RECOMBINANT EXPRESSION AND PURIFICATION OF LARP6 PROTEINS FROM *ARABIDOPSIS THALIANA*

by

Francisco C. Betancourt, 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 2017

Committee Members:

Karen A. Lewis, Chair

L. Kevin Lewis

Steven T. Whitten

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

Fair Use

Duplication Permission

ACKNOWLEDGEMENTS

To start, I would like to thank my advisor Dr. Karen Lewis for the tremendous amount of support throughout my years at Texas State, including the opportunity to work in her lab. I feel as though I have learned more science through her classes and lab interactions than I have throughout the rest of my life. I would consider myself lucky to work for another boss of such high caliber. I would also like to thank the South Texas Doctoral Bridge Program, through support by the National Institutes of Health – National Institute for General Medical Studies, GM102783, for funding my education and research, as well as providing several means to explore the scientific community. Thank you to the staff and faculty of Texas State that have assisted in my education and growth as a professional. Thank you to my lab members Daniel Horn, Jose Castro, Ellie Pena, Natalie Bonner, Jennifer White, and Eliseo Salas for their support and contributions. Lastly, and certainly not least, I'd like to thank my parents for their unfaltering support and everything they've done for me throughout my life.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

I. INTRODUCTION

The order of amino acids in a protein sequence, also called the primary structure of a protein, is the basis for the three-dimensional structure and biological function of proteins. Amino acids possess diverse chemical functionalities, because of their side chains, in addition to their amino and carboxylic acid functional groups. Hydrogen bonding between the amino and carboxylic acid groups stabilize the secondary structure of a protein, while side chain interactions are primarily responsible for stabilizing the three-dimensional, or tertiary structure (Segrest *et al*., 1992). These side chains can be grouped by chemical property, such as hydrophobic, polar, or charged. The tertiary structure of a cytosolic protein is primarily dictated by hydrophobic collapse of the secondary structure elements. The structure is then refined by hydrogen bonding within the peptide backbone and interactions between side chains (Rossmann and Liljas, 1974).

Protein sequence provides the organism with a powerful regulatory tool. In addition to directing proper folding pathways, protein sequence motifs serve as critical signaling elements. Motifs can localize proteins within the cell or identify other molecules to bind (Böhmer *et al*., 2013). Therefore, the specificity of protein sequences is critical; just one amino acid substitution in a protein can give rise to myriad negative consequences for the organism. This specificity also allows the mature protein to potentially serve a unique purpose and/or to be very selective with interacting molecules.

Naturally-occurring homopolymeric amino acid motifs

If every amino acid has a 1/20 chance of being coded for at any residue location, then a homopolymeric tract of five contiguous identical amino acids would occur $(1/20)^5$, or one in about ten million. Similarly, a contiguous tract of seven amino acids would have roughly a one in a trillion chance of occurring. The human transcription factor TAFIIB contains a stretch of seven contiguous histidine residues though, and is just one of many prevalent proteins containing amino acid repeat tracts (Salichs, *et al*., 2009). In fact, homopolymeric tracts are overrepresented in eukaryotic genomes, and are found in about 18-20% of proteins (Salichs *et al.,* 2009). This suggests that these sequences were subjected to a positive evolutionary selective pressure. In other words, they likely have an important function(s) that provides the organism with an evolutionary advantage. For example, poly-Q and poly-N tracts are known to form coiled-coil structures that mediate important protein-protein interactions (Fiumara *et al*., 2010). However, these repeat sequences are also highly susceptible to expansion during DNA replication, in which the template strand slips and re-anneals to the elongating strand in a different register. Expansion of poly-Q/N tracts can extend their coiled-coil structure, which impairs function and promotes protein aggregation (Fiumara *et al*., 2010). In fact, homopolymeric glutamine and asparagine tracts are implicated in several human diseases (Smalley *et al*., 2016). A notable example is Huntington's disease, in which a CAG trinucleotide repeat sequence codes for a poly-Q tract in the Huntingtin protein. When this tract is expanded, the mutated protein can aggregate into homooligomers and form plaques (Fiumara *et al*., 2010).

Poly-His tags are widely used in biotechnology for labeling proteins of interest for either isolation or detection (Bornhorst and Falke, 2000). The imidazole ring found in histidine can form ion-dipole bonds with an immobilized metal-affinity column, allowing the column to selectively retain poly-His containing molecules (Bornhorst and Falke, 2000). However, the recent advances in genome sequencing have shown that sequences of contiguous histidine residues exist in a significant number of endogenous proteins present in prokaryotes, eukaryotes, and even viruses (Faux *et al*., 2005). Homopolymeric "poly-His tracts" are defined as at least five contiguous histidine residues (Salichs *et al*., 2009). "Extended poly-His tracts" contain at least one contiguous His_5 tract, with additional His residues at the start and end of the tract, and with other amino acids constituting less than 50% of the total tract sequence (Salichs *et al*., 2009). Both poly-His tracts and extended poly-His tracts exhibit a high affinity for divalent metal ions (Watly *et al*., 2015). This metal binding activity is well-known to play a role in bacterial response to environmental metal. For example, the expression of the bacterial protein Hpn, a nickel storage protein in *Helicobacter pylori* containing many poly-histidine tracts, is upregulated when the nickel sensing protein NikR is activated by binding of a Ni^{2+} ion (Witkowska *et al*., 2012). Recently, a protein in the venom of the eukaryote *Atheris squamigera (*African bush viper) was found to have an endogenous His_9 tract. This sequence binds several divalent metal ions with high affinity, the strongest being Cu^{2+}, determined through circular dichroism and computational studies (Watly *et al*., 2015). Watly and colleagues conclude that these divalent metal ions (Ni^{2+}, Cu^{2+}, Zn^{2+}) can bind to multiple sites on the poly-His sequence, and form various polymorphic states much like the His_6 tags used for recombinant protein purification (Watly *et al*., 2015).

This endogenous metal-binding poly-His tract may be important for inhibiting metalloproteases that degrade snake-venom proteins (Watly *et al*., 2015).

The role of poly-His tracts in eukaryotic cellular proteins is less well characterized. In the human genome, there are 86 proteins out of around 20,000 that contain poly-His tracts (Salichs *et al*., 2009). In a surprising discovery, Salichs and colleagues recently demonstrated that polyhistidine tracts are necessary and sufficient for the localization of a cytosolic protein to nuclear speckles in HeLa cells. This discovery suggests an alternative to the canonical nuclear localization pathway, in which conserved nuclear localization signal sequences (NLS) within "cargo" proteins are bound by "transporter" proteins, such as Importin α, allowing influx of protein into the nuclear membrane (Lim *et al*., 2011).

Salichs and colleagues conducted multiple experiments using four proteins with naturally-occurring poly-histidine tracts, in which the tract sequences are added or deleted by directed mutation (Fig. 1-1, Panel A).

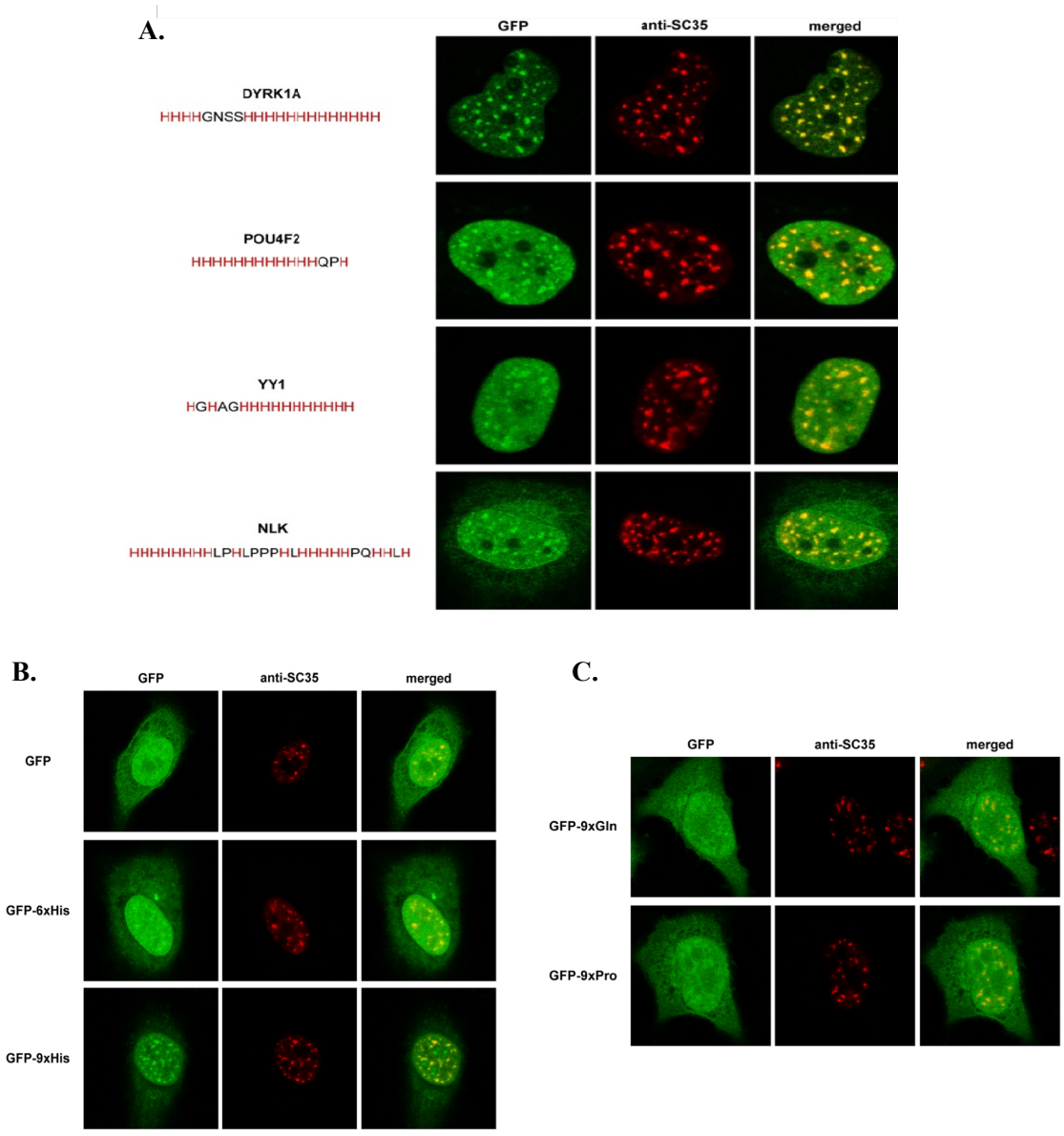

Figure 1-1. Cellular localization of poly-His proteins. (A) HeLa cells were transfected with plasmids that expressed GFP fused to a poly-His sequence from an endogenously expressed protein. Cells were immunostained for SC35, a known splicing factor in nuclear speckles, to visualize localization to speckles (B) GFP-Poly-His fusion products are shown to localize to nuclear speckles, with longer poly-His tracts increasing the amount of localization. (C) GFP-poly-Q/P fusion products show little to no localization to nuclear speckles, suggesting localization occurs specifically with poly-His tracts. All three panels were adapted from (Salichs *et al*., 2009).

When the endogenous histidine tract was selectively deleted, the same proteins failed to localize to speckles (Salichs *et al*., 2009). Paralogous proteins that do not contain the natural poly-His tract failed to localize (Salichs *et al*., 2009). When increasing

lengths of a poly-His tract were artificially added to Green Fluorescent Protein (GFP), the GFP localized to nuclear speckles (Fig 1-1, Panel B). Poly-His tracts were the only single-amino acid repeat to show this localization, when compared to poly-Pro and poly-Gln tracts (Figure 1-1, Panel C). This work demonstrated that poly-His tracts can and do work to localize proteins to cellular structures (Salichs *et al*., 2009).

Proteins and RNAs are regulated by sub-cellular localization

Sub-cellular localization of proteins has been observed to fluctuate when the organism is under stress, and may serve as regulators of gene expression to help the organism survive the stressor (Wek *et al*., 2006). Organisms must be able to adapt during periods outside ideal conditions. At the cellular level, stress can impede DNA synthesis and repair, repress transcription of housekeeping genes, and disrupt production of protein chaperones and protein synthesis. Cells adapt to stress using several mechanisms, including enzymatic antioxidant systems, activating promoters for transcription of mRNA crucial to mitigate the stressor (or repress transcripts that may not be necessary), and sequestration of mRNA for control of protein synthesis.

Translation of RNAs is controlled during stress responses by increasing the expression of translation-inhibiting RNA-binding proteins, and down-regulation of initiation factors by various complexes of proteins that bind to the mRNA, or "ribonucleoprotein complexes" (RNPs) (Buchan and Parker, 2009). This mechanism allows for rapid regulation of protein synthesis to cope with the stressor, but is also reversible when the stressor subsides (Buchan and Parker, 2009). As an example, eIF2

kinases can regulate translation depending on a variety of stress conditions, including viral infection, oxidative stress, and heat shock (Wek RC. 2006). By controlling the overall amount of translation occurring within cells, the cell can conserve energy.

Three particular kinds of RNPs are P-bodies, stress granules, and nuclear speckles (Buchan and Parker, 2009). There is a cytoplasmic mRNP cycle in which mRNPs can be moved between P-bodies, stress granules, and actively-translating polysomes (Buchan and Parker, 2009). Many of the proteins found in these complexes can form reversible amyloid protein conformations (such as those found in the proteins hnRNPA1, hnRNPA2, and Sup35).

Processing bodies, or P-bodies, are cytoplasmic RNP complexes that are always present and contain the translation repressors or mRNA decay machinery (Buchan and Parker, 2009). These P-bodies are known to interact with other mRNPs during various RNA lifecycles, for examples, exchanging components such as mRNAs when associating with stress granules (Buchan and Parker, 2009). Stress granules are also regulatory mRNPs, and serve as compartments for non-translating mRNA, translation initiation or repressor components, and other proteins that affect mRNA function (Buchan and Parker, 2009). Stress granules are an example of a cytoplasmic RNP granule formed through specific RNA-protein interactions, and are tightly regulated (Buchan and Parker, 2009). These mRNPs contain translation initiation factors, non-translating mRNAs, and several other proteins affecting mRNA function (Buchan and Parker, 2009). Disease-associated

mutations of hnRNPA1/A2 results in excess incorporation of these RNPs into stress granules as a mechanism of pathology (Kim *et al*., 2013).

Unlike P-bodies and stress granules, the third type of RNP granule exists in the nucleus. Nuclear speckles form throughout the nucleoplasm, generally in regions with little DNA content. Speckles may contain a small number of genes, but can be very active transcription sites. These speckles are thought to act as centers to organize active genes in proximity to one another to form cohesive euchromatic systems (Wek *et al*., 2006). Other evidence suggests that speckles function to store, assemble, and/or modify splicing factors (Wek *et al*., 2006).

Naturally-occurring poly-Histidine motifs

The discovery that poly-His motifs can localize proteins to nuclear speckles strongly suggests that poly-His tracts may represent a novel mechanism of directing proteins and their binding partners to cellular compartments (Salichs *et al*., 2009).

In the model plant organism *Arabidopsis thaliana*, there are 29 proteins that contain endogenous poly-His tracts. We hypothesize that the poly-His tracts in plant proteins are critical for the physiological function of these proteins. These endogenous poly-His tracts could function by metal-sensing/binding, as signal sequences for cellular localization, and/or bind or assist in binding other macromolecular ligands. One of these 29 proteins is the putative mRNA-binding protein LARP6b, which contains both a poly-His tract and an extended poly-His tract. Additionally, the *Arabidopsis* genome also

encodes two paralogs of LARP6b, LARP6a and LARP6c, which do not contain poly-His tracts. The extended poly-His tract of the LARP6b paralog is conserved across multiple plant species (Figure 1-2). We predict that the poly-His tracts of LARP6b mediate interactions with other proteins and/or RNAs to direct the cellular localization of LARP6b and its RNA ligands to stress granules. This work will use the three LARP6 paralogs of *Arabidopsis thaliana*, LARP6a, LARP6b, and LARP6c, to directly test this hypothesis.

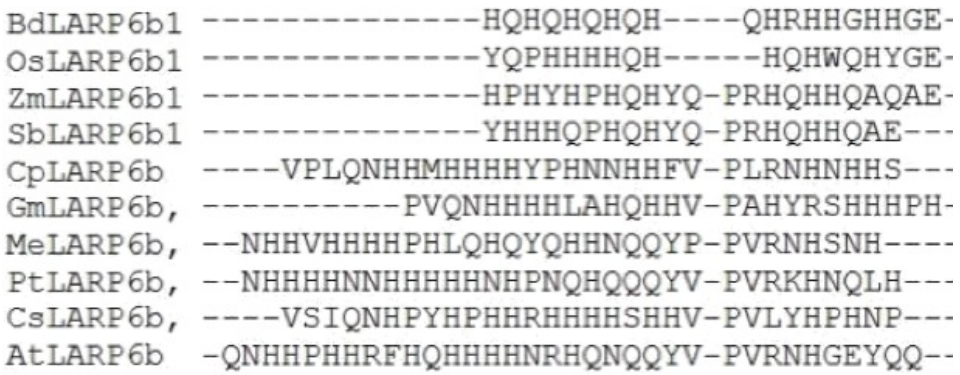

Figure 1-2. Extended poly-His tracts are found in most plant LARP6b proteins. A multiple sequence alignment of LARP6b proteins from plants shows that a histidine-rich motif is an evolutionarily conserved element of LARP6b. *Bd: Brachypodium distachyon, Os: Oryza sativa, Zm: Zea mays, Sb: Sorghum bicolor, Cp: Carica papaya, Gm: Glycine max, Me: Manihot esculenta, Pt: Populus trichocarpa, Cs: Cucumis sativus, At: Arabidopsis thaliana*

The La-related superfamily of RNA-binding proteins

The LARP6 proteins are members of a larger superfamily, called the La-related proteins (LARPs). All LARPs are RNA binding proteins that are associated with the processing and regulation of select RNAs (Bousquet-Antonelli and Deragon, 2009). LARPs are found in both animals and plants (Merret *et al*., 2013). Members of this family are characterized by two protein domains: a highly conserved La motif and an

adjacent RNA recognition motif (RRM). Each subfamily of LARPs has evolved its own unique RRM variant, which may reflect the distinct classes of RNA ligands bound by each subfamily (Figure 1-3). The founding member of this family is the "genuine La protein", which binds the 3'-UUU-OH motifs of pre-tRNAs that were transcribed by RNA Polymerase III, to both protect against 3'exonucleolytic degradation and chaperone the processing of these critical cellular molecules (Bousquet-Antonelli and Deragon, 2009). Genuine La proteins can also bind to viral RNAs, and may increase their survivability by increasing their rate of translation (Costa-Mattioli *et al*., 2004), or even by protecting these transcripts from host immune systems (Bitko *et al*., 2008). A close homolog, LARP7, also binds to 3'-UUU-OH RNA Pol III transcripts and negatively regulates RNA polymerase II transcription through interactions with 7SK snRNP (Stavraka and Blagden, 2015).

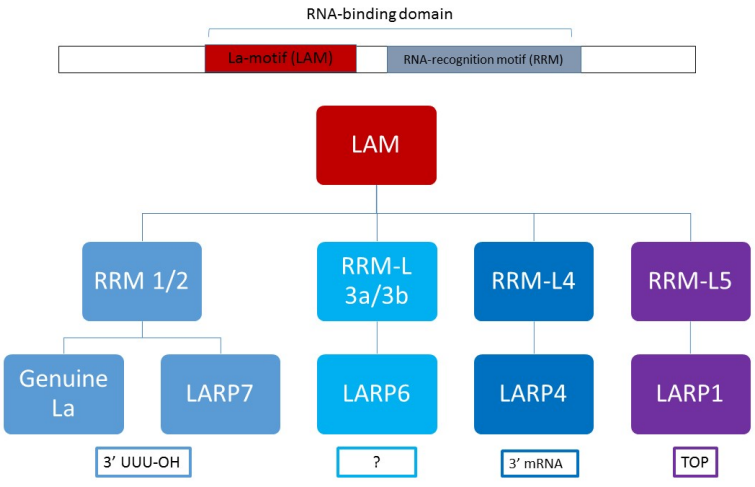

Figure 1-3. Divergence of RNA-binding domains across LARP proteins. Known ligands of each RNA-recognition motif are shown below each LARP.

LARP1 has been shown to both directly and indirectly associate with the untranslated regions of mRNAs *in vitro*, and is involved in cellular response to heat shock (Bousquet-Antonelli and Deragon, 2009, Merret *et al*., 2013). LARP4 also interacts with mRNA untranslated regions, and promotes mRNA stability through interaction with the poly (A)-binding protein (Schaffler *et al.,* 2010).

The LARP6 Subfamily

In addition to the La motif and RRM, LARP6 contains a conserved motif in its C-terminus, called the SUZ-C or LSA motif (for "La- and SUZ-C-associated" motif) (Bousquet-Antonelli and Deragon, 2009). This sequence is found in other RNA binding proteins and may be involved in intermolecular interactions and subcellular localization (Stavraka and Blagden, 2015). In mammals, LARP6 regulates the translation of the collagen type I α subunit (Martino *et al.*, 2015). Type I collagen is the most abundant protein in the human body, making up most muscle tissue, both skeletal and smooth (Smith and Rennie, 2007). With such a wide range of tissues affected, it is strongly suspected that LARP6 plays a role in the development of fibrosis, a disease characterized by normal tissue damage through excessive production of extracellular matrix components (Wynn, TA. 2008).

However, LARP6 is also conserved in eukaryotes that do not produce collagen, such as plants. Thus, the global cellular role of LARP6 proteins is not yet fully understood. In contrast to the one LARP6 gene found in most animals (*larp6*), some

plants contain three LARP6 genes (*larp6a, larp6b*, and *larp6c*). Only the LARP6b proteins have evolved an endogenous poly-His tract (Figure 1-4).

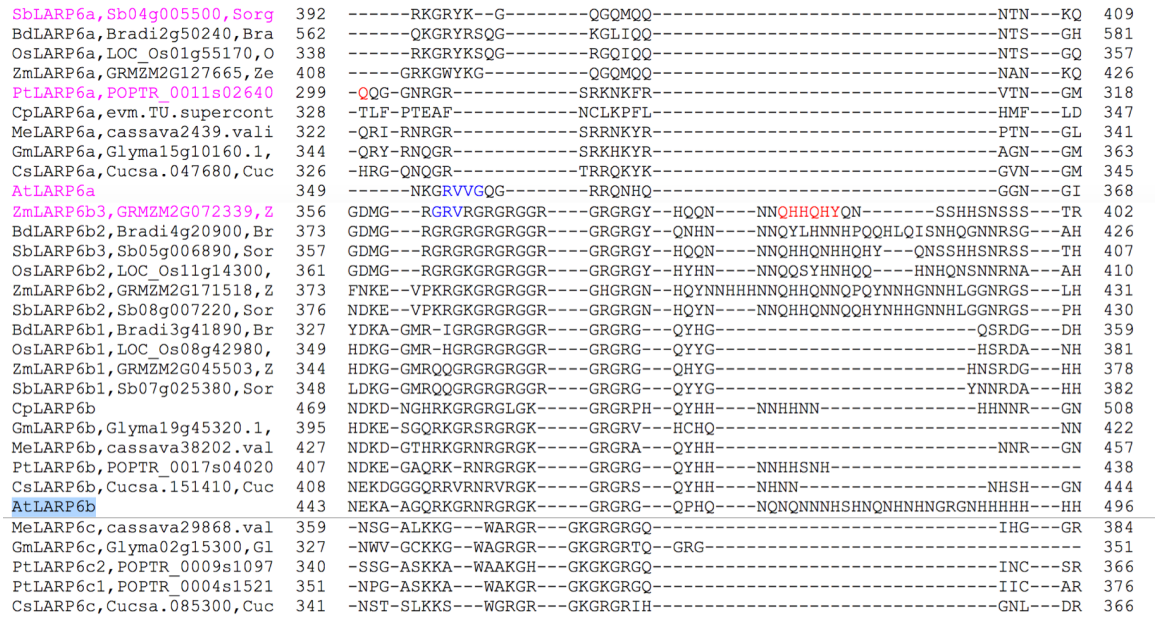

Figure 1-4. Multiple sequence alignment of LARP6 across species. Sequence alignment of LARP6a, LARP6b, and LARP6c from multiple species. Histidine-rich tracts are only found in LARP6b proteins. *Bd: Brachypodium distachyon, Os: Oryza sativa, Zm: Zea mays, Sb: Sorghum bicolor, Cp: Carica papaya, Gm: Glycine max, Me: Manihot esculenta, Pt: Populus trichocarpa, Cs: Cucumis sativus, At: Arabidopsis thaliana.*

The LARP6b of the model organism *Arabidopsis thaliana* has both a poly-His and an extended poly-His tract. The contiguous poly-His region is near the C-terminus of the protein sequence (residues 490-496), while the extended poly-His tract consisting of 10 histidine residues within a 16-residue motif is closer to the N-terminus (residues 111-126).

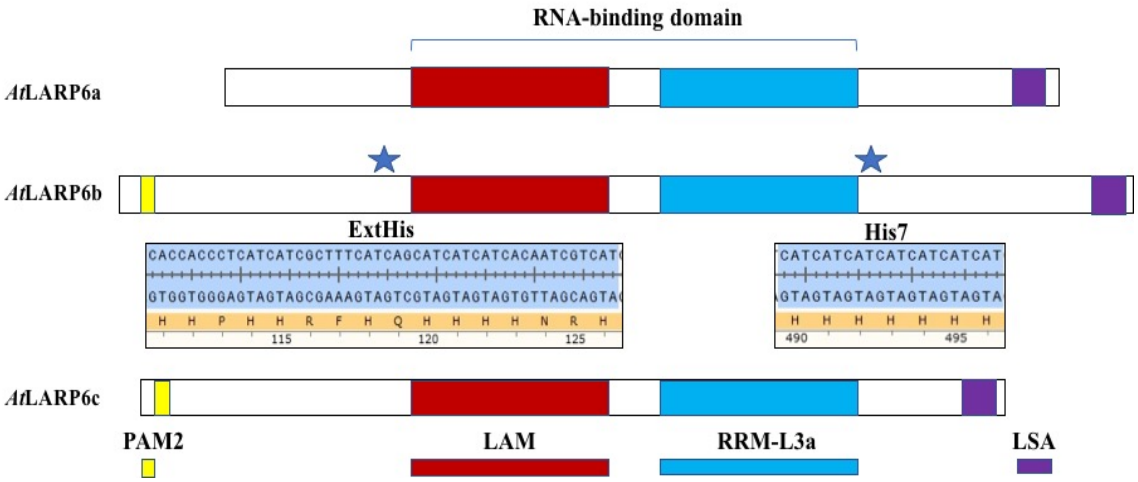

Figure 1-5. The *At*LARP6 proteins. *A.thaliana* expresses three paralogues of LARP6, each of which has a La motif, RNA recognition motif, and an LSA motif. LARP6b and LARP6c also possess a poly-A Binding Protein motif (PAM2). Only LARP6b contains poly-His tracts.

Another difference between the LARP6 paralogs of *Arabidopsis* is the presence of a Poly-A Binding Protein Association Motif (PAM2) in the N-terminus of both LARP6b and LARP6c. This PAM2 motif may play a role in mRNA regulation, as the mammalian LARP6 has been implicated in collagen mRNA regulation (Stefanovic *et al*., 2014, Merret *et al*., 2013).

Known characteristics of plant LARP6 proteins

Merret *et al.* were the first to study RNA binding activity of plant LARP6 proteins by conducting isothermal titration calorimetry (ITC) experiments. Merret and colleagues purified the La modules of LARP6a and LARP6c, but were unable to reliably express the LARP6b La module (Merret *et al*., 2013). Using 10-nt and 20-nt homopolymeric RNA tracts for A, U, G, and C nucleotides as ligands for ITC, they found that LARP6a binds the 20-nt poly(A) ligand with a fairly high affinity (K_d = 0.3 μM). In contrast, LARP6c

shows strong binding to the poly(U) ligand (K_d = 1.7 μM). This result suggests either different binding mechanisms and/or preferred target ligands (Merret *et al*., 2013).

In the studies described above, only the La modules of *At*LARP6a and *At*LARP6c have been previously purified. The work described in this thesis builds on this research by producing the full-length *At*LARP6 proteins. Several experiments were conducted to optimize the expression and solubility of each protein, leading to the novel purification of each full-length *A. thaliana* LARP6 protein. We began characterization of the full-length proteins through limited proteolysis and intrinsic disorder algorithms to probe for preliminary structural data. We also began to develop a platform for studying the role of poly-His tracts *in vivo*. An *Atlarp6b-egfp* fusion product was designed for future expression in *A. thaliana,* to be visualized through fluorescent microscopy.

II. MATERIALS AND METHODS

AtLARP6 Sequences

*At*LARP6a

MSSLPLRSGEKMESPSISDAVPLHAPEDATADFSQPQSPLHEVDSFPVTESSDD
VVVNVSEIPNLSPSDDDFDHERNSGEDRDQDHGENPVETDGVVVPIDELNQKII
RQVEYYFSDENLPTDKFLLNAMKRNKKGFVPISTIATFHKMKKLTRDHALIVSA
LKESSFLVVSADEKKVKRLSPLPEIRDPKIFTVLVENLPEDHSNENIREIFGKA
GSIKSVSICDPNAVEESEKGGKKENFIRTRLHAFVEYETVEAAEKAAATLNNEQ
DWRNGLRVKLLEQAAGKFAQRRPARREVDKEKDTTGRVHDQTGGEKNKKTREHQ
NHRLHHSDNPADDDGGNHQKDKNGNKGRVVGQGRRQNHQGGNGIGHGTASSSSH
PNYHPVEVSKRPPGPRMPDGTRGFTMGRGKAIPPPTSTQTSHEV

- Amino acids: 422
- Molecular weight: 46.84183 kDa
- Theoretical pI: 6.11
- Molar Extinction coeffiecient: 11460

*At*LARP6b

MADQQTLDSSTPPPTQSDDLSHSHSTSSTTSASSSSDPSLLRSLSLSLSRLNAGAP
EFVPGRTTPPLPQPPRMIIPPPPPHGMLHMYHHQPPFNTPVLGPVPIQPHLVPV
QNHHPHHRFHQHHHHNRHQNQQYVPVRNHGEYQQRGGVGGEQEPDLVSKKNDRR
DHSKRESKNDQVTETGASVSIDSKTGLPEDSIQKIVNQVEYYFSDLNLATTDHL
MRFICKDPEGYVPIHVVASFKKIKAVINNNSQLAAVLQNSAKLFVSEDGKKVRR
ISPITESAIEELQSRIIVAENLPEDHCYQNLMKIFSTVGSVKNIRTCQPQNNGS
GAPPAARSAAKSDGTLFSNKVHAFVEYEIVELAERAVTELSEAGNWRSGLKVRL
MLKHQTKEPKQGQGRGRKGHDADVEHEEDDATTSEQQPIEKQSDDCSGEWDTHM
QEQPIGEDQGNEKAAGQRKGRNRGRGKGRGRGQPHQNQNQNNNHSHNQNHNHNG
RGNHHHHHHHQVGTQPSNNPMNNMEQPGMGKQQPPGPRMPDGTRGFSMGRGKPV
MVQAE

- Amino acids: 545
- Molecular weight: 60.58897 kDa
- Theoretical pI: 7.9
- Extinction coefficients: 23170 native, 22920 reduced

*At*LARP6c

MAQMQREEVESVTTEKKRLDGGGGSSGAQATAFKFNAQAPEFVPRSHTTAPAPQ
VSPVSGYFYPCFHYNGGCIGGCGGGVCGGGGTGVGTQSSDWIYVGGGDPTAQHQ
HVHDPAAAFYISNPAVQFPASQNSSSSSKNLLSDDLRLKIVKQVEYQFTDMSLL
ANESISKHISKDPEGYVPVSYIASTKKIKALTSNHHLVSLALRSSSKLVVSEDG
KKVKRTSQFTDRDREELQGRTVVAENLPDDHSYQNLEKIFGVVGNVKAIRICHP
PESNSSRPKGDFLMSNKIHALIEYDNTVIADKAVEKLNDERNWRKGLRVRLLLR
CSPKSVLKNRRNFDGILIDDELPSYESGEDSPRLHLTESQLDNDGDDNNVGGLW
GKGRGKGRGRSPRSYAVGGGGRSFGIGLGVSLGIPSLGSHESSSPKTATKGPRM
PDGTRGFTMGRGKPSISLSPNNL

- Amino acids: 455
- Molecular weight: 48.95874 kDa
- Theoretical pI: 9.02
- Molar extinction coefficients: 34755 native, 34380 reduced

Transformations

DH5α ultra-competent cells were thawed on ice from -70˚ C storage. 0.5 μL pET28-SUMO His_6 *At*LARP6 expression was added to DH5α cells and incubated on ice for 30 minutes. Cells were then heat-shocked at 37˚ C for 1 minute and 30 seconds. Cells were placed on ice for 2 minutes following heat-shock for recovery. 700 μL of LB was added to cells, and cells were placed in a 37˚ C incubator at 250 rpm for 1 hour. Cells were removed and plated on antibiotic selection LB-Agar plates and placed in 37˚C incubator overnight.

Overnight cultures

For DNA minipreps: 5 μL of antibiotic was added into 5 mL of LB media, colonies from plates were scraped onto a pipette tip and dropped into glass tube containing media/antibiotic. Tubes were placed in the 37˚ C incubator shaking at 250 rpm overnight.

For large-scale expressions: 50 μL of primary and secondary antibiotics were added to 50 mL of LB media, colonies from plates were scraped onto a pipette tip and dropped into flask containing media/antibiotic. Flasks were placed in the 37˚ C incubator shaking at 250 rpm overnight.

Recombinant protein expression

Flasks containing 1L LB media were inoculated with 10 mL of overnight culture and placed in a 37˚ C incubator under shaking conditions. OD600 was monitored one hour after incubation, and subsequently monitored every 30 minutes afterwards until reading was between 0.4 – 0.8. Cultures were removed and placed on ice for 10 minutes,

with occasional shaking. 1 mL of 1M ITPG was added to cultures, and flasks were placed in a 16˚C incubator under shaking conditions overnight. Cultures were removed and spun down in 250 mL increments at 5000x*g* for 8 minutes at 4˚ C. Pellets were stored at -20˚C.

DNA electrophoresis and extraction

PCR products were verified and extracted via agarose gel electrophoresis. 0.7% agarose gels were loaded with samples and run for 1 hour at 90 V. DNA was visualized under UV light for presence of intercalated ethidium bromide. DNA bands were excised from gel and purified using the Omega E.Z.N.A Gel Extraction Kit via spin protocol.

Cloning

Cloning vectors and primers were designed electronically using SnapGene restriction site cloning. Sequence for the pER-HA plasmid was provided by Dr. Hong-gu Kang. Restriction sites XhoI and SnaBI were built into the primers to amplify the *At*LARP6b-eGFP fusion products. Completed SnapGene plasmids were created through SnapGene's restriction site cloning feature.

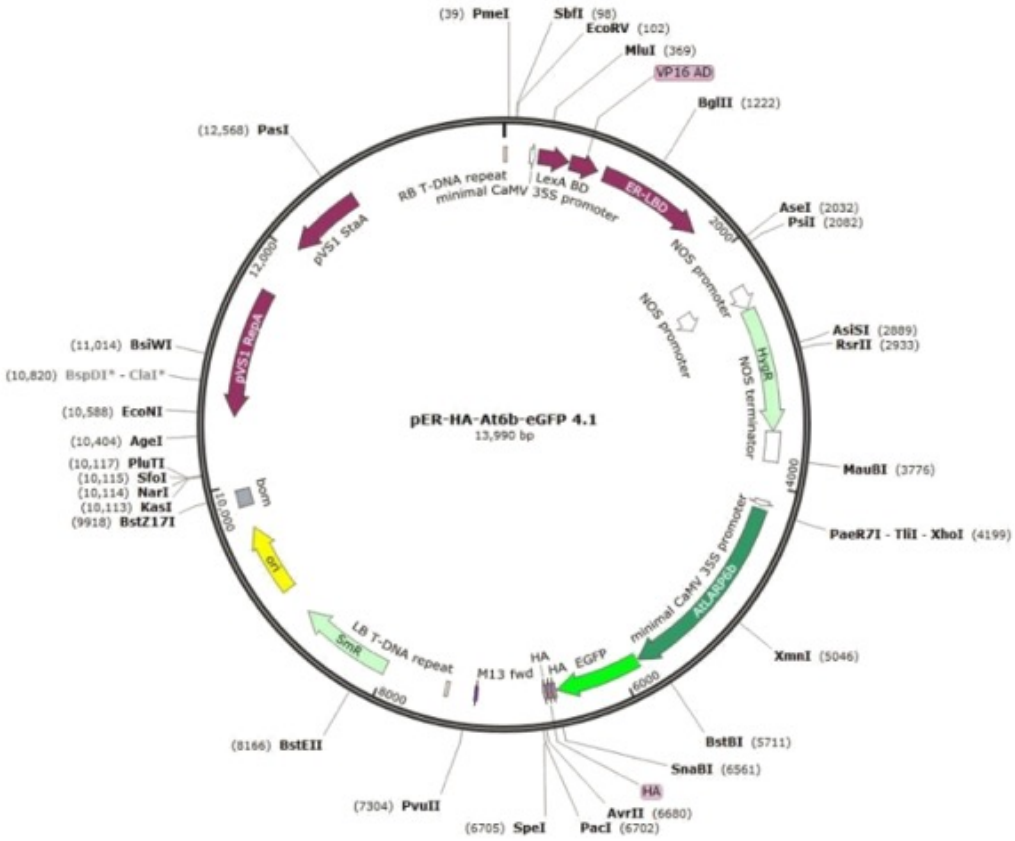

Figure 2-1. ***Arabidopsis thaliana*** **cloning vector:** SnapGene created vector of the pER-HA-AtLARP6b-eGFP construct to be transfected into *A. thaliana* cells.

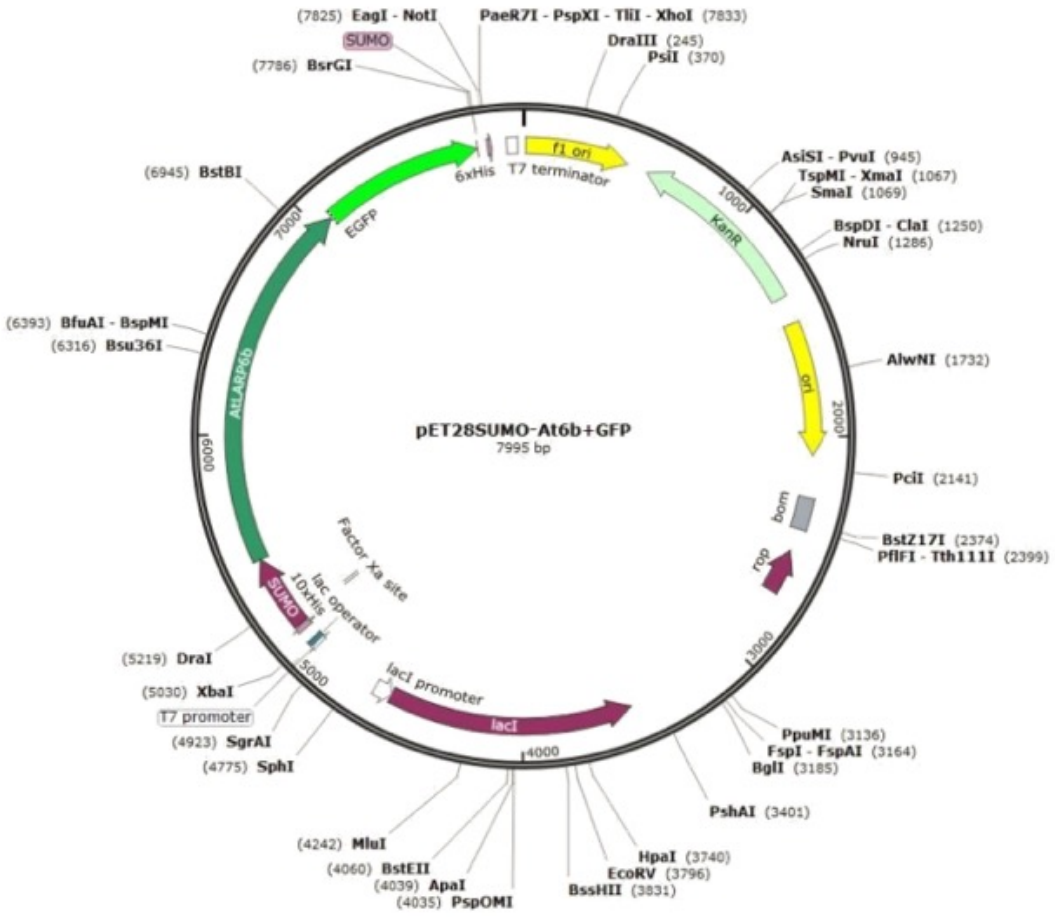

Figure 2-2. Proposed vector for bacterial cloning of *AtLARP6*b-eGFP gene product fusion. SnapGene created vector of the pET28SUMO-*AtLARP6b-eGFP* construct to transfect *E. coli* cells.

Purified DNA fragments were digested using 1 µL of Restriction Enzyme per 20 µL of DNA, with CutSmart Buffer and incubated at 37˚ C for 1 hour. Products were extracted and purified by agarose gel electrophoresis and Omega E.Z.N.A kit.

Table 2-1. Primers used in cloning procedures

Primer	Sequence	Function
KAL056	CATGCTCGAGATGGCTGATCAACAAACCCTAGATTCCTCTACTCCTCCTCCTACTCAATC	PCR fwd: AtLARP6b for insertion into pER-HA (XhoI)
KAL057	GTTCCATGGTCTCTGCCTGAACCATAACCGGCTTTCCCCGTCCCATTGAG	PCR rev: AtLARP6b for insertion into pER-HA with C-term eGFP (NcoI)
KAL058	CATCCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGG	PCR fwd: eGFP for insertion into pER-HA, C-term to NcoI site
KAL059	GTTGATTACGTAGCCTTGTACAGCTCGTCCATGCCGAGAGTGATCCCG	PCR rev: eGFP for insertion into pER-HA, N-term and in frame with 3xHA tag (SnaBI)
KAL075	CAGCAGTCGAGGTAAGATTAG	Sequencing fwd: pER-HA (upstream of CaMV 35S promoter)
KAL076	GTGACTGAAACTGGAGCTTCTG	Sequencing fwd: AtLARP6b(nt 518-540, aa174-181)
KAL077	GGGAGGCCTGGATCGACTAGTTAATTAA	Sequencing rev: pER-HA (downstream of 3xHA)
KAL078	CATGCTCGAGATGGCTGATCAACAAACCC	PCR fwd: AtLARP6b for insertion into pER-HA (XhoI), for OE-PCR
KAL079	CCCTTGCTCACCATCTCTGCCTGAA	PCR rev: AtLARP6b for overlap with eGFP for OE-PCR
KAL080	TGGTTCAGGCAGAGATGGTGAGCAAG	PCR fwd: eGFP for overlap with AtLARP7b, for OE-PCR
KAL081	GTTGATTACGTAGCCTTGTACAGCTCGTCCATGCC	PCR rev: eGFP for insertion into pER-HA (SnaBI), for OE-PCR
KAL082	GCCTGATGGGACCAGAGGATTCTCAATGGGACGGGGAAAGCCGGTTATGGTTCAGGCA GAGATGGTGAGCAAGGGCGAGGAGC	PCR fwd: eGFP for overlap with AtLARP7b, for OE-PCR (take 2, longer overlap)
KAL083	ATGGGCACCACCCCGGTGAACAGCTCCTCGCCCTTGCTCACCATCTCTGCCTGAACCAT AACCGGC	PCR rev: AtLARP6b for overlap with eGFP for OE-PCR (take 2, longer overlap)
KAL094	GCAGTCGAGGTAAGAT	Sequencing: AtLARP6 Forward
KAL095	CTCCAGTTTCAGTCAC	Sequencing: AtLARP6 Midway Reverse
KAL096	GCCCTTGCTCACCAT	Sequencing: AtLARP6 Reverse
KAL097	TATGGTTCAGGCAGAGA	Sequencing: eGFP Forward
KAL098	GGACGAAAGCTGGGA	Sequencing: eGFP Reverse (including 3xHA tag) in pER-HA

Ligation

Digested and purified DNA fragments were ligated together using T4 Quick Ligase and T4 Quick Ligase buffer. Samples were incubated at room temperature for 10 minutes and then placed on ice.

Solubility screening assay

Cell pellets were thawed on ice, and 25 mL of 1x Standard Buffer (50 mM sodium phosphate mono/di basic, 200 mM NaCl pH 8.0) with protease inhibitor (ThermoFisher) was added. Cells were sonicated while on ice for 2 minutes (20 seconds of 1 pulse, 30 seconds of rest). Sonicated product was transferred to a 50 mL Oak Ridge centrifuge tube and centrifuged at 18,000x*g* for 25 minutes at 4˚C. 10 µL of supernatant was used for each sample, which ultimately contained 10 µL 10x standard buffer, 10 µL, 10x Additive (Table 2), 80 µL MiliQ H20. (Control contained 10 µL 10x standard buffer and 90 µL MQ H20). Samples were incubated at 4˚ C for 2 hours. Solutions were transferred into an Amicon Ultra 0.5 mL centrifugal filter system and centrifuged for 20 minutes at 15,000x*g* at 4˚C. Aggregated protein in the retentate fraction was re-suspended by vigorously pipetting across the membrane surface with 30 µL of MQ H2O. 20 µL of filtrate and retentate fractions were treated with 5 µL of 5x SDS Sample buffer and incubated at 90˚ C for 5 minutes. Samples were then loaded onto a 0.75mm 10% SDS polyacrylamide gel and electrophoresed at 200 V for 45-60 minutes. Proteins were transferred to a nitrocellulose membrane using a Trans-Blot Turbo on high molecular weight setting. For detection of His-tagged proteins, the membrane was placed in 2% BSA blocking solution in 1x TBS and incubated overnight at 4˚C with shaking. The membrane was then probed with a 1:5000 dilution anti-His probe conjugated to horseradish peroxidase (HRP) (Pierce) lyophilized in 0.1M MES pH: 5.5 for 1 hour under shaking conditions at 4˚C. The membrane was then washed with 1x TBS. The bound HRP was detected with homemade chemiluminescence solution (25 mL 1 M Tris, pH: 8.8, 1.25 mL 250 mM luminol, and 5.56 mL 90 mM 4IPBA) and 12.5 µL hydrogen

peroxide for 1 minute (Mruk and Cheng, 2011), and then imaged with a ChemiDoc XRS+ Molecular Imager.

Table 2-2. Chemical Additives for each solubility screening assay

Solubility Screening Assay:	*At*LARP6b	*At*LARP6a/ *At*LARP6c
Run 1	300 mM NaCl 10% Glycerol 0.5 M Urea 500 mM L-Arginine 0.5 M TMAO 1% Nonidet P40	50 mM Na_2SO_4 200 mM NaCl 200 mM KCl 100 mM $MgSO_4$ 50 mM $(NH_4)_2SO_4$
Run 2	50 mM Na_2SO_4 200 mM NaCl 200 mM KCl 100 mM $MgSO_4$ 50 mM $(NH_4)_2SO_4$	
Run 3	25 mM $(NH_4)_2SO_4$ 50 mM $(NH_4)_2SO_4$ 100 mM $(NH_4)_2SO_4$ 150 mM $(NH_4)_2SO_4$ 200 mM $(NH_4)_2SO_4$	
Run 4	5 mM $(NH_4)_2SO_4$ 10 mM $(NH_4)_2SO_4$ 15 mM $(NH_4)_2SO_4$ 20 mM $(NH_4)_2SO_4$ 25 mM $(NH_4)_2SO_4$	

Purification of recombinant AtLARP6 proteins

Table 2-3. Buffers used for Ni-NTA Elution and SEC

Buffers used :	*At*LARP6a	*At*LARP6b	*At*LARP6c
SEC Buffer	50 mM NaH_2SO_4/Na_2HSO_4 pH: 8.0 200 mM NaCl 50 mM Na_2SO_4	50 mM NaH_2SO_4/Na_2HSO_4 pH: 8.0 200 mM NaCl 10 mM $(NH_4)_2SO_4$	50 mM NaH_2SO_4/Na_2HSO_4 pH: 8.0 200 mM NaCl 50 mM Na_2SO_4
Lysis/Wash 1 Buffer	50 mM NaH_2SO_4/Na_2HSO_4 pH: 8.0 200 mM NaCl 50 mM Na_2SO_4 10 mM Imidazole pH: 8.0 5 mM 2-Mercaptoethanol	50 mM NaH_2SO_4/Na_2HSO_4 pH: 8.0 200 mM NaCl 10 mM $(NH_4)_2SO_4$ 10 mM Imidazole pH: 8.0 5 mM 2-Mercaptoethanol	50 mM NaH_2SO_4/Na_2HSO_4 pH: 8.0 200 mM NaCl 50 mM Na_2SO_4 10 mM Imidazole pH: 8.0 5 mM 2-Mercaptoethanol
Wash 2 Buffer	50 mM NaH_2SO_4/Na_2HSO_4 pH: 8.0 200 mM NaCl 50 mM Na_2SO_4 50 mM Imidazole pH: 8.0 5 mM 2-Mercaptoethanol	50 mM NaH_2SO_4/Na_2HSO_4 pH: 8.0 200 mM NaCl 10 mM $(NH_4)_2SO_4$ 50 mM Imidazole pH: 8.0 5 mM 2-Mercaptoethanol	50 mM NaH_2SO_4/Na_2HSO_4 pH: 8.0 200 mM NaCl 50 mM Na_2SO_4 50 mM Imidazole pH: 8.0 5 mM 2-Mercaptoethanol
Elution Buffer	50 mM NaH_2SO_4/Na_2HSO_4 pH: 8.0 200 mM NaCl 50 mM Na_2SO_4 350 mM Imidazole pH: 8.0 5 mM 2-Mercaptoethanol	50 mM NaH_2SO_4/Na_2HSO_4 pH: 8.0 200 mM NaCl 10 mM $(NH_4)_2SO_4$ 350 mM Imidazole pH: 8.0 5 mM 2-Mercaptoethanol	50 mM NaH_2SO_4/Na_2HSO_4 pH: 8.0 200 mM NaCl 50 mM Na_2SO_4 350 mM Imidazole pH: 8.0 5 mM 2-Mercaptoethanol

Ni-NTA affinity chromatography

Cells were thawed on ice and re-suspended in 30 mL of lysis/wash 1 buffer with added EDTA-free protease tablet. Cells were then sonicated at 30% amplitude at one pulse for a total of 2 minutes. Lysate was transferred to an Oakridge centrifuge tube and centrifuged at 18,000x*g* for 20 minutes at 4˚ C.

Ni-NTA beads suspended in 20% ethanol were equilibrated for binding by first mixed with MilliQ water, pelleted by centrifugation, and re-suspended in MilliQ water three times before repeating the process with two volumes of lysis buffer. Beads were then re-suspended in lysis buffer and added to the supernatant from the centrifuged lysate. The mixture was incubated for one hour at 4˚C under shaking conditions.

The bead/lysate mixture was decanted into a glass flex column, and the beads allowed to settle. The solution in the column was collected in a conical vial as the flow-through fraction. The beads were washed with 20 mL of lysis buffer, which was collected as a single fraction. A second wash was performed with 24 mL of Wash 2 buffer, which was collected in four 6 mL fractions. Finally the His-tagged proteins were eluted with 24 mL of Elution buffer, which was collected in 4 mL fractions. Fractions were analyzed by 10% SDS-PAGE and Coomassie blue staining to identify fractions that contained protein of interest.

Ni-NTA elutions were pooled in a Vivaspin Turbo 50 mL conical vial with 30,000 MWCO filter and initially concentrated at 4 ˚C for 10 minutes at 4,000x*g*. Additional concentration was conducted by increments of 5-10 minutes until the retentate volume was 1.5-3 mL. Retentate was further filtered through a Fisher 0.2 μm syringe filter and injected into the Sephadex S200 gel filtration column for size exclusion chromatography (SEC). Elution from the SEC was collected in 2 mL fractions at a flow rate of 1 mL/min.

Size exclusion chromatography

FPLC fractions corresponding to the major peaks of absorbance at 280 nm were analyzed by SDS-PAGE gel to determine fractions containing purified protein. Protein-containing fractions were pooled and treated with 450 μL 24.7 μM ULP1 enzyme (enough enzyme to account for the unknown protein concentration) and incubated for 2 hours at 16˚C. The solution was again concentrated via centrifugal filter at 4˚C and

filtered through a 0.2-micron syringe filter and injected into S200 column for fraction collection at 1 mL/min flow rate to separate the full-length LARP6 protein from SUMO and ULP1.

Eluted FPLC fractions were analyzed by SDS-PAGE and Coomassie blue staining. Purified and cleaved protein fractions were pooled and snap frozen by liquid nitrogen into 50 µL aliquots prior to storage at -70 ºC.

Limited proteolysis

Limited proteolysis by trypsin was conducted with *At*LARP6 protein on ice, using 240 µL of protein and 4 µL of 1 mM trypsin. 20 µL samples collected at set time points and immediately boiled at 90˚C. After 3 hours, samples collected over the time course were subjected to SDS-PAGE and subsequently prepared for silver staining. SDS-PAGE gels were fixed in 50% ethanol for 30 minutes before staining with silver solution for one hour. SDS-PAGE gels were washed with 200 mL of MilliQ water three times before being exposed to developing solution. SDS-PAGE gels were shaken by hand in developing solution until bands were clearly visible. Kill solution was added to arrest further development of the SDS-PAGE gels, and gels were imaged after 30 minutes of incubation with kill solution.

SLIDER disorder prediction data

Sequences of *At*LARP6 proteins were entered into the SLIDER intrinsic disorder predictor, and compared to the human (*Hs*) HNRNPA1 RNA-binding protein, which is relatively well characterized when compared to the LARP6 proteins (Table 2-4). A SLIDER score above 0.538 is indicative of predicted disorder within the protein (Peng *et al.,* 2014).

Table 2-4. SLIDER Scores

Prot. ID,SEQ,SLIDER score (the higher the more likely a protein has long disorder segment)
*At*LARP6a (0.8700570293766754)
*At*LARP6b (0.8772569036081675)
*At*LARP6c (0.7311140512396886)
*Hs*HNRNPA1 (0.6392310511576451) MSKSESPKEPEQLRKLFIGGLSFETTDESLRSHFEQWGTLTDCVVMRDPNTKRSRGFGF VTYATVEEVDAAMNARPHKVDGRVVEPKRAVSREDSQRPGAHLTVKKIFVGGIKEDTEE HHLRDYFEQYGKIEVIEIMTDRGSGKKRGFAFVTFDDHDSVDKIVIQKYHTVNGHNCEV RKALSKQEMASASSSQRGRSGSGNFGGGRGGGFGGNDNFGRGGNFSGRGGFGGSRGGGG YGGSGDGYNGFGNDGGYGGGGPGYSGGSRGYGSGGQGYGNQGSGYGGSGSYDSYNNGGG GGFGGGSGSNFGGGGSYNDFGNYNNQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQ GGYGGSSSSSSYGSGRRF

III. EXPRESSION AND PURIFICATION OF *ARABIDOPSIS THALIANA* LARP6 PROTEINS

Introduction

The isolated RNA-binding La modules of *At*LARP6a and *At*LARP6c have been previously successfully expressed and purified as recombinant proteins in *E. coli* (Merret *et al*., 2013). However, the full-length proteins have not been purified or biochemically characterized, which is critical for understanding the intramolecular interactions that may regulate LARP6 function. Preliminary work in our own laboratory indicated that the full-length *At*LARP6a and *At*LARP6c proteins could be produced in *E. coli* when fused at the N-terminus to the SUMO protein (D. A. Horn and K. A. Lewis, unpublished). However, previous attempts to purify either the La module of *At*LARP6b (Merret *et al*., 2013) or the full length *At*LARP6b (by Daniel Horn in our own laboratory) were unsuccessful.

We hypothesized that the addition of solubility-enhancing chemical additives to the purification buffer conditions would prevent protein aggregation and improve protein yield. Churion and Bondos devised a protocol for identifying ideal chemical additives to improve solubility of proteins, especially those that may be intrinsically disordered (Churion and Bondos, 2012). The assay starts by testing osmolytes of varying chemical characteristics to identify what category of chemical interaction is the most effective. These chemicals affect the entropy of the solvent to improve protein stability, allowing for greater protein purification yields (Churion and Bondos, 2012).

Solubility screening assays were conducted on cell lysates containing the proteins of interest to determine the ability of various classes of chemicals to enhance fusion protein solubility. The screening assays were then repeated using various chemicals from that helpful class to identify which specific chemical was most effective, and at what concentration. The optimal concentration of the identified chemical additive was then added to the Ni-NTA and SEC buffers for large-scale purification. To obtain biochemical quantities of LARP6 proteins, the N-terminal SUMO-tagged *At*LARP6 proteins were separated from crude cell lysate first by Ni-NTA affinity chromatography, and further by size exclusion chromatography (SEC). After confirming the presence of the protein by SDS-PAGE, the SUMO tags were cleaved from the LARP6 protein of interest using the enzyme ULP1, and the cleaved SUMO separated from LARP6 using SEC.

Results

Calibration of SEC S200 Column

Before carrying out any protein purification, it was necessary to generate a standard curve for elution from the Sephadex 200 gel filtration column (GE Healthcare). Size exclusion chromatography (SEC) separates molecules by molecular weight using porous beads that accommodate smaller molecules, causing larger molecules to elute more quickly. Because each size exclusion column can pack slightly differently, it is important to calibrate this particular column in order to precisely calculate the apparent molecular weights of eluted proteins.

To establish a standard curve, a set of proteins (and a sugar, Blue Dextrose) with known molecular weights ranging from 6.5 -2000 kDa were applied to the column. By eluting various combinations of these proteins from the S200 column, the average elution volume (K_{av}) was determined for each protein using Equation 1.

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

where V_e is the elution volume of the protein, V_c is the total column volume (our column volume is 120 mL), and V_o is the void volume (empirically determined as described below). To ensure baseline resolution of the standard proteins, molecular weight standards were analyzed using three separate SEC runs. Generally, in size exclusion chromatography, elution volume is inversely related to molecular weight, as proteins with higher molecular weight elute from the column first and smaller molecular weight proteins elute later. The molecules used to create the standard curve were grouped into high and low molecular weight groups, eluting as the blue and gray peaks, respectively (Figure 3-1, panel A). Blue Dextran was used to identify the void volume, as it would not interact with any of the column beads due to the large molecular weight (2,000 kDa).

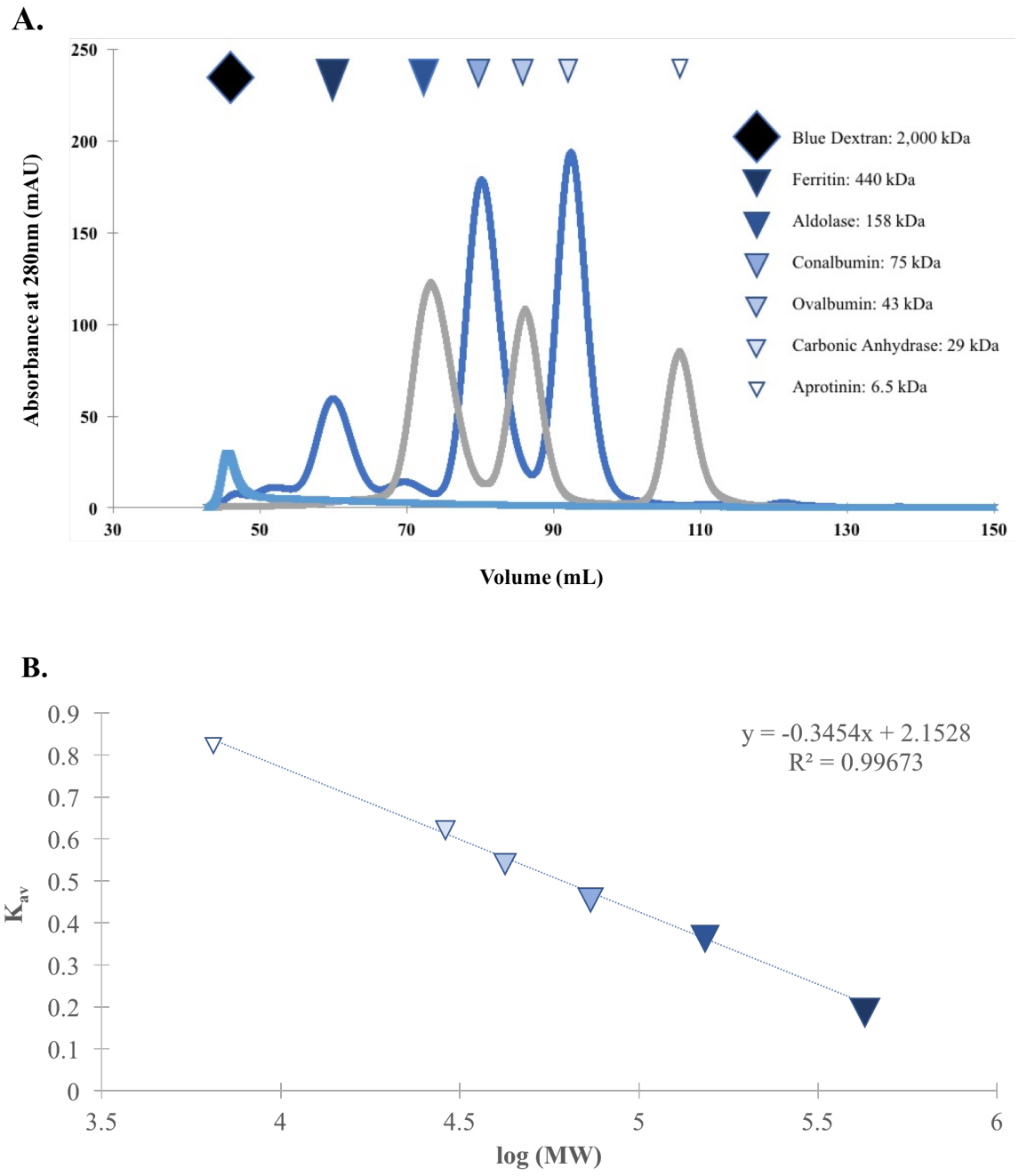

Figure 3-1. Calibration of S200 column with molecular weight standards. (A) Overlaid elution profiles of the GE Molecular Weight standard kit. Peaks of each protein are marked above the UV chromatography curve, with unique markers corresponding to each protein. (B) K_{av} was calculated per Eq. 1, and plotted as a function of the logarithm of molecular weight (in Daltons). The data were fit with least-squares linear regression line.

Our lab has previously attempted to obtain the full-length *At*LARP6b-SUMO protein, but it appeared that LARP6b aggregated during purification, as demonstrated by indistinguishable elution peaks from the calibrated S200 column (Figure 3-2).

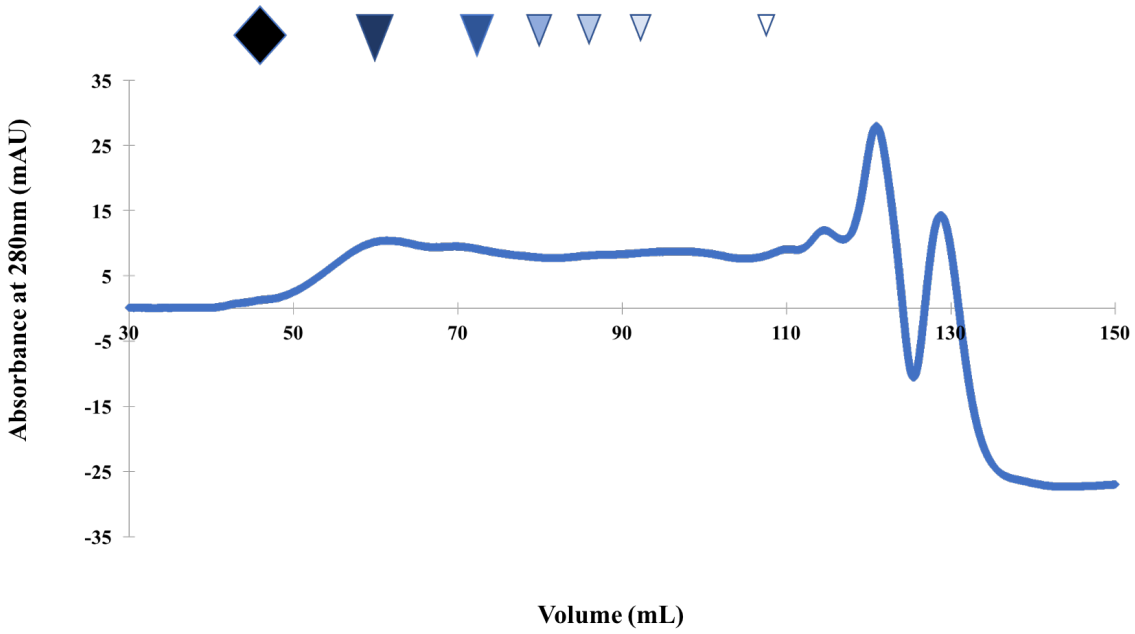

Figure 3-2. **Initial purification of *At*LARP6b.** *At*LARP6b chromatogram of a Sephadex 200 column elution lacks a sharp peak, indicating a poly-disperse solution, which is consistent with protein aggregation. This result prompted solubility screening using different additives to prevent protein aggregation and to promote protein stability.

Optimization of protein expression conditions

To successfully purify recombinant *At*LARP6b protein, the first task was to optimize protein expression. Sufficient production of recombinant protein by bacteria is vital to the process of retrieving purified protein. Bacterial plasmids containing the gene for the protein of interest were transformed into competent cells and grown to an optimal optical density value (OD_{600} = 0.5-0.8) to ensure bacteria were in the logarithmic phase of growth when protein expression was induced. Optimal expression of proteins will

contribute to an improved yield of purified protein by producing an adequate amount of starting material before the purification process begins. The β-galactoside analog isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to the cultures to induce production of the product protein by activating the *lac* promoter, which controls expression of the T7 RNA polymerase that in turn transcribes the desired gene (Studier, F. 2005). Cells were harvested by centrifugation after completion of the expression period and stored as frozen pellets prior to protein purification.

Identifying water sources for bacterial culture – At the time that protein expression was being tested, the department had recently upgraded the house deionized water system to an Evoqua-brand system, which produced >18.0 MΩ water. To test if the upgraded water was suitable for bacterial growth and protein expression cultures, growth curves were measured with four cultures of *E. coli* grown in LB prepared with either MilliQ-polished diH_2O (2 cultures, "MilliQ A" and "MilliQ B") or house deionized water LB (2 cultures, "House A" and "House B"). 10 mL of the same overnight culture was used to inoculate each liter of LB, and the cultures grown in the same shaker at 37ºC. Bacterial growth was measured by OD_{600} over several hours (Figure 3-3). Out of the four cultures, the house deionized "B" and MilliQ "A" cultures grew the fastest, possibly because they were positioned nearest to the heat vents inside the incubator shaker. Overall, though, OD_{600} readings were very similar, indicating that house deionized water was suitable for bacterial expression.

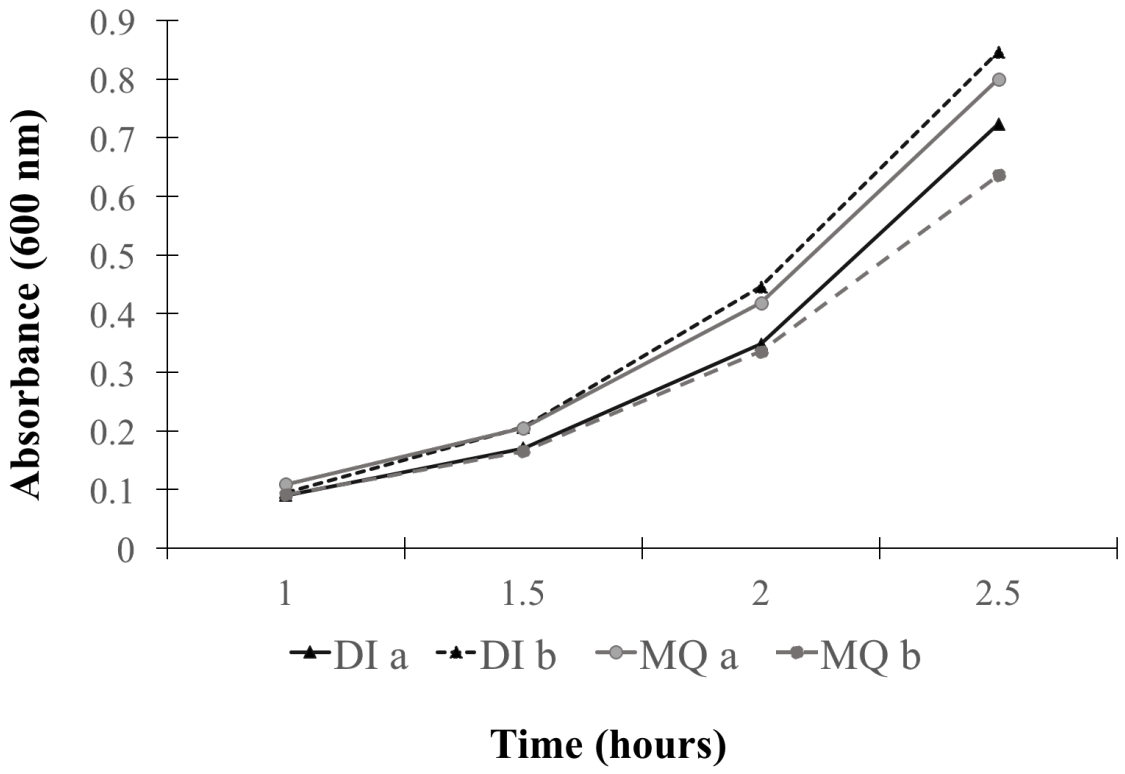

Figure 3-3. Testing bacterial growth using two deionized water sources. Growth of *E. coli* containing pET28:SUMO-*At*LARP6b plasmid before induction of recombinant protein expression with IPTG. Two independent cultures were grown in LB media prepared with MilliQ-polished water, and two were grown in LB media prepared with Evoqua house deionized water. All cultures were grown at 37ºC. Beginning at one hour after inoculation, OD_{600} measurements were taken every thirty minutes. No significant differences in cell growth were observed between the different water sources.

Expression of *At*LARP6b – Previously in the lab, *At*LARP6b was robustly expressed at 23 ˚C for 4 hours (D.A. Horn and K.A. Lewis, unpublished). Repeated expression trials showed that expressing protein for 16 hours at 16 ˚C yielded much stronger bands than 4 hours at 23 ˚C, though degradation products were also more apparent (Figure 3-4).

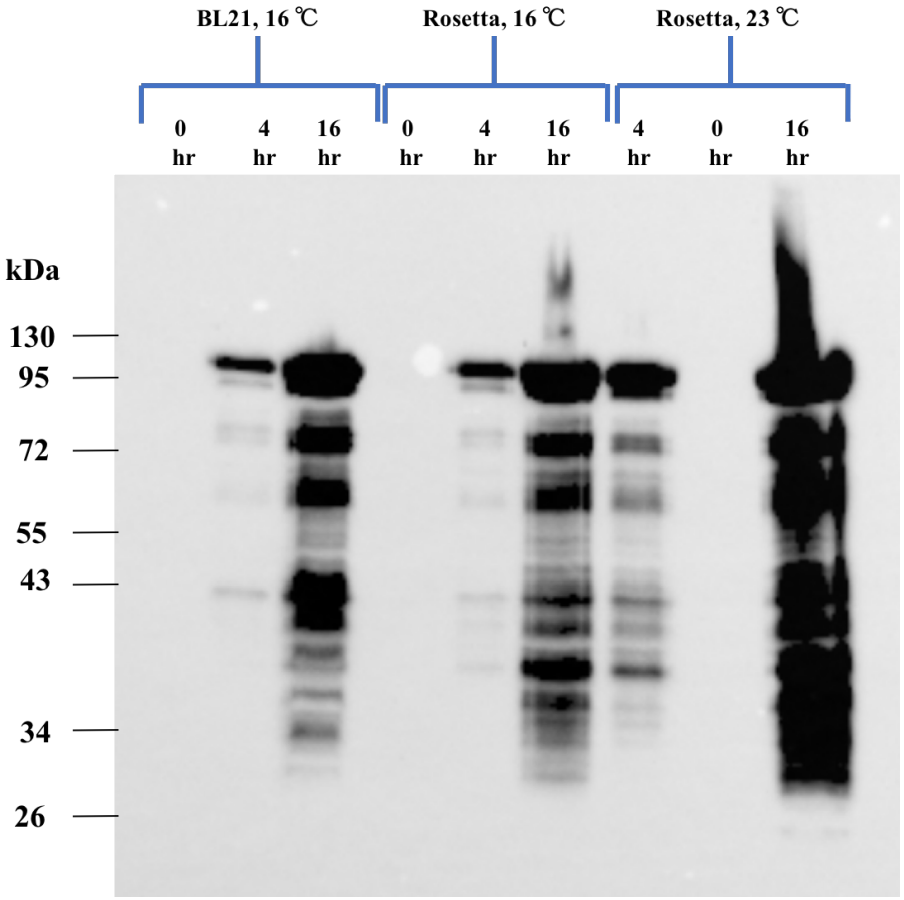

Figure 3-4. Small-scale expression trial of *At*LARP6b. Western blot of small scale expression conducted by DAH shows protein expression levels across varying times, temperatures, and cell lines (BL21 vs Rosetta *E. coli* cells).

Expression of *At*LARP6c – Optimal expression of *At*LARP6c was another concern when attempting to purify the full-length protein. Daniel Horn had previously expressed *At*LARP6c in the Rosetta *E. coli* strain at 23˚C for 6 hours, which produced bands when visualized on a SDS-PAGE gel after Ni-NTA elution, but failed to produce workable concentrations of protein eluted from the SEC (Figure 3-5).

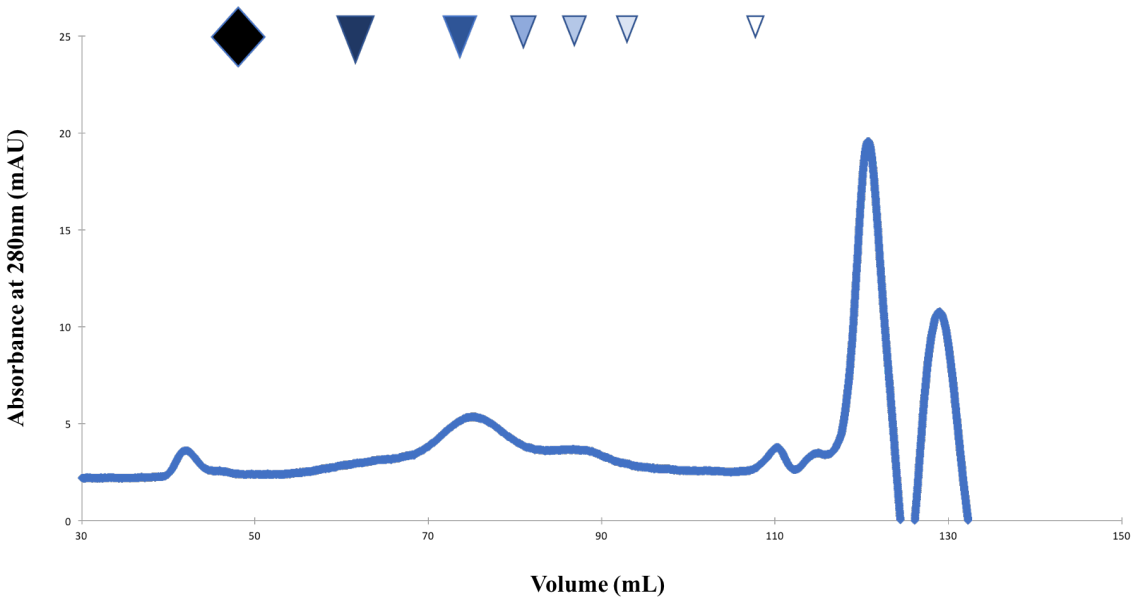

Figure 3-5. Initial purification of *At*LARP6c. Elution profile of *At*LARP6c-SUMO over S200 did not yield any sharp peaks where protein would be expected. Only a small, broad peak was observed near 75 mL, where the protein might be present.

To remedy this, just like for SUMO-*At*LARP6b above, SUMO-*At*LARP6c was expressed overnight at 16 ˚C, and then harvested. As discussed in detail below, a large amount of protein was recovered from the cell lysate, which was suitable for further purification by SEC.

Expression of *At*LARP6a – Similarly, previous strong expression of *At*LARP6a had been observed when expressed at 23 ˚C for 6 hours. However, attempts to reproduce expression with this protocol were unsuccessful, yielding very low amounts of protein recovered from Ni-NTA elutions (data not shown) and low, broad peaks observed on the SEC elution profile (Figure 3-6).

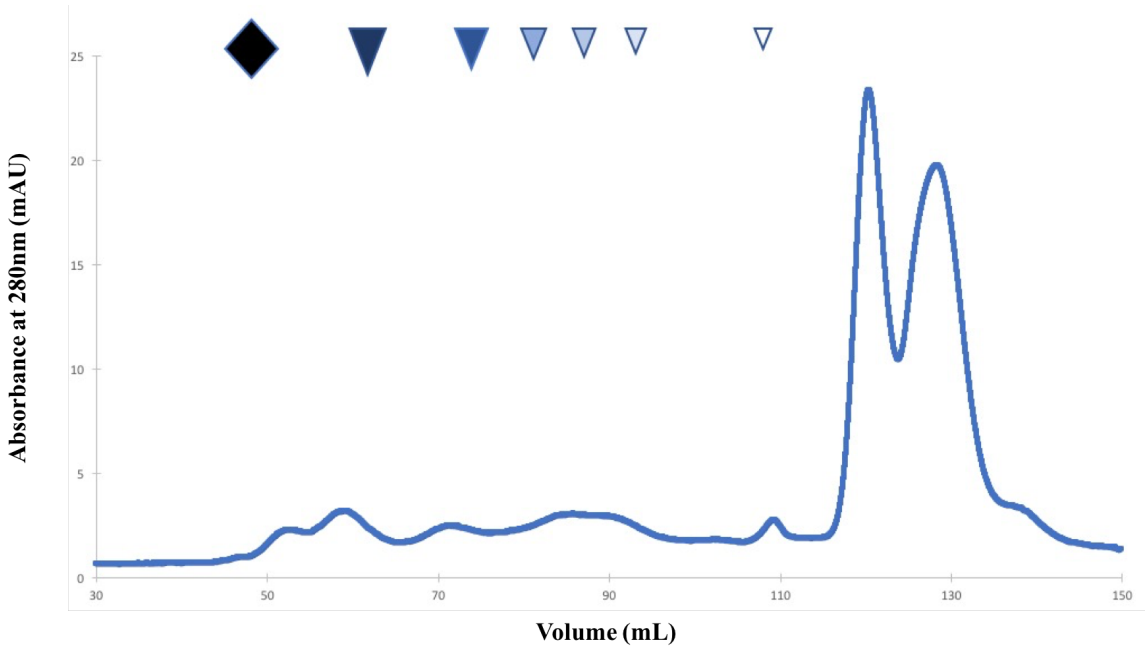

Figure 3-6. Initial purification of *At*LARP6a. Elution profile of *At*LARP6c-SUMO over S200 did not yield any sharp peaks where protein would be expected. Only a small, broad peak is observed near 60 mL, before where the protein would be expected.

Therefore, small-scale expression trials were conducted to monitor the production of protein at various times throughout expression. In contrast to our full-scale expression cultures that used 1 L liquid broths, small-scale expressions were conducted in 100 mL LB. *At*LARP6a was expressed at 16 ˚C, with one culture containing an additional 1% glucose to slow the rate of protein production by lowering cyclic AMP (cAMP) levels and decreasing binding of the cyclic AMP receptor protein upstream of the *lac* promoter (Studier, F. 2005). The hypothesis was that the lower temperature would enhance protein folding and stability, and that the glucose may additionally enhance expression yields. Expressing *At*LARP6a at 16˚C for 23 hours seemed to be the most effective, producing the most prominent bands where the full-length protein is expected when analyzed by

SDS-PAGE, in the presence or absence of glucose (Figure 3-7). Future large-scale expressions were conducted in the absence of glucose, for simplicity.

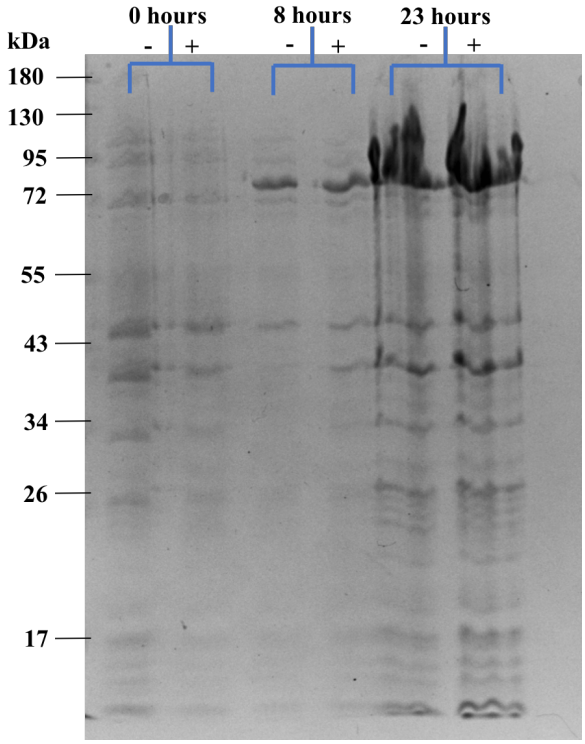

Figure 3-7. Small-scale expression trial of *At*LARP6a. Expression at 16 ˚C both in the presence (+) and absence (-) of 1% glucose. Samples taken from the cultures at 23 hours had produced much larger amounts of protein than those taken earlier in the expression. Glucose did not appear to improve expression.

Solubility Screening Assays

We hypothesized that buffer conditions could be identified that would improve the solubility and prevent the aggregation of *At*LARP6b, resulting in improved protein yields during purification and increased stability of the protein during biochemical characterization experiments. A solubility screening assay was used to determine chemical additives that may improve the solubility of *At*LARP6b (Churion and Bondos, 2012). The first iteration screened multiple categories of additives, including kosmotropic

and chaotropic osmolytes, to identify the best category for solubilization. Kosmotropes are chemicals that improve the stability of protein-protein interactions, including intramolecular bonding, primarily through the formation of ion-dipole bond networks between the protein and the solvent. Chaotropes improve the stability of proteins by moderating the interactions between hydrophobic parts of proteins and the otherwise ordered hydrogen bonding networks in the aqueous solvent.

The solubility screening assays were conducted using cell lysates, where samples of lysate were incubated with different chemical additives and then centrifuged through a filter. SDS-PAGE and Western blotting, using an anti-His antibody, were then used to measure the protein content in the filter retentate and flowthrough fractions ("R" and "F", respectively).

Based on the Western blot obtained after the first iteration of the assay, it was determined that NaCl (a kosmotrope) and amino acids (acting as osmolytes) would be the most effective solubility enhancing agents (Figure 3-8). While other osmolytes such as glycerol and L-arginine also improved solubility, glycerol produced a small band in the retentate lane, and the amino acids such as L-arginine might be more reactive in changing pH conditions.

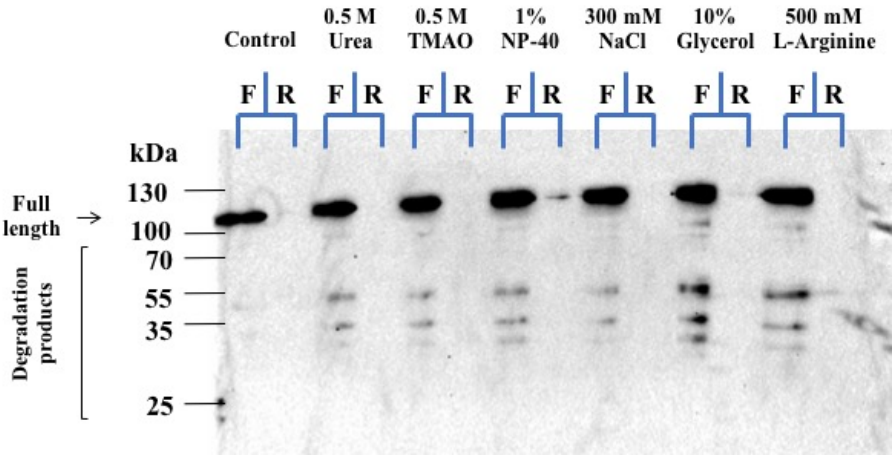

Figure 3-8. Initial solubility screening of *At*LARP6b. Western blot of solubility screening assay using an anti-His antibody. Chemical additives were added to the supernatant of pET28SUMO-*At*LARP6b transformed cells. Flowthrough (F) and Retentate (R) lanes are labeled for each chemical additive. 300 mM NaCl was chosen as the initial group of chemicals to assess based on the lack of signal present in the retentate lane and conservative nature of the additive to the standard buffer.

The next iteration of the screen evaluated several kosmotropes (Figure 3-9). The incubation and centrifugation protocol was repeated using ammonium sulfate, potassium chloride, magnesium sulfate, sodium chloride, and sodium sulfate as additives. Of these, ammonium sulfate yielded the most intense bands in the flow-through sample and minimal signal in the retentate lane.

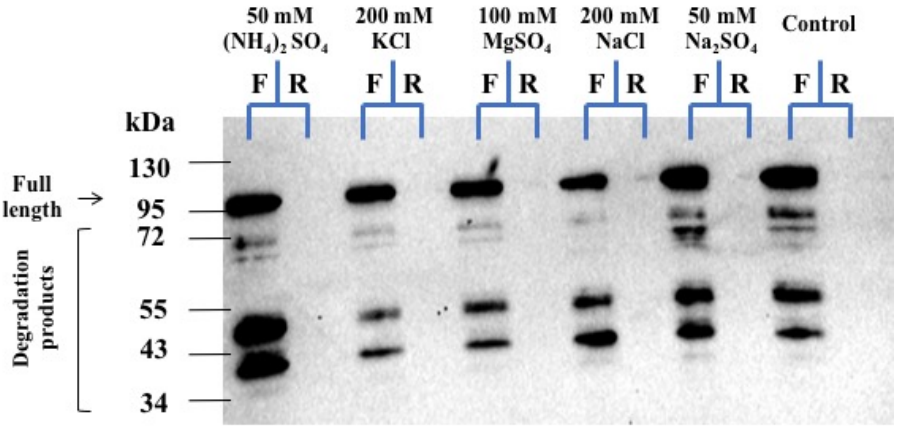

Figure 3-9. Kosmotrope screening of *At*LARP6b. Western blot of solubility screening assay using an anti-His antibody. Kosmotropic chemical additives as suggested by Bondos were used in the same solubility screening assay. Flowthrough (F) and Retentate (R) lanes are labeled for each chemical additive. 50 mM ammonium sulfate appeared to increase the intensity of degradation product bands when compared to the control.

Finally, the filter assay protocol was then used to identify the optimum concentration of ammonium sulfate for solubility improvement (Figure 3-10). Initial testing showed 25 mM ammonium sulfate as the most effective at improving solubility. However, a second assay was performed to observe the effects of concentrations closer to 25 mM ammonium sulfate. This subsequent solubility assay identified 10 mM ammonium sulfate as producing a much stronger signal than either 25 mM ammonium sulfate or the control. Thus, 10 mM ammonium sulfate was determined to be the most effective, and would be used in purification buffers for large-scale purification of the recombinant SUMO-*At*LARP6 protein, as described below.

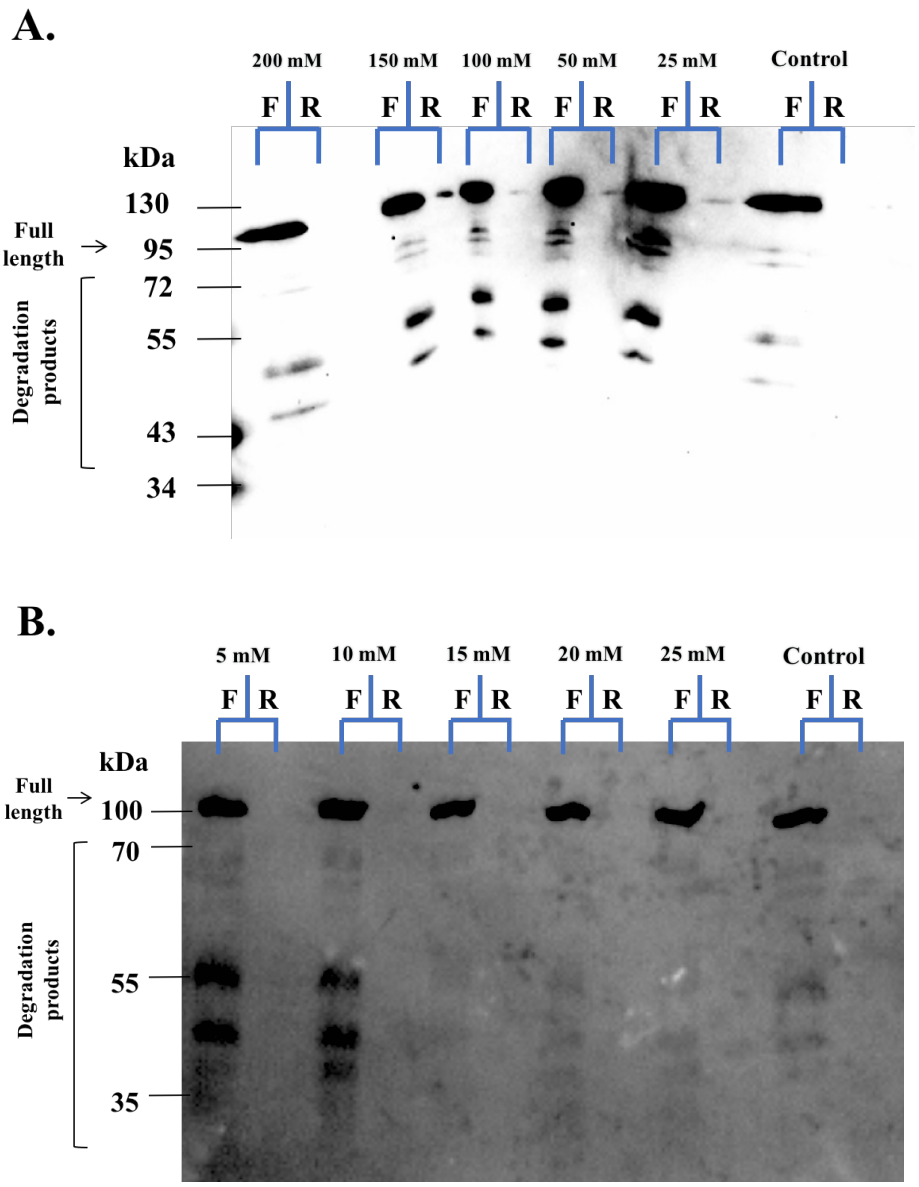

Figure 3-10. Optimizing ammonium sulfate concentration. Western blots of solubility screening assay using an anti-His antibody. (A) Differing concentrations of ammonium sulfate were tested in a subsequent round of solubility screening. Flowthrough (F) and retentate (R) lanes are labeled for each chemical additive. Lower concentrations of ammonium sulfate, 50 and 25 mM, appeared to improve the intensity of both the full-length and degradation product bands when compared to the control. (B) The solubility assay protocol was repeated using smaller concentrations of ammonium sulfate to observe an optimal concentration. 10 mM ammonium sulfate provided the most intense bands where the full-length protein and degradation products were observed, with notable improvement when compared to the control.

The success of the solubility screening assay for *At*LARP6b suggested that it could also be helpful for increasing yields of the other two LARP6 paralogs. While the SUMO-tagged *At*LARP6a and *At*LARP6c had been previously purified, attempts to reproduce those results encountered solubility and aggregation problems like those experienced when attempting to purify *At*LARP6b. Therefore, the same solubility screening assay was used to identify conditions that would improve protein yields for both *At*LARP6a and *At*LARP6c.

For SUMO-*At*LARP6c, the initial solubility screen was conducted with only kosmotropes (Figure 3-11), as these chemicals were most successful at enhancing the solubility of *At*LARP6b (Figure 3-9). *At*LARP6b and *At*LARP6c contain three of the same conserved domains (Figure 1-5), so chemical additives that have shown success in improving the solubility of *At*LARP6b would be thought to have a similar effect on *At*LARP6c. Sodium sulfate was found to produce the most intense full-length protein band with no detectable aggregated SUMO-*At*LARP6c protein in the retentate.

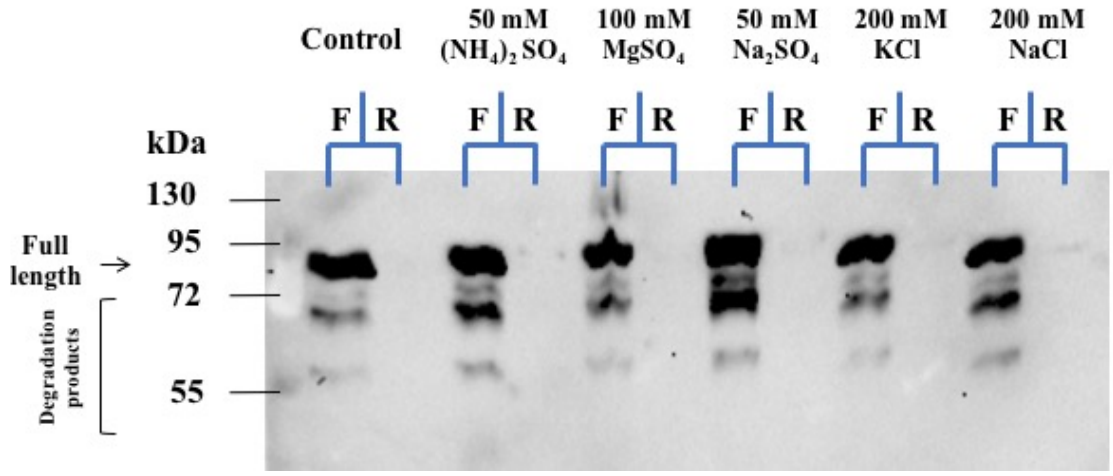

Figure 3-11. *At*LARP6c solubility screening assay. Western blot of solubility screening assay using an anti-His antibody. Solubility assay using kosmotropic chemical additives on *At*LARP6c showed an increase in full-length and degradation product band intensity when 50 mM sodium sulfate was added. Flowthrough (F) and Retentate (R) lanes are labeled for each chemical additive.

Similar to both SUMO-*At*LARP6b and SUMO-*At*LARP6c, kosmotropes also improved solubility of SUMO-*At*LARP6a (Figure 3-12). Also like SUMO-*At*LARP6c, sodium sulfate was found to produce the most intense full-length protein band with no detectable aggregated SUMO-*At*LARP6a protein in the retentate.

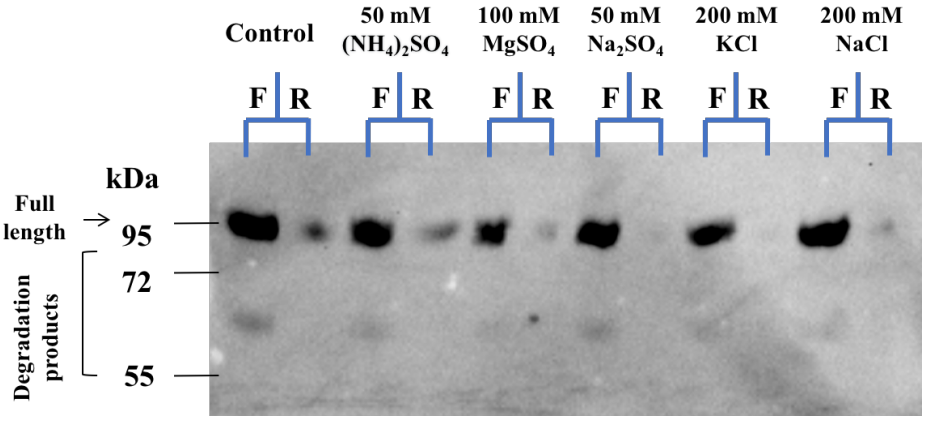

Figure 3-12. *At*LARP6a solubility screening assay. Western blot of solubility screening assay using an anti-His antibody. Solubility assay using kosmotropic chemical additives on *At*LARP6a showed a decrease in the intensity of the retentate band when compared to the control. Flowthrough (F) and Retentate (R) lanes are labeled for each chemical additive. 50 mM Na_2SO_4 appeared to improve solubility the most when compared to the control.

Protein Purification

His-tagged proteins contain a short chain of contiguous histidine residues able to bind divalent metal cations. Ni^{2+}-NTA beads are used to separate His-tagged proteins from others in the supernatant for further purification through a size-exclusion column. These proteins bind to the nickel beads on the column, and are eluted by competition with high concentrations of imidazole in the elution buffer. The eluted protein fractions are then concentrated and subjected to a size exclusion column. This procedure might separate pure SUMO-tagged protein from degradation products and imidazole present in the elution fractions. This initial affinity chromatography successfully isolated His_6-SUMO-*At*LARP6b (Figure 3-13), His_6-SUMO-*At*LARP6c (Figure 3-14), and His_6-SUMO-*At*LARP6a (Figure 3-14).

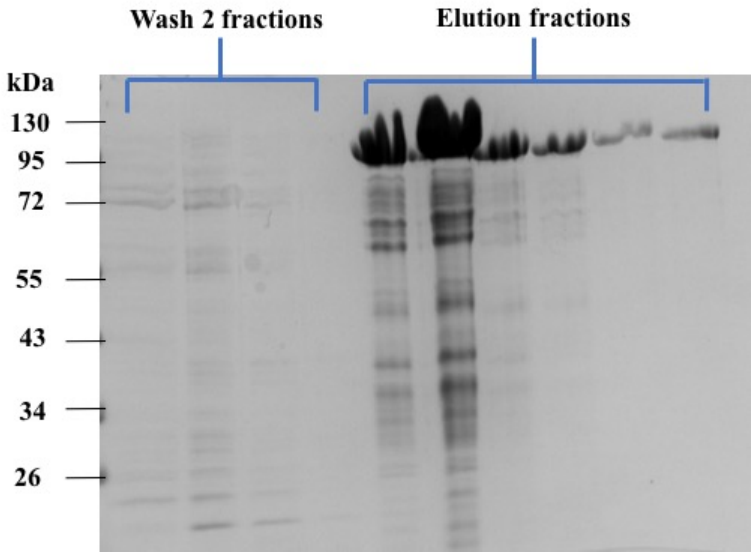

Figure 3-13. Ni-NTA elution of pET28SUMO-*At*LARP6b. Coomassie-blue stained SDS-PAGE gel of fractions collected from a Ni-NTA column during the purification of *At*LARP6b. The first four elution fractions showed the highest concentration of protein present to continue with SEC.

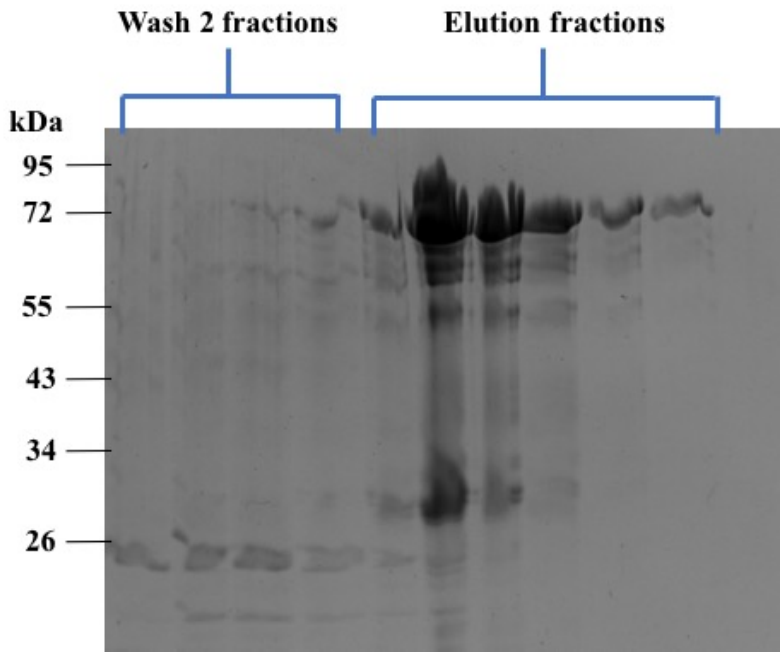

Figure 3-14. Ni-NTA elution of pET28SUMO-*At*LARP6c. Coomassie-blue stained SDS-PAGE gel of fractions collected from a Ni-NTA column during the purification of *At*LARP6c. Elution fractions 2,3, and 4 contained the greatest amount of protein that was carried forward to the SEC.

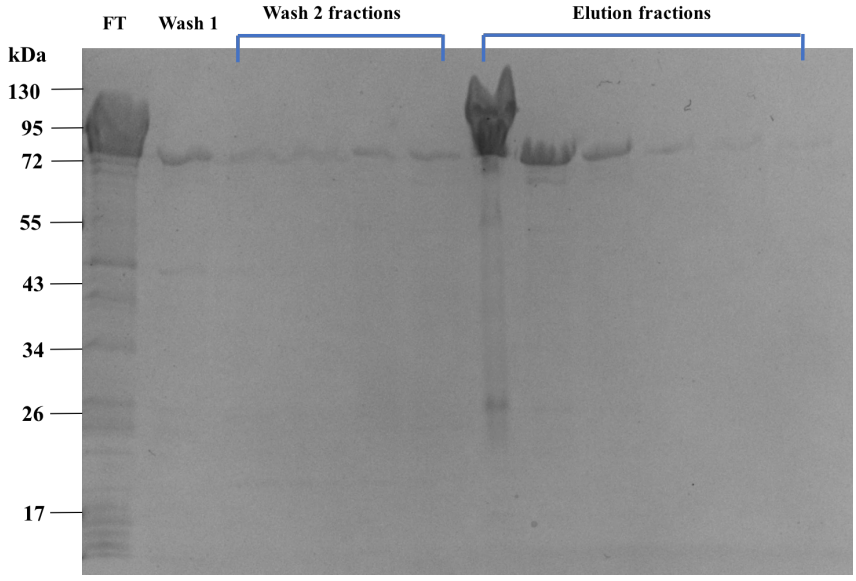

Figure 3-15. Ni-NTA elution of pET28SUMO-*At*LARP6a. Coomassie-blue stained SDS-PAGE gel of fractions collected from a Ni-NTA column during the purification of *At*LARP6a. Elution fractions 1, 2, and 3 showed the highest amount of protein present to continue with SEC.

The initial size exclusion chromatography of *At*LARP6b using the buffers containing ammonium sulfate proved successful, although the amount of recovered protein was less than desired. The elution volume was lower than expected, at a molecular weight corresponding to a dimer. When fractions were checked under reducing and denaturing conditions, however, molecular weight was at the expected value of around 72 kDa. The purification of *At*LARP6b was repeated using the original protocol, except using pellets that were induced for 16 hours at 16˚C. The initial concentration of protein present in Ni-NTA elution fractions was noticeably stronger when compared to previous purification attempts, demonstrating that the solubility screening improved purification conditions (Figure 3-16).

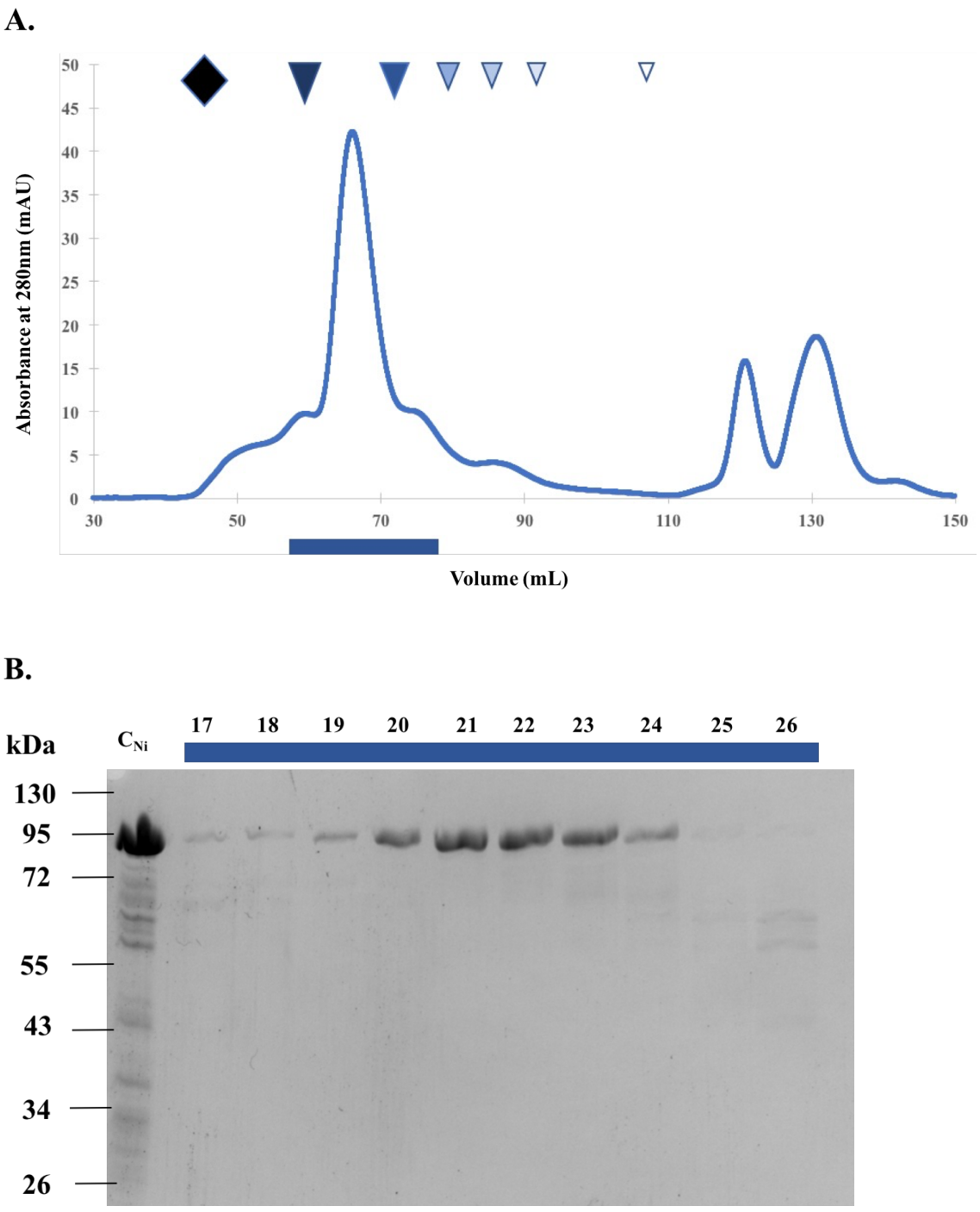

Figure 3-16. Purification of SUMO-*At*LARP6b. (A) Elution profile of *At*LARP6b-SUMO after size exclusion column over the S200 column. The large peak present at 69 mL is the elution of *At*6b-SUMO. Concentration of protein obtained, approximated by mAU at 280 nm, was much higher than previous attempts. Standard molecular weight markers are depicted above the UV chromatography curve, and fractions verified by SDS-PAGE are marked by a blue bar. (B) Coomassie-blue stained SDS-PAGE gel of the marked fractions. C_{Ni}, the control, was the retentate from the concentrated Ni-NTA elution, before being injected into the FPLC. Numbers correspond to the eluted fraction number from the FPLC.

Digestion of fusion protein by ULP1

Small Ubiquitin-like modifier proteins are used to tag proteins for purification. These small proteins (molecular weight around 12 kDa) are covalently attached to the N-terminus of target proteins, and have been shown to improve solubility of their target proteins (although the mechanism is still not well understood). Optimizing the ULP-1 cleavage reaction should remove the SUMO tag from proteins of interest, with minimal degradation of the target protein.

ULP1 digestion time trials were conducted to determine optimal incubation time of protein with ULP1 enzyme for cleavage of the SUMO tag. Ideally, target protein to ULP1 enzyme concentration ratio would be 1000:1, but it was difficult to measure protein concentration using either intrinsic fluorescence or a BCA assay. A set volume of protein to ULP1 ratio was used instead, 100 µL of protein to 5 µL of 24.7 µM ULP1. The SUMO tag was successfully cleaved after 2 hours of incubation at 16ºC, indicated by the reduction in band size compared to the control and appearance of a second band between 26-35 kDa (Figure 3-17). Samples incubated for longer periods of time appeared to degrade (Figure 3-17).

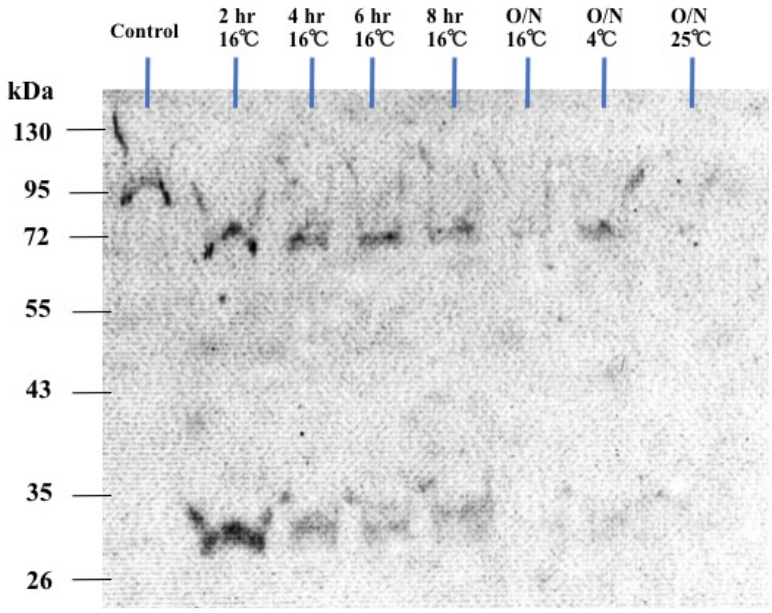

Figure 3-17. ULP1 test digest of *At*LARP6b. Coomassie-blue stained SDS-PAGE gel of SUMO-*At*LARP6b treated with ULP1. Cleavage of SUMO from *At*LARP6b appeared to be successful after two hours of incubation with ULP1 at 16˚C.

In the full-scale purification, the SEC-purified fractions of His_6-SUMO-*At*LARP6b were incubated with ULP1 enzyme, and transferred to a Vivaspin Turbo conical for concentration and then purification over the size exclusion column. The SEC elution profile yielded promising results, notably the two sharp peaks observed around 73 mL and 93 mL, representing full-length *At*LARP6b and ULP1-SUMO, respectively (Figure 3-18, panel A). Fractions corresponding to those peaks were run on an SDS-PAGE gel, confirming the presence of both full-length *At*LARP6b and ULP1-SUMO (Figure 3-18, panel B). The absorbance corresponding to the full-length *At*LARP6b peak was much higher than previous attempts, indicating the concentration of recovered protein would be sufficient for further experimentation. As *At*LARP6b was the greatest challenge in purification of all three *A. thaliana* LARP6 paralogues, attention could now

be shifted to purifying and ULP1 digestion of the remaining *At*LARP6 proteins, starting with *At*LARP6c.

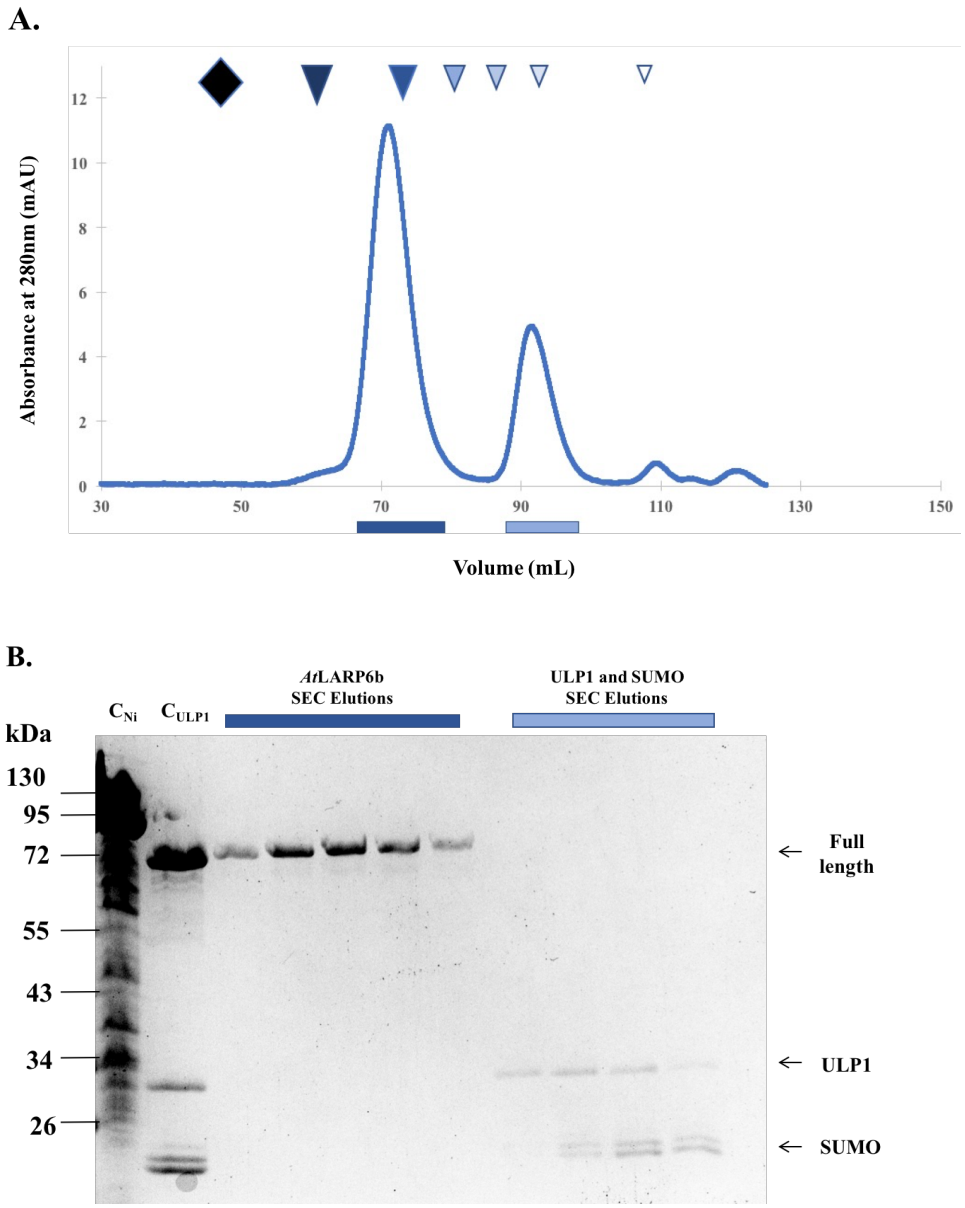

Figure 3-18. Fully purified full-length *At*LARP6b. (A) Elution profile of *At*LARP6b after SUMO cleavage by ULP-1 after size exclusion column over the S200 column. The large peak present at 70 mL is the elution of full-length *At*6b, the peak to the right is the elution of the ULP1 enzyme with SUMO bound. Standard molecular weight markers are depicted above the UV chromatography curve, and fractions verified by SDS-PAGE are marked by the blue bars. (B) Coomassie-blue stained SDS-PAGE gel of *At*LARP6b after ULP1 digest and elution from Sephadex 200 column. Samples from the original retentate before S200 elution and retentate after ULP1 digest were used as controls, C_{Ni} and C_{ULP1} respectively. Elutions of *At*LARP6b and ULP1/SUMO were selected based on peaks shown on the SEC chromatogram.

Nickel-affinity purification of *At*LARP6c was conducted using essentially the same protocol used for *At*LARP6b, but with buffers containing 50 mM sodium sulfate instead of 10 mM ammonium sulfate (Figure 3-14). Centrifugal concentration of the nickel elution fractions that contained the most intense full-length protein bands produced a highly viscous solution, of which 1.5 mL was injected onto the S200 column. The chromatogram of this initial SEC run showed a high concentration of *At*LARP6c recovered, as well as a high concentration of aggregated protein that eluted in the void (Figure 3-19, panel A). The initial viscosity of the injected solution may have reflected concentration-dependent protein aggregation.

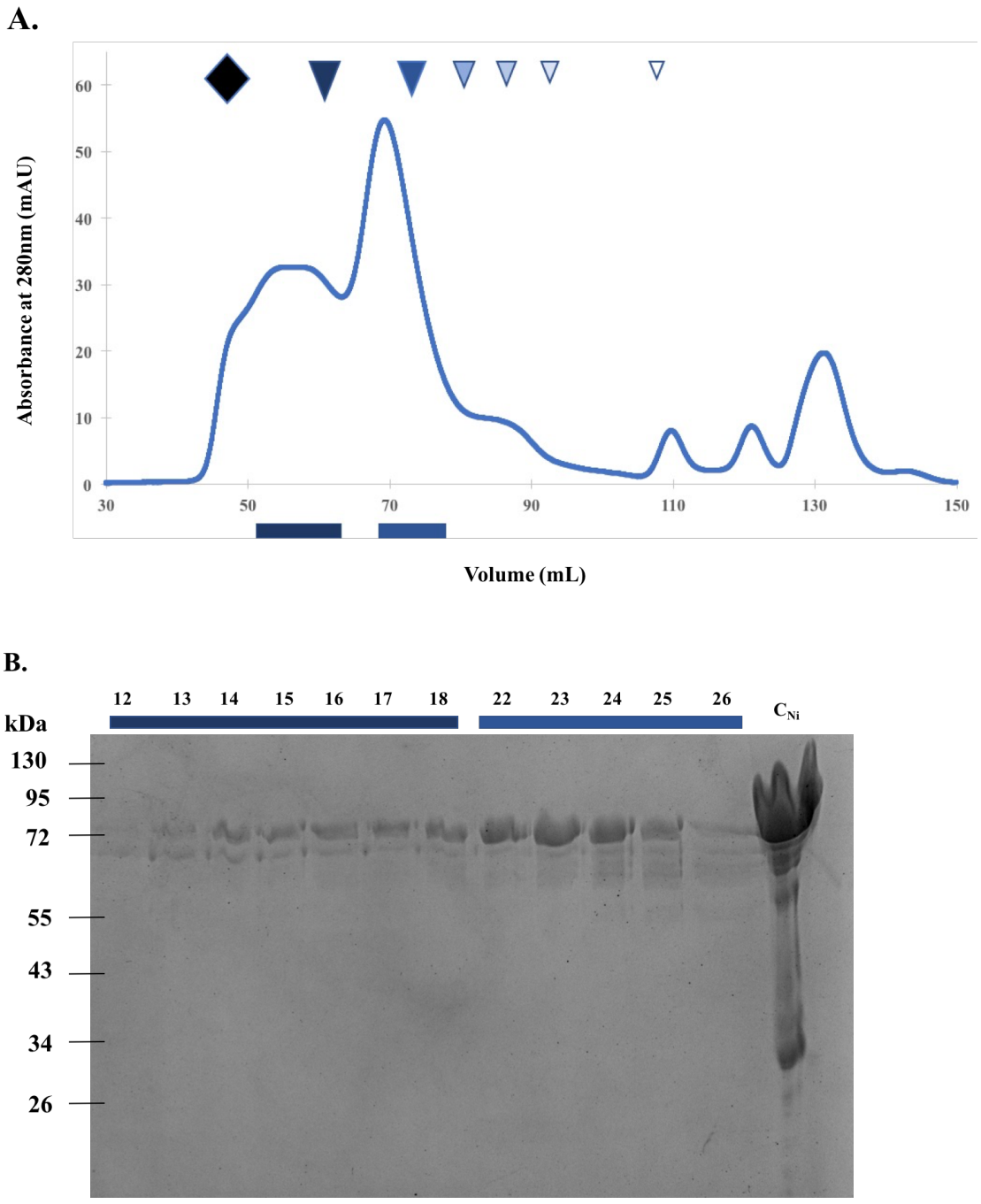

Figure 3-19. Purification of SUMO-*At*LARP6c. (A) Elution profile of *At*LARP6c-SUMO after size exclusion column over the S200 column. The large peak present at 72 mL is the elution of *At*LARP6c-SUMO. Concentration of protein obtained, approximated by mAU, was much higher than previous attempts. Standard molecular weight markers are depicted above the UV chromatography curve, and fractions verified by SDS-PAGE are marked by the blue bars. (B) Coomassie-blue stained SDS-PAGE gel of the marked fractions. C_{Ni}, the control, was the retentate from the concentrated Ni-NTA elution, before being injected into the FPLC. Numbers correspond to the eluted fraction number from the FPLC.

Soluble His_6-SUMO-*AtLARP6*c, which was considered to include fractions 22-26 (Figure 3-19, panel B), was pooled and incubated with ULP1 to cleave the SUMO tag from *At*LARP6c. The elution profile obtained from the SEC run after ULP1 digestion yielded similar results to the profile of *At*LARP6b. Sharp peaks were observed for both full-length *At*LARP6c and ULP1-SUMO, with the *At*LARP6c peak detected slightly to the right of the *At*LARP6b peak (Figure 3-20, panel A and Figure 3-18, panel A, respectively). This was expected, as the molecular weight of *At*LARP6c is less than that of *At*LARP6b, around 49 kDa to around 60 kDa of *At*LARP6b. The peak representing the ULP1-SUMO elution is identical to that previously seen in the *At*LARP6b ULP1 digest SEC run (Figure 3-20, panel A and Figure 3-18, panel A, respectively). SEC fractions corresponding to the peaks were analyzed via SDS-PAGE, which confirmed the presence of both full-length *At*LARP6c and ULP1-SUMO (Figure 3-20, panel B).

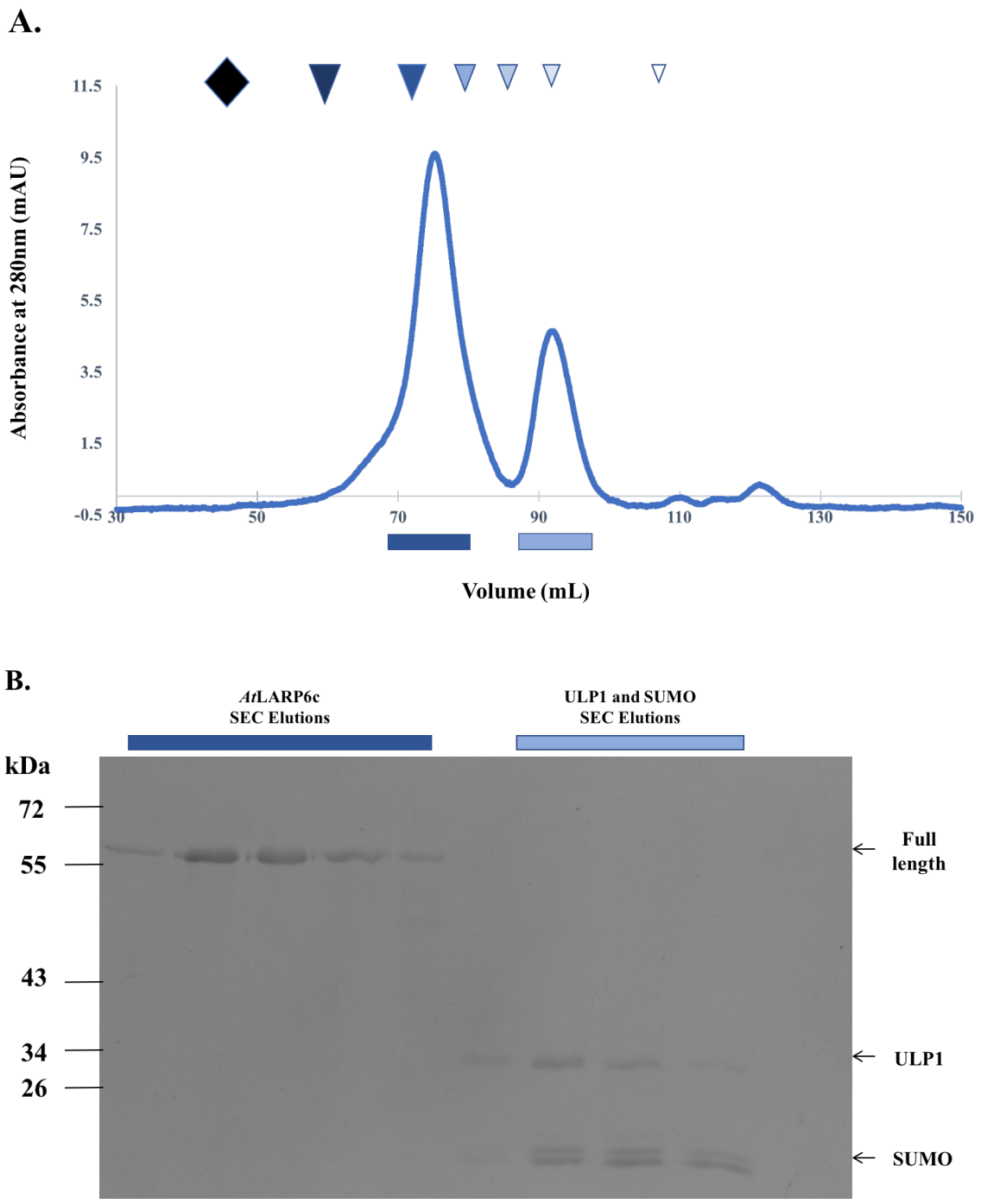

Figure 3-20. Fully purified full-length *At*LARP6c. (A) Elution profile of *At*LARP6c after SUMO cleavage by ULP-1 after size exclusion column over the S200 column. The large peak present at 75 mL is the elution of full-length *At*LARP6c, the peak to the right is the elution of the ULP1 enzyme with SUMO bound. Standard molecular weight markers are depicted above the UV chromatography curve, and fractions verified by SDS-PAGE are marked by the blue bars. (B) Coomassie-blue stained SDS-PAGE gel of the marked fractions.

With fully-purified full-length *At*LARP6b and *At*LARP6c obtained, efforts were shifted to express and purify *At*LARP6a. Conducting expression trials and the solubility assay with His_6-SUMO-*At*LARP6a led to a successful purification of *At*LARP6a. The primary changes to the protocol were to express at 16˚C for 23 hours and the addition of 50 mM sodium sulfate to Ni-NTA and SEC purification buffers (Figure 3-21).

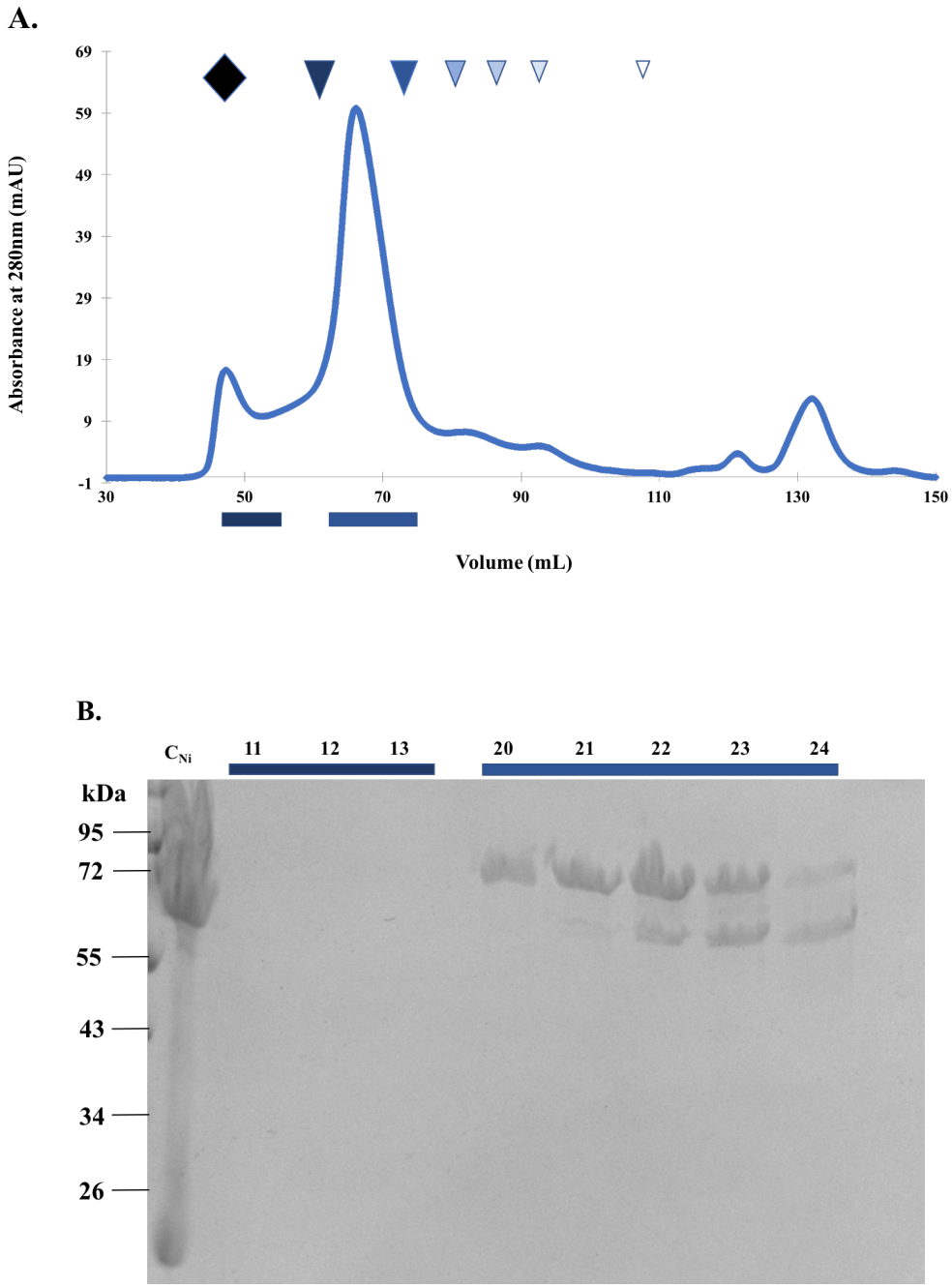

Figure 3-21. Purification of SUMO-*At*LARP6a. (A)Elution profile of *At*LARP6a-SUMO after size exclusion column over the S200 column. The large peak present at 69 mL is the elution of *At*LARP6a-SUMO. The protein concentration, as measured by mAU (280 nm), was much higher than previous attempts. Standard molecular weight markers are depicted above the UV chromatography curve, and fractions verified by SDS-PAGE are marked by the blue bars. (B) Coomassie-blue stained SDS-PAGE gel of the marked fractions. C, the control, was the retentate from the concentrated Ni-NTA elution, before being injected into the FPLC. Numbers correspond to the eluted fraction number from the FPLC.

Two distinct peaks were observed on the SEC elution profile for the purification of His_6-SUMO-*At*LARP6a (Figure 3-21, panel A). The first peak is consistent with the void volume of the column, possibly indicating aggregated protein, but the second peak at around 69 mL could contain His_6-SUMO-*At*LARP6a (Figure 3-21, panel A). Fractions from the initial purification of *At*LARP6a were visualized by SDS-PAGE, and those that appeared to contain His_6-SUMO-*At*LARP6a (Figure 3-21, panel B) were pooled and digested with ULP1 for two hours. Pooled fractions were run through the SEC after ULP1 digest, as done for the other *At*LARP6 proteins. The elution profile yielded results similar to the previous ULP1 digested protein purifications, with two peaks indicating the presence of full-length, cleaved *At*LARP6a and ULP1-SUMO (Figure 3-22, panel A). Fractions corresponding to the peak observed on the elution profile were analyzed by SDS-PAGE, confirming the presence of full-length *At*LARP6a and ULP1-SUMO, producing similar results as the two other *At*LARP6 proteins (Figure 3-22, panel B).

All purified recombinant proteins were aliquoted, snap-frozen in liquid nitrogen, and stored at -70 ºC for further characterization.

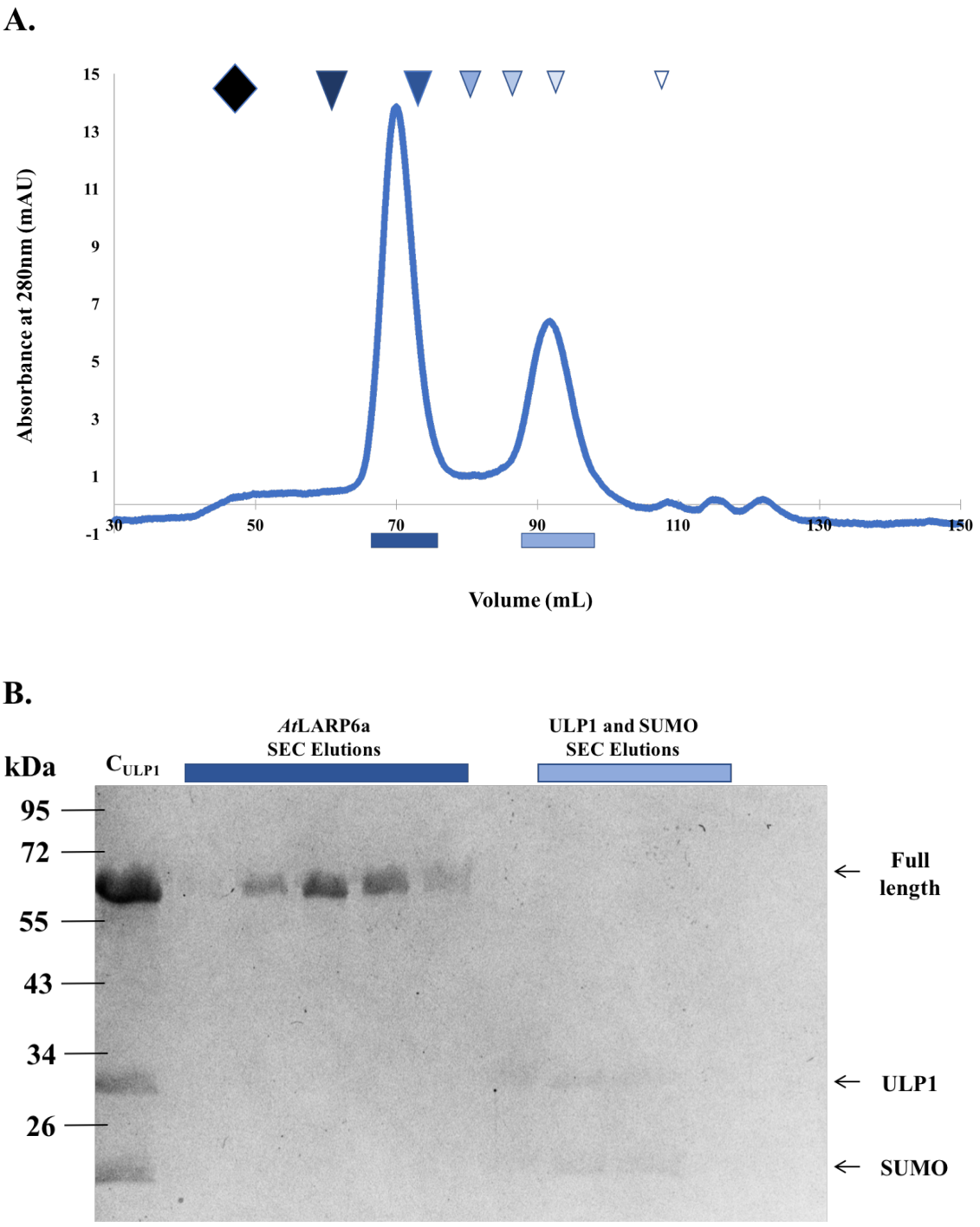

Figure 3-22. Fully purified full-length *At*LARP6a. (A) Elution profile of *At*LARP6a after SUMO cleavage by ULP-1 after size exclusion column over the S200 column. The large peak present at 70 mL is the elution of full-length *At*LARP6a, the peak to the right is the elution of the ULP1 enzyme with SUMO bound. Standard molecular weight markers are depicted above the UV chromatography curve, and fractions verified by SDS-PAGE are marked by the blue bars. (B) SDS-PAGE gel of the marked fractions. C_{ULP1}, the control, was the retentate from the concentrated SEC fractions, before being injected into the FPLC.

Limited Proteolysis

Trypsin is a proteolytic enzyme, or protease, found in eukaryotes that hydrolyzes the peptide backbone C-terminal to the positively-charged residues arginine and lysine. In general, trypsin will only degrade a stably-folded protein slightly, due to the inaccessible arginine and lysine residues within the protein folds. Data from limited proteolysis can indicate the presence or absence of stable folded domains in proteins. Limited proteolysis by trypsin was conducted with *At*LARP6 protein on ice, with 20 µL samples collected at set time points. After 4 hours, samples collected over the time course were subjected to SDS-PAGE and subsequently silver stained for detailed analysis. Each *At*LARP6 protein appeared to degrade very differently over the course of the limited proteolysis. Unlike *At*LARP6b and *At*LARP6c, *At*LARP6a appeared to aggregate upon thawing. White flocculent material appeared after thawing *At*LARP6a protein samples from – 70˚C storage, but were able to be readily dissolved before continuing with the limited proteolysis experiment. This aggregation could be responsible for the multiple intense bands observed within the first ten minutes of trypsin exposure (Figure 3-23, panel A). *At*LARP6b was quickly degraded by trypsin, the intensity of the full-length band noticeably faded after the first twenty minutes (Figure 3-23, panel B). *At*LARP6c, however, remained relatively stable throughout the course of the experiment, with the full-length band losing signal only slightly, until the three-hour mark (Figure 3-23, panel C). Overall, all three proteins showed low stability, as each limited proteolysis was conducted on ice, and no bands appeared to survive the digest. This may also indicate the lack of stable domains within the *At*LARP6 proteins. Additional experiments probing structural data are needed.

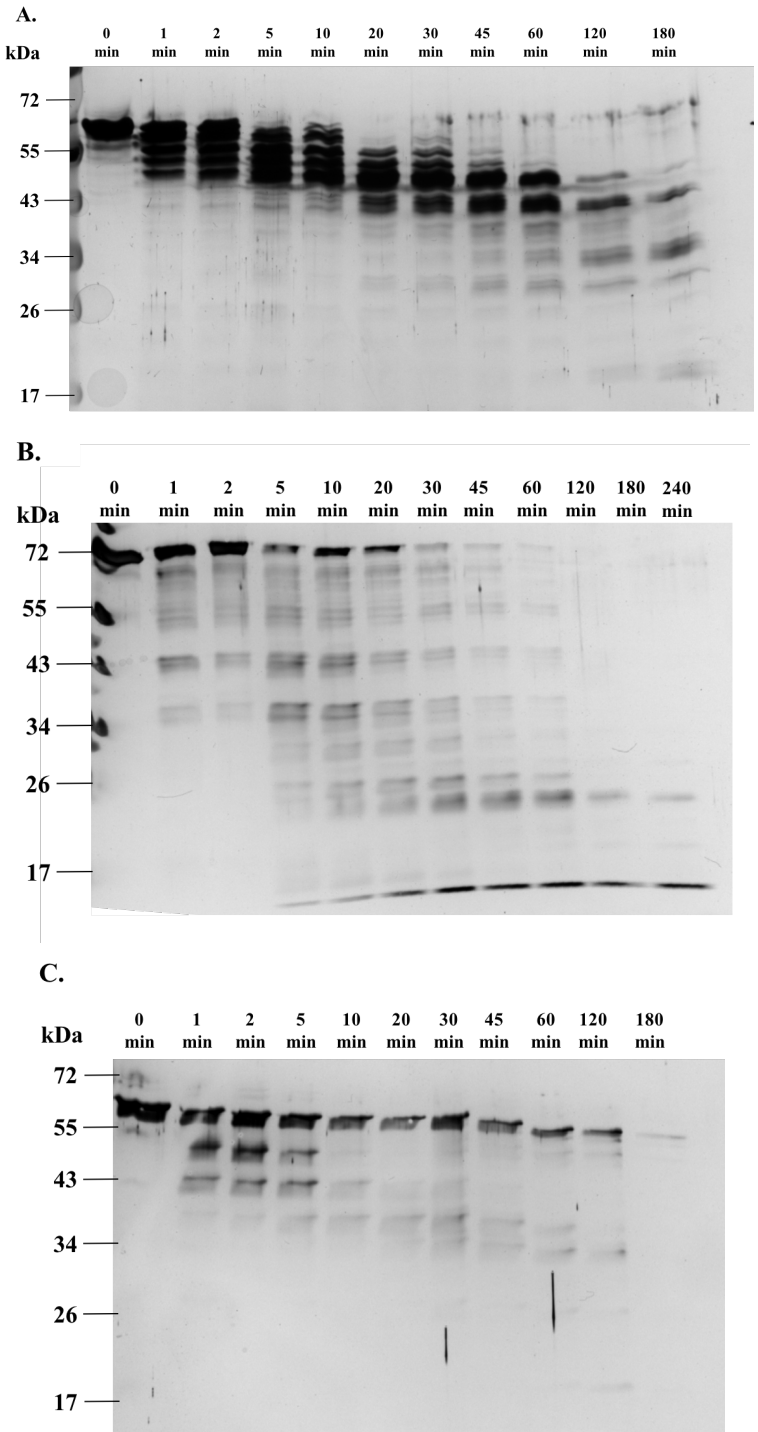

Figure 3-23. Limited proteolysis of *At*LARP6 proteins. Silver-stained SDS-PAGE gels: (A) Limited proteolysis by trypsin of *At*LARP6a over time with samples on ice. Time measured by the amount of time the sample was exposed to trypsin before boiling. (B) Limited proteolysis by trypsin of *At*LARP6b over time with samples on ice. Lane labels denote amount of time the sample was exposed to trypsin before boiling to inactivate the enzyme. (C) Limited proteolysis by trypsin of *At*LARP6c over time with samples on ice. Lane labels denote amount of time the sample was exposed to trypsin before boiling to inactivate the enzyme.

Discussion

Previous work by Merret and colleagues provided valuable insight on the *At*LARP6 proteins, but many questions remain concerning the characterization of not only the RNA-binding domains, but also the full-length *At*LARP6 proteins. Our experiments have built on this previous work, with a focus on obtaining and characterizing the full-length *At*LARP6 proteins.

First, recombinant expression conditions for each protein were optimized to increase protein yield, which would translate to a higher protein yield during purification. Each of the expressions were carried out at a lower temperature than previously attempted, which could have improved yield of full-length protein by decreasing the effect of thermal denaturation on protein folding. While the higher temperatures used for earlier expression trials would not completely denature the protein, heat could still cause slight unfolding of hydrophobic regions, making them more vulnerable to aggregation or degradation. Future experiments could observe if lower expression temperatures would further improve protein yield, or further confirm 16˚C to be an optimal temperature. Daniel Horn's initial successes with higher expression temperatures may also suggest that comparable amounts of stable protein could be obtained at slightly higher expression temperatures, for shorter periods of time.

Successful expression of proteins is important for further purification and analyses on pure protein samples. Merret *et al*. was able to purify the RNA-binding

domains of *At*LARP6a and *At*LARP6c, but was unable to retrieve a stable *At*LARP6b binding domain. We also initially encountered issues obtaining *At*LARP6b, observing SEC elution profiles consistent with protein aggregation. We proceeded to conduct solubility screening assays as described by Churion and Bondos to remedy the aggregation issue. Solubility was improved for all three proteins by conducting filter-based solubility screens with various chemicals. Ammonium sulfate for *At*LARP6b, and sodium sulfate for *At*LARP6a and *At*LARP6c noticeably increased solubility when compared to the control. *At*LARP6b was the most extensively experimented upon, with four rounds of screening. Each round of screening narrowed down the potentially helpful additives, of which 10 mM ammonium sulfate would prove most effective. It is important to note that each solubility screen did identify other potential additives that were not pursued, and may be useful in future studies. L-arginine also showed promise on the initial solubility screening, but was not used in the subsequent screens due to the limited availability of reagents necessary to carry out the amino acid screen. L-arginine has been successfully used before to prevent protein aggregation, due to its structural-stabilizing ability (Tischer *et al*., 2010).

Abbreviated solubility screenings of *At*LARP6a and *At*LARP6c were conducted by only testing kosmotropic additives. Since 50 mM sodium sulfate appeared to improve solubility in *At*LARP6c, the buffer additive was directly tested at that concentration in SEC and Ni-NTA buffers. *At*LARP6a was also only tested with the kosmotropic group of additives, where 50 mM sodium sulfate was again found to be effective. Future screening could reveal other chemical additives that also increase solubility of these proteins. The

selected buffer additives aided with the purification of each protein by size exclusion chromatography.

Size exclusion chromatography produced elution profiles of each protein showed higher concentrations of protein recovered than previously achieved, especially with *At*LARP6b and *At*LARP6c. Initial attempts at purifying *At*LARP6b were unsuccessful, as indicated by the lack of peaks present on the elution profile of the SEC run. After the addition of 10 mM ammonium sulfate, a sharp peak at 69 mL was observed, and fractions corresponding to that peak contained *At*LARP6b. While the observed elution volume did not match the expected volume for a monomeric protein, *At*LARP6b could have eluted as a dimer.

Previous attempts at purifying *At*LARP6c were unsuccessful due to the lack of concentrated protein obtained. While a peak was observed on the elution profile of *At*LARP6c, it was of low absorbance, and fractions visualized on a gel confirmed a low protein yield. The same elution profile also contained a notable broad peak at the void volume of the column, which could be due to *At*LARP6c aggregating. Like *At*LARP6b, AtLARP6c purification was repeated using the solubility-enhancing buffer additive, 50 mM sodium sulfate for *At*LARP6c, which yielded much better results than previously obtained. Concentration of protein was noticeably high during the elution of protein from the Ni-NTA beads, where very viscous drops were coming off the column. Concentrating the amount of protein recovered from Ni-NTA beads was also challenging, as viscosity increased even further. Due to the high concentration of protein present injected into the

SEC, a large broad peak was observed at the void volume, presumably from concentration-dependent aggregation. Despite the aggregation, a large sharp peak was also observed for purified *At*LARP6c, which was confirmed by visualization on a SDS-PAGE gel.

*At*LARP6a proved challenging to purify at first, with multiple attempts failing to yield workable concentrations of protein. Even when using sodium sulfate to increase solubility, as indicated by the solubility screening, protein yield was still low. However, elution of *At*LARP6a from the Ni-NTA beads appeared to be low when visualized on a SDS-PAGE gel, suggesting a lack of protein being expressed in the first place. Small scale expression trials conducted with *At*LARP6a showed that overnight expression of *At*LARP6a at 16˚C dramatically increased the amount of protein the bacteria were able to produce, which led to a more successful full scale expression. Elution from nickel beads using supernatant from the full-scale expression conducted overnight was noticeably stronger, and resulting protein yield from the SEC was also much higher. The high yield of *At*LARP6a-SUMO allowed for follow-up ULP1 digest and recovery of SUMO-cleaved *At*LARP6a from a subsequent SEC run.

Purified full-length proteins were cleaved from their SUMO tag by ULP1 digest, and obtained after another size exclusion column run. ULP1 digests of each tagged-protein yielded ample cleavage of the SUMO tag when visualized on a gel. One band was observed for fractions containing the cleaved protein, with no other visible traces of the tag present. Visualization of the full-length, SUMO-cleaved *A. thaliana* LARP6 proteins

on a SDS-PAGE gel confirmed the presence of the target proteins. Each protein eluted from the size exclusion column at a higher molecular weight than what would be expected for a monomeric protein, but matched the expected molecular weight on SDS-PAGE gels. Therefore, we have now produced biochemical quantities of pure *Arabidopsis thaliana* LARP6a, LARP6b, and LARP6c proteins.

With these proteins in hand, novel characterization of full-length *At*LARP6 proteins is now possible. Limited proteolysis provided insight on protein stability when digested with trypsin. *At*LARP6a appeared to aggregate upon thawing, and full-length protein remained intact for only about ten minutes (Figure 3-23, panel A). *At*LARP6b was quickly degraded, with a noticeable decrease in band intensity after twenty minutes of trypsin exposure (Fig 3-23, panel B). In contrast, the full-length *At*LARP6c band degraded at a much slower rate (Fig 3-23, panel C). In all cases, no degradation products remained stable throughout the digest.

The relationship between molecular weight and SEC elution volume is complicated by the presence of intrinsic disorder in proteins, where the hydrodynamic radius is skewed (English, L. 2016). SLIDER was used to predict disordered regions within the *A. thaliana* LARP6 proteins to reference when evaluating SEC chromatograms and limited proteolysis data (Table 2-4). All three *A. thaliana* LARP6 proteins are predicted to have at one long disordered segment, at least 30 amino acids in length (Peng *et al.,* 2014).

While the SLIDER data can only provide predictions on disordered segments, this data may help to explain the lack of stability observed with all three *At*LARP6 proteins from the limited proteolysis experiments. Further experimentation is needed to account for the lack of stable degradation products observed in each limited proteolysis. The predicted disorder may also explain the higher molecular weight observed from the SEC elution profiles of each protein. Now that the full-length proteins have been purified, further data explaining these phenomena could be obtained in the near future.

Future Directions

The SUMO/ULP1 tag and cleavage system was sufficient for our initial experiments, but it should be noted that our lab also has stocks of expression vectors for the *A. thaliana* LARP6 proteins as fusions to a chitin-binding domain (using the IMPACT protein expression system and pTXB1 vector) if an alternative system would be required in the future. This chitin-binding domain system could be useful in collaborating with other labs that may prefer utilizing that system. As Merret *et al*. tested RNA-binding activity of the La module and RRM for *At*LARP6a and *At*LARP6c, we could also test RNA-binding activity of the full-length *At*LARP6 proteins through electromobility shift assays (EMSA) as a comparison. Additional biophysical information about the stability of each LARP6 protein should be obtained using either thermal or chemical denaturation as well as possible mass spectrometry to identify the boundaries of proteolytically-protected fragments.

The endogenous poly-His tracts present in *At*LARP6b also raise questions as to why these tracts are only found in LARP6b, and what potential cellular function they may have. Probing the functions of poly-His tracts both *in vitro* and *in vivo* may elucidate the functions of plant LARP6, which are still largely unknown. Creating an *A. thaliana* expression construct would serve as one of the first steps to further research on both the role of endogenous poly-His tracts and the LARP6 family in plants.

IV. A LARP6-GFP PLATFORM FOR RECOMBINANT EXPRESSION IN PLANTS

Introduction

Arabidopsis thaliana LARP6 proteins have been shown to localize to the nucleolus and to sub-cytoplasmic foci under hypoxic stress when transiently expressed in onion epidermal cells (Merret *et al*., 2013). *At*LARP6a is mostly present in the nucleolus, and does not aggregate to sub-cytoplasmic foci, most likely stress granules, under stress. *At*LARP6b and *At*LARP6c are also present in the nucleolus, but contrary to *At*LARP6a, do localize to sub-cytoplasmic foci under hypoxic stress. The difference in localization under stress may be attributed to the presence of the PAM2 motif in *At*LARP6b and *At*LARP6c. However, the poly-His tracts in *At*LARP6b may also be contributing to localization mechanisms as these motifs have been previously shown to guide localization (Salichs *et al*., 2009).

To test the significance of the poly-His tract in *At*LARP6b, a GFP-*At*LARP6b fusion product was created by cloning into an *A. thaliana* expression vector. PCR cloning of gene products, followed by restriction enzyme digestion and ligation of GFP and AtLARP6b products was used to create the fusion product to be inserted in the vector. This vector can then be transfected into *A. thaliana* cells, and the fusion product can be visualized under normal and hypoxic stress conditions by confocal microscopy *in vivo.* Site-directed mutagenesis can then be used to selectively delete the poly-His tracts present in *At*LARP6b, and the same transfection/visualization of localization experiments

repeated to observe any difference in localization in the presence and absence of the poly-His tracts.

Results

PCR Amplification

The two genes that were to be inserts, *larp6b* and *egfp,* were synthesized by PCR synthetic DNA oligonucleotide primers, which contained complementary restriction sites (Table 1). The template for *larp6b* amplification was *Arabidopsis* cDNA, a kind gift of Dr. Hong-gu Kang (Texas State University), and PCR carried out with primers KAL056 and KAL057. The template for *egfp* amplification was a yeast expression plasmid that encoded a GFP-tagged version of the human epithelial sodium channel (ENaC), a kind gift of Dr. Rachell E. Booth (Texas State University), and PCR was carried out with primers KAL058 and KAL059. PCR products were visualized on an agarose-TAE gel. Amplified gene products were observed at their expected molecular weights, *Atlarp6b is* 1635 bp, and *egfp* at 717 bp (Figure 4-1).

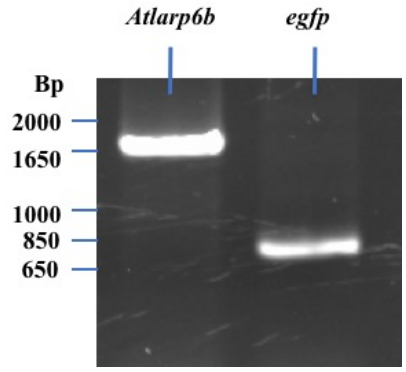

Figure 4-1. Amplification of *AtLarp6b* and *egfp.* Agarose gel of PCR-amplified *AtLARP6b* and *egfp*. Sharp bands of amplified gene products were observed at their expected molecular weights (*Atlarp6b* at 1635 bp, and *egfp* at 717 bp).

The PCR products were then digested with the restriction enzyme NcoI, to create the compatible ends on each product for ligation. The digestion products were gel-purified, and then ligated together. A second round of PCR was carried out using primers KAL056 and KAL059, to amplify only the ligated insert, *At*LARP6b-eGFP. The production of the "full" insert was visualized on an agarose gel. This product would be gel-purified, digested with CutSmart, and ligated to the pER-HA plasmid.

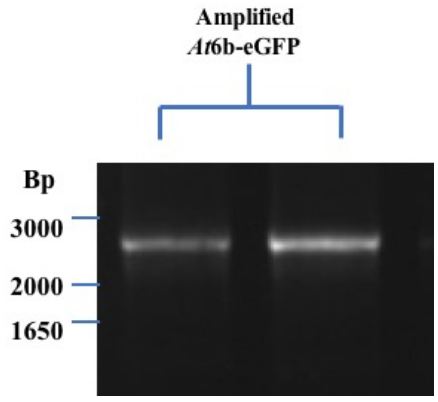

Figure 4-2. *AtLARP6*b-eGFP fusion product. Ligation of restriction-site digested *AtLARP6*b and eGFP was amplified with the outer primers to create the amplified product, shown in the expected position on an agarose gel.

The completed plasmid was subjected to a double digest using the same restriction enzymes initially used to cut the ends of the complete insert. Transformation of ligated insert + plasmid into bacterial cells appeared to be successful, as the product appeared to be present when visualized on an agarose gel at the expected molecular weight of 2,352 bp in lane 3 (Figure 4-3).

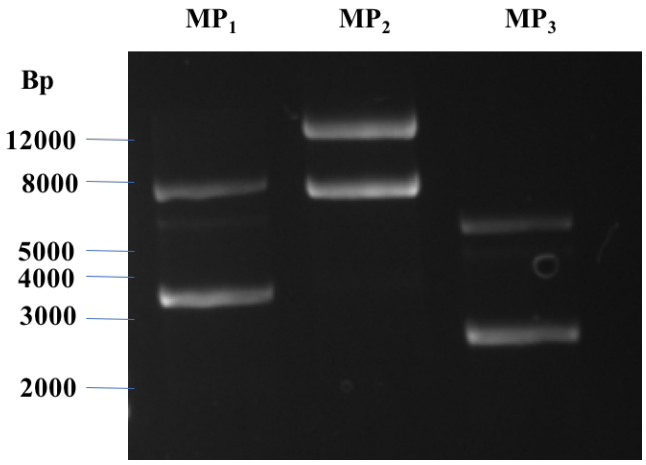

Figure 4-3. Mini-preps of DH5α transformed with pER-HA-*At*LARP6b-eGFP. Mini-preps (MP) of transformed DH5a cells were double digested with XhoI and SnaBI to test for the presence of the *AtLARP6*b-eGFP fusion products, visualized on an agarose gel. Lane 3 (MP_3) was thought to contain the expected product.

Overlap Extension PCR cloning

The overlap extension PCR method was used as an alternative to the previously attempted restriction site cloning. Cloning through PCR overlap involves the design of large primers that not only amplify one target gene, but also amplify a piece of the template and adjoining insert sequence. These overlapping products are further amplified by large primers that synthesize the entire insert through PCR. The insert also has overlapping regions with the template, and another primer is used to directly amplify the insert and the template, creating a recombinant plasmid. Primers were designed for these stages of PCR amplification, and used with *larp6* and *egfp* template plasmids to create products. This procedure appeared to be successful for the first set of amplifications, but subsequent steps yielded no products.

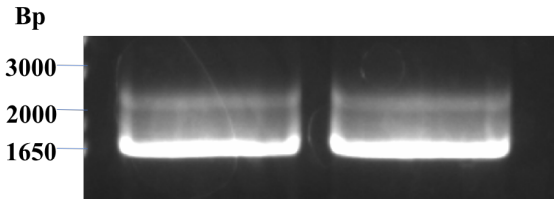

Figure 4-4. Final phase of overlap PCR cloning. PCR overlap products of the 3rd phase of overlap cloning. Fusion product should have been distinctly present at around 2400 bp, but was instead observed as a smear throughout the 1700-2500 bp range.

Discussion

Poly-histidine tracts have been shown to bind divalent metal ions (Walty *et al*., 2015) and help localize proteins to sub-nuclear compartments (Salichs *et al*., 2009). However, the function of endogenous poly-histidine tracts in plant proteins remains elusive. Experimentation with *A. thaliana* LARP6b for the use of studying an endogenous poly-histidine tract was put on hold after multiple unsuccessful sequencing attempts. Troubleshooting by using different cloning methods and primers also proved unsuccessful. However, PCR amplification of *At*LARP6b, eGFP, and pER-HA appeared to be successful by agarose gel analysis. The use of restriction enzymes and ligase also appeared to be successful when samples were run by gel electrophoresis. The final product of a pER-HA vector containing a *At*LARP6b-eGFP fusion could not be obtained, and may require new cloning strategies to progress the project. There are multiple alternative cloning methods that could be utilized to create the target vector. An alternative *A. thaliana* vector could also be used, and primers redesigned for compatibility with restriction sites present in a new plasmid. The fusion product could also be sub-cloned into a bacterial vector, which would be easy to manipulate due to our lab's resources. This would also allow the fusion protein to be expressed, which would

glow green if indeed present in colonies, adding another verification step before being sent for sequencing.

The poly-histidine tracts in *At*LARP6b may also be causing instability within the protein. Comparing the limited proteolysis gels of *At*LARP6a, *At*LARP6b, and *At*LARP6c, one can see that *At*LARP6b quickly degrades, while full-length *At*LARP6c remains more intact and *At*LARP6a degradation products also appear more prevalent over time. *At*LARP6b and *At*LARP6c contain 3 out of 4 conserved sequences: the PAM2 motif, the La motif, and the LSA motif, yet exhibit different characteristics when exposed to the same enzyme. This could be tested by creating a fusion product with a mutant *At*LARP6b gene, where the poly-His tracts have been selectively deleted. The mutant fusion product would be inserted into the same pER-HA vector and sequenced to compare to the unsuccessful sequencing attempts. Further experimentation with recombinant plasmids containing *larp6a* and *larp6c* could also prove useful, not only to compare the success of fusion products but also to visualize localization within the cell through confocal microscopy experimentation. Localization of the full-length *A. thaliana* LARP6 proteins could be compared to the localization experiments done by Merret, as those were conducted with only the RNA-binding domains of the LARP6 proteins, and utilized transiently expressed onion cells.

Future Directions

The project to study the poly-histidine tracts in *At*LARP6b has already been set-up, but requires revision to retrieve a workable *A. thaliana* vector. The issues with eGFP cloning could also be remedied by using a different fluorescent protein. PCR amplification of genes was successful, creation of the fusion product was successful, restriction digest of the plasmid was successful, but sequencing of the final product was repeatedly unsuccessful. Troubleshooting this aspect and obtaining a complete *A. thaliana* vector would allow the project to continue. This vector could be used to transfect *A. thaliana* plants *in vivo*, allowing for hypoxic stress tests and visualization of protein localization by fluorescent microscopy. Since poly-His tracts have been shown to bind divalent metal ions (Walty *et al.*, 2015), conducting metal binding experiments on the His_7 and extended poly-His tract as found in *At*LARP6b could reveal potential mechanisms behind their function *in vivo*. By studying the cellular localization of GFP-tagged *At*LARP6b, data from these experiments could further our understanding of how the RNA binding proteins in the plant LARP6 family function, as well as the physiological role of all LARP6 proteins across species.

REFERENCES

Bitko, V., A. Musiyenko, M. A. Bayfield, R. J. Maraia, and S. Barik. "Cellular La Protein Shields Nonsegmented Negative-Strand Rna Viral Leader Rna from Rig-I and Enhances Virus Growth by Diverse Mechanisms." J Virol 82, no. 16 (Aug 2008): 7977-87. http://dx.doi.org/10.1128/JVI.02762-07.

Bohmer, F., S. Szedlacsek, L. Tabernero, A. Ostman, and J. den Hertog. "Protein Tyrosine Phosphatase Structure-Function Relationships in Regulation and Pathogenesis." FEBS J 280, no. 2 (Jan 2013): 413-31. http://dx.doi.org/10.1111/j.1742-4658.2012.08655.x.

Bornhorst, J. A. and J. J. Falke. "Purification of Proteins Using Polyhistidine Affinity Tags." Methods Enzymol 326 (2000): 245-54. https://www.ncbi.nlm.nih.gov/pubmed/11036646.

Bousquet-Antonelli, C. and J. M. Deragon. "A Comprehensive Analysis of the La-Motif Protein Superfamily." RNA 15, no. 5 (May 2009): 750-64. http://dx.doi.org/10.1261/rna.1478709.

Buchan, J. R. and R. Parker. "Eukaryotic Stress Granules: The Ins and Outs of Translation." Mol Cell 36, no. 6 (Dec 25 2009): 932-41. http://dx.doi.org/10.1016/j.molcel.2009.11.020.

Cai, L., D. Fritz, L. Stefanovic, and B. Stefanovic. "Binding of Larp6 to the Conserved 5' Stem-Loop Regulates Translation of Mrnas Encoding Type I Collagen." J Mol Biol 395, no. 2 (Jan 15 2010): 309-26. http://dx.doi.org/10.1016/j.jmb.2009.11.020.

Chen, D. F., C. Lin, H. L. Wang, L. Zhang, L. Dai, S. N. Jia, R. Zhou, R. Li, J. S. Yang, F. Yang, J. S. Clegg, H. Nagasawa, and W. J. Yang. "An La-Related Protein Controls Cell Cycle Arrest by Nuclear Retrograde Transport of Trnas During Diapause Formation in Artemia." BMC Biol 14 (Mar 03 2016): 16. http://dx.doi.org/10.1186/s12915-016-0239-4.

Churion, K. A. and S. E. Bondos. "Identifying Solubility-Promoting Buffers for Intrinsically Disordered Proteins Prior to Purification." Methods Mol Biol 896 (2012): 415-27. http://dx.doi.org/10.1007/978-1-4614-3704-8_28.

Costa-Mattioli, M., Y. Svitkin, and N. Sonenberg. "La Autoantigen Is Necessary for Optimal Function of the Poliovirus and Hepatitis C Virus Internal Ribosome Entry Site in Vivo and in Vitro." Mol Cell Biol 24, no. 15 (Aug 2004): 6861-70. http://dx.doi.org/10.1128/MCB.24.15.6861-6870.2004.

English, L. R., E. C. Tilton, B. J. Ricard, and S. T. Whitten. "Intrinsic Alpha Helix Propensities Compact Hydrodynamic Radii in Intrinsically Disordered Proteins." Proteins (Dec 09 2016). http://dx.doi.org/10.1002/prot.25222.

Faux, N. G., S. P. Bottomley, A. M. Lesk, J. A. Irving, J. R. Morrison, M. G. de la Banda, and J. C. Whisstock. "Functional Insights from the Distribution and Role of Homopeptide Repeat-Containing Proteins." Genome Res 15, no. 4 (Apr 2005): 537-51. http://dx.doi.org/10.1101/gr.3096505.

Fiumara, F., L. Fioriti, E. R. Kandel, and W. A. Hendrickson. "Essential Role of Coiled Coils for Aggregation and Activity of Q/N-Rich Prions and Polyq Proteins." Cell 143, no. 7 (Dec 23 2010): 1121-35. http://dx.doi.org/10.1016/j.cell.2010.11.042.

Kim, H. J., N. C. Kim, Y. D. Wang, E. A. Scarborough, J. Moore, Z. Diaz, K. S. MacLea, B. Freibaum, S. Li, A. Molliex, A. P. Kanagaraj, R. Carter, K. B. Boylan, A. M. Wojtas, R. Rademakers, J. L. Pinkus, S. A. Greenberg, J. Q. Trojanowski, B. J. Traynor, B. N. Smith, S. Topp, A. S. Gkazi, J. Miller, C. E. Shaw, M. Kottlors, J. Kirschner, A. Pestronk, Y. R. Li, A. F. Ford, A. D. Gitler, M. Benatar, O. D. King, V. E. Kimonis, E. D. Ross, C. C. Weihl, J. Shorter, and J. P. Taylor. "Mutations in Prion-Like Domains in Hnrnpa2b1 and Hnrnpa1 Cause Multisystem Proteinopathy and Als." Nature 495, no. 7442 (Mar 28 2013): 467-73. http://dx.doi.org/10.1038/nature11922.

Kimura, N., M. Uchida, S. Nishimura, and H. Yamaguchi. "Promotion of Polypeptide Folding by Interactions with Asn-Glycans." J Biochem 124, no. 4 (Oct 1998): 857-62. https://www.ncbi.nlm.nih.gov/pubmed/9756634.

Lim, C., J. Lee, C. Choi, V. L. Kilman, J. Kim, S. M. Park, S. K. Jang, R. Allada, and J. Choe. "The Novel Gene Twenty-Four Defines a Critical Translational Step in the Drosophila Clock." Nature 470, no. 7334 (Feb 17 2011): 399-403. http://dx.doi.org/10.1038/nature09728.

Martino, L., S. Pennell, G. Kelly, B. Busi, P. Brown, R. A. Atkinson, N. J. Salisbury, Z. H. Ooi, K. W. See, S. J. Smerdon, C. Alfano, T. T. Bui, and M. R. Conte. "Synergic Interplay of the La Motif, Rrm1 and the Interdomain Linker of Larp6 in the Recognition of Collagen Mrna Expands the Rna Binding Repertoire of the La Module." Nucleic Acids Res 43, no. 1 (Jan 2015): 645-60. http://dx.doi.org/10.1093/nar/gku1287.

Martino, L., N. J. Salisbury, P. Brown, G. Kelly, R. A. Atkinson, and M. R. Conte. "(1)H, (15)N and (13)C Chemical Shift Assignments of the La Motif and Rrm1 from Human Larp6." Biomol NMR Assign 9, no. 2 (Oct 2015): 337-40. http://dx.doi.org/10.1007/s12104-015-9605-3.

Merret, R., L. Martino, C. Bousquet-Antonelli, S. Fneich, J. Descombin, E. Billey, M. R. Conte, and J. M. Deragon. "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 19, no. 1 (Jan 2013): 36-50. http://dx.doi.org/10.1261/rna.035469.112.

Mruk, D. D. and C. Y. Cheng. "Enhanced Chemiluminescence (Ecl) for Routine Immunoblotting: An Inexpensive Alternative to Commercially Available Kits." Spermatogenesis 1, no. 2 (Apr 2011): 121-22. http://dx.doi.org/10.4161/spmg.1.2.16606.

Peng Z, Mizianty MJ, Kurgan LA. Genome-scale prediction of proteins with long intrinsically disordered regions. Proteins: Structure, Function, and Bioinformatics (2014) 82(1): 145-158.

Rossman, M. G. and A. Liljas. "Letter: Recognition of Structural Domains in Globular Proteins." J Mol Biol 85, no. 1 (May 05 1974): 177-81. https://www.ncbi.nlm.nih.gov/pubmed/4365123.

Salichs, E., A. Ledda, L. Mularoni, M. M. Alba, and S. de la Luna. "Genome-Wide Analysis of Histidine Repeats Reveals Their Role in the Localization of Human Proteins to the Nuclear Speckles Compartment." PLoS Genet 5, no. 3 (Mar 2009): e1000397. http://dx.doi.org/10.1371/journal.pgen.1000397.

Schaffler, K., K. Schulz, A. Hirmer, J. Wiesner, M. Grimm, A. Sickmann, and U. Fischer. "A Stimulatory Role for the La-Related Protein 4b in Translation." RNA 16, no. 8 (Aug 2010): 1488-99. http://dx.doi.org/10.1261/rna.2146910.

Segrest, J. P., M. K. Jones, H. De Loof, C. G. Brouillette, Y. V. Venkatachalapathi, and G. M. Anantharamaiah. "The Amphipathic Helix in the Exchangeable Apolipoproteins: A Review of Secondary Structure and Function." J Lipid Res 33, no. 2 (Feb 1992): 141-66. https://www.ncbi.nlm.nih.gov/pubmed/1569369.

Smalley, J. L., C. Breda, R. P. Mason, G. Kooner, R. Luthi-Carter, T. W. Gant, and F. Giorgini. "Connectivity Mapping Uncovers Small Molecules That Modulate Neurodegeneration in Huntington's Disease Models." J Mol Med (Berl) 94, no. 2 (Feb 2016): 235-45. http://dx.doi.org/10.1007/s00109-015-1344-5.

Smith, K. and M. J. Rennie. "New Approaches and Recent Results Concerning Human-Tissue Collagen Synthesis." Curr Opin Clin Nutr Metab Care 10, no. 5 (Sep 2007): 582-90. http://dx.doi.org/10.1097/MCO.0b013e328285d858.

Stavraka, C. and S. Blagden. "The La-Related Proteins, a Family with Connections to Cancer." Biomolecules 5, no. 4 (Oct 16 2015): 2701-22. http://dx.doi.org/10.3390/biom5042701.

Stefanovic, L., L. Longo, Y. Zhang, and B. Stefanovic. "Characterization of Binding of Larp6 to the 5' Stem-Loop of Collagen Mrnas: Implications for Synthesis of Type I Collagen." RNA Biol 11, no. 11 (2014): 1386-401. http://dx.doi.org/10.1080/15476286.2014.996467.

Studier, F. W. "Protein Production by Auto-Induction in High Density Shaking Cultures." Protein Expr Purif 41, no. 1 (May 2005): 207-34. https://www.ncbi.nlm.nih.gov/pubmed/15915565.

Tischer, A., H. Lilie, R. Rudolph, and C. Lange. "L-Arginine Hydrochloride Increases the Solubility of Folded and Unfolded Recombinant Plasminogen Activator Rpa." Protein Sci 19, no. 9 (Sep 2010): 1783-95. http://dx.doi.org/10.1002/pro.465.

Uversky, V. N. "Intrinsically Disordered Proteins and Their (Disordered) Proteomes in Neurodegenerative Disorders." Front Aging Neurosci 7 (2015): 18. http://dx.doi.org/10.3389/fnagi.2015.00018.

Watly, J., E. Simonovsky, N. Barbosa, M. Spodzieja, R. Wieczorek, S. Rodziewicz-Motowidlo, Y. Miller, and H. Kozlowski. "African Viper Poly-His Tag Peptide Fragment Efficiently Binds Metal Ions and Is Folded into an Alpha-Helical Structure." Inorg Chem 54, no. 16 (Aug 17 2015): 7692-702. http://dx.doi.org/10.1021/acs.inorgchem.5b01029.

Wek, R. C., H. Y. Jiang, and T. G. Anthony. "Coping with Stress: Eif2 Kinases and Translational Control." Biochem Soc Trans 34, no. Pt 1 (Feb 2006): 7-11. http://dx.doi.org/10.1042/BST20060007.

Witkowska, D., R. Politano, M. Rowinska-Zyrek, R. Guerrini, M. Remelli, and H. Kozlowski. "The Coordination of Niii and Cu(Ii) Ions to the Polyhistidyl Motif of Hpn Protein: Is It

as Strong as We Think?" Chemistry 18, no. 35 (Aug 27 2012): 11088-99. http://dx.doi.org/10.1002/chem.201200780.

Wynn, T. A. "Cellular and Molecular Mechanisms of Fibrosis." J Pathol 214, no. 2 (Jan 2008): 199-210. http://dx.doi.org/10.1002/path.2277.