

THE N-TERMINAL REGION OF THE *ARABIDOPSIS THALIANA* LARP6C  
PROTEIN CONTRIBUTES TO STRUCTURAL STABILITY

by

Julia E. Roberts, B.S.

A thesis submitted to the Graduate Council of  
Texas State University in partial fulfillment  
of the requirements for the degree of  
Master of Science  
with a Major in Biochemistry  
May 2020

Committee Members:

Karen A. Lewis, Chair

Sean M. Kerwin

Steven T. Whitten

**COPYRIGHT**

by

Julia E. Roberts

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, Julia E. Roberts, authorize duplication of this work, in whole or in part, for educational or scholarly purposes only.

## **ACKNOWLEDGMENTS**

As I am now in the final stages of writing, I can only think about the tremendous amount of dedication and time that went into this project. However, it was not just my own, I had the immense support of my family, friends, and mentors. First, I would like to thank my parents, Mark and Shizue, I could not have done this without your love and encouragement, which I felt constantly from thousands of miles away. To my sister Emma, I am so thankful for your unwavering encouragement which lifted me up when I needed it most.

To my dear friends in both my undergraduate and graduate cohorts, I cannot imagine how I would have made it through this degree without you. Alejandro Oviedo, thank you for all the coffee/anxiety-fueled study sessions at Mochas. Elisia “Ari” Paiz, thank you for your positive energy and scientific incite at both Gibbs and Boulder. Kathryn Banks, I will miss our hallway chats and the results of your ever creative and delicious stress-baking. Jesse Durham and Amy Baker, you both have been with me since undergrad and I am so excited to see what we accomplish.

To my fellow LewisKAN and most decidedly my greatest ally in this journey, Brianna Godinez. In this short time, we have both grown so much as people, scientists, and as friends. From teaching labs, conducting research, the late nights, and long talks, you have been there with me through it all. I am so

sad we will not be seeing each other as often, but know I am so thankful and appreciative of your friendship. Leticia “Lety” Gonzalez, you will always be my lab mom. You provided so much support and advice as I reached new goals, but you were there doubly so when I missed the mark. I will always cherish our time together, and I will take everything you taught me as I move forward in life.

Finally, and most importantly, none of this would have been possible without the efforts and mentorship of my PI, Dr. Karen A. Lewis. As an undergrad with no research experience, you saw potential and believed in me. You inspired me to pursue my goals and to keep reaching upwards and onwards. Thank you for pushing me to be greater than I ever thought possible.

## TABLE OF CONTENTS

	<b>Page</b>
ACKNOWLEDGMENTS .....	iv
LIST OF TABLES .....	vii
LIST OF FIGURES .....	viii
ABSTRACT .....	xi
CHAPTER	
I. INTRODUCTION .....	1
II. MATERIALS AND METHODS .....	8
III. RECOMBINANT EXPRESSION AND PURIFICATION OF <i>AtLaRP6C</i> PROTEIN VARIANTS .....	19
IV. STRUCTURE AND FUNCTION OF <i>AtLaRP6C</i> VARIANTS .....	61
V. CONCLUSIONS AND FUTURE DIRECTIONS .....	82
APPENDIX SECTION .....	84
REFERENCES .....	88

## LIST OF TABLES

<b>Table</b>	<b>Page</b>
1. Molecular weight standards for S200 column .....	14
2. Molecular weight standards for S75 column .....	14
3. Calculated stability parameters for lysozyme.....	67
4. Calculated stability parameters for <i>AtLaRP6C</i> proteins.....	72

## LIST OF FIGURES

Figure	Page
1. Domain topologies of LaRP families.....	2
2. The three <i>AtLaRP6</i> paralogs .....	4
3. <i>AtLaRP6C</i> constructs used in this project.....	19
4. Expression trials of frameshifted <i>AtLaRP6C</i> FL.....	21
5. Expression trials of resolved <i>AtLaRP6C</i> FL.....	22
6. Expression trials of <i>AtLaRP6C</i> $\Delta$ CTD and $\Delta$ PAM2 $\Delta$ CTD .....	24
7. Expression trials of <i>AtLaRP6C</i> NTR.....	25
8. Expression trials of <i>AtLaRP6C</i> $\Delta$ PAM2 NTR.....	26
9. Expression trials of <i>AtLaRP6C</i> $\Delta$ LSA .....	27
10. Purification of <i>AtLaRP6C</i> Full-length .....	31
11. Purification of <i>AtLaRP6C</i> $\Delta$ PAM2.....	34
12. Purification of His <sub>10</sub> -SUMO- <i>AtLaRP6C</i> La Module .....	37
13. Purification of <i>AtLaRP6C</i> La Module .....	38
14. Purification of His <sub>10</sub> -SUMO- <i>AtLaRP6C</i> $\Delta$ CTD .....	41
15. Purification of <i>AtLaRP6C</i> $\Delta$ CTD .....	42

16. Purification of His <sub>10</sub> -SUMO- <i>At</i> LaRP6C ΔPAM2 ΔCTD.....	45
17. Purification of <i>At</i> LaRP6C ΔPAM2 ΔCTD .....	46
18. Purification of His <sub>10</sub> -SUMO- <i>At</i> LaRP6C NTR.....	49
19. Purification of <i>At</i> LaRP6C NTR.....	50
20. Purification of His <sub>10</sub> -SUMO- <i>At</i> LaRP6C ΔPAM2 NTR.....	53
21. Purification of <i>At</i> LaRP6C ΔPAM2 NTR.....	54
22. Checking for presence of His <sub>10</sub> -SUMO tag in purified proteins .....	56
23. <i>At</i> LaRP6C La Module vs 2 – 7.25 M Guanidinium Hydrochloride at 25°C....	62
24. <i>At</i> LaRP6C La Module vs 1 – 7.8 M Guanidinium Hydrochloride with 30°C ..	63
25. Location of tryptophan in <i>At</i> LaRP6C La Module.....	65
26. Thermofluor assay of lysozyme .....	67
27. Thermofluor assay of <i>At</i> LaRP6C proteins .....	69
28. Normalized data from thermofluor assay of <i>At</i> LaRP6C proteins .....	70
29. Comparison of <i>At</i> LaRP6C protein T <sub>m</sub> obtained from thermal shift assay.....	72
30. Biotinylated RNA in gel shift assay .....	74
31. Detection limit of SYBR Gold using poly(A) RNA .....	75

32. Detection limit of 5'-FAM poly(A) RNA.....	76
33. <i>At</i> LaRP6C La Module vs 2 nM 5'-FAM CB1 .....	77
34. <i>At</i> LaRP6C La Mod and $\Delta$ CTD vs 0.5 nM CB1, no BSA .....	78

## ABSTRACT

The La-related proteins (LaRPs) are a superfamily of RNA-binding proteins that are distinguished by a core RNA binding domain called the “La Module”. This family is highly conserved across eukaryotes and exerts diverse functions in RNA processing and function. The genetic model vascular plant, *Arabidopsis thaliana* (*At*), has three paralogs of LaRP6, denoted “A”, “B”, and “C”. Of these paralogs, *AtLaRP6B* and *AtLaRP6C* are more closely related to each other as both have an N-terminal sequence motif known to associate with other RNA binding proteins, the “PAM2w” motif. Previous work evaluated the RNA binding activity of the isolated *AtLaRP6C* La Module. Recent work on vertebrate LaRP4A suggests that N-terminal PAM2 motifs are important for higher-order assembly of regulatory complexes. We have generated a set of recombinant constructs deleting the C-terminal domain (CTD) to test the role of both the N-terminal region (NTR) and PAM2w motif in RNA binding activity. These protein variants are stably expressed and have been highly purified for use in electrophoretic mobility shift assays (EMSAs) to measure RNA binding activity. Although binding does occur, quantifiable data has not been obtained by this methodology. However, the effect of these domains on melting temperature ( $T_m$ ), a stability parameter, have been obtained. Deletion of the PAM2w does not affect these parameters, whereas deletion of the CTD greatly impacts these values.

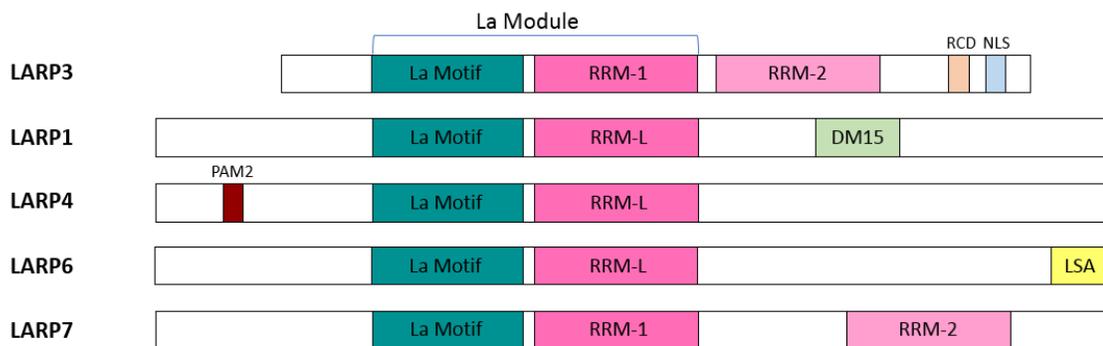
## I. INTRODUCTION

### RNA Binding Proteins

The Central Dogma is the simplified model of the flow of genetic information from DNA to RNA to proteins.<sup>1</sup> In this model, the RNA is specifically messenger RNAs (mRNAs), which are directly transcribed from DNA and carry the genetic information required for the expression of genes. The regulation of gene expression is modulated at various levels, but nearly all post-transcriptional regulation is governed by the action of RNA binding proteins (RBPs). Post-transcriptional modifications can affect localization, turnover, transport, and modifications of RNAs.<sup>2</sup> Though the functionality of these proteins is highly diverse and specific, there are few structural modules required for RNA binding. To bind the ligand, RBPs use RNA binding domains (RBDs), which include zinc fingers, KH, RGG box, PAZ, and the RNA recognition motif (RRM).<sup>2</sup> The ligand binding specificity of these proteins can be conferred by the arrangement of multiple RBDs, whether made of the same domain or a combination of different domains.<sup>3</sup> The linker between domains can also affect binding affinity and specificity depending upon the length, content, and flexibility of the linker itself.<sup>3</sup> The different structural combinations create diverse binding surfaces which define specificity and can further regulate targets of other enzymatic domains.<sup>3</sup> Of the RBDs the RRM is the most abundant in vertebrate proteins and is structurally well characterized.<sup>4</sup> The RRM is typically between 80-90 amino acids in length and consists of a four-stranded anti-parallel beta-sheet and two alpha helices, which form a canonical  $\beta\alpha\beta\alpha\beta$  topology.<sup>4</sup>

## La Related Proteins (LaRPs)

The La-related proteins (LaRPs) are a superfamily of RBPs that are distinguished by a bipartite “La Module” RNA binding domain, comprised of a highly conserved La motif (“LaM”) and an RRM/RRM-like (RRM-L) domain.<sup>5</sup> This family is maintained across eukaryotes and is separated into major families consisting of LaRP1, LaRP4, LaRP6 (originally called Acheron), LARP7, and LaRP3 (also called Genuine La and SS-B).<sup>5</sup> LaRPs have diverse functions and ligands, attributable to variations within the La Module as well as additional domains characteristic of each LaRP family (Figure 1).<sup>6</sup>



**Figure 1: Domain topologies of LaRP families.** LaRPs have a conserved La Module consisting of a La motif (LaM) and an RNA Recognition Motif (RRM). Each family is characterized by separate domains, which may correlate to different functionality. Abbreviations: RNA recognition-like motif (RRM-L); RNA chaperon domain (RCD); nuclear localization signal (NLS); Poly(A) binding protein motif (PAM2); DM15-repeat containing region (DM15); La and S1 associated motif (LSA). (Modified from Stavraka and Blagden, 2015)<sup>7</sup>

Of all the LaRP families, the least studied is LaRP6. The LaRP6 proteins are characterized by a conserved C-terminal La and S1-associated (LSA) motif, though the function of the LSA is currently unknown.<sup>6</sup> However, cold-shock response protein 1 (CSP1) also contains a similarly conserved region and there

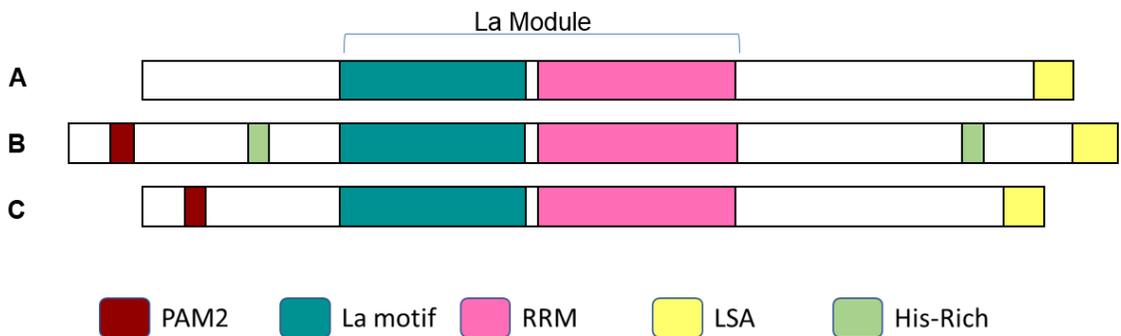
are indications that the LSA may support the selective binding of CSP1 and LaRP6 to other protein binding partners.<sup>6</sup> In humans, the LaRP6 protein (*HsLaRP6*) binds to a stem-loop structure in the 5' untranslated region of type 1 collagen mRNAs.<sup>8</sup> This interaction is important for the upregulation of collagen type I synthesis.<sup>8</sup> Collagen is critical for normal growth and development in vertebrates, although excessive or insufficient production can lead to fibroproliferative disorders such as pulmonary fibrosis, systemic sclerosis, or osteogenesis imperfecta.<sup>9</sup>

### ***Arabidopsis thaliana* LaRP6**

Plant genomes contain three paralogs of LaRP6, denoted “A”, “B”, and “C”.<sup>10</sup> In the genetic model plant *Arabidopsis thaliana* (*At*), *AtLaRP6A* is ubiquitously expressed in all tissues, while *AtLaRP6B* and *AtLaRP6C* are expressed in mutually exclusive patterns.<sup>10</sup> *AtLaRP6B* is expressed everywhere except pollen, while *AtLaRP6C* is expressed only in pollen.<sup>10</sup> *AtLaRP6A* is most closely related to human LaRP6 (“*HsLaRP6*”) by sequence similarity, whereas *AtLaRP6B* and *AtLaRP6C* are more closely related to each other due to a common N-terminal “PAM2” sequence motif that binds poly-A binding proteins (PABP) (Figure 2).<sup>10</sup> The substrate of PABP are the 3' poly(A) tails of mRNA which interact with the eukaryotic initiation factor-4E (EIF4E) which further associates with eIF4G, a binding factor for other translation factors.<sup>5</sup> The PAM2 motifs of *AtLaRP6B* and *AtLaRP6C* were found to bind the mademoiselle (MLLE) motif of the *AtPAB2* protein (i.e., the PABP in *A. thaliana*).<sup>10</sup> Together, these data

suggest that *AtLaRP6B* and *AtLaRP6C* binding may stabilize polyadenylated RNA, in concert with other translation-related RBPs.

Notably, LaRP6 proteins from non-plant species do not contain the PAM2 sequence; in those organisms, the PAM2 motif is found in LaRP4. In these species, LaRP4 has been found to bind PABP and localize to stress granules and can regulate translation of mRNA.<sup>10</sup> Additionally, recent work has shown the NTR of vertebrate LaRP4 binds poly-A RNA independently from the La Module.<sup>11</sup> However, as LaRP4 is not found in plants and vertebrate LaRP6 does not contain the PAM2 sequence, the plant LaRP6B and LaRP6C proteins may be the result of a necessary neofunctionalization of plant LaRP6 paralogs to carry out a similar function and ligand binding mechanism as the vertebrate LaRP4.<sup>10</sup>



**Figure 2: The three *AtLaRP6* paralogs.** *AtLaRP6A* is most closely related to human LaRP6 by sequence similarity. *AtLaRP6B* and *AtLaRP6C* are more closely related to each other due to an N-terminal PAM2 sequence. *AtLaRP6B* also contains polyhistidine regions which are not found in the other paralogs.

To determine the cellular localization of the *AtLaRP6* proteins, they were transiently expressed in onion epidermal cells as recombinant fusions to fluorescent proteins.<sup>10</sup> When visualized with fluorescence confocal microscopy, the proteins were localized to the nucleolus. Under hypoxic conditions,

*AtLaRP6B* and *AtLaRP6C* form subcytoplasmic aggregates, which are likely to be stress granules.<sup>10</sup> These stress granules are speculated to be involved in the storage, stability, degradation, or translation of mRNAs during conditions of increased environmental stress.<sup>12</sup> As *AtLaRP6A* does not form these stress granules, the PAM2 motif in *AtLaRP6B* and *AtLaRP6C* may contribute to their formation. It was also hypothesized that the acquisition of the PAM2 sequence resulted in a structural change of the La Module, resulting in a change of RNA binding specificity compared to *AtLaRP6A*. The isolated La Modules of *AtLaRP6A* and *AtLaRP6C* were assessed for RNA binding affinity against single-stranded 20-nt homopolymeric RNAs (polyA, polyC, polyU, and polyG). The *AtLaRP6A* La Module bound to all homopolymers tested with a preference for poly(A) RNA. In contrast, the *AtLaRP6C* La Module was found to only significantly bind to poly(U) RNA. Together, these data further suggest that the acquisition of the PAM2 sequence did change the RNA binding specificity of the protein.<sup>10</sup>

Since this work, there has been some new findings in the field with respect to *AtLaRP6C*. Unpublished work by our collaborator, Dr. Cécile Bousquet-Antonelli, identified two potential consensus sequences for ligands of full-length *AtLaRP6C*. One set of ligands was U-rich, and the other was A-rich (C.B-A., personal communication). The U-rich consensus sequence is consistent with the ligand specificity of the isolated La Module described above.<sup>10</sup> However, the A-rich sequence is inconsistent with the established specificity profile of the isolated La Module, which does not show any binding activity for poly(A) RNA.<sup>10</sup>

Additionally, recent work from the K.A. Lewis group has shown the CTD of *AtLaRP6C* is mostly disordered (unpublished, E. Hackler). As previously described, recent work in vertebrate LaRP4 has shown the PAM2 containing NTR has independent RNA binding activity and may have important implications for higher order assembly. Therefore, we hypothesize that the N-terminal region of *AtLaRP6C* may contribute additional binding activity for A-rich RNA ligands.<sup>10,13</sup>

These developments in the field generated two major questions. First, does the NTR physically interact with the La Module, as is seen in other LaRP6 proteins? Second, does the NTR modulate the RNA binding activity in a PAM2-dependent manner, as is observed in other LaRPs? To fully test the hypothesis that the NTR is important for binding A-rich sequences, additional protein constructs are needed. Previous work on the plant LaRP6 proteins only focused on the isolated La Modules, due to challenges in the recombinant expression of the full-length *AtLaRP6* proteins. Recent work in the K.A. Lewis group successfully expressed and purified all three full-length *Arabidopsis* LaRP6 paralogs.<sup>13</sup> These reagents enable the biophysical analysis of various constructs derived from the full-length protein, which will enable direct testing of the hypotheses that the NTR interacts with the La Module to modulate structure and function of plant LaRP6. Recombinant expression constructs for the full-length *AtLaRP6C* as well as a  $\Delta$ PAM2 mutant and the isolated La Module were cloned, expressed, and initially purified as N-terminal SUMO fusions by former lab members. To directly test the effect of the NTR on the La Module, a " $\Delta$ CTD"

construct will be synthesized, which will be comprised of the NTR and the La Module. The importance of the PAM2 will be evaluated using the  $\Delta$ PAM2 mutant of this truncated protein. To test the NTR for RNA binding activity, we also need to generate a construct comprised of only the isolated NTR. These new constructs will be recombinantly expressed in *E. coli* and purified using immobilized metal affinity chromatography followed by size exclusion chromatography. The goal will be to purify all of these constructs to >99% purity and to greater than 10  $\mu$ M stock concentration for use in biochemical analyses.

To determine the role of the NTR and specifically the PAM2w motif in protein stability, a modified differential scanning fluorimetry (DSF) protocol will be utilized to extract protein stability information of the constructs, specifically the melting temperature ( $T_m$ ).<sup>14</sup> To determine the role of the NTR and PAM2w motif in RNA binding activity, electrophoretic mobility shift assays (EMSAs) against identified ligands will be used. Building on the methods in the K.A. Lewis lab to recombinantly express and purify various constructs of *At*LaRP6C, this work will directly test these hypotheses. By characterizing the stability and RNA binding activity of these proteins, we will gain significant insight into the function of individual domains within this highly conserved protein.

## II. MATERIALS AND METHODS

### Site-directed mutagenesis of pET28-SUMO-A $\alpha$ LARP6C full-length constructs

Mutagenic primers were designed to alter the full-length constructs by mutating specific base pairs. A typical SDM reaction was made using 0.5  $\mu$ L of template DNA, 0.5  $\mu$ M forward primer, 0.5  $\mu$ M reverse primer, 1 $\times$  Phusion HF buffer, 0.2 mM dNTPs, and 1.0 unit Phusion polymerase with a final volume of 50  $\mu$ L. The thermocycling parameters were construct-dependent and are listed in Table S1. The SDM reactions were treated with 1.0 unit of *DpnI* for 1 hour at 37°C. The digested samples were then transformed into DH5 $\alpha$  *E. coli* cells following the standard protocol with the following changes: 8  $\mu$ L of SDM reaction/ 50  $\mu$ L DH5 $\alpha$  cells.

### Transformation of plasmid DNA into *Escherichia coli* cells

To 50  $\mu$ L of thawed DH5 $\alpha$  cells, 0.5  $\mu$ L of plasmid DNA was added and incubated on ice for 30 minutes. The samples were then heat-shocked at 37°C for 1.5 minutes and cooled for 2 minutes on ice. Cells were recovered by the addition of 700  $\mu$ L LB and a 1 hour incubation at 37°C with shaking. The cells were plated on LB-agar plates supplemented with kanamycin (35  $\mu$ g/mL final concentration) and were incubated for 16 – 18 hours at 37°C. DNA was isolated from the transformed cells using the QIAprep Spin Miniprep Kit (Qiagen) and sequence verified by Sanger sequencing (Genewiz). The sequence verified DNA was transformed into competent Rosetta™ *E. coli* cells for expression using the

same protocol listed previously with the following variations: 0.5  $\mu\text{L}$  of template DNA was added to 100  $\mu\text{L}$  thawed cells, heat-shock occurred at 42°C for 45 seconds, and the LB-agar plates were supplemented with kanamycin (35  $\mu\text{g}/\text{mL}$  final concentration) and chloramphenicol (34  $\mu\text{g}/\text{mL}$  final concentration).

### **Expression of His<sub>10</sub>-SUMO-A $\alpha$ LARP6C Variants (small-scale and large-scale)**

For each construct transformed into Rosetta™ cells an isolated colony was used to inoculate LB supplemented with kanamycin (35  $\mu\text{g}/\text{mL}$  final concentration) and chloramphenicol (34  $\mu\text{g}/\text{mL}$  final concentration). The inoculates were incubated for 16 – 18 hours with shaking at 37°C. After this initial growth, a 1:100 ratio of inoculum to fresh Miller's broth supplemented with the same antibiotics was grown to an OD<sub>600</sub> of 0.6-0.7 at 37°C with shaking. The cell solution was cold-shocked for 15 minutes with agitation and expression was induced with IPTG (100 mM). The culture was incubated at 16°C for 16-20 hours with shaking. To ensure expression 1 mL culture pellets were collected throughout expression, and stored at -20°C. Small-scale expressions were stopped without harvesting the cells. Once expression was confirmed, large-scale expressions were carried out in the same manner, as 1 L cell cultures. Cells were harvested by pelleting the cell culture at 5,000  $\times g$ , 4°C for 10 minutes in a Sorvall LYNX 6000 centrifuge. The resulting pellets were stored at -20°C.

### **Verification of Expression**

The 1 mL pellets collected during expression were resuspended in 300  $\mu\text{L}$  1 $\times$  SDS-Sample Buffer (50 mM Tris (pH 6.8), 2% SDS, 10% glycerol, 1%

$\beta$ -mercaptoethanol) and heated to 90°C for 5 minutes. Each sample, was analyzed by SDS-PAGE on duplicate identical gels, and electrophoresed in 1× Tris-Gly-SDS Buffer (3 g Tris, 14.4 g glycine, 1 g SDS, dissolved in ultrapure polished water and brought to 1 L) at 200 V for 1 hour at room temperature. These 'sister' gels were used in two different detection methods. One gel was stained for 30 minutes with Coomassie Stain (50% methanol, 10% acetic acid, 0.1% Coomassie brilliant blue R-250), and de-stained for 30 minutes using Coomassie De-stain (40% methanol, 10% acetic acid). This gel was imaged using the BioRad ChemiDoc XRS+ system using the "Protein > Coomassie" built-in setting. The other gel was transferred to a BioRad TransBlot Turbo Mini-size nitrocellulose membrane using the Bio-Rad TransBlot transfer system using the Mixed Molecular Weights setting (1.3 A, 25 V, 7 min). Tris-buffered saline (TBS) was made as a 1 L 10× solution by dissolving 24 g Tris and 88 g NaCl in ultrapure polished water and made to pH 7.4. The membrane was incubated for 1 hour in Blocking Solution (5% BSA in 1× TBS -T (0.05% Tween-20)), and then incubated for 1 hour with 1:5000 HisProbe-HRP in 1× TBS-T. The membrane was subsequently washed twice with 1× TBS-T, and twice with 1× TBS for 10 minutes each. The membrane was incubated with 20 mL of enhanced chemiluminescence detection solution (a.k.a. "The Juice"; 1 M Tris (pH 8.8), 250 mM luminol (in DMSO), 90 mM 4-IPBA (in DMSO)) and 12  $\mu$ L 30% hydrogen peroxide. The membrane was imaged using the BioRad ChemiDoc XRS+ system using the "Blots > Chemi" built-in setting.

## **Purification of AtLaRP6C variants**

### *Resuspension of cell pellet and sonication*

A thawed 1 L cell pellet was resuspended in 30 mL Wash-1 buffer with the addition of 1 EDTA-free protease inhibitor tablet. Cells were lysed by sonication at 37% amplitude for a total time of 1.5 mins with 20 s of sonication followed by 30 s of rest. The lysate was centrifuged at 18,000  $\times g$  at 4°C for 15 min.

### *Nickel immobilized metal affinity chromatography (IMAC)*

For each 1 L cell pellet, 2 mL packed HisPur Ni-NTA resin (Thermo Fisher Scientific) was equilibrated in Wash-1 buffer + protease inhibitor. The cleared lysate was combined with the equilibrated Ni-NTA resin and incubated for 1 hour at 4°C with shaking. The combined mixture was added to a gravity flow column and the flowthrough was collected. The resin was subsequently washed with 20 mL Wash-1, 24 mL Wash-2, and 24 mL Elution buffer and fractions from each solution was collected. Presence of the protein of interest in the fractions was determined by SDS-PAGE and Coomassie stain. Fractions containing the protein of interest were pooled and concentrated using Sartorius VivaSpin Ultrafiltration units.

### *Size-exclusion chromatography (SEC)*

Approximately 2.5 mL of nickel-pure concentrated protein was loaded on to a Sephadex S75 or S200 exclusion chromatography (SEC) column equilibrated with SEC buffer. Fractions containing the protein of interest were

identified by peaks on a chromatogram measuring absorbance at 280 nm and verified by SDS-PAGE with Coomassie stain.

#### *His<sub>10</sub>-SUMO tag cleavage and subsequent IMAC and SEC*

Fractions containing the protein of interest were pooled and ULP1 was added at a molar ratio of 1:100 and incubated at 16°C for two hours to remove the His<sub>10</sub>-SUMO tag. The protein was separated from the cleaved SUMO tag by nickel IMAC utilizing the same wash and elution buffers, and fractions containing the cleaved protein were identified, pooled, concentrated, and re-applied to the sizing column. Presence of protein was determined by chromatogram and purity determined by SDS-PAGE with Coomassie stain. The purified protein solution was brought to 5% glycerol, aliquoted, snap-frozen, and stored at -70°C.

#### *Determining molecular weight from SEC*

Both S75 and S200 columns were calibrated using molecular weight standards (Table 1, 2). Using the elution volume of these standards and their respective molecular weights, the molecular weight of subsequent proteins loaded on the columns could be determined. Blue dextran as the largest standard is used to identify the void volume ( $V_0$ ) of the column. During purification, the  $K_{av}$  (average distribution constant) is calculated using the equation:

$$K_{av} = \frac{V_e - V_0}{V_c - V_0}$$

where  $V_e$  is the elution volume of the protein of interest, and  $V_c$  is the total volume of the column (120 mL). The  $K_{av}$  values of the standards were plotted against their respective elution volumes, and a linear regression was performed to generate an equation to calculate an apparent molecular weight ( $MW_{app}$ ), where  $x$  is the  $\log_{10}$  of the molecular weight (Da) of the standard protein, and  $y$  is  $K_{av}$ . The columns were re-calibrated during the course of this project, so there are two sets of equations for the S200 column.

S200 Equations:

$$\text{FL, } \Delta\text{PAM2}\Delta\text{CTD: } y = -0.3782x + 2.265$$

$$\Delta\text{PAM2, } \Delta\text{CTD: } y = -0.3373x + 2.0997$$

S75 Equation:

$$\text{La Mod, NTR, } \Delta\text{PAM2NTR: } y = -0.4962x + 2.5122$$

**Table 1: Molecular weight standards for S200 column.**

Symbol	Standard	MW (kDa)	Elution Volume (mL)
◇	Blue Dextran	2,000	47.03
◆	Ferritin	440	60.62
◆	Aldolase	158	72.98
◆	Conalbumin	75	80.28
◆	Ovalbumin	43	85.15
◆	Carbonic Anhydrase	29	91.91
◆	RNase A	13.7	99
◆	Aprotinin	6.5	105.34

**Table 2: Molecular weight standards for S75 column.**

Symbol	Standard	MW (kDa)	Elution Volume (mL)
◇	Blue Dextran	2,000	47.6
◆	Conalbumin	75	55.36
◆	Ovalbumin	43	60.97
◆	Carbonic Anhydrase	29	69.24
◆	RNase A	13.7	81.725
◆	Aprotinin	6.5	92.245

### **Electrophoretic mobility shift assays (EMSAs)**

Within a normal environment RNases are found which easily degrade RNA. To mitigate degradation of the RNA used for these experiments, extreme care was used while handling and preparing the RNA. This included cleaning surfaces and tools with RNase Zap (Invitrogen) to remove RNases and using only sterile DNase and RNase-free pre-packaged filter-tips.

Protein aliquots were thawed from  $-70^{\circ}\text{C}$  and centrifuged at  $16,000 \times g$  for 15 minutes at  $4^{\circ}\text{C}$  to pellet any aggregates. The top  $\sim 90\%$  of the protein solution was moved to a new 1.5 mL microcentrifuge tube. Before serial dilution, RNA ligands were heated at  $80^{\circ}\text{C}$  for 10 minutes or 2 minutes for biotinylated and FAM-labelled RNA respectively. Reactions were prepared in  $1\times$  Binding Buffer (20 mM Tris-HCl (pH 7.25), 200 mM KCl, 15% glycerol, 1 mM DTT, 0.1 mg/mL BSA). Proteins were serially diluted to  $2\times$  final concentrations 0 – 40  $\mu\text{M}$  and mixed 1:1 v/v with labelled RNA ligands (final concentrations 0.5 – 2 nM) in 1.5 mL microcentrifuge tubes with a total reaction volume of 24  $\mu\text{L}$  and allowed to equilibrate for 1 hour on ice. From each reaction 20  $\mu\text{L}$  was loaded on a 6.5% native polyacrylamide gel (29:1 acrylamide:bisacrylamide (ProtoGel),  $1\times$  Tris-borate-EDTA (TBE: 1 L of  $5\times$  TBE made with 54 g Tris, 28 g boric acid, and 40 mL 0.5 M EDTA (pH 8.0), 5% glycerol) with  $1\times$  TBE buffer and separated at 200 V for 20 minutes. To absorb heat from the buffer,  $-70^{\circ}\text{C}$  ice-packs were added to the tank.

#### *Biotinylated RNA ligands*

RNA ligands were ordered from IDT and had been previously biotinylated using the Pierce™ RNA 3' End Biotinylation Kit by former lab members, and stored at  $-20^{\circ}\text{C}$  for less than 1 year. After separation, the gel was transferred to a  $1\times$  TBE equilibrated Hybond(+) membrane (GE Biosciences) using the BioRad TransBlot Turbo system (25 V, 1.0 A, 30 min). The membrane was then crosslinked with a UV oven at  $120 \text{ mJ}/\text{cm}^2$  for 45 seconds. The membrane was

then dried overnight or detected immediately using the Chemiluminescence Nucleic Acid Detection Kit (ThermoScientific). The membranes were imaged using the BioRad ChemiDoc XRS+ on the Blots>Chemi setting.

#### *5'-FAM-labelled RNA ligands*

RNA ligands were ordered with 5'-FAM labels from IDT. FAM-labelled ligands were resuspended in 0.5× Tris-EDTA (TE) to preserve fluorescence and stored at -20°C. When EMSAs were performed with these ligands, reactions were kept in the dark as much as possible to minimize photobleaching of FAM. The binding assays were performed as normal, but after separation on the gel they were immediately imaged on the Pharos gel imager on the FITC setting at 50 μm, with imaging area made as tightly around the gel as possible.

#### **Guanidinium hydrochloride protein denaturation**

Guanidinium hydrochloride (GndHCl) was dissolved in 50 mM NaH<sub>2</sub>PO<sub>4</sub>/Na<sub>2</sub>HPO<sub>4</sub> (pH 8.0) with the highest concentration at 8.5 M. After making the GndHCl solution, 1 mL aliquots were made and stored at -70°C to prevent degradation. Before each denaturation reaction, the concentration of GndHCl was determined by refractometer and using the formula:  $57.147(\Delta N) + 38.68(\Delta N)^2 - 91.6(\Delta N)^3$  from C. N. Pace<sup>15</sup>. Varying concentrations of GndHCl were mixed with 5 μM protein at a final volume of 250 μL and allowed to incubate for 1 hour at the desired temperature. From each reaction 200 μL was transferred to a sub-micro quartz fluorometer cuvette (Starna Cells, Inc) cleaned with ultrapure polished water and 100% ethanol. The cuvettes were placed in a Cary

fluorometer multicell holder excitation at 280 nm and emission spectrum collected between 300 – 460 nm. Data was first processed by plotting the intensity by the fluorescence. The fluorescence at 340 nm was then plotted against the concentration of GndHCl. All processing was done in Excel (Microsoft).

### **Thermofluor assay using SYPRO™ Orange**

Reactions were made with 5× SYPRO™ Orange (Fisher Scientific), 5 μM protein, 20 mM MOPS (pH 7.0), and 150 mM NaCl with a final volume of 225 μL. Before dilution, proteins were centrifuged at 16,000 ×g for 15 minutes at 4°C and the top ~90% of solution was moved to a clean 1.5 mL microcentrifuge tube. SYPRO™ Orange was diluted from the manufacturer's 5000× stock immediately prior to reactions, using nuclease-free water (IDT). From each reaction, 200 μL was transferred to a sub-micro masked quartz cuvette (Starna Cells, Inc) and placed in a Cary fluorometer multicell holder that is temperature-regulated with a Peltier device. The thermal setting was used with the following specifications: excitation 470 nm, emission 570 nm, excitation slit width 2.5 nm, emission slit width 10 nm, 1.0 s averaging time, 0.5 data interval, 0.5°C/min, end-temp at 95°C, start-temp at 20°C. Teflon lids were placed on each cuvette to prevent extensive evaporation of solution. Independent replicates were performed on different days and in different slots within the multicell holder. The cuvettes were cleaned with 1 M HCl to remove any denatured and/or adsorbed biomolecules from the interior of the cuvettes.

### *Data processing*

Fluorescence intensity at each temperature was obtained as raw data from the Cary fluorometer, and was manipulated in Excel (Microsoft). To determine melting temperature ( $T_m$ ) for each protein, a two-state model for denaturation was assumed, and the fraction of folded protein ( $P_f$ ) was calculated at each temperature using the formula:

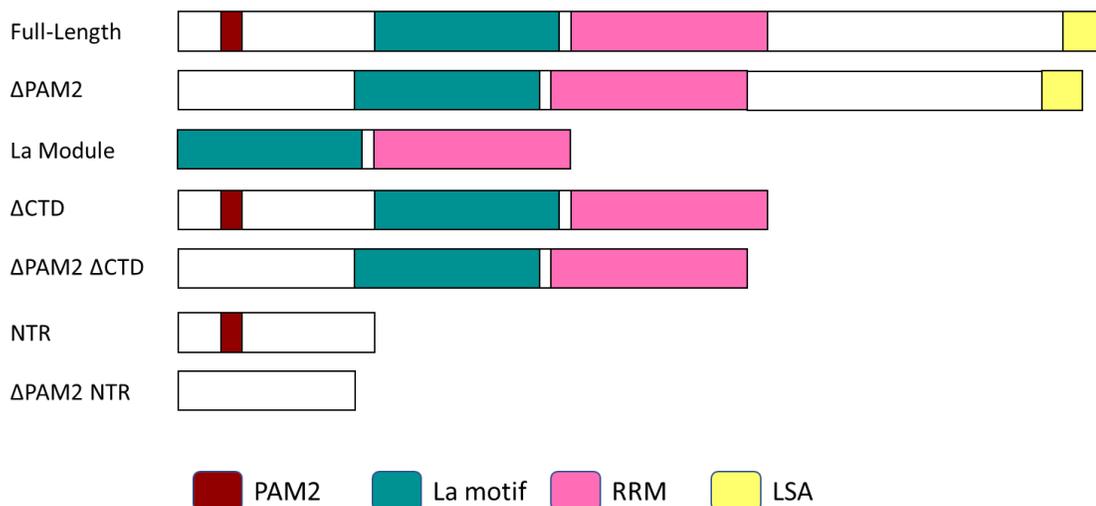
$$P_f = 1 - \frac{F - F_{min}}{F_{max} - F_{min}}$$

where  $F$  is the fluorescence at a specific temperature,  $F_{min}$  is the minimum fluorescence value, and  $F_{max}$  is the maximum fluorescence value.  $F_{min}$  and  $F_{max}$  values were limited to the increasing portion of the curves.  $T_m$  values were identified where  $P_f = 0.5$ . To plot the normalized data, the fraction of unfolded protein ( $P_u$ ) was calculated using the formula below, and plotted against the temperature:

$$P_u = 1 - P_f$$

### III. RECOMBINANT EXPRESSION AND PURIFICATION OF *AtLaRP6C* PROTEIN VARIANTS

To test the hypotheses about the structure and RNA binding activity of *AtLaRP6C*, a full suite of protein variants were developed. In addition to the full-length *AtLaRP6C* and the isolated La Module, the deletion of the PAM2 motif in the N-terminal region (NTR) is needed to test its role in protein structure and RNA binding. Similarly, the deletion of the C-terminal domain (CTD) will allow for the evaluation of how the NTR contributes to structure and function.



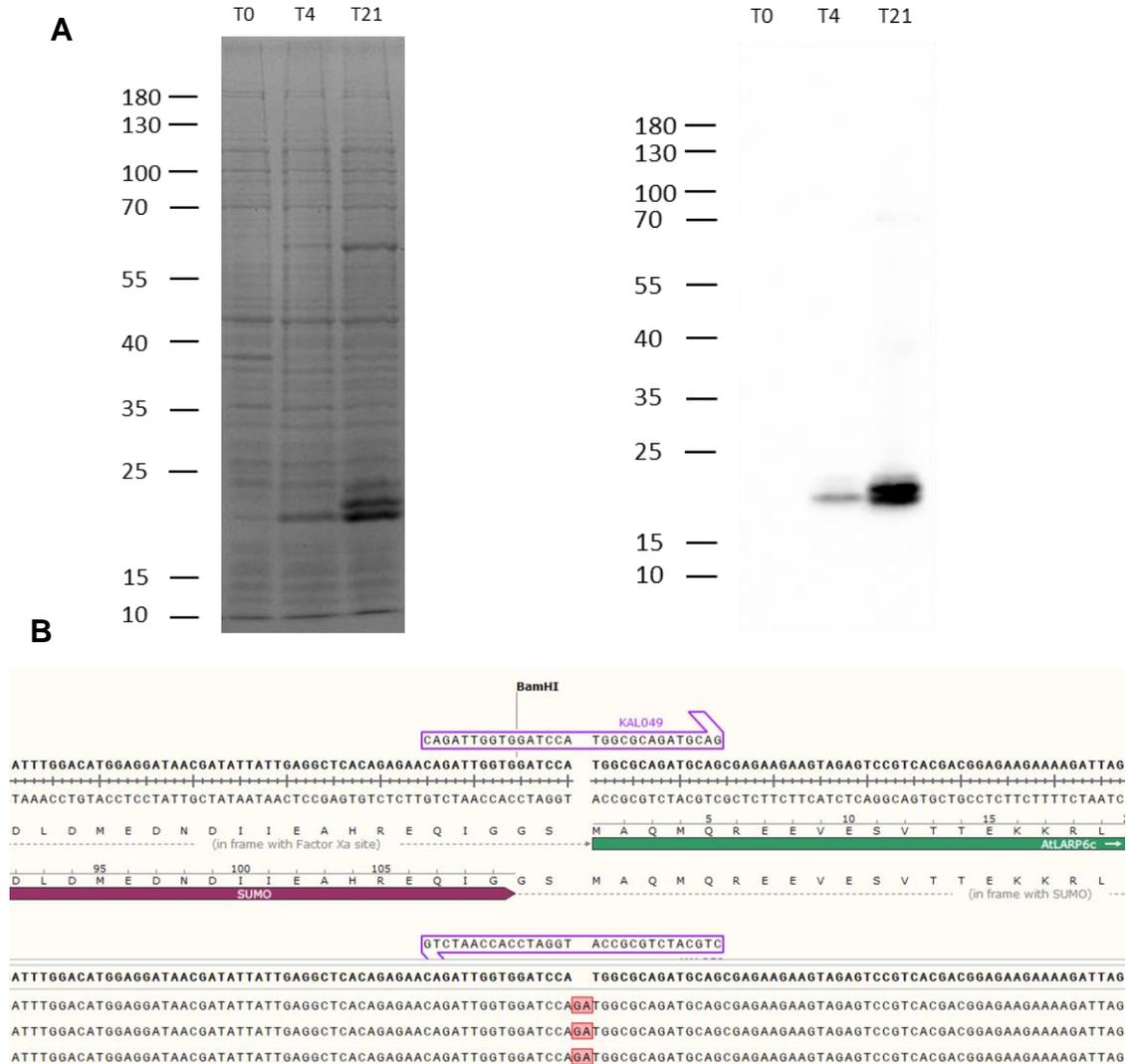
**Figure 3: *AtLaRP6C* constructs used in this project.** The Full-Length, ΔPAM2, and La Module were previously expressed and purified. The ΔCTD and NTR constructs were made to directly test the importance of the N-terminal region on stability and RNA binding activity. The ΔPAM2ΔCTD and ΔPAM2 NTR constructs were made to directly test the effect of the PAM2 domain on RNA binding activity.

#### Expression of *AtLaRP6C* Constructs

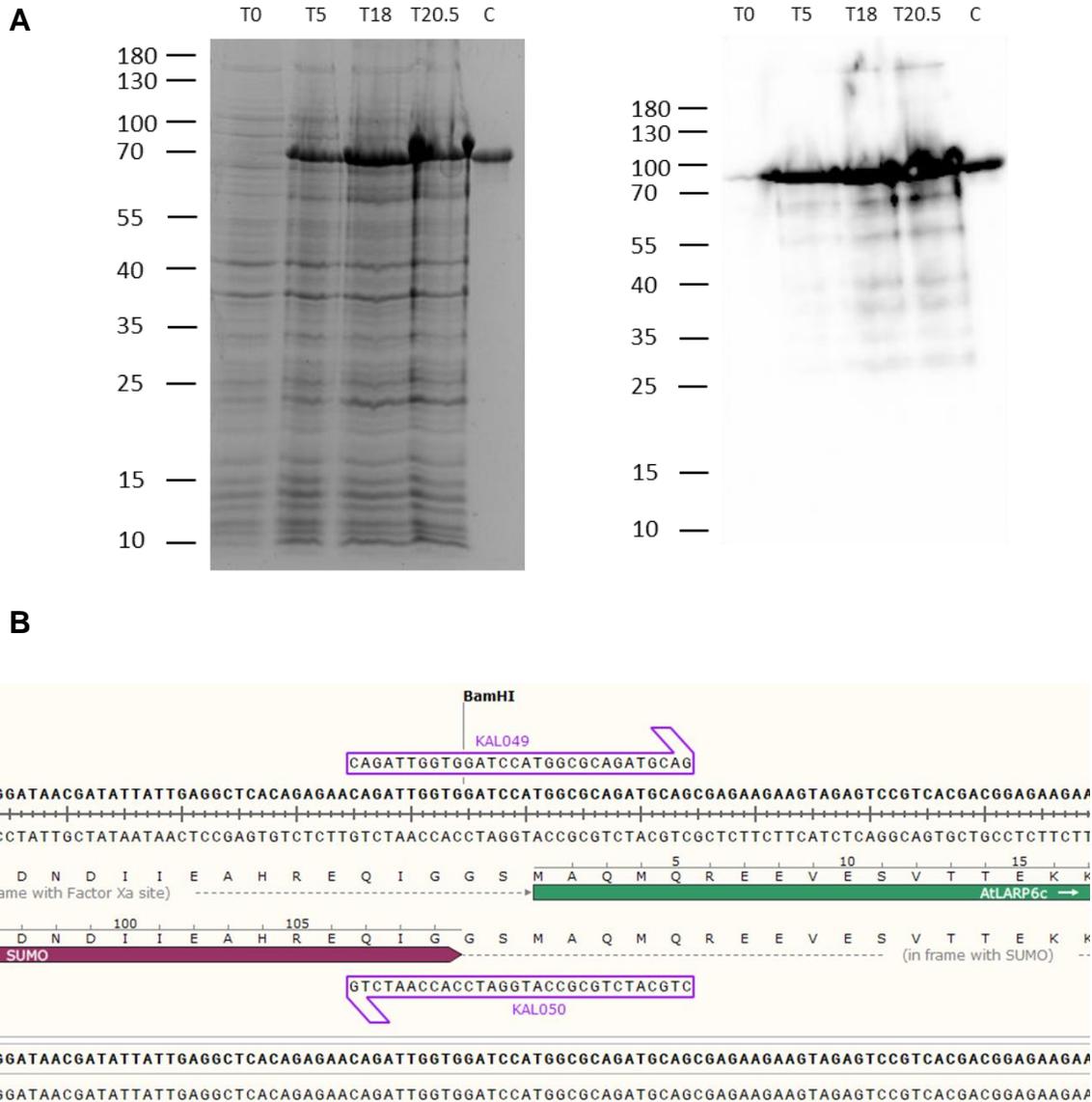
The full-length (FL), ΔPAM2, and La Module constructs were previously expressed and purified by former members of the lab. To make fresh protein

preparations, these constructs were expressed following the established expression protocols<sup>13</sup>(Foster, J & Foster, C, unpublished) . As described in Methods, aliquots were removed from the expression cultures at regular time intervals following induction and analyzed for protein content. These data revealed that the FL construct did not express as expected. When fused to the His<sub>10</sub>-SUMO tag, the expected molecular weight of the full-length construct is 63 kDa. In both the Coomassie stained gel and the anti-His western blot, a band of increasing intensity is seen over time for both constructs at 20 kDa (Figure 4A). As this was not the correct molecular weight, a second expression was carried out using a colony picked from a fresh transformation of the plasmid DNA into Rosetta™ cells. However, the same result was obtained (data not shown). All plasmid DNA stocks were sent for Sanger sequencing (Genewiz) and a 2 nt insertion between the SUMO tag and gene was identified (Figure 4B). This insertion caused a frameshift in the reading frame of the LaRP6C sequence which lead to a nonsense mutation. This frameshift was remedied by site-directed mutagenesis (SDM), and successful mutagenesis was confirmed by Sanger sequencing. The resulting plasmid was used to create new plasmid stocks and to carry out a new expression in Rosetta™ cells as described above. The expression of the rectified construct produced a band at 70 kDa that increased in intensity over time, as evaluated by both Coomassie stain and anti-His western blot (Figure 5). From previous experience in the K. A. Lewis lab, it is well established that the LaRP6 constructs migrate ~10 kDa higher in SDS-

PAGE gels than the expected molecular weight.<sup>16</sup> The *AtLaRP6C* proteins behave in the same manner. This is likely due to the acidic nature of the proteins.



**Figure 4: Expression trials of frameshifted *AtLaRP6C* FL.** pET28-SUMO plasmids containing the *AtLaRP6C* were transformed into *E. coli* Rosetta™ (DE3) competent cells. Expression was induced with 1 mM IPTG and samples were taken at regular time intervals as stated by the numbers above the bands. Cell pellets were resuspended in 1× SDS-PAGE sample buffer and separated by gel electrophoresis on 10% SDS-PAGE gels and followed by either (left) Coomassie stain or (right) anti-His Western blot. C = His<sub>10</sub>-SUMO-*AtLaRP6A* purified by Daniel Horn **(A)** Expression of full-length protein prior to site-directed mutagenesis. **(B)** Sequencing data showing 2 nucleotide insertion between SUMO tag and gene. Expected molecular weight = 63 kDa.

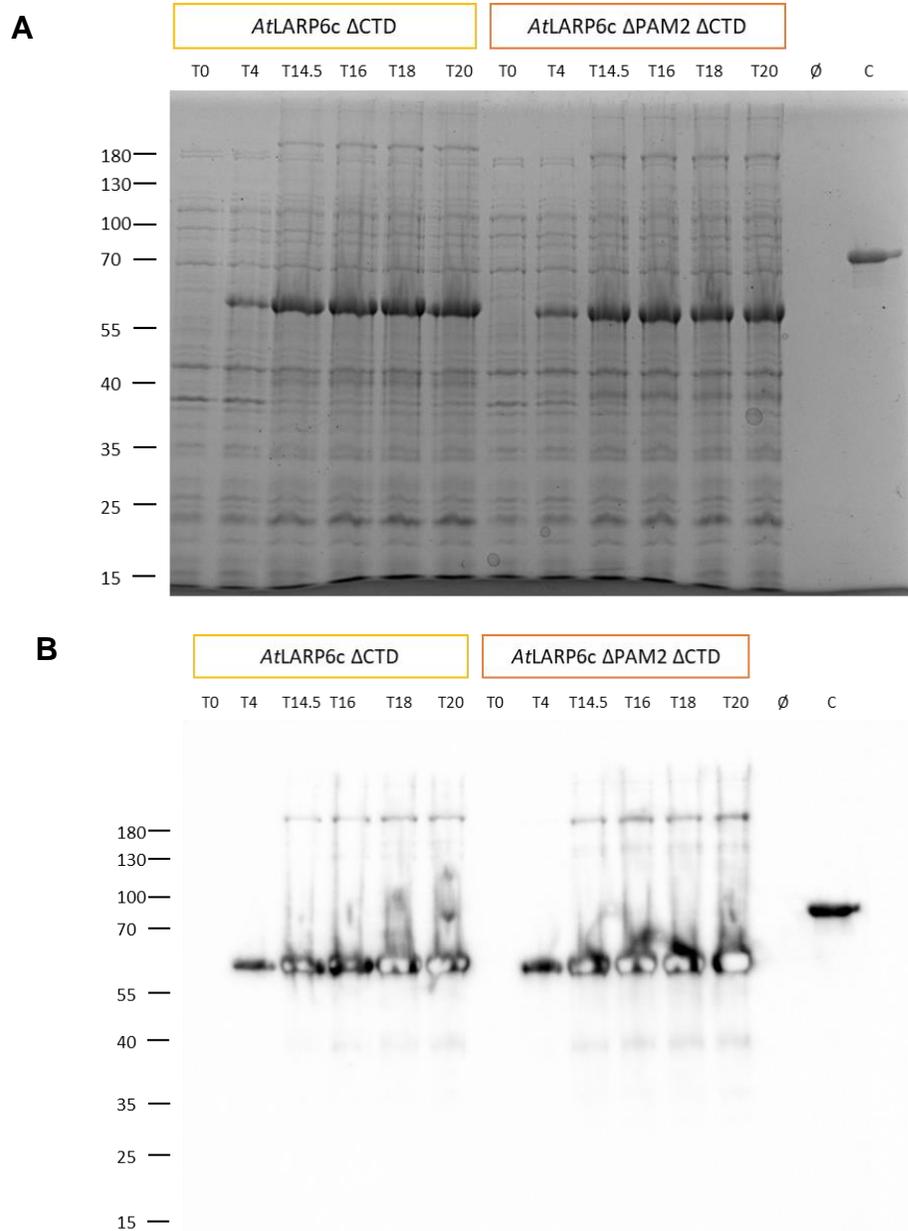


**Figure 5: Expression trials of resolved AtLaRP6C FL.** Sequence verified pET28-SUMO-AtLaRP6C was transformed into *E. coli* Rosetta™ (DE3) competent cells. Expression was induced with 1 mM IPTG and samples were taken at regular time intervals as stated by the numbers above the bands. Cell pellets were resuspended in 1× SDS-PAGE sample buffer and separated by gel electrophoresis on 10% SDS-PAGE gels and followed by either (left) Coomassie stain or (right) anti-His Western blot. C = His<sub>10</sub>-SUMO-AtLaRP6A purified by Daniel Horn **(A)** Expression of full-length protein after site-directed mutagenesis. **(B)** Sequencing data showing successful removal of 2 nucleotide insertion. Expected molecular weight = 63 kDa.

## Expression of *At*LaRP6C $\Delta$ CTD Constructs

The  $\Delta$ CTD constructs were created via SDM for both the wildtype and  $\Delta$ PAM2 sequences. Both SDMs were confirmed via Sanger sequencing (data not shown).

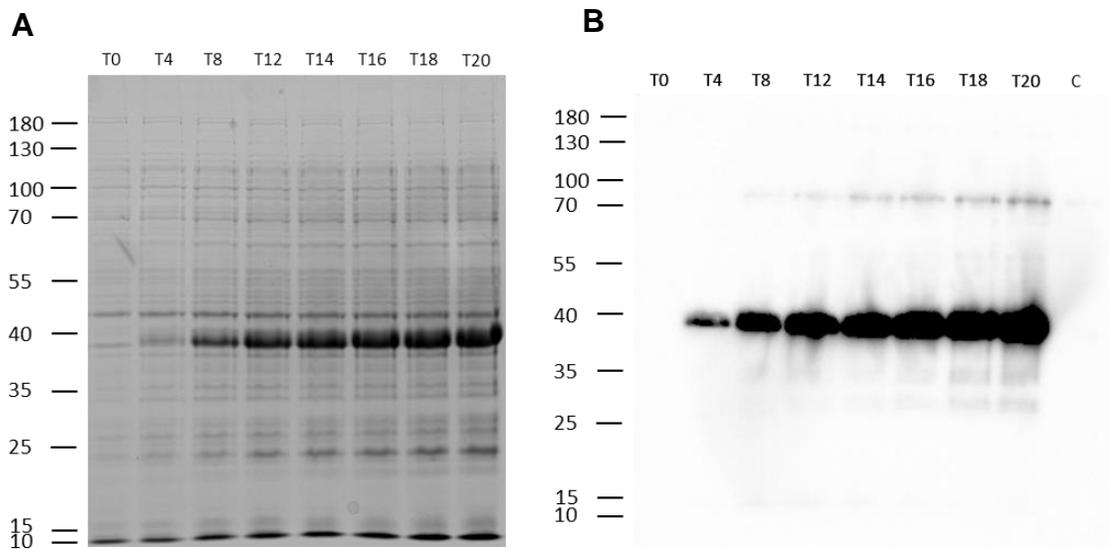
As described above for the other constructs, cell culture aliquots were taken at regular time intervals during expression and analyzed for protein content. The  $\Delta$ CTD construct with the His<sub>10</sub>-SUMO tag had an expected molecular weight of 50 kDa. The  $\Delta$ PAM2  $\Delta$ CTD construct with the His<sub>10</sub>-SUMO tag had an expected molecular weight of 48.4 kDa. In both the Coomassie stained gel and the anti-His western blot, a band of increasing intensity is seen over time for both constructs at 60 kDa (Figure 6). Protein expression was confirmed by the signal present in the anti-His western blot. Although similar in size the  $\Delta$ CTD construct runs slightly higher than the  $\Delta$ PAM2  $\Delta$ CTD construct.



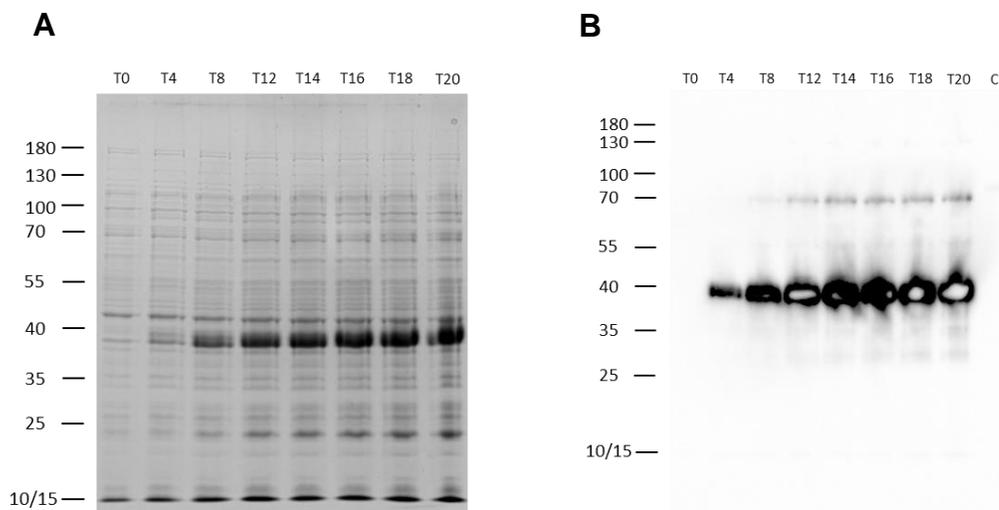
**Figure 6: Expression trials of AtLaRP6C  $\Delta$ CTD and  $\Delta$ PAM2  $\Delta$ CTD.** pET28-SUMO-AtLaRP6C  $\Delta$ CTD and pET28-SUMO-AtLaRP6C  $\Delta$ PAM2  $\Delta$ CTD were transformed into *E. coli* Rosetta™ (DE3) competent cells separately. Expression was induced with 1 mM IPTG and samples were taken at regular time intervals as stated by the numbers above the bands. Cell pellets were resuspended in 1 $\times$  SDS-PAGE sample buffer and separated by gel electrophoresis on 10% SDS-PAGE gels and followed by either **(A)** Coomassie stain or **(B)** anti-His western blot. C = His<sub>10</sub>-SUMO-AtLaRP6A purified by Daniel Horn. Expected molecular weight of  $\Delta$ CTD = 50 kDa. Expected molecular weight of  $\Delta$ PAM2  $\Delta$ CTD = 48.4 kDa.

## Expression of *AtLaRP6C* NTR Constructs

The NTR constructs were created via SDM for both the wildtype and  $\Delta$ PAM2 sequences. Both SDMs were confirmed via Sanger sequencing (data not shown). As previously described, the expression of both NTR constructs were analyzed for protein content. The NTR construct with the His<sub>10</sub>-SUMO tag had an expected molecular weight of 27.8 kDa. The  $\Delta$ PAM2 NTR construct with the His<sub>10</sub>-SUMO tag had an expected molecular weight of 26.2 kDa. In both the Coomassie stained gel and the anti-His Western blot, a band at approximately 40 kDa increases in intensity over time for both the wildtype SUMO-NTR (Figure 7) and the  $\Delta$ PAM2-NTR (Figure 8). Protein expression was confirmed by the signal present in the anti-His western blot.



**Figure 7: Expression trials of *AtLaRP6C* NTR.** pET28-SUMO-*AtLaRP6C* NTR was transformed into *E. coli* Rosetta™ (DE3) competent cells. Expression was induced with 1 mM IPTG and samples were taken at regular time intervals as stated by the numbers above the bands. Cell pellets were resuspended in 1× SDS-PAGE sample buffer and separated by gel electrophoresis on 10% SDS-PAGE gels and followed by either (A) Coomassie stain or (B) anti-His western blot. C = His<sub>10</sub>-SUMO-*AtLaRP6A* purified by Daniel Horn. Expected molecular weight = 27.8 kDa.

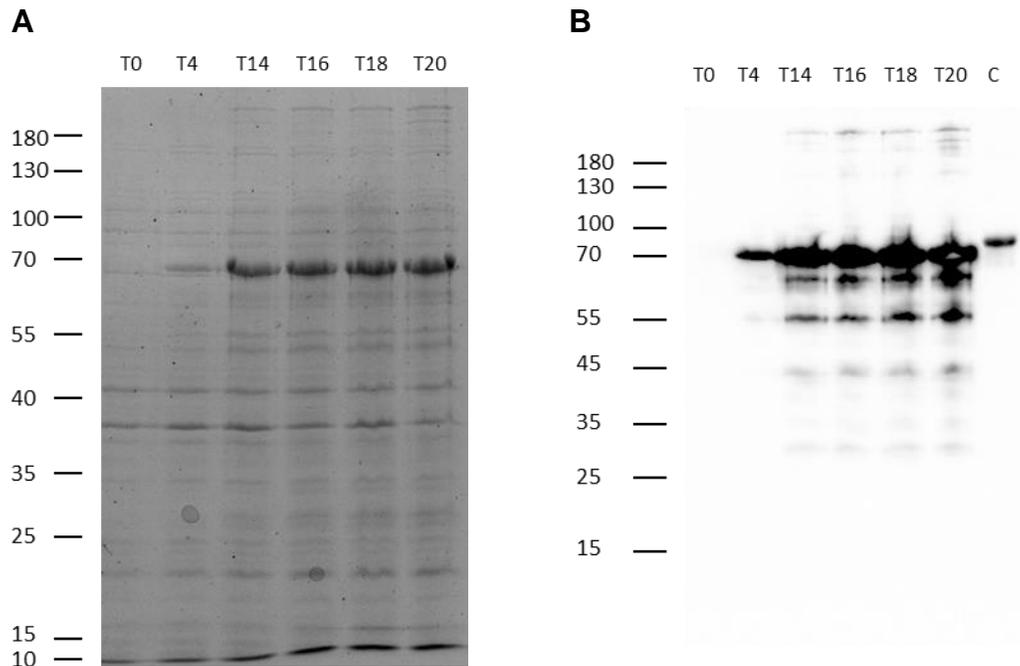


**Figure 8: Expression trials of *AtLaRP6C*  $\Delta$ PAM2 NTR.** pET28-SUMO-*AtLaRP6C*  $\Delta$ PAM2 NTR was transformed into *E. coli* Rosetta™ (DE3) competent cells. Expression was induced with 1 mM IPTG and samples were taken at regular time intervals as stated by the numbers above the bands. Cell pellets were resuspended in 1× SDS-PAGE sample buffer and separated by gel electrophoresis on 10% SDS-PAGE gels and followed by either **(A)** Coomassie stain or **(B)** anti-His western blot. C = His<sub>10</sub>-SUMO-*AtLaRP6A* purified by Daniel Horn. Expected molecular weight = 26.2 kDa.

### Expression of *AtLaRP6C* $\Delta$ LSA Construct

Although not necessary for this project, the  $\Delta$ LSA construct was synthesized by SDM and expressed in *E. coli*. Previous work in the lab had focused on the effect of deleting the C-terminal LSA region of the *AtLaRP6* proteins. In this work, deletion of the LSA in *AtLaRP6C* was thought to destabilize the protein, due to the expression of an anti-His-reactive protein of ~20 kDa (C. Toner, data not shown). However, upon the discovery of the frameshifted wildtype SUMO-*AtLaRP6C* sequence, this finding was revisited. In particular, the molecular weights of the frameshifted FL construct and the  $\Delta$ LSA construct appeared to be the same (data not shown). Again using SDM, the

rectified wildtype plasmid DNA was used as the backbone to insert stop codons upstream of the LSA, effectively deleting the LSA motif. Expression of this construct produced a band with an observed molecular weight  $\sim 70$  kDa, which matched the expected molecular weight of 60 kDa (Figure 9). Therefore, the previous finding of protein destabilization via the deletion of the LSA is likely incorrect and arose from using the frameshifted *AtLaRP6C* construct as a template. These data show that the deletion of the LSA does not impede the expression of the SUMO-tagged *AtLaRP6C*- $\Delta$ LSA protein.



**Figure 9: Expression trials of *AtLaRP6C*  $\Delta$ LSA.** pET28-SUMO-*AtLaRP6C*  $\Delta$ LSA was transformed into *E. coli* Rosetta<sup>™</sup> (DE3) competent cells. Expression was induced with 1 mM IPTG and samples were taken at regular time intervals as stated by the numbers above the bands. Cell pellets were resuspended in 1 $\times$  SDS-PAGE sample buffer and separated by gel electrophoresis on 10% SDS-PAGE gels and followed by either (A) Coomassie stain or (B) anti-His western blot. C = His<sub>10</sub>-SUMO-*AtLaRP6A* purified by Daniel Horn. Expected molecular weight = 60 kDa

## **Purification of AtLaRP6C Proteins: Overview**

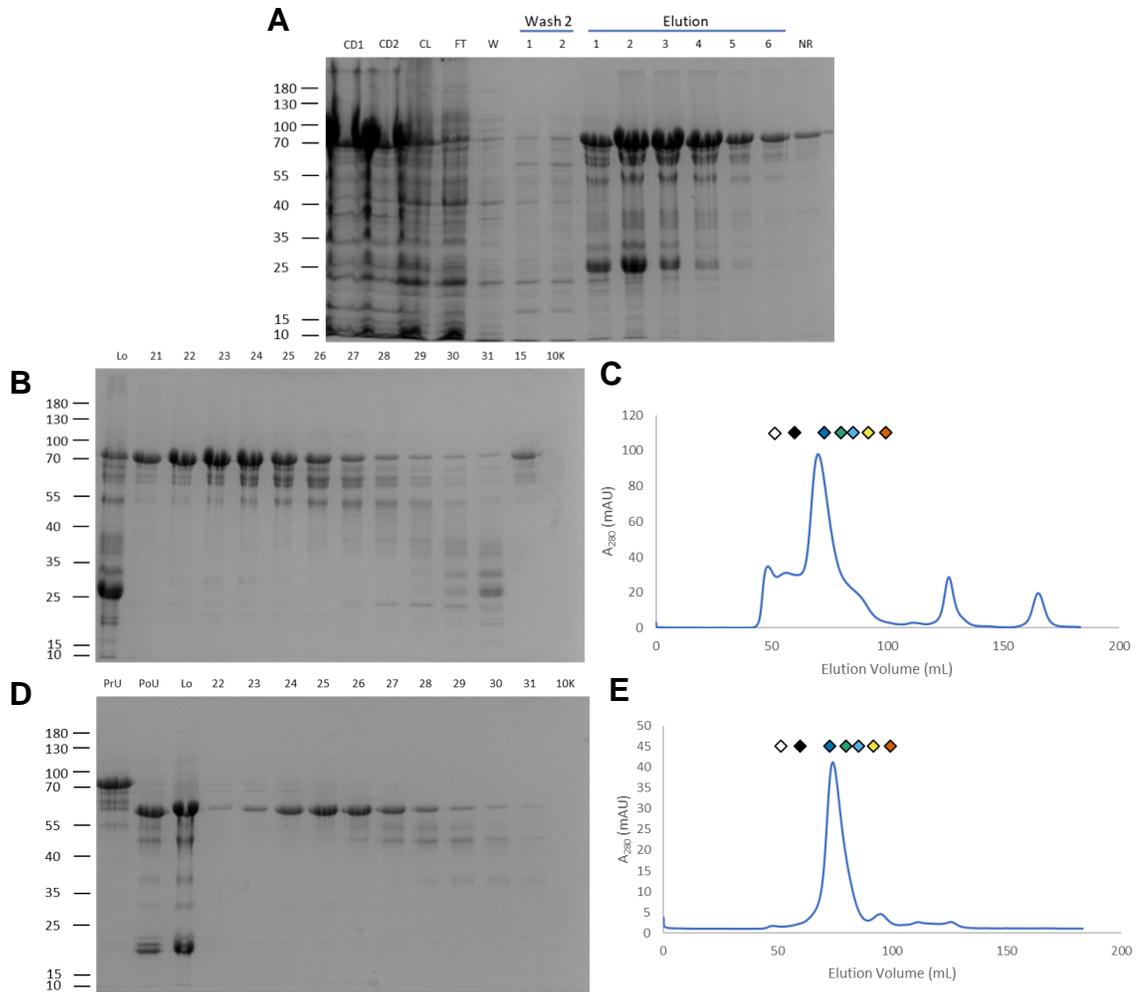
Each of the protein constructs were expressed as 1 L cultures at 16°C for 16-18 hours. Cell pellets were collected, lysed by sonication, and initially purified by nickel affinity chromatography. The presence of protein after each chromatography step was confirmed by SDS-PAGE gel followed by Coomassie staining. Protein containing fractions were then pooled, concentrated, and loaded on an S75 or S200 sizing column depending upon the molecular weight of the protein. Following ULP1 digest to remove the His<sub>10</sub>-SUMO tag, the protein was purified again by nickel affinity chromatography and size exclusion chromatography. After the final sizing column, protein containing fractions were pooled, concentrated, snap-frozen, and stored at -70°C. Details for each protein preparation are described below.

## **Purification of Full-Length AtLaRP6C**

To obtain higher amounts of the FL protein, 2 – 1 L cell pellets were used for the purification. As 50 mL of lysis buffer was prepared for each purification, each cell pellet was resuspended separately in 20 mL of cold lysis buffer and sonicated separately before centrifugation to remove cell debris. The cleared lysate was then pooled for further purification. The expected molecular weight of the His<sub>10</sub>-SUMO tagged protein was 63 kDa, but a band at 70 kDa band was observed in the analyzed SDS-PAGE gel post-nickel affinity chromatography (Figure 10A). This band agreed with the previous expression data. Some protein was lost within the flowthrough and initial wash steps, but most of the protein was eluted after addition of the elution buffer. This was expected as the elution buffer

contains higher concentrations of imidazole which out-competes the poly-histidine tag on the protein. Not all protein was fully eluted from the nickel resin, as can be seen by the presence of a band in the nickel resin sample (Figure 10A). Elution fractions 1 – 6 were pooled and concentrated on a 10,000 MWCO concentrator to 2.5 mL before loading on to a S200 sizing column. To mitigate protein degradation and improve protein stability, 2% glycerol was added to the SEC buffer. Protein elution was monitored by absorbance at 280 nm, and the peaks on the chromatogram were used to identify the presence of protein within the SEC fractions. A small shoulder was observed at the void-volume of the column, and a larger peak was observed at 70.06 mL ( $MW_{app} = 187.2$  kDa) (Figure 10C). Fractions from these peaks were analyzed for protein content and the FL protein along with some degradation products were observed from the fractions directly under the major peak. Some protein was also observed in the void-volume fraction, indicative of protein aggregation. Flowthrough from the 10,000 MWCO concentrator was also analyzed to check for protein content; no protein was observed within this sample. Fractions 21 – 26 with the apparent highest amount of FL protein were pooled and subjected to ULP1 digest (Figure 10B). After ULP1 digest, the solution was concentrated as previously described and loaded again on the S200 sizing column. As the FL protein is much larger than the tag itself, there was no need for a second nickel column, as the cleaved protein should be readily separated from the tag by SEC. The chromatogram shows, a major peak at 74.01 mL ( $MW_{app} = 131.8$  kDa) and the fractions under this peak were evaluated for protein content (Figure 10E). The cleaved FL

protein has an expected molecular weight of 49 kDa and a band was observed at 60 kDa. By comparing the bands between the pre-ULP1 and post-ULP1 fractions, a decrease in size from 70 kDa to 60 kDa is observed. The post-ULP1 fraction also contains additional bands at 20 kDa at the expected molecular weight of the of the His<sub>10</sub>-SUMO tag, confirming successful cleavage of the tag (Figure 10D). Fractions 24 – 27 were pooled and concentrated (to ~20 μM), then brought to 5% v/v glycerol as a cryoprotectant prior to snap-freezing. The protein was snap-frozen as 40 μL aliquots and stored at -70°C.

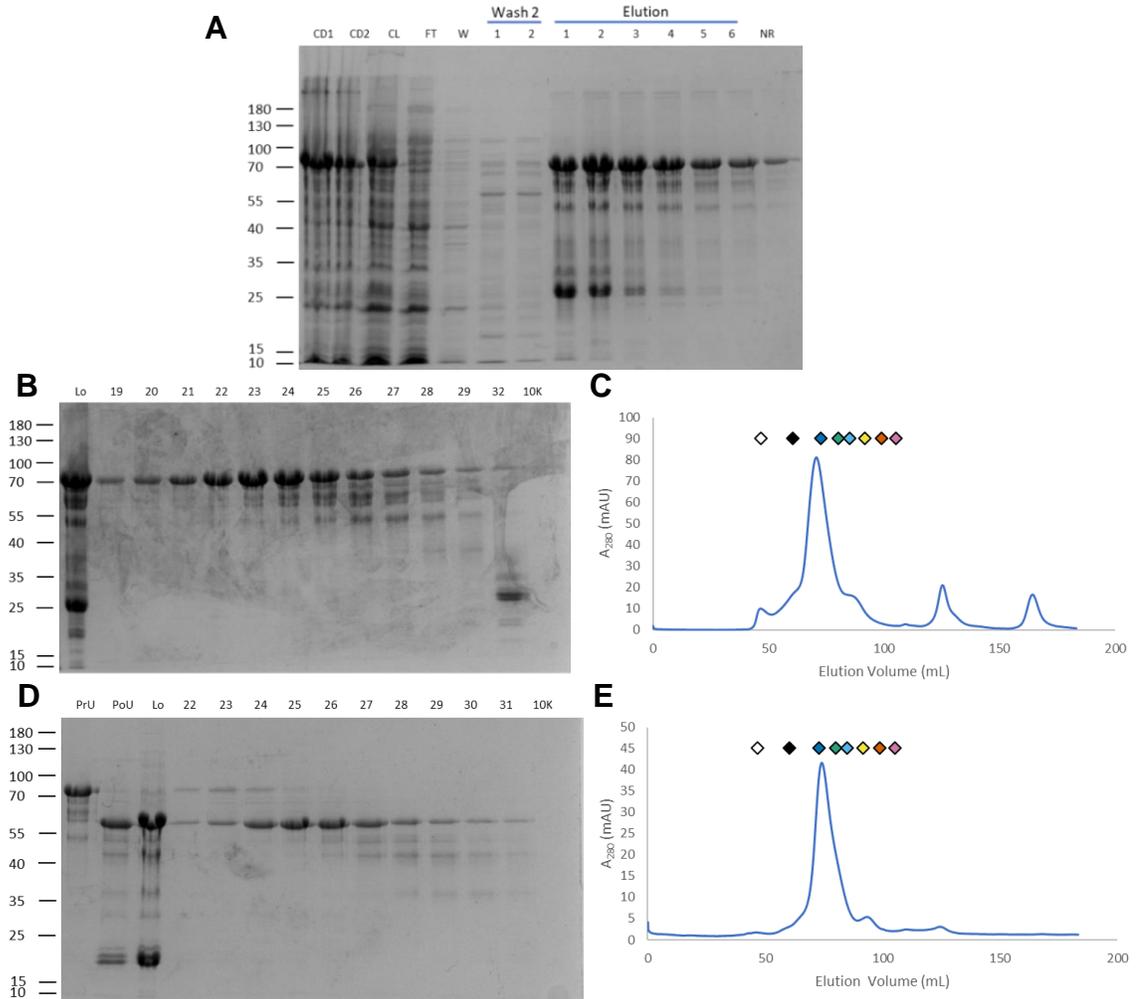


**Figure 10: Purification of AtLaRP6C Full-length.** **(A)** The protein was separated from cleared bacterial cell lysate by a Ni-NTA column. All six Elution fractions were pooled and loaded on an S200 Sephadex column. CD1 = Cell debris from pellet 1, CD2 = Cells debris from pellet 2, CL = Cleared lysate, FT = Flowthrough, W = Wash, NR = Nickel resin **(B)** Protein content within fractions under the peaks as indicated by the chromatogram were visualized by Coomassie stain. Fractions 21 – 26 containing the protein were pooled for ULP1 digest. Lo = protein loaded on sizing column. Expected molecular weight = 63 kDa. **(C)** Chromatogram from the FPLC indicating presence of proteins by peaks. Molecular weight standards are shown above. **(D)** Protein content of fractions under the peaks as indicated by the chromatogram were visualized by Coomassie stain. Fractions 24 – 27 were kept for storage. Expected molecular weight = 49 kDa. PrU = Pre-ULP1 digest, PoU = Post-ULP1 digest, Lo = protein loaded on sizing column, 10K = flowthrough from 10,000 MWCO Sartorius VivaSpin concentrator. **(E)** After incubation with ULP1, the protein was pooled and loaded on an S200 Sephadex column to separate cleaved protein from the His10-SUMO tag and un-cleaved protein. Chromatogram from the FPLC indicating presence of protein by peaks. Molecular weight standards are shown above.

## Purification of AtLaRP6C $\Delta$ PAM2

Following the success of the FL purification, the  $\Delta$ PAM2 was also purified using 2 – 1 L pellets as mentioned before with 2% glycerol in the SEC buffer. Previously this protein was purified in pH 8.0 buffer, but as the pI is 7.83 (tagged protein), pH 8.2 was used in this work to keep the protein from being neutrally charged. The expected molecular weight of the His<sub>10</sub>-SUMO tagged protein was 61.3 kDa and a band at 70 kDa was observed in the first affinity chromatography gel. Some protein remained in the cell debris of both pellets, but very little protein was lost in the flowthrough and wash steps. As expected, the protein was eluted with the elution buffer and again some protein did not come off the nickel resin as can be seen by the presence of a band at 70 kDa in the nickel resin sample (Figure 11A). Elution fractions 1 – 6 were pooled and concentrated on a 10,000 MWCO concentrator to 2.5 mL and filtered (0.2  $\mu$ m) before loading on a S200 sizing column. Protein elution was monitored by absorbance at 280 nm, and the peaks on the chromatogram were used to identify the presence of protein within the SEC fractions. Fractions under the major peak at 70.7 mL ( $MW_{app} = 183.4$  kDa) were run on an SDS-PAGE gel followed by Coomassie staining (Figure 11B,C). Fractions 20 – 27 were chosen to move forward with the ULP1 digest. Following tag cleavage, the second nickel affinity column was omitted, and the protein was loaded on to the sizing column. Following the peaks on the chromatogram (Figure 11E), the fractions under the peak at 74 mL ( $MW_{app} = 134.7$  kDa) was run on an SDS-PAGE gel to determine purity of the fractions (Figure 11D). After cleavage, the expected molecular weight was 47.4 kDa and a

band at 60 kDa was observed. Cleavage of the tag was not 100% effective as can be seen by the 70 kDa band in the pre-ULP1 sample and in fractions 22 – 25 (Figure 11D). Due to this contamination of the tagged protein, fractions 25 – 29 were pooled and concentrated (to ~20.5  $\mu$ M) before being brought to 5% glycerol to prevent degradation after snap-freezing. The protein was snap-frozen as 40  $\mu$ L aliquots and stored at -70°C.



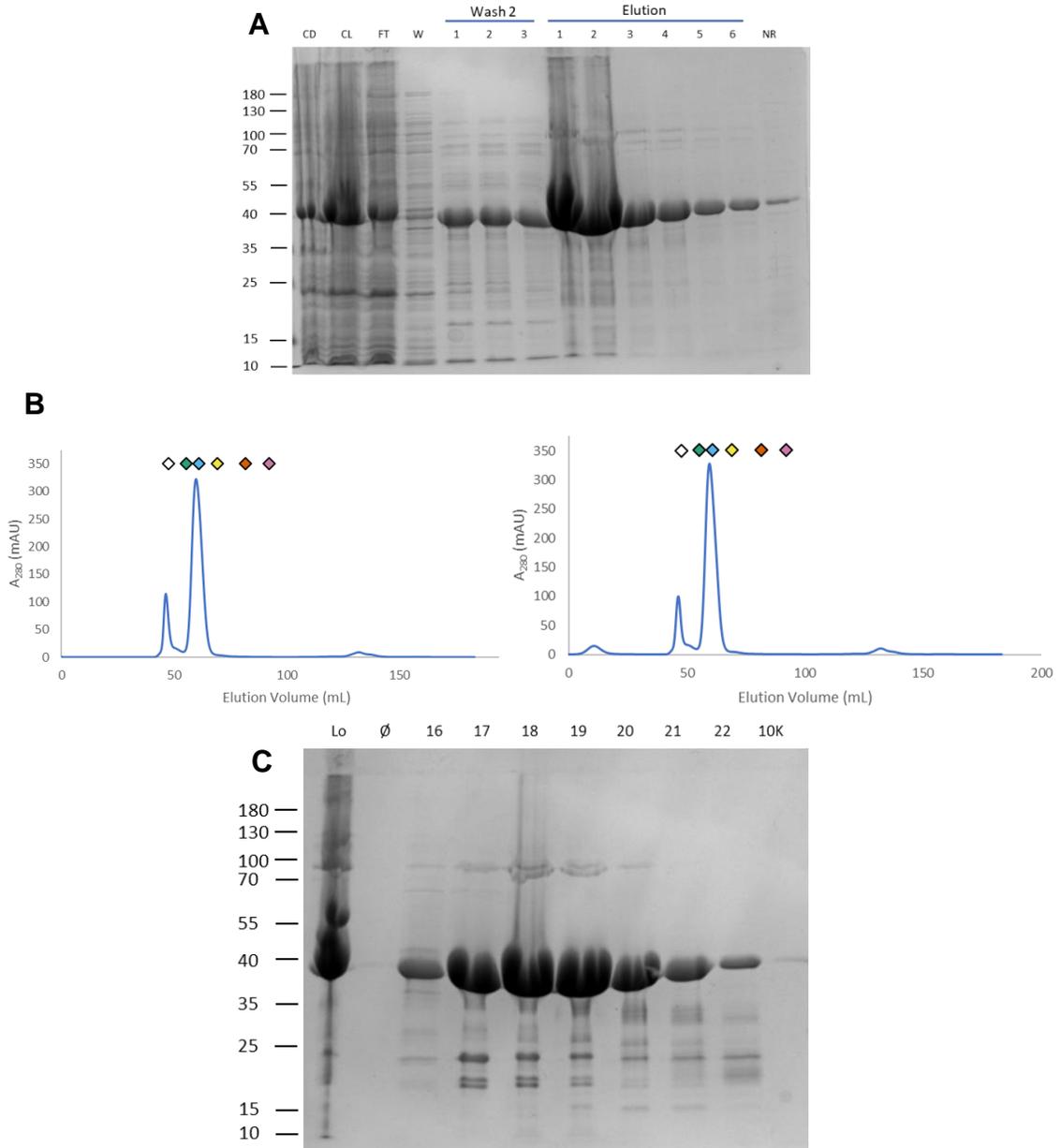
**Figure 11: Purification of *AtLaRP6C*  $\Delta$ PAM2.** (A) The protein was separated from cleared bacterial cell lysate by a Ni-NTA column. All six Elution fractions were pooled and loaded on an S200 Sephadex column. CD1 = Cell debris from pellet 1, CD2 = Cells debris from pellet 2, CL = Cleared lysate, FT = Flowthrough, W = Wash, NR = Nickel resin (B) Protein content within fractions under the peaks as indicated by the chromatogram were visualized by Coomassie stain. Fractions 20 – 27 containing the protein were pooled for ULP1 digest. Lo = protein loaded on sizing column, 10K = flowthrough from 10,000 MWCO Sartorius VivaSpin concentrator. Expected molecular weight = 61.3 kDa. (C) Chromatogram from the FPLC indicating presence of proteins by peaks. Molecular weight standards are shown above. (D) Protein content of fractions under the peaks as indicated by the chromatogram were visualized by Coomassie stain. Fractions 25 – 29 were chosen for storage. PrU = Pre-ULP1 digest, PoU = Post-ULP1 digest, Lo = protein loaded on sizing column. Expected molecular weight = 47.4 kDa. (E) After incubation with ULP1, the protein was pooled and loaded on an S200 Sephadex column to separate cleaved protein from the His10-SUMO tag and un-cleaved protein. Chromatogram from the FPLC indicating presence of protein by peaks. Molecular weight standards are shown above.

## Purification of AtLaRP6C La Module

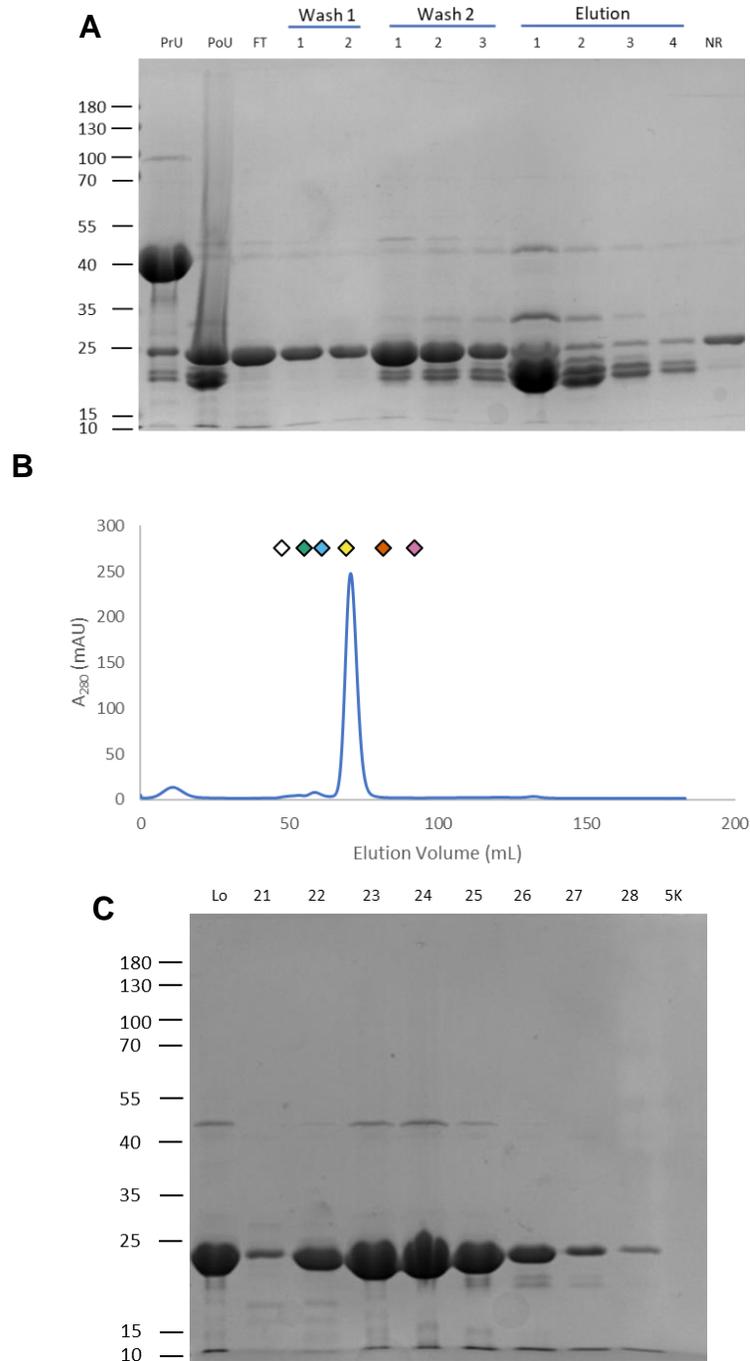
Previous purifications of the La Module were very successful, but storage conditions led to eventual degradation in the freezer as seen by a decrease in concentration over time. The current work optimized storage conditions for the FL and  $\Delta$ PAM2 constructs by adding 2% glycerol, which appears to prevent the apparent decrease in protein concentration during storage. Therefore, new AtLaRP6C protein preparations were carried out and modified to include the addition of glycerol.

Following the general purification protocol, a 1 L cell pellet was resuspended in 30 mL cold lysis buffer and lysed by sonication. In comparison to the FL and  $\Delta$ PAM2 purifications, much more protein was lost in the flowthrough and wash steps, specifically wash 2. Smearing and intense bands seen in elution fractions 1 – 2 were indicative of high protein content. As expected, a band at 40 kDa was seen through the first SDS-PAGE gel (Figure 12A). Very little protein is present in the nickel resin sample. Elution fractions 1 – 6 were pooled and concentrated on a 10,000 MWCO concentrator to 5 mL, filtered, and loaded on an S75 sizing column in two separate injections. The normal 2 mL plastic collection tubes were replaced with 6 mL lime-glass collection tubes to allow fractions from both injections to be collected in the same tubes. Protein elution was monitored by absorbance at 280 nm, and the peaks on the chromatogram were used to identify the presence of protein within the SEC fractions. Fractions under the major peak at 59.7 mL ( $MW_{app} = 53.2$  kDa) in the chromatogram (Figure 12B) were run on an SDS-PAGE gel to check for protein content (Figure

12C). Fractions 16 – 22 were pooled for ULP1 digest. The second nickel column was imperative in this purification, as the size of the cleaved protein and the His<sub>10</sub>-SUMO tag are similar in size at 26 kDa and 20 kDa, respectively, and cannot be separated by size exclusion. The nickel affinity column bound the un-cleaved protein, the His<sub>10</sub>-SUMO tag, and His<sub>6</sub>-ULP1 (Figure 13A). As the cleaved La Module was not expected to bind the nickel resin, it was anticipated to be in the flowthrough and wash fractions. The cleaved La Module was seen in all fractions, but the wash 2 fractions also contained other bands at 20 kDa similar to the bands seen in the elution fractions. As a band at this size is representative of the His<sub>10</sub>-SUMO tag, only the flowthrough and wash 1 fractions were pooled. The pooled volume was concentrated on a 5,000 MWCO concentrator to 2.5 mL and loaded on the sizing column. A very sharp peak at 70.66 mL ( $MW_{app} = 26.4$  kDa) was seen in the chromatogram (Figure 13B), and the fractions under this peak were tested for protein content. Large band at 25 kDa are visible, but additional bands under the major expected band and a band at 50 kDa were observed (Figure 13C). The band at 50 kDa is larger than the molecular weight of the tagged protein, and remains unidentified. Even so, fractions 22 – 26 were pooled, concentrated (to ~80  $\mu$ M), and brought to 5% v/v glycerol. The protein was snap-frozen as 50  $\mu$ L aliquots and stored at -70°C.



**Figure 12: Purification of His<sub>10</sub>-SUMO-AflLaRP6C La Module. (A)** The protein was separated from cleared bacterial cell lysate by a Ni-NTA column. All six Elution fractions were pooled and loaded on an S75 Sephadex column. CD = Cell debris from pellet, CL = Cleared lysate, FT = Flowthrough, W = Wash, NR = Nickel resin **(B)** Chromatograms from the FPLC indicating presence of proteins by peaks. First injection (left) and second injection (right) of protein on sizing column. Molecular weight standards are shown above. **(C)** Protein content within fractions under the peaks as indicated by the chromatogram were visualized by Coomassie stain. Fractions 16 – 22 containing the protein were pooled for ULP1 digest. Lo = protein loaded on sizing column. Expected molecular weight = 36.2 kDa.

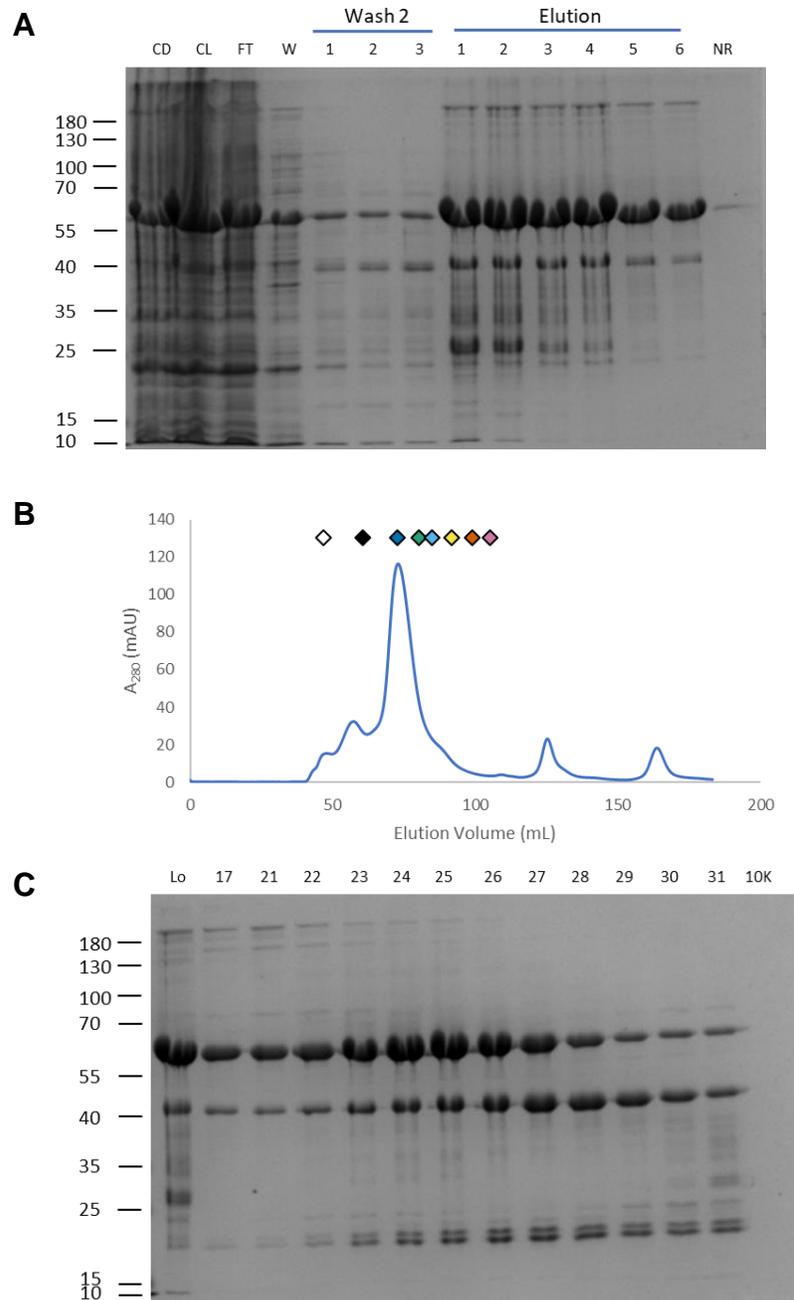


**Figure 13: Purification of AtLaRP6C La Module (A)** After incubation with ULP1, the protein was separated from the His10-SUMO tag by a Ni-NTA column. FT, Wash 1, and Wash 2 fractions 1-3 were pooled and loaded on an SEC column. PrU = Pre-ULP1 digest, FT = flowthrough, NR = Nickel resin. **(B)** Chromatogram from the FPLC indicating presence of protein by peaks. Molecular weight standards are shown above. **(C)** Protein content of fractions under the peaks as indicated by the chromatogram were visualized by Coomassie stain. Fractions 22 – 26 were chosen for storage. Expected molecular weight = 23 kDa.

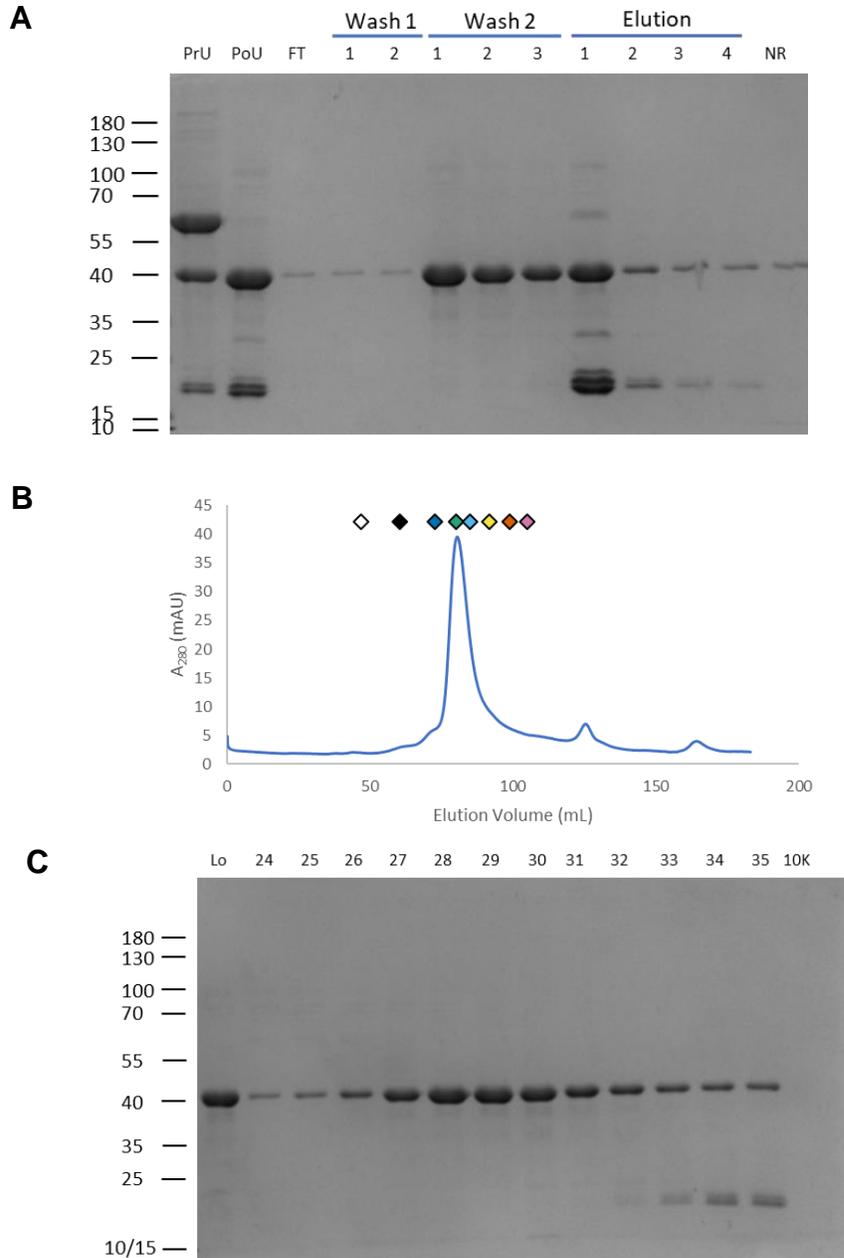
## Purification of AtLaRP6C $\Delta$ CTD

Beginning from the standard purification protocol, a 1 L cell pellet was prepped and lysed by sonication. After loading onto a nickel affinity column fractions were collected, and all samples were run on an SDS-PAGE gel visualized with Coomassie stain. The majority of the protein was released into the cleared lysate, but some protein was still visible with the cell debris. A loss of protein was seen in the flowthrough and wash fractions, but bands of equal intensity were seen in the elution fractions while very minimal protein remained on the nickel resin (Figure 14A). Elution fractions 1 – 6 were concentrated on a 10,000 MWCO concentrator to 2.5 mL and filtered (0.2  $\mu$ m) before loading on an S200 column. Protein elution was monitored by absorbance at 280 nm, and the peaks on the chromatogram were used to identify the presence of protein within the SEC fractions. A shoulder at the void-volume of the column was seen in the chromatogram with a major peak at 72.98 mL ( $MW_{app} = 148.1$  kDa) (Figure 14B). Fractions under these peaks were run on an SDS-PAGE gel and visualized with Coomassie stain (Figure 14C). An expected band at 60 kDa was seen with a secondary band at 40 kDa. Fractions 21 – 28 were pooled for ULP1 digest and subjected to another nickel affinity column to remove any un-cleaved protein, tags, and the ULP1 protease. Comparison of the pre-ULP1 and post-ULP1 fractions show the loss of the major band at 60 kDa and an increase in the band at 40 kDa (Figure 15A). The post-ULP1 sample also contained bands at 20 kDa associated with the SUMO tag. As the cleaved protein was expected to not bind to the nickel column, the flowthrough and wash 1 and wash 2 fractions were

pooled, concentrated as previously mentioned, and loaded again on to the sizing column. The cleaved protein came off the nickel column most readily after the addition of wash 2 as can be seen by the intensity of the 40 kDa bands in these samples. Additionally, the bands at 20 kDa associated with the SUMO tag was seen as expected in the elution fractions, although some cleaved protein was also lost in these fractions. Following the second sizing column, the fractions under the peak at 80.5 mL ( $MW_{app} = 73.3$  kDa) (Figure 15B) were tested for protein content by SDS-PAGE gel (Figure 15C) and fractions 26 – 32 were pooled, concentrated (to  $\sim 50$   $\mu$ M), and brought to 5% v/v glycerol. The protein was snap-frozen as 20  $\mu$ L aliquots and stored at  $-70^{\circ}\text{C}$ .



**Figure 14: Purification of His<sub>10</sub>-SUMO-A $\beta$ LARP6C  $\Delta$ CTD. (A)** The protein was separated from cleared bacterial cell lysate by a Ni-NTA column. All six Elution fractions were pooled and loaded on an S75 Sephadex column. CD = Cell debris from pellet, CL = Cleared lysate, FT = Flowthrough, W = Wash, NR = Nickel resin **(B)** Chromatogram from the FPLC indicating presence of proteins by peaks. Molecular weight standards are shown above. **(C)** Protein content within fractions under the peaks as indicated by the chromatogram were visualized by Coomassie stain. Fractions 21 – 28 containing the protein were pooled for ULP1 digest. Lo = protein loaded on sizing column. Expected molecular weight = 50 kDa.

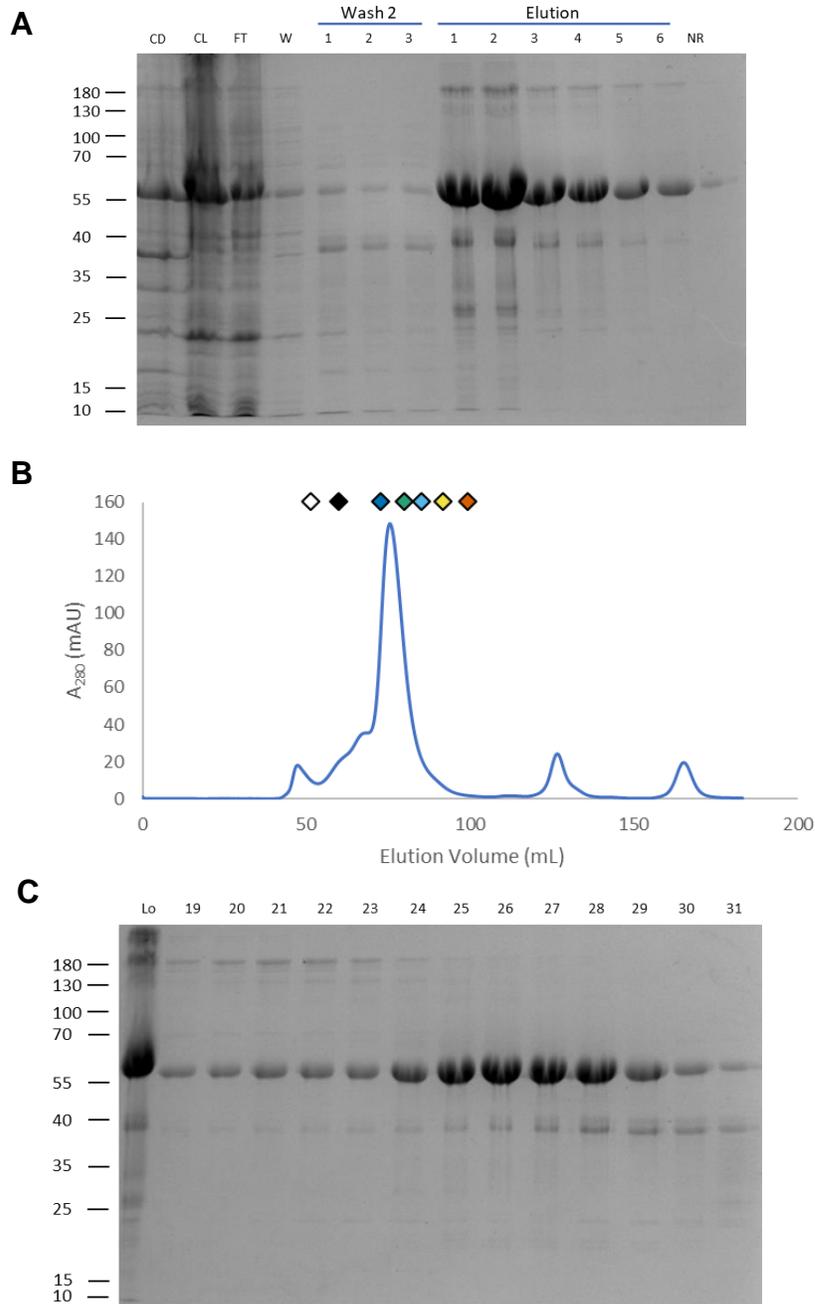


**Figure 15: Purification of *AtLaRP6C*  $\Delta$ CTD (A)** After incubation with ULP1, the protein was separated from the His10-SUMO tag by a Ni-NTA column. FT, Wash 1, and Wash 2 fractions 1-3 were pooled and loaded on an SEC column. PrU = Pre-ULP1 digest, FT = flowthrough, NR = Nickel resin. **(B)** Chromatogram from the FPLC indicating presence of protein by peaks. Molecular weight standards are shown above. **(C)** Protein content of fractions under the peaks as indicated by the chromatogram were visualized by Coomassie stain. Fractions 26 – 32 were chosen for storage. Expected molecular weight = 36.2 kDa.

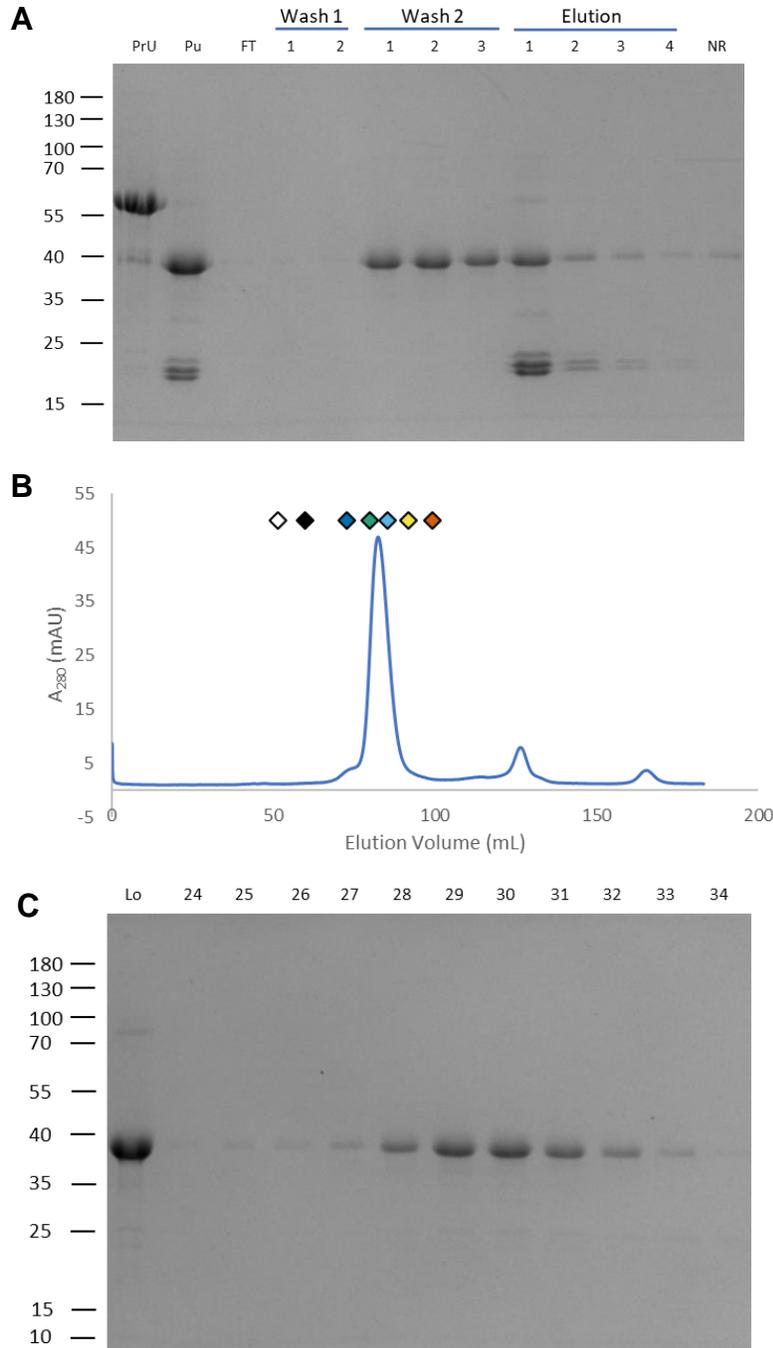
## Purification of AtLaRP6C $\Delta$ PAM2 $\Delta$ CTD

The purification of the  $\Delta$ CTD and  $\Delta$ PAM2  $\Delta$ CTD constructs were very similar to one another, suggesting that the contribution of the PAM2 motif is not essential to protein integrity. The standard purification protocol was used, as described above. As expected, the protein eluted from the nickel column in the elution fractions (Figure 16A). Elution fractions 1 – 6 were concentrated on a 10,000 MWCO concentrator to 2.5 mL and filtered (0.2  $\mu$ m) before loading on an S200 column. Protein elution was monitored by absorbance at 280 nm, and the peaks on the chromatogram were used to identify the presence of protein within the SEC fractions. A shoulder at the void-volume of the column was seen in the chromatogram with a major peak at 75.57 mL ( $MW_{app} = 114.6$  kDa) (Figure 16B). Fractions under these peaks were run on an SDS-PAGE gel and visualized with Coomassie stain (Figure 16C). An expected band at 60 kDa was seen with a secondary band at 40 kDa. Fractions 24 – 29 were pooled for ULP1 digest and subjected to another nickel affinity column to remove any un-cleaved protein, tags, and the ULP1 protease. Comparison of the pre-ULP1 and post-ULP1 fractions show the loss of the major band at 60 kDa and an increase in the band at 40 kDa (Figure 17A). The post-ULP1 sample also contained bands at 20 kDa associated with the SUMO tag. As the cleaved protein was expected to not bind to the nickel column, the flowthrough and wash 1 and wash 2 fractions were pooled, concentrated as previously mentioned, and loaded again on to the sizing column. The cleaved protein came off the nickel column most readily after the addition of wash 2 as can be seen by the intensity of the 40 kDa bands in these

samples. Additionally, the bands at 20 kDa associated with the SUMO tag was seen as expected in the elution fractions, although some cleaved protein was also lost in these fractions. Following the second sizing column, the fractions under the peak at 82.41 mL ( $MW_{app} = 62.4$  kDa) (Figure 17B) were tested for protein content by SDS-PAGE gel (Figure 17C) and fractions 27 – 33 were pooled, concentrated (to  $\sim 30$   $\mu$ M), and brought to 5% v/v glycerol. The protein was snap-frozen as 20  $\mu$ L aliquots and stored at  $-70^{\circ}\text{C}$ .



**Figure 16: Purification of His<sub>10</sub>-SUMO-AfLaRP6C  $\Delta$ PAM2  $\Delta$ CTD. (A)** The protein was separated from cleared bacterial cell lysate by a Ni-NTA column. All six Elution fractions were pooled and loaded on an S75 Sephadex column. CD = Cell debris from pellet, CL = Cleared lysate, FT = Flowthrough, W = Wash, NR = Nickel resin **(B)** Chromatogram from the FPLC indicating presence of proteins by peaks. Molecular weight standards are shown above. **(C)** Protein content within fractions under the peaks as indicated by the chromatogram were visualized by Coomassie stain. Fractions 24 – 29 containing the protein were pooled for ULP1 digest. Lo = protein loaded on sizing column. Expected molecular weight = 48.4 kDa.

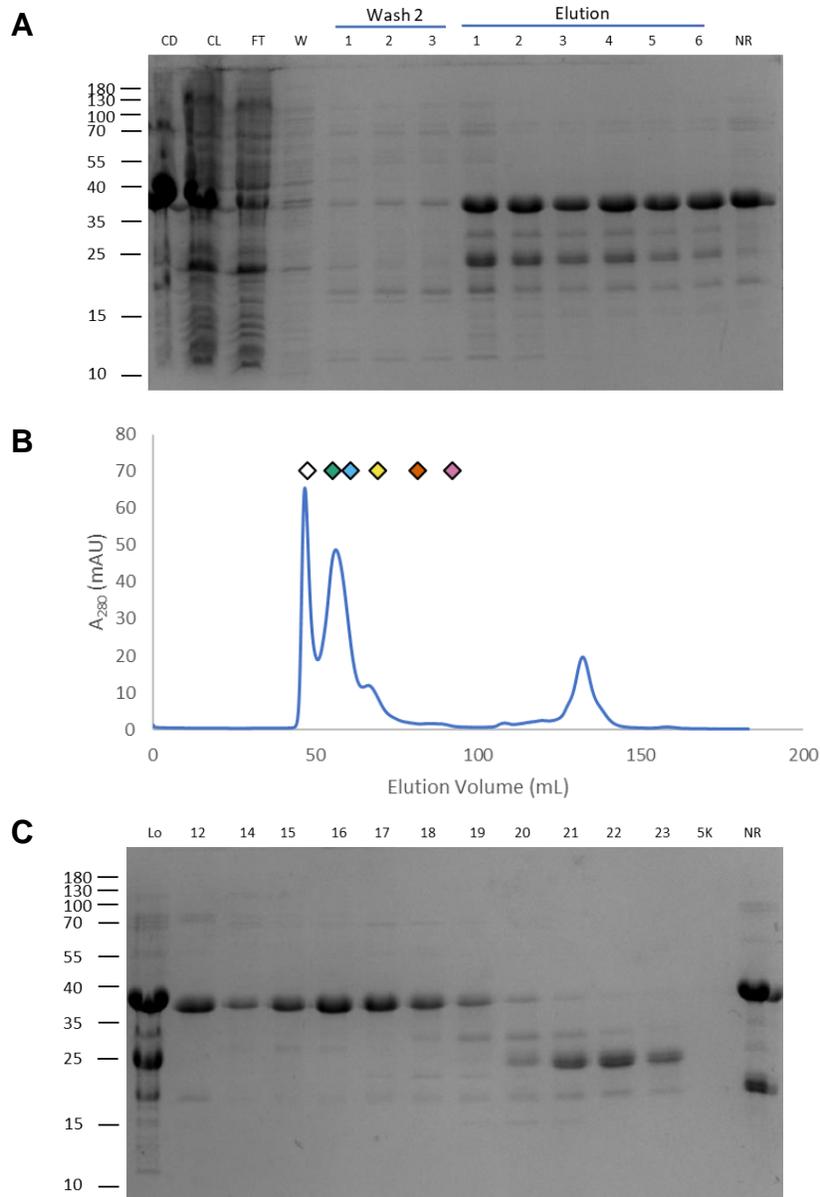


**Figure 17: Purification of *AtLaRP6C*  $\Delta$ PAM2  $\Delta$ CTD. (A)** After incubation with ULP1, the protein was separated from the His10-SUMO tag by a Ni-NTA column. FT, Wash 1, and Wash 2 fractions 1-3 were pooled and loaded on an SEC column. PrU = Pre-ULP1 digest, FT = flowthrough, NR = Nickel resin. **(B)** Chromatogram from the FPLC indicating presence of protein by peaks. Molecular weight standards are shown above. **(C)** Protein content of fractions under the peaks as indicated by the chromatogram were visualized by Coomassie stain. Fractions 27 – 33 were chosen for storage. Expected molecular weight = 34.6 kDa.

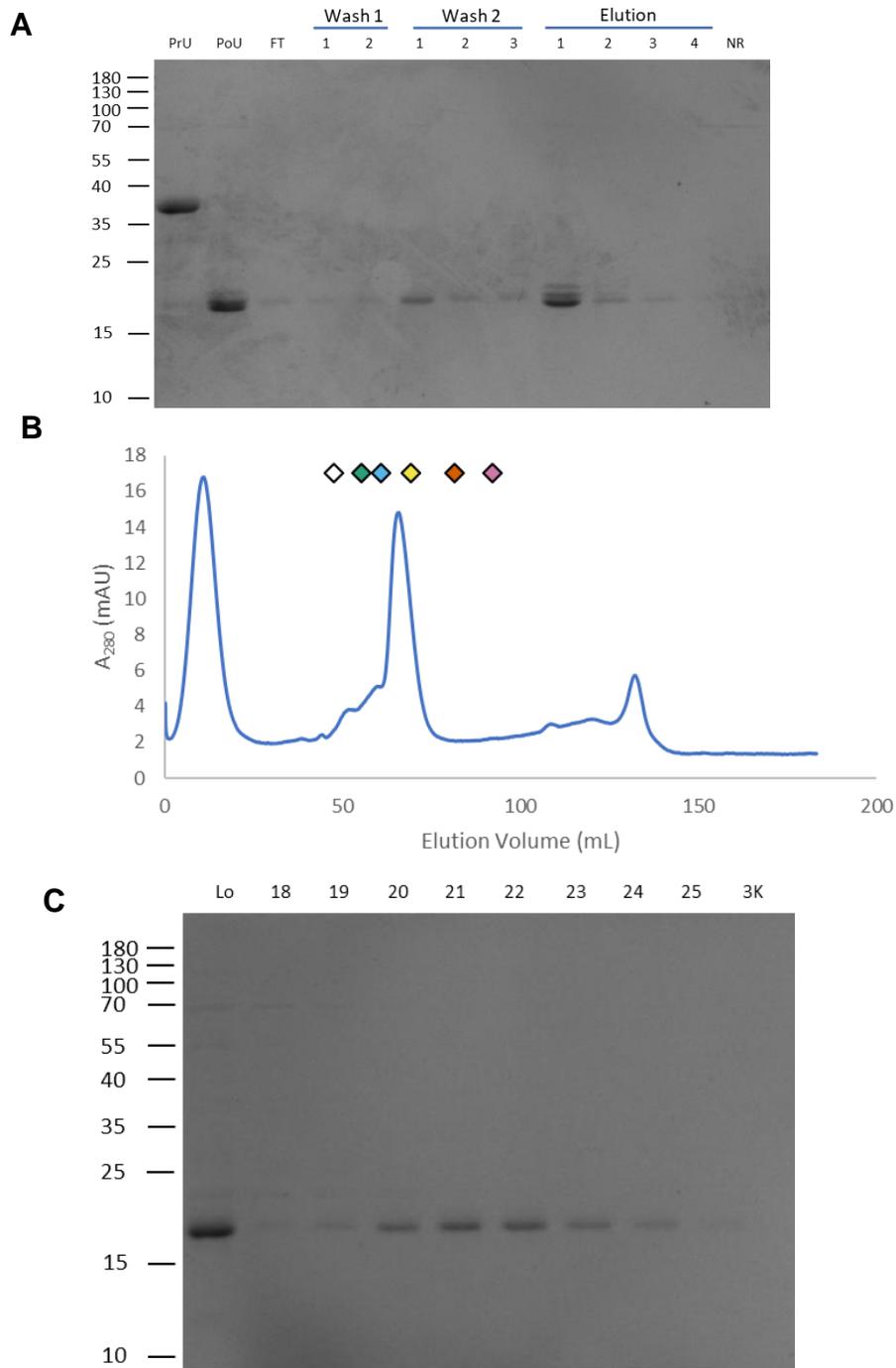
## Purification of AtLaRP6C NTR

Purification began using the typical method, and the cleared lysate was subjected to nickel affinity chromatography. Notably, each band in the elution fractions were equivalent in intensity and a large amount of protein remained on the nickel resin (Figure 18A). Elution fractions 1 – 6 were concentrated on a 5,000 MWCO concentrator to 2.5 mL and filtered (0.2  $\mu$ m) before loading on an S75 column. Protein elution was monitored by absorbance at 280 nm, and the peaks on the chromatogram were used to identify the presence of protein within the SEC fractions. A large peak at the void-volume of the column was seen in the chromatogram with a major peak at 56.33 mL ( $MW_{app} = 66$  kDa) (Figure 18B). Fractions under these peaks were run on an SDS-PAGE gel and visualized with Coomassie stain (Figure 18C). An expected band at 38 kDa was seen with a secondary band at 25 kDa. A 500 mM imidazole solution was used to elute any remaining protein off the column and is seen in the nickel resin sample (Figure 18C). Fractions 14 – 19 were pooled for ULP1 digest and subjected to another nickel affinity column to remove any un-cleaved protein, tags, and the ULP1 protease. Comparison of the pre-ULP1 and post-ULP1 fractions show the loss of the major band at 38 kDa and an increase in a band at 20 kDa (Figure 19A). This band at 20 kDa could not be distinguished as the isolated NTR or the tag. As the cleaved protein was expected to not bind to the nickel column, the flowthrough, wash 1, and wash 2 fractions were pooled, concentrated as previously mentioned, and loaded again on to the sizing column. Following the second sizing column, the fractions under the peak (65.81 mL,  $MW_{app} = 35.9$  kDa)

(Figure 19B) were tested for protein content by SDS-PAGE gel (Figure 19C) and fractions 19 – 24 were pooled, concentrated (to ~20  $\mu$ M), and brought to 5% v/v glycerol. The protein was snap-frozen as 40  $\mu$ L aliquots and stored at -70°C.



**Figure 18: Purification of His<sub>10</sub>-SUMO-A<sub>t</sub>LaRP6C NTR. (A)** The protein was separated from cleared bacterial cell lysate by a Ni-NTA column. All six Elution fractions were pooled and loaded on an S75 Sephadex column. CD = Cell debris from pellet, CL = Cleared lysate, FT = Flowthrough, W = Wash, NR = Nickel resin **(B)** Chromatogram from the FPLC indicating presence of proteins by peaks. Molecular weight standards are shown above. **(C)** Protein content within fractions under the peaks as indicated by the chromatogram were visualized by Coomassie

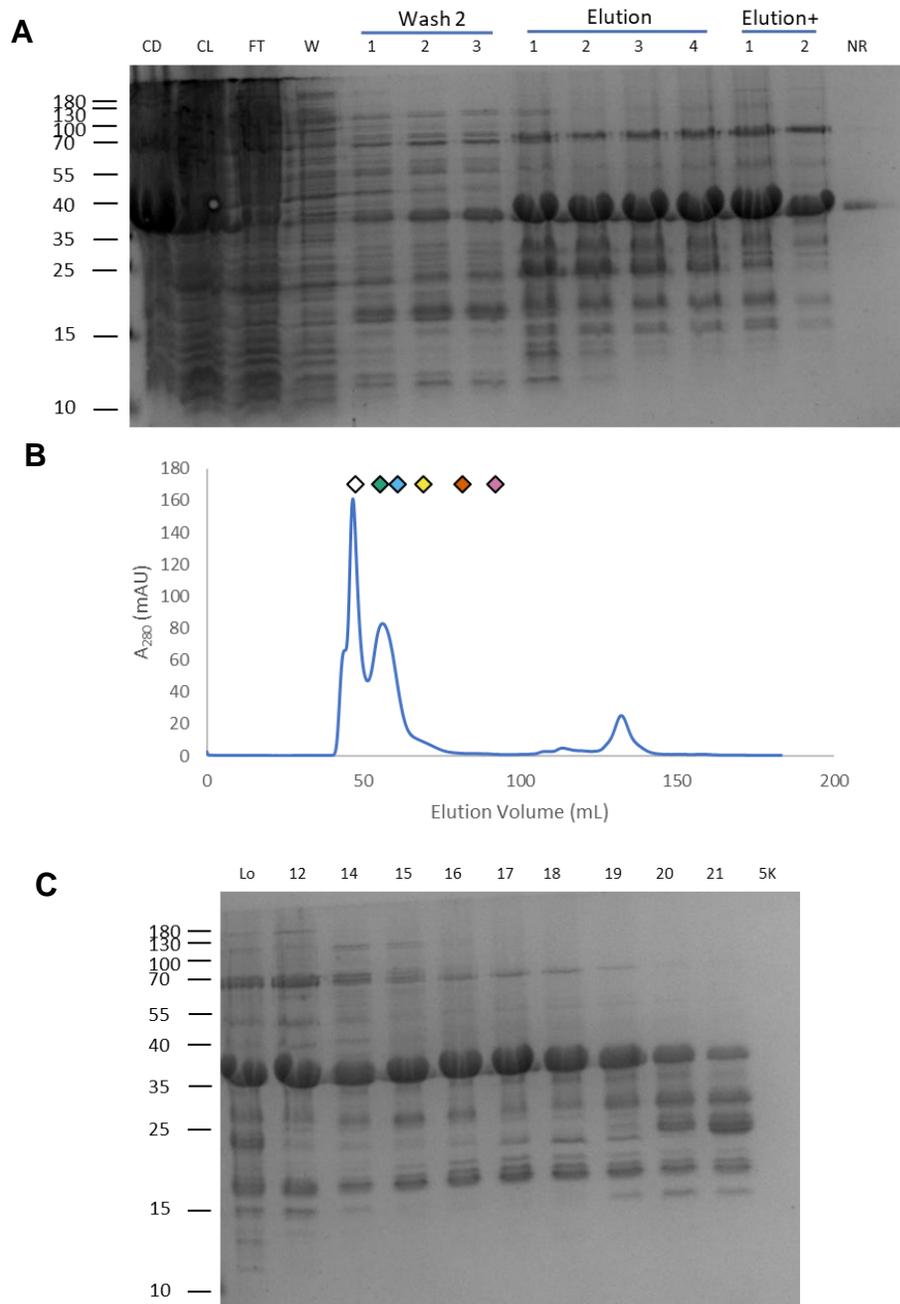


**Figure 19: Purification of AtLaRP6C NTR. (A)** After incubation with ULP1, the protein was separated from the His10-SUMO tag by a Ni-NTA column. FT, Wash 1, and Wash 2 fractions 1-3 were pooled and loaded on an SEC column. PrU = Pre-ULP1 digest, FT = flowthrough, NR = Nickel resin. **(B)** Chromatogram from the FPLC indicating presence of protein by peaks. Molecular weight standards are shown above. **(C)** Protein content of fractions under the peaks as indicated by the chromatogram were visualized by Coomassie stain. Fractions 19 – 24 were chosen for storage. Expected molecular weight = 14 kDa.

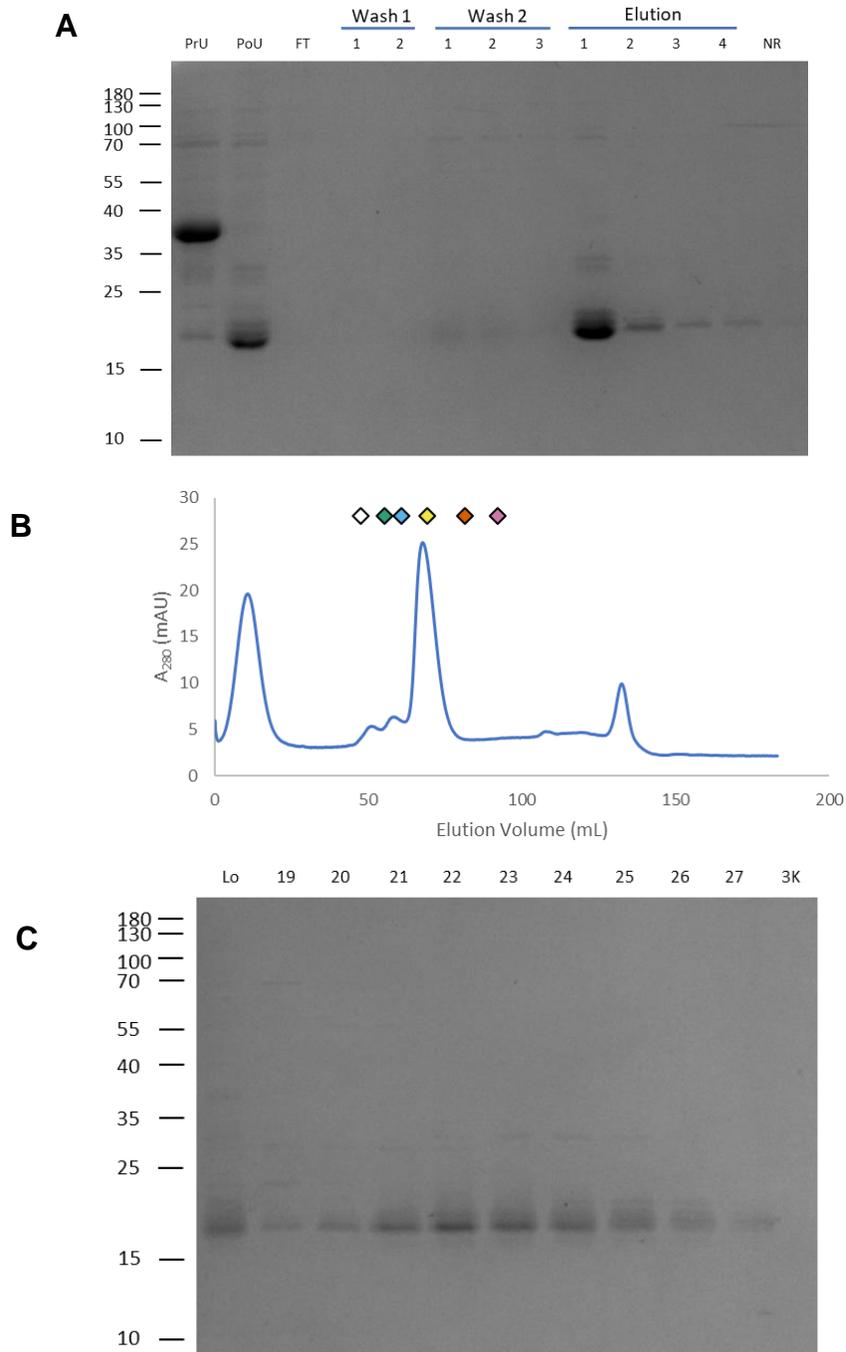
## Purification of AtLaRP6C ΔPAM2 NTR

Following the purification of the NTR, an “elution+” buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>/Na<sub>2</sub>HPO<sub>4</sub> (pH 8.0), 200 mM NaCl, 50 mM Na<sub>2</sub>SO<sub>4</sub>, 1 mM DTT, 500 mM imidazole (pH 8.0)) was prepared in an attempt to fully elute the ΔPAM2 NTR protein from the nickel resin in the event that it did not fully elute. The extremely high total protein content caused smearing in the gel for the cell debris, cleared lysate, and flowthrough samples (Figure 20A). Some ΔPAM2 NTR protein was lost in the wash steps, and as expected, most protein eluted during the first elution buffer. The intensity of the bands are the same in the elution 1 – 4 fractions, and the additional “elution+” fractions also contain bands of the same intensity. Unlike the previous NTR purification, the protein was mostly eluted from the nickel resin. Elution fractions 1 – 4 and elution+ fractions 1-2 were concentrated on a 5,000 MWCO concentrator to 2.5 mL and filtered (0.2 μm) before loading on a S75 column. A large peak at the void-volume of the column was seen in the chromatogram with a major peak at 56.18 mL ( $MW_{app} = 66.7$  kDa) (Figure 20B). Fractions under these peaks were run on an SDS-PAGE gel and visualized with Coomassie stain (Figure 20C). An expected band at 38 kDa was seen with a secondary bands at 28 and 17 kDa. Fraction 12 is within the void-volume of the column, the band at 38 kDa is the tagged ΔPAM2 NTR showing the protein is contained in both peaks. Fractions 14 – 19 were pooled for ULP1 digest and subjected to another nickel affinity column to remove any un-cleaved protein, tags, and the ULP1 protease. Comparison of the pre-ULP1 and post-ULP1 fractions show the loss of the major band at 38 kDa and an increase

in a band at 20 kDa (Figure 21A). This band at 20 kDa could not be distinguished as the isolated  $\Delta$ PAM2 NTR or the tag. Notably, the intensity of the band in the pre-ULP1 sample is much darker than the band in the post-ULP1 sample. As the cleaved protein was expected to not bind to the nickel column, but due to the intensity of the bands, only wash 2 fractions were pooled, concentrated as previously mentioned, and loaded again on to the sizing column. An intense band is seen in the first elution fraction, but this could be either the protein of interest or the tag. Following the second sizing column, the fractions under the peak at 67.6 mL ( $MW_{app} = 32$  kDa) (Figure 21B) were tested for protein content by SDS-PAGE gel (Figure 21C) and fractions 20 – 24 were pooled, concentrated (to  $\sim 28$   $\mu$ M), and brought to 5% v/v glycerol. The protein was snap-frozen as 30  $\mu$ L aliquots and stored at  $-70^{\circ}\text{C}$ .



**Figure 20: Purification of His<sub>10</sub>-SUMO-AtlLaRP6C ΔPAM2 NTR.** (A) The protein was separated from cleared bacterial cell lysate by a Ni-NTA column. All six Elution fractions were pooled and loaded on an S75 Sephadex column. CD = Cell debris from pellet, CL = Cleared lysate, FT = Flowthrough, W = Wash, NR = Nickel resin (B) Chromatogram from the FPLC indicating presence of proteins by peaks. Molecular weight standards are shown above. (C) Protein content within fractions under the peaks as indicated by the chromatogram were visualized by Coomassie stain. Fractions 14 – 19 containing the protein were pooled for ULP1 digest. Lo = protein loaded on sizing column. Expected molecular weight = 26.2 kDa.



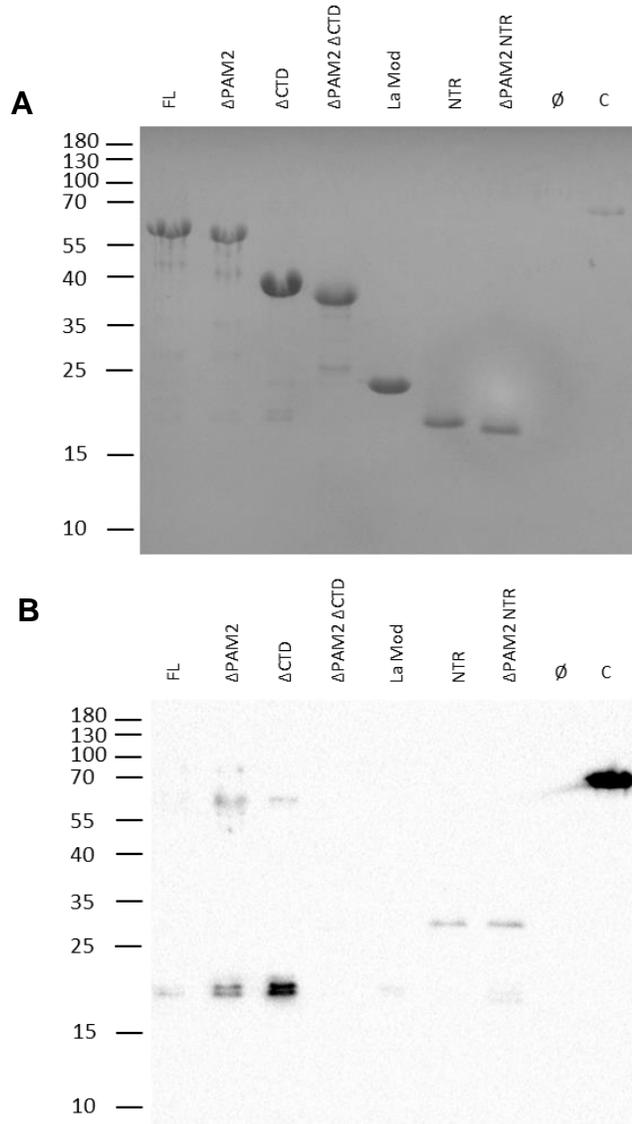
**Figure 21: Purification of *AtLaRP6C*  $\Delta$ PAM2 NTR.** (A) After incubation with ULP1, the protein was separated from the His10-SUMO tag by a Ni-NTA column. FT, Wash 1, and Wash 2 fractions 1-3 were pooled and loaded on an SEC column. PrU = Pre-ULP1 digest, FT = flowthrough, NR = Nickel resin. (B) Chromatogram from the FPLC indicating presence of protein by peaks. Molecular weight standards are shown above. (C) Protein content of fractions under the peaks as indicated by the chromatogram were visualized by Coomassie stain. Fractions 20 – 24 were chosen for storage. Expected molecular weight = 12.4 kDa.

## Determining Purity of Purified Proteins

The NTR constructs have approximately the same molecular weight as the His<sub>10</sub>-SUMO tag. Therefore, after the purification of the NTR constructs, all proteins intended for biochemical analyses were evaluated by anti-His western blot. Ponceau stain confirmed that all proteins were present and migrated in the SDS-PAGE at their expected molecular weights with minimal degradation after storage (Figure 22A). The Ponceau stain was washed away and the membrane then blocked and probed for His-tagged proteins (Figure 22B). For this analysis, sample load volumes were not normalized for protein amount.

A His-tagged positive control was used in the far-right lane, and is only faintly visible in the Ponceau-stained membrane. As expected, it presents an intense band at the same molecular weight when detected by chemiluminescence. For all of the purified *At*LaRP6C proteins, there are anti-His-reactive bands at various molecular weights within each sample. However, the relative intensity of these bands in comparison to the intensity of the major Ponceau-stained bands are minimal. In particular, the  $\Delta$ PAM2 $\Delta$ CTD and La Module constructs appear extremely pure. Additionally, the NTR constructs were purified successfully and were not the His<sub>10</sub>-SUMO tag. The relative percent purity of each sample in the Ponceau stained membrane was quantified by ImageLab (Biorad): full-length at 89%,  $\Delta$ PAM2 at 83%,  $\Delta$ CTD at 94%,  $\Delta$ PAM2  $\Delta$ CTD at 90%, NTR at 100%,  $\Delta$ PAM2 NTR at 100%, La Module at 100%. For the purposes of preliminary structural stability and RNA binding activity assays, it

was determined that these protein preparations were acceptably pure to continue with further biochemical analyses.



**Figure 22: Checking for presence of His<sub>10</sub>-SUMO tag in purified proteins.** 5× SDS sample buffer was added to purified protein aliquots to a final concentration of 1×. Samples were run on a 13% SDS-PAGE gel and transferred to a 1× transfer buffer equilibrated nitrocellulose membrane. After transfer the membrane was treated with 1× Ponceau Red (A) (Brightness +20%, Contrast +20%). Excess stain was removed, and detection was continued as normal to visualize any His<sub>10</sub>-SUMO tag (B). C = AtLaRP6A purified by Daniel Horn.

## Discussion and Conclusions

Although all mutations were successfully produced in the expression constructs, the efficiency of the molecular biology to generate the mutations was never 100%. Some mutations contained repeated segments of the primer due to oligomerization in the thermocycler, while some did not contain any mutations and were copies of the template DNA. Optimizing the best parameters was difficult because even though the same template was being utilized, each mutation required different conditions. This issue stems from the complementarity of the primers to each other as opposed to the template DNA. Another member in lab created a different set of mutagenic primers which had mutagenic complementarity but a 5' overhang in either direction to prevent the primers from annealing to each other. This worked with much greater efficiency and this method should be used in future mutagenesis protocols that use the Phusion DNA polymerase.

The most significant changes to the established expression protocols for the previously prepared *AtLaRP6C* constructs was increasing the optimal OD600 from 0.5 – 0.8 to ~0.7. Additionally, all expression data indicated that the proteins expressed stably between 16 – 20 hours at 16°C with very little degradation products or changes in protein yield, in contrast to the prior protocols that expressed for 21 hours. This allowed for a significant increase in protein expression with more protocol flexibility in the overall large-scale expression of these proteins. There was no noticeable difference in the expression profile of any of the constructs

Previous purifications of the FL and  $\Delta$ PAM2 proteins resulted in little to no protein. With a goal of obtaining higher amounts of protein, two 1-L cell pellets were used per purification. Resuspending and sonicating each cell pellet separately was determined to be the best way to ensure proper sonication of the cells. While initial purification protocols for these proteins used a sonication amplitude of 30-40%, 37% was found to release more protein from the cells without compromising protein recovery. Although higher amplitudes could be better, these were not tested as optimization of these conditions were not the major focus of this project.

During the purification of the  $\Delta$ CTD and  $\Delta$ PAM2  $\Delta$ CTD constructs, it was observed that the His<sub>10</sub>-SUMO-tagged proteins obtained a reddish hue during the centrifugal concentration preparation for size exclusion chromatography. This was distinct from the untagged proteins and is hypothesized to be some metal binding activity, potentially iron. This color was visualized most vividly in the first concentrating step but could faintly be seen throughout the purification. Even the untagged La Module, which had the highest overall protein content during purification, was completely void of color throughout the entire purification protocol.

All of the *At*LaRP6C proteins used in this project have some intrinsic affinity for the nickel resin, as 50 mM imidazole must be used to elute the protein from the nickel column after SUMO-cleavage. Cleaved protein is expected to elute in the flowthrough and wash 1, but the greatest protein elution was seen after the addition of wash 2. This inherent attraction for the metal resin is also

observed in the amount of cleaved, untagged protein that was eluted from the second nickel column under high concentrations of imidazole. Therefore, extreme care must be taken to maximize the elution of the cleaved protein while simultaneously maximizing the separation from the His-tag containing proteins in the mixture (His<sub>10</sub>-SUMO, un-cleaved tagged protein, and the ULP1 protease).

Compared to the longer proteins, the purification of the NTR and  $\Delta$ PAM2 NTR was more challenging due to their smaller size. Interestingly, these proteins exhibited strong intrinsic affinity for the nickel resin, as evidenced by the sustained intensity of the bands over the course of the elution fractions. Although these proteins seemed to interact with the nickel resin more readily, there was no noticeable color change in the concentrated proteins, like that observed for the  $\Delta$ CTD proteins. The chromatograms from the sizing columns also showed a large peak in the void-volume of the column, even higher than the peaks used for further purification. This indicates aggregation of the protein, and indeed significant amounts of the protein were lost during the first sizing column. In future purifications, this phenomenon may need to be mitigated by decreasing the degree of concentration and instead carrying out multiple injections.

In all purifications, the ULP1 digest is performed at 16°C for 2 hours. During this time, almost 100% cleavage is accomplished, but is accompanied by a loss in recovery of the LaRP protein. A reasonable explanation is that the cleavage of the SUMO tag compromises the structural stability of the protein; alternatively, the increase in temperature of the protein solution from 4°C to 16°C may destabilize the protein. This phenomenon was most readily seen in the NTR

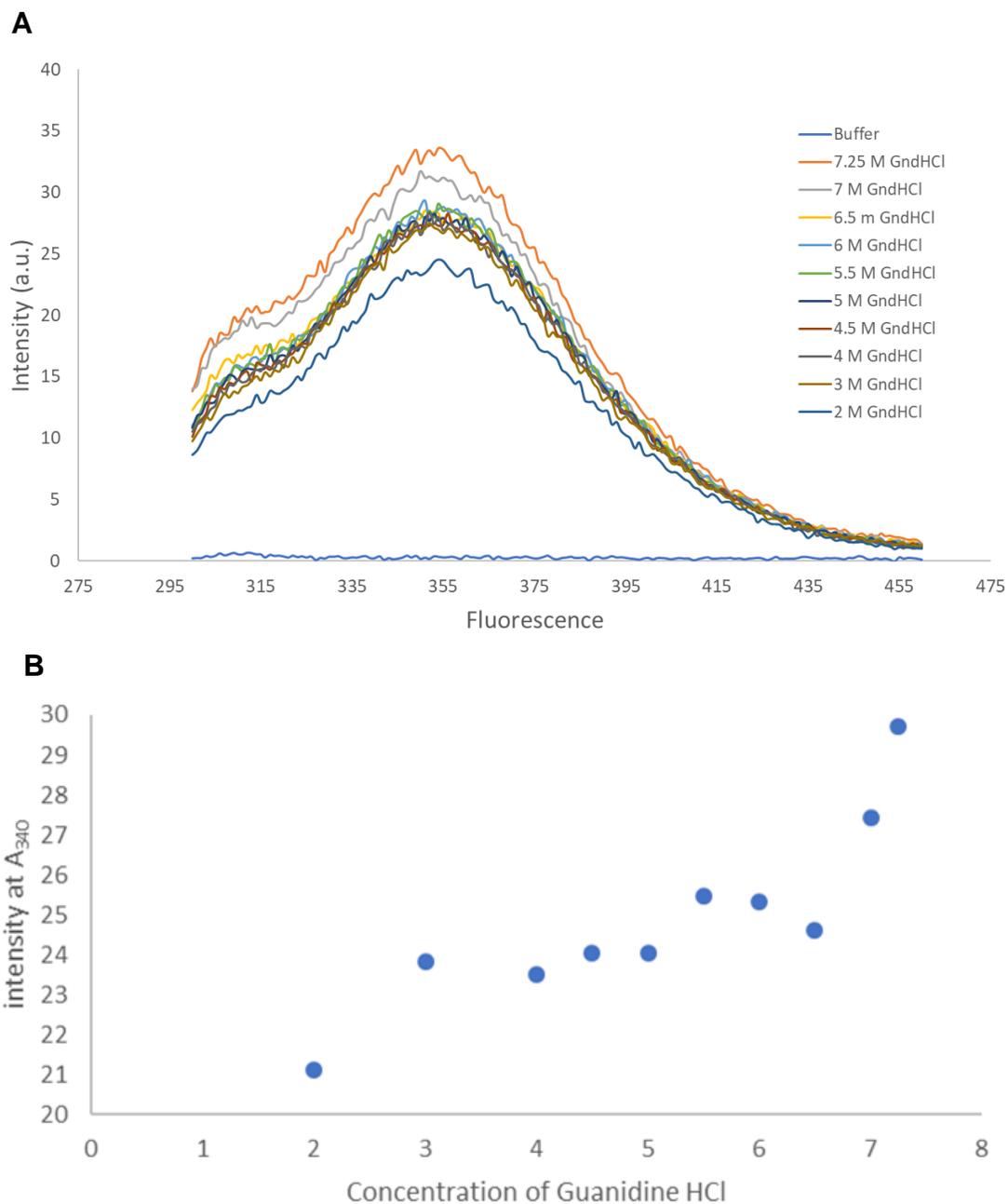
and  $\Delta$ PAM2 NTR purifications, where a significant decrease in the overall intensity of the bands is observed when comparing the pre-ULP1 and post-ULP1 samples in the SDS-PAGE gel (Figures 19A, 21A). In the future, a longer ULP1 digest at a lower temperature should be explored to determine if this protein loss can be mitigated.

#### **IV. STRUCTURE AND FUNCTION OF *AtLaRP6C* VARIANTS**

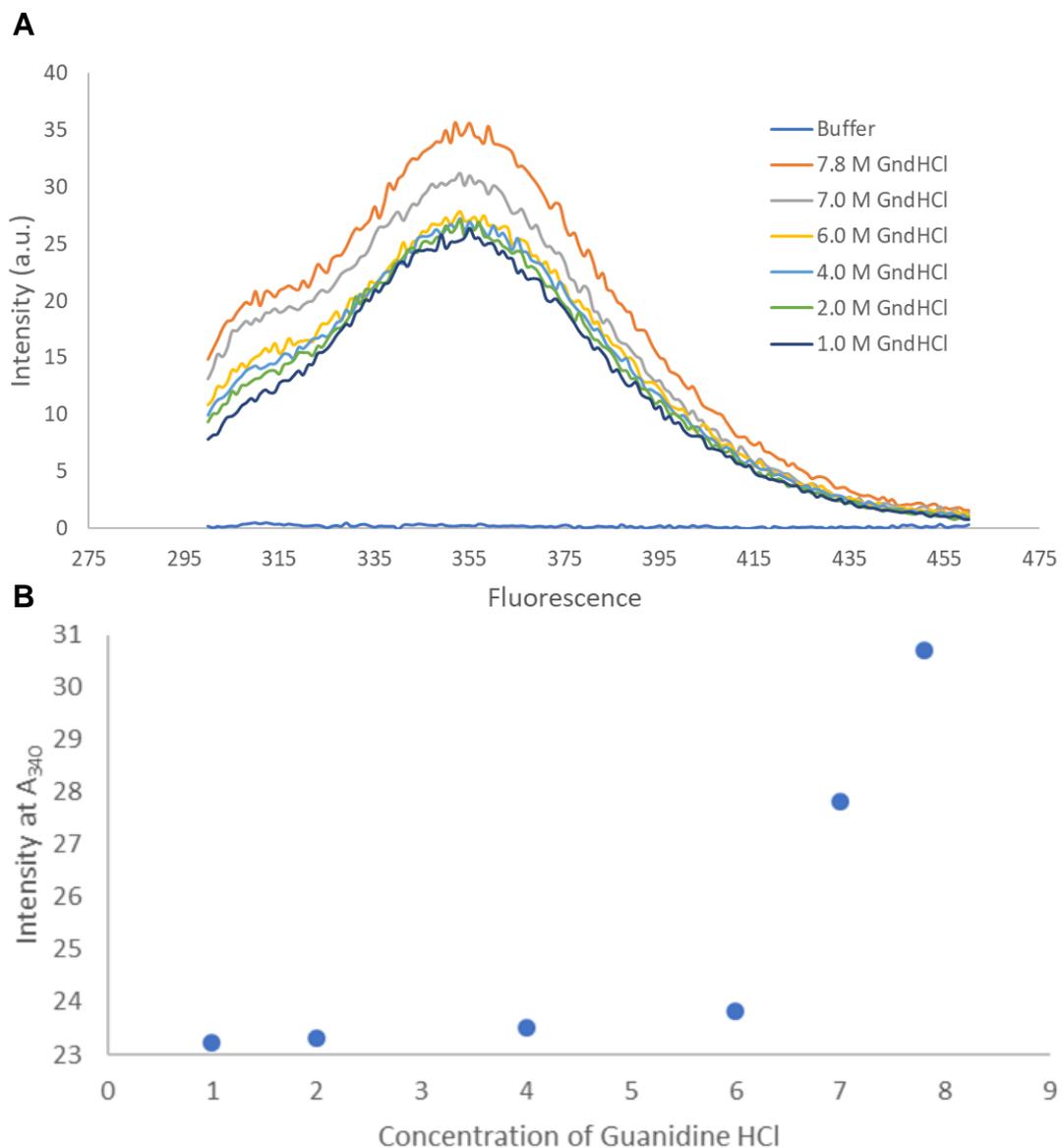
There are two aims with the purified *AtLaRP6C* proteins. First, we sought to determine if the NTR and/or the PAM2 motif contribute to overall protein stability. The second goal, was to test whether either element modulates the RNA binding activity and specificity of *AtLaRP6C*.

##### **Guanidinium Hydrochloride Protein Denaturation Assays**

As the purified La Module was obtained in the greatest amounts, and due to the previous work in the field, it was used to develop the structural stability assays. Guanidinium hydrochloride (GndHCl) is a strong chaotropic agent and denatures proteins over time. The La Module was incubated for 1 hour at 25°C with various concentrations of GndHCl, and an emission spectra from 300-460 was obtained (Figure 23A). The intensity at 340 nm was chosen as it is the expected emission maximum of tryptophan. The fluorescence was plotted as a function of the concentration of GndHCl, with the expectation of a sigmoidal curve as protein unfolding is a cooperative process. In the first test against the La Module, an increasing intensity was seen with a non-sigmoidal correlation to concentration of GndHCl (Figure 23B). Most importantly, saturation of the denatured state is not occurring which would typically be indicated by flattening of the signal at increasing concentrations of GndHCl. To further destabilize the protein, higher concentrations of GndHCl were used and the reactions were incubated at 30°C. The results of this experiment were the same as the previous experiment, with a non-sigmoidal shape and no saturation of the denatured protein state (Figure 24B).

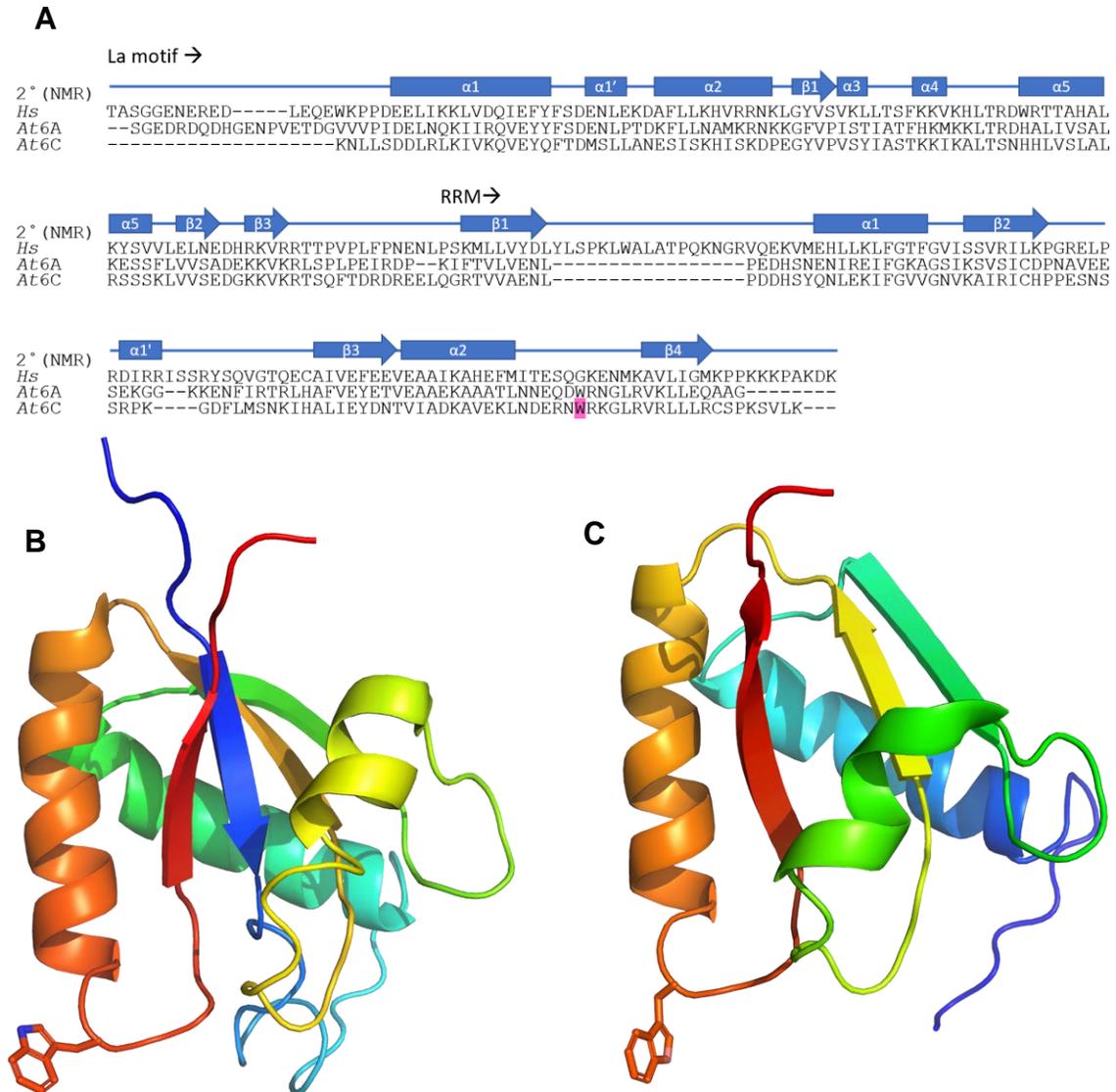


**Figure 23: AtLaRP6C La Module vs 2 – 7.25 M Guanidinium Hydrochloride at 25°C.** The La Module was incubated with varying concentrations of guanidinium hydrochloride and excited at 280 nm, fluorescence was measured between 300 and 460 nm. **(A)** Raw fluorescence spectra. **(B)** Intensity at 340 nm plotted against guanidinium hydrochloride concentrations.



**Figure 24: AtLaRP6C La Module vs 1 – 7.8 M Guanidinium Hydrochloride with 30°C.** The La Module was incubated with varying concentrations of guanidinium hydrochloride and excited at 280 nm, fluorescence was measured between 300 and 460 nm. **(A)** Raw fluorescence spectra. **(B)** Intensity at 340 nm plotted against guanidinium hydrochloride concentrations.

This method relies on the intrinsic fluorescence of tryptophans in the protein, but also on the number and location of these residues. Looking at the sequence of the La Mod, a single tryptophan was identified in the N-terminal of the sequence. Using a sequence alignment of the *AtLaRP6C* La Mod against the *HsLaRP6* La Mod, and the 2° structure based on the solution NMR structure of the *HsLaRP6* La Motif and RRM, the most likely location of the tryptophan was determined (Figure 25A). To further investigate the location of this tryptophan, the equivalent position in the human RRM was mutated to a tryptophan and modeled (Figure 25B). As the sequence of the *AtLaRP6C* La Mod is different than the human, the *AtLaRP6C* sequence was threaded onto the structure of the human RRM (Figure 25C). From these data, it was determined the tryptophan most likely lies in the loop between the final  $\alpha$ -helix and the final  $\beta$ -sheet in the RRM. As the tryptophan is likely surface-exposed and not folded in a secondary structure, the chemical environment of the residue was minimally changed as GndHCl concentrations increased. This is likely why the GndHCl denaturation assay did not provide quantifiable results.

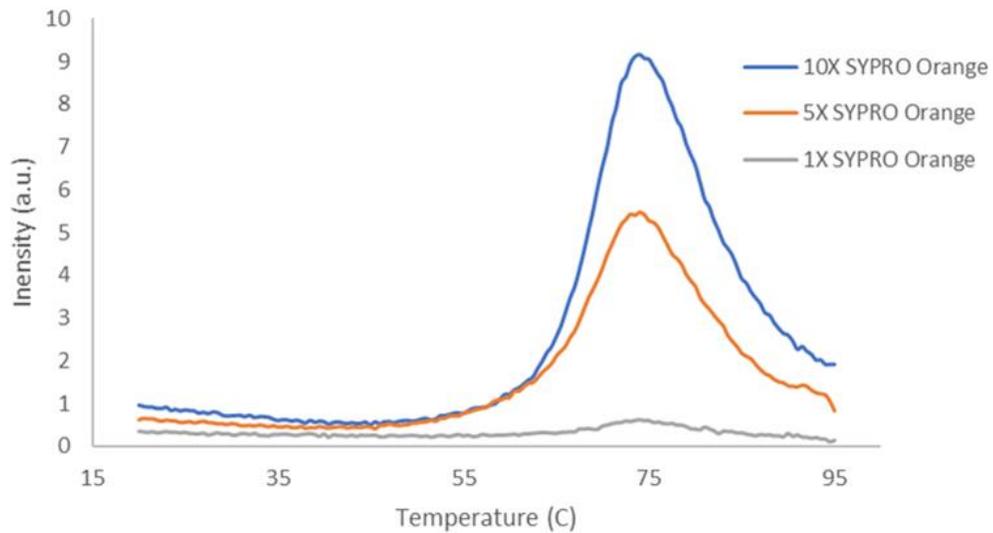


**Figure 25: Location of tryptophan in *AtLaRP6C* La Module. (A)** A sequence alignment of the *AtLaRP6A* and *AtLaRP6C* La Modules against the *HsLaRP6* La Module, and the 2° structure based on the solution NMR structure of the human La Module and RRM. The tryptophan is highlighted in pink in the *AtLaRP6C* sequence. **(B)** Modeled equivalent position of the *AtLaRP6C* La Module tryptophan on the human LaRP6 RRM (PDB: 2MTG). **(C)** *AtLaRP6C* RRM sequence threaded onto the *HsLaRP6* RRM structure (PDB: 2MTG, PHYRE).

## SYPRO™ Orange Thermal Melt Assays

As chemical denaturation while monitoring intrinsic tryptophan fluorescence was not a tractable approach, a thermal shift assay was developed

instead. The Cary fluorometer software package includes a thermal melt program, which allows variations in ramping of heat at specific rates with precise start and stop points, while continuously monitoring fluorescence emission. Prior to using any of the A $\beta$ LaRP6C proteins, the protocol was tested against lysozyme using 1 $\times$ , 5 $\times$ , and 10 $\times$  SYPRO™ Orange (Figure 26A). The fluorescence values produced from the 1 $\times$  SYPRO™ Orange were too noisy to use, however the fluorescence values from both the 5 $\times$  and 10 $\times$  SYPRO™ Orange were found to be comparable to each other. From the 5 $\times$  and 10 $\times$  SYPRO™ Orange fluorescence spectra the  $T_m$  of lysozyme was calculated to be 67.5°C and 68°C respectively. The  $T_m$  of lysozyme is approximately 70°C at pH 7, as determined by Fourier Transform Infrared (FTIR) Spectroscopy, and the calculated values from this assay agree with previous studies.<sup>17</sup> While the 5 $\times$  and 10 $\times$   $T_m$  values were very similar (Table 3), the 5 $\times$  was chosen for the standard protocol as it produced robust absorbance values and reduced reagent usage.



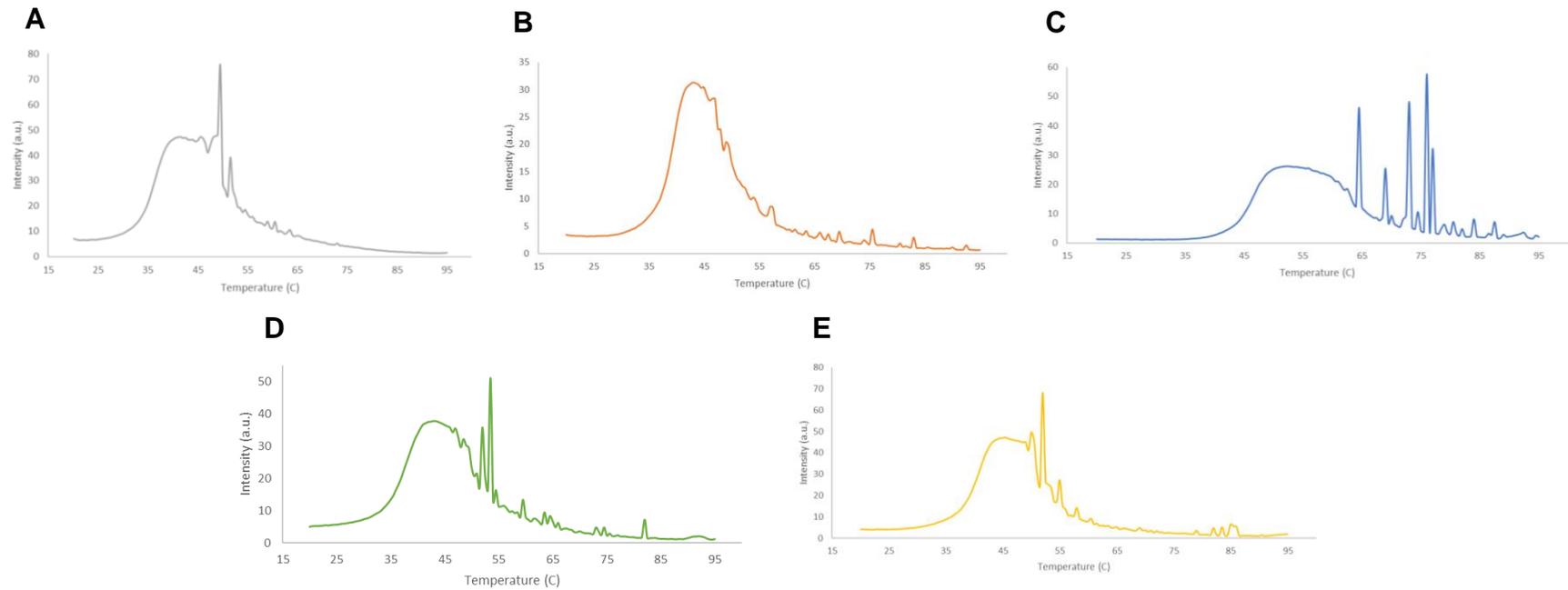
**Figure 26: Thermofluor assay of lysozyme.** Intensity of fluorescent signal of lysozyme mixed with varying concentrations of SYPRO™ Orange.

**Table 3: Calculated stability parameters for lysozyme.**

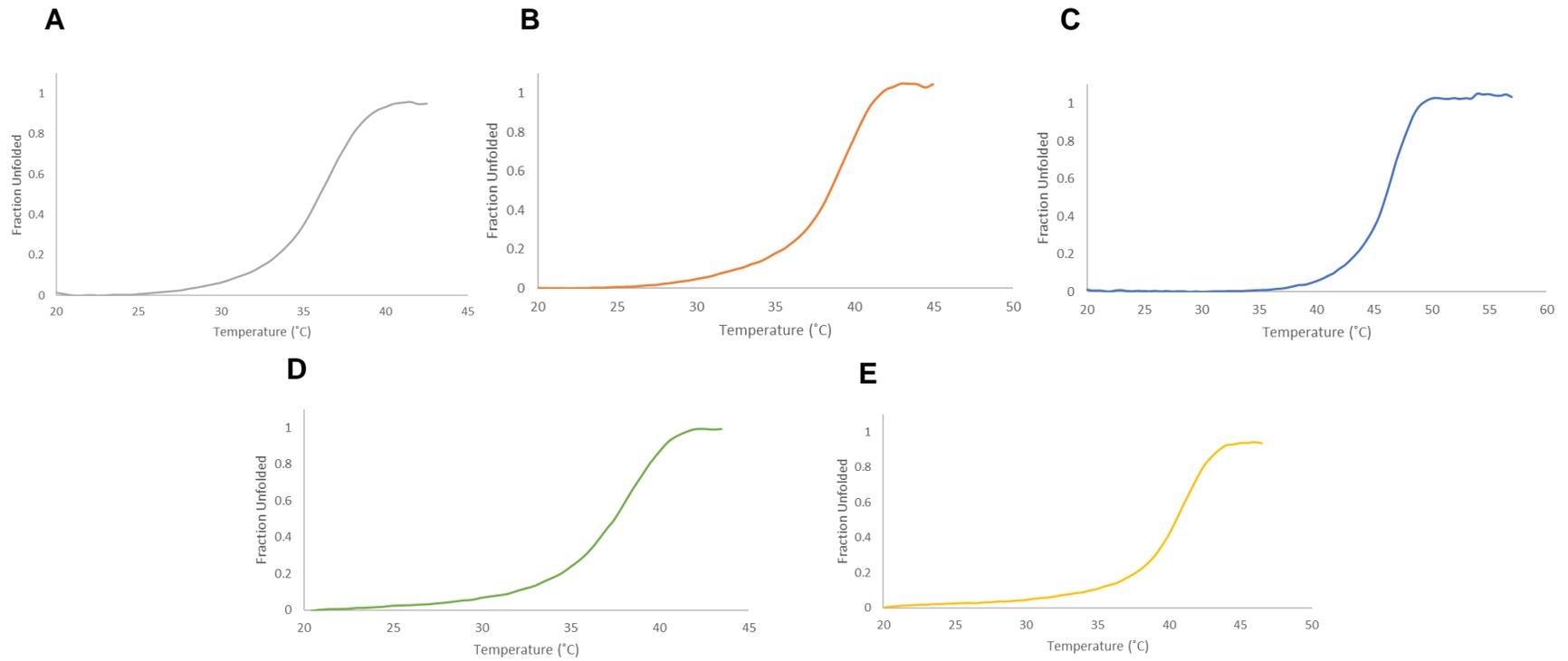
[SYPRO™ Orange]	T <sub>m</sub> (°C)
10x	68
5x	67.5

Using the protocol developed with lysozyme, the melting temperatures were determined for the *At*LaRP6C constructs FL, ΔPAM2, ΔCTD, ΔPAM2ΔCTD, and La Module. Proteins were mixed with 5x SYPRO™ Orange and incubated stepwise from 20°C to 95°C while monitoring SYPRO™ Orange fluorescence (Figure 28). As the proteins denature, hydrophobic surfaces are exposed and bound by SYPRO™ Orange and fluoresces. As denaturation continues, the denatured protein forms aggregates and precipitates out of solution, while SYPRO™ Orange is excluded leading to a reduced signal. From

this curve, multiple parameters that describe protein stability, such as  $T_m$ , can be determined.<sup>14</sup> Each protein was evaluated in triplicate with the exception of the FL, which was only analyzed in duplicate due to limited protein stocks. The NTR and  $\Delta$ PAM2 NTR were excluded from these analyses, as too little protein was obtained from these purifications for use in this analysis.



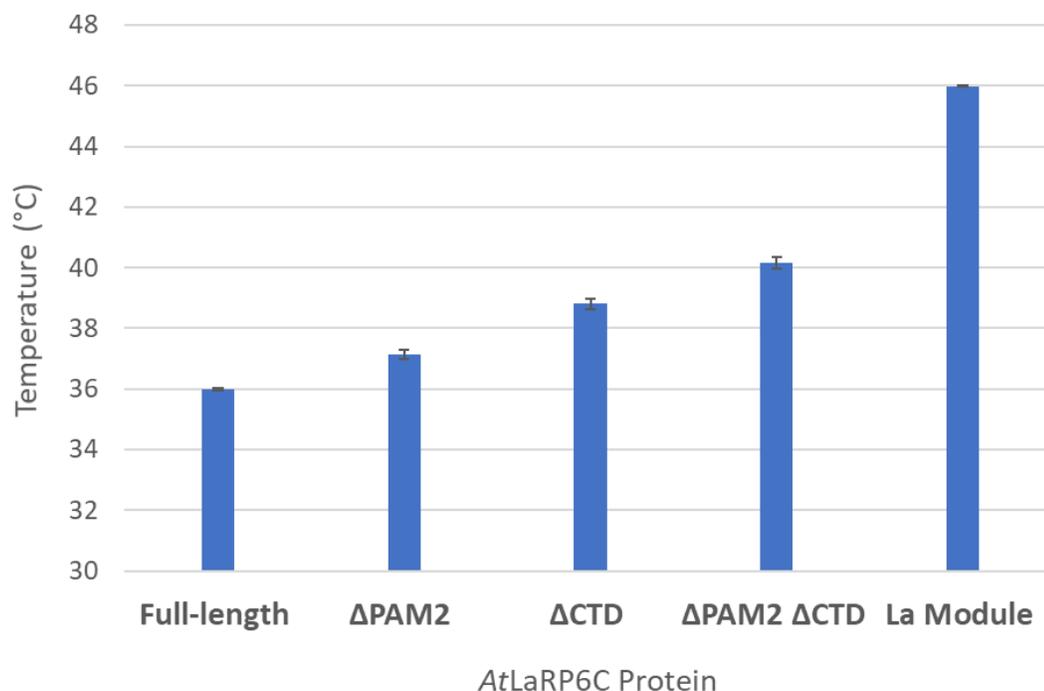
**Figure 27: Thermofluor assay of AflaRP6C proteins.** Proteins were combined with 5 $\times$  SYPRO<sup>TM</sup> Orange and heated from 20°C to 95°C. Excitation/Emission is 470 nm/ 570 nm. These are representative graphs of the obtained fluorescence spectra. **(A)** Full-Length **(B)**  $\Delta$ CTD **(C)** La Module **(D)**  $\Delta$ PAM2 **(E)**  $\Delta$ PAM2  $\Delta$ CTD



**Figure 28: Normalized data from thermofluor assay of AflLaRP6C proteins.** Data was normalized as described in methods, these are representative graphs. **(A)** Full-Length **(B)**  $\Delta$ CTD **(C)** La Module **(D)**  $\Delta$ PAM2 **(E)**  $\Delta$ PAM2  $\Delta$ CTD

These data demonstrate that the size of the *AtLaRP6C* protein construct inversely correlates with melting temperature ( $T_m$ ). The smallest construct evaluated, the La Module, has the highest  $T_m$  at 46°C, while the longest construct (the full-length protein) has the lowest  $T_m$  at 36°C. Deletion of the CTD increases the  $T_m$  by 3°C (compare  $\Delta$ CTD to FL); similarly, deleting the NTR from the  $\Delta$ CTD construct increases the  $T_m$  by 8 °C (compare La Module to  $\Delta$ CTD) (Figure 29, Table 4).

These data also provide insight into the role of the PAM2 motif on the stability of the proteins. Deletion of the PAM2 motif from the full-length protein only increases the  $T_m$  by 1°C, which is small but still potentially significant, as 1°C is greater than the SEM variance for each measurement (Table 4). As observed for the wildtype sequence, the deletion of the CTD on the  $\Delta$ PAM2 background increases the  $T_m$  by ~3°C. Together, these data suggest that the PAM2 motif is not a major contributor to the overall stability of the *AtLaRP6C* protein; this is consistent with the individual folded domains of the protein functioning as “beads on a string”, rather than forming stabilizing intramolecular interactions.



**Figure 29: Comparison of AtLaRP6C protein T<sub>m</sub> obtained from thermal shift assay.** Each protein was combined with 5× SYPRO™ Orange and heated at 0.5°C/min from 20°C to 95°C. From the thermal data, the T<sub>m</sub> was calculated and plotted as a bar graph; n = 3 for all constructs except for the FL, for which n = 2. Error bars are standard error of mean (for n = 3) or standard deviation (for n = 2).

**Table 4: Calculated stability parameters for AtLaRPC proteins.**

Protein	T <sub>m</sub> (°C)
Full-Length*	36.0 ± 0.04
ΔPAM2	37.1 ± 0.15
ΔCTD	38.8 ± 0.17
ΔPAM2ΔCTD	40.2 ± 0.18
La Module	46.0 ± 0.01

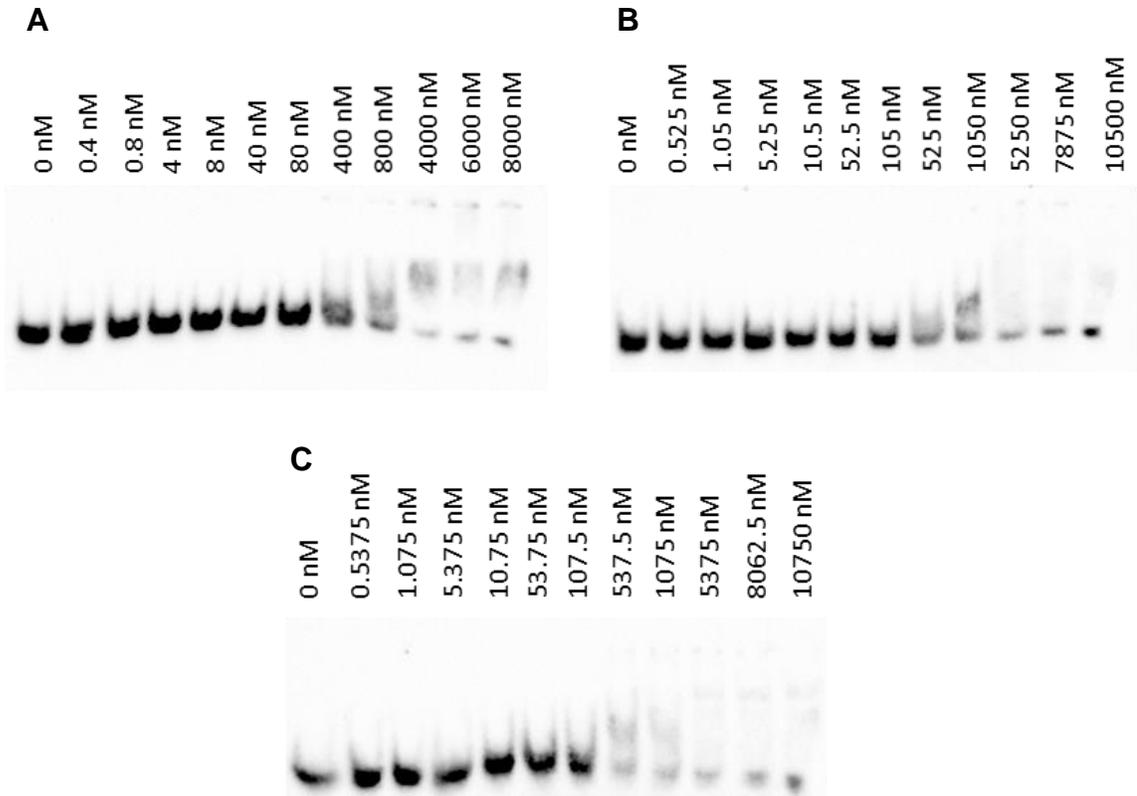
\*FL error value is standard deviation; all others are standard error of the mean

## **RNA Binding Activity Assays**

To evaluate the RNA binding activity of the AtLaRP6C proteins, electrophoretic mobility shift assays were conducted. Two different methods of ligand detection were used for these assays. The first utilized a biotinylated RNA ligand, as had been previously established in the lab. During the course of this work, however, an alternative labeling and detection method was developed that used a fluorescently-labelled RNA ligand. The binding reactions themselves were performed the same.

## **Biotinylated Ligands**

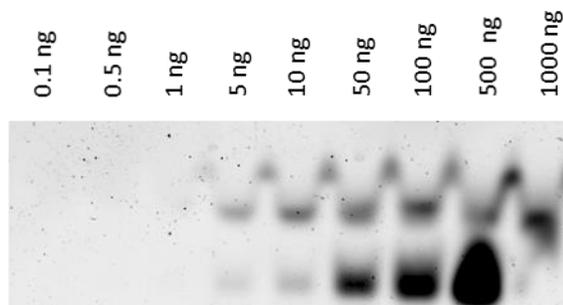
As described in Methods, serially-diluted protein was combined with RNA and allowed to equilibrate for 1 hour. The binding reactions were separated on native gels and transferred to a Hybond(+) membrane, which was then analyzed to detect biotinylated RNA. Using a U-rich RNA ligand ("CB1"), the La Module and  $\Delta$ PAM2  $\Delta$ CTD were tested for RNA binding activity. As protein concentration increased, a reduction in the signal of free RNA ligand was observed; however, a clearly shifted band is not visible (Figure 30 A, B). Similarly,  $\Delta$ PAM2  $\Delta$ CTD was incubated with biotinylated poly-U RNA. Similarly, free RNA ligand is visibly reduced under higher protein concentrations, but there were no quantifiable shifted bands (Figure 30 C).



**Figure 30: Biotinylated RNA in gel shift assays.** Reactions were kept on ice and equilibrated for 1 hour before separated on a 6.5% native gel. After separation, the gel was transferred to a Hybond(+) membrane, detected, and imaged. **(A)** La Module vs 1 nM CB1 exposed for 31.1 sec **(B)**  $\Delta$ PAM2  $\Delta$ CTD vs 2 nM CB1 exposed for 61.0 sec **(C)**  $\Delta$ PAM2  $\Delta$ CTD vs 20 nM poly-U exposed for 64.7 sec.

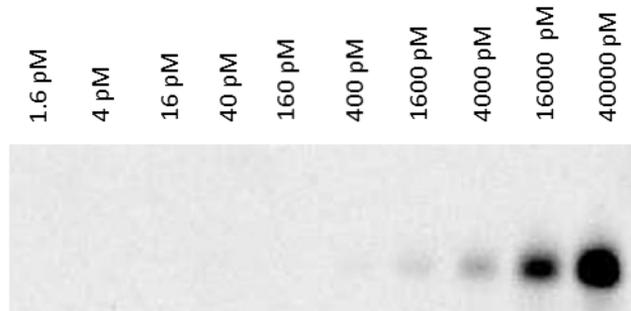
### Alternative RNA Detection Method Development

To determine whether post-electrophoresis staining of RNA could be used to primarily detect RNA, poly(A) RNA was serially diluted with 0.5 $\times$  TE buffer with 15% glycerol. The RNA was then separated on a 6.5% native gel and stained with 2 $\times$  SYBR Gold for 20 minutes. At a minimum, 5 – 10 ng of RNA was visible by this method, which was not low enough for the binding assays (Figure 31).



**Figure 31: Detection limit of SYBR Gold using poly(A) RNA.** Serially diluted poly-A RNA was run on a 6.5% native gel and stained with 2× SYBR Gold for 20 minutes. The gel was imaged using the BioRad ChemiDoc XRS+ with the SYBR Gold setting.

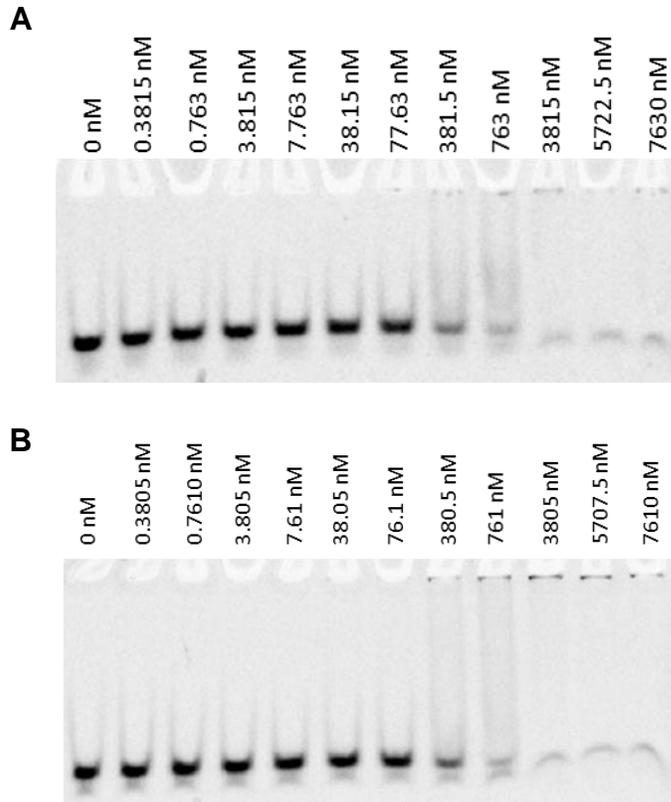
Therefore, RNA ligands of interest were ordered as 5'-FAM labelled oligonucleotides (Integrated DNA Technologies, Inc.) and similarly evaluated to identify the lowest possible RNA concentration for EMSAs. The 5'-FAM poly(A) ligand was serially diluted and electrophoresed under the same conditions as a typical EMSA on a 6.5% native gel. The RNA was then detected using the BioRad Pharos molecular imager on the FITC setting (excitation/emission is 488 nm/530 nm). The lower limit of detection was 1.6 nM (Figure 32), and so this approach was determined to be sufficient for EMSAs in which the upper limit of RNA ligand concentration was ~ 20 nM.



**Figure 32: Detection limit of 5'-FAM poly (A) RNA.** Serially diluted 5'-FAM poly(A) RNA was run on a 6.5% native gel and detected using the BioRad Pharos molecular imager on the FITC setting.

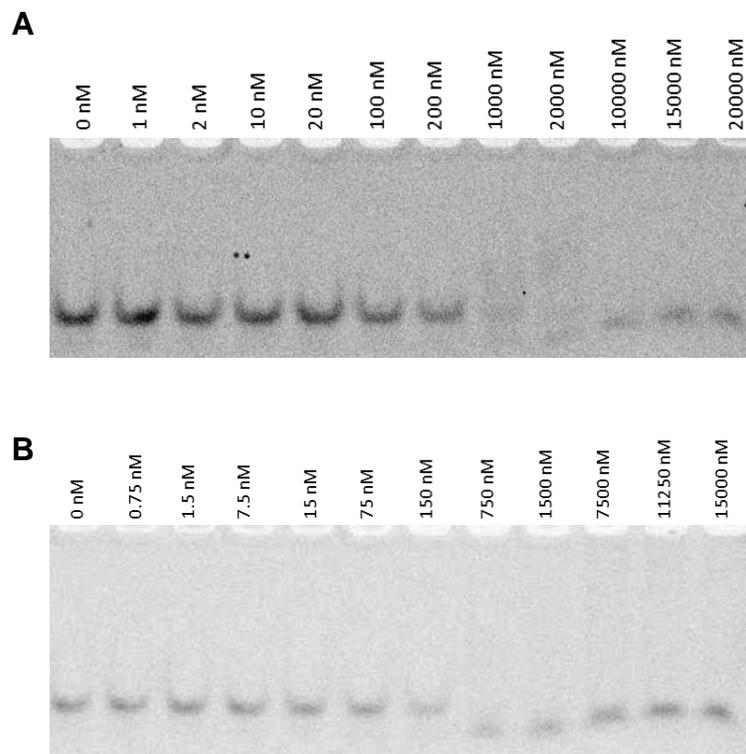
### EMSA with fluorescently labelled RNA

The RNA binding activity of the La Module was first tested against the fluorescently-labelled U-rich RNA, CB1, at a final concentration of 2 nM. As was observed with the biotinylated ligand, binding is occurring as can be seen by the loss of free ligand bands in the wells with the highest protein concentration (Figure 33A). Although shifted bands can be seen, they have not migrated into the gel and seem to be stuck in the wells. To test if the bound complex would migrate in a lower-percentage gel, this experiment was repeated using a 5.25% native gel. While the free RNA ligand traveled further, the bound complex still did not enter the gel (Figure 33B).



**Figure 33: AtLaRP6C La Module vs 2 nM 5'-FAM CB1. (A)** Binding reactions of the La Module against CB1 on a 6.5% native gel. **(B)** Binding reactions of the La Module against CB1 on a 5.25% native gel.

The 1× binding buffer used in the binding reactions contained BSA to prevent non-specific binding. To test the hypothesis that this additional protein was preventing the bound complex from entering the gel, fresh 1× binding buffer was prepared without BSA. The La Module and ΔCTD proteins were evaluated for binding to 0.5 nM 5'-FAM CB1 using the 1× BSA-free binding buffer (Figure 34). Again, binding was detected, but no quantifiable shifted bands were observed.



**Figure 34: AtLaRP6C La Module and  $\Delta$ CTD vs 0.5 nM CB1, no BSA.** Binding reactions were performed as previously mentioned using  $1\times$  binding buffer with no BSA. **(A)** La Module vs 0.5 nM 5'-FAM CB1 separated on a 6.5% native gel. **(B)**  $\Delta$ CTD vs 0.5 nM 5'-FAM CB1 separated on a 6.5% native gel.

## Discussion and Conclusions

The data obtained from plotting the absorbance at 340 nm by the concentration of GndHCl resulted in a positive line with direct correlation to the concentration of GndHCl. This was unexpected as most denaturation assays result in a sigmoidal curve with an observed natured and denatured state of the protein. A singular tryptophan is present in the La Module sequence, located at the C-terminus of the RRM; using the human LaRP6 RRM structure and multiple sequence alignment, this tryptophan is predicted to be in the C-terminal extension of the RRM. Therefore, there is a strong possibility that this tryptophan

is already considerably exposed to solvent, minimizing its potential to report on local or global changes in structure. Therefore, a new protocol was developed using SYPRO™ Orange to report on global changes in structure.

SYPRO™ Orange is a fluorescent dye which has an excitation/emission profile of 470/570. SYPRO™ Orange fluorescence increases upon binding to hydrophobic surfaces. As proteins denature, hydrophobic surfaces can become exposed allowing for SYPRO™ Orange to bind and fluoresce. These two features have been combined by others into a “thermal melt assay” (a.k.a. “thermofluor assay”), in which protein/SYPRO™ Orange mixtures are monitored over a temperature range in a qPCR machine. We adapted this approach for the departmental Cary Eclipse fluorescence spectrophotometer, which is equipped with a four-position Peltier sample holder. This allows for the simultaneous evaluation of four samples over a wide temperature range. Building on the chemical denaturation experiments described above, this approach provided a robust approach to denature the *AtLaRP6C* protein constructs.

As expected, the La Module has the highest  $T_m$ , while the Full-Length proteins (wildtype and  $\Delta$ PAM2) had the lowest  $T_m$ . Interestingly, deletion of the PAM2 lead to a slight increase in  $T_m$  in both the FL and  $\Delta$ CTD constructs. Overall, the La Mod is extremely stable in comparison to the other constructs, suggesting that the entropy contained in the intrinsically-disordered NTR and CTD may contribute to less structural stability of the full-length proteins.

Unfortunately, none of the attempts at measuring RNA binding activity were quantifiable. However, we were able to establish that these constructs do

exert RNA binding activity, indicating that the proteins were folded and functional. This work did contribute to the development of a novel RNA detection method in the research group. The data presented here demonstrate the feasibility of using fluorescently-labeled RNA ligand for direct detection of the RNA, rather than the secondary detection via biotinylation and chemiluminescence of HRP-streptavidin.

In an effort to remove the transfer step in our EMSA protocol, and to find a more robust and less resource-consuming RNA binding assay method, SYBR Gold was considered. As a highly sensitive nucleic acid reporter, we believed it could be used to stain the native gels to determine RNA presence and amount. Before use in EMSAs, the detection limit of SYBR Gold on our RNA ligands was determined. The detection limit of SYBR Gold proved to be too high for our needs at the concentration of RNA we require for our binding assays. Another issue with this method, was the incubation of the native gels in liquid for a prolonged time. This has been shown to lead to band diffusion, and there was some concern this would lead to quantification errors. However, this was a useful experiment as it led us to consider fluorescently labelled RNA ligands.

To fit our lab and instrument needs, a 5'-FAM label was found to be the most effective fluorescent reporter and all ligands were ordered. The detection limit of the FAM labelled RNA was found to be sufficient for our needs and provided a more robust method of detecting RNA. The fluorescently tagged RNA allowed us to eliminate the transfer step and visualize free-ligand and bound complexes by imaging the gel itself in a fluorescence imager. The subsequent

fluorescent EMSAs (fEMSAs) showed a new problem with the RNA binding of the *At*LaRP6C proteins. The isolated La Module was bound against the 5'-FAM-CB1 ligand and although binding did occur, the bands appeared stuck in the wells of the native gel. To alleviate this, the percentage of the gel was reduced from 6.5% to 5.25%. The free RNA ligand moved much further through the gel, however the bound complex still seemed to be stuck in the wells and shifted bands were not quantifiable. With the isolated La Module as the core RNA-binding region of the protein and as one of the smallest constructs, we were concerned with the comparability of the RNA binding activities of the various proteins.

We then hypothesized the age of the La Module could be the problem, as it had been purified over a year earlier, and was thought to be aggregating in the wells. Another purification of the protein was prepared, but additional fEMSAs showed no difference in movement through the gel. We then considered the 1× binding buffer recipe which was obtained from our collaborators. In this buffer, bovine serum albumin (BSA) was added to prevent non-specific binding. We believed this additional protein could have been preventing the bound complex from entering the gel in our system. To test this, fresh 1× Binding Buffer was made without the addition of BSA. This did not change the result for either the La Module or the  $\Delta$ CTD construct. At this point, it was decided to step away from the RNA binding portion of the project and to focus on other biophysical characteristics of the proteins.

## V. CONCLUSIONS AND FUTURE DIRECTIONS

The proteins constructs that are needed to probe the effect of the NTR on RNA binding activity are highly expressible and purifiable. However, effective testing of the RNA binding activity of these proteins has not been determined. Of the proteins used in this project, the La Mod has the greatest stability with the FL having the lowest. Deletion of the PAM2 does slightly increase the stability and  $T_m$  of the proteins, but not significantly. Deletion of the CTD increases the stability of the protein.

To better understand the contribution of the NTR towards binding activity and stability, a mutant deleting the NTR should be synthesized. With the observed association of the NTR and  $\Delta$ CTD constructs with the nickel resin during purification, these proteins should be tested for metal binding activity. This activity may help to elucidate the function of the NTR *in vivo*, and lead to understanding additional functionality of the protein.

In order to successfully complete the goal of testing the RNA binding affinity of these purified AtLaRP6C protein constructs against the four ligands that were identified by the Bousquet-Antonelli CLIP-Seq assay (CB1, CB2, CB3, and CB4), the current EMSA protocol must be optimized. Alternatively, another method of determining binding affinity should be used. We expect that because CB1 and CB2 are U-rich, they will bind readily to all protein constructs that contain the La Module, regardless of the presence of the PAM2 motif. In contrast, CB3 and CB4 are A-rich, and therefore we expect that they will only be robustly bound by protein constructs that contain the N-terminal region. These RNA

ligands have been commercially synthesized with 5'-FAM labels for detection in electrophoretic mobility shift assays <sup>16, 18</sup>. The conditions necessary for the binding of the purified recombinant proteins and synthetic ligands will be established based on the previous work performed by the Bousquet-Antonelli Lab.

This would directly test the hypothesis that the N-terminal region of *AtLaRP6C* confers binding activity against A-rich RNA sequences, and whether the PAM2 site is necessary for this functionality. The analysis of this protein and its domains will be used to further characterize functions of LaRP6 proteins within other species.

## APPENDIX SECTION

### Protein Sequences

1. **His<sub>10</sub>-SUMO (cleaved – terminal S remains on LaRP6C protein):**  
HHHHHHHHHSSGHIEGRHMASMSDSEVNQEAKPEVKPEVKPETHINL  
KVSDGSSEIFFKIKKTTPLRRLMEAFKRQKGEMDSLRFlyDGIRIQADQ  
TPEDLDMEDNDIIEAHREQIGG
2. **AtLaRP6C:**  
SMAQMQRREEVESVTTEKKRLDGGGGSSGAQATAFKFNAQAPEFVPRS  
HTTAPAPQVSPVSGYFYPCFHYNGGCIGGCGGGVCGGGGTGVGTQSS  
DWIYVGGGDPTAQHQHVHDPAAAFYISNPAVQFPASQNSSSSSKNLLS  
DDLRLKIVKQVEYQFTDMSLLANESISKHISKDPEGYVPVSYIASTKKIKAL  
TSNHHLVSLALRSSKLVVSEDGKKVKRTSQFTDRDREELQGRTVVAEN  
LPDDHSYQNLEKIFGVVGNVKAIRICHPPESSSRPKGDFLMSNKIHALIE  
YDNTVIADKAVEKLNDRNWRKGLRVRLLLRCSPKSVLKNRRNFDGILID  
DELPSYESGEDSPRLHLTESQLDNDGDDNNVGGLWGKGRGKGRGRSP  
RSYAVGGGGRSFGIGLGVSLGIPSLGSHESSPKTATKGPRMPDGTRG  
FTMGRGKPSISLSPNNL
3. **AtLaRP6C ΔPAM2:**  
SMAQMQRREEVESVTTEKKRLDGGGGSSGAQATAHTTAPAPQVSPVSG  
YFYPCFHYNGGCIGGCGGGVCGGGGTGVGTQSSDWIYVGGGDPTAQ  
HQHVHDPAAAFYISNPAVQFPASQNSSSSSKNLLSDDLRLKIVKQVEYQ  
FTDMSLLANESISKHISKDPEGYVPVSYIASTKKIKALTSNHHLVSLALRSS  
SKLVVSEDGKKVKRTSQFTDRDREELQGRTVVAENLPDDHSYQNLEKIF  
GVVGNVKAIRICHPPESSSRPKGDFLMSNKIHALIEYDNTVIADKAVEKLN  
DRNWRKGLRVRLLLRCSPKSVLKNRRNFDGILIDDELPSYESGEDSP  
RLHLTESQLDNDGDDNNVGGLWGKGRGKGRGRSPRSYAVGGGGRSFG  
IGLGVSLGIPSLGSHESSPKTATKGPRMPDGTRGFTMGRGKPSISLS  
PNNL
4. **AtLaRP6C La Module:**  
SKNLLSDDLRLKIVKQVEYQFTDMSLLANESISKHISKDPEGYVPVSYIAS  
TKKIKALTSNHHLVSLALRSSKLVVSEDGKKVKRTSQFTDRDREELQG  
RTVVAENLPDDHSYQNLEKIFGVVGNVKAIRICHPPESSSRPKGDFLM  
SNKIHALIEYDNTVIADKAVEKLNDRNWRKGLRVRLLLRCSPKSVLK

5. **AtLaRP6C ΔCTD:**  
SMAQMQRREEVESVTTEKKRLDGGGGSSGAQATAFKFNAQAPEFVPRS  
HTTAPAPQVSPVSGYFYPCFHYNNGGCIGGCGGGVCGGGGTGVGTQSS  
DWIYVGGGDPTAQHQHVHDPAAAFYISNPAVQFPASQNSSSSSKNLLS  
DDLRLKIVKQVEYQFTDMSLLANESISKHISKDPEGYPVPSYIASTKKIKAL  
TSNHHLVSLALRSSKLVVSEDGKKVKRTSQFTDRDREELQGRTVVAEN  
LPDDHSYQNLEKIFGVVGNVKAIRICHPPESSSRPKGDFLMSNKIHALIE  
YDNTVIADKAVEKLNDERNWRKGLRVRLLLRCSPKSVLK
  
6. **AtLaRP6C ΔPAM2 ΔCTD:**  
SMAQMQRREEVESVTTEKKRLDGGGGSSGAQATAHTTAPAPQVSPVSG  
YFYPCFHYNNGGCIGGCGGGVCGGGGTGVGTQSSDWIYVGGGDPTAQ  
HQHVHDPAAAFYISNPAVQFPASQNSSSSSKNLLSDDLRLKIVKQVEYQ  
FTDMSLLANESISKHISKDPEGYPVPSYIASTKKIKALTSNHHLVSLALRSS  
SKLVVSEDGKKVKRTSQFTDRDREELQGRTVVAENLPDDHSYQNLEKIF  
GVVGNVKAIRICHPPESSSRPKGDFLMSNKIHALIEYDNTVIADKAVEKL  
NDERNWRKGLRVRLLLRCSPKSVLK
  
7. **AtLaRP6C NTR:**  
SMAQMQRREEVESVTTEKKRLDGGGGSSGAQATAFKFNAQAPEFVPRS  
HTTAPAPQVSPVSGYFYPCFHYNNGGCIGGCGGGVCGGGGTGVGTQSS  
DWIYVGGGDPTAQHQHVHDPAAAFYISNPAVQFPASQNSSSSS
  
8. **AtLaRP6C ΔPAM2 NTR:**  
SMAQMQRREEVESVTTEKKRLDGGGGSSGAQATAHTTAPAPQVSPVSG  
YFYPCFHYNNGGCIGGCGGGVCGGGGTGVGTQSSDWIYVGGGDPTAQ  
HQHVHDPAAAFYISNPAVQFPASQNSSSSS
  
9. **AtLaRP6C ΔLSA:**  
SMAQMQRREEVESVTTEKKRLDGGGGSSGAQATAFKFNAQAPEFVPRS  
HTTAPAPQVSPVSGYFYPCFHYNNGGCIGGCGGGVCGGGGTGVGTQSS  
DWIYVGGGDPTAQHQHVHDPAAAFYISNPAVQFPASQNSSSSSKNLLS  
DDLRLKIVKQVEYQFTDMSLLANESISKHISKDPEGYPVPSYIASTKKIKAL  
TSNHHLVSLALRSSKLVVSEDGKKVKRTSQFTDRDREELQGRTVVAEN  
LPDDHSYQNLEKIFGVVGNVKAIRICHPPESSSRPKGDFLMSNKIHALIE  
YDNTVIADKAVEKLNDERNWRKGLRVRLLLRCSPKSVLKNRRNFDGILID  
DELPSYESGEDSPRLHLTESQLDNDGDDNNGGLWGKGRGKGRGRSP  
RSYAVGGGGRSFGIGLGVSLGIPSLGSHE

**Table S1: Site-directed mutagenesis primers and conditions to form AflLaRP6C variants.**

<b>Construct/mutation/template</b>	<b>Mutagenic Primer</b>	<b>Cycling Parameters</b>
pET28(a)-SUMO-AflLaRP6C remove 2 nucleotide insertion causing frameshift	F: CAGATTGGTGGATCCATGGCGCAGATGCAG  R: CTGCATCTGCGCCATGGATCCACCAATCTG	95°C – 30s 25 cycles: 95C – 30s 50C – 45s 72C – 4 min 72C – 6 min 4C – Hold
pET28(a)-SUMO-AflLaRP6C <b>ΔLSA</b> Template: pET28(a)-SUMO-AflLaRP6C	F: GGGTCACACGAATCTTAGTGACCTAAAACAGCAACAAAGGG  R: CCCTTTGTTGCTGTTTTAGGTCACTAAGATTCGTGTGACCC	94°C – 30s 25 cycles: 94C – 30s 56C – 45s 65C – 11 min 65C – 15 min 4C – Hold
pET28(a)-SUMO-AflLaRP6C <b>ΔCTD</b> (Template: pET28(a)-SUMO-AflLaRP6C)  pET28(a)-SUMO-AflLaRP6C <b>ΔPAM2 ΔCTD</b> (Template: pET28(a)-SUMO-AflLaRP6C ΔPAM2)	F: CCAAATCGGTGCTCAAGTAATGAAGAACTTC  R: CGAAGTTTCTTCATTACTTGAGCACCGATTTTGG	98°C – 30s 30 cycles: 98C – 10s 72C – 3 min 72C – 5 min 4C – Hold
pET28(a)-SUMO-AflLaRP6C <b>NTR</b> Template: pET28(a)-SUMO-AflLaRP6C  pET28(a)-SUMO-AflLaRP6C <b>ΔPAM2 NTR</b> mutate double stop at 136 Template: pET28(a)-SUMO-AflLaRP6C	F: GTCTTCTTCGTGCTGATAACTGCTTTCCGATG  R: CATCGGAAAGCAGTTATCACGACGAAGAAGAC	94°C – 1 min 20 cycles: 94C – 45s 56C – 45s 65C – 11 min 65C – 15 min 4C – Hold

**Table S3: Purification buffers for AtLaRP6C constructs.**

<b>Construct</b>	<b>SEC Buffer</b>	<b>Lysis/Wash 1</b>	<b>Wash 2</b>	<b>Elution</b>	<b>Elution +</b>
<b>FL La Module ΔCTD ΔPAM2 ΔCTD NTR</b> (SUMO-tagged and cleaved)	50 mM NaH <sub>2</sub> PO <sub>4</sub> /Na <sub>2</sub> HPO <sub>4</sub> (pH 8.0) 200 mM NaCl 50 mM Na <sub>2</sub> SO <sub>4</sub> 1 mM DTT 2% Glycerol	50 mM NaH <sub>2</sub> PO <sub>4</sub> /Na <sub>2</sub> HPO <sub>4</sub> (pH 8.0) 200 mM NaCl 50 mM Na <sub>2</sub> SO <sub>4</sub> 1 mM DTT 10 mM Imidazole (pH 8.0)	50 mM NaH <sub>2</sub> PO <sub>4</sub> /Na <sub>2</sub> HPO <sub>4</sub> (pH 8.0) 200 mM NaCl 50 mM Na <sub>2</sub> SO <sub>4</sub> 1 mM DTT 50 mM Imidazole (pH 8.0)	50 mM NaH <sub>2</sub> PO <sub>4</sub> /Na <sub>2</sub> HPO <sub>4</sub> (pH 8.0) 200 mM NaCl 50 mM Na <sub>2</sub> SO <sub>4</sub> 1 mM DTT 350 mM Imidazole (pH 8.0)	N/A
<b>ΔPAM2 NTR</b> (SUMO-tagged and cleaved)	50 mM NaH <sub>2</sub> PO <sub>4</sub> /Na <sub>2</sub> HPO <sub>4</sub> (pH 8.0) 200 mM NaCl 50 mM Na <sub>2</sub> SO <sub>4</sub> 1 mM DTT 2% Glycerol	50 mM NaH <sub>2</sub> PO <sub>4</sub> /Na <sub>2</sub> HPO <sub>4</sub> (pH 8.0) 200 mM NaCl 50 mM Na <sub>2</sub> SO <sub>4</sub> 1 mM DTT 10 mM Imidazole (pH 8.0)	50 mM NaH <sub>2</sub> PO <sub>4</sub> /Na <sub>2</sub> HPO <sub>4</sub> (pH 8.0) 200 mM NaCl 50 mM Na <sub>2</sub> SO <sub>4</sub> 1 mM DTT 50 mM Imidazole (pH 8.0)	50 mM NaH <sub>2</sub> PO <sub>4</sub> /Na <sub>2</sub> HPO <sub>4</sub> (pH 8.0) 200 mM NaCl 50 mM Na <sub>2</sub> SO <sub>4</sub> 1 mM DTT 350 mM Imidazole (pH 8.0)	50 mM NaH <sub>2</sub> PO <sub>4</sub> /Na <sub>2</sub> HPO <sub>4</sub> (pH 8.0) 200 mM NaCl 50 mM Na <sub>2</sub> SO <sub>4</sub> 1 mM DTT 500 mM Imidazole (pH 8.0)
<b>ΔPAM2</b> (SUMO- tagged and cleaved)	50 mM NaH <sub>2</sub> PO <sub>4</sub> /Na <sub>2</sub> HPO <sub>4</sub> (pH 8.2) 200 mM NaCl 50 mM Na <sub>2</sub> SO <sub>4</sub> 1 mM DTT 2% Glycerol	50 mM NaH <sub>2</sub> PO <sub>4</sub> /Na <sub>2</sub> HPO <sub>4</sub> (pH 8.2) 200 mM NaCl 50 mM Na <sub>2</sub> SO <sub>4</sub> 1 mM DTT 10 mM Imidazole (pH 8.0)	50 mM NaH <sub>2</sub> PO <sub>4</sub> /Na <sub>2</sub> HPO <sub>4</sub> (pH 8.2) 200 mM NaCl 50 mM Na <sub>2</sub> SO <sub>4</sub> 1 mM DTT 50 mM Imidazole (pH 8.0)	50 mM NaH <sub>2</sub> PO <sub>4</sub> /Na <sub>2</sub> HPO <sub>4</sub> (pH 8.2) 200 mM NaCl 50 mM Na <sub>2</sub> SO <sub>4</sub> 1 mM DTT 350 mM Imidazole (pH 8.0)	N/A

## REFERENCES

1. Crick, F., Central dogma of molecular biology. *Nature* **1970**, 227 (5258), 561-3.
2. Glisovic, T.; Bachorik, J. L.; Yong, J.; Dreyfuss, G., RNA-binding proteins and post-transcriptional gene regulation. *FEBS letters* **2008**, 582 (14), 1977-1986.
3. Lunde, B. M.; Moore, C.; Varani, G., RNA-binding proteins: modular design for efficient function. *Nature reviews. Molecular cell biology* **2007**, 8 (6), 479-90.
4. Chen, Y.; Varani, G., Engineering RNA-binding proteins for biology. *The FEBS journal* **2013**, 280 (16), 3734-54.
5. Maraia, R. J.; Mattijssen, S.; Cruz-Gallardo, I.; Conte, M. R., The La and related RNA-binding proteins (LARPs): structures, functions, and evolving perspectives. **2017**, 8 (6), e1430.
6. Bousquet-Antonelli, C.; Deragon, J. M., A comprehensive analysis of the La-motif protein superfamily. *Rna* **2009**, 15 (5), 750-64.
7. Stavraka, C.; Blagden, S., The La-Related Proteins, a Family with Connections to Cancer. *Biomolecules* **2015**, 5 (4), 2701-2722.
8. 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. *Journal of molecular biology* **2010**, 395 (2), 309-26.
9. Wynn, T. A., Cellular and molecular mechanisms of fibrosis. *The Journal of pathology* **2008**, 214 (2), 199-210.
10. Merret, R.; Martino, L.; Bousquet-Antonelli, C.; Fneich, S.; Descombin, J.; Billey, E.; Conte, M. R.; Deragon, J.-M., The association of a La module with the PABP-interacting motif PAM2 is a recurrent evolutionary process that led to the neofunctionalization of La-related proteins. *RNA (New York, N.Y.)* **2013**, 19 (1), 36-50.
11. Cruz-Gallardo, I.; Martino, L.; Kelly, G.; Atkinson, R. A.; 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 research* **2019**, 47 (8), 4272-4291.
12. Anderson, P.; Kedersha, N., Stress granules. *Current biology : CB* **2009**, 19 (10), R397-8.

13. Betancourt, F. Recombinant expression and purification of LARP6 proteins from *Arabidopsis thaliana*. Texas State University 2017.
14. Wright, T. A.; Stewart, J. M.; Page, R. C.; Konkolewicz, D., Extraction of Thermodynamic Parameters of Protein Unfolding Using Parallelized Differential Scanning Fluorimetry. *The Journal of Physical Chemistry Letters* **2017**, *8* (3), 553-558.
15. Pace, C. N., Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods in enzymology* **1986**, *131*, 266-80.
16. 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 expression and purification* **2017**, *134*, 147-153.
17. Venkataramani, S.; Truntzer, J.; Coleman, D. R., Thermal stability of high concentration lysozyme across varying pH: A Fourier Transform Infrared study. *J Pharm Bioallied Sci* **2013**, *5* (2), 148-153.
18. Altschuler, S. E.; Lewis, K. A.; Wuttke, D. S., Practical strategies for the evaluation of high-affinity protein/nucleic acid interactions. *Journal of nucleic acids investigation* **2013**, *4* (1), 19-28.