Structural and Functional Studies of AP4 and its Accessory Protein Tepsin

By

Meredith N. Frazier

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

In

Biological Sciences

May 11, 2018

Nashville, Tennessee

Approved:

Todd R. Graham, Ph.D.

Kendal S. Broadie, Ph.D.

Brandt F. Eichman, Ph.D.

Charles R. Sanders, Ph.D.

Lauren P. Jackson, Ph.D.

“There is a sea, and I am a captain
Of something unknown, waves high as mountains
There is a key, and there is a light
Here’s to times that I catch it
May it last through the night”
The Avett Brothers

ACKNOWLEDGEMENTS

I owe my success in this endeavor to many people who have encouraged and assisted me throughout the years, from Cannon, to Elon, to Vanderbilt. First, I would like to thank my friends and colleagues at Vanderbilt who have supported me by reading drafts, lending me reagents last-minute, letting me vent, and helping me troubleshoot. My support network of scientists has been great, even if many were not structural biologists and did not understand the unique appeal of determining structures.

I would especially like to thank Erin Palmer, my Elon roommate, fellow Vanderbilt scholar, and editor. Your eye for organization and attention to style helped make this thesis much more readable, and I am very grateful. I also must thank Dr. Jim Patton for connecting me with a community of Vanderbilt baseball fans. I appreciate all the games I enjoyed from behind home plate with fellow baseball enthusiasts, sharing both my love for keeping score and practicing my basic science communication skills. I believe very strongly in the necessity of a good work-life balance to manage stress, stimulate new ideas, and be a well-rounded person, and Vanderbilt baseball filled that niche for me.

The Jackson lab and the Center for Structural Biology have been a great place to spend my graduate career. The wealth of knowledge and expertise in the CSB made collaborations easy and helped shape my project greatly. I have enjoyed starting the Jackson lab as the first graduate student and watching it grow. Amy Kendall, you are the best lab manager—you keep the lab stocked and running smoothly, and are always in the know. Natalie and Allie, there's nobody else I would want to share the annex with. Thanks for always helping out, putting up with my artistic shenanigans, and being so quotable. I know you both will go on to do great things. And I would definitely not be nearly as successful as I am today without the mentorship and friendship of Tara Archuleta. I could always count on your knowledge and assistance, Tara, and you kept me from making many mistakes. You also shared many time-saving tips and tricks, for which I am forever thankful.

Thank you to my committee: Dr. Chuck Sanders, Dr. Brandt Eichman, Dr. Kendal Broadie, and my chair Dr. Todd Graham, for their constructive criticism and advice throughout the years. Thank you to Dr. Walter Chazin, for being an outstanding training grant director and source of support. Thank you to my collaborators in the Robinson, Chazin, Graham, and Creanza labs.

I would like to thank my advisor, Dr. Lauren Jackson, for accepting me into the lab and pushing me to accomplish my goals. You were always available to discuss my latest ideas or problems, and you always went to bat for me. Thank you for training me in structural biology, encouraging me to branch out, and understanding spontaneous trips to Omaha, Nebraska, for the College World Series.

And last but certainly not least, I must acknowledge my family. My parents have invested so much into my education, and I am forever grateful for your love and support. You have given me the gift of a world class education and instilled a love of learning in me early—I can't imagine many mothers encourage their second-grade daughter to read the dictionary and pick out vocabulary words like 'loquacious' and 'zoology'. You may not understand much about what I do now, but you try! And most importantly, my parents go above and beyond to give me a hand when I need it. My father guided me through the process of buying a home during my preparation for my qualification exam (excellent timing on my part), which was very helpful and much appreciated. My brother has also always supported me—and even stepped in and used his Illustrator skills when I couldn't get my own versions to look just right—so many thanks for saving me frustration, and being a sounding board for creative ideas. Much love to you all!

ACKNOWLEDGEMENT OF FUNDING

This work was supported by the Vanderbilt Molecular Biophysics Training Grant, NIH T32GM008320, trainee Meredith Frazier; NIGMS R35GM119525 to Lauren Jackson; and funding from the Pew Charitable Trusts to Lauren Jackson. Chapter III was supported in part by the Vanderbilt Postdoctoral Program in Functional Neurogenomics, T32MH65215, trainee Tara Archuleta.

PUBLICATIONS

The work in Chapter II appears in *Traffic,* 2017 Sep;18(9):590-603. The work in Chapter III appears in *Traffic*, 2016 Apr;17(4):400-15. Reuse permissions have been obtained from the journal for both articles: license 4266120722095 (Frazier *et al.,* 2016) and license 4266120782432 (Archuleta*, Frazier* *et al.,* 2017).

Additional publications are as follows:

Xu P, Hankins HM, MacDonald C, Erlinger SJ, Frazier MN, Diab NS, Piper RC, Jackson LP, MacGurn JA, Graham TR. COPI mediates recycling of an exocytic SNARE by recognition of a ubiquitin sorting signal. *Elife*, 2017 Oct 23;6:e28342.

Frazier MN, Jackson LP. Watching real-time endocytosis in living cells. *J Cell Biol*, 2017 Jan 2;216(1):9-11.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AAK1	α-appendage binding kinase 1
AMPAR	α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor
ANTH	AP-180 N-terminal homology domain
AP	adaptor protein, originally assembly polypeptide
AP1	adaptor protein complex 1
AP2	adaptor protein complex 2
AP3	adaptor protein complex 3
AP4	adaptor protein complex 4
AP5	adaptor protein complex 5
APP	amyloid precursor protein
APS	Advanced Photon Source
ARF	ADP-ribosylation factor
Atg	autophagy related proteins
BMRB	Biological Magnetic Resonance Bank
BSA	bovine serum albumin
C-ERMAD	C-terminal ERM-association domain
CALM	clathrin assembly lymphoid myeloid leukemia protein
CCC	CCDC93, CCDC22, COMMD
CCV	clathrin coated vesicle
CD	circular dichroism
CI-MPR	cation-independent mannose-6-phosphate receptor
CIA	cytosolic iron-sulfur cluster assembly proteins
CIC-3	synaptic vesicle chloride channel 3

CME	clathrin mediated endocytosis
COPI	coat protein complex I
CRISPR	clustered, regularly interspaced, short palindromic repeat
CSP	chemical shift perturbation
ECL	electrochemiluminescence
ENTH	epsin N-terminal homology domain
ER	endoplasmic reticulum
FCS	fetal calf serum
FERM	band4.1, ezrin, radixin, moesin
GAK	cyclin-G-associated kinase
GAP	GTPase activating protein
GEF	guanine exchange factor
GFP	green fluorescent protein
GGAs	Golgi-localized, γ-ear containing, ADP ribosylation factor binding family
GPI	glycosylphosphatidylinositol
GST	glutathione S-transferase
HRP	horseradish peroxidase
HSP	hereditary spastic paraplegia
HSQC	heteronuclear single quantum correlation
IP	immunoprecipitation
IRP1	aconitase/iron regulatory protein 1
ISC	iron-sulfur cluster assembly proteins
ISCA1	iron-sulfur cluster assembly 1
ISCA1 ΔMTS	iron-sulfur cluster assembly 1 (13-129)
ITC	isothermal titration calorimetry

LC3	microtubule-associated proteins 1A/1B light chain 3B
LDLR	low density lipoprotein receptor
LROs	lysosome related organelles
NF-L	neurofilament light chain
NMR	nuclear magnetic resonance
PASTA	Practical Alignments using SATé and TrAnsitivity
PBS	phosphate buffered saline
PBS-T	phosphate buffered saline with Tween
PDB	Protein Data Bank
PI(4,5)P2	phosphatidylinositol 4,5-bisphosphate
PI(4)P	phosphatidylinositol 4-phosphate
PIP	phosphatidylinositol phosphate
PICALM	phosphatidylinositol binding clathrin assembly protein
RMSD	root mean square deviation
RUSC2	Run and SH3 domain containing 2
SeMet	selenomethionine
SNARE	soluble N-ethylmaleiide-sensitive factor attachment protein receptors
SNX17	sorting nexin 17
SPG47	spastic paraplegia gene 47
SRP	signal recognition particle
TARPs	transmembrane AMPA receptor regulatory proteins
TBS	Tris buffered saline
TBS-T	Tris buffered saline with Tween
tENTH	tepsin ENTH domain
TfR	transferrin receptor

TGN	trans-Golgi network
TRIM	tripartite motif
tVHS	tepsin VHS/ENTH-like domain
Ub	monoubiquitin
UBDs	ubiquitin binding domains
UIMs	ubiquitin interacting motifs
VAMP	vesicle associated membrane protein
VGAT	vesicular GABA transporter
VGLUT1	vesicular glutamate transporter 1
VHS	VPS27, Hrs, Stam domain
VPS	vacuolar protein sorting associated proteins
VPS35L	VPS35-like
VPS35L Y2H ID	VPS35-like (438-688)
VPS35L ΔN-term	VPS35-like (87-1050)
WASH	WASP and SCAR homologue
Y2H	yeast two-hybrid

CHAPTER I

INTRODUCTION

The intracellular trafficking of proteins and lipids is essential for eukaryotic life. Many different mechanisms have evolved to ensure proper trafficking, localization, and regulation throughout the cell. In addition to controlling cargo localization, vesicle trafficking also plays a key role in controlling the size and composition of the membrane-bound compartments in the cell. These trafficking pathways can broadly be classified into two types, the biosynthetic and the endocytic pathways (Figure 1-1).

Biosynthetic trafficking pathway

The biosynthetic trafficking pathway coordinates protein movement from synthesis, translocation into the endoplasmic reticulum (ER), passage through the Golgi complex, and sorting to the appropriate organellar destination (Figure 1-1). Approximately one third of the human proteome is trafficked through this pathway (reviewed in DeMatteis and Luini, 2011). Vesicles carry cargoes between compartments; the presence of signal peptides and cargo motifs helps trafficking machinery direct proteins along the proper pathways to the correct destination.

The endoplasmic reticulum

The ER, an extended tubular network throughout the cell that makes contacts with the plasma membrane and other organelles (West *et al.,* 2011; Alpy *et al.,* 2013; Csordás *et al.,* 2006; reviewed in Philips and Voeltz, 2016), has many functions including protein synthesis and folding, assembly of protein complexes, and lipid biosynthesis and metabolism. Proteins are directed to the ER by

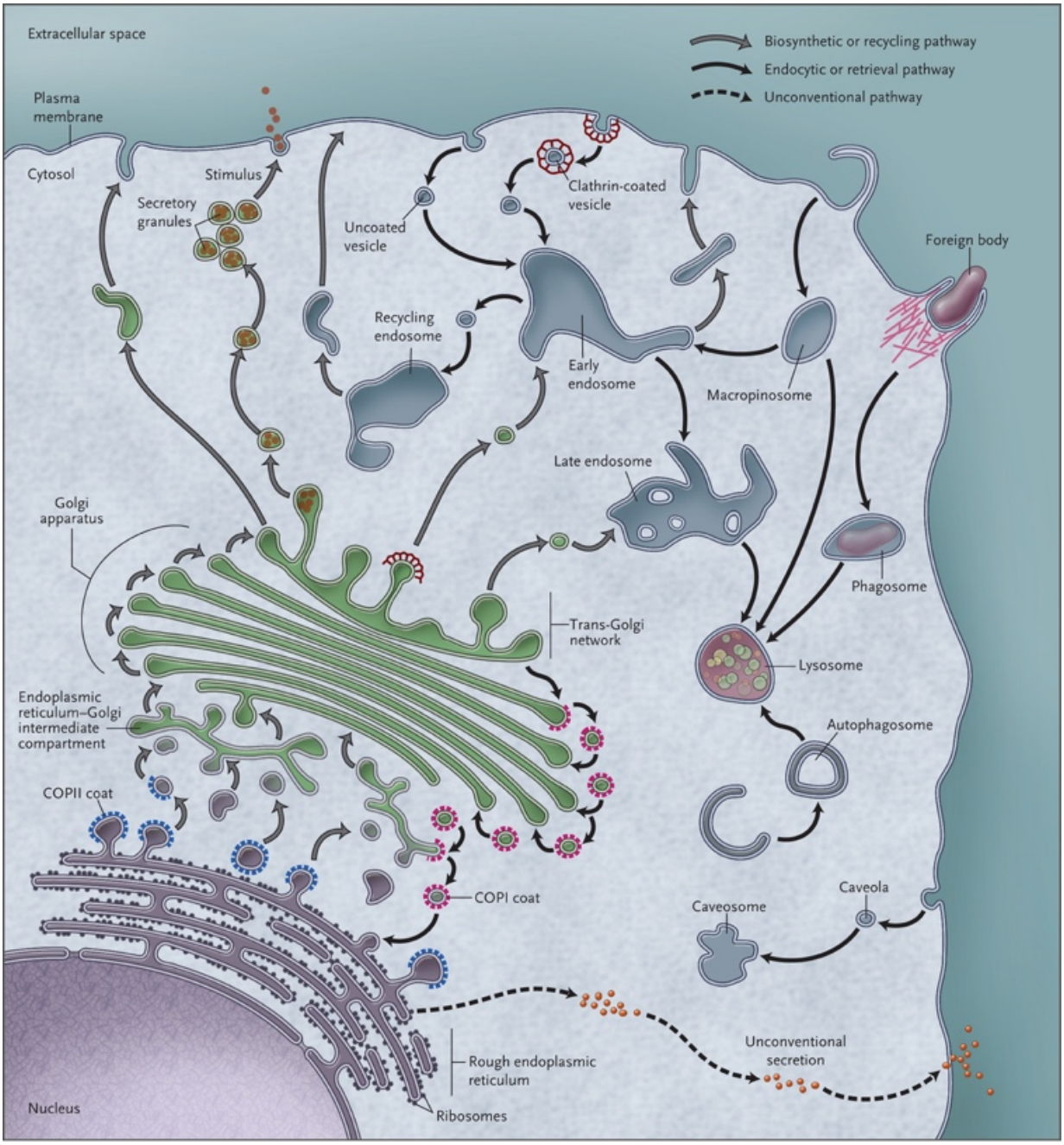


Figure 1-1: Schematic of intracellular trafficking pathways from DeMatteis and Luini, 2011.

an N-terminal signal peptide, which is recognized by the signal recognition particle (SRP; reviewed in Aviram and Schuldiner, 2017). The SRP machinery inserts the nascent polypeptide chain into the ER. Soluble proteins are released into the ER lumen following translation, while trans-membrane proteins stay in the ER membrane. To exit the ER, proteins must pass quality control mechanisms to ensure they are folded properly (reviewed in Tannous *et al*., 2015). Molecular chaperones modulate the energy landscape so proteins can reach their native fold more easily. Enzymes like protein disulfide isomerase catalyze proper disulfide bond formation. If proteins do not pass quality control, they will be retained in the ER and eventually targeted for degradation by the proteasome.

The Golgi complex

After exiting the ER, proteins make their way through the *cis*-, *medial*- and *trans*-Golgi cisternae, acquiring any further post-translational modifications they may need for their final function (glycosylation, phosphorylation, etc; De Matteis and Luini, 2011). The *trans*-Golgi Network (TGN) acts as the main sorting site for proteins and lipids in the biosynthetic pathway, with numerous originating paths for transporting cargoes in vesicles to their final destinations in the endosomal system or at the plasma membrane.

Biosynthetic trafficking in polarized cells

Many eukaryotic cells are polarized, even if only transiently as is the case with migrating fibroblasts. To create and maintain polarity, proteins must be differentially trafficked to establish domains with distinct properties (reviewed in Stoops and Caplan, 2014). For example, in polarized epithelial cells, which line the digestive system, proteins can be trafficked to the apical or basolateral plasma membrane. Specialized trafficking pathways transport proteins to a polarized domain, either from the TGN or distinct endosomal populations for basolateral or apical sorting (Sheff *et al.,* 2002; Ang *et al.,* 2004; Cresawn *et al.,* 2007). Signals for basolateral sorting resemble endocytic signals: short, linear cytosolic motifs that often have a hydrophobic residue (tyrosine, leucine) as a crucial determinant (Casanova *et al.,* 1991; Matter *et al.,* 1992; Weise *et al.,* 2010).

Apical sorting relies on modifications such as a glycosylphosphatidylinositol (GPI) anchor, or *N*- or *O*-linked glycosylation (Brown *et al.,* 1989; Lisanti *et al.,* 1988; Lisanti *et al.,* 1989; Yeaman *et al.,* 1997). Other factors in maintaining polarity include the rearrangement of the cytoskeleton and polarity markers (such as the proteins Crumbs, Par, and Scribble; reviewed in St Johnston and Ahringer, 2010).

Neuronal trafficking

Neurons are highly specialized—and highly polarized—cell types consisting of a cell body with dendrites and an axon, where signals are received at the dendrites and propagated down the axon. The dendrites can be further polarized with apical versus basolateral regions. Developing neurons establish polarity early and then maintain it throughout their long lifetime by regulation of cytoskeleton dynamics and membrane trafficking (reviewed in Bentley and Banker, 2016). As with polarized trafficking to the basolateral membrane in epithelial cells, tyrosine- and dileucine-based motifs appear to be important for dendritic trafficking (Brown *et al.,* 1997; Rivera *et al.,* 2003). With the axon extending far from the cell body, the long distance would not be efficiently trafficked for high turnover proteins. As such, the axon contains translation machinery and produces certain secretory proteins locally (Gonzalez *et al.,* 2016).

Vesicle formation

Vesicle formation must be precisely organized to ensure all the necessary components are packaged—selecting only the correct cargoes, excluding others, and incorporating the correct v-SNARE protein for fusion at the acceptor membrane. The well-defined clathrin coated vesicle (CCV) model consists of a three-layered vesicle: the membrane and cargoes make up the inner layer, adaptor and accessory proteins comprise the middle layer,

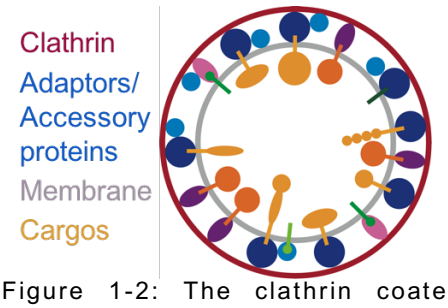


Figure 1-2: The clathrin coated vesicle model is comprised of three layers: the internal layer of cargoes and the membrane, the middle layer of adaptors and accessory proteins, and the external layer of clathrin, which forms a cage to support the vesicle. Figure courtesy of the Owen lab.

and the coat scaffold—clathrin—forms the outer layer (Figure 1-2). However, recent electron tomography work shows that the adaptor layer and structural scaffold layer are not separate in the related coatomer protein I (COPI) coat (Dodonova *et al.,* 2015), which suggests that how coats form on vesicles is not conserved.

Vesicle formation can be broken down into several steps: coat assembly, mechanical deformation, and vesicle scission. The vesicle is then transported along the cytoskeleton to the acceptor membrane, where it is recognized by Rabs/tethers and fused via the SNARE machinery (Figure 1-3).

The formation of the coat typically consists of recruitment of proteins that perform one or more of the following jobs: selecting cargo, binding the membrane, or structurally supporting the vesicle shape. SNARE and cargo selection occurs through the recognition of short, linear motifs, or surfaces on these proteins by the coat proteins—the adaptors, accessory proteins, and scaffolds (reviewed in Kelly and Owen, 2011).

Coats appear in two forms: spherical vesicular coats (most commonly COPI-, COPII-, and clathrin-coated vesicles) and tubular coats (such as the retromer complex). COPII can also form tubular carriers in addition to the traditional spherical vesicles. While COPI functions *en bloc* (the α/β'/ε B-subcomplex is the scaffold and the β/δ/γ/ζ F-subcomplex functions as an adaptor complex; Hara-Kuge *et al.*, 1994), clathrin coated vesicles utilize numerous adaptor proteins. Two of the most common adaptor families are the protein family of heterotetrameric complexes (discussed below) and the monomeric Golgi-localized, γ-ear containing, ADP ribosylation factor binding family (GGAs; reviewed in Robinson, 2004).

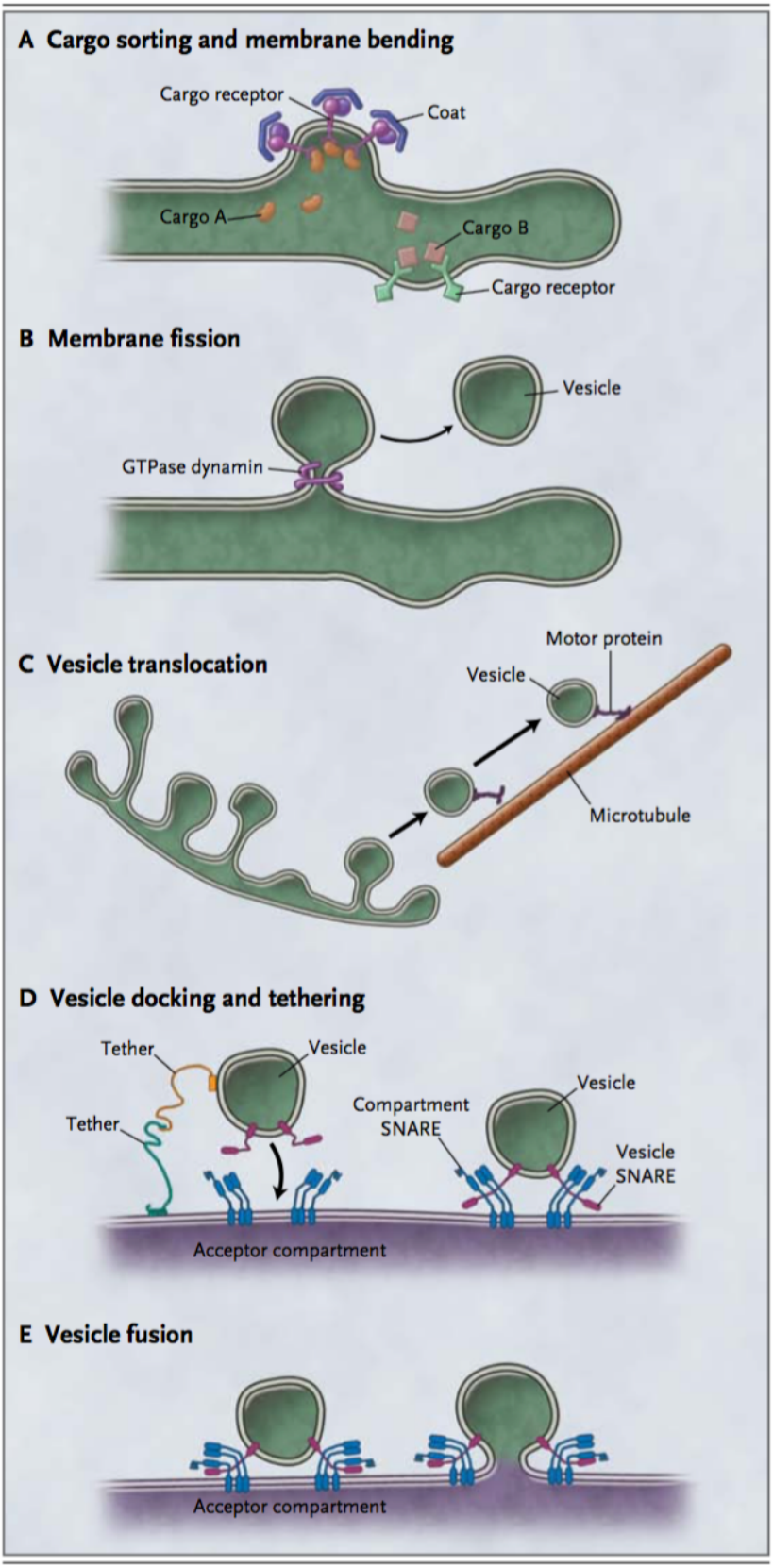


Figure 1-3: The steps of vesicle formation. A) Coat proteins select cargo, the clustering of which contributes to membrane bending. B) The vesicle buds off when the membrane is pinched off by the work of dynamin. C) The vesicle moves along the cytoskeleton with the help of motor proteins. D) The vesicle is recognized at the acceptor membrane first by tethers, which allow it to dock. E) Then the SNARE complex drives fusion of the vesicle with the donor membrane. Figure from DeMatteis and Luini, 2011.

The concentration of cargoes and coat proteins on one surface of the membrane leads to deformation of the membrane, which is further shaped by membrane bending proteins, such as those in the F-BAR family. The scaffold of the coat complex, like the well-known clathrin molecule, forms lattices or cages that stabilize and support the extreme curvature of the membrane. Clathrin, COPI, and COPII all consist of α-solenoid and β-propeller domains but assemble differently on membranes (Figure 1-4; Dodonova *et al.*, 2015; Stagg *et al.*, 2008; Fotin *et al.*, 2004). The shared structural motifs suggest a common evolutionary origin; novel scaffolds may feature similar domain structures as well. Finally, scission results in the vesicle being pinched off from the donor membrane. In endocytosis, scission is carried out by the large GTPase dynamin, which couples GTP hydrolysis with restriction of the vesicle neck, creating a force that pinches off the vesicle (reviewed in Ferguson and De Camilli, 2012; Kirchhausen *et al.,* 2014; Robinson, 2015).

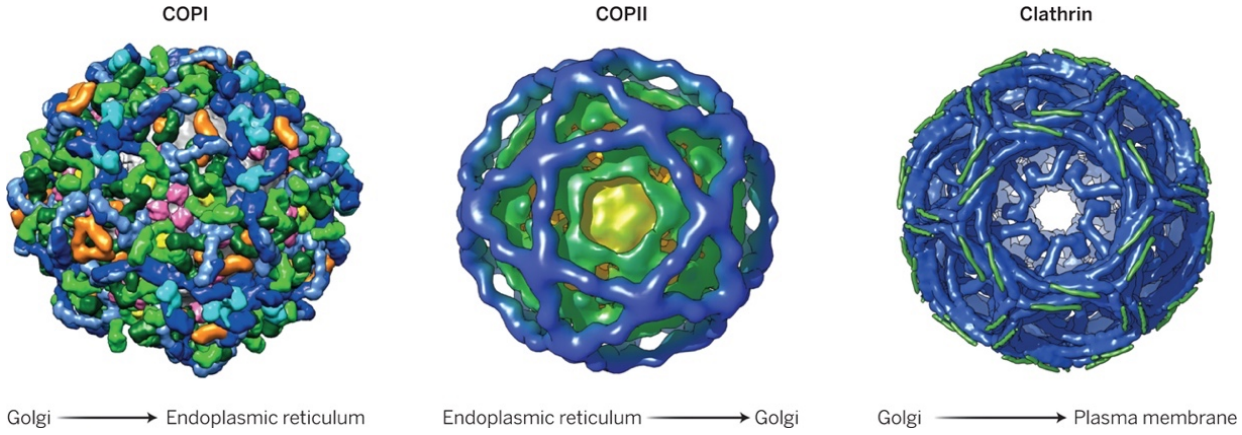


Figure 1-4: EM models reveal COPI, COPII, and clathrin assemble differently on membranes, despite sharing similar folds and a common evolutionary history. COPI [EMD-2985, EMD-2986, EMD-2987, EMD-2988, EMD-2989, and PDB-5A1U, PDB-5A1V, PDB-5A1W, PDB-5A1X, PDB-5A1Y (Dodonova *et al.,* 2015)], COPII [Electron Microscopy Data Bank accession code EMD-1511 (Stagg *et al.,* 2008)] and, clathrin [PDB-1XI4 (Fotin *et al.,* 2004)] are shown. Figure adapted from Nobel and Stagg, 2015.

The adaptor protein family

Originally identified as 'assembly polypeptides' that promoted clathrin assembly (Zaremba and Keen, 1983), further purification determined the assembly polypeptides were two distinct complexes (AP1 and AP2; Pearse and Robinson, 1984; Keen, 1987). Today AP stands for adaptor protein, and the family consists of seven structurally and evolutionarily related protein complexes: APs 1-5, the F-subcomplex of coat protein complex I (COPI), and TSET (reviewed in Robinson, 2004; Owen *et al.,* 2004; Hirst *et al.,* 2013). Members of the AP family serve as adaptors for vesicle formation as they interact with cargo, accessory proteins, and scaffolds. As such, they are hubs of protein-protein interactions. COPI traffics in a retrograde manner within the Golgi and from the Golgi to the ER (reviewed in Arakel and Schwappach, 2018). TSET, an ancient complex that may be the evolutionary link between COPI and the APs, likely functions at the plasma membrane (Hirst *et al.,* 2014). The APs participate in the late secretory (post-Golgi) and endocytic pathways (Figure 1-5).

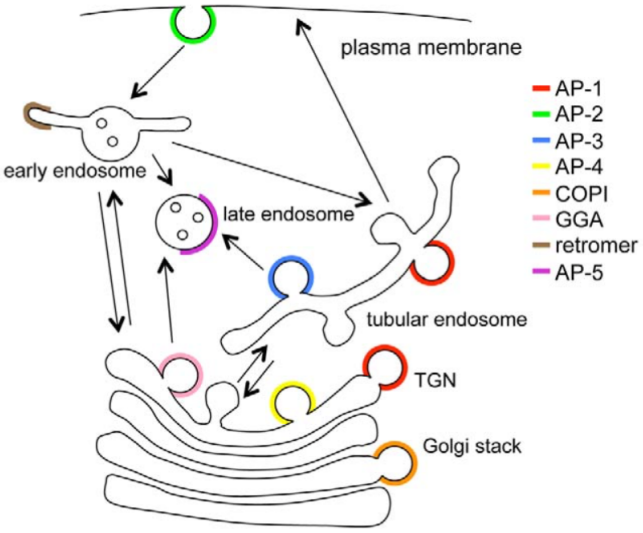


Figure 1-5: Diagram of post-Golgi trafficking pathways from Hirst et al., 2013. The AP family members, with the exception of TSET, are featured. Also included are the monomeric GGA adaptors which interact with clathrin and sort from the TGN, as well as the tubular carrier retromer, which localizes to endosomes.

Structurally, the AP complexes are heterotetramers with two large (α/γ/δ/ε/ζ, β; ~100 kDa), one medium (μ; ~50 kDa), and one small (σ; ~20 kDa) subunits (Robinson, 2004; Owen *et al.,* 2004). The large subunits consist of a N-terminal trunk domain connected to C-terminal appendage domains by an unstructured flexible linker, or 'hinge' (Figure 1-6). The medium and small subunits and the N-terminal trunks of the large subunits make up the

core complex (Collins *et al.,* 2002). The AP1/2 core typically interacts with the membrane and cargoes, while the appendage domains recruit additional accessory proteins. A clathrin binding box motif (LΦxΦ[DE], where x is any amino acid, Φ is a bulky hydrophobic) is harbored in the β hinge of AP1/AP2/AP3 (Dell'Angelica *et al.,* 1998; ter Haar *et al.,* 2000; reviewed in Kirchhausen *et al.,* 2000). Once recruited to the donor membrane by phosphoinositides or members of the small GTPase ADP ribosylation factor (Arf) family, the APs undergo a conformational change from an inactive, closed form in the cytosol to an active, open form capable of binding cargo motifs (Collins *et al.,* 2002; Dodonova *et al.,* 2015; Jackson *et al.,* 2010; Ren *et al.,* 2013; Yu *et al.,* 2012).

Two main cargo selection motifs have been identified and well characterized in the AP1 and AP2 systems: YXXΦ and [DE]XXXL[L/I] motifs, where X is any amino acid, and Φ is a bulky

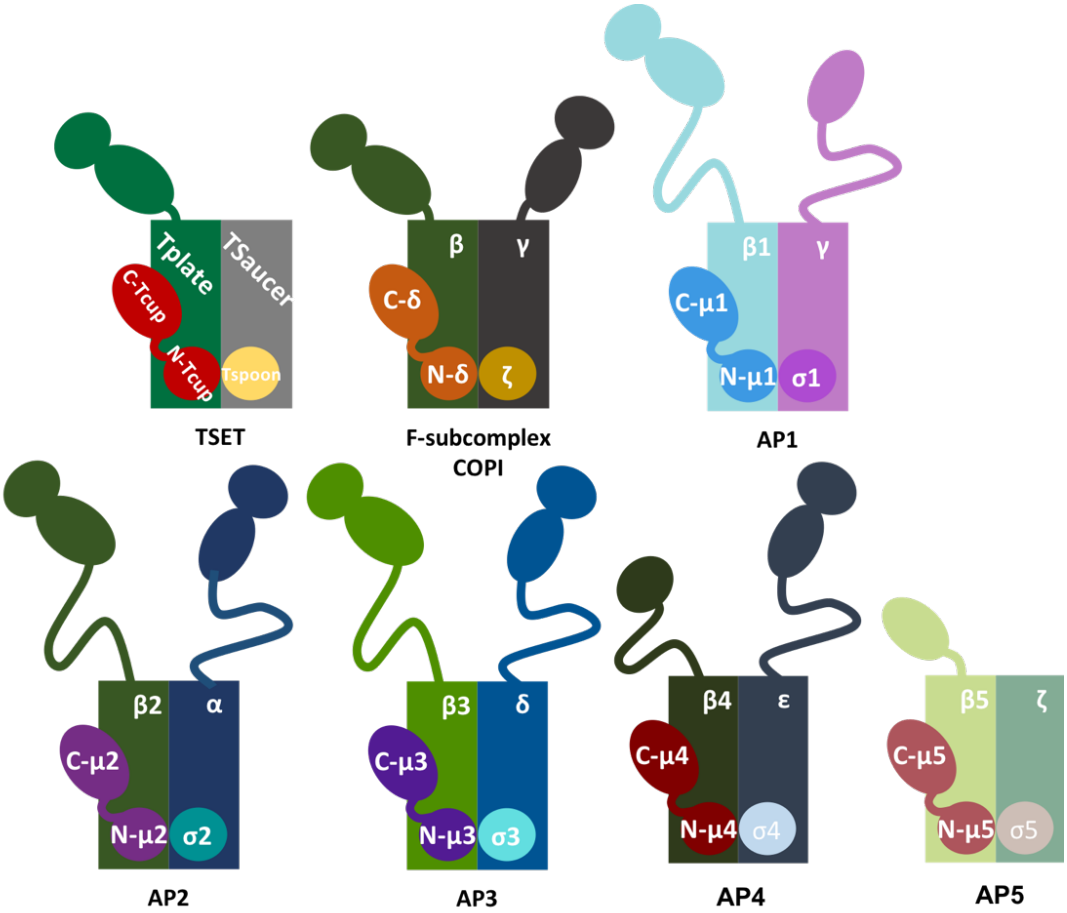


Figure 1-6: Diagram of AP family members. The 'core complex' consists of the N-terminal trunk domains of the two large subunits, as well as the medium and small subunits. The C-terminal appendage domains are connected to the trunk domains by flexible hinges.

hydrophobic residue (Ohno *et al.,* 1996; Rapaport *et al.,* 1998). YXXΦ motifs are recognized by the μ subunits of AP complexes (Owen and Evans, 1998), while [DE]XXXL[L/I] motifs bind the γ/α/δ/ε/ζ-σ subunit interface (Kelly *et al.,* 2008). Binding these motifs appears to be a conserved property of AP complexes (reviewed in Bonifacino and Traub, 2003). However, AP4 binds traditional YXXΦ motifs only weakly, and μ5 does not contain the conserved patch that binds YXXΦ motifs (Hirst *et al.,* 1999; Aguilar *et al.,* 2001; Hirst *et al.,* 2011).

AP1 traffics from the TGN

Discovered along with AP2, AP1 localizes to the TGN and traffics to the endosomes. Several AP1 isoforms exist with different tissue expression levels; the epithelial AP1-B isoform participates in basolateral trafficking of low density lipoprotein receptor (LDLR) and transferrin receptor (TfR; Folsch *et al.,* 1999). While it has a conserved phosphoinositide binding site on the γ subunit with a preference for phosphatidylinositol 4-phosphate (PI(4)P), which is enriched at the TGN, PI(4)P is not sufficient for recruitment to the membrane (Wang *et al.,* 2003; Heldwein *et al.,* 2004; reviewed in Di Paulo and De Camilli, 2006). Indeed, the Arf1 GTPase is necessary for recruiting AP1 to the TGN membrane (Stamnes and Rothman, 1993; Traub, 1993). As a GTPase, Arf1 has an active, GTP-bound conformation and an inactive, GDP-bound conformation. A guanine exchange factor (GEF) facilitates the GDP to GTP conversion, while a GTPase activating protein (GAP) catalyzes GTP hydrolysis to GDP. In its active conformation, an N-terminal myristolated amphipathic helix is exposed, which inserts into the membrane as an anchor (reviewed in Donaldson and Jackson, 2011). In this way, Arf1(GTP) can recruit AP1 to the TGN for vesicle initiation. Structural work by the Hurley group revealed two binding sites for Arf1 on AP1. The guanine nucleotide sensitive switch I/II regions of Arf1 interacts with β1, while a binding site on the 'back side' of ArfI interacts with γ (Ren *et al.,* 2013). The γ-Arf1 interaction is not necessary for membrane recruitment, but is necessary for complete allosteric activation. This work provides not only a model for membrane recruitment, but also a mechanism for stabilizing the open conformation of AP1, suggesting Arf1 plays a role in both membrane recruitment and

the closed to open AP conformational change necessary to expose phosphoinositide and cargo binding sites (Figure 1-7).

AP2 is essential in clathrin mediated endocytosis

AP2 was the first adaptor protein identified, characterized by Scottie Robinson in Barbara Pearse's group (EMBO J 1984). Much of what we know about clathrin and vesicle trafficking in general comes from the study of AP2 and clathrin mediated endocytosis (CME). AP2 localizes to the plasma membrane and interacts with clathrin to internalize many proteins responsible for maintaining the plasma membrane, transducing signals across the membrane, and participating in synaptic transmission. AP2 is recruited to the membrane by binding phosphatidylinositol 4,5-bisphosphate ($PI(4,5)P_2$; Jost *et al.,* 1998; Gaidarov and Keen, 1999; Rohde *et al.,* 2002; Motley *et al.,* 2006), and possibly Arf6 (Krauss *et al.,* 2003; Paleotti *et al.,* 2005). Interactions with $PI(4,5)P_2$ and cargo motifs trigger and stabilize the allosteric open form of the complex (Figure 1-7). In the cytosol, the closed form occludes cargo binding sites, hides the clathrin binding box on the β2 subunit, and therefore prevents non-productive polymerization away from the membrane (Kelly *et al.,* 2014).

AP2 μ2 harbors a threonine (Thr156) in a loop region that becomes structured upon shifting to the open conformation (Jackson *et al.,* 2010). This threonine can be phosphorylated by AAK1 (α-appendage binding kinase 1), which promotes cargo interactions, and likely contributes to vesicle initiation (Höning *et al.,* 2005; Ricotta *et al.,* 2002; Olusanya *et al.,* 2001). This threonine is conserved in AP1 μ1; thus, the phosphorylation mechanism may act in AP1 vesicle formation as well. Recent work from the Schmid group sought to determine the hierarchy of events that occurs during clathrin coated vesicle formation (Figure 1-7; Kadlecova *et al.*, 2017). They found higher levels of μ2 Thr156 phosphorylation in cells expressing a μ2 construct deficient for $PI(4,5)P_2$ binding compared to wild type, but decreased phosphorylation in cells expressing an α construct that was $PI(4,5)P_2$ binding deficient. They were unable to study the effects of ablating μ2 phosphorylation due to poor construct expression, but concluded that μ2 phosphorylation can

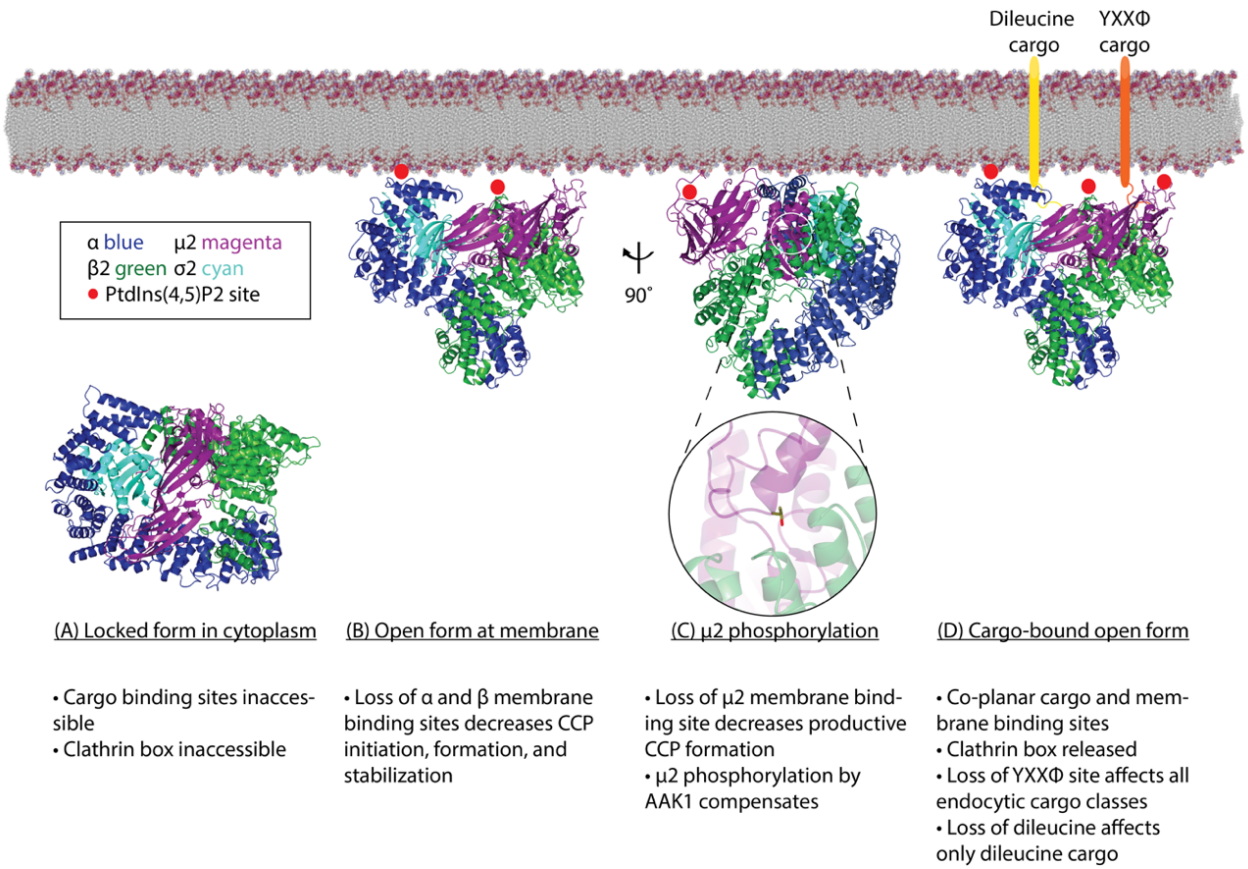


Figure 1-7: An overview of AP2 allostery and the hierarchy of binding events, as discovered in Kadlecova et al., 2017. This figure is printed in Frazier and Jackson, 2017. (A) Structural models suggest AP2 exists in a closed form in the cytoplasm, preventing futile interactions with clathrin in the absence of membrane-embedded cargo. (B) The authors demonstrate AP2 recruitment to the plasma membrane depends on the PIP2 binding sites on the α and β2 subunits (shown as red filled circles), as predicted by x-ray crystal structures and biophysical data. When these sites are absent, coated pit initiation and stabilization is significantly decreased, as is productive vesicle formation. (C) Loss of the PIP2 binding site on μ2 (red filled circle) increases coated pit initiation events but decreases the number of productive structures. The authors suggest μ2 phosphorylation by AAK1 compensates for loss of the PIP2 binding site. (D) At the plasma membrane, all phosphoinositide and cargo binding sites are coplanar on the same surface of AP2, and the clathrin box motif is released. The authors show that loss of the YxxΦ binding site (orange cargo) results in decreased uptake of all classes of endocytic cargo, whereas loss of the dileucine site (yellow cargo) affects only dileucine cargo.

trigger the AP2 conformational change and provide some compensation if phosphoinositide binding is not robust. However, since AAK1 is recruited by clathrin, μ2 phosphorylation cannot overcome the AP2 recruitment deficit if the α binding site is compromised (Kadlecova *et al.,* 2017). Thus, the PI(4,5)P_2 binding events can be placed at the top of the hierarchy, followed by the μ2 phosphorylation event. Interestingly, the importance of the cargo binding sites varies: loss of the YXXΦ site affects all classes of AP2 endocytic cargoes, including FXNPXY cargo and the EGF

receptor, while loss of the dileucine motif affects only dileucine cargo uptake. This suggests the YXXΦ site has a broader role in regulating vesicle formation.

AP3 functions in the endolysosomal system

Unlike AP1 and AP2, AP3 appears to function in clathrin-dependent and clathrin-independent pathways, as numerous *in vitro* and *in vivo* studies have produced conflicting results depending on the system studied (Newman *et al.,* 1995; Simpson *et al.,* 1996; Dell'Angelica *et al.,* 1998; Faundez *et al.,* 1998; Drake *et al.,* 2000b). AP3 primarily traffics from tubular endosomes to late endosomes, lysosomes, and lysosome-related organelles (LROs), and is recruited to the membrane by Arf1. Mutations in AP3 cause Hermansky-Pudlak syndrome, which manifests because of defects with melanosome biogenesis and platelet dense granules, types of LROs.

AP3 has additional tissue-specific isoforms: β3b, μ3b, and σ3b, which are highly expressed in the brain. Interestingly, AP3 appears to have a role in sorting synaptic membrane proteins in neurons. The *mocha* (AP3 δ knockout) mouse line exhibits epilepsy and hyperactivity and shows a defect in trafficking ZnT3, a zinc transporter, which leads to a lack of zinc at synaptic terminals (Kantheti *et al.,* 2003; Salazar *et al.,* 2004a). Other synaptic membrane proteins mistrafficked in AP3 depletions include synaptic vesicle chloride channel 3 (CIC-3), vesicular glutamate transporter 1 (VGLUT1), vesicular GABA transporter (VGAT), and the SNARE VAMP7 (Salazar *et al.,* 2004b; Salazar *et al.,* 2005; Nakatsu *et al.,* 2004; Salazar *et al.,* 2006). AP2-dependent pathways carry out most synaptic vesicle trafficking, but under high frequency stimulation, AP3-dependent pathways begin to contribute (Voglmaier *et al.,* 2006).

AP4 and AP5 and their role in disease

AP4 and AP5 are the most recently discovered APs (Dell'Angelica *et al.,* 1999; Hirst *et al.,* 1999; Hirst *et al.,* 2011). While ubiquitously expressed, they are much less abundant than AP2. This lower expression may have contributed to difficulty in isolation and identification (Hirst *et al.,* 2012). Furthermore, AP4 and AP5 do not associate with clathrin, because they lack the clathrin

binding box in the β hinge. The large subunits of AP1/AP2/AP3 each contain C-terminal hinge regions followed by bilobal appendage domains. AP4 and AP5 appear to have slightly different structures from the other APs. The β4 appendage domain contains only one lobe, or subdomain, meaning it has less surface area to bind motifs (Hirst *et al.,* 1999; Dell'Angelica *et al.,* 1999). In AP5, ζ lacks the hinge and appendage domain entirely, while β5 lacks the unstructured hinge (Hirst *et al.,* 2011). While AP4 binds no known scaffolds, AP5 may interact with a novel coat. Two AP5 associated proteins, SPG11 (spactasin) and SPG15 (spastizin/ZFYVE26/FYVE-CENT), have predicted secondary structure elements like those in the clathrin heavy chain and α- and β'-COPI subunits (Hirst *et al.,* 2011). AP5 localizes to endosomes (Hirst *et al.,* 2011; Hirst *et al.,* 2018), while AP4 localizes to the TGN (Hirst *et al.,* 1998; Yap *et al.,* 2003).

These complexes are important in neurological development as mutations in AP4 and AP5 lead to hereditary spastic paraplegias (HSPs). Mutations in any subunit of AP4 lead to HSPs, and all AP4 subunits have been given spastic paraplegia gene numbers: SPG47 (β4), SPG50 (μ4), SPG51 (ε), SPG52 (σ4). So far, only one subunit of AP5 has been identified as HSP causative (SPG48, ζ), although mutations in the proposed scaffold proteins SPG11 and SPG15 do cause disease (hence their designation as spastic paraplegia genes; Hirst *et al.,* 2012). Hereditary spastic paraplegias are a complex group of clinically and genetically diverse disorders that cause progressive spasticity of the lower limbs. Mutations in trafficking proteins or membrane shaping/organelle developing proteins are a common link between the many types of HSPs (reviewed in Blackstone, 2012). The most common forms of HSPs are caused by mutations in spastin, an ATP-dependent microtubule severing protein, and atlastin-1, a GTPase that functions in the ER (Hazan *et al.,* 1999; Namekawa *et al.,* 2006). Cases of HSP due to AP4 and AP5 mutations are rare, because these mutations are recessive. Cases of HSP due to these mutations are identified more frequently in consanguineous families from the Middle East than elsewhere in the world (Abdollahpour *et al.,* 2015; Abou Jamra *et al.,* 2011; Bauer *et al.,* 2012; Lamichhane *et al.,* 2013; Schlipf *et al.,* 2014; Tan *et al.,* 2016; Tüysüz *et al.,* 2014). Recently, however, several

children in the US have been diagnosed with SPG47 mutations (Ebrahimi-Fakhari *et al.,* 2017). Due to the similarity between symptoms of HSPs and cerebral palsy, and the necessity of whole exome sequencing to identify the AP4/5 mutation, the number of cases of AP4- and AP5-related HSPs may be underestimated.

Further evidence of the importance of AP4 in the brain is its role in trafficking AMPA and δ2 glutamate receptors as well as amyloid precursor protein (APP; Burgos *et al.,* 2010; Matsuda *et al.,* 2008; Matsuda and Yuzaki, 2008; Yap *et al.,* 2003). While a knockout β4 mouse model displayed no overt phenotypes in the gait or development, the cellular morphology of neurons was affected (Matsuda *et al.,* 2008). The authors noted enlarged autophagosome structures in the axons, which contained aberrantly sorted AMPA receptors. The authors further noted that δ2 glutamate receptors and the low-density lipoprotein (LDL) receptors were also missorted from the somatodendritic domain to axons in β4 knockout conditions (Matsuda and Yuzaki, 2008).

Precisely how AP4 sorts cargoes is not currently well understood. The YXXΦ motif recognition by μ subunits is well established, but μ4 binds these motifs only weakly (Hirst *et al.,* 1999; Aguilar *et al.,* 2001). Work by the Hurley and Bonifacino groups identified a non-canonical YXXΦ motif, YKFFE, in APP by yeast two-hybrid (Y2H), and showed in cells that mutation of the motif or depletion of μ4 shifted APP localization from endosomes to the TGN (Burgos *et al.,* 2010).

Since the secretases responsible for cleaving APP have different localizations, the steady state localization of APP is a determinant of the amount of amyloidogenic Aβ formed. AP4 mediated transport of APP from the TGN to endosomes could therefore reduce the negative cleavage reaction. The AP4 mediated trafficking of APP is controversial, though, because the binding site identified on μ4 is not the canonical YXXΦ binding site and would only be available for binding the YKFFE motif in the ‘closed’ form of the AP4 complex (Figure 1-8). While the allosteric conformational change of APs at the membrane has only been demonstrated for AP1 and AP2, it is assumed to be a conserved feature of all AP complexes.

AP4 likely has a role in polarized sorting generally, not just in the nervous system. In epithelial cells, AP4 appears to sort cargoes basolaterally. When μ4 was depleted in MDCK cells, LDLR and MPR46 were missorted to the apical surface, indicating a role for AP4 in basolateral trafficking (Simmen *et al.,* 2002). New work from the Bonifacino, Robinson, and Kittler groups suggests AP4 might also have a role in sorting the autophagy protein Atg9a, as all groups independently identified Atg9a as an AP4 cargo (Mattera *et al.,* 2017; Davies *et al.,* 2017; Ivankovic *et al.,* 2017).

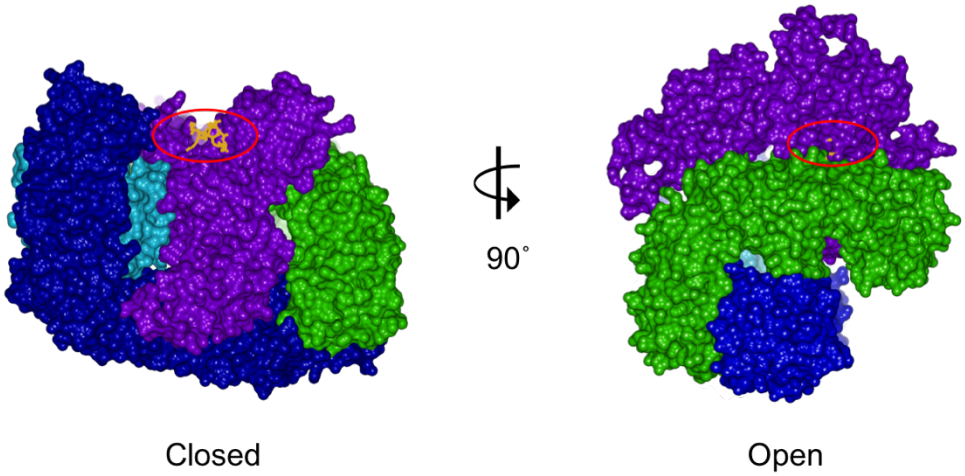


Figure 1-8: The μ4-YKFFE interaction in the context of membrane based allostery. The YKFFE motif binding pocket is only available in the closed AP conformation. C-μ4 with YKFFE bound (3I81; Burgos *et al.,* 2010) was superposed with the closed AP2 complex (2VGL; Collins *et al.,* 2002) and the open AP2 complex (2XA7; Jackson *et al.,* 2010). The subunits are shown in surface representation and colored according to the color scheme depicted in Figure 1-6. The YKFFE motif is colored gold, and highlighted by a red circle.

Accessory proteins in vesicle formation

Accessory proteins often bind additional cargoes or regulate coat assembly and disassembly, including AP180/CALM, auxilin/GAK, and epsin1. CALM and its neuronal form AP180 bind $PI(4,5)P_2$ through ANTH (AP180 N-terminal homology) domains, recruiting clathrin to the membrane and assisting in polymerization (Ford *et al.,* 2001). CALM also appears to have a role in regulating vesicle size (Zhang *et al.,* 1998; Miller *et al.,* 2015). Cyclin-G-associated kinase (GAK) and its tissue-specific form auxilin function as cofactors for Hsc70, recruiting this enzyme

to vesicles to promote the disassembly of clathrin and uncoating of vesicles during transport to donor membranes (Ungewickell *et al.,* 1995; Greener *et al.,* 2000).

The epsin family

Epsins, present across eukaryotes, are defined by their N-terminal domain (Epsin N-Terminal Homology, ENTH) and otherwise contain a largely unstructured C-terminus that harbors motifs for interacting with proteins such as the AP complexes, clathrin, and ubiquitin (Figure 1-9). Epsins may contribute to membrane curvature or may ensure incorporation of additional proteins, such as cargo and SNAREs, into forming vesicles.

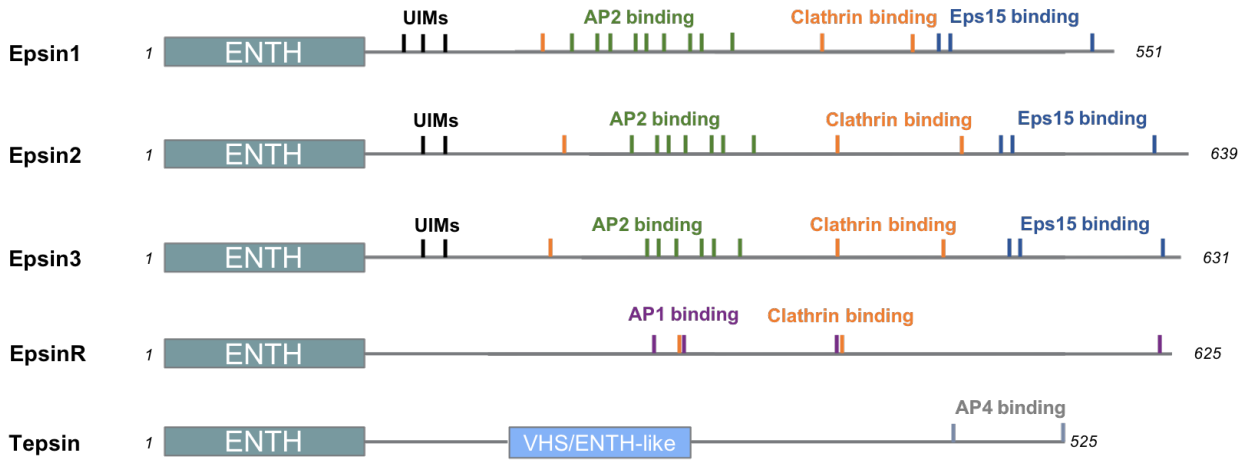


Figure 1-9: Domain structure of epsin family members. Epsin1/2/3 contain 2-3 UIMs, and variable copies of clathrin, AP2, and Eps15 binding motifs. EpsinR harbors clathrin and AP1 binding motifs. Tepsin has a second folded domain, the VHS/ENTH-like domain, as well as AP4 binding motifs.

Epsin1 (and its isoforms epsin2 and epsin3) interacts with AP2, clathrin, and ubiquitin through the C-terminus and with PI(4,5)P_2 through the ENTH domain (Ford *et al.*, 2002). It was originally identified as an Eps15 interacting protein, binding via NPF motifs in the C-terminus, hence the name epsin (Chen *et al.,* 1998). Via its ubiquitin interacting motifs (UIMs), epsin can bring additional cargoes into the forming vesicle; ubiquitination of proteins at the plasma membrane is a common internalization signal. Indeed, epsin1 appears to select for K63-linked polyubiquitin chains (Hawryluk *et al.,* 2006). Epsin1 is highly expressed in the brain (Chen *et al.,* 1998), and epsin3 is specifically expressed in keratinocytes (Spradling *et al.,* 2001).

Epsin-related protein (epsinR) binds AP1 and clathrin through the C-terminus, but does not bind Eps15, hence its naming as epsin-related (Mills *et al.,* 2003; Hirst *et al.,* 2003). The epsinR ENTH domain has a preference for PI(4)P, which explains its localization to the TGN. The ENTH domain of epsinR also interacts with the SNARE vti1b on the adjacent side to the membrane, which suggests that ENTH domains may have additional functions beyond binding lipids (Miller *et al.,* 2007; Chidambaram *et al.,* 2004; Hirst *et al.,* 2004). Indeed, follow up work using yeast homologues identified two more members of the vti1p (the yeast vti1b) endosomal SNARE complex as ENTH interactors and dependent on Ent3p (the yeast epsinR) for proper localization, which suggests a broader role for epsinR mediated trafficking of endosomal SNAREs (Chidambaram *et al.,* 2008).

Tepsin, the first AP4 accessory protein, contains an ENTH domain at its N-terminus but also a second domain downstream of the ENTH domain, named VHS/ENTH-like for its similarity to the ENTH/ANTH/VHS protein superfamily (discussed below). Unlike other epsins, tepsin lacks clathrin and ubiquitin binding motifs in its unstructured C-terminus. While the other epsin family members are recruited independently to their membrane of action, tepsin requires AP4 to be recruited to the membrane (Borner *et al.,* 2012). This suggests the function of tepsin may be different than the other epsins. Furthermore, Borner *et al.* used BLAST searches to demonstrate that AP4 and tepsin are highly linked: when AP4 is lost from an organism, tepsin is coordinately lost as well.

ENTH/ANTH/VHS protein superfamily

This familial protein fold is all α-helical superhelical fold, typically consisting of eight helices. While ENTH/VHS domains are roughly the same size, ANTH domains have about twice

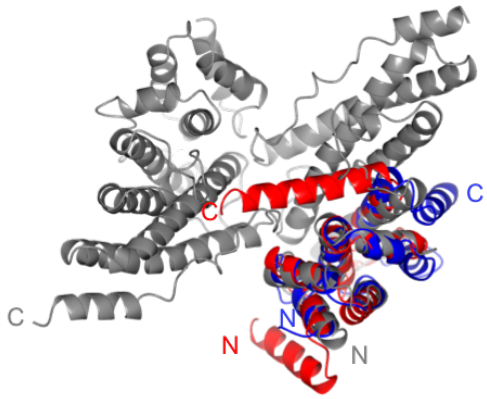


Figure 1-10: Comparison of ENTH/ANTH/VHS domains by superposition of representative domains. ENTH/ANTH/VHS domains share a common helical core (helices 1-7). Helix 8 packs differently against the core depending on the domain. ANTH domains are much larger than ENTH/VHS domains, while ENTH domains contain an additional helix 0. The N- and C-termini are labeled for each domain. The VHS domain (GGA1; 1JWF) is blue, the ENTH domain (epsin1; H0A) is red, and the ANTH domain (CALM; 3ZYK) is gray.

as many residues and more helices (Figure 1-10). However, an eight α-helical core is revealed when superposing ANTH and ENTH domains. ENTH domains also possess an amphipathic helix 0 which becomes ordered in the presence of lipids and inserts into the membrane, potentially driving membrane curvature (Ford *et al.,* 2002). However, Stachowiak *et al.* have demonstrated that the effect of protein crowding on a membrane (highly reminiscent of vesicle formation) can create membrane curvature regardless of special membrane bending mechanisms (Stachowiak *et al.,* 2010; Stachowiak *et al.,* 2012). The ENTH helix 0, along with an adjacent pocket of basic residues, explains the ability of ENTH domains to bind phosphatidylinositol species (Mills *et al.,* 2003). ANTH domains also bind phosphatidylinositol species, but through a different mechanism: a basic surface motif (KxxKxH) interacts with the lipid headgroup (Ford *et al.,* 2001; Mills *et al.,* 2003).

VHS (Vps27, Hrs, Stam) domains are found in at least 60 proteins related to trafficking, always at the very N-terminus of the protein, which suggests an important function related to this positioning. VHS and ENTH domains align very well, with only about 1.8 Å RMSD (root mean square deviation of the protein C-α backbone). VHS domains primarily function to sort cargoes. In GGAs, VHS domains recognized acidic dileucine motifs in receptors like sortilin and cation-independent mannose 6-phosphate receptor (Nielsen *et al.,* 2001; Puertollano *et al.,* 2001; Takatsu *et al.,* 2001; Misra *et al.,* 2003). In ESCRT-0, which selects cargo for lysosomal degradation via the multivesicular body pathway (Raiborg *et al.,* 2008), Hrs and STAM VHS domains bind ubiquitin (Mizuno *et al.,* 2003; Hong *et al.,* 2009; Ren *et al.,* 2010). Ren *et al.*

concluded that binding ubiquitin is a general property of VHS, as they detected an interaction between ubiquitin and other VHS domains in addition to the ESCRT-0 proteins (Ren *et al.,* 2010). Most VHS domain containing proteins have additional ubiquitin binding domains (UBDs) or UIMs downstream of the VHS domain, suggesting the relatively weak VHS-ubiquitin interaction could be strengthened by avidity. The multiple interactions between ubiquitin and VHS domain containing proteins likely confer selectivity for different types of ubiquitin chains (Ren *et al.,* 2010).

Research objective

This research seeks to understand the interaction between tepsin and AP4 and to characterize tepsin structurally and functionally. Chapter II details my work identifying a C-terminal motif in tepsin that interacts with the β subunit of AP4. I used isothermal titration calorimetry, circular dichroism, and nuclear magnetic resonance chemical shift perturbation experiments to identify the motif, quantify the binding affinity, and map the binding surface on β4. This was the first structural and biochemical description of any AP4 accessory protein. Chapter III focuses on the structure and function of the two folded domains in tepsin, the ENTH domain and the VHS/ENTH-like domain. Chapter IV contains my work on binding partners for the tepsin domains using commercial yeast two-hybrid screens followed by biochemical validation and cell culture studies. Chapter V features a discussion of this work as well as future directions for the study of AP4 and tepsin.

CHAPTER II

MOLECULAR BASIS FOR THE INTERACTION BETWEEN AP4 β4 AND ITS ACCESSORY PROTEIN, TEPSIN

Meredith N. Frazier[1,2], Alexandra K. Davies[3], Markus Voehler[2,4], Amy K. Kendall[1,2], Georg H. H. Borner[5], Walter J. Chazin[2,4], Margaret S. Robinson[3], and Lauren P. Jackson[1,2]

[1]Department of Biological Sciences, Vanderbilt University, Nashville, TN, USA
[2]Center for Structural Biology, Vanderbilt University, Nashville, TN, USA
[3]Cambridge Institute for Medical Research, Department of Clinical Biochemistry, University of Cambridge, Hills Road, Cambridge, United Kingdom
[4]Department of Biochemistry and Chemistry, Vanderbilt University, Nashville, TN, USA
[5]Max Planck Institute of Biochemistry, Department of Proteomics and Signal Transduction, Martinsried, Germany
Correspondence to lauren.p.jackson@vanderbilt.edu

This article has been published under the same title in *Traffic*, 2016 Apr;17(4):400-15.

Synopsis

The adaptor protein (AP) complex family mediates membrane trafficking events, but the molecular mechanisms of the non-clathrin AP4 coat remain poorly understood. We identify a conserved sequence in the AP4 accessory protein, tepsin, that directly interacts with the β4 appendage domain, and we map the tepsin binding site on the β4 surface. Mutations of key residues in the tepsin sequence or on β4 demonstrate their importance for binding both *in vitro* and in cultured cells. These data provide the first detailed molecular glimpse of how AP4 interacts with an accessory protein.

Abstract

The adaptor protein 4 (AP4) complex (ε/β4/μ4/σ4 subunits) forms a non-clathrin coat on vesicles departing the *trans*-Golgi network (TGN). AP4 biology remains poorly understood, in stark contrast to the wealth of molecular data available for the related clathrin adaptors AP1 and AP2. AP4 is important for human health because mutations in any AP4 subunit cause severe neurological problems, including intellectual disability and progressive spastic para- or tetraplegia. We have used a range of structural, biochemical, and biophysical approaches to determine the molecular basis for how the AP4 β4 C-terminal appendage domain interacts with tepsin, the only known AP4 accessory protein. We show that tepsin harbors a hydrophobic sequence, LFxG[M/L]x[L/V], in its unstructured C-terminus, which binds directly and specifically to the C-terminal β4 appendage domain. Using NMR chemical shift mapping, we define the binding site on β4 appendage by identifying residues on the surface whose signals are perturbed upon titration with tepsin. Point mutations in either the tepsin LFxG[M/L]x[L/V] sequence or in its cognate binding site on β4 abolish binding *in vitro*. In cells, the same point mutations greatly reduce the amount of tepsin that interacts with AP4. However, they do not abolish the binding between tepsin and AP4 completely, suggesting the existence of additional interaction sites between AP4 and

tepsin. These data provide one of the first detailed mechanistic glimpses at AP4 coat assembly and should provide an entry point for probing the role of AP4 coated vesicles in cell biology, and especially in neuronal function.

Introduction

Large coat protein complexes play central roles in many membrane trafficking pathways by driving vesicle coat formation at specific organelle membranes. The mammalian AP (Assembly Polypeptide) adaptor protein complexes (APs 1–5, COPI F-subcomplex) are a family of heterotetrameric complexes that recognize membrane components, transmembrane protein cargo, and additional machinery required for vesicle coat assembly. While the clathrin adaptors AP1 and AP2 are well-characterized mechanistically and functionally (Owen *et al.,* 1998; Owen *et al.,* 1999; Traub *et al.,* 1999; Owen *et al.,* 2000; Collins *et al.,* 2002; Heldwein *et al.,* 2004; Kelly *et al.,* 2008; Jackson *et al.,* 2010; Kelly *et al.,* 2014; Boucrot *et al.,* 2010; Kirchausen *et al.,* 2014; Antonescu *et al.,* 2011; Hirst *et al.,* 2012; Ren *et al.,* 2013; Shen *et al.,* 2015), the molecular structures and cellular functions of non-clathrin adaptors like AP4 and AP5 remain poorly understood. The AP4 complex (ε/β4/μ4/σ4 subunits) (Figure 2-1A) is recruited to the *trans*-Golgi network (TGN) by the small GTPase, Arf1, in its GTP-bound state (Boehm *et al.,* 2001). AP4 has been implicated in polarized cargo sorting in both epithelial cells (Simmen *et al.,* 2002) and neurons (Yap *et al.,* 2003; Matsuda *et al.,* 2008), and the C-terminus of μ4 can bind a YKFFE motif found in amyloid precursor protein (APP; Burgos *et al.,* 2010). Although AP4 is ubiquitously expressed (Hirst *et al.,* 1999; Hirst *et al.,* 2013), the most striking phenotypes are associated with the brain. A β4 knockout mouse exhibits mis-sorting of LDL, AMPA, and δ2 glutamate receptors from somatodendrites to axons (Matsuda *et al.,* 2008) but shows no further significant anatomical abnormalities. In contrast, human patients with mutations in any one of the four AP4 subunits (Abrou Jamra *et al.,* 2011; Abdollahpour *et al.,* 2015; Hardies *et al.,* 2015; Tüysüz *et al.,* 2014)

suffer from a severe “AP-4 deficiency syndrome” characterized by early onset of severe intellectual disability, growth retardation, stereotypic laughter, progressive spasticity and an inability to walk (Abou Jamra *et al.,* 2011; reviewed in Hirst *et al.,* 2012). AP4 thus likely plays a key role in neurological development and function, but the underlying mechanism remains unclear. It is therefore important to identify and understand the machinery required to form an AP4 coat, specifically the nature of the full complement of proteins in this coat, in order to uncover both the fundamental cellular function of AP4 and its role in human disease.

Unlike the AP1 and AP2 clathrin adaptors, an understanding of AP4 biology using traditional genetic and biochemical approaches has been hampered for two main reasons. During evolution, AP4 was lost in multiple eukaryotic lineages (Field *et al.,* 2007; Hirst *et al.,* 2014), including model organisms like yeast, worms, and flies. In addition, AP4 is much less abundant than either AP1 or AP2 (Hirst *et al.,* 2012). This has made standard fractionation-based proteomic analysis of AP4 vesicles unfeasible. However, progress was made using a comparative proteomic profiling method in which coated vesicle fractions were prepared under different conditions (Borner *et al.,* 2012). Similar patterns of enrichment across different vesicle preps were used to predict association, and this led to the identification of tepsin as the first AP4 accessory protein. Full-length tepsin was shown to bind specifically to the β4 C-terminal appendage domain (residues 612-739; Borner *et al.,* 2012).

Tepsin (Figure 2-1A) is a member of the epsin family of mammalian post-Golgi trafficking proteins, which includes the plasma membrane epsins1–3 and TGN/endosomal epsinR. All epsins contain an epsin N-terminal homology (ENTH) domain followed by mostly unstructured C-termini. Tepsin is distinct among family members in possessing a second folded structural domain, predicted to be most similar to a VHS domain (Borner *et al.,* 2012). In clathrin coated vesicles (CCVs), epsins function as accessory proteins by binding phosphoinositides via their ENTH domains (Ford *et al.,* 2002; Hirst *et al.,* 2003; Mills *et al.,* 2003); recognizing cargo via ubiquitin interacting motifs (UIMs) in epsins1–3 (Oldham *et al.,* 2002; Hawryluk *et al.,* 2006*)* or

folded structural domains in epsinR (Miller *et al.,* 2007); weakly binding endocytic SNAREs (Messa *et al.,* 2014); and interacting with clathrin (Rosenthal *et al.,* 1999; Drake *et al.,* 2000a) and either AP1 or AP2 via short linear amino acid motifs (Owen *et al.,* 1999; Rosenthal *et al.,* 1999; Praefcke *et al.,* 2004) or short stretches of secondary structure (Edeling *et al.,* 2006a). However, quantitative proteomics (Borner *et al.,* 2012) suggest epsin1 is present at only very low levels in endocytic CCVs, while tepsin is a major component in AP4 vesicles. Tepsin lacks characterized motifs for binding ubiquitin, clathrin, and AP1 or AP2. Furthermore, tepsin requires AP4 for its membrane recruitment (Borner *et al.,* 2012); epsins1–3 are recruited to membranes directly through binding phosphoinositides (Ford *et al.,* 2002).

Since tepsin is the first established AP4 accessory protein, we set out to identify and characterize the mechanism of the interaction between β4 and tepsin in order to obtain the first mechanistic information about AP4 coat formation. We show the β4 C-terminal appendage domain interacts specifically with a short, hydrophobic sequence in the unstructured tepsin C-terminus; this sequence is one of the first candidate AP4-specific binding motifs. The interaction is specific to the β4 C-terminus, because the tepsin motif cannot bind the β appendage domains of AP1 or AP2. Mutation of conserved residues in the tepsin motif abolishes binding to β4 *in vitro*. We use structural and biophysical methods combined with mutagenesis to identify and confirm key tepsin binding residues on the surface of the β4 appendage domain. Finally, we demonstrate the functional relevance of this interaction in cultured cells. Specific point mutations in either the β4 appendage domain or the tepsin motif greatly reduce the amount of tepsin bound to AP4 in HeLa cell lines. However, the mutations do not completely abolish the interaction between tepsin and AP4 and do not prevent the recruitment of tepsin to TGN membranes. Together, these data uncover one mechanism by which AP4 interacts with tepsin and further suggest the presence of at least one additional binding site on AP4 for tepsin in cells.

Results

β4 appendage domain binds a conserved hydrophobic motif in the tepsin C-terminus

Previous work demonstrated the β4 C-terminal appendage domain interacts directly with full-length GFP-tagged tepsin (Borner *et al.,* 2012). In other interactions between APs and epsins, the unstructured epsin C-terminus harbors one or more short amino acid motifs that directly bind an AP appendage domain. We created a panel of short GST-tagged fusion constructs corresponding to unstructured regions of tepsin; this included the region between the ENTH and VHS-like folded domains, as well as short regions of the entire C-terminus (Figure 2-1B). GST pulldown experiments using these recombinant purified GST-tepsin and β4 appendage domain constructs identified a stretch of ~50 amino acid residues in the tepsin C-terminus (residues 450-500) that binds β4 *in vitro* (Figure 2-1B). This interaction was visible by Coomassie staining on an

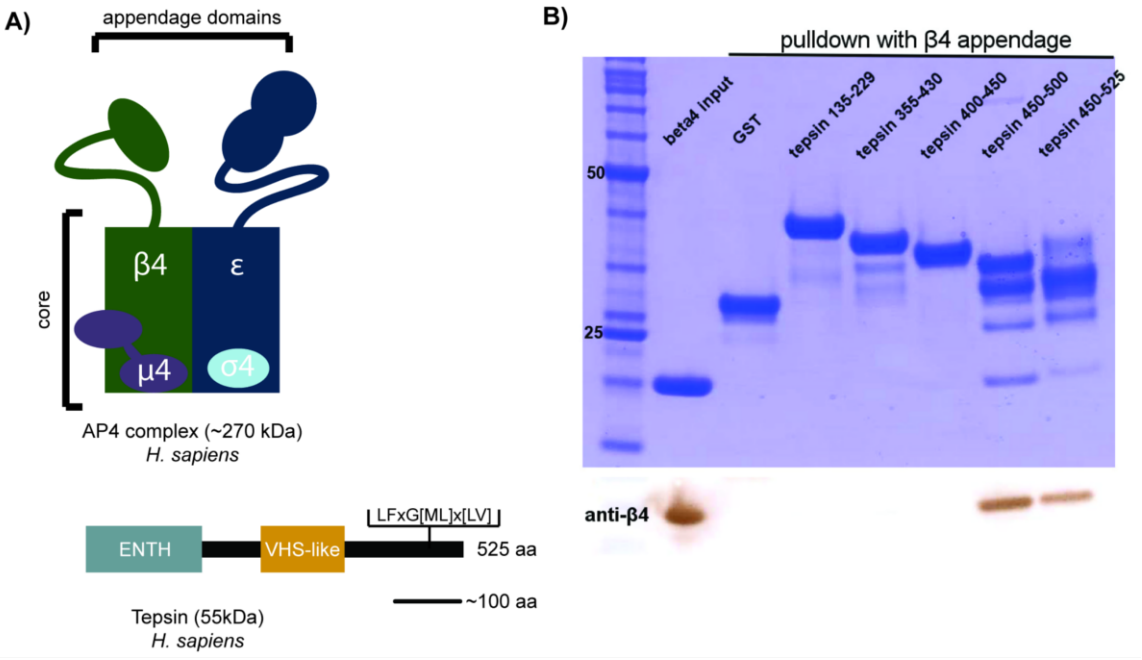
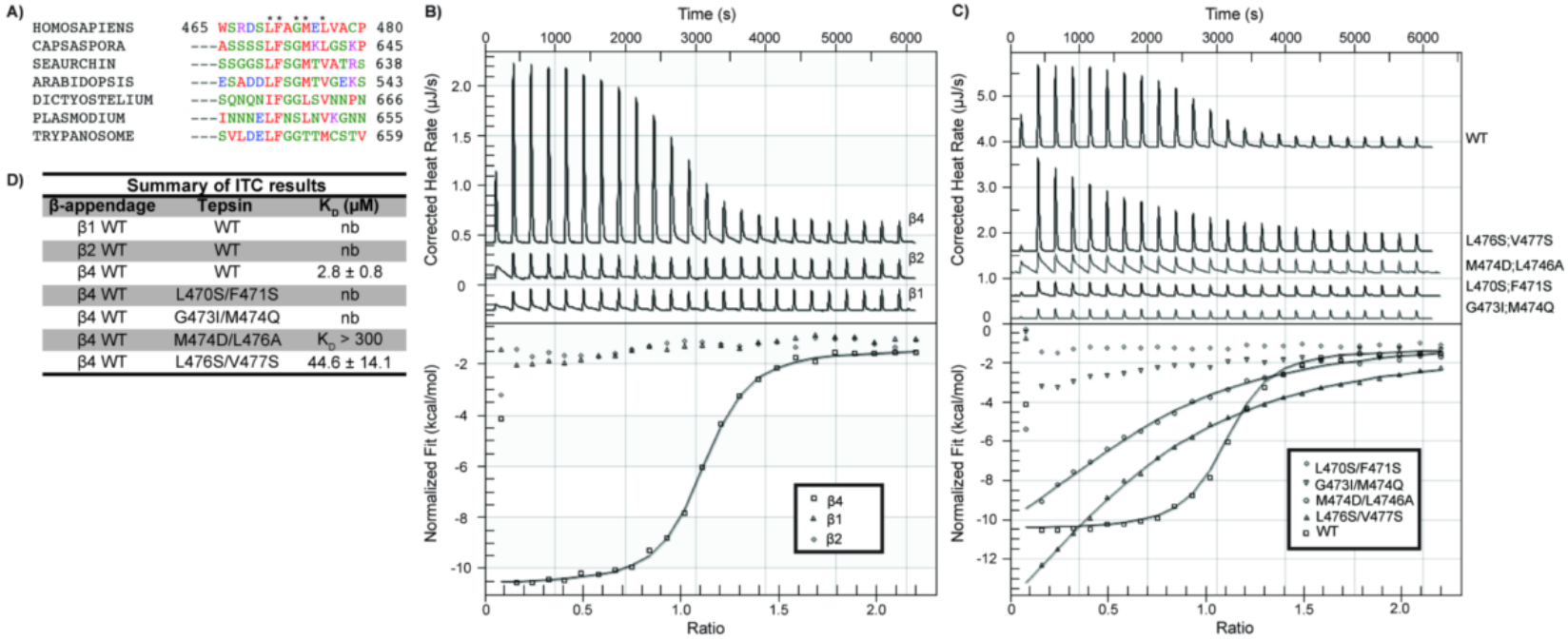


Figure 2-1. The AP4 β4 appendage domain interacts directly with the tepsin C-terminus. A) Schematics of AP4 coat protein complex (top, ε/β4/μ4/σ4 subunits) and tepsin (bottom). Tepsin contains structured ENTH and VHS-like domains at its N-terminus, together with a mostly unstructured C-terminus. The conserved hydrophobic sequence for binding to β4 identified in this work (LFxG[M/L]x[L/V]) is highlighted. B) Pulldown experiments using recombinant purified proteins identified a region in the tepsin C-terminus that binds β4; a panel of GST-tepsin constructs was used as bait and β4 appendage domain as prey. (Top: Coomassie-stained SDS-PAGE gel with tepsin construct amino acid residue ranges marked; bottom: Western blot using anti-β4 antibody).

SDS-PAGE gel and was further confirmed using an antibody raised specifically against β4 (Hirst *et al.,* 1999) (Figure 2-1B). A longer construct containing tepsin residues 450–525 also exhibited binding to β4 appendage domain but was pulled down to a lesser degree; this construct was unstable and prone to proteolysis, so we did not pursue it further, since the shorter fragment exhibited strong binding to β4. We then performed a sequence alignment of this region of tepsin across species and identified a highly conserved stretch of 8 amino acids (Figure 2-2A). We hypothesized this sequence, LFxG[M/L]x[L/V], might constitute a unique motif that allows tepsin to bind specifically to β4. To test this hypothesis, we used isothermal titration calorimetry (ITC) to test directly whether β4 appendage domain could bind both to a short peptide (SSRDSLFAGMELVACS) containing just the motif (Figure A1-1) and to a recombinant tepsin fragment containing the motif (residues 450–500). In both cases, β4 binds the tepsin sequence (Figure A1-1, 2-2B) with 1:1 stoichiometry and a low micromolar affinity (K_D= 2.9±0.8 μM, 10 independent experiments), which is typical of the strength of interactions between APs and accessory proteins (Edeling *et al.,* 2006b). This interaction is specific to β4, because neither the β1 or β2 appendage domains bind the motif *in vitro* (Figure 2-2B).

Using site-directed mutagenesis and ITC, we next tested whether highly conserved residues within the putative motif were key in β4 binding. All tepsin point mutants were expressed as GST-fusion proteins and purified from *E. coli*; the GST tag was removed prior to all ITC runs so that binding was conducted on untagged and unlabeled proteins. We introduced double point mutations in order to observe an effect *in vitro*; a requirement to introduce two mutations suggests a large interface and multiple side chains are involved in binding. Each of our tepsin point mutants exhibited either significantly reduced or abolished measurable binding by ITC (K_D >300 μM) (Figure 2-2C, 2-2D). The most highly conserved residues in the motif are L470, F471, G473, and M474. Simultaneous mutation of both L470 and F471 to serine residues completely abrogated binding, suggesting these residues form important hydrophobic interactions with the surface of



D)

Summary of ITC results

β-appendage	Tepsin	K_D (µM)
β1 WT	WT	nb
β2 WT	WT	nb
β4 WT	WT	2.8 ± 0.8
β4 WT	L470S/F471S	nb
β4 WT	G473I/M474Q	nb
β4 WT	M474D/L476A	K_D > 300
β4 WT	L476S/V477S	44.6 ± 14.1

Figure 2-2. Mutagenesis of a conserved hydrophobic sequence in tepsin reduces or abolishes binding to β4. A) Sequence alignment of tepsin from major eukaryotic super groups reveals a conserved stretch of 8 residues in the unstructured C-terminus. Asterisks mark highly conserved amino acids. B) Wild-type β4 binds a recombinant tepsin fragment (residues 450-500) containing the conserved sequence with a K_D of 2.9±0.8 µM by isothermal titration calorimetry (ITC, 10 independent experiments). In contrast, the tepsin fragment cannot bind purified β1 or β2 appendages found in the clathrin adaptors AP1 or AP2, respectively. C) Representative ITC traces of tepsin mutants with WT β4. Tepsin residues marked by asterisks in A) were mutated to test their importance for binding. All mutants either reduced or abrogated measurable binding to β4 *in vitro*. D) ITC summary results table. All K_D values are represented as averages ± standard deviation; nb= no measurable binding.

β4. The highly conserved glycine may suggest a position in the motif at which backbone flexibility or a unique backbone conformation is required, since glycine can adopt backbone torsion angles forbidden to other amino acids. Introducing a bulkier isoleucine side chain at the glycine position, together with a polar glutamine residue in place of the conserved methionine (G473I/M474Q), also completely abrogates measurable binding. Mutating less conserved hydrophobic residues (L476S/V477S) has only a modest effect on binding: this mutant exhibits ~15-fold weaker binding. Combining the L476S mutation with a point mutant having a charged aspartate at the methionine position (M474D/L476S) substantially reduces binding, suggesting the methionine forms important contacts with β4. The Network Protein Sequence Analysis secondary structure prediction algorithm (Combet *et al.,* 2000) suggested the tepsin motif may form a helix, which is reminiscent of the induced helix found in a $[DE]_nX_{1-2}FXX[FL]XXXR$ motif when it binds the β2 appendage domain (see below and Discussion; Edeling *et al.,* 2006a). In the tepsin motif, the spacing of critical conserved residues identified via sequence alignment and confirmed by mutagenesis does not seem to fit the register required for a helix, so we further tested the possibility of helix induction using circular dichroism (CD) spectroscopy. Formation of the β4-tepsin complex did not exhibit any additional α-helical character (Figure A1-2) beyond that already present in the two isolated molecules. Combined with our mutagenesis data, this suggests the tepsin motif does not adopt a helical conformation when bound to β4 but instead may bind in an extended conformation on the surface of β4 appendage; however, additional structural data will be required to confirm this prediction.

Chemical shift NMR data reveal the tepsin binding surface on β4 appendage domain

We next set out to identify residues on the surface of β4 appendage that bind the tepsin motif. All attempts to crystallize the β4 appendage, alone or together with a peptide corresponding

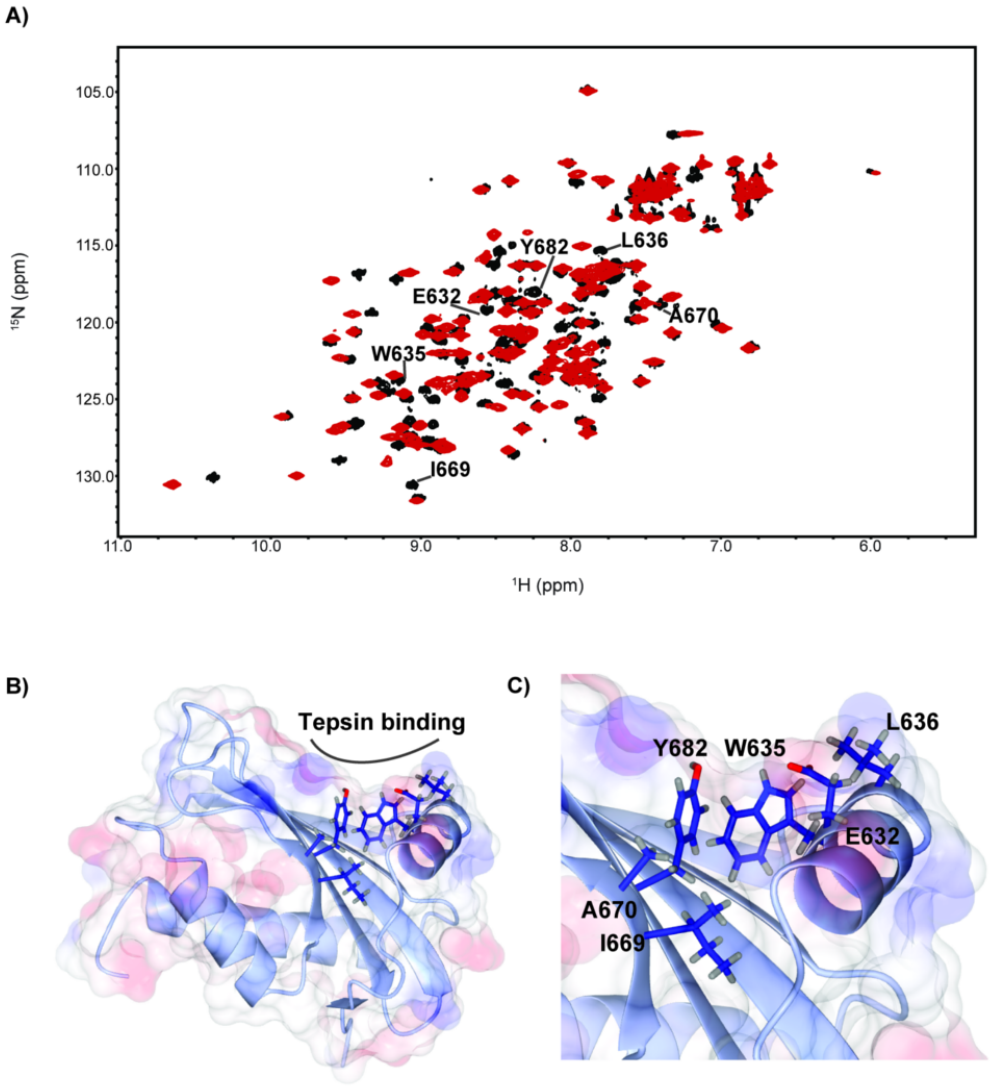


Figure 2-3. NMR chemical shift perturbations reveal residues on β4 surface involved in tepsin binding. A) ^{15}N-^{1}H HSQC spectra of the initial (black) and final (red) titration points show chemical shifts resulting from binding upon addition of unlabeled recombinant tepsin (residues 450-500) to labeled β4. Residues identified for further analysis are highlighted below. B) From the HSQC titration, residues that exhibited large chemical shift perturbations were mapped onto the structure (PDB 2MJ7): E632, W635, L636, I669, A670, Y682. C) Close-up view of binding interface residues identified by chemical shift perturbations.

to the tepsin motif, failed. Because there is an unpublished nuclear magnetic resonance (NMR) structure of β4 deposited in the PDB (2MJ7), we instead turned to NMR to identify the binding site experimentally. A NMR chemical shift perturbation (CSP) experiment was conducted monitoring changes in the ^{15}N-^{1}H heteronuclear single quantum correlation (HSQC) spectrum of ^{15}N-labeled

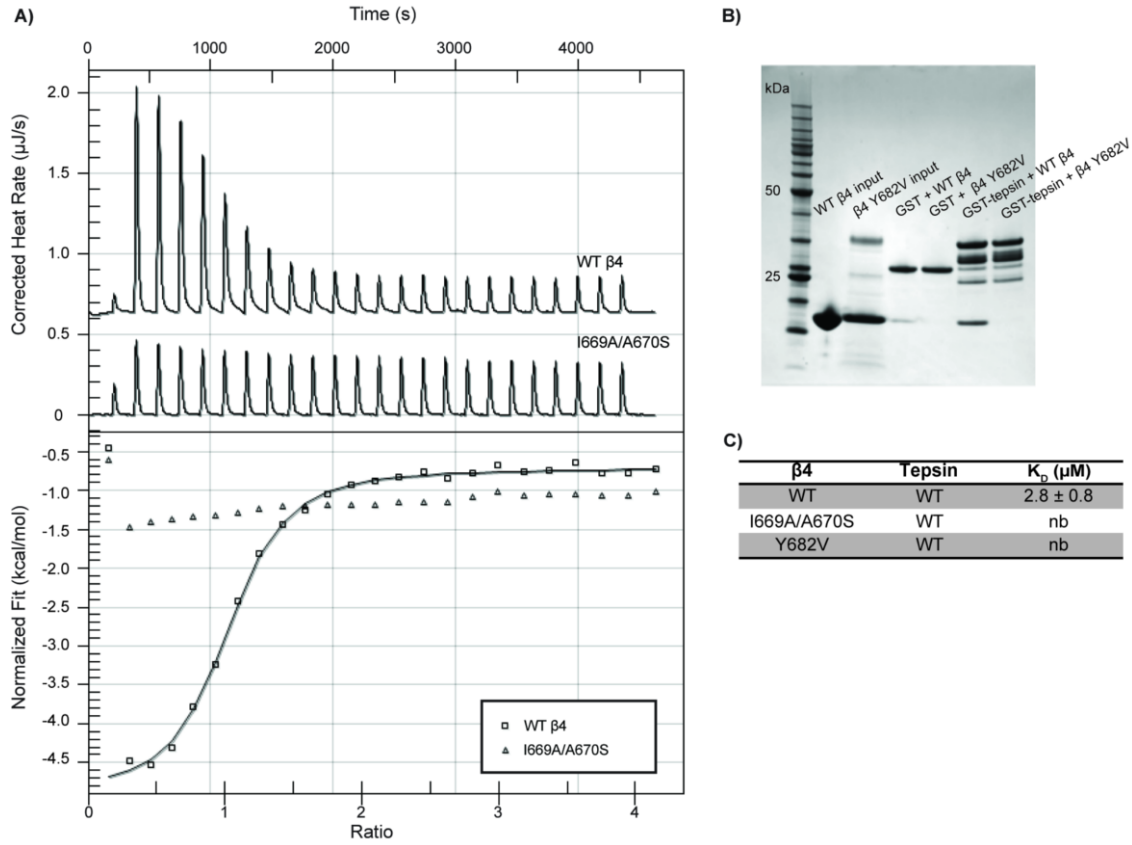
β4 appendage domain upon titration with unlabeled tepsin (residues 450–500). The chemical shift assignments for free tepsin were obtained from the Biological Magnetic Resonance Bank (BMRB). Assignments for the complex were transferred by following the chemical shift changes over the course of the titration for signals exhibiting fast exchange on the NMR chemical shift time scale, and on the basis of the closest new signal for signals in the intermediate to slow exchange regime. Observation of signals in the intermediate and slow exchange regimes is consistent with the K_D value in the low micromolar range as measured by ITC (Ziarek *et al.,* 2011). The spectrum obtained at a 2:1 molar ratio of tepsin:β4 was used to identify those peaks with substantial CSPs (Figure 2-3A). The analysis revealed ~20 peaks with substantial CSPs ($\delta\Delta(^{15}N) > 0.5$ ppm, $\delta\Delta(^{1}H) > 0.05$ ppm). To visualize the binding site on the β4 appendage, the large CSPs were mapped onto the NMR structure (Figure A1-3A).

We also compared the β4 structure with X-ray crystal structures of β2 appendage domain alone and in the presence of the $[DE]_nX_{1-2}FXX[FL]XXXR$ motif (PDB ID: 2G30) found in epsin1, ARH, and arrestin. β4 appendage is distinct among β appendages in having only one of the two subdomains found in other appendage domains (Owen *et al.,* 1999; Traub *et al.,* 1999; Owen *et al.,* 2000): β4 possesses the C-terminal platform subdomain but lacks the N-terminal sandwich subdomain. Structure conservation mapping in ConSurf (Figure A1-3B) highlighted specific residues in a binding patch that are highly conserved between the two domains. In addition, many of the same β4 residues found in this patch were independently identified in our NMR experiments as being critical for binding the tepsin motif. We superposed a representative conformer of the β4 NMR ensembles (PDB ID: 2MJ7) onto the β2 crystal structure (Figure A1-3C). The NMR structure confirms that residues 618–739 comprise the β4 appendage domain; residues prior to 618 are present in the construct but are disordered in the NMR structure. β4 superposes on the β2 platform subdomain with an RMSD of 1.9 Å in CCP4MG (McNicholas *et al.,* 2011). Furthermore, spatially equivalent residues in β2 mediate binding to the $[DE]_nX_{1-2}FXX[FL]XXXR$ motif through

the F and FL pockets (Edeling *et al.,* 2006a) that accommodate crucial hydrophobic residues in the motif. It thus seemed likely the tepsin motif binds β4 in an equivalent position to $[DE]_nX_{1-2}$FXX[FL]XXXR motif binding on β2, though there are subtle structural differences between the two appendage domains (see Discussion). Based on our NMR chemical shift data and residues conserved between the β2 and β4 binding pockets, we selected several candidate β4 residues for structure-based mutagenesis: E632, W635, L636, I669, A670, and Y682 (Figure 2-3B,C). W635, I669, and A670 are equivalent to β2 W841, I876, and A877 in the F pocket, while Y682 is equivalent to β2 Y888 in the FL pocket. E632 and L636 are not conserved between β4 and β2 but demonstrated large chemical shift perturbations.

We then used a combination of pulldowns and ITC to test the importance of these β4 residues for binding tepsin (Figure 2-4). All β4 mutants used in pulldown or ITC experiments were determined to be folded using a combination of gel filtration profiles (data not shown) and CD spectroscopy (Figure A1-4). Our ITC and pulldown data indicate the most important residues for binding tepsin are I669, A670, and Y682. In contrast, the E632A/L636A mutant bound the tepsin motif as strongly as wild-type (data not shown). Although this suggests the glutamate and leucine residues do not interact directly with the motif but move out of the way to accommodate tepsin binding, high resolution structural data will be required to confirm this. In contrast, the I669A/A670S double mutant exhibited no measurable binding by ITC (Figure 2-4A). We replaced the bulky tyrosine side chain at Y682 with a smaller valine moiety; although this mutant was folded, we did not obtain enough material to undertake ITC experiments. Instead, pulldowns with the GST-tepsin fragment confirm this single point mutation is sufficient to substantially weaken binding in a pulldown assay (Figure 2-4B,C). Finally, we attempted to test the role of W635 because its equivalent residue in the β2 appendage is an important component of the F pocket. In β4, this residue exhibited only a small chemical shift perturbation upon the addition of tepsin. Unfortunately, our W635A mutant was completely insoluble when expressed in *E. coli*, suggesting

this mutant was unfolded. In the absence of a high resolution structure, we cannot rule out the importance of this residue in binding the tepsin motif, but the small chemical shift perturbation may suggest it is not as important as other residues, in contrast to its equivalent residue in β2 appendage that plays a key role in recognizing $[DE]_nX_{1-2}FXX[FL]XXXR$ motifs.



β4	Tepsin	K_D (μM)
WT	WT	2.8 ± 0.8
I669A/A670S	WT	nb
Y682V	WT	nb

Figure 2-4. Structure-based mutagenesis of key β4 residues eliminates *in vitro* interaction with tepsin. A) The β4 mutant I669A/A670S exhibits no measurable binding to the tepsin motif by ITC (representative trace). B) Wild-type GST-tepsin (residues 450-500) pulls down wild-type β4 (positive control) but fails to pull down the β4 mutant Y682V. GST (lane 4) was used as a negative control. C) Table summarizing β4 mutant results from ITC and pulldown experiments.

Mutation of key residues in β4 appendage domain or the C-terminal tepsin motif disrupts the interaction between tepsin and AP4 in cultured cells

Based on our *in vitro* characterization of the tepsin binding surface on the β4 appendage domain, we sought to test the functional relevance of this interaction in cultured human cells. To provide a clean background for expression of structure-based point mutants, we used clustered, regularly interspaced, short palindromic repeat (CRISPR) technology to inactivate all alleles of *AP4B1* in HeLa M cells. Knockout of *AP4B1* was confirmed by Western blotting and sequencing (Figure 2-5A and A1-5A). Patient studies have shown that disruption of one AP4 subunit can lead to reduced expression of the others (Hirst *et al.,* 2013; Hardies *et al.,* 2015; Borner *et al.,* 2012). Similarly, there is a reduction in the amount of AP4 ε in the β4 knockout cell line (Figure 2-5A). However, as is also the case in AP4 patient cell lines, there is no change in tepsin expression

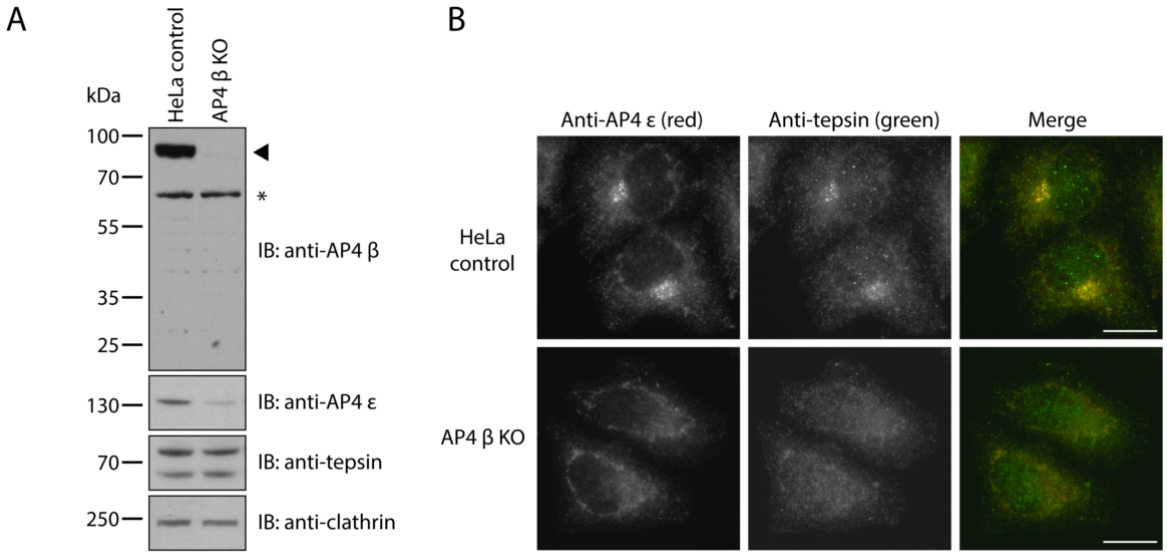


Figure 2-5. Tepsin is not recruited to the membrane in AP4 β4 knockout (KO) HeLa cells generated using CRISPR technology. A) Western blots of whole cell lysates from wild-type and AP4 β KO HeLa cells, probed with antibodies against AP4 β, ε, and tepsin (n.b. tepsin has two isoforms). AP4 β (marked by the arrowhead) is undetectable in the AP4 β KO cells. The band marked with the asterisk is non-specific. While the amount of AP4 ε in the AP4 β KO cells is reduced, the expression level of tepsin is unchanged. An antibody against clathrin was used as a loading control. B) Immunofluorescence double labeling for AP4 ε and tepsin in wild-type and AP4 β KO HeLa cells. In wild-type cells, tepsin labeling is punctate and concentrated in the perinuclear region where it colocalizes extensively with AP4. This pattern was absent in the AP4 β KO cells. Scale bars are 20 μm.

level. In immunofluorescence analysis of wild-type cells, tepsin has a punctate pattern in the TGN region that co-localizes extensively with AP4 (Figure 2-5B; Borner *et al.,* 2012). As expected, this pattern is absent in the β4 knockout cells; tepsin appears to be purely cytosolic, demonstrating that tepsin depends on AP4 for membrane recruitment (Figure 2-5B).

In order to test whether residues I669, A670, and Y682 in the β4 appendage domain are important for tepsin binding *in vivo*, we created stable rescue cell lines on the β4 knockout background, using either full-length wild-type β4 or one of several mutants: earless (residues 1-612); I669A/A670S; and Y682V. The earless mutant lacked the entire β4 appendage domain and was used to determine whether the β4 appendage domain is necessary for tepsin recruitment by AP4. We immunoprecipitated AP4 from extracts of these cells with antibodies against either β4 or ε, and we determined how much tepsin co-immunoprecipitated by Western blotting (Figure 2-6A). Immunoprecipitates from knockout cells rescued with full-length wild-type β4 contained a similar amount of tepsin to immunoprecipitates from original wild-type parental HeLa cells. In contrast, substantially less tepsin co-immunoprecipitated with AP4 from all three mutant cell lines. The amount of tepsin in the immunoprecipitates from the I669A/A670S and Y682V cells was comparable to that from the earless cells, providing evidence that these residues are crucial for the direct interaction between tepsin and the β4 appendage domain *in vivo*. However, a small amount of tepsin still came down with AP4 in the immunoprecipitates from all three cell lines rescued with mutant β4. This is in contrast to the β4 knockout cell line where no tepsin could be detected in the AP4 immunoprecipitates and strongly suggests the existence of at least one additional tepsin binding site elsewhere on the AP4 complex. The β4 constructs in the four rescue cell lines were expressed at similar levels (Figure A1-5B) and blots of the immunoprecipitates probed with antibodies against β4 and ε

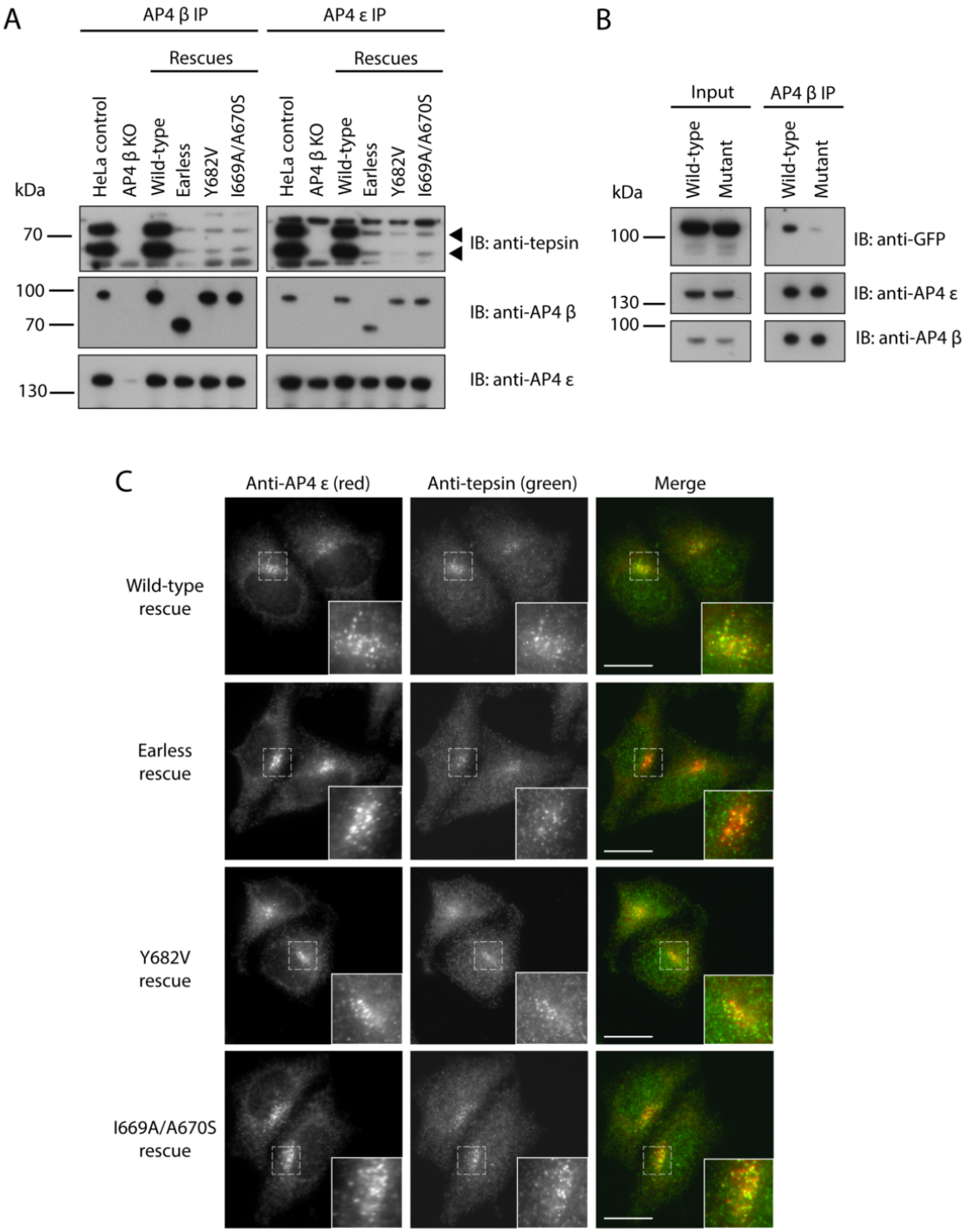

A
AP4 β IP
AP4 ε IP
Rescues
Rescues
HeLa control
AP4 β KO
Wild-type
Earless
Y682V
I669A/A670S
kDa
70
100
70
130
IB: anti-tepsin
IB: anti-AP4 β
IB: anti-AP4 ε
B
Input
AP4 β IP
Wild-type
Mutant
kDa
100
130
100
IB: anti-GFP
IB: anti-AP4 ε
IB: anti-AP4 β
C
Anti-AP4 ε (red)
Anti-tepsin (green)
Merge
Wild-type rescue
Earless rescue
Y682V rescue
I669A/A670S rescue

Figure 2-6. Disruption of the β4 appendage-tepsin interaction *in vivo* greatly reduces, but does not abolish, tepsin binding to AP4. A) Western blots of immunoprecipitates of AP4 β or ε from extracts of control (wild-type HeLa), AP4 β KO, or AP4 β KO cells stably rescued with full-length wild-type or mutant (earless, Y682V, or I669A/A670S) β4, probed with antibodies against AP4 β, ε, and tepsin (marked by arrow heads). No tepsin could be detected in the immunoprecipitates from the AP4 β KO cells, but the immunoprecipitates from the KO cells rescued with wild-type β4 contained a similar amount of tepsin to immunoprecipitates from the wild-type control cells. A small amount of tepsin was detected in the immunoprecipitates from the KO cells rescued with mutant β4. B) Western blots of immunoprecipitates of AP4 β from extracts of HeLa cells stably expressing either wild-type or mutant tepsin-GFP (L470S/F471S), probed with antibodies against GFP, AP4 β and ε. Mutant tepsin-GFP co-immunoprecipitates with AP4 less than wild-type tepsin-GFP. C) Immunofluorescence double labeling for AP4 ε and tepsin in AP4 β KO HeLa cells stably rescued with wild-type or mutant (earless, Y682V or I669A/A670S) β4. Tepsin co-localizes with AP4 in the perinuclear region in all four rescued cell lines. Scale bars are 20 μm.

confirmed that immunoprecipitation of the AP4 complex was equally efficient from all four rescue cell lines (Figure 2-6A).

To confirm the role of the C-terminal tepsin motif (LFxG[M/L]x[L/V]) *in vivo*, we created stable HeLa cell lines expressing either wild-type tepsin C-terminally tagged with GFP (tepsin-GFP), or a tepsin-GFP construct carrying two point mutations, L470S and F471S, which had been shown to abolish binding to the β4 appendage domain *in vitro*. Immunoprecipitation of AP4 using an antibody against β4 from each cell line was performed, and co-immunoprecipitation of tepsin-GFP was assayed by probing Western blots with anti-GFP (or anti-tepsin; data not shown). These experiments demonstrated greatly reduced co-immunoprecipitation of mutant tepsin-GFP relative to wild-type tepsin-GFP (Figure 2-6B), supporting an important role for the C-terminal tepsin motif in binding to AP4. However, the interaction between L470S/F471S mutant tepsin-GFP and AP4 was not completely abolished, providing further evidence to support the existence of one or more additional tepsin-AP4 interaction sites.

Finally, we tested whether the interaction between β4 appendage and the C-terminal tepsin motif is necessary for the recruitment of tepsin to TGN membranes. Given that disruption of this interaction site had a severe effect on tepsin binding in the immunoprecipitation assays, we hypothesized that the residual tepsin binding observed would be insufficient to mediate

recruitment of tepsin. Wild-type and mutant β4 rescue cells were fixed and double-labeled with antibodies against AP4 ε and tepsin (Figure 2-6C). The anti-AP4 ε labeling confirmed that all three mutant β4 proteins are able to successfully form AP4 complexes on the membrane. To our surprise, punctate co-localization of AP4 ε and tepsin could be seen in the wild-type rescue cells and the three mutant cell lines, again suggesting that other parts of the AP4 complex are involved in tepsin binding and recruitment to the membrane.

Discussion

The data presented here provide molecular insight into how the AP4 β4 appendage domain interacts specifically and directly with the major AP4 accessory protein, tepsin. The low micromolar K_D, 1:1 stoichiometry, and presence of a short motif are all reminiscent of trafficking protein interactions between AP complexes and accessory proteins found in the middle layer of AP1- or AP2-containing clathrin coated vesicles (CCVs). AP4 forms a non-clathrin coat but seems to have adopted similar strategies for interacting with accessory proteins. However, the tepsin motif binds specifically to β4 appendage and not to β appendages found in the clathrin adaptors AP1 or AP2. One notable feature of the β4 appendage domain is that it possesses only one subdomain; most other appendages, and all other β appendages, contain both an N-terminal sandwich and C-terminal platform subdomain (Owen *et al.,* 1999; Traub *et al.,* 1999; Owen *et al.,* 2000), while β4 contains only the platform subdomain. The sandwich subdomain of AP2 β2 provides a second binding site for [FL]xxG[FL]xDF motifs found in eps15 (Schmid *et al.,* 2006). Thus, β2 harbors 'top' and 'side' binding sites for multiple motifs in different proteins. In contrast, because β4 lacks the sandwich subdomain, it lacks an equivalent 'side' binding site for other accessory proteins. This suggests the predominant role of β4 appendage domain may be to recruit tepsin by binding the LFxG[M/L]x[L/V] motif.

The surface patch we identified on β4 is broadly conserved with the β2 patch that interacts with the $[DE]_nX_{1-2}FXX[FL]XXXR$ helical motif found in epsin1, ARH, and arrestin. The two surface patches share several conserved residues, including β4 W635, I669, A670, and Y682 (equivalent to β2 residues W841, I876, A877, andY888). In β2, Y888 and W841 are located in the F and FL pockets, respectively, that accommodate critical hydrophobic residues in the $[DE]_nX_{1-2}FXX[FL]XXXR$ motif. How then can we explain the specificity of the tepsin motif for β4, as observed in our ITC data? Despite very similar overall folds, the shape of the β2 and β4 surfaces differs in one important feature: β2 appendage features a deep pre-formed groove (Figure A1-3B) that can accommodate the helical motif. In contrast, the β4 surface is relatively flat (Figure A1-3C). Together with our CD data and the motif pattern, these relatively subtle differences in surface shape further support the idea that the tepsin motif likely binds β4 in an extended conformation.

Comparison of the β2 and β4 platform binding interfaces suggests that evolution modified a conserved binding patch in order to "tune" specificity for AP binding partners at different membranes. The tepsin motif is mostly hydrophobic, appears to lack key charged residues, and likely binds in an extended conformation. In contrast, the β2-binding $[DE]_nX_{1-2}FXX[FL]XXXR$ motif combines an important charged moiety with key hydrophobic residues and adopts a helical conformation. Despite having similar overall three-dimensional folds and conserved residues, β2 and β4 use shape to discriminate easily between their respective epsins. This is even more important for the AP1 β1 appendage domain, since both AP1 and AP4 mediate different trafficking pathways from the TGN.

One important question is whether or not the LFxG[M/L]x[L/V] motif we have identified in tepsin is found in other proteins. To address this question, we used the Genome Net Motif (Kanehisa *et al.,* 2002) online search tool with the KEGG GENES database to search for other human proteins containing our motif. We obtained fifty-nine hits, and we scanned the entries for

motifs located in the C-termini of these proteins. Interesting candidates included transmembrane proteins such as solute carrier family members that mediate membrane transport; subunits of the vacuolar ATPase; and a transporter implicated in Niemann-Pick disease. In the future, it will be important to test these candidate proteins to determine if they can interact with AP4.

Our data in HeLa cells suggest the LFxG[M/L]x[L/V] motif binding to β4 appendage domain is one mechanism for tepsin to bind AP4 *in vivo*. Specific disruption of this interaction, by mutation of either tepsin or β4, results in a large reduction in the amount of tepsin co-immunoprecipitated with AP4. However, it does not completely abolish the association of tepsin with AP4, suggesting the existence of at least one further interaction between AP4 and tepsin. This is also supported by immunofluorescence data from our β4 rescue cell lines: cells lacking the β4 appendage or β4 appendage mutants that cannot bind tepsin *in vitro* by ITC or IP still exhibit significant tepsin membrane recruitment. While preparing this manuscript, Mattera *et al*. (2015) independently used yeast two-hybrid screening to identify the LFxG[M/L]x[L/V] motif in tepsin. In addition, their peptide screen identified a second tepsin binding motif, S[A/V]F[S/A]FLN, that binds the AP4 ε appendage domain. The presence of this second motif may explain why tepsin is still recruited to membranes in our β4 knockout and mutant rescue cell lines. However, in cells, the reported low micromolar affinity (3 μM; Mattera *et al*., 2015) between S[A/V]F[S/A]FLN and the ε appendage would likely be insufficient to recruit tepsin on its own. Furthermore, we note that β4 is more highly expressed in our rescue cell lines than in the parental wild-type HeLa cell line (see Figure A1-5B). In other experiments not discussed here, we have observed that overexpression of one AP4 subunit is sufficient to drive increased AP4 membrane recruitment and vesicle formation. This might explain how tepsin is still recruited to membranes in the mutant β4 rescue cell lines: with more AP4 present on membranes, a weaker second tepsin-AP4 interaction may be sufficient for tepsin recruitment. Alternatively, there may be additional interactions between tepsin and AP4 that have not yet been identified. For example, Mattera *et al*. (2015) also found that deleting the tepsin

ENTH and VHS-like domains together greatly reduced binding to AP4 in cell lysates, while deleting only the ENTH domain had no effect.

Our data have important implications for the role of tepsin in AP4 coat assembly and in understanding both AP4 cell biology and disease pathogenesis. AP4 is distinct among AP complexes in lacking a structural or scaffolding protein like clathrin. This raises the mechanistic question of how AP4 coats assemble, since they are unlikely to form a layered coat in the absence of clathrin. Furthermore, tepsin and AP4 are found or lost concomitantly together in eukaryotes, and tepsin depends upon AP4 for its membrane recruitment (Borner *et al.,* 2012). This suggests a strong interaction between these two proteins across evolutionary history. What then is the role of tepsin in an AP4 coat? Tepsin may be a cargo adaptor; both ENTH (Miller *et al.,* 2007; Hirst *et al.,* 2004) and VHS (Misra *et al.,* 2002) domains fulfill this role in other trafficking proteins, including epsins and the monomeric GGA coats. Another possibility is that tepsin may be a structural component that is required to polymerize AP4 complexes, especially in the absence of a structural or scaffolding protein. Indeed, tepsin could fulfill both of these roles simultaneously. We have accounted for only a small portion of tepsin, and much work remains to be done to understand its full role in the cell.

Materials and methods

Reagents

Unless otherwise noted, all chemicals were purchased from Sigma (St. Louis, MO, USA).

The following antibodies were used in this study: rabbit anti-AP4 β, rabbit anti-AP4 ε (for western blotting and immunoprecipitation; both in-house; Hirst *et al*., 1999), mouse anti-AP4 ε (for immunofluorescence; 612019; BD Transduction Labs), rabbit anti-clathrin (in-house; Simpson *et al*., 1996), rabbit anti-GFP (gift from Matthew Seaman, Cambridge Institute for Medical Research, UK), and rabbit anti-tepsin (in-house; Borner *et al*., 2012). Horseradish peroxidase (HRP)–conjugated secondary antibodies were purchased from Sigma-Aldrich, and fluorescently labelled

secondary antibodies were from Invitrogen. For Western blotting of immunoprecipitates where the protein band of interest was close to an IgG band, protein-A-HRP (BD Biosciences) was used in the place of HRP-conjugated secondary antibody.

Molecular biology and cloning

For structural and biochemical analyses, part of the tepsin C-terminus (residues 450–500) was subcloned into the pGEX-6P-1 backbone (GE Healthcare) using BamHI/SalI sites, to create a N-terminal GST tagged protein. β4 appendage domain (residues 612-739) was subcloned into the pGEX-4T-2 backbone (GE Healthcare) using BamHI/NotI sites, resulting in a N-terminal GST tagged protein. A two-stage quick-change mutagenesis protocol adapted from Wang and Malcolm (1999) was used to introduce mutations in the tepsin motif and the β4 appendage domain. Briefly, mutagenic primers (Sigma) were created for the desired mutations. In the first step, two PCR reactions, with either the mutagenic 5' or 3' primer, amplified around the plasmid. The two reactions were then combined in an additional PCR step, followed by a DpnI digest and transformation. Resultant colonies were sequenced to confirm mutagenesis.

To generate the constructs for the AP4 β rescue cell lines, full-length (residues 1–739) and earless (1–612) AP4B1 were amplified by PCR from a full-length IMAGE clone of β4 (2906087). Cloning sites were added (5' SalI and 3' NotI), and the PCR products were cloned into a modified version of the retroviral expression vector pLXIN (gift from Andrew Peden, University of Sheffield, UK). Full-length pLXIN_AP4B1 was modified by site-directed mutagenesis to give pLXIN_AP4B1[Y682V] and pLXIN_AP4B1[I669A/A670S], as described above. For the tepsin-GFP mutant, the L470S and F471S point mutations were introduced into a tepsin-GFP plasmid reported previously (Borner *et al.,* 2012) by site-directed mutagenesis. Wild-type and mutant tepsin-GFP were subsequently amplified by PCR from the wild-type and mutant tepsin-GFP plasmids, respectively, and were inserted into the HpaI site of the modified pLXIN vector using

Gibson Assembly Master Mix (New England Biolabs)-directed cloning. The sequences of all constructs described here were verified by Sanger DNA sequencing.

GST pulldown assays

GST or GST-tepsin proteins (50 µg) were immobilized on glutathione sepharose resin for 1 hour on ice. The resin was incubated for 2 hours on ice with wild-type or Y682V β4 (75 µg) in 20 mM HEPES (pH 7.5), 200 mM NaCl, and 2 mM DTT, with additional buffer added at 1 hour. Samples were washed three times with the same buffer plus 1 mM EDTA and 0.1% v/v Triton X-100. Proteins were eluted from the resin using the wash buffer plus 30 mM reduced glutathione following a 30 minute incubation on ice. Gel samples were prepared from the supernatant following elution, and the assay was analyzed by Coomassie staining of SDS-PAGE gels. When gels were further analyzed by Western blotting, rabbit anti-AP4 β antibody (Hirst *et al.,* 1999) was used.

Protein expression and purification

Constructs were expressed in BL21(DE3)pLysS cells (Invitrogen) for 16-20 hours at 22˚C after induction with 0.4 mM IPTG. Wild-type β4 appendage domain and all tepsin constructs were purified in 20 mM HEPES (pH 7.5), 200 mM NaCl, and 2 mM βME. Mutant β4 appendage domain constructs were purified in 20 mM HEPES (pH 7.5), 500 mM NaCl, and 2 mM βME. Cells were lysed by a disruptor (Constant Systems Limited, Daventry, UK), and proteins were affinity purified using glutathione sepharose (GE Healthcare) in the purification buffer. GST-tagged β4 appendage domains were cleaved overnight with thrombin (Recothrom, The Medicines Company, Parsippany, NJ) at room temperature and eluted in batch. GST-tagged tepsin was cleaved overnight with recombinant 3C protease at 4˚C and eluted in batch. All proteins were further purified by gel filtration on a Superdex S200 preparative or analytical column (GE Healthcare).

NMR titration experiments

Uniformly enriched ^{15}N-labelled β4 appendage domain was produced in minimal media (M9) with 0.5g/L $^{15}NH_4Cl$ (Cambridge Isotope Laboratory) as the sole nitrogen source. Otherwise, expression and purification proceeded as described above. Samples were prepared in gel filtration buffer with 10% v/v D_2O at a concentration of 300 μM. Unlabeled tepsin was added in 0.1 or 0.3 mol equivalents until a 2:1 tepsin:β4 ratio was reached. Standard 2D ^{15}N-^{1}H HSQC NMR spectra were collected at 25˚C on a 600 MHz Bruker Avance III spectrometer with a TCI triple resonance cryoprobe (Bruker BioSpin). Data were processed initially using TopSpin 3.2 (Bruker BioSpin), with linear prediction in the indirect dimension plus one-fold zero filling, and squared sine bell apodization in both dimensions. The full titration was further analyzed using NMRViewJ (Version 8.0.3, One Moon Scientific). Resonance assignments for β4 were obtained from the BMRB database (19709) and were transposed to our HSQC spectra manually. To analyze the titration data, peaks were classified as did not move, moved slightly (detectable but $\leq$ 0.5 ppm in ^{15}N or $\leq$ 0.05 ppm in ^{1}H), or moved greatly (> 0.5 ppm in ^{15}N or $\geq$ 0.05 ppm in ^{1}H).

Isothermal titration calorimetry

ITC experiments were conducted on a NanoITC instrument (TA Instruments) at 20˚C. Molar peptide concentration in the syringe was at least 6.25 times that of protein in the cell. In experiments using the synthetic tepsin peptide, β4 appendage domain constructs were gel filtered into 50 mM HEPES (pH 7.5), 100 mM NaCl, and 0.5 mM TCEP. In experiments using recombinant tepsin, β4 appendage and tepsin constructs were gel filtered into 20 mM HEPES (pH7.5), 100 mM NaCl, and 0.5 mM TCEP. Incremental titrations were performed with either initial baseline of 180s and injection intervals of 180s, or initial baseline of 120s and injection intervals of 250s. Titration data were analyzed in NanoAnalyze (TA Instruments) to obtain a fit and values for stoichiometry (n) and equilibrium association constant (K_a).

Circular dichroism

Spectra were collected on a Jasco J-810 CD spectropolarimeter, using a cell with 1 mm pathlength. Protein samples were diluted to 0.1 mg/ml in MilliQ water. Data was collected with standard sensitivity, scanning from wavelengths of 260 to 190 nm with a data pitch of 1 nm or 0.5 nm. Under continuous scanning mode the scan speed was 100 nm/min with a response time of 2s, and 5 accumulations were collected.

CRISPR knockout of AP-4 β subunit

We inactivated all copies of the *AP4B1* gene in HeLa M cells using the 'double nickase' CRISPR/Cas9 system (Cong *et al.,* 2013; Ran *et al.,* 2013). This system minimizes the off-target effects associated with the use of wild-type Cas9 by using a mutant Cas9 enzyme (D10A) which can only make single-strand breaks. In combination with paired guide RNAs (gRNAs), this allows the specific introduction of a compound double-strand break at a site of interest. The Zhang online CRISPR design tool (http://crispr.mit.edu/; Hsu *et al.,* 2013) was used to identify suitable paired gRNA targets in *AP4B1*: GACCCCAATCCAATGGTGCG and TGCACAGCGTATTGATGGCC (both in exon 2 of transcript ENST00000369569). Each gRNA was ordered as a pair of complementary oligos with the sequences 5'CACCG$N_{19/20}$-3' and 5'AAAC$N_{19/20}$C-3'. Due to the requirement of the human U6 promoter for a G at the transcriptional start site, an extra G was added in front of the second gRNA target sequence. Oligo pairs were annealed and cloned into the BbsI site of pX335 (Addgene). HeLa M cells were transfected with both pX335 plasmids and pIRESpuro (Clontech) in a ratio of 2:2:1 using a TransIT-HeLaMONSTER transfection kit (Mirus Bio LLC). Forty-eight hours later, untransfected cells were killed off by a 4-day selection in 1 µg/ml puromycin. Single cell clones were isolated by serial dilution of the cells and tested for knockout of AP4 β by Western blotting and immunofluorescence. Clone x2A3 was negative for AP4 β expression in both assays and was further validated by sequencing. Genomic DNA was harvested using a High Pure PCR Template Purification Kit (Roche) and PCR was used to amplify a 633 base pair region around the target sites. The PCR products were then blunt-end cloned (Zero

Blunt PCR Cloning Kit; Invitrogen), and 24 clones were sent for Sanger sequencing with the M13_F primer (Beckman Coulter Genomics).

Cell lines and tissue culture

HeLa M (Tiwari *et al.,* 1987) cells were maintained in RPMI 1640, supplemented with 10% v/v foetal calf serum (FCS), 100 units/ml penicillin, and 100 µg/ml streptomycin. Stable cell lines were additionally maintained with 500 µg/ml G418 (PAA). Stable cell lines were created using retrovirus made in HEK 293ET cells transfected using TransIT-293 Transfection Reagent (Mirus Bio LLC). pLXIN plasmids were mixed with the packaging plasmids pMD.GagPol and pMD.VSVG in a ratio of 10:7:3. Viral supernatants were harvested after 48 hours and used to transduce HeLa M (for the tepsin-GFP cell lines) or HeLa AP4 β knockout clone x2A3 (for the AP4 β rescue cell lines) cells. Transduced cells were selected in 500 µg/ml G418 48 hours post-transduction.

Fluorescence microscopy

Cells were grown onto 13 mm glass coverslips and fixed in ice-cold methanol. Fixed cells were blocked in 0.5 % bovine serum albumin (BSA; in PBS [137 mM NaCl, 2.7 mM KCl, 10 mM Na2HPO4, and 1.76 mM KH2PO4, pH 7.4]). Primary antibody (diluted in BSA block) was added for 45 minutes at room temperature. Coverslips were washed three times in BSA block and then fluorochrome-conjugated secondary antibody was added in block for 30 minutes at room temperature. Coverslips were then washed three times in PBS, followed by a final wash in dH_2O, before being mounted in ProLong Gold Antifade Reagent (Life Technologies). Antibody fluorescence was visualised with an Axio Imager II microscope (Carl Zeiss), controlled by AxioVision software (Carl Zeiss).

Immunoprecipitations and Western blotting

Estimations of protein concentrations were made using a Pierce BCA Protein Assay Kit (Thermo Fisher Scientific). For immunoprecipitations of AP4 β and AP4 ε, cells in 100 mm plates were lysed in 1 ml 1% Triton TX-100 (in PBS), supplemented with 200 µM AEBSF protease inhibitor. Lysates were cleared by centrifugation and, if required, lysates were adjusted to the

same concentration with 1% Triton TX-100. A portion of each lysate was retained as input and the remainder pre-cleared by incubation with 30 µl packed protein-A-sepharose (GE Healthcare) for 1 hour. To each pre-cleared lysate, 5 µl antibody were added for 3-hour incubation and then 30 µl packed protein-A-sepharose were added to recover immunoprecipitated complexes. The sepharose was then washed three times with 1% Triton TX-100, two times with PBS, and boiled in NuPAGE LDS Sample Buffer (Life Technologies) to prepare for Western blot analysis.

Unless prepared for immunoprecipitation, cells were lysed for Western blot analysis in 2.5% (wt/vol) SDS/50 mM Tris, pH 8.0. Lysates were passed through a QIAshredder column (Qiagen) to shred DNA, incubated at 65°C for 3 minutes, and then boiled in NuPAGE LDS Sample Buffer. Samples were loaded at equal protein amounts for SDS-PAGE, performed on NuPAGE 4-12% Bis-Tris gels in NuPAGE MOPS SDS Running Buffer (Life Technologies). PageRuler Plus Prestained Protein Ladder (Thermo Scientific) was used to estimate the molecular size of bands. Proteins were transferred to nitrocellulose membrane by wet transfer and membranes were blocked in 5% w/v milk in PBS with 0.1% v/v Tween-20 (PBS-T). Primary antibodies (diluted in 5% milk) were added for at least 1 hour at room temperature, followed by washing in PBS-T, incubation in secondary antibody (also in 5% milk) for 30 minutes at room temperature, washing in PBS-T, and finally PBS. Chemiluminescence detection of HRP-conjugated secondary antibody/protein-A was carried out using Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare) and X-ray film.

Acknowledgements

The authors thank Linton Traub for the β1 construct; Kate Mittendorf for helpful NMR discussions; and David Owen for the β2 construct, helpful discussion, and critical reading of the manuscript. LPJ and MNF are supported by funds from Vanderbilt University College of Arts and Science to LPJ. MNF is supported on a Molecular Biophysics Training Grant (NIH

5T32GM008320). AKD and MSR are supported by a Wellcome Trust Principal Research Fellowship to MSR. AKD is supported by a National Institute for Health Research Cambridge Biomedical Research Centre PhD Fellowship. GHHB is supported by the Max Planck Society for the Advancement of Science. NMR instrumentation is supported in part by grants from the NSF (0922862), NIH (S10 RR025677), and Vanderbilt University matching funds. The authors declare no conflicts of interest.

CHAPTER III

STRUCTURE AND EVOLUTION OF ENTH and VHS/ENTH-LIKE DOMAINS IN TEPSIN

Tara L. Archuleta*[1,2], Meredith N. Frazier*[1,2], Anderson E. Monken[1,2], Amy K. Kendall[1,2], Joel Harp[2], Airlie J. McCoy[3], Nicole Creanza[1], and Lauren P. Jackson[1,2,4†]

[1]Department of Biological Sciences, Vanderbilt University, Nashville, TN, USA
[2]Center for Structural Biology, Vanderbilt University, Nashville, TN, USA
[3]Cambridge Institute for Medical Research, Department of Clinical Biochemistry, University of Cambridge, Hills Road, Cambridge, United Kingdom
[4]Department of Biochemistry, Vanderbilt University, Nashville, TN, USA

*Equal contributions
[†]Correspondence to lauren.p.jackson@vanderbilt.edu

Division of work: In this co-first author paper, Tara Archuleta performed all biochemical and structural experiments regarding the tENTH domain, with the exception of the tENTH NMR experiment. I performed all biochemical and structural experiments regarding the tVHS domain.

This article has been published under the same title in *Traffic,* 2017 Sep;18(9):590-603.

Synopsis

Tepsin is an accessory protein in AP4 coated vesicles, but its biological role remains unknown. Crystal structures of both folded domains (ENTH and VHS/ENTH-like) reveal these domains harbor structural features distinct from other ENTH/ANTH/VHS superfamily proteins. Phylogenetic and comparative genomics data show how tepsin diverged away from other epsins early in evolutionary history. Together these results imply tepsin has diverged away to undertake a distinct biological role.

Abstract

Tepsin is currently the only accessory trafficking protein identified in adaptor-related protein 4 (AP4) coated vesicles originating at the *trans*-Golgi network (TGN). The molecular basis for interactions between AP4 subunits and motifs in the tepsin C-terminus have been characterized, but the biological role of tepsin remains unknown. We determined X-ray crystal structures of the tepsin ENTH and VHS/ENTH-like domains. Our data reveal unexpected structural features that suggest key functional differences between these and similar domains in other trafficking proteins. The tepsin ENTH domain lacks helix 0, helix 8, and a lipid binding pocket found in epsin1/2/3. These results explain why tepsin requires AP4 for its membrane recruitment and further suggest ENTH domains cannot be defined solely as lipid binding modules. The VHS domain lacks helix 8 and thus contains fewer helices than other VHS domains. Structural data explain biochemical and biophysical evidence that tepsin VHS does not mediate known VHS functions, including recognition of dileucine-based cargo motifs or ubiquitin. Structural comparisons indicate the domains are very similar to each other, and phylogenetic analysis reveals their evolutionary pattern within the domain superfamily. Phylogenetics and comparative genomics further show tepsin within a monophyletic clade that diverged away from epsins early in evolutionary history (~1,500 million years ago). Together, these data provide the first detailed molecular view of tepsin and suggest tepsin structure and function diverged away from other

epsins. More broadly, these data highlight the challenges inherent in classifying and understanding protein function based only on sequence and structure.

Introduction

Large multi-subunit coat protein complexes drive vesicle or tubule formation at specific organelle membranes. The mammalian AP (Assembly Polypeptide) adaptor protein complex family (APs 1–5, COPI F-subcomplex) is a family of heterotetrameric complexes implicated in Golgi-ER and post-Golgi trafficking pathways. Coat components are recruited by lipids (Höning *et al.,* 2005) and membrane-associated small G-proteins (Yu *et al.,* 2012; Ren *et al.,* 2012); they recognize and incorporate transmembrane protein cargo into forming vesicles or tubules (Owen and Evans, 1998; Kelly *et al.,* 2008; Jackson *et al.,* 2010; Kelly *et al.,* 2011; Jackson *et al.,* 2012); and they recruit accessory proteins and additional machinery (Miller *et al.,* 2015) required for assembly and maturation (Robinson *et al.,* 2015). The molecular structures, assembly mechanisms, and cellular functions of non-clathrin associated AP complexes remain poorly understood. The AP4 complex ($\varepsilon/\beta4/\mu4/\sigma4$ subunits) is recruited to the *trans*-Golgi network (TGN) by the small GTPase, Arf1, in its GTP-bound state (Boehm *et al.,* 2001). AP4 has been implicated in polarized cargo sorting in epithelial cells (Simmen *et al.,* 2002) and neurons (Matsuda *et al.,* 2008). While AP4 is ubiquitously expressed, the most striking phenotypes are associated with the brain and central nervous system. A $\beta4$ knockout mouse exhibits missorting of glutamate receptors from the somatodendritic region to axons (Matsuda *et al.,* 2008) but shows no significant anatomical abnormalities. In contrast, human patients with mutations in any of the four AP4 subunits (Abou Jamra *et al.,* 2011; Abdollahpour *et al.,* 2015; Hardies *et al.,* 2015; Tüysüz *et al.,* 2014) suffer from the hereditary spastic paraplegias (Abou Jamra *et al.,* 2011; reviewed in Hirst *et al.,* 2013). AP4 thus plays a key role in neurological development and function, but the underlying mechanism for this role remains unclear. Identifying and understanding the protein

components and molecular mechanisms required to form AP4 coats remain paramount to uncover the function of AP4 both in the cell and in human brain disease.

Tepsin (Figure 3-1A) was the first accessory protein identified in AP4 coated vesicles (Borner *et al.,* 2012); it is a member of the epsin family of post-Golgi trafficking proteins. In mammals, epsins serve as key accessory proteins and cargo adaptors in clathrin coated vesicle (CCV) formation at the plasma membrane and TGN. Members of the epsin family are defined by the presence of an Epsin N-Terminal Homology (ENTH) domain followed by a mostly unstructured C-terminus. The ENTH domain has been identified and characterized in both mammals (Rosenthal *et al.,* 1999) and yeast (Kay *et al.,* 1999). Structural, biochemical, and cell biological data indicate that ENTH domains interact directly with phosphoinositide head groups (Ford *et al.,* 2002) and with SNARE proteins (Hirst *et al.,* 2004; Chidambaram *et al.,* 2004; Miller *et al.,* 2007). The unstructured C-termini of many epsins contain ubiquitin interacting motifs (UIMs; Aguilar *et al.,* 2003) and short linear (Owen *et al.,* 2000; Brett *et al.,* 2002) or secondary structural motifs (Edeling *et al.,* 2006a) for binding AP1/AP2 or clathrin. Epsins thus interact directly with membranes in multiple ways: they directly bind phosphoinositides and cargo, and indirectly bind other trafficking machinery. In contrast, tepsin lacks motifs for binding ubiquitin, clathrin, AP1, or AP2. Tepsin instead contains two short linear motifs in its C-terminus that directly and specifically bind the AP4 ε and β4 appendage domains (Frazier *et al.,* 2016; Mattera *et al.,* 2015). Tepsin is the only epsin known to depend upon its AP complex for membrane recruitment (Borner *et al.,* 2012) and is unique among family members in possessing a second internal folded module, the VHS/ENTH-like domain (Borner *et al.,* 2012).

We determined high resolution X-ray crystal structures of the human tepsin ENTH (tENTH) and the horse tepsin VHS/ENTH-like (tVHS) domains. For clarity, we have shortened the name of the VHS/ENTH-like domain to tVHS. Both domain structures reveal important differences from other members of the ENTH/ANTH/VHS superfamily, because they lack key helices at both the N- and C-termini. tENTH lacks both helix 0 and helix 8. Loss of helix 0 precludes formation of a

lipid binding pocket and explains why tENTH cannot interact directly with phosphoinositide head groups, which we confirm biochemically. Our structural data suggest the tENTH domain could exist as a dimer or tetramer, but we have not found evidence of oligomer formation in solution. The tVHS structure lacks the C-terminal helix 8 that differentiates VHS from ENTH domains within the ENTH/ANTH/VHS superfamily. Important functions of other VHS domains include cargo motif or ubiquitin recognition and binding, but the structure explains our experimental observations that tVHS cannot perform any of these known functions. One striking observation from our data is that the tepsin ENTH and VHS domains are structurally similar. Our phylogenetic analyses indicate that tepsin forms a monophyletic clade within the epsin family but do not support the idea that tepsin ENTH and VHS domains arose from a gene duplication event. Our structural data provide the first detailed molecular view of tepsin, and evolutionary data suggest that ENTH and VHS domains, as well as tepsin itself, have evolved to serve different functions over the course of evolution.

Results

Structure of the tepsin ENTH domain

Other ENTH domains, including those in epsin1 and epsinR, can be expressed and purified in the absence of the first 10–12 residues that constitute helix 0 (Ford *et al.,* 2002; Miller *et al.,* 2007). Initial attempts to purify the tepsin ENTH domain with and without its putative helix 0 failed. However, a construct containing the full predicted ENTH domain (residues 1–136) was successfully purified, indicating the N-terminus was absolutely required to obtain soluble protein (data not shown). We determined a high resolution structure (beyond 1.4 Å) of human tENTH residues 1–136 using molecular replacement methods by placing individual α-helices into the density (further details in Methods). The crystal form contained two molecules in the asymmetric unit; both copies show clear and well-ordered density from residues 3–133 in chain A and residues 4–133 in chain B. There is no significant difference between the two copies; they overlay with an

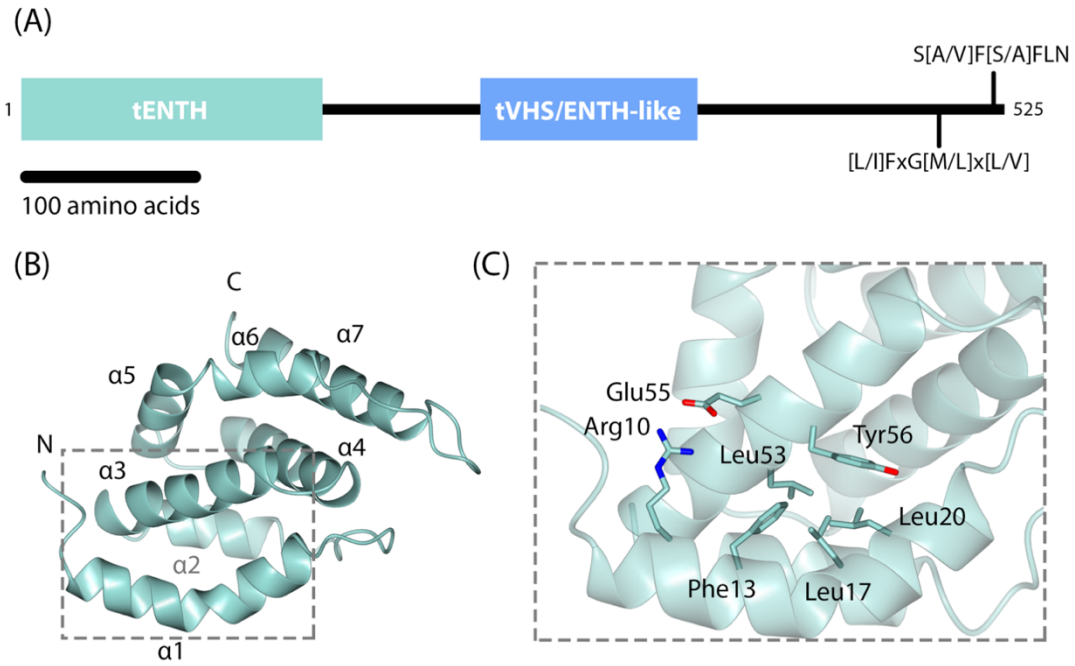


Figure 3-1. Tepsin architecture and crystal structure of tepsin ENTH domain. (A) Overall domain architecture of human tepsin, which contains an N-terminal ENTH (tENTH) and internal VHS/ENTH-like (tVHS) domains. The unstructured C-terminus contains two motifs for binding C-terminal appendage domains of AP4 ε and β4 subunits. (B) Crystal structure of human tENTH residues 1-136 at 1.38 Å resolution. tENTH contains only seven α-helices; it lacks both helix 0 and helix 8 found in other ENTH domains. (C) View of key interactions that facilitate packing between helices α1 and α3 in tENTH, including an ion pair (Arg10 and Glu55) and multiple hydrophobic interactions.

r.m.s.d. of 0.524 Å. Following several rounds of refinement, we obtained a model with excellent geometry (0% Ramachandran and rotamer outliers, 98% Ramachandran favored; Table A2-1).

The tENTH domain is all α-helical and contains seven helices connected by loops of varying length (Figure 3-1B; Table A2-1). The first α-helix, which we have named helix α1 (discussed further below), extends from residues 7–24 and contains a 'kink' in the middle resulting from the presence of a proline at residue 18. From the structure, we can explain our biochemical result that deleting the first helix results in completely insoluble protein. Helix α1 packs against and makes important contacts with specific residues in helix α3; residues Phe13, Leu17, and Leu20 in helix α1 make hydrophobic interactions with Leu53 and Tyr56 in helix α3, while Arg10 in helix1 forms a salt bridge with Glu55 in helix3 (distance=2.9 Å) (Figure 3-1C).

ENTH domains normally contain nine α-helices (discussed further below), but both secondary structural prediction servers (Network Protein Sequence Analysis; Combet *et al.,* 2000) and our crystal structure indicate tENTH contains only seven helices. To test this experimentally, we also crystallized and determined the structure of a second construct containing tepsin residues 1–153 (Figure A2-2). The electron density indicated no additional α-helical or other secondary structure located beyond helix α7 (Figure A2-2). Together these data suggest that additional amino acids beyond residue 136 in the longer construct are disordered and the tepsin ENTH domain is indeed smaller than other published ENTH structures by two α-helices.

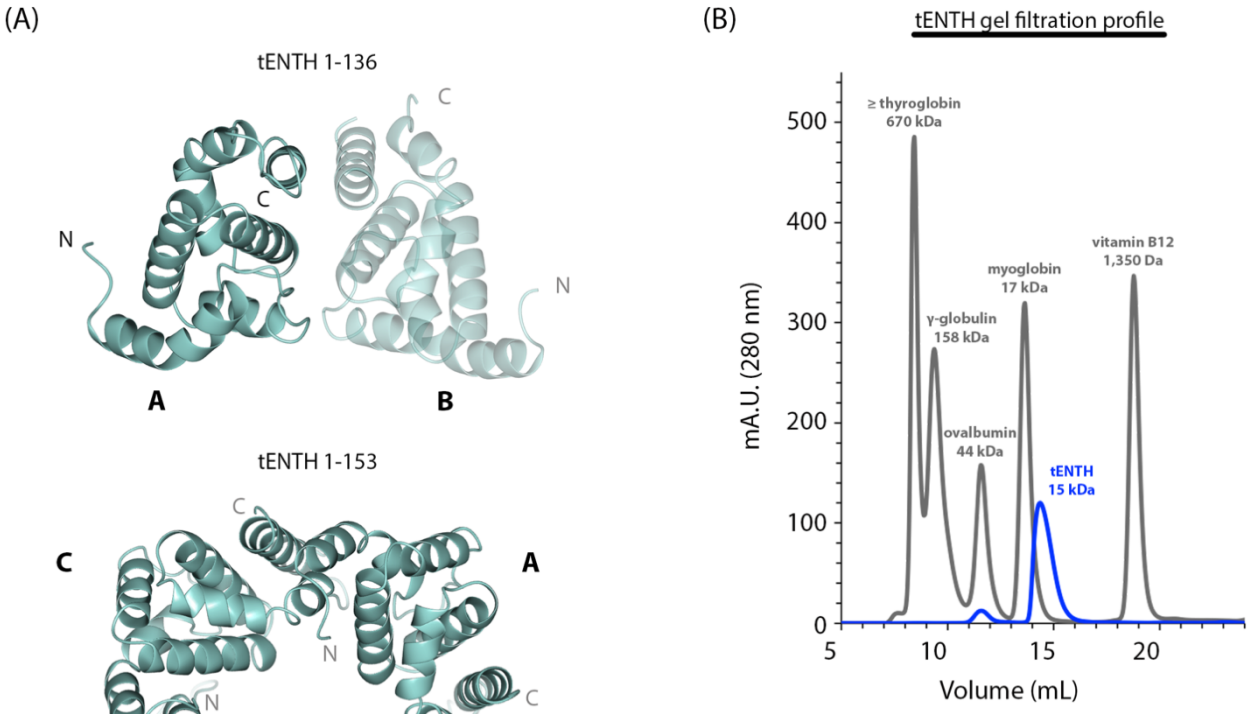


Figure 3-2. tENTH is a multimer in crystals but a monomer in solution. (A) Top panel: The tENTH 1-136 construct crystallized as a dimer (labeled A and B) in the asymmetric unit. Bottom panel: The tetramer observed in the crystal lattice of the tENTH 1-153 structure. (B) Gel filtration profile of the tENTH domain (blue trace) with standards (grey trace), which are consistent with tENTH existing as a monomer in solution.

The tENTH 1–136 construct crystallized as a dimer with two molecules in the asymmetric unit (Figure 3-2A). In this crystallographic dimer, residues in the loop between helices α1 and α2, together with helices α4 and α7, interact to form the interface between chains A and B; this dimer buries 725 $Å^2$ of solvent accessible surface area as calculated by PISA (Krissinel *et al.,* 2007). PISA did not identify this dimer as a potential biological interface but instead identified a tetramer with four-fold circular symmetry (4,637 $Å^2$ buried surface area). The longer tENTH construct (residues 1–153) contains one copy in the asymmetric unit, and four monomers pack to form a tetramer with four-fold circular symmetry in the crystal lattice (Figure 3-2B; 4,765 $Å^2$ buried solvent accessible surface area). The tetramer observed in the tENTH 1–153 lattice is the same tetramer predicted by PISA based on the tENTH 1–136 structure.

Based on these data, we tested whether tENTH can form a dimer and/or tetramer in solution. We turned to gel filtration using a Superdex75 analytical column (10/300 GL, GE Healthcare) using protein standards for comparison. Multiple runs indicated the tENTH domain ran as a monomer around its predicted molecular weight of 15 kDa (Figure 3-2B), close to the myoglobin standard peak (17 kDa). We therefore find no evidence that tENTH exists as an oligomer in solution but cannot rule out the possibility that multimers may form when tepsin is concentrated on membranes in the cell.

Structural and functional comparison of ENTH domains

ENTH domains have been characterized as lipid binding modules (Itoh *et al.,* 2006). Biochemical and structural data from multiple ENTH domains indicate helix α0 is an amphipathic helix that is conformationally dynamic and becomes ordered only in the presence of its binding partner. The epsin1 ENTH binds the phosphoinositide Ptd(Ins)4,5P_2 (Ford *et al.,* 2002) at the plasma membrane, where the first eighteen residues of the ENTH domain form helix α0 only in the presence of the PIP head group. Helix α0 folds and contributes basic residues to a binding pocket that specifically accommodates the Ptd(Ins)4,5P_2 head group. Helix α0 has thus been

hypothesized to facilitate or drive membrane curvature by inserting partially into the membrane (Ford *et al.,* 2002). Several ENTH structures have been determined in the absence and presence of phosphoinositide and protein binding partners. Examples include both epsin1/2 and epsinR ENTHs: yeast Ent1 (PDB: 5LOZ), zebrafish epsin1 (PDB: 5LP0), and mammalian epsin1 (1EDU, 1H0A, 1EYH); yeast Ent3 (PDB: 3ONK, 3ONL) and mammalian epsinR (PDB: 2QY7). These ENTHs adopt compact globular structures containing nine α-helices, and helix 0 is often removed to improve domain solubility when producing proteins *in vitro*.

The first notable difference between tENTH and other ENTH domains is its size. Our data indicate tENTH contains only seven helices, which we have numbered helix α1-α7. In contrast to other ENTHs, first helix in tENTH is required for protein folding and solubility. The extensive contacts between tENTH helix α1 and helix α3 preclude conformational flexibility of the first helix in our structure; we predict this helix can never be disordered and thus cannot be defined as an amphipathic helix 0. Overlaying tENTH with epsin1 ENTH (Figure 3-3A) indicates that helices α1–7 overlay nicely (r.m.s.d. 1.56 Å), except that tENTH helix α1 is longer by seven residues. The tENTH domain overlays with a variety of ENTH domains and is structurally most similar to yeast Ent1 (Figure A2-3A), as measured by r.m.s.d. scoring in CCP4 Superpose (Krissinel *et al.,* 2004).

A second consequence of the missing helix α0 is that tENTH lacks a basic binding pocket for binding a phosphoinositide (Figure 3-3B). The tENTH surface is mostly hydrophobic in character with no major basic (or acidic) pockets or regions, while in contrast, the epsin1 ENTH possesses a highly basic binding pocket. We also considered the possibility that our domain might be similar to the N-terminal portion of an ANTH domain. A sequence alignment using ClustalW (data not shown; Sievers *et al.,* 2011) between the tENTH and CALM ANTH (residues 1–260) revealed that none of the Ptd(Ins)4,5P_2 binding residues in CALM (K38, K40, H41) are conserved between the two domains. We also tested experimentally whether the tENTH domain could bind a panel of phosphoinositide head groups (Figure 3-3B). We incubated recombinant purified

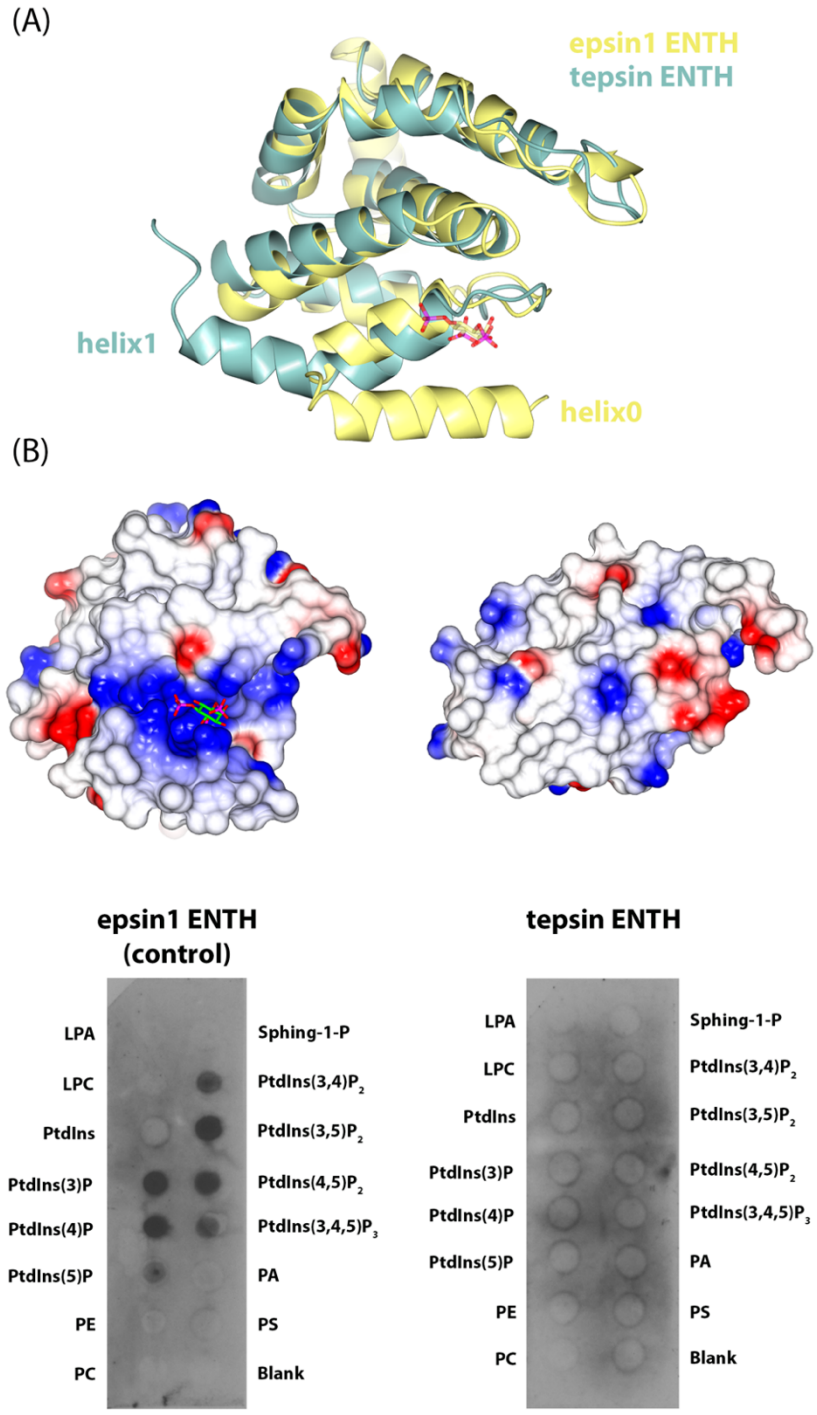


Figure 3-3. Structural and functional comparison of epsin1 and tepsin ENTH domains. (A) Overlay of epsin1 ENTH (yellow; PDB: 1H0A) and tENTH (green) domains. Tepsin lacks the amphipathic helix 0 found in other ENTHs and instead contains an elongated helix1. (B) Equivalent electrostatic surface views of epsin1 ENTH (left) and tENTH (right). Epsin1 contains a basic binding pocket to accommodate a phosphoinositide head group, while tepsin lacks this binding pocket because of the absence of helix 0. Biochemical data using recombinant ENTH domains on PIP strips confirms the structural predictions that epsin1 ENTH recognizes phosphoinositides, while tepsin ENTH does not.

tENTH-H6 protein with PIP strips™, using epsin1 ENTH-H6 as a positive control, and blotted against the 6xHis tag. As predicted by our structure, the tepsin ENTH does not bind any phosphoinositide head groups (Figure 3-3B). Finally, the CALM ANTH domain contains a short helix α0 that does not bind a phospholipid head group directly but has been shown to insert into membranes (Miller *et al.,* 2015). Structural overlays (data not shown) again indicate tENTH is different: tENTH contains an elongated helix α1, as opposed to the short helix α0 plus α1 in CALM. Although one face of tENTH helix α1 is hydrophobic, nearly all these residues participate in packing interactions with helix α3 (cf. Figure 3-1) to stabilize the structure of the domain.

Recently, ENTH domains in yeast Ent1 and zebrafish epsin1 have been reported to bind very weakly to ubiquitin (~2mM K_D; Levin-Kravets *et al.,* 2016). We thus tested whether recombinant tENTH could interact with ^{15}N-labelled mono-ubiquitin in an NMR chemical shift perturbation experiment, because NMR is the most sensitive method for detecting weak interactions. We detect no interaction between tENTH and ubiquitin (Figure A2-3B).

Structure of the tepsin VHS-like domain

Multiple attempts to crystallize human tepsin VHS-like (tVHS) domain failed. We instead used sequence alignments (Figure A2-4A) to identify and undertake trials with five other species that contained 75-90% sequence identity to the human domain: cow (*Bos taurus*), mouse (*Mus musculus*), horse (*Equus caballus*), rat (*Rattus norvegicus*), and marmoset (*Callithrix jacchus*). We synthesized five constructs (GenScript), then expressed and purified these proteins in *E. coli*. We obtained crystals from the horse domain (residues 306-437), which diffracted beyond 1.85 Å. However, we could not obtain a solution to the phase problem using molecular replacement in Phaser (McCoy *et al.,* 2007) with a variety of VHS structures as input models. Instead, we expressed and purified a horse VHS-like derivative that contained two selenomethione residues at positions 319 and 379; these crystals diffracted beyond 2 Å. We used AutoSol in PHENIX (Adams *et al.,* 2010) for automated experimental SAD phasing and automated model building (full

details in Methods). We obtained excellent initial maps and a preliminary model containing 100 residues. We then used this model to obtain a molecular replacement solution for our native crystals, which diffracted to slightly higher resolution. The horse domain is 80% identical to its human counterpart and thus provides an excellent mammalian model.

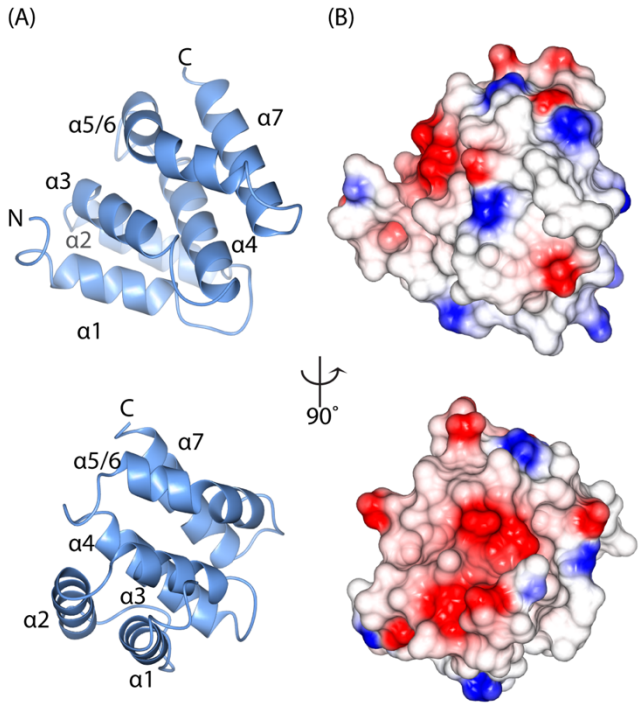


Figure 3-4. Crystal structure of the tepsin VHS/ENTH-like domain. (A) Ribbon diagrams of tVHS domain structure determined at 1.8 Å resolution; the views are rotated by 90 degrees. Unlike other VHS domains, tVHS contains only six α-helices. For consistency with other VHS domains, the helices are labelled α1-4, α5/6 (see Results), and α7. (B) Electrostatic surface representations of tVHS, shown in the same orientations as (A). tVHS contains a deep acidic pocket or groove on one surface.

The tVHS domain (Figure 3-4A; Figure A2-4) is an all α-helical protein containing six helices, which we have numbered α1-α5/6 and α7 (discussed in next section). The most notable feature in both the native and SeMet structures is that we do not observe density for the first twelve residues. Secondary structure prediction programs indicated tVHS contains seven α-helices, and all our constructs were designed to include this predicted helix but it is completely disordered in our structures. Instead, bioinformatics analysis using PSIPRED (Buchan *et al.,* 2013) suggests the first helix may constitute a 'disordered helix'. VHS domains normally contain eight α-helices, so like its ENTH counterpart, the tepsin VHS domain contains fewer α-helices than expected. However, the first α-helix observed in the density clearly corresponds to helix α1 found in other VHS domains (discussed below) based on structural comparisons. We used circular dichroism spectroscopy thermal denaturation experiments (data not shown) to ascertain whether

or not the α-helix is folded in solution. We conducted experiments using our crystallization construct (tVHS residues 306-437) and a shorter version (residues 321-437) that lacked the predicted helix. We could find no evidence to support an additional folded helix; if the helix ever folds, it may do so when the domain encounters a binding partner in the cell.

Structural and functional comparison of VHS domains

VHS domains are found in other trafficking proteins, including the Golgi-localized, γ-adaptin ear homology domain, Arf-binding protein (GGA) adaptor protein; Hrs/STAM subunits of the ESCRT-0 complex; and Tom1. Here we compare published VHS domains to tVHS and discuss functional implications. Overall, the first notable difference in tepsin is the position of the VHS-like domain in the middle of the protein. All other VHS domains are found at the N-terminus of proteins, and published VHS structures contain a right-handed superhelix of eight helices numbered α1-α8. However, our tVHS domain contains only six helices. In general, the tVHS structure overlays well with other VHS domains at helices α1-4 (Figure 3-5A, Figure A2-5); note that superpositions indicate the tVHS contains helix α1, so a possible disordered helix would be a new

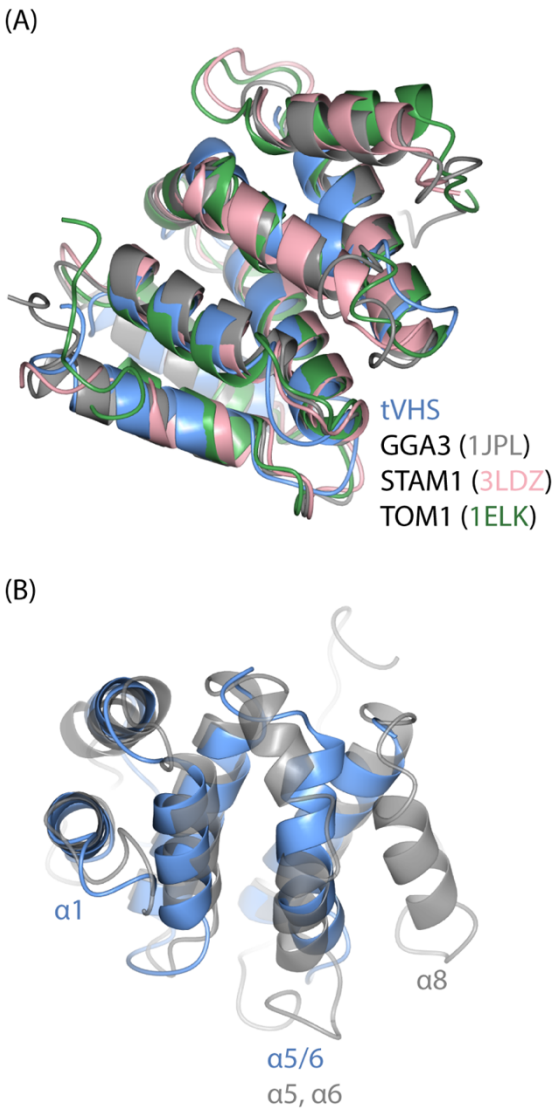


Figure 3-5. Structural comparison of VHS domains. (A) Overlay of VHS domain structures from tepsin, GGA3 (PDB: 1JPL), STAM1 (PDB: 3LDZ), and TOM1 (PDB: 1ELK). (B) Overlay of tepsin and GG3 VHS domains to highlight key structural differences; view is rotated 90 degrees relative to (A). tVHS aligns well with other VHS domains at helices 1-4. tVHS contains a bent α-helix (α5/6), while other VHS domains contain two separate helices (α5, α6). Helix 8 is absent in tVHS.

feature in VHS domains. Other VHS domains then have a short helix α5 followed by a flexible loop and helix α6. tVHS instead has one long, 'kinked' helix that we have named α5/6, because the first half overlays with helix α5 and the second half with helix α6 in other VHS domains (Figure 3-5B). tVHS helix α5/6 contains a proline at position 406, which causes the kinked structure. tVHS then overlays well once again at helix α7, and it lacks helix α8 altogether, based on both secondary structural predictions (data not shown) and our structural data.

GGA adaptor proteins. GGAs are monomeric adaptor proteins that localize to the TGN. The VHS domain of GGAs binds short acidic dileucine motifs (DxxL[L/I]; Misra *et al.,* 2002) found in the cytosolic portion of transmembrane protein cargo; one example is recognition of the cation-independent mannose-6-phosphate receptor (CI-MPR) by GGA1 and GGA3. One might predict that cargo recognition and binding is a possible function of the tepsin VHS domain within the AP4 coat. However, our structure explains why tVHS cannot recognize dileucine-based cargoes. In the GGA3 VHS (PDB ID: 1JPL), helices α6 and α8 are required for motif binding: both helices provide side chains that engage the dileucine motif (buried surface area of 542 $Å^2$), especially the conserved Leu residues, in shallow hydrophobic pockets. Because tVHS lacks helix α8 (Figure 3-5B; 3-6A), it cannot provide enough surface area to engage a dileucine motif. Indeed, we confirmed this experimentally using a model dileucine cargo peptide in isothermal titration calorimetry experiments with recombinant purified tVHS domain (Figure 3-6A).

Hrs/STAM. Hrs and STAM constitute two subunits of the ESCRT-0 complex, which targets ubiquitinated proteins to the proteasome for degradation. Hrs and STAM VHS domains bind mono-ubiquitin (Ub) *in vitro* (Ren *et al.,* 2010), and published surface plasmon resonance experiments (Ren *et al.,* 2010) have suggested all VHS domains may bind mono-ubiquitin with a range of affinities (from low millimolar to high micromolar). We thus considered the possibility that tepsin interacts with Ub. We tested experimentally whether recombinant tVHS could interact with ^{15}N-labelled Ub in an NMR chemical shift perturbation experiment (Figure 3-6B), because this

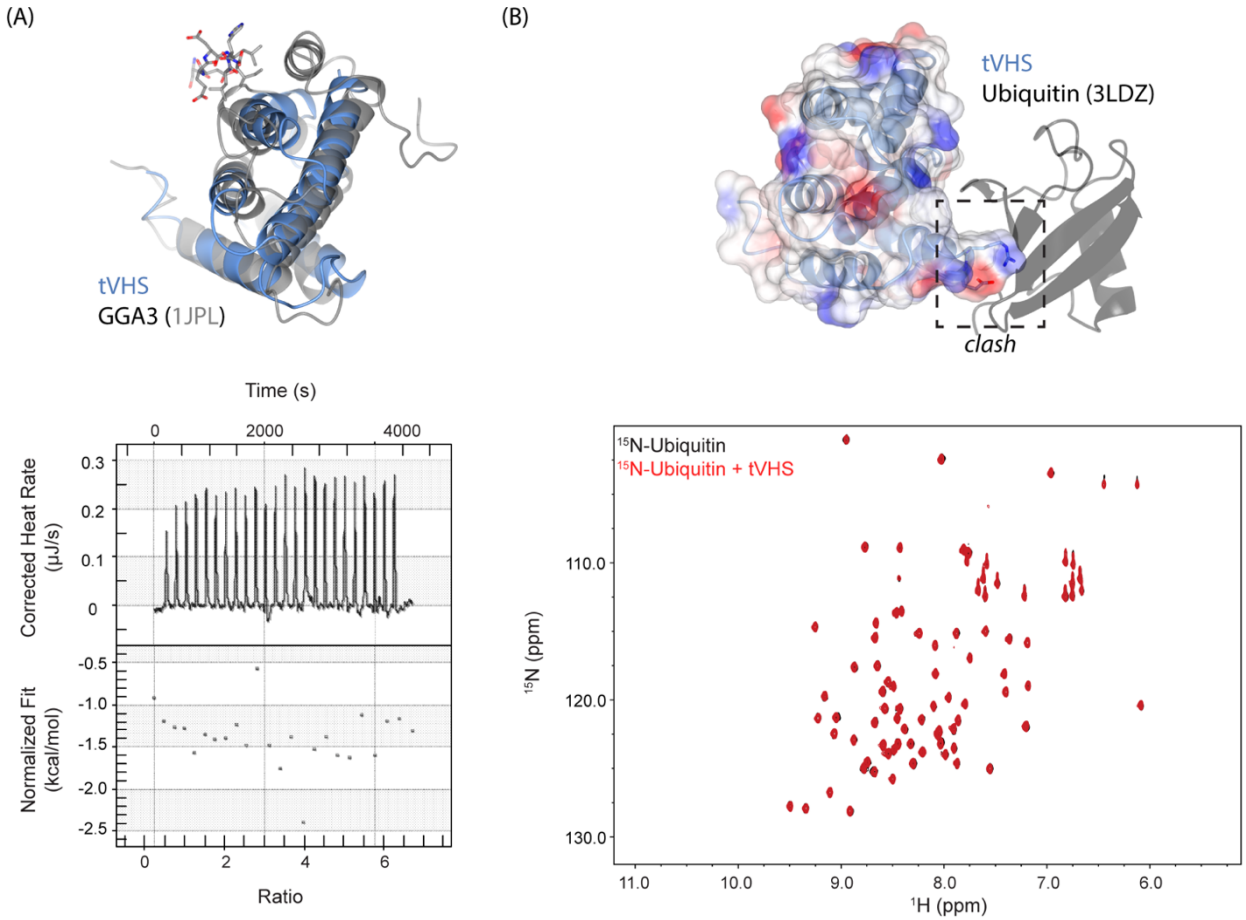


Figure 3-6. tVHS structure explains functional divergence from other VHS domains. (A) Top panel: overlay of tepsin (blue) and GGA3 (grey) VHS domains. The acidic dileucine motif recognized by GGA3 is shown in grey stick figures. GGA3 VHS helix 8 contributes key surface area to accommodate dileucine motif binding; loss of this helix prevents binding by tVHS. Bottom panel: ITC run with recombinant tVHS protein and 10x molar excess of a model acidic dileucine motif (DSVIL), demonstrating no binding. (B) tVHS ribbon and transparent electrostatic surface modeled by overlaying the STAM1-Ub structure (PDB: 3LDZ). tVHS lacks key residues that form a Ub binding surface, and an arginine present in tVHS would clash with ubiquitin. Bottom panel: HSQC experiment of ^{15}N-labeled ubiquitin (black) spectrum overlaid with a spectrum collected in the presence of a ten times molar excess of tVHS domain. We observe no chemical shifts, suggesting that tVHS cannot bind ubiquitin.

technique can detect weak (high millimolar) interactions. We observed no binding between tVHS and Ub; in contrast, STAM1 VHS binds Ub with a K_D ~220 µM (Ren *et al.,* 2010). Our structure again explains why tVHS cannot engage Ub: an arginine at tVHS residue 344 would directly clash with Ub (Figure 3-6B).

Structural comparison of tepsin ENTH & VHS-like domains

Once we determined both structures, a surprising feature emerged. Based on r.m.s.d. values, the tepsin ENTH and VHS-like domains appear more similar to each other than to most other domains in either family (Figure A2-5). In other words, at the structural level, tVHS "looks" more like certain ENTH domains than VHS domains. That tVHS appears similar to ENTHs might be explained by the lack of helix 0 in many published ENTH structures. For example, deposited epsinR structures either lack helix 0 altogether (PDB: 2QY7) or fail to resolve helix 0 in the absence of its phosphoinositide (PDB: 1XGW). We also fail to resolve the first predicted helix in our tVHS construct, and tepsin is unusual in having an internal folded domain at all. Together, these data raise the question of whether the tVHS domain is actually an ENTH domain that arose from a gene duplication event. To address this question, we also tested and verified biochemically that recombinant tVHS does not bind a panel of phosphoinositides (data not shown). Furthermore, our tVHS structure reveals a very acidic patch (Figure 3-4B) that would strongly repel most anionic phospholipid head groups found in membranes.

In yeast, ENTH and ANTH proteins have been shown to interact with each other in the presence of a phosphoinositide ligand (Skruzny *et al.,* 2012). We considered the possibility that tENTH and tVHS domains could function using a similar principle, although unlike the yeast proteins, we find neither tENTH nor tVHS domains bind phosphoinositides on their own. We tested for direct binding using pulldown experiments (GST-tENTH with tVHS-H6 and GST-VHS with tENTH-H6). We observed no interactions between these domains at either the Coomassie or Western levels (data not shown).

Phylogenetics and comparative genomics

Our structural, biochemical, and biophysical data raised several questions about tepsin evolution. We thus decided to use phylogenetics and comparative genomics first to analyze how tepsin ENTH and VHS domains evolved, and second, to understand the evolution of tepsin within the epsin family. Our structural data strengthened the possibility that the tENTH and tVHS

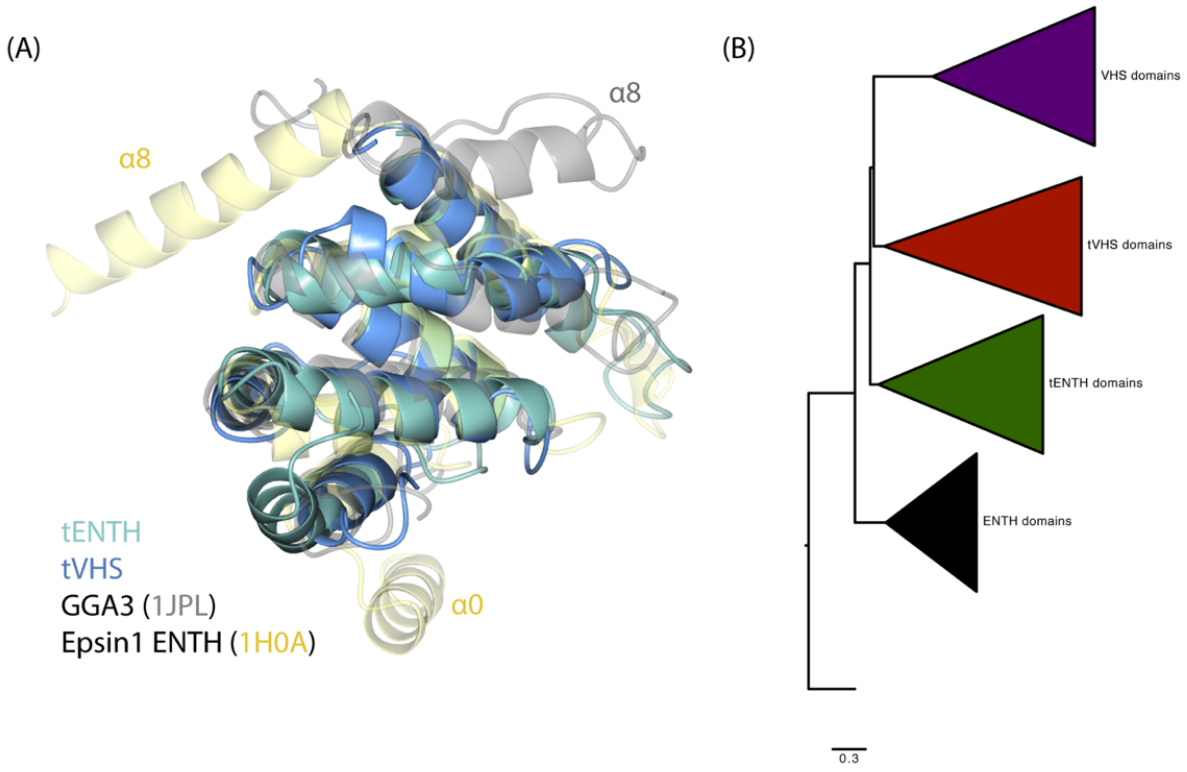


Figure 3-7. Evolution of ENTH and VHS domains. (A) Structural overlay of tENTH, tVHS, epsin1 ENTH, and GGA3 VHS domains. All four domains demonstrate a conserved structural core containing helices α1-7. (B) Phylogenetic tree of ENTH, tepsin ENTH (tENTH), VHS, and tepsin VHS-like (tVHS) domains, with ANTH domains as an outgroup.

domains arose from a gene duplication event. In addition, the smaller size of both domains might imply they constitute a 'minimal core' structure of the ENTH/ANTH/VHS superfamily (Figure 3-7A), in which helix 0 and/or helix 8 are absent. We conducted phylogenetic analyses on ENTH and VHS domains from a variety of species representing five eukaryotic supergroups (Figure 3-7B). We used ANTH domains as an outgroup (DeCrane *et al.,* 2012): ANTH domains are about twice the size of ENTH or VHS domains but they contain the same core of 6–8 α-helices. Our analysis reveals that ENTH, tENTH, VHS, and tVHS domains each form monophyletic clades as expected, with the VHS and tVHS domains more closely related to one another than to other domains. The tVHS domains in two species (*Aplysia californica* and *Crassostrea gigas*) appear to be particularly divergent and do not occur in a consistent location on the phylogeny (Figure A2-

7). The phylogenetic analysis further reveals that VHS domains are generally acquiring mutations at a higher rate than ENTH domains, with non-tepsin VHS domains showing the highest within-group genetic distance (Table A2-9).

We also conducted a phylogenetic analysis across the epsin family to learn how tepsin fits into the tree. Previously published trees (DeCraene *et al.,* 2012; Gabernet-Castello *et al.,* 2009) did not include tepsin, likely because it had not yet been discovered or annotated in most genomes. We obtained sequences representing species across the five eukaryotic supergroups, again using CALM/PI-CALM (an ANTH-containing protein) as an outgroup (DeCraene *et al.,* 2012). Our tree (Figure 3-8, Figure A2-8) reveals tepsin forms its own monophyletic clade within the family.

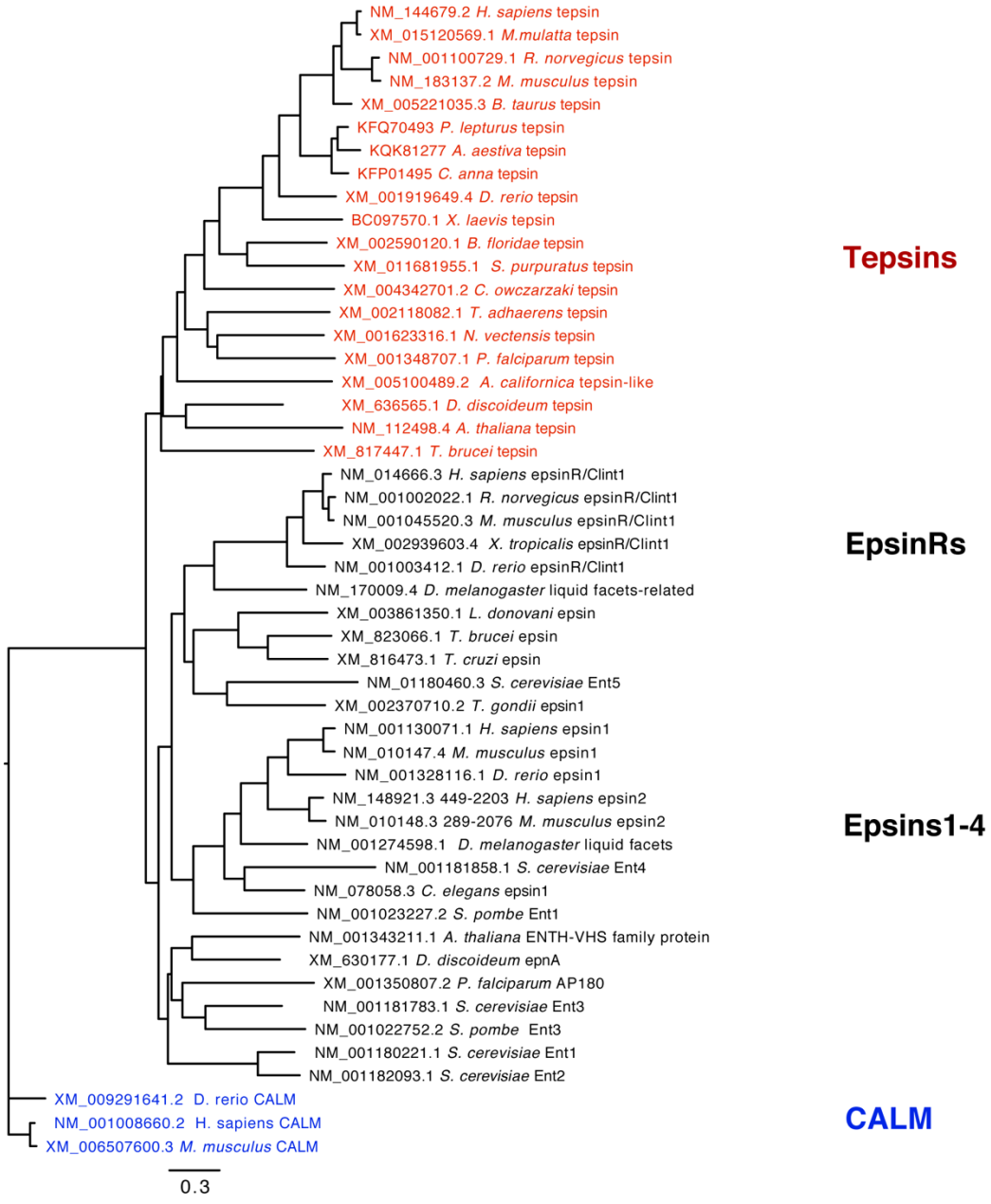


Figure 3-8. Epsin family phylogenetic tree. Tepsin forms a monophyletic clade that likely diverged early in the evolutionary history of eukaryotes. (CALM sequences, which contain ANTH domains, were used as an outgroup.)

Discussion

Our phylogenetic data do not support the idea of a straightforward gene duplication event giving rise to the unique second folded domain in tepsin; they instead suggest a more complicated duplication event. It is not possible to interpret whether our domain tree supports the idea of tepsin domains constituting a 'minimal core' structure. One might argue that a common ancestor contained only seven helices: tepsin domains maintained this structure, while other ENTHs or VHS domains picked up additional helices (helix 0 and/or helix 8) as they evolved to undertake different functions. Alternatively, the ancestral domain may have contained nine helices in which a single loss of helix 0 occurred in tENTHs and a loss of helix 8 in both tepsin domains. Both scenarios are equally parsimonious on the current tree. But many other permutations are possible, and it is also possible we have not fully sampled the landscape.

In contrast, our phylogenetic data reveal a much clearer picture of tepsin in the context of the epsin family. This tree indicates tepsin diverged away from other epsins early in the evolutionary history of eukaryotes (~1,500 million years ago), and we predict that tepsin diverged away to support different biological functions. The monophyletic tepsin clade includes sequences from plants, animals, and algae; it is thus parsimonious to suggest a secondary loss of tepsin in yeast and insects as opposed to multiple evolutionary origins of tepsin-like proteins. Our phylogenetic analysis is consistent with our biochemical and biophysical data demonstrating how both tepsin domains fail to undertake known functions of other ENTH or VHS domains. Other ENTH and VHS domains are implicated in phosphoinositide recognition (Ford *et al.,* 2002; Boal *et al.,* 2015); ubiquitin binding (Levin-Kravets *et al.,* 2016; Ren *et al.,* 2010); or dileucine-based cargo motif binding (Misra *et al.,* 2002). We predict both tepsin domains will likely engage a protein partner. One possible binding partner for the ENTH domain is a SNARE protein; epsinR specifically recognizes the SNARE Vti1b (Hirst *et al.,* 2004; Chidambaram *et al.,* 2004). There is currently no evidence that tepsin ENTH engages a SNARE protein (Borner *et al.,* 2012; Hein *et*

al., 2015). However, we note that extension of tENTH helix α1 compared to other ENTH domains could provide additional surface area for engaging a SNARE or other protein binding partner.

The VHS domain has been proposed to interact directly with AP4, based on co-immunoprecipitation (IP) experiments using a variety of deletion constructs in HeLa cells (Mattera *et al.,* 2015). We could find no evidence for an interaction in HeLa cell lysates using a C-terminally GFP-tagged construct containing the ENTH and VHS domains (residues 1-356; Figure A2-6); a construct containing only the VHS domain (residues 225-356) repeatedly failed to express upon transfection, while our positive control efficiently immunoprecipitated AP4. The micromolar interactions between specific tepsin motifs and the AP4 appendage domains (Frazier *et al.,* 2016; Mattera *et al.,* 2015) are likely most important. If there is a tertiary interaction between tVHS and AP4, it must be very weak (beyond high millimolar) and likely only relevant in an assembled coat when avidities are high.

Our structures raise broader questions about protein classification, specifically regarding the differences between ENTH and VHS domains. ENTH and VHS domains show clear differences when classified at the sequence level, and one anticipates that structural evidence would explain these differences. Previously published structures suggest two concrete structural differences: ENTHs contain helix 0 while VHS domains do not; and the position of helix 8 differs between ENTH and VHS domains. However, tepsin domains lack all features that normally differentiate ENTHs from VHS domains. Surprisingly, phylogenetic data indicate tepsin VHS sequences cluster with VHS domains, while tENTH and ENTH domains are ancestral to this VHS/tVHS clade. It is difficult for us to rationalize or explain precisely what features generally differentiate ENTH from VHS domains in light of new structural data from tepsin. Others have noted the importance of sampling broadly across evolutionary space to understand gene and protein function. One relevant trafficking example is the role of ENTH/ANTH proteins in trypanosomes (*T. brucei*). Here, TbCALM and TbEpsinR function with clathrin but independently of AP2 (Manna *et al.,* 2015). Our data further support the idea that we miss important features

and functions by sampling narrowly across eukaryotic lineages. Perhaps we cannot fully classify tepsin domains, or any protein domain, in the absence of functional data.

Materials and methods

Reagents

Unless otherwise noted, all chemicals were purchased from Sigma (St. Louis, MO, USA). The following antibodies were used in this study: rabbit anti-AP4 β, rabbit anti-AP4 ε (for western blotting and immunoprecipitation; both in-house; Hirst *et al*., 1999), mouse anti-AP4 ε (for immunofluorescence; 612019; BD Transduction Labs), rabbit anti-clathrin (in-house; Simpson *et al*., 1996), rabbit anti-GFP (gift from Matthew Seaman, Cambridge Institute for Medical Research, UK), and rabbit anti-tepsin (in-house; Borner *et al*., 2012). Horseradish peroxidase (HRP)–conjugated secondary antibodies were purchased from Sigma-Aldrich, and fluorescently labelled secondary antibodies were from Invitrogen. For Western blotting of immunoprecipitates where the protein band of interest was close to an IgG band, protein-A-HRP (BD Biosciences) was used in the place of HRP-conjugated secondary antibody.

Molecular biology and cloning

All GST-fusion proteins of the tepsin ENTH domain were sub-cloned from full-length tepsin (Borner *et al.,* 2012) into the BamHI/SalI sites of pGEX-6P-1. Because there are no tryptophan residues in the ENTH domain, a single tryptophan was added at the C-terminus to facilitate light absorbance measurements and estimates of protein concentration. The following constructs were made in this way: GST-tENTH (residues 1–136W); GST-tENTH-H6 (residues 1–136W with a C-terminal 6xHis tag; and GST-tENTHlong (residues 1–153W). Horse tepsin VHS-like domain (NCBI reference sequence: XM_001489994.4) was synthesized by Genscript and sub-cloned into BamHI/SalI sites of pGEX-6P-1.

Protein expression and purification

Human tepsin ENTH and horse VHS-like domain constructs were expressed in BL21(DE3)pLysS cells (Invitrogen) for 16–20 hours at 22˚C after induction with 0.4 mM IPTG. Native horse tVHS-like domain was expressed in BL21(DE3)pLysS cells (Invitrogen) for 16–20 hours at 22˚C after induction with 0.4 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) at OD_{600}=1.0. Selenomethionine labeled (SeMet) tVHS-like domain was expressed in BL21(DE3)pLysS cells (Invitrogen) with incorporation by metabolic inhibition. Briefly, cells were grown in minimal media to OD_{600}=0.3, when amino acid supplements (Lys, Phe, Thr, Leu, Ile, Val, and SeMet) were added to induce metabolic inhibition of methionine synthesis and supply selenomethionine. Cells were induced with 0.4 mM IPTG at OD_{600}=0.8, and expression occurred for 16 hours at 22˚C.

Tepsin ENTH was purified in 20 mM Tris (pH 8.5), 250 mM NaCl, 2 mM βME. Human tepsin VHS-like was purified in 20 mM Tris (pH 8.7), 200 mM NaCl, 2 mM βME. Native horse tVHS was purified in 20 mM HEPES, pH 7.5, 200 mM NaCl, 2 mM DTT buffer. SeMet horse tVHS was purified in 20 mM HEPES, pH 7.5, 200 mM NaCl, 10 mM DTT buffer. AEBSF protease inhibitor (Calbiochem) was used at all early stages of purification. Cells were lysed by a disruptor (Constant Systems Limited, Daventry, UK), and proteins were affinity purified using glutathione sepharose (GE Healthcare) in relevant purification buffers. GST-fusion proteins were cleaved overnight at 4˚C using in-house recombinant GST-3C protease and eluted in batch. All proteins were further purified by gel filtration on a Superdex S200 preparative or analytical column (GE Healthcare).

Crystallization, structure determination, and structural comparisons

tENTH domains. Purified tENTH (residues 1–136W) was concentrated to 3-5 mg/ml and crystallized in 12% (w/v) PEG3350, 4% (v/v) tacsimate pH 5.0. Crystallization trays were set up using 200nL drops on a Mosquito robot (TTP LabTech). Crystals were harvested directly from 96-well plates into 500nL drops in well buffer plus 25% glycerol for cryo-protection. Data were

collected at Diamond Light Source, beamline I04-1, on a Pilatus 2M detector and integrated using Xia2 and Mosflm. Crystals diffracted beyond 1.4 Å resolution and were of space group C2 2 21 a=80.8 Å, b=84.8 Å, c= 80.9 Å, α= 90˚, β=90˚, γ=90˚, in Pointless (Evans, 2011). There were two molecules in the asymmetric unit, and no translational NCS was detected in either Scala or Phaser. Because of the high resolution, the structure was determined by molecular replacement in Phaser by individually placing α-helices in the density to identify the seven α-helices in the ENTH domain. Initial building was done in ARP/wARP (Perrakis *et al.,* 2001). Rounds of refinement and rebuilding were undertaken in phenix.refine (Afonine *et al.,* 2005) and Coot (Emsley *et al.,* 2010), respectively. The longer tENTH construct comprising residues 1–153W was concentrated to 3 mg/ml and crystallized in 150 mM MES monohydrate pH 6.0, 15% (w/v) PEG6000. Crystallization trays were set up and harvested as described for the tENTH 1–136W construct. Data were collected at Advanced Photon Source (APS) beamline 21-ID-F (0.979 Å) using a Rayonix MX300 CCD detector. Crystals diffracted to 1.8 Å resolution and were of spacegroup P4 21 2 with cell dimensions a= 90.2 Å, b=90.2 Å, c= 42.8 Å, α=90˚, β=90˚, 90˚. The structure was determined using molecular replacement in Phaser with the tENTH 1-136 structure as a model. Rounds of refinement and rebuilding were undertaken in phenix.refine and Coot, respectively. All structure figures were generated using CCP4MG (McNicholas *et al.,* 2001).

tVHS-like domain. Native tVHS-like domain was concentrated to 5-7 mg/ml and spin-filtered (Millipore). SeMet tVHS-like domain was concentrated to 8.75 mg/ml and spin-filtered (Millipore). A Mosquito robot (TTP Labtech, Cambridge, MA) was used to set the protein in sitting drops. Multiple hits were obtained from the PEG/Ion HT screen (Hampton Research, Aliso Viejo, CA). The plates were stored and imaged at room temperature by Rock Imager (Formulatrix, Bedford, MA). Crystals were cryoprotected using 25% glycerol or perfluoropolyether cryo oil (Hampton Research) by flash freezing in liquid nitrogen. The best native data were collected from crystals grown in 0.2M sodium tartrate dibasic dehydrate, 20% w/v polyethylene glycol 3,350, pH

7.2, and cryoprotected with perfluoropolyether cryo oil. The best SeMet data were collected from crystals grown in the same condition, using 25% glycerol as a cryo-protectant.

Native data were collected at the Advanced Photon Source (APS) beamline 21-ID-G (0.97857 Å) using a MAR300 CCD detector. SeMet data were collected on beamline 21-ID-D (0.97910 Å) using a Dectris Eiger 9M detector. Native crystals diffracted to 1.85 Å and were of space group P61. The unit cell dimensions were a = 58.828 Å, b = 59.929 Å, c = 69.078 Å and α = 90˚, β = 90˚, γ = 120˚. SeMet crystals diffracted to 1.95 Å and were of space group P61. The unit cell dimensions were a = 58.775 Å, b = 58.775 Å, c = 69.522 Å, α = 90˚, β = 90˚, γ = 120˚.

Both native and SeMet data were integrated using HKL2000, then processed further using either the CCP4 or Phenix suites. We used Autosol in Phenix for automated SAD phasing and model building; the initial model contained 109 of 133 residues. Additional rounds of model building was undertaken in Coot with iterative rounds of refinement in phenix.refine. The native structure was determined by molecular replacement using the final SeMet model. Final refinement and validation runs were performed using phenix.refine and MolProbity.

Despite extensive manual inspection of the density maps, twelve N-terminal residues (306–317 in the horse sequence) could not be placed in the electron density. Although secondary structural prediction programs suggested these residues formed a helix, we observe no visible density that would allow us to confidently place these N-terminal residues. We suspect this disordered region contributes to our R-factors being higher than expected at this resolution.

Structural comparisons. Superpose (CCP4) was used to compare structures of ENTH and VHS domains deposited in the PDB. The SSM algorithm was used to align the structures and to determine the RMSD, Q-score, and number of residues aligned between structures.

Immunoprecipitations and Western blotting.

PIP strips™ (Life Technologies) were used to probe ENTH and VHS domain binding to phosphoinositides. PIP strips were blocked for 1 hour at 23˚C and then incubated with 2 µg/mL of Epsin1-H6, tENTH-H6 or tVHS-H6 recombinant purified proteins for 1 hour at 23˚C. PIP strips

were subsequently washed and incubated with a 1:3000 dilution of anti-6X His tag® [GT359] HRP antibody (abcam®) at 4˚C overnight. PIP strips were visualized using Amersham ECL Western blotting detection reagents (GE Healthcare).

HeLa cells (ATCC®) were grown on T-75 cm2 flasks to 80-90% cell density. Cells were then transfected with EGFP-N1, tepsin FL (1-525)-GFP, tENTH (1-136)-GFP, or tENTH+tVHS (1-356)-GFP using LipfectamineTM 3000 Reagent (Invitrogen). Cells were imaged and harvested 24 hours post transfection. Cells were resuspended and lysed in 1 mL 10 mM HEPES pH 7.5, 150 mM NaCl, 0.5% NP-40, 0.02 mg/mL AEBSF-HCl (EMD Millipore), and one cOmplete™ Mini EDTA-free Protease Inhibitor Cocktail (Roche). Cell slurry was incubated on ice for 30 minutes and mixed every ten minutes. Cell slurry was centrifuged at 20,000 x RCF for 15 minutes, removed, and soluble fraction saved. GFP-Trap_A resin (ChromoTek, gta-20) batch equilibrated with 10 mM HEPES pH 7.5, 100 mM NaCl, 0.5 mM EDTA, 0.5% NP-40. Resin slurry (60 uL) was added to each soluble cell fraction. Samples were then incubated and rotated for 1 hour at 4°C. Samples were centrifuged at 2500 x RCF for 5 minutes. Supernatant was removed, and resin was washed with 1 mL of wash buffer for a total of three washes. After final removal of supernatant fraction, 80uL 2XSDS loading buffer was added to each sample and boiled at 95°C for 10 minutes. Boiled samples were centrifuged at 5000 x RCF for 5 minutes, then supernatant was remove and transferred to a new tube. The samples were used for western blotting and probed with anti-GFP-HRP at 1:5000 (abcam) and anti-AP4ε at 1:1000 (gift from Margaret Robinson lab).

Isothermal titration calorimetry.

ITC experiments were conducted on a NanoITC instrument (TA Instruments) at 20˚C. A model dileucine peptide (DSVIL) was dissolved in 20 mM sodium phosphate (pH 6.5), 200 mM NaCl, 0.5 mM TCEP. A PD-10 desalting column (GE Healthcare) was used to buffer exchange human tVHS (residues 220–360) into the same buffer as the peptide. Incremental titration was performed with an initial baseline of 180 s and injection intervals of 180 s. Peptide was titrated in to a ratio of 6x molar excess. Data were analyzed in NanoAnalyze (TA Instruments).

NMR spectroscopy.

Uniformly enriched ^{15}N-labeled mono-ubiquitin (gift from Natalja Pashkova) was diluted in 50 mM sodium phosphate pH 7.0, 1 mM DTT to 30 μM, with 10% v/v D_2O. Samples containing a ten times molar excess of either human tVHS (residues 220-360) or human tENTH (1-136W) were prepared by diluting in 50 mM sodium phosphate pH 7.0, 1 mM DTT to 30 μM, with 10% v/v D_2O. Standard two-dimensional ^{15}N-^{1}H HSQC NMR spectra were collected at 25˚C on a 800-MHz Bruker Avance III spectrometer with a CPTCI triple resonance cryoprobe (Bruker BioSpin). Data were processed in Topspin 3.2 (Bruker BioSpin), with zero filling in the indirect dimension and squared sine bell apodization in both dimensions.

Phylogenetic analyses.

To construct phylogenetic trees, we first assembled nucleotide sequences that encode tepsin and epsin proteins. Sequences were identified as tepsin or epsin based on (1) references in the published literature (Borner *et al.,* 2012; DeCreane *et al.,* 2012; Gabernet-Castello *et al.,* 2009); (2) annotations and sequence names in UNIPROT or NCBI (tepsin annotations also include ENTHD2, and epsin annotations also include clint1, epn, and ent); and (3) sequence homology to known proteins, assessed via tBlastn). We also included CALM/PI-CALM (an ANTH-containing protein) sequences from three species as an outgroup clade (DeCreane *et al.,* 2012). In total, we compiled 50 sequences for this analysis. Sequences were aligned using the E-INS-i algorithm in MAFFT (Katoh *et al.,* 2013), which is optimized for sequences with sections of homology that are potentially separated by stretches of non-homology. We made 100 bootstrap replicates of the aligned sequences using *seqboot* in the Phylip package (Felsenstein, 2005), built a maximum-likelihood tree for each replicate using *dnaml*, and obtained a consensus sequence using *consense*. To further refine this tree and to estimate branch lengths, we used the consensus phylogeny as a starting tree for 100 iterations of the PASTA (Practical Alignments using SATé and TrAnsitivity) algorithm (Mirarab *et al.,* 2014). On each iteration, this algorithm splits the tree into subsets, aligns the subsets with the MAFFT algorithm, merges the subsets with the MUSCLE

algorithm (Edgar, 2004), and re-analyzes the tree with the RaxML algorithm (GTR-GAMMA model; Stamatakis, 2014), ultimately returning the tree that optimizes the maximum-likelihood score. In addition, we compiled nucleotide sequences from specific domains: ENTH and VHS domains from epsin sequences, tENTH and tVHS domains from tepsin sequences, and ANTH domains from CALM/PI-CALM sequences. We constructed a phylogeny for 121 domain-specific sequences using the same methods described above.

Acknowledgements

The authors thank Linton Traub (Pittsburgh) for the rat epsin1 GST-ENTH-H6 construct; Natalja Pashkova and Rob Piper for ^{15}N-labeled ubiquitin; Jenny Hirst, Alexandra Davies, and Margaret Robinson for AP4 antibodies; and LS-CAT beamline scientists Elena Kondrashkina and Zdzislaw Wawrzak for assistance with remote data collection. 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). The crystallization trials described here used the Vanderbilt robotic crystallization facility supported by NIH S10 RR026915. TLA, MNF, AM, AKK, and LPJ are supported by NIH 1R35GM119525, and LPJ is a Pew Scholar in the Biomedical Sciences, supported by the Pew Charitable Trusts. MNF was supported in part by NIH Molecular Biophysics 2T32GM008320-26. The authors declare they have no conflicts of interest.

CHAPTER IV

IN SEARCH OF BINDING PARTNERS FOR THE TEPSIN ENTH AND VHS/ENTH-LIKE DOMAINS

Introduction

Since neither tepsin domain performed any of the canonical ENTH or VHS functions ascribed in the literature (see Chapter III), I turned to screening methods to identify potential protein binding partners for these two domains. Commercial yeast two-hybrid screens (Y2H; Hybrigenics) were performed using the ULTImate Y2H human adult brain libraries and either the tENTH domain or the tVHS domain as baits. The tENTH domain was prepared as a C-terminal fusion to LexA or Gal4, and the tVHS domain was used as a N-terminal fusion to LexA or Gal4. Hybrigenics scores the positive hits on an A–F scale, where A is a very robust hit with high confidence. We focused on A or B hits where the identified protein had some relevance to trafficking or appeared frequently in other published screens.

tENTH domain candidates

The top results from the Y2H screen as well as SILAC mass spectrometry based proteomics data (Tara Archuleta, unpublished) included several TRIM (tripartite motif) proteins, TRIM2 and TRIM23, as well as c16orf62. TRIM proteins contain three zinc finger domains—usually followed by a coiled coil domain—and have RING E3 ubiquitin ligase activity. TRIM2 is highly expressed in the nervous system and ubiquitinates neurofilament light chain (NF-L; Balastik *et al.,* 2008). TRIM23/ARD1 has one RING, one B-box type 1, and one B-box type 2 zinc fingers—which give rise to E3 ligase activity—as well as a GTPase activating domain, ADP-ribosylation factor domain, and a guanine nucleotide binding domain (Mishima *et al.,* 1993; Vitale *et al.,* 1996; Vichi *et al.,* 2005). TRIM23 is known to localize to the Golgi and lysosomes, and may play a role in intracellular trafficking (Vitale *et al.,* 1998). Recently, TRIM23 has been identified as essential for autophagy in response to viral infection (Sparrer *et al.,* 2017). TRIM2 and TRIM23 both received B, or high confidence, interaction scores from Hybrigenics. The interacting region was residues 97–294 for TRIM2 and residues 263–372 for TRIM23. In both cases these selected interaction domains cover all or part of the B-box zinc finger domains.

c16orf62 was recently identified bioinformatically as part of the large multisubunit protein complex Commadner, which likely functions in endosomal protein sorting (Mallam *et al.,* 2017). Furthermore, it was identified and characterized in cells as a core subunit in the retromer-like complex retriever, which sorts proteins from the same endosomal compartments as retromer (McNally *et al.,* 2017). c16orf62 appears to be most similar to the VPS35 subunit of retromer, and it has been renamed to VPS35L for VPS35-like, which we have adopted. VPS35 is the largest subunit in retromer, acting primarily as a structural scaffold for VPS26 and VPS29; it forms a right handed, α-helical solenoid (Haft *et al.,* 2000; Hierro *et al.,* 2007). The core heterotrimer of retriever consists of VPS35L, DSCR3 (a paralogue of VPS26; Koumandou *et al.,* 2011), and VPS29. The Y2H analysis scored VPS35L an A, or very high confidence, hit and revealed residues 474–634 as the likely interacting domain.

tVHS domain candidates

The top results from the Y2H screen included FERM (Band4.1, ezrin, radixin, moesin) domain containing proteins ezrin and moesin as well as ISCA1, an iron sulfur cluster assembly protein. FERM domain containing proteins are believed to link the membrane to the cytoskeleton through interactions between the FERM domain and membrane associated proteins (Yonemura *et al.,* 1998) and between the C-ERMAD (C-terminal ERM-association domain) and F-actin (Turonen *et al.,* 1994). Thus, these proteins could potentially organize specialized membrane domains (reviewed in Fehon *et al.,* 2010). Ezrin and moesin were rated B interactions from the Y2H screen. In ezrin, residues 139–278 were identified as the interacting region, which covers part of the FERM domain. For moesin, the likely interacting region is residues 82–329, which partially covers the FERM domain and the beginning of the coiled-coiled domain.

ISCA1 is described as a mitochondrial iron sulfur cluster assembly protein (Kaut *et al.,* 2000), but one group also published evidence it exists in the cytosol (Song *et al.,* 2009). However, this cytosolic localization was subsequently disputed by Roland Lill's group as an artifact of tag

placement or overexpression, as their experiments with endogenous ISCA1 showed no detectable amounts in the cytosol (Sheftel *et al.,* 2012). Iron sulfur clusters are inorganic cofactors important for many biological processes, functioning in electron transfer and catalysis reactions as well as providing a structural role in proteins (reviewed in Braymer and Lill, 2017). Iron and sulfide concentrations are tightly regulated due to their redox potential, so a complex Fe/S biogenesis pathway has evolved. Fe/S cluster assembly (ISC) and cytosolic Fe/S protein assembly (CIA) proteins comprise this biogenesis assembly pathway. *De novo* cluster assembly occurs on scaffold proteins; the cluster is then passed along Fe/S biogenesis machinery to acceptor proteins that need the cluster to function (Braymer and Lill, 2017). These transfer processes are less well understood. Recent work determined ISCA1 is required for mitochondrial Fe_4S_4 protein biogenesis and can participate in a heterocomplex with ISCA2 (Beilschmidt *et al.,* 2017).

This hit appeared numerous times in our Y2H report and an older Y2H screen with full-length tepsin (Jenny Hirst, personal communication). A tepsin-ISCA1 interaction is also noted in several different proteomics screens as well (both published and deposited online; Beilschmidt *et al.,* 2017; Hein *et al.,* 2015; Huttlin *et al.,* unpublished). The interacting region identified by the Y2H screen encompasses the entire 129 amino acid protein, with high confidence in the hit (A grade). The selected tENTH and tVHS Y2H hits are summarized in Table 4-1, below.

Table 4-1: Summary of selected protein binding partners for tENTH and tVHS domains from Hybrigenics Y2H screen.

Protein hit	Interacting region (residues)	Strength of hit	Tepsin domain partner
TRIM2	97-294	B	tENTH
TRIM23	263-372	B	tENTH
C16orf62 (VPS35L)	474-634	A	tENTH
Ezrin	139-278	B	tVHS
Moesin	82-329	B	tVHS
ISCA1	1-129	A	tVHS

Materials and methods

Cloning

TRIM2 ordered from HarvardPlasmID database was subcloned via SalI and NotI into pGEX6P1. The N-terminus (RING, B box, CC domains; residues 1–290) was subcloned into pGEX6P1, using BamHI and SalI. TRIM23 was also ordered from HarvardPlasmID database and subcloned into pGEX6P1 using BamHI and SalI sites. The N-terminus (RING, B Box, CC domains; residues 1–385) was also subcloned using BamHI and SalI into pGEX6P1, and a C-terminal 6x His tag was added.

VPS35L constructs were synthesized by Genscript (New Jersey, USA) in pGEX6P1 and pcDNA3.1_N-eGFP using BamHI and XhoI restriction sites. Subcloning into the *in vitro* wheat germ vector (pF3A WG) used SgfI and PmeI restriction sites to generate the full-length (residues 1–1050), ΔN-term (residues 87–1050), and Y2H_ID (residues 438–688) constructs with N-terminal myc tags (C-terminal myc tag for the ΔN-term construct).

ISCA1 constructs were synthesized by Genscript into pGEX6P1 and pcDNA3.1_N-eGFP using BamHI/SalI and BamHI/NotI, respectively. Constructs were either full-length or missing the mitochondrial targeting sequence (ΔMTS, residues 13–129). The pGEX6P1 constructs contained a C-terminal 6x His tag. Genscript synthesized an ISCA1 construct in pcDNA3.1 with an internal myc tag between K70 and G71.

Yeast two-hybrid assays

Direct Y2H experiments were performed as previously described by our collaborator Sally Gray at the Cambridge Institute for Medical Research (Harbour *et al*., 2010; Kent *et al*., 2012; Pryor *et al*., 2008). Several direct Y2H systems were used to test the interactions between tepsin and potential binding partners; the vectors and the placement of the activation and DNA binding domains are listed in Table 4-2. Auxotrophic selection using media lacking histidine and adenine ensured only conditions with a positive interaction would grow.

Table 4-2: Table describing the Y2H systems used to test tENTH and tVHS interactions. The vectors and tags they have are listed. Table courtesy of Sally Gray.

System	Matt Seaman	Kate Bowers	Mate & Plate	Hybrigenics
Vectors	pGAD-C pGBT9	pGAD-C pGBDU-C	pGADT7 pGBKT7	pP6 pB29
Tags	All have N-terminal tags: pGAD N-term Gal4 activation domain pGB N-term Gal4 DNA-binding domain			N-terminal + C-terminal pP6 N-term Gal4 activation domain pB29 C-term LexA

Protein purification

Constructs were expressed in BL21(DE3)pLysS cells (Invitrogen) for 16–20 hours at 22˚C following induction with 0.4 mM IPTG at OD_{600}=0.8. For TRIM constructs, 10 μM zinc chloride was added at induction. TRIM constructs were purified in 20 mM Tris pH 8.7, 200 mM NaCl, 1 mM DTT. 10 μM zinc chloride was added at the beginning of the purification. ISCA1 constructs were purified in 20 mM Hepes pH 8.0, 200 mM NaCl, 2 mM βME. AEBSF protease inhibitor (Calbiochem) was used in the early stages of purification. Cells were lysed by a disruptor (Constant Systems Limited), and proteins were affinity purified using glutathione sepharose (GE Healthcare). For proteins that were cleaved, recombinant GST-3C protease was added overnight at 4˚C. Proteins were eluted in batch (by purification buffer if cleaved; 30 mM reduced glutathione if uncleaved) and further purified by gel filtration on a Superdex S200 preparative or analytical column (GE Healthcare).

VPS35L constructs were expressed using the TNT SP6 High-Yield Wheat Germ Protein expression system (Promega), according to the manufacturer's protocol. Briefly, up to 4 μg of DNA was added to the provided master mix and incubated at 25˚C for 2 hours. For co-expression of VPS35L and VPS29, 2 μg of DNA of each construct was added to one reaction. For the equivalent single construct controls, only 2 μg of DNA was added. The wheat germ reaction was directly used in pulldowns.

Pull down assays

For GST or His tag pulldowns, 50 µg of bait with the appropriate tag was used along with 5x molar excess of the prey constructs. Bait and prey proteins were incubated with glutathione sepharose or cobalt resin for 1 hour at 4˚C. The unbound fraction was removed, and samples were washed three times. The wash buffer was 20 mM Tris pH 8.5, 200 mM NaCl, 0.5 mM TCEP, 0.5% NP40 or 20 mM Tris pH 8.7, 200 mM NaCl, 0.5 mM TCEP, 1% NP40. Proteins were eluted from the resin using the wash buffer plus 30 mM reduced glutathione or 400 mM imidazole following a 10-minute incubation on ice. Gel samples were prepared from the supernatant, and analyzed by PageBlue (Thermo Scientific) staining of SDS-PAGE gels and Western blotting.

Cell culture

HeLa cells were cultured under standard conditions. Cells were plated on 6 well plates the night before a transfection (400,000 cells/ well) in MEMα media (GIBCO). During transfection, this media was aspirated and replaced with Opti-MEM Reduced Serum for the 4-hour incubation period. The manufacturer's protocol for Lipofectamine 2000 transfection was used to transfect tepsin, VPS35L, and ISCA1 constructs with 2.5 µg of DNA used per transfection. After four hours, the Opti-MEM media was aspirated and replaced with full serum media for overnight incubation. After 21–24 hours, cells were imaged to gauge transfection efficiency. FuGENE (Promega) was also used for VPS35L transfection, following the manufacturer's protocol, but no significant improvement in transfection was observed.

Cells were harvested and lysed, and lysates were normalized for use in immunoprecipitations. Briefly, cells were washed twice with non-sterile PBS then scraped and collected using a small volume of lysis buffer (15mM HEPES, pH 7.5, 150mM NaCl, 0.5% EDTA, 1% NP40, 1 Roche EDTA free complete protease tablet). The cell slurry was lysed by syringe (21g and 23g needles). Lysed cells were placed on ice for a 30-minute incubation prior to

centrifugation to remove cell debris. Total protein concentration was calculated using Precision Red (Cytoskeleon, Inc).

Immunoprecipitations

Lysates (1000 µg total protein) were incubated with GFP-Trap-A resin (Chromtek) for 1.5 hours at 4˚C. Following incubation, the unbound supernatant was removed and the resin washed three times with wash buffer (20mM HEPES, pH 7.5, 150mM NaCl, 2mM DTT, 1% NP40). Protein was eluted from the resin by boiling in SDS loading buffer.

Western blotting

SDS-PAGE gels were transferred to PVDF under high-voltage conditions. Membranes were blocked in 5% milk for 1 hour, followed by overnight primary antibody incubation at 4˚C. If the primary antibody was HRP-conjugated, membranes were washed with Tris-buffed saline containing 0.1% Tween 20 (TBS-T) and prepared for detection by electrochemiluminescence (ECL) using a kit (Fisher). If a secondary antibody was necessary, the membranes were washed and placed in secondary for 1.5 hours at room temperature prior to ECL development.

The following antibodies were used: anti-β actin (Santa Cruz; sc-47778), anti-c16orf62/VPS35L (abcam; ab97889), anti-DSCR3 (Millipore; ABN87), anti-GFP (abcam, ab6663), anti-GST (abcam, ab3416), anti-6X His tag (abcam, ab184607), anti-ISCA1 (ThermoFisher, PA5-60121), anti-moesin (abcam, ab52490), anti-myc (Millipore, 16-213), anti-rabbit (Sigma, A9169), anti-TRIM23 (abcam, ab192032), anti-VPS29 (abcam, ab98929), anti-VPS35 (abcam, ab157220).

Results

We employed a parallel approach using direct Y2H assays, pulldown assays with purified proteins, and IPs from HeLa cell lysates to test the ability of our candidate binding partners to bind individual tepsin domains and full-length tepsin.

TRIM2 and TRIM23

We saw no interaction between tENTH and TRIM2 or TRIM23 using a direct Y2H assay (Figure 4-1). Both TRIM2 and TRIM23 expressed and purified poorly from *E. coli,* and neither GST-tagged protein pulled down tENTH above background (data not shown). Because of the negative direct Y2H and pulldown data, we did not attempt to IP TRIM2 or TRIM23 from HeLa cells transfected with tepsin-GFP. While the original Hybrigenics hit appeared strong, and these are proteins involved in trafficking, it does not appear that the interaction between tepsin and TRIM2 and TRIM23 exists in the conditions we tested.

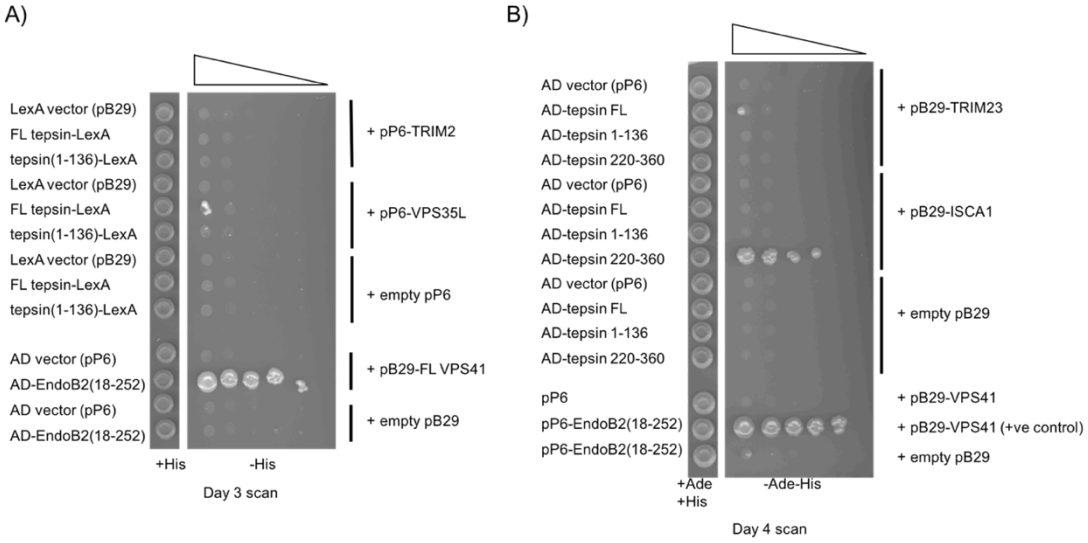


Figure 4-1: Direct Y2H assaying interactions between tepsin and our top four candidates: TRIM2, VPS35L, TRIM23, and ISCA1. A) No interaction is seen between tepsin and TRIM2. A very minor interaction is seen in the full-length tepsin and VPS35L condition. B) A weak interaction is seen in the full-length tepsin and TRIM23 condition. A strong interaction is seen between tVHS and ISCA1. The positive control was endophilin B2 and VPS41, a known Y2H interaction. Figure courtesy of Sally Gray.

VPS35L

Direct Y2H using the Hybrigenics system recapitulated the tENTH-VPS35L (438–688) interaction seen in the commercial screen (Figure 4-2A, Sally Gray, CIMR). We did not see a convincing interaction between the full-length versions of tepsin and VPS35L nor an interaction between full-length tepsin and VPS35L (438–688). Only minor growth was observed after day 3 between the full-length versions (Figure 4-1A); it took fifteen days to observe a minor colony population in the tepsin-VPS35L (438–688) condition (Figure 4-2B). We only observed this interaction with the same Hybrigenics system used in the commercial screen. The other Y2H systems tested did not show an interaction or had protein autoactivation (data not shown).

Pulldowns with purified VPS35L and tepsin constructs revealed variability depending on the experimental set up. Since VPS35L did not express and purify well in *E. coli*, we used an *in vitro* wheat germ translation system to express VPS35L

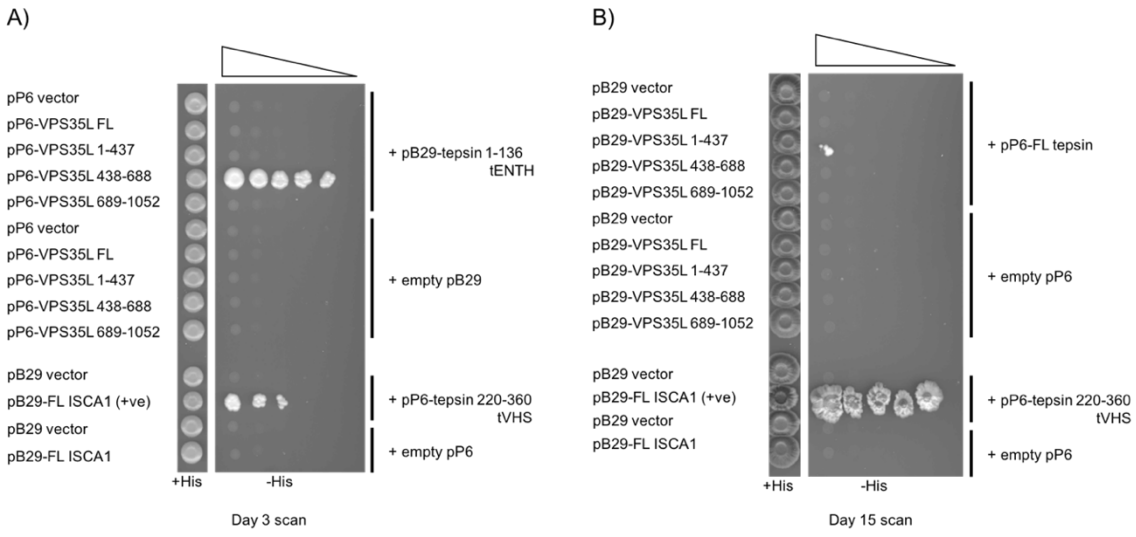


Figure 4-2: Direct Y2H analysis of tepsin and VPS35L using the Hybrigenics system. A) Growth on plates lacking histidine reveals an interaction between tENTH and VPS35L(438-688). A Day 3 experimental scan is shown. B) No significant growth is seen after 15 days between full length tepsin and VPS35L. The tVHS-ISCA1 interaction is used as a positive control. The B29 vector has a C-terminal LexA, the P6 vector has an N-terminal Gal4 AD. Adapted from a figure courtesy of Sally Gray.

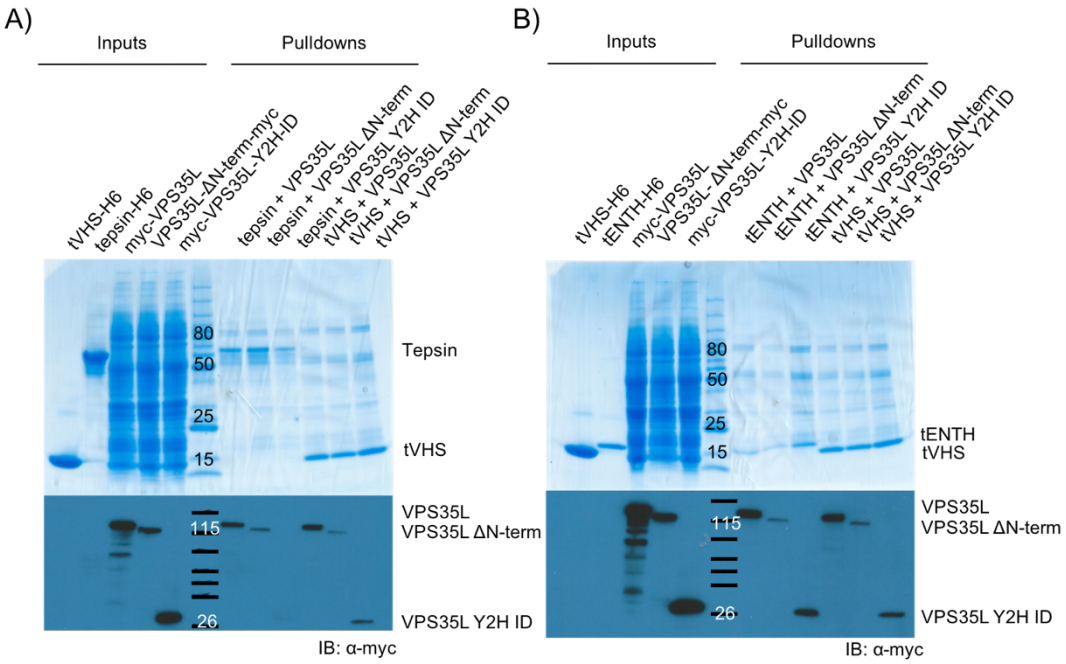


Figure 4-3: Tepsin and VPS35L interact in His-pulldown assay. SDS-PAGE gels and α-myc Western blots are shown. A) Pulldown of full-length VPS35L, VPS35L ΔN-term (residues 87-1050), VPS35L Y2H ID (residues 438-488) by tepsin compared to tVHS. B) Pulldown of full-length VPS35L, VPS35L ΔN-term, VPS35L Y2H ID by tENTH compared to tVHS.

constructs with myc tags. When we used GST-tagged tepsin constructs and tried to pull down VPS35L, we saw no interaction (Figure A3-1).

When we used His-tagged tepsin constructs instead, we pulled down VPS35L constructs in every condition, including our negative controls (Figure 4-3, Figure 4-4). 1% NP40 detergent in the wash buffer did not lead to a decrease in binding. Initially, we used tVHS as the negative control for the VPS35L interaction with tENTH and full-length tepsin (Figure 4-3). Because VPS35L constructs bound to tVHS under the experimental conditions, we used the first β-propeller from β'COP, one of the outer-coat like subunits of COPI, as the negative control (Figure 4-4). Densitometry analysis of the Western blots indicates VPS35L interacts with tENTH above the negative control background (Table 4-3).

The GST- and His-tagged pulldowns were conducted under similar conditions (pH 8.5, pH 8.7) with the tags always placed at the C-terminus of tepsin constructs. Since the tags immobilize the protein on the resin, they should not interfere with binding. The size of the tags is the greatest difference between the two experimental set ups and perhaps contributes to the variability seen.

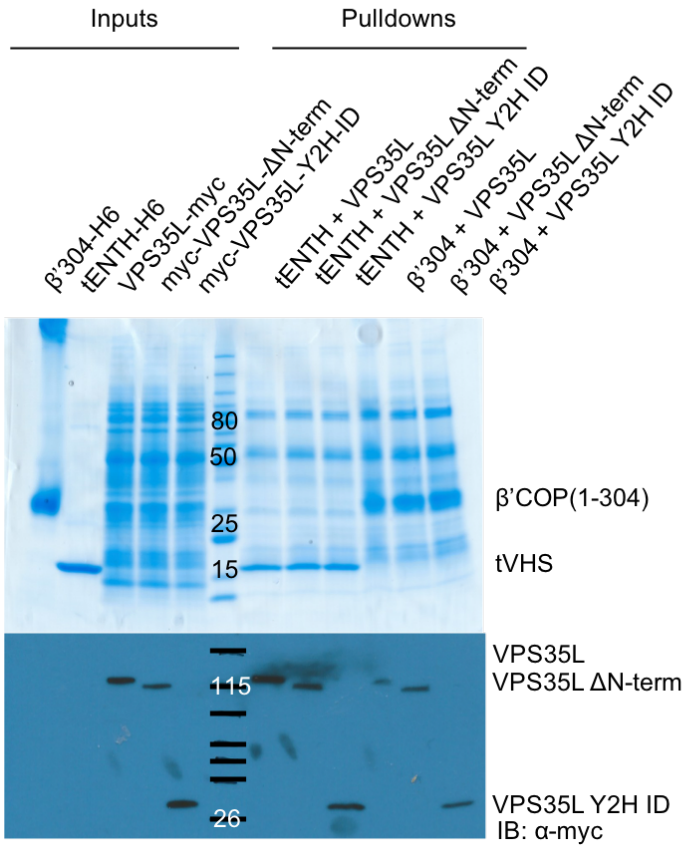


Figure 4-4: tENTH and VPS35L interact in a His-pulldown assay. Pulldown of full-length VPS35L, VPS35L ΔN-term, VPS35L Y2H ID by tENTH compared to β'COP(residues 1-304; negative control). SDS-PAGE gel (top) and Western blot for α-myc (bottom).

Table 4-3: Densitometry analysis of Western blots from Figures 4-3 and 4-4. A) tENTH pulls down VPS35L and VPS35L Y2H ID 1.6 times more than tVHS, but does not pull down the ΔN-term construct more than tVHS. B) Tepsin does not interact with VPS35L above tVHS background signal. C) tENTH interacts with VPS35L more strongly than β'COP. ImageJ was used for densitometry analysis.

(A) (n=1)	Normalized to tVHS	**(B)** (n=1)	Normalized to tVHS	**(C)** (n=1)	Normalized to β'COP
tENTH + VPS35L	1.6	tepsin + VPS35L	0.5	tENTH + VPS35L	24.8
tENTH + VPS35L ΔN-term	0.6	tepsin + VPS35L ΔN-term	1.0	tENTH + VPS35L ΔN-term	8.2
tENTH + VPS35L Y2H ID	1.6	tepsin + VPS35L Y2H ID	0.1	tENTH + VPS35L Y2H ID	6.0
tVHS + VPS35L	1.0	tVHS + VPS35L	1.0	B'304 + VPS35L	1.0
tVHS + VPS35L ΔN-term	1.0	tVHS + VPS35L ΔN-term	1.0	B'304 + VPS35L ΔN-term	1.0
tVHS + VPS35L Y2H ID	1.0	tVHS + VPS35L Y2H ID	1.0	B'304 + VPS35L Y2H ID	1.0

Immunoprecipitations testing the tepsin-VPS35L interaction also revealed a dependency on experimental approach. When we transfected GFP-VPS35L and performed an IP from HeLa cell lysate, endogenous tepsin did not co-IP with VPS35L (Figure 4-5A). Our positive control,

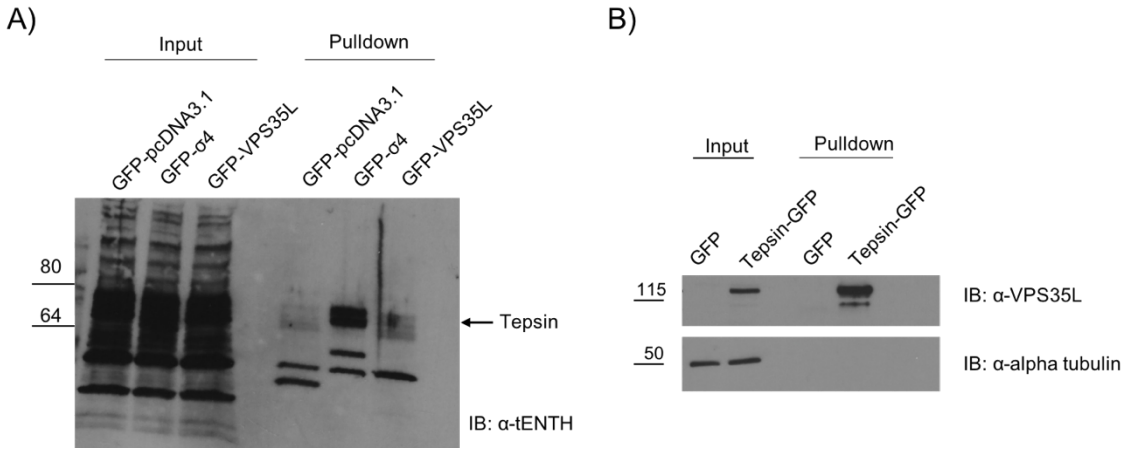


Figure 4-5: IPs from Hela cells testing the tENTH-VPS35L interaction vary depending on which construct is transfected. A) Endogenous tepsin does not immunoprecipitate with VPS35L significantly, compared to our positive control AP4, using an GFP-tagged σ4. n=4. B) Endogenous VPS35L does immunoprecipitate with tepsin-GFP. n=4. Adapted from a figure provided by Allie Isabelli.

GFP-σ4 showed a strong interaction between AP4 and endogenous tepsin (Figure 4-5A). When we transfected tepsin-GFP and performed an IP with HeLa lysate, endogenous VPS35L did co-IP strongly with tepsin (Figure 4-3B). We consistently saw these results across four biological replicates. This may be an artifact of GFP tag placement, and further work is needed to clarify these results.

Moesin and ezrin

The tVHS candidate binding partners ezrin and moesin have not been validated to the same extent as our other hits. We have only tested the interaction between tepsin and moesin by IP from HeLa cell lysates. Moesin did not co-IP with tepsin-GFP (Figure A3-2). Future work will employ the same experiments described above for these proteins.

ISCA1

Direct Y2H analysis confirmed the interaction between tVHS and ISCA1 seen in the Hybrigenics commercial screen (Figure 4-1). We saw no interaction between full-length tepsin and ISCA1, in contrast to the proteomics work that used either full-length tepsin or ISCA1 as bait

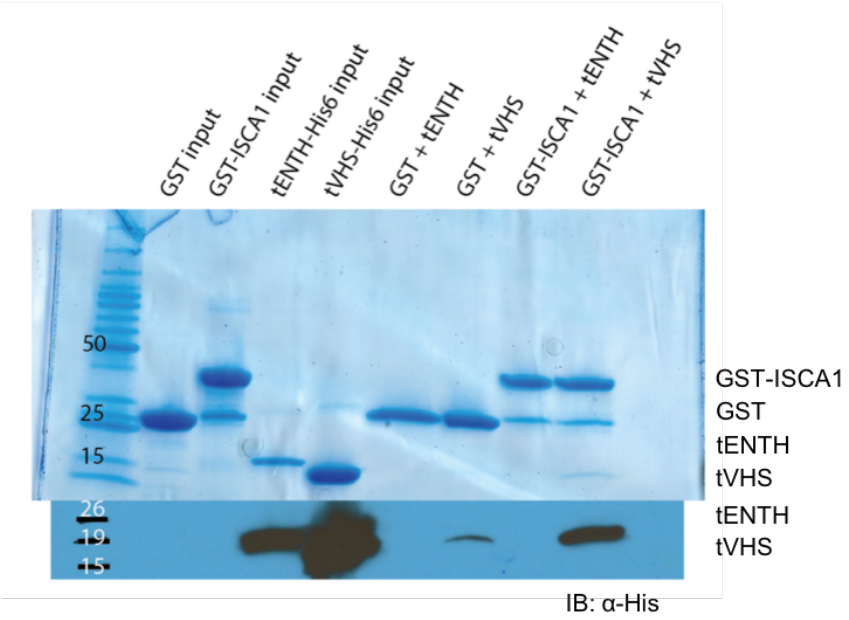


Figure 4-6: GST-ISCA1 specifically pulls down tVHS compared to tENTH. Above is the PAGEBlue stained SDS-PAGE gel, below is the Western blot using anti-His.

for affinity capture mass spectrometry (Beilschmidt *et al.,* 2017; Hein *et al.,* 2015, Huttlin *et al.,* unpublished).

GST-pulldowns using GST-ISCA1 show an interaction between tVHS and ISCA1 (Figure 4-6). Our tepsin-H6 construct cannot be distinguished well when probed with our anti-His antibody, but may interact with ISCA1 above the GST background (data not shown). While we consistently see the interaction between tVHS and ISCA1 when GST-ISCA1 is the bait, we do not see the interaction when tepsin constructs are GST tagged and H6-ISCA1 serves as prey (Figure 4-7A).

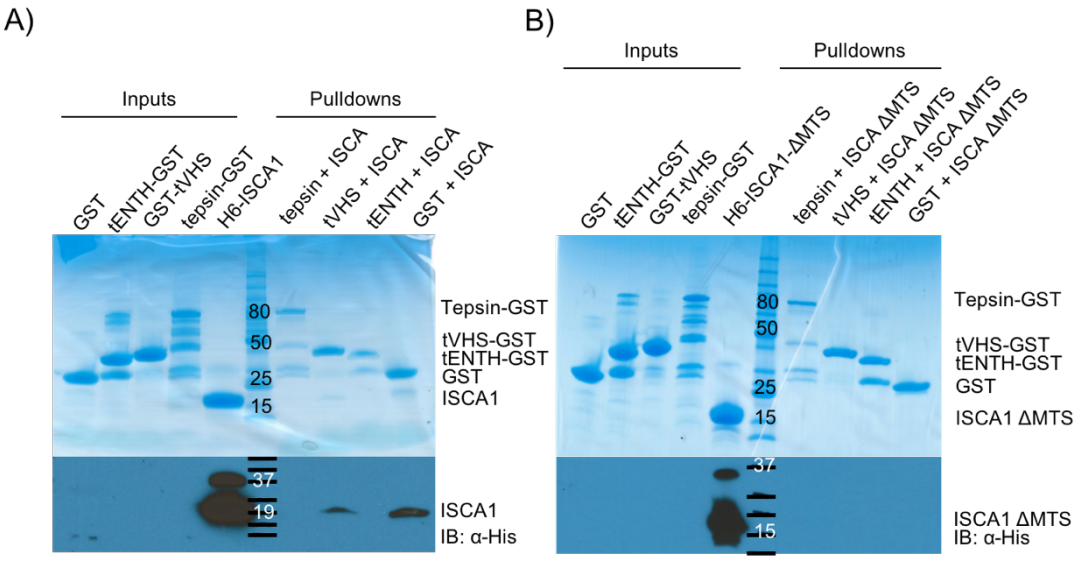


Figure 4-7: No interaction seen between tepsin and ISCA1 ΔMTS by GST-pulldown assay. SDS-PAGE gels and Western blots (α-His) are shown. A) GST-tagged tVHS pulls down H6-ISCA1 weakly, below the background of GST. B) ISCA1 ΔMTS is not pulled down by GST-tagged tepsin, tVHS or tENTH constructs.

These results could be an artifact of the tag placement, however ISCA1 uses its C-terminus for dimerization, and C-terminally tagged ISCA1 does not purify well. H6-ISCA1 was expressed as an N-terminal GST fusion. The weak interaction with H6-ISCA1 may be because purified ISCA1 does not behave well without a GST tag. Until cleavage the protein purifies well, but upon cleavage and further purification it loses its Fe-S cluster. A visible loss of color to the protein in solution indicates this loss, and spectroscopic measurements at 350 and 415 nm

wavelengths support this conclusion (Table A3-1). Absorbance at these wavelengths is characteristic of Fe-S clusters (Gryzb *et al.,* 2012). These data suggest ISCA1 loses its Fe-S cluster throughout the aerobic purification; following cleavage, very few copies of the protein contain an intact cluster. The fact that tVHS interacts with GST-ISCA1, but not His-ISCA1, implies that this interaction depends in part on the state of the Fe-S cluster in ISCA1. If the interaction is sensitive to Fe-S cluster status, the purification may need to proceed with higher concentrations of reducing agent to protect against oxidation and degradation of the intact cluster.

If the tVHS-ISCA1 interaction is biologically relevant, it would likely have to occur in the cytosol. We wanted to test if tVHS had a preference for binding an ISCA1 construct without the mitochondrial targeting sequence (ISCA1-ΔMTS). Since we were not concerned about ISCA1

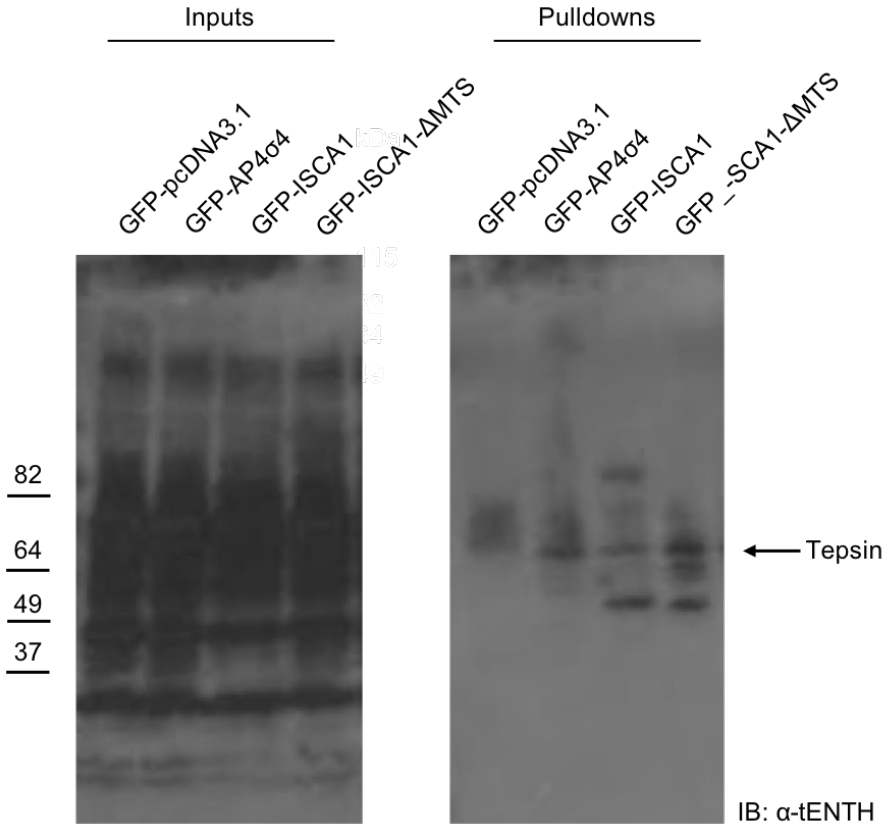


Figure 4-8: IPs from HeLa cells suggest an interaction between tepsin and ISCA1, with a slight preference for ISCA1 ΔMTS. An antibody against tENTH (Poconos rabbit farm) was used to probe endogenous tepsin. The specific tepsin band is denoted by the arrow. n=1. Adapted from a figure provided by Allie Isabelli.

getting to the mitochondria, we used an N-terminal GFP tag. The resulting IPs from this experiment showed that ISCA1 can co-IP endogenous tepsin (Figure 4-8). We observed a slight preference for the ISCA1-ΔMTS construct, but cannot draw conclusions from one experiment without replicates. GST-pulldown of ISCA1-ΔMTS by tVHS or tepsin did not occur, but this could be because of the poorly behaved nature of ISCA1 without a GST tag (Figure 4-7B; see above). An ISCA1 construct with a myc tag placed in an unstructured loop will avoid the pitfalls of disrupting mitochondrial localization (N-terminal tag placement) and disrupting oligomerization (C-terminal tag placement) in cells.

Discussion

To test the hits from the commercial Hybrigenics Y2H screen, we used several complementary approaches: direct Y2H analysis, pulldown assays with purified proteins, and IPs from HeLa cell lysates.

The direct Y2H assays tested the interactions in three Y2H systems (Table 4-2). Interactions were only observed in the Hybrigenics system, the same system used in the commercial screen (Figure 4-1, 4-2). This is not surprising in the case of full-length tepsin and the tENTH domain, because C-terminally tagged tepsin is more stable in cells than N-terminally tagged tepsin (personal communication, Georg Borner). As the tENTH domain is at the very N-terminus of tepsin, we believe its N-terminus needs to be unhindered by tags for proper function. The other Y2H systems have N-terminal tag placement for both the activation and DNA binding domains; among the systems tested, only the Hybrigenics system has a C-terminal tag option.

We also believe full length tepsin may have an autoinhibited state, although we have no direct evidence of this yet. The lack of direct Y2H interactions between our candidate binding partners and full-length tepsin further supports this idea. Tepsin could be inhibited by the C-terminus binding back to the N-terminal domains. In this case, adding β4 to the pulldown reactions of tepsin and candidate binding partner would shift tepsin to the uninhibited conformation by

binding to the C-terminus of tepsin. Binding under those conditions would provide more direct evidence of an autoinhibited state for tepsin.

The direct Y2H assay and pulldowns with purified proteins did not confirm an interaction between tepsin and TRIM2 or TRIM23. Recent work on TRIM23 suggests it has an important role in autophagy in response to viral infection (Sparrer *et al.,* 2017). It may be worthwhile to try to co-IP tepsin and TRIM23 from untreated HeLa cells and starved HeLa cells to see if the interaction becomes more robust under starved conditions, when autophagy is induced.

Recent work characterized Atg9a, a transmembrane protein essential for induction of autophagy, as a cargo of AP4 (Mattera *et al.,* 2017; Davies *et al.,* 2017; Ivankovic *et al.,* 2017). Thus, it may be important to look at all the tepsin candidate binding partners, not just the TRIMs, in the context of basal and starved (autophagy-inducing) conditions. Additionally, we have identified a putative LIR (LC3 interacting region) motif in tepsin, perhaps another link between AP4 vesicles and autophagy (iLIR database, Jacomin *et al.,* 2016). LIR motifs target autophagy receptors and adaptors to lipidated Atg8 family members (including LC3) anchored in the forming phagophore (Lamarck *et al.,* 2009).

IPs and pulldowns showed binding between tENTH/tepsin and VPS35L only under certain conditions. Purified proteins reveal an interaction only between His-tagged tepsin constructs and VPS35L constructs (Figure 4-3, Figure 4-4). Tepsin-GFP can co-IP endogenous VPS35L, but we did not observe the opposite (Figure 4-5). The relevance of these inconsistent results bears further validation by co-localization in cells.

Our preliminary work uncovered a potential role for the N-terminus of VPS35L, which was not observed in the original Hybrigenics screen. The direct Y2H analysis showed a minor interaction between tepsin and the N-terminus of VPS35L (Figure 4-2B), and pulldown of the VPS35L ΔN-term construct was weaker than full-length VPS35L (Figure 4-4, Table 4-3).

The interaction between VPS35L and tepsin also raises the question of whether this interaction occurs independently of AP4 trafficking. Does tepsin function in another trafficking

pathway? The planned co-localization studies can address the role of AP4 in the tepsin-VPS35L interaction. To test whether tepsin or AP4 plays a role in VPS35L function, we can look for changes in VPS35L steady state localization in tepsin and AP4 KO cell lines. However, since VPS35L is cytosolic, we do not expect to see changes in VPS35L steady state localization.

Similar to the inconsistent results seen with VPS35L, the ISCA1 Y2H interaction can be recapitulated by direct Y2H screen, but no Y2H interaction is seen between full-length tepsin and ISCA1 (Figure 4-1B). Additionally, while GST-ISCA can pull down tVHS, the reverse pulldown does not occur (Figure 4-6, Figure A3-1). Co-localization studies in HeLa cells are therefore also necessary to validate this interaction. Our collaborators are using direct Y2H to map the ISCA1-tVHS interaction, starting with ISCA1 ΔMTS (residues 13–129) and tVHS ΔN-term (residues 235–356). We also identified two conserved surface patches to mutate in tVHS: a large acidic patch (D242S/C243A/E279S/Q283S), and a small basic patch (K340S/K343S). We believe these experiments will narrow down the interacting regions on ISCA1 and tVHS.

One potential explanation for the ISCA1-tVHS interaction is that it functions as an iron sensor in the cytosol. If ISCA1 levels rise in the cytosol due to disrupted mitochondria, tepsin could bind ISCA1, perhaps resulting in LIR-mediated trafficking for degradation by autophagy. If this were the case, ISCA1 would have undergone the cleavage of its MTS upon import into the mitochondria. We tested whether tepsin had a preference for ISCA lacking the MTS and found a slightly stronger interaction (Figure 4-8), although this needs further replication. The report of cytosolic ISCA1 (Song *et al.,* 2009) is likely due to overexpression (Sheftel *et al.,* 2012); however, our own co-localization studies would also provide insight into that discussion as well.

Our preliminary examination of hits from tENTH and tVHS Y2H screens suggest that VPS35L and ISCA1 may be binding partners for these respective domains. More work is needed to uncover the biological relevance of these interactions. Further validation is needed to determine whether tENTH binding partner candidates TRIM2 and TRIM23, or tVHS binding partner candidates ezrin and moesin, lead to meaningful interactions.

Acknowledgements

Thanks to Sally Gray and Paul Luzio for the direct Y2H assays and advice on the project. Tara Archuleta performed the SILAC experiments with tENTH and full-length tepsin, as well as the initial biochemical experiments with TRIM2, TRIM23, and VPS35L. Allie Isabelli performed the cell culture work and immunoprecipitation experiments.

CHAPTER V

DISCUSSION AND FUTURE DIRECTIONS

Discussion

Interactions between APs and epsins

In addition to the tepsin-AP4 interaction, epsin family members function as accessory proteins in AP1 and AP2 vesicle formation. EpsinR interacts with AP1 by a DFxD[FW] motif in its C-terminus, which binds the γ appendage domain with highest affinity (Mills *et al.,* 2003). EpsinR binds the N-terminal sandwich subdomain of the γ appendage (Mills *et al.,* 2003) since it lacks the C-terminal platform subdomain (Kent *et al.,* 2002; Nogi *et al.,* 2002), which AP2 and AP4 use—albeit in the β subunit—to interact with epsin1 (Owen *et al.,* 1999; Traub *et al.,* 1999; Brett *et al.,* 2002; Edeling *et al.,* 2006a) and tepsin (Chapter II) respectively (Figure 5-1).

Epsin1 binds AP2 α and β2 appendage domains using different motifs: DPF/W motifs interact with the α appendage domain, while a $[DE]_nX_{1-2}Fxx[FL]xxR$ motif binds the β2 appendage domain (Figure 5-1). The $[DE]_nX_{1-2}Fxx[FL]xxR$ motif is unstructured, except in the presence of β2

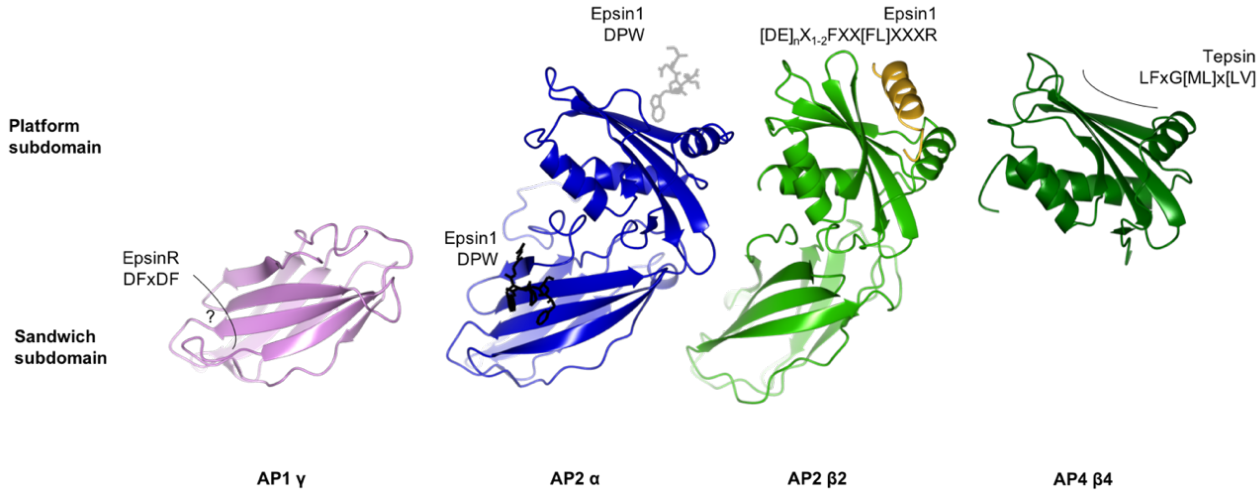


Figure 5-1: Comparison of AP and epsin interactions. Structures of AP1 γ appendage (light purple; 1GYU; Kent *et al.,* 2002), AP2 α appendage in complex with epsin1 DPW motifs (blue; 1KY6, 1KYD; Brett *et al.,* 2002), AP2 β2 appendage in complex with epsin1 helical motif (light green; 2G30; Edeling *et al.,* 2006) and AP4 β4 (dark green; 2MJ7) are shown.

when it forms a helix to fit in the β2 binding pocket (Edeling *et al.,* 2006a). The binding surface used by β2 is conserved in β4 (Chapter II); with epsin1 and β2, Edeling *et al.* named the β2 surface pockets after key ligand residues—the F, [FL], and R pockets. The β4 binding motif I discovered (LFxG[ML]x[LV]) uses conserved residues from the F and [FL] pockets of β2. As identified by NMR chemical shift perturbation and confirmed by mutagenesis, I669, A670, and Y682 in β4 mediate the tepsin interaction. Y682 is equivalent to Y888 in β2, which contributes to both the F and [FL] pockets. I669 and A670 are equivalent to β2 I876 and A877 in the F pocket. Although these important residues are identical between β4 and β2, the surrounding surfaces are conserved but not identical. This leads to differences in shape that create the specificity of interactions, as shown by the lack of β2 binding tepsin (Chapter II).

The ε binding motif in tepsin and its corresponding surface on ε have not been structurally or biochemically characterized by mutagenesis, so we do not know exactly where tepsin binds the ε appendage; nor do we have a structure of the ε appendage domain. It will be interesting to see if the ε binding motif in tepsin interacts with ε via conserved patches—whether that is where epsin1 binds the α appendage domain, or where epsinR binds the γ appendage domain. The distance between the two motifs in tepsin is only 38 amino acids, presumably entirely unstructured. Estimating 2.5 Å per amino acid, that would give a distance of 95 Å between motifs. The surface of β4 tepsin binds across measures approximately 30 Å; depending on how the ε binding motif engages its appendage domain, it may be possible that both motifs could engage β4 and ε of different AP4 complexes at the same time. A GST-pulldown assay could test whether tepsin can pull down both β4 and ε, but only careful analysis of the stoichiometry could provide insight into whether both motifs can engage the appendages at the same time. Structural studies would provide the best answer to the question of whether tepsin can crosslink different AP4 complexes.

Work in HeLa cells by our collaborators in the Robinson group (Chapter II), as well as work from the Bonifacino group (Mattera *et al.,* 2015), raise the possibility of a weaker third

interaction between AP4 and tepsin. Tepsin and AP4 still showed punctate co-localization even when tepsin lacking β4 and ε motifs, or β4 with mutations that ablated the interaction, was transfected in HeLa cells. However, IPs from these mutant backgrounds pulled down only residual AP4 or tepsin, respectively, suggesting the β4 binding motif is the main driver of the AP4-tepsin interaction. Evolutionary analysis of the ε binding motif from Mattera *et al.* supports this conclusion as well. The full motif they defined (S[AV]F[SA]FLN) is only found in vertebrates. They determined a degenerate motif is conserved more broadly (FxF[LIMV]), but it seems unlikely that a single copy of a smaller motif could provide the same affinity as the longer β4 motif. Although the ε binding motif is unlikely to be as important as the β4 binding motif, the avidity effect of multiple weak interactions is an important contribution of the ε binding motif.

Mattera *et al.* posits there is an interaction between the tVHS domain and AP4 based on evidence from pulldowns with a series of deletion constructs. However, they do not conclusively show that the interaction relies on the tVHS domain, rather than the linker between the tENTH and tVHS domain. I used a direct approach to test binding between the tVHS domain and AP4, as opposed to the deletion construct approach used by Mattera *et al.,* but I failed to detect an interaction (data not shown). However, I could not definitively say such an interaction with the tVHS domain does not exist; since we are currently incapable of purifying the AP4 complex from *E. coli* or mammalian cells, I used an *in vitro* wheat germ translation system to produce small quantities of the full-length subunits as heterodimers (β4/μ4 and ε/σ4). Heterodimers have proven to be stable for biochemical studies in other AP complexes, but only the β4/μ4 homodimer behaved well in pulldown experiments. I was able to biochemically confirm an interaction between β4/μ4 and Arf1(GTP), which recruits AP4 to the membrane (Figure A4-1). ε/σ4 appeared indiscriminately “sticky” in pulldowns, which suggests either a problem with folding in the wheat germ system, or exposure of hydrophobic patches due to lack of the full complex (data not shown).

AP4 and tepsin likely coevolved to support new functions

Tepsin appears to be much more functionally intertwined with AP4 than other epsins and their AP interacting partners. Tepsin relies on AP4 for recruitment to the membrane, while other epsins are recruited independently, and there appears to be a concomitant loss of AP4 and tepsin in several eukaryotic lineages (Borner *et al.,* 2012). Tepsin also has a unique structural organization compared to other epsins: a second domain downstream of the ENTH domain, no UIMs, and no clathrin binding motifs.

The domain organization of tepsin, with both ENTH and VHS domains, remains intriguing. Why does tepsin need two domains that share a common fold and general function in trafficking? Our phylogenetic analyses (Chapter III) does not support the gene duplication hypothesis; on a sequence level, the tVHS domain appears to be a VHS domain and not a duplicated tENTH domain. The literature reveals no other internal VHS domains. With over 60 proteins that have a VHS domain at the very N-terminus, the field has assumed that positioning plays a large role in the function of the domain.

The structural work described in Chapter III identified key differences in the tENTH and tVHS domain compared to other ENTH and VHS domains: both lack critical functional helices (helix 0 and helix 8 in tENTH; helix 8 in tVHS), which explains why neither domain performs the common ENTH and VHS functions. Instead, it appears tepsin evolved to support novel functions to assist in AP4 trafficking.

The crystal structure of the tENTH domain revealed no amphipathic helix 0 or basic binding patch exists for interacting with phosphoinositides, which explains why tepsin is not recruited to the membrane independently. This suggests independent membrane binding is not required for the function of tepsin. Ford *et al.* proposed epsins might drive membrane curvature through insertion of the ENTH domain amphipathic helix 0 (Ford *et al.,* 2002). More recent work suggests that protein crowding on a membrane is enough to drive curvature (Stachowiak *et al.,* 2012). Regardless of the exact mechanism, membrane curvature does not seem a likely function

for tepsin if it cannot independently interact with the membrane. This led to our initial hypothesis that tepsin functions mainly as a cargo adaptor just as epsin1 uses UIM motifs downstream of the ENTH domain to bind ubiquitinated cargoes and endocytose them. This served as our motivation to find novel binding partners for the tENTH and tVHS domain (Chapter IV).

The possible disordered N-terminal helix in the tVHS domain also generates further questions. The first 12 amino acids in our crystallization construct were disordered in the electron density maps, and one bioinformatics server annotates the predicted helix in that region as disordered (PSIPRED, Buchan *et al.,* 2013). If the disordered helix exists and we can find evidence of it upon interaction with a protein binding partner, does that resurrect the argument that this domain might have significant characteristics of an ENTH domain? Or does it simply highlight the fact that at some point fine classification of protein domains is not possible and, frankly, not particularly useful?

To quickly probe the existence of the disordered helix, I proposed using circular dichroism to compare thermal melting curves of the tVHS crystallization construct and a construct lacking the proposed helix. I would expect an increase in secondary structure to correspond with a greater melting temperature due to increased molecular interactions. The circular dichroism experiments, carried out by Anderson Monken (unpublished data, see Figure A4-1), showed no difference in helical content as assessed by melting curves. However, without a binding partner to stabilize this helix, I would expect the helix to be sampled in conformational space only occasionally. Thus, it is not a surprise that there were no significant differences in melting temperature. This negative data does not disprove the existence of the disordered helix, it only fails to confirm it.

The CD experiment, while far from perfect, was the best option to quickly probe the existence of the disordered helix without requiring labeled protein. The human sequence contains a high proportion of charged residues (5/12), while the horse sequence has slightly fewer (3/12). Neither sequence appears to fall in a register that would lead to an amphipathic helix, with a

charged side and hydrophobic side, reminiscent of helix 0 in ENTH domains (Figure A4-2; Gautier *et al.,* 2008).

Likely the only definitive answer will come with finding a binding partner that utilizes the disordered helix for interacting with the tVHS domain. Once binding partners have been validated, one relatively straightforward way to assay for tVHS interaction surfaces would be a pulldown assay. This assay would test the interaction between binding partners and tVHS constructs with and without the N-terminal disordered helix. A difference in binding would suggest involvement of the N-terminus, and further structural work could be done to identify whether the interaction induced helicity in the tVHS N-terminus.

Tepsin as an autophagy effector

A β4 knockout mouse model (Matsuda *et al.,* 2008) revealed missorting of AMPA receptors (AMPARs) to axons, and their accumulation in autophagosomes (as identified by electron microscopy and LC3-positive labeling). LC3A/B/C is a member of the mAtg8 family of autophagy proteins that undergoes a change from a cytosolic form (LC3-I) to a lipidated form (LC3-II). The lipidation of LC3 and its binding to the forming phagophore is part of the maturation process of autophagosomes (Kabeya *et al.,* 2004). Matsuda *et al.* found an increase in LC3B-II as well as a decrease in p62 (an autophagy receptor degraded during autophagy; Bjørkøy *et al.,* 2005) in cerebellar lysates. While the authors' conclusions focused on the role of AP4 in sorting AMPARs to the somatodendritic domain via interactions between μ4 and TARPs (transmembrane AMPA receptor regulatory proteins), this study was the first to identify a potential link between AP4 and autophagy (Matsuda *et al.,* 2008).

Recent work by the Kittler, Robinson, and Bonifacino groups identified new AP4 cargoes and accessory proteins (Mattera *et al.,* 2017; Davies *et al.,* 2017; Ivankovic *et al.,* 2017). They all identified Atg9a, a transmembrane protein critically involved in the formation of the pre-autophagosome structure, as a cargo of AP4 vesicles. Atg9a is likely responsible for bringing membrane to the forming structure (reviewed in Pavel and Rubinsztein, 2017). Mattera *et al.*

identified Atg9a as an AP4 cargo using affinity mass spectrometry, and validated the hit by Y2H, IPs, and co-localization studies. They found that Atg9a interacts with μ4 through a noncanonical YXXΦE motif and that depletion of AP4 results in accumulation of Atg9a at the TGN (Mattera et al., 2017). Subsequently, a reduction in LC3B lipidation was observed. However, LC3B expression increased in AP4 ε knockout cells, and the size and number of LC3B-labeled structures increased, in agreement with data from the $\beta4^{-/-}$ mouse (Matsuda *et al.,* 2008). However, Mattera *et al.* found no changes in p62 levels. Combined with increased expression of Atg9a and LC3B in AP4 ε knockout cells, this data suggests that the mistrafficking of Atg9a and decrease in LC3B lipidation due to loss of AP4 can be overcome despite evidence of abnormal maturation (Mattera *et al.,* 2017).

Davies *et al.* identified Atg9a, RUSC2, and several other proteins as AP4 cargoes and accessory proteins through the creation of dynamic organellar maps (Itzhak *et al.,* 2016; Itzhak *et al.,* 2018). This is an unbiased mass spectrometry-based approach that identifies proteins whose subcellular localization shifts in cells with different genetic backgrounds (such as AP4 knockout cell lines). The Atg9a and RUSC2 hits were validated by proximity labeling mass spectrometry, co-IPs, and co-localization studies. They saw retention of Atg9a at the TGN when AP4 ε has been knocked out; but in a wild-type AP4 background, overexpression of RUSC2 drove Atg9a to punctate peripheral structures (Davies *et al.,* 2017). RUSC2 is believed to be involved in plus-end microtubule transport. Davies *et al.* also demonstrated an interaction (likely indirect) between RUSC2 and Atg9a that occurs in wild-type AP4 cells, but not in AP4 knockout cells. This suggests that RUSC2 is an accessory protein of AP4 necessary for peripheral sorting of Atg9a. In agreement with Matsuda *et al.* and Mattera *et al.,* Davies *et al.* observed increased LC3B, but did not see any impairment in lipidation (LC3B-II), as seen in Mattera *et al*. In fact, these authors saw increased LC3B-II in basal and starved AP4 ε and β4 knockout cell lines (Davies *et al.,* 2017). Davies *et al.* saw an increase in LC3-labeled puncta size, but not a difference in number of puncta.

Ivankovic *et al.* (2017) also identified Atg9a as an AP4 cargo during characterization of an AP4 ε knockout mouse. This $\varepsilon^{-/-}$ mouse showed corpus callosum thinning and ventricular enlargement that resembles phenotypes seen in patients with AP4-causative HSP. They attributed thinning of axonal tracts to neuronal defects in axonal branching and extension (Ivankovic *et al.,* 2017). They also noticed distal swelling in the axons due to aberrant autophagosome biogenesis and clearance from axons. As with the studies discussed above, the authors noted TGN retention of Atg9a. In neurons, they found that TGN retention due to AP4 loss specifically lead to a reduction of Atg9a vesicles in axons, but not in somatodendrites. Additionally, the retrograde movement of autophagosomes (labeled by LC3) back to the soma was reduced, leading to accumulation in axons.

These data taken together point to dysregulation of autophagy in the absence of a functional AP4 complex, due to the retention of Atg9a in the TGN. While flux through the pathway persists, marked differences in the size of the structures—and in the case of neurons, the retrograde trafficking—indicate defects in autophagy. None of the groups showed a direct role for tepsin in mediating this Atg9a sorting event; indeed, Mattera *et al.* showed loss of tepsin does not affect Atg9a sorting. However, we identified a putative LIR (LC3 interacting region) motif in the linker between the tENTH and tVHS domains when we input the tepsin sequence into the iLIR database tool (Jacomin *et al.,* 2016). None of the other epsin family members have any LIR motifs according to this database.

LIR motifs target autophagy receptors and adaptors to Atg8 family members (including LC3) anchored in the forming phagophore membrane. Atg8 family members are comprised of a ubiquitin-like C-terminal core plus two additional N-terminal helices (reviewed in Birgisdottir *et al.,* 2013). Like ubiquitin, Atg8 members are regulated by conjugating and deconjugating enzymes systems (Birgisdottir *et al.,* 2013). Other epsins contain UIMs to facilitate sorting of ubiquitinated proteins, a common modification in endocytic pathways. In that vein, the presence of a LIR motif in tepsin suggests it may support the function of AP4 as a vesicle carrier for autophagy.

Given the emerging pattern of AP4 involvement in sorting Atg9a, perhaps the tVHS-ISCA1 interaction fits into an autophagy framework. The biological relevance of this interaction has always been confusing, despite its common and robust appearance in binding partner screens, including Y2H and affinity capture mass spectrometry experiments (Beilschmidt *et al.,* 2017; Hein *et al.,* 2015, Huttlin *et al.,* unpublished). The major supported localization and function of ISCA1 is in the mitochondria. Why would it to interact with tepsin? When would it encounter tepsin? Mitochondrial sorting relies on an N-terminal sorting signal to target proteins to the organelle (reviewed in Backes and Hermann, 2017). Iron is carefully regulated within cells, as excess iron is cytotoxic (Braymer and Lill, 2017). The presence of cytosolic ISCA1 could indicate misregulation of iron or a problem with mitochondrial integrity, and could act as a signal for the upregulation of autophagy. While purely speculative, conceptual support for this idea comes from work done with aconitase/iron regulatory protein 1 (IRP1) as an iron sensor. Cytosolic IRP1 has two conformations that depend on the status of its iron sulfur cluster. IRP1 contains an iron sulfur cluster under high iron conditions and has aconitase activity in the cytosol. Under low iron conditions, IRP1 loses its iron sulfur cluster, changes conformation, and binds RNAs that inhibit an iron storage pathway, which leads to increased iron available for incorporation in iron sulfur clusters (Haile *et al.,* 1992; Basilion *et al.,* 1994; reviewed in Rouault and Maio, 2017).

Involvement of tepsin in other trafficking pathways

Although the evidence presented so far suggests AP4 and tepsin are strongly functionally linked, the tENTH interaction with VPS35L (Chapter IV) raises the possibility that it could function in trafficking pathways independently of AP4. VPS35L is a member of the newly discovered retriever complex that resembles retromer and mediates recycling of sorting nexin 17 (SNX17) cargoes from endosomes (McNally *et al.,* 2017). Through different targeting mechanisms, retriever and retromer both occupy the same endosomal subdomain and recycle distinct classes of proteins (McNally *et al.,* 2017). Retriever does not have any membrane binding domains, and interactions with the CCC (CCDC93, CCDC22, COMMD) and WASH (WASP and SCAR

homologue) complexes localize the complex to the endosome (McNally *et al.,* 2017). Normally tepsin has a puncate perinuclear localization, and it becomes cytosolically diffuse when AP4 is knocked out (Borner *et al.,* 2012; Chapter II). Currently, there is no evidence tepsin co-localizes with endosomes, but that question has not been specifically addressed. Future work in the lab will determine if VPS35L and tepsin co-localize as well.

Future Directions

My work has focused on understanding the molecular mechanisms of AP4 vesicle formation, specifically by characterizing the individual interacting motifs and domains in tepsin. Future work should aim to capture images of AP4 vesicles, to provide insight into what an "uncoated" vesicle looks like, and to provide information on the role of tepsin in vesicle formation. Super-resolution fluorescent microscopy and immunolabeled electron tomography experiments would be the first steps to answering those questions.

To gain a better understanding of how non-traditional AP coats form, it will be necessary to study interactions with the entire AP4 complex. AP4 cannot be expressed and purified as a complex from *E. coli*, and our pilot studies in mammalian cell culture have not produced the full complex yet. I have shown that the β4/μ4 heterodimer produced by *in vitro* wheat germ translation is stable and binds tepsin and Arf1 (Figure A4-1), but the ε/σ4 heterodimer does not behave as well. The ability to reconstitute AP4 *in vitro* will allow for budding assays using liposomes to be conducted. Studying the morphology of buds with and without tepsin could help answer whether tepsin plays a structural scaffold role in AP4 vesicle formation. Using tepsin constructs with one or both C-terminal AP4 binding motifs mutated (Chapter II) would allow for more precise dissection of the role of tepsin in AP4 coat assembly. Structural studies with the AP4 complex would reveal whether tepsin can interact with multiple AP4 complexes. Likely EM experiments would be necessary to answer this question as the flexibility of the hinges and appendages of the

large subunits make crystallizing full AP complexes difficult. However, x-ray crystallography could be used to determine if AP4 exhibits the same membrane-based allostery as AP1 and AP2.

As the founder of the AP family, and the most well-known and well-characterized member, AP2 is often viewed as the model for AP structure and function. The generalization of AP2 discoveries is upheld by studies with AP1, with the caveat that AP1 structures were determined by molecular replacement with the AP2 structures as templates, which could induce some bias. But perhaps there are fundamental structural differences between APs that interact with clathrin (AP1/AP2/AP3) and APs that do not (AP4/AP5). The membrane-based allostery that AP1 and AP2 exhibit elegantly explains how these complexes only bind cargo and polymerize clathrin at the correct membrane: the YXXΦ and clathrin binding sites are available only in the open conformation. Additionally, work from the Schmid group suggests the YXXΦ motif has a broader role in vesicle initiation dynamics beyond incorporating YXXΦ-containing cargoes (see Chapter I; Kadlecova *et al.,* 2017).

AP4 binds traditional YXXΦ motifs weakly (Hirst *et al.,* 1999; Aguilar *et al.,* 2001), and no *bona fide* AP4 cargoes with a YXXΦ motif have been discovered. In the papers that identify YXXΦ, AP4 depletion does not change the localization of the proposed cargoes. This suggests AP4 does not drive their trafficking. The YKFFE motif in APP discovered by the Bonifacino group (Burgos *et al.,* 2010) does not bind in the conserved tyrosine pocket, but rather on a surface on the opposite side of μ4, which is obscured in the open, "active" form of AP4 (Figure 1-8). AP5 lacks conserved residues that comprise the tyrosine pocket altogether, suggesting it is incapable of binding YXXΦ motifs (Hirst *et al.,* 2011). Although that does not rule out the possibility AP5 evolved to bind another type of motif with that surface.

The apparent lack of relevant YXXΦ cargo binding in the non-clathrin associated APs raises several questions. Is the primary function of the membrane-based allostery to prevent non-productive clathrin polymerization? Perhaps then AP4 and AP5 did not have such a selective pressure on the conformational change and did not need to utilize the membrane-based allostery,

leading to development of other cargo binding sites outside of the YXXΦ binding pocket. If tepsin functions as a scaffold or polymerization factor for AP4 vesicle formation, utilizing its ability to bind AP4 by the ε and β appendage domains (and a yet undetermined region), the closed conformation may not prevent polymerization. Structures of AP complexes with the appendages have not been determined, but work has shown the β2 hinge is buried in the closed complex (Kelly *et al.,* 2014). Thus, the appendages may bind back to the core, but it is not known which surfaces would be occluded.

Multiple motifs for interactions with AP complexes is not unusual (epsin1 has multiple DPW motifs for interacting with α; Figure 1-9), and may not indicate tepsin helps polymerize AP4 vesicles. The crystal packing of the tENTH domain (Chapter II) suggests the possibility for oligomerization, which could help drive the process of clustering AP4 at the membrane to efficiently form vesicles. However, we never found evidence of tepsin multimers in solution. Clathrin provides structural support for the extreme curvature of vesicle membranes, but it also helps vesicle budding initiation by crosslinking and stabilizing APs, leading to greater chances of successful CCV formation. With no known candidates for such a role in AP4 vesicles, the dynamics of vesicle formation remain unknown. Future work should test how tepsin influences the rates of AP4 vesicle budding.

Perhaps the primary function of AP family membrane-based allostery is to create a co-planar surface of multiple membrane- and cargo-binding sites, thus stabilizing the AP complex at the membrane? How does the YKFFE binding pocket fit into this model? When the YKFFE binding pocket is exposed on μ4, AP4 would be in the closed, inactive conformation—lacking the coplanar arrangement of membrane- and cargo- binding sites. However, while some of the α/β2/μ2 residues implicated in phosphoinositide interactions (Jackson *et al.,* 2010) are conserved in ε/β4/μ4, it is not known whether they play a significant role in interacting with the membrane, since Arf1 is necessary for AP4 recruitment. In AP1, Arf1 recruits the complex and drives the conformational change (Ren *et al.,* 2013). Previous Y2H work (Boehm *et al.,* 2001) suggests AP4

might use a different mechanism to interact with Arf1. Structural studies of AP4 interacting with Arf1 could provide further insight into whether AP4 possesses a membrane-based conformational change.

With no robust AP4 cargoes known, I could not assay the function of tepsin in AP4 vesicles because I had no read out for normal AP4 trafficking. Now, armed with the new knowledge of how AP4 and tepsin interact (Chapter II; Mattera *et al.,* 2015) and a well-described cargo and mistrafficking phenotype (Mattera *et al.,* 2017; Davies *et al.,* 2017; Ivankovic *et al.,* 2017), future work can probe how tepsin affects AP4 vesicle assembly. Tepsin or β4 mutants that abrogate the interaction could be introduced into cell lines, and the localization of Atg9a could serve as a readout of AP4 trafficking efficiency. Additionally, the rates of AP4 vesicle budding could be measured in these cell lines to answer whether tepsin affects AP4 assembly.

BIBLIOGRAPHY

Abdollahpour H, Alawi M, Kortüm F, *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(2):256-259. DOI: https://doi.org/10.1038/ejhg.2014.73.

Abou Jamra R, Philippe O, Raas-Rothschild A, *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(6):788-795. DOI: https://doi.org/10.1016/j.ajhg.2011.04.019.

Adams PD, Afonine PV, Bunkóczi G, *et al*. PHENIX: a comprehensive, Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*. 2010;66:213-221.

Afonine P, Grosse-Kunstleve RW, Adams PD. phenix.refine. *CCP4 News*. 2005; 42: Contribution 8.

Aguilar RC, Boehm M, Gorshkova I, *et al*. Signal-binding Specificity of the μ4 Subunit of the Adaptor Protein Complex AP-4. *J Biol Chem*. 2001;276:13145–13152.

Aguilar RC, Watson HA, Wendland B. The yeast Epsin Ent1 is recruited to membranes through multiple independent interactions. *J Biol Chem*. 2003;278(12):10737-10743. DOI: https://doi.org/10.1074/jbc.M211622200.

Alpy F, Rousseau A, Schwab Y, *et al*. STARD3 or STARD3NL and VAP form a novel molecular tether between late endosomes and the ER. *J Cell Sci*. 2013;126:5500 LP-5512.

Ang AL, Taguchi T, Francis S, *et al*. Recycling endosomes can serve as intermediates during transport from the Golgi to the plasma membrane of MDCK cells. *J Cell Biol*. 2004;167:531–543. A

Antonescu C, Aguet F, Danuser G, *et al*. Phosphatidylinositol-(4, 5)-bisphosphate regulates clathrin-coated pit initiation, stabilization, and size. *Mol Biol Cell.* 2011;22:2588 – 2600.

Arakel EC, Schwappach B. Formation of COPI-coated vesicles at a glance. *J Cell Sci*. 2018;131.

Aviram N, Schuldiner M. Targeting and translocation of proteins to the endoplasmic reticulum at a glance. *J Cell Sci*. 2017;130:4079 LP-4085.

Backes S, Herrmann JM. Protein Translocation into the Intermembrane Space and Matrix of Mitochondria: Mechanisms and Driving Forces. *Front Mol Biosci.* 2017;4:83.

Balastik M, Ferraguti F, Pires-da Silva A, *et al*. Deficiency in ubiquitin ligase TRIM2 causes accumulation of neurofilament light chain and neurodegeneration. *Proc Natl Acad Sci*. 2008;105:12016–12021.

Basilion JP, Kennedy MC, Beinert H, *et al.* Overexpression of Iron-Responsive Element-Binding Protein and Its Analytical Characterization as the RNA-Binding Form, Devoid of an Iron-Sulfur Cluster. *Arch Biochem Biophys*. 1994;311:517–522.

Bauer P, Leshinsky-Silver E, Blumkin L, *et al*. Mutation in the AP4B1 gene cause hereditary spastic paraplegia type 47 (SPG47). *Neurogenetics.* 2012;13:73–76.

Beilschmidt LK, Ollagnier de Choudens S, Fournier M, *et al*. ISCA1 is essential for mitochondrial Fe4S4 biogenesis in vivo. *Nat Commun.* 2017;8:15124.

Bentley M, Banker G. The cellular mechanisms that maintain neuronal polarity. *Nat Rev Neurosci*. 2016;17:611.

Birgisdottir AB, Lamark T, Johansen T. The LIR motif – crucial for selective autophagy. *J Cell Sci*. 2013;126:3237–3247.

Bjørkøy G, Lamark T, Brech A, *et al*. p62/SQSTM1 forms protein aggregates degraded by autophagy and has a protective effect on huntingtin-induced cell death. *J Cell Biol*. 2005;171:603 LP-614.

Blackstone C. Cellular Pathways of Hereditary Spastic Paraplegia. *Annu Rev Neurosci*. 2012;35:25–47.

Boal F, Mansour R, Gayral M, *et al*. TOM1 is a PI5P effector involved in the regulation of endosomal maturation. *J Cell Sci*. 2015;128(4):815-827. DOI: https://doi.org/10.1242/jcs.166314.

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(22):6265-6276. DOI: https://doi.org/10.1093/emboj/20.22.6265.

Bonifacino JS, Traub LM. Signals for Sorting of Transmembrane Proteins to Endosomes and Lysosomes. *Annu Rev Biochem*. 2003;72:395–447.

Borner GHH, Antrobus R, Hirst J, *et al*. Multivariate proteomic profiling identifies novel accessory proteins of coated vesicles. *J Cell Biol*. 2012;197(1):141-160. DOI: https://doi.org/10.1083/jcb.201111049.

Boucrot E, Saffarian S, Zhang R, *et al*. Roles of AP-2 in clathrin-mediated endocytosis. *PLoS One*. 2010;5:e10597.

Braymer JJ, Lill R. Iron-Sulfur Cluster Biogenesis and Trafficking in Mitochondria. J Biol Chem 2017; 292,12754-12763

Brett TJ, Traub LM, Fremont DH. Accessory protein recruitment motifs in clathrin-mediated endocytosis. *Structure*. 2002;10(6):797-809.

Brown DA, Crise B, Rose JK. Mechanism of membrane anchoring affects polarized expression of two proteins in MDCK cells. *Science*. 1989;245:1499 LP-1501.

Brown MD, Banker GA, Hussaini IM, Gonias SL, VandenBerg SR. Low density lipoprotein receptor-related protein is expressed early and becomes restricted to a somatodendritic domain during neuronal differentiation in culture. *Brain Res*. 1997;747:313–317.

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-W357. DOI: https://doi.org/10.1093/nar/ gkt381.

Burgos PV, Mardones GA, Rojas AL, *et al*. Sorting of the Alzheimer's disease amyloid precursor protein mediated by the AP-4 complex. *Dev Cell.* 2010;18:425 – 436.

Casanova JE, Apodaca G, Mostov KE. An autonomous signal for basolateral sorting in the cytoplasmic domain of the polymeric immunoglobulin receptor. *Cell*. 1991;66:65–75.

Chen H, Fre S, Slepnev VI, *et al*. Epsin is an EH-domain-binding protein implicated in clathrin-mediated endocytosis. *Nature*. 1998;394:793.

Chidambaram S, Müllers N, Wiederhold K, Haucke V, von Mollard GF. Specific interaction between SNAREs and epsin N-terminal homology (ENTH) domains of epsin-related proteins in trans-Golgi network to endosome transport.. 2004;279(6):4175-4179. DOI: https://doi.org/10.1074/jbc. M308667200.

Chidambaram S, Zimmermann J, von Mollard GF. ENTH domain proteins are cargo adaptors for multiple SNARE proteins at the TGN endosome. *J Cell Sci.* 2008;121:329 LP-338.

Collins BM, McCoy AJ, Kent HM, *et al*. Molecular Architecture and Functional Model of the Endocytic AP2 Complex. *Cell*. 2002;109:523–535.

Combet C, Blanchet C, Geourjon C, Deléage G. NPS@: network protein sequence analysis. *Trends Biochem Sci*. 2000;25(3):147-150.

Cong L, Ran FA, Cox D, *et al*. Multiplex genome engineering using CRISPR/Cas systems. *Science.* 2013;339:819 – 823.

Cresawn KO, Potter BA, Oztan A, *et al*. Differential involvement of endocytic compartments in the biosynthetic traffic of apical proteins. *EMBO J*. 2007; 26:3737–3748.

Csordás G, Renken C, Várnai P, *et al*. Structural and functional features and significance of the physical linkage between ER and mitochondria. *J Cell Biol*. 2006;174:915 LP-921.

Davies AK, Itzhak DN, Edgar JR, Archuleta TL, Hirst J, Jackson LP, Robinson MS, Borner GHH. AP-4 vesicles unmasked by organellar proteomics to reveal their cargo and machinery. *bioRxiv*. 2017. [pre-pub]

De Craene J-O, Ripp R, Lecompte O, Thompson JD, Poch O, Friant S. Evolutionary analysis of the ENTH/ANTH/VHS protein superfamily reveals a coevolution between membrane trafficking and metabolism. *BMC Genomics*. 2012;13(1):297. DOI: https://doi.org/10.1186/1471- 2164-13-297.

De Matteis MA, Luini A. Mendelian Disorders of Membrane Trafficking. *N Engl J Med*. 2011;365:927–938.

Dell'Angelica EC, Klumperman J, Stoorvogel W, *et al*. Association of the AP-3 adaptor complex with clathrin. *Science*. 1998;280:431–434.

Dell'Angelica EC, Mullins C, Bonifacino JS. AP-4, a novel protein complex related to clathrin adaptors. *J Biol Chem.* 1999;274:7278–7285.

Di Paolo G, De Camilli P. Phosphoinositides in cell regulation and membrane dynamics. *Nature.* 2006;443:651.

Dodonova SO, Diestelkoetter-Bachert P, von Appen A, *et al*. A structure of the COPI coat and the role of coat proteins in membrane vesicle assembly. *Science*. 2015;349:195–198.

Donaldson JG, Jackson CL. ARF family G proteins and their regulators: Roles in membrane transport, development and disease. *Nat Rev Mol Cell Biol.* 2011;12:362–375.

Drake MT, Downs MA, Traub LM. Epsin binds to clathrin by associating directly with the clathrin-terminal domain. Evidence for cooperative binding through two discrete sites. *J Biol Chem.* 2000;275:6479 – 6489.

Drake MT, Zhu Y, Kornfeld S. The Assembly of AP-3 Adaptor Complex-containing Clathrin-coated Vesicles on Synthetic Liposomes. *Mol Biol Cell*. 2000;11:3723–3736.

Ebrahimi-Fakhari D, Cheng C, Dies K, *et al*. Clinical and genetic characterization of AP4B1-associated SPG47. *Am J Med Genet Part A.* 2018;176:311–318.

Edeling MA, Mishra SK, Keyel PA, *et al*. Molecular switches involving the AP-2 beta2 appendage regulate endocytic cargo selection and clathrin coat assembly. *Dev Cell*. 2006;10(3):329-342. DOI: https:// doi.org/10.1016/j.devcel.2006.01.016.

Edeling MA, Smith C, Owen D. Life of a clathrin coat: insights from clathrin and AP structures. *Nat Rev Mol Cell Biol.* 2006;7:32 – 44.

Edgar, RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32(5):1792-1797. DOI: https://doi.org/10.1093/nar/gkh340.

Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*. 2010;66(Pt 4):486-501. DOI: https://doi.org/10.1107/S09074449 10007493.

Evans PR. An introduction to data reduction: space-group determination, scaling and intensity statistics, *Acta Crytsallogr*., 2011, D67, 282-292.

Faúndez V, Horng J-T, Kelly RB. A Function for the AP3 Coat Complex in Synaptic Vesicle Formation from Endosomes. *Cell*. 1998;93:423–432.

Fehon RG, McClatchey AI, Bretscher A. Organizing the Cell Cortex: The role of ERM proteins. *Nat Rev Mol Cell Biol.* 2010;11:276–287.

Felsenstein J. PHYLIP (Phylogeny Inference Package) version 3.6. *Distrib by author*. 2005; Department.

Ferguson SM, De Camilli P. Dynamin, a membrane remodelling GTPase. *Nat Rev Mol Cell Biol*. 2012;13:75–88.

Field MC, Gabernet-Castello C, Dacks JB. Reconstructing the evolution of the endocytic system: insights from genomics and molecular cell biology. *Adv Exp Med Biol.* 2007;607:84 – 96.

Fölsch H, Ohno H, Bonifacino JS, *et al*. A Novel Clathrin Adaptor Complex Mediates Basolateral Targeting in Polarized Epithelial Cells. *Cell*. 1999;99:189–198.

Ford MGJ, Mills IG, Peter BJ, *et al*. Curvature of clathrin-coated pits driven by epsin. *Nature*. 2002;419(6905):361-366. DOI: https://doi.org/10.1038/nature01020.

Fotin A, Cheng Y, Sliz P, *et al*. Molecular model for a complete clathrin lattice from electron cryomicroscopy. *Nature*. 2004;432:573.

Frazier MN, Davies AK, Voehler M, *et al*. Molecular basis for the interaction between AP4 β4 and its accessory protein, Tepsin. *Traffic*. 2016;17(4):400-415. DOI: https://doi.org/10.1111/tra.12375.

Gabernet-Castello C, Dacks JB, Field MC. The single ENTH-domain protein of trypanosomes; endocytic functions and evolutionary relationship with epsin. *Traffic*. 2009;10(7):894-911. DOI: https://doi.org/10.1111/j.16000854.2009.00910.x.

Gaidarov I, Keen JH. Phosphoinositide–Ap-2 Interactions Required for Targeting to Plasma Membrane Clathrin-Coated Pits. *J Cell Biol*. 1999;146:755 LP-764.

Gautier R, Douguet D, Antonny B *et al*. HELIQUEST: a web server to screen sequences with specific α-helical properties. *Bioinformatics*. 2008 Sep 15;24(18):2101-2.

González C, Cánovas J, Fresno J, *et al*. Axons provide the secretory machinery for trafficking of voltage-gated sodium channels in peripheral nerve. *Proc Natl Acad Sci USA.* 2016;113:1823–1828.

Greener T, Zhao X, Nojima H, *et al*. Role of Cyclin G-associated Kinase in Uncoating Clathrin-coated Vesicles from Non-neuronal Cells. *J Biol Chem.* 2000;275:1365–1370.

Grzyb J, Xu F, Nanda V, *et al*. Empirical and computational design of iron-sulfur cluster proteins. *Biochim Biophys Acta – Bioenerg*. 2012;1817:1256–1262.

Haft CR, Sierra M de la L, Bafford R, *et al*. Human Orthologs of Yeast Vacuolar Protein Sorting Proteins Vps26, 29, and 35: Assembly into Multimeric Complexes. *Mol Biol Cell*. 2000;11:4105–4116.

Haile DJ, Rouault TA, Harford JB, *et al*. Cellular regulation of the iron-responsive element binding protein: disassembly of the cubane iron-sulfur cluster results in high-affinity RNA binding. *Proc Natl Acad Sci U S A*. 1992;89:11735–11739.

Hara-Kuge, S, Kuge O, Orci L, *et al.* * En Bloc incorporation of coatomer subunits during the assembly of COP-coated vesicles. *J Cell Biol.* 1994; 124: 883-892.

Harbour ME, Breusegem SYA, Antrobus R, *et al*. The cargo-selective retromer complex is a recruiting hub for protein complexes that regulate endosomal tubule dynamics. *J Cell Sci.* 2010;123:3703–3717.

Hardies K, May P, Djemie T, *et al*. Recessive loss-of-function mutations in AP4S1 cause mild fever-sensitive seizures, developmental delay and spastic paraplegia through loss of AP-4 complex assembly. *Hum Mol Genet*. 2015;24:1-10. DOI: https://doi.org/10.1093/hmg/ ddu740.

Hawryluk MJ, Keyel PA, Mishra SK, *et al*. Epsin 1 is a polyubiquitin-selective clathrin-associated sorting protein. *Traffic.* 2006;7:262 – 281.

Hazan J, Fonknechten N, Mavel D, *et al*. Spastin, a new AAA protein, is altered in the most frequent form of autosomal dominant spastic paraplegia. *Nat Genet.* 1999;23:296.

Hein MY, Hubner NC, Poser I, *et al*. A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell*. 2015;163(3):712-723. DOI: https://doi.org/10.1016/j. cell.2015.09.053.

Heldwein EE, Macia E, Wang J, *et al*. Crystal structure of the clathrin adaptor protein 1 core. *Proc Natl Acad Sci USA*. 2004;101:14108–14113.

Hierro A, Rojas AL, Rojas R, *et al*. Functional architecture of the retromer cargo-recognition complex. *Nature*. 2007;449:1063–1067.

Hirst J, Bright NA, Rous B, *et al*. Characterization of a fourth adaptor-related protein complex. *Mol Biol Cell.* 1999;10:2787 – 2802.

Hirst J, Motley A, Harasaki K, *et al*. EpsinR: an ENTH domain-containing protein that interacts with AP-1. *Mol Biol Cell*. 2003;14:625 – 641.

Hirst J, Miller SE, Taylor MJ, *et al*. EpsinR is an adaptor for the SNARE protein Vti1b. *Mol Biol Cell*. 2004;15(12):5593-5602. DOI: https://doi.org/10.1091/mbc.E04-06- 0468.

Hirst J, D. Barlow L, Francisco GC, *et al*. The Fifth Adaptor Protein Complex. *PLoS Biol.* 2011;9:e1001170.

Hirst J, Borner GHH, Antrobus R, *et al*. Distinct and overlapping roles for AP-1 and GGAs revealed by the “knocksideways” system. *Curr Biol.* 2012;22:1711 – 1716.
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(2):153-164. DOI: https://doi.org/10. 1111/tra.12028.

Hirst J, Schlacht A, Norcott JP, *et al*. Characterization of TSET, an ancient and widespread membrane trafficking complex. *Elife*. 2014;3:e02866.

Hirst J, Itzhak DN, Antrobus R, *et al*. Role of the AP-5 adaptor protein complex in late endosome-to-Golgi retrieval. *PLOS Biol*. 2018;16:e2004411.

Hong Y-H, Ahn H-C, Lim J, *et al*. Identification of a novel ubiquitin binding site of STAM1 VHS domain by NMR spectroscopy. *FEBS Lett*. 2009;583:287–292.

Höning S, Ricotta D, Krauss M, *et al*. Phosphatidylinositol-(4,5)-bisphosphate regulates sorting signal recognition by the clathrin- associated adaptor complex AP2. *Mol Cell*. 2005;18(5):519-531. DOI: https://doi.org/10.1016/j.molcel.2005.04.019.

Hsu PD, Scott DA, Weinstein JA, *et al*. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol.* 2013;31:827 – 832.

Huttlin EL, Ting L, Bruckner LJ, *et al*. The BioPlex Network of Human Protein Interactions: Additional Unpublished AP-MS Results (Pre-Publication). *Biogrid*. https://thebiogrid.org/166968/publication/the-bioplex-network-of-human-protein-interactions-additional-unpublished-ap-ms-results.html

Itoh T, De Camilli P. BAR, F-BAR (EFC) and ENTH/ANTH domains in the regulation of membrane-cytosol interfaces and membrane curvature. *Biochim Biophys Acta*. 2006;1761(8):897-912. DOI: https://doi.org/10.1016/j.bbalip.2006.06.015.

Itzhak DN, Tyanova S, Cox J, *et al*. Global, quantitative and dynamic mapping of protein subcellular localization. *Elife*. 2016;5:e16950.

Itzhak DN, Davies C, Tyanova S, *et al*. A Mass Spectrometry-Based Approach for Mapping Protein Subcellular Localization Reveals the Spatial Proteome of Mouse Primary Neurons. *Cell Rep*. 2018;20:2706–2718.

Ivankovic D, López-Doménech G, Drew J, Tooze SA, Kittler JT. AP-4 mediated ATG9A sorting underlies axonal and autophagosome biogenesis defects in a mouse model of AP-4 deficiency syndrome. *bioRxiv*. 2017. [pre-pub]

Jackson LP, Kelly BT, McCoy AJ, *et al*. A large-scale conformational change couples membrane recruitment to cargo binding in the AP2 clathrin adaptor complex. *Cell.* 2010;141(7):1220-1229. DOI: https://doi.org/10.1016/j.cell.2010.05.006.

Jackson LP, Lewis M, Kent H, *et al*. Molecular basis for recognition of dilysine trafficking motifs by COPI. *Dev Cell*. 2012;23(6):1255-1262.
DOI: https://doi.org/10.1016/j.devcel.2012.10.017.

Jacomin A-C, Samavedam S, Promponas V, *et al*. iLIR database: A web resource for LIR motif-containing proteins in eukaryotes. *Autophagy*. 2016;12:1945–1953.

Jost M, Simpson F, Kavran JM, Lemmon MA, Schmid SL. Phosphatidylinositol-4,5-bisphosphate is required for endocytic coated vesicle formation. *Curr Biol*. 1998;8:1399–1404.

Kabeya Y, Mizushima N, Yamamoto A, *et al*. LC3, GABARAP and GATE16 localize to autophagosomal membrane depending on form-II formation. *J Cell Sci*. 2004;117:2805 LP-2812.

Kadlecova Z, Spielman SJ, Loerke D, *et al*. Regulation of clathrin-mediated endocytosis by hierarchical allosteric activation of AP2. *J Cell Biol.* 2017;216:167 LP-179.

Kanehisa M, Goto S, Kawashima S, *et al*. The KEGG databases at GenomeNet. Nucleic Acids Res 2002;30:42 – 46.

Kantheti P, Qiao X, Diaz ME, *et al*. Mutation in AP-3 δ in the mocha Mouse Links Endosomal Transport to Storage Deficiency in Platelets, Melanosomes, and Synaptic Vesicles. *Neuron*. 1998;21:111–122.

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772-780. DOI: https://doi.org/10.1093/molbev/mst010.

Kaut A, Lange H, Diekert K, *et al*. Isa1p Is a Component of the Mitochondrial Machinery for Maturation of Cellular Iron-Sulfur Proteins and Requires Conserved Cysteine Residues for Function. *J Biol Chem.* 2000;275:15955–15961.

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(2):435-438. DOI: https://doi. org/10.1110/ps.8.2.435.

Keen JH. Clathrin assembly proteins: affinity purification and a model for coat assembly. *J Cell Biol*. 1987;105:1989–1998.

Kelly BT, McCoy AJ, Späte K, *et al*. A structural explanation for the binding of endocytic dileucine motifs by the AP2 complex. *Nature*. 2008;456(7224):976-979. DOI: https://doi.org/10.1038/nature07422.

Kelly BT, Owen DJ. Endocytic sorting of transmembrane protein cargo. *Curr Opin Cell Biol*. 2011;23:1-9. DOI: https://doi.org/10.1016/j.ceb.2011.03.004.

Kelly BT, Graham SC, Liska N, *et al*. AP2 controls clathrin polymerization with a membrane-activated switch. *Science*. 2014;345:459–463.

Kent HM, McMahon HT, Evans PR, *et al*. γ-Adaptin Appendage Domain. *Structure.* 2002;10:1139–1148.

Kent HM, Evans PR, Schäfer IB, *et al*. Structural Basis of the Intracellular Sorting of the SNARE VAMP7 by the AP3 Adaptor Complex. *Dev Cell.* 2012;22:979–988.

Kirchhausen T. Clathrin. *Annu Rev Biochem*. 2000;69:699–727.

Kirchhausen T, Owen D, Harrison SC. Molecular structure, function, and dynamics of clathrin-mediated membrane traffic. *Cold Spring Harb Perspect Biol.* 2014;6:a016725.

Koumandou VL, Klute MJ, Herman EK, *et al*. Evolutionary reconstruction of the retromer complex and its function in Trypanosoma brucei. *J Cell Sci*. 2011;124:1496–1509.

Krauss M, Kinuta M, Wenk MR, *et al*. ARF6 stimulates clathrin/AP-2 recruitment to synaptic membranes by activating phosphatidylinositol phosphate kinase type Iγ. *J Cell Biol.* 2003;162:113 LP-124.

Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr*. 2004;60(12):2256-2268. DOI: https://doi.org/10.1107/S0907444904026460.

Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol*. 2007;372:774-797.

Lamark T, Kirkin V, Dikic I, Johansen T. NBR1 and p62 as cargo receptors for selective autophagy of ubiquitinated targets. *Cell Cycle*. 2009;8:1986–1990.

Lamichhane D. New AP4B1 mutation in an African-American child associated with intellectual disability. *J Pediatr Genet.* 2013;2:191–195.

Levin-Kravets O, Tanner N, Shohat N, *et al*. A bacterial genetic selection system for ubiquitylation cascade discovery. *Nat Methods*. 2016;13(11):945-952. DOI: https://doi.org/10.1038/nmeth. 4003.

Lisanti MP, Sargiacomo M, Graeve L*, et al.* Polarized apical distribution of glycosyl-phosphatidylinositol-anchored proteins in a renal epithelial cell line. *Proc Natl Acad Sci USA*. 1988;85:9557–9561.

Lisanti MP, Caras IW, Davitz MA, *et al*. A glycophospholipid membrane anchor acts as an apical targeting signal in polarized epithelial cells. *J Cell Biol*. 1989;109:2145–2156.

Mallam AL, Marcotte EM. Systems-wide Studies Uncover Commander, a Multiprotein Complex Essential to Human Development. *Cell Syst*. 2018;4:483–494.

Manna PT, Gadelha C, Puttick AE, Field MC. ENTH and ANTH domain proteins participate in AP2-independent clathrin-mediated endocytosis. *J Cell Sci*. 2015;128(11):2130-2142. DOI: https://doi. org/10.1242/jcs.167726.

Matsuda S, Miura E, Matsuda K, *et al*. Accumulation of AMPA receptors in autophagosomes in neuronal axons lacking adaptor protein AP-4. *Neuron*. 2008;57(5):730-745. DOI: https://doi.org/10.1016/j. neuron.2008.02.012.

Matsuda S, Yuzaki M. AP-4: Auto-phagy 4 mislocalized proteins in axons. *Autophagy*. 2008;4:815–816.

Matter K, Hunziker W, Mellman I. Basolateral sorting of LDL receptor in MDCK cells: The cytoplasmic domain contains two tyrosine-dependent targeting determinants. *Cell.* 1992;71:741–753.

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(52):30736- 30749. DOI: https://doi.org/10.1074/jbc.M115.683409.

Mattera R, Park SY, De Pace R, Guardia CM, Bonifacino JS. AP-4 mediates export of ATG9A from the *trans*-Golgi network to promote autophagosome formation. *Proc Natl Acad Sci.* 2017;114:E10697 LP-E10706.

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Cryst*. 2007;40:658-674.

McNally KE, Faulkner R, Steinberg F, *et al*. Retriever is a multiprotein complex for retromer-independent endosomal cargo recycling. *Nat Cell Biol*. 2017;19:1214–1225.

McNicholas S, Potterton E, Wilson KS, Noble MEM. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr D Biol Crystallogr*. 2011;67(4):386-394. DOI: https://doi.org/10.1107/ S0907444911007281.

Messa M, Fernández-Busnadiego R, Sun EW, *et al*. Epsin deficiency impairs endocytosis by stalling the actin-dependent invagination of endocytic clathrin-coated pits. *Elife*. 2014;3:e03311.

Miller SE, Collins BM, McCoy AJ, Robinson MS, Owen DJ. A SNARE-adaptor interaction is a new mode of cargo recognition in clathrin-coated vesicles. *Nature*. 2007;450(7169):570-574. DOI: https://doi. org/10.1038/nature06353.

Miller SE, Mathiasen S, Bright NA, *et al*. CALM regulates clathrin-coated vesicle size and maturation by directly sensing and driving membrane curvature. *Dev Cell.* 2015;33(2):163-175. DOI: https://doi.org/10.1016/j.devcel.2015.03.002.

Mills IG, Praefcke GJK, Vallis Y, *et al*. EpsinR: an AP1/clathrin interacting protein involved in vesicle trafficking. *J Cell Biol.* 2003;160:213 – 222.

Mirarab S, Nguyen N, Warnow T. PASTA: ultra-large multiple sequence alignment. *Proc RECOMB*. 2014;22:377-386.

Mishima K, Tsuchiya M, Nightingale MS, *et al*. ARD 1, a 64-kDa guanine nucleotide-binding protein with a carboxyl-terminal ADP-ribosylation factor domain. *J Biol Chem.* 1993;268:8801–8807.

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(6874):933-937. DOI: https://doi.org/10. 1038/415933a.

Mizuno E, Kawahata K, Kato M, *et al*. STAM Proteins Bind Ubiquitinated Proteins on the Early Endosome via the VHS Domain and Ubiquitin-interacting Motif. *Mol Biol Cell*. 2003;14:3675–3689.

Motley AM, Berg N, Taylor MJ, *et al*. Functional Analysis of AP-2 α and μ2 Subunits. *Mol Biol Cell.* 2006;17:5298–5308.

Nakatsu F, Okada M, Mori F, *et al*. Defective function of GABA-containing synaptic vesicles in mice lacking the AP-3B clathrin adaptor. *J Cell Biol.* 2004;167:293–302.

Namekawa M, Ribai P, Nelson I, *et al*. SPG3A is the most frequent cause of hereditary spastic paraplegia with onset before age 10 years. *Neurology*. 2006;66:112 LP-114.

Newman LS, McKeever MO, Okano HJ, *et al*. β-NAP, a cerebellar degeneration antigen, is a neuron-specific vesicle coat protein. *Cell*. 1995;82:773–783.

Nielsen MS, Madsen P, Christensen EI, *et al*. The sortilin cytoplasmic tail conveys Golgi–endosome transport and binds the VHS domain of the GGA2 sorting protein. *EMBO J*. 2001;20:2180 LP-2190.

Noble AJ, Stagg SM. COPI gets a fancy new coat. *Science*. 2015;349:142 LP-143.

Nogi T, Shiba Y, Kawasaki M, *et al*. Structural basis for the accessory protein recruitment by the γ-adaptin ear domain. *Nat Struct Biol.* 2002;9:527.

Ohno H, Fournier MC, Poy G, *et al*. Structural determinants of interaction of tyrosine-based sorting signals with the adaptor medium chains. *J Biol Chem.* 1996;271:29009–29015.

Oldham CE, Mohney RP, Miller SLH, *et al*. The ubiquitin-interacting motifs target the endocytic adaptor protein epsin for ubiquitination. *Curr Biol.* 2002;12:1112 – 1116.

Olusanya O, Andrews PD, Swedlow JR, *et al*. Phosphorylation of threonine 156 of the μ2 subunit of the AP2 complex is essential for endocytosis in vitro and in vivo. *Curr Biol*. 2001;11:896–900.

Owen DJ, Evans PR. A structural explanation for the recognition of tyrosine-based endocytotic signals. *Science*. 1998;282:1327-1332.

Owen DJ, Vallis Y, Noble MEM, *et al*. A structural explanation for the binding of multiple ligands by the α-adaptin appendage domain. *Cell*. 1999;97:805–815.

Owen DJ, Vallis Y, Pearse BM, *et al*. The structure and function of the β2-adaptin appendage domain. *EMBO J*. 2000;19(16):4216-4227. DOI: https://doi.org/10.1093/emboj/19.16. 4216.

Owen DJ, Collins BM, Evans PR. ADAPTORS FOR CLATHRIN COATS: Structure and Function. *Annu Rev Cell Dev Biol*. 2004;20:153–191.

Paleotti O, Macia E, Luton F, *et al*. The Small G-protein Arf6GTP Recruits the AP-2 Adaptor Complex to Membranes. *J Biol Chem.* 2005;280:21661–21666.

Pavel M, Rubinsztein DC. Mammalian autophagy and the plasma membrane. *FEBS J*. 2017;284:672–679.

Pearse BMF, Robinson MS. Purification and properties of 100-kd proteins from coated vesicles and their reconstitution with clathrin. *EMBO J*. 1984;3:1951–1957.

Perrakis A, Harkiolaki M, Wilson KS, Lamzin VS. ARP/wARP and molecular replacement. *Acta Crystallogr D Biol Crystallogr*. 2001;57(10):1445-1450.

Phillips MJ, Voeltz GK. Structure and function of ER membrane contact sites with other organelles. *Nat Rev Mol Cell Biol*. 2015;17:69. DOI: http://dx.doi.org/10.1038/nrm.2015.8

Praefcke GJK, Ford MGJ, Schmid EM, *et al*. Evolving nature of the AP2 alpha-appendage hub during clathrin-coated vesicle endocytosis. *EMBO J.* 2004;23:4371 – 4383.

Pryor PR, Jackson L, Gray SR, *et al*. Molecular Basis for the Sorting of the SNARE VAMP7 into Endocytic Clathrin-Coated Vesicles by the ArfGAP Hrb. *Cell*. 2008;134:817–827.

Puertollano R, Aguilar RC, Gorshkova I, *et al*. Sorting of Mannose 6-Phosphate Receptors Mediated by the GGAs. *Science*. 2001;292:1712 LP-1716.

Raiborg C, Malerød L, Pedersen NM, *et al*. Differential functions of Hrs and ESCRT proteins in endocytic membrane trafficking. *Exp Cell Res.* 2008;314:801–813.

Ran FA, Hsu PD, Lin C-Y, *et al*. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*. 2013;154:1380 – 1389.

Rapoport I, Chen YC, Cupers P, *et al*. Dileucine-based sorting signals bind to the β chain of AP-1 at a site distinct and regulated differently from the tyrosine-based motif-binding site. *EMBO J*. 1998;17:2148 LP-2155.

Ren X, Hurley JH. VHS domains of ESCRT-0 cooperate in high-avidity binding to polyubiquitinated cargo. *EMBO J*. 2010;29(6):1045-1054. DOI: https://doi.org/10.1038/emboj.2010.6.

Ren X, Farías GG, Canagarajah BJ, *et al*. Structural basis for recruitment and activation of the AP-1 clathrin adaptor complex by Arf1. *Cell*. 2013;152(4):755-767. DOI: https://doi.org/10.1016/j.cell.2012.12.042.

Ricotta D, Conner SD, Schmid SL, *et al*. Phosphorylation of the AP2 μ subunit by AAK1 mediates high affinity binding to membrane protein sorting signals. *J Cell Biol.* 2002;156:791 LP-795.

Rivera JF, Ahmad S, Quick MW, *et al*. An evolutionarily conserved dileucine motif in Shal K+ channels mediates dendritic targeting. *Nat Neurosci*. 2003;6:243.

Robinson MS. Adaptable adaptors for coated vesicles. *Trends Cell Biol*. 2004;14:167–174.

Robinson MS. Forty years of clathrin-coated vesicles. *Traffic*. 2015;16:1210-1238. DOI: https://doi.org/10.1111/tra.12335.

Rohde G, Wenzel D, Haucke V. A phosphatidylinositol (4,5)-bisphosphate binding site within μ2-adaptin regulates clathrin-mediated endocytosis. *J Cell Biol.* 2002;158:209 LP-214.

Rosenthal JA, Chen H, Slepnev VI, *et al*. The epsins define a family of proteins that interact with components of the clathrin coat and contain a new protein module. *J Biol Chem*. 1999;274(48):33959-33965.

Rouault TA, Maio N. Biogenesis and functions of mammalian iron-sulfur proteins in the regulation of iron homeostasis and pivotal metabolic pathways. *J Biol Chem.* 2017;292:12744–12753.

Salazar G, Love R, Werner E, *et al*. The Zinc Transporter ZnT3 Interacts with AP-3 and It Is Preferentially Targeted to a Distinct Synaptic Vesicle Subpopulation. *Mol Biol Cell.* 2004;15:575–587.

Salazar G, Love R, Styers ML, *et al*. AP-3-dependent Mechanisms Control the Targeting of a Chloride Channel (ClC-3) in Neuronal and Non-neuronal Cells. *J Biol Chem*. 2004b;279:25430–25439.

Salazar G, Craige B, Love R, *et al*. Vglut1 and ZnT3 co-targeting mechanisms regulate vesicular zinc stores in PC12 cells. *J Cell Sci.* 2005;118:1911 LP-1921.

Salazar G, Craige B, Styers ML, *et al*. BLOC-1 Complex Deficiency Alters the Targeting of Adaptor Protein Complex-3 Cargoes. *Mol Biol Cell.* 2006;17:4014–4026.

Schlipf NA, Schüle R, Klimpe S, *et al*. AP5Z1/SPG48 frequency in autosomal recessive and sporadic spastic paraplegia*. Mol Genet Genomic Med.* 2014;2:379–382.

Schmid EM, Ford MGJ, Burtey A, *et al*. Role of the AP2 beta-appendage hub in recruiting partners for clathrin-coated vesicle assembly. PLoS Biol 2006;4:e262.

Sheff DR, Kroschewski R, Mellman I. Actin Dependence of Polarized Receptor Recycling in Madin-Darby Canine Kidney Cell Endosomes. *Mol Biol Cell.* 2002;13:262–275.

Sheftel AD, Wilbrecht C, Stehling O, *et al*. The human mitochondrial ISCA1, ISCA2, and IBA57 proteins are required for [4Fe-4S] protein maturation. *Mol Biol Cell.* 2012;23:1157–1166.

Shen Q-T, Ren X, Zhang R, *et al*. HIV-1 Nef hijacks clathrin coats by stabilizing AP-1:Arf1 polygons. *Science.* 2015;350:aac5137.

Sievers F, Wilm A, Dineen D, *et al*. Fast, scalable generation of high- quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011;7:539. DOI: https://doi.org/10.1038/msb. 2011.75.

Simmen T, Höning S, Icking A, Tikkanen R, Hunziker W. AP-4 binds basolateral signals and participates in basolateral sorting in epithelial MDCK cells. *Nat Cell Biol.* 2002;4(2):154-159. DOI: https://doi.org/ 10.1038/ncb745.

Simpson F, Bright N, West M, *et al*. A novel adaptor-related protein complex. *J. Cell Biol*. 1996;133, 749 LP-760.

Skruzny M, Brach T, Ciuffa R, *et al*. Molecular basis for coupling the plasma membrane to the actin cytoskeleton during clathrin-mediated endocytosis. *Proc Natl Acad* Sci USA. 2012;109(38):E2533-E2542. DOI: https://doi.org/10.1073/ pnas.1207011109.

Song D, Tu Z, Lee FS. Human ISCA1 Interacts with IOP1/NARFL and Functions in Both Cytosolic and Mitochondrial Iron-Sulfur Protein Biogenesis. *J Biol Chem.* 2009;284:35297–35307.

Sparrer KMJ, Gableske S, Zurenski MA, *et al*. TRIM23 mediates virus-induced autophagy via activation of TBK1. *Nat Microbiol.* 2017;2:1543–1557.

Spradling KD, McDaniel AE, Lohi J, *et al*. Epsin 3 Is a Novel Extracellular Matrix-induced Transcript Specific to Wounded Epithelia. *J Biol Chem.* 2001;276:29257–29267.

Stachowiak JC, Hayden CC, Sasaki DY. Steric confinement of proteins on lipid membranes can drive curvature and tubulation. *Proc Natl Acad Sci.* 2010;107:7781 LP-7786.

Stachowiak JC, Schmid EM, Ryan CJ, *et al*. Membrane bending by protein-protein crowding. *Nat Cell Biol.* 2012;14:944–949.

Stagg SM, LaPointe P, Razvi A, *et al*. Structural Basis for Cargo Regulation of COPII Coat Assembly. *Cell*. 2008;134:474–484.

Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312- 1313. DOI: https://doi.org/10.1093/bioinformatics/btu033.

Stamnes MA, Rothman JE. The binding of AP-1 clathrin adaptor particles to Golgi membranes requires ADP-ribosylation factor, a small GTP-binding protein. *Cell*. 1993;73:999–1005.

St Johnston D, Ahringer J. Cell Polarity in Eggs and Epithelia: Parallels and Diversity. *Cell*. 2010;141:757–774.

Stoops EH, Caplan MJ. Trafficking to the Apical and Basolateral Membranes in Polarized Epithelial Cells. *J Am Soc Nephrol*. 2014;25:1375–1386.

Takatsu H, Katoh Y, Shiba Y, *et al*. Golgi-localizing, γ-Adaptin Ear Homology Domain, ADP-ribosylation Factor-binding (GGA) Proteins Interact with Acidic Dileucine Sequences within the Cytoplasmic Domains of Sorting Receptors through Their Vps27p/Hrs/STAM (VHS) Domains. *J Biol Chem.* 2001;276:28541–28545.

Tan CA, Topper S, del Gaudio D, *et al*. Characterization of patients referred for non-specific intellectual disability testing: the importance of autosomal genes for diagnosis. *Clin Genet.* 2016;89:478–483.

Tannous A, Pisoni GB, Hebert DN, *et al*. N-linked sugar-regulated protein folding and quality control in the ER. *Semin Cell Dev Biol*. 2015;41:79–89.

ter Haar E, Harrison SC, Kirchhausen T. Peptide-in-groove interactions link target proteins to the β-propeller of clathrin. *Proc Natl Acad Sci USA* [2000;97:1096–1100.

Tiwari RK, Kusari J, Sen GC. Functional equivalents of interferon-mediated signals needed for induction of an mRNA can be generated by double-stranded RNA and growth factors. *EMBO J.* 1987;6:3373 – 3378.

Traub LM, Downs MA, Westrich JL, *et al*. Crystal structure of the alpha appendage of AP-2 reveals a recruitment platform for clathrin-coat assembly. *Proc Natl Acad Sci USA.* 1999;96:8907 – 8912.

Traub LM, Ostrom JA, Kornfeld S. Biochemical dissection of AP-1 recruitment onto Golgi membranes. *J Cell Biol*. 1993;123:561–573.

Turunen O, Wahlström T, Vaheri A. Ezrin has a COOH-terminal actin-binding site that is conserved in the ezrin protein family. *J Cell Biol*. 1994;126:1445–1453.

Tüysüz B, Bilguvar K, Koçer N, *et al*. Autosomal recessive spastic tetraplegia caused by AP4M1 and AP4B1 gene mutation: expansion of the facial and neuroimaging features. *Am J Med Genet A*. 2014;164A(7):1677-1685. DOI: https://doi.org/10.1002/ajmg.a. 36514.

Ungewickell E, Ungewickell H, Holstein SEH, *et al*. Role of auxilin in uncoating clathrin-coated vesicles. *Nature*. 1995;378:632.

Vichi A, Payne DM, Pacheco-Rodriguez G, *et al*. E3 ubiquitin ligase activity of the trifunctional ARD1 (ADP-ribosylation factor domain protein 1). *Proc Natl Acad Sci USA*. 2005;102:1945 LP-1950.

Vitale N, Moss J, Vaughan M. ARD1, a 64-kDa bifunctional protein containing an 18-kDa GTP-binding ADP-ribosylation factor domain and a 46-kDa GTPase-activating domain. *Proc Natl Acad Sci USA.*1996;93:1941–1944.

Vitale N, Horiba K, Ferrans VJ, *et al*. Localization of ADP-ribosylation factor domain protein 1 (ARD1) in lysosomes and Golgi apparatus. *Proc Natl Acad Sci USA*. 1998;95:8613–8618.

Voglmaier SM, Kam K, Yang H, *et al*. Distinct Endocytic Pathways Control the Rate and Extent of Synaptic Vesicle Protein Recycling. *Neuron*. 2006;51:71–84.

Wang WY, Malcolm BA. Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange Site Directed Mutagenesis. *Biotechniques.* 1999;26:680 – 682.

Wang YJ, Wang J, Sun HQ, *et al*. Phosphatidylinositol 4 Phosphate Regulates Targeting of Clathrin Adaptor AP-1 Complexes to the Golgi. *Cell*. 2003;114:299–310.

Weise C, Erbar S, Lamp B, *et al*. Tyrosine Residues in the Cytoplasmic Domains Affect Sorting and Fusion Activity of the Nipah Virus Glycoproteins in Polarized Epithelial Cells . *J Virol*. 2010;84:7634–7641.

West M, Zurek N, Hoenger A, Voeltz GK. A 3D analysis of yeast ER structure reveals how ER domains are organized by membrane curvature. *J Cell Biol*. 2011;193:333 LP-346.

Yap CC, Murate M, Kishigami S, *et al*. Adaptor protein complex-4 (AP-4) is expressed in the central nervous system neurons and interacts with glutamate receptor delta2. *Mol Cell Neurosci.* 2003;24:283 – 295.

Yeaman C, Gall AH Le, Baldwin AN, *et al*. The O-glycosylated Stalk Domain Is Required for Apical Sorting of Neurotrophin Receptors in Polarized MDCK Cells . *J Cell Biol*. 1997;139:929–940.

Yonemura S, Hirao M, Doi Y, *et al*. Ezrin/Radixin/Moesin (ERM) Proteins Bind to a Positively Charged Amino Acid Cluster in the Juxta-Membrane Cytoplasmic Domain of CD44, CD43, and ICAM-2 . *J Cell Biol*. 1998;140:885–895.

Yu X, Breitman M, Goldberg J. A structure-based mechanism for Arf1-dependent recruitment of coatomer to membranes. *Cell*. 2012;148(3):530-542. DOI: https://doi.org/10.1016/j.cell.2012.01.015.

Zaremba S, Keen JH. Assembly polypeptides from coated vesicles mediate reassembly of unique clathrin coats. *J Cell Biol*. 1983;97:1339 – 1347.
Zheng L, Cash VL, Flint DH, *et al*. Assembly of Iron-Sulfur Clusters: IDENTIFICATION OF AN iscSUA-hscBA-fdx GENE CLUSTER FROM AZOTOBACTER VINELANDII. *J Biol Chem*. 1998;273:13264–13272.

Zhang B, Koh YH, Beckstead RB, *et al*. Synaptic Vesicle Size and Number Are Regulated by a Clathrin Adaptor Protein Required for Endocytosis. *Neuron.* 1998;21:1465–1475.

Ziarek JJ, Peterson FC, Lytle BL, *et al*. Binding site identification and structure determination of protein-ligand complexes by NMR a semiautomated approach. *Methods Enzymol.* 2011;493:241 – 275.

APPENDIX I

Supplementary Materials from Chapter II

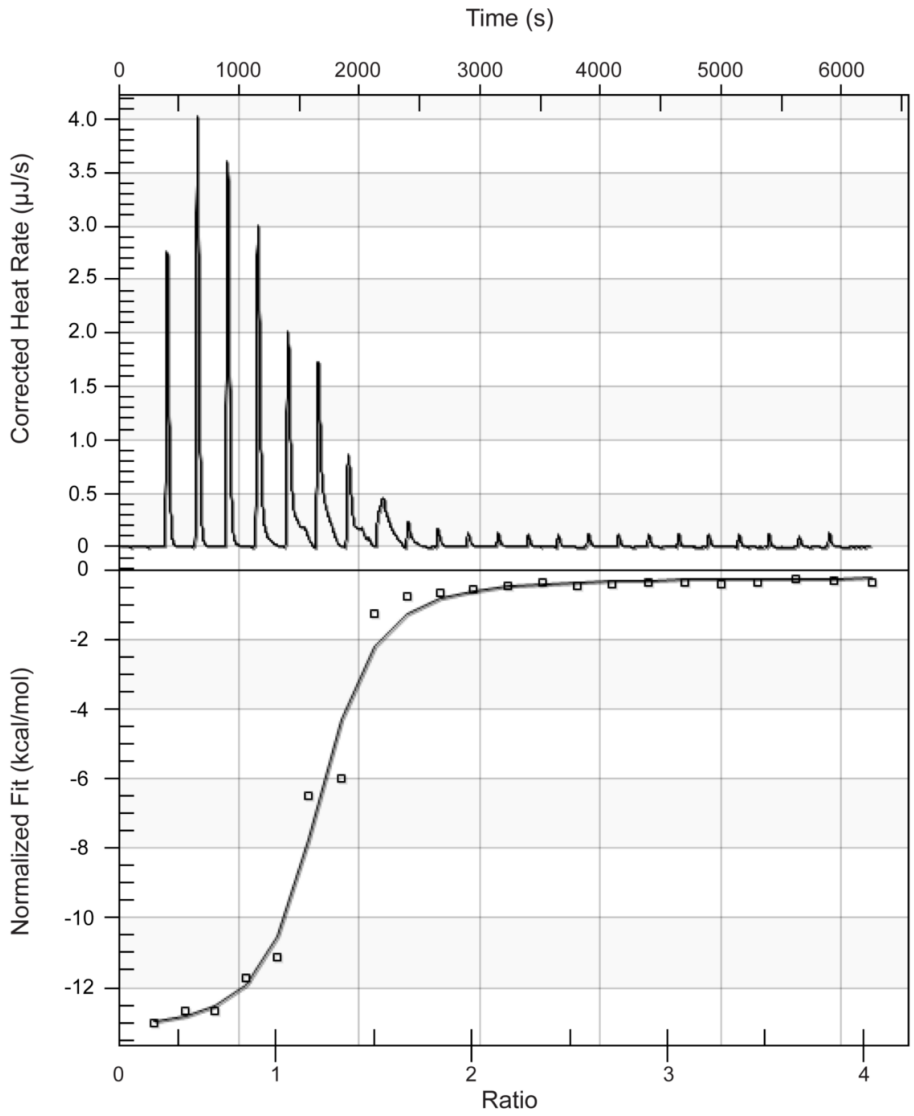


Figure A1-1. Tepsin peptide binding to β4 appendage domain. A short synthesized peptide corresponding to the conserved hydrophobic tepsin sequence (SSRDSLFAGMELVACS, GenHunter) exhibits binding to β4 with similar K_D and stoichiometry (K_D=1.47 µM, n = 1.2) as the recombinant tepsin fragment (residues 450-500) used predominantly in this work.

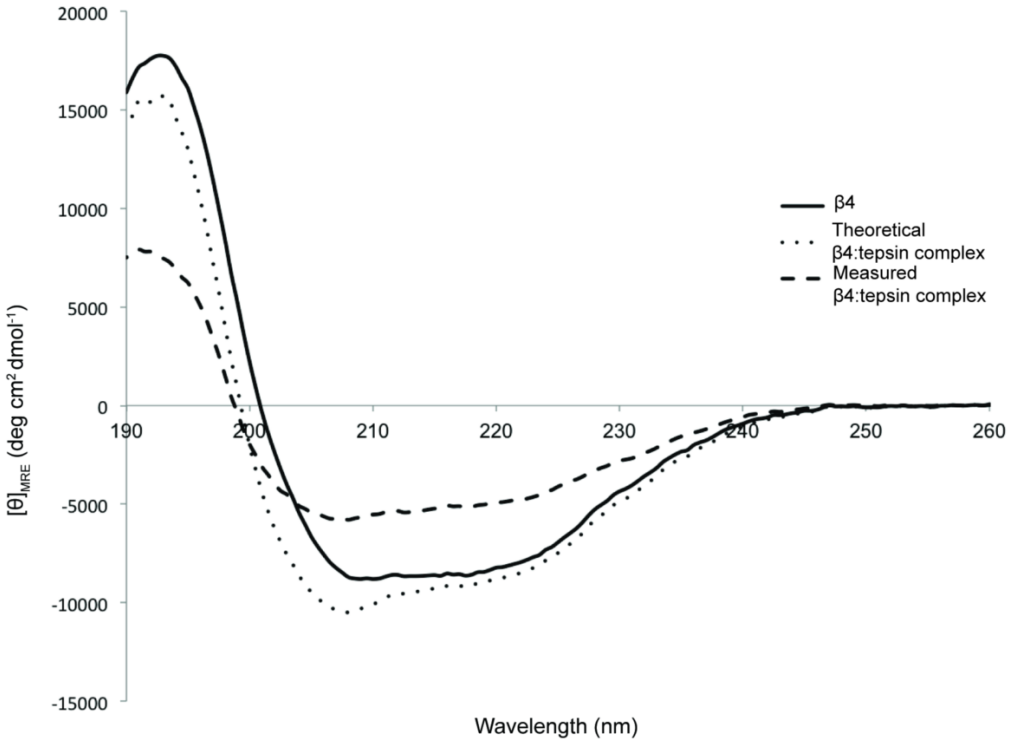
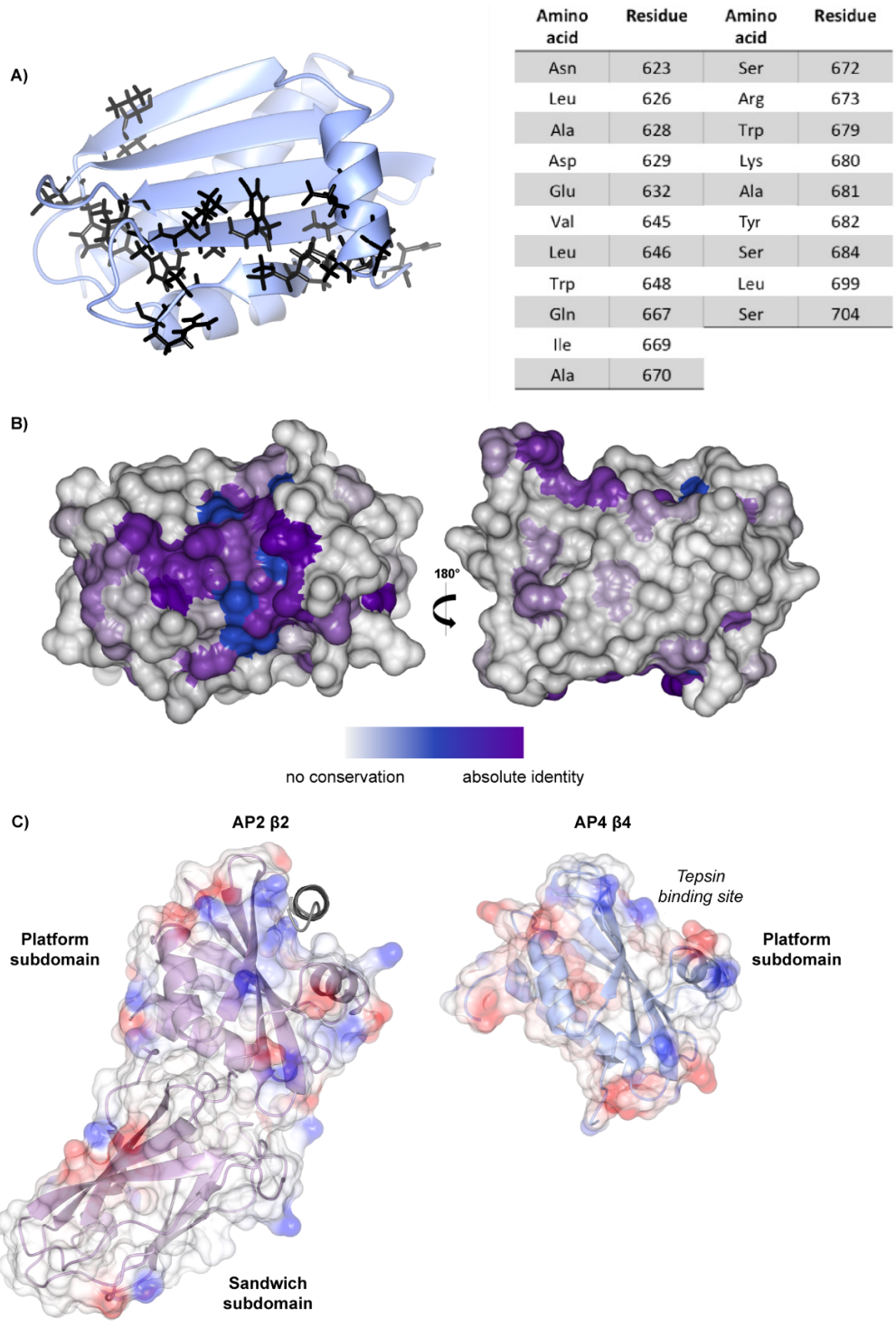


Figure A1-2. The tepsin fragment does not gain helicity upon binding β4. Circular dichroism of β4 alone (solid line) and a 1:1 β4-tepsin complex (long dashed line) indicate there is no additional helicity present in the complex. Characteristic α-helix minima occur at λ=208 and 222 nm; neither wavelength showed increased CD signal, suggesting the tepsin fragment does not form an α-helix when binding β4. The theoretical spectrum for a β4-tepsin complex is shown for comparison (short dashed line).



Amino acid	Residue	Amino acid	Residue
Asn	623	Ser	672
Leu	626	Arg	673
Ala	628	Trp	679
Asp	629	Lys	680
Glu	632	Ala	681
Val	645	Tyr	682
Leu	646	Ser	684
Trp	648	Leu	699
Gln	667	Ser	704
Ile	669		
Ala	670		

Figure A1-3. Structural details of β4 and β2 appendage domains. A) Residues exhibiting large chemical shift perturbations were mapped onto the β4 structure (PDB ID: 2MJ7), shown as a ribbon diagram. The table (right) highlights 20 residues corresponding to these peaks. B) Sequence conservation between β2 and β4 appendage domains across species mapped onto the structure of β4 in ConSurf. This model highlights a highly conserved binding patch between the β2 and β4 appendage domains; the β4 patch contains residues identified by chemical shift perturbation mapping. Left-hand view is equivalent to ribbon diagram in A). Rotating the view by 180° (right-hand view) indicates much less conservation on this side of the protein domain.

C) Left-hand: electrostatic surface and ribbon diagram of the β2 appendage domain (PDB ID: 2G30) with N-terminal sandwich and C-terminal platform subdomains marked. The $[DE]_nX_{1-2}FXX[FL]XXXR$ helical motif found in epsin1, ARH, and arrestin binds a pre-formed groove on the surface of the β2 platform subdomain. Right-hand: equivalent views of electrostatic surface and ribbon diagram of β4 appendage domain, which consists of only the platform subdomain. In contrast to β2, β4 has a flat surface that accommodates the LFxG[M/L]x[L/V] motif identified in tepsin.

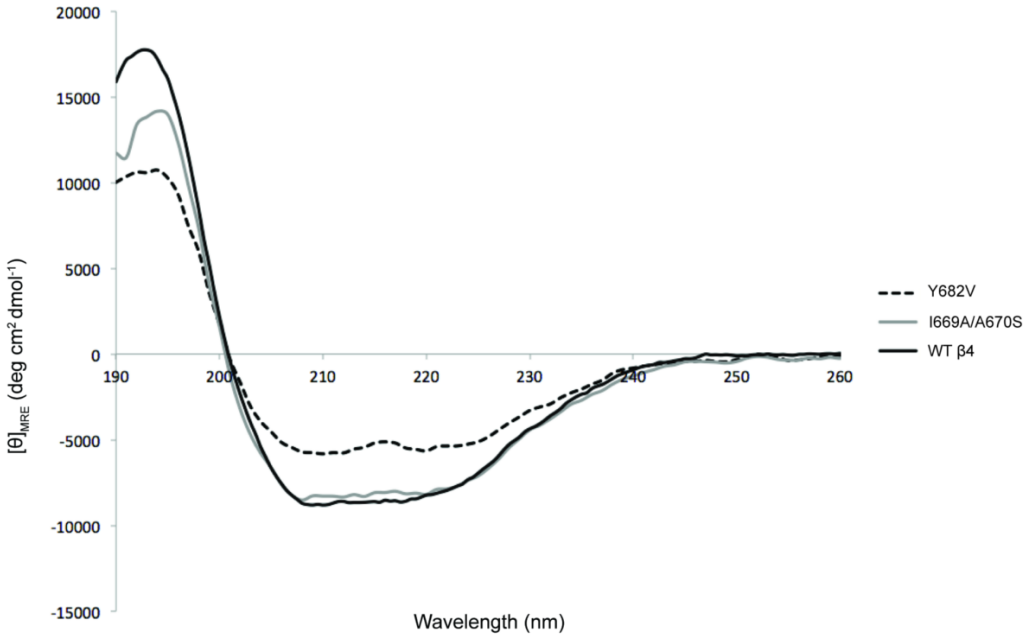


Figure A1-4. β4 point mutants are folded. Circular dichroism spectroscopy comparing wild-type β4 and mutants shows the mutations do not grossly disturb overall protein fold and structure.

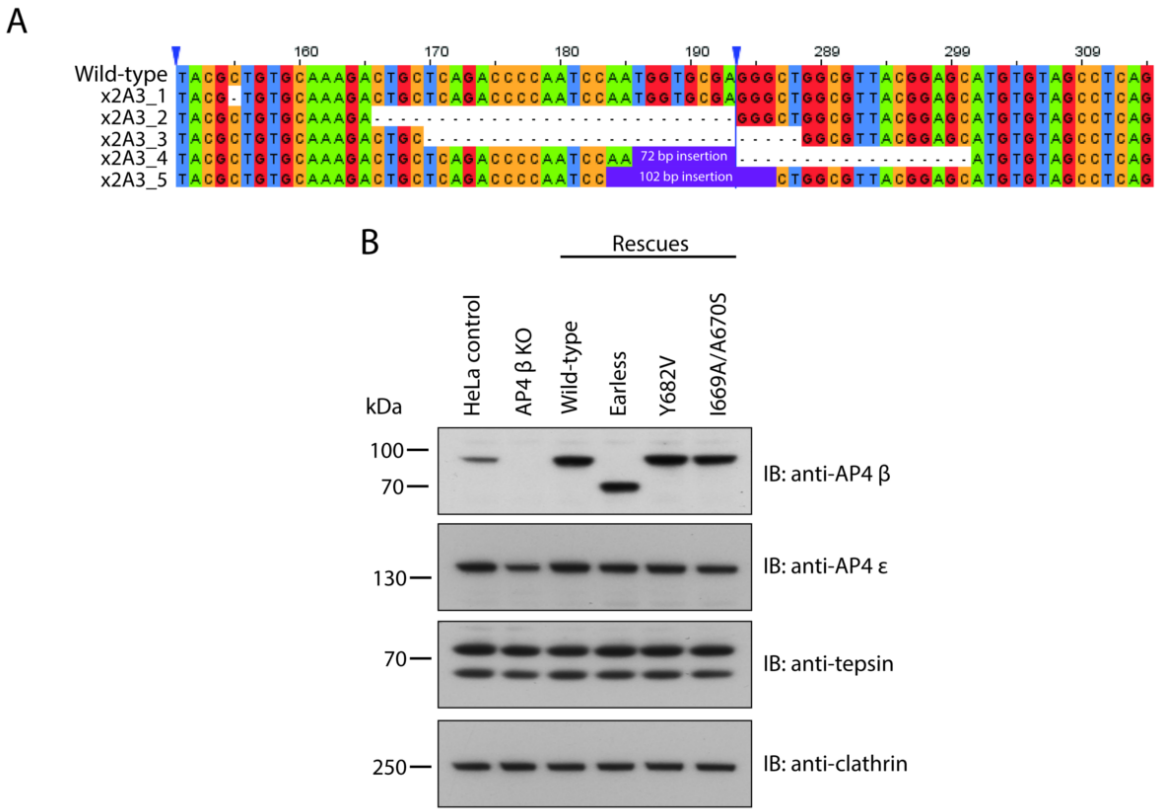


Figure A1-5. Characterisation of AP4 β knockout and β4 rescued cell lines. A) Multiple sequence alignment (Jalview) of wild-type *AP4B1* exon 2 and the mutant sequences recovered from clone x2A3. Allele 1 (1 bp deletion) was recovered five times, allele 2 (28 bp deletion) four times, allele 3 (29 bp deletion) six times, allele 4 (26 bp deletion plus 72 bp insertion) four times and allele 5 (13 bp deletion plus 102 bp insertion) five times. The alignment shows only the area of exon 2 surrounding the mutations (the rest of exon 2 was the same in the mutants as the wild-type sequence) and parts of the insertions in alleles 4 and 5 are hidden from view. The nucleotides are numbered based on the longest allele (allele 5), starting from 1 for the first nucleotide of exon 2. All mutant alleles result in frameshift and the introduction of a premature termination codon in exon 2. B) Western blots of whole cell lysates from control (wild-type HeLa), AP4 β KO, or AP4 β KO cells stably rescued with full-length wild-type or mutant (earless, Y682V or I669A/A670S) β4, probed with antibodies against AP4 β, ε, tepsin and clathrin (loading control). In the rescued cell lines β4 is over-expressed relative to the level of expression in the control wild-type HeLa cell line.

APPENDIX II

Supplementary Materials from Chapter III

tENTH 1-136	
Data collection	
Space group	$C222_1$
Cell dimensions	
a, *b*, *c* (Å)	80.8, 84.8, 80.9
α, β, γ (°)	90.0, 90.0, 90.0
Beamline	Diamond I04-1
Wavelength (Å)	0.92
Resolution (Å)	28.59-1.38 (1.38)
R_{merge}	0.083 (1.042)
I / σ(I)	14.6 (2.7)
Completeness (%)	99.6 (99.8)
Multiplicity	12.1 (12.3)
Refinement	
Resolution (Å)	19.82-1.38
No. reflections	56,820
R_{work} / R_{free}	0.1561/0.1862
No. atoms	
Protein	2021
Ligand/ion	0
Water	178
B-factors	
Protein	23.00
Ligand/ion	-
Water	30.79
R.m.s deviations	
Bond lengths (Å)	0.012
Bond angles (°)	1.05
Ramachandran	
Outliers	0.00%
Allowed	1.93%
Favored	98.07%

tENTH 1-153	
Data collection	
Space group	$P42_11$
Cell dimensions	
a, *b*, *c* (Å)	90.2, 90.2, 42.8
α, β, γ (°)	90.0, 90.0, 90.0
Beamline	LSCAT-21-ID-F
Wavelength (Å)	0.979
Resolution (Å)	50.00-1.92 (1.95-192)
CC ½	0.997 (0.644)
I / σ(I)	16.9 (1.3)
Completeness (%)	97.9 (93.9)
Multiplicity	7.5 (5.7)
Refinement	
Resolution (Å)	45.09-1.8
No. reflections	30,390
R_{work} / R_{free}	0.188/0.222
No. atoms	
Protein	1921
Ligand/ion	0
Water	165
B-factors	
Protein	28.33
Ligand/ion	-
Water	38.15
R.m.s deviations	
Bond lengths (Å)	0.008
Bond angles (°)	0.903
Ramachandran	
Outliers	0.00%
Allowed	1.55%
Favored	98.45%

Table A2-1. Tepsin ENTH domain structure determination. Data collection and refinement statistics for tENTH crystal structures. Values in parentheses represent the highest resolution shell.

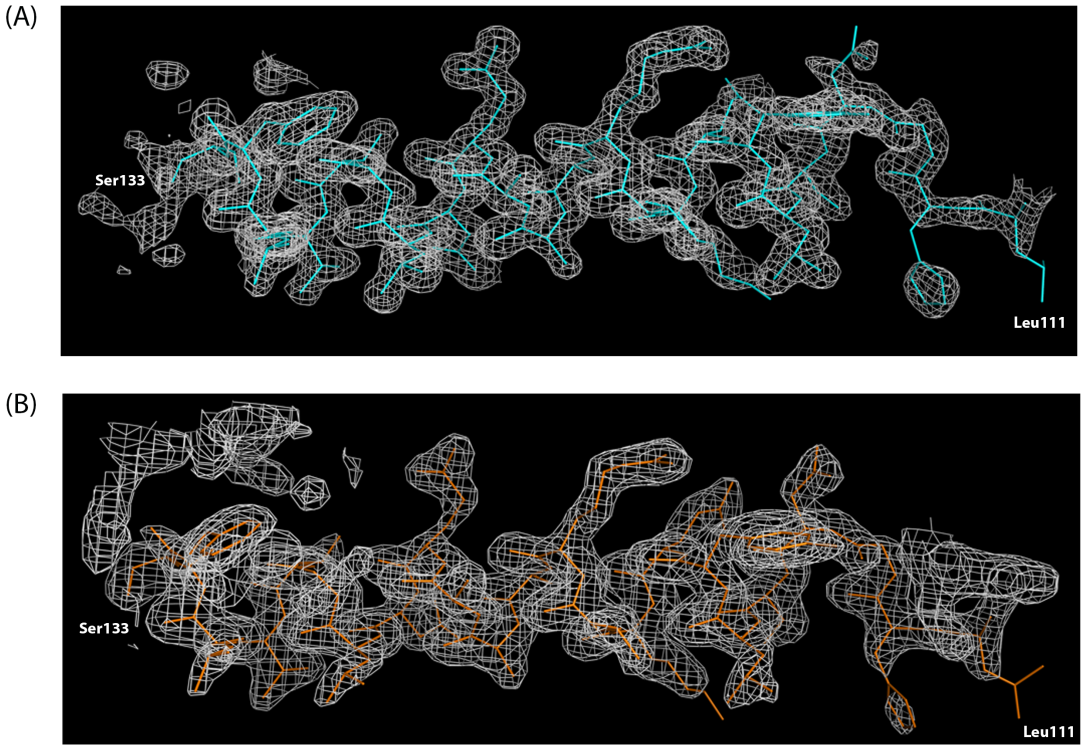


Figure A2-1. The tepsin ENTH domain lacks helix 8. (A) $2mF_{obs}$-mF_{calc} density for tENTH construct 1-136 showing residues Leu111-Ser133 from the final model. (B) Equivalent view for tENTH construct 1-153. No additional density was observed after residue Asp134 in the longer construct, supporting the idea that helix α7 is the final helix in the tepsin ENTH domain.

(A)

tENTH *H. sapiens* vs.	RMSD (Å)	Quality Q score (0-1)	Residues aligned (aa)	PDB ID
ENT1 *S. cerevisiae*	1.44	0.579	114	5LOZ
Epsin1 ENTH *R. norvegicus*	1.52	0.541	116	1EDU
tENTH *A. thaliana*	1.54	0.601	119	1VDY
Epsin1 ENTH *D. rerio*	1.56	0.549	114	5LPO
EpsinR ENTH *H. sapiens*	1.62	0.557	114	2QY7

(B)

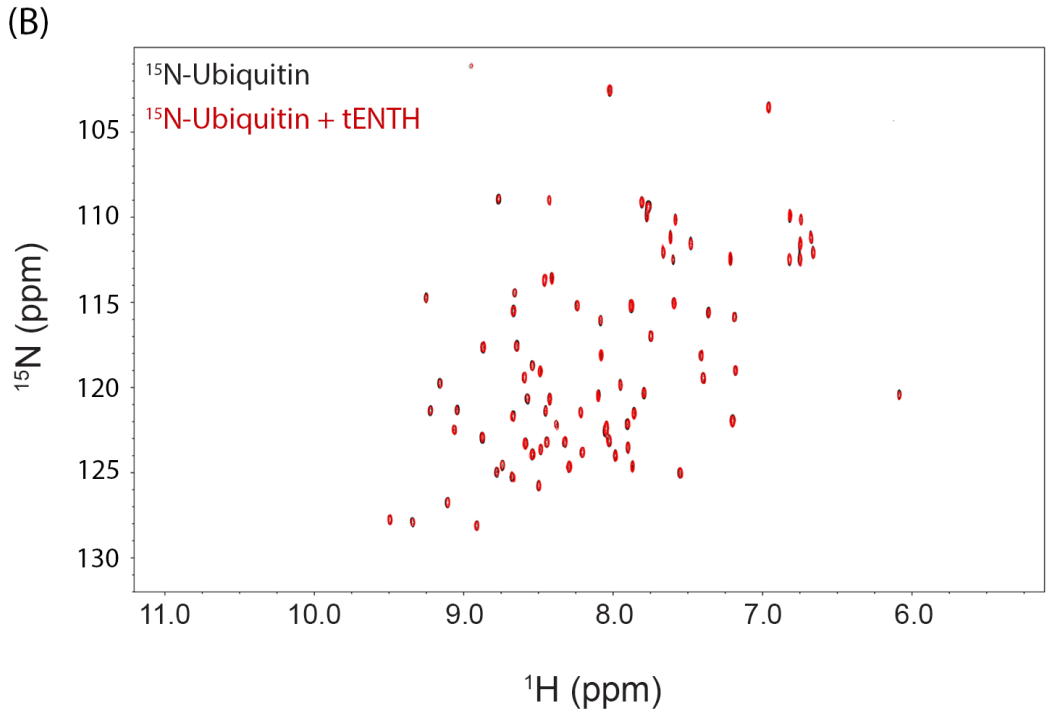


Figure A2-2. Structural and functional comparison of ENTH domains. (A) The tepsin ENTH structure was compared to other ENTH domains found in a variety of species and deposited in the PDB. Analysis was carried out using Superpose (CCP4). (B) HSQC spectrum of ^{15}N-labelled ubiquitin (black peaks) and ^{15}N-labelled ubiquitin with a ten-fold molar excess of tENTH (red peaks). No chemical shifts are observed, implying that tENTH does not bind ubiquitin *in vitro*.

(A)

```
rat           -----SREPGDLAERAEGMAPNDCQQELNLVRTVTQGPRVFLSREETQHFIKECGLLNCE      346
mouse         -----SREPGDLAERAEATPPNDCQQELNLVRTVTQGPRVFLSREETQHFIKECGLLNCE      346
human         ------REPGDLAERVEVVALSDCQQELSLVRTVTRGPRAFLSREEAQHFIKACGLLNCE      279
marmoset      SHSGTSRELGDLAERAEVMALSDCQQELGLVRTVTRGPRAFLSREEAQHFIRACGLLNCE      365
horse         -----SWAPSDLAERAEAMTLSDCQQELSLVQTVTRGSRAFLSREEAQHFVKECGLLNCE      360
cow           ---GASRAPSDLAERVEAVTLSDCQQELSLVRTVTQGQHTFLSREEVQHFVKECGLLNCE      215
                        .*****.*    .******.**:***:* :.******.***:: *******

rat           AVLELLLQQLVGTSECEQMRALCAIASFGSADLLPQEHILLLCRQQLQELGAGSPGPVTN      406
mouse         AVLELLLRQLVGTSECEQMRALCAIASFGSADLLPQEHVLLLCRQQLQELGAGSPGPVTN      406
human         AVLQLLTCHLRGTSECTQLRALCAIASLGSSDLLPQEHILLRTRPWLQELSMGSPGPVTN      339
marmoset      AVLQLLTCHLHGTSECTQLRALCAIASLGSADLLPQEHILLRTRRQLQELSTGSPGPVTN      425
horse         AVLELLICHLGGTSECVQMRALGAVASLGCTDLLPQEHILLLTRPRLQELSAGSPGPVTN      420
cow           AVLELLIRHLAATSERVQMRALGAIASLGCTDLLSQERVLLLTRPRLQELSAGSPGPVTN      275
              ***:**  :* .***  *:*** *:**:*.:*** **::**  *  ****. ********

rat           KATKILRHLEASCGQQF--       423
mouse         KATKILRHFEASCGQQLPT       425
human         KATKILRHFEASCGQLS--       356
marmoset      KATKILRHFEASCGELP--       442
horse         KATKILRHFEASCRQRP--       437
cow           KATKILRHFEASCRQWP--       292
              ********:**** :
```

(B)

Data Collection	tVHS (Native)	SeMet derivative
Beamline	LSCAT-21-ID-G	LSCAT-21-ID-D
Wavelength (Å)	0.97857	0.97910
Space group	P65	P65
Cell dimensions		
a, b, c (Å)	58.828, 58.828, 69.078	58.775, 58.775, 69.522
α, β, γ (˚)	90, 90, 120	90, 90, 120
Resolution (Å)	50.00-1.85 (1.88-1.85)	50.0-1.95 (1.98-1.95)
R_{meas}	0.126 (1.217)	0.138 (0.692)
R_{pim}	0.032 (0.311)	0.035 (0.309)
$I/\sigma I$	25.88 (2.69)	28.14 (2.25)
Completeness (%)	99.96 (100.00)	99.00 (99.73)
Redundancy	15.4 (15.2)	6.4 (2.4)

Refinement	tVHS (Native)	SeMet derivative
Resolution (Å)	41.00-1.85 (1.92-1.85)	29.39-1.95 (2.10-2.02)
Unique reflections	11,617 (1,149)	9904 (989)
R_{work}/R_{free}	0.1909/0.2179 (0.2585/0.3285)	0.1856/0.2189 (0.2650/0.2765)
No. atoms		
Protein	878	868
Water	52	67
B-factors		
Protein	33.27	29.90
Water	38.80	35.92
R.m.s. deviations		
Bond lengths (Å)	0.012	0.007
Bond angles (˚)	1.04	0.93
Ramachandran plot (%)		
Favored	98.25	96.46
Allowed	1.75	2.65
Outliers	0	0.88
Rotamer outliers	0	0
Clashscore	0.56	2.28

Figure A2-3. Tepsin VHS-like domain structure determination. (A) Sequence alignments of tVHS across five species that demonstrate 70-90% sequence identity with human tVHS domain. We cloned, expressed, and purified these constructs to determine the most suitable protein for crystallization. (B) Data collection and refinement statistics for native and selenomethione-labeled horse tVHS domain.

tVHS (Ec) vs.	RMSD (Å)	Quality Q score (0-1)	Residues aligned (aa)	PDB ID
tENTH *H. sapiens*	2.11	0.408	97	-
tENTH *A. thaliana*	2.30	0.428	105	1VDY
EpsinR ENTH *H. sapiens*	1.99	0.432	99	2QY7
Epsin1 ENTH *D. rerio*	2.03	0.382	95	5LPO
Epsin1 ENTH *R. norvegicus*	2.15	0.368	98	1EDU
Epsin1 ENTH *S. cerevisiae*	2.37	0.372	98	5LOZ
Tom1 VHS *H. sapiens*	2.33	0.394	106	1ELK
Tom1like VHS *H. sapiens*	2.34	0.411	104	3RRU
Hrs VHS *D. melanogaster*	2.34	0.393	103	1DVP
GGA3 VHS *H. sapiens*	2.44	0.397	110	1JPL
STAM VHS *H. sapiens*	2.62	0.416	109	3LDZ
CALM ANTH *R. norvegicus*	2.25	0.185	95	3ZYK

Figure A2-4. Structural comparison of tepsin VHS/ENTH-like domain. The tepsin VHS-like domain was compared to both ENTH and VHS domains from a variety of species and depositing in the PDB. Analysis was carried out using Superpose (CCP4). Based on RMSD values and Q-scores, these data suggest tVHS may be more similar to ENTH than to VHS domains.

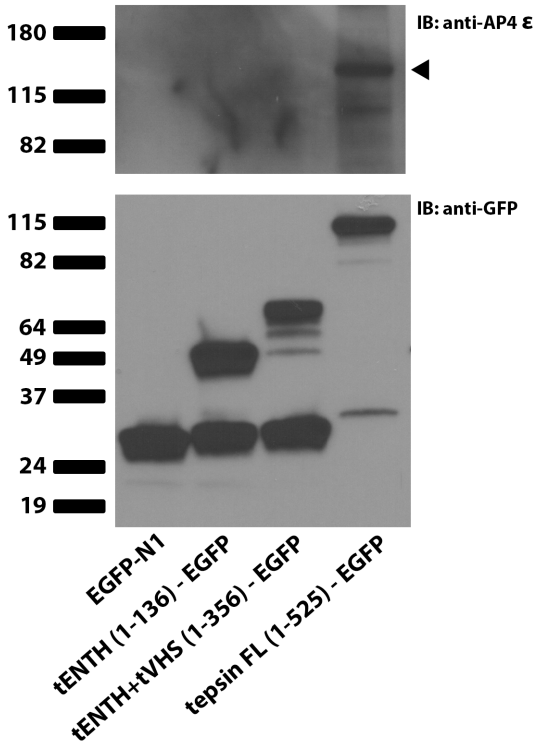


Figure A2-5. The tepsin N-terminus is insufficient for AP4 immunoprecipitation. HeLa cells were transfected with empty vector (EGFP-N1) or tepsin constructs corresponding to the ENTH domain (tepsin-1-136-GFP); both N-terminal domains (tepsin 1-356-GFP); or full-length tepsin (tepsinFL-GFP). Only full-length tepsin efficiently immunopreciptated AP4 from lysates, as judged by immunoblotting against the ε subunit. We were unable to transfect a construct containing on the tepsin VHS domain.

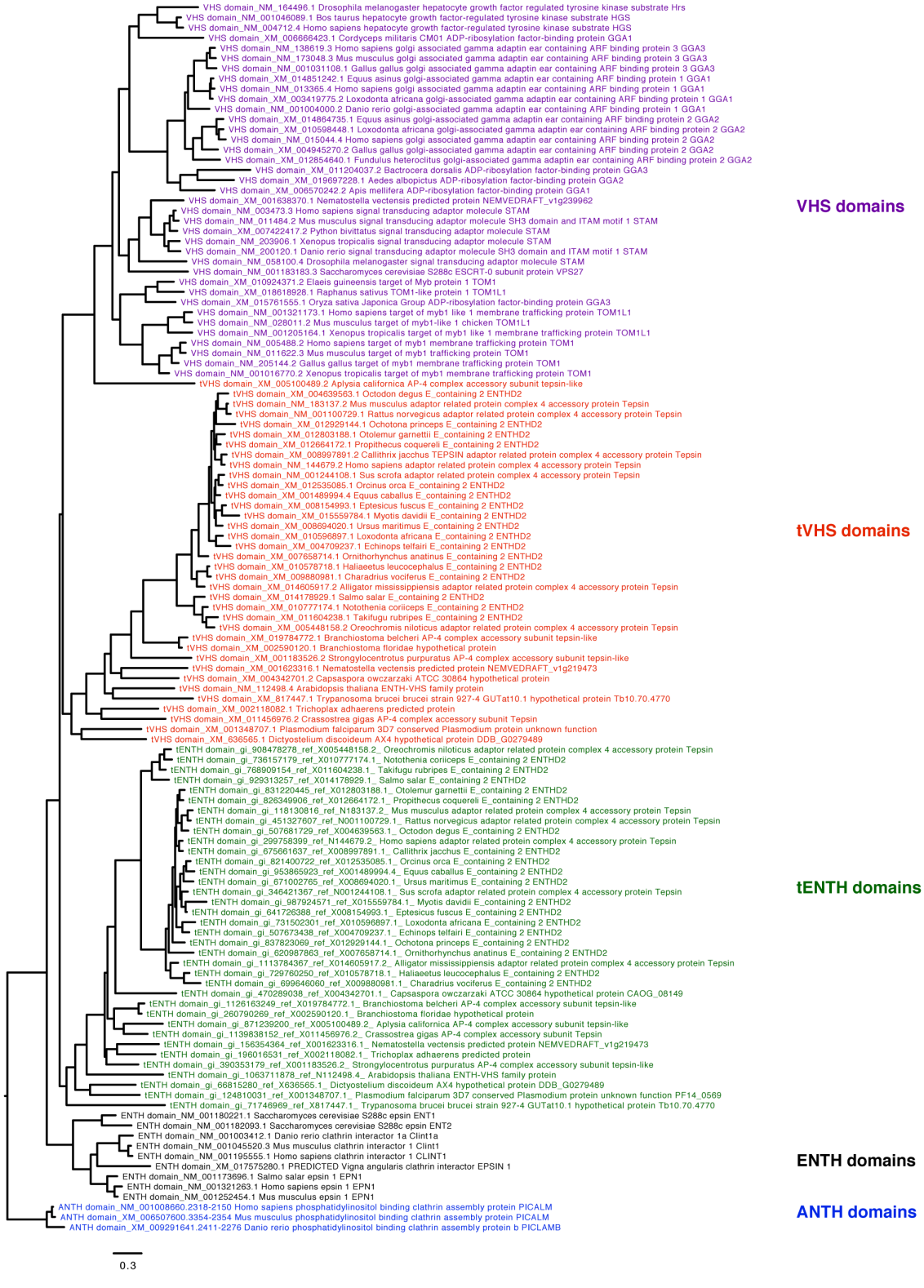


Figure A2-6. Phylogenetic analysis of tepsin ENTH and VHS-like domains. Maximum-likelihood phylogenetic tree of ANTH (outgroup), ENTH, tENTH, tVHS, and VHS domains. Note that the tVHS domains for *Aplysia californica* and *Crassostrea gigas* appear to have diverged

significantly from other tVHS domains and do not reliably form a clade with the other tVHS sequences. The tree with the best maximum likelihood score, pictured here, places these sequences outside of the VHS and tVHS clade. NCBI/GenBank accession numbers are included for reference.

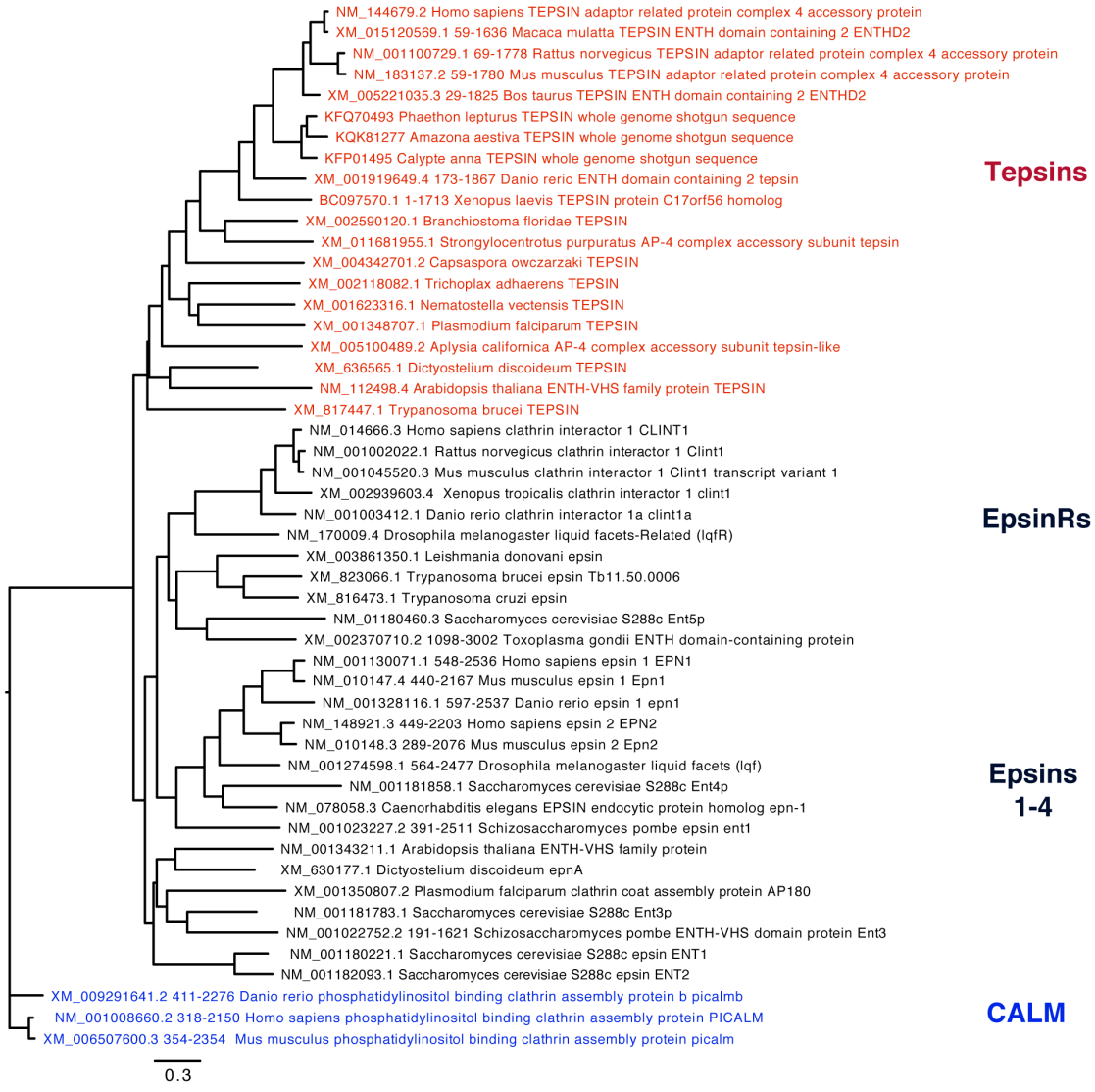


Figure A2-7. Phylogenetic analysis of the epsin family.
Maximum-likelihood phylogenetic tree of tespsin, epsin, and PI-CALM (outgroup) sequences. Sequences annotated as tepsin, epsin, and PI-CALM each form respective monophyletic clades. NCBI/GenBank accession numbers are included for reference.

Domain	Number of sequences	Average percentage of sites at which pairs of sequences differ
ANTH	3	15.93%
ENTH	9	42.57%
VHS	37	52.79%
tENTH	36	42.41%
tVHS	36	47.65%

Table A2-2. Genetic distance between members of different ENTH/ANTH/VHS domains. Genetic distance was calculated for each pair of sequences as the proportion of differences (number of nucleotide differences divided by the number of sites compared). Gaps were not considered in the distance calculation (pairwise deletion of gaps).

Appendix III

Supplementary figures from Chapter IV

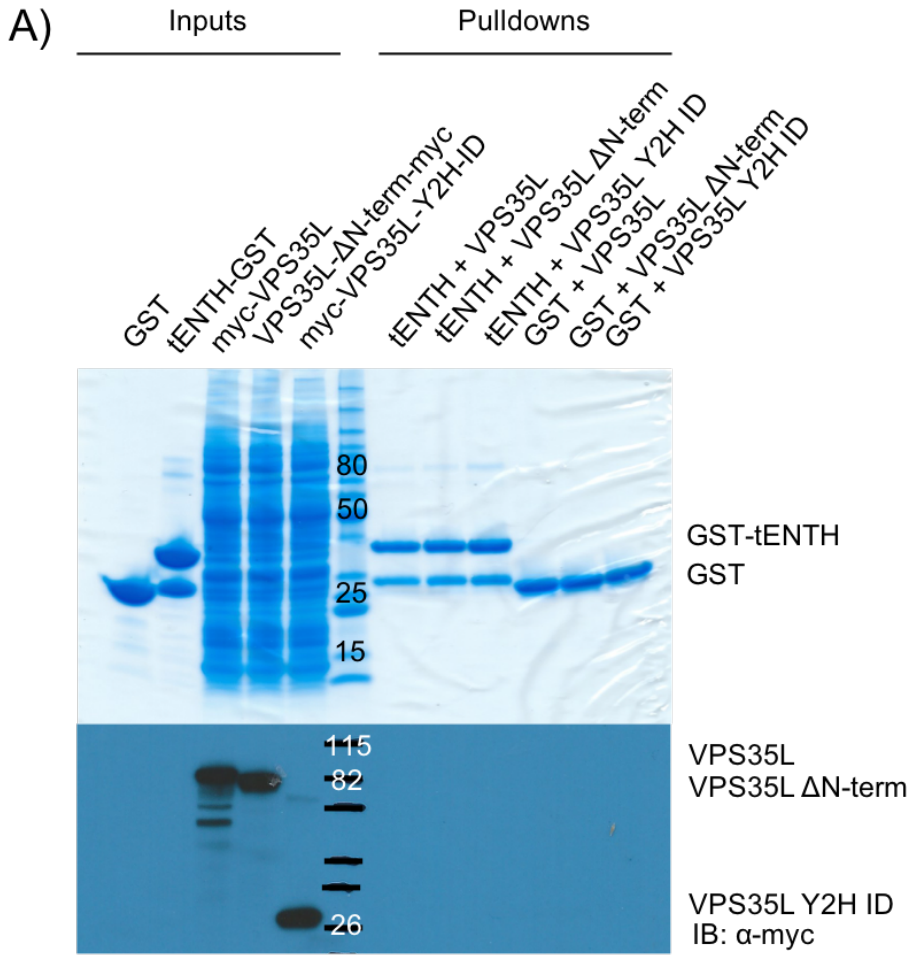


Figure A3-1: No interaction between VPS35L constructs and tENTH was observed when tENTH was C-terminally tagged with GST. SDS-PAGE (top) and α-myc Western blot (bottom).

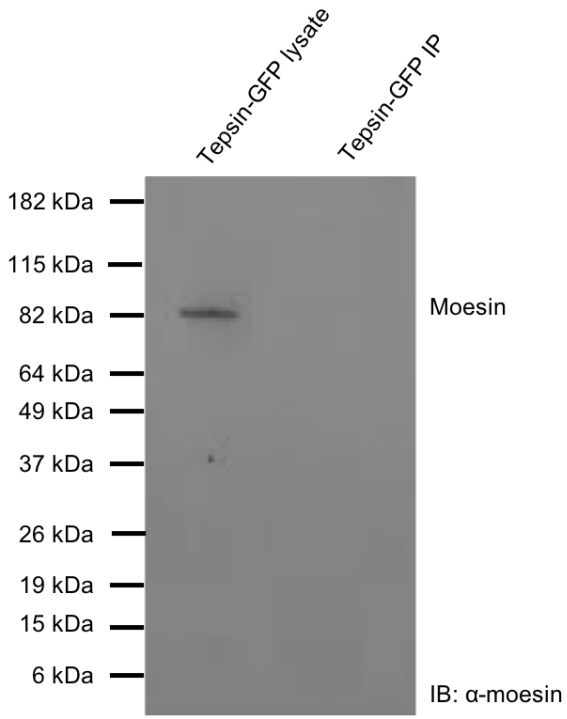


Figure A3-2: Moesin does not immunoprecipitate with tepsin-GFP. The expected molecular weight of moesin is 75 kDa.

	280 nm	*350 nm*	*415 nm*
GST-ISCA1	5.6	0.66	0.41
H6-ISCA1	8.2	0.28	0.04

Table A3-1: Spectroscopic comparison of GST-ISCA1 and H6-ISCA1 following purification. Three wavelengths were measured using a UV spectrophotometer: 280 nm, where protein absorbs; 350 nm, Fe/S cluster absorption/iron bound to protein; and 415 nm, Fe/S cluster absorption (Grzyb *et al.,* 2012). The reported numbers are relative, but show loss of absorbance at the Fe/S related wavelengths in the H6-ISCA1 sample, even though the protein is at higher concentration. n=1.

Appendix IV

Supplementary figures from Chapter V

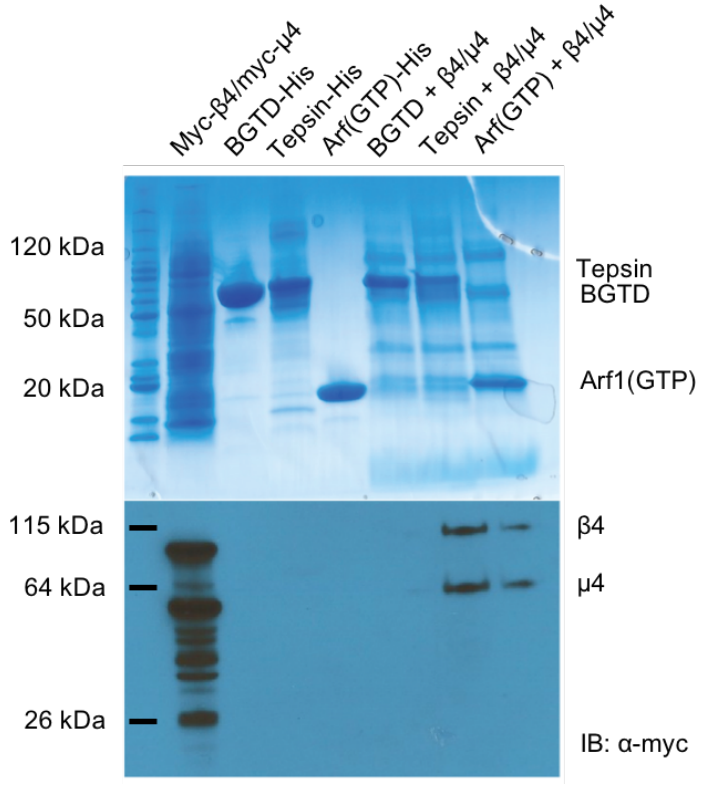


Figure 4-1: Full length β4 and μ4 were successfully expressed using *in vitro* wheat germ translation and shown to interact with tepsin and Arf(GTP) by His-pulldown. SDS-PAGE gel (top) and Western blot for anti-myc (bottom). The *C. difficile* B toxin glucosyl transferase domain (BGTD) was used as a negative His-tagged control for this experiment.

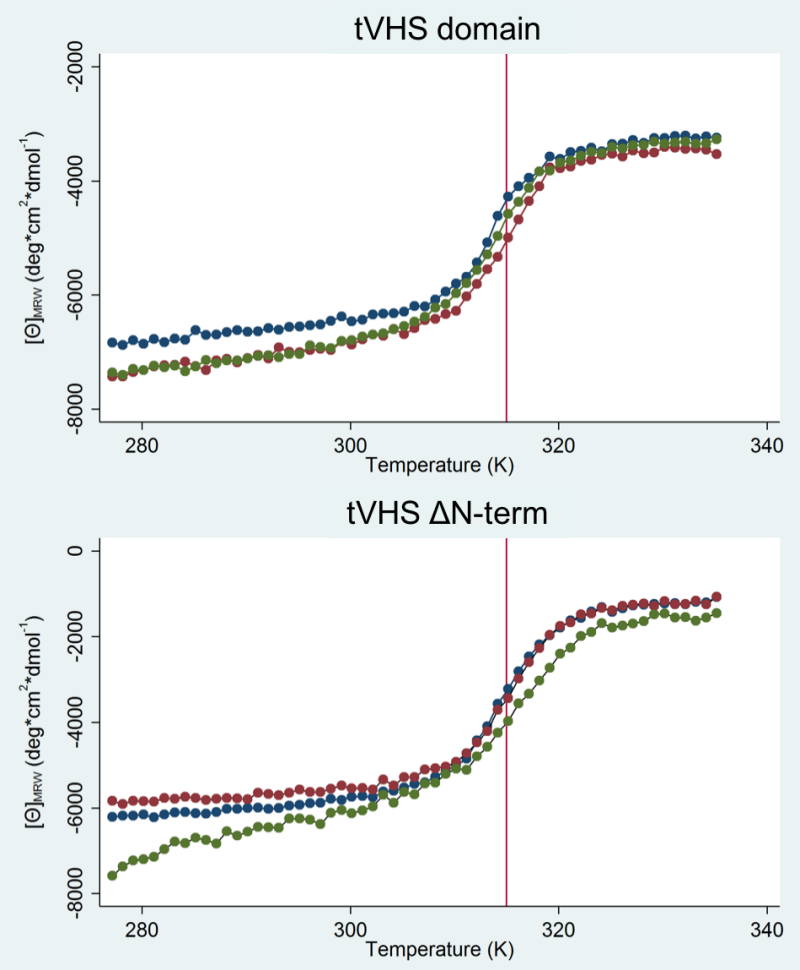


Figure A4-1: Thermal denaturation experiments of our tVHS crystallization construct (top) and tVHS domain without the putative N-terminal disordered helix (tVHS ΔN-term; bottom), show no differences in T_M. Figure adapted from image provided by Anderson Monken.

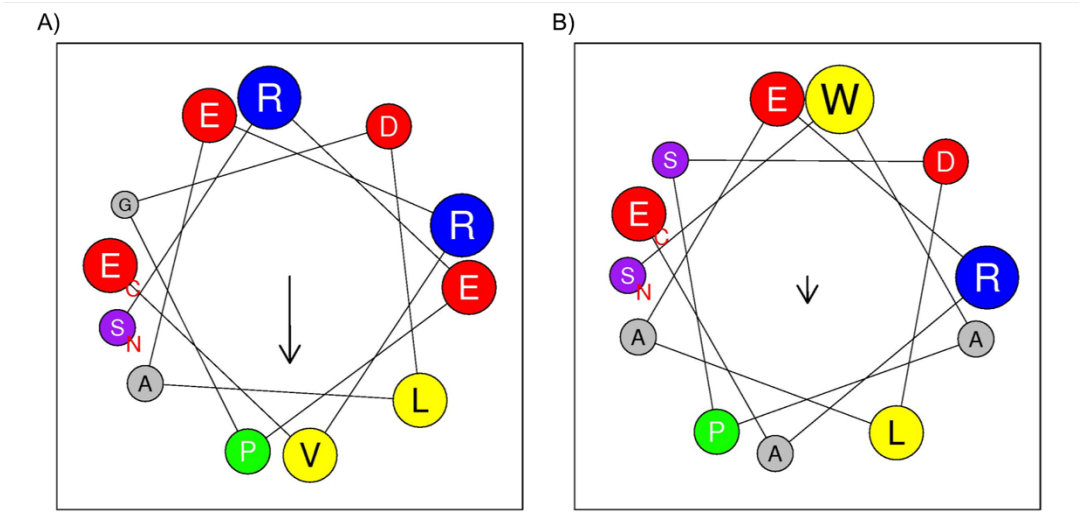


Figure A4-2: Helical wheel plots (Heliquest; Gautier *et al.,* 2008) of the possible tVHS N-terminal disordered helix using the human sequence (A) or horse sequence (B). Neither helical plot has a hydrophobic face, although the hydrophobic moment (μH) is larger for human compared to horse, as denoted by the size of the arrow in the panels. The size of the amino acid corresponds to the volume of the residue; N- and C-terminals are labeled with red letters.