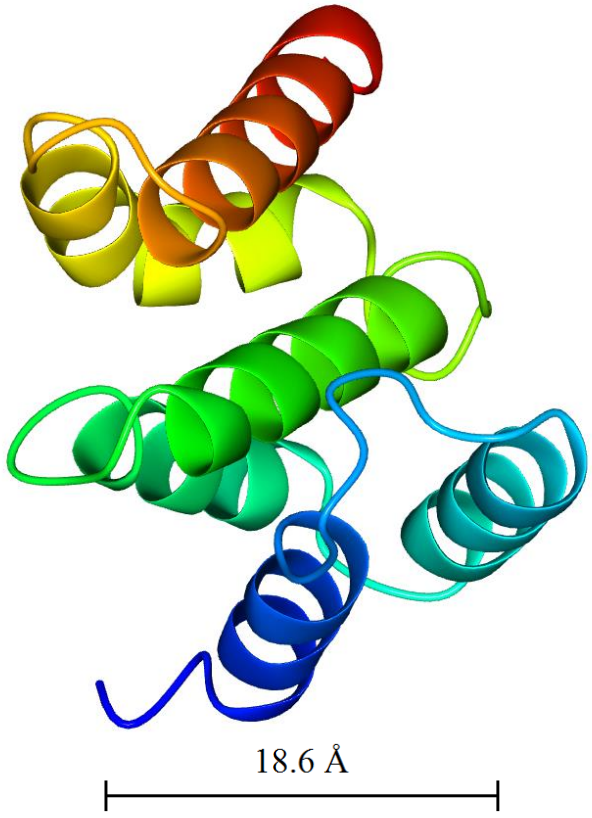
Structure Determination of the VHS-Like Domain of Tepsin



Chemistry Honors Thesis

Anderson Monken
Jackson Lab, Vanderbilt University

April 21, 2017

Table of Contents

Abstract

Adaptor protein 4 (AP4) is a critical intracellular trafficking protein complex in mammalian organisms that also plays a key role in the nervous system. A mutation in any of the four subunits of AP4 leads to a syndrome of debilitating diseases called hereditary spastic paraplegias. Tepsin is the only known accessory protein involved with AP4 in intracellular trafficking. Tepsin is composed of two ordered domains, one N-terminal and one central, and two C-terminal conserved motifs. These motifs are used to interact with the ε and β4 subunits of AP4 in an essential manner such that tepsin and AP4 exist synchronously and have evolved together. The biological function of tepsin is not presently known. As part of the Jackson Lab, I was on the team that determined the high-resolution x-ray crystal structure of the VHS-like domain of tepsin (tVHS) to a resolution of 1.86 Å. Circular dichroism spectroscopy (CD) was used to further test the structural features of tVHS. While there are significant similarities to other intracellular trafficking proteins that are involved in protein cargo binding and ubiquitin recognition, specific structural features on tVHS explain the biochemical experimental results that the Jackson Lab had previously conducted. While tVHS has structural similarities to other VHS domains it does not bind ubiquitin or acidic dileucine motifs.

Introduction

The intracellular trafficking system is essential to the life of the cell. The efficient and specific movement of proteins, which takes place in vesicles, requires an elaborate network of coat, adapter, and regulator proteins. Cargo-packaged vesicles are mediated by coat proteins that assist in forming the composition and curvature of the vesicle membrane [1]. Coat proteins act as

scaffolding for the vesicle, while adapter proteins bind specifically with various cargo motifs and drive the accuracy of protein trafficking.

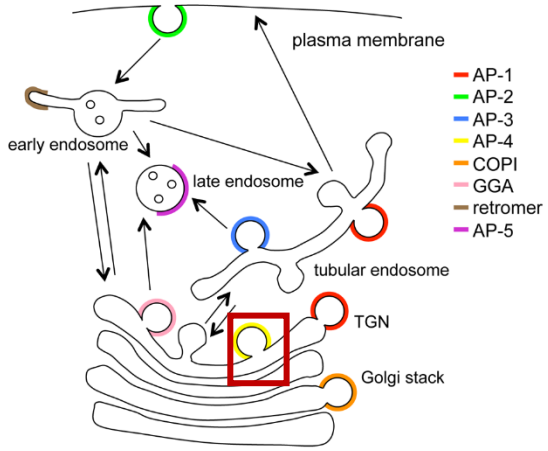


Figure 1. Overview of intracellular trafficking machinery
The path of each complex is designated by arrows. Some of the complexes do not have a known function. The red box highlights the role of AP4 [7].

Adaptor protein (AP) complexes play an essential role in the proper maintenance of intracellular trafficking. This family of heterotetrameric complexes includes five known AP complexes, each designated with a number from one to five, and the COPI F-subcomplex. Although adaptor protein complexes were named for their role between clathrin and the membrane layer, AP4 and AP5 are believed to function independently of clathrin [2]. Each AP consists of a core with two large subunits that vary for each AP, a medium-sized subunit, and a small subunit [3,4]. AP1 is responsible for trafficking between the *trans* Golgi network (TGN) and the endosome, AP2 mediates endocytosis, AP3 moves cargo from the endosome to the lysosome, AP4 vesicles originate from the TGN, and AP5 is involved in endosomal trafficking (Figure 1) [5–7]. The destinations of AP4 and AP5 coated vesicles are not fully understood [5,6].

AP4 adheres to the same general structure as other APs with ε, β4, σ4, and μ4 subunits. AP4 is of particular interest in the medical community because its proper function is critical for brain development in humans [5]. Mutations in any subunit of AP4 cause severe neurological disability including but not limited to intellectual disability, growth retardation, stereotypic laughter, progressive spasticity, and inability to walk [8,9]. These prognoses have led to clinically recognizable syndrome of hereditary spastic paraplegias [10].

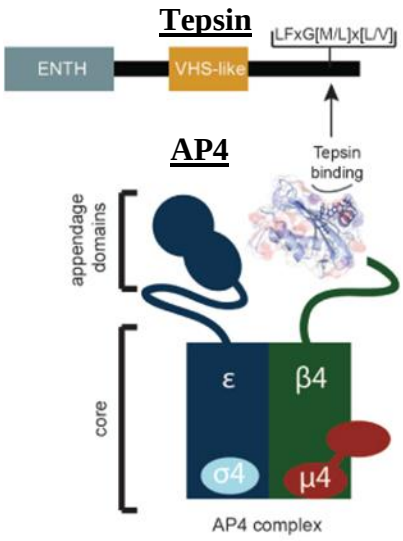


Figure 2. Schematic of the AP4 complex its binding partner, tepsin. [3]

There is little known about other cytosolic proteins that interact with the AP4 complex. In addition to the GTPase ADP-ribosylation factor 1 (Arf1), only one binding partner has been identified to date: tepsin [11,12]. Tepsin is a 525-amino acid protein consisting of two all α-helical domains: the ENTH, found at the N-terminus, and the VHS-like domain, located in the middle of the protein (Figure 2). The ε and β4 subunits each have a large trunk domain, that is part of the core, and an appendage domain, connected by a flexible linker capable of binding accessory proteins (Figure 2). Tepsin interacts with the β4 and ε appendage domain via linear motifs LFxG[M/L]x[L/V] and S[AV]F[SA]FLN, respectively [3,12,13]. Tepsin is not recruited to the membrane in the absence of AP4. The evolution of several model organisms has resulted in the loss of tepsin and AP4 simultaneously, suggesting that physiological function requires both the accessory protein and the complex in these organisms [12].

The focus of my work has been on the VHS-like domain of the accessory protein tepsin. It has been predicted that the VHS-like domain in tepsin is structurally similar to both the ENTH domain and the VHS domain [12].

The acronym ENTH stands for epsin N-terminal homology domain. The ENTH domain is highlighted by a super helix of α-helices and is found in proteins that are involved in trafficking signaling and in actin regulation [14]. This domain is one of the most conserved portions of the epsin family; ENTH has been found to bind to membrane lipids such as phosphatidylinositol-4,5-bisphosphate (PtdIns(4,5)P_2) which can lead to membrane curvature

[15]. This suggests that ENTH could play a role in vesicle budding or assist in complex formation.

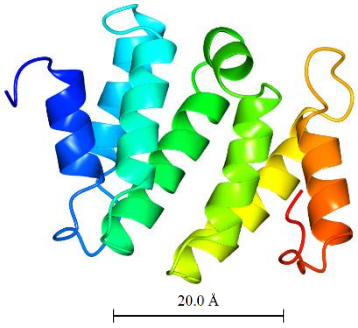


Figure 3. X-ray crystal structure of the VHS domain in Hrs protein (PDB ID: 1DVP) [16].

The VHS domain is a conserved domain with eight α-helices forming a helical super secondary structure (Figure 3) [16]. VHS is an acronym for three VHS-containing proteins, Vps27, Hrs, and STAM. VHS domains are known to bind ubiquitin and acidic dileucine motifs [17,18]. The Vps27 (vacuolar protein sorting) protein is critical for the multivesicular body sorting pathway, and binds to proteins and ubiquitinated cargo on the endosome [19]. The hepatocyte growth factor-regulated tyrosine kinase substrate (Hrs) is an endosomal-associated protein that may have a role in early endosome fusion and can form complexes with proteins, including STAM1, that can interact with ubiquitin moieties [20,21]. Signal-transducing adaptor molecule (STAM) has a VHS domain that drives the specificity and efficiency of a deubiquitinating enzyme [22]. The VHS domain in another family of proteins – the Golgi-localized, γ-adaptin era-containing, ADP-ribosylation factor-binding (GGA) family – can recognize an acidic dileucine motif (DxxLL) on protein cargo [17,23].

The goal for this project was express, purify, crystallize, and obtain a high-resolution structure of the tepsin VHS-like domain (tVHS). Constructs from several species with varying degrees of sequence identity to *H. sapiens* tVHS were analyzed using crystallographic techniques. Through the use of x-ray crystallography, we determined an atomic resolution structure of mammalian tVHS and found that the tVHS domain is, in fact, quite similar to both VHS and ENTH domains.

Methods

Construct Design

The amino acid sequence of human tepsin was obtained from the GenBank database (accession no. NP_653280.1) [24]. The VHS-like domain was defined as amino acid residues 220-360 based on a Network Protein Sequence (NPS) analysis consensus model, which is used to predict secondary structures [25]. This 140 residue sequence was processed in NCBI Protein BLAST analysis to identify species with high degrees of sequence identity [26,27]. A number of candidate species were chosen based on BLAST, and a multiple sequence alignment was performed using ClustalX to determine the degree of sequence conservation across species [28]. ClustalX confirmed that the region spanning amino acids 220 to 360 in *Homo sapiens* were an appropriate limit for the VHS-like domain, and the corresponding construct in each species served as the range limits for secondary structure prediction (Appendix: Figure 3) [25,28]. Five species were chosen (Table 2) in order to examine a range of identities with human tVHS. NPS analysis consensus models determined the predicted domain boundaries for each species which was the basis for choosing the construct boundaries of the homolog tVHS domains [25]. An example of the NPS analysis is shown in Figure 4 and the limits of all the constructs are shown in Appendix: Figure 3. Protein analysis using ProtParam on the ExPASy server provided key characteristics of each construct (Table 2) [29–31].

Shortened constructs that were made for circular dichroism were chosen based on PRISPRED secondary structure prediction to remove potential disordered portions of the tVHS domain [32,33]. Constructs were designed using the same procedure described above.

Species	Common name	Accession Number	Identities	pI	Molecular Weight (kDa)	ε all Cys bridges ($M^{-1}cm^{-1}$)	ε no Cys bridges ($M^{-1}cm^{-1}$)
Homo sapiens	Humans	NP_653280.1	100%	6.88	14.52	5875	5500
Callithrix jacchus	White-tufted-ear Marmoset	XP_008996139.1	91%	6.58	15.05	375	0
Equus caballus	Horse	XP_001490044.3	80%	6.07	14.54	5875	5500
Rattus norvegicus	Rat	NP_001094199.1	80%	5.02	14.62	375	0
Mus musculus	House mouse	NP_898960.2	78%	5.18	14.84	375	0
Bos taurus	Cow	NP_001095424.1	76%	6.90	14.89	5750	5500

Table 2. Species candidates for homology modelling. Accession numbers and sequence identity based on NCBI BLAST [26]. pI, mass, and ε obtained from ExPASy [29–31].

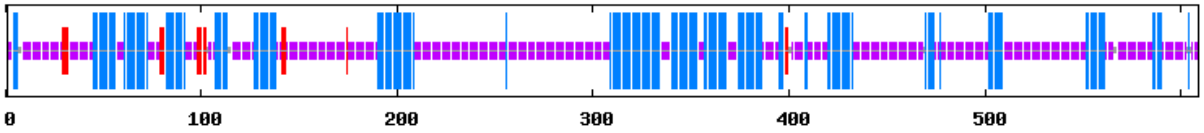


Figure 4. Secondary structure prediction for the VHS-like domain in *Equus caballus*. Long blue lines are predicted α-helices, red lines are predicted β-sheets, while short pink lines are predicted disordered regions. The arrows with numbers refer to the residue number in the *E. caballus* tepsin sequence. The boundary of the construct was chosen so that there were 4-5 disordered residues on the end of the α-helical secondary structures (306-442) [25].

Expression and Purification

Native Protein

DNA sequences for the constructs described in Table 2 were obtained from NCBI and cloned (GenScript; Piscataway, NJ) into a pGEX6P1 vector with an N-terminal GST tag at restriction sites BamHI and SalI. DH5α *E. coli* cells were used for DNA production. The tVHS constructs were expressed in BL21(DE3)pLysS cells grown in 2XTY media for 16-20 h at 22 °C after induction with 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were spun down at 3880 xg for 25 min in a JLA 8.100 rotor (Beckman-Coulter; Brea, CA), then

resuspended and frozen at -80 °C in a solution containing 20 mM HEPES (pH 7.5), 200 mM NaCl, 2 mM βME, 0.03 mM DNase (Sigma-Aldrich; St. Louis, MO), 2.5 mM $MgCl_2$, and 0.75 mM $MnCl_2$. After thawing, the cells were homogenized with a Dounce homogenizer and lysed using high pressure with rapid decompression in a disruptor (Constant Systems Limited; Daventry, United Kingdom). The lysed sample was centrifuged at 104,350 xg in a Ti45 rotor (Beckman-Coulter; Brea, CA) for one h at 4 °C. The constructs were affinity purified with glutathione sepharose resin (GE Healthcare; Little Chalfont, United Kingdom) and purification buffer. The buffer chosen was based on the theoretical pI of the construct to ensure the pH of the solution was one unit away from the pI of the protein (Table 3) [29,31]. GST-tagged protein was cleaved overnight at 4 °C with 0.55 mg of GST-3C protease and 2 mM DTT. Cleaved protein was then run over fresh glutathione sepharose resin and concentrated to ~5 mg/mL. The protein was further purified using size exclusion gel chromatography on a Superdex S200 preparative or analytical column using an ÄKTA System (GE Healthcare; Little Chalfont, United Kingdom).

Species	Purification Buffer
Rattus norvegicus	20 mM Tris (pH 8.7), 200 mM NaCl, 2 mM βME
Callintrhix jacchus	20 mM HEPES (pH 8.0), 200 mM NaCl, 2 mM βME
Equus caballus	20 mM HEPES (pH 7.5), 200 mM NaCl, 2 mM βME
Shortened *E. caballus*	20 mM Tris (pH 8.0), 200 mM NaCl, 2 mM βME

Table 3. Purification buffer chosen for purified tepsin VHS-like domain constructs.

Samples were analyzed by UV spectroscopy using a spectrometer (Eppendorf; Hamburg, Germany) at a wavelength of 280 nm. Aliquots from each step of the purification process were run on 4-20% gel gradient SDS-PAGE gels in 1x SDS-PAGE running buffer (Bio-Rad; Hercules, CA). Each lane being analyzed in the SDS-PAGE gel contained a 5 to 10 µL aliquot of a solution containing 25 µL of sample, 9 µL of SDS-PAGE dye and 4 µL of DTT. For constructs without aromatic residues and thus with an extinction coefficient ε=0 at 280 nm (Table 2), SDS-

PAGE protein bands were compared to BSA protein standards of 2.5, 5.0, and 10. mg/mL to determine protein concentration. Constructs lacking tryptophan and tyrosine residues were analyzed using a spectrometer (Eppendorf; Hamburg, Germany) at 260 nm with an extinction coefficient of 200 per Phenylalanine residue.

Selenomethionine-incorporated protein

Meredith Frazier, graduate student in the Jackson Lab, provided the selenomethionine-incorporated protein. *E. caballus* tVHS was expressed by in BL21(DE3)pLysS in minimal media using feedback inhibition and supplemental selenomethionine. The protein was purified using the same method described for native tVHS; however, 10 mM concentrations of DTT or βME were used as the reducing agent.

Crystallization Trials

The samples for crystallography were concentrated to ~7 mg/mL and prepared for crystal trials by spinning at 37,565 xg for 13 min at 4 °C followed by the addition of 2 mM DTT and 1x AEBSF. The protein samples were tested in a variety of commercial crystallization screens including JSCG-*plus* (Molecular Dimensions; Altamonte Springs, FL), Wizard Cryo HT (Rigaku Reagents, Inc.; Bainbridge, WA), Hampton Screen HT, and Hampton PEG/Ion Screen (Hampton Research; Aliso Viejo, CA). Protein was dispensed into 96-well MRC 2-well sitting drop crystallization plates using a Mosquito Crystal auto-dispenser (TTP Labtech; Cambridge, MA) that transferred volumes of 300 - 500 nL of protein and mother liquor into wells. Native crystals were soaked in CH_3HgCl, $Hg(OAc)_2$, or K_2PtCl_4 for a variety of times including 5 min, 15 min, 1

h, and 1 day to ensure good incorporation of the heavy-atom into the crystal solvent channels. Crystal trays were stored in a Rock Imager (Formulatrix; Bedford, MA) at 25 ºC, and images were collected in bright field and UV nine times over the course of 30 days to track the growth of crystals. Cryoprotection was achieved by the addition of 25% glycerol, ethylene glycol, or PEG600; some crystals were instead swiped through perfluoropolyether cryo oil (Hampton Research; Aliso Viejo, CA). All crystals were picked using 50 - 300 μm crystal loops (Hampton Research; Aliso Viejo, CA) and frozen in liquid nitrogen.

Data Collection

X-ray diffraction data from *E. caballus* native protein crystals was collected in the Center for Structural Biology (CSB) on a Bruker Microstar microfocus rotating-anode X-ray generator (Bruker; Madison, WI) with Bruker Proteum PT135 CCD area detector and Bruker KryoFlex cryostat device (Figure 8). Diffraction data from native, selenomethionine-incorporated, and heavy-atom soaked crystals was collected at the Life Sciences Collaborative Access Team beamlines 21-ID-D and 21-ID-G at the Advanced Photon Source, Argonne National Laboratory.

Structure Determination

All datasets were processed, integrated, and scaled using the HKL2000 package [34]. Solving the native data phasing problem was first attempted using molecular replacement with various ensembles of search models with the highest levels of sequence identity.

Search models were chosen by HHPred analysis and included some VHS domains among other structures with multiple α-helices (PDB IDs: 4QMI, 4L7M, 4G3A, 1EYH, 4GZC) [35–37].

ClustalX was used to align the search models to *E. caballus* tVHS [28]. CCP4i suite of programs was used to trim the search models, superimpose the models on one another, and perform the molecular replacement [38–41].

The SeMet scaled dataset was processed using the PHENIX Suite [42]. The SeMet derivative was analyzed using *xtriage* and run through AutoSol to provide phase information and attempt to build an initial model [43–46]. Autobuild was then used to iteratively build residues in the model [42,47]. Manual model building was undertaken in Coot to build some of the 20 missing amino acids on the N-terminus [48]. The model was then iteratively refined using phenix.refine to minimize clashes, rotamer outliers, and φ-ψ angles [42,47,49]. The best SeMet model became the search model for molecular replacement to determine the phase for the native dataset using the program PHASER [41,42,50]. The native model built using PHASER went through a final round of refinement through phenix.refine [42,47,49]. Rendering of protein models was performed on CCP4MG [38,51,52].

Circular Dichroism

tVHS constructs used for circular dichroism spectroscopy (CD) were buffer exchanged into a solution containing 10 mM phosphate pH 7.9 and 1mM DTT using a PD-10 column (GE Healthcare; Chicago, IL).

Circular dichroism spectra were collected on a J-720 spectropolarimeter (Jasco; Easton, MD). Far-UV and (190-250 nm) and near-UV (250-310 nm) CD measurements were performed in a 1 mm path length quartz cuvette at 4 ºC. The protein concentration for the full-length (306-442) and shortened (321-442) tVHS horse constructs were diluted to 0.5 – 1.5 mg/mL in 10 mM

phosphate pH 7.9 and 1 mM DTT. CD data were collected using standard sensitivity, scanning wavelengths of 260-190 nm with a data pitch of 1 nm. Using the continuous scanning mode, the speed was 50 nm/min with a response time of 2 sec, a bandwidth of 1 nm, and 2 or more accumulations were collected. Spectra Manager Suite was used to collect data (Jasco; Easton, MD).

Thermal scans of tVHS constructs were collected using a J-720 spectropolarimeter (Jasco; Easton, MD). Samples were heated from 4 ºC to 63 ºC using a Jasco programmable Peltier element at a heating rate of 1 ºC/min. Spectra was collected at 222 nm with a data pitch of 1ºC after a delay time of 90 sec using standard sensitivity (100mdeg), 4 sec response time, bandwidth of 1 nm, and one accumulation per concentration level. Three concentrations per construct were collected and analyzed individually.

Each spectrum was corrected for solvent contribution for the entire temperature range. The resulting ellipticity in mdeg was expressed in mean residue molar ellipticity (MRME) with an amino acid residue count of 123 and 138 for the shortened and full-length versions, respectively. The linear transformation is defined in Equation 1.

Data analysis of the thermal denaturation experiment to determine melting temperature and van't Hoff enthalpy was performed using STATA and Microsoft Excel, using the differential melting curve technique [53–55]. The first difference was taken for all data points to determine the maximum of the first difference which corresponds to the inflection point of the curve. I used a regression analysis in STATA using Equation 2 as the nonlinear fit function against the set of first differences. Summary statistics were taken for the resulting melting temperatures for each construct. T-testing was performed to determine if the average melting temperature for the two constructs are statistically equal.

$$MRME = \frac{m^{\circ} * MW}{10 * L * N * c}$$

Equation 1: Linear transformation of ellipticity in millidegrees (m°) to MRME using molecular weight in g/mol: (MW), path length in cm: (L), number of residues: (N), and concentration in mg/mL (c) [54,55]**.**

$$\frac{d(signal)}{dT} = A\frac{e^{\left[\frac{\Delta H_{vH}}{R}\left(\frac{1}{T_m}-\frac{1}{T}\right)\right]}}{1+e^{\left[\frac{\Delta H_{vH}}{R}\left(\frac{1}{T_m}-\frac{1}{T}\right)\right]}}(1-\frac{e^{\left[\frac{\Delta H_{vH}}{R}\left(\frac{1}{T_m}-\frac{1}{T}\right)\right]}}{1+e^{\left[\frac{\Delta H_{vH}}{R}\left(\frac{1}{T_m}-\frac{1}{T}\right)\right]}})(T^2)$$

Equation 2. Fit equation for differential melting curve analysis. Adjustable parameters are scaling factor: A, van't Hoff enthalpy: (ΔH_{vH}), and melting temperature: (T_m) [54].

Structure Comparison

The final best lab model, which Meredith Frazier completed, was compared with the PDB database using Protein structure comparison service PDBeFold at European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/ssm), authored by E. Krissinel and K. Henrick [40]. Structures with low root mean squared deviation (RMSD), high Q-score, and a high number of aligned residues are highlighted in the discussion. CCP4MG was used to superpose models and prepare figures that were used in the Discussion section of the paper [51,52].

Results

Expression and Purification

Five tVHS constructs were expressed in *E. coli* induced with IPTG. The bacterial cells were homogenized, lysed, and the mammalian proteins affinity purified using glutathione sepharose resin. An SDS-PAGE gel from the successful purification of *E. caballus* tVHS is shown in Figure 5. The elution lane contains a single isolated band at 15 kDa indicating a clean product at the expected size. The horse tVHS construct was further purified with an S200 Superdex preparative column and with fractions analyzed by UV absorbance (Figure 6).

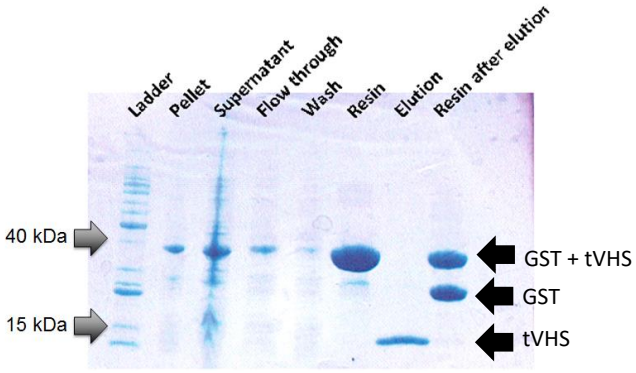


Figure 5. Purification of *E. caballus* tVHS. Cleaving the GST-tagged protein on the column resulted in a clean sample band at ~15 kDa in the elution lane.

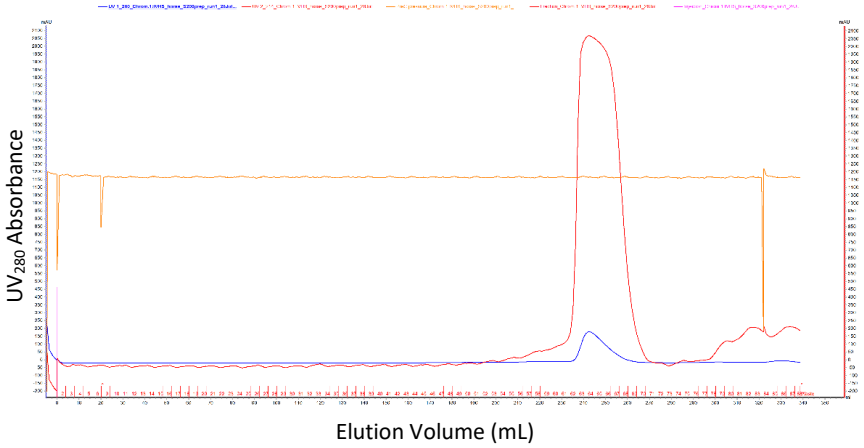


Figure 6. AKTA Absorbance Trace from *E. caballus* tVHS Purification. The cleaved protein sample exhibited a single major peak at 214 nm absorbance.

Crystallization Trials and Data Collection

Crystal trays were laid within two days of protein purification at concentrations ranging from 2 to 7 mg/mL in a variety of crystallization screens. Various combinations of mother liquor and protein volumes were tested to find the ideal crystal growth conditions. Protein crystal growth was monitored over the course of four weeks. tVHS protein from *R. norvegicus* and *C. jacchus* did not yield any crystals. *E. caballus* protein produced rod-like crystals appearing within one day under multiple conditions; the crystals exhibited a most consistent growth in 20% w/v PEG3350 (Figures 7 and 8).

Protein crystals were picked using crystal loops and cryoprotected by a variety of cryoprotectants including perfluoropolyether cryo oil. To ensure that the *E. caballus* native tVHS crystals were indeed protein crystals, preliminary x-ray diffraction data was collected at the Center for Structural Biology (CSB) at Vanderbilt University using a Bruker rotating-anode X-ray generator with a Bruker PT135 detector and a Bruker cryostat device. A snapshot of the diffraction pattern for a protein crystal is shown in Figure 9.

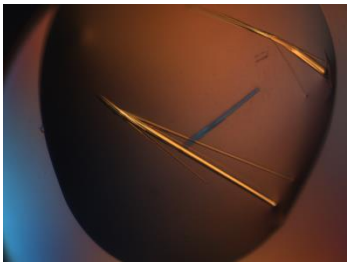


Figure 7. *E. caballus* tVHS crystals. An example of crystals from Hampton PEG/Ion screen. Mother liquor contained 0.2M Potassium Fluoride and 20% w/v PEG3350.

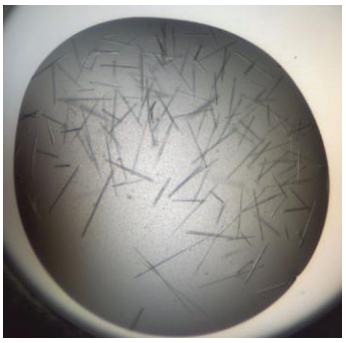


Figure 8. Selenomethionine-incorporated protein crystals of tVHS from *E. caballus*. Mother liquor containing 0.07 M Na_3Cit pH 2.3, 0.07 M BIS-TRIS propanol pH 9.7, and 20% w/v PEG3350.

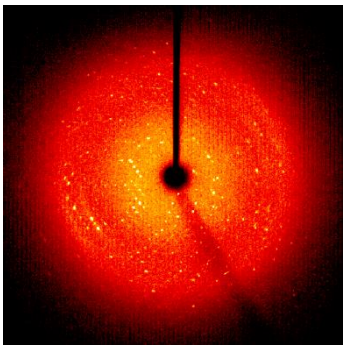


Figure 9: Diffraction pattern from crystals of tVHS from *E. caballus*. The presence of relatively tight spots indicates a single protein crystal.

Protein crystals were sent to the Argonne National Laboratory for detailed dataset collection. The crystals diffracted beyond 2 Å on the 21-ID-D beamline. Figure 10 shows a screenshot of a diffraction pattern from a native horse tVHS protein crystal. Appendix: Table 1 contains statistics from the processed datasets including space group, cell constants, completeness, and resolution. I was unable to find a compelling solution using molecular replacement with the native dataset due to sequence identity below 30% between tVHS and the most similar protein data base (PDB) models (PDB IDs: 4QMI, 4L7M, 4G3A, 1EYH, 4GZC).

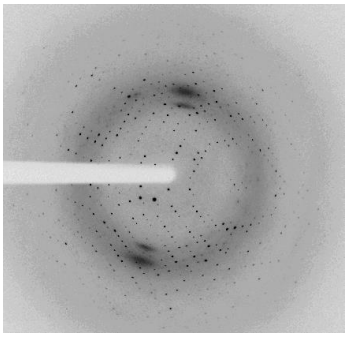


Figure 10. Diffraction pattern from crystals of tVHS from *E. caballus*. Image taken from HKL2000 viewer

Solving the phasing problem was approached from two directions. The first approach was to soak native crystals in heavy-atom derivatives (CH_3HgCl, $Hg(OAc)_2$, or K_2PtCl_4) for multiple isomorphous replacement (MIR) diffraction; this approach only yielded a single isomorphous replacement (SIR) data set from the $Hg(OAc)_2$ soaked crystals. The second approach was to grow selenomethionine incorporated protein for single-wavelength anomalous dispersion (SAD) diffraction. Meredith Frazier, graduate student in the Jackson Lab, expressed and purified the selenomethionine constructs. The three methionine residues on the construct should have been replaced by selenomethionine. Both approaches utilized the tunable beam line 21-ID-G to identify an anomalous signal and provide phasing information.

Structure Determination

Figure 11. X-ray crystal structure of *E. caballus* tVHS at 1.86 Å [52].

The resulting dataset from SAD diffraction was processed using HKL2000 contained phasing information from the anomalous signal from one or two selenomethionines; this provided information sufficient to elucidate the *E. caballus* tVHS structure [34]. An initial model of 109 residues was built and additional residues were added using Coot with further refinement using the PHENIX Suite (Figure 11) [42,48]. Our model fits the experimental data well; the validation statistics in PHENIX indicate that measures of clashes, deviations, and uncertainty are lower for our model than for other PDB models at similar resolutions (Figure 12) [42]. The polygon scoring consists of multiple one-dimensional histograms that compare statistics of our model to other protein structures deposited in the PDB of similar resolution [56]. This serves as a way to quickly estimate the model quality. Other geometric validations such a Ramachandran plot check the φ and ψ angles of the peptide bonds to ensure that they conform to normal combinations with low degrees of steric hinderance (Figure 13). Summary statistics for the refined model and refined experimental data appear in Appendix: Table 2.

Meredith Frazier elucidated another four residues at the N-terminus after using our initial structure as the search model for molecular replacement on the original native dataset. There are four residues on the C-terminus and twelve on the N-terminus of the protein that could not be confidently built into the experimental map. This could be simply because that portion of the protein is unstructured, but given NPS secondary structure prediction, it is likely that some type of helix should exist [25]. A different secondary structure predictor, PSIPRED, suggested the N-terminal helix of *E. caballus* would be disordered, which would explain its absence in the density

[32,33]. The chosen construct lengths and secondary structure prediction results are shown in Appendix: Figure 4. Mac Castro, rotation student in the Jackson Lab suggested the disordered helix could lead to decreased stability and thus a decrease in heat capacity and a lower melting temperature.

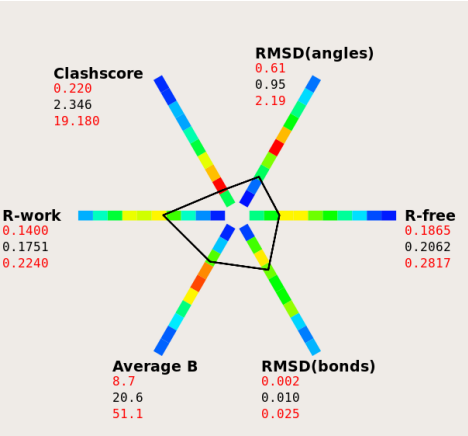


Figure 12. Polygon scoring of the experimental model. Compares my model with PDB models of similar resolution based on clash, geometric, and correlation statistics [56] .

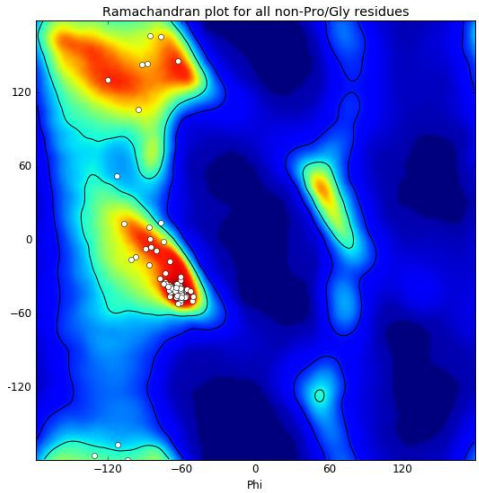


Figure 13. Ramachandran plot from PHENIX. Dots non-blue regions signify Ramachandran favored rotations. Our model has 96.36% Ramachandran favored angles.

Circular Dichroism

The disordered nature of the N-terminus of tVHS was then tested using circular dichroism on the full-length *E. caballus* tVHS construct and on a shortened construct lacking the theorized disordered helix. Both the full-length (306-442) and shortened (321-442) constructs were analyzed using circular dichroism spectroscopy to determine spectral scans. Based on the spectral scans, 222 nm was chosen for single-wavelength thermal denaturation experiments (see Appendix Figures 1 and 2 for thermal denaturation curves and first difference curves). The inclusion or deletion of the disordered helix had the potential to change the enthalpy of the protein and change its melting temperature (T_m). Nonlinear function fitting resulted in melting

temperatures as shown in Table 1. The t-testing indicated that there is not a statistically significant difference between the melting temperature for the full-length construct and the shortened construct.

Construct	Melting Temperature (K)	Number of Runs
Full-length *E. caballus* tVHS	314.48 ± 0.68 K	3
Shortened *E. caballus* tVHS	315.57 ± 0.68 K	3

Table 1. Circular dichroism regression results.

Discussion

Circular Dichroism

The circular dichroism experiments helped explain the role that the first twelve residues have in the *E. caballus* tVHS structure. Contrary to the NPS results, the first twelve residues do not fold in with the rest of the protein in the crystallized domain construct. There is no increase in melting temperature going from the shortened to the full-length construct, indicating that there was not a significant change in the stability of the protein. An increase in stability would have been expected from an additional helix folded into the domain.

Structural Comparison

The α-helical bundle found in *E. caballus* tVHS is ubiquitous to all VHS domains. There are six α-helices in this structure rather than the eight that are typical of VHS domains [16]. There is a significant degree of similarity based on RMSD, Q-Score, and the number of aligned residues between tVHS and other VHS and ENTH domains (Table 4). The most structurally similar protein domain to *E. caballus tVHS*, found in the list in Table 4, is in the Ent5 protein

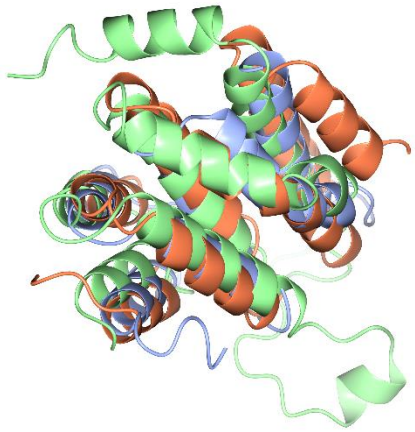


Figure 14. Horse tVHS domain (blue), N-terminus of Ent5 (green), CID Domain (orange).

which is a clathrin adaptor that supports the function of clathrin-coated vesicle (CCV) mediated traffic in yeast cells and has structural similarities to ENTH domains (Figure 14) [57]. The least similar protein domain is in RPRD1A; it is a CTD-interaction domain (CID), where CTD refers to the C-terminal domain of RNA polymerase II. This protein is involved with the transcription process (Figure 14) [58].

E. caballus tVHS comparison with:	PDB ID	RMSD (Å)	Q-Score [0-1]	Number of aligned residues
ENTH domain of Tepsin	Unpublished	2.12	0.41	97
ENTH Domain of Human Epsin	2QY7	2.29	0.34	96
VHS domain of GGA3	1JPL	2.28	0.38	105
VHS domain of *H. sapiens* STAM1	3LDZ	2.65	0.40	107
VHS domain of *H. sapiens* GGA1	1JWF	2.41	0.39	102
N-terminal domain of *S. cerevisiae* Ent5	5J08	1.92	0.43	100
CID domain of *H. sapiens* RPRD1A	4JXT	3.02	0.35	103

Table 4. Results from PDBeFold analysis of horse tVHS. Protein domains with similar structures as scored using RMSD, which is a measure of error in the two models, Q-score, which is a percentile ranking of similarity, and aligned residues [40].

E. caballus tVHS has a high degree of similarity to tepsin ENTH domain and ENTH domain of epsin. Figure 15 is a superposition of my experimental *E. caballus* tVHS the unpublished structure of the tepsin ENTH domain (Tara Archuleta, unpublished date) and the ENTH human epsin. The ENTH domain of human epsin is involved in clathrin-mediated endocytosis and plays a role in actin cytoskeleton organization in endocytosis [59]. ENTH

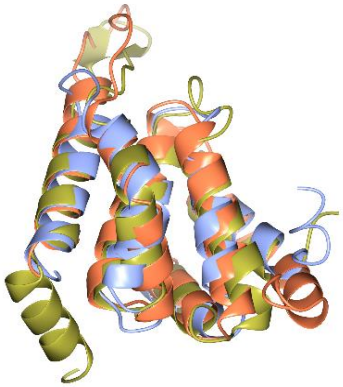


Figure 15. Horse tVHS domain (blue), tepsin ENTH (orange), epsin ENTH (gold).

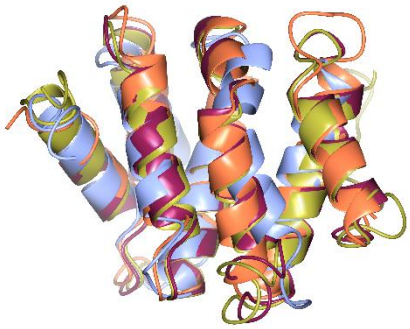


Figure 16. Horse tVHS domain (blue), VHS GGA3 (gold), VHS STAM1 (orange), VHS (purple).

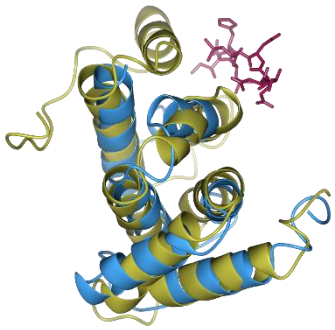


Figure 17. Superposition of tVHS (blue), VHS GGA3 (gold), and acidic dileucine peptide (purple). The lack of the final helix at the top of the figure means that tVHS will not bind well with a dileucine motif. Image courtesy of Meredith Frazier.

domains are known to function as membrane-interacting domains, binding phosphatidylinositol 4,5-bisphosphate (PIP_2) and other phospholipids [15,60]. Due to the high degree of structure similarity, the functional roles of tENTH and tVHS could be similar.

The VHS domain in GGA proteins and STAMs show up multiple times with high degrees of similarities (Figure 16). VHS domains in GGA proteins are responsible for cargo recognition in vesicular trafficking [61]. The sorting of mannose 6-phosphate receptors (MPRs) from the trans-Golgi network to endosomes is mediated by the VHS domain of GGAs using an acidic-cluster-dileucine motif [17,62]. Could tepsin VHS-like domain perform a similar function to the VHS dom ains of GGA proteins? Meredith Frazier performed a three-dimensional alignment of tVHS with the VHS domain of GGA3 shows that tVHS lacks a helix that is critical to forming the binding pocket for the dileucine motif (Figure 17). This explains previous research conducted by Meredith Frazier which determined that tVHS does not bind acidic dileucine.

Our structure is also similar to the VHS domain of (signal-transducing adapter molecule 1) STAM1, a protein that forms complexes with ubiquitin and other proteins to sort ubiquitinated cargo proteins from early endosomes to the endosomal sorting complex [63]. Ubiquitin binding is a common regulatory post-translational modification in many types of cellular processes. Meredith Frazier compared the electrostatic potential maps of VHS STAM1 and tVHS with a monoubiquitin. Due to the presence of Arg344, tVHS is unable to bind ubiquitin. The residue would sterically clash with the region of the protein that is similar to the region that VHS STAM1 uses to bind ubiquitin (Figure 18).

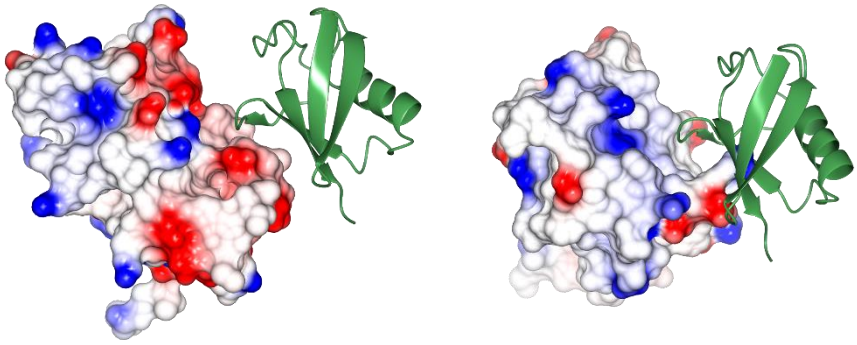


Figure 18. The electrostatic potential mapping for 3LDZ (left) and tVHS (right) with ubiquitin (green). Notice the basic patch of Arginine residue 344 that clashes with where ubiquitin would bind. Image courtesy of Meredith Frazier.

The structural information about tVHS serves as the background for further analysis about its binding partners and regulation pathways. The implications of a better understanding of tepsin and the AP4 trafficking could lead to substantial translational research discoveries.

The results of the structural alignment point to the importance of tepsin in the AP4 system. Tepsin and AP4 evolve together, and thus tepsin plays a critical role in the operation of this trafficking system [12]. Candidate binding partners, ubiquitin and an acidic dileuicine motif, were proven to not interact with tVHS through previous biochemical expirments in the Jackson Lab and confirmed by the structural analysis of tVHS. While tVHS is similar to many VHS and ENTH domains, key structural differences lead to tVHS not functioning in the same manner.

One question of intense interest are the roles of the ENTH and VHS-like domains of tepsin and whether they might have similar responsibilities in the trafficking framework. Tepsin as an accessory protein that does not have a known function. By elucidating the function of tVHS, we will be able to learn about its role in the AP4 trafficking system.

Acknowledgements

Thank you to everyone who provided advice, support, and assistance towards my project. I would like to show great appreciation for Meredith Frazier who was my partner and mentor throughout this entire project. We spent countless days planning experiments, discussing results, and writing things up. She was involved in practically every aspect of the day-to-day experimentation that was described above.

Jackson Lab: Dr. Jackson, Amy Kendall, Tara Archuleta, Rodger Burcham, Betty Xie, Mac Castro, and Allison Isabelli

CSB: Joel Harp and Ben Spiller

LS-CAT: Elena Kondrashkina and Zdzislaw Wawrzak

Thank you to members of my committee: Dr. Todd Graham, Dr. Lauren Jackson, Dr. Adam List, and Dr. Carmelo Rizzo. I appreciate your taking time out of their busy schedules to hear about my research as an undergraduate and give me feedback on my written and presented work.

This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (Grant 085P1000817).

References

• 1. Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., and Walter P. Molecular Biology of the Cell. 2015.
• 2. Robinson MS. Adaptable adaptors for coated vesicles. *Trends Cell Biol.* 2004;**14**:167–74.
• 3. Frazier MN, Davies AK, Voehler M, Kendall AK, Borner GHH, Chazin WJ, et al. Molecular Basis for the Interaction Between AP4 β4 and its Accessory Protein, Tepsin. *Traffic*. 2016;**17**:400–15.
• 4. Dell'Angelica EC, Mullins C, Bonifacino JS. AP-4, a novel protein complex related to clathrin adaptors. *J. Biol. Chem.* 1999;**274**:7278–85.
• 5. Hirst J, Irving C, Borner GHH. Adaptor Protein Complexes AP-4 and AP-5: New Players in Endosomal Trafficking and Progressive Spastic Paraplegia. *Traffic*. 2013;**14**:153–64.
• 6. Hirst J, Bright NA, Rous B, Robinson MS. Characterization of a fourth adaptor-related protein complex. *Mol. Biol. Cell*. 1999;**10**:2787–802.
• 7. Hirst J, D. Barlow L, Francisco GC, Sahlender DA, Seaman MNJ, Dacks JB, et al. The Fifth Adaptor Protein Complex. Schmid SL, editor. *PLoS Biol.* 2011;**9**:e1001170.
• 8. Abou Jamra R, Philippe O, Raas-Rothschild A, Eck SH, Graf E, Buchert R, et al. Adaptor Protein Complex 4 Deficiency Causes Severe Autosomal-Recessive Intellectual Disability, Progressive Spastic Paraplegia, Shy Character, and Short Stature. *Am. J. Hum. Genet.* 2011;**88**:788–95.
• 9. Moreno-De-Luca A, Helmers SL, Mao H, Burns TG, Melton AMA, Schmidt KR, et al. Adaptor protein complex-4 (AP-4) deficiency causes a novel autosomal recessive cerebral palsy syndrome with microcephaly and intellectual disability. *J. Med. Genet.* 2011;**48**:141–4.
• 10. Abdollahpour H, Alawi M, Kortüm F, Beckstette M, Seemanova E, Komárek V, et al. An AP4B1 frameshift mutation in siblings with intellectual disability and spastic tetraplegia further delineates the AP-4 deficiency syndrome. *Eur. J. Hum. Genet.* 2015;**23**:256–9.
• 11. Boehm M, Aguilar RC, Bonifacino JS. Functional and physical interactions of the adaptor protein complex AP-4 with ADP-ribosylation factors (ARFs). *EMBO J.* 2001;**20**:6265–76.
• 12. Borner GHH, Antrobus R, Hirst J, Bhumbra GS, Kozik P, Jackson LP, et al. Multivariate proteomic profiling identifies novel accessory proteins of coated vesicles. *J. Cell Biol.* 2012;**197**.
• 13. Mattera R, Guardia CM, Sidhu SS, Bonifacino JS. Bivalent Motif-Ear Interactions Mediate the Association of the Accessory Protein Tepsin with the AP-4 Adaptor Complex. *J. Biol. Chem.* 2015;**290**:30736–49.
• 14. De Camilli P, Chen H, Hyman J, Panepucci E, Bateman A, Brunger AT. The ENTH domain. *FEBS Lett.* 2002;**513**:11–8.
• 15. Ford MGJ, Mills IG, Peter BJ, Vallis Y, Praefcke GJK, Evans PR, et al. Curvature of clathrin-coated pits driven by epsin. *Nature*. 2002;**419**:361–6.
• 16. Mao Y, Nickitenko A, Duan X, Lloyd TE, Wu MN, Bellen H, et al. Crystal Structure of the VHS and FYVE Tandem Domains of Hrs, a Protein Involved in Membrane Trafficking and Signal Transduction. *Cell*. 2000;**100**:447–56.
• 17. Misra S, Puertollano R, Kato Y, Bonifacino JS, Hurley JH. Structural basis for acidic-cluster-dileucine sorting-signal recognition by VHS domains. *Nature*. 2002;**415**:933–7.
• 18. Ren X, Hurley JH. VHS domains of ESCRT-0 cooperate in high-avidity binding to polyubiquitinated cargo. *EMBO J.* 2010;**29**:1045–54.
• 19. Katzmann DJ, Stefan CJ, Babst M, Emr SD. Vps27 recruits ESCRT machinery to endosomes during MVB sorting. *J. Cell Biol.* 2003;**162**:413–23.

• 20. Pullan L, Mullapudi S, Huang Z, Baldwin PR, Chin C, Sun W, et al. The Endosome-Associated Protein Hrs Is Hexameric and Controls Cargo Sorting as a “Master Molecule.” *Structure*. 2006;**14**:661–71.
• 21. Sun W, Yan Q, Vida TA, Bean AJ. Hrs regulates early endosome fusion by inhibiting formation of an endosomal SNARE complex. *J. Cell Biol.* 2003;**162**:125–37.
• 22. Baiady N, Padala P, Mashahreh B, Cohen-Kfir E, Todd EA, Du Pont KE, et al. The Vps27/Hrs/STAM (VHS) Domain of the Signal- transducing Adaptor Molecule (STAM) Directs Associated Molecule with the SH3 Domain of STAM (AMSH) Specificity to Longer Ubiquitin Chains and Dictates the Position of Cleavage *. *J. Biol. Chem.* 2015;**291**:2033–42.
• 23. Santonico E, Mattioni A, Panni S, Belleudi F, Mattei M, Torrisi MR, et al. RNF11 is a GGA protein cargo and acts as a molecular adaptor for GGA3 ubiquitination mediated by Itch. *Oncogene*. 2015;**34**:3377–90.
• 24. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GenBank. *Nucleic Acids Res.* 2007;**35**:D21–5.
• 25. Combet C, Blanchet C, Geourjon C, Deléage G. NPS@: Network Protein Sequence Analysis. Trends Biochem. Sci. 2000. p. 147–50.
• 26. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;**25**:3389–402.
• 27. Altschul SF, Wootton JC, Gertz EM, Agarwala R, Morgulis A, Schäffer AA, et al. Protein database searches using compositionally adjusted substitution matrices. *FEBS J.* 2005;**272**:5101–9.
• 28. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 1997;**25**:4876–82.
• 29. Bjellqvist B, Hughes GJ, Pasquali C, Paquet N, Ravier F, Sanchez J-C, et al. The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis*. 1993;**14**:1023–31.
• 30. Bjellqvist B, Basse B, Olsen E, Celis JE. Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. *Electrophoresis*. 1994;**15**:529–39.
• 31. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, et al. Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol.* 1999;**112**:531–52.
• 32. Buchan DWA, Minneci F, Nugent TCO, Bryson K, Jones DT. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res.* 2013;**41**:W349–57.
• 33. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 1999;**292**:195–202.
• 34. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 1997;**276**:307–26.
• 35. Soding J. Protein homology detection by HMM-HMM comparison. *Bioinformatics*. 2005;**21**:951–60.
• 36. Söding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 2005;**33**:W244-8.
• 37. Meier A, Söding J. Automatic Prediction of Protein 3D Structures by Probabilistic Multi-template Homology Modeling. Ben-Tal N, editor. *PLOS Comput. Biol.* 2015;**11**:e1004343.
• 38. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of

the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 2011;**67**:235–42.
• 39. Schwarzenbacher R, Godzik A, Grzechnik SK, Jaroszewski L, W. G, W. M, et al. The importance of alignment accuracy for molecular replacement. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 2004;**60**:1229–36.
• 40. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 2004;**60**:2256–68.
• 41. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ, et al. Phaser crystallographic software. *J. Appl. Crystallogr.* 2007;**40**:658–74.
• 42. Adams PD, Afonine P V., Bunkóczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 2010;**66**:213–21.
• 43. Zwart PH, Grosse-Kunstleve RW, Adams PD. Xtriage and Fest: automatic assessment of X-ray data and substructure structure factor estimation. *CCP4 Newsl.* 2005;**43**.
• 44. Terwilliger TC, Adams PD, Read RJ, McCoy AJ, Moriarty NW, Grosse-Kunstleve RW, et al. Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 2009;**65**:582–601.
• 45. Wang JW, Chen JR, Gu YX, Zheng CD, Jiang F, Fan HF, et al. SAD phasing by combination of direct methods with the *SOLVE* / *RESOLVE* procedure. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 2004;**60**:1244–53.
• 46. Kabsch W, Sander C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* 1983;**22**:2577–637.
• 47. Terwilliger TC, Grosse-Kunstleve RW, Afonine P V., Moriarty NW, Zwart PH, Hung L-W, et al. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 2008;**64**:61–9.
• 48. Emsley P, Lohkamp B, Scott WG, Cowtan K, IUCr, W. G-KR, et al. Features and development of *Coot*. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 2010;**66**:486–501.
• 49. Headd JJ, Echols N, Afonine P V., Moriarty NW, Gildea RJ, Adams PD, et al. Flexible torsion-angle noncrystallographic symmetry restraints for improved macromolecular structure refinement. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 2014;**70**:1346–56.
• 50. Bunkóczi G, Echols N, McCoy AJ, Oeffner RD, Adams PD, Read RJ, et al. Phaser.MRage: automated molecular replacement. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 2013;**69**:2276–86.
• 51. Potterton L, McNicholas S, Krissinel E, Gruber J, Cowtan K, Emsley P, et al. Developments in the *CCP* 4 molecular-graphics project. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 2004;**60**:2288–94.
• 52. McNicholas S, Potterton E, Wilson KS, Noble MEM, IUCr, H. LW, et al. Presenting your structures: the *CCP* 4 *mg* molecular-graphics software. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 2011;**67**:386–94.
• 53. StataCorp. Stata Statistical Software: Release 14. 2015.
• 54. John DM, Weeks KM. van't Hoff enthalpies without baselines. *Protein Sci.* 2000;**9**:1416–9.
• 55. Greenfield NJ. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat. Protoc.* 2006;**1**:2527–35.
• 56. Urzhumtseva L, Afonine P V, Adams PD, Urzhumtsev A. Crystallographic model quality at a glance. *Acta Crystallogr. D. Biol. Crystallogr.* 2009;**65**:297–300.
• 57. Zhang F, Song Y, Ebrahimi M, Niu L, Teng M, Li X. Structural and functional insight into

the N-terminal domain of the clathrin adaptor Ent5 from Saccharomyces cerevisiae. *Biochem. Biophys. Res. Commun.* 2016;**477**:786–93.
• 58. Ni Z, Xu C, Guo X, Hunter GO, Kuznetsova O V, Tempel W, et al. RPRD1A and RPRD1B are human RNA polymerase II C-terminal domain scaffolds for Ser5 dephosphorylation. *Nat. Struct. Mol. Biol.* 2014;**21**:686–95.
• 59. Koshiba S, Kigawa T, Kikuchi A, Yokoyama S. Solution structure of the epsin N-terminal homology (ENTH) domain of human epsin. *J. Struct. Funct. Genomics*. 2002;**2**:1–8.
• 60. Kay BK, Yamabhai M, Wendland B, Emr SD. Identification of a novel domain shared by putative components of the endocytic and cytoskeletal machinery. *Protein Sci.* 1999;**8**:435–8.
• 61. Lohi O, Poussu A, Mao Y, Quiocho F, Lehto V-P. VHS domain - a longshoreman of vesicle lines. *FEBS Lett.* 2002;**513**:19–23.
• 62. Puertollano R, Aguilar RC, Gorshkova I, Crouch RJ, Bonifacino JS. Sorting of Mannose 6-Phosphate Receptors Mediated by the GGAs. *Science (80-.).* 2001;**292**.
• 63. Hong Y-H, Ahn H-C, Lim J, Kim H-M, Ji H-Y, Lee S, et al. Identification of a novel ubiquitin binding site of STAM1 VHS domain by NMR spectroscopy. *FEBS Lett.* 2009;**583**:287–92.

Appendix

Statistic		Native Data	SeMet
Beamline		LSCAT-21-ID-G	LSCAT-21-ID-D
Wavelength (Å)		0.97857	0.97910
Space group		P61	P65
Cell Constants			
	a, b, c (Å)	58.86, 58.86, 69.263	57.99, 57.99, 82.21
	α, β, γ (°)	90, 90, 120	90, 90, 120
Cell volume (Å^3)		207786.2	2077991.2
Resolution (Å)[a]		50.00 - 1.83 (1.90 - 1.83)	50.00 - 1.92 (1.99 - 1.92)
R_{sym}[a]		0.065 (1.621)	0.138 (0.594)
I/σ(I)[a]		36.2 (2.14)	17 (2.6)
Completeness (%)		100.	98.8 (89.4)
Redundancy		14.4 (17.8)	12.6 (~5)

[a]Values in parentheses are for highest-resolution shell.

Table 1: Data collection statistics for crystals collected at Argonne National Laboratory.

	My tVHS Best Model		
Wavelength			
Resolution range	29.41 - 1.86 (1.927 - 1.86)		
Space Group	P 65		
Unit cell	58.828 58.828 69.078 90 90 120	**Protein residues**	112/133
Total reflections	206259 (20632)	**RMS(bonds)**	0.01
Unique reflections	11459 (1153)	**RMS(angles)**	0.95
Multiplicity	18.0 (17.9)	**Ramachandran favored (%)**	96.36
Completeness (%)	99.95 (100.00)	**Ramachandran allowed (%)**	3.64
Mean I/sigma(I)	26.11 (2.39)	**Ramachandran outliers (%)**	0
Wilson B-factor	14.32	**Rotamer outliers (%)**	0
Reflections used in refinement	11456 (1153)	**Clashscore**	2.35
Reflections used for R-free	1130 (111)	**Average B-factor**	19.47
R-work	0.1747 (0.2628)	**macromolecules**	18.09
R-free	0.2073 (0.3106)	**solvent**	29.29

Table 2: Table of the experimental model and refined experimental data.
Statistics for the highest-resolution shell are shown in parentheses.

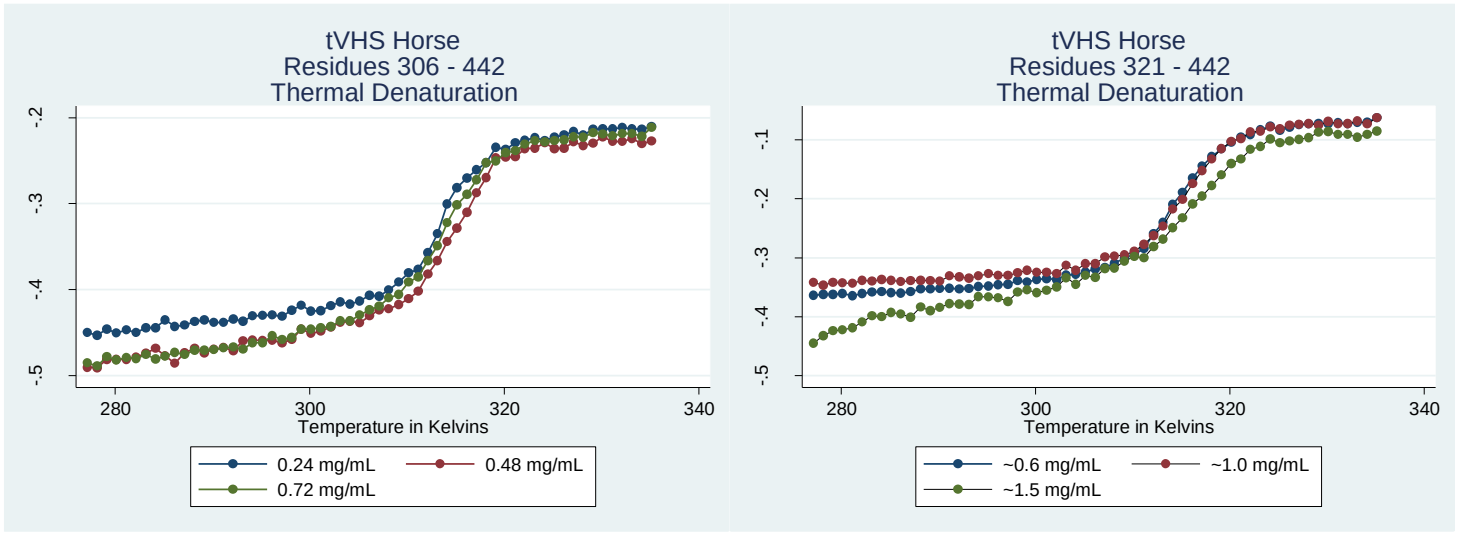


Figure 1: Thermal denaturation curves for full-length tVHS (on the left) and shortened tVHS (on the right).

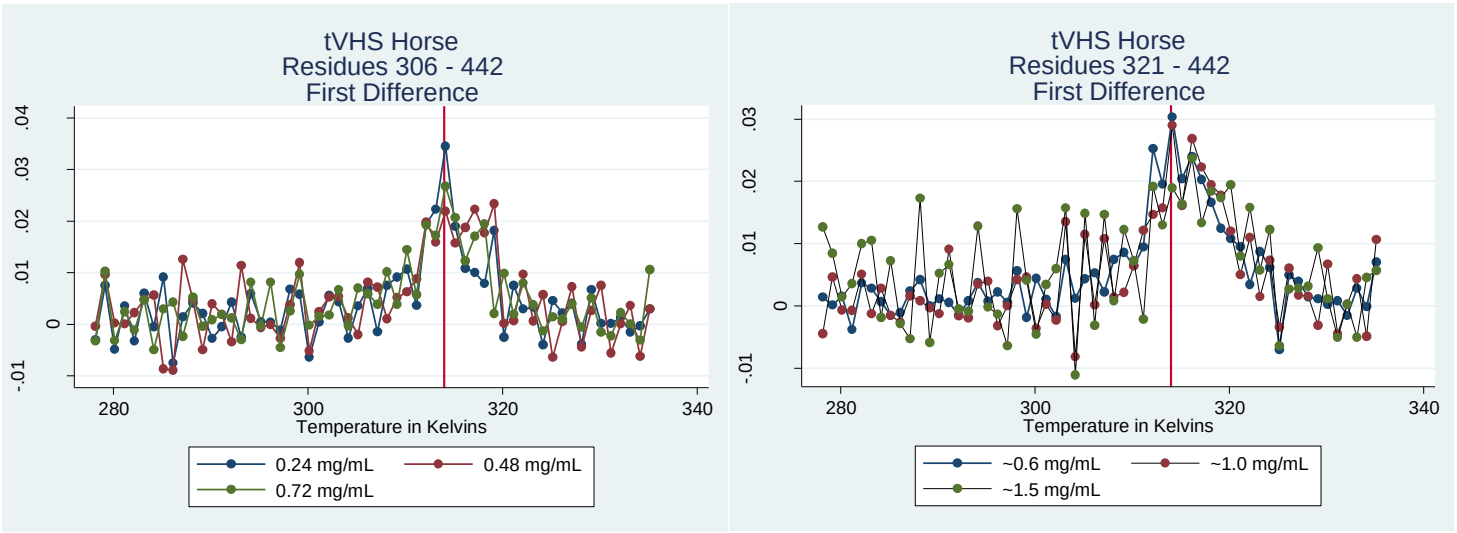


Figure 2: First difference curves for full-length tVHS (on the left) and shortened tVHS (on the right). The red line is at 314 K, which is close to the maximum of the first difference curves.