DIVERGENCE OF THE RNA RECOGNITION MOTIF

IN VERTEBRATE LARP6 PROTEINS

By

Melissa Carrizales, B. S.

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 2019

Committee Members:

Karen A. Lewis, Chair

Raquel Salinas, Co-Chair

Steven T. Whitten, Co-Chair

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

Fair Use

Duplication Permission

DEDICATION

Para mi familia

ACKNOWLEDGEMENTS

I have been lucky to be surrounded by a supportive cohort of people in both my personal and professional life throughout my scientific journey, which I would like to acknowledge. Firstly, I must thank my father, José Carrizales, my siblings Johana, Samantha, and José (Alex) Carrizales and my grandparents, José (Pepé) and Imelda Carrizales for their unwavering support, and for the well-needed laughs during stressful moments. I would like to thank my mentor and PI, Dr. Karen A. Lewis, who not only motivated and inspired me, but taught me to have the confidence to believe in myself. I hope to be like you one day Dr. Lewis, challenge accepted!

Furthermore, I am grateful for my lab team. Each member taught me something new. I would like to thank, Leticia "Lety" Gonzalez, who kindly took me under her wing and showed me the proper way to science. Eliana Peña, José Castro, Eliseo Salas and Lance English who first helped me find my way during my early days of research. Julia Roberts, thank you for all your unconditional support and help, I cannot thank you enough! And of course, Samantha Zepeda, for her support and wisdom and in who I also found a life-long friend. I need you all, keep in touch!

My Bridges family. Thank you. I would like to acknowledge Dr. Raquel Salinas for her guidance, support and dedication to help me reach my goals. Few are fortunate enough to meet someone like Dr. Salinas, and I am so grateful I was. Dr. Nicquet Blake

and Dr. Babatunde Oyajobi, your wisdom and encouragement pushed me to reach higher, thank you. I would also like to thank the Bridges scholars for their support, and for the fun time! In particular, I would like to thank Brandie Taylor and George Parra, for their support and friendship. We began this journey together, struggling all the way through, but we reached our goals. We have been there for each other, and I hope we continue to be on the next phases of our paths.

Lastly, I would like to thank Gabriella “Gaby” Herrera and Juan “Tony” Rodriguez. Gaby has always been a supportive friend and a listening ear. Tony, thank you for everything you have done to help me throughout this process. You both have played an important role, along with everyone else, in helping me become a better version of myself.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

ABSTRACT

The La-related proteins (LARPs) share a conserved RNA-binding unit composed of a La Motif and an RNA Recognition Motif (RRM), which recognizes diverse RNA substrates. LARPs use different surfaces on their RRMs to bind their respective RNA ligands. LARP6, the least understood LARP subfamily, binds to the stem loop (SL) structure found in the 5' untranslated region of collagen types I and III. Previous work showed that the exchange of the LARP6 RRM for the RRM from Genuine La (LARP3) protein disrupts RNA binding activity. Across the vertebrate LARP6 proteins, there are regions of localized divergence within the RRMs, raising the question whether LARP6 RRMs are exchangeable. To understand the role of RRM sequence divergence in the structure and function of LARP6, chimeric proteins were constructed in which the RRMs are exchanged between human and fish species. The chimeric proteins were cloned as three constructs: full-length proteins, C-terminal deletions (ΔCTD), and isolated "La Modules". The chimeras were recombinantly expressed in *E. coli* with N-terminal histidine tags for detection and affinity purification. The full-length and ΔCTD constructs of the human/fish chimeras (consisting of human sequence flanking a fish RRM) required a solubility tag for recombinant expression. In contrast, the human/fish chimeric La Module constructs did not. This result suggests that the N-terminal region may form species-specific interactions with the RRM, that are disrupted by the chimera. The chimeric La Modules were recombinantly expressed and purified. They were

characterized for protein stability via limited trypsinolysis assays and RNA binding activity with electrophoretic mobility shift assays. As predicted, the stability and RNA binding behavior of the chimeras appears to depend on the RRM. Specifically, the chimeras composed of human La Motif and fish RRM are more similar to the native fish La Modules than the native human La Module. This work suggests the sequence divergence within the RRM defines the structural stability and RNA binding behavior of the La Module.

I. INTRODUCTION

RNA Binding by Proteins

RNA often forms secondary structures that act as binding sites for proteins[1]. RNA-binding proteins (RBPs) contribute to different aspects of RNA processing, from the regulation of co- and post-transcriptional gene expression to RNA modification, transport, translation, localization and turnover[1,2]. Interactions between RBPs and their RNA targets are mediated by one or more globular RNA-binding domains (RBDs), which provide both recognition and functional diversity within these proteins[1]. Multiple RBDs can combine to create diverse and versatile RNA-binding surfaces to define the specificity and affinity of the protein[3]. For example, multiple domains can cooperate to create a larger binding surface to recognize longer RNA sequences. Often, this expanded interaction surface increases the affinity and specificity for a target[3,4]. When RNA is not present, the individual domains can move independently, adopt a pre-formed structure similar to the RNA-bound conformation, or form a closed structure that inhibits interaction with RNA[4]. Upon binding the substrate, RBDs can interact with ligands in multiple ways, including contacting different RNA sequence motifs non-cooperatively, or undergo induced conformational changes that mediate RNA contact[4].

The specificity of each, individual RBD is not the only determinant of binding affinity and specificity to the RNA target[3]. The spatial arrangement of the domains relative to one another and the structure of inter-domain linkers are also important factors[3]. The length and rigidity of the linkers between RBDs can affect RNA affinity, as well as influencing whether the protein can bind a single or multiple RNA molecules[3].

Generally, longer linkers are more disordered and exert a smaller effect on affinity than shorter linkers.[3] Shorter linkers can drastically affect RNA binding affinity by restricting the relative orientation of the RBDs[3]. Linker sequences have also shown to be conserved.[3] As a result, RBDs and inter-domain linkers can co-evolve into particular arrangements that exert a specific function[3].

The RNA Recognition Motif

The most common RBD is the RNA recognition motif (RRM), which was first identified when mRNA precursors (pre-mRNA) and heterogeneous nuclear RNAs (hnRNAs) were found in complex with proteins[5]. The RRM is found in all kingdoms of life and is often found as multiple copies within a single protein and/or in tandem with other RNA-binding domains[5]. The canonical RRM is 80-90 amino acids long, arranged into a four-stranded, anti-parallel β-sheet that is flanked by two α-helices[5,6]. This gives the domain an α/β sandwich structure, in which RNA recognition mostly occurs on the β-sheet surface, partly due to the presence of the conserved ribonucleoprotein motifs 1 and 2 (RNP-1 and RNP-2). However, the β-sheet surface is not used to the same extent in all RNA/RRM complexes[5]. One or up to all four of the strands may be used to contact the RNA[5]. Additional conserved residues in the RRM are located in the hydrophobic core of the domain[5]. As described above, some RRMs employ additional structural elements to recognize RNA ligands, including exposed loops, additional β strands, or extensions of the N- and/or C-termini[5,6]. In particular, loops can vary in length and are often disordered in the free form[5]. The RRM is an extremely adaptable domain, in both structure and

function, and further biochemical and structural studies are needed to understand its full complement of structural and functional roles[5].

The La-Related Proteins (LARPs)

The La-related proteins (LARPs) comprise an evolutionarily conserved superfamily of RBPs (Figure 1)[7,8]. These proteins are characterized by a bipartite, RNA-binding unit called the "La Module" [7, 8, 9]. The two domains within the La Module are the La Motif and an RRM, connected by a short linker sequence that varies across the subfamilies[9, 7]. Individual subfamilies are characterized by the domain composition of the C-termini as seen in Figure 1. For example, the founding member of the family, the "genuine La protein" (now called LARP3) has a second RRM, as well as an RNA chaperone domain (RCD)[9,8].

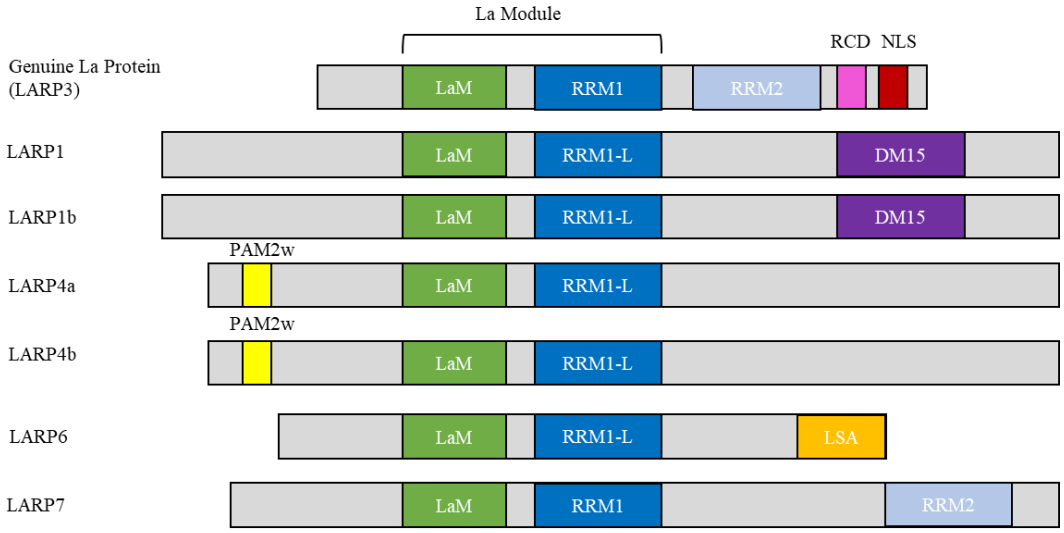

Figure 1. Schematic Representation of the LARPs.
All LARPs share a bipartite La Module composed of a La Motif (LaM) and a RNA Recognition Motif (RRM). Some LARPs carry an additional RRM domain (RRM2). The RRM1 domains show sequence divergence across the LARP superfamily and are denoted as RRM-like domains (RRM-L). In addition, the LARPs have additional domains that distinguishes them into subfamilies and may contribute to different functions. Abbreviations: DM15-repeat containing region (DM15); nuclear localization signal (NLS); an atypical Poly(A) Binding Protein Motif (PAM2w); RNA chaperone domain (RCD); and the

uncharacterized La and S1 associated motif (LSA). (Modified from Stavraka and Blagden, 2015. Reproduced with permission under Creative Commons CC BY license.)[8]

Multiple sequence alignments of the La Modules of the human LARPs revealed higher sequence conservation across the La Motif, in contrast to the RRM[10]. However, the length of the La Modules does differ across subfamilies (Figure 2)[10]. Since the discovery of LARP3, the field has identified LARPs 1, 4, 6, and 7[7, 8, 9, 10]. To maintain the contemporary nomenclature of the LARP subfamily members, which is based on protein topology, LARP2 and LARP5 were renamed to LARP1b and LARP4b, respectively, to reflect their group of origin.[10] Though all LARPs contain a La Module, the La Modules recognize a wide variety of substrates, ranging from pre-tRNAs to untranslated regions of mRNA[8,9]. While other functional domains present in each LARP family may additionally contribute to the functional diversity of these LARPs, the divergence within the La Module suggests that the RRMs have structurally and functionally adapted to execute family-specific binding activity[7, 8, 9 11, 12, 13, 14.]

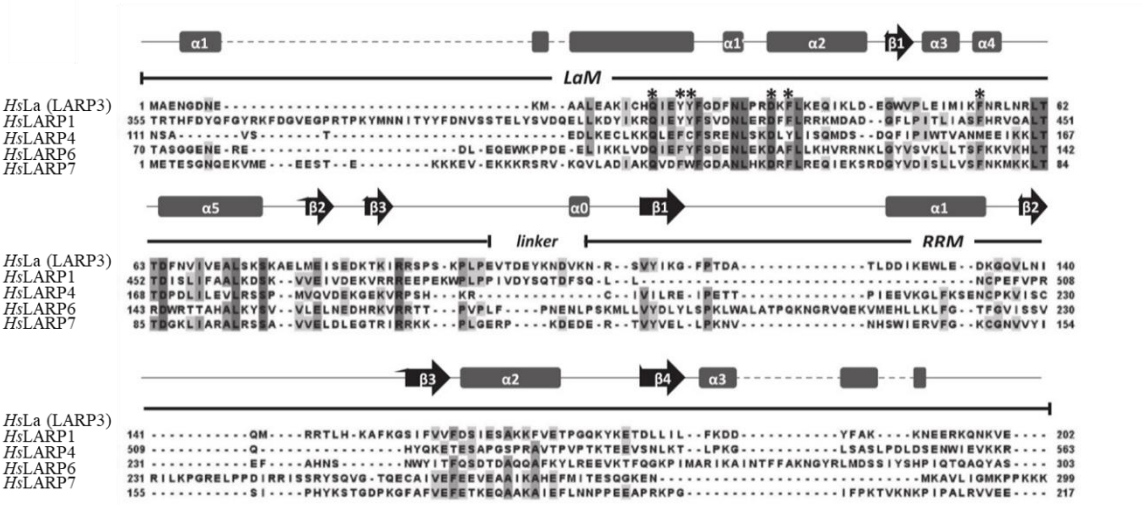

Figure 2. Sequence Alignment of La Modules in Human LARPs. Human La (LARP3) La Module secondary structures are used to visualize regions of divergence between the sequences of LARP La Modules. Conserved residues are shaded. The six highly conserved residues within the La Motif, which are involved in oligoU RNA binding in human LARP3, are indicated with asterisks (Image from Martino, *et al*., 2015. Reproduced with permissions under Creative Commons CC BY license)[10]

The human LARP3 protein was first identified as a protein expressed in autoimmune disorders, such as Sjogren's syndrome, progressive systemic sclerosis (PSS), and systemic lupus erythematosus (SLE)[16, 17]. LARP3 was found in one of the earliest branches from the last common ancestor of eukaryotes in which the La Motif co-evolved with an RRM[7]. LARP3 preferentially interacts with the UUU-3'-OH motif characteristic of RNA polymerase III transcripts, but also associates with other cellular RNAs, viral RNAs, and UUU-3'OH-lacking mRNAs[8, 10, 17, 18, 19]. The most well-established functions of LARP3 are to provide 3' end protection from exonuclease digestion and promote proper folding of pre-tRNAs, though it also acts as a pre-tRNA chaperone to non-UUU-3'-OH bearing substrates and contributes to miRNA processing[9, 17, 18, 19]. The La Motif and RRM cooperate to form a fixed conformational shape around the UUU-3'-OH substrate but move independently in the absence of RNA[18, 19]. This induced V-shaped conformation forms a sequence-specific binding pocket tailored for the UUU-3'OH motif, in which interactions with both the La Motif and RRM occur[18, 19]. Interactions with six residues, crucial for RNA recognition of LARP3, have been identified and are highly conserved across the LARPs[10, 18, 19]. The expected binding surfaces, such as the β-sheet surface of the RRM, were found to not be involved in UUU-3'-OH recognition, but instead interact with the RRM2 domain to recognize different ligand sequences[8, 9]. Thus, LARP3 contains multiple binding surfaces for multiple ligands [9]. LARP7, the closest relative of LARP3, specifically recognizes the UUU-3'-OH motifs of 7SK snRNAs[9, 13]. LARP7 adopts a similar mode of binding to that of LARP3, and likewise contains a similar RRM2 domain[9, 20].

In contrast to LARP3 and LARP7 which are mainly found in the nucleus, LARPs 1, 4, and 6 are mostly cytoplasmic and are mRNA-associated proteins[9, 13, 18, 20, 21]. Both LARP 1 and LARP4 interact with 3'-polyA rich tails in pre-mature mRNAs and are thought to play a role in post-transcriptional regulation[9]. LARP1 and LARP4 also have additional domains (the PAM2 and DM15 domains respectively) that contribute to mRNA binding activity. Notably, neither LARP4 nor LARP6 recognize the short poly(U) substrates that are recognized by LARP3 and LARP7[10, 12].

LARP6 (also called Acheron)

LARP6, originally named Acheron, was first identified in a moth species, where it was found to be induced during programmed cell death of intersegmental muscles[22]. Sequence analysis identified the human ortholog and its relationship to LARP3[22]. Like the moth ortholog, the human LARP6 is mainly expressed in neurons, cardiac muscle, and skeletal muscle, and contains a La motif and a RRM[8, 21, 22]. High-resolution structures of the isolated subdomains of the La Module, and phylogenetic sequence analysis, identified two elements that distinguish a LARP6 protein: a unique helix within the RRM and a novel "LSA motif," at the extreme C-terminus (Figure 1)[7, 9, 10]. The function of this LSA motif has yet to be characterized, but due to its evolutionary conservation it is possible the motif has a functional role[10]. The cellular function of LARP6 is the least understood out of all the LARPs, with only one endogenous ligand identified[14]. Previous studies showed that LARP6 binds to a stem loop (SL) structure that is conserved in the 5'-untranslated region of collagen type I and III mRNAs (Figure 3)[10, 14, 15].

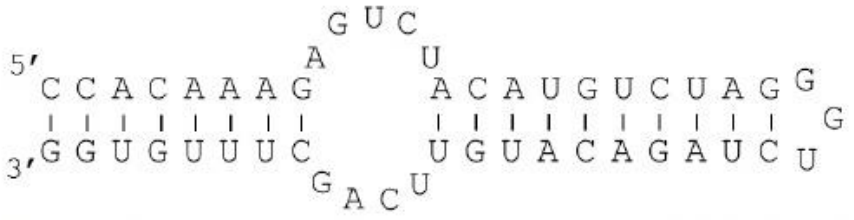

Figure 3. Human Collagen Type I mRNA 5' Untranslated Region (UTR). The sequence around the start codon in the mRNAs that encode for collagen 1(I), 2(I) and 1(III) mRNAs adopts a stem loop secondary structure. Shown here is the sequence from the mRNA encoding collagen type 1.[14]

The LARP6 La Module is necessary and sufficient for binding of the 5'SL sequence[14]. This association appears to recruit factors to promote the translation of these collagen mRNAs, as well as prolong their half-life, to drive the high collagen production in tissue repair and fibroproliferative disorders, which are characterized by uncontrolled synthesis of collagen[18, 10, 21]. Knockdown studies of LARP6 in cell culture showed that in the absence of LARP6, the expression of collagen α1(I) and α2(I) polypeptides were reduced. Therefore, LARP6 plays a principal role in upregulating the synthesis of collagen type I[18]. It has also been observed that LARP6 interacts with other proteins, including the differentiation factor MyoD, which has a role in differentiation of myotubes in myoblasts[23]. The knockdown of LARP6 in zebrafish embryos using morpholinos demonstrated that LARP6 expression is a critical factor in development, but the direct molecular mechanism was not identified[22, 23]. LARP6 has also been seen to interact with the proteins CASK-C and Id (inhibitor of differentiation), though it is unclear if these interactions require the RNA binding activity of LARP6[24]. Together, these observations suggest that LARP6 plays an important role in collagen synthesis and proper muscle

development. As such, LARP6 is an attractive therapeutic target for fibroproliferative disorders[10, 15]. Since LARP6 interacts with collagen mRNAs, there have been recent efforts, namely by Stefanovic *et al*. (2019), to find LARP6 inhibitors as antifibrotic drugs.[30]

Evolutionary Divergence of the La Modules

To date, no molecular structures have been solved for any of the full length LARPs. However, previous studies have produced high-resolution structures of the La Module or its subdomains from three human LARPs[9, 10, 18, 19, 20]. The isolated La Motifs of human LARP3 and human LARP6 have been solved by solution NMR (Figure 4)[10, 9, 18, 19]. Similarly, the isolated RRMs of human LARP3 and human LARP6 were also solved by solution NMR (Figure 5)[9, 10, 18, 19]. Finally, X-ray crystallography was used to solve the RNA-bound structure of the intact La Module from human LARP7[20].

Generally, the La Motif is a canonical, winged-helix domain, as seen in both the La Motifs of human LARP6 and human LARP3, (Figure 4)[10]. Though the La Motifs of these subfamily members are highly similar, there are some notable minor differences[10]. Helix α1 is shorter in LARP6 than in LARP3; additionally, the loop between α1 and β1 ("Loop 1") is longer and less structured in LARP6 than in LARP3[10]. The LARP6 wing 2 is longer, perhaps compensating for a shorter La Motif-RRM interdomain linker. The interdomain linkers are seen shorter in LARP6 subfamily, in contrast to the longer interdomain linker of the LARP3 subfamily[9, 10]. For the LARP superfamily in general, these types of differences are expected to contribute to the different relative orientations of the La Module component domains as well as ligand specificity. These differences are

thought to contribute to the different orientations the La Modules of the LARPs can adopt[9].

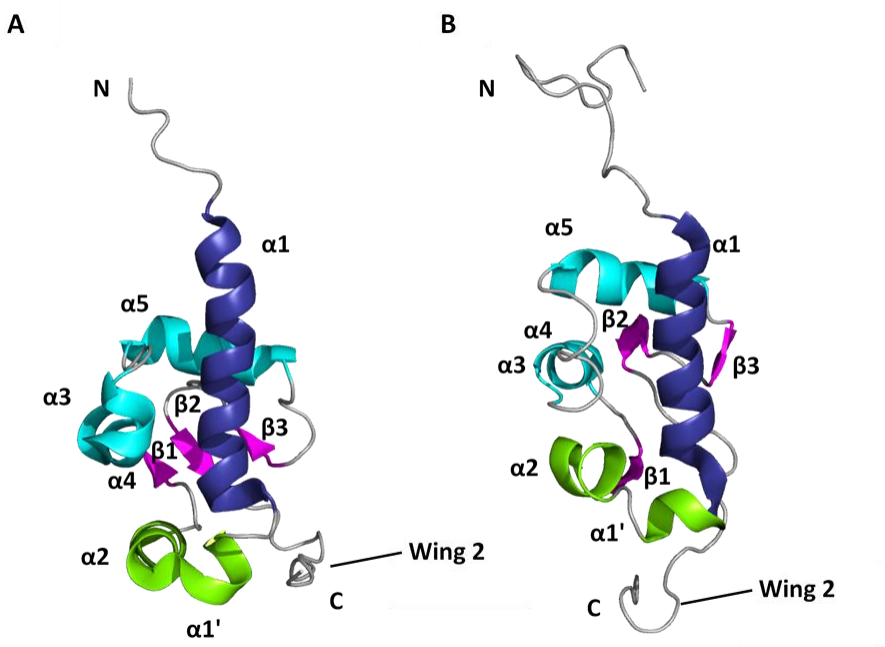

Figure 4. Structures of Human LARP3 vs. Human LARP6 La Motifs. A) *Hs*LARP3 La Motif NMR (PDB: 1S7A) compared to B) *Hs*LARP6 La Motif (PDB: 2MTF). The overall structures display high structural similarity. However, *Hs*LARP6 La motif has a shorter α1 helix (blue), and a less structured, longer loop 1 (green). Structures generated with PyMOL[35].

In contrast to the generally conserved sequence and structure of the La Motif, the RMM is much more variable across the LARPs (Figure 5)[7, 10]. The LARP6 RRM was discovered to adopt a novel variation of the RRM fold. It retains the canonical four-stranded, β-sheet core, and so it is more similar to LARP3 than LARP7. This core fold is augmented by two LARP6 unique helices called α0' and α1' [9, 10]. Interestingly, α1' obscures the canonical RRM RNA binding site in a manner reminiscent of LARP3, but

with critical differences. This helix adopts a similar three-dimensional position to that of helix α3 of LARP3 but is located in a different position within the primary sequence [9, 10]. More specifically, helix α1' is found within the longer Loop 3 of LARP6, when compared to LARP3-Loop 3 (10). Helix α0' is found within the longer loop 1 of LARP6, when compared to the shorted loop 1 of LARP3[10]. Additionally, there are elements found in LARP3 and LARP7 that are not present in LARP6, and the LARP6 RRM three-dimensional structure is more similar to the La Module RRM of LARP3 than to that of LARP7[9, 10]. Together these data suggests that LARPs use different surfaces of their RRMs to bind their respective RNA ligands.

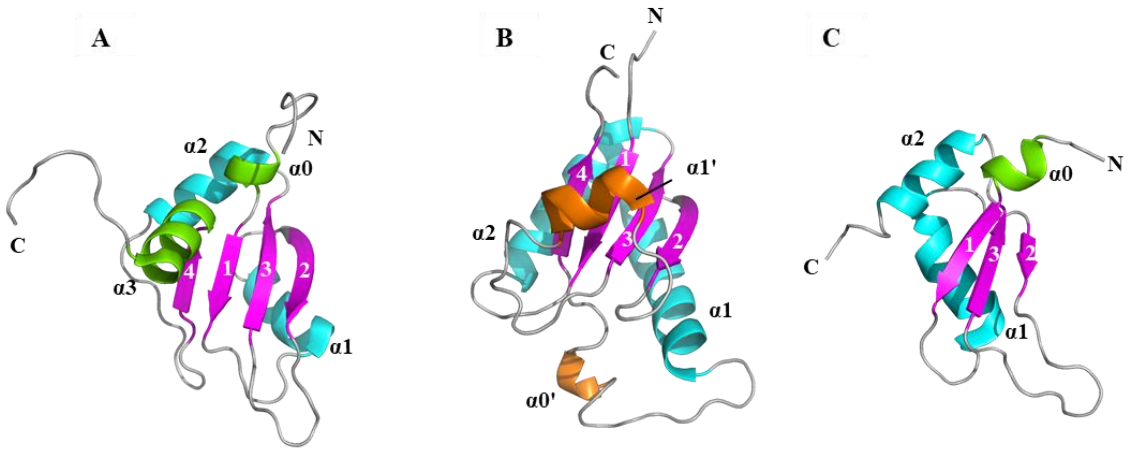

Figure 5. NMR Structures of Human LARP3 vs. Human LARP6 vs Human LARP7 RRMs. A) *Hs*LARP3 RRM (PDB: 1S79) compared to B) *Hs*LARP6 RRM (PDB: 2MTG) also compared to C) *Hs*LARP7 RRM (PDB: 4WKR). The overall structures display high variability. *Hs*LARP3 RRM and *Hs*LARP7 RRM have additional elements (green) not seen in *Hs*LARP6. However, *Hs*LARP6 RRM has two unique helices (orange). Structures generated with PyMOL[35].

To test if the similar structures of the LARP6 and LARP3 La Modules confer similar function, Martino *et al.*, study carried out extensive analysis based on the alanine

scanning mutagenesis of LARP3[10]. First, sequence variation in the La Motif was evaluated by focusing on six conserved residues in the hydrophobic pocket of the La Motif that are responsible for RNA interaction in LARP3[10]. When mutated to alanine, five out of these six residues prevented LARP6 from binding to its human collagen stem-loop ligand (*COL1A1*)[10]. The mutation of the sixth residue, the aspartic acid at position 112, weakened RNA binding affinity, but did not impede it[10]. Therefore, the hydrophobic pocket in the La Motif was suggested to play a key role in RNA binding in LARP6, much as it does in LARP3[10, 18, 19].

Next, amino acid substitutions were made at residues in the RRM that are analogous to sites involved in the RNA binding activity of LARP3. They targeted multiple sets of residues; within the canonical RRM motifs RNP-1 and RNP-2 on the β-sheet surface; Loops 1 and 3; and the unique conserved α1' motif. None of these mutations disrupted RNA binding[10]. In fact, alanine mutations of lysine 187 (found in RNP-2) or glutamic acid 262 (found in RNP-1) actually increased binding for the RNA ligand[10].

Finally, Martino *et al.* investigated the effects of swapping larger sequence motifs between LARP subfamilies. They constructed chimeric proteins comprised of LARP6 La Motif and the LARP3 RRM and tested for RNA binding ability[10]. The RRM-chimera completely abolished RNA binding activity. A similar phenomenon was seen when LARP7-RRM was swapped out with the LARP3-RRM[9]. Collectively, these results demonstrate that RRMs have evolved within the La Modules in a LARP-specific manner, and protein coding sequences for this domain cannot be swapped between LARP

subfamily members[9, 10, 25]. A second inter-family chimera was also made, in which the short interdomain linker of LARP6 was replaced with the longer linker of LARP3. This mutation decreased RNA binding affinity by three orders of magnitude. Overall, these data suggest that a specific combination of La Motif/interdomain linker/RRM is required for RNA binding activity, and that component sequences cannot be exchanged between LARP subfamilies[10]. This indicates that each subfamily La Module employs different mechanisms for RNA binding.

It is still unknown whether binding mechanisms are conserved within a specific LARP subfamily of proteins. If so, then protein domains of a LARP subfamily member from two different species should be able to be exchanged with minimal effect on structure and RNA binding activity[10].

Divergence within LARP6 proteins

Past studies have used animal models to study the interaction between LARP6 and its mRNA target due to its potential as a therapeutic agent for fibrotic diseases[8, 10, 15, 22]. Zebrafish (*Danio rerio*) have homologous gene sequences of collagen type I and LARP6. As described above, previous studies demonstrated that the expression of LARP6 in zebrafish was important for embryonic development[22]. As such, zebrafish is an ideal genetic model to study the function of LARP6 proteins and offer the opportunity for future genetic experiments in the whole organism. However, zebrafish carry two paralogues of the LARP6 proteins (*Dr*LARP6a and *Dr*LARP6b)[10]. The zebrafish LARP6a is more homologous in sequence to the single LARP6 in other vertebrates. To fill in the gap between zebrafish that carry two genes for the LARP6 protein and humans,

which only carry one gene for the protein, the widely accepted model, platyfish (*Xiphophorous maculatus*), which also only expresses one LARP6 protein (*Xm*LARP6), was used[24].

A multiple sequence alignment of LARP6 proteins from different species identified clade-specific features, (Figure 6). The known secondary structure of the LARP6 La Modules are represented as rectangles (α helices) and arrows (β-sheets) as well as domain boundaries for the La Motif, linker, and RRM. These alignments demonstrate regions of conservation, mainly in the La Motif, RRM and LSA motif, although localized divergence is present within these regions. However, conservation is much higher between the La Motifs than the RRMs (Appendix A, Table S1). Sequence divergence is primarily located in the loops between the beta strands of the RRM (Loops 1 and 3) and interdomain linkers.[10]

Figure 6. Multiple Sequence Alignments of LARP6 Sequences from Different Species. The known secondary structures of HsLARP6 La Module are depicted above the amino acid sequences in which boxes represent α helices and arrows represent β sheets. Residues in dark gray represents 100% conservation, medium gray is ≥ 40% identity, and the light gray shade indicates a 30% identity. Image generated using Clustal Omega and modified with Jalview (adapted from Martino *et al*. (2015); Reproduced with permission under Creative Commons CC BY license).

To study the binding mechanisms of LARP6 and the effects of evolutionary sequence divergence, previous work in our laboratory cloned, expressed, and purified recombinant human LARP6 (*Hs*LARP6), platyfish LARP6 (*Xm*LARP6), and zebrafish LARP6a proteins (*Dr*LARP6a), Figure 7[26]. The recombinant proteins were then biochemically characterized by measuring binding affinity to the 5' SL motif from the human LARP6 collagen type I mRNA[26]. Results showed that both fish LARP6 proteins bound to the stem-loop structure with higher apparent affinity than the human protein[26]. Structural characterization of these LARP6 proteins through limited proteolysis showed that the fish proteins contain a 40 kDa, protease-resistant domain[26]. While a similarly-sized domain is present in the human LARP6 protein, it is much less resistant to protease digestion than the non-mammalian vertebrate LARP6 domains[26]. Therefore, the 40 kDa domain in the fish proteins remains stably folded against trypsin digestion, as the band persists throughout the digestion reaction, in contrast to the similarly-sized human domain[26]. This stable domain was identified via mass spectrometry to be composed of the N-terminus and the La Module[26], suggesting that the C-terminus in all three species is disordered, but the N-terminus may interact with the La Module and potentially contribute to the RNA binding activity of the protein.

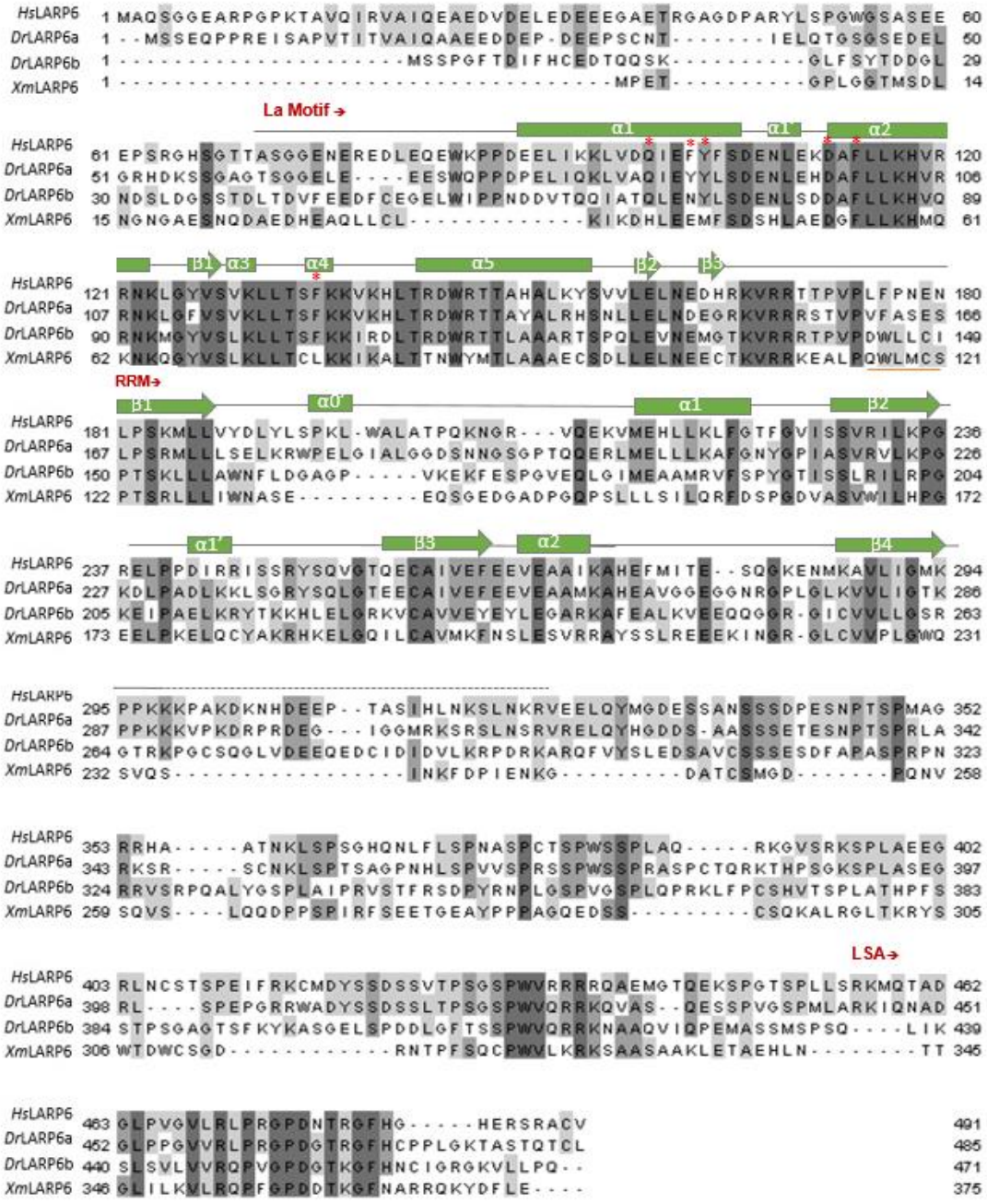

Figure 7. Sequence Alignments of *Hs*LARP6, *Xm*LARP6, *Dr*LARP6A, and *Dr*LARP6B LARP6 Proteins. The known secondary structures of *Hs*LARP6 La Module are depicted above the amino acid sequences. Conserved residues are shaded. The six highly conserved residues of the La Motif are indicated by red asterisks. The interdomain linkers are indicated by underlined residues. Multiple sequence alignment generated through Clustal Omega. Image modified with Jalview. Dark gray represents 100% conservation, medium gray is ≥ 40% identity, and the light gray shade indicates a 30% identity.

To identify regions of the RRM that are critical for LARP6 structure and function, this thesis project takes advantage of the sequence divergence between human and fish LARP6 proteins to create interspecies chimeras of LARP6 in which the RRMs are exchanged. The two specific aims of this project will address the following questions: (1) Can the RRM domains be exchanged between LARP6 proteins from mammalian and non-mammalian vertebrate species and produce stable chimeric proteins? (2) Do interspecies chimeras demonstrate changes in RNA binding activity?

To address the first question, a set of chimeric proteins will be constructed by exchanging the linker and RRM sequences between eutherian (human) LARP6 and non-mammalian vertebrate (fish) LARP6. First, we will attempt to create full-length chimeras in which the linker and RRM protein coding sequences of *Xm*LARP6, and *Dr*LARP6a are inserted in *Hs*LARP6 to create *Hs*LARP6*Xm*(RRM), *Hs*LARP6*Dr*A(RRM), and *Hs*LARP6*Dr*B(RRM). The complementary chimeras will also be generated using the full-length *Xm*LARP6 and *Dr*LARP6A proteins, to create *Xm*LARP6*Hs*(RRM) and *Dr*LARP6A*Hs*(RRM). These full-length chimeras will be used to generate two other sets of shorter subdomain constructs. One set will be the chimeras as described above in which the C-terminus is deleted ("ΔCTD"), and the second set of constructs that will express just the isolated chimeric La-Modules. All proteins will be recombinantly expressed in *E. coli*. If expression is successful, the chimeras will be purified through affinity chromatography via an N-terminal polyhistidine tag, followed by size exclusion chromatography. The effect of the domain swap on protein stability will be measured using limited proteolysis.

To address the second aim, all successfully purified chimeras will be subjected to biochemical characterization through electrophoretic mobility shift assays (EMSAs) against the stem-loop RNA structure from human collagen type I mRNA and compared to their wildtype counterparts to determine whether the RRM exchanges cause any difference in RNA binding activity. Together, these studies will identify how sequence divergence within the LARP6 RRM affects the structure and RNA binding activity of LARP6 and enable future studies that identify shared and unique elements of binding mechanisms within LARP subfamilies.

II. MATERIALS AND METHODS

Calibration of Size Exclusion Chromatography Columns

The Sephadex S75 size exclusion column was calibrated by eluting a set of standard proteins over the column. The elution volume of each protein was recorded and correlated with a known molecular weight (Table 1). Blue dextran, the largest protein, eluted first and identified the void volume (V_0). The calibration showed that the column efficiently separated proteins of different sizes, marked by elution volumes, which makes it useful for protein purification. The sets of standards used for the S75 columns are listed in Table 1.

Table 1. Size Exclusion Chromatography S75 Standards

Standard	MW (kDa)	Elution Volume (mL)
Blue Dextran	2,000	45.03
Conalbumin	75	53.29
Ovalbumin	44	58.75
Carbonic Anhydrase	29	67.16
RNAse A	13.7	79.72
Aprotinin	6.5	92.87

For each protein, K_{av} (the "available value") of the distribution coefficient K_D was calculated using the equation:

$$K_{av} = (V_e - V_0)/(V_c - V_0)$$

where V_e is the protein of interest's elution volume, V_0 is the column void volume (determined by Blue Dextran), and V_c is the column's total volume (120 mL).

The K_{av} values were plotted as a function of the elution volumes from Table 1, and linear regression analysis was performed to generate an equation to model the apparent molecular weight.

$$y = - 0.519x + 2.6075$$

where x is the $\log_{10}$ of the molecular weight of the standard protein (Da) and y is K_{av}.

Cloning of Expression Constructs

The human LARP6 (pET28a-*Hs*LARP6_FL), zebrafish LARP6 (pET28-SUMO-*DrA*LARP6_FL) and the platyfish LARP6 (pET28-SUMO-*Xm*LARP6_FL) expression plasmids were cloned by Castro (2017)[26]**.** These proteins served as templates to generate wildtype controls for comparison against the chimeras. The expression plasmids for the SUMO fusion system (pET28-SUMO) was a gift of Dr. Christopher Lima (Rockefeller University).

Cloning and amplification of the full-length chimeric constructs *Hs*LARP6*Xm*(RRM)_FL, *Hs*LARP6*DrA*(RRM)_FL, *Xm*LARP6*Hs*(RRM)_FL and *DrA*LARP6*Hs*(RRM)_FL were cloned by overlap extension PCR as described by Nelson and Fitch (2011)[26,27]. Each full length chimera was used as the template to produce the "ΔCTD" constructs: *Hs*LARP6*Xm*(RRM)_ΔCTD, *Hs*LARP6*DrA*(RRM)_ ΔCTD, *Xm*LARP6*Hs*(RRM)_ ΔCTD and *DrA*LARP6*Hs*(RRM)_ ΔCTD. The isolated La Modules (La Module) of each chimeric protein were also cloned using PCR and

restriction-enzyme based cloning (*Hs*LARP6*Xm*(RRM)_ LaMod, *Hs*LARP6*DrA*(RRM)_ La Module, *Xm*LARP6*Hs*(RRM)_ La Module, and *DrA*LARP6*Hs*(RRM)_ La Module). The products were amplified using primers that contained either *Nde*I/*Xho*I, *Nco*I/*Xho*I or *Bam*HI/*Xho*I restriction sites (Appendix A, Table S2). The amplified products were then inserted into suitably-digested pET28-SUMO or pET28a plasmid. The set of primers and expression plasmids used are summarized in Table S2. Each expression construct includes an N-terminal polyhistidine tag for affinity purification. All cloned constructs were screened with colony PCR and then confirmed by commercial Sanger sequencing.

Cell Transformations

Escherichia coli DH5α competent cells were used for cloning. The cells and DNA were incubated for 30 min on ice, heat shocked for 90 s at 37 ºC, and then incubated for 2 min on ice. 700 µL of autoclave-sterile Luria Broth (LB) was added and cultures were incubated at 37 ºC for 1 h with shaking. Recovered cells were plated onto a LB-agar plate containing the Kanamycin antibiotic (Appendix A, Table S3 and Table S4) and incubated at 37 ºC overnight. A single colony from each LB-agar plate was used to inoculate 3 mL LB broth and incubated at 37 ºC for ~16-18 hours. Plasmid DNA was isolated using QIAprep Spin Miniprep Kit (Qiagen). The manufacturer's protocol was followed, using the optional wash step, using the manufacturer's wash buffer, and eluted DNA in 40 µL of elution buffer.

For protein expression, sequenced-verified constructs were transformed into *E. coli* Rosetta™ (DE3) pLysS competent cells. The cells and DNA were incubated on ice for 30 min, heat shocked for 45 s at 42 ºC, followed by incubation on ice for 2 min. Next,

700 µL of autoclave-sterile LB was added and samples were incubated at 37 ºC for 1 h with shaking. The samples were plated onto LB agar plates containing Kanamycin and Chloramphenicol (Appendix A, Table S3 and Table S4) and incubated overnight at 37 ºC.

Colony PCR

For this set of PCR reactions, the colonies from cell transformations were used as the source for the template DNA. A 5 mL sample of LB with the appropriate antibiotics was prepared first. Then using a sterile, pipette tip, a single colony was picked and swabbed vigorously at the bottom of an autoclave-sterile tube. The tip was then ejected into the 5 mL LB sample prepared prior, and the culture was grown overnight at 37 ºC. As such, each PCR reaction corresponded to a culture tube. The PCR reactions were prepared as to detect the inserts of interest and analyzed by agarose gel electrophoresis. If the expected band showed on the gel, indicating the presence of the construct of interest, then the corresponding culture was miniprepped and sent for commercial Sanger sequencing to confirm the insert.

Protein Expression Trials

A 5 mL culture was prepared by inoculating a single colony from the Rosetta™ transformation plates into 5 mL LB containing the appropriate antibiotics and grown at 37 ºC overnight with shaking. 1 mL aliquots of these cultures were transferred into 100 mL of autoclaved-sterile LB with appropriate antibiotics. 100 mL cultures were grown until an OD_{600} between 0.5 – 0.7 was reached at 37 ºC with shaking. The cultures where then placed on ice for 5 – 10 minutes with occasional shaking. A 1 mL aliquot was then removed from the cultures and pelleted by centrifugation at 16,00 ×g for 1 min to serve

as a pre-induced sample subsequent evaluation of protein expression. Protein expression was induced with 1 mM isopropyl b-D-1-thiogalactopyranoside (IPTG) and placed at 18 ºC with shaking overnight. To track expression, 1 mL samples were taken at 2 h, 4 h, 6 h, 8 h, and overnight (~18 hrs). The cells in these aliquots were pelleted by centrifugation at 16,000 ×g and stored at -20 ºC. Each cell pellet was resuspended in 250 µL 1× SDS sample buffer (5× buffer: 0.5 M Tris [pH 7.5 at 4 ºC], 4 mM β-ME, 0.4 M SDS) and incubated for 5 min before immediate loading onto two 10% SDS-PAGE gels (37.5:1 acrylamide: bis-acrylamide; ProtoGel, National Diagnostics). The gels were electrophoresed in 1× Tris-Gly Running buffer (50 mM Tris and 0.5 M glycine, 0.4 M SDS) at 200 V for about 1 h. The PageRuler Pre-stained Ladder (Thermo Scientific) was loaded to determine apparent molecular weights. These gels were analyzed by either Coomassie staining or an anti-His Western blot.

Coomassie Blue Staining

Total protein was detected using a Coomassie Brilliant Blue staining solution (0.05% (w/v) Coomassie Brilliant blue, 40% (v/v) methanol, 10% (v/v) glacial acetic acid, and 50% (v/v) with Milli-Q polished water) for 30 min at room temperature with shaking. Gels were then destained in Destain solution (40% (v/v) methanol, 10% (v/v) glacial acetic acid, and 50% (v/v) Milli-Q polished water) and imaged using BioRad ChemiDoc XRS+ molecular imager using the Coomassie stain setting.

Silver Staining

Silver stains were performed for sensitive detection of protein content. In a glass container, gels were fixed using a 50% ethanol solution for 15-30 min at room

temperature with shaking. The gels were then incubated in staining solution (7.56% NaOH, 1.5% (v/v) NH_4OH, 4.7M $AgNO_3$) for 15-30 min with shaking. Gels were rinsed 3 times with ultrapure water. After rinsing, the gels were incubated in developing solution (2.5% citric acid, 37% formaldehyde) and gently mixed until the desired intensity was reached. Immediately 100- 200 mL of kill solution (45% v/v methanol and 2% v/v acetic acid) was added and incubated for 1 h with shaking. Gels were imaged with the BioRad ChemiDoc XRS + molecular imager using the silver stain setting.

Anti-His Western Blot

To confirm expression of the protein of interest, anti-His Western Blots were performed. The Bio-Rad Trans-Blot Turbo Transfer System was used to transfer the proteins from the gel to a 0.2 μm nitrocellulose membrane using the Mixed Molecular Weights program (1.3 A, 25 V, 7 min). A 1× Tris-buffered saline solution (1× TBS) solution was prepared. After transfer, the membrane was incubated for 1 h in Blocking solution (10% bovine serum albumin [BSA] in 1× TBS containing 0.05% Tween-20 [TBS-T]) at room temperature with shaking. The membrane was then incubated for 1 h with anti-His probe (Pierce, lot number 15165) in TBS-T at room temperature with shaking. The membrane was washed twice with 1× TBS-T for 10 min each. Lastly, the membrane was washed twice for 10 min with 1× TBS. The membrane was then incubated with 20 mL of chemiluminescence solution (100 mM Tris-HCl [pH 8.8 at 4 ºC], 1.25 mM luminol in DMSO, 2 mM 4-Iodophenylboronic acid ([IPBA] in DMSO) and 12 μL of 30% H_2O_2[28]. The membrane was then imaged using the BioRad ChemiDoc XRS + molecular imager using the chemiluminescence setting.

Protein Expression and purification

1 L of autoclaved-sterile LB with appropriate antibiotics was inoculated with 16 mL of overnight culture of transformed Rosetta™ cells grown at 37 ºC to a target OD_{600} of 0.5-0.7. Cells were induced with 0.1 M IPTG and incubated overnight with shaking at 18 ºC. Cells were then collected by centrifugation at 5,000 ×g, 4 ºC for 10 min. Cell pellets were then stored at -20 ºC until purification.

Specific purification buffers for each protein construct can be found in Appendix A, Table S5. Briefly, cells were thawed on ice and resuspended using Lysis/Wash Buffer #1 (Table S5) with a dissolved protease inhibitor tablet (Pierce, PIA32965). The cells were then lysed by sonication at 30% amplitude in an ice water bath at one pulse per second for 20 pulses followed by a 30 s rest. This was done for a total of six cycles. The supernatant was then collected by centrifugation at 18,000 ×g, 4 ºC for 15 min.

Nickel affinity chromatography

Ni^{2+}- agarose beads (4 mL of a 50% slurry per 1 L expression culture) were washed with MilliQ-polished water and equilibrated with ~10 mL of Lysis/Wash #1 buffer + protease inhibitor. The equilibrated beads were then added to the cleared lysate and incubated for 1 h at 4 ºC with shaking. The beads/lysate mix was then transferred to a glass column and allowed to settle. The flow-through was collected and the beads washed with 20 mL of Lysis/Wash Buffer #1 (Table S5) without protease inhibitor. The column was washed a second time with 40 mL Wash Buffer #2 (Table S5) to dislodge weakly-bound proteins. Tightly bound proteins were eluted with 24 mL Elution Buffer. All fractions collected were analyzed via SDS-PAGE and Coomassie Blue staining.

Size exclusion chromatography

A size exclusion column (Sephadex 75) was equilibrated with Storage Buffer (Table S5) which had been filtered through a 0.2 μm nitrocellulose membrane and degassed. Elution fractions from the Ni^{2+} affinity column were pooled based on SDS-PAGE band intensity and concentrated to 2.5 – 3 mL using a pre-rinsed Vivaspin™ centrifugal concentrator (Sartorius) with a molecular weight cut off of 10,000 by centrifugation at 4,000 ×g, 4 ºC for 10 -20 min cycles. The concentrated protein was then filtered with a 0.2 μm syringe filter before loading into the AKTA Pure Fast Protein Liquid Chromatography (FPLC) system. The column was set to elute 180 mL at a flow rate of 1 mL/min. Fractions were collected based on the chromatogram generated from A280 absorbance measurements. These fractions were then analyzed by SDS-PAGE and Coomassie staining to identify band intensity and purity of each fraction. The pooled fractions were aliquoted into 50 μL aliquots that were snap-frozen in liquid N_2 and stored at -70 ºC.

Limited Proteolysis

To probe the structural topology of the wildtype La Module and the chimeric La Module proteins, limited proteolysis analysis with trypsin was performed. The stored proteins were thawed on ice and a 20 μL aliquot was removed and combined with 5× SDS sample buffer, incubated at 37 ºC, boiled at 95 ºC for 5 min and stored at -20 ºC as the "0 s" timepoint sample. Trypsin (0.1 mg/mL stock concentration) was added to the remaining protein at molar ratio of 500:1, protein: trypsin. At the indicated timepoints, a 20 μL aliquot was removed from the reaction (0 - 180 min) and treated with 5x SDS

sample buffer, incubated at 37 ºC, boiled at 95 ºC and stored at -20 ºC. Samples were analyzed by SDS-PAGE and silver staining.

Biotinylation of *COL1A1* ligand

Extra care was taken when working with RNA due to RNases found in the day-to-day environment. These steps include cleaning surfaces and tools used for this experiment with RNase Zap (Invitrogen) to remove RNases, wearing gloves at all times and using only filtered, sterile tips.

Biotinylation reactions were prepared using commercially synthesized RNA (IDT, Inc) and the Pierce™ RNA 3' End Biotinylation Kit, following the manufacturer's instructions. Each reaction started with an RNA+DMSO mix, to prevent non-desirable secondary structures, that were incubated at 85 ºC for 2 min. Immediately after, the RNA+DMSO mixes were placed in an ice water bath for 10 min. A 24:1 mixture of chloroform:isoamyl alcohol was prepared. The micropellet was resuspended in 20 µL of 0.5x Tris-EDTA (TE) buffer. The RNA was then stored as 2 µL aliquots at -20 ºC.

RNA Labeling Efficiency

To serve as our control, the Biotinylated Control IRE RNA from the Pierce 3'-End Biotinylation kit, was used and diluted to make serial standards per manufacturer's instructions with the following adjustments. Each experimental RNA was diluted to 1 nM as the starting concentration for blotting. Prior to blotting, a Hybond N+ membrane was incubated with 1x TBE for at least 10 min. The membrane was slightly blotted. The semi-wet membrane was then placed on a piece of plastic wrap and 2 µL of each dilution was spotted to the membrane.

Electrophoretic Mobility Shift Assays (EMSAs)

Electrophoretic mobility shift assays (EMSAs) were performed to test the RNA binding of the La Module proteins against the stem-loop from the *COL1A1* mRNA. A 10x Binding Buffer stock (100 mM Tris-HCl (pH 7.4 at 4 ºC), 200 mM KCl, 10 mM $MgCl_2$, 1000 mM NaCl) was prepared and stored at -20 ºC. The final binding buffer was prepared by diluting the 10× Binding Buffer to1× Binding Buffer in nuclease-free water (IDT) and adding 15% glycerol was prepared from the stock. Excess Binding Buffer was stored at -20 ºC for future use.

Protein aliquots were thawed on ice. Serial protein dilutions were prepared in 0.5 mL autoclave-sterile, low-binding microcentrifuge tubes (USA Scientific). The protein concentrations were designed to be 2× the intended concentration for the final binding reaction.

Biotinylated RNA was thawed on ice and heated at 80 ºC for 2 min and then placed in an ice-water bath for 10 min. In low-binding tubes, the biotinylated RNA stock was serially diluted to a concentration that is twice the concentration of the final binding reaction using binding buffer with RNase inhibitor (Invitrogen, 1900300).

The binding reactions were prepared by a 1:1 mixture of the 2x protein dilutions and the 2x RNA dilution. The reactions were incubated on ice for 1 h to reach equilibrium. Reactions were loaded onto 6.5% polyacrylamide (29:1, acrylamide:bisacrylamide; ProtoGel) Tris-borate-EDTA (TBE) native gels with 5% glycerol[29]. The gels were electrophoresed in cold 1x TBE without glycerol at 200 V for 15 min with -70 ºC ice blocks to absorb excess heat. A 50% sucrose/bromophenol blue solution was used to

track gel progress. The separated RNAs were transferred to a Hybond N+ membrane (GE Biosciences) using the Transblot Turbo (25 V, 1.0 A, 30 min) in 1x TBE, then crosslinked with a UV oven for 45 seconds at 120 mJ/cm^2. The membrane was allowed to dry overnight in the dark.

Detection of Biotinylated RNA

Detection of the biotinylated RNA was performed using the LightShift RNA EMSA Chemiluminescence Nucleic Acid Detection Module (ThermoScientific). Two membranes were rehydrated with 20 mL of Nucleic Acid Detection Blocking Buffer for 15 min with gentle shaking. The standard manufacturer's protocol was followed. After the last incubation, the membrane was blotted dry through one corner of the membrane with a paper towel. The membrane was then placed RNA face side up on clean plastic wrap and covered with 5 mL of 1:1 mixture of luminol enhancer solution and peroxidase solution for 5 minutes. The membrane was blotted again ad placed RNA face down on a fresh piece of plastic wrap. The membranes were then imaged using the chemiluminescence protocol of the BioRad ChemiDoc XRS + molecular imager.

EMSA Quantification and determination of $K_{D,app}$

Quantification was done using the ImageLab™ software installed in the BioRad ChemiDoc XRS+ imaging system. Volume tool boxes outlined the boundaries of unbound RNA and bound RNA. Quantification of the pixels found in the boundaries for both the unbound and bound states were given as numerical values and exported as an Excel file.

The RNA control, U1, which was not incubated with protein, was used to denote the maximum volume signal that unbound RNA should have in an EMSA**.** This volume box was then copied and placed around the remaining unbound RNA bands, hypothesized to be U1 – U14. The bound volume box boundaries were measured using the highest signal in space located in the gel. In this analysis, U28 shows the boundaries for bound RNA; this volume box was copied and aligned with the boxes from unbound RNA. U14 and U15 include membrane background signal and was subtracted from each measurement.

The counts in the unbound and bound set of boxes in each lane were added to calculate a total signal value. The fractional saturation was determined by dividing the bound signal for each column by the total signal of that same column. Completely unbound protein is represented by a value of zero and completely saturated protein is represented by a value of one.

The fractional saturation and protein concentrations were plotted and fit to 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. If the concentration of total protein ($[P]_T$) and total ligand ($[L]_T$) are known, and the protein bound to the ligand can be measured, the apparent dissociation constant, $K_{D,app}$, can be determined. The factors S is a saturation offset and factor O is a background offset[36]. The standard

error of the mean (SEM) was calculated by first calculating the standard deviation of all apparent K_D values and dividing that value by the square root of the number of trials. All reported $K_{D,app}$ values are derived from at least three independent replicates.

III. RESULTS

Cloning of the *Human Backbone* Chimeric Expression Constructs

These recombinant proteins were designed as follows: the linker-RRM primary sequence of the human LARP6 protein was removed from the full-length protein and replaced with either the RRM-linker sequence of *Xiphophorus maculatus (platyfish)* or *Danio rerio (zebrafish)*. Three template plasmids were used and a series of primers (Table S2) for each set of PCR reactions to clone the individual fragments with complementary sticky ends to create the chimeric sequences. The chimeras were then cloned into either pET28a or pET28-SUMO plasmids. Each chimera set is comprised of three different constructs, a full-length protein (FL), with a C-terminal domain deletion (ΔCTD) and the isolated La Module (La Module). Figure 8 shows a representative schematic for each chimera set. Chimeric sequences can be found in the Appendix Section.

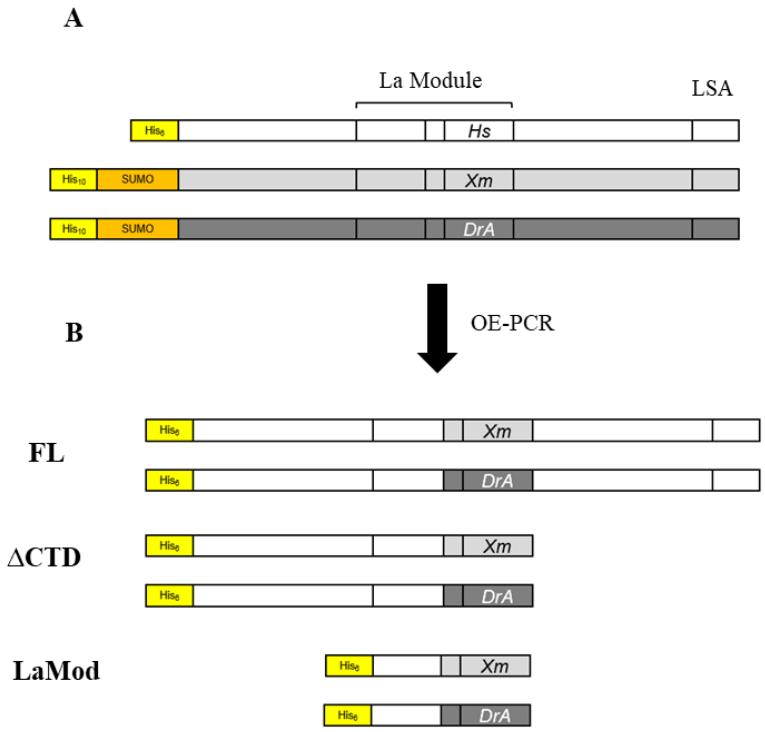

Figure 8. Human Backbone Chimera Expression Constructs Schematic. The LARP6 topology is represented with the presence of the La Module (La Motif – Linker – RRM1) and the LSA motif. The N-terminal His tag (His_6 or His_{10} for those constructs with the SUMO tag) is shown in yellow. The SUMO tag is shown in orange. A) At the top, the wildtype human LARP6 (*Hs*) in white, platyfish LARP6 (*Xm*) in light gray and zebrafish LARP6 (*DrA*) in dark grey. B) The full-length human backbone chimeras retain the N-terminus, La Motif and C-terminal region of *Hs*LARP6 with linker-RRM sequences from either *Xm*LARP6 or *DrA*LARP6; The constructs with C-terminal deletions (ΔCTD); isolated chimeric La Module constructs (La Mod).

I. *Hs*LARP6*DrA*(RRM) Chimera Set

This set of chimeras was designed as follows: the linker-RRM primary sequence from the human LARP6 protein was replaced with the zebrafish linker-RRM sequence. Figure 9A shows the amplified fragments (1A, 2A, and 3A) to form *Hs*LARP6*DrA*(RRM)_FL, the full length construct (corresponding to amino acids = 1-516) on a 0.8% agarose gel. The individual fragments were designed as follows:

Fragment 1 is composed of *Hs*LARP6 (coding for amino acids 1-170), Fragment 2 is composed of *DrA*LARP6 (corresponding amino acids 158-290), and Fragment 3 is composed of *Hs*LARP6 (encoding amino acids 301-491). The encoding DNA fragments were expected at 585 bp, 421 bp and 590 bp, respectively. These fragments were pooled for overlap extension PCR to form the full-length chimeric sequence. The OE-PCR directly purified using the QIAquick® PCR Purification Kit (Qiagen). Finally, a set of primers that included the appropriate restriction sites were used to amplify the purified, chimeric DNA sequence through PCR **(**Figure 9B). A band was expected at 1551 bp to indicate the presence of the full-length chimeric insert ("123A"), and a band was seen between 1650 bp and 1000 bp, indicating the presence of *Hs*LARP6*DrA*(RRM)_FL. Figure 9C shows the gel after excision of the double-digested insert and target plasmid. Bands were expected around 5369 bp and 1551 bp, respectively. Figure 9D shows the PCR product to confirm the presence of the fish RRM. The presence of the insert in the correct reading frame was confirmed by commercial Sanger sequencing.

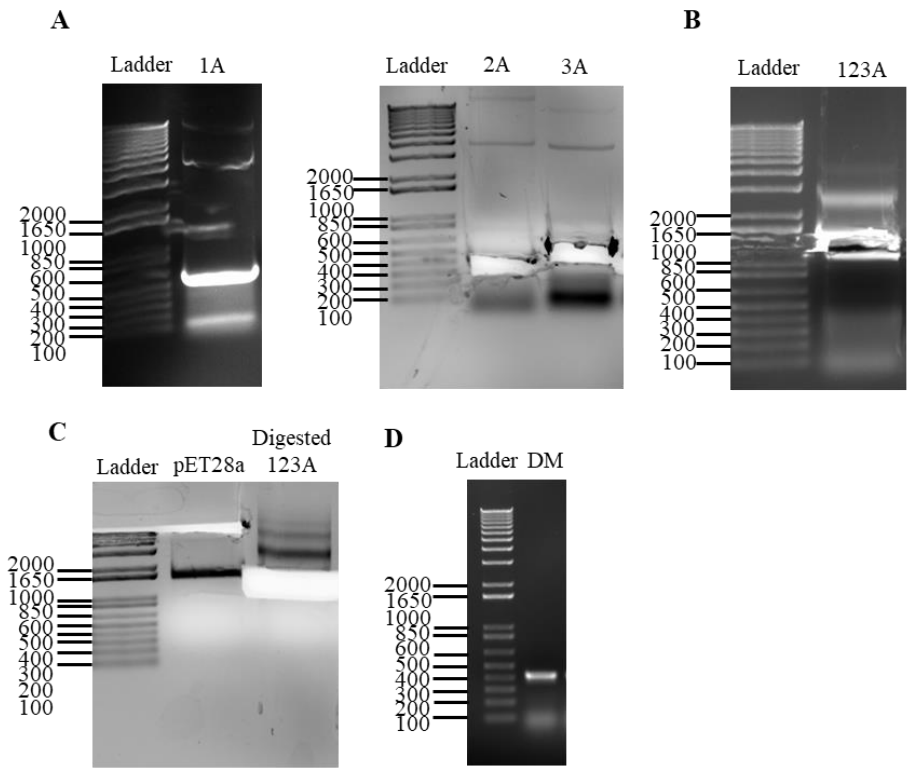

Figure 9. PCR Products for Cloning of pET28a-*Hs*LARP6*DrA*(RRM)_FL Expression Construct. A) *Lane 1* shows the 1 kb+ ladder (for every gel in this figure); *Lane 2* shows amplified fragment 1 (***H*****s**LARP6 1-170) using primers from Table S2. A band was expected at 585 bp. *Lane 2* and *lane 3* show amplified 2A (*DrA*LARP6 158-290) and 3A (*Hs*LARP6 301 - 491) fragments, respectively. Bands were expected at 421 and 590 bp. B) *Lane 2* shows the amplified product after OE-PCR and the PCR purification of the chimeric construct ("123A"). A band was expected at 1551 bp, representing *Hs*LARP6*DrA*(RRM)_FL. C) *Lane 2* shows the digest product for pET28a plasmid and *Lane 3* shows the digest product for *Hs*LARP6*DrA*(RRM)_FL ("Digested 123A"). Bands were expected at around 5369 bp and 1551 bp., respectively. D) Amplification products of colony PCR. *Lane 2* shows the amplified zebrafish RRM to confirm the presence of our construct of interest ("Detected 123A"). A band was expected at 421 bp, to confirm the presence of the zebrafish RRM, therefore confirming the presence of our insert.

Using the sequence-verified, full-length *Hs*LARP6*DrA*(RRM) expression plasmid as a template, the ΔCTD construct (corresponding to amino acids = 1-338) *(left)* and the *Hs*LARP6*DrA*(RRM)_LaMod construct (amino acids = 89 - 338) (*right)* were generated by PCR (Figure 10A). Bands were expected at 999 bp and 735 bp. Successful PCR products were double-digested *Hs*LARP6*DrA*(RRM)_ΔCTD with *Nco*I/*Xho*I and

*Hs*LARP6*DrA*(RRM)_La Mod with *Nde*I/*Xho*I (Figure 10B**)** double digestion, respectively. Figure 10C shows the double digest product of the pET28a expression plasmid with either *Nco*I/*Xho*I or *Nde*I/*Xho*I. The recovered ligated constructs were confirmed by commercial Sanger sequencing.

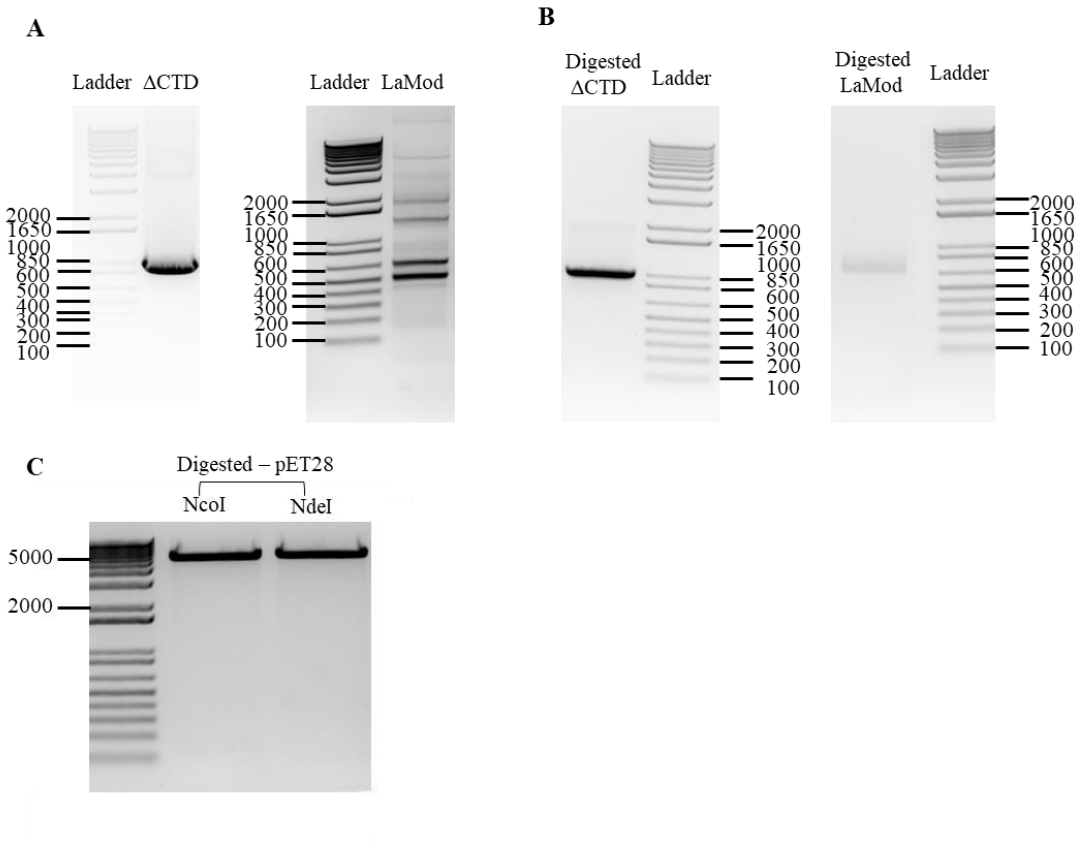

Figure 10. Amplification Products for Cloning of *Hs*LARP6*DrA*(RRM)_ΔCTD and *Hs*LARP6*DrA*(RRM)_La Modulr Expression Construct. A) *Left* shows the amplified product after removal of the C-terminal domain from the full-length version of the protein. A band was expected at 999 bp. *Right* shows the amplified product after isolating the chimeric La Module. A band was expected at 735 bp. B) *Left* shows the digest product of HsLARP6DrA(RRM)_ ΔCTD. A band was expected around 999 bp. *Right* shows the digest product of HsLARP6DrA(RRM)_La Module. A band was expected around 735 bp. C) *Lane 1* shows the digest product for pET28a after NcoI/XhoI digestion. *Lane 2* shows the NdeI/XhoI digestion product. Bands were expected around 5369 bp.

The constructs *Hs*LARP6*DrA*(RRM)_FL and *Hs*LARP6*DrA*(RRM)_ΔCTD were additionally cloned behind the pET28-SUMO expression plasmids. Figure 11A shows the product after amplification of these constructs with different restriction sites compatible with pET28-SUMO on a 0.8% agarose gel. Figure 11B shows the double digest products of the chimeric protein and pET28-SUMO, after undergoing *Bam*HI/*Xho*I digestion. Figure 11C shows the pET28-SUMO-*Hs*LARP6*DrA*(RRM)_FL and pET28-SUMO-*Hs*LARP6*DrA*(RRM)_ΔCTD colony PCR products to test for the presence of the inserts of interest before sending out for commercial Sanger sequencing to confirm the inserts.

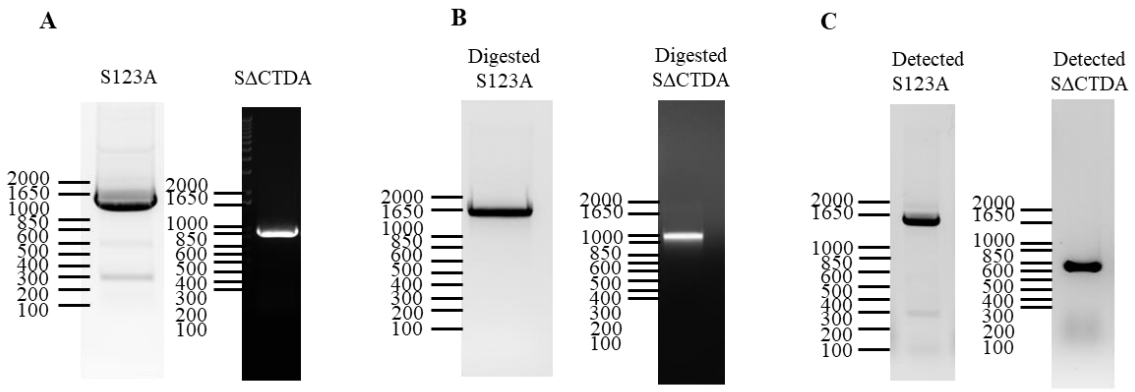

Figure 11. Amplification Products for Cloning of *Hs*LARP6*DrA*(RRM)_FL and *Hs*LARP6*DrA*(RRM)_ΔCTD Expression Construct Behind pET28-SUMO. A) *Left* shows the amplification product of the *Hs*LARP6*DrA(*RRM) full-length chimeric protein with BamHI/XhoI restriction sites ("S123A"). *Right* shows the amplification product of *Hs*LARP6*DrA*(RRM)_ΔCTD ("SΔCTDA). Bands were expected at 1551 bp and 999 bp, respectively. B) *Left*, shows *Hs*LARP6*DrA(*RRM)_FL with restriction digested "sticky" ends compatible with pET28-SUMO expression vector. *Right*, *Hs*LARP6*DrA*(RRM)_ ΔCTD digest product with SUMO-compatible ends C) Colony PCR product after amplification of full-length construct to confirm its presence.

II. *Hs*LARP6*Xm*(RRM) Chimera Set

In this set of the constructs, the linker-RRM sequence from the human LARP6 protein is removed and exchanged for the platyfish linker-RRM sequence**.** Figure 12A shows the amplified fragments (1X, 2X and 3X) that were used for OE-PCR to form *Hs*LARP6*Xm*(RRM)_FL, the full length chimeric construct (corresponding to amino acids = 1 – 510). The individual fragments are as follows: 1X is comprised of *Hs*LARP6 amino acids (1-170), 2X is composed of *Xm*LARP6 amino acids (150-292) and 3X is composed of *Hs*LARP6 amino acids (315-491). Bands are expected at 582 bp, 455 bp, 548 bp, respectively for each fragment. Figure 12B shows the OE-PCR product after PCR purification and amplification with primers that include the appropriate restriction sites on a 0.8% agarose gel. A band is expected at 1533 bp, "123X". Figure 12C shows the double digest product of the final chimeric insert with the *Nco*I/*Xho*I restriction sites for cloning into the pET28a expression plasmid previously digested. A band was expected at around 1533 bp ("Digested 123X"). The samples were then sent for commercial Sanger sequencing to confirm the insert after transformation.

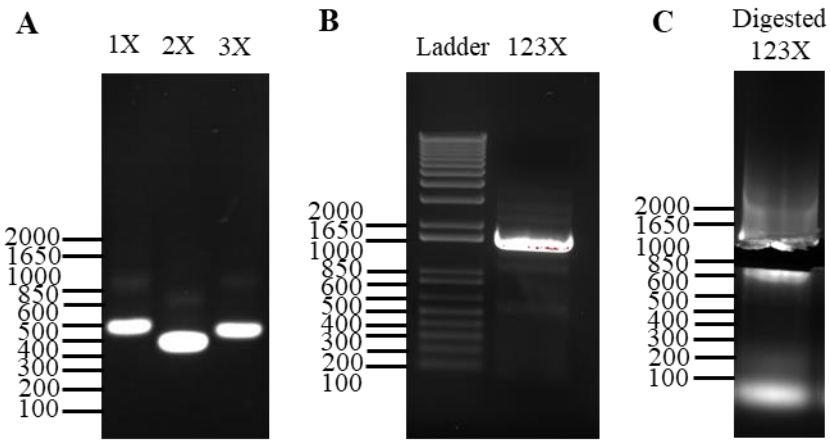

Figure 12. Amplification Products for Cloning of pET28a-*Hs*LARP6*Xm*(RRM)_FL Expression Construct. A) *Lane 1* shows amplified fragment 1X (*Hs*LARP6 amino acids 1-170) using primers from table S2. A band was expected at 582 bp. *Lane 2* and *lane 3* show amplified 2X (*Xm*LARP6 amino acids 150-292) and 3X (*Hs*LARP6 amino acids 315-491) fragments, respectively. Bands were expected at 455 bp and 548 bp. B) *Lane 1* shows the amplified product after OE-PCR and PCR purification, *Hs*LARP6*Xm*(RRM)_FL ("123X"). A band was expected at 1533 bp, representing *Hs*LARP6*Xm*(RRM)_FL. C) *Lane 1* shows the digest product for *Hs*LARP6*Xm*(RRM)_FL. A band was expected around 1533 bp.

Figure 13A shows the *Hs*LARP6*Xm*(RRM)_ΔCTD construct (amino acids = 1 - 336) (*left*) and the *Hs*LARP6*Xm*(RRM)_LaMod construct (amino acids = 91 - 336) *(right)* after amplification using the sequenced verified pET28a-*Hs*LARP6*Xm*(RRM)_FL plasmid as the template. Bands are expected at 1020 bp and 735 bp, respectively. Figure 13B shows the double-digested products. Figure 13C shows the colony PCR product to confirm the presence of our insert of interest. A band was expected at 735 bp. Colony PCR was not performed on the ΔCTD version of the protein. The samples were then sent for commercial Sanger sequencing.

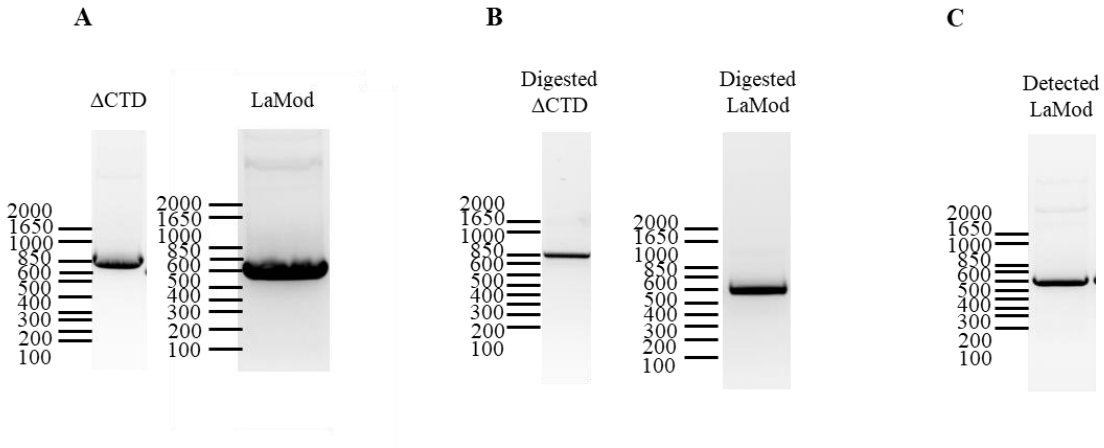

Figure 13. Amplification Products for Cloning of *Hs*LARP6*Xm*(RRM)_ΔCTD and *Hs*LARP6*Xm*(RRM)_La Module Expression Constructs. A) *Left,* shows the amplified products after removal of the C-terminal domain from the full-length version of the protein (amino acids 1 - 336). A band was expected at 1050 bp. *Right,* shows the amplified product after isolation of the chimeric La Module (amino acids 91 – 336). A band was expected at 735 bp. B) *Left,* shows the digest product of *Hs*LARP6*Xm*(RRM)_ ΔCTD and *Right,* HsLARP6*Xm*(RRM)_La Module. Bands were expected at around 1050 bp and around 735 bp, respectively. C) Colony PCR products of *Hs*LARP6*Xm*(RRM)_La Module.

The constructs *Hs*LARP6*Xm*(RRM)_FL and *Hs*LARP6*Xm*(RRM)_ΔCTD were also cloned behind the pET28-SUMO expression plasmids. Figure 14A shows the product after amplification of these constructs with restriction sites *Bam*HI/*Xho*I. Bands were expected at 1507 bp and 1050 bp. Figure 14B shows the double-digested products of the inserts. Figure 14C shows pET28-SUMO- *Hs*LARP6*Xm*(RRM)_FL and pET28-SUMO-*Hs*LARP6*Xm*(RRM)_ΔCTD colony PCR products to confirm the presence of the inserts before sending the samples out for commercial Sanger sequencing.

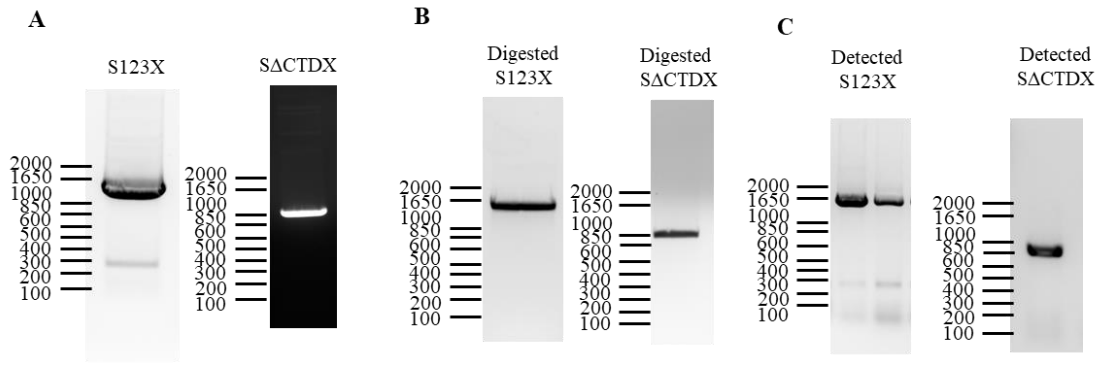

Figure 14. Amplification Products for Cloning of *Hs*LARP6*Xm*(RRM)_FL and *Hs*LARP6*Xm*(RRM)_ΔCTD Expression Constructs Behind pET28-SUMO. A) *Left*, *Hs*LARP6Xm(RRM)_FL with BamHI/XhoI restriction sites ("S123X"). A band was expected at 1507 bp. *Right*, *Hs*LARP6*Xm*(RRM)_ΔCTD with BamHI/XhoI restriction sites ("SΔCTDX"). A band was expected at around 1050 bp. B) *Left*, the double digest product after BamHI/XhoI digestions of *Hs*LARP6*Xm*(RRM)_FL ("Digested S123X"). *Right*, the double digest product after BamHI/XhoI digestions of *Hs*LARP6*Xm*(RRM)_ ΔCTD ("Digested SΔCTDX"). C) *Left* shows the band that confirmed the presence of *Hs*LARP6*Xm*(RRM)_FL after colony PCR (" Detected S123X"). Right, colony PCR product for *Hs*LARP6*Xm*(RRM)_ ΔCTD ("Detected SΔCTDX").

Cloning of the *Fish Backbone* Chimeric Expression Constructs

These recombinant proteins were composed of either the platyfish or zebrafish LARP6 protein that had the linker-RRM primary sequence exchanged for the human linker-RRM primary sequence. Three template plasmids were used and a series of primers, refer to Table 3, for each set of PCR reactions to clone the chimeric sequences into pET28-SUMO plasmids. Each chimera set is comprised of three different constructs: full-length, ΔCTD, and La Module. A schematic is shown in Figure 15.

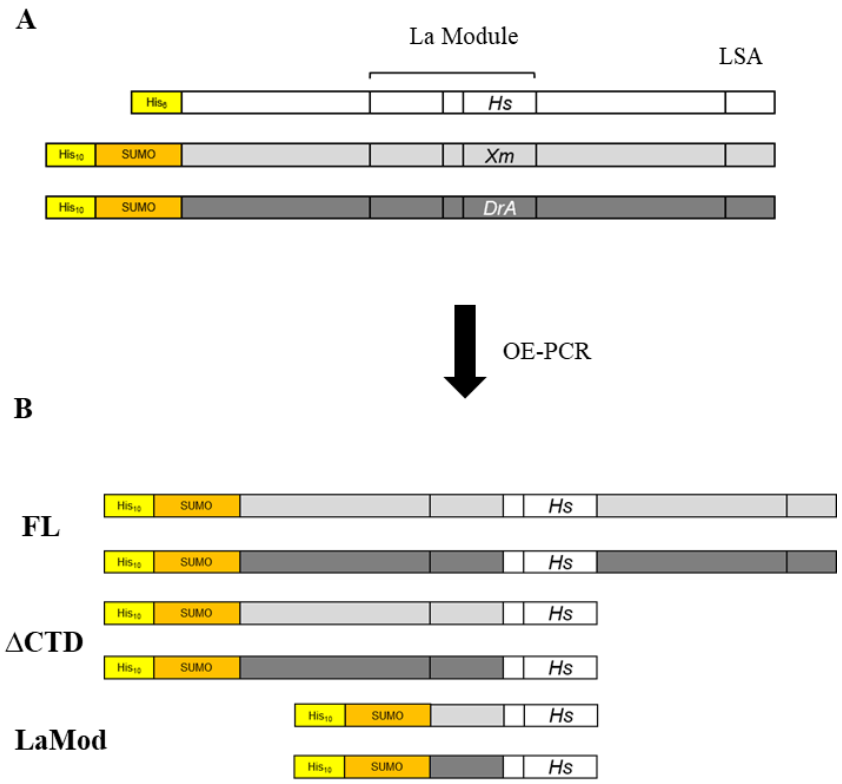

Figure 15. Fish Backbone Chimera Schematic. The LARP6 topology is represented with the presence of the La Module (La Motif – Linker – RRM1) and the LSA motif. The N-terminal His tag (His_6 or His_{10} for those constructs with the SUMO tag) is shown in yellow. The SUMO tag is shown in orange. A) At the top, the wildtype human LARP6 (*Hs*) in white, platyfish LARP6 (*Xm*) in light grey and zebrafish LARP6 (*DrA*) in dark grey. B) The full-length fish backbone chimeras retain the N-terminus, La Motif and C-terminal region of either *Xm*LARP6 or *DrA*LARP6 with linker-RRM1 sequences from *Hs*LARP6; The constructs with C-terminal deletions (ΔCTD); isolated chimeric La Module constructs (LaMod)

I. DrALARP6Hs(RRM) Chimera Set

Figure 16A shows the amplification products to each fragment to form *DrA*LARP6*Hs*(RRM)_FL, the full length construct (corresponding to amino acids =1 - 600) on a 1.5% agarose gel. The pET28-SUMO-*DrA*LARP6_FL and the pET28a-*Hs*LARP6_FL plasmids were used as templates for these PCR reactions. The fragments were designed as follows: 1D was of composed *DrA*LARP6 (corresponding residues 1 –

160), 2D was composed of *Hs*LARP6 (coding for amino acids 172 - 300), and 3D was composed of *DrA*LARP6 (corresponding to amino acids 290 – 464). Bands were expected at 480 bp, 412 bp and 601 bp, respectively. Figure 16B shows the expected band at 1465 bp for the amplified product of overlap-extension PCR, which is the full-length chimera. This product was double-digested with *Bam*HI/*Xho*I for insertion into pET28-SUMO (Figure 16C). Recovered colonies were screened by colony PCR to identify successful ligations to make the pET28-SUMO-*DrA*LARP6*Hs*(RRM)_FL expression construct **(**Figure 16D). As above, all DNA products were confirmed by commercial Sanger sequencing.

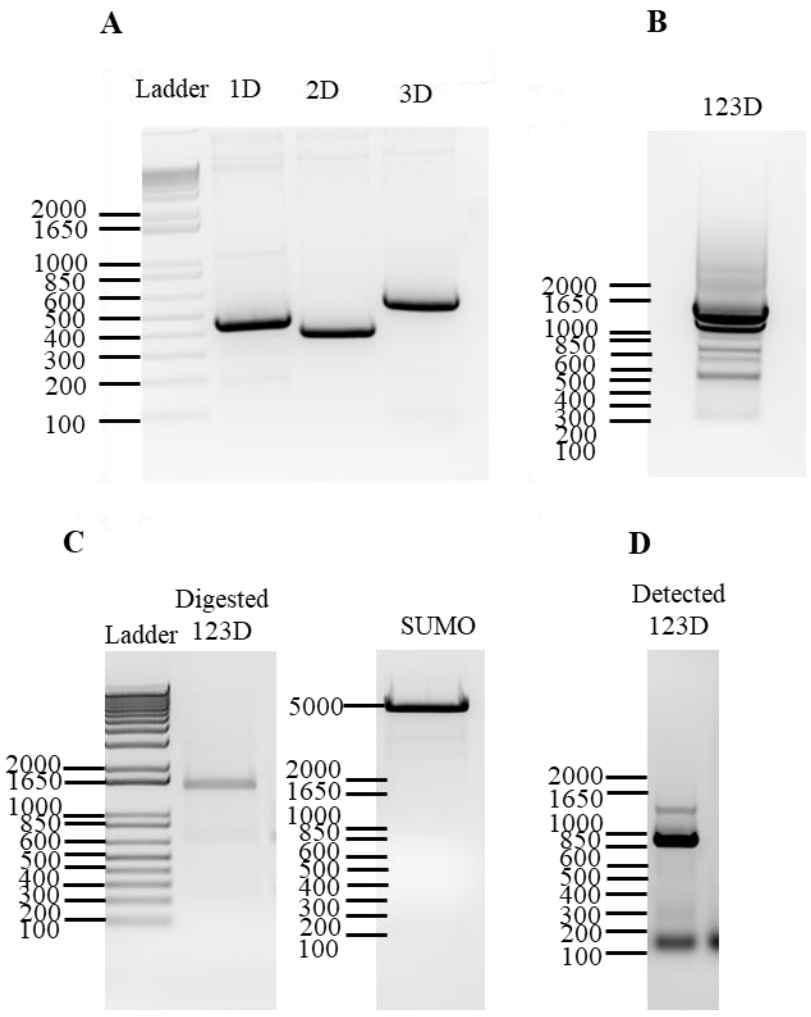

Figure 16. Amplification Products for Cloning of pET28-SUMO-*Dr*ALARP6*Hs*(RRM)_FL Expression Construct. The first lanes are taken up by the same 1 kb+ ladder. A) *Lane 2* shows amplified fragment 1D (*Dr*ALARP6 amino acids 1 - 160) using primers from Table S2. A band was expected at 480 bp. B) *Lane 2* and *lane 3* show amplified 2D (*Hs*LARP6 amino acids 172 - 300) and 3A (*Dr*ALARP6 amino acids 290 - 464) fragments. Bands were expected at 412 bp and 601 bp. B) Shows the amplified product after OE-PCR and PCR purification of the chimeric construct ("123D"). A band was expected at 1469 bp representing *Dr*ALARP6*Hs*(RRM)_FL. C) *Left* shows the digest product for *Dr*ALARP6*Hs*(RRM)_FL. A band was expected at around 1469 bp. *Right,* shows the pET28-SUMO digest product with *Bam*HI/*Xho*I digestion (SUMO). D) Amplification products of colony PCR for pET28-SUMO-*Dr*ALARP6*Hs*(RRM)_FL. A band was expected at 1465 bp, which confirmed the presence of the chimeric insert.

As for the human backbone chimeras, the ΔCTD (chimeric amino acids 1 - 407) and La Module (chimeric amino acids 181 - 407) constructs were also cloned for the zebrafish/human chimeric proteins. Figure 17A shows the amplified products after PCR with primers that include the appropriate restriction sites at the expected sizes of 883 bp and 706 bp, respectively. Following double-digest (Figure 17B), the inserts were ligated to pET28-SUMO and successful ligations identified via colony PCR (Figure 17C).

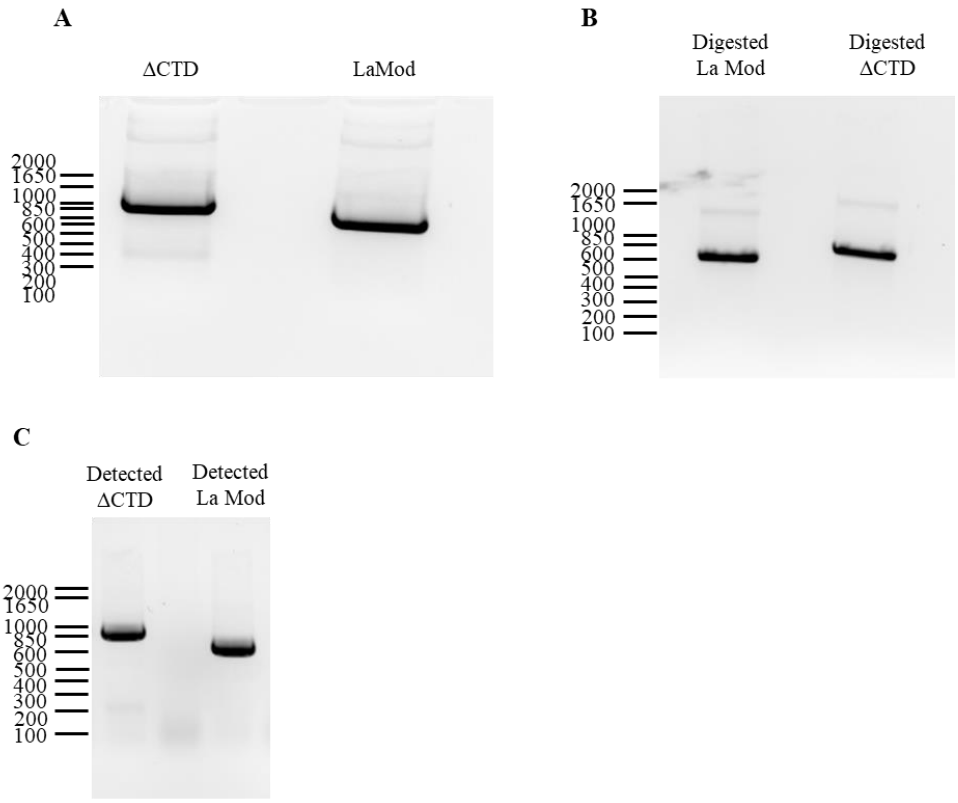

Figure 17. Amplification products for cloning of *Dr*ALARP6*Hs*(RRM)_ΔCTD and *Dr*ALARP6*Hs*(RRM)_La Module expression constructs. Gels compared to 1 Kb+ ladder. A) *Lane 1* shows the amplified products after removal of the C-terminal domain from the full-length version of the protein. A band was expected at 883 bp. *Lane 2* shows the amplified isolated chimeric La Module. A band was expected at 706 bp. B) *Lane 1* shows the digest product of *Dr*ALARP6*Hs*(RRM)_ΔCTD and *Dr*ALARP6*Hs*(RRM)_La Module. C) *Lane 1* shows the colony PCR product of *Dr*ALARP6*Hs*(RRM)_ΔCTD. *Lane 2* shows the colony PCR product of *Dr*ALARP6*Hs*(RRM)_La Module.

II. XmLARP6Hs(RRM) Chimera Set

Finally, the same approach was taken to generate the platyfish-backbone set if chimeras. OE-PCR was again used to form *Xm*LARP6*Hs*(RRM)_FL, the full length construct (corresponding to amino acids = 1 – 586). The fragments were designed as follows: 1M was composed of *Xm*LARP6 (encoding amino acids 1 – 148), 2M was composed of *Hs*LARP6 (corresponding to amino acids 172 – 300) and 3M was composed of *Xm*LARP6 (corresponding to amino acids 290 – 464). The expected bands were produced at 454 bp, 410 bp, and 579 bp, respectively (Figure 18A). The amplified, gel-purified chimeric insert from the OE-PCR was observed at 1398 bp as expected (Figure 18B). The chimeric inset was double-digested with *Bam*HI/*Xho*I (Figure 18C). Purified DNA from recovered colonies of the ligation reaction was sent for commercial Sanger sequencing to confirm the insert.

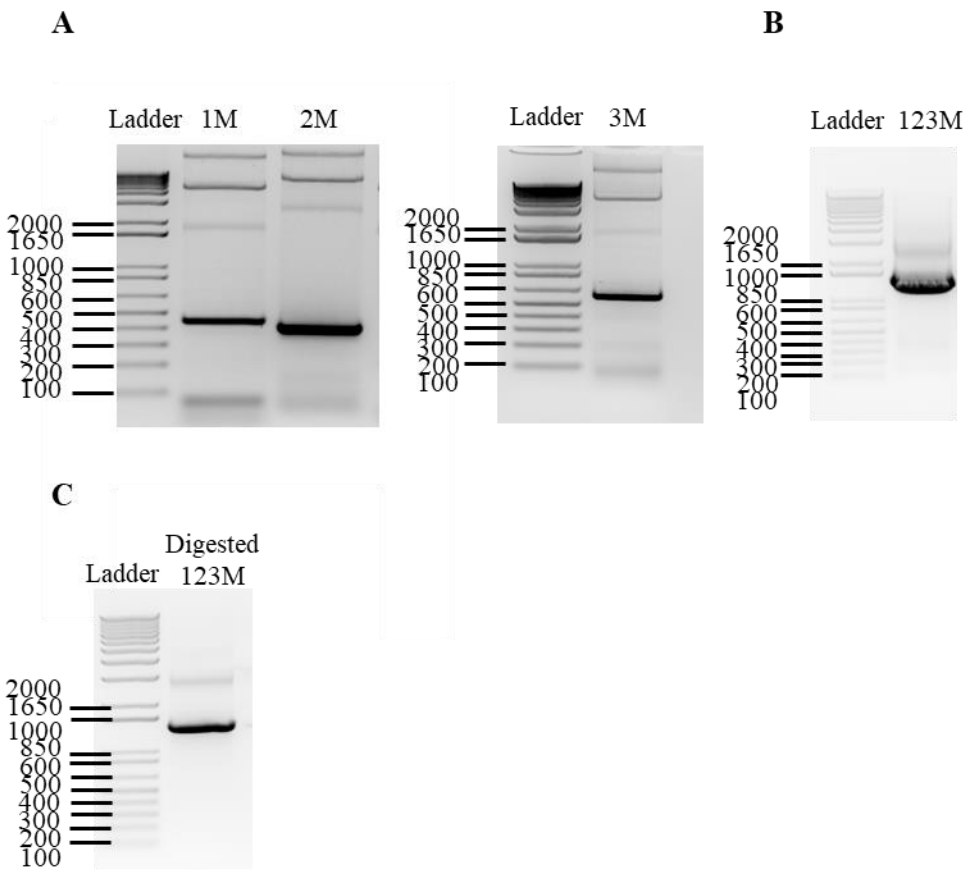

Figure 18. Amplification products for cloning of *Xm*LARP6*Hs*(RRM)_FL expression constructs. *Lane 1* of these gels are occupied by the same 1 Kb+ ladder. A) *Lane 2* shows the amplified fragment, 1M (*Xm*LARP6 amino acids 1-148), Lane 3 shows fragment 2M (*Hs*LARP6 amino acids 172 - 300). To the right, fragment 3M is shown (*Xm*LARP6 amino acids 290 - 464). Bands were expected at 454 bp, 410 bp, and 579 bp, respectively. B) Shows the amplified product after OE-PCR and PCR purification. A band is expected at 1398 bp to confirm the presence of the full-length chimera. C) The double digested product for *Xm*LARP6*Hs*(RRM)_FL.

As described above, the ΔCTD and La Module constructs were also cloned for the platyfish/human chimera. The amplified inserts for *Xm*LARP6*Hs*(RRM)_ΔCTD (chimeric amino acids 1 -398) and *Xm*LARP6*Hs*(RRM)_LaMod (chimeric amino acids 171 – 398) were observed at the expected sizes at 857 bp and 709 bp, respectively (Figure 19A). These products were double-digested in preparation for ligation into

pET28-SUMO (Figure 19B). Finally, ligated DNA was screened by colony PCR for the La Module construct (Figure 19C). Miniprepped DNA was confirmed by Sanger sequencing.

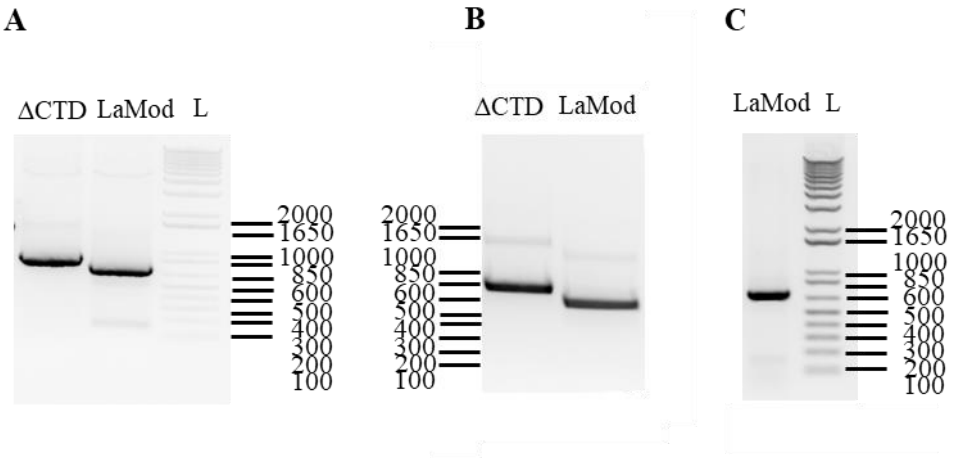

Figure 19. Amplification Products for Cloning of *Xm*LARP6*Hs*(RRM)_ΔCTD and *Xm*LARP6*Hs*(RRM)_La Module expression constructs. . Gels compared to 1 Kb+ ladder. A) *Lane 1* shows the amplified products after removal of the C-terminal domain from the full-length version of the protein (amino acids 1 - 398). A band was expected at 857 bp. *Lane 2* shows the amplified isolated chimeric La Module (171 - 398). A band was expected at 709 bp. B) *Lane 1* shows the digest product of *Xm*LARP6*Hs*(RRM)_ΔCTD and *Lane 2 Xm*LARP6*Hs*(RRM)_La Module. C) *Lane 1* shows the colony PCR product of *Xm*LARP6*Hs*(RRM)_LaModule; "L" is the ladder.

Expression of Wild Type Proteins

Transformed Rosetta ™ (DE3) pLysS competent cell culture aliquots for each construct were analyzed for protein expression. The wildtype sequences of all constructs were cloned and/or expressed by previous members of the laboratory, José Castro, Eliana Peña and Leticia Gonzalez (Table 2)[26, 31]; there were freezer stocks of all proteins except *Xm*LARP6_La Module.

Table 2. Source of Wildtype Proteins

Construct	Generator(s)
HsLARP6_FL	Mark Bayfield (York University)
DrALARP6_FL	José Castro
XmLARP6_FL	José Castro
HsLARP6_ΔCTD	Hatice Kulkoyluoglu
DrALARP6_ΔCTD	Hatice Kulkoyluoglu
XmLARP6_ΔCTD	Hatice Kulkoyluoglu
HsLARP6_LaMod	Eliana Peña; Leticia Gonzalez; this work
DrALARP6_LaMod	Hatice Kulkoyluoglu; Eliana Peña; Leticia Gonzalez
XmLARP6_LaMod	José Castro; this work

The La Module wildtype protein, *Xm*LARP6_La Module, was transformed in the same cell line. Cultures were grown using colonies from these transformations and time point aliquots were analyzed by SDS-PAGE. The *Xm*LARP6_La Module was expected to migrate at ~ 30 kDa. The protein expressed robustly indicated by the time-dependent increase in band intensity in the Coomassie-stained gel (Figure 20). The expression of this protein was confirmed by an anti-His western blot that reacts with the N-terminal His tag of all proteins analyzed in this project.

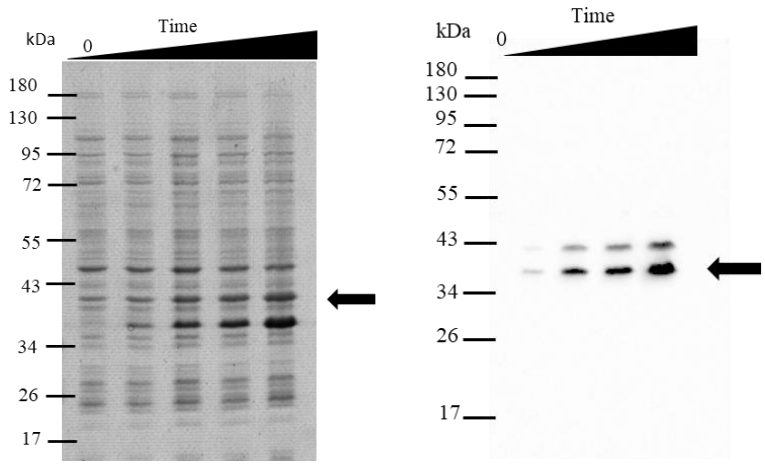

Figure 20. Expression Trials of *Xm*LARP6_La Module. pET28a-XmLARP6 La Module plasmid was transformed into *E. coli* Rosetta ™ (DE3) pLyS competent cells. Following induction with 1 mM IPTG, expression was tracked at regular time intervals of 2,4,6,8 h and overnight (~18 hrs). Cell pellets were lysed in 1x SDS-PAGE sample buffer and subjected to gel electrophoresis, followed by either (A) Coomassie blue stain or (B) anti-His western blot. Rosetta cells shows expression by the time-dependent increase of band intensity.

Expression of Human Backbone Chimeras

Transformed Rosetta™ (DE3) pLysS competent cell culture aliquots for each chimera were analyzed for protein expression. Each of the full-length chimeras were predicted to have apparent molecular weight of ~70 kDa, except for the SUMO-tagged fusion proteins that were predicted to migrate at a higher apparent molecular weight of ~90 kDa. Similarly, the ΔCTD constructs have an expected molecular weight of ~40 kDa, but are expected to migrate higher if the SUMO tag is present (~50 kDa). The chimeric La Modules were expected at molecular weights of ~30 kDa. Figure 21 shows the expression profiles for the zebrafish chimeras. The FL and ΔCTD (Figure 21A – B) constructs did not show any time-dependent increase in band intensity at the expected sizes indicating lack of expression. No His-tagged proteins were detectable, despite extensive screening of expression conditions and times (data not shown). However, the

SUMO-tagged versions of FL and ΔCTD (Figure 21D – E) showed stable expression. In contrast, the chimeric La Modules expressed without requiring the additional stability conferred by the SUMO tag (Figure 21C).

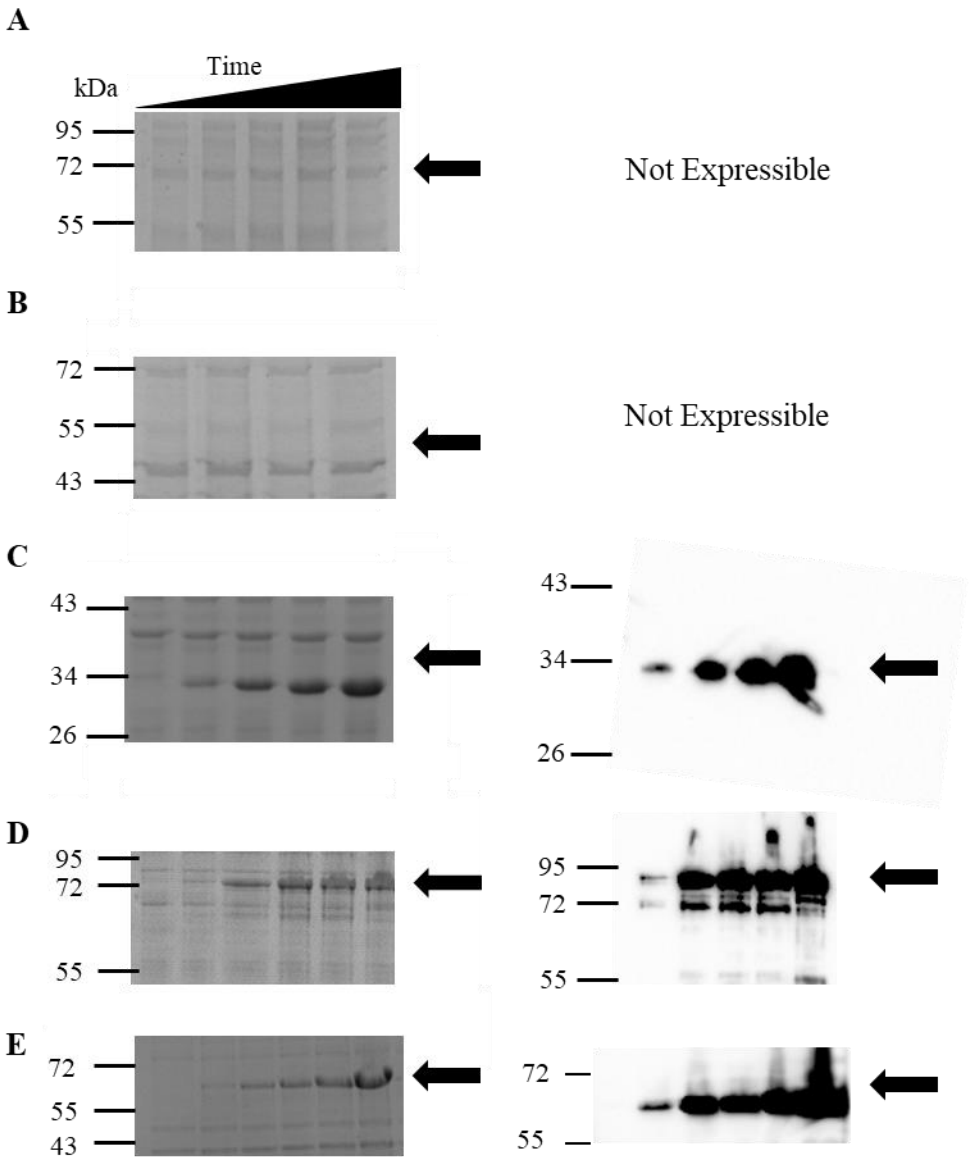

Figure 21. Expression Trials of Human Backbone-Zebrafish Chimeras. pET28 plasmids containing the human backbone chimeric sequences were transformed into E. coli Rosetta ™ (DE3) pLyS competent cells. Following induction with 1 mM IPTG, expression was tracked at regular time intervals of 2,4,6,8 h and overnight (~18 hrs). Cell pellets were lysed in 1x SDS-PAGE sample buffer and subjected to gel electrophoresis, followed by either *left* Coomassie blue stain or *right* anti-His western blot. Rosetta cells shows expression by the time-dependent increase of band intensity. A) *Hs*LARP6*DrA*(RRM)_FL,

B)*Hs*LARP6*DrA*(RRM)_ΔCTD, C) *Hs*LARP6*DrA*(RRM)_La Module, D)SUMO-*Hs*LARP6*DrA*(RRM)_FL and E) SUMO-*Hs*LARP6*DrA*(RRM)_ΔCTD.

A similar phenomenon occurred with the human/platyfish chimeras. The full-length and ΔCTD constructs were not expressible without the SUMO tag (Figure 22). As seen with the human/zebrafish chimeric La Module, the human/platyfish La Module expressed readily without the aid of the SUMO tag (Figure 22).

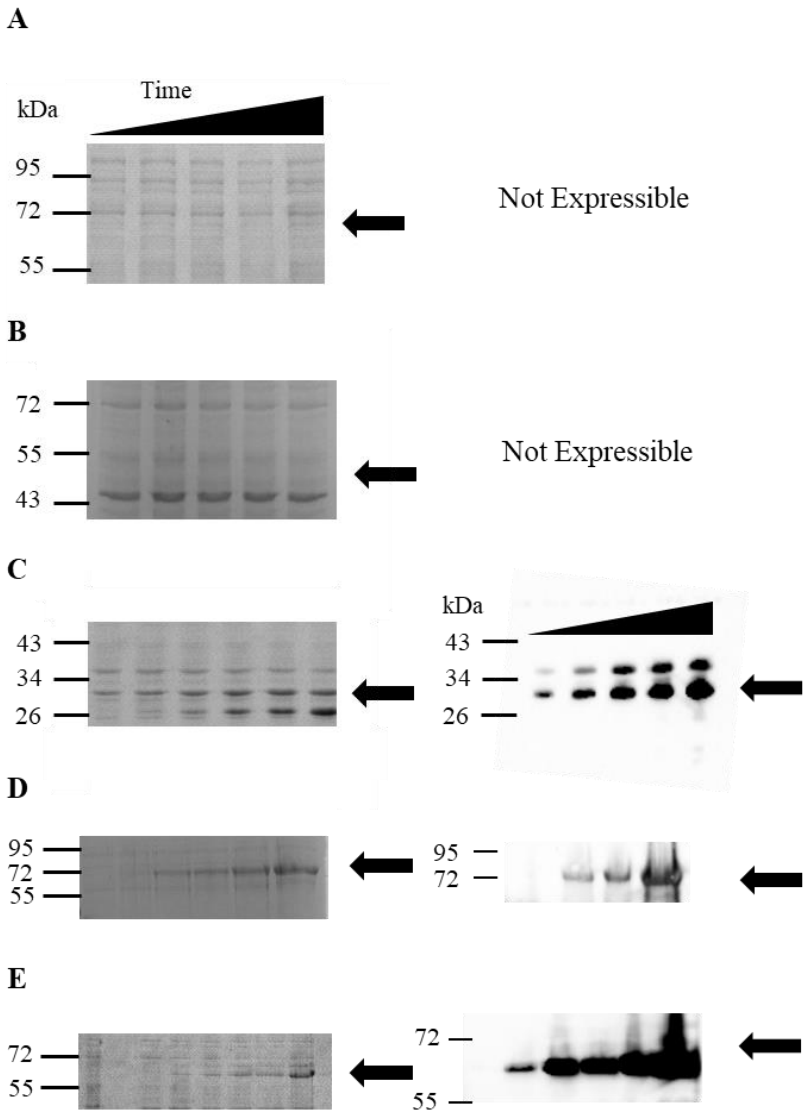

Figure 22. Expression Trials of Human Backbone-Platyfish Chimeras. pET28 plasmids containing the human backbone chimeric sequences were transformed into E. coli Rosetta ™ (DE3) pLyS competent cells. Following induction with 1 mM IPTG, expression was tracked at regular time intervals of 2,4,6,8 h and overnight (~ 18 hrs). Cell pellets were lysed in 1× SDS-PAGE sample buffer and subjected to gel electrophoresis, followed by either *left* Coomassie blue stain or *right* anti-His western blot. Rosetta cells shows expression by the time-dependent increase of band intensity. A) *Hs*LARP6*Xm*(RRM)_FL, B)*Hs*LARP6*Xm*(RRM)_ΔCTD, C) *Hs*LARP6*Xm*(RRM)_LaModule, D)SUMO-*Hs*LARP6*Xm*(RRM)_FL and E) SUMO-*Hs*LARP6*Xm*(RRM)_ΔCTD.

Since the human chimeric La Modules, pET28a-*Hs*LARP6*DrA*(RRM)_La Module and pET28a-*Hs*LARP6*Xm*(RRM)_La Module displayed stable expression they were chosen for large-scale purification. Together, these data suggest that the human/fish chimeras that contain the N-terminal region have transient instability that can be rescued by the N-terminal SUMO tag.

Expression of Fish Backbone Chimeras

The platyfish/human and zebrafish/human chimera sets were cloned as N-terminal fusions to the SUMO tag, and also expressed in Rosetta (DE3) *E. coli*. Culture aliquots for these fish/human chimeras were analyzed as described above (Figure 23). Each of the full-length chimeras had an expected molecular weight of ~90 kDa. The ΔCTD chimeras were expected at a molecular weight of ~50 kDa and the La Module chimeras were expected at a molecular weight of ~40 kDa. All six proteins showed expression as indicated by the increase in intensity at the expected size for each, and these bands were positive in the anti-His western blots that confirmed protein expression.

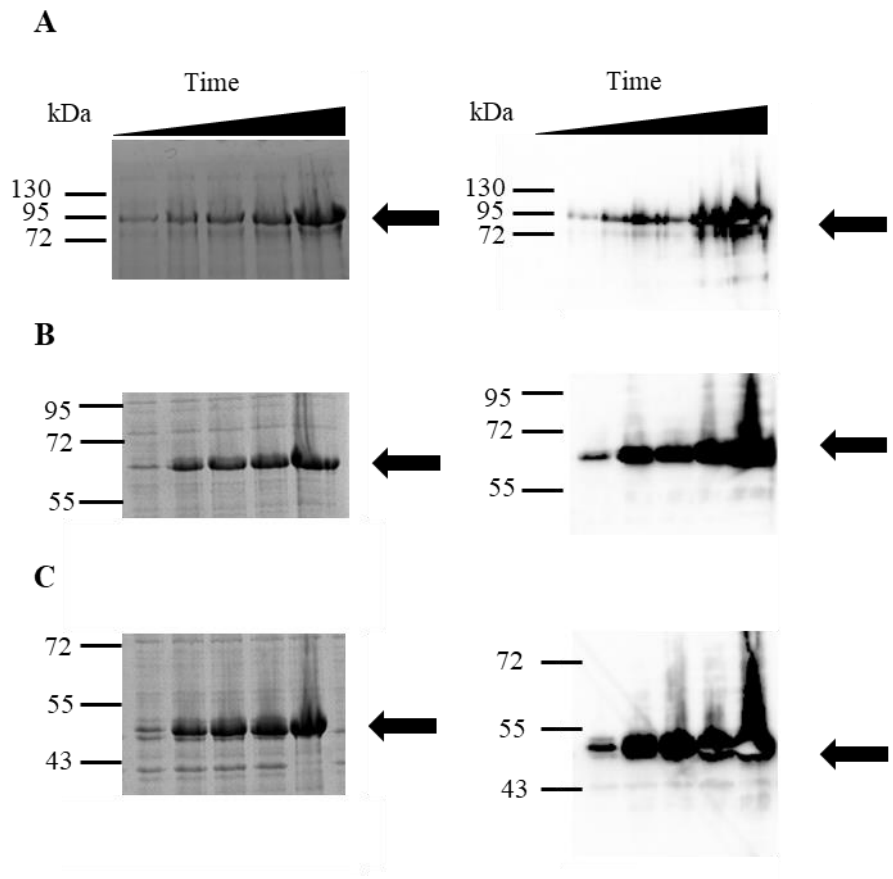

Figure 23. Expression trials of Fish Backbone Chimeras. pET28 plasmids containing the human backbone chimeric sequences were transformed into E. coli Rosetta ™ (DE3) pLyS competent cells. Following induction with 1 mM IPTG, expression was tracked at regular time intervals of 2,4,6,8 h and overnight (~18 hrs). Cell pellets were lysed in 1X SDS-PAGE sample buffer and subjected to gel electrophoresis, followed by either (A and D) Full length versions of the chimeric proteins (B and E) ΔCTD versions (C and F) La Module versions. *Left:* Coomassie blue stain. *Right:* anti-His western blot. Rosetta cells shows expression by the time-dependent increase of band intensity.

Purification of Wild Type and Chimeric La Modules

Each of the La Modules (*Xm*LARP6_La Module, *Hs*LARP6*DrA*(RRM)_La Module, and *Hs*LARP6*Xm*(RRM)_La Module was expressed in a 1 L overnight at 18 ºC. The expressed cell pellet was collected, lysed by sonication and subjected to nickel affinity chromatography as described in Methods. The elution fractions were analyzed by SDS-PAGE and Coomassie staining. The appropriate elution fractions were then pooled

and subjected to a second round of purification by size exclusion chromatography (SEC) on a Sephadex S75 column. The collected fractions were then analyzed by SDS-PAGE and the appropriate fractions were pooled, snap-frozen and stored at -70 ºC.

Wildtype La Modules - The affinity chromatography elution fractions of the wild type La Modules (*Hs*LARP6_La Module and *Xm*LARP6_La Module) were analyzed by gel electrophoresis as seen in Figure 24 and 25. As expected, the gels showed bands that migrated near the expected molecular weight and removal of other proteins occurred as indicated by the decrease of different bands. The elution fractions that showed the highest yield for each La Module were collected and prepared for size exclusion chromatography. The elution fractions 1-2 were pooled for *Hs*LARP6_La Module **(**Figure 24). The elution fractions 1-3 were pooled for *Xm*LARP6_La Module (Figure 25). The nature of the second band present at a higher molecular weight in *Xm*LARP6_La Module is unknown. The *DrA*LARP6_La Module was purified by a different member of the laboratory, Leticia Gonzalez, unpublished (Table 2). Elution of protein at Wash Buffer 2 ("W2") is expected as imidazole concentration was increased in the buffer, allowing for disruption of the interaction between the N-terminal His-tag and the nickel resin.

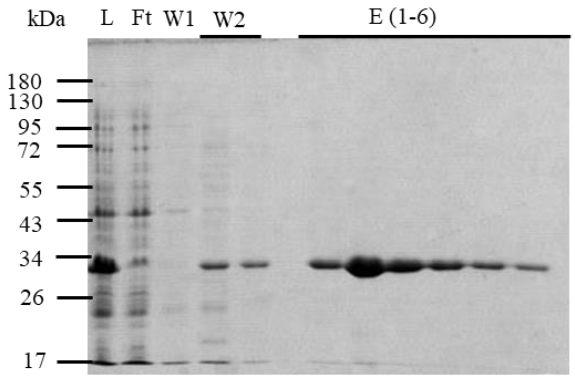

Figure 24. Affinity Chromatography Fractions for *Hs*LARP6_La Module (amino acids [70-300]). The cell lysate was purified using nickel affinity chromatography ("L", represents the expression culture cell pellet, "Ft", is the flow-through, "W1", is Wash 1, "W2", is Wash 2 and "E" is elution). Aliquots were separated by gel electrophoresis and stained with Coomassie blue. The protein was expected to migrate ~30 kDa.

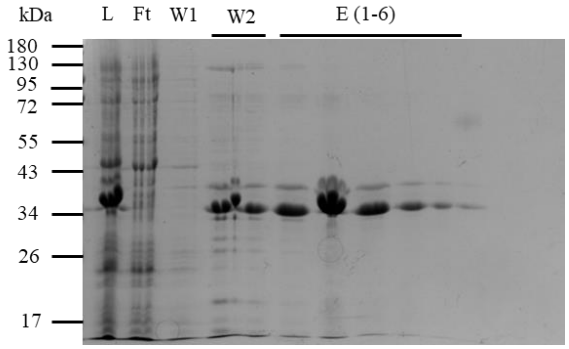

Figure 25. Affinity Chromatography Fractions for *Xm*LARP6_La Module (amino acids 170 – 292). The cell lysate was purified using nickel affinity chromatography ("L", represents the expression culture cell pellet, "Ft", is the flow-through, "W1", is Wash 1, "W2", is Wash 2 and "E" is elution). Aliquots were separated by gel electrophoresis and stained with Coomassie blue. The protein was expected to migrate ~ 30 kDa.

Similarly, the His-La Module chromatogram displayed separation of the protein from other protein species over the S75 Sephadex as monitored by UV absorbance at 280 nm. (Figure 26). *Hs*LARP6_La Module elution profile displayed one major peak at 60 mL. Based on the standards and the molecular weight calculation from the regression, it was inferred that the peak at 60 mL contained the human La Module. When calculating

the molecular weight by elution profile, the protein collected had an apparent molecular weight of 55 kDa, when its predicted molecular weight is ~30 kDa. Fractions under the peak were analyzed by gel electrophoresis and Coomassie staining to decide which fractions to pool and store at -70 ºC (Figure 27). The fractions analyzed showed the expected bands that correspond to the His-LARP6_La Module protein. In addition, these fractions showed higher purity. Fractions 16-20 were pooled for *Hs*LARP6_La Module. These fractions were concentrated and stored as 50 µL aliquots at -70 ºC for further experiments.

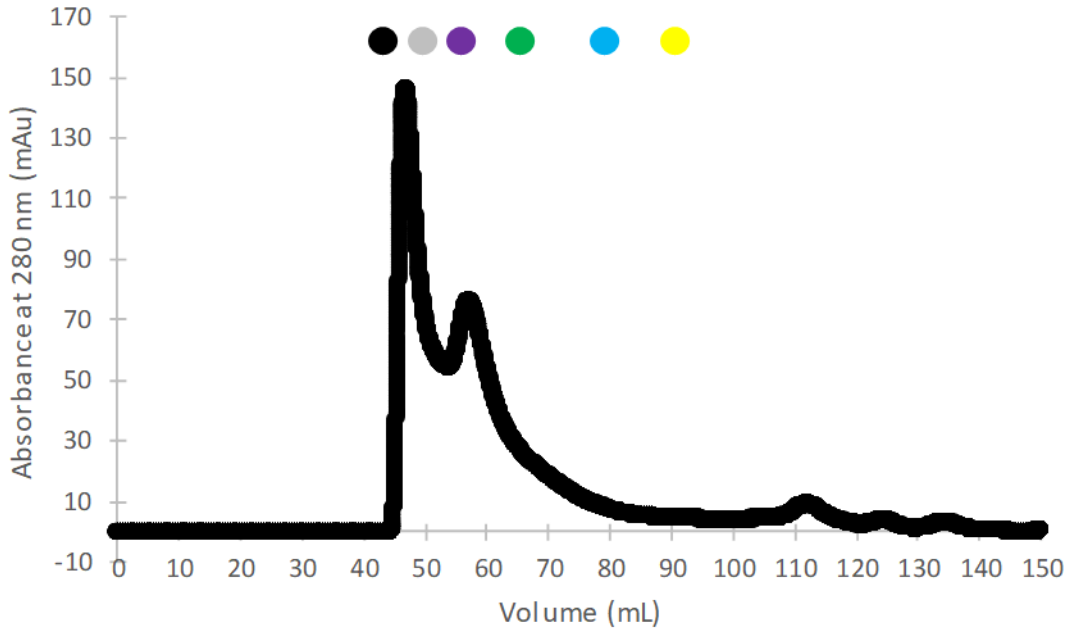

Figure 26. S75 Size Exclusion Chromatography of Human LARP6 La Module. The proteins pooled from affinity chromatography were loaded to the FPLC which was monitored by UV absorbance at 280 nm at a 1 mL/min flowrate. Fractions were collected in 2 mL aliquots automatically. The fractions were analyzed by gel electrophoresis and Coomassie staining. Colors above chromatogram represent molecular weight standards (Table 1).

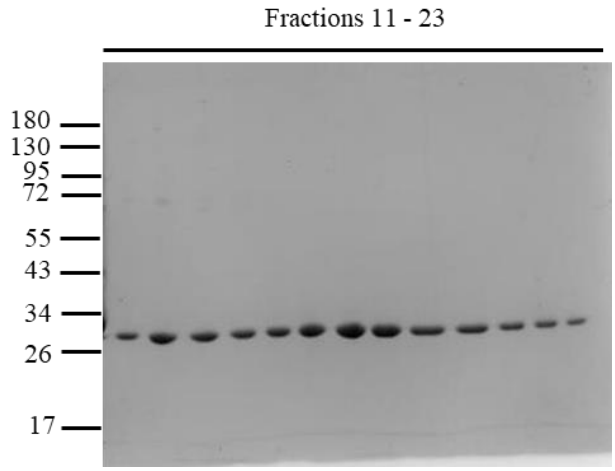

Figure 27. SDS-PAGE Analysis of Fractions from Size Exclusion Chromatography of the Human La Module. The pooled elution from affinity chromatography was filtered through a 0.2 μm filter and separated on a S75 Sephadex column. Fractions 11-23, correspond to the eluted volumes that contain the *Hs*LARP6_La Module proteins. Fractions 16 – 20 were pooled for further experimentation.

In contrast, the *Xm*LARP6_La Module chromatogram displayed one major peak at 60 mL (Figure 28). Based on the standards and the molecular weight calculation from the regression, it was inferred that the peak at 60 mL contained the platyfish La Module. When calculating the molecular weight by elution profile, the protein collected had an apparent molecular weight of 40 kDa, when its predicted molecular weight is ~30 kDa. The fractions under the peak were analyzed by gel electrophoresis and Coomassie staining to decide which fractions to pool at store at -70 ºC (Figure 29). SEC fractions 18-19 were pooled for *Xm*LARP6_La Module. These fractions were concentrated and stored as 50 μL aliquots at -70 ºC for further experiments.

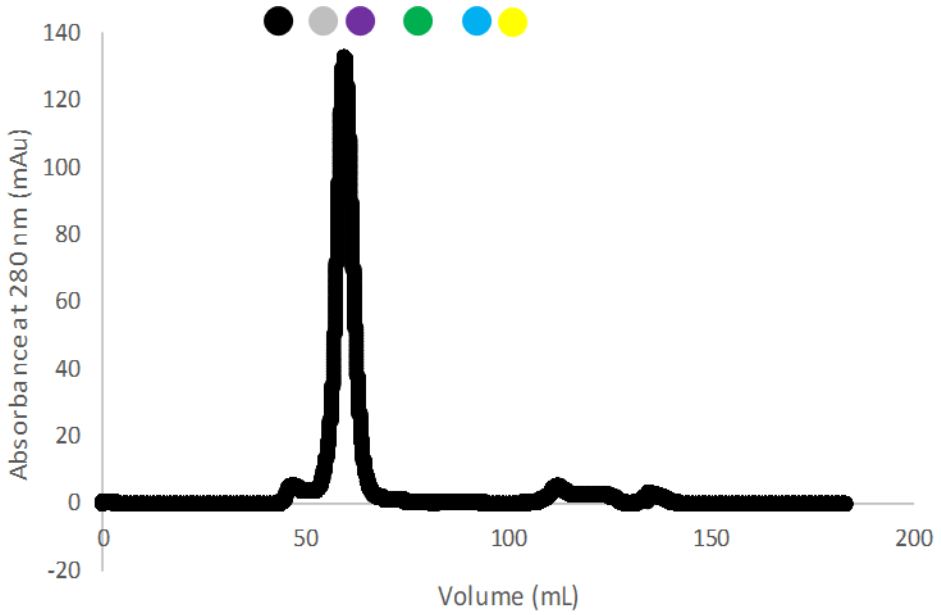

Figure 28. S75 Size Exclusion Chromatography of Platyfish LARP6 La Module. The proteins pooled from affinity chromatography were loaded to the FPLC which was monitored by UV absorbance at 280 nm at a 1 mL/min flowrate. Fractions were collected in 2 mL aliquots automatically. The fractions under the peak were analyzed by gel electrophoresis and Coomassie staining. Colors above chromatogram represent molecular weight standards (Table 1).

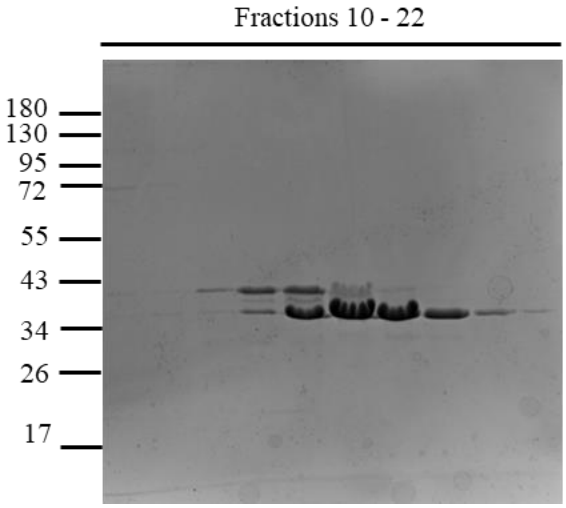

Figure 29. SDS-PAGE Analysis of Fractions from Size Exclusion Chromatography of the Platyfish La Module. The pooled elution from affinity chromatography was filtered through a 0.2 uM filter and separated on a S75 Sephadex column. Fractions 10-22 contain the *Xm*LARP6_La Module proteins. Fractions 18-19 were stored in 50 uL aliquots at -70 C for further experiments.

The fractions analyzed showed the expected bands that correspond to the His-LARP6_La Module protein. In addition, these fractions showed higher purity. Fractions 16-20 were pooled for *Hs*LARP6_La Module and fractions 18-19 were pooled for *Xm*LARP6_La Module. These fractions were concentrated and stored as 50 μL aliquots at -70 ºC for further experiments.

Purification of Human LARP6 Backbone Chimeric La Modules

The nickel affinity column elution fractions of the chimeric La Modules were analyzed as described above (Figures 30 and 31). The gels showed bands that migrated near the expected molecular weight (~ 30 kDa) and removal of other proteins occurred as indicated by the decrease of additional bands. The elution fractions that showed the highest yield for each La Module were collected and prepared for size exclusion chromatography. For each of the human/fish La Modules, elution fractions 1-2 were pooled. The nature of the second band present at a higher molecular weight in *Hs*LARP6*Xm*(RRM)_La Module is unknown, but was also present in the affinity purification profile of *Xm*LARP6_La Module and is reactive with an anti-His probe.

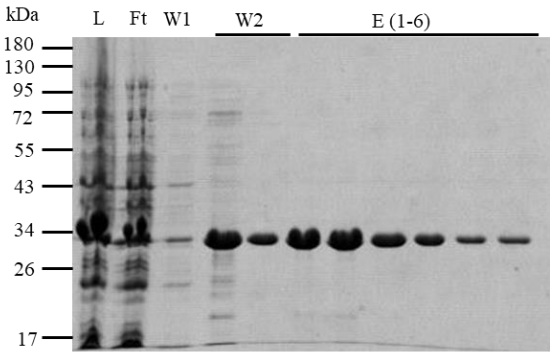

Figure 30. Affinity Chromatography Fractions for *Hs*LARP6*DrA*(RRM)_La Module. The cell lysate was purified using nickel affinity chromatography ("L", represents the expression culture cell pellet, "Ft", is the flow-through, "W1", is Wash 1, "W2", is Wash 2 and "E" is elution). Aliquots were separated by gel electrophoresis and stained with Coomassie blue. Aliquots E1-E2 were pooled for size exclusion chromatography.

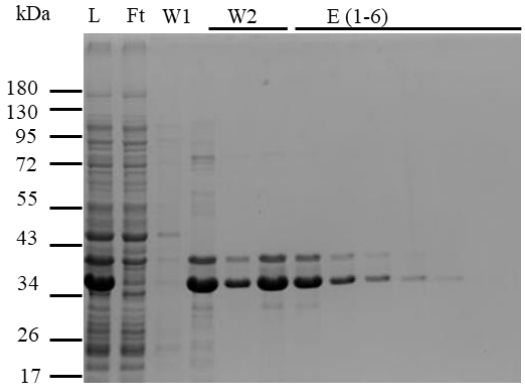

Figure 31. Affinity Chromatography Fractions for *Hs*LARP6*Xm*(RRM)_La Module. The cell lysate was purified using nickel affinity chromatography ("L", represents the expression culture cell pellet, "Ft", is the flow-through, "W1", is Wash 1, "W2", is Wash 2 and "E" is elution). Aliquots were separated by gel electrophoresis and stained with Coomassie blue. Aliquots E1-E2 were pooled for size exclusion chromatography.

The SEC chromatogram effectively separated His-tagged La Module chimeras from other species **(**Figure 32). *Hs*LARP6*DrA*(RRM)_La Module elution profile displayed one major peak at 62 mL. Based on the standards and the molecular weight calculation from the regression, it was inferred that this peak contained the

human/zebrafish La Module chimera. When calculating the molecular weight by elution profile, the protein eluted at an apparent molecular weight of 40 kDa, when its predicted molecular weight is ~30 kDa. Fractions under the peak were analyzed by gel electrophoresis to decide which fractions to pool and store (Figure 33). Fraction 18-21for *Hs*LARP6*DrA*(RRM)_La Module were pooled concentrated, filtered and stored at -70 ºC.

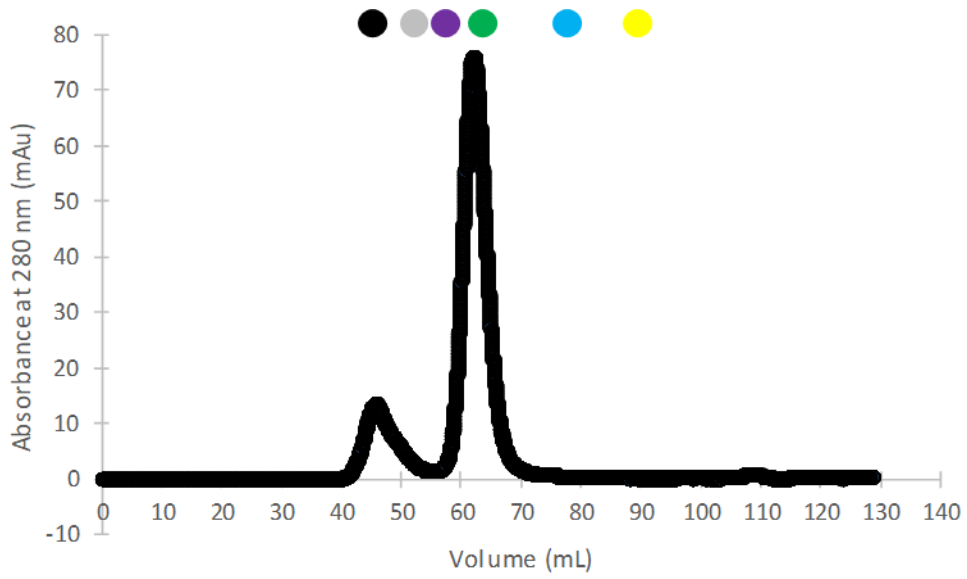

Figure 32. S75 Size Exclusion Chromatography of *Hs*LARP6*DrA*(RRM)_La Module. The proteins pooled from affinity chromatography were loaded to the FPLC which was monitored by UV absorbance at 280 nm at a 1 mL/min flowrate. Fractions were collected in 2 mL aliquots automatically. The fractions were analyzed by gel electrophoresis and Coomassie staining. Colors above chromatogram represent molecular weight standards (Table 1).

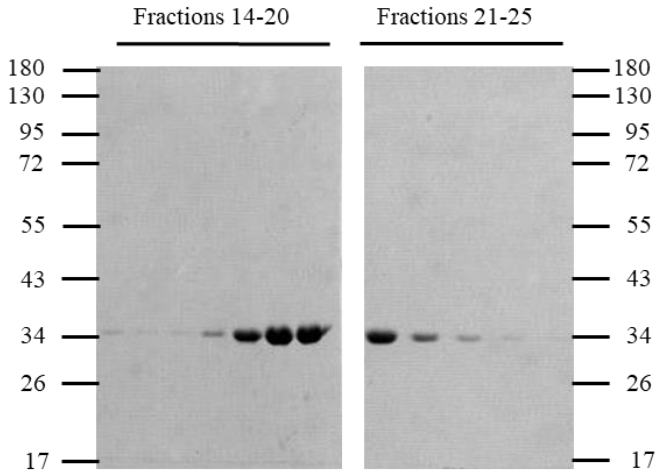

Figure 33. SDS-PAGE Analysis of Fractions from Size Exclusion Chromatography of the *Hs*LARP6*DrA*(RRM)_LaModule. The pooled elution from affinity chromatography was filtered through a 0.2 uM filter and separated on a S75 Sephadex column. Fractions 14-25 (top) contain *Hs*LARP6*DrA*(RRM)_La Module. Fractions were stored in 50 µL aliquots at -70 ºC for further experiments.

For *Hs*LARP6*Xm*(RRM)_La Module, the elution profile displayed one major peak 60 mL (Figure 34). Based on the standards and the molecular weight calculation form the regression, it was inferred that this peak contained the human/platyfish La Module Chimera. When calculating the molecular weight by elution profile, the protein collected had an apparent molecular weight of 44 kDa, when its predicted molecular weight is ~30 kDa. Fractions under the peak were analyzed by gel electrophoresis to decide which fractions to pool and store (Figure 35). The presence of the additional band in *Hs*LARP6*Xm*(RRM)_La Module was not removed by SEC (Figure 35). To combat this, the fractions with less detectible amounts of that higher band were pooled for storage. Fractions 18-20 were stored for *Hs*LARP6*Xm*(RRM)_La Module.

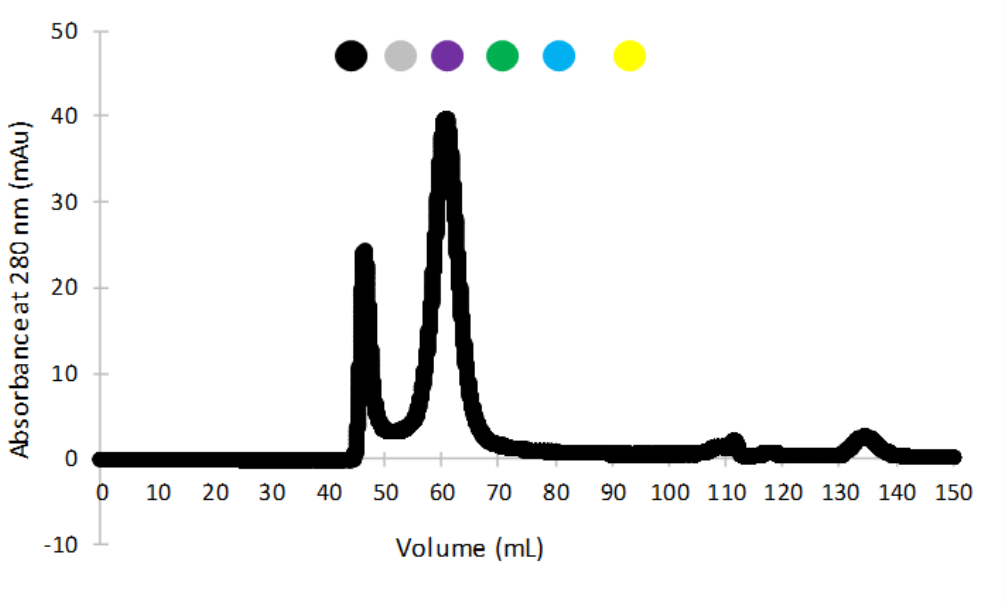

Figure 34. S75 Size Exclusion Chromatography of *Hs*LARP6*Xm*(RRM)_La Module. The proteins pooled from affinity chromatography were loaded to the FPLC which was monitored by UV absorbance at 280 nm at a 1 mL/min flowrate. Fractions were collected in 2 mL aliquots automatically. The fractions were analyzed by gel electrophoresis and Coomassie staining. Colors above chromatogram represent molecular weight standards (Table 1).

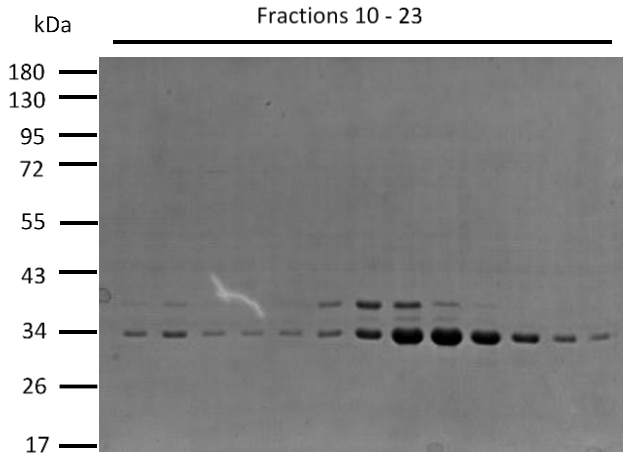

Figure 35. SDS-PAGE Analysis of Fractions from Size Exclusion Chromatography of the *Hs*LARP6*Xm*(RRM)_La Module. The pooled elution from affinity chromatography was filtered through a 0.2 µm filter and separated on a S75 Sephadex column. Fractions 14-25 (top) contain *Hs*LARP6*DrA*(RRM)_La Module. Fractions were stored in 50 µL aliquots at -70 ºC for further experiments.

Probing Protein Topology of La Modules with Limited Proteolysis

After purification, trypsin digestions of the La Module proteins was performed to probe for topology. Samples were collected at regular time intervals throughout the trypsin digestion reaction and analyzed by denaturing gel electrophoresis and silver stain (Figure 36**)**.

The trypsin digestion products of *Hs*LARP6_La Module, shows multiple accessible cleavage sites (lysines/arginines)(Figure 36A).. Three domains appear first upon time-dependent digestion at ~30 kDa, ~26 kDa and ~17 kDa. Additional fragments appear, about mid-way through the digestion.

Fewer cleavage sites are accessible in *Xm*LARP6_La Module, as evidenced by fewer degradation products being formed during the incubation (Figure 36B). This result suggests *Xm*LARP6_La Module is a more protease-resistant protein, indicating a higher stability that prevents trypsin digestion. The additional His-reactive band previously seen in the expression western blots disappears rapidly during digestion.

Finally, Figure 36C shows the digestion products for *DrA*LARP6_La Module. As seen, there are the fewest number of accessible cleavage sites in the zebrafish La Module, when compared to the other wild type La Modules. Though there is a small amount of cleavage at the later time points, the La Module mostly stays intact through the incubation. Therefore, its structure can be considered the most refractory to trypsin digestion.

The chimeric La Modules were similarly subjected to trypsin. Figure 36D shows the products after trypsin digestion for *Hs*LARP6*DrA*(RRM)_La Module. This chimera displays similar stability to the zebrafish wild type La Module with no detectable signs of

cleavage. Figure 36E shows the digest products for *Hs*LARP6*Xm*(RRM)_La Module. Interestingly, this chimera displayed similar stability to its counterpart *Xm*LARP6_La Module. We note that both the *Xm*LARP6_La Module and the *Hs*LARP6*Xm*(RRM)_La Module chimera exhibit a secondary band at about 40 kDa in early time points. This band was readily degraded during the digestion time course with *Xm*LARP6_La Module (Figure 36B). In contrast, this band persisted in the *Hs*LARP6*Xm*(RRM)_La Module chimera (Figure 36C). This suggests that the nature of this band is distinct between the two constructs and should be identified and removed for future structural characterization of these proteins.

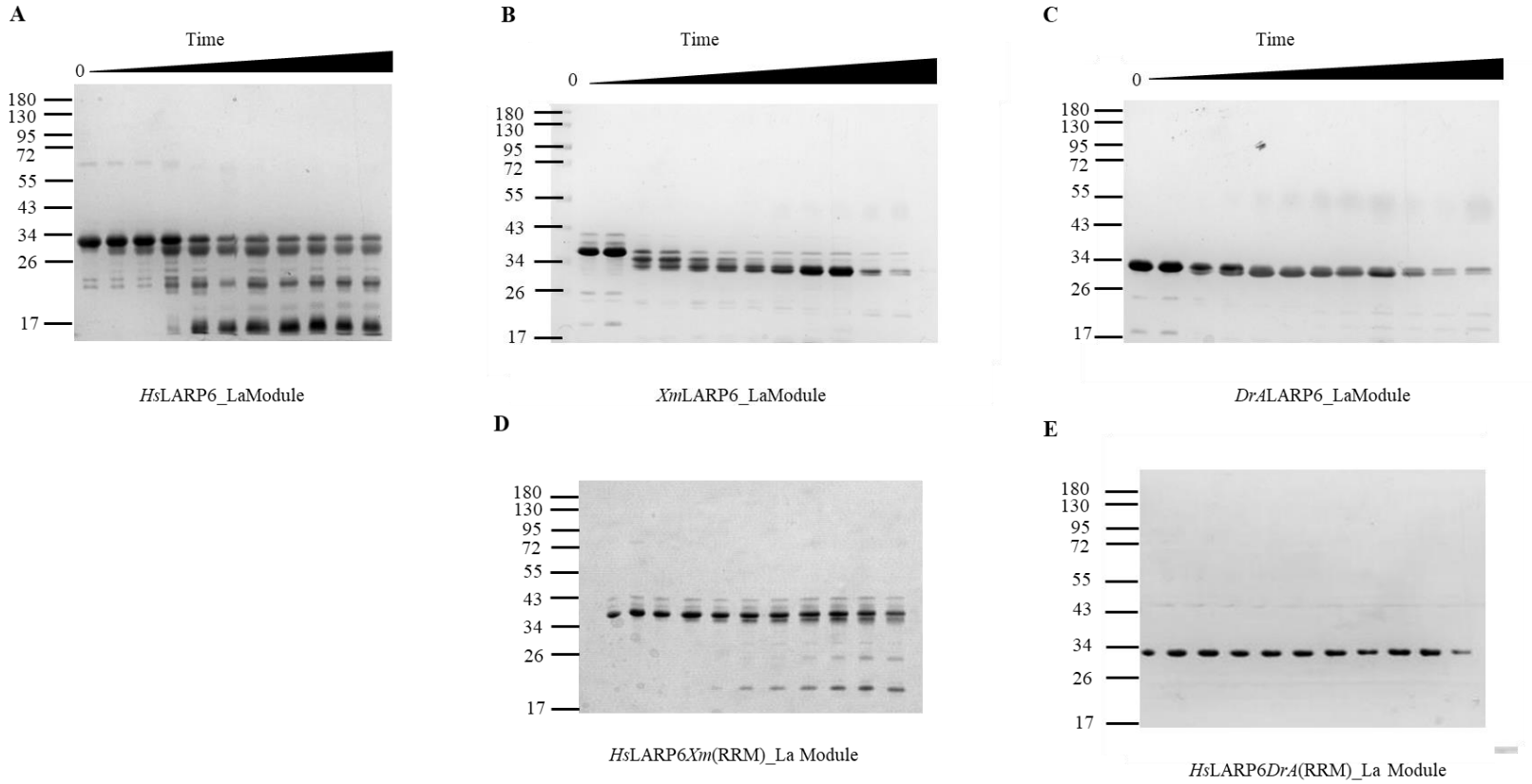

Figure 36. Topology Studies of Wild Type and Chimeric LARP6_La Module Proteins by Limited Proteolysis. Limited Trypsinolysis of La Module proteins (n = 1). Protein were incubated with a molar protein:trypsin ratio of (500:1) at 37 ºC. Aliquots were removed at 0, 1, 2, 5, 10, 20, 30, 45, 60, 120, and 180, min. Samples were subjected to gel electrophoresis and silver stained. A) *Hs*LARP6_La Module, B) *Xm*LARP6_La Module, C) *DrA*LARP6_La Module, D) *Hs*LARP6*DrA*(RRM)_La Module, E) *Hs*LARP6*Xm*(RRM)_La Module.

RNA Binding Assays: RNA Biotinylation

Biotinylated cytidine bisphosphate was conjugated to the 3' end of the human collagen type I mRNA ligand (*Hs*COL1A1). The biotinylated RNA was then analyzed through by blotting dots and then detecting the biotin with a horseradish peroxidase+streptavidin complex (HRP) which gives a quantifiable chemiluminescent signal.

Figure 37 shows the results after spotting 2 µL of biotinylated RNA in a semi-wet Hybond – N+ membrane and crosslinked. Columns A – E contain the serially diluted controls (100%, 75%, 50%, 25%, 0% biotinylated). *Hs*COL1A1 was also serially diluted (Column G). The control RNA was also biotinylated under the same conditions and serially diluted (Column I).

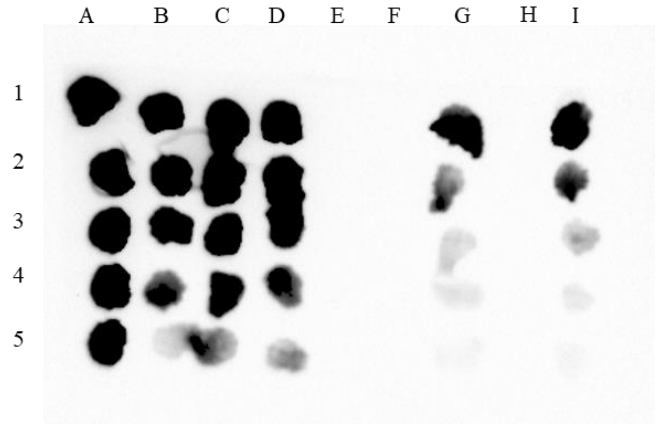

Figure 37. Biotinylation Efficiency of *Hs*COL1A1. (G) *Hs*COL1A1 compared to a series of positive biotinylated RNA control dilutions (A-E) with an exposure for 30 seconds using the ChemiDoc+ XRS system. (I) Control RNA. Rows 1 – 5 represent the sequential 1:2 dilutions for all RNAs made (from 1 to 2, from 2 to 3 etc.) (F) and (H) are empty lanes.

RNA Binding Activity Against *Hs*COL1A1

To measure RNA binding affinity of the Wild Type and chimeric LARP6 La Module proteins, electrophoretic mobility shift assays (EMSAs) using the stem-loop sequence from the 5'UTR of *Hs*COL1A1 mRNA. Protein dilutions were equilibrated on ice for 1 h with biotinylated *Hs*COL1A1 stem-loop RNA. The RNA-protein complex was separated from free RNA on a 6.5% native polyacrylamice gels with ice packs to reduce heat. Then, the RNAs were transferred to a Hybond N+ membrane and UV-crosslinked, followed by detection of biotin as described in Materials and Methods.

Figure 38 shows a representative EMSAs for the LARP6 La Modules. The bands that appear at a higher apparent molecular weight correspond to the LARP6:mRNA complex. The lower apparent molecular weight bands are unbound RNA. Qualitatively, the complex seems to be more stable in the *Xm*LARP6-containing proteins as suggested by the persistence of a sharp Bound RNA band. The LARP6:mRNA complex seems less stable in *DrA*LARP6-containing proteins as the Bound band signal smears across the lane, and appears the least stable for the native human LARP6:mRNA.

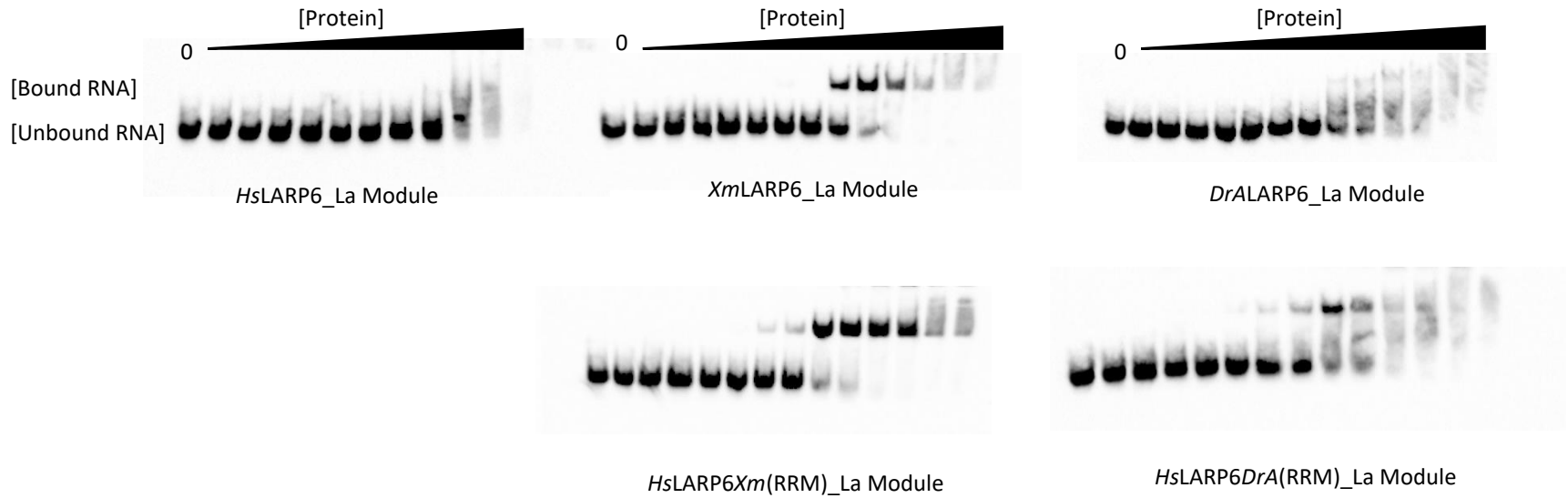

Figure 38. Representative Images of the Binding Activity of Wild Type and Chimeric LARP6_La Module Proteins with Human Collagen 1A1 mRNA Ligand. To measure the apparent binding activities of the LARP6_La Module proteins against the *Hs*COL1A1 mRNA stem loop, EMSAs were performed. Different protein concentrations were set to reach equilibrium with a defined RNA concentration. Complexes were subjected to gel electrophoresis by native gel transferred to Hybond N+ membrane and cross-linked by UV. Images are representative EMSAs of at least three independent replicates.

Using the ChemiDoc XRS+ molecular imager volume tools, pixel intensity was measured and then used to determine fractional saturation as seen in Figure 39. Fractional saturation data were plotted against protein concentration for each EMSA replicate ($N \geq 3$), and then fit with the binding isotherm equation as described in Materials and Methods (Figure 40). The human LARP6 La Module protein displayed an average $K_{D,app} = 63 \pm 22$ nM, while both fish wild type La Modules displayed tighter binding for the stem loop ligand (*Dr*ALARP6_La Module: $K_{D,app} = 37 \pm 8$ nM, *Xm*LARP6_La Module: $K_{D,app} = 9 \pm 2$ nM). The chimeric LARP6 La Module proteins both displayed tighter binding when compared to human La Module protein, but $K_{D,app}$ values were higher for the chimeric LARP6 La Modules (*Hs*LARP6*DrA*(RRM)_La Module: $K_{D,app} = 11 \pm 2$ nM, *Hs*LARP6*Xm*(RRM)_La Module: $K_{D,app} = 8 \pm 1$ nM) than for the fish La Modules. Upon exchanging the linker-RRM of the fish LARP6 La Modules for the human sequence, the fish La Modules decrease in affinity for the stem loop ligand.

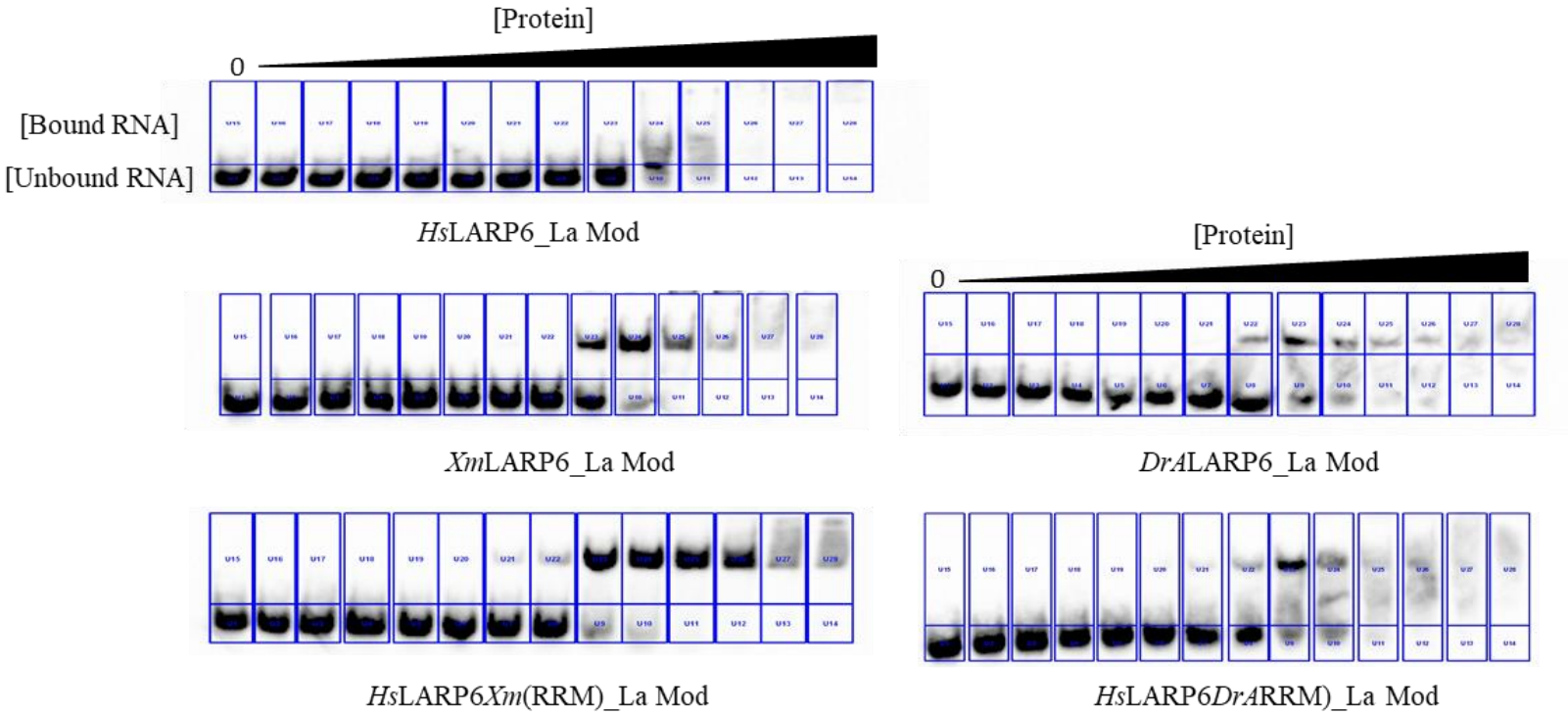

Figure 39. RNA Quantification Schemes of EMSAs Using Image Lab™ Software. Visual demonstration of quantified data that corresponds to the unquantified versions of Figure 39.

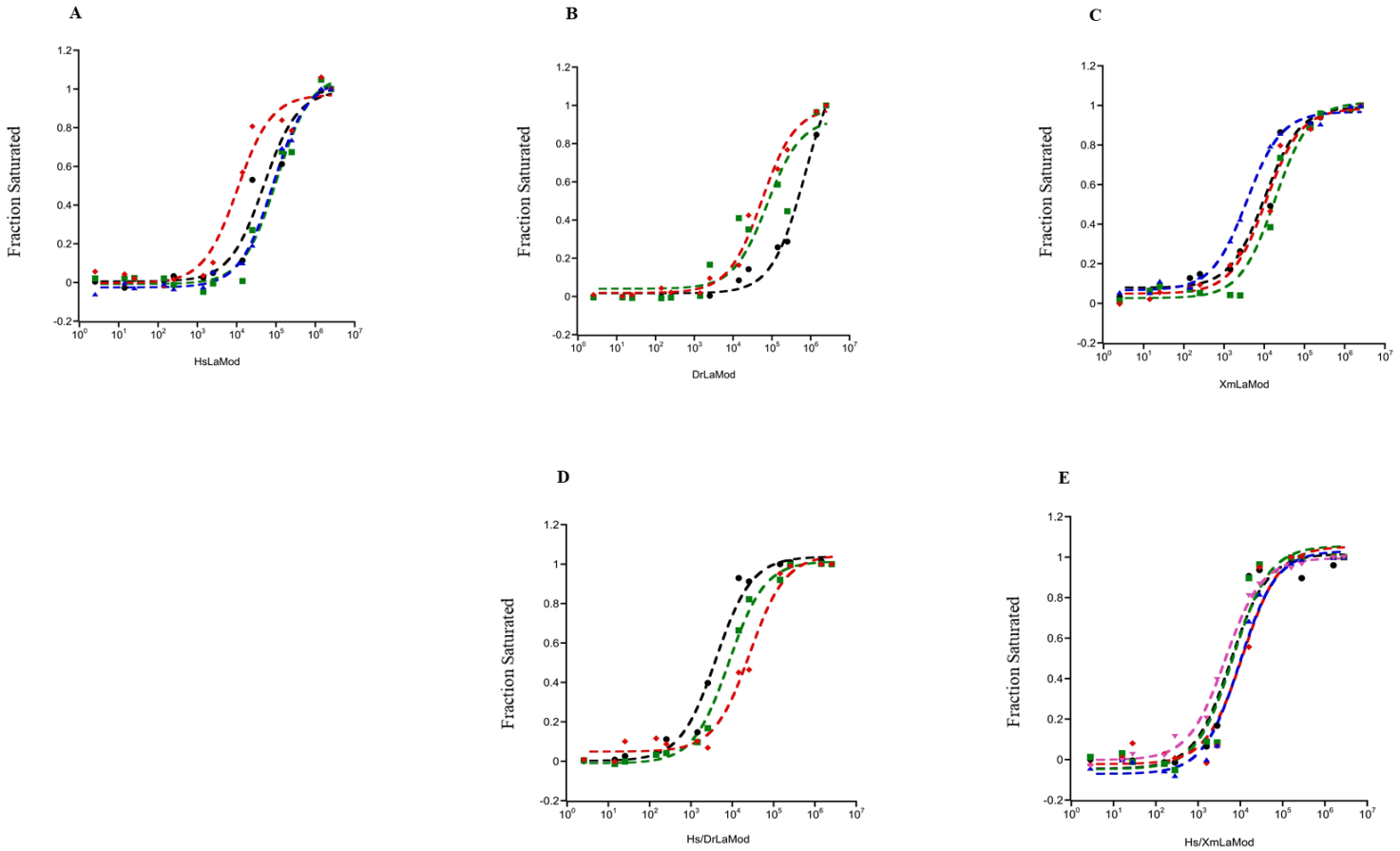

Figure 40. RNA Binding Data and Fitted Curves for Wild Type and Chimeric La Modules. The known human RNA ligand was used in these assays. RNA binding activity was observed for all three recombinant La Module proteins. Each color/symbol combination is an independent replicate. A) *Hs*LARP6_La Module has a K_{Dapp} = 63 ± 22 nM (n = 4)*;* B) *DrA*LARP6_La Module has a K_{Dapp} = 37 ± 8 nM. (n = 3) C) *Xm*LARP6_La Module has a K_{Dapp} = 9 ± 2 (n = 4); D) *Hs*LARP6*DrA*(RRM)_La Module has a K_{Dapp} = 11 ± 3 nM (n = 3); E) *Hs*LARP6*Xm*(RRM)_La Module has a K_{Dapp} = 8 ± 1 nM. (n = 5).

IV. DISCUSSION

Expression of Human LARP6 Backbone Chimeras

In this study six chimeric LARP6 constructs (*Hs*LARP6*DrA*(RRM)_FL, *Hs*LARP6*DrA*(RRM)_ΔCTD, *Hs*LARP6*DrA*(RRM)_La Module, *Hs*LARP6*Xm*(RRM)_FL, *Hs*LARP6*Xm*(RRM)_ ΔCTD, *Hs*LARP6*Xm*(RRM)_La Module) were transformed into Rosetta™ cells, expressed and purified as previously described for the full-length constructs[26]. The full-length constructs were expected to migrate at ~70 kDa, the ΔCTD chimeras were expected to migrate ~40 kDa, and the chimeric La Module proteins were expected to migrate at ~30 kDa in SDS-PAGE.

In the initial expression trials, four human/fish chimeras did not show detectable bands, by either Coomassie staining or anti-His western blotting, at the expected molecular weight in SDS-PAGE: *Hs*LARP6*DrA*(RRM)_FL, *Hs*LARP6*Xm*(RRM)_FL, *Hs*LARP6*DrA*(RRM)_ΔCTD and *Hs*LARP6*Xm*(RRM)_ΔCTD. This result indicated that these constructs lacked expression. In contrast, both the *Hs*LARP6*DrA*(RRM)_La Module and *Hs*LARP6*Xm*(RRM)_La Module proteins showed expression indicated by the presence of bands that increased in intensity and migrated near the expected molecular weights in SDS-PAGE. These bands were also reactive with an anti-His probe, which confirmed that these bands represented the His-tagged proteins of interest. The lack of expression of the full-length and ΔCTD constructs, along with a previous finding of the lab in which a novel interaction between the N-terminal domain of LARP6 and the La Module was discovered[26], indicated that there may be some species specific interactions between the N-terminal domain and the linker-RRM in *Hs*LARP6 that

stabilize the full-length protein. We hypothesize that these stabilizing interactions are disrupted in *Hs*LARP6 upon the exchange of the linker-RRM regions for the fish linker-RRM regions. To test this hypothesis, the full-length and ΔCTD constructs were cloned behind pET28-SUMO since the SUMO tag has been seen to help with expression of recombinant proteins[26, 30]. Upon expression trials as described before, both SUMO-tagged full-length and ΔCTD constructs show time-dependent increase in band intensity and anti-His-reactive bands which indicated expression of these proteins. The bands were expected to migrate at ~90 kDa and ~50 kDa, respectively, due to added weight of the SUMO tag. This leads us to infer that if species-specific interactions were disrupted by the chimeric sequence, the disruption is minor enough to be rescued by the SUMO solubility tag. The mechanism by which the SUMO tag facilitates protein folding and solubility are currently unknown and require further exploration but have been observed to increase recombinant expression in other proteins.[31, 32]

Expression of Fish LARP6 Backbone Chimeras

These fish/human chimeric proteins (*DrA*LARP6*Hs*(RRM)_FL, *DrA*LARP6*Hs*(RRM)_ΔCTD, *DrA*LARP6*Hs*(RRM)_La Module, *Xm*LARP6*Hs*(RRM)_FL, *Xm*LARP6*Hs*(RRM)_ ΔCTD, *Xm*LARP6*Hs*(RRM)_La Module) were transformed into Rosetta™ cells, expressed and purified as described before. Each of these constructs includes the SUMO tag. The FL proteins were expected to migrate at ~90 kDa, the ΔCTD at ~50 kDa, and the La Module proteins at ~40 kDa. All fish backbone chimeras stably expressed as indicated by time-dependent increase in band intensity and the His reactive bands at the expected molecular weights. However,

each of these proteins includes the SUMO tag. As non-SUMO tagged versions of these constructs have yet to be generated, it is unclear if there are similar disruptions of species-specific interactions upon the linker-RRM swaps as seen with the previous set of "human LARP6 Backbone chimeras."

Purification of Wild Type LARP6 La Modules.

In this study, the human LARP6 La Module had been previously expressed and stored as a 1L culture pellet by a previous member of the lab, Eliana Peña.[31] This pellet was used for purification. The *Hs*LARP6_La Module was predicted by sequence to have a molecular weight of ~30 kDa, which is consistent the SDS-PAGE results. However, the protein is eluting from a sizing column at an apparent molecular weight (MW_{app}) of 55 kDa. This difference in values can be due to *Hs*LARP6_La Module not being perfectly globular, since the La Module includes a flexible linker allowing for extended structures to form. These extended structures may mimic the size of a larger protein. Overall, Coomassie stained gels showed the removal of different bands (that were not *Hs*LARP6_La Module) through affinity chromatography and SEC, which supported proper purification of the protein.

The platyfish LARP6 La Module was expressed as a 1 L culture pellet and stored at – 20 ºC until prepared for purification. The *Xm*LARP6_La Module was predicted by sequence to have a molecular weight of ~30 kDa, which is consistent with the migration seen in SDS-PAGE. However, an additional band that migrates at a higher molecular weight persists throughout the purification process. The nature of this band is unknown but was also seen to be His-reactive. The protein eluted from the sizing column at an

MW_{app} = 55 kDa. To combat the contamination from the second persistent band, the SEC fractions with less detectable amounts of that contaminant were pooled and stored for further experimentation.

Purification of Human LARP6 Backbone Chimeric La Modules

As before, *Hs*LARP6*DrA*(RRM)_La Module and *Hs*LARP6*Xm*(RRM)_La Module were expressed as 1 L culture pellets and stored until prepared for purification. Like the previous La Module proteins, the chimeric La Modules were also predicted by sequence to have a molecular weight of ~30 kDa. Like before, the proteins are eluting at a slightly larger size than expected (*Hs*LARP6*DrA*(RRM)_La Module MW_{app} = 40 kDa; *Hs*LARP6*Xm*(RRM)_La Module MW_{app} = 44 kDa). Similar to *Xm*LARP6_La Module, *Hs*LARP6*Xm*(RRM)_La Module also displayed that additional band at a higher molecular weight. The same strategy was used as before to combat this contaminant in the final stored protein aliquot. The nature of this band is unknown but due to its presence in only *Xm*-containing proteins, it can be suggested that the contaminant is due to some element in the *Xm*LARP6_La Module.

Physical Characterization of LARP6_La Module Proteins

The purified La Module proteins were subjected to trypsin digest to probe the general domain topology and stability at 37 ºC (body temperature). Previous studies, such as the one conducted by Daniel, *et al*. (1982), supported a correlation between resistance to proteolysis and thermostability, in which observations suggested that factors that contribute to thermostability may protect against proteolysis.[33] Furthermore, Grutter *et al*., (1979) showed that amino acid differences can enhance or decrease thermal stability

of a protein by monitoring the melting temperature of lysozyme in its native form and after single mutations.[34] *Hs*LARP6_La Module is the least stable against trypsin digestion as it showed to have the most accessible cleavage sites indicated by the appearance of multiple bands after digestion began. There are two prominent bands between 17 kDa and 26 kDa, indicating two similarly-sized domains were produced after digestion. Due to the size of the two domains, we hypothesize that the cleavage is occurring within the linker region of the La Module. A third band is also seen near the expected molecular weight of *Hs*LARP6_La Module. This indicates cleavage of a small part of the La Module, only enough to slightly decrease its migration in the gel. Whether this cleavage occurs at the N-terminus or C-terminus is unclear but could be determined either via western blot or mass spectrometry.

The *Xm*LARP6_La Module is the second-most stable construct with fewer proteolytically-susceptible conformations when compared to *Hs*LARP6_La Module. The protein is digested quickly, producing a smaller version of the protein. It is unclear if this cleavage occurs from N-terminal or C-terminal sides. Interestingly, the contaminant band at the higher molecular weight was digested quickly, suggesting it is an unstable construct. These results show different digestion products to those seen in the same reaction but conducted at 4 ºC, as performed by José Castro (2017)[35]. These differences can be explained by temperature induced conformational changes caused at 37 ºC, which likely exposed more cleavage sites for trypsin, resulting in additional digestion products. The *Dr*ALARP6_La Module is the most stable protein used in this work as there are no clear detectable signs of trypsin digestion. The fainter bands may be due to loading

discrepancies. This suggests that the *DrAL*ARP6_La Module adopts the most resistant conformation to proteolysis.

Upon the linker-RRM sequence exchange, the *Hs*LARP6*Xm*(RRM)_ La Module displayed similar behavior to the *Xm*LARP6_La Module digestion process. However, the *Hs*LARP6*Xm*(RRM)_La Module was more stable as the intact La Module protein band remained even after digestion. In addition, smaller protein fragments appeared later during the digestion process. This may be due to the dynamic nature of proteins in solution and the presence of a flexible linker may have provided additional conformations that provided trypsin accessible sites, or unspecific cleavage. Curiously, the contaminant band was not digested away as seen in *Xm*LARP6_La Module. It can be inferred that the human La Motif in *Hs*LARP6*Xm*(RRM) provided interactions that shielded away trypsin-accessible sites. *Hs*LARP6*DrA*(RRM)_La Module showed similar results to the *DrA*LARP6_La Module digestion products. There were no consistent, detectable bands of trypsin cleavage, indicating that the potential trypsin cleavage sites were sequestered into ordered, protease-resistant structural elements.

Biochemical characterization of LARP6_La Module Protein

To test for functional protein, RNA binding assays were performed. The stem loop ligand was biotinylated at an efficiency high enough to detect binding. EMSAs were performed to measure apparent dissociation constants. All purified La Module proteins in this work were tested for binding ability. Both the wild type and chimeric La Modules retained RNA binding ability, suggesting that the linker-RRM1 exchange did not abolish RNA binding activity. The EMSA blots seen in Figure 38 are representative of the La

Module behavior. The *Hs*LARP6:mRNA complex is the least stable as indicated by the smearing of the Bound band signal. The disruption of the complex may be attributed to the stress induced by gel electrophoresis or transfer onto the membrane. The *Xm*LARP6:mRNA complex was the most stable as represented by the persistent Bound bands. Interestingly, the *Hs*LARP6*Xm*(RM)_La Module:mRNA complex showed similar results to platyfish La Module, and the *Hs*LARP6*DrA*(RM)_La Module:mRNA complex showed similar results to zebrafish La Module. This suggests that linker-RRM1 regions of the La Module may be modulating the binding activity of these proteins.

*DrA*LARP6_La Module, *Xm*LARP6_La Module, *Hs*LARP6*DrA*(RRM)_La Module and *Hs*LARP6*Xm*(RRM)_La Module bound to the stem-loop with a higher affinity than *Hs*LARP6_La Module (63 nM). As such, affinity for the stem loop is increased upon exchanging the linker-RRM1 regions of the human La Module for either the platyfish or zebrafish linker-RRM1 regions. It can be said the human linker-RRM region has significant contributions to the binding activity of the human LARP6 La Module.

Overall, the sequence divergence of the RRM seems to play a critical role in successful protein expression of the full-length human LARP6 protein by providing stabilizing interactions perhaps through the regions of localized divergence detailed here. This sequence divergence, however, did not impede RNA binding activity to the natural ligand, but it did affect the affinity and perhaps the kinetics of the complex formation/dissociation.

V. CONCLUSION

The primary aims of the study were to 1) Determine if the RRM1 domains could be exchanged between mammalian and non-mammalian vertebrate species and produce stable chimeric proteins and 2) Determine if expressible chimeric proteins exhibited RNA binding activity. The findings in this works show that 1) Exchanging the linker-RRM regions of the human LARP6 for either the platyfish or zebrafish linker-RRM regions, destabilize the full-length and ΔCTD versions of the human backbone chimeras and require the added stability of the SUMO tag to be expressible. The contributions from the solubility tag, SUMO, are still unclear.[30,31] The human backbone chimeric La Modules, however, did not require the SUMO tag to express. This suggests that in the presence of the N-terminal domain, species-specific interactions are disrupted in the chimeric proteins that are being compensated by the SUMO tag. Overall, only the isolated La Modules stably express upon the linker-RRM exchange and 2) This linker-RRM1 exchange does not abolish RNA binding activity of the human backbone La Module chimeric proteins. Furthermore, the zebrafish and platyfish La Modules show higher affinity for the *Hs*COL1A1 stem loop ligand than the human La Module. Interestingly, the human/fish chimeric La Modules, also showed higher affinity for the *Hs*COL1A1 stem loop when compared to the human La Module. The proteins that displayed this higher affinity all had a fish linker-RRM. These differences in binding affinities leads us to conclude that the linker-RRM regions of the human La Module play an important role in modulating its binding activity.

Additionally, the human La Module was discovered to be the most susceptible to trypsin digestion and the least stable of the La Module proteins studied in this work. Upon the linker-RRM exchange for a fish linker-RRM, there were less accessible cleavage sites for trypsin, indicating a possible increase in stability. Furthermore, the human/platyfish chimera displayed similar digestion products as the native platyfish La Module. The human/zebrafish chimera displayed a similar profile to the native zebrafish La Module, in which minimal digestion was observed which marked the zebrafish proteins as the most stable constructs in this work. These results further support that the linker-RRM regions play a critical role in stability, structure and function of the human LARP6 La Module.

In conclusion, the sequence divergence of the RRM in vertebrate LARP6 proteins seem to play roles that affect the stability of the full-length human LARP6 protein and the structure and function of the human LARP6 La Module.

VI. FUTURE DIRECTIONS

To test the hypothesis that it is the linker-RRM regions that modulate the binding activity of the human LARP6 La Module, the reverse fish/human La Module chimeras (*Dr*ALARP6*Hs*(RRM)_La Module and *Xm*LARP6*Hs*(RRM)_La Module) need to be purified and characterized for topology and RNA binding activity. These results should then be compared to the findings presented in this thesis. Furthermore, thermal shift assays should be performed to further investigate the structural stability of these proteins. This assay would identify the stability of the native La Modules, and the chimeric La Modules, to then compare and identify the effects of the linker-RRM sequence exchange on thermostability of the La Modules. Furthermore, the assay should be repeated with the native and chimeric La Modules in the presence of the stem loop ligand to observe thermal effects of the binding reaction. Due to the additional interactions between the RNA ligand and the La Module protein, the complex is hypothesized to require more energy for unfolding. As such, the conformational stability of the protein is expected to increase.

In addition, the full-length and ΔCTD chimeras should be also be investigated to determine contributions from the N-terminal domain to the binding activity of the LARP6 proteins. This should be investigated since a novel finding of the lab discovered a possible interaction between the N-terminal domain and the La Module in the fish full-length proteins[26]. Each full-length and ΔCTD chimera needs to be purified and characterized for topology and RNA binding activity against the stem loop ligand, as done in this thesis work.

APPENDIX SECTION

Table S1. Percent Identity and Similarity to human LARP6 La Module.

	La Motif		RRM	
Species	**Identity (%)**	**Similarity (%)**	**Identity (%)**	**Similarity (%)**
Equus caballus (Ec)	99	98	98	99
Rattus norvegicus (Rn)	95	97	94	97
Canis familiaris (Cf)	94	97	97	98
Ailuropoda melanoleuca (Am)	94	96	94	96
Mus musculus (Mm)	94	97	97	98
Gallus gallus (Gg)	81	92	70	86
Danio rerio A (DrA)	80	87	54	75
Danio rerio B (DrB)	66	85	36	58
Drosophila melanogaster (Dm)	60	82	36	55
Nematostella vectensis (Nv)	59	78	27	50
Arabidopsis thaliana A (At)	53	73	25	44
Xiphophorus maculatus (Xm)	54	71	28	55
Arabidopsis thaliana B (Atb)	42	72	26	48

Table S2. Template and Primer Combinations for Chimera Constructs

Human Backbone Chimeras	HsLARP6Xm(RRM)_FL	pET28-*Hs*LARP6	ATGGGCAGCAGCCATCATCATCATCATCACAGC	Fragment 1 & Hybrid: Forward
			TGGCACCGCGGTGGTCCTCCTCACCTTCCGGTGGT	Fragment 1: Reverse
		pET28-SUMO-*Xm*LARP6	GGACCACCGCGGTGCCAGTCTTTGCCAG	Fragment 2: Forward
			GTTCAGGTGGATCCCCCCCTCCTCGCG	Fragment 2: Reverse
		pET28-*Hs*LARP6	AGGAGGGGGGGATCCACCTGAACAAGTCCCTGAACAAGAAG	Fragment 3: Forward
			TTATACACAGGCCCTGCTCCTCTCATGG	Fragment 3 & Hybrid: Reverse
		*Hs*LARP6*Xm*(RRM)	GTACCATGGATGGGCAGCAGCCATCATCATCAT	Restriction Sites Forward: NcoI
			CTAGCTCGAGTCATTATACACAGGCCCTGCTCC	Restriction Sites Reverse: XhoI
	HsLARP6Xm(RRM)_ΔCTD	pET28-*Hs*LARP6*Xm*(RRM)	GCAACTCGAGTCATTAGCCCCCCCTCCTCGCGAGGTCG	Restriction sites Forward: NcoI
			GTACCATGGATGGGCAGCAGCCATCATCATCAT	Restriction sites Reverse: XhoI
	HsLARP6Xm(RRM)_LaMod	pET28-*Hs*LARP6*Xm*(RRM)	AGCCATATGATGACTGCAAGTGGAGGTGAGAACGAGCG	Restriction sites Forward: NdeI
			CAACTCGAGTCATTAGCCCCCCCTCCTCGCGAGGTCG	Restriction sites Reverse: XhoI
	HsLARP6DrA(RRM)_FL	pET28-*Hs*LARP6	ATGGGCAGCAGCCATCATCATCATCATCACAGC	Fragment 1 & Hybrid: Forward
			TGGCACGGTGGGGGTGGTCCTCCTCACCTTCCGGTGGT	Fragment 1: Reverse
		pET28-SUMO-*Dr*ALARP6	CCACCCCCACCGTGCCAGTTTTCGCCAGTGAATCT	Fragment 2: Foward
			TTTTTGTCTTTGGCTTTCTTCGGTGGCTTGGTGCCGATTAGCACC	Fragment 2: Reverse
		pET28-*Hs*LARP6	AGCCACCGAAGAAAGCCAAAGACAAAAATCATGACGAGGAGCCCACTGCGAGCATCCAC	Fragment 3: Forward
			TTATACACAGGCCCTGCTCCTCTCATGG	Fragment 3 & Hybrid: Reverse
		pET28-*Hs*LARP6*Xm*(RRM)	GTACCATGGATGGGCAGCAGCCATCATCATCAT	Restriction Sites Forward: NcoI
			CTAGCTCGAGTCATTATACACAGGCCCTGCTCC	Restriction Sites Reverse: XhoI
	HsLARP6DrA(RRM)_ΔCTD	pET28-*Hs*LARP6*DrA*(RRM)_FL	GTACCATGGATGGGCAGCAGCCATCATCATCAT	Restriction Sites Forward: NcoI
			GCACTCGAGCTATTATTTCTTCGGTGGCTTGGTGCCGATTAGCAC	Restriction Sites Reverse: XhoI
	HsLARP6DrA(RRM)_LaMod	pET28-*Hs*LARP6*DrA*(RRM)_FL	AGCCATATGATGACTGCAAGTGGAGGTGAGAACGAGCG	Restriction Sites Forward: NdeI
			GCACTCGAGCTATTATTTCTTCGGTGGCTTGGTGCCGATTAGCAC	Restriction Sites Reverse: XhoI

Fish Backbone Chimeras	XmLARP6Hs(RRM)_FL	pET28-SUMO-*Xm*LARP6	ATGGGGACGGTGACGATCACCGTGG	Fragment 1 & Hybrid: Forward
			GTGGGACGGGCGCAGATTTGCGGCGTACCTTACGACCC	Fragment 1: Reverse
		pET28-*Hs*LARP6_delBamHI	CAAATCTGCGCCCGTCCCACTGTTCCCCAACGAG	Fragment 2: Forward
			GCATCCCCCCCTCAGGTTTCTTTTTGGGTGGCTTCATACCAATCAGGACAGC	Fragment 2: Reverse
		pET28-SUMO-*Xm*LARP6	AAAAAGAAACCTGAGGGGGGGATGCGCAAAAGTCGTTCT	Fragment 3: Forward
			TTACTGTGTGGCAGCAGTCTTTCCCCTCTCGG	Fragment 3 & Hybrid: Reverse
		pET28-*Xm*LARP6Hs(RRM)	CATGGATCCATGGGGACGGTGACGATCACCGTGGCGATTCAGGC	Restriction Sites Forward: BamHI
			CGTACTCGAGTCATTACTGTGTGGCAGCAGTCTTTCCCCTCTCGGCAACTG	Restriction Sites Reverse: XhoI
	XmLARP6Hs(RRM)_ΔCTD	pET28-SUMO-*Xm*LARP6*Hs*(RRM)_FL	CATGGATCCATGGGGACGGTGACGATCACCGTGGCGATTCAGGC	Restriction Sites Forward :BamHI
			GCAACTCGAGTCATTAAGGTTTCTTTTTGGGTGGCTTCATACCAATCAGGACAG	Restriction Sites Reverse: XhoI
	XmLARP6Hs(RRM)_LaMod	pET28-SUMO-*Xm*LARP6Hs(RRM)_FL	CATGGATCCGGCACCAGTGGAGGGGAGTTGGAGG	Restriction Sites Forward: BamHI
			GCAACTCGAGTCATTAAGGTTTCTTTTTGGGTGGCTTCATACCAATCAGGACAG	Restriction Sites Reverse: XhoI
	DrALARP6Hs(RRM)_FL	pET28-SUMO-*Dr*ALARP6	ATGAGCAGCGAGCAGCCGCCG	Fragment 1 & Hybrid: Forward
			TGGGACGGGCGACCTGCGGCGCAC	Fragment 1: Reverse
		pET28-*Hs*LARP6	GCAGGTCGCCCGTCCCACTGTTCCCCAACGAG	Fragment 2: Forward
			GGCCGATCTTTAGGAACAGGTTTCTTTTTGGGTGGCTTCATACCAATCAGGACAGC	Fragment 2: Reverse
		pET28-SUMO-*Dr*ALARP6	ACCCAAAAAGAAACCTGTTCCTAAAGATCGGCCGAGAGATGAAGGAATCGGC	Fragment 3: Forward
			TTATAGACAGGTCTGTGTGGACGCAGTTTTGCCG	Fragment 3 & Hybrid: Reverse
		*Dr*ALARP6*Hs*(RRM)_FL	CATGGATCCATGAGCAGCGAGCAGCCGCCG	Restriction Sites Forward: BamHI
			CCGGCTCGAGTCATTATAGACAGGTCTGTGTGGACGCAGTTTTGCCGAGCGG	Restriction Sites Reverse: XhoI
	DrALARP6Hs(RRM)_ΔCTD	*Dr*ALARP6*Hs*(RRM)_FL	CATGGATCCATGAGCAGCGAGCAGCCGCCG	Restriction Sites Forward: BamHI
			GCAACTCGAGTCATTAGAGGTTTCTTTTTGGGTGGCTTCATACCAATCAGGACAG	Restriction Sites Reverse: XhoI
	DrALARP6Hs(RRM)_LaMod	*Dr*ALARP6*Hs(*RRM)_FL	TACGGATCCGGCACCAGCGGTGGCGAGCTG	Restriction Sites Forward: BamHI
			GCAACTCGAGTCATTAGAGGTTTCTTTTTGGGTGGCTTCATACCAATCAGGACAG	Restriction Sites Reverse: XhoI

Table S3. Cell line and Antibiotic Combinations

Cell Line	Antibiotic
DH5α	N/A
Escherichia coli Rosetta™ (DE3) pLysS	Chloramphenicol

Table S4. Plasmid and Antibiotic Combinations

Protein	Plasmid	Antibiotics
XmLARP6_La Module	pET28	Kanamycin
*Hs*LARP6*DrA*(RRM)_FL		
*Hs*LARP6*DrA*(RRM)_ΔCTD		
*Hs*LARP6*DrA*(RRM)_La Module		
*Hs*LARP6*Xm*(RRM)_FL		
*Hs*LARP6*Xm*(RRM)_ΔCTD		
*Hs*LARP6*Xm*(RRM)_La Module		
*Hs*LARP6*DrA*(RRM)_FL	pET28-SUMO	Kanamycin
*Hs*LARP6*DrA*(RRM)_ΔCTD		
*Hs*LARP6*Xm*(RRM)_FL		
*Hs*LARP6*Xm*(RRM)_ΔCTD		
*DrA*LARP6*Hs*(RRM)_FL		
*DrA*LARP6*Hs*(RRM)_ΔCTD		
*DrA*LARP6*Hs*(RRM)_La Module		
*Xm*LARP6*Hs*(RRM)_FL		
*Xm*LARP6*Hs*(RRM)_ΔCTD		
*Xm*LARP6*Hs*(RRM)_La Module		

Table S5. Purification Buffers

Proteins	Lysis/Wash buffer #1	Wash buffer #2	Elution buffer	SEC buffer
HsLARP6_LaMod XmLARP6_LaMod HsLARP6Xm(RRM)_LaMod HsLARP6DrA(RRM)_LaMod	50 mM NaH2PO4 [pH 7.4] 300 mM NaCl 10 mM Imidazole [pH 8]	50 mM NaH2PO4 [pH 7.4] 300 mM NaCl 75 mM Imidazole [pH 8]	50 mM NaH2PO4 [pH 7.4] 300 mM NaCl 200 mM Imidazole [pH 8]	Tris-HCl [pH 7.5 @ 4ºC] 100 mM NaCl 1% (v/v) Glycerol

List of Chimeric Sequences

1. ***Hs*LARP6*Xm*(RRM)_Full length**

MGSSHHHHHHSSGLVPRGSHMAQSGGEARPGPKTAVQIRVAIQEAEDVDELEDEEEGAETRGAGDPARYLSPGWGSASEEEPSRGHSGTTASGGENE
REDLEQEWKPPDEELIKKLVDQIEFYFSDENLEKDAFLLKHVRRNKLGYVSVKLLTSFKKVKHLTRDWRTTAHALKYSVVLELNEDHRKVRRTTAVP
VFASESLPSRMLLLSDLQKWPELAALTKDNGSNEGGATQQEQLMKLLLKAFGTYGAIASVRVLKPGKDLPADLKRLSGRYAQLGNEECAIVEFEEVEA
AVKANEAVGGEDGGTGSLGLKVVLIGTKPPKKKVLKERPREEGGIHLNKSLNKRVEELQYMGDESSANSSSDPESNPTSPMAGRRHAATNKLSPSGHQ
NLFLSPNASPCTSPWSSPLAQRKGVSRKSPLAEEGRLNCSTSPEIFRKCMDYSSDSSVTPSGSPWVRRRRQAEMGTQEKSPGTSPLLSRKMQTADGLPV
GVLRLPRGPDNTRGFHGHERSRACV

2. ***Hs*LARP6*Xm*(RRM)_ΔCTD**

MGSSHHHHHHSSGLVPRGSHMAQSGGEARPGPKTAVQIRVAIQEAEDVDELEDEEEGAETRGAGDPARYLSPGWGSASEEEPSRGHSGTTASGGENE
REDLEQEWKPPDEELIKKLVDQIEFYFSDENLEKDAFLLKHVRRNKLGYVSVKLLTSFKKVKHLTRDWRTTAHALKYSVVLELNEDHRKVRRTTAVP
VFASESLPSRMLLLSDLQKWPELAALTKDNGSNEGGATQQEQLMKLLLKAFGTYGAIASVRVLKPGKDLPADLKRLSGRYAQLGNEECAIVEFEEVEA
AVKANEAVGGEDGGTGSLGLKVVLIGTKPPKKKVLKERPREEGG

3. ***Hs*LARP6*Xm*(RRM)_La Module**

MGSSHHHHHHSSGLVPRGSHMMTASGGENEREDLEQEWKPPDEELIKKLVDQIEFYFSDENLEKDAFLLKHVRRNKLGYVSVKLLTSFKKVKHLTRD
WRTTAHALKYSVVLELNEDHRKVRRTTAVPVFASESLPSRMLLLSDLQKWPELAALTKDNGSNEGGATQQEQLMKLLLKAFGTYGAIASVRVLKPG
KDLPADLKRLSGRYAQLGNEECAIVEFEEVEAAVKANEAVGGEDGGTGSLGLKVVLIGTKPPKKKVLKERPREEGG

4. ***Hs*LARP6*DrA*(RRM)_Full Length**

MGSSHHHHHHSSGLVPRGSHMAQSGGEARPGPKTAVQIRVAIQEAEDVDELEDEEEGAETRGAGDPARYLSPGWGSASEEEPSRGHSGTTASGGENE
REDLEQEWKPPDEELIKKLVDQIEFYFSDENLEKDAFLLKHVRRNKLGYVSVKLLTSFKKVKHLTRDWRTTAHALKYSVVLELNEDHRKVRRTTPTVP
VFASESLPSRMLLLSELKRWPELGIALGGDSNNGSGPTQQERLMELLLKAFGNYGPIASVRVLKPGKDLPADLKKLSGRYSQLGTEECAIVEFEEVEAA
MKAHEAVGGEGGNRGPLGLKVVLIGTKPPKKAKDKNHDEEPTASIHLNKSLNKRVEELQYMGDESSANSSSDPESNPTSPMAGRRHAATNKLSPSGH
QNLFLSPNASPCTSPWSSPLAQRKGVSRKSPLAEEGRLNCSTSPEIFRKCMDYSSDSSVTPSGSPWVRRRRQAEMGTQEKSPGTSPLLSRKMQTADGLP
VGVLRLPRGPDNTRGFHGHERSRACV

***Hs*LARP6*DrA*(RRM)_ΔCTD**

MGSSHHHHHHSSGLVPRGSHMAQSGGEARPGPKTAVQIRVAIQEAEDVDELEDEEEGAETRGAGDPARYLSPGWGSASEEEPSRGHSGTTASGGENE
REDLEQEWKPPDEELIKKLVDQIEFYFSDENLEKDAFLLKHVRRNKLGYVSVKLLTSFKKVKHLTRDWRTTAHALKYSVVLELNEDHRKVRRTTPTVP
VFASESLPSRMLLLSELKRWPELGIALGGDSNNGSGPTQQERLMELLLKAFGNYGPIASVRVLKPGKDLPADLKKLSGRYSQLGTEECAIVEFEEVEAA
MKAHEAVGGEGGNRGPLGLKVVLIGTKPPKK

5. ***Hs*LARP6*DrA*(RRM)_La Module**

MGSSHHHHHHSSGLVPRGSHMMTASGGENEREDLEQEWKPPDEELIKKLVDQIEFYFSDENLEKDAFLLKHVRRNKLGYVSVKLLTSFKKVKHLTRD
WRTTAHALKYSVVLELNEDHRKVRRTTPTVPVFASESLPSRMLLLSELKRWPELGIALGGDSNNGSGPTQQERLMELLLKAFGNYGPIASVRVLKPGK
DLPADLKKLSGRYSQLGTEECAIVEFEEVEAAMKAHEAVGGEGGNRGPLGLKVVLIGTKPPKK

6. ***Xm*LARP6*Hs*(RRM)_Full length**

MGSSHHHHHHSSGLVPRGSHMAQSGGEARPGPKTAVQIRVAIQEAEDVDELEDEEEGAETRGAGDPARYLSPGWGSASEEEPSRGHSGTTASGGENE
REDLEQEWKPPDEELIKKLVDQIEFYFSDENLEKDAFLLKHVRRNKLGYVSVKLLTSFKKVKHLTRDWRTTAHALKYSVVLELNEDHRKVRRTTAVP
VFASESLPSRMLLLSDLQKWPELAALTKDNGSNEGGATQQEQLMKLLLKAFGTYGAIASVRVLKPGKDLPADLKRLSGRYAQLGNEECAIVEFEEVEA
AVKANEAVGGEDGGTGSLGLKVVLIGTKPPKKKVLKERPREEGGIHLNKSLNKRVEELQYMGDESSANSSSDPESNPTSPMAGRRHAATNKLSPSGHQ
NLFLSPNASPCTSPWSSPLAQRKGVSRKSPLAEEGRLNCSTSPEIFRKCMDYSSDSSVTPSGSPWVRRRRQAEMGTQEKSPGTSPLLSRKMQTADGLPV
GVLRLPRGPDNTRGFHGHERSRACV

7. ***Xm*LARP6*Hs*(RRM)_ΔCTD**

MGHHHHHHHHHHSSGHIEGRHMASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQA
DQTPEDLDMEDNDIIEAHREQIGGSMGTVTITVAIQAAEDEEPEEEHPGNVECLRGSCSEDELGRHDKSRHSGAGTSGGELEEESWQPPDTELIQKLVTQ
IEFYLSDENLEHDAFLLKHVRRNKLGFVSVKLLTSFKKVKHLTRDWRTTAYALKHSKILELNDEGRKVRRKSAPVPLFPNENLPSKMLLVYDLYLSPKL
WALATPQKNGRVQEKVMEHLLKLFGTFGVISSVRILKPGRELPPDIRRISSRYSQVGTQECAIVEFEEVEAAIKAHEFMITESQGKENMKAVLIGMKPPK
KKP

8. ***Xm*LARP6*Hs*(RRM)_La Module**

MGHHHHHHHHHHSSGHIEGRHMASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQA
DQTPEDLDMEDNDIIEAHREQIGGSGTSGGELEEESWQPPDTELIQKLVTQIEFYLSDENLEHDAFLLKHVRRNKLGFVSVKLLTSFKKVKHLTRDWRT
TAYALKHSKILELNDEGRKVRRKSAPVPLFPNENLPSKMLLVYDLYLSPKLWALATPQKNGRVQEKVMEHLLKLFGTFGVISSVRILKPGRELPPDIRRI
SSRYSQVGTQECAIVEFEEVEAAIKAHEFMITESQGKENMKAVLIGMKPPKKKP

9. ***Dr*LARP6A*Hs*(RRM)_Full length**

MGHHHHHHHHHHSSGHIEGRHMASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQA
DQTPEDLDMEDNDIIEAHREQIGGSMSSEQPPREISAPVTITVAIQAAEEDDEPDEEPSCNTIELQTGSGSEDELGRHDKSSGAGTSGGELEEESWQPPDPE
LIQKLVAQIEYYLSDENLEHDAFLLKHVRRNKLGFVSVKLLTSFKKVKHLTRDWRTTAYALRHSNLLELNDEGRKVRRRSPVPLFPNENLPSKMLLVY
DLYLSPKLWALATPQKNGRVQEKVMEHLLKLFGTFGVISSVRILKPGRELPPDIRRISSRYSQVGTQECAIVEFEEVEAAIKAHEFMITESQGKENMKAV
LIGMKPPKKKPVPKDRPRDEGIGGMRKSRSLNSRVRELQYHGDDSAASSSETESNPTSPRLARKSRSCNKLSPTSAGPNHLSPVVSPRSSPWSSPRASPCT
QRKTHPSGKSPLASEGRLSPEPGRRWADYSSDSSLTPSGSPWVQRRKQVASQESSPVGSPMLARKIQNADGLPPGVVRLPRGPDGTRGFHCPPLGKTAS
TQTCL

10. ***Dr*LARP6A*Hs*(RRM)_ΔCTD**

MGHHHHHHHHHHSSGHIEGRHMASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQA
DQTPEDLDMEDNDIIEAHREQIGGSMSSEQPPREISAPVTITVAIQAAEEDDEPDEEPSCNTIELQTGSGSEDELGRHDKSSGAGTSGGELEEESWQPPDPE
LIQKLVAQIEYYLSDENLEHDAFLLKHVRRNKLGFVSVKLLTSFKKVKHLTRDWRTTAYALRHSNLLELNDEGRKVRRRSPVPLFPNENLPSKMLLVY
DLYLSPKLWALATPQKNGRVQEKVMEHLLKLFGTFGVISSVRILKPGRELPPDIRRISSRYSQVGTQECAIVEFEEVEAAIKAHEFMITESQGKENMKAV
LIGMKPPKKKP

11. ***Dr*LARP6A*Hs*(RRM)_La Module**

HHHHHHHHHHSSGHIEGRHMASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQT
PEDLDMEDNDIIEAHREQIGGSGTSGGELEEESWQPPDPELIQKLVAQIEYYLSDENLEHDAFLLKHVRRNKLGFVSVKLLTSFKKVKHLTRDWRTTAY
ALRHSNLLELNDEGRKVRRRSPVPLFPNENLPSKMLLVYDLYLSPKLWALATPQKNGRVQEKVMEHLLKLFGTFGVISSVRILKPGRELPPDIRRISSRY
SQVGTQECAIVEFEEVEAAIKAHEFMITESQGKENMKAVLIGMKPPKKKP

REFERENCES

1. Glisovic, T., et al. (2008). RNA-binding proteins and post-transcriptional gene regulation. *FEBS Letters*. 582, 1977-1986.
2. Ray, D., et al. (2013). A compendium of RNA-binding motifs for decoding gene regulation. *Nature.* 499, 172.
3. Lunde, B. M., et al. (2007). RNA-binding proteins: modular design for efficient function. *Nature reviews*. 8, 479-490.
4. Mackereth, C. D. and M. Sattler (2012). Dynamics in multi-domain protein recognition of RNA. *Current Opinion in Structural Biology.* 22, 287-296.
5. Maris, C., et al. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS Journal.* 272, 2118-2131.
6. Nagai, K. (1996). RNA—protein complexes. *Current Opinion in Structural Biology* 6, 53-61.
7. Bousquet-Antonelli, C. and J.-M. Deragon (2009). A comprehensive analysis of the La-motif protein superfamily. *RNA.* 15, 750-764.
8. Stavraka, C. and S. Blagden (2015). The La-Related Proteins, a Family with Connections to Cancer. *Biomolecules.* 5, 2701-2722.
9. Maraia, R. J., et al. (2017). The La and related RNA-binding proteins (LARPs): structures, functions, and evolving perspectives. *Wiley Interdisciplinary Reviews: RNA.* 8, 1430.
10. Martino, L., et al. (2015). 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 Research.* 43, 645-660.
11. Burrows, C., et al. (2010). The RNA binding protein Larp1 regulates cell division, apoptosis and cell migration. *Nucleic Acids Research*. 38, 5542-5553.
12. Yang, R., et al. (2011). 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. *Molecular and Cellular Biology.* 31, 542-556.
13. Markert, A., et al. (2008). The La-related protein LARP7 is a component of the 7SK ribonucleoprotein and affects transcription of cellular and viral polymerase II genes. *EMBO Reports.* 9, 569-575.
14. Cai, L., et al. (2010). Binding of LARP6 to the Conserved 5’ Stem-Loop Regulates Translation Of mRNAs Encoding Type I Collagen. *Journal of molecular biology.* 395, 309-326.
15. Stefanovic, L., et al. (2014). Characterization of binding of LARP6 to the 5’ stem-loop of collagen mRNAs: Implications for synthesis of type I collagen. *RNA Biology*. 11, 1386-1401.
16. Alspaugh, M. A., et al. (1976). Differentiation and characterization of autoantibodies and their antigens in sjögren's syndrome. *Arthritis & Rheumatism*. 19, 216-222.

17. Alfano, C., et al. (2004). Structural analysis of cooperative RNA binding by the La motif and central RRM domain of human La protein. *Nature Structural &Amp; Molecular Biology*. 11, 323.
18. S. L. W. and T. Cedervall (2002). The La Protein. *Annual Review of Biochemistry.* 71, 375-403.
19. Kotik-Kogan, O., et al. (2008). Structural Analysis Reveals Conformational Plasticity in the Recognition of RNA 3′ Ends by the Human La Protein. *Structure*. 16, 852-862
20. Uchikawa, E., et al. (2015). Structural insight into the mechanism of stabilization of the 7SK small nuclear RNA by LARP7. *Nucleic Acids Research***.** 43, 3373-3388.
21. Wang, H. and B. Stefanovic (2014). Role of LARP6 and Nonmuscle Myosin in Partitioning of Collagen mRNAs to the ER Membrane. *PLoS ONE.* 9, e108870.
22. Valavanis, C., et al. (2007). Acheron, a novel member of the Lupus Antigen family, is induced during the programmed cell death of skeletal muscles in the moth Manduca sexta. *Gene.* 393, 101-109.
23. Glenn, H. L., et al. (2010). Acheron, a Lupus antigen family member, regulates integrin expression, adhesion, and motility in differentiating myoblasts.*" American journal of physiology, Cell physiology.* 298, C46-55.
24. Weng, H., et al. (2009). Acheron, a novel LA antigen family member, binds to cask and forms a complex with id transcription factors. *Cellular & Molecular Biology Letters*. 14, 273-287.
25. He, N., et al. A La-Related Protein Modulates 7SK snRNP Integrity to Suppress P-TEFb-Dependent Transcriptional Elongation and Tumorigenesis. *Molecular Cell.* 29, 588-599.
26. Castro, J. M., et al. (2017). Recombinant expression and purification of the RNA-binding LARP6 proteins from fish genetic model organisms. *Protein Expression and Purification.* 134, 147-153.
27. Nelson, M. D. and D. H. A. Fitch (2011). Overlap Extension PCR: An Efficient Method for Transgene Construction. Molecular Methods for Evolutionary Genetics. V. Orgogozo and M. V. Rockman. Totowa, NJ, Humana Press**:** 459-470.
28. Haan, C. and Behrmann (2007) A cost effective non-commercial ECL solution for Western blot detections yielding strong signals and low background. Jour. Of Immuno. Methods 318, 11—19.
29. Altschuler, S. E., Lewis, K. A. and Wuttke, D. S. (2013) Practical strategies for the evaluation of high-affinity protein/nucleic acid interactions. J. Nucleic Acids Investig. 4, 19—28.
30. Stefanovic, B., et al. (2019). "Discovery and evaluation of inhibitor of LARP6 as specific antifibrotic compound." Scientific Reports 9, 326
31. Eliana Peña (2018). Structural and Biochemical Characterization of Domains in the Posttranscriptional Regulator LARP6. A thesis published at Texas State University.
32. Mossessova, E. and C. D. Lima (2000). Ulp1-SUMO Crystal Structure and Genetic Analysis Reveal Conserved Interactions and a Regulatory Element Essential for Cell Growth in Yeast. Molecular Cell 5, 865-876.

33. Marblestone, J. G., et al. (2006). Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. Protein science: a publication of the Protein Society 15, 182-189.
34. Daniel, R. M., et al. (1982). "A correlation between protein thermostability and resistance to proteolysis." The Biochemical journal 207, 641-644.
35. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.