

THE STRUCTURAL AND FUNCTIONAL ROLE OF VERTEBRATE LaRP6 N-
TERMINAL REGION

by

Brianna Norbury Godinez, B.A.

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 2020

Committee Members:

Karen A. Lewis, Chair

Steven T. Whitten

Xiaoyu Xue

COPYRIGHT

by

Brianna Norbury Godinez

2020

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

Fair Use

This work is protected by the Copyright Laws of the United States (Public Law 94-553, section 107). Consistent with fair use as defined in the Copyright Laws, brief quotations from this material are allowed with proper acknowledgement. Use of this material for financial gain without the author's express written permission is not allowed.

Duplication Permission

As the copyright holder of this work I, Brianna Norbury Godinez, authorize duplication of this work, in whole or in part, for educational or scholarly purposes only.

DEDICATION

For my parents and grandma, who have endlessly supported my ambitions.

For my sister Emily, my study buddy and fellow Buff.

For Tiff, who supported me moving to Texas and starting this new chapter of my life.

For Ana and Lauren, my rocks in life.

And, for my husband Pablo, my motivation to be better. Even from over 3,000 miles away, your incessant love, support and encouragement has been a vital component of my success in this program.

ACKNOWLEDGEMENTS

My time and training in this program has been invaluable for my growth as not only a student, but as a researcher, a teacher, and most importantly, a science colleague.

I would first like to thank my P.I. and mentor, Dr. Karen A. Lewis. Your faith in my potential and willingness to mentor me even before I officially started at Texas State helped me hit the ground running into this new terrain. I would also like to thank my committee members Dr. Steven T. Whitten and Dr. Xiaoyu Xue for helpful discussions and use of equipment/resources when needed for my project.

What I did not realize before joining Texas State was the lifelong friendships I would make here. To my fellow LewisKAn grad student, Julia Roberts, you have truly become one of my dearest friends in this short amount of time. Talking about anything from science (both in lab and teaching) to the personal things has made lab feel like home. And, to Leticia Gonzalez, who has been an amazing mentor, friend, and saving grace for science gone awry.

Outside of the lab, I want to acknowledge Alejandro Oviedo, with whom I have shared many adventures in San Marcos and beyond. I also want to acknowledge Elisia (Ari) Paiz and Kathryn Banks, my vegan buddies and fellow get-together hostesses, which helped to keep our cohort sane through the craziness!

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS.....	v
LIST OF TABLES	vii
LIST OF FIGURES.....	viii
ABSTRACT	x
CHAPTER	
I. INTRODUCTION	1
II. MATERIALS & METHODS.....	19
III. RECOMBINANT EXPRESSION & STRUCTURAL ANALYSIS OF LaRP6 N-TERMINAL REGION	49
IV. ROLE OF LaRP6 NTR IN RNA BINDING ACTIVITY	76
V. CONCLUSIONS.....	84
VI. FUTURE DIRECTIONS	86
APPENDIX SECTION.....	95
REFERENCES.....	98

LIST OF TABLES

Table	Page
1. Primers Used.....	20
2. Cloning Methods Used for Each Plasmid	22
3. Plasmids Used.....	24
4. Cell Lines	25
5. SDS-PAGE Gel Percentages Used.....	27
6. Sequences of Proteins in this Work.....	28
7. Large Scale Expression Conditions.....	32
8. Calibration Standards for S75 Sephadex	33
9. Sonication Parameters	35
10. List of Sizing Columns Used and Purification Buffers.....	38
11. RNA Sequences.....	40
12. CD Preparation for His ₆ -DrLaRP6A-LaModule and His ₆ -DrLaRP6A-ΔCTD.....	47
13. Protein Extinction Coefficients for ϵ_{205} Experiments	48
14. Dilutions Prepared for ϵ_{205} Experiments	48
15. A Comparison of Protein Molecular Weights.....	58
16. BSA Data at 280 nm for ϵ_{205} Experiments	65
17. BSA Data at 205 nm for ϵ_{205} Experiments	65
18. His ₆ -DrLaRP6A-ΔCTD Data at 280 and 205 nm for ϵ_{205} Experiments	65

LIST OF FIGURES

Figure	Page
1. A classical example of the RRM.....	2
2. The general topology of each LaRP family.....	5
3. A simplified schematic of LaRP6.....	7
4. 5'-UTR of human type 1 collagen mRNA is predicted to form a stem-loop.....	9
5. NMR Structures of Human La vs. Human LaRP6 RRMs.....	10
6. LaRP1 constructs used to test the binding capabilities of truncation variants.....	12
7. Sequence alignment of <i>H. sapiens</i> , <i>X. maculatus</i> , and <i>D. rerio</i> LaRP6 proteins.....	14
8. Limited trypsinolysis of the three LaRP6 models.....	15
9. LC-MS fragments of protected LaRP6 bands from Figure 8.....	16
10. A comparison of full-length LaRP6 to the tagged constructs.....	17
11. SDM primer designs for pET28-SUMO-GG.....	23
12. Thymine insertions between the SUMO tag and <i>Drl</i> LaRP6A- Δ CTD insert.....	50
13. Cloning of pET28-SUMO- <i>Drl</i> LaRP6A- Δ CTD.....	52
14. SDS-PAGE analysis of first His ₁₀ -SUMO- <i>Drl</i> LaRP6A-NTR expression.....	53
15. Cloning of pET28-SUMO- <i>Drl</i> LaRP6A-NTR.....	54
16. Trial expressions stained with Coomassie stain and verified by anti-His probe.....	57
17. SEC chromatogram of His ₆ - <i>Drl</i> LaRP6A- Δ CTD under non-reducing conditions.....	60
18. SDS-PAGE analysis and chromatogram from His ₆ - <i>Drl</i> LaRP6A- Δ CTD purification under reducing conditions.....	62
19. Pilot purification of His ₁₀ -SUMO- <i>Drl</i> LaRP6A-NTR.....	67

20. Silver stains of <i>Drl</i> LaRP6A-NTR.....	68
21. Second purification of His ₁₀ -SUMO- <i>Drl</i> LaRP6A-NTR, leaving the tag on.....	70
22. Pilot purification of His ₁₀ -SUMO-GG.....	72
23. Circular dichroism spectra of <i>Drl</i> LaRP6A constructs and the SUMO control.....	75
24. Fluorescent electrophoretic mobility shift assays of His ₆ - <i>Drl</i> LaRP6A-La Module and His ₆ - <i>Drl</i> LaRP6A-ΔCTD versus FAM-tagged hompolymeric RNA	78
25. Fluorescent electrophoretic mobility shift assays of His ₁₀ -SUMO-GG and His ₁₀ -SUMO- <i>Drl</i> LaRP6A-NTR versus FAM-tagged hompolymeric RNA	79
26. Example of EMSA: His ₆ - <i>Drl</i> LaRP6A-ΔCTD vs. <i>Hs</i> COL1A1 RNA.....	80
27. Fluorescent electrophoretic mobility shift assays of His ₆ - <i>Drl</i> LaRP6A-La Module and His ₆ - <i>Drl</i> LaRP6A-ΔCTD versus FAM-tagged <i>Hs</i> COL1A1 stem loop mRNA	81
28. Fluorescent electrophoretic mobility shift assays of His ₁₀ -SUMO-GG and His ₁₀ -SUMO- <i>Drl</i> LaRP6A-NTR versus FAM-tagged <i>Hs</i> COL1A1 stem loop mRNA	81
29. A representative gel quantification using the volume tool settings on the Pharos FX™ Plus molecular imager.....	82
30. Quantifications of fluorescent electrophoretic mobility shift assays.....	83
31. A comparison of LaRP6-NTR sequences.....	87
32. Cloning of His ₁₀ -SUMO- <i>Hs</i> LaRP6-tNTR.....	88
33. Expression of His ₁₀ -SUMO- <i>Hs</i> LaRP6-tNTR.....	89
34. Pilot purification of His ₁₀ -SUMO- <i>Hs</i> LaRP6-tNTR.....	90
35. Second purification of His ₁₀ -SUMO- <i>Hs</i> LaRP6-tNTR	93

ABSTRACT

La-related protein 6 (LaRP6) belongs to a superfamily of evolutionarily related RNA-binding proteins. While other LaRP families have been characterized more thoroughly, LaRP6 is still poorly understood. The majority of the work to date has primarily studied the RNA binding domain, the “La Module”, in isolation. However, these studies may not reflect how the full-length protein behaves *in vivo*. Previous work in our lab using the *Danio rerio* (zebrafish) LaRP6 protein as a model system demonstrated that a domain adjacent to the La Module, called the “N-terminal region” (NTR), may play an important role in the full-length protein. To study the effect of this domain on the RNA binding activity, we created a new protein construct that contained only the NTR and the La Module while omitting the C-terminal domain, dubbed the “ Δ CTD.” The tagged NTR has also been generated, along with the isolated tag to serve as a negative control for biochemical analyses. RNA binding activity was measured using electrophoretic mobility shift assays and showed that, unexpectedly, the Δ CTD exhibited more stable binding as compared to the isolated La Module for all the ligands tested. Both the La Module and Δ CTD bind to the endogenous stem-loop ligand *Hs*COL1A1 an order of magnitude more tightly than the homopolymers. However, while the Δ CTD binds to Poly-A and Poly-U, no binding was observed to Poly-C, showing that the NTR not only modulates RNA binding activity, but also its ligand specificity. Additionally, the isolated NTR was found to bind independently to *Hs*COL1A1, albeit weakly. Circular dichroism of the Δ CTD and tagged NTR supports that the NTR is a random coil.

I. INTRODUCTION

RNA Binding Proteins

RNA binding proteins interact with many kinds of RNA, including transfer RNA (tRNA), ribosomal RNA (rRNA), and messenger RNA (mRNA). The mechanisms of RNA-protein interaction are also varied; while some RNA binding proteins (RBPs) bind to specific RNA sequences, others recognize structure rather than a particular stretch of nucleotides. Indeed, some RBPs bind to a variety of sequences *and* structures.¹

With over 1,500 different RNA binding domains (RBDs) that have been analyzed to date, RBPs use a wide variety of structural means to engage their ligand.² In eukaryotes, the most abundant RBD is the RNA recognition motif (RRM).² Originally called the “ribonucleoprotein domain” (RNP), the RRM is the most studied and therefore the most understood biochemically and structurally.³ This motif usually consists of a four-stranded antiparallel beta sheet with one to three helices on the backside, the loops between which are named by the secondary elements they connect.³

Since its discovery in the 1980s, this fold has been found to interact with RNA ligands in a multitude of ways, demonstrating its remarkable plasticity and ability to be adapted for seemingly innumerable RNA interaction mechanisms.³ Systematic analysis of various RRM-nucleic acid interactions has led to the hnRNPA1’s binding mechanism often being considered the “canonical” behavior.³ It is capable of interacting with both single-stranded DNA or RNA ligand via binding to the face of the core β -sheet, where the bases of the ligand stack with conserved aromatic residues on the RRM (Figure 1).⁴

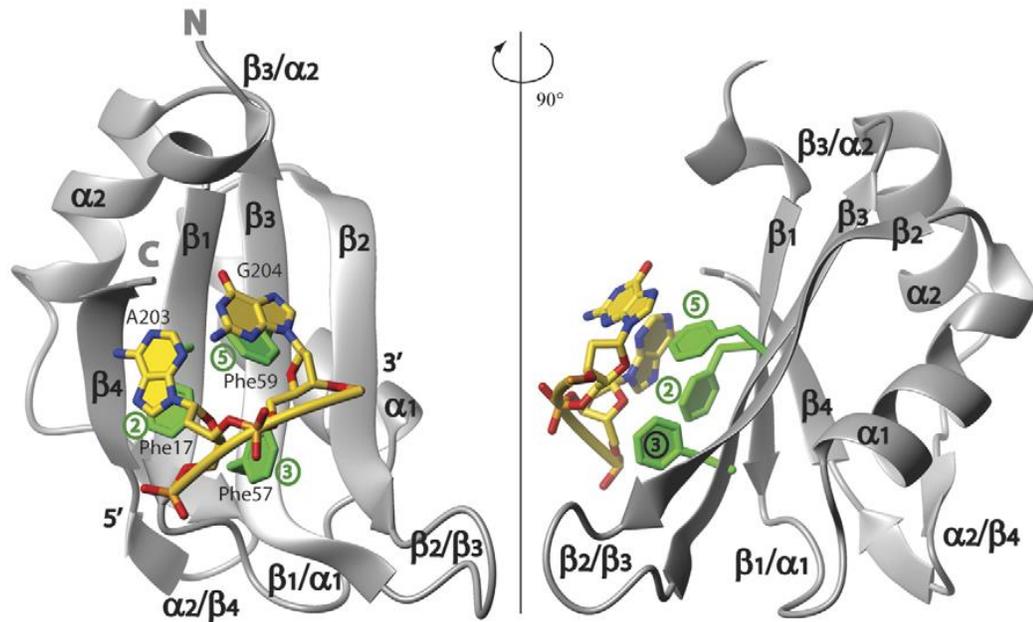


Figure 1: A classical example of the RRM. The hnRNPA1 RRM2 (gray) is shown complexed with single stranded DNA (yellow) as a model of single-stranded nucleic acids binding to the RRM. The conserved aromatic residues (Phe17, Phe57, and Phe59) are shown in green sticks. Reproduced with permission.⁴ PDB ID: 2UP1⁵

For many years, this was accepted as the sole mechanism by which RRMs engaged ligands. However, over the last ten years, a list of “variant” RRMs have been identified that elaborate from this canonical mechanism of binding. The number of “variant” RRMs now outnumber “canonical” RRMs, generating a new model of the RRM being a highly plastic domain that has been adapted via evolution to interact with a variety of RNA ligands using a variety of mechanisms. Components of the RRM:RNA interaction that can be adapted include the location and size of the binding surface on the protein, length of the ligand, and even the involvement of additional RNA binding domains (RBDs).⁴ An RBP can exert specific binding capabilities and affinities through the cooperation of multiple RBDs, even if individual domains in isolation only binds small stretches of RNA fairly weakly.¹ Because this

cooperativity confers considerable advantages both genetically and physically, it is quite common for an RNA-binding protein to contain multiple RBDs – for example, human high density lipoprotein-binding protein (HDLBP) contains fourteen repeats of the K-homology RNA-binding domain.²

Interestingly, while the RRM is the most common RBD across eukaryotes, most RBPs actually lack an identifiable RNA-binding domain.¹ What's more, many newly discovered RBPs lacking canonical RBDs have not been observed to participate in RNA biology, despite possessing binding capabilities. Many RBPs also contain low-complexity regions that are intrinsically disordered.¹ This vastly broadens the scope of how we understand this class of proteins, and how we can go about characterizing new ones.

A subclass of RBPs that bind to mRNA in particular constitute almost 45% of classified RBPs; the ribonucleoprotein complex formed by these components are called mRBPs.² These RBPs can control a variety of fates for the mRNA ligands, including translation and degradation.²

The La-related proteins (LaRPs) constitute a superfamily of RNA-binding proteins that use both an RRM and an auxiliary RBD, the La Motif, to bind a variety of RNA ligands that are involved in the regulation of translation. For example, the genuine La protein and LaRP7 bind to pre-tRNAs in the nucleus; in the cytosol, LaRP4 promotes mRNA stability whereas LaRP1 is implicated in mRNA decay via P bodies.^{6,7} Though these proteins belong to the same superfamily of proteins, they have very different mechanisms of engaging their RNA ligands, reflecting the diversity of RNA binding proteins on all levels.

Genuine La

The namesake of the La-related proteins (LaRPs), “genuine La,” was first discovered in the 1970s as an autoantigen in patients with Sjögren’s syndrome (giving La the alternative name of “SS-B”) and neonatal lupus.^{8,9} La interacts with a wide array of RNA ligands.¹⁰ Examples include nascent RNA polymerase III transcripts, pre-tRNAs, and even RNAs encoded from viruses.¹¹ The most supported role for La in the body is to protect nascent RNA transcripts from 3'-endonucleases, with varying outcomes depending on the specific transcript being chaperoned.¹² Genuine La is highly abundant in the body, being present in approximately the same amounts as a ribosomal protein.¹¹ Most of La appears to be localized to the nucleus, though it has been observed in the cytosol in certain conditions, such as during apoptosis.¹¹

Structurally, La is composed of a specific RNA binding domain that contains a La motif, which is the most homologous portion of the protein amongst different species, and an RRM (Figure 2).¹¹ In human La, a second RRM is found in the C-terminus that contributes to the RNA binding activity and localization of the protein.¹³ The function of RRM2 in mammalian La may also account for the difference in length and sequence between La C-termini of different species.¹³

La-Related Proteins (LaRPs)

Conservation of the La Motif or homology across the La Module as a whole has led to the classification of this superfamily of proteins.¹⁴ There are four families: LaRP1, LaRP4 (comprised of LaRP4A and LaRP4B, which used to be called LaRP5), LaRP6, and LaRP7 (Figure 2).^{12,14} They were numbered according to the order of discovery of the human

sequence for each family.¹² A brief overview of the LaRP families will be followed by an in-depth discussion of LaRP6.

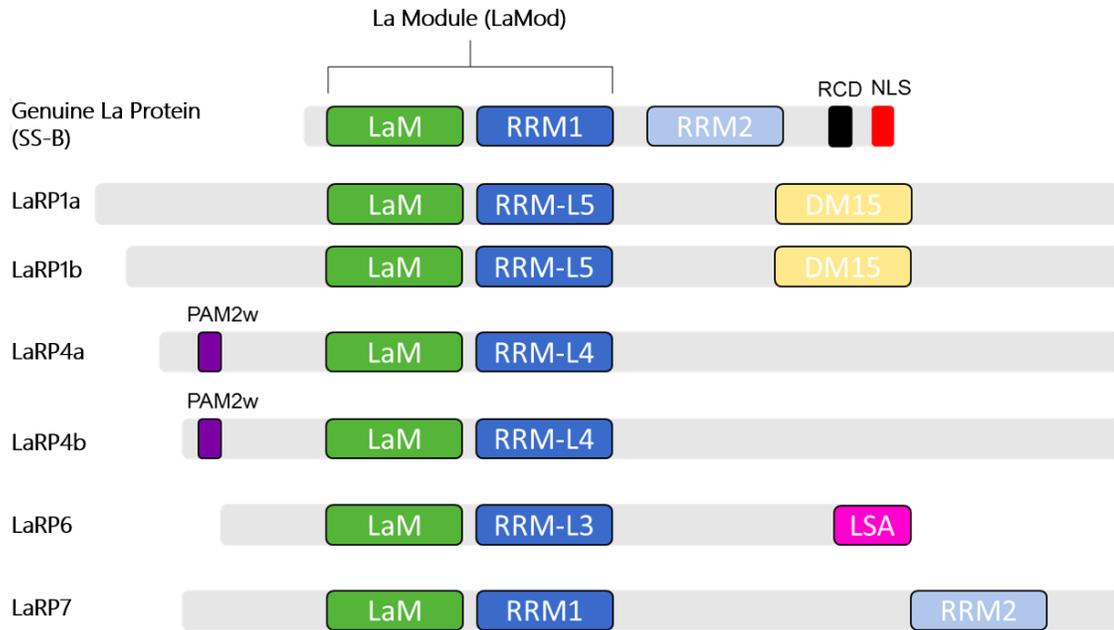


Figure 2: The general topology of each LaRP family. Adapted from Stavrika and Blagden, 2015 (reproduced with permission under Creative Commons Attribution License).¹⁴

LaRP1:

Comprised of LaRP1A and LaRP1B (the latter of which has also been called “LaRP2”), this family of LaRPs is estimated to have over 3,000 different mRNA targets. It is putatively involved in several forms of cancer, including cervix and liver.¹⁴ LaRP1 has a highly conserved C-terminal domain called DM15 that binds to the 5’ terminal oligopyrimidine (TOP) motif of ribosomal protein mRNAs.¹⁵ Recently, the LaRP1 La Module was found to bind to different homopolymeric ligands, including Poly-A, Poly-U and Poly-G.¹⁶ This study also elucidated the importance of the polypyrimidine region of

TOP-motif ligands, as deletion of this RNA structural feature greatly reduced or abolished binding altogether.¹⁶

LaRP4:

To date, LaRP4 is the most diverged LaRP from Genuine La.¹⁰ Currently, its only known endogenous ligand is single-stranded Poly-A.¹⁴ Despite being closely related in sequence, LaRP4A may be a tumor suppressor, while LaRP4B is a potential proto-oncogene.¹⁴ Unlike the other LaRP families, LaRP4 does not contain a unique domain in the C-terminus, but instead contains a variant PAM2w motif in the N-terminus, which may interact with Poly-A Binding Protein (PABP).¹⁷ A recent study found that the N-terminal region is the primary contact for RNA binding, despite it being disordered and lacking an established binding domain.¹⁸ This discovery supports a unique conformational selection model among LaRPs, in which LaRP1 interacts with the MLLE (“mademoiselle”) domain of PABP in one conformation, and a Poly-A RNA ligand in an alternate conformation.¹⁸

LaRP7:

Other names for this protein include PIP7S and HDCMA18.¹⁴ This LaRP has the highest homology with La, and for this reason has been studied the most despite its relatively late discovery.^{10,14} Also like La, this protein contains a second RRM motif in the C-terminus, although it is situated farther away from the La Module in the primary sequence.¹² This second RRM has independent RNA binding activity, and concerted RNA binding events between the La Module and RRM2 have been suggested to contribute to binding very large RNAs like the 7SK lncRNA.¹⁹

All LaRP families except for LaRP7 appear to be directly involved in translational processes.¹⁴ It is important to note that despite the structural and functional differences between these families, the La motif and RRM work together as a single functional unit (the La Module) to bind RNA ligands. Phylogenetic analysis of the La Modules of all known LaRPs supports the hypothesis that these two domains have co-evolved within the context of each LaRP family.¹²

LaRP6 (Acheron)



Figure 3: A simplified schematic of LaRP6.

The La-related protein 6 (LaRP6) was discovered in the early 2000s in a study on the skeletal muscles of the tobacco hawkmoth, *Manduca sexta*.²⁰ Because of its putative involvement in programmed cell death, it was first named Acheron after the river in Greek mythology that leads to the realm of the dead.²⁰

In multiple animal species (including humans and the tobacco hawkmoth), the LaRP6 gene was mapped to chromosome 15.²⁰ Human LaRP6 appears to have the highest mRNA expression in skeletal muscle, the heart, and the testes.²⁰ It contains putative nuclear export sequence (NES) and nuclear localization sequence (NLS) regions, potentially allowing the protein to translocate in and out of the nucleus.²⁰ However, the protein has been found to reside most often in the cytoplasm.²¹ It has several putative phosphorylation sites, in which a phosphorylation cascade within the C-terminal domain of LaRP6 has been implicated in regulatory protein:protein interactions.²² Its estimated concentration when

expressed in and purified from *E. coli* cells is similar to that expected of human La in human cells, about 100 nM.²³

LaRP6 lies in the middle of the LaRP homology spectrum, as it has greater sequence identity to genuine La than to LaRP4, but is less similar to genuine La than LaRP7.¹⁰ LaRP6 is distinguished from other families by a proline- and arginine-rich motif in the C-terminus, called the La- and S1-like-proteins-associated (LSA) motif.¹² Though the function of the LSA motif is still not known, it may be involved in protein:protein interactions.¹⁷ Though all LaRPs are far less studied as a whole compared to genuine La, a lack of a clear phenotype or close association with disease has resulted in LaRP6 has been especially overlooked until recent years.¹²

LaRP6 Regulates Collagen Expression in Vertebrates

Extensive evidence supports a role for LaRP6 in the regulation of collagen type I and type III synthesis in mammals. These collagens are part of the fibrillar collagen family, which comprise the extracellular matrix that holds cells in tissues and organs as well as the majority of scar tissue in response to injury.²⁴ This makes LaRP6 a putative target for fibroproliferative disease therapy. Collagen type I is the most abundant protein in the human body; overproduction of collagen type I results in fibrotic disorders that experts estimate is responsible for approximately 45% of mortalities in the United States alone.^{25,26} Examples of these diseases include from major organ cirrhosis, atherosclerosis, and fibrosis induced from chemotherapeutic drugs.

Currently, the only known endogenous ligand for LaRP6 is a stem-loop structure in the 5' untranslated region of type I collagen mRNA (Figure 4).²⁶ Five residues within the bulge of the stem-loop were found to be vital for binding to LaRP6.²⁷ In order for wildtype

levels of collagen type I to be synthesized in response to tissue injury, both this stem-loop and LaRP6 must be present.²⁶ This LaRP6-mediated up-regulation of collagen synthesis has also been countered with evidence that in some instances, excessive amounts of LaRP6 was found to inhibit the synthesis of this mRNA.²⁶ One potential mechanism by which LaRP6 may facilitate the initiation of translation is by tethering the mRNA to RNA helicase A, which may remodel secondary structures in the untranslated regions and/or open reading frame.²⁷ Interestingly, a very recent development in this protein's putative disease involvement identified a small molecule that may work to inhibit LaRP6's negative mRNA regulation.²⁸

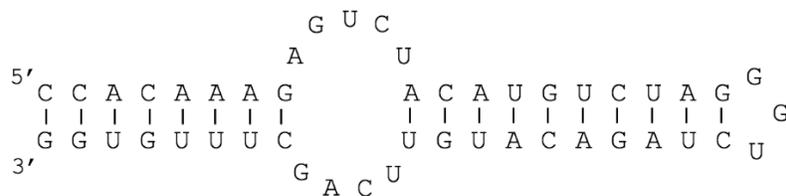


Figure 4: 5'-UTR of human type 1 collagen mRNA is predicted to form a stem-loop. This is the only endogenously known ligand *HiLaRP6* has been observed to bind.²⁶

Novel Structure, Novel Binding?

Biochemical work demonstrated that not only does LaRP6 RNA binding activity require the bipartite La Module, but the short yet flexible linker between the two domains is also an active element in RNA binding.¹⁷ Compared to genuine La, the LaRP6 RRM has unique sequence and structural elements that deviate from the “canonical” RRM, and it was therefore dubbed an RNA Recognition Motif-Like domain (RRM-L) (Figure 5).²³ In particular, the loop between $\beta 2$ and $\beta 3$ (“L3”) is significantly extended in LaRP6 and

includes a novel helix ($\alpha 1'$) that packs against the canonical β -sheet binding face of the RRM.¹⁷

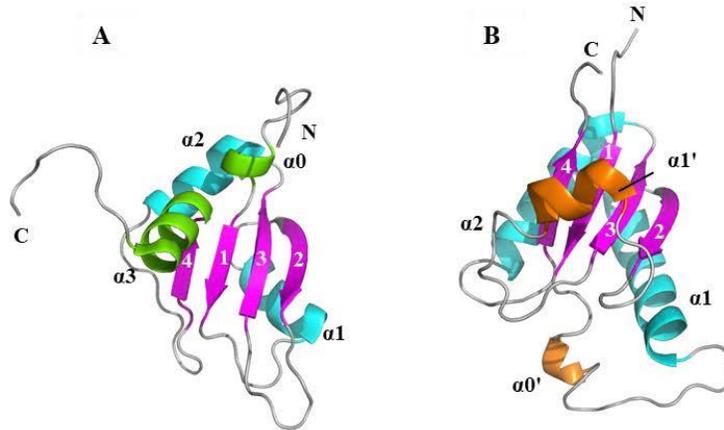


Figure 5: NMR Structures of Human La vs. Human LaRP6 RRMs.²⁹ **A:** *HsLa* RRM (PDB: 1S79) compared to **B:** *HsLaRP6* RRM (PDB: 2MTG) *HsLa* RRM has additional elements (green) not seen in *HsLaRP6*. However, *HsLaRP6* RRM has two unique helices (orange).

Despite LaRP6's observed physiological target being collagen mRNAs, human LaRP6 actually has the highest homology in the La Motif with human La regarding the residues involved in binding the ligand Poly-U.²⁶ Accordingly, LaRP6 does exhibit binding to Poly-U, as well as Poly-A and Poly-G RNA sequences, albeit more weakly than the collagen 1A1 mRNA ligand.¹⁷

The role of LaRP6 RNA binding activity in the cell remains to be fully elucidated. The strongest argument is for localization of the target collagen type I to the ER membrane for synthesis and translocation into the ER lumen.²² However, *in vitro* evidence that LaRP6 can remodel double-stranded RNAs suggests LaRP6 may also exert RNA chaperone activity that could facilitate ribosomal translocation along the mRNA.²³

Intramolecular Modulation of Nucleic Acid Binding

As mentioned above, in each LaRP family, the RRM and La Motif appear to co-evolve.¹² The combination of these two domains is thought to give each LaRP family unique specificities for their RNA ligands, as well as potential dynamicity for how each LaRP interacts with these particular ligands. Interestingly, in RNA-binding proteins, this modularity can be enhanced or affected by the evolution of multiple binding domains within a single polypeptide.³⁰

There are several benefits to modularity, particularly for RNA-binding. It increases the length of ligand that can be recognized and bound, or allows for the recognition of vastly different ligands.³⁰ Even the linker between the La motif and the RRM can play a role in ligand recognition and potentially increased avidity for binding.^{17,30}

This modularity can extend beyond the core RNA binding domain. For example, while studying the LaRP6-La Module in isolation provides a solid foundation for building an understanding of the protein's binding and structure, it is not reflective of the structure and functionality of LaRP6 in its native, full-length state *in vivo*.

Studies of other LaRP families have already demonstrated that other domains can alter or modulate the RNA binding activity of the isolated binding domain. For example, the La Module of p65 (a member of the LaRP7 family) binds to ligand in isolation.³¹ However, the C-terminal region also exhibited independent binding activity.³¹ By comparing individual domain activities to that of the full-length protein, it can be inferred that though the C-terminal domain is not required for ligand binding, it does contribute to the process in some way.³¹

A similar study was conducted with LaRP1 using full-length protein, an N-terminal construct and a C-terminal construct (Figure 6). Similar conclusions were drawn: both of the

truncated proteins were able to bind independently, but compared to the wild-type protein, neither fragment exhibited full binding activity.⁷ Additionally, it was observed that the ligand specificity for each construct was different; while the isolated C-terminal fragment appeared to bind both Poly-U and Poly-G ligands and with similar affinity, the N-terminal fragment only bound Poly-G.⁷

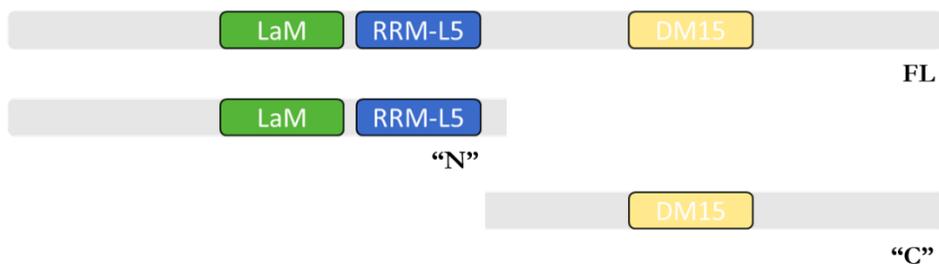


Figure 6: LaRP1 constructs used to test the binding capabilities of truncation variants.⁷ **FL:** Full-length protein. **N:** N-terminal construct. **C:** C-terminal construct. Adapted from Stavrika and Blagden, 2015 (reproduced with permission under Creative Commons Attribution License).¹⁴

In addition to the recent data on LaRP4's disordered N-terminal region driving binding to its ligand, this data further supports the potential functional importance of the N-terminal region in other LaRPs.¹⁸

Probing Domain Topology of LaRP6

In human LaRP6, it was previously thought that the minimal domain needed for sufficient RNA binding activity is the La Module.^{17,26} Though intramolecular modulation has been studied in other LaRP families, no such studies had been performed for LaRP6. Recombinant fish LaRP6 proteins were used to study the domain topology of the full-length vertebrate LaRP6. Conditions and purification parameters were optimized for three different LaRP6 species: human (*Homo sapiens*, *Hs*), zebrafish (*Danio rerio*, *Dr*), and platyfish (*Xiphophorus maculatus*, *Xm*).³² These fish in particular are ideal model organisms to use, as

one can study macromolecular structure, biochemistry and physiology all within one system.³³ A comparison of LaRP6 sequence between these and other species are in Figure 7.

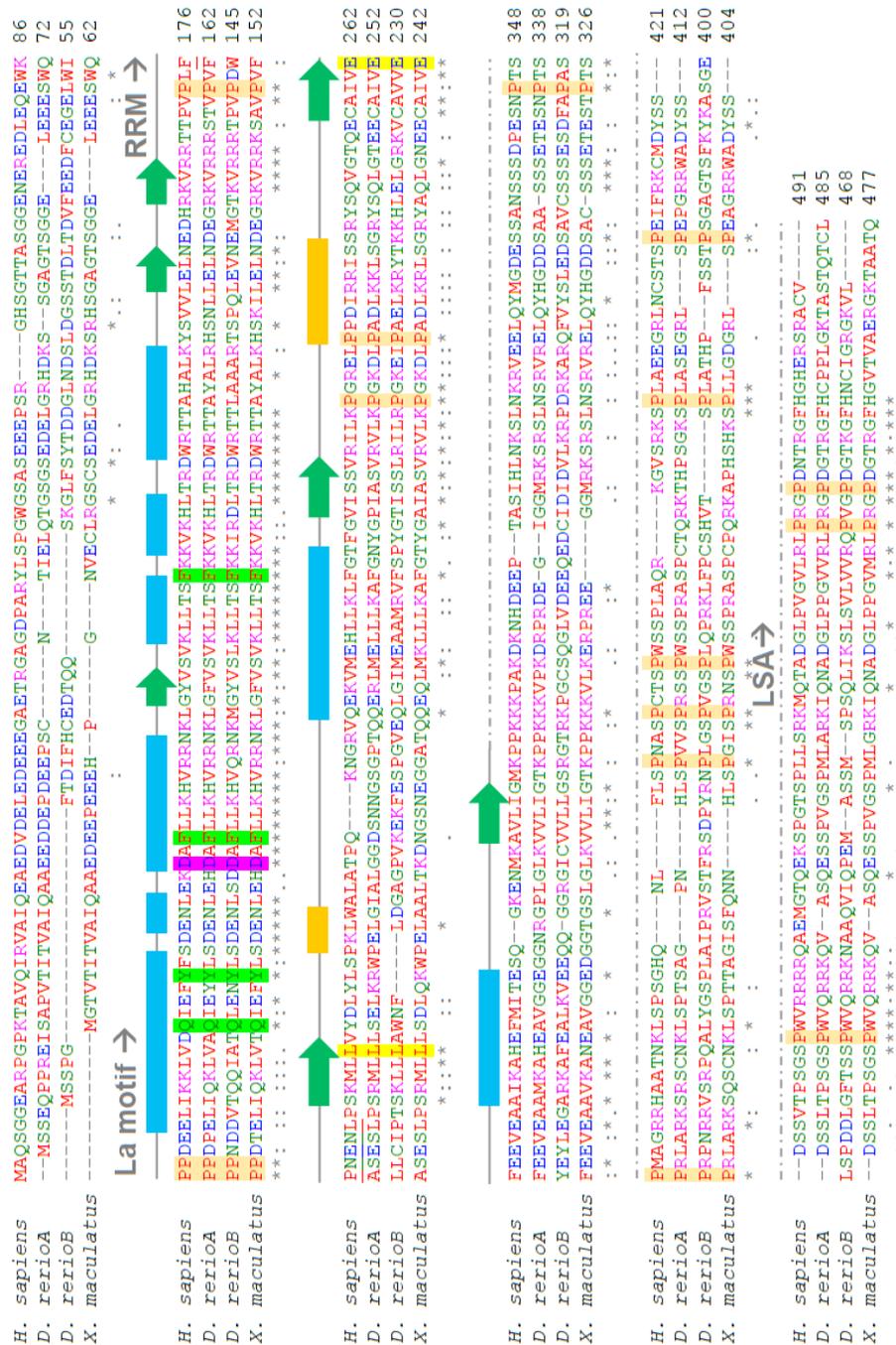


Figure 7: Sequence alignment of *H. sapiens*, *X. maculatus*, and *D. rerio* LaRP6 proteins. ³⁴ The known secondary structures of *Hs*LaRP6 La Motif and RRM are aligned to the sequences, with the underlined sequence being the linker between them. Residues in the La motif that completely inhibit RNA binding when mutated to alanine are highlighted in green. When the aspartate that is highlighted in pink is mutated to alanine, RNA binding affinity is reduced. The residues in the RRM that are highlighted in yellow increase binding affinity when mutated to alanine. All conserved prolines are highlighted in orange. Their conservation strongly suggests that they serve as structural elements for the vertebrate LaRP6 proteins.

Once protein was purified, the three different domains (N-terminus, La Module, and C-terminus) in each species were probed.³² By subjecting each full-length protein to the enzyme trypsin and tracking degradation over four hours, a protease resistant band in SDS-PAGE analysis was observed in each species (Figure 8).

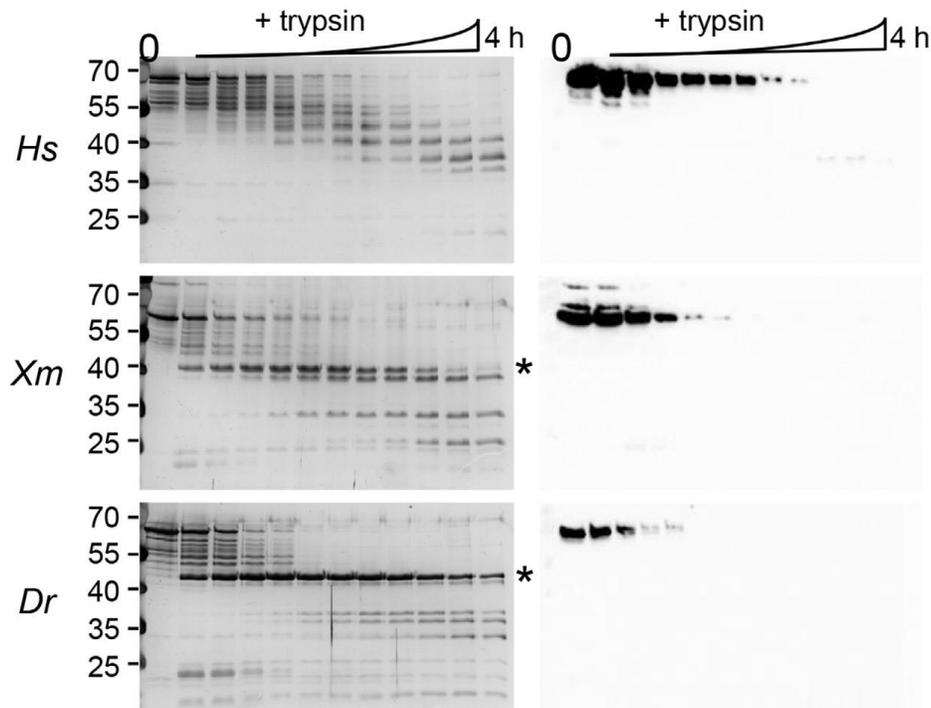


Figure 8: Limited trypsinolysis of the three LaRP6 models. Timed aliquots were run on SDS-PAGE gels and either stained with silver stain (left panels) or as a Western blot against the C-terminus of LaRP6.³² The blot tracks the loss of the C-terminus over the course of the experiment.

The starred bands from zebrafish and platyfish were fully digested and subjected to analysis by liquid chromatography coupled with mass spectrometry (LC-MS), which identified that the protease-protected bands contained not only the La Module, but also the N-terminus (Figure 9).³²

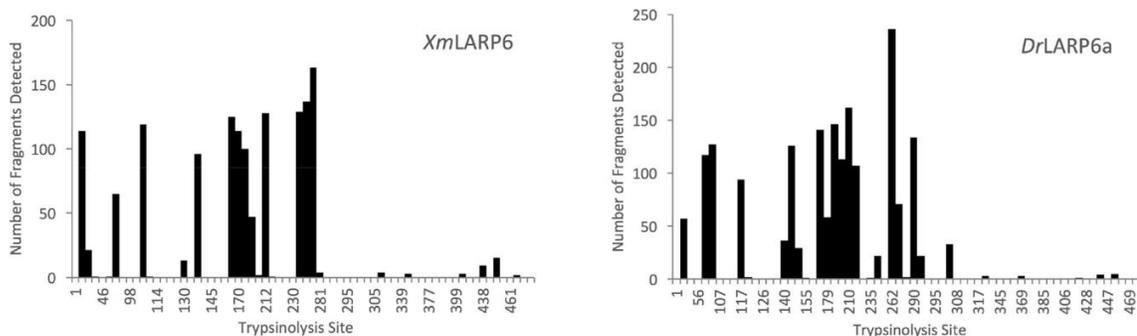


Figure 9: LC-MS fragments of protected LaRP6 bands from Figure 8. As can be observed, the fragments aligned mostly with the N-terminus and La Module of the two fish species³².

This was unexpected, as the Predictor of Natural Disordered Proteins (PONDR) analysis of LaRP6 predicted that both the N-terminus and the C-terminus would be highly disordered, and therefore they were expected to be susceptible to digestion.³² However, despite having trypsin recognition sites, the N-terminus was protected from digestion by the protease, suggesting that it either folds or interacts with the La Module.³² These data have raised questions regarding the N-terminus of LaRP6 specifically, and how it may modulate protein function and structure.

It can be noticed that the zebrafish band in the trypsinolysis experiment appears to be the most protease resistant. Because of this enhanced stability of the truncated and full-length zebrafish proteins compared to human and platyfish, this organism was chosen as the model for further structural studies in this specific project. Diving further, aligning its LaRP6 sequence with human LaRP6 reveals that the “A” paralog is more conserved than the “B” paralog.¹⁷ Thus, *DrLaRP6A* will be the model organism for the next phases of this project.

A previous K.A. Lewis lab member designed and cloned a construct called the Δ CTD in the zebrafish model, which is comprised of just the NTR and the La Module (Figure 10).³⁵ The vertebrate NTR will be characterized using this construct along with the novel NTR.

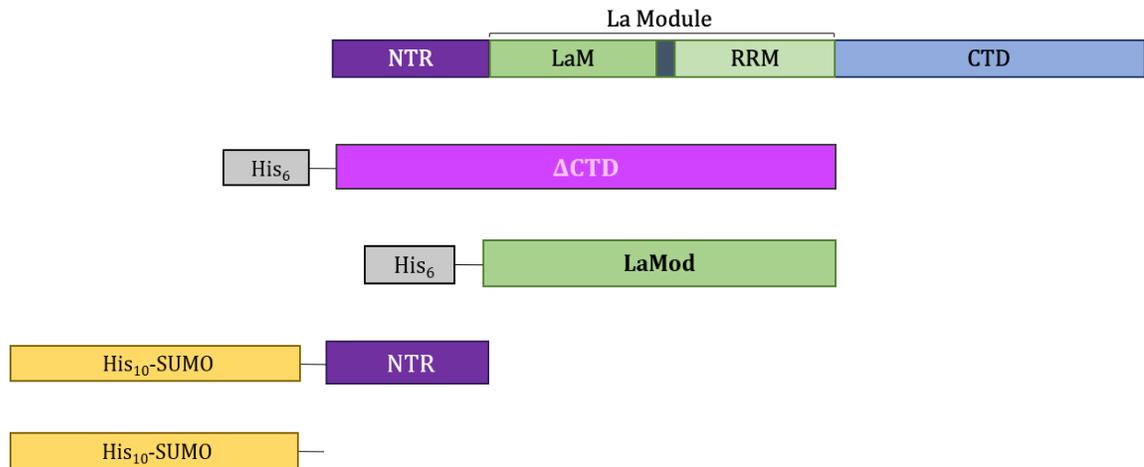


Figure 10: A comparison of full-length LaRP6 to the tagged constructs. From top to bottom: The full length protein, His₆-ΔCTD (pink), His₆-LaModule (green), His₁₀-SUMO-NTR (purple), and the isolated His₁₀-SUMO tag (yellow).

Hypotheses

In this project, two main hypotheses will be tested:

Hypothesis 1: The NTR forms a stably-folded, globular domain.

We will first clone the *DrlaRP6A*-NTR construct as N-terminal fusions to the His₁₀-SUMO tag. The final proteins will be purified using nickel-NTA affinity and size exclusion chromatography (SEC). The structure of the purified proteins will be evaluated quantitatively by circular dichroism (CD). CD data will be collected of the isolated NTR and compared to the La Module and ΔCTD.

Hypothesis 2: The NTR imparts ligand discrimination to the full-length LaRP6 protein, and exerts independent RNA binding activity.

The ability of the *DrlaRP6A*-NTR to bind RNA independently of the La Module will be evaluated using the previously studied RNA ligands *HsCOL1A1* stem-loop, Poly-A,

and Poly-U, as well as a new ligand, Poly-C, in electrophoretic mobility shift assays (EMSAs). These ligands will be tested since they have not only been tested against other LaRPs, but they are also found in nature (Poly-A in the tails of mRNAs, Poly-U in nascent pre-tRNA transcripts, and Poly-C in i-motifs, a four-stranded secondary structure known in DNA but recently discovered in RNA as well).³⁶⁻³⁸ Binding assays with the isolated La Module and Δ CTD will also be conducted against the same set of ligands to observe putative ligand discrimination.

II. MATERIALS & METHODS

Cloning and Expression Constructs

The coding sequences for *DtLaRP6A* and *DtLaRP6B* were first commercially synthesized as inserts into pcDNA3.1 (GenScript).³² The expression plasmids pET28-SUMO and pET28-ULP1, used for the SUMO fusion proteins, were kindly gifted by Dr. Christopher Lima (Rockefeller University) and Cornell University.

The primary cloning method used for this work was restriction enzyme (traditional) cloning. Particularly, directional cloning was used by designing forward primers to have the cleavage site for one restriction endonuclease, and the reverse primer having another cleavage site for a different enzyme. This allows for the insert to be ligated into the correct orientation into the vector post-digestion.

The insert of interest was first amplified from template plasmid DNA using the designed forward and reverse primers (Table 1). PCR reagents and conditions are listed in Table 19 (Appendix). Both the vector and insert were digested with the two required enzymes (*Bam*HI and *Xho*I for this work) for 1 hour at 37 °C. The vector was also treated with Antarctic Phosphatase (NEB) for 1 hour to prevent recircularization. The digested products were ligated (ratio of insert:vector was adjusted until successful ligation occurred) using T4 QuickLigase and Ligation Buffer (NEB) for 5 minutes at room temperature, before immediately transforming into DH5 α cells.

Table 1: Primers Used.

Protein/Construct	Primer Name	Sequence (5' → 3')	Description
His ₁₀ -SUMO- <i>DrLaRP6A</i> -ΔCTD, His ₁₀ -SUMO- <i>DrLaRP6A</i> -NTR	KAL 223	CATGGATCCATGAGCAGCGAGCAGCCGCCG	PCR fwd: <i>Dr6A</i> ΔCTD and <i>Dr6A</i> NTR into pET28-SUMO (<i>Bam</i> HI)
His ₁₀ -SUMO- <i>DrLaRP6A</i> -ΔCTD, His ₁₀ -SUMO- <i>DrLaRP6A</i> -LaMod	KAL 135	GCACTCGAGCTATTATTTCTTCGGTGGCTTGGTGCCGATTAGCAC	PCR rev: <i>Dr6A</i> ΔCTD and SUMO- <i>Dr6A</i> LaMod with double stop for insertion into pET28 (<i>Xba</i> I)
His ₁₀ -SUMO- <i>DrLaRP6A</i> -LaMod	KAL 222	TACGGATCCGGCACCAGCGGTGGCGAGCTG	PCR fwd: <i>Dr6A</i> LaMod into pET28-SUMO (<i>Bam</i> HI)
His ₁₀ -SUMO- <i>DrLaRP6A</i> -NTR	KAL 257	GGCACTCGAGTCATTACGCGCCACTGGATTTGTTCGTGTC	PCR rev: <i>Dr6A</i> NTR with double stop for insertion into pET28-SUMO (<i>Xba</i> I)
His ₁₀ -SUMO- <i>HsLaRP6</i> -NTR*	KAL 170	GATGGATCCATGGCCAGTCCGGCGGG	PCR fwd: <i>Hs</i> NTR with double stop for insertion into pET28-SUMO (<i>Bam</i> HI)
His ₁₀ -SUMO- <i>HsLaRP6</i> -NTR*	KAL 277	CTCGAGTATCAGATCAACTCCTCATCCGGGGGCTTC	PCR rev: <i>Hs</i> NTR with double stop for insertion into pET28-SUMO (<i>Xba</i> I)
His ₁₀ -SUMO- <i>HsLaRP6</i> -tNTR	KAL 255	CTAGCAGGATCCATGGCTCAAAGTGGTGGCGAAG	PCR fwd: <i>Hst</i> NTR clone into pET28-SUMO (<i>Bam</i> HI)
His ₁₀ -SUMO- <i>HsLaRP6</i> -tNTR	KAL 256	CGTACTCGAGCTAAGTAGATGGTGCAGTAGAAG	PCR rev: <i>Hst</i> NTR clone into pET28-SUMO (<i>Xba</i> I)

Table 1 continued.

His ₁₀ -SUMO-GG	KAL 273	ACAGATTGGTGGATGATAGTTCGAGCTCCGTC	SDM fwd, complementary: truncate after the Ser in ULP1 cleavage site to create negative SUMO control
His ₁₀ -SUMO-GG	KAL 274	GACGGAGCTCGAACTATCATCCACCAATCTGT	SDM rev, complementary: truncate after the Ser in ULP1 cleavage site to create negative SUMO control
His ₁₀ -SUMO-GG	KAL 275	TCACAGAGAACAGATTGGTGGATGATAGTTCGAGCTCC	SDM fwd, 5'-overhang: truncate Ser in ULP1 cleavage site to create negative SUMO control
His ₁₀ -SUMO-GG	KAL 276	AAGCTTGTCGACGGAGCTCGAACTATCATCCACCAAT	SDM rev, 5'-overhang: truncate Ser in ULP1 cleavage site to create negative SUMO control
His ₁₀ -SUMO- <i>Dr</i> LaRP6ANTR-Trp*	KAL 271	ATTGGTGGATCCTGGATGAGCAGCGAG	SDM fwd: Insert Trp after SUMO-tag cleavage site, before NTR sequence
His ₁₀ -SUMO- <i>Dr</i> LaRP6ANTR-Trp*	KAL 272	TAACCACCTAGGACCTACTCGTCGCTC	SDM rev: Insert Trp after SUMO-tag cleavage site, before NTR sequence

*Cloning incomplete

Site-directed mutagenesis was used as a secondary cloning method. Specially designed primers were used to create truncation mutants by inserting two consecutive stop codons into full-length template DNA. Both two-step and three-step PCR were conducted, though only the three-step protocol was successful. PCR products were then treated with *DpnI* (NEB) to digest any remaining template DNA.

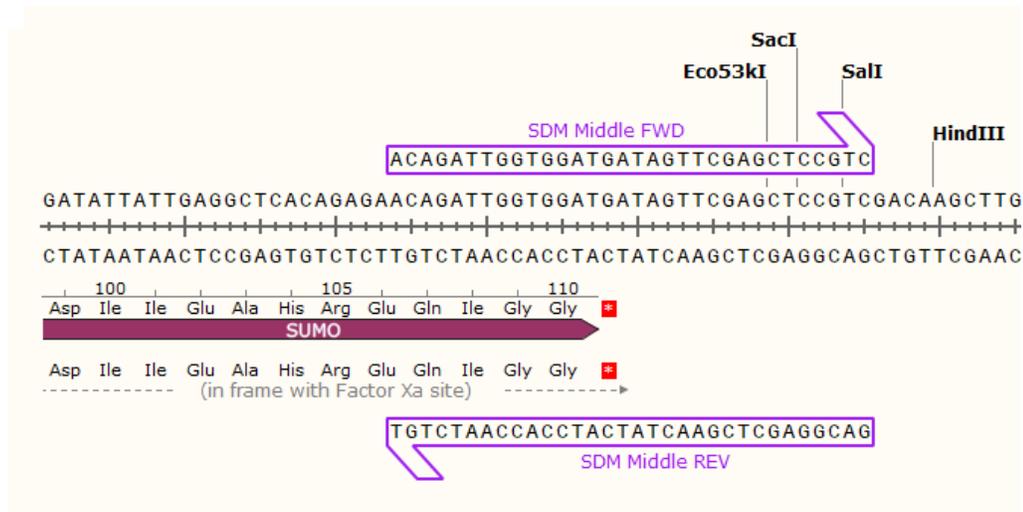
Table 2: Cloning Methods Used for Each Plasmid.

Restriction Cloning	Site-Directed Mutagenesis
pET28-SUMO- <i>Drla</i> RP6A- Δ CTD	pET28-SUMO-GG
pET28-SUMO- <i>Drla</i> RP6A-LaMod	pET28-SUMO- <i>Drla</i> RP6A-NTR-Trp*
pET28-SUMO- <i>Drla</i> RP6A-NTR	
pET28-SUMO- <i>Hsla</i> RP6-NTR (Isoform 1)*	
pET28-SUMO- <i>Hsla</i> RP6-tNTR (Isoform 2)	

*Cloning incomplete

For pET28-SUMO-GG, two different SDM primer sets were designed to test the efficacy of primer design in creating the mutation. The first set had the desired mutation centered or “sandwiched” in the middle in both primer sequences (Figure 11, panel A). The second primer set had the desired mutation off-center on each primer, leaving a 5’-overhang on both sequences (Figure 11, Panel B).

A



B

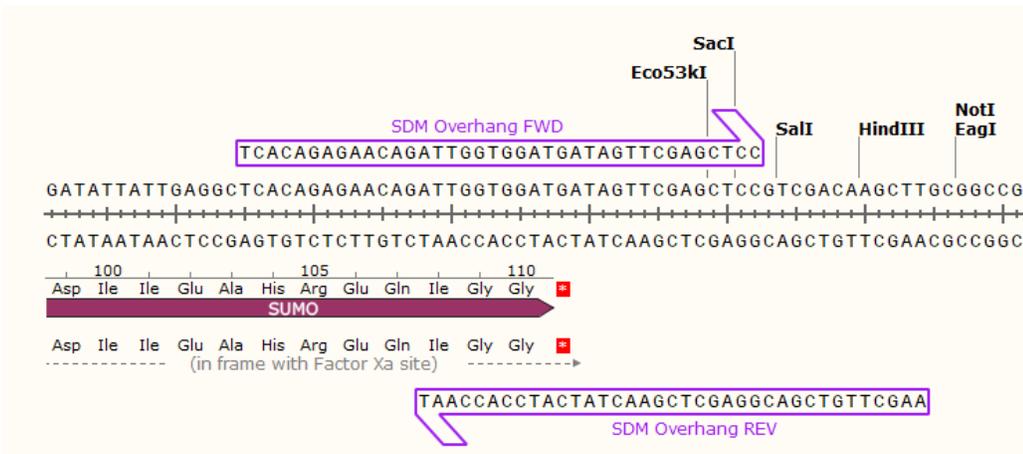


Figure 11: SDM primer designs for pET28-SUMO-GG. A: “Middle” primer set, with mutation placed in middle of primer sequences. **B:** 5'-overhang primer sequences.

Table 3: Plasmids Used.

Protein	Plasmid	Antibiotics
<i>DtLaRP6A</i>	pET28-SUMO	Kanamycin
His ₆	pET28	Kanamycin
His ₁₀ -SUMO	pET28-SUMO	Kanamycin
<i>HsLaRP6</i>	pET28-SUMO	Kanamycin

Cell Transformations

All cloning was done using *Escherichia coli* DH5 α ultracompetent cells. Miniprep DNA and 50 μ L of cells were mixed and allowed to incubate on ice for 20 minutes, heat shocked for 90 seconds at 37 °C, placed back on ice for 2 minutes, then recovered with 700 μ L of sterile Luria broth (LB [RPI]) at 37 °C for one hour, with shaking. Recovered cells were then plated onto an LB-agar plate with 1 \times kanamycin (35 μ g/mL), and incubated at 37 °C overnight (usually 18 hours). For isolating plasmid DNA, an overnight culture for each construct was prepared using sterile LB, kanamycin and a single colony from the plate, allowed to grow overnight, and a 3-mL pellet saved. DNA was then extracted using the QIAprep Spin Miniprep Kit (QIAGEN), following the manufacturer's protocol without the optional Buffer PB wash step, and eluting with 40 μ L of Buffer EB.

All protein expression was done using *Escherichia coli* RosettaTM (DE3) pLysS competent cells (EMD Millipore). Plasmid DNA was mixed with 100 μ L of cells and incubated on ice for 20 minutes. Samples were then heat shocked at 42 °C for 45 seconds, then placed back on ice for 2 minutes. Cells were recovered with the addition of 700 μ L of sterile LB and incubated at 37 °C for one hour, with shaking. Samples were plated on LB-

agar plates with 1× kanamycin and 1× chloramphenicol (34 µg/mL), and grown overnight at 37 °C.

After growing overnight, LB-agar plates were stored at 4 °C for up to 6 weeks.

Table 4: Cell Lines.

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

Agarose Electrophoresis and DNA Gel Extraction

The PCR amplification of the target fragment was verified and purified via agarose gel. Depending on the size of the DNA of interest, either 1.2% or 0.8% agarose gels were prepared, weight:volume, using agarose (Fisher Scientific) and 1× Tris-acetate-EDTA (TAE) buffer. Gels were prepared at 50 mL total volume with 1 µL/mL of ethidium bromide swirled in so as to visualize product with UV light. The 1 Kb+ Ladder (Invitrogen™) was used as standards.

After imaging, bands of interest were excised from the gel using a sterile razor and placed into 1.5 mL microcentrifuge tubes. To the tube containing the excised band, 100 µL of Binding Buffer for every 0.1 g of gel mass was added from the E.Z.N.A.® Gel Extraction Kit (Omega Bio-Tek) and the DNA recovered following the manufacturer's protocol, including the optional SPW Wash Buffer step and using 40 µL of Elution Buffer. After elution into a new 1.5 mL tube, DNA concentration was measured via absorbance at 260 nm on the Implen NanoPhotometer® N60.

Protein Expression Time Trials

An overnight culture for the protein of interest was made by inoculating 5 mL of sterile LB supplemented with 1× kanamycin and chloramphenicol with a single colony from the LB-agar plate containing the Rosetta™ cells. The tube was incubated with shaking for ~16-18 hours at 37 °C until the stationary phase was reached. A new, 100-mL flask of LB was then inoculated with 1 mL of the overnight culture, along with more antibiotics to maintain 1× concentration. Cultures were then grown further at 37 °C, shaking at 225 – 250 rpm until an OD₆₀₀ value of 0.5 – 0.8 (e.g., mid-log phase) was achieved. Cultures were placed briefly on ice (~5-10 minutes) with occasional shaking. A 1 mL aliquot of each culture pre-induction was taken and pelleted at 16,873 rcf (max speed) for 1 minute to monitor basal expression. To induce, the cultures were brought to a final concentration of 1 mM IPTG (Fisher Scientific) and placed at the desired expression temperature. Induction was allowed to proceed overnight with shaking, and aliquots were taken every few hours to track expression. All aliquots were pelleted at 16,873 rcf (max speed) for 1 minute and stored at -20 °C. Aliquots were briefly thawed and resuspended in 250 µL of 1× SDS Sample Buffer (5× buffer: 50% glycerol, 0.25 M Tris [pH 6.8], 0.7 M β-mercaptoethanol, 1% SDS, bromophenol blue). Samples were then denatured at 90 °C for 5 minutes, then immediately loaded onto duplicate (37.5:1 acrylamide to bis-acrylamide, ProtoGel, National Diagnostics) 0.75 mm SDS-PAGE gels (Table 5). Electrophoresis was run in 1× Tris-Glycine buffer (50 mM Tris and 0.5 M glycine, 0.4 M SDS) at 200 V for approximately 1 hour, or until the dye front reached the bottom of the gel. One gel was analyzed using Coomassie blue staining, and its sister gel was subjected to anti-His western blot.

Table 5: SDS-PAGE Gel Percentages Used.

Gel Percentage	Proteins
10%	His ₆ -DrLaRP6A-LaModule, His ₆ -DrLaRP6A-ΔCTD, His ₁₀ -SUMO-DrLaRP6A-ΔCTD, His ₁₀ -SUMO-DrLaRP6A-NTR
13%	DrLaRP6A-NTR, His ₁₀ -SUMO-GG, His ₁₀ -SUMO- HsLaRP6-tNTR
15%	DrLaRP6A-NTR
4-20% gradient (pre-cast)	DrLaRP6A-NTR

Table 6: Sequences of Proteins in this Work.

His₁₀-SUMO-DrLaRP6A-LaMod:

MGHHHHHHHHHHSSGHIEGRHMASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRF
LYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGSGTSGGELEEEESWQPPDPELIQKLVAQIEYYLSDENLEHDAFLLKHVRRNKLGFVSV
KLLTSFKKVKHLTRDWRTTAYALRHSNLELNDEGRKVRRRSTVPVFASESLPSRMLLSELKRWPELGIALGGDSNNGSGPTQQRMLMEL
LLKAFGNYGPIASVRVLKPGKDLPADLKKLSGRYSQLGTEECAIVEFEEVEAAMKAHEAVGGEGGNRGPLGLKVVVIGTKPPKK

His₁₀-SUMO-DrLaRP6A-ΔCTD:

MGHHHHHHHHHHSSGHIEGRHMASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRF
LYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGSMSEQPPREISAPVTTTVAIQAAEEDDEPDEEPCNTIELQTGSGSEDELGRHDKSSG
AGTSGGELEEEESWQPPDPELIQKLVAQIEYYLSDENLEHDAFLLKHVRRNKLGFVSVKLLTSFKKVKHLTRDWRTTAYALRHSNLELNDE
EGRKVRRRSTVPVFASESLPSRMLLSELKRWPELGIALGGDSNNGSGPTQQRMLMELLLKAFGNYGPIASVRVLKPGKDLPADLKKLSGR
YSQLGTEECAIVEFEEVEAAMKAHEAVGGEGGNRGPLGLKVVVIGTKPPKK

His₆-DrLaRP6A-ΔCTD[†]:

MGSSHHHHHHSSGLVPRGSHMMSSEQPPREISAPVTTTVAIQAAEEDDEPDEEPCNTIELQTGSGSEDELGRHDKSSGAGTSGGELEEEES
WQPPDPELIQKLVAQIEYYLSDENLEHDAFLLKHVRRNKLGFVSVKLLTSFKKVKHLTRDWRTTAYALRHSNLELNDEGRKVRRRSTVP
VFASESLPSRMLLSELKRWPELGIALGGDSNNGSGPTQQRMLMELLLKAFGNYGPIASVRVLKPGKDLPADLKKLSGRYSQLGTEECAIV
EFEEVEAAMKAHEAVGGEGGNRGPLGLKVVVIGTKPPKK

His₆-DrLaRP6A-LaMod[†]:

MGSSHHHHHHSSGLVPRGSHMGTSGGELEEEESWQPPDPELIQKLVAQIEYYLSDENLEHDAFLLKHVRRNKLGFVSVKLLTSFKKVKHLT
RDWRTTAYALRHSNLELNDEGRKVRRRSTVPVFASESLPSRMLLSELKRWPELGIALGGDSNNGSGPTQQRMLMELLLKAFGNYGPIAS
VRVLKPGKDLPADLKKLSGRYSQLGTEECAIVEFEEVEAAMKAHEAVGGEGGNRGPLGLKVVVIGTKPPKK

His₁₀-SUMO-DrLaRP6A-NTR:

MGHHHHHHHHHHSSGHIEGRHMASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRF
LYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGSMSEQPPREISAPVTTTVAIQAAEEDDEPDEEPCNTIELQTGSGSEDELGRHDKSSG

A

His₁₀-SUMO-*HsLaRP6*-tNTR (Isoform 2):

MGHHHHHHHHHHSSGHIEGRHMASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFQKQKEMDSLRF
LYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGSMAQSGGEARPGPKTAVQIRVAIQEAEDVDELEDEEEGAETRGAGDPARYLSPGWG
SASEEPPSRGHRNRSSVNSRTMLASFIVSSAPSTAPST

His₁₀-SUMO-GG:

MGHHHHHHHHHHSSGHIEGRHMASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFQKQKEMDSLRF
LYDGIRIQADQTPEDLDMEDNDIIEAHREQIGG

His₁₀-SUMO-*HsLaRP6*-NTR (Isoform 1)*:

MGHHHHHHHHHHSSGHIEGRHMASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFQKQKEMDSLRF
LYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGSMAQSGGEARPGPKTAVQIRVAIQEAEDVDELEDEEEGAETRGAGDPARYLSPGWG
SASEEPPSRGHS GTTASGENEREDLEQEWKPPDEELI

His₁₀-SUMO-*DrLaRP6A*-NTR-Trp*:

MGHHHHHHHHHHSSGHIEGRHMASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFQKQKEMDSLRF
LYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGSWMSSEQPPREISAPVTTTVAIQAAEEDDEPDEEPSCNTIELQTGSGSEDELGRHDKSS
GA

Bovine Serum Albumin (BSA)#:

MKWVTFISLLLLFSSAYSRGVFRRDTHKSEIAHRFKDLGEEHFKGLVLIAFSQYLQQCPFDEHVKLVNELTEFAKTCVADESHAGCEKSLH
TLFGDELCKVASLRETYGDMADCCCKQEPERNECFLSHKDDSPDLPKLPDPNTLCDEFKADEKFWGKYLEIARRHPYFYAPELLYY
ANKYNGVFQECCQAEDKGACLLPKIETMREKVLASSARQLRCASIQKFGERALKAWSVARLSQKFPKAEFVEVTCLVTDLTKVHKECC
HGDILLECADDRADLAKYICDNQDTISSKLKECCDKPLLEKSHCIAEVEKDAIPENLPPLTADFAEDKDVCKNYQEAKDAFLGSFLYEYSR
RHPEYAVSVLLRLAKEYEATLEECCAADDPHACYSTVFDKLLKHLVDEPQNLIKQNCQFEKLGEGYGFQNALIVRYTRKVPQVSTPTLVEV
SRLGKVGTRCCTKPESERMPCTEDYLSLILNRLCVLHEKTPVSEKVTCCCTESLVNRRPCFSALTTPDETYVPKAFDEKLFTFHADICTLPDT
EKQIKKQTALVELLKHKPKATEEQKLTVMENFVAFVDKCCAADDKEACFAVEGPKLVVSTQTALA

† Expressed by previous lab members; purified proteins used for biochemical analyses

*Constructs have not been expressed yet; cloning still in progress

Used in ϵ_{205} studies³⁹

Coomassie Blue Staining

Total protein content was observed using Coomassie Brilliant Blue staining (0.05% (w/v) Coomassie Brilliant blue, 40% (v/v) methanol and 10% (v/v) glacial acetic acid) for at least 20 minutes with shaking at room temperature. Excess stain was then removed from the gel using destaining solution (40% (v/v) methanol, 10% (v/v) glacial acetic acid) for again at least 20 minutes with shaking, along with Kimwipes to absorb excess dye molecules. The destained gels were then imaged using the Coomassie stain setting on the ChemiDoc™ XRS+ molecular imager.

Anti-His Western Blot

This technique was used to confirm the presence of the His-tagged protein of interest in expression samples, as well as determine expression efficiency. Proteins were transferred from the SDS-PAGE gel to a nitrocellulose membrane using 1× Transfer Buffer and the Bio-Rad Trans-Blot® Turbo™ transfer system using the Mixed Molecular Weights program (1.3 A, 25 V, 7 minutes). A 1× solution of Tris-buffered saline (TBS) was (10×: 0.2 M Tris-HCl [pH 7.4 at 4 °C], 1.5 M NaCl). After transferring, the membrane was incubated in blocking solution (5% bovine serum albumin [BSA] in 1× TBS containing 0.05% Tween 20 ["TBS-T"]) for at least 1 hour at room temperature with shaking. The HisProbe™-HRP Conjugate (Thermo Scientific™, lot number 15165) was then added straight to the blocking solution to a final concentration of 1:5000, and the membrane was again incubated for at least 1 hour. The membrane was then rinsed with 1× TBS-T and washed for 10 minutes twice with fresh reagent each time. Two final washes of 1× TBS were done. The membrane was then incubated with 20 mL of homemade enhanced chemiluminescence reagent (a.k.a.

“The Juice”; 100 mM Tris-HCl [pH 8.8 at 4 °C], 1.25 mM luminol in DMSO, 2 mM 4-Iodophenylboronic acid [IPBA] in DMSO) and 12 μ L of 30% H₂O₂⁴⁰. The corner of the membrane was briefly blotted onto paper towels to remove excess liquid, wrapped in clingwrap, and imaged using the Chemi function from the blot settings on the ChemiDoc™ XRS+ molecular imager.

Silver staining

This staining was performed in lieu of Coomassie staining when increased sensitivity of total protein content was required. In a glass container, 0.75 mm SDS-PAGE gels (see Table 5 for percentages) were fixed with 50% ethanol for 15 – 30 minutes at room temperature with light shaking. The fixing solution was aspirated and staining solution added (7.56% NaOH, 1.5% (v/v) NH₄OH, 4.7 M Ag(NO₃)₂) in which the gels incubated for 15 – 30 minutes at room temperature with light shaking. The fixing solution was carefully aspirated, and the gels washed three times with enough ultrapure polished H₂O to submerge the gels. After water was removed, the developing solution (2.5% citric acid, 37% formaldehyde) was added and mixed by hand until the desired band intensity was reached. The same volume of kill solution (45% v/v methanol, 2% v/v acetic acid) was immediately added to cease further developing and incubated for 1 hour. Gels were imaged using the silver stain setting on the ChemiDoc™ XRS+.

Large-Scale Protein Expression

Based on the small-volume expression trials, each protein was expressed in large culture volume (1 L), according to Table 7.

Table 7: Large Scale Expression Conditions.

Protein	Temperature (°C)	Time (Hours)
His ₁₀ -SUMO-DrLaRP6A-LaModule	18	8
His ₁₀ -SUMO-DrLaRP6A-ΔCTD	18	10
His ₁₀ -SUMO-DrLaRP6A-NTR	18	8
His ₁₀ -SUMO-HsLaRP6-tNTR (Isoform 2)	18	10
His ₁₀ -SUMO-GG	18	7

E. coli Rosetta™ (DE3) cells were used for all large-scale expressions. From a 50 mL overnight culture, 10 mL was used to inoculate 1 L of Miller's LB Broth (RPI) containing 50 µg/mL kanamycin and 34 µg/mL chloramphenicol. Cells were grown to an OD₆₀₀ value of 0.5 – 0.8 at 37 °C with shaking before placing on ice for 5 – 10 minutes, with occasional shaking. A 1 mL, pre-induction aliquot for basal expression was taken. The cells were brought to a final concentration of 1 mM IPTG and expressed at the desired temperature and time, as listed above. Two more 1 mL aliquots, one during the expression and one right before harvesting cells, were taken to verify expression. Cells were collected by centrifugation at 5,000 rcf at 4 °C for 10 minutes. Cell pellets were scooped into clean 50-mL conical vials and stored at -20 °C until use.

Cleaning and Calibration of Sephadex S75 Column

The sizing column used in this work is regularly cleaned and calibrated as part of regular laboratory protocols to maintain cleanliness and condition of the media, as well as ensure accurate molecular weight determination via A₂₈₀. This regularly-scheduled cleaning occurred partway through this thesis project.

A single 500 mM NaOH wash was sandwiched between ultrapure polished water washes, after which the column was equilibrated into running buffer consisting of 50 mM NaH₂PO₄/Na₂HPO₄ (pH 7.2) and 150 mM NaCl. Three independent mixtures of

calibration-grade protein standards were separated on the cleaned column. The column was then equilibrated extensively with water and finally placed into 20% ethanol for storage.

Due to the timing of this cleaning, the chromatograms presented in this thesis may have two slightly different standard annotations. The standards below will be present on every chromatogram at the elution volumes that were appropriate for the chromatogram being shown (i.e., whether the specified protein preparation was performed prior to cleaning or after cleaning).

Table 8: Calibration Standards for S75 Sephadex.

Standard	Molecular Weight (kDa)	Previous V_e (mL)	New V_e (mL)
Blue Dextran (V_0)	2,000	46.25	47.6
Conalbumin	75	53.75	55.36
Ovalbumin	44	59.37	60.97
Carbonic Anhydrase	29	67.16	69.24
RNAse A	13.7	80.305	81.725
Aprotinin	6.5	93.265	92.245

A standard curve was fit to the data to determine the relationship between elution volume and protein molecular weight. First, the K_{av} (representing the “available value” of the distribution coefficient K_D) was determined for each protein:

$$K_{av} = \frac{V_e - V_0}{V_c - V_0} \quad (\text{eq. 1})$$

where V_e is the elution volume of the protein of interest, V_0 is the void volume as identified by Blue Dextran, and V_c is the column volume (120 mL). K_{av} was plotted against the \log_{10} of

the known molecular weights of each protein in Da, and a linear regression line was fitted to the data, yielding the equation:

Previous Calibration	Current Calibration
$y = -0.519x + 2.6075$	$y = -0.4962x + 2.5122$
$R^2 = 0.9929$	$R^2 = 0.9948$

These equations were used to calculate the apparent molecular weight (MW_{app}) of all purified proteins in this work, as appropriate.

Protein Purification

One protease inhibitor tablet (Pierce, PIA32965) was dissolved in 50 mL of cold Lysis/Wash Buffer 1 (buffers listed in Table 10). A 1 L cell pellet was briefly thawed on ice, then resuspended in 30 mL of the cold buffer-tablet solution. The cell slurry was then sonicated in a 50 mL conical in an ice-water bath using a probe sonicator fitted with a 1/4" microtip (Fisher Scientific). The optimal sonication parameters for each protein are listed below. Cell lysate was collected via centrifugation at 18,000 rcf at 4 °C for 15 minutes. A sample of the cell debris pellet was taken as a control for SDS-PAGE analysis to ensure adequate lysis.

From there, the general purification scheme is as follows; each step is described in detail below.

1. Immobilized metal affinity chromatography (IMAC). His-tagged constructs bind to the nickel resin.
2. Size exclusion chromatography (SEC), for buffer exchange and evaluation of monodispersity and apparent MW.

- a. If retaining the His₁₀-SUMO tag attached to protein of interest: pool, concentrate, and store.
- b. If cleaving tag, proceed to step 3.
3. Cleave His₁₀-SUMO tag with His₁₀-ULP1.
4. Second IMAC, in which the cleaved tag binds to nickel resin, and tag-free protein elutes in the flowthrough.
5. Second SEC, for buffer exchange and evaluation of monodispersity and MW.
6. Pool, concentrate, and store.

For SDS-PAGE analysis, samples were processed as described above, except prepared with 5× SDS Sample Buffer. Note: for the trial purification of the cleaved *DrlaRP6A-NTR*, bromophenol blue was left out of the sample buffer, as it was suspected the protein would run in line with the dye front.

Table 9: Sonication Parameters.

Protein	Amplitude (%)	Time on and Off (sec)	Number of Cycles	Total Sonication Time (min)
His ₆ - <i>DrlaRP6ΔCTD</i>	20	20 on, 30 off	6	2
His ₁₀ -SUMO- <i>DrlaRP6ANTR</i>	37	20 on, 30 off	4.5	1.5
His ₁₀ -SUMO- <i>HslaRP6tNTR</i> (Isoform 2)	38	20 on, 30 off	7.5	2.5
His ₁₀ -SUMO-GG	35	20 on, 30 off	4.5	1.5

Immobilized Metal Affinity Chromatography (IMAC)

For each 1 L cell pellet, approximately 4 mL of a 50% Ni²⁺-agarose bead slurry (HisPur™ Ni-NTA resin, 2 mL packed [Thermo Scientific™]) were washed three times with 10 mL of ultrapure polished H₂O and equilibrated with 10 mL of the above-mentioned Lysis/Wash 1 Buffer-protease inhibitor tablet solution. The equilibrated beads were added to the cleared cell lysate, and incubated at 4 °C for 1 hour with light shaking. The bead-lysate mixture was then transferred to a Kimble Flex-Column glass chromatography column and briefly allowed to settle. The flowthrough was collected, followed by a 40 mL wash of cold Lysis/Wash 1 buffer without protease inhibitor. To elute weakly bound proteins, 24 mL of Wash 2 buffer was added, and three 8 mL fractions collected. Tightly bound proteins (including the protein of interest) were eluted using 24 mL of Elution Buffer, with six 4 mL fractions collected. A sample of the beads was also taken for SDS-PAGE analysis to ensure protein of interest eluted adequately.

For purifications involving tag cleavage, the protein of interest should be in the flowthrough for the second nickel column, as the cleaved tag will stick to the nickel beads, but the cleaved protein should not.

Size exclusion chromatography

A Sephadex size exclusion column (S75 or S200; see Table 10) was equilibrated with cold, filtered (0.2 µm cellulose acetate membrane; Sartorius) and degassed SEC Buffer. Chosen fractions from the Ni²⁺ affinity column containing the protein of interest (oftentimes all of the elution fractions) were pooled, concentrated to a minimum of ~2.5 mL using a pre-rinsed Vivaspin™ centrifugal concentrator (Sartorius) by centrifugation at 4,000 rcf at 4 °C for 10-20 minute cycles. The concentrated protein was filtered through a 0.2 µm syringe

filter before loading at least 2 mL into the ÄKTA Pure Fast Protein Liquid Chromatography (FPLC) system (GE Healthcare Life Sciences). Oftentimes, more than 1 injection was done as the concentration step had to be halted before reaching the 2.5 mL target volume, either because of aggregation or solution viscosity due to high protein concentrations. The column flow rate was set to 1 mL/min, with a total elution of 180 mL, in 2 mL fractions. Fractions of interest were chosen based the A_{280} absorbance peaks in the chromatogram. Samples from the chosen fractions were analyzed via SDS-PAGE analysis and Coomassie staining to identify presence and purity of the protein of interest. Desired fractions were pooled, concentrated as before until desired volume/concentration was reached, and aliquoted into 40 or 50 μ L aliquots. For the pilot purification for each protein, four different storage conditions were tested: (1) No glycerol + snap-freezing in N_{20} , stored at -70 °C, (2) 2% glycerol + snap-freezing in N_{20} , stored at -70 °C, (3) No glycerol + no snap-freezing, stored at 4 °C, (4) 2% glycerol + no snap-freezing, stored at 4 °C. Once identified, the best storage condition was used for future purifications of each respective protein.

Table 10: List of Sizing Columns Used and Purification Buffers.

Protein	Sizing Column Used	Lysis/Wash 1 Buffer	Wash 2 Buffer	Elution Buffer	SEC Buffer
His ₆ -DrLaRP6A- Δ CTD	S200	50 mM sodium phosphate, pH 8.0; 200 mM NaCl; 300 mM NaI; 500 mM glucose; 10 mM imidazole; 2 mM β ME	50 mM sodium phosphate, pH 8.0; 200 mM NaCl; 300 mM NaI; 500 mM glucose; 30 mM imidazole; 2 mM β ME	50 mM sodium phosphate, pH 8.0; 200 mM NaCl; 300 mM NaI; 500 mM glucose; 300 mM imidazole; 2 mM β ME	50 mM sodium phosphate, pH 8.0; 200 mM NaCl
His ₁₀ -SUMO-DrLatP6A-NTR	S75	50 mM sodium phosphate, pH 8.0; 200 mM NaCl; 10 mM imidazole; 1 mM TCEP	50 mM sodium phosphate, pH 8.0; 200 mM NaCl; 30 mM imidazole; 1 mM TCEP	50 mM sodium phosphate, pH 8.0; 200 mM NaCl; 300 mM imidazole; 1 mM TCEP	50 mM sodium phosphate, pH 8.0; 200 mM NaCl; 1 mM TCEP
His ₁₀ -SUMO-GG, His ₁₀ -SUMO-HsLaRP6-tNTR (Isoform 2)	S75	50 mM sodium phosphate, pH 8.0; 200 mM NaCl; 10 mM imidazole; 2 mM β ME	50 mM sodium phosphate, pH 8.0; 200 mM NaCl; 30 mM imidazole; 2 mM β ME	50 mM sodium phosphate, pH 8.0; 200 mM NaCl; 300 mM imidazole; 2 mM β ME	50 mM sodium phosphate, pH 8.0; 200 mM NaCl; 1 mM TCEP

ULP1 Cleavage of SUMO tag

For biochemical analyses, the His₁₀-SUMO tag was left on the *DLaRP6A*-NTR protein construct. However, several attempts were made to produce a tag-free protein. After the initial SEC purification, fractions of interest as determined by SDS-PAGE analysis and Coomassie staining were pooled and incubated with 1:100 ULP1: His₁₀-SUMO-*DLaRP6A*-NTR for 2 hours at 16 °C. A second IMAC column and a second sizing column were then carried out with the digested protein, as mentioned in the general purification scheme. Elution fractions for each column were analyzed as described above.

Electrophoretic Mobility shift Assays (EMSAs)

Electrophoretic mobility shift assays (EMSAs) were used to test the RNA binding activity of the purified proteins. At the beginning of this project, the established methodology in the group was to use biotinylated single-stranded RNA as the ligand, detecting RNA using a commercial chemiluminescence detection kits. However, as described below, the bulk of this work used an improved detection method that relied on 5'-end labeled 6-FAM (Fluorescein)-labeled ligands.

Table 11: RNA Sequences.

Ligand	Sequence
3' Biotinylated A ₂₀ (PolyA)	5'- AAAAAAAAAAAAAAAAAAAAAAAAAA - Biotin - 3'
3' Biotinylated HsCOL1A1 stemloop	5' - CCACAAAGAGUCUACAUGUCUAGGGUCUAGACAUGUUCAGCUUUGUGG - Biotin - 3'
FAM-A ₂₁ (PolyA)	5'- 6-FAM-AAAAAAAAAAAAAAAAAAAAAAAAA - 3'
FAM-U ₂₀ (PolyU)	5'- 6-FAM-UUUUUUUUUUUUUUUUUUUUUU - 3'
FAM-UC ₁₉ (PolyC)	5'- 6-FAM-UCCCCCCCCCCCCCCCCCCCCC - 3'
FAM-HsCOL1A1 stemloop	5'- 6-FAM-CCACAAAGAGUCUACAUGUCUAGGGUCUAGACAUGUUCAGCUUUGUGG - 3'

RNA Preparation

Biotinylated RNA ligands (Dharmacon) were prepared by previous lab members using Pierce™ 3' End Biotinylation Kit, and diluted to 2.5 μ M prior to aliquoting.²⁹

The fluorescently tagged RNAs were commercially synthesized (IDT) and received as lyophilized pellets. Pellets were resuspended into sterile-filtered 0.5 M Tris-EDTA (TE), pH 7.25 to IDT's recommended stock concentration of 100 μ M.

Biotinylated RNA

A 10 \times Binding Buffer stock (100 mM Tris-HCl, pH 7.4 at 4° C; 200 mM KCl; 10 mM MgCl₂; 1000 mM NaCl) was prepared and stored at -20° C. As needed, this stock was diluted to 1 \times with 15% glycerol and 1 mM β -mercaptoethanol, in nuclease-free water (IDT).

Protein aliquots were thawed on ice. Serial protein dilutions were prepared in 1.5 mL, autoclave-sterile microcentrifuge tubes (USA Scientific). The protein concentrations were calculated to be 2X the final concentration to be incubated with the RNA.

Biotinylated RNA was thawed, aliquoted as necessary, and heated at 80° C for 2 minutes, followed by at least 10 minutes in an ice bath. RNA was diluted to 2 \times the final concentration for the binding reaction with the protein, along with a final concentration of 0.028 U/ μ L of Ribolock RNase Inhibitor (Thermo Scientific™, EO0381).

The binding reactions were prepared by mixing a 1:1 mixture of the 2 \times protein and 2 \times RNA solutions in 1.5 mL, autoclave-sterile microcentrifuge tubes (USA Scientific).

Reactions were incubated on ice for 1 hour to reach equilibrium, after which 20 μ L of each reaction was loaded onto 1.5 mm, 6.5% polyacrylamide (29:1, acrylamide:bisacrylamide; Accugel) Tris-borate-EDTA (TBE) native gels with 5% glycerol.⁴¹ Electrophoresis was run

in cold 1× TBE without glycerol at 200 V for 20 minutes with -70° C icepacks to absorb excess heat. A 50% solution of sucrose and bromophenol blue was used to track electrophoresis.

The separated binding products were transferred to a Hybond N+ membrane (GE Biosciences) using the Bio Rad Trans-Blot Turbo (25 V, 1.0 A, 30 minutes) in 1× TBE without glycerol, then crosslinked with ultraviolet light for 45 seconds at 120 mJ/cm². The membrane was either allowed to dry overnight in the dark, or immediately detected using the LightShift™ RNA EMSA Chemiluminescence Nucleic Acid Detection Module (Thermo Scientific). Two membranes were rehydrated in the same container with 16 mL of Nucleic Acid Detection Blocking Buffer for 15 minutes with gentle shaking, and then processed following the manufacturer's recommended protocol. After the final incubation, the corner of each membrane was blotted on a paper towel until the membrane was dry. The membrane was then faced RNA face side up on a clean sheet of plastic wrap and covered with 5 mL of a 1:1 mixture of Luminol/Enhancer solution and Stable Peroxide solution for 5 minutes without shaking. The membrane was briefly blotted on a paper towel and placed wet, RNA face side down, onto a clean sheet of plastic wrap. The membranes were imaged soon after, using the chemiluminescence protocol for the Bio Rad ChemiDoc™ XRS+ molecular imager.

5' 6-FAM (Fluorescein) Labeled RNA

These fluorescent EMSAs (fEMSAs) allowed for reduction in cost, time, and a large increase in both efficiency and quantification capabilities. Generally, the same procedure was followed as for the biotinylated RNA, with the following modifications:

- 1) When thawing protein aliquots and measuring A_{280} to calculate protein concentration, it was noted that the His₆-tagged constructs seemed to be more prone to aggregation than the His₁₀-SUMO tagged proteins, causing large variations in absorbance readings. Therefore, these samples were first centrifuged at 10,000 rpm for 20 minutes, the top 90% of supernatant was transferred into a new 0.5 mL tube, and then readings were taken.
- 2) After testing assays with RNase inhibitor and without, no observable difference was noted, so for the reported fEMSAs, this reagent was omitted.
- 3) Detection of RNA ligand. After electrophoresis finished, the gel itself was imaged directly using the Pharos FX™ Plus molecular imager, using the Quantity One software. The imaging platform was cleaned thoroughly before placing gel on, and the gel was handled as minimally as possible. The settings chosen based on optimal intensity and clarity of bands are as follows:

Fluorophore → FITC → Low. (The excitation and emission wavelengths [495 nm and 520 nm, respectively] are the same for FAM.)

Resolution: 50 μm .

Imaging area was narrowed in on gel bands as much as possible.

EMSA Quantification and Determination of the $K_{D,app}$

Quantification was done using Pharos FX™ Plus molecular imager, using the Quantity One software. Volume tool boxes outlined the boundaries of unbound (free) RNA and bound RNA. Quantification of the pixels found within each boundary for unbound and bound species was given as numerical values and exported as a text file.

Data was first processed in Excel for internal background correction, in which the lowest pixel count in boxes U1 – U12 was subtracted for the free RNA bands, and the lowest pixel count among boxes U13 – U24 subtracted for any bound species. Fraction saturated was then calculated (completely unbound = 0; completely bound = 1) for each lane by dividing fraction bound by the total signal for that lane. Then, fraction saturated versus protein concentration (pM) was plotted as XY pairs in SigmaPlot, with the x-axis was plotted in log scale.

The plot was fit to a simplified form of the binding isotherm:

$$\frac{[PL]}{L_T} = S \left(\frac{[P]_T}{[P]_T + K_{D,app}} \right) + O \quad (\text{eq. 2})$$

which assumes a two-state binding system. This means that is the total ligand (L_T) and total protein ($[P]_T$) are known, the apparent K_D ($K_{D,app}$) can be determined. The other variables represent saturation (S) and background offset (O).⁴¹ The error reported for the binding data is the standard error of the mean (SEM), which is found by first calculating the standard deviation of each $K_{D,app}$ value and then dividing by the square root of the number of replicates for that binding experiment. All reported $K_{D,app}$ values are derived from at least three independent replicates.

Circular Dichroism (CD)

CD spectroscopy allowed for analysis of protein secondary structure, particularly giving insight to regions of more ordered structures (α -helix or β -sheet, for example) and those which are more disordered (random coil).

The day prior to conducting CD, a Hellma Analytics High Precision quartz cell with a 1-mm pathlength was cleaned with alternative washes of 1 M NaOH and 1 M HCl, with

MQ H₂O, 70% and 100% ethanol rinses between each (allowing the cuvette to dry via vacuum post-ethanol) for a total of 2 base and 2 acid washes. Lens paper and some ethanol were used to clean any residue off the outside of the cuvette, and for general handling. The cuvette was stored in lens paper in a clean 50 mL conical vial until use the following day.

Protein samples were thawed, spun to pellet aggregates if indicated, and concentrations taken using the Implen NanoPhotometer® N60. An adequate volume of protein was thawed on ice per volume and concentration calculations.

Any calculations that needed to be done for standardizing buffer conditions between different protein samples (for example, varying salt concentrations, varying reducing agents, and so forth) were conducted (see tables below). Dilutions were made independently, a few minutes before CD was to be conducted. Dilutions were allowed to sit on the bench for a few minutes to get to room temperature to avoid bubbles forming in the cuvette in the 16° C sample chamber.

The cuvette was filled with 298 µL of a buffer blank and placed into the sample chamber of a JASCO J-710 spectropolarimeter. Compressed N₂ gas (Airgas) was used to purge the sample chamber for 15 minutes at ~22,000 PSI. The spectropolarimeter was then turned on, and after warming up, a Geoglobal Partners FP155 water pump was also engaged. The JASCO PFD-425S Peltier was also turned on. The JASCO Spectra Manager software on the attached desktop was opened and the following parameters set:

- 300 – 195 nm (or 198 nm if voltage was too high at 195)
- 20 nm/min
- Accumulation: 10
- Peltier selected; set to 16 °C
- Data collected: HT and CD

After allowed to equilibrate to temperature for at least 10 minutes, the N₂ PSI was reduced to ~15,000 and a blank spectrum was created in the absorption range of 195 – 300 or 198 – 300 nm (Table 12). It was necessary to increase the minimum wavelength for some samples to prevent the HT voltage from exceeding 800 V. After 10 runs of the sample were completed, the system was shut down in reverse order in which it was turned on: spectra manager, Peltier, water pump, and spectropolarimeter. The chamber was purged for another 15 minutes as the lamp cooled down, then the N₂ was turned off and the cuvette removed.

Only ultrapure polished H₂O and 100% ethanol were used to clean the cuvette between samples. The cuvette was completely dried via vacuum filtration while fresh protein dilution was being made.

The next sample was run the same way as the buffer blank, while selecting the “Baseline Subtract” option in the spectra manager to select and subtract the buffer blank that was previously conducted. All protein dilutions were at a final concentration of 0.17 mg/mL.

All CD spectra measurements were converted to mean residue ellipticity (MRE) by using the following formula:

$$[\theta]_{MRE} = \frac{CD \cdot MW}{l \cdot c \cdot n} \text{ deg cm}^2 \text{ dmol}^{-1} \text{ res}^{-1} \quad (\text{eq. 3})$$

where CD is circular dichroism (degrees), MW is molecular weight (g/dmol), l is pathlength (cm), n is number of residues, and c is concentration (g/mL). MRE was then plotted versus wavelength.

Difference spectra that represents the isolated *Drla*RP6A-NTR were plotted by subtracting His₆-La Module from His₆-ΔCTD and separately, His₁₀-SUMO-GG from His₁₀-SUMO-*Drla*RP6A-NTR.

Table 12: CD Preparation for His₆-DrLaRP6A-LaModule and

His₆-DrLaRP6A-ΔCTD.

Protein	SEC Purification Buffer	Buffer Blank	Spectrum Range (nm)
His ₆ -DrLaRP6A-LaModule	50 mM sodium phosphate, pH 8.0; 150 mM NaCl*	50 mM sodium phosphate, pH 8.0; 200 mM NaCl; 1 mM TCEP	195 – 300
His ₆ -DrLaRP6A-ΔCTD	50 mM sodium phosphate, pH 8.0; 200 mM NaCl; 1 mM TCEP		
His ₁₀ -SUMO-GG	50 mM sodium phosphate, pH 8.0; 200 mM NaCl; 1 mM TCEP	50 mM sodium phosphate, pH 8.0; 200 mM NaCl; 1 mM TCEP	198 – 300
His ₁₀ -SUMO-DrLaRP6A-NTR			

*Normalized for difference in [NaCl] and [TCEP] for each trial

Analytical Ultracentrifugation

Protein and the SEC purification buffer (50 mM sodium phosphate, pH 8.0, 200 mM NaCl, 1mM TCEP) were shipped to Dr. Chad Brautigam at UT-Southwestern Macromolecular Biophysical Resource. The buffer was used for the reference cells, and for preparing protein dilutions. The protein was thawed on ice for approximately one hour. Dilutions with target concentrations of 0.8, 0.25, and 0.08 OD were prepared and incubated overnight at 4 °C. After this incubation, the samples were centrifuged at 14,800 rcf for 10 min. Sample volumes of 400 μL were pipetted into the sample cells Microfine Green manufactured centerpieces. The same volume of the reference buffer was dispensed into the reference sectors. The cells were placed in an An50-Ti rotor and incubated in the centrifuge under vacuum at 20 °C for 2.5 hours. After incubating, the samples were centrifuged at 50,000 rpm. Absorbance data at 280 nm were collected and analyzed using the $c(s)$ methodology in SEDFIT. Figures were created in GUSI, and extinction coefficients were calculated using ProtParam.⁴²

UV-Vis Absorbance Methods A_{280} vs. A_{205} for DrLaRP6A-NTR Concentration

To verify this method of calculating protein concentration could be replicated prior to using it for DrLaRP6A-NTR concentration, a protein that was included in one study, Bovine Serum Albumin, was used.⁴³ Two separate dilutions were made, one in MQ H₂O and the other in a standard SEC purification buffer (50 mM sodium phosphate, pH 8.0; 200 mM NaCl; 1 mM TCEP) to see whether the buffer components absorbed highly at 205 nm.

First, a DrLaRP6A protein that could absorb at 280 nm, His₆-Dr6AΔCTD, was used. Concentrations were taken both within and outside of the mentioned absorbance range, again using both ultrapure polished water and the same SEC buffer to make various dilutions.

Table 13: Protein Extinction Coefficients for ϵ_{205} Experiments.

Protein	ϵ_{280}^{42} M⁻¹ cm⁻¹	ϵ_{205}^{43} M⁻¹ cm⁻¹
BSA	47,790	2,301,960
His ₆ -DrLaRP6A-ΔCTD	23,950	1,113,680

Table 14: Dilutions Prepared for ϵ_{205} Experiments.

Protein	Solvent	Dilution Factor
BSA	SEC buffer	3×
	SEC buffer	>25×
	Ultrapure H ₂ O	3×
	Ultrapure H ₂ O	>25×
His ₆ -DrLaRP6A-ΔCTD	SEC buffer	3×
	SEC buffer	>25×
	Ultrapure H ₂ O	3×
	Ultrapure H ₂ O	>25×

III. RECOMBINANT EXPRESSION & STRUCTURAL ANALYSIS OF LaRP6 N-TERMINAL REGION

Cloning of pET28-SUMO-D Δ LaRP6A- Δ CTD

Previous cloning in the lab of the *D Δ LaRP6A- Δ CTD* resulted in an insertion of two thymines between the protein sequence and the SUMO tag (Fig 12). To correct this mutation, site-directed mutagenesis (SDM) was attempted to remove these two bases and bring the protein sequence into the correct reading frame.

However, after several attempts of unsuccessful mutagenesis, new primers were designed to create the Δ CTD construct from the full-length sequence via restriction cloning. The primers were designed to insert two stop codons to create nonsense mutations (a “double-stop”) after residue 290 to ensure truncation of the sequence. The insert was amplified via PCR as described in Table 19 (Appendix) with an expected size of 894 base pairs. The insert was run on a gel to confirm this size, and was gel purified as described in the methods (Fig 13A).

The gel-purified insert and the pET28-SUMO vector were then both digested with restriction endonucleases *Bam*HI and *Xho*I, creating products with expected sizes of 870 base pairs and 5,608 base pairs, respectively. The vector was also treated with Antarctic Phosphatase (NEB) to prevent recircularization before ligation. Both digests were gel purified as above (Fig 13B). The remaining vector digest was stored at -20 °C and used for future cloning work.

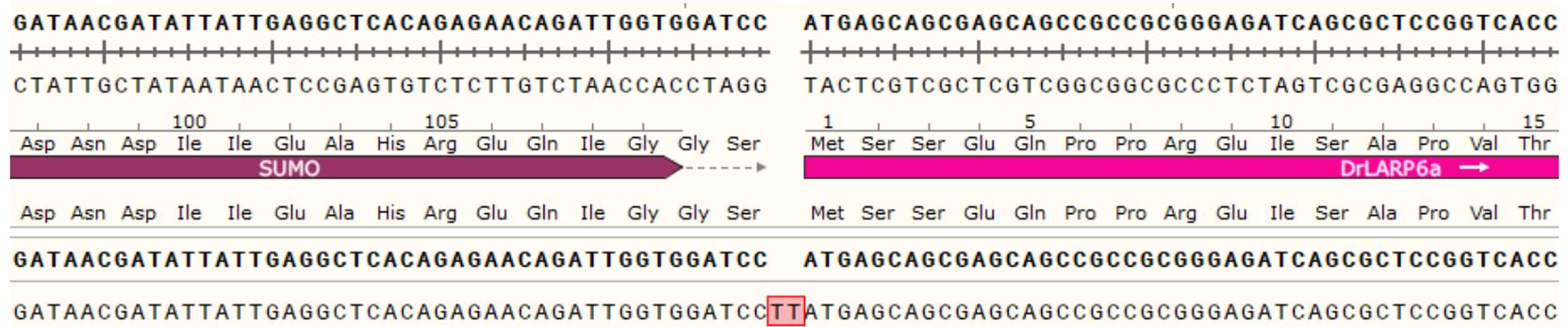


Figure 12: Thymine insertions between the SUMO tag and *DrLARP6A-ΔCTD* insert. This causes a frameshift. SDM was attempted to remove the two bases and revert the sequence back to the correct reading frame.

The digest fragments were ligated and immediately transformed into DH5 α cells. A negative control transformation with the digested vector but no insert was also plated to ensure proper digestion of the vector. Of the colonies that grew from the ligation reaction, seven were screened via colony PCR (Fig 13C). Samples 3, 4 and 5 were then sent for Sanger sequencing, from which all three sequences were verified.

A
Ladder *Drl*6A Δ CTD

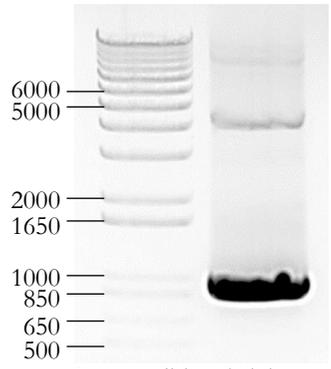
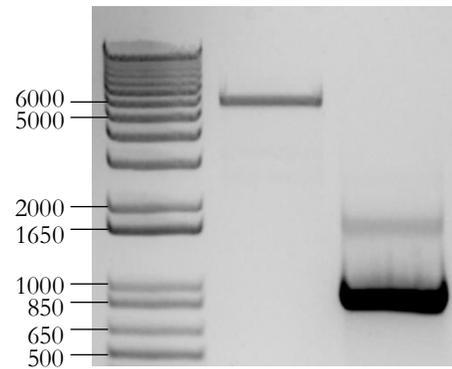


Image editing: brightness
-20%, contrast -40%

B
Ladder pET28-SUMO *Drl*6A Δ CTD



C
Ladder *Drl*6A Δ CTD

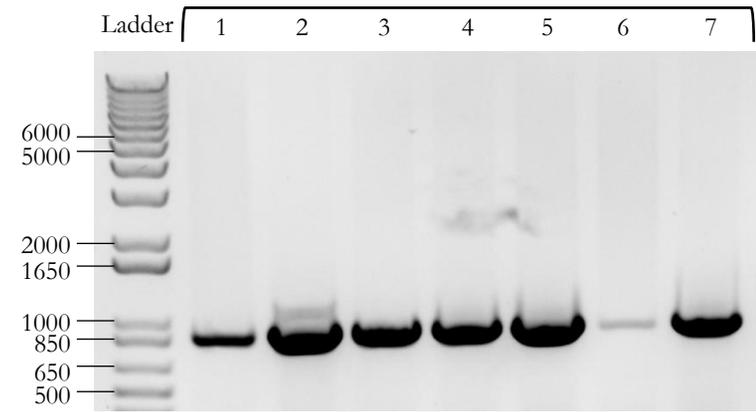


Figure 13: Cloning of pET28-SUMO-*Drl*6A Δ CTD. **A:** Insert amplification of *Drl*6A Δ CTD. **B:** Double digest of pET28-SUMO vector and *Drl*6A Δ CTD insert with *Bam*HI and *Xho*I. **C:** Colony PCR of *Drl*6A Δ CTD insert. Colonies 3, 4 and 5 were sent for Sanger sequencing.

Cloning of pET28-SUMO-D α LARP6A-N-Terminal-Region (NTR)

The first attempts at cloning the isolated *D α LARP6A-NTR* employed site-directed mutagenesis to create a “double-stop” in the full-length *D α LARP6A* sequence after residue 60. However, this approach was not successful despite multiple attempts with several adjustments of annealing and extension parameters.

Therefore, new primers were designed for restriction cloning. However, after cloning into pET28-SUMO, transforming into DH5 α cells, and collecting a cell pellet with the expressed protein, the molecular weight observed via SDS-PAGE analysis was approximately 40 kDa instead of the expected 20.4 kDa (Fig 14).

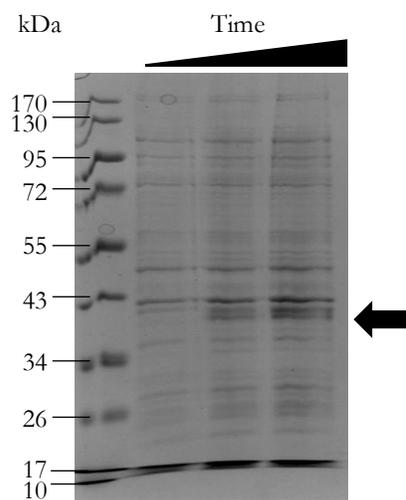


Figure 14: SDS-PAGE analysis of first His₁₀-SUMO-*D α LARP6A-NTR* expression. Arrow indicates product of interest.

Recombinant LARP6 generally migrates higher in SDS-PAGE, as both the full-length protein and its truncations are acidic in nature and thus run more slowly in SDS-PAGE, resulting in a higher apparent molecular weight.^{35,44} However, close evaluation *in silico* identified that the reverse primer originally designed for this cloning did not contain a stop codon at the C-terminal end of the amplified product. This was problematic because as our

template DNA (full-length *DrLaRP6A*) contains a C-terminal His-tag, this resulted in the NTR construct being double His-tagged, one on either terminus of the protein. Therefore, a new reverse primer had to be designed. The correctly sized insert was amplified at the expected 202 base pairs (Fig 15A). The insert was then gel-purified, double digested with *Bam*HI and *Xho*I, running at the expected size of 177 base pairs (Fig 15B). The digest was gel-purified and ligated into the digested pET28-SUMO vector, then transformed into DH5 α cells. The two colonies that grew were screened via colony PCR (Fig 15C), and the correct sequence and frame of the new insert was verified via Sanger sequencing (GeneWiz).

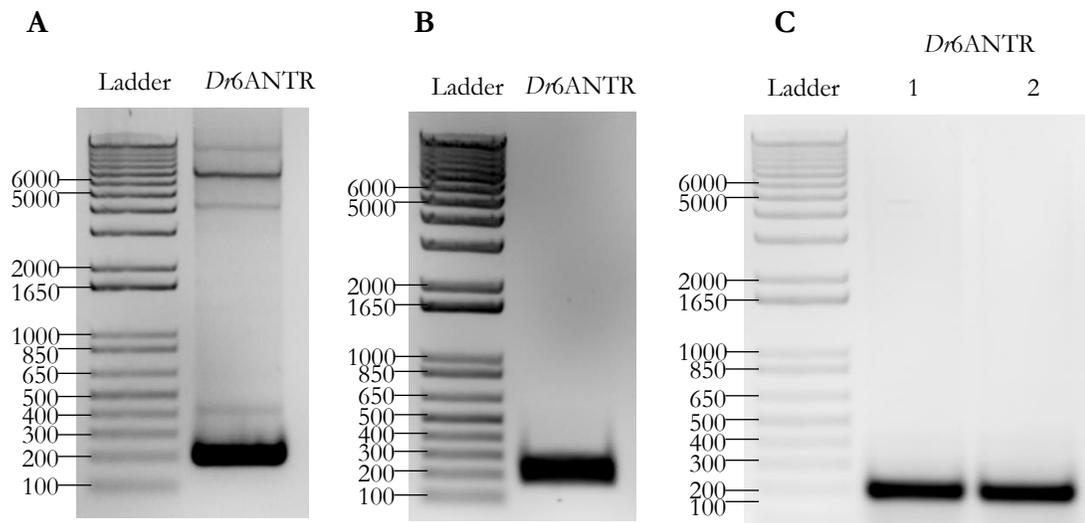


Figure 15: Cloning of pET28-SUMO-*DrLaRP6A*-NTR. **A:** Insert amplification of *DrLaRP6A*-NTR (expected size: 202 base pairs). **B:** Double digest of *DrLaRP6A*-NTR insert with *Bam*HI and *Xho*I (expected size: 177 base pairs). **C:** Colony PCR of *DrLaRP6A*-NTR insert. Both colonies were sent for Sanger sequencing.

Cloning of pET28-SUMO-GG: Negative Control for Tagged Dr6A-NTR

ULP1 is a cysteine protease that is used to cleave the His₁₀-SUMO tag of fusion proteins. ULP1 recognizes the SUMO tertiary structure, and catalyzes the hydrolysis of the peptide backbone within a Gly-Gly-Ser cleavage site, between the second glycine and the serine. All the cleaved proteins are thus left with a single serine residue on the N-terminus.

To create a true negative control protein, site-directed mutagenesis was used to insert a double-stop after Gly-Gly on the full-length pET28-SUMO template, hence the name “His₁₀-SUMO-GG”.

Just as with the His₁₀-SUMO-NTR construct, all attempts at site-directed mutagenesis to make His₁₀-SUMO-GG were unsuccessful. We hypothesized that the problem may have been due to the design of the primers, which were 32 nt long and 100% complementary. To test this hypothesis, two alternative sets of primers were created. The first set was designed in a similar fashion as the previous SDM primers, with the desired mutation centered within each primer sequence (this set was dubbed the “Middle” primer set). The second set was designed to be of similar length, GC content, and T_M, but to only share 25 nt of complementarity; the remainder of each primer were 5'-overhangs, in which the mutation of interest was offset towards the 3'-end of each primer (this set was called the “Overhang” set).

The template used was the vector used in the cloning for the previous constructs, full-length pET28-SUMO. After PCR amplification, and digestion with *DpnI* (NEB) to digest the parental DNA, the PCR products were immediately transformed into DH5 α cells. While the reaction using the “Middle” primer set did not produce any colonies, the “Overhang” PCR reaction yielded eight colonies. Plasmid was extracted from all colonies and analyzed by Sanger sequencing; all eight colonies had the mutation of interest. Perhaps the double-stranded DNA-binding domain of Phusion® polymerase prevents efficient melting of 100% complementary primers.

Expression of Protein Constructs

Sequence-verified constructs were expressed in Rosetta™ (DE3) cells, as described in Chapter 2: Materials and Methods. Initial trial expressions were conducted to find the optimal growth temperature and time for each protein construct. All proteins were expressed at 18 °C to start, as many recombinant proteins (full-length LaRP6 included) have improved solubility at colder temperatures. The optimal length of expression was chosen based off the timed aliquot that produced the strongest band at the expected molecular weight in an SDS-PAGE gel, with the least amount of degradation products. These final times and temperatures are listed in Table 7 in the methods section.

All expected and observed molecular weights via SDS-PAGE analysis are listed in Table 15, along with the calculated apparent molecular weight (MW_{app}) from size exclusion chromatography.

Expression efficiency was determined by running two sister gels of the timed aliquots. The first was stained with Coomassie stain to show all proteins present (including bacterial), and the second was used in an anti-His Western Blot to detect expression of the His-tagged protein of interest specifically (Fig 16). All constructs successfully expressed.

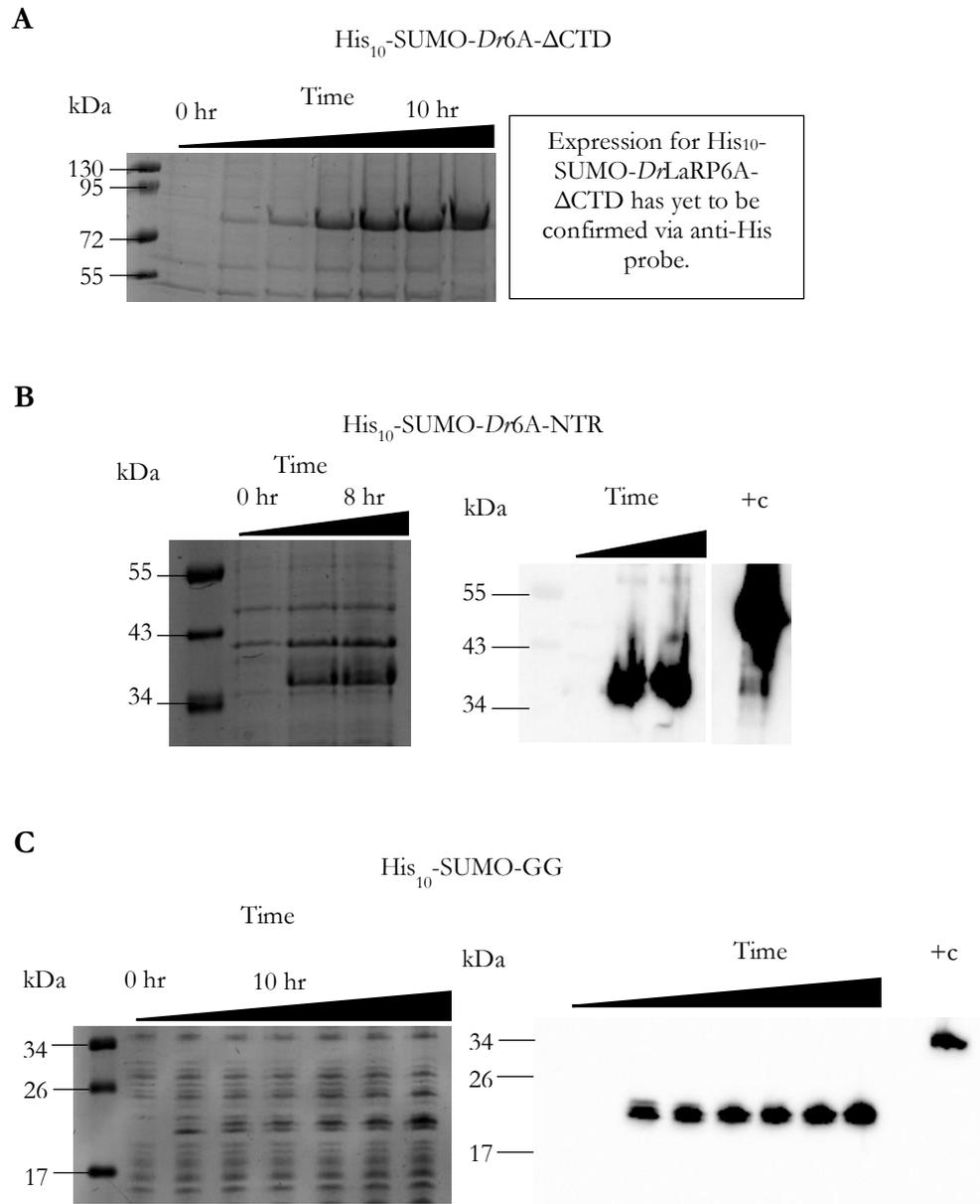


Figure 16: Trial expressions stained with Coomassie stain and verified by anti-His probe. Coomassie: left images. Anti-His blot: right images. The labeled hours were determined to be the optimal expression length for each construct. **A:** His₁₀-SUMO-Dr6A-ΔCTD. **B:** His₁₀-SUMO-Dr6A-NTR. **C:** His₁₀-SUMO-GG.

Protein expression cultures for preparative purifications were carried out in 1 L culture volumes (a.k.a. “large-scale expression”), during which three aliquots were taken. The prepared samples were run on SDS-PAGE and stained with Coomassie stain to verify protein expression prior to purification.

Table 15: A Comparison of Protein Molecular Weights.

Protein	Expected Molecular Weight (ProtParam) (kDa)	Approximate Observed Molecular Weight via SDS-PAGE (kDa)	MW_{app} (Calculated from Size Exclusion Chromatography) (kDa)
His ₆ -DrLaRP6A-ΔCTD	34.028	50	77.328
His ₁₀ -SUMO-DrLaRP6A-NTR	20.383	34	47.100
His ₁₀ -SUMO-GG	14.072	22	19.009
His ₁₀ -SUMO-HsLaRP6-tNTR	23.913	34	45.715

His₆-DrLaRP6AΔCTD is Sensitive to Red-Ox Conditions

Cloning and expression of this construct was originated by a previous lab member³⁵. To make a 1-L cell pellet for purification, sequence-verified DNA was transformed into Rosetta cells, followed by a large-scale expression. The 1-L cell pellet was sonicated and purified via Ni-NTA IMAC.

When purified under non-reducing conditions, the SEC chromatogram reveals two distinct species, one eluting at 71.5 mL and the second at 80.1 mL (H.K. & L.G., Fig. 17A). It was suspected that a dimer was being formed, while the second peak corresponding to lower molecular weight was monomer. To confirm this, the first peak was re-injected to see if it re-distributed into the dimer and monomer (Fig 17A, gray trace). As predicted, the first peak correlated to dimerized ΔCTD, and the second monomer.

This preparation was analyzed by analytical ultracentrifugation (Dr. Chad Brautigam, UT-Southwestern Macromolecular Biophysical Resource) (Fig 17B). Two peaks were observed, one with a calculated molecular weight of ~77 kDa and the other ~157 kDa – implying the presence of both monomer and dimer. Together, the SEC and AUC data indicated that *D α* LaPR6A- Δ CTD was forming a covalent dimer. Analysis of the N-terminal region sequence revealed a single cysteine (Table 6), suggesting that a covalent dimer may be formed by a disulfide bond via the NTR.

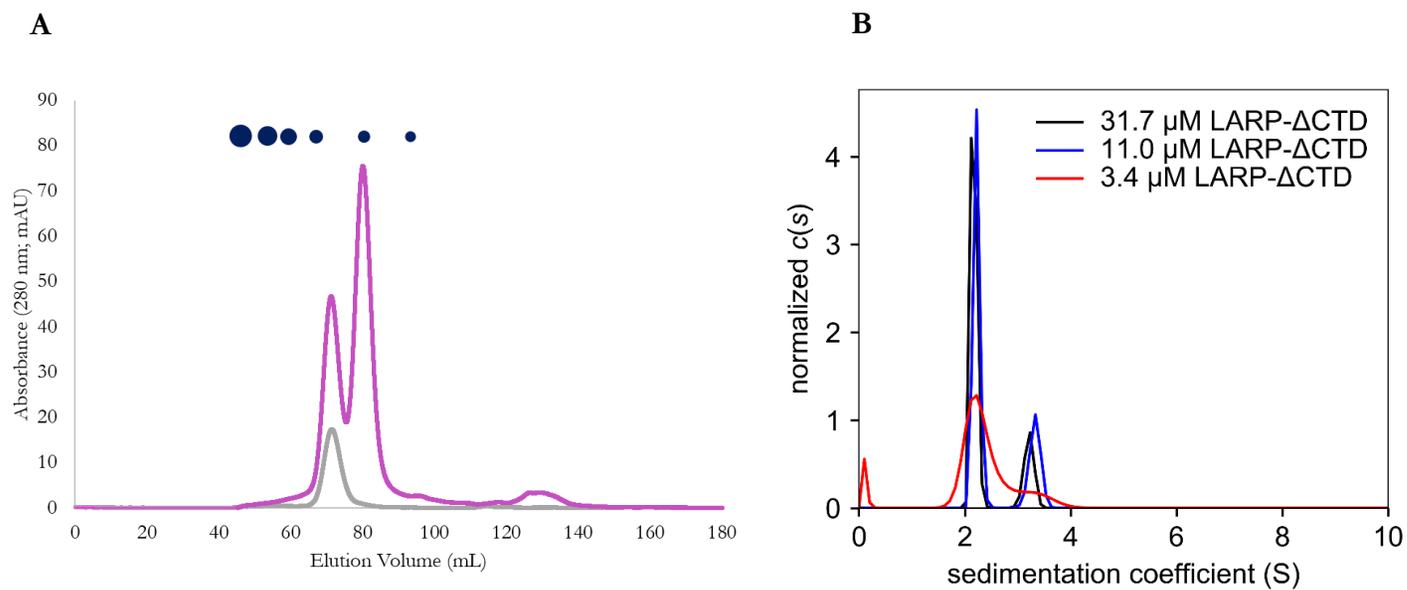


Figure 17: SEC chromatogram of His₆-DrLARP6A-ΔCTD under non-reducing conditions. A: S200 chromatogram from size exclusion chromatography, the initial purification (pink) and the suspected dimer re-injected (gray). **B:** AUC data confirming monomer and dimer.

To test this hypothesis, a second purification of Δ CTD was performed under reducing conditions (2 mM β -mercaptoethanol [β ME] in all buffers). The IMAC fractions E1 and E2 were pooled as they contained the expected protein size as assessed by SDS-PAGE gel analysis (Fig 18A). These pooled fractions were then subjected to size exclusion chromatography. One peak was observed at an elution volume of 80 mL with a calculated apparent molecular weight (MW_{app}) of \sim 77 kDa (Fig 18B). Fractions of interest were chosen from the chromatogram to analyze via SDS-PAGE (Fig 18C). Fractions 27, 28, 29 and 30 were collected, pooled, concentrated to 35 μ M (\sim 3.5 mg protein). The concentrated protein was partitioned into 50 μ L aliquots and stored at -70 $^{\circ}$ C.

In the presence of 2 mM β ME in all purification buffers (IMAC and size exclusion), only one peak was observed at 80.1 mL. This elution volume is almost exactly the same elution volume as the supposed “monomer” species in the non-reducing preparation, supporting the intermolecular disulfide bond hypothesis. To confirm monodispersity and more accurately measure molecular weight, this reduced Δ CTD preparation was also analyzed by AUC. Only one peak was observed, and its intensity increased with protein concentration. This confirmed that the reduced Δ CTD preparation was a solution of solely monomeric protein (Fig 18D).

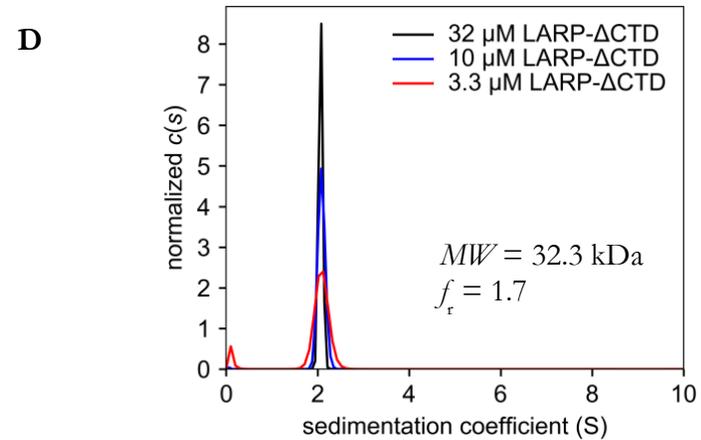
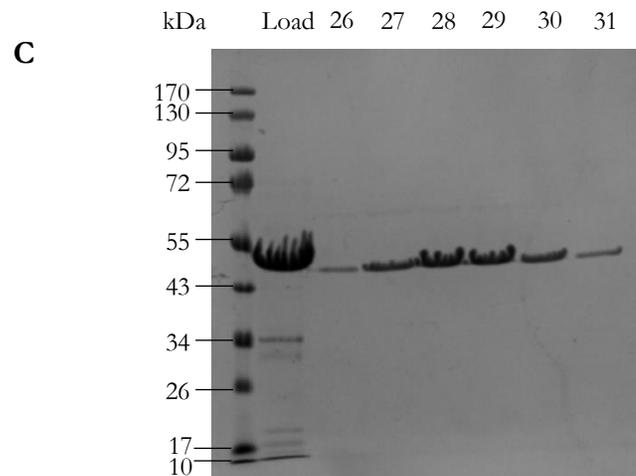
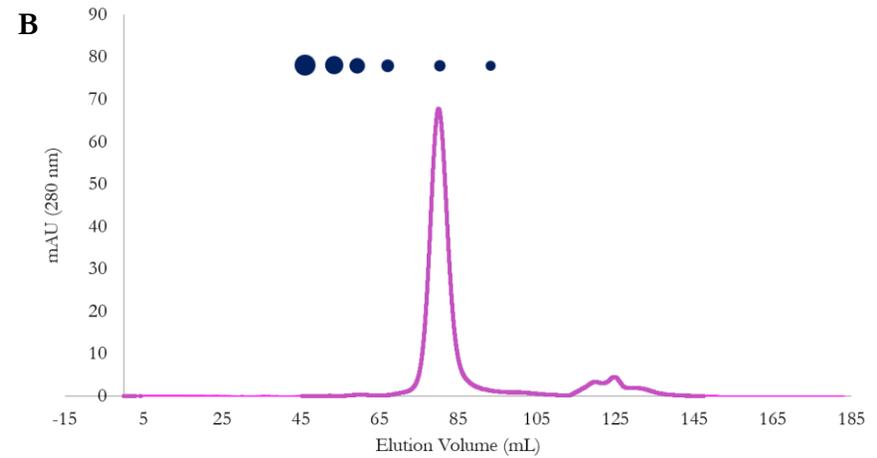
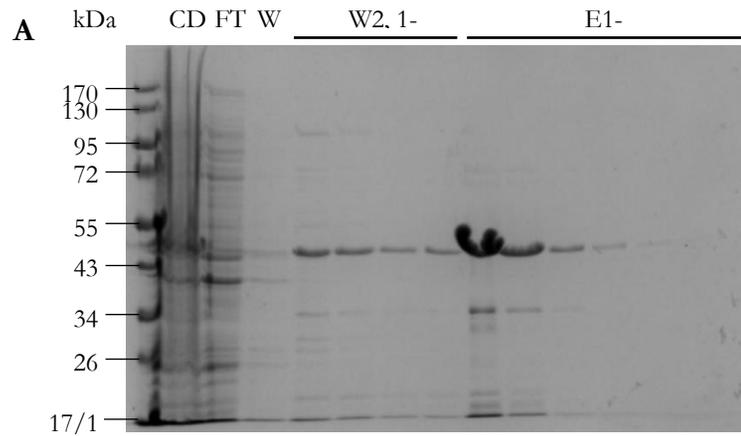


Figure 18: SDS-PAGE analysis and chromatogram from His₆-DrLARP6A-ΔCTD purification under reducing conditions. A: SDS-PAGE analysis of IMAC fractions. **B:** S-200 chromatogram from size exclusion chromatography. **C:** SDS-PAGE analysis of SEC fractions. **D:** AUC data confirming monomer species. It can be noted the frictional ratio is 1.7, indicated a more elongated and less globular protein.

Purification of DrLaRP6A-NTR

For all work described in this thesis, protein concentration was determined via Beer's law. The A_{280} for each sample was taken using Implen NanoPhotometer® N60 and the extinction coefficient of the protein at 280 nm as calculated via ProtParam.⁴² However, the calculated extinction coefficient of His₁₀-SUMO-DrLaRP6A-NTR is zero after cleavage of the tag. This is primarily due to the lack of tryptophan, tyrosine, and phenylalanine residues in the NTR. This means that after cleaving the His₁₀-SUMO tag, the protein concentration could not be taken using traditional intrinsic tryptophan absorbance at 280 nm. Additionally, the second sizing column that is typically used after cleavage would have to be adapted, as monitoring at 280 nm (the only wavelength monitor on the FPLC) would be futile. There was also a concern that the cleaved protein may approach the lower limit of the S75 column, as the column's lowest calibration marker, aprotinin, is approximately 6.5 kDa, while the tag-cleaved DrLaRP6A-NTR is 6.4 kDa.

Exploration for alternative wavelengths for calculating an accurate protein concentration led to a study by Anthis and Clore, using absorbance at 205 nm.⁴³ This is a logical alternative wavelength to 280 nm, as the protein backbone absorbs strongly here. Anthis and Clore created their own algorithm for calculating the molar extinction coefficient of polypeptide chains at 205 nm by using an averaged value of the protein backbone (which relies on the length of your peptide) and the sum of the residues which are first multiplied by their individual ϵ_{205} values.⁴³

It is important to note that the authors state absorbance values well outside of the range of the spectrometer they used; whether the absorbances were calculated from a dilution factor or not is unclear. The manual of the spectrometer used in this work, the Implen NanoPhotometer® N60, states the photometric range for a 10 mm pathlength is

0.02 – 300 AU, so data presented here were collected both within and outside of the assumed linear region of the instrument (0.10 – 2.00 AU).

One protein used in the study, bovine serum albumin (BSA), was first used in this work to test the repeatability of calculating protein concentration with 205 nm (Tables 16 and 17). A comparison study was then conducted at 280 nm and 205 nm of a *D_rLaRP6A* protein that can be monitored at both wavelengths, the Δ CTD (Table 18).

Several initial trials found that one sample dilution could not give absorbances at both 205 and 280 nm that allowed both readings to fall within the linear range of the instrument. Therefore, dilutions were created when necessary, and the initial concentration calculated using the dilution factors. The background absorbance spectra of the solvents, even after blanking, were also taken to ensure they were not significantly contributing to the absorbance.

Table 16: BSA Data at 280 nm for ϵ_{205} Experiments.

Sample	A280 Average	Concentration	Back-Calculated Concentration
BSA in SEC	1.32	27.63 μM	82.89 μM
BSA in MQ	1.31	27.34 μM	82.01 μM

Table 17: BSA Data at 205 nm for ϵ_{205} Experiments.

Sample	A205 Average	Concentration	Back-Calculated Concentration
BSA in SEC	1.54	0.67 μM	72.81 μM
BSA in MQ	1.55	0.67 μM	73.25 μM

Table 18: His₆-DrLaRP6A- Δ CTD Data at 280 and 205 nm for ϵ_{205} Experiments.

Sample	Average Absorbance	Concentration*
His ₆ -DrLaRP6A- Δ CTD at 280 nm	1.34	55.81 μM
His ₆ -DrLaRP6A- Δ CTD at 205 nm	1.05	25.24 μM

*Stock protein was already in range for 280 nm, so didn't back calculate, but did for 205 nm

Simultaneously, a pilot purification was carried out to produce untagged NTR. Despite the A_{280} limitation, the first trial purification did include tag cleavage to observe how pure the product would be after the second Ni-NTA column, since the second sizing column was to be omitted. A 1 L cell pellet was sonicated and purified via the first Ni-NTA column, in which fractions E1 and E2 were chosen to move forward with the purification

(Fig 19A). The chromatogram produced via SEC showed one large peak with an elution volume of 59.7 mL and a calculated MW_{app} of ~ 47 kDa (Fig 19B). Fractions 17- 20 were pooled to proceed (Fig 19C).

The pooled protein was incubated with ULP1 to cleave the His₁₀-SUMO tag as described in Materials and Methods. Afterwards, the solution was run over a fresh Ni-NTA column, onto which the cleaved tag was expected to bind to the resin while the free protein of interest should elute in the flowthrough.

The tag-cleaved protein was expected to run at a higher apparent molecular weight in an SDS-PAGE gel due to its heightened acidity.⁴⁴ There was therefore a concern that its bands would run in line with those of the cleaved His₁₀-SUMO tag. Conversely, if it did not run that high, a second worry was that it would run along with the dye front. To try and prevent these issues, a higher percentage of gel was used (15%), as well as a 5× sample buffer containing no dye. However, no protein was observed in the flowthrough or wash fractions (Fig 19D).

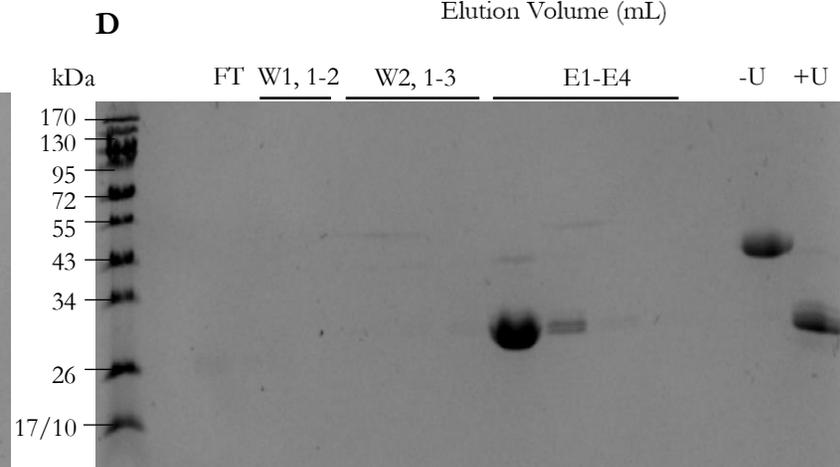
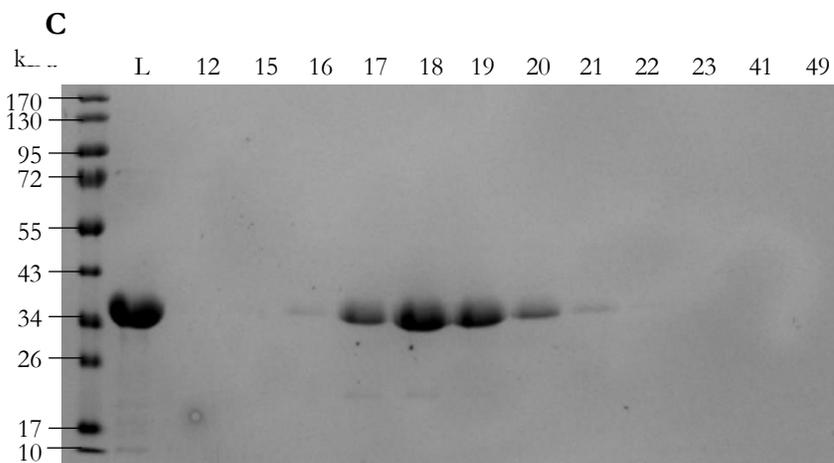
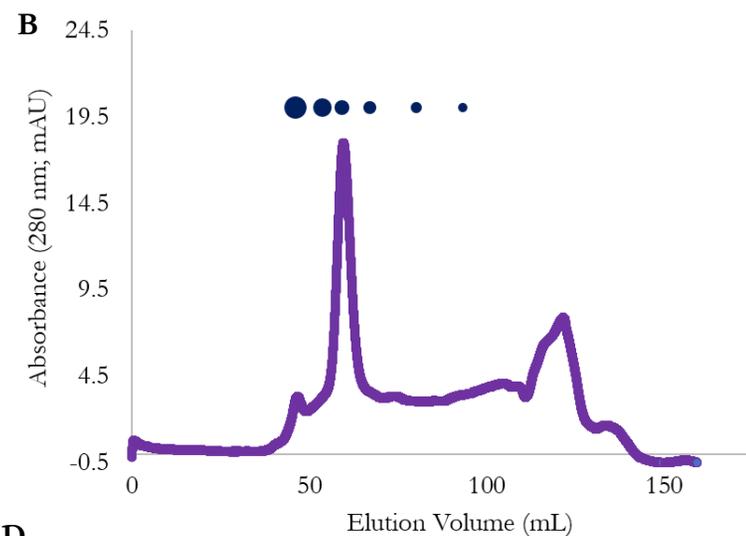
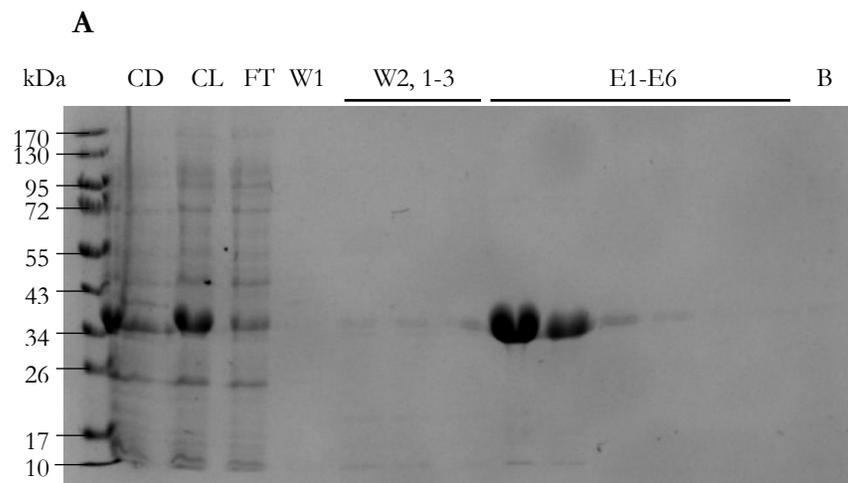


Figure 19: Pilot purification of His₁₀-SUMO-DrLARP6A-NTR. **A:** SDS-PAGE analysis of first IMAC fractions. **B:** Chromatogram from size exclusion chromatography. **C:** SDS-PAGE analysis of size exclusion chromatography fractions. **D:** SDS-PAGE analysis of second IMAC.

To further assess purification products, new gels were run and subjected to silver staining. With the first silver stain, all IMAC fractions where the protein was expected to possibly elute were compared on a 15% SDS-PAGE gel, along with samples of the protein pre- and post-tag cleavage. (Fig 20A). In a second attempt to elucidate the location of the protein of interest, a 4-20% gradient gel was used, and several more samples were run along with some of the previous (Fig 20B).

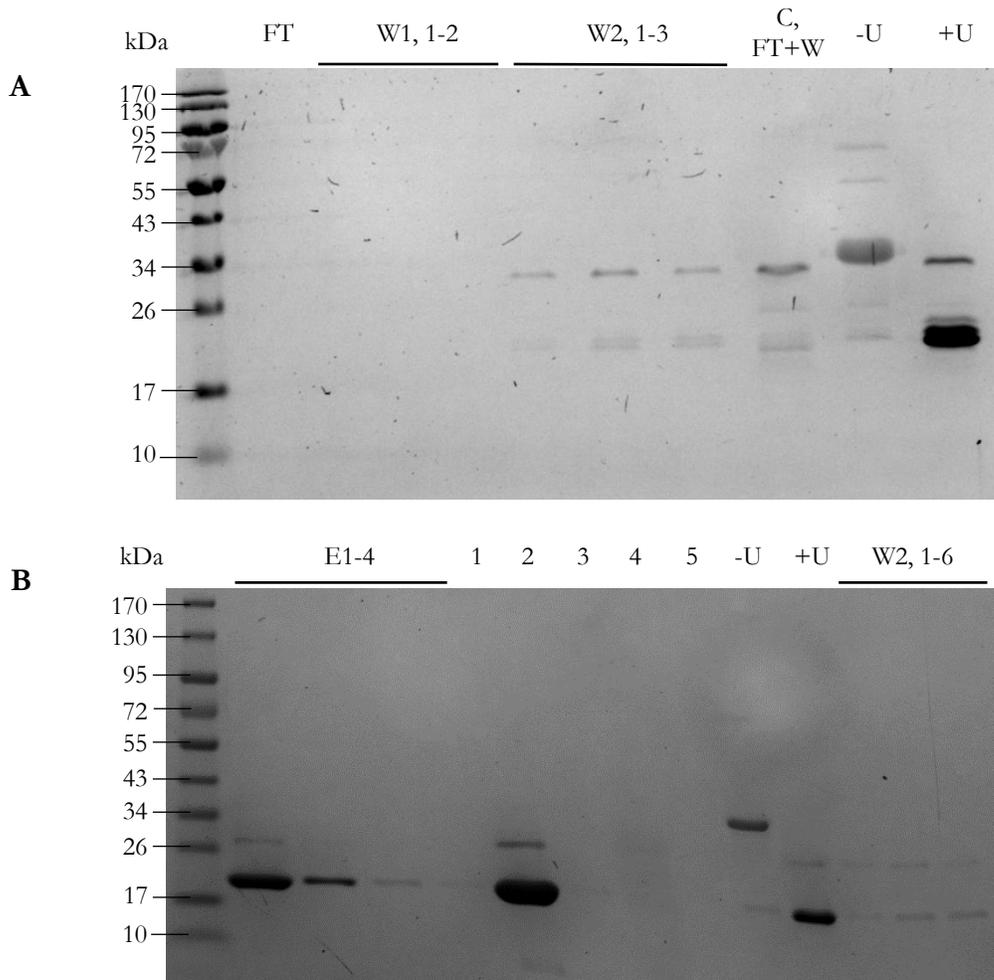


Figure 20: Silver stains of *DrLaRP6A-NTR*. A: First silver stain. **B:** Second silver stain running more samples.

-U: Pre-ULP1 digestion. +U: Post-ULP1 digestion.

1: Concentrated E1.

2: Concentrator flowthrough of concentrated E1.

3: Concentrated FT + washes.

4: Concentrator flowthrough of concentrated FT + washes.

5: empty.

Note: Gel ran at a slant because the tape at the bottom of the pre-cast gel was not removed prior to electrophoresis.

Overall, the location of the NTR protein was unclear. The protein itself may be degrading upon cleavage from the His₁₀-SUMO tag. Additionally, the absence of data from the FPLC chromatogram for comparison significantly compromised the confidence in this purification. Given this data was inconclusive and the final product not pure enough without the second sizing column purification, the His₁₀-SUMO tag was left on the NTR for the second purification. This therefore eliminated the ULP1 digestion step, as well as both the second Ni-NTA column and second sizing column. From the only round of IMAC in the second purification, fractions E1 and E2 contained the highest amount of protein of interest (Fig 21A). These fractions were thus pooled and run over the sizing column (Fig 21B). There was one peak observed with an elution volume of 59.7 mL, and a calculated MW_{app} of ~47 kDa. Fractions 17-20 contained the protein of interest (Fig 21C) were pooled and concentrated to 80 μM (~9 mg of protein). The concentrated protein was divided into 40 μL aliquots and screened for four different storage conditions (see below).

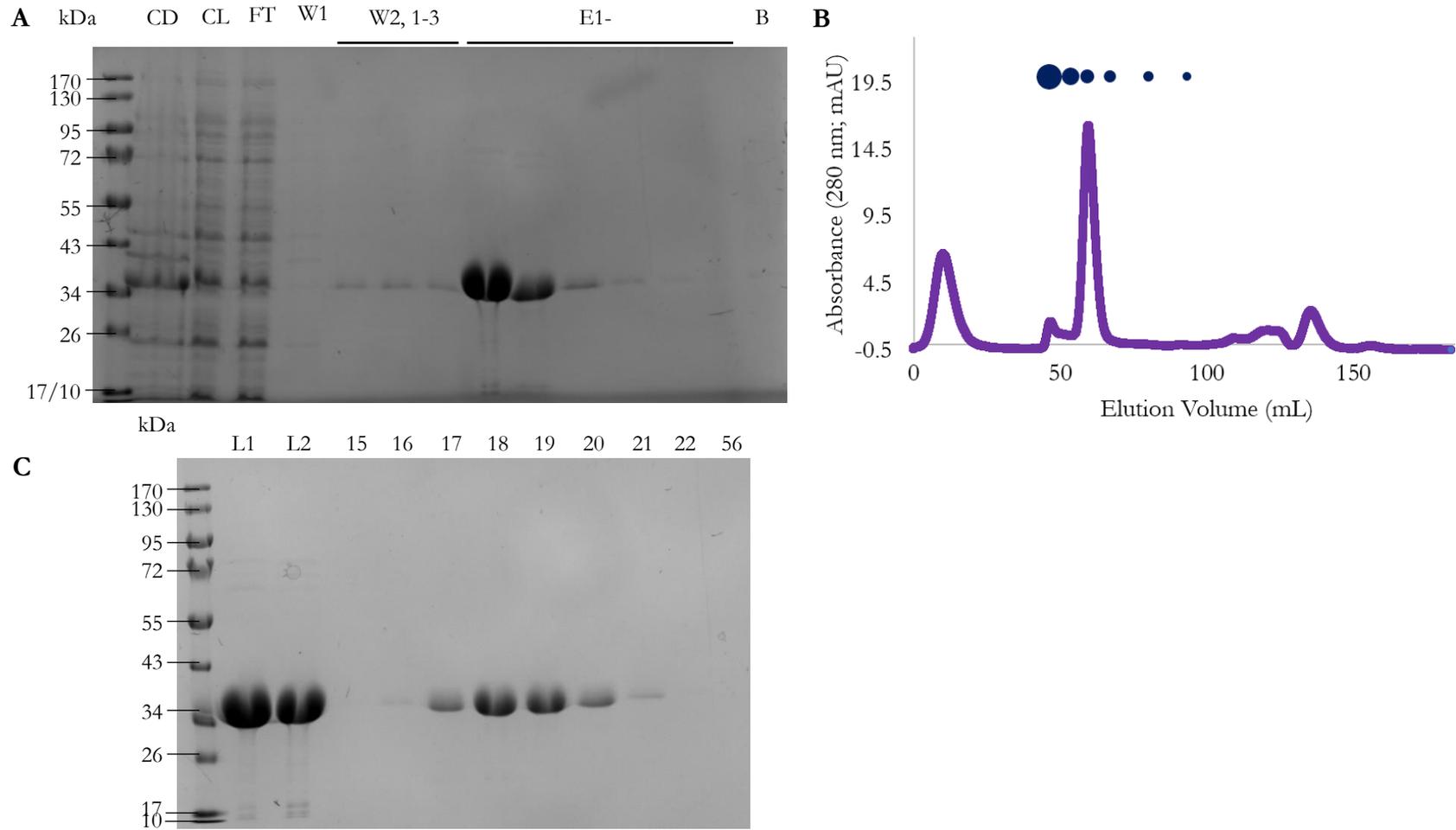


Figure 21: Second purification of His₁₀-SUMO-*Dt*LaRP6A-NTR, leaving the tag on. A: SDS-PAGE analysis of IMAC fractions B: Chromatogram from size exclusion chromatography. C: SDS-PAGE analysis of size exclusion chromatography fractions.

Purification of His₁₀-SUMO-GG (Negative Control for Tagged DrNTR)

A 1-L cell pellet was sonicated and tandem purified as with the other constructs above. The Ni-NTA fractions E1 and E2 were selected to move forward with the purification (Fig 22A).

The S75 sizing column produced two peaks. As SUMO is a globular protein, it was deduced to have eluted at a MW_{app} closest to its expected molecular weight (~14 kDa). Therefore, the second peak at 74.8 mL is the correct elution peak, with a calculated MW_{app} of ~19 kDa (Fig 22B). From the sizing column fractions, 24-28 were pooled and concentrated to 50 μM (~2.5 mg of protein) (Fig 22C). Then, 50 μL aliquots were made and tested with four different storage conditions (see below).

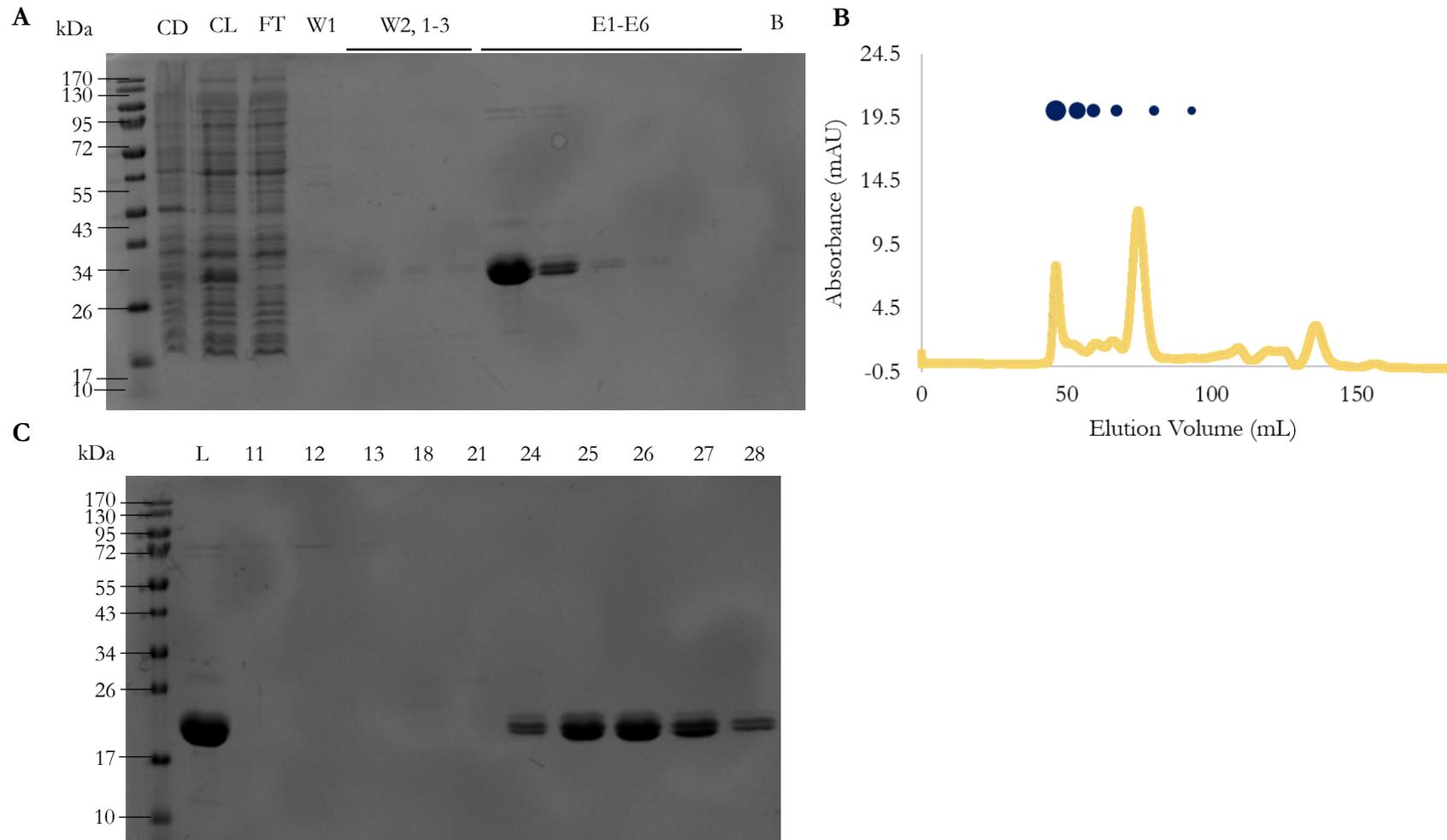


Figure 22: Pilot purification of His₁₀-SUMO-GG. A: SDS-PAGE analysis of IMAC fractions. **B:** Chromatogram from size exclusion chromatography. **C:** SDS-PAGE analysis of size exclusion chromatography fractions.

Storage Conditions for Purified Proteins

For most of the pilot purifications listed above, different storage conditions were tested to find the optimal storage temperature, as well as to see if glycerol can help preserve the protein:

- 1) -70 °C, no glycerol
- 2) -70 °C, 2% glycerol
- 3) 4 °C, no glycerol
- 4) 4 °C, 2% glycerol

Aliquots of each protein were assessed for changes in concentration using Beer's law and absorbance at 280 nm on the Implen NanoPhotometer® N60. After comparing the concentrations of each construct at each condition as time passed, -70 °C was found to be the optimal temperature. Additionally, glycerol does not appear to be needed for storage of these constructs.

Circular Dichroism

Circular dichroism spectra were collected for the Δ CTD and La Module proteins and converted to mean residue ellipticity (MRE) as described in the Methods. Three independent replicate spectra were averaged and plotted versus wavelength (Fig 23A). Spectra were similarly collected for the SUMO and SUMO-NTR constructs (Fig 23B). Since CD is an additive technique, the difference spectrum for each of these pairs was also plotted, which reflects the secondary structural characteristics of the sequence that differs between the two – i.e., the N-terminal region.

The local minimum at 222 nm and global minimum at 208 nm for the Δ CTD and isolated La Module are reflective of the expected mixed α/β structure of the La Module. For

SUMO and SUMO-tagged NTR, while there is a local minimum at 222 nm, there is a much more pronounced global minimum at 200-205 nm, consistent with strong β -strand content of SUMO.

The difference spectra for both sets of proteins reports on the structure of the NTR sequence in both contexts. This analysis reveals a global minimum at 198, wholly consistent with a random coil polypeptide (Fig 23C). In fact, the two difference spectra are very similar, disproving the original hypothesis that the NTR adopts a folded, globular structure.

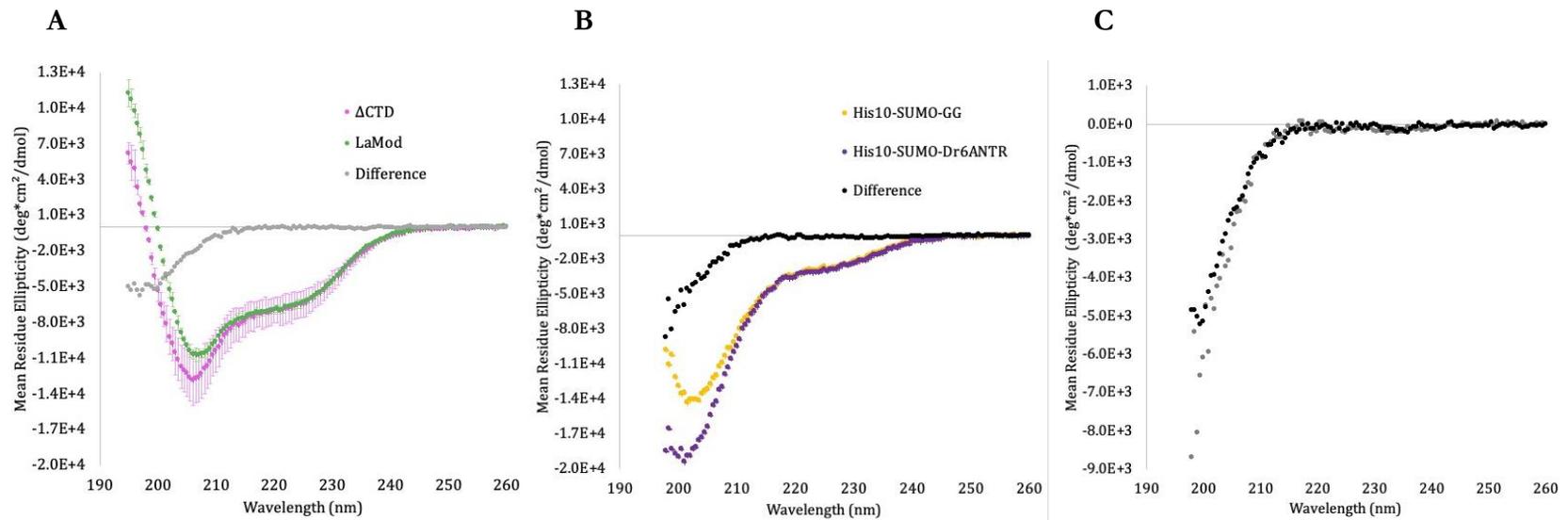


Figure 23: Circular dichroism spectra of *DrLaRP6A* constructs and the SUMO control. A: *His₆-DrLaRP6A-ΔCTD* (pink) and *His₆-DrLaRP6A-LaMod* (green). The difference spectrum is in gray. **B:** *His₁₀-SUMO-GG* (yellow) and *His₁₀-SUMO-Dr6ANTR* (purple). The difference spectrum is in black. **C:** The two difference spectra from A and B overlaid.

IV. ROLE OF LaRP6 NTR IN RNA BINDING ACTIVITY

As mentioned in the introduction, the importance of the NTR in protein function was observed in several other LaRP families. To test the hypothesis that the NTR in LaRP6 can modulate RNA binding and potentially even bind ligand independently, a series of electrophoretic mobility shift assays were conducted and compared.

RNA Binding Assays Against Homopolymers

RNA binding activity was tested using three different fluorescent RNA homopolymers: A₂₁ (Poly-A), U₂₀ (Poly-U), and UC₁₉ (Poly-C). Running electrophoretic mobility shift assays (EMSAs) allowed for analysis of the binding capacity for each protein against each ligand.

After several trials with 1 nM RNA, a final concentration of 0.5 nM was selected. Protein dilutions were allowed to equilibrate for 1 hour on ice, in the dark with 5'-FAM-tagged ligand. Free ligand was separated from the bound RNA-protein complex(es) on 6.5% native polyacrylamide gels, which were run with cold buffer and ice packs to reduce heat. The gels were then immediately imaged on the Pharos FX™ Plus molecular imager, as described in the methods, ensuring cleanliness of the imager and minimal handling of the gel. Representative gels will be shown for each protein:ligand experiment. The lower bands correspond to unbound RNA, while the higher bands represent LaRP6 bound to the ligand.

Qualitatively, the bands observed for His₆-D_rLaRP6A-ΔCTD (“ΔCTD” hereafter) bound to ligand show distinct shifting behavior as compared to the His₆-D_rLaRP6A binding (“La Module” hereafter) for Poly-A and Poly-U, the former forming more discrete and less smeary shifts of unbound to bound RNA (Fig 24, Panels A and B). The smearing seen with

the La Module may indicate a faster off-rate in the system, until a more stable complex is formed at higher protein concentrations.

It can be noted that in the gels with no observed binding, the intensities of free RNA bands varied across the gel, but differed in pattern across multiple independent replicates, supporting that the differences in intensity is not a result of binding.

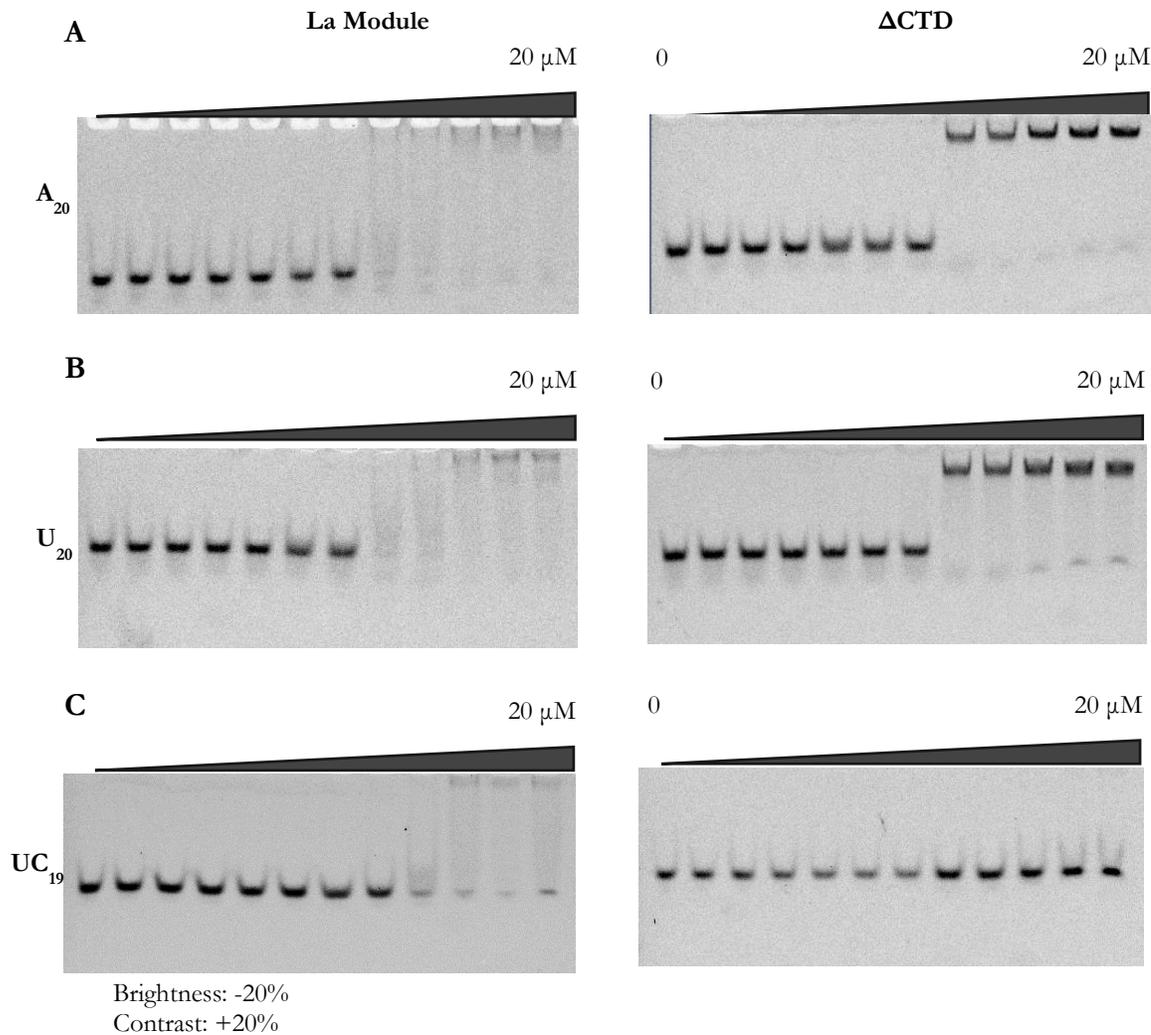


Figure 24: Fluorescent electrophoretic mobility shift assays of His₆-DrlARP6A-La Module and His₆-DrlARP6A- Δ CTD versus FAM-tagged homopolymeric RNA. Left images: La Module. Right images: Δ CTD. **A:** A₂₀ (Poly-A). **B:** U₂₀ (Poly-U). **C:** UC₁₉ (Poly-C).

Intriguingly, while binding was observed with Poly-C and the La Module (albeit more weakly than the other two homopolymers), the Δ CTD did not bind to this ligand, although it bound to Poly-A and Poly-U more stably than the La Module (Fig 24C).

Unsurprisingly, the negative control for the NTR, His₁₀-SUMO-GG, did not bind to any of the ligands (Fig 25). The NTR did not exhibit binding with the homopolymers tested, either (Fig 25).

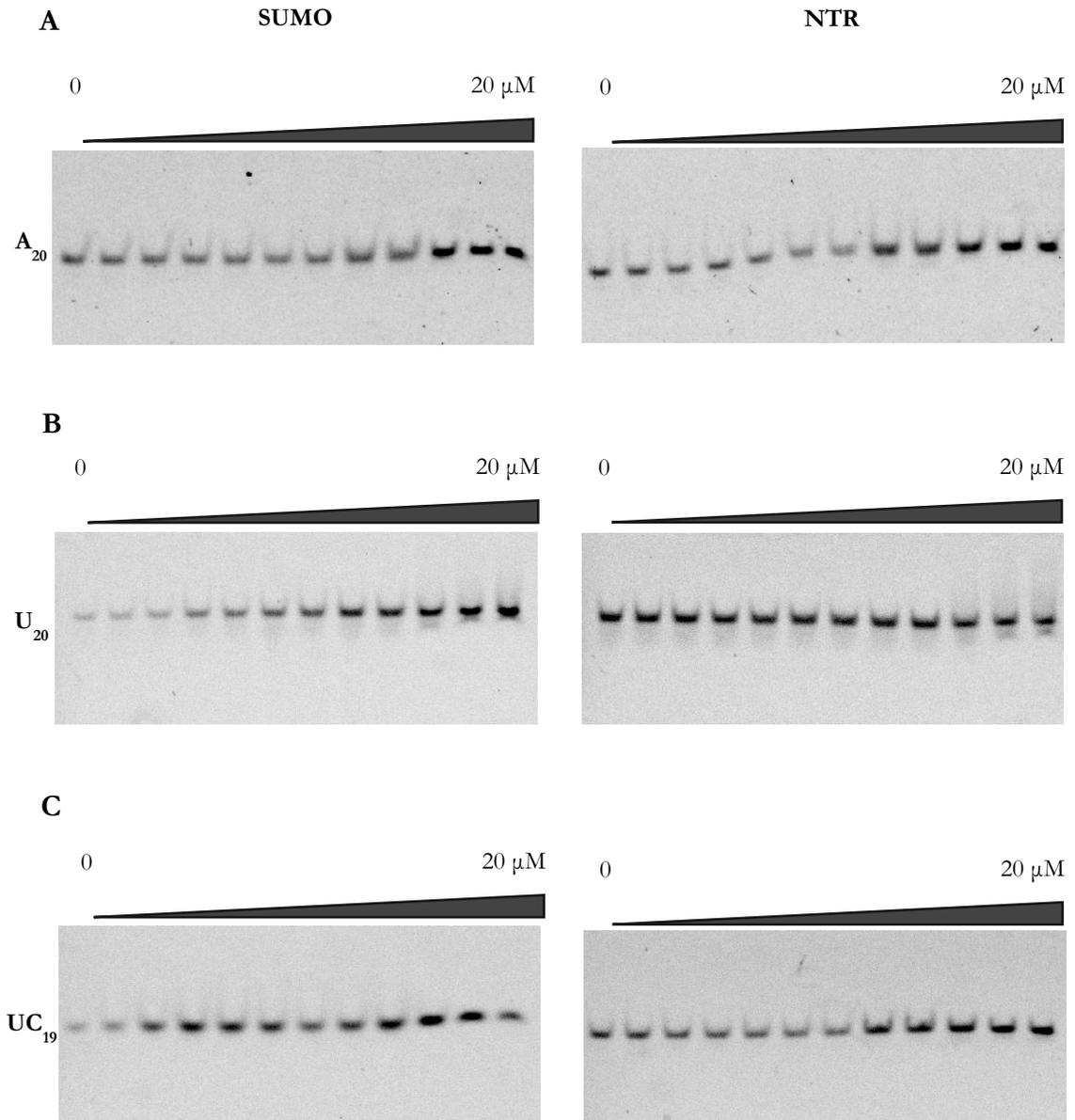


Figure 25: Fluorescent electrophoretic mobility shift assays of His₁₀-SUMO-GG and His₁₀-SUMO-*DrlARP6A*-NTR versus FAM-tagged homopolymeric RNA. Left images: SUMO. Right images: NTR. **A: A₂₀ (Poly-A). **B:** U₂₀ (Poly-U). **C:** UC₁₉ (Poly-C).**

RNA Binding Assays Against HsCOL1A1

As discussed above, the only known endogenous ligand for LaRP6 is the 5' untranslated region of collagen type 1 mRNA. Initial binding assays were conducted using biotinylated *HsCOL1A1* stem loop mRNA, as described in the methods (Fig 26). It was noted that resolution is lost during the extra transfer step, as can be seen in the lanes with higher protein concentration. Other major drawbacks with this detection method include length of time and cost of reagents.

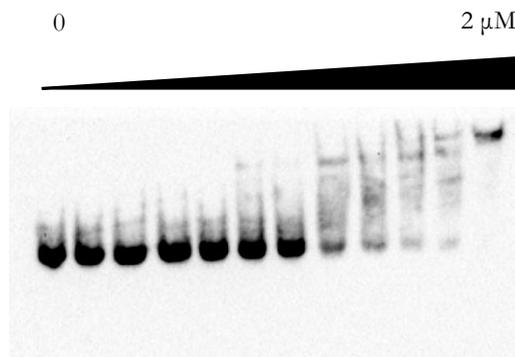


Figure 26: Example of EMSA: His₆-DrlaRP6A-ΔCTD vs. *HsCOL1A1* RNA.

Therefore, a new methodology was developed to carry out the RNA binding assays using fluorescently labeled *HsCOL1A1*, enabling direct detection of ligand. For the La Module, an interesting double shift is seen: the first shift occurs at a concentrations of 10 nM to 160 nM, with a supershift that retains a complex close to the well occurs from 1 μM to 20 μM (Fig 27A). Conversely, for the ΔCTD, a cleaner shift is observed with a $K_{D,app}$ of ~ 51 nM (Fig 27B). Both constructs appear to need less protein to bind to this ligand as compared to the homopolymers.

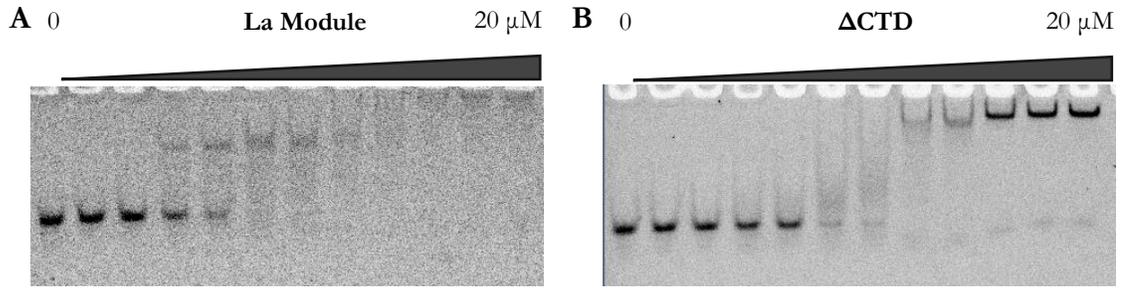


Figure 27. Fluorescent electrophoretic mobility shift assays of His₆-DrlaRP6A-La Module and His₆-DrlaRP6A-ΔCTD versus FAM-tagged *HsCOL1A1* stem loop mRNA. **A:** La Module. **B:** ΔCTD.

As expected, His₁₀-SUMO-GG also does not bind to *HsCOL1A1* (Fig 28A). Surprisingly, the NTR does appear to bind this ligand at high protein concentrations ($\sim 20 \mu\text{M}$) (Fig 28B).

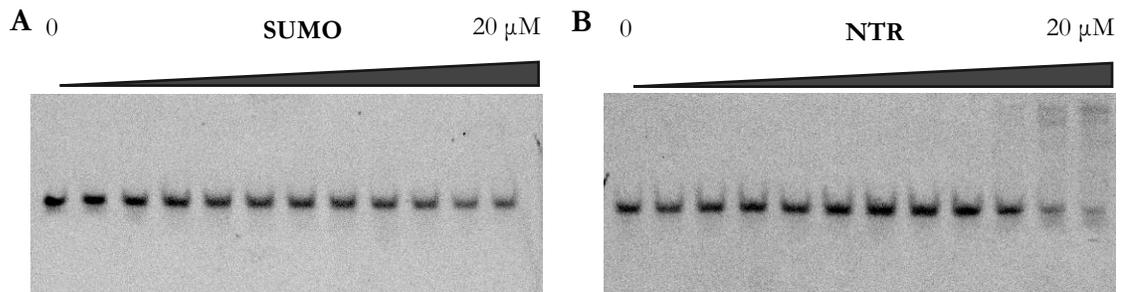


Figure 28: Fluorescent electrophoretic mobility shift assays of His₁₀-SUMO-GG and His₁₀-SUMO-DrlaRP6A-NTR versus FAM-tagged *HsCOL1A1* stem loop mRNA. **A:** SUMO. **B:** SUMO-NTR.

Apparent K_D s of Tested Ligands

To quantify binding, volume analysis tools on the Pharos FX™ Plus molecular imager were used to measure pixel intensity, after which fractional saturation was calculated (Fig 29). Plotting this data against protein concentration for each replicate ($N \geq 3$) and fitting to the simplified binding isotherm stated in the methods allowed for extrapolation of the apparent K_D ($K_{D,app}$) for each protein:ligand combination (Fig 30).

Overall, the La Module and Δ CTD bind to *HsCOL1A1* with about 35-fold and 8-fold greater affinity, respectively, than the homopolymers. The Δ CTD binds more tightly than the La Module to Poly-A and Poly-U, though the difference is not statistically significant in this work. The La Module binds more tightly to *HsCOL1A1* compared to the Δ CTD (though again, within the limit of error). The La Module also binds to Poly-C with a much weaker affinity than Poly-A or Poly-U. Since no binding was observed for the Δ CTD versus Poly-C, no $K_{D,app}$ could be determined. While the NTR binds to *HsCOL1A1*, it could not be quantified as the $K_{D,app}$ was much larger than 20 μ M.

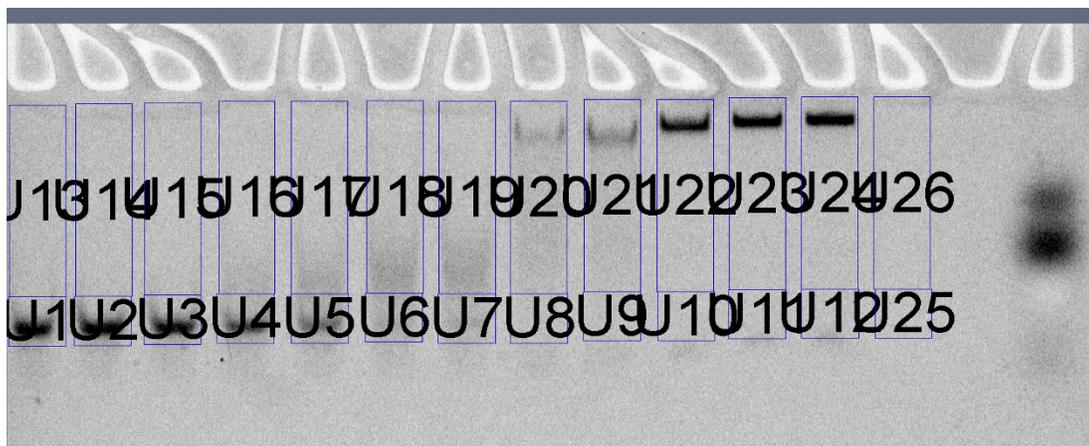


Figure 29: A representative gel quantification using the volume tool settings on the Pharos FX™ Plus molecular imager. Boxes U1 – U12 encapsulate any free ligand, while U13 – U24 capture bound species. U25 and U26 were taken to compare internal versus external background correction.

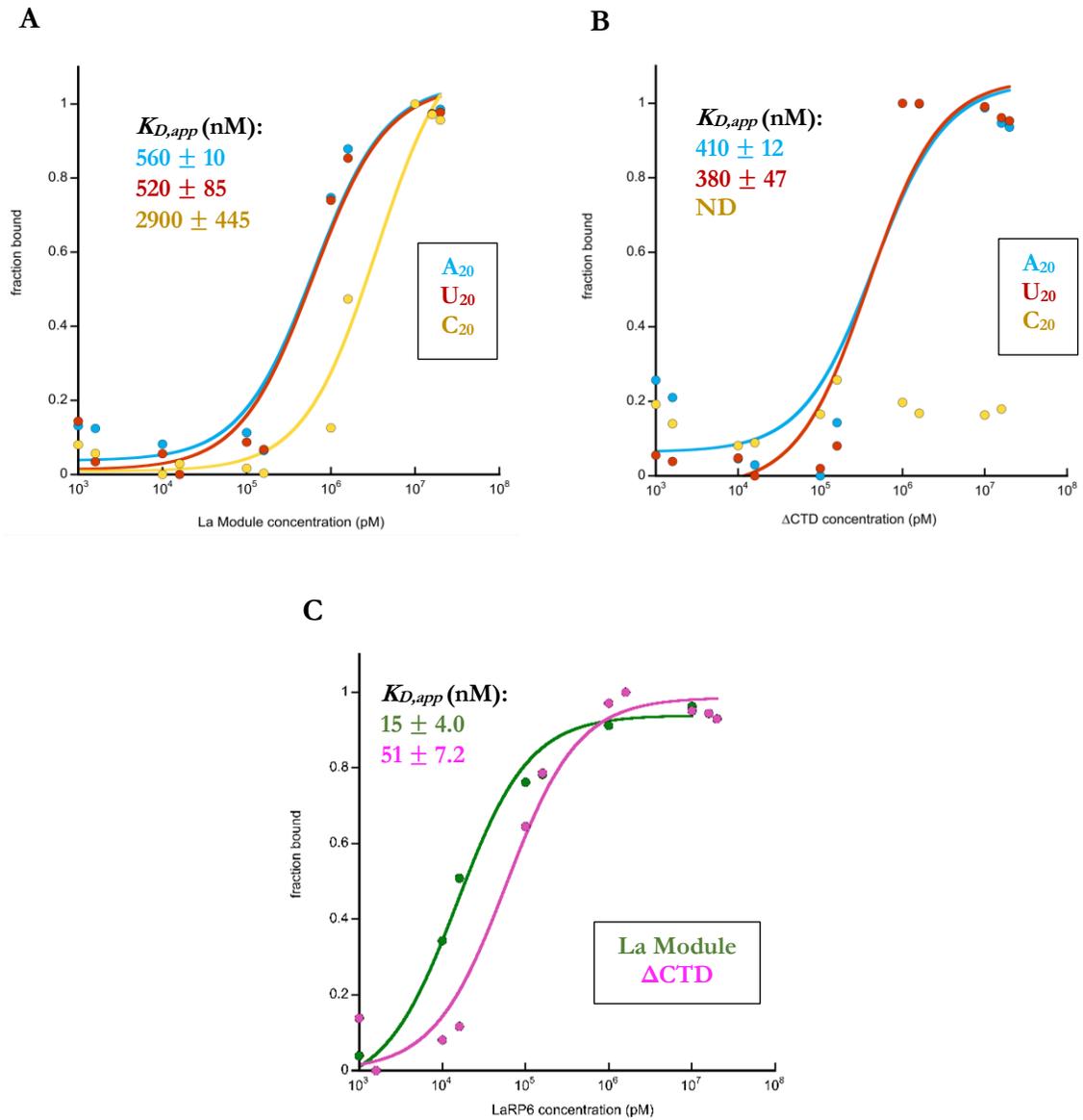


Figure 30: Quantifications of fluorescent electrophoretic mobility shift assays. A: La Module vs. homopolymers. **B:** Δ CTD vs. homopolymers. **C:** La Module and Δ CTD vs. *Hs*COL1A1.

V. CONCLUSIONS

The primary goals of this study were to: 1) Successfully clone, express and purify the isolated *D₇LaRP6A-NTR*, 2) probe its gross secondary structure and 3) test for differences in binding between the *D₇LaRP6-ΔCTD* and -La Module, as well as putative independent binding activity of the NTR.

The first major finding from this study was the inability to purify the isolated NTR, as the extinction coefficient at 280 nm becomes zero upon tag cleavage. Leaving the His₁₀-SUMO tag on allowed for the project to proceed, but required the creation of the isolated tag as a negative control for biochemical analyses.

Structurally, purifying the ΔCTD in the absence of reducing agent resulted in unwanted dimerization, as confirmed by AUC. Further analysis found this is most likely due to a cysteine residue in the NTR. The frictional ratio calculated from AUC also supports the ΔCTD is more elongated than globular. Purifying the ΔCTD and NTR resulted in a larger calculated MW_{app} than expected, which further supports this elongated nature. Probing gross secondary structure via CD spectroscopy supports the NTR being a random coil, though further structural analysis without the His₁₀-SUMO tag is needed to fully disprove the initial hypothesis of the NTR being a stably folded, globular domain.

Through various EMSAs using fluorescently labeled RNA ligands, the isolated La Module and ΔCTD each exhibited distinct binding behavior with all ligands tested (the three homopolymers and the endogenous ligand, *HsCOL1A1*). Overall, the ΔCTD appears to bind more stably as indicated by the cleaner shift. With the La Module, along with smearing and bound protein being somewhat retained in the wells, a second bound species is observed when binding *HsCOL1A1*. Additionally, the ΔCTD does not bind Poly-C. The NTR also

binds to *HsCOL1A1*, albeit too weak to accurately quantify. These differences support that the NTR plays a role in ligand binding, both in terms of binding stability and ligand specificity.

VI. FUTURE DIRECTIONS

Human LaRP6 NTR

From this work, the importance of the N-terminal region in the zebrafish LaRP6 model has prompted further questions of the NTR's role in the human protein species. The novel *HsCOL1A1* binding data particularly implicates the NTR in binding to collagen mRNA *in vivo*.

Our collaborators have identified a splicing isoform of the human LaRP6 that is highly expressed in the testes.²⁰ This form of LaRP6 contains canonical residues 1 – 66 but then differs for the following 27 amino acids (aside from one residue; highlighted in Figure 5), as they are encoded over a splice junction.⁴⁵ Because of its particular tissue localization and that the sequence variation occurs mostly in the N-terminal region, this isoform has therefore been named “tNTR.” However, the structure and function of this isoform has yet to be characterized either biochemically or physiologically.

As an addition to the zebrafish work, cloning has therefore been started for both the canonical *Hs*-NTR and *Hs*-tNTR, the splicing isoform.

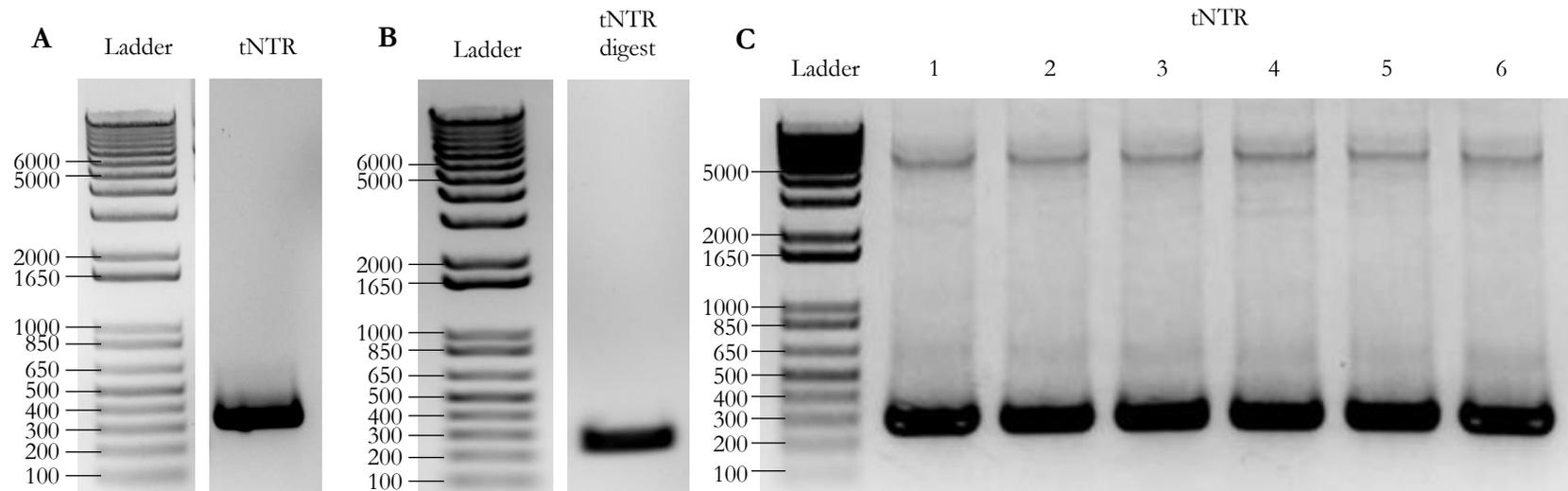


Figure 32: Cloning of His₁₀-SUMO-*HsLaRP6*-tNTR. **A: Insert amplification of *HsLaRP6*-tNTR (expected size: 304 base pairs). **B:** Double digest of tNTR insert with *Bam*HI and *Xba*I (expected size: 282 base pairs). **C:** Colony PCR of tNTR insert. All 6 samples were sent for Sanger sequencing.**

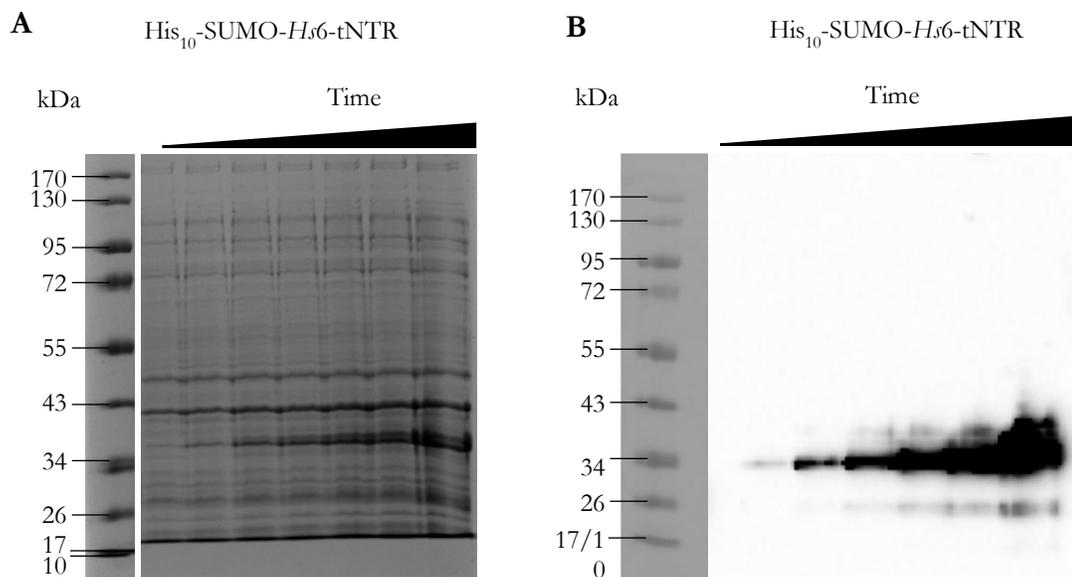
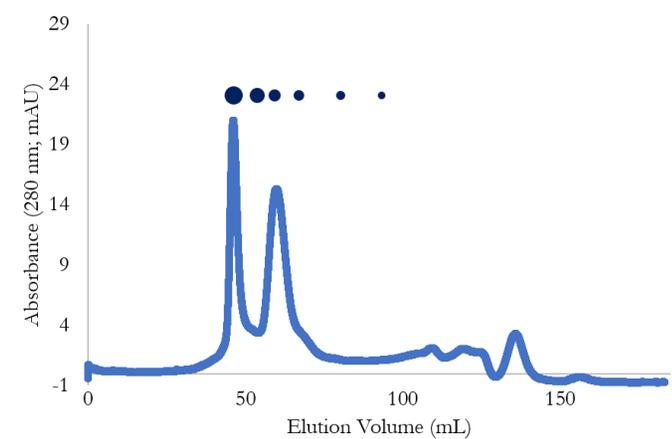
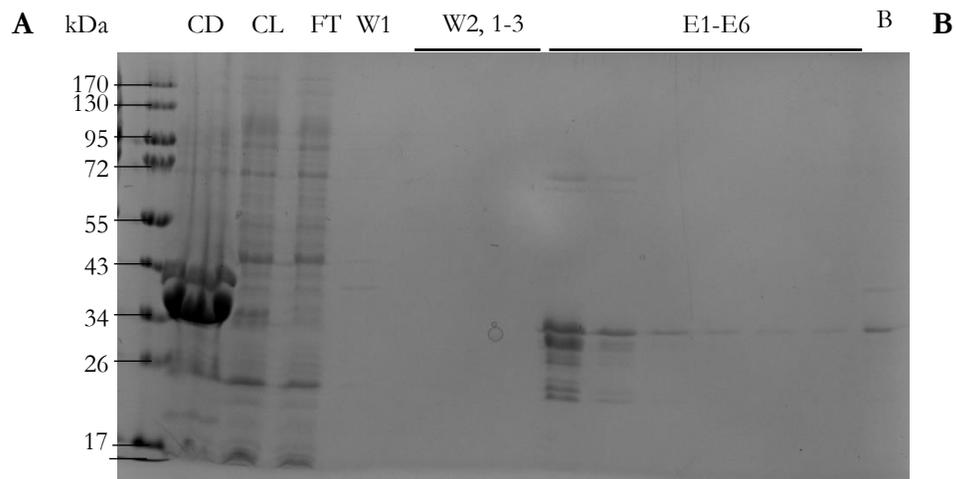


Figure 33: Expression of His₁₀-SUMO-*HsLaRP6*-tNTR. **A:** SDS-PAGE analysis of expression aliquots, stained with Coomassie blue staining. **B:** Anti-His Western blot to confirm expression of His-tagged protein.

Purification of His₁₀-SUMO-*HsLaRP6*-tNTR

A 1-L cell pellet was sonicated and purified via IMAC as previously described. All six elution fractions were pooled to proceed with size exclusion chromatography (Fig 34A). Compared to other proteins purified in this work, it can be seen that the protein largely remains in the cell debris. The purification proceeded as usual. Two peaks were observed. The suspected protein of interest eluted at 59.98 mL, with a calculated MW_{app} of ~46 kDa (Fig 34B). Fraction 17 alone was concentrated to 23 μM (~0.3 mg of protein), brought to 50% glycerol and split up into 50 μL aliquots. Half the aliquots were stored at 4 °C, and half at -70 °C (Fig 34C).



06

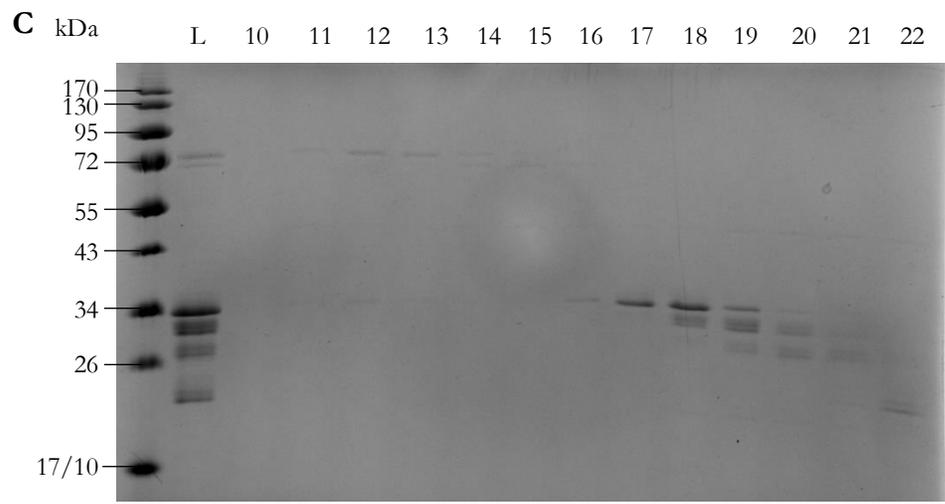


Figure 34: Pilot purification of His₁₀-SUMO-*Hs*LaRP6-tNTR. **A:** SDS-PAGE analysis of IMAC fractions. **B:** Chromatogram from size exclusion chromatography. **C:** SDS-PAGE analysis of size exclusion chromatography fractions

For the second purification, in order to test if the *Hs*-tNTR could be solubilized from the cell debris, the cell debris was incubated with urea before proceeding with additional fractionation. The cell pellet was sonicated and purified as before, and the purification proceeded as stated. However, instead of discarding the cell debris, it was resuspended in 10 mL of Lysis Buffer that was brought to 6 M urea and incubated at room temperature overnight. The slurry was centrifuged the next day and the new supernatant decanted. Both the original cell debris and lysate samples were run side-by-side the new samples (Fig 35D). By the number of bands in the soluble fraction of the urea treatment, it can clearly be seen that the urea helped to solubilize more protein from the cell debris.

From the SEC of the initially soluble fraction, two peaks were once again observed, with the average elution volume being 60.39 mL (Fig 35B). Again, all elution fractions from the Ni-NTA column were pooled to proceed to the sizing column (Fig 35A). Additionally, only fraction 17 was saved again at the end of the purification (Fig 35C). Interestingly, it was noticed that the protein began to aggregate during the second purification much sooner in the concentrating step than the first purification, and thus had to be stopped at a lower concentration of 10 μ M (\sim 0.2 mg of protein). The protein was partitioned into 100 μ L aliquots and stored at 4 °C without glycerol.

Solubilizing Potential Inclusion Bodies with Urea

During the trial purification of His₁₀-SUMO-*Hs*tNTR (isoform 2), a majority of the protein was observed to be stuck in the cell debris. During the second purification, the cell debris pellet was therefore saved and incubated overnight at room temperature in 10 mL of a urea solution (Lysis/Wash 1 Buffer brought to 6 M urea) in attempts to solubilize the protein.⁴⁶ The mixture was then centrifuged at 18,000 rcf at 22 °C for 15 minutes, and the

lysate decanted. An additional SDS-PAGE gel was run comparing the initial lysate and cell debris versus the new lysate and debris samples. It can be seen a large amount of protein was recovered from the cell debris using this method. Future work will require a refolding protocol to accompany this purification and solubilization methodology.

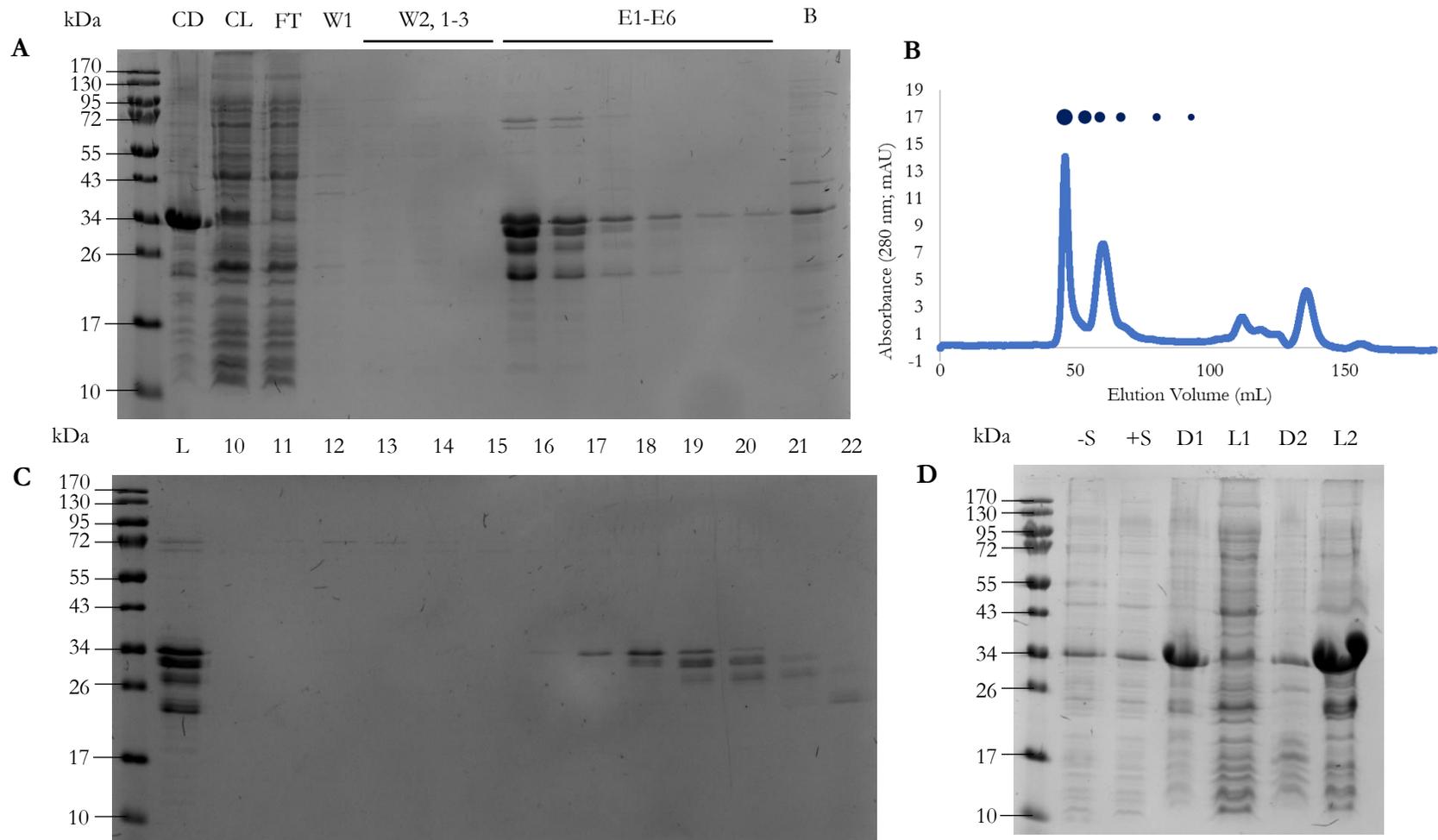


Figure 35: Second purification of His₁₀-SUMO-*HsLaRP6*-tNTR. **A:** SDS-PAGE analysis of IMAC fractions. **B:** Chromatogram from size exclusion chromatography. **C:** SDS-PAGE analysis of size exclusion chromatography fractions. **D:** Comparison of cell debris and lysate before and after urea solubilization. -S: Pre-sonication. +S: Post-sonication. D1: cell debris from first sonication. L: Cleared lysate from first round of centrifugation. D2: Cell debris after pelleting post-urea incubation. L2: Cleared lysate after centrifuging post-urea incubation.

Cloning in Progress: pET28-SUMO-HsLaRP6-NTR (Isoform 1), DrLaRP6-NTR-Trp, and DrLaRP6B Constructs

Restriction cloning was attempted for the canonical isoform of the human *DrLaRP6A*-NTR. Additionally, SDM to insert a tryptophan right after the serine in the ULP1 cleavage site in the zebrafish *LaRP6*-NTR sequence was started. The purpose of the latter construct is to eventually produce a tag-free NTR that can be monitored via 280 nm. However, all cloning attempts for both constructs to date have been unsuccessful.

Previously, cloning and expression were completed on various *DrLaRP6B* constructs. Work will continue on this paralog to probe the difference between the two zebrafish homologs.

Collaborations on various techniques will help expand our understanding of the NTR and its role in full-length vertebrate *LaRP6* protein. These techniques include ITC to obtain true K_{DS} , as well as SAXS and NMR for structural data.

APPENDIX SECTION

Table 19: Polymerase Chain Reaction Reagents and Protocols.

Construct	Template DNA	Reagents (for a 50 μ L reaction; in order of addition)	Parameters
<p>pET28-SUMO-<i>Drla</i>RP6ALaMod, pET28-SUMO-<i>Drla</i>RP6AΔCTD</p>	<p>pET28-SUMO-<i>Drla</i>RP6A, FL (from Melissa Carrizales)</p>	<ul style="list-style-type: none"> • 31.5 μL IDT H₂O • 1\times Phusion® HF Buffer • 0.2 mM dNTPs • 0.4 mM FWD/REV primer • 448 ng template DNA • 1 U Thermo Scientific™ Phusion® High-Fidelity DNA Polymerase 	<p>98 °C – 1 min 30 cycles of:</p> <ul style="list-style-type: none"> • 98 °C – 10 sec • 60 °C – 1 min • 72 °C – 1 min <p style="padding-left: 40px;">30 sec</p> <p>72 °C – 6 min 30 sec Hold at 4 °C</p>

pET28-SUMO- <i>Drl</i> LaRP6ANTR	pET28-SUMO- <i>Drl</i> LaRP6A, FL (from Melissa Carrizales)	<ul style="list-style-type: none"> • 33.5 μL IDT H₂O • 1\times Phusion® HF Buffer • 0.2 mM dNTPs • 0.4 mM FWD/REV primer • 281.175 ng template DNA • 1 U Thermo Scientific™ Phusion® High-Fidelity DNA Polymerase 	98 °C – 1 min 30 cycles of: <ul style="list-style-type: none"> • 98 °C – 10 sec • 65 °C – 1 min • 72 °C – 1 min 30 sec 72 °C – 6 min 30 sec Hold at 4 °C
pET28-SUMO-GG	pET28-SUMO, FL (from Leticia Gonzalez)	<ul style="list-style-type: none"> • 36.5 μL IDT H₂O • 1\times Phusion® HF Buffer • 0.2 mM dNTPs • 0.32 mM FWD/REV primer • 31.74 ng template DNA 1 U Thermo Scientific™ Phusion® High-Fidelity DNA Polymerase	94 °C – 30 sec 20 cycles of: <ul style="list-style-type: none"> • 94 °C – 30 sec • 62 °C – 45 sec • 72 °C – 5 min 42 sec 72 °C – 7 min 42 sec Hold at 4 °C

pET28-SUMO- <i>HsLaRP6tNTR</i> (Isoform 2)	gBlocks® Gene Fragment – IDT	<ul style="list-style-type: none"> • 31.5 µL IDT H₂O • 1× Phusion® HF Buffer • 0.2 mM dNTPs • 0.4 mM FWD/REV primer • 50 ng template DNA • 1 U Thermo Scientific™ Phusion® High-Fidelity DNA Polymerase 	98 °C – 1 min 10 cycles of: <ul style="list-style-type: none"> • 98 °C – 10 sec • 60 °C – 1 min • 72 °C – 1 min 30 sec 72 °C – 6 min 30 sec Hold at 4 °C
pET28-SUMO- <i>HsLaRP6NTR</i> (Isoform 1)*	pET28-SUMO- <i>HsLaRP6ΔBamHI</i> (from Leticia Gonzalez)	Need to be optimized	Need to be optimized
pET28-SUMO- <i>DrlaRP6NTR</i> - Trp*	pET28-SUMO- <i>DrlaRP6ANTR</i> (from this work)	Need to be optimized	Need to be optimized

†Cloned by previous lab members; purified proteins used for biochemical analyses

*Cloning still in progress

REFERENCES

1. Beckmann, B. M.; Castello, A.; Medenbach, J. The Expanding Universe of Ribonucleoproteins: Of Novel RNA-Binding Proteins and Unconventional Interactions. *Pflügers Arch. Eur. J. Physiol.* **2016**, *468* (6), 1029–1040. <https://doi.org/10.1007/s00424-016-1819-4>.
2. Gerstberger, S.; Hafner, M.; Tuschl, T. A Census of Human RNA-Binding Proteins. *Nat Rev Genet* **2014**, *15* (12), 829–845.
3. Maris, C.; Dominguez, C.; Allain, F. H. T. The RNA Recognition Motif, a Plastic RNA-Binding Platform to Regulate Post-Transcriptional Gene Expression. *FEBS J.* **2005**, *272* (9), 2118–2131. <https://doi.org/10.1111/j.1742-4658.2005.04653.x>.
4. Cléry, A.; Blatter, M.; Allain, F. H. T. RNA Recognition Motifs: Boring? Not Quite. *Curr. Opin. Struct. Biol.* **2008**, *18* (3), 290–298. <https://doi.org/10.1016/j.sbi.2008.04.002>.
5. Ding, J.; Hayashi, M. K.; Zhang, Y.; Manche, L.; Krainer, A. R.; Xu, R. M. Crystal Structure of the Two-RRM Domain of HnRNP A1 (UP1) Complexed with Single-Stranded Telomeric DNA. *Genes Dev.* **1999**, *13* (9), 1102–1115. <https://doi.org/10.1101/gad.13.9.1102>.
6. Yang, R.; Gaidamakov, S. A.; Xie, J.; Lee, J.; Martino, L.; Kozlov, G.; Crawford, A. K.; Russo, A. N.; Conte, M. R.; Gehring, K.; Maraia, R. J. La-Related Protein 4 Binds Poly(A), Interacts with the Poly(A)-Binding Protein MLE Domain via a Variant PAM2w Motif, and Can Promote mRNA Stability. *Mol. Cell. Biol.* **2011**, *31* (3), 542–556. <https://doi.org/10.1128/mcb.01162-10>.

7. Nykamp, K.; Lee, M.; Kimble, J. To Germline P Bodies and Attenuates Ras-MAPK Signaling during Oogenesis. **2008**, 1378–1389.
<https://doi.org/10.1261/rna.1066008.recognition>.
8. Mattioli, M.; Reichlin, M. Heterogeneity of RNA Protein Antigens Reactive with Sera of Patients with Systemic Lupus Erythematosus Description of a Cytoplasmic Nonribosomal Antigen. *Arthritis Rheum.* **1974**, *17* (4), 421–429.
<https://doi.org/10.1002/art.1780170413>.
9. Reichlin, M. Current Perspectives on Serological Reactions in SLE Patients. *Clin. Exp. Immunol.* **1981**, *44* (1), 1–10.
10. Bayfield, M. A.; Yang, R.; Maraia, R. J. Conserved and Divergent Features of the Structure and Function of La and La-Related Proteins (LARPs). *Biochim. Biophys. Acta - Gene Regul. Mech.* **2010**, *1799* (5–6), 365–378.
<https://doi.org/10.1016/j.bbagr.2010.01.011>.
11. Wolin, S. L.; Cedervall, T. The La Protein. *Annu. Rev. Biochem.* **2002**.
<https://doi.org/10.1146/annurev.biochem.71.090501.150003>.
12. Bousquet-Antonelli, C.; Deragon, J. M. A Comprehensive Analysis of the La-Motif Protein Superfamily. *Rna* **2009**, *15* (5), 750–764. <https://doi.org/10.1261/rna.1478709>.
13. Birney, E.; Kumar, S.; Krainer, A. R. Analysis of the RNA-Recognition Motif and RS and RGG Domains: Conservation in Metazoan Pre-mRNA Splicing Factors. *Nucleic Acids Res.* **1993**, *21* (25), 5803–5816. <https://doi.org/10.1093/nar/21.25.5803>.
14. Stavrouka, C.; Blagden, S. The La-Related Proteins, a Family with Connections to Cancer. *Biomolecules* **2015**, *5* (4), 2701–2722. <https://doi.org/10.3390/biom5042701>.

15. Lahr, R. M.; Mack, S. M.; Héroux, A.; Blagden, S. P.; Bousquet-Antonelli, C.; Deragon, J. M.; Berman, A. J. The La-Related Protein 1-Specific Domain Repurposes HEAT-like Repeats to Directly Bind a 5'TOP Sequence. *Nucleic Acids Res.* **2015**, *43* (16), 8077–8088. <https://doi.org/10.1093/nar/gkv748>.
16. Al-Ashtal, H. A.; Rubottom, C. M.; Leeper, T. C.; Berman, A. J. The LARP1 La-Module Recognizes Both Ends of TOP MRNAs. *RNA Biol.* **2019**, *00* (00), 1–11. <https://doi.org/10.1080/15476286.2019.1669404>.
17. Martino, L.; Pennell, S.; Kelly, G.; Busi, B.; Brown, P.; Atkinson, R. A.; Salisbury, N. J. H.; Ooi, Z. H.; See, K. W.; Smerdon, S. J.; Alfano, C.; Bui, T. T. T.; Conte, M. R. Synergic Interplay of the La Motif, RRM1 and the Interdomain Linker of LARP6 in the Recognition of Collagen mRNA Expands the RNA Binding Repertoire of the La Module. *Nucleic Acids Res.* **2015**, *43* (1), 645–660. <https://doi.org/10.1093/nar/gku1287>.
18. Cruz-Gallardo, I.; Martino, L.; Kelly, G.; Andrew Atkinson, R.; Trotta, R.; De Tito, S.; Coleman, P.; Ahdash, Z.; Gu, Y.; Bui, T. T. T.; Conte, M. R. LARP4A Recognizes PolyA RNA via a Novel Binding Mechanism Mediated by Disordered Regions and Involving the PAM2w Motif, Revealing Interplay between PABP, LARP4A and MRNA. *Nucleic Acids Res.* **2019**, *47* (8), 4272–4291. <https://doi.org/10.1093/nar/gkz144>.
19. Eichhorn, C. D.; Yang, Y.; Repeta, L.; Feigon, J. Structural Basis for Recognition of Human 7SK Long Noncoding RNA by the La-Related Protein Larp7. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (28), E6457–E6466. <https://doi.org/10.1073/pnas.1806276115>.
20. Valavanis, C.; Wang, Z.; Sun, D.; Vaine, M.; Schwartz, L. M. 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* **2007**, *393* (1–2), 101–109. <https://doi.org/10.1016/j.gene.2007.01.033>.

21. Maraia, R. J.; Mattijssen, S.; Cruz-Gallardo, I.; Conte, M. R. The La and Related RNA-Binding Proteins (LARPs): Structures, Functions, and Evolving Perspectives. *Wiley Interdiscip. Rev. RNA* **2017**, *8* (6). <https://doi.org/10.1002/wrna.1430>.
22. Zhang, Y.; Stefanovic, B. LARP6 Meets Collagen mRNA: Specific Regulation of Type I Collagen Expression. *Int. J. Mol. Sci.* **2016**, *17* (3).
<https://doi.org/10.3390/ijms17030419>.
23. Hussain, R. H.; Zawawi, M.; Bayfield, M. A. Conservation of RNA Chaperone Activity of the Human La-Related Proteins 4, 6 and 7. *Nucleic Acids Res.* **2013**, *41* (18), 8715–8725.
<https://doi.org/10.1093/nar/gkt649>.
24. Ricard-Blum, S. The Collagen Family. *Cold Spring Harb. Perspect. Biol.* **2011**, *3* (1), 1–19.
<https://doi.org/10.1101/cshperspect.a004978>.
25. Wynn, T. A. Fibrotic Disease and the TH1/TH2 Paradigm. *Nat. Rev. Immunol.* **2004**, *4* (8), 583–594. <https://doi.org/10.1038/nri1412>.
26. Cai, L.; Fritz, D.; Stefanovic, L.; Stefanovic, B. Binding of LARP6 to the Conserved 5' Stem-Loop Regulates Translation of MRNAs Encoding Type I Collagen. *J. Mol. Biol.* **2010**, *395* (2), 309–326. <https://doi.org/10.1016/j.jmb.2009.11.020>.
27. Stefanovic, L.; Longo, L.; Zhang, Y.; Stefanovic, B. Characterization of Binding of LARP6 to the 5' Stem-Loop of Collagen MRNAs: Implications for Synthesis of Type I Collagen. *RNA Biol.* **2014**, *11* (11), 1386–1401.
<https://doi.org/10.1080/15476286.2014.996467>.
28. Stefanovic, B.; Manojlovic, Z.; Vied, C.; Badger, C. D.; Stefanovic, L. Discovery and Evaluation of Inhibitor of LARP6 as Specific Antifibrotic Compound. *Sci. Rep.* **2019**, *9* (1), 1–15. <https://doi.org/10.1038/s41598-018-36841-y>.

29. Carrizales, M. Divergence of the RNA Recognition Motif in Vertebrate LaRP6 Proteins (Master's Thesis). *Dep. Chem. Biochem. Grad. Coll. Sci. Eng. Texas State Univ.* **2019**.
30. Lunde, B. M.; Moore, C.; Varani, G. RNA-Binding Proteins: Modular Design for Efficient Function. *Nat. Rev. Mol. Cell Biol.* **2007**, *8* (6), 479–490.
<https://doi.org/10.1038/nrm2178>.
31. O'Connor, C. M.; Collins, K. A Novel RNA Binding Domain in Tetrahymena Telomerase P65 Initiates Hierarchical Assembly of Telomerase Holoenzyme. *Mol. Cell. Biol.* **2006**, *26* (6), 2029–2036. <https://doi.org/10.1128/mcb.26.6.2029-2036.2006>.
32. Castro, J. M.; Horn, D. A.; Pu, X.; Lewis, K. A. Recombinant Expression and Purification of the RNA-Binding LARP6 Proteins from Fish Genetic Model Organisms. *Protein Expr. Purif.* **2017**, *134*, 147–153. <https://doi.org/10.1016/j.pep.2017.04.004>.
33. Scharl, M. Beyond the Zebrafish: Diverse Fish Species for Modeling Human Disease. *DMM Dis. Model. Mech.* **2014**, *7* (2), 181–192. <https://doi.org/10.1242/dmm.012245>.
34. Castro, J. M. Biochemical Analysis of Evolutionary Divergent Vertebrate LARP6 Species, Texas State University, 2017.
35. Külköylüoğlu, H. Characterization of the N-Terminal Domain in the RNA-Binding LaRP6 Proteins from Fish (Master's Thesis). *Grad. Sch. Nat. Appl. Sci. Dep. Biol. T. C. Ondokuz Mayıs Univ.* **2018**.
36. Scitable. Definition: poly-a tail. <https://www.nature.com/scitable/definition/poly-a-tail-276/>.
37. Li, S.; Xu, Z.; Sheng, J. tRNA-Derived Small RNA: A Novel Regulatory Small Non-Coding RNA. *Genes (Basel)*. **2018**, *9* (5). <https://doi.org/10.3390/genes9050246>.
38. Snoussi, K.; Nonin-Lecomte, S.; Leroy, J. L. The RNA I-Motif. *J. Mol. Biol.* **2001**, *309* (1), 139–153. <https://doi.org/10.1006/jmbi.2001.4618>.

39. UniProt. Serum albumin <https://www.uniprot.org/uniprot/P02769>.
40. Haan, C.; Behrmann, I. A Cost Effective Non-Commercial ECL-Solution for Western Blot Detections Yielding Strong Signals and Low Background. *J. Immunol. Methods* **2007**, *318* (1–2), 11–19. <https://doi.org/10.1016/j.jim.2006.07.027>.
41. Altschuler, S. E.; Lewis, K. A.; Wuttke, D. S. Practical Strategies for the Evaluation of High-Affinity Protein/Nucleic Acid Interactions. **2014**, *4* (1), 19–28. <https://doi.org/10.4081/jnai.2013.e3.Practical>.
42. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. *Proteomics Protoc. Handb.* **2005**, 571–608. <https://doi.org/10.1385/1592598900>.
43. Anthis, N. J.; Clore, G. M. Sequence-Specific Determination of Protein and Peptide Concentrations by Absorbance at 205 Nm. *Protein Sci.* **2013**, *22* (6), 851–858. <https://doi.org/10.1002/pro.2253>.
44. Guan, Y.; Zhu, Q.; Huang, D.; Zhao, S.; Jan Lo, L.; Peng, J. An Equation to Estimate the Difference between Theoretically Predicted and SDS PAGE-Displayed Molecular Weights for an Acidic Peptide. *Sci. Rep.* **2015**, *5* (August), 1–11. <https://doi.org/10.1038/srep13370>.
45. UniProt. LaRP6 Isoform 2.
46. Jevševar, S.; Gaberc-Porekar, V.; Fonda, I.; Podobnik, B.; Grdadolnik, J.; Menart, V. Production of Nonclassical Inclusion Bodies from Which Correctly Folded Protein Can Be Extracted. *Biotechnol. Prog.* **2005**, *21* (2), 632–639. <https://doi.org/10.1021/bp0497839>.