CHARACTERIZING THE LIGAND SPECIFICITY OF THE RNA-BINDING PROTEIN LARP6

by

Eliseo Salas

A thesis submitted to the Graduate Council of
Texas State University in partial fulfillment
of the requirements for the degree of
Master of Science
with a Major in Biochemistry
May 2017

Committee Members:

Karen Lewis, Chair

Ronald Walter

Lisa Warner

Steven Whitten

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

Fair Use

Duplication Permission

DEDICATION

To my mother, Martina Lopez. For always motivating me to reach farther than the eye can see.

ACKNOWLEDGEMENTS

Throughout my life I have walked several paths, which led me to the individual I am today. Some paths led me to some dead ends while others opened the door to some really cool experiences. Along the way there are times when I have felt tired or simply stopped halfway and dozed off thinking about the future instead of running towards it. Luckily, I have had people who remind me of which way I'm going and motivate me to continue moving forward.

I am grateful for everyone. To say there is a beginning to the path would be an injustice. Therefore, I will mention those individuals who have kept me in check in the path which led me to the completion of my master's degree and who encouraged me to further advance my studies.

A very special thanks to my P.I., Dr. Karen Lewis, for not only teaching me science but also what it means to be a member of the scientific community. Dr. Lewis has been a unique person in my path. I can honestly say that I had never encountered someone like her before. She is very passionate about her work in lab and in class, but someone who I have come to appreciate and cherish over the years. She is the person who I think has kept me sane throughout this higher education process by being a boss when she had to be, and also being a guide for things I had never experienced before. She has been not only a boss, but also a great mentor, and one of the most wonderful human beings I have had the pleasure of knowing, all fused into one. When I left my hometown several years ago, I had

high hopes of finding unique people who were well rounded and could offer me a way of looking at the world through my own eyes with a new perspective. I can say I found that person in Dr. Lewis.

The opportunity to pursue a master's degree which would bridge the gap between a master's and a doctoral program was offered to me by Dr. Nicquet Blake and Dr. Babatunde Oyajobi. I will never forget their encouragement when I expressed my desire to continue learning. I would also like to thank Dr. Rachell Booth and Dr. Ron Walter for always guiding me in the right direction when I was wondering off the path. These two individuals acted as a Ying and Yang during my professional development. Thanks to these four individuals I owe the completion of my master's education and the ability to pursue a Ph.D. education.

I would like to thank Dr. Linda Roman, who was a great mentor. It was her questioning and unending curiosity approach to science which made me fall in love with research. Her patience with me was more than I could have ever asked for. Without her, I would not have had the determination or ambition to continue my career. Not only did Dr. Roman give me my first real laboratory experience, but she also introduced me to some interesting people who share a real passion for unearthing answers hidden inside our cells. These individuals included Dr. Satya Panda and Dr. Rekha Kharr, who were always eager to teach me new methods with the most heartwarming smile in their face; Dr. Martino Adamo, who always showed an interest in what I was learning and helped me look at things

in the big picture; and to Mrs. Kathryn Hinchee-Rodriguez, who motivated me to get out of my shell and strive to become the best that I could ever be.

I want to tip my hat to the following people who I suffered, grew up, and spent more than $20 on overpriced coffee with. First off, Jose Castro, who has been one of my closest friends and an accomplice whenever something went wrong in lab. I am grateful for always hearing me and always being down to try new restaurants and activities with me. E.L. Peña was always around for hugs and jokes, and makes the best vegan tacos a real Mexican could eat. Even Francisco Betancourt will make this list since he was always around to give a mature point of view whenever I was over childish. In class, Ms. Elizabeth McIvor is who I looked up to when I needed to focus on studying and who always posed the hardest questions for me to answer. She showed me the value of understanding your own research and the value of the knowledge you acquire from reading on subject with which you are unfamiliar. I would also like to thank Ms. Cally Anne Moore for being a good friend and for teaching me that there should be a balance in your life which you should never compromise, and when things go bad you should work extra hard to achieve your goals. Thanks to Mrs. Kimberly Ramsdell who was always smiling in lab even when her experiments didn't work. That was a huge motivation to keep going at the beginning of every optimization experiment I had to do.

Lastly, I would like to thank my only family, my mother who despite being a single mother worked hard to see me succeed, and who never doubted me even when I doubted

myself. She is the person who is always happy to hear me and who is everlasting in my life.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

ABSTRACT

The La-related proteins are a superfamily of RNA-binding proteins characterized by a unique RNA binding domain that is comprised of two subdomains, the La motif (LAM) and the RNA recognition motif (RRM), which are connected by a flexible linker. It is still unclear whether ligand recognition by LARP6 is sequence-specific or dependent on formation of specific three-dimensional structures. The lack of a clear binding mechanism by LARP6 limits the ability of the field to identify other ligands for this protein. The LARP6 protein is only known to bind to a hairpin structure that is found in the 5' untranslated region (UTR) of collagen type I mRNA (*COL1A1*). Interestingly, a similar hairpin structure with a high degree of sequence conservation is found in the 5'UTRs of three other fibrillar human collagen mRNAs (*COL1A2*, *COL3A1* and *COL5A2*). The goal of this work was to determine the affinity and specificity of LARP6 binding to this set of collagen mRNAs. To accomplish this, we first optimized a native electrophoretic mobility shift assay (EMSA) using biotinylated ligands. Using this system, we have shown that LARP6 binds with high affinity (in the low nanomolar range) to all of the collagen sequences. Unexpectedly, LARP6 was also observed to bind with similar affinity to a random RNA sequence with a predicted secondary structure very different from the collagen mRNA ligands. These data suggest that LARP6 is able to bind RNAs due to nonspecific ionic interactions with the RNA phosphate backbone. This hypothesis was tested by evaluating the dependence of the apparent binding affinity on salt concentration, ranging from 20 mM to 150 mM (physiological concentration) salt. We found that the binding affinity of some RNAs demonstrate a linear dependence on ionic strength, while other ligands seem to be minimally affected by salt concentration. The predicted secondary structures of the collagen mRNA ligands that were minimally affected by salt indicate that *Hs*LARP6 recognizes RNA ligands at least in part based on structure as well as sequence.

1.INTRODUCTION

There are many types of cellular RNAs, ranging in size from a dozen to thousands of nucleotides. The types of cellular RNAs include micro-RNAs, messenger RNAs, transfer RNAs, ribosomal RNAs, and others. These RNAs may contribute to regulating the expression of genes through complex control mechanisms that govern the quantity and location of protein synthesis.

Gene expression regulatory mechanisms also include many RNA binding proteins (RBPs). RBPs may contribute to regulation of gene expression in many ways, including; transportation of mRNA from inside the nucleus to the cytoplasm, binding of the ribosomes to the mRNA, and degradation of the mRNA. RBPs may contain a number of different structural motifs, that include the RNA recognition motif (RRM), dsRNA binding motif (RBM) and zinc finger motif among others.[9] Multiple motifs can interact with one another and lead to cooperative mechanisms of ligand binding.

Proteins are made of various motifs which can act together to perform a function. The way in which a proteins motif associates with a ligand is overall dependent on the quantity of specific interactions that can arise between the motif and the ligand. The amount of the essential requirements that are necessary to be present is known as specificity. Sometimes proteins can be highly specific (requirements for binding are high), and sometimes a protein can be non-specific (requirements are small or non-existent) depending on the interactions that are present. RBP motifs bind different RNAs through hydrogen bonds, Van der Waals, and ionic interactions which can be present in different combinations. The structures which RNA can adopt make it suitable to act as a dynamic ligand.

RNA Structure

RNA can form complex intermolecular and intramolecular interactions that are distinct from those formed by DNA. These structures are based on differences in DNA and RNA chemistry (Figure 1). While both DNA and RNA are composed of a ribose ring, a nitrogenous base, and a phosphate backbone, the hydroxyl group at the 2' position of the ribose distinguishes RNA from DNA, which has only a hydrogen at this position. This change in chemistry produces significant differences in both the structure and function between RNA and DNA.

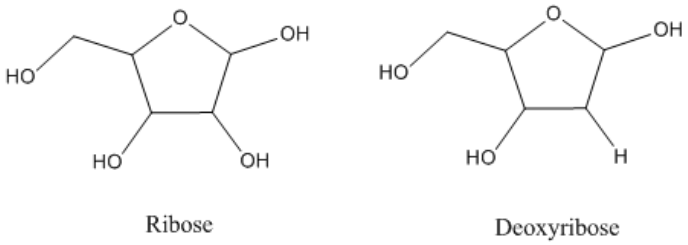


Figure 1. Comparison between the ribose and deoxyribose sugars found in RNA and DNA, respectively.

Hydrogen bonding between base pairs depends on the composition of the actual nucleotides, whereas stacking depends on composition and the overall sequence arrangement in three-dimensional space. The Watson-Crick base-pairing model underlies the B form structure of double-stranded nucleic acids (Figure 2). B form DNA and RNA have a wide major groove and narrow minor groove with almost perfect perpendicular alignment between the sugar ring and the nitrogenous base. DNA predominantly forms a B-form structure *in vitro*. The A-form structure has a narrow deep major groove and a wide

shallow minor groove. Double stranded RNA primarily adopts the A-form structure, as hydrogen bonding that includes the 2'OH group makes a more flexible backbone.

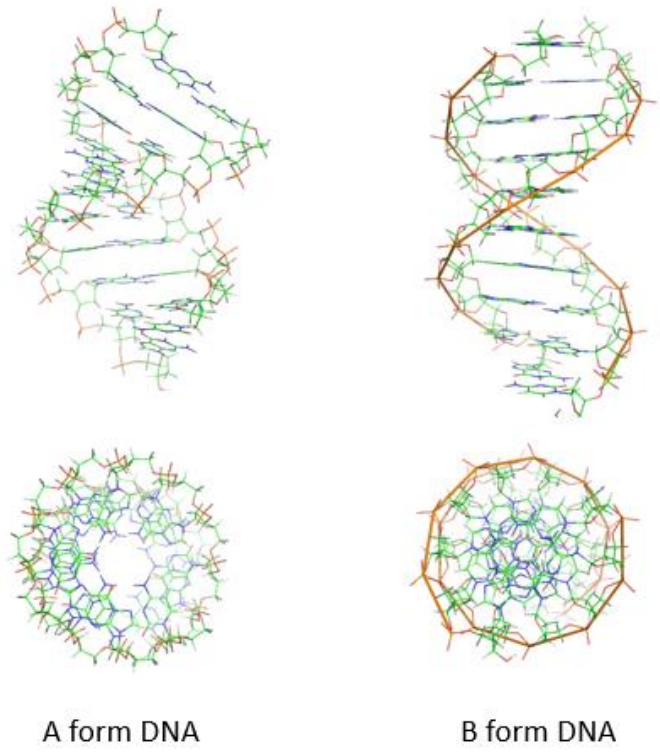


Figure 2. 3D view of A-form (PDB: 440D) and B-form (PDB: 1BNA) helical structures of DNA (side and top view). RNA primarily adopts the A-form structure. Structures generated using PyMOL.

In both DNA and RNA, the bases can form base pairs, either intra-strand (by folding on itself) or inter-strand (by interacting with a separate molecule). RNA generally forms the Watson-Crick base pairs C:G and A:U. However, the Watson-Crick model assumes a rigid nucleotide system in which the bases are in their most probable tautomeric form.[10] Alternative base pairing can also occur. One form of alternative base pairing is wobble base pairing, the most common type which occurs is G:U (Figure 3). The flexible backbone in RNA can accommodate wobble base pairing in addition to base stacking. The existence

of different base pairing possibilities besides Watson-Crick base pairs makes it challenging to accurately predict RNA secondary and tertiary structures.

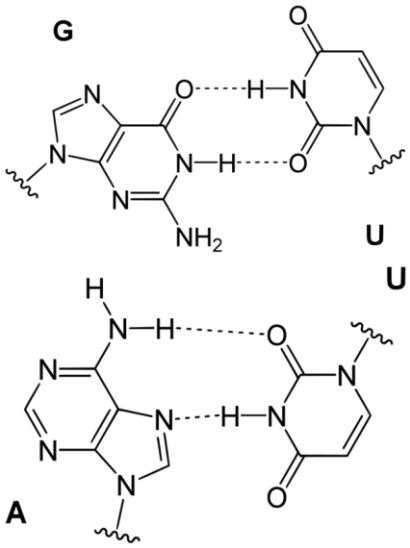


Figure 3. Wobble base pair G-U compared to Watson-Crick base pair A-U.

As a result, the tertiary structure of RNA is very different than that of DNA. RNAs normally occur as single stranded polynucleotides, therefore it is not surprising that when complimentary bases come together there are some bases left unpaired.[11] Figure 4 summarizes a few of the different regions within an RNA that can arise when incomplete complementary base pairing occurs. Inside the cells most RNAs have been shown to experience this duality of double stranded and single stranded regions which can then fold to form dynamic and non-linear tertiary structures.

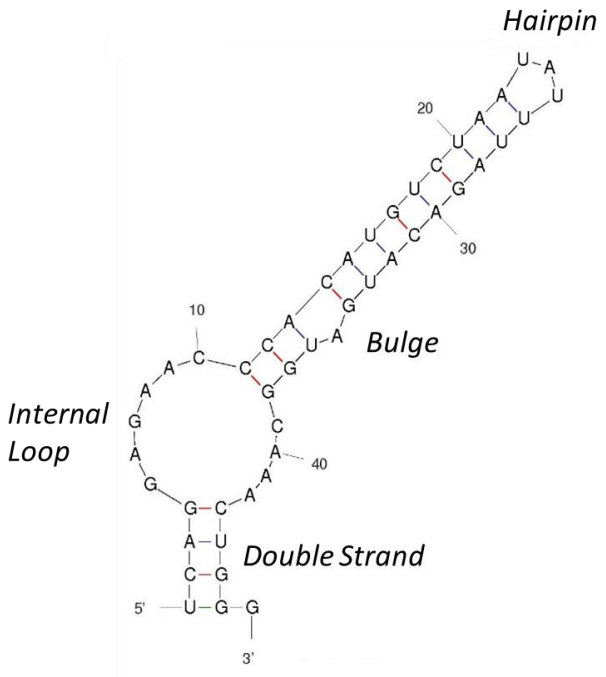


Figure 4. Secondary structures in RNA. Non-complimentary regions can arise in RNA that allow for a more dynamic RNA molecule.

RNA-protein interactions can occur through multiple mechanisms. Proteins can bind to RNA through interaction between nitrogenous bases and aromatic amino acid residues. The stability of aromatic stacking interactions increases in the following order for the natural nitrogenous bases: phenylalanine, histidine, tyrosine, and tryptophan.[12] In addition to base stacking, ionic interactions can occur between charged amino acid residues and the phosphate backbone of RNA to promote RNA-protein complex stability. Overall, however, the major determinant of stability in most RNA-protein complexes is the amount of van der Waals interactions.[13]

Roles of Ions in RNA Structure & Function

Addition of metal cations is often essential for RNA folding.[14] Metal cations interact with the electrostatic field in RNA and are non-specific to sequence (Figure 5).[14] Folding of the RNA molecule into its tertiary forms a negative electric potential which can be stabilized by metal cations.[14] The interaction between the cations and RNA can depend on the surrounding solvent and the negative electric potential of the RNA.

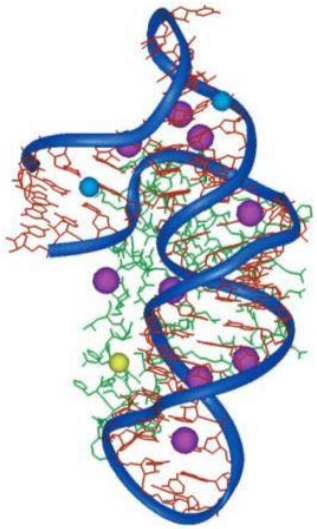


Figure 5. RNA being stabilized by metal cations. Ribosomal S15 protein bound to rRNA is stabilized by Mg^{2+} (purple), Na^{+} (blue), and K^{+} cations. Reproduced with direct permission from Pyle.[7]

Because RNA is a negatively-charged molecule, intermolecular interactions often involve complementary positively-charged binding sites on the binding partner molecule. As a result, ions can also play an active role in the protein-RNA complexes that form *in vitro*.[15] Studies between small peptides and single stranded nucleic acids have shown that the magnitude of binding decreases as salt concentration increases.[16] One possible reason for this effect relies on the theory of counter ion condensation.[16] Under this theory, a linear

nucleic acid is modeled as a cylinder with a net negative charge that contains a certain number of binding sites for positive charges.[16] In the presence of positive ions, such as those provided by the salt, the nucleic acid is saturated. One of the assumptions in this theory is that the nucleic acid is a perfect cylinder with a perfectly distributed charge, and that according to the charge of the ion used there is a limit to counter ion condensation.[16] Due to various asymmetrical structures found in RNAs (*i.e.*, uneven charge on the RNA surface), the counter ion condensation theory had to be revised to accommodate this change in its assumptions.[16] The major revision to the theory was the inclusion of polyions that would then be responsible to cover the whole RNA structure and act as a cohesive unit, which could then disperse the localize charge. Non-specific RNA binding may be observed when there is a large net opposing charge between the RNA and the protein.[15] It is important to experimentally quantify binding in physiological salt concentrations.

The stability of RNA structures is most commonly described in terms of Gibbs free energy, and thus depends on temperature, enthalpy and entropy of the system. The secondary structure of RNA is the spontaneous base pairing that occurs when RNA is either synthesized or heated and allowed to cool, and occurs within a matter of seconds.[17] An RNA sequence may be able to form two or more secondary structures that are favored to occur in solution (Figure 6). For example, the random RNA ligand used in this thesis *RNA Str8*, which is expected to be single stranded (straight) at 30 °C, has two predicted conformation at 16 °C while it only has one at 4 °C (Figure 6). This specifically presents a problem when trying to determine the structure that is functionally active as a ligand for binding, because there are many possible folded structures plus the unfolded state.

However, RNA folding pathways can in some cases be manipulated using temperature or chemical environments to favor a desired secondary structure.

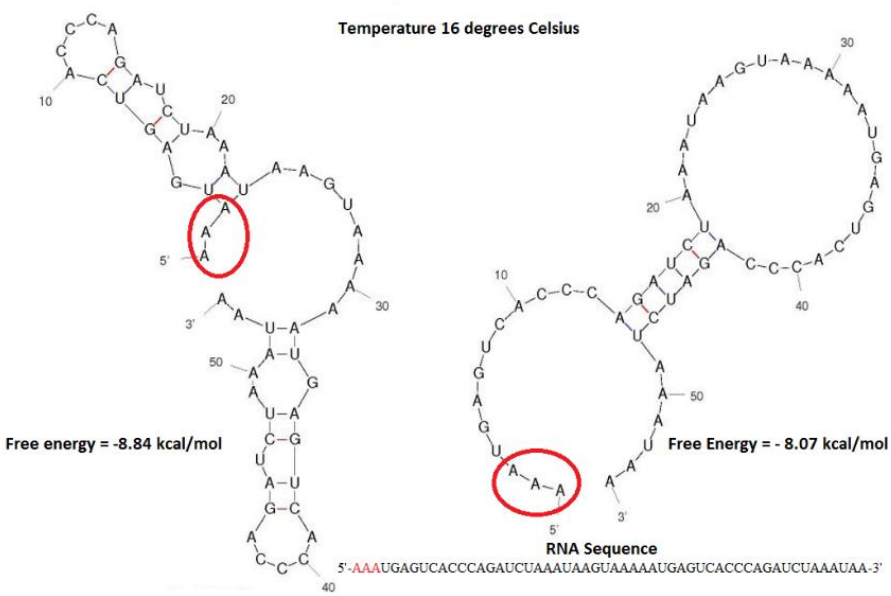


Figure 6. Predicted secondary structures for a random RNA using mFOLD. Red circle denotes the 5'end of the RNA.

Challenges of predicting tertiary structure

With predicted secondary structure elements, the tertiary conformation of small (<100 nt) RNAs can be computationally predicted. The dynamic range in which RNA can fold is as vast as the conformations in which a protein can be found, and as discussed above, RNA structure can change depending on the environment. This presents a particular challenge when trying to replicate *in vivo* studies of very large (>500 nt) RNAs using *in vitro* synthetic RNAs.[17] Although there are currently efforts to predict RNA tertiary configuration with computer programs, only small RNA molecules are able to be accurately predicted (Figure 7).[4] The best way to characterize RNA tertiary structure today

is the use of x-ray crystallography or NMR, which can be time consuming and expensive. In addition to this, RNAs may change tertiary conformation when bound to a protein, or vice versa, which can further complicate the ability to accurately predict physiologically functional RNA structures.[18]

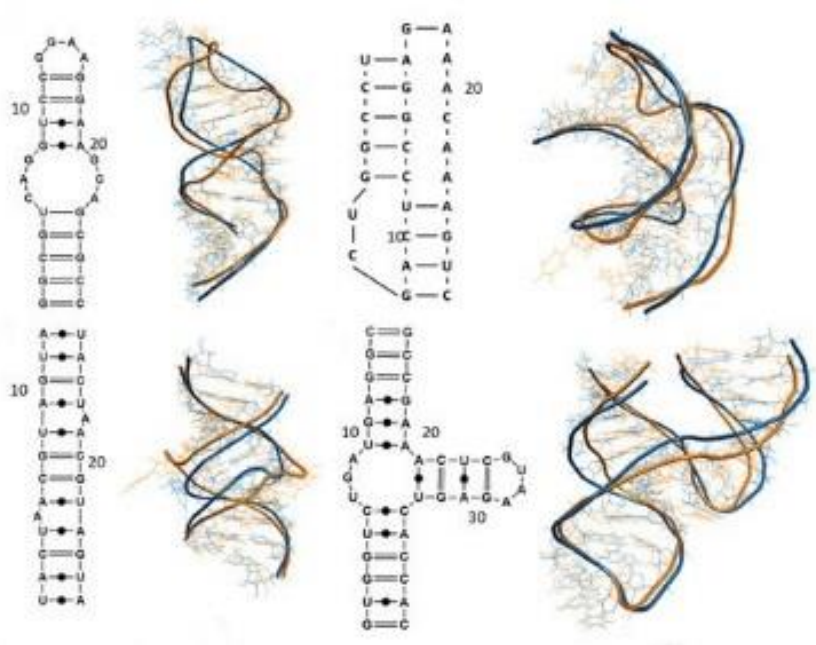


Figure 7. Comparison of RNAs tertiary structure experimentally (orange) and computer generated 3D prediction (blue). Reproduced in full with permission from Zhao et. al.[4]

RNA Sequences and Structures in Translation

The sequence recognition of mRNA needed for translation is different in eukaryotes and prokaryotes. Herein, we will reference five parts that make up a typical mRNA macromolecule (Figure 8): the 5' cap, 5' untranslated region (5'UTR), coding sequence (or "open reading frame"), 3' untranslated region (3'UTR), and a polyadenosine (polyA) tail. Both the 5' cap and the polyA tail serve as recruiting agents to begin translation. The 5' and 3' UTRs, are non-coding sequences that flank the coding sequence.

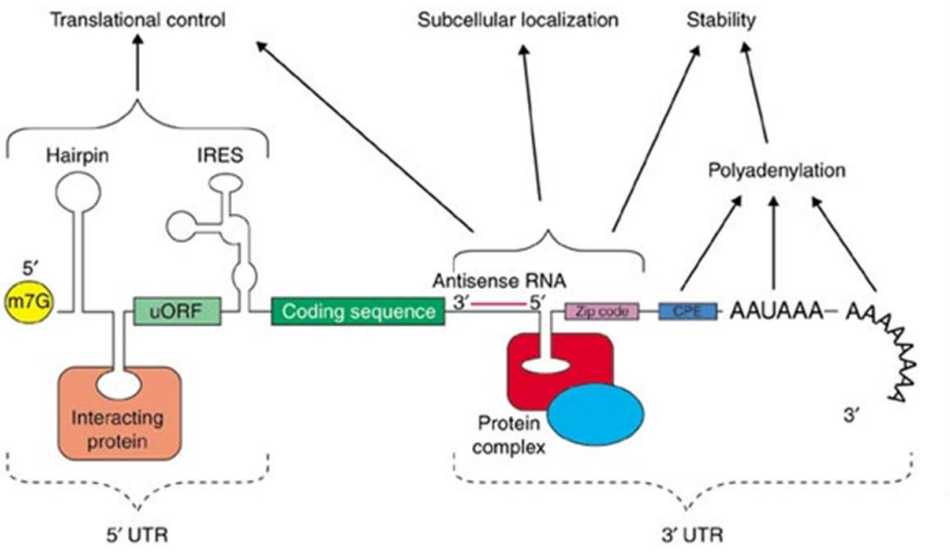


Figure 8. The different regions that make up eukaryotic mRNA.[5] Reproduced in full with permission from Mignone et.al,[8] under the Creative Commons Attribution License.

Riboswitches

UTRs were regarded as space filling sequences until recently, when it was shown they are able to regulate gene expression and are conserved throughout vertebrate species.[19] The complex tertiary structures that form in the 5'UTRs of mRNAs that can bind to metabolites and/or proteins are known as riboswitches (Figure 9).[20] Riboswitches interact directly with naturally occurring metabolites such as amino acids and different inorganic ions that can induce a change in the UTR.[21] The riboswitches were first discovered in bacteria, were relatively small, and were affected by a change in the metabolite concentration to induce structural changes.[21]

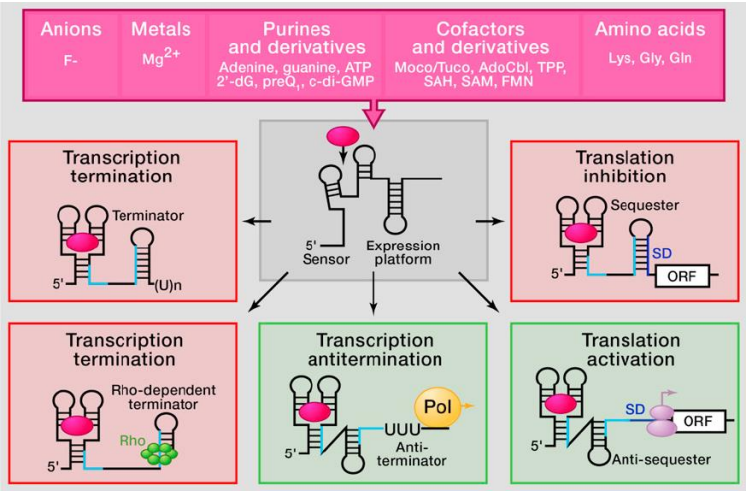


Figure 9. Examples of different forms of regulation in which RNA can participate. Adapted with direct permission from Serganov et. al.[3]

In contrast to riboswitches, which directly bind small molecules to regulate translation, other RNA structures serve as binding sites for regulatory proteins. Characterization of the molecular mechanisms of interaction between nucleic acids and proteins is necessary to understand the overall mechanism and stability of RNA-protein complexes. There are several factors that affect stability, which may include non-polar, polar, and ionic interactions that may greatly influence the dynamic RNA-protein complex.

RNA-Binding Proteins

RNA binding proteins (RBPs) are a group of proteins that participate in many different cellular pathways. By binding in a variety of ways to different RNAs, RBPs can participate in multiple pathways of post-transcriptional gene regulation.[9] RBPs are first classified by the type of RNA they bind to, including but not limited to mRNAs, pre-rRNAs, tRNAs, snRNAs, snoRNAs, and ncRNA.[9] Perhaps the most important of these is mRNA due to its direct participation in one of the most important pathways in the cell:

protein synthesis. Messenger RNA binding proteins (mRBPs) are the most abundant of all RBPs.[9] It has been hypothesized that mRBPs are able to rapidly adapt evolutionarily to different RNA ligands due to the many potential mechanisms for recognizing RNA.[9]

Three RNA-binding domains found in mRBPs that have been studied closely are: the K homology domain, the double stranded RNA (dsRNA) binding motif, and the RNA recognition motif (RRM).[22] The K homology domain binds to both ssRNA and ssDNA, and has the ability to interact with multiple mRBP domains[22]. The dsRNA binding motif recognizes dsRNA by binding to the major and minor grooves of RNA.[22] The RRM is the most commonly found RNA binding domain in mRBPs.[9] The RRM has a conserved peptide sequence that is ~90 amino acids long and was first observed in heterogeneous nuclear ribonucleoprotein complexes (hnRNPs).[2] A multiple sequence alignment of RRM containing proteins shows two highly conserved regions in the RRM motifs, termed RNP1 and RNP2, which consist of 8 central residues that are mainly aromatic and positively charged (Figure 10).[2]

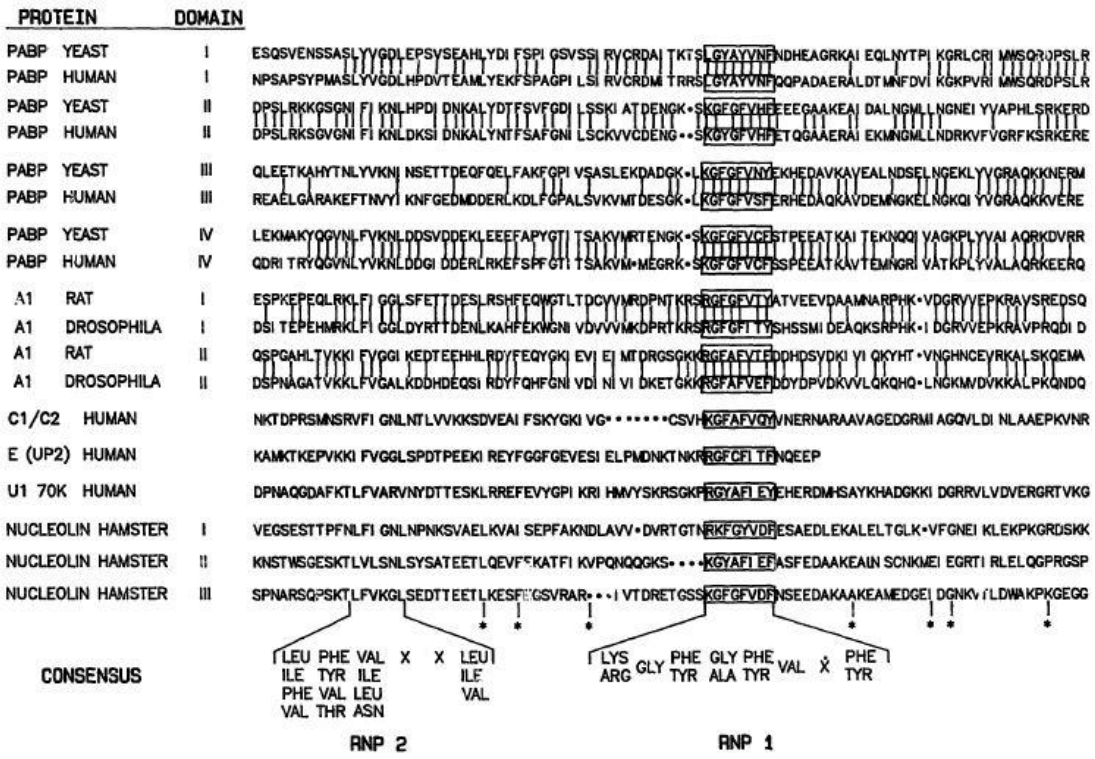


Figure 10. Alignment of different RRM-containing proteins show conserved regions denoted RNP1 and RNP2. Adapted with direct permission from Dreyfuss et al.[2]

The typical structure of the RRM motif consists of four antiparallel beta strands that form a central beta sheet, packed against two alpha helices.[6] In the primary sequence, the order of these secondary structure elements are β1, α1, β2, β3, α2, β4.[6] The points of contact between the RRM in hnRNP and nucleic acids were the conserved sequences, RNP1 and RNP2, which are located on the β-sheet surface (Figure 11).[23]

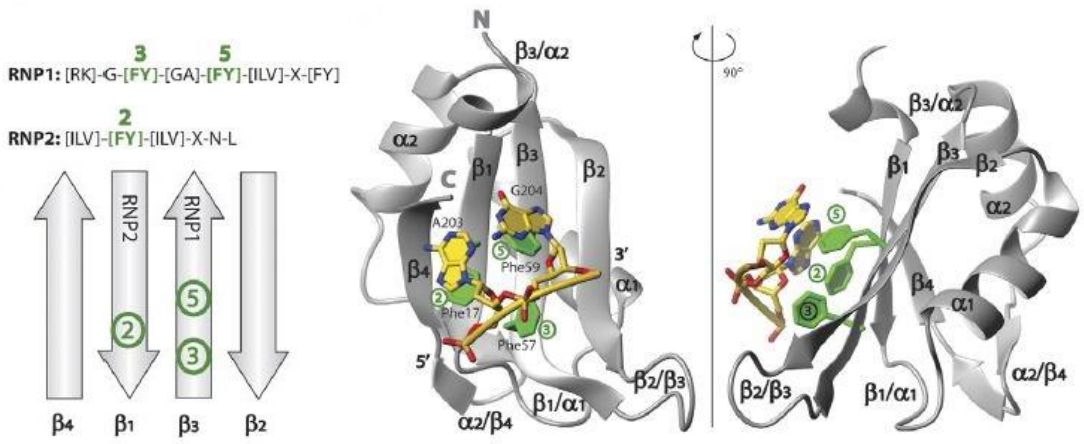


Figure 11. Binding of a DNA oligo to a canonical RRM found in hnRNP. In canonical RRMs a series of conserved regions containing hydrophobic amino acid residues (RNP1 and RNP2) are responsible for recognizing nucleic acids through base stacking. Adapted with direct permission from Clery et. al.[6]

The La-Related Protein Superfamily

The discovery that different RBPs can affect gene regulation by binding to RNA molecules has led to an intensive study of RBPs. The La-related protein (LARP) superfamily is characterized by a bipartite RNA binding domain that is comprised of two subdomains, the La motif and the RNA recognition motif (RRM), that are connected by a flexible linker. The RNA binding domains of the different LARP subfamilies bind a wide variety of RNAs (*e.g.* tRNA and mRNA) (Figure 12).[24] While the La motif is highly conserved, the amino acid sequence of the RRM varies greatly between LARP6 proteins across vertebrate species.[25]

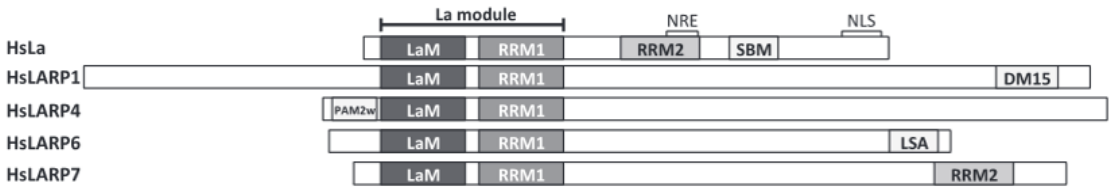


Figure 12. Alignments of the relative sizes and composition of different La related proteins. The La proteins contain the following motifs: La motif (LaM), RNA recognition motif (RRM), La associated motifs (LSA), Poly A binding protein interaction motif (PAM2w), DM15 repeat containing region (DM15), and sumo binding motif (SBM). Adapted with direct permission from Martino et. al. 2014.[1]

Genuine La and LARP7

The genuine La protein (also called the Lupus antigen protein) contains a La motif (LaM), and a RRM, which together make up the La module that is commonly found in most eukaryotic species.[25] The genuine La protein mainly helps chaperone polymerase III transcripts.[26] Such transcripts include pre-tRNA and snRNAs. The La module binds with high affinity to the 3' UUU-OH terminus in these transcripts to protect against 3'-exonucleolytic degradation.[24] In addition to this function, the La motif has been shown to aid proper folding of tRNA, and improves translation efficiency in some organisms.[25]

In the genuine La protein, the La motif and the RRM are capable of being expressed independently in vitro.[27] The RRM looks almost identical to other observed RRMs in RBPs.[27] Experiments showed the La motif alone cannot bind RNA, nor can the RRM.[27] The structure of the La protein was derived from NMR experiments with a short RNA ligand that was bound on the surface of the overall protein.[27]

LARP7 is similar to the genuine La protein in its function of binding to 3' UUU-OH end regions with the exception that it binds specifically to 7SK snRNA. 7SK snRNA is an essential non-coding RNA that is responsible for regulating P-TEFb, a positive elongation

transcription factor.[28] By forming a protein complex, LARP7 aids in circulating 7SK snRNA which helps protect 7SK snRNA digestion from exonucleases.[25]

LARP1 and LARP4

LARP1 has two distinctive paralogues that are termed LARP1a and LARP1b, from which the former is the most studied. A role of LARP1a is to aid formation of the mTORC1 complex that associates the 5' cap in mRNA and the poly A tail to allow for proper mRNA translation.[29] It is also known to act as a regulator in several translation process.[30]

LARP4 stimulates mRNA translation by interacting with scaffold proteins and is the least conserved of the LARPs having two variants, LARP4a and LARP4b. One of these proteins is the poly A binding protein (PABP) which is responsible for binding to the poly A tail in mRNA.[31] LARP4a binds to poly A tails in mRNA and uses its PAM2w sequence to associate with PABP to recruit other 5'cap translation proteins.[32] LARP4 stimulates mRNA translation by interacting with scaffold proteins and is the least conserved LARP protein.[30]

LARP6

Of the LARP proteins, the cellular function of LARP6 is the least understood. The LARP6 protein has only one known binding target *in vivo*. It binds to a hairpin structure that is formed in the 5' UTR region of collagen type I mRNA and enhance the translation of collagen type I protein.[33] LARP6 has also been reported to remodel dsRNA by denaturing base pairing.[24] Previous experiments using LARP6 mutants identified the

regions within the La motif that are involved in binding RNA, which are consistent with other LARP subfamilies.[1]

Although structurally the RRM in human LARP6 resembles the canonical RRM structure, the presence of two extra motifs makes it unique. The first deviation occurs between β1and α1, there is a short sequence with no designated secondary structure followed by an alpha helix denoted α0'. The second deviation is similar to the first and happens between β2 and β3 with a slightly longer alpha helix denoted α1' (Figure 13).[1] Although these differences in the RRM of human LARP6 are minimal they seem to greatly influence the binding activity of human LARP6. For example, no RNA binding is observed when the human LARP6 RRM domain was replaced with the RRM from the genuine La protein. Theoretically, swapping the RRMs should have resulted in equal binding of the RNA ligand since the RRM structure remained the same. However, because no binding is observed, the expected relationship between common structure to common function was in this case null. To further investigate the deviation that the RRM in human LARP6 has from the canonical La protein, a set of single point mutations to the RNP1 and RNP2 motifs were made. Due to previously being known as the critical points of recognition in the RRM, the point mutations were hypothesized to ablate binding of RNA. However, binding was significantly reduce, but not totally ablated as predicted.[1]

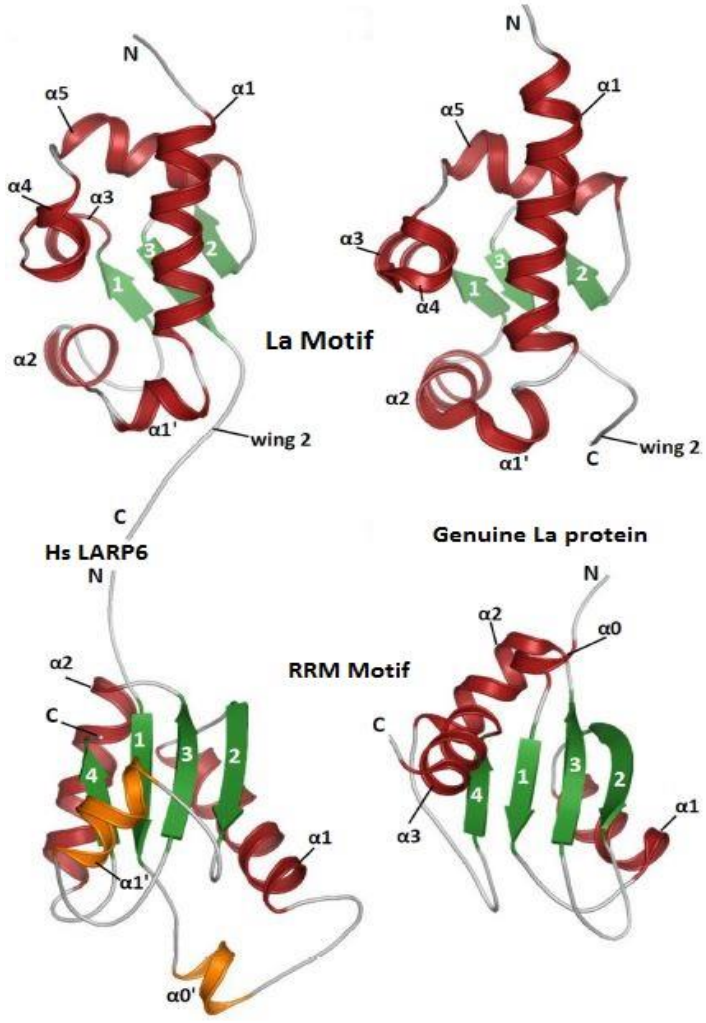


Figure 13. NMR structures for HsLARP6 and genuine La protein divided into their La motifs (top) and their RRM motifs (bottom). Adapted with direct permission from Martino et al 2014.[1]

Overall, these results suggest there are unique interactions between the LARP6 RRM domain and RNA substrate that arise from the mutual interaction with the La motif. It is important to mention that changes to the linker also have significant effects in RNA binding.[1] Most of the studies described above have been made with the isolated La module, not the full-length LARP6 protein. Therefore, the field has not explored unique interactions that may arise using the full length protein. Unlike the other LARP proteins, the LARP6

interactome has not been characterized, either *in vivo* or *in vitro*. Importantly, endogenous targets of LARP6, other than the collagen type I mRNA, have not been identified.

A multiple sequence alignment of LARP6 sequences from eukaryotes shows that the La domain is highly conserved. In addition, the RRM sequence appears mostly conserved with several short regions of evolutionary divergence (Figure 14). The most notable divergence occurs in β1, where there is a high degree of sequence conservation between vertebrates and eutherians, which may be due to some unknown functional adaptation. Another important aspect to notice in this alignment is the absence of the α0' in plants and invertebrates. This short helix may have been evolved to aid in binding but this hypothesis still has not been tested.

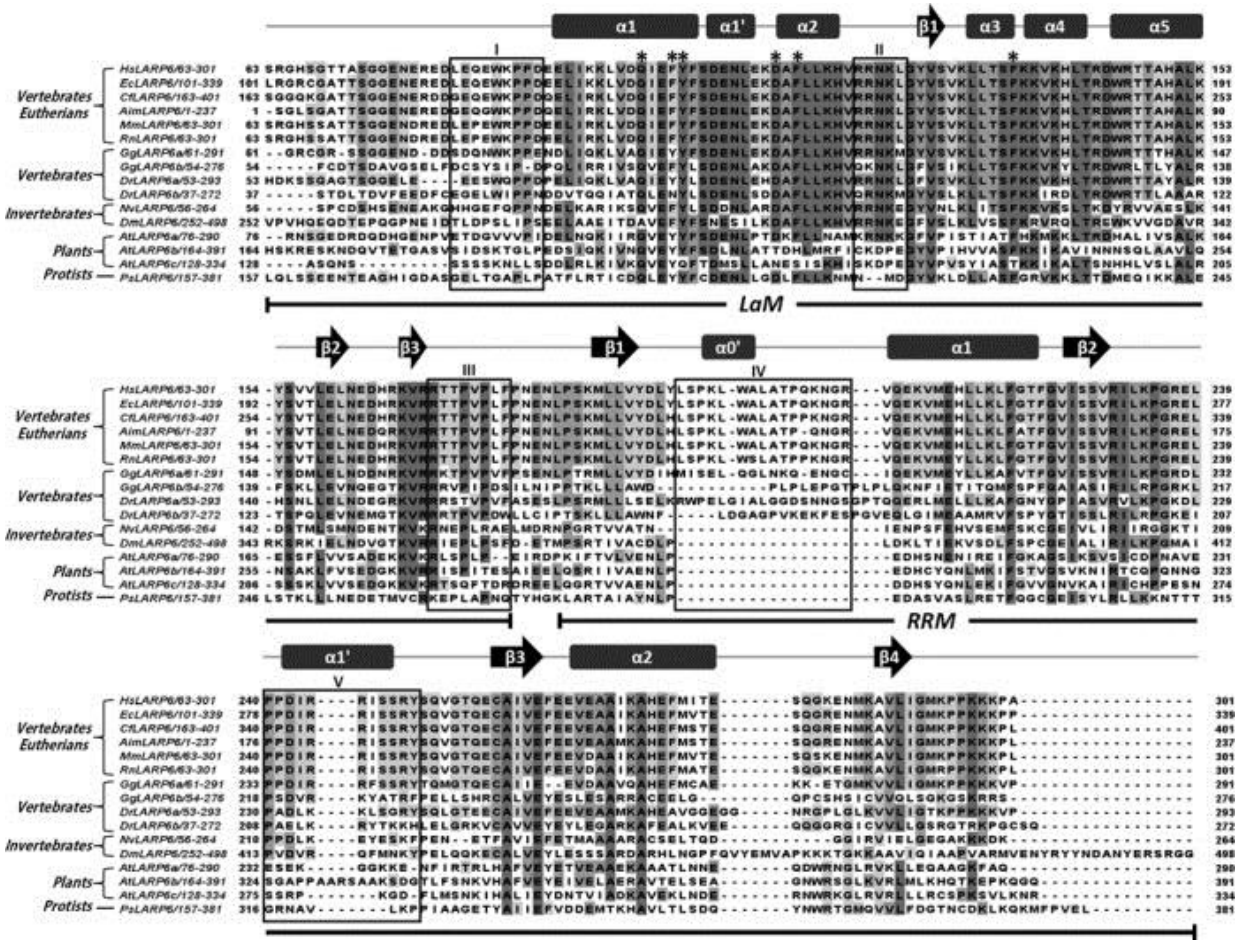


Figure 14. Alignment of different LARP6 sequences of various species. Adapted with direct permission from Martino et al 2014.[1]

Collagen Coding mRNAs Share Conserved UTR Elements

As mentioned previously, the only known ligand for *Hs*LARP6 is the 5'UTR of collagen type I mRNA. The collagen protein family is divided into four major categories; fibrillar collagens, fibril-associated collagens, network forming collagens, and unconventional collagens. Type I collagen is part of the fibrillar collagens, which also include type II, III, V, and XI.[34] An alignment of a section of 5'UTRs of most of these collagens are presented in Figure 15. Secondary structure prediction of these 5'UTR sequences show the formation of similar stem loop structures (Figure 16A).

```
Ligand                    Sequence
HsCOL1A1    -CACAAAGAGUCUACAUGUCUAGGGUCUAGACAUGUUCAGCUUUGUGG
HsCOL1A2    -CACAAGGAGUCUGCAUGUCUAAGUGCUAGACAUGCUCAGCUUUGUGG
HsCOL3A1    GCACAAAGAGUC-UCAUGUCUGAUAUUUAGACAUGAUGAGCUUUGUGC
HsCOL5A2    -UCAGGAGAACCCACAUGUCUAAUAUUUAGACAUGAUGGCAAACUGGG
                    **   *   *******      ******** *          *
RNA Str8    AAAUGAGUCACCCAGAUCUAAAUAAGUAAAAAUGAGUCACCCAGAUCUAAAUAA
```

Figure 15. Alignment of several collagen encoding ligands found in *Homo sapiens (Hs)* and their conservation. At the bottom of the alignment lies the sequence for RNA Str8, a random RNA which was donated by Dr. Kevin Lewis from Texas State University.

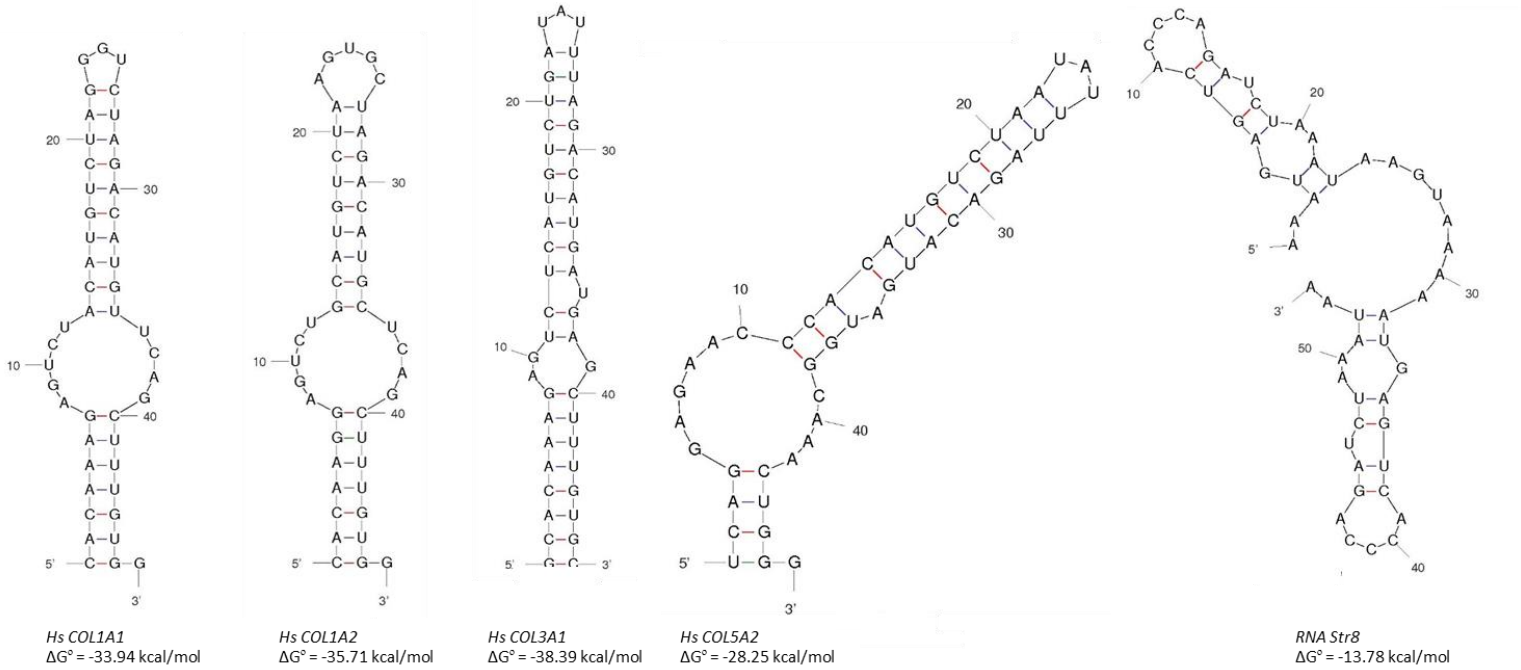


Figure 16. Secondary structure predictions at 4°C of various collagen coding UTRs and a random RNA using mFOLD. The structure names and Gibbs free energy of folding are under each respective RNA. "RNA Str8" was a gift of Dr. Kevin Lewis (Texas State Biology) and was used as a random ligand to test sequence and structure specificity.

Outstanding questions about LARP6 function

This project seeks to better understand the specificity of the human LARP6 protein to various RNA ligands.

Does human LARP6 primarily bind to RNA with a defined secondary structure?

To date, only one collagen-encoding mRNA has been studied as a natural ligand for human LARP6. Contrary to predictions, the RRM in human LARP6 does not use RNPs to recognize collagen mRNA, but the domain is still necessary for binding.[1] It is hypothesized that extended loops found on the surface of the RRM throughout vertebrate species may be responsible for RNA recognition by ionic interactions (Figure 14 regions IV and V). It is crucial to characterize the mechanism by which these two RNA binding motifs recognize their mRNA ligand to further investigate the overall role of LARP6 inside the cell. Can LARP6 bind to other RNAs besides collagen type I mRNA, and if so, what is LARP6 recognizing, structure or sequence? We propose to begin with collagen coding sequences with similar primary and secondary structures to the human collagen type I mRNA and determine if binding occurs. For this we will use conserved UTR regions found in different types of collagen coding mRNAs (type I, III, and V) found in humans (*Homo sapiens)* due to minor changes that occur in overall primary and secondary structure (instead of mutating human collagen type I mRNA) (Figure 15). In addition to the highly conserved human mRNAs, we will test other non-cognate RNAs with similar lengths to explore if any nonspecific binding occurs and to what degree (Figure 16).

Are the interactions between mRNA and Hs*LARP6 dependent on ionic interactions?*

The independent NMR structures of the La motif and RRM give us a base in which to start predicting how the RRM binds to RNA, but fail to give us a direct answer due to the possibility of extensive dynamics introduced by the interdomain linker. The overall length of the RNA required for strong binding (~50 nucleotides) to LARP6 would have made it complicated to have an NMR structure of LARP6 bound to its natural ligand.

As discussed above, RNA is negatively charged and can form nonspecific ionic interactions with binding proteins like LARP6. To investigate the degree in which nonspecific ionic interactions are responsible for ligand binding, we tested a series of salt concentration studies ranging from a minimal to physiological range.

2. MATERIALS AND METHODS

General Methodology

Electrophoretic mobility shift assays were performed to separate free RNA from bound RNA to assess the magnitude of binding. The screening of diverse mRNA sequences had two goals: first, could *Hs*LARP6 bind to them at all, and second, could any of these candidates bind with an equal or greater affinity than the known collagen type I ligand.[35] *Hs*LARP6 was recombinantly expressed in Rosetta *E. coli* cells, and RNA ligands were commercially synthesized by and purchased from IDT.

Expression of Human LARP6

Recombinant human LARP6 was expressed in *E. coli* cells from the pET28-His_6-LARP6 plasmid (a kind gift of Dr. Mark Bayfield, York University; Figure 17). Briefly, the expression plasmid was transformed into *E. coli* Rosetta cells and stored as glycerol stocks at –70 °C. An overnight culture was started from the *Hs*LARP6 -70 ºC glycerol stock and contained 50 µg/mL kanamycin and 34 µg/mL chloramphenicol. In the morning, 1 L of LB at room temperature was inoculated with the overnight cells. The cells grew at 37 °C until a 0.5-0.6 OD_{600} was reached. Cells were then cold-shocked on ice for 15 minutes with constant swirling, and induced with 1 mM IPTG and grown for 8 hours at 18 °C. Cells were then harvested at 4 °C at 5,000 x*g* and stored at -20 °C.

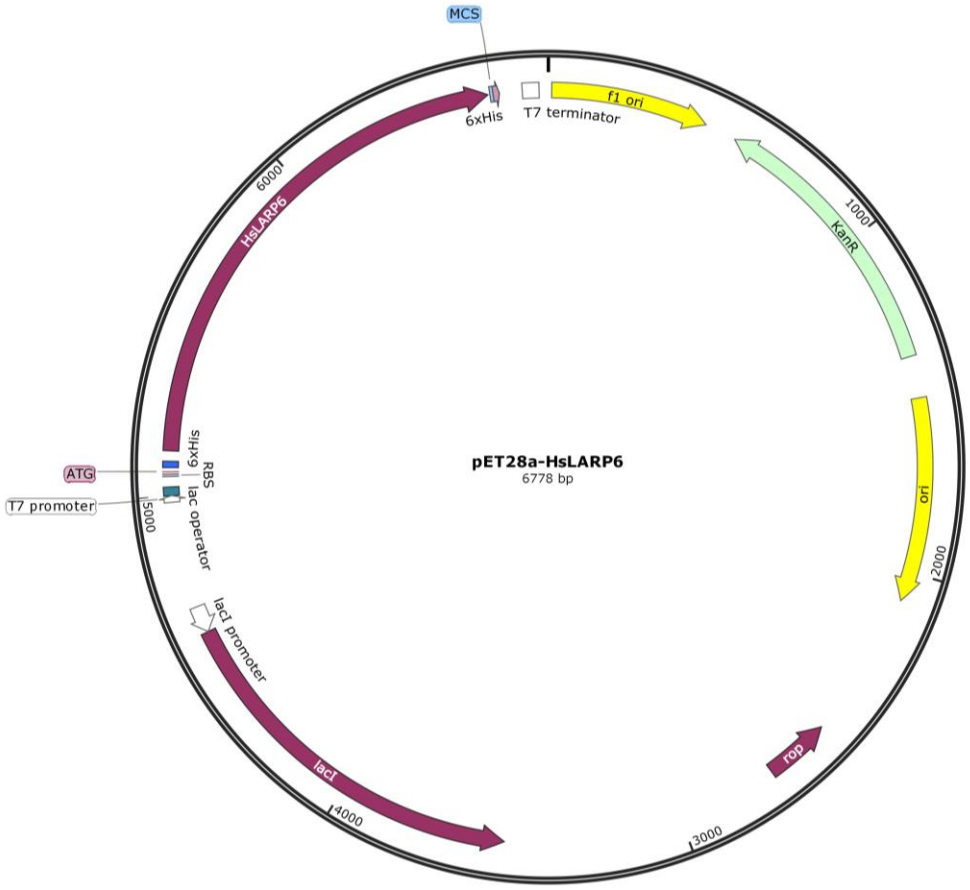


Figure 17. pET28-His$_6$-*Hs*LARP6 plasmid donated by Dr. Bayfield.

Purification of Human LARP6

Before starting purification, the Sephadex 200 size exclusion column was equilibrated with Storage Buffer (Table 1). The cell pellet was thawed on ice for 10 min. prior to resuspension. An EDTA-free Pierce protease inhibitor tablet was dissolved in 50 mL of lysis buffer (Table 1), from which 30 mL was used to resuspend the cell pellet to prevent protein degradation and the remainder held on ice for later use. Resuspended cells were sonicated using a Fisher Scientific Model 505 Sonic Dismembrator with the following parameters: 50% amplitude, 1 sec. pulses, for a total of 20 pulses followed by 30 sec. rest in ice water baths, and a total of 6 cycles. Lysate was transferred to an Oakridge tube, and

centrifuged at 4 °C at 18,000 xg for 15 min. Meanwhile, 4 mL of 50% HisPur™ Ni^{2+}-NTA resin beads where poured into a clean but not sterile 15 mL conical tube and water was added to ~10 mL. The tube containing the nickel beads was then spun for about 30 sec. and the water was removed. This process was repeated 2 more times using what was left of the lysis + protease buffer and only ~2 mL of solution was retained. The Oakridge tubes were then remove from the centrifuge and the supernatant was decanted into a clean but not sterile 50 mL conical tube. Resuspended Ni^{2+}-NTA beads were then added to the cleared lysate and incubated at 4 ºC, while shaking, for 1 hr. The lysate plus nickel beads mixture was transferred to a clean flex column with an attached valve in the 4 ºC fridge. The beads settled for ~15-20 min. The flowthrough was collected in a 50 mL conical tube (F) for further analysis. Special attention was given to ensure that that the nickel column did not run dry and always had ~1 mL of liquid left. The column was washed with 40 mL of Wash Buffer #1 which was then collected in a 50 mL conical tube (W1) as well as with 40 mL of Wash Buffer #2 which was collected in two 20 mL aliquots in 50 mL tubes (W2A, W2B) after running through the column. The protein was eluted into six 4 mL aliquots stored in 15 mL conical tubes using 24 mL of Elution Buffer (E1-E6).

Table 1. Buffers used in purification of *Hs* LARP6.

Buffer	Final Concentration
Lysis/Wash #1 Buffer	
Sodium phosphate (pH 7.4)	50 mM
NaCl	300 mM
Imidazole (pH 8.0)	10 mM
Wash Buffer #2	
Sodium phosphate (pH 7.4)	50 mM
NaCl	300 mM
Imidazole (pH 8.0)	30 mM
Elution Buffer	
Sodium phosphate (pH 7.4)	50 mM
NaCl	300 mM
Imidazole (pH 8.0)	300 mM
Storage Buffer	
Tris-HCl (pH 7.5 @ 4ºC)	50 mM
NaCl	150 mM
Glycerol	5% (v/v)

Molecular weight analysis of all the elutions (F, W1, W2A, W2B, E1-E6) from the column was performed by running a 10% SDS-PAGE acrylamide gel. Samples were prepared by mixing 20 µL of sample with 5 µL of 5X SDS Sample Buffer which were then heated at 90 ºC for 5 minutes and allowed to cool down for 2 min. prior to loading. The gel was run at 200V for 1 hr and loaded with PageRuler Prestained Protein Ladder and 12 µL of each sample. The gel was then stained for 20 min using Coomasie blue and destain for 1 hr to identify the fractions with the highest protein concentration. The fractions with the highest protein concentrations were then concentrated using a VivaSpin Turbo (MWCO 30,000 kDa, pre-rinsed with MilliQ water) to a final volume of ~ 2.2 mL. The concentrated protein was filtered through a 0.22 µm filter into a 2 mL microcentrifuge tube. The protein was then loaded onto the Sephadex 200 using a 3 mL Luer-Lock syringe and eluted at a flow rate of 1 mL/min at 4 ºC.

The next day, the FPLC chromatogram was used to identify possible fractions which could contain *Hs*LARP6. Those fractions were then analyzed via SDS-PAGE acrylamide gel using the same procedure described above. Fractions containing both a high amount and high purity of protein were then concentrated as above to a concentration between 2-10 µM. Protein concentration was measured using an IMPLEN nano-spectrophotometer, with Storage Buffer as a blank and Beer's Law to calculate concentration.

RNA oligonucleotide preparation

All of the RNA oligonucleotides were commercially synthesized by Integrated DNA Technologies (IDT) and consisted of ~50 nucleotides each. Oligonucleotides were first diluted to 1 mM stock solutions, from which aliquots of 10 µM working solutions were made and stored at -20 °C. Oligonucleotide working stocks underwent a maximum of three freeze-thaw cycles to minimize RNA degradation.

RNA secondary structure formation was predicted using the mfold Web Server hosted by The RNA Institute, College of Arts and Sciences, University at Albany. The mfold program is a computer algorithm that calculates the minimum free energy of RNA folding by using experimental data collected over the years that include information on base stacking, single mismatch and corrections for varying salt concentration.[36] The algorithm is also able to incorporate temperature into secondary structure prediction. As detailed later in this chapter, RNAs were incubated with protein on ice for 1 hr, therefore we used 4 °C as our reference for secondary structure prediction.

3' End Biotinylation

Sterile technique was followed during RNA biotinylation due to abundance of RNases on the human body surface. Only filter tips were used when working with RNA and gloves were always worn.

Biotinylation of RNA was performed using the Pierce™ RNA 3' End Biotinylation Kit, following the manufacturer's instructions. Before ligation reactions were started, all kit components except 30% PEG and DMSO were thawed on ice. The DMSO was thawed at room temperature and the 30% PEG in a 37 ºC water bath for 5 – 10 min. until the entire volume was melted. The small heating block was then adjusted to 85 ºC.

Table 2. Components of biotinylation ligation reaction.

Reagent	**Volume (µL)**	**Final Concentration**
Nuclease-free water	3	
10X RNA Ligase Reaction Buffer	3	1X
RNase Inhibitor	1	
Unlabeled RNA + DMSO mix (heated and snap-cooled)	5	1.67 µM (50 pmol)
Biotinylated cytidine bisphosphate	1	33.3 µM (1nmol)
T4 RNA ligase	2	1.33 U/µL (40 U)
30% PEG	15	
Total reaction volume	30 µL	

The ligation reaction was started by transferring 50 pmol of non-labeled control or experimental RNA to sterile 1.5 mL micro-centrifuge tubes. To melt any non-desired secondary structures that may form, 1.25 µL of 100% DMSO was added to each individual tube and heated at 85 ºC for 2 min. The tubes were then immediately removed to an ice water bath and left to chill for 10 min. To a new set of sterile 1.5 mL microcentrifuge tubes,

the following components were added in the order given (Table 2). Special care was taken when transferring PEG due to its viscosity. The reactions were then placed in a 30 ºC heating block and left overnight for ~16 hrs.

The next day, the RNA was recovered by phase separation. To begin, sterile 5 M NaCl solution, 70% ethanol, and 100% ethanol were cooled on ice, and the glycogen stock was thawed. A 24:1 mixture of chloroform:isoamyl alcohol was made in a volume that was sufficient for the number of reactions plus one (*e.g.*, for 5 labeling reactions, 600 µL was prepared). The ligation reactions were then removed from the 30 ºC heating block and 70 µL of nuclease-free water was added, as well as 100 µL of 24:1 chloroform: isoamyl mixture. Each tube was then vortexed at max speed for 10 sec., and centrifuged at max speed (>16,000 x*g*) at room temperature for 3 min. While the reactions were being centrifuged, a new set of sterile 1.5 mL microcentrifuge tubes were labeled. The reactions were then very carefully removed from the centrifuge, and the aqueous phase of each reaction transferred to its appropriate new labeled tube, and the non-aqueous phase discarded into a designated waste container. The RNA was precipitated by adding 10 µL of 5 M NaCl, 1 µL of glycogen, and 300 µL of ice-cold 100% ethanol to each reaction, and left for 1 hr. at -20 ºC. The mixtures were then centrifuged for 15 min. at >13,000 x*g* at 4 ºC. The supernatant was then removed very carefully from the centrifuge so as to not disturb the micropellet at the bottom. When removing the supernatant, careful observation was critical since the RNA pellet was usually microscopic and appeared as a speck of white "webs". To ensure that the micropellet was not disturbed, ~20 µL of 100% ethanol was left for the next step. The micropellet was washed with 300 µL of ice-cold 70% ethanol, and centrifuged for 5 min. at >13,000 x*g* at 4 ºC. The micropellet was then clearly visible

at the bottom of the tube, making the removal of all excess ethanol possible. To ensure that all the ethanol evaporated and only RNA was left in the tubes, the tubes were left open to air dry for ~2 min. The RNA was resuspended in 10 µL of 0.5X TE buffer. The RNA was then transferred into 2 µL aliquots and stored at -20 ºC.

RNA labeling efficiency protocol

Before detecting biotinylated RNA, 1 L of 1X TBE was made and placed at 4 ºC for at least 3 hrs. A Hybond N+ membrane (GE Healthcare Life Sciences) was equilibrated using 1X TBE for at least 10 min. Biotinylated Control IRE RNA (5'-UCCUGCUUCAACAGUGCUUGGACGGAAC-3'-Biotin; Pierce 3'-End Biotinylation Kit) was diluted to make serial standards: 50% (6 µL in 54 µL IDT H_2O), 25% (3 µL in 57 µL IDT H_2O), and 0% biotinylation. Each experimental RNA was diluted to 2 nM (200 fmol in 100 µL) as a starting concentration for blotting. A 96-well conical bottom plate was used to dilute both the control RNA and the experimental RNA to create a series of different RNA concentrations. The initial IRE RNA control dilutions were placed in A1-A3, while the tested RNAs were placed in A4-AX (X being any additional RNAs that are being detected) as seen in Table 3. IDT nuclease-free H_2O was added to wells B-E for every sample, and each sample was serially diluted two-fold for a total of 3 times. For example, if 1A had 50 µL, then 25 µL would be transferred to 1B, and so on. After the dilutions were made, one of the lower corners of the Hybond membrane was cut to preserve orientation, and the membrane placed on a piece of plastic wrap. To the semi-wet membrane, 2 µL of each dilution was spotted to the membrane, and then crosslinked using

a Hoefer UVC 500 UV crosslinker for 45 sec. at 120 mJ. The membrane was then either set aside to fully dry or detected immediately after crosslinking.

Table 3. Representative plate layout for biotinylation efficiency reaction

	1	2	3	4	5	6	7
A	**50%**	**25%**	**0%**	**2 fmol**	**2 fmol**	**2 fmol**	**2 fmol**
B	**25%**	**12.5%**	**0%**	**1 fmol**	**1 fmol**	**1 fmol**	**1 fmol**
C	**12.5%**	**6.25%**	**0%**	**0.5 fmol**	**0.5 fmol**	**0.5 fmol**	**0.5 fmol**
D	**6.25%**	**3.12%**	**0%**	**0.25 fmol**	**0.25 fmol**	**0.25 fmol**	**0.25 fmol**
E	**3.12%**	**1.56%**	**0%**	**0.125 fmol**	**0.125 fmol**	**0.125 fmol**	**0.125 fmol**

Electrophoretic mobility shift assays

At least 3 hrs before starting any EMSA, 2 L of 1X TBE and 5.5% native gels were made and stored at 4 °C, and ice packs were placed at -70 °C. In addition, 10X salt binding buffers were made (Table 3) and diluted to 1X concentration + 15% glycerol. All binding buffers were stored at -20 °C between uses.

Table 4. Final concentrations of 10X binding buffers.

10X Binding Buffer	Tris HCl, pH 7.46 @ 4 °C	KCl	$MgCl_2$	NaCl	DTT
0 mM NaCl	100 mM	200 mM	10 mM	0 mM	10 mM
50 mM NaCl	100 mM	200 mM	10 mM	500 mM	10 mM
100 mM NaCl	100 mM	200 mM	10 mM	1000 mM	10 mM
130 mM NaCl	100 mM	200 mM	10 mM	1300 mM	10 mM

To begin EMSA protocol, a heating block was set at 80 °C and an ice-cold water bath was prepared. The RNA being tested was thawed on ice for 5 minutes and heated at 80 °C for 2 min. During this time, a series of 0.65 mL microcentrifuge tubes were labeled according to the concentration of protein they would contain. After the 2 min. were over, the RNA was quickly placed in the cold ice-water bath for 10 min. to ensure rapid RNA folding. Dilutions of *Hs*LARP6 that ranged from 0 – 8.5 µM were then made using 1X binding buffer + 15% glycerol, always working on ice. RNA stocks were then diluted with 1X TE to 25 nM, and a second dilution was made to 500 pM with 1X binding buffer. After all dilutions were complete, the diluted RNA was added to the protein dilutions and incubated for 1 hr on ice. Before the incubation was complete, the wells on the gels were marked using a VWR marker to help with gel loading. Cold 1X TBE and -70 °C ice packs were used to set up the electrophoresis apparatus. After incubation, 20 µL of protein-RNA mixtures was loaded into the wells of cold native 5.5% acrylamide gels (19:1 acrylamide:bis-acrylamide) containing 5% glycerol. A 50% bromophenol blue-sucrose mixture was used as a dye for gel orientation. Gels were then run at 200 V for 15 min. Meanwhile, Hybond N+ membranes were equilibrated using cold 1X TBE for 10 min.

The RNA was transferred from the gel onto the Hybond N+ using a Bio-Rad Trans-Blot® Turbo™ Transfer system. Whatman #3 paper was cut in similar dimensions as the HyBond N+ membrane, and 6 pieces used for each membrane transferred (3 pieces below and 3 pieces above the transfer sandwich). Transfer was performed at 25 V and 1.0 A for 30 min. The RNA was crosslinked to the HyBond membrane using a Hoefer UVC 500 UV crosslinker for 45 sec. at 120 mJ. The membrane was either set aside to fully dry or detected immediately after crosslinking.

Detection of biotinylated RNA

The following detection protocol was used to detect biotinylated RNA in either the biotinylation efficiency experiment or transferred EMSAs. It was carried out for either one or two membranes at a time. To begin detection, the nucleic acid detection buffer and 4X wash buffer were placed in a 37 °C water bath for 3 min. while the substrate equilibration buffer was left at room temperature. Pincers were always used when handling the membrane, and no other direct contact was made with the membrane. To re-equilibrate a dried membrane, it was placed in a "RNA only" container and washed with 20 mL of nucleic acid detection buffer for 15 min. with gentle shaking. While the membrane was shaking, 20 mL of nucleic acid blocking buffer was mixed with 67 µL of a solution streptavidin-fused horseradish peroxidase and inverted 3 times. After the 15 min., the nucleic acid detection buffer was removed and placed in a 50 mL conical tube to be reused 2 more times. The buffer with streptavidin+horseradish peroxidase was then poured into the membrane and left for 15 min. with constant shaking. During this time period, the 4X wash buffer was diluted to a working 1X wash buffer with a total volume of 160 mL (*i.e.*,

40 mL of 4X wash buffer + 120 mL MQ water). After the 15 min., the buffer+ streptavidin+ horseradish peroxidase was discarded in a waste container, and 20 mL of 1X wash buffer was added for 5 min. The membrane was washed 2 more times with 1X wash buffer, and then with 20 mL of substrate equilibration buffer for 5 min. Each membrane required 5 mL of a 1:1 combination of luminol/enhancer solution (termed Substrate Working Solution). After 5 min., one of the corners in lower part of the membrane was used to gently blot the membrane with a paper towel. The membrane was then placed face up in a sheet on Saran Premium Wrap. Before the membrane would dry up, 5 mL of Substrate Working Solution were poured onto the membrane creating and adhesion surface which was left untouched for 5 min. The membrane was then carefully blotted again and placed in a dry sheet of Saran Premium Wrap, which was then folded so the membrane would be protected on both sides. Detection of biotinylated RNA and SDS gels was possible by using a ChemiDoc XRS+ imaging system which uses CCD cameras to capture signal intensity and store the data as an electronic file. Unquantified files were stored and backed up in the lab external hard drive.

Quantification and curve-fitting to determine $K_{D,app}$

Quantification for biotinylation efficiency was done qualitatively by comparing single signals emitted by the tested RNA and the biotinylated control RNA.

Quantification of EMSAs was done by using the ImageLab™ software that accompanies the ChemiDoc XRS+ imaging system. Volume boxes were drawn around unbound RNA, and around bound RNA (Figure 18). Quantification of pixels found in the bound and unbound states was then represented as numerical values in an Excel file.

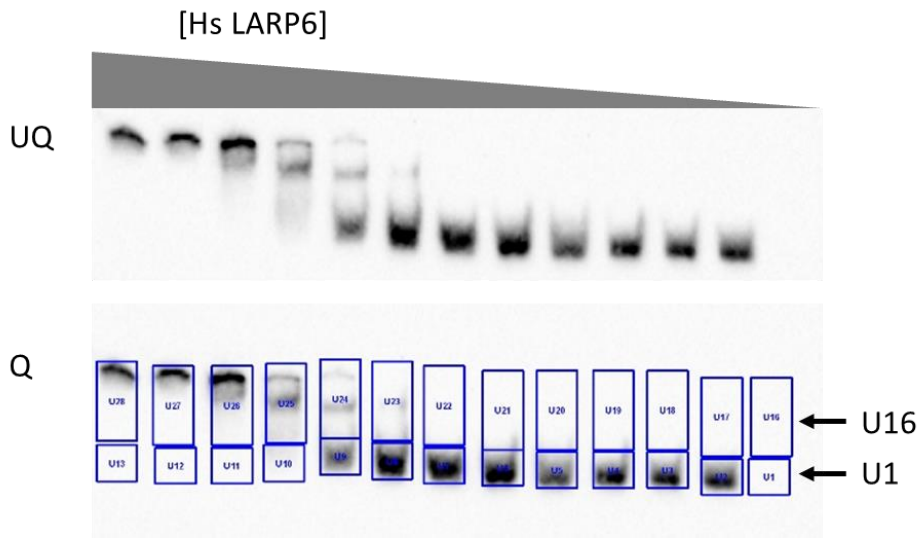


Figure 18. Quantification of incubated biotinylated *HsCOL1A1* to *Hs*LARP6 using Image Lab™ software at 0 mM total salt. UQ represent unquantified data, and Q is quantified data. U1 and U16 represent blanks for background noise.

RNA that was not incubated with *Hs*LARP6 was taken as a reference to represent the migration distance that unbound RNA should have in an EMSA and as the maximum volume signal that unbound RNA should have (*e.g.*, Figure 18, Box U2). A volume box was then drawn around U2 to make it a reference. This volume box was then duplicated and placed around the rest of the columns where unbound RNA was hypothesized to be (Figure 18 U2-U13). The second volume box pertaining to the bound RNA was acquired by measuring where in the gel the highest signal in space was located. In Figure 18 that point would correspond to U28 which was then duplicated and aligned with the volume boxes from the unbound RNA. In addition, membrane background signal was quantified which encompassed the same volume being measure for bound and unbound RNA and subtracted from each measurement (U1 and U16).

For each column, the bound and unbound signal were added to get a total signal value. The fractional saturation was determined by dividing the bound signal for each column by the total signal of that column. A value of 0 fractional saturation represent a completely unsaturated protein and a value of 1 corresponds to a completely saturated protein.

The plotted data were fit with the simplified form of the binding isotherm equation,

$$\frac{[PL]}{[L]_T} = S\left(\frac{[P]_T}{[P]_T + K_{D,app}}\right) + O$$

in which a two-state binding system is used as a model for binding. Provided the concentration of total protein ($[P]_T$) and total ligand ($[L]_T$) are known, and the protein bound to ligand can be measured, the $K_{D,app}$ can be derived. Factors S and O are constants that change from trial to trial. For each ligand at individual salt concentrations, the standard error of the mean (SEM) was calculated by first taking the standard deviation of all apparent K_D values and dividing it by the square root of the number of trials which in every case were ≥3. The SEM was graphed together with the mean of the apparent K_Ds for each ligand at its respective salt concentration.

3. RESULTS

Human LARP6 Purification

*Hs*LARP6 was overexpressed in Rosetta *E. coli* cells and purified on a HisPur™ Ni^{2+}-NTA column. Figure 19 represents a Coomassie-stained gel of fractions collected from the affinity column. A series of washes with high sodium chloride and low imidazole concentrations were done to elute any unwanted proteins that might have bound to the column and reduce non-specific binding (lanes W1- W2B). As expected, the His_6 tag on *Hs*LARP6 was displaced from the nickel column by a higher concentration of imidazole (300 mM), which can be seen in lanes E1-E6 of Figure 19.

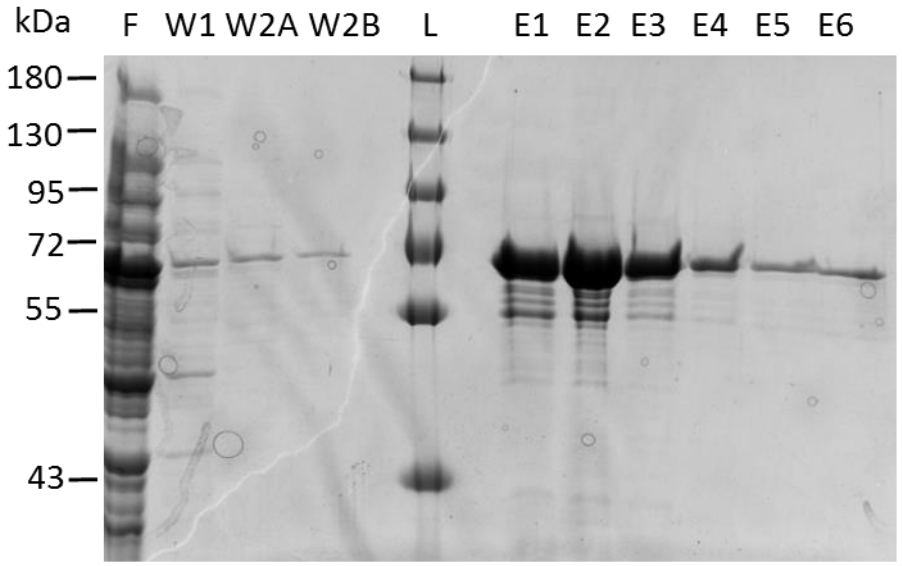


Figure 19. The flowthrough (F) and the elutions following, Wash 1 (W1), Wash 2 a & b (W2A& W2B), and Elution buffer (E1-E6) of one liter recombinant His_6-*Hs*LARP6.

After detecting where *Hs*LARP6 eluted, fractions E1-E4 were concentrated to ~2.3 mL and injected into the Sephadex S200 size exclusion column to further separate any other molecules that might have been bound to the nickel affinity column. Figure 20 shows the elution profile of *Hs*LARP6. In addition, this step also served to exchange the buffer to get rid of the imidazole. The size exclusion column had been previously calibrated to determine elution volumes of various proteins of different molecular weights (black circles above *Hs*LARP6 chromatogram, Figure 20). This served as a baseline to predict the volume where *Hs*LARP6 would elute from the column. According to the column calibration, *Hs*LARP6 (57 kDa molecular weight) is expected to elute at a volume between 80 and 85 mL. However, experimentally the tallest peak was at 71 mL with a molecular weight of 140 kDa.

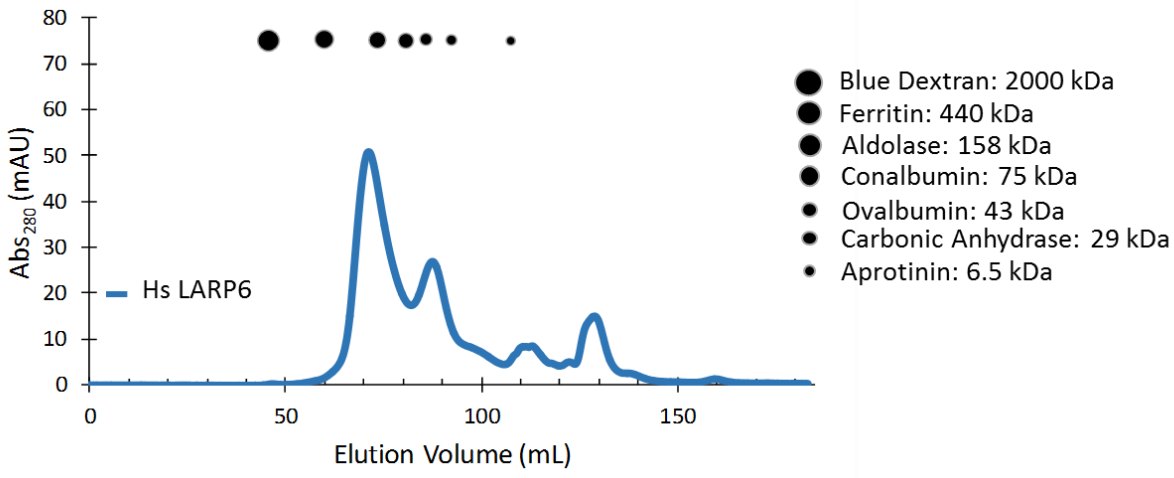


Figure 20. Chromatogram of *Hs*LARP6 elution from Sephadex S200 size exclusion column. Expression culture was one liter of LB.

Analysis by SDS-PAGE (Figure 21) shows that the tallest peak in the chromatogram is where the majority of *Hs*LARP6 is present (Fractions 22-34). The corresponding fractions (Fractions 22-28) were collected and concentrated to 5 µM and stored at -70 °C. From Figure 21 we can see additional bands both under the fractions that were collected and under the fractions that were not. These bands may represent degradation products or other proteins molecules that were bound to *Hs*LARP6.

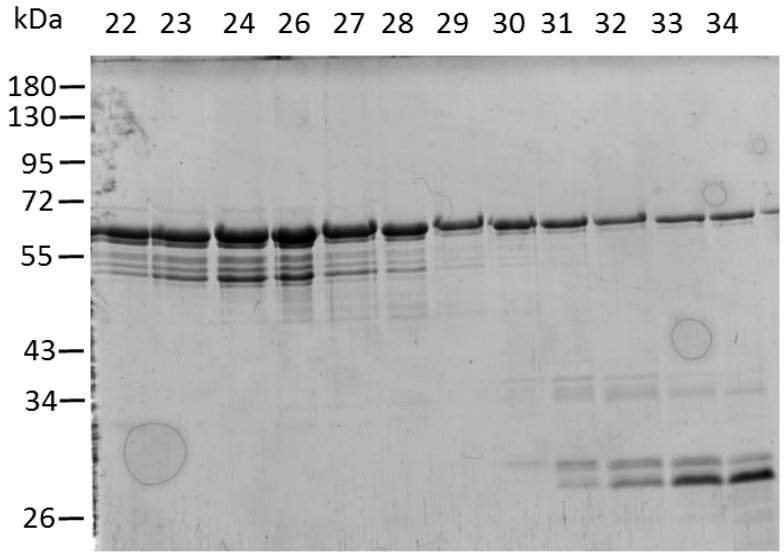


Figure 21. SDS denaturing gel showing fractions 22-34 post size exclusion column corresponding to elution volumes of 65-95 mL, purified from *Hs*LARP6 expressed in one liter of LB.

After expressing *Hs*LARP6 in one liter of LB, we decided to test whether two liters could be expressed and purified with the established protocol. In addition to being able to have a higher concentration of protein, the resulting chromatogram could be compared to the one liter chromatogram. Just as in the previous expression, most of the *Hs*LARP6 eluted from the affinity column in fractions E1- E4, with a small quantity of protein eluting in the initial washes (Figure 22).

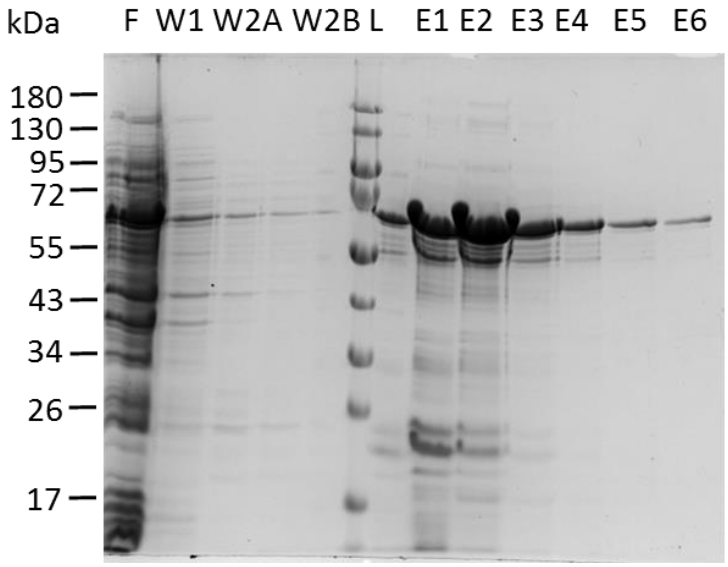


Figure 22. The flowthrough (F) and the elutions following, Wash 1 (W1), Wash 2 a & b (W2A& W2B), and Elution buffer (E1-E6) of two liter expression culture of recombinant *Hs*LARP6.

The S200 elution chromatogram of two liters of expressed *Hs*LARP6 can be seen in Figure 23, in which the tallest peak occurs at an elution volume of 60-75 mL. Fractions 21—33, corresponding to elution volumes of 65—90 mL, were again analyzed via SDS-PAGE and stained using Coomassie blue (Figure 24). Fractions 21—30 were collected due to the abundance of protein and concentrated to 8.5 µM and stored at -70 °C.

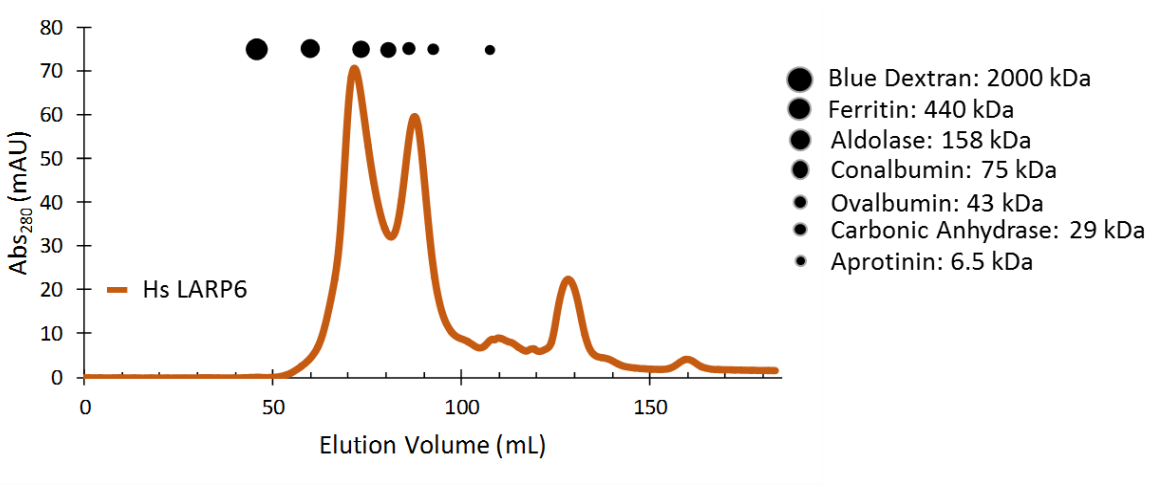


Figure 23. Elution profile of Sephadex S200 size exclusion column for *Hs*LARP6 expressed in two liters of LB.

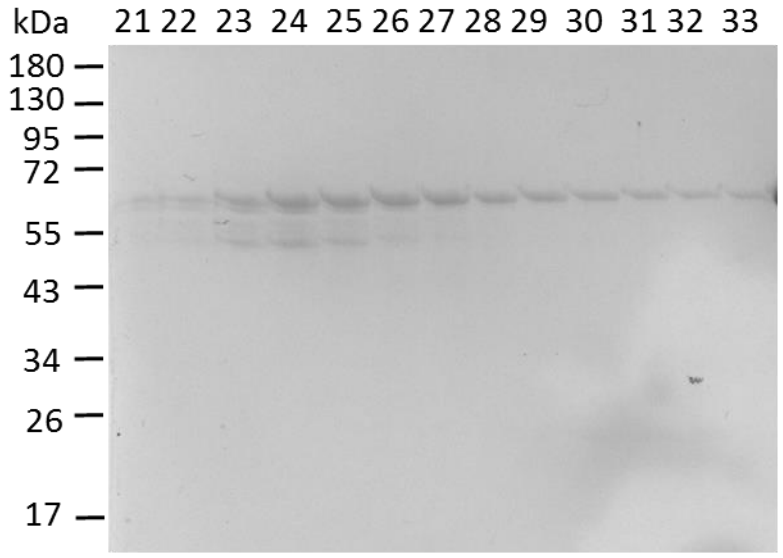


Figure 24. SDS denaturing gel showing fractions 21-33 post Size Exclusion Column corresponding to an elution volume of 65-90 mL from *Hs*LARP6 expressed in two liters of LB.

Between the two purifications, a similar degree of protein content can be seen in both SEC elutions (compare Figures 20 and 23). In both, fractions 22—26 were collected, concentrated, and stored for later use in both purifications. Similarly, an overlay of the S200 chromatograms of the 1 L (blue line) and 2 L (orange line) expressions show a similar shift in predicted elution volume of *Hs*LARP6 (Figure 25).

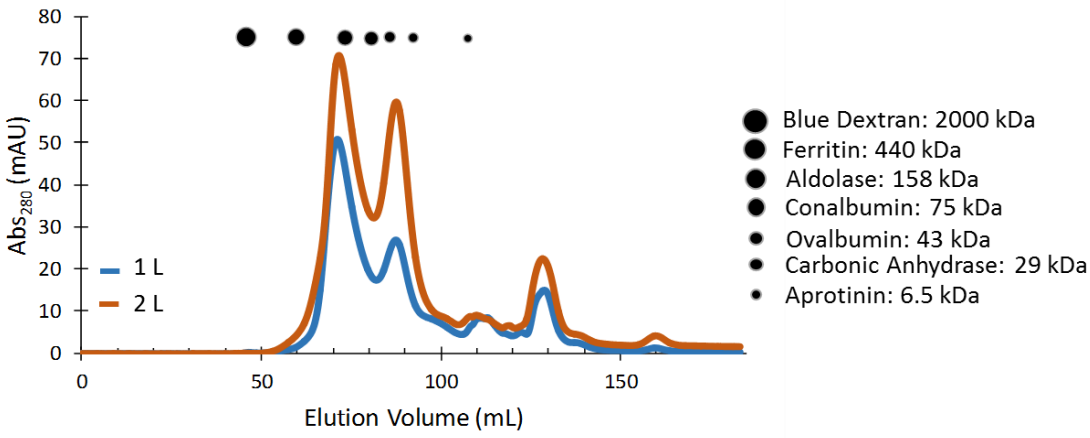


Figure 25. Overlay of elution chromatograms from Sephadex S200 size exclusion columns for both expression conditions of *Hs*LARP6.

RNA Biotinylation

Tagging of RNA with a radioisotope to study binding through the use of EMSAs can often cause changes in binding rates which reflect equilibrium.[37] Recently, methods have been developed that tag the 3' end of RNA without radioactivity and allow for similar sensitivity with a more eco-friendly approach to perform binding studies. In these experiments, biotinylated cytidine bisphosphate was conjugated to the 3' end of each putative RNA ligand through the use of T4 RNA ligase, as described in Materials and Methods. The biotinylated RNA was then analyzed via dot blots by reacting the biotin with

a horseradish peroxidase+streptavidin complex which resulted in quantifiable chemiluminescence.

Each experimental RNA was serially diluted, as can be observed in Figures 27-31. Each control was diluted to make a visual table which could serve as a standard of different degrees of percent biotinylation. Figure 27-31 show the biotinylation efficiencies of *HsCOL1A1*, *HsCOL1A2*, *HsCOL3A1*, *HsCOL5A2*, and *RNA Str8*.

For *HsCOL1A1* (Figure 26 C), 2 µL of biotinylated RNA were spotted in a semi-wet Hybond -N+ membrane and crosslinked at 120 mJ for 45 seconds. Columns A, B, and C contain a series of controls (75%, 50%, and 25% biotinylated RNA, respectively) that were serially diluted in a 1:2 ratio (rows 1-5) to create wide range of RNA biotinylation. The experimental RNA ligand *HsCOL1A1* (column D) was also serially diluted. A control RNA was also biotinylated under the same conditions and serially diluted (column E).

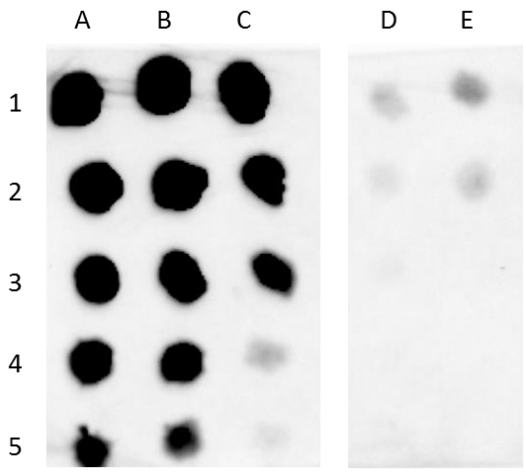


Figure 26. Biotinylation efficiency of *HsCOL1A1* (D) compared to a series of positive biotinylated RNA control dilutions (A-C) with an exposure of 23 seconds using a ChemiDoc XRS+ System. Control RNA (E) was also spotted in membrane. 1:2 dilutions were made going down (from 1 to 2, from 2 to 3, etc.) for all RNAs.

Similar analysis was carried out for the biotinylation of *HsCOL1A2* (Figure 27, column C). Columns A and B contained biotinylation controls (50% and 25%, respectively) that were serially diluted (rows 1-5). In addition to the biotin ligated RNA controls, a non-biotin labeled RNA was biotinylated under the same conditions as *Hs COL1A2* (Figure 27, column D). The positive controls in Columns A and B show considerable smearing, because the spotted solutions were not allowed to fully dry before moving the membrane.

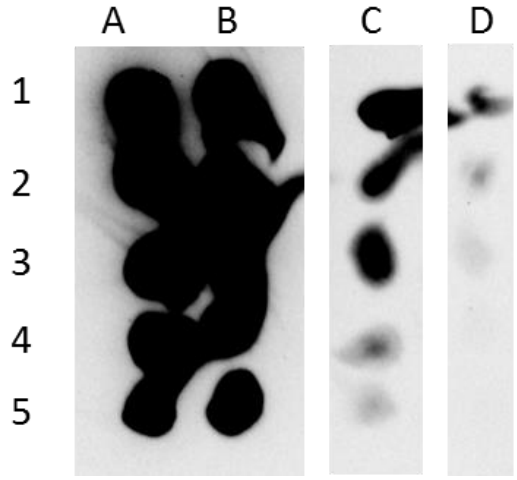


Figure 27. Biotinylation efficiency of *Hs COL1A2* (C) compared to a series of positive biotinylated RNA control dilutions (A-B) with an exposure of 180 seconds using a Chemi Doc XRS System. Control RNA (D) was also spotted in membrane. 1:2 dilutions were made going down (from 1 to 2, from 2 to 3, etc.) for all RNAs.

For the biotinylation of *HsCOL3A1* (Figure 28), columns A and B again contained control biotinylated RNA (50% and 25%, respectively) that were serially diluted (rows 1-5) to create varying concentration of RNA biotinylation. *HsCOL3A1* was serially diluted in Column C, and a non-biotin labeled RNA was biotinylated under the same conditions as *HsCOL3A1* (Column D). Even with a 30 second incubation to dry the last 2 µL of biotinylated RNA, some smearing is visible.

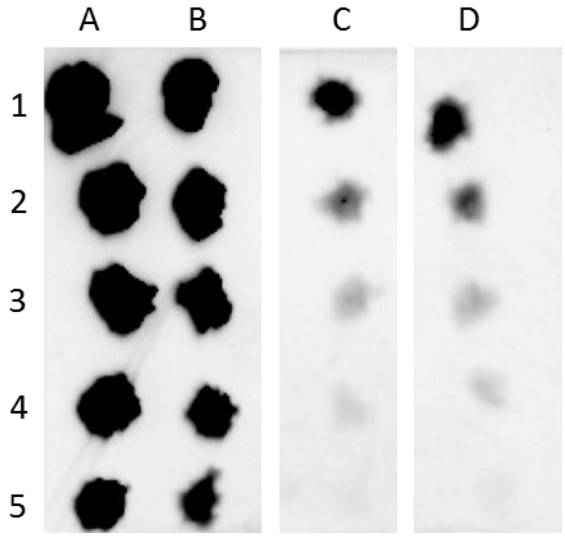


Figure 28. Biotinylation efficiency of *HsCOL3A1* (C) compared to a series of positive biotinylated RNA control (A-B) was evaluated with an exposure of 22 seconds using a ChemiDoc XRS System. Control RNA (D) was also spotted in membrane. 1:2 dilutions were made going down (from 1 to 2, from 2 to 3, etc.) for all RNAs.

For *HsCOL5A2* (Figure 29), columns A and B again contained 50% or 25% biotinylated control RNA, respectively, that were serially diluted (rows 1-5) to create varying concentration of successful RNA biotinylation. In addition to the positive RNA controls, the unlabeled RNA control was biotinylated under the same conditions as *HsCOL5A2* (column D). A 30 second incubation after the last 2 µL of biotinylated RNA was spotted in the membrane was helpful to reduce smearing.

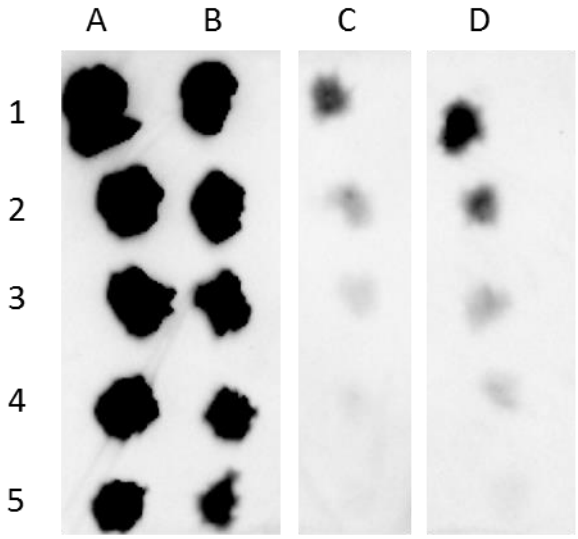


Figure 29. Biotinylation efficiency of *Hs COL5A2* (C) compared to a series of positive biotinylated RNA control dilutions (A-B) with an exposure of 22 seconds using a ChemiDoc XRS System. Control RNA (D) was also spotted in membrane. 1:2 dilutions were made going down (from 1 to 2, from 2 to 3, etc.) for all RNAs.

The biotinylation of *RNA Str8* is shown in Figure 30 (column B). Column A consists of a 25% biotinylated control that was serially diluted (rows 1-5) to create varying concentrations of RNA biotinylation. The unlabeled RNA was biotinylated under the same conditions as *RNA Str8* (column C).

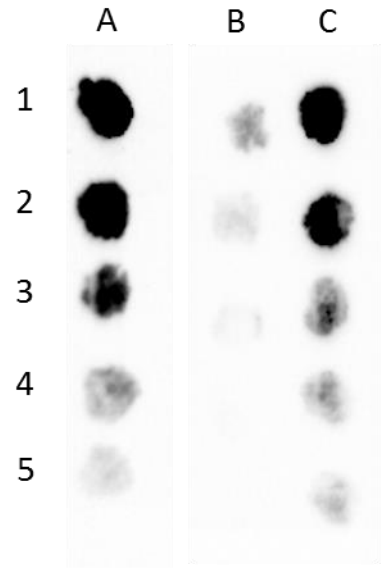


Figure 30. Biotinylation efficiency of *RNA Str8* (B) compared to a positive biotinylated RNA control (A) with an exposure of 20 seconds using a ChemiDoc XRS System. Control RNA (C) was also spotted in membrane. 1:2 dilutions were made going down (from 1 to 2, from 2 to 3, etc.) for all RNAs.

In summary, all five putative RNA ligands for LARP6 were successfully biotinylated, with varying apparent efficiencies. Table 5 lists the estimated biotinylation efficiency of each ligand.

Table 5. Comparison of the different estimated biotinylation efficiencies for the different ligands.

Ligand	Estimated % Biotinylation
Hs 1A1	**3.12**
Hs 1A2	**6.25**
Hs 3A1	**1.56**
Hs 5A2	**1.56**
RNA Str8	**3.12**

LARP6 Binding to *HsCOL1A1*

Electrophoretic mobility shift assays were done to measure binding of *Hs*LARP6 to *HsCOL1A1* at varying salt concentrations. The different binding buffer conditions were obtained by diluting the 10X salt buffers (Table 4) to a 1X concentration with 15% glycerol. All buffers contained a base of 20 mM KCl, and ionic strength was increased by the addition of NaCl. Therefore, "0 mM NaCl" is actually 20 mM total salt concentration due to the initial 20 mM KCl present in the binding buffer. Four concentrations of NaCl were assayed: 0 mM, 50 mM, 100 mM, and 130 mM; because of the 20 mM KCl present in all buffers, the final ionic strength of these buffers is 20 mM, 70 mM, 120 mM, and 150 mM, respectively.

Figure 31 shows representative EMSAs of each binding reaction under each of the salt concentrations. These images were chosen based on how close their calculated apparent dissociation constants were to the mean of all replicates ($n > 3$) and how well they represented the rest of the EMSAs in the group. Briefly, in each trial, increasing concentrations of *Hs*LARP6 were incubated for one hour at 4 °C with 250 pM RNA and then separated on 5.5 % native polyacrylamide gels. RNA was transferred to a Hybond N+

membrane and crosslinked at 120 mJ for 45 seconds. Detection of biotin was performed by binding HRP-conjugated streptavidin and then chemically reacting with a luminol-peroxide solution.

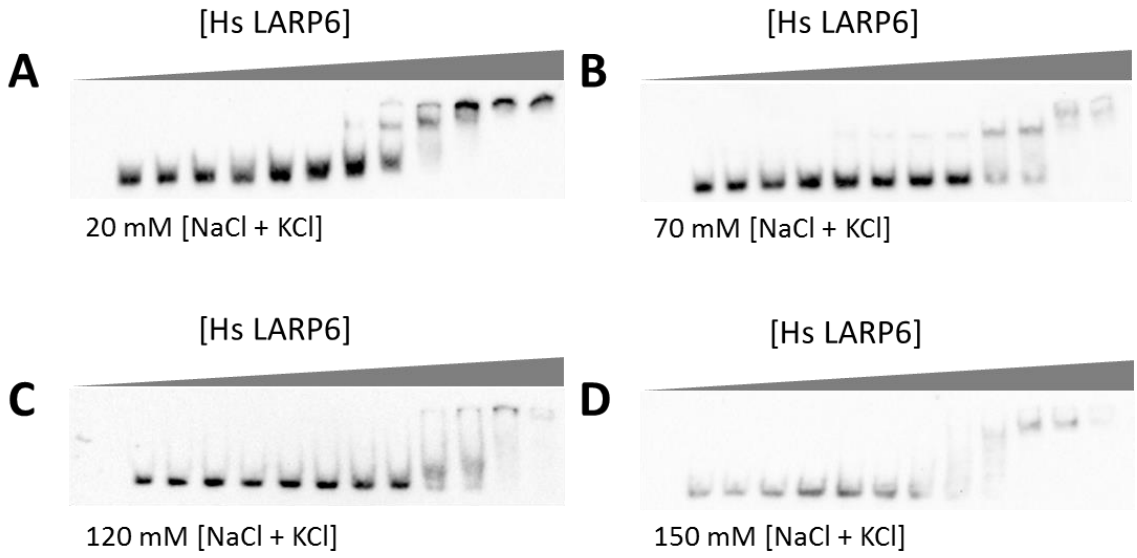


Figure 31. Representative electrophoretic mobility shift assays of *Hs*LARP6 bound to *HsCOL1A1* at 20 mM (A), 70 mM (B), 120 mM (C), and 150 mM (D) total salt.

The intensity of the signals recovered from each EMSA was then quantified using the ImageLab™ software as described in Materials and Methods. Each column in an EMSA had two components, bound and unbound RNA. The plot for the quantitated EMSAs in Figure 31 can be seen in Figure 32. In some cases, the concentration of *Hs*LARP6 was changed to include more data points in the inflection point of the sigmoidal curve (for example, as in the 150 mM total salt concentration).

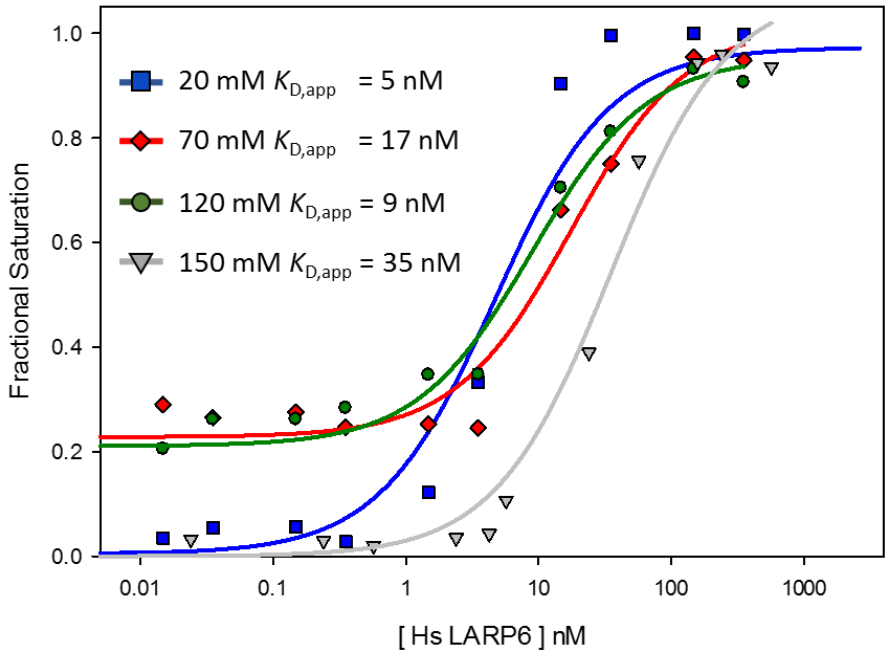


Figure 32. Plot of fractional saturation against protein concentration for representative *HsCOL1A1* EMSAs using SigmaPlot. Data were fitted with the simplified form of the binding isotherm equation as described in Materials and Methods.

To determine the effect that salt concentration has on binding affinity of *Hs*LARP6 to *HsCOL1A1*, the means of the apparent dissociation constants ($K_{D,app}$) obtained from all EMSA trials were plotted against the salt concentration (Figure 33). This plot was then fitted with a linear regression line to explore the correlation between the two. The means for 20, 70, 120, and 150 mM total salt concentration were 9, 57, 17, and 37 nM respectively.

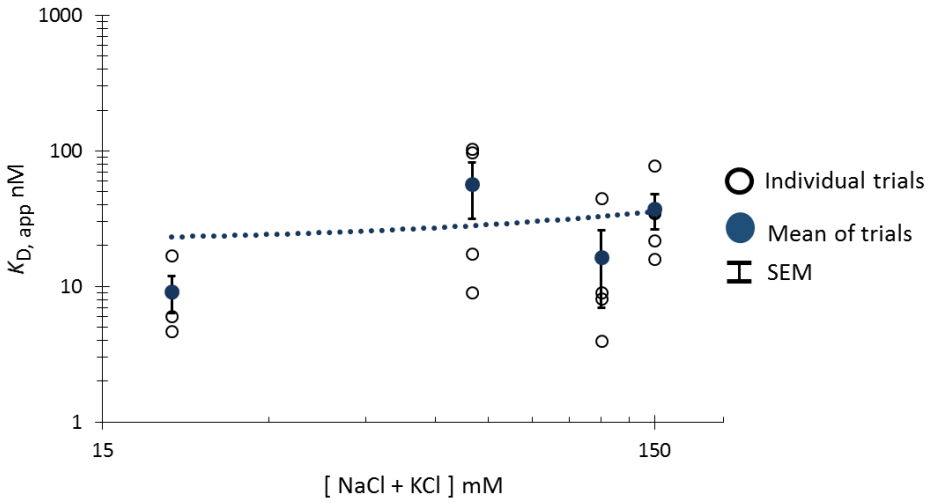


Figure 33. Relationship between increasing salt concentration and $K_{D,app}$ (≥3 individual trials) of *Hs*LARP6 bound to *HsCOL1A1*. Open circles represent the individual EMSA trials, while close circles represent the mean of the individual trials at a given salt concentration. The equation of the line is y = 0.0975x + 21.283 with an R^2 value of 0.0663 where y is the $K_{D,app}$ and x is the total salt concentration.

LARP6 Binding to *HsCOL1A2*

The same process for quantification and selection of the most representative EMSAs was followed with *HsCOL1A2* as was done with *HsCOL1A1*. In Figure 34, the transition between the free and bound RNA can be seen through all four salt concentrations.

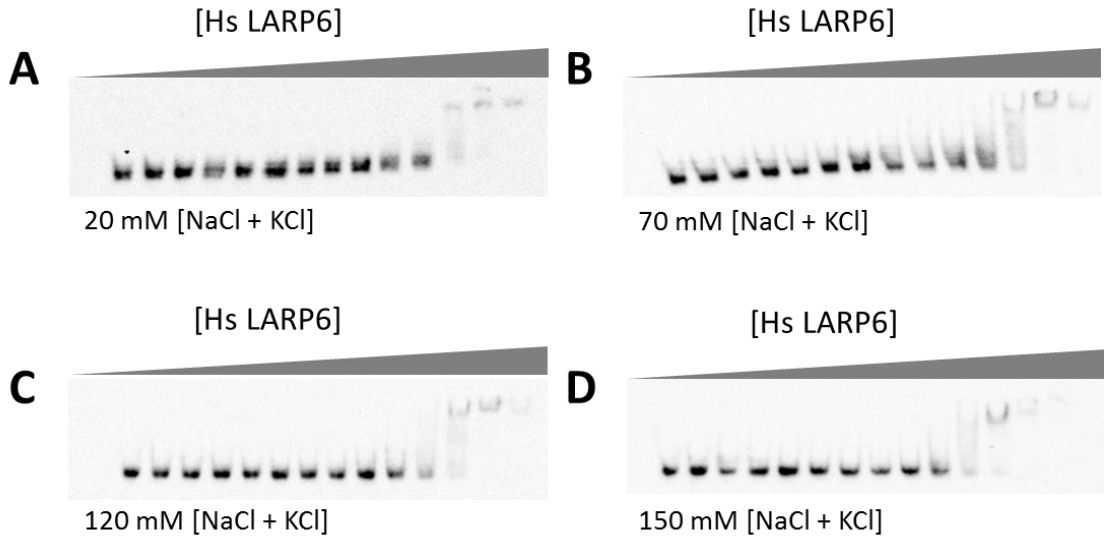


Figure 34. Representative electrophoretic mobility shift assays of *Hs*LARP6 bound to *Hs COL1A2* at 20 mM (A), 70 mM (B), 120 mM (C), and 150 mM (D) total salt.

For *HsCOL1A2*, the apparent dissociation constants appear to remain relatively the same after 20 mM total salt (Figure 36). The quantitated plot for the EMSAs in Figure 34 can be seen in Figure 35. In some cases, the concentrations of *Hs*LARP6 was adjusted to include more data points near the inflection point of the sigmoidal curve (for example, as seen in the 150 mM total salt concentration).

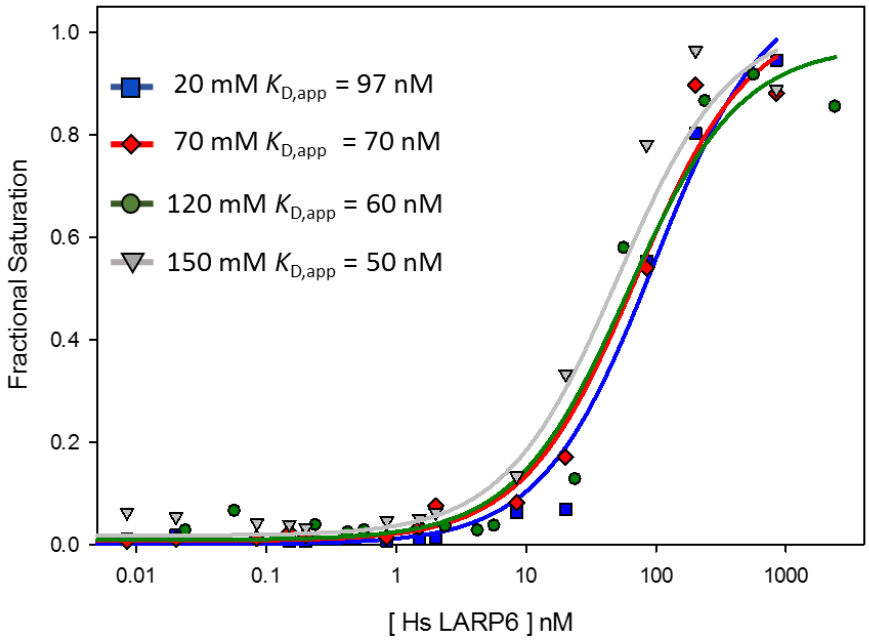


Figure 35. Visual representation of fractional saturation binding curves for representative EMSAs for *HsCOL1A2* using SigmaPlot and fitted with the reduced form of the binding isotherm equation.

To determine the effect that salt concentration has on binding affinity of *Hs*LARP6 to *HsCOL1A2*, the mean of the apparent dissociation constants was plotted as a function of the salt concentration (Figure 36). This plot was then fitted with a linear regression line to explore the correlation between the two. The means for 20, 70, 120, and 150 mM total salt concentration were 71, 55, 74, and 77 nM respectively.

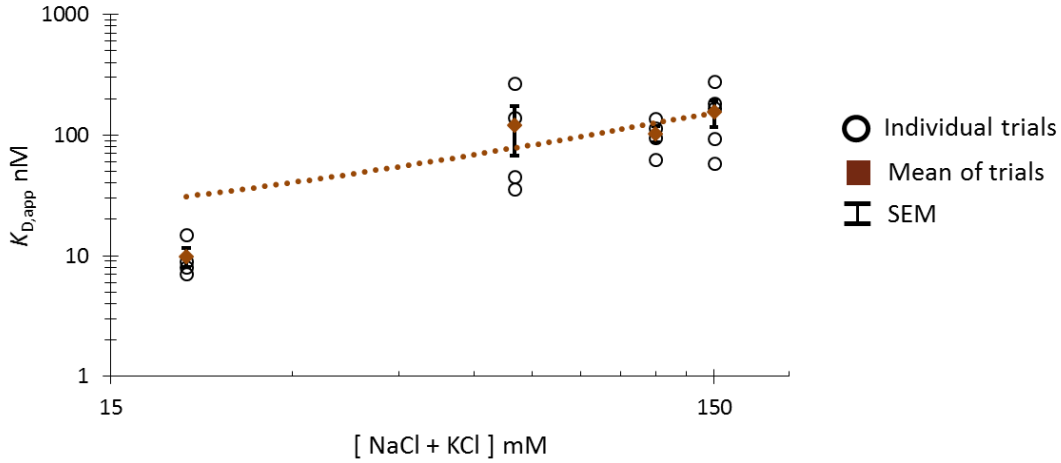


Figure 36. Relationship between increasing salt concentration and $K_{D,app}$ (≥3 individual trials) of *Hs*LARP6 bound to *HsCOL1A2*. Open circles represent the individual EMSA trials, while closed circles represent the mean of the individual trials at a given salt concentration. The equation of the line is y = 0.0794x + 62.094 with an R^2 value of 0.2232 where y is the $K_{D,app}$ and x is the total salt concentration.

LARP6 Binding to *HsCOL3A1*

The same process for quantification and selection of the most representative EMSAs was followed with *HsCOL3A1* as was done with *HsCOL1A1 and HsCOL1A2* (Figure 37).

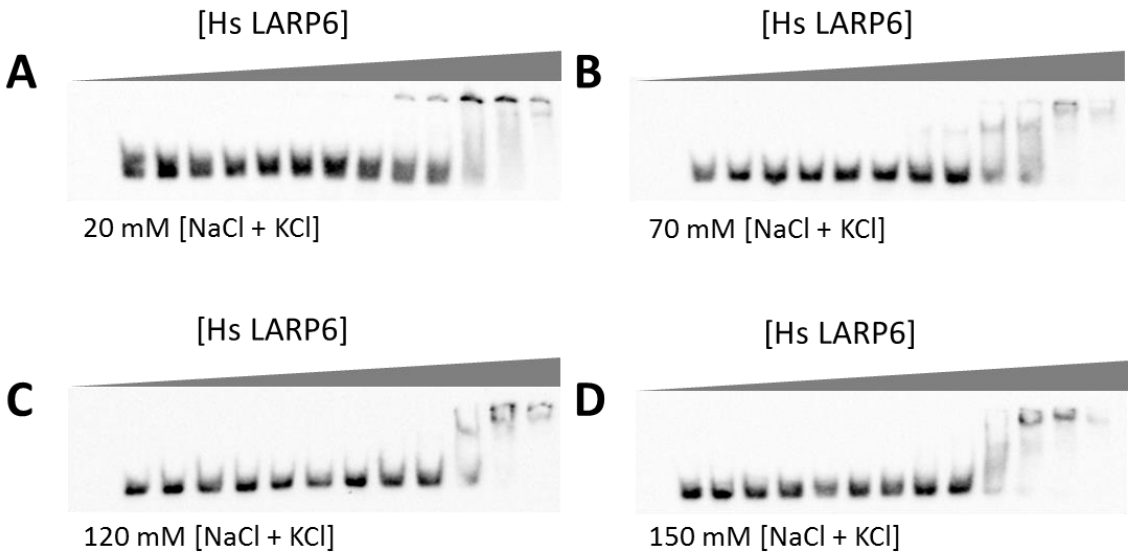


Figure 37. Representative electrophoretic mobility shift assays of *Hs*LARP6 bound to *HsCOL3A1* at 20 mM (A), 70 mM (B), 120 mM (C), and 150 mM (D) total salt.

For *HsCOL3A1*, the apparent dissociation constants appear to remain relatively the same throughout the different salt concentration trials (Figure 38). The quantitation of the EMSAs in Figure 37 is presented in Figure 38.

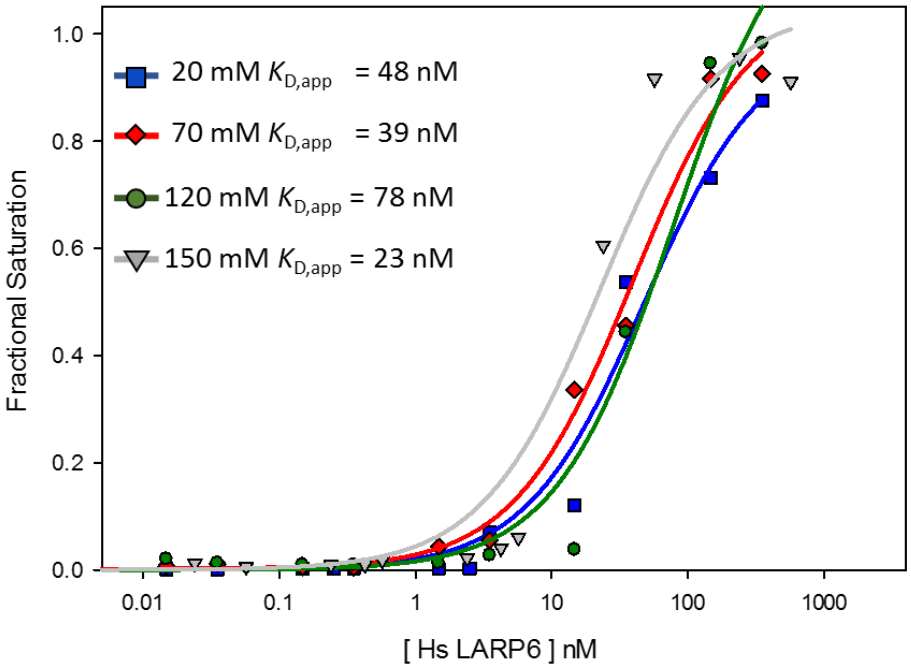


Figure 38. Visual representation of unbound to bound binding curves for representative EMSAs for *HsCOL3A1* using SigmaPlot and fitted with the reduced form of the binding isotherm equation.

To determine the effect that salt concentration has on binding affinity of *Hs*LARP6 to *HsCOL3A1*, the mean of the dissociation constants retrieved from all EMSA trials at a specific salt concentration was plotted against the different salt concentration (Figure 39). This plot was then fitted with a linear regression line to explore the correlation between the two. The means for 20, 70, 120, and 150 mM total salt concentration were 34, 52, 64, and 19 nM respectively.

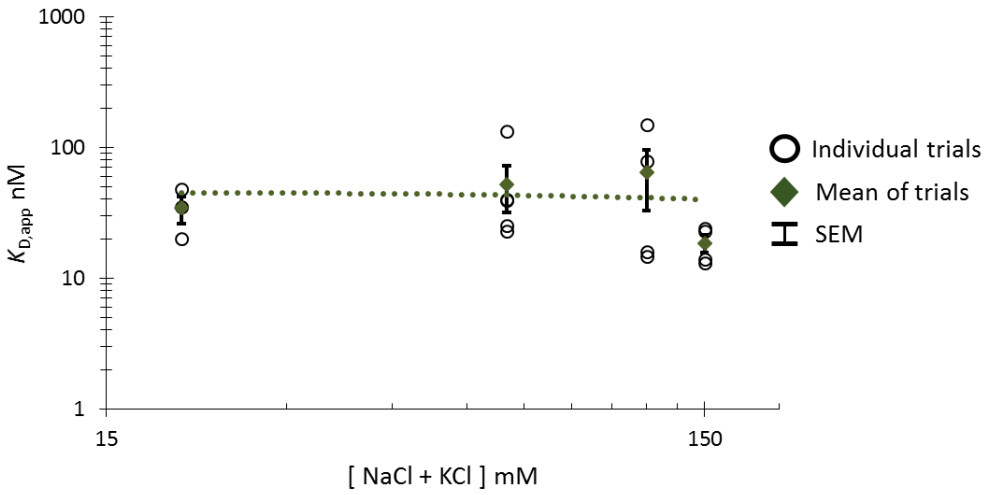


Figure 39. Relationship between increasing salt concentration and $K_{D,app}$ (≥ 3 individual trials) of *Hs*LARP6 bound to *HsCOL3A1*. Open circles represent the individual EMSA trials, while closed circles represent the mean of the individual trials at a given salt concentration. The equation of the line is y = - 0.0406x + 45.969 with an R^2 value of 0.0133 where y is the $K_{D,app}$ and x is the total salt concentration.

LARP6 Binding to *HsCOL5A2*

The same process for quantification and selection of the most representative EMSAs was followed with *HsCOL5A2* as was done with the previous ligands. However, for the *HsCOL5A2* trials, *Hs*LARP6 from the second purification was used (original culture volume of 2 L). In Figure 40, the transition between the free and bound RNA can be seen through all four salt concentrations.

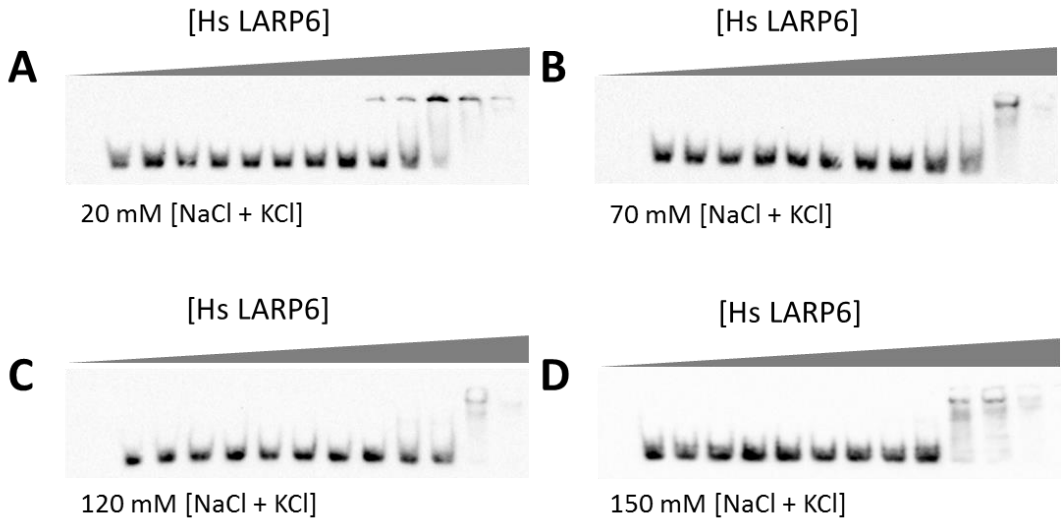


Figure 40. Representative electrophoretic mobility shift assays of *Hs*LARP6 bound to *HsCOL5A2* at 20 mM (A), 70 mM (B), 120 mM (C), and 150 mM (D) total salt.

For *HsCOL5A2*, there is a two-fold increase in apparent dissociation constants with increasing salt concentration (Figure 41). The quantitated plot for the EMSAs in Figure 40 can be seen in Figure 41.

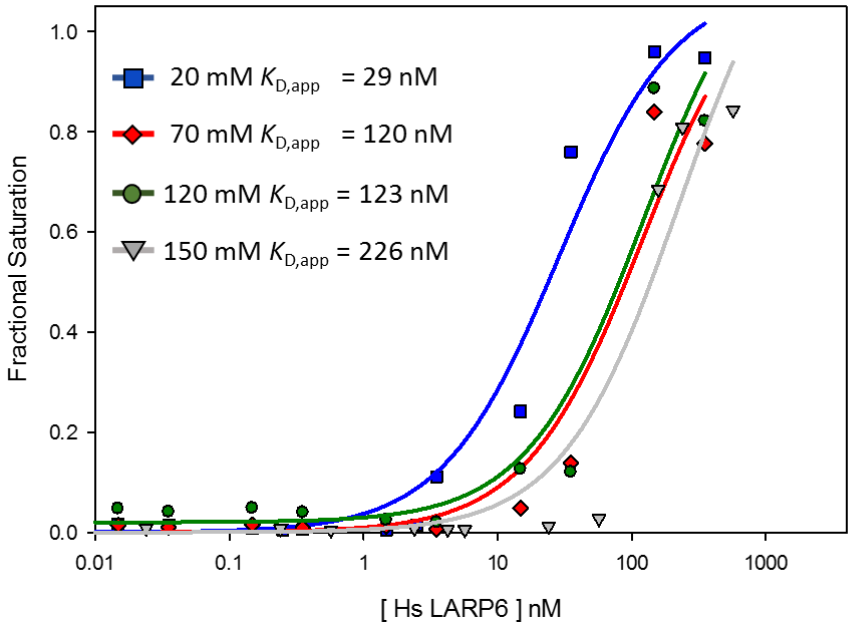


Figure 41. Visual representation of unbound to bound binding curves for representative EMSAs for *HsCOL5A2* using SigmaPlot and fitted with the reduced form of the binding isotherm equation.

To determine the effect that salt concentration has on binding affinity of *Hs*LARP6 to *HsCOL5A2*, the mean of the apparent dissociation constants obtained from all EMSA trials was plotted against the salt concentration (Figure 42). This plot was then fitted with a linear regression line to explore the correlation between the two. The means for 20, 70, 120, and 150 mM total salt concentration were 16, 114, 137, and 216 nM respectively.

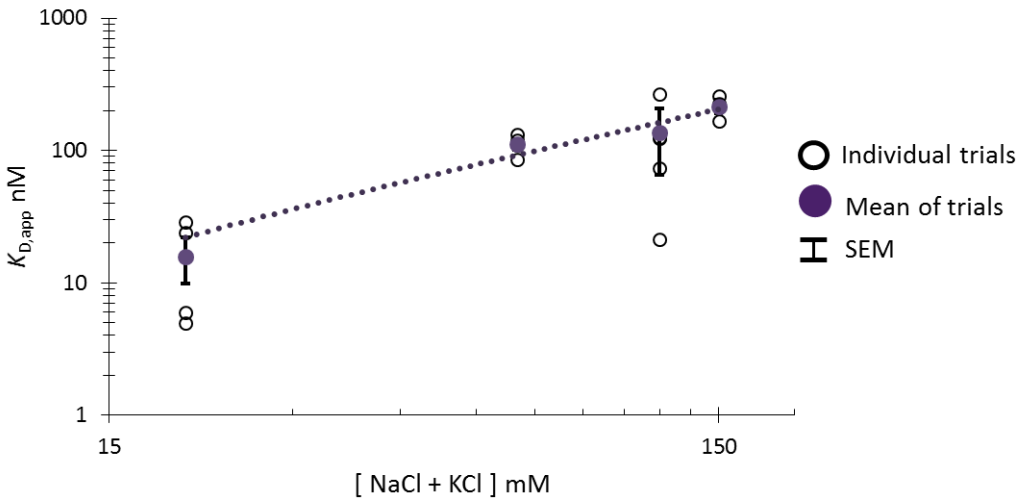


Figure 42. Relationship between increasing salt concentration and $K_{D,app}$ (≥3 individual trials) of *Hs*LARP6 bound to *HsCOL5A2*. Open circles represent the individual EMSA trials, while closed circles represent the mean of the individual trials at a given salt concentration. The equation of the line is y = 1.4044x - 6.1056 with an R^2 value of 0.943 where y is the $K_{D,app}$ and x is the total salt concentration.

LARP6 Binding to *RNA Str8*

The same process for quantification and selection of the most representative EMSAs was followed with *RNA Str8* as was done with *HsCOL1A1*. As with *HsCOL5A2*, for the *RNA Str8* trials *Hs*LARP6 from the second purification was used. In Figure 43 the transition between the free and bound RNA can be seen through all four salt concentrations. Due to an increase in apparent dissociation constants, higher LARP6 concentrations were used at 150 mM total salt concentration.

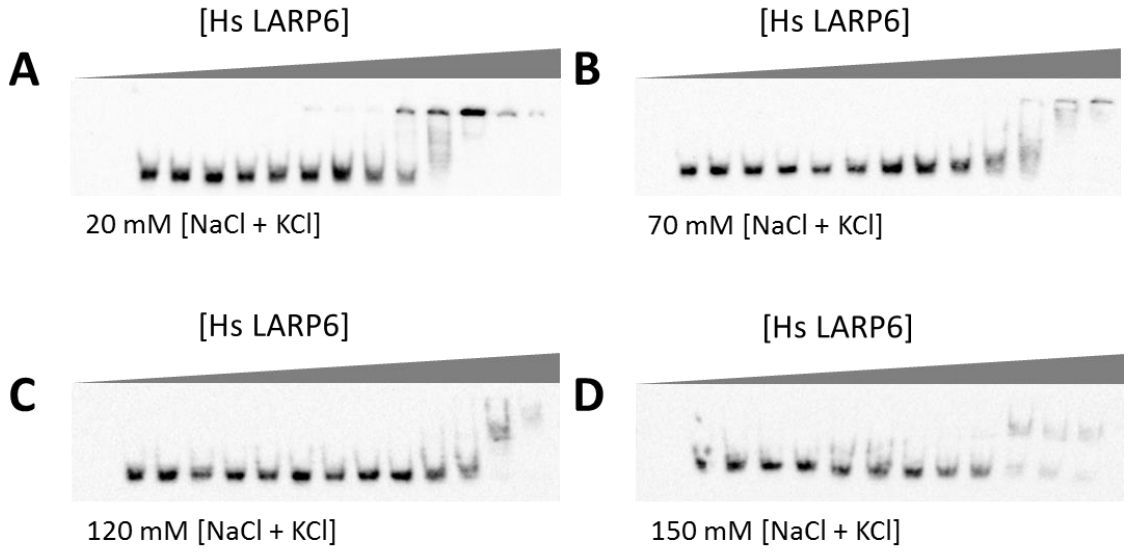


Figure 43. Representative electrophoretic mobility shift assays of *Hs*LARP6 bound to *RNA Str8* at 20 mM (A), 70 mM (B), 120 mM (C), and 150 mM (D) total salt.

For *RNA Str8*, there is a three-fold increase in apparent dissociation constants with increasing salt concentration (Figure 44). The quantitated plot for the EMSAs in Figure 43 can be seen in Figure 44. Furthermore the apparent dissociation constant for RNA Str8 at 150 mM NaCl may be even higher than calculated, due to an absence of a band showing completely bound RNA (Figure 43 D).

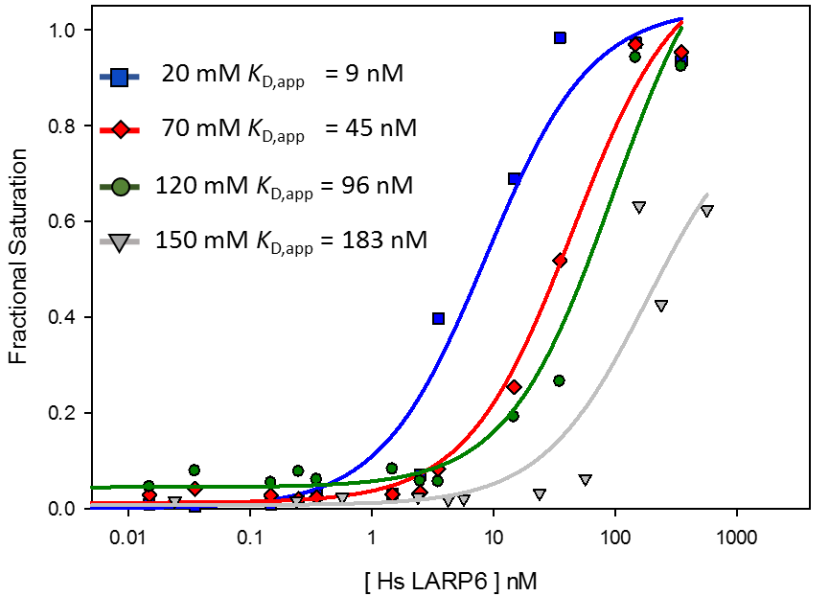


Figure 44. Visual representation of unbound to bound binding curves for representative EMSAs for *RNA Str8* using SigmaPlot and fitted with the reduced form of the binding isotherm equation.

To determine the effect that salt concentration has on binding affinity of *Hs*LARP6 to *RNA Str8*, the means of the apparent dissociation constants ere plotted against the salt concentration (Figure 45). This plot was then fitted with a linear regression line to explore the correlation between the two. The means for 20, 70, 120, and 150 mM total salt concentration were 10, 121, 102, and 155 nM respectively.

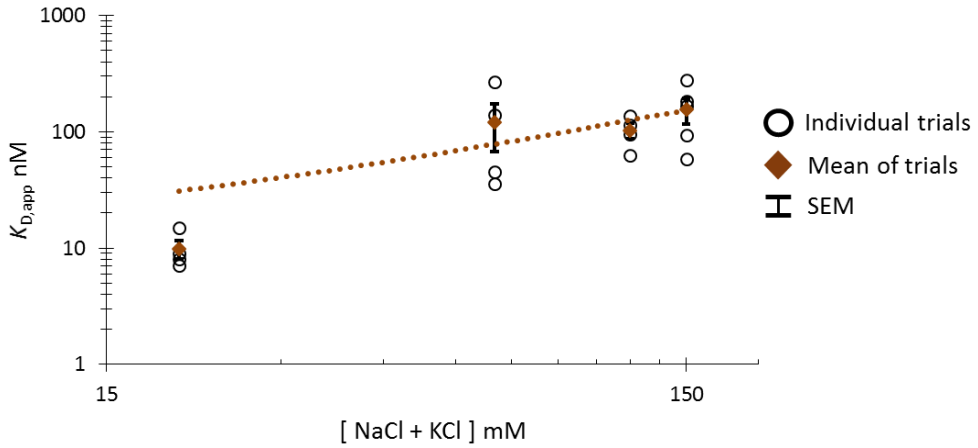


Figure 45. Relationship between increasing salt concentration and $K_{D,app}$ (≥3 individual trials) of *Hs*LARP6 bound to *RNA Str8*. Open circles represent the individual EMSA trials, while closed circles represent the mean of the individual trials at a given salt concentration. The equation of the line is y = 0.9462x + 12.066 with an R^2 value of 0.7549 where y is the $K_{D,app}$ and x is the total salt concentration.

To better understand the results and relationships that were obtained from the salt binding studies, several combination plots were prepared. The ligands *HsCOL1A1*, *HsCOL1A2,* and *HsCOL3A1* were minimally affected by increasing salt concentration (Figure 46). Each individual dot represents the mean of each ligand at their respective salt concentrations. Error bars reflect the standard error of the mean.

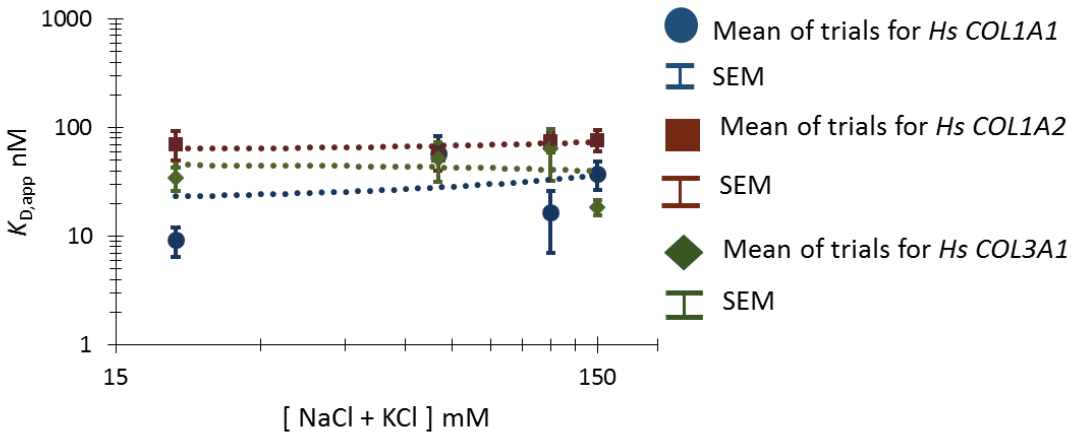


Figure 46. Effects of increasing salt concentration on *HsCOL1A1*, *HsCOL1A2*, and *HsCOL3A1* ligands upon binding to *Hs*LARP6.

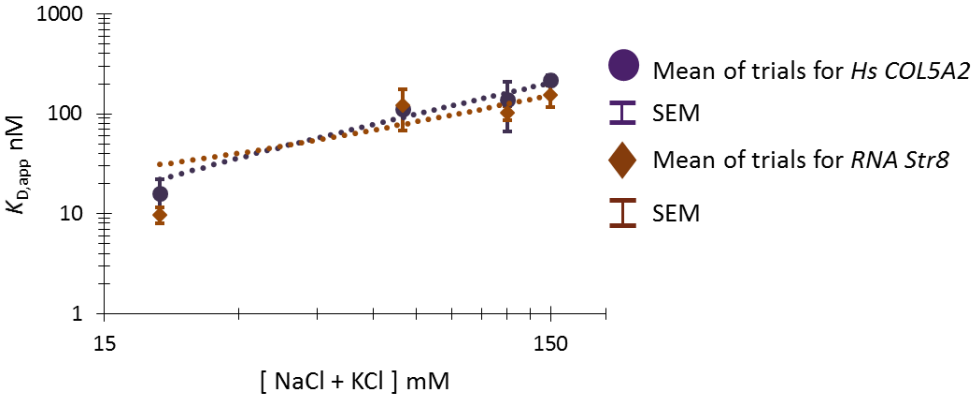


Figure 47. Effects of increasing salt concentration on *HsCOL5A2* and *RNA Str8* upon binding to *Hs*LARP6.

To better show the similar relationship between *HsCOL5A2* and *RNA Str8* under increasing salt concentrations, they were graphed together as seen in Figure 47. The mean for each trial was represented as well as the respective standard error of the mean to show an approximate range of values that could overlap.

Discussion of Results

In the present study, we have successfully purified *Hs*LARP6 with a high yield and minimal degradation. By purifying protein from a 2 L expression culture, the concentration of *Hs*LARP6 can be stored at a concentration of up to ~8 µM and retain its activity. As previously discussed, the elution volumes of the control standards, ovalbumin (43 kDa) and conalbumin (75 kDa), indicate that *Hs*LARP6 should be eluting at a volume between 80 and 85 mL. However, in our study, *Hs*LARP6 eluted from the column at a larger predicted molecular weight, which may be due to the non-globular structure of *Hs*LARP6. The interdomain linker between the La motif and RRM provides flexibility, which can add to the overall radius of the protein. As a result, this could possibly make the protein elute at a higher predicted molecular weight. A high *Hs*LARP6 concentration is desired when performing binding studies, especially when the RNA ligands may bind with low affinity. In this case a higher protein concentration would be needed to see the full RNA retardation using EMSAs.

Biotinylation of RNA ligands for this study were conducted under the same conditions with varying success rates. In this set of experiments the percent biotinylation of each ligand was enough to assess binding. However, a higher percentage of biotinylation would be ideal. EMSAs were run on 6% native polyacrylamide gels, which resulted in smearing and poor differentiation between bands within the gel. Band resolution improved by lowering the polyacrylamide concentration to 5.5%, however, the gel itself was less firm. This posed a problem when transferring the gel onto a membrane using the Transblot Turbo system. By compressing the filter paper before transfer, shifting of the gel during transference to the membrane was significantly decreased.

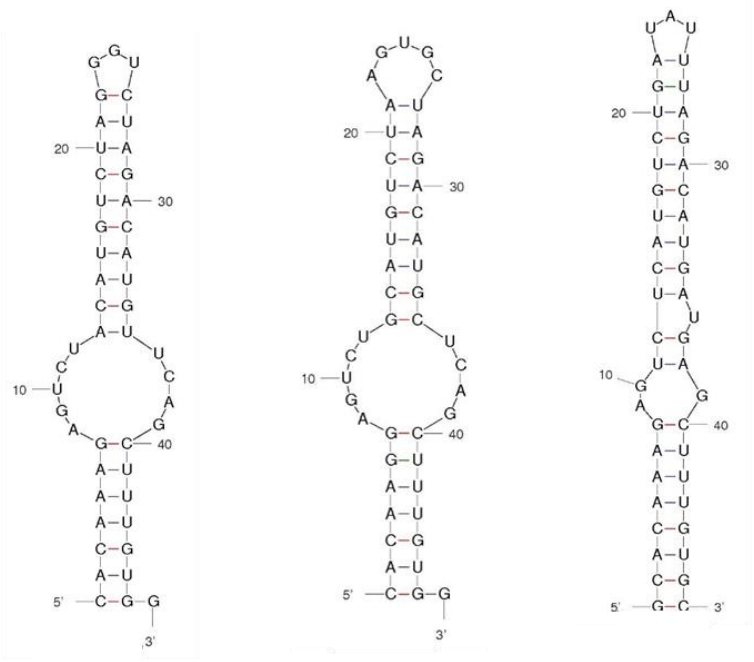
Effects of increasing salt concentration on apparent dissociation constants where graphed on an x-y axis logged scale to more adequately show the relationship between the two variables (comparable to Suryawanshi *et al.*).[38] In these experiments specific binding refers to ligands whose dissociation constants were not affected by salt concentration. For ligands *HsCOL1A1*, *HsCOL1A2* and *HsCOL3A1*, the effects of ionic solvation on both the protein and RNAs seem minimal with very low slope values, which would indicate no effect on binding. In contrast, ligands whose affinity to *Hs*LARP6 was affected by salt concentration are referred to as experiencing non-specific binding. For ligands *HsCOL5A2* and *RNA Str8*, the slope values reflect a strong positive correlation, indicating salt dependency. Currently, there are studies that show that the knockdown of *Hs*LARP6 effects collagen production. However, there are currently no experiments directly showing the mechanism behind this observation. One of the predominant hypothesis is that *Hs*LARP6 denatures the stem loop found in the 5'UTRs, in which the start codon is sequestered. By denaturing the stem loop the start codon is then free for recruitment allowing for the translation of collagen. When comparing the binding affinity of these ligands, it is still unclear whether the affinity that *Hs*LARP6 has to *HsCOL5A2* and *RNA Str8* is significant *in vivo*. The adequate affinity that a protein should have for its ligands greatly depends on each individual protein, therefore it is difficult to pinpoint with certainty the range in which a protein should bind. However, because *RNA Str8* was a random RNA, the nonspecific binding it experiences at physiological conditions could indicate the threshold that *Hs*LARP6 should have to perform its function, which is in the range of 0—100 nM.

4. CONCLUSIONS

At the beginning of this study our primary aim was to: 1) Assess if *Hs*LARP6 binds primarily to a defined secondary structure, and 2) to what extent are ionic interactions responsible for binding. Experimental results led to two main discoveries of this thesis: (1) *Hs*LARP6 appears to be recognizing secondary structure of RNA (Figure 48 and 49), and (2) ionic interactions appear to have a minimal effect on binding to some RNAs but a large effect on non-specific binding to other RNAs (Figure 46 and 47).

Our studies revealed that at low salt concentrations (20 mM total salt), *Hs*LARP6 bound RNA promiscuously in the low nanomolar range. The $K_{D,app}$ values of *HsCOL1A1* and *RNA Str8* which were vastly different, both in sequence and structure at 20 mM total salt, were 9 and 10 nM respectively. This strong degree of binding may have been the result of interactions between the negative phosphate backbone of RNA and a net positively charged protein. It was therefore hypothesized that under these conditions, *Hs*LARP6 interacts with RNA through non-specific ionic interactions. We directly tested this hypothesis through the use of ion solvation. As salt concentration increased, the apparent binding affinity of *Hs*LARP6 for two ligands, *HsCOL5A2* and *RNA Str8*, decreased by a factor of ~2. Thus, the only ligand that has been confirmed *in vivo*, *HsCOL1A1*, seemed to be minimally affected by the change in salt concentration. Similarly, the apparent dissociation constants of ligands *HsCOL1A2* and *HsCOL3A1* were also minimally affected by increased salt concentration (Figure 46). The secondary structures of these ligands (*HsCOL1A1*, *HsCOL1A2* and *HsCOL3A1)* are similar, and have high sequence conservation (Figure 48). Only one secondary structure is predicted at 4 °C for each ligand evaluated in this study.[36] In addition, the ligands have a minimum Gibbs free energy less

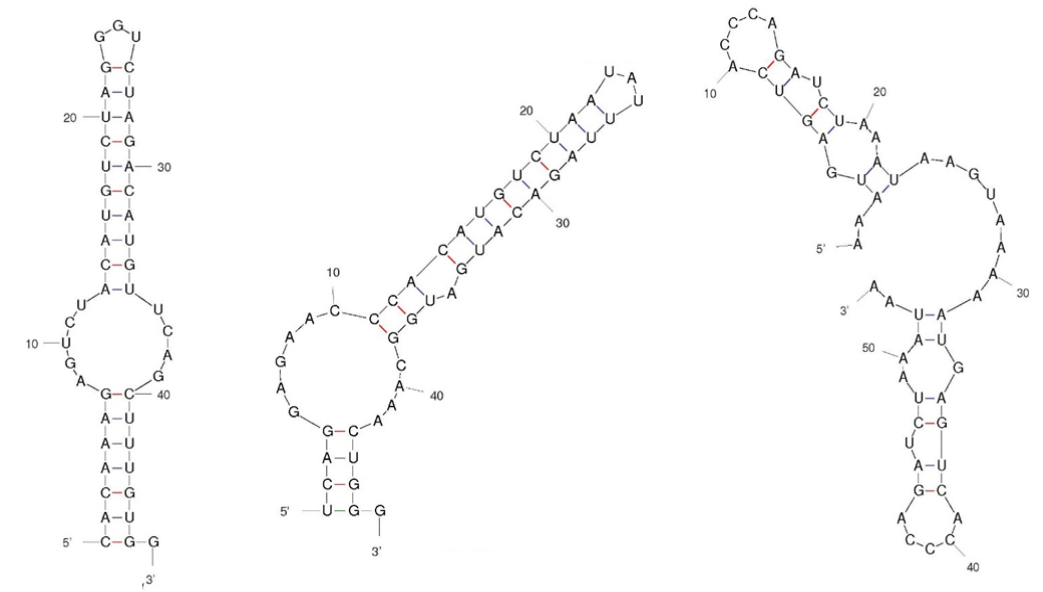
than -10 kcal/mol at 4°C, which indicates considerable stability of the secondary structure.[36] Alignment of the RNA ligands that were minimally affected by salt show a higher degree of similarity (Figure 48) than those that were greatly affected (Figure 49).




```
HsCOL1A1    CACAAAGAGUCUACAUGUCUAGGGUCUAGACAUGUUCAGCUUUGUGG
HsCOL1A2    CACAAGGAGUCUGCAUGUCUAAGUGCUAGACAUGCUCAGCUUUGUGG
            ***** ****** ******** *  ********* ************

HsCOL1A1    -CACAAAGAGUCUACAUGUCUAGGGUCUAGACAUGUUCAGCUUUGUGG
HsCOL3A1    GCACAAAGAGUCU-CAUGUCUGAUAUUUAGACAUGAUGAGCUUUGUGC
             ************ *******    * ******** * *********
```

Figure 48. Comparison between the predicted secondary structures of *HsCOL1A1*, *HsCOL1A2*, and *HsCOL3A1*. Individual alignments between *HsCOL1A1* and *HsCOL1A2* and *HsCOL3A1* are shown to highlight the high degree of sequence conservation.



Hs COL1A1
ΔG° = -33.94 kcal/mol

Hs COL5A2
ΔG° = -28.25 kcal/mol

RNA Str8
ΔG° = -13.78 kcal/mol

```
HsCOL1A1    CACAAAGAGUCUACAUGUCUAGGGUCUAGACAUGUUCAGCUUUGUGG
HsCOL5A2    UCAGGAGAACCCACAUGUCUAAUAUUUAGACAUGAUGGCAAACUGGG
                 ***  * *********   * ******** *         **

HsCOL1A1    CACAAAGAGUCUA-CAUGUCUAGG-GUCUAGACAUG-----UUCAG--CUUUGUGG
RNA Str8    AAAUG--AGUCACCCAGAUCUAAAUAAGUAAAAAUGAGUCACCCAGAUCUAAAUAA
             *     ****   **  ****      ** * ***       ***  **   *
```

Figure 49. Comparison between the predicted secondary structures of *HsCOL1A1*, *HsCOL5A2*, and *RNA Str8*. Individual alignments between *HsCOL1A1* and *Hs COL5A2* and *RNA Str8* are shown to highlight the difference in sequence.

Future Directions

One of the challenges encountered in these experiments was a low biotinylation efficiency. One of the possible reasons why biotinylation was low may be due to RNA structure (*i.e.*, biotin could not be attached due to inaccessibility of the 3' terminus). To increase biotinylation, it may be helpful to further stabilize the RNA structure during the biotinylation reaction. To increase the accessibility of the 3' terminus of the RNA to the RNA ligase, a set of experiments with varying concentrations of KCl during ligation could

be done to measure the effects salt has on biotinylation. The temperature at which the RNA is incubated overnight to be biotinylated could also be changed to further increase or decrease the amount of collision required for the ligation reaction of RNA to biotin to proceed forward.

The data presented here suggest that LARP6 is recognizing RNA ligands on the basis of structure. To ensure that the RNA molecules were not forming intermolecular structures, all RNAs were heated and snapped cooled immediately before every experiment. However, as discussed in the Introduction, ions can significantly contribute to the structure and stability of RNA molecules. To further investigate the role of RNA in these set of experiments, the structure of each RNA should be measured at each salt concentration. One way to do this would be to take the circular dichroism spectrum of each RNA and monitor the absorbance at a range between 260-280 nm. Previous experiments have been able to record CD spectrum for RNA at high salt concentrations.[39]

The majority of the published work on LARP6 focuses on the RNA binding activity of the La Module (Figure 50, red box). Recently, Martino *et al* performed isothermal calorimetry (ITC) binding studies between the La Module of *Hs*LARP6 and *HsCOL1A1*.[1] In their study, the reaction was carried out at 125 mM total salt concentration and the dissociation constant was 48 nM.[1] This is in agreement with the binding activity of the full-length LARP6 binding to *HsCOL1A1*, which was measured by EMSA to have a $K_{D,app}$ = 17 nM at 120 mM total salt. Because *HsCOL1A1* exhibited salt-independent binding behavior similar to the binding activity of *HsCOL1A2* and *HsCOL3A1*, we hypothesize that all three of these ligands are specifically binding to the La Module.

The present study also identified nonspecific, salt-dependent binding activity for the ligands *HsCOL5A2* and *RNA Str8*. One question that needs to be addressed is whether non-specific binding activity is being carried out by the La Module. Due to the minimal characterization of the different regions in *Hs*LARP6 it is difficult to predict whether there is an additional binding site for RNA in *Hs*LARP6. To test this hypothesis, one possible course of action is to purify the isolated La Module and to assess the degree of binding to *HsCOL5A2* and *RNA Str8*. If no binding is observed, then that would indicate that both *HsCOL5A2* and *RNA Str8* are bound non-specifically by another part of the protein that is outside the La Module. This would also indicate a third domain or region responsible for RNA binding. A more direct approach, however, would be to make the F102A point mutation in the highly conserved La motif, which has been shown to result in total loss of RNA binding activity in the La Module (the red X in Figure 50).[1]

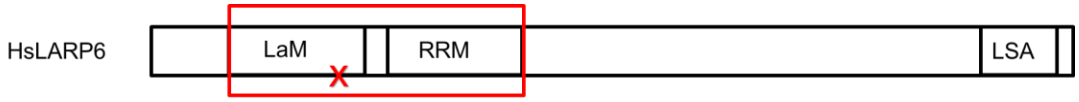


Figure 50. Diagram of the domain structure of *Hs*LARP6. The La Module is boxed in red. The proposed point mutation of F102A is marked by a red X.

The sole identified physiological function of LARP6 is to modulate the production of collagen type I. The *in vitro* binding affinity of *Hs*LARP6 for *HsCOL1A2* and *HsCOL3A1* is comparable to that of *HsCOL1A1*. To test whether these ligands could potentially be targets for *Hs*LARP6 *in vivo*, a knockout of LARP6 in zebrafish could be produced, and the levels of the different collagen proteins in different tissues could be measured.

In this study, we identified two sets of RNA ligands for *Hs*LARP6: one that is unaffected by increased ionic strength, and one that is subject to ionic interactions. Now we know ionic interactions are not solely responsible for RNA binding by *Hs*LARP6. Thus, future studies aimed at elucidating the specific mechanisms by which *Hs*LARP6 binds to RNA may be performed. Understanding these mechanisms will allow the field to also understand how *Hs*LARP6 exerts control over collagen expression. These discoveries could help address diseases that are caused by abnormal collagen production, such as scleroderma.

REFERENCES

1. Martino, L.; Pennell, S.; Kelly, G.; Busi, B.; Brown, P.; Atkinson, R. A.; Salisbury, N. J. H.; Ooi, Z. H.; See, K. W.; Smerdon, S. J.; Alfano, C.; Bui, T. T. T.; Conte, M. R., Synergic interplay of the La motif, RRM1 and the interdomain linker of LARP6 in the recognition of collagen mRNA expands the RNA binding repertoire of the La module. *Nucleic Acids Res* **2015,** *43* (1), 645-660.

2. Dreyfuss, G.; Swanson, M. S.; Pinolroma, S., Heterogeneous Nuclear Ribonucleoprotein-Particles and the Pathway of Messenger-Rna Formation. *Trends Biochem Sci* **1988,** *13* (3), 86-91.

3. Serganov, A.; Nudler, E., A Decade of Riboswitches. *Cell* **2013,** *152* (1-2), 17-24.

4. Zhao, Y. J.; Huang, Y. Y.; Gong, Z.; Wang, Y. J.; Man, J. F.; Xiao, Y., Automated and fast building of three-dimensional RNA structures. *Sci Rep-Uk* **2012,** *2*.

5. Pesole, G.; Luini, S.; Grillo, G.; Saccone, C., Structural and compositional features of untranslated regions of eukaryotic mRNAs. *Gene* **1997,** *205* (1-2), 95-102.

6. Clery, A.; Blatter, M.; Allain, F. H. T., RNA recognition motifs: boring? Not quite. *Curr Opin Struc Biol* **2008,** *18* (3), 290-298.

7. Pyle, A. M., Metal ions in the structure and function of RNA. *J Biol Inorg Chem* **2002,** *7* (7-8), 679-690.

8. Mignone, F.; Gissi, C.; Liuni, S.; Pesole, G., Untranslated regions of mRNAs. *Genome Biol* **2002,** *3* (3).

9. Gerstberger, S.; Hafner, M.; Tuschl, T., A census of human RNA-binding proteins. *Nat Rev Genet* **2014,** *15* (12), 829-845.

10. Watson, J. D.; Crick, F. H. C., Molecular Structure of Nucleic Acids - a Structure for Deoxyribose Nucleic Acid. *Nature* **1953,** *171* (4356), 737-738.

11. Doty, P.; Boedtker, H.; Fresco, J. R.; Haselkorn, R.; Litt, M., Secondary Structure in Ribonucleic Acids. *Proc Natl Acad Sci U S A* **1959,** *45* (4), 482-99.

12. Rutledge, L. R.; Campbell-Verduyn, L. S.; Wetmore, S. D., Characterization of the stacking interactions between DNA or RNA nucleobases and the aromatic amino acids. *Chem Phys Lett* **2007,** *444* (1-3), 167-175.

13. Luscombe, N. M.; Laskowski, R. A.; Thornton, J. M., Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res* **2001,** *29* (13), 2860-2874.

14. Woodson, S. A., Metal ions and RNA folding: a highly charged topic with a dynamic future. *Curr Opin Chem Biol* **2005,** *9* (2), 104-109.

15. Mascotti, D. P.; Lohman, T. M., Thermodynamic Extent of Counterion Release Upon Binding Oligolysines to Single-Stranded Nucleic-Acids. *P Natl Acad Sci USA* **1990,** *87* (8), 3142-3146.

16. Anderson, C. F.; Record, M. T., Polyelectrolyte Theories and their Applications to DNA. *Annual Review of Physical Chemistry* **1982,** *33* (1), 191.

17. Moore, P. B., Structural motifs in RNA. *Annu Rev Biochem* **1999,** *68*, 287-300.

18. Williamson, J. R., Induced fit in RNA-protein recognition. *Nat Struct Biol* **2000,** *7* (10), 834-837.

19. Serganov, A.; Patel, D. J., Molecular recognition and function of riboswitches. *Curr Opin Struc Biol* **2012,** *22* (3), 279-286.

20. Nahvi, A.; Sudarsan, N.; Ebert, M. S.; Zou, X.; Brown, K. L.; Breaker, R. R., Genetic control by a metabolite binding mRNA. *Chem Biol* **2002,** *9* (9), 1043-1049.

21. Breaker, R. R., Prospects for Riboswitch Discovery and Analysis. *Mol Cell* **2011,** *43* (6), 867-879.

22. Gerstberger, S.; Hafner, M.; Ascano, M.; Tuschl, T., Evolutionary Conservation and Expression of Human RNA-Binding Proteins and Their Role in Human Genetic Disease. *Adv Exp Med Biol* **2014,** *825*, 1-55.

23. Ding, J. Z.; Hayashi, M. K.; Zhang, Y.; Manche, L.; Krainer, A. R.; Xu, R. M., Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Gene Dev* **1999,** *13* (9), 1102-1115.

24. Hussain, R. H.; Zawawi, M.; Bayfield, M. A., Conservation of RNA chaperone activity of the human La-related proteins 4, 6 and 7. *Nucleic Acids Res* **2013,** *41* (18), 8715-8725.

25. Bousquet-Antonelli, C.; Deragon, J. M., A comprehensive analysis of the La-motif protein superfamily. *Rna* **2009,** *15* (5), 750-764.

26. Maraia, R. J.; Intine, R. V. A., Recognition of nascent RNA by the human La antigen: Conserved and divergent features of structure and function. *Mol Cell Biol* **2001,** *21* (2), 367-379.

27. Alfano, C.; Sanfelice, D.; Babon, J.; Kelly, G.; Jacks, A.; Curry, S.; Conte, M. R., Structural analysis of cooperative RNA binding by the La motif and central RRM domain of human La protein. *Nat Struct Mol Biol* **2004,** *11* (4), 323-329.

28. Peterlin, B. M.; Brogie, J. E.; Price, D. H., 7SK snRNA: a noncoding RNA that plays a major role in regulating eukaryotic transcription. *Wiley Interdiscip Rev RNA* **2012,** *3* (1), 92-103.

29. Belotti, D.; Paganoni, P.; Manenti, L.; Garofalo, A.; Marchini, S.; Taraboletti, G.; Giavazzi, R., Matrix metalloproteinases (MMP9 and MMP2) induce the release of vascular endothelial growth factor (VEGF) by ovarian carcinoma cells: Implications for ascites formation. *Cancer Res* **2003,** *63* (17), 5224-5229.

30. Stavraka, C.; Blagden, S., The La-Related Proteins, a Family with Connections to Cancer. *Biomolecules* **2015,** *5* (4), 2701-2722.

31. Yang, R. Q.; Gaidamakov, S. A.; Xie, J. W.; Lee, J.; Martino, L.; Kozlov, G.; Crawford, A. K.; Russo, A. N.; Conte, M. R.; Gehring, K.; Maraia, R. J., La-Related Protein 4 Binds Poly(A), Interacts with the Poly(A)-Binding Protein MLLE Domain via a Variant PAM2w Motif, and Can Promote mRNA Stability. *Mol Cell Biol* **2011,** *31* (3), 542-556.

32. Qiang, Z.; Yik, J. H. N., The Yin and Yang of P-TEFb regulation: Implications for human immunodeficiency virus gene expression and global control of cell growth and differentiation. *Microbiology and Molecular Biology Reviews* **2006,** *70* (3), 646-+.

33. Cai, L.; Fritz, D.; Stefanovic, L.; Stefanovic, B., Binding of LARP6 to the Conserved 5 ' Stem-Loop Regulates Translation of mRNAs Encoding Type I Collagen. *J Mol Biol* **2010,** *395* (2), 309-326.

34. Ricard-Blum, S.; Ruggiero, F., The collagen superfamily: from the extracellular matrix to the cell membrane. *Pathol Biol (Paris)* **2005,** *53* (7), 430-42.

35. Dalby, A. B.; Goodrich, K. J.; Pfingsten, J. S.; Cech, T. R., RNA recognition by the DNA end-binding Ku heterodimer. *Rna* **2013,** *19* (6), 841-851.

36. Zuker, M., Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **2003,** *31* (13), 3406-3415.

37. Cassano, A. G.; Anderson, V. E.; Harris, M. E., Understanding the transition states of phosphodiester bond cleavage: insights from heavy atom isotope effects. *Biopolymers* **2004,** *73* (1), 110-29.

38. Suryawanshi, H.; Sabharwal, H.; Maiti, S., Thermodynamics of peptide-RNA recognition: the binding of a Tat peptide to TAR RNA. *J Phys Chem B* **2010,** *114* (34), 11155-63.

39. Martino, L.; Pennell, S.; Kelly, G.; Bui, T. T.; Kotik-Kogan, O.; Smerdon, S. J.; Drake, A. F.; Curry, S.; Conte, M. R., Analysis of the interaction with the hepatitis C virus mRNA reveals an alternative mode of RNA recognition by the human La protein. *Nucleic Acids Res* **2012,** *40* (3), 1381-94.