Stony Brook University

Structural and Biochemical Analysis of the Hnrnpab Tandem RRM RNA Binding Domain

A Thesis Presented

by

Daniel Richard Catt

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Master of Science

in

Biochemistry and Cell Biology

Stony Brook University

December 2013

Stony Brook University

The Graduate School

Daniel Richard Catt

We, the thesis committee for the above candidate for the

Master of Science degree, hereby recommend

acceptance of this thesis.

Dr. Kevin Czaplinski – Thesis Advisor
Assistant Professor
Department of Biochemistry and Cell Biology

Dr. Steven O. Smith – Second Reader
Professor
Department of Biochemistry and Cell Biology

This thesis is accepted by the Graduate School

Charles Taber

Dean of the Graduate School

Abstract of the Thesis

Structural and Biochemical Analysis of the Hnrnpab Tandem RRM RNA Binding Domain

by

Daniel Richard Catt

Master of Science

in

Biochemistry and Cell Biology

Stony Brook University

2013

Post-transcriptional trafficking of mRNA within the cell provides for spatial regulation of protein expression. *Cis*-acting localization elements known as zip codes located often in 3' untranslated regions (UTRs) of mRNA transcripts play a strong role in directing distribution within the cell through recognition by RNA-binding proteins. Hnrnpab is a protein with putative roles in this process with RNA-binding properties likely directed through tandem RNA Recognition Motifs (RRMs) that form a cooperative RNA Binding Domain (RBD). We purified this Hnrnpab RBD and carried out electrophoretic mobility shift assays (EMSAs) using the zip code region from the ActB gene to demonstrate high specificity binding. In order to determine the structural determinants of this specificity we purified and ^{13}C, ^{15}N-labeled Hnrnpab RBD for analysis by solution NMR to begin determining a structure for the RRMs both with and without high specificity RNA partners. Further characterization of how Hnrnpab recognizes its targets and what structural motifs determine this type of binding will facilitate greater understanding of the spatial regulation of protein products necessary for cellular and organismal development.

Table of Contents

List of Figures

List of Abbreviations

mRNA	Messenger RNA
snRNPs	Small Nuclear Ribonucleoprotein Particles
PAP	Poly-Adenosine Polymerase
UTR	Untranslated Region
ARE	AU-Rich Element
ARE-BP	AU-Rich Element Binding Protein
hnRNPs	Heterogeneous Nuclear Ribonucleoproteins
hnRNAs	Heterogeneous Nuclear RNAs
RBP	RNA Binding Protein
RRM	RNA Recognition Motif
RNP	Ribonucleoprotein Domain Motif
NMR	Nuclear Magnetic Resonance
ABBP-1	APOBEC1-binding protein 1
MEFs	Mouse Embryonic Fibroblasts
RBD	RNA Binding Domain
TBE	Tris-Borate EDTA
Actb	Beta-actin gene

Acknowledgements

I would like to express gratitude for Kevin Czaplinski for his guidance and support throughout my project. In the lab I want to recognize John Sinnamon for his valuable advice and training in a wide array of areas. Steve Smith was more than accommodating in contributing time and resources to the project most especially with donations of heavy isotopes for the NMR samples and, importantly, time on the NMR machines. Martine Ziliox was instrumental in running the NMR samples and is also deserving of praise.

I. Introduction

1. Post-Transcriptional Regulation of mRNA

In the cell, information encoded in messenger RNA (mRNA) provides the intermediate molecular blueprints for functional protein products. Complex transcriptional and post-translational controls have evolved to dynamically respond to environmental changes. However, the unique positioning of information encoded within the alternatively spliced, trafficked, modified, and degraded mRNA macromolecules provides a rich source of post-transcriptional regulation.

1.1 Pre-mRNA Splicing, Processing, and Stability

The existence of protein-coding exons along with non-coding introns in the genes of most eukaryotic organisms creates an initial form of regulation on the information ultimately contained within an mRNA transcript. Recruitment of small nuclear ribonucleoprotein particles (snRNPs) along with other proteins to nascent pre-mRNA strands initiates formation of the spliceosome, which is responsible for the remarkably orderly removal of introns and joining of exons during the production of mature mRNA. The process of splicing regularly involves the possibility of an alternative splicing event. Depending on the gene in question, alternatively spliced products missing either single or many exons or containing unique permutations from exon groups are reliably produced in predictable patterns dictated often by the strength of the splice sites found upstream and downstream of the candidate segment [1].

Other elements of pre-mRNA processing involve co-transcriptional addition of the 7-methylguanosine cap through a triphosphate bridge to the 5’ end of the mRNA and the poly-adenosine extension of the 3’ end by conserved, template-independent poly-adenosine-polymerases (PAPs) [1]. Both of these structures are crucial for mRNA stability within the cell as removal of either modification leads to rapid degradation of the message. Regulating access to these structures provides the potential for a wide range of transcript stabilities as well as the opportunity to change stability upon demand. In addition, initiation of most eukaryotic translation events depends upon these structures, which suggests that their addition indicates the transcript is full-length [2]. Ensuring the

transcript is in the proper reading frame and full-length provides a quality control mechanism on protein synthesis in the cell and prevents potentially deleterious effects of incomplete or incorrect products [2].

The 5' and 3' untranslated regions (UTRs), which are located upstream before the start codons and downstream after the stop codons, respectively, contain RNA sequences that can impart stabilizing or destabilizing effects. Most notably, one class of A-U rich elements (AREs) within the sequence of many 3' UTRs have been shown to recruit deadenylases to degrade the poly-A tail on mRNA transcripts, thus shortening the half-lives of these polymers [3]. For a given transcript, the AREs may associate with multiple different ARE-binding proteins (ARE-BPs) that possess varying roles in mRNA regulation though commonly affect stability [4]. ARE-mediated degradation of mRNAs through the recruitment of ARE-BPs is notably influenced by environmental conditions making regulation of mRNA stability dynamic [4,5]. This observation highlights the important role that RNA sequence elements acting in *cis* play in determining an mRNAs fate. *Cis*-acting elements dictate protein binding partners that will associate with the transcript during each stage of processing and can direct the mRNA to particular locations within the cell. This important aspect of cellular organization is still not fully understood.

1.2 mRNA Localization and Trafficking

Maintaining local levels of a protein in the cell has been the subject of research for some time. It was known that mechanisms within the cell allowed for active transport of organelle cargo [6]. These mechanisms turned out to involve the ATP-dependent unidirectional locomotion of motor proteins like kinesins and dyneins along cytoskeletal protein polymers. It was observed that such mechanisms could transport protein complexes and membrane-bound vesicles or organelles to localized areas of the cell based on association with particular binding partners [6].

However, in dealing with proteins that the cell needs to spatially distribute, localization of encoding mRNA in lieu of the protein product would save transportation costs for the cell [7]. The energy input for moving a single mRNA would theoretically save on subsequent transport for the many proteins translated from the one transcript.

Support for this model was found while probing for mRNAs encoding the actin cytoskeletal proteins [8]. It was found that actin mRNA was preferentially localized to the lammelipodia regions of cultured fibroblasts where actin polymerization is mobilized for cell spreading.

Studies on yeasts, *Drosophila* and *Xenopus* ooctyes, neurons, and oligodendrocytes have indicated that many mRNA transcripts actively define their spatial distribution [9]. Importantly, it was found that both mRNA trafficking and localized stabilization can contribute to the patterning of mRNA transcripts within the cell. Once again, *cis*-acting elements within the transcripts have consistently been tied to guiding the message toward a particular location within the cell for physiologically vital reasons [9]. The *cis*-acting mRNA localization elements within an mRNA have been called zipcodes.

2. RNA Binding Proteins

The *cis*-acting elements within mRNA transcripts are often recognized by *trans*-acting factors in the form of RNA-binding proteins. These proteins are linked to every process mentioned above including alternative splicing, polyadenylation and modification, mRNA localization, and mRNA turnover along with other important roles in nuclear export and translation [10]. RNA binding proteins are therefore a very large and diverse set of proteins that encompass multiple different sub-families, each with homologous protein sequence and structural elements.

2.1 The hnRNPs

Heterogeneous nuclear ribonucleoproteins (hnRNPs) are named for their characteristic association with pre-mRNAs as well as generic heterogeneous nuclear RNAs (hnRNAs) within the nucleus – playing roles in mRNA splicing, processing, and nuclear export. However, the constituent members of the family have been additionally linked to post-export mRNA trafficking, stability, and translation [11]. The cause of this multi-functionality comes largely from conserved structural motifs that are included in varying numbers or with important sequence variations [11].

2.2 The Highly Conserved RRM

One of the highly conserved RNA binding domains first identified in the hnRNPs but widespread in RNA-binding proteins in general is the RNA recognition motif (RRM). A generic, single RRM is made up of a single β-sheet consisting of four antiparallel β-strands against which two α-helices are stacked (Figure 1). Within the consensus RRM, β-strands 1 and 3 are most highly conserved and each of these consists of predominantly hydrophobic stretches of 6-7 residues referred to as ribonucleoprotein domain motifs 1 and 2 (RNP1 and RNP2). Base-stacking interactions that provide part of the binding capacity of the RRM are made in the hydrophobic core made up of residues within the RNPs. Additionally, basic amino acids located proximal to the RNPs make important connections to the phosphate backbone.

There is variety in functionality among proteins containing RRMs that comes from several sources. In many cases, proteins will contain multiple RRMs that can either work in tandem or as separate, independently functioning domains. In some cases, each RRM will have unique properties that cause it to bind nucleic acids in novel ways [12]. For proteins whose RRMs work in tandem, the peptide that links the two domains

Figure 1 A two-tandem RRM domain. Purple arrows show beta strands forming a single sheet binding site with red alpha helices providing structural support.

can confer unique conformations of the RRMs that is important for the protein to function. Other portions of the protein outside the RRMs can also contribute to the function of the protein as well. Importantly, RRMs can also mediate protein-protein interactions that are coupled to the domain's nucleic acid binding [12].

2.2.1 Structural Studies of RRM-Containing Proteins

At least a few dozen structures have been solved for RRM-containing proteins both by Nuclear Magnetic Resonance (NMR) and X-ray crystallography. In some cases, structures have been determined in the presence of bound nucleic acids to clarify the exact location of binding. The results of these studies have revealed a remarkable diversity in binding action for these proteins.

The hnRNP A1 protein is a prototypical two-tandem RRM containing protein. The structure of the tandem RRMs has been shown to represent a highly cooperative relationship between the two domains, which are separated by a short, structured linker region [13]. Aspartate and arginine residues located on the second α-helix within the RRMs forms a salt bridge that binds the two together in a head-on fashion. This orientation creates a discontinuous antiparallel-binding platform that has been shown to form loops in bound nucleic acid strands in other proteins, including those with different RNA-binding elements such as the K-homology domain [13,14]. Looping of RNA has been linked to involvement in the regulation of alternative splicing, but in the case of hnRNP A1 this looping capability has not yet been shown [13].

Another two-tandem RRM hnRNP, hnRNP D/AUF1, possesses two RRMs whose functions in relation to each other are not fully understood [15]. AUF1 has been linked to binding AREs in mRNA decay pathways as well as unwinding guanosine-tetraplexes formed at terminal telomeric DNA repeats [15,16]. The structure of each of the RRMs has been solved individually. These structures yield very specific residue interactions to bases within telomeric repeats in a single-stranded context [15]. The researchers used this structural information to define the unwinding activity of the RRM – the simplistic rationale with which many similar structural studies have been performed. However, the conservation of the tandem RRMs, as well as the linker

between them, suggests that the two domains work simultaneously or cooperatively to form a single binding site.

Other structural studies into a variety of proteins have yielded many alternative methods for two RRM interactions. The RRMs of HuD, Hrp1, and Nucleolin have been shown to orient independently prior to nucleic acid binding, upon which they form a continuous binding platform [13]. Additional studies on other multi-RRM proteins reveal a multitude of binding behaviors such that interactions between the individual domains can be binding dependent or independent and in some cases no interactions occur in either state [13]. The extent of these successive studies illustrates such a great diversity in RRM-mediated nucleic acid binding that it is difficult to predict how an individual RRM protein will behave from its sequence alone.

3. Hnrnpab

The protein on which this work is focused is Hnrnpab, a conserved RNA binding protein. The protein is conserved among eukaryotes, orthologous proteins are found in vertebrates (such as mice, rats, frogs), and a similar homologous protein is found in fruit flies (called Squid). The RRMs are the most highly conserved region of the protein, and the Hnrnpab RRMs are paralogous to those of the AUF1 family of proteins and also more distantly related to the hnRNP A1 family [17] (Figure 2).

Figure 2 .Schematic of the two Hnrnpab isoforms

3.1 Isoforms and Sequence Characteristics

Full-length isoform is Hnrnpab1 is a 36.2 kD protein with two RRM motifs and a C-terminal predicted ATP/GTP binding motif, although no other associated ATP/GTP hydrolysis motifs are present and neither ATP nor GTP binding has ever been demonstrated (Figure 3). An alternative splicing event excludes the seventh exon to

yield the 30.6 kD Hnrpab2 isoform. Work within the lab has shown more promiscuous binding character of the Hnrpab2 isoform and the nature of this apparent difference is of interest since the tandem RRM domains are identical. The Hnrnpab1 specific exon is not within the RRMs, however, so differences in binding and specificity between these isoforms may be due to the influence of this additional peptide on the activity of the RRMs. If so this would represent a novel mechanism of RNA binding regulation.

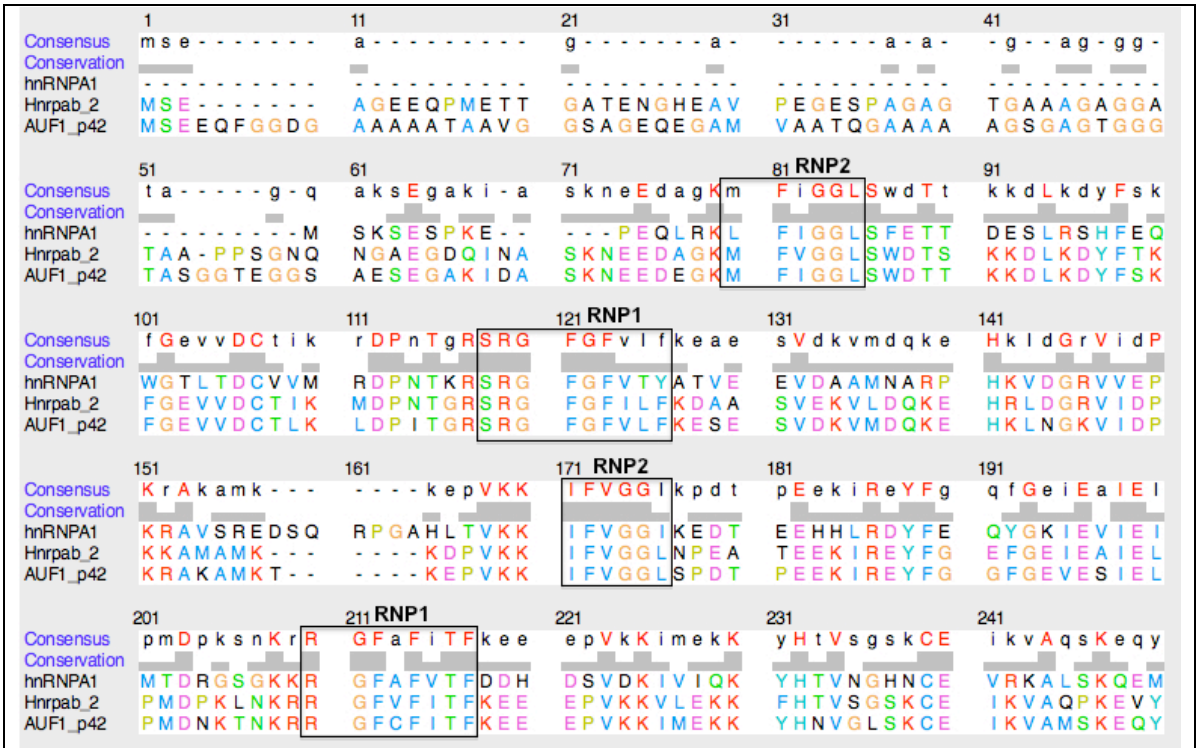

Figure 3 Alignment of tandem RRM domains from hnRNP A1, Hnrpab, and AUF1 showing heightened areas of conservation in RNP motifs at putative RNA binding sites. Boxes outline paradoxically numbered conserved RNP motifs in the first and second RRMs

3.2 Hnrnpab is an Actb Zipcode Binding Protein

Several studies have linked Hnrnpab function to processes from transcription and splicing to localization and translation [18]. Stronger evidence from studies of the frog ortholog 40LoVe and the *Drosophila* homolog Squid have shown it to function in localization of mRNA transcripts of Vg1 and grk, respectively, during development [19,20]. In ongoing work in the Czaplinski laboratory, Hnrnpab1 was purified by its ability to specifically recognize the *cis*-acting zipcode mRNA localization element from Actb

mRNA, which codes for the highly conserved cytoskeletal protein β-actin. Actb mRNA distribution is altered in mouse embryonic fibroblasts (MEFs) from Hnrnpab$^{-/-}$ mice while distribution of the mRNA encoding the γ-actin protein isoform is unaffected, and re-expression of Hrnpab1 restores normal Actb distribution. Biochemical description of the interaction between Hnrnpab1 with the Actb zipcode is an important step in understanding the role of Hnrnpab1 in the mechanism of Actb mRNA localization at the molecular level. Furthermore these studies will provide the basis to understand how the exon 7 sequence functions to regulate the Hnrnpab isoforms binding specificity.

II. Materials and Methods

1. Expression of ^{15}N-Labeled Hnrnpab RNA Binding Domain (RBD) for NMR

An approximate 20 kDa segment of Human Hnrnpab consisting of the RRMs and linker region we will call the RBD. This domain was inserted into pET-22 vector with a 6 Histidine tag for Ni-NTA purification (provided by Jeffrey Chao, Albert Einstein College of Medicine) and transformed into *E. coli* strain DH5α and subsequently miniprepped then diagnostically digested to test for correct plasmid (Figure 4 shows amino acid sequence of construct). This vector was transformed into expression strain Rosetta™ DE3 for one liter growth in M9 media (8.6 mM NaCl, 22 mM KH_2PO_4, 18.4 mM $^{15}NH_4Cl$, 47.7 mM Na_2HPO_4 (heptahydrate), 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 20 mL 20% glucose; MilliQ H_2O to 1 liter). One liter inoculated with cells was grown at 37 ˚C and 250 rpm until optical absorbance at 600 nm reached 0.6. Expression was induced with 0.1 mM IPTG and cells were incubated at 23 ˚C and 250 rpm over night (about 16 hours).

```
  1- MHHHHHHGENLYFQGGSKMFVGGLSWDTSKKDLKDYFTKFGEVVDCTIKM
 51- DPNTGRSRGFGFILFKDAASVEKVLDQKEHRLDGRVIDPKKAMAMKKDPV
101- KKIFVGGLNPEATEEKIREYFGEFGEIEAIELPMDPKLNKRRGFVFITFK
151- EEEPVKKVLEKKFHTVSGSKCEIKVAQPK
```

Figure 4 Sequence of Human Hnrnpab RBD as expressed

Cells were pelleted at 10,000 *x g* at 4 ˚C and resuspended in lysis buffer (1X PBS pH 7.4, additional 370 mM NaCl, 5 mM imidazole, 0.5 mM PMSF). Lysis was performed by sonication on ice and crude lysate was cleared by centrifugation 15,000 *x g* at 4 ˚C. The soluble fraction was added to 1 mL lysis buffer-equilibrated Ni-NTA resin in 15 mL Falcon tube and left at 4 ˚C on rotary mixer for 4 hours. Resin was sedimented out and resuspended in wash buffer (1X PBS pH 7.4, additional 500 mM NaCl, 5 mM imidazole, 0.5 mM PMSF) and mixed for 1 hour. Subsequent washes occurred on column to 20 column volumes. Elution buffer (1X PBS pH 7.4, additional 500 mM NaCl, 250 mM imidazole, 0.5 mM PMSF) was titrated in and multiple 0.5 mL fractions were taken while protein concentration was monitored by Bradford assay.

Elution fractions were run on 15% polyacrylamide:bis-acrylamide (at 37.5:1) SDS-PAGE gel and stained with 0.1% Coomassie Blue in 50% methanol and 10% acetic acid.

1.1 Mass Spectrometry of Purified RBD

To check for full-length protein, 1 µg of pure protein was run on an SDS-PAGE gel and stained as before. The gel was delivered to the Stony Brook University Proteomics Center where the bands were cut out and digested for MALDI-TOF/MS analysis using a Voyager-DE *STR* instrument (Applied Biosystems).

1.2 ^{15}N-HSQC NMR with Labeled RBD Sample

The purified sample was concentrated to approximately 1 mM and subsequently dialyzed against 0.1 mM potassium phosphate buffer at pH 7.6 and was run at 30 ˚C for one week on the 700 MHz (Bruker Instruments) machine in Stony Brook University Structural Biology NMR Core Facility. HSQC was collected to determine suitability of construct for structural determination

2. Double-Labeling of Hnrnpab RBD with ^{13}C and ^{15}N for 3D NMR Experiments

Expression was conducted identically to ^{15}N-labeled protein purification with the exception that 2 g of U-^{13}C-glucose was added to M9 media in lieu of 2 mL 20% unlabeled glucose.

2.1 3D NMR Experiments for Structural Determination of Hnrnpab RBD

The sample was run on a 850 MHz machine in Stony Brook University Structural Biology Core NMR Facility and the following data sets were collected: N-HSQC, CACBCONH, NHCACB, HAHBCONH, and C-HSQC (aliphatic and aromatic). N-HSQC, CACBCONH, and NHCACB were analyzed in Sparky and peaks were picked and lists were uploaded to the MANI PINE Server (http://pine.nmrfam.wisc.edu/) for first pass analysis. Subsequently, manual peak assignments were completed.

3. Electrophoretic Mobility Shift Assays with Hnrnpab RBD and Actb 3' UTR

3.1 Preparation of Hnrnpab RBD Protein Sample

Samples of Hnrnpab RBD were purified as before but without heavy isotopes in growth media. Purified samples were dialyzed against storage buffer (100 mM Tris-Cl, 50 mM NaCl, 50% glycerol, 1 mM DTT) to a concentration near 1 mM.

3.2 Radio-Labeling of Oligonucleotide Probes

Probes were prepared by run-off transcription reactions from plasmid pGEM-5Z containing the complete 3' UTR of Actb. Plasmid was digested with FspI to cut at the end of the first 54 nucleotides of the 3' UTR that corresponds to the putative zip code. As a control probe, the 154 nucleotides 3' of the zip code that is not expected to bind Hnrnpab was produced. Phage T3 RNA polymerase (Roche) was used for *in vitro* transcription reaction using 0.5 µg plasmid template with 0.667 µM α-32P-labeled GTP (Perkin-Elmer) along with 0.5 mM ATP, CTP, and UTP and 50 µM unlabeled GTP. Transcripts were Turbo DNase I (Ambion) doubly phenol/chloroform extracted, purified from unincorporated nucleotide by Micro Bio-Spin® 6 (BIO-RAD), and ethanol precipitated to concentrate. A few microliters of concentrated transcript was boiled with RNA loading dye (with formamide) and loaded onto 15% polyacrylamide: bis-acrylamide (at 37.5:1) 8 M urea denaturing gel. Gel was transferred to film cassette and exposed to film with intensifying screen for 2-4 hours. The film was subsequently developed.

Concentration of transcript was estimated by scintillation counter (Wallac DSL 1409) using known amount of dilute, labeled nucleotide as reference. Concentration of incorporated labeled nucleotide was transformed using ratio of labeled to unlabeled in transcription reaction to yield total guanosine, which was used to calculate molarity based on guanosine content per transcript.

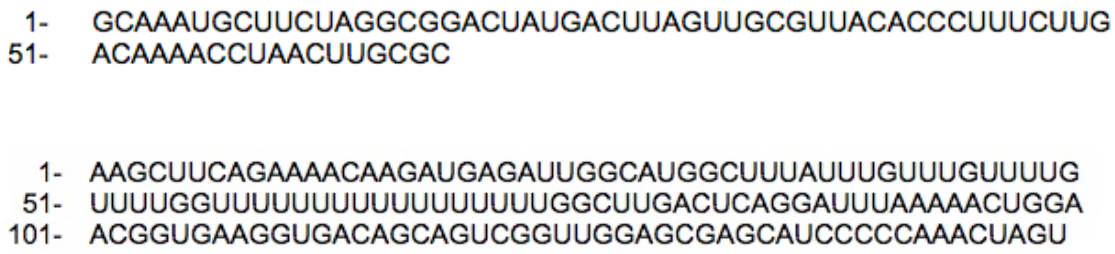

Figure 5 Sequences for RNA transcripts synthesized as radiolabeled probes. The transcript containing the first 54 nucleotides is at top while the middle 150 of the zipcode is at bottom

3.3 Electrophoretic Mobility Shift Assay

3.3.1 Binding Conditions

Binding buffer was prepared at 2x concentration (20 mM Tris, 100 mM NaCl, 0.2 mM EDTA, 20% glycerol, 100 µg/mL heparin, and 0.2% IGEPAL CA-630) and adjusted to pH 7.5. Protein was diluted to working concentrations and added to binding mixture with approximately 100 pM of labeled probe in 10 µL reaction and equilibrated for 45 minutes to 1 hour.

3.3.2 Gel Electrophoresis

Native electrophoresis gels (4% polyacrylamide: bis-acrylamide at 37.5:1, 1X Tris Borate EDTA (TBE) electrophoresis buffer at pH 7.5, and 0.5% glycerol) were cast between glass plates to 1 mm thickness and 13 cm length. Gels were pre-run for 30 to 45 minutes before addition of samples. 2 µL of Bromophenol Blue:Xylene Cyanol loading dye with 50% glycerol was added to the binding reactions and mixture was transferred in total onto native gel. Electrophoresis was performed with 1X TBE buffer in

cold room at 4 ˚C and run at 25 milliamps. Upon completion of the run, the gel was removed from glass plates and transferred to Whatman chromatography paper to be dried at 80 ˚C for approximately 1 hour under vacuum using BIO-RAD Model 583 Gel Dryer in conjunction with Savant Gel Pump GP110.

After drying, gel was exposed to Super Resolution type phosphor screen (medium size; Perkin-Elmer) for 24 hours at room temperature and visualized using a Perkin-Elmer Cyclone® Plus Storage Phosphor System.

III. Results

1. **Expression of ^{15}N-Labeled Hnrnpab RBD for NMR**

The protein was stably expressed in culture and approximately 20 mg of pure RBD was obtained from the 1 L culture. A gel showing the elution fractions from the Ni-NTA column is shown in Figure 6 where it can be observed that multiple bands persist in the purified fractions. In Figure 7, an aliquot from the dialyzed and concentrated RBD demonstrates a persistent doublet pattern observed when running the sample in a denaturing gel. The upper band shown in the pure RBD lane seems to be running near the 20.5 kDa predicted size of the protein while the lower is estimated to be nearer to 16

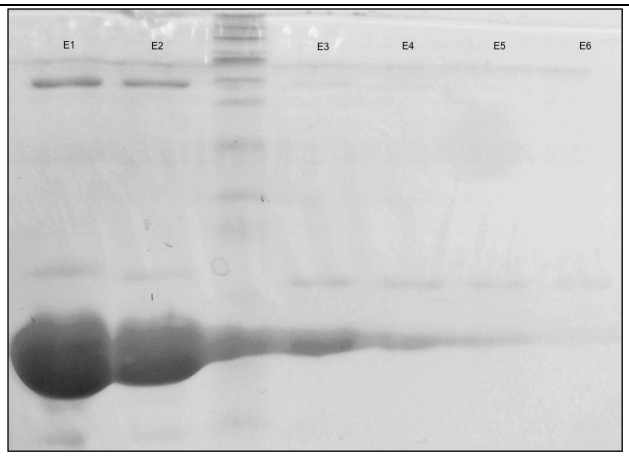

Figure 6 Elution fractions containing substantial protein content (E1-6) from expression and purification. Thick band corresponds to pure RBD with ColorPlus™ Prestained protein ladder in lane 3.

or 17 kDa. Higher order homo-multimeric complexes of the purified protein are often visible near 40 kDa, and even higher. This result is representative of multiple gels with stringent treatment of the sample prior to loading and during the run.

This observation in previous expressions of unlabeled protein prompted addition of broad-spectrum protease inhibitors to all buffers in order to treat for specific degradation of the protein though no change was seen in response to this.

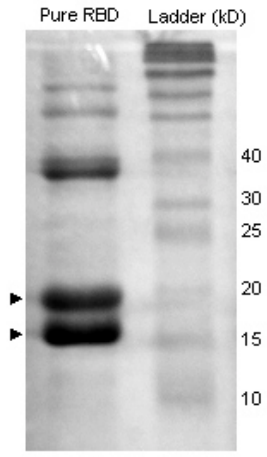

Figure 7 Aberrant Hnrnpab RBD migration pattern in SDS-PAGE gel with upper and lower band (arrows) indicating either degradation, truncation, or alternative conformations resilient in denaturing conditions

1.1 Mass spectrometry of Purified Hnrnpab RBD

The spectra and analysis obtained from the Proteomics Center indicated that the 20.5 kDa and 17 kDa bands seen in our gels were the same product based on the digestion pattern as observed by mass spectrometry. The results also confirmed that the bands corresponded to our fully-tagged RBD construct. Figure 8 shows the spectra for the upper and lower bands that are visually and empirically highly identical. As an additional note, testing on this ^{15}N-labeled sample also indicated that we had achieved high heavy isotopic incorporation in the sample, which is helpful for strong NMR results.

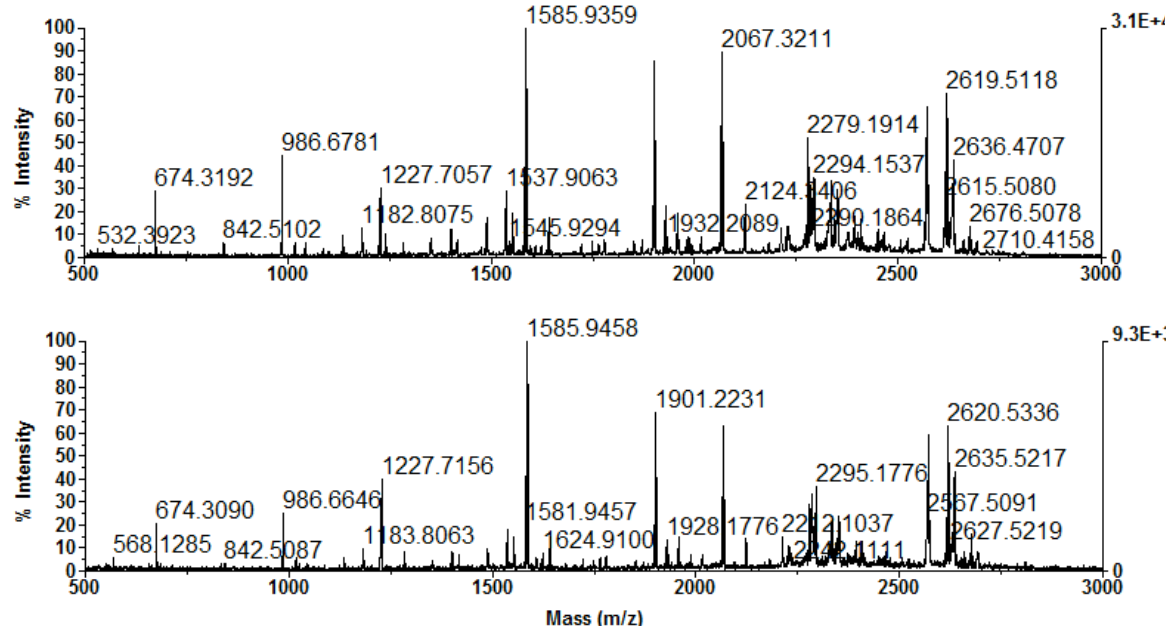

Figure 8 Mass spectrometry results showing identical spectra for upper and lower bands (top and bottom, respectively).

1.2 ^{15}N-HSQC NMR with Labeled RBD sample

The ^{15}N-HSQC spectra showed strong, well-defined peaks that seemed to indicate a strong consensus structure is present in our Hnrnpab RBD sample. The spectrum in Figure 9 was acquired at 30 ˚C is depicted with some clusters of peaks present. On the vertical axis the nitrogen chemical shifts are plotted between 100 and 130 ppm. The particular chemical shift is dependent on the magnetic field strength of the NMR instrument and is sensitive to the chemical environment that surrounds the nucleus being measured. Along the horizontal axis the proton chemical shifts are plotted. The use of two dimensions (i.e. ^{1}H and ^{15}N) allows isolation of peaks representing single residues in the sample. Here, fewer clear peaks are observed compared to the actual number of residues in the protein, likely due to overlap.

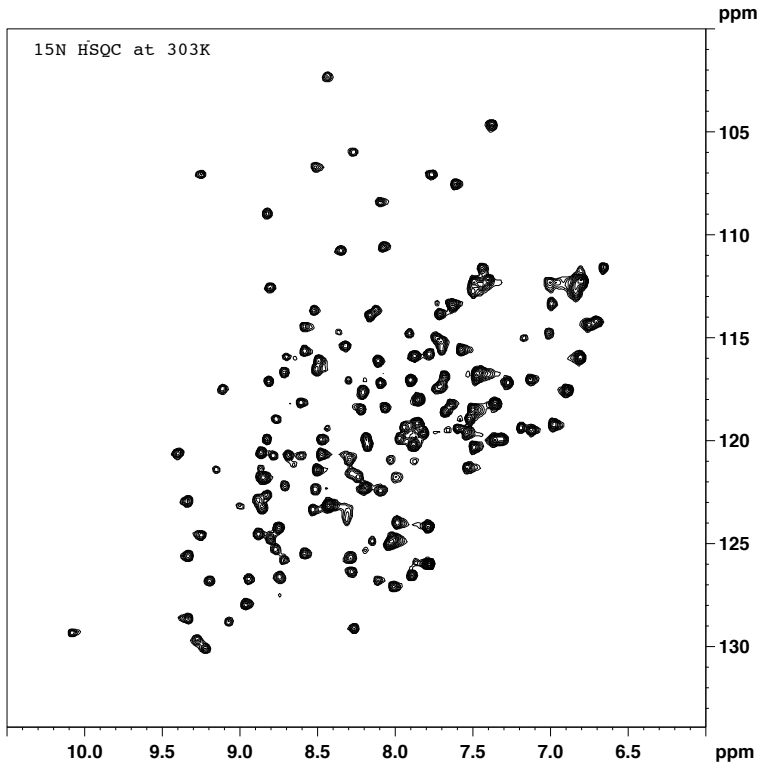

Figure 9 ^{15}N-HSQC of pure Hnrnpab RBD with nitrogen chemical shifts plotted on the ordinate against proton chemical shifts.

2. Double Labeling of Hnrnpab RBD with ^{13}C and ^{15}N for 3D NMR Experiments

2.1 3D NMR Experiments for Structural Determination of Hnrnpab RBD

Upon preparation of a new sample, the RBD was run on the 850 MHz magnet, which resulted in the ^{15}N-HSQC shown in Figure 10, which is visually identical to the HSQC collected previously with strong and well-spread peaks. In some areas of the spectra, clusters of peaks are better resolved, but the overall pattern appears the same. In order to turn this population of peaks into a structure, experiments that provide a third dimension were performed to assign particular backbone residues to each peak. Such 3D experiments include the ^{13}C heavy nucleus and a utilize a series of electromagnetic pulses to generate information about the carbons associated with each amide nitrogen

and proton in the peptide bond, tracing through bonds to the beta carbons to identify amino acid side chains.

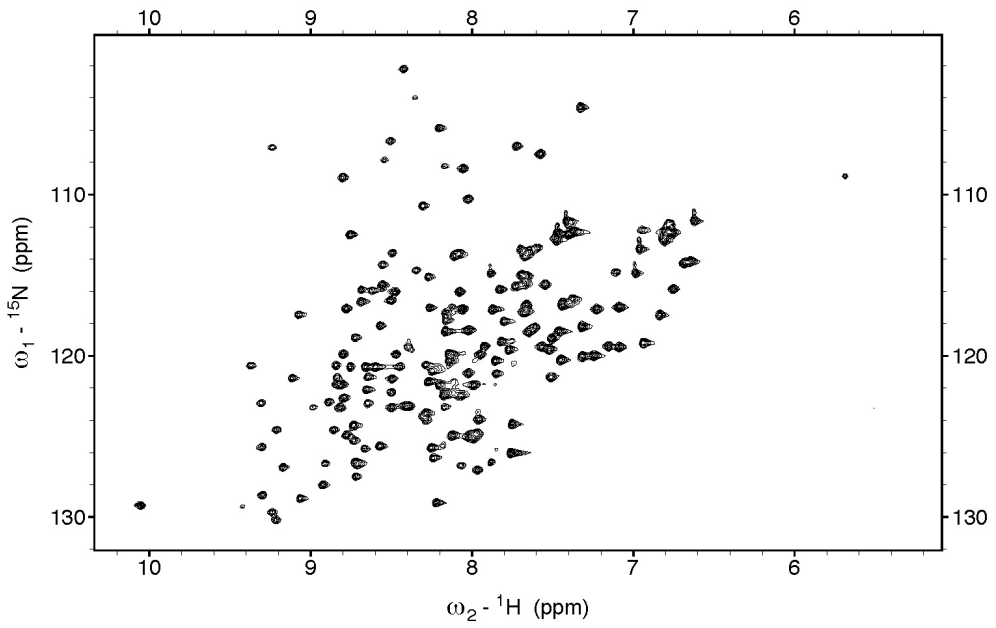

Figure 10 ^{15}N-HSQC for ^{13}C/^{15}N-Labeled Hnrnpab RBD

In Figure 11, a sample view of a CBCACONH experiment window in the Sparky software is shown. In this experiment, the alpha and beta carbons are identified that are N-terminal in relation to a particular amide proton-nitrogen pair. Using these data and the HNCACB experiment, which provides information about both the previous and current residue, enables a sequential “walk” down the peptide backbone. A sample view at the position in the HNCACB window as before in the CBCACONH shows the alpha carbons and beta carbons color-coded for identification (Figure 12).

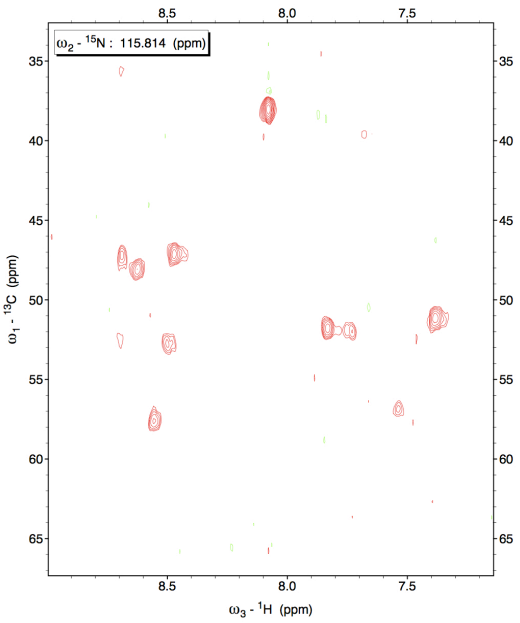

Figure 11 Sample view of CBCACONH Sparky window showing peaks representing alpha and beta carbons from position n-1 residue

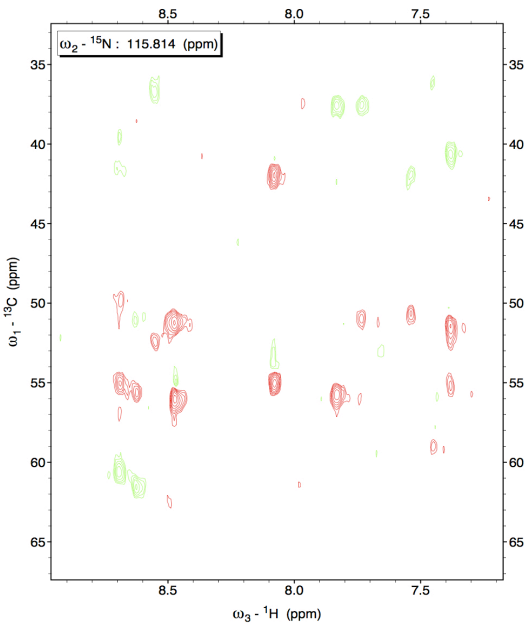

Figure 12 Sample view of HNCACB Sparky window showing peaks representing alpha carbons (red) and beta carbons (green) of position n-1 and n residues

In general, the 3D experiments provided clear peaks that allowed for partial backbone assignment for the RBD sample. Using the PINE server, uploaded peaks and a sequence of the protein were translated into an approximation of the assignments based on characteristic chemical shifts and sequential relationships determined from the 3D experiments. The results of this analysis with estimated likelihood of match per residue is summarized graphically in Figure 13.

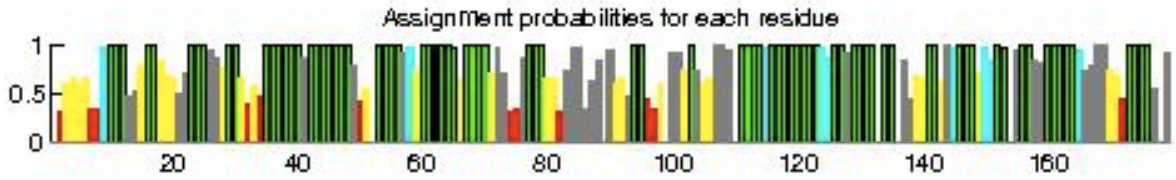

Figure 13 Probability for residue assignments from automated PINE analysis.

Of note, the N-terminal 6-histidine tag shows low success in assignments likely due jointly to the degeneracy of each residue and the relatively unstructured nature of this segment. The relative accuracy in particular stretches of the protein superficially

corresponds to conserved portions of the binding domain, though manual review residue-by-residue of the assignments reveal a remarkable number of implausible matches.

3. Electrophoretic Mobility Shift Assays with Hnrnpab RBD and Actb 3' UTR

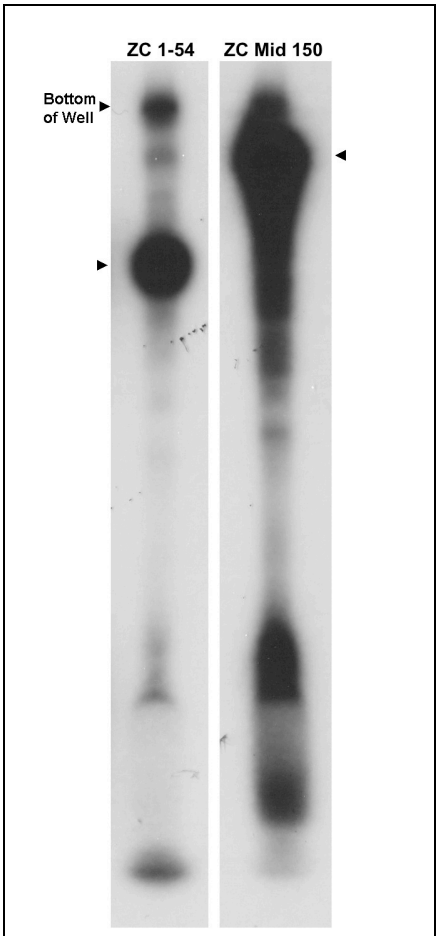

Figure 14 Denaturing gel of labeled RNAs from *in vitro* transcription reaction

Radiolabelled RNA was synthesized by runoff *in vitro* transcription from a plasmid DNA template as indicated in the materials and methods and as shown in denaturing gel (Figure 14). These RNAs were combined with purified Hnrnpab RBD in EMSA RNA binding experiments. A representative EMSA with the 1-54 zipcode is shown in Figure 15. The lower band corresponds to unbound RNA (Figure 15, lane 1). A slower migrating diffuse RNA-protein complex appears as protein is titrated in beginning at approximately 4 nM RBD concentration (Figure 15, lanes 2-7). At roughly

100 nM the titration curve is approaching 50% assortment into upper and lower bands, reflecting the apparent dissociation constant ($K_d$app).

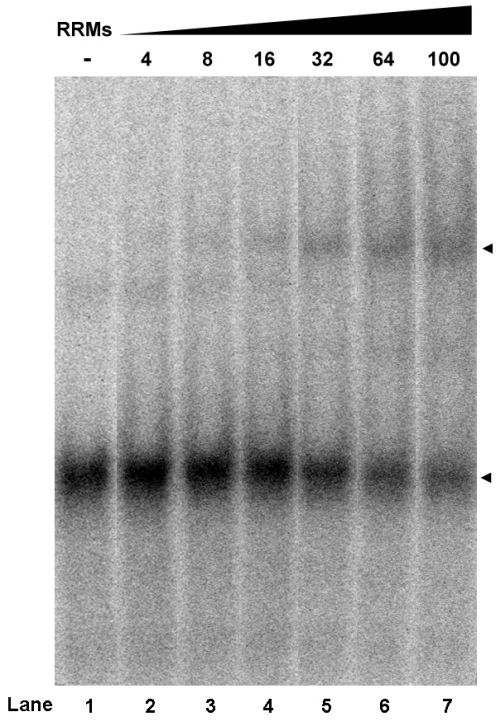

Figure 15 EMSA of zipcode 1-54 from 3' UTR of ActB with varied concentrations (nM) of Hnrnpab RBD. Free probe is labeled with the lower of the two arrowheads on the right, while the RBD-RNA complex is indicated by the upper of the two arrowheads

Figure 16 shows a comparison of binding to the first 54 residues versus the 150 nt sequence 3' of the zip code in the Actb 3' UTR. In this experiment the RNA-protein complex is observed at 8nM RBD, and increases with more protein (Figure 16, compare lanes 3 – 6 to lane 1). In contrast the same protein concentrations do not bind to the 150 nt 3' of the zip code (Figure 16, compare lane 1-6 to lanes 7-12). The appearance of a more diffuse signal at the 100 nM concentration makes it unclear whether binding to this substrate begins at this concentration, but the clear lack of binding at lower concentrations allows us to conclude that the K_d for the 3' 150nt is much larger than for the zip code.

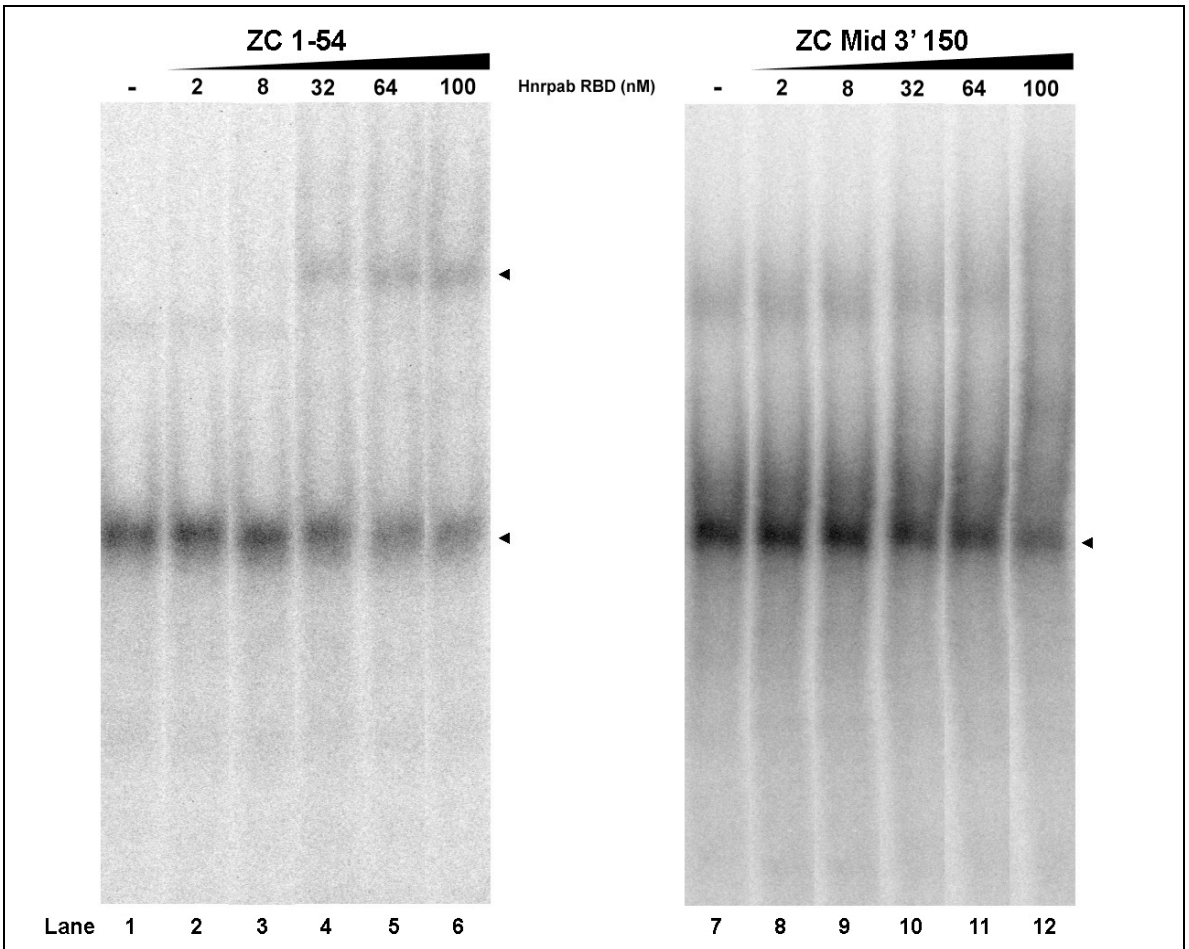

Figure 16 EMSA of zip code 1-54 and middle 150 bases from 3' UTR of Actb with varied concentrations of Hnrnpab RBD

.

IV. Discussion

1. Expression of ^{15}N-Labeled Hnrnpab RBD for NMR

The goal of this series of experiments was to obtain an NMR spectrum of the Hnrnpab RBD to determine whether the three dimensional structure of the RBD could be solved using NMR. The purified RBD product appeared smaller in denaturing SDS-PAGE gels, and we presumed that truncation or specific degradation was responsible. However, mass spectrometry analysis of the different protein bands determined that both bands were identical in peptide content (Figure 8). We hypothesize that the size discrepancy results from a population of the protein that adopts an alternative conformation of the protein that is resistant to denaturation by SDS and DTT at 95 °C. This observation is interesting since both isoforms of Hnrnpab protein migrates in gels at a higher apparent size [18,21]. Interestingly, we also began to see protein bands approximately twice the size of the Hnrnpab RBD protein on SDS-PAGE gels as depicted in Figure 7. The size is consistent with dimerization of the RBDs, and it has been observed that the related AUF1 protein commonly forms dimers to construct a binding site for ribonucleic targets, though it isn't clear whether these dimers are sensitive to denaturation [22]. Study of the dimerization of the RBDs on their own, or upon RNA binding is approachable using relatively simple NMR techniques and will likely be pursued in future research. We continued to the next steps in the project since despite finding multiple proteins observed on an SDS-PAGE gel, they are all the same polypeptide.

The ^{15}N-HSQC spectrum of the purified RBD captured in Figure 9 demonstrated clearly defined and spaced peaks, and this result is a good indication that a structural determination in solution by NMR would be possible and justified further experiments.

2. Double-Labeling of Hnrnpab RBD with ^{13}C and ^{15}N for 3D NMR Experiments

To determine the structure of the Hnrnpab RBD, the most difficult steps involved are assigning the peaks to particular residues in the sequence. Once the backbone is assigned, additional experiments that utilize couplings through space instead of through direct bonds enable contacts between residues to be resolved in the third dimension.

The data collected in the experiments represented by Figures 10-12 showed clear peaks at many positions but without consistent clarity and resolution between most peaks it is difficult to approach a high quality backbone assignment. Through manual and automated processes we were able to assign about 30% of the residues to peaks with high confidence. Building on the data we have, future NMR studies can be adjusted to yield better spectra that may allow for further progress on the structural determination project.

3. Electrophoretic Mobility Shift Assays with Hnrnpab RBD and Actb 3' UTR

Experiments in the Czaplinski lab have shown a specific association of Hnrnpab1 with the Actb 3' UTR using RNA column chromatography and RNP-co-immunoprecipitation. A preliminary experiment indicated that the Hnrnpab RBD on its own directly bound the 54 nucleotide zip code segment with an apparent K_d of about 42 nM (Jeffrey Chao, unpublished data). A K_d^{app} in this range indicates a physiologically relevant interaction between this construct and its RNA target.

Importantly, I demonstrated specific binding of the RBD to the zip code in comparison to the 150 nucleotides 3' to the zip code sequence. This demonstrates that the RBD indeed harbors specificity for a particular sequence, not just RNA in general. In our hands, this level of affinity has not been consistent, but is close to the aforementioned K_d. There are several differences in the RNA binding assays that could explain the differences observed. The RNA substrate in our assays is radiolabeled instead of fluorescently labeled, and the format of my gels is polyacrylamide slab gels versus agarose submarine gels. Repeated binding with the first 54 bases of the zipcode reveals nM affinity, though sufficiently quantitative assays have not yet been completed to pinpoint $K_d$app.

Examination of Figure 16 gives an indication that some protein-dependent shifting of the unbound probe is occurring which is likely of not insignificant affinity. Previous research on the binding properties of Hnrnpab are limited to a bit more than a handful and only one paper written shortly after its isolation nearly 30 years ago yielded estimates of affinity through competitive filter retention assays [23]. Hnrnpab1 was tested with homopoly- and oligonucleotides as well as random copolymers of nucleotides in

varying proportions against tritium-labeled heterogeneous nuclear RNA extracted from HeLa cells. Low retention of the [^{3}H]-hnRNA was interpreted as high competition of the tested polymers and it was found that Hnrnpab1 bound best to random poly U:G (1:1), poly A:G(1:0.64), and poly A:U:G (1.6:1:1.4) RNA [22]. These experiments did not result in binding constants; however, they seem to agree with consensus sites for some members of the similar hnRNP family of proteins based on SELEX experiments that reveal a high affinity for sequences containing 'UAGGG' [24].

It is vital to separate the binding activity of the RBD from what seems to be relatively simplistic guanosine-tetraplex unwinding manifesting as high affinity for G-rich sites. This has been repeatedly found in the literature and is perhaps mediated through the ATP/GTP motif in the C-terminal as suggested by sequential deletion of binding sites in the RBD [25]. Since Hnrnpab1 and Hnrnpab2 binding to G-rich sequences has been shown to be RBD-independent, we are interested to find whether this G binding is absent in RRM-only constructs. In order to test binding through the conventional RBD contacts, single-site missense mutants directed toward abolishing base-stacking capacity are being developed as comparisons against wild-type RBD. Deletion of highly conserved phenylalanine residues is projected to severely limit binding affinity for RBD targets and if successful can be implemented to test full-length protein binding with inactive RRMs.

Following on these results the 54 nucleotide zipcode will be selectively deleted to map high affinity sites for the Hnrnpab RBD. This expected minimal site is likely to be based on a specific sequence given the binding characteristics of similar proteins and the rarity of RRM-containing proteins that strongly target RNA secondary structure as per high-throughput screens [23]. We plan on using the highest affinity construct in NMR RNA titration experiments wherein peak shifts visualized in the ^{15}N HSQC spectrum will reveal evidence of spatial residue rearrangement upon binding. This requires assignment of the peptide backbone using a doubly-labeled RBD sample and 3D experiments before such data is usable. If successful, further experiments would involve titrating in purified exon 7 peptide to observe whether binding can be reduced by what may be a novel regulatory mechanism in proteins of this type.

V. Summary

Our lab focuses on understanding regulators of RNA processing and trafficking in higher organisms. Previous work in the Czaplinski lab has shown an important role for Hnrnpab in mRNA localization and my work has importantly demonstrated that this can be through a direct recognition of the zipcode sequences through Hnrnpab's conserved RRMs. Determining the structural and binding properties of the RBD alone will enable further research into the activity of the alternatively spliced exon 7 and a potential novel form of RNA regulation as well as provide a coherent explanation for the diverse range of biochemical analyses that have complicated the activity of this protein.

Understanding the structure-function relationship in this conserved RNA binding protein will provide greater insight into the logistics of mRNA localization and processing as mediated by protein binding partners. Further, it will cast new light on defects in localization that may result in disease states that are in some ways conventionally difficult to understand [17,18].

VI. References

1. Darnell, J. E. (2013) Reflections on the History of Pre-mRNA Processing and Highlights of Current Knowledge: A Unified Picture. *RNA*. 19:443-460.
2. Malys, N. and McCarthy, J. E. G. (2011) Translation Initiation: Variations in the Mechanism Can Be Anticipated. *Cell. Mol. Life. Sci.* 68:991-1003.
3. Chen, C.-Y. A and Shyu, A.-B. (2011) Mechanisms of Deadenylation-Dependent Decay. *WIREs RNA*. 2:167-183.
4. Barreau, C., Paillard, L., and Osborne, H. B. (2005) AU-Rich Elements and Associated Factors: Are There Unifying Principles? *Nucleic Acids Res*. 33:7138-7150.
5. Brennan, C. M. and Steitz, J. A. (2001) HuR and mRNA Stability. *Cell. Mol. Life Sci.* 58:266-277.
6. Vale, R. D. (2003) The Molecular Motor Toolbox for Intracellular Transport. *Cell*. 112:467-480.
7. Du, T.-G., Schmid, M., and Jansen, R.-P. (2007) Why Cells Move Messages: The Biological Functions of mRNA Localization. *Semin. Cell Dev. Bio.* 18:171-177.
8. Lawrence, J. B. and Singer, R. H. (1986) Intracellular Localization of Messenger RNAs for Cytoskeletal Proteins. *Cell*. 45:407-415.
9. Martin, K. C. and Ephrussi, A. (2009) mRNA Localization: Gene Expression in the Spatial Dimension. *Cell*. 136:719-730.
10. Glisovic, T., Bachorik, J. L., Yong, J., and Dreyfuss, G. (2008) RNA-Binding Proteins and Post-Transcriptional Gene Regulation. *FEBS Lett*. 582:1977-1986.
11. Chaudhury, A., Chander, P., and Howe, P. H. (2010) Heterogeneous Nuclear Ribonucleoproteins (hnRNPs) in Cellular Processes: Focus on hnRNP E1's Multifunctional Regulatory Roles. *RNA*. 16:1449-1462.
12. Maris, C., Dominguez, C., and Allain, F. H.-T. (2005) The RNA Recognition Motif, a Plastic RNA-Binding Platform to Regulate Post-Transcriptional Gene Expression. *FEBS J*. 272:2118-2131.
13. Barraud, P. and Allain, F. H.-T. (2013) Solution Structure of the Two RNA Recognition Motifs of hnRNP A1 Using Segmental Isotope Labeling: How the Relative Orientation Between RRMs Influences the Nucleic Acid Binding Topology. *J. Biomol. NMR*. 55:119-138.
14. Chao, J. A., Patskovsky, Y., Patel, V., Levy, M., Almo, S. C., and Singer, R. H. (2010) ZBP1 Recognition of β-actin Zipcode Induces RNA Looping. *Gene. Dev.* 24:148-158.
15. Enokizono, Y., Konishi, Y., Nagata, K., Ouhashi, K., Uesugi, S., Ishikawa, F., and Katahira, M. (2005) Structure of hnRNP D Complexed with Single-stranded

Telomere DNA and Unfolding of the Quadruplex by Heterogeneous Nuclear Ribonucleoprotein D. *J. Biol. Chem*. 280:18862-18870.

16. Trojanowicz, B., Dralle, H., and Hoang-Vu, C. (2011) AUF1 and HuR: Possible Implications of mRNA Stability in Thyroid Function and Disorders. *J. Thyroid Res.* 4(Suppl 1):S5.

17. Akindahunsi, A. A., Bandiera, A., and Manzini, G. (2005) Vertebrate 2xRBD hnRNP Proteins: A Comparative Analysis of Genome, mRNA and Protein Sequences. *Comput. Biol. Chem.* 29:13-23.

18. Sinnamon, J. R., Waddell, C. B., Nik, S., Chen, E. I., and Czaplinski, K. (2012) Hnrpab Regulates Neural Development and Neuron Cell Survival After Glutamate Stimulation. *RNA*. 18:704-719.

19. Czaplinski, K., Köcher, T., Shelder, M., Segref, A., Wilm, M., and Mattaj, I. W. (2005) Identification of 40LoVe, a *Xenopus* hnRNP D Family Protein Involved in Localizing a TGF-β-Related mRNA During Oogenesis. *Dev. Cell*. 8:505-515.

20. Norvell, A., Kelley, R. L., Wehr, K., and Schüpbach, T. (1999) Specific Isoforms of Squid, A Drosophila hnRNP, Perform Distinct Roles in Gurken Localization During Oogenesis. *Gene. Dev*. 13:864-876.

21. Khan, F. A., Jaiswal, A. K., and Szer, W. (1991) Cloning and Sequence Analysis of a Human Type A/B hnRNP Protein. *FEBS Lett.* 290:159-161.

22. Zucconi, B. E. and Wilson, G. M. (2013) Assembly of Functional Ribonucleoprotein Complexes by AU-rich Element RNA-binding Protein 1 (AUF1) Requires Base-Dependent and –Independent RNA Contacts. *J. Biol. Chem*. 288:28034-28048.

23. Kumar, A., Sierakowska, H., and Szer, W. (1987) Purification and RNA Binding Properties of a C-type hnRNP Protein from HeLa Cells. *J. Biol. Chem.* 262:17126-17137.

24. Ray, D., Kazan, H., Cook, K. B., et al. (2013) A Compendium of RNA-Binding Motifs for Decoding Gene Regulation. *Nature*. 499:172-177.

25. Weisman-Shomer, P., Cohen, E., and Fry, M. (2002) Distinct Domains in the CArG-Box Binding Factor A Destabilize Tetraplex Forms of the Fragile X Expanded Sequence d(CGG)$_n$. *Nucleic Acids Res*. 30:3672-3681.