2017

Ice Buckets to Proteins: Investigating the Role of matrin 3 in ALS

Elizabeth DiLoreto
Assumption College

Ice Buckets to Proteins; Investigating the Role of matrin 3 in ALS

Regarding the Potential Presence of an Intermediate State in the Folding Free-Energy Landscape

Elizabeth DiLoreto

. . .

Faculty Supervisors:

Dr. Michele Lemons, Assumption College

Dr. Jill Zitzewitz, University of Massachusetts Medical School

. . .

Natural Sciences

A Thesis Submitted to Fulfill the Requirements of the Honors Program at Assumption College

April 2017

TABLE OF CONTENTS

ABSTRACT

Amyotrophic Lateral Sclerosis, ALS, is a neurodegenerative disease characterized by the dysfunction of motor neurons in the spinal column, which prevents muscle movements and eventually, breathing. The prognosis is death typically within 2 to 5 years with only one drug available for treatment, Riluzole. This drug can only help select ALS patients as ALS is associated with many different protein mutations. One protein frequently found in ALS patient samples is TDP-43. These samples are from stress granules that form when there is toxicity in the cell. TDP-43 has been studied in an isolated context of its RNA-Recognition Motifs, RRM. These RRM domains are approximately 90 amino acids long, typically containing 2 alpha helices and 3 beta pleated sheets. Using this information, the secondary structure of the protein can be examined. The RRMs are highly specific regions that only bind to a certain DNA sequence, RNA sequence, or protein. Having an ordered region such as the RRM, allows for probing of the folding free-energy landscape, or the conformations that the protein takes as it is denatured to an unfolded state. This landscape is studied in hopes of locating an intermediate state; a state where either the secondary or tertiary structure is maintained, even though the protein continues to unfold. This intermediate conformation would be targeted for therapeutic development with the hope of forcing proteins in this state back to a native and functioning conformation. This would decrease the amount of misfolded protein and hopefully slow stress granule formation. RNA-binding ALS-linked protein matrin 3 is typically found alongside TDP-43 in stress granules of ALS patients. MATR3 gene mutations have been associated with RNA mismanagement. Repeating the isolated experiments performed with TDP-43, matrin 3 was examined for the presence of an intermediate state. Like TDP-43, matrin 3 has 2 RRMs and a tethered complex of the two RRMs. Unlike TDP-43, an intermediate state was not found. This brings to question the biological purpose of intermediate state conformations, which is investigated here.

INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is a dysfunction of motor neurons in the spinal column that prevents muscle movements, causing trouble swallowing, speaking, and eventually breathing. Currently, there is no known cure, and the progressive nature of the disease leads to fatality, typically within 2 to 5 years of diagnosis (ALS Association, 2016).

Approximately 10% of ALS cases have been linked to genetic factors (e.g. familial cases) while over 90% are sporadically occurring without a genetic link or family history. This lack of a genetic factor is concerning to researchers and the general population because ALS can strike at any time before age 60, with fatality guaranteed to follow (Xu and Yang, 2014). Even with genetic testing, the lack of biomarkers makes it difficult to determine the presence of ALS until it is too late for treatment to be effective. The diagnosis of ALS typically comes once all other neurodegenerative causes are ruled out (ALS Association, 2016). There is only one drug available for treating ALS. Riluzole works by inhibiting glutamate receptors of damaged neurons. Generally, the drug only allows for a 2 to 3 month extension of life; a substantial but still very limited amount of time for a patient that is only given 2 years to live (Miller et al., 2012).

The presented research focuses on matrin 3, an RNA-binding protein, encoded from the gene MATR3. The matrin 3 protein is a member of a family of RNA-binding proteins and has been associated with ALS by its abundance in protein aggregates of ALS patients post-mortem. matrin 3 is typically found in conjunction with TDP-43, trans-activation response (TAR) deoxyribonucleic acid (DNA) Binding Protein 43 (Wang, 2015; Blatter et al., 2015). TDP-43 has been studied in a purified, isolated state to examine the folding free-energy landscape of the protein as a denaturing agent is added to the TDP-43 solution. Examining secondary and tertiary

structure of TDP-43 helped uncover the existence of an intermediate state in the second RiboNucleic Acid Recognition Motif (RRM) between the native and unfolded conformations (Mackness et al., 2014).

Within the class of RNA-binding ALS-linked proteins, matrin 3 was examined for the presence of an intermediate state along its folding free-energy landscape. The ultimate goals of this research are: 1) to discover an intermediate conformation state of a protein that is found aggregated in patient samples and 2) to develop a therapeutic that could force protein in this state back to a native conformation.

History of ALS Research

ALS is known as an orphan disease. Orphan diseases afflict less than 1 in 1600 people, making it a rare disease. Often, research in these fields is limited, slowing development of treatments. However, most diseases have been linked to some sort of genetic mutation or a genetic biomarker that has been developed (107th Congress, 2002). Even though ALS was identified nearly 200 years ago, it is still difficult to diagnose. In most cases there is no genetic marker, which results in a late diagnosis of ALS. Sporadic ALS (SALS) accounts for 90% of ALS cases versus 10% of Familial cases (FALS) (Xu and Yang, 2014; Wu et al., 2012). Researchers identified many proteins in patient biopsies and began to study these proteins to find a link to ALS. To replicate conditions and phenotypes of disease, different model organisms were designed. Although many research techniques are used to study ALS, it has proven difficult to identify a marker that indicates ALS before the onset of disease. This introduction will expand upon historic findings of ALS, protein structure and function, and what proteins have been identified in association with this disease.

The first case of ALS was reported by Charles Bell in France in 1824. At the time, there was no name for the disease because little was known about it, as it was only identified as a motor disorder. After additional investigation, Jean-Martin Charcot had named the disease by 1874. Identifying the disease that was first reported 50 years prior proved difficult because many of the phenotypes, such as muscle weakness and degeneration (atrophy), were attributed to other degenerative diseases. Charcot was able to correctly create a timeline of disease progression through patient observation (Rowland, 2001).

Effects begin with immobilization but not loss of sensation in the upper limbs, controlled by lower motor neurons, over the course of several months. The atrophy then spreads to the lower limbs, controlled by upper motor neurons. Eventually walking is impossible because the muscles can no longer function without fibrillary tremors. The last step in the timeline is bulbar symptoms, localized in the mouth. For some patients this is the location of disease onset. Paralysis begins in the tongue before moving to the uvula, lips and finally the vagus cranial nerve. This paralysis initiates with trouble swallowing, pronunciation, closing the mouth, and, if the vagus nerve is paralyzed, trouble breathing which causes imminent death (Rowland, 2001). Even with limitations of technology, Charcot was able to identify the origin of the disease (lateral spine) and the related phenotypes (muscle malnourishment by hardening or deadening of neurons), to create an accurate name. According to the Greek origin, it literally means no muscle nourishment (ALS Association, 2016).

Throughout the next 60 years, little progress was made in identifying the cause of the disease. In 1939, all-star Yankees player Lou Gehrig was diagnosed with the disease, eventually killing him two years later. This celebrity diagnosis spread the name of ALS; the two were so intertwined that ALS is sometimes called Lou Gehrig's disease (RVW Foundation, 2015). There

was no prescription to help stave off the onset of the disease and there were no indicators developed to provide an early diagnosis. The only course of treatment was light exercise to treat unaffected muscles (NINDS, 2016). Gehrig was unable to benefit from this protocol because the disease had already taken hold of his body and his physical condition continued to diminish rapidly after his retirement (RVW Foundation, 2015). Stephen Hawking was also diagnosed with ALS back in 1963 with some difficulty determining the disease because of symptoms similar to other neurodegenerative disease. He is an atypical patient with ALS because he is still fighting 54 years after diagnosis (Weebly, 2014). His survival of 10 to 27 times the expected number of years cannot be explained because the disease itself has yet to be explained.

Moderate amounts of research on ALS continued, but few significant strides were made until the turn of the century. In 1983, the Orphan Drug Act was passed to increase research on drug development for orphan diseases, but could not be sustained for more than a decade. The Rare Disease Act of 2002 created more research opportunities for orphan diseases because funds were established for those fields of research. It defined a rare disease as one actively affecting fewer than 200,000 Americans. The Congress who passed the act addressed the fact that large pharmaceutical companies were not funding research for orphan disease drug development because they would not have a large audience to market. This Act of 2002 established a new wave of research and funding (107th Congress, 2002).

Currently, there is only one drug available for treating ALS, Riluzole. It was first identified as an ALS treatment in 2003. This drug prolongs life for a mere 2 months during the course of the 2-5 year terminal prognosis. It may also increase the quality of life as it staves off the use of tracheostomy and mechanical ventilation. Riluzole (Rilutek) acts by inhibiting glutamate receptors in the neurons. These receptors were reportedly overexcited in cases of ALS,

which caused early neuron death. The use of Riluzole caused researchers to investigate the glutamate receptor in relation to ALS (Miller et al., 2012).

No other drug has been used to treat ALS because little is known about what factors are implicated in the disease and what can be done to slow disease progression. Recently, ALS has become better known from the ALS Ice Bucket Challenge, which started in the summer of 2014. The movement of dousing participants in ice water has raised millions for the ALS Association as well as other foundations, benefitting patient support, outreach and treatment. A portion of money raised by the ALS Ice Bucket Challenge goes to funding drug development.

A Protein's Structure Drives its Functionality

The role of proteins in survival becomes clear as they are needed to carry out many functional duties in the cell. Not being able to function properly is linked to disease. Proteins often play a role in disease propagation because of their relation to DNA, the genetic code of mammalian systems. DNA is the genetic code because it is the starting point of development. DNA transcribes RNA and RNA regulates the translation of a specific protein, or group of proteins. This statement relating DNA to RNA to protein is what is known as the central dogma in biology. Due to this relationship, proteins may be implicated in disease if there is a coding error in the DNA it comes from.

Proteins contain four different types of structure. Primary structure is the order in which the amino acids are connected. In humans, there are 20 amino acids, the building blocks of proteins, that can be in any order or length to give the sequence of the protein. Secondary structure comes from the bonds that hydrogens make throughout the local amino acid sequence; this can take the form of beta pleated sheets (β) or alpha helices (α). Tertiary structure occurs

when other bonds are formed, such as sulfur-sulfur bonds, causing three-dimensional interactions in the protein. Typically, proteins also contain quaternary structure, the interactions of multiple amino acid sequences, allowing the formation of larger protein complexes. Whatever the sequence, many proteins need to fold to function, and may contain some or all of the types of structure listed above.

Proteins can produce alterations in structure as different conformations- or shapes- are formed. This is akin to a slinky stretching and compressing. The movement of the protein allows it to adapt to different conditions such as heat, pH levels, and salt concentrations. After RNA translates a protein, the protein must fold into the shape it has been coded for. When a slinky is first made, it is fully stretched and changes shape to suit the purpose it was designed for. These conformational changes can be essential for protein function and it helps to bring the protein to a stable conformation, promoting better functioning of the protein. The most stable state of a slinky is when it is closed. Proteins have different amounts of energy stored in their conformation depending upon the amount of energy required to make and maintain their bonds and structure. A stretched slinky contains a lot of energy, as does an unfolded protein. The different conformations of proteins can be plotted in an energy landscape; a depiction of how energetically stable the different conformations of proteins are. The more favorable conformations of proteins exist at lower energy.

Some proteins have an intermediate state; a partially folded shape that a protein can hold while another segment of the protein folds. The number of intermediates per protein depends upon its folding pathway and the function the intermediate may serve within the cell. These intermediates are hypothesized to aid in overall protein folding because they exist in the middle of the energy landscape. The folding free-energy pathway is visual representation of the different

conformations a protein can take and the amount of free energy stored in that conformation mapped among the other conformations and the amount of energy needed to get to them. Often it looks like an actual landscape with local maxima and minima as peaks and valleys in the two or even three dimensional map (*Figure 1*).

An imbalance in homeostatic conditions may cause the protein to misfold, thereby taking on a non-native conformation. Misfolding can bring the protein beyond the lowest energetic state on the native energy landscape. This state is so stable that it cannot refold to gain functionality in its native state, which occurs in cases of protein aggregation. If a slinky is misfolded, its structure is damaged so it cannot return to its normal function. Normal protein structures exist around the native conformation as the local minima, while misfolding states surround other minima. Often the free energy stored in the aggregated conformations is much lower that of the native conformations so the protein cannot overcome the barrier to return to a native state. This lack of functionality makes the state immediately before the point of no return very important; it is the precipice before a long fall. Researchers examine the importance and propagation of the intermediate in the folding and misfolding or function and dysfunction of a protein (Cooper, 2000).

The RRM domains are approximately 90 amino acids long, typically containing β-1, α-1, β-2, β-3, α-2, β-4 secondary structure. They are highly specific regions, meaning that they will only bind to a certain DNA sequence, RNA sequence, or protein (Afroz, 2015). Binding to an RNA sequence allows the protein to perform different functions within the cell.

Research of ALS-linked Proteins

In 1993, researchers began looking at potential molecules that contribute to ALS. Cu-Zn Superoxide Dismutase (SOD1) was the first enzyme identified in cases of ALS during this

movement (ALS Association, 2016). Mutated SOD1 was reported in 20% of FALS cases with 150 different cases of mutations in the course of its 153 amino acid sequence. These mutant enzymes would misfold (changing from their normal, native state conformation) and be ubiquitinated. Ubiquitin is a molecule that tags a protein for transportation into protease so the protein can be degraded. In cases of ALS, this ubiquitinated protein builds up, potentially causing the lack of neuronal functionality because excess protein is unable to be broken down (Miller et al., 2012). The genetic link of FALS and SOD1 launched investigation into SOD1 and why it builds up in diseased cells.

Trans-activation response [TAR] deoxyribonucleic acid [DNA] binding protein 43 (TDP-43) was not implicated in ALS until 2006. This protein contains 414 amino acids and 2 RRM domains, a nuclear exportation sequence (NES), a nuclear localization sequence (NLS), and a glycine rich domain (Gly-rich) (Mackenzie et al., 2010) (*Figure 2 A,D*). Like SOD1, ubiquitinated TDP-43 was found to be built up (aggregated) in cases of ALS (Wu et al., 2012). This discovery was incredibly significant because TDP-43 was reported in 95% of SALS cases (Xu and Yang, 2014). This protein linked ALS to Frontotemporal Lobular Degeneration (FTLD), establishing a correlation between the two diseases. Some patients showed pure ALS character (motor degeneration) while others showed pure FTLD character (cognitive degeneration), however the majority shared some combination of the two (Robberechet and Philips, 2013).

Inappropriate levels of TDP-43 has been found to cause motor dysfunction in mammalian models (Stoica et al., 2014). These levels of TDP-43 are influenced by the amount of misfolded protein in aggregates, puncta, or stress granules. The misfolding of TDP-43 in ALS has been associated with mutations in the largely disordered glycine rich region. Studying the secondary and tertiary structures of isolated TDP-43 RRM enables investigation into the biophysics of

protein folding. Only a select region of the full-length protein was examined to discover if there was a folding intermediate, which may play an important role in disease as it may toggle between the two energy landscapes of folded and misfolded aggregation. The structured RRM have a stable native conformation with secondary structure that is distinct from its unfolded state, which lacks all but primary structure. An intermediate state has been identified in RRM2 or TDP-43 (Mackness et al., 2014). The role of the intermediate in RRM2 is under investigation because it alters the shape of the protein, potentially aiding in the folding process (Mackness, 2016).

Fused in sarcoma (FUS) is another protein that is highly investigated in cases of ALS although it is typically not found in patients that contain ubiquitinated clusters of TDP-43 (Mackenzie et al., 2010). FUS is 526 amino acids long. However, for such a lengthy protein only a few mutants have been found to relate to ALS. Similar to TDP-43, it contains an NLS, NES, and a Gly-rich region. It varies in that it has only one RRM and contains two arginine-glycine-glycine repeating regions (RGG-Rich), a zinc finger (ZnF), and a glutamine-glycine-serine-tyrosine rich region (QGSY-Rich) (*Figure 2 B,E*). Unlike SOD1 and TDP-43, ubiquitination was not found in protein aggregates of FUS, neither was any other modification to the protein (Neumann, et al., 2006; Zhou et al., 2014). This difference is interesting because there was no mutation or conformational change to the protein, making it unclear what mechanism caused protein accumulation or disease propagation. Creating a biomarker to identify when the disease is propagating would be helpful in identifying SALS cases because they are the majority of all ALS cases.

Research Involving matrin 3

Like FUS and TDP-43, matrin 3 is an ALS-linked RNA-binding protein. Compared to the other proteins, matrin 3 has largely been uninvestigated until recently. It is composed of 847 amino acids and while it is longer than the other proteins previously mentioned, it also has more ordered domains; 2 RRMS (mRRM1, mRRM2, and a tethered depiction of the two in mtRRM), 2 ZnF's, an NES and an NLS (*Figure 2 C,F*). matrin 3 has 4 mutations associated with ALS which have been identified through studies of stress granules and droplet formation in patient samples (Taylor et al., 2016). TDP-43 has been associated with some stress granules which contain matrin 3 (Wang, 2015). Even with this association, the majority of matrin 3 ALS cases are found in familial ALS, while TDP-43 mutations are associated with more sporadic cases of ALS.

MATR3 gene mutations have been associated with RNA mismanagement. This interaction with RNA has been localized to the interaction of mRRM2, which is also involved in mRNA stabilization (Salton et al., 2011). The RNA binding affinity of this protein will also be discussed because, although it is in the family of RNA-binding proteins, its main function seems to be as a scafolding protein because of its zinc fingers. These features also allow the protein to bring together mRNA and DNA for interaction. The RNA binding will be examined by an Electromobility Gel Shift Assay (EMSA).

The presence of different missense mutations result in different disease phenotypes. For example a serine 85 to cysteine mutation results in slow disease progression while a phenylalanine 115 to cysteine mutation results in respiratory failure and fatality within 5 years (Johnson et al., 2014). This difference in disease progression was also seen when a mouse model using human wildtype matrin 3 resulted in hind limb and forelimb muscle atrophy, and human

symptoms with MATR3 mutation resulted in autosomal dominant distal myopathy (vocal cord and pharyngeal weakness) (Moloney, 2016).

The effects of matrin 3 can also be considered globally. First, in the sense of the whole body. Mutations in matrin 3 have been shown to not change its localization, meaning that it continues to stay in the nucleus while mutated (Gallego-Iradi et al., 2015). This localization is relative as matrin 3 has been examined as it acts as a nuclear protein that can translocate to the cytoplasm. The microtubule system by which matrin 3 travels allows its movement to become a tool in its ability to act a tumor suppressor. In breast cancer samples, lower quantities of matrin 3 were found in patient samples while the expression of matrin 3 resulted in lower tumor expression (Subbarayalu et al., 2015). In a truly international sense, the localization of certain ALS mutations can be considered around the world. The ALSOD data base allows this reporting and tracking of protein mutations across countries (Institute, 2007). Using this and other resources, it was identified that a high percentage of French patients with matrin 3 mutations had familial ALS (Millecamps et al., 2014). matrin 3 mutants found in Italian ALS patients occurred in both familial and sporadic cases. This more recent study identified 5 missense mutations; Q66K, G153C, E664A, S707L, and N787S (Marangi et al., 2016).

Considering the established base of knowledge concerning matrin 3 and other RNA-binding ALS-linked proteins, matrin 3 was examined in its RRMs. This was done as it was for TDP-43 because of the innate order of these regions (Mackness et al., 2014). Since the goal is to identify if there is an intermediate state, there needs to be a clear transition from folded to unfolded. The RRM regions typically contain 2 alpha helices and 3 beta pleated sheets. From this the secondary structure of the protein can be examined. This can be performed by Circular

Dichroism (CD), which examines the absorption of polarized light from two directions as light passes through a sample (Greenfield, 2009).

To compliment the information gathered on secondary structure, the tertiary structure of a protein can be identified by the location of the aromatic groups, tryptophan and tyrosine. The segment mRRM1 contains 4 tyrosines while mRRM2 contains 4 tyrosines and 1 tryptophan (*Figure 3*). Therefore, mtRRM contains 8 tyrosines and a tryptophan. The method used to study tertiary structure is Fluorescence Spectroscopy (FL). Similar to CD, FL uses light to determine the location of protein in a liquid sample. The difference arises in the wavelength of light used. FL intentionally uses a wavelength of light that will excite the aromatic amino acids present in the sample.

Using the compilation of CD and FL data, the structure of the protein can be determined as it relates to the folding free-energy landscape (*Figure 1*). After identifying the intermediate state of TDP-43, the idea of therapeutic development targeting this intermediate state was formed. These therapeutics would aid protein folding like molecular chaperones do within the cell. These proteins help other proteins properly fold and shuttle proteins through cell membranes by binding to folding intermediates and redirecting them. They act as the 'big brother' in the cell (Muchowski, 2002). The proteins are studied in isolation so that chaperones do not affect the unfolding of the protein. The aim is to first understand the biophysics of protein folding from a basic science perspective before advancing into other mechanisms.

MATERIALS AND METHODS

The following protocol was tested to optimize the growing and purification conditions utilizing the equipment available in the University of Massachusetts Medical School laboratory of Dr. Jill Zitzewitz. The experiments involving matrin 3 were divided into 3 RRM segments; the first mRRM1, the second mRRM2, and the tethered complex mtRRM (*Figure 4*). Unless otherwise stated, the method was used for all three segments.

Preparation

Transformation

Matrin 3 was purchased in a puc57 vector and transformed into a PET 3d vector for optimal growth and expression. Four nanograms of the matrin 3 DNA was resuspened in 40µL of RoH_2O (a water free of metal and other contaminants) resulting in a final concentration of 100ng/mL. This solution was spun in a mini centrifuge to ensure the DNA was in the RoH_2O. Two microliters of this solution was added to 50-75µL of XL1 blue bacteria cells. The bacteria were stored at -80°C and before adding the DNA solution; they were moved onto ice to thaw. The DNA was mixed into the cells and stored on ice for 15 minutes. Heat shock was not needed as the cells were z-competent. The cells were then plated on ampicillin (AMP [100mg/1mL H_2O) plates and incubated at 37°C for 12-16 hours. A plate of stock protein in PET 3d was also prepared in this manner. This step grew the desired matrin 3 plasmid and the PET 3d vector simultaneously.

After 12-16 hours, round bottom tubes with 5mL of LB broth and 5µL of AMP were prepared. A sterile pipette tip was used to select one grown colony of matrin 3 and it was

inserted into the LB broth. The same was done for all of the mRRMs and PET 3d. These tubes were placed in a rotating incubator at 37°C for 12-16 hours.

DNA Mini Prep

Once the cells had grown in the LB broth overnight, a Qiagen DNA mini prep was performed to isolate the DNA. The cells were spun at 3000g for 15 minutes. The LB was decanted, leaving the cells. The cell sample was prepared using the Qiagen reagents, resulting in 850µL sample which was spun at 1300rpm for 10 minutes to remove the cell membranes. The supernatant containing the DNA and was decanted into a spin column where the DNA bound to the resin. The flow-through was discarded and the resin was rinsed to remove all non-DNA particulates, by centrifugation at 13000rpm for 1 minute. The resin was then dried by centrifugation without any buffer. The DNA was then eluted off of the resin into a clean Eppendorf with a buffer that raises the pH of sample to unbind the DNA from the resin. This mini prep was repeated with all of the cells samples.

The concentration of the DNA was measured with a NANO drop machine. A sample of the read out is *Figure 5*. The concentration of the DNA must be at least 100ng/µL in order to put the plasmid into the correct expression vector.

Enzyme Digest

Once the required concentration of DNA was obtained, an enzyme digest reaction was completed to cut the plasmid from the vector. The reaction contained the DNA, dH_2O, Biolabs Cutsmart buffer, and restriction enzymes. The enzymes used were high fidelity BamH1 and Nco1. These enzymes must be kept cold and mixed well before use, as they are stored in

glycerol. The volumes of each reagent used are reported in *Table 1*. After combining the reagents, the mixture was incubated at 37°C for 1 hour. The solution was then run on a low melting agarose or a Lonza flash gel as described below. One type of gel may prove more successful for DNA extraction but the goal and obtaining of DNA is true for both.

Table 1 Reaction volumes for cutting the matrin 3 plasmid and vector. This table was used to calculate the volumes needed to complete the enzymatic reaction combining the matrin 3 plasmids into the PET 3d vector.

Reaction Volume	**50µL**	**20µL**	**Control**	**Enzyme**
Cutsmart Buffer (µL)	5	2	0.5	0.5
DNA (µL)	4	10	1	1
Nco1-HF (µL)	1	1	0	0 or 0.25
BamH1-HF (µL)	1	1	0	0.25 or 0
dH_2O (µL)	39	6	3.5	3.25

Gel Extraction

A 1% Agarose gel was used to separate the plasmid and the vector that were just cut in the enzyme digest described above. This gel was prepared by adding 600mg agarose to 60mL TAE Buffer and heating until the agarose dissolved. The mixture was then cooled before adding 4µL of ethidium bromide (EtBr). The gel was then poured into a 14-16 lane mold and cooled until hardened. The digests were prepared by adding 1µL of gel loading dye to 5µL digest. These samples and a DNA ladder of known size was loaded onto the cooled gel. The gel was then run at 160V for 20 minutes in 1x TAE buffer. Once finished, the gel was scanned with a Typhoon FLA-9000 to locate the DNA fragments. The EtBr allows the bands of separated vector and plasmid to be visualized (*Figure 6*).

The DNA band of interest was then cut out of the gel using a UV light box. The DNA of the matrin 3 plasmids were cut and the band of the PET 3d vector was also cut. The rest of the gel was discarded. The mass of the cut gel blocks was measured to perform the following

experiments. The gel was then dissolved to remove the DNA by using a Qiagen kit for gel extraction. Following the manufacturer's instructions, the gel was dissolved by addition of Buffer A. The solution was run over a mini spin column to isolate the DNA. The pH of the DNA was raised with a buffer to elute it into a clean Eppendorf. Like before, the concentration of the DNA was measured via Nano-drop.

Ligation

After isolation and purification of the vector and plasmid DNA, a ligation reaction was performed to insert the plasmid into the correct vector. This reaction was performed at different volumes and concentrations, *Table 2*. Depending on the size of the vector and plasmid and their concentrations, different proportions of vector to plasmid may be more optimal. The exact ratio can be determined through trial and error. Once the reactions were made, half of the mixture was incubated at 20°C for 1 hour and the other half was incubated for more than 4 hours. This process of trial and error was repeated throughout the transformation process.

Table 2 Ligation ratios for inserting the matrin 3 plasmid into PET 3d. The different reactions corresponded to the concentration of the DNA and the size of the plasmid or vector. The ratios of insert:vector influence the volume of reagents used in the reactions. The base pair lengths for the matrin 3 RRMs were 600bp for mtRRM and 300bp for mRRM1 and mRRM. The vector PET 3d was 5000bp. The concentration of vector was 50ng or at least 3µL. The Biolabs Quick Ligase kit was used for the buffer and enzyme in these 10µL reactions.

Reactant	**1:9 Ratio**		**1:18 Ratio**	
Plasmid Size (bp)	300bp	600bp	300bp	600bp
Reaction Buffer (µL)	5	5	5	5
Enzyme (µL)	1	1	1	1
Vector (µL)	1	1	1	1
Plasmid (µL)	1	.5	2	1
dH_2O (µL)	2	2.5	1	2

The following equation was used to calculate the volume of insert (plasmid) to use.

$$[Vector] * \frac{bp\ insert}{bp\ vector} * ratio\ of\ \frac{insert}{vector} * \frac{insert\ ng}{[insert]} = \mu L\ insert$$

Six microliters of the ligation mixture was transferred into 50-75μL of DH5α cells and incubated on ice for 30 minutes. Then, 10μL of the reaction was plated on an LB+AMP plate. It was then incubated for 12 to 16 hours at 37°C. One of these grown colonies was picked and added to 6mL of LB broth with 6μL of AMP and the mixture was placed in a rotating incubator for 12 to 16 hours at 37°C. The DNA mini prep, see above, was performed on these cells. A 10μL sample of the isolated DNA was sent to Eton Bio or Genewiz for sequencing. If the sequence matches *Figure 4*, then the ligation and enzyme reactions were successful.

E. coli Stocks

If the sequencing matched the known sequence of the protein, then the DNA was inserted into cells to make glycerol stocks. BL21 DE3 cells are used for growing and protein expression while DH5α cells are used for cloning vectors. A volume of 2.0-3.0μL of DNA of each RRM was placed into these cell types. The cells were incubated with the DNA on ice for 30 minutes before plating onto LB+AMP plates. These plates were incubated at 37°C for 12-16 hours. A colony from each plate was selected and placed into 5mL of LB Broth containing 5μL of AMP for 16-18 hours. The next day, 800μL of cells were added to 500μL of glycerol for 12 hours. This mixture (75μL) was then aliquoted into Eppendorf tubes, frozen in liquid nitrogen and then stored at -80°C.

Purification

Growing Conditions

The protein was first grown in a strain of *Escherichia coli* that is resistant to AMP and chloramphenicol, CAM. The volume of *E. coli* grown depends on the amount of protein needed, so the following volumes are what was typically used for experiments.

Two 14mL round bottom tubes were prepared near a flame with 5mL of LB and 5µL of AMP (1mL/1L). A pipette tip containing the Bl21 DE3 *E. coli* cells was obtained from the freezer and placed into the broth. This was repeated for both tubes. The tubes were then placed in a 37°C rotating incubator for 8 hours. Then 250µL of the grown cells were transferred to a 2 L beveled Erlenmeyer flask with 1L LB and 1mL AMP. This was grown 12-16 hours at 37°C in a shaker, 180rmp. The cells were equally divided into 6 3L beveled flasks with 1.5L LB and 1.5mL AMP. These flasks were grown in the shaker until absorbance of a sample of the LB is OD_{600}=0.6-0.8, typically 6-8 hours. The cells were then induced with 1mL/1L of 1M IGTP and shaken 12-16 hours at 20°C.

After growing the cells, the broth was transferred to 750mL containers to spin at 400rpm for 10 minutes at 4°C. This process was repeated to all the broth containing the same cell line. Each time, the supernatant was decanted, leaving only the cells. Once all the broth has been removed, the mass of the cells was obtained. At this point the cells were frozen or used in further purification.

Lysing expressed protein out of *E. coli* cells

mRRM1: In a ratio of 2mL/g of cells, a native lysis buffer was added to thawed cells. The buffer for mRRM1 was 20mM NaPi, 300mM NaCl, and 30mM imidazole, pH 7.4. In addition to the

buffer, DNAse and a protease inhibitor tablet were added to the protein, not to the stock buffer, and stirred until all components were dissolved. This mixture was then placed in a container on ice. A sonicator, Ultrasonic Processor XL, was used to break apart the cells. The cells were placed in this machine for a cycle of 10seconds of sonication and 45seconds of rest for a total of 6 minutes of sonication. The cells were then spun at 18000rpm in an SS34 rotor for 45minutes at 4°C.

mRRM2: This protein domain was treated similarly to mRRM1 except that its native lysis buffer was pH 8.0.

mtRRM: The mtRRM pellet was resuspended at 2mL/g in a urea lysis buffer at pH 7.4 containing 20mM NaPi, 300mM NaCl, 30mM Imidazole, and 6M Urea. Like before, a protease inhibitor tablet and DNAse are added to the protein pellet which was lysed and spun as described above.

Nickel Affinity Chromatography

mRRM1: A 75mL Ni+ column was cleaned with 3 column volumes (CV) of 20mM MES, 300mM NaCl pH5. It was then washed with 5CV of RoH_2O. Next, it was equilibrated with 2CV of native lysis buffer used for sonication. The buffer was run at a flow rate of 1.5mL/minute. The cell supernatant was poured over the equilibrated column. The flow-through was collected for testing. After the protein was loaded onto the column, it was imperative that the column was never allowed to dry. The column was washed with 15CV of the native lysis buffer. The protein was then eluted with 2CV native elution buffer. This buffer was 20mM NaPi, 25mM NaCl, and 300mM Imidazole, pH 7.4. The protein was eluted with 1.5L 20mM Tris, pH 6.5 or NaPi plus 1mM DTT. Once the protein was fully eluted, 10mL of precision protease was added to the

protein to cleave the His-tag. The solution was continuously stirred and incubated at 4°C for 20 to 24 hours.

mRRM2: Like mRRM1, the Ni+ column was washed and equilibrated with the buffer unique to mRRM2. The protein was loaded, washed, and then eluted with 2CV of a buffer at pH 7.0 containing 20mM NaPi, 25mM NaCl and 300mM Imidazole into 1.5 liters of 20mM Tris and 1mM DTT at pH 6.5. This was then cut with precision protease for 20 to 24 hours.

mtRRM: The Ni+ column was washed and equilibrated with the buffer unique to mtRRM. The protein was loaded, washed, and then eluted with 2CV of a buffer at pH 7.0 containing 20mM NaPi, 300mM Imidazole, and 6M urea. The eluted protein was dripped slowly via separator funnel into 2L of pH 6.5 protein containing 20mM Tris and 1mM DTT. It was then cut with precision protease as previously stated.

Further Purification

After the incubation with precision protease, the protein was filtered with 20micron filter to remove any aggregates. The 5mL S-column was cleaned with 5CV filtered RoH_2O, 5CV of filtered 1M NaOH, and then rinsed with 5CV of filtered RoH_2O. It was then equilibrated with 3CV of filtered 20mM NaPi, pH 6.5. The filtered protein was then loaded onto the column at 1.5mL/minute at 4°C. The column was washed with 5CV of filtered NaPi. The column was then linked to a Biorad DuoFlow, which elutes the protein in a gradient by combining two buffers. The two buffers are a low salt filtered 20mM NaPi and a filtered high salt 1M NaCl, 20mM NaPi, both at pH 6.5. Once the protein was eluted in a gradient, the resulting graph is analyzed.

The graph displays the absorbance of different aliquots at varying concentrations of the two buffers, *(Figure 7*). Samples of the peaks were collected and run on a gel to determine the

location of the protein. Once this was determined, the protein was pooled and dialyzed in 3kDa tubing for the individual RRMs and 6-8kDa tubing for mtRRM. The dialysis buffer is 1L of pH 7.2 10mM KPi, 150mM KCl, and 1mM BME. The tubing was moved into another container of this buffer after 6 hours to fully buffer exchange the protein. The protein was then filtered and ready for experiments. The concentration of the protein was calculated from *Beer's Law Adaptation*.

Beer's Law Adaptation

To find the protein concentration a UV spectrometer was used. The wavelengths collected in this scan were from 350nm-250nm. The absorbance of protein is at 280nm. A sample of 1mL LB Broth was inserted into a UV Spec cuvette. This sample was scanned as the baseline for the absorbance. Once the baseline was established, 1mL of the protein sample was inserted into the cuvette and scanned. The resulting graph was used to calculate the following:

$$\frac{A_{280} - A_{320}}{E * l} = C$$

The absorbance (A) of the protein at wavelengths of 280 and 320 is used to baseline correct the scan. The extinction coefficient (E) is specific to different proteins at 280nm. For matrin 3 the values are as follows: mRRM1=6970, mRRM2=3840, and mtRRM=10810. The path length (l) is the length of the cuvette, 1cm. The factors are used to calculate the concentration of the protein.

<u>Electrophoresis glycerol gel to test protein purity</u>

Samples at each step of the purification were collected for gel analysis. To run this gel, a dilute 16µL sample was prepared. Each sample was combined with 6µL of SDS-page running buffer. This dye, like EtBr for DNA, allows the protein sample to be seen on the gel. These

samples were then heated to 98°C for 7minutes. A 15% Tris gel was used for these runs. The running buffer used in this process prepared in 4L to create a 5x stock at pH 8.3, containing 376g glycine, 60.4g Tris base, and 20g SDS. To use the running buffer, the stock is diluted to 1x with RoH_2O. After the samples were heated, 15μL of sample was loaded into the 14 well gel after the gel tray had been filled with the running buffer. The gel was then run at 140V for 75minutes. See *Figure 8* for a sample gel.

Size Exclusion Chromatography for mRRM2

To run this size exclusion column, the protein must be concentrated because only 10mL of sample can be used per each 8 hour run. Similar to the gradient elution, this column eluted the protein in small aliquots. The buffer used on the column was filtered 20mM Tris and 1mM DTT at pH 8.8. This buffer was appropriate for mRRM2 purification because of the buffer's pH..

Once the size exclusion column was connected to the Biorad Duoflow system, 5mL of concentrated protein was injected. After 16 minutes, the other 5mL was injected into the system. There was this delay because the column can only handle 5mL of protein every 4 hours and so this delayed injection allows this to occur.

After the sample was run for 8 hours, a conductance graph was generated. From the graph, the location of the protein was determined. To check that this was the correct sample, an electrophoresis gel was run. Once the identity of the protein was confirmed, the concentration of the sample was measured via NanoDrop. After this, the protein was dialyzed in 3kDa dialysis tubing into 10mM KPi and 150mM KCl at pH 7.2 (the physiologically relevant buffer). It took approximately 16 hours and 2L of buffer for a full buffer exchange to occur. Once the protein was in this physiological buffer, it was ready for experiments.

mRRM1-ICC

Initial studies were done with mRRM1, *Figure 9*. However, a paper was released indicating that the residues IKK at the beginning of the linker actually forms an Interacting C-terminal Coil (ICC) (Blatter et al., 2015). Knowing this, these residues were added back in using the protocol stated in **Preparation**. This new RNA Recognition Motif was tested as the others were. It was found to be best purified by the method of mRRM1, using a native buffer. This resulted in the following sequence.

Data Collection

Equilibrium Unfolding Experiments

After obtaining purified samples in the physiological buffer, it was important to know their concentrations. This measure is stated in *Beer's Law Adaptation*. Calculations were then performed to create stocks of protein with denaturant and without denaturant. The exact volumes of sample needed, concentration of protein required, and the concentration of the denaturant required depended on the experiment. The following conditions were used.

Sample of tRRM
Volume of each sample: 0.50mL
Concentration of Protein/Sample: 4.20uM
Number of Samples: 41
Denaturant (Urea) Gradient: 0-8M
Aliquots every: 0.2M urea

Measured Stocks:
Concentration of Protein: 68.70uM
Concentration of Denaturant: 9.80M

Knowing these conditions, 2 stocks were made. The first had no denaturant, the second had the highest concentration. In calculating the total volume of these samples, it was important to

consider the total volume required, in this case 20.5mL, and add a 'dead' volume, 5.50mL was used in these calculations. With these parameters, each stock was 13.00mL.

Calculations:

$$M_1V_1 = M_2V_2$$

$$V_2 = volume\ of\ concentrated\ protein = \frac{M_1V_1}{M_2} = \frac{4.20uM * 13mL}{68.70uM} = 0.795mL\ per\ stock$$

$$M_1V_1 = M_2V_2$$

$$V_2 = volume\ of\ concentrated\ denaturant = \frac{M_1V_1}{M_2} = \frac{8M * 13mL}{9.8M} = 10.612mL$$

Therefore the compositions of the stocks are as follows:

Low Denaturant: 0.795mL protein, 12.205mL physiological buffer

High Denaturant: 0.795mL protein, 10.612mL denaturant, 1.593mL physiological buffer

These buffers were then aliquoted into 0.5mL samples while increasing the denaturant concentration by 0.2M in each sample. This was done by hand with further calculations like the ones listed above or by an auto-sampler program. Once made, the samples sat in light-proof wrapping for 12-24 hours before testing.

Circular Dichroism

This method allows for the observation of secondary protein structure. Following the sample used in the above calculations, a 2mm cuvette is used. This cuvette required 0.5mL of sample. The size of the cuvette required depended upon the concentration of the protein. If the concentration is too high for a large volume cuvette, then the proteins will cause the light to scatter, interrupting the data gathered. This protein concentration to cuvette size ratio was established through several trials.

As eluded to, circular dichroism requires light to be passed through a cuvette. The recording is taken over a range of wavelengths. Typically this range is 195-260nm, but if trials found that the protein produced a signal that scattered the lights too much to get a reading, then the range was 215nm to 245 or 260nm. The data was manipulated to show the MRE value. The data coming off of the machine was interpreted with a lesser degree of intensity. To understand the secondary structure of a protein, a few things must be understood. Firstly, it that a protein is made of alpha helices and beta-pleated sheets. The interpretation of this data is shown in *Figure 10.*

Far-UV CD

The analysis thus far used Far-UV. This type of CD was useful for analyzing the secondary structure of the protein. It does this by scanning across the ultraviolet light spectrum. It was used across a denaturation gradient because it allowed analysis as to at what concentration the protein loses its secondary structure. This is evident because there were no longer any peaks in the CD reading.

Near-UV CD

A Near-UV reading occurs at a spectrum higher than that of Far-UV. In this experiment, a range of 320-245nm was used. It was used to study the more disordered regions of the protein. The cuvette was larger at 5cm and it holds 10mL of sample. A larger sample at a lower protein concentration was used to reduce the amount of light scattered. The samples was prepared by the same method as discussed.

Refractometer

The concentration of denaturant was checked by a refractometer. Using a blank with no denaturant, the refractometer will be able to determine the concentration of denaturant based on the refraction of light. Even though the samples were calculated to have a certain concentration, the refractometer reading was considered to be the actual denaturant concentration for the sample as incubation of the samples overnight may have changed the concentration should some denaturant crystalize.

Fluorescence Spectroscopy

Fluorescence spectroscopy was used to measure the tertiary structure of the protein. It is able to do this by analyzing aromatics, tryptophan and tyrosine which are excited at 295nm and 274nm, respectively. The spectroscopy uses these fluorescing residues to measure their location in space and also their fluorescence intensity. From the same sample used with CD, 70uL of sample was needed in the special fluorescence cuvette. Depending on the composition of the RRM, a different excitation wavelength was used. The tryptophan excitation was used for mRRM2 and mtRRM rather than the tyrosine excitation, which was used for mRRM1, because the fluorescence intensity signal for tryptophan is 4 times that of tyrosine.

Calcium Binding Assay

A paper published by the Liu lab in 2007 indicates that matrin 3 may bind to calcium. Specifically it is mRRM2 that binds to calmodulin (Valenci et al., 2007). Knowing this, a calcium binding prepared for mRRM2. The calcium buffer contained 20mM HEPES, 150mM KCl, 1mM $CaCl_2$, and 1mM BME. The control buffer without calcium contained 10mM KPi,

150mM KCl, and 1mM BME. Using these buffers, samples both with and without calcium were made at protein concentrations of 0uM, 2.5uM, 5uM, 7.5uM, and 10uM. The samples were then scanned by FL and CD, as previously stated.

Electomobility Shift Assay (EMSA)

The protein matrin 3 was selected in this investigation because it is an RNA-binding protein. Previous projects had investigated other RNA-binding ALS-linked proteins such as TDP-43, hnRNP A1, SOD-1, and FUS. To test the RNA binding of the protein, an RNA sequence was selected that matrin 3 has an affinity for (Yamazaki et al., 2014).

5'- CUUCUC**ACUACU**GCACUUG**ACUAGU**CU - 3'

The bolded regions are the parts of the sequence that the protein binds to.

To perform this assay, the concentration protein must be measured. Based on the concentration, a serial dilution was made by changing the protein concentration, either by half or a third. A 96 well plate is used to make the dilutions. The dilution was made by adding 30uL of the highest concentration protein into the first well. Then 15uL of physiological buffer was added to all the 20 wells. After mixing the first cell, 15uL was removed and added to the second cell. This too was mixed and 15uL was removed and added to the third cell. This repeats until the last cell was reached, the 15uL removed was discarded.

To make the RNA part of the sample, calculations are made taking into account that 45uL will go into each well (*Table 3*).

Require 2200uL total= 20 wells*45uL+1300uL dead volume

Buffer:

$$M_1V_1 = M_2V_2$$

$$Volume = \frac{1x * 2200uL}{10x} = 220uL$$

tRNA and IGEPAL:

$$Volume = \frac{1x * 2200uL}{100x} = 22uL * \frac{4}{3} = 29.3uL$$

DTT:

$$Volume = \frac{2mM * 2200uL}{1000mM} = 4.4uL * \frac{4}{3} = 5.87uL$$

RNA:

Measured concentration by Nanodrop: 1120nM

$$Volume = \frac{3nM * 2200uL}{1120nM} = 5.89uL * \frac{4}{3} = 7.86uL$$

dH_2O:

$$Volume = 2200uL - 2200uL - (29.3uL * 2) - 5.87ul - 7.86uL = 1907.8uL$$

Table 3 EMSA sample Buffer volumes. Following the above calculations, the following table summarizes the volumes needed to create the desired concentrations for the reagents.

Stock	**[Stock]**	**[Desired]**	**Calculated Volume (uL)**
Physiological buffer	10x	1x	220
tRNA	100x	1x*4/3	29.3
IGEPAL	100x	1x*4/3	29.3
DTT	1M	2mM*4/3	5.87
RNA	1120nM	3nM*4/3	7.86
dH_2O	-	-	1907.8

The buffer was mixed, save for adding the DTT and RNA. The mixture was heated for 5 minutes and then DTT and RNA are added once it has cooled a bit. Then, 45uL of this buffer was added to each well. The plate then equilibrates for 2 hours.

During that time, the acrylamide gel was made. This gel is 8% acrylamide and is made of 90mL acrylamide, 450mL dH_2O, 45mL 10x running buffer (TB buffer for 1L 55g Boric acid, 108g Tris base), 4.5mL 10% APA (450mg ammonium persulfate), and 450mL TEMED. The TEMED and APS were added simultaneously to the rest of the mixture in a graduated cylinder and it was covered and inverted. A gel mold with 20 50uL wells was prepared and washed with RoH_2O. Once the mixture was made, it was poured into the mold to set.

Once the samples have equilibrated and the gel set, 6uL of 1uM bromocresol gel shift dye was added to each well. Only 50uL of the sample was loaded into each well. The 10x running buffer was diluted to 1x to run this gel. The gel was then run by electrophoresis at 140V for 60minutes. After the run, the gel was scanned using a fluorescence machine called Typhon FLA. This measured the fluorescence of the protein and RNA bands in this neurotoxic gel. The images were then be analyzed according the Hill equation to measure the cooperativity of the binding.

RESULTS

From the methods discussed in **Data Collection**, the following results have been obtained and interpreted. The data was presented to show the raw data from CD, FL, and EMSA and what can be drawn from this information. The data was then expanded upon in the interpretation of graphs and by using analysis software to manipulate the data. Throughout the data representation, the same coloring scheme is used: mRRM1 is light blue, mRRM1-ICC is blue, mRRM2 is orange, and mtRRM is grey. If additional information is displayed, the legend and caption contain details. Some figures have multiple panels, preventing the caption from being on the same page. In cases like this, the caption was placed after the figure.

Initial CD Studies

In the far-UV CD Scan (*Figure 11*), mRRM1 has defined helixes indicated by its distinct divots at 222nm and 208nm. mRRM2 is more random because it's helixes are less defined, yet it doesn't completely resemble beta structure. When denaturant is added, mRRM1 loses all structure. mtRRM resembles both of them with more defined helixes than mRRM1.

In CD signal of the tethered domains as mtRRM can be compared to the signals of mRRM1+mRRM2. If the CD signal of the additive structure does not shift the minima compared

to the mtRRM signal, then the tethered complex provides no additional support in joining the two domains together (*Figure 12*).

Protein Concentration CD Experiments

To analyze the data to test the effects of denaturant at different concentrations of mRRM2, SAVUKA can be used. The graphs demonstrate this analysis in addition to the importance of protein concentration. To start the analysis, the CD signal is obtained across a spectrum of wavelengths (*Figure 13 A*). Initial CD gathers data across a spectrum of wavelengths.

With different samples of protein with different concentrations of denaturant, not only was there data regarding the secondary structure across wavelengths but the data from the different samples were compiled to look at one wavelength across different denaturant concentrations. The wavelength selected was the one where the data points best fit a modeled two-state or three-state curve. A model two-state unfolding curve represents a protein going right from a native state to an unfolded state $N \leftrightarrow U$. In a three-state model, a protein would unfold to an intermediate state before unfolding, $N \leftrightarrow I \leftrightarrow U$ (*Figure 13 B*).

Finally, the data was fit to a Fraction Apparent Plot (*Figure 13 C*). This type of plot takes the data of MRE and flattens the base lines because the protein is either folded/native (0 Fraction Apparent Unfolded) or unfolded (1 Fraction Apparent Unfolded). Free energy of the systems equilibrium is calculated based on midpoint of the FAP curve. The differences between concentrations of protein are determined by the location of the positive slope relative to the other slopes. This analysis determined that the higher the concentration of denaturant needed to reach the midpoint of this slope, the more stable to the complex.

Initial FL Studies

In order to examine tertiary structure, fluorescence spectroscopy was performed. In the native FL scan, looking at the 4 tyrosines in mRRM1 a scan was performed to excite at 274nm (*Figure 14 A*). mRRM2 contains 4 tyrosines, like mRRM1, and a tryptophan. Therefore, the mtRRM contains 8 tyrosines and a tryptophan, so the scan used excited the tryptophan at 295nm (*Figure 14 B*).

Addition of the Interacting C-terminal Coil to mRRM1

The addition of the Interacting C-terminal Coil did affect the stability of the motif (*Figure 15*). Therefore, the results beyond this point will display mRRM1-ICC as the results because it more accurately represents the motif.

Near- UV CD

In Near-UV CD, the more disordered domains were probed. The spectra occurred at a higher wavelength than the Far-UV CD (*Figure 16*) and so the MRE wavelength used was 280nm.

Comparative CD and FL signaling across and between domains

After obtaining CD and FL data, the results were compared to the Far-UV CD MRE and against the FL intensity. Typically the FL signal will go from high to low across the denaturant gradient and the CD will go from low to high. This doesn't appear to be the case for any of the motifs, even though all were properly fit to a two-state model (*Figure 17*).

Comparing the FAP of all the motifs reveals forces in each domain stabilizing itself without the tethering to another domain (*Figure 18*). Adding the ICC to mRRM1 stabilized the domain so it could withstand higher concentrations of denaturant before unfolding. Each individual RRM domain and the tethered RRMs of matrin 3 reveal two-state equilibrium unfolding transitions, providing no evidence for an intermediate state.

Calcium Binding of mRRM2

In order to examine calcium binding to mRRM2, a calcium buffer was created. The experiments shown in *Figure 19* were replicated to compare to this new set of data. This domain has been shown to bind calcium. The Calcium binding loop of the protein contains the single tryptophan present in mRRM2 and without the Ca^{+2}, it is switching between conformations, lending to the 'unfolded' looking signal when it is supposed to be native baseline.

RNA-Binding Experiments

matrin 3 was initially identified for this investigation because it is an RNA-binding protein associated with ALS. The affinity of the RNA-Recognition Motifs for RNA was studied by FCS. Preliminary assays by EMSA suggest that mRRM1 and mRRM2 do not have strong RNA binding ability (*Figure 20 A, B*). If the protein had bound to RNA, then it would have created a larger complex that moved more slowly through the gel. Such a complex could be seen with mtRRM (*Figure 20 C*). It showed a higher affinity of RNA binding, so EMSA and FCS was run to test this binding and affinity. In the replication of mtRRM RNA assays at even higher protein concentrations, there was binding but there was also free protein (*Figure 20 D*). This may

hint at a higher order species but FCS reveals no higher order species created when there is an increase in protein concentration; therefore mtRRM is just a weak binder.

DISCUSSION

The presence of an intermediate state is marked by a difference in CD and FL midpoints. This difference would indicate that there was a conformation formed in which the tertiary or secondary structure is changing while the other type of structure is not., meaning the CD or FL data fits a three-state model ($N \leftrightarrow I \leftrightarrow U$). For several reasons, an intermediate state was not located in the matrin 3 RNA-Recognition Motifs. With the inconclusive fluorescence data because it had a positive rather than the expected negative slope, it was not possible to determine the difference in midpoints between CD and FL data (*Figure 17*). Both the data gathered by Far and Near-UV CD fit a two-state model ($N \leftrightarrow U$) (*Figure 16, 17*). Even though there was no intermediate state found, these results could be used to create a prediction for the presence of an intermediate state among other RNA-binding proteins. There was an intermediate state identified in FUS RRM and TDP-43 RRM2 (Mackness et al., 2014). TDP-43 RRM1, mRRM1, and mRRM2 all lacked an intermediate state. Further analysis into the sequences and structures of these domains may unveil some common characteristic differentiating between the intermediate contain group and the one that lacks it.

Beyond the main intermediate investigation, supplemental investigations were performed. Studies had said that matrin 3 was a calcium binding protein because of the interaction of mRRM2 and ZnF 2 (Valenci et al., 2007). Current experimentation showed that in isolation, while there was a stabilization of the domain in the presence of calcium, it was not feasible to

perform the same experiments that had been done with TDP-43 (*Figure 19*). Also, the difference with the calcium buffer was not significant enough to pursue this deviation of methodology.

As an RNA-binding protein, matrin 3 was expected to bind to RNA. Well, it would appear as though the RNA-Recognition Motifs of matrin 3 do not play a role in the RNA binding. At least, this statement is true given the sequence of RNA used (5'-CUUCUC*ACUACU*GCACUUG*ACUAGU*CU – 3') which had been identified in previous research to be the sequence that isolated matrin 3 binds to (italicized regions) (Yamazaki et al., 2014). It is not required that the RRM bind to RNA, only that it is able to interact with it in some way. This connects back to the phenotype of matrin 3 ALS mutations being RNA mismanagement (Salton et al., 2011). The exact influence of the mutation depends on its location (Johnson et al., 2014). This brings into the light how the location of ALS mutations occur in disordered domains of the proteins (*Figure 1*).

Mutations to the RRMs have been used to induce conformational changes. Currently, missense mutations are being investigated for TDP-43 to mimic the conformation of the intermediate state. This is not maintained for long and only occurs in 0.2% of the protein population at a time (Mackness, 2016). The naturally occurring rate of this conformation would make it difficult to develop a therapeutic to treat it. Therefore, mutations can be used to mimic these states. Mutants of matrin 3 can also be made to simulate the ALS patient's mutants. These methods could be used to further therapeutic development, which is the next step of this research.

Another step that could be taken is investigating other RNA-binding ALS-linked proteins. There is a whole class of these proteins and only some of them have been studied in the context used in this experiment. If the prediction of an intermediate state is not clear from the

current base of research, then more research must be done to fill this gap of knowledge. The ability to make prediction of an intermediate state and eventually being able to create a model for this conformation (seen in *Figure 1* for TDP-43) would allow for therapeutic testing to begin at a later stage than the biophysical experiments seen in these trial. Knowing the shapes of the target conformation and the therapeutic could invite computer simulations to mimic these interactions. Large scale prediction software has been established in programs like Folding@home. Designed by the Pande lab, this program can be downloaded to devices connected to the internet to simulate folding patterns of proteins until the proper conformation is formed (Stanford, 2013).

The marriage of simulation and wet lab work can be continued with further research and development from both points of reference. This research has been enabled with national and crowd funding, as seen in the Ice Bucket Challenge. There is power in awareness of the public to the research that is being done on disease. Although cures are not popping up every day, research is always happening to make people's lives better in any way scientists can. Action arises from the ability to understand the cycle of research for the continuation of the pursuit of a better future.

REFERENCES

Afroz T, Cienikova Z, Clery A, Allain F. (2015). One, two, three four! How multiple RRMs read the genome sequence. *Methods in Enzymology*. 558: 235-278.

ALS Association. Washington, DC: ALS Association [2016]. alsa.org

Blatter M, Dunin-Horkawicz S, Grishina I, et al. (2015). The signature of the five-stranded vRRM fold defined by functional, structural, and computational analysis of the hnRNP L protein. *Journal of Molecular Bio*. 427(19):2997-3000.

Cooper, GM. (2000). The Cell: A Molecular Approach. 2nd edition. Sunderland (MA): Sinauer Associates; https://www.ncbi.nlm.nih.gov/books/NBK9957/

Gallego-Iradi MC, Clare A, Brown H, et al. (2015). Subcellular Localization of Matrin 3 containing mutations associated with ALS and distal myopathy. *PLOS.*

Greenfield, M. (2009). Using circular dichroism spectra to estimate protein secondary structure. *PMC.*

Institute of Psychiatry, Psychology, and Neuroscience. London: Amyotrophic Lateral Sclerosis Online Databank. [2007]. http://alsod.iop.kcl.ac.uk/

Johnson, J. et al. (2014). Mutations in the *matrin 3* gene cause familial amyotrophic lateral sclerosis. *Nat Neuro.* 17(5):664-6.

Mackenzie I, Rademakers R, Neumann M. (2010) TDP-43 and FUS in amyotrophic lateral sclerosis and frontotemporal dementia. *Lancet Neurology.* (9); 995-1000.

Mackness, Brian. (2016). The identification and targeting of partially-folded conformations on the folding free-energy landscapes of ALS-linked proteins for therapeutic intervention. PhD Thesis, University of Massachusetts Medical School.

Mackness B, Tran M, McClain S, Matthews CR, Zitzewitz JA. (2014). Folding of the RNA Recognition Motif (RRM) Domains of the ALS -linked Protein TDP -43 Reveals an Intermediate State. *Journal of Bio. Chem.*

Marangi G, Lattante S, Doronzio P, et al. (2016). Matrin 3 variants are frequent in Italian ALS patients. *Neurobio of Aging.* 49:1-7.

Millecamps S, Septenville A, Teyssoum E, Daniau, Camuzat A, Albert M, LeGuern E. Galimberti D. (2014). Genetic analysis of matrin 3 in French amyotrophic lateral

sclerosis patients and frontotemporal lobar degeneration with amyotrophic lateral sclerosis patients. *Neurobio. Of Aging.* 35(12): 13-5.

Miller, RG, Mitchell, JD, Moore, DH. (2012). Riluzole for amyotrophic lateral sclerosis (ALS)/motor neuron disease (MND). *Cochrane Database of Systematic Reviews*. 3: CD001447.

Moloney C, Rayaprolu S, Howard J, et al. (2016). Transgenic mice overexpressing the Las-linked protein Matrin 3 develop a profound muscle phenotype. *Neuro of Disease*. 4(122).

Muchowski P. (2002). Protein Misfolding, Amyloid Formation, and Neurodegeneration: A Critical Role for Molecular Chaperones? *Cell Press*. 35: 9-12.

Neumann M, Sampathu D, Kwong L, et al. (2006). Ubiquinated TDP-43 in Frontotemporal Lobar Degeneration and Amyotrophic Lateral Sclerosis. *Science.* 314:130-3.

NINDS. Bethesda: Amyotrophic Lateral Sclerosis (ALS) fact sheet [2016]. http://www.ninds.nih.gov/disorders/amyotrophiclateralsclerosis/detail_ALS.htm

Robberechet W, Philips T. (2013). The changing scene of amyotrophic lateral sclerosis. *Nature.* 248-64.

Rowland LP. (2001). How Amyotrophic Lateral Sclerosis Got Its Name: The Clinical-Pathologic Genius of Jean-Martin Charcot. *Arch Neurol.* 58(3):512-5.

RVW Foundation. Hillsdale: Lou Gehrig. [2015]. lougehrig.org

Subbarayalu, P, et al. (2015). Abstract P4-05-09: Matrin 3: A novel micro-tubule associated RNA binding protein that acts as a potent tumor suppressor. *Cancer Research*. 75(9).

Salton M, Elkon R, Borodina T, Davydov A, Yaspo M-L, Halperin E, Shiloh Y. (2011). Matrin 3 binds and stabilizes mRNA. *PLOS.*

Stanford: Folding@Home [2013]. folding.stanford.edu

Stoica R, De Vos K, Paillusson S, et al. (2014). ER-mitochondria associations are regulated by the VAPB-PTPIP51 interaction and are disrupted by ALS/FTD-associated TDP-43. *Nature.* 5(3996):1-12.

Taylor P, Brown R., Cleveland D. (2016). Decoding ALS: from genes to mechanism. *Nature*. 539: 197-206.

Valencia CA, Ju W, Liu R. (2007). Matrin 3 is a Ca^{2+}/calmodulin-binding protein cleaved by caspases. *Biochem. And Biophys. Research Comminications* 361: 281-6.

Wang H. (2015). The role of matrin 3 in the pathogenesis of amyotrophic lateral sclerosis. MS Thesis, Boston University.

Weebly. United Kingdom: Stephen Hawking. [2014]. hawking.org.uk

Wu L-S, Cheng W-C, Shen C-K J. (2012). Targeted Depletion of TDP-43 Expression in the Spinal Cord Motor Neurons Leads to the Development of Amyotrophic Lateral Sclerosis-like Phenotypes in Mice. *Journal of Bio. Chem*. 287(33):27335-44.

Xu Z, Yang C. (2014) TDP-43- The key to understanding amyotrophic lateral sclerosis. *Rare Diseases*. 2(1):e944443.

Yamazaki F, Kim HH, Lau P, et al. (2014). pY RNA1-s2: A highly retina-enriched small RNA that selectively binds matrin 3 (Matr3). *PLOS.*

Zhou Y, Liu S, Ozturk A, Kicks G. (2014). FUS-regulated RNA metabolism and DNA damage repair: Implication for amyotrophic lateral sclerosis and frontotemporal dementia pathogenesis. *Rare Diseases.* 2:e29515.

107th Congress. (2002). Rare Disease Act of 2002. *National Institute of Health.* http://history.hin.gov/research/downloads/PL107-280.pdf

ABBREVIATION KEY

Amyotrophic Lateral Sclerosis (ALS)

Familial ALS (FALS)

Sporadic ALS (SALS)

Ribonucleic Acid Recognition Motif (RRM)

Cu-Zn Superoxide Dismutase (SOD1)

Trans-activation response [TAR] deoxyribonucleic acid [DNA] binding protein 43 (TDP-43)

Fused in sarcoma (FUS)

matrin 3 1st RRM domain (mRRM1)

matrin 3 2nd RRM domain (mRRM2)

matrin 3 tethered RRM domains (mtRRM)

Nuclear Exportation Sequence (NES)

Nuclear Localization Sequence (NLS)

Glycine rich domain (Gly-rich)

Arginine-Glycine-Glycine repeating regions (RGG-Rich)

Zinc finger (ZnF)

Glutamine-Glycine-Serine-Tyrosine rich region (QGSY-Rich)

Protein Databank (PDB)

Tris base, acetic acid, and EDTA buffer (TAE)

Ethylenediaminetetraacetic acid (EDTA)

Escherichia coli (*E. coli*)

Ampicillin (AMP)

Chloramphenicol (CAM)

Lysogeny Broth (LB)

Column Volume (CV)

2-(N-morpholino)ethanesulfonic acid (MES)

Sodium Phosphate (NaPi)

Electromobility Gel Shift Assay (EMSA)

Interacting C-terminal Coil (ICC)

Circular Dichroism (CD)

Secondary (2°)

Molar residue ellipticity (MRE)

Fluorescence Spectroscopy (FL)

Tertiary (3°)

Tryptophan (Trp)

Tyrosine (Tyr)

Fraction Apparent Plot (FAP)

APPENDIX

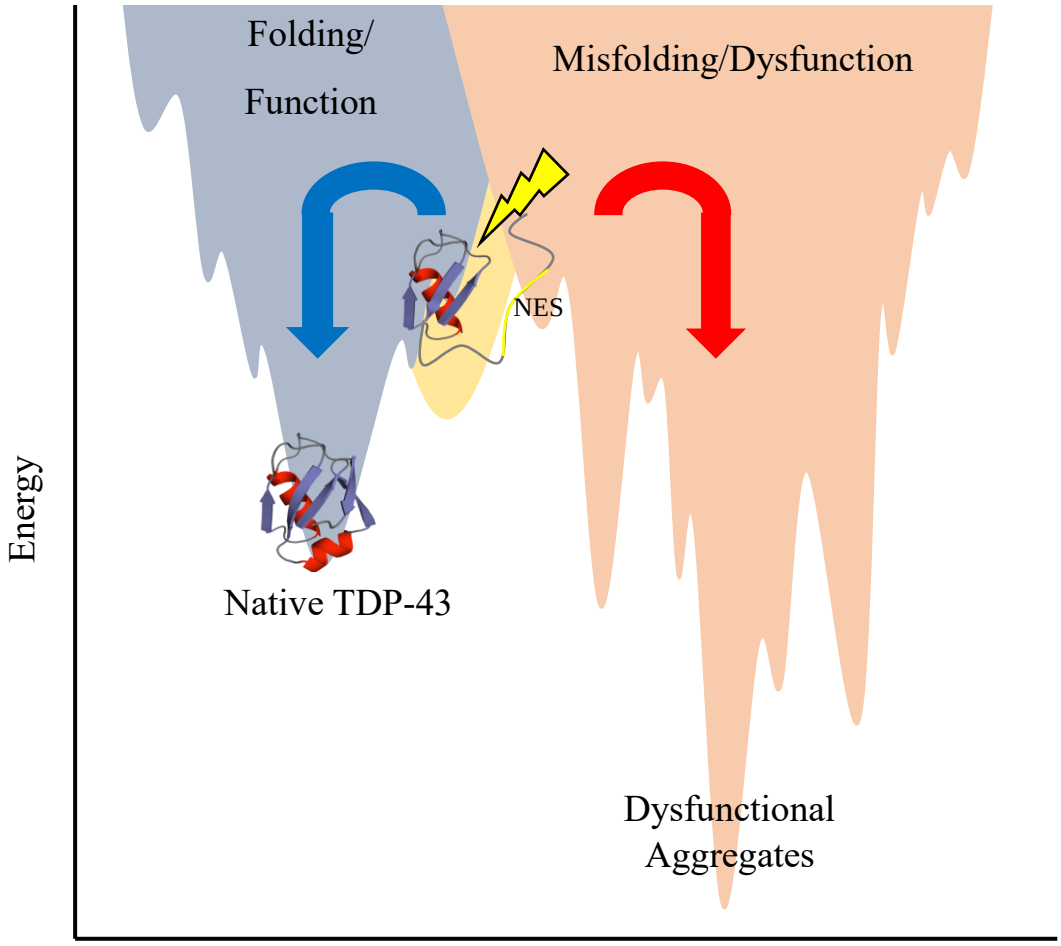
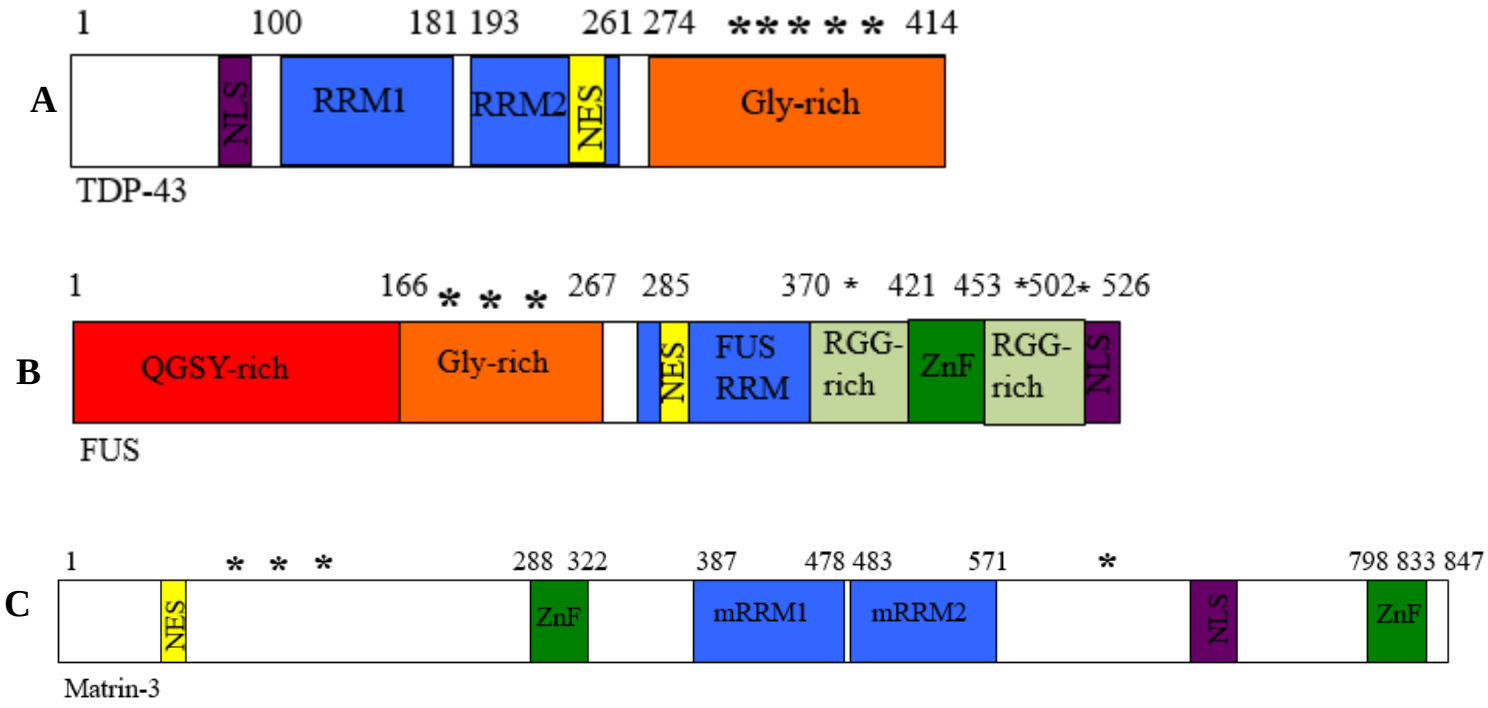


Figure 1 TDP-43 folding free-energy landscape. After the identification of the intermediate state in RRM2-yellow-, it was possible to identify the structure of this state. In this intermediate, the NES became exposed. Reaching this point, the protein could either fold to its native state- blue, or it could aggregate-red. Once a protein aggregates, it is difficult to force it back up the folding pathway as it requires too much energy to overcome the transitional barriers.



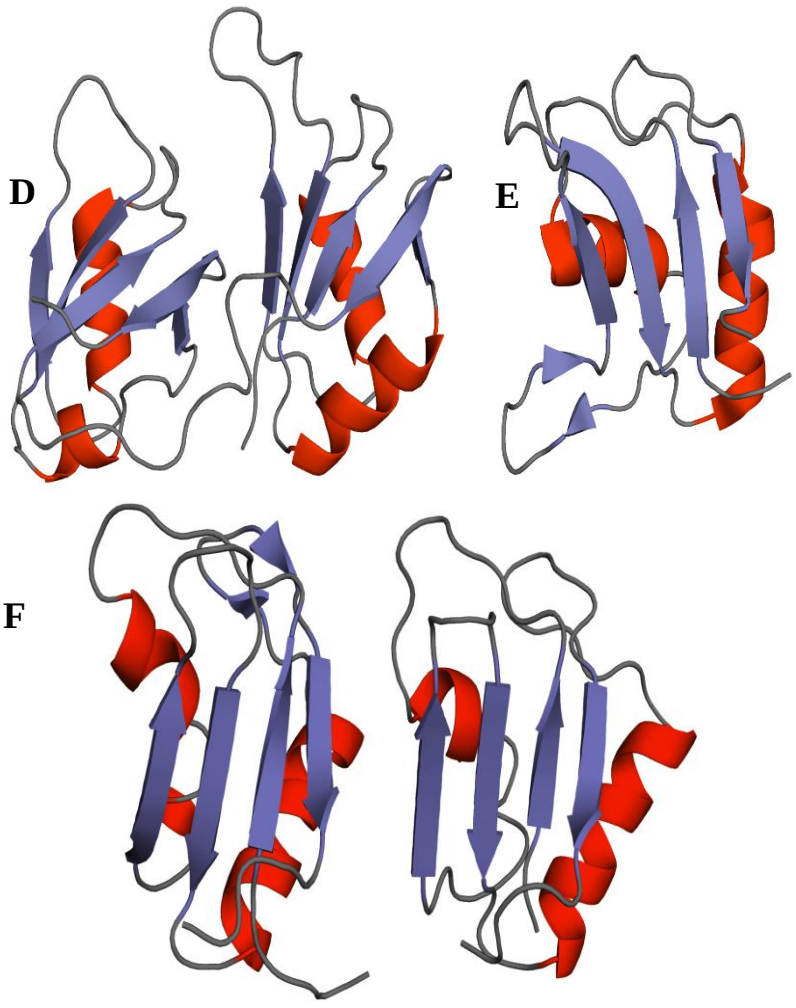


Figure 2 ALS-linked proteins sequence and topology comparison. The sequences are colored according to the following key: ribonucleic acid (RNA) Recognition Motif (RRM) domains-blue, nuclear exportation sequence (NES)- yellow, nuclear localization sequence (NLS)-purple, glycine rich domain (Gly-rich)- orange, arginine-glycine-glycine repeating regions (RGG-Rich)- light green, zinc finger (ZnF)-green, glutamine-glycine-serine-tyrosine rich region (QGSY-Rich)- red, ALS mutation *. **A-C:** The sequences of TDP-43, FUS, and matrin 3, respectively.

The RRM topologies are colored with alpha helices in red and beta-pleated sheets in blue. **D:** In TDP-43, RRM2 is on the left and RRM1 is on the right as the tethered complex assembles in this manner (PDB: 4BS2). **E:** FUS RRM is just a single complex. (PDB: 2LA6) **F:** The crystalized structure of the tethered matrin RRMs has not been identified so the mRRM1 (PDB: 1X4D) and the mRRM2 (PDB: 1X4F) are displayed to from right to left.

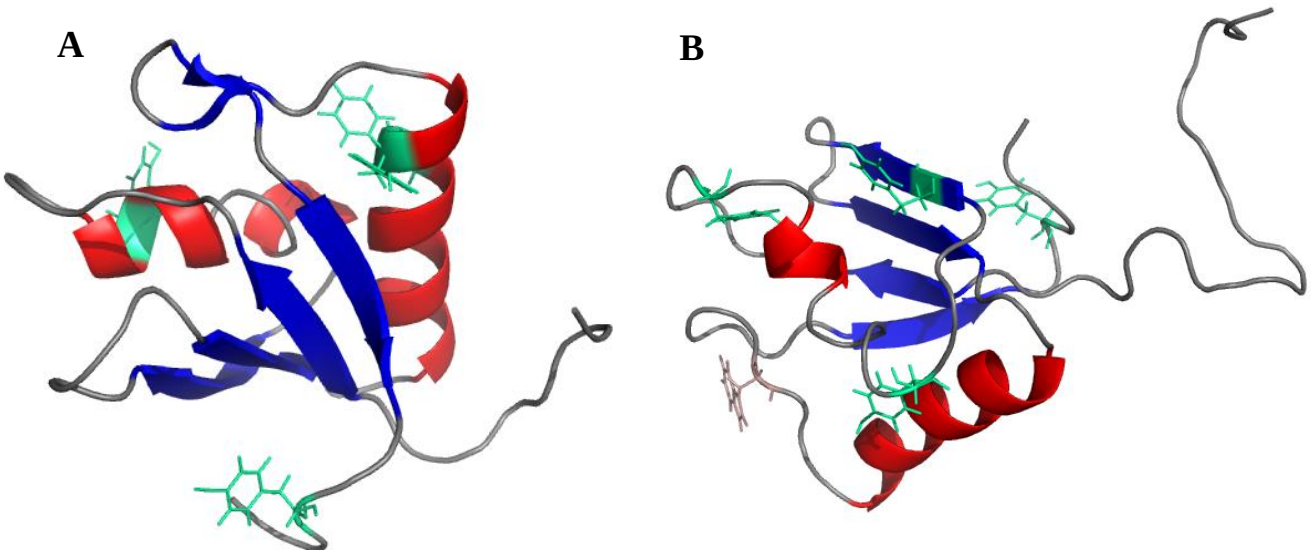


Figure 3 mRRM1 and mRRM2 aromatic amino acids. A: mRRM1 contains 4 tyrosines, teal. **B:** mRRM2 contains 4 tyrosines and a tryptophan, pink. These aspects of tertiary structure are studied because they can be excited at particular wavelengths of light.

mtRRM
*MGHHHHHHGLEVLFQGP*VETSRVVHIMDFQRGKNLRYQLLQLVEPFGVISNHLILNKIN
EAFIEMATTEDAQAAVDYYTTTPALVFGKPVRVHLSQKYKRIKKPEGKPDQKFDQKQE
LGRVIHLSNLPHSGYSDSAVLKLAEPYGKIKNYILMRMKSQAFIEMETREDAMAMVDH
CLKKALWFQGRCVKVDLSEKYKKL

mRRM1
*MGHHHHHHGLEVLFQGP*VETSRVVHIMDFQRGKNLRYQLLQLVEPFGVISNHLILNKIN
EAFIEMATTEDAQAAVDYYTTTPALVFGKPVRVHLSQKYKR

mRRM2
*MGHHHHHHGLEVLFQGP*KQELGRVIHLSNLPHSGYSDSAVLKLAEPYGKIKNYILMRM
KSQAFIEMETREDAMAMVDHCLKKALWFQGRCVKVDLSEKYKKL

Figure 4 Amino acid sequences of matrin 3 RRMs. These amino acid sequences were ordered and used for the experiments. The linker of the two RRM domains is underlined. The His-tag used in the sequence that is later cleaved is italicized. Without the 17 amino acid His-tag, mtRRM is 182 amino acids long, mRRM1 is 83, mRRM1-ICC is 85, and mRRM2 is 84.

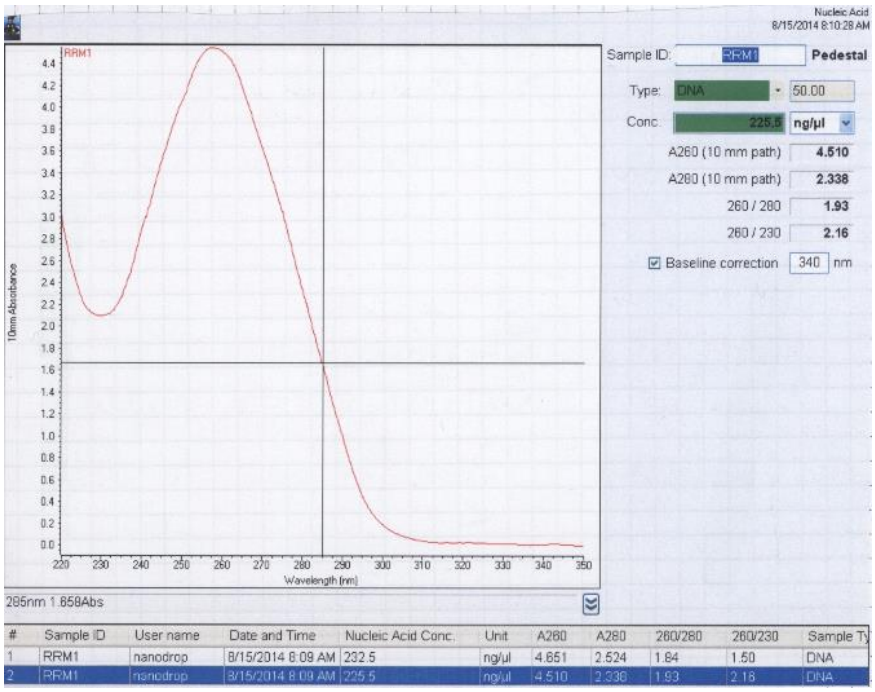


Figure 5 Nucleic acid concentration of mRRM1. The Nanodrop reading can calculate the concentration of the DNA based on the absorbance of the sample and Beer's Law. The number considered in the nucleic acid concentration in ng/µL.

Figure 6 Agarose gel scan of PET 3d vector and mtRRM plasmid. The left most column is the ladder for measuring the density of the sample. The two highest bands are the vectors as they are larger and move through the gel slower. The four lower bands are the vectors for matrin 3 tRRM. These vectors are smaller so they move through the gel faster. The mtRRM plasmid should be in the rectangle, according the ladder used.

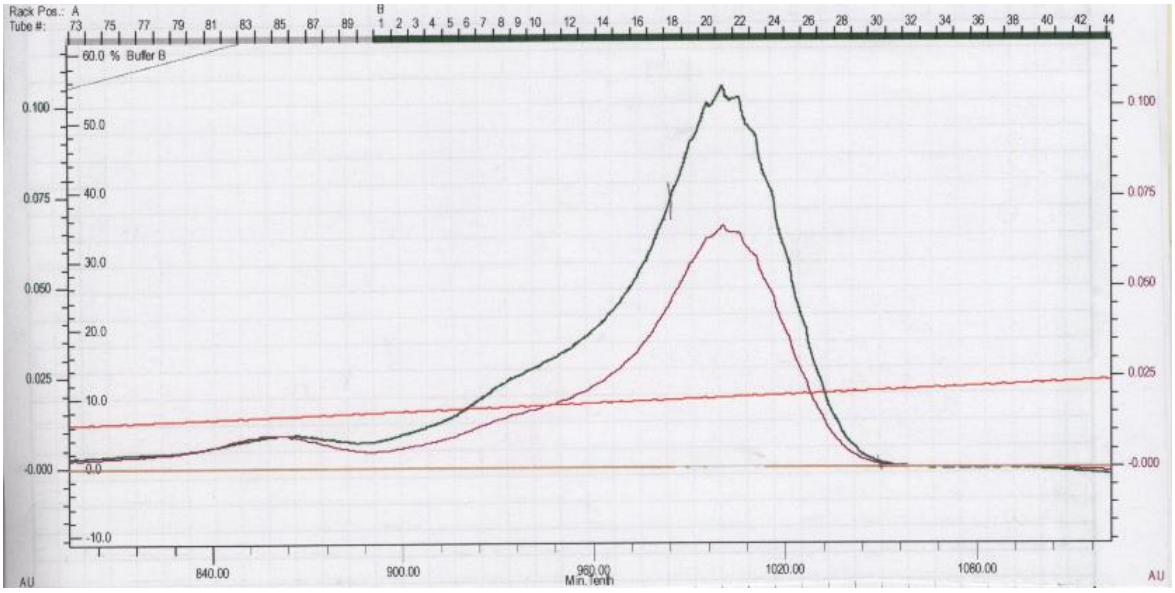


Figure 7 S-column conductivity readout of mRRM2. This figure demonstrated the reading obtained from the Biorad gradiation. The green line, topmost line, represents the absorbance of the sample at 280nm, the wavelength that protein absorbs light. The top axis shows the tube number in which this sample is. Samples are collected along this peak to determine which tubes have the correct protein, as determined by an electrophoresis gel. The tubes that contain the protein are then pooled.

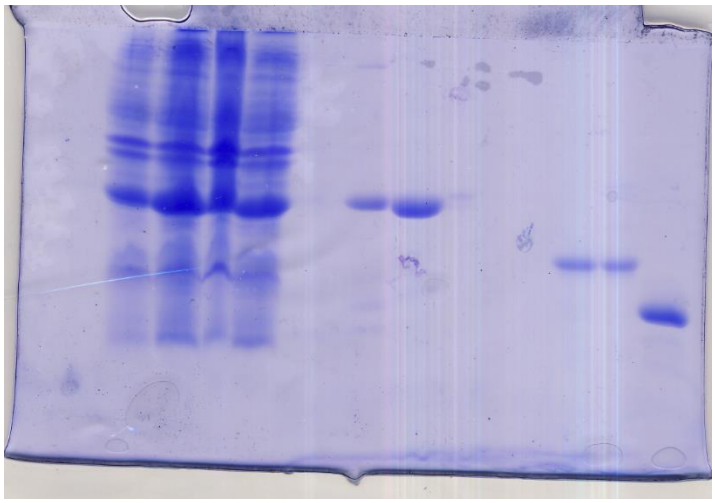


Figure 8 mtRRM gel with samples from the purification process. This 14 lane gel is a sample of the matrin 3 protein purification. The subsequent rows contain the lysis before sonication, the lysing supernatant after spinning, the lysing pellet after spinning, the flow through of the nickel column, the wash flow through of the nickel column, the elution flow through off of the nickel column without urea, the eluate with urea, a sample after the precision protease was added, and the s-column flow through. These samples contained steps where protein should and should not have been present. Sampling at these points helped to monitor the purification and to ensure protein was still present in the samples.

mtRRM
*MGHHHHHHGLEVLFQGP*VETSRVVHIMDFQRGKNLRYQLLQLVEPFGVISNHLILNKIN
EAFIEMATTEDAQAAVDYYTTTPALVFGKPVRVHLSQKYKRIKKPEGKPDQKFDQKQE
LGRVIHLSNLPHSGYSDSAVLKLAEPYGKIKNYILMRMKSQAFIEMETREDAMAMVDH
CLKKALWFQGRCVKVDLSEKYKKL

mRRM1
*MGHHHHHHGLEVLFQGP*VETSRVVHIMDFQRGKNLRYQLLQLVEPFGVISNHLILNKIN
EAFIEMATTEDAQAAVDYYTTTPALVFGKPVRVHLSQKYKR

mRRM1- ICC
*MGHHHHHHGLEVLFQGP*VETSRVVHIMDFQRGKNLRYQLLQLVEPFGVISNHLILNKIN
EAFIEMATTEDAQAAVDYYTTTPALVFGKPVRVHLSQKYKRIKK

Figure 9 Adding in the ICC of mRRM1. When this project first began it was believed that the segment that was cut out in mtRRM to make mRRM1 and mRRM2 was simply the linker between the two segments. This is highlighted. The first three residues of this highlighted segment were added back in to make mRRM1-ICC.

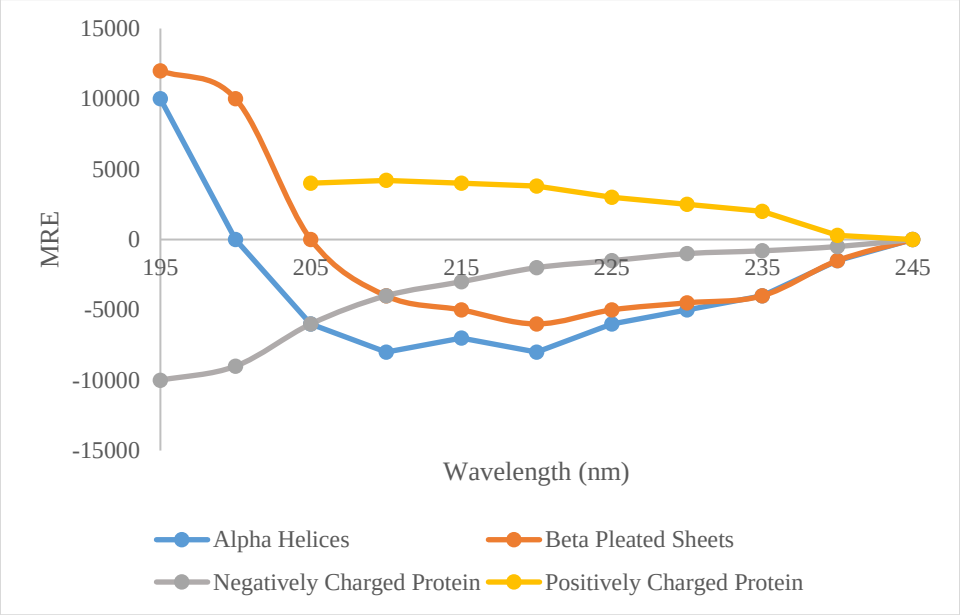


Figure 10 Sample protein CD analysis by MRE. The mass analysis of CD data was done using a computer program called SAVUKA. The resulting graph was used to understand secondary structure. The above graph is a demonstration of the different secondary structures CD results. The alpha helical structure can be inferred from the 222nm/208nm ratio.

Molar residue ellipticity (MRE) is a measure of intensity by which the secondary structure of the protein can be understood. Using the signal obtained by circular dichroism, analysis was performed using the following equation.

$$MRE = \frac{0\ signal}{10 * length\ of\ cuvette * amino\ acid\ number * concentration}$$

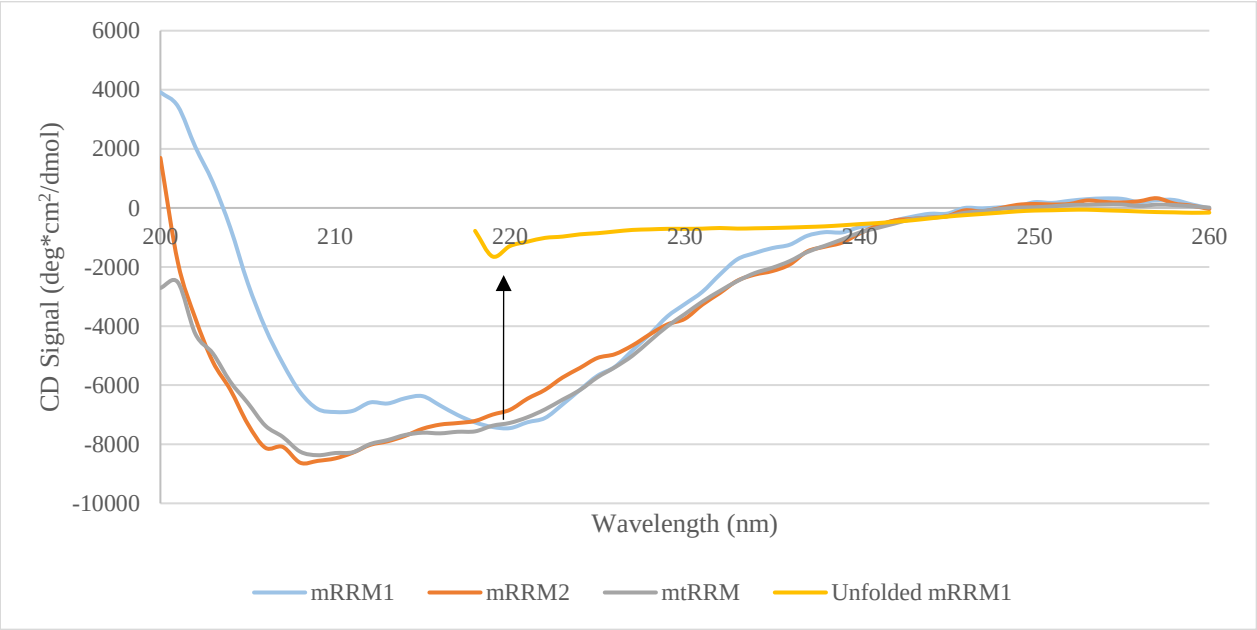


Figure 11 RRMs native and unfolded secondary structure. CD uses a spectrum of light that the sample is scanned across. The ratios between certain wavelengths can be used to determine the presence of secondary structure (see *Figure 10*). With no denaturant, in this case guanidine-hydrochloride, the protein is folded and there are local minima between -6000 and -9000 CD signal. The intensity of this signal depends on the RRM tested. This graph also displays a completely unfolded scan of mRRM1, blue to yellow. This requires denaturant, 7M GDN-HCl. The loss of folded structure reduces the signal as there is no secondary structure left. The scan does not go far beyond 220nm as the signal beyond this point scatters the light too much.

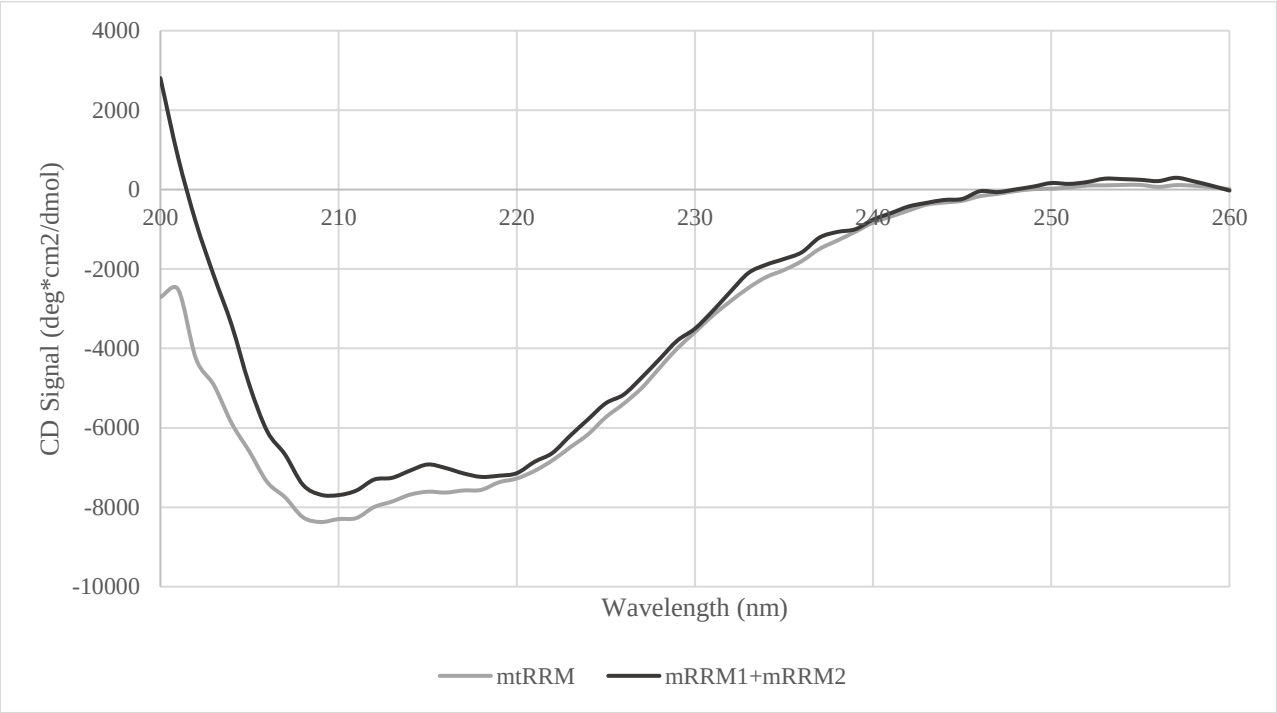
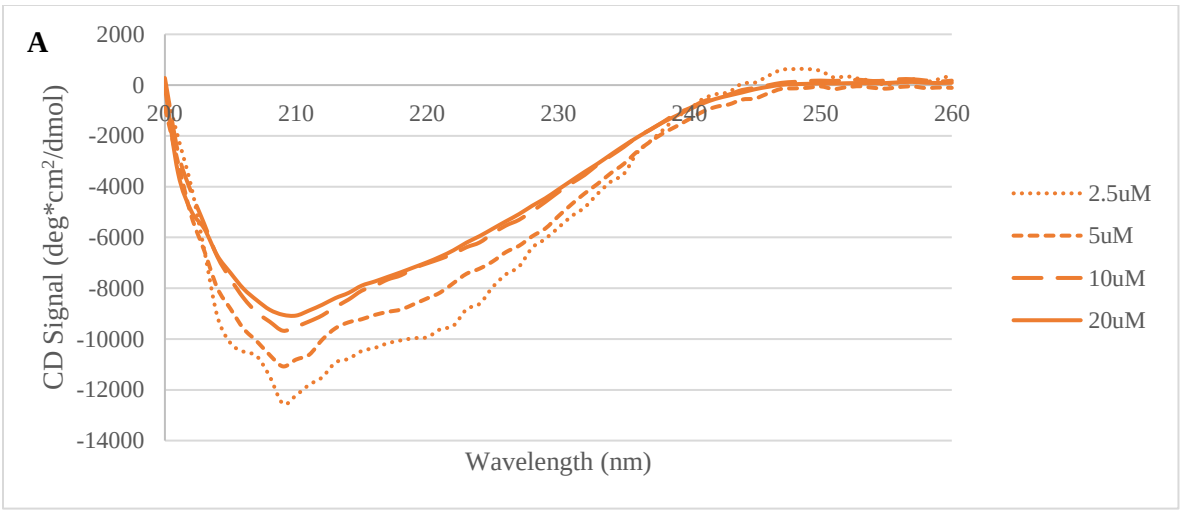
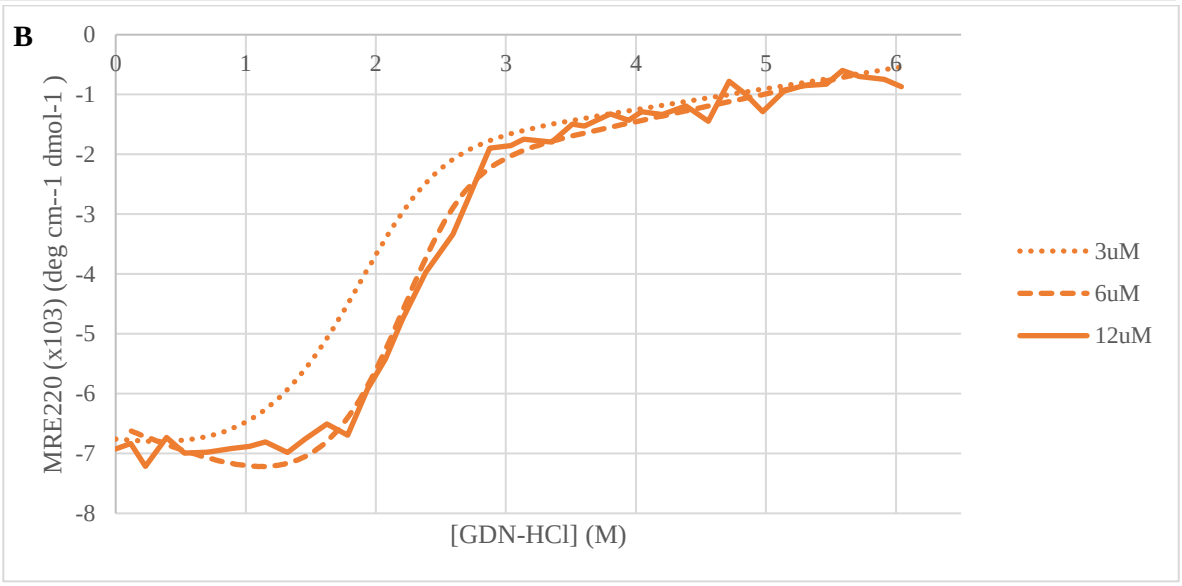


Figure 12 mtRRM CD signal versus the additive mRRM1+mRRM2 signal. This graph demonstrates how the tethered domain of mRRM1 and mRRM2 *in vitro* provides additional structure and stability to the complex in terms of secondary structure as compared to the signals of mRRM1 and mRRM2 added together. The signal of the tethered domain is lower, indicating that it has more secondary structure than the additive complex.


A
2000
0
-2000
-4000
-6000
-8000
-10000
-12000
-14000
200
210
220
230
240
250
260
CD Signal (deg*cm²/dmol)
Wavelength (nm)
2.5uM
5uM
10uM
20uM


B
0
-1
-2
-3
-4
-5
-6
-7
-8
0
1
2
3
4
5
6
MRE220 (x103) (deg cm--1 dmol-1)
[GDN-HCl] (M)
3uM
6uM
12uM

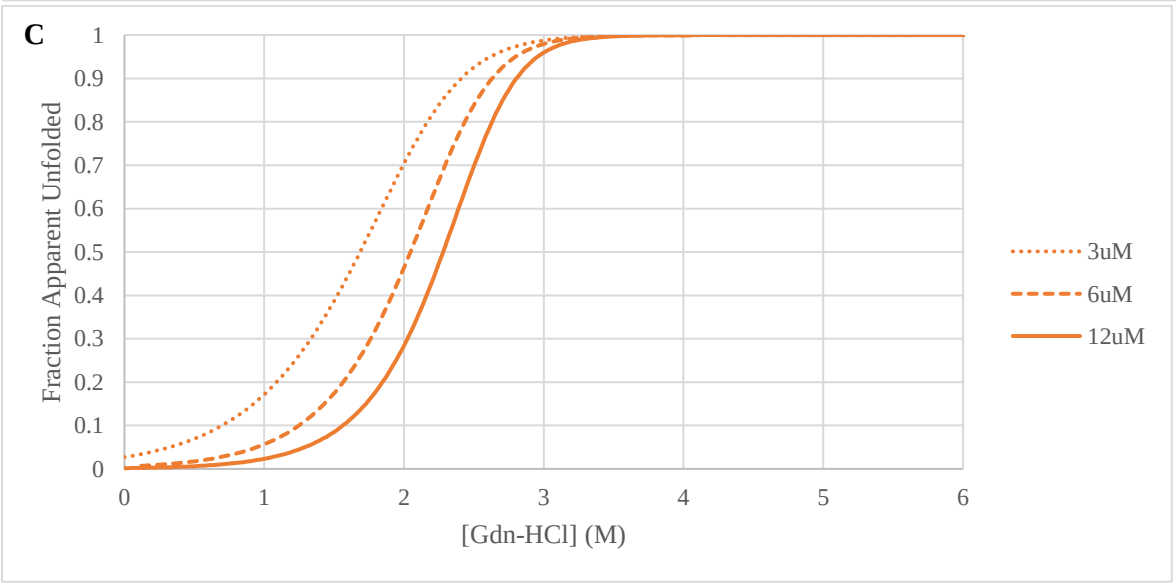

C
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
1
2
3
4
5
6
Fraction Apparent Unfolded
[Gdn-HCl] (M)
3uM
6uM
12uM

Figure 13 mRRM2 concentration analysis from CD signal to FAP. A: The higher the concentration of protein, the higher the CD signal. A test like this is done to determine the least amount of protein required to produce the clearest signal. The data in the graph is contains no denaturant. **B:** The protein best fit a two-state unfolding curve at 220 nm and so the MRE at this point was calculated. **C:** The stability and signal of the samples is determined by the concentration of denaturant needed to unfold the sample. The highest concentration protein is the most stable, as it requires 2.29M GDN-HCl to be at equilibrium with its unfolding process. The concentration dependence suggesting may be an artifact of probing the isolated mRRM2 domain in the absence of mRRM1 rather than suggesting a functional role for dimerization.

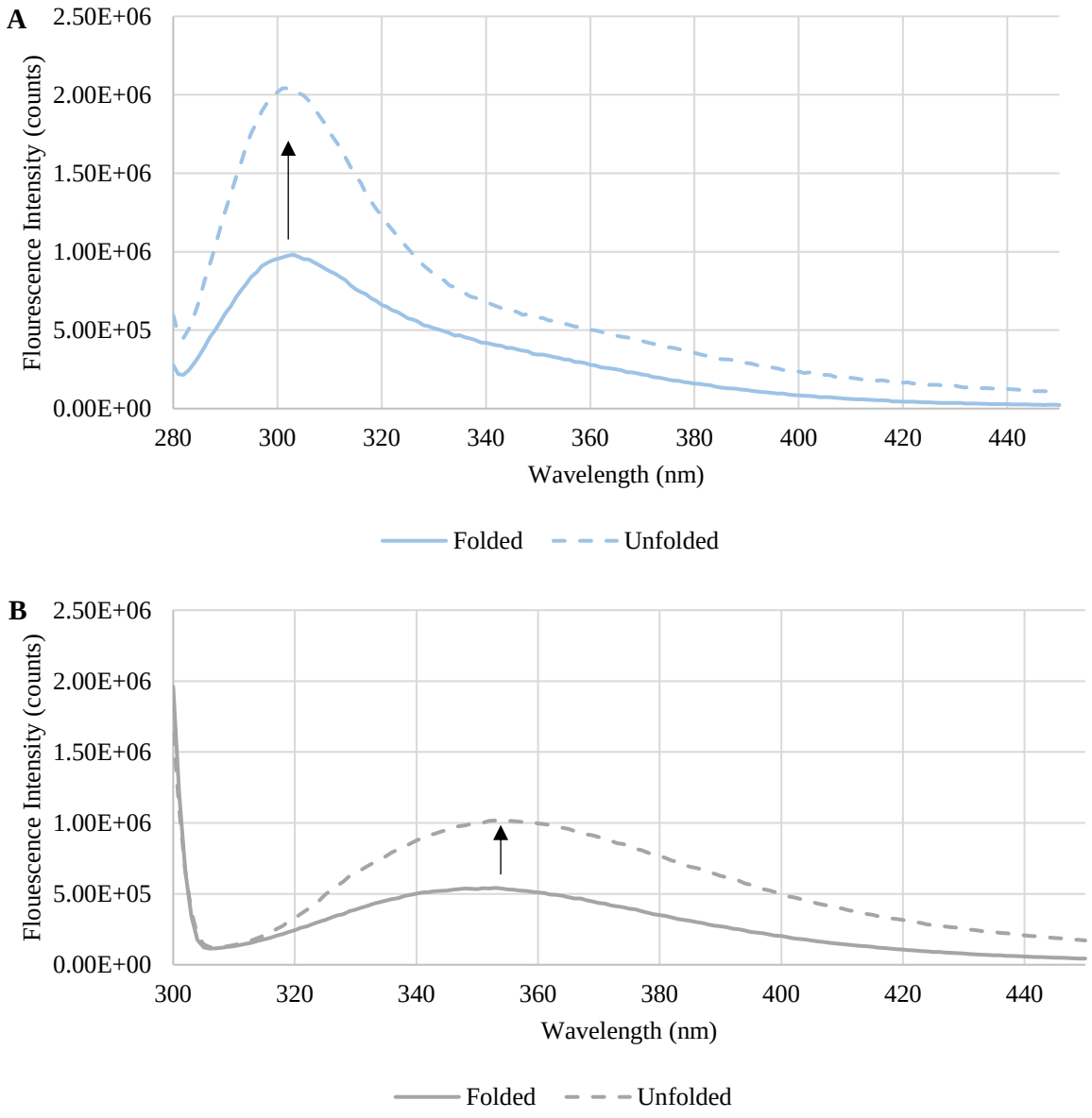


Figure 14 Tertiary structure unfolding of a tyrosine and tryptophan fluorescence excitation. A: mRRM1 was excited at 274nm for its presence of tyrosine. The light was then released at 302nm. Unfolding the motif with guanidine hydrochloride revealed an increase in fluorescence intensity as the Tyr spell this out moved further from each other. **B:** mtRRM excitation occurred at a 295nm to excite the tryptophan, which released an emission at 354nm. In unfolding the motif, the signal increased as the tryptophan moves further from the other fluorescent aromatics.

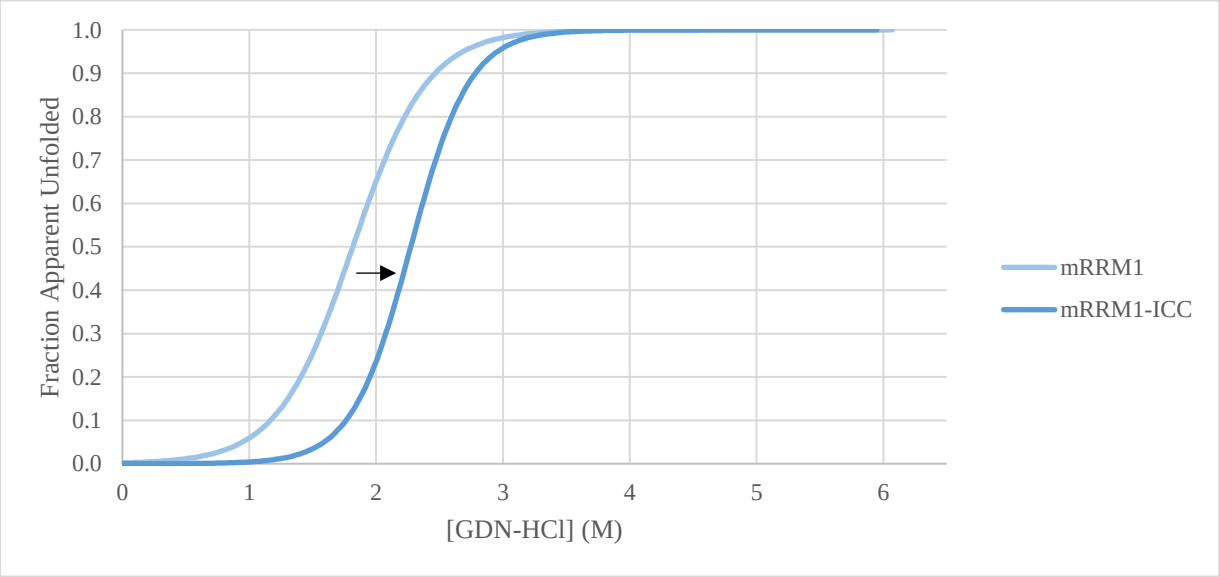


Figure 15 Secondary structure stability of mRRM1 and mRRM1-ICC denaturation. The addition of ICC inducing residues increases the stability of mRRM1 by 1.88kcal (ΔG 3.46 to 5.34). This stability increase can be seen in the shift to a higher concentration of denaturant (1.9M to 2.4M).

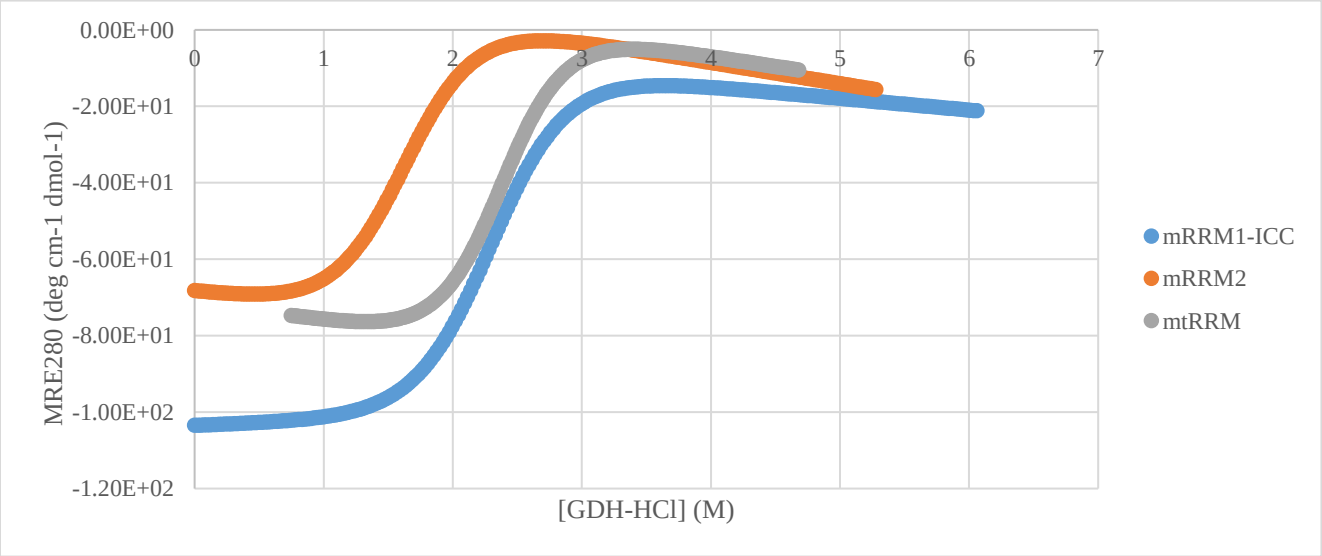
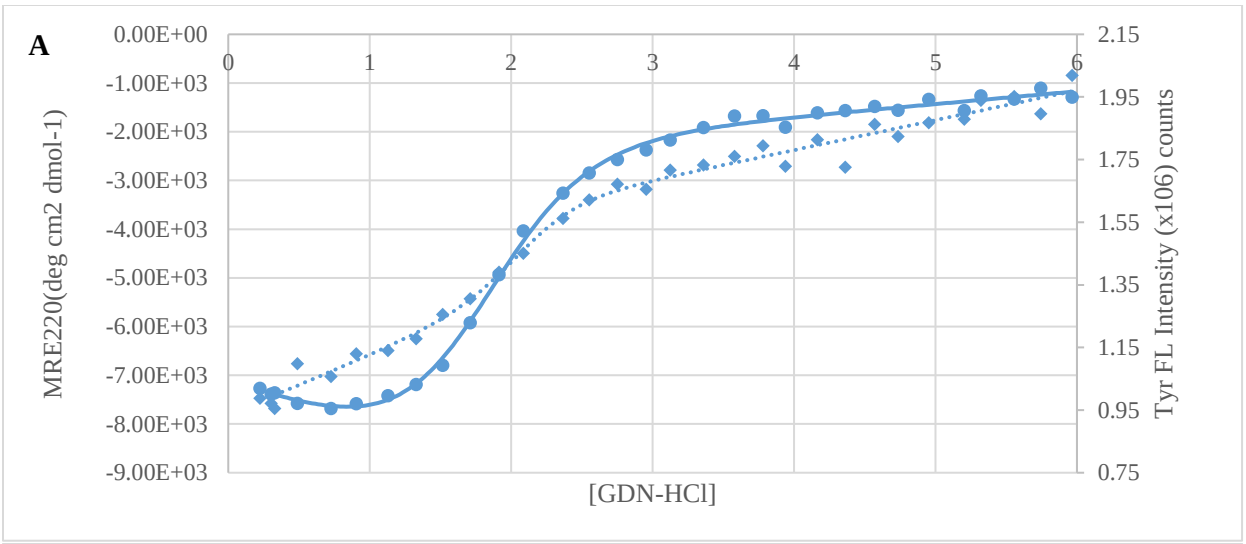
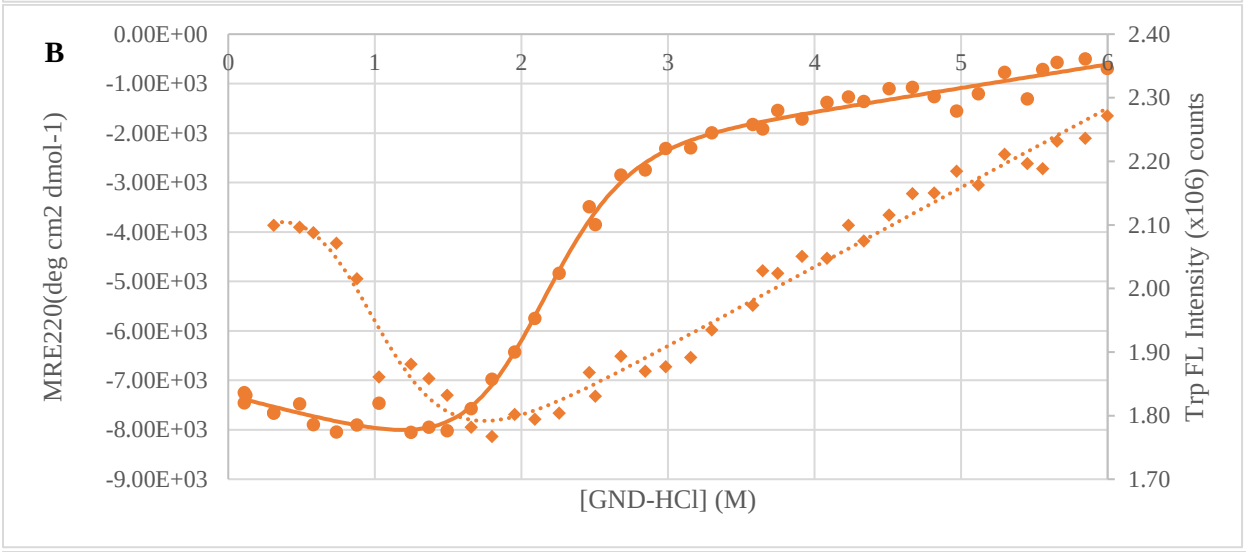


Figure 16 Near-UV CD of matrin 3 RRM disordered domains. Investigating the RRMs individual revealed no intermediate state, fitting a two-state folding landscape. The fluorescence studies did not yield useful results in regards to the intermediate studies.


A
MRE220(deg cm2 dmol-1)
0.00E+00
-1.00E+03
-2.00E+03
-3.00E+03
-4.00E+03
-5.00E+03
-6.00E+03
-7.00E+03
-8.00E+03
-9.00E+03
0
1
2
3
4
5
6
2.15
1.95
1.75
1.55
1.35
1.15
0.95
0.75
Tyr FL Intensity (x106) counts
[GDN-HCl]


B
MRE220(deg cm2 dmol-1)
0.00E+00
-1.00E+03
-2.00E+03
-3.00E+03
-4.00E+03
-5.00E+03
-6.00E+03
-7.00E+03
-8.00E+03
-9.00E+03
0
1
2
3
4
5
6
2.40
2.30
2.20
2.10
2.00
1.90
1.80
1.70
Trp FL Intensity (x106) counts
[GND-HCl] (M)

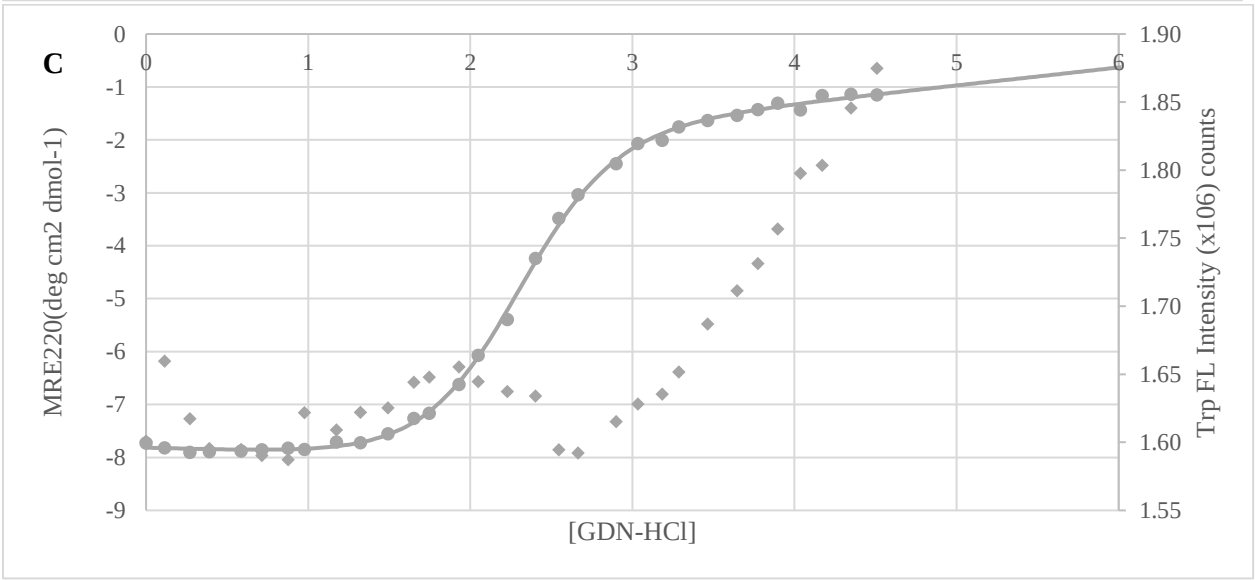

C
MRE220(deg cm2 dmol-1)
0
-1
-2
-3
-4
-5
-6
-7
-8
-9
0
1
2
3
4
5
6
1.90
1.85
1.80
1.75
1.70
1.65
1.60
1.55
Trp FL Intensity (x106) counts
[GDN-HCl]

Figure 17 Far-UV CD and FL comparison to analyze the intermediate state. In the graphs, the Far-UV CD model is represented with a solid line and the data with circles. The FL model is a dashed line (mtRRM does not have a model) and the data is squares. **A:** mRRM1-ICC fits a two-state plot. **B:** mRRM2 fits a two-state plot. FL for mRRM2 is interesting because it does not have a native baseline, indicating that it was a folded state going to an unfolded state. Instead it's this checkmark shape which is interesting, looking like it's already partially unfolded. **C:** mtRRM fits a two-state plot.

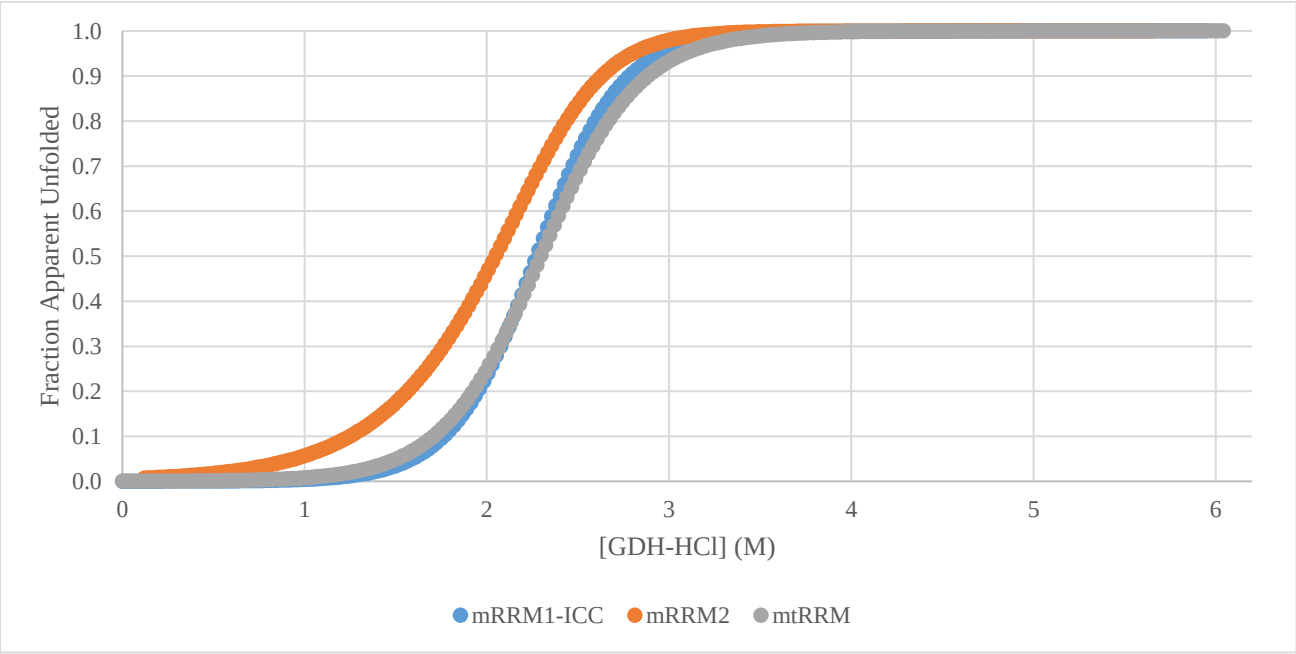


Figure 18 FAP unfolding comparison of mRRMs. mRRM1-ICC is shifted to the right of where mRRM1 initially had FAP signal. Therefore the midpoint is also shifted. All 3 have similar slopes as it unfolds. Tethering the two RRM domains indicates that mRRM1 and mRRM2 may fold coincidentally as the m-value of mtRRM was not the sum of the individual m-values for mRRM1 and mRRM2; nor was it the average.

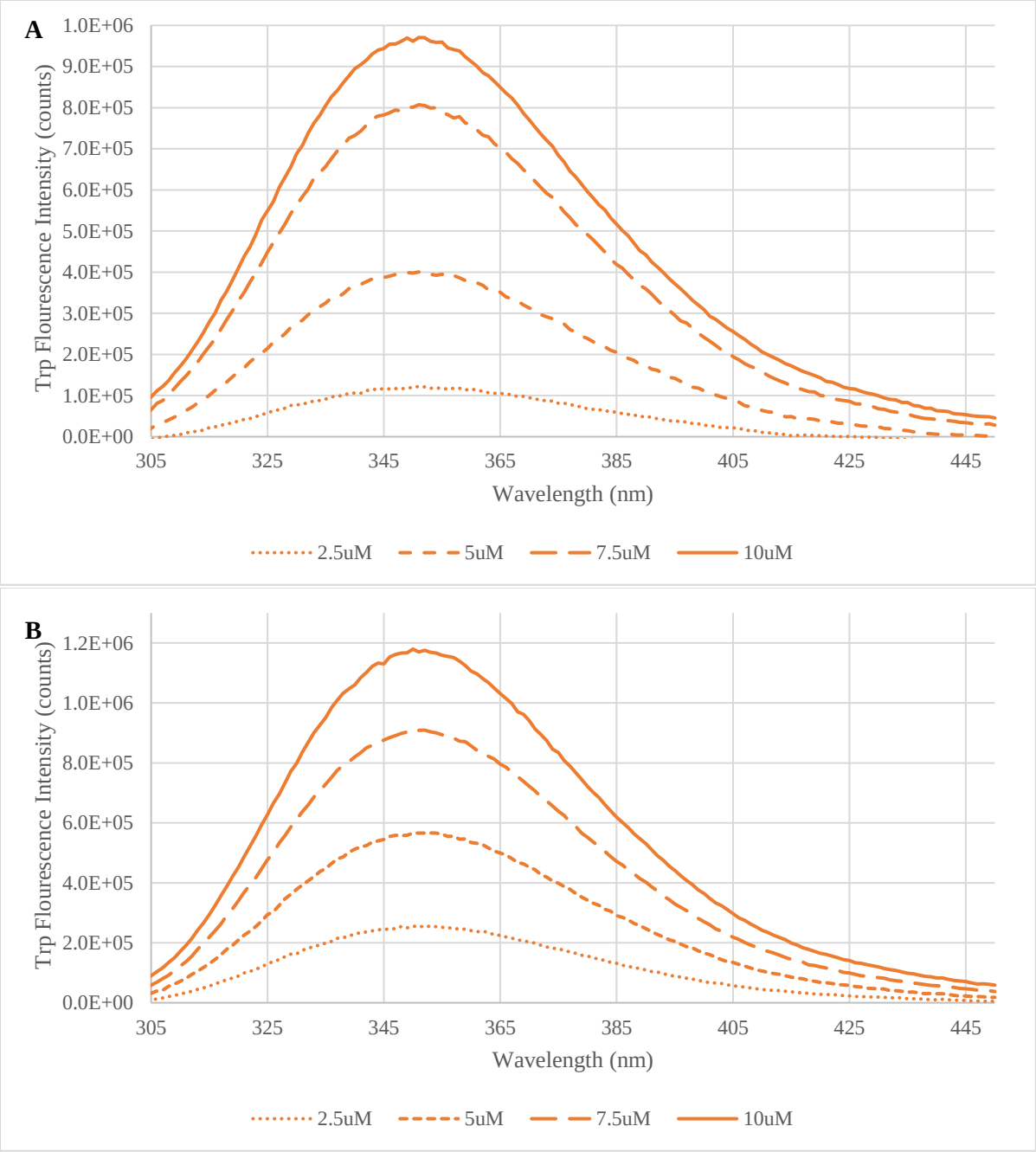


Figure 19 Calcium binding of mRRM2 effects on FL intensity. A: different concentrations of mRRM2 and their respective fluorescence intensities. **B:** Calcium binding lower the intensity; at 2.5uM there is a two-fold difference. Though the intensity of FL was lowered, using calcium buffer inhibited the ability to use CD as the signal scattered too much.

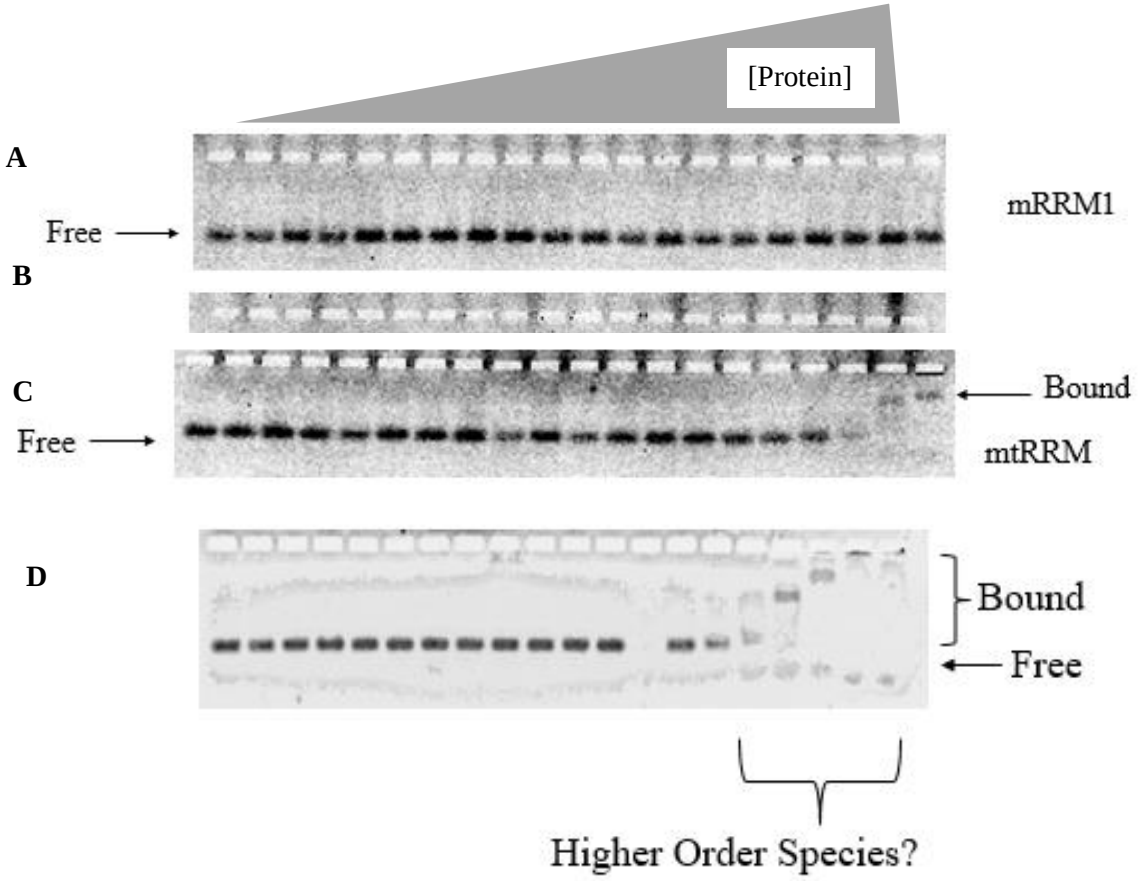


Figure 20 mtRRM RNA binding to 5' - CUUCUC*ACUACU*GCACUUG*ACUAGU*CU - 3'. **A:** mRRM1 and **B:** mRRM2 travelled to the bottom of the lane. This protein is considered free, meaning that it did not bind to the RNA. **C:** mtRRM did have some binding at higher protein concentrations, leading to a larger and slower complex running through the gel. **D:** Replication of this RNA binding assay at higher concentrations of protein before the serial dilution of protein revealed interesting binding behaviors at high protein concentrations. This protein ran lower than the free protein at lower protein concentrations so it suggests the possibility of a higher order species forming in RNA binding.

ACKNOWLEDGEMENTS

Thanks are required for Dr. Jill Zitzewitz as she allowed me to complete this research in her laboratory over the past three years. She trusted one such as myself, who had no skills or experience in lab, to be released unto her research. Also in the lab was (the now Dr.) Brian Mackness, whose impeccable notebooks and directions enabled this research. Turning toward Assumption College, Dr. Michele Lemons should be thanked for her ever present patience even when I was less that forthcoming with all my thesis progress. I would also like to thank the Honors Program as a whole for presenting me with the opportunity to complete this thesis and Professors Laura Marcotte and Kimberly Schandel for serving on my thesis committee. To my roommates who know nothing of science but are interested in who I am as a person, thank you for coming along on this ride. And lastly, thanks to my family. When I was little, I knew I could be anything but I always knew I wanted to pursue my passion. You allowed and nurtured that drive and while I don't know what's to come next, it was this start in life that will keep me moving forward.