Summer 2022

Sortase-Mediated Ligation to Investigate Large IDR-Containing Cytoskeletal Regulators

Melissa Oueld Es Cheikh
Western Washington University, oueldemelissa@gmail.com

Sortase-Mediated Ligation to Investigate Large IDR-Containing Cytoskeletal Regulators

By

Melissa Oueld Es Cheikh

Accepted in Partial Completion

Of the Requirements for the Degree

Master of Science

ADVISORY COMMITTEE

Dr. Sergey Smirnov, Chair

Dr. John Antos

Dr. Spencer Anthony-Cahill

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.

Melissa Oueld Es Cheikh

August 1st , 2022

Sortase-Mediated Ligation to Investigate Large IDR-Containing cytoskeletal Regulators

A Thesis
Presented to
The Faculty of
Western Washington University

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science

By
Melissa Oueld Es Cheikh
August 2022

Abstract:

About 43% of eukaryotic proteins contain intrinsically disordered regions (IDRs) of 40 residues or longer. These proteins modulate key cellular functions such as transcription, translation, and signal transduction. Despite their vital roles, little is known about the properties of large disordered proteins due to the difficulties encountered studying them. Solution nuclear magnetic resonance (NMR) spectroscopy is a unique tool allowing for their analysis but traditionally requires uniform protein labeling with ^{15}N and ^{13}C isotopes resulting in convoluted spectra which obscure any meaningful information on their structures and interactions. To remedy this issue, we aim at developing a procedure to segmentally label large IDR-containing proteins by synthesizing them as separate fragments before ligating them together using sortase enzymes. This method would allow for the recording of $^{15}N/^{13}C$-HSQC experiments of smaller IDR/IDP segments at a time, avoiding spectral overlap while retaining the proteins' native biochemical environment. We focus our work on cytoskeletal regulators, dematin (*Homo sapiens*) and villin 4 (*Arabidopsis thaliana*) each comprising a large IDR. Both dematin and villin 4 bind and control the bundling of filamentous actin (F-actin) but their mechanism of action remains unclear. We plan to use segmental labeling (SL) to gain insight into their respective properties which would help advance both the medical (dematin) and agricultural (villin 4) fields.

Acknowledgments:

I would like to take the time to thank the National Science Foundation (NSF) for funding the first two aims described in this thesis (award # 2004237) as well as Western Washington University's office of Research and Sponsored Program (RSP) for funding the experiments encompassing the third aim of this project.

I would also like to thank Dr. Smirnov and Dr. Antos from the bottom of my heart for all the support and mentorship they provided for me throughout my undergraduate and graduate studies. Be it in a classroom or research setting, their high expectations and challenging coursework has allowed me to develop a strong foundation in both Biochemistry and Organic Chemistry. Under their co-supervision, I was always encouraged to strive for excellence, they believed in me and gave me the opportunity to realize my full potential during the covid-19 pandemic when many would have been reluctant to do so.

A huge thank you to Dr. Anthony-Cahill for having agreed to be part of my graduate committee in addition to being an exceptional professor. One of my biggest regret is not having taken more classes for which he was the instructor due to scheduling conflicts. I think I can speak for all graduate students when I say that no M.S Biochemistry thesis would have been written this year without the cookies and candy Dr. Anthony-Cahill frequently brought to the lab as fuel for our motivation.

Finally, I would like to thank all my lab mates for their wonderful teamwork and dependability. Specifically, I would like to thank Erich Walkenhauer, Micah Lund, and Darmon Ghanbari for their friendships which made fourteen-hour workdays go by very quickly.

Table of Contents:

List of Figures and Tables:

Chapter I: Introduction

1.1 Intrinsically Disordered Regions/Proteins (IDRs/IDPs):

Structural biologists long considered a tertiary structure indispensable for proteins to be able to perform their various functions[1]. However, numerous polypeptides studied since this model was established in the 1960s have come to challenge this notion. Intrinsically disordered regions/proteins (IDRs/IDPs) are characterized by the absence of a stable defined three-dimensional structure. In fact, they often adopt a variety of conformational states which are in dynamic equilibrium with one another under physiological conditions[1,2]. Moreover, recent estimates report that close to 43% of proteins within eukaryotes contain IDRs larger than 40 residues in length[3]. Their flexible nature, capacity to bind several partners simultaneously with great specificity yet low affinity, along with their tendency to undergo post-translational modifications allow IDRs/IDPs to act as mediators in a multitude of signaling pathways through allosteric changes[4]. These same features make IDRs/IDPs accessory proteins which aid in the organization of macromolecule complexes such as chromatin, the ribosome, and the cytoskeleton[4].

In addition to their roles as central protein hubs, IDRs/IDPs make up a large portion of the mRNA splicing and post-translational silencing machineries[4,5]. Finally, unstructured proteins are known to promote phase separation which leads to the formation of membrane-less organelles in the cytoplasm[6], thus contributing to the necessary compartmentalization of biochemical processes in the cell.

Because of their vital cellular functions, mutations in the amino acid sequence of IDRs which alter their level of disorder are associated with several cancers and neurodegenerative disorders such as such as Alzheimer's, Parkinson's, and Huntington's diseases[1,7,8].

1.2 Challenges of Studying IDRs/IDPs:

In spite of their biological significance, IDRs/IDPs remain a mysterious class of proteins. IDRs' dynamic nature exposes the limitation of many characterization methods such as X-ray crystallography[1,9], able to only capture a single conformation at a time, an incomplete and often misleading picture as it is not representative of the various structures IDRs adopt *in vivo,* not to mention IDR's inherent inability to form crystals. Unlike X-ray crystallography, solution NMR spectroscopy is a powerful tool able to characterize the various forms and interactions of IDRs[9]. However, some limitations do persist: long recording time, especially for 3D heteronuclear NMR which is problematic because IDRs are prone to degradation via proteolytic cleavage[10].

NMR experiments also require protein expression in minimal media which often results in low yield[9]. If a purification tag is included, it may also interfere with function of these disordered segments. Therefore, careful consideration should be given to the construct design and purification scheme.

Although the manipulation of experimental parameters may help mitigate some of the above inconveniences, one major hurdle continues to pose a serious challenge to the acquisition of meaningful NMR data on IDRs, spectral overlap[9,11]. Because the entire IDR/IDP is typically isotopically labeled, the resulting spectrum is plagued with significant signal overlap, impeding

the peak assignment and thus structure characterization process. One way to circumvent the problem would be to label small segments within the IDR at a time, leaving the rest of the protein unlabeled, allowing for a considerable decrease in the number of peaks that would appear on the spectrum while preserving the amino acid context of studied segment. To produce a segmentally labeled construct, the IDR/IDP needs to be synthesized in separate fragments (**Figure 1.2.1**). The fragments would then be covalently linked through a variety of mechanisms. Segmental labeling was successfully accomplished to analyze the properties of folded proteins[12]. To the best of our knowledge, segmental labeling (SL) has yet to be applied to the characterization of IDRs/IDPs.

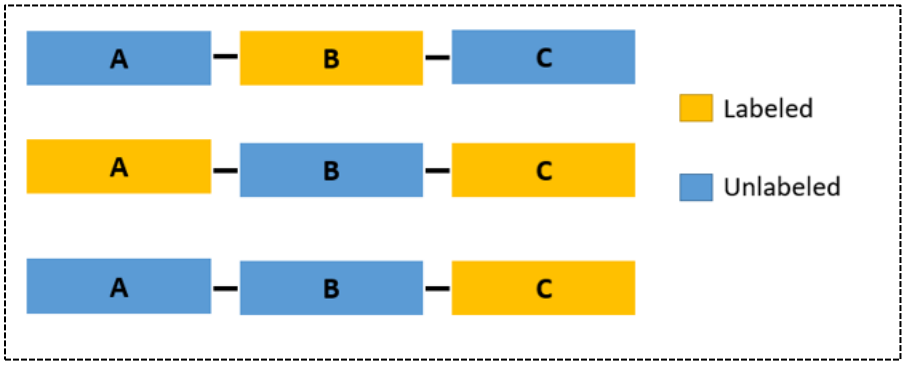


Figure 1.2.1. Segmental labelling (SL) of protein segments (A, B, or C) according to experimental needs. Labeled fragments are indicated in yellow, blue shows unlabeled protein portions.

Traditionally, SL is achieved through a combination of native chemical ligation (NCL)[13], expressed protein ligation (EPL)[13,14], or protein trans-splicing (PTS) mechanisms[12–14]. These established strategies possess a number of limitations in terms of their design. For both NCL and EPL, the presence of a N-terminal cysteine and a C-terminal thioester is critical to the success of the ligation reaction. When introducing a cysteine to a wild-type sequence lacking it to allow for the ligation to occur, the potential for unwanted disulfide interactions increases,

leading to possible alteration of the protein structure and function[15,16]. While desulfurization may be a suitable solution to this problem, it will likely disturb the network of native disulfide bonds, ultimately affecting protein shape and activity[15–17]. Finally, many purification steps associated with the above methods result in significant losses in yield[17], emphasizing the need for novel strategies to address these current hurdles. Thus, we propose sortase-mediated ligation (SML) as an additional SL procedure. Our current work focuses on using SML to segmentally label various IDRs/IDPs under investigation. As an eventual expansion to our project, we plan to incorporate EPL to our synthesis scheme to allow for the isotopic labelling of internal protein segments (**Figure 1.3.2**).

1.3 Sortase-Mediated Ligation (SML):

Sortase enzymes are transpeptidases found in bacterial cells[18]. They recognize and covalently attach surface proteins onto the peptidoglycan layer in gram-positive bacteria, thus playing a key role in the virulence of these pathogens. Depending on the organism from which they are derived, sortase variants perform different roles and recognize different sorting signals[19,20]. Sortase A, isolated from *Staphylococcus aureus*, targets the LPXTG sequence motif and cleaves between the threonine (T) and glycine (G) residues[19] (**Figure 1.3.1**). It proceeds to catalyze the reversible ligation reaction between the C-terminal end of this motif (LPXT) and an N-terminal glycine nucleophile present at the surface of the bacterial cell wall. Many known sortase proteins have been genetically engineered to enhance their affinity to particular substrates as well as their specificity and ligation efficiency[20–22]. In this thesis, we utilize wild-type sortase A (WT-srtA) and the hepta-mutant variant (7m-srtA) which compared to WT-srtA, does not depend on calcium for its activity and has been shown to have greater catalytic activity[21]. If the

IDR/IDP of interest contains a sortase recognition sequence either natively or through the introduction of conservative/minimally-disruptive mutations, SML could be utilized to synthesize a segmentally labeled sample suitable for NMR studies. Additionally, when combined with other ligation methods, such as EPL, it would allow for greater flexibility in terms of which fragment gets labeled (**Figure 1.3.2**).

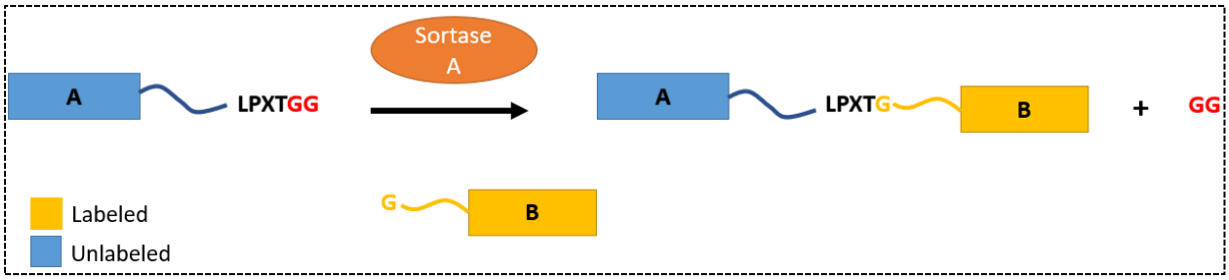


Figure 1.3.1. Synthesis of a segmentally isotopically labeled protein via SML using sortase A (*Staphylococcus aureus).*

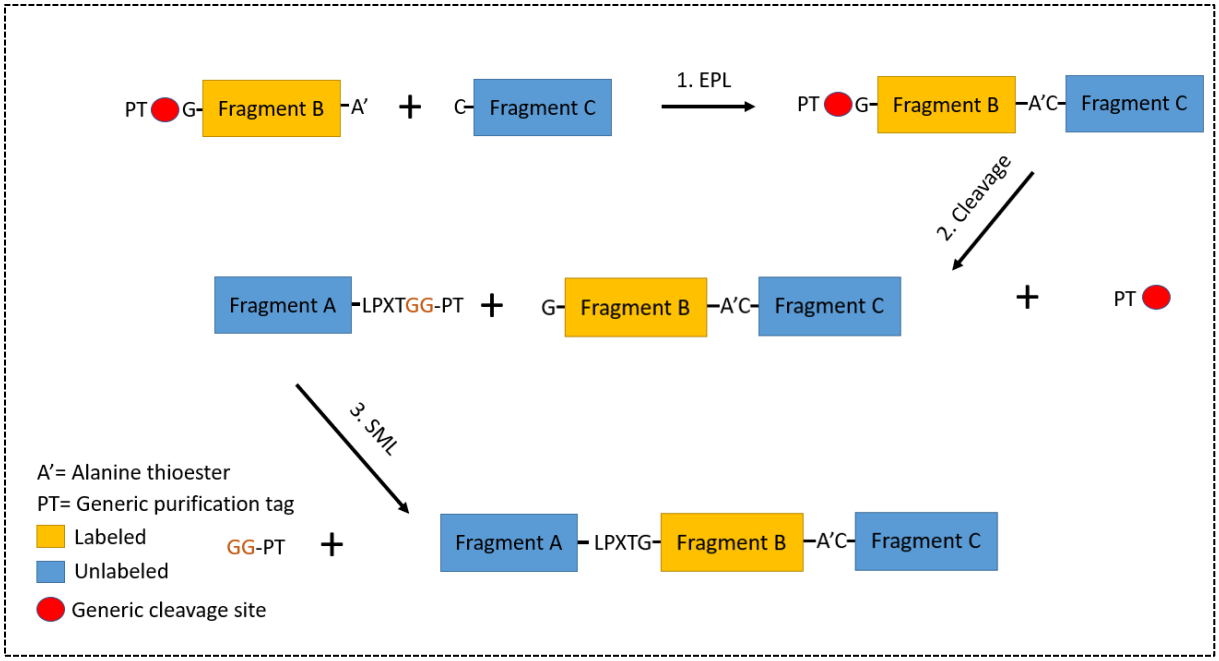


Figure 1.3.2. Tentative synthesis scheme to segmentally label a generic protein using expressed protein ligation (EPL) in combination with sortase-mediated ligation (SML). EPL requires an N-terminal cysteine and a C-terminal thioester at the ligation site. A cleavage step following the EPL step allows for the exposure of an N-terminal glycine nucleophile necessary for SML. Sortase enzymes used in our lab cleave between the threonine and glycine within the LPXTG motif, allowing for the N-terminal glycine generated in step 2 required for SML to occur.

1.4 The Cytoskeleton:

In eukaryotic cells, the cytoskeleton is the structure responsible for maintaining cell shape and internal organization[23]. It controls a multitude of important cellular processes from motility and adhesion to cytokinesis and cell polarity. It comprises microtubules, microfilaments, and intermediate filaments[24]. Microfilaments are composed of two interweaved strands of globular actin (G-actin) monomers[23]. Because of the way G-actin proteins are oriented, the resulting actin filaments (F-actin) are endowed with polarity and distinguishable plus (+) and minus (-) ends, a feature which determines the direction of myosin movement in the cell. Such an interaction between F-actin and myosin provides the molecular basis for muscle contraction[25].

Actin filaments are highly dynamic, constantly assembling and disassembling depending on the concentration of free G-actin proteins[23]. Upon aggregation of a few G-actin monomers, the spontaneous elongation of F-actin is initiated at both ends of the polymer. G-actin subunits also bind and hydrolyze ATP upon their association such that the G-actin-ADP complex disassociates more easily than the G-actin-ATP pair, resulting in treadmilling. This phenomenon refers to the difference in the polymerization rate between the two extremities as the plus end grows faster than the minus end.

In the cell, actin filaments are either arranged into bundles or networks[23] (**Figure 1.4.1**). In bundles, F-actin is crosslinked in a parallel pattern whereas in networks, it is loosely crosslinked in a perpendicular fashion, forming a gel-like meshwork. The formation of these structures is modulated by actin binding proteins (ABP) which assemble F-actin into these different configurations in response to changing environmental stimuli. In fact, ABPs control nearly all aspects of microfilament formation. Examples of actin-binding proteins include the villin family and its homologs.

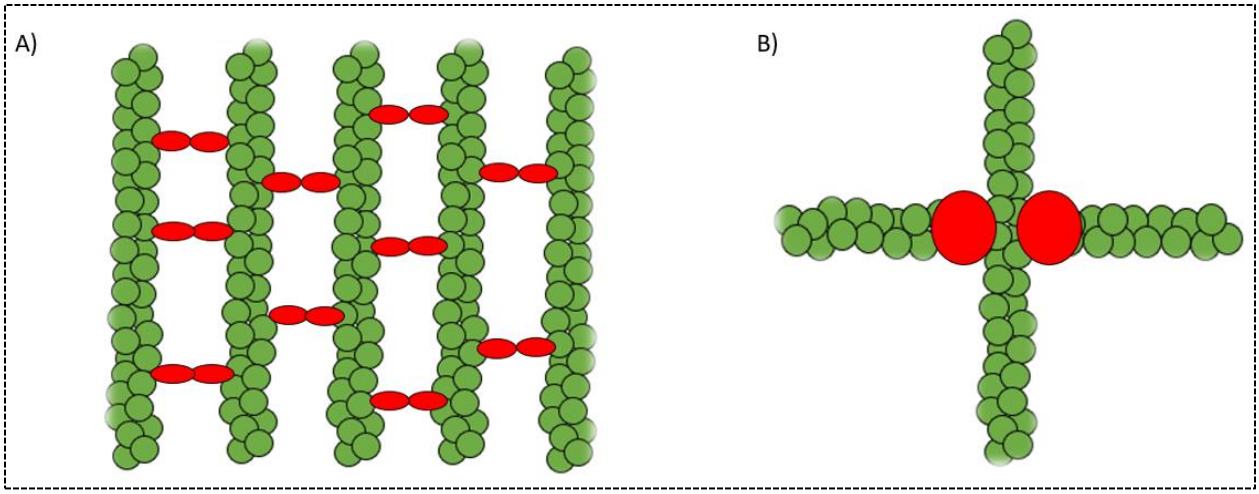


Figure 1.4.1[23]**.** Schematic representation of F-actin organization in eukaryotic cells. A) Parallel F-actin bundles held by various ABPs. Typically, a space of 14 nm separates microfilaments from each other[23]. B) F-actin cross-linked into perpendicular arrangements through the catalytic activity other ABPs.

1.5 Microfilament Regulators:

There are numerous proteins that function as F-actin regulators. This section provides a general introduction to a few ABPs relevant to the content presented in this document: gelsolin, supervillin, villin, and dematin.

1.5.1 Gelsolin:

Gelsolin regulates F-actin assembly and disassembly through its ability to sever and cap filamentous actin[26]. In animals lacking the gelsolin gene, significant disruptions in wound healing and clotting mechanisms were observed, further asserting the vital role this protein plays in many cellular processes. Gelsolin is composed of six repeat domains, S1-S6 (**Figure 1.5.1).** The C-terminal portion of gelsolin binds F-actin when calcium levels exceed 1 µM and acts as a modulator to the N-terminal moiety capable of capping and severing F-actin, irrespective of calcium concentrations. X-ray crystallography studies revealed that in the

absence of calcium the S4 and S6 segments combine into a single beta sheet but upon calcium introduction, the secondary structure disappears as S4 detaches from S6 and binds to S5 and actin. These findings support the role calcium plays in inducing the conformational changes necessary to trigger gelsolin's capping and severing functions. Interestingly, through initiating F-actin depolymerization, gelsolin may be a major a player in F-actin nucleation as its activity results in an increase of F-actin minus ends and free-floating actin monomers.

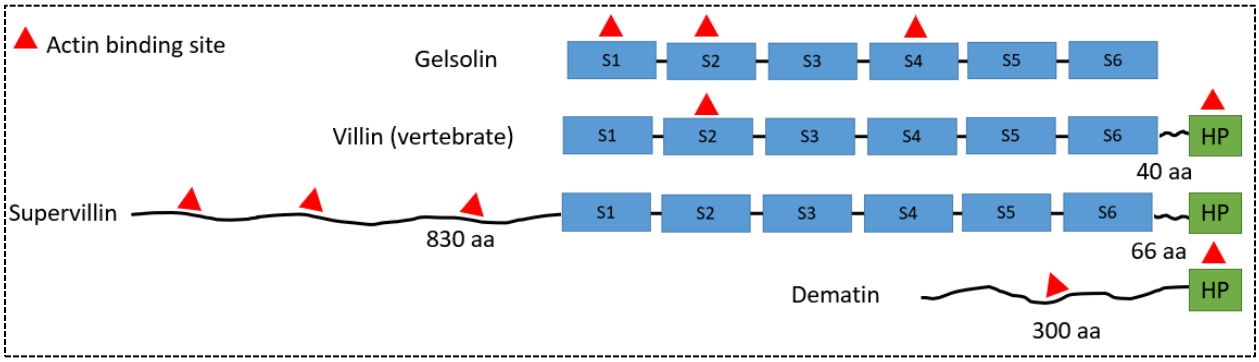


Figure 1.5.1. Comparison of some actin-binding proteins belonging to the villin family. Villin, gelsolin, and supervillin all possess the characteristic 6 tandem gelsolin domains while dematin lacks them. With the exception of gelsolin, all villin-like proteins contain a folded headpiece domain capable of binding actin. In supervillin however, F-actin binding sites are located within its exceptionally-large 830-residue IDR.

1.5.2 Villin:

The villin class belongs to a group of actin-binding proteins known as the villin/gelsolin/fragmin family[27]. These proteins share significant sequence homology in terms of their 6 successive gelsolin domains, S1-S6 (**Figure 1.5.1**). Unlike gelsolin, villin proteins are distinguishable by the presence of a folded headpiece domain on their C-terminal end, also found in supervillin and dematin. The headpiece is attached to the signature gelsolin core via an unstructured segment

(IDR). This folded domain allows villin to assemble actin filaments into bundles in a calcium-sensitive manner. In vertebrates, villin is found in microvilli of epithelial cells both in the small intestine and kidneys[28] . In plants, however, different villin isoforms are distributed throughout various tissues and contribute to maintaining larger structures than their vertebrate counterparts[27,28]. Despite having a similar role, plant villin proteins share little sequence homology with their vertebrate equivalents in the IDR connecting their gelsolin core to their headpiece domain. It is unclear how these differences affect their respective mode of operation.

1.5.2a. Bovine Supervillin:

Supervillin, a member of the villin family, controls many aspects of cell motility, division, adhesion, and communication[23,29]. In addition to possessing the six gelsolin domains (S1-S6) and the characteristic villin headpiece, it is composed of a large N-terminal IDR of 830 residues[29] (**Figure 1.5.1**). While it retains its capacity to bundle microfilaments, it does so through its N-terminal segment because a change in sequence inhibits such activity at the C-terminal end, the typical F-actin bundling site for other villin proteins. Along with filamentous actin, supervillin interacts with Myosin II and around 70 other signaling macromolecules. Computational analysis combined with actin co-sedimentation assays established the presence of three actin binding sites on the A1, A2, and A3 segments, as well as two myosin II sites within the M portion of the IDR (**Figure 1.5.2**).

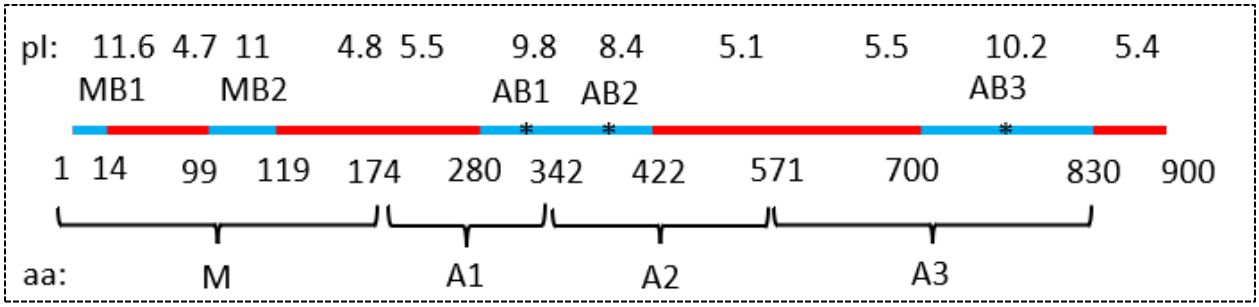


Figure 1.5.2[29]**.** Location of F-actin and myosin I binding sites on the 830-residue N-terminal IDR of bovine supervillin. Segments containing F-actin binding sites (AB1, AB2, AB3) are depicted in blue along with Myosin I binding sites (MB1, MB2), all are composed of basic amino acids. Proposed actin-binding sites are highlighted with asterisks (*). Theoretical pI values for relevant segments as well as the amino acid positions are indicated above and below the IDR, respectively. Image reproduced based on Dr. Sergey Smirnov's original figure.

Both Actin and myosin II are predicted to carry an overall negative charge at physiological pH (**Figure 1.5.3**), given their respective pI values of 5.5 and 4.3[29]. Contrastingly, bovine supervillin remains positively charged under the same conditions. In fact, supervillin fragments shown to bind actin were characterized by pI values of 9.8, 8.4, and 10.2 for AB1, AB2, AB3, respectively, making them quite positively charged at neutral pH (**Figure 1.5.2**). This difference in charges suggests an electrostatic interaction as the main mechanism through which supervillin binds F-actin. Such a hypothesis is equally plausible in the context of Myosin II association. Indeed, progressive increase of salt concentrations has been shown to lower supervillin's affinity for Myosin II, further supporting the electrostatic model.

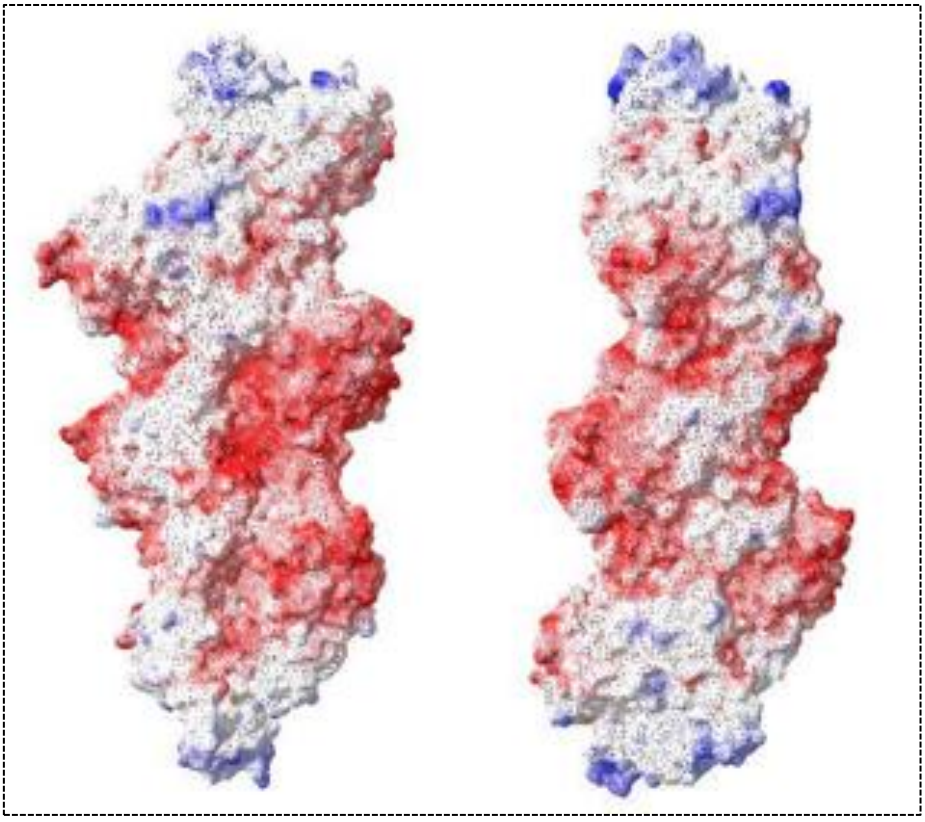


Figure 1.5.3[29]**.** Charge partioning on the surface of microfilaments at physiological pH. Acidic residues are shown in red, basic amino acids are represented in blue. Neutral residues are depicted in gray. Image curtesy of Dr. Sergey Smirnov, WWU.

1.5.2b. Villin 4 (*Arabidopsis thaliana*):

Arabidopsis thaliana contains 5 villin variants expressed in a ubiquitous manner throughout the plant[27]. Villin isoform 4 (AtV4) retains the gelsolin core and the characteristic C-terminal headpiece domain linked together by a 190-residue IDR (**Figure 1.5.4**). AtV4 performs the same functions attributed to the villin family but is especially important to the growth of plant root hair. Plants lacking the villin 4 gene show a decrease in the formation of actin cables, leading to significantly fewer and shorter root hairs. AtV4 binds actin filaments irrespective of the presence of calcium. Its severing and bundling activity however does depend on calcium levels in the cell. In normal physiological conditions (<10 μM Ca^{2+}) AtV4 promotes the formation of

actin bundles[27,28]. On the other hand, when the levels of calcium increase to about 50-200 μM, its severing function is favored, and the bundles found at higher Ca^{2+} concentration were significantly shorter, suggesting that villin 4 regulates not only the formation but also the length of actin bundles in response to stressful stimuli[27].

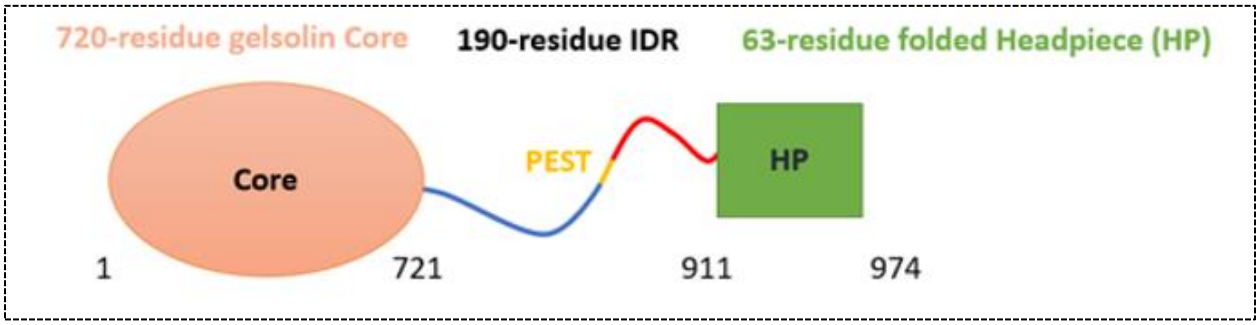


Figure 1.5.4. Structure of villin 4 (*Arabidopsis thaliana*). Six gelsolin domains are shown as the core which spans residues 1-720. The core is linked to the folded headpiece domain (residues 911-974) via a disordered region (721-910). This IDR is composed of a basic segment (blue, residues 721-869) separated by a predicted PEST motif (~870-879) to the acidic portion (red, residues ~880-911)

Analysis by Dr. Smirnov, from the Chemistry department at Western Washington University, of the IDR sequence of villin 4 reveals a distinct charge partitioning. Indeed, the IDR N-terminal portion is enriched in basic residues while its C-terminal segment has a preponderance of acidic amino acids. These two fragments are separated by a predicted PEST motif which is recognized by the cell as a cleavage signal (**Figure 1.5.4**). This same charge partioning is found in the large 830-residue N-terminal IDR of supervillin[29], which was shown to contain Myosin II and F-actin binding sites (**Figure 1.5.2**). In supervillin, as discussed in the above section, these interactions were hypothesized to be driven by electrostatic forces. Because of the similarities in charge separation between supervillin's and villin 4's respective IDRs, comparable intermolecular forces may be responsible for villin 4's capacity to bind F-actin.

The first step in investigating this hypothesis would be to synthesize and purify the full-length villin 4 linker then subject it to an actin pull-down assay. A difficult task which faces all the hurdles associated with the production and analysis of large disordered protein segments mentioned in section 1.2. The propensity of the villin 4 IDR to degradation is heightened by the presence of a predicted PEST motif between its acidic and basic portions which makes it a target for proteases inside the host bacterial cells where it is expressed[30]. Although *in vitro* synthesis could be considered to solve this issue, it is time-consuming; methods such as solid phase peptide synthesis (SPPS) are best suited to produce peptides less than 70 residues long[31]. Furthermore, the frequent incidence of side reactions and multiple purification steps reduce the desired product's yield. An optimized procedure that may utilize SML biotechnology for the synthesis and purification of the large villin 4 IDR is needed. Once produced, several studies can be conducted to decipher how proteins such as villin 4 operate to modulate F-actin assembly, relying on solution NMR spectroscopy and fluorescent microscopy, among other instrumentation techniques.

Previous work our lab conducted on villin 4 (*Arabidopsis thaliana*) perfectly illustrates the great capabilities of sortase within the context of NMR-driven structural characterization of IDPs. The isotopically labeled acidic IDR portion in villin 4 was successfully ligated to an unlabeled villin 4 folded headpiece using sortase enzymology[32]. ^{15}N-HSQC spectra were then recorded at various temperatures (15 ℃, 25 ℃, and 45 ℃). Analysis of the spectra obtained indicated unequal temperature sensitivity across the acidic IDR sequence. Due to the known temperature dependence plants show during various stages of their growth processes, such observations raise several questions on the role of AtV4 IDR in plant maturation. The data collected in these

experiments would have been difficult to interpret had the entire protein been labeled with ^{15}N isotopes due to the challenges outlined in section 1.2.

Surprisingly, the presence of the acidic stretch (EDEED) at the N-terminal end of LPXTG sortase site significantly hindered the ligation efficiency. However, introduction of a glycine spacer between the EDEED sequence and the sortase motif was sufficient to resolve the issue (**Figure 1.5.5**). This unexpected finding calls for further investigation into the effects of various types of neighboring amino acid residues on SML efficiency, specifically in the context of real IDR sequences, the focus of this thesis.

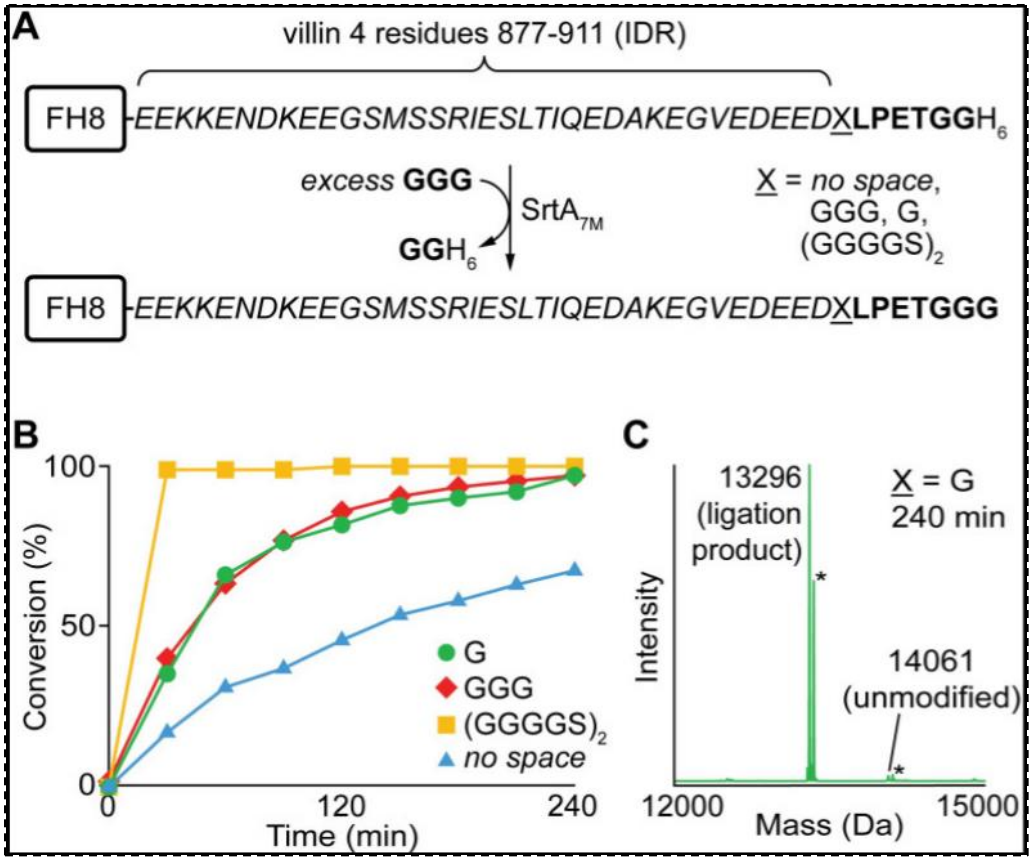


Figure 1.5.5[32]**.** Effect of N-terminal acidic residues on SML efficiency. A) SML between the acidic IDR stretch in villin 4 (*Arabidopsis thaliana*) and a poly-glycine N-terminal nucleophile using a hepta-mutant variant of sortase A. FH8 and His_6 tags are included for purification and solubility enhancement (FH8) purposes. B) SML reaction in A was conducted at room temperature for 4 hours and monitored via LC-ESI-MS. The reaction efficiency increases with the insertion of glycine spacers between the LPXTG ligation site and the IDR acid stretch EDEED. C) Mass spectrum of ligation product containing one glycine as the spacer (13296 Da).

1.5.3 Dematin 2 (*Homo sapiens*):

Dematin (*Homo sapiens*) is an F-actin-binding and bundling protein originally identified in the spectrin-actin junction of erythrocytes[33]. It is also present in human platelets (acting as a calcium modulator) heart, brain, kidney, and lens tissue[34]. It is known to control cell adhesion and motility in these various organs though the negative regulation of the Rho GTPase activation pathway[35]. Dematin (isoform 2) is composed of a 315-residue N-terminal disordered region and a folded C-terminal domain 68-residues long, homologous to the headpiece domain found in the villin family[35,36] (**Figure 1.5.6**). Dematin is regulated through the phosphorylation of its headpiece domain at the Ser381 position by a cAMP-dependent protein kinase[36]. In this manner, dematin differs from other large-IDR-containing cytoskeletal regulators which tend to undergo covalent modifications on their unstructured segments. It also lacks the folded N-terminal core characteristic of villin-like proteins but still manages to interact with F-actin through two binding sites, one located within its IDR and the other found on its headpiece. While its phosphorylation state does not impact its F-actin-binding properties, its F-actin-bundling capability is abolished upon headpiece phosphorylation[34–36].

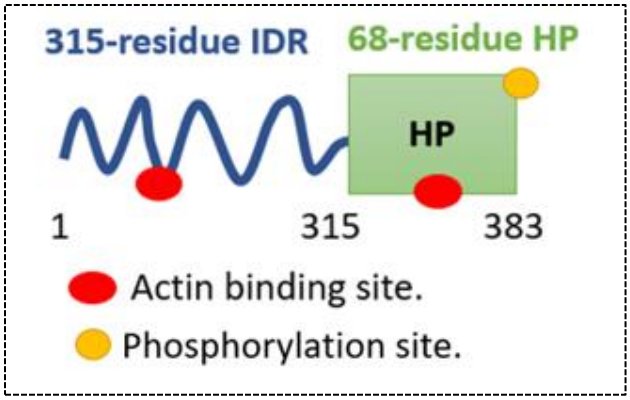


Figure 1.5.6. Structure of dematin (*Homo sapiens*). A) Dematin has preserved the folded headpiece domain characteristic of villin proteins but lacks the gelsolin domain. It has a large disordered region consisting of the first 315 N-terminal residues. It has two known actin-binding sites, one on its IDR and one on the headpiece domain. One way to regulate its activity is through phosphorylation of S381. Point mutation S381E was shown to mimic the phosphorylated state in terms of both structure and function.

Recent NMR studies of the dematin suggest that the phosphorylation of the headpiece may trigger an intramolecular binding event between the headpiece and its IDR[23] but the binding interface remains uncharacterized (1.5.7). Such an interaction is further suggested by the presence of the headpiece signature peaks irrespective of its phosphorylation state when its 102 N-terminal IDR portion is truncated. Peaks which proceed to disappear upon analysis of the phosphorylation mimic and only do so in the presence of the full dematin IDR (**Figure 1.5.8**).

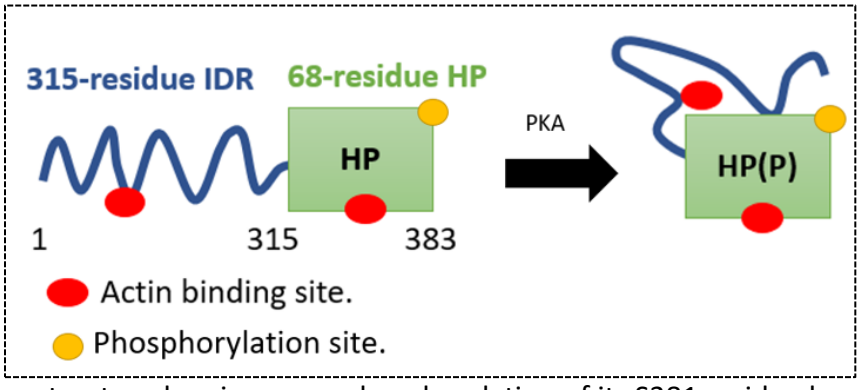


Figure 1.5.7. Hypothesized intramolecular interaction within dematin. Based on ^{15}N-HSQC data[36], the headpiece domain may be involved in an intramolecular binding interaction with the unstructured region upon phosphorylation of its S381 residue by protein kinase A (PKA). The nature of this interaction, as well as the binding interface, remains unknown.

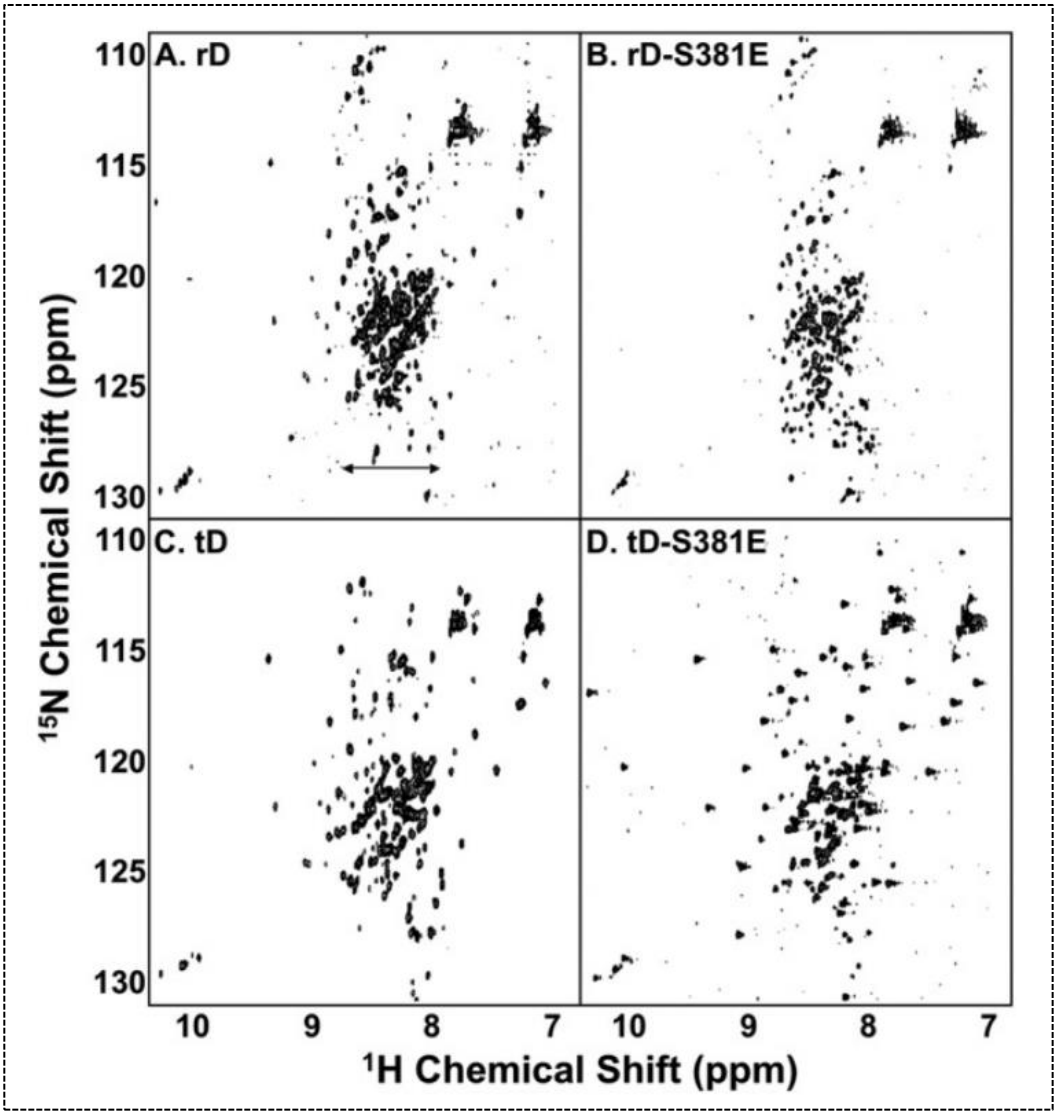


Figure 1.5.8[36]**.** ^{15}N-HSQC of uniformly labeled dematin *2* (*Homo sapiens*). All spectra were collected at 500MHz and 20 ℃. Dematin concentrations ranged from 20-50 µM. Buffer composition for all samples consisted of 200 mM NaCl, 10 mM phosphate 10% D_2O at pH 7.0. A) Full-length wild-type (WT) dematin (recombinant, rD). B) full-length phosphorylation mimic dematin (rD-S381E). C) WT dematin lacking the first 102 N-terminal IDR residues (truncated, tD). D) phosphorylation mimic dematin missing the same 102 N-terminal portion (tD-S381E). Headpiece signature peaks disappear upon phosphorylation only when the 102 N-terminal IDR segment is present**.**

Disruptions of the dematin gene is associated with multiple health disorders. Mice in which the dematin headpiece was truncated produced osmotically fragile red blood cells with compromised semi-permeability leading to their developing hemolytic anemia[37]. Moreover, prostate, breast, colon, and bladder cancer patients often lack the dematin genes[38]. Recent findings have demonstrated that reintroduction of the dematin gene reverted cancerous prostate cells to their normal healthy state[1,7,8,38]. Because of its role in tumor suppression, characterizing dematin's intramolecular interface will enrich our current understanding of cancer and contribute to the advancement of biomedical research.

1.6 Thesis Aims:

This works aims to achieve the following goals in investigating the properties, function, and interactions of large IDR-containing cytoskeleton regulators:

I. Develop and optimize procedures to synthesize large IDR-containing proteins via Sortase-Mediated Ligation (SML).

II. Segmentally label dematin (*Homo sapiens*) via SML to characterize its intramolecular binding interface.

III. Optimize the *in vitro* synthesis of the full-length villin 4 IDR (*Arabidopsis thaliana*) through SML.

Chapter II: Materials and Methods:

2.1 Chemical Supplies:

All buffers and solutions were made from chemical reagents obtained from commercial vendors. The water used in these preparations was filtered with a Milli-Q Advantage A10 system (Millipore).

2.2 Plasmid Design:

All Aim I proteins, dematin, and villin 4 constructs discussed were provided in pET-24a(+) vectors by GenScript. Genes for wild-type and hepta-mutant sortase A (7m-srtA) were incorporated into pET-30b(+) plasmids and issued by the Hidde Ploegh Laboratory. Aim I hexapeptides were generated via solid-phase peptide synthesis and provided by Antos lab, Department of Chemistry.

2.2a Design of Aim I Constructs:

Starting from the N-terminal end, all Aim I proteins contain an FH8 solubility and purification tag, a TEV cleavage site, a poly-glycine/serine spacer, the authentic IDR sequence under investigation (five residues constituting the N-terminal flanking region relative to the sortase site), the LPXTG motif, an additional glycine to improve the SML reaction rate, and a histidine affinity tag for purification purposes (**Figure 2.2.1**)

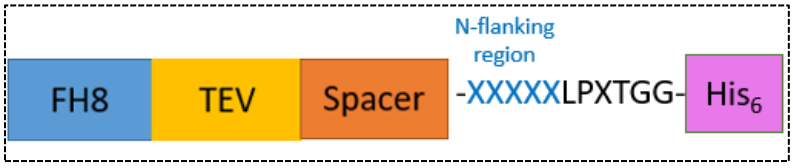


Figure 2.2.1. Structure of the SML N-terminal ligation partner.

All Aim I hexapeptides are composed of an N-terminal glycine nucleophile, followed by an authentic IDR sequence (five residues representing the C-terminal flanking region to the LPXTG site), and 2,4-dinitro-phenyllysine or K(DNP) as the chromophore (**Figure 2.2.2**).

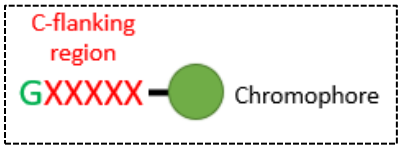


Figure 2.2.2. Structure of the SML C-terminal ligation partner.

All authentic IDR sequences were found in the DisProt database. To generate the LPXTG sortase site, two point-mutations were generated within these true IDR sequences. These mutations differ depending on the construct but are indicated appropriately in the corresponding figures throughout this document. The extra glycine and His_6 found in aim I proteins are absent from the final ligation product.

2.2b Design of Aim II Constructs:

Dematin IDR: From the N-terminal end, contains a twin strep tag, glycine spacer, TEV cleavage site, IDR sequence(1-300), the LPSTG sortase site, originally LQSTE in WT dematin, an additional glycine, and a His_6 affinity tag (**Figure 2.2.3**).

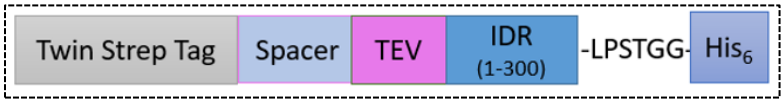


Figure 2.2.3. Structure of dematin IDR construct.

Dematin HP (P): Dematin headpiece (phosphorylation mimic) is composed of an N-terminal His_6 followed by a TEV cleavage site, and the headpiece sequence containing the S381E mutation to mimic phosphorylation by protein kinase A (PKA) (**figure 2.2.4**).

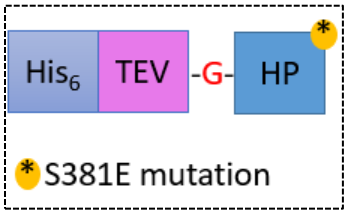


Figure 2.2.4. Structure of the mutant dematin headpiece construct. The S381E mutation mimics the phosphorylated state in terms of structure and function. This segment encompasses residues 306-383.

Dematin HP: Wild-type dematin headpiece (unphosphorylated). Same design as the dematin HP (P) but contains the native serine at position 381 (**Figure 2.2.5**).

Figure 2.2.5. Structure of the wild-type dematin headpiece construct. This segment encompasses residues 306-383.

Dematin C-IDR-HP: Starting from the N-terminus, a His_8 followed by a TEV cleavage site and the C-terminal end of dematin IDR attached to the wild-type headpiece sequence (residues 161-383) (**Figure 2.2.6**).

Figure 2.2.6. Structure of C-IDR-HP. This SML C-terminal ligation partner contains the C-terminal portion of the dematin IDR attached to the wild-type headpiece. This segment encompasses residues 161-381.

Dematin N-IDR: From the N-terminus, twin strep tag followed by a TEV cleavage site, then the N-terminal portion of dematin IDR (residues 1-160). This fragment contains two mutations, S156L and K160G (**Figure 2.2.7**).

Figure 2.2.7. Structure of N-IDR. This SML C-terminal ligation partner contains the N-terminal portion of the dematin IDR. This segment encompasses residues 1-160 (when counting amino acids in the SML site).

As in Aim I constructs, the last glycine and His_6 at the C-terminal end of the IDR fragment do not appear in the ligation product.

2.2c Design of Aim III Constructs:

AtV4IDR: Villin 4 full-length IDR from the N- to C-terminus includes a twin strep tag, a poly-glycine spacer, the IDR sequence, a diglycine spacer, a His_8 affinity tag. Residues of suspected PEST motif within positions 866-873 changed from SPAPESNS to AAAGGGAG (**Figure 2.2.8**).

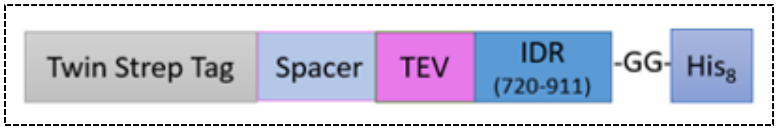


Figure 2.2.8. Structure of AtV4IDR.

AtV4IDR-B: The basic portion of the IDR flanked on its N-terminal side by a twin strep tag, followed by a poly-glycine spacer, and a TEV cleavage site. At the C-terminal side, the suspected PEST motif was mutated from APESNS to LPETGG followed by His_8. The mutations aimed at creating a sortase recognition site and minimizing proteolytic cleavage. The second glycine and His_6 are not part of the final ligation product (**Figure 2.2.9**).

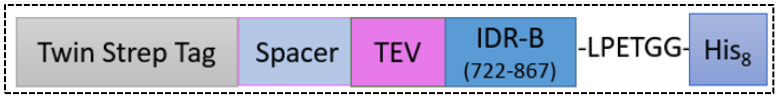


Figure 2.2.9. Structure of AtV4IDR-B.

AtV4IDR-A-HP: Contains the acidic portion of the V4 IDR attached to the folded headpiece domain, included to provide stability. The N-terminus of this protein fragment includes a twin strep tag, a poly-glycine spacer, a His_8 tag, and a GG sequence, initially NS in the suspected PEST motif, and a TEV cleavage site (**Figure 2.2.10**).

Figure 2.2.10. Structure of AtV4IDR-A-HP.

2.2d Enzymes Used in the Experimental Protocols:

See Appendix section 1 for enzymes used and corresponding amino acid sequences.

2.3 Bacterial Transformation:

All constructs referenced were transformed into BL21 (DE3) *E. coli* cells obtained from GOLDBIO. The competent cells (50 µL) were first thawed on ice for 15 minutes upon retrieval from the -80C freezer. Following the addition of ~1 pg of plasmid DNA, the cells were incubated on ice for 30 minutes before undergoing heat shock at 42 ℃ for 45 seconds. Subsequently, the cells were placed on ice for 2 minutes, mixed with 950 µL of Luria Broth (LB) media (10g/L Bacto tryptone, 10g/L NaCl, 5g/L yeast extract) then incubated in a shaker at 37 ℃ and 210 rpm for 1 hour to recover prior to being plated onto solid agar media (LB composition, 15 mg/mL

granulated agar) containing kanamycin (10 μg/mL) as the selective agent. These plates were then incubated at 37 ℃ overnight (12-17 hours).

2.4 Bacterial Growth and Protein Expression:

2.4a Regular Growth and Expression:

All constructs were expressed in BL21 (DE3) *E. coli* cells. These cells were grown in 1 L LB media (37 ℃, 210 rpm) and induced with 1 mL of 0.8 M IPTG when an optical density (O.D.) of 0.5-0.7 was reached. 4-6 hours following IPTG induction, the cell culture was centrifuged at 4 ℃, 4629 x g for 20 minutes, the LB was discarded and the pellet was stored at -80 ℃.

2.4b Bacterial Growth in Minimal Media (M9):

When isotopically labeled constructs are generated, they will be made in BL21 (DE3) *E. coli* strains cultured in minimal media according to our established procedure: cells are first added to LB media (37 ℃, 210 rpm). Upon reaching an O.D. of 0.4-0.5, they are centrifuged at 4000 x g for 20 minutes at 4℃ then resuspended in 500 mL of salt wash for each 6 L LB media growth (6 g/L Na_2HPO_4 , 3 g/L KH_2PO_4, 0.5 g/L NaCl, 200 g/L vitamin B1, 120 g/L $MgSO_4$, 28 g/L $CaCl_2$, 10 mg/L kanamycin). Upon resuspension, the cells are subject to another centrifugation* step at 4000 x g for 20 minutes at 4℃ before being suspended in minimal media M9 which contains the sources for carbon and nitrogen (identical salt wash composition, 3 g/L $^{15}NH_4Cl$, 6 g/L ^{13}C-D-Glucose). Following 1 hour of incubation at 37 ℃, they are induced with 2 mL of 0.8M IPTG per liter of culture and incubated for 4 hours at 210 rpm, 37℃. Finally, the culture is spun down for 15 minutes at 4629 x g and 4℃. The supernatant is discarded and the pellets are stored at either -80℃ or -20℃ awaiting lysis.

*All centrifugation steps are performed using a Sorvall Lynx 4000 centrifuge, Thermo Scientific, rotor: F12-6x500 LEX, PN 096-062375.

2.5 Lysate Preparation:

2.5a Aim I and III protein lysate:

10mL of lysis buffer* was added to pellets harvested from a 1L growth. Upon resuspension, 10 µL of stock lysozyme (50 mg/mL) was added and agitated at 15 rpm 4℃ for 30 minutes. This incubation was followed by three sonication rounds of 30 seconds each at 50% duty cycle using a Branson Sonifer instrument. DNase I was added to the lysate to a final concentration of 0.1 units/mL. After an incubation period of 30 minutes at room temperature, the lysate was subject to centrifugation for 30 minutes at 4℃ and 39,375 x g. The supernatant was passed through a 0.22µm filter in preparation for purification.

*Lysis buffer composition for Aim I proteins: pH 7.5, 150mM NaCl, 50mM Tris. Lysis buffer used with Villin 4 fragments: pH 8.0, 300mM NaCl, 50mM NH_2PO_4, 10mM Imidazole.

2.5b Aim II Protein Lysate:

10 mL of lysis buffer* was added to pellets obtained from a 1L growth. The lysate was sonicated for four 10 seconds intervals at 50% duty cycle with 10 seconds of cooling between each sonication round. The lysate was left overnight stirring at 4 ℃ then spun down at 4 ℃ and 39,375 x g or 1 hour. Like all lysate, it was filtered with 0.22µm filter prior to chromatography.

*When IMAC was performed immediately following the lysate preparation IMAC solubilizing buffer was used to lyse the bacterial cells (6M Urea, 50mM NaH2PO4, 300mM NaCl, pH 8).

When Strep Tactin XT affinity purification was performed first, Buffer W (6M urea, 300 mM NaCl, 50mM NaH2PO4, 20mM Imidazole, pH 8) served as the lysis buffer instead.

2.6 Protein Purification:

2.6a Aim I Proteins:

All Aim I proteins were purified using immobilized metal affinity chromatography (IMAC). All IMAC purification was done with QIAGEN nickel nitrilotriacetic acid (Ni-NTA) Superflow resin. 1 column volume (CV) or 3mL* of Ni-NTA resin was left agitating gently at 15 rpm with the filtered lysate for 1 hour at 4℃. Once the flow-through was collected, the resin was washed with 8 CV of IMAC wash buffer (pH 7.5, 150 mM NaCl, 50 mM Tris-base, 20 mM imidazole). The protein of interest was then eluted with 10 CV of IMAC elution buffer (pH 7.5, 150 mM NaCl, 50mM Tris-base, 300 mM imidazole)

*The column volume mentioned applies to a lysate obtained from a 1L expression growth.

2.6b. Aim II Purification:

All dematin constructs containing a strep and a $His_{6/8}$ affinity tag are purified via IMAC immediately followed by Strep affinity chromatography (SAC). IMAC was performed following the procedure described in section 2.6a. For dematin constructs however, 15 CV of IMAC wash buffer (pH 8, 2 M urea, 300 mM NaCl, 50mM NaH_2PO_4, 20 mM imidazole) and 9 CV of IMAC elution buffer (pH 8, 2 M urea, 50 mM NaH_2PO_4, 250 mM imidazole, 300 mM NaCl) were used. SAC purification was performed using IBA Strep Tactin XT 4flow resin. 1 CV or 2 mL* of strep tactin XT was mixed with the IMAC elutions. After collecting the flow through, the resin was washed with 15 CV of buffer W (pH 8, 2 M urea, 100mM Tris-HCl, 150mM NaCl, 1mM EDTA)

and the protein was eluted in 10 CV of Buffer BXT (pH 8, 2 M urea, 100mM Tris, 150mM NaCl, 1mM EDTA, 100 mM biotin). The resin bed should not be disturbed or agitated in anyway. No incubation period with the lysate is required for binding to occur.

Dematin headpiece fragments are purified via two rounds of IMAC using the above-mentioned volumes for wash and elution fractions. Thus, IMAC was done before and after the TEV-cleavage reaction to which they are subject prior to their ligation with the rest of the dematin protein (the IDR). Moreover, the relative stability of both headpieces eliminates the need of urea in the buffers used for their purification, storage, and analysis.

*The column volume mentioned applied to a lysate obtained from a 1L expression growth.

2.6c Aim III Purification:

All villin 4 fragments follow an identical purification regimen as dematin fragments except that no urea is used in the IMAC and strep buffers.

2.7 TEV-Cleavage Reactions:

In order to expose the N-terminal glycine nucleophile required for sortase-mediated ligation, all dematin and villin 4 headpiece (HP) constructs underwent a TEV cleavage. The total reaction volume ranged from 1-5 mL. The final reaction mixture contained 1 mM DTT, 0.5 mM EDTA, 1 µM of TEV protease for every 90 µM of headpiece. The reaction was incubated at room temperature for 16 hours in PIPES buffer (pH 6.8, 50 mM NaCl, 20 mM PIPES) then purified via IMAC using 250 µL of Ni-NTA resin. The flow-through was collected and the resin was washed with 8 CV of PIPES buffer followed 4 CV of IMAC elution buffer (pH 8, 50 mM NaH_2PO_4, 250 mM imidazole, 300 mM NaCl). The flow through and the wash were typically combined as they contained the pure cleaved Headpiece constructs which were ready for ligation.

2.8 Sortase-Mediated Ligation (SML):

2.8a Aim I SML:

All ligation reactions utilized 20 µM of wild-type sortase A (Staphylococcus *aureus*) and were monitored for 8 hours at room temperature and pH 7.5, in 150 mM NaCl and 50 mM Tris-base buffer spiked with 10 mM Ca^{2+} and 1 mM TCEP. The protein substrate and hexapeptide nucleophile were present at 50 µM and 250 µM, respectively. The total reaction volume was 100 µL.

2.8b Aim II SML:

All ligation reactions between dematin's IDR and TEV-cleaved headpiece fragments utilized 5 µM of hepta-mutant sortase A (7m-srtA). They were incubated for 4-48 hours at room temperature and pH 8 in 20 mM imidazole, 50 mM NaH_2PO_4, 300 mM NaCl and 0-2 M urea buffer. The reaction volume varied between 0.5-3 mL. The molarity of the TEV- cleaved headpiece construct exceeded that of the IDR by a factor of 5 as the headpiece levels were kept around 120 µM while the IDR was measured to be about 24 µM. Throughout these optimization experiments, the only data collected was for the ligation of the TEV-cleaved phosphorylated headpiece to the IDR.

2.8c Aim III SML:

The ligation between AtV4IDR-B and AtV4IDR-A-HP was incubated for 24 hours at room temperature and pH 8 in 100 mM Tris-base, 150 mM NaCl, 1 mM EDTA, and 100 mM biotin. 6 µM of 7m-srtA was used with 40 µM AtV4IDR-A-HP and 30 µM TEV- cleaved AtV4IDR-B Purification of the ligation product from other reaction components was attempted using IMAC and SAC. The only difference resides in the volume of resin used. To isolate the ligation product,

200 µL of Ni-NTA and strep tactin XT were typically used for reaction mixtures ranging from 1-5 mL in volume. Size-exclusion chromatography (SEC) was also utilized as either an alternative or a complement to IMAC and SAC.

2.9 Size-Exclusion Chromatography (SEC):

Occasionally, the proteins expressed were further purified through SEC in which a HiPrep™ 16/60 Sephacryl™ S-100 HR column (120 mL CV, GE Healthcare) was attached to an ÄKTAprime plus FPLC system (GE Healthcare). Proteins were eluted in PIPES buffer (pH 6.8, 20 mM PIPES, 50 mM NaCl) which contained 2 M urea for dematin constructs at a flow rate of 0.5mL/min.

2.10 Buffer-Exchanging Protein Samples:

Aim I proteins were buffer-exchanged into 150 mM NaCl, 50mM Tris-base at pH 7.5 preceding ligation with their corresponding hexapeptides using disposable Econo-Pac 10DG desalting columns from BIO-RAD. The column was equilibrated with 20 mL of 150 mM NaCl, 50mM tris-base buffer at pH 7.5. A 3 mL-sample was then added to the column. After the flow through was collected, the protein was eluted in 4-12 mL of 150 mM NaCl, 50mM Tris-base buffer at pH 7.5. All other buffer-exchange procedures involving dematin or villin 4 fragments were performed via dialysis.

2.11 Protein Quantity and Purity Assessment:

Protein quantification was achieved through measuring absorbances at 280 nm with a Nanodrop™ ND-1000 spectrophotometer (ThermoFisher).

Their extinction coefficients were obtained in Expasy ProtParam and their molarity calculated using Beer-Lambert law. Protein concentration was achieved through Thermo Scientific Pierce spin concentrators with 3000 Da MWCO (all samples) spun in the Thermo scientific Sorvall

Legend X1R centrifuge (rotor# 75003181). In addition to using spin concentrators, dematin (P) LP was concentrated via lyophilization with Labconco FreeZone Lyophilizer 2.5 L. Protein purity was determined via 15% SDS-PAGE and their identity was confirmed through Mass Spectrometry (MS). In this document, concentrations given for the protein fragments discussed are typically for the most concentrated fraction obtained from a 1-3L bacterial growth.

2.12 Identity Verification: Mass Spectrometry (MS)

2.12a Liquid Chromatography Electrospray Ionization Mass Spectrometry (LC-ESI-MS)[32,39]: Identity of protein samples was confirmed through liquid chromatography electrospray ionization mass spectrometry (LC-ESI-MS). Such experiments were done using Advion CMS expression^{L} mass spectrometer and a Dionex Ultimate 3000 HPLC instrument to which a Phenomenex Aeris 3.6 µm WIDEPORE C4 200 Å column (100 x 2.1 mm) was attached. The mobile phase consisted of 95% H_2O, 5% MeCN, and 0.1% formic acid and organic MeCN and flowed at a rate of 0.3 mL/min. For villin 4 and dematin headpiece fragments (settings: keep at 10% organic for 1.0 min, linear gradient of 10–90% organic from 1.0–7.0 min, keep at 90% organic 7.0–9.0 min, linear gradient of 90–10% organic 9.0–9.1 min, re-equilibrate at 10% organic 9.1–12.0 min). The protein charge ladders were deconvoluted with Advion Data Express or Analyst 1.4.2 software. This same method is also described in Erin Rosenkranz' thesis[39] (OCLC# 1277514139) as well as our latest published manuscript (PMID# 34710113)[32]

2.12b Liquid Chromatography Quadrupole Time-of-Flight Mass Spectrometry (LC-QTOF-MS)[39]: For Dematin's IDR segment as well as the final ligation product, identity was checked using liquid chromatography quadrupole time-of-flight mass spectrometry (LC-QTOF-MS). This analysis was carried through with Agilent AdvanceBio 6545XT LC/QTOF mass spectrometer and

an Agilent 1290 Infinity II UHPLC to which a Phenomenex AerisTM 3.6 µm WIDEPORE C4 200 Å column (100 x 2.1 mm) was attached. The mobile phase consisted of 95% H_2O, 5% MeCN, and 0.1% formic acid and organic MeCN at a rate of 0.3 mL/min (settings: keep at 10% organic for 1.0 min, linear gradient of 10–90% organic from 1.0–7.0 min, keep at 90% organic 7.0–9.0 min, linear gradient of 90–10% organic 9.0–9.1 min, re-equilibrate at 10% organic 9.1–12.0 min). The samples' charge ladders were deconvoluted with Agilent BioConfirm software. This same method is also described in Erin Rosenkranz' thesis[39] (OCLC# 1277514139)

2.13 Solution Nuclear Magnetic Resonance (NMR) Spectroscopy:

When protein samples are ready for NMR analysis, they will be buffer-exchanged into PIPES buffer (pH 7.5, 20 mM PIPES, 50-200 mM NaCl) and spiked with EDTA, D_2O, and NaN_3 to a final concentration of 0.5mM, 8%, and 0.02%, respectively. ^{15}N-HSQC spectra will be acquired through Bruker Avance III HD NMR spectrometer operating at a magnetic field of 11.7 Tesla.

Chapter III: Results and Discussion:

3.1 Aim I Results and Discussion:

Aim I: *Develop and Optimize Procedures to Synthesize Large IDR-Containing Proteins via Sortase-Mediated Ligation (SML).*

The IDR sequences to be reconstructed via SML were found in the DisProt database which contains over 625 IDRs greater than 100 residues (table 3.1.1) . Using SLIDRs (sortase ligation for intrinsically disordered regions), a software co-developed by Dr. Sergey Smirnov and Dr. John Antos from the Chemistry department at Western Washington University, potential SML sites were identified within the IDR sequences analyzed. Within the 625 large IDRs found in DisProt, only 4 native LPXTG sites were found. Upon specifying in SLIDRs the number of point-

mutations required to generate an LPXTG motif, 301 additional sites became available through the incorporation of 1 point-mutation, this number further increased to 8147 if up to 2 mutations were to be introduced into these IDR sequences.

Table 3.1.1. SML Sites Identified in DisProt IDRs using novel SLIDRs software.

# Large IDRs (>100 residues)	625
# Native LPXTG sites in all IDRs	4
# LPXTG sites accessible through 0-1 mutations	305
# LPXTG sites accessible through 0-2 mutations	8147
# IDR residues per LPXTG site	~50

In the case of the IDR constructs chosen for this study, we limited the number of amino acid substitutions to 2 in order to minimize the impact these changes may have on structure, function, and behavior of the IDRs under investigation. Thus, 12 ligation systems were initially designed to be compared in terms of their SML conversion rate from substrate to product under the same reaction conditions. Specifically, our efforts aim at assessing how the identity of residues (acidic, basic, aromatic aliphatic, etc.) surrounding the LPXTG site affects SML efficiency as we attempt to reconstitute IDR sequences (**Figure 3.1.2**). Each system is composed of a protein as the N-terminal ligation partner and a hexapeptide as a C-terminal substrate. All proteins include an LPXTG motif preceded by a five-residue IDR sequence representing the N-terminal flanking region to sortase site (**Figure 3.1.1**). All hexapeptides synthesized include an N-terminal glycine nucleophile followed by a five-residue IDR motif constituting the C-terminal flanking region to the sortase recognition signal. In addition to these ligation systems, a positive glycine control was designed to serve as our reference. In this scenario, the N- and C-terminal flanking regions are composed of a pentaglycine motif. Indeed, sortase enzymes typically

demonstrate a high degree of efficiency when glycine residues surround the LPXTG sequence[32,40]. Due to the collaborative nature of this work, only some systems which reflect the author's direct contribution to this aim will be discussed in this document.

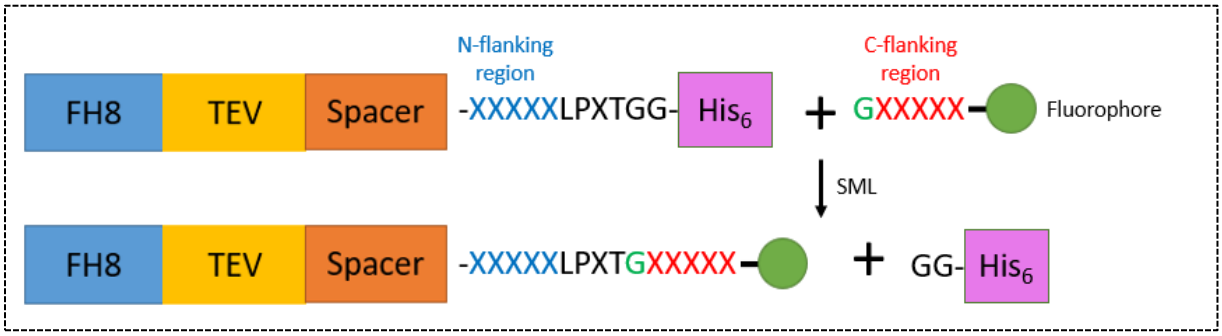


Figure 3.1.1. Evaluation of SML reaction efficiency under various N- or C-terminal contexts. Authentic IDR sequences are reconstructed *in vitro* through SML. 0-2 point mutations are performed in these true IDR sequences to create the LPXTG sortase recognition sites. The N-terminal IDR portion is expressed in *E .coli* BL21 (DE3) and includes the N-terminal flanking residues. TEV-cleavable FH8 and His_6 tags are incorporated to allow for purification and improve solubility (FH8). The hexapeptide constitutes the C-terminal ligation partner and contains an N-terminal glycine nucleophile, the C-terminal flanking region, and a chromophore to aid in tracking and yield determination. The flanking regions assessed for their impact on SML reaction rate were chosen for their acidic, basic, aromatic, aliphatic, hydrophobic, polar, or proline-rich properties. All ligation reactions in aim I were performed with Wild-type sortase A (staphylococcus *aureus*)

3.1.1 Proline-rich C-terminus (ProC):

The proline-rich C-terminus (ProC) system evaluates the impact a pentaproline stretch downstream of the LPXTG motif makes on SML rate. Here, valine and alanine found in the wild-type IDR sequence were mutated to proline and glycine, respectively to create the sortase A recognition signal (**Figure 3.1.2**). On average 35% of the protein substrate was ligated to the corresponding hexapeptide, around 2.5-fold lower efficiency relative to the glycine control which was found to produce 80% ligation product (**Figure 3.1.3**).The control was designed to contain poly-glycine stretches based on the known high performance of sortase enzymes in the

presence of glycine residues surrounding the LPXTG sorting motif[40]. In the ProC system, 23 % of the protein substrate was lost to competing reactions. As suggested by LC-ESI-MS, hydrolysis and potential intramolecular events are suspected to occur.

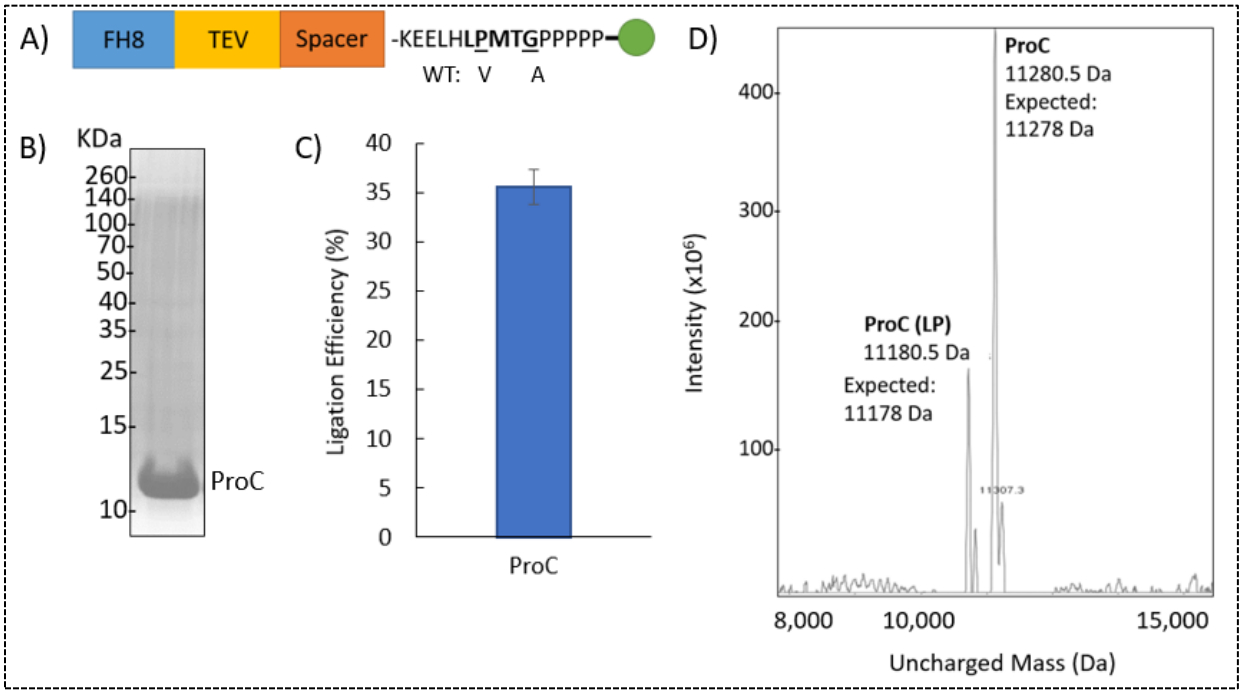


Figure 3.1.2. Sortase-mediated ligation in the presence of a C-terminal pentaproline motif. A) Structure of the ligation product, ProC (LP). Mutated residues are underlined, wild type amino acids (aa) are denoted directly underneath the substitution site. B) 15 % SDS-PAGE of N-terminal ligation partner ProC following IMAC purification. C) Percent ProC successfully ligated to the corresponding hexapeptide to form ProC (LP). D) LC-ESI-MS spectrum of ProC and ProC (LP).

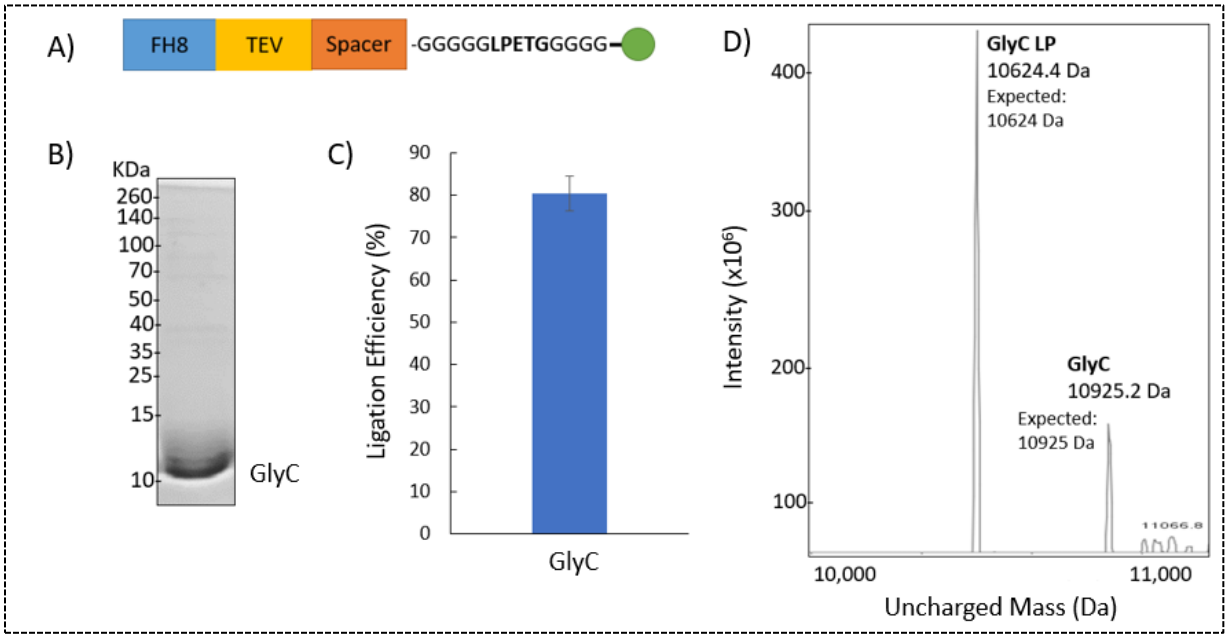


Figure 3.1.3. Sortase-mediated ligation of glycine control. A) Structure of ligation product GlyC (LP). B) 15% SDS-PAGE of glycine N-terminal ligation partner. C) Percent glycine protein ligated to C-terminal pentaglycine substrate. D) LC-ESI-MS spectrum of GlyC and GlyC (LP).

3.1.2 Aromatic C-terminus (AroC) and Aromatic N-terminus (AroN):

In Aromatic C (AroC), the native lysine and cysteine were changed into leucine and threonine to generate the LPXTG sequence which is followed by aromatic amino acids (YYHWD) in the ligation product (**Figure 3.1.4**). A conversion rate of 5 % was determined after 8 hours of monitoring. Given identical reaction conditions, the 17-fold decrease in ligation efficiency compared to the glycine control (**Figure 3.1.3**) can be attributed to the amino acid composition around the LPXTG site. Whether the N- or C-terminus context is responsible for the massive drop in substrate conversion rate is still unclear. An unexpected peak around 11146 Da was detected and may be the product of a side reaction that has yet to be identified.

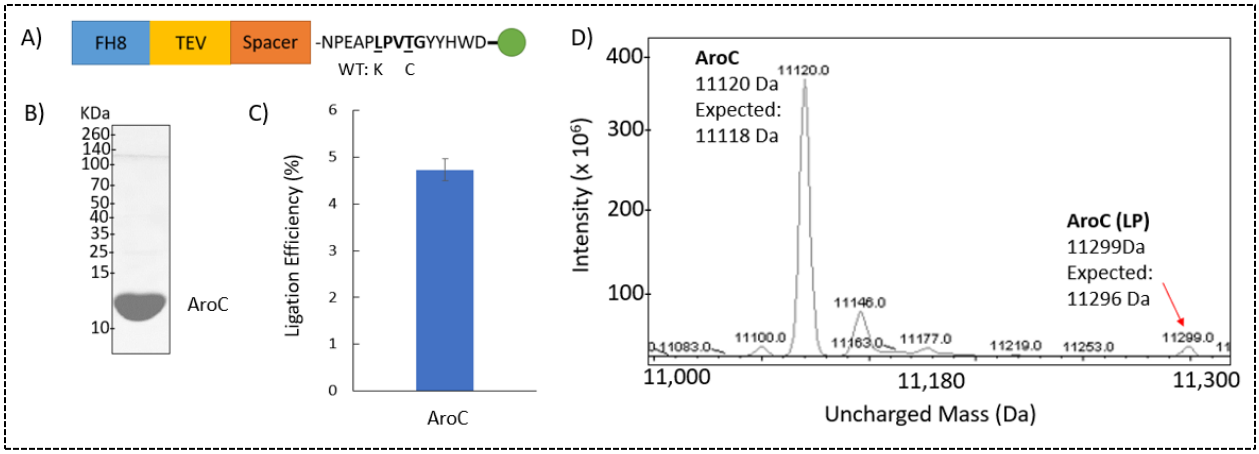


Figure 3.1.4. Sortase-mediated ligation in the presence of aromatic residues downstream of the LPXTG site. A) structure of the ligation product, AroC (LP). Mutated residues are underlined, wild type amino acids (aa) are denoted directly underneath the substitution site. B) 15% SDS-PAGE of N-terminal ligation partner AroC following IMAC purification. C) Percent AroC converted to ligation product AroC (LP). D) LC-ESI-MS spectrum of AroC and AroC (LP).

In contrast to AroC, aromatic N-terminus (AroN) contained an aromatic motif (WHIWW) upstream of the LPXTG site, originally LAIVG in the wild-type IDR sequence (**Figure 3.1.5**). No ligation product was detected for this system. No undesired products were made. Similar to AroC, our current data is insufficient to determine with certainty which of the N- or C-terminal motif is impeding the SML reaction.

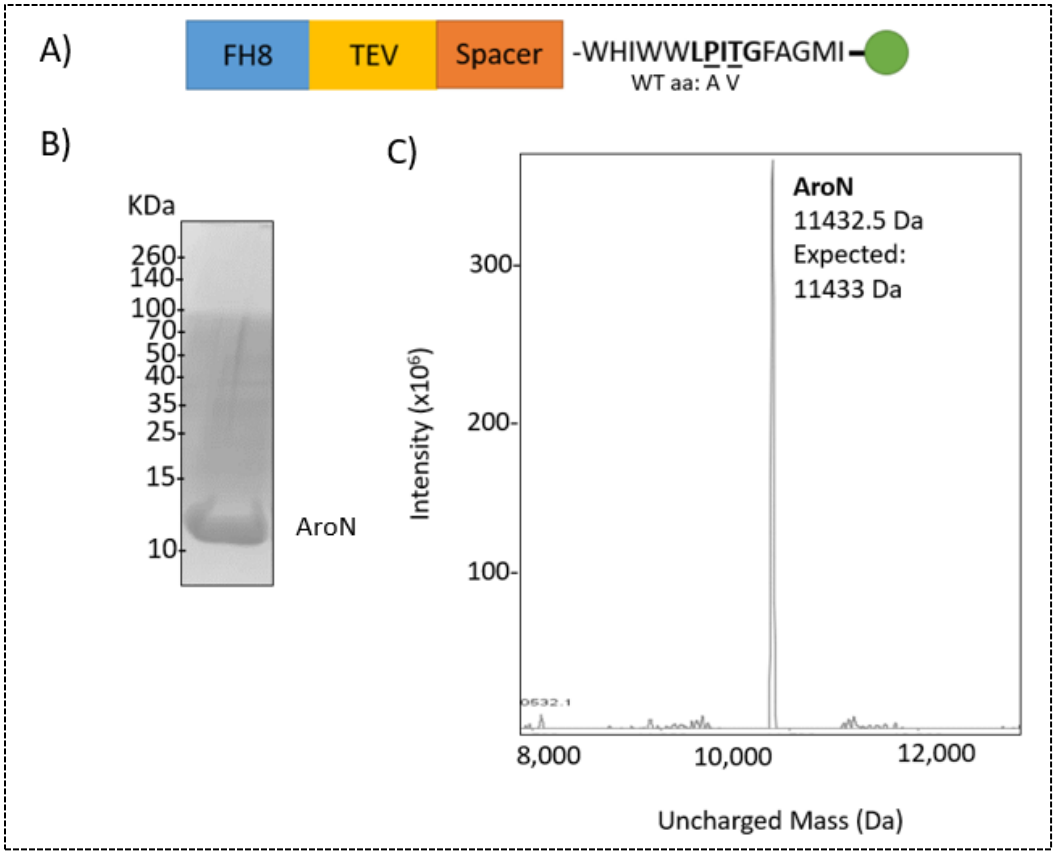


Figure 3.1.5. Sortase-mediated ligation in the presence of an aromatic motif upstream of the LPXTG site. A) Structure of potential ligation product AroN (LP). Mutated residues are underlined, wild type amino acids (aa) are denoted directly underneath the substitution site. B) 15 % SDS-PAGE of N-terminal ligation partner AroN following IMAC purification. C) LC-ESI-MS spectrum of sample reaction mixture, no AroN (LP) detected.

3.1.3 Polar-Charged N-terminus (PolN):

Polar charged N-terminus (PolN) incorporated polar charged residues (DRIKE) immediately before the sortase site in which a threonine and glycine have replaced the native leucine and tryptophan, respectively (**Figure 3.1.6**). While this system performed better than AroC and AroN (section 3.1.3), it remained suboptimal compared to the glycine control (**Figure 3.1.3**). Indeed, in the polyglycine reference reaction, the extent of ligation using WT-srtA was found to

be be 3.5-fold higher relative to the PolN context, which exhibited an overall conversion of only 22%. Competing reactions contributed to the loss of ~13 % of the initial PolN protein substrate. Like in ProC, these undesired side reactions are suspected to be hydrolysis of the protein substrate and some uncharacterized intramolecular event.

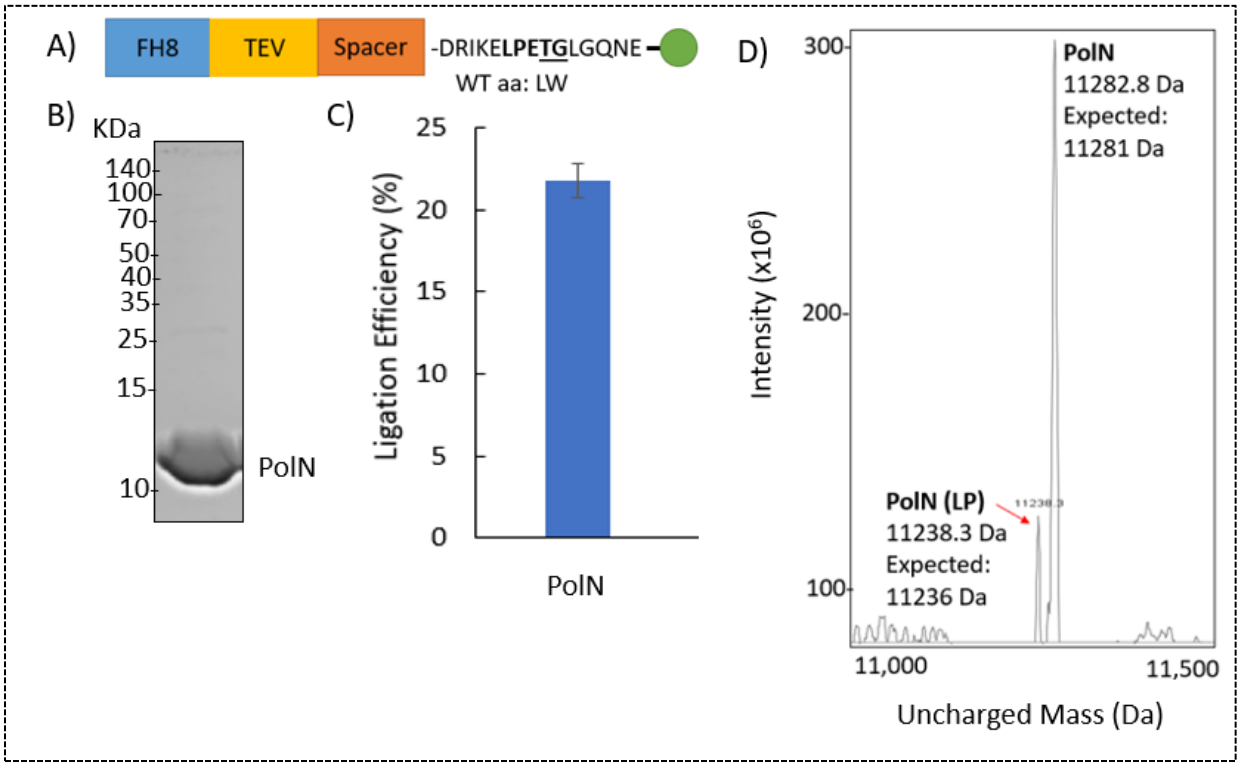


Figure 3.1.6. Sortase-mediated ligation in the presence of N-terminal polar charged residues before the LPXTG site. A) Structure of ligated product PolN (LP). Mutated residues are underlined, wild type amino acids (aa) are denoted directly underneath the substitution site. B) B) 15% SDS-PAGE of N-terminal ligation partner PolN following IMAC purification. C) Percent PolN converted to ligation product PolN (LP). D) LC-ESI-MS spectrum of PolN and PolN (LP).

Interestingly, upon ligating the N-terminal protein partner with a pentaglycine control as the C-terminal substrate at 37 °C, the reaction conversion reached 92% after 8 hours (**Figure 3.1.7**), exceeding that of the glycine control performed at room temperature (**Figure 3.1.3**). The near 4-fold increase in ligation efficiency cannot be attributed to the temperature change alone,

because when the initial experimental constructs (**Figure 1.3.6, panel A**) were ligated at 37 ℃, there was no significant change in the success of the reaction. Thus, the substitution of the C-terminal flanking region LGQNE into a pentaglycine stretch likely caused the increase in ligation.

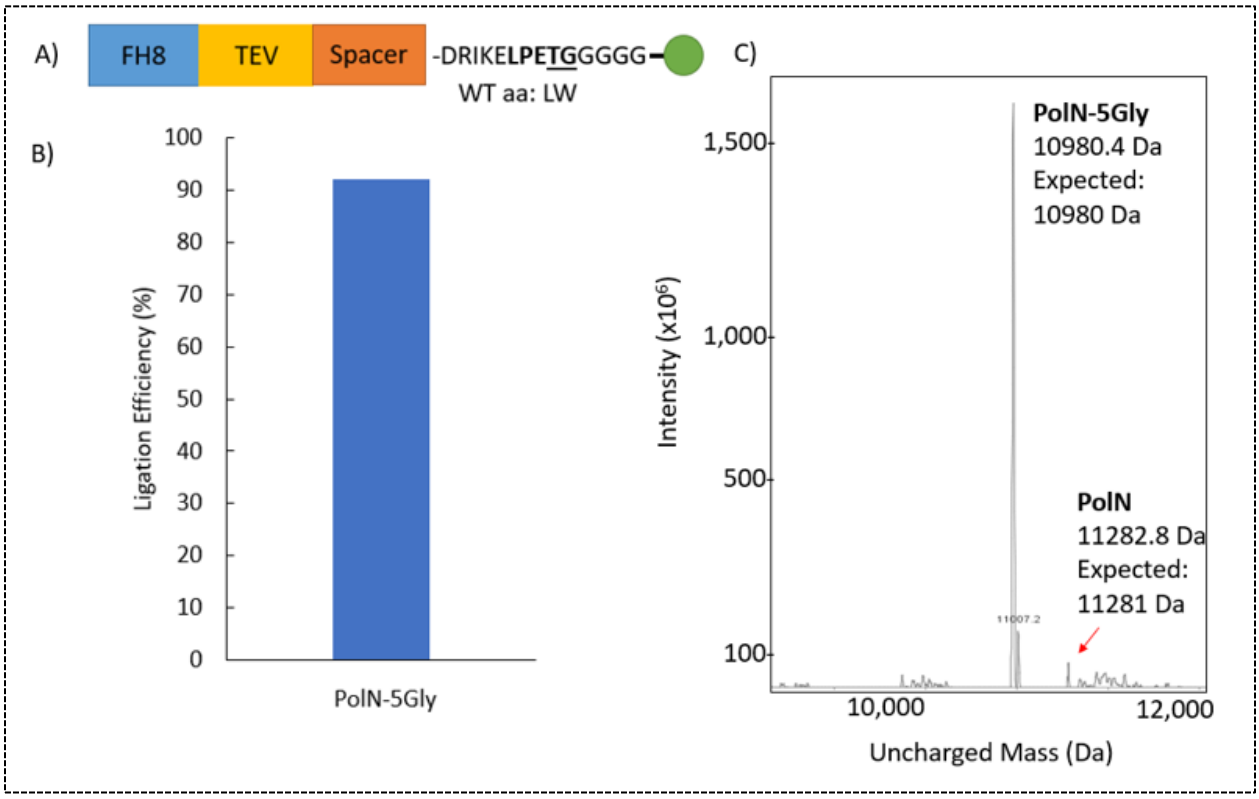


Figure 3.1.7. Sortase-mediated ligation between N-terminal polar residues and C-terminal pentaglycine. A) Structure of ligation product PolN-5Gly. Mutated residues are underlined, wild-type amino acids (aa) are denoted directly underneath the substitution site. B) Percent conversion of PolN into polN-5Gly. C) LC-ESI-MS spectrum of PolN and PolN-5Gly.

To further support this conclusion, a protein containing a pentaglycine as the N-terminal flanking region instead of the native DRIKE IDR sequence was ligated to the GLGQNE hexapeptide. With these reactants, only 14% of the protein substrate was converted to the product (5GlyN-PolN) via SML (**Figure 3.1.8**), confirming that the presence of the C-terminal

sequence LGQNE is the culprit for poor sortase performance. To note, the reactions depicted in **Figures 3.1.7** and **3.1.8** have been performed only once, unlike all other ligations discussed which were done in triplicates.

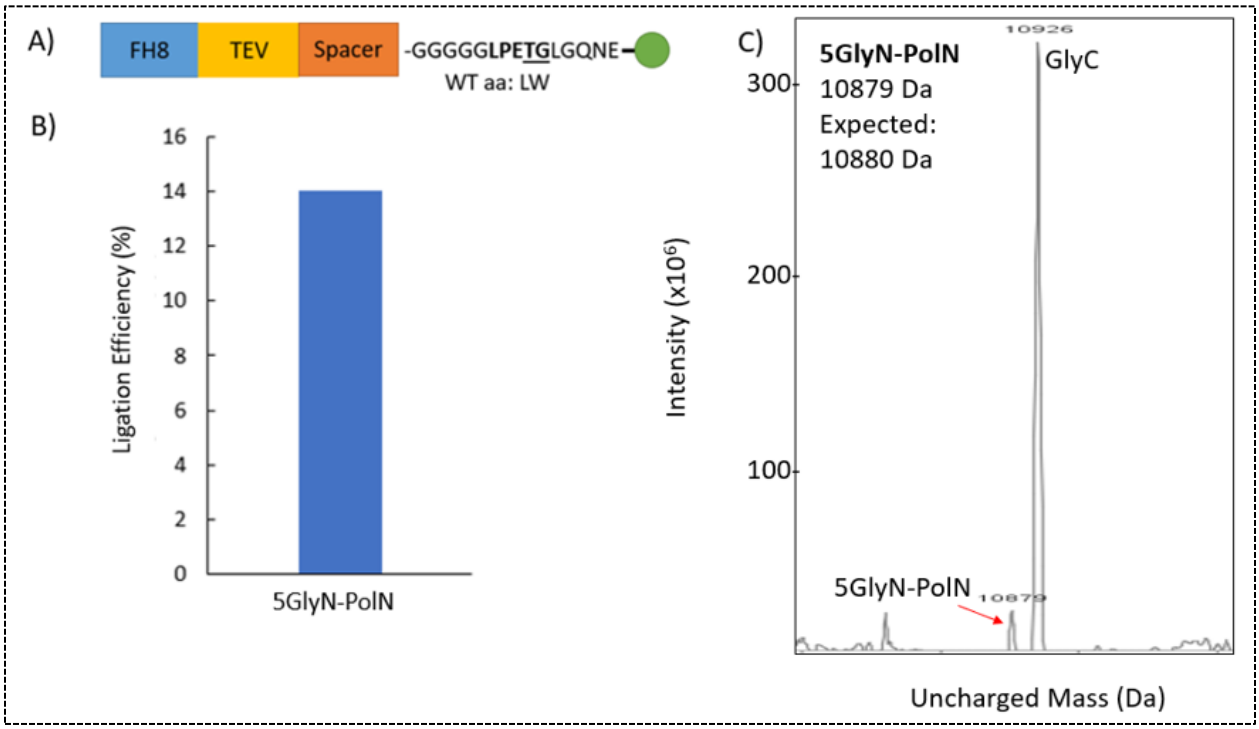


Figure 3.1.8. Sortase-mediated ligation between N-terminal pentaglycine and C-terminal PolN ligation partners. A) Structure of ligation product 5GlyN-PolN. Mutated residues are underlined, wild-type amino acids (aa) are denoted directly underneath the substitution site. B) Percentage of N-terminal glycine control that ligated with the C-terminal partner normally associated with PolN. C) LC-ESI-MS spectrum of ligation product 5GlyN-PolN.

3.1.5 Future Works in Aim I:

WT-srtA appears to operate at significantly suboptimal levels in all systems currently under investigation compared to the positive glycine control. Because all reactions, including the controls, were carried out under identical conditions, any deviation from the mean reference conversion of 80% can only be explained by the differences in sequence surrounding the

sortase ligation site. In order to determine which amino acid(s) in the N- or the C-terminal flanking region is(are) responsible for the decrease in ligation efficiency, each SML partner (protein and hexapeptide) will eventually be ligated to a glycine control acting as either the N-terminal or the C terminal ligation partner. Additionally, the use of increasingly larger glycine spacers between the ligation site and the problematic flanking regions may be warranted to assess the proximity necessary for these residues to impede the ligation reaction. Other parameters such as buffer composition, pH, and temperature may be altered as well in an effort to improve the product yield. The use of different sortase variants which may offer higher substrate affinity or the ability to recognize different ligation sites[19,20,22] are currently under consideration but issues related to competing hydrolysis are expected[41].

3.2 Aim II Results and Discussion:

Aim II: *Segmentally Label Dematin (Homo sapiens) via SML to characterize its intramolecular binding interface.*

Through analysis of the dematin sequence via SLIDRS, multiple potential sortase ligation sites were identified when up to two amino acid substitutions are allowed (**Figure 3.2.1**). Examples of these possible sortase motifs include positions 78-82, 156-160, and 301-305. The constructs discussed in this work will focus on the LPXTG site at position 301-305. To this end, the native sequence underwent two point mutations, Q302P and E305G. Because of the location of this ligation site, dematin's IDR ends up representing the N-terminal ligation partner while the folded headpiece constitutes the C-terminal SML substrate thus providing the required N-terminal glycine. Two different headpiece constructs were generated, the wild-type and the

phosphorylation mimic. The phosphorylation mimic contains an additional mutation S381E, emulating the protein's behavior, structure, and function when it is phosphorylated[41]. While ligation reactions will eventually also involve the wild-type headpiece and the IDR, the phosphorylation mimic headpiece was used throughout these optimization experiments because of the stability of the final ligation product.

A)

```
  1 MERLQKQPLT SPGSVSPSRD SSVPGSPSSI VAKMDNQVLG YKDLAAIPKD KAILDIERPD
 61 LMIYEPHFTY SLLEHVELPR QRERSLSPKA AAGGGAGKVW ADSRSPGIIS QASAPRTTGT
121 PRTSLPHFHH PETSRPDSNI YKKPPIYKQR ESVGGSPQTK HLIEDLIIES SKFPAAQPPD
181 PNQPAKIETD YWPCPPSLAV VETEWRKRKA SRRGAEEEEE EEDDDSGEEM KALRERQREE
241 LSKVTSNLGK MILKEEMEKS LPIRRKTRSL PDRTPFHTSL HQGTSKSSSL PAYGRTTLSR
301 LQSTEFSPSG SETGSPGLQI YPYEMLVVTN KGRTKLPPGV DRMRLERHLS AEDFSRVFAM
361 SPEEFGKLAL WKRNELKKKA S*LF
```

B)

SML site	Wild-type	Mutations
78-82	LPRQR	LPR**TG**
156-160	SPQTK	**L**PQT**G**
301-305	LQSTE	L**P**ST**G**

*Phosphorylation site.

Figure 3.2.1. Synthesis of a segmentally labeled Dematin. A) sequence of dematin isoform 2 (*Homo sapiens*). Highlighted motifs represent potential SML sites upon introduction of 2 point mutations. Underlined residues are part of the folded headpiece domain, the remaining residues constitute the IDR. B) Table listing positions of possible SML sites, showing both wild-type and corresponding mutant sequences.

Both the wild-type (**Figure 3.2.2**) and mutant headpiece (**Figure 3.2.3**) were successfully expressed, purified, and TEV-cleaved to expose their N-terminal glycine nucleophiles required for SML. 15-25mL of ~ 0.45 mM could generally be produced from a 3L-bacterial growths. The IDR was produced separate from the headpiece and stored in 2M urea (**Figure 3.2.4**). Typically 2-4 mL of dematin's IDR at 40-50 µM could easily be obtained from a 1L- bacterial growth.

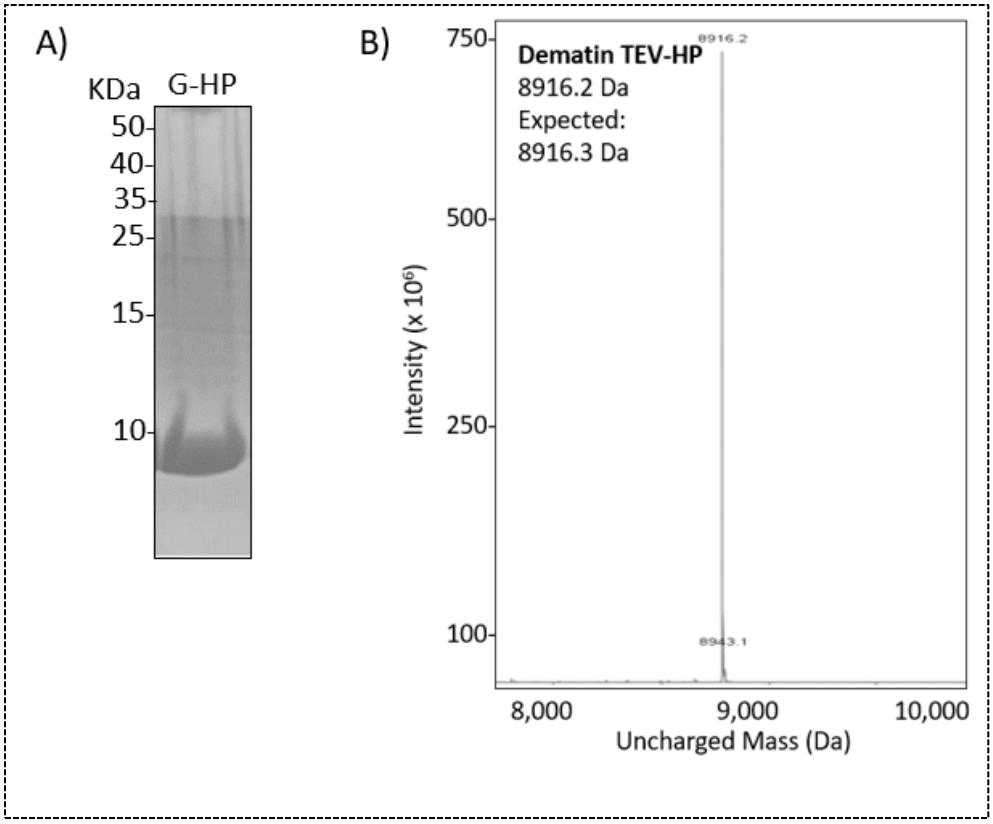


Figure 3.2.2. Dematin's wild-type TEV-cleaved headpiece domain (TEV-HP). A) 15% SDS-PAGE of TEV-HP. B) LC-ESI-MS spectrum of dematin TEV-HP

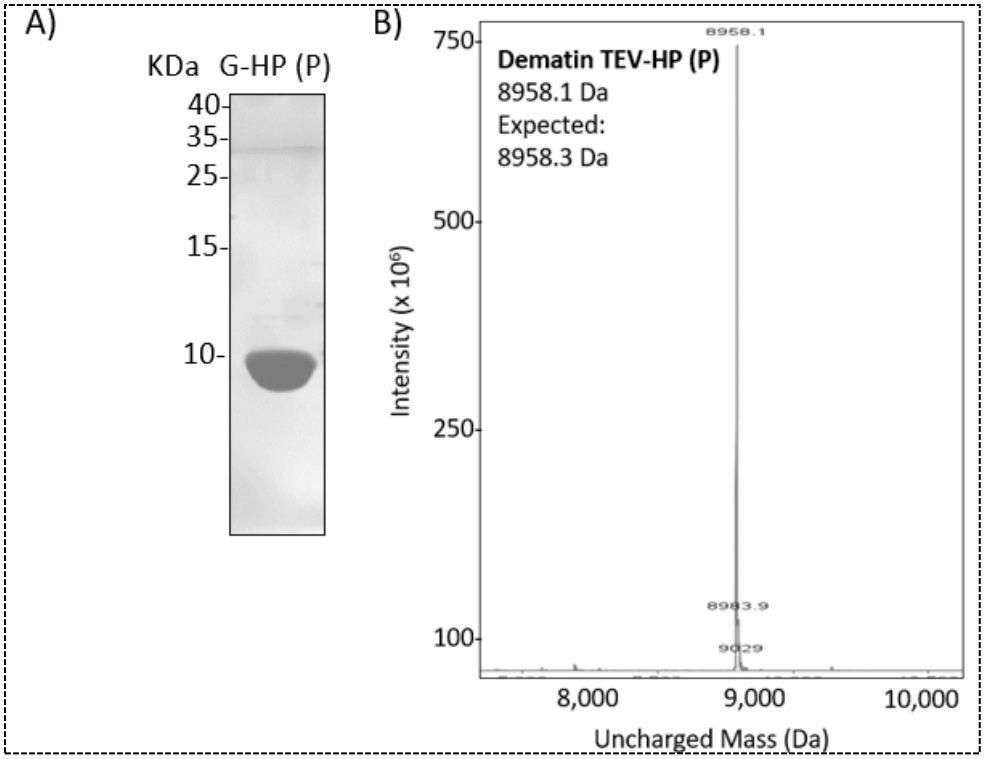


Figure 3.2.3. Dematin's mutant TEV-cleaved headpiece domain (TEV-HP (P)). This headpiece contains S381E mutation to mimic phosphorylation (P). A) 15% SDS- PAGE of TEV-HP (P). B) LC-ESI-MS spectrum of dematin TEV-HP (P).

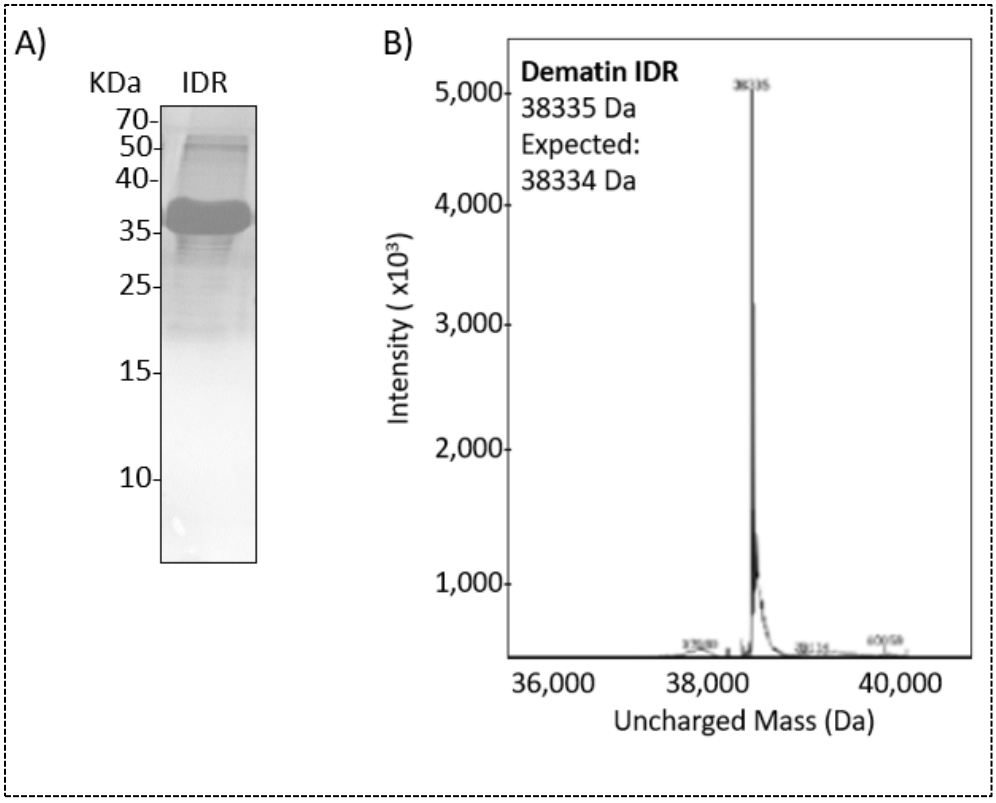


Figure 3.2.4. Dematin's IDR construct. A) 15% SDS-PAGE of dematin's IDR after strep and IMAC tandem purifications. B) LC-ESI-MS spectrum of dematin's IDR.

All SML reactions were performed using hepta-mutant sortase A (7m-srtA) because of its enhanced activity relative to WT-srtA. Urea up to 2 M in concentration was often used as a denaturing agent to minimize proteolytic degradation of the IDR-containing SML substrate and product without having any major effect on ligation efficiency. (See **Appendix Figure 2.3** for SML control reaction).

Figure 3.2.5. Synthesis and purification of full-length dematin (P) LP construct (phosphorylation mimic). A) Structure of dematin (P) (LP). B) SML reaction between the mutant dematin headpiece (S381E) and IDR was incubated for 24 hours. The reaction mixture was then purified via IMAC and SAC. The pure LP is indicated with a red arrow.

While the reaction was successful (**Figure 3.2.5**), difficulties were encountered when attempting to purify the ligation product. To achieve this end, a two-step process was designed: IMAC followed by Strep-Tag purification. While this was a promising purification method, a significant amount of ligation product remained bound to both the strep and Ni-NTA resin, leading to significant losses in yield (**Figure 3.2.6**). Furthermore, the majority of the ligation product collected could not be separated from the excess unreacted headpiece constructs and residual sortase enzyme. These issues were exacerbated when the purification steps were done

in the absence of urea, for reasons that may be related to the poor solubility of full-length dematin.

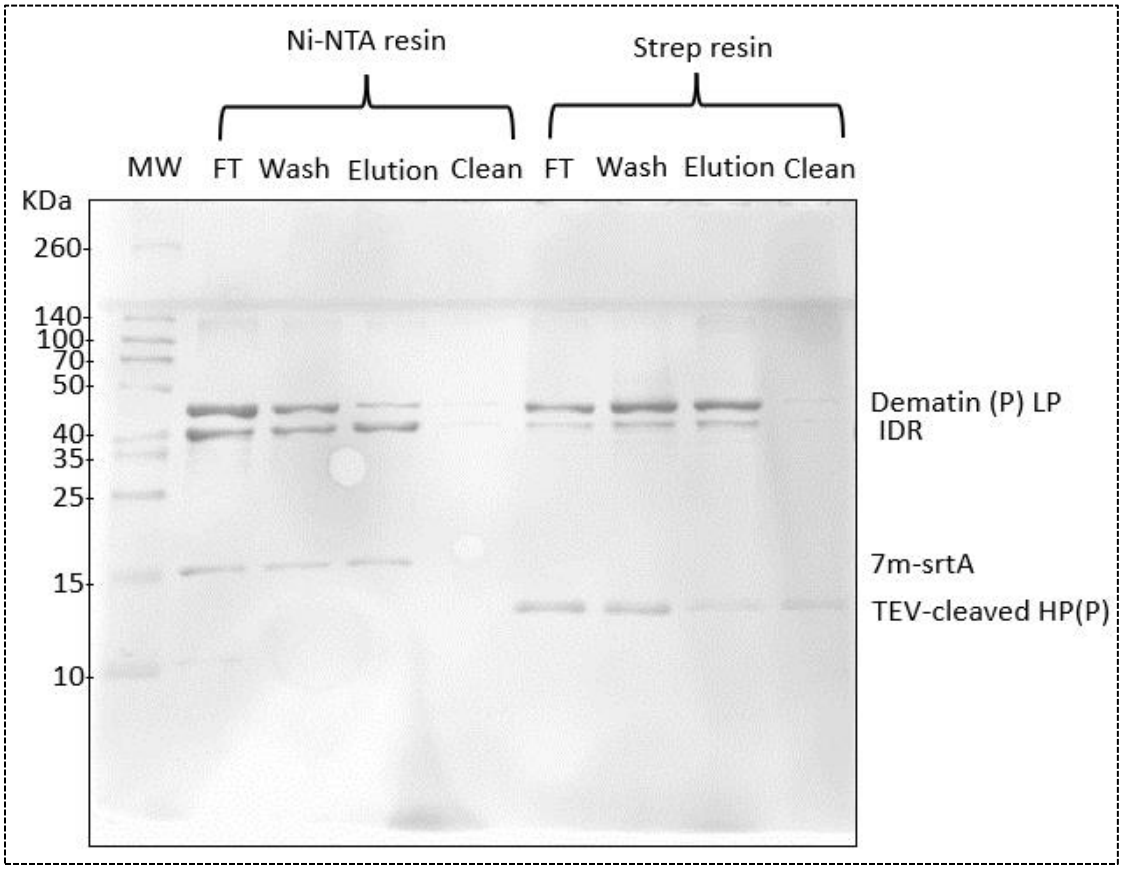


Figure 3.2.6. 15% SDS-PAGE of Ni-NTA and Strep resin following IMAC and strep purification of dematin (P) LP in SML reaction mixture. Ni-NTA and strep resin was sampled after each step performed as part of IMAC and Strep-Tag purification protocols (flow through (FT), wash, and elution collection).

Because only 3 μM dematin (P) LP could be isolated from this two-step purification process, size-exclusion chromatography (SEC) was attempted in order to separate the components of the reaction mixture in an effort to improve product yield (**Figures 3.2.7** and **3.2.8**).

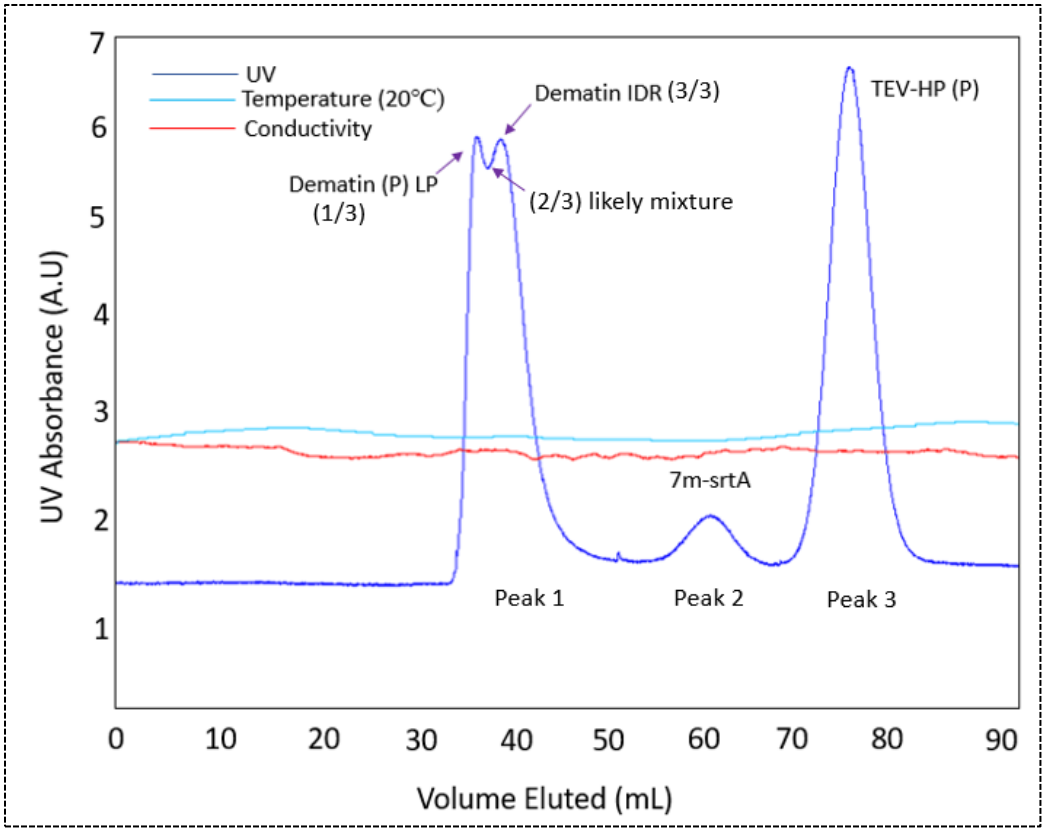


Figure 3.2.7 Purification of dematin SML reaction mixture via size-exclusion chromatography (SEC). SEC chromatogram of SML reaction mixture. The protein samples are eluted at pH 6.8 in 2 M urea, 20 mM PIPES 50 mM NaCl. Each peak corresponds to a specific component indicated in the above graph. Peak 1 is collected over 3 separate fractions (1/3, 2/3, 3/3) while peak 2 and 3 span two fractions each (1/2, 2/2).

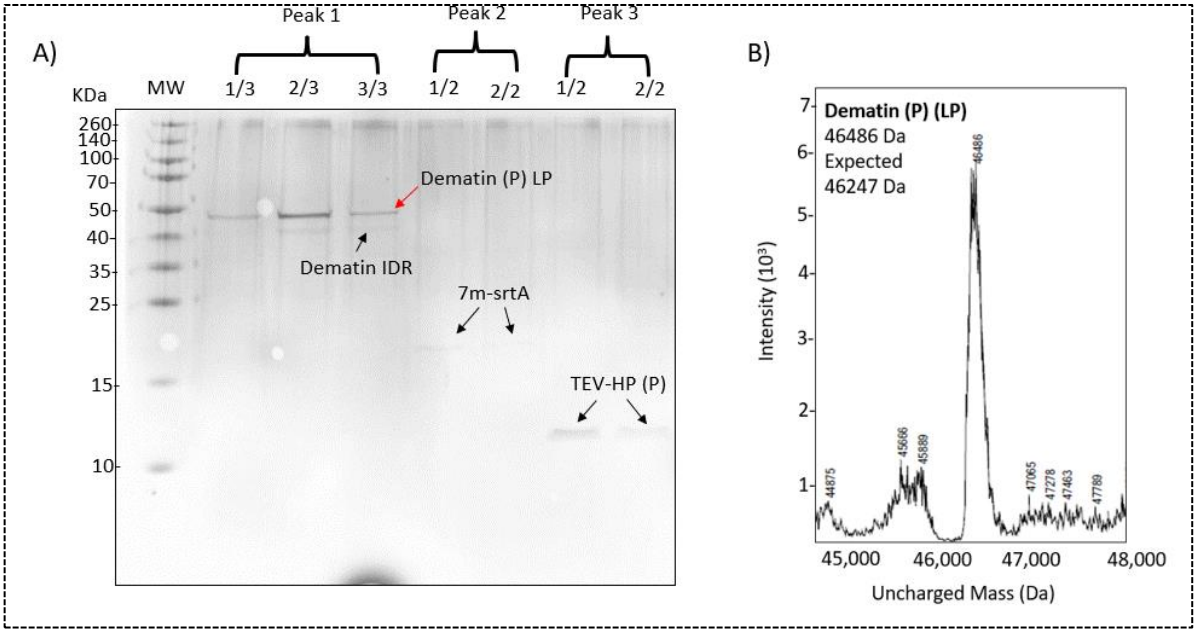


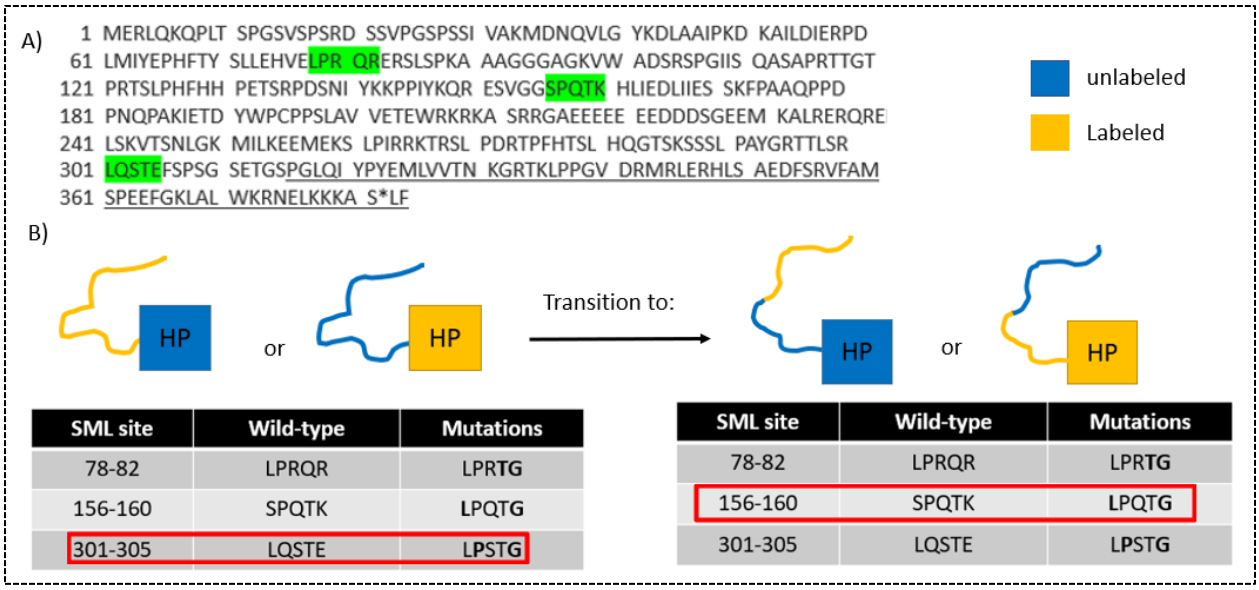
Figure 3.2.8. 15% SDS-PAGE of collected SEC fractions. See **Figure 3.2.7**. A) Peak 1 is collected over 3 separate fractions (1/3, 2/3, 3/3) while peak 2 and 3 span two fractions each (1/2, 2/2). B) LC-ESI-MS spectrum of fraction containing peak 1 (1/3)

While SEC successfully separated sortase and headpiece from the ligation product, it could not isolate it from the unreacted IDR construct due to proximity in molecular weight (38.4 KDa IDR, 46.4 KDa LP). Another issue involves the dramatic dilution incurred by the ligation product. Initially in a 1-mL-reaction, it was now collected as part of a 25-mL-fraction. LC-QTOF-MS analysis of the SEC fraction containing the first third of peak 1 (**Figures 3.2.7** and **3.2.8**) detected a peak at 46486 Da which is 239 Da over the expected mass of 46247 Da for the ligation product. While we suspect the presence of non-covalent adducts (e.g. Na^+) to contribute to this mass discrepancy, instrument accuracy may be low due to overall low signal intensity. Attempts to concentrate the ligation product in the above-mentioned SEC fraction was made: centrifugation in Thermo Scientific Pierce spin concentrators as well as lyophilization-rehydration experiments were executed. Both cases resulted in an unexplained loss of the

protein product. In the case of lyophilization, salts contained int the storage buffer (urea, PIPES, and NaCl) were likely concentrated beyond the compatibility range for dematin (P) LP[37,42,43] , leading to its inevitable destabilization. Indeed, proteins are often unable to redissolve after being lyophilized, suggesting they were denatured throughout this process[44].

3.2.1 Future Works in Aim II:

The use of SEC allowed for the efficient separation of the headpiece construct and sortase enzyme from the ligation product. This promising outcome led to the design of new fragments utilizing the LPXTG site at 156-160 position (**Figure 3.2.5**).



SML site	Wild-type	Mutations
78-82	LPRQR	LPR**TG**
156-160	SPQTK	L**PQTG**
301-305	LQSTE	L**PSTG**

SML site	Wild-type	Mutations
78-82	LPRQR	LPR**TG**
156-160	SPQTK	L**PQTG**
301-305	LQSTE	L**PSTG**

Figure 3.2.9. Change of dematin constructs utilized. A) wild-type (WT) dematin sequence. Potential LPXTG sites are highlighted in green. The headpiece sequence is underlined. B) Change of LPXTG site used from positions 301-305 to 156-160 leading to novel segmentally labeled dematin constructs possible. The LPXTG motifs shown are obtained upon 2 point-mutations to the WT sequences indicated.

In this new scenario, a serine and a lysine are replaced by a leucine and a glycine, respectively. With this change in mind, the fragment including the headpiece and the C-terminal IDR portion (C-IDR-HP, 25 KDa when TEV-cleaved to expose the N-terminal glycine nucleophile) and the one

composed of the N-terminal IDR segment (N-IDR, 21 KDa) will form a ligation product of about 46 KDa. Because 7m-srtA is around 18 KDa, this design should allow for easy separation of the dematin LP from other SML components through SEC. Both dematin C-IDR-HP (~8 mL at about 97 µM) and N-IDR (~4 mL at about 32 µM) have been successfully produced from separate 1L-bacterial growths (**Figures 3.2.6**). They are currently undergoing optimization studies in term of their yield, purity, and TEV-cleavage (for C-IDR-HP) before SML can take place. At the time of writing this thesis, Mass spectrometry has yet to be performed on these two new fragments.

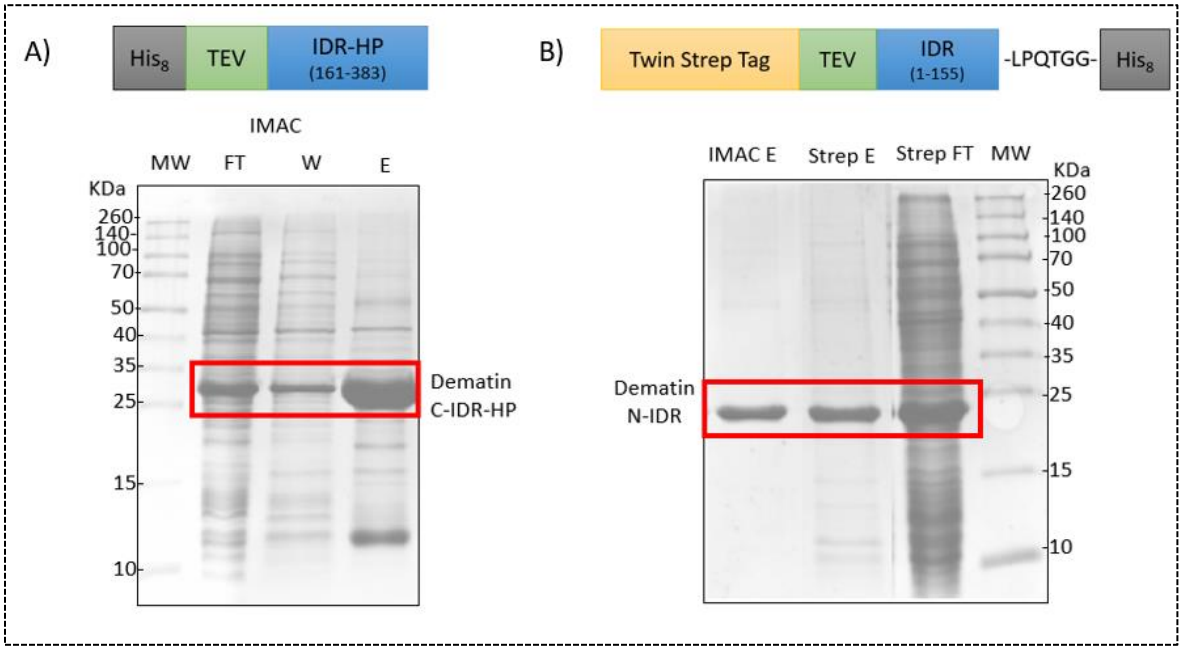


Figure 3.2.10. Production of New Dematin fragments, C-IDR-HP and N-IDR. A) structure and 15% SDS-PAGE of dematin C-IDR-HP purified through IMAC. The flow through (FT), wash (W) and elution (E) fractions are shown. B) Structure of dematin N-IDR segment purified through strep affinity chromatography (SAC), elution fractions collected from SAC underwent IMAC purification.

Desalting the ligation product may also be considered prior to lyophilization to minimize sample loss during the concentration process[42,44]. When successful isolation of the ligation product at adequate yield (50-100 µM) is achieved, a segmentally labeled sample can then be made for

analysis via solution NMR spectroscopy to investigate the suspected intramolecular interaction between the phosphorylated headpiece and IDR.

3.3 Aim III Results and Discussion:

Aim III: Optimize the *in vitro* synthesis of the Full-Length Villin 4 IDR (*Arabidopsis thaliana*) through SML.

To perform any structure characterization studies or actin binding assays, the full-length villin 4 IDR first needs to be synthesized and purified. Our ultimate goal is to decipher the mechanism through which this disordered segment contributes to villin 4's F-actin binding and bundling activity. Such knowledge would help elucidate the way plant root growth is regulated. To produce this IDR, two avenues were explored in parallel. The first one consisted of separately expressing the acidic and basic stretch of the linker and ligating the two IDR portions via SML. This strategy was initially favored because of the possibility of generating a segmentally labeled sample for solution NMR analysis. The acidic portion attached to the folded headpiece (AtV4IDR-A-HP) was successfully produced as ~3mL at 112 µM were made from a 1L-bacterial growth (**figure 3.3.1**) and TEV-cleaved (**Figure 3.3.2**). While the TEV-cleaved AtV4IDR-A-HP (TEV-AtV4IDR-A-HP) appears as 15-KDa band on a gel, analysis of this same sample via LC-ESI-MS shows it to have a mass of 11984.7 Da which matches our expected calculated mass of 11983 Da when instrumental error is taken into account. Thus, in our lab, SDS-PAGE data is typically only used to assess protein purity while MS is utilized confirm our constructs' identity. The basic part of the villin 4 IDR (AtV4IDR-B) was expressed and purified (~3 mL at around 88 µM/ 1L bacterial growth) in parallel to TEV-AtV4IDR-A-HP (**Figure 3.3.3**).

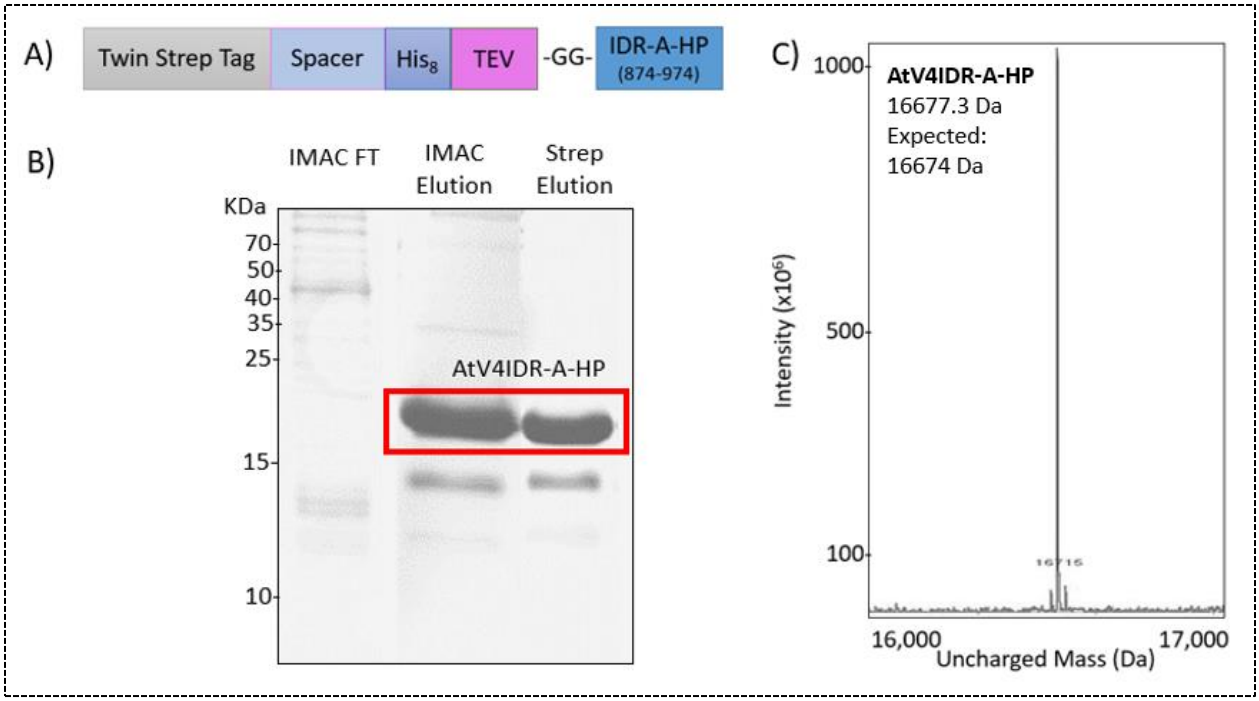


Figure 3.3.1. Synthesis and purification of AtV4IDR-A-HP. A) structure of the AtV4IDR-A-HP segment. B) 15% SDS-PAGE of AtV4IDR-A-HP. C) LC-ESI-MS of AtV4IDR-A-HP.

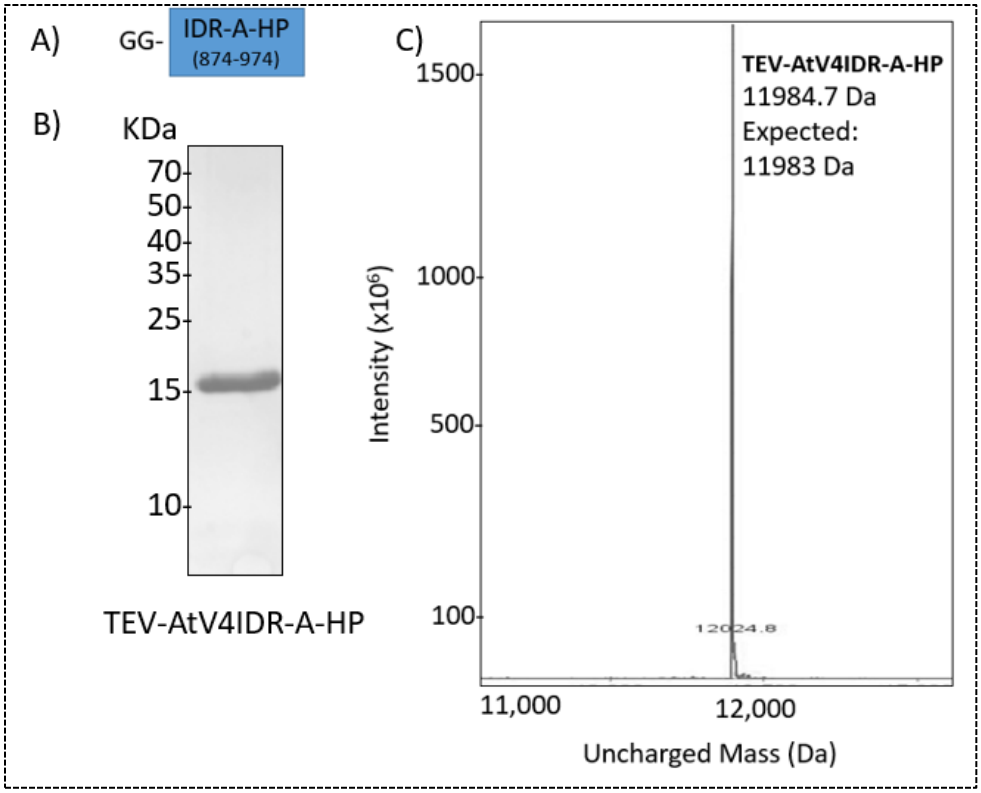


Figure 3.3.2. Production of TEV-AtV4IDR-A-HP. A) structure of the TEV-AtV4IDR-A-HP segment. B) 15% SDS-PAGE of TEV-AtV4IDR-A-HP. C) LC-ESI-MS of TEV-AtV4IDR-A-HP.

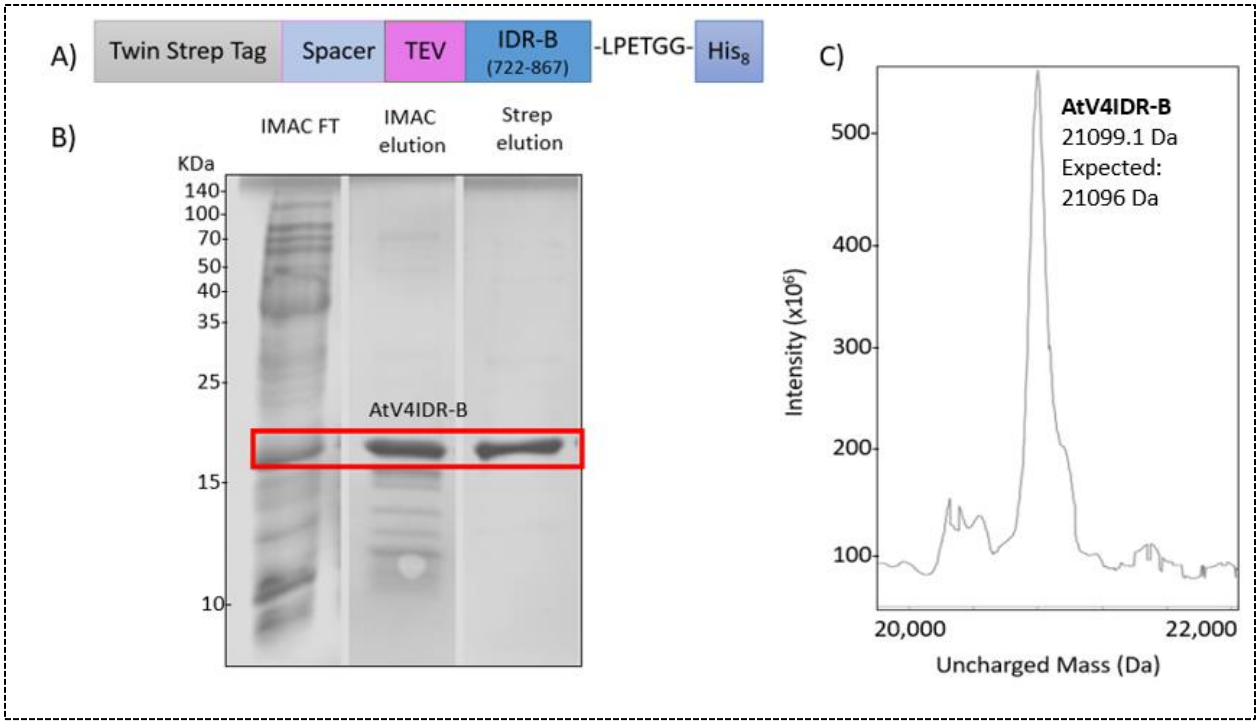


Figure 3.2.3. Synthesis and purification of AtV4IDR-B. A) structure of the AtV4IDR-B segment. B) 15% SDS-PAGE of AtV4IDR-B. C) LC-ESI-MS of AtV4IDR-B.

SML between AtV4IDR-B and AtV4IDR-A-HP utilized 7m-srtA and was monitored for 48 hours. The ligation product has an expected molecular weight of 31.7KDa but 15% SDS-PAGE analysis revealed one main band at around ~38 KDa absent in the controls and at the zero-hour timepoint (**Figure 3.3.4**).

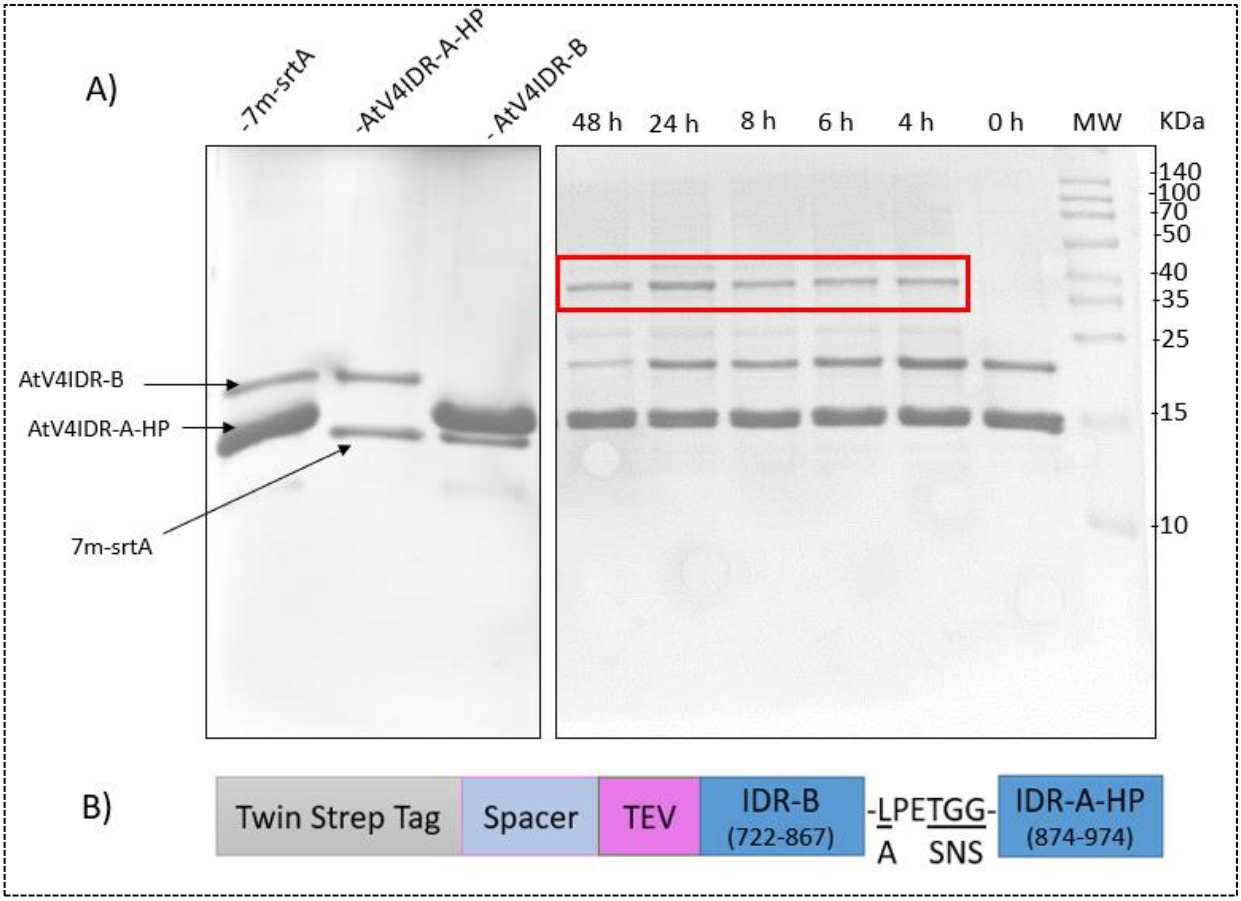


Figure 3.3.4. Sortase-mediated ligation between AtV4IDR-B and TEV-AtV4IDR-A-HP. A) 15% SDS- PAGE of SML monitored periodically for 48 hours. The reaction was performed alongside controls at room temperature using 7m-srtA. B) Structure of the expected ligation product. Residues within the PEST motif were mutated to create the LPXTG site with an extra glycine to improve the reaction rate[40]. Wild-type amino acids are specified directly underneath the substitution site.

This band is suspected to be the ligation product (LP) despite its molecular weight not matching our expected calculated LP mass. This discrepancy is not unusual when performing SDS-PAGE analysis. Indeed, 7m-srtA which is ~18 KDa and TEV-AtV4IDR-A-HP which is ~11.9 KDA both appear as ~15-KDa bands. Furthermore, the absence of this 38-KDa band in the controls and at the zero-hour time point suggests that the ligation product did indeed form. The presence of other faint bands (e.g. at ~25KDa) correspond to impurities which can be observed as part of

the AtV4IDR-B sample (**Figure 3.3.3**) and thus are unlikely to be the ligation product. LC-ESI-MS analysis of the reaction mixture however failed to confirm the presence of any of the SML components including the villin 4 LP likely due to very low signal-to-noise ratio. These results are thus preliminary. More optimization is needed to synthesize then isolate the villin 4 IDR via SML. The use of other sortase variants, different molar ratios to have excess of substrates in order to minimize reversibility, as well as manipulation of temperature, buffer conditions, and pH are currently under consideration to address this issue. Because problems with the SML scheme was anticipated, efforts were invested in parallel to the above experiments to express and purify the full-size linker as a single construct (AtV4IDR), bypassing the need for SML steps (**Figure 3.3.5**). While this alternative prevents any possibility for segmental labeling of IDR portions, it would still allow us to subject the IDR to F-actin binding assays. A one-liter expression growth yielded 5 mL of AtV4IDR at 25 μM. By increasing the volume of bacterial culture, enough AtV4IDR was produced for actin-binding and covalent modification studies performed by fellow graduate students Jake Heins (thesis expected to be published in Spring 2023) and Derek McCaffrey (thesis expected to be published in Summer 2022), respectively.

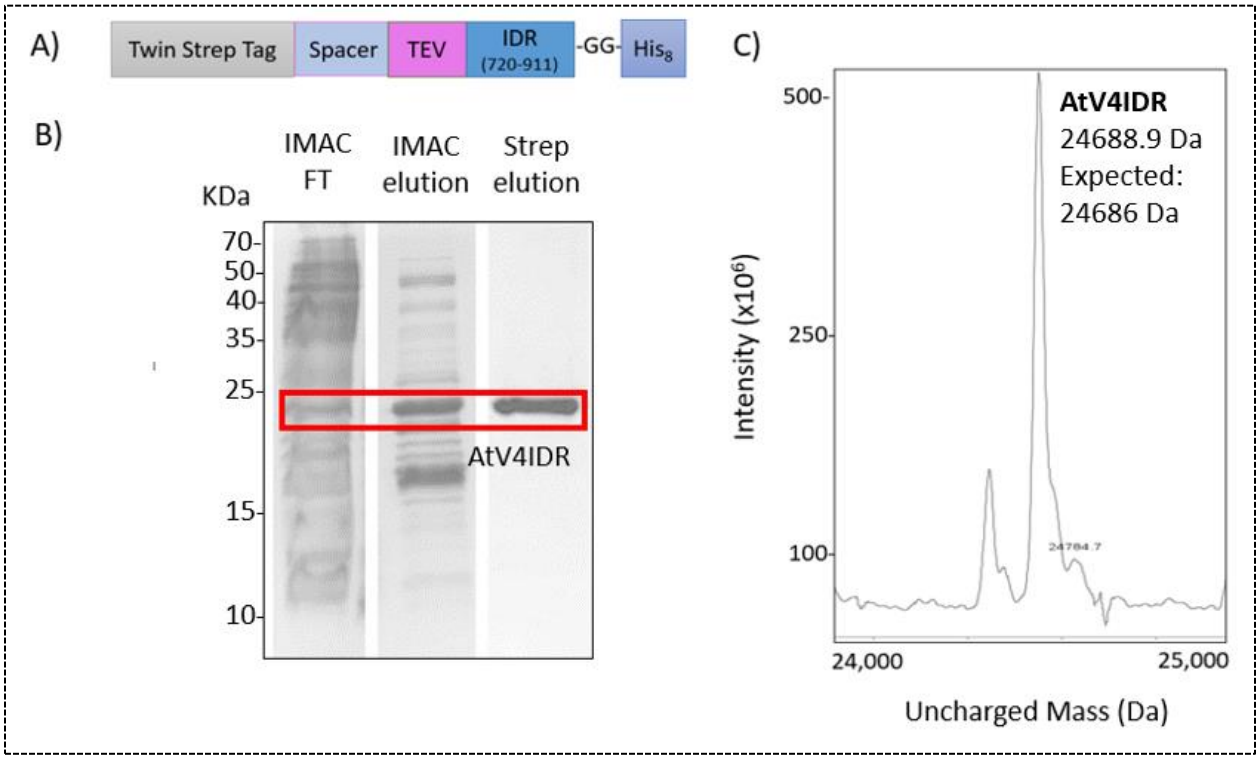


Figure 3.1.5. Synthesis of full-length villin 4 IDR (AtV4IDR). A) Structure of AtV4IDR including twin strep and histidine tags for purification purposes. B) 15% SDS-PAGE of AtV4IDR. C) LC-ESI-MS chromatogram of AtV4IDR

3.3.1 Future Works in Aim III:

Expression of AtV4IDR in minimal media (M9) for solution NMR spectroscopy studies is being optimized. We intend to produce a uniformly ^{15}N-labeled sample and collect corresponding ^{15}N-HSQC spectra to gain a better understanding of the IDR's various conformational changes. A new villin 4 IDR construct containing TEV-cleavable tags is being designed such that twin strep and His_6 could both be removed prior to any NMR analysis to not affect the resonances of native IDR residues. The segmental labeling of the villin 4 linker may be revisited through the use of different sortase variants, buffer conditions, or substrate ratios.

Chapter IV: Conclusion

Large IDRs/IDPs play vital roles in cellular processes. Their dynamic nature allows them to successfully perform various functions but this same characteristic represents one of the main challenges when attempting to study their interactions and mode of operation. 2D heteronuclear solution NMR spectroscopy has proven to be a unique tool with the capacity to circumvent such difficulties, but the overlap of peaks observed in ^{15}N-HSQC spectra collected complicates the analysis process. Segmental labeling (SL), achieved through sortase-mediated ligation reactions, aims to address the spectral convolution while maintaining the native amino acid environment of the protein domain under investigation. Because SML has yet to be performed in the context of IDR sequences, there is a great need to optimize the reaction conditions under such unique circumstances. Indeed, our work indicates that in many cases, a sortase-mediated ligation between two halves of authentic IDR segments is problematic likely because of the identity of the amino acids surrounding the LPXTG site from the N- and C-terminal ends. Strategies to improve the ligation efficiency are currently under development in our lab. The addition of glycine spacers, the use of different sortase variants or manipulation of reaction parameters such as buffer composition, pH, and temperature are presently being tested to arrive at a robust optimized protocol, universally applicable to all large IDRs. Two model cytoskeletal regulator proteins were used as molecular test subjects to our SML optimization studies, villin 4 (*Arabidopsis thaliana*) and dematin (*Homo sapiens*). In the case of dematin, SML which occurred within its large IDR was successful in synthesizing the full-length protein. This construct used contained the S381E mutation to mimic a phosphorylated headpiece state. However, failure to isolate the ligation product from other reaction

components in high enough yield currently prevents the collection of adequate ^{15}N-HSQC data. New dematin fragments designed to significantly differ in terms of their molecular weight from the ligation product should allow for successful purification via SEC. A segmentally labeled dematin sample should then be synthesized and analyzed via solution NMR spectroscopy to determine whether the previously suspected intramolecular interaction occurs between dematin's IDR and its phosphorylated headpiece. If such binding is observed, identify of the residues involved could be inferred from the location of ^{15}N HSQC peaks. Using SML to synthesize the full-length villin 4 IDR was not as successful as the dematin ligation reaction. Fortunately a full-length villin 4 IDR was produced as one fragment. While segmental labeling may not be possible for this construct, its ability to bind filamentous actin and undergo covalent modifications was successfully demonstrated by fellow graduate students Jake Heins and Derek McCaffrey under Dr. Sergey Smirnov's supervision. Efforts to generate a uniformly isotopically labeled AtV4IDR sample are underway to study the various conformational changes this IDR is capable of adopting.

To ensure that the mutations introduced to create the LPXTG motif do not impact structure, function, and interactions of the relevant protein constructs, ^{15}N-HSQC of wild-type sequences will eventually be collected and compared with the spectra obtained from our modified sequences. Actin-binding assays shall also be conducted to ensure that dematin's function is preserved.

Because villin proteins are implicated in every developmental phase of root hair from the initial bulge formation to hair elongation[45], understanding villin 4's mode of operation would lay the foundations necessary to allow us to eventually control every step of such a process in order to

engineer plants with longer, more abundant root hair with larger surface areas, capable of absorbing more water and nutrients, ultimately producing crops at a higher yield which would also be suited for high salinity and dry soil conditions. While our work on villin 4 aims to enrich the agricultural field, our study of dematin may shed some light on the way through which it interacts with filamentous actin in erythrocytes but also in prostate and breast tissues, bringing us one step closer to understanding its cancer-suppressing properties[1,7,8,38] as this protein represents a potential alternative or complement to current cancer therapeutics.

Bibliography

1. Babu MM. The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochemical Society Transactions*. 2016;44(5):1185-1200. doi:10.1042/BST20160172

2. Forman-Kay JD, Mittag T. From Sequence and Forces to Structure, Function, and Evolution of Intrinsically Disordered Proteins. *Structure*. 2013;21(9):1492-1499. doi:10.1016/J.STR.2013.08.001

3. Bogatyreva NS, Finkelstein A v., Galzitskaya O v. TREND OF AMINO ACID COMPOSITION OF PROTEINS OF DIFFERENT TAXA. *http://dx.doi.org/101142/S0219720006002016*. 2011;4(2):597-608. doi:10.1142/S0219720006002016

4. Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. *Nature Reviews Molecular Cell Biology 2015 16:1*. 2014;16(1):18-29. doi:10.1038/nrm3920

5. Korneta I, Bujnicki JM. Intrinsic disorder in the human spliceosomal proteome. *PLoS Comput Biol*. 2012;8(8). doi:10.1371/JOURNAL.PCBI.1002641

6. Lin Y, Protter DSW, Rosen MK, Parker R. Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Molecular Cell*. 2015;60(2):208-219. doi:10.1016/J.MOLCEL.2015.08.018/ATTACHMENT/2CF70017-4ABD-472C-B13D-9BEECD87AB63/MMC5.MP4

7. Vacic V, Iakoucheva LM. Disease mutations in disordered regions—exception to the rule? *Molecular BioSystems*. 2011;8(1):27-32. doi:10.1039/C1MB05251A

8. Uversky VN, Oldfield CJ, Dunker AK. Intrinsically Disordered Proteins in Human Diseases: Introducing the D2 Concept. *http://dx.doi.org/101146/annurev.biophys37032807125924*. 2008;37:215-246. doi:10.1146/ANNUREV.BIOPHYS.37.032807.125924

9. Graether SP. Troubleshooting guide to expressing intrinsically disordered proteins for use in NMR experiments. *Frontiers in Molecular Biosciences*. 2019;5(JAN):118. doi:10.3389/FMOLB.2018.00118/BIBTEX

10. Uversky VN. Paradoxes and wonders of intrinsic disorder: Stability of instability. *https://doi.org/101080/2169070720171327757*. 2017;5(1):e1327757. doi:10.1080/21690707.2017.1327757

11. Nováček J, Janda L, Dopitová R, Žídek L, Sklenář V. Efficient protocol for backbone and side-chain assignments of large, intrinsically disordered proteins: transient secondary structure analysis of 49.2 kDa microtubule associated protein 2c. *J Biomol NMR*. 2013;56(4):291-301. doi:10.1007/S10858-013-9761-7

12. Busche AEL, Aranko AS, Talebzadeh-Farooji M, Bernhard F, Dötsch V, Iwaï H. Segmental Isotopic Labeling of a Central Domain in a Multidomain Protein by Protein Trans-Splicing Using Only One Robust DnaE Intein. *Angewandte Chemie International Edition*. 2009;48(33):6128-6131. doi:10.1002/ANIE.200901488

13. Mikula KM, Krumwiede L, Plückthun A, Iwaï H. Segmental isotopic labeling by asparaginyl endopeptidase-mediated protein ligation. *Journal of Biomolecular NMR 2018 71:4*. 2018;71(4):225-235. doi:10.1007/S10858-018-0175-4

14. Minato Y, Ueda T, MacHiyama A, Shimada I, Iwaï H. Segmental isotopic labeling of a 140 kDa dimeric multi-domain protein CheA from Escherichia coli by expressed protein ligation and protein trans-splicing. *Journal of Biomolecular NMR*. 2012;53(3):191-207. doi:10.1007/S10858-012-9628-3/TABLES/3

15. Muralidharan V, Muir TW. Protein ligation: an enabling technology for the biophysical analysis of proteins. *Nature Methods 2006 3:6*. 2006;3(6):429-438. doi:10.1038/nmeth886

16. Debelouchina GT, Muir TW. A molecular engineering toolbox for the structural biologist. *Q Rev Biophys*. 2017;50. doi:10.1017/S0033583517000051

17. Ma J, Zeng J, Wan Q. Postligation-desulfurization: a general approach for chemical protein synthesis. *Top Curr Chem*. 2015;363:57-102. doi:10.1007/128_2014_594

18. Spirig T, Weiner EM, Clubb RT. Sortase enzymes in Gram-positive bacteria. *Molecular Microbiology*. 2011;82(5):1044-1059. doi:10.1111/J.1365-2958.2011.07887.X

19. Freund C, Schwarzer D. Engineered Sortases in Peptide and Protein Chemistry. *Chembiochem*. 2021;22(8):1347. doi:10.1002/CBIC.202000745

20. Podracky CJ, An C, DeSousa A, Dorr BM, Walsh DM, Liu DR. Laboratory evolution of a sortase enzyme that modifies amyloid-β protein. *Nature Chemical Biology 2021 17:3*. 2021;17(3):317-325. doi:10.1038/s41589-020-00706-1

21. 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. doi:10.1002/BIOT.201500012

22. Wu Q, Ploegh HL, Truttmann MC. Hepta-Mutant Staphylococcus aureus Sortase A (SrtA7m) as a Tool for in Vivo Protein Labeling in Caenorhabditis elegans. *ACS Chemical Biology*. 2017;12(3):664-673. doi:10.1021/ACSCHEMBIO.6B00998/SUPPL_FILE/CB6B00998_SI_003.XLSX

23. Structure and Organization of Actin Filaments - The Cell - NCBI Bookshelf. Accessed July 3, 2022. https://www.ncbi.nlm.nih.gov/books/NBK9908/

24. Hohmann T, Dehghani F. The Cytoskeleton—A Complex Interacting Meshwork. *Cells 2019, Vol 8, Page 362*. 2019;8(4):362. doi:10.3390/CELLS8040362

25. Actin, Myosin, and Cell Movement - The Cell - NCBI Bookshelf. Accessed July 3, 2022. https://www.ncbi.nlm.nih.gov/books/NBK9961/

26. Sun HQ, Yamamoto M, Mejillano M, Yin HL. Gelsolin, a Multifunctional Actin Regulatory Protein *. *Journal of Biological Chemistry*. 1999;274(47):33179-33182. doi:10.1074/JBC.274.47.33179

27. 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. doi:10.1111/J.1469-8137.2010.03632.X

28. Miears HL, Gruber DR, Horvath NM, et al. Plant Villin Headpiece Domain Demonstrates a Novel Surface Charge Pattern and High Affinity for F-Actin. *Biochemistry*. 2018;57(11):1690-1701. doi:10.1021/ACS.BIOCHEM.7B00856/SUPPL_FILE/BI7B00856_SI_001.PDF

29. Fedechkin SO, Brockerman J, Luna EJ, Lobanov MY, Galzitskaya O v., Smirnov SL. An N-terminal, 830-residue Intrinsically Disordered Region of the Cytoskeleton-regulatory Protein Supervillin Contains Myosin II- and F-actin- Binding Sites. *J Biomol Struct Dyn*. 2013;31(10):1150. doi:10.1080/07391102.2012.726531

30. Sekhar KR, Freeman ML. PEST sequences in proteins involved in cyclic nucleotide signalling pathways. *Journal of Receptor and Signal Transduction Research*. 1998;18(2-3):113-132. doi:10.3109/10799899809047740

31. Fields GB. Introduction to Peptide Synthesis. *Current Protocols in Protein Science*. 2001;26(1):18.1.1-18.1.9. doi:10.1002/0471140864.PS1801S26

32. Boyko K v., Rosenkranz EA, Smith DM, et al. Sortase-mediated segmental labeling: A method for segmental assignment of intrinsically disordered regions in proteins. *PLoS One*. 2021;16(10). doi:10.1371/JOURNAL.PONE.0258531

33. Khanna R, Chang SH, Andrabi S, et al. Headpiece domain of dematin is required for the stability of the erythrocyte membrane. *Proc Natl Acad Sci U S A*. 2002;99(10):6637-6642. doi:10.1073/PNAS.052155999/ASSET/DC5C36D6-1450-4418-B649-F9363D9ACBCB/ASSETS/GRAPHIC/PQ1021559005.JPEG

34. Chen L, Jiang ZG, Khan AA, Chishti AH, McKnight CJ. Dematin exhibits a natively unfolded core domain and an independently folded headpiece domain. *Protein Science*. 2009;18(3):629-636. doi:10.1002/PRO.59

35. Mohseni M, Chishti AH. Regulatory models of RhoA suppression by dematin, a cytoskeletal adaptor protein. *http://dx.doi.org/104161/cam327375*. 2009;3(2):191-194. doi:10.4161/CAM.3.2.7375

36. Chen L, Brown JW, Mok YF, Hatters DM, McKnight CJ. The Allosteric Mechanism Induced by Protein Kinase A (PKA) Phosphorylation of Dematin (Band 4.9) *. *Journal of Biological Chemistry*. 2013;288(12):8313-8320. doi:10.1074/JBC.M112.438861

37. Lu Y, Hanada T, Fujiwara Y, et al. Gene disruption of dematin causes precipitous loss of erythrocyte membrane stability and severe hemolytic anemia. *Blood*. 2016;128(1):93-103. doi:10.1182/BLOOD-2016-01-692251

38. Lutchman M, Pack S, Kim AC, et al. Loss of Heterozygosity on 8p in Prostate Cancer Implicates a Role for Dematin in Tumor Progression. *Cancer Genetics and Cytogenetics*. 1999;115(1):65-69. doi:10.1016/S0165-4608(99)00081-3

39. Rosenkranz E. Engineering Segmentally Labeled Intrinsically Disordered Proteins. *WWU Graduate School Collection*. Published online January 1, 2021. Accessed July 29, 2022. https://cedar.wwu.edu/wwuet/1060

40. Pritz S, Wolf Y, Kraetke O, Klose J, Bienert M, Beyermann M. Synthesis of Biologically Active Peptide Nucleic Acid−Peptide Conjugates by Sortase-Mediated Ligation. *The Journal of Organic Chemistry*. 2007;72(10):3909-3912. doi:10.1021/JO062331L/SUPPL_FILE/JO062331LSI20070404_030014.PDF

41. Morgan HE, Turnbull WB, Webb ME. Challenges in the use of sortase and other peptide ligases for site-specific protein modification. *Chemical Society Reviews*. 2022;51(10):4121-4145. doi:10.1039/D0CS01148G

42. Parida AK, Das AB, Mittra B, Mohanty P. Salt-stress induced alterations in protein profile and protease activity in the mangrove Bruguiera parviflora. *Zeitschrift fur Naturforschung C, Journal of biosciences*. 2004;59(5-6):408-414. doi:10.1515/ZNC-2004-5-622

43. Sinha R, Khare SK. Protective role of salt in catalysis and maintaining structure of halophilic proteins against denaturation. *Frontiers in Microbiology*. 2014;5(APR):165. doi:10.3389/FMICB.2014.00165/BIBTEX

44. Maintaining protein stability through buffers, freezing, and lyophilization. Accessed July 3, 2022. https://opsdiagnostics.com/notes/ranpri/ProteinStability.html

45. Pei W, Du F, Zhang Y, He T, Ren H. Control of the actin cytoskeleton in root hair development. *Plant Science*. 2012;187:10-18. doi:10.1016/J.PLANTSCI.2012.01.008

Appendix

Section 1: Amino Acid Sequences of Samples Discussed.

Proline-Rich C-terminus (ProC): DisProt ID DP00775
N-terminal ligation partner: protein substrate
Plasmid name: ProC_AroN_pET-24a(+)
Extinction coefficient: 1490 M^{-1} cm^{-1}
MPSVQEVEKLLHVLDRNGDGKVSAEELKAFADDSKCPLDSNKIKAFIKEHDKNKDGKLDLKELVSILSSGTSEN
LYFQGGGGGSGGGGSKEELHLPMTGGHHHHHH
C-terminal ligation partner: hexapeptide GPPPPPK(DNP)
ProC ligation product (LP)
MPSVQEVEKLLHVLDRNGDGKVSAEELKAFADDSKCPLDSNKIKAFIKEHDKNKDGKLDLKELVSILSSGTSEN
LYFQGGGGGSGGGGSKEELHLPMTGPPPPPK(DNP)

Aromatic N-terminus (AroN): DisProt ID DP00088
N-terminal ligation partner: protein substrate
Plasmid name: AroN_AroN_pET-24a(+)
Extinction coefficient: 17990 M^{-1} cm^{-1}
MPSVQEVEKLLHVLDRNGDGKVSAEELKAFADDSKCPLDSNKIKAFIKEHDKNKDGKLDLKELVSILSSGTSEN
LYFQGGGGGSGGGGSWHIWWLPITGGHHHHHH
C-terminal ligation partner: hexapeptide: GFAGMIK(DNP)
AroN ligation product (LP):
MPSVQEVEKLLHVLDRNGDGKVSAEELKAFADDSKCPLDSNKIKAFIKEHDKNKDGKLDLKELVSILSSGTSEN
LYFQGGGGGSGGGGSWHIWWLPITGFAGMIK(DNP)

Aromatic C-terminus (AroC): DisProt ID DP02488
N-terminal ligation partner: protein substrate
Plasmid name: AroC_AroN_pET-24a(+)
Extinction coefficient: 1490 M^{-1} cm^{-1}
MPSVQEVEKLLHVLDRNGDGKVSAEELKAFADDSKCPLDSNKIKAFIKEHDKNKDGKLDLKELVSILSSGTSEN
LYFQGGGGGSGGGGSNPEAPLPVTGGHHHHHH
C-terminal ligation partner: hexapeptide: GYYHWDK(DNP)
AroC ligation product (LP):
MPSVQEVEKLLHVLDRNGDGKVSAEELKAFADDSKCPLDSNKIKAFIKEHDKNKDGKLDLKELVSILSSGTSEN
LYFQGGGGGSGGGGSNPEAPLPVTGYYHWDK(DNP)

Polar Charged N-terminus (PolN): DisProt ID DP00356
N-terminal ligation partner: protein substrate

Plasmid name: PolN_AroN_pET-24a(+)
Extinction coefficient: 1490 M^{-1} cm^{-1}
MPSVQEVEKLLHVLDRNGDGKVSAEELKAFADDSKCPLDSNKIKAFIKEHDKNKDGKLDLKELVSILSSGTSEN
LYFQGGGGGSGGGGSDRIKELPETGGHHHHHH
C-terminal ligation partner: hexapeptide: GLGQNEK(DNP)
PolN ligation product (LP):
MPSVQEVEKLLHVLDRNGDGKVSAEELKAFADDSKCPLDSNKIKAFIKEHDKNKDGKLDLKELVSILSSGTSEN
LYFQGGGGGSGGGGSDRIKELPETGLGQNEK(DNP)
PolN system Controls:
PolN-5Gly ligation product (LP)
MPSVQEVEKLLHVLDRNGDGKVSAEELKAFADDSKCPLDSNKIKAFIKEHDKNKDGKLDLKELVSILSSGTSEN
LYFQGGGGGSGGGGSDRIKELPETGGGGGK(DNP)
5GlyN-PolN ligation product (LP)
MPSVQEVEKLLHVLDRNGDGKVSAEELKAFADDSKCPLDSNKIKAFIKEHDKNKDGKLDLKELVSILSSGTSEN
LYFQGGGGGSGGGGSGGGGGLPETGLGQNEK(DNP)

Glycine Control (GlyC)
N-terminal glycine control: protein substrate (also used as N-terminal ligation partner for 5GlyN-PolN LP)
Plasmid name: GGGGG_AroN_pET-24a(+)
Extinction coefficient: 1490 M^{-1} cm^{-1}
MPSVQEVEKLLHVLDRNGDGKVSAEELKAFADDSKCPLDSNKIKAFIKEHDKNKDGKLDLKELVSILSSGTSEN
LYFQGGGGGSGGGGSGGGGGLPETGGHHHHHH
C-terminal glycine control: pentapeptide (also used as C-terminal ligation partner for PolN-5Gly LP) GGGGGK(DNP)
MPSVQEVEKLLHVLDRNGDGKVSAEELKAFADDSKCPLDSNKIKAFIKEHDKNKDGKLDLKELVSILSSGTSEN
LYFQGGGGGSGGGGSGGGGGLPETGGGGGK(DNP)

Dematin (HP)
Plasmid name: Dem_SML301-305_Cterm_pET-24a(+)
Extinction coefficient: 9970 M^{-1} cm^{-1}
MHHHHHHENLYFQGFSPSGSETGSPGLQIYPYEMLVVTNKGRTKLPPGVDRMRLERHLSAEDFSRVFAMSP
EEFGKLALWKRNELKKKASLF

Dematin TEV-HP
Extinction coefficient: 8480 M^{-1} cm^{-1}

GFSPSGSETGSPGLQIYPYEMLVVTNKGRTKLPPGVDRMRLERHLSAEDFSRVFAMSPEEFGKLALWKRNEL
KKKASLF

Dematin (P) HP

Plasmid name: Dem_SML301-305_Cterm_Phos_pET-24a(+)

Extinction coefficient: 9970 M^{-1} cm^{-1}

MHHHHHHENLYFQGFSPSGSETGSPGLQIYPYEMLVVTNKGRTKLPPGVDRMRLERHLSAEDFSRVFAMSP
EEFGKLALWKRNELKKKAELF

Dematin (P) TEV-HP

Extinction coefficient: 8480 M^{-1} cm^{-1}

GFSPSGSETGSPGLQIYPYEMLVVTNKGRTKLPPGVDRMRLERHLSAEDFSRVFAMSPEEFGKLALWKRNEL
KKKAELF

Dematin IDR

Plasmid name: Strep-IDR-SML301-305-His6_pET-24a(+)

Extinction coefficient: 39420 M^{-1} cm^{-1}

MWSHPQFEKGSGGASWSHPQFEKGGGENLYFQSERLQKQPLTSPGSVSPSRDSSVPGSPSSIVAKMDNQV
LGYKDLAAIPKDKAILDIERPDLMIYEPHFTYSLLEHVELPRQRERSLSPKAAAGGGAGKVWADSRSPGIISQAS
APRTTGTPRTSLPHFHHPETSRPDSNIYKKPPIYKQRESVGGSPQTKHLIEDLIIESSKFPAAQPPDPNQPAKIET
DYWPCPPSLAVVETEWRKRKASRRGAEEEEEEEDDDSGEEMKALRERQREELSKVTSNLGKMILKEEMEKSL
PIRRKTRSLPDRTPFHTSLHQGTSKSSSLPAYGRTTLSRLPSTGGHHHHHH

Dematin (P) LP (utilizing the LPSTG site at position 301-305)

Extinction coefficient: 47900 M^{-1} cm^{-1}

MWSHPQFEKGSGGASWSHPQFEKGGGENLYFQSERLQKQPLTSPGSVSPSRDSSVPGSPSSIVAKMDNQV
LGYKDLAAIPKDKAILDIERPDLMIYEPHFTYSLLEHVELPRQRERSLSPKAAAGGGAGKVWADSRSPGIISQAS
APRTTGTPRTSLPHFHHPETSRPDSNIYKKPPIYKQRESVGGSPQTKHLIEDLIIESSKFPAAQPPDPNQPAKIET
DYWPCPPSLAVVETEWRKRKASRRGAEEEEEEEDDDSGEEMKALRERQREELSKVTSNLGKMILKEEMEKSL
PIRRKTRSLPDRTPFHTSLHQGTSKSSSLPAYGRTTLSRLPSTGFSPSGSETGSPGLQIYPYEMLVVTNKGRTKLP
PGVDRMRLERHLSAEDFSRVFAMSPEEFGKLALWKRNELKKKAELF

Dematin C-IDR-HP

Plasmid name: Dem_HP_nophos_156-160_ pET-24a(+)

Extinction coefficient: 23950 M^{-1} cm^{-1}

MHHHHHHHHENLYFQGHLIEDLIIESSKFPAAQPPDPNQPAKIETDYWPCPPSLAVVETEWRKRKASRRGAE
EEEEEEDDDSGEEMKALRERQREELSKVTSNLGKMILKEEMEKSLPIRRKTRSLPDRTPFHTSLHQGTSKSSSLP
AYGRTTLSRLQSTEFSPSGSETGSPGLQIYPYEMLVVTNKGRTKLPPGVDRMRLERHLSAEDFSRVFAMSPEE
FGKLALWKRNELKKKASLF

Dematin N-IDR
Plasmid name: DualStrep_IDR-156-160_ pET-24a(+)
Extinction coefficient: 25440 M^{-1} cm^{-1}
MWSHPQFEKGSGGASWSHPQFEKGGGENLYFQGERLQKQPLTSPGSVSPSRDSSVPGSPSSIVAKMDNQV
LGYKDLAAIPKDKAILDIERPDLMIYEPHFTYSLLEHVELPRSRERSLSPKAAAGGGAGKVWADSRSPGIISQAS
APRTTGTPRTSLPHFHHPETSRPDSNIYKKPPIYKQRESVGGLPQTGGHHHHHHHH

Dematin LP (utilizing LPQTG site at position 156-160)
Extinction coefficient: 47900 M^{-1} cm^{-1}
MWSHPQFEKGSGGASWSHPQFEKGGG ENLYFQGERLQKQPLTSPGSVSPSRDSSVPGSPSSIVAKMDNQ
VLGYKDLAAIPKDKAILDIERPDLMIYEPHFTYSLLEHVELPRSRERSLSPKAAAGGGAGKVWADSRSPGIISQA
SAPRTTGTPRTSLPHFHHPETSRPDSNIYKKPPIYKQRESVGGLPQTGHLIEDLIIESSKFPAAQPPDPNQPAKIE
TDYWPCPPSLAVVETEWRKRKASRRGAEEEEEEEDDDSGEEMKALRERQREELSKVTSNLGKMILKEEMEKS
LPIRRKTRSLPDRTPFHTSLHQGTSKSSSLPAYGRTTLSRLQSTEFSPSGSETGSPGLQIYPYEMLVVTNKGRTKL
PPGVDRMRLERHLSAEDFSRVFAMSPEEFGKLALWKRNELKKKASLF

AtV4IDR
Plasmid name: sABh_pET-24a(+)
Extinction coefficient: 13980 M^{-1} cm^{-1}
MWSHPQFEKGSGGASWSHPQFEKGGGDSSKSAMHGNSFQRKLKIVKNGGTPVADKPKRRTPASYGGR
ASVPDKSQQRSRSMSFSPDRVRVRGRSPAFNALAATFESQNARNLSTPPPVVRKLYPRSVTPDSSKFAPAPKS
SAIASRSALFEKIPPQEPSIPKPVKASPKTPEAAAGGGAGKEQEEKKENDKEEGSMSSRIESLTIQEDAKEGVED
EEDGGHHHHHHHH

AtV4IDR-A-HP
Plasmid name: VLN4_B_ pET-24a(+)
Extinction coefficient: 22460 M^{-1} cm^{-1}
MWSHPQFEKGSGGASWSHPQFEKGGGHHHHHHHHHENLYFQGGKEQEEKKENDKEEGSMSSRIESLTIQE
DAKEGVEDEEDLPAHPYDRLKTTSTDPVSDIDVTRREAYLSSEEFKEKFGMTKEAFYKLPKWKQNKFKMAVQ
LF
AtV4IDR-B
Plasmid name: VLN4_A_ pET-24a(+)
Extinction coefficient: 15470 M^{-1} cm^{-1}
MWSHPQFEKGSGGASWSHPQFEKGGGENLYFQSSKSAMHGNSFQRKLKIVKNGGTPVADKPKRRTPASY
GGRASVPDKSQQRSRSMSFSPDRVRVRGRSPAFNALAATFESQNARNLSTPPPVVRKLYPRSVTPDSSKFAP
APKSSAIASRSALFEKIPPQEPSIPKPVKASPKTPESPLPETGGHHHHHHHH

WT-SrtA

Extinction coefficient: 14440 M^{-1} cm^{-1}

MRGSSHHHHHHSSGLVPRGSHMQAKPQIPKDKSKVAGYIEIPDADIKEPVYPGPATPEQLNRGVSFAEENES
LDDQNISIAGHTFIDRPNYQFTNLKAAKKGSMVYFKVGNETRKYKMTSIRDVKPTDVGVLDEQKGKDKQLTLI
TCDDYNEKTGVWEKRKIFVATEVK

7m-SrtA

Extinction coefficient: 14440 M^{-1} cm^{-1}

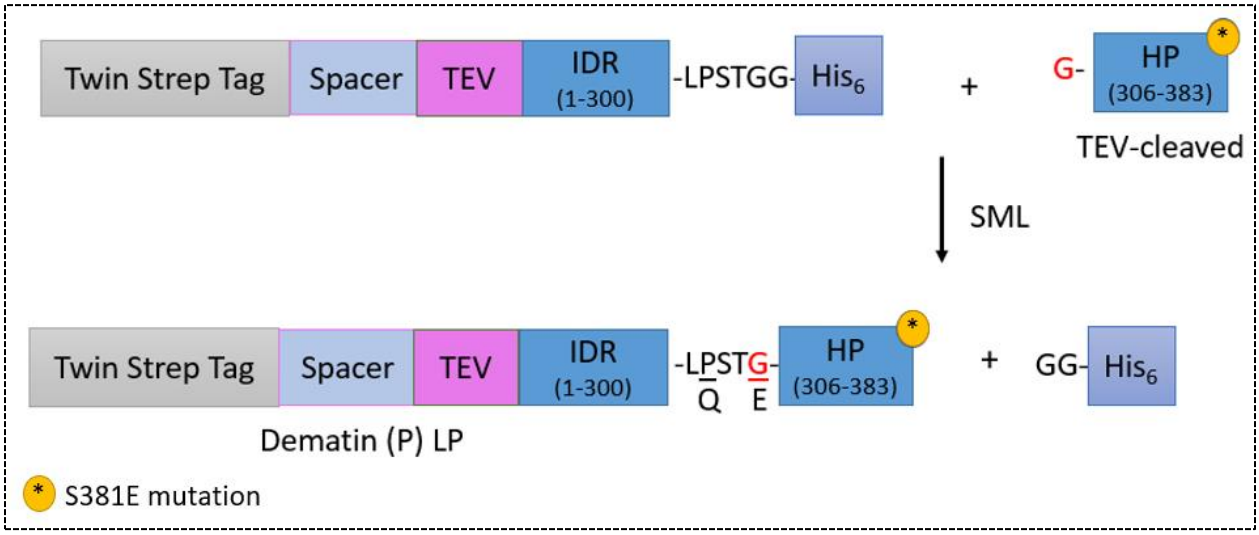
MQAKPQIPKDKSKVAGYIEIPDADIKEPVYPGPATREQLNRGVSFAKENQSLDDQNISIAGHTFI
DRPNYQFTNLKAAKKGSMVYFKVGNETRKYKMTSIRNVKPTAVEVLDEQKGKDKQLTLITCDD
YNEETGVWETRKIFVATEVKLEHHHHHH

TEV-Protease

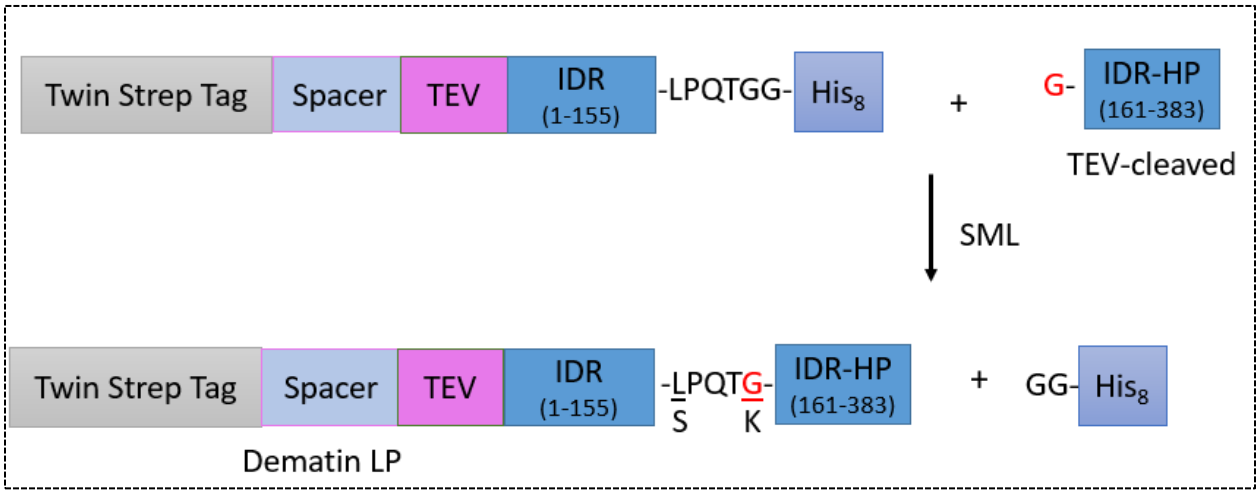
Extinction coefficient: 40005 M^{-1} cm^{-1}

GHIVWPDYANILKEVFGGARMACVTSAHMAGANGSILKKAEETSRATMHKPVIFGEDYVTEADL
PYTPLHLEVNAEMERMYYLGRRALTHGKRRKVSVNNKRNRRRKVAKTYVGRDSIVEKIVVPHT
ERKVDTTTAVKDTCNEVSTQLVHNSMPKRKKQKNFLPATSLSNVYAQTWSIVRKRHMQVEIISK
KSVRAKVKRFEGSVQLFASVRHMYGERKRVDLRIDNWQQKTLLDLAKRFKNERVDQSKLTFG
SSGLVLRQGSYAPAHWYRHGMFIVRGRSDGMLVDARAKVTFAVCYSMTHYHHHHHH

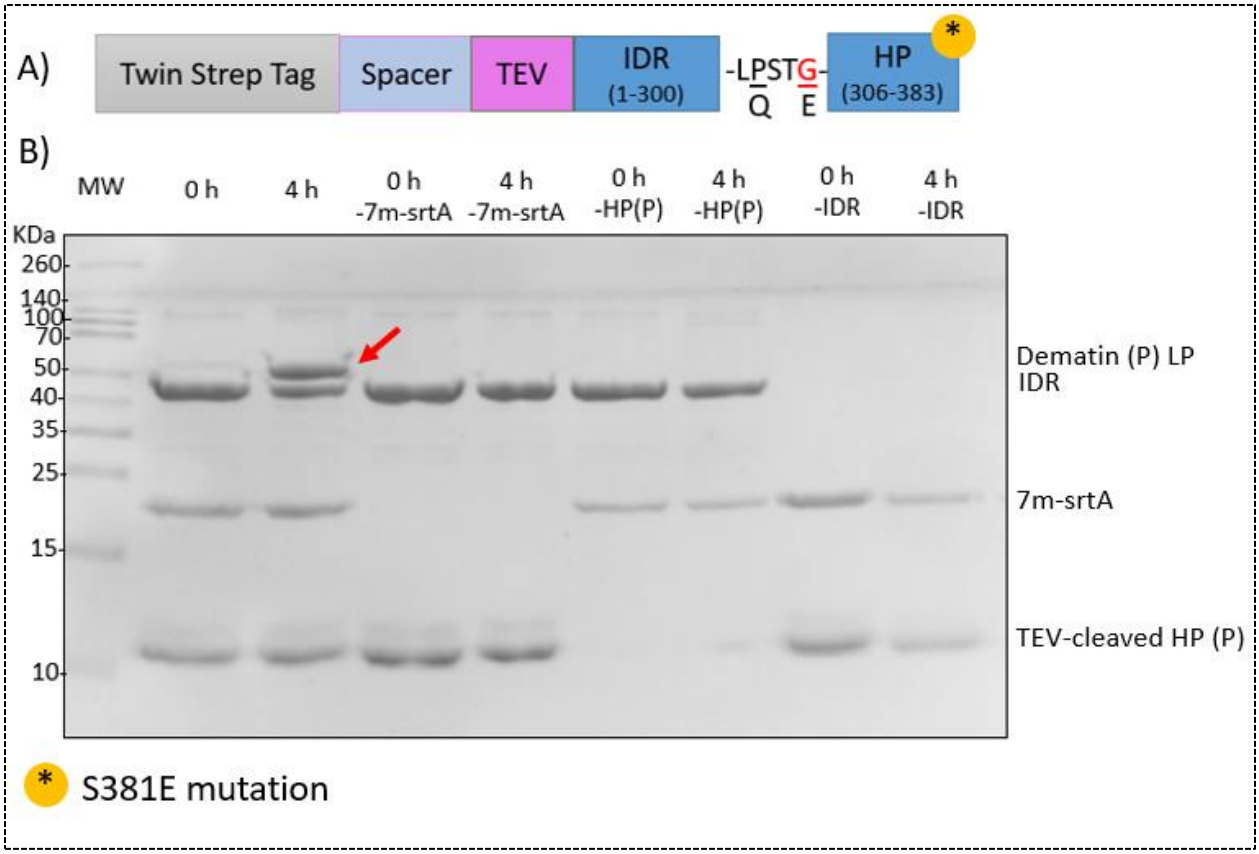
Section 2: SML Reaction Schemes.



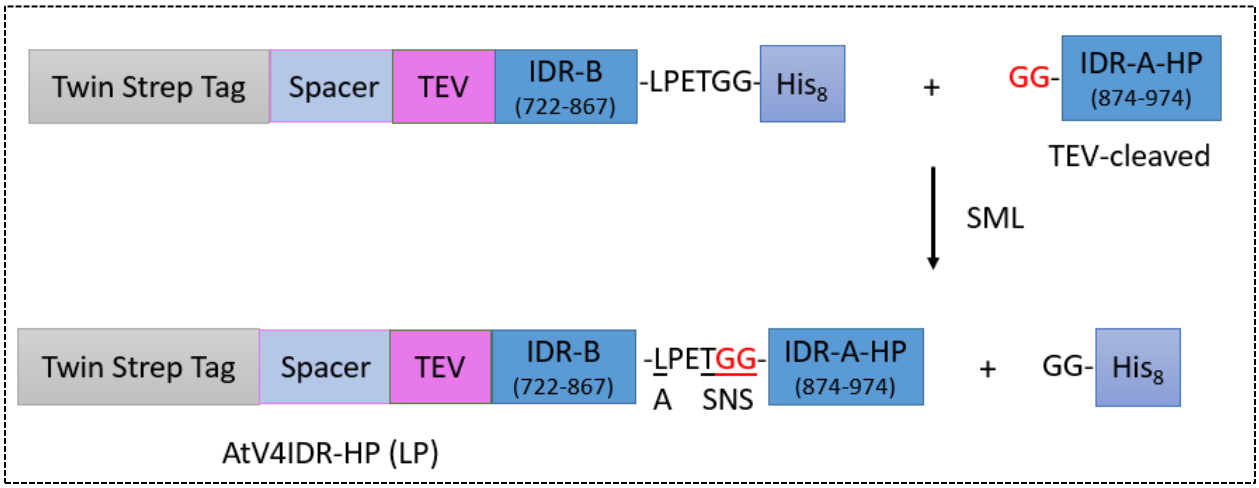
Appendix Figure 2.1. Aim II SML Scheme 1. SML to make full-length dematin (P) LP utilizing LPXTG site at positions 301-305.



Appendix Figure 2.2. Aim II SML Scheme 2. SML to make full-length dematin LP utilizing LPXTG site at positions 156-160. The reaction has yet to be run.



Appendix Figure 2.3. Aim II SML Control Reaction. The LPXTG site is at positions 301-305. SML was run for 4 hours under the same conditions described in the Materials and Methods section 2.8b. Control reactions are missing one different SML component each, indicated above corresponding lanes. Image also found in Erin Rosenkranz's thesis[39] (OCLC# 1277514139).



Appendix Figure 2.4. Aim III SML Scheme. Mutated residues within predicted PEST motif are underlined. Wild-type amino acids are shown directly underneath the substitution site.