state. This observation supports the presence of FRET quenching because if only contact quenching is present, the acceptor fluorescence would have increased or remained unchanged. It also suggests that FRET quenching is the dominant quenching mode in MBNL1 bound RNA. It is important to use a fluorescent quencher instead of a dark quencher that was used in our previous report (8) in order to rule out the possibility that the formation of the protein-RNA complex only disrupts the static interaction between donor and acceptor but does not melt the stem-loop RNA. In addition, the relatively long Förster radius ($R_0$) of fluorescein/TAMRA pair enables us to estimate the distance between the donor and acceptor.

To determine the maximal GST-MBNL1 induced energy transfer efficiency, dual and singly labeled RNAs were separately incubated with GST-MBNL1 at protein concentrations above the dissociation constant, $K_d$. The energy transfer efficiency and distance between the donor and acceptor were calculated to be 54 Å using the method described in the Materials and Methods. The fluorescein/TAMRA Förster radius, 50 Å, is similar to the determined distance, which is required for an accurate calculation of the distance. These results suggest that the two strands in the stem region are completely melted upon binding to the protein.
Figure 3.1. A. Schematic illustration of the structure of the RNA labeled with fluorescein and TAMRA. The fluorescein and TAMRA are the donor and acceptor, respectively (emission/excitation maxima fluorescein 520/495 nm, TAMRA 580/565 nm); B. Curve fitting for fluorescence intensity of the donor fluorophore against protein concentration; C. Thermal melting analysis of the dual labeled RNA; D. Steady-state emission spectra of 20 nM labeled RNA in the presence and absence of GST-MBNL1 (blue line and green line). The pink line shows the emission spectrum of the labeled RNA with antisense DNA in a 1:10 ratio; E. Quantification of fluorescence emission intensity of different labeled RNA for FRET analysis at 520 nm with excitation wavelength of 485 nm. The bars labeled with FAM only and TAMRA only indicate the fluorescence intensity obtained with donor- or acceptor- singly labeled RNA.

3.2.2 The Stability of RNA Secondary Structure is Correlated with the Binding Affinity of MBNL1
Figure 3.2. A. Structures of RNA substrates used; B. Representative gel pictures of the electrophoresis mobility shift assay of RNA substrates interacting with MBNL1.

The results of the FRET analysis led us to examine whether stem-loop stability of the RNA is correlated with its binding affinity with MBNL1. Thus, three stem-loop RNA targets, SL1, SL2 and SL3, with varying secondary structure stability were designed, synthesized, and used in affinity assays with MBNL1 (Figure 3.2 A). These oligonucleotides contain identical MBNL1 binding sites closed by three different base pairs to stabilize the stem-loop structure. The stability of SL2 was modified compared to SL1 by replacing the G-C base pair closing the bulge in SL1 with an A-U base pair in SL2. Two base pairs in SL1 were similarly replaced to
form SL3. The dissociation constants of the RNA-MBNL1 complexes were determined by an electrophoresis mobility shift assay at 4 °C, (Figure 3.2 B) and the RNA folding free energies at 4 °C were determined by UV melting analysis (Figure 3.3).

**Table 3.1.** Equilibrium dissociation constants of complexes formed between RNA substrates and MBNL1 and RNA folding free energies

<table>
<thead>
<tr>
<th></th>
<th>$K_d$ (nM)</th>
<th>$\Delta G_{\text{binding}}$ (kcal/mol)</th>
<th>$\Delta G_{\text{folding}}$ (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SL1</td>
<td>308±71</td>
<td>-8.2</td>
<td>-8.6</td>
</tr>
<tr>
<td>(most stable)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SL2</td>
<td>130±40</td>
<td>-8.6</td>
<td>-7.1</td>
</tr>
<tr>
<td>SL3</td>
<td>10.1±1.7</td>
<td>-10.1</td>
<td>-4.3</td>
</tr>
<tr>
<td>(least stable)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The data shown in Table 3.1 suggest that the secondary structure of SL1 RNA is more stable than that of SL2 RNA, which is more stable than that of SL3 RNA. Conversely, the binding affinity of SL1 RNA for MBNL1 is the weakest while the binding affinity of SL3 RNA is the strongest. Based on these data, we conclude that the RNA folding free energies are inversely correlated with MBNL1 binding energies. The RNA target sites with low GC content in the stem bind with higher affinity to MBNL1 than the RNA target sites with high GC content. The difference in stem-loop stability ($\Delta \Delta G$) appears to follow this same trend because the difference in RNA folding energy between the two is greater when the difference in binding energy between the two RNA substrates is greater.
3.2.3 Both Association and Dissociation Rates are Correlated with Stem-loop Stability

Based on the equilibrium analysis, increasing the stabilities of the RNA secondary structure increases the dissociation constant. The equilibrium dissociation constant is the ratio of off-rate and on-rate constants. Therefore, we probed whether the off-rate or the on-rate or only both are affected by the stability of the secondary structure of the target site RNA. Stopped-flow kinetics experiments were performed to obtain the rate constants for binding reactions. The three RNA constructs, SL1, SL2 and SL3, which were used in the equilibrium experiments described
above were covalently labeled with fluorescein at the 5’ ends and used in experiments to probe the binding kinetics with MBNL1. Fluorescein labeled RNA and GST-MBNL1 were mixed and the fluorescence intensity change was monitored using a 495 nm band-path filter. In addition, dissociation constants of the fluorescein labeled RNA with GST-MBNL1 were determined by steady state fluorescence measurements (Table 3.2). It is worth noting that the dissociation constants obtained here were different from what we reported with gel shift assays because the GST tag was not removed and the RNA was labeled with fluorescein. In spite of these differences, both sets of data suggest similar trends for dissociation constants in response to changing the stability of the stem.

Table 3.2 Summary of association and dissociation constants of fluorescein labeled RNA interacting with GST-MBNL1

<table>
<thead>
<tr>
<th></th>
<th>$K_d$/µM</th>
<th>$k_{on}$</th>
<th>$k_{off}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SL1 (most stable)</td>
<td>1.7(±0.5)</td>
<td>1.33(±0.11)×10^4</td>
<td>0.023(±0.007)</td>
</tr>
<tr>
<td>SL2</td>
<td>1.4(±0.4)</td>
<td>2.66(±0.19)×10^4</td>
<td>0.037(±0.011)</td>
</tr>
<tr>
<td>SL3 (least stable)</td>
<td>0.29(±0.06)</td>
<td>1.16(±0.13)×10^6</td>
<td>0.336(±0.079)</td>
</tr>
</tbody>
</table>

In our measurements, we observed that the formation of the RNA-protein complex resulted in an increase in fluorescence intensity, probably because guanine can quench the fluorescein when RNA is folded. Binding of protein melts the stem, dissociates the G-C base pair and therefore, separates the fluorescein from guanine. In all of the kinetic data obtained, the time courses of fluorescence increase follows a single exponential equation with $R^2>0.95$ suggesting that the binding reaction follows first-order or pseudo-first-order kinetics. Performing experiments with different protein concentrations revealed that the rate constants were linearly correlated with protein concentrations (Figure 3.4 B and C). These results suggest that the
increase in fluorescence intensity results from a pseudo-first-order reaction and involves the interactions of two species, the protein and the RNA.

![Graphs showing fluorescence intensity over time for SL1, SL2, and SL3 RNA reacting with GST-MBNL1.](image)

**Figure 3.4 A.** Plot of stopped-flow kinetic data for 20nM of SL1, SL2, or SL3 RNA reacting with 14µM of GST-MBNL1. Data are fit with eq(6) described in materials and methods in this chapter; B. and C. Apparent rate constants for association between SL1(B.)/SL3(C.) RNA and MBNL1 protein were determined and the rate constants were shown to be linearly correlated with the protein concentration used.

After fitting the curves, we calculated the dissociation and association rate constants using the observed rate constants and the equilibrium dissociation constants obtained from steady-state fluorescence measurements using eq8 and eq9 described in materials and methods of this chapter. As can be seen in Table 3, both association and dissociation constants are correlated.
with the RNA secondary structure stability. SL1 RNA, the RNA target with the most stable secondary structure, gives the smallest rate constants, while SL3 RNA, the one with weakest stem-loop strength, gives the greatest rate constants among the three RNA substrates. In addition, the amplitude increase for association constants appears to be always greater than the amplitude increase for dissociation constants when the strength of the stem-loop becomes weaker. This may suggest that the association reaction is affected more by the stability of the RNA secondary structure than is the dissociation reaction.

3.2.4 RNA Secondary Structure Stabilizer as a Potential Inhibitor of MBNL1-RNA Interactions

Because both the association and dissociation rates of MBNL1-RNA complexes are inversely correlated with the RNA stem-loop stability, we speculated that a small molecule that stabilizes the RNA secondary structure might destabilize MBNL1-RNA interactions. Specifically, we chose to investigate the ability of mitoxantrone to destabilize the MBNL1-RNA complex. Mitoxantrone is a classic intercalating reagent with promising potential in cancer chemotherapy. It has been shown to intercalate into both double stranded RNA and DNA (11) and increase the melting temperature of RNA. Mitoxantrone is soluble in aqueous solution and exhibits no fluorescence or UV absorption within the range of wavelengths used in experiments involving fluorescein.
The RNA target site CUG₁₂ was used to evaluate the ability of mitoxantrone to destabilize MBNL1-RNA complexes. CUG₁₂ RNA can fold into a stable hairpin structure and the melting temperature ($T_m$) was determined to be 72 °C by UV-melting analysis. (Figure 3.5). Upon incubation of the RNA with mitoxantrone, the melting temperature ($T_m$) of the RNA increased in a dose-dependent manner suggesting that mitoxantrone can stabilize the RNA secondary structure. (Figure 3.6 B) Gel shift assays performed with CUG₁₂ RNA and MBNL1 protein revealed that supplementation of mitoxantrone increased the dissociation constant 3-fold. (Figure 3.6 C) The binding of mitoxantrone to CUG₁₂ was confirmed in steady state fluorescence experiments in which the incubation of fluorescein labeled RNA with mitoxantrone led to a decrease in signal. (Figure 3.6 D)
Figure 3.6 A. Structure of Mitoxantrone. B. Increase in melting temperature of CUG₁₂ RNA supplemented with mitoxantrone of different concentrations. C. CUG₁₂ incubated w/o mitoxantrone (50µM) interacts with MBNL1 protein in gel electrophoresis mobility shift assay. D. Analysis of the binding between CUG₁₂ RNA and mitoxantrone by fluorescence.

3.2.5 Evaluation of the Binding Affinities and RNA Stabilization Effect of Small Molecule Ligands for CUG Repeat RNA

As part of the collaboration work with Zimmerman’s group in University of Illinois, we analyzed the small molecule induced RNA melting temperature changes for a series of potential small molecule inhibitors for MBNL1-RNA interaction as well as the RNA binding affinities for these molecules. These small molecule inhibitors were rationally designed based on a simple ligand that selectively binds to stem-loop RNA with U-U mismatches (12). Dimmerization of the
previously reported ligand is expected to give rise to better affinity and selectivity. Therefore, a library of dimer molecules with different connectivities and linker lengths were prepared and evaluated here for their activities (Table 3.3 and Figure 3.7). Amin Haghighat Jahromi in University of Illinois prepared the dimer molecules.

**Table 3.3** Melting temperature change induced by small molecule ligands and binding affinities of small molecule ligands to (CUG)\(_{12}\) RNA\(^a\)

<table>
<thead>
<tr>
<th>Ligand</th>
<th>T(_m) (°C)(^b,c)</th>
<th>K(_d) (μM)</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA only</td>
<td>64.5± 1.1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>AJ1</td>
<td>66.6± 0.4 (1 eq.), 66.8± 0. (3 eq.)</td>
<td>23± 4</td>
<td>5% DMSO</td>
</tr>
<tr>
<td>AJ2</td>
<td>N. D.</td>
<td>N. D.</td>
<td>insoluble</td>
</tr>
<tr>
<td>AJ3</td>
<td>N. D.</td>
<td>70± 15</td>
<td>5% DMSO</td>
</tr>
<tr>
<td>AJ4</td>
<td>65.7± 0.6 (1 eq.), 66.1± 0.5 (3 eq.)</td>
<td>221± 58</td>
<td>5% DMSO</td>
</tr>
<tr>
<td>AJ6</td>
<td>N. D.</td>
<td>201± 74</td>
<td>5% DMSO</td>
</tr>
<tr>
<td>AJ8</td>
<td>70.3± 4.6 (1 eq.), 77.8± 2.9 (3 eq.)</td>
<td>2.03± 0.68</td>
<td>aqueous</td>
</tr>
<tr>
<td>AJ10</td>
<td>67.3± 0.6 (1 eq.), N. D. (3 eq.)</td>
<td>0.31± 0.10</td>
<td>aqueous</td>
</tr>
<tr>
<td>AJ11</td>
<td>65.7± 0.2 (1 eq.), 66.0± 0.4 (3 eq.)</td>
<td>N. D.</td>
<td>insoluble</td>
</tr>
<tr>
<td>AJ12</td>
<td>N. D.</td>
<td>N. D.</td>
<td>insoluble</td>
</tr>
<tr>
<td>AJ14</td>
<td>73.5± 1.4 (1 eq.), N.D. (3 eq.)</td>
<td>2.88± 0.82</td>
<td>aqueous</td>
</tr>
<tr>
<td>AJ15</td>
<td>66.4± 0.2 (1 eq.), 69.3± 0.8 (3 eq.)</td>
<td>&gt;250</td>
<td>aqueous</td>
</tr>
<tr>
<td>AJ16</td>
<td>72.2± 2.1 (1 eq.), 76.0± 1.4 (3 eq.)</td>
<td>3.14± 1.09</td>
<td>aqueous</td>
</tr>
</tbody>
</table>

a, N.D. stands for not determined; b. eq. stands for equivalent; c, UV melting analysis is carried out in 1 X PBS buffer.

The melting temperatures were determined by incubating indicated equivalents of small molecule ligand with (CUG)\(_{12}\) RNA. The increase in RNA melting temperature suggests that the RNA secondary structures were stabilized upon interaction with small molecule ligands. The binding affinities of small molecule ligands with (CUG)\(_{12}\) RNA were evaluated by steady state fluorescence measurements using TAMRA labeled CUG repeat RNA. Based on the data presented in Table 3.3, small molecule ligands that show high affinities towards CUG repeat RNA generally induce great melting temperature changes for (CUG)\(_{12}\) RNA. These ligands
appear to be promising candidates for inhibitors of MBNL1-CUG repeat RNA interaction.

Further experiments have been carried out by Amin Haghighat Jahromi using SPR to evaluate the inhibitory activity of the ligands by SPR.

Figure 3.7 Structures of small molecule ligands for CUG repeat RNA. The corresponding names of the ligands are given at bottom of the structure.
3.3 Discussion and Conclusions

3.3.1 Target RNA Provides a Flexible Conformation for MBNL1 Recognition

The increase in end-to-end distance of the RNA molecule measured by FRET suggests that the binding of MBNL1 proteins to stem-loop RNA results in a large conformational change of the RNA structure. The end to end distance of the bound RNA construct is longer than the distance between the two zinc fingers in one tandem pair (7). Therefore, it is possible that the two 5’YGCY3’ binding sites in the RNA structure are bound independently by the two zinc fingers in MBNL1. The flexibility of the non-base paired RNA may allow the MBNL1 protein to interact with two 5’YGCY3’ motifs simultaneously without introducing a large conformational change. However, more evidence is required to confirm this hypothesis.

3.3.2 RNA is Partially Melted upon Binding

The inverse correlation between RNA folding energy and RNA binding affinity indicates that MBNL1 must overcome the energy barrier of melting the RNA upon binding. It is interesting to note that the ΔΔG of RNA folding is greater than the ΔΔG of binding, which may imply that the base-stacking interactions within the stem region are not fully disrupted or the stem is not fully melted upon binding of RNAs to the protein (Table 3.4). In fact, some of the discrepancies in reports about MBNL1 binding preferences may arise from the partial melting of RNA secondary structure by MBNL1. Some reports showed that MBNL1 has a very good affinity towards less structured RNA target sites (7,13-15), but there are other reports suggesting that MBNL1 recognizes the stem-loop structure of RNA (4-5). It is possible that there are only
2-3 open base pairs at the binding sites, while the other base pairs in the stem-loop RNA remain closed with the base stacking interactions maintained.

<table>
<thead>
<tr>
<th>Table 3.4 Comparison of differences in folding and binding energies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binding (kcal/mol)</td>
</tr>
<tr>
<td>ΔΔG(SL1-SL2)</td>
</tr>
<tr>
<td>ΔΔG(SL2-SL3)</td>
</tr>
</tbody>
</table>

Besides thermodynamic analysis, the kinetic studies with different RNA substrates also suggest that a stable stem-loop structure can slow down the binding reactions possibly by blocking MBNL1-RNA interactions. The ability of mitoxantrone to destabilize MBNL1-RNA interactions, presumably by stabilizing the secondary structure, is consistent with these findings. Although an intercalator has been shown to be a promising inhibitor for MBNL1-RNA complex, it is worth noting that the intercalator itself cannot be used as a drug for treatment of myotonic dystrophy due to the lack of binding specificity. A library of intercalator derivatives could be screened to find an ideal intercalator that selectively interacts with CUG repeat RNA for treatment of myotonic dystrophy. For example, the two basic side chains of mitoxantrone (Figure 3.6 A) could be ideal positions for introducing modifications and making derivatives to modulate binding specificity and affinity. In this regard, our research presents a new perspective for drug identification of myotonic dystrophy.

3.3.3 RNA Melts During the Course of MBNL1 Binding

In addition to the conformational change of the RNA and the correlation between the RNA secondary structure and binding affinity, what we sought to determine is whether RNA melts during the course of MBNL1-RNA binding or before binding occurs. Three possible
mechanisms are shown in Figure 3.8. In mechanism (1), MBNL1 is directly involved in melting the RNA. In mechanisms (2) and (3), MBNL1 captures the single-stranded RNA after melting. The difference between mechanisms (2) and (3) lies in whether melting of the RNA or binding of MBNL1 is rate-determining.

Figure 3.8 Schematic illustration of three possible mechanisms for MBNL1 interacting with stem loop RNA

Based on the results described in this thesis, only the one step mechanism (1) is supported by our kinetic data. If the binding of MBNL1 follows mechanism (2), the dissociation rate should be independent of RNA secondary structure stability because the refolding of RNA is after the rate determining step. Yet, as we have observed, the dissociation rate constants of MBNL1 interacting with stem-loop RNA are correlated with the stem stability, thus ruling out mechanism (2). If the binding reaction follows mechanism (3), the association reaction should be a “true” first order reaction only with respect to RNA concentration. However, our data have suggested that the association reaction is a pseudo-first order reaction, ruling out mechanism (3). Thus, mechanism (1) is consistent with the current data suggesting that MBNL1 is directly involved in the process of melting the stem-loop RNA.

3.3.4 Conclusions
The data presented here provide insight into the role of target RNA secondary in MBNL1-RNA interactions. MBNL1 induces a large conformational change of RNA upon binding and the RNA secondary structure stabilities are correlated with the MBNL1 binding properties. Although we only tested a few artificially designed RNA substrates in this study, it is highly probable that MBNL1 could also change the secondary structure of its natural pre-mRNA targets and the hairpin structures of poly(CUG) RNA in vivo. The structural changes of the pre-mRNA or poly(CUG) RNA may potentially have impact on the interactions of the these RNA with other protein molecules and therefore contribute to the change of splicing pattern or formation of protein-RNA aggregates observed in myotonic dystrophy. In future, it would be helpful to probe the kinetic process of MBNL1 interacting with long poly(CUG) RNA to study the formation of nuclei foci in myotonic dystrophy patients’ cells. In addition, small molecules that stabilize RNA structure have been shown to be promising candidates for drugs against myotonic dystrophy.

3.4 Materials and Methods

3.4.1 Protein and RNA

The protein was prepared as described in Chapter 2.4.4. The expressed protein GST-MBNL1 is comprised of four zinc finger domains (1-260aa) from human MBNL1 and contains a GST tag at the N-terminus. For EMSA analysis, the GST tag was removed by site-specific protease cleavage.

The PAGE purified oligonucleotides used in FRET analysis were purchased from Dharmacon. It was labeled with fluorescein and TAMRA at the 5’- and 3’- ends, respectively.
For fluorescence stopped-flow measurements and electrophoresis mobility shift assays, the oligonucleotides were purchased from Integrated DNA Technology and were HPLC purified.

3.4.2 Thermal Denaturation Study

![Graph showing melting curves](image)

**Figure 3.9** Representative melting curves showing ligand induced melting temperature change of RNA

The melting temperatures of the RNAs were measured on a Shimadzu UV2450 spectrophotometer equipped with a temperature controller. The path length of the cuvettes used was 1 cm. Small molecule ligands of different concentrations were mixed with the RNA solutions (3.3 µM RNA in 1X PBS buffer). The absorbances of each RNA solution were recorded at 260 nm with a slit width of 1 nm from 10 °C to 95 °C at a ramp rate of 0.5 °C/min. Each profile for melting temperature analysis was generated by subtracting the absorbances of small molecule ligand only from that of RNA plus small molecule. Melting temperatures were determined by fitting the melting curve using Meltwin 3.5. ([http://www.meltwin.com/](http://www.meltwin.com/)).
3.4.3 Steady State Fluorescence Measurements

Steady state fluorescence measurements were obtained on a FluoroMax-2 fluorometer (Horiba Jobin Yvon) equipped with a 150W xenon lamp and modified Czerny-Turner spectrometers in both the excitation and emission position utilizing Datamax Spectroscopy Software. The instrument was interfaced with a Neslab RTE-111 temperature controller with a remote sensor. The fluorophore-labeled RNAs (40-80 nM) were dissolved in fluorescence binding buffer (20 mM Tris-Cl pH=7.5, 100 mM NaCl, 5 mM MgCl$_2$, 1 mM BME). The RNA solution was mixed in a 1:1 ratio with either blank buffer or proteins dialyzed in the same buffer. The mixture was transferred into a dried 3 mm $\times$ 3 mm path length quartz fluorescence microcuvette (Starna Cells, Inc). Before measurements, all the RNA samples were allowed to incubate with either protein or blank buffer for 10 mins in the dark at 20°C. The RNAs labeled with fluorophore were excited at 485 nm with a spectral slit width of 4 nm. All fluorescence scans were recorded at 20°C using 1 nm wavelength increments. For FRET measurements, the fluorescence intensity was measured at 520 nm by incubating 20 nM fluorophore labeled RNA with 6.5 µM protein. The concentrations of donor- and acceptor-labeled RNA, acceptor-labeled RNA, and donor-labeled RNA were determined by measuring the absorbance at 90 °C. The extinction coefficients were calculated as described by Richards (16).
Figure 3.10. Representative curves for determining the dissociation constants of fluorescein labeled RNA substrates with GST-MBNL1 by steady-state fluorescence measurements.

To obtain a dissociation constant for binding to the fluorophore-labeled RNA, proteins or small molecules were added to the RNA in a series of different final concentrations. A new aliquot RNA was used every time for each measurement to avoid denaturation of the proteins from repeated stirring. Single point fluorescence intensity was obtained at 485 nm excitation wavelength and 520 nm emission wavelength with both excitation and emission spectral slit widths set to 4 nm. The fluorescence intensity was fit to the following equation:

\[
F = \frac{F_{\text{max}} - F_0}{1 + \frac{K_d}{[P]}} + F_0
\]

eq(1)

or
\[ F = \frac{F_{\text{max}} - F_0}{K_d} + F_0 \quad \text{eq(2)} \]

where \( K_d \) is the dissociation constant, \([P]\) is the total protein concentration, \([L]\) is the total small molecule concentration, and \(F_0\) and \(F_{\text{max}}\) are the fluorescence intensity of free and fully bound RNA respectively. The concentrations of the protein and small molecule are approximately equal to total protein and total small molecule concentrations, because their concentrations are more than 50-fold greater than RNA concentration.

### 3.4.4 Stopped-flow Fluorescence Measurements

Stopped-flow fluorescence measurements were performed using a SX-20 Series KinTek spectrophotometer. The fluorescein-labeled RNA was excited at 485 nm and fluorescence emission was monitored through a 495 nm long-pass cut-off filter with the mechanical slit of the monochromater set to 0.5 mm. The path length of the optical cell was set to 2 mm. The fluorescence scans were collected in 5s or 20s data files and each file is consisted of 1000 data points. All the measurements were carried out at 20°C. Fluorescein labeled RNA (40nM) from one syringe was mixed with either blank buffer or protein solution in another syringe. The proteins were dialyzed into the same buffer (20 mM Tris-Cl pH=7.5, 100mM NaCl, 5 mM MgCl\(_2\), 1 mM BME) in which RNA was dissolved. All kinetic measurements were performed under pseudo-first order condition by using protein concentrations at least 100 fold higher than the concentration of labeled RNA. Different protein concentrations ranging from 3.5 µM to 14 µM were used to determine the concentration-dependence of the rate constants. No significant
photobleaching of the fluorescein was observed during the course of the stopped flow experiments. At least five measurements were obtained for each sample to give one data set.

3.4.5 Data Analysis

FRET efficiency was calculated by following formula:

\[
E = 1 - \frac{F_{DA} - F_A}{F_D} \tag{3}
\]

where \(F_{DA}, F_A, \) and \(F_D\) are values for the fluorescence intensity of the donor- and acceptor-labeled RNA, acceptor-labeled RNA, and donor-labeled RNA. In our experiment, the donor and acceptor are fluorescein and TAMRA, respectively. The theoretical fluorescence intensity of fully bound RNA was calculated using following equation:

\[
F = p \times F_{\text{max}} + (1 - p) \times F_0 \tag{4}
\]

where \(F\) is the observed fluorescence intensity, \(p\) is the fraction of RNA bound to the protein which is calculated based on dissociation constant, \(F_0\) and \(F_{\text{max}}\) are fluorescence intensity of free RNA and fully bound RNA, respectively. In the FRET efficiency calculation, the fluorescence intensity of fully bound RNA was used. The distance between the donor and acceptor in the FRET pair was calculated with following equation using the FRET efficiency obtained:

\[
E = \frac{R_0^5}{R_0^5 + R^5} \tag{5}
\]

where \(E\) is the FRET efficiency, \(R_0\) is the distance at energy transfer efficiency of 50% and \(R\) is the distance between the donor and acceptor. \(R_0\) equals to 50 Å for the Fluorescein/TAMRA pair (17).
For the kinetic analysis, a pseudo-first order model was used because more than 100 fold molar excess of protein was mixed with the RNA. The apparent dissociation constant was obtained by fitting the fluorescence intensity to following equation:

\[ F = (F_{\text{max}} - F_0) \times e^{-kt} + F_0 \]  

where \( F \) is the fluorescence intensity, \( F_{\text{max}} \) is the fluorescence intensity upon saturation, \( F_0 \) is the initial fluorescence intensity, \( k \) is the apparent rate constant, \( t \) is the time, and \( e \) is the base of the natural logarithm. The observed rate constant is the sum of the association constant and dissociation constant because the binding reaction is reversible. The association and dissociation constants were calculated using following equations:

\[ K_d = \frac{k_{\text{off}}}{k_{\text{on}}} \]  
\[ k = k_{\text{on}}[P] + k_{\text{off}} \]

where \([P]\) is the protein concentration, \(K_d\) is the equilibrium dissociation constant, \(k\) is the observed rate constant and \(k_{\text{on}}\) and \(k_{\text{off}}\) are apparent association and dissociation rate constants, respectively.

3.5 References


CHAPTER 4

Kinetic Analysis of the Interaction between MBNL1 and CUG Repeat RNA

4.1 Introduction

The interaction between MBNL1 and CUG repeat RNA is the key step in the pathogenesis of myotonic dystrophy (1-2). For treatment of myotonic dystrophy, the MBNL1-CUG repeat RNA complex appears to be an ideal target and studies are required to understand the molecular basis of MBNL1-RNA interactions. In the cell, timing of association and dissociation of MBNL1 with RNA targets can affect the localization of the protein and the ability of inhibitors to destabilize the complex. Therefore, the formation of the MBNL1-CUG repeat RNA complex should be studied as a dynamic process. Herein, I will present my efforts in studying dynamic processes involved in the association of MBNL1-CUG repeat RNA complexes using short CUG repeat RNA mimics by both stopped-flow fluorescence spectroscopy and surface plasmon resonance methods. The data presented reveal presence of multiple phases and suggest negative cooperativity of MBNL1-CUG repeat RNA interactions.

4.2 Results

4.2.1 Binding Occurs in Multiple Kinetic Phases

To investigate the kinetics of MBNL1-CUG repeat RNA interactions, the oligonucleotide (CUG)_{12} was selected as a mimic of disease-causing RNA. Based on previously reported crystal structures (3-4), the number of repeats we used here is sufficient to maintain the structural features of poly(CUG) RNA and reflects the kinetic process of the protein-RNA interaction. For
the convenience of fluorescence analysis, fluorescein was attached to the 3’ end of the oligonucleotide. The apparent dissociation constant of (CUG)$_{12}$ RNA with MBNL1 protein was determined to be $150 \pm 20$ nM by electrophoresis mobility shift assay and to be $9.6 \pm 3.0 \mu$M by steady state fluorescence analysis. The discrepancy between the dissociation constants determined by the two methods is likely due to the presence of multiple binding sites, which will be discussed in section 4.3. Stopped-flow fluorescence experiments were performed by mixing the GST-MBNL1 solution with labeled RNA solution. The fluorescence intensity change of the labeled RNA was recorded over a 20 s or 200 s time frame to probe the kinetic processes of the interaction on different time scales.
Figure 4.1 Representative stopped-flow kinetic curves for MBNL1-(CUG)$_{12}$ interactions. The time scale and concentration of the protein used are labeled on top of each curve.

Three phases were observed during the course of MBNL1-(CUG)$_{12}$ binding. (Table 4.1 and Figure 4.1) The relaxation times of the three phases were 3.7s, 17.4s and 217s respectively.
for sample with 10 μM MBNL1. The relaxation times are negatively correlated but not proportional to the protein concentration. The relationship between the relaxation times and protein concentration can not be unambiguously determined due to the magnitude of experimental error. At low protein concentrations (≤ 2.5 μM), data fitting with triple exponential equations did not converge properly and only two phases were observed. This is probably because the fluorescence signal change intensity was too small for the third phase when protein concentration was low.

Table 4.1 Relaxation times determined by fluorescence stopped-flow analysis

<table>
<thead>
<tr>
<th>Protein Concentration (μM)</th>
<th>r1 (s)</th>
<th>r2 (s)</th>
<th>r3 (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>3.7 ± 1.2</td>
<td>17.4 ± 1.9</td>
<td>217 ± 19</td>
</tr>
<tr>
<td>5</td>
<td>4.3 ± 1.3</td>
<td>20.4 ± 2.0</td>
<td>241 ± 19</td>
</tr>
<tr>
<td>2.5</td>
<td>9.8 ± 1.9</td>
<td>92 ± 24</td>
<td>N.D.</td>
</tr>
</tbody>
</table>

*N.D. stands for not determined.

4.2.2 Analysis of the MBNL1-CUG Repeat RNA Interactions by Surface Plasmon Resonance

To further investigate the interaction between MBNL1 and CUG repeat RNA, we performed kinetic analysis using surface plasmon resonance (SPR) biosensors. In all the SPR experiments discussed below, ligand refers to the component that was first immobilized on the sensor chip and analyte refers to the other binding partner that will be captured on the sensor chip during kinetic analysis. As an alternative method for kinetic analysis, surface plasmon resonance served to verify our findings with fluorescence stopped-flow experiments. An advantage of SPR over fluorescence measurements in determining the RNA binding kinetics is that the resonance unit change in SPR is largely due to the immobilization of analyte on
biosensor chip surface. Usually, the intrinsic conformational change of RNA or protein will only lead to a much smaller signal change compared with that resulting from analyte immobilization, if the molecular weight of the analyte is large enough (5). Therefore, if proteins are used as the analyte, the curves obtained by SPR will solely reflect the kinetic of binding rather than the conformational change of RNA. In addition, SPR requires very little ligand, which allows us to study the kinetic interaction between long CUG repeat RNA and MBNL1, because long RNA transcripts are technically difficult to generate in large quantities.

With a Biacore system for SPR, we first obtained the kinetic curves using (CUG)_{12} as ligand and MBNL1 as analyte (Figure 4.2 B). Based on the curves obtained, the binding occurs in at least two phases. It is worth mentioning that the resolution of the Biacore system has prevented us from accurately identifying any fast phase with relaxation time less than 4s. Nevertheless, the relaxation times determined for the other phases are similar to those determined in stopped-flow fluorescence measurements (Table 4.2). In addition to association curves, dissociation curves were also recorded by the Biacore system. However, MBNL1 protein does not fully dissociate from the chip surface immobilized with RNA and the dissociation curves are unreliable for quantitative analysis. In addition to (CUG)_{12} RNA, we also used (CUG)_{74} RNA for kinetic measurements. The relaxation times show no significant change with changing the length of CUG repeat RNA.
Table 4.2 Comparison between relaxation times obtained by stopped-flow fluorescence measurements and SPR for MBNL1-(CUG)$_{12}$ interactions.$^a$

<table>
<thead>
<tr>
<th>Experimental Methods</th>
<th>r1 (s)</th>
<th>r2 (s)</th>
<th>r3 (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stopped-flow fluorescence</td>
<td>3.7 ± 1.2</td>
<td>17.4 ± 1.9</td>
<td>217 ± 19</td>
</tr>
<tr>
<td>SPR ((CUG)$_{12}$ as ligand)</td>
<td>N.D.</td>
<td>9.6 ± 4.9</td>
<td>265 ± 30</td>
</tr>
<tr>
<td>SPR (MBNL as ligand)</td>
<td>N.D.</td>
<td>18 ± 6</td>
<td>-</td>
</tr>
</tbody>
</table>

a. The analyte concentrations are all 10µM; 
b. N.D. stands for not determined.

Figure 4.2 Association kinetics for MBNL1 binding to (CUG)$_{12}$ RNA. A MBNL1 as ligand; B. (CUG)$_{12}$ as ligand.

Taking the SPR and stopped-flow fluorescence data together, we speculate that MBNL1 does not bind to CUG repeat RNA monovalently and binding events are not independent to each other. The presence of multiple phases is due to the multiple binding of MBNL1 to CUG repeat RNA target. To verify this speculation, we switched the analyte and ligand in the SPR experiment, making MBNL1 the ligand and (CUG)$_{12}$ the analyte (Figure 4.2 A). Only one phase was observed after switching the analyte and ligand. The relaxation time is determined to be 18 s, which is shorter than the second relaxation time observed in previous SPR experiments. This result is consistent with our expectations because the data were collected under pseudo-first
order conditions with RNA in large excess and the second phase should not be present, if binding is in a one-to-one stoichiometry.

<table>
<thead>
<tr>
<th></th>
<th>τ2</th>
<th>τ3</th>
</tr>
</thead>
<tbody>
<tr>
<td>CUG_{12}(SPR)</td>
<td>9.6 ± 4.9 s</td>
<td>265 ± 30 s</td>
</tr>
<tr>
<td>CUG_{74}(SPR)</td>
<td>13 ± 2 s</td>
<td>310 ± 27 s</td>
</tr>
</tbody>
</table>

**4.3 Discussion and Conclusions**

The data presented above suggest that the number of kinetic phases observed during the course of MBNL1-CUG repeat RNA interactions is dependent on the stoichiometry of binding. On the one hand, there are multiple kinetic phases under the pseudo-first order conditions in the presence of excess MBNL1 protein. On the other hand, only one phase can be observed in presence of excess CUG repeat RNA. Although the binding stoichiometry of MBNL1 with CUG repeat RNA is unknown, previous studies suggest that the 4-nt sequence 5’YGCY3’ is the minimal MBNL1 binding site (6), and MBNL1 can bind tightly to an RNA construct with only four CUG repeats (7). Our kinetic data indicate that at least two MBNL1 protein molecules interact with one single (CUG)_{12} RNA construct. Consistent with this proposal, single molecule fluorescence studies carried out by Spies and coworkers (unpublished work) at the University of Illinois, Urbana-Champaign show multiple FRET states when labeled fluorescent MBNL1 is incubated with fluorescent (CUG)_{12} RNA. Unfortunately, our efforts in determining the maximal stoichiometry of the protein-RNA complex at saturation were not successful due to the solubility limits of the MBNL1 protein. The difficulty in preparing protein in high concentration also
prevented an accurate determination of the overall dissociation constant by surface plasmon resonance or fluorescence stopped-flow analysis.

It is worth mentioning that the dissociation constant determined by the steady state fluorescence analysis ($K_d=9.6 \pm 3.0 \, \mu M$) is much greater than that determined by the electrophoresis mobility shift assay ($K_d=150 \pm 20 \, nM$). We speculate that the $K_d$ obtained in gel shift assay might only describe the first binding event in the equilibrium, as no streaks or multiple bands have been observed. By comparison, the $K_d$ obtained from fluorescence measurements reveals the overall apparent dissociation constant, in which one RNA molecule is treated as multiple independent binding sites.
The presence of multiple kinetic phases not only suggests the presence of multiple binding events but also indicates that these binding events do not share the same association rates. Cooperativity can be seen when MBNL1 interacts with CUG-repeat RNA. To illustrate the cooperativity, we normalized the bound RNA concentration at equilibrium to 1 and plotted \[
\frac{d[R]}{d[R]} \]
against time, where [R] is the free RNA concentration. As can be seen in Figure 4.3, the binding shows negative cooperativity, because the apparent rate constant decreases as time increases. In other words, the association rate constant is becoming smaller and smaller when
more and more binding sites are occupied. The negative cooperativity can be also illustrated in another way by plotting \( \frac{d[R]}{dt} \) against (1-[R]). The slope of the curve is becoming smaller as the fraction bound increases (Figure 4.4). One possible explanation for the negative cooperativity is that binding of one protein molecule could sterically hinder the binding of the other, making the overall rate constant smaller. Furthermore, the conformational change of CUG repeat RNA at one binding site may somehow affect the conformation of its nearby binding sites, leading to a change in the activation energy of the association reaction.

![Graph showing the association rate decreases as the fraction of bound RNA increases.](image)

**Figure 4.4** The association rate decreases as the fraction of bound RNA increases.

It is interesting to note that changing the length of the CUG repeat does not affect the binding kinetic significantly. This might be attributed to the distance between binding sites on the CUG repeat RNA. If the two binding sites are far away from each other along the RNA sequence, binding at one site would have little steric effect on binding at the other site. Only the nearby binding sites would affect the activation energy of each other. Therefore, the shape of the
kinetic curve is independent of the length of the CUG repeat RNA as long as it is sufficiently long.

In summary, based on the data presented above, CUG repeat RNA can interact with multiple MBNL1 proteins. The binding of MBNL1 protein shows negative cooperativity. Furthermore, the kinetics of MBNL1-CUG repeat RNA interactions are not repeat-length dependent. Our study provides insight into the kinetics of MBNL1-CUG repeat RNA interactions. However, the interaction between MBNL1 and CUG repeat RNA could be more complicated under physiological conditions. In the future, it would be helpful to characterize the kinetic process of formation of CUG repeat RNA nuclei foci within living cells.

4.4 Materials and Methods

4.4.1 Plasmid

The vector pSP72-CTG74 was obtained from Maurice S. Swanson (8) (University of Florida, College of Medicine, Gainesville, FL, USA). The vector is a T7 promoter based vector for RNA transcription with a 74 CTG repeat expansion followed by a BamHI restriction site.

4.4.2 Protein Expression, Electrophoresis Mobility Shift Assay and Stopped-Flow Fluorescence Analysis


Figure 4.5 A comparison of single and triple exponential fits shows the lack-of-fit of the kinetic curve obtained by stopped-flow fluorescence analysis fit by the single exponential equation.

Protein expression and stopped-flow fluorescence analysis followed the same protocol as described in Chapter 2.4.4, Chapter 2.4.6 and Chapter 3.4.4, respectively. For data collection in stopped-flow fluorescence analysis, 1000 data points were recorded in each time frame. To reduce systematic bias in global fitting, all data points have equal weights in data analysis no matter in what time frame they were collected. The stopped fluorescence data were fit to a triple exponential equation instead of the single exponential equation described in Chapter 3.4.5. It is
worth noting that the protein concentration reported here is determined by Bradford protein assay. Therefore, it does not necessarily reflect the true active protein concentration in solution.

4.4.3 *In vitro* Transcription of CUG Repeat RNA

The DNA template for *in vitro* transcription of RNA was prepared by incubating the pSP72-CTG74 vector with BamHI (Invitrogen) for 6 hours at 37°C. The subsequent linear DNA template was treated with Proteinase K and gel purified to remove the uncleaved DNA and possible contamination of RNase. Transcription reactions were carried out in a 20 µl volume which contained 2 µg of DNA template, 7.5 mM rNTPs, 40 Unit of RNasin (Promega), 200 U of T7 RNA polymerase (Ambion), 10 mM DTT, 40 mM Tris-HCl (pH 7.9), 6 mM MgCl₂, 2 mM spermidine and 10 mM NaCl. The reaction was performed at 37°C overnight. The transcript was purified by phenol extraction and ethanol precipitation, followed by gel purification on a 10% denaturing polyacrylamide gel. (Figure 4.6)

![Figure 4.6](image_url)  
*Figure 4.6* Representative denaturing acrylamide gel for CUG₇₄ RNA transcript. Lane 1 shows the RNA Century Marker (Invitrogen) with number of nucleotide of each band labeled on right
on the lane. Lane 2 shows the RNA transcript. The theoretical size of the RNA transcript is 260(nt.

4.4.4 Biotinylation of RNA Transcript

The biotinylation of the RNA transcript was carried out using T4 RNA ligase to attach a single biotin modified nucleotide to the 3’ end of the RNA transcript. The labeling reaction was performed following the protocol provided with the RNA 3’ End Biotinylation Kit (Pierce). (Figure 4.7 A) The subsequent biotinylated RNA was isolated from the modified nucleotide by chloroform/phenol extraction and ethanol precipitation. Dot blotting was performed to determine the labeling efficiency. In dot blotting, diluted RNA samples were spotted on positively charged nylon membrane and immobilized covalently by UV cross-linking. After blocking, the nylon membrane was then allowed to incubate with streptavidin-horseradish peroxidase conjugate and washed 4 times with PBST buffer. Blots were treated with the indicated Western HRP Chemiluminescent Substrate (Pierce) and exposed to X-ray for 30 s for visualization. Synthetic biotin modified RNA was used to plot a standard curve for quantification the biotinylation percentage of CUG repeat RNA transcript. (Figure 4.7 B)
4.4.5 Surface Plasmon Resonance

Surface plasmon resonance experiments were performed on BIACORE 3000 biosensors (Biacore, Inc., Piscataway, NJ). The biotin modified RNA targets were either chemically synthesized (Integrated DNA Technology) or prepared following the protocol described above. When RNA was used as ligand, the RNA in Tris buffer (20 mM Tris-Cl pH=7.5, 100mM NaCl, 5 mM MgCl₂, 1 mM BME) was first heated up to 90 °C and cooled to room temperature to allow
annealing of the stem. The diluted RNA solution was injected at a flow rate of 10 μl/min over the streptavidin-coated sensor chip surface. Approximately 20 RU of RNA was captured on the chip surface through biotin-streptavidin interactions. When MBNL1 was used as ligand, the protein (~5 μM) was at a flow rate of 10 μl/min and captured on a chip surface with nitrilotriacetic acid (NTA) which is designed to bind to histidine-tagged protein molecules. The binding of the histidines relies on a NTA-chelated nickel atom, so Ni^{2+} (~1 mM) was allowed to be captured on the surface prior to immobilization of protein. Around 800 RU of MBNL1 was captured on the chip surface. The immobilization level of protein is much greater than that of RNA, because the signal change in kinetic analysis with RNA as analyte is very small.

For all the SPR experiments, Tris buffer (20 mM Tris-Cl pH=7.5, 100mM NaCl, 5 mM MgCl\textsubscript{2} and 1 mM BME) was used as the running buffer. All binding experiments were carried out at 20 °C and a flow rate of 40 μl/min. No mass transport effect was observed during the course of analysis. A combination of NaOH (50mM) and SDS (0.5%) solution was used for regeneration of the sensor chip SA surface to remove the remaining protein after the dissociation phase. EDTA (350mM) solution was used to remove both protein and RNA from the NTA sensor chip. Therefore, MBNL1 had to be immobilized on the NTA sensor chip before every measurement, but the RNA only needed to be immobilized once on an SA sensor chip. For data analysis, global fitting was carried out with following exponential equations using OriginLab 8.0:

\[
\text{Resonance Unit} = A_1 e^{-\frac{t}{r_1}} + A_2 e^{-\frac{t}{r_2}} + A_3 e^{-\frac{t}{r_3}}
\]
where $\tau_1$, $\tau_2$ and $\tau_3$ are relaxation times.

### 4.5 References


