Summer 2017

Structural and Functional Characterization of the F-actin Binding Headpiece Domain of Plant Villin 4

Heather Miears
Western Washington University, miearsh@wwu.edu

Structural and Functional Characterization of the F-actin Binding Headpiece Domain of Plant Villin 4

By

Heather Miears

Accepted in Partial Completion
Of the Requirements for the Degree
Master of Science

Kathleen L. Kitto, Dean of the Graduate School

ADVISORY COMMITTEE

Chair, Dr. Serge Smirnov

Dr. Spencer Anthony-Cahill

Dr. P. Clint Spiegel

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.

Heather Miears

July 26th, 2017

Structural and Functional Characterization of the F-actin Binding Headpiece Domain of Plant Villin 4

A Thesis

Presented to

The Faculty of

Western Washington University

In Partial Fulfillment

Of the Requirements for the Degree

Master of Science

By

Heather Miears

July 2017

Abstract

The villin family of proteins are filamentous actin (F-actin) regulators, which play key roles in cytoskeleton organization regulation in eukaryotes. Recently discovered villin 4, from *A. thaliana* has been shown to bundle F-actin in root hairs. However, previous work found that mutations to villin 4 result in shorter and sparser root hairs which lead to a weaker capacity for water absorption by *A. thaliana*. It is previously known that many members of the vertebrate villin protein family contain a highly conserved F-actin binding site within the C-terminal headpiece domain. In order to investigate the headpiece of villin 4, we predicted where the folded domain began based on sequence homology. Interestingly, in plant villins, a long linker (>100 residues) region exists between the folded core domain and the C-terminal headpiece. While it is known that vertebrate villin contains a 40-residue linker, it shows no homology to that of the plant linkers and, therefore, the linker-headpiece boundary in villin 4 was unclear. To define this boundary, we designed two constructs: the first, atVHP76, we believed would contain some residues of the linker as well as the headpiece; the second construct, atVHP60, was predicted to contain just the folded headpiece domain. Furthermore, we proposed to investigate whether this portion of the linker alters the F-actin binding affinity of this domain. Presented here is the structural and functional study of the villin 4 headpiece domain (V4HP63). Our solution NMR analysis demonstrates that V4HP63 possesses a canonical villin headpiece fold, with the exception of a merged variable-loop (V-loop) and a newly discovered high-mobility loop (HM-loop). Through an actin pull-down assay, we showed that both atVHP76 and atVHP60 binds specifically and with high affinity to F-actin (K_d of 3.0 ± 0.7 µM and of 2.0 ± 1.0 µM, respectively), indicating the N-terminal portion of the linker does not alter the affinity. Comparison of the electrostatic surface potential maps of V4HP63 and high-affinity vertebrate F-actin binding headpiece domains, we propose a modification to the currently accepted rules for predicting high-affinity binding of villin headpieces. This includes a reduction of the previously described residues in the “charged crown” to K65,

R37, and the C-terminal carboxylate. This work demonstrates that the plant villin 4 headpiece is evolutionarily related to the vertebrate homologs.

Acknowledgments

I would like to express my sincere gratitude to Dr. Serge Smirnov for giving me the opportunity to become a member of his lab in January of 2013. He was always willing to assist me with lab and computer work, even if that meant working into the night. But most of all, I greatly appreciated is sincere encouragement throughout my undergraduate and graduate work. Working in this lab has been the highlight of my time at Western Washington University.

I would also like to thank my committee members, Dr. Clint Spiegel and Dr. Spencer Anthony-Cahill for all their guidance on my thesis work. This project would not have been completed without the generous use of shared lab space and instruments that both Dr. Spiegel and Dr. Anthony-Cahill provided. I would also like to thank Dr. Antos for his generosity in allowing us to use his HPLC and assisting us with analysis.

In addition, I would like to thank the past and current lab members who helped support this work. Specifically, I would like to thank the fellow graduate students in the lab, Joanna Hoppins and David Gruber, for not only their friendship, but for all their support along this project. Special thanks also goes out to Johann Sigurjonsson who not only supported me through this work but was always there for help and encouragement when it was needed most.

Table of Contents

List of Figures and Tables

Chapter I: Introduction

Actin Cytoskeleton

F-actin generation

Eukaryotic cells contain a cytoskeleton which comprises three filamentous proteins that provide the architecture of cells: microfilaments, intermediate filaments, and microtubules[1]. Specifically, microfilaments are formed from a cytoplasmic pool of proteins, which allow the filaments to perform versatile and essential roles such as cell motility, cytokinesis and cell shape regulation[1-3]. These filaments are composed of a protein called actin, which exists in two principle forms: globular monomer (G-actin) and polarized filaments (F-actin). F-actin contains two dynamic ends, a barbed (+) and a pointed (-), which grow and depolymerize at differing rates (Figure 1.1). The formation of F-actin is followed by the binding of ATP by G-actin, which in turn non-covalently binds primarily to the barbed end. The filaments then elongate at a rate that is proportional to the concentration of G-actin (µM) in the cells, which varies by cell type[4]. Once an equilibrium is reached between the concentration of free G-actin and F-actin, elongation will continue until either a steady state of polymerization and depolymerizaton occurs or the barbed end is capped. This is done primarily through end-capping proteins (K_d ~0.1nM) which are regulated by cellular signals[4]. New F-actin generation can then be driven by the formation of *de novo* barbed ends through nucleation, severing, or un-capping.

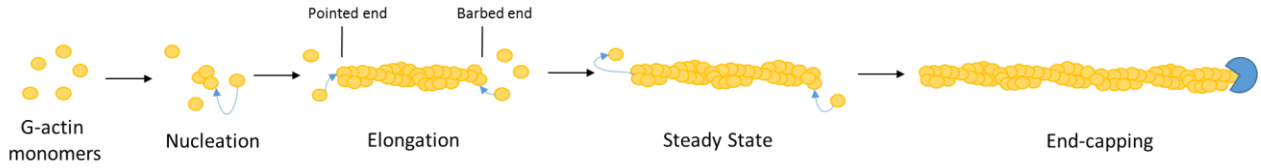


Figure 1.1. Free G-actin monomers in solution once bound to ATP, form a nucleus which initiates elongation. Elongation occurs at both ends of the filament, but at an increased rate at the barbed end. Steady state is reached once the concentration of free G-actin and F-actin reaches equilibrium and the subunits incorporated into the filament hydrolyze ATP to reach a more stable ADP- F-actin state. This steady state process can then be capped by an end-capping protein.

Actin binding proteins

Due to the dynamic nature of F-actin, all organization into higher-order networks is regulated by actin binding proteins (ABPs)[5]. ABPs are classified into seven functional categories: (1) G-actin monomer binding, (2) filament-depolymerizing, (3) filament end-capping, (4) filament severing, (5) filament crosslinking, (6) filament stabilizers, (7) and motor proteins which use F-actin as a track[5]. Many of these ABPs perform in more than one of the groups.

When filament crosslinking is performed by ABPs and there are two main organizations of filaments: bundles and crosslinking networks (Figure 1.2). Bundles are parallel actin filaments that are held together by crosslinking ABPs that either have two actin binding sites within the monomer, or through formation of a dimer which has only one actin binding site on each monomer. F-actin bundles are usually found in cells that utilize polarized filaments such as in microvilli[7]. Actin networks are filaments that branch and cross over each other to have a gel like property in the cytosol. This provides cells with the capability of altering the cell membrane for a variety of essential cellular processes such as motility and endocytosis[4]. APBs that arrange filament networks have long distances between their actin binding sites, which in turn allows filaments to cross over each other.

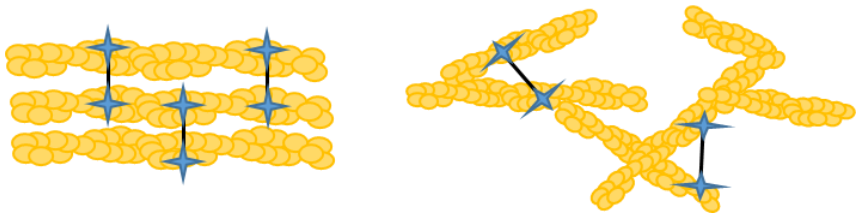


Figure 1.2. Depiction of F-actin bundles (left) or networks (right).

Villin family of proteins and homologs

Due to the high complexity of domains and proteins that are discussed in this work, the following section will give a brief overview of villin members that are related to the work in this thesis: gelsolin, villin, supervillin, dematin and protovillin.

Regulators of F-actin

Villin homologs are a family of eukaryotic ABPs that are found in various cell/tissue types and have several functions for actin regulation such as binding, bundling, severing, capping actin filaments, sequestering actin monomers, and nucleating actin filaments[5-7]. Most homologs' actin re-modelling functions are regulated by biological signal such as Ca^{2+} levels, tyrosine phosphorylation, and phosphatidylinositol-4,5-bisphosphate (PIP2)[6, 8, 9].

Structurally, a majority of the villin family shares six homologous domains (G1-G6) referred to as the gelsolin core. Each domain comprises 97-188 residues that fold into a 5- or 6-stranded β-sheet that is sandwiched between two α-helices. ABPs within the villin family contain a variety of multiples of G domains, or are depleted of the domain entirely (Figure 1.3). Initially, only three or six domain repeats were found within the family so it was suggested that they evolved from a single gelsolin domain that was first triplicated, to form the three G-domain variants, and then the trimer was duplicated to form the six domain repeat seen in a majority of members[7]. However, as more members were discovered, a variety of novel architectures has been identified, such as two, four, five, or no gelsolin domains, suggesting the evolution of this homologous domain is much more complex than it was previously thought.

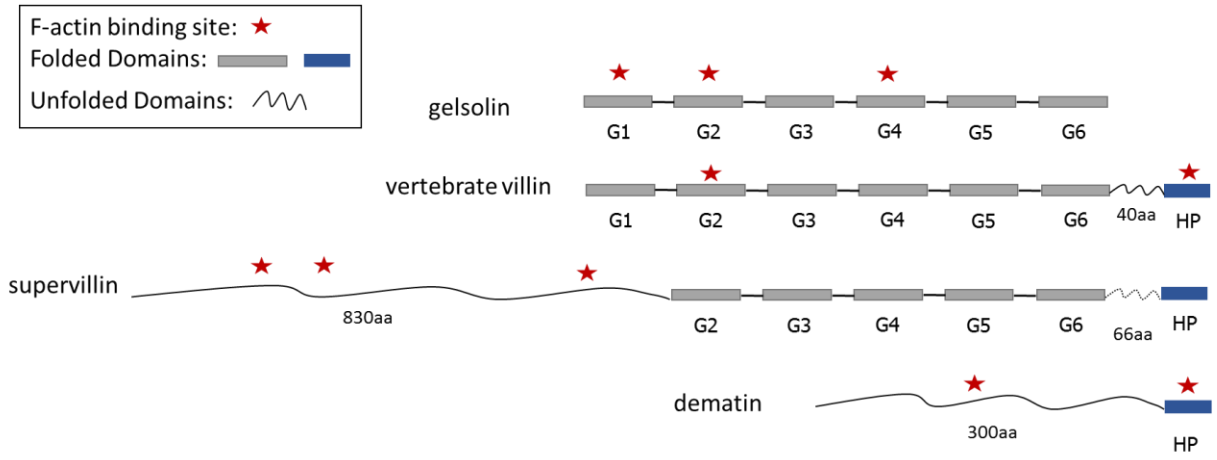


Figure 1.3. Schematic of the domain arrangement of villin family members. There is a variety of F-actin binding sites on this family, including those that are in intrinsically disordered regions.

A unique feature to some villin homologs is the headpiece domain. This domain is strictly seen on villin family members and is unrelated to any other actin binding domain in the GenBank database[10]. It is a small, ~67 residue, α-helical domain that is seen exclusively at the C-terminus of villin homologs. This domain has been found to commonly contain actin-binding properties as well as the unique ability to retain full actin binding activity when expressed alone in *vitro*[10, 11]. However, some members of the villin family contain a headpiece domain that does not bind actin. There is little known about the role of the headpiece domain in those proteins. The linker that attaches the headpiece to the gelsolin core varies in length, charge, and secondary structure among villins and is known to be unstructured in villin[12]. This linker varies in length, charge, and secondary structure (Figure 1.3). There is also evidence of a cryptic actin binding site within this unstructured linker in villin[12].

Gelsolin

Gelsolin is a multifunctional actin regulator in the villin family. It was first identified in 1979 as an agent that could transform macrophage extracts from a viscous gel-like solutions to a more traditional fluid[7]. There are two well studied isoforms in mammals, cytoplasmic and plasma gelsolin. Both of these isoforms are capable of binding, severing, capping and sequestering actin monomers[7]. Cytoplasmic gelsolin remains in an inactive state until calcium levels rise (from 0.1μM to 10 μM) which then activates it to sever actin filaments. This can either expose a pool of shortened oligomers with newly exposed barbed ends, which then can further elongate and heighten cell movement, or completely sever the oligomers to depolymerize fully into the monomeric G-actin pool[7]. In plasma, where calcium levels are consistently high, gelsolin further severs any fragments of actin filaments that may have been released into the blood plasma from cell rupture[7].

Gelsolin contains the six folded core domains (G1-G6) but lacks a C-terminal headpiece. The major actin binding sites are within the G1, G2 and G4 domain (Figure 1.3). There is also a proposed interaction between gelsolin and actin in the linker connecting G3-G4 as this region displays high homology to other ABPs[7]. The G1 and G4 domains contain filament capping regions which compete for the same site on actin. This competitive binding is the proposed model of actin severing by gelsolin as it is a consequence of a pinching movement completed by the simultaneous binding[7].

Villin

Vertebrate villin is exclusively found in epithelial cells of the small intestine and kidney[8]. In the intestinal tract, these cells have small finger-like extensions, called microvilli, that protrude into the gut to provide a high surface area for efficient water and nutrient uptake[13]. Through calcium ion signaling, villin along with ezrin, fimbrin and other APBs help regulate the morphology of microvilli by either bundling or severing F-actin[14, 15]. Under normal cell conditions the concentration of calcium is low (<10 μM) which allows villin to bundle actin filaments and support microvilli[7]. Conversely, in stressed or starving conditions

these calcium levels rise (~100-200 µM) which switches villins role to sever the filaments, causing destabilization of microvilli[7] (Figure 1.4).

Vertebrate villin contains the six domain gelsolin-like core followed by a 40 residue acidic unstructured linker that is attached to the C-terminal headpiece domain[12]. There are at least two F-actin binding sites, one on the G2 core domain and one on the headpiece. Another proposed binding site in the linker between G6 and headpiece domains has not been fully characterized[12]. The mechanism of filament bundling by villin is still under discussion but is suggested to be completed by a single villin spanning parallel filaments via the core and headpiece domains or through the dimerization of villin with each headpiece binding to a filament[14-17].

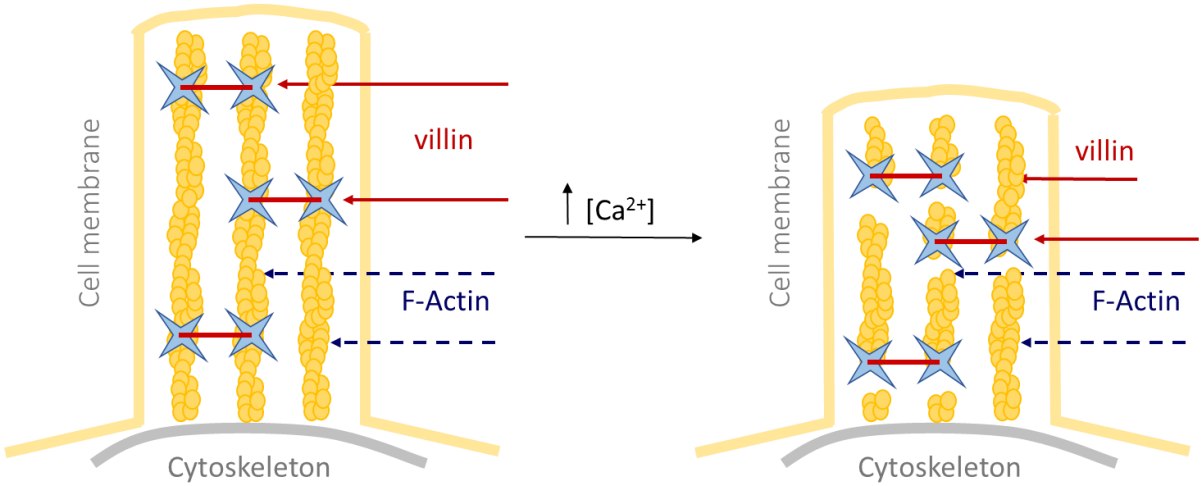


Figure 1.4. Villin bundles F-actin in sub-µM concentrations of Ca^{2+} in microvilli which allows them to protrude into the gut. Once [Ca^{2+}] rises, villin then severs F-actin into short filaments which in turn destabilizes the structure of the microvilli.

Supervillin

Supervillin is a membrane-associated protein that is expressed throughout the human body. It binds to principle components of the cytoskeleton, such as F-actin and myosin II, but it also interacts to ~70 other signaling and cytoskeletal proteins[18-20]. The current well-known functions of supervillin are that it is

required for proper cytokinesis by binding to myosin II and it promotes cell survival by suppressing p53[21, 22]. A unique feature of supervillin is the 830 residue unstructured N-terminus that contains the sites of many interactions including binding to F-actin and mysosin II[23]. This large, intrinsically disordered region (IDR), is what makes supervillin extremely difficult to study *in vivo*. In fact, there is little known about molecular mechanisms by which IDRs interact with their binding partners. It is estimated that ~12% of eukaryotic proteins are entirely unfolded and ~43% of eukaryotic proteins contain an IDR of more 40 residues[24].

The N-terminal IDR is followed by a truncated gelsolin core (Figure 1.3). Then, following the core domain is a 60 residue linker and the C-terminal headpiece domain. Most other G6-headpiece linkers in the villin family are predicted or known to be unstructured, however, the supervillin linker may contain a folded region due to the high content of aromatic residues and prolines. In addition, despite 50% sequence identity to that of villin, the supervillin headpiece domain does but does not retain F-actin binding capability[25]. Currently, there is no understanding of the role of this headpiece domain.

Dematin

Dematin is a cytoskeletal associated protein in erythrocytes and plays an important role in maintaining the shape of red blood cells[26]. In humans, erythrocytes have a life span of ~120 days and must be able to withstand extensive cell membrane deformations as they go through the circulatory track[27]. Thus, these cells require extensively controlled cytoskeleton regulation. The membrane skeleton is composed of an $\alpha_2\beta_2$-spectrin tetramer scaffold that is connected to other APBs and F-actin. At this junction, dematin exists as a trimer, composed of one of the 2 isoforms (48 or 52 kDa), and facilitates the interaction between spectrin and the actin filaments through direct binding[27]. This interaction can be disrupted by phosphorylation of dematin by protein kinase A (PKA), resulting in reduced membrane stability[28]. Dematin has also been shown to bind and bundle F-actin, although mature erythrocytes to not contain actin bundles so a physiological function of this is poorly understood[27, 29].

Dematin has a fully disordered, actin-binding, N-terminus, reminiscent to supervillin, but does not possess a gelsolin core The N-terminal IDR is connected to an actin-binding C-terminal headpiece domain that is homologous to the headpiece found in villin[30]. Currently, this headpiece is the only one known that is phosphorylated *in vivo.* Phosphorylation by PKA at serine 74 near the C-terminus results in the headpiece domain binding to the N-terminal IDR. Dephosphorylated dematin can utilize both its binding domains to bundle filaments and, upon phosphorylation, the headpiece no longer functions as a binding domain[31].

Protovillin

Protovillin is thought to be the "oldest" evolutionary member of the villin family. It was first found in *Dictyostelium discoideum* amoebae and its only characterized function is the inhibition of F-actin elongation through capping[32]. Protovillin contains the full gelsolin core as well as the C-terminal headpiece. Currently, there is no experimentally determined structural data on any domain of this protein, however, a model of the headpiece was completed by threading the sequence onto that of villins headpiece and computationally minimizing the structure[33].

Table 1. Selected villin family members. Organism, molecular weight, structural domains, actin binding domains, isoelectric points of domains and size of linkers are listed. Residues in the linker of supervillin are protovillin are predictions.

Name	Organism	MW (kDa)	Core Domain	Headpiece	Known Actin Binding Domains	pI of HP	pI of Linker	Residues in Linker
villin	*Gallus gallus*	90	yes	yes	HP and Core	7.18	4.4	40
supervillin	*Homo sapiens*	200	yes	yes	Unstructured N-terminus	4.92	4.8	66
dematin	*Homo sapiens*	42	No IDR?	yes	HP and Unstructured Core	9.9	6.7	330
protovillin	*Dictyostelium discoideum*	100	yes	yes	HP and ???	8.92	10.7	85

Plant villin homologs

In recent years there has been significant progress in both sequencing and characterizing plant genomes. Through these investigations, it was discovered that many plant species contain villin homologs, that reside in a most tissues. An overview is presented here on both actin regulation by villin homologs in *Arabidopsis thaliana*, as well as an introduction to the investigated protein in this thesis, villin 4.

Cytoskeleton regulators

Unlike vertebrate villin family members, which are restricted to absorbent tissue, villin isoforms in plants are found in a variety of tissue types such as pollen, leaves, stems, siliques and root hairs[34]. So far, many villin analogs appear to be vertebrate-specific as they are absent in the studied plants, e.g. dematin, gelsolin and supervillin. This leaves plant villins as the major regulators of actin dynamics and therefore they play a more general and potentially diverse role in plants than they do in vertebrates.

The overall primary structures of plant villins resemble that of vertebrate analogs: they contain the homologous gelsolin core, a C-terminal headpiece domain, and a linker connecting the two. This linker is predicted to be disordered and is the longest linker, at 100-200 residues, seen in any villin homolog. Uniquely, this linker demonstrates charge partitioning: a basic, isoform-specific and variable-length N-terminus (60-150 residues) and an acidic C-terminus (~40-residue). There is no significant sequence homology between the ~40 residues vertebrate villin linker and the ~40 residue C-terminal end of plant villin linkers except that they both are acidic regions. Another novel feature found in plant villin linkers is a PEST motif positioned at the boundary between the basic and acidic linker regions. The PEST sequence is rich in proline (P), glutamic acid (E), serine (S) and threonine (T) and is hypothesized to be a signal for degradation[35].

Villin 4

Arabidopsis thaliana (A. thaliana), the model organism of the plant world, was the second plant discovered to contain a villin homolog, after lily (*Lilium longforum*)[36]. There are five isoforms of villin found in *A. thaliana*, and are named villin 1-5. Villin 1 and 5 are found primarily in pollen and facilitate pollen tube growth through the calcium dependent organization of F-actin [37, 38]. Villin 2 and 3 are also involved in F-actin regulation and are located in sclerenchyma, cells located in the leaves and stems of plants which provide the main structural support to the plant[39, 40]. Lastly, villin 4 is the only isoform known to be primarily located in the root hairs. Similar to microvilli, root hairs are extended cellular structures that are responsible for absorbing nutrients and water for the plant. Villin 4 is found in many plants including tobacco (*Nicotiana tabacum*), rice (*Oryza sativa*), red wild einkorn which is a grass species related to wheat (*Triticum uratu*), and potato (predicted) (*Solanum tuberosum*)[41].

It was predicted, through sequence analysis, that villin 4 contains both F-actin binding and bundling capability. Through a high-speed co-sedimentation assay with full length *A. thaliana* villin 4, it was shown that villin 4 could bind F-actin in both low (0.5µM) and high (200 µM) Ca^{2+} concentrations with an average K_d of 0.24 ± 0.05µM[42] . Villin 4 also showed F-actin bundling activity at high and low Ca^{2+} concentrations, and severing or depolymerizing activity in a Ca^{2+} dependent manner[42].

To investigate the function of villin 4 in *A. thaliana*, two T-DNA insertion lines (atVLN4-1 and atVLN4-2), which prevented the expression of full-length villin 4 transcripts, were produced and no marked phenotypic changes were seen in the main root of the plant. However, the root hairs were much sparser and their growth was significantly reduced in comparison to WT[42]. Further analysis, through actin staining, showed fewer actin polymers organized into higher-order cables in both atVLN4-1 and atVLN4-2 lines in comparison to WT[42] All data combined suggests that villin 4 is responsible for the regulation of F-actin in root hairs and is necessary for proper root hair morphology.

Drought resistant crops

Currently, water shortages are responsible for the greatest crop losses around the world and are projected to be a continual problem due to climate changes[43]. Furthermore, with the increasing world population and global warming, plant breeders are unable to keep up with the agricultural demands[44]. Agriculture is one of the largest consumers of water on the planet and as water becomes more scarce, there is an increased need to improve the efficiency and sustainability of water use[45]. Therefore, there is a need for drought-resistant crop engineering through either conventional breeding methods or genetic modification. Conventional breeding relies on identifying genetic variability to drought among different crop varieties, which must have suitable agronomic characteristics. Although this method has shown some success, it is extremely slow and limited by the gene pools already available. Genetic engineering allows us to identify a specific genetic target that confers drought resistance. This method can also be broadly applicable to many plant varieties that may not have naturally occurring drought resistant types. Understanding the mechanisms of water and nutrient uptake by root hairs and the proteins involved in these processes is important to help focus genetic engineering strategies. Since plant villins are the main regulators of the ability for root hairs to extend out into the soil, they are valuable target for further study as well as potential genetic engineering targets.

Comparison of villin homolog headpiece domains

Sequence similarity

The headpiece domains in the villin family are composed of an N- and C-terminal subdomain. The C-terminal subdomain is identified by similarity to the last 35 residues in chicken villin headpiece (VHP) and contains the highest conserved sequence identity between villin homologs (Figure 1.5). Among known actin binding headpieces, VHP and dematin headpiece (DHP) have 45% and 38% sequence identity to *A. thaliana* headpiece villin 4 (atVHP) respectively, with most of this conservation in the C-terminal subdomain. Whereas the non-actin binding headpiece from supervillin (SVHP) has 39% sequence identity

with atVHP. Among the five isoforms of atVHP all headpieces show >45% sequence identity, again with the most discrimination in the N-terminal subdomain.

```
                                  10         20         30         40         50         60         70      76
Vertebrate Villin HP67             P TKLETFPLDV LVNTAAEDLP RGVDPSRKEN HLSDEDFKAV FGMTRSAFAN LPLWKQQNLK KEKGLF
Plant Villin HP76          QEDAKEGVED EEDLPAHPYD RLKTTSTDPV SDIDVTRREA YLSSEEFKEK FGMTKEAFYK LPKWKQNKFK MAVQLF
Plant Villin HP60                    HPYD RLKTTSTDPV SDIDVTRREA YLSSEEFKEK FGMTKEAFYK LPKWKQNKFK MAVQLF
Supervillin HP67                   M LAKLCKTIYP LADLLARPLP EGVDPLKLEI YLTDEDFEFA LDMTRDEYNA LPAWKQVNLK KAKGLF
Dematin HP68                      PG LQIYPYEMLV VTNKGRTKLP PGVDRMRLER HLSAEDFSRV FAMSPEEFGK LALWKRNELK KKASLF
Protovillin HP66                     PTTITTFYPL SVLKQKTNLP NDIDKSCLHL YLSDEEFLST FKMTKEIFQK TPAWKTKQLR VDNGLF
```

Figure 1.5. Sequence comparison of villin homolog headpiece domains. The underlined portion in HP67 is the highly stable and compact C-terminal subdomain named HP35.

It is a general feature that the N-terminal subdomains of villin headpieces contain the least sequence similarity. However, they contain a partially conserved region known as the variable loop (V-loop) sequence which is 5, 8 and 10 residues long for SVHP, VHP and DHP respectively. This loop was first identified through manual alignments of the headpiece domains because computational alignment algorithms failed to recognize it, as there are extensive changes in the number and types of residues here between homologs[25]. This loop begins at residue 16 for VHP and is summarized as: F/Y X L/Y X X L X X ------ L/P X G/D V/I D/N X X R/K X E/D X H/Y, where X can be any residue and "------" represents a loop that may contain any or no residues[10]. The length and composition of the loop have yet to be determined as critical for the headpiece structure but the correct positioning of the conserved N-terminal hydrophobic residues flanking this loop are essential for proper N-terminal subdomain folding[10].

Headpiece structure and sub-domains

As stated previously, headpiece domains contain two subdomains that are distinguished by their differing folds and ability to fold independently of one-another. The C-terminal subdomain on chicken villin headpiece (VHP35) is a small compact 35-residue subdomain that comprises three α-helices and is stabilized by a hydrophobic core. In fact, VHP35 is the smallest known cooperatively folding protein domain and is used as model of protein folding through computational and experimental methods[46-49].

The structure of VHP35 in isolation is highly similar to VHP35 in the full headpiece, only differing at the terminal residues and those that directly interact with the N-terminal subdomain[10, 11].

The N- and C-terminal subdomains are connected via a salt bridge and a contiguous hydrophobic core (Figure 1.6). The 32-residue N-terminal subdomain has little secondary structure with only some headpieces containing a short helix formed by L18-V22[50] . The long aliphatic region on the side chain of K38 forms part of the hydrophobic core that bridges these domains. Only a lysine or leucine is seen here across all headpiece homologs which suggests the presence of a long side chain at this position is a conserved element needed for structural stability. Along with a hydrophobic core, a buried salt bridge, E39-K70, connects these two subdomains and is found in all headpiece homologs. This salt bridge is unique: that the most frequent solvent accessibility of salt bridges is 20-32% and in VHP it is 2.3%. In addition, most bridges are only 4 residues away, while the bridge in VHP and homologs is 31 residues away. Highlighting the unique design of this domain. Although, it should be noted, in DHP the salt bridge is now considered a cluster of charged residues, E39, R66 and K70, which have less than 10% solvent accessibility[51]. Through mutational studies, it was shown that this salt bridge is an essential requirement for proper folding of the N-terminal subdomain and F-actin binding capability[52]. In fact, no mutation was tolerated for proper folding or actin binding on VHP even when the residues were switched (E39-K70 → K39-E70)[52].

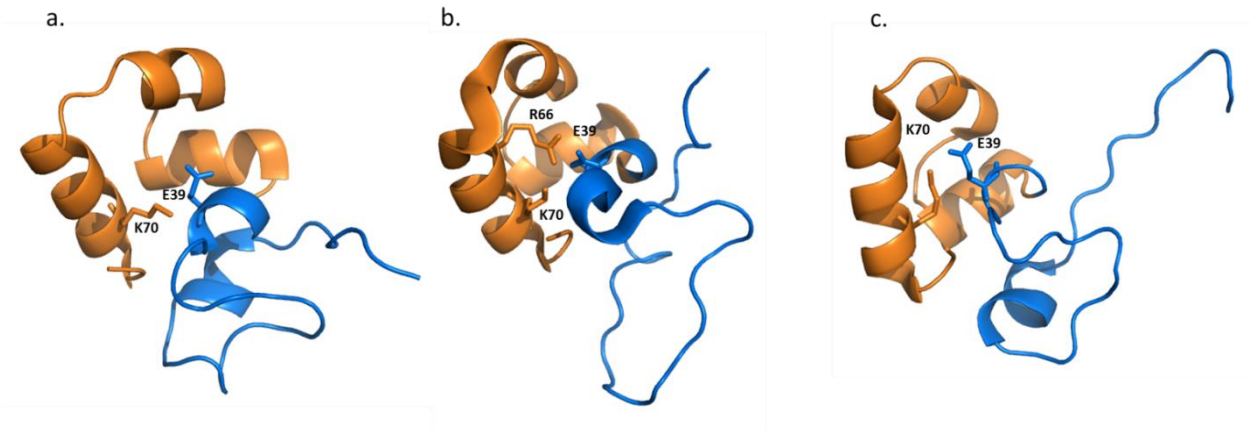


Figure 1.6. NMR solution structures of (a) VHP (PDB ID: 1QQV), (b) DHP (PDB ID:1QZP) and (c) SVHP(PDB ID: 2K6N). The highly structured C-terminal subdomain is shown is orange and the less folded N-terminal subdomain is shown in blue. The essential charged residues responsible for connecting the N- and C-terminal subdomains are labled and shown as sticks.

As stated previously, the V-loop is a general feature of the headpiece domain (Figure 1.7). In VHP, the loop brings the flanking hydrophobic residues F16, L18, L21, L29 and P30 in close proximity to the hydrophobic core. ^{15}N-relaxation analysis was utilized to probe the mobility of the V-loop and was shown that the 5 residue V-loop in SVHP showed no mobility in this region, while an increased mobility is observed for both the VHP and DHP V-loops. As SVHP still contains a properly folded N-terminus it was determined that five residues is the absolute shortest sequence that is tolerated in the V-loop for proper N-terminal subdomain folding[25].

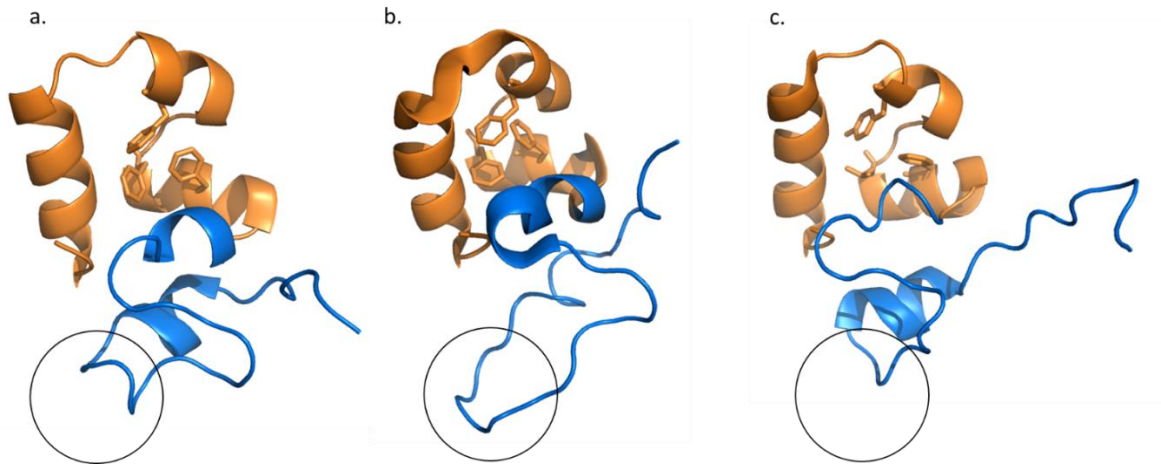


Figure 1.7. NMR solution structures of (a) VHP(PDB ID: 1QQV), (b) DHP(PDB ID: 1QZP) and (c) SVHP(PDB ID: 2K6N). The highly conserved phenylalines are shown as well as the variable loops (circled).

Actin binding by headpiece domains

Through mutagenesis studies and computational modeling on the VHP domain it was determined that all essential F-actin binding residues map to a single face and have three criteria for F-actin[11]. These are (1) a hydrophobic cap, primarily formed from L63 and W64; (2) alternating charged "crown" below the cap composed of K65, E72, K71, R37, and the C-terminal carboxylate group on F76; and (3) a positive patch below the crown composed of K38 and K73[11] (Figure 1.8).

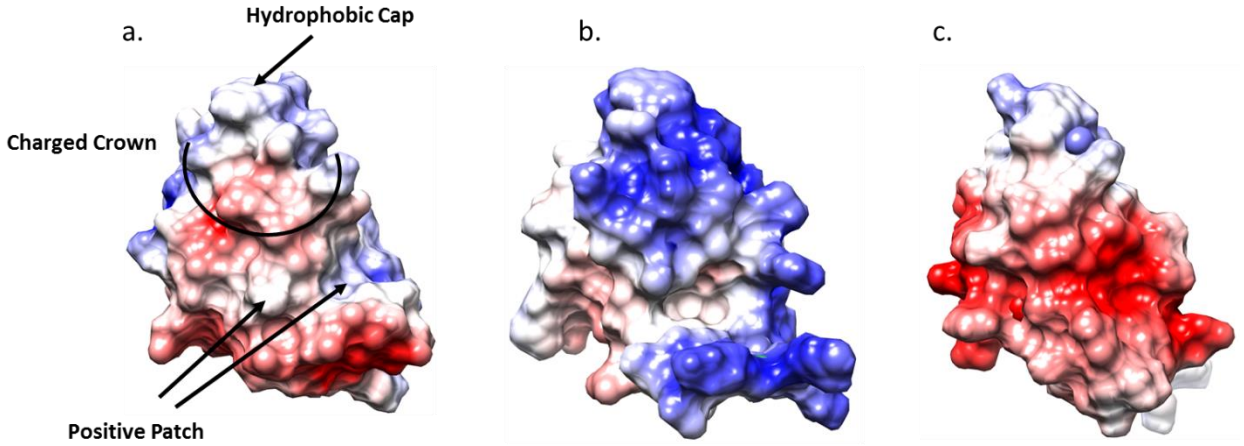


Figure 1.8. Surface charge potentials on (a) VHP(PDB ID: 1QQV), (b) DHP (PDB ID:1QZP), and (c) SVHP SVHP(PDB ID: 2K6N). Surface charge potentials were produced by electrostatic calculations from Adaptive Poisson-Boltzmann Solver (APBS)[53] and are colored by electrostatic potential from red (-6) through white (0) to blue (6).

Through actin sedimentation assays it was determined that the affinity for F-actin can be grouped into three categories: high (3-10 μM), low (30-50 μM), and non-specific binding (>100 μM)[33]. The highest affinity headpiece was shown to be VHP (3.9 ± 0.8 μM). Both un-phosphorylated DHP (3.3 ± 1.4 μM) and phosphorylated DHP (DHP-P; 10.3 ± 2.4 μM) were shown to have high affinity binding.

The low affinity headpiece PV (54.6 ± 15.2 μM) showed modest increased affinity when it was modified by introducing the "charged crown" sequence (PV_{KKEK}; 39.9 ± 34.4 μM). A K_d value was not determined for SVHP and VD as they are estimated to be >100 μM, which indicates all HP pulled down through the sedimentation assay is due to non-specific binding or trapping by the F-actin filaments.

To determine whether the N-terminal subdomain was required for high affinity binding, truncated variants HP67, HP45, and HP38 were tested. HP67 does not contain the N-terminal 9 residues in VHP which are shown to be unfolded and was found to have high affinity binding (2.5 ± 0.5 μM). HP45 has loss of structure for the N-terminal 10 residues, which include residue E39 that forms a salt-bridge to the C-terminal subdomain residue K70 and was shown to have low affinity (37.8 ± 13.6 μM). Lastly, HP36, which

contains solely the C-terminal subdomain which lacks the positive patch on the F-actin binding surface, K38, was not found to bind actin.

Looking at the truncated versions of VHP shows the importance of retaining the N-terminal subdomain residues. The HP45 construct, which lacks formation of the salt-bridge, indicates that this motif is essential for high affinity binding, and further truncation to remove the entire N-terminal subdomain (HP35) abolishes any affinity.

Both DHP and SVHP have slight alterations of the required residues for actin binding but DHP retains the proper electrostatic charges. SVHP, has shown an attenuated positive patch and a smaller hydrophobic cap due to a K38L and L63A mutation respectively, which leads to the inability to bind F-actin. Furthermore, it has been shown that recovery of the positive patch in SVHP, through a single point mutation of L38K, can recover specific of F-actin binding[25].

Although the hydrophobic tryptophan is conserved in atVHP, the residues that map to the charged crown and positive patch region appear to be replaced with aliphatic amino acids of varying lengths. Thus, the investigation of atVHP will help give further insight onto the important residues for F-actin binding because of the altered surface charge distribution of the headpiece.

Core-to-HP linker

In vertebrate villin, the span between domain 6 and the headpiece is the longest inter-domain linker (40 residues). Previous studies have shown this linker to be acidic (pI 4.4) and unstructured[12]. Plant villin linkers are uniquely long (100-200 residues), and in all *A.thaliana* isoforms, have two distinct portions, an N-terminal basic region (87-140residues, pI >10) and a shorter C-terminal acidic region (41-56 residues, pI ~4). One exception to this charge partitioning is in villin 1, where the linker is more like the IDR in dematin. The plant villin linkers are predicted to be unstructured because all isoforms have a low content of

aromatics for the entire linker, and contain several regions of high proline content as well as triplicate proline stretches.

Research Aims

The goal of the presented research was to characterize, for the first time, the structure and function of the headpiece domain from villin 4, a villin homolog recently discovered in plants root hairs.

Specifically, our first aim was to characterize the tertiary structure of the folded headpiece domain, and therefore, identify where the linker ends and the folded headpiece begins. We employed solution NMR for structure determination.

The second aim was to characterize the F-actin binding capability of the headpiece domain. We utilized an F-actin pull-down assay to determine both the affinity of the headpiece domain with a short linker element (atVHP76) and without any linker (atVHP60).

These two aims helped answer the question whether the plant headpiece domains belong to any headpiece classes found in vertebrates or form a group of their own in terms of their structure and F-actin binding function.

Chapter II: NMR Experimental Design

Nuclear Magnetic Resonance Spectroscopy

Solution nuclear magnetic resonance spectroscopy (NMR) is a powerful technique used to characterize three-dimensional structures and conformational dynamics of biological molecules. The method is built on the foundation that different nuclei in biological molecules (e.g. ^{1}H, ^{13}C, ^{15}N) resonate at different frequencies in a magnetic field, referred to as a chemical shift measured in units of parts-per-million (ppm), which are dependent on the environment around the nuclei.

There are a few reasons why solution NMR was the best option for structure determination for this work. The first being that NMR allows biological molecules to reside in physiologically-relevant solution environments. As proteins are not fixed in solution, molecular dynamic experiments can be utilized to understand the motions of individual atoms or residues in the protein. Understanding the dynamics of the villin 4 fragments was important to determine the boundaries between the linker and the folded headpiece (Specific Aim 2). Understanding the junction between a folded domain and a potentially unstructured region is not currently possible utilizing X-ray crystallography because it is difficult to crystallize proteins with large flexible regions. Secondly, villin headpieces have been extensively characterized through NMR, as they are small (<15 kDa) domains that have some flexible regions[46-49]. Currently, there are many villin-family headpiece domains, in both native and mutant forms, that have been previously solved by both NMR and X-ray crystallography. Here, we solved, the first villin headpiece structure from a non-vertebrate.

Experimental design

To solve the solution structure of a protein, one can utilize both homonuclear (^{1}H) and heteronuclear (^{1}H-^{15}N or ^{1}H-^{13}C) strategies, depending on the size of the protein. In general, using exclusively homonuclear experiments is possible to fully assign proteins up to 12kDa, if there is little overlap in the spectral lines

and if the proteins stay well folded in solution[54]. There are both one dimensional (1D) ^{1}H experiments, as well as, two dimensional (2D) ^{1}H NOESY, TOCSY, and COSY experiments, which can further separate overlapping peaks in the 1D spectrum. Proteins larger >10kDa will best be approached from heteronuclear NMR experiments which requires an additional NMR active nucleus such as ^{15}N or ^{13}C. As the number of hydrogen atoms scales approximately linearly with the molecular mass, peaks begin to overlap in the spectra. The rotational correlation time also increases with mass, which in turn, causes line or peak broadening in the data[54]. This exacerbates the ambiguity of the spectra in homonuclear experiments of larger proteins. Having, both ^{15}N and ^{13}C labels, the spectral resolution can be further increased by going from homonuclear 2D experiments to heteronuclear 3D. An introduction to the 3D experiments used in this work is presented later.

The atVHP76 construct is ~9kDa, so by literature standards it may be possible to fully assign through only homonuclear experiments. However, extensive overlap in both the amide proton and aliphatic proton regions is observed in the 1D spectrum of this protein (Figure 2.1). Thus, we chose to use heteronuclear experiments for full assignment of this protein. A combination of both through-bond and through-space experiments was used in this work to get complete assignment of backbone and sidechain residues, as well as distance restraints. Due to the high complexity of the 3D NMR experiments, here, a brief overview is given on the experiments we used to determine the solution structure of atVHP.

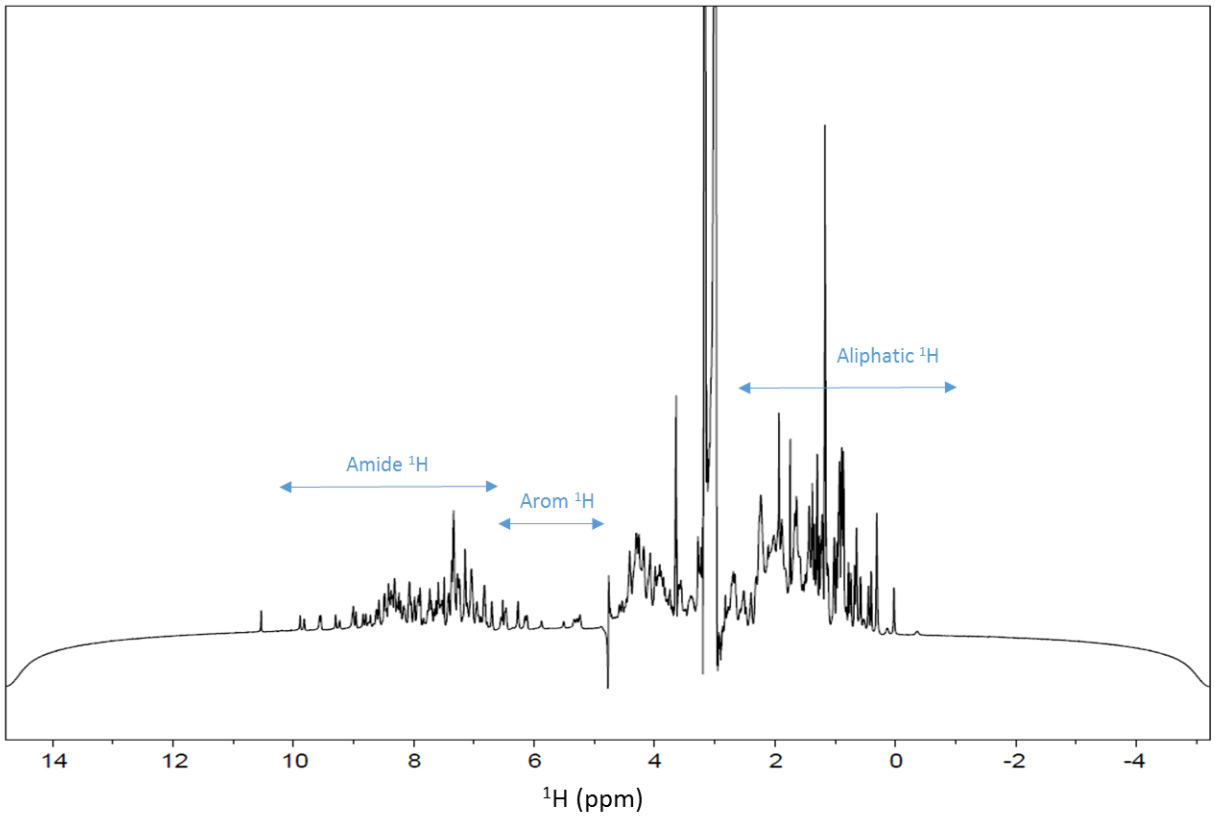


Figure 2.1. 1D-^{1}H spectrum of atVHP76 taken on a 850 MHz Bruker Avance III HD spectrometer equipped with helium-cooled (TCI) probe at 25°C. Arrows indicate regions of amide, aromatic and aliphatic protons.

Backbone experiments and assignments

The assignment of a protein backbone is achieved by "through-bond" correlations. This probes resonances/nuclei that are connected through several covalent bonds in the primary sequence. A general strategy for this assignment, and the one used in this work, uses the following NMR experiments: HNCACB and CBCA(CO)NH as well as other pairs of complementary datasets.

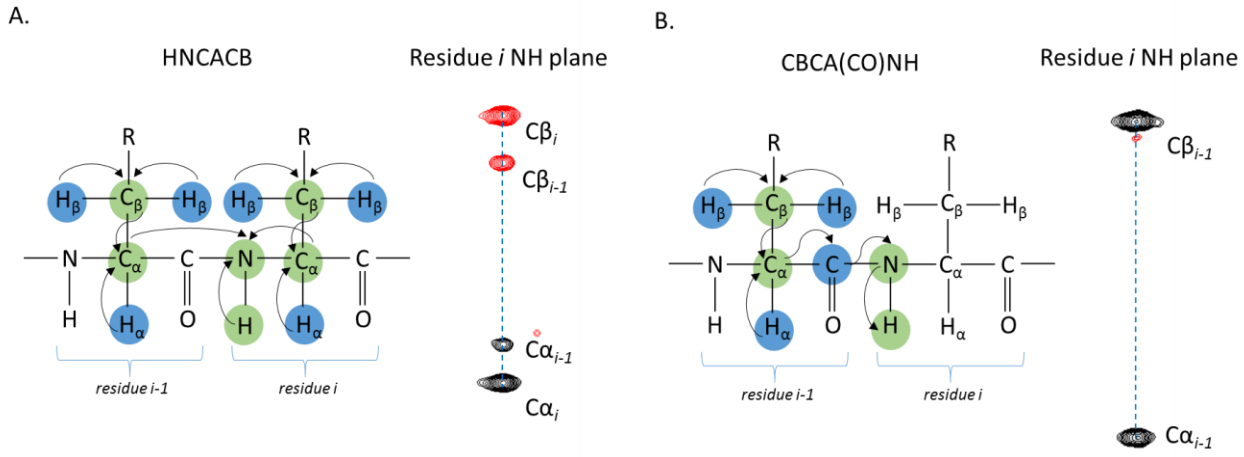


Figure 2.2. Magnetization transfer for both the (A) HNCACB and (B) CBCA(CO)NH backbone assignment experiments. Arrows indicate where the magnetization is transferred and chemical shifts are evolved on atoms designated by green circles. For the HNCACB both the C_α and C_β from residue *i* and *i-1* whereas the CBCA(CO)NH only residue *i-1* shows its C_α and C_β resonances. For the NH planes, the vertical axis (indirect dimension) is ^{13}C chemical shift and the horizontal axis (direct dimension) is ^{1}HN chemical shifts.

These two experiments allow assignment of the Cα, Cβ, and NH atoms by "walking" along the backbone from the C- to N-terminus provided no ambiguous spectral overlaps are encountered (Figure 2.2). In HNCACB, magnetization is first transferred from the ^{1}H-α and ^{1}H-β's to their respective carbons ($^{13}C\alpha$ and $^{13}C\beta$) and then from $^{13}C\beta$ to $^{13}C\alpha$. Next, it is transferred to the ^{15}N and finally to ^{1}HN for detection[54]. Magnetization can occur from $^{13}C\alpha_{i-1}$ to both $^{15}N_{i-1}$ and $^{15}N_i$ and the opposite is true as well, where transfer occurs from both $^{13}C\alpha_i$ and $^{13}C\alpha_{i-1}$ to $^{15}N_i$. Thus, on the ^{15}N plane of a spectrum that corresponds to a particular NH group of the backbone (position i) contains four peaks: the $^{13}C\alpha$ and $^{13}C\beta$ of its own residue (*i*) and $^{13}C\alpha$ and $^{13}C\beta$ of the preceding residue (*i-1*) (Figures 2.2). For the CBCA(CO)NH, magnetization is transferred to $^{13}C\alpha$ and $^{13}C\beta$ in the same manner as above, however, rather than going straight to ^{15}N it is transferred through ^{13}CO then to ^{15}N. This allows only the preceding residue to have evolved peaks in the spectrum[54]. These experiments also allow the assignment of both ^{15}NH and ^{1}HN atoms by choosing the HN plane that contains the ^{13}C peaks with the highest intensity.

The peaks on the HNCACB are phased in a way to allow all $^{13}C\beta$ to be negative and $^{13}C\alpha$ to be positive or vice versa. With both data sets opened together on the same ^{15}NH plane one can decipher which $^{13}C\alpha$ and $^{13}C\beta$ are intra-residue as they will have no corresponding partner in the CBCA(CO)NH (Figure 2.3). Those that do show a ^{13}C shift of the same resonance on both data sets are $^{13}C\alpha$ and $^{13}C\beta$ that are on the residue preceding the NH group.

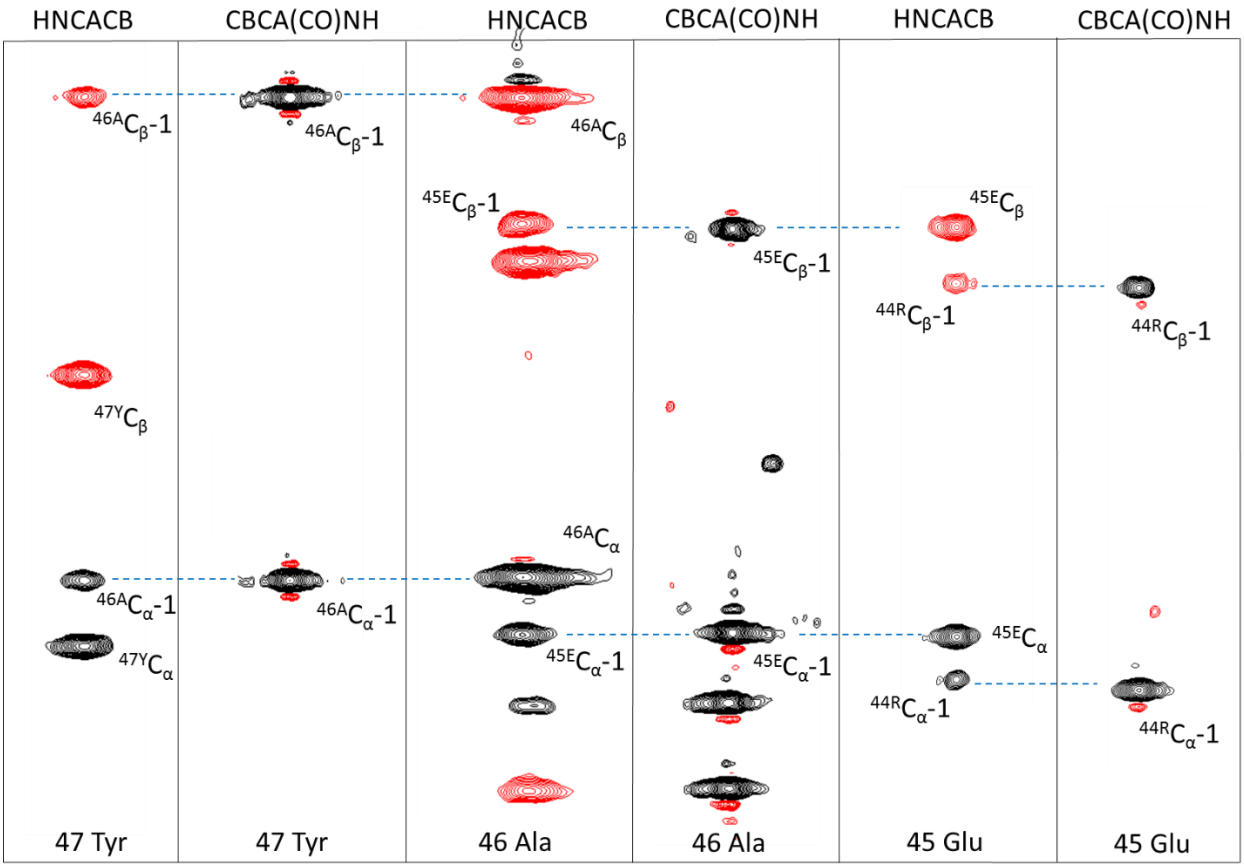


Figure 2.3. Backbone assignments of $^{13}C\alpha_i$, $^{13}C\beta_i$, $^{13}C\alpha_{i-1}$, and $^{13}C\beta_{i-1}$ on atVHP76. Both of these spectra were taken on an 850 MHz Bruker Avance III HD spectrometer equipped with helium-cooled (TCI) probe at 25°C.

The last atomic resonance required to assign in the backbone is the ^{13}CO and that is acquired through two experiments similar to those above: HN(CA)CO and HNCO (Figure 2.4.). From the previous two data sets, the chemical shifts of ^{15}NH and ^{1}HN act as a guide to find ^{13}CO resonances. In HN(CA)CO, magnetization is

first transferred from ^{1}H to ^{15}N and then to the $^{13}C\alpha$ of its own and preceding residue via the J-coupling between ^{15}N-$^{13}C\alpha$. Next, transfer occurs from $^{13}C\alpha$ to ^{13}CO (via the J-coupling), and finally, it is then transferred back the same way to the ^{1}H for detection[54]. For every NH group (or rather its ^{15}N plane) there will be two carbonyl groups observed in the spectrum, *i* and *i -1*. In HNCO, magnetism is only passed from ^{1}H to ^{15}N and then to the ^{13}CO via the ^{15}N-^{13}CO J-coupling. Again, magnetization is transferred back the same path for detection of a single carbonyl group for every NH group of the preceding reside[54]. Using these two data sets side by side allows the assignment of sequential CO groups by a similar method to that above. With the HN(CA)CO and HNCO open together on the same NH plane, one can decipher the CO of the preceding residue as it will contain a peak at the same chemical shift for both the data sets. The remaining peak on the HN(CA)CO will be the CO from the corresponding NH residue (*i*).

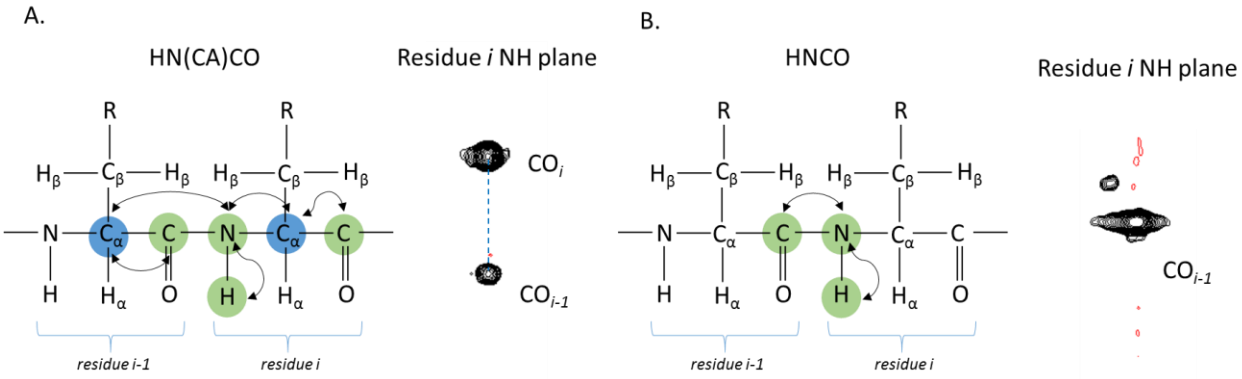


Figure 2.4. Magnetization transfer for both the (A) HN(CA)CO and (B) HNCO data sets. Arrows indicate where the magnetization is transferred and chemical shifts are evolved on atoms designated by green circles. For the NH planes, the vertical axis (indirect dimension) is ^{13}C chemical shift and the horizontal axis (direct dimension) is ^{1}HN chemical shifts.

Side-chain experiments and assignments

There are many experiments available for side-chain assignments. Selection of the optimal experiment is based upon the size of the protein and the available spectrometer time. Due to the small size of headpiece domains we chose an approach that is generally considered the simplest method and collected three data sets for aliphatic side chains: HBHA(CO)NH, H(CCO)NNH, and CC(CO)NNH (Figure 2.5). These allow the

assignment of the hydrogens and carbons along aliphatic side chains (*i-1*) with respect to the C-terminal NH group (*i*). First the HBHA(CO)NH data set shows both Hα and Hβ for the residue preceding the NH group. These two protons are easily distinguished in this data set due to the difference in average chemical shifts between Hα (~4 ppm) and Hβ (~1.5-2ppm) protons. Once these two protons chemical shifts are known, they become a starting place to use for the H(CCO)NH data set. For every NH group there should be peaks from protons Hα, Hβ and beyond on the side chain preceding that residue. There is a potential for some longer residues to not have all their protons present on this spectrum, as well as some ambiguous protons that share similar chemical shifts such as Hβ and Hδ. The last experiment, CC(CO)NH allows the assignment of all carbons along the aliphatic side chain, including the Cα. Again, this is very similar to the H(CCO)NH data set where peaks appear for carbons on the residue preceding the NH group. As a quality control, we also checked these carbon assignments through ^{13}C-HSQC spectra.

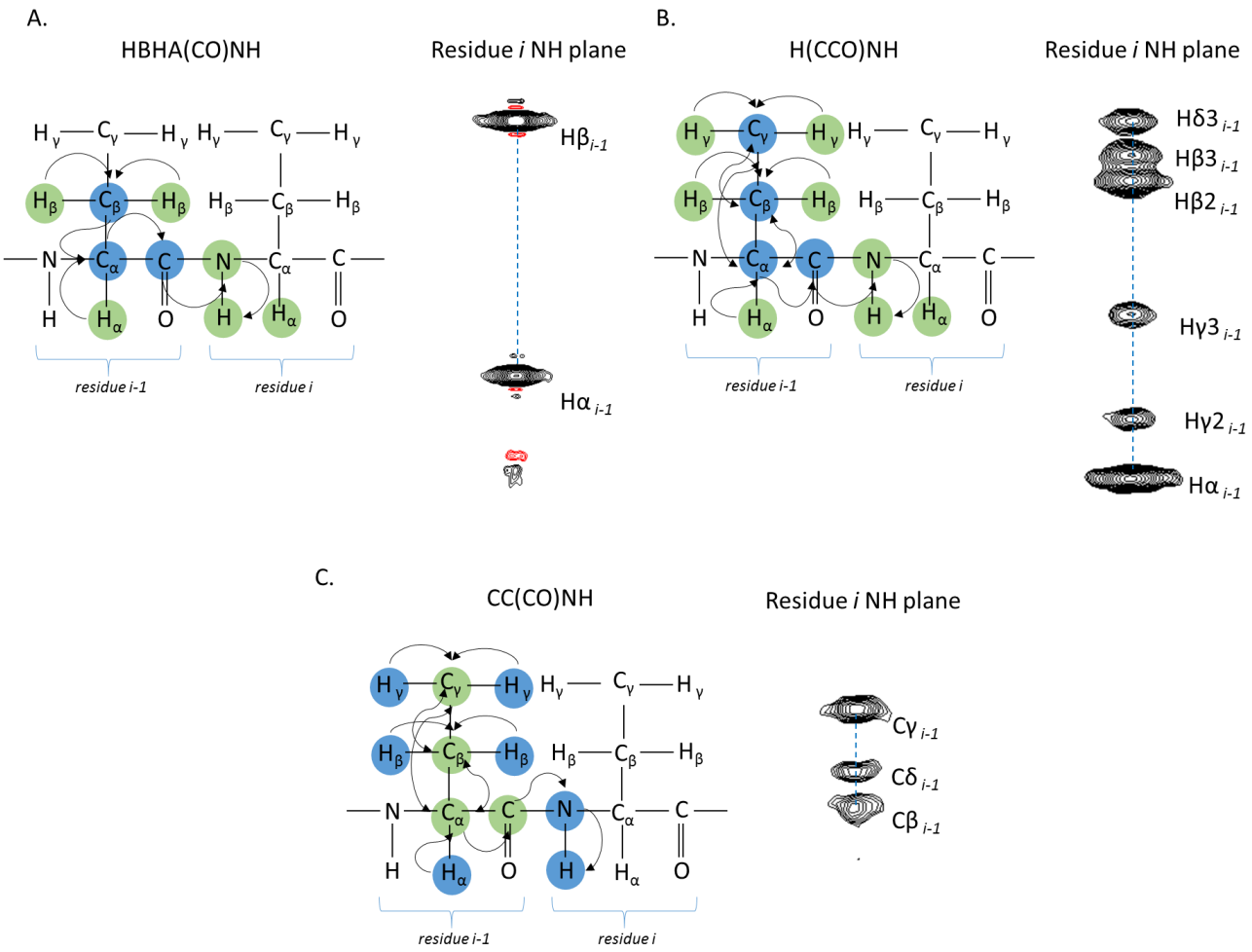


Figure 2.5. Magnetization transfer for the three side-chain experiments used in this work: (A) HBHA(CO)NH, (B) H(CCO)NH and (C) CC(CO)NH. Resonances that are evolved in the spectra are highlighted with green circles. For the NH planes, the vertical axis (indirect dimension) is ^{13}C chemical shift and the horizontal axis (direct dimension) is ^{1}HN chemical shifts.

The high chemical shift values (>100ppm) associated with carbons in aromatic residues do not allow these to be assigned using the methods listed above. In order to assign aromatic carbons such as Cδ, Cε, and Cζ a ^{13}C-Aromatic HSQC can be used. Every peak in this data set refers to an aromatic carbon residue directly attached to a proton. Then, to assign the aromatic protons, there are two specific data sets: (Hβ)Cβ (Cγ)CδHδ and (Hβ)Cβ (CγCδ)CεHε. These 2D datasets allow assignment of side chain Hδ and Hε by utilizing resonance transfer from the Cβ of that residue.

Distance restraints

To determine a solution NMR structure the distances between protons <5Å apart need to be measured within the protein. These distances then become restraints for a computational structure calculation. Specifically, a Nuclear Overhauser Effect (NOE) is used to calculate these distances with the formula $\frac{1}{r^6}$, where r is the internuclear distance. An NOE comes from a set of data called a NOESY, which probes "through-space" correlations. A 2D ^{1}H-NOESY show two protons that are <5Å apart in space which can give rise to hundreds or thousands of peaks. However, many of these cannot be used for structure calculation due to the high level of spectral overlap, especially in the aliphatic region (0-2.5ppm) (Figure 2.6).

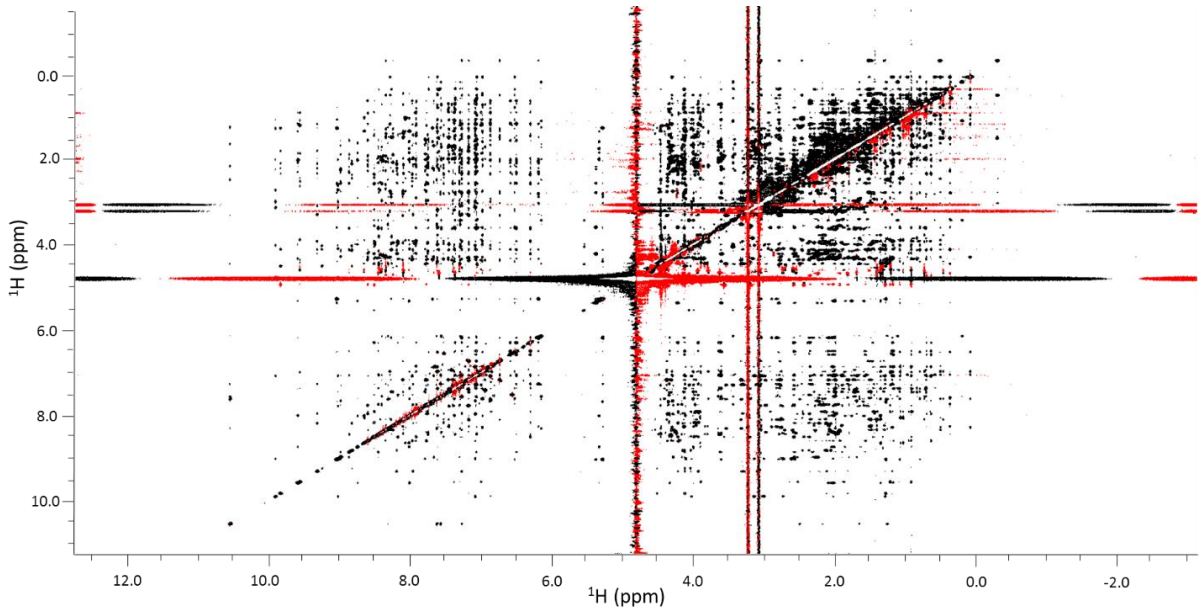


Figure 2.6. 2D ^{1}H-NOESY of atVHP76 taken on an 850 MHz Bruker Avance III HD spectrometer equipped with helium-cooled (TCI) probe at 25°C. A total of 1216 peaks were picked from this data set and determined to be amenable for structure calculations.

To further increase spectral resolution, there are three types of heteronuclear 3D NOESY experiments: ^{13}C-aliphatic NOESY, ^{13}C-aromatic NOESY, and a ^{15}N-NOESY[54]. An aliphatic NOESY will only contain peaks from protons, one of which is bound to an aliphatic carbon. Similarly, in aromatic NOESY one proton must be bound to an aromatic carbon. The 3D ^{15}N-NOESY only shows peaks of protons in space that are <5Å to a proton directly bound to a ^{15}N, or the NH. Both 2D and 3D NOESY experiments are required for structural determination. Combining distance restraints from all four of these NOESY data sets gives rise to thousands of distance restraints for the structure calculation.

Chapter III: Materials and Methods

Buffers

Minimal Media: 22 mM Na^+/K^+ phosphate buffer (pH.7.0), 85 mM NaCl, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 1 mg/ml B Vitamins (J.T. Baker)

Lysis Buffer: 50 mM sodium phosphate (pH 8.0), 300 mM NaCl, 10 mM Imidazole

Ni-column Wash Buffer: 50 mM sodium phosphate (pH 8.0), 300 mM NaCl, 20 mM Imidazole

Ni-column Elution Buffer: 50 mM sodium phosphate (pH 8.0), 300 mM NaCl, 250 mM Imidazole

PIPES Buffer: 20 mM PIPES (pH 6.8), 50 mM NaCl

General Actin Buffer: 5 mM Tris-HCl (pH 8.0), 0.2 mM $CaCl_2$

Actin Polymerization Buffer: 100 mM Tris-HCl (pH 7.5), 50 mM KCl, 2 mM $MgCl_2$, 1 mM ATP

F-actin Buffer: 79 mM Tris-HCl (pH 8.0), 50 mM KCl, 0.2 mM $CaCl_2$, 2 mM $MgCl_2$, 1 mM ATP

NMR Buffer: 10% $^2H_20$, 20 mM PIPES (pH 6.8), 50 mM NaCl, 0.02% NaN_3

Construction of atVHP76 and atVHP60 vector

The atVHP76 fragment was designed to include the 76 C-terminal residues, and atVHP60 included the 60 C-terminal residues, of *Arabidopsis thaliana* villin-4 fused with an N-terminal $(His)_6$-tag. The constructs were cloned into the pET-24a vector (Novagen) using commercial services (Genscript) with primers containing an NdeI site at the 5′ end and a HindIII site at the 3′ end of the TAG stop codon.

Expression of atVHP76 and atVHP60

Expression of atVHP76 and atVHP60 was produced according to procedures previously adapted by WWU students Lucian Burns, Alex Nelson, Danielle Pfaff and Stas Fedechkin. The *E.coli* strain BL21-(DE3) was transformed with the pET-24a plasmid containing the atVHP76 or atVHP60 sequence using heat shock transformation. The cells were then plated on Kanamycin Luria-Bertani broth (LB) plates. An isolated colony was used to inoculate a seed culture (~100ml), which was then allowed to grow overnight at 37 °C. This starter culture was used to inoculate a large scale growth of the transformed strain. The strain was grown in 6L of LB at 37 °C, containing 10µg/ml of kanamycin. The growth rate was monitored by taking OD_{600} readings on a V-1200 Spectrophotometer (VWR) until an OD_{600} of ~0.6 was reached. Protein expression was then induced with 0.4 mM ispropyl β-D-1-thiogalactopyranoside (IPTG) and continued for 5 hours at 37 °C. Cells were pelleted by centrifugation on a Sorvall Lynx 4000 centrifuge (Thermo Scientific) at 6,000 x g for 10 minutes and stored at -20°C until purification.

Expression of $^{13}C/^{15}N$ atVHP76 for solution structure determination

Expression of isotopically labeled atVHP76 was also produced according to procedures previously adapted by Lucian Burns, Alex Nelson, Terry Webb, Danielle Pfaff, and Stas Fedchkin. The *E.coli* strain BL21-(DE3) was transformed with the pET-24a plasmid containing the atVHP76 or atVHP60 sequence using heat shock transformation and the cells were plated on Kanamycin Luria-Bertani broth (LB) plates. An isolated colony was used to inoculate a seed culture (~100ml), which was then allowed to grow overnight at 37 °C. This starter culture was used to inoculate a large-scale growth of the transformed strain. The strain was grown

in 6L of LB at 37 °C, containing 10µg/ml of kanamycin. The growth rate was monitored by taking OD_{600} readings on a V-1200 Spectrophotometer (VWR) until a OD_{600} of ~0.6 was reached. Cells were then pelleted via centrifugation on a Sorvall Lynx 4000 centrifuge (Thermo Scientific) for 10min at 4,000 x g and resuspended/washed in minimal media (M9) containing no nutrient sources. The cells were again pelleted by centrifugation and resuspended in minimal media containing ^{15}N and ^{13}C labeled nutrients (2.4g $^{15}NH_4CL$, 3.0g $^{13}C_6$-D-Glucose). Cells were allowed to equilibrate in the media for 1 hour at 37 °C and then induced with 0.8 mM IPTG and incubated for 5 hours at 37 °C. Cells were then pelleted via centrifugation for 30 minutes at 4,000 x g and stored at -20°C until purification.

Affinity purification of both $^{15}N/^{13}C$ and unlabeled atVHP76 and atVHP60

Cell pellets were resuspeneded in lysis buffer along with addition of lysozyme (50 µg/mL) and were left to shake at 4 °C for 30 minutes. Cells were then lysed via sonication in three 30-second intervals (Branson sonifer, 50% duty cycle). DNase I (1 µg/mL) was added to the lysed cells for incubation at room temperature for 30 min. Lysed cells were then pelleted at 36,000 x g for 30 minutes on a Sorvall Lynx 4000 centrifuge (Thermo Scientific) to pellet insoluble cell debris. The supernatant was then filtered through successive 5 µM and 0.45 µM filters. Nickel-nitrilotriacetic acid (Ni-NTA) Superflow resin (QIAGEN) was equilibrated with lysis buffer and then clarified lysate was added for incubation with the resin for one hour at 4°C with shaking. The solution was added to a gravity flow Ni-NTA column and after elution of the flow-through, the column was rinsed with four column volumes of wash buffer. $(His)_6$-tagged atVHP76 or atVHP60 was then eluted with elution buffer in 1-mL fractions. All fractions were analyzed for the presence of protein by 15% sodium dodecyl sulfate polyacrylamide gels (SDS-PAGE).

Size exclusion purification of atVHP76 and atVHP60

For the final purification, pooled fractions that contained $(His)_6$-tagged atVHP76 or atVHP60 were loaded onto a HiLoadTM 16/60 Superdex 75 preparatory grade size-exclusion chromatography column (120 mL, GE Healthcare). The column was pre-equilibrated with PIPES buffer and attached to an ÄKTAprime Plus liquid chromatography system (GE Healthcare). The column was eluted at 1 ml/min and the major peak was collected and protein purity was analyzed using 15% SDS-PAGE. The identity of the purified protein fragments was confirmed with mass spectroscopy (Advion expression LCMS). Pure samples were concentrated to ~1mM and confirmed by UV-Vis A_{280} and the Beer-Lambert law, where the molar extinction coefficient was estimated using ProtParam[55]. The samples were fractioned into 100 ul aliquots and stored at -80° for future experiments.

F-actin pull-down assay of atVHP76 and atVHP60

To investigate binding of actin filaments by either atVHP76 or atVHP60 an Actin Binding Protein Spin-Down Assay Biochem Kit™: Muscle Actin (Cytoskeleton Inc.) was employed. Lyophilized rabbit skeletal actin was resuspended to a concentration of 1 mg/mL with general actin buffer and allowed to equilibrate at 4 °C for 30 minutes before the addition of actin polymerization buffer for F-actin stock (21 µM) generation. The mixture then was equilibrated for one hour at room temperature for F-actin polymerization. Lyophilized bovine serum albumin (BSA) and α-actinin were used as negative and positive controls for F-actin binding respectively, and were resuspended to the stock concentrations of 3.4 mg/mL and 1 mg/mL respectively. A clarification step for atVHP76 or atVHP60 and BSA was performed by centrifugation at 150,000 x g in a Sorvall MX 150 micro-ultra centrifuge (Thermo Scientific) for one hour at 4 °C. Headpiece samples at a stock concentration of 10, 25, 50, 100, 150, 200, 250, 300, 600 µM (10 µl in PIPES buffer, concentration of 2, 5, 10, 20, 30, 40, 50, 60, or 120 µM after dilution) were incubated with F-actin stock (40 µl, actin concentration 16.8 µM after dilution) for 1 hour at room temperature before centrifugation at 150,000 x g for 1.5 hours at 24 °C. The reaction buffer mixture (PIPES and Tris) meets

conditions for F-actin polymerization. Supernatant was removed and pellets were gently washed with F-actin buffer. Pellets were then resuspended with 7% acetic acid solution and an aliquot of all samples, including controls, was removed for protein analysis by HPLC. A Luna 5U C18 150 x 4.6 mm (Phenomenex) column connected to an Ultimate 3000 HPLC (Thermo Scientific) was pre-equilibrated with 10% solution of buffer A (0.1% TFA, ddH_20) and samples were separated using a 10-90% gradient of buffer A to B (0.1% TFA, 90% Acetonitrile). Chromatograms were analyzed with Chromeleon software. The F-actin binding activity was also tested through 15% SDS-PAGE analysis of the pull-down mixtures.

To determine the concentration of headpiece pulled down with F-actin, a calibration curve was completed for headpiece concentrations of 1-5 uM, which allowed the quantification of all concentrations assayed for atVHP76 and atVHP60, excluding the lowest concentration of 2 uM for atVHP60, as we could not get an accurate measurement for this concentration. All calibration curves had an R^2 of > 0.999.

The binding data was fit to a published ligand-receptor interaction model for villin headpiece domains[33,25]. The system includes an equilibrium dissociation constant (K_d), a maximal binding constant (B_{max}), and a non-specific parameter (NS), which includes any trapping or non-specific interactions with the actin filaments. The ligand-receptor interaction formula is as follows:

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

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

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

$$a = 1 - NS,$$

$$b = K_d + NS \cdot K_d + [HP_{total}] + 2NS[HP_{total}] + B_{max},$$

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

NMR Data collection and processing for $^{13}C/^{15}N$ atVHP76

NMR samples contained 0.3-1.0 mM of atVHP76 and atVHP60 (unlabeled as well as labeled with ^{15}N and ^{13}C as appropriate) and were suspended in NMR buffer. No correction for the effect of $^{2}H_2O$ was made during the pH adjustment. 2D and 3D NMR spectra for atVHP76 were acquired at 25 °C on the Bruker Avance III HD spectrometers operating at magnet strength values of 850 MHz, 600 MHz and 500 MHz (1H frequency) and equipped with helium-cooled (TCI) or room-temperature (SmartProbe) probes (University of British Columbia, Vancouver, BC, Canada; Boston University School of Medicine, Boston, MA; Western Washington University, Bellingham, WA). All spectra were processed using NMRPipe[56] and analyzed using NMRViewJ[57] (One Moon Scientific).

NMR resonance assignments

NMR resonance assignments for the backbone (^{1}H, ^{15}N, and ^{13}C) were produced by the standard approach from a combined use of 2D ^{15}N HSQC and the following 3D data sets: HNCACB, CBCA(CO)NH, HNCO, HN(CA)CO, HNCA, and HCA(CO)N. The HCA(CO)N and HCAN experiments were used to confirm backbone assignments for the first 20 residues. The side chain resonance (aliphatic ^{1}H, ^{15}N and ^{13}C) assignments were produced from: HBHA(CO)NH, H(CCO)NH, and CC(CO)NH. Aromatic side chain ^{1}H assignments were derived from ^{13}C HSQC, ^{15}N HSQC, (Hβ)Cβ(CγCδ)Hδ, (Hβ)Cβ(CγCδCε)Hε, (Hβ)Cβ(CγCarom)Harom, and ^{15}N-TOCSY. Peak visualization and manual assignment of the backbone and side chains were achieved using NMRViewJ. All samples were referenced internally with water ^{1}H resonance at 25 °C.

Solution NMR structure calculation

The following experiments were collected for distance restraints: ^{15}N-NOESY (mixing time 80 ms), ^{13}C-NOESY aliphatic (mixing time 100 ms), and ^{13}C NOESY aromatic (mixing time 100 ms). Peak visualization and volume measurements were performed with NMRViewJ. The solution structure of atVHP76 was determined using CYANA 2.1 using the NMR-based distance and dihedral angle restraints. Distance

restraints were derived from heteronuclear 3D and proton 2D NOESY data recorded in 90% H_2O. The dihedral angle restraints were produced prior to the CYANA simulations with PREDITOR[58] utilizing the backbone chemical shift values of ^{1}H, ^{15}N, $^{13}C_{\alpha}$, ^{13}CO, and $^{12}C_{\beta}$. Initially, CYANA simulations utilized only the 3D NOESY (^{15}N NOESY, ^{13}C NOESY aliphatic, and ^{13}C NOESY aromatic) sets to assign automatically during the iterative simulations. Each simulation was conducted according to a standard CYANA protocol (*noeassign* macro) of iteratively determining the solution structure while concurrently assigning the maximally high number of NOE cross-peaks. The simulations started with 1000 test conformers and employed 10,000 steps of torsion angle dynamics. The 10 conformers with the lowest target function were retained for analysis. Every simulation was performed with more than one seed value, and the common assignments from the simulations were used as an assigned set of NOEs for the next simulations. After no more of the 3D NOESY peaks were assigned, the peaks/volumes from the 2D ^{1}H NOESY were added and assigned by the same iterative process during CYANA runs while all the previously obtained 3D NOE assignments were retained. Every 2D or 3D NOE peak generated distance restraint >0.5 Å at the end of a run was manually erased for the next simulation after investigation of the peak validity. The final ensemble of 10 structures with the lowest penalty energy was submitted to the PDB (entry 5VNT) and BMRB (entry 30289) repositories.

Structure representation and modeling

The structure models for presentation were generated with Chimera[59] and PyMOL[60] , solvent accessibility area values calculated with DSSP database[61, 62]. Homology modeling of the headpiece domains in plant villins 1,2, 3 and 5 was performed via the Phyre2 online server[63].

NMR investigation of ^{15}N atVHP76 for ion sensitivity

A 1 mM sample of ^{15}N atVHP76 was prepared by standard procedures, and buffer exchanged through a Amicon Ultra-15 centrifugal filter device with a 3kDa MW cutoff (Millipore) to lower the NaCl concentration from 50 mM to 145 µM. A ^{15}N-HSQC spectrum was collected after the sample was brought

up to the following concentrations of NaCl: 145 µM, 3.6, 50, 91, 100, 200, and 300mM. All spectra were processed and analyzed by overlaying the data sets to determine if any peaks show chemical shift changes.

The sample was buffer exchanged again via a filtration device to the standard 50 mM NaCl buffer and then was equilibrated with first 100mM then 200mM KCl. A ^{15}N-HSQC spectrum was collected after each equilibration step and processed and analyzed by overlaying the data sets.

Solution NMR dynamics measurements

Longitudinal, R_1, transverse, R_2, and heteronuclear NOE ^{1}H-^{15}N, relaxation rates of backbone amide nitrogens in 0.9 mM atVHP76 were measured using standard pulse sequences [64] on a Bruker 500 MHz spectrometer with the Avance III console and equipped with a cryoprobe. The ^{1}H-^{15}N NOE rates were measured from pairs of spectra recorded with (NOE) and without (control) proton saturation during the recycle delay. The saturation period was five seconds in the NOE experiment and the recycle delay was five seconds in the control experiment. Each NOE experiment was repeated three times. Relaxation delays were the following for the R_1 experiment (0.01, 0.07, 0.14, 0.23, 0.35, 0.77, 1 s) and for the R_2 experiment (0.004, 0.038, 0.078, 0.126, 0.184, 0.260 s). The recycle delay of 2.5 sec and 1024x80 complex points were employed. Relaxation rates were analyzed via NMRViewJ and were plotted in Origin 2016.

Molecular dynamics on V4HP63

Molecular dynamics package AMBER was utilized at every stage of the simulation[65].

System initialization: The starting structure was the solution NMR structure of V4HP63. Each structure was parameterized with the AMBER ffSB14 force field. Each system was neutralized with 3 Cl^- ions using Joung and Cheatham parameters[66] with CPPTRAJ[67] from AmberTools 15. 3058 explicit solvent molecules were added using the TIP3P water model in truncated octahedral box.

System minimization and equilibration: Initially, the protein atoms were held fixed. To eliminate van der Waals clashes, the solvent molecules and ions were subjected to 1000 steps of the steepest descent

minimization. Subsequently, 1000 steps of conjugate gradient minimization were carried out, with a force constant of 25 (kcal/mol)/Å^2 applied to the solute molecule. Next, the whole system was minimized with 1000 steps of the steepest descent followed by 1500 steps of conjugate gradient minimization without restraints. Heating was completed over 10 ps at constant volume from 100 K to 300 K with weak positional restraints (25 (kcal/mol)/Å^2) applied to the protein molecule. A 2 fs time step was used and a weak coupling thermostat was used to control temperature, with a collision frequency of 1.0 ps^{-1}. SHAKE[68] was used to constrain bonds involving hydrogen atoms, with a tolerance of 0.002 Å. A 10 Å cutoff was used for non-bonded interactions and the Particle Mesh Ewald (PME) method was used to compute long range electrostatics. A five step minimization protocol followed heating in which the restraints applied to the protein were gradually reduced from 5.0 to 0.5 (kcal/mol)/Å^2. At each step, minimization was carried out with 1000 steps of the steepest descent followed by 1500 steps of conjugate gradient minimization. A Berendsen coupling constant of 0.2 ps was used. Finally, 100 ps of fMD was run at 300K with no restraints and constant pressure to relax the density of water.

Production molecular dynamics: Production MD simulations were run for 1.1 µs using graphics processing code (GPU) code with the PMEMD.cuda implementation of SANDER from Amber14 on an NVIDIA GTX 970 GPU. Simulations were held at constant pressure (1 atm) periodic boundaries and 300 K using Berendsen coupling constants of 5.0 ps. Long range interactions were calculated with PME. A 2 fs integration time step was used. Simulation coordinates were recorded every 1 ps. Trajectories analysis was performed with CPPTRAJ, where, the structural stability of the simulation was examined in terms of RMSD vs simulation time. RMSD figures were generated and running averages were calculated using the XMGRACE program[69].

Chapter IV: Results and Discussion

Preparation of atVHP76 and atVHP60

Genes encoding the headpiece sequence atVHP76 or atVHP60 were cloned separately into the pET-24a plasmid and then transformed into BL21 *E.coli* cells. We included an N-terminal $(His)_6$-tag to our headpiece sequences to aid in purification. Transformed bacterial cells were grown and induced to allow protein expression, and then harvested and lysed by standard methods (Materials and Methods). The filtered lysate was incubated with Ni-NTA superflow resin, which binds with high affinity to the N-terminal $(His)_6$-tag. After incubation, the column was washed with a low concentration of imidazole to remove non-specific interactors on the Ni-NTA resin. A high concentration of imidazole is used to compete for the binding sites on the Ni-NTA resin, dissociating the $(His)_6$-tag headpiece domain. Eluent from each step was analyzed by SDS-PAGE to determine which eluted fractions contained the headpiece fragment (Figure 4.1). In order to remove the trace impurities a final step of size exclusion chromatography (SEC) was employed (Materials and Methods). Fractions of each peak at the retention volumes of 75 ml for atVHP76 and 95 ml for atVHP60, were collected and assessed by SDS-PAGE to determine purity. A single band was present for the main peak indicating a purified single species of headpiece fragment (Figure 4.2, 4.3).

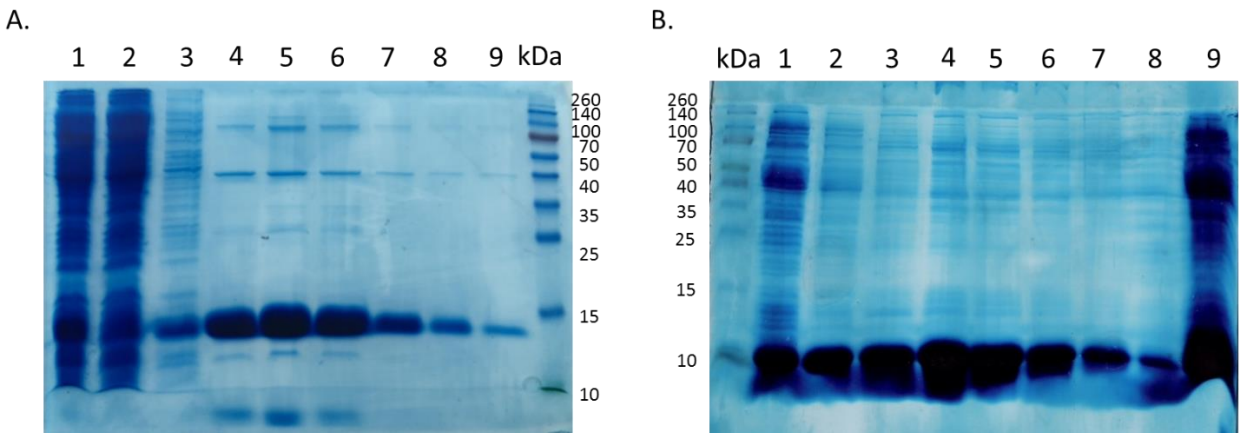


Figure 4.1. Ni-NTA purification of atVHP76 (A) and atVHP60 (B). Lanes: (A) 1. Clarified lysate; 2. Flow-through; 3. 20 mM Imidazole Wash; 4-9. Elution 1-6. (B) 1. Flow-through; 2-8. Elution fractions 1-7; 9. Clarified lysate.

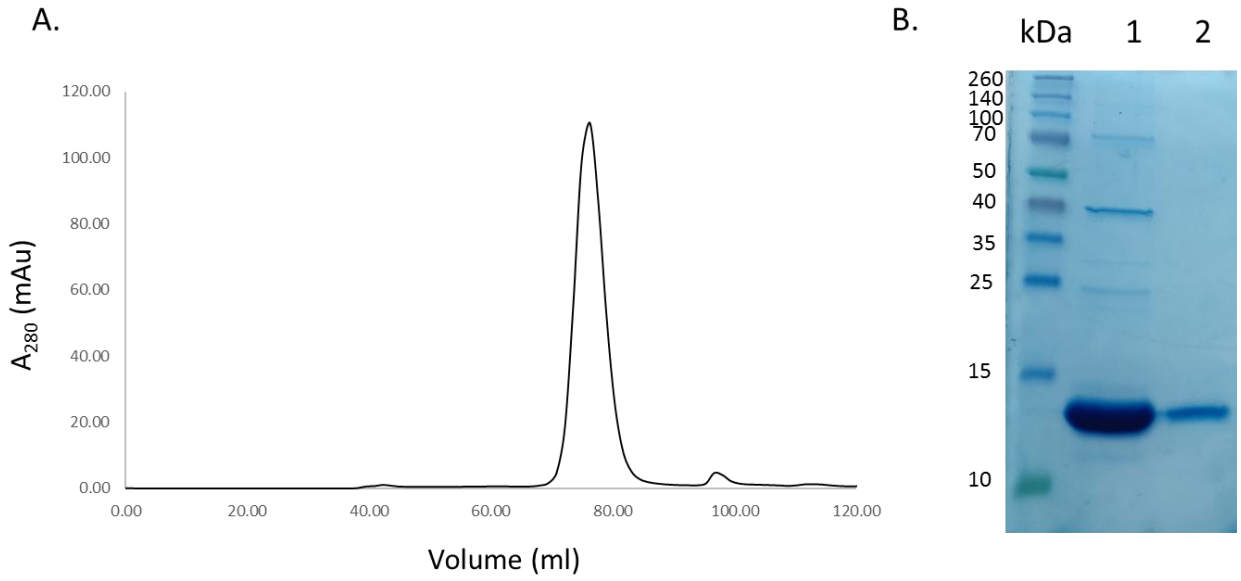


Figure 4.2. Purification by SEC of atVHP76. A) Chromatogram of IMAC purified atVHP76 B) SDS-PAGE of. (Lanes: 1. IMAC eluent; 2. SEC fraction at 75 ml retention volume)

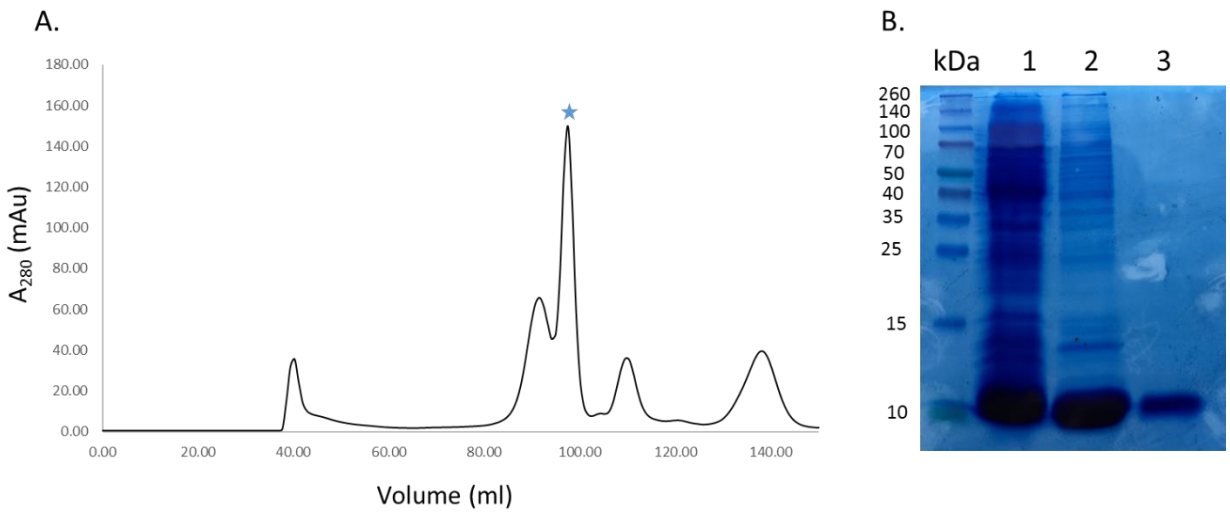


Figure 4.3. Purification by SEC of atVHP60. A) Chromatogram of IMAC purified atVHP60 B) SDS-PAGE of atVHP60 (Lanes: 1. atVHP60 clarified lysate; 2. IMAC eluent; 3. SEC fraction at 95 ml retention volume)

Preparation of $^{15}N/^{13}C$ atVHP76

In order to enrich for ^{15}N and ^{13}C isotopes, atVHP76 was overexpressed in *E. coli* using an established minimal media expression protocol[70] (Materials and Methods). The two-step purification system in Materials and Methods was sufficient to purify labeled $^{15}N/^{13}C$ atVHP76 (Figure 4.4, 4.5). Samples that contained pure headpiece were concentrated up to 1.5 mM and exchanged into 10% D_2O buffer for NMR experiments. Enrichment efficiency for $^{15}N/^{13}C$ nuclei was assessed by a standard ^{15}N-HSQC.

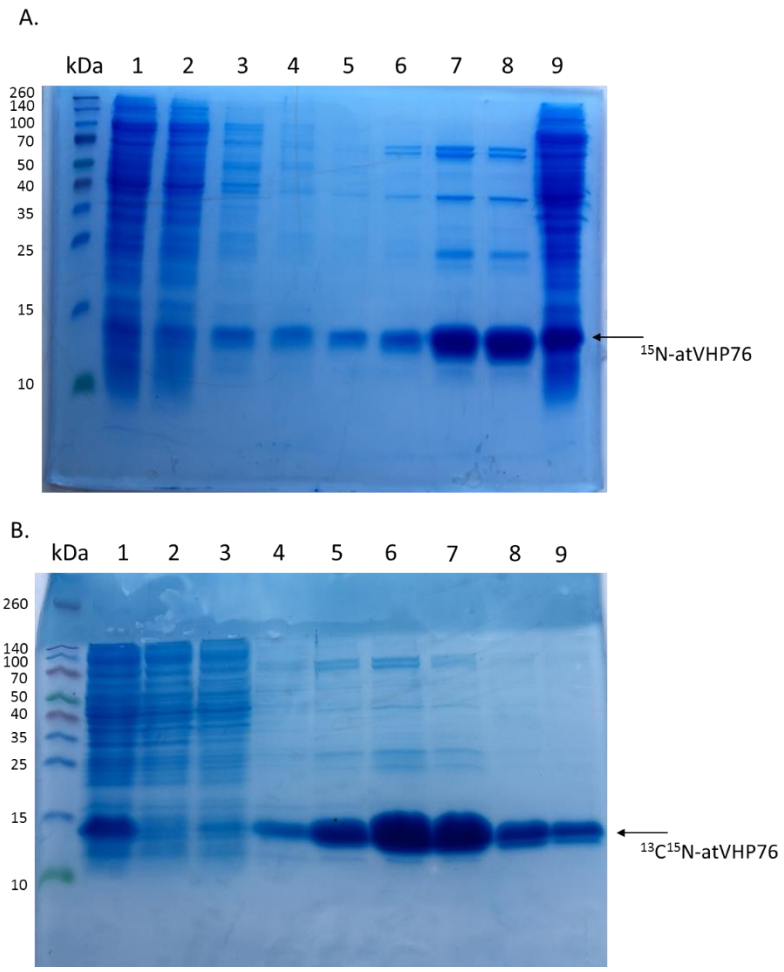


Figure 4.4. SDS-PAGE of ^{15}N-atVHP and ^{13}C/^{15}N-atVHP. A) Ni-NTA purification (Lanes: 1. Flow-through; 2-5. Imidazole washes; 6-8. Elution fractions; 9. Filtered lysate. B) Ni-NTA purification fractions of filtered lysate (1), loading flow-through (2), imidazole washes (3-4), and elution fractions (5-9).

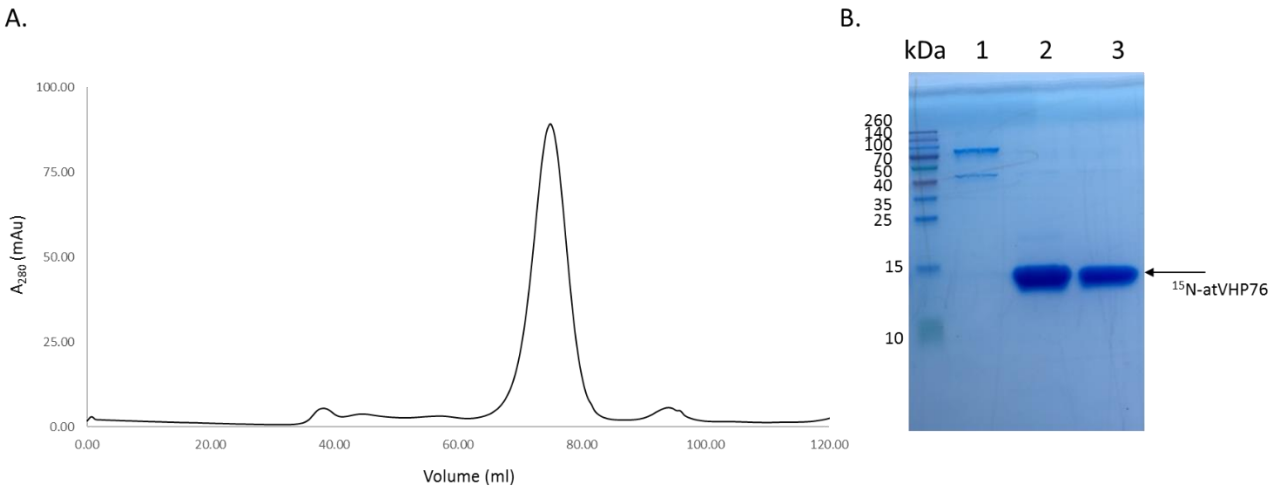


Figure 4.5. Purification by SEC of ^{15}N-atVHP76. A) Chromatogram of IMAC purified ^{15}N-atVHP76. B) SDS-PAGE of ^{15}N-atVHP76 (Lanes: 1. SEC 60 ml retention volume; 2,3. SEC fraction at 75 ml retention volume)

NMR Resonance Assignment of atVHP76

The ^{15}N HSQC spectrum of atVHP76 contains a set of well-dispersed cross-peaks characteristic of a mostly folded polypeptide (Figure 4.6). A small cluster of cross-peaks containing ^{1}H chemical shifts values between 8.0-8.5 ppm was observed, which may indicate exposure to solvent and structural disorder. This was expected for atVHP76, as it contains residues in the linker that are predicted to be disordered. However, there was a relatively low number of peaks within this region, suggesting a folded domain.

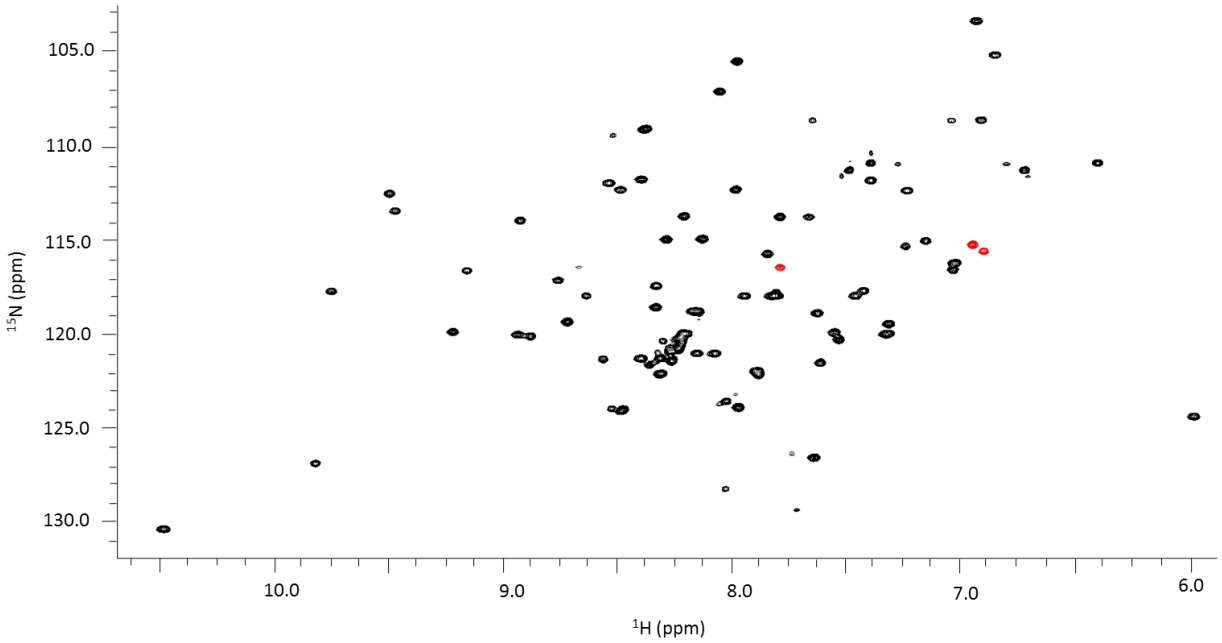


Figure 4.6. ^{15}N HSQC of atVHP76 taken on a 500 MHz Bruker Avance III HD spectrometer equipped with a room temperature probe at 25°C.

Backbone resonance assignments were completed using the following pairs of spectra: HNCACB and CBCACO; HNCO and HNCACO. To begin the assignment, a sequential TTST was identified in the primary structure, as these residues give characteristic large downfield chemical shift values for their Cβ atoms. From this starting point, sequential assignments were elaborated by methods described in Chapter II.

Backbone assignments identified as the N-terminal 14 residues, as well as the N-terminal $(His)_6$-tag, showed ^{1}H chemical shift values within 8.0-8.5 ppm, suggesting residue 14 as the boundary between the N-terminal linker fragment and the folded domain.

Following backbone assignments, side chains were assigned using methods described in Chapter II. Peaks that were ambiguous were left unassigned and not utilized in the structure calculation. Finally, a combination of four different NOESY experiments (listed in Materials and Methods) were picked and their volumes were integrated to be converted into distance restraints via automatic volume-to-distance calibration (see Chapter II).

Villin 4 headpiece has the canonical “villin headpiece” fold in solution

The solution NMR structure was determined using distance restraints from 2D and 3D NOESY spectra and dihedral angle restraints from chemical shift values. The ensemble of the 10 best lowest-penalty NMR structures contained backbone R.M.S deviation of 0.18 Å (Figure 4.7, A). Further statistics for the ensemble are shown in Table 2. During structural calculations only intra-residual or sequential NOE restraints were observed for the first 14 residues (residues 7-19) and no NOEs were observed between this stretch and folded headpiece domain. Therefore, the final solution structure only contained the last 63 residues of atVHP76 and was newly named V4HP63 (as such, numbering is maintained from atVHP76).

Table 2. NMR Statistics of V4HP63 Structure Calculation

no. of NOE distance constraints	
Total	625
intraresidue (I=j)	155
sequential (|i-j| = 1)	156
medium-range (1 < |i-j| < 5)	174
long-range (|i-j| ≥ 5)	140
average constraints per residue	10
no. of dihedral angle constraints	
	215
no. of distance violations per model[a]	
< 0.4 Å	59
>0.4 Å	0
no. of dihedral angle violations per model	
1.0-9.0°	10
>9.0°	0
average RMSD from the mean for the 10 best models	
all backbone atoms (residues 23-82)	0.18 Å
all heavy atoms (residues 23-82)	0.67 Å

[a]Average values

V4HP63 forms a fold typical of many villin headpiece domains, which consist of exclusively α helices, loops and turns (Figure 4.7). Like other headpiece domains, V4HP63 is composed of two approximately equally

sized subdomains[10]: a 34-residue C-terminal subdomain (residues 49-82) composed of three α-helices (A: K69-A78; B: K61-K66 and C: residues S50-F57) and a 29-residue N-terminal subdomain (residues 20-48). The N-terminal subdomain shows only regular secondary structure within a kinked α-helix (D: residues V41-Y47).

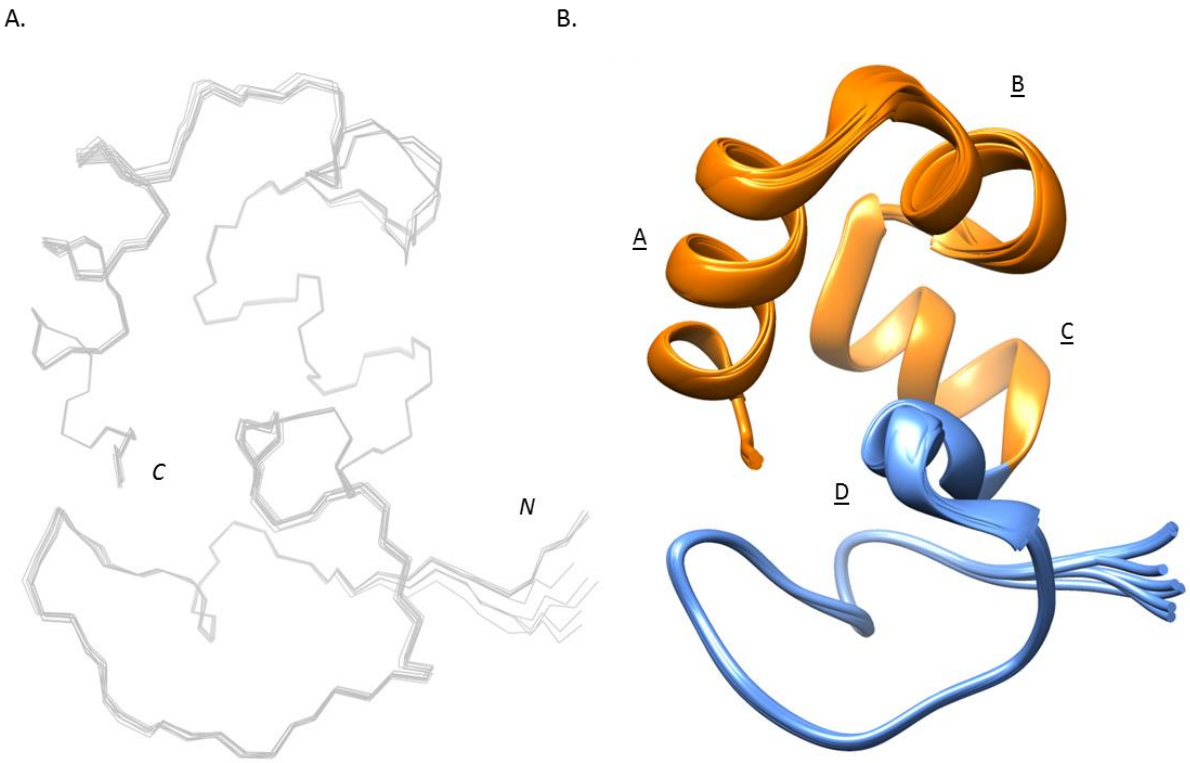


Figure 4.7. V4HP63 from the 10 best solution NMR structures. (A) The backbone depicted in wireframe and (B) Ribbon model of V4HP63 colored by subdomain: C-terminal (orange) and N-terminal (Blue) with the conserved α-helices (A-D) labeled.

The hydrophobic core of V4HP63 was classified utilizing DSSP[59, 60], which calculates the area of solvent accessibility of residues measured in Å^2. The hydrophobic core of VHP is composed of residues <15% solvent accessible. However, when this constraint was applied to V4HP63, residues that were clearly solvent exposed were calculated to be within the hydrophobic core, indicating that this percentage was too high. Therefore, we lowered the cutoff to residues <10% solvent accessible to account for the smaller total surface area of V4HP63. These residues included F53, F57, A63, F64, Y65, L67, K76, and L81 within

the C-terminal subdomain and A22, H23, Y25, L28, E45, Y47 and L48 within the N-terminal subdomain (Figure 4.8, A and B). The three aromatic side chains (F53, F57 and F64) face one another to form an aromatic core of the C-terminal subdomain also positioned closely to the subdomains' interface (Figure 4.8, C). This positioning of the three Phe side chains is supported by 28 NOE distance restraints.

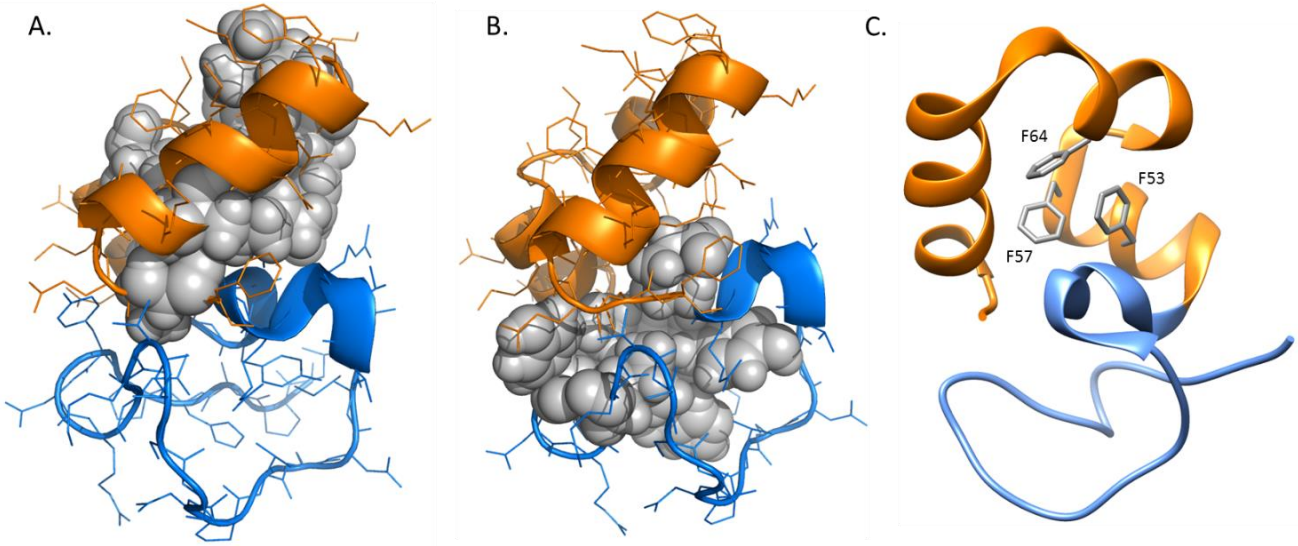


Figure 4.8. Hydrophobic packing of the V4HP63 NMR solution structure. (A) C-terminal subdomain (B) N-terminal subdomain. Residues that make up the hydrophobic core are space filled, while all others are depicted in wire. (C) The aromatic core within the C-terminal subdomain.

Structural analysis showed that V4HP63 contained additional headpiece characteristics: a buried salt bridge at the subdomain interface, and the conserved V-loop. Recall, both VHP and SVHP subdomains are stabilized by an intra-domain salt bridge, while DHP contains a salt "cluster" (see Introduction). Similar to VHP and SVHP, this structure suggests V4HP63 contains a E45-K76 salt bridge between the N- and C-terminal subdomains (Figure 4.9). The E45 O^{ε} to K76 H^{ζ} distance ranges within 2.0-3.5 Å for 8 models out of our 10 lowest energy structures similar to VHP[10]. The other two models show a different E45 side chain orientation with the E45 O^{ε} and K76 H^{ζ} distance ranging within 4.1-5.5 Å. Interestingly, the V4HP63 salt bridge is significantly less solvent exposed than in VHP (6 $Å^2$ and 21 $Å^2$, respectively) as measured by DSSP.

The full solvent exposed surface area is highly similar for both V4HP63 at 4636 $Å^2$ compared to VHP at 4642 $Å^2$.

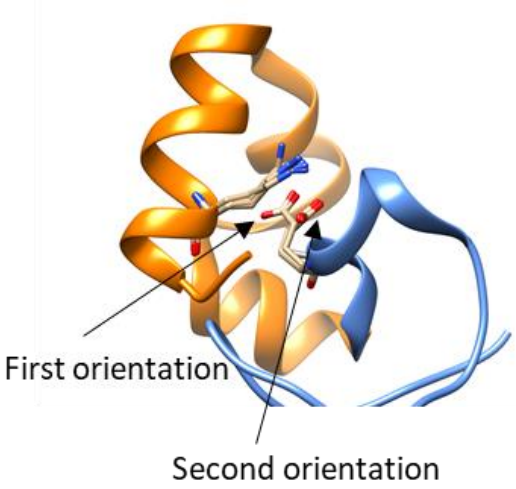


Figure 4.9. Intra-domain salt bridge formed from E45 and K76. Two orientations were seen within the 10 models.

Analysis of the N-terminal subdomain reveals a V-loop that is conserved in all other characterized headpiece domains. In order to identify this loop, modifications to the standard V-loop sequence identifier were observed. The first, rather than an initial L/P in the N-terminal flanking sequence (see Introduction), a histidine was observed for V4HP63. The second is a serine insertion in the following C-terminal flanking sequence, the (L/P) X S (G/D), unlike any other characterized headpiece domain. In fact, villins 1-5 seen in *A. thaliana* contain either a serine or a threonine insertion as this position while the flanking histidine is only seen in villin 4. Using the sequence criterion with these two modifications, this variable loop of V4HP includes residues $K_{29}T_{30}T_{31}S_{32}T_{33}D_{34}$. (Figure 4.10)

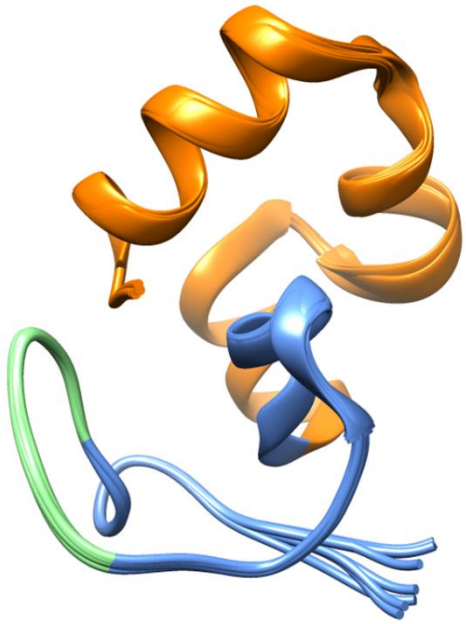


Figure 4.10. The conserved V-loop in V4HP63 depicted in green.

Backbone alignment of V4HP63 to VHP and DHP shows the highly similar folds between these villin headpiece domains. Overall V4HP63 has an RMSD of 2.6 Å and 2.7 Å across all 63 residues for VHP and DHP respectively. The alignment of V4HP63 and SVHP is omitted as the overlap RMSD (6.1 Å) suggest non-homologous structures. The C-terminal subdomains are nearly superimposable, with all secondary structural element overlapped for the three considered headpiece domains (RMSD of 1.0 Å and 1.2 Å for VHP and DHP respectively) (Figure 4.11). However, the N-terminal subdomain, composed mainly of loops and turns, shows the most variability between the structures. Specifically, the length of Helix D is longer in V4HP63 than in VHP, and appears more bent than helix D of DHP (Figure 4.11, B). Inspection of the general fold of the N-terminal subdomains suggests V4HP63 is more similar to that of DHP rather than VHP. In conclusion, V4HP63 possesses of all the distinguishing characteristics found in other characterized headpiece domains.

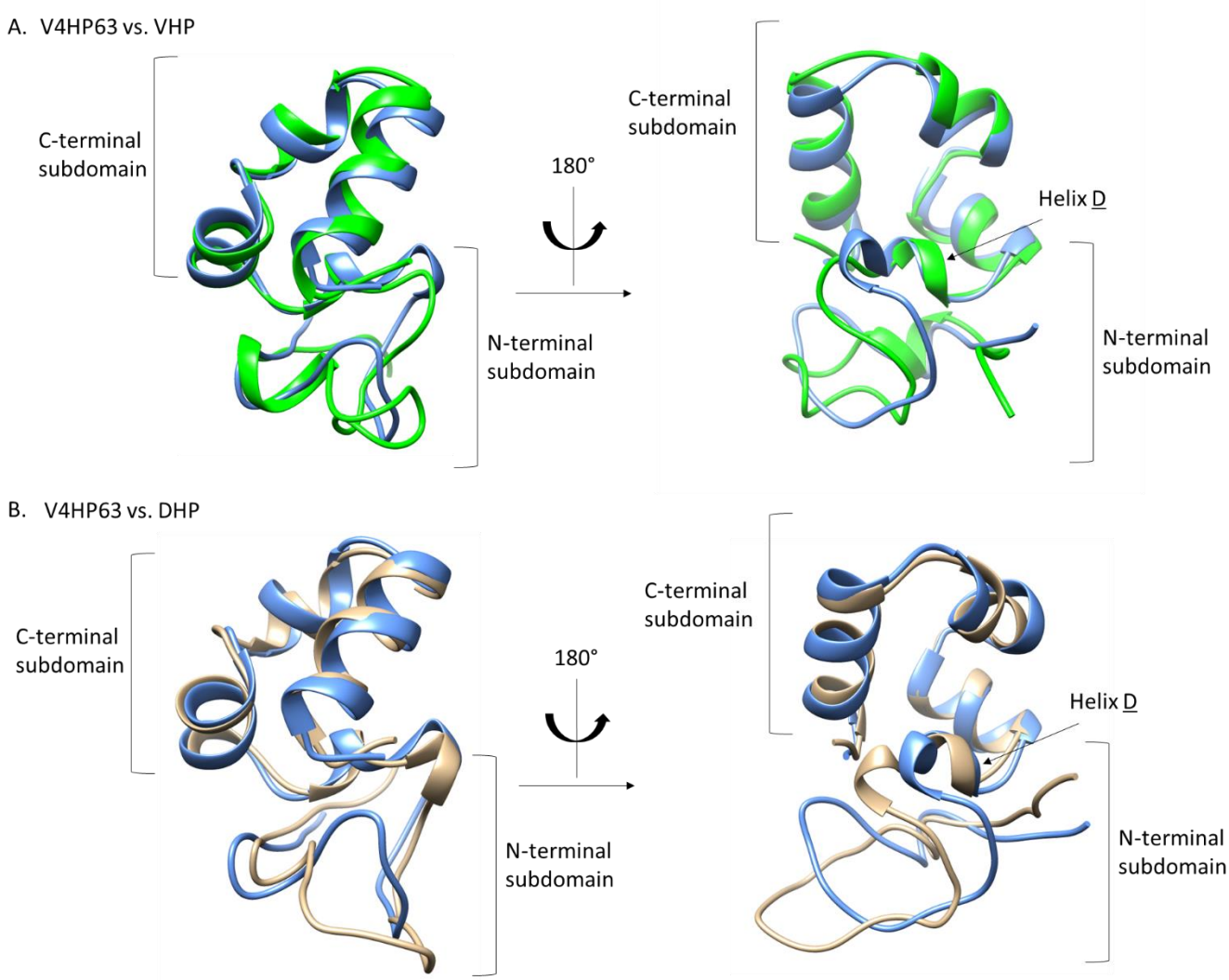


Figure 4.11. Structure comparison of (A) V4HP63 (PDB ID: 5VNT, blue) and VHP (PDB ID: 1QQV, green) and (B) V4HP63 (blue) and DHP (PDB ID: 1QZP, tan).

Backbone dynamics of atVHP76 shows novel mobility features within the N-terminal subdomain

Solution NMR backbone dynamics of atVHP76 was probed by the R1 (as 1/T1) and R2 (as 1/T2) relaxation rates and {^{1}H-^{15}N}-NOE measurements (shown as T1/T2 ratio for this work). It is evident in Figure 4.12 that the first ~20 residues of atVHP76 are unprotected from backbone amide ^{1}H exchange with the solvent. This along with dynamics data, further supports a fully solvent exposed linker region.

The T1/T2 ratio, a representation of internal backbone mobility, demonstrates higher values (≥4.4) for residues 25-82 and lower values (≤2.6) for residues 10-20 (Figure 4.12). This is consistent with structural analysis: a folded domain (residues 20-82) and a disordered linker fragment (1-20).

The {^{1}H-^{15}N}-NOE ratio also indicates a folded polypeptide for residues 22-82 (the NOE ratio > 0.58) and disordered for residues 10-19 (<0.20 and negative NOE ratio values; Figure 4.12).

The predicted V-loop region (residues 29-34) does not demonstrate any outstanding T1/T2, and {^{1}H-^{15}N}-NOE ratio values, which indicates similar internal mobility of the region with respect to the rest of the V4HP63 domain.

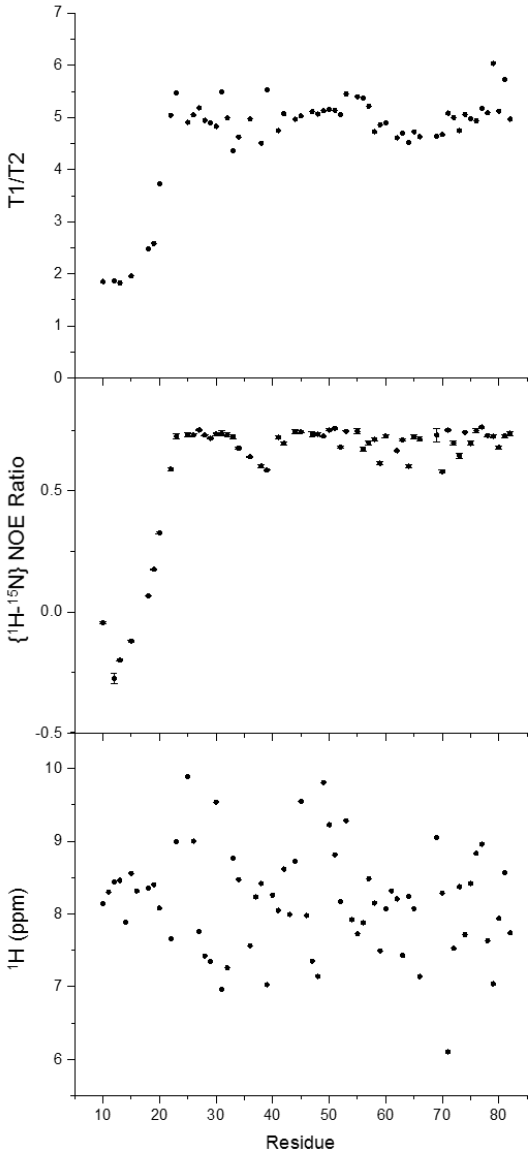


Figure 4.12. Relaxation rates, {^{1}H-^{15}N} NOE ratios and ^{1}HN chemical shift absolute values for atVHP76.

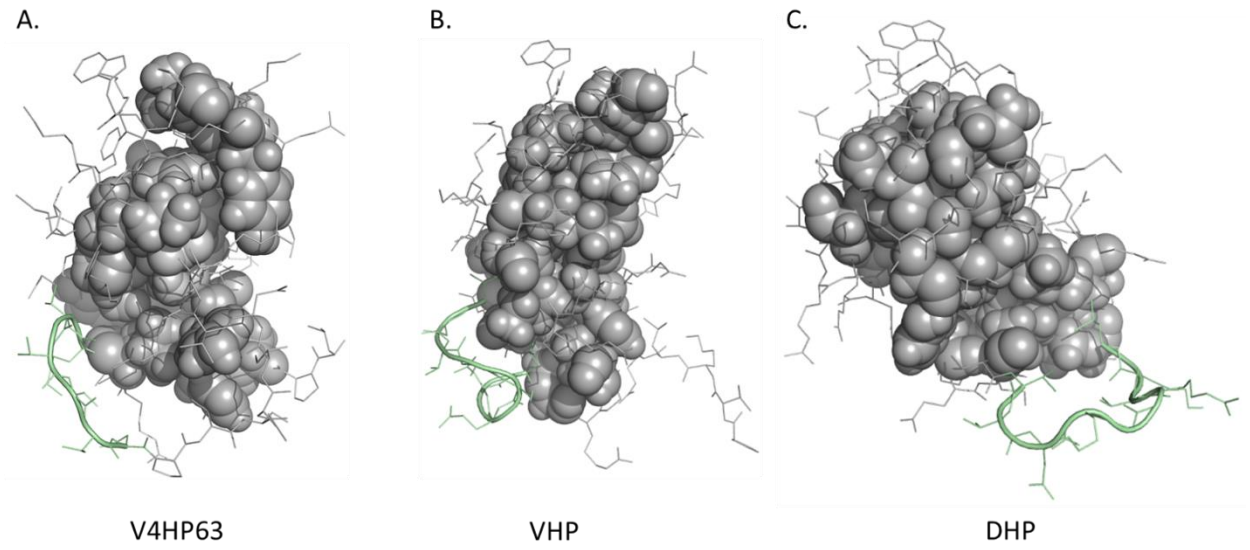


Figure 4.13. Surface representation of the hydrophobic core for (A) V4HP63, (B) VHP (PDB ID: 1QQV) and (C) DHP (1QZP). The conserved V-loop is highlighted in green.

It is suggested by Vardar *et. al.* that the composition or length of the V-loop region shows no specific implications on the fold of headpiece domains, but is essential in supporting a proper N-terminal fold. Flanking both the N- and C-terminal regions of the V-loop in both VHP and DHP are hydrophobic residues that are within the hydrophobic core of these domains (Figure 4.13). Therefore, the V-loop may act as a connector to bridge these two hydrophobic regions for proper N-terminal folding.

Upon closer inspection, it is apparent that the V-loop in V4HP63 protrudes out into the solvent and does not loop back into the hydrophobic core of the headpiece (Figure 4.13). This is the first V-loop identified that does not act as a bridge between flanking hydrophobic residues. In fact, this V-loop continues into a larger loop that connects back to the core domain ten residues later. The five-residue region C-terminal to the predicted V-loop has decreased $\{^{1}H\text{-}^{15}N\}$-NOE ratios relative to the adjacent region, indicative of elevated mobility (Figure 4.12). We refer to this stretch ($P_{35}V_{36}S_{37}D_{38}I_{39}$) as a the high-mobility loop (HM-loop) in order to distinguish it from the canonical vertebrate villin V-loop (Figure 4.14). While only the

HM-loop shows any marked mobility of this region within the N-terminal subdomain, it is evident that this loop and the V-loop are not contributing to a proper hydrophobic core orientation.

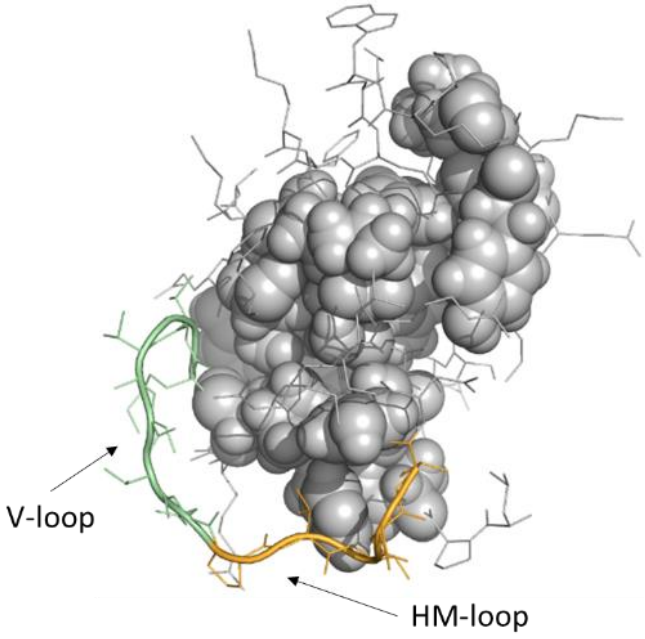


Figure 4.14. Surface representation of the hydrophobic core of V4HP63 with the V-loop (green) and HM-loop highlighted (orange).

Molecular dynamics of V4HP63 show greater amplitudes within the N-terminal subdomain

Unrestrained molecular dynamics simulation of V4HP63 was employed by a current graduate student, David Gruber, to investigate residue specific motions to corroborate our NMR dynamics data. The simulation was run for 1 μs in explicit solvent using the V4HP63 NMR structure as a starting conformation. Analysis of the molecular dynamics (MD) trajectory is reported as RMSD of the entire fragment over the time-course of the simulation. These simulations indicate that V4HP63 is structurally stable and remains close to the solution NMR structure as shown by Figure 4.15, A. Comparison of the RMSD between the N- and C-terminal domains demonstrate a more stable C-terminal subdomain (RMSD 0.9 ± 0.13 Å) than the N-terminal subdomain (RMSD 2.63 ± 0.24 Å) (Figure 4.15, B).

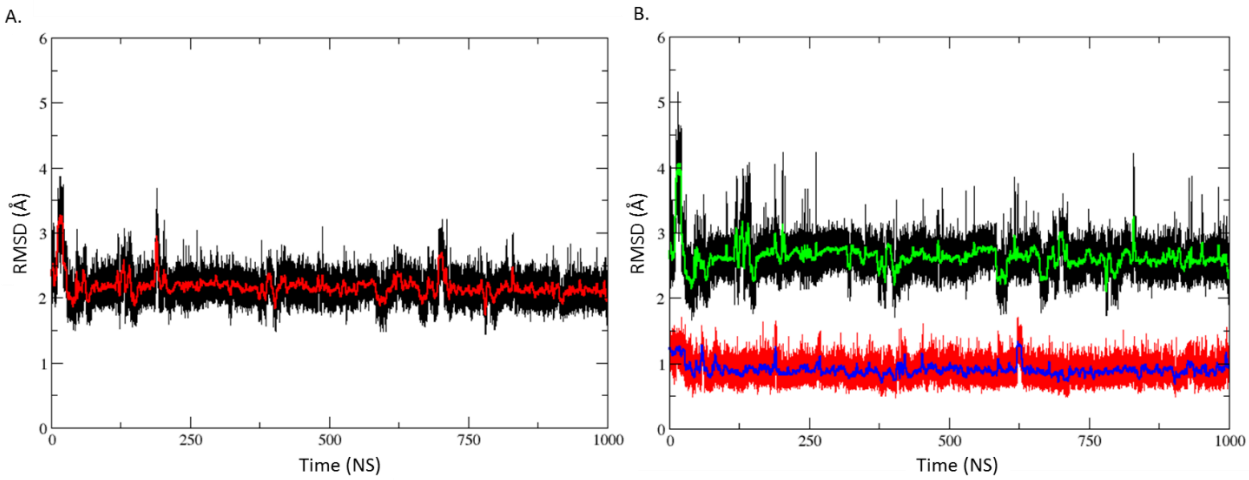


Figure 4.15. RMSD vs. time for the molecular dynamics simulation of V4HP63. A) RMSD for the entire V4HP63 domain. B) RMSD for the N-terminal subdomain (top) and the C-terminal subdomain (bottom). Bright middle lines are running averages over 2 ns.

Principal component analysis (PCA) allowed us to identify patterns in backbone motions contained within the most dominant modes. Figure 4.16, A, shows a scree plot for the proportion of variance that each resulting principal component contributes to the overall motion. The five most dominant modes combined account for ~50% of the overall backbone structural variability. We visualized the motion of the most dominant principle components using a porcupine plot, where an arrow representing the direction and amplitude of its motion accompanies each atom (Figure 4.16, B). The most dominant mode (eigenvalue rank 1) showed increased structural flexibility at the N-terminus and within residues 33-36, shared between the V-loop and HM-loop.

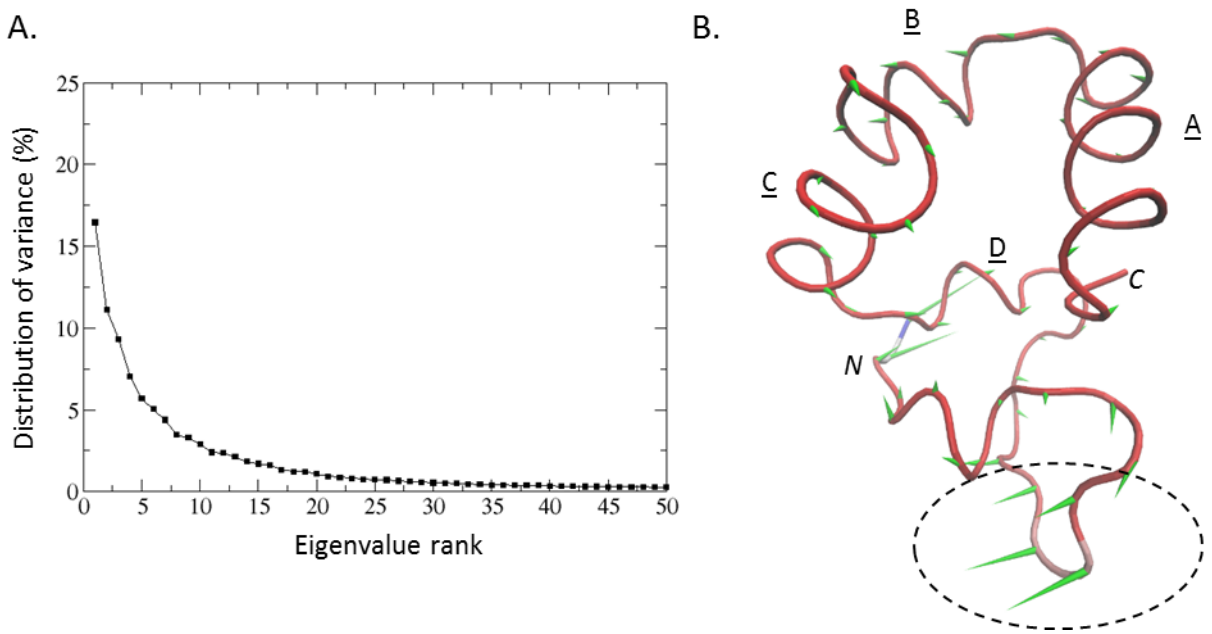


Figure 4.16. A) Ranked eigenvalues (weights) of principle components extracted from the MD simulation of V4HP63. B) The porcupine plot representing motions of the principal component 1 for V4HP63. The high-amplitude, coordinated motions of the joined V-loop and HM-loop are represented with long arrows within the dashed oval. The other high-amplitude motions occur at the N-terminus of the fragment. The helices A, B, C, and D are labeled for reference.

Backbone fold of atVHP60 is predicted to be highly similar to V4HP63

The atVHP60 construct, which lacks 16 N-terminal residues of atVHP76 (thus, three N-terminal residues of V4HP63), shows a ^{15}N-HSQC cross-peak distribution highly similar to that of atVHP76 for the same residues (Figure 4.17). The ^{1}H and ^{15}N chemical shift values for 47 out of 60 residues in atVHP60 differ from those in atVHP76 by less than the threshold values associated with noticeable structural perturbations: 0.05 ppm for ^{1}H and 0.2 ppm for ^{15}N dimensions.

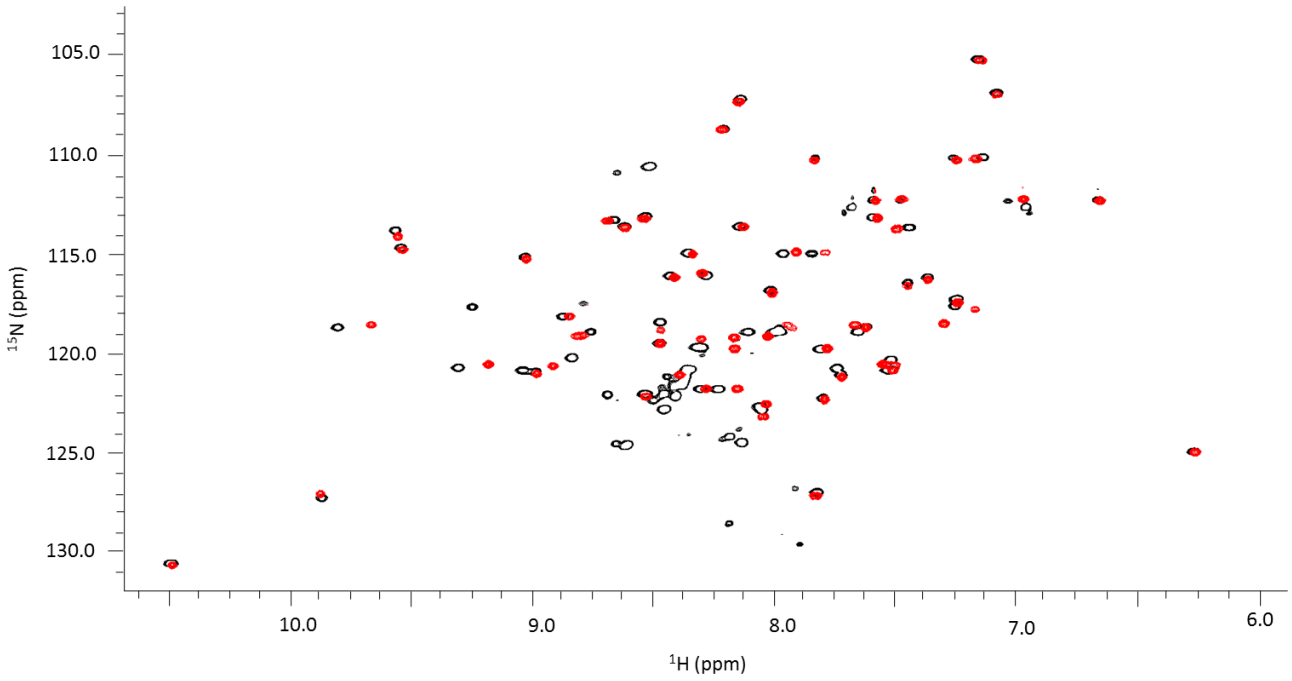


Figure 4.17. ^{15}N HSQC of atVHP76 (black) and atVHP60 (red) taken on a 500 MHz Bruker Avance III HD spectrometer equipped with a room temperature probe at 25°C.

Out of 13 residues whose cross-peaks moved by greater values than the thresholds, 8 are located in the N-terminal subdomain (Y25, R27, V36, D38, R44, E45, A46, Y47) and five are located in the C-terminal subdomain (S49, E52, F53, K56, A63) (Figure 4.18). Four residues in atVHP60 were unassigned: S37, D40, S60, and K69. However, the assignment of atVHP60 was done by comparison to an assigned atVHP76 spectrum, thus, there are multiple unassigned peaks in this spectrum as the chemical shift perturbation was too great to be accurately assigned. These peaks are likely candidates for the four missing residues as well as the $(His)_6$-tag.

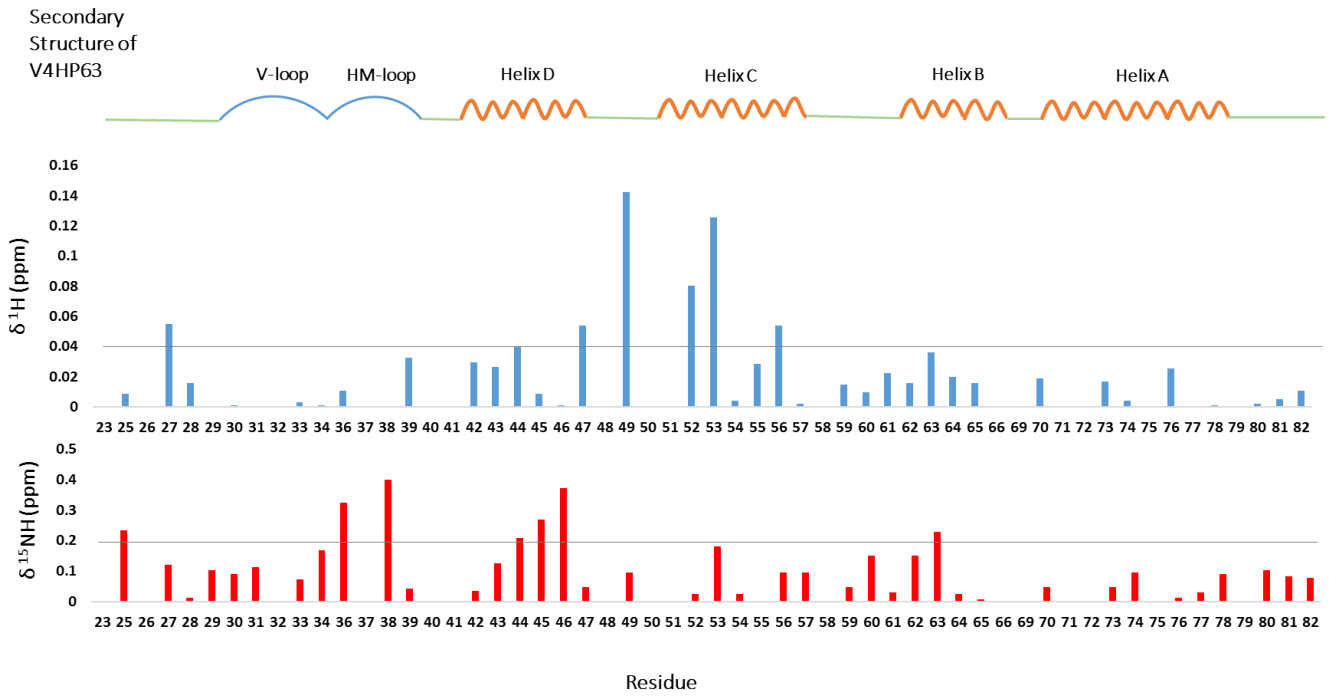


Figure 4.18. Chemical shift differences between atVHP76 and atVHP60. The line depicts the threshold value cut offs of 0.05 ppm for ^{1}H and 0.2 ppm for ^{15}N dimensions. Secondary structure shown with α-helices in orange and loops in blue.

A majority of the residues that show the largest chemical shift perturbation (^{15}N: 0.32-0.40 ppm ^{1}H: 0.08-0.14 ppm) can be rationalized as being located near the N-terminus of atVHP60 (H23), as well as the N-terminal $(His)_6$-tag (Figure 4.19). While the remaining residues (Y25, R44, K56, and A63) are distant from the last residue in atVHP60, the chemical shift perturbations are minimal (^{15}N: 0.21-.23 ppm and ^{1}H: 0.05ppm), indicating minor structural differences (Figure 4.19).

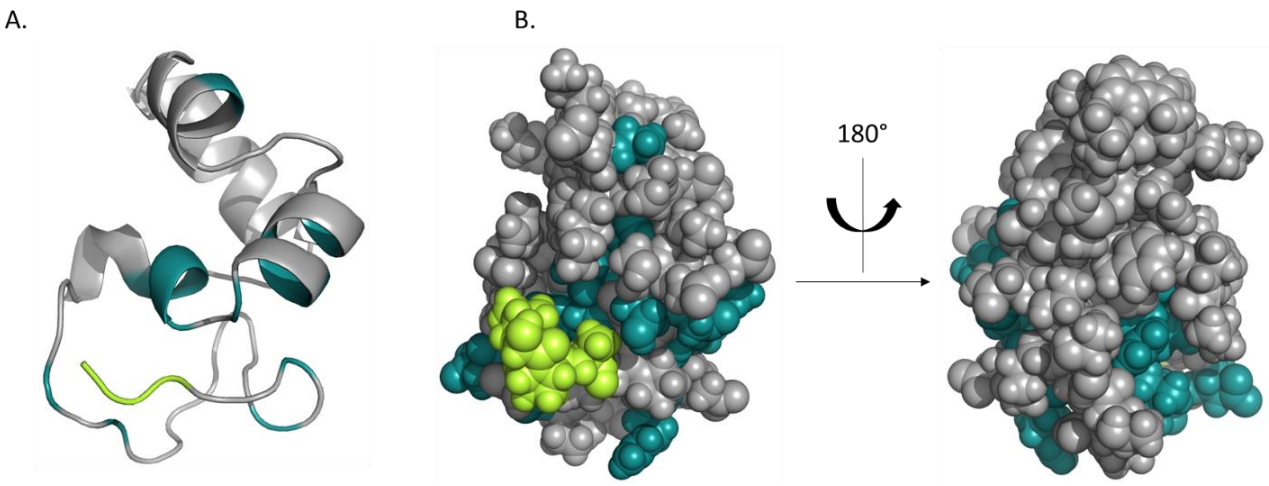


Figure 4.19. (A) Ribbon structure and (B) surface representation of V4HP63 highlighting residues in atVHP60 that showed structural perturbations in blue, and residues that are absent in atVHP60 are shown in green.

The purpose of using both the atVHP76 and atVHP60 designs was to probe the headpiece domain with a short linker (atVHP76) and headpiece in isolation (atVHP60). From these results, we concluded that atVHP60 likely adopts a highly similar solution structure to that of atVHP63, with the majority of structural perturbations to be those around the N-terminus.

atVHP76 showed no detectable sensitivity to calcium, sodium and potassium ions

It was previously reported by Zhang, Y. *et. al.*, that full length villin 4 could bundle and sever actin filaments in a calcium-dependent manner. Due to the variability of F-actin binding site locations and allosteric ion binding sites throughout villin-like proteins (see Introduction), we wanted to test the atVHP76 headpiece for ion binding sites. Calcium, as well as, sodium and potassium are known regulators in plant cells and, as such, their interactions with atVHP76 were studied[42]. Collecting a series of ^{15}N-HSQC spectra at varying ion concentrations should identify putative ion binding residues by changes in chemical shift at varying salt concentrations. The atVHP76 construct showed no sensitivity to the three types of ions within the concentration ranges tested: Ca^{2+} (0 - 5 mM) (Figure not shown), Na^{+} (0 - 300 mM) (Figure 4.20) and K^{+} (0 – 100 mM) (Figure 4.21). In these trials, ^{15}N-HSQC resonance line movements by more than 0.04 ppm in

^{1}H and 0.2 ppm in ^{15}N dimensions were not detected. Therefore, the data presented here suggests the headpiece domain in villin 4 contains no ion binding site for calcium, potassium or sodium.

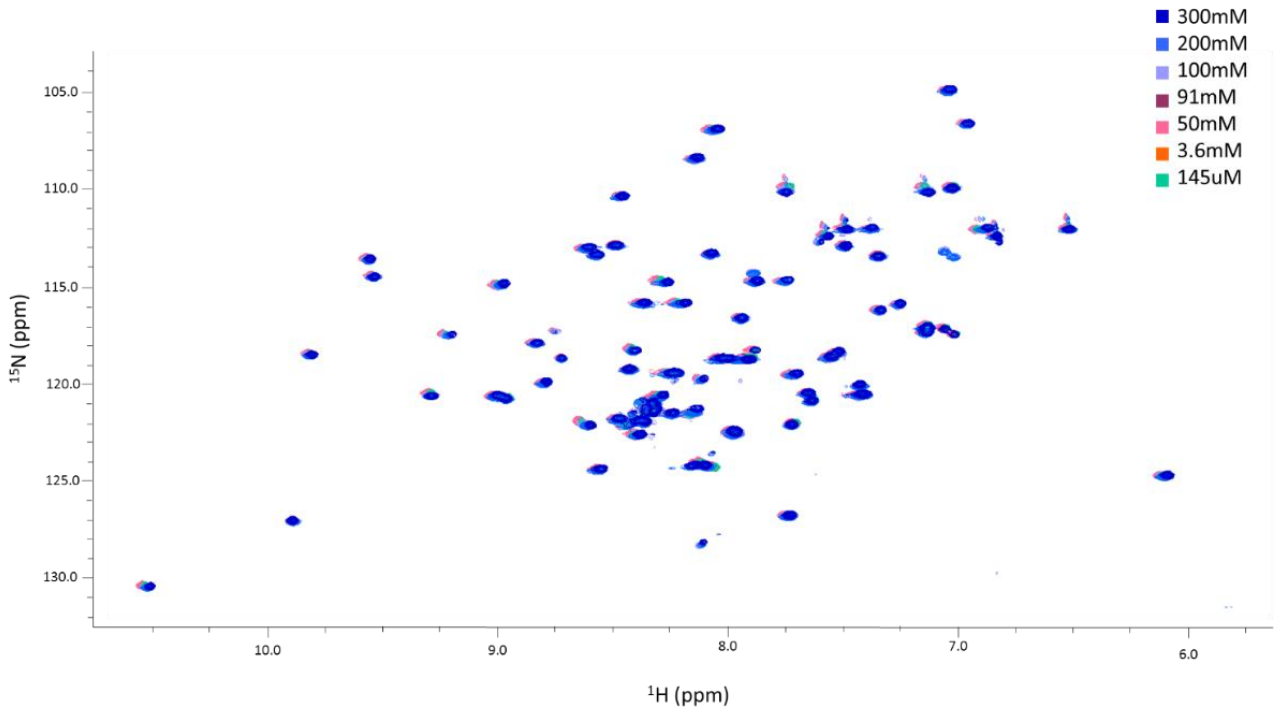


Figure 4.20. ^{15}N HSQC of atVHP76 with varying sodium ion concentrations taken on a 500 MHz Bruker Avance III HD spectrometer equipped with a room temperature probe at 25°C.

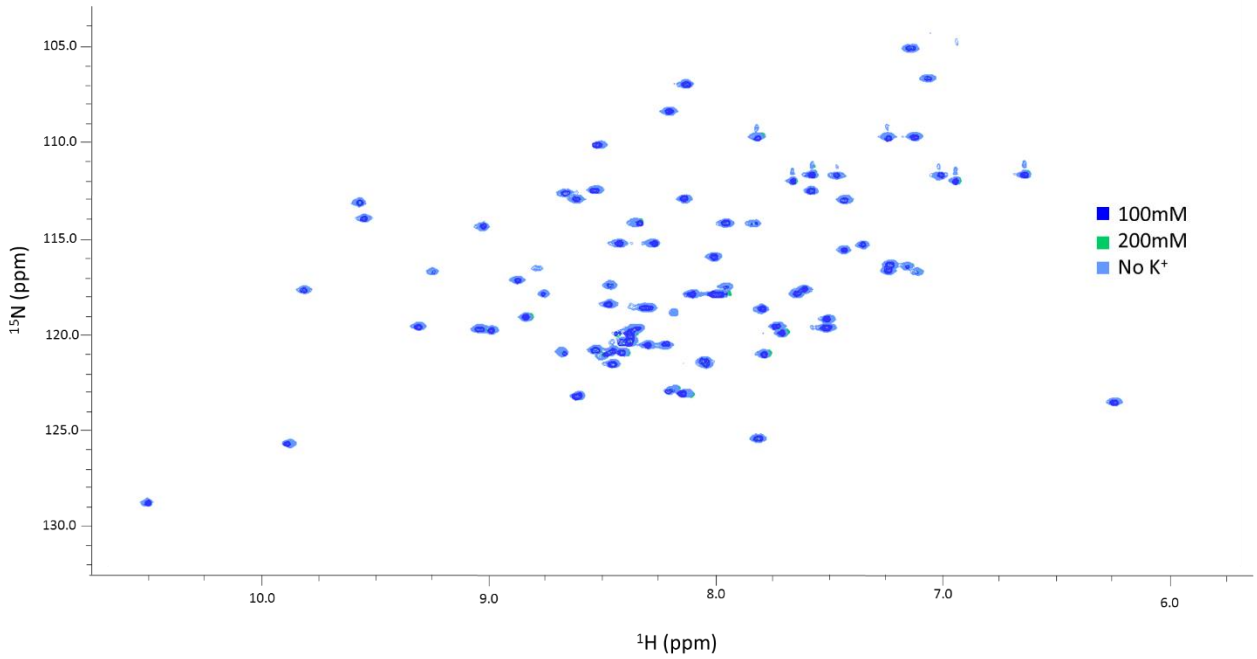


Figure 4.21. ^{15}N HSQC of atVHP76 with varying potassium ion concentrations taken on a 500 MHz Bruker Avance III HD spectrometer equipped with a room temperature probe at 25°C.

atVHP76 and atVHP60 display strong specific binding to F-actin

Recall, that a full-length villin 4 construct was shown to bind to F-actin with a K_d of 0.24 ± 0.05 µM[42]. In order to determine if an F-actin binding site was located on the headpiece domain, like the vertebrate analogs, we employed a standard pull-down assay. Total headpiece concentration was varied from 2 µM to 120 µM for atVHP76 and 5 µM to 120 µM for atVHP60. Each data point was collected in triplicate. Pellets from the pull-down were first washed with F-actin buffer and subsequently suspended in 7% acetic acid solution for analysis by reverse-phase liquid chromatography (see Materials and Methods). Concentrations of bound headpiece were determined by integration of chromatogram peak areas monitored at 280 nm and a calibration curve created using isolated headpiece (Appendix Figure 2). Replicate trials were averaged and fitting was performed as described above: both specific (hyperbolic) and non-specific (linear) modes of binding were used.

The binding curve for atVHP76 was fit to a hyperbolic model, indicating specific binding to F-actin (Figure 4.22). The dissociation constant (K_d) was determined to be 3.0 ± 0.7 µM for the atVHP76 construct. The extent of the non-specific component (NS) of binding was measured to be 4.0 ± 0.8%. The saturation binding concentration, B_{max}, was determined to be 14.1 ± 0.7 µM. The atVHP60 construct shows a similar binding specificity (K_d of 2.0 ± 1.0 µM) and low non-specific binding (NS = 0.0%). B_{max} was determined to be 19.0±1.5 µM (Figure 4.22).

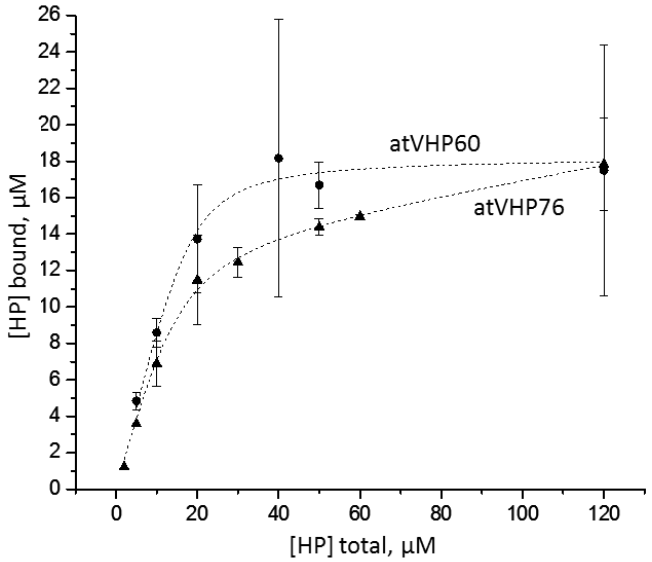


Figure 4.22. F-actin binding by atVHP76 (triangles) and atVHP60 (circles). Fitting and plotting was done with Origin 6.0.

Both the atVHP60 and atVHP75 K_d values overlap, indicating there is no significant difference in F-actin binding affinity between the folded headpiece domain (atVHP60) compared to the headpiece with 16 residues from the linker (atVHP76). There is a small percentage of non-specific binding (4.0 ± 0.8%) for atVHP76, indicating some trapping probably due to the first 16 unstructured residues. It can be concluded from B_{max} that there is only one binding site within the villin-4 headpiece domain as this value

is approximately equal to the overall actin concentration (17 µM), indicating 1:1 binding of atVHP76 or atVHP60 to actin. Furthermore, these values indicate that the N-terminal 14 residues next to the headpiece do not disrupt the binding site.

The binding affinity of villin-4 headpiece is consistent with a high-affinity headpiece domain. Recall, that the K_d for VHP and DHP is 2.9 ± 0.5 µM and 3.3 ± 1.4 µM respectively. It can be concluded that atVHP fits into the vertebrate family definition of high-affinity villins and, in conjunction with the solution structure, can help expand the investigation of the headpiece binding interface.

Solution structure of V4HP63 helps expand the criteria for high-affinity F-actin binding

One fundamental motivation for this study was a lack in confidence for predicting the capacity of villin 4 headpiece to bind F-actin. This difficulty arose from the observation that, when compared to vertebrate villin headpieces (VHP, DHP, and SVHP), atV4HP displays dramatically different surface charge and hydrophobic distribution patterns in the primary structure. The solution NMR structure of V4HP63 corroborates these patterns, therefore the work done here may help to expand the criteria for high affinity binding.

Our structural data indicate that the secondary and tertiary structures of V4HP63 are highly similar to those of its vertebrate relatives in chicken villin, supervillin and dematin. However, structural similarity is highly uneven between the two subdomains. In the C-terminal subdomain, nearly the entire backbone conforms to a structure that is isomorphic to VHP, DHP, and SVHP (Table 3). Additional features of the vertebrate C-terminal subdomains are conserved, including the presence and location of the tryptophan cap, conserved aromatic core, and a buried salt bridge. To the contrary, less than half of the N-terminal subdomain backbone residues are structurally homologous to VHP or DHP (Table 3). This structural diversity within the N-terminal subdomain of V4HP63 is further highlighted by the five residue HM-loop, C-terminal to the V-loop, not present in analogous regions of VHP and DHP. Other differences are seen

from sequence alignment of the residues that are central to F-actin binding in VHP and DHP, which readily shows that these residues in V4HP63 often have different charged states. In spite of these differences, atVHP76 and atVHP60 bind F-actin specifically with approximately the same affinity as their counterparts in chicken villin and human dematin. This finding allows us to expand the set of criteria for high-affinity F-actin binding by headpiece domains.

Table 3. Sequence similarity to V4HP63 and its C- and N- subdomains to headpiece sequences.

Headpiece	RMSD (Cα) vs. V4HP63*	RMSD (Cα) vs. C-terminal subdomain of V4HP63	RMSD (Cα) vs. N-terminal subdomain of V4HP63
VHP	1.3 Å (41 Cα pairs)	1.0 Å (30 Cα pairs)	1.3 Å (14 Cα pairs)
DHP	1.3 Å (48 Cα pairs)	1.2 Å (34 Cα pairs)	1.3 Å (15 Cα pairs)
SVHP	1.2 Å (30 Cα pairs)	1.2 Å (30 Cα pairs)	Below detection**

*According to automatic RMSD Chimera analysis
**Insufficient coverage of the structural match: only 5 Cα pairs selected, no structural similarity

The currently accepted criteria for specific and high-affinity binding to F-actin by headpiece domains are: a) a hydrophobic cap; b) alternating charged "crown" below the cap and; c) a positive patch below the crown[10, 33]. In V4HP63, however, the hydrophobic cap has been reduced to a single tryptophan residue, as a lysine replaced the corresponding leucine of VHP. In addition, the patch/crown tripeptide K71-E72-K73 has been replaced with M-A-V (in V4HP63) and K-K-A (in DHP) (Figure 4.23). Suggesting that the charge "crown" can be reduced to K65, R37, and the C-terminus. Thus, with the addition of V4HP63 to the VHP/DHP high-affinity group, we observed that certain residues listed above are not present and, therefore, may not be required for strong specific F-actin binding. Thus, the binding criteria can be relaxed (Figure 4.24).

```
                                  10          20          30          40          50          60          70      76
Vertebrate Villin HP67             P TKLETFPLDV LVNTAAEDLP RGVDPSRKEN HLSDEDFKAV FGMTRSAFAN LPLWKQQNLK KEKGLF
Dematin HP68                     PG LQIYPYEMLV VTNKGRTKLP PGVDRMRLER HLSAEDFSRV FAMSPEEFGK LALWKRNELK KKASLF
Plant Villin HP76       QEDAKEGVED EEDLPAHPYD RLKTTSTDPV SDIDVTRREA YLSSEEFKEK FGMTKEAFYK LPKWKQNKFK MAVQLF

Supervillin HP67                 M LAKLCKTIYP LADLLARPLP EGVDPLKLEI YLTDEDFEFA LDMTRDEYNA LPAWKQVNLK KAKGLF
Protovillin HP66                   PTTITTFYPL SVLKQKTNLP NDIDKSCLHL YLSDEEFLST FKMTKEIFQK TPAWKTKQLR VDNGLF
```

Figure 4.23. Aligned sequences of the high-affinity actin binding headpieces, followed by the non-actin binding domains. The residue numbering corresponds to that in VHP. Residues implicated in F-actin binding are bolded. Those that are underlined form the charged crown, those left black are the hydrophobic cap, and those colored blue with no underline contribute to the positive patch.

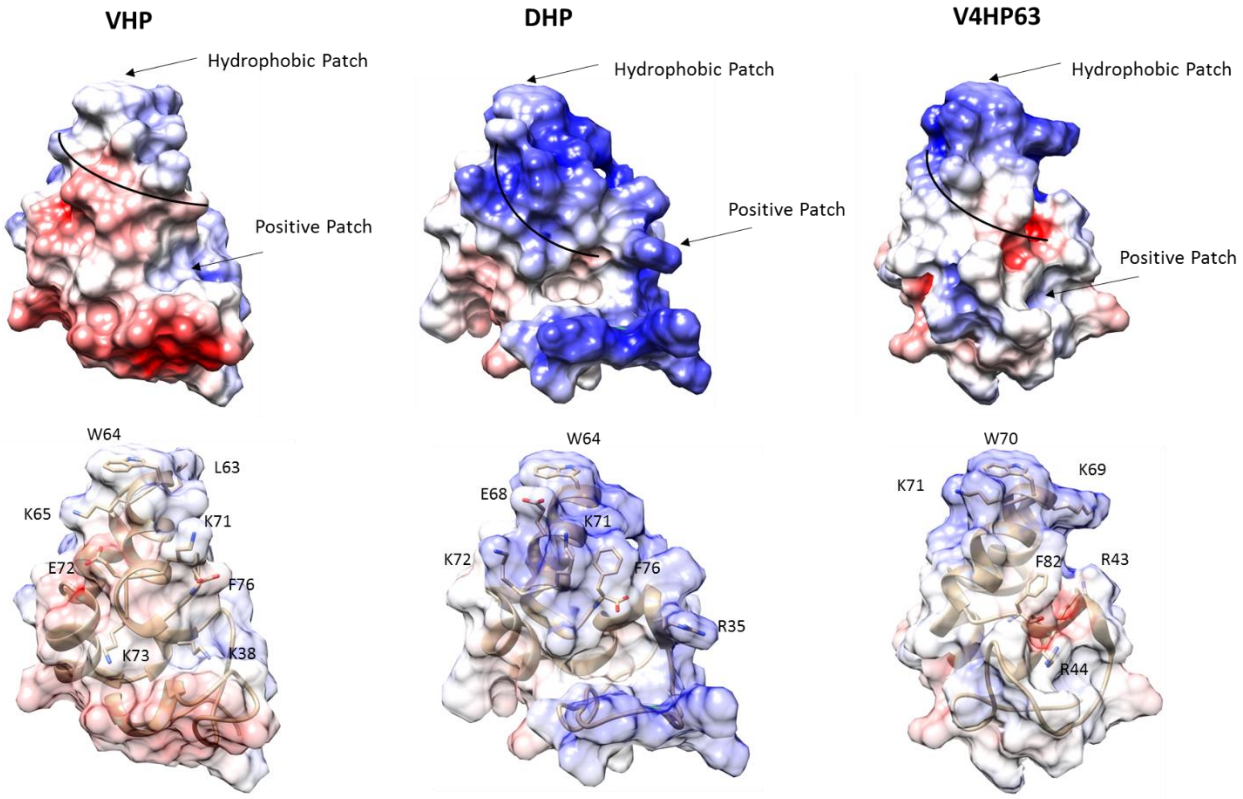


Figure 4.24. Surface charge distributions of the three characterized headpiece domains that show high-specific F-actin binding. The surface charge was calculated and modeled by Chimera with red, blue, white as negative, positive and neutral charge respectively. The three criteria described for high affinity binding are highlighted with arrows and the charged crown is shown with a curved line. The specific residues which contribute to these criteria are labeled in the structures below. The R37 in VHP is hidden for simplicity.

Lessons from supervillin headpiece can also help update the set of criteria for F-actin binding. Like V4HP63, SVHP retains a hydrophobic W64 and some features of a charged crown. However, unlike V4HP63, the positive patch is completely absent and replaced with negatively charged residues, rendering this SVHP "inactive" (Figure 4.25). In fact, Brown *et al.* were able to introduce a positive region with a single point mutation (L38K) that allowed SVHP to bind F-actin with moderate affinity. Likewise, protovillin headpiece has no F-actin binding capability because it lacks residues that would contribute to a positive charge as well as the conserved salt bridge.

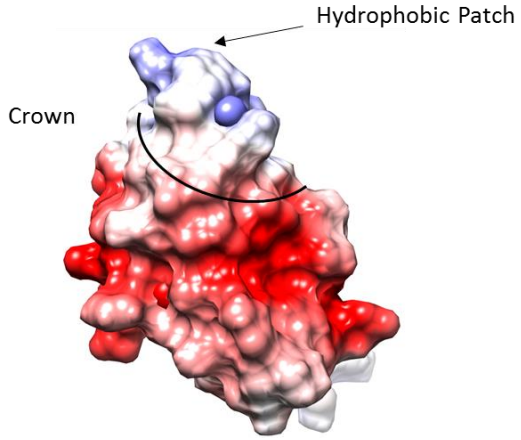


Figure 4.25. Surface charge distribution of SVHP (PDB 2K6N), calculated by Chimera. The charge coloring from red, blue to white is negative, positive to neutral respectively.

It would seem that in addition to updating the criteria for specific F-actin binding by headpiece domains, a weight can also be set on each rule. Lessons from SVHP and protovillin suggest the most critical factor for high-affinity binding is the positive patch. Furthermore, structural analysis of V4HP and DHP indicates the second most critical factor is a charged crown, which we showed previously can be reduced to three specific residues.

All headpiece domains from villin isoforms in *A. thaliana* are predicted to bind with high affinity to F-actin

Utilizing our updated criteria for F-actin binding, we predict that all the headpiece domains of *A. thaliana* should bind actin filaments with high affinity. As this work has shown, the headpiece from villin 4 binds F-actin specifically and strongly. The other four headpiece domains have high sequence similarity to V4HP63, including a high level of conservation within the C-terminal subdomain and a 100% conserved salt-bridge between the N- and C- terminal subdomains (Figure 4.26). The elements required for specific, strong F-actin binding described above are all present in all *A. thaliana* villin headpiece domains.

```
                          20        30        40        50        60        70      76
Plant Villin HP63     LPAHPYD RLKTTSTDPV SDIDVTRREA YLSSEEFKEK FGMTKEAFYK LPKWKQNKFK MAVQLF

Plant Villin V1HP     SLAYSYE QLRVDSQKPV TDIDATRREA YLTEKEFEER FGMAKSEFYA LPKWKQNKLK ISLHLF
Plant Villin V2HP     GATFTYE QLRAKSENPV TGIDFKRREA YLSEEEFQSV FGIEKEAFNN LPRWKQDLLK KKFDLF
Plant Villin V3HP     GVTFTYE RLQAKSEKPV TGIDFKRREA YLSEVEFKTV FGMEKESFYK LPGWKQDLLK KKFNLF
Plant Villin V5HP     LPTFPYE RLKTDSEDPV SDVDLTRREA YLTSVEFKEK FEMTKNEFYK LPKWKQNKLK MSVNLF
```

Figure 4.26. Aligned sequences of headpiece domains in *A. thaliana*. The residue numbering corresponds to that in VHP. Residues implicated in F-actin binding are bolded. Those that are underlined form the charged crown, those left black are the hydrophobic cap, and those colored blue with no underline contribute to the positive patch.

Furthermore, the sequences of villin 1, 2, 3 and 5 headpieces were modeled onto the published structure of V4HP63 (PDB ID: 5VNT) by the Phyre2 web server (Figure 4.27). The Phyre2 web server allows the input of sequences of proteins with no known structure and it then models these sequences onto structures submitted to the PDB with high sequence similarity[63]. Based on the surface charge representations of these models, the plant villin homologs are predicted to include a hydrophobic cap, charged crown, and positive patch formed by R38 (R44 of the submitted PDB) (Figure 4.27).

To test this prediction of appropriate surface charge distribution a further step of modeling the plant villin headpiece sequences onto V4HP63 was performed. The villin 1,2,3 and 5 headpiece sequences were all predicted by Phyre2 to contain the highest sequence homology to our structure of V4HP63 and was then subsequently modeled onto this structure (Figure 4.27). It is clear from these models that the surface charge representations do contain a hydrophobic cap, charged crown and positive patch formed by R44 in V4HP63 residue numbering (or R38 in VHP residue numbering).

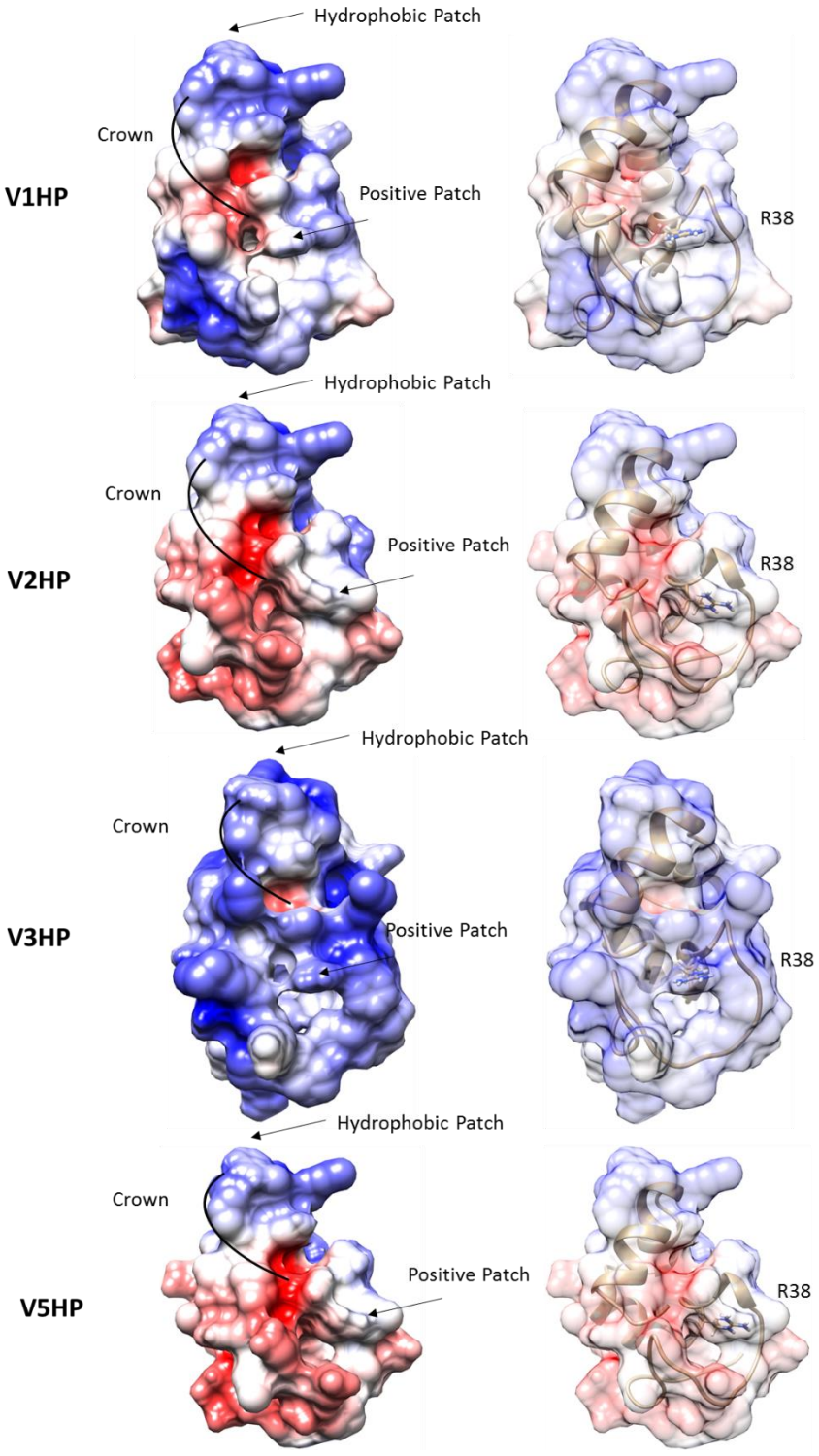


Figure 4.27. Phyre2 model predictions of headpiece domains from villin isoforms 1,2,3 and 5 in *A. thaliana.* The surface charge was calculated and modeled by Chimera with red, blue, white as negative, positive and neutral charge respectively.

Chapter V: Conclusions

The work presented here is the first biochemical characterization of a non-vertebrate villin headpiece. As we expected, this domain of *A. thaliana* villin 4 adopts a fold that is canonical to those found in vertebrate villins. However, unique to villin 4, the N-terminal subdomain fold contains a V-loop flanked by a distinct HM-loop, not observed in vertebrate villins. With this work, we demonstrate that the structure and activity of villin-type headpiece domains are conserved within vertebrate actin regulatory family of villins and in plants.

On the basis of the presented structural and functional data we suggest a model for F-actin binding that differs from the previously proposed model[10,33]. We define the most critical factor for F-actin binding is a positive patch, primarily from residue R44, homologous to K38 in VHP. While, the C-terminal subdomain contains the tryptophan cap seen in all headpieces, we did not find the residues consistent with the "charged crown", that previously was thought to be required for high-affinity binding.

This work supports the ongoing effort in understanding a residue-specific binding interface between headpiece domains and F-actin that is currently not possible to study due to the dynamic nature of F-actin. Not only is this interaction of interest for fundamental science, it provides insight to the working mechanics of F-actin regulation within the cell. It is difficult to propose modifications to low-affinity headpiece domains in the effort to increase binding without knowledge of the required residues. This work narrows down the residues that are critical for high-affinity F-actin binding.

Actin is highly homologous between plants, humans and yeast and is considered one of the most evolutionary conserved proteins within eukaryotes[3]. In fact, it is hypothesized that this conservation was driven by the idea that actin is a hub for many actin binding partners, and therefore has little chance to evolve without compromising one or more interactor, or biological process[3]. From this work, we can

hypothesize that the villins from all eukaryotes also evolved with this evolutionary constraint, which conserved F-actin binding, fold and affinity.

References

1. Wickstead, B., and Gull, K. (2011) The evolution of the cytoskeleton, *J Cell Biol 194*, 513-525.
2. Lemiere, J., Valentino, F., Campillo, C., and Sykes, C. (2016) How cellular membrane properties are affected by the actin cytoskeleton, *Biochimie 130*, 33-40.
3. Gunning, P. W., Ghoshdastider, U., Whitaker, S., Popp, D., and Robinson, R. C. (2015) The evolution of compositionally and functionally distinct actin filaments, *J Cell Sci 128*, 2009-2019.
4. Suarez, C., and Kovar, D. R. (2016) Internetwork competition for monomers governs actin cytoskeleton organization, *Nat Rev Mol Cell Biol 17*, 799-810.
5. dos Remedios, C. G., Chhabra, D., Kekic, M., Dedova, I. V., Tsubakihara, M., Berry, D. A., and Nosworthy, N. J. (2003) Actin binding proteins: regulation of cytoskeletal microfilaments, *Physiological reviews 83*, 433-473.
6. Bretscher, A., and Weber, K. (1980) Villin is a major protein of the microvillus cytoskeleton which binds both G and F actin in a calcium-dependent manner, *Cell 20*, 839-847.
7. Nag, S., Larsson, M., Robinson, R. C., and Burtnick, L. D. (2013) Gelsolin: the tail of a molecular gymnast, *Cytoskeleton (Hoboken) 70*, 360-384.
8. Khurana, S., and George, S. P. (2008) Regulation of cell structure and function by actin-binding proteins: villin's perspective, *FEBS Lett 582*, 2128-2139.
9. Kumar, N., Zhao, P., Tomar, A., Galea, C. A., and Khurana, S. (2004) Association of villin with phosphatidylinositol 4,5-bisphosphate regulates the actin cytoskeleton, *J Biol Chem 279*, 3096-3110.
10. Vardar, D., Buckley, D. A., Frank, B. S., and McKnight, C. J. (1999) NMR structure of an F-actin-binding "headpiece" motif from villin, *J Mol Biol 294*, 1299-1310.
11. Vermeulen, W., Vanhaesebrouck, P., Van Troys, M., Verschueren, M., Fant, F., Goethals, M., Ampe, C., Martins, J. C., and Borremans, F. A. (2004) Solution structures of the C-terminal headpiece subdomains of human villin and advillin, evaluation of headpiece F-actin-binding requirements, *Protein Sci 13*, 1276-1287.
12. Smirnov, S. L., Isern, N. G., Jiang, Z. G., Hoyt, D. W., and McKnight, C. J. (2007) The Isolated Sixth Gelsolin Repeat and Headpiece Domain of Villin Bundle F-Actin in the Presence of Calcium and Are Linked by a 40-Residue Unstructured Sequence, *Biochemistry 46*, 7488-7496.
13. Lange, K. (2011) Fundamental role of microvilli in the main functions of differentiated cells: Outline of an universal regulating and signaling system at the cell periphery, *J Cell Physiol 226*, 896-927.
14. Friederich, E., Vancompernolle, K., Louvard, D., and Vandekerckhove, J. (1999) Villin function in the organization of the actin cytoskeleton. Correlation of in vivo effects to its biochemical activities in vitro, *J Biol Chem 274*, 26751-26760.
15. George, S. P., Wang, Y., Mathew, S., Srinivasan, K., and Khurana, S. (2007) Dimerization and actin-bundling properties of villin and its role in the assembly of epithelial cell brush borders, *J Biol Chem 282*, 26528-26541.
16. Finidori, J., Friederich, E., Kwiatkowski, D. J., and Louvard, D. (1992) In vivo analysis of functional domains from villin and gelsolin, *J Cell Biol 116*, 1145-1155.
17. Pope, B., Way, M., Matsudaira, P. T., and Weeds, A. (1994) Characterisation of the F-actin binding domains of villin: classification of F-actin binding proteins into two groups according to their binding sites on actin, *FEBS Lett 338*, 58-62.
18. Pestonjamasp, K. N., Pope, R. K., Wulfkuhle, J. D., and Luna, E. J. (1997) Supervillin (p205): A novel membrane-associated, F-actin-binding protein in the villin/gelsolin superfamily, *J Cell Biol 139*, 1255-1269.

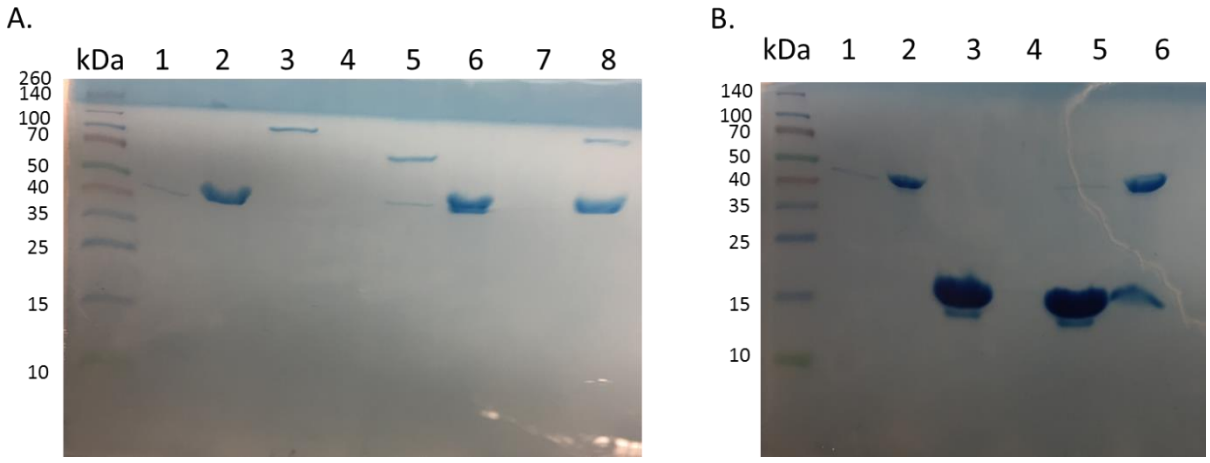
19. Pope, R. K., Pestonjamasp, K. N., Smith, K. P., Wulfkuhle, J. D., Strassel, C. P., Lawrence, J. B., and Luna, E. J. (1998) Cloning, characterization, and chromosomal localization of human supervillin (SVIL), *Genomics 52*, 342-351.
20. Smith, T. C., Fang, Z., and Luna, E. J. (2010) Novel interactors and a role for supervillin in early cytokinesis, *Cytoskeleton (Hoboken) 67*, 346-364.
21. Fang, Z., and Luna, E. J. (2013) Supervillin-mediated suppression of p53 protein enhances cell survival, *J Biol Chem 288*, 7918-7929.
22. Smith, T. C., Fridy, P. C., Li, Y., Basil, S., Arjun, S., Friesen, R. M., Leszyk, J., Chait, B. T., Rout, M. P., and Luna, E. J. (2013) Supervillin binding to myosin II and synergism with anillin are required for cytokinesis, *Mol Biol Cell 24*, 3603-3619.
23. Fedechkin, S. O., Brockerman, J., Luna, E. J., Lobanov, M. Y., Galzitskaya, O. V., and Smirnov, S. L. (2013) 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 31(10)*, 1150-1159.
24. Bogatyreva, N. S., Finkelstein, A. V., and Galzitskaya, O. V. (2006) Trend of amino acid composition of proteins of different taxa, *J Bioinform Comput Biol 4*, 597-608.
25. Brown, J. W., Vardar-Ulu, D., and McKnight, C. J. (2009) How to arm a supervillin: designing F-actin binding activity into supervillin headpiece, *J Mol Biol 393*, 608-618.
26. Rana, A. P., Ruff, P., Maalouf, G. J., Speicher, D. W., and Chishti, A. H. (1993) Cloning of human erythroid dematin reveals another member of the villin family, *Proc Natl Acad Sci U S A 90*, 6651-6655.
27. Koshino, I., Mohandas, N., and Takakuwa, Y. (2012) Identification of a novel role for dematin in regulating red cell membrane function by modulating spectrin-actin interaction, *J Biol Chem 287*, 35244-35250.
28. Cohen, C. M., and Gascard, P. (1992) Regulation and post-translational modification of erythrocyte membrane and membrane-skeletal proteins, *Seminars in hematology 29*, 244-292.
29. Siegel, D. L., and Branton, D. (1985) Partial purification and characterization of an actin-bundling protein, band 4.9, from human erythrocytes, *J Cell Biol 100*, 775-785.
30. Chen, L., Jiang, Z. G., Khan, A. A., Chishti, A. H., and McKnight, C. J. (2009) Dematin exhibits a natively unfolded core domain and an independently folded headpiece domain, *Protein Sci 18*, 629-636.
31. Chen, L., Brown, J. W., Mok, Y. F., Hatters, D. M., and McKnight, C. J. (2015) The allosteric mechanism induced by protein kinase A (PKA) phosphorylation of dematin (band 4.9), *J Biol Chem 290*, 17808.
32. Hofmann, A., Noegel, A. A., Bomblies, L., Lottspeich, F., and Schleicher, M. (1993) The 100 kDa F-actin capping protein of Dictyostelium amoebae is a villin prototype ('protovillin'), *FEBS Lett 328*, 71-76.
33. Vardar, D., Chishti, A. H., Frank, B. S., Luna, E. J., Noegel, A. A., Oh, S. W., Schleicher, M., and McKnight, C. J. (2002) Villin-type headpiece domains show a wide range of F-actin-binding affinities, *Cell Motil Cytoskeleton 52*, 9-21.
34. Huang, S., Qu, X., and Zhang, R. (2015) Plant villins: versatile actin regulatory proteins, *Journal of integrative plant biology 57*, 40-49.
35. Rechsteiner, M. (1990) PEST sequences are signals for rapid intracellular proteolysis, *Seminars in cell biology 1*, 433-440.
36. Yokota, E., Vidali, L., Tominaga, M., Tahara, H., Orii, H., Morizane, Y., Hepler, P. K., and Shimmen, T. (2003) Plant 115-kDa actin-filament bundling protein, P-115-ABP, is a homologue of plant villin and is widely distributed in cells, *Plant Cell Physiol 44*, 1088-1099.

37. Zhang, H., Qu, X., Bao, C., Khurana, P., Wang, Q., Xie, Y., Zheng, Y., Chen, N., Blanchoin, L., Staiger, C. J., and Huang, S. (2010) Arabidopsis VILLIN5, an actin filament bundling and severing protein, is necessary for normal pollen tube growth, *Plant Cell 22*, 2749-2767.
38. Khurana, P., Henty, J. L., Huang, S., Staiger, A. M., Blanchoin, L., and Staiger, C. J. (2010) Arabidopsis VILLIN1 and VILLIN3 have overlapping and distinct activities in actin bundle formation and turnover, *Plant Cell 22*, 2727-2748.
39. Bao, C., Wang, J., Zhang, R., Zhang, B., Zhang, H., Zhou, Y., and Huang, S. (2012) Arabidopsis VILLIN2 and VILLIN3 act redundantly in sclerenchyma development via bundling of actin filaments, *Plant J 71*, 962-975.
40. van der Honing, H. S., Kieft, H., Emons, A. M., and Ketelaar, T. (2012) Arabidopsis VILLIN2 and VILLIN3 are required for the generation of thick actin filament bundles and for directional organ growth, *Plant physiology 158*, 1426-1438.
41. Coordinators, N. R. (2017) Database Resources of the National Center for Biotechnology Information, *Nucleic Acids Res 45*, D12-D17.
42. Zhang, Y., Xiao, Y., Du, F., Cao, L., Dong, H., and Ren, H. (2011) Arabidopsis VILLIN4 is involved in root hair growth through regulating actin organization in a Ca2+-dependent manner, *The New phytologist 190*, 667-682.
43. Shanker, A. K., Maheswari, M., Yadav, S. K., Desai, S., Bhanu, D., Attal, N. B., and Venkateswarlu, B. (2014) Drought stress responses in crops, *Functional & integrative genomics 14*, 11-22.
44. Hu, H., and Xiong, L. (2014) Genetic engineering and breeding of drought-resistant crops, *Annual review of plant biology 65*, 715-741.
45. Fang, Y., and Xiong, L. (2015) General mechanisms of drought response and their application in drought resistance improvement in plants, *Cell Mol Life Sci 72*, 673-689.
46. Harpole, K. W., O'Brien, E. S., Clark, M. A., McKnight, C. J., Vugmeyster, L., and Wand, A. J. (2016) The unusual internal motion of the villin headpiece subdomain, *Protein Sci 25*, 423-432.
47. Huggins, D. J. (2016) Studying the role of cooperative hydration in stabilizing folded protein states, *J Struct Biol 196*, 394-406.
48. Wang, E., Wang, J., Chen, C., and Xiao, Y. (2015) Computational evidence that fast translation speed can increase the probability of cotranslational protein folding, *Sci Rep 5*, 15316.
49. Hsu, W. L., Shih, T. C., and Horng, J. C. (2015) Folding stability modulation of the villin headpiece helical subdomain by 4-fluorophenylalanine and 4-methylphenylalanine, *Biopolymers 103*, 627-637.
50. Meng, J., Vardar, D., Wang, Y., Guo, H. C., Head, J. F., and McKnight, C. J. (2005) High-resolution crystal structures of villin headpiece and mutants with reduced F-actin binding activity, *Biochemistry 44*, 11963-11973.
51. Frank, B. S., Vardar, D., Chishti, A. H., and McKnight, C. J. (2004) The NMR structure of dematin headpiece reveals a dynamic loop that is conformationally altered upon phosphorylation at a distal site, *J Biol Chem 279*, 7909-7916.
52. Packer, L. E., Song, B., Raleigh, D. P., and McKnight, C. J. (2011) Competition between intradomain and interdomain interactions: a buried salt bridge is essential for villin headpiece folding and actin binding, *Biochemistry 50*, 3706-3712.
53. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome, *Proc Natl Acad Sci U S A 98*, 10037-10041.
54. Cavanagh, J., Fairbrother, W.J., Palmer A.G., Rance, M., Skelton, N.J. *Protein NMR Spectroscopy, Principles and Practice Second Edition. Acedemic Press; 2007*.

55. Gasteiger, E., Hoogland C., Gattiker A., Duvaud, S., Wilkins M.R., Appel R.D., Bairoch, A. (2005) *Protein Identifrication and Analysis Tools on the ExPASy Server; The Proteomics Protocols Handbook*, Humana Press
56. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J Biomol NMR 6*, 277-293.
57. Johnson, B. A., and Blevins, R. A. (1994) NMRView: A computer program for the visualization and analysis of NMR data, *J Biomol NMR 4*, 603–614.
58. Berjanskii, M. V., Neal, S., and Wishart, D. S. (2006) PREDITOR: a web server for predicting protein torsion angle restraints, *Nucleic Acids Res 34*, W63-69.
59. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis, *J Comput Chem 25*, 1605-1612.
60. The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrodinger, LLC.
61. Touw, W. G., Baakman, C., Black, J., te Beek, T. A., Krieger, E., Joosten, R. P., and Vriend, G. (2015) A series of PDB-related databanks for everyday needs, *Nucleic Acids Res 43*, D364-368.
62. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers 22*, 2577-2637.
63. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis, *Nat Protoc 10*, 845-858.
64. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation, *Biochemistry 33*, 5984-6003.
65. Case, D. A., Cheatham, T. E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs, *J Comput Chem 26*, 1668-1688.
66. Joung, I. S., and Cheatham, T. E., 3rd. (2008) Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations, *J Phys Chem B 112*, 9020-9041.
67. Roe, D. R., and Cheatham, T. E., 3rd. (2013) PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data, *Journal of chemical theory and computation 9*, 3084-3095.
68. Ryckaert, J. P., Ciccotti, G., Berendsen, H. J. . (1997) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, *Journal of Computational Physics 23*, 327-341.
69. Gulans, A., Kontur, S., Meisenbichler, C., Nabok, D., Pavone, P., Rigamonti, S., Sagmeister, S., Werner, U., and Draxl, C. (2014) Exciting: a full-potential all-electron package implementing density-functional theory and many-body perturbation theory, *J Phys Condens Matter 26*, 363202.
70. Marley, J., Lu, M., and Bracken, C. (2001) A method for efficient isotopic labeling of recombinant proteins, *J Biomol NMR 20*, 71-75.

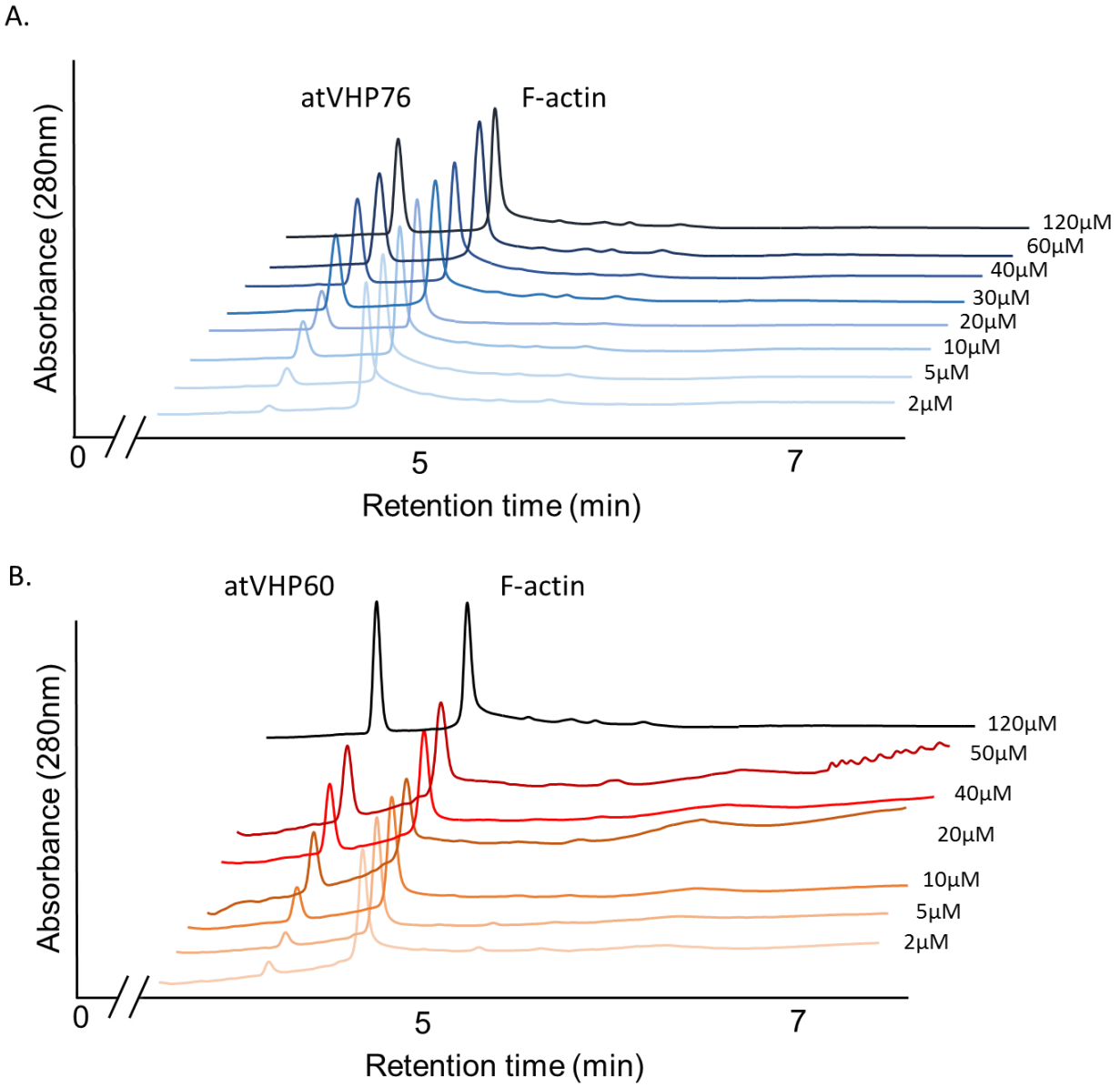
Appendix

Figure 1. SDS-PAGE of actin pull-down controls.



Lanes: (A) 1. F-actin only supernatant; 2. F-actin only pellet; 3. α-actinin only supernatant; 4. α-actinin only pellet; 5. BSA and F-actin supernatant; 6. BSA and F-actin pellet; 7. α-actinin and F-actin supernatant; 8. α-actinin and F-actin pellet (B) 1. F-actin only supernatant; 2. F-actin only pellet; 3. atVHP76 supernatant; 4. atVHP76 pellet; 5. atVHP76 and F-actin supernatant; 6. atVHP76 and F-actin pellet

Figure 2. HPLC chromatograms of F-actin pull down assay pellets.



(A) atVHP76 and F-actin pellet HPLC chromatograms. Peaks at 4.16 and 4.85 minutes are atVHP76 and F-actin respectively. Increasing concentrations of the headpiece are shown. (B) atVHP60 and F-actin pellet HPLC chromatograms. Peaks at 4.19 and 4.85 minutes are atVHP60 and F-actin respectively. Increasing concentrations of the headpiece are shown.

Table 1. Multidimensional heteronuclear NMR experiments and their respective dimensions performed on atVHP76

	^{1}H	^{13}C	^{15}N
Experiments	**Center/ Spec. Width/ NP**	**Center/ Spec. Width/ NP**	**Center/ Spec. Width/ NP**
HNCACB	4.77/ 13586.96/ 2816	43.06/ 13698.63/ 128	86.17/ 2498.75/ 54
CBCA(CO)NH	7.78/ 13586.96/ 2560	43.06/ 13698.63/ 128	117.96/ 2498.75/ 54
HNCO	4.77/ 7978.72/ 2560	176.60/ 1382.36/ 128	118.0/ 1418.04/ 64
HN(CA)CO	4.78/ 13586.96/ 2560	176.00/ 2352.94/ 128	117.96/ 2498.75 /54
HNCA	4.77/ 9578.54/ 3072	56.00/ 4222.97 / 64	121.00/ 2188.18/ 64
HCA(CO)N	4.77/ 9578.54/ 3072	56.00/ 4222.97 / 64	121.00/ 2188.18/ 64
CC(CO)NH	4.83/ 6684.49/ 2048	45.67/ 8038.59/ 96	118.00/ 1519.30/ 30
HCAN	4.77/ 9578.54/3072	56.00/ 4222.97/ 64	121.00/ 2188.18/ 64
^{13}C aliphatic HSQC	4.77/ 13586.96/ 2816	80.05/ 34246.58/ 512	
^{13}C aromatic HSQC	4.77/ 13586.96/ 4096	126.04/ 6839.95/ 186	
(Hβ)Cβ(CγCδ)Hδ	4.78/ 13736.26/ 4096	35.56/ 5133.47/ 86	
(Hβ)Cβ(CγCδCε)Hε	4.78/ 13736.26/ 4096	35.57/ 4812.32/ 80	
Hβ)Cβ(CγCarom)Harom	4.77/ 13736.26/ 4096	35.56/ 4812.32/ 74	
^{15}N HSQC	4.77/ 8012.82/ 2560		117.96/ 1620.75/ 256
	^{1}H (HA)	**^{1}H (HB)**	**^{15}N**
HBHA(CO)NH	4.77/ 8012.82/ 2560	4.77/ 3498.95/ 104	118.00/ 1418.04/ 54
H(CCO)NH	4.77/ 7978.72/ 2560	4.77/ 3498.95/ 128	118.00/ 1418.01/ 54
^{15}N TOCSY	4.76/ 13586.96/ 8192	4.76/ 11037.53/ 600	
^{15}N NOESY (250ms)	4.76/ 13586.96/ 8182	4.76/ 11037.53/ 600	
^{15}N NOESY (100ms)	4.77/ 13586.96/ 4096	4.77/ 10706.64/ 256	117.95/ 2498.75/ 54
	^{1}H (HA)	**^{1}H (HB)**	**^{13}C**
^{13}C NOESY aliphatic	4.18/ 13736.26/ 2048	4.74/ 10893.25/ 180	43.06/ 5133.47/ 64
^{13}C NOESY aromatic	4.45/ 13586.96/ 4096	4.78/ 10893.25/ 256	126.00/ 5988.02/ 14

Table 2. ^{15}N and ^{1}H resonance assignment for backbone amide groups in V4HP63.

Residue Number	Residue Name	^{1}H δ (ppm)	^{15}N δ (ppm)	Residue Number	Residue Name	^{1}H δ (ppm)	^{15}N δ (ppm)
10	ALA	8.139	124.231	**47**	TYR	7.352	113.569
11	LYS	8.299	120.673	**48**	LEU	7.142	117.442
12	GLU	8.440	122.100	**49**	SER	9.807	118.438
13	GLY	8.460	110.450	**50**	SER	9.224	117.461
14	VAL	7.889	118.346	**51**	GLU	8.813	120.073
15	GLU	8.560	124.451	**52**	GLU	8.170	121.594
16	ASP	8.316	121.347	**53**	PHE	9.283	120.530
18	GLU	8.351	121.991	**54**	LYS	7.923	118.751
19	ASP	8.401	122.670	**55**	GLU	7.726	119.631
20	LEU	8.081	124.267	**56**	LYS	7.881	114.757
22	ALA	7.659	120.610	**57**	PHE	8.483	112.936
23	HIS	8.994	120.661	**58**	GLY	8.149	108.538
25	TYR	9.889	127.099	**59**	MET	7.492	112.944
26	ASP	9.000	114.983	**60**	THR	8.069	107.025
27	ARG	7.761	114.798	**61**	LYS	8.316	120.908
28	LEU	7.424	120.652	**62**	GLU	8.208	115.809
29	LYS	7.346	116.281	**63**	ALA	7.430	120.154
30	THR	9.537	114.485	**64**	PHE	8.243	121.584
31	THR	6.962	106.745	**65**	TYR	8.073	113.436
32	SER	7.260	115.998	**66**	LYS	7.141	117.158
33	THR	8.767	117.298	**69**	LYS	9.052	125.934
34	ASP	8.473	121.866	**70**	TRP	8.287	114.747
36	VAL	7.562	118.787	**71**	LYS	6.105	124.762
37	SER	8.234	119.532	**72**	GLN	7.526	118.464
38	ASP	8.419	118.212	**73**	ASN	8.373	115.906
39	ILE	7.027	109.976	**74**	LYS	7.715	122.083
40	ASP	8.257	119.538	**75**	PHE	8.420	119.278
41	VAL	8.047	118.748	**76**	LYS	8.833	117.941
42	THR	8.615	113.064	**77**	MET	8.959	120.778
43	ARG	7.992	122.446	**78**	ALA	7.633	120.922
44	ARG	8.724	118.736	**79**	VAL	7.039	104.990
45	GLU	9.547	113.587	**80**	GLN	7.938	116.665
46	ALA	7.978	122.633	**81**	LEU	8.570	113.448
				82	PHE	7.741	126.845

Table 3. Backbone assignments of V4HP63.

		hncacb CB	hncacb CB-1	hncacb CA	hncacb CA-1	hncaco CO	hncaco CO-1	hnco CO-1	N-hsqc	H shift	N Shift
1	HIS										
2	HIS										
3	HIS										
4	HIS										
5	HIS										
6	HIS										
7	GLN										
8	GLU										
9	ASP										
10	ALA	19.24964	41.10037	52.56906	54.35255	177.60306	175.85651	175.87334	67	8.1187	124.05491
11	LYS	32.98063	19.23592	56.30844	52.5452	176.48811	177.60388	177.57573	64	8.2797	120.67268
12	GLU	30.24162	32.99278	56.55609	under 80?	176.91541	176.41901	176.43135	61	8.44039	122.21276
13	GLY	-	30.295	45.368	56.714	173.88969	176.90338	176.89693	30	8.46021	110.45002
14	VAL	32.8056	-	61.97953	45.35009	176.15981	173.8972	173.95905	71	7.90683	118.64157
15	GLU	30.29956	32.84351	56.53412	61.97473	under 16?	176.08968	176.16669	68	8.55955	124.45111
16	ASP	41.27946	30.41344	54.4274	56.53747	under 73?	176.09668	176.08711	66	8.31569	121.3466
17	GLU										
18	GLU	under 100?	30.6354	under 93?	56.5193	175.76917	176.27008	176.28828	63	8.35095	121.99078
19	ASP	40.98947	30.65277	53.97604	*56.560*	175.41403	175.77284	175.79137	51	8.40109	122.67031
20	LEU	41.54288	40.9206	52.82634	54.03979	174.8118	175.41379	175.44133	49	8.06556	124.32894
21	PRO	-	-	-	-	-	-	-	-		
22	ALA	19.70426	32.1405	50.09473	62.59546	178.08414	174.07307	174.13533	53	7.65187	120.58844
23	HIS	33.724	19.758	53.149	50.098	173.12671	178.08943	178.05138	8	9.01146	120.70972
24	PRO	32.62933	-	62.3033	-	-	-	-	-		
25	TYR	38.07651	32.63809	62.56361	under	176.87061	178.12767	178.09128	1	9.88171	127.17023
26	ASP	40.92207	38.0665	57.41888	62.77797	177.88901	176.87093	176.87254	7	9.00259	115.04874
27	ARG	31.350	41.008	57.215		175.61771	177.88333	177.8569	25	7.76328	114.87011
28	LEU	46.28229	31.35119	54.68781	-	176.37724	175.62859	175.64565	86	7.42956	120.68519
29	LYS	32.07504	46.27949	55.42498	under 266?	179.08994	-	176.40097	44	7.34795	116.31592
30	THR	67.882	32.083	65.114	55.446	179.09492	179.09492	179.09492	4	9.53954	114.55032
31	THR	68.15238	-	60.22265	-	174.82677		175.74684	74	6.96362	106.74451
32	SER	64.25492	68.23246	59.05026	-	176.59222	174.83264	174.88048	45	7.26084	116.04079
33	THR	68.702	64.286	61.869	58.991	173.97874	176.59718	176.60672	12	8.75144	117.34435
34	ASP	41.00827	68.67233	51.99955	61.8922	171.74806	173.99202	174.05188	60	8.4831	121.89255
35	PRO	32.55225	-	62.65376	-				-		
36	VAL	32.336	Under 215?	61.056	62.716	175.17253	178.56918	178.52553	29	7.56256	118.785
37	SER	64.481	32.276	58.458	61.056	173.94418	175.17143	175.21208	16	8.24546	119.55721
38	ASP	39.230	64.481	54.840	58.432	175.15709	173.96147	174.01341	14	8.41844	118.30138
39	ISO	41.81253	39.25893	59.50172	54.81794	175.02628		175.19466	42	7.02441	110.03165
40	ASP	40.19088	41.76025	52.77114	59.38547	177.63594		175.0639	85	8.26083	119.59836
41	VAL	30.629	40.176	63.982	52.765	177.28229	177.63602	177.6113	27	8.0377	118.78848

42	THR	69.340	30.718	63.939	63.278	175.40631	177.28119	177.26315	20	8.62067	113.14934
43	ARG	30.70042	69.35	54.28491	63.10971	176.77303			84	7.98607	122.56135
44	ARG	32.097	30.652	61.112	54.073	177.66187	176.77818	176.77769	13	8.71872	118.77028
45	GLU	27.979	32.116	57.967	61.095	176.27782	177.66646	177.63531	3	9.5635	113.68636
46	ALA	18.77187	27.98393	53.7313	57.94883	178.12178	176.28212	176.29341	57	7.9785	122.6635
47	TYR	38.80993	18.78871	58.34459	53.61856	174.23439	178.13052	178.09308	36	7.34013	113.52559
48	LEU	44.05148	38.80539	54.238	58.39435	178.44214	174.23433	174.2995	47	7.14323	117.46578
49	SER	64.469	43.999	57.138	54.251	176.87061	178.12767	178.09128	2	9.81324	118.55776
50	SER	-	64.460	62.551	57.121	-	174.6884	174.7372	6	9.23102	117.52685
51	GLU	29.387	-	60.037	62.605	179.33211	179.61591	179.53072	10	8.80098	120.07729
52	GLU	30.04586	under 109?	under 104?	60.03378	178.45703	179.21306	179.14944	59	8.16678	121.65244
53	PHE	39.586	30.157	62.657	59.988	177.60072	178.46416	178.42183	5	9.29136	120.57645
54	LYS	32.01712	39.55682	59.19567	under 204	under 1636	177.57819	177.60507	72	7.91383	118.7476
55	GLU	29.73004	31.92198	59.3306	under 179?	178.27043	180.19586	169.09821	70	7.72591	119.63093
56	LYS	31.128	29.863	55.922	59.350	177.64467	178.27562	178.23616	24	7.88672	114.83577
57	PHE	40.061	31.110	59.049	55.899	176.61426	177.65181	177.62018	22	8.48138	112.96487
58	GLY	-	40.083	46.287	59.026	172.80388	176.61623	176.62416	31	8.14655	108.58177
59	MET	35.40403	-	53.41564	46.31161	172.5367	172.84006	172.90247	35	7.49695	113.01279
60	THR	70.99953	35.41761	60.56368	53.43303	175.8692	172.54137	172.64352	38	8.07097	107.07727
61	LYS	32.28461	71.00694		60.32399	177.27818	-	-	65	8.31417	120.94696
62	GLU	29.035	32.199	59.724	under 277?	178.85483	177.2822	177.26608	19	8.22121	115.94639
63	ALA	19.13377	29.01958	54.68781	59.63164	180.54175	178.85452	178.80569	55	7.41937	120.18295
64	PHE	40.01559	19.14469	61.25572	54.56643	177.24512	180.54851	169.44394	58	8.24308	121.62894
65	TYR	36.741	39.995	60.255	61.208	176.63931	177.24342	177.23381	23	8.07572	113.46812
66	LYS	33.25329	36.71875	56.68362	60.19489	177.87141	176.65053	176.64494	46	7.13389	117.11061
67	LEU	40.85688	33.21571	53.22383	56.8792	175.11691	177.8772	177.84172	82	7.42617	120.69416
68	PRO	32.38845	-	62.34087							
69	LYS	31.32855		59.17527		178.14455		178.14455	83	9.05226	125.93414
70	TRP	26.839	31.232	59.228	under 288?	178.39349	under 179?	178.11639	18	8.30035	114.81687
71	LYS	32.37136	26.81369	59.01613	under 24?	177.54434	178.39851	178.35835	48	6.10909	124.80316
72	GLN	28.209	32.312	59.138	under 221?	179.42383	177.55016	177.52309	28	7.52893	118.53119
73	ASN	37.353	28.223	56.001	59.196	177.26236	179.42929	179.35571	17	8.37551	115.958
74	LYS	32.55383	37.36021	59.91261	55.9601	180.15596	177.26617	177.24828	52	7.71359	122.10235
75	PHE	38.322	32.547	under 190?	60.084	179.6156	180.16179	169.06425	15	8.42006	119.33997
76	LYS	32.098	38.299	60.844	under 233?	179.60934	179.60934	179.54118	11	8.83953	117.9931
77	MET	next to 130?	32.278	59.742	60.946	179.33211	179.61591	179.53072	9	8.9587	120.80808
78	ALA	18.57989	32.66835	54.84204	59.70195	178.68436	179.33473	179.26744	54	7.63368	120.96033
79	VAL	31.12546	18.57725	59.47046	54.81775	176.31517	178.68901	178.64502	73	7.03976	105.06889
80	GLN	24.992	31.133	58.749	under 263?	174.74106	176.32527	176.33369	26	7.94132	116.69062
81	LEU	43.341	24.966	53.111	58.627	174.23929	174.74699	174.79588	21	8.57073	113.47303
82	PHE	42.590	43.411	60.971	53.141				56	7.73797	126.90311

Table 4. Aliphatic side chain assignments of V4HP63.

		H shift	N Shift	HA PPM	HB2 PPM	HB3 PPM	hcconh HA	hcconh HB2	hcconh HB3	hcconh HD2	hcconh HD3	hcconh HG2	hcconh HG3	hcconh HE2	hcconh HE3
1	HIS														
2	HIS														
3	HIS														
4	HIS														
5	HIS														
6	HIS														
7	GLN														
8	GLU														
9	ASP														
10	ALA	8.1187	124.05491												
11	LYS	8.2797	120.67268	4.28815	1.78109		4.2889	1.78512						2.98457	
12	GLU	8.44039	122.21276	4.26165	1.98495		4.2637	1.97133				2.24309			
13	GLY	8.46021	110.45002	4.20529	2.69552										
14	VAL	7.90683	118.64157	4.14962	2.05652		4.1548	2.06386				7.87676			
15	GLU	8.55955	124.45111	4.53442	2.59922		4.2542					2.22887			
16	ASP	8.31569	121.3466												
17	GLU														
18	GLU	8.35095	121.99078	4.19245	1.92309		4.1911	1.94941				2.18111			
19	ASP	8.40109	122.67031	4.58126	2.6817	2.49388	4.5823	2.67908	2.49812						
20	LEU	8.06556	124.32894												
21	PRO														
22	ALA	7.65187	120.58844	5.237	8.21641	NA	5.2373	8.21793							
23	HIS	9.01146	120.70972												
24	PRO														
25	TYR	9.88171	127.17023	3.98148	3.08201	2.95729	3.9788	3.09157	2.95047						
26	ASP	9.00259	115.04874	4.06093	2.62182		4.0595	2.6576							
27	ARG	7.76328	114.87011	4.26424	1.94684	1.7838	4.2652	1.93588		2.93795		1.45107			
28	LEU	7.42956	120.68519	4.0877	1.62506	0.24947	4.0862	1.62514		7.73888		8.02241			
29	LYS	7.34795	116.31592	4.53558	1.81702	8.03404	4.5385	1.81861		8.06155		7.85484			
30	THR	9.53954	114.55032	4.3876	3.84652	NA	4.3907	3.84583				8.16178			
31	THR	6.96362	106.74451	4.52198	4.32707	NA	4.5127	4.33048				8.15231			
32	SER	7.26084	116.04079	3.97011	3.88924		3.9633	3.89424							
33	THR	8.75144	117.34435	4.5331			4.4124						8.26316		
34	ASP	8.4831	121.89255												
35	PRO														
36	VAL	7.56256	118.785	4.07955	1.45148		4.0779					7.31018			
37	SER	8.24546	119.55721	4.40379	3.76977	2.26097	4.4068	3.77282							
38	ASP	8.41844	118.30138	4.43691	2.68776		4.4391	2.68872							
39	ISO	7.02441	110.03165	4.25415	8.12806		4.2542	8.12275				7.46016	7.02323		
40	ASP	8.26083	119.59836	4.76958	3.17704		4.7689	3.1771	2.6987						
41	VAL	8.0377	118.78848	3.85356	2.40479		3.8537	2.39297				7.99458			
42	THR	8.62067	113.14934	4.57999			4.5736	4.41391				1.29384			
43	ARG	7.98607	122.56135	5.28214	1.94707	1.70197	5.2866	1.9552	1.77553	1.53025		3.04029			
44	ARG	8.71872	118.77028	3.95846	1.96487	1.7986	3.9575	1.97259	1.80197	1.57462		3.56208	2.86144		
45	GLU	9.5635	113.68636	4.29103	2.0057		4.2856	2.00224				2.25303			
46	ALA	7.9785	122.6635	4.06838	1.31107	NA	4.0660	1.3031							
47	TYR	7.34013	113.52559		3.55813		4.3422	3.55903							
48	LEU	7.14323	117.46578	4.88548	2.29069	1.75601	4.8831	2.29033	1.75349	0.22754		2.03247			
49	SER	9.81324	118.55776	4.46524	4.07609		4.4679	4.07479							
50	SER	9.23102	117.52685	3.99144	2.05961		4.0662								
51	GLU	8.80098	120.07729	4.10794	2.03109		4.1060	2.03234				2.35955			
52	GLU	8.16678	121.65244	4.17143	2.1076		4.1696	2.06682				2.50648	2.22821		
53	PHE	9.29136	120.57645	3.9306	2.14719		3.93274								
54	LYS	7.91383	118.7476	4.29593	1.92888		4.2986	1.92994		1.60182		1.38057			
55	GLU	7.72591	119.63093	3.92614	2.11611		3.9257	2.09519				2.32405			
56	LYS	7.88672	114.83577	3.87967	1.39846	7.84531	3.8799	1.37755	8.12491	7.84061		7.59926			
57	PHE	8.48138	112.96487	4.17273	2.52817	2.30662	4.1717	2.52737	2.30575						
58	GLY	8.14655	108.58177	3.91063	NA	NA	3.9116								

59	MET	7.49695	113.01279	4.77014	2.28559	2.07519	4.7696				2.75999				
60	THR	8.07097	107.07727				4.6823								
61	LYS	8.31417	120.94696	2.98461	1.41378	8.23601	2.9873	1.43335	8.22551	7.44722	6.63506				
62	GLU	8.22121	115.94639	3.91474	2.01794	1.84325	3.9144	2.03986	1.84273		2.2363				
63	ALA	7.41937	120.18295	3.95066	1.42676	NA	3.9489	1.42917							
64	PHE	8.24308	121.62894	4.09124	2.80503		4.0895	2.80894							
65	TYR	8.07572	113.46812	4.33838	3.26754	2.71864	4.3215	3.2715	2.71877						
66	LYS	7.13389	117.11061	4.42783	2.22895	1.48958	4.4260								3.35866
67	LEU	7.42617	120.69416												
68	PRO			4.40736	2.43703	8.2032	4.4112	2.435			1.59943	8.19871			
69	LYS	9.05226	125.93414	3.59538	1.97361		3.5984	1.97678		1.71712	1.56014		3.11225		
70	TRP	8.30035	114.81687	4.40788	3.48285	3.26695	4.4101	3.4827	3.26556						
71	LYS	6.10909	124.80316	3.38436	1.50798	2.42835	3.3856						2.74569		
72	GLN	7.52893	118.53119	3.55311	1.95048		3.5564	1.95137							
73	ASN	8.37551	115.958	4.56379	2.9215	2.59026	4.5663	2.92377	2.5285						
74	LYS	7.71359	122.10235	4.11774	2.03821	1.90011	4.1158	2.03791	1.87932	1.69015	1.48521		3.06647		
75	PHE	8.42006	119.33997	4.62858	3.57468		4.6281	3.57635							
76	LYS	8.83953	117.9931	4.07691	2.25886	1.84	4.0745	2.26074	1.84417	1.51774			2.51204		
77	MET	8.9587	120.80808	4.24522	2.49243		4.2487	2.49355							
78	ALA	7.63368	120.96033	4.24815	1.63451	NA	4.2501	1.63462							
79	VAL	7.03976	105.06889	4.69555	2.29344	NA	4.6963	2.29203			7.61593				
80	GLN	7.94132	116.69062	4.31587	2.5736	2.46955	4.3169	2.57591			2.35062				
81	LEU	8.57073	113.47303	4.88118	2.1197	1.50585	4.8800		1.51147	7.86099	8.11974				
82	PHE	7.73797	126.90311			NA									

Table 5. Aliphatic carbon side chain assignments of V4HP63.

		H shift	N Shift	CA	CB	ccconh CA	ccconh CB	ccconh CG1	ccconh CG2	ccconh CD1	ccconh CD2
1	HIS										
2	HIS										
3	HIS										
4	HIS										
5	HIS										
6	HIS										
7	GLN										
8	GLU										
9	ASP										
10	ALA	8.1187	124.05491	52.56906	19.24964						
11	LYS	8.2797	120.67268	56.30844	32.98063		33.12687	24.66414		29.20282	
12	GLU	8.44039	122.21276	56.55609	30.24162	56.71387		36.26568			
13	GLY	8.46021	110.45002	45.368	-	45.33115					
14	VAL	7.90683	118.64157	61.97953	32.8056	61.85783		20.86082			
15	GLU	8.55955	124.45111	56.53412	30.29956			36.29447			
16	ASP	8.31569	121.3466	54.4274	41.27946						
17	GLU					56.62978		36.30856			
18	GLU	8.35095	121.99078	under 93?	under 100?	56.53763		36.3776			
19	ASP	8.40109	122.67031	53.97604	40.98947	53.73365	41.03509				
20	LEU	8.06556	124.32894	52.82634	41.54288						
21	PRO			-	-		32.35305	27.25238			
22	ALA	7.65187	120.58844	50.09473	19.70426	49.95553	19.64946				
23	HIS	9.01146	120.70972	53.149	33.724						
24	PRO			62.3033	32.62933		32.75565	27.79258			
25	TYR	9.88171	127.17023	62.56361	38.07651		38.16276				
26	ASP	9.00259	115.04874	57.41888	40.92207	57.23592	40.96712				
27	ARG	7.76328	114.87011	57.215	31.350			29.63916			
28	LEU	7.42956	120.68519	54.68781	46.28229		46.17896	26.37301			
29	LYS	7.34795	116.31592	55.42498	32.07504						
30	THR	9.53954	114.55032	65.114	67.882				21.69105		
31	THR	6.96362	106.74451	60.22265	68.15238				22.32942		
32	SER	7.26084	116.04079	59.05026	64.25492		64.19583				
33	THR	8.75144	117.34435	61.869	68.702		68.77366		21.83232		
34	ASP	8.4831	121.89255	51.99955	41.00827						
35	PRO			62.65376	32.55225		32.63336	27.1198			
36	VAL	7.56256	118.785	61.056	32.336	61.15688		20.23206			
37	SER	8.24546	119.55721	58.458	64.481		64.5651				
38	ASP	8.41844	118.30138	54.840	39.230	54.72799	39.25254				
39	ILE	7.02441	110.03165	59.50172	41.81253						
40	ASP	8.26083	119.59836	52.77114	40.19088	52.44854	40.30254				

41	VAL	8.0377	118.78848	63.982	30.629			21.72761	17.43301		
42	THR	8.62067	113.14934	63.939	69.340				21.84848		
43	ARG	7.98607	122.56135	54.28491	30.70042			25.78899		44.42611	
44	ARG	8.71872	118.77028	61.112	32.097		32.16539	24.08316			
45	GLU	9.5635	113.68636	57.967	27.979			35.65386			
46	ALA	7.9785	122.6635	53.7313	18.77187	53.55313	18.69047				
47	TYR	7.34013	113.52559	58.34459	38.80993	58.20525	38.84379				
48	LEU	7.14323	117.46578	54.238	44.05148		43.84122	27.28399			
49	SER	9.81324	118.55776	57.138	64.469	57.08432	64.50151				
50	SER	9.23102	117.52685	62.551	-	63.16403					
51	GLU	8.80098	120.07729	60.037	29.387			36.73242			
52	GLU	8.16678	121.65244	under 104?	30.04586	39.12154					
53	PHE	9.29136	120.57645	62.657	39.586		39.63983				
54	LYS	7.91383	118.7476	59.19567	32.01712		32.06629	24.92884		29.39776	
55	GLU	7.72591	119.63093	59.3306	29.73004			35.96333			
56	LYS	7.88672	114.83577	55.922	31.128		30.7641	23.88157		26.66585	
57	PHE	8.48138	112.96487	59.049	40.061		40.08052				
58	GLY	8.14655	108.58177	46.287	-	46.17957					
59	MET	7.49695	113.01279	53.41564	35.40403			29.33087			
60	THR	8.07097	107.07727	60.56368	70.99953				22.17896		
61	LYS	8.31417	120.94696		32.28461		32.20349	25.12032		30.16883	
62	GLU	8.22121	115.94639	59.724	29.035			36.50034			
63	ALA	7.41937	120.18295	54.68781	19.13377	54.45688	19.03649				
64	PHE	8.24308	121.62894	61.25572	40.01559	61.10578	40.09498				
65	TYR	8.07572	113.46812	60.255	36.741		36.77419				
66	LYS	7.13389	117.11061	56.68362	33.25329			25.55314			
67	LEU	7.42617	120.69416	53.22383	40.85688						
68	PRO			62.34087	32.38845						
69	LYS	9.05226	125.93414	59.17527	31.32855		31.24956	24.25367		28.47565	
70	TRP	8.30035	114.81687	59.228	26.839	59.14135	26.81719				
71	LYS	6.10909	124.80316	59.01613	32.37136			24.95642		29.69463	
72	GLN	7.52893	118.53119	59.138	28.209			32.71919			
73	ASN	8.37551	115.958	56.001	37.353	55.91936	37.32058				
74	LYS	7.71359	122.10235	59.91261	32.55383			25.00421			
75	PHE	8.42006	119.33997	under 190?	38.322		38.30784				
76	LYS	8.83953	117.9931	60.844	32.098		32.05557			28.4868	
77	MET	8.9587	120.80808	59.742	next to 130?		33.11279				
78	ALA	7.63368	120.96033	54.84204	18.57989		18.49225				
79	VAL	7.03976	105.06889	59.47046	31.12546			20.86067	19.69536		
80	GLN	7.94132	116.69062	58.749	24.992			34.73978			

81	LEU	8.57073	113.47303	53.111	43.341	28.08659		23.57999
82	PHE	7.73797	126.90311	60.971	42.590			