Summer 2023

A Functional Study of Plant Protein Villin4 and The Application of Sortase Mediated Ligation in Intrinsically Disordered Protein Substrates

Jake Heins
Western Washington University, jake.heins47@gmail.com

A Functional Study of Plant Protein Villin4 and The Application of Sortase Mediated Ligation in Intrinsically Disordered Protein Substrates

By Jake Heins

Accepted in Partial Completion

Of the Requirements for the Degree

Masters of Science

ADVISORY COMMITTEE

Dr. Serge Smirnov, Chair

Dr. John Antos

Dr. Jeffery Young

GRADUATE SCHOOL

David L. Patrick, Dean

MASTER'S THESIS

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Western Washington University, I grant to Western Washington University the non-exclusive royalty-free right to archive, reproduce, distribute, and display the thesis in any and all forms, including electronic format, via any digital library mechanisms maintained by WWU.

I represent and warrant this is my original work, and does not infringe or violate any rights of others. I warrant that I have obtained written permissions from the owner of any third party copyrighted material included in these files.

I acknowledge that I retain ownership rights to the copyright of this work, including but not limited to the right to use all or part of this work in future works, such as articles or books.

Library users are granted permission for individual, research and non-commercial reproduction of this work for educational purposes only. Any further digital posting of this document requires specific permission from the author.

Any copying or publication of this thesis for commercial purposes, or for financial gain, is not allowed without my written permission.

Jake S. Heins

July 27, 2023

A Functional Study of Plant Protein Villin4 and The Application of Sortase Mediated Ligation in Intrinsically Disordered Protein Substrates

A Thesis

Presented to

The Faculty of

Western Washington University

Accepted in Partial Completion

Of the Requirements for the Degree

Master of Science

By

Jake Heins

July 2023

Abstract

Villin proteins are a family of filamentous actin (F-actin) regulators, playing vital roles in the organization of cytoskeletons in eukaryotes. Villin4, found in *A. thaliana* has a distinct structure that contains a folded C-terminal headpiece, a 192-residue disordered "linker", and a folded N-terminal core domain. Villin4 has been shown to bind and bundle F-actin in the C-terminal headpiece domain. This disordered linker, or intrinsically disordered region (IDR), lacks the defined three-dimensional structure often found in proteins. IDRs are not novel. In fact, over 30% of eukaryotic proteins contain a sizable intrinsically disordered region over 50 amino acids in length. These IDRs are multifunctional and play a variety of crucial roles in the function of proteins.

Because of the dynamic nature of IDRs, most standard biochemical methods of structural determination fall short. We have developed methods specifically suited to gain insight into proteins containing these IDRs, including segmentally labeled sortase mediated ligation, pull down assays and covalent modification studies. Through these efforts we have formulated a robust method of characterizing and optimizing the relative reactivity of sortase mediated ligation sites within authentic IDRs. We probed the backbone structure and dynamics of Villin4's linker using NMR, as well as calculating the specific and nonspecific binding capacity by formulating a holistic actin binding curve. We also probed predicted regions of the linker for specific covalent modifications in plant cell extract over a variety of time scales. Our work with sortase mediated ligation introduces a new method for researching IDRs. Using the knowledge gained in our covalent modification studies and our pull-down assays we lay the foundation for developing more robust crops.

Acknowledgments

I would like to begin by expressing my deepest gratitude to Dr. Serge Smirnov for providing me with the opportunity to become a member of his lab back in Spring of 2019. Even though I was completely new to the field of biochemistry, Dr. Smirnov took a chance on me and gave me the resources I needed to thrive. The past 5 years of research have been a whirlwind of ups and downs and without his guidance and encouragement, I can honestly say I have no idea where I would be today. Being able to conduct meaningful research has been one of my highlights during my tenure at Western Washington University.

I would also like to thank my MS thesis committee members Dr. John Antos and Dr. Jeff Young for their continued support. I am incredibly grateful that they put up with my never-ending questions and for the time they found for me. Without their permission to use their lab spaces and instruments, my work would never have been completed. Funding for much of this work was provided by grant 2004237 of the National Science Foundation.

In addition, I would like to thank the past and present members of the Antos and Smirnov Labs who supported and assisted me during my work. Specifically, Caleb Howerton for dedicating a significant portion of his time to assist me. I owe a huge thank you to the all the students who worked alongside me in CB470, for the endless entertainment and the genuine sense of comradery. Finally, a special thank you goes out to Irina Reinhardt for always believing in me and Paul Sage who unknowingly put me on this path.

Table of Contents

Abstract iv

Acknowledgements v

List of Figures and Tables xi

Chapter I Introduction 1

Intrinsically disordered regions and intrinsically disordered proteins 1

- **Challenges with studying sizable IDRs/IDPs** 1
- **Sortase-Mediated Ligation** 2

Actin and cytoskeleton regulation 4

- **F-actin Formation** 4
- **Actin Binding Proteins** 5

Villin 6

- **Plant Villin4** 7
- **Stress resistant plants** 8

Covalent modifications of plant villin in cell extract 9

- **Degradation/Proteolysis** 9
- **Phosphorylation** 10

Research Aims 11

Chapter II Materials and Methods 12

Plant Villin4 Constructs 12

- **Construct atVHP76** 12
- **Construct VL147N** 12
- **Construct VL192** 12

Sortase Mediated Ligation Constructs 13

Construct Design 13

Original Constructs 14

Additional Glycine Constructs 14

Protein Production 15

Plasmid Design 15

Plasmid Transformation 16

Protein Expression 16

IPTG Induction Optimization 17

Expression of ^{15}N Labeled Proteins 17

Soluble Protein Extraction 18

Insoluble Protein Extraction 19

Protein Purification 19

Ni-NTA Purification Gravity Column 19

Cleaning and Regeneration of Ni-NTA Resin 20

Gradient Ni-NTA Purification FPLC 20

Strep Purification Gravity Column 21

Cleaning and Regeneration of Strep-Tactin Resin 22

Size Exclusion Chromatography 23

Protein Identification and Quantification 23

SDS PAGE 23

LC-ESI-MS 24

LC-QTOF-MS 25

NMR 25

HPLC/FPLC UV-Vis 26

Nanodrop 26

Buffer Exchange and Protein Concentration 26

Desalting Columns 26

Dialysis Tubing 27

Spin Concentration 27

TEV Protease-Cleavage Reaction 28

Sortase Mediated Ligation 29

SML with WTsrtA 29

SML with 7mutA 29

SML with 7mutA + $NiSO_4$ 30

F-actin Assay and Calibration Curve 30

Calibration Curve 30

Preliminary F-actin Spin-down Assay 31

F-actin Spin-down Assays for Actin Binding Curve 33

Construction of the Binding Curve 35

***Arabidopsis Thaliana* Extract Assay** 35

Preparation of Plant Cell Extract 35

Exposure to Plant Cell Extract 36

Chapter III Results and Discussion 38

Aim 1 38

Preparation of Original Constructs 38

Results of SML on Original Constructs 38

Preparation of Additional Glycine Constructs 40

Results of SML on Additional Glycine Constructs 41

Aim 1 Discussion 43

Aim 2 44

Preparation of VL147N 44

Preparation of ^{15}N VL147N 45

HSQC of ^{15}N VL147N 47

Constructing the Calibration Curve 48

Actin Binding for VL147N and VL192 49

Actin Binding Curve for VL147N 50

Aim 2 Discussion 52

Aim 3 53

Preparation of VL192 53

Covalent modification of VL192 54

Aim 3 Discussion 57

Chapter IV Conclusions 59

Sortase Mediated Ligation 59

F-actin Binding of VL147N and VL192 60

Preliminary Probing of VL147N Backbone Dynamics 60

Covalent Modifications of VL192 and VL147N 61

References 62

Appendix I. Purification of additional glycine constructs **66**

Appendix II. FH8 tags do not bind F-actin **74**

Appendix III. VL147N and VL192 bind actin and Missing O-ring **75**

Appendix IV. HSQC of ^{15}N VL147N **78**

Appendix V. Deconvoluted spectra of VL192 after exposure to A. thaliana at time point 2.5 and 3 minutes **79**

Appendix VI. Sequences for villin constructs **81**

Appendix VII. Sequences for SML constructs **82**

List of Figures and Tables

Figure 1.1 Use of segmental labeling with NMR samples. **2**

Figure 1.2 Sortase as a cleavage and ligation enzyme. **3**

Table 1.1 Summary of SLIDR algorithm & DisProt data. **4**

Figure 1.3 Filamentous actin formation. **5**

Figure 1.4 Branching and bundling of filamentous actin. **6**

Figure 1.5 Plant villin structure. **7**

Table 1.2 Villin constructs to study. **8**

Figure 1.6 Potential ubiquitination and phosphorylation sites in VL192. **9**

Figure 2.1 Plant Villin4 construct design. **13**

Table 2.1 Original Aim 1 Constructs. **14**

Table 2.2 Additional Glycine Constructs. **15**

Table 2.3 Contents of micro-ultra centrifuge tubes for preliminary actin binding. **32**

Figure 3.1 Percent Ligation of the Original N-terminal Constructs. **39**

Figure 3.2 Complex formed by $NiSO_4$ and $GGHis_6$. **40**

Figure 3.3 15% SDS-PAGE of Additional Glycine Constructs. **41**

Figure 3.4 Percent Ligation of AcidN Constructs. **42**

Figure 3.5 Percent Ligation of AroN Constructs. **42**

Figure 3.6 Relative purity of VL147N sample. **45**

Figure 3.7 Relative purity of ^{15}N VL147N sample. **46**

Figure 3.8 HSQC of ^{15}N VL147N. **47**

Figure 3.9 Calibration Curve for VL147N. **49**

Figure 3.10 15% SDS-PAGE of VL147N and VL192 binding F-actin. **50**

Figure 3.11 Actin Binding Curve of VL147N. **51**

Figure 3.12 15% SDS-PAGE of Strep purification of VL192. **53**

Figure 3.13 15% SDS-PAGE of VL192 exposed to A. *thaliana* cell extract. **55**

Figure 3.14 HPLC Chromatogram of VL192 and Elution 1. **56**

Figure 3.15 15% SDS-PAGE of VL147N exposed to A. *thaliana* cell extract. **57**

Chapter I Introduction

Intrinsically disordered proteins and intrinsically disordered regions

Lacking a distinct structure, originally believed to be crucial for proteins, intrinsically disordered proteins/regions (IDPs/IDRs) challenge the fundamental theories of structural biology.[1-4] IDRs are prevalent in nature; with over 30% of eukaryotic proteins containing a sizable IDR of over 50 amino acids in length.[1-4] Unimpinged by secondary or tertiary structure, IDRs are inherently dynamic, capable of altering their conformation.[2-3,5] Subsequently IDRs are multifunctional, influenced by their environment, able to undergo post-translational modification and ultimately responsible of performing many key cellular functions.[2,4-5]

Challenges with studying sizable IDPs/IDRs

The same characteristics which make IDPs so interesting, are those which introduce a litany of challenges for biochemists trying to study them. X-ray crystallography is only capable of capturing a single conformation, which would form an incomplete representation of the IDP due to its dynamic nature.[1,6,7] Furthermore, some IDPs/IDRs are inherently incapable of forming a crystalline structure.[7] Standard solution NMR also falls short, as it can be difficult to express large IDRs at sufficient concentration, they are prone to degradation, and are dynamic in nature (results in severe spectral overlap).[6,7]

To circumvent the issue of spectral overlap, segmentally labeling manageable portions of the IDR can allow for clean interpretable peaks.[6,7] Segmental labeling (SL) is the process by which different portions of the native protein are isotopically labeled, while the other parts are left

unlabeled. Though this decreases the spectral overlap, significantly more time is needed to prepare each portion of the protein and to stitch together the labeled and unlabeled portions.[6] This process must be repeated for however many segments you have (Figure 1.1). SL has been utilized frequently when dealing with folded proteins, however only in the past few decades has SL been attempted with IDPs/IDRs.[6,7]

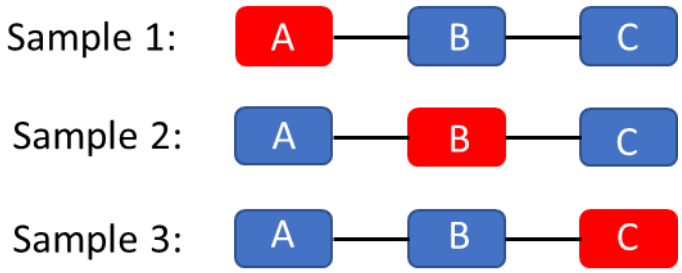


Figure 1.1 Example of how to utilize 3 different NMR samples with segmental labeling to study proteins. Red segments are ^{15}N labeled, blue segments are unlabeled.

One can surmise that if SL is difficult with folded proteins which primarily express more readily than IDRs, the task in which we have been attempting is truly daunting. Furthermore, the more standard ligation methods, including native chemical ligation, expressed protein ligation, and protein trans-splicing, have extreme limitations.[6,8] These standard methods must be at least partially completed *in vivo*.[6,8] Also, disulfide interactions due to the introduction of cysteines and loss of yield through purification plague standard methods of ligation.[8] As such, we have been exploring sortase mediated ligation (SML) as an improved method of SL.

Sortase Mediated Ligation

Sortase enzymes are cysteine transpeptidases found in gram-positive bacteria.[9] These enzymes can consistently cleave precise locations in a protein, as well as link two otherwise

distinct peptides together in what is termed sortase mediated ligation (SML).[9-11] Sortase enzymes have been implemented in a variety of protein engineering applications, such as SL, phosphorylation studies, and FRET, due to their robust nature and their relative abundence.[6,10,11] Also of note, sortase enzymes allow for all of this to be done *in vitro*.[6,9-11]

Sortase A (srtA), isolated from *Staphylococcus aureus* specifically recognizes and targets the LPXTG (X being any amino acid) motif.[6,10] The srtA enzyme facilitates a reversible cleavage between the threonine (T) and glycine (G) residues, found in the LPXTG motif (Figure 1.2a).[6,10] To utilize srtA in SL, the N-terminal construct must contain a LPXTG site, while the C-terminal construct must contain a glycine residue on its N-terminus.[6,10] Once the srtA enzyme cleaves between the T and G residues, a thioester-linked acyl enzyme intermediate undergoes a nucleophilic attack from a nucleophile with a N-terminal glycine residue (Figure 1.2b).[6] A nucleophilic attack then removes the sortase and ligates the two distinct fragments together via an amine bond (Figure 1.2b).[6]

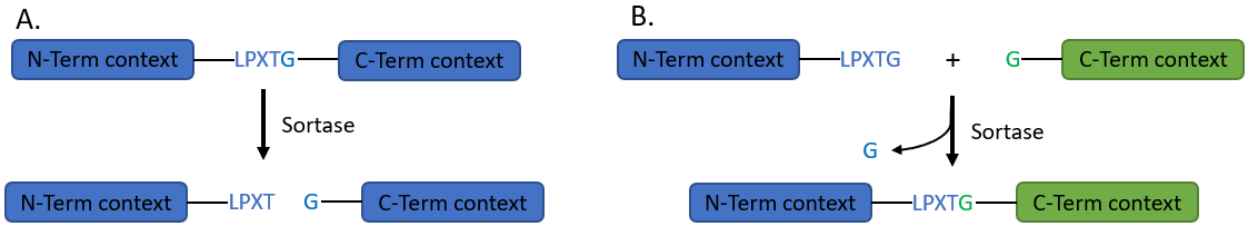


Figure 1.2 Sortase as a cleavage and ligation enzyme. **(A)** Schematic as to how Sortase can cleave a protein into two pieces. **(B)** Schematic as to how sortase can be used to ligate two distinct contexts together.

Though wild-type sortase A (WTsrtA) is a robust and effective catalyst for SML, biochemists have genetically engineered it to increase ligation speed and efficiency.[6,11] One such mutant, hepta-mutant sortase A (7mutA), has been the primary comparison to WTsrtA in our

trials.[11] On top of having faster catalytic activity, 7mutA does not depend on calcium for its activation.[11] Both sortase enzymes target the LPXTG motif, which is common in proteins.[6,11] With minimal point mutations, these sites become even more abundant (Table 1.1).[6,11]

Instance Type	Number of Instances
Proteins with large IDRs	625
Native LPXTG sites	4
LPXTG sites with 0-2 mutations	8147

Table 1.1 Summation of SLIDR algorithm & DisProt data. DisProt is a database which contains experimentally verified proteins over 100 a.a. in length. SLIDR is a computational algorithm proposed and designed by Dr. John Antos and Dr. Serge Smirnov. Its purpose was to find potential SML sites within the DisProt server. SLIDR found over 8000 LPXTG sites accessible using two point mutations.

Actin and cytoskeleton regulation

F-actin Formation

The cytoskeleton in a eukaryotic cell is comprised of three types of filamentous proteins, microfilaments, intermediate filaments, and microtubules.[12-15] These microfilaments perform essential, versatile roles in the construction of a cell, including maintaining cell morphology, and polarity, as well as assisting in motility and cellular division.[14-16] Within these microfilaments, a protein known as actin resides.[15-18] Actin exists in two primary forms, monomeric globular particles (G-actin) and polymerized, polarized filaments (F-actin).[16,18,19]

The formation of F-actin is a multi-stage process where ATP facilitates the binding of G-actin monomers into F-actin.[14,16,19] The first stage, nucleation, results in the formation of a nucleus. This nucleus is lengthened in a process called elongation.[14,16,19] Elongation refers to

subsequent G-actin being bound to either end of the nucleus.[14,16,19] Eventually an equilibrium between G-actin, F-actin and ATP is reached. This is called the steady state, where the F-actin no longer grows or shrinks (Figure 1.3).[14,16,19] The equilibrium and size of F-actin can further fluctuate depending on the available ATP.[20-22]

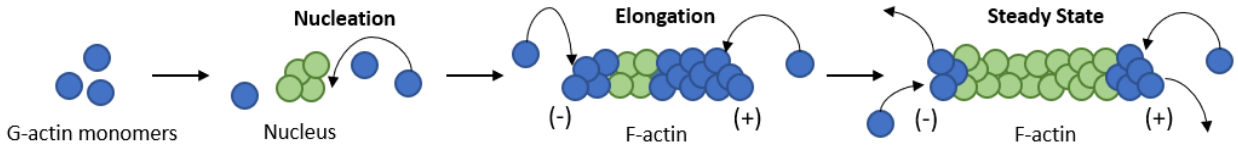


Figure 1.3 Process of F-actin formation from G-actin monomers. In elongation, more G-actin binds to the positively charged "barbed" side than the negatively charged side. At steady state, G-actin is gained and lost at an equal rate. Green portions refer to what has happened in the past stage, blue is the recent/ongoing process.

Actin Binding Proteins

F-actin is a dynamic protein and must be tightly regulated to ensure cohesion in the cell.[14,16,18] Actin binding proteins (ABPs) are necessary for actin to be organized into complex networks and higher-order stuctures.[18,23] Actin binding proteins can serve many functions in the regulation of F-actin such as: polymerization, depolymerization, end-capping, severing, crosslinking, stabilizing, and motoring.[17,18,23]

F-actin adopts two primary architectures: branched F-actin networks and linear F-actin bundling (Figure 1.4).[14-16,24-26] Branched F-actin networks are formed when actin binding proteins regulate F-actin to form orthogonal arrays with large, flexible spacer regions.[14-16,23,26] The spacer regions enable F-actin to branch in an almost perpendicular manner.[15-17,23] Linear F-actin bundling can occur when actin binding proteins consist of two or more distinct actin binding

locations.[15-17,23] Bundling is a result of parallel-like alignment of F-actin in a linear array, with either loose or tight bundles.[15-17,23] F-actin bundles are often found in cells that utilize polarized filaments, such as, microvilli.[14-17,24-26]

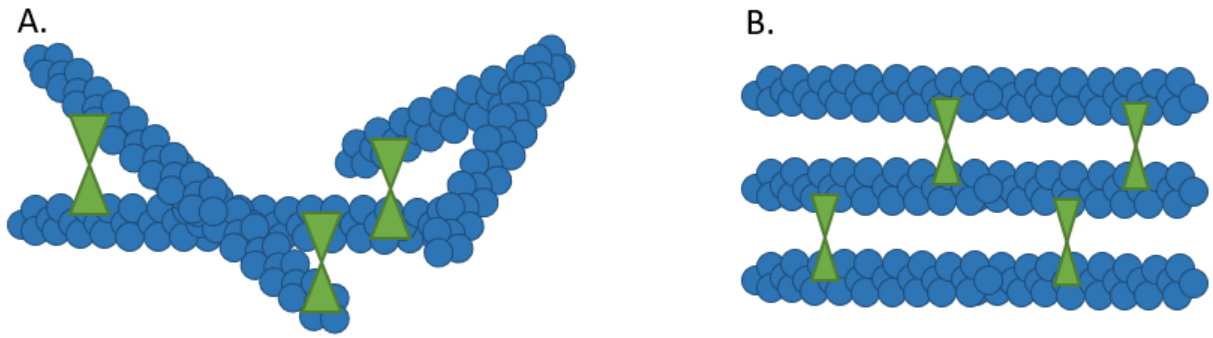


Figure 1.4 Branching and bundling of F-actin. **(A)** F-actin branched network. **(B)** F-actin linear bundling.

To some extent all actin binding proteins regulate F-actin, however in the work to be detailed, villin gene family members, specifically Villin4 is the focus. Villin homologs are a family of eukaryotic actin binding proteins found in numerous cell and tissue types.[14-18,20,21] Villin gene family members regulate F-actin through, binding, bundling, severing, capping, nucleating, and sequestering.[15,24,26-29] Regulation is primarily achieved through biological signaling of Ca^{2+}, tyrosine phosphorylation, and phosphatidylinositol-4,5-biphosphate.[15,29]

Villin

Most villins share a similar structure with six N-terminal homologous domains, referred to as the gelsolin core.[18,19,24,29] Some of these domains contain F-actin binding sites.[18,19,24,29] Plant villin also contain the gelsolin core, however in addition to a C-terminal headpiece domain, they also contain large IDRs which are often referred to as linkers (Figure 1.5).[24,27,29,30] This linker

is a large IDR stretching up to 192 residues in length with significant charge partitioning (basic N terminus and acidic C terminus).[29,30]

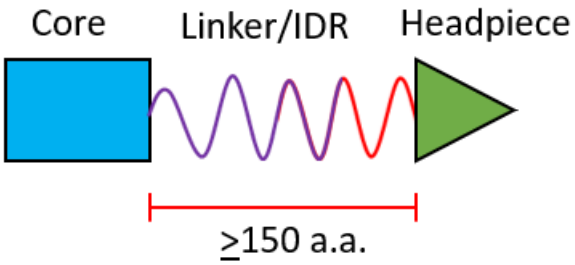


Figure 1.5 Mockup of a plant villin's structure. Red portion of the linker/IDR is basic. Purple portion of the linker is acidic. The split between basic and acidic charged portions are not equal or uniform but are depicted as such for simplicity.

Plant Villin4

In *Arabidopsis thaliana* (*A. thaliana*), five different gene family members of villin have been discovered.[31] Villin4, is primarily located in root hairs, which like microvilli are variable structures responsible for absorbing nutrients and water.[29-31] Villin4 can also be found in a variety of other plants, including many crops.[24,26,27,30] It has been predicted in sequence analysis and preliminary studies that Villin4 contains actin binding and bundling capabilities.[28-30] In 2018, the Smirnov lab showed that the headpiece region of Villin4 specifically binds F-actin with a K_d of 3.0 $\pm$ 0.7 µM.[30] To examine the linker of Villin4 for specific F-actin binding capabilities a variety of constructs were created (Table 1.2).

Construct Names	Constructs Contents
VL192	Strep - GSGGAS - Strep - G3 - IDR192 - G2 - His8
VL147N	Strep - GSGGAS - Strep - G3 - His10 - G3 - TEV - IDR147 - GRA
VHP101C	Strep - GSGGAS - Strep - G3 - His8 - TEV - G2 - IDR38 - HP63
atVHP76	His6 - IDR13 - HP63

Table 1.2 Strep stands for a strep tag consisting of WSHPQFEK, G# corresponds to the number of consecutive glycine's, His# corresponds to number of consecutive histidine's, TEV corresponds to ENLFYQ motif which TEV protease recognizes. VL192 is Villin4's full length linker with strep and His tags on either side of it. VL147N is the first 147 residues of Villin4's linker starting from the N terminal side. VHP101C is the 101 residues starting from the C terminus, including the HP.

Stress resistant plants

With the continuous increase in human population and the ever-looming issue of global warming, agricultural demands are exacerbated.[32] Water shortages are responsible for the greatest crop losses around the world, as such there is an obvious need to improve our sustainability and efficiency when it comes to our use of water.[32] Conventional agricultural practices, with sufficient knowledge of weather cycles have served us well in the past, but the current variety of crops cannot keep up with the ever-increasing fluctuation in temperature. Genetic engineering can allow us to tip the scales in the favor of our crops.[32-34] One such method is to increase the capabilities of our crops to absorb nutrients and water. To do so we must understand the structure and function of the proteins responsible for the health and development of root systems. This necessity gives credence to the study of plant Villin4 and F-actin, particularly due to plant villin being the primary regulators for root system development and function.

Covalent modifications of plant villin in cell extract

Covalent modifications are enzyme catalyzed alterations to proteins, which influence protein function and signaling.[35,36] Examples include degradation through ubiquitination, proteolysis, and phosphorylation.[35] IDRs are particularly prone to degradation though proteolysis, as well as through ubiquitination. Phosphorylation can produce conformational changes, which in many cases alter the function of an IDR.[36]

Degradation/Proteolysis

There are numerous instances when targeted degradation serves a desired function in enzymology. However, when dealing with IDRs degradation is often a curse. With their large, exposed regions IDRs can be degraded enzymatically through proteolysis. This often occurs through the ubiquitin-mediated pathway.[34,35] Though ubiquitin can serve other roles, its primary function is that of targeting proteins for degradation.[35] This pathway involves ubiquitin to be activated by ATP and target lysines.[35] VL192 contains 21 distinct lysine residues, 6 of which are predicted to be ubiquitinated (Figure 1.6).[34]

VL192

```
1   WSHPQFEKGS GGASWSHPQF EKGGGDSSKS AMHGNSFQRK LKIVKNGGTP  50
51  VADKPKRRTP ASYGGRASVP DKSQQRSRSM SFSPDRVRVR GRSPAFNALA 100
101 ATFESQNARN LSTPPPVVRK LYPRSVTPDS SKFAPAPKSS AIASRSALFE 150
151 KIPPQEPSIP KPVKASPKTP EAAAGGGAGK EQEEKKENDK EEGSMSSRIE 200
201 SLTIQEDAKE GVEDEEDGGH HHHHHHH                         227
```

Figure 1.6 Potential phosphorylation and ubiquitination sites in VL192. Yellow highlights are locations of predicted ubiquitination, green highlights are locations of predicted phosphorylation, red text are Strep tags, green text are generic spacers, blue text are His tags.

Phosphorylation

Conformational changes found in IDPs/IDRs are often strong indications of multifunctionalality.[34] Phosphorylation is often responsible for these changes. As phosphorylation alters the electric charge of the targeted amino acid, the overall pI value can change.[34-36] Phosphorylation also has an effect on a proteins capacity to engage in internal and external noncovalent interactions.[34-36] Phosphate groups binding to amino acids can act similar to a light switch, turning on and off a specific function of a protein.[36] VL192 has 3 sites that were predicted to undergo phosphorylation depending on the kinases present (Figure 1.6).[34]

Research Aims

I. Characterize the relative reactivity of sortase mediated ligation sites within authentic intrinsically disordered region sequences and create guidelines for optimization of reaction conditions.

II. Probe backbone structure and dynamics and generate an actin binding curve of plant protein Villin4 linker, to quantify the specific and nonspecific binding capacity of filamentous actin.

III. Probe predicted regions of plant protein Villin4 linker for specific covalent modifications in plant extract using bioinformatic mass spectrometry.

Chapter II Materials and Methods

Plant Villin4 Constructs

Construct atVHP76

atVHP76 was designed and studied by previous WWU graduate student Heather Miears. From N to C this construct has a $His_{(6)}$ tag, a 13 residue IDR and the 63 residues pertaining to the headpiece region (Figure 2.1 A).

Construct VL147N

VL147N was designed to examine the basic region of plant protein Villin4's linker. From N to C this construct has a strep tag, a glycine/serine/alanine spacer, a second strep tag, a glycine spacer, a $His_{(10)}$ tag, a TEV cleavage site, 147 residue IDR, and a glycine/arginine/alanine spacer (Figure 2.1 B).

Construct VL192

VL192 was designed to include the entire 192 residue linker found in plant protein Villin4, with purification tags on either end to provide a level of protection to the 192 residue IDR. From N to C this construct has a strep tag, a glycine/serine/alanine spacer, a second strep tag, a glycine spacer, 192 residue IDR, a glycine spacer and a $His_{(8)}$ tag (Figure 2.1 C).

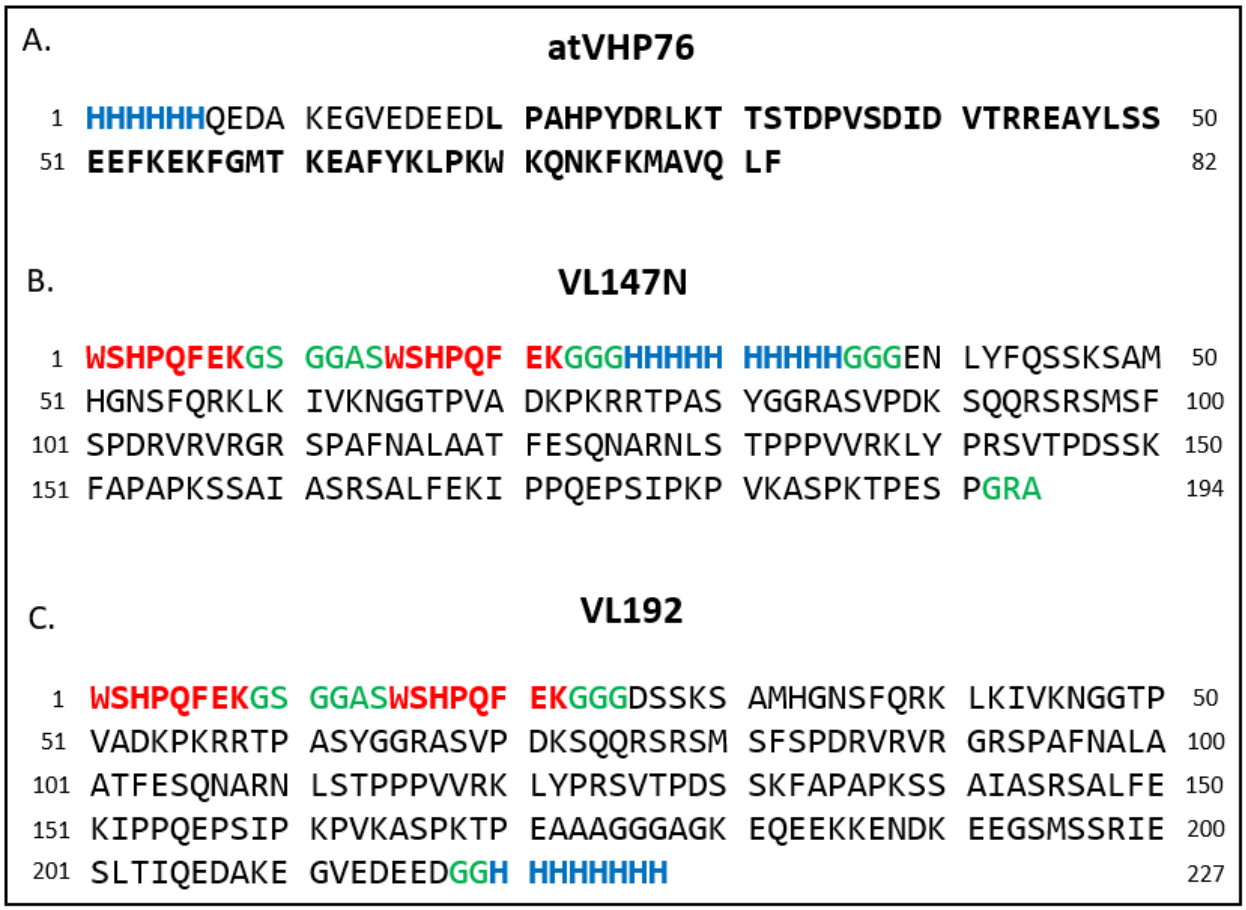


Figure 2.1 Plant Villin4 construct design. Blue texts are His tags, bolded black text is folded headpiece, red text are Strep tags, green text are generic spacers.

Sortase Mediated Ligation Constructs

Construct Design

The sortase mediated ligation constructs were designed to test the efficiency of SML in six authentic extreme scenarios. The six scenarios were proline, aromatic, acidic, basic, aliphatic, and polar rich residues flanking the LPXTG site. A control was also used with five glycine's in a row. From N to C these constructs had a 70 residue FH8 tag, a TEV cleavage site, a glycine/serine spacer, an extreme scenario, a LPXTG site, a glycine spacer and a $His_{(6)}$ tag.

Original Constructs

Originally 13 constructs were created, one control, the six extreme scenarios on the N-terminal side, and the six extreme scenarios on the C terminal side (Table 2.1). Ligation partners were made by the Antos lab.

Construct Name	Description	Protein Construct Sequence	Ligation Partner Peptide Sequence
*ProN	Proline rich N-term	FH8-ENLYFQG-GGGGSGGGGS-**PPPPP**-LPTTG-G-His_6	G – **ASTSE** – K(DNP)
ProC	Proline rich C-term	FH8-ENLYFQG-GGGGSGGGGS-**KEELH**-LPMTG-G-His_6	G – **PPPPP** – K(DNP)
*AroN	Aromatic rich N-term	FH8-ENLYFQG-GGGGSGGGGS-**WHIWW**-LPITG-G-His_6	G – **FAGMI** – K(DNP)
AroC	Aromatic rich C-term	FH8-ENLYFQG-GGGGSGGGGS-**NPEAP**-LPVTG-G-His_6	G – **YYHWD** – K(DNP)
*AcidN	Acidic N-term	FH8-ENLYFQG-GGGGSGGGGS-**DDEED**-LPSTG-G-His_6	G – **ELEDW** – K(DNP)
*AcidC	Acidic C-term	FH8-ENLYFQG-GGGGSGGGGS-**SPANA**-LPNTG-G-His_6	G – **DDDDD** – K(DNP)
*BasN	Basic N-term	FH8-ENLYFQG-GGGGSGGGGS-**KRKRR**-LPVTG-G-His_6	G – **DTNTK** – K(DNP)
BasC	Basic C-term	FH8-ENLYFQG-GGGGSGGGGS-**GDPNQ**-LPRTG-G-His_6	G – **RKRKT** – K(DNP)
AliN	Aliphatic N-term	FH8-ENLYFQG-GGGGSGGGGS-**LVMVI**-LPPTG-G-His_6	G – **FMDTP** – K(DNP)
AliC	Aliphatic C-term	FH8-ENLYFQG-GGGGSGGGGS-**VGVRT**-LPDTG-G-His_6	G – **LLVLL** – K(DNP)
*PolN	Polar N-term	FH8-ENLYFQG-GGGGSGGGGS-**DRIKE**-LPETG-G-His_6	G – **LGQNE** – K(DNP)
*PolC	Polar C-term	FH8-ENLYFQG-GGGGSGGGGS-**GFFGN**-LPQTG-G-His_6	G – **DKLKD** – K(DNP)
*GGGGG	Glycine rich N-term	FH8-ENLYFQG-GGGGSGGGGS-**GGGGG**-LPETG-G-His_6	G – **GGGGG** – K(DNP)

Table 2.1 The original 13 constructs designed to study SML. Bolded text are the amino acids being studied. Protein constructs with an asterisk by their name are those which I helped produce.

Additional Glycine Constructs

After the initial set of trials, it was determined that more information was needed on the proline, aromatic, acidic, and polar examples, specifically the N terminal constructs. As such the "additional glycine constructs" were created. The 1G variants mimicked the original constructs in design with only the amino acid closest to the N terminal side of the leucine being mutated to

a glycine. The 2G variants were the same, except the two amino acids closest to the N terminal side of the leucine were mutated to glycine's (Table 2.2). Ligation partners were made by the Antos lab.

Construct Name	Description	Protein Construct Sequence	Ligation Partner Peptide Sequence
ProN G	Proline rich N-term with a mutated glycine	FH8-ENLYFQG-GGGGSGGGGS-**PPPPG**-LPTTG-G-His_6	G – **G**STSE – K(DNP)
ProN 2G	Proline rich N-term with two mutated glycine's	FH8-ENLYFQG-GGGGSGGGGS-**PPPGG**-LPTTG-G-His_6	G – **GG**TSE – K(DNP)
AroN G	Aromatic rich N-term with a mutated glycine	FH8-ENLYFQG-GGGGSGGGGS-**WHIWG**-LPITG-G-His_6	G – **G**AGMI – K(DNP)
AroN 2G	Aromatic rich N-term with two mutated glycine's	FH8-ENLYFQG-GGGGSGGGGS-**WHIGG**-LPITG-G-His_6	G – **GG**GMI – K(DNP)
AcidN G	Acidic N-term with a mutated glycine	FH8-ENLYFQG-GGGGSGGGGS-**DDEEG**-LPSTG-G-His_6	G – **G**LEDW – K(DNP)
AcidN 2G	Acidic N-term with two mutated glycine's	FH8-ENLYFQG-GGGGSGGGGS-**DDEGG**-LPSTG-G-His_6	G – **GG**EDW – K(DNP)
PolN G	Polar N-term with a mutated glycine	FH8-ENLYFQG-GGGGSGGGGS-**DRIKG**-LPETG-G-His_6	G – **G**GQNE – K(DNP)
PolN 2G	Polar N-term with two mutated glycine's	FH8-ENLYFQG-GGGGSGGGGS-**DRIGG**-LPETG-G-His_6	G – **GG**QNE – K(DNP)
GGGGG	Glycine rich N-term	FH8-ENLYFQG-GGGGSGGGGS-**GGGGG**-LPETG-G-His_6	G – **GGGGG** – K(DNP)

Table 2.2 The 8 additional constructs and the original pentaglycine control designed to further study SML. Bolded text are the amino acids being studied, highlighted amino acids are the point mutations made to the original constructs. I was responsible for the production of all these protein constructs.

Protein Production

Plasmid Design

All constructs were cloned into a pET-24a vector (Novagen) using commercial services (GenScript) using chemically competent BL21 E. *coli* cells unless explicitly stated otherwise.

Plasmid Transformation

Transformation of all constructs were produced according to procedures adapted from WWU graduate students Derek McCaffery and Melissa Oueld Es Cheikh master's theses. The competent BL21(DE3) E*. coli* cells were transformed with the corresponding pET-24a plasmid provided by GenScript (pET-30b(+) provided by Hidde Ploegh Laboratory in the case of WTsrtA and 7mutA) via heat shock transformation. **I found the greatest success heat shocking the cells for 30 seconds at 42 ºC**. The cells were then plated and spread onto kanamycin (50 ug/mL) Luria-Berani (LB) agar plates (15 g/L agar, 10 g/L Bacto tryptone, 10 g/L NaCl, 5 g/L yeast extract) using sterile procedure. Once the plates had visibly absorbed the cells, the plates were placed upside down in an incubator for 12-16 hours at 37ºC. Photos were then taken of the plates for record keeping.

Protein Expression

A "seed" growth was performed by taking an isolated transformant with a sterile pipette tip and ejected into 100 mL of LB (5 g/L Bacto yeast extract, 10 g/L Bacto tryptone, 10 g/L NaCl) and incubated for 12-16 hours at 37ºC, shaken at 210 rpm. 25-50 mL of the seed growth was added to 1 L of LB and incubated at 37ºC at 210 rpm until an optical density between 0.5 and 0.6 was reached at 600 nm (2-4 hours). **If 50 mL of seed growth was added the optical density would reach 0.6 in 2 hours. This process can be further expedited by warming the 1 L of LB to 37 ºC prior to adding the seed growth.** The remaining seed growth was mixed in a 3:1 ratio mixture with 100% glycerol, flash frozen with liquid nitrogen and stored in the -80ºC freezer as

cell stock. It was imperative to fully mix the cell stock and immediately flash freeze before the glycerol had a chance to separate.

IPTG Induction Optimization

All villin based constructs besides atVHP63 and VL192 were induced to a final concentration of **0.4 mM IPTG**, while all other constructs including atVHP63 and VL192were induced to a final concentration of **0.8 mM IPTG**. The 1L growths would continue to incubate for 5 hours at 37^{o}C and 210 rpm. The growths were then centrifuged for 30 minutes at 4,500 x g at 4^{o}C. The supernatant was poured off and sterilized via 20% bleach and the cell pellets were scrapped into a 50 mL falcon tube, weighed and flash frozen with liquid nitrogen before storing them at -20^{o}C.

Expression of ^{15}N Labeled Proteins

The protocol used to express ^{15}N labeled proteins was adapted from WWU graduate student Heather Miears master's thesis and optimized by Caleb Howerton and myself. Following the methods described in "Protein Expression" the contents of the 1L flasks were centrifuged for 30 minutes at 5,000 x g at 4^{o}C. The supernatant was disposed of, and the pellet is resuspended in 1L (for a 6L growth) of sterilized M9T minimal media (6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl, pH. 7.0) containing no nutrients. The mixture was then centrifuged again for 30 minutes at 5,000 x g at 4^{o}C and the supernatant was disposed. The pellet was carefully resuspended in 1L (for a 6L growth) of nutrient rich M9T media (6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 3 g/L

$^{15}NH_4Cl$, pH. 7.0) and allowed to equilibrate on a shaker for up to 1 hour set to 210 rpm and 37°C. **In our experience it took 5-10 minutes to get the last remnants of the pellet to solubilize. Usually, you want to check the OD of the media to insure it's at 0.5-0.6, however for every ^{15}N growth that we did it had already exceeded the desired absorbance and did not need to equilibrate longer.** The media was spiked to a final concentration of 0.8 mM IPTG, incubated for 5 hours at 210 rpm at 37°C. The contents were then centrifuged for 30 minutes at 5,000 x g and 4°C. The pellet was collected and weighed before we proceeded with the protocol described in "Insoluble Protein Extraction".

Soluble Protein Extraction

The cell pellet was resuspended in 6-8 mL lysis buffer/1 gram of cell pellet using the relevant lysis buffer. **Villin based and Aim 1 constructs used 50 mM NaH_2PO_4, 300 mM NaCl, 10 mM Imidazole, pH 8.0; WTsrtA and 7mutA used 50 mM Tris, 150 mM NaCl, pH 7.6.** Once the pellet was fully resuspended, 50 mg/mL Lysozyme was added to a final concentration of 0.5 mg/mL and shaken at 4°C for 30 minutes at ~70 rpm. The cells were then sonicated in three 60 second intervals (Branson sonifer, 50% duty cycle, 15 mm tip). DNase (10 units/µL) was added to a final concentration of 10 µg/mL and incubated at room temperature for 30 minutes, ~70 rpm. The lysate was then transferred to centrifuge bottles and centrifuged for 30 minutes at 40,000 x g at 4°C. The supernatant was then filtered with a 0.45 µm filter and a 0.22 µm filter, before proceeding to the relevant purification protocol. The cell material was transferred into a 50 mL falcon tube and stored at -20°C, in the scenario that a significant amount of protein was found in the inclusion bodies of the cells.

Insoluble Protein Extraction

The cell pellet is resuspended in 6-8 mL Solubilizing Buffer (6 M Urea, 50 mM NaH_2PO_4, 300 mM NaCl, 10 mM Imidazole, pH. 8.0)/1 gram of cell pellet and left mixing for 12-16 hours at 4°C. The lysate was then sonicated in three 60 second intervals (Branson sonifer, 50% duty cycle, 15 mm tip). The lysate was then transferred to centrifuge bottles and centrifuged for 30 minutes at 40,000 x g at 4°C. The supernatant was filtered with a 0.45 μm filter and a 0.22 μm filter, before proceeding to the relevant purification protocol. **This protocol is utilized if significant protein is found in the inclusion bodies of the cell. This protocol can also be used instead of the "Soluble Protein Extraction" in applicable scenarios.**

Protein Purification

Ni-NTA Purification Gravity Column

Ni-NTA purification using a gravity column was preformed using procedures adapted from WWU graduate students Derek McCaffery and Melissa Oueld Es Cheikh master's theses. 3 mL of QIAGEN nickel nitrilotriacetic acid (Ni-NTA) resin was the standard column volume (CV). The filtered lysate was allowed to equilibrate with the resin for 30 minutes at 4°C. Three washes (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM Imidazole, pH 8.0), each 3 CVs were collected into a 15 mL falcon tube. Following the washes 0.5 CVs of IMAC Elution Buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM Imidazole, pH 8.0) was added to the column and allowed to equilibrate for 30 minutes at 4°C before collecting the elution in an Eppendorf tube. A total of 5 CVs of IMAC Elution

buffer, added in 1 mL fractions, was pipetted onto the column, and collected in Eppendorf tubes. After the initial addition of IMAC Elution Buffer, it was not necessary for equilibration.

Cleaning and Regeneration of Ni-NTA Resin

After every use the Ni-NTA resin was cleaned with 2 CVs of IMAC Elution Buffer, 2 CVs of nanopure water, 2 CVs of MES Buffer (20 mM MES, 200 mM NaCl, pH 6.0), 2 CVs of nanopure water, and 2 CVs of 20% ethanol prior to being stored in 20% ethanol at 4°C. After every five uses the Ni-NTA resin was regenerated with 10 CVs of nanopure water, 10 CVs of 100 mM EDTA, 10 CVs of nanopure water, 10 CVs of IMAC Wash Buffer, 10 CVs of nanopure water, 10 CVs of 100 mM $NiSO_4$, 10 CVs of nanopure water, 10 CVs of 20% ethanol and then stored in 20% ethanol at 4°C. **I found that 10 CVs was excessive, and 2-4 CVs was sufficient. However, I would recommend watching for the color changes for confirmation.** The color changes were from blue/teal to white during the addition of EDTA and from white to blue/teal during the addition of $NiSO_4$.

Gradient Ni-NTA Purification FPLC

Gradient Ni-NTA purifications were occasionally required. These were performed with an FPLC on a Biorad NGC Chromatography System Quest 10 Plus using a 5 mL prepacked Bio-Scale Mini Nuvia IMAC Ni-Charged column from Bio-Rad. Solutions of 20% ethanol, nanopure water, IMAC Wash Buffer and IMAC Elution buffer were prepared, as well as 1.5 mL Eppendorf tubes with the caps cut off for the gradient elution. Alternatively, Tris Wash Buffer and Tris Elution

Buffer could be substituted for the IMAC buffers, with no discernible adverse effect to the protein, instrument, column, or efficacy. The automated program named "Affinity IMACFeb 2020_1.5mL frac_5mL_loop", stored in ChromoLab 4.0 was used. Prior to executing the gradient purification, the system was manually prepped by flushing all the lines and pumps with nanopure water. The program was then initiated, which flushed the lines and column with 3 CVs of IMAC Wash Buffer. The system then prompts you to inject the sample. It is imperative that you do not introduce air bubbles or microparticles. To prevent this, filter your sample with a 0.22 µm filter and remove all air bubbles from the syringe. It may take multiple attempts to remove the air bubbles from the syringe. Following the injection, resume the program on the computer. The instrument will then wash the column with 3 CVs of IMAC Wash Buffer and follow a linear gradient from 0-100% of IMAC Elution Buffer over 5 CVs. These elution's will be collected by the system in the Eppendorf tubes provided. The system will then wash the column with 5 CVs of IMAC Wash Buffer. All automated steps performed by the instrument were done at flowrate of 5 mL/min. After the system is finished the instrument must be manually instructed to wash the lines and pumps with 5 CVs of nanopure water and 5 CVs of 20% ethanol. As with all purifications, ensure the sample is in a compatible loading buffer (IMAC Lysis Buffer, Tris Lysis Buffer, etc.) prior to injection.

Strep Purification Gravity Column

Strep affinity chromatography using a gravity column was optimized for my villin constructs from the procedures of former WWU students Derek McCaffery, Melissa Oueld Es

Cheikh and Erin Rosenkranz. 2 mL of Strep-Tactin XT® 4flow High-capacity resin* was the standard column volume used. **This resin has a binding capacity of 16 mg/mL (mg of protein/mL of resin).** The resin was drained and then rinsed with 2 CVs of Strep Wash Buffer AKA Buffer W (100 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, pH 8.0). The resin was then suspended in either lysate or the pooled fractions from a previous purification. **If the solution being added to the column does not contain biotin, no buffer exchange is needed. I found that it was always best to perform an IMAC purification prior to a Strep purification.** The resin was allowed to mostly settle before the flow through was collected. The resin was then washed with two 2.5 CVs of Strep Wash Buffer for a total of 5 CVs using 1 mL additions. The protein was then eluted in 1 mL fraction for a total of 6 CVs of Strep Elution Buffer AKA Buffer BXT (100 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 100 mM Biotin, pH 8.0). **Minimize the duration which Strep Elution Buffer is kept at room temperate, putting it on ice during the purification is advised.**

Cleaning and Regeneration of Strep-Tactin Resin

The resin was cleaned by washing 5 CVs of Strep Wash Buffer. The resin was stored in this buffer at 4°C. Note that the manufacturer's manual is somewhat vague as to the frequency at which one should regenerate the resin. They claim that the resin can be used up to 5 times before needing to be regenerated. **I suggest regenerating the resin after every 3 uses. You can calculate the theoretical amount of resin free of biotin with the provided binding capacity of the resin (16 mg/mL).** The resin was regenerated with 15 CVs of 3M $MgCl_2$ following 8 CVs of

* From IBA life sciences

Strep Wash Buffer. **I suggest rinsing the resin with 5 CVs of nanopure water prior to washing the resin with 8 CVs of Strep Wash Buffer to insure all the of the $MgCl_2$ is removed.**

Size Exclusion Chromatography

Should a construct need further purification after using Ni-NTA, Gradient Ni-NTA, and/or Strep purification Size Exclusion Chromatography can be used. Refer to previous WWU graduate student Heather Miears's master's thesis for the protocol. **Note this is a lengthy process that seems to be ill suited for large IDRs similar to the ones being studied in this thesis.**

Protein Identification and Quantification

SDS PAGE

Efficacy of purification and overall protein purity was determined via 15% SDS-PAGE. Electrophoresis was run at 90 volts until the protein front crossed the resolving line and then increased to 180 volts. For a molecular weight standard, 5 µL of Thermo Fisher's Scientific Spectra Multicolor Broad Range Protein Ladder was added to each gel. The standard procedure was to mix 15 µL of sample with 5 µL of 4X stain (100 µL βME, 900 µL 4x Laemmli Sample Buffer) in an Eppendorf and pipette 15 µLs of this sample into a well.

If the sample was not soluble (pre lysate or cell material), Cell Cracking Buffer (SDS, Tris-HCl, Bromophenol Blue, Urea, DTT) should be used. **It is highly recommended that you only load 5 µL of insoluble samples to your gels as they are usually highly concentrated.** To prepare these samples 1 mL would be spun down in an Eppendorf at ~15000 rpm for 3 minutes on a bench top

centrifuge, the supernatant was removed by violently flicking your wrist, 100 µL of Cell Cracking Buffer was added and the pellet was disrupted/mixed with the pipette tip. The sample was then heated on a heat block set to 80^{o}C for 30-60 minutes. After the heating the sample should be vortexed and checked for "stringy/stickiness" by pipetting. If the sample was easy to pipette it was ready to be loaded onto a gel, if it was sticky, it required another round of heating and vortexing.

Liquid Chromatography Electrospray Ionization Mass Spectrometry (LC-ESI-MS)

The identity of pure protein samples was confirmed by LC-ESI-MS. A Phenomenex Aeris 3.6 µm WIDEPORE C4 200 Å column (100 x 2.1 mm) was attached to a Dionex Ultimate 3000 HPLC instrument and routed to an Advion Expression cms^{L} mass spectrometer. These instruments were controlled with Chromeleon CDS software and Advion Software. 80 µL of protein was diluted with 20 µL of nanopure water in an HPLC vial. Separation was achieved with the HPLC running a gradient elution with an aqueous mobile phase (95% nanopure water, 5% MeCN, 0.1% Formic Acid) and an organic phase (100% MeCN, 0.1% Formic Acid). The default method used a flow rate of 0.3 mL/min starting with 10% organic from 0.0-1.0 mins, a linear gradient of 10-90% organic from 1.0-7.0 mins, 90% organic from 7.0-9.0 mins, a linear gradient from 90-10% organic from 9.0-9.1 mins and 10% organic from 9.1-12.0 mins. The protein charge ladder was then deconvoluted using Advion Data Express software.

Liquid Chromatography Quadruple Time-of-Flight Mass Spectrometry (LC-QTOF-MS)

Depending on the instrument's availability, some of the identities of protein samples was confirmed using LC-QTOF-MS. **To become an authorized user of this instrument you must contact Sarina Kiesser in SciTech.** A Phenomenex Aeris 3.6 µm WIDEPORE C4 200 Å column (100 x 2.1 mm) was attached to a Agilent 1290 Infinity II UPLC instrument and routed to a Agilent AdvanceBio 6545XT LC/QTOF mass spectrometer. These instruments were controlled with a series of Agilent software. 80 µL of protein was diluted with 20 µL of nanopure water in an HPLC vial. Separation was achieved with the UPLC running a gradient elution with an aqueous mobile phase (95% nanopure water, 5% MeCN, 0.1% Formic Acid) and an organic phase (100% MeCN, 0.1% Formic Acid). The default method used a flow rate of 0.3 mL/min starting with 10% organic from 0.0-1.0 mins, a linear gradient of 10-90% organic from 1.0-7.0 mins, 90% organic from 7.0-9.0 mins, a linear gradient from 90-10% organic from 9.0-9.1 mins and 10% organic from 9.1-12.0 mins. The protein charge ladder was then deconvoluted using Agilent's BioConfirm software.

NMR

^{15}N labeled NMR samples of VL147N were buffer exchanged via dialysis into PIPES buffer (20 mM PIPES, 50 mM NaCl, pH 6.8) and spiked with NaN_3 (0.02% of the solution). No corrections were made for the effects that ^2H_2O have on the pH. Structural data was collected using a Bruker Avance III HD NMR spectrometer operating at 11.7 Tesla (^1H resonance frequency of 500 MHz) equipped with a room temperature broadband (^1H-X) SmartProbe (Western Washington University, Bellingham, WA, USA). ^{15}N-HSQC spectra were acquired via uniform sampling

recording with X indirect points at 10°C. Data was processed with NMRPipe software, and the spectra was analyzed via NMRViewJ software (One Moon Scientific Inc.).

HPLC/FPLC UV-Vis

Ultraviolet-visible spectroscopy was used to identify which elution fractions from gradient purifications contained the purest protein as well as analyzing the results of F-actin pull down assays. The Dionex Ultimate 3000 HPLC instrument collected the absorption at 200-900 nm, while the Biorad NGC Chromatography System by default collected the absorption at 280 nm.

Nanodrop

The concentration of protein samples was determined by measuring their absorbance at 280 nm on a Thermo Scientific Nanodrop 1000 Spectrophotometer. Finding the calculated extinction coefficients was done using Expasy ProtParam software and plugging those values into the Beer-Lambert Law.

Buffer Exchange and Protein Concentration

Desalting Columns

Disposable Econo-Pac 10DG desalting columns from BioRad were used. These are proprietary columns designed to capture and slow salts and small molecules in the beads of its resin while the larger sample proteins pass around the resin. These columns were used on all Aim 1 constructs, as it is a rapid way to perform a buffer exchange. The cap of the desalting

column was removed, and the storage buffer solution (sodium azide) was poured off into waste. 20 mL of the desired buffer (50 mM Tris, 150 mM NaCl, pH 7.6) was added to the column and the plastic tip was snapped off to allow the buffer to flow. Once the column was completely drained (it cannot run dry) 3 mL of sample was added to the column and allowed to flow through. **If you have less than 3 mL of sample, you must dilute to final volume of 3 mL.** Once the column has drained 8 mL of the desired buffer was added and the elution's were collected in a 15 mL falcon tube.

Dialysis Tubing

Dialysis tubing with a 3.5 kDa cut-off was used for all other scenarios where buffer exchange was needed. The tubing containing the sample was submerged in ~40 times the volume of the desired buffer at 4^{o}C with a stir bar for 12-16 hours. **This process can be sped up in numerous ways, one is to use significantly more of the desired buffer (100 X), another is to split your sample into multiple small volumes each in their own tubing, the other is to switch out the desired buffer multiple times. These strategies can be combined, however, to calculate exactly how long the sample must remain in the buffer, the rate of diffusion, the kDa cut-off, and surface area of the desalting tubing must be accounted for.**

Spin Concentration

Spin concentrators were used for the primary purpose of concentrating our proteins. Size cut-offs at 30 kDa, 15 kDa, 10 kDa and 3 kDa were used. **The general recommendation for spin concentrators is to select a cut-off size that is two times less than the size of the folded protein**

(10 kDa concentrator for a 22 kDa protein). The protein was added to the top of the concentrator, and placed in a benchtop centrifuge, with a counterbalance and spun at 4000 x g for 5-10 minutes at 4^{o}C until the volume was at the desired level and concentration. **Interestingly, VL147N responded best to a 10 kDa cut-off, even though it is a higher cut-off than is recommended.** In some cases, spin concentrators were used as a form of buffer exchange to lower/remove imidazole. This was achieved by spinning the protein sample down until there was 1.5 mL sample, adding 15 mL of desired buffer, and repeating 3 times to the final desired volume.

TEV Protease-Cleavage Reaction

To mitigate past reviewers' concerns over affinity tags contributing to F-actin binding, it became standard to cleave off all affinity tags prior to conducting any F-actin pull-down assays. This was achieved and optimized using Super TEV Protease (TEV protease). **An excel calculator has been set up in the Smirnov Google Drive in the Protein Database under the tab "TEV Calculation".** To maximize productivity the ratio of TEV protease to protein was optimized for a 2-hour reaction. **We found that a 14:1 protein to TEV protease was best for this 2-hour window**. The other final reaction concentrations were 0.5 mM EDTA, and 1 mM DTT. **For VL147N I discovered that the reaction worked best when incubated at room temperature with gently shaking at ~60 rpm.** Immediately following the 2-hour window the mixture was subjected to Ni-NTA purification with 0.5 mL of resin as a CV. The desired product will be found in the Flowthrough and Washes. The Ni-NTA purification removed any remaining unreacted reactants and undesired contaminants.

Sortase Mediated Ligation

All Aim 1 constructs were subjected to a Sortase Mediated Ligation (SML) under several conditions, to assist in developing a robust method of ligations. This included comparing WTsrtA, 7mutA, and the effects of adding $NiSO_4$.

SML with WTsrtA

The parameters of this reaction were to mix 50 µM of the protein with 250 µM of the ligation partner peptide and 20 µM of WTsrtA. All ligation reactions ran for 8 hours at room temperature. The protein stocks were in Tris Lysis Buffer and were spiked with 1 mM TCEP and 10 mM Ca^{2+}. The total reaction volume was 100 µL. Every hour on the hour for eight hours a 5 µL injection at room temperature was auto injected by the Dionex Ultimate 3000 HPLC instrument into a C4 column attached to the Advion Expression cms^{L} mass spectrometer (LC-ESI-MS). The method was titled "C8 30-90 (30 min) 250 µL flow (6min double ramp)" on the Chromeleon CDS Software. The data points were deconvoluted using Advion Software. Each SML reaction was done in triplicate.

SML with 7mutA

The parameters of this reaction were to mix 50 µM of the protein with 250 µM of the ligation partner peptide and 10 µM of 7mutA. All ligation reactions ran for 8 hours at room temperature. The protein stocks were in Tris Lysis Buffer and were spiked with 1 mM TCEP. The total reaction volume was 100 µL. Every hour on the hour for eight hours a 5 µL injection at room temperature was auto injected by the Dionex Ultimate 3000 HPLC instrument into a C4 column

attached to the Advion Expression cms^{L} mass spectrometer (LC-ESI-MS). The method was titled "C8 30-90 (30 min) 250 µL flow (6min double ramp)" on the Chromeleon CDS Software. The data points were deconvoluted using Advion Software. Each SML reaction was done in triplicate.

SML with 7mutA + $NiSO_4$

The parameters of this reaction were to mix 50 µM of the protein with 250 µM of the ligation partner peptide and 10 µM of 7mutA. All ligation reactions ran for 8 hours at room temperature. The protein stocks were in Tris Lysis Buffer and were spiked with 1 mM TCEP and 100 µM $NiSO_4$. The total reaction volume was 100 µL. Every hour on the hour for eight hours a 5 µL injection at room temperature was auto injected by the Dionex Ultimate 3000 HPLC instrument into a C4 column attached to the Advion Expression cms^{L} mass spectrometer (LC-ESI-MS). The method was titled "C8 30-90 (30 min) 250 µL flow (6min double ramp)" on the Chromeleon CDS Software. The data points were deconvoluted using Advion Software. Each SML reaction was done in triplicate.

F-actin Assay and Calibration Curve

Calibration Curve

Identify at least 7 concentrations that you wish to use in your actin binding curve. Prepare 50 µL of each concentration in an HPLC snap cap tube and run the samples on the Dionex Ultimate 3000 HPLC instrument with the Phenomenex Aeris 3.6 µm WIDEPORE C4 200 Å column (100 x 2.1 mm) with blanks between each run. If the retention time for the protein matches that of the

mass spectrophotometer data, run the samples twice more to complete your curve. If the retention time does not match, troubleshooting with a Phenomenex Luna 5U C18 (150 x 4.6 mm) column is advised.

For VL147N the final concentrations of 2, 3, 5, 7, 14, 40, 70, and 80 µM were selected as such stocks at 10, 15, 25, 35, 70, 200, 350, and 400 µM of VL147N was needed. 10 µL of each stock concentration was diluted with 40 µL of nanopure water to the final concentrations and ran in triplicates through the Dionex Ultimate 3000 HPLC instrument with the Phenomenex Aeris 3.6 µm WIDEPORE C4 200 Å column (100 x 2.1 mm) with blanks between each sample. The method used was "C4 RP-HPLC DAD analytical ultra 10-90 300 µL flow (15min-Actin)".

Preliminary F-actin Spin-down Assay

To investigate the capacity of the villin constructs to bind actin filaments an Actin Binding Protein Spin-Down Assay Biochem Kit™: Muscle Actin produced by Cytoskeleton Inc. was used. Frozen stocks of Bovine Serum (BSA), α-actinin, Actin Polymerization Buffer were rapidly defrosted via water bath and then immediately placed on ice. Lyophilized rabbit skeletal actin was resuspended with 250 µL of General Actin Buffer and equilibrated at 4°C for 30 minutes. The actin was then spiked with 30 µL Actin Polymerization Buffer forming F-actin Stock (17.5 µM). The mixture was equilibrated at room temperature for one hour. While this equilibrated the of 15 µL BSA and 55 µL of the test protein were clarified via centrifugation at 150,000 x g in a Sorvall MX 150 micro-ultra centrifuge **(MAKE SURE THERE ARE TWO RUBBER O-RINGS, ONE SMALL AND ONE LARGE IN THE ROTOR FORMING AN AIRTIGHT SEAL)** for one hour at 4°C with counterbalances. Ten micro-ultra centrifuge tubes were then prepared following Table 2.3

When the test protein was added to ensure immersion, the mixture was pipetted up and down slowly three times. The micro-ultra centrifuge tubes were then incubated at room temperature for 30 minutes prior to centrifugation at 150,000 x g for 1.5 hours at 24^{o}C **(MAKE SURE THERE ARE TWO RUBBER O-RINGS, ONE SMALL AND ONE LARGE IN THE ROTOR FORMING AN AIRTIGHT SEAL)**. The supernatants were removed carefully as to not disturb the pellets that formed and stored in labeled individual Eppendorf's. **You will want to record the volume of the supernatant collected, especially for tubes 3, 4, 5, 6-9. If the supernatant volume is less than 45 µL, double check that the small and large O-rings in the rotor are properly seated and inform Dr. Serge L. Smirnov, prior to repeating this assay or any subsequent trials.** The pellets were gently washed with 70 µL F-actin Buffer (630 µL General Actin Buffer and 70 µL of Actin Polymerization Buffer). The wash was discarded into waste. The pellets were then resuspended with 50 µL of 7% acetic acid, mixed and extracted into labeled individual Eppendorf's.

Tube 1	40 µL F-actin Stock, 10 µL General Actin Buffer
Tube 2	40 µL F-actin Buffer, 10 µL α-actinin, 2 µL 1M Tris-HCl pH 6.5
Tube 3	40 µL F-actin Stock, 10 µL α-actinin, 2 µL 1M Tris-HCl pH 6.5
Tube 4	40 µL F-actin Stock, 10 µL BSA, 2 µL nanopure water
Tube 5	40 µL F-actin Buffer, 10 µL Test Protein
Tube 6	40 µL F-actin Stock, 10 µL Test Protein
Tube 7	40 µL F-actin Stock, 10 µL Test Protein
Tube 8	40 µL F-actin Stock, 10 µL Test Protein
Tube 9	40 µL F-actin Stock, 10 µL Test Protein (1/2 concentration)
Tube 10	50 µL nanopure water

Table 2.3 Contents of each micro-ultra centrifuge tube for preliminary F-actin Spin-down Assay.

To draw a conclusion as to if the test protein bound actin filaments an 15% SDS-PAGE was used. **I recommend taking 10 µL from both the supernatants and pellets Eppendorf's corresponding to micro-ultra centrifuge tubes 3, 4, 5, and 6 and mixing each in 10 µL of 4X stain.**

This will show all the controls necessary to draw a conclusion. You can also take the pellet and supernatant samples from tubes 7-9 and run them through the HPLC to have extra data for a binding curve if applicable.

F-actin Spin-down Assays for Actin Binding Curve

To investigate the capacity of the villin constructs to bind actin filaments an Actin Binding Protein Spin-Down Assay Biochem Kit™: Muscle Actin produced by Cytoskeleton Inc. was used. Three aliquots of Actin Polymerization Buffer was rapidly defrosted via water bath and then immediately placed on ice. Two aliquots of Lyophilized rabbit skeletal actin were resuspended, each with 250 µL of General Actin Buffer and equilibrated at 4°C for 30 minutes. Each aliquot of actin was then spiked with 30 µL Actin Polymerization Buffer forming F-actin Stock (17.5 µM). The mixture was equilibrated at room temperature for one hour. While this equilibrated 40 µL of the test protein at each stock concentration was clarified via centrifugation at 150,000 x g in a Sorvall MX 150 micro-ultra centrifuge **(MAKE SURE THERE ARE TWO RUBBER O-RINGS, ONE SMALL AND ONE LARGE IN THE ROTOR FORMING AN AIRTIGHT SEAL)** for one hour at 4°C with counterbalances. 14 micro-ultra centrifuge tubes were then prepared with 40 µL of F-actin Stock and 10 µL of clarified stock protein at each concentration. **There should be duplicates of each concentration in this first run.** When the test protein was added to ensure immersion, the mixture was pipetted up and down slowly three times. The micro-ultra centrifuge tubes were then incubated at room temperature for 30 minutes prior to centrifugation at 150,000 x g for 1.5 hours at 24°C **(MAKE SURE THERE ARE TWO RUBBER O-RINGS, ONE SMALL AND ONE LARGE IN THE ROTOR FORMING AN AIRTIGHT SEAL)**. After the centrifugation a third aliquot of Lyophilized

rabbit skeletal actin was resuspended with 250 µL of General Actin Buffer and equilibrated at 4°C for 30 minutes. While it was equilibrating the supernatants were removed carefully as to not disturb the pellets that formed and stored in labeled individual HPLC snap cap vials. **Record the volume of each supernatant.** The pellets were gently washed with 70 µL F-actin Buffer (630 µL General Actin Buffer and 70 µL of Actin Polymerization Buffer). The wash was discarded into waste. The pellets were then resuspended with 50 µL of 7% acetic acid, mixed and extracted into labeled individual HPLC snap cap vials. Eight final micro-ultra centrifuge tubes were prepared, one with 50 µL of nanopure water, the rest with 40 µL of F-actin Stock and 10 µL of clarified stock protein at each concentration. The mixture was pipetted up and down slowly three times to insure immersion. The micro-ultra centrifuge tubes were then incubated at room temperature for 30 minutes prior to centrifugation at 150,000 x g for 1.5 hours at 24°C **(MAKE SURE THERE ARE TWO RUBBER O-RINGS, ONE SMALL AND ONE LARGE IN THE ROTOR FORMING AN AIRTIGHT SEAL)**. The original batch of HPLC snap cap vials are placed into the Dionex Ultimate 3000 HPLC instrument with the Phenomenex Aeris 3.6 µm WIDEPORE C4 200 Å column (100 x 2.1 mm) and run with blanks between each sample. The method used was called "C4 RP-HPLC DAD analytical ultra 10-90 300 µL flow (15min-Actin)". The final set of micro-ultra centrifuge tubes were subjected to the same process and the HPLC snap caps for the pellet and supernatants were added to the HPLC list and run using the same method.

Note that this many samples will occupy the HPLC for almost 48 hours, you must make sure that you have reserved enough time, that the buffers are sufficiently full, and that waste level is not exceeded. I suggest running this over a weekend and checking in on it every 12 hours.

Construction of the Binding Curve

Using methods adapted from previous WWU graduate student Heather Miears master's thesis the binding curve was created. The system includes an equilibrium dissociation constant (K_d), a maximal binding constant (B_{max}), and a non-specific parameter (*NS*), which includes any trapping or non-specific interactions with the actin filaments. The ligand-receptor interaction formula is as follows:

$$[P_{bound}] = \frac{B_{max}\,([P_{total}] - [P_{bound}])}{K_d + ([P_{total}] - [P_{bound}])} + NS([P_{total}] - [P_{bound}])$$

To fit the binding data, the following function was used in the nonlinear regression protocol of the Origin 2016 software to determine the K_d, B_{max} and the percentage of NS:

$$[P_{bound}] = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \text{ where}$$

a = $1 - NS$,

b = $K_d + NS(K_d) + [P_{total}] + 2NS[P_{total}] + B_{max}$

c = -$[P_{total}](NS(K_d) + NS[P_{total}] + B_{max})$

Arabidopsis Thaliana Extract Assay

Preparation of Plant Cell Extract

A clean mortar and pestle were chilled in a freezer set to -20°C for 2 hours. **This can be expedited by pouring liquid nitrogen into the mortar and letting it evaporate naturally 2 times.**

Let the mortar and pestle sit in the -20°C freezer for 5-10 minutes to warm back up before continuing. Frozen A. *thaliana* plant material weighed out and placed into the chilled mortar and ground with room temperature PIPES buffer (20 mM PIPES, 50 mM NaCl, pH 6.8). For every 1.0 gram of plant material 1.0 mL of PIPES should be used. Once the plant material became a homogeneous liquid, it was centrifuged at 40000 RCF for 30 minutes at 4°C. The supernatant was then decanted and filtered with a 0.22 µm filter. The filtered supernatant was aliquoted into Eppendorf's with 300 µL in each. This plant cell extract was always stored at -20°C and no aliquot was used more than once.

Exposure to Plant Cell Extract

VL192 and VL147N were subjected to A. *thaliana* plant cell extract and monitored for changes via 15% SDS-PAGE and LC-QTOF-MS. For all assays conducted, 7 µL of ~150 µM protein was mixed with 10 µL of A. *thaliana* plant cell extract with verifying amounts of PMSF. The usual time frame of exposure was 0 - 1.5 hours all conducted at room temperature. In each assay two lanes were dedicated as controls, one for the protein and one for the plant cell extract. These controls were diluted to match the relative intensity of the bands in the rest of the assay.

Some exposure assays included repurification of VL192 after exposure to A. *thaliana* plant cell extract using protocols as described in "Ni-NTA Purification Gravity Column". However, with a smaller volume 100 µL of resin and similarly small volume for washes and elutions were used. **If the use of PMSF was used in the assay, PMSF should also be added to the wash and elution buffer. I found that a final concentration of 10 mM was sufficient and did not seem to adversely**

affect anything. Timing was critical in these assays, and I strongly recommend pre-tuning the QTOF and prepping your worklist so that the moment you collect your elutions you can transfer whatever samples you desire into HPLC snap caps insert the samples into the QTOF and start the run.

Chapter III Results and Discussion

Aim 1

Characterize the relative reactivity of sortase mediated ligation sites within authentic intrinsically disordered region sequences and create guidelines for optimization of reaction conditions.

Preparation of Original Constructs

The original constructs (**Table 2.1**) were prepared in the following order of operations: transformation via heat shock, expression, purification using Ni-NTA gravity flow columns, and buffer exchange into Tris Lysis Buffer using desalting columns (see relevant procedures). There were two such instances where further purification was required, both AroN and BasN needed gradient Ni-NTA purifications using the FPLC to generate pure samples.

Results of SML on Original Constructs

Originally it was our intent to test all 13 original constructs with WTsrtA, however, it was quickly determined that we did not have the bandwidth to optimize all of these constructs. We decided to focus on the N-terminal constructs, as the influence of these residues had not been thoroughly explored in the literature. The result of our work is summarized in **Figure 3.1.** Each ligation was run at least three times and analyzed by past and present WWU students in the Antos Lab. The WTsrtA runs were conducted by WWU undergraduates Micah Lund and Liam Quille. The 7mutA runs were conducted by WWU undergraduates, Emma Nusbaum, Micah Lund,

Liam Quille, Miriam Gold, and Noah Goodwin-Rice. The $NiSO_4$ runs were conducted by WWU undergraduates Miriam Gold, Leo Salcedo, Mikale Llaneza, and Noah Goodwin-Rice.

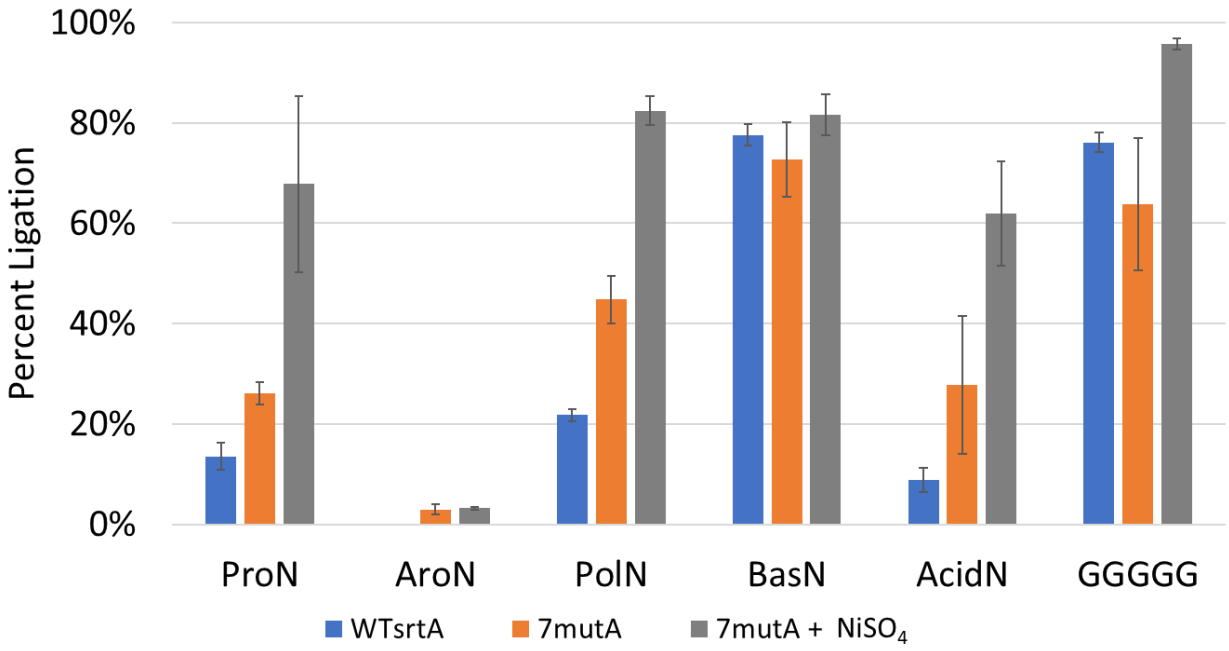


Figure 3.1 Percent Ligation of the Original N-terminal Constructs. Depicted is a summary of the ligation efficiency measured in percentage of the original N-terminal constructs using the three different methods of SML. Percent ligation for WTsrtA and 7mutA runs were monitored for 8 hours and data displayed corresponds to the 8 hours mark. Percent ligation for 7mutA + $NiSO_4$ runs were monitored for 12 hours but data displayed corresponds to hour mark 11, 11, 2, 4, 12, and 4 respectively.

We found a significant improvement in ligation efficacy with 7mutA, a sortase A mutant which was expected to increase the speed and yield SML, however it was noted by that many constructs suffered from hydrolysis and intramolecular interactions.[37] By implementing $NiSO_4$ we hoped to drive the reaction in the forward direction by deactivating the loose $GGHis_6$ which was produced during the cleavage of the ligation (**Figure 3.2**).[38] For all but AroN we saw a positive impact when the $NiSO_4$ additive was included.

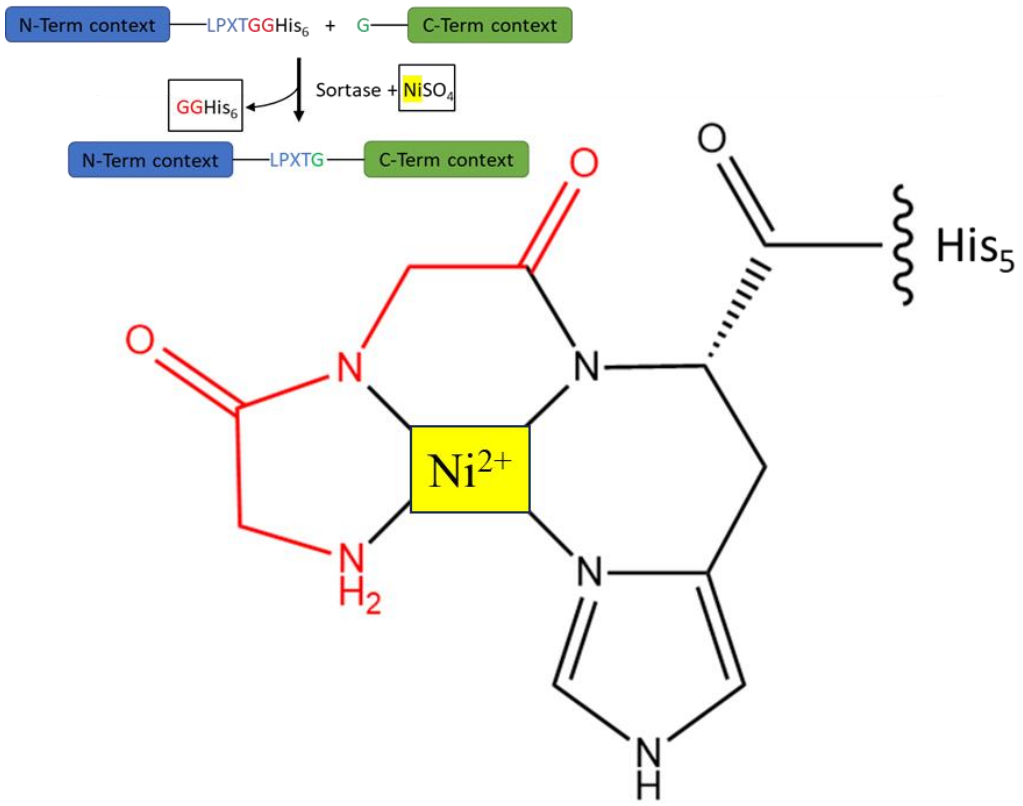


Figure 3.2 Complex formed by $NiSO_4$ and $GGHis_6$. Depicted is the complex which is presumably formed by Ni^{2+} when the $GGHis_6$ is cleaved by the sortase enzyme during the ligation process. In red are the glycine residues, in black are the histidine's and in yellow is the Ni^{2+} from the $NiSO_4$.

Preparation of Additional Glycine Constructs

The additional glycine constructs (**Table 2.2**) were prepared in the following order of operation: transformation via heat shock, expression, purification using Ni-NTA gravity flow columns, and dialysis into Tris Lysis Buffer using desalting columns (see relevant procedures). AroNG and AroN2G both required gradient Ni-NTA purifications using the FPLC to generate pure samples. The final purity of each additional glycine construct can be seen in **Figure 3.3**.

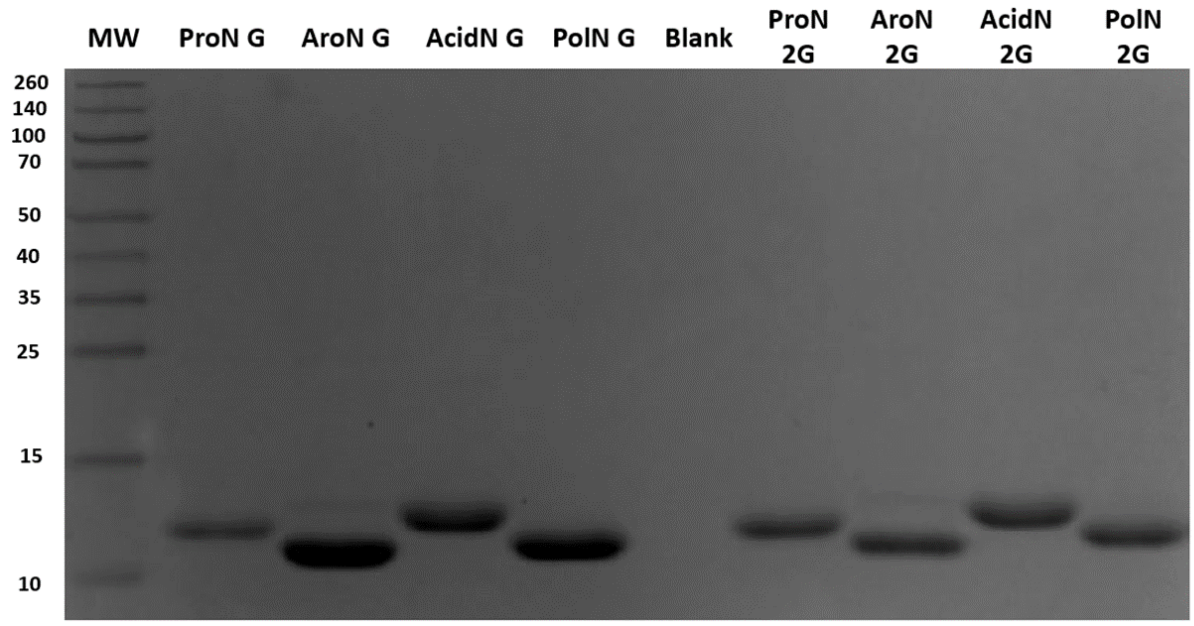


Figure 3.3 15% SDS-PAGE of Additional Glycine Constructs.

Results of SML on Additional Glycine Constructs

AroNG, AroN 2G, AcidNG, and AcidN 2G are the only constructs which have been studied at the time of this publication, some of which have yet to be completed in triplicate. The duration of these SMLs varied from construct to construct, however most were monitored for 12 hours. The actual SML trials and analysis of the data was performed by WWU undergraduate students Leo Salcedo, Noah Goodwin-Rice, and Mikale Llaneza.

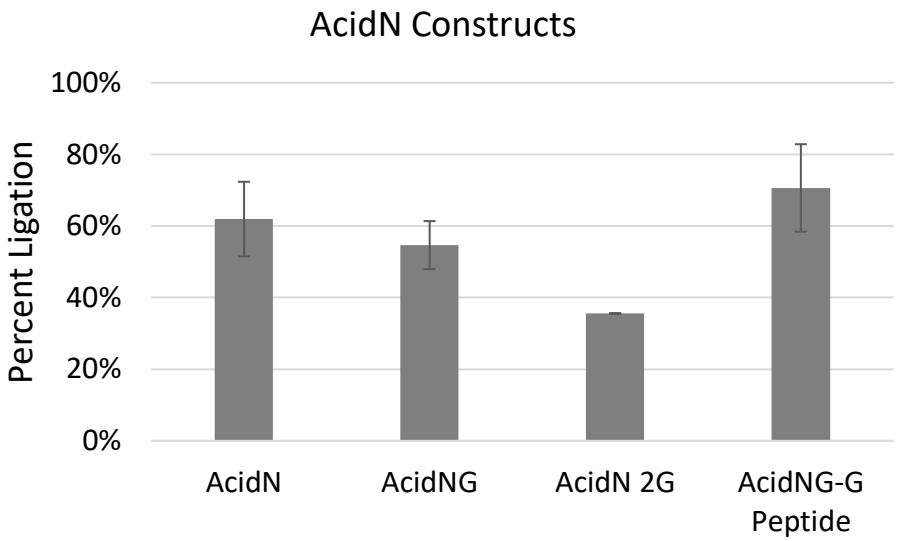


Figure 3.4 7mutA with $NiSO_4$ Maximum Percent Ligation of AcidN Constructs. Depicted is a summary of the ligation efficacy measured in percentage of the AcidN constructs at the 12-hour mark.

The efficacy of SML for AcidN seemed to rely on the identity of the C-terminal context (peptide partner), however as this dataset is not complete, we cannot draw any conclusions. Such a claim could only be made after more trials with AcidN 2G with its original peptide partner and trials with AcidN 2G with its additional glycine peptide partner.

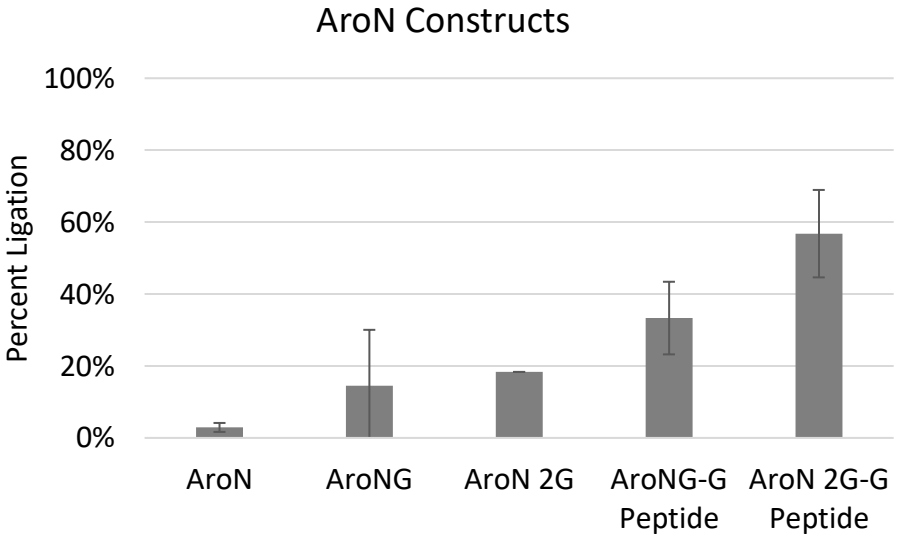


Figure 3.5 7mutA with $NiSO_4$ Maximum Percent Ligation of AroN Constructs. Depicted is a summary of the ligation efficacy measured in percentage of the AroN constructs at varying timepoints.

AroN, AroNG, AroNG-G Peptide, and AroN 2G-G Peptide were all monitored for 12 hours, while the one trial conducted on AroN 2G was only monitored for 10 hours. The use of the additional glycine peptide partners has a massive positive impact with AroN constructs. We see that with the additional glycine peptide partner the ligation efficacy for AroNG doubles and the ligation efficacy for AroN 2G almost triples. To complete this data set two more trials need to be conducted with AroN2G and the original peptide partner.

Aim 1 Discussion

This was a collaboration between past and present students from the Smirnov and Antos labs. Together we have made significant strides in documenting the reactivity of sortase mediated ligation sites within authentic IDR segments and have begun creating guidelines for the optimization of reaction conditions. The introduction of $NiSO_4$ had a significant impact as to maximal ligation level achieved. However, over time almost all constructs suffered from intramolecular interaction and hydrolysis, which in turn decreased the percentage of ligation product.

Furthermore, through substitution (additional glycine constructs) we see adding spacers in the form of a glycine increases the overall efficacy of the SML reactions. We hypothesize that by performing these substitutions we are making it easier for the sortase enzyme to bind the LPXTG site and perform its ligation.

Aim 2

Probe backbone structure and dynamics and generate an actin binding curve of plant protein Villin4 linker, to quantify the specific and nonspecific binding capacity of filamentous actin.

Preparation of VL147N

VL147N was prepared in the following order of operations: transformation via heatshock, expression using 0.4 mM IPTG, a purification using Ni-NTA gravity flow column, a purification using Strep gravity column, a TEV cleavage reaction at a 14:1 protein to TEV ratio for 2 hours at room temperature, a third and final purification using Ni-NTA gravity flow column, and spin concentration with a 10 kDA cutoff (see relevant procedures). This construct should always be stored at 4 ^{o}C and once the TEV reaction occurs, used within 5-7 days. After 5-7 days the construct begins to degrade rapidly.

To complete the F-actin binding curve, 0.5 mL at 400 µM of VL147N was required. As with most sizable IDRs, to generate this amount of pure protein was a demanding task. A three-liter growth was used to achieve this concentration. VL147N required extensive purification, utilizing both of its purification tags and often it was not until after the TEV cleavage and a third Ni-NTA purification in which a pure product was present. If this protein had not been such a large IDR SEC may have been a very useful tool. The purity of the sample was confirmed with LC-QTOF-MS and 15% SDS-PAGE (**Figure 3.6**).

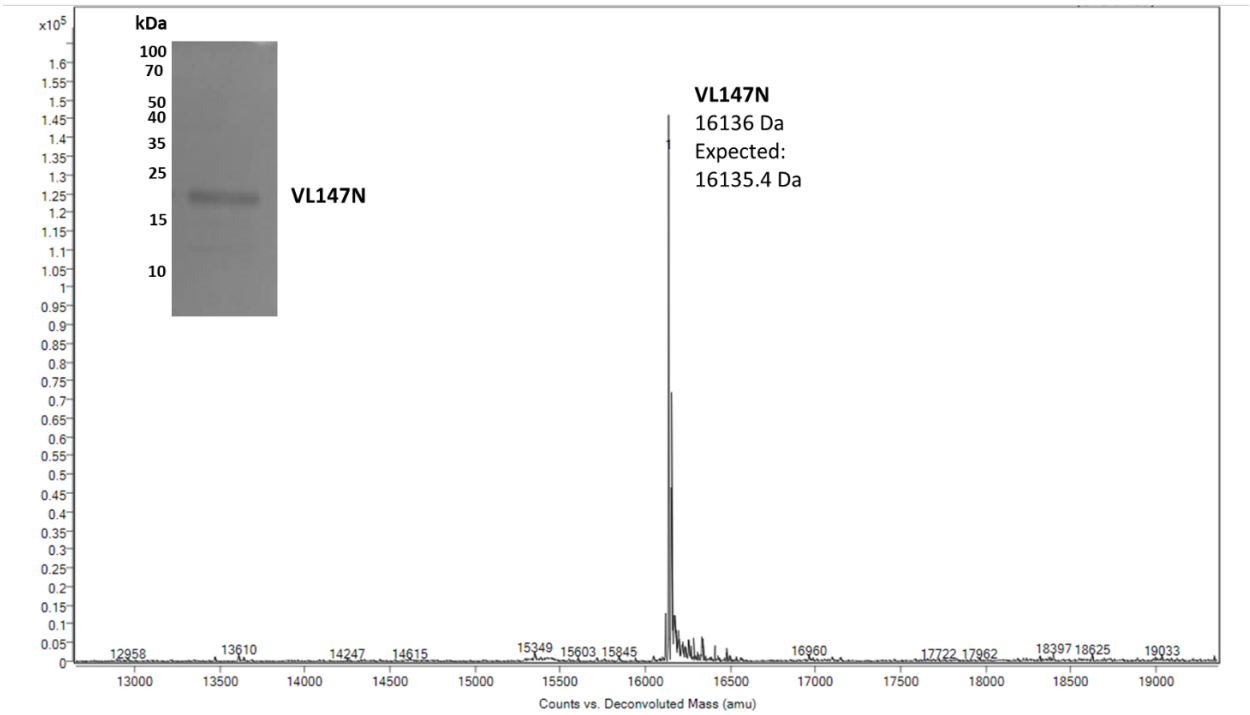


Figure 3.6 Relative purity of VL147N sample and the deconvoluted spectra of VL147N.

Preparation of ^{15}N VL147N

Uniformly labeled ^{15}N VL147N was prepared in the following order of operations: transformation via heatshock, protein expression, pellet washing via M9T minimal media, resuspension in nutrient rich M9T media, induction at 0.8 mM IPTG, extraction via solubilizing buffer, purification using Ni-NTA gravity column, purification using Strep gravity column, TEV cleavage reaction at a 14:1 protein to TEV ratio for 2 hours at room temperature, a third purification using Ni-NTA gravity flow column, spin concentration with a 10 kDA cutoff, and a buffer exchange into PIPES buffer using dialysis tubing (see relevant procedures).

For the ^{15}N-HSQC NMR recording the sample needed to be uniformly labeled at concentrations exceeding 100 µM, a high value for most large, labeled IDRs. A six-liter growth, with one-liter of M9T minimal media, and one-liter of nutrient rich M9T media was used. Though

this sample required multiple purifications, using both tags, the final purification after TEV cleavage was primarily used to remove the TEV protease and the un-cleaved product. The purity of the sample was confirmed with LC-QTOF-MS and 15% SDS-PAGE (**Figure 3.7**). The spectra indicated that ^{15}N labeling efficiency was 68%. This was estimated by dividing the mass difference between the measured mass of ^{15}N VL147N and VL147N and the mass difference between the expected mass of ^{15}N VL147N and VL147N $\frac{(16286-16136)}{(16356-16136)} \times 100 = 68.2\%$.

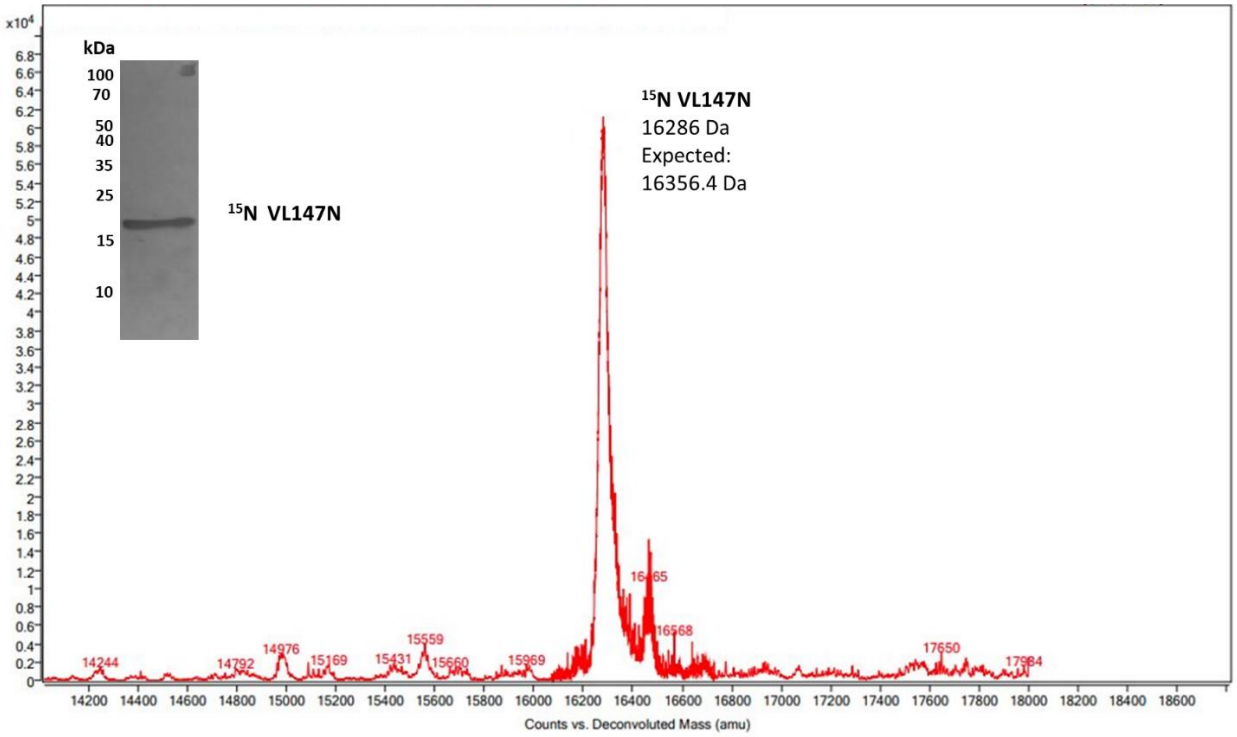


Figure 3.7 Relative purity of ^{15}N VL147N sample and the deconvoluted spectra of ^{15}N VL147N.

HSQC of ^{15}N VL147N

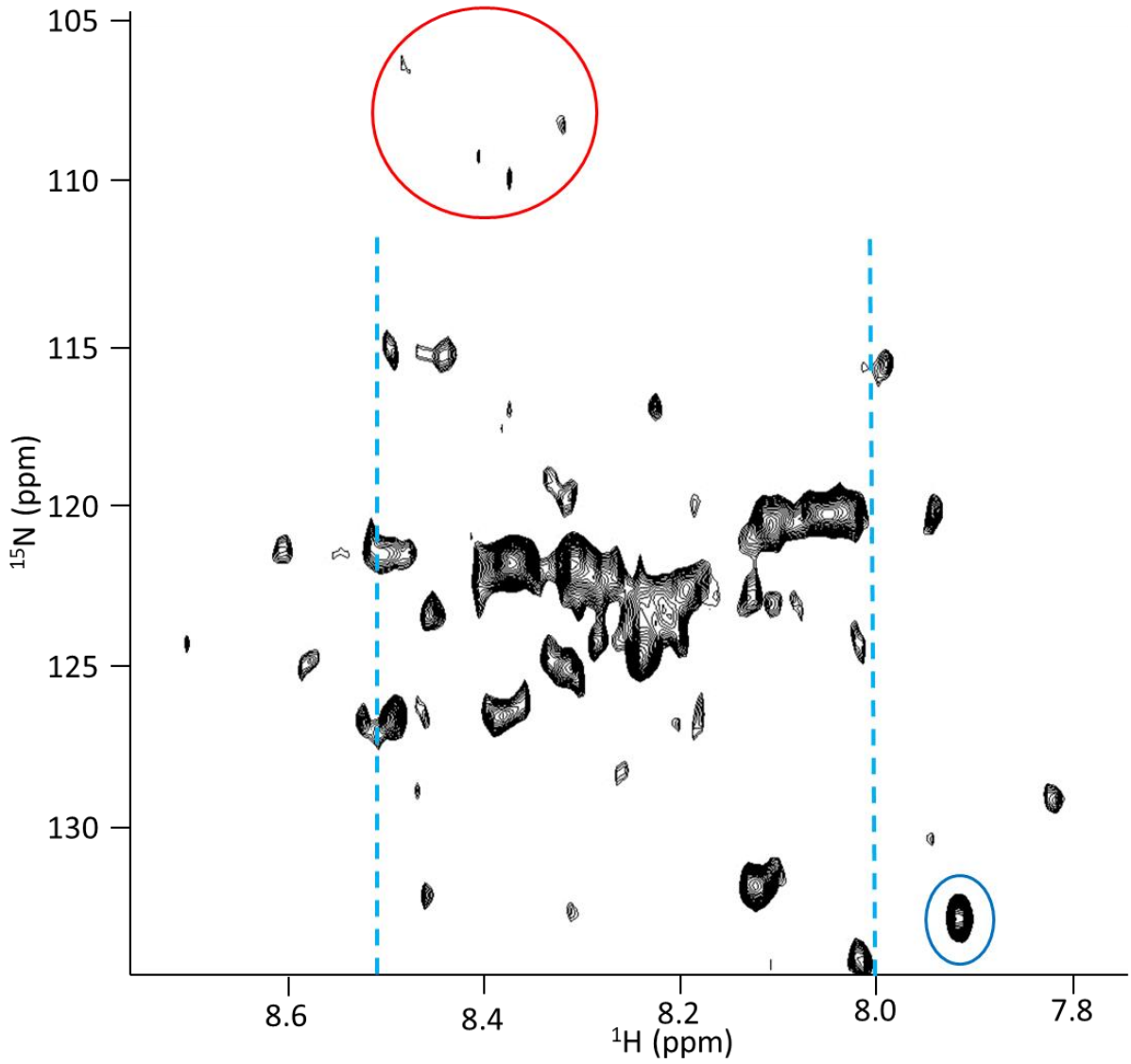


Figure 3.8 HSQC of ^{15}N VL147N. Four glycine residues are circled in red, a peak attributed to imidazole is circled in blue, and the region often attributed to intrinsically disordered proteins are bracketed in the dashed light blue lines. Concentration was 331 µM at 10 °C in PIPES buffer with 0.2% NaN_3 at pH 6.8. The HSQC was taken on a 500 MHz Bruker Avance III HD spectrometer equipped with a temperature-controlled probe.

Based on the ^{15}N-HSQC spectrum, it appeared that VL147N is a mostly disordered highly dynamic IDR, with non-uniform backbone dynamic rates. This is concluded as most of the peaks

are within 8.0-8.5 ppm, which is the region where peaks correlating to IDRs generally reside. In this region we can clearly see that peak intensity and width differ by up to a factor of ten. We can observe what appears to be four glycine residues clustered around (8.4, 108), as well as a handful of peaks outside the 8.0-8.5 ppm range which are often indicative of some sort of folded region. We also identified that the peak at (7.9, 133) is a peak from imidazole, a buffer reagent that was not completely removed during dialysis. The full spectrum from this recording can be found in appendix IV.

Constructing the Calibration Curve

A stock of VL147N was concentrated to 400 µM and dilutions at 350, 200, 70, 35, 25, 15, and 10 µM were prepared. These stocks were then further diluted and run on the HPLC in triplicate as described in "Calibration Curve". The following equation emerged Y = 1.6966X – 2.1944 with an R^2 value of 0.9998, where Y is the area under the peak (210 nm) and X is the concentration of VL147N (**Figure 3.9**).

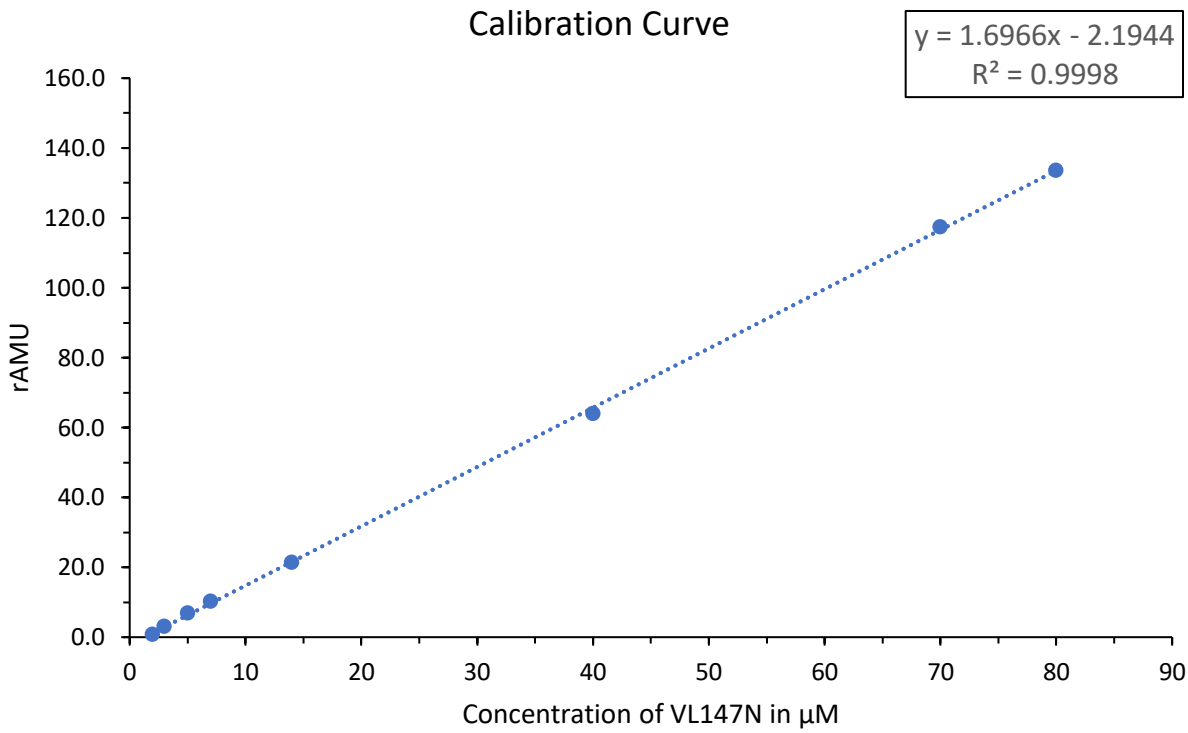


Figure 3.9 Calibration curve for VL147N. Trendline and R^2 value were calculated with Excel. The R^2 value indicates a high degree of reliability of the calibration data.

Actin Binding for VL147N and VL192

VL147N and VL192 were shown to bind filamentous actin using standard pull-down assays and 15% SDS-PAGE (see relevant procedures). However, these gels only provide qualitative evidence as to their ability to bind F-actin. VL147N was elected to continue onto further studies as we could remove the purification tags from the construct.

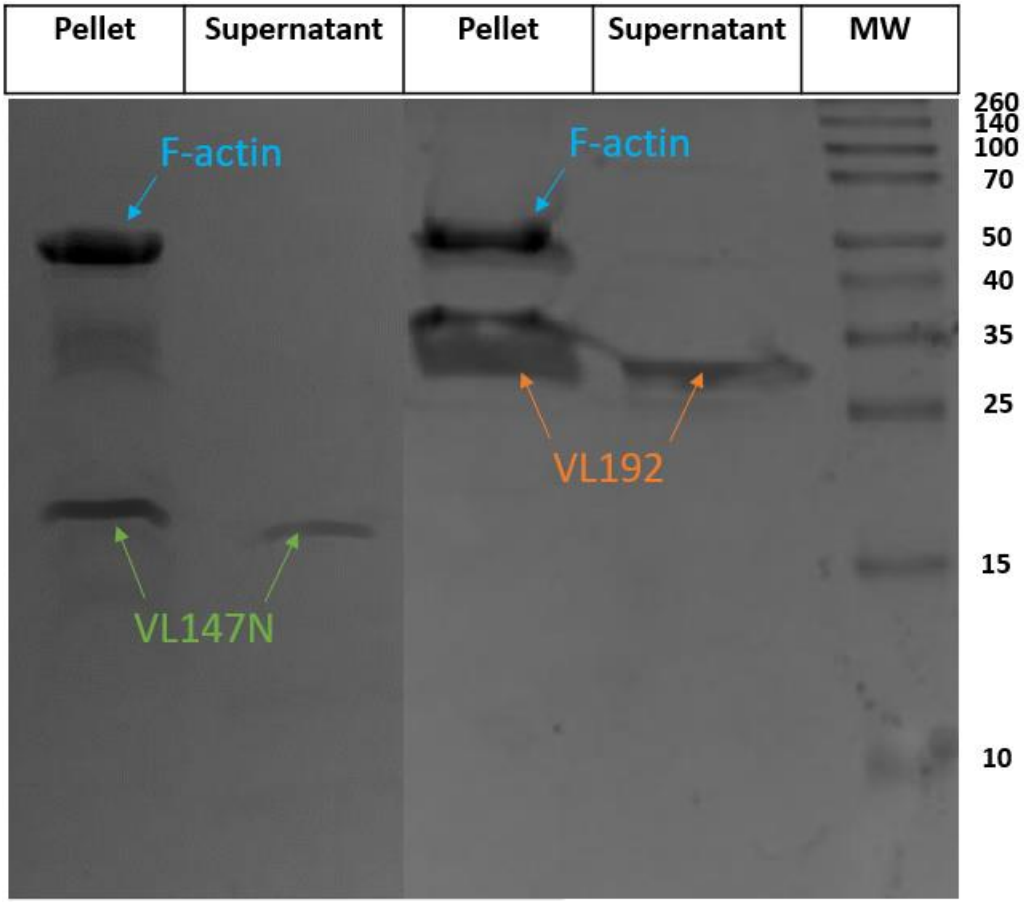


Figure 3.10 15% SDS-PAGE's of VL147N and VL192 binding F-actin. These are two different gels showing that both VL147N and VL192 bind F-actin. Confirmation of binding exists when the protein and F-actin are found in the "pellet" lane. The full gels of each binding assays can be found in Appendix III.

Actin Binding Curve for VL147N

Using methods as described in "F-actin Spin-down Assays for Actin Binding Curve" we employed standard pull-down assays to determine the binding affinity of VL147N. The binding curve was fit to a hyperbolic model, which indicated specific binding to F-actin (**Figure 3.11**). The dissociation constant (K_d) was determined to be <0 µM, indicating a sub micromolar range of affinity, however, to confirm more trials at sub 2 µM VL147N would be necessary. The non-

specific binding (NS) was measured to be 12.2%. The saturation binding concentration (B_{max}) was 13.70 µM.

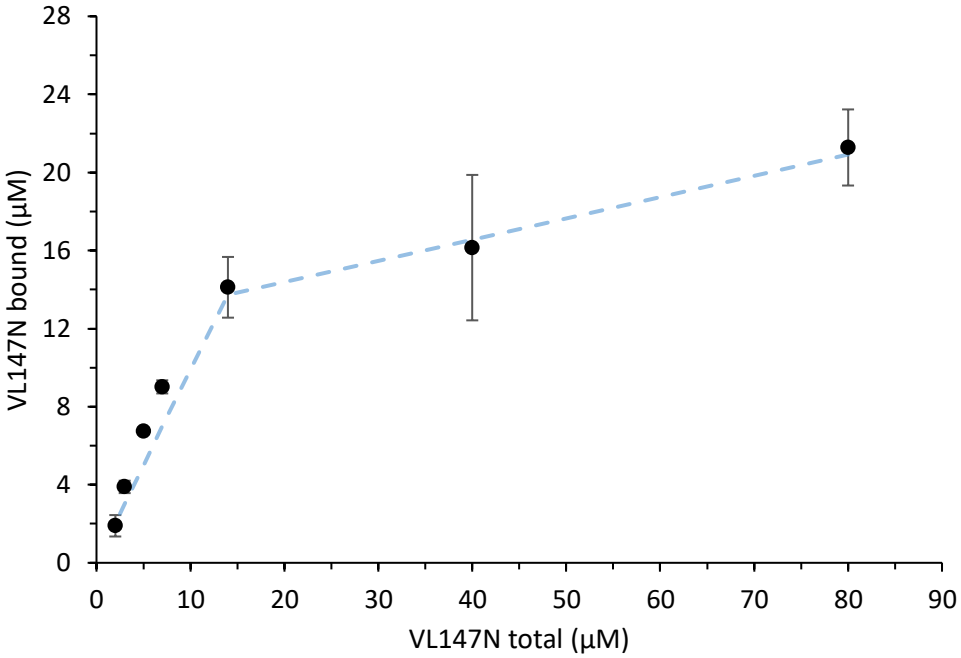


Figure 3.11 Actin Binding Curve of VL147N. Summary of the triplicate of pull-down assays used to determine VL147N's affinity to bind F-actin. Fitting and plotting of the trendline was done with Excel.

A mechanical issue with the rotor used to perform the pull-down assays was discovered after the completion of these trials and renders this data largely unpublishable. To summarize, these pull-down assays are conducted using a micro-ultra centrifuge, spinning at 150,000 x g in a vacuum. To keep the water in your samples from evaporating, two rubber O-rings should exist in the rotor to form an airtight seal (appendix III). Unbeknownst to myself and the others who operate this instrument, one of these O-rings was missing for an extended period of time, in which I conducted my binding assays. Unfortunately, by the time I noticed the rotor had been compromised, it was too late to remake more VL147N and complete the binding trials in triplicate.

Aim 2 Discussion

Full-length plant protein Villin4 was shown to bind F-actin with a K_d of 0.24 $\pm$ 0.05 μM.[29] atVHP76, the plant protein Villin4 construct which previous WWU graduate student Heather Miears studied had a K_d of 3.0 $\pm$ 0.7 μM.[30] We also confirmed that VL192 binds F-actin, however no curve was established based on previous reviewer's comments regarding the ability for purification tags to bind or bundle F-actin. Though we found VL147N to have a K_d of <0 μM, indicating sub micromolar binding affinity, further studies need to be completed to confirm these findings. At the lower concentrations of the binding curve, specifically 3 and 2 μM it was difficult to quantify the amount of protein bound. A HPLC with a better UV detector would most likely help rectify this, though I believe it would be worth looking into alternative methods by which one can measure the amount of F-actin a protein can bind.

VL147N is from the basic side of Villin4's IDR and has large positively charged regions. In the headpiece region of villin variants, a pattern emerged that indicates strong specific binding of F-actin. Specifically, they have a hydrophobic cap, an alternating charged "crown" below the cap and a positively charged patch below the crown.[30] Currently, the other villin variant linkers have not been studied in enough detail to identify a pattern that would suggest strong specific binding of F-actin. Instead of studying these other variants, however, I would suggest examining smaller regions of VL147N and comparing their relative binding affinity. Because we already have a tried-and-true method of producing VL147N, I would like to think that more time could be saved by adapting the methods described in this thesis, instead of developing new methods of producing other large IDRs. I would also suggest that the acidic side of Villin4's linker should be studied briefly to confirm that no binding occurs.

Aim 3

Probe predicted regions of plant protein Villin4 linker for specific covalent modifications in plant extract using bioinformatic mass spectrometry

Preparation of VL192

A two-liter growth of VL192 was prepared in the following order of operations: transformation via heatshock, over-night expression for 16 hours at 20 ^{o}C using 0.8 mM IPTG, a purification using Ni-NTA gravity flow column, and a purification using Strep gravity column. This construct should primarily be stored at 4 ^{o}C, however it can be stored at -20 or -80 ^{o}C for extended periods of time. The stock that I prepared has lasted with minor degradation for two years and been used in various studies besides my own.

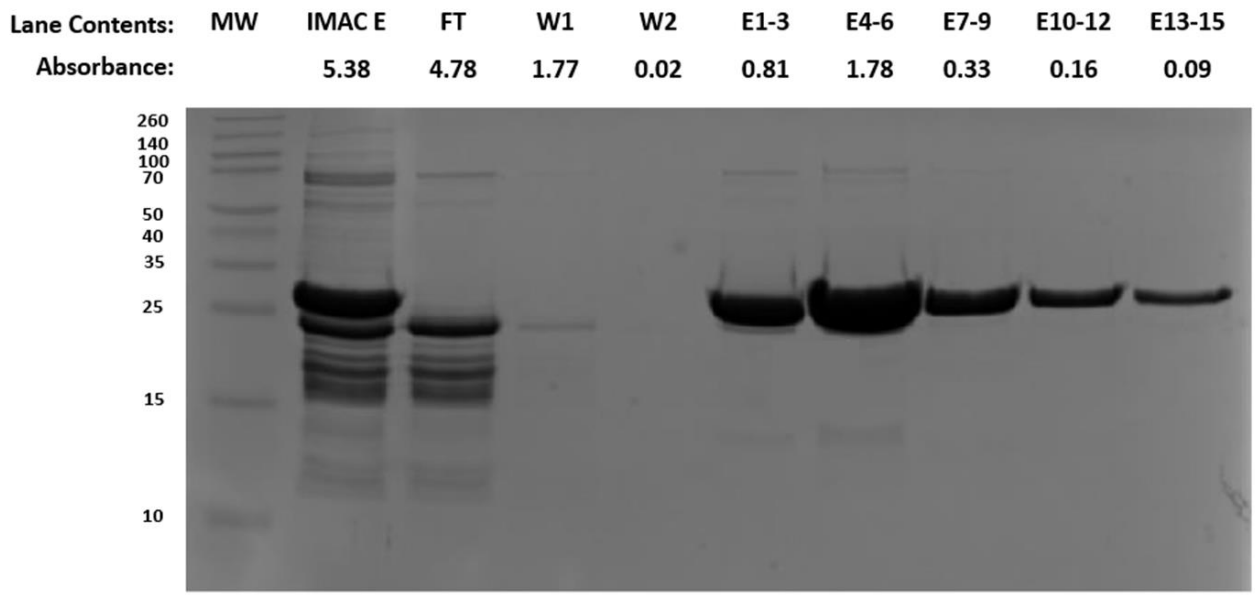


Figure 3.12 15% SDS-PAGE of the Strep purification performed on VL192. IMAC E is a sample of VL192 prior to Strep purification, FT is the flowthrough, W stands for wash, E stands for elution. Each elution is 1 mL so, E4-6 for example is 3 mL. Absorbance was measured at 280 nm using a Thermo Scientific Nanodrop 1000 Spectrophotometer. The corresponding concentration for the absorbance of 1.78 is measured to be 127 µM.

Covalent modification of VL192

Prior work completed by WWU graduate student Derek McCaffery showed that VL192 when exposed to A. *thaliana* cell extract rapidly undergoes covalent modification, in what appears to be proteolysis.[34] In fact, it was shown that in 5 minutes of exposure, over 90% of VL192 is gone.[34] To combat this proteolytic activity phenylmethylsulfonyl fluoride (PMSF) a serine and cystine protease inhibitor was used. Though the proteolysis was slowed, even using a final concentration of 100 mM PMSF all VL192 was gone after 1.5 hours.[34] My goal was too subject VL192 to A. *thaliana* cell extract, extract VL192 back out using gravity flow chromatography and monitor the sample for any signs of covalent modification beyond proteolysis using LC-QTOF-MS.

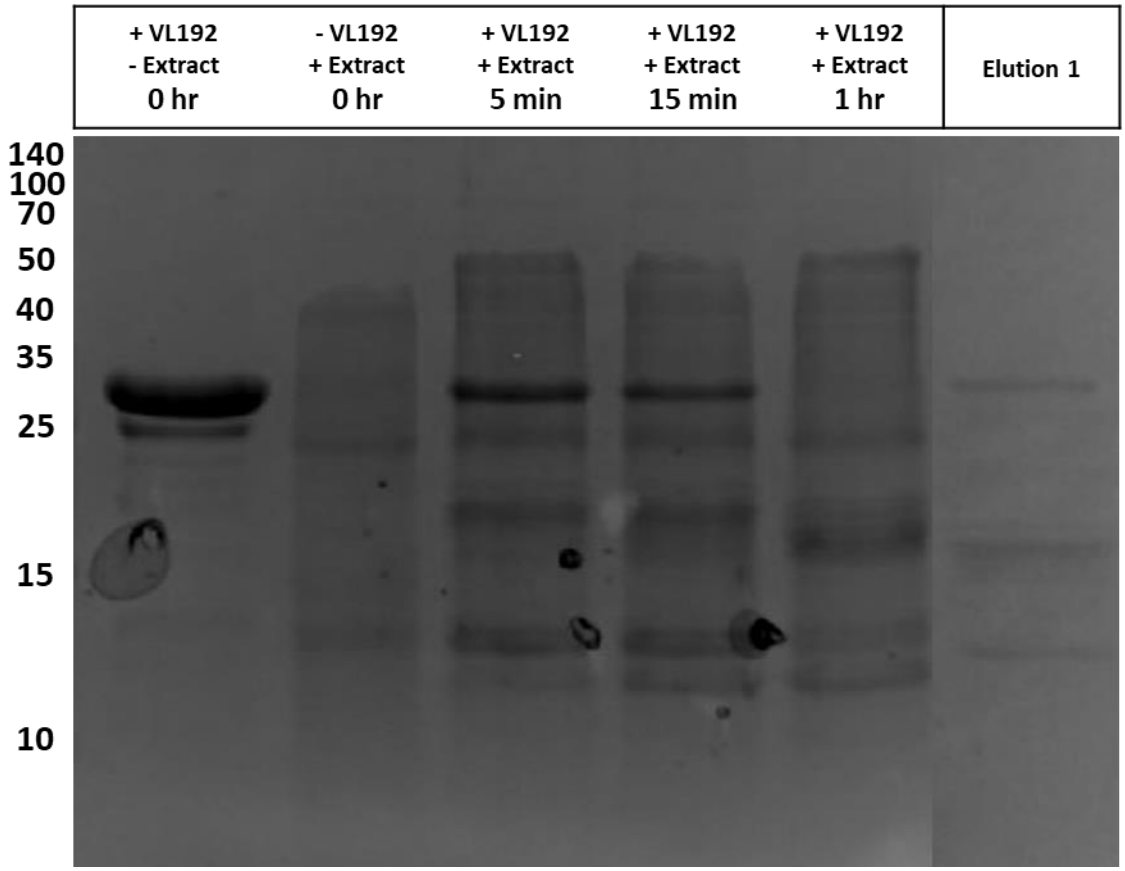


Figure 3.13 15% SDS-PAGE of VL192 exposed to A. *thaliana* cell extract over 1 hour in 25 mM PMSF and the elution of a Ni-NTA gravity flow purification extracting VL192 and any covalent modified product after 5 minutes of exposure.

Elution 1 from this study was immediately subjected to LC-QTOF-MS and examined for masses larger than 24687 Da (molecular weight of unmodified VL192). Figure 3.14 shows the difference in chromatograms after VL192 is exposed to and extracted back out of A. *thaliana* cell extract. We clearly see a change in hydrophobicity, and because a C4 column was used we can assume that any modification which occurred would make VL192 more polar.

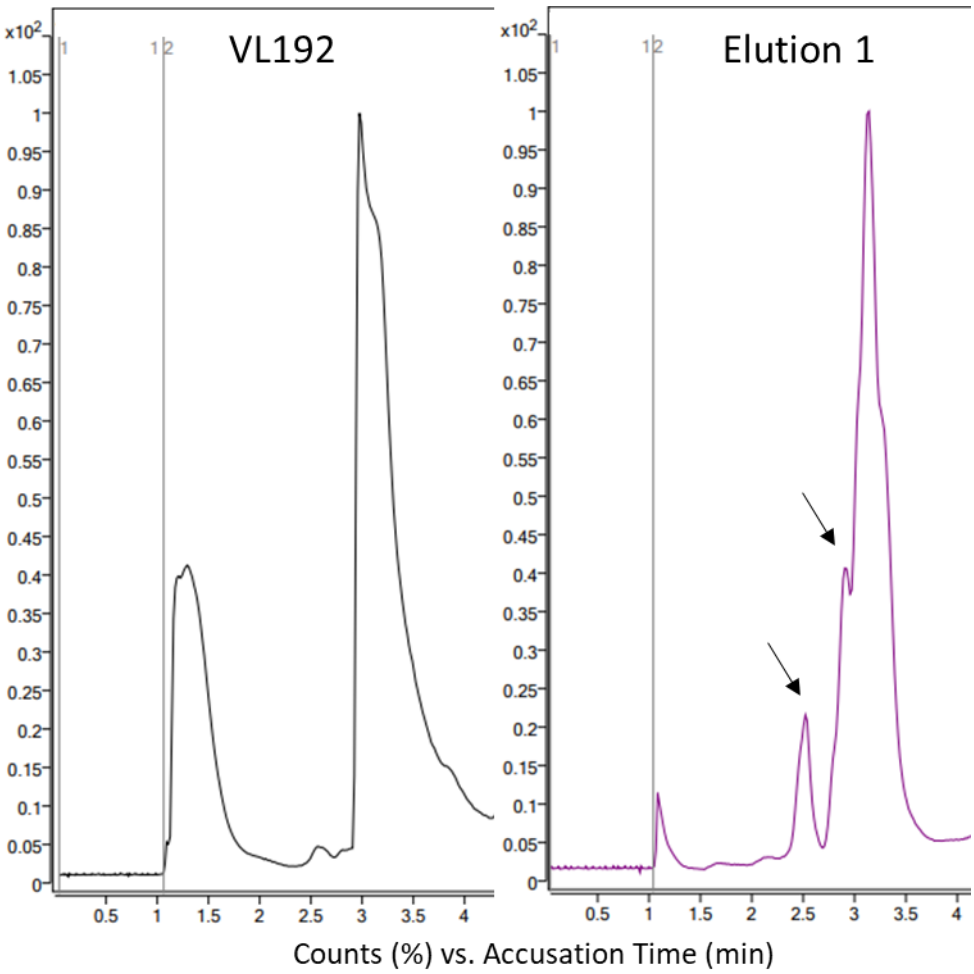


Figure 3.14 HPLC Chromatogram of VL192 and Elution 1. The black arrows show the change in hydrophobicity of the sample, which suggests that these peaks correspond to VL192 that has been covalently modified.

When deconvoluted, the peak at 2.5 minutes shows three predominate masses above that of VL192, specifically 26620, 32363, and 42626 Da (appendix V). At time point 2.6-3.5 we see two predominant masses above that of VL192, specifically 25225 and 27832 Da (appendix V). Though Figure 3.14 shows that the sample repurified out of A. *thaliana* cell extract exhibits a change in hydrophobicity, none of these mass shifts correlate to predicted covalent modifications. The two predicted covalent modifications were ubiquitination and phosphorylation, which correspond to a mass shift of 8,800 and 80 Da respectively. The results

were consistent across all three trials taken to the LC-QTOF-MS and VL192 continued to interact with A. *thaliana* in the same manner demonstrated originally in McCaffery's prior work.[34]

VL147N was also subjected to one covalent modification trial without the presence of PMSF and similarly to VL192, underwent what appeared to be rapid proteolysis (**Figure 3.15**).

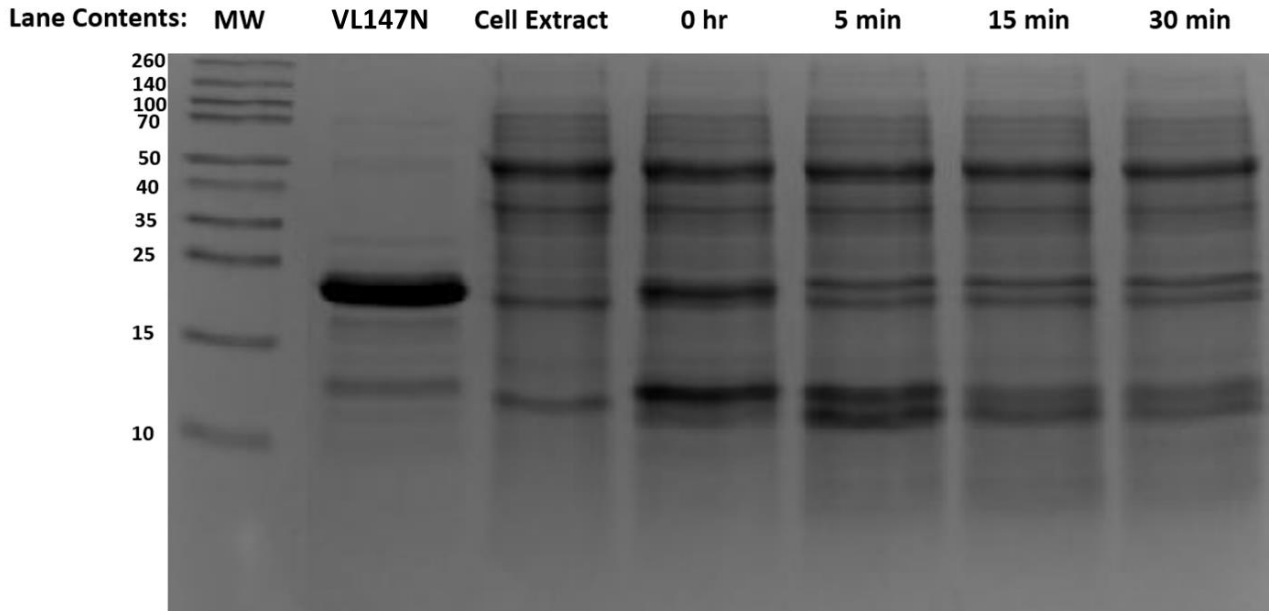


Figure 3.15 15% SDS-PAGE of VL147N exposed to A. *thaliana* cell extract over the duration of 30 minutes. VL147N for this trial had its purification tags intact.

Aim 3 Discussion

As it stands, we cannot identify the specific covalent modifications that VL192 or VL147N undergo when subjected to A. *thaliana* cell extract. We see rapid degradation of the target bands in Figures 3.13 and 3.15 and though we have successfully repurified VL192 back out of the cell extract none of the predominant masses which we have observed with LC-QTOF-MS match with the projected mass shifts corresponding to the predicted covalent modifications.

To improve the likelihood of identifying covalent modifications which occur when villin constructs are exposed to A. *thaliana* cell extract, I would suggest truncating the constructs to significantly smaller sizes. If we exposed a 5-10 amino acid IDR to cell extract, there would be fewer locations to which covalent modifications could occur. Alternatively, instead of grinding up plant material from the entire A. *thaliana* plant, only use material from a specific area (leaves, root, stem, etc.). With this information you could cross-reference bioinformatics to identify potential proteasomes or other covalent modifiers found in that specific region of the plant. In addition to either of these suggestions, the use of protease cocktails would vastly increase the time in which one has to work with the sample prior to it degrading fully.

Chapter IV Conclusions

Sortase Mediated Ligation

The expression and purification for all 13 original constructs and the 8 addition glycine constructs has been optimized for conducting sortase mediated ligation studies. At the time of publication, of the 15 N-terminal constructs studied, we have optimized the ligation efficacy of 10 N-terminal constructs. We see that using 7mutA is effective in increasing maximal ligation efficacy in most scenarios when compared to the results of WTsrtA. The addition of $NiSO_4$ has the most prominent effect, in increasing ligation efficacy. However, AroN saw no major change between using 7mutA alone and using 7mutA with $NiSO_4$.

With the use of additional glycine spacers on both the N-terminal context and the C-terminal peptide partner, we noticed drastic improvements, particularly with AroN. Using 7mutA and the addition of $NiSO_4$ the ligation efficacy of AroN increases from 3% to 14 % and 18% with AroNG and AroN 2G respectively. Using the additional glycine peptide partners for AroNG and AroN 2G, we are reaching ligation efficacy of 33% and 57%. The C-terminal context seems to have the greatest effect on the AcidN constructs, with AcidNG suffering from worse ligation efficacy than AcidN. Trials with AcidN 2G and its additional glycine peptide partner would complete this data set.

It is still preemptive to discuss patterns without the entire data set completed. However, it appears as though the use of additional glycine spacers on both the N-terminal context and the C-terminal context may have a significant positive impact on the ligation efficacy of specific

constructs. It is clear that each construct behaves differently, further justifying the need for this research.

F-actin Binding of VL147N and VL192

Both the full sized intrinsically disordered region of Villin4 (VL192) and the basic side of the region (VL147N) definitively bind actin filaments. The data strongly suggests that VL147N has a binding affinity in the sub micromolar range. However, until another binding assay is conducted on VL147N, we cannot confirm the validity of this claim due to the mechanical issue that occurred during the binding studies. The specific binding capacity of VL192 has yet to be determined.

This work continues the ongoing effort to understand the specific binding interactions which take place between the large intrinsically disordered regions of villin gene family members and that of filamentous actin. Actin is highly homologous between yeast, plants and humans and is one of the most evolutionarily conserved proteins within eukaryotes.

Preliminary Probing of VL147N Backbone Dynamics

The HSQC of ^{15}N VL147N confirms that the basic part of Villin4's linker is a highly dynamic, largely disordered structure. This data shows that this fragment does have some small, folded regions. However, with the peak width and intensity differing greatly, we cannot yet solve this structure. A uniformly labeled ^{13}C/^{15}N sample of VL147N would likely be necessary for this

endeavor. It is completely possible that this would not be enough, and we would have to resort to segmental labeling or using a more powerful NMR instrument.

To produce a uniformly ^{15}N labeled IDR of 147 residues at over 300 µM is a massive feat. One which would never have been possible without Caleb Howerton's persistence and diligent work. The methods we have produced will be adapted and used in future projects which require ^{15}N labeled samples of large IDRs.

Covalent Modifications of VL192 and VL147N

When exposed to A. *thaliana* cell extract both VL192 and VL147 undergo rapid covalent modification. There is clear evidence of proteolysis, even with high concentrations of PSMF being added to the mixture. From this, we can conclude that there are some other proteases beside those that target serines. Using Ni-NTA gravity purification, we were able to repurify VL192 back out of the cell extract mixture. However, when probed with LC-QTOF-MS we could not identify specific modifications that would correspond to the resulting shift in mass.

As it stands there are too many potential sites for covalent modification in VL192 for us to accurately identify specific modifications. By truncating VL192 down to a couple of 10 amino acid sequences, it would be much easier to identify the specific modifications which occur when proteins are exposed to A. *thaliana*. This is a practical exercise that would produce rapid actionable results.

References

(1) Forman-Kay, J. D.; Mittag, T. From Sequence and Forces to Structure, Function, and Evolution of Intrinsically Disordered Proteins. *Structure* **2013**, *21* (9), 1492–1499. https://doi.org/10.1016/j.str.2013.08.001.

(2) Pazos, F.; Pietrosemoli, N.; García-Martín, J.; Solano, R. Protein Intrinsic Disorder in Plants. *Frontiers in Plant Science* **2013**, *4*. https://doi.org/10.3389/fpls.2013.00363.

(3) Uversky, V. N. Paradoxes and Wonders of Intrinsic Disorder: Stability of Instability. *Intrinsically Disord Proteins* **2017**, *5* (1), e1327757. https://doi.org/10.1080/21690707.2017.1327757.

(4) van der Lee, R.; Buljan, M.; Lang, B.; Weatheritt, R. J.; Daughdrill, G. W.; Dunker, A. K.; Fuxreiter, M.; Gough, J.; Gsponer, J.; Jones, D. T.; Kim, P. M.; Kriwacki, R. W.; Oldfield, C. J.; Pappu, R. V.; Tompa, P.; Uversky, V. N.; Wright, P. E.; Babu, M. M. Classification of Intrinsically Disordered Regions and Proteins. *Chem. Rev.* **2014**, *114* (13), 6589–6631. https://doi.org/10.1021/cr400525m.

(5) Bah, A.; Forman-Kay, J. D. Modulation of Intrinsically Disordered Protein Function by Post-Translational Modifications. *Journal of Biological Chemistry* **2016**, *291* (13), 6696–6705. https://doi.org/10.1074/jbc.R115.695056.

(6) Boyko, K. V.; Rosenkranz, E. A.; Smith, D. M.; Miears, H. L.; Cheikh, M. O. es; Lund, M. Z.; Young, J. C.; Reardon, P. N.; Okon, M.; Smirnov, S. L.; Antos, J. M. Sortase-Mediated Segmental Labeling: A Method for Segmental Assignment of Intrinsically Disordered Regions in Proteins. *PLOS ONE* **2021**, *16* (10), e0258531. https://doi.org/10.1371/journal.pone.0258531.

(7) Mao, H.; Hart, S. A.; Schink, A.; Pollok, B. A. Sortase-Mediated Protein Ligation: A New Method for Protein Engineering. *J. Am. Chem. Soc.* **2004**, *126* (9), 2670–2671. https://doi.org/10.1021/ja039915e.

(8) Mikula, K. M.; Krumwiede, L.; Plückthun, A.; Iwaï, H.; Segmental Isotopic Labeling by Asparaginyl Endopeptidase-Mediated Protein Ligation. *Journal of Biomolecular NMR* **2018**, *71* (4), 225–235. https://doi.org/10.1007/s10858-018-0175-4.

(9) Spirig, T.; Weiner, E. M.; Clubb, R. T. Sortase Enzymes in Gram-Positive Bacteria. *Molecular Microbiology* **2011**, *82* (5), 1044–1059. https://doi.org/10.1111/j.1365-2958.2011.07887.x.

(10) Kruger, R. G.; Otvos, B.; Frankel, B. A.; Bentley, M.; Dostal, P.; McCafferty, D. G. Analysis of the Substrate Specificity of the Staphylococcus Aureus Sortase Transpeptidase SrtA. *Biochemistry* **2004**, *43* (6), 1541–1551. https://doi.org/10.1021/bi035920j.

(11) Wu, Q.; Ploegh, H. L.; Truttmann, M. C. Hepta-Mutant Staphylococcus Aureus Sortase A (SrtA7m) as a Tool for in Vivo Protein Labeling in Caenorhabditis Elegans. *ACS Chem. Biol.* **2017**, *12* (3), 664–673. https://doi.org/10.1021/acschembio.6b00998.

(12)
Wickstead, B.; Gull, K. The Evolution of the Cytoskeleton. *Journal of Cell Biology* **2011**, *194* (4), 513–525. https://doi.org/10.1083/jcb.201102065.

(13)
Moseley, J. B. An Expanded View of the Eukaryotic Cytoskeleton. *MBoC* **2013**, *24* (11), 1615–1618. https://doi.org/10.1091/mbc.e12-10-0732.

(14)
Suarez, C.; Kovar, D. R. Internetwork Competition for Monomers Governs Actin Cytoskeleton Organization. *Nature Reviews. Molecular Cell Biology* **2016**, *17* (12), 799–810. https://doi.org/10.1038/nrm.2016.106.

(15)
Dos Remedios, C. G.; Chhabra, D.; Kekic, M.; Dedova, I. V.; Tsubakihara, M.; Berry, D. A.; Nosworthy, N. J. Actin Binding Proteins: Regulation of Cytoskeletal Microfilaments. *Physiological Reviews* **2003**, *83* (2), 433–473. https://doi.org/10.1152/physrev.00026.2002.

(16)
Gunning, P. W.; Ghoshdastider, U.; Whitaker, S.; Popp, D.; Robinson, R. C. The Evolution of Compositionally and Functionally Distinct Actin Filaments. *Journal of Cell Science* **2015**, *128* (11), 2009–2019. https://doi.org/10.1242/jcs.165563.

(17)
Dominguez, R.; Holmes, K. C. Actin Structure and Function. *Annu. Rev. Biophys.* **2011**, *40* (1), 169–186. https://doi.org/10.1146/annurev-biophys-042910-155359.

(18)
Friederich, E.; Vancompernolle, K.; Louvard, D.; Vandekerckhove, J. Villin Function in the Organization of the Actin Cytoskeleton: CORRELATION OF IN VIVO EFFECTS TO ITS BIOCHEMICAL ACTIVITIES IN VITRO *. *Journal of Biological Chemistry* **1999**, *274* (38), 26751–26760. https://doi.org/10.1074/jbc.274.38.26751.

(19)
Bretscher, A.; Weber, K. Villin Is a Major Protein of the Microvillus Cystoskeleton Which Binds Both G and F Actin in a Calcium-Dependent Manner. *Cell* **1980**, *20* (3), 839–847. https://doi.org/10.1016/0092-8674(80)90330-X.

(20)
Lemière, J.; Valentino, F.; Campillo, C.; Sykes, C. How Cellular Membrane Properties Are Affected by the Actin Cytoskeleton. *Biochimie* **2016**, *130*, 33–40. https://doi.org/10.1016/j.biochi.2016.09.019.

(21)
Hsiao, A.-S.; Wang, K.; Ho, T.-H. D. An Intrinsically Disordered Protein Interacts with the Cytoskeleton for Adaptive Root Growth under Stress. *Plant Physiol* **2020**, *183* (2), 570–587. https://doi.org/10.1104/pp.19.01372.

(22)
Shanker, A. K.; Maheswari, M.; Yadav, S. K.; Desai, S.; Bhanu, D.; Attal, N. B.; Venkateswarlu, B. Drought Stress Responses in Crops. *Functional & Integrative Genomics* **2014**, *14* (1), 11–22. https://doi.org/10.1007/s10142-013-0356-x.

(23)
Winder, S. J.; Ayscough, K. R. Actin-Binding Proteins. *Journal of Cell Science* **2005**, *118* (4), 651–654. https://doi.org/10.1242/jcs.01670.

(24)
Smirnov, S. L.; Isern, N. G.; Jiang, Z. G.; Hoyt, D. W.; McKnight, C. J. The Isolated Sixth Gelsolin Repeat and Headpiece Domain of Villin Bundle F-Actin in the Presence of Calcium and Are Linked by a 40-Residue Unstructured Sequence,. *Biochemistry* **2007**, *46* (25), 7488–7496. https://doi.org/10.1021/bi700110v.

(25)
Csizmok, V.; Follis, A. V.; Kriwacki, R. W.; Forman-Kay, J. D. Dynamic Protein Interaction Networks and New Structural Paradigms in Signaling. *Chem. Rev.* **2016**, *116* (11), 6424–6462. https://doi.org/10.1021/acs.chemrev.5b00548.

(26)
Yokota, E.; Vidali, L.; Tominaga, M.; Tahara, H.; Orii, H.; Morizane, Y.; Hepler, P. K.; Shimmen, T. Plant 115-KDa Actin-Filament Bundling Protein, P-115-ABP, Is a Homologue of Plant Villin and Is Widely Distributed in Cells. *Plant and Cell Physiology* **2003**, *44* (10), 1088–1099. https://doi.org/10.1093/pcp/pcg132.

(27)
Chen, Y.; Takizawa, N.; Crowley, J. L.; Oh, S. W.; Gatto, C. L.; Kambara, T.; Sato, O.; Li, X.; Ikebe, M.; Luna, E. J. F-Actin and Myosin II Binding Domains in Supervillin*. *Journal of Biological Chemistry* **2003**, *278* (46), 46094–46106. https://doi.org/10.1074/jbc.M305311200.

(28)
Pope, B.; Way, M.; Matsudaira, P. T.; Weeds, A. Characterisation of the F-Actin Binding Domains of Villin: Classification of F-Actin Binding Proteins into Two Groups According to Their Binding Sites on Actin. *FEBS Letters* **1994**, *338* (1), 58–62. https://doi.org/10.1016/0014-5793(94)80116-9.

(29)
Zhang, Y.; Xiao, Y.; Du, F.; Cao, L.; Dong, H.; Ren, H. Arabidopsis VILLIN4 Is Involved in Root Hair Growth through Regulating Actin Organization in a Ca2+-Dependent Manner. *New Phytologist* **2011**, *190* (3), 667–682. https://doi.org/10.1111/j.1469-8137.2010.03632.x.

(30)
Miears, H. L.; Gruber, D. R.; Horvath, N. M.; Antos, J. M.; Young, J.; Sigurjonsson, J. P.; Klem, M. L.; Rosenkranz, E. A.; Okon, M.; McKnight, C. J.; Vugmeyster, L.; Smirnov, S. L. Plant Villin Headpiece Domain Demonstrates a Novel Surface Charge Pattern and High Affinity for F-Actin. *Biochemistry* **2018**, *57* (11), 1690–1701. https://doi.org/10.1021/acs.biochem.7b00856.

(31)
Klahre, U.; Friederich, E.; Kost, B.; Louvard, D.; Chua, N.-H. Villin-like Actin-Binding Proteins Are Expressed Ubiquitously in Arabidopsis. *Plant Physiology* **2000**, *122* (1), 35–48. https://doi.org/10.1104/pp.122.1.35

(32)
Fang, Y.; Xiong, L. General Mechanisms of Drought Response and Their Application in Drought Resistance Improvement in Plants. *Cellular and Molecular Life Sciences* **2015**, *72* (4), 673–689. https://doi.org/10.1007/s00018-014-1767-0.

(33)
Tocmo, R.; Veenstra, J. P.; Huang, Y.; Johnson, J. J. Covalent Modification of Proteins by Plant-Derived Natural Products: Proteomic Approaches and Biological Impacts. *PROTEOMICS* **2021**, *21* (3–4), 1900386. https://doi.org/10.1002/pmic.201900386.

(34)
McCaffery, Derek, "Covalent Modifications as Targets and Means for Research in Intrinsically Disordered Proteins" **2022**. WWU Graduate School Collection. 1120. https://cedar.wwu.edu/wwuet/1120.

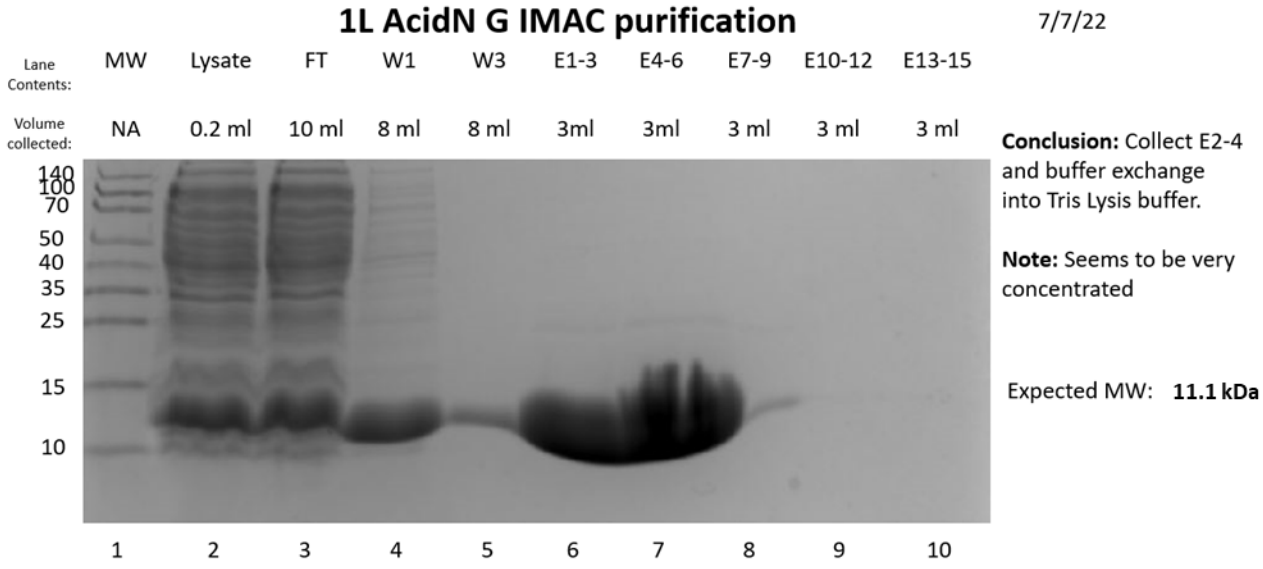
(35)
Xu, G.; Jaffrey, S. R. Proteomic Identification of Protein Ubiquitination Events. *Biotechnology and Genetic Engineering Reviews* **2013**, *29* (1), 73–109. https://doi.org/10.1080/02648725.2013.801232.

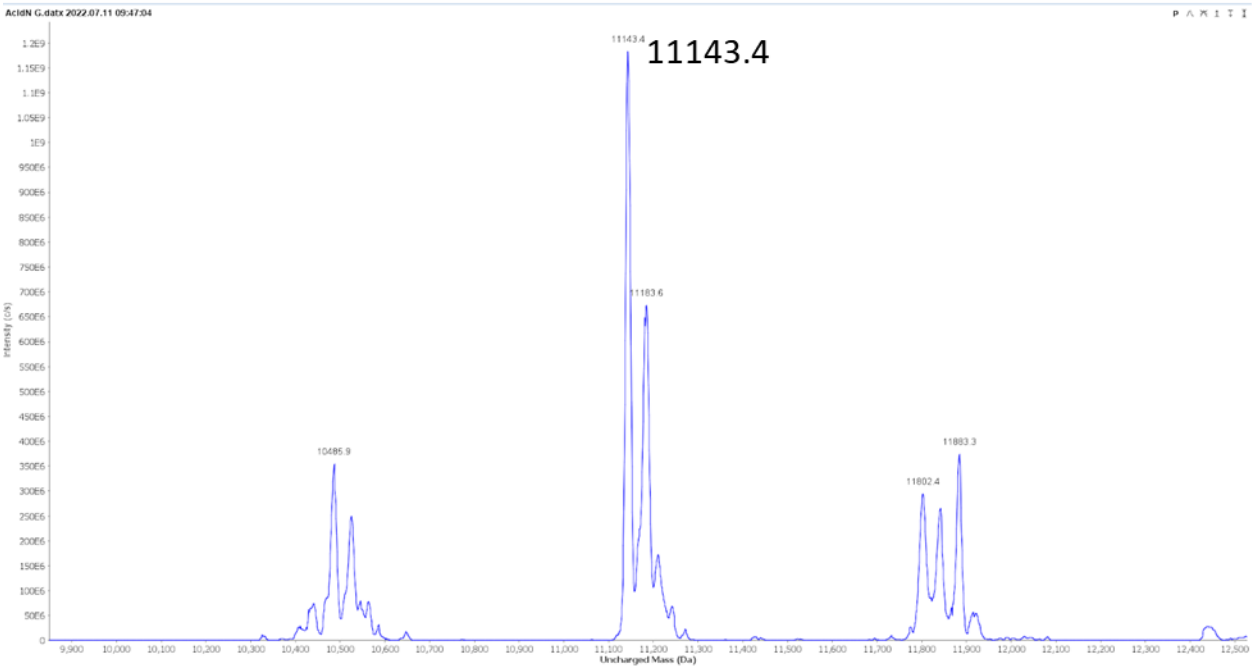
(36) Dephoure, N.; Gould, K. L.; Gygi, S. P.; Kellogg, D. R. Mapping and Analysis of Phosphorylation Sites: A Quick Guide for Cell Biologists. *MBoC* **2013**, *24* (5), 535–542. https://doi.org/10.1091/mbc.e12-09-0677.

(37) Hirakawa, H.; Ishikawa, S.; Nagamune, T. Ca2+-Independent Sortase-A Exhibits High Selective Protein Ligation Activity in the Cytoplasm of Escherichia Coli. *Biotechnology Journal* **2015**, *10* (9), 1487–1492. https://doi.org/10.1002/biot.201500012.

(38) Reed, S. A.; Brzovic, D. A.; Takasaki, S. S.; Boyko, K. V.; Antos, J. M. Efficient Sortase-Mediated Ligation Using a Common C-Terminal Fusion Tag. *Bioconjug Chem* **2020**, *31* (5), 1463–1473. https://doi.org/10.1021/acs.bioconjchem.0c00156.

Appendix I. Purification of additional glycine constructs

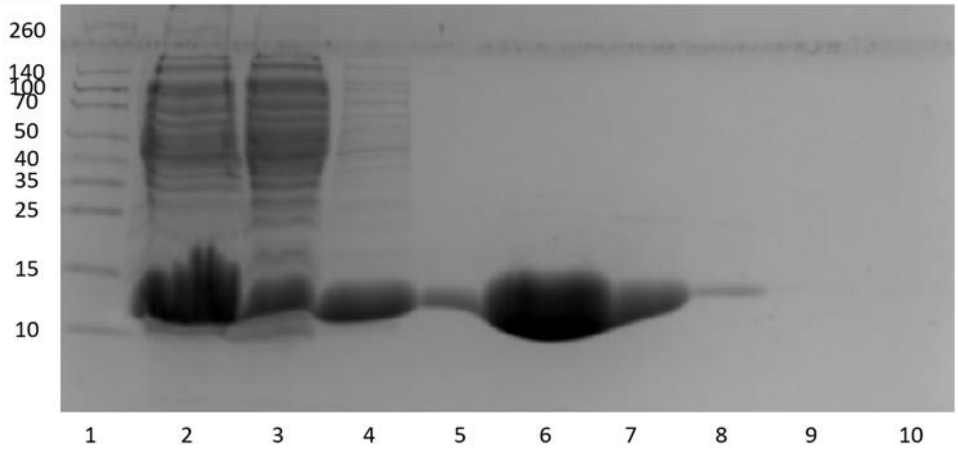




1L AcidN 2G IMAC purification

7/7/22

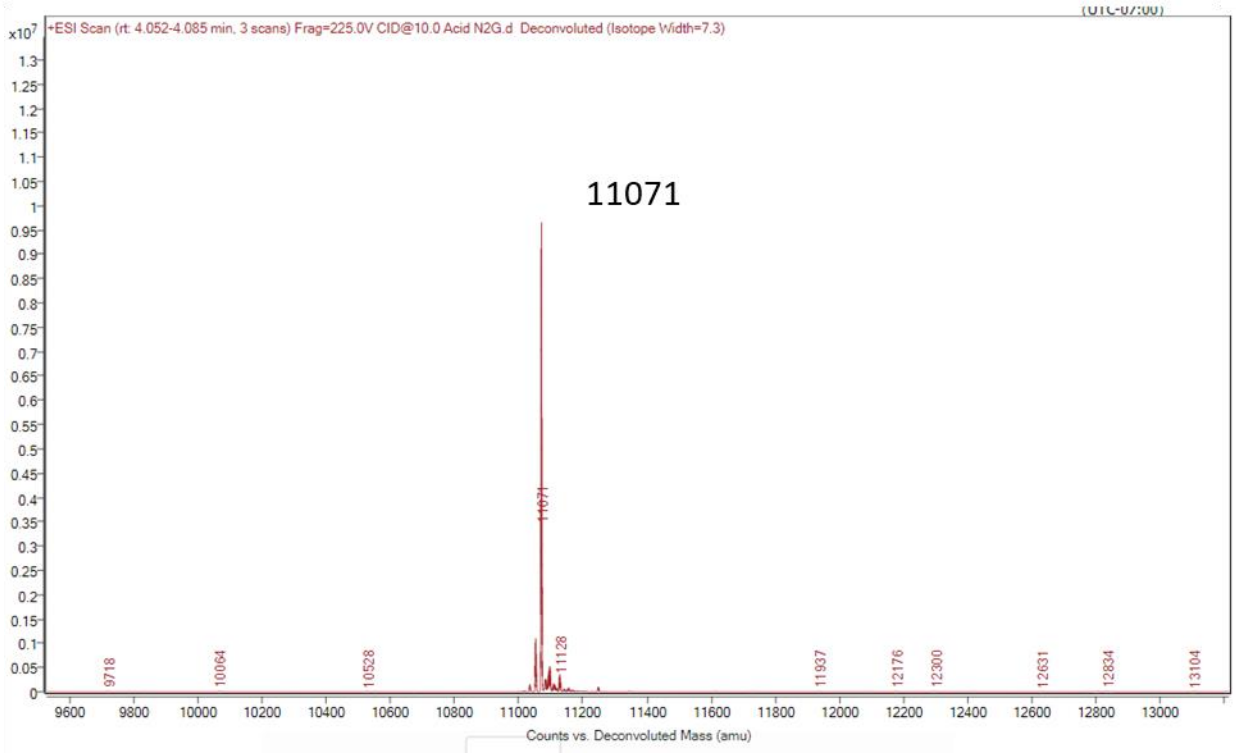
Lane Contents:	MW	Lysate	FT	W1	W3	E1-3	E4-6	E7-9	E10-12	E13-15
Volume collected:	NA	0.2 ml	10 ml	8 ml	8 ml	3ml	3ml	3 ml	3 ml	3 ml



Conclusion: Collect E1-3 and buffer exchange into Tris Lysis buffer.

Note: Nice and concentrated.

Expected MW: **11.1 kDa**

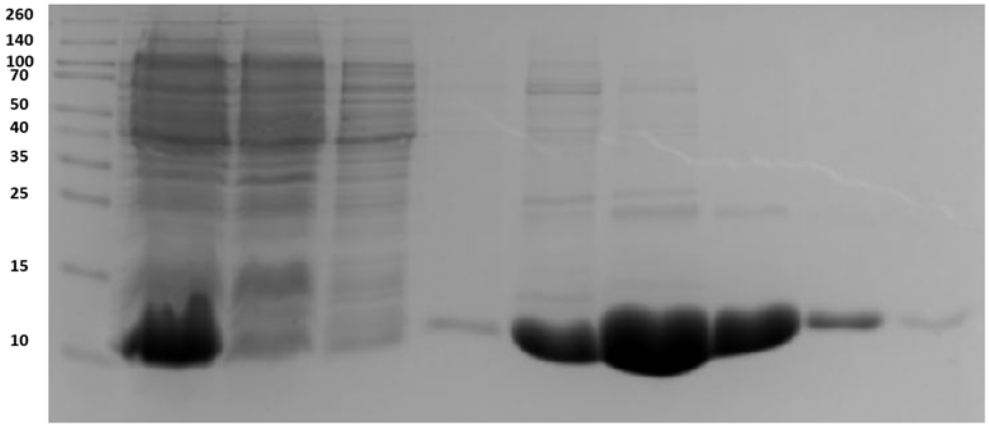


Expected MW: 11.3 kDa

1L AroN G Ni-NTA Purification

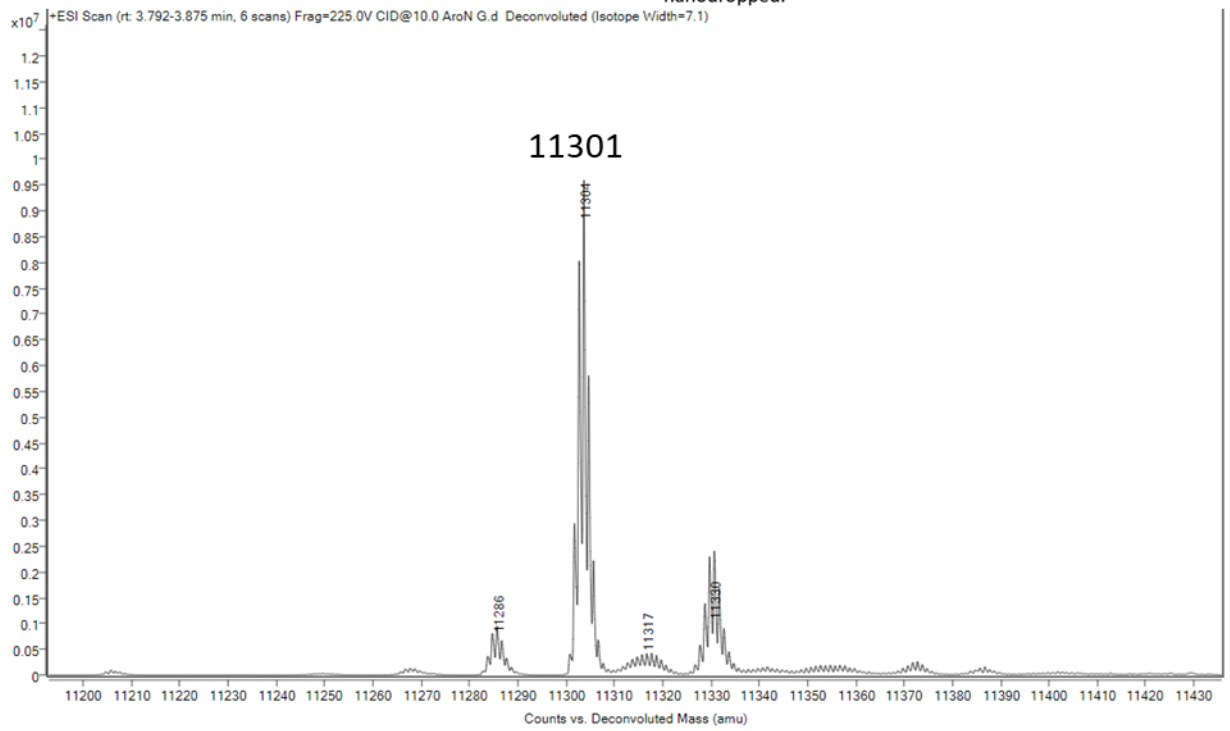
7/15/22

Lane Contents:	MW	Lysate	FT	W1	W3	E1-3	E4-6	E7-9	E10-12	E13-15
Volume Collected:		0.2 mL	10 mL	8 mL	8 mL	3 mL	3 mL	3 mL	3 mL	3 mL



Conclusion: E 4-6 will be collected, buffer exchanged, and nanodropped.

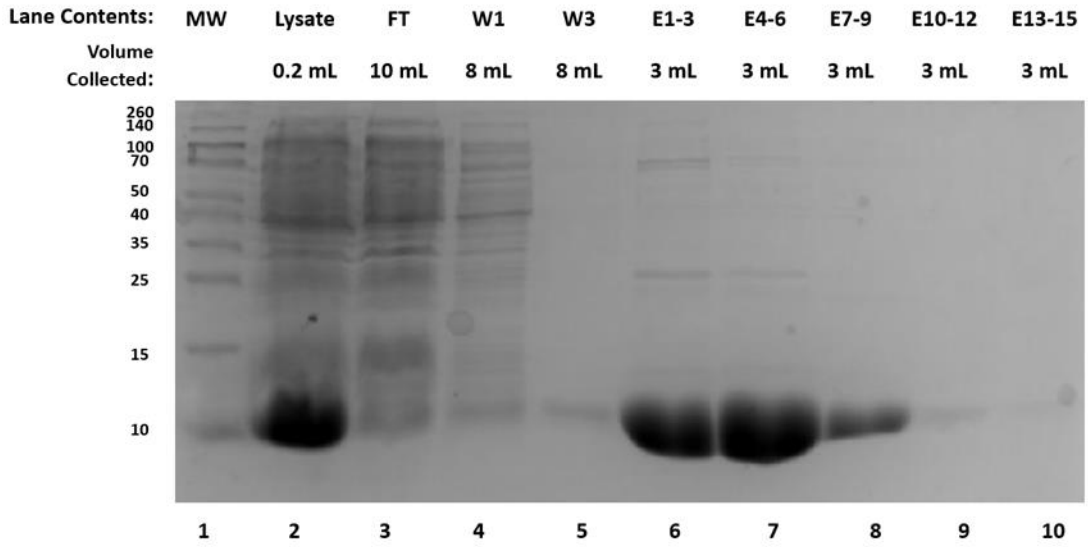
15 µL sample + 5 µL 4X per sample



Expected MW: 11.1 kDa

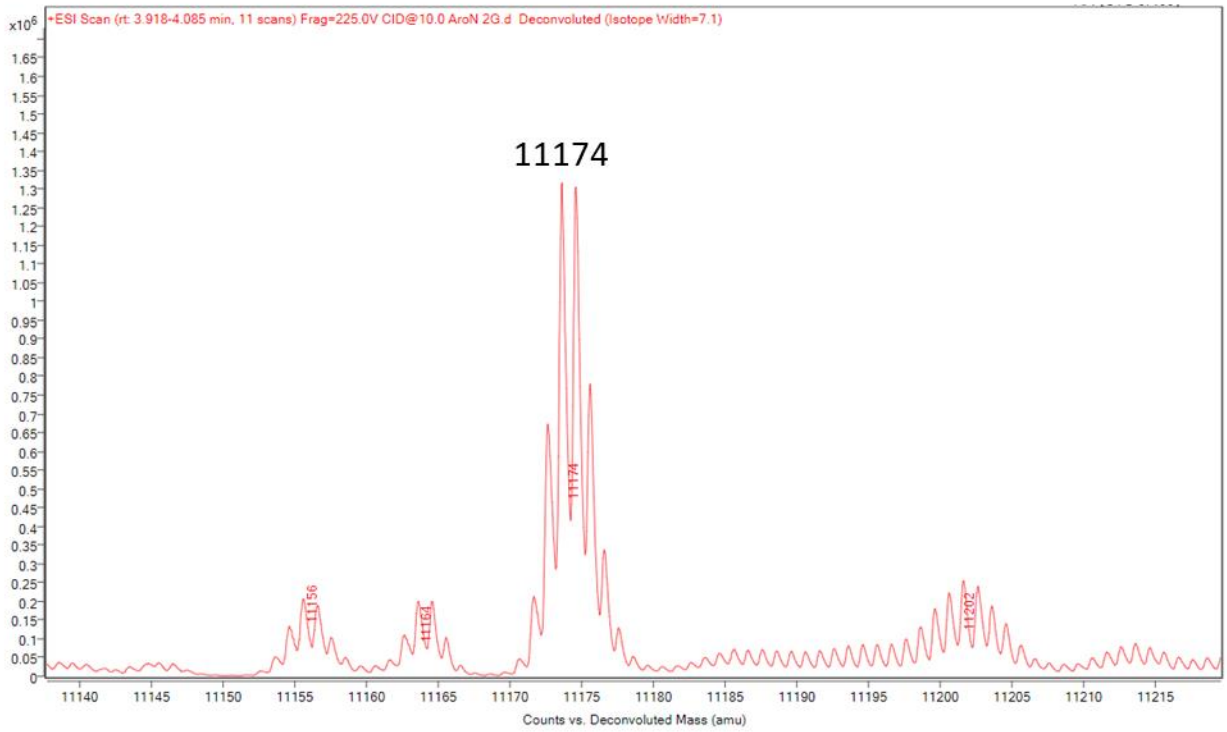
1L AroN 2G Ni-NTA Purification

7/15/22



15 µL sample + 5 µL 4X per sample

Conclusion: E 4-6 will be collected, buffer exchanged, and nanodropped.

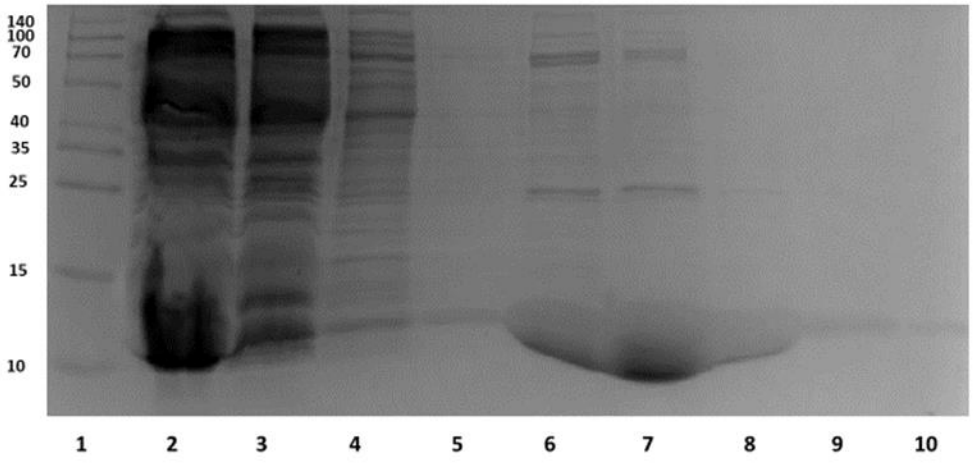


Expected MW: 11.1 kDa

1L ProN G Ni-NTA Purification

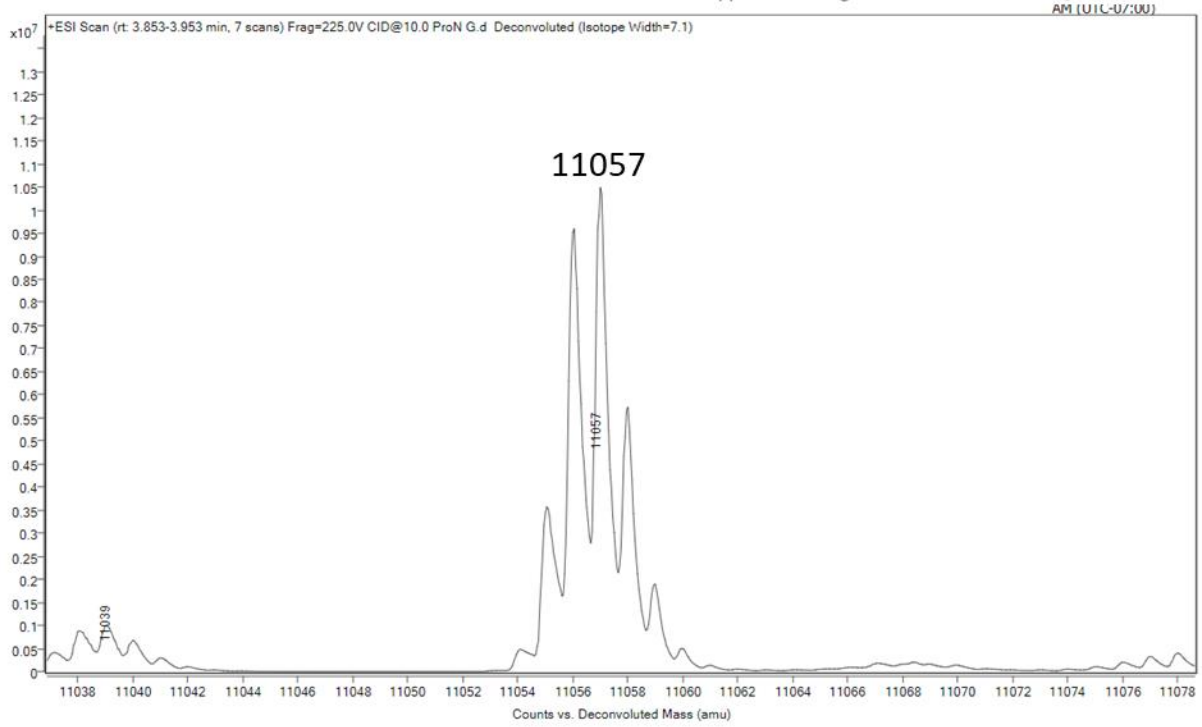
7/21/22

Lane Contents:	MW	Lysate	FT	W1	W3	E1-3	E4-6	E7-9	E10-12	E13-15
Volume Collected:		0.2 mL	10 mL	8 mL	8 mL	3 mL	3 mL	3 mL	3 mL	3 mL



15 µL sample + 5 µL 4X per sample

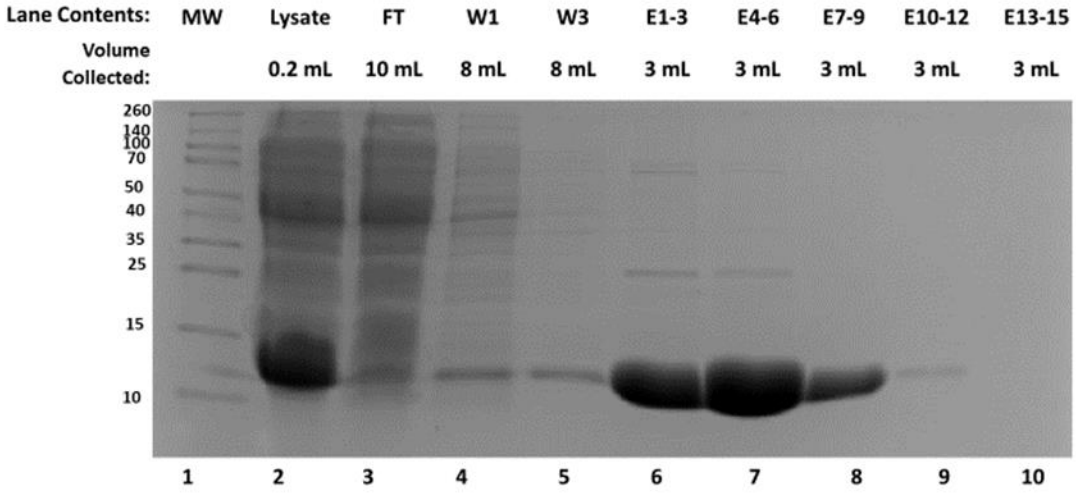
Conclusion: E 4-6 will be collected, buffer exchanged, and nanodropped. Resulting concentration was **8 mL @ 424 uM**



Expected MW: 11.02 kDa

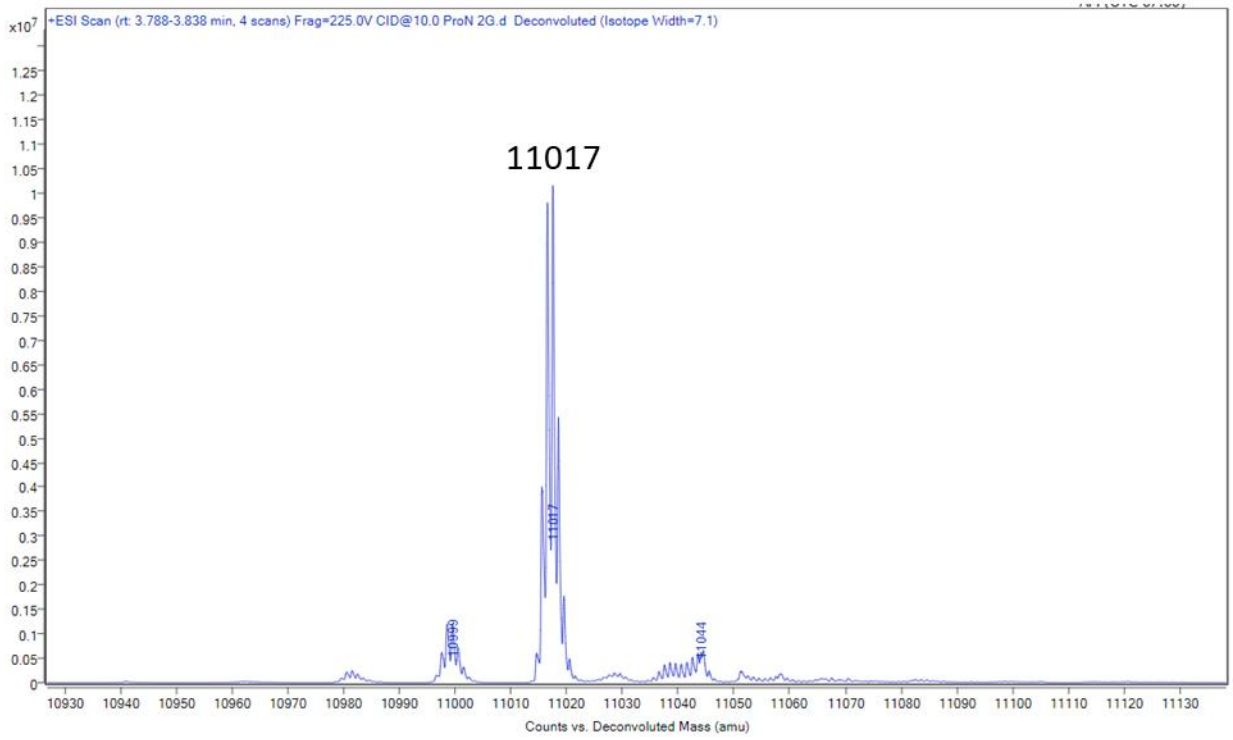
1L ProN 2G Ni-NTA Purification

7/21/22



15 µL sample + 5 µL 4X per sample

Conclusion: E 4-6 will be collected, buffer exchanged, and nanodropped. Resulting concentration was **8 mL @ 401 uM**

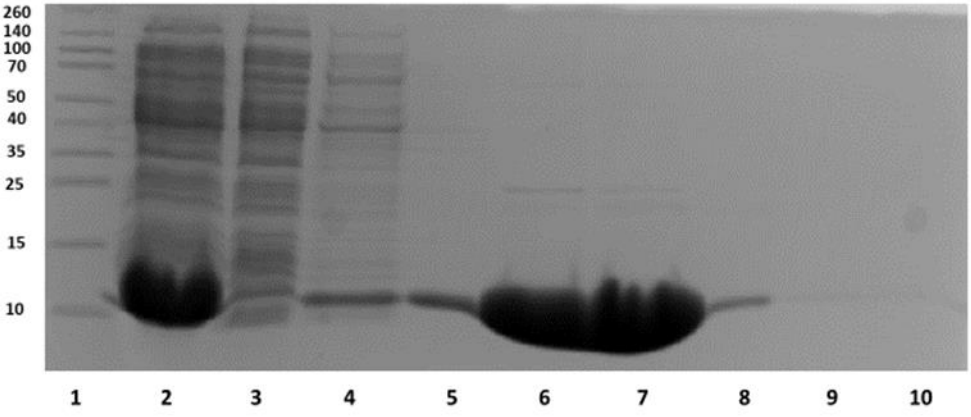


Expected MW: 11.2 kDa

1L PolN G Ni-NTA Purification

8/11/22

Lane Contents:	MW	Lysate	FT	W1	W3	E1-3	E4-6	E7-9	E10-12	E13-15
Volume Collected:		0.2 mL	10 mL	8 mL	8 mL	3 mL	3 mL	3 mL	3 mL	3 mL
OD:		N/A	N/A	N/A	N/A	1.97	2.37	0.13	0.03	0.03

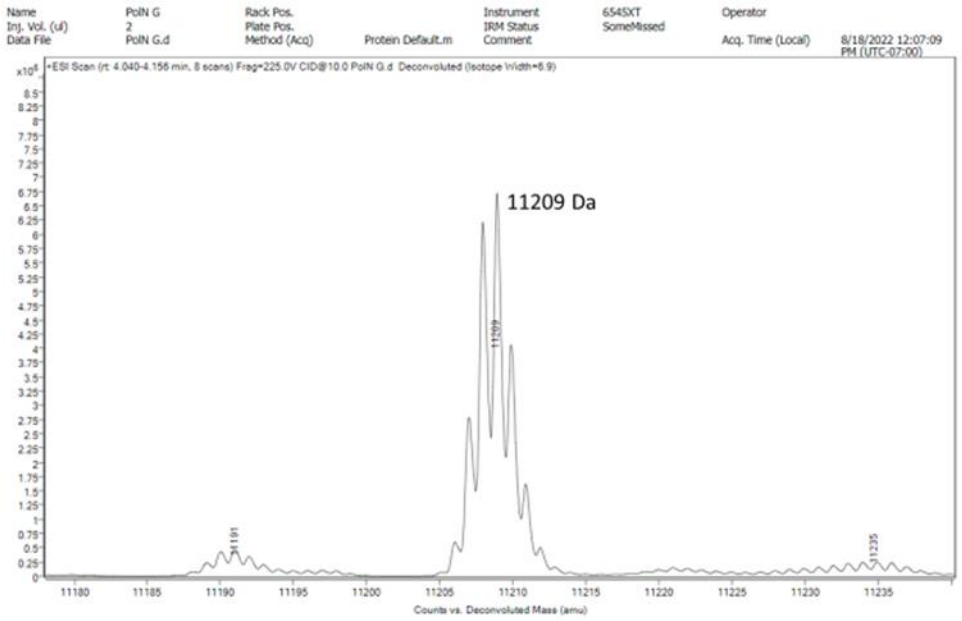


Conclusion: Elution's 3-4 were collected, and buffer exchanged into Tris Lysis buffer. **8 mL of 503 uM** was collected and flash frozen.

15 µL sample + 5 µL 4X per sample

Expected MW: 11209 Da

PolN G Spectra

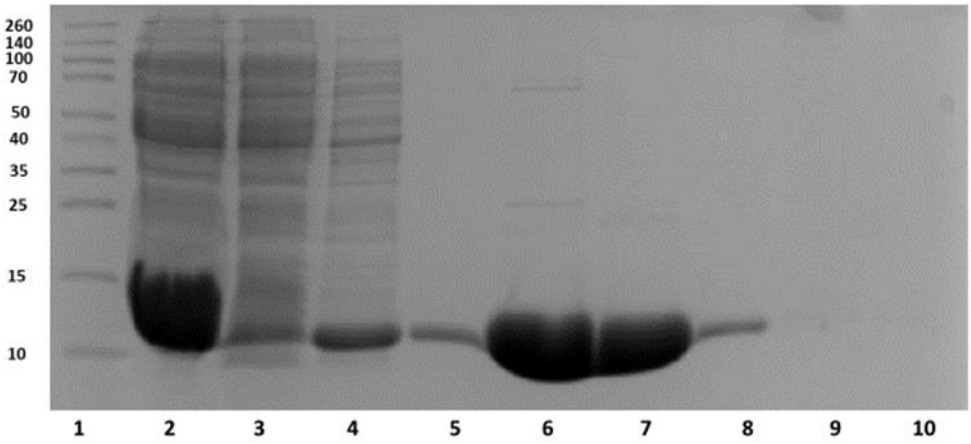


Expected MW: 11.1 kDa

1L PolN 2G Ni-NTA Purification

8/11/22

Lane Contents:	MW	Lysate	FT	W1	W3	E1-3	E4-6	E7-9	E10-12	E13-15
Volume Collected:		0.2 mL	10 mL	8 mL	8 mL	3 mL	3 mL	3 mL	3 mL	3 mL
OD:		N/A	N/A	N/A	N/A	1.86	1.14	0.11	0.02	0.02

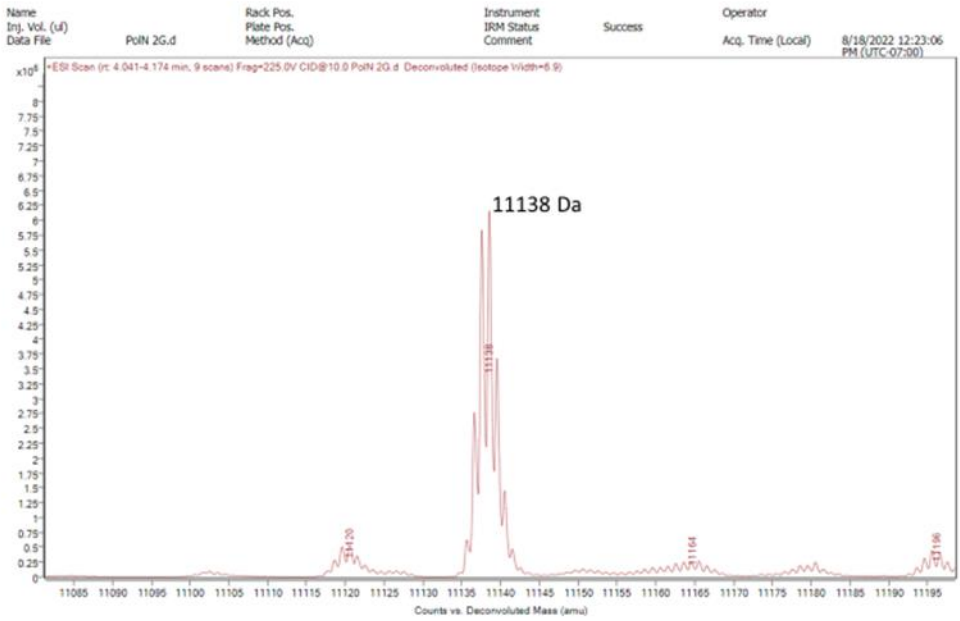


Conclusion: Elution's 2-4 were collected, and buffer exchanged into Tris Lysis Buffer. **8 mL of 573 uM** was collect and flash frozen.

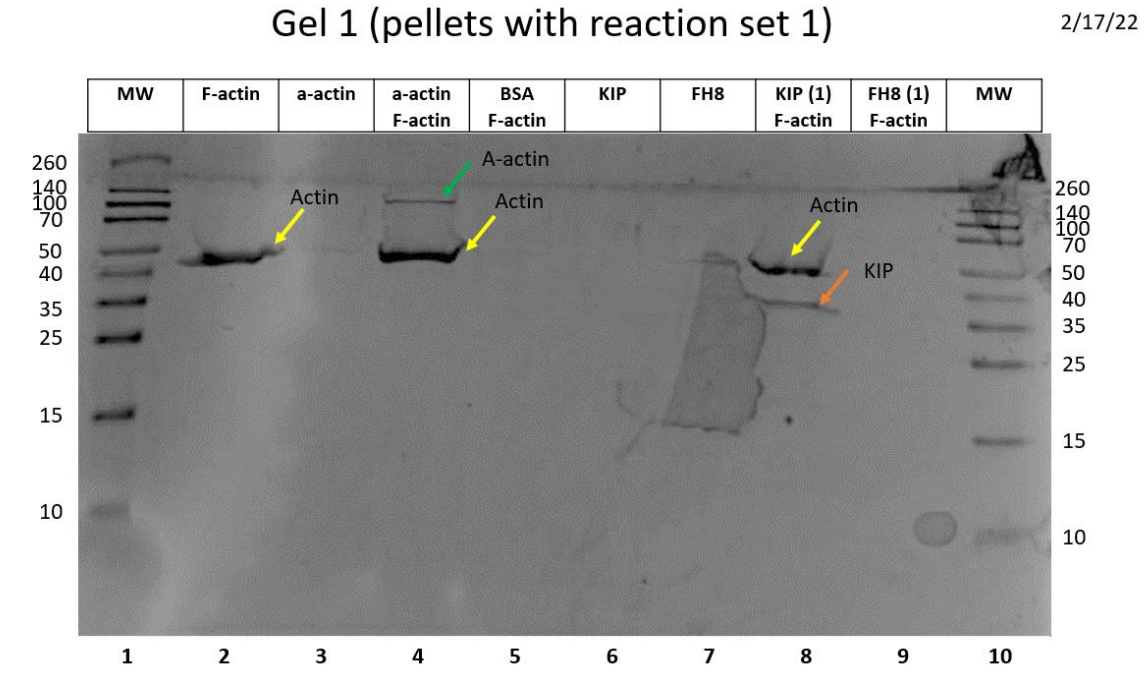
15 µL sample + 5 µL 4X per sample

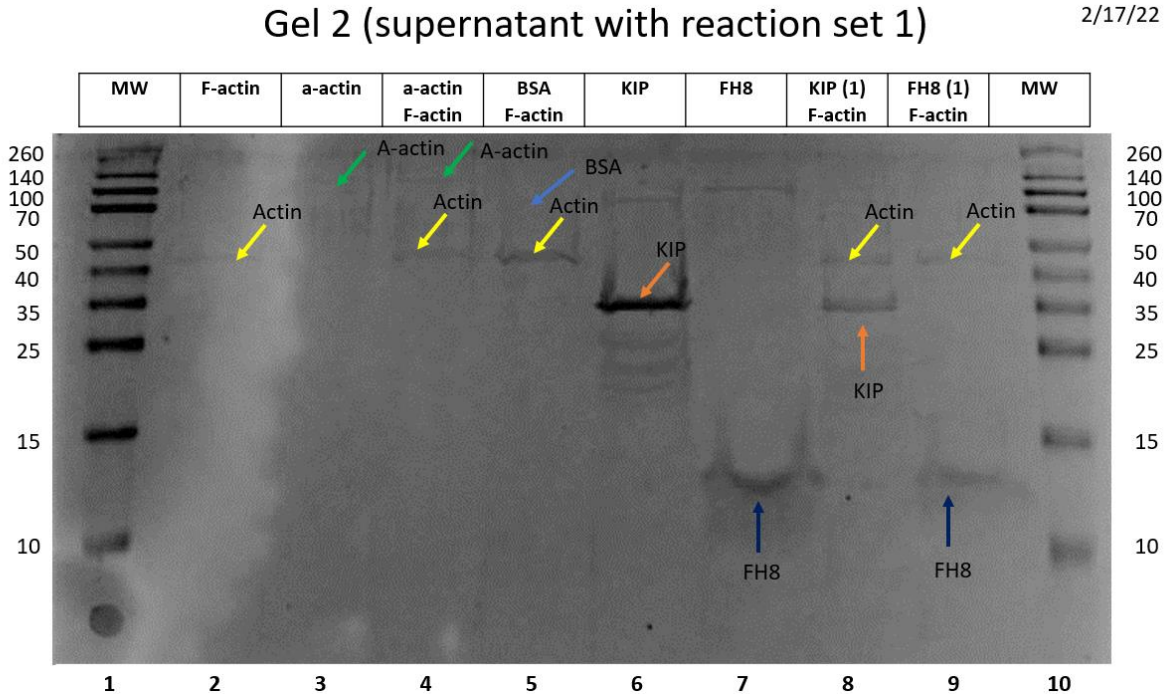
Expected MW: 11138 Da

PolN 2G Spectra



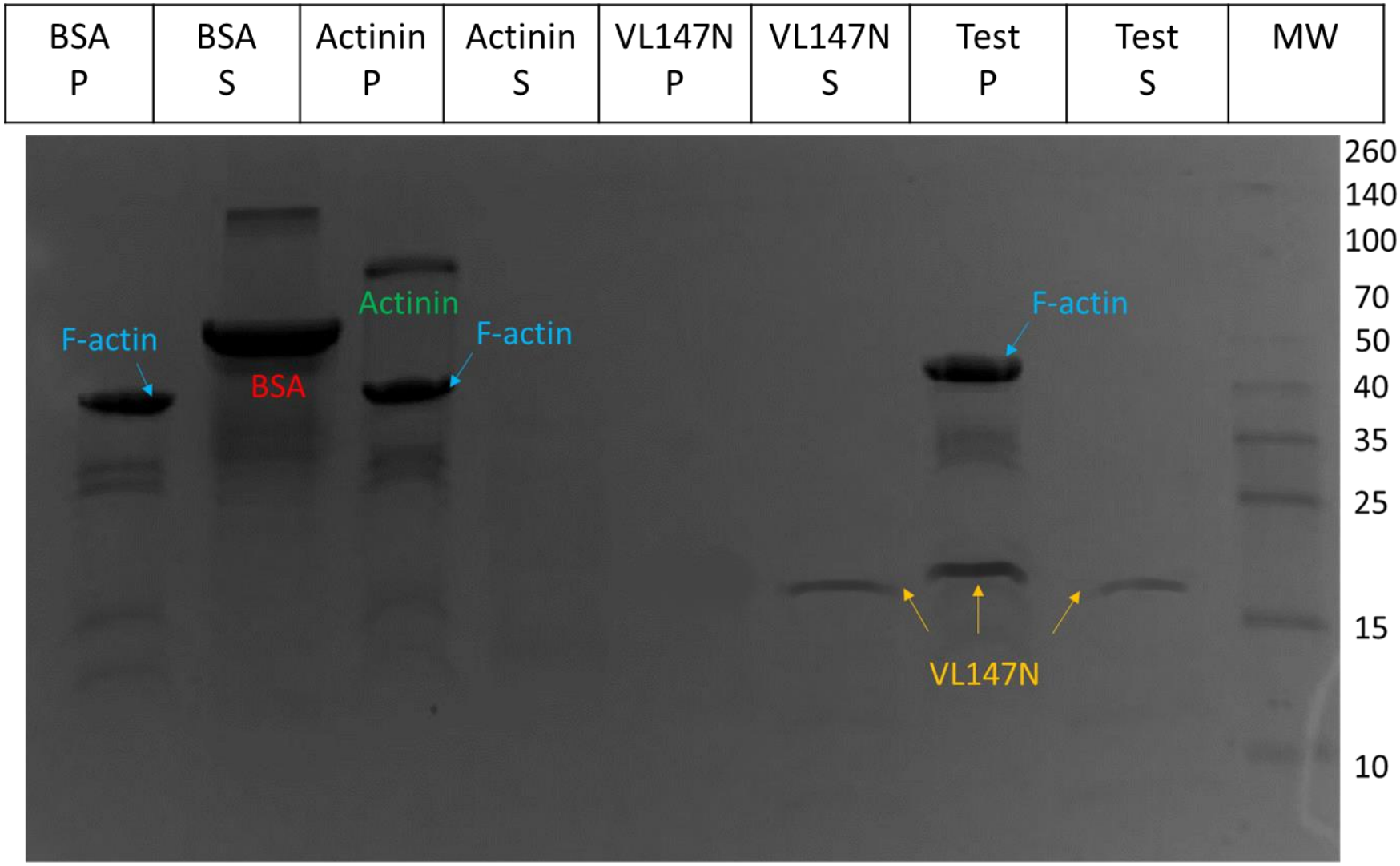
Appendix II. FH8 tags do not bind F-actin



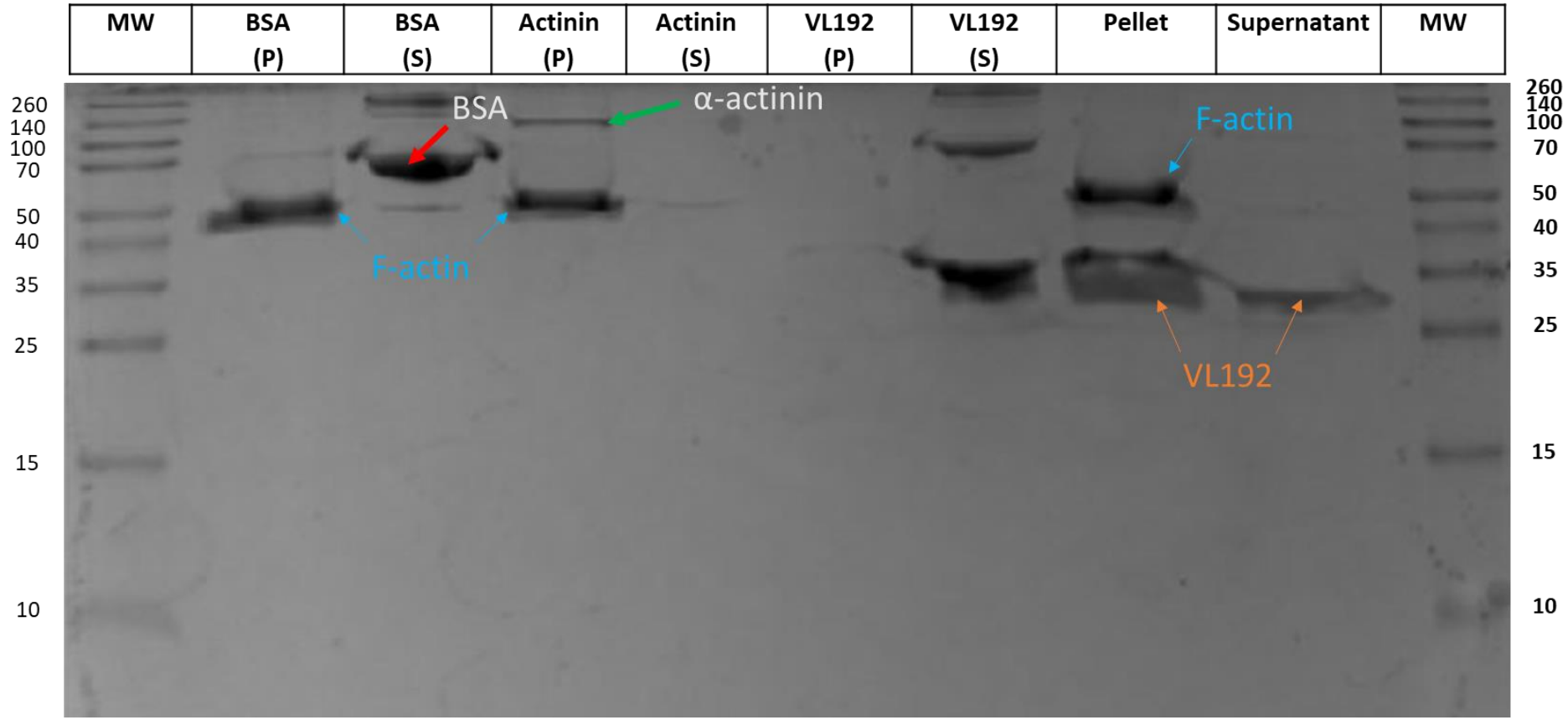


15% SDS PAGE showing that the construct known as KIP binds F-actin and that FH8 purification tag does not bind F-actin.

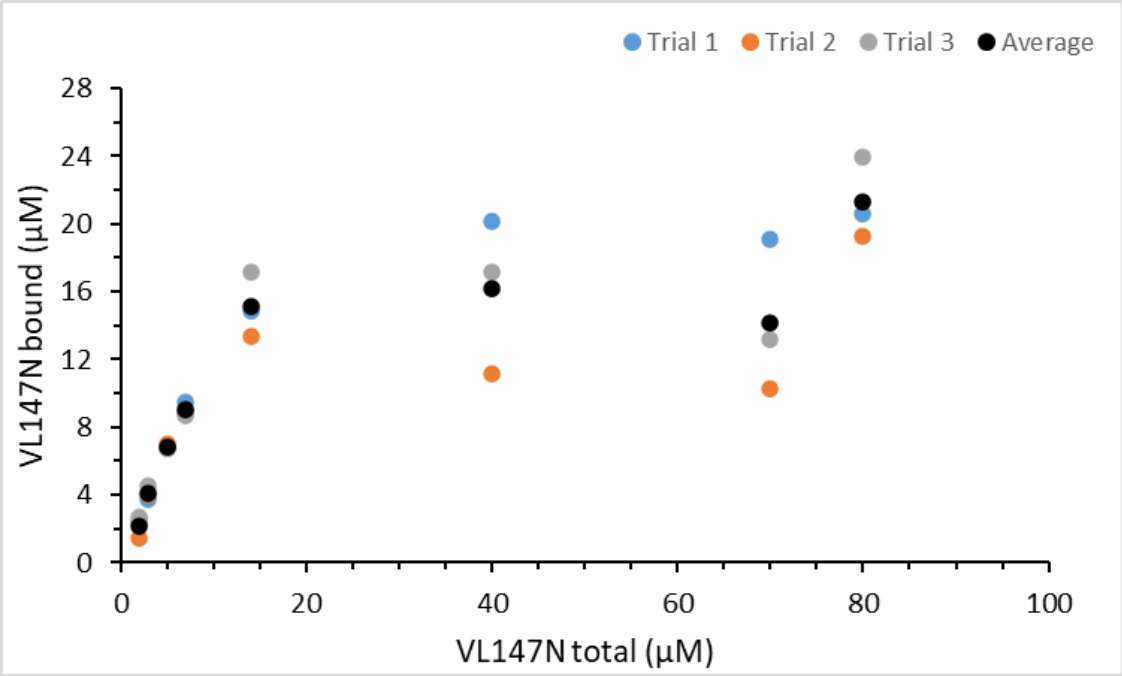
Appendix III. VL147N and VL192 bind actin and Missing O-ring



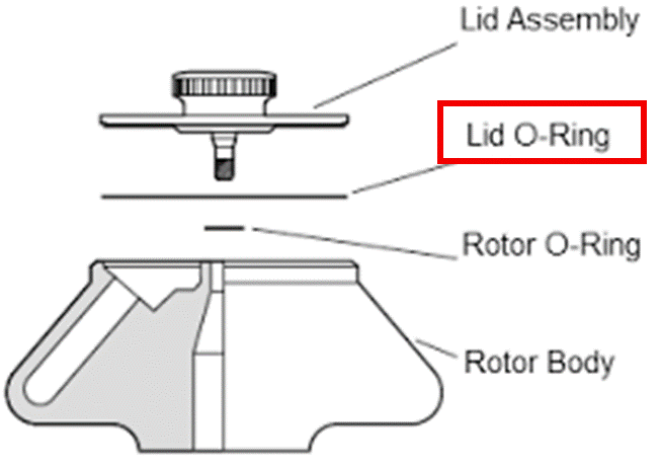
15% SDS PAGE showing the controls and that VL147N binds F-actin. BSA is the negative control, α-actinin is the positive control.



15% SDS PAGE showing the controls and that VL192 binds F-actin. BSA is the negative control, α-actinin is the positive control.

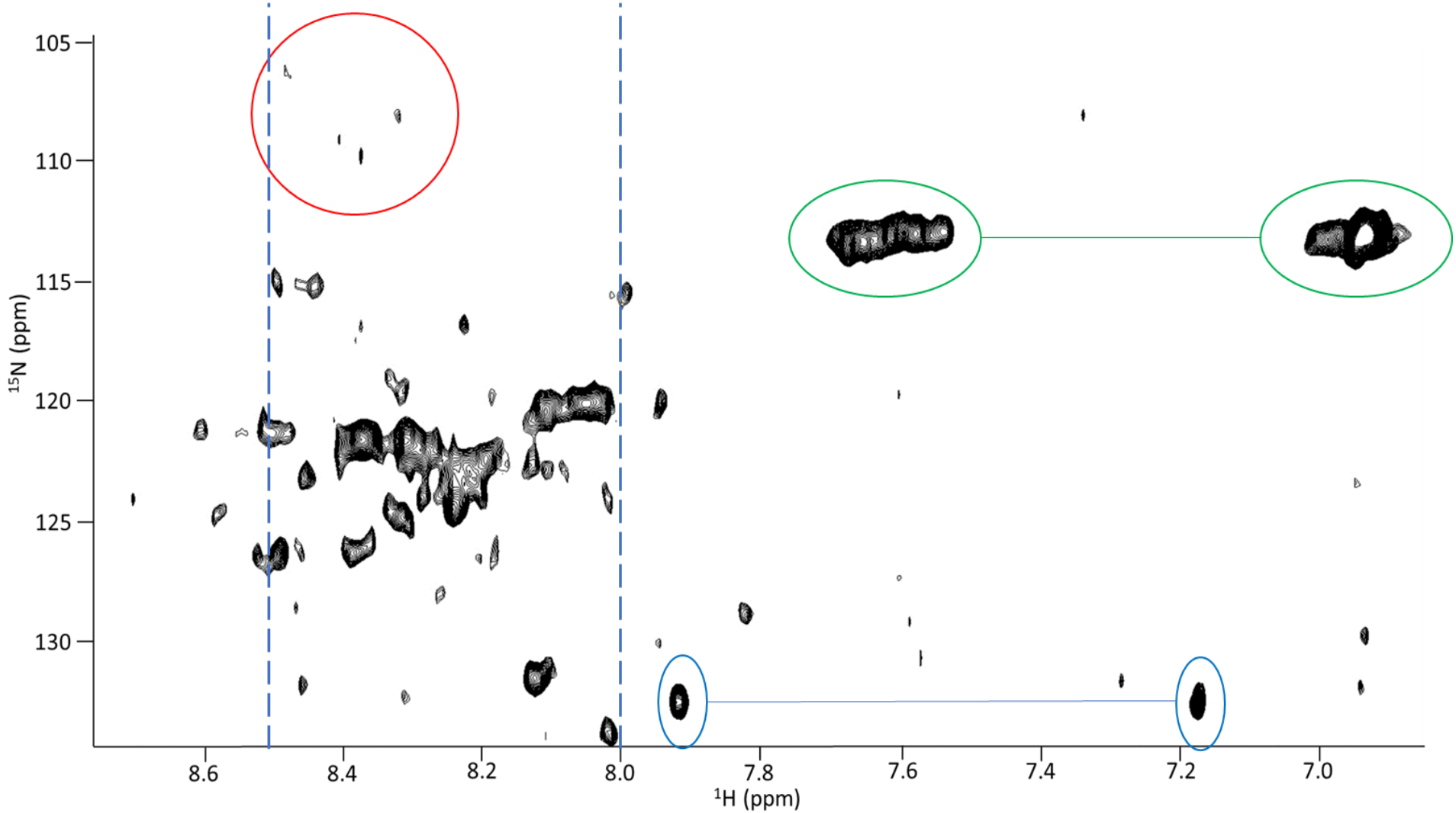


Entirety of the actin binding curve in triplicate including averages.



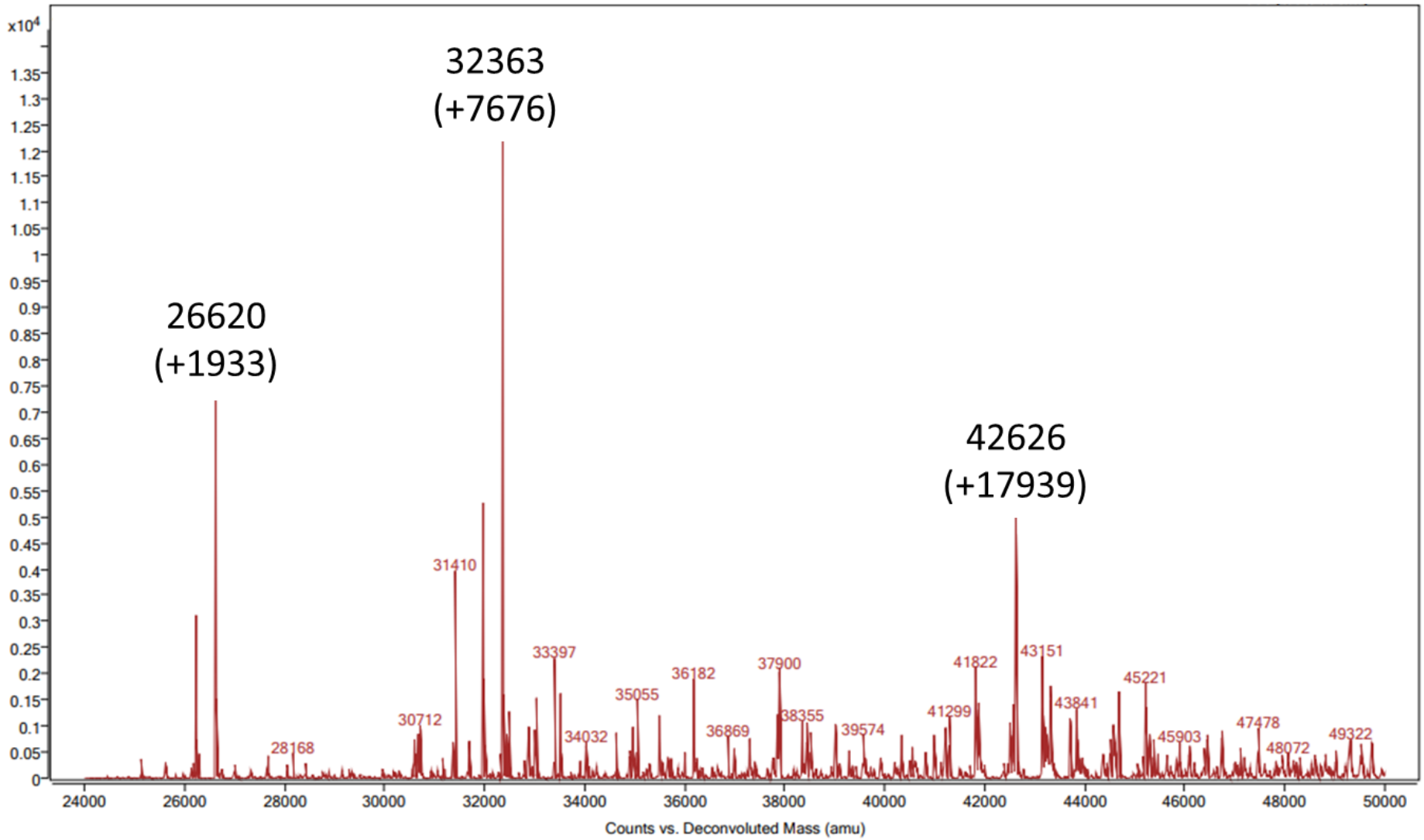
Schematic of the rotor used to conduct the pull-down assay. In red is the rubber O-Ring which I discovered was missing.

Appendix IV. HSQC of ^{15}N VL147N

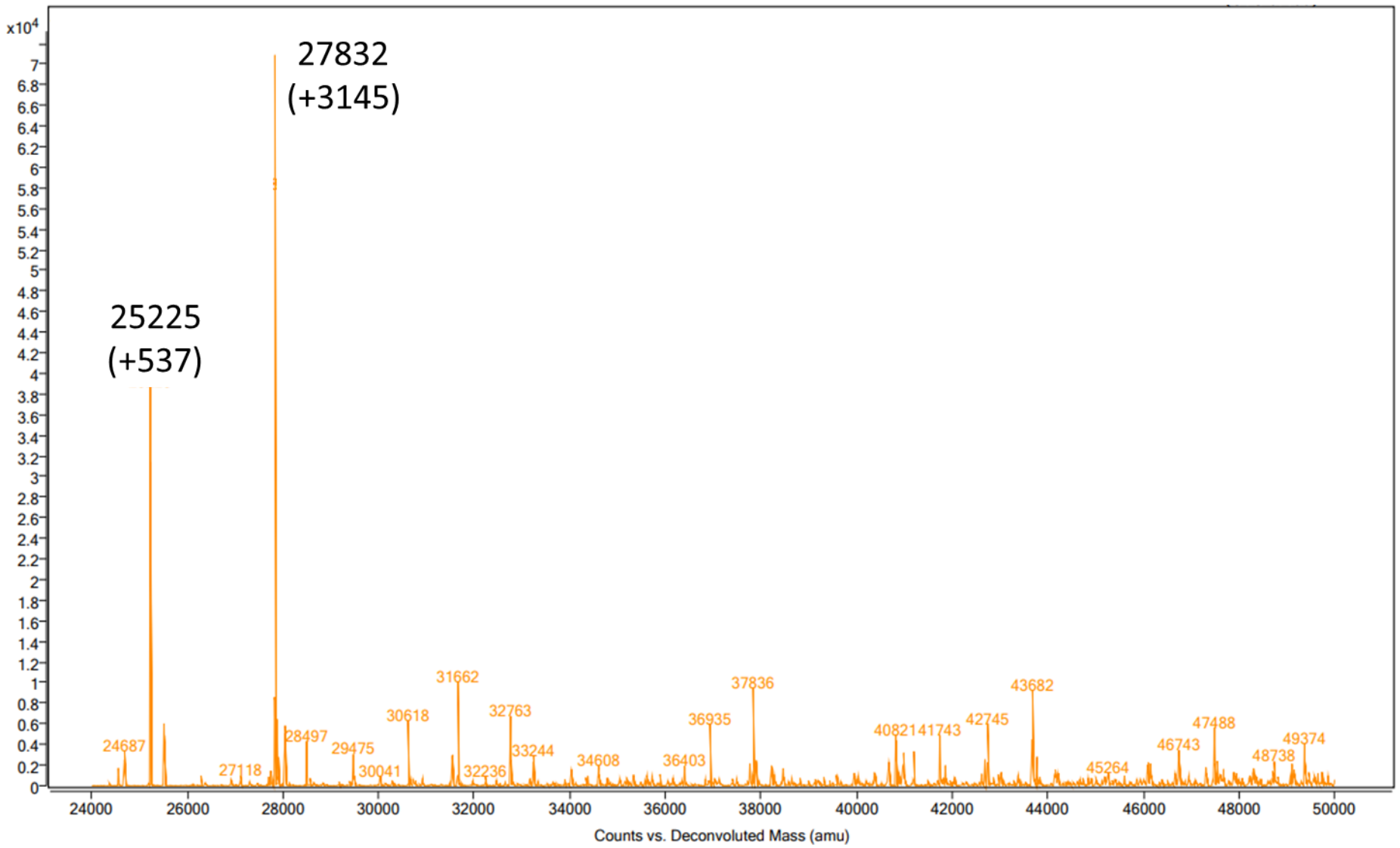


Full HSQC of ^{15}N VL147N. Red circled area is the location of glycine residues, green circled area is the location of the NH_2 backbone peaks, blue circled area is the location of two imidazole peaks, blue dashed lines indicate the area which is often corresponded with disordered substrates.

Appendix V. Deconvoluted spectra of VL192 after exposure to A. *thaliana*



Deconvoluted spectra of VL192 after exposure to *A. thaliana* at timepoint 2.5 minutes of the chromatogram.



Deconvoluted spectra of VL192 after exposure to *A. thaliana* at timepoint 3 minutes of the chromatogram

Appendix VI. Sequences for villin constructs

Abbreviated Name	Full Sequence	Protein Design	Purification Tags
atVHP76	MHHHHHHQED AKEGVEDEED LPAHPYDRLK TTSTDPVSDI DVTRREAYLS SEEFKEKFGM TKEAFYKLPK WKQNKFKMAV QLF	His6 - IDR13 - HP63	His6
VL147N*	MMWSHPQFEK GSGGASWSHP QFEKGGGHHH HHHHHHHGGG ENLYFQSSKS AMHGNSFQRK LKIVKNGGTP VADKPKRRTP ASYGGRASVP DKSQQRSRSM SFSPDRVRVR GRSPAFNALA ATFESQNARN LSTPPPVVRK LYPRSVTPDS SKFAPAPKSS AIASRSALFE KIPPQEPSIP KPVKASPKTP ESPGRA (Extra methanine on the sequence)	Strep - GSGGAS - Strep - G3 - His10 - G3 - TEV - IDR147 - GRA	Strep, His10
VL147N tagless	SSKSAMHGNS FQRKLKIVKN GGTPVADKPK RRTPASYGGR ASVPDKSQQR SRSMSFSPDR VRVRGRSPAF NALAATFESQ NARNLSTPPP VVRKLYPRSV TPDSSKFAPA PKSSAIASRS ALFEKIPPQE PSIPKPVKAS PKTPESPGRA	IDR147 - GRA	N/A
VL192*	MMWSHPQFEK GSGGASWSHP QFEKGGGDSS KSAMHGNSFQ RKLKIVKNGG TPVADKPKRR TPASYGGRAS VPDKSQQRSR SMSFSPDRVR VRGRSPAFNA LAATFESQNA RNLSTPPPVV RKLYPRSVTP DSSKFAPAPK SSAIASRSAL FEKIPPQEPS IPKPVKASPK TPEAAAGGGA GKEQEEKKEN DKEEGSMSSR IESLTIQEDA KEGVEDEEDG GHHHHHHHH (Extra methanine on the sequence)	Strep - GSGGAS - Strep - G3 - IDR192 - G2 - His8	Strep, His8

*The di-methionine in this sequence came about during a miscommunication with the vendor who provided the plasmids. In mass analysis two species will populate regardless of how the constructs are transformed, expressed, and purified. One will be the mass of the construct without the initial methionine's and the other will be the mass of the construct with one methionine (+131.2 Da).

Appendix VII. Sequences for SML constructs

Abbreviated Name	Full Sequence	Protein Design	Purification Tags
ProN	MPSVQEVEKL LHVLDRNGDG KVSAEELKAF ADDSKCPLDS NKIKAFIKEH DKNKDGKLDL KELVSILSSG TSENLYFQGG GGGSGGGGSP PPPPLPTTGG HHHHHH	FH8 - TEV - (G4S)2 - PPPPP - LPTTG - G - His6	FH8, His6
ProC	MPSVQEVEKL LHVLDRNGDG KVSAEELKAF ADDSKCPLDS NKIKAFIKEH DKNKDGKLDL KELVSILSSG TSENLYFQGG GGGSGGGGSK EELHLPMTGG HHHHHH	FH8 - TEV - (G4S)2 - KEELH - LPMTG - G - His6	FH8, His6
AroN	MPSVQEVEKL LHVLDRNGDG KVSAEELKAF ADDSKCPLDS NKIKAFIKEH DKNKDGKLDL KELVSILSSG TSENLYFQGG GGGSGGGGSW HIWWLPITGG HHHHHH	FH8 - TEV - (G4S)2 - WHIWW - LPITG - G - His6	FH8, His6
AroC	MPSVQEVEKL LHVLDRNGDG KVSAEELKAF ADDSKCPLDS NKIKAFIKEH DKNKDGKLDL KELVSILSSG TSENLYFQGG GGGSGGGGSN PEAPLPVTGG HHHHHH	FH8 - TEV - (G4S)2 - NPEAP - LPVTG - G - His6	FH8, His6
AcidN	MPSVQEVEKL LHVLDRNGDG KVSAEELKAF ADDSKCPLDS NKIKAFIKEH DKNKDGKLDL KELVSILSSG TSENLYFQGG GGGSGGGGSD DEEDLPSTGG HHHHHH	FH8 - TEV - (G4S)2 - DDEED - LPSTG - G - His6	FH8, His6
AcidC	MPSVQEVEKL LHVLDRNGDG KVSAEELKAF ADDSKCPLDS NKIKAFIKEH DKNKDGKLDL KELVSILSSG TSENLYFQGG GGGSGGGGSP SANALPNTGG HHHHHH	FH8 - TEV - (G4S)2 - PSANA - LPNTG - G - His6	FH8, His6
BasN	MPSVQEVEKL LHVLDRNGDG KVSAEELKAF ADDSKCPLDS NKIKAFIKEH DKNKDGKLDL KELVSILSSG TSENLYFQGG GGGSGGGGSK RKRRLPVTGG HHHHHH	FH8 - TEV - (G4S)2 - KRKRR - LPVTG - G - His6	FH8, His6
BasC	MPSVQEVEKL LHVLDRNGDG KVSAEELKAF ADDSKCPLDS NKIKAFIKEH DKNKDGKLDL KELVSILSSG TSENLYFQGG GGGSGGGGSG DPNQLPRTGG HHHHHH	FH8 - TEV - (G4S)2 - GDPNQ - LPRTG - G - His6	FH8, His6
AliN	MPSVQEVEKL LHVLDRNGDG KVSAEELKAF ADDSKCPLDS NKIKAFIKEH DKNKDGKLDL KELVSILSSG TSENLYFQGG GGGSGGGGSL VMVILPPTGG HHHHHH	FH8 - TEV - (G4S)2 - LVMVI - LPPTG - G - His6	FH8, His6
AliC	MPSVQEVEKL LHVLDRNGDG KVSAEELKAF ADDSKCPLDS NKIKAFIKEH DKNKDGKLDL KELVSILSSG TSENLYFQGG GGGSGGGGSV GVRTLPDTGG HHHHHH	FH8 - TEV - (G4S)2 - VGVRT - LPDTG - G - His6	FH8, His6
PolN	MPSVQEVEKL LHVLDRNGDG KVSAEELKAF ADDSKCPLDS NKIKAFIKEH DKNKDGKLDL KELVSILSSG TSENLYFQGG GGGSGGGGSD RIKELPETGG HHHHHH	FH8 - TEV - (G4S)2 - DRIKE - LPETG - G - His6	FH8, His6
PolC	MPSVQEVEKL LHVLDRNGDG KVSAEELKAF ADDSKCPLDS NKIKAFIKEH DKNKDGKLDL KELVSILSSG TSENLYFQGG GGGSGGGGSG FFGNLPQTGG HHHHHH	FH8 - TEV - (G4S)2 - GFFGN - LPQTG - G - His6	FH8, His6
GGGGG	MPSVQEVEKL LHVLDRNGDG KVSAEELKAF ADDSKCPLDS NKIKAFIKEH DKNKDGKLDL KELVSILSSG TSENLYFQGG GGGSGGGGS GGGGGLPETGG HHHHHH	FH8 - TEV - (G4S)2 - GGGGG - LPETG - G - His6	FH8, His6
ProN G	MPSVQEVEKL LHVLDRNGDG KVSAEELKAF ADDSKCPLDS NKIKAFIKEH DKNKDGKLDL KELVSILSSG TSENLYFQGG GGGSGGGGSP PPPGLPTTGG HHHHHH	FH8 - TEV - (G4S)2 - PPPPG - LPTTG - G - His6	FH8, His6
AroN G	MPSVQEVEKL LHVLDRNGDG KVSAEELKAF ADDSKCPLDS NKIKAFIKEH DKNKDGKLDL KELVSILSSG TSENLYFQGG GGGSGGGGSW HIWGLPITGG HHHHHH	FH8 - TEV - (G4S)2 - WHIWG - LPITG - G - His6	FH8, His6

AcidN G	MPSVQEVEKL LHVLDRNGDG KVSAEELKAF ADDSKCPLDS NKIKAFIKEH DKNKDGKLDL KELVSILSSG TSENLYFQGG GGGSGGGGSD DEEGLPSTGG HHHHHH	FH8 - TEV - (G4S)2 - DDEEG - LPSTG - G - His6	FH8, His6
PolN G	MPSVQEVEKL LHVLDRNGDG KVSAEELKAF ADDSKCPLDS NKIKAFIKEH DKNKDGKLDL KELVSILSSG TSENLYFQGG GGGSGGGGSD RIKGLPETGG HHHHHH	FH8 - TEV - (G4S)2 - DRIKG - LPETG - G - His6	FH8, His6
ProN 2G	MPSVQEVEKL LHVLDRNGDG KVSAEELKAF ADDSKCPLDS NKIKAFIKEH DKNKDGKLDL KELVSILSSG TSENLYFQGG GGGSGGGGSP PPGGLPTTGG HHHHHH	FH8 - TEV - (G4S)2 - PPPGG - LPTTG - G - His6	FH8, His6
AroN 2G	MPSVQEVEKL LHVLDRNGDG KVSAEELKAF ADDSKCPLDS NKIKAFIKEH DKNKDGKLDL KELVSILSSG TSENLYFQGG GGGSGGGGSW HIGGLPITGG HHHHHH	FH8 - TEV - (G4S)2 - WHIGG - LPITG - G - His6	FH8, His6
AcidN 2G	MPSVQEVEKL LHVLDRNGDG KVSAEELKAF ADDSKCPLDS NKIKAFIKEH DKNKDGKLDL KELVSILSSG TSENLYFQGG GGGSGGGGSD DEGGLPSTGG HHHHHH	FH8 - TEV - (G4S)2 - DDEGG - LPSTG - G - His6	FH8, His6
PolN 2G	MPSVQEVEKL LHVLDRNGDG KVSAEELKAF ADDSKCPLDS NKIKAFIKEH DKNKDGKLDL KELVSILSSG TSENLYFQGG GGGSGGGGSD RIGGLPETGG HHHHHH	FH8 - TEV - (G4S)2 - DRIGG - LPETG - G - His6	FH8, His6