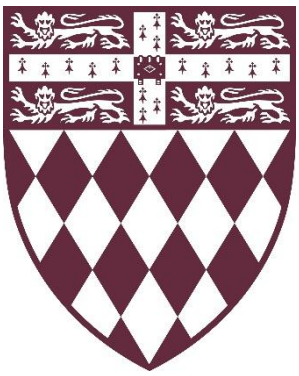
UNIVERSITY OF CAMBRIDGE
School of Clinical Medicine

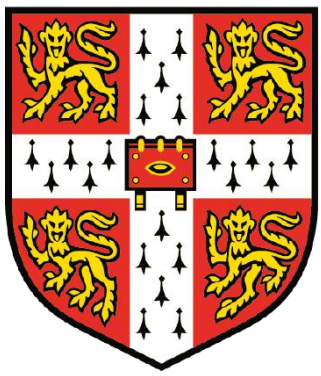
VIRAL SUBVERSION OF HOST CELL MEMBRANE TRAFFICKING

Julia Muenzner

Fitzwilliam College, University of Cambridge

June 2017





Supervisor: Dr. Stephen C. Graham

Department of Pathology, Division of Virology

University of Cambridge

Tennis Court Road

Cambridge CB2 1QP

This dissertation is submitted for the degree of Doctor of Philosophy.

Viral subversion of host cell membrane trafficking

Julia Muenzner

Enveloped viruses acquire their membrane coat from the plasma membrane or intracellular organelles and rely on cellular machinery to facilitate envelopment and egress of virus progeny. This thesis examines egress-related interactions between host cell factors and proteins of two different enveloped viruses: hepatitis D virus (HDV) and herpes simplex virus 1 (HSV-1). HDV is a small RNA virus causing fulminant hepatitis or severely aggravating cirrhosis and hepatocellular carcinoma. HSV-1 is a large DNA virus infecting epithelial and neuronal cells. Infection with HSV-1 not only triggers the development of recurring sores on oral or genital mucosa, but can also cause severe disease in neonates and immunocompromised patients.

The interaction between the large antigen of HDV (HDAg-L) and the N-terminal domain (NTD) of clathrin, a protein crucial for endocytosis and intracellular vesicular trafficking, was examined by structural, biochemical and biophysical techniques. Co-crystal structures of NTD bound to HDAg-L peptides derived from different HDV genotypes revealed that HDV interacts with multiple binding sites on NTD promiscuously, prompting re-evaluation of the binding between cellular peptides and NTD. Surprisingly, co-crystal structures and pull-down capture assays showed that cellular peptides containing clathrin-binding motifs can also bind multiple sites on the surface of NTD simultaneously. In addition, the structures of viral and cellular peptides bound to NTD enabled the molecular characterization of the fourth peptide binding site on NTD, the “Royle box”, and led to the identification of a novel binding mode at the “arrestin box” peptide binding site on NTD. The work in this thesis therefore not only identifies the molecular basis of HDV:clathrin interactions, but also furthers our understanding of basic clathrin biology.

Even though many HSV-1 proteins have been implicated in the envelopment and egress of viral particles, only few interactions between HSV-1 and cellular proteins promoting these processes have been described. Therefore, the HSV-1 proteins gE, UL21 and UL56 were selected and characterized bioinformatically and/or biochemically. Cellular proteins interacting with UL56 were identified by yeast two-hybrid screening and quantitative mass spectrometry. Co-immunoprecipitation and pull-down experiments confirmed the Golgi-trafficking protein GOPC, components of the mammalian trafficking protein particle complex, and the ubiquitin ligase NEDD4 as novel binding partners of UL56, thereby suggesting exciting new avenues for the investigation of cellular mechanisms contributing to HSV-1 envelopment and egress.

Für Ferdinand

Bleib' wissensdurstig!

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. In accordance with the Biology guidelines, this thesis does not exceed 60,000 words excluding bibliography, figure legends, tables and appendices.

Signed:

Date:

Julia Muenzner BSc MRes

Acknowledgements

First and foremost, I am deeply grateful to my supervisor Dr. Stephen Graham for giving me the opportunity to join his research group and work on this project. Stephen has been an inspiring teacher, freely sharing his knowledge and spreading his enthusiasm about science (selflessly, even through sick leave!). Without his encouragement, patience and mentorship, this PhD would not have been possible. I am especially indebted to Stephen for providing exceptional support when I was on maternity leave and then came back to the lab with a small baby at home - I am sure not every supervisor would have handled such a situation as professionally and with such understanding as him. Thanks a lot for making my time in the lab both enjoyable and challenging and the lab outings so much fun!

Furthermore, I thank my friends and colleagues in Stephen's lab and in the Division of Virology, especially Danielle, Edward, Morag, Chen and Tim for creating such a friendly and collaborative atmosphere at work and for great hours spend together away from the bench. Thanks are also due to the staff supporting the work in the Division, in particular to Susanna Colaco, Dr. Heather Coleman, Deb Walsh and Emily-Jane Brand, as well as Giles Lewis in CIMR for running the crystallography facility, and the staff at Diamond Light Source for their support during beamtime. I am also grateful to Fitzwilliam College for general support and for financial assistance of my attendance at numerous conferences and workshops.

I would also like to thank our collaborators Dr. Bernard Kelly, Dr. Colin Crump, Dr. Janet Deane, Dr. Edward Emmott and Prof. Linton Traub for their time, advice and ideas. I especially enjoyed the joined meetings with the Deane group and hope the members of both our and their lab will forgive my endless questions.

I would not have been able to embark on and to complete this PhD without the encouragement and support of my family, and I am especially grateful to my mum who took excellent care of my little boy whilst I hid in the library to write up. Thanks also to my good friends Jonas, Martin M. and Zenobia for trying to understand my experimental struggles and for providing an understanding ear.

Last but not least, I would like to thank my amazing partner Martin for his unconditional love, honesty and never-ending patience, for listening to my late night experimental ideas (and that

after an hour of cello practice), motivating me after setbacks, and actually reading my "sciency" thesis as a scholar of Medieval Jewish literature. Thank you also for supporting me during distressing personal events and for being a fantastic father who makes our time with Ferdinand so awesome (and that includes reciting Latin poems during Ferdinand's late evening wake ups – “I'll be there soon, just one more sentence!”). Ferdinand, thanks for all your little smiles and big grins, and thanks for being so patient when Mami had to finish her thesis.

This work was funded by the Wellcome Trust and the School of Clinical Medicine of the University of Cambridge.

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 Overview

From viral entry and intracellular transport to replication, assembly and egress — viruses exploit host cells at various points during their life cycle to achieve efficient production and spread of progeny particles. Most non-enveloped viruses cause cell lysis to facilitate release from the host cell, but enveloped viruses have to acquire a membrane coat and be directed to the cell surface to ensure the release of virions. Many enveloped viruses, for example influenza virus, measles virus or human immunodeficiency virus, bud from the plasma membrane into the extracellular space, whereas other enveloped viruses such as Dengue virus or herpes simplex virus bud into intracellular compartments to obtain an envelope (5). In the latter case, enveloped viral particles have to be transported in secretory vesicles to the plasma membrane, where fusion of these transport vesicles with the plasma membrane leads to the release of the virus progeny. For almost all viruses, cellular machinery facilitates these budding events (6), and trafficking of capsids, envelope components and virus-filled transport vesicles is likewise mediated by host cell proteins (7). Study of virus:host interactions involved in the envelopment and egress of viral particles therefore does not only shed light on the final stages of virus life cycles, potentially identifying new drug targets for treatment of often debilitating diseases, but also furthers our understanding of cellular membrane functions and trafficking pathways.

In this thesis, I investigate egress-related interactions between host cell factors and two enveloped viruses: hepatitis D virus (HDV) and herpes simplex virus 1 (HSV-1). I sought to identify the molecular basis of the interaction between the large antigen of HDV (HDAg-L) with the N-terminal domain (NTD) of clathrin and compared it to binding between clathrin and cellular proteins. My results do not only provide a structural explanation for the interaction between HDAg-L and NTD, but they also prompted me to revisit the basic molecular determinants of cellular clathrin recruitment. I studied protein:protein interactions of HSV-1 tegument and envelope proteins, and present hypotheses suggesting that HSV-1 extensively remodels the protein architecture of intracellular membranes to promote viral envelopment and/or egress.

In this chapter, I first introduce the cellular endomembrane system with a particular focus on clathrin-mediated trafficking and ESCRT-mediated budding (section 1.2). I discuss two cellular clathrin-interacting proteins, adaptor protein 2 and amphiphysin (section 1.2.3), as well as the structure and the interaction sites of clathrin NTD (section 1.2.4). Next, I highlight the importance of clathrin-mediated trafficking and ESCRT-mediated budding for viruses by providing some examples of viruses hijacking these pathways (section 1.3). I then give an overview of the general virology and life cycles of HDV (section 1.4) and HSV-1 (section 1.5), presenting the current state of knowledge on the egress mechanisms employed by these viruses. Lastly, I introduce the viral proteins studied in this thesis: HDV protein HDAg-L (section 1.4.4) and HSV-1 proteins UL56, gE and UL21 (section 1.5.4, 1.5.5 and 1.5.6).

1.2 The cellular endomembrane system

1.2.1 Organelles and trafficking pathways

One hallmark of eukaryotic cells is their compartmentalization into membrane-confined organelles. Compartmentalization creates different biochemical milieus within the cell, thus allowing the co-existence of different catabolic or anabolic reaction environments, enables enrichment and storage of molecules, and provides means for directed uptake of extracellular and release of intracellular material (8). The classical organelles of the eukaryotic endomembrane system are the endoplasmic reticulum (ER), the Golgi apparatus, early, late and recycling endosomes, endolysosomes, and lysosomes (Figure 1). Material is transported between these organelles and to/from the plasma membrane by vesicles. Commonly, two main routes of intracellular trafficking are described. The endocytic pathway regulates the uptake of extracellular material into vesicular or tubular structures, delivering it to endosomal compartments for recycling, or to lysosomal compartments to mediate degradation (9). The secretory pathway controls the transport of proteins and lipids from the ER, via the cis- and trans-Golgi apparatus towards the plasma membrane (10). However, one should not think of

these pathways as linear or "one-way-streets" since extensive exchange and crosstalk between the compartments takes place (10-12) (Figure 1).

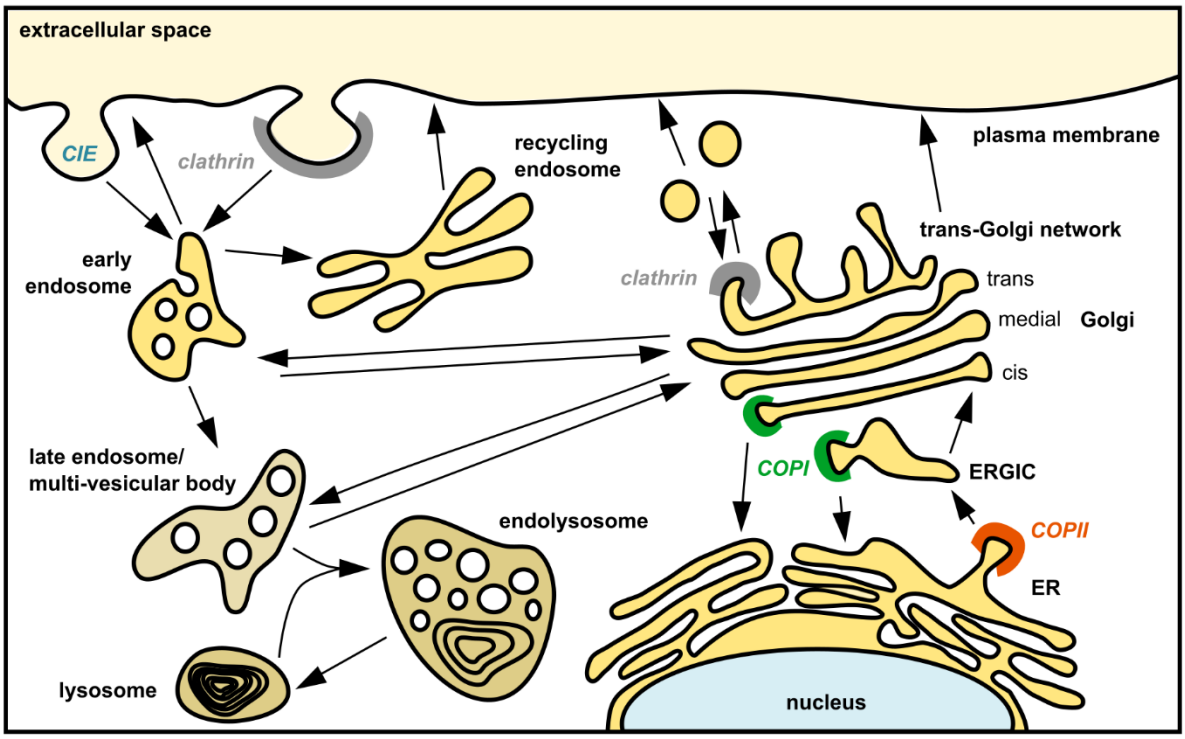


Figure 1 The endomembrane system of a eukaryotic cell. Cellular organelles are labeled (ER: endoplasmic reticulum, ERGIC: ER-Golgi intermediate compartment). The three main vesicle coats are depicted at the membranes they assemble on (clathrin: grey, COPI: green, COPII: orange). Selected vesicular trafficking pathways are indicated with arrows. CIE: clathrin-independent endocytosis.

Vesicular transport occurs in four tightly regulated steps: budding from a membrane, directed or diffuse transport, tethering to the target compartment, and the final fusion of the vesicle membrane with the target membrane (13). During budding, vesicles acquire a structured protein coat that helps both to select cargo and shape the vesicle, although coat-independent budding mechanism such as clathrin-independent endocytosis or micropinocytosis have been described as well (12, 14). One of the most important regulators of intracellular transport are the Rab proteins. Over 60 small GTPases known as Rab proteins are present in eukaryotic cells (15, 16). Switching between GTP- and GDP-bound states and associating with effector proteins, these Rab proteins can direct cargo molecules into transport vesicles, convey target compartment specificity, and even interact with motor proteins to facilitate movement of vesicles along the cytoskeleton through the cytosol (13, 15-17). Large multi-protein complexes such as HOPS, CORVET or TRAPP tether the vesicles to the target compartment, and also contribute to the specificity of

fusion events (13). Fusion of the tethered vesicle with the target membrane is mediated by the interaction of SNARE proteins present on both the vesicle and the acceptor membrane, which form highly organized helical bundles and thereby pull the vesicle towards the target compartment, allowing lipid bilayer mixing and fusion of the lumen of the two compartments (18).

1.2.2 Principle mechanisms and topology of membrane budding

Notably, budding of vesicles into intracellular compartments or at the plasma membrane into the extracellular space can be viewed as budding *away* from the cytosol. Therefore, these budding events are fundamentally different from the ones occurring during endocytosis or during formation of intracellular transport vesicles, which are directed *towards* the cytosol (Figure 2). Consequently, the cellular machineries involved in these two types of budding differ considerably.

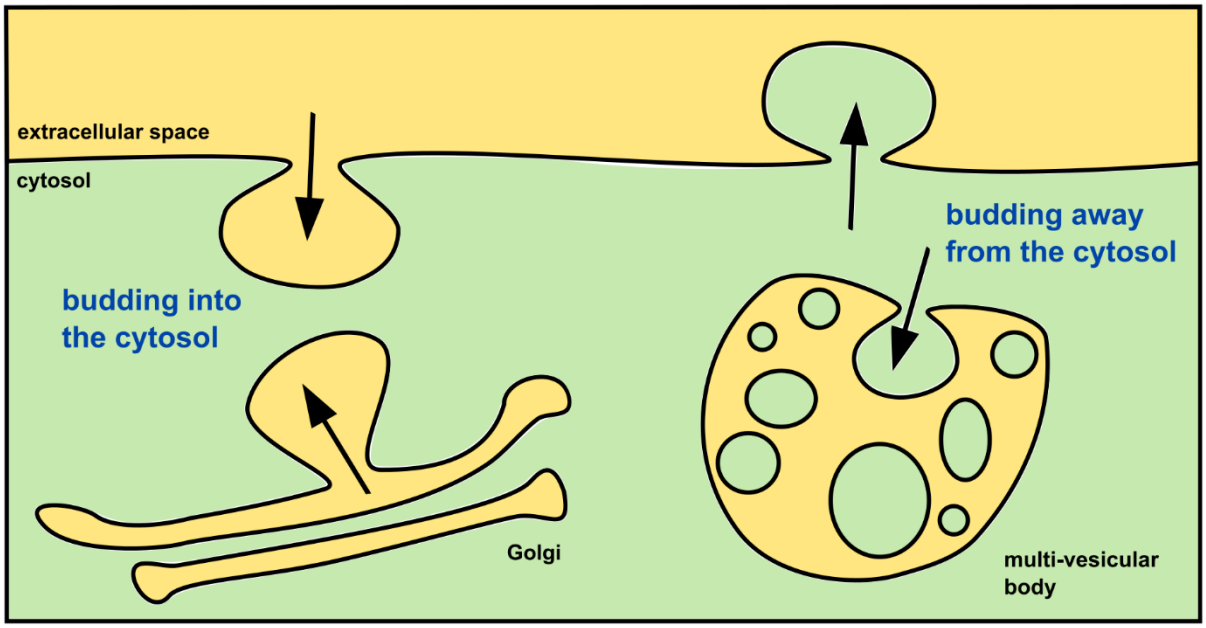


Figure 2 Topology of budding events. Formation of endocytic vesicles or budding from intracellular compartments to form transport vesicles occurs through budding into the cytosol (left), whereas formation of multi-vesicular bodies or budding into the extracellular space is directed away from the cytosol (reverse topology budding, right).

Although membranes can form buds in the absence of any proteins due to lipid phase separation (19), most cellular budding events involve a complex interplay between membrane lipids and proteins at the budding vesicle (14, 20). Some proteins, such as amphiphysin, arfaptin2 or RICH1 (21, 22), insert an amphipathic helix into the membrane, thus inducing local changes in membrane curvature. Transmembrane proteins can potentially intrinsically bend membranes due

to their conical shape (23, 24). Crowding of proteins (25, 26) and incorporation of intrinsically disordered proteins with large hydrodynamic radii lead to increased steric pressure (27), thus driving membrane budding. Another important mechanism of inducing and/or stabilizing membrane curvature towards the cytosol is the interaction of Bin/amphiphysin/Rvs (BAR) domain proteins with membranes. These domains form crescent- or banana-shaped dimers that bind to membranes via electrostatic interactions, thus deforming membranes to match the shape of bound molecule (22, 28). However, probably the best characterized mechanisms of membrane budding into the cytosol involve the formation of a complex protein scaffold called a vesicle coat around membrane invaginations (9, 12, 29). Three main vesicle coats have been described: the COPI coat, regulating vesicular transport within the Golgi and retrograde transport from the Golgi to the ER (30), the COPII coat, mediating trafficking from the ER to the Golgi or ER-Golgi intermediate compartments (31), and the clathrin coat, involved in endocytosis and intracellular trafficking (9, 32, 33).

Membrane budding *away* from the cytosol, also described as "reverse topology budding", is mediated by the cellular endosomal sorting complexes required for transport (ESCRT) machinery (34). Originally discovered to facilitate the formation of multi-vesicular bodies (35), ESCRT proteins also mediate reverse topology membrane budding events observed during cytokinetic abscission, formation of exosomes, reformation of the nuclear envelope, and during budding of at least 40 different viruses, including HIV-1 and other retroviruses, Ebola virus, rabies virus, and herpes simplex virus 1 (HSV-1) (34, 36-38).

1.2.3 Example of budding into the cytosol: clathrin-mediated trafficking

Clathrin-mediated trafficking is an essential cellular process, as for example inactivation of adaptor protein 2 (39) or knock-out of dynamin (40) are embryonic lethal. Clathrin-mediated endocytosis is key to the stimulated and selective internalization of metabolic receptors such as transferrin receptor, to the regulation of cell signaling by trafficking of signaling receptors, and for synaptic vesicle recycling (41). Intracellular (non-plasma membrane-located) clathrin vesicle budding has been investigated much less than clathrin-mediated endocytosis. Nevertheless, it has been shown to participate for example in trafficking from the late endosomes to the trans-Golgi network (42) and in rapid endocytic vesicle recycling (43).

Clathrin is a triskelion-forming protein that can multimerize and forms lattices or cage-like structures in vivo and in vitro (44-46). The protein was first purified by Barbara Pearse in 1975, who named it after its clathrate-like appearance observed in her vesicle preparations (47). Each triskelion contains three large 180 kDa molecules, the so-called clathrin heavy chains, which associate with three clathrin light chains (44). Clathrin heavy chains consist of an N-terminal domain (NTD), which serves as the main interaction hub for binding to clathrin adaptors (see section 1.2.4), and a long α-solenoid region containing eight motifs of ten helices connected by loops (48, 49). These repeated motifs form the so-called ankle, distal leg, knee, and the proximal leg of a clathrin monomer (49). The C-terminal domain of clathrin heavy chain molecules mediates its trimerization (50). During assembly of clathrin cages, interactions between the proximal and distal legs of different clathrin triskelia lead to the formation of the typical pentagons and hexagons "coating" a vesicle (50, 51) (Figure 3A).

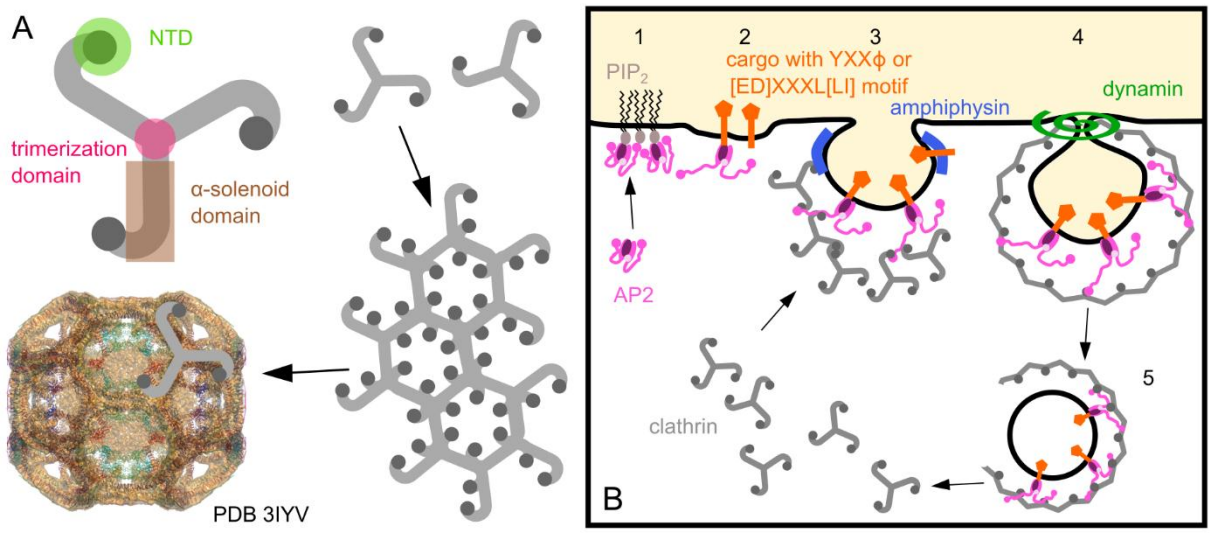


Figure 3 Assembly of clathrin cages and process of clathrin-mediated endocytosis. (A) Cartoon of a clathrin triskelion. Three clathrin heavy chains, each containing a trimerization domain (pink), the N-terminal domain (NTD, green) and the α-solenoid domain (brown), and three clathrin light chains (not pictured) assemble to form a trimeric complex. Multiple triskelia assemble into lattices and form cages around transport vesicles. A full clathrin cage is depicted (PDB 3IYV, brown), and a single clathrin triskelion highlights the positioning of triskelions in this cage. (B) Simplified schematic of the steps of clathrin-mediated endocytosis. 1: Adaptor protein AP2 binds to PIP_2, starting the nucleation of a membrane pit. 2: AP2 binds to endocytic motifs of cargo proteins such as the YXXΦ motif or the acidic dileucine motif [ED]XXXL[LI]. 3: Clathrin is recruited to the site of budding via interaction with clathrin-binding motifs present in various adaptor proteins, such as AP2 or amphiphysin. Amphiphysin binds to both curved membranes and clathrin, further stabilizing the formation of a clathrin-coated vesicle. 4: Scission of the vesicle from the plasma membrane is mediated by dynamin. 5: The budded vesicle is uncoated inside the cytosol.

Clathrin coats assemble around membranes during endocytosis at the plasma membrane and during vesiculation events at intracellular organelles (33, 41). Five steps of clathrin-coated vesicle formation have commonly been described: nucleation of a membrane pit, selection of cargo, assembly of the clathrin coat, dynamin-dependent scission of the coated vesicle from the membrane, and uncoating of the trafficking vesicle (32, 41) (Figure 3B). Dozens of proteins have been shown to interact with the invaginating membrane, cargo molecules, or clathrin itself (32, 41, 52-54), and a comprehensive review of these factors is beyond the scope of this thesis. However, two key players, the adaptor protein AP2 and the BAR domain-containing protein amphiphysin shall be introduced here.

Assembly polypeptide 2, also named adaptor protein 2 (this name is used throughout this thesis) and abbreviated as AP2, is one of five adaptor protein complexes (55) involved in sorting of cargo into vesicles for intracellular transport. It consists of four subunits: two large subunits (α and β2), also known as α-adaptin and β-2adaptin, one medium subunit (μ2), and one small subunit (σ2) (32, 54). The N-terminal "trunk" domains of subunits α and β2, together with the μ2 and σ2 subunits, form the core of AP2, to which the appendage domains of α and β2 are connected via flexible hinge regions (56) (Figure 4A). The C-terminal domain of μ2 (c-μ2) subunit interacts with YXXΦ motifs ("X" denoting any amino acid, and "Φ" a bulky hydrophobic amino acid) present in cytosolic domains of cargo proteins (57), and the σ2 domain recognizes so-called acidic di-leucine motifs, [ED]XXXL[LI], likewise mediating endocytic cargo selection (58, 59). The α appendage domain binds to DP[FW] and FXDXF motifs of accessory proteins such as auxilin or Eps15 (54), and the appendage domain and hinge region of AP2 β2 interact with clathrin (60, 61). The β2 hinge region contains a canonical clathrin-binding motif (CBM), LLNLD, which mediates binding to the N-terminal domain of clathrin (62) (see section 1.2.4). It has been shown that interaction of both the β2 appendage and hinge domain to clathrin are necessary for tight binding (63). The AP2 complex localizes to the plasma membrane (64, 65), where it binds to phosphatidylinositol 4,5-bisphosphate (PIP_2) (66, 67). Binding to PIP_2 triggers large conformational changes in AP2, thereby exposing the binding sites on μ2 and σ2 for cargo proteins (56, 58) and the clathrin-binding motif of the β2 hinge region (68). These conformational rearrangements trigger both the loading of cargo into vesicles and facilitate the recruitment of clathrin triskelia to the site of budding, identifying AP2 as a crucial regulator of clathrin-mediated endocytosis.

Humans possess two amphiphysin paralogues. Amphiphysin was first identified in brain tissues (amphiphysin 1) as a component of synaptic vesicles (69), but is ubiquitously expressed in multiple splice forms with different cellular functions (amphiphysin 2) (70). Amphiphysin 1 consists of an N-terminal BAR domain and a C-terminal Src homology region 3 (SH3) domain linked by an unstructured region (54) (Figure 4B). The BAR domain of amphiphysin consists of three long α-helices and dimerization of BAR domains of two amphiphysin molecules leads to the formation of a banana-shaped dimer that can bind to membranes via its positively charged, concave surface (22). The linker region contains a canonical CBM as well as an additional motif for clathrin, the W box motif, which both bind to the N-terminal domain of clathrin and therefore contribute to recruitment and/or dwelling of clathrin triskelia to/at membranes (4) (see section 1.2.4). Amphiphysin also interacts with dynamin and the α subunit of AP2 (71, 72) via the SH3 domain and the linker region, respectively (54, 73). Binding of dynamin to the SH3 domain displaces clathrin (71), even though the binding sites are situated in different parts of amphiphysin. The SH3 domain controls polymerization of dynamin into a ring around the neck of a budding vesicle (73), therefore regulating the scission of cargo-loaded vesicles from membranes.

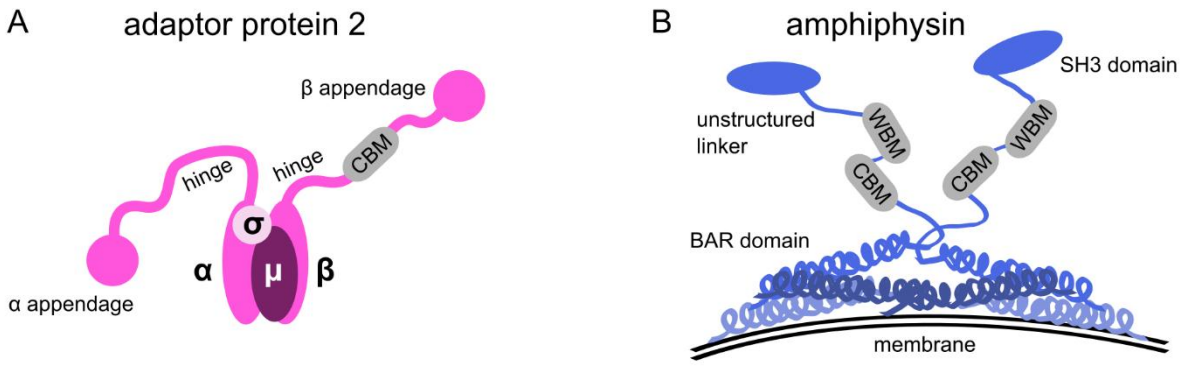


Figure 4 Schematic representations of AP2 and amphiphysin. (A) AP2 is a tetrameric complex consisting of four subunits (α, β, σ, μ). The hinge region of the β subunit of AP2 harbors a canonical clathrin-binding motif (CBM). (B) Amphiphysin consists of a BAR domain, an unstructured linker containing a canonical CBM and a W box binding motif (WBM), and an SH3 domain. Two molecules of amphiphysin associate and their respective BAR domains form a banana-shaped dimer which binds to curved membranes.

1.2.4 Clathrin N-terminal domain

In 1998, ter Haar and colleagues solved the X-ray crystal structure of clathrin NTD, demonstrating that it adapts a seven-bladed β-propeller fold (74) (Figure 5). Two years later,

the same group published the first structures of NTD in complex with binding partners, two peptides derived from the clathrin-binding proteins β-arrestin 2 (also termed arrestin2) and the β subunit of adaptor protein 3 (AP3) (62). Both peptides contained clathrin-binding motifs, LIEFE for β-arrestin 2 and LLDLD for AP3-β3, that had been previously shown to mediate the interaction with clathrin. (75, 76). In agreement with a study investigating the binding of β-arrestin to recombinantly expressed NTD containing alanine scanning mutations between residues 85-100 (77), the β-arrestin 2 peptide bound to a surface groove between blades 1 and 2 of the NTD β-propeller (62) (Figure 5). Surprisingly, the AP3-β3 peptide bound to the same region on NTD, and the authors proposed that proteins containing the consensus sequence L[LI][DEN][LF][DE] generally interact with clathrin via peptides at this site (62, 75). This motif can be found in a vast number of mammalian clathrin-binding proteins, including amphiphysin, and is now known as the "canonical" clathrin-binding (or clathrin-box) motif with the consensus sequence LΦXΦ[DE] (78).

Following the discovery of the "clathrin box" binding site on NTD, several biochemical studies showed that amphiphysin contains not only a canonical CBM, but also a second conserved sequence downstream of the LΦXΦ[DE] motif, spanning residues PWDLW (79-81). This motif is sufficient for an interaction with clathrin and enhances binding to clathrin cooperatively with the canonical CBM (79, 80). In 2004, Miele et al. solved the crystal structure of NTD in complex with a PWDLW-containing peptide. They demonstrated that this motif binds to a distinct site on clathrin NTD, the surface around the central "hole" of the β-propeller that is oriented towards the membrane during vesicle budding, and termed it "W box" (4) (Figure 5). Subsequently, Collette et al. investigated the importance of these binding sites for clathrin function in vivo using yeast as a model system (82). Surprisingly, when clathrin heavy chain mutated at both the clathrin box and the W box was expressed as the only source of clathrin in yeast, yeast growth was not affected and the mutated clathrin showed no disruption of endocytosis or intracellular clathrin trafficking. These results suggested that additional binding sites on clathrin NTD might exist (82, 83).

The conclusion from these functional experiments in yeast was supported by structural and biochemical data published shortly after the yeast study. Kang et al. presented a crystal structure of short and long isoforms of β-arrestin 2 (arrestin2S and arrestin2L, respectively) bound to

clathrin NTD (1). The structures clearly showed that the canonical CBM of arrestin2, present in both isoforms, bound to the clathrin box as predicted, whereas a loop formed by an eight amino acid insert only present in arrestin2L interacted with a third site on NTD. This third binding site, referred to as "β-arrestin splice loop site" or "arrestin box", is located between blades 4 and 5 of the NTD β-propeller, and biochemical experiments indicated [LI][LI]GXL as the consensus binding motif (1) (Figure 5).

Based on these studies, Willox and Royle performed functional experiments trying to elucidate which of the three binding sites of NTD contribute to the recruitment of AP2 to clathrin in vivo (2). They assessed the uptake of transferrin receptor, a well-established model for endocytosis, in mammalian cells expressing mutated clathrin in place of the endogenous protein by confocal fluorescence microscopy and flow cytometry. Unexpectedly, their experiments demonstrated that even upon mutation of all three known binding sites — the clathrin box, the W box and the arrestin box — together, endocytosis of transferrin receptor is undisturbed, indicating that a fourth binding site on NTD might be sufficient to sustain endocytosis. Willox and Royle then elegantly used sequence conservation analysis to predict the putative location of this fourth binding site to be at blade 7 of the NTD β-propeller (Figure 5) and showed that disruption of the three known sites plus this putative fourth site together blocks transferring receptor uptake. In addition, their experiments demonstrated that each binding site on NTD on its own is sufficient for endocytosis. However, even though they predicted the location of the fourth binding site, they neither identified a consensus binding motif responsible for interactions at this site, nor demonstrated which adaptor or accessory proteins can contact NTD via this site.

Taken together, over the span of roughly 20 years, the interaction sites of clathrin NTD have been identified and three of four have been structurally characterized. Early studies indicated that a single, distinct consensus motif mediates the binding of linear peptides to each site on NTD. However, recent functional studies suggest a high degree of functional redundancy either on part of the adaptor proteins recruited to clathrin, or on part of the involved binding sites (2, 82).

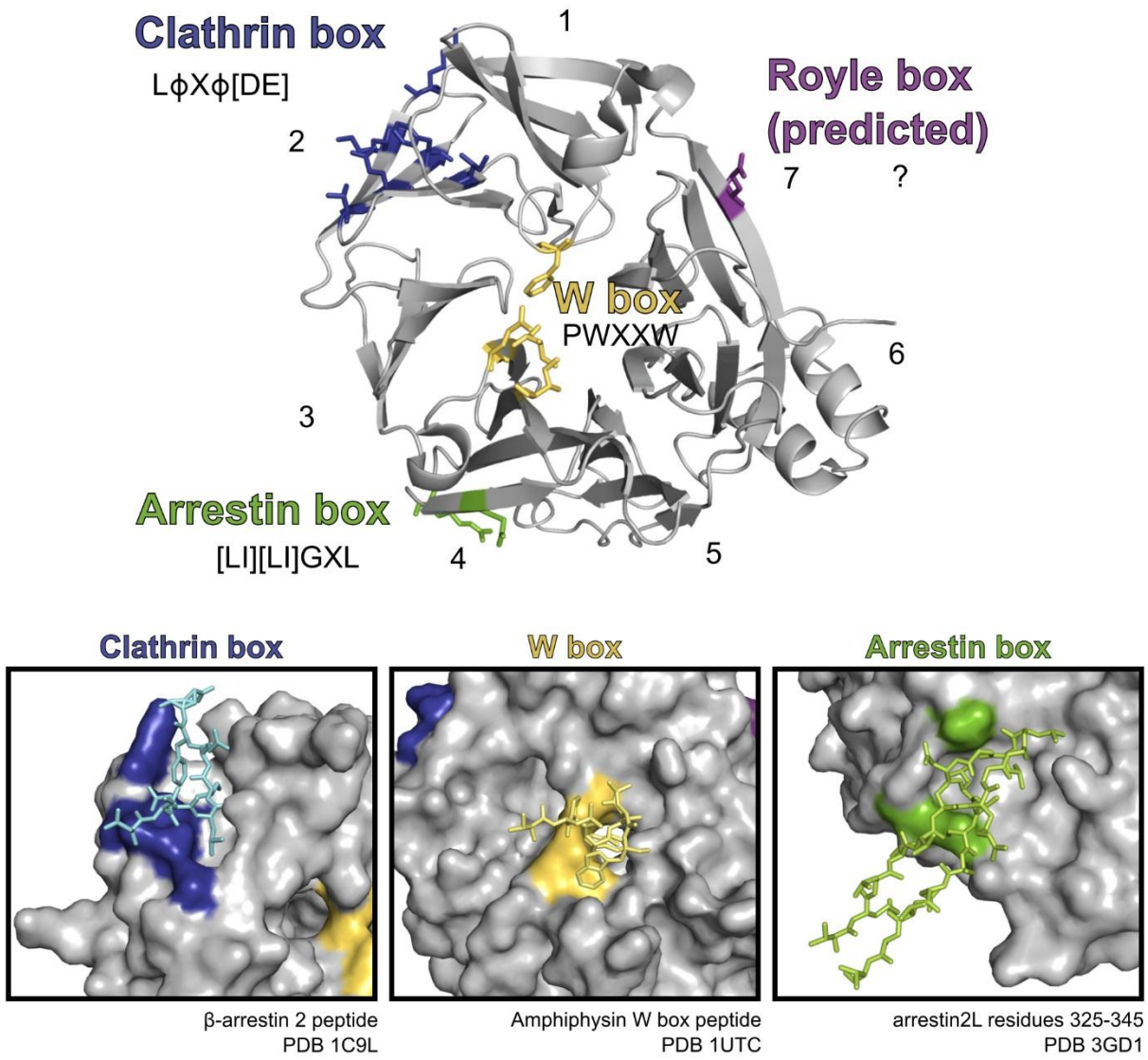


Figure 5 Binding sites on clathrin N-terminal domain. Clathrin NTD is shown in grey as a cartoon (upper panel) or as a surface model (bottom inserts). The seven blades of NTD are numbered. Key residues of NTD at the four different binding sites are shown as colored sticks or colored surfaces, respectively, indicating the location of the binding sites on NTD. Consensus binding motifs for the clathrin box, the W box and the arrestin box are noted. The inserts show peptides derived from β-arrestin 2 (blue sticks), amphiphysin (yellow sticks) and arrestin2L (green sticks) bound at their respective interaction sites. The PDB IDs of the original structures are listed.

1.2.5 Example of reverse topology budding: ESCRT-mediated formation of multi-vesicular bodies

The characteristic membrane remodeling process facilitated by the ESCRT machinery is the invagination of endosomal membranes into the lumen of the endosome, resulting in the formation of vesicle-containing organelles called multi-vesicular bodies (MVBs). Four multimeric protein complexes, ESCRT-0, ESCRT-I, ESCRT-II and ESCRT-III, and the ATPase Vps4 have been

described and interact at endosomal membranes in a regulated fashion to promote reverse topology budding (Figure 6).

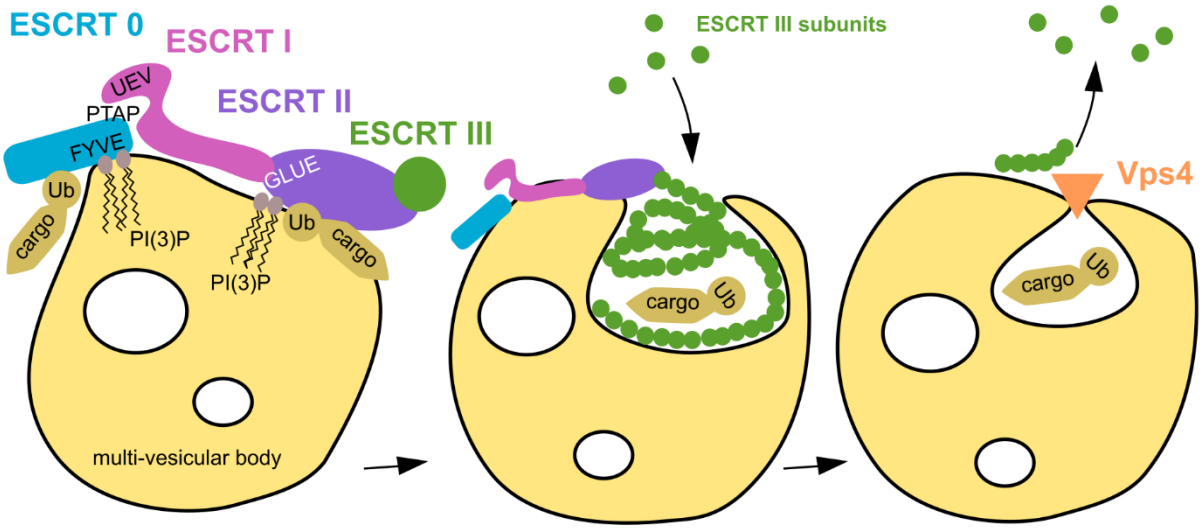


Figure 6 Budding into of multi-vesicular bodies mediated by the ESCRT machinery. The ESCRT0 complex (blue) recognizes PI(3)P via its FYVE domain and also binds to ubiquitinated cargo (khaki). ESCRTI (pink) binds via its UEV domain to a short linear motif (PTAP) of ESCRT0. ESCRTII (purple) binds to ESCRTI via a GLUE domain, and also recognizes PI(3)P and ubiquitinated cargo. ESCRTIII (green) interacts with ESCRTII. Multiple subunits of ESCRTIII associate with each other and form long filament which probably promote membrane invagination. Vps4 (orange) "pulls" the ESCRTIII subunits out of the budding vesicle and thus contributes to membrane fission.

ESCRT-0 contains a FYVE (Fab-1, YGL023, Vps27, and EEA1) zinc finger domain, which binds to the endosomal lipid phosphatidylinositol 3-phosphate (PI(3)P) (84) and thus determines the targeting of the ESCRT machinery to endosomes for MVB biogenesis. Additionally, ESCRT-0 binds to and clusters ubiquitinated cargo proteins (85), thus sorting them to the site of budding. However, it should be noted that also non-ubiquitinated proteins can be incorporated into intraluminal vesicles if they bind to components of the ESCRT machinery (86). One of the subunits of ESCRT-0 also harbors a PTAP-like motif, which binds to the ubiquitin E2 variant (UEV) domain of ESCRT-I and thus recruits it to the endosome (87). The opposite end of the ESCRT-I multimer binds to the ESCRT-II GLUE (GRAM-like ubiquitin-binding in EAP45) domain, which also recognizes PI(3)P and thus further contributes to endosomal targeting of the ESCRT complexes (88). The ESCRT-III complex binds to ESCRT-II via interaction of CHMP6 and EAP20 subunits, respectively (89).

The assembly of these ESCRT proteins serves as a "nucleation complex" (38), to which more ESCRT-III monomers can bind and form long polymerized filaments. These flexible ESCRT-III

filaments are generally thought to assemble as spirals or helices on the inside of the necks of invaginating vesicles, where interactions between the filaments and with the endosomal membrane stabilize the polymers and contribute to membrane deformation (38, 90-92). The ESCRT-III complex lastly recruits Vps4 (91) to the budding site, which disassembles the filaments and pulls monomers through the vesicle neck out of the vesicle lumen (93, 94). The exact mechanism of abscission of the vesicle into the organelle lumen is still debated; it has been speculated that fission of the membranes occurs either spontaneously due to the high curvature of membranes at the tip of the vesicle neck, or that disassembly of the ESCRT-III polymers by VPS4 actively contributes to membrane fission (38, 93, 95).

1.3 Exploitation of the endomembrane system by viruses

Enveloped viruses acquire their coat from either the plasma membrane or from compartments of the endomembrane system. For viruses that assemble at the plasma membrane, budding represents the final step of the intracellular viral life cycle. For viruses that bud into intracellular compartments, transport of virion-filled vesicles to the plasma membrane via the secretory pathway and fusion of these vesicles with the plasma membrane must occur to release virus progeny into the extracellular space in the absence of cell lysis. In addition, viruses entering host cells via binding to specific surface receptors indirectly exploit clathrin-mediated endocytosis or other endocytosis mechanisms during virus uptake (Figure 7). Examples of viruses hijacking clathrin- or ESCRT-mediated trafficking mechanisms are outlined below.

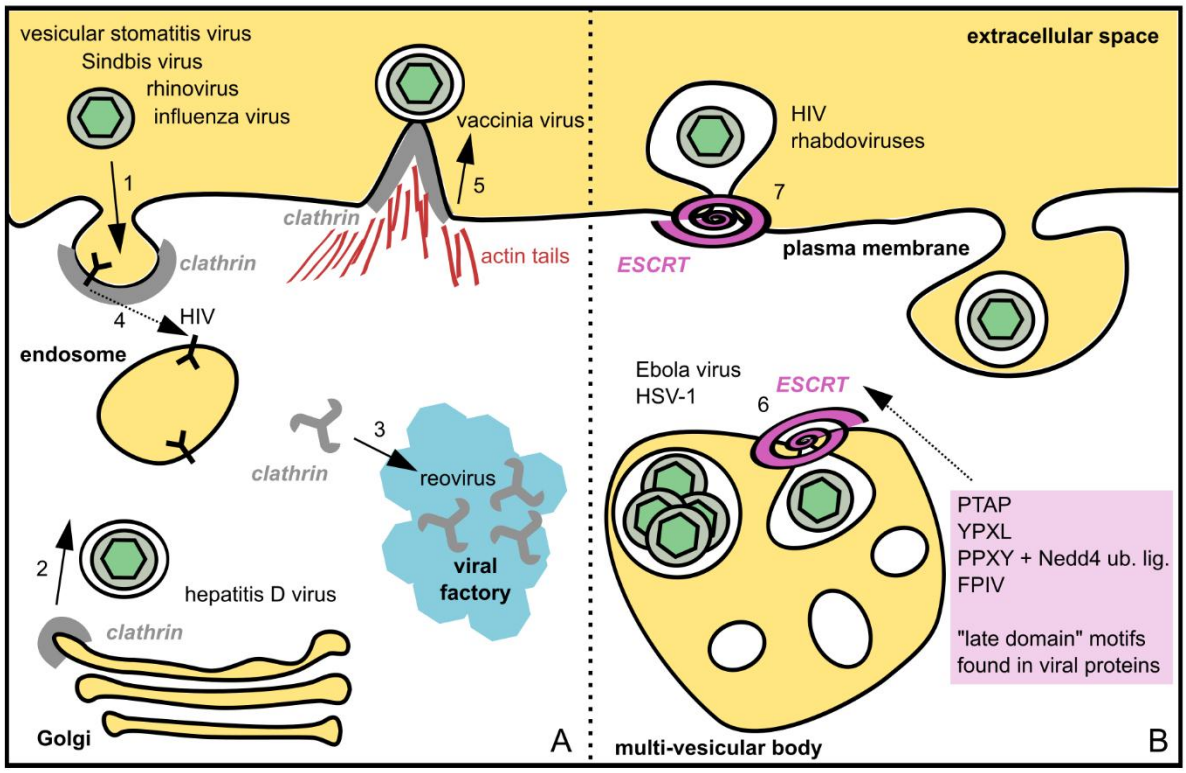


Figure 7 Examples of viruses hijacking the cellular clathrin trafficking machinery or the ESCRT machinery. (A) 1: Viruses entering host cells via clathrin-mediated endocytosis. Clathrin is depicted in grey. 2: Hepatitis D virus exploits clathrin during viral egress after envelopment at the trans-Golgi compartment. 3: Reovirus sequesters clathrin to viral factories. 4: Human immunodeficiency virus 1 (HIV) influences endocytosis of cell surface receptors. 5: Vaccinia virus recruits clathrin to promote formation of actin tails. (B) 6: Viruses budding into intracellular compartments (e.g. multi-vesicular bodies) by using the ESCRT machinery (pink). 7: Viruses redirect the ESCRT machinery to the plasma membrane to facilitate budding into the extracellular space. Late domain motifs interacting with components of the ESCRT machinery or with members of the Nedd4-family of ubiquitin ligases (Nedd4 ub. lig) are highlighted (pink).

Early discovered examples of viruses entering host cells via clathrin-mediated endocytosis were Semliki Forest virus, Sindbis virus and human rhinovirus 14 (96). Another important example is influenza virus, which triggers de novo assembly of clathrin-coated pits during viral entry (97). Remarkably, even very large viruses such as vesicular stomatitis virus, an enveloped virus with particles measuring 70 nm × 200 nm, can enter cells via clathrin-mediated endocytosis by assembling clathrin cages at the tip of an endocytic structure and polymerizing actin filaments to internalize the whole virion (98). Viruses also interfere with clathrin trafficking to influence levels of cell surface receptors in order to promote virus spread or to avoid the host cell immune response. For example, human immunodeficiency virus 1 (HIV-1) downregulates the cell receptors CD4 and major histocompatibility complex class I (99, 100), and mammalian reovirus sequesters

clathrin to viral factories, thereby depleting the cellular clathrin pool and disrupting physiological clathrin functions like endocytosis and regulated secretion (101).

In contrast, the role of clathrin during viral egress has been investigated less extensively (102). Clathrin and adaptor protein 1 have been implicated in the egress of hepatitis C virus particles from cells (103) and in budding of HIV-1 at the plasma membrane (104, 105). Clathrin:virus interactions also contribute to the formation of virus-like particles of hepatitis D virus (106, 107) (see section 1.4.4). Yet another way of clathrin manipulation is employed by vaccinia virus. Clathrin and AP2 are reported to assemble on the cytoplasmic domain of the vaccinia virus protein A36, supporting the formation of actin tails that propel exocytosed vaccinia particles forward to enhance viral spread (102, 108, 109).

Whereas clathrin-mediated viral egress is not well understood, exploitation of the ESCRT system by viruses undergoing envelopment at the plasma membrane or at intracellular organelles has been studied in more detail (6, 36, 110). An interaction between ESCRT and viral budding was first discovered in retroviruses. Göttlinger et al. and Huang et al. showed that the scission of fully formed HIV-1 virions at the plasma membrane is inhibited when HIV-1 protein p6 is either truncated or a short peptide motif of p6 spanning the four residues Pro-Thr-Ala-Pro (PTAP) is mutated (111, 112). Similar motifs were also found in other viruses, multiple studies supported the notion that PTAP-like motifs recruits host cell proteins to mediate viral budding, and the term “late motif” or “late domain” was coined to indicate the importance of these motifs for late steps during viral life cycles (110). It was shown that the cellular binding partner of these late motifs is the UEV domain of ESCRT-I component Tsg101 and that interaction between Tsg101 and viral proteins containing these motifs recruits the ESCRT machinery to the site of viral budding, facilitating viral envelopment and scission of virions (113-116).

Subsequently, different late domain motifs and their cellular interaction partners were identified. The YPXL (or YPX_nL, where n symbolizes multiple inserted residues) motif binds to the cellular ESCRT adaptor protein ALIX, which itself is recruited to ESCRT-III components. The importance of YPXL motifs for viral budding was discovered in equine infectious anemia virus (117), but interactions between this type of late domain and ALIX have now been described for many virus families (36). The third late domain motif, PPXY, binds to the WW domain of cellular HECT (homologous to the E6AP carboxyl terminus) domain-containing ubiquitin ligases

of the Nedd4 family (118, 119). PPXY motifs have been implicated in the budding of human T-cell leukemia virus type 1 (120), Ebola virus (121), rhabdoviruses (122) and many others (6, 36). The exact link between PPXY motifs, the Nedd4 ubiquitin ligases and the ESCRT machinery has not been established yet. However, it has been suggested that PPXY-mediated budding of viruses requires the catalytic ubiquitin ligase activity of the HECT domain of Nedd4 ligases (119, 123), and that arrestin-related trafficking proteins could be involved in the PPXY:ESCRT pathway (124). A fourth late domain motif, "FPIV", has also been described (36). In summary, although some viruses can form buds without involvement of cellular factors (6, 125), ESCRT-mediated budding into the extracellular space or into intracellular organelles is probably the main mechanism employed by enveloped viruses during membrane wrapping.

1.4 Hepatitis D virus

1.4.1 Transmission, infection scenarios and disease burden

Hepatitis D virus (HDV), also referred to as hepatitis delta virus or simply deltavirus, can only maintain infections in hosts also infected with hepatitis B virus (HBV) since it relies not only on host cell machinery, but also on HBV envelope surface antigens to facilitate virus reproduction (see sections 1.4.2 and 1.4.3). Therefore, HDV is considered a satellite virus of HBV. So far, HDV is the only described mammalian satellite virus (126). HDV is hepatotropic and therefore causes liver pathologies. The transmission of HDV occurs via parenteral routes, for example by contact with contaminated blood, sharing of syringes, or a yet unknown insect vector (127). Unlike HBV, HDV is not commonly transmitted perinatally (128).

Due to its dependence on HBV, two infection scenarios can be distinguished. Infection with HDV can either occur at the same time as infection with HBV, known as co-infection, or subsequent to establishment of chronic HBV infection in the host, known as super-infection. Clearance of an acute co-infection of a naïve individual is dependent on the host immune response to HBV. In most cases, HBV/HDV co-infection is self-limiting, but this mode of infection leads more often to the development of acute hepatitis (extensive necrosis of the liver, liver failure, and possibly death) than mono-infection with HBV (129). Super-infection of a chronic HBV carrier causes acute fulminant hepatitis which can be cleared, but in 80% of all cases, the disease progresses to

chronicity (130). Chronic hepatitis D severely aggravates liver diseases like cirrhosis and increases the likelihood development of hepatocellular carcinoma (131).

It is estimated that around 15-20 million people are chronically infected with HDV, which equals approximately 5% of the population of chronic HBV carriers (132-134). However, since one third of the world's population encounter HBV during their lifetime, equaling 2 billion people world-wide (135), and routine testing for HDV is not common (127), this number could be underestimated. Due to dependency of HDV on an underlying HBV infection, the highly effective HBV vaccine protects against HDV as well. However, all chronically infected HBV carriers are at risk of HDV superinfection. HDV is also still a major health concern in poorer countries, as well as in the Western world in drug user communities and populations of immigrants from high-prevalence regions (136). Currently, the only treatment option is interferon α, which is associated with severe side effects, clears the HDV infection in only up to 50% of treated patients, and often does not resolve the underlying HBV infection (127). New drugs targeting the host cell farnesyltransferase or the HDV/HBV entry receptor are undergoing clinical trials (127). Overall, infections with HDV contribute to an extremely high liver disease burden world-wide and current treatment options are very limited.

1.4.2 Virus structure

Hepatitis D virus is an enveloped virus measuring only 36 nm in diameter (127) with a circular, negative-sense single-stranded RNA genome of around 1700 bases (b), one of the smallest known viral genomes (137-139). Because of its size, circular genome structure and mechanism of replication, HDV is often compared to viroids; small, non-coding circular RNA pathogens infecting plants (138-141).

In contrast to viroids, the HDV genome encodes one protein, which occurs in two isoforms arising from RNA-editing of the virus genome: small and large hepatitis D antigen (HDAg-S and HDAg-L). The small HDAg has a molecular weight of 24 kDa and contains 195 amino acids, whereas HDAg-L contains 214 or 215 amino acids (depending on the virus genotype), corresponding to a molecular weight of 27 kDa (142). The large HDV antigen possesses a CXXQ farnesylation motif at its C terminus and is farnesylated, whereas HDAg-S is not (143). HDV antigens assemble around the RNA genome to form the ribonucleoprotein complex (144, 145) (Figure 8). The viral RNA is highly self-complementary, and thus, adopts extensive secondary structures (139). A

crystal structure of residues 12-60 of HDAg showed that HDV antigens dimerize via an anti-parallel coiled-coils and suggested that these dimers subsequently form octameric complexes with a 50 Å basic central hole that could accommodate a rod-like RNA molecule (146).

The envelope of HDV is formed by antigens of HBV (Figure 8). It has been shown that HDAg-L and the small surface antigen of HBV (S-HBsAg) are required for the formation of HDV particles, whereas other HBV antigens are not (144, 147). The interaction between HDAg-L and S-HBsAg is direct and farnesylation of HDAg-L is necessary, but not sufficient, for binding (148). A tryptophan in the second cytosolic loop of S-HBsAg has been implicated in the recruitment of HDAg-L to S-HBsAg (149, 150). The incorporation of the large surface antigen of HBV (L-HBsAg) is necessary for infectivity of HDV particles since L-HBsAg mediates viral entry of both HBV and HDV (151). However, the medium surface antigen of HBV (M-HBsAg) is required neither for assembly nor for entry of HDV (147, 152).

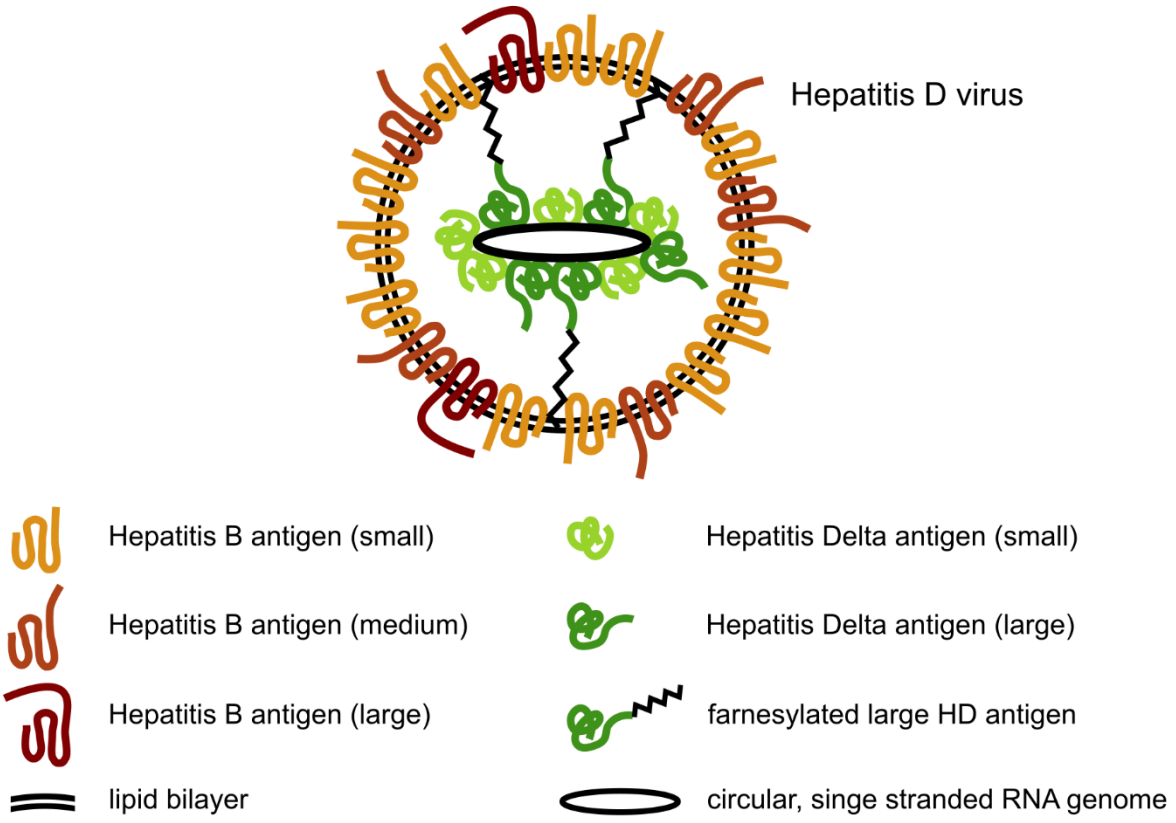


Figure 8 Cartoon of a hepatitis D virus particle. The HDV antigens associate with the viral genome, forming the nucleocapsid, whereas the viral envelope is formed by HBV proteins. The small antigen of HBV and farnesylated large antigen of HDV interact.

1.4.3 Viral life cycle

Since HDV and HBV particles are coated with the same virus envelope proteins, namely the small, medium and large surface antigens of HBV, HDV enters host cells using the same mechanism as HBV. Deltavirus particles attach with low affinity to heparan sulfate proteoglycans expressed on the surface of hepatocytes (153). Recent studies have identified the high-affinity surface receptor of HBV and HDV in mammalian cells. Yan et al. and Ni et al. demonstrated that L-HBsAg interacts specifically with sodium taurocholate co-transporting polypeptide (NTCP), mediating cell entry (154, 155). NTCP is a transmembrane protein responsible for the transport of bile salts (156, 157), and its assumed liver-specific expression explains the hepatotropism of HDV and HBV (158). The pathway by which HDV particles are taken up into cells, as well as the site of fusion of the viral particles with cellular membranes, have not yet been elucidated (127, 159).

After release of the HDV nucleocapsid into the cytosol, it is most likely targeted to the nucleus via the genome-bound small and large antigens, which both contain nuclear localization signals (160, 161). However, the specific cellular factors involved in transport of the capsid to and into the nucleus have not yet been defined (159). The replication of the HDV genome takes places in the nucleus by cellular polymerases. First, host cell RNA polymerase II binds to the genome and amplifies it by a rolling circle mechanism (162), producing antigenomic, linear RNA containing multiple copies of the genome, which is auto-catalytically cleaved by the HDV ribozyme (163). Both the genome and the antigenome of HDV encode ribozymes that adapt a similar secondary structure (164). The cleaved, antigenomic linear RNA is then ligated to form a circular RNA molecule. Even though it has been reported that the ribozyme can mediate this ligation step in the absence of any proteins in vitro (165), this self-ligation is very slow and it has therefore been proposed that cellular ligases facilitate formation of the circular RNA in vivo (166). The circular antigenomic RNA serves as a template for RNA polymerase II-dependent rolling-circle replication, yielding linear genomic RNA, which is again ligated to form circular genomic RNA (167).

The genomic RNA serves as the template for mRNA production and HDAg expression. At later points during infection, the amber stop codon of HDAg-S (UAG) on the antigenomic RNA is modified by the small isoform of the host enzyme ADAR1 (adenosine deaminase that acts on

RNA) during transcription (168). This modification results in deamination of the adenosine of the amber stop codon to inosine (codon UIG). Upon transcription of the genome from this modified antigenome, the inosine pairs with cytosine (ACC). Subsequently, mRNA transcribed from the modified genomic RNA contains the codon UGG instead of the amber stop codon. This change results in a tryptophan base being incorporated in place of the stop codon, and thereby, the production of the large antigen of HDV.

The small and the large antigen of HDV have opposite roles during replication: HDAg-S enhances transcription, whereas expression of HDAg-L inhibits replication and promotes the later stages of the viral life cycle (169, 170). Antigenomic mRNAs encoding the two isoforms of hepatitis D virus antigen are transcribed throughout the viral life cycle, 3' polyadenylated, 5' capped and transported into the cytoplasm, where both isoforms of HDAg-L are produced using the host cell translation machinery (161, 167, 171, 172). During infection, genomic RNA, but not antigenomic RNA is also exported into the cytoplasm (173), where assembly and subsequent egress of viral particles takes place, completing the viral life cycle (Figure 9).

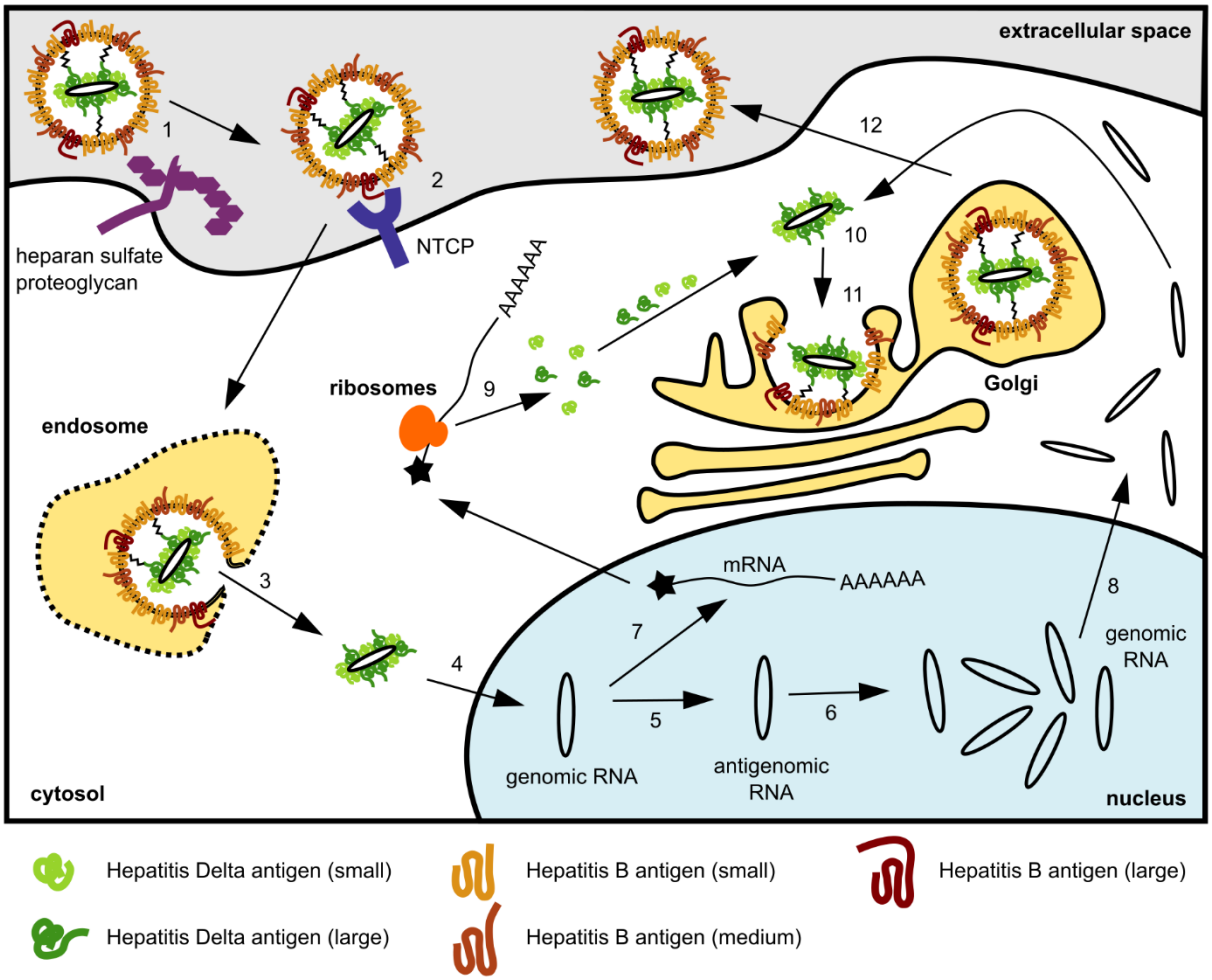


Figure 9 Life cycle of hepatitis D virus. 1: HDV attaches to cell surface via binding to heparan sulfate. 2: HDV interacts with NTCP and is taken up into host cell. 3: The viral nucleocapsid is released into the cytosol. 4: The viral nucleocapsid is transported into the nucleus. 5: The viral genome is transcribed to produce antigenomic RNA. 6: Antigenomic RNA serves as a template for production of more viral genomic RNA. 7: mRNAs encoding the small and the large antigen of HDV are transcribed from the genomic circular RNA and transported into the cytosol. 8: Copies of the genomic circular RNA are also transported into the cytosol. 9: Cellular machinery translates the viral mRNAs and therefore produces small and large HDV antigens. 10: Genomic RNA and viral antigens assemble and form the nucleocapsid. 11: The nucleocapsid buds into the trans-Golgi and thereby acquires the viral envelope, including the hepatitis B virus antigens. 12: Transport vesicles filled with HDV particles are trafficked to the plasma membrane, where the progeny virus is released into the extracellular space.

1.4.4 Egress of HDV and interactions of HDAg-L with clathrin

The egress of HDV particles is poorly characterized. Since HDV and HBV share the same viral envelope proteins — the small, medium and large HBsAgs — it is conceivable that they acquire the envelope and are released from host cells using the same cellular mechanisms. Envelopment of infectious HBV particles, also called Dane particles, has been reported to involve formation of MVBs using the cellular ESCRT system (174, 175) (Figure 10A). HBV also releases numbers of non-infectious, capsid-less subviral particles into the extracellular space, which probably serve as

a decoy for the host immune system. These subviral particles are assembled in the ER and the ER-Golgi intermediate compartment (175, 176) (Figure 10B). However, in contrast to these two HBV envelopment pathways, it has been suggested that the Golgi network is the budding site for HDV and that release of HDV particles requires the interaction of HDAg-L with clathrin (106, 107, 177) (Figure 10C).

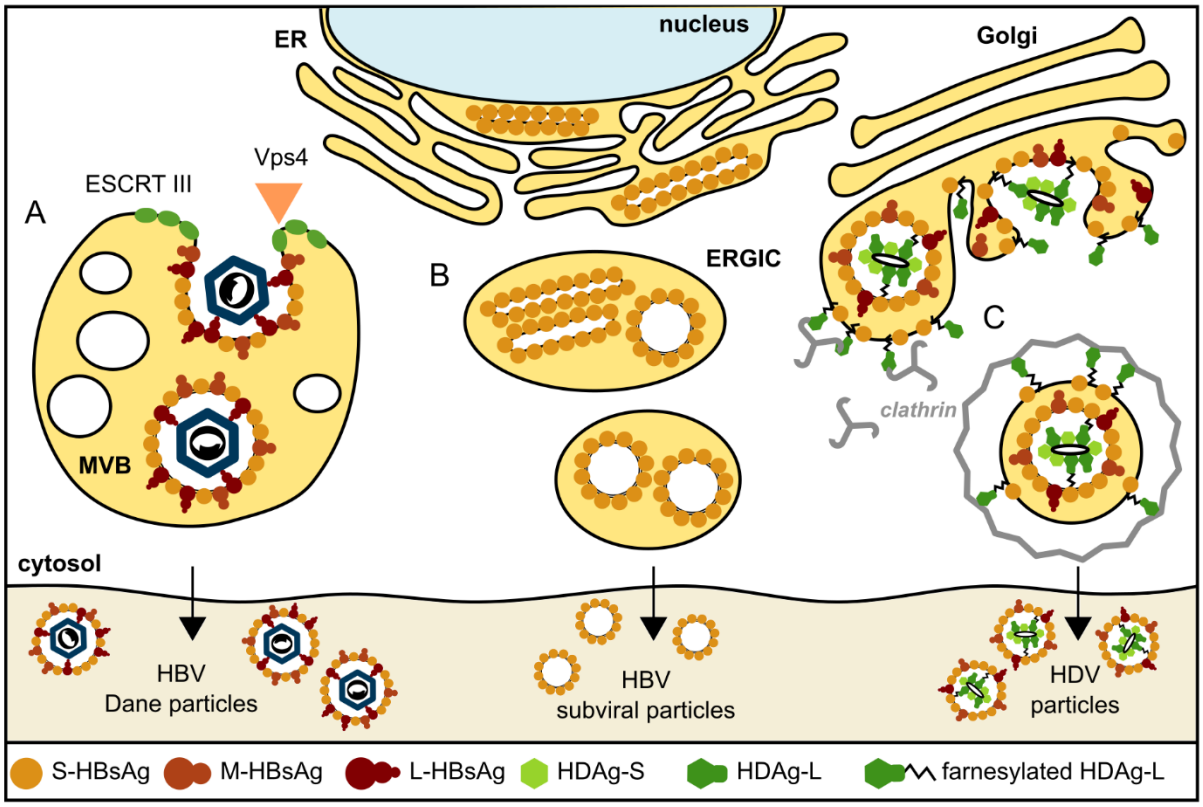


Figure 10 Comparison of proposed egress pathways of HBV and HDV particles. (A) Infectious HBV particles (Dane particles) exploit the ESCRT machinery to bud into multi-vesicular bodies (MVB). (B) Non-infectious subviral particles of HBV devoid of capsid are thought to assembly in the ER and the ERGIC via an ESCRT-independent mechanism and are released via the secretory pathway. (C) HDV buds into the trans-Golgi, acquiring its HBV envelope proteins. A cytosolic protein, HDAg-L, binds to the membrane-localized small antigen of HBV (S-HBsAg) and is therefore also found on the cytosolic face of the Golgi, where it can interact with clathrin (grey). Particles of HDV are then transported in a clathrin-coated secretory vesicle to the plasma membrane and released.

Currently, HDV isolates are grouped into eight clades (or genotypes) (178-180). This classification is based on sequence alignments and phylogenetic analysis of the whole RNA genome of HDV isolates. Wang et al. analyzed the C-terminal extension of HDAg-L from 43 HDV isolates and discovered that these C-terminal "tails" of HDAg-L can be categorized into three groups (177). The C-terminal extensions of HDAg-L belonging to group 1 (corresponding to clade/genotype HDV-1) contain residues LFP[ASV]D, a motif resembling the canonical

clathrin box binding motif LΦXΦ[DE]. Likewise, HDAg-L of HDV clades 2 and 4-8 contain the motif LPLLE, resembling to the canonical CBM, and form group 2. Only the C-terminal tail of HDAg-L of HDV clade 3 does not possess an obvious clathrin-binding motif. However, the C-terminal extension of this clade resembles other genotypes in that it is similarly rich in hydrophobic residues. In this thesis, the three consensus sequence groups of C-terminal tails of the large hepatitis D virus antigen are referred to as HDAg-L1, HDAg-L2 and HDAg-L3, respectively (Figure 11).

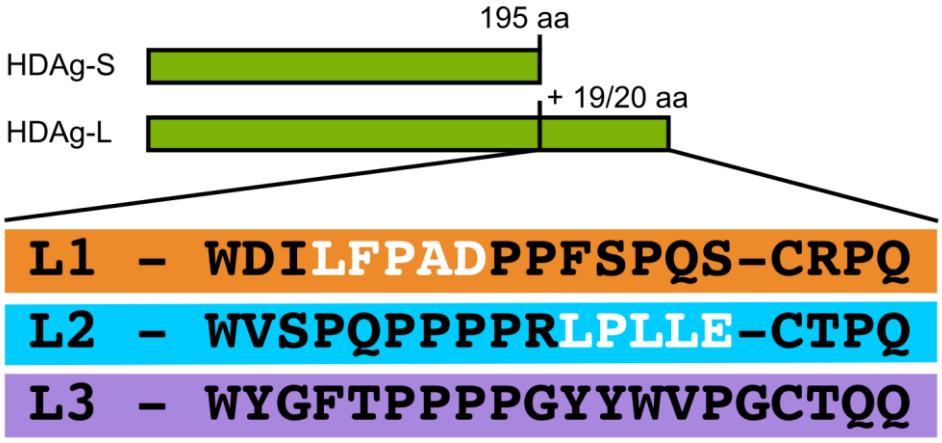


Figure 11 The C-terminal extensions of HDAg-L of different HDV genotypes. All known HDV C-terminal tail amino acid sequences can be classified into one of the three depicted groups. Sequences of the "L1" and the "L2" group harbor a putative clathrin-binding motif (white).

In 2007, Huang and colleagues identified clathrin heavy chain (CHC) as an interaction partner of HDAg-L1 in co-immunoprecipitation experiments and demonstrated co-localization of HDAg-L1 and clathrin at the trans-Golgi network (TGN), but not the ER or the PM (107). Huang et al. furthered their findings by showing that mutation of key residues L199 and D203 of the putative clathrin box binding motif contained in the C terminus of HDAg-L1 ($_{199}$LFPAD$_{203}$) does not only inhibit HDAg-L1 binding to CHC, but also abolishes the release of HDV virus-like particles from cells (106). Moreover, they demonstrated that shRNA knockdown of CHC leads to a significant drop in the number of released HDV particles but does not reduce the number of released HBV subviral particles (106).

Huang et al. also reported an interaction between clathrin and HDAg-L2, which contains a different clathrin box motif (LPLLE), albeit binding seemed less efficient than between HDAg-L1 and clathrin (106). In contrast, Wang and colleagues, who assessed the interaction between a

purified fragment of the clathrin NTD and all three groups of HDAg-L, claimed that HDAg-L2 binds more strongly to CHC than HDAg-L1 (177). Taken together, these results indicate that even though HDV and HBV share the same envelope proteins, their respective sites of envelopment are different and that clathrin plays a key role in HDV, but not in HBV, assembly.

Currently, it is unknown which step of the envelopment or egress of HDV is promoted by the HDAg-L:clathrin interaction. Presumably, clathrin participates in the budding of transport vesicles containing enveloped HDV particles off the Golgi. At first sight, it seems counterintuitive how cytosolic HDAg-L could recruit clathrin to the Golgi membrane to promote vesicle budding. However, farnesylated HDAg-L binds to the small antigen of HBV (S-HBsAg), which is a multi-pass transmembrane protein. This interaction could explain how HDAg-L is specifically enriched at the trans-Golgi and could therefore direct clathrin to the site of viral envelopment (Figure 10C).

1.5 Herpes simplex virus 1

1.5.1 Herpes simplex – an underestimated disease

Herpes simplex virus I (HSV-1) belongs to the subfamily *Alphaherpesvirinae* of the order of *Herpesvirales* (181). The virus is probably best known as the causative agent of orofacial ulcers ("cold sores"). However, HSV-1 can also cause genital and esophageal lesions, ocular herpes keratitis, encephalitis, severe disease in immunocompromised patients, and neonatal herpes (182-186). It is estimated that 67% of the population are infected with HSV-1, and 140 million people worldwide suffer from genital HSV-1 infection (186). HSV-1 is transmitted enterally by oral or genital contact, and risk of infection is highest when active sores are present (186). However, the virus can also be infrequently transmitted intrauterine or during birth, leading to neonatal herpes associated with high morbidity and mortality (187).

Primary infection with HSV-1 is mostly acquired during childhood or puberty, and is often asymptomatic (188). The virus infects epithelial cells of non-keratinized mucosal tissues in the oral cavity (e.g. inner lip) or genital area, though it can also infect basal layers of keratinized epithelium if lesions such as small cuts in the skin are present (189). After infection of epithelial cells, the virus spreads to sensory nerve terminals and is retrogradely transported to the cell bodies of neurons of the peripheral ganglia, establishing lifelong latent infection (190) (Figure

12). Both viral and host factors contribute to viral latency (191). The virus can reactivate and subsequently spreads from the latently infected neuronal cells by anterograde transport to keratinized epithelial tissues, leading to active lytic infection at these sites and the typical development of blisters or ulcers containing large amounts of highly contagious infectious virus. Less commonly, HSV-1 can spread anterogradely from the peripheral ganglions to the central nervous system, causing severe neurological disease (190) (Figure 12). Reactivation of HSV-1 can be triggered by fever, emotional stress, tissue damage or suppression of the immune system, although the molecular mechanisms behind reactivation are currently only poorly understood (188, 191).

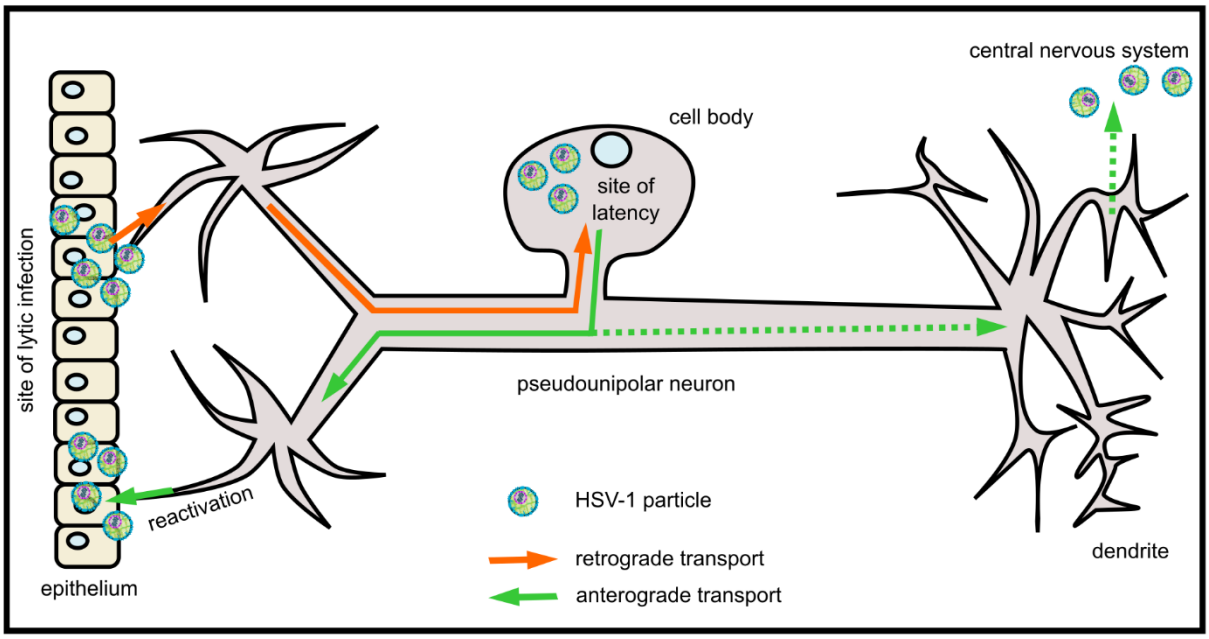


Figure 12 Sites of lytic and latent HSV-1 infection and directions of transport events during HSV-1 spread. HSV-1 infects epithelial (yellow) cells during lytic infection, establishes latency in the cell bodies of sensory neurons (light chocolate brown), and can spread to the central nervous system in rare cases. Directions of transport inside the neuron are indicated with arrows.

The standard treatment of epithelial HSV-1 lesions involves topical application of aciclovir or precursor drugs. Oral or intravenous therapy with aciclovir is usually reserved for treatment of immunocompromised patients or neonates where it reduces morbidity and improves survival (188). Aciclovir is converted into aciclovir triphosphate by the viral thymidine kinase and host cell kinases and subsequently blocks the HSV-1 DNA polymerase without affecting cellular DNA synthesis (192). However, these treatments only target the symptoms of acute lytic infection and

cannot resolve latency of HSV-1. There are few side effects of these antiviral drugs and their toxicity is low. However, resistance to aciclovir can develop and lead to severe disease or even death in immunocompromised individuals (188).

Taken together, although symptomatic treatment of lytic HSV-1 infection is well established, research into novel drug targets is necessary to provide better therapies for the more severe cases of HSV-1-caused disease, especially in neonates and the immunocompromised. Moreover, there is currently no vaccine against HSV-1 available. Importantly, investigation of basic HSV-1 biology does not only further our understanding of the life cycle of HSV-1, but discoveries can potentially also be translated to other herpesviruses affecting humans, such as herpes simplex virus 2 (HSV-2), varicella zoster virus, human cytomegalovirus (HCMV), Epstein-Barr virus, or to herpesviruses infecting livestock, such as suid herpesvirus 1 (pseudorabies virus) or equine herpesviruses. Infection with HCMV can cause devastating congenital disease and is also a major problem for solid organ transplant recipients and therefore, development of an HCMV vaccine is a priority for public health.

1.5.2 Virus structure and life cycle

Herpes simplex virus 1 is an enveloped virus with a linear, double-stranded DNA genome of 154 kb encoding at least 84 proteins (193). The monopartite genome is contained in an icosahedral nucleocapsid with a diameter of 125 nm (194). The capsid is surrounded by an amorphous proteinaceous layer called tegument. The viral envelope is studded with 600-750 copies of approximately 12 glycoproteins (195, 196). Cryo-electron tomography revealed that HSV-1 particles measure 170-225 nm in diameter and are not symmetric; the nucleocapsid is off-set within the tegument layer (193, 195) (Figure 13).

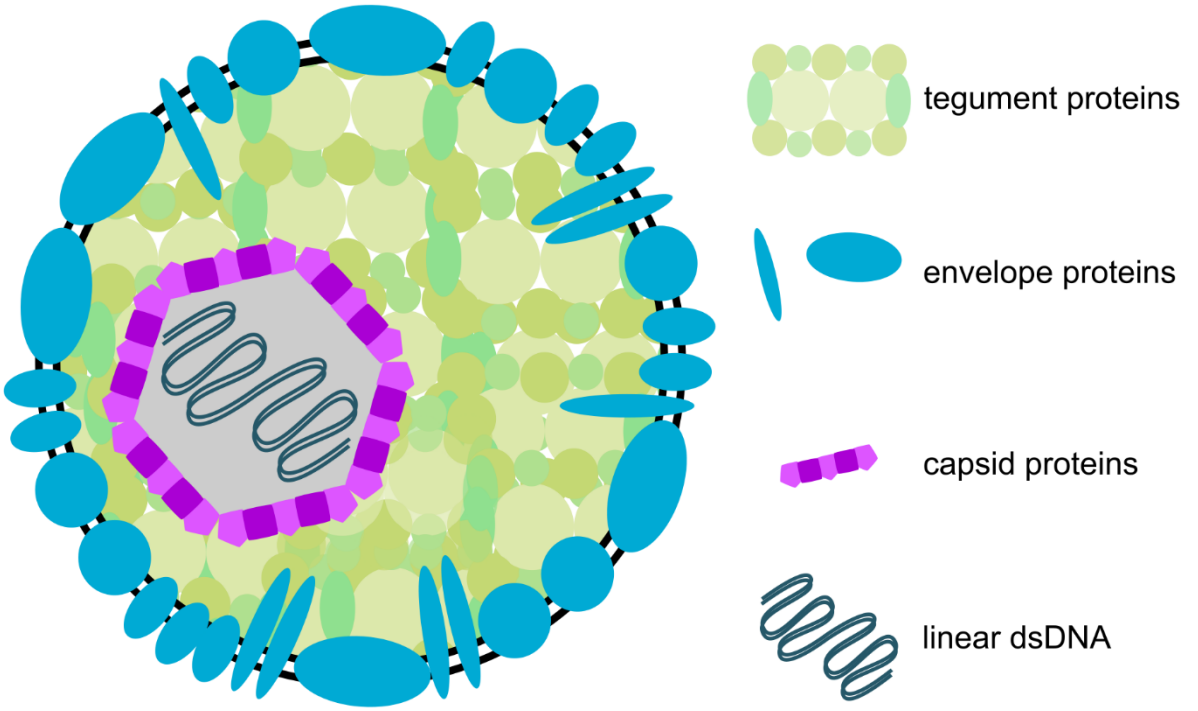


Figure 13 Cartoon of an HSV-1 virion. Capsid proteins assemble around the linear, double-stranded DNA to form the nucleocapsid. Between the nucleocapsid and the glycoprotein-studded viral envelope is an amorphous protein layer called the tegument.

The replication cycle of HSV-1 is comparatively complex (Figure 14). Entry of HSV-1 into host cells requires five of the twelve glycoproteins — glycoprotein B (gB), glycoprotein C (gC), glycoprotein D (gD), and the glycoprotein H/L dimer (gH/gL) (197, 198). Glycoproteins gB and gC mediate binding of the virus particle to the host cell surface via interaction with heparan sulfate, but such "attachment" to the host cell is not strictly necessary for infection, at least in cultured cells (198, 199). Entry receptors bound by gD are protein herpesvirus entry mediator (HVEM), immunoglobulins nectin-1 and nectin-2, as well as modified heparan sulfate (197, 198). The exact contribution of these receptors to HSV-1 entry in vivo has still to be elucidated. Nonetheless, it has been shown that nectin-1 is required for infection of neuronal cells in vivo (200). Fusion of the viral envelope with the plasma membrane is mediated by gB and gH/gL (197). It should be noted that several studies suggest that in certain cell types, for example keratinocytes, HSV-1 entry involves endocytosis and subsequent fusion of the viral envelope with endosomal membranes (201). Interestingly, ultrastructural studies showed that the "pole" of HSV-1 virions containing less tegument fuses with the membrane during entry, whereas the pole containing a thick layer of tegument is distal to the fusion site (202).

After entry, the outer layers of the tegument are disassembled and the capsid transported retrogradely along microtubules to the nucleus (203-205). The nucleocapsid then binds to the nuclear pore complex, where the viral DNA is uncoated and ejected into the nucleus (206). Transcription of viral DNA is facilitated by the host cell DNA-dependent RNA polymerase II, and DNA replication is performed by the viral DNA polymerase. Both processes take place in the nucleus and are aided by extensive reorganization of nuclear structures by HSV-1 (193). Expression of viral proteins is highly regulated and occurs in three cascades (193). First, the immediate early genes are transcribed. These genes encode proteins that activate the expression of the early genes, which in turn encode proteins involved in the replication of the viral DNA, for example the viral DNA polymerase. During the last phase of replication, the late genes are transcribed, which encode structural proteins involved in virus particle assembly and egress. In cultured cells, the entire replicative cycle takes place within approximately 16 h.

The viral DNA is packaged and the capsid assembled in the nucleus (207). Since the nucleocapsid is around 125 nm large, it cannot be transported into the cytoplasm via the nuclear pore, but instead undergoes envelopment at the inner nuclear membrane, buds into the perinuclear space, and fuses again with the outer nuclear membrane, releasing so-called “naked” capsids into the cytoplasm (207-209). The naked capsids then acquire the tegument layer, undergo secondary envelopment by budding into intracellular organelles, and are exocytosed (210) (section 1.5.3). The vast majority of mature HSV-1 virions remain cell-associated, and only very few are released from the cell surface (211).

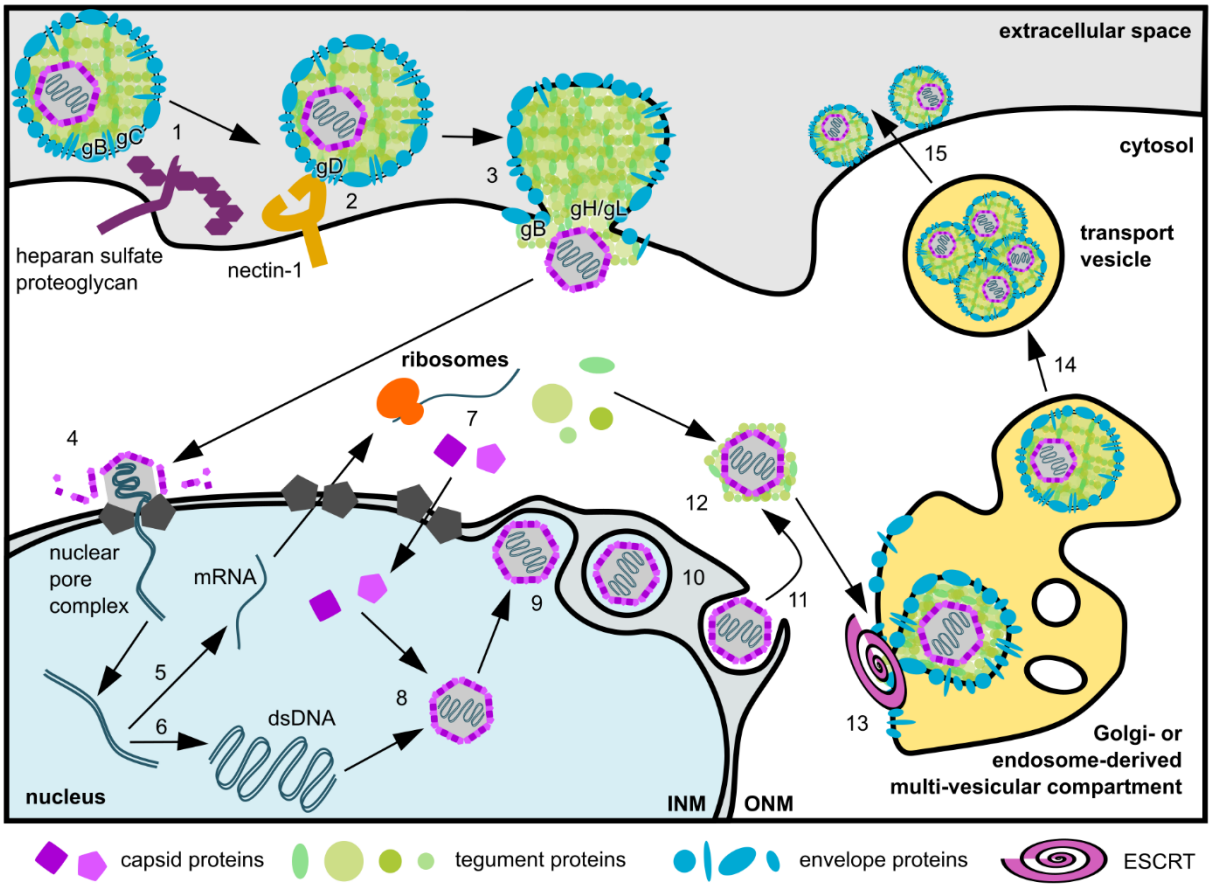


Figure 14 Life cycle of herpes simplex virus 1. 1: Glycoproteins gB and gC interact with heparan sulfate, mediating attachment of a virus particle to the host cell. 2: Glycoprotein gD binds to cell surface proteins such as nectin-1, initiating viral entry. 3: Glycoproteins gB and gH/gL facilitate fusion of the viral envelope with the cellular membranes and the nucleocapsid is released into the cytosol. 4: The nucleocapsid is transported to the nuclear pore complex, where capsid proteins dissociate and the viral DNA is injected into the nucleus. 5: Host cell DNA-dependent RNA polymerase II transcribes viral genomic DNA to mRNAs encoding viral proteins. 6: The viral DNA polymerase replicates the viral dsDNA genome. 7: Cellular machinery translates viral mRNAs and therefore synthesizes viral proteins. Capsid proteins are transported into the nucleus. 8: Capsid proteins and the viral genome assemble and form the nucleocapsid. 9: Primary envelopment. The nucleocapsid buds into the perinuclear space. 10: The nucleocapsid has acquired an envelope from the inner nuclear membrane (INM). 11: De-envelopment. The membrane of the nucleocapsid fuses with the outer nuclear membrane (ONM) and the nucleocapsid is released into the cytosol. 12: Cytosolic tegument proteins assemble around the nucleocapsid. 13: Re-envelopment/secondary envelopment. The tegument-coated nucleocapsids associates with proteins of the outer tegument layer and buds into a multi-vesicular compartment using the cellular ESCRT machinery, acquiring an enveloping membrane spiked with glycoproteins. 14: Transport vesicles filled with virions are trafficked to the plasma membrane. 15: Progeny virus is released into the extracellular space.

1.5.3 Current models of herpes simplex virus secondary envelopment and egress

After being released from the nucleus into the cytosol, a process referred to as primary envelopment/de-envelopment, cytosolic tegument proteins assemble around the nucleocapsid (207, 210). The viral envelope and HSV-1 glycoproteins are acquired when capsids bud into intracellular vesicles (re-envelopment, also known as secondary envelopment). Extensive

interactions are formed between tegument proteins themselves and between tegument proteins and envelope proteins (210), and therefore, assembly of the outer layers of the tegument might occur during envelopment (212).

It is thought that the compartments exploited by HSV-1 for secondary envelopment are derived from the trans-Golgi network (212-214) or endosomes (211, 215). These envelopment sites might seem controversial at first sight, given that in the "classical" view of cellular membrane trafficking the Golgi network and endosomes are key compartments for opposing directions of vesicular transport processes. However, trafficking within the endocytic and secretory pathway is highly dynamic (Figure 1), and herpesviruses extensively re-model the endomembrane system during late stages of infection, leading to redistribution of Golgi-, trans-Golgi and endosomal membranes (216, 217). Therefore, the compartment exploited during HSV-1 envelopment might be heterogeneous, explaining seemingly conflicting results obtained by tracking classical endosomal or Golgi markers.

The laboratory of Colin Crump has identified the cellular machinery facilitating secondary envelopment by demonstrating that components of the ESCRT machinery, specifically ESCRT III and Vps4, are required for envelopment of HSV-1 capsids in cultures of non-neuronal cells (218, 219). However, the ESCRT-I component Tsg101 and the ESCRT adaptor protein ALIX, which are known to bind to the viral late motifs PTAP and YPXL, respectively, are dispensable for ESCRT-mediated budding of HSV-1 (219). Therefore, Crump and colleagues suggested that recruitment of ESCRT complexes might be mediated by viral proteins such as UL56, UL52, UL19 or others, which contain PPXY motif(s), interacting with E3 ubiquitin ligases of the Nedd4 family (219). Indeed, interactions between HSV-2 UL56 and the ubiquitin ligases NEDD4 and ITCH has been previously demonstrated (220-222), but HSV-2 UL56 is dispensable for virus growth in cultured cells (221). Therefore, it seems likely that multiple HSV-1 proteins recruit the cellular ESCRT machinery redundantly and/or cooperatively to mediate virus envelopment.

Once HSV-1 capsids have budded into an intracellular compartment, virion-filled vesicles have to be trafficked to the plasma membrane to release infectious virus particles into the extracellular space. HSV-1 infects both endothelial and neuronal cells, and egress mechanisms probably vary between those cells. During lytic infection of endothelial cells, virus particles have to be transported to the surface of the host cell to accumulate in mucosal blisters, to infect neighboring

cells, or to spread from infected mucosa to sensory neuron endings (Figure 15A). Another cellular transport pathway is probably exploited when HSV-1 travels anterogradely from the soma of neuronal cells to the distal tips of axons, for example during reactivation (Figure 15B). Two mechanism of anterograde transport in axons have been proposed: the "married" and the "separate" model (223). In the "married" model, viral particles bud into intracellular compartments located in the neuronal cell body and are trafficked inside vesicles to the axon tip, whereas in the "separate" model, HSV-1 envelope proteins and the nucleocapsid are transported to the axon tip separately and assemble only there to form complete virions. Ultrastructural and immunofluorescence studies yield conflicting results and coordinated experiments are needed to elucidate the physiological relevance of these two models in cultured cells and in vivo (223).

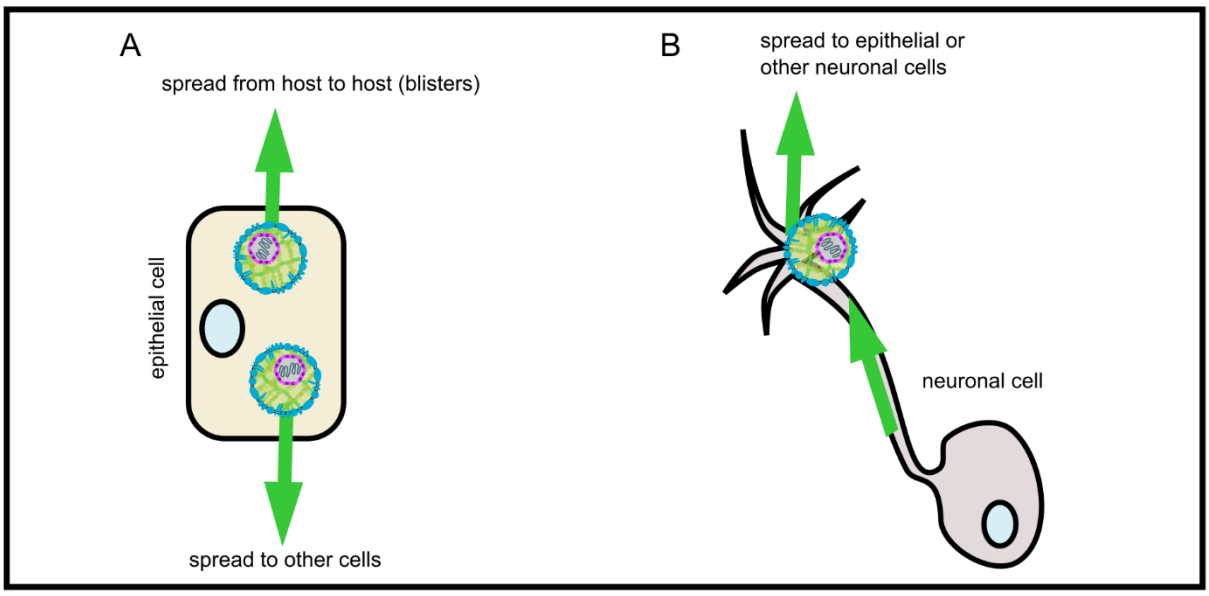


Figure 15 Herpes simplex virus 1 infects and therefore spreads from different cell types. (A) HSV-1 particles can egress at the membrane of epithelial cells to infect cells adjacent to the epithelium or to accumulate in mucosal blisters, which burst and lead to virus spread from host to host. (B) HSV-1 can also spread from neuronal cells to epithelial cells or the central nervous system, and therefore, virions have to be transported anterogradely inside axons and egress at axon tips.

Taken together, our current knowledge about egress of HSV-1 particles after envelopment is incomplete. Several host proteins such as Rab GTPases (210) or kinesins (204) have been implicated in this process, and various viral tegument and envelope proteins could potentially hijack cellular machinery to facilitate trafficking of virus particles (see Chapter 5 for an overview). In this thesis, three HSV-1 proteins previously linked to viral egress — glycoprotein E, envelope protein UL56, and tegument protein UL21 — were selected and studied in more detail.

1.5.4 HSV-1 membrane protein UL56

The HSV-1 protein UL56 is 234 amino acids in length and anchored to the viral envelope via its C-terminal transmembrane domain (224). Homologs of UL56 can be found throughout the subfamily of alphaherpesviruses, but the sequence is highly divergent even between HSV-1 and HSV-2 (225). UL56 was first described in the early 1990s as a virulence factor in intraperitoneal rodent models of infection: HSV-1 strain HFEM, which does not express UL56 naturally, and strain F carrying a deletion including UL56, did not exhibit intraperitoneal pathogenicity in tree shrews and mice (226, 227), even though replication in cell culture was normal (227). The importance of UL56 as a virulence factor was questioned by Nash and Spivack (228). However, it was later demonstrated that re-insertion of UL56 from strains F or 17 into strain HFEM restored pathogenicity and that the transmembrane domain of UL56 is necessary for pathogenicity (229).

Most of the more recent work on UL56 has been performed on the protein from HSV-2. Koshizuka and colleagues demonstrated that UL56 is a potentially phosphorylated, tail-anchored, type II transmembrane protein (224). It is expressed from 6 hours post infection onwards (221, 222, 224, 230) and can be found in intra- and extracellular virions (196, 231, 232). In cells infected with HSV-2, UL56 is found in the cytosol, first in a punctate pattern, but at later time points also with diffuse localization. When transfected into cells, UL56 localizes mainly to a compact perinuclear compartment, but is also found throughout the cytoplasm in both diffuse and speckled patterns (224). Deletion of the transmembrane domain of UL56 leads to its mislocalization to the nucleus (224, 233), whereas deletion of the N-terminal 20 residues does not affect its distribution within the cell (224). Herpes simplex virus 2 carrying a complete deletion of the UL56 locus replicates normally in cells (221), indicating that UL56 is a non-essential gene. Nonetheless, it has also been suggested that deletion of UL56 decreases the release of infectious particles from cells even when the amount of intracellular virus is not affected (220), hinting towards a potential involvement of UL56 in trafficking or release of virus particles from host cells.

Protein UL56 possesses a large cytosolic domain (amino acids 211-233) that is predicted to be unstructured. Therefore, it is conceivable that it binds to both viral and cellular proteins to facilitate assembly of the virion or to exploit cellular functions. Indeed, GST pull-down and co-

localization experiments show that UL56 interacts with the membrane-associated tegument protein UL11 (230) and the cellular kinesin motor protein KIF1A (233). Intriguingly, a yeast two-hybrid screen performed by Ushijima et al. in 2008 identified several HECT domain-containing E3 ubiquitin ligases, including NEDD4 and ITCH, as potential binding partners of HSV-2 UL56 (221). Follow-up experiments demonstrated that both NEDD4 and ITCH both co-immunoprecipitate with UL56 in cells transfected with UL56 or infected with HSV-1. Both E3 ligases were also found to co-localize with UL56 in transfected cells and during infection (220-222).

The N-terminal domain of UL56 contains three late domain PPXY motifs which are known to facilitate binding to E3 ubiquitin ligases and have been strongly associated with viral budding (36, 118). By mutating these motifs in UL56, Ushijima et al. demonstrated them to be necessary for the interaction and co-localization with NEDD4 (220, 221). Interestingly, both NEDD4 and ITCH were found to be rapidly degraded in the presence of HSV-2 infection, probably via the proteasomal degradation pathway (221, 222). ITCH was also found to be partially degraded via the lysosome (222). On the contrary, during infection with HSV-1, NEDD4 and WWP2, a further E3 ubiquitin ligase of the NEDD4 family, are not downregulated (222). It should be noted that UL56 does not contains any lysine residues and therefore, cannot itself be ubiquitinated. Ushijima and colleagues concluded that HSV-2 UL56 regulates and relocalizes the two ubiquitin ligases NEDD4 and ITCH, probably in order to control virus budding or transport within the host cell.

In contrast with the potential involvement of UL56 in intracellular trafficking, a recent study using PRV revealed that deletion of UL56 does not affect virus replication or spread in epithelial cells and that UL56 is not required for retrograde axonal transport in cultured dorsal root ganglion neurons (234). In addition, it was shown that UL56 is not necessary for anterograde axonal transport in cultured neurons or in a mouse model (234). However, the same paper also demonstrated that UL56 is a virulence factor in vivo since survival times of mice were prolonged when infected with UL56 deletion vs. wild-type virus. Yet another potential function of UL56 was highlighted when studies on equine herpesviruses 1 and 4 discovered that cell surface proteins, including the major histocompatibility complex class I (MHC I), are downregulated in cells infected with these viruses in a UL56-dependent manner (235-237).

In summary, no coherent model describing the function of UL56 during virus infection has emerged so far. Studies are complicated by differential behavior of UL56 in cell lines of epithelial or neuronal origin, and it seems that even though UL56 is conserved across the alphaherpesviruses (Figure 16), results obtained from work on HSV-2, PRV or Equine Herpesviruses cannot be straightforwardly translated to HSV-1. Nonetheless, there is strong evidence for UL56 being a virulence factor in vivo, and thus, investigation of its interaction with host cell proteins presents a promising avenue for further studies.

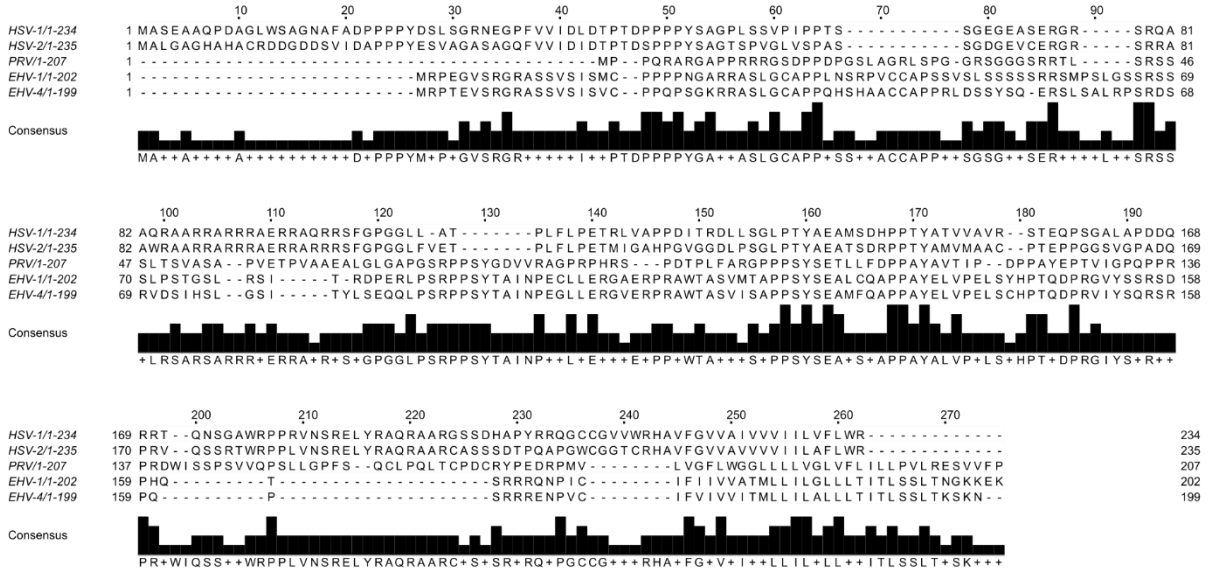


Figure 16 Sequence alignment of UL56 from HSV-1, HSV-2, PRV, EHV-1 and EHV-4 showing conservation of UL56 across these different alphaherpesviruses.

1.5.5 HSV-I glycoprotein gE

Glycoprotein gE, also known as US8, is a 550 amino acid transmembrane protein and one of the glycoproteins that constitute the HSV-1 envelope. Protein gE binds to glycoprotein gI, forming a heterodimer which serves as an Fc receptor for host IgG antibodies (238, 239). The transmembrane domain of gE spans approximately residues 422-444. The N-terminal domain is extracellular and contains a β-sheet rich, Ig-like Fc-binding domain (240), whereas the C-terminal domain is predicted to be largely unstructured and located in the cytosol. The gE/gI dimer is known to interact with the membrane protein US9 via gE, forming a tripartite complex (241).

One of the first observations hinting towards a role of gE/gI in HSV-1 cell-to-cell spread or axonal transport was made by studying mutant viruses lacking gE or gI, which exhibit a small plaque phenotype in cultured cells and are less pathogenic in mouse ocular infection models (242). Herpes simplex virus infects and replicates in epithelial cells during the lytic phase of its life cycle, but during latency, resides in neuronal cells. Therefore, it soon became clear that mechanisms for envelopment, intracellular transport and viral spread may differ between these cell types and studies had to be designed carefully to separate the context-dependent functions of gE/gI.

In epithelial tissues, the gE/gI dimer is reported to participate in the sorting of virions to lateral cell surfaces and the complex binds to cell junction proteins, thereby contributing to virus cell-to-cell spread (243, 244). In neuronal cells, gE/gI work together with US9 to sort viral glycoproteins and assembled virions into axons (245, 246). The complex is thought to stabilize the interaction between US9 and the kinesin motor protein KIF1A (247, 248), thereby facilitating transport of virions from cell soma into proximal axons and virus spread to non-neuronal cells (249). Interestingly, even though the extracellular domains of gE/gI are located inside the lumen of cytoplasmic vesicles and thus cannot interact with cytoplasmic proteins, they have been implicated in cell-to-cell spread (250), axonal transport and spread from neurons to epithelial cells (251). Such interactions could, for example, occur via binding of the extracellular domain of gE to luminal domains of cellular transmembrane proteins. Most recently, it was shown that HSV-1 viruses lacking both gE and US9 accumulate as non- or incompletely enveloped particles in the cytosol of neuronal cells, but not during infection of keratinocytes (252). Such a defect in virus particle envelopment provides an alternative explanation for the loss of anterograde axonal transport observed for HSV-1 lacking gE. Taken together, these studies show that the gE/gI complex plays a vital role during axonal sorting, anterograde axonal transport and cell-to-cell spread of HSV-1 particles.

1.5.6 HSV-I tegument protein UL21

The 535 amino acid protein UL21 is a tegument protein of HSV-1. It is conserved across alpha-, beta- and gammaherpesviruses (253) and is incorporated into extracellular virions (196). Its three-dimensional structure was solved recently and shows UL21 to consist of N-terminal and C-terminal folded domains (see section 5.3) (254, 255). Multiple studies assessing the phenotype of

HSV-1 lacking UL21 in cell culture have been conducted, but these studies reach conflicting results. In 1994, Baines and colleagues did not observe a replication defect in Vero cells infected with HSV-1 missing the first 484 amino acids of UL21. More recent work suggests that deletion of UL21 delays the production of infectious virions, but leads only to a slight reduction of virus titer after 24 hours (256). Yet another report demonstrated UL21 deletion viruses to have reduced intra- and extracellular titers of infectious virions in a human glioblastoma cell line, but not in Vero cells at five days post infection (257). The same study also observed a small plaque phenotype for the UL21 deletion virus in both Vero and glial cells. Interestingly, Le Sage and colleagues showed that, in contrast to HSV-1 and PRV, UL21 is essential for HSV-2 replication in cell culture (258).

Not much is known about the function of UL21, but it is likely that is contributes to virulence in vivo. For example, in the alphaherpesvirus PRV, UL21 highly increases virulence, since one-week-old piglets infected with PRV strain Bartha, harboring a UL21 gene which encodes an inactive protein due to several point mutations, develop no disease, whereas wild-type virus-infected pigs die within eight days (259). In agreement with this report, PRV Bartha has impaired neuronal spread in rat models of infection (260). In transfected cells, HSV-1 UL21 localizes predominantly to the nucleus, but can also be found in the cytoplasm (261). Upon overexpression in cultures of non-neuronal cells, Takakuwa et al. observed co-localization of UL21 with microtubules and formation of long, neuron-like processes (262). However, such stark changes in cell morphology upon transfection with UL21 have not been reported by any other group. Protein UL21 has also been implicated in nuclear egress (HSV-2) (258) and is involved in packaging of other viral proteins into virions (PRV) (263). Co-immunoprecipitation experiments demonstrated that UL21 interacts with the tegument proteins UL16 and UL11 in both HSV-1 and PRV (264, 265). This complex has also been reported to associate with the cytoplasmic tail of glycoprotein gE, supporting and directing its cellular trafficking and functions (261). Additionally, UL21 has also been suggested to bind RNA (254). Together, it is evident that although UL21 has been implicated in several processes that may contribute to virulence or viral spread, its exact role during the viral life cycle and its interplay with host cell proteins are largely unclear.

1.6 Project aims

In this thesis, I study the egress of hepatitis D virus and herpes simplex virus 1 by investigating interactions between viral and host cell proteins. Clathrin-mediated intracellular trafficking has been previously implicated in the production of HDV virus-like particles (106, 107, 177). It was shown that an interaction between clathrin NTD and a short, linear peptide motif present in the C-terminal extension of the large antigen encoded by HDV is necessary for the egress of virus-like particles from cells. However, it is not known how exactly HDAg-L and NTD interact. In particular, it needs to be investigated whether HDAg-L binds to the same sites on NTD as cellular proteins, and whether the binding affinities of cellular and viral proteins for NTD are comparable. Binding of viral peptides at different sites of NTD than those occupied by cellular binding motifs could present opportunities for structure-guided design of drugs that specifically interfere with the viral life cycle without affecting cellular clathrin trafficking. In addition, previous functional (2) and in vitro studies (266) hinted towards degeneracy of clathrin binding by peptide motifs. However, high-resolution structural evidence for promiscuous binding of peptides to NTD is not available, and it has not been proved whether and to what extent binding to different sites on NTD contributes to overall NTD:peptide interactions.

Only very few interactions between HSV-1 and cellular proteins involved in intracellular trafficking have been described. Therefore, our understanding of the late stages of the HSV-1 life cycle is limited and cellular pathways that HSV-1 could exploit to facilitate envelopment or egress of viral particles from host cells have yet to be identified. Investigation of protein:protein interactions between HSV-1 and host cell proteins will not only help to elucidate the mechanisms employed by the virus during host cell exit, potentially highlighting novel targets for treatment of lytic or latent HSV-1 infection, but could also further our knowledge about basic cell biology.

I therefore address the following questions in this thesis:

What is the molecular basis of interactions between clathrin and HDV large antigen?

I solved the crystal structure of NTD in complex with short viral peptides derived from the large antigen of HDV to assess which binding sites of NTD are occupied by the viral proteins. Surprisingly, peptide binding could be observed at two or three interaction sites of NTD simultaneously. Therefore, I investigated using biochemical binding assays whether interactions observed in crystallo can be confirmed in vitro.

Do cellular CBM peptides bind to multiple sites on NTD?

I assessed binding between clathrin NTD and cellular peptides derived from AP2 or amphiphysin by X-ray crystallography. As observed for the viral peptides, and in contrast to previously published crystal structures of NTD in complex with peptides, I observed that cellular peptides bound NTD promiscuously. I therefore designed and characterized a panel of NTD proteins containing mutations at one, two, three, or all four binding sites of NTD and used these proteins in biochemical or biophysical binding assays to probe these observations and to disentangle the contribution of binding at the different sites to the overall interaction. Taken together, these experiments enabled me to structurally characterize the fourth binding site on NTD and shed new light on the molecular basis of clathrin recruitment by viral and cellular proteins.

Which cellular proteins or machineries does HSV-1 protein UL56 hijack to facilitate envelopment or egress of viral particles?

Based on my study of the current literature and bioinformatic analysis of selected HSV-1 proteins, I decided to investigate the cellular interactomes of glycoprotein E, tegument protein UL21, and envelope protein UL56 via unbiased interaction screens. Since gE and UL21 proved to be unsuitable for yeast two-hybrid screening due to autoactivation, only UL56 was investigated further. Using two different unbiased screening approaches, yeast two-hybrid and affinity

purification coupled with quantitative mass spectrometry, I identified multiple previously undescribed cellular binding partners of UL56. Confirming some of these novel interactions by co-immunoprecipitation and GST pull-down experiments, I present exciting hypotheses ascribing Golgi trafficking proteins, as well as the Nedd4-family of ubiquitin ligases and the ESCRT system functional roles during HSV-1 envelopment and egress.

2 MATERIALS AND METHODS

2.1 General procedures

2.1.1 Buffers

All buffers were made up using 0.2 µm filtered, distilled water demineralized to a resistivity greater than 18.2 MΩcm (MQW) and were again 0.2 µm filtered before use. Buffers containing dithiothreitol (DTT) were used within 48 h of preparation.

2.1.2 Quantification of protein and DNA

Unless otherwise indicated, the concentration of protein was determined by measuring the absorbance at 280 nm and calculating the protein concentration according to the Beer-Lambert law. Theoretical extinction coefficients were calculated using the ExPASy protparam online service (267). Measurements of protein concentrations were performed in triplicate and the mean was calculated.

Bicinchoninic acid (BCA) assays were performed using a BCA protein reagent kit (Pierce). Standard curves of protein concentrations were obtained from dilution series of bovine serum albumin (BSA). Protein standards and samples were incubated in clear 96-well plates for 30 min at 37 °C with 200 µL of BCA reagent prepared according to the manufacturer's instructions. Absorption was measured at 570 nm on a SpectraMax i3 spectrophotometer (Molecular Devices).

DNA concentrations were determined by measuring UV absorbance at 260 nm.

2.1.3 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Gels for electrophoresis were prepared as indicated in Table 1 and Table 2 and stored at 4 °C for a maximum of one week. A 5x SDS-PAGE buffer stock (0.25 M Tris-HCl pH 6.8, 0.25% bromophenol blue, 25% glycerol, 10% sodium dodecyl sulfate) was kept at ambient temperature. Per 1 mL, 50 µL of 2-mercaptoethanol or 1 M DTT were added to the 5x stock as a reducing agent before use. Samples were diluted in 1x SDS-PAGE buffer, denatured at 95 °C for 5 min and applied to the polyacrylamide gel. Samples were separated at 200 V, 20-50 mA/gel for 30 to 60 min. Separated proteins were visualized by staining with Instant Blue Coomassie stain (Stratech Scientific or Expedeon) or immunoblotting (Western blotting, section 2.1.4).

Table 1 Recipes for pouring two 0.75 mm thick gels for electrophoresis.

Acrylamide percentage	**9%**	**10%**	**12%**	**15%**	**18%**
MQW (mL)	4.3	4.0	3.3	2.3	1.3
1.5 M Tris.HCl pH 8.8 (mL)	2.5				
30% acrylamide (37.5:1 acrylamide:bis-acrylamide) (mL)	3.0	3.3	4.0	5.0	6.0
10% SDS (µL)	100				
10% ammonium persulfate (APS) (µL)	100				
tetramethylethylenediamine (TEMED) (µL)	10				

Table 2 Stacking gel recipe for two 0.75 mm thick stacking gels.

	all gel percentages
4% stacking gel stock (mL)	2
10% APS (µL)	10
TEMED (µL)	10

Table 3 Stacking gel stock recipe.

Acrylamide percentage	**4%**
MQW (mL)	60.7
0.5 M Tris.HCl pH 6.8 (mL)	25.0
10% SDS (mL)	1.0
30% acrylamide (37.5:1 acrylamide:bis-acrylamide) (mL)	13.3

2.1.4 Immunoblotting

Immunoblotting, in this thesis referred to as Western blotting, was performed using the wet transfer method with the Towbin buffer (25 mM Tris-HCl pH 8.3, 192 mM glycine, 20% methanol) (268). Gels were blotted onto Protran nitrocellulose membranes with a pore size of 0.45 µM (PerkinElmer) at either 100 V for 60 min, 70 V for 180 min, or at 18 V for 18-24 h. The membranes were blocked for 2-72 h using 3% bovine serum albumin (BSA) in 1x phosphate-buffered saline (PBS) (for membranes to be probed with the α-His antibody) or 5% (w/v) skim milk powder in 1x PBS (5% milk, all other membranes). The membranes were then incubated with primary antibody for 1.5-72 h. The antibodies used in this thesis and the dilution used for Western blotting are listed in Table 4. All primary antibody dilutions were prepared in 5% milk in 1x PBS supplemented with 0.05% Tween-20 (1x PBS-T), except for the α-His antibody, which was prepared in 3% BSA in 1x PBS-T. After the primary antibody incubation, membranes were washed three times for at least 5 min with 5% milk in 1x PBS-T. Membranes were then incubated for 30-60 min with suitable species-reactive near-infrared fluorescent LI-COR secondary

antibodies (1:10000 dilution in 5% milk in 1x PBS-T), washed twice for 5 min with 5% milk in 1x PBS-T and once for 5 min in 1x PBS-T. Three final rinses in 1x PBS were then performed, the membranes dried and imaged using an Odyssey LI-COR machine. Some membranes incubated with α-clathrin NTD primary antibody were visualized using chemiluminescence. These membranes were incubated with a rabbit α–mouse horseradish peroxidase-conjugated secondary antibody (1:10000 in 5% milk in 1x PBS-T, Sigma), washed and rinsed as above, and incubated with enhanced chemiluminescence reagents (Geneflow). Membranes were exposed onto Fuji RX NF X-Ray film, developed and scanned.

Table 4 Antibodies used for Western blotting (WB).

Antibody	Description	Source	Dilution for WB	Reference
α-clathrin NTD	monoclonal	mouse	1:1000	abcam, ab11331
α-His	monoclonal	mouse	1:1000	Qiagen, 34660
α-myc	monoclonal	mouse	1:4000	Millipore, 05-724
α-GFP	polyclonal	rabbit	1:5000	Sigma, G1544
α-GAPDH	monoclonal	mouse	1:20000	Ambion, AM4300
α-BECN1	monoclonal	mouse	1:500	abcam, ab114071
α-SENP6	monoclonal	mouse	1:500	Abnova, H00026054-M01
α-WDR61	monoclonal	mouse	1:500	abcam, ab57840
α-FNIP2	polyclonal	rabbit	1:500	Sigma, SAB3500007
α-GOPC	monoclonal	rabbit	1:1000	Millipore, MABC731
α-PAWR	monoclonal	rabbit	1:1000	abcam, ab109174
α-TRAPPC9	polyclonal	rabbit	1:1000	NovusBio, NB100-55735
α-TRAPPC10	monoclonal	mouse	1:500	Abnova, H00007109-M01
α-TRAPPC10*	polyclonal	mouse	1:500	Abnova, H00007109-B01P
α-NEDD4	polyclonal	rabbit	1:10000	abcam, ab14592
α-NEDD4L	monoclonal	rabbit	1:10000	abcam, ab131167

2.1.5 Data analysis, molecular graphics and figures

Data analysis and graph plotting were performed using MATLAB (Mathworks) and Prism 7 (GraphPad Software). Molecular graphics were generated using PyMOL (269) and ligplot+ (3). Figures were assembled using Inkscape (https://inkscape.org/).

2.2 Molecular biology

2.2.1 Generation of competent *E. coli* cells

100 mL of lysogeny broth (LB) medium were inoculated with 1 mL of an overnight culture of competent *E. coli* and incubated at 37 °C, 200 rpm until an optical density (A_{600}) of 0.3 was

reached. The cells were cooled on ice for 15 min and then pelleted (3300 g, 10 min, 4 °C). The supernatant was discarded, the cell pellet resuspended in 40 mL ice-cold 0.1 M $CaCl_2$ and incubated for further 30 min on ice. The cells were pelleted again (3300 g, 10 min, 4 °C) and the pellet resuspended in 6 mL ice-cold 75 mM $CaCl_2$, 15% glycerol. Cells were frozen as 50 µL aliquots on dry ice and stored at -80 °C.

2.2.2 Transformation and liquid cultures of bacterial cells

Plasmid DNA (approximately 50 ng for purified and sequence-verified plasmids) was added to 50 µL of competent *E. coli* cells and the cells incubated on ice for 30 min. Cells were heat-shocked for 40 s in a 42 °C water bath and immediately placed on ice again for 2 min. Cells were then recovered in 200 µL 2xTY medium for 60 min (200 rpm, 37 °C) and subsequently plated onto 2xTY agar plates supplemented with suitable antibiotics (100 µg/mL of ampicillin and/or 25 µg/mL of chloramphenicol and/or 50 µg/mL of kanamycin). Plates were incubated for 16-24 h at 37 °C or up to 72 h at 20 °C.

Liquid bacterial cultures were grown in 2xTY medium supplemented with suitable antibiotics (concentrations as above) at 37 °C, 200 rpm unless otherwise indicated.

2.2.3 Agarose gel electrophoresis

Agarose gels were prepared by dissolving 0.5-2.0% agarose in 45 mM Tris-borate, 1 mM ethylenediaminetetraacetic acid (EDTA) (0.5x TBE) by boiling. Gels were stained with 5 µL SYBRsafe DNA gel stain (Life Technologies) per 50 mL. DNA samples were mixed with loading dye (0.04% (w/v) Orange G, 1 mM EDTA pH 8.0, 5% (v/v) glycerol) and separated by electrophoresis (100 V, 100 mA for approximately 60 min) in 0.5x TBE buffer. DNA was visualized using UV light or visible blue light if the DNA was used for downstream applications.

2.2.4 Purification of plasmid DNA, PCR products and agarose gel fragments

For preparations of plasmid DNA, competent DH5α E. *coli* were transformed with approximately 50 ng of plasmid DNA and incubated on agar plates supplemented with suitable antibiotics. Suspension cultures of the transformed bacteria (5 mL for small-scale preparations, 100 mL for medium-scale preparations) were grown for 16-20 h and subsequently pelleted (10 min, 4000 g, 4 °C). Plasmid DNA was purified using variations of the alkaline lysis method (270). Small-scale plasmid DNA preparations, as well as polymerase chain reaction (PCR) product and excised

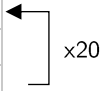
agarose gel fragments were purified using pre-packed silica membrane "EconoSpin" columns (Epoch Life Science) following the manufacturer's protocol. Samples of purified plasmid DNA were eluted in 70-100 µL nuclease-free water, whereas purified PCR products and gel-extracted DNA samples were eluted in 20 µL nuclease-free water. Medium-scale plasmid DNA preparations were purified using a PureLink HiPure Plasmid Filter Midiprep Kit (ThermoFisher) or a Hi-Speed plasmid midi kit (Qiagen) following the manufacturer's instructions. Sanger sequencing (Biochemistry Sequencing Facility, University of Cambridge) was used to verify the sequence of plasmids.

2.2.5 PCR amplification of DNA

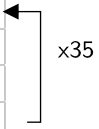
Oligonucleotides were obtained from Sigma-Aldrich. DNA from plasmid templates or from HeLa cell cDNA (kindly provided by Yvonne Hackmann) was amplified using KOD Hot Start DNA polymerase. Reactions were assembled and the PCR was executed as below. For amplifications from cDNA, a range of annealing temperatures was scanned using a gradient PCR cycler. The elongation time was adjusted to the size of the product to be amplified. For products smaller than 500 bases (b), an elongation time of 10 s/kb was chosen; for products between 0.5-1 kb, the elongation time was 15 s/kb; for products between 1-3 kb, the elongation time was 20 s/kb; and for products larger than 3 kb, the elongation time was 25 s/kb. In some cases, 5-10% dimethyl sulfoxide (DMSO) or 1 M betaine were added to the reactions to prevent formation of primer secondary structures or primer dimers. Samples were purified and/or treated with 1 µL of the methylated DNA nuclease Dpn1 per reaction before further use.

Per reaction:	
5 µL	10x KOD buffer
3 µL	25 mM $MgSO_4$
0.5 µL	25 mM dNTP mix
1 µL	10 ng/µl template *OR* cDNA
1.5 µL	forward primer
1.5 µL	reverse primer
36.5 µL	nuclease-free water
1 µL	KOD Hot start DNA polymerase
50 µL	total

PCR cycle for amplification of plasmid DNA:		
	2 min	95 °C
melting	20 s	95 °C
annealing	10 s	55 °C
elongation	_ s	70 °C
	hold	10 °C



PCR cycle for amplification from cDNA:		
	2 min	95 °C
melting	20 s	95 °C
annealing	10 s	55-65 °C (gradient)
elongation	_ s	70 °C
	hold	10 °C



2.2.6 Restriction digest of DNA

Restriction endonucleases were purchased from New England Biolabs (NEB) or Roche. Restriction digests of vectors and inserts were performed in a final volume of 50 µL following manufacturer's instructions. Generally, DNA was incubated with restriction enzymes for 30-60 min at 37 °C. Digest buffers were chosen according to the specific enzymes used. Samples digested with Sgf1 and Pme1 were incubated at 65 °C for 20 min after the restriction digest to heat-inactivate the enzymes. All cut inserts were PCR-purified. Cut vectors were treated with 1 µL alkaline shrimp phosphatase (Roche) for 60 min at 37 °C to prevent vector re-ligation, analyzed on agarose gels and subsequently gel-purified before further use.

2.2.7 Ligation of DNA

Generally, ligation reactions were performed using a 6:1 molar ratio of insert:vector with a total amount of DNA of approximately 100 ng. Appropriate amounts of insert and vector were mixed and nuclease-free water added to obtain a reaction volume of 8.5 µL. Samples were incubated at 42 °C for 5 min and then immediately placed on ice. 1 µL of 10x T4 DNA ligase buffer (NEB) and 0.5 µL T4 DNA ligase (NEB) were added to the mixture and samples incubated for 2 h at 16 °C, or for up to 24 h at 14 °C. 50 µL of DH5α competent *E. coli* cells were transformed with 3-6 µL of the ligation mixture. If high numbers of colonies were observed on negative control plates assessing vector re-ligation, colonies were screened by bacterial colony PCR (2.2.10) to identify positive clones. All plasmids were purified and sequence-verified before further use.

2.2.8 Annealed oligonucleotide insertion of small DNA fragments

For the insertion of small DNA fragments encoding short tags or peptide sequences, complementary oligonucleotides with 5' and 3' overhangs suitable for ligation into the cut target vector were designed. Primers were first phosphorylated and annealed as below. The phosphorylated primers were diluted ten-fold using nuclease-free water and ligation was performed using 0.5 µL diluted, annealed primers (150 fmol) and 25 fmol of vector cut with appropriate enzymes.

Per reaction:	
5 µL	10x T4 DNA ligase buffer (NEB)
40 µL	nuclease-free water
1.5 µL	100 µM forward primer
1.5 µL	100 µM reverse primer
2 µL	DNA polynucleotide kinase (NEB)
50 µL	total

PCR cycle for phosphorylation and annealing:	
30 min	37 °C
5 min	98 °C
cool as slow as possible to	10 °C

2.2.9 Site-directed mutagenesis

Primer pairs for site-directed mutagenesis were designed using the PrimerX online tool (http://www.bioinformatics.org/primerx/). A primer mix containing 18 µL nuclease-free water, 1 µL of 100 µM forward primer and 1 µL 100 µM reverse primer was prepared. Reactions were assembled and the PCR cycling conducted as below. Extension times were adjusted to the size of the plasmid using 90 s per kb. Samples were then digested at 37 °C for 120 min using 1 µL Dpn1 per reaction to remove *Dam*-methylated template DNA. Control PCR reactions without primers were prepared to assess the efficiency of the Dpn1 digest. 50 µL DH5α cells were transformed with 1-3 µL of the digested reaction mix, a single colony picked, the plasmid purified and insertion of the desired mutation confirmed by Sanger sequencing.

Per reaction:	
5 µL	10x PFU buffer (Agilent)
0.5 µL	25 mM dNTP mix
2 µL	50 mM $MgCl_2$
1 µL	DMSO
2 µL	10 ng/µL plasmid template
2 µL	primer mix
36.5 µL	nuclease-free water
1 µL	PFU turbo DNA polymerase (Agilent)
50 µL	total

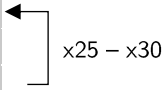
PCR cycle:		
30 s	98 °C	
30 s	98 °C	x5
60 s	52 °C	
_ s	68 °C	
30 s	98 °C	x7
60 s	55 °C	
_ s	68 °C	
300 s	68 °C	
hold	10 °C	

2.2.10 Bacterial colony PCR

Forward and reverse PCR primers annealing to the vector backbone, but not the insert to be confirmed, were chosen. Colonies to be screened were picked and each swirled into 20 µL nuclease-free water. Reaction mixtures were assembled as below and distributed into separate tubes (10 µL per colony to be screened). 2.5 µL of the colony suspension were added to the reaction mix and PCR cycling carried out as below. The annealing temperature was selected based on the melting temperatures of the forward and reverse primers. Generally, the annealing temperature was 5 K lower than the lower of the two primer melting temperatures. The elongation time was adjusted to the length of the insert to be confirmed using 90 s per kb. Samples were analyzed by agarose gel electrophoresis to confirm the presence of absence of a PCR product of the correct size, indicating successful insertion of the desired DNA sequence.

Per 10 reactions:	
61.25 µL	nuclease-free water
12.5 µL	10x Taq buffer (- $MgCl_2$) (Invitrogen)
5 µL	50 mM $MgCl_2$
3.75 µL	10 µM forward primer
3.75 µL	10 µM reverse primer
1 µL	25 mM dNTP mix
1.25 µL	Taq polymerase (Invitrogen)
100 µL	total

PCR cycle:		
	5 min	94 °C
melting	15 s	94 °C
annealing	15 s	_ °C
elongation	_ s	72 °C
	5 min	72 °C
	hold	10 °C



2.3 Construct design and cloning

Constructs were designed and cloned as listed in Table 5 using the primers listed in Table 6.

Vector pOPT3G is a bacterial expression vector encoding a 3C protease-cleavable GST-tag N-terminal to the cloning site (271). Vectors pOPTH and pOPTnH are modified pOPT vectors encoding non-cleavable N- and C-terminal hexahistidine tags. Vector pF3A-WG is an in vitro wheat germ protein expression vector with ampicillin drug selection (pF3A WG (BYDV) Flexi® Vector, Promega). Vectors pF3A-myc-WG and pF3A-myc-s were derived from the pF3A-WG vector to encode a non-cleavable myc-tag N-terminal to the modified cloning site. Vector pF5K-myc is a mammalian protein expression vector derived from pF5K CMV-neo (Promega) containing a non-cleavable myc-tag N-terminal to the cloning site. pEGFP-N1 and pEGFP C1 are mammalian protein expression vectors encoding eGFP-tags C- and N-terminally to the cloning site, respectively.

The following constructs were kindly provided by others: GST-NTD, VPS18 beta-propeller, His-alpha-SNAP by Stephen C. Graham (University of Cambridge, UK); GST-AP2-HA, His-c-µ2, GST-YXXΦ by Bernard T. Kelly (Cambridge Institute for Medical Research, UK), His-NTD-NEMO (272), GST-Amph4T1 by Linton M. Traub (University of Pittsburgh, USA); BBRF2-His, His-ORF55 by Danielle J. Owen (University of Cambridge); pGBKT7-UL21, pGBKT7-gE,

pGBKT7-UL56(1-215), eGFP-UL56(1-234), UL56(1-234)-eGFP by Colin M. Crump (University of Cambridge).

Table 5 Constructs designed for this thesis.

Construct	Cloning method	Source of DNA / source construct	Target vector	Primers
GST-HDAg-L1	annealed oligo insertion into target vector, followed by site-directed mutagenesis to introduce missing stop codons	primers	pOPT3G cut with Nde1/BamH1	5, 6 and 11, 12
GST-HDAg-L2	annealed oligo insertion into target vector, followed by site-directed mutagenesis to introduce missing stop codons	primers	pOPT3G cut with Nde1/BamH1	7, 8 and 13, 14
GST-HDAg-L3	annealed oligo insertion into target vector, followed by site-directed mutagenesis to introduce missing stop codons	primers	pOPT3G cut with Nde1/BamH1	9, 10 and 15, 16
GST-HDAg-L2-2	annealed oligo insertion into target vector	primers	pOPT3G cut with Nde1/BamH1	17, 18
GST-HDAg-L2-3	annealed oligo insertion into target vector	primers	pOPT3G cut with Nde1/BamH1	19, 20
GST-HDAg-L2-4	annealed oligo insertion into target vector	primers	pOPT3G cut with Nde1/BamH1	21, 22
GST-HDAg-L2-C251S	site-directed mutagenesis	GST-HDAg-L2	-	23, 24
GST-HDAg-L2-C251Stop	site-directed mutagenesis	GST-HDAg-L2	-	25, 26
GST-AP2CBM	annealed oligo insertion into target vector	primers	pOPT3G cut with Nde1/BamH1	27, 28
GST-AmphCBM	site-directed mutagenesis	GST-AmphCBM-long	-	33, 34
GST-AmphCBM-long	annealed oligo insertion into target vector	primers	pOPT3G cut with Nde1/BamH1	29, 30
GST-Wbox	annealed oligo insertion into target vector	primers	pOPT3G cut with Nde1/BamH1	31, 32

Construct	Cloning method	Source of DNA / source construct	Target vector	Primers
His-NTD-NEMO Q89M	site-directed mutagenesis	His-NTD-NEMO	-	35, 36
His-NTD-NEMO F91A	site-directed mutagenesis	His-NTD-NEMO	-	37, 38
His-NTD-NEMO Q89A+F91K	site-directed mutagenesis	His-NTD-NEMO	-	39, 40
His-NTD-NEMO Q192Y	site-directed mutagenesis	His-NTD-NEMO	-	47, 48
His-NTD-NEMO W164E	site-directed mutagenesis	His-NTD-NEMO	-	49, 50
His-NTD-NEMO L183K+Q192A	site-directed mutagenesis introducing the two mutations step-wise	His-NTD-NEMO	-	51, 52 and 53, 54
His-NTD-NEMO Q152L	site-directed mutagenesis	His-NTD-NEMO	-	41, 42
His-NTD-NEMO I154Q	site-directed mutagenesis	His-NTD-NEMO	-	43, 44
His-NTD-NEMO Q152L+I154Q	site-directed mutagenesis	His-NTD-NEMO	-	45, 46
His-NTD-NEMO I7K	site-directed mutagenesis	His-NTD-NEMO	-	55, 56
His-NTD-NEMO F9E	site-directed mutagenesis	His-NTD-NEMO	-	57, 58
His-NTD-NEMO F9W	site-directed mutagenesis	His-NTD-NEMO	-	59, 60
His-NTD-NEMO E11K	site-directed mutagenesis	His-NTD-NEMO	-	61, 62
His-NTD-NEMO Q14D+Q16M+N17S	site-directed mutagenesis	His-NTD-NEMO	-	63, 64
His-NTD-NEMO Q89A+F91K+Q192Y	site-directed mutagenesis	His-NTD-NEMO Q89A+F91K	-	47, 48
His-NTD-NEMO Q89A+F91K+F9W	site-directed mutagenesis	His-NTD-NEMO Q89A+F91K	-	59, 60
His-NTD-NEMO Q192Y+F9W	site-directed mutagenesis	His-NTD-NEMO Q192Y	-	59, 60
His-NTD-NEMO Q89A+F91K+Q192Y+F9E	site-directed mutagenesis	His-NTD-NEMO Q89A+F91K+Q192Y	-	57, 58
His-NTD-NEMO Q89A+F91K+Q192Y+F9W	site-directed mutagenesis	His-NTD-NEMO Q89A+F91K+Q192Y	-	59, 60
His-NTD-NEMO Q89A+F91K+Q192Y+E11K	site-directed mutagenesis	His-NTD-NEMO Q89A+F91K+Q192Y	-	61, 62
His-NTD-NEMO Q89A+F91K+Q192Y+ Q152L+I154Q	site-directed mutagenesis	His-NTD-NEMO Q89A+F91K+Q192Y	-	45, 46
His-NTD-NEMO Q89A+F91K+Q192Y+ Q152L+I154Q+F9W	site-directed mutagenesis	His-NTD-NEMO Q89A+F91K+Q192Y+F9W	-	45, 46

Construct	Cloning method	Source of DNA / source construct	Target vector	Primers
His-GST	site-directed mutagenesis, inserting the His-tag N-terminal to GST, making vector pOPT3HG	primers	pOPT3G	65, 66
His-GST-AmphCBM-long	annealed oligo insertion into target vector	primers	pOPT3HG cut with Nde1/BamH1	29, 30
His-GST-L1	site-directed mutagenesis, inserting the His-tag N-terminal to GST	primers	GST-HDAg-L1 (in pOPT3G)	65, 66
His-GST-L2	site-directed mutagenesis, inserting the His-tag N-terminal to GST	primers	GST-HDAg-L2 (in pOPT3G)	65, 66
His-GST-L3	site-directed mutagenesis, inserting the His-tag N-terminal to GST	primers	GST-HDAg-L3 (in pOPT3G)	65, 66
His-GST-Amph4T1	annealed oligo insertion into target vector	primers	pOPT3HG cut with Nde1/BamH1	67, 68
His-GST-AmphCBM	site-directed mutagenesis, inserting the His-tag N-terminal to GST	primers	GST-HDAg-AmphCBM (in pOPT3G)	65, 66
His-GST-AP2CBM	annealed oligo insertion into target vector	primers	pOPT3HG cut with Nde1/BamH1	27, 28
His-GST-Wbox	annealed oligo insertion into target vector	primers	pOPT3HG cut with Nde1/BamH1	31, 32
His-UL21(1-535)	amplification from source construct, cutting insert, ligation into target vector	pGBKT7-UL21	pOPTH cut with Nde1/BamH1	71, 72
UL21(1-535)-His	amplification from source construct, cutting insert, ligation into target vector	pGBKT7-UL21	pOPTnH cut with Nde1/BamH1	71, 73
His-gE(445-550)	amplification from source construct, cutting insert, ligation into target vector	pGBKT7-gE	pOPTH cut with Nde1/BamH1	74, 75
gE(445-550)-His	amplification from source construct, cutting insert, ligation into target vector	pGBKT7-gE	pOPTnH cut with Nde1/BamH1	74, 76

Construct	Cloning method	Source of DNA / source construct	Target vector	Primers
His-UL56(1-207)	amplification from source construct, cutting insert, ligation into target vector	pGBKT7-UL56(1-215)	pOPTH cut with Nde1/BamH1	77, 78
UL56(1-207)-His	amplification from source construct, cutting insert, ligation into target vector	pGBKT7-UL56(1-215)	pOPTnH cut with Nde1/BamH1	77, 79
His-UL56(1-215)	amplification from source construct, cutting insert, ligation into target vector	pGBKT7-UL56(1-215)	pOPTH cut with Nde1/BamH1	77, 80
UL56(1-215)-His	amplification from source construct, cutting insert, ligation into target vector	pGBKT7-UL56(1-215)	pOPTnH cut with Nde1/BamH1	77, 81
pGBKT7-UL56(1-207)	site-directed mutagenesis introducing stop codon	pGBKT7-UL56(1-215)	-	69, 70
GST-UL56(1-207)	amplification from source construct, cutting insert, ligation into target vector	pGBKT7-UL56(1-215)	pOPT3G cut with Nde1/BamH1	77, 78
myc-BECN1(128-450)	amplification from cDNA, cutting insert, ligation into target vector	HeLa cDNA	pF3A-myc-WG cut with HindIII/Pme1	82, 83
myc-DDX59(115-619)	amplification from cDNA, cutting insert, ligation into target vector	HeLa cDNA	pF3A-myc-WG cut with HindIII/Pme1	84, 85
myc-FNIP2(870-1114)	amplification from PCR product adding myc-tag, cutting insert, ligation into target vector	positive clone from yeast two-hybrid screen	pF3A-WG cut with Sgf1/Pme1	86, 87
myc-PAWR(217-340)	amplification from PCR product, cutting insert, ligation into target vector	positive clone from yeast two-hybrid screen	pF3A-myc-WG cut with HindIII/Pme1	88, 89
myc-PSD3(12-243)	amplification from cDNA, cutting insert, ligation into target vector	HeLa cDNA	pF3A-myc-s cut with Sgf1/Pme1	95, 96
myc-PSD3(1-243)	site-directed mutagenesis, inserting amino-terminal 11 residues	primer	myc-PSD3(12-243)	97, 98
myc-SENP6(809-1112)	amplification from cDNA, cutting insert, ligation into target vector	HeLa cDNA	pF3A-myc-WG cut with HindIII/Pme1	90, 91

Construct	Cloning method	Source of DNA / source construct	Target vector	Primers
myc-SENP6(972-1112)	amplification from cDNA, cutting insert, ligation into target vector	HeLa cDNA	pF3A-myc-WG cut with HindIII/Pme1	92, 91
myc-WDR61(1-305)	amplification from cDNA, cutting insert, ligation into target vector	HeLa cDNA	pF3A-myc-WG cut with HindIII/Pme1	93, 94
UL56(1-207)-eGFP	amplification from source construct, cutting insert, ligation into target vector	pGBKT7-UL56(1-215)	pEGFP-N1 cut with BamH1/HindIII	99, 100
myc-hTRAPPC9(1-1148) pF5K	amplification from source construct, cutting insert, ligation into target vector	myc-hTRAPPC9(1-1148) pF3A	pF5K-myc cut with Sgf1/Pme1	104, 105
myc-hTRAPPC10(1-1259) pF5K	amplification from source construct, cutting insert, ligation into target vector	myc-hTRAPPC10(1-1259) pF3A	pF5K-myc cut with Sgf1/Pme1	106, 107
myc-GOPC(1-454) (myc-GOPC-iso2(1-454))	amplification from cDNA, cutting insert, ligation into target vector	HeLa cDNA	pF3A-myc-s cut with Sgf1/Pme1	101, 102
myc-PAWR(1-340)	amplification from cDNA, cutting insert, ligation into target vector	HeLa cDNA	pF3A-myc-s cut with Sgf1/Pme1	103, 89
myc-hTRAPPC9(1-1148) pF3A	amplification from cDNA, cutting insert, ligation into target vector	HeLa cDNA	pF3A-myc-s cut with Sgf1/Pme1	104, 105
myc-hTRAPPC10(1-1259) pF3A	amplification from cDNA, cutting insert, ligation into target vector	HeLa cDNA	pF3A-myc-s cut with Sgf1/Pme1	106, 107
myc-GOPC-iso2(1-275)	amplification from source construct, cutting insert, ligation into target vector	myc-GOPC-iso2(1-454)	pF3A-myc-s cut with Sgf1/Pme1	101, 108
myc-GOPC-iso2(276-362)	amplification from source construct, cutting insert, ligation into target vector	myc-GOPC-iso2(1-454)	pF3A-myc-s cut with Sgf1/Pme1	109, 110
myc-GOPC-iso2(270-363)	amplification from source construct, cutting insert, ligation into target vector	myc-GOPC-iso2(1-454)	pF3A-myc-s cut with Sgf1/Pme1	111, 112
myc-GOPC-iso2(272-367)	amplification from source construct, cutting insert, ligation into target vector	myc-GOPC-iso2(1-454)	pF3A-myc-s cut with Sgf1/Pme1	113, 114

Construct	Cloning method	Source of DNA / source construct	Target vector	Primers
myc-GOPC-iso2(276-454)	amplification from source construct, cutting insert, ligation into target vector	myc-GOPC-iso2(1-454)	pF3A-myc-s cut with Sgf1/Pme1	109, 102
eGFP-UL56(1-207)	site-directed mutagenesis introducing stop codon	eGFP-UL56(1-234)	-	115, 116
eGFP-UL56(1-157)	site-directed mutagenesis introducing stop codon	eGFP-UL56(1-234)	-	117, 118
eGFP-UL56(1-104)	site-directed mutagenesis introducing stop codon	eGFP-UL56(1-234)	-	119, 120
eGFP-UL56(1-74)	site-directed mutagenesis introducing stop codon	eGFP-UL56(1-234)	-	121, 122
eGFP-UL56(105-234)	amplification from source construct, cutting insert, ligation into target vector	eGFP-UL56(1-234)	pEGFP-C1 cut with Hind3/EcoR1	123, 124
eGFP-UL56(105-207)	site-directed mutagenesis introducing stop codon	eGFP-UL56(105-234)	-	115, 116
eGFP-UL56(105-157)	site-directed mutagenesis introducing stop codon	eGFP-UL56(105-234)	-	117, 118

Table 6 Primers used in this thesis.

#, Name		Sequence 5' → 3'
1	pGADT7-f	GATGAAGATACCCCACCAAACC
2	pGADT7-r	TGGCCAAGATTGAAACTTAGAGG
3	pGADT7-long-f	CGATGATGAAGATACCCCACCAAACC
4	pGADT7-long-r	AGAGGTTACATGGCCAAGATTGAAACTTAGAGG
5	HDAg-L1-f	TATGCCGTGGGACATCCTGTTCCCGGCGGACCCGCCGTTCTCTCCGCAG TCTTGCCGTCCGCAGG
6	HDAg-L1-r	GATCCCTGCGGACGGCAAGACTGCGGAGAGAACGGCGGGTCCGCCGGGA ACAGGATGTCCCACGGCA
7	HDAg-L2-f	TATGCCGTGGGTTTCTCCGCAGCCGCCGCCGCCGCGTCTGCCGCTGCTG GAATGCACCCCGCAGG
8	HDAg-L2-r	GATCCCTGCGGGGTGCATTCCAGCAGCGGCAGACGCGGCGGCGGCGGCT GCGGAGAAACCCACGGCA
9	HDAg-L3-f	TATGCCGTGGTACGGTTTCACCCCGCCGCCGCCGGGTTACTACTGGGTT CCGGGTTGCACCCAGCAGG
10	HDAg-L3-r	GATCCCTGCTGGGTGCAACCCGGAACCCAGTAGTAACCCGGCGGCGGCG GGGTGAAACCGTACCACGGCA
11	GST L1 stop fwd	CTTGCCGTCCGCAGTGATCCAGATCTAAG

#, Name		Sequence 5' → 3'
12	GST L1 stop rev	CTTAGATCTGGATCACTGCGGACGGCAAG
13	GST L2 stop fwd	GAATGCACCCCGCAGTGATCCAGATCTAAG
14	GST L2 stop rev	CTTAGATCTGGATCACTGCGGGGTGCATTC
15	GST L3 stop fwd	GTTGCACCCAGCAGTGATCCAGATCTAAG
16	GST L3 stop rev	CTTAGATCTGGATCACTGCTGGGTGCAAC
17	HDAg_L2-2-f	TATGCCGTGGGTTCACCCGTCTCCGCCGCAGCAGCGTCTGCCGCTGCTG GAATGCACCCCGCAGTGAG
18	HDAg_L2-2-r	GATCCTCACTGCGGGGTGCATTCCAGCAGCGGCAGACGCTGCTGCGGCG GAGACGGGTGAACCCACGGCA
19	HDAg_L2-3-f	TATGCCGTGGGGTAACACCCCGCCGCGTCCGCCGCGTCTGCCGCTGCTG GAATGCACCCCGCAGTGAG
20	HDAg_L2-3-r	GATCCTCACTGCGGGGTGCATTCCAGCAGCGGCAGACGCGGCGGACGCG GCGGGGTGTTACCCCACGGCA
21	HDAg_L2-4-f	TATGCCGTGGGTTTCTCCGGGTTCTCCGTCTCCGCGTCTGCCGCTGCTG GAATGCACCCCGCAGTGAG
22	HDAg_L2-4-r	GATCCTCACTGCGGGGTGCATTCCAGCAGCGGCAGACGCGGAGACGGAG AACCCGGAGAAACCCACGGCA
23	HDAg_L2-1-C251S-f	CCGCTGCTGGAAAGCACCCCGCAGTG
24	HDAg_L2-1-C251S-r	CACTGCGGGGTGCTTTCCAGCAGCGG
25	HDAg_L2-1-C251Stop-f	CGCTGCTGGAATGAACCCCGCAGTGATC
26	HDAg_L2-1-C251Stop-r	GATCACTGCGGGGTTCATTCCAGCAGCG
27	AP2_cbm-long-f	TATGGGTGACCTGCTGAACCTGGACCTGGGTCCGCCGGTTTGAG
28	AP2_cbm-long-r	GATCCTCAAACCGGCGGACCCAGGTCCAGGTTCAGCAGGTCACCCA
29	Amph1_cbm-long-f	TATGGAAACCCTGCTGGACCTGGACTTCGACCCGTTCAAATGAG
30	Amph1_cbm-long-r	GATCCTCATTTGAACGGGTCGAAGTCCAGGTCCAGCAGGGTTTCCA
31	Amph1-Wbox-f	TATGACCCTGCCGTGGGACCTGTGGACCACCTCTTGAG
32	Amph1-Wbox-r	GATCCTCAAGAGGTGGTCCACAGGTCCCACGGCAGGGTCA
33	qc-Amphi-F359Stop-f	GACTTCGACCCGTGAAAATGAGGATCC
34	qc-Amphi-F359Stop-r	GGATCCTCATTTTCACGGGTCGAAGTC
35	NTD-Q89M-f	CACTAAAAGCTGGGAAAACTCTCATGATATTTAACATTGAAATGAAAA G
36	NTD-Q89M-r	CTTTTCATTTCAATGTTAAATATCATGAGAGTTTTCCCAGCTTTTAGT G
37	NTD-F91A-f	GCTGGGAAAACTCTCCAGATAGCGAACATTGAAATGAAAAGTAA
38	NTD-F91A-r	TTACTTTTCATTTCAATGTTCGCTATCTGGAGAGTTTTCCCAGC
39	NTD-Q89A+F91K-f	ATTGCACTAAAAGCTGGGAAAACTCTCGCGATTAAAAACATTGAAATGA AAAGTAAAATGAAGG
40	NTD-Q89A+F91K-r	CCTTCATTTTACTTTTCATTTCAATGTTTTTAATCGCGAGAGTTTTCC CAGCTTTTAGTGCAAT
41	NTD-Q152L-f	CCTCGCAGGGTGCCTGATTATCAATTATCG
42	NTD-Q152L-r	CGATAATTGATAATCAGGCACCCTGCGAGG
43	NTD-I154Q-f	CGCAGGGTGCCAGATTCAGAATTATCGAACTGATGC
44	NTD-I154Q-r	GCATCAGTTCGATAATTCTGAATCTGGCACCCTGCG
45	NTD-Q152L+I154Q-f	CCACTCTAGCCTCGCAGGGTGCCTCATCCAGAATTATCGAACTGATGCA AAGC
46	NTD-Q152L+I154Q-r	GCTTTGCATCAGTTCGATAATTCTGGATGAGGCACCCTGCGAGGCTAGA GTGG
47	NTD-Q192Y-f	GTAGATAGGAAAGTGTCTTACCCCATTGAGGGACACGC
48	NTD-Q192Y-r	GCGTGTCCCTCAATGGGGTAAGACACTTTCCTATCTAC
49	NTD-W164E-f	CGAACTGATGCAAAGCAAAAGGAATTACTTCTGACTGGCATATC
50	NTD-W164E-r	GATATGCCAGTCAGAAGTAATTCCTTTTGCTTTGCATCAGTTCG
51	NTD-L183K-f	GTGTGGTGGGAGCTATGCAGAAATATTCTGTAGATAGGAAAG

#, Name		Sequence 5' → 3'
52	NTD-L183K-r	CTTTCCTATCTACAGAATATTTCTGCATAGCTCCCACCACAC
53	NTD-Q192A-f	GATAGGAAAGTGTCTGCGCCCATTGAGGGAC
54	NTD-Q192A-r	GTCCCTCAATGGGCGCAGACACTTTCCTATC
55	NTD-I7K-f	GCCCAGATTCTGCCAAAACGTTTTCAGGAGCATC
56	NTD-I7K-r	GATGCTCCTGAAAACGTTTTGGCAGAATCTGGGC
57	NTD-F9E-f	CAGATTCTGCCAATTCGTGAACAGGAGCATCTTCAGCTC
58	NTD-F9E-r	GAGCTGAAGATGCTCCTGTTCACGAATTGGCAGAATCTG
59	NTD-F9W-f	GATTCTGCCAATTCGTTGGCAGGAGCATCTTCAGC
60	NTD-F9W-r	GCTGAAGATGCTCCTGCCAACGAATTGGCAGAATC
61	NTD-E11K-f	CTGCCAATTCGTTTTCAGAAACATCTTCAGCTCCAGAACC
62	NTD-E11K-r	GGTTCTGGAGCTGAAGATGTTTCTGAAAACGAATTGGCAG
63	icq-NTD-R5-f	TTTCAGGAGCATCTTGATCTCATGAGCCTGGGTATAAAC
64	icq-NTD-R5-r	ATCAAGATGCTCCTGAAAACGAATTGGCAGAATCTGG
65	pOPT3HG fwd	GAAGGAGATATACATAATATGCATCACCATCACCATCACGGGTCCTCCC CTATACTAGGTTATTG
66	pOPT3HG rev	CAATAACCTAGTATAGGGGAGGACCCGTGATGGTGATGGTGATGCATA TTATGTATATCTCCTTC
67	aoi-Amphi-old-f	TATGGAAACCCTGCTGGACCTGGACTTCCTGGAACGTCCGCACCGTGAC TGAG
68	aoi-Amphi-old-r	GATCCTCAGTCACGGTGCGGACGTTCCAGGAAGTCCAGGTCCAGCAGGG TTTCCA
69	UL56-qc-T624A-f	GCGACAGGGCTGATGTGGCGTGGTG
70	UL56-qc-T624A-r	CACCACGCCACATCAGCCCTGTCGC
71	UL21-pOPT-1-f	GGAAGTCATATGGAGCTTAGCTACGCCACC
72	UL21-pOPT-535-r	GGAAGTGGATCCTCACACAGACTGTCCGTGTTG
73	UL21-pOPTnH-535-r	GGAAGTGGATCCCACAGACTGTCCGTGTTGGGAG
74	gE-pOPT-445-fwd	GGAAGTCATATGCGCAGGCGTGCCTGGCGGGC
75	gE-pOPT-550-rev	GGAAGTGGATCCTCACCAGAAGACGGACGAATCG
76	gE-pOPTn-550-rev	GGAAGTGGATCCCCAGAAGACGGACGAATCG
77	UL56-pOPT-1-fwd	GGAAGTCATATGGCTTCGGAGGCG
78	UL56-pOPT-207- rev	GGAAGTGGATCCTCAGCCCTGTCGCCGGTATG
79	UL56-pOPTn-207-rev	GGAAGTGGATCCGCCCTGTCGCCGGTATG
80	UL56-pOPT-215-rev	GGAAGTGGATCCTCAATGGCGCCACACCACG
81	UL56-pOPTn-215-rev	GGAAGTGGATCCATGGCGCCACACCACG
82	hBECN1(128-450)-WG-f	GGAAGAAGCTTCCGGCCAGACAGATGTGGATCAC
83	hBECN1(128-450)-WG-r	GGAAGTTTAAACTCATTTGTTATAAAATTGTGAGGACACCCAAG
84	hDDX59-iso1(115-619)-WG-f	GGAAGAAGCTTCCGGAGAGTATATCTGTGATAAGACAGATGAAG
85	hDDX59-iso1(115-619)-WG-r	GGAAGTTTAAACTCATTTCTGAGAATTACTTTTATCATGTTTTCTG
86	hFNIP2(870-1114)-WG-f	GGAAGCGATCGCCATGGAGCAAAAGCTCATTTCTGAAGAGGACTTGAAT ATCCCCTGTGGGGATGAC
87	hFNIP2(870-1114)-WG-r	GGAAGTTTAAACTCATAAGAGTATTTGAGCCACATAAGGAG
88	hPAWR(217-340)-WG-f	GGAAGAAGCTTCCGAGCCACCTAGAACAGTTTC
89	hPAWR(217-340)-WG-r	GGAAGTTTAAACTCACCTGGTCAGCTGACCCAC
90	hSENP6-iso1(809-1112)-WG-f	GGAAGAAGCTTCCAGTGAAATGGAGAGTTGTTCACAAAACTC
91	hSENP6-iso1(809-1112)-WG-r	GGAAGTTTAAACTCAATCTGAGATACTATTGACATATTGTTCTGTTC
92	hSENP6-iso1(972-1112)-WG-f	GGAAGAAGCTTCCTGTATCCTACTTATGGACTCACTCCG
93	hWDR61(1-305)-WG-f	GGAAGAAGCTTCCACCAACCAGTACGGTATTCTCTTC

#, Name		Sequence 5' → 3'
94	hWDR61(1-305)-WG-r	GGAAGTTTAAACTCAAATTGGACAATCATAGATGTG
95	hPSD3(12-243)-WG-f	GGAAGCGATCGCAATGCTGGGGTGAAAACAACAC
96	hPSD3-iso3(1-243)-WG-r	GGAAGTTTAAACTCAAATGTCCAAAAATGGGTTAGTAGTACTTC
97	iqc-hPSD3(2-11)-f	AATGGCGATCGCGGCTCTTCTTGGTGTCTGTATGGTTGCTGCAATGCT GGGGTGAAAAC
98	iqc-hPSD3(2-11)-r	GCAGCAACCATACAGACACCAAGAAGAGCCGCGATCGCCATTCAAGTCC TCTTCAGAAATGAGC
99	UL56 1 GFP-N fwd	GGAAAAGCTTGCCGCCACCATGGCTTCGGAGGCG
100	UL56 207 GFP-N rev	GGAAGGATCCGAGCCCTGTCGCCGGTATG
101	hGOPC(1-462)-WG-f	GGAAGGCGATCGCATGTCGGCGGGCGGTC
102	hGOPC(1-462)-WG-r	GGAAGTTTAAACTCAATAAGATTTTTTATGATACAGAGTGTGCAG
103	hPAWR(1-340)-WG-f	GGAAGGCGATCGCATGGCGACCGGTGGCTAC
104	hTRAPPC9(1-1148)-WG-f	GGAAGGCGATCGCATGAGCGTCCCTGACTACATG
105	hTRAPPC9(1-1148)-WG-r	GGAAGTTTAAACTCAGGCCTGCGCCTCCAG
106	hTRAPPC10(1-1259)-WG-f	GGAAGGCGATCGCATGGACGCCTCTGAGGAGC
107	hTRAPPC10(1-1259)-WG-r	GGAAGTTTAAACTCATGTTACACTGACTTCCAGGACG
108	hGOPC_2(1-275)-WG-r	GGAAGTTTAAACTCAAACACCTTGGCTTTTCTTTAGGGAATC
109	hGOPC_2(276-362)-WG-f	GGAAGGCGATCGCGGTCCAATTAGAAAAGTTCTCCTCCTTAAG
110	hGOPC_2(276-362)-WG-r	GGAAGTTTAAACTCACACATAAACTACTTCAAATTCAATCTCTCCTC
111	hGOPC_2(270)-WG-f	GGAAGGCGATCGCAAGAAAAGCCAAGGTGTTGGTC
112	hGOPC_2(363)-WG-r	GGAAGTTTAAACTCAAGCCACATAAACTACTTCAAATTCAATCTC
113	hGOPC_2(272)-WG-f	GGAAGGCGATCGCAGCCAAGGTGTTGGTCCAATTAG
114	hGOPC_2(367)-WG-r	GGAAGTTTAAACTCAATCCACTTCAGGAGCCACATAAAC
115	qc-UL56(C208Stop)-f	GCGACAGGGCTGATGTGGCGTGGTG
116	qc-UL56(C208Stop)-r	CACCACGCCACATCAGCCCTGTCGC
117	qc-UL56(Q158Stop)-f	GTTCGTTCGACCGAATAACCGTCCGGGGCTTTG
118	qc-UL56(Q158Stop)-r	CAAAGCCCCGGACGGTTATTCGGTCGAACGAAC
119	qc-UL56(G105Stop)-f	CCGGAGTTTTGGCCCTTGAGGGTTATTGGCAACC
120	qc-UL56(G105Stop)-r	GGTTGCCAATAACCCTCAAGGGCCAAAACTCCGG
121	qc-UL56(R75Stop)-f	GGCGAGGCGTCGGAGTGAGGCCGCTCACGCCAAG
122	qc-UL56(R75Stop)-r	CTTGGCGTGAGCGGCCTCACTCCGACGCCTCGCC
123	eGFPC1-UL56(105)-f	GGAAGAAGCTTCCGGCGGGTTATTGGCAACC
124	eGFPC1-UL56(234)-r	GGAAGAATTCTCACCGCCACAGGAATACCAGAATAATG

2.4 Protein expression and purification

2.4.1 Clathrin N-terminal domain (NTD) and full-length clathrin

Glutathione S-transferase- (GST-) tagged NTD was expressed in *E. coli* BL21(DE3) pLysS (Life Technologies). Large-scale cultures were grown in 2xTY at 37 °C, 200 rpm. Upon reaching an OD_{600} of 0.8-1.0, the temperature was dropped to 22 °C and protein expression was induced by addition of 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). After 12-18 h, cells were harvested by centrifugation (6000 g, 15 min, 4 °C). All pellets were frozen and then thawed and used immediately, or stored at -80 °C.

The cell pellet was resuspended in lysis buffer (20 mM Tris pH 7.5, 200 mM NaCl, 0.05% Tween-20, 0.5 mM $MgCl_2$, 1.4 mM β-mercaptoethanol) with added bovine pancreas DNase I (200-400 U/L cell culture, Sigma) and EDTA-free protease inhibitor cocktail (100-200 µL/L cell culture, Sigma). Cells were lysed using a cell disruptor (Constant Systems) at 24 kpsi. The lysate was cleared by centrifugation (40000 g, 30 min, 4 °C) and purified by affinity chromatography using glutathione sepharose 4B beads (GE Healthcare). The beads were incubated with the protein for 60 min at 4 °C and then washed using 40 column volumes of wash buffer (20 mM Tris pH 7.5, 200 mM NaCl, 1 mM DTT). The GST-tag of GST-NTD was then liberated by overnight incubation of the beads with thrombin (125 U, Serva) in 30 mL thrombin cleavage buffer (20 mM Tris pH 7.5, 200 mM NaCl, 1 mM $CaCl_2$). The flow-through containing cleaved NTD was collected, whereas the cleaved GST and any uncleaved GST-NTD remained on the beads. Cleaved NTD was further purified by size exclusion chromatography on an ÄKTA Pure fast protein liquid chromatography (FPLC) using HiLoad Superdex 200 PG gel filtration columns (GE Healthcare) equilibrated in gel filtration buffer (10 mM Tris pH 7.5, 50 mM NaCl, 4 mM DTT). The purity of the protein was assessed by SDS-PAGE. Small aliquots (20-100 µL) of purified NTD were snap-frozen in liquid nitrogen and stored at -80 °C (273).

Full-length clathrin was purified from pig brain lysate by Bernard T. Kelly as described previously (46).

2.4.2 Wild-type and mutant His-NTD-NEMO

All wild-type and mutant His-NTD-NEMO variants were expressed in *E. coli* B834(DE3). Large-scale cultures were grown, harvested and frozen as above (section 2.4.1). Cell pellets were resuspended in lysis buffer (20 mM Tris pH 7.5, 500 mM NaCl, 0.05% Tween-20, 0.5 mM $MgCl_2$, 1.4 mM β-mercaptoethanol) supplemented with approximately 10 mg hen egg white lysozyme (Sigma), bovine pancreas DNase I (400 U/L cell culture) and EDTA-free protease inhibitor cocktail (200 µL/L cell culture). The cells were lysed, the lysate was cleared as described above and then 0.2 µm filtered using a syringe filter and stored on ice. The solution was then injected onto a 1 mL HisTrap Excel Nickel Sepharose affinity column (GE Healthcare). The column was washed with wash buffer (20 mM Tris pH 7.5, 500 mM NaCl, 12.5 mM imidazole pH 7.5) and the protein eluted using elution buffer (20 mM Tris pH 7.5, 500 mM NaCl, 250 mM imidazole pH 7.5). The eluate from the affinity purification was injected onto a HiLoad Superdex 200 PG

gel filtration column (GE Healthcare) equilibrated in gel filtration buffer (20 mM Tris pH 7.5, 200 mM NaCl, 1 mM DTT) for purification by size exclusion chromatography. The purified protein eluted in distinct peaks, which were collected and concentrated separately, yielding oligomeric (His-NTD-NEMO) and monomeric (His-NTD) proteins. Both oligomeric and monomeric NTD(-NEMO) variants were snap-frozen in liquid nitrogen as 10-100 µL aliquots and stored at -80 °C.

2.4.3 GST, His-GST and (His-)GST fusion proteins

GST, His-GST, GST-YXXΦ, GST-AP2-HA and His-GST-AP2-HA, and all GST and His-GST fusion proteins of cellular or viral clathrin-binding motifs were expressed in *E. coli* BL21(DE3) pLysS. Large-scale cultures were grown, harvested and frozen as above (section 2.4.1). The cell pellet was resuspended in lysis buffer (20 mM HEPES pH 7.5, 300 mM NaCl, 0.05% Tween-20, 0.5 mM $MgCl_2$, 1.4 mM β-mercaptoethanol) supplemented with bovine pancreas DNase I (200-400 U/L cell culture) and EDTA-free protease inhibitor cocktail (100-200 µL/L cell culture). Cells were lysed and the lysate clarified as above. Proteins were first purified by affinity chromatography by incubation with glutathione sepharose 4B resin for 60 min at 4 °C. The column was washed with 40 column volumes of wash buffer (20 mM HEPES pH 7.5, 300 mM NaCl, 1 mM DTT) and the protein eluted using wash buffer supplemented with 25 mM reduced glutathione adjusted to pH 7.5 with sodium hydroxide. Proteins were further purified by size exclusion chromatography using HiLoad Superdex 75 PG or HiLoad Superdex 200 PG gel filtration columns (GE Healthcare) equilibrated in gel filtration buffer (20 mM HEPES pH 7.5, 200 mM NaCl, 1 mM DTT). Purified proteins were analyzed by SDS-PAGE, concentrated, mixed 1:1 with 100% (v/v) glycerol and stored at -20 °C. An exception was protein His-GST-AP2-HA, which proved not to be stable as a 50% glycerol stock and was therefore snap-frozen in 100 µL aliquots and stored at -80 °C.

2.4.4 His-c-µ2

His-c-µ2 was expressed in *E. coli* BL21(DE3) pLysS. Large-scale cultures were grown, harvested and frozen as above (section 2.4.1). A cell pellet was resuspended in lysis buffer (10 mM HEPES pH 7.5, 500 mM NaCl, 20 mM imidazole pH 7.5, 0.05% Tween-20, 0.5 mM $MgCl_2$, 1.4 mM β-mercaptoethanol) supplemented with approximately 5 mg hen egg white lysozyme, bovine pancreas DNase I (400 U/L cell culture) and EDTA-free protease inhibitor cocktail (200 µL/L

cell culture). Cells were lysed and the lysate cleared as above. The protein was first purified by affinity chromatography using Ni-nitrilotriacetic acid (Ni-NTA) resin (Qiagen). The cleared lysate was incubated with Ni-NTA beads for 60 min at 4 °C. The beads were subsequently washed using 60 column volumes of wash buffer (10 mM HEPES pH 7.5, 500 mM NaCl, 20 mM imidazole pH 7.5) and proteins were eluted using a high imidazole elution buffer (10 mM HEPES pH 7.5, 500 mM NaCl, 250 mM imidazole pH 7.5). The protein was then further purified by size exclusion chromatography into gel filtration buffer (10 mM HEPES pH 7.5, 350 mM NaCl) using a HiLoad Superdex 200 PG gel filtration column. The purified protein was analyzed on an SDS-PAGE gel, concentrated, snap-frozen and stored in small aliquots at -80 °C.

2.4.5 High-throughput expression and purification screen of HSV-1 proteins

When possible, steps of this expression and purification screen were performed in 24- or 96-well plates to streamline sample handling. N- and C-terminally His-tagged constructs of the HSV-1 proteins gE, UL56 and UL21, as well as constructs of His-tagged control proteins were transformed into BL21(DE3) pLysS, Rosetta2(DE3) pLysS and B834(DE3) competent *E. coli* cells. A single, isolated colony of each construct in each bacterial cell strain was inoculated into 3 mL 2xTY supplemented with appropriate antibiotics and these starter cultures were grown for 3.5-18 h at 37 °C, 200 rpm. Fresh 3.5 mL 2xTY medium containing antibiotics was then inoculated with 3.5 µL of each starter culture and these cultures were grown until they reached an A_{600} of 0.6-1.0. The incubation temperature was dropped to 22 °C and protein expressing induced by adding 17.5 µL of 0.4 M IPTG to each culture. Cultures were grown for 20 h at 22 °C, 200 rpm. Cells were harvested by centrifugation (4000 g, 30 min, 4 °C), the supernatant discarded and the cell pellets frozen on dry ice.

The cell pellets were thawed and lysed in 250 µL lysis buffer (50 mM sodium phosphate pH 8.0, 300 mM NaCl, 10 mM imidazole pH 7.5, 1% (v/v) Tween-20 supplemented with hen egg white lysozyme, bovine pancreas DNase I and EDTA-free protease inhibitors) by shaking on a fast-rotating table-top shaking platform. The cell lysates were subsequently clarified by centrifugation (4000 g, 30 min, 4 °C). 20 µL of magnetic Ni-NTA beads per sample were dispensed into a 96-well flat-bottomed plate and washed twice with 200 µL wash buffer (50 mM sodium phosphate pH 8.0, 300 mM NaCl, 20 mM imidazole pH 7.5, 0.05% Tween-20) by first shaking the plate on the shaking platform, then placing it on a magnet, thus collecting the beads, and then aspirating

the supernatant. The cleared lysates (200 µL of each sample) were added to the beads and the mixture incubated for 45 min at 20 °C whilst rotating on the shaking platform. The plate was placed on the magnet again and the supernatant containing unbound proteins was discarded. The beads were washed twice using 200 µL of wash buffer as above. Subsequently, bound proteins were eluted by adding 50 µL of elution buffer (50 mM sodium phosphate pH 8.0, 300 mM NaCl, 250 mM imidazole pH 7.5, 0.05% Tween-20) to each well, incubating the beads for 1 min on the rocker, then collecting the beads with the magnet and aspirating 48 µL of the supernatant. 12 µL of 5x SDS-PAGE buffer was added to each elution sample and the samples were boiled at 95 °C for 5 min. Samples were separated on 12 % SDS-PAGE gels and expression and purification of the different constructs in each *E. coli* strain were assessed using Coomassie staining.

2.4.6 UL21(1-535)-His and His-UL56(1-207)

For large-scale purifications, UL21(1-535)-His and His-UL56(1-207) were expressed in *E. coli* BL21(DE3) pLysS competent cells. Large-scale cultures were grown, harvested and frozen as above (section 2.4.1). The protein pellets were lysed in lysis buffer (20 mM Tris pH 7.5, 500 mM NaCl, 20 mM imidazole pH 7.5, 0.05% Tween-20, 0.5 mM $MgCl_2$, 1.4 mM β-mercaptoethanol, supplemented with 400 U/L cell culture of bovine pancreas DNase I and 200 µL/L cell culture of EDTA-free protease inhibitor cocktail) as described above. The proteins were purified by affinity chromatography followed by size exclusion chromatography as described in section 2.4.4, but using the following buffers: wash buffer (20 mM Tris pH 7.5, 500 mM NaCl, 20 mM imidazole pH 7.5), elution buffer (20 mM Tris pH 7.5, 500 mM NaCl, 250 mM imidazole pH 7.5), and gel filtration buffer (20 mM Tris pH 7.5, 200 mM NaCl, 1 mM DTT). For UL21(1-535)-His, a HiLoad Superdex 200 PG gel filtration column, and for His-UL56(1-207), a HiLoad Superdex 75 PG gel filtration column were used during size exclusion chromatography. The purity of the proteins after size exclusion chromatography was assessed by SDS-PAGE. The purified proteins were concentrated, snap-frozen in liquid nitrogen and stored at -80 °C.

2.4.7 GST-UL56(1-207)

GST-UL56(1-207) was expressed in *E. coli* BL21(DE3) pLysS competent cells. Large-scale cultures were grown, harvested and frozen as above (section 2.4.1). Cell pellets were lysed and clarified, and the protein purified by affinity chromatography followed by size exclusion chromatography as described in section 2.4.3. A HiLoad Superdex 200 PG gel filtration column

was used for size exclusion chromatography. The purified protein was assessed by SDS-PAGE. Due to DNA contamination, only an approximate protein concentration could be measured. The protein was snap-frozen in liquid nitrogen and stored in 30 µL aliquots at -80 °C.

2.5 Crystallization, structure determination and analysis

2.5.1 Sparse-matrix screening

Sparse-matrix crystallization experiments ("broad screens") of clathrin NTD in complex with peptides, c-µ2 in complex with HDAg-L3$_{pep}$, and full-length UL21 were attempted using an Innovadyne Screenmaker liquid handling platform. Crystallization experiments were performed in 96-well sitting drop vapor diffusion plates using drops containing 200 nL of protein or protein:peptide solutions and 200 nL reservoir, drops being equilibrated against 80 µL of reservoir at 20 °C.

2.5.2 Crystallization of NTD in complex with peptides

All crystals were grown at 20 °C using sitting drop vapor diffusion. Clathrin NTD(1-363) was co-crystallized in complex with an HDAg-L1 peptide (SDILFPADS), an HDAg-L2 peptide (SPRLPLLES), an Amph4T1 peptide (ETLLDLDFLE), an AmphCBM peptide (ETLLDLDFDP), AmphCBM-long peptide (ETLLDLDFDPFK) or an AP2CBM peptide (CGDLLNLDLG). The HDAg-L1, HDAg-L2 and AP2CBM peptides were purchased from Designer Bioscience. All other peptides were purchased from Genscript. All peptides were prepared as 10 mM stock solutions in 10 mM Tris pH 7.5, 50 mM NaCl, 4 mM DTT and stored at -20 °C.

The crystal used for the determination of the NTD:HDAg-L1$_{pep}$ structure was grown in a 600 nL drop (400 nL of 14 mg/mL NTD and 3.4 mM HDAg-L1$_{pep}$, plus 200 nL of the reservoir solution) equilibrated against a 80 µL reservoir of 1.21 M sodium malonate pH 7.0 (Hampton Research). The crystal used for the solution of the NTD:HDAg-L2$_{pep}$ structure was grown in a 600 nL sitting drop (200 nL of 20 mg/mL NTD and 3.55 mM HDAg-L2$_{pep}$, plus 400 nL of the reservoir) equilibrated against a 80 µL reservoir of 1.75 M sodium malonate pH 7.0. The crystal used for the determination of the NTD:Amph4T1$_{pep}$ structure was grown in a 2 µL sitting drop (1 µL of 14 mg/mL NTD and 3.4 mM Amph4T1$_{pep}$, plus 1 µL of the reservoir) equilibrated against a 200 µL reservoir of 1.1 M sodium malonate pH 8.0. The crystal used for the solution of the

NTD:AmphCBM$_{pep}$ structure was grown in a 3 µL sitting drop (1 µL of 14 mg/mL NTD and 3.4 mM AmphCBM$_{pep}$, plus 2 µL of the reservoir) equilibrated against a 200 µL reservoir of 0.85 M sodium malonate pH 7.5. The crystal used for the determination of the NTD:AmphCBM-long$_{pep}$ structure was grown in a 3 µL drop (1 µL of 14 mg/mL NTD and 3.4 mM AmphCBM-long$_{pep}$, plus 2 µL of 1.04 M sodium malonate pH 7.1, 0.2 M sodium perchlorate (Jena Bioscience)) equilibrated against a 200 µL reservoir of 1.15 M sodium malonate pH 7.1. The crystal used for the determination of the NTD:AP2CBM$_{pep}$ structure was grown in a 3 µL sitting drop (1 µL of 14 mg/mL NTD and 3.4 mM AP2CBM$_{pep}$, plus 2 µL of the reservoir) equilibrated against a 200 µL reservoir of 0.94 M sodium malonate pH 6.7.

All protein crystals were cryoprotected by transferring them quickly into a drop consisting of 55% mother reservoir, 25% glycerol and 20% of the respective 10 mM peptide stock solution. The peptide was added to the cryoprotective solution in order to prevent back-soaking of the peptides from the co-crystal into the cryoprotectant. The crystals were then immediately flash-cooled by plunging into liquid nitrogen.

2.5.3 Crystallization experiments with HDAg-L3 peptides

Crystallization experiments were performed with an HDAg-L3 peptide (DGYYWVPGS) and a longer HDAg-L3-long peptide (DQFPWYGFTPPPPGYYWVPGD). Both peptides were purchased from Genscript, prepared as 10 mM stock solutions in 10 mM Tris pH 7.5, 50 mM NaCl, 4 mM DTT and stored at -20 °C.

For broad-screen co-crystallization experiments, NTD (at 10 mg/mL, 12 mg/mL, 14 mg/mL or 20 mg/mL) and the peptides (in 7- or 10-fold molar excess) were both present during the crystallization. These crystals were cryoprotected in a drop of 55% mother reservoir, 25% glycerol (or 25% PEG-400) and 20% of the respective peptide before flash-cooling in liquid nitrogen. For soaking experiments, needle-shaped NTD crystals without peptide were grown using vapor diffusion against reservoirs containing PEG-3350 (19.0-35.5% w/v), 0.1 M Tris (pH 7.0, 7.5, 7.7, 8.0 and 8.3) and 0.15 M potassium acetate (adapted from (274)). Screening these crystals for an optimal cryoprotectant identified PEG-400 as the most suitable reagent. Subsequently, these NTD crystals were soaked for up to 24 h in drops containing their respective mother reservoir solution with addition of 25% PEG-400 and up to 7.5 mM of peptide and then flash-cooled in liquid nitrogen. Conditions from co-crystallization and peptide soaking experiments were

improved by screening for chemical compounds that promoted crystallization using the JBScreen Plus additive screen (Jena Bioscience).

2.5.4 Data collection, structure determination, refinement and analysis

Diffraction data for all crystals were recorded at the Diamond Light Source, Oxford, at beamlines I-02 (NTD:HDAg-L1$_{pep}$, NTD:HDAg-L2$_{pep}$, NTD:AmphCBM-long$_{pep}$, NTD:HDAg-L3$_{pep}$, NTD:HDAg-L3-long$_{pep}$), I-04 (NTD:HDAg-L3$_{pep}$, NTD:HDAg-L3-long$_{pep}$), and I-04-1 (NTD:Amph4T1$_{pep}$, NTD:AmphCBM$_{pep}$, NTD:AP2CBM$_{pep}$).

The data for the NTD:HDAg-L1$_{pep}$, NTD:HDAg-L3$_{pep}$, NTD:AmphCBM-long$_{pep}$ and the NTD:Amph4T1$_{pep}$ structures were processed using the xia2 automated processing pipeline using XDS, XSCALE, and Aimless for indexing and integration, scaling and merging, respectively (275-281). The data for the NTD:AmphCBM$_{pep}$ and the NTD:AP2CBM$_{pep}$ structure were processed using the xia2 pipeline using DIALS (282, 283) and Aimless. The data for the NTD:HDAg-L2$_{pep}$ structure were indexed and integrated using iMOSFLM (284) and then scaled and merged using Aimless. Accumulated radiation dose damaged the crystals during data collection and the best compromise between completeness and multiplicity of the data sets and lower data quality due to radiation damage was reached by truncating the data sets accordingly.

The structures of the NTD:HDAg-L1$_{pep}$, NTD:HDAg-L2$_{pep}$, NTD:Amph4T1$_{pep}$, NTD:AmphCBM$_{pep}$ and NTD:AmphCBM-long$_{pep}$ complexes were in the same crystal form as a high-resolution ligand-free model of NTD (PDB 1C9I)(62). They were solved by rigid-body refinement in REFMAC5 (285, 286). The structures of the NTD:AP2CBM$_{pep}$ and NTD:HDAg-L3$_{pep}$ complexes were solved by molecular replacement with Phaser (287) using a single protein main chain model of NTD (PDB 1C9I, chain A). Manual model building was performed using COOT (288) and the models were refined using REFMAC5. In all structures, the peptides were modeled after initial improvement of the peptide-free structure, and with cycles of refinement in between modeling of the different peptide binding sites. The geometric quality of the models was improved by consulting the validation tools of COOT, as well as the programs MolProbity (289) and whatcheck (290). Evolutionary conservation of amino acids was estimated with ConSurf (default parameters) using chain A of the refined NTD:Amph4T1$_{pep}$ structure (291).

2.6 Mammalian cell culture

2.6.1 General maintenance of cells

Human embryonic kidney cells (HEK293T) were grown in Dulbecco's modified eagle medium (DMEM)(Sigma) supplemented with 10% heat-inactivated fetal bovine serum, 2 mM glutamine, 100 U/ml penicillin and 100 µg/ml streptomycin (complete DMEM). Cells were maintained at 37 °C, 5% CO_2 in humidified atmosphere. Thawing, passaging and counting of cells were performed by Susanna Colaco.

2.6.2 Transfection of cells

Cells were seeded into 9 cm dishes at a density of 5×10^6 cells/dish and grown in complete DMEM for 18-24 h to reach 80% confluency. The medium was replaced with fresh DMEM supplemented with 10% fetal bovine serum, 2 mM glutamine, but lacking antibiotics. Cells were then transfected using lipofectamine 2000 (Invitrogen). Per dish, a total amount of 7.7 µg DNA was used. Both lipofectamine and DNA were diluted in Opti-MEM reduced serum medium (gibco) and incubated for 5 min separately, before being mixed and incubated a further 20 min together. The lipofectamine-DNA mixture was added dropwise onto the cells and the plates were gently rocked to distribute the transfection mix evenly. Cells were then incubated for 16-24 h before further use. Transfection efficiency and expression of constructs was assessed by measuring eGFP fluorescence using an Olympus IX70 microscope equipped with a QIMAGING Retiga 2000R camera.

2.7 Biochemical interaction assays

2.7.1 Glutathione S-transferase (GST) pull-down using non-magnetic beads

The protocol described here was optimized through extensive troubleshooting (see Chapter 4). Variations of this "standard protocol" are described in Chapter 4 or the figure legends of the respective experiments. The tips of disposable 10 µL pipette tips were crushed with tweezers and used throughout the experiment for aspirating solutions to minimize loss of beads.

40 µL of non-magnetic glutathione sepharose 4B bead suspension (GE Healthcare) were dispensed into Eppendorf tubes and washed twice with 200 µL assay buffer (25 mM HEPES-KOH, 125 mM potassium acetate, 5 mM magnesium acetate, 2 mM EDTA, 2 mM triethylene glycol diamine

tetraacetic acid (EGTA), 1 mM DTT, pH 7.2) by gentle inversion for 2 min before pelleting (10000 g, 2 min, 4 °C) and discarding the supernatant. GST or GST-tagged bait proteins (500 µg) were prepared in assay buffer and 400 µL of the bait protein solution added to the washed beads. From here, all steps of the pull-down experiment were performed at 4 °C. The beads were incubated with the bait protein for 120 min on a rotating wheel. Non-immobilized bait protein was removed by pelleting the samples (10000 g, 2 min). The supernatant was removed and discarded, and the beads washed three times as above with 235 µL assay buffer. Prey proteins (0.1 mg/mL wild-type or mutant His-NTD-NEMO, 0.1 mg/mL wild-type His-NTD, or 0.4 µM purified clathrin) were prepared in assay buffer to a final volume of 300 µL (input) and incubated with the washed beads for 120 min on a rotating wheel. The samples were pelleted as above and the supernatant containing any unbound prey protein kept for analysis. The beads were then washed four times with 1.4 mL of 1x PBS by centrifugation as above. After the last wash, the supernatant was removed completely, the bead pellet was resuspended in 80 µL 1x SDS-PAGE buffer and the samples were eluted by boiling for 5 min at 95 °C. Input and supernatant samples were prepared in 1x SDS-PAGE buffer. Samples were loaded onto 12% SDS-PAGE gels (0.33% of the prey input samples, 11.25% of the eluted pellet samples, 0.6% of the supernatant kept after the prey incubation), resolved by electrophoresis and visualized using Western blotting with an anti-clathrin NTD antibody.

2.7.2 Preparation of liposomes and liposome sedimentation assays

Lipids were obtained from Avanti polar lipids. Liposomes were prepared by mixing chloroform-solubilized 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC), 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE) and 1,2-dioleoyl-sn-glycero-3-[(N-(5-amino-1-carboxypentyl)iminodiacetic acid)succinyl] (nickel salt) (DOGS-NTA-Ni) in a 70:20:10 mass proportion in a glass vial. The mixture was briefly vortexed and then dried under argon whilst rotating until a dry film of lipid was formed. The lipids were resuspended by gentle shaking and vortexing in 20 mM HEPES pH 7.5, 120 mM potassium acetate, 5 mM magnesium acetate to form a milky solution of large liposomes at a final lipid concentration of 2 mg/mL. Liposomes were stored at 4 °C and used within two weeks.

Liposome sedimentation assays were performed in assay buffer (20 mM HEPES pH 7.5, 120 mM potassium acetate, 5 mM magnesium acetate) at 25 °C. In experiments using purified NTD, this

assay buffer was supplemented with 10 mM imidazole to reduce non-specific binding. First, 30 µg of liposomes and 1.8 µM of His6-tagged adaptor protein were mixed in low-binding Eppendorf tubes to a total volume of 50 µL and incubated at 220 rpm for 30 min. The liposomes were then pelleted (3000 g, 5 min, 4 °C) and the supernatant was removed carefully. Purified full-length clathrin (1.0 µM in 50 µL assay buffer) or purified clathrin N-terminal domain (0.25 µM in 50 µL assay buffer) were added and the liposome pellet resuspended. After incubation at 220 rpm for 30 min, the tubes were centrifuged (16000 g, 10 min, 4 °C) and the supernatant was removed. The liposome pellet was resuspended in 50 µL 1x SDS-PAGE loading buffer and heated gently at 37 °C for 30 min whilst shaking to break up clathrin lattices. All samples were then heated at 65 °C for 10 min before boiling them at 95 °C for 5 min. Samples were run on a 10% SDS-PAGE gel (18% of pellet sample, 9% of input, unless otherwise stated) and visualized using Instant Blue Coomassie stain.

2.7.3 Electron microscopy

Electron microscopy of liposomes was performed by Bernard E. Kelly as described previously (68).

2.7.4 GST pull-downs using proteins expressed in wheat germ extract

DNA encoding myc-tagged prey proteins (2-4 µg in 20 µL MQW) was mixed with 30 µL cell-free TNT SP6 wheat germ extract (Promega) and incubated for 120 min at 25 °C to express the prey proteins. The reactions were then mixed thoroughly and 4 µL of each reaction were reserved as input samples (mixed with SDS-PAGE loading buffer to a final volume of 20 µL). The mixtures containing prey protein were diluted with 154 µL of assay buffer (20 mM Tris pH 7.4, 200 mM NaCl, 0.1% NP-40, 1 mM DTT, 1 mM EDTA) and kept on ice. During the wheat germ expression incubation time, magnetic glutathione beads (Pierce) were loaded with bait proteins. Per pull-down, 20 µL of magnetic beads were dispensed into a flat-bottomed 96-well plate and washed twice with assay buffer by placing the plate on a magnet, thus collecting the beads, and aspirating the supernatant. Bait proteins were prepared in assay buffer (2.5 µM, 200 µL final volume) and incubated with the beads for 15 min on a fast-rotating table-top shaking platform. The beads were then washed three times with 200 µL of assay buffer as above. After the last wash, 100 µL of assay buffer and 100 µL of the diluted prey protein sample were added to each well containing bait protein-loaded magnetic beads and incubated for the times indicated. After the prey

incubation, the beads were again collected using the magnet, the supernatant containing unbound proteins was aspirated and discarded, and the beads were washed four times using 200 µL of wash buffer as above. Bound proteins (pull-down) were competitively eluted for 2 min using 48 µL of elution buffer (50 mM reduced glutathione, pH 7.5 in assay buffer). 12 µL of 5x SDS-PAGE buffer were added to each bound (pellet) sample. Samples were separated by SDS-PAGE (3.5 µL of input samples, 14.3 µL of pull-down samples) and visualized by Western blotting using an anti-myc antibody.

2.7.5 Co-immunoprecipitation experiments

HEK293T cells were transfected with eGFP or eGFP fusion proteins and cells were grown for 16-24 h at 37 °C, 5% CO_2. Transfection was confirmed by fluorescence microscopy. Cells were harvested into the medium, pelleted (220 g, 5 min, 4 °C) and washed three times with cold 1x PBS. Cells were then lysed at 4 °C in 1 mL lysis buffer (10 mM Tris/HCl pH 7.5, 150 mM NaCl, 2 mM $MgCl_2$, 0.5% Triton X-100, 1:100 Sigma protease inhibitors, 50 U/mL Sigma benzonase) for the times indicated. In some experiments, a mixture of bovine pancreas DNase I (Sigma) and RNase cocktail (Ambion) was used instead of benzonase. The cell lysate was pelleted (20000 g, 10 min, 4 °C) and the supernatant transferred to fresh tubes. A small sample of the cleared lysate was mixed 1:1 with 2x SDS-PAGE buffer and kept for analysis (input sample). GFP-Trap A beads (ChromoTek, 20 µL per sample) were washed three times in 800 µL wash buffer (10 mM Tris/HCl pH 7.5, 150 mM NaCl, 2 mM $MgCl_2$, 0.05% Triton X-100) by centrifugation (2500 g, 2 min, 4 °C). The beads were then resuspended in 100 µL lysis buffer per sample. Up to 950 µL of cleared cell lysate was added to 90 µL of resuspended beads and the mixture was incubated at 4 °C on a rotating wheel for the times indicated. The sample was then spun down (2500 g, 2 min, 4 °C) and the supernatant removed. A fraction of the supernatant was mixed 1:1 with 2x SDS-PAGE buffer and kept for analysis by SDS-PAGE (supernatant sample). The beads were washed twice with 1 mL wash buffer by spinning (2500 g, 2 min, 4 °C). After washing, all supernatant was carefully removed from the beads, 2x SDS-PAGE buffer was added and the mixture boiled at 95 °C for 10 min to elute proteins bound to the beads. The samples were then centrifuged again to sediment the beads (20000 g, 2 min, 4 °C) and the supernatant (bound sample) was transferred to a fresh tube. Input, supernatant and bound (pellet) samples were separated by SDS-PAGE and analyzed by Western blot. Relative loading of input, supernatant

and bound samples dependent on the experiment and is stated accordingly. If lysis and wash buffers used differed from the here stated ones, it is indicated in the respective figures legends.

2.8 Biophysics

2.8.1 Circular dichroism

Circular dichroism spectra were recorded on a Jasco J-810 spectropolarimeter under nitrogen atmosphere. All measurements were performed at 20 °C using a Hellma 106-QS quartz cuvette. Protein samples were measured at a concentration of 1 mg/mL in 50 mM phosphate buffer, pH 7.4. Eight spectra per sample were recorded (50 nm/min, 1 nm bandwidth, 260-190 nm), averaged and the curves were smoothed using the means-movement method, convolution width 15, using the Jasco analysis software. Buffer blank spectra were recorded identically.

2.8.2 Multi-angle light scattering

Experiments were performed at 22 °C using a Superdex 200 10/300 gel filtration column (GE Healthcare) equilibrated in 20 mM Tris pH 7.5, 200 mM NaCl, 1 mM DTT. Samples (100 µL) were injected at a flow rate of 0.5 mL/min and size exclusion was followed by inline measurement of static light scattering (DAWN 8+, Wyatt Technology), differential refractive index (Optilab T-rEX, Wyatt Technology) and 280 nm UV absorbance (Agilent 1260 UV, Agilent Technologies). The data were analyzed using Astra6 (Wyatt Technology).

2.8.3 Fluorescence polarization

Fluorescence polarization experiments were performed at 30 °C on a SpectraMax i3 fluorimeter using 20 mM Tris pH 7.5, 200 mM NaCl, 1 mM DTT as the assay buffer. The fluorescein isothiocyanate- (FITC-) labeled amphiphysin and YXXΦ peptides were provided by Bernard T. Kelly. Dilution series of His-GST, wild-type and mutant His-NTD were prepared in assay buffer. 50 µL of 20 nM labeled peptide in assay buffer were mixed with 50 µL of each protein dilution and transferred into black 96-well half-area non-binding plate (Corning). The plate was centrifuged (1000 g, 1 min) and incubated at 27 °C, 220 rpm for 45 min before measurement. The parallel and perpendicular fluorescence intensities of all samples were corrected for intrinsic protein fluorescence before anisotropy values were calculated.

2.8.4 Differential scanning fluorimetry

Due to equipment availability, two different experimental setups had to be employed. Experiments were either performed in 48-well white PCR plates (BioRad) using SYPRO Orange dye (Molecular Probes) on a MiniOpticon real-time PCR system (BioRad), or in 96-well clear PCR plates (Axygen Scientific) using protein thermal shift dye (Life Technologies) on a Viia7 real-time PCR system (Applied Biosystems). The PCR plates were centrifuged before reading (1000 g, 3 min, 25 °C). On the MiniOpticon system, samples were heated from 20 °C to 90 °C in 1 K/min increments, and on the Viia7 system, samples were heated from 25 °C to 95 °C at a rate of 1 K/20 s. Throughout this thesis, the platform used for specific experiments is indicated.

Experiments involving wild-type or mutated His-NTD or His-NTD-NEMO were performed using 20 mM HEPES-KOH, 120 potassium acetate, pH 7.5 as assay buffer, and the protein and dye concentrations indicated in final volumes of 50 µL (MiniOpticon) or 20 µL (Viia7). Differences between the T_M of wild-type His-NTD and mutants could be determined consistently using either platform. For the NTD binding quantitation experiments, the HDAg-L1 or Amph4T1 peptide (same peptides as used in crystallization experiments) was dissolved in assay buffer, the pH adjusted to 7.5, and dilution series were prepared. Experiments were performed using these peptide dilutions, and the indicated wild-type or mutant His-NTD and dye concentrations in final volumes of 50 µL (MiniOpticon) or 20 µL (Viia7).

Experiments with UL21(1-535)-His were performed on the MiniOpticon system using 20 mM Tris pH 7.5, 200 mM NaCl, 1 mM DTT as assay buffer, and protein and dye concentrations as indicated in a total reaction volume of 50 µL. For buffer screening, reactions were assembled to contain 40 µL of the above buffer or buffers from a commercial screen (JBS solubility screen, Jena Bioscience), 10 µg of purified UL21 and 10x SYPRO Orange dye in a final reaction volume of 50 µL.

The melting temperature T_M was calculated by non-linear least-squares fitting of the melting curves to the following equation describing an unfolding transition (292):

$$y = \frac{(m_f \cdot x + b_f) + (m_u \cdot x + b_u) \cdot e^{m \cdot (x - T_M)}}{1 + e^{m \cdot (x - T_M)}}$$

The parameters b_f and b_u describe the equilibrium signals of the folded and unfolded state, respectively, m_f and m_u the slopes of the folded and unfolded regimes, m the slope at the midpoint of unfolding and T_M is the temperature at the midpoint of unfolding. Curve fitting, melting temperature calculations and plotting were performed using MATLAB.

2.8.5 Isothermal titration calorimetry

Isothermal titration calorimetry experiments were performed using a nano-isothermal titration calorimeter (TA Instruments). Peptides (AmphCBM-long$_{pep}$, AmphCBM$_{pep}$, HDAg-L3$_{pep}$, same as used for crystallization experiments) and NTD were resuspended in the indicated buffers. Peptides were dialyzed using a 100 µL reservoir Biodialyzer with a 500 Da pore size membrane (Sigma). Buffers were supplemented with tris(2-carboxyethyl)phosphine hydrochloride (TCEP, Sigma) as a reducing agent. Titrations of peptide into buffer, buffer into protein, or peptide into protein were performed at 25 °C using 2 µL injections with 180 s or 300 s intervals (25 injections total).

2.9 Bioinformatic analysis of HSV-1 proteins and interaction partners

Sequence-based prediction of structural features and disorder of selected HSV-1 proteins (Table 7) was performed using the tools listed in Table 8. Sequence homology-based 3D structure modeling of UL56 interaction partners was performed using HHpred (293) (accessed via https://toolkit.tuebingen.mpg.de/hhpred#). Multiple sequence alignments were performed using Clustal Omega (294) and visualized using Jalview 2 (295).

Table 7 HSV-1 proteins used in bioinformatic analyses.

Protein	HSV-1 strain	Uniprot ID	Size (amino acids)
UL11	KOS (JQ673480.1)	F8REU4	96
UL14	KOS (JQ673480.1)	H9E920	219
UL16	KOS (JQ673480.1)	G8HBC9	373
UL21	KOS (JQ673480.1)	F8RG07	535
UL49	KOS (JQ673480.1)	Q6GYI9	301
UL56	KOS (JQ673480.1)	H9E964	234
US9	17 (JN555585.1)	P06481	90
gE	KOS (JQ673480.1)	D3YPM6	550

Table 8 Bioinformatic tools used for analysis of HSV-1 proteins.

Name	Feature predicted	Accessed via	References
TMHMM v. 2.0	range of transmembrane domains	http://www.cbs.dtu.dk/services/TMHMM/	(296)
RONN v3.2	protein disorder	https://www.strubi.ox.ac.uk/RONN	(297)
DISOPRED3	protein disorder	http://bioinf.cs.ucl.ac.uk/psipred/	(298, 299)
NetSurfP v. 1.1	secondary structure	http://www.cbs.dtu.dk/services/NetSurfP/	(300)
PSIPRED v. 3	secondary structure	http://bioinf.cs.ucl.ac.uk/psipred/	(298, 301)
MEMSAT3	transmembrane protein topology	http://bioinf.cs.ucl.ac.uk/psipred/	(298, 302)

2.10 Yeast cell culture and yeast two-hybrid screening

Yeast two-hybrid interaction screening was performed using the Matchmaker Gold Yeast Two-Hybrid System (Takara) with a pre-transformed, universal human normalized Mate&Plate prey library (Clontech, 630480, Lot. No. 1103584A). Agar, yeast nitrogen base without amino acids and amino acid supplements (Kaiser formulation) were obtained from Formedium. All yeast growth medium and agar plates were prepared, sterilized (or sterile filtered) and stored at 4 °C. Liquid cultures and plated yeast cells were incubated at 30 °C. Nutritional marker selection media and plates used are listed in Table 9.

Table 9 Yeast media and plates used in screen.

Medium/plate	Explanation	Selection of
-Trp	yeast growth medium lacking tryptophan	yeast transformed with pGBKT7 plasmid (bait)
-Leu	yeast growth medium lacking leucine	yeast transformed with pGADT7 (prey)
-Trp -Leu (double dropout, DDO)	yeast growth medium lacking tryptophan and leucine	successfully mated yeasts (bait and prey)
-Ade -His -Leu -Trp (quadruple dropout, QDO)	yeast growth medium lacking adenine, histidine, tryptophan and leucine	positive interaction of bait and prey fusion proteins
QDO with added Aureobasidin A and X-α-Gal (QDO/X/A)	yeast growth medium lacking adenine, histidine, tryptophan and leucine, supplemented with Aureobasidin A and X-α-Gal (5-Bromo-4-chloro-3-indolyl-α-D-galactopyranoside)	positive interaction of bait and prey fusion proteins (high stringency)

2.10.1 Resuscitation of yeast strains

Small scrapes of frozen yeast strains (Y2H Gold and Y187) were dropped onto a fresh 1x yeast extract peptone dextrose (YPD) plates, thawed on the plates and spread. Colonies were ready for use after five days.

2.10.2 Transformation of yeast

A yeast strain culture was grown overnight in 1x YPD at 220 rpm. The optical density (OD, A_{600}) of the culture was measured and adjusted with 1x YPD to obtain a 60 mL culture with an OD of 0.2. This starter culture was incubated at 220 rpm and its OD monitored at regular intervals until it reached 0.4-0.6. The volume of cells needed for ten transformations (40-60 mL) was pelleted (700 g, 5 min). The supernatant was removed and the pellet resuspended in 20 mL sterile MQW. The suspension was centrifuged again (700 g, 5min), the supernatant removed and the pellet resuspended in 1.5 mL 1.1x TE/LiAc buffer (1.1x Tris-HCl/EDTA pH 8.0, 0.11 M lithium acetate pH 7.5) and transferred into a 1.5 mL tube. The tube was pelleted at high speed (20000 g, 15 s), the supernatant discarded and the resulting pellet resuspended in 0.5 mL 1.1x TE/LiAc buffer and kept on ice.

For each transformation, 5 µL of Yeastmaker carrier DNA (Clontech) were added to 50 µL of cells resuspended in 1.1x TE/LiAc and the mixture was transferred to a fresh tube containing 0.5 mL 40% (w/v) PEG3350, 1x TE, 0.1 M lithium acetate pH 7.5. Carrier DNA was boiled at 100 °C and then chilled on ice immediately before use. Per transformation, 100 ng of plasmid DNA were mixed with 0.5 mL of the prepared cell suspension and mixed gently by inversion. The transformation reaction was incubated at 30 °C for 30 min and mixed every 10 min. Per tube, 20 µL of DMSO were added and the cells heat shocked for 15 min at 42 °C. Immediately afterwards, the reactions were centrifuged briefly (20000 g, 15 s) and the supernatant was discarded. The pellet was resuspended in 1 mL 1x YPD medium and incubated for 1 h at room temperature on a rocker. Cells were spun down again (20000 g, 15 s), the supernatant discarded and the pellet resuspended in 50 µL MQW. The cells were plated onto suitable dropout plates and colonies appeared after two to three days.

2.10.3 Autoactivation screen

Yeast strain Y2H Gold was transformed with bait plasmids containing the proteins of interest (pGBKT7-gE(445-550), pGBKT7-UL21(1-535), pGBKT7-UL56(1-207), pGBKT7-UL56(1-215)) or positive and negative bait control plasmids (pGBKT7-53, pGBTK7-lam, pGBKT7 empty vector). Yeast strain Y187 was transformed with prey plasmids (pGADT7-T, pGADT7 empty vector). Colonies were picked and resuspended in 50 µL sterile MQW. Bait and prey yeast resuspensions were then mixed in a 1:1 ratio and plated as 2 µL drops onto a 1x YPD dish.

Colonies were grown for several days and then the plate was replicated onto a DDO plate. After three days, the replicated plate was patched onto a second DDO plate and incubated for four days. The patched DDO plate was then replicated onto a QDO plate. Autoactivation was monitored over several days and the final result obtained twelve days after replication onto the QDO plate.

2.10.4 Mating of yeast with prey cDNA library and yeast two-hybrid screen

A colony of yeast transformed with pGBKT7-UL56(1-207) was picked into 65 mL -Trp medium and incubated at 220 rpm overnight until an OD of 0.8 was reached. 50 mL of this culture were pelleted (1000 g, 10 min), the supernatant was discarded and the pellet resuspended in 4.5 mL -Trp medium. The cell suspension was transferred into a large conical flask containing 45 mL 2x YPD medium with added kanamycin (50 µg/mL). One vial of pre-transformed yeast expressing a human cDNA library (Mate&Plate) was quickly thawed, mixed and 198 µL added to the bait culture. The culture was incubated for 24 h at 40 rpm to allow mating of the yeast cells expressing the bait construct with library yeast cells. The titer of the library was determined by using the remainder 2 µL of the library to make serial dilutions and plating them onto -Leu plates.

After 24 h, the mating culture was pelleted (1000 g, 10 min) and the supernatant discarded. The conical flask was washed with 45 mL 0.5x YPD with added kanamycin (50 µg/mL) and the rinse used to resuspend the yeast cell pellet. The suspension was pelleted again (1000 g, 10 min) and the flask wash step repeated. The final cell pellet was resuspended in 7.5 mL 0.5x YPD with added kanamycin (50 µg/mL). 50 µL of this cell suspension were removed and used to set up serial dilutions, which were plated onto -Leu, -Trp and DDO plates to calculate the number of diploids screened and the mating efficiency. The remainder of the mated yeast cells were plated onto 15 cm QDO plates (380 µL per plate).

The QDO plates were incubated for one week. Colonies were then picked at regular intervals and resuspended into 250 µL yeast freezing medium (85% 1x YPD, 15% glycerol), keeping 1 µL of this cell suspension for streaking onto QDO/X/A plates. Colonies were grown again for several days and re-streaked onto fresh QDO/X/A plates. Colonies which grew on these plates were picked and sequenced.

2.10.5 Yeast colony PCR and analysis of hits

Per reaction, the PCR mix was assembled as below. A single, medium-sized yeast colony was picked and swirled in the mixture. The PCR was performed as below and samples were analyzed by agarose gel electrophoresis. Samples were then PCR-purified and sequenced using primers pGADT7-f and pGADT7-r.

Sequences of positive clones were searched against a human genomic + transcript database using megablast (part of the BLAST/blastn suite). Mapping of open reading frames was performed manually.

Per reaction:	
67 µl	MQW
10 µL	10x Rxn buffer (- MgCl2)
4 µL	50 mM MgCl2
8 µL	dNTP mix (2.5 mM each)
5 µL	10 µM primer pGADT7-long-f
5 µL	10 µM primer pGADT7-long-r
1 µL	Taq polymerase
100 µL	total

PCR cycle:		
5 min	94 °C	
15 s	94 °C	◄┐
15 s	67 °C	x35
3.5 min	72 °C	┘
hold	10 °C	

2.11 Stable isotope labeling by amino acids in cell culture (SILAC) interaction screening

2.11.1 Cell culture, transfection and co-immunoprecipitation

HEK293T cells were grown in SILAC medium (high glucose DMEM lacking arginine and lysine, Life Technologies) supplemented with 10% dialyzed fetal bovine serum (10 kDa cutoff), 2 mM glutamine, 100 U/ml penicillin and 100 µg/ml streptomycin. Unlabeled arginine and lysine were added to the "light", Arg6 ($^{13}C_6$) and Lys4 ($^{2}H_4$) to the "medium", and Arg10 ($^{13}C_6$, $^{15}N_4$) and

Lys8 ($^{13}C_6$, $^{15}N_2$) to the "heavy" SILAC medium. Cells were maintained SILAC medium for at least five passages before use to ensure complete labeling.

Cells were seeded into 9 cm dishes (5 × 106 cells/dish) and grown for 24 h. Transfection with eGFP, eGFP-UL56(1-234) or UL56(1-207)-eGFP was performed according to the standard lipofectamine procedure (section 2.6.2), but using labeled SILAC light, medium or heavy medium supplemented as above to make up the transfection reagents instead of Opti-MEM. Cells were incubated for further 24 h and transfection was confirmed by fluorescence microscopy.

Cells lysis and co-immunoprecipitation were performed as before (section 2.7.5) with 70 min lysis and 40 min bead incubation time. A BCA assay was performed to ensure equal loading of lysate onto the GFP Trap A beads (total protein input per sample: 3.33 mg). After elution, 8 µL of each light, medium and heavy labeled samples were mixed in a 1:1:1 ratio and submitted for mass spectrometry. Samples were analyzed by Western blot as described before (section 2.7.5).

2.11.2 Mass spectrometry

Mass spectrometry analysis was performed by the proteomics facility of the University of Bristol (UK).

2.11.3 Data processing and bioinformatic analysis

The raw data files were processed using MaxQuant v. 1.5.6.0 (303). The in-built Andromeda search engine (304) was used to search against the human proteome containing canonical and isoform entries (uniport proteome up000005640, dated 11.09.2016) plus a custom FASTA file containing the full-length HSV-1 strain KOS UL56 sequence (uniprot ID H9E964). Trypsin/P digestion, standard modifications (oxidation, N-terminal acetylation) were selected as group-specific parameters and SILAC quantification was performed using light (Arg0, Lys0), medium (Arg6, Lys4) and heavy (Arg10, Lys8) labels. Re-quantification, razor protein false discovery rate, and second peptide options were enabled for the processing. Minimum peptides was set to 1, minimum peptide length was set to 7, and maximum false discovery rates were set to 0.1. The quantified data were analyzed with Perseus v. 1.5.6.0 (305) using the normalized ratios obtained by MaxQuant. Proteins only identified by site or against the reverse database, as well as common experimental contaminants such as keratins (specified in the MaxQuant contaminants file) were removed and the experiments grouped by biological repeat. Only proteins identified in at least

two of the three biological repeats were considered for analysis. A one-sample, two-sided t-test with a threshold p-value of 0.05 was performed on the groups to identify significantly enriched proteins. Proteins with a $\log_2$ fold change greater than 1 and a p value smaller than 0.05 were designated as "hits". Hits were screened against a contaminant repository that lists proteins identified in negative control affinity purification followed by mass spectrometry experiments, the CRAPome (306). Interaction network analysis of hits was carried out using the STRING server v. 10.0 (307) with a minimum required interaction score of 0.4 (medium confidence).

3 STRUCTURAL CHARACTERIZATION OF VIRAL AND CELLULAR PEPTIDES BOUND TO CLATHRIN N-TERMINAL DOMAIN

The large antigen of hepatitis D virus (HDAg-L) contains a putative clathrin-binding motif (CBM) in its C-terminal domain and has been reported to exploit clathrin in order to facilitate egress of viral particles from host cells (106, 107). I wanted to characterize the molecular basis of the interaction between NTD and HDAg-L and compare it to binding of cellular clathrin adaptors like amphiphysin and β2 adaptin, a subunit of the AP2 complex, to further our understanding of NTD:adaptor proteins, and to potentially identify targets for novel therapeutics.

In this chapter, I first briefly describe the purification of clathrin NTD (section 3.1). Then, I present structures of peptides encompassing CBMs from two different HDV genotypes bound to NTD (section 3.2). Crystallization of peptides representing a third group of HDV C-terminal extensions was also attempted, but proved unsuccessful (section 3.2.2). Surprisingly, both viral peptides were found to associate with multiple binding sites on NTD simultaneously. Therefore, I decided to reassess binding between known adaptors and clathrin and present structures of four cellular CBMs, originating from amphiphysin and AP2, bound to clathrin NTD (section 3.3). Excitingly, these structures also demonstrated binding of the CBM peptides to multiple sites on NTD. In addition, the structures of viral and cellular peptides on NTD enabled me not only to identify the fourth binding site on clathrin, which we termed the "Royle box", but also to shed new light on the binding mode of peptides at the arrestin box (sections 3.4 and 3.5).

Note: This chapter and the following one are based on and expand the paper "Cellular and viral peptides bind multiple sites on the N-terminal domain of clathrin" published in 2017 (272).

3.1 Purification of clathrin N-terminal domain

Clathrin N-terminal domain was expressed overnight as an N-terminal GST fusion protein in BL21(DE3) pLysS *E. coli* cells. Affinity purification using glutathione sepharose resin was performed, the GST-tag cleaved using thrombin, and the protein further purified by size exclusion chromatography into 10 mM Tris pH 7.5, 50 mM NaCl, 4 mM DTT (Figure 17). Using this

method, highly pure protein was routinely obtained and could easily be concentrated to up to 50 mg/mL.

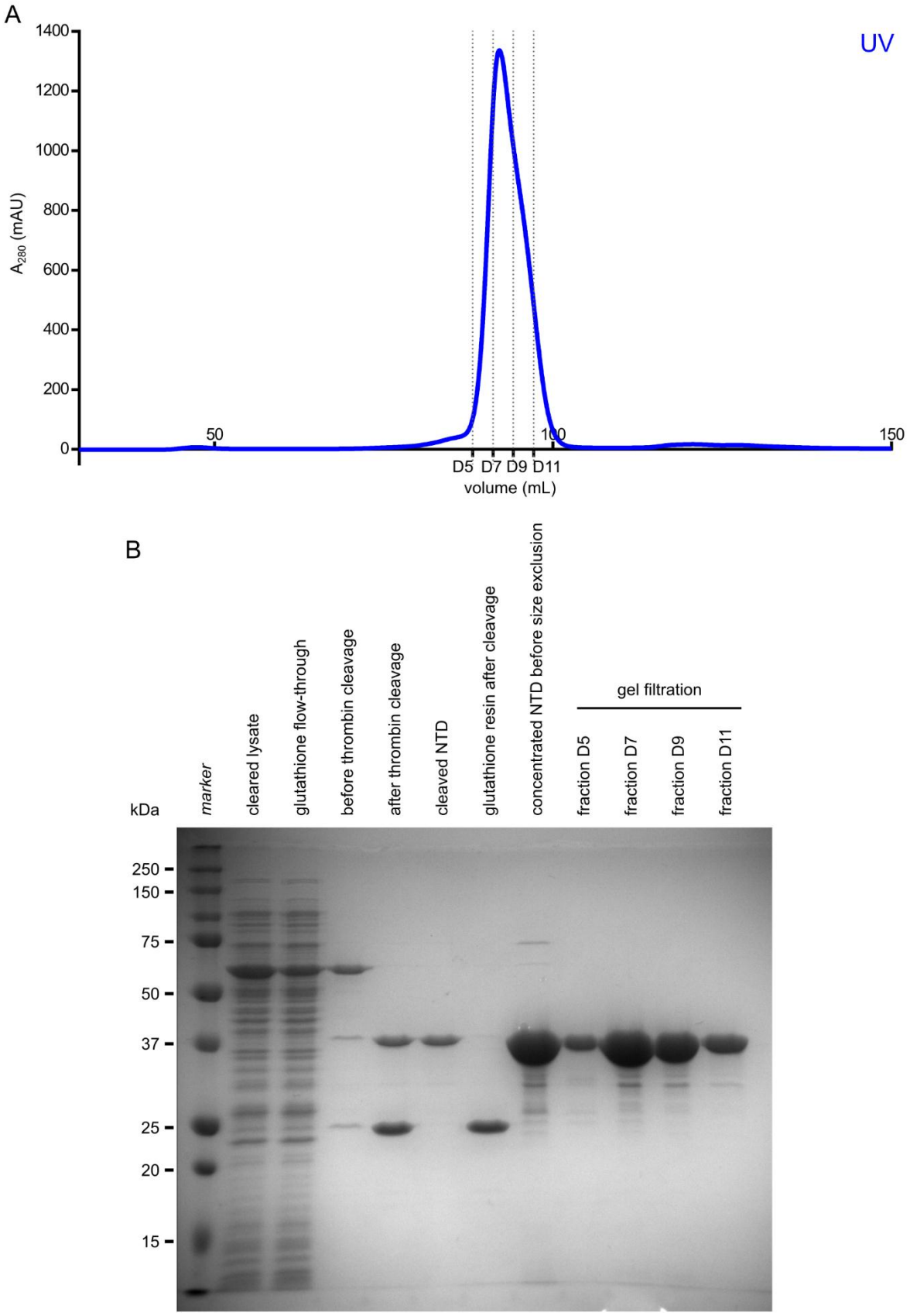


Figure 17 Purification of clathrin NTD. The protein was expressed with an N-terminal GST-tag in BL21(DE3) pLysS competent *E. coli* cells and first purified by affinity chromatography using a glutathione sepharose 4B resin. The GST-tag was cleaved overnight by incubation with thrombin and the cleaved product further purified by size exclusion. (A) Size exclusion elution profile of NTD on an S200 16/600 gel filtration column (column volume: 120 mL). (B) Coomassie-stained 12% SDS-PAGE gel showing samples from key steps of the purification process and selected fractions from the size exclusion chromatography.

3.2 Crystallization of NTD with viral peptides

3.2.1 Co-crystallization of NTD with HDAg-L1$_{pep}$ and HDAg-L2$_{pep}$

To crystallize NTD in complex with HDAg-L1$_{pep}$ (SDILFPADS) and HDAg-L2$_{pep}$ (SPRLPLLES), a co-crystallization strategy was followed. Purified NTD was mixed with the respective peptides and used in crystallization experiments. First, numerous commercial crystallization screens were attempted using NTD concentrations ranging from 12 mg/mL up to 20 mg/mL mixed with HDAg-L1$_{pep}$ or HDAg-L2$_{pep}$.in a 1:10 or 1:7 molar ratio. Such high concentrations of peptide were chosen to ensure saturation of the peptide binding sites. In these broad screens, thin, plate-shaped crystals grew in multiple wells containing sodium malonate. Therefore, sodium malonate was taken forward as the lead condition and further optimized. Close screens varying the pH and the concentration of sodium malonate, the concentration of the protein:peptide mixtures, the molar ratio between NTD and the peptides, as well as the drop ratio between reservoir and protein:peptide mixture used for preparing the sitting drops were performed. Crystals grew reproducibly in these conditions and diffracted to a resolution between 2.0 Å and 2.9 Å. Data sets from numerous crystals were collected for both NTD:HDAg-L1$_{pep}$ and NTD:HDAg-L2$_{pep}$. The data sets were processed to evaluate the quality of the obtained data by comparing the resolution, $CC_{1/2}$, R_{merge} and completeness. The structures were solved by isomorphous or molecular replacement and preliminary models built.

For NTD:HDAg-L1$_{pep}$, the crystal yielding the best diffraction data was grown in a drop containing 400 nL of protein:peptide mixture (14 mg/mL NTD, 3.4 mM HDAg-L1 peptide) plus 200 nL of the reservoir solution (1.21 M sodium malonate pH 7.0) equilibrated against an 80 µL reservoir (Figure 18, left panel). For NTD:HDAg-L2$_{pep}$, the best diffraction data were obtained from a crystal grown in a drop containing 200 nL of protein:peptide mixture (20 mg/mL NTD and 3.55 mM HDAg-L2 peptide) plus 400 nL of the reservoir (1.75 M sodium malonate pH 7.0) against an 80 µL reservoir (Figure 18, right panel). The structures of the two co-crystals were solved by isomorphous replacement using a high resolution ligand-free model of NTD (PDB ID: 1C9I, (62)) and subsequently refined to 2.2 Å and 2.0 Å, respectively (Table 10).

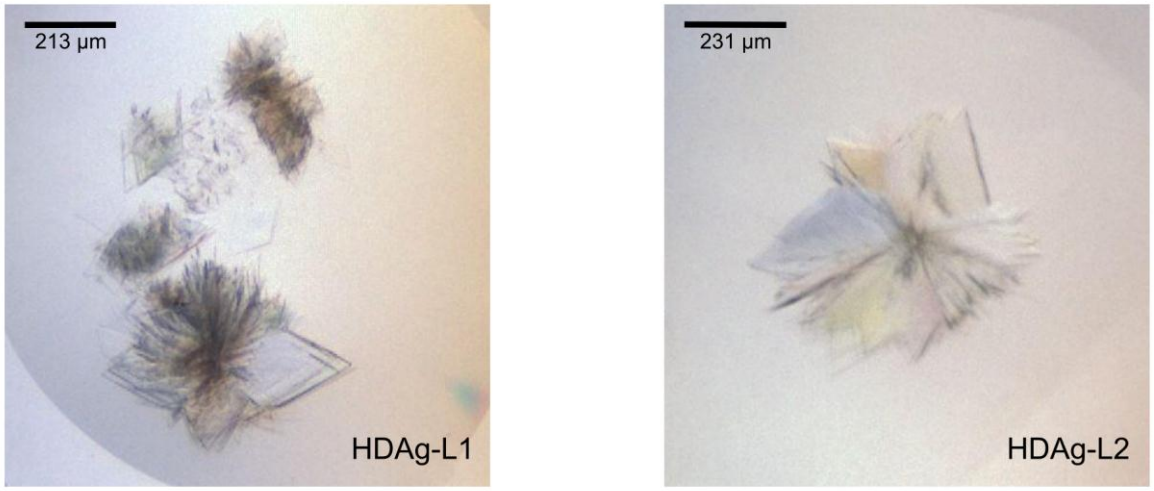


Figure 18 Crystals of NTD in complex with HDAg-L1 and HDAg-L2. A crystal from each of these drops was used for the determination of the respective structures of the complexes.

Table 10 Data collection and refinement statistics of structures of NTD in complex with viral peptides. Values for the highest resolution shell are shown in parentheses. One crystal was used for each structure.

NTD:	**HDAg-L1$_{pep}$**	**HDAg-L2$_{pep}$**
Data collection		
Space group	C2	C2
Cell dimensions		
a, b, c (Å)	136.2, 131.2, 77.9	136.9, 131.2, 78.5
α, β, γ (°)	90, 115.6, 90	90, 115.9, 90
Resolution (Å)	48.4-2.2 (2.21-2.15)	39.2-2.0 (2.00-1.96)
R_{merge}	0.101 (0.930)	0.081 (0.581)
$<I/\sigma I>$	9.6 (1.5)	7.2 (1.5)
$CC_{1/2}$	0.996 (0.501)	0.996 (0.563)
Completeness (%)	99.7 (99.0)	94.6 (95.9)
Redundancy	4.5 (4.3)	2.5 (2.3)
Refinement		
Resolution (Å)	48.4-2.2 (2.21-2.15)	39.2-2.0 (2.01-1.96)
No. of reflections (work/free)	63,339/3381	79,976/4224
R_{work}/R_{free}	0.175/0.207	0.169/0.193
Ramachandran favored regions (%)	98.2	99.1
Ramachandran outliers (%)	0.0	0.0
No. of atoms		
Protein	5634	5634
Glycerol	6	6
Peptide ligands	200	240
Water	437	583
Average B-factors		
Protein	39.8	32.0
Glycerol	32.7	20.9
Peptide ligands	70.2	61.0
Water	45.6	38.2
r.m.s. deviations		
Bond lengths (Å)	0.020	0.019
Bond angles (°)	1.897	1.917
PDB ID	**5M5U**	**5M5V**

3.2.2 Crystallization attempts with HDAg-L3

Hepatitis D virus isolates belonging to HDV clade 3 do not carry a peptide motif resembling the canonical CBM in their C-terminal extension (HDAg-L3), but there is evidence that they bind to clathrin nevertheless (177). Since many hydrophobic amino acids and a tryptophan residue are present in this C-terminal region, it could be possible that the interaction between HDAg-L3 and clathrin occurs not at the clathrin box, but at the W box, for which the consensus motif PWXXW has been suggested (4). Therefore, crystallization of NTD with a peptide representing this group of HDV C-terminal domains, HDAg-L3$_{pep}$ (DGYYWVPGS) was attempted.

Crystallization was attempted using various commercial broad screens and different concentrations of NTD mixed with according amounts of HDAg-L3$_{pep}$ to reach 1:10 or 1:7 molar ratios. Some thin, hexagonal crystals were found in conditions containing 30% Jeffamine M-600, 0.1 M MES pH 6.5, 0.05 M CsCl or 20% Jeffamine M-600, 0.1 M HEPES pH 7.5 and could be reproduced in optimization screens. The crystals diffracted weakly, but data processing revealed the unit cell size to be too small to fit a single NTD molecule. Therefore, the crystals most likely contained crystallized peptide only and thus, these conditions were not further pursued. Thin, needle-shaped, crystals were found in wells containing 28% PEG400, 0.1 M HEPES pH 7.5, 0.2 M $CaCl_2$, or 24% PEG1500, 20% glycerol, or 30% PEG MME 2000, 0.1 M potassium thiocyanate, or 25% PEG3350, 0.1 M Bis-Tris pH 5.5, 0.2 M NaCl. Other conditions showing low-quality crystals or crystal-like precipitate contained 20% PEG8000, 0.1 M HEPES pH 7.5, or 20% PEG3350, 0.2 M LiCl, or 10% isopropanol, 0.1 M HEPES pH 7.5, 0.2 M NaCl. These conditions were screened by varying the concentration of the precipitants and drop ratio between the protein:peptide mixture and reservoir. Hair-thin needles grew in these optimization screens, were harvested and cryoprotected carefully. However, only very weak data diffracting to low resolution were obtained from these crystals and could not be processed.

It seemed possible that the chosen peptide was interfering with crystallization. Therefore, a second HDV peptide, HDAg-L3-long$_{pep}$ (DQFPWYGFTPPPPGYYWVPGD), was tested. This peptide was much longer than the first one and encompassed all of the C-terminal extension of HDAg of HDV clade 3 bar the terminal four amino acids in order to maximize the chances of it containing the putative binding motif for interaction with NTD, thus promoting crystallization. Broad screens of different concentrations of NTD mixed with HDAg-L3-long$_{pep}$ in a 1:10 molar ratio were attempted and some conditions containing promising precipitate were optimized further, but no crystals could be obtained.

Since co-crystallization of NTD and HDAg-L3$_{pep}$ or HDAg-L3-long$_{pep}$ was not successful, a soaking strategy was pursued next. In order to perform soaking experiments, crystals of NTD without added peptide had to be produced reliably. An NTD crystallization condition similar to the one published by von Kleist et al. (20% PEG 3350, 0.15 M potassium acetate, 0.1 M Tris, pH 8.0) (274) was used as a starting point and produced thin needles (condition: 19%-35% PEG3350, 0.1 M Tris pH 7.1, 0.15 M potassium acetate). These needles did not diffract well and the condition

was therefore further optimized by varying the pH, drop ratios and by adding 5% glycerol to potentially slow down crystal growth and promote formation of bigger crystals. The condition found to deliver the best and most reproducible crystals in these optimization screens was 28% PEG3550, 0.1 M Tris pH 8.0, 0.15 M potassium acetate, with a 1:2 protein:peptide mixture to reservoir drop ratio and no added glycerol. Therefore, an additive screen was set up with this condition. Various of the many crystals obtained in this additive screen were selected for soaking and harvested. High concentrations of HDAg-L3$_{pep}$ were used to soak the crystals during cryoprotection.

Diffraction data were collected for multiple of these crystals and showed that various volatile components improved the diffraction. For two of them, grown in 25.2% PEG3350, 0.09 M Tris pH 8.0, 0.135 M potassium acetate and 1.8% ethanol or 2.4% isopropanol, respectively, data processing and molecular replacement could be performed. These crystals diffracted to 2.4 Å and 2.0 Å, respectively and therefore could be confidently assessed. However, no differences in electron density indicating the presence of peptides were found at any of the peptide binding sites on NTD.

3.2.3 Structures of viral peptides bound to NTD

Structures of HDAg-L1$_{pep}$ and HDAg-L2$_{pep}$ bound to NTD were solved and binding of peptides to the interaction sites on NTD was assessed. Surprisingly, electron density indicating the presence of peptides was found not only at the clathrin box, where the canonical CBM found in HDAg-L1 and HDAg-L2 would be predicted to bind, but also at the arrestin box (in the NTD:HDAg-L2$_{pep}$ complex only) and at a third site between blades 6 and 7 of the β-propeller where peptide binding had not previously been structurally characterized (Figure 19). Since this region matches the location of the fourth binding site on NTD described by Willox and Royle (2), it is called “Royle box” throughout this thesis.

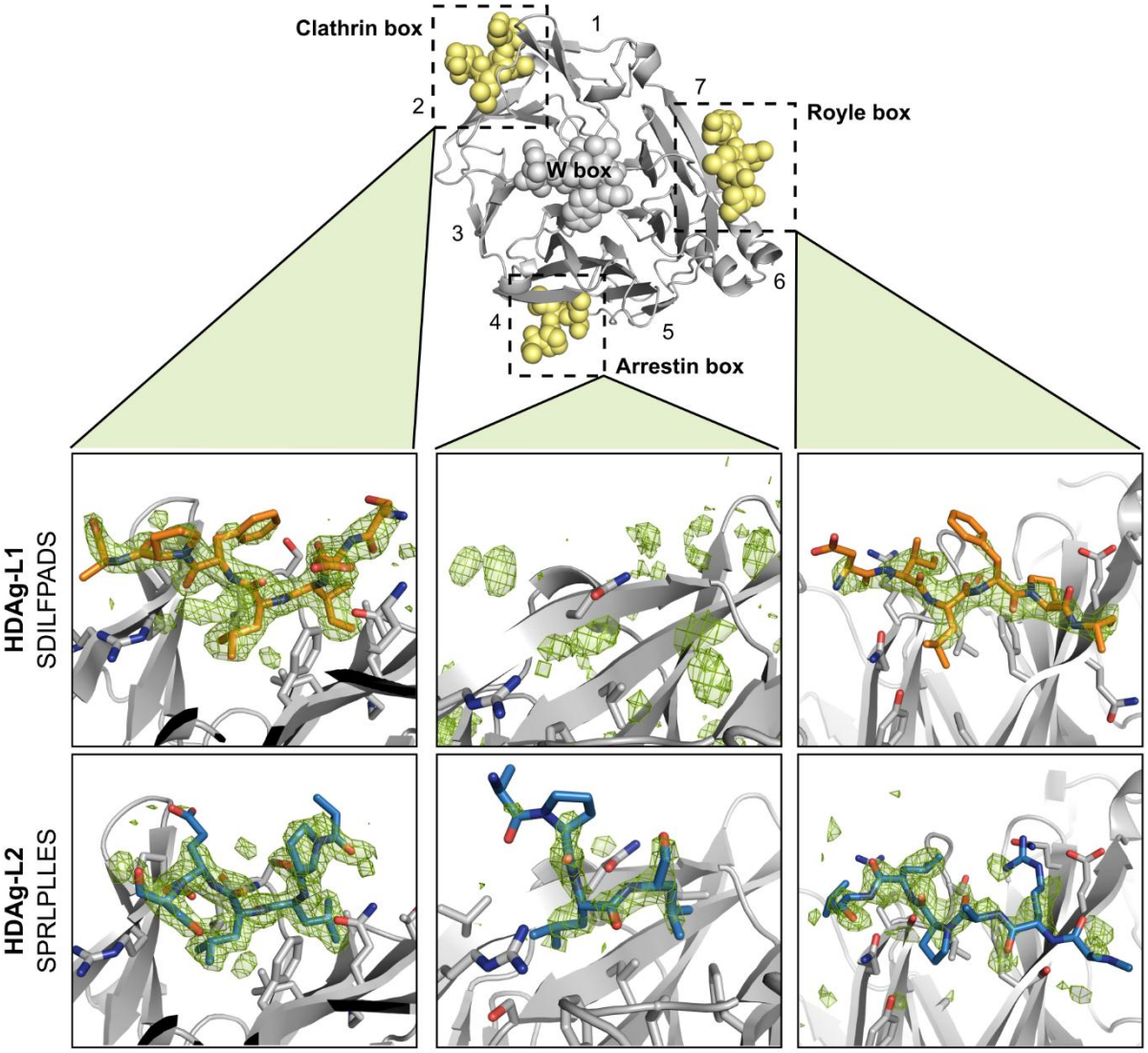


Figure 19 Structures of the viral peptides HDAg-L1 and HDAg-L2 bound to clathrin NTD. The unboxed image shows a cartoon of clathrin NTD (grey). The seven blades of the β-propeller fold are numbered. Spheres represent peptides bound at the four interaction sites on NTD. Zoomed in, boxed images illustrate the viral peptides (HDAg-L1: SDILFPADS, orange sticks; HDAg-L2: SPRLPLLES, light blue sticks) bound at the clathrin box, arrestin box and Royle box on clathrin NTD (grey ribbons, selected side chain atoms shown as sticks). Unbiased F_O-F_C difference electron density (green mesh, shown at 3 σ) was calculated before the peptides were added to the respective models and used to place the peptides.

At the clathrin box, distinct peptide density enabled unambiguous modeling of the peptide orientation and the side chains of the bound residues. The backbone trajectories of the two viral peptides agree very well with a previously published structure of a β-arrestin 2 peptide binding at the clathrin box (PDB ID: 1C9L) (62). Hydrogen bonds formed between the side chain oxygen of Gln89, the backbone carbonyl oxygen of Pro65 and the backbone amide nitrogen of Ser67 with respective peptide backbone oxygen and nitrogen atoms are conserved between the bound viral peptides and the β-arrestin 2 peptide (Figure 20).

Another conserved key feature between these three structures is the occupation of a hydrophobic groove formed by blades 1 and 2 of NTD by two hydrophobic residues of the peptides. In the structure of NTD in complex with the β-arrestin 2 peptide (CBM: $_{373}$LIEFE$_{377}$), residue Leu373 binds to a hydrophobic pocket between residues Thr87, Leu82 and Ala84 of NTD, and the residue Ile374 binds to a deep, highly hydrophobic pocket framed by the side chains of NTD residues Val50, Ile66, Ile80, Leu82, Phe91 and the hydrophobic part of the side chain of Arg64. In the NTD:HDAg-L1$_{pep}$ structure (peptide $_{1}$SDILFPADS$_{9}$), these two adjacent hydrophobic pockets are occupied by the peptide residues Ile3 and Leu4, and for the NDT:HDAg-L2$_{pep}$ complex (peptide $_{1}$SPRLPLLES$_{9}$), peptide residues Leu6 and Leu7 bind to this region (Figure 19 and Figure 20). Interestingly, the residues of the viral peptides binding to these pockets do not match the CBM consensus sequence. The CBM of HDAg-L1 predicted by Huang et al. was LFPAD (106, 107), but instead of residues Leu and Phe binding to the hydrophobic cleft, the isoleucine residue N-terminal to the consensus motif (Ile3 of the peptide used for crystallization) is involved in binding, resulting in peptide residues ILFPA occupying the positions of the CBM. For HDAg-L2, Huang et al. predicted the CBM to be LPLLE (106). However, it is not the first two residues Leu and Pro, but the adjacent leucine doublet (Leu6 and Leu7 of the peptide used for crystallization) that binds to the hydrophobic pockets.

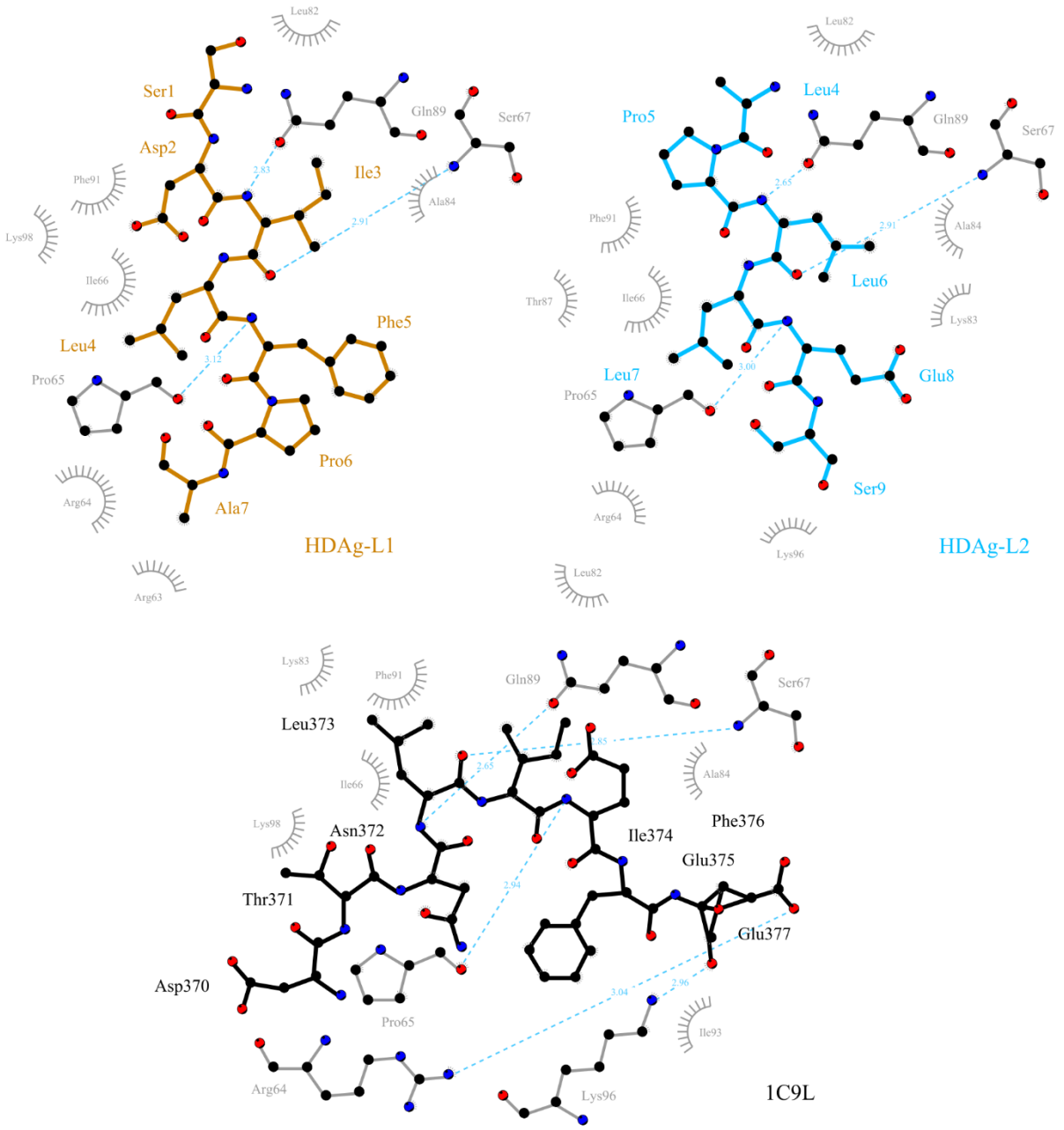


Figure 20 Schematic representation of interactions between viral peptide ligands and NTD side chains at the clathrin box. Viral peptides (HDAg-L1$_{pep}$, orange sticks and HDAg-L2$_{pep}$, blue sticks) are compared to a peptide of β-arrestin 2 (PDB ID: 1C9L, black sticks). The orientation of hydrophobic residues of NTD (grey spiky segments) and hydrogen bonds formed between NTD and the peptides (dashed blue lines) are displayed. Side chains of NTD residues binding to the ligands are shown as grey sticks. Representations were generated using ligplot+ (3).

At the arrestin box, no difference density showing the presence of peptide was observed with HDAg-L1$_{pep}$. For HDAg-L2$_{pep}$, four residues (LPLL) could be modeled into the electron density. The side chains of the consecutive leucine residues Leu6 and Leu7 bind to the hydrophobic cleft between blades 4 and 5 of the NTD β-propeller. The side chain oxygen atom of Gln192 of NTD forms a hydrogen bond with a backbone amide nitrogen of the peptide (Figure 21). Strikingly,

the mode of binding of HDAg-L2$_{pep}$ to the arrestin box is distinct from the previously published observed interaction of NTD with the surface loop of the arrestin 2 long isoform (arrestin2L) (1). Both arrestin2L and HDAg-L2 interact with NTD via key leucine residues. However, the arrestin2L loop is tightly bent and binds to the hydrophobic pocket via residues Leu335 and Leu338, whereas HDAg-L2$_{pep}$, the peptide binds in a linear fashion and the two interacting leucine residues are consecutive (see also section 3.5 and Figure 27).

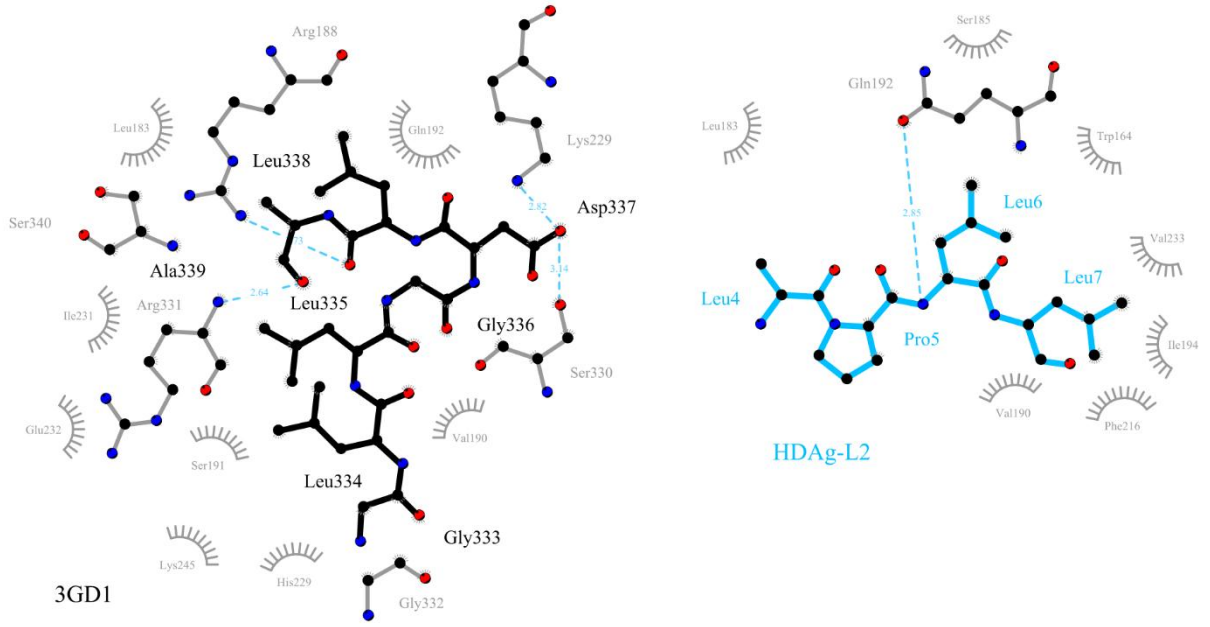


Figure 21 Schematic representation of interactions between HDAg-L2$_{pep}$ and NTD at the arrestin box. The viral peptide (blue sticks) is compared to residues 333-339 of arrestin2 (PDB ID: 3GD1, black sticks). The orientation of hydrophobic residues of NTD (grey spiky segments) and hydrogen bonds formed between NTD and the peptides (dashed blue lines) are displayed. Side chains of NTD residues binding to the ligands are shown as grey sticks.

Both viral peptides were found to bind to a hydrophobic patch between blades 6 and 7 of clathrin NTD, corresponding to the predicted fourth binding site on NTD (2), which we termed the Royle box (Figure 19). The difference density at this binding site was less well resolved than at the clathrin and arrestin box sites, and the peptides modeled into this density were less ordered than the surrounding residues. Therefore, the peptides were modeled in different orientations and conformations. However, these analyses showed clearly that only a single orientation could be found that fitted the electron density. Surprisingly, the HDAg-L1 and -L2 peptides were found to bind to the Royle box in opposite directions. Despite this, the defining feature of binding at the Royle box, three residues wrapping around Phe9 of NTD (HDAg-L1$_{pep}$: “LFP”, HDAg-L2$_{pep}$:

"RLP", see Figure 19 and Figure 25B), was conserved between the two structures (see also section 3.4).

3.3 Crystallization of NTD with cellular peptides

The structures of NTD in complex with HDAg-L1$_{pep}$ and HDAg-L2$_{pep}$ showed convincing binding of the viral CBMs to multiple sites on NTD. This observation was in stark contrast to the findings of published literature, which assumes a "one motif binds to one specific site on NTD" model (1, 4, 54, 62). Therefore, although crystal structures of cellular peptide motifs bound to clathrin NTD at the clathrin box, the arrestin box and the W box have been solved previously (1, 4, 62), binding of cellular CBMs to NTD was re-evaluated.

3.3.1 Co-crystallization of NTD with AP2CBM$_{pep}$ and Amph4T1$_{pep}$

The CBMs of the clathrin-interacting proteins amphiphysin (LLDLD) and of the hinge region of the AP2 β2 subunit (LLNLD) were selected for further analysis and peptides encompassing these motifs and flanking residues were designed. For AP2, the peptide CGDLLNLDLG (AP2CBM$_{pep}$) was already available in our laboratory. For amphiphysin, a peptide with sequence ETLLDLDFLE (Amph4T1$_{pep}$) was designed. This peptide not only contains the CBM of amphiphysin plus the flanking phenylalanine residue, but two additional C-terminal amino acids (LE) as well. These residues do not occur in to the sequence of human amphiphysin, but were included in the peptide to agree with a construct which was previously used for biochemical NTD interaction studies (308) (see also Chapter 4).

Co-crystals of NTD:HDAg-L1$_{pep}$ and NTD:HDAg-L2$_{pep}$ grew in very similar conditions using sodium malonate as a precipitate (section 3.2.1). Therefore, it seemed likely that NTD would also crystallize with other CBM peptides in sodium malonate-based conditions. Thus, co-crystallization experiments were attempted by manually pipetting reservoir and protein:peptide drops and equilibrating against 200 µL reservoirs. The concentration and pH of sodium malonate, as well as the drop ratio between protein:peptide mixture and the reservoir were varied. Plate-shaped crystals of similar morphology and thickness as observed when co-crystallizing NTD with HDAg-L1$_{pep}$ and HDAg-L2$_{pep}$ grew reproducibly under these conditions. The crystal used for solving the structure of NTD:AP2CBM$_{pep}$ was grown in a sitting drop containing 1 µL protein:peptide mixture (14 mg/mL NTD, 3.4 mM AP2CBM peptide) plus 2 µL of reservoir

(0.94 M sodium malonate pH 6.7). Unlike all other structures presented in this thesis, the space group of this data set was $C222_1$, and therefore, the structure had to be solved using non-isomorphous molecular replacement. For NTD:Amph4T1$_{pep}$, the crystal used for structure solution was grown in a sitting drop containing 1 µL of protein:peptide mixture (14 mg/mL NTD, 3.4 mM Amph4T1 peptide) plus 1 µL or reservoir solution (1.1 M sodium malonate pH 8.0). This structure was solved by isomorphous replacement. Both structures were refined to high resolution (1.8 Å for NTD:AP2CBM$_{pep}$, 1.7 Å for NTD:Amph4T1$_{pep}$). Data collection and refinement statistics are shown in Table 11.

3.3.2 Co-crystallization of NTD with AmphCBM$_{pep}$ and AmphCBM-long$_{pep}$

In order to evaluate binding of the CBM of amphiphysin to NTD without possible interference of the non-natural C-terminal "LE" residues of NTD:Amph4T1$_{pep}$ and to assess a potential stabilizing effect of residues C-terminal to the CBM, two further amphiphysin peptides (AmphCBM$_{pep}$: ETLLDLDFDP and AmphCBM-long$_{pep}$: ETLLDLDFDPFK) were crystallized with NTD. Crystallization experiments screening sodium malonate conditions were attempted as above and again, flat, thin plate-shaped crystals grew reproducibly. For AmphCBM$_{pep}$, the crystal used for structure solution was grown in a sitting drop containing 1 µL protein:peptide mixture (14 mg/mL NTD and 3.4 mM AmphCBM peptide) plus 2 µL reservoir solution (0.85 M sodium malonate pH 7.5) against a 200 µL reservoir. For NTD:AmphCBM-long$_{pep}$, an additive screen was performed to further improve the diffraction of the crystals. The crystal used for solution of the structure of NTD:AmphCBM-long$_{pep}$ was harvested from a 3 µL drop comprising 1 µL of 14 mg/mL NTD, 3.4 mM AmphCBM-long peptide plus 2 µL of 1.04 M sodium malonate pH 7.1, 0.2 M sodium perchlorate. Structures were solved by isomorphous replacement and refined to 1.9 Å (NTD:AmphCBM$_{pep}$) and 1.8 Å (NTD:AmphCBM-long$_{pep}$) (Table 11).

Table 11 Data collection and refinement statistics of structures of NTD in complex with cellular peptides. Values for the highest resolution shell are shown in parentheses. One crystal was used for each structure.

NTD:	**AP2CBM$_{pep}$**	**Amph4T1$_{pep}$**	**AmphCBM$_{pep}$**	**AmphCBM-long$_{pep}$**
Data collection				
Space group	$C222_1$	C2	C2	C2
Cell dimensions				
a, b, c (Å)	108.1, 133.2, 77.9	137.8, 131.0, 79.1	140.0, 134.1, 78.0	138.1, 131.2, 77.8
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 116.2, 90.0	90.0, 115.1, 90.0	90.0, 115.4, 90.0
Resolution (Å)	57.1-1.8 (1.81-1.76)	33.6-1.7 (1.74-1.70)	67.1-1.9 (1.93-1.88)	39.7-1.8 (1.89-1.84)
R_{merge}	0.053 (1.538)	0.055 (0.751)	0.149 (1.092)	0.084 (1.232)
$<I/\sigma I>$	15.8 (1.2)	13.1 (1.6)	6.5 (1.3)	12.8 (1.5)
$CC_{1/2}$	1.000 (0.672)	0.999 (0.564)	0.993 (0.568)	0.999 (0.528)
Completeness (%)	100.0 (100.0)	98.2 (94.5)	100.0 (100.0)	99.9 (99.5)
Redundancy	7.5 (7.0)	3.9 (3.4)	5.1 (4.4)	6.8 (6.2)
Refinement				
Resolution (Å)	57.1-1.8 (1.81-1.76)	33.6-1.7 (1.74-1.70)	67.1-1.9 (1.93-1.88)	39.7-1.8 (1.89-1.84)
No. of reflections (work/free)	52,951/2740	128,743/6564	99,868/5291	102,809/5312
R_{work}/R_{free}	0.176/0.205	0.158/0.185	0.204/0.234	0.162/0.182
Ramachandran favored regions (%)	98.7	98.8	98.6	98.4
Ramachandran outliers (%)	0.0	0.0	0.0	0.0
No. of atoms				
Protein	2836	5796	5845	5779
Glycerol	-	6	18	6
Peptide ligands	104	340	268	336
Water	403	1019	803	785
Average B-factors				
Protein	36.1	26.2	22.0	31.9
Glycerol	-	21.8	36.5	27.2
Peptide ligands	52.4	48.7	45.1	57.8
Water	53.5	42.3	37.8	46.5
r.m.s. deviations				
Bond lengths (Å)	0.016	0.014	0.010	0.012
Bond angles (°)	1.719	1.619	1.396	1.548
PDB ID	**5M5R**	**5M5T**	**5M5S**	**5M61**

3.3.3 Structures of NTD in complex with cellular peptides

Strikingly, as observed with the viral peptides (section 3.2.3), the structures of NTD in complex with cellular peptides revealed binding of peptides to three sites on NTD (Figure 22). All four peptides, AP2CBM$_{pep}$, Amph4T1$_{pep}$, AmphCBM$_{pep}$ and AmphCBM-long$_{pep}$, were found to bind to both the clathrin box and the arrestin box. For the Amph4T1 peptide, difference density

consistent with peptide binding was also observed at the Royle box. No binding at the W box was detected for any of the peptides.

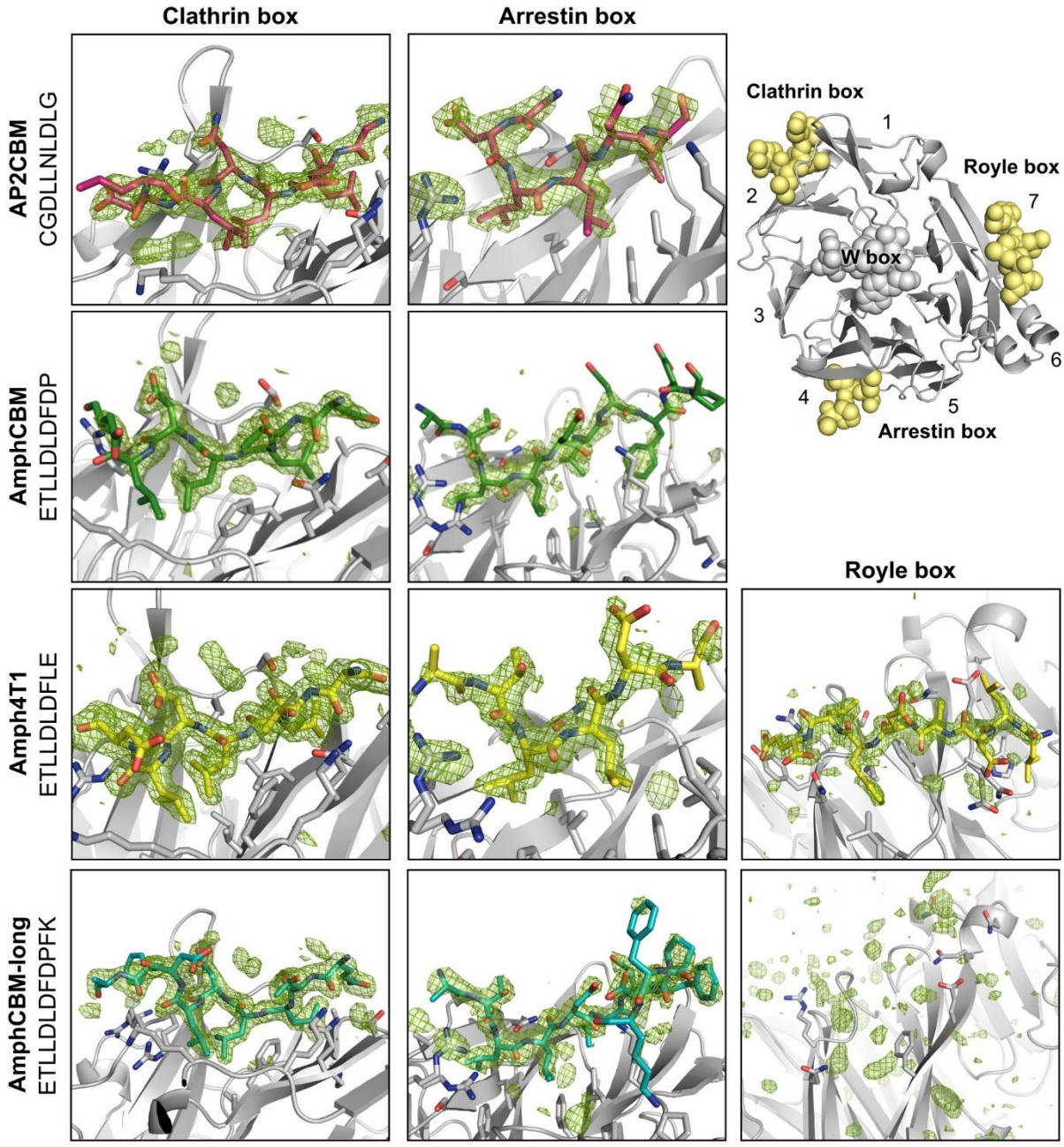


Figure 22 Structures of cellular peptides of AP2 and Amphiphysin bound to clathrin NTD. The unboxed image shows a cartoon of clathrin NTD (grey) with the four binding sites indicated by spheres. Boxed images illustrate the cellular peptides (AP2CBM$_{pep}$, CGDLLNLDLG, magenta sticks; AmphCBM$_{pep}$, ETLLDLDFDP, dark green sticks; Amph4T1$_{pep}$, ETLLDLDFLE, yellow sticks; AmphCBM-long$_{pep}$, ETLLDLDFDPFK, cyan sticks) bound at the clathrin box, arrestin box and Royle box on clathrin NTD (grey ribbons, selected side chain atoms shown as sticks). Unbiased F_O-F_C difference electron density (green mesh, shown at 3 σ) was calculated before the peptides were added to the respective models and used to place the peptides.

At the clathrin box, binding of all cellular peptides was very similar to the binding mode observed for HDAg-L1$_{pep}$, HDAg-L2$_{pep}$ and β-arrestin 2 (62). The residues LLDLD of the amphiphysin peptides, and residues LLNLD of the AP2 peptide, each corresponding to the consensus CBM motif LΦXΦ[D/E], bind to the hydrophobic groove between blades 1 and 2 of the clathrin NTD β-propeller. The two N-terminal leucine residues protrude into the hydrophobic pockets formed by residues Thr87/Leu82/Ala84 and Val50/Arg64/Ile66/Ile80/Leu82/Phe91 (described in section 3.2.3), respectively. The third leucine of the CBM motif binds between Phe91 and the hydrophobic face of the sidechain of Lys96.

The formation of hydrogen bonds between residues of NTD and the bound peptides was also conserved between viral and cellular peptides (Figure 20 and Figure 23). In all structures, the backbone amide nitrogen of Ser67 and the backbone carbonyl oxygen of Pro65 bind to backbone oxygen and nitrogen atoms of the peptides, respectively. In the NTD:AP2CBM$_{pep}$ and NTD:AmphCBM$_{pep}$ structures, additional hydrogen interactions are formed between the backbones of the peptides and side chain nitrogen atoms of the NTD residue Arg64. For all peptides, side chain oxygen and nitrogen atoms of NTD Gln89 form hydrogen bonds to backbone nitrogen and oxygen atoms of the peptide. These hydrogen bonds are formed with backbone atoms of the peptide that lie N-terminal to the consensus binding motif residues (LLDLD/LLNLD). Since these N-terminal residues differ between the used amphiphysin and the AP2 peptide (“ET” and “GD”, respectively), it seems likely that residues adjacent to the CBM contribute to binding by stabilizing the interaction in a sequence-independent manner.

Interestingly, although the final residue of the CBM motif (“D”) could be modeled for all cellular peptides, it points away from NTD and does not form any hydrogen bonds via its side chain, indicating that the nature of the amino acid ad this position of the CBM might not be relevant for binding. The steric positioning of this final residue of the CBM motif together with previously published biochemical data (76) suggest that although this amino acid is conserved across CBMs of different clathrin-binding proteins (309), it might not be an integral part of the consensus motif.

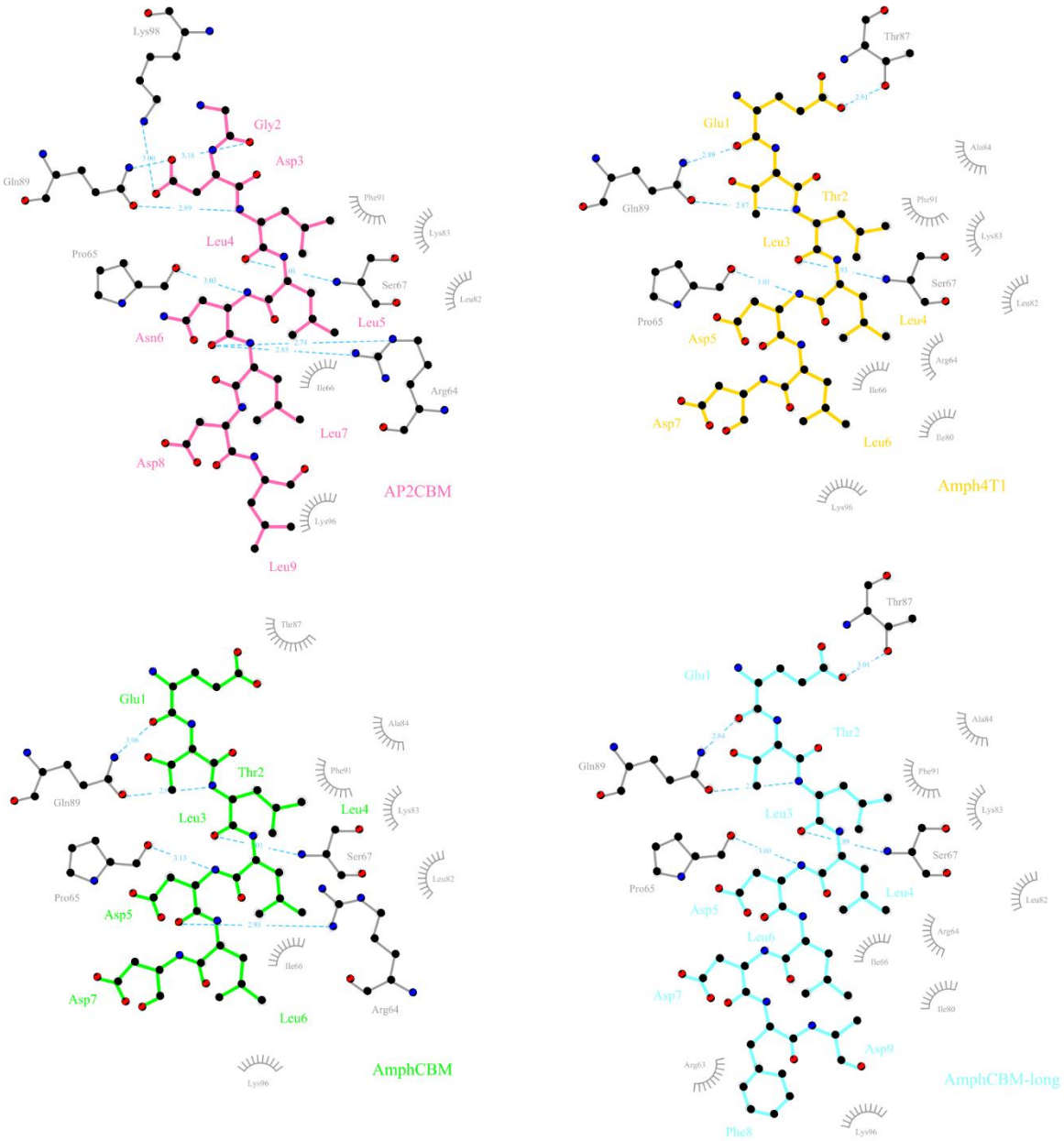


Figure 23 Schematic representation of interactions between NTD and cellular peptide ligands derived from amphiphysin and the β2 subunit of AP2 at the clathrin box. Cellular peptides are shown as colored sticks ($AP2CBM_{pep}$, pink, $Amph4T1_{pep}$, yellow, $AmphCBM_{pep}$, green and $AmphCBM\text{-}long_{pep}$, cyan). The orientation of hydrophobic residues of NTD (grey spiky segments), formed hydrogen bonds (dashed blue lines), as well as side chains of NTD residues interacting with the ligands (grey sticks) are displayed.

At the arrestin box, all four peptides bind in the same orientation to a large hydrophobic patch between blades 4 and 5 of NTD (Figure 22, see also section 3.5). Similar to their behavior at the clathrin box, the three leucine residues of the CBM peptides (LLDLD for the amphiphysin peptides, LLNLD for the hinge region of the β2 subunit of the AP2 peptide) are the key amino acids mediating the interaction. The N-terminal leucine side chain projects into a pocket lined

by residues Val90, Trp164 and Leu183 of NTD, the adjacent second leucine of the CBM binds close to residues Leu183, Phe216, Ile231 and Val233, and the third leucine residue is framed by residues Ile194 and Phe218 of NTD. Extensive hydrogen bonds are formed between the peptide backbones and NTD (Figure 24). In all four structures, the side chain oxygen of NTD residue Gln192 forms two hydrogen bonds to backbone amide nitrogen atoms of the bound peptides. For AmphCBM-long$_{pep}$, the nitrogen side chain atom of Gln192 forms an additional bond with the peptide backbone. Other hydrogen bonds can be observed between the side chain nitrogen atoms of Arg188 or Lys245 and the AP2CBM$_{pep}$ backbone, and between one side chain nitrogen atom of Lys245 and two backbone carbonyl oxygens of AmphCBM$_{pep}$ and AmphCBM-long$_{pep}$. The last residue of the clathrin box motif in all peptides ("D") is resolved in the three co-structures of NTD with amphiphysin peptides, but does not form any interactions with NTD via its side chains.

It should be noted that all twelve residues of the extended amphiphysin peptide used for crystallization, AmphCBM-long$_{pep}$ (ETLLDLDFDPFK), could be modeled at the arrestin site (Figure 22, bottom central panel). The C terminus of this peptide contains two phenylalanine residues and one proline, and these residues form a hydrophobic intra-peptide stack in the crystal structure. These stacked peptide residues are located approximately 10 Å away from the arrestin box and form hydrogen bonds and hydrophobic interactions with the second NTD molecule in the asymmetric unit (related by non-crystallographic symmetry). The relative orientation of the two NTD molecules is not consistent with their presumed binding to AP2-coated membranes in vivo, where the NTD discs lie parallel to the place of the membrane. Therefore, I assumed that the observed binding of the C terminus of the AmphCBM-long peptide to NTD is a crystallographic artifact and it was thus not further investigated.

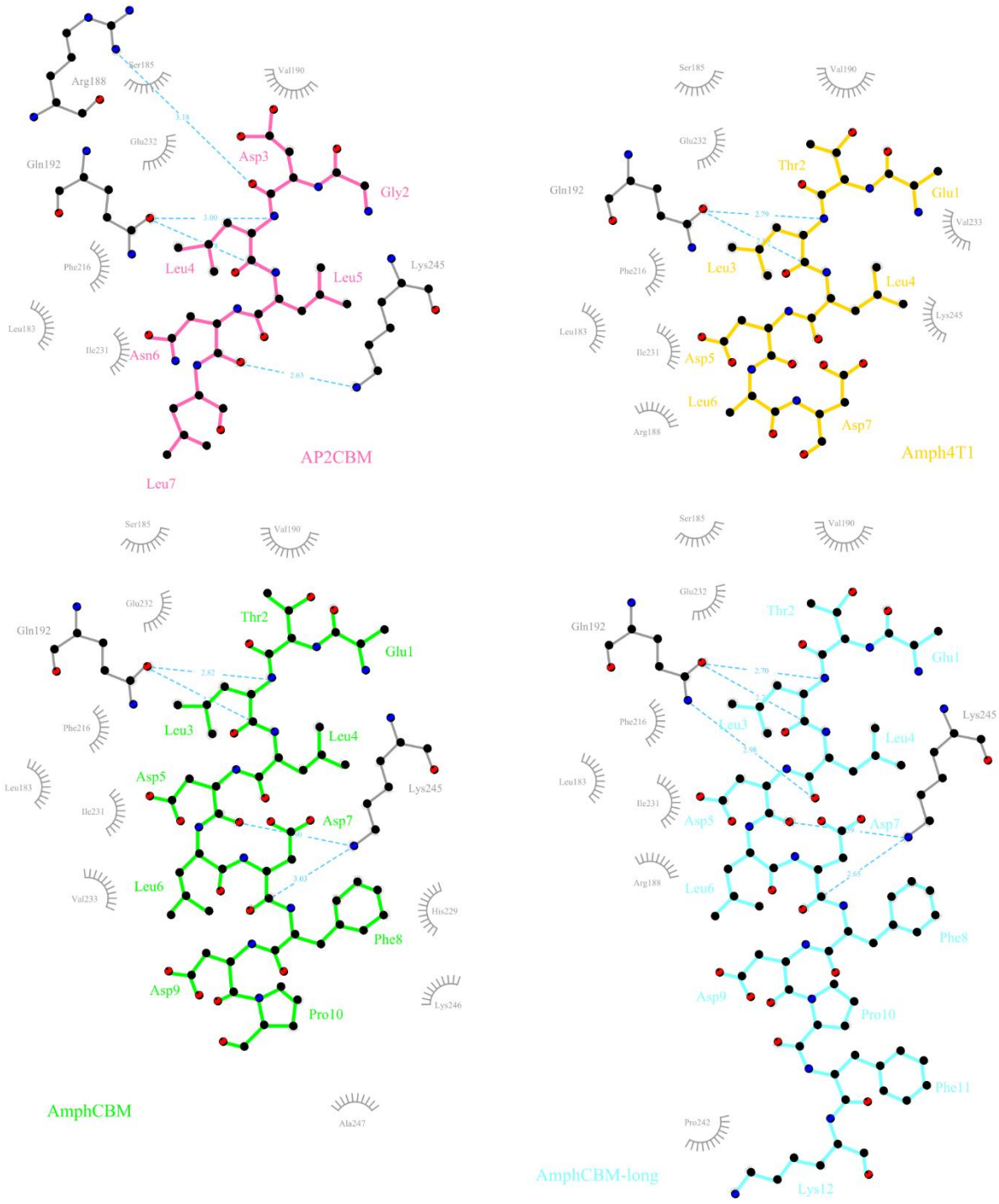


Figure 24 Schematic representation of interactions between NTD and cellular peptide ligands derived from amphiphysin and the β2 subunit of AP2 at the arrestin box. Peptides are shown as sticks ($AP2CBM_{pep}$, pink, $Amph4T1_{pep}$, yellow, $AmphCBM_{pep}$, green and $AmphCBM\text{-}long_{pep}$, cyan). The orientation of hydrophobic residues of NTD (grey spiky segments), formed hydrogen bonds (dashed blue lines), as well as side chains of NTD residues interacting with the ligands (grey sticks) are displayed.

Like $HDAg\text{-}L1_{pep}$ and $HDAg\text{-}L2_{pep}$, the Amph4T1 peptide (ETLLDLDFLE) was also found to bind to the NTD Royle box (Figure 22). As observed with the HDV-derived peptides, binding centered around the NTD residue Phe9 and was predominantly mediated by hydrophobic

interactions. However, residues "LE" were involved in binding, forming hydrophobic interactions with residues 297-299 of NTD. These C-terminal peptide residues do not occur in amphiphysin naturally. No binding of the native amphiphysin peptide AmphCBM$_{pep}$, which has the same length as Amph4T1$_{pep}$ (ten amino acids), to the Royle box was detected. This observation presented the motivation for crystallizing NTD in complex with the longer amphiphysin peptide (AmphCBM-long$_{pep}$). The long amphiphysin peptide contains two additional C-terminal residues that could potentially stabilize the interaction at the Royle box. However, as for AmphCBM$_{pep}$, no electron density indicating binding of AmphCBM-long$_{pep}$ to the Royle box was observed.

3.4 Characterization of the Royle box

In total, two viral and one amphiphysin-derived peptide were found to bind the fourth binding site on NTD, the Royle box. In agreement with the prediction by Willox and Royle (2), all three peptides bind to a highly conserved surface on NTD (Figure 25A). Although HDAg-L1$_{pep}$ binds in the opposite direction to HDAg-L2$_{pep}$ and Amph4T1$_{pep}$ (Figure 25B), important binding features are common in the three structures. All peptides bind to a groove between blades 6 and 7 of the NTD β-propeller. Residue Phe9 of NTD seems pivotal for the interaction since the peptides wrap around this residue and form extensive hydrophobic interactions with both faces of its benzyl group. A deep hydrophobic pocket formed by residues Leu5, Ile7, Phe9, Ile282, Asn296 and Val327 is occupied by a hydrophobic residue of each peptide, most prominent for Amph4T1$_{pep}$, where the phenylalanine residue fills the whole pocket and stacks against Phe9 of NTD. Further hydrophobic interactions are formed by leucine (Amph4T1$_{pep}$ and HDAg-L2$_{pep}$) or isoleucine (HDAg-L1$_{pep}$) residues of the peptides and a patch adjacent to the backbone of Arg297 of NTD (Figure 25B,C). Interestingly, these interactions seem to be at least partially mediated by residues outside the consensus CBM sequence. In all three structures, the backbone carbonyl oxygen of Arg297 forms a hydrogen bond with backbone amide nitrogen atoms of the peptides. Other common backbone:backbone hydrogen bonds are found between the peptides and NTD residues Phe9 and Glu11 in the NTD:Amph4T1$_{pep}$ and NTD:HDAg-L1$_{pep}$ structures (Figure 26). Taken together, these results clearly demonstrate the localization of the Royle box on NTD and highlight the importance of hydrophobic residues for interactions at this site.

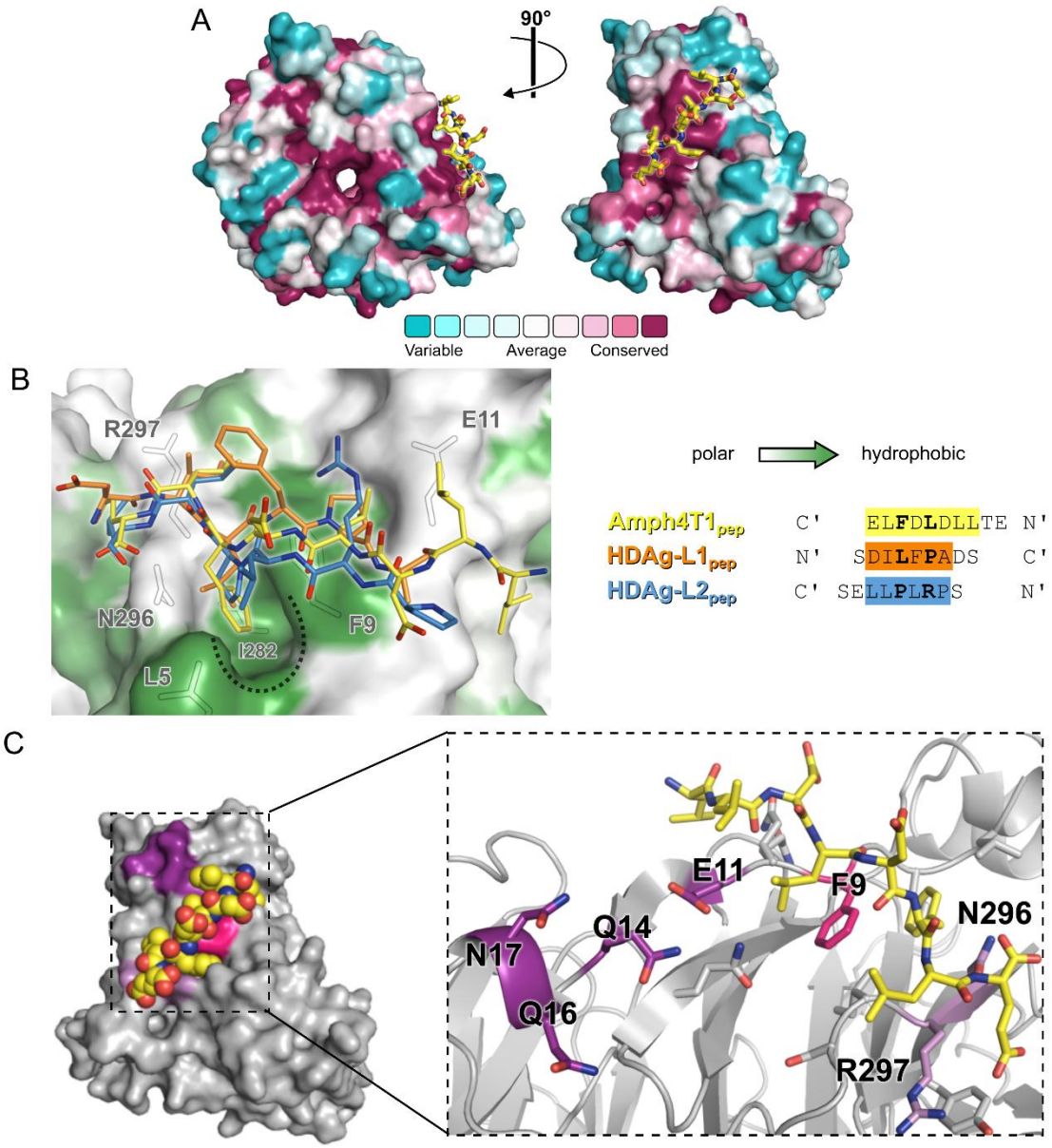


Figure 25 Structural characterization of the Royle box. (A) Surface model of clathrin NTD with amino acids colored by conservation (cyan: most variable, magenta: most conserved). The Amph4T1 peptide, binding to NTD at the conserved Royle box, is shown as yellow sticks. The model is oriented as in Figure 19 and Figure 22 (left) and rotated by 90° (right). (B) Close-up view of three different peptides, Amph4T1$_{pep}$ (yellow sticks), HDAg-L1$_{pep}$ (orange sticks) and HDAg-L2$_{pep}$ (blue sticks), binding to the Royle box on NTD. The surface of NTD is shown with residues colored according to their hydrophobicity (low: white, high: green). Outlines of selected NTD residues are depicted in grey. A strongly hydrophobic pocket occupied by hydrophobic residues of all three peptides is marked with a dotted black line. The peptide sequences used for co-crystallization are structurally aligned next to the panel. Residues that could be confidently modeled into the electron density are color-highlighted, and residues projecting into surface pockets are printed in bold. The directionality of the bound peptides is indicated. (C) Surface model of NTD (grey) with residues that are implicated in binding to the Royle box (2) highlighted in light or dark purple. Residue Phe9, which forms extensive hydrophobic interactions with the bound peptides, is highlighted in pink. The Amph4T1 peptide binding at the Royle box is shown as yellow spheres. The zoomed inset shows details of Amph4T1$_{pep}$ (yellow sticks) at the binding site. Residues that disrupt transferrin uptake when mutated are shown in dark purple, whereas residues which do not impact transferrin update are shown in light purple (2).

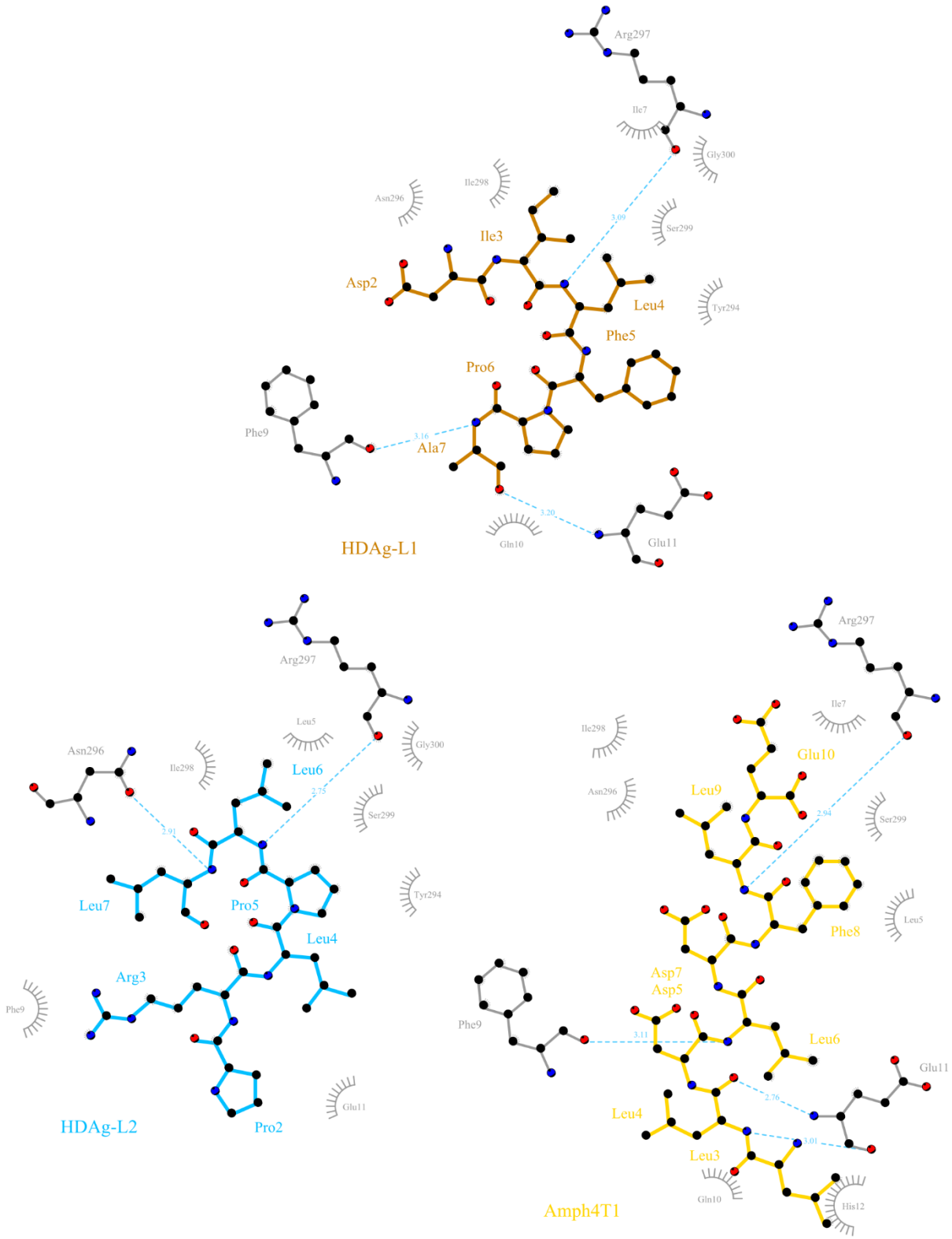


Figure 26 Schematic representation of interactions between peptide ligands and NTD at the Royle box. Two viral peptides (HDAg-L1$_{pep}$, orange sticks and HDAg-L2$_{pep}$, blue sticks) and the non-native Amph4T1 peptide (yellow sticks) bind to NTD at this interaction site. The orientation of hydrophobic residues of NTD (grey spiky segments), hydrogen bonds formed between the peptides and NTD (dashed blue lines) and residues of NTD (grey sticks) interacting with the peptides are shown.

3.5 Identification of a novel binding mode at the arrestin box

The viral HDAg-L2 peptide and all four cellular peptides bound to the arrestin box on NTD (sections 3.2.3 and 3.3.3). Located between blades 4 and 5 of the β-propeller (Figure 27A), this binding site was identified by Kang et al. as the second interaction site of the long isoform of arrestin2 (1). In the NTD:arrestin2L complex, residues 332-340 of arrestin2L form a tight loop that binds to the NTD surface groove (Figure 27C). In contrast, the peptides investigated in this thesis adapted a linear conformation at the arrestin box, similar to the peptide conformation observed at the clathrin box (Figure 22 and Figure 27B). In addition, the backbones of the linear peptides bind in the opposite direction to that observed in the NTD:arrestin2L structure.

Despite these differences, the key interaction feature, leucine residues binding to hydrophobic pockets, is conserved across all structures. For $AP2CBM_{pep}$ and the amphiphysin-derived peptides, the two adjacent leucine residues corresponding to the first two residues of the CBM (e.g. Leu4 and Leu5 of the AP2CBM peptide used for crystallization, Figure 27B), form hydrophobic interactions at the arrestin site, whereas in the structure of arrestin2L bound to NTD, residues Leu335 and Leu338 occupy the same hydrophobic pocket (Figure 27C). To accommodate the additional two amino acids between these two leucine residues, the arrestin2L loop bulges away from the NTD surface. A third leucine of $AP2CBM_{pep}$, $AmphCBM_{pep}$ and $AmphCBM\text{-}long_{pep}$, corresponding to the third Leu of the CBM (e.g. Leu7 of $AP2CBM_{pep}$), interacts with a highly hydrophobic patch, extending the binding interface. Residue Leu334 of arrestin2L binds to this hydrophobic patch as well, but cannot project into the pocket as deeply due to the conformation of the peptide backbone. Interestingly, although only four residues of $HDAg\text{-}L2_{pep}$ could be modeled at the arrestin box, the two consecutive leucine residues were resolved and bound to the same hydrophobic pocket as the first two leucine residues of the cellular CBM peptides.

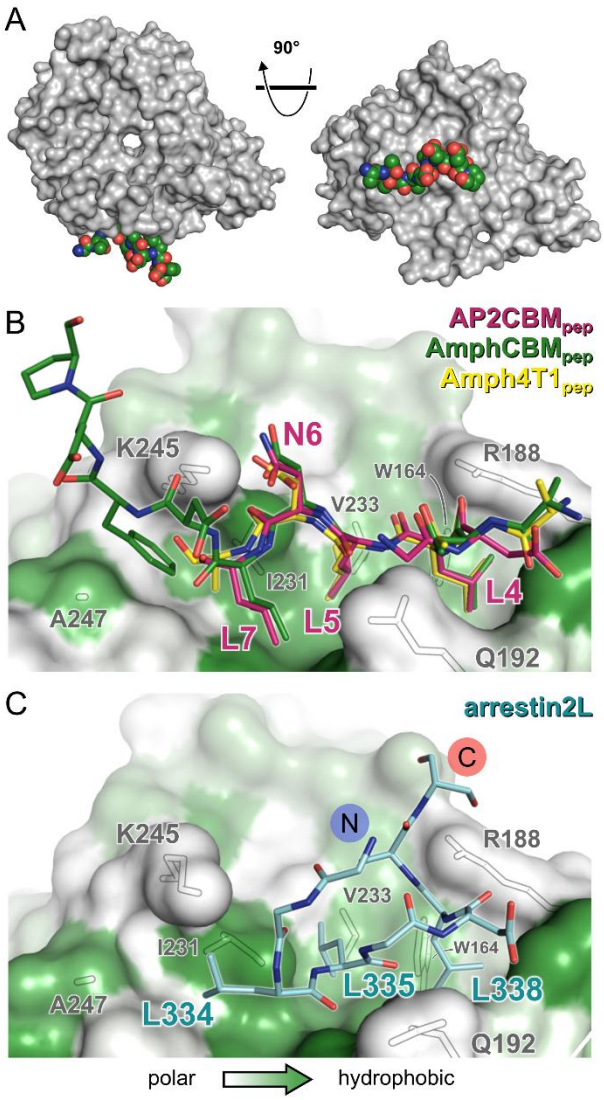


Figure 27 Comparison of binding of linear peptides and the arrestin2L loop at the arrestin box. (A) Surface model of clathrin NTD (grey) with the AmphCBM peptide (green spheres) bound at the arrestin box. The model is oriented as in Figure 19 and Figure 22 (left) and rotated by 90° (right). (B) Detailed view of the cellular peptides (shown as colored sticks) binding at the arrestin box. The surface on NTD is shown with residues colored by their hydrophobicity (low: white, high: green). Outlines of selected NTD residues are depicted in grey. Residues forming the CBM motif of the AP2CBM peptide (magenta sticks) are numbered according to their position in the peptide used for crystallization. (C) Close-up view of the surface loop of arrestin2L bound at the arrestin site (PDB 3GD1, (1)). The surface of NTD is illustrated as in (B) and residues 332-340 of arrestin2L are shown as light blue sticks. Leucine residues of the loop interacting with hydrophobic groove are numbered as in the published structure. N and C terminus of the bound peptide are indicated.

3.6 Discussion

3.6.1 Peptides containing a CBM bind to multiple sites on NTD simultaneously

Clathrin NTD was crystallized in complex with two viral and four cellular peptides, each encompassing the canonical clathrin binding motif LΦXΦ[D/E]. Surprisingly, binding of these

peptides to more than one site on NTD was observed in all structures. The viral peptides, derived from the C terminus of the large antigen of different hepatitis D virus genotypes, and the cellular peptides containing CBMs of the clathrin-interacting proteins β2 adaptin and amphiphysin, all bound to the clathrin box of NTD. All peptides except HDAg-L1$_{pep}$ were also found to interact with a second site on NTD, the arrestin box. No peptides were observed at the W box, but both viral peptides and the Amph4T1 peptide bound to a region on NTD corresponding to the fourth binding site identified by Willox and Royle in functional assays (2). In this chapter, this fourth interaction site was structurally characterized and termed the "Royle box".

The consensus across the literature is that each binding site on NTD can only interact with peptides containing a specific motif. At the clathrin box, the well-studied motif LΦXΦ[D/E] mediates binding (62, 78), for the arrestin box, [LI][LI]GXL has been suggested (1), and at the W box, residues PWXXW are proposed to facilitate the interaction (4, 79, 80). All these interactions have not only been assessed biochemically, but also by crystallographic studies. To date, published structures of NTD in complex with peptides spanning the CBM of β-arrestin 2 or of the hinge of β3-adaptin demonstrate binding of these peptides only at the clathrin box (62), the splice loop of arrestin2L was observed only at the arrestin box (with the clathrin box being already occupied by the canonical CBM of arrestin2L) (1), and an amphiphysin peptide containing residues PWDLW bound solely at the W box (4). However, HDAg-L2$_{pep}$ and the cellular peptides used for co-crystallization in this study bound to both the clathrin box and the arrestin box via key leucine residues of the CBM, raising the question why such binding was not observed by ter Haar et al. (62). The incongruity between the previous work and the results presented in this chapter can probably be attributed to differences during crystallization and cryoprotection. Ter Haar and colleagues used a 2- to 4-fold molar excess of peptide during crystallization, whereas for all the structures presented in this chapter, 7- to 10-fold molar excess of peptide was used. In addition, the solution used for cryoprotection of the crystals used in this study contained a high concentration (2 mM) of the respective peptides to prevent backsoaking of weakly bound peptides. In contrast, ter Haar et al. did not report using any such measures and followed a cryoprotection protocol including long soaks during which peptides could easily dissociate from NTD.

Recently, Zhuo et al. studied the interaction of NTD with long peptides containing CBMs derived from AP2 and AP180 by nuclear magnetic resonance and isothermal titration calorimetry (266). In agreement with the results presented in this chapter, they observed promiscuous, low affinity binding of the peptides to the three sites on NTD, namely the clathrin box, arrestin box and W box. One of their peptides encompassed residues 619–644 of AP2 β2 subunit (CQVIPSQGDLLGDLLNLDLGPPVNVPQ, CBM highlighted) and is therefore similar to the AP2CBM peptide used in this study (CGDLLNLDLG). However, whereas the crystallographic data described in this chapter demonstrate binding of AP2CBM$_{pep}$ to the clathrin box and the arrestin box, Zhuo and colleagues report additional binding of their peptide at the W box. The disparity between the binding preferences between those two peptides could potentially be attributed to non-canonical W box binding motifs present in the much longer peptide used by Zhuo et al. in their biophysical experiments.

3.6.2 Viral peptides bind to the same sites on NTD as cellular peptides

Interestingly, the comparison of the structures of NTD:HDAg-L1$_{pep}$ and NTD:HDAg-L2$_{pep}$ with the structures of NTD in complex with cellular peptides revealed that the HDV-derived peptides bind to the same sites on NTD as the cellular ones, specifically the clathrin box, the Royle box, and, for HDAg-L2$_{pep}$, also the arrestin box. It has been demonstrated previously that mutation of the HDAg-L1 CBM residues Leu199 or Asp203 to alanine reduced the production of virus-like particles and association with clathrin in co-immunoprecipitation experiments, implicating the interaction between HDAg-L1 and clathrin in viral egress (106). In agreement with these observations, the crystal structures presented in this chapter show that residue Leu199 of HDAg-L1 and the corresponding leucine residue of HDAg-L2 are pivotal for binding at the clathrin box and, for HDAg-L2, the arrestin box as well. On the other hand, residue Asp203 is consistently disordered and is thus probably not directly involved in binding between HDAg-L1 and clathrin. Therefore, while loss of virus-like particle production by mutating Leu199 could be due to loss of NTD binding, the effect of mutating Asp203 is probably not due to loss of NTD binding, but mediated by secondary effects.

The data presented in this chapter suggest that the interaction between HDV and clathrin is not a suitable target for drug design since the same binding sites on NTD are occupied by viral and cellular proteins. Targeting these binding sites could therefore interfere with physiological

clathrin trafficking. Based on their ITC results, Zhuo et al. suggested that peptides bind to all sites on NTD with similar affinities (266). However, comparison of the structures presented in this chapter with the structures published by ter Haar et al. (62) suggest that peptide binding to one or more sites on clathrin might depend on the molar excess of peptide present during crystallization and/or on possible backsoaking of bound peptide into cryoprotectant solutions, arguing against the hypothesis that all interaction sites on NTD bind peptides with similar affinities. To explore the contribution of the separate binding sites to the overall interaction between NTD and viral or cellular CBM-containing peptides, robust biochemical or biophysical data are required. Such experiments have been conducted and are described in Chapter 4.

3.6.3 Binding determinants at the clathrin box are highly conserved

Alignments of predicted or confirmed clathrin box-binding sequences strongly suggest LΦXΦ[D/E] as the clathrin box consensus binding motif (62, 75, 78). The importance of these residues for clathrin binding has also been demonstrated in GST pull-down and clathrin cage binding assays (75, 76). In all six structures presented in this chapter, binding of peptides at the clathrin box was observed. For the cellular peptides containing residues LLDLD (amphiphysin peptides) or LLNLD (AP2CBM$_{pep}$), conforming to the canonical CBM LΦXΦ[D/E], the first, second, and partially, also the third leucine residues formed hydrophobic interactions at the clathrin box. The two viral peptides also bind to the clathrin box via two consecutive hydrophobic residues ("IL" for HDAg-L1$_{pep}$ and "LL" for HDAg-L2$_{pep}$), but these residues "are out of register" compared to the CBM (Figure 28). In addition, both viral peptides do not seem to require a third hydrophobic residue such as the third leucine of the cellular peptides for binding. The interaction between peptides and NTD is stabilized by hydrogen bonds formed between the backbone of the bound peptide and residues Pro65, Ser67 and Gln89 of NTD, determining the backbone conformation of the bound peptide.

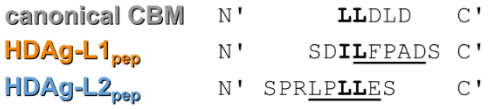


Figure 28 Sequence alignment of a canonical clathrin-binding motif and viral peptides based on the obtained structural data. The consecutive hydrophobic residues binding to the clathrin box are printed in bold. The originally proposed CBMs of HDAg-L1 and HDAg-L2 are underlined. Binding of these peptides to the clathrin box is "out of register" compared to the canonical CBM.

Although the structures confirm the contribution of the first, second and fourth residue of the CBM (LΦXΦ[D/E]) to binding at the clathrin box, it was noted that the final acidic residue (LΦXΦ[D/E]) did not form any direct interactions with NTD. This residue is conserved across most cellular proteins predicted or shown to bind to NTD via the clathrin box (309) and upon mutation of this residue in arrestin3, binding of clathrin to arrestin3 in vitro is moderately (30%) disturbed (76). Given the data presented here, the functional relevance of the final acidic residue for clathrin:adaptor interactions may need to be reassessed.

Taken together, these results show that the key determinants of binding at the clathrin box, namely interactions of leucine or isoleucine residues of the bound peptide with the hydrophobic pocket between blades 1 and 2 of the β-propeller of NTD, as well as several hydrogen bonds and the resulting overall backbone conformation, are conserved between viral and cellular peptides. However, it seems that the stipulated consensus binding motif is less stringent than previously thought.

3.6.4 The Royle box is a conserved, hydrophobic interaction site on NTD

The structures of NTD in complex with HDAg-L1$_{pep}$, HDAg-L2$_{pep}$ and Amph4T1$_{pep}$ demonstrate the location of the Royle box at a conserved surface patch between blades 6 and 7 of clathrin NTD, as predicted by Willox and Royle (2). Similar to binding at the arrestin and clathrin boxes, the interactions between the peptides and NTD are mainly hydrophobic. In particular, the deep hydrophobic pocket next to residue Phe9 of NTD seems to be crucial for binding of the peptides. However, unlike at the arrestin and clathrin boxes, residues outside the canonical CBM mediate binding and several natural cellular peptides do not bind NTD at this site. Assessment of the only non-viral peptide bound at the Royle box, Amph4T1$_{pep}$, is complicated by two residues at its C terminus that are not native to the amphiphysin sequence, but nevertheless participate in the interaction. Bioinformatic screening performed to identify other cellular proteins that could potentially bind to the Royle box was unsuccessful due to the degeneracy and varying N-to-C-terminal orientation of the peptide sequences found at the Royle box in the structures presented here.

Willox and Royle observed significantly reduced uptake of transferrin receptor in functional assays when the clathrin box, the arrestin box, the W box, and residue Glu11 of the Royle box were mutated (2). Since the side chain of Glu11 does not bind directly to the peptides, it is not

clear how mutation of this residue would influence interactions at the Royle box. Interestingly, residues Glu14, Glu16 and Asn17, which are also reported to disrupt endocytosis of transferrin receptor when mutated together with the clathrin box, the arrestin box and the W box (2), likewise do not form direct contacts with the peptides bound at the Royle box (Figure 25C). It is possible that in vivo, extended peptides bind at the Royle box and therefore interact with these residues. Alternatively, mutation of these residues could lead to destabilization and misfolding of NTD, thus inhibiting endocytosis. Identification of a native cellular Royle box-binding peptide and/or biochemical characterization of the mutants E11K and Q14D+Q16M+N17S could help to elucidate the role of these residues. Biochemical and biophysical properties of these mutants were thus investigated (see Chapter 4).

On the contrary, mutation of residues Asn296 and Arg297 to alanine and glutamate, respectively, did not impact transferrin receptor uptake (2), even though both residues are part of the Royle box interaction site (Figure 25C). However, it should be noted that binding between Arg297 and the peptides is not mediated by the arginine side chain, but by backbone:backbone hydrogen bonds. Therefore, mutation of Arg297 to glutamate, which would not disturb the formation of these hydrogen bonds, would potentially not influence binding of peptides to the Royle box. Similarly, mutation of residue Asn296, located at the edge of the hydrophobic binding pocket, to a small hydrophobic residue like alanine would probably not destroy and may even enhance the binding pocket. Therefore, the seemingly contradictory findings can possibly be attributed to the specific mutations chosen by Willox and Royle (2).

In summary, further studies are needed to determine a consensus binding motif for the Royle box and to find a bona fide cellular peptide interacting with NTD at the Royle box. In addition, biochemical or biophysical experiments confirming the NTD residues highlighted as important for binding based on the structural data have to be performed to substantiate the relevance of described Royle box described here for clathrin:adaptor interactions (see Chapter 4).

3.6.5 The arrestin box can bind to both linear and non-linear peptides

In 2009, a structure of NTD in complex with arrestin2 showed that its CBM interacts with the clathrin box and that the eight amino acid long loop present only in one long isoform (arrestin2L) of this alternatively spliced protein binds to a second site on clathrin, termed arrestin box, via the motif [LI][LI]GXL (1). The crystal structures presented in this chapter demonstrate that

peptides containing the canonical clathrin-binding motif LΦXΦ[D/E] can also bind to the arrestin site, but in an extended, linear conformation. Leucine residues of both the arrestin2 splice loop and the peptides used in this study occupy largely similar positions within the hydrophobic cleft between blades 4 and 5 of the NTD β-propeller, but the direction of the backbone is reversed between arrestin2L and the linear peptides.

It therefore seems likely that interactions at the arrestin box are dependent on the context in which a peptide is presented to NTD. If a constrained loop such as the splice loop of arrestin2 is present, non-consecutive leucine residues can form interactions at the arrestin box. Yet when elongated protein chains containing a CBM as for example found in AP1, AP2, AP3, amphiphysin, epsin 1 and 2 and many other proteins (54, 78) bind to clathrin, the linear CBM can also bind to the arrestin box. The [LI][LI]GXL published consensus seems unique to the case of the arrestin2L loop, and linear CBM binding at this site may represent the more general mode of peptide interactions at the arrestin box.

3.6.6 Summary

In this chapter, I presented crystal structures of clathrin N-terminal domain in complex with two peptides originating from the large antigen of different genotypes of hepatitis D virus, and with four cellular peptides derived from amphiphysin and AP2. All six peptides bound to multiple sites on clathrin promiscuously. The binding modes of viral and cellular peptides at the different interaction sites were largely conserved, making clathrin trafficking a difficult target for the design of compounds interfering with the HDV life cycle. Interactions at the clathrin box and the arrestin box were found to be mediated by key leucine or isoleucine residues forming part of the canonical clathrin-binding motif. In addition, both viral peptides and one of the amphiphysin peptides also bound to a third site, which corresponds to the "fourth and final binding site" of clathrin NTD delineated by Willox and Royle (2). Analysis of these structures enabled me to conducted the molecular characterization of the "Royle box". In addition, I described a novel binding mode for linear peptides at the arrestin box. In order to confirm the contribution of the different binding sites to overall clathrin:adaptor interactions in a non-crystallographic context, I next attempted to assess the binding between viral or cellular peptides and NTD by biochemical and biophysical techniques (Chapter 4).

4 BIOCHEMICAL AND BIOPHYSICAL PORTRAIT OF INTERACTIONS AT THE N-TERMINAL DOMAIN OF CLATHRIN

In the previous chapter, I demonstrated that viral and cellular peptides can interact with multiple sites on clathrin NTD promiscuously. The binding sites on NTD are known to be functionally redundant (2), making similar redundancy on the part of the adaptor proteins plausible. In order to test whether the interactions between NTD and peptides can be validated in a non-crystallographic context, and to disentangle the contribution of binding to separate sites to the overall interaction, I developed biochemical and biophysical methods to monitor clathrin:adaptor interactions.

In this chapter, I first describe how I established a GST pull-down assay to monitor binding between NTD and GST fusion proteins of NTD-binding motifs (section 4.1). Since interactions between clathrin and its adaptors are generally very weak, I used a construct of NTD fused to the NEMO oligomerization domain (His-NTD-NEMO) to provide avidity of binding (section 4.1.2). In these pull-down experiments, I show that peptides derived from the hepatitis D virus large antigen only bind clathrin NTD very weakly compared to cellular proteins (section 4.1.4). Having established a reliable pull-down protocol, I next designed and characterized an extensive panel of His-NTD-NEMO mutants carrying amino acid substitutions at single binding sites of NTD and combinations of these binding sites (section 4.2). These mutants were then used in GST pull-down assays to determine whether all interactions observed in the crystal structures contribute to binding (section 4.3). I also investigated whether cellular and viral peptide motifs presented on liposomes can bind purified clathrin and facilitate the formation of clathrin buds on liposomes (section 4.4). Finally, I explored multiple biophysical techniques to quantitate the binding between peptides and NTD, but these attempts proved unsuccessful, mostly due to the low affinities of the interactions (section 4.5).

Note: This chapter and the previous one are based on and expand the paper “Cellular and viral peptides bind multiple sites on the N-terminal domain of clathrin” published in 2017 (272).

4.1 Development of a biochemical assay to measure binding between NTD and binding motifs

4.1.1 Pull-down of clathrin from pig brain cytosol

Three viral peptides derived from the C-terminal tail of the large antigen of different genotypes of hepatitis D virus (HDAg-L1/2/3), one peptide encompassing the canonical clathrin-binding motif (CBM) of amphiphysin (Amph4T1), and the hinge-appendage domain of the β2 subunit of AP2 (AP2-HA) (Figure 29) were expressed overnight as glutathione s-transferase (GST) fusion proteins in BL21(DE3) pLysS *E. coli*. Amphiphysin and AP2-HA were used as positive controls because they contain canonical CBMs and are known to interact with NTD (61, 68, 79, 80, 310). All GST-fusion proteins were purified by affinity chromatography using glutathione sepharose 4B resin followed by size exclusion chromatography into storage buffer (20 mM HEPES pH 7.5, 200 mM NaCl, 1 mM DTT). High yields of purified protein were obtained routinely using this method.

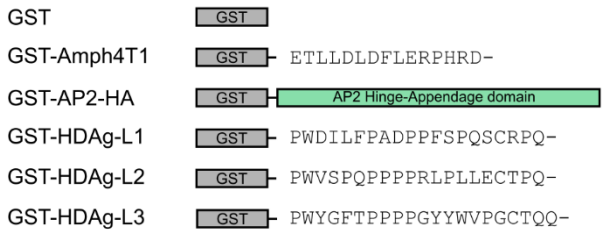


Figure 29 Glutathione S-transferase (GST) fusion proteins used during the development of the GST pull-down assays. Amph4T1: clathrin binding motif (CBM) of human amphiphysin I plus additional residues derived from the expression vector; AP2-HA: hinge-appendage domain of the AP2 β2 subunit; HDAg-L1: C-terminal extension of clade I Hepatitis D virus large antigen; HDAg-L2: C-terminal extension of clade II Hepatitis D virus large antigen; HDAg-L3: C-terminal extension of clade III Hepatitis D virus large antigen. GST alone served as a negative control.

In order to test whether the viral CBMs could bind to clathrin, a GST pull-down with pig brain cytosol, a clathrin-rich cell lysate, was performed using these GST-tagged clathrin-binding motifs as bait (Figure 30). No binding between HDAg-L1/2/3 and clathrin was detected, and the CBM of construct Amph4T1, which has been previously used as a positive control in pull-down experiments (308), was also unable to pull down clathrin. The hinge-appendage domain of AP2 was the only protein capable of capturing clathrin, indicating that the interaction between a single binding motif and NTD might not be strong enough to isolate clathrin from cytosolic extracts under the conditions tested.

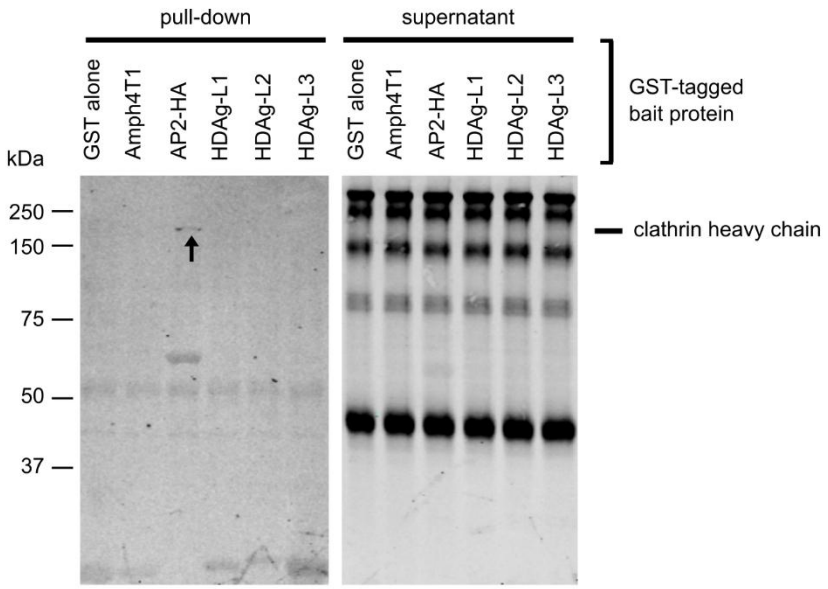


Figure 30 Pull-down assessing the ability of GST-tagged cellular and viral clathrin-binding motifs to capture clathrin from pig brain cytosol. Magnetic glutathione beads (20 µL suspension) were pre-loaded with 200 µg purified GST fusion proteins or GST as a negative control (in assay buffer: 25 mM HEPES-KOH, 125 mM potassium acetate, 5 mM magnesium acetate, 2 mM EDTA, 2 mM EGTA, 1 mM DTT, 0.1% Triton X-100, pH 7.2). The beads were washed with assay buffer and then incubated with 200 µL pig brain cytosol for 120 min. The supernatant was collected, the beads were washed (wash buffer: 1x PBS, 0.1% Triton X-100, 1 mM DTT) and bound proteins competitively eluted in 40 µL elution buffer (50 mM reduced glutathione in wash buffer). Bound (pull-down) and supernatant samples were separated on a 10% SDS-PAGE gel (10 µL of pull-down sample, 5 µL of supernatant sample per lane) and Western blotting using an α-clathrin heavy chain antibody was performed. The expected molecular weight of clathrin heavy chain is indicated on the right side, and an arrow highlights clathrin pulled down by the GST-AP2-HA fusion protein.

4.1.2 Purification and biochemical characterization of His-NTD-NEMO, a dimeric NTD construct

Since the interaction between HDAg-L1/2/3 or Amph4T1 and NTD was not strong enough to affinity-purify endogenous clathrin from pig brain lysate, a more sensitive technique to detect clathrin:adaptor interactions had to be developed. One option would have been to purify clathrin from brain cytosol and use it at high concentrations in pull-downs. However, the use of recombinantly expressed clathrin NTD seemed like a more sensible choice because recombinantly expressed protein can straightforwardly be mutagenized, a factor important for the planned studies trying to elucidate the role of the separate binding sites on NTD. Pull-downs using purified clathrin N-terminal domain have been conducted previously (1, 4, 77, 311). However, the interaction between NTD and CBM-containing peptides are likely very weak. Therefore, a construct of NTD fused to the oligomerization domain of nuclear factor κB (NF-κB) essential modulator, termed His-NTD-NEMO throughout this thesis, was explored. Using an oligomerizing

construct in GST pull-downs could increase the avidity of NTD:peptide interactions and thus enhance the capture and subsequent detection of clathrin NTD.

His-NTD-NEMO was expressed overnight in B834(DE3) *E. coli*. The protein was affinity-purified on Ni-NTA agarose resin, competitively eluted using a high-imidazole buffer and then further purified by size exclusion chromatography into a 20 mM Tris pH 7.5, 200 mM NaCl, 1 mM DTT buffer (Figure 31). The size exclusion elution profile of His-NTD-NEMO showed three slightly overlapping peaks. Analysis of the fractions by SDS-PAGE indicated that the first peak contained a protein running slightly above 50 kDa as the main component, whereas the third peak consisted of a protein running at around 37 kDa. The second peak was not only smaller than the first and third, but also contained much less protein that was quite heterogeneous in size. Full-length His-NTD-NEMO has an expected molecular weight of 58 kDa, of which 15.5 kDa can be attributed to the NEMO oligomerization domain. It was therefore proposed that during expression and purification of His-NTD-NEMO, proteolytic events took place, cleaving the NEMO domain from the full-length His-NTD-NEMO protein. Accordingly, the three peaks observed during size exclusion would correspond to His-NTD-NEMO (first peak), an intermediate cleavage product (second peak), and His-NTD (third peak). Of these constructs, only His-NTD-NEMO should be able to oligomerize.

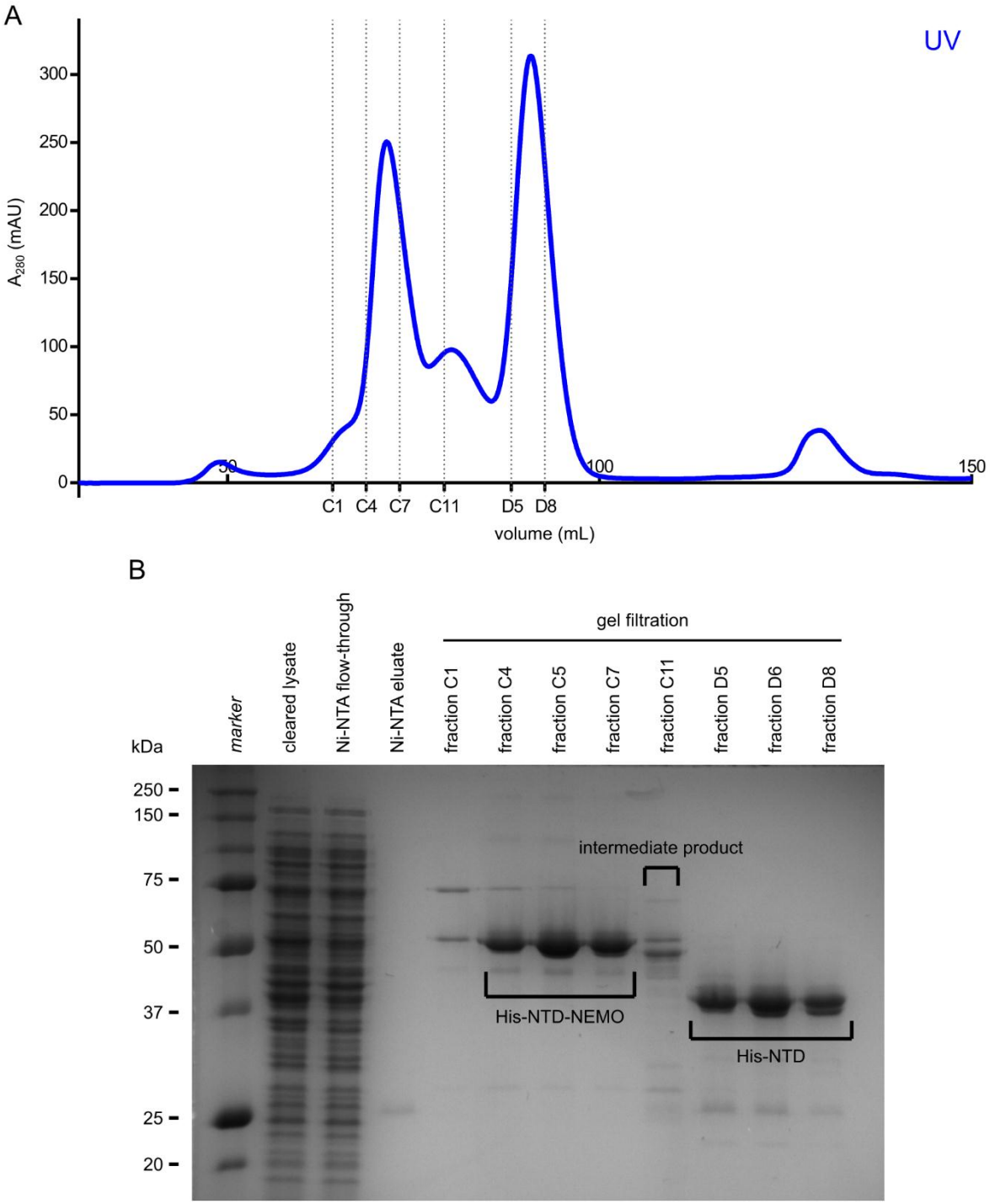


Figure 31 Purification of wild-type His-NTD-NEMO. The protein was purified by affinity chromatography using a Ni-NTA resin and then further purified by size exclusion chromatography. (A) Elution profile of His-NTD-NEMO subjected to size exclusion chromatography on an S200 16/600 column (column volume: 120 mL). (B) Samples from key steps of the purification process as well as selected fractions collected during gel filtration visualized on a Coomassie-stained 12% SDS-PAGE gel. Fractions containing protein samples assumed to be oligomerized His-NTD-NEMO, an intermediate product and monomeric His-NTD are highlighted.

In order to test this hypothesis, fractions of the first and third peak were concentrated and analyzed by multi-angle light scattering to accurately determine the molecular weight (and thus the oligomerization state) of the proteins. The proposed monomeric fraction eluted as a distinct peak with a weight-averaged molar mass of 42 kDa (Figure 32A), consistent with the predicated molecular weight of His-NTD (42.5 kDa). The fraction assumed to contain full-length His-NTD-NEMO eluted as two SEC-MALS peaks with molar masses of 111 kDa and 66 kDa (Figure 32A). Since the predicted molecular weight of one monomer of His-NTD-NEMO is 58 kDa, it may be possible that only the second SEC-MALS peak contained His-NTD-NEMO, or that the first SEC-MALS peak also contained dimerized His-NTD-NEMO.

To determine in which SEC-MALS peak His-NTD-NEMO was present, fractions were collected during SEC-MALS of this protein, separated on an SDS-PAGE gel and probed with an α-clathrin NTD antibody (Figure 32B). Only the second SEC-MALS peak with a molar mass of 66 kDa, eluting between 12 to 14 mL, was recognized by the antibody, implying that the first peak contained a contaminant co-eluting with His-NTD-NEMO during the preparative size exclusion experiment. Since the size exclusion column used for MALS was a higher resolution column, this co-purifying contaminant could be separated.

The disparity between the experimentally determined and predicted molar masses of His-NTD-NEMO (66 kDa and 58 kDa, respectively) could be caused by concentration-dependent oligomerization. To address this question, SEC-MALS was performed with the presumed His-NTD-NEMO fraction at different concentrations (Figure 32C). Interestingly, at low concentrations the weight-averaged molar mass of the eluted protein was consistent with the expected molar mass of monomeric His-NTD-NEMO. With increasing concentration, His-NTD-NEMO eluted at lower volumes and showed an increase in weight-averaged molar mass, indicating that His-NTD-NEMO indeed exists in concentration-dependent oligomerization equilibrium.

To further investigate the nature of the putative His-NTD protein, it was analyzed by circular dichroism. The oligomerization domain of NEMO, spanning residues 246-365, is mainly α-helical (312). Evaluation of the circular dichroism spectrum of His-NTD revealed that it contained less α-helical secondary structures than full-length His-NTD-NEMO (Figure 33A). In addition, this protein was detected by an α-clathrin NTD antibody (Figure 33B). Taken together, these data

clearly support the hypothesis that the third peak observed during purification of His-NTD-NEMO (Figure 31) represents monomeric His-NTD.

Having demonstrated that fusion with the NEMO domain leads to oligomerization of NTD, it was next tested whether use of this construct would enhance binding between NTD and peptide in GST pull-downs. A GST pull-down using a GST-tagged peptide encompassing the AP2 CBM motif (GST-GDLLNLDLGPPV) as bait and equal concentrations of oligomerizing His-NTD-NEMO or monomeric His-NTD as prey was conducted[1] (Figure 33B). Capture of His-NTD-NEMO was roughly seven times more efficient than capture of His-NTD, presumably due to oligomerization and thus avidity-enhanced binding, indicating that His-NTD-NEMO is much better suited to study weak clathrin:adaptor interactions than NTD on its own.

[1] This experiment was performed concurrently with the optimization of pull-downs described in the next section.

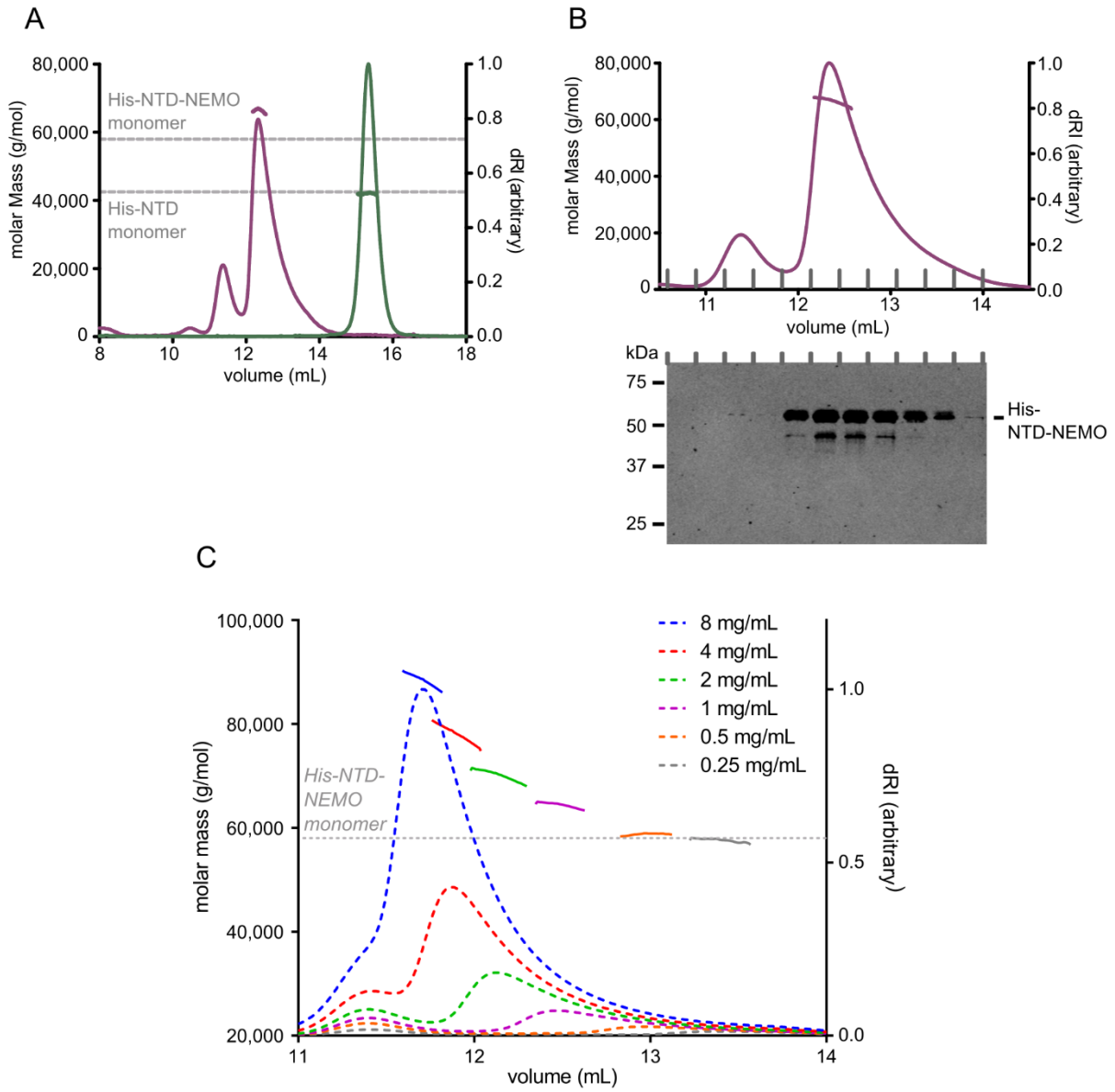


Figure 32 Analysis of His-NTD-NEMO oligomerization by size exclusion multi-angle light scattering (SEC-MALS). (A) Comparison of the SEC-MALS profiles of His-NTD-NEMO (purple) and His-NTD (green) proteins, monitored using the solvent differential refractive index (dRI). Weight-averaged molar masses calculated from the dRI and light scattering are shown as lines across the elution profiles. The expected molar masses for a His-NTD-NEMO and His-NTD monomers are shown as dashed grey lines. (B) Western blot analysis of the two main elution peaks observed during SEC-MALS of His-NTD-NEMO. The weight-averaged molar mass is shown as a line across the elution profile (upper panel, purple lines). Fractions (0.31 mL) were collected throughout the experiment (grey ticks on horizontal axis) and 0.24% of each fraction were loaded onto a 12% SDS-PAGE gel and analyzed by Western blot with an anti-clathrin NTD antibody (lower panel). (C) Analysis of concentration-dependent oligomerization of His-NTD-NEMO by SEC-MALS. Elution profiles of His-NTD-NEMO injected at different concentrations (0.25-8 mg/mL) were monitored by measuring the dRI (dashed colored lines). Weight averaged molar masses are shown as solid colored lines across the elution profiles. The expected molar mass of monomeric His-NTD-NEMO is indicated (grey dotted line).

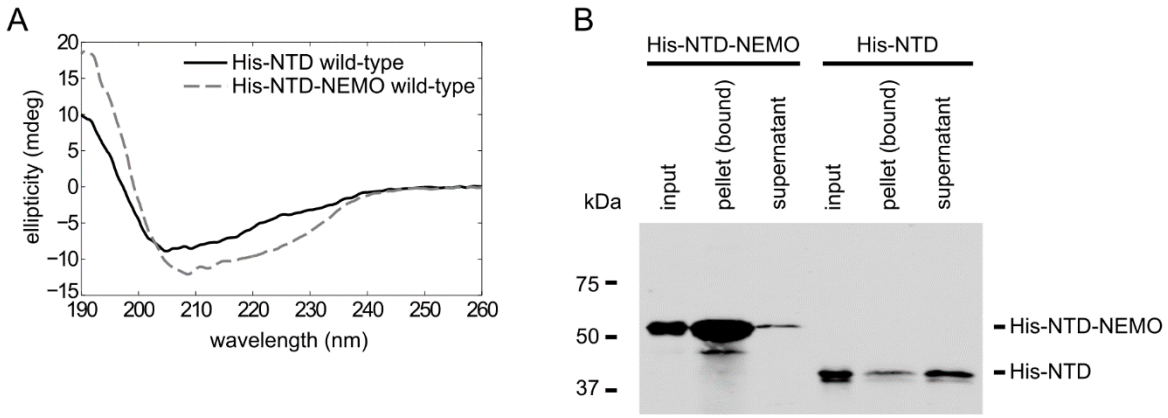


Figure 33 Comparison of circular dichroism spectra and binding affinities of His-NTD-NEMO and His-NTD. (A) Circular dichroism spectra of wild-type His-NTD and His-NTD-NEMO. His-NTD-NEMO showed increased ellipticity at 190-200 nm and a more pronounced minimum at 208 nm, consistent with increased α-helical content compared to His-NTD. (B) GST pull-down experiment comparing the efficiency with which the monomeric His-NTD and the oligomeric His-NTD-NEMO construct bound to a GST-tagged bait. Glutathione sepharose beads loaded with 500 µg of GST-AP2CBM (described in section 4.1.4) bait protein were incubated with 0.1 mg/mL of either His-NTD-NEMO or His-NTD. The mixture was centrifuged, the unbound prey protein (supernatant) removed and the bound sample eluted using 1x SDS-PAGE buffer (pellet). The samples were resolved on a 12% SDS-PAGE gel and analyzed using Western blot with an anti-clathrin NTD antibody. Gel loading: 0.33% of the input, 11.25% of the pellet and 0.6% of the supernatant samples.

4.1.3 Troubleshooting of GST pull-down assays

Having established His-NTD-NEMO as a suitable prey protein, GST pull-downs were performed in order to assess the interaction between NTD and viral clathrin binding motifs derived from hepatitis D virus large antigen. Initial experiments using magnetic glutathione beads as a resin delivered promising results, showing binding of the positive controls Amph4T1 and AP2-HA, as well as binding of the viral peptides HDAg-L1, -L2 and -L3 to His-NTD-NEMO (Figure 34).

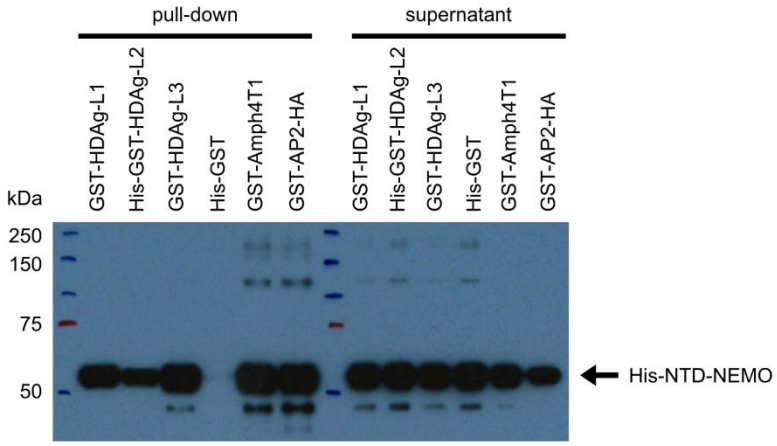


Figure 34 Initial GST pull-down experiment assessing binding between NTD and viral peptides. Magnetic glutathione sepharose beads (20 µL suspension) were loaded with 125 µg of GST- or His-GST-tagged bait protein in assay buffer (20 mM HEPES, 120 mM potassium acetate, 1 mM DTT, 0.1% Triton X-100, pH 7.5). His-GST alone served as a negative control. The beads were washed with assay buffer and incubated for 85 min with 0.1 mg/mL of His-NTD-NEMO in a final volume of 200 µL. The supernatant was collected, the beads were washed three times with wash buffer (20 mM HEPES pH 7.5, 200 mM NaCl, 1 mM DTT 0.1% Triton X-100) and bound proteins were eluted in 40 µL elution buffer (50 mM reduced glutathione in wash buffer, pH 7.5). Supernatant and bound samples (pull-down) were separated on a 12% SDS-PAGE gel (gel loading: 10 µL per lane. Western blotting was performed using an anti-clathrin NTD primary antibody and a horseradish peroxidase-conjugated anti-mouse secondary antibody. Bands were visualized on film by chemiluminescence.

However, the results of these initial experiments could not be reproduced with a fresh batch of purified His-NTD-NEMO. Therefore, a side-by-side pull-down comparing the original ("old") and a new batch of His-NTD-NEMO was performed (Figure 35A). When the original batch of prey protein was used, binding between His-NTD-NEMO and HDAg-L3 or GST-AP-2 HA could be observed, whereas the GST alone did not pull down His-NTD-NEMO. In contrast, when the fresh batch of His-NTD-NEMO was used, no interaction between HDAg-L3 and His-NTD-NEMO was detected. It was therefore concluded that the original batch of purified His-NTD-NEMO may have contained aggregated protein due to inappropriate storage conditions (-20 °C, long and slow thawing followed by re-freezing due to freezer breakdown). Such aggregated protein would greatly enhance avidity of binding, explaining the results obtained with the "old" batch of His-NTD-NEMO. During numerous subsequent purifications of His-NTD-NEMO, no such variability in protein quality was observed and proteins were stored consistently at -80 °C. In addition, none of the experiments presented in section 4.1.2 were conducted with the original "old" batch of His-NTD-NEMO, but with reproducibly purified protein.

It should be noted that the initial experiments were visualized using chemiluminescence. Accurate quantification and reproducibility of Western blots on photographic film are difficult due to effects such as substrate availability or depletion, differences in exposure time, the limited dynamic range of film, or the reactivity of the chemiluminescence reagents. Therefore, all subsequent Western blots were imaged using quantitative infrared fluorescence. As demonstrated in Figure 35A, the mode of detection did not influence the results obtained from different batches of His-NTD-NEMO.

Even after identifying and resolving the issue of an inconsistent batch of His-NTD-NEMO that led to irreproducible results, generating a reliable GST pull-down protocol proved difficult. An exhaustive set of pull-down conditions was explored to assess binding of clathrin NTD to GST (negative control), GST-tagged positive controls (Amph4T1, AP2-HA) and viral peptides (GST-HDAg-L1/2/3). Multiple factors that could potentially influence the interaction, including assay buffer, wash buffer, bait and prey incubation times, concentration of the prey protein, addition of bovine serum albumin (BSA) as a blocking agent, as well as elution conditions, were scrutinized. In addition, NTD alone or clathrin purified from pig brain cytosol were used as alternative prey proteins during these troubleshooting experiments. Some of the investigated conditions were adapted from the literature (79, 81), and all explored experimental protocols are summarized in Table 12. However, none of the tested assay setups (#1 - #16) delivered satisfying results. Binding between GST-AP2-HA and the prey protein was consistently observed, but binding to other bait proteins was very weak and/or irreproducible (representative blot: Figure 35B). In other protocols, non-specific binding of the prey was observed (representative blot: Figure 35C).

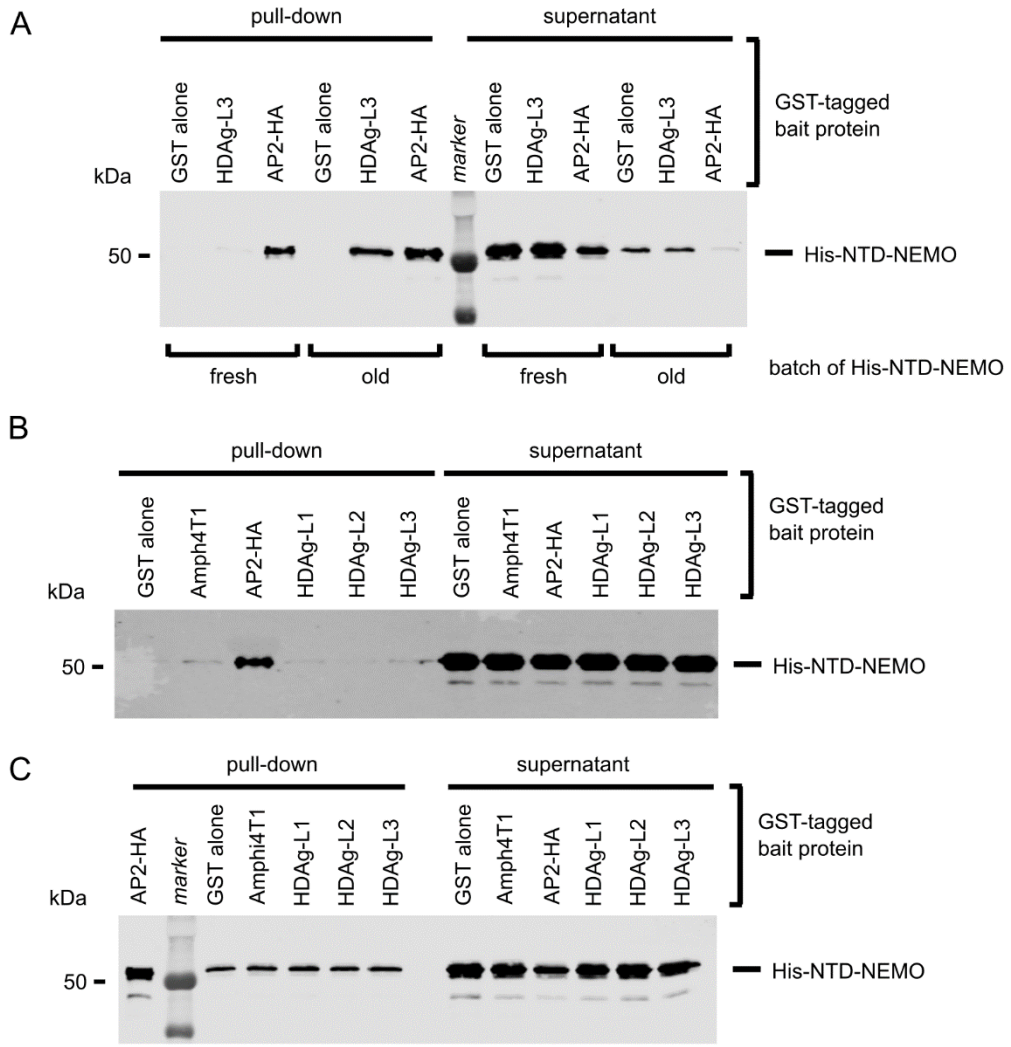


Figure 35 Representative results of unsuccessful pull-down experiments performed whilst establishing a reliable GST pull-down protocol. (A) Comparison of the binding behavior of the first batch of purified His-NTD-NEMO (old) to a subsequent purification batch (fresh). Experimental conditions: as in #4, Table 12, but with 90 min bait incubation time and 60 min prey incubation time. (B) Typical result of pull-down conducted with conditions enabling the detection of binding between His-NTD-NEMO and the positive control GST-AP2-HA, but not between His-NTD-NEMO and other bait proteins. Experimental conditions: #11, Table 12. (C) Typical result of pull-down performed using conditions under which all bait proteins bind non-specifically to His-NTD-NEMO. Experimental conditions: #6, Table 12.

All troubleshooting experiments were performed using magnetic glutathione beads (20 µL suspension). In comparison to non-magnetic agarose beads, magnetic beads can be used for performing pull-down assays in a high-throughput 96-well plate format, replacing time-consuming centrifugation steps with magnetic capture of the beads. Such "magnetic GST pull-downs" are routinely performed in our laboratory and usually yield excellent results under standard conditions (see also Chapters 6 and 7).

As a last resort for troubleshooting the GST pull-downs with His-NTD-NEMO, a pull-down using non-magnetic glutathione sepharose 4b beads was attempted (conditions: #17, Table 12). Moreover, similar to the assay described by Drake and Traub (79), increasing amounts of the positive control bait protein Amph4T1 were immobilized on the beads (Figure 36). Astonishingly, a completely different picture than the one observed during previous troubleshooting attempts emerged. With increasing amounts of GST-Amph4T1, His-NTD-NEMO was pulled down more efficiently, with 500 µg of immobilized bait giving the strongest signal. In contrast, 500 µg of GST alone showed no binding to His-NTD-NEMO. Multiple repeats of the pull-down using non-magnetic beads reproduced this result.

The assay was then repeated using magnetic glutathione beads with identical conditions and concentrations of bait and prey proteins. However, the GST alone negative control showed as much binding to the prey as Amph4T1, indicating that His-NTD-NEMO probably interacts non-specifically with the magnetic glutathione resin. Taken together, despite the reduced convenience, the non-magnetic agarose resin yielded clearly superior results and a reliable protocol for demonstrating interactions between His-NTD-NEMO and CBM peptides had been established ("standard protocol", see #17, Table 12, and Chapter 2).

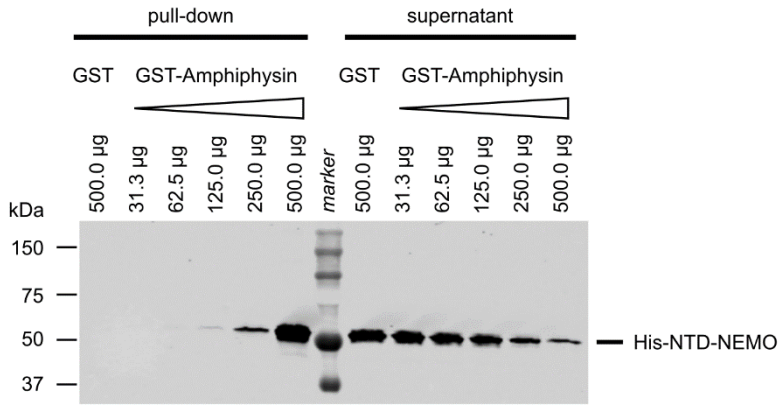


Figure 36 "Breakthrough" GST pull-down experiment demonstrating that very high concentrations of bait protein are required to pull-down His-NTD-NEMO efficiently. Increasing amounts of GST-tagged Amphiphysin (GST-Amph4T1) were immobilized on non-magnetic glutathione beads and incubated with His-NTD-NEMO. A high concentration of GST was used as the negative control. Experimental conditions are as in #17, Table 12, but with 60 min prey incubation time. Samples (9 µL per lane) were separated on a 12% SDS-PAGE gel and visualized by Western blotting using an anti-clathrin NTD antibody.

Table 12 Conditions tested during troubleshooting of GST pull-down experiments. BSA: bovine serum albumin.

#	Assay buffer	Wash buffer	Elution buffer	Bait incubation (min)	Prey incubation (min)	Prey (mg/mL)	BSA (mg/mL)	Result
1	25 mM HEPES-KOH pH 7.2, 125 mM potassium acetate, 5 mM magnesium acetate, 2 mM EDTA, 2 mM EGTA, 0.1% NP-40	1x PBS, 0.1% NP-40	SDS-PAGE buffer	30	60	Clathrin, 0.02	0.1	Only strong binding of AP-2 HA to clathrin, no binding of other adaptors to clathrin
2	as in 1	as in 1	50 mM reduced glutathione in wash buffer	60	240	NTD, 0.1	0.1	Only binding of AP-2 HA to NTD
3	20 mM HEPES, 120 mM potassium acetate, 1 mM DTT, 0.1% Triton, pH 7.5	20 mM HEPES pH 7.5, 200 mM NaCl, 0.1% Triton, 1 mM DTT	50 mM reduced glutathione in wash buffer	60	800	NTD, 0.1	0.1	Only binding of AP-2 HA to NTD
4	as in 3	as in 3	as in 3	60	110	His-NTD-NEMO, 0.1	0.1	Only binding of AP-2 HA and very weak binding of Amph4T1 to NTD
5	as in 3	as in 3	as in 3	150	60	NTD, 1.0 or His-NTD-NEMO, 1.0	0.1	Background of non-specifically bound prey in all pull-down samples due to high NTD or His-NTD-NEMO concentration
6	as in 1, but 0.1% Triton instead of 0.1% NP-40	1x PBS, 0.1% Triton	SDS-PAGE buffer	120	160	His-NTD-NEMO, 0.1	0.1	Non-specific binding of His-NTD-NEMO to all bait proteins including negative control
7	20 mM HEPES, 120 mM potassium acetate, 0.1 mM DTT, 0.1% Triton, pH 7.2	same as assay buffer	boiling in SDS-PAGE buffer	60	120	His-NTD-NEMO, 0.1	-	as in 6
8	as in 7	as in 3	50 mM reduced glutathione	75	150	His-NTD-NEMO, 0.1	0.1	Only binding of AP-2 HA to His-NTD-NEMO

#	Assay buffer	Wash buffer	Elution buffer	Bait incubation (min)	Prey incubation (min)	Prey (mg/mL)	BSA (mg/mL)	Result
			in wash buffer					
9	as in 7	as in 3 plus 120 mM potassium acetate	50 mM reduced glutathione in wash buffer	75	150	His-NTD-NEMO, 0.1	0.5	Only binding of AP-2 HA and Amph4T1 to His-NTD-NEMO, weak nonspecific background binding to other baits including negative control
10	as in 6	1x PBS, 1 mM DTT, 0.1% Triton	50 mM reduced glutathione in wash buffer	210	170	His-NTD-NEMO, 0.1	-	Only binding of AP-2 HA to His-NTD-NEMO
11	as in 10	as in 10	as in 10	75	90	His-NTD-NEMO, 0.1	0.1	Only binding of AP-2 HA to His-NTD-NEMO
12	80 mM Tricine pH 7.5, 150 mM NaCl, 1 mM DTT, 0.1% Triton	1x PBS, 1 mM DTT, 0.1% Triton	50 mM reduced glutathione in wash buffer	30	175	His-NTD-NEMO, 0.1	0.1	Only binding of AP2-HA to His-NTD-NEMO
13	as in 12	same as assay buffer	50 mM reduced glutathione in wash buffer	30	175	His-NTD-NEMO, 0.1	0.1	Only binding of AP2-HA to His-NTD-NEMO
14	as in 12	80 mM Tricine pH 7.5, 500 mM NaCl, 1 mM DTT, 0.1% Triton	50 mM reduced glutathione in wash buffer	30	175	His-NTD-NEMO, 0.1	0.1	Only binding of AP2-HA to His-NTD-NEMO
15	80 mM potassium phosphate pH 7.5, 150 mM NaCl, 1 mM DTT, 0.1% Triton	1x PBS, 1 mM DTT, 0.1% Triton	50 mM reduced glutathione in wash buffer	40	130	His-NTD-NEMO, 0.2	0.1	Non-specific binding of His-NTD-NEMO to all bait proteins including negative control
16	20 mM HEPES, 120 mM potassium acetate, 0.1 mM DTT, 0.1% Triton, pH 7.2	20 mM HEPES pH 7.5, 200 mM NaCl, 0.1% Triton, 1 mM DTT	50 mM reduced glutathione in wash buffer	35	120	His-NTD-NEMO, 0.1	0.1	Only binding of AP2-HA to His-NTD-NEMO

#	Assay buffer	Wash buffer	Elution buffer	Bait incubation (min)	Prey incubation (min)	Prey (mg/mL)	BSA (mg/mL)	Result
17 non-magnetic beads	25 mM HEPES-KOH, 125 mM potassium acetate, 5 mM magnesium acetate, 2 mM EDTA, 2 mM EGTA, 1 mM DTT, pH 7.2	1x PBS	boiling in SDS-PAGE buffer	120	120	His-NTD-NEMO, 0.1	0.1	Binding between His-NTD-NEMO and all positive controls, no binding between His-NTD-NEMO and negative control → standard protocol

4.1.4 Determination of relative binding affinities of viral and cellular peptides

Using the protocol developed in section 4.1.3, a GST pull-down was performed to compare the binding affinity of HDAg-L1, -L2 and -L3 (see Figure 29) with the affinity of Amph4T1 for NTD (Figure 37A). Interestingly, all three viral peptides demonstrated weaker binding to His-NTD-NEMO than did Amph4T1. For HDAg-L3, the interaction could barely be detected, and for HDAg-L2, binding was not reproducibly discernible over binding observed with the negative control (GST alone).

To investigate whether the weak binding between HDAg-L2 and NTD was caused by incidental peculiarities of the L2 sequence fused to GST, a set of alternative HDAg-L2 fusion proteins was designed (Figure 37B). The HDAg-L2 sequence used so far was derived from the C-terminus of the large antigen of HDV clade 4, e.g. isolate Taiwan-Tw-2b. Three other C-terminal extensions of the large antigen with a slightly different amino acid sequence were explored: HDAg-L2-2, derived from isolate Taiwan-3 (HDV clade 2), HDAg-L2-3, derived from HDV clade 6 and 7, e.g. isolates dFr48 and dFr45, and HDAg-L2-4, derived from isolate dFr2005 (HDV clade 5) (177). The large antigens of all clades of HDV contain a C-terminal cysteine as the fourth to last residue, which contributes to virus assembly via its farnesylation (143, 148). Since unwanted disulfide bridge formation or post-translational modification of this residue during expression in *E. coli* could potentially interfere with the interaction between HDAg-L2 and NTD, two more constructs of HDAg-L2 were designed, one carrying a cysteine to serine mutation (HDAg-L2-C251), and a construct that was truncated by introducing a stop codon to replace the C-terminal cysteine residue (HDAg-L2-C251Stop).

All alternative GST-HDAg-L2 constructs expressed well in BL21(DE3) pLysS *E. coli* and were purified as described previously (section 4.1.1). However, like HDAg-L2, none of these new GST fusion proteins bound strongly to His-NTD-NEMO in GST pull-down assays (Figure 37C). These results suggest that the interaction between HDAg-L2 and NTD, even though it could be observed in the crystal structures, is too weak to be reliably detected by GST pull-downs.

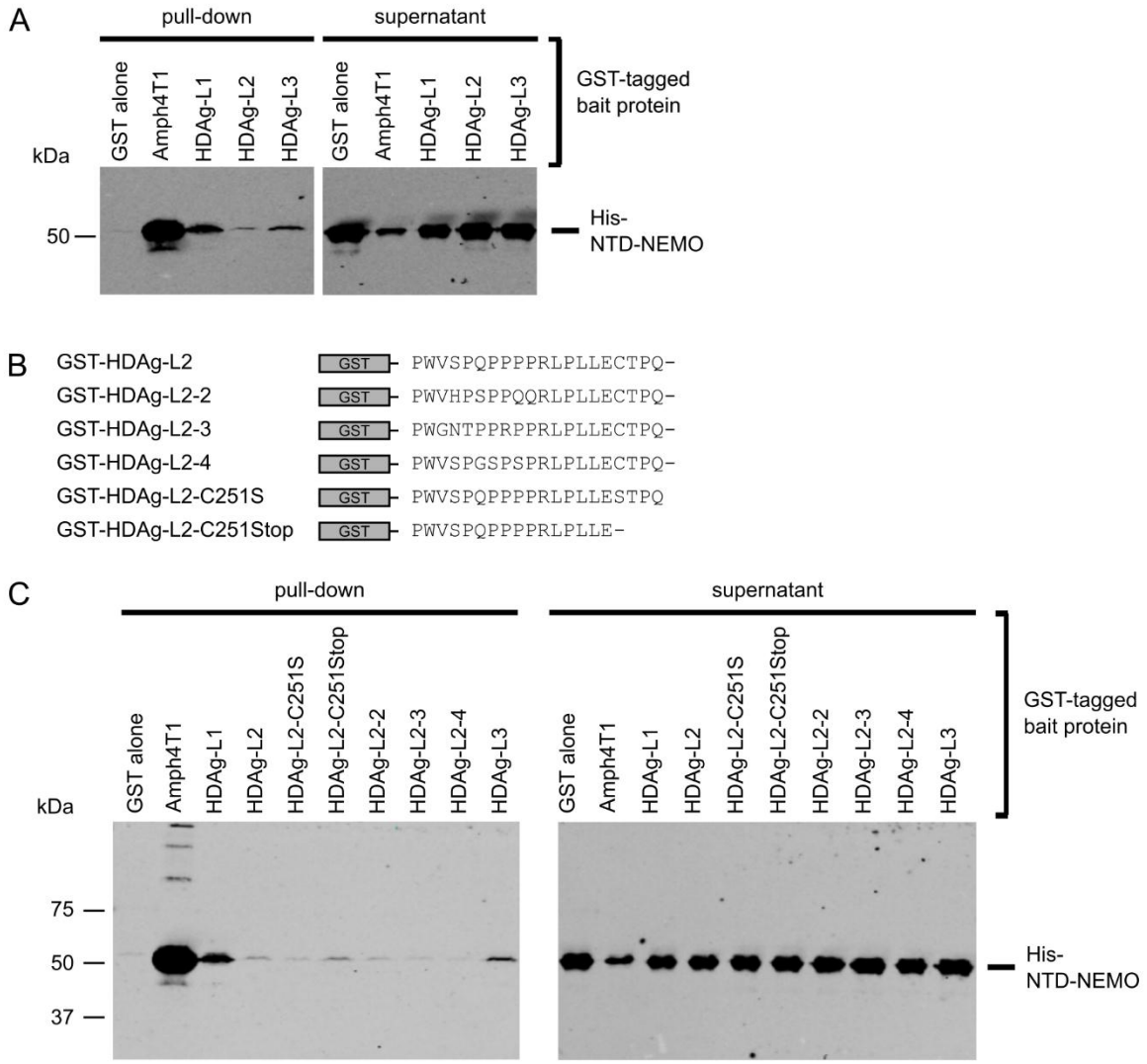


Figure 37 Evaluation of relative NTD binding affinities of HDV large antigen peptides. (A) A GST pull-down was performed using the standard protocol with 165 min of bait and 100 min of prey incubation. Amph4T1 was used as the positive control and GST alone as the negative control. Samples loading: 10 µL per lane for both pull-down and supernatant. (B) Additional GST fusion proteins of peptides derived from the C-terminal tail from HDAg-L2. HDAg-L2: original HDAg-L2 construct, HDAg-L2-2/3/4: constructs based on different isolates of HDV, HDAg-L2-C251S: HDAg-L2 construct with residue Cys251 mutated to serine, HDAg-L2-C251Stop: HDAg-L2 construct with Cys251 replaced by a stop codon. (C) A pull-down with the additional HDAg-L2 constructs was performed (standard protocol, 150 min bait incubation, 60 min prey incubation). Controls and sample loading as in (A).

As described in Chapter 3, crystallization experiments of NTD and peptides representing the C-terminal extension of HDAg-L3 were not successful. In addition, no immunoprecipitation or virus-like particle production assays that could demonstrate an interaction between HDAg-L3 and clathrin have been conducted previously. Since the GST pull-downs gave only weak evidence for a direct interaction between His-NTD-NEMO and HDAg-L3, it was hypothesized that HDAg-L3 might not recruit clathrin on its own, but via association with another cytosolic protein. The C-terminal tail of HDAg-L3 contains a potential YXXΦ motif (WYGFTPPPPGYYWVPGCTQQ). Cytosolic YXXΦ motifs bind to the medium subunits (μ) of adaptor protein complexes, thus mediating selection and packing of cargo into clathrin coated vesicles (57). Wang et al. previously speculated that HDAg-L3 might form a complex with AP2 and clathrin heavy chain (177).

To test whether HDAg-L3 associates with the C-terminal domain of the μ2 subunit of AP2 (c-μ2), a pull-down assay using purified GST-HDAg-L3 as bait and His-tagged c-μ2 as prey was performed (Figure 38). A canonical YXXΦ peptide fused to GST was used as a positive control, and GST alone, GST-Amph4T1, GST-HDAg-L1 and -L2 served as negative controls. Even though background binding was observed between the negative controls and c-μ2, GST-YXXΦ clearly demonstrated enhanced binding to c-μ2 in this assay. However, binding between c-μ2 and HDAg-L3 was not stronger than the detected background binding, indicating that AP2 (or presumably AP1) does not facilitate a potential interaction between HDAg-L3 and clathrin via binding to the YXXΦ motif of HDAg-L3.

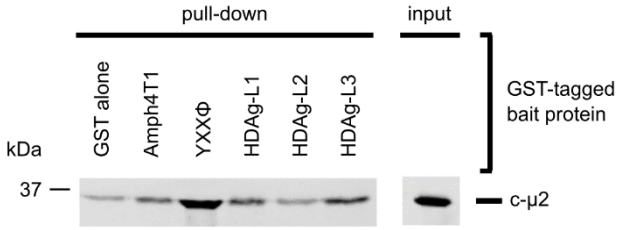


Figure 38 GST pull-down probing a potential association between HDAg-L3 and the C-terminal domain of the medium subunit of adaptor protein complex 2 (c-μ2). Magnetic glutathione beads (20 μL suspension) were loaded with 125 μg of GST-tagged bait proteins in assay buffer (10 mM HEPES, 350 mM NaCl, 1 mM DTT, 1 mM EDTA, 0.1% Triton X-100, pH 7.5) and incubated for 40 min. The beads were washed three times in assay buffer and incubated with 0.1 mg/mL of c-μ2 prey protein (in assay buffer) for 70 min. Beads were then washed again three times in assay buffer and bound proteins were competitively eluted (50 mM reduced glutathione in assay buffer). Samples were separated on a 12% SDS-PAGE gel (25% of pull-down samples, 5% of total input) and Western blotting using a mouse α-His antibody was performed to visualize c-μ2.

For the crystallization of NTD with cellular CBMs (Chapter 3), multiple versions of synthesized peptides of amphiphysin or AP2 were used, differing in the residues flanking the CBM. Therefore, GST fusion proteins of these peptides were designed (Figure 39A), expressed overnight BL21(DE3) pLysS *E. coli* and purified as described in section 4.1.1. The W box binding motif of amphiphysin (PWDLW) fused to GST was used as an additional positive control. Pull-downs using these GST-fused peptides as baits were performed and revealed broadly similar binding affinities of all cellular canonical CBMs and the W box motif for His-NTD-NEMO (Figure 39B,C). Taken together, of the six peptides found to bind to multiple sites on clathrin NTD in the crystallization experiments (Chapter 3), all but the HDAg-L2 peptide were found to bind reproducibly to His-NTD-NEMO in vitro when presented as GST fusion proteins.

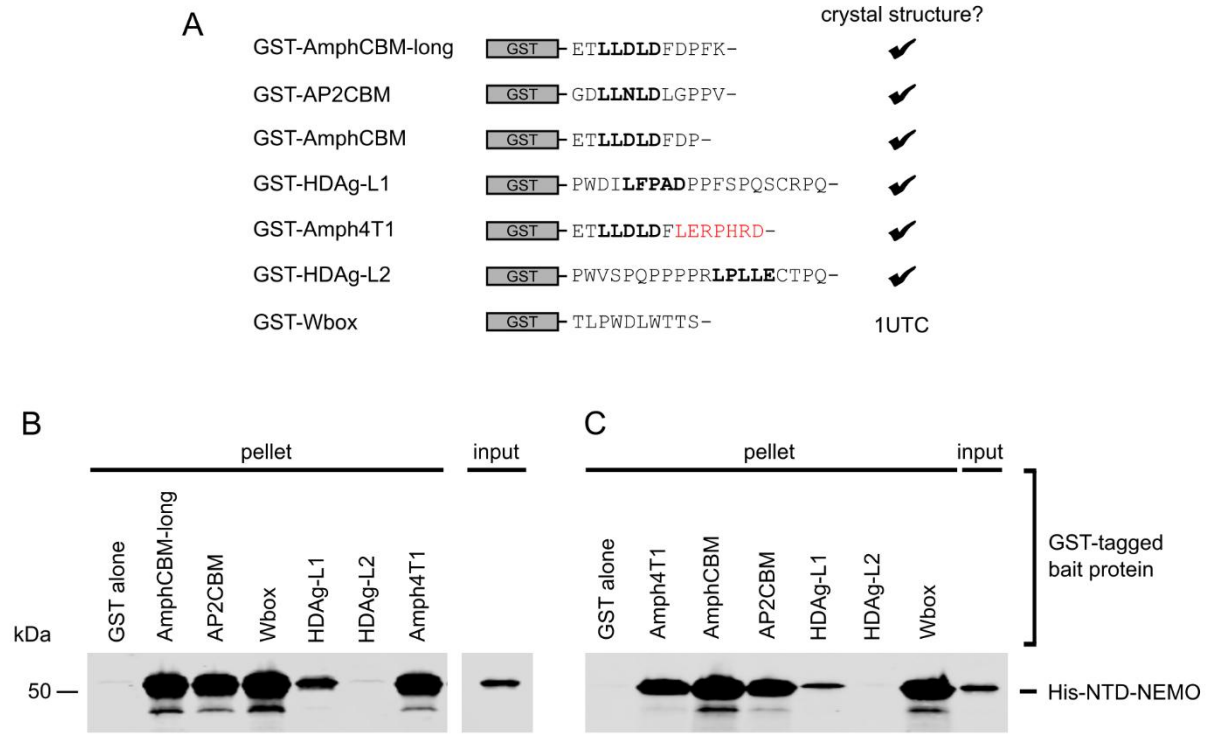


Figure 39 Relative binding affinities of cellular peptides for NTD. (A) Overview of GST fusion constructs used in pull-down experiments. Clathrin-binding motifs are highlighted in bold. Artificial residues of the Amph4T1 construct originating from the XhoI cloning site are marked red. Peptides which have been co-crystallized with NTD (as synthesized peptides, see Chapter 3) in this thesis are indicated, and the PDB ID of NTD in complex with a W box peptide is shown (4). (B, C) Comparison of binding affinities between cellular GST fusion peptides derived from amphiphysin (AmphCBM, AmphCBM-long, Wbox, Amph4T1), Hepatitis D virus (HDAg-L1, -L2) and hinge region of the β2 subunit of AP2 (AP2CBM). GST alone served as the negative control. GST pull-downs and Western blotting were performed using the standard protocol.

4.2 Design and characterization of His-NTD-NEMO constructs carrying mutations at the putative binding sites of NTD

The crystal structures presented in Chapter 3 showed that viral and cellular peptides can bind to multiple sites on NTD simultaneously. It was also demonstrated that cellular peptides and HDAg-L1, but not HDAg-L2 or -L3, interact with His-NTD-NEMO strongly enough to be reliably detected in GST pull-downs (section 4.1.4). Therefore, a number of His-NTD-NEMO constructs were designed in which each binding site of NTD was abolished either on its own, or in combination with the other binding sites ("compound mutants") (Figure 40 and Table 13). These proteins could then be used in GST pull-downs to investigate whether all binding sites contribute to NTD:peptide interaction in vitro and, potentially, to dissect the contribution of each binding site to the overall interaction.

At the clathrin box, mutations previously reported to interrupt binding are Q89M and F91A (77). At the arrestin box, mutation W164E has been described to abolish binding (1). For the W box, mutations of residues Gln152 and Ile154 either alone or in combination disrupt binding (2, 4). Mutations E11K and Q14D+Q16M+N17S are reported to affect binding at the Royle box (2). Based on the structural data obtained with NTD and peptides bound at the different interaction sites (Chapter 3), additional mutations were designed.

At the clathrin box, hydrogen bonds between side chain oxygen and nitrogen atoms of NTD residue Gln89 and the backbone of bound peptides were observed. In addition, Phe91 was one of the key residues forming the hydrophobic binding pocket. Therefore, a clathrin box mutant in which Gln89 and the bulky hydrophobic residue Phe91 were substituted to alanine and lysine, respectively, was constructed (Q89A+F91K). At the arrestin box, hydrogen bonds to the bound peptides originated from NTD residue Gln192, and one of the hydrophobic residues involved in forming the interaction site was Leu183. Thus, two more arrestin box mutants were designed, one where the ability of Gln192 to form this hydrogen bond was disrupted (Q192Y), and another where the hydrophobic Leu183 was mutated to a charged residue and the ability of Gln192 to form hydrogen bonds was inhibited (L183K+Q192A). Lastly, at the Royle box, residue Phe9 seemed to be a key interface residue. Therefore, two mutations perturbing the physicochemical properties of this residue with the aim of perturbing the hydrophobic binding pocket were

constructed by replacing Phe9 with bulkier or charged residues (F9E and F9W). An additional mutation targeted residue Ile7 and replaced it with a charged residue (I7K).

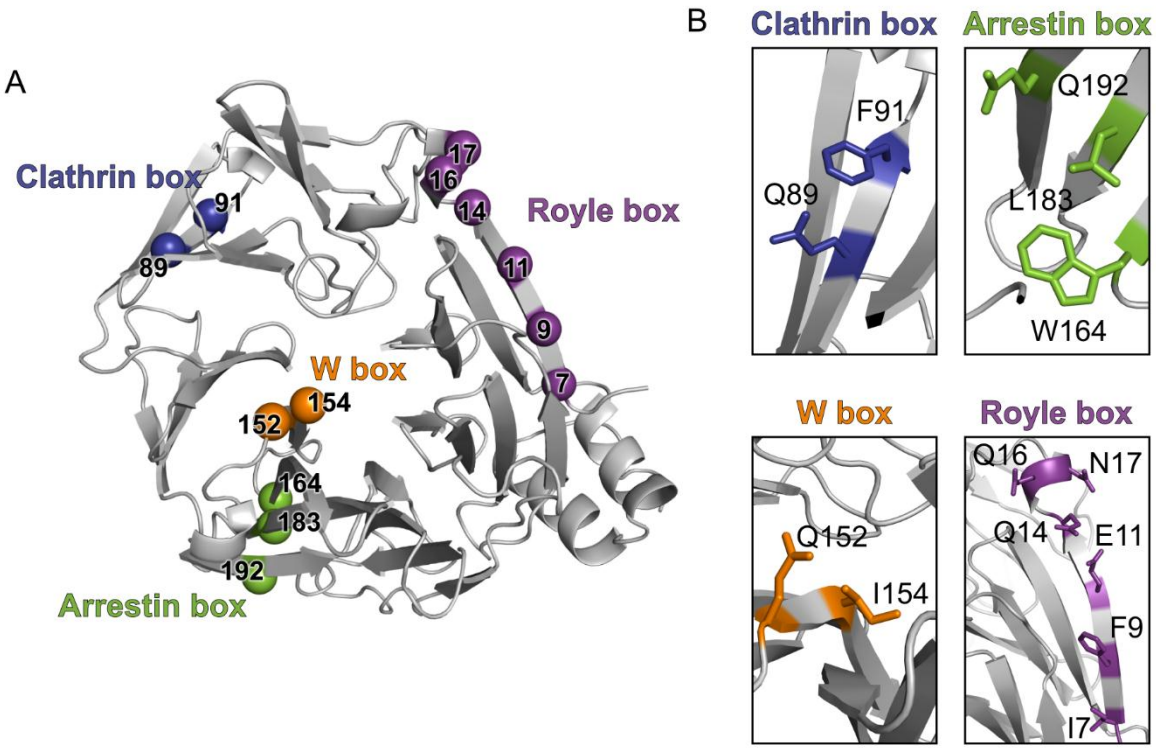


Figure 40 Overview of NTD residues chosen for the generation of mutagenized His-NTD-NEMO constructs. (A) Cartoon of NTD (grey) showing the localization of residues that were mutated on their own or in combination to disrupt binding at the Clathrin box site (blue), the Arrestin site (green), the W box site (orange) and the Royle box site (purple). (B) Close-up view of the mutated residues.

Table 13 Clathrin NTD mutations introduced into His-NTD-NEMO.

Abbreviation	Mutation(s)	Site(s)	Reference
C1	Q89M	Clathrin box	(77)
C2	F91A	Clathrin box	(77)
C3	Q89A+F91K	Clathrin box	This work
A1	Q192Y	Arrestin box	This work
A2	W164E	Arrestin box	(1)
A3	L183K+Q192A	Arrestin box	This work
W1	Q152L	W box	(4)
W2	I154Q	W box	(4)
W3	Q152L+I154Q	W box	(2)
R1	I7K	Royle box	This work
R2	F9E	Royle box	This work
R3	F9W	Royle box	This work
R4	E11K	Royle box	(2)
R5	Q14D+Q16M+N17S	Royle box	(2)
C3A1	Q89A+F91K+Q192Y	Clathrin box + Arrestin box	This work
C3R3	Q89A+F91K+F9W	Clathrin box +Royle box	This work
A1R3	Q192Y+F9W	Arrestin box + Royle box	This work
C3A1R2	Q89A+F91K+Q192Y+F9E	Clathrin box + Arrestin box + Royle box	This work
C3A1R3	Q89A+F91K+Q192Y+F9W	Clathrin box + Arrestin box + Royle box	This work
C3A1R4	Q89A+F91K+Q192Y+E11K	Clathrin box + Arrestin box + Royle box	This work
C3A1W3	Q89A+F91K+Q192Y+Q152L+I154Q	Clathrin box + Arrestin box + W box	This work
C3A1W3R3	Q89A+F91K+Q192Y+Q152L+I154Q+ F9W	Clathrin box + Arrestin box + W box + Royle box	This work

The mutations were introduced into the His-NTD-NEMO construct by site-directed mutagenesis and mutant proteins were expressed and purified as described for wild-type (wt) His-NTD-NEMO (section 4.1.2). For all but one mutant, His-NTD-NEMO I7K (R1), the size exclusion chromatography elution pattern was similar to the one of wild-type His-NTD-NEMO, showing three slightly overlapping peaks (Figure 41). Therefore, for all mutants except R1, both oligomerizing His-NTD-NEMO and monomeric His-NTD lacking the NEMO domain could be purified and concentrated. For mutant R1, only the His-NTD-NEMO fraction was present and thus collected.

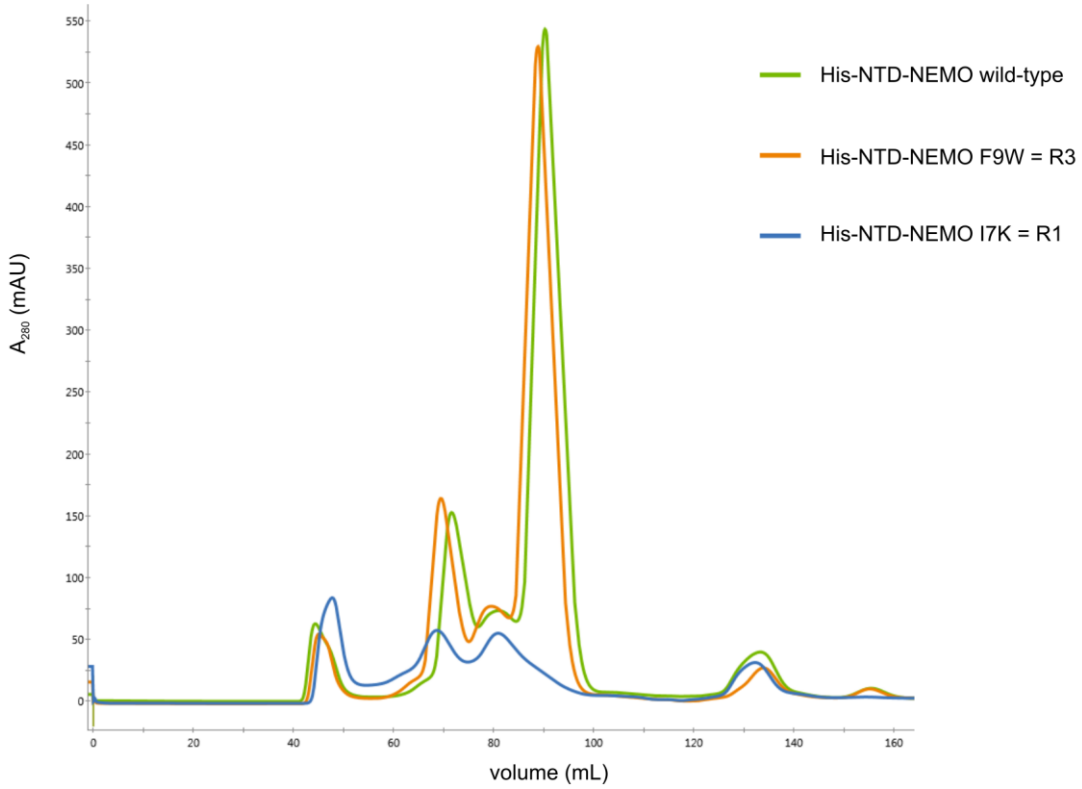


Figure 41 Comparison of size exclusion elution profiles of wild-type and mutant His-NTD-NEMO. The proteins were purified by Ni-NTA affinity chromatography followed by size exclusion chromatography using an S200 16/600 column (column volume: 120 mL). For wild-type His-NTD-NEMO (green) and all mutant versions of His-NTD-NEMO except for I7K = R1 (mutant F9W shown in orange as an example), three peaks eluted from the size exclusion column (compare to Figure 31). For mutant R1 (blue), only two peaks were detected.

Introduction of point mutations, even at surface residues, can affect protein folding and stability. Even though some of the designed NTD mutations have been described by others (Table 13), no evaluation of potential disruption of the protein secondary structure or overall fold has been reported. Therefore, a comprehensive characterization of all mutated His-NTD-NEMO variants was performed. Circular dichroism spectroscopy was performed using monomeric His-NTD (except for mutant R1=I7K) and showed that none of the mutants had significantly different CD spectra, consistent with wild-type and mutant His-NTD having highly similar secondary structure content (Figure 42).

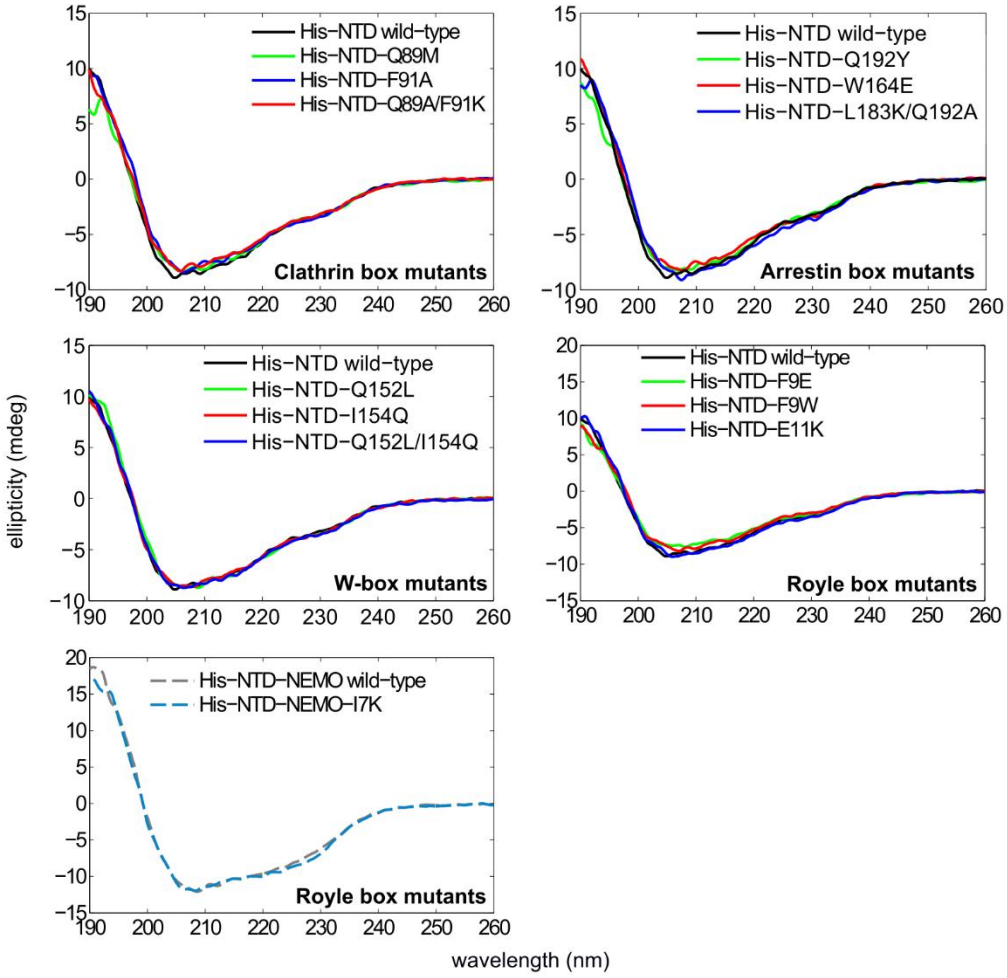


Figure 42 Circular dichroism spectra of wild-type and single-site mutants of clathrin NTD. Spectra of single site mutants (colored) are compared to wild-type His-NTD (black). For mutation R1=I7K (light blue), only the oligomerizing His-NTD-NEMO protein could be measured and is compared to wild-type His-NTD-NEMO (grey).

As measuring the overall secondary structure of a protein may not identify local destabilization, a second assay was employed to evaluate the effect of the mutations introduced into NTD. The thermal stability of the mutants was assessed by differential scanning fluorimetry (Figure 43B and Figure 46). Since the melting temperatures of monomeric His-NTD and oligomerizing His-NTD-NEMO did not differ (Figure 43A), the monomeric construct was used for these assays. None of the mutations introduced at the clathrin box or the W box affected the thermal stability of His-NTD, but mutations A2 (W164E), A3 (L183K+Q192A), R2 (F9E) and R4 (E11K) destabilized His-NTD by more than 3 K. Mutant R1 (I7K) could only be measured as the oligomerizing construct and was found to be significantly less stable than wild-type His-NTD-NEMO.

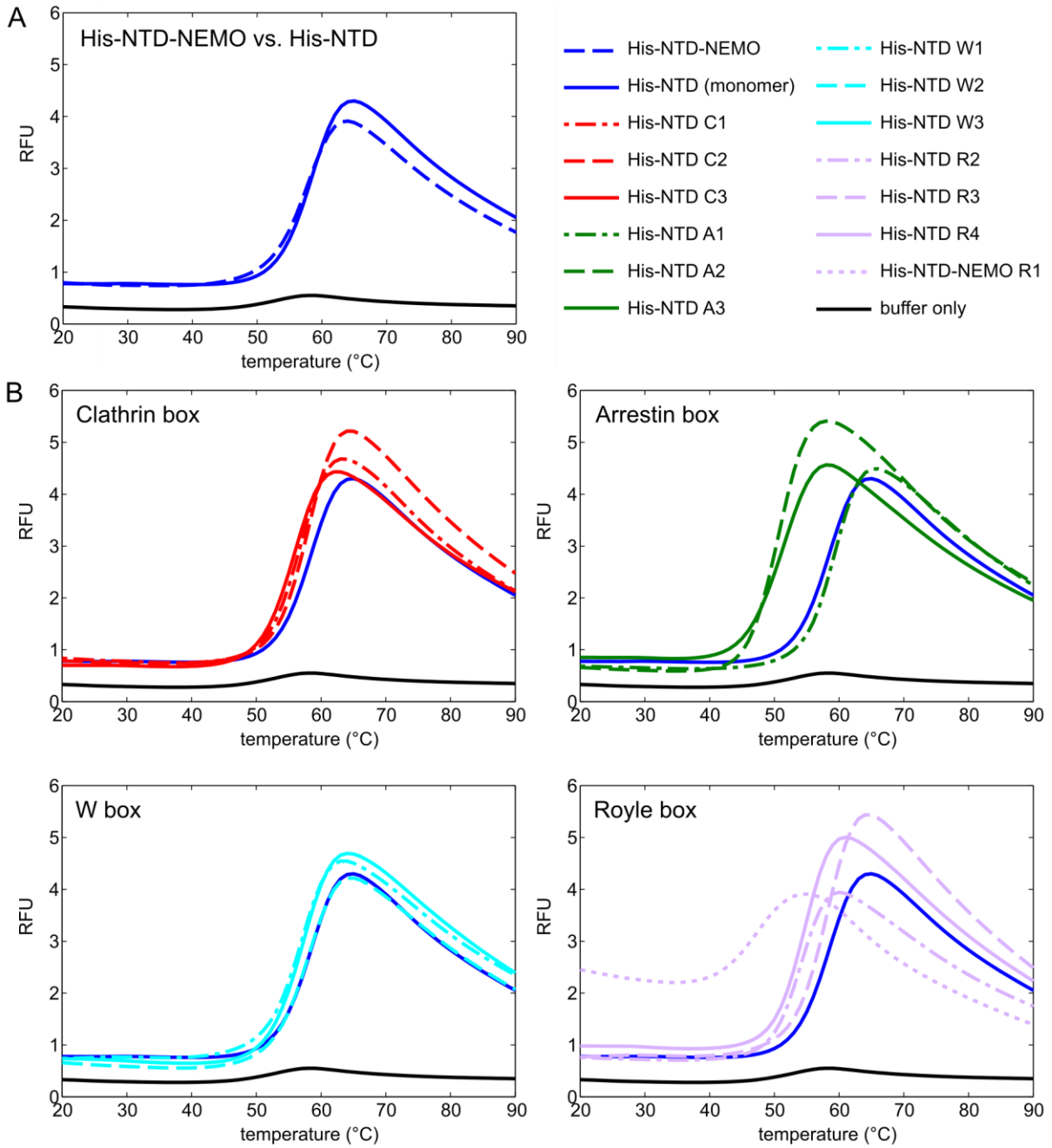


Figure 43 Differential scanning fluorimetry melting curves of wild-type and single-site mutated NTD. (A) The melting curve of wild-type His-NTD (solid dark blue line) or wild-type His-NTD-NEMO (dashed dark blue line) are virtually identical. (B) The melting temperature of wild-type His-NTD was compared to melting temperatures of NTD variants carrying mutations at the clathrin box (red), the arrestin box (green), the W box (cyan) or the Royle box (light purple). Data were obtained on a MiniOpticon real-time PCR system using 10x SYPRO Orange dye and 10 ng of protein in a final volume of 50 µL. The melting curve of the Royle box mutant R5 (Q14D+Q16M+N17S) is not displayed since it was obtained on the Viia7 real-time PCR platform, but the melting temperature of R5 is compared to all other determined ones in Figure 46. Data are shown as relative fluorescence units (RFU).

Taking the thermal stability data obtained for the single site mutants into account, "compound mutants" carrying mutations at two, three or all four binding sites of NTD were designed (Table

13). Since in the crystal structures, binding of peptides to the clathrin box, the arrestin box and the Royle box was observed, "double mutants" contained combinations of the Q89A+F91K mutation of the clathrin box, the Q192Y mutation of the arrestin box and the F9W mutation of the Royle box. In the "triple" mutant, clathrin box, arrestin box and Royle box were disrupted at the same time. Two additional triple mutants containing the R2 (F9E) or R4 (E11K) mutation instead of the R3 (F9W) mutation of the Royle box were also constructed. Lastly, a triple and a quadruple mutant containing the W3 (Q152L+I154Q) mutation of the W box in addition to the C3A1 double or the C3A1R3 triple mutation were also designed. All compound mutants were expressed and purified as before (section 4.1.2).

As observed with the mutations at single sites, the CD spectra of the compound mutants did not differ from the one of wild-type His-NTD (Figure 44). Differential scanning fluorimetry demonstrated that all double mutants and the triple mutant involving the R3=F9W mutation, but not the triple mutants containing the R2 or R4 mutation, had melting temperatures within 3 K of wild-type His-NTD. The triple and the quadruple mutants disrupted at the W box were slightly destabilized, their melting temperature being 3.3 or 3.6 K lower than that of wild-type His-NTD (Figure 45 and Figure 46).

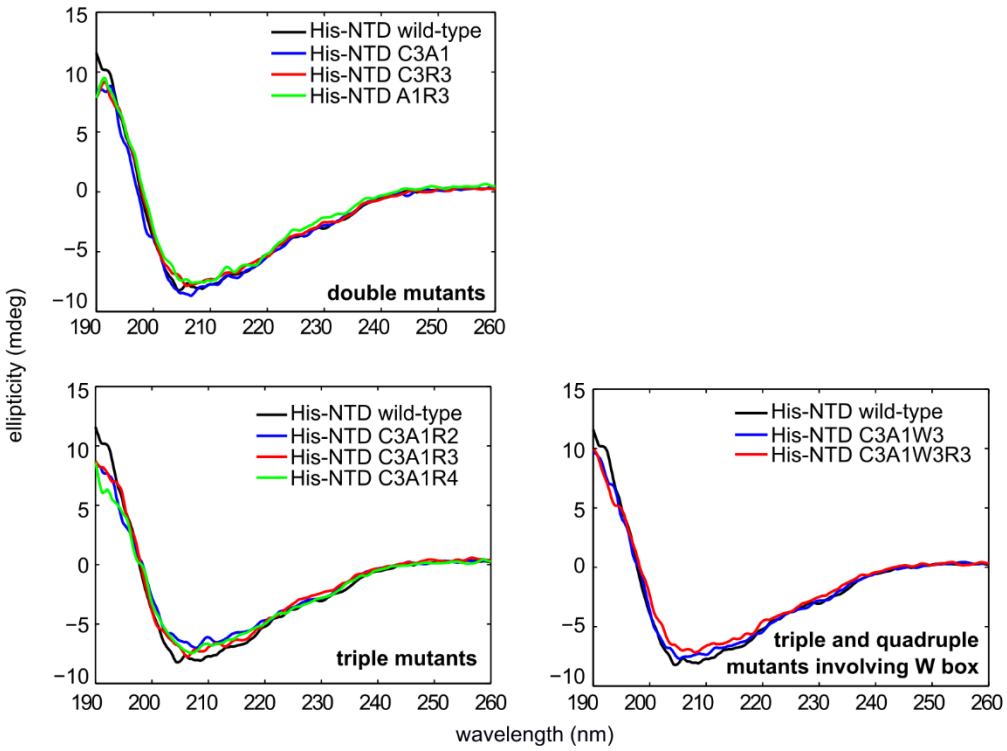


Figure 44 Circular dichroism spectra of wild-type and compound mutants of clathrin NTD. Spectra of compound mutants (colored) are compared to wild-type His-NTD (black).

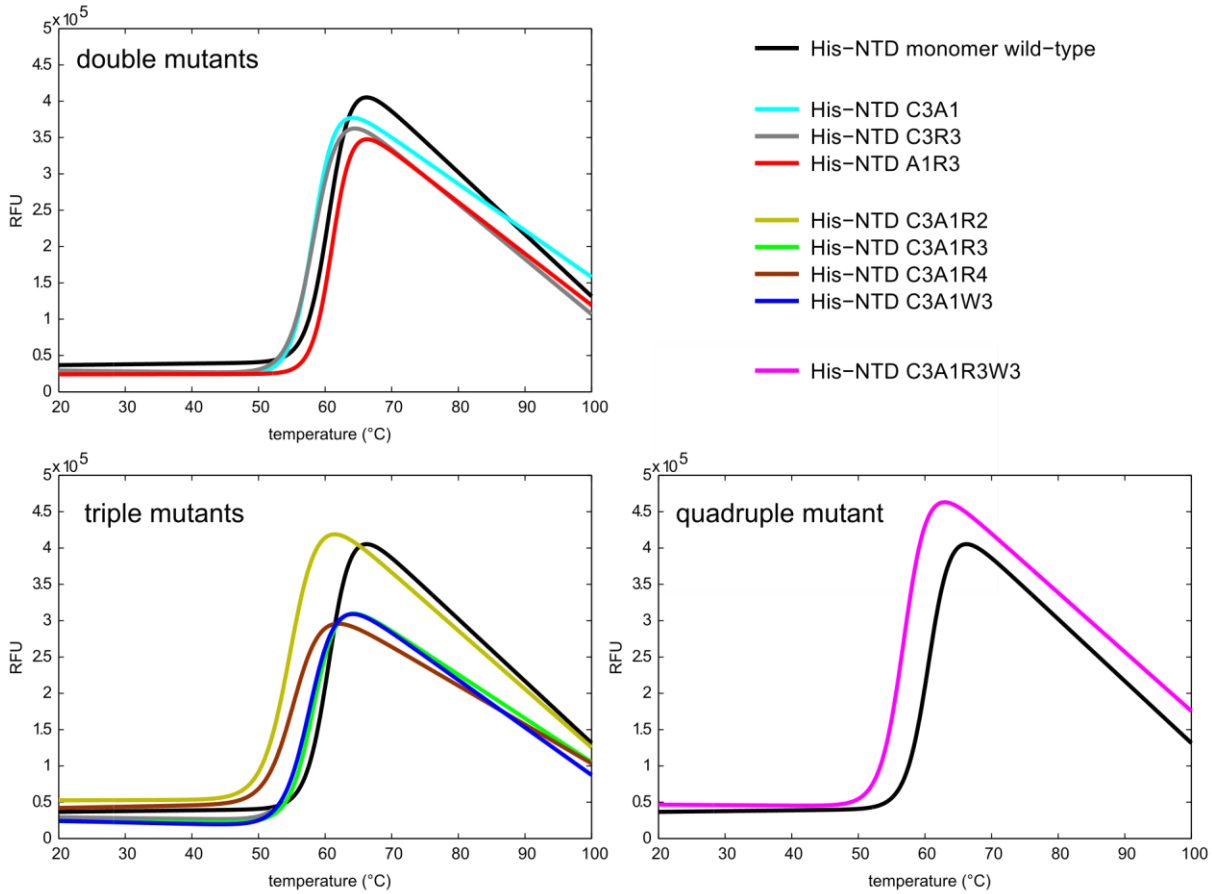


Figure 45 Differential scanning fluorimetry melting curves of wild-type and compound mutants of His-NTD. The melting curve of His-NTD (black line) is compared to that of His-NTD carrying mutations at two, three or all four binding sites (colored lines). Data were obtained on a Viia7 real-time PCR system using thermal 2 ng of protein and 1x shift dye (Life Technologies) in a final volume of 20 µL. Data are shown as relative fluorescence units (RFU).

In summary, CD spectroscopy showed that mutating one or multiple surface residues of NTD does not affect the overall secondary structure composition of the protein. However, as demonstrated by differential scanning fluorimetry, these mutations can have a profound impact on the thermal stability of NTD, indicating that the proper folding of these mutants was disturbed. The data presented here enabled the selection of a set of stably folded NTD single site and compound mutants. Only His-NTD-NEMO variants with a melting temperature differing by less than 3 K compared to wild-type were employed in further experiments (Figure 46).

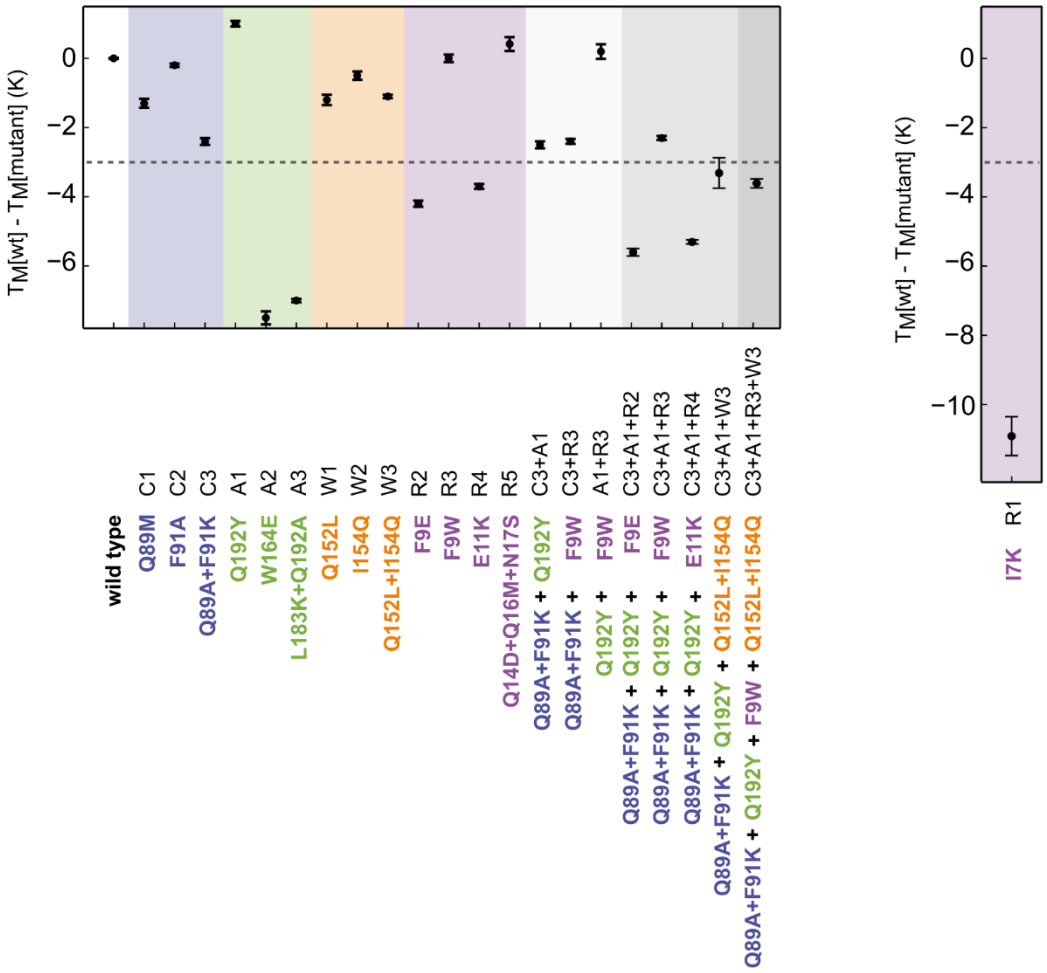


Figure 46 Summary of thermal stability screening of NTD single site and compound mutants. The melting temperatures (T_M[mutant]) of His-NTD with mutations affecting the clathrin box (blue), the arrestin box (green), the W box (orange), the Royle box (purple), or NTD with mutations affecting multiple binding sites (grey shades) are plotted relative to the melting temperature of wild-type NTD (T_M[wt]). Mutation I7K (R1) is displayed in a separate panel since it was measured against the His-NTD-NEMO construct. Mutations affecting the melting temperature of NTD by more than 3 K (dotted line) were viewed with caution. Samples were measured in triplicate; error bars show the standard deviation.

4.3 Relative contribution of separate NTD binding sites to overall NTD:peptide interactions

4.3.1 Binding between fusion peptides and single site mutants of NTD

Having identified mutations of the clathrin box, the arrestin box, the Royle box and the W box of NTD that do not interfere with protein folding, GST pull-downs were performed to investigate which binding sites contribute to NTD:peptide interactions in vitro. Using GST fusion proteins of cellular and viral peptides as bait, pull-downs were performed to assess binding between the peptides and single site mutants of NTD (Figure 47). Only one arrestin box (A1 = Q192Y) and

two Royle box mutants (R3 = F9W and R5 = Q14D+Q16M+N17S) were found to be well-folded (Figure 46), and only NTD residue Phe9 was implicated in binding according to the crystal structures. Thus, for the arrestin box and the Royle box, only one mutant each could be investigated (R3 and A1) in pull-down experiments. Since all three clathrin box mutants and all three W box mutants were well-folded (Figure 46), two mutants for each site (C1 and C3 for the clathrin box, W2 and W3 for the W box) were selected as prey for the pull-down experiments.

None of the prey proteins was found to interact with the GST negative control. In agreement with previous work (4), mutations of the W box on NTD greatly disturbed binding of His-NTD-NEMO to the Wbox peptide. Interestingly, none of the single site mutations completely disrupted the interaction of AP2CBM or amphiphysin-derived CBM peptides (Amph4T1, AmphCBM, AmphCBM-long) with NTD. All amphiphysin peptides bind to the clathrin box, the arrestin box and, for Amph4T1, also the Royle box on NTD in the crystal structures. For these three peptides, no differences in affinity between mutated and wild-type His-NTD-NEMO could be detected. This observation suggested that disruption of one binding site of clathrin alone is not sufficient to reduce binding of His-NTD-NEMO to the amphiphysin CBM. For AP2CBM, which binds to the clathrin box and the arrestin box in the crystal structures, mutation of the clathrin box (C3) or the arrestin box (R1) decreased, but did not abolish binding.

Although the viral peptides HDAg-L1/2/3 showed only very low or no affinity for His-NTD-NEMO in pull-downs (Figure 37), they were still tested with the single site His-NTD-NEMO mutants. As expected, HDAg-L2 and -L3 did not bind strongly to wild-type or mutant His-NTD-NEMO. Wild-type His-NTD-NEMO and HDAg-L1 interacted weakly, and similar binding was observed between HDAg-L1 and mutants C3, W2, W3 and R3. However, mutations Q89M (C1) and Q192Y (A1) seemed to increase the affinity of His-NTD-NEMO for HDAg-L1, suggesting that these mutations had an unexpected stabilizing effect on the interaction between HDAg-L1 and NTD. A similar slight stabilization could also be observed with the Q89M mutation in the GST-Wbox or GST-AmphCBM-long pull-downs, and the second tested clathrin box mutation, Q89A+F91K (C3), did not show such effects. Since the band of the His-NTD-NEMO variant Q89M additionally migrated anomalously for unknown reasons, this mutant was not studied further. For mutation Q192Y (A1), the case was more difficult. Protein His-NTD-NEMO Q192Y was the only stably folded arrestin box mutant, the other two investigated mutations showing a

melting temperature shift of at least -7 K compared to wild-type (Figure 46), excluding them from use as prey proteins in pull-down experiments. However, since no stabilizing effects of the Q192Y (A1) mutation were observed with other bait proteins, the His-NTD-NEMO A1 protein was considered suitable for further studies.

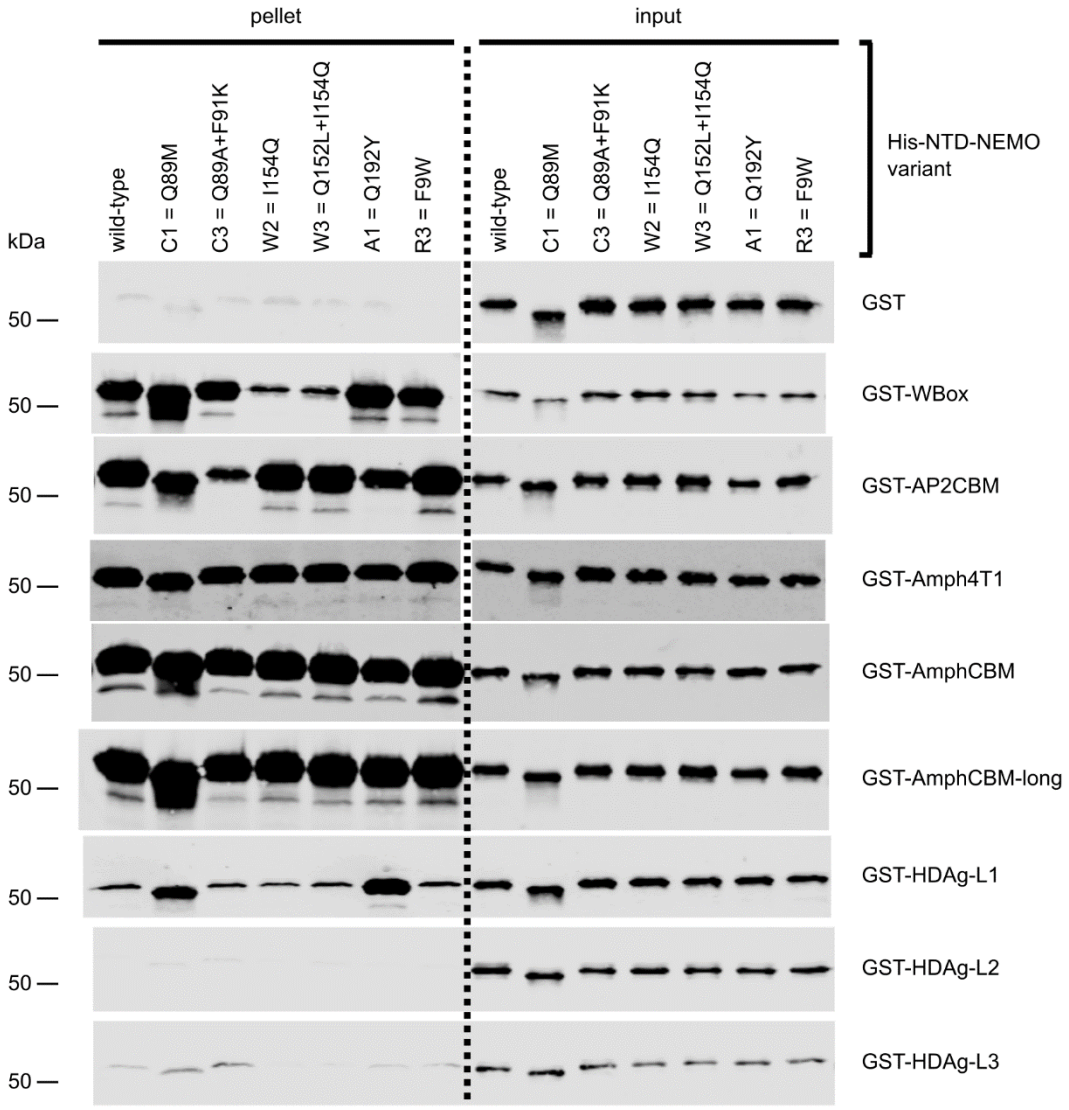


Figure 47 Binding between single site mutants of His-NTD-NEMO and GST fusions of cellular or viral peptides. Pull-downs using GST-tagged bait proteins as bait and the indicated single site mutants of His-NTD-NEMO as prey were performed using the standard protocol. All experiments were performed at least twice and representative blots are shown.

4.3.2 Binding between fusion peptides and compound mutants of NTD

It was next investigated how mutation of more than one binding site of NTD affected binding of peptides in GST pull-downs. None of the compound mutants bound to the negative control GST (Figure 48). In the NTD:AP2CBM$_{pep}$ crystal structure, binding of the AP2CBM peptide to both

the clathrin box and the arrestin box had been observed (Chapter 3). GST pull-downs using GST-AP2CBM and His-NTD-NEMO carrying mutations at these binding sites showed that the peptide was no longer able to bind to NTD mutated at both sites, a result consistent with the crystallographic observation (Figure 49).

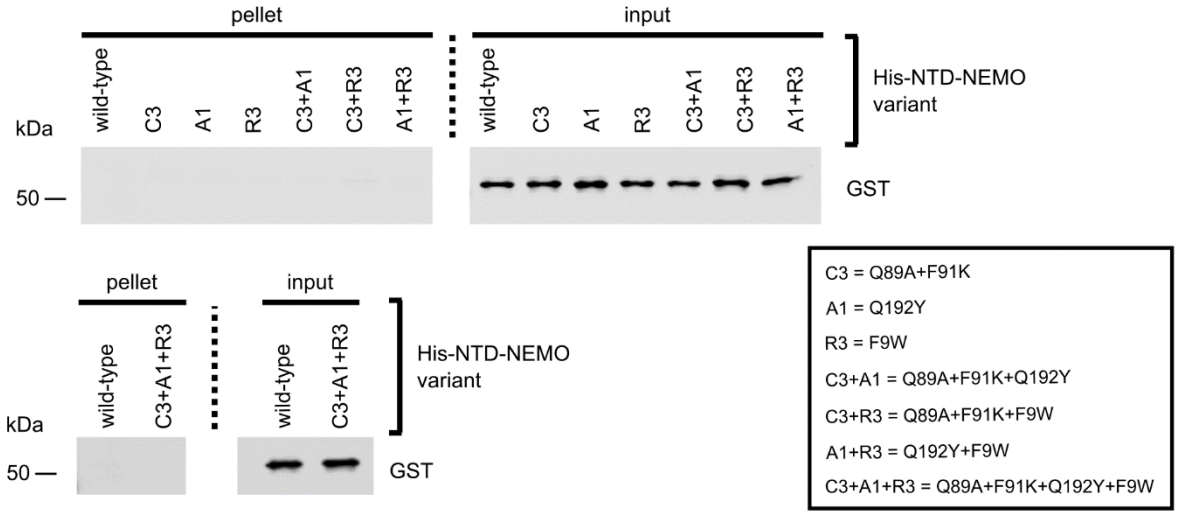


Figure 48 Control experiments assessing background binding between GST and His-NTD-NEMO compound mutants. Pull-downs using GST alone as bait and single, double or triple compound mutants of His-NTD-NEMO as prey were carried out using the standard protocol. Abbreviations of mutations are explained in the boxed inset.

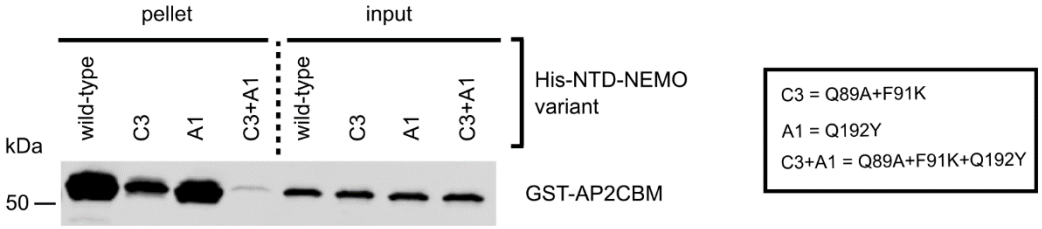


Figure 49 Binding of AP2CBM to His-NTD-NEMO mutated at clathrin box and arrestin box. The pull-down was performed following the standard protocol using the prey proteins indicated. Abbreviations of mutations are explained in the boxed inset.

The Amph4T1 peptide was observed to bind to the clathrin box, the arrestin box and the Royle box in the crystal structures. Simultaneous mutation of the clathrin box and the arrestin box (C3+A1), or the clathrin box and the Royle box (C3+R3), but not the arrestin box and the Royle box (A1+R3), markedly reduced binding. When all three binding sites were mutated (C3+A1+R3), binding was almost completely abolished (Figure 50).

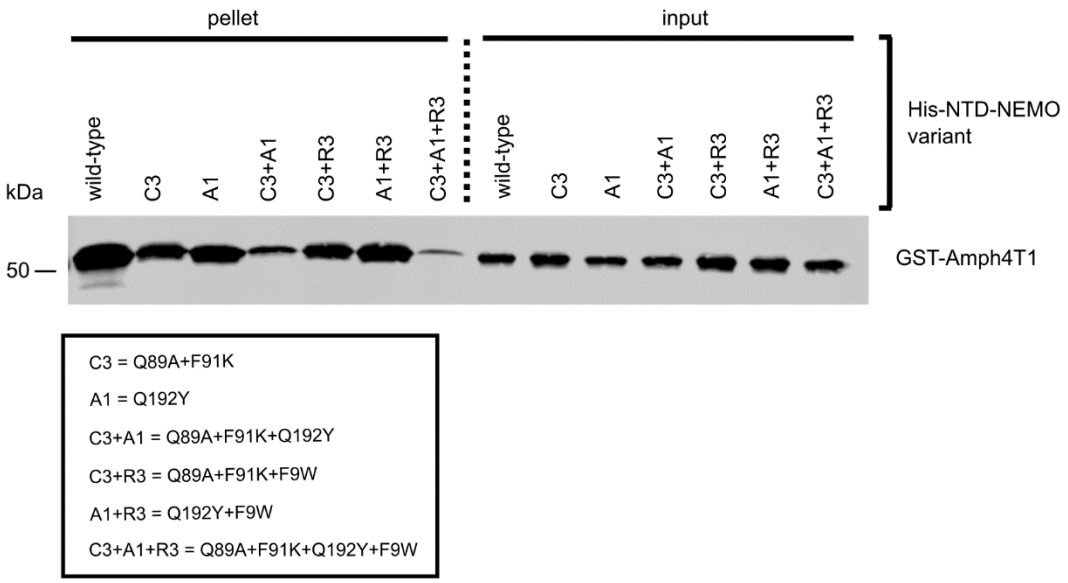


Figure 50 Binding between Amph4T1 and compound mutants of His-NTD-NEMO. Pull-downs using GST-tagged Amph4T1 as bait and single, or double and the triple compound mutants of His-NTD-NEMO as preys were carried out using the standard protocol. Abbreviations of mutations are explained in the boxed inset.

In contrast to the Amph4T1 peptide, which contained two non-natural amino acids at its C terminus that contributed to binding at the Royle box, the AmphCBM peptide was found to only bind to the clathrin and arrestin boxes of NTD in the crystal structures. Pull-downs performed with GST-tagged AmphCBM and compound mutants of His-NTD-NEMO showed that simultaneous mutation of the clathrin box and the arrestin box (C3+A1) reduces binding of the peptide to NTD, whereas NTD mutated at the clathrin box and the Royle box (C3+R3), at the arrestin box and the Royle box (A1+R3), or the triple mutant (C3+A1+R3) bound to AmphCBM with similar affinity as the wild-type prey protein (Figure 51). The nature of the residual binding could not be resolved, but could potentially be attributed to non-specific "stickiness" of the used AmphCBM peptide.

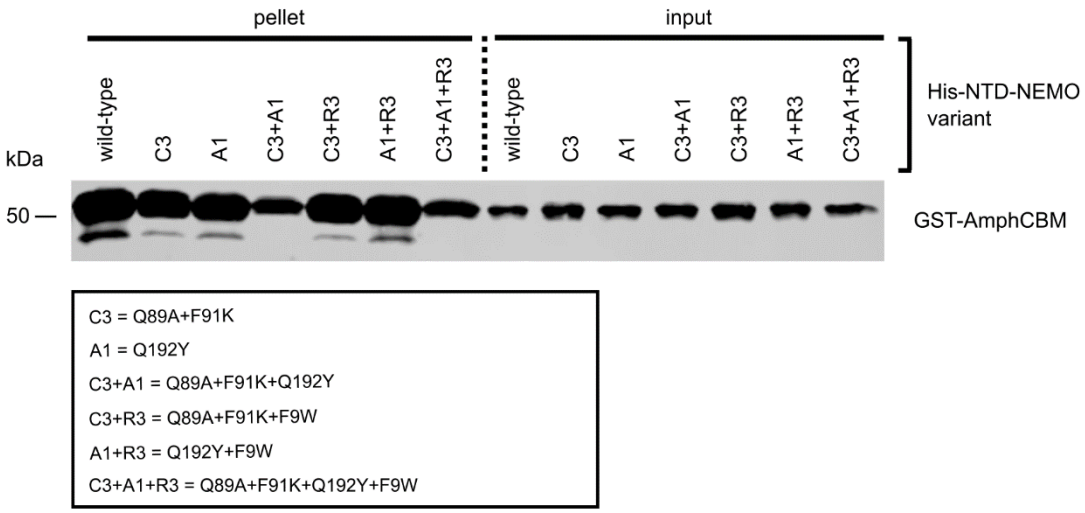


Figure 51 Binding between AmphCBM and compound mutants of His-NTD-NEMO. Pull-downs using GST-AmphCBM as bait and selected single, as well as double or the triple interaction site compound mutants of His-NTD-NEMO as preys were performed using the standard protocol. Abbreviations of mutations are explained in the boxed inset.

The AmphCBM-long peptide, interacting with NTD at the clathrin box and the arrestin box in the crystal structures, was also used as a GST fusion protein in pull-downs with compound mutants of His-NTD-NEMO. However, across numerous repeats of the experiments, equal binding of AmphCBM-long to all compound mutants independent of the introduced mutations was observed (Figure 52). This result suggests that the AmphCBM-long peptide is able to non-specifically bind and potentially, cross-link NTD molecules in non-physiological spatial orientations as observed in the crystal structure.

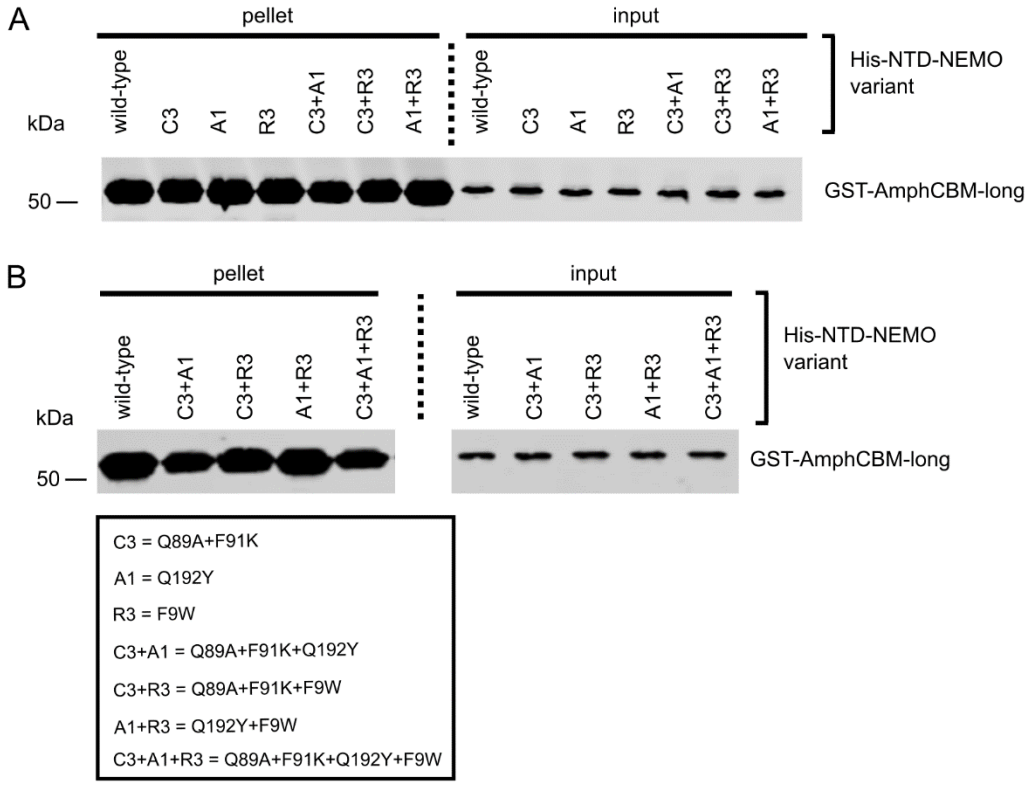


Figure 52 Binding between AmphCBM-long and His-NTD-NEMO compound mutants. Pull-downs were carried out using the standard protocol with GST-tagged AmphCBM-long as bait and (A, B) double or (B) triple binding site compound mutant of His-NTD-NEMO as preys. Abbreviations of mutations are explained in the boxed inset.

Lastly, pull-down experiments with compound mutants of His-NTD-NEMO were also performed with GST-HDAg-L1 as bait. However, as observed during pull-downs with single site NTD mutants (Figure 47), HDAg-L1 exhibited aberrant high affinity for compound mutants involving mutation Q192Y (Figure 53). This peculiar effect and the low binding affinity of HDAg-L1 for wild-type His-NTD-NEMO made it impossible to determine the contribution of the separate binding sites to the overall interaction between NTD and this viral peptide.

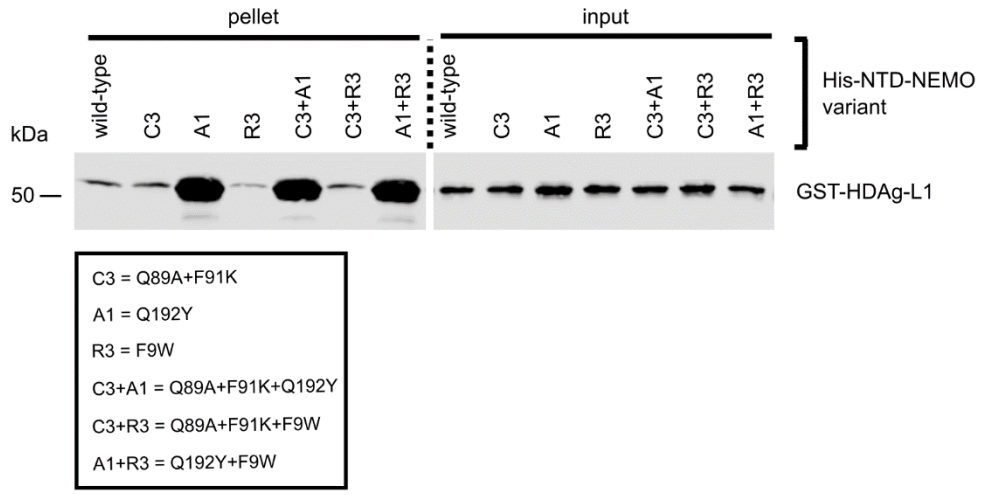


Figure 53 Binding between HDAg-L1 and compound mutants of His-NTD-NEMO. The pull-down was performed using the standard protocol with GST-tagged HDAg-L1 as bait and single or double mutants of His-NTD-NEMO as prey proteins. Abbreviations of mutations are explained in the boxed inset.

4.4 Investigation of clathrin assembly on liposomes

The interaction between cellular adaptor proteins such as AP2 and clathrin is crucial for the recruitment of clathrin to cytosolic membranes and the subsequent formation of clathrin-coated pits (52). As shown in the crystal structures and GST pull-down experiments presented in this thesis, small peptide motifs such as the canonical clathrin binding motif of AP2 or amphiphysin can bind to multiple sites on clathrin N-terminal promiscuously. A GST pull-down demonstrated that cellular peptides derived from amphiphysin or the hinge region of the β2 subunit of AP2 can bind to clathrin purified from pig brain, whereas viral peptides derived from HDV (HDAg-L1 and -L2) do not bind to purified clathrin in this system (Figure 54). Based on this result, it was investigated whether the interactions between cellular and potentially, viral peptide motifs and clathrin are sufficient for the assembly of clathrin on liposomes and for the formation of clathrin-coated pits in vitro.

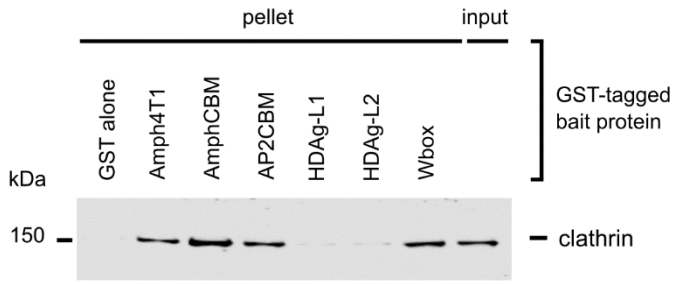


Figure 54 Binding between cellular and viral peptides and clathrin purified from pig brain. The pull-down was performed according to the standard protocol, using 0.4 µM purified clathrin as prey.

Constructs containing viral and cellular peptide motifs fused to an N-terminal hexa-histidine-GST tag were designed (Figure 55). The fusion proteins were expressed overnight in BL21(DE3) pLysS *E. coli* and purified in the same way as the earlier described GST-tagged fusion peptides (section 4.1.1). Liposomes containing a mixture of dioleoyl-phosphocholine, dioleoyl-phosphoethanolamine and a dioleoyl lipid carrying a nickel-chelated nitrilotriacetic acid (Ni-NTA) modified headgroup were prepared and loaded with the His-GST fusion proteins. These liposomes were then incubated with recombinantly expressed and purified NTD or clathrin purified from pig brain. Recruitment of NTD or clathrin was analyzed after sedimenting the liposomes (Figure 56 and Figure 57). The hinge-appendage domain of the β2 subunit of AP2, which has been previously reported to efficiently recruit clathrin onto liposomes (68), was used as a positive control, whereas unloaded liposomes and liposomes loaded with His-GST only served as negative controls.

Efficient binding of the His-GST-tagged peptides to the liposomes was observed, although the fusion peptides were not completely depleted from the incubation mixture (Figure 57A). When purified clathrin NTD was incubated with the liposomes, high non-specific association between the liposomes and NTD was observed (Figure 56). Therefore, recombinantly expressed NTD could not be used to assess binding between peptides and clathrin in this assay. When clathrin purified from pig brain was used as a prey, weak non-specific binding between the liposomes and clathrin was detected, but could be clearly distinguished from binding between clathrin and the positive control His-GST-AP2-HA. Interestingly, all cellular peptides containing a canonical CBM (Amph4T1, AmphCBM, AmphCBM-long and AP2CBM), but not the viral peptides derived from HDV large antigen were able to recruit clathrin onto liposomes (Figure 57). In addition, the W box motif of amphiphysin could not bind to clathrin in this context (Figure

57B), even though in GST pull-down experiments using purified clathrin as prey such an interaction was observed (Figure 54).

Since not only AP2-HA, but also liposome-bound CBM peptides were able to interact with clathrin, it was tested whether the interaction also promoted formation of clathrin buds on the liposomes. Electron microscopy analysis revealed that clathrin buds were only present on liposomes loaded with AP2-HA, not on liposomes loaded with viral peptides or the amphiphysin-derived AmphCBM-long peptide (Figure 58).

Figure 55 Overview of hexa-histidine-tagged GST fusion proteins used in liposome sedimentation assays. Peptides fused to His-GST were derived from the large antigen of HDV (HDAg-L1, -L2, -L3), from AP2 (AP2CBM) and from amphiphysin (Amph4T1, AmphCBM, AmphCBM-long, Wbox). The hinge-appendage domain of the adaptor protein 2 β2 subunit (AP2-HA) fused to His-GST was used as an additional positive control, and His-GST alone served as a negative control.

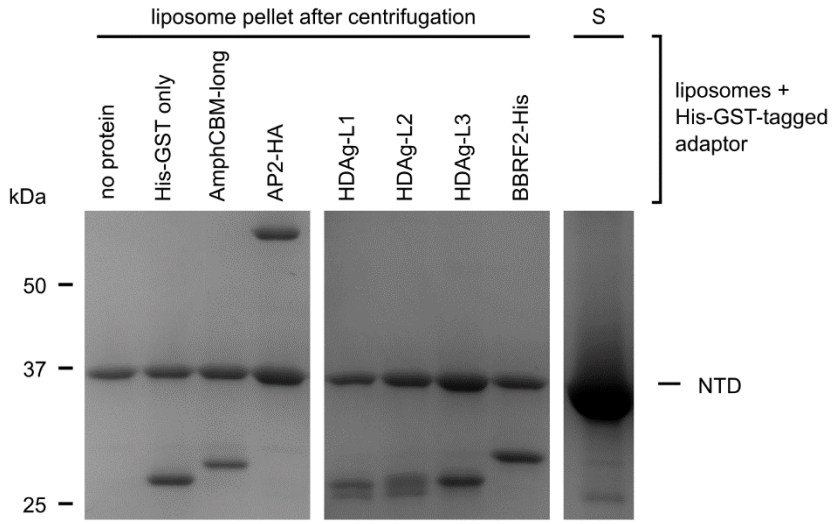


Figure 56 Liposome sedimentation assays assessing the interaction between purified NTD and cellular or viral peptides. Liposomes containing a Ni-NTA-modified dioleoyl-sn-glycero-lipid were loaded with His-GST-tagged peptides and incubated with purified NTD. His-tagged BBRF2 (cytoplasmic envelopment protein 1 from Epstein-Barr virus) was used as an additional negative control. Samples of the liposome pellet (18% of total eluted sample) and of unbound NTD from the His-GST-L1 sample (S, 4% of total supernatant) were analyzed on a 10% SDS-PAGE gel by Coomassie staining.

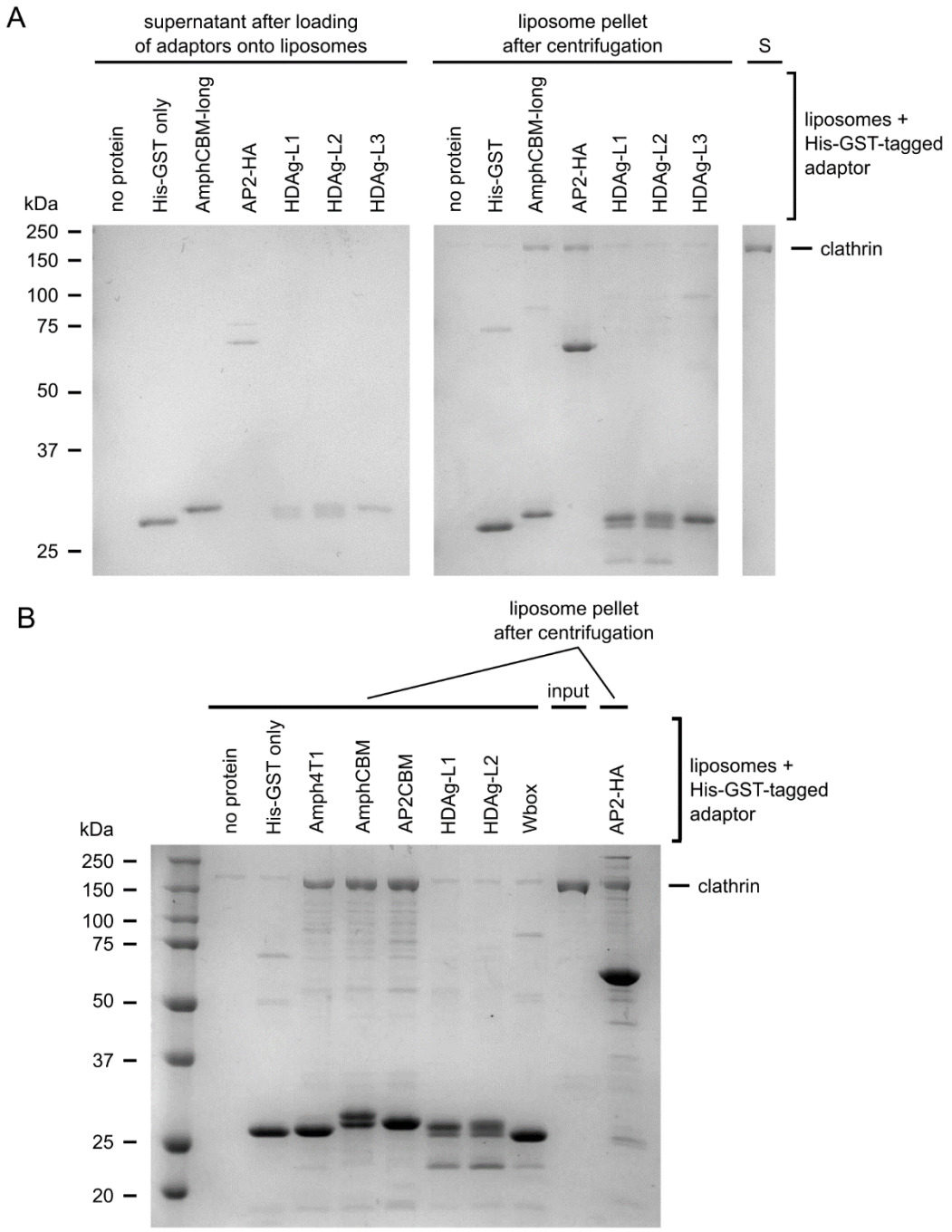


Figure 57 Interaction between liposome-bound peptide motifs and clathrin in liposome sedimentation assays. (A) Liposomes were loaded with His-GST-tagged peptides and incubated with clathrin purified from pig brain. Samples of the supernatant containing unbound peptides (18% of total supernatant), of the liposome pellet after the incubation with clathrin (18% of total eluted sample), and of unbound clathrin from the sample containing unloaded liposomes only (S, 8% of total supernatant) were separated on a 10% SDS-PAGE gel and proteins visualized by Coomassie staining. (B) Liposomes were pre-loaded with His-GST-tagged peptides and subsequently incubated with purified clathrin. Samples of the liposome pellet (18% of total eluate) and the input (9% of total input) were analyzed on a Coomassie-stained SDS-PAGE gel.

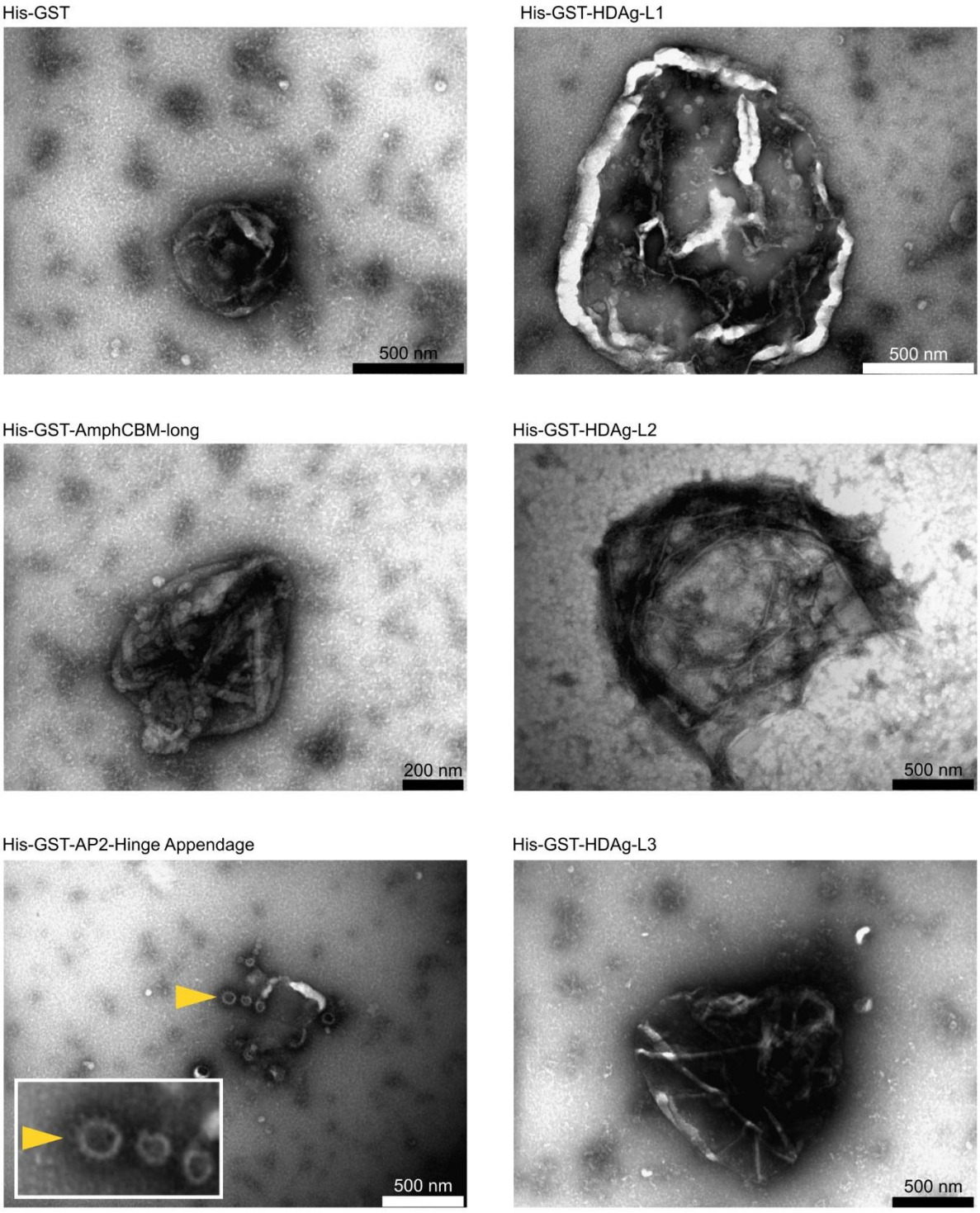


Figure 58 Negative stain electron microscopy analysis of clathrin budding on peptide-loaded liposomes. A liposome sedimentation experiment was performed using His-GST-tagged peptide fusion proteins as immobilized "baits" on liposomes containing a Ni-NTA-modified dioleoyl-sn-glycero-lipid. Samples of the liposome pellet after incubation with purified clathrin were pipetted onto electron microscopy grids and stained with 2% uranyl acetate. The formation of clathrin lattices or buds was then examined by electron microscopy. The yellow arrowhead shows a clathrin-coated bud.

4.5 Exploring biophysical techniques to quantify interactions between NTD and peptides

In GST pull-down assays (sections 4.1.4 and 4.3), high concentrations of bait proteins had to be immobilized onto the agarose beads in order to facilitate efficient capture of His-NTD-NEMO, indicating that the interactions between clathrin NTD and its binding partners are generally weak. Nevertheless, in agreement with the presented structural data, the results of these pull-down experiments showed that CBM peptides can bind to multiple sites on NTD. In order to quantify the contribution of binding to the separate sites to the overall affinity of the interaction, different biophysical assays were explored. Previous estimates of NTD binding affinities differ vastly, ranging from 22 µM (4) to 800-900 µM for CBM-containing peptides (266). Therefore, quantification of the interaction between NTD and viral or cellular peptides could also help to clarify these discrepancies.

4.5.1 Fluorescence polarization

When a fluorophore is excited with plane-polarized light, the emitted light is also polarized. If the fluorophore is not mobile, the absorption and emission dipoles are parallel and 60% of the emitted light stays in the same plane as the excitation light (high anisotropy, A=0.4) (313). If the fluorophore rotates, the dipoles are no longer aligned and less than 60% of the emitted light is found in the excitation polarization plane. In fluorescence polarization experiments, the fluorescent molecule is excited with polarized light and the emitted light is detected in two planes (parallel, $\parallel$, and perpendicular, $\perp$). The light intensities, I, measured in the two planes can be used to calculate the fluorescence anisotropy A (313):

$$A = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}}$$

The quicker a fluorophore tumbles, the less polarized the emitted light is and the lower the anisotropy (A). Therefore, fluorescence polarization can be used to monitor binding between a small fluorescent ligand, for example a labeled peptide, and proteins. A fluorophore-labeled peptide is comparatively small and will rotate freely in solution, thus emitting depolarized light (low A). Upon binding to a protein with a higher molar mass and therefore a larger molecular volume, the protein:peptide complex will rotate much slower and emit polarized light (high A). Measuring the fluorescence anisotropy during titration of a binding partner to a solution

containing a constant concentration of fluorescently labeled ligand (<< K_D) enables calculation of the dissociation constant (K_D) of an interaction.

Advantages of fluorescence polarization assays include the low protein sample consumption, low costs, their suitability for high-throughput setups, and the broad range of K_D values that can be determined. However, autofluorescence of proteins, fluorescence quenching and the need for fluorescent tagging of the ligand limit their use. In addition, ligand and protein must have a large enough size difference to enable detection of changes in anisotropy (313, 314).

Fluorescence polarization experiments were performed using a fluorescein isothiocyanate-labeled amphiphysin CBM peptide (FITC-Amphiphysin) or a FITC-labeled control peptide and wild-type (Figure 59A) or mutant (Figure 59B) monomeric His-NTD. With increasing concentrations of wild-type His-NTD, an increase of the anisotropy was observed, which indicated binding between FITC-Amphiphysin and NTD. When an unrelated control peptide was used, an increase of anisotropy was only observed at higher protein concentrations, suggesting that binding between wild-type His-NTD and FITC-Amphiphysin is at least partially specific in the presence of significant background (Figure 59A). In agreement with the pull-down experiments and structural data, mutation of single binding sites on NTD did not affect the interaction between NTD and the amphiphysin CBM peptide (Figure 59B). However, in all experiments, saturation of the anisotropy was not reached at the protein concentrations used (up to 225 µM), suggesting either a very weak interaction or high non-specific binding between NTD and the peptide. In addition, FITC-Amphiphysin was also found to non-specifically interact with the His-GST control. Due to this non-specific binding of the fluorescent peptides to NTD or His-GST at high protein concentrations and unattainability of binding saturation, fluorescence polarization was deemed unsuitable as a technique to quantify NTD:peptide interactions.

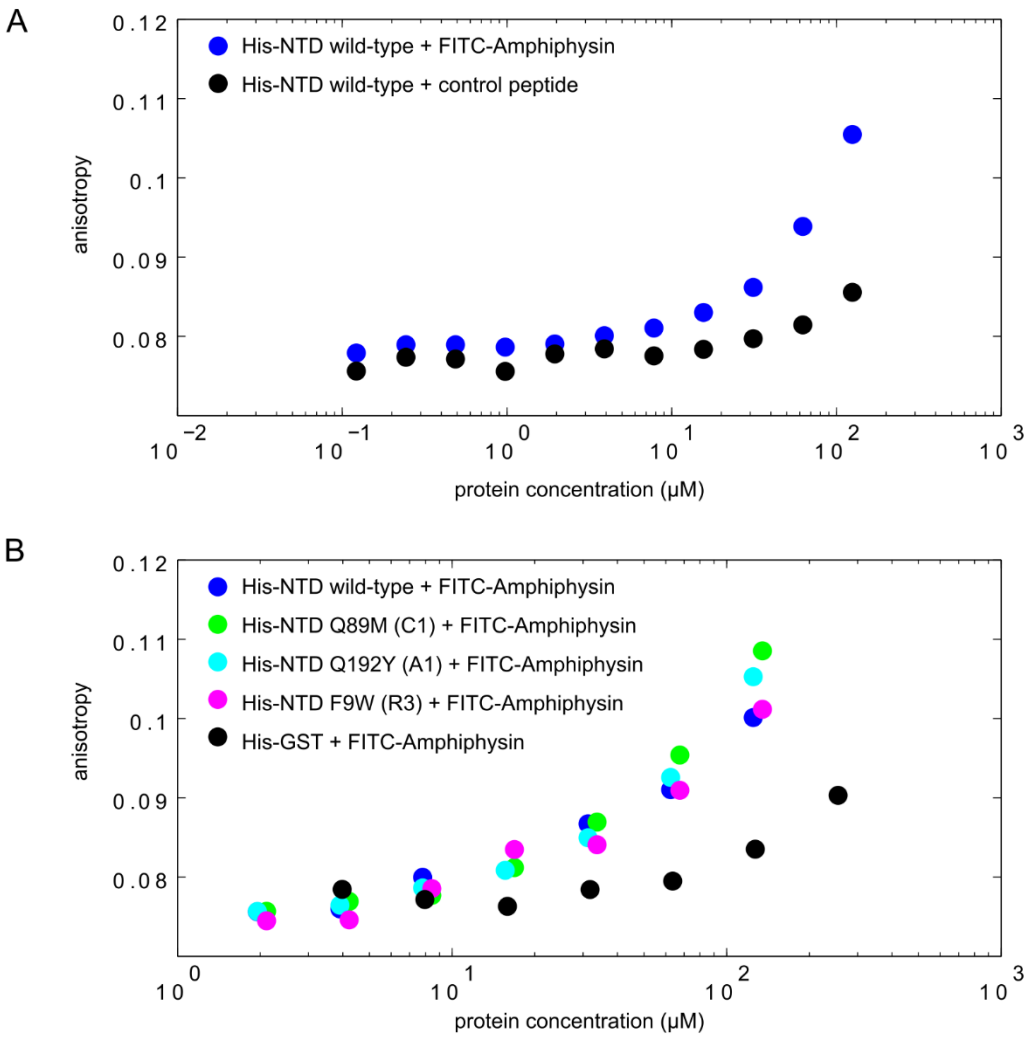


Figure 59 Association between NTD and a fluorescently labeled amphiphysin peptide measured by fluorescence polarization. (A) Increasing concentrations of wild-type His-NTD (up to 225 μM) were incubated with 10 nM of FITC-Amphiphysin (blue dots) or an unrelated FITC-labeled control peptide (black dots) and the fluorescence anisotropy was measured. (B) Wild-type and single site mutants of His-NTD (colored dots, C1 – clathrin box mutant, A1 – arrestin box mutant, R3 – Royle box mutant), or His-GST as a control protein (black dots) were incubated with 10 nM of FITC-Amphiphysin and the fluorescence anisotropy measured.

4.5.2 Differential scanning fluorimetry

In folded proteins, most hydrophobic residues are hidden in the core of the protein. Upon unfolding, hydrophobic residues are exposed to the solvent. Differential scanning fluorimetry (DSF), also known as thermal shift assay or thermofluor, exploits this behavior by monitoring the temperature-induced unfolding of a protein with a fluorescent dye which is quenched in hydrophilic environments, but emits light when bound to hydrophobic parts of a protein. The midpoint of the protein unfolding curve is defined as the melting temperature (T_M) and a measure of protein stability. Thermal shift assays can for example be used to assess the impact of

mutations on the protein fold (as described in section 4.2), to find buffer conditions stabilizing a protein, making it more amenable to crystallization, or to evaluate stabilizing or destabilizing effects of ligands (315). Differential scanning fluorimetry is a cost-effective high-throughput method, does not require labeling of protein or ligands, and consumes a few micrograms of protein per tested condition. However, even though some studies report DSF as a method for determining binding affinities (316), extensive thermodynamic analysis of DSF data is complicated since enthalpic and entropic effects on melting temperature shifts cannot be easily predicted (315).

It was attempted to use DSF to monitor interactions between clathrin NTD and peptides, using shifts of the T_M as a measure for binding. The thermal unfolding curves of wild-type or mutant His-NTD in the presence of increasing concentrations of CBM-containing peptides (HDAg-$L1_{pep}$ or Amph4$T1_{pep}$) were recorded and the T_M of each condition determined (Figure 60 and Figure 61). Control experiments showed that, compared to wild-type or mutant His-NTD without added peptide, the peptides alone exhibited only minimal background fluorescence even at high concentrations (Figure 60A and Figure 61A). Surprisingly, the melting temperature of both wild-type and mutant (F91A) His-NTD decreased dramatically with increasing concentrations of HDAg-$L1_{pep}$, differing by approximately 9 K between samples incubated with 0 mM and 1 mM peptide (Figure 60B). On the contrary, upon incubation with increasing concentrations of Amph4$T1_{pep}$, the T_M of wild-type and the quadruple compound mutant of His-NTD increased (Figure 61B). These opposing trends of the melting temperatures could genuinely be due to differences between the HDAg-L1 and the Amph4T1 peptides, indicating that binding of HDAg-$L1_{pep}$ to NTD destabilizes the protein, whereas binding of Amph4$T1_{pep}$ stabilizes NTD. However, it should be noted that the DSF experiments had to be performed on two platforms using two dyes (SYPRO Orange vs. protein thermal shift dye, both from Life Technology) due to instrument availability. Since it is not known whether the "thermal shift dye" is different from SYPRO Orange and the exact interplay between the environmentally sensitive dyes and hydrophobic residues of peptides and proteins is not understood (315), the results obtained from the two experimental setups cannot be directly compared.

Nonetheless, for both HDAg-$L1_{pep}$ and Amph4$T1_{pep}$, no differences of the melting temperature shift could be observed between wild-type and mutant His-NTD incubated in the presence of increasing concentrations of the respective peptides. In the crystal structures, HDAg-$L1_{pep}$ was

found to bind to both the clathrin box and the Royle box on NTD. It could thus be argued that no differences between wild-type His-NTD and His-NTD F91A were observed because the HDAg-L1 peptide can still bind to NTD via the Royle box even if the clathrin box is mutated at residue Phe91 of NTD. On the other hand, as demonstrated by both structural data and in GST pull-downs, Amph4T1$_{pep}$ bound to three sites on NTD, the clathrin box, the arrestin box and the Royle box. However, even when all four binding sites on NTD were disrupted (quadruple compound mutant, C3A1W3R3), binding of the Amph4T1 peptide was not diminished compared to wild-type His-NTD (Figure 61B). The quadruple mutant of NTD was used for the DSF experiments to ensure complete disruption of all four binding sites, even though this NTD variant is destabilized compared to wild-type NTD (Figure 46). A GST pull-down showed that binding between GST-Amph4T1 and wild-type NTD or the quadruple compound mutant of NTD is indeed abolished when all four sites of NTD are mutated (Figure 62).

These results indicate that in DSF experiments, high concentrations of peptide probably bind to NTD non-specifically and lead to non-specific stabilization and/or destabilization of the protein. Alternatively, contaminations from the peptide synthesis present in the reaction mixtures such as trifluoroacetic acid could interfere with NTD folding. These experimental problems rendered DSF unsuitable for biophysical characterization of NTD:peptide interactions.

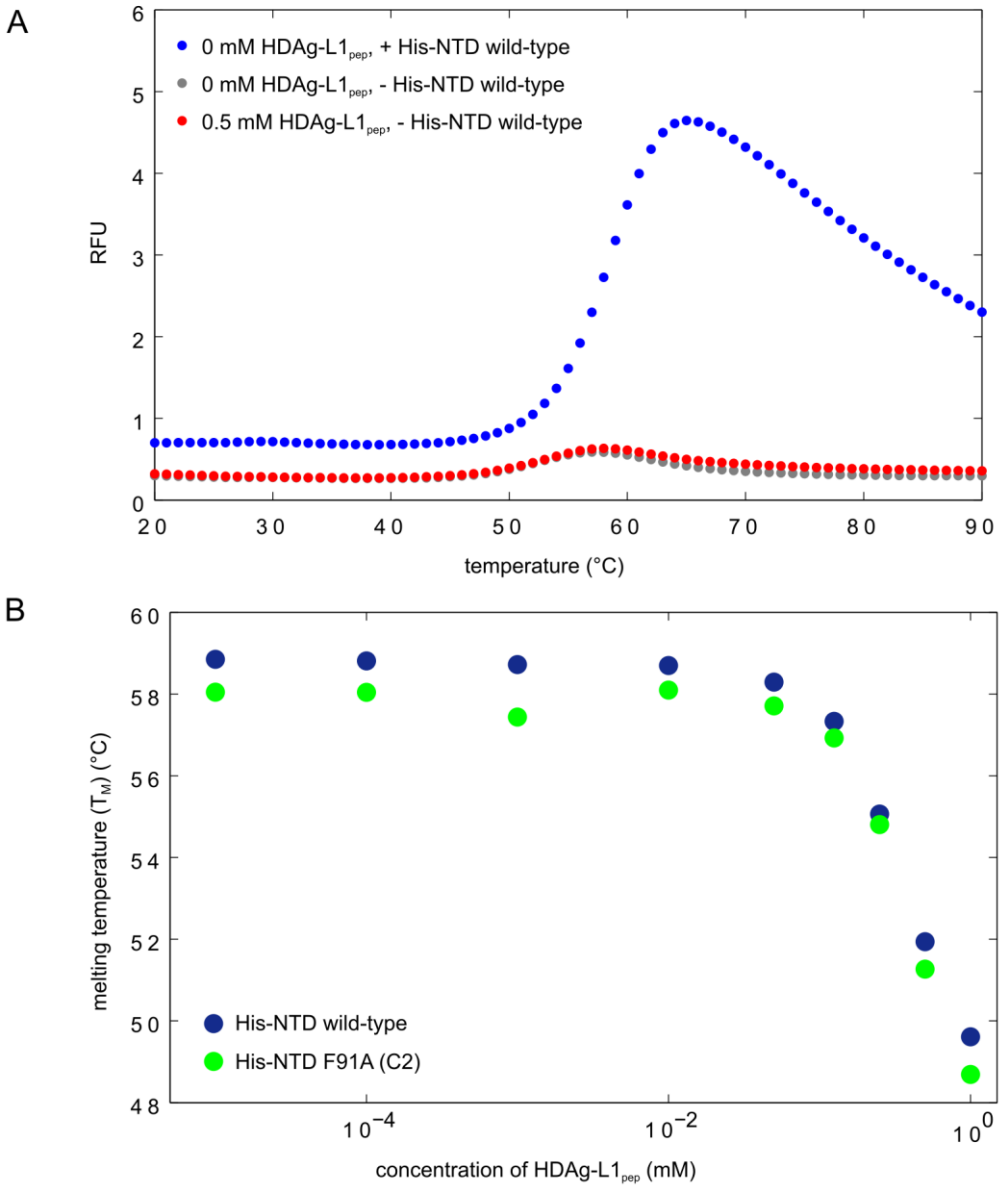


Figure 60 Binding between wild-tpe or mutant His-NTD (F91A) and the HDAg-L1 peptide assessed by differential scanning fluorimetry. (A) Control experiments were carried out to monitor unfolding of wild-type His-NTD (2 mg/mL) in the absence of peptide (blue), and to assess background fluorescence of the HDAg-L1 peptide (red) and SYPRO Orange (grey) without added protein. (B) Wild-type (dark blue dots) or mutated His-NTD (green dots) at 2 mg/mL were incubated with increasing concentrations of HDAg-L1$_{pep}$ and the melting temperatures calculated. All experiments were performed using 10x SYPRO Orange on the MiniOpticon real-time PCR system.

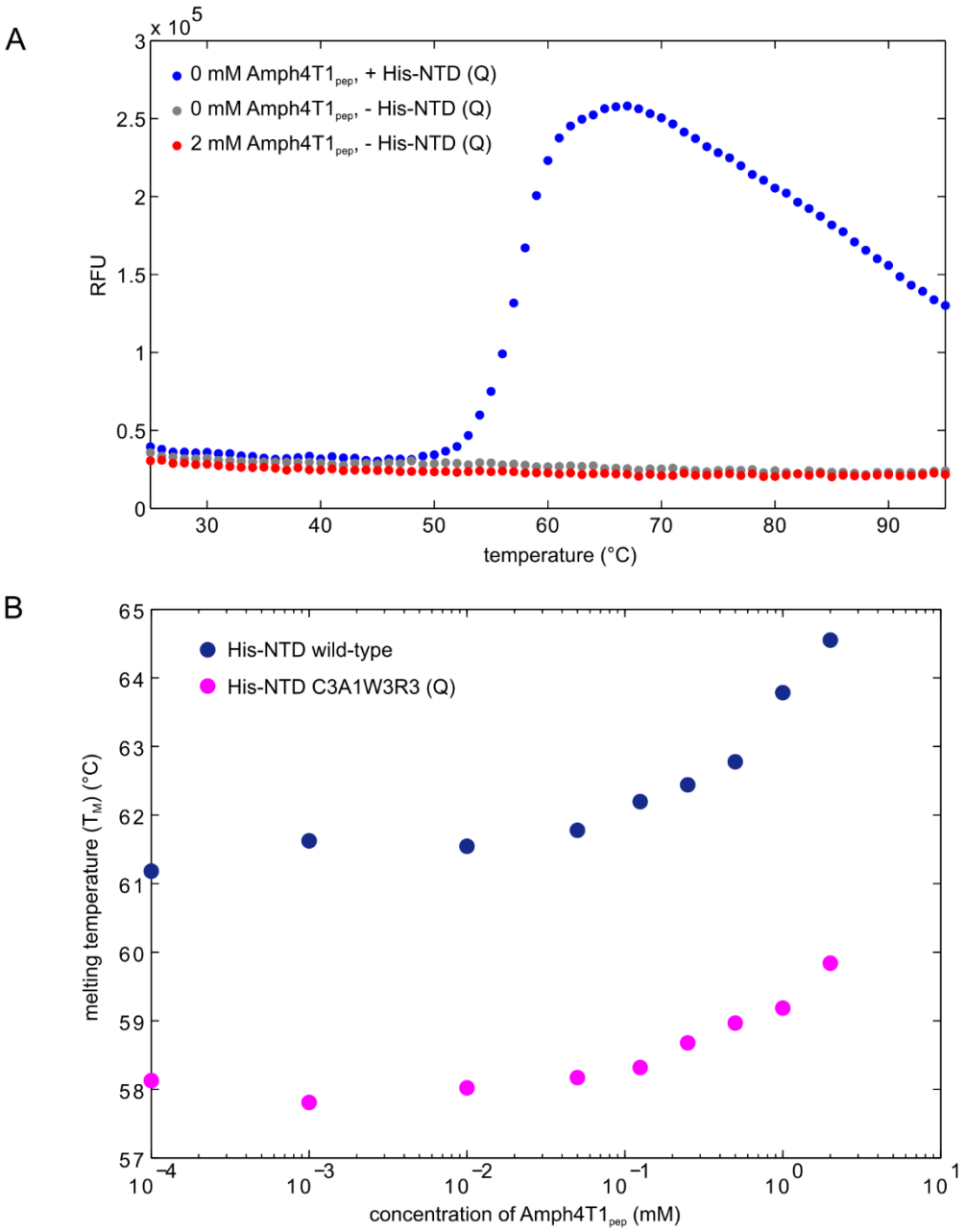


Figure 61 Differential scanning fluorimetry measurements of binding between wild-type or the quadruple compound mutant of His-NTD and a CBM-containing amphiphysin peptide. (A) Control experiments were performed to test unfolding of the quadruple compound mutant of His-NTD (0.1 mg/mL) in the absence of peptide (blue) and to evaluate background fluorescence of the Amph4T1 peptide (red) or the thermal shift dye (grey) in the absence of protein. (B) Wild-type His-NTD (dark blue) or His-NTD mutated at all four binding sites (Q, pink) at 0.1 mg/mL were incubated with increasing concentrations of $Amph4T1_{pep}$, the unfolding curves analyzed and the melting temperatures calculated. The experiments were performed using 1x thermal shift dye on the Viia7 real-time PCR platform.

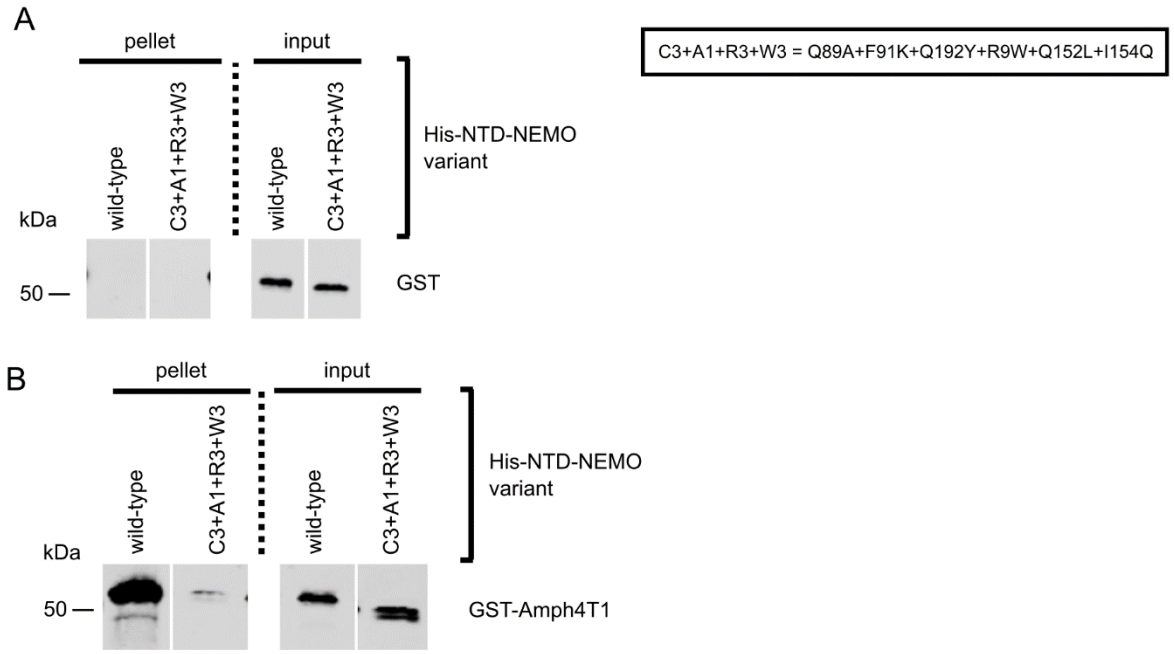


Figure 62 Assessment of binding between GST-Amph4T1 and wild-type or the quadruple mutant of His-NTD-NEMO by GST pull-down. Pull-downs using (A) GST as a negative control as bait or (B) the GST-tagged Amph4T1 peptide were performed using the standard protocol.

4.5.3 Isothermal titration calorimetry

Interactions between proteins and ligands such as small molecules, peptides or other proteins can not only be characterized by the dissociation constant, K_D, but also by thermodynamic parameters such as changes in enthalpy (ΔH_D) and entropy (ΔS_D) of the binding reaction. The Gibbs energy of dissociation, G_D, is dependent on the enthalpy and entropy changes and related to the dissociation constant via

$$\Delta G_D = -RT \cdot \ln K_D = \Delta H_D - T\Delta S_D$$

with T being the absolute temperature and R the gas constant (1.9872 cal/K*mol) (317). Formation or breaking of specific interactions such as hydrogen bonds, salt bridges or van der Waals interactions contribute to enthalpy changes during formation of a protein:ligand complex, whereas entropy changes are related to changes in solvation or conformational freedom (317, 318). Isothermal titration calorimetry (ITC) is a biophysical technique that allows measurement of these thermodynamic parameters and the dissociation constant of binding reactions by directly measuring the heat development during binding.

In a typical ITC experiment, small aliquots of a ligand at a known concentration are injected into a sample cell containing a protein solution at lower, known concentration. The sample cell and a reference cell, held at a constant temperature, are set in an adiabatic chamber. Any heat developed during the exo- or endothermic binding reaction is measured by applying power to the sample cell to keep sample and reference cell at the same temperature. Multiple injections of ligand into the sample cell are carried out until binding is saturated and no further protein:ligand complex can be formed, and thus no further heat development corresponding to the protein:ligand binding can be observed. The area under the detected heat peaks of each injection is integrated and non-linear regression fitting of a model is carried out using the molar ratios of protein and ligand concentrations at each injection to determine K_D and ΔH (317). The Gibbs energy of the reaction and the entropy change of the reaction can then be calculated using the equation above.

Isothermal titration calorimetry is a label-free and immobilization-independent method for the determination of binding constants. The sizes of the participating molecules also do not influence the experiments. With ITC, not only the K_D and stoichiometry, but also the thermodynamic parameters of an interaction can be determined. However, sample throughput is low even in automated setups. In addition, both ligand and protein concentration have to be determined accurately before the experiment to ensure correct calculation of the K_D and the thermodynamic parameters of the interaction. Importantly, the buffers in which protein and ligand are present have to be perfectly matched to avoid development of solvation heat of buffer components (314, 319). Therefore, experimental preparations for ITC can prove challenging.

As a third biophysical method for determining the affinity of peptide ligands for NTD, ITC was attempted. Isothermal titration calorimetry has been used previously to determine the affinity of peptide ligands for NTD (266). Solutions of NTD (0.5 mM, sample cell) and AmphCBM-long$_{pep}$ (15 mM, titration syringe) were prepared in a 50 mM sodium phosphate pH 7.5, 2 mM TCEP buffer, matching the concentrations of protein and peptide used by Zhuo and colleagues (266). A control titration of buffer into the protein solution showed only negligible heat development (Figure 63, top graph). During titration of the AmphCBM-long peptide into the protein solution, binding could be observed, but the reaction seemed to be partly exothermic, partly endothermic, and the protein precipitated partially during the experiment (Figure 63, bottom graph). Therefore, a control titration of peptide into buffer was performed. A high heat development was

observed during the experiment (Figure 63, middle graph), prompting reassessment of the buffer conditions used.

During solubilization of AmphCBM-long$_{pep}$, acidification of the used buffer to pH 6.5 was observed. Since pH differences between the solutions used during titration can result in non-specific heat development, it was attempted to eliminate these pH differences. First, the AmphCBM-long peptide was dissolved in a higher molarity sodium phosphate pH 7.5 buffer (100 mM) to buffer the acidification during peptide solubilization better. However, as with the 50 mM sodium phosphate buffer, a high heat development was detected during titration of the peptide into buffer (Figure 64A). Therefore, as a next step, a different peptide, HDAg-L3$_{pep}$, was solubilized in 25 mM sodium phosphate buffer and the pH of the solution was adjusted manually to 7.5. To prevent non-specific heat development due to mismatches in buffer composition after this pH adjustment, the peptide solution was dialyzed extensively against a 25 mM sodium phosphate buffer pH 7.5. In addition, a third peptide, AmphCBM$_{pep}$, was solubilized in 25 mM sodium phosphate, 50 mM NaCl, matching the buffer used by Zhuo et al. (266), and subsequently dialyzed in that buffer. Nevertheless, even upon titration of these dialyzed peptides into the respective buffers, high heat development was observed (Figure 64B,C). According to Zhuo and colleagues, no such heats of dilution were observed during their ITC experiments with NTD and peptides (266), which could be attributed to the much longer peptide used (26 amino acids in the published vs. 9-12 amino acids in this study). Given the thorough troubleshooting performed with the peptides of interest for this study, it seemed likely that the observed heats of dilution are intrinsic to the peptides used and it was decided that ITC be not pursued further.

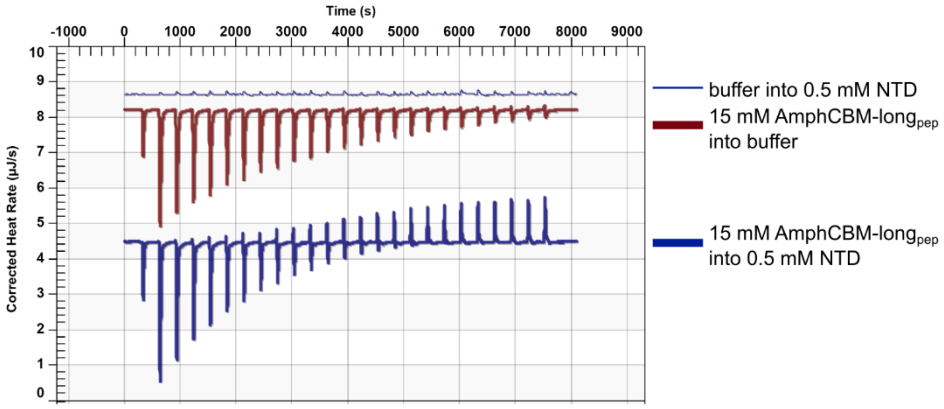


Figure 63 Evaluation of interaction between NTD and AmphCBM-long$_{pep}$ by isothermal titration calorimetry. Experiments were carried out using purified untagged NTD and an amphiphysin-derived peptide (AmphCBM-long$_{pep}$) in a 50 mM sodium phosphate pH 7.5, 2 mM TCEP buffer. Titrations of NTD into buffer (blue), peptide into buffer (bold red) and peptide into protein (bold blue) were performed. For clarity, graphs are shown offset.

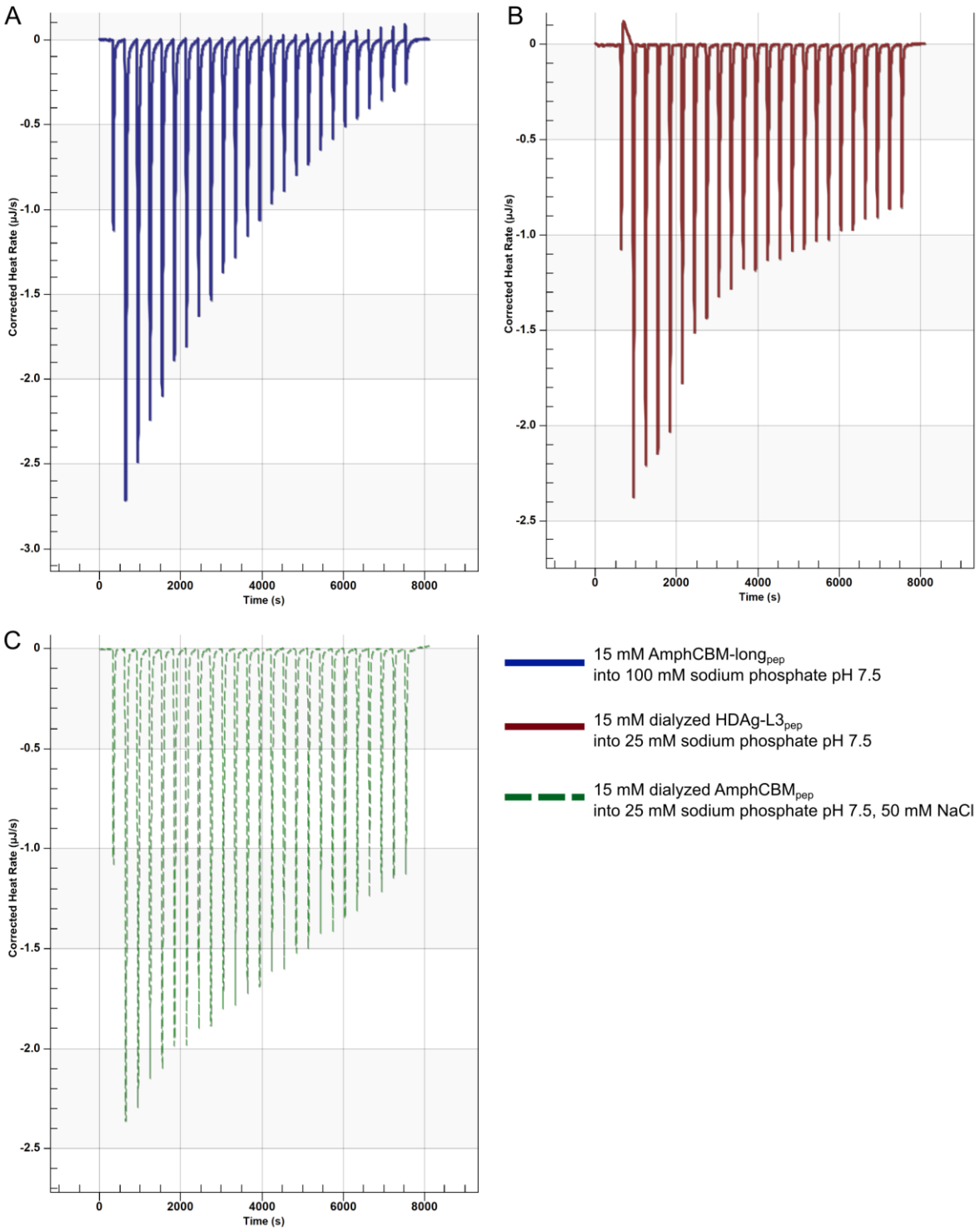


Figure 64 Heat of dilution of peptides titrated into different buffers assessed by ITC. Three peptides (AmphCBM-long$_{pep}$, blue; HDAg-L3$_{pep}$, red; HDAg-L3-long$_{pep}$, dashed green) were titrated into sodium phosphate pH 7.5 buffers of different molarity and ionic strength. The HDAg-L3 peptides were also dialyzed in the respective buffer before the experiments were performed.

4.6 Discussion

Structural analysis of NTD in complex with peptides derived from cellular and viral clathrin-binding proteins revealed that these peptides can bind to multiple sites on NTD simultaneously (Chapter 3). In this chapter, different biochemical and biophysical assays were explored to assess the interaction between NTD and peptides in a non-crystallographic context.

4.6.1 The promiscuous interaction between cellular peptides and NTD binding sites can be validated by a biochemical pull-down assay

Pull-down experiments have been routinely performed to characterize binding between clathrin (61, 68, 79, 81, 106, 308, 320) or purified NTD (1, 4, 81) and adaptor proteins or peptides. In particular, binding between the GST-tagged amphiphysin peptide Amph4T1 and clathrin from rat brain cytosol has been reported (79). Initially, these pull-down results could not be replicated with purified His-NTD-NEMO, a dimerizing construct of the N-terminal domain of clathrin. However, after numerous painstaking attempts of troubleshooting GST pull-down experiments, it was discovered that magnetic glutathione beads interact non-specifically with NTD. In contrast, non-magnetic glutathione agarose beads show no such non-specific binding of NTD. In agreement with previous studies (79, 308), it was demonstrated that high concentrations of GST-tagged bait protein must be immobilized onto the agarose beads for effective binding of the prey. Most likely, the binding between His-NTD-NEMO and peptides in the GST pull-down is supported by avidity effects between the oligomerizing prey protein and the high density of bait proteins. Therefore, a reliable protocol for GST pull-downs using recombinantly expressed and purified His-NTD-NEMO was established.

Multiple GST fusion proteins of cellular amphiphysin and AP2 peptides encompassing CBMs were used as baits in pull-downs and found to reproducibly bind to wild-type His-NTD-NEMO to similar extents. Therefore, pull-down experiments using GST-tagged cellular peptides and mutants of His-NTD-NEMO in which the four interaction sites were systematically disrupted (compound mutants) could be carried out. The results of these pull-downs largely support the observations made in the crystal structures of NTD in complex with the cellular peptides. The CBM peptide from AP2 bound to the clathrin box and the arrestin box in the crystal structures, and NTD mutated at both these sites did not interact with a GST-tagged version of the AP2CBM peptide anymore. Likewise, structural data demonstrated that an amphiphysin-derived peptide

containing two non-native amino acids at its C terminus (Amph4T1) bound to three sites on NTD: the clathrin box, the arrestin box and the Royle box. The pull-downs with the GST-Amph4T1 fusion peptide showed that all three sites contribute to this interaction in a biochemical context since only low levels of binding between Amph4T1 and NTD mutated at all three sites were observed. The importance of NTD residue Phe9, predicted by analysis of the structures of peptides bound to the Royle box (Chapter 3), for interactions at the Royle was confirmed in these assays.

For the two other variants of the amphiphysin peptide lacking the non-natural C-terminal residues, AmphCBM and AmphCBM-long, some non-specific binding was detected during the pull-downs. The contribution of the clathrin and arrestin binding site to the interaction between AmphCBM and NTD could still be demonstrated in the GST pull-downs using compound mutants. However, for AmphCBM-long, non-specific binding was too pronounced and thus, the crystallographic observations could not be assessed with the chosen biochemical context. It is possible that the non-specific are facilitated by the exposed C-terminal hydrophobic residues of the AmphCBM-long peptide (ETLLDLDFDPFK), similar to the bridging observed between NTD molecules in the asymmetric unit in the NTD:AmphCBM-long$_{pep}$ crystal structure.

Interestingly, the structural and biochemical data presented in this thesis directly question a study by von Kleist and colleagues in which two compounds, Pitstop 1 and Pitstop 2, were designed to selectively inhibit binding of ligands at the clathrin box of NTD. Both compounds inhibit the association between clathrin and amphiphysin with IC_{50} values of 18 µM and 12 µM, respectively (274). Crystal structures of NTD in complex with these compounds show that they bind at the clathrin box, at first sight providing a rationale for the observed inhibition of clathrin-mediated endocytosis upon treatment of cells with the compounds. However, these observations are in stark contrast to the results presented here which show that the CBM of amphiphysin can bind to both the clathrin box and arrestin box on NTD. In addition, Willox and Royle demonstrated in functional assays that any binding site of clathrin NTD on its own can mediate the uptake of transferrin, casting further doubt on how a compound specific for one interaction site of clathrin could block endocytosis (2). The work by von Kleist et al. has subsequently been heavily criticized (83) and it has been shown that Pitstop 2 inhibits clathrin-mediated endocytosis non-specifically (321) and also inhibits clathrin-independent endocytosis (322). The work

presented here adds to this critique, supporting the argument that the originally proposed mode of action of Pitstop molecules at the clathrin box of NTD alone cannot explain the disruption of clathrin-mediated endocytosis.

4.6.2 Viral peptide motifs bind to NTD with very low affinity

Pull-downs performed using GST fusion peptides encoding CBMs of amphiphysin, AP2 and the large antigen of HDV showed that all cellular peptides bind NTD equally well, whereas only one of the viral peptides investigated, HDAg-L1, bound reproducibly to NTD, but with a much lower apparent affinity than the cellular peptides. This result stands partially in contrast to the structural observations since both HDAg-L1 and -L2 were found to bind to multiple sites on NTD in the crystal structures. Interestingly, these structures revealed that the viral peptides do not interact with NTD via the predicted CBMs (106, 107), but with an "offset" motif (Chapter 3). In addition, they lack the third leucine residue present in the canonical CBM of e.g. amphiphysin or AP2. These differences in binding between viral and cellular peptides to NTD could contribute to the lower apparent affinity of HDAg-L1 and -L2.

For HDAg-L3, structural and pull-down data were consistent in showing no interaction between this sequence variant of the HDV antigen and NTD. Unfortunately, previous studies describing the binding between HDV large antigen and clathrin did not compare the binding affinity of the viral proteins and cellular clathrin adaptors (106, 107, 177). However, in cell-based assays, Huang and colleagues demonstrated that the interaction between cellular clathrin and HDAg-L1 or HDAg-L2 is necessary for the release of HDV virus-like particles (VLPs) and that mutations of key residues of HDAg-L1 (L199A, D203A, P205A) prevent the release of VLPs (106). In previous studies, pull-down assays were also performed to confirm direct binding between clathrin and HDAg-L (106, 107, 177). However, in these experiments, a truncated construct of NTD containing only residues 1-107, and thus lacking the arrestin box and the Royle box, was used. In addition, since NTD forms a seven-bladed β-propeller, it is unlikely that the truncated construct folds correctly and these biochemical experiments should therefore be viewed with caution. In summary, the structural and biochemical data presented in this thesis suggest that viral CBMs can directly bind to the same sites on NTD as cellular peptides, but with a much lower affinity. Therefore, further studies have to be conducted to elucidate the contribution of

this direct binding and the potential involvement of other cellular clathrin adaptor proteins during maturation and release of HDV.

4.6.3 Characterization of oligomerizing wild-type and mutant NTD constructs

Since binding between NTD and peptides is weak, an oligomerizing construct of NTD was designed by fusing the NEMO oligomerization domain to the C terminus of NTD. Using multi-angle light scattering, it was demonstrated that His-NTD-NEMO forms concentration-dependent oligomers. There have been disagreements over the oligomerization state of NEMO, with earlier studies based on crosslinking experiments suggesting di-, tri- and even tetramerization (323-325), and more recent structural and biochemical work showing dimerization of NEMO (312, 325, 326). The MALS study of a NEMO fusion protein performed in this thesis demonstrates that the NEMO oligomerization domain can serve as a useful "building block" for the design of recombinant proteins, enabling moderate-affinity multimerization of otherwise monomeric proteins in a concentration-dependent manner. For example, in the GST pull-downs performed to assess binding between NTD and peptides, low concentrations of His-NTD-NEMO fusion proteins were used as prey. The enrichment of the His-NTD-NEMO monomers on glutathione beads presumably resulted in high concentrations of His-NTD-NEMO in a local environment, promoting dimerization and supporting binding between NTD and peptides by avidity effects.

Importantly, this study provides a comprehensive characterization of mutations of surface residues at the four binding sites of NTD. Circular dichroism and DSF were used to evaluate the effect of previously published and newly designed NTD mutations (see Table 13) on the overall secondary structure and thermal stability of the protein. Even though the CD spectra were virtually identical for all tested mutants, large differences in thermal stability were observed. These findings suggest that DSF is more sensitive than single-temperature CD measurements for detecting changes in protein stability induced by mutation.

One of the previously published mutations, W164E at the arrestin box, was found to have a highly destabilizing effect on NTD, potentially bringing into question the results of the biochemical experiments with this mutant performed by Kang et al. to confirm the localization of the arrestin box (1). However, both the structures of short linear peptides bound at the arrestin box (Chapter 3) and the results obtained in GST pull-down experiments with a different arrestin

box mutant of NTD (Q192Y) presented in this thesis clearly demonstrate the importance of the arrestin box for NTD:peptide interactions.

In the functional study by Willox et al. which identified the fourth binding site of NTD ("Royle box"), mutations E11K and Q14D+Q16M+N17S at the Royle box disrupted transferrin receptor uptake when combined with mutations of the other three NTD binding sites (2). In the crystal structures, the side chain of Glu11 did not form direct interactions with peptides bound at the Royle box (see Chapter 3), but hydrogen bonds between the backbone atoms of Glu11 and the HDAg-L1 and Amph4T1 peptide backbones were observed. Since mutation E11K reduced the melting temperature of NTD by 3.7 K, it is conceivable that destabilization of the local protein fold is the reason for disrupted binding at the Royle box in vivo. However, no destabilization compared to wild-type NTD was observed for the Q14D+Q16M+N17S Royle box mutation. One hypothesis to explain this apparent discrepancy is that, in cells, bona fide Royle box interaction motifs bind a larger region of NTD that includes residues 14-17. All other previously published NTD mutations (Q89M, F91A, Q152L, I154Q, Q152L+I154Q) did not have considerable effects on the thermal stability of NTD, validating their suitability for biochemical or functional analyses.

In summary, the biochemical characterization of a panel of mutated NTD proteins enabled the selection of stably folded mutants for pull-down experiments. In addition, previous work in which mutated NTD was used (1, 2, 4, 77) can now be more clearly evaluated and future studies using NTD mutated at one or more interaction sites can be informed by the thermal stability data presented in this chapter.

4.6.4 Isolated clathrin-binding motifs do not promote clathrin budding in vitro

It has been previously demonstrated that the hinge-appendage domain of the β2 subunit of the clathrin adaptor protein AP2 is sufficient to promote formation of clathrin buds on liposomes (68). The hinge region, containing a canonical CBM, and the appendage of AP2 can bind to clathrin independently, but both interactions together are required for strong clathrin binding (60, 61, 63). In this thesis, it was demonstrated that isolated CBMs, including the one derived from the hinge of AP2, do not only bind to NTD but also to purified clathrin in GST pull-downs. Therefore, it was investigated whether a single CBM is sufficient for clathrin budding in vitro. Using liposomes loaded with CBM-containing peptides, it was shown that the canonical CBMs

of amphiphysin and of the AP2 β2 subunit hinge region can recruit clathrin onto liposomes. As in the GST pull-downs, the CBMs of large antigens of different HDV genotypes did not bind clathrin in this assay. A liposome-immobilized peptide containing the W box motif of amphiphysin was also not able to interact with clathrin in the sedimentation assays, even though it was demonstrated to bind purified clathrin in GST pull-down experiments. It is possible that the W box peptide folds back onto or interacts non-specifically with the liposomes, but not agarose beads, providing a potential explanation for the discrepancy in clathrin binding observed between the different assays.

Despite purified clathrin bound to liposomes loaded with peptides containing cellular CBMs, EM revealed that only the complete β2 hinge-appendage domain of AP2 facilitated budding, whereas HDAg-L1, -L2, -L3, or the CBM of amphiphysin were not sufficient for budding of clathrin-coated vesicles from the liposomes. In summary, these results suggest that canonical CBMs, though binding to clathrin in different biochemical assays, are not able to promote clathrin budding on liposomes in vitro. However, it seems likely that proteins containing multiple simultaneous clathrin-binding epitopes can promote clathrin cage assembly. These findings are consistent with the model of clathrin budding suggested by Kelly et al. (68), who propose that clathrin-coated pits are formed after AP2 is recruited to phosphatidylinositol 4,5-bisphosphate-containing cellular membranes and transitions to an "open" conformation in which both the hinge and appendage region of the β2 subunit, and thus two clathrin-interacting epitopes, are accessible for binding of clathrin.

4.6.5 Binding between NTD and short peptides is very weak and cannot be quantified by various biophysical techniques

Interactions between clathrin NTD and other clathrin coat components are thought to be generally weak (54). In this chapter, the contribution of multiple binding sites to the interaction of NTD with peptides was demonstrated in GST pull-downs. However, even the most careful handling of samples during the execution of pull-down experiments cannot completely prevent the loss of beads. Therefore, the evaluation of relative NTD binding affinities by this biochemical method can only be qualitative and plenty of repeat experiments are required to gain confidence in the observed differences in binding. Measuring NTD binding affinities by biophysical techniques could overcome these limitations of biochemical experiments. So far, ITC (266),

nuclear magnetic resonance chemical shift analysis (266, 327), analytical ultracentrifugation (327) and a fluorometric assay detecting changes in fluorescence intensity of peptides upon binding to NTD (4) have been used to quantify the affinity of peptides to NTD. However, the binding affinities of CBM peptides reported by these studies differ considerably with a K_D of 22 µM determined by fluorometry (4), approximately 250 µM by analytical ultracentrifugation (327), 470 µM by ITC (266) and 200-900 µM by chemical shift analysis (266, 327).

In this work, three biophysical techniques, fluorescence polarization, differential scanning fluorimetry and isothermal titration calorimetry were employed to study the interaction of short peptides with NTD. Unfortunately, in FP and DSF experiments, non-specific binding was observed at the high concentrations of protein or peptide needed to measure weak affinities with these methods. Isothermal titration calorimetry experiments could not be performed either since even extensively dialyzed peptides showed high heat of dilution development upon titration into buffer. Interestingly, Zhuo and colleagues did not report such difficulties in their experiments measuring the affinity of a peptide containing 26 amino acids for NTD (266). The discrepancy between the ITC experiments performed in this study and the previously published one was observed even when experimental conditions were fully matched and could thus not be resolved. Taken together, the binding affinities of short peptides for NTD still have to be elucidated.

4.6.6 Summary

In the previous chapter, I showed that in crystallo, different viral and cellular peptides can bind to multiple sites on NTD. The data obtained from the crystal structures prompted me to re-visit the current dogma that one specific peptide motif binds to exactly one site on clathrin NTD. I thus decided to investigate interactions between NTD and peptides further by biochemical and biophysical techniques. In this chapter, I described the development of a robust GST pull-down assay which was used to compare the binding affinities of cellular and viral peptides for NTD. The results from these experiments indicate that peptides derived from amphiphysin and AP2 bind to NTD with low affinity, whereas the interaction between HDV peptides and NTD is too weak to be detected reliably. I investigated a range of different biophysical techniques in an attempt to quantify the NTD:peptide interactions, but due to the low affinity of the interactions or unresolvable experimental artifacts, I was not to able obtain dissociation constants or other binding parameters. Nevertheless, using a panel of NTD proteins in which all four binding sites

were systematically mutated, either alone or in combination, I next could probe the contribution of binding at the distinct peptide binding sites to the overall interaction. Importantly, the results from pull-downs with these mutant proteins substantiate the structural data, showing that peptide interactions at the clathrin box, the arrestin box and, for one peptide, also at the Royle box, all contribute to NTD binding in vitro. Furthermore, I demonstrated that peptides containing a single, isolated canonical clathrin-binding motif can recruit purified full-length clathrin both onto agarose beads and liposomes, but are not sufficient for budding of clathrin-coated vesicles from liposomes. Finally, I here present the first comprehensive characterization of mutations of binding sites of clathrin NTD. I identify mutations which destabilize the protein, and describe mutations which do not affect protein folding but efficiently disrupt binding at each of the interaction sites, therefore providing a valuable tool for further in vitro and in vivo studies of clathrin:adaptor protein binding.

5 CHARACTERIZATION OF HERPES SIMPLEX VIRUS I PROTEINS FOR UNBIASED INTERACTION SCREENS

Herpes simplex virus 1 acquires its envelope by budding into *trans*-Golgi network membranes, relying on cellular machinery to facilitate this budding event and subsequently direct trafficking of virus-containing vesicles to the cell surface. Many HSV-1 tegument and envelope proteins have been implicated in the final stages of the virus life cycle, but few interactions between these viral proteins and their cellular counterparts have been described (210, 214, 328). In addition, the structures of only a handful of HSV-1 proteins are known and thus, no structural framework exists which could be used for designing specific mutations to probe interactions of these multi-functional proteins. Therefore, the identification and structural characterization of novel protein:protein interactions between viral and host cell proteins is of great importance to our understanding of secondary envelopment and egress of HSV-1 particles.

In this chapter, I describe how I selected three HSV-1 proteins, UL21, UL56 and the glycoprotein gE, from a subset of tegument and envelope proteins as promising candidates for structural studies and unbiased screens for novel cellular interaction partners (section 5.1). Attempts to recombinantly express and purify these proteins, as well as exploratory biophysical characterizations of UL21 are also presented (section 5.2).

5.1 Selection of herpes simplex virus I proteins involved in viral egress

Viral proteins involved in budding of capsids into cellular organelles or associated with trafficking of virus particle-containing vesicles are likely to be either localized in the cytosol or, for transmembrane proteins, possess a cytosolic domain to directly interact with host cell proteins. Topologically, the cytosol is equivalent to the inside of the virion (Figure 65). Therefore, of the 19 envelope and 26 tegument proteins encoded by the HSV-1 genome (196), a subset of eight proteins was chosen, all of which possessed a sufficiently sized intravirion domain and had been previously implicated in viral egress or trafficking. Of the eight proteins, gE, US9 and UL56 are membrane proteins, protein UL11 is a tegument proteins reported to be associated with the viral envelope via palmitoyl and myristoyl groups (329, 330), and proteins UL14, UL16, UL21 and UL49 are tegument proteins (210) (Figure 65).

In order to find three candidates to take forward for interaction screening and structural studies, a literature review on the role of the proteins in viral budding, egress and/or spread was performed (Table 14). Multiple bioinformatic tools were employed to predict secondary structure elements and overall disorder of the proteins. Programs used were TMHMM and MEMSAT3 for the prediction of membrane protein topology and of the range of residues spanning transmembrane domains, RONN and DISOPRED3 for the prediction of protein disorder, as well as NetSurfP and PSIPRED for the prediction of protein secondary structure. The results of these bioinformatic analyses are summarized in Table 14 and exemplary disorder and secondary structure predictions are shown in Figure 66, Figure 67 and Figure 68.

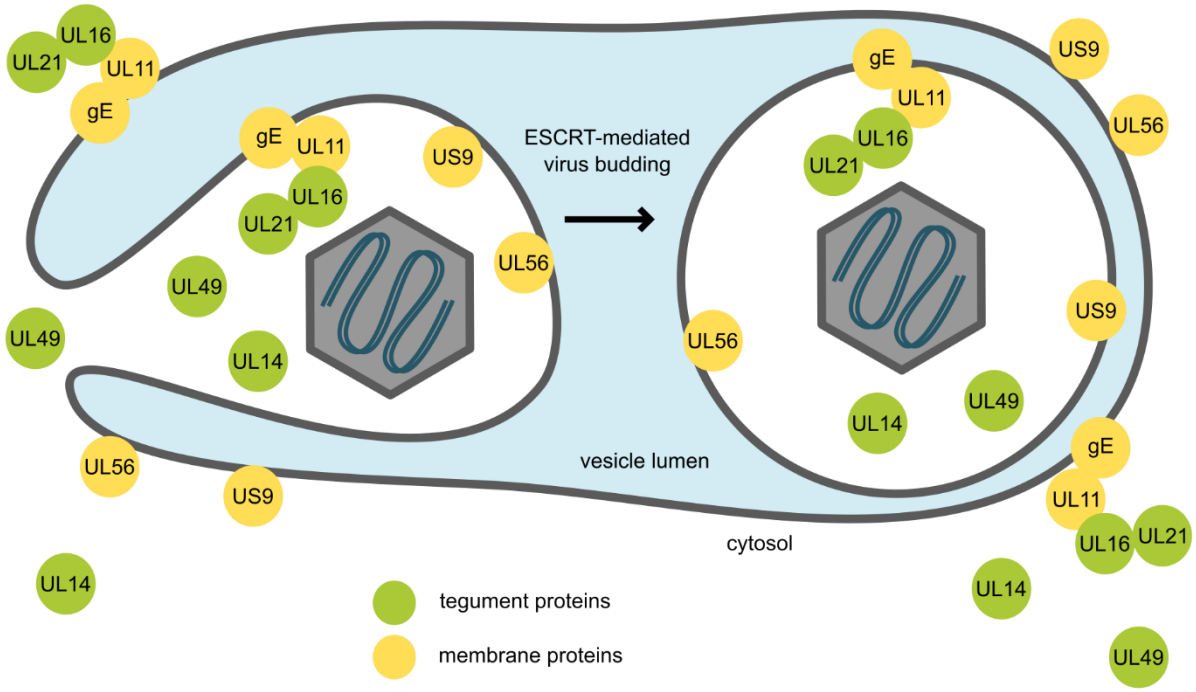


Figure 65 Schematic of an HSV-1 capsid budding into an intracellular vesicle. Proteins implicated in virus egress that were selected for further investigation are represented as circles. Localization of proteins inside the virion (right) is topologically equivalent to cytosolic localization (left).

The results of the literature review, summarized in Table 14, revealed extensive evidence for the involvement of proteins gE and US9 in anterograde axonal transport and sorting of virus particles, or their components such as glycoproteins, from neuronal cell bodies into proximal axons (245, 246, 248, 329, 331-334). In addition, gE is also implicated in cell-to-cell spread in epithelial cells (243, 244). Protein UL56 appears to be linked to cellular ubiquitination pathways and the endosomal sorting complexes required for transport (ESCRT) machinery via its

interaction with E3 ubiquitin ligases (220-222). The ESCRT pathway is exploited by a range of enveloped viruses to facilitate fission of membranes during virus budding (36). Therefore, gE and UL56 were selected for further characterization based on the evidence in the literature of their importance for viral envelopment and egress (see also Chapter 1 for more detail).

Both gE and UL56 are HSV-1 envelope proteins with single-pass transmembrane domains predicted to span residues 422-444 and 211-233, respectively. The cytoplasmic domains of gE (residues 445-550) and UL56 (residues 1-210) were both predicted to be largely disordered (Figure 66 and Figure 67). To increase the chances of performing biochemical and structural studies on one of the candidate proteins in parallel to the unbiased protein:protein interaction screen, the selection of the third candidate protein was based on secondary structure and disorder predictions. Ideally, a protein suitable for structural studies would contain folded secondary structure elements and/or low predicted disorder. Of the eight proteins assessed, UL16 and UL21 exhibited the lowest estimated overall disorder with 30% and 20%, respectively. Protein UL16 contains 20 cysteine residues which could form aberrant disulfide bridges during protein expression, making it a potentially difficult target for structural studies. In comparison, UL21 contains only six cysteines and is predicted to consist of an N-terminal domain containing mostly β-sheets and a C-terminal α-helical domain separated by a short central, unstructured linker (Figure 68). At the time, the structure of UL21 was unknown, so UL21 was selected as the third candidate protein for the unbiased interaction screens and structural studies.

Table 14 Predicted structural elements and functional roles of HSV-1 proteins likely to mediate virus assembly and/or egress, as of June 2014. Some of studies investigated HSV-1 protein orthologs in pseudorabies virus (PRV), a neurotropic alphaherpesvirus closely related to HSV-1, or HSV-1 protein orthologs in herpes simplex virus 2 (HSV-2).

Integral membrane proteins and membrane-associated proteins			
Protein	Structural predictions	Implication in HSV-1 egress	References
US9	disordered N terminus, ordered C terminus partially (40%) disordered	- role in anterograde neural transport of viral particles and spread from pre- to postsynaptic neurons; sorts virus particles from soma into axons (PRV) - interacts with KIF1a (motor protein involved in anterograde axonal transport) and thus mediates virus particle transport in axons (PRV) - important for sorting and transport of glycoproteins and virions into axons (along with gE)	(329, 331-333) (247) (245, 246)
gE	disordered cytoplasmic tail	- mediates sorting of virions to cell junctions and facilitates cell-to-cell spread - involved in anterograde axonal transport of virus particles - necessary for efficient KIF1a:US9 co-purification - important for sorting and transport of glycoproteins and virions into axons (along with US9)	(243, 244) (248, 334) (248) (245, 246)
UL11	central β-sheets, largely (60%) disordered	- UL11 deletion virus shows strong decrease in infectious titer and increase of naked capsids in cytoplasm - UL11 deletion virus has small plaque phenotype - capsids and distorted cytosolic membranes accumulate in cells infected with UL11 deletion virus (PRV)	(335) (336) (337)
UL56	largely (80%) disordered, little regular secondary structure aside from C-terminal transmembrane domain	- interacts with KIF1a (HSV-2) - interacts with UL11 (HSV-2) - localizes to *trans*-Golgi network (HSV-2) - interacts with Nedd4, Nedd4-L, Itch and other E3 ubiquitin ligases (link to ubiquitination pathways and ESCRT machinery) (HSV-2)	(233) (230) (220) (220-222)
Tegument proteins			
Protein	Structural predictions	Implication in HSV-1 egress	References
UL14	disordered C-terminus, N-terminus appears α-helical, overall partially (40%) disordered	- UL14 deletion mutant shows small plaque phenotype and infected cells accumulate naked capsids in cytoplasm	(338)
UL16	short disordered parts in N and C terminus, contains many cysteine residues, partially (30%) disordered	- co-purifies with Rich1 (host protein associated with vesicle trafficking) - deletion of the UL16 gene results in 90% reduction of infectious virion production following infection - interacts directly with UL11 - interacts directly with UL21 - cells infected with UL16 deletion virus contain virions with multiple capsids	(339) (340) (341) (264) (342)
UL21	short disordered central region, otherwise ordered (80%)	- required for transneuronal spread of virus (PRV) - co-localizes with microtubules - interacts directly with UL16	(260) (262) (264)
UL49	largely (80%) disordered	- implicated in spread of HSV-1	(343)

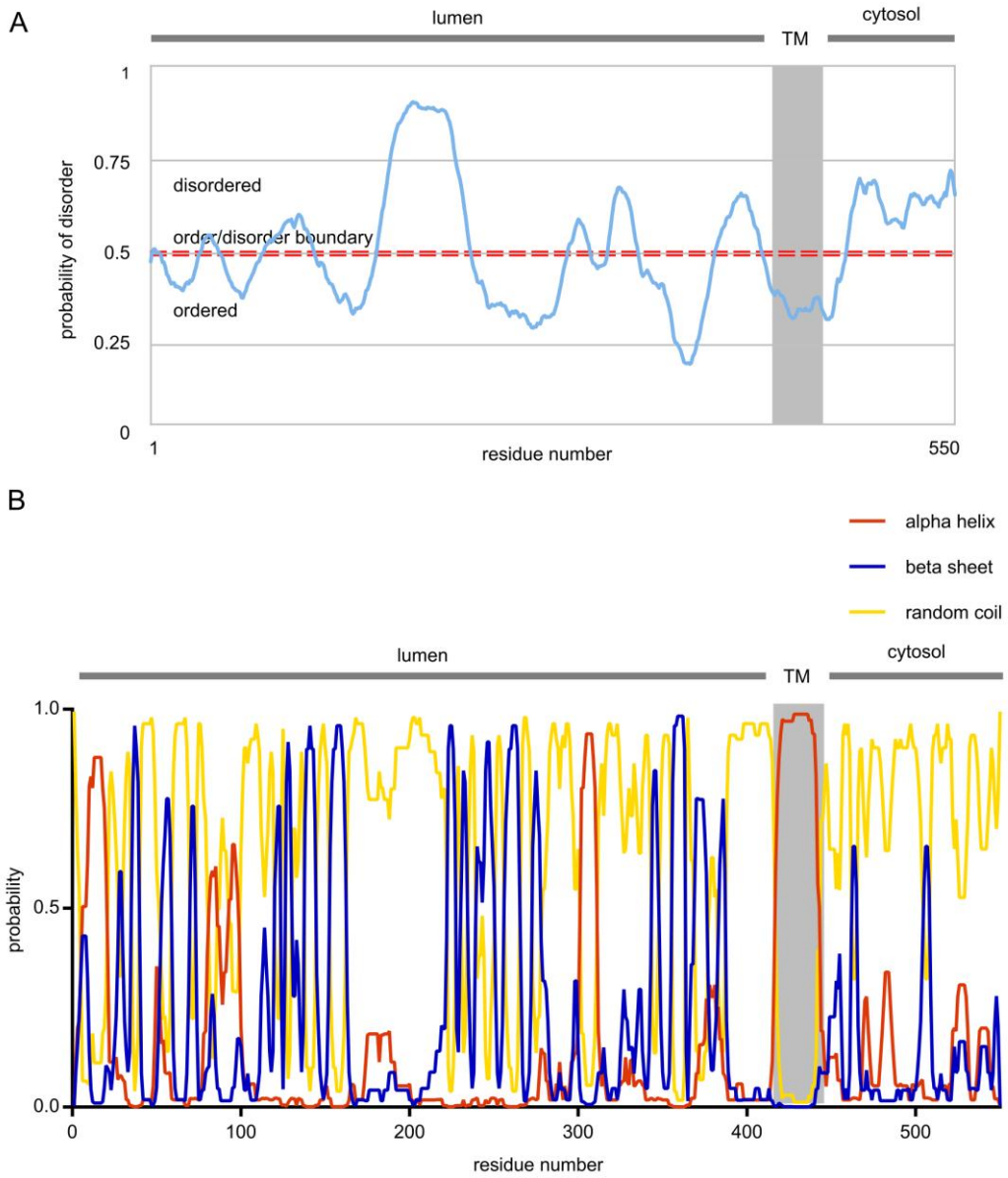


Figure 66 (A) Predicted disorder and (B) secondary structure of glycoprotein gE (550 amino acids, from HSV-1 strain KOS, Uniprot ID D3YPM6). The transmembrane domain (TM) of gE is formed by residues 422-444 (shaded grey, calculated by TMHMM). All residues N-terminal of this domain are extracellular and residues 445-550 form the intracellular tail of the protein. Protein disorder and secondary structure were predicted using RONN and NetSurfP, respectively.

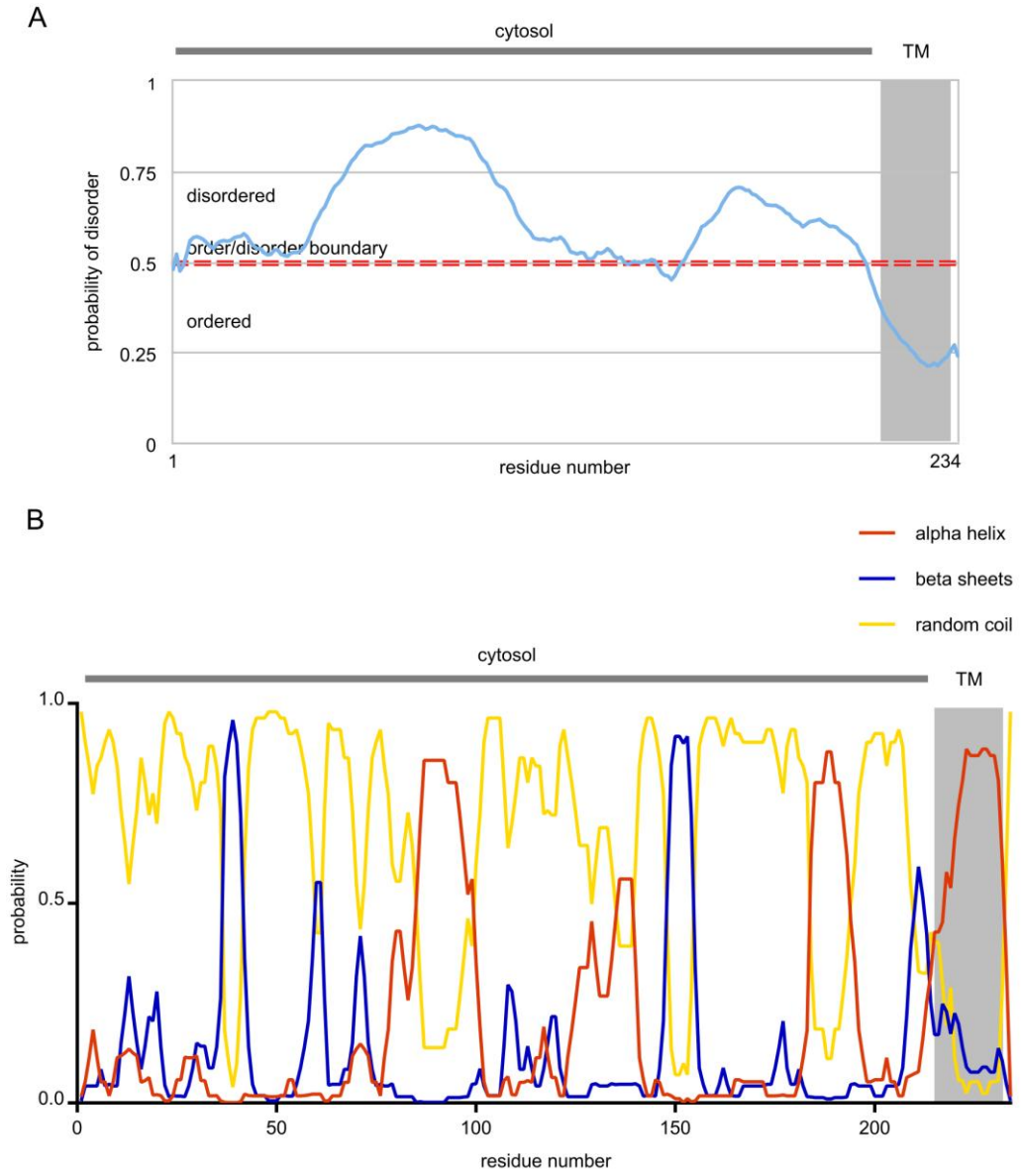


Figure 67 (A) Predicted disorder and (B) secondary structure of UL56 (234 amino acids, from HSV-1 strain KOS, Uniprot ID H9E964). The transmembrane domain (TM9 of UL56 is located at its very C terminus between residues 211-233 (shaded grey, calculated by TMHMM), with residues 1-210 forming the cytosolic domain. Protein disorder and secondary structure were predicted using RONN and NetSurfP, respectively.

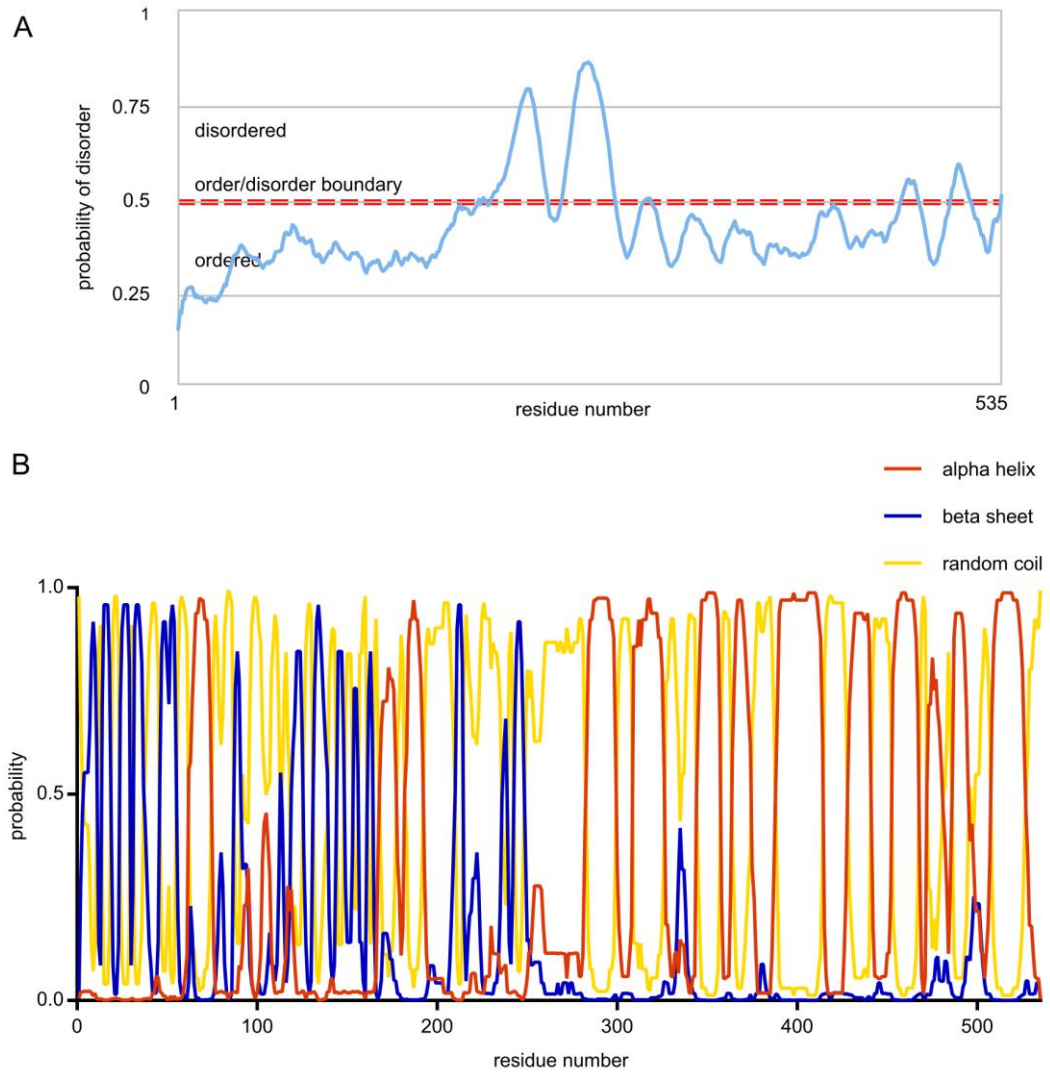


Figure 68 (A) Predicted disorder and (B) secondary structure of UL21 (535 amino acids, from HSV-1 strain KOS, Uniprot ID F8RG07). Protein disorder and secondary structure were predicted using RONN and NetSurfP, respectively.

5.2 Exploratory expression, purification and characterization of candidate proteins

5.2.1 High-throughput expression and purification screen of gE, UL21 and UL56

High yields of expressed and purified protein are desirable for any biophysical and structural studies, but the expression of a target protein can depend considerably on the construct, the expression system and the cell type used. In order to find conditions favorable for the recombinant *E. coli* expression of the candidate proteins gE, UL21 and UL56, a high-throughput expression and purification screen was performed. Full-length UL21 (residues 1-535), the cytosolic domain of gE (residues 445-550), and two constructs of the cytosolic domain of UL56 (residues 1-207 and 1-215) were cloned into bacterial expression vectors encoding N- or C-terminal His-tags (pOPTH and pOPTnH, respectively). Two constructs of UL56 were used because UL56 contains two adjacent cysteine residues (C208 and C209) immediately preceding its transmembrane domain. These cysteine residues could potentially form aberrant disulfide bonds during expression, thus leading to misfolding or aggregation of UL56. Three constructs unrelated to this project were employed as positive expression controls: α-SNAP and ORF55, two strongly expressing proteins, and VPS18 β-propeller, a weakly expressing protein. The expression and subsequent purification of all constructs was assessed in three different strains of *E. coli* using inducible overnight expression followed by affinity purification (Figure 69).

As expected, the three positive control constructs were expressed in and could be purified from all cell strains tested. High amounts of both N-terminally and C-terminally tagged UL21 could be purified from all *E. coli* cell strains. Virtually no glycoprotein gE could be purified from BL21(DE3) pLysS and B834(DE3) cells, but a band corresponding to the size of gE(445-550) was observed for the C-terminally tagged construct gE(445-550)-His expressed in Rosetta2(DE3) pLysS cells. Of the four constructs tested for UL56, only N-terminally tagged His-UL56(1-207) could be purified and the yields were comparable for all three cell strains used. It should be noted that the band for His-UL56(1-207) was observed at around 30 kDa, not at the expected molecular weight of around 24 kDa. Since both UL21 and UL56 could be purified to high yields following *E. coli* expression, large-scale purification and characterization of these two proteins were attempted.

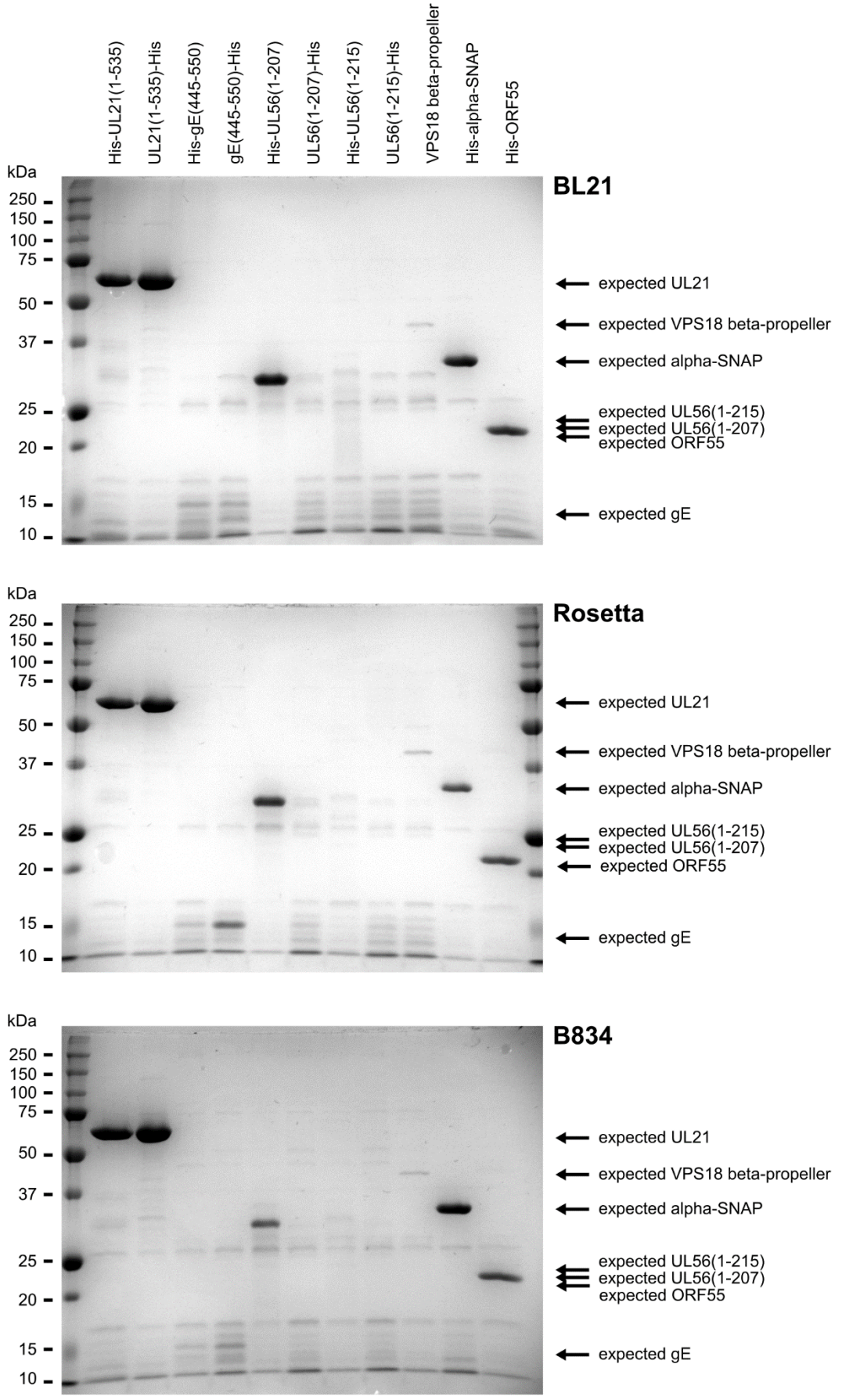


Figure 69 High throughput expression and purification trial of N- or C-terminally His-tagged UL21(full-length), cytoplasmic domain of gE (residues 445-550), and cytoplasmic domain of UL56 (residues 1-207 or 1-215). Constructs were expressed overnight in BL21(DE3) pLysS, Rosetta2(DE3) pLysS or B834(DE3) competent *E. coli* cells. Bacterial

cell pellets were harvested, lysed and the supernatant from the cell lysates was collected. The His-tagged proteins were then affinity purified by incubation of the cleared lysate with Ni-NTA agarose beads, eluted and separated on 12% SDS-PAGE gels. Constructs Vps18 β-propeller, α-SNAP and ORF55 served as positive expression controls. The expected molecular weight of the expressed constructs is indicated.

5.2.2 Purification, solubility screening and crystallization trials of UL21

In the high-throughput screen, higher yields of full-length UL21 could be captured when it was C-terminally tagged. Therefore, construct UL21(1-535)-His was used for a large-scale purification attempt. After overnight expression in BL21(DE3) pLysS *E. coli* cells, UL21(1-535)-His was purified by affinity chromatography using a Ni-NTA agarose resin followed by size exclusion chromatography (Figure 70). As assessed by SDS-PAGE, UL21 was approximately 99% pure after size exclusion chromatography. However, UL21 was not very soluble in the buffer used for gel filtration (20 mM Tris pH 7.5, 200 mM NaCl, 1 mM DTT) and could only be concentrated to a maximum of 3.2 mg/mL before precipitating. It also appeared to be temperature-sensitive, exhibiting higher solubility at lower temperatures, and was therefore handled on ice when possible.

To assess the thermal stability of UL21, differential scanning fluorimetry experiments using the fluorescent dye SYPRO Orange were performed. In the absence of SYPRO Orange, no emission could be detected (Figure 71A). The melting temperature of UL21 was not strongly influenced by the concentration of SYPRO Orange in the reaction mix, but a slight melting temperature decrease of less than 3 K could be observed with increasing amounts of UL21 (Figure 71C). On average, the T_M of UL21 was approximately 54 °C. Notably, when no UL21 was present, some emission of SYPRO Orange was still recorded. The absolute intensity of this background emissions was at least 80% lower than the in the presence of UL21 (Figure 71B).

Since UL21(1-535)-His precipitated easily upon concentration, a thermal shift buffer screen was carried out to find buffer conditions in which UL21 would be more stable (Figure 72). Overall, only little variation of the melting temperature of UL21 in these different buffers was observed, with the lowest measured T_M 49.5 °C and the highest T_M 54.5 °C. Within the panel of conditions tested, the buffer UL21 was filtered into during the final purification step (20 mM Tris pH 7.5, 200 mM NaCl, 1 mM DTT) provided one of the most stabilizing conditions (T_M around 54 °C).

Crystallization of UL21(1-535)-His at 3.2 mg/mL was attempted using sparse-matrix screens, but yielded no crystals. Whilst work on this thesis was in progress, E. Heldwein and co-workers solved the structure of first the N-terminal domain of UL21 (residues 1-198) and then the C-terminal domain (residues 279-530) (254, 255). Therefore, no further crystallization trials of full-length UL21 were pursued.

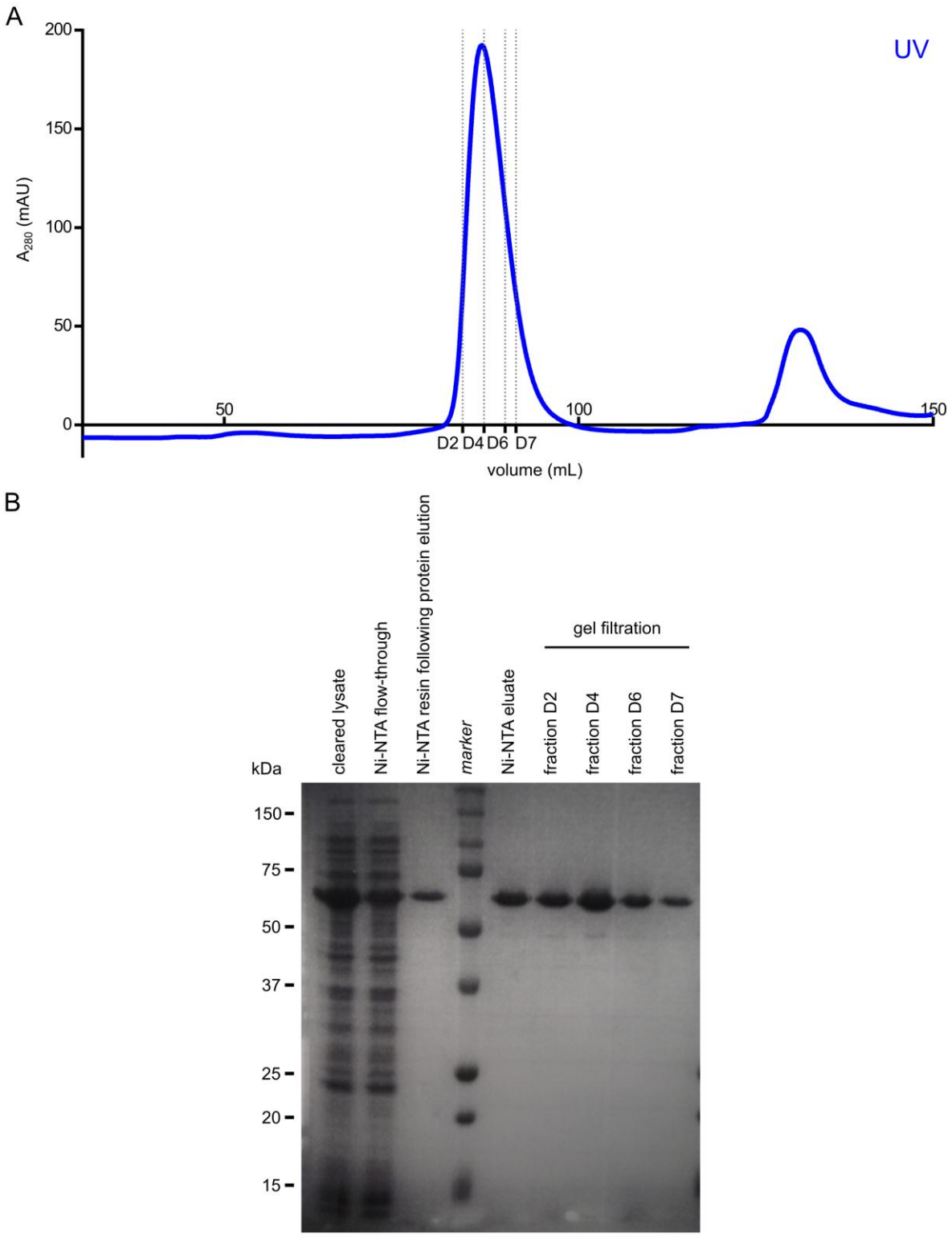


Figure 70 Purification of UL21(1-535)-His. The protein was first purified by Ni-NTA affinity chromatography followed by size exclusion. (A) Size exclusion elution profile of UL21(1-535)-His on an S200 16/600 gel filtration column (column volume: 120 mL). (B) Coomassie-stained 12% SDS-PAGE gel showing samples from key steps of the purification process as well as selected fractions collected during the size exclusion chromatography. Fractions D2-D7 were pooled, concentrated to 3.2 mg/mL and used for crystallization and differential scanning fluorimetry experiments.

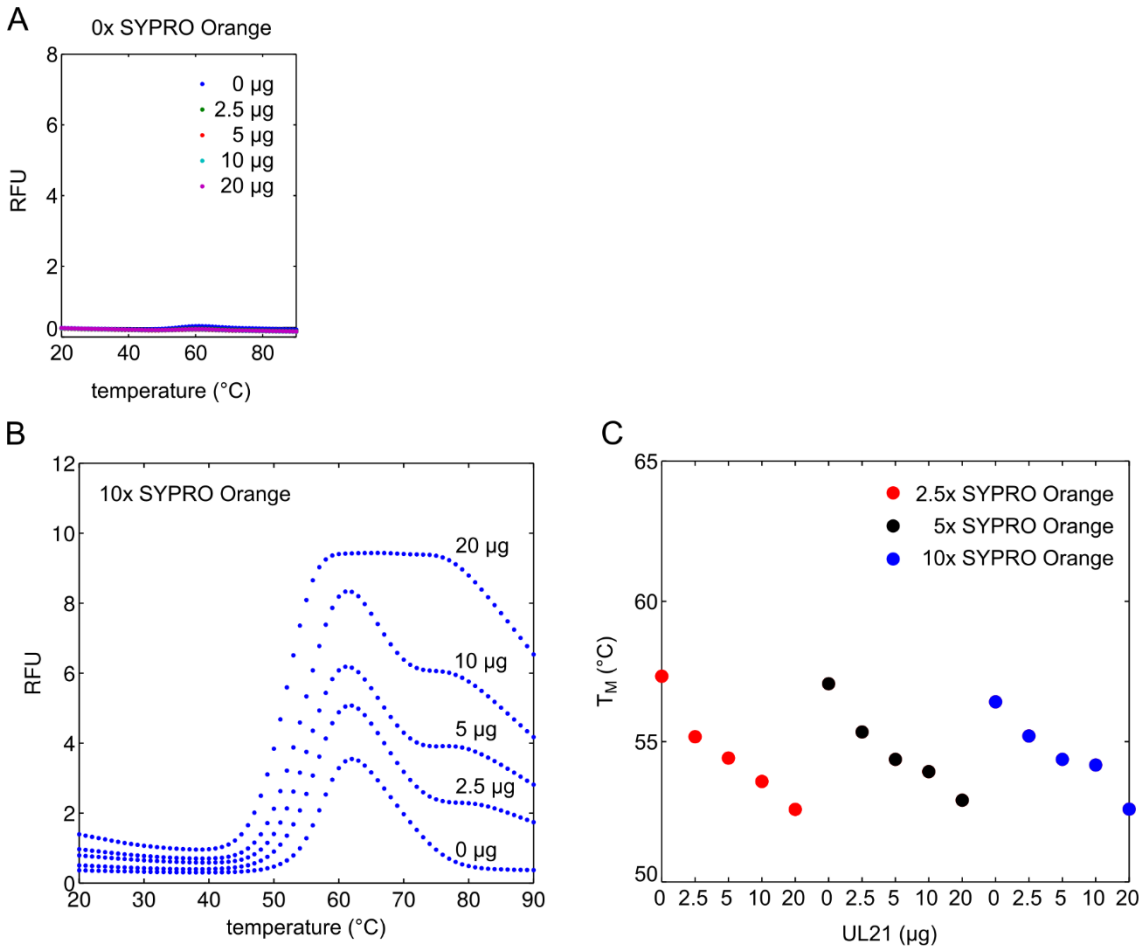


Figure 71 Optimization of a thermal shift assay to measure the thermal stability of UL21 using varying quantities of UL21 and different concentrations of SYPRO Orange as indicated. (A) Purified UL21(1-535)-His in the absence of dye does not exhibit any fluorescence. (B) Examples of melting curves used to calculate the T_M in (C). (C) Calculated melting temperatures (T_M) of UL21 at different concentrations of SYPRO Orange. Data were obtained on a MiniOpticon real-time PCR system.

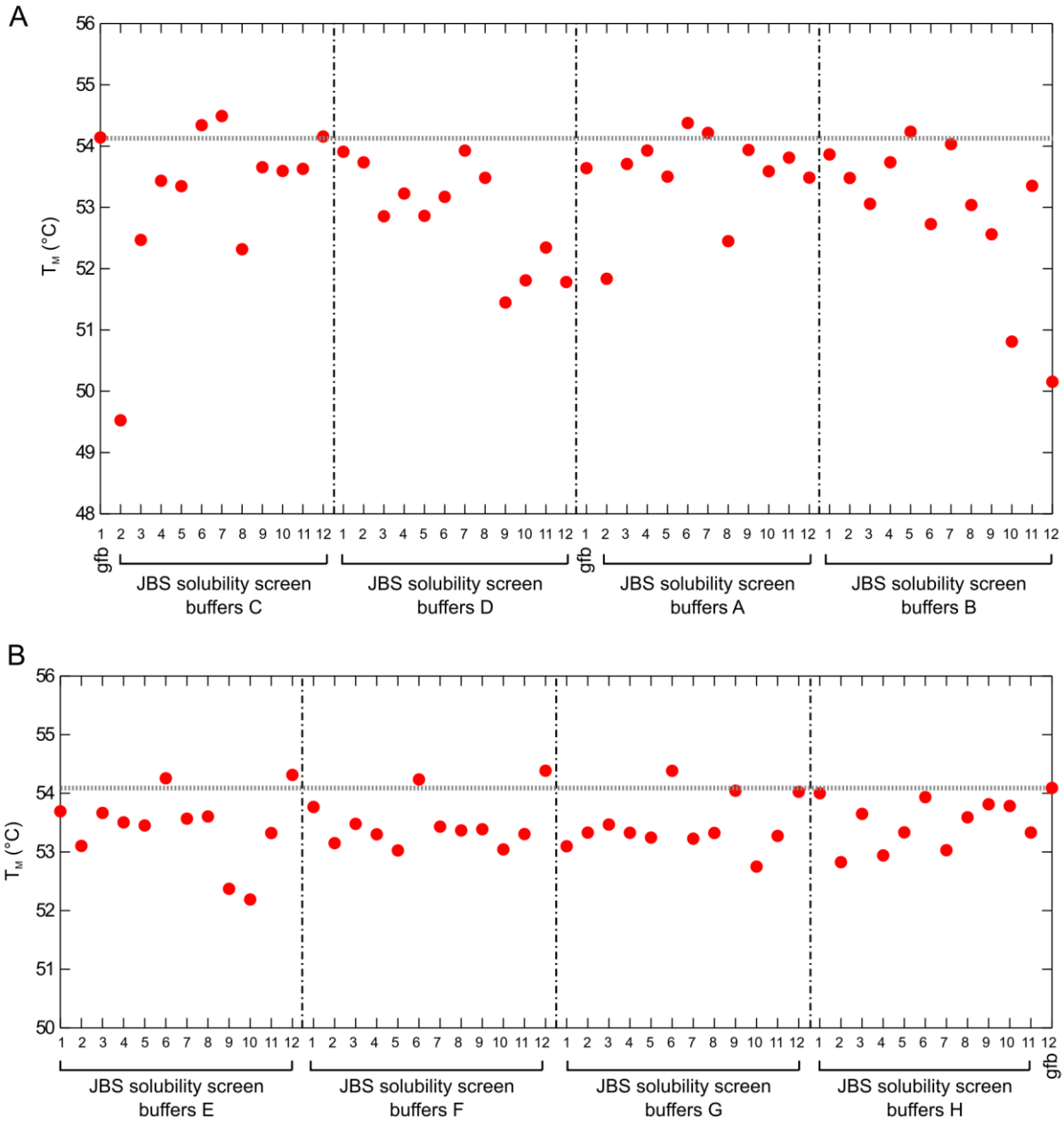


Figure 72 Differential scanning fluorimetry screen to find buffer conditions promoting stability of UL21(1-535)-His. Buffers from a commercial buffer solubility kit (JBS solubility screen, Jena Biosciences; (A) buffers A2-D12, (B) buffers E1-H11) were compared against the buffer used during the gel filtration step of the UL21 purification (gfb, gel filtration buffer, 20 mM Tris pH 7.5, 200 mM NaCl, 1 mM DTT). The melting temperature of UL21(1-535)-His is indicated by a horizontal dashed line. Data were obtained on a MiniOpticon real-time PCR system.

5.2.3 Purification of UL56

The only UL56 construct that could be purified in the *E. coli* high throughput expression and purification screen was His-UL56(1-207), an N-terminally tagged version of the protein truncated before the C-terminal cysteine doublet. Even though UL56 is predicted to be largely disordered (Table 14 and Figure 67), large-scale purification was attempted. As indicated by the expression and purification screen, construct His-UL56(1-207) was expressed overnight in BL21(DE3) pLysS *E. coli*. The protein was purified first by Ni-NTA affinity chromatography. Upon concentrating the eluate from the affinity chromatography column, formation of small, pebble-like precipitates was noticed. Inspection under a light microscope showed these pebbles to consist of a shiny, gel-like outer shell and a dense, brown precipitated core. This precipitate was removed and the protein further purified by size exclusion chromatography (Figure 73). No distinct peak eluted from the size exclusion column. Instead, after a high void peak (around 50 mL), very broad and overlapping low intensity peaks were observed, indicative of unfolded protein. Fraction ranges across the elution profile were pooled, concentrated and analyzed on an SDS-PAGE gel (Figure 73B). All fraction ranges contained an enriched protein of roughly 30 kDa, consistent with the size of the overexpressed protein found for His-UL56(1-207) during the high throughput expression screen. Interestingly, the absorption spectra for these combined and concentrated fractions exhibited strong absorption at 260 nm (data not shown), suggesting possible nucleic acid contamination of His-UL56(1-207).

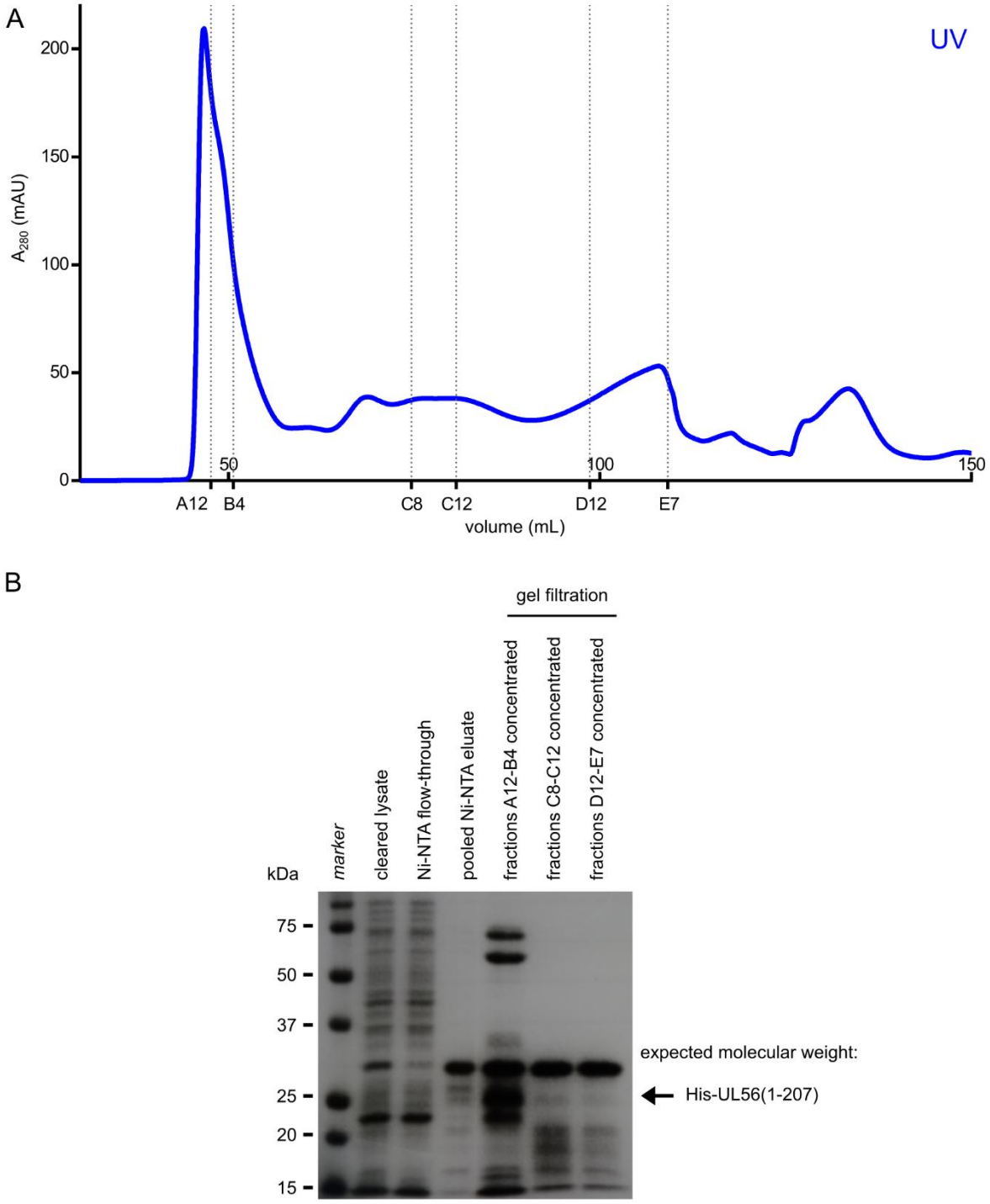


Figure 73 Purification of His-UL56(1-207). The protein was purified by Ni-NTA affinity chromatography followed by size exclusion chromatography. (A) Size exclusion elution profile of His-UL56(1-207) on an S75 16/600 gel filtration column (column volume: 120 mL). (B) Coomassie-stained 12% SDS-PAGE gel showing samples from key steps of the purification process as well as pooled size exclusion chromatography fraction ranges (A12-B4, C8-C12 and D12-E7).

5.3 Discussion

Envelope and tegument proteins of HSV-1 play critical roles during membrane acquisition and intracellular trafficking of virus particles. Glycoprotein gE, membrane protein UL56 and tegument protein UL21 have been implicated in these steps of the viral life cycle previously and were therefore chosen as candidate proteins for protein:protein interaction screens to identify their cellular interaction partners, and for structural and biophysical studies.

Expression and purification conditions were found for all three proteins, with gE and UL56 being much more effectively purified when tagged at their C or N termini, respectively. Of the four different UL56 constructs tested, only the shorter His-UL56(1-207) expressed well, which is probably due to the longer constructs containing a cysteine doublet before the truncated transmembrane domain that potentially leads to the formation of aberrant disulfide bonds and subsequent protein aggregation. Protein His-UL56(1-207) was found to run approximately 6 kDa higher on an SDS-PAGE gel than expected. This is in agreement with a previous report of UL56 purification from bacterial cells (232). UL56 has been reported to be phosphorylated (224, 236) and protein phosphorylation can affect migration during SDS-PAGE. However, since phosphorylation is unlikely to occur in bacterial expression systems, the reason for UL56 running at a higher molecular weight than expected is currently unclear. Large-scale purification of UL56 proved to be difficult due to nucleic acid contamination. This strong association with nucleic acid could also potentially influence screening for interaction partners of UL56, delivering DNA- or RNA-binding proteins as false positive hits. The elution behavior during size exclusion supports the prediction of UL56 being a largely disordered protein. Therefore, UL56 is not a suitable target for crystallization studies on its own, though it remains possible that certain regions of UL56 fold upon interaction with its binding partners.

Full-length UL21 was readily expressed and purified and was found to be stable at lower concentrations in a wide range of buffers. At concentrations above around 3 mg/mL, UL21 precipitated rapidly, consistent with the observation of UL21 exhibiting a lower thermal stability at higher concentrations. Very few structures of HSV-1 proteins are known, and bioinformatic predictions of the secondary structure and overall order of UL21 highlighted it as a good target for crystallographic studies. Heldwein and colleagues published structures of the N- and C-terminal domains of UL21 in 2015 and 2016, at 2.0 Å and 2.7 Å, respectively (254, 255). In

agreement to the structure predictions presented in this chapter, the structures of UL21 show that it consists of two stable domains, a mainly α-helical, "dragonfly-shaped" C-terminal domain and a "sail-like" N-terminal β-sheet domain. In agreement with the low stability of UL21 at concentrations above 3 mg/mL shown in this chapter, both domains were crystallized at comparatively low protein concentrations (3 mg/mL for the C-terminal domain, 5 mg/mL for the N-terminal domain). Interestingly, Heldwein et al. reported that full-length UL21 underwent spontaneous proteolytic cleavage first during expression and, for the C-terminal domain, a second time during the crystallization experiment, leading to the formation of the two domains they crystallized. Such proteolysis was not observed during the purification of full-length UL21 described in this chapter, even though the same *E. coli* expression strain (Rosetta) was used. Since structures of both ordered domains of UL21 are now available, I chose not to pursue determination of the full-length structure in isolation since such a structure is unlikely to be more informative on its own. Nonetheless, investigation of the orientation and dynamics of the two domains with respect to each other or upon binding of UL21 interaction partners could yield valuable information and the purification method and buffer stability screening presented in this chapter could prove useful for such studies.

6 YEAST TWO-HYBRID SCREENING OF NOVEL UL56 INTERACTION PARTNERS

The previous chapter described the selection of three HSV-1 proteins for further study that modulate envelopment and intracellular trafficking of virus particles. In order to identify cellular interaction partners of these proteins, high throughput methods such as yeast two-hybrid screens can be employed. First, I assessed proteins gE, UL21 and UL56 for their suitability as bait proteins in yeast two-hybrid assays. Two of the three candidate proteins, gE and UL21, proved to be unsuitable for screening due to autoactivation (section 6.1). Therefore, I performed a comprehensive yeast two-hybrid screen with only the third candidate protein, UL56, and present its putative interaction partners in this chapter (section 6.2). GST pull-down and co-immunoprecipitation experiments were performed in an attempt to confirm the interaction between UL56 and selected putative binding partners (section 6.3).

6.1 Yeast two-hybrid construct design and autoactivation screen

Yeast two-hybrid assays can be used to detect protein:protein interactions. A protein of interest (bait protein) is fused to the binding domain of the yeast promoter protein GAL4, and a library of potential prey proteins to the activation domain of GAL4. If a prey protein interacts with the bait protein in yeast, binding and activation domain of GAL4 are brought together and transcription of reporter genes is triggered. Reporter genes can confer nutritional selection, yeast colony color change as well as drug resistance. In this study, the Matchmaker® Gold Yeast Two-Hybrid System was used, in which bait proteins are expressed in the Y2HGold yeast strain and the prey library is expressed in strain Y187. Selection is achieved via the nutritional markers adenine and histidine, plus resistance to Aureobasidin A and a blue color change in the presence of the chromogenic substrate X-α-Gal due to expression of α-galactosidase. Positive clones can be isolated and sequenced following yeast colony PCR, enabling the identification of the interaction partner from the prey library.

Proteins gE, UL21 and UL56 were cloned into the pGBKT7 vector to be used as bait proteins in a yeast two-hybrid screen. Constructs were designed in accordance with the high throughput expression and purification screen (Chapter 5). For UL21, the full-length protein and for gE, the

residues of the cytoplasmic domain (445-550) were cloned. For the envelope protein UL56, two constructs without the transmembrane domain were cloned, UL56(1-215) and UL56(1-207), the shorter one lacking the C-terminal cysteine doublet that could potentially form aberrant disulfide bonds when expressed recombinantly.

In order to test whether the chosen bait constructs do not activate the expression of the reporter genes in the absence of prey protein, an autoactivation test was performed (Figure 74). In addition to the described bait vectors, three control bait vectors, pGBKT7-53, pGBKT7-lam and pGBKT7, as well as two control prey vectors, pGADT7-T and pGADT7 were used. The construct pGBKT7-53 encodes the mouse protein p53 and pGADT7-T encodes the Simian Vacuolating Virus 40 large T antigen (SV40-T). These two proteins serve as a positive control for reporter gene activation in the autoactivation test since they are known to interact in the yeast two-hybrid assay performed (344). Construct pGBKT7-lam encodes lamin and serves as a negative control since it does not interact with SV40-T.

Yeast were transformed, success of which was confirmed using the nutritional selection markers tryptophan and leucine for pGBKT7 and pGADT7, respectively. To test for autoactivation, the yeast strains were mated and then subjected to selection via the nutritional markers adenine, histidine, the antibiotic Aureobasidin A and blue/white color change mediated by X-α-Gal. Strong autoactivation was observed for the positive control (pGBKT7 + pGADT7-T), whereas no yeast colonies grew when the negative control pGBKT7-lam was mated with pGADT7-T and plated onto selective media or when empty bait or prey vectors were used (Figure 74, left). Of the four bait constructs of HSV-1 proteins tested, pGBKT7-gE and pGBKT7-UL21 showed strong autoactivation. Weak autoactivation could also be observed for the pGBKT7-UL56(1-215) construct, whereas the shorter UL56 construct pGBKT7-UL56(1-207) did not autoactivate (Figure 74, right).

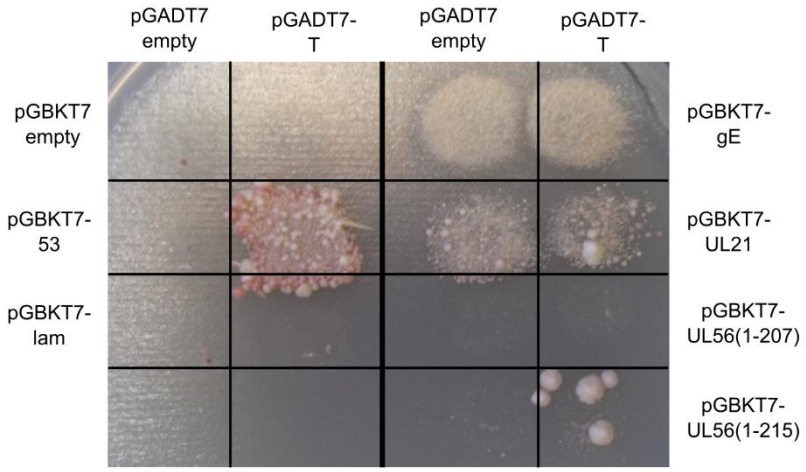


Figure 74 Yeast two-hybrid autoactivation screen with gE, UL21 and UL56 constructs. Bait- (pGBKT7, rows) and prey- (pGADT7, columns) expressing yeast cells were mated and plated onto non-selective agar. Subsequently, the plate was replicated onto selective agar (shown here) on which cells can only grow if bait and prey interact and therefore, all reporter genes are expressed. Positive and negative controls: left columns; HSV-1 candidate proteins: right columns.

6.2 Identification of UL56 interaction partners

Since construct pGBKT7-UL56(1-207) showed no signs of autoactivation, a yeast two-hybrid screen could be performed to identify interaction partners of UL56. As the prey library, a universal, normalized human Mate&Plate™ cDNA library derived from a collection of human whole tissues with a cDNA size range of 0.7-2.8 kb, a titer of 5×10^7 cfu/mL and 7.5×10^6 independent clones was used. Yeast strain Y2HGold transformed with pGBKT7-UL56(1-207) was mated with strain Y187 pre-transformed with the prey library. The mating efficiency was 15.9% and the number of mated yeasts screened was 3.8×10^6, exceeding the Matchmaker® Gold Yeast Two-Hybrid System recommendations (1.0×10^6 diploids screened, mating efficiency at least 2-5%). After mating, yeast were spread onto selection plates lacking the nutritional markers adenine, histidine, leucine and tryptophan.

More than 1000 positive colonies grew over a course of five weeks. At regular intervals, a total of 302 colonies were picked and streaked onto higher stringency selective plates containing the above nutritional markers, Aureobasidin A and the chromogenic substrate X-α-galactosidase in order to confirm the positive phenotype. In total, 263 colonies grew on these plates and were re-streaked onto fresh high stringency plates to ensure that each colony derived from a single yeast cell. Colony PCRs were performed for all 261 colonies re-grown on these plates, yielding 183 PCR

products that were used for Sanger sequencing. In total, 180 reactions could be sequenced, leading to an overall positive clone identification rate of 60% (180/302 colonies) (Figure 75).

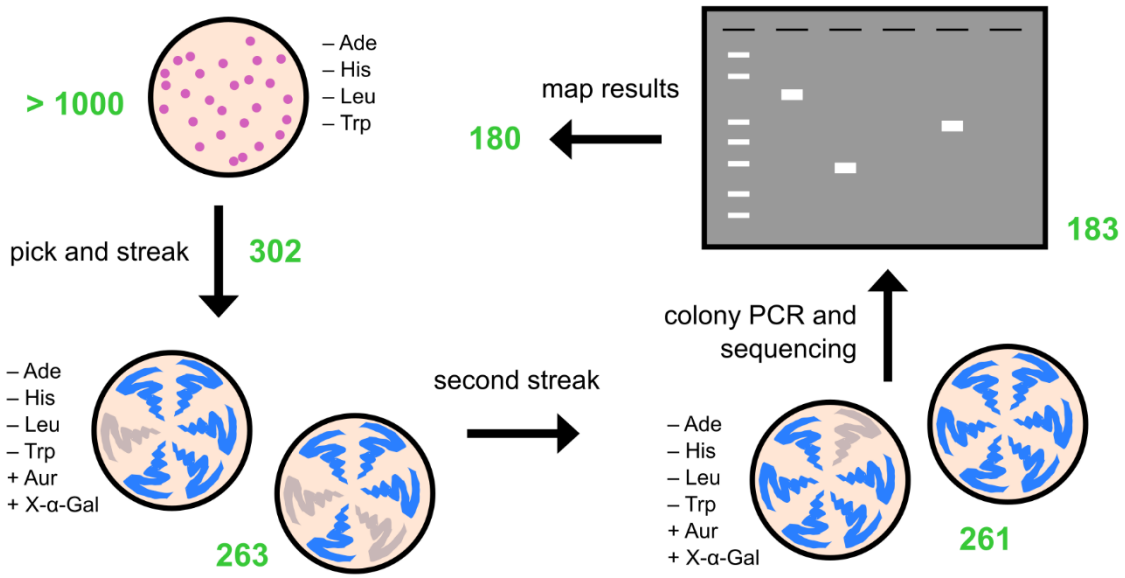


Figure 75 Illustration of yeast two-hybrid positive clone selection process. Yeast cells expressing UL56(1-207) fused to the GAL4 binding domain were mated with yeast cells expressing a human cDNA library fused to the activation domain of GAL4 and plated onto selection plates lacking adenine, histidine, leucine and tryptophan (-Ade, -His, -Leu, -Trp). Interaction of UL56 with one of the library prey proteins leads to the expression of reporter genes, enabling growth on the selective medium. Positive clones were streaked onto higher stringency plates containing Aureobasidin A (+Aur) and X-α-galactosidase (+X-α-Gal) in addition to the nutritional markers, re-streaked to ensure each colony grew from a single yeast cell, subjected to colony PCR and sequenced. Numbers (green) of grown colonies, successful PCR reactions and sequencing results mapped to coding regions of human mRNA are indicated.

Overall, 146 of the 180 sequenced positive clones could be mapped to coding regions of human mRNA in the correct strand sense. The remaining 34 clones mapped to non-coding RNA or chromosomal DNA. All sequenced clones which were identified in at least two positive colonies are listed in Table 15. Even though yeast two-hybrid assays ideally identify bimolecular, direct protein:protein interactions, false positive hits are commonly observed (345) and any interaction of interest should be validated by complementary biochemical or biophysical methods. Consequently, the most promising positive clones obtained as potential UL56 interaction partners in the yeast two-hybrid screen were selected for independent confirmation.

Several potential false positive clones such as phosphoglucomutase 1, some zinc finger proteins and DNA-binding proteins (zinc finger, AN1-type domain 1; ZFP1 zinc finger protein; zinc finger with KRAB and SCAN domains 4; cAMP responsive element modulator; TOX high mobility group box family member 4) were identified and not considered for follow-up. These proteins had

either been seen in other yeast two-hybrid screens in our laboratory, or are DNA-binding proteins. Similarly, proteins with tissue specificity such as fibrinogen-like 1 (liver-specific), cysteine-rich secretory protein 2 (sperm protein), sperm associated antigen 5 and spermatogenesis associated protein 22 (highly expressed in testis) were not analyzed further.

The most striking positive hit was PRKC, apoptosis, WT1, regulator (PAWR), identified with four independent overlapping clones in six positive colonies. Protein PAWR is a pro-apoptopic protein (346) and could therefore potentially be targeted by HSV-1 to prevent apoptosis of host cells. Pleckstrin and Sec7 domain containing protein 3 (PSD3) was another interesting hit since it is predicted to be a potential guanine nucleotide exchange factor for ADP-ribosylation factor 6 (ARF6), an intracellular protein involved in membrane trafficking during endocytosis (347). Therefore, these two proteins were selected as top hits for further biochemical validation.

SUMO1/sentrin specific peptidase 6 (SENP6), a protein involved in post-translational modification by cleaving the small ubiquitin-like modifier (SUMO) from cellular proteins (348), and the little-studied proteins folliculin interacting protein 2 (FNIP2), WD repeat domain 61 (WDR61) and probable ATP-dependent RNA helicase DDX59 were chosen as four additional hits for confirmation. Lastly, Beclin-1, a protein with key roles in autophagy and previously reported to interact with HSV-1 protein ICP34.5, resulting in inhibited autophagosome generation (349), was selected for further characterization, although only a single positive clone was identified in the yeast two-hybrid screen.

Table 15 Putative UL56 interaction partners identified in the yeast two-hybrid screen. Hits selected for biochemical confirmation are indicated by green shading and in the last column. (n.d.: not determined)

NCBI identifier	Gene	Protein name	Hits	Independent clones	Follow-up
NM_002583.2	PAWR	PRKC, apoptosis, WT1, regulator	6	4	yes
NM_014828.2	TOX4	TOX high mobility group box family member 4	4	1	-
NM_004467.3	FGL1	fibrinogen-like 1	3	1	-
NM_024699.2	ZFAND1	zinc finger, AN1-type domain 1	3	n.d.	-
NM_153688.2	ZFP1	ZFP1 zinc finger protein	3	n.d.	-
NM_000082.3	ERCC8	DNA excision repair protein ERCC-8	2	n.d.	-
NM_001031725.4	DDX59	Probable ATP-dependent RNA helicase DDX59	2	1	yes
NM_001170699.1	SPATA22	spermatogenesis associated protein 22	2	n.d.	-
NM_001261822.1	CRISP2	cysteine-rich secretory protein 2	2	n.d.	-
NM_002633.2	PGM1	phosphoglucomutase 1	2	n.d.	-
NM_005875.2	EIF1B	eukaryotic translation initiation factor 1B	2	1	-
NM_006461.3	SPAG5	sperm associated antigen 5	2	n.d.	-
NM_015571.2	SENP6	SUMO1/sentrin specific peptidase 6	2	1	yes
NM_019110.3	ZKSCAN4	zinc finger with KRAB and SCAN domains 4	2	n.d.	-
NM_020840.1	FNIP2	folliculin interacting protein 2	2	1	yes
NM_025234.1	WDR61	WD repeat domain 61	2	1	yes
NM_182769.2	CREM	cAMP responsive element modulator	2	n.d.	-
NM_206909.2	PSD3	pleckstrin and Sec7 domain containing 3 (isoform 3)	2	1	yes
NM_003766.3	BECN1	beclin 1, autophagy related	1	1	yes

6.3 Validation of yeast two-hybrid screening hits

In the previous section, it was described how proteins PAWR, DDX59, SENP6, FNIP2, WDR61, PSD3 (isoform 3) and BECN1 were identified as potential novel interaction partners of the HSV-1 protein UL56 in a yeast two-hybrid screen. To confirm these interactions, two parallel biochemical capture methods were tested. First, the prey proteins were expressed in a cell-free wheat germ expression system and tested for binding to purified GST-UL56(1-207). Interactions confirmed by this method were assessed by co-immunoprecipitation of endogenous protein from cells with eGFP-tagged UL56.

6.3.1 Purification of GST-UL56(1-207)

Since only N-terminally His-tagged UL56(1-207) was expressed in the high throughput expression and purification screen (Chapter 3), the protein was tagged at the N terminus with GST for use in GST pull-down assays (GST-UL56(1-207). The protein was expressed overnight in the *E. coli*

strain BL21(DE3) pLysS. The protein was first affinity-purified using glutathione sepharose resin followed by size exclusion chromatography (Figure 76) into a mild, near-neutral buffer (20 mM HEPES pH 7.5, 200 mM NaCl, 1 mM DTT).

As during the purification of His-tagged UL56(1-207), GST-UL56(1-207) did not elute in a sharp, single peak, but across the column, suggestive of UL56 being intrinsically disordered. Additionally, as observed by the A_{260}/A_{280} ratio, GST-UL56(1-207) was found to be contaminated with nucleic acid which could not cleared by the initial DNase treatment of the cell lysate. Cation exchange chromatography was attempted to remove the nucleic acid contamination, but no GST-tagged UL56(1-207) could be recovered. For this reason, the fractions of GST-UL56(1-207) collected during size exclusion (B6-C8, Figure 76) were pooled and concentrated to an apparent concentration of 6.8 mg/mL as estimated by measuring the absorbance of the sample at 280 nm. Even though this concentration is likely grossly overestimated due to the contaminating nucleic acids absorbing strongly at 280 nm, semi-purified GST-UL56(1-207) was used for pull-down assays. without problems. As none of the putative interaction partners of UL56 had an overall basic charge or known nucleic acid-binding affinity (except for DDX59), I decided to proceed with GST pull-down interaction assays using this semi-purified GST-UL56 sample.

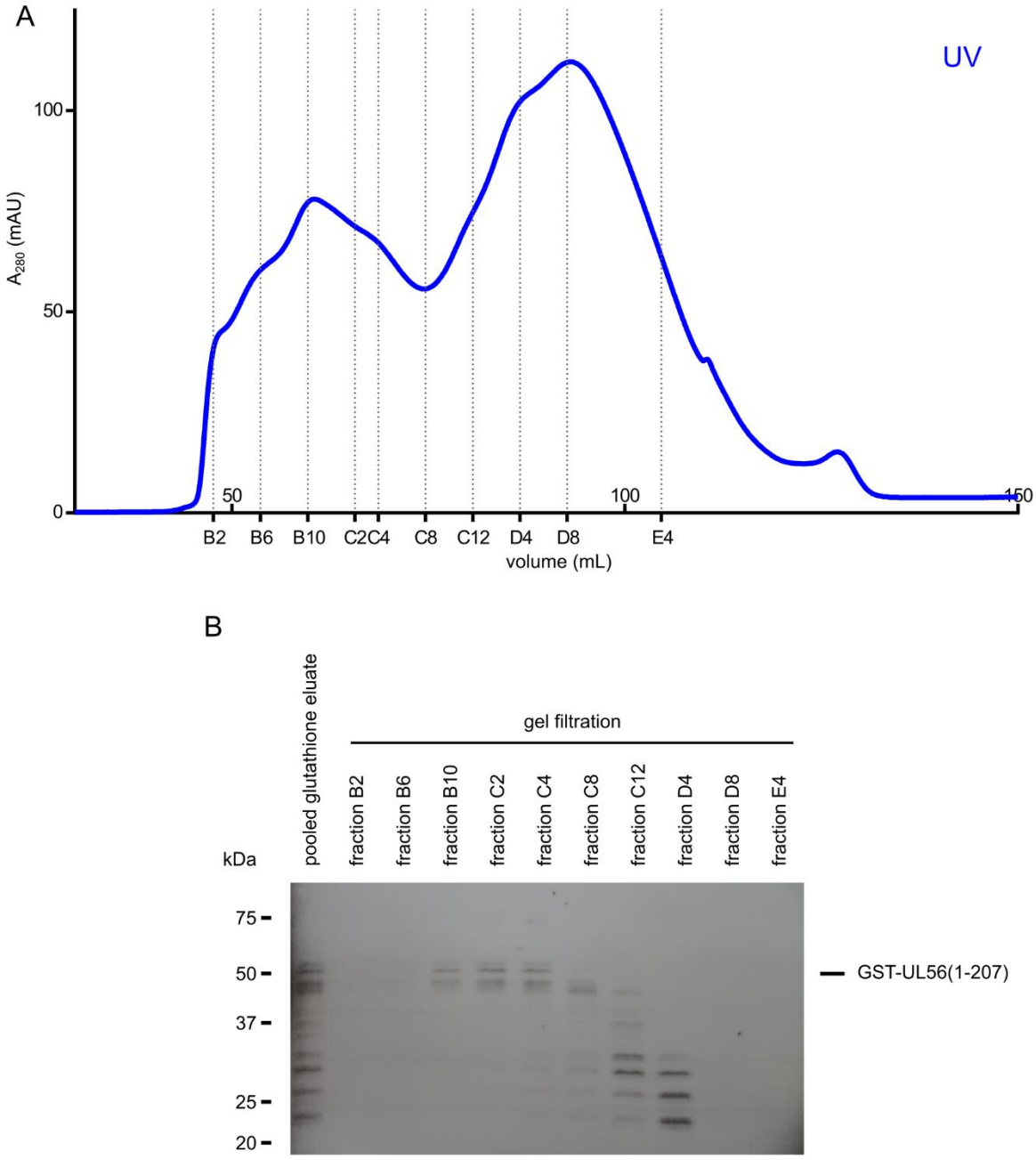


Figure 76 Purification of GST-UL56(1-207). The protein was first purified by glutathione sepharose affinity chromatography followed by size exclusion chromatography. (A) Elution profile of GST-UL56(1-207) on an S200 16/600 size exclusion column (column volume: 120 mL). (B) Coomassie-stained 12% SDS-PAGE gel showing samples from the purification process as well as selected fractions collected during gel filtration chromatography. Note the low amount of protein visible on the gel compared to the relatively high UV absorbance observed during size exclusion, presumably due to nucleic acid contamination.

6.3.2 Confirmation of direct interactions between UL56 and yeast two-hybrid hits by GST pull-down experiments

To test whether the interactions between UL56 and PAWR, DDX59, SENP6, FNIP2, WDR61, PSD3 and BECN1 could be replicated outside of yeast cells in an in vitro context, the yeast two-hybrid hits were cloned into a myc-tag-encoding vector suitable for expression in wheat germ extract (pF3A-myc) for use in GST pull-down experiments. Incomplete transcripts can be present in the prey cDNA library screened in the yeast two-hybrid assay, resulting in prematurely terminated proteins fused to the GAL4 activation domain. For some transcripts, termination could occur within a folded protein domain, leading to aberrantly folded proteins binding to the bait. Therefore, the sequences cloned into the wheat germ expression vector were designed to take folded protein domains into account and to encompass at least the interacting domain as sequenced from the yeast two-hybrid positive clones (Table 16).

The structures of BECN1 (residues 105-450) and WDR61 (residues 2-305) are known. For DDX59, FNIP2, PAWR, PSD3 (isoform 3) and SENP6, structure predictions based on homology detection were carried out using HHpred. For BECN1, DDX59 and PSD3 (isoform 3), residues 128-450, 115-619 and 1-243, respectively, were cloned into pF3A-myc since the cDNA library fusion protein terminated in a folded protein domain. Proteins FNIP2 and PAWR were cloned as found in the sequenced yeast two-hybrid clones as no reliable structural information was available to inform construct design, whereas WDR61 was cloned as the full-length protein. Structure predictions for SENP6 revealed an insertion splitting its catalytic domain compared to related SUMO proteases. A similar insertion is observed in SENP7, but the exact function of these insertion is currently unknown (350). The start of the interacting domain found in the yeast two-hybrid screen lies within the large catalytic domain insertion of SENP6. Therefore, two constructs were cloned for SENP6, one from the start of the domain identified in the yeast two-hybrid screen up to the very C-terminus of SENP6 (809-1112), the other one encompassing only the predicted folded C-terminal catalytic domain after the insertion (972-1112).

Table 16 Wheat germ expression constructs designed for GST pull-downs.

Protein	Full length (amino acids)	Interacting domain from yeast two-hybrid (amino acids)	Region cloned into pF3A-myc (amino acids)
BECN1	450	128-372	128-450
DDX59	619	115-347	115-619
FNIP2	1114	870-1114	870-1114
PAWR	340	217-340	217-340
PSD3 (isoform 3)	513	1-173	1-243
SENP6	1112	809-1075	809-1112 972-1112
WDR61	305	71-305	1-305

All prey protein constructs expressed well using a cell-free wheat germ extract expression system (Figure 77, input lanes). Glutathione-S-transferase pull-down assays using magnetic glutathione beads were performed to test for direct binding between GST-UL56(1-207) and the myc-tagged prey proteins. Purified GST served as a negative control bait protein. Due to nucleic acid contamination of GST-UL56(1-207) (section 6.3.1), the amount of protein loaded onto the beads was based on the apparent protein concentration and therefore, underestimated. Nevertheless, interactions between UL56 and several prey proteins were observed, whereas no binding to GST alone was detected for any prey protein. Glutathione-S-transferase-tagged UL56 was found to bind to myc-tagged BECN1, FNIP2, both constructs of SENP6, and WDR61, potentially confirming these yeast two-hybrid positive clones as bona fide interaction partners of HSV-1 UL56 (Figure 77).

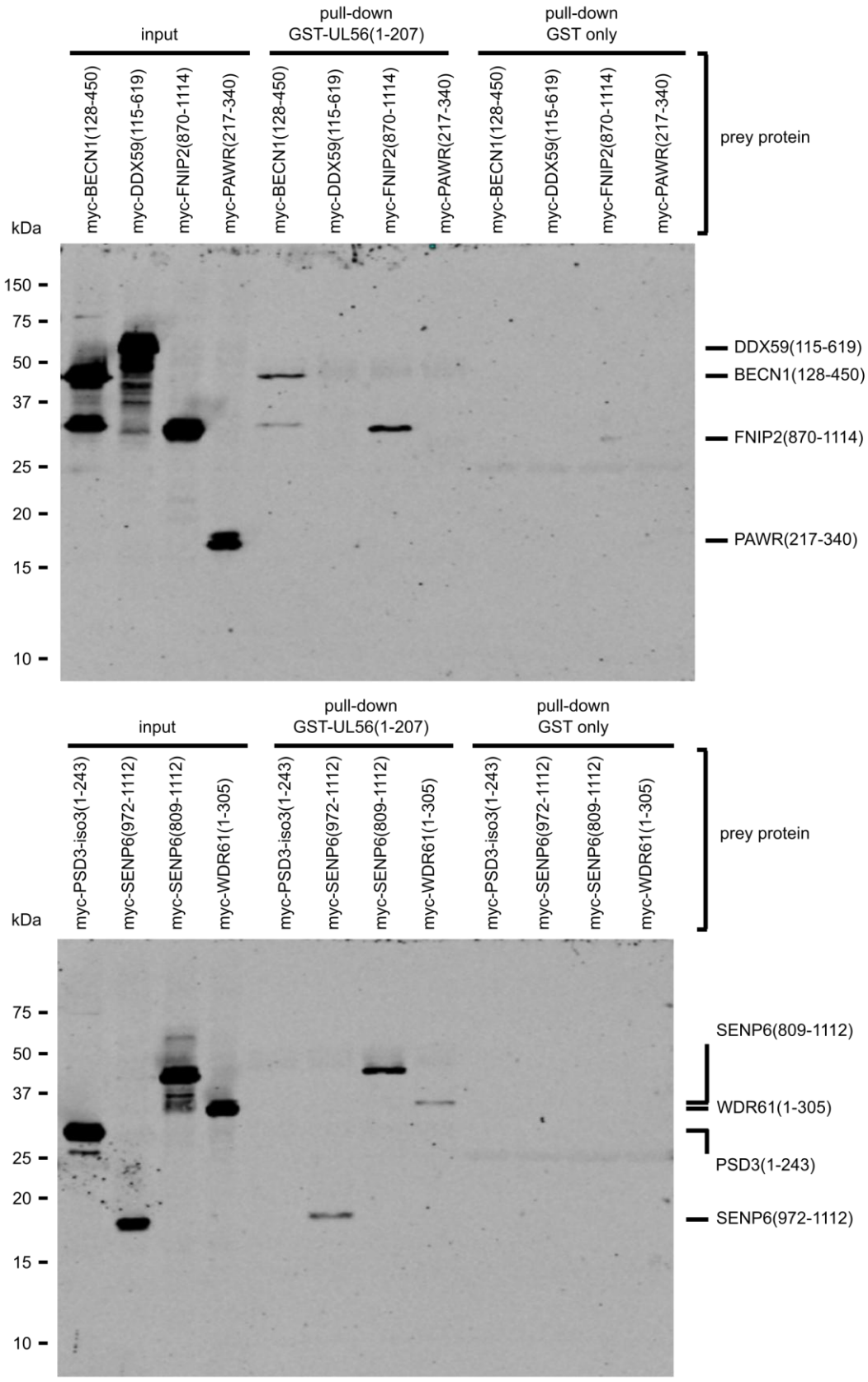


Figure 77 Confirmation of direct interactions between UL56 and yeast two-hybrid hits by GST pull-down. Myc-tagged prey proteins were expressed using wheat germ extract (input) and incubated with magnetic glutathione beads pre-loaded with purified GST-UL56(1-207) or GST as a negative control. After 75 min, the beads were washed and the protein eluted using reduced glutathione. Input and bound samples (pull-down) were separated on 15% SDS-PAGE gels. Western blotting using an α-myc antibody was performed to visualize inputs and prey proteins bound to GST-UL56(1-207) or GST. Expected molecular weights of the prey proteins are indicated by arrows (right).

6.3.3 Optimization of a co-immunoprecipitation assay to confirm interactions

In GST pull-down assays, direct binding between UL56 and four of the seven hits identified in the yeast two-hybrid screen, namely BECN1, FNIP2, SENP6 and WDR61, was observed. In order to verify that these interactions do not only occur in yeast and using purified proteins in vitro, but also in the context of human cells, co-immunoprecipitation experiments were performed. Since no monoclonal antibody against HSV-1 UL56 is available, a C-terminally eGFP-tagged version of the protein was designed (UL56(1-207)-eGFP) by cloning UL56 into a mammalian expression vector. Co-immunoprecipitation was performed using cell lysate from human embryonic kidney cells 293 expressing the SV40-T antigen (HEK293T cells) transfected with UL56(1-207)-eGFP and assessed by Western blotting (Figure 78).

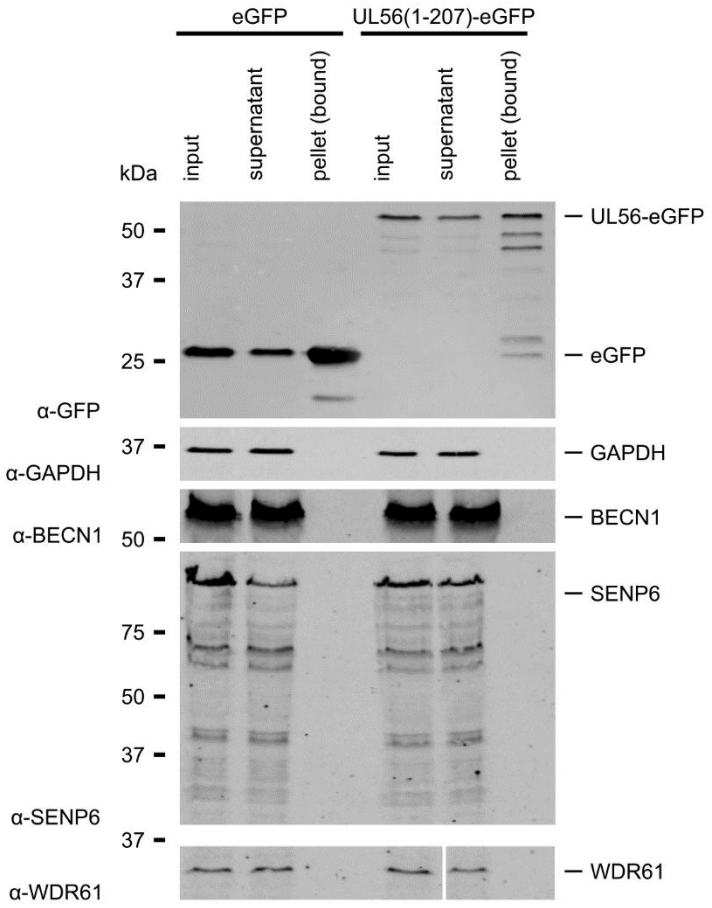


Figure 78 Co-immunoprecipitation experiment of UL56(1-207)-eGFP with its binding partners identified in GST pull-down assays. Lysates of HEK293T cells transfected with UL56(1-207)-eGFP or an eGFP-only control were collected and incubated with affinity resin GFP-Trap-A (lysis buffer: 10 mM Tris/HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 0.5% NP-40, 1:100 Sigma protease inhibitors). After 40 min, the beads were washed and proteins bound to the beads were dissociated by boiling at 95 °C (wash buffer: 10 mM Tris/HCl pH 7.5, 0.5 mM EDTA). After separation of the samples on 12% SDS-PAGE gels, Western blotting using target-specific antibodies as indicated was performed. Loading of samples: eGFP, GAPDH: 0.34% of total input and supernatant, 11% of total pellet sample; BECN1, SENP6, WDR61: 0.68% of total input and supernatant, 22% of total pellet sample.

Both eGFP and UL56(1-207)-eGFP, but not Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) were enriched in the bound samples, showing effective immunoprecipitation of eGFP and eGFP-tagged UL56. However, none of the previously identified interaction partners of UL56 were found in the bound samples. This observation could either be due to the putative interactions not occurring in cells, or due to experimental artifacts. In order to exclude the latter, the co-immunoprecipitation protocol was optimized in multiple aspects.

Firstly, UL56 lacking its transmembrane domain is reported to localize to mainly the nucleus in transfected cells (224, 233), and such localization was observed upon overexpression of UL56(1-207)-eGFP in HEK293T cells in this work (Figure 80). Since UL56 is a transmembrane protein

targeted to the trans-Golgi network and perinuclear membranes (220), it is possible that the mislocalization of the construct with a truncated transmembrane domain interferes with binding to its cytosolic interaction partners. To alleviate such effects, two full-length UL56 constructs with eGFP tags at the N and C termini, eGFP-UL56(1-234) and UL56(1-234)-eGFP, respectively, were transfected into HEK293T cells and their expression and binding to GFP-Trap-A beads assessed (Figure 79). C-terminally tagged UL56(1-234) expressed only very weakly and could not be captured efficiently. On the contrary, N-terminally tagged UL56(1-234) expressed well, localized to the cytoplasm (Figure 80) and could be enriched on GFP-Trap beads, making it a suitable construct for further co-immunoprecipitation interaction studies.

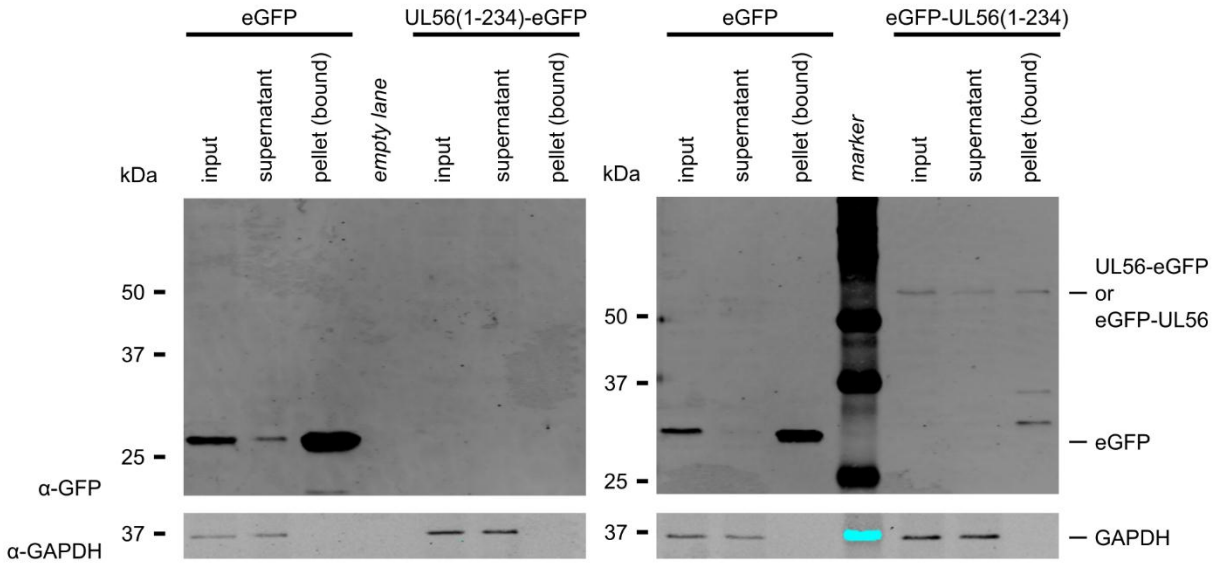


Figure 79 Evaluation of expression and GFP-Trap capture of C- and N-terminally eGFP-tagged, full-length UL56. Cells were transfected with eGFP, UL56(1-234)-eGFP or eGFP-UL56(1-234), lysed and the lysate incubated with GFP-Trap-A beads for 30 min (lysis buffer: 10 mM Tris/HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 0.5% NP-40, 1:100 Sigma protease inhibitors). The beads were washed (wash buffer: 10 mM Tris/HCl pH 7.5, 0.5 mM EDTA), bound proteins were eluted and separated on 12% SDS-PAGE gels. Tagged proteins were detected by Western blot using an α-GFP antibody. The house-keeping gene GAPDH served as a loading and unspecific capture control. Loading of samples: 0.6% of total input and supernatant, 11% of total pellet sample.

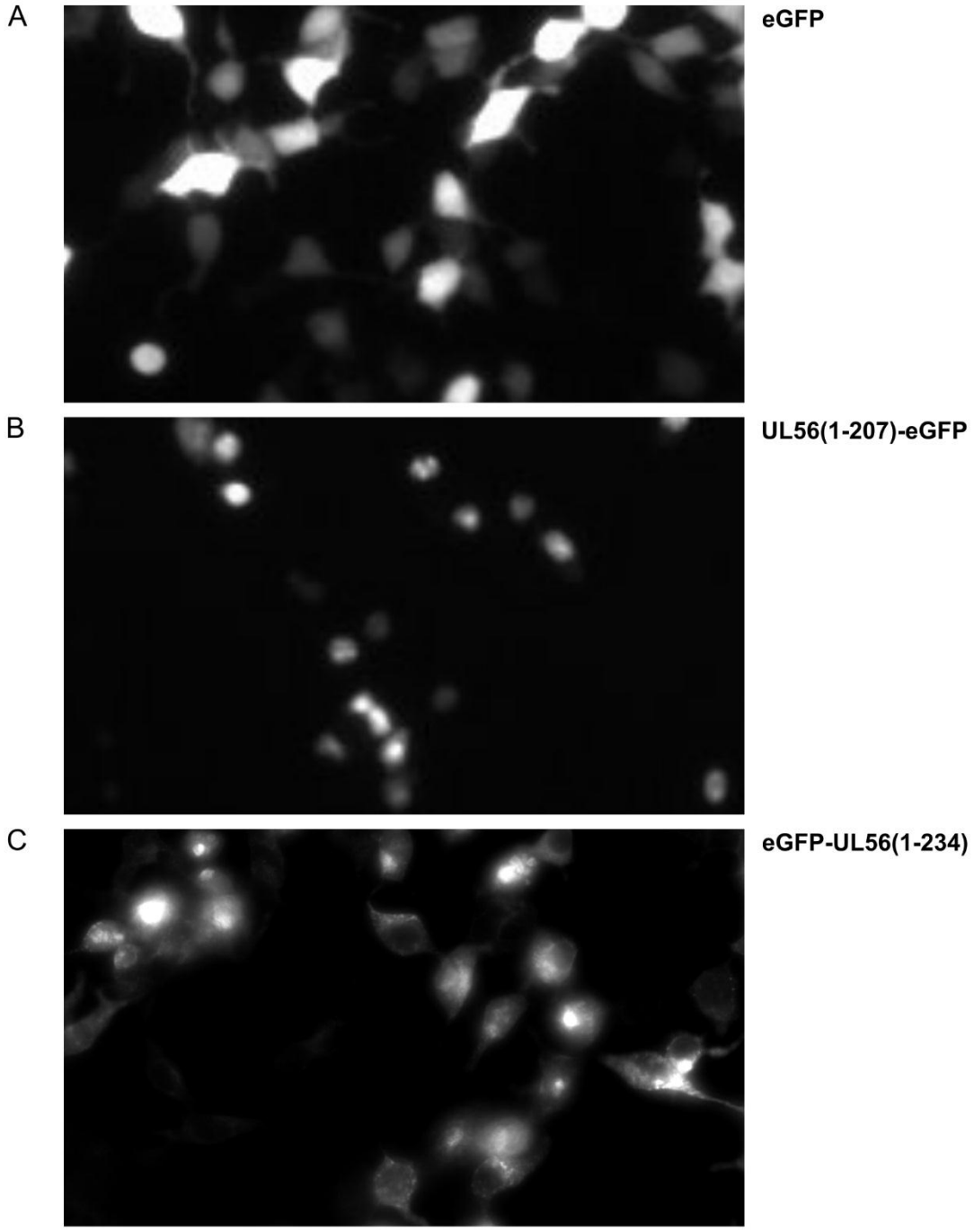


Figure 80 Localization of overexpressed UL56 constructs in HEK293T cells. Cells were transfected with a mammalian expression vector encoding (A) eGFP, (B) UL56 lacking its C-terminal transmembrane region (UL56(1-207)-eGFP), or (C) full-length UL56 (eGFP-UL56(1-234) and incubated for 18 hours. GFP fluorescence was assessed by fluorescence microscopy. eGFP and full-length UL56 show cytoplasmic localization, whereas truncated UL56 localizes to much smaller spots, consistent with nuclear localization. Images in (A) and (B) were kindly provided by Stephen C. Graham.

Secondly, despite broadly similar transfection rates and expression in cells, both truncated and full-length UL56 appeared to be captured much less efficiently on GFP-Trap beads than eGFP alone. If little GFP-tagged UL56 is immobilized, the detection of UL56-interacting proteins is considerably hindered. Therefore, optimization of the immunoprecipitation protocol was

attempted, trying to find conditions which would enable the extraction of more UL56 from cells whilst being gentle enough not to disturb potential interactions. As the abundance of eGFP-UL56(1-234) in cell lysates was low, four different lysis buffers testing two different detergents at two different concentrations were tested (Table 17). Corresponding wash buffers were chosen to contain the same detergent as the lysis buffer at one tenth of the concentration. Additionally, since UL56 exhibited a strong propensity to bind nucleic acids when expressed in *E. coli*, a mixture of an DNase and an RNase cocktail were used during lysis in addition to protease inhibitors.

Surprisingly, no prominent differences in either solubilization (input) or immobilization (pellet) of UL56 could be observed between the different buffer conditions tested (Figure 81). However, both full-length and truncated construct were extracted and captured more efficiently when using buffers containing nucleases (Figure 78 and Figure 79). Noticeably, full-length eGFP-UL56 seemed to be degraded less than UL56(1-207)-eGFP. Taken together, these results indicate that immunoprecipitation of UL56 was improved by addition of nucleases during the lysis and capture step of the experiment.

Table 17 Lysis and wash buffer conditions tested for the optimization of immunoprecipitation of eGFP-tagged UL56.

Condition	Lysis and binding buffer	Wash buffer
1	10 mM Tris/HCl pH 7.5, 150 mM NaCl, 0.5 mM $MgCl_2$ 0.5% NP-40 1:100 protease inhibitors 1 µg/100 mL DNase, 1.25 U/mL RNase cocktail	10 mM Tris/HCl pH 7.5 150 mM NaCl 0.05% NP-40
2	10 mM Tris/HCl pH 7.5, 150 mM NaCl, 0.5 mM $MgCl_2$ 1.0% NP-40 1:100 protease inhibitors 1 µg/100 mL DNase, 1.25 U/mL RNase cocktail	10 mM Tris/HCl pH 7.5 150 mM NaCl 0.05% NP-40
3	10 mM Tris/HCl pH 7.5, 150 mM NaCl, 0.5 mM $MgCl_2$ 0.5% Triton X-100 1:100 protease inhibitors 1 µg/100 mL DNase, 1.25 U/mL RNase cocktail	10 mM Tris/HCl pH 7.5 150 mM NaCl 0.05% Triton X-100
4	10 mM Tris/HCl pH 7.5, 150 mM NaCl, 0.5 mM $MgCl_2$ 1.0% Triton X-100 1:100 protease inhibitors 1 µg/100 mL DNase, 1.25 U/mL RNase cocktail	10 mM Tris/HCl pH 7.5 150 mM NaCl 0.05% Triton X-100

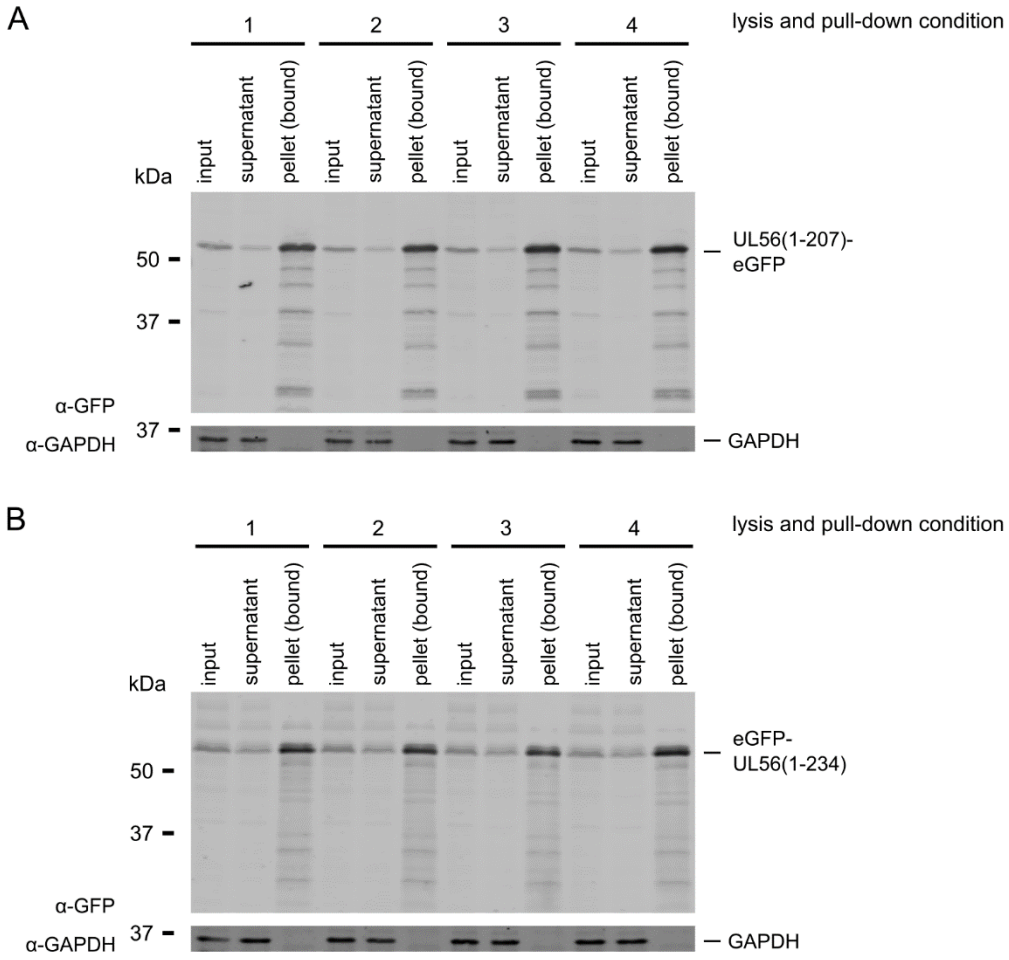


Figure 81 Screening of different lysis buffer conditions to optimize immunoprecipitation of eGFP-tagged UL56. Cells were transfected with UL56(1-207)-eGFP (A) or eGFP-UL56(1-234) (B). Cells were lysed in lysis buffers 1-4 (Table 17) and incubated with GFP-Trap-A beads. After 60 min, beads were washed, proteins were eluted and separated on 12% SDS-PAGE gels. Western blot using an α-GFP antibody was performed to assess the amount of UL56 in the lysate (input) and captured (pellet). Loading of samples: 1.1% of total input and supernatant, 40% of total pellet sample.

Validation of the yeast two-hybrid hits BECN1, FNIP2, SENP6 and WR61 by co-immunoprecipitation with eGFP-UL56(1-234) was then re-attempted. Following optimization of the eGFP-UL56(1-234) immunoprecipitation, co-immunoprecipitation of BECN1, FNIP2, SENP6 and WDR61 was re-attempted using 0.5% Triton X-100 and benzonase (50 U/mL), an endonuclease with activity against both DNA and RNA, during lysis and incubation steps. To support benzonase activity, 2 mM $MgCl_2$ were used in the lysis/binding buffer. Furthermore, to increase the capture of eGFP-UL56(1-234) by the GFP-Trap A beads, lysate and beads were incubated for a longer time (75 min) before elution. Both eGFP alone and full-length eGFP-UL56(1-234), but not GAPDH, were enriched on the GFP-Trap beads. However, despite using the optimized protocol, none of the four interaction partners of UL56 identified in the yeast two-

hybrid screen and confirmed by GST pull-downs were found to co-immunoprecipitate with eGFP-UL56(-1234) (Figure 82).

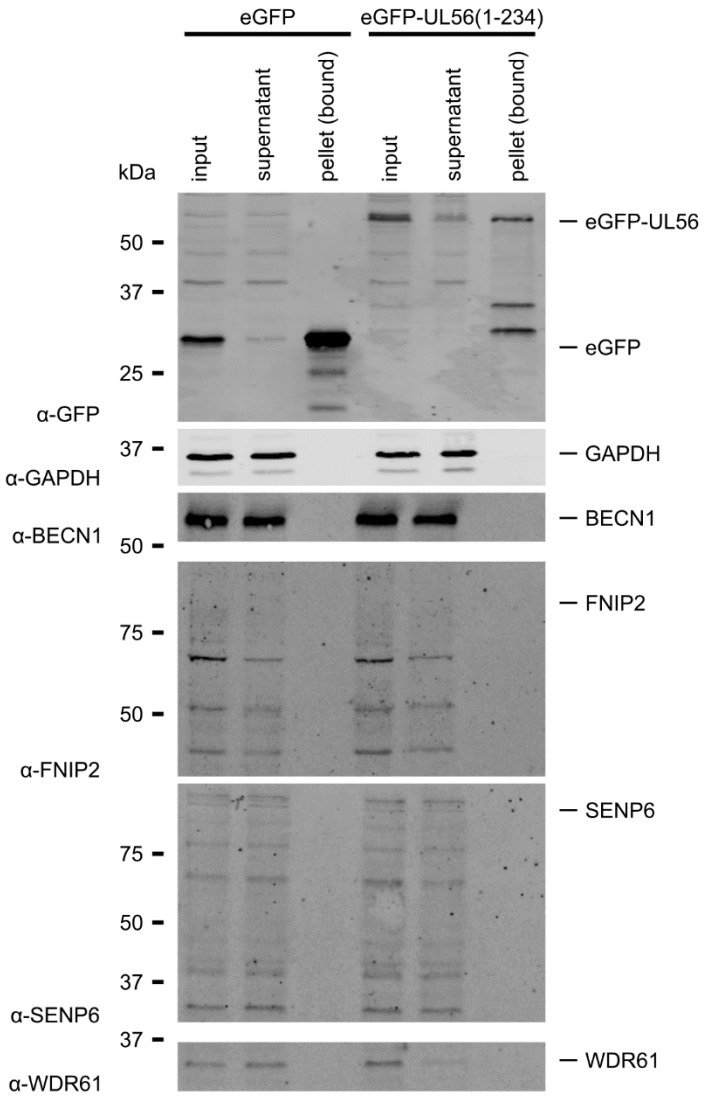


Figure 82 Co-immunoprecipitation experiment of eGFP-UL56(1-234)-eGFP with BECN1, FNIP2, SENP6 and WDR61 using optimized lysis and capture conditions. Cells were lysed in buffer containing 50 U/mL benzonase nuclease and 0.5% Triton X-100. The lysate was applied to GFP-Trap-A beads and incubated for 75 min. The beads were then washed, the bound samples eluted and separated on SDS-PAGE gels. Western blotting against GFP, GAPDH and the putative UL56 interaction partners was performed using the antibodies indicated. Loading of samples: 0.33% of total input and supernatant, 44% of total pellet sample.

6.4 Discussion

6.4.1 Yeast two-hybrid as a tool for identifying novel interactions between herpes simplex virus 1 and cellular proteins

Yeast two-hybrid assays are a well-established technique for the discovery of novel protein:protein interactions. However, the propensity of bait proteins to autoactivate reporter gene expression in this assay limits the range of bait proteins that can be used. Multiple HSV-1 proteins have been investigated in yeast two-hybrid screens previously. For example, Wu et al. performed a screen for interaction partners of the tegument protein UL14, but did not publish any data on autoactivation of the used construct (351). Similarly, full-length UL21 was used previously in a yeast two-hybrid screen by Muto et al. to identify novel cellular binding partners, but it is not evident that the authors conducted any autoactivation tests (257). On the contrary, full-length UL14 and UL21 were also investigated in a yeast two-hybrid screen assessing interactions between 13 tegument proteins, and both proteins were found to autoactivate (352). The same study also reported autoactivation of UL16, UL17, and UL48. In this thesis, it was demonstrated that two of three HSV-1 proteins investigated, gE and UL21, showed strong autoactivation in one of the most advanced and stringent yeast two-hybrid systems (Matchmaker® Gold). This finding is in agreement with the previously described autoactivation of full-length UL21 (352) and highlights the necessity of autoactivation tests before screening for cellular interaction partners is conducted.

For the third protein investigated in this thesis, UL56, two shorter constructs lacking the transmembrane domain were investigated, and only the truncated construct spanning residues 1-207 of HSV-1 UL56 was found not to autoactivate. Two studies previously used yeast two-hybrid assays to identify cellular interaction partners of herpes simplex virus 2 UL56. Ushijima et al. describe using UL56(1-210) in a yeast two-hybrid assay, but unfortunately, do not report on any autoactivation tests or show any of their yeast two-hybrid data (221). Koshizuka and colleagues used a construct spanning residues 1-217 and did not observe any autoactivation (233). However, they only assessed binding between their construct of UL56 fused to the binding domain of GAL4 (pGBKT7-UL56(1-217)) and an empty activation domain-containing yeast vector.

In agreement with their published data, the work described in this chapter did also not observe autoactivation for pGBKT7-UL56(1-215) and an empty pGADT7 vector. However, as soon as a

prey protein was fused to the activation domain (pGADT7-T), weak autoactivation of the UL56 construct spanning residues 1-215 could be detected. It is possible that the observed autoactivation arose due to residues lying within the transmembrane domain of UL56, which is predicted to span amino acids 211-233. Differences observed in autoactivation between HSV-1 and HSV-2 UL56 could also be explained by the slight differences of the protein sequences or the yeast two-hybrid systems employed. The overall sequence identity between HSV-1 and HSV-2 UL56 is around 63% (similarity: 72%). Despite these incongruences, construct pGBKT7-UL56(1-207), used in this work, showed no autoactivation and therefore, seemed suitable for yeast two-hybrid assays.

For the yeast two-hybrid screen with UL56(1-207), a normalized library was used, in which highly abundant cDNAs were removed to enhance the detection of uncommon transcripts and reduce the number of false positives. Still, more than 1000 positive yeast clones were found, of which only around 300 were picked, re-streaked and sequenced due to time constraints. Such a high number of positive clones is unusual and can probably be attributed to the high intrinsic disorder and potentially exposed hydrophobic regions of UL56 (345). In total, 18 proteins were found in two or more positive clones, but only one of these, PAWR, was present with multiple different open reading frames (independent clones).

The two previous yeast two-hybrid assays with UL56 identified human KIF1A (233) and several E3 ubiquitin ligases, NEDD4, NEDD4L, HECW1, HECW2, WWP1, WWP2 and ITCH (221), as interaction partners of UL56. None of these genes were among the positive clones sequenced in the screen described in this work. Given the high number of positive clones and low number of genes detected in multiple clones for UL56(1-207), this disparity could simply be due to chance. Alternatively, although the HSV-1 and HSV-2 UL56 are largely conserved, even slight differences in sequence could lead to differential interactomes. However, since Koshizuka and Ushijima do not report on the number of positive colonies screened, no conclusive comparison of the screen presented in this work with the previously published screens can be made.

On the whole, these results manifest the suitability of yeast two-hybrid screens for the identification of cellular interaction partners of the viral protein UL56. However, it is also apparent that thoughtful selection and rigorous testing of the bait construct is mandatory. Even though the yeast two-hybrid assay can be considered a high-throughput method, it is labor-

intensive, and due to intrinsic restrictions of the yeast system such as the high false-positive rate or missing post-translational modifications, careful validation of hits should be performed.

6.4.2 Biochemical validation of yeast two-hybrid hits yields ambivalent results

The exact role of UL56 during the HSV-1 life cycle is unclear. Thus, not only the two hits with immediate links to intracellular trafficking, PSD3 and BECN1, but also additional potentially interesting putative interactions partners of UL56, PAWR, DDX59, SENP6, FNIP2, WDR61, were selected for biochemical follow-up experiments. Purified GST-UL56 interacted with four of these proteins when expressed in wheat germ lysate, BECN1, FNIP2, SENP6 and WDR61, confirming the interactions. For SENP6, binding of UL56 to both truncation constructs was detected, indicating that UL56 might interact with the C-terminal catalytic domain of SENP6. Proteins DDX59, PAWR and PSD3 were not found to bind to GST-UL56(1-207). It is possible that for these hits, the interaction with UL56 in the yeast two-hybrid assay was mediated by additional proteins present in yeast cells, or that the hits are non-specifically binding to UL56 or the GAL4 binding domain in yeast.

The pull-down experiments were not impeded by the nucleic acid contamination of GST-UL56(1-207) noticed during the purification process and the hence lower apparent concentration of the bait protein. However, interactions with such nucleic acid-contaminated protein could be simply charge-mediated instead of being specific for UL56. Therefore, the pI of the confirmed prey proteins was calculated and revealed that all four myc-tagged constructs have a predicted pI lower than 6 (BECN1: 4.8, FNIP2: 5.1, SENP6: 5.1, WDR61: 4.9). This finding indicates that the demonstrated binding between UL56 and the prey proteins is not likely to be mediated by nucleic acid, but rather due to direct protein:protein interactions.

To further validate these interactions in a cellular context, co-immunoprecipitation experiments were conducted. At first, a truncated construct (residues 1-207) was used and found not to interact with BECN1, SENP6 or WDR61. As described previously by Koshizuka and colleagues (224, 233), this version of UL56, lacking its transmembrane domain, localized to the nucleus. In contrast, full-length UL56 is reported to localize predominantly to perinuclear membrane regions and the trans-Golgi network in both transfected in infected cells (220). It has been suggested that in the absence of the transmembrane region, the arginine-rich region of UL56 serves as an alternative nuclear localization signal, explaining this observation (224). Since proteins BECN1

and WDR61 are known to localize to the cytoplasm (353, 354), they may have been unable to interact with UL56 sequestered to the nucleus. However, SENP6 is reported to be localized mainly in the nucleoplasm (348) and was also not found to interact with UL56(1-207), suggesting that the interaction between UL56 and SENP6 observed in vitro does not occur in HEK293T cells. Therefore, as a next step, full-length UL56 (residues 1-234), which would not be expected to reside in the nucleus, was tested for interaction with BECN1, FNIP2, SENP6 and WDR61.

Both N- and C-terminally eGFP-tagged UL56 constructs were assessed for their suitability for co-immunoprecipitation, but only N-terminally tagged full-length UL56 expressed well. Most likely, for UL56(1-234)-eGFP, the GFP tag interferes with expression due to the membrane insertion of the C-terminal transmembrane region. As expected, eGFP-UL56(1-234) exhibited cytoplasmic localization. Nevertheless, neither BECN1, SENP6, WDR61, nor FNIP2 were found to bind to this construct of UL56, although BECN1, WDR61 and FNIP2 (355) are reported to be expressed in the cytoplasm.

Altogether, despite four of the putative binding partners of UL56 identified in the yeast two-hybrid screen could be confirmed by GST pull-down, none of them co-immunoprecipitated with eGFP-tagged UL56. It was notable that the α-FNIP2 and α-SENP6 antibodies used for Western blotting showed high background and did not detect distinct bands of appropriate size for their respective targets. Therefore, it might be possible that the interaction between UL56 and FNIP2 or SENP6 occurs in cells, but could not be confirmed with the antibodies tested due to poor recognition of the target antigen and/or low antigen expression levels. On the other hand, the antibodies against BECN1 and WDR61 worked well, but no binding to UL56 was observed. Conceivably, the lysis conditions required to extract sufficient eGFP-tagged UL56(1-234) from cellular membranes could be too harsh and disrupt the interactions with the pull-down validated hits. However, Ushijima et al. confirmed interactions between UL56 and the ubiquitin ligases NEDD4 (221) or ITCH (222) by co-immunoprecipitation using a lysis buffer with comparable ionic strength and a similarly high concentration of detergent (1% Nonident P-40) as the one used in this work. Thus, it seems most likely that the interaction between UL56 and BECN1 or WDR61 found in the yeast two-hybrid screen and in GST pull-downs is truly not taking place in transfected HEK293T cells.

It has been suggested that UL56 is phosphorylated, although the exact phosphorylation sites were not determined (224). One of the intrinsic limitations of the yeast two-hybrid system is the lack of tyrosine phosphorylation in yeast. Protein UL56 contains six tyrosine residues, and due to its disordered nature, these easily could be phosphorylated in vivo. Missing phosphorylation of UL56 expressed in yeast or in *E. coli* could lead to aberrant interactions. In addition, the largely unfolded protein chain of UL56 exposes highly charged patches such as the arginine-rich region (residues 75-100), as well as smaller, hydrophobic stretches of residues. These regions could potentially be shielded when UL56 is in a cellular context, illustrating an alternative explanation for the discrepancy observed between the interaction results obtained by the yeast two-hybrid assay or the GST pull-downs and the co-immunoprecipitation experiments.

Finally, it is possible that interactions between UL56 and the proteins identified as novel cellular binding partners occur only during of infection of mammalian cells, for example due to different expression levels of the prey proteins and/or different subcellular localization in the presence of other viral proteins. Since there is currently no α-UL56 antibody available, co-immunoprecipitation experiments of untagged UL56 during transfection or HSV-1 infection of mammalian cells cannot be performed. Construction of a HSV-1 virus encoding tagged UL56 would enable testing of the interactions identified by yeast two-hybrid in the context of infection.

6.4.3 Summary

In this chapter, I investigated whether the three HSV-1 proteins, gE, UL21 and UL56, could be used as bait proteins in yeast two-hybrid assays. Only UL56 was suitable for yeast two-hybrid analysis as it did not autoactivate. A screen to identify putative interaction partners of UL56(1-207) was performed. Multiple interesting hits, including two proteins with direct links to intracellular membrane trafficking, BECN1 and PSD3, were found. I confirmed four of the selected hits, proteins BECN1, FNIP2, SENP6 and WDR61, to bind to purified GST-tagged UL56 in pull-down experiments. To validate these hits in a cellular context, I optimized a cell lysis and co-immunoprecipitation protocol and showed that nuclease treatment of cell lysates is crucial for efficient extraction and enrichment of eGFP-tagged UL56. However, even using this optimized protocol, none of the four hits were found to co-immunoprecipitate with eGFP-UL56. Whereas it is possible that these interactions do occur in the context of infection and/or in cell types naturally infected by HSV-1 (keratinocytes, fibroblasts and neurons), it is also possible

that they are not biologically relevant. Therefore, an alternative method for the discovery of potential cellular interaction partners of HSV-1 UL56 had to be explored.

7 IDENTIFICATION AND VALIDATION OF UL56 INTERACTION PARTNERS USING QUANTITATIVE PROTEOMICS

Yeast two-hybrid assays are an established and very effective method for the identification of protein:protein interactions, but a high number of false positive hits can often be observed in even the most carefully designed systems (356). For example, in yeast two-hybrid screens, interactions can be observed between proteins that would normally be found in different cellular compartments. Missing or incorrect post-translational modification such as glycosylation or phosphorylation is another problem inherent to standard yeast two-hybrid interaction assays (357). In Chapter 6, I described how several putative interaction partners of the HSV-1 protein UL56 were found in a yeast two-hybrid screen. Although four of these binding partners could be confirmed in GST pull-down assays as interactors of UL56, none of these four hits were found to co-immunoprecipitate with UL56, casting doubt on the physiological relevance of the identified interactions.

Therefore, I performed a second, independent screen for UL56 interaction partners using a SILAC-based, high-throughput quantitative proteomics approach. In this screen, cellular proteins such as the human trafficking protein particle (TRAPP) complex, the Golgi-associated PDZ and coiled-coil motif-containing protein (GOPC), and several E3 ubiquitin ligases, were identified as UL56-interacting proteins (section 7.1). Multiple putative interaction partners could be validated by both in vitro GST pull-down assays and by co-immunoprecipitation (section 7.2). Since the interaction between UL56 and GOPC seemed to be strong and direct, I also conducted pull-down and co-immunoprecipitation assays to identify the regions of UL56 and GOPC required for binding (section 7.3).

7.1 Characterization of the UL56 interactome

7.1.1 Description of SILAC method for the detection of protein:protein interactions

Proteomic research has benefitted greatly from the development of quantitative mass spectrometry techniques. Mass spectrometry is inherently non-quantitative, but enzymatic, metabolic or chemical labeling of samples before chromatography, or label-free methods

measuring ion intensity or spectral counts allow for quantification and comparison of different samples (358). Stable isotope labeling by amino acids in cell culture (SILAC) is a quantification method based on the introduction of metabolic labels. Cells are grown in media supplemented with specific heavy atom-labeled amino acids, most often lysine and arginine residues labeled with ^{13}C, ^{2}H and ^{15}N isotopes. Combinations of different isotopes allow for multiplexing of experiments, enabling, for example, the comparison of drug treatments or the effect of inducible gene overexpression.

For the discovery of novel protein:protein interactions, co-immunoprecipitation followed by mass spectrometry is a particularly powerful approach (359) (Figure 83). A control protein and a protein of interest, possibly tagged, are expressed in cells growing in light (unlabeled, Arg0-Lys0) and heavy (labeled, Arg10-Lys8) SILAC medium and then immunoprecipitated using protein- or tag-specific antibodies. Light and heavy samples can be mixed before liquid chromatography/mass spectrometry (LC-MS/MS). Multiple proteins of interest or the effect of mutations in the protein of interest can be compared by introducing a "medium" label (Arg6-Lys4). Comparison of sample ratios (heavy over light, medium over light, etc.) will identify proteins binding specifically to the protein of interest since enrichment of interaction partners will be observed during co-immunoprecipitation with the protein of interest, but not with the control. Ideally, experiments are performed in triplicate with swapped labels (if medium and heavy labels are used) to account for any differences in protein expression in cells grown in the labeled medium (359).

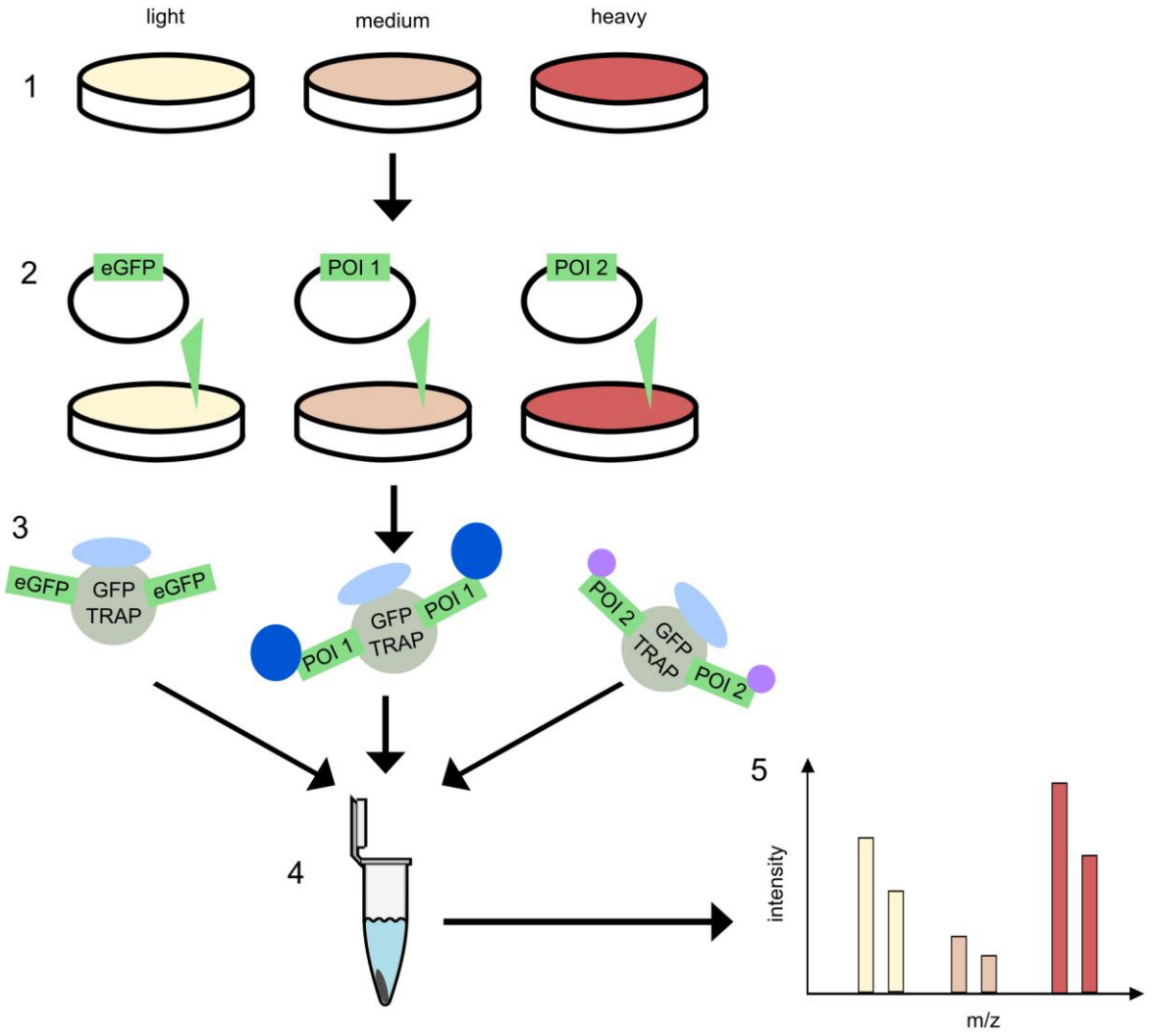


Figure 83 Schematic of a SILAC-based mass spectrometry experiment for the identification of protein:protein interactions. (1) Cells are grown in labeled medium. (2) A control plasmid (eGFP) and proteins of interest (POI) tagged with eGFP are transfected into the cells. (3) Cells are lysed and an immunoprecipitation using GFP TRAP resin is performed. Some cellular proteins will bind non-specifically (light blue), whereas others will specifically bind to the proteins of interest (dark blue, purple). (4) Light, medium and heavy samples are combined and (5) submitted for quantitative mass spectrometry.

In comparison to yeast two-hybrid methods, post-translational modifications and cellular localization of the protein of interest and potential interaction partners are retained when SILAC methods are used. The quantitative dimension of a SILAC interaction screen also allows for easy distinction of contaminants and non-specifically interacting host cell proteins from bona fide interaction partners (359). In addition, SILAC offers higher accuracy and reproducibility over other labeling techniques used for quantification (358).

7.1.2 Expression and immunoprecipitation of UL56 using SILAC-labeled media

Co-immunoprecipitation followed by SILAC quantitative mass spectrometry was performed to identify novel interactions partners of HSV-1 UL56. For the experiment, human embryonic kidney cells (HEK293T) were grown in unlabeled (light, Lys0-Arg0) and labeled SILAC medium (medium, Lys4-Arg6; heavy, Lys8-Arg10) and seeded into 9 cm dishes. Cells were transiently transfected with either eGFP as a control (light) or UL56 (medium and heavy). Two constructs, full-length eGFP-UL56(1-234) and UL56 without its transmembrane domain, UL56(1-207)-eGFP, were used. Cells lysis and co-immunoprecipitation using GFP-Trap A beads were performed according to the optimized protocol described in Chapter 6. To ensure equal amounts of protein were incubated with the beads during the immunoprecipitation step, a BCA assay was performed and the protein concentrations of the lysates were normalized. Bound proteins were eluted, and light, medium and heavy samples were mixed in a 1:1:1 ratio and submitted for mass spectrometry. Experiments were performed in triplicate with labels swapped in experiments 2 and 3 (Table 18) to minimize the chances of identifying false-positive interaction partners of UL56, to increase coverage of the cellular proteome, and to circumvent bias due to the expression in labeled medium.

Table 18 Experimental setup of biological replicates for SILAC interaction screen. Constructs transfected into HEK293T cells growing in the indicated medium are shown.

Medium	Experiment 1	Experiment 2	Experiment 3
L (Lys0-Arg0)	eGFP	eGFP	eGFP
M (Lys4-Arg6)	eGFP-UL56(1-234)	UL56(1-207)-eGFP	UL56(1-207)-eGFP
H (Lys8-Arg10)	UL56(1-207)-eGFP	eGFP-UL56(1-234)	eGFP-UL56(1-234)

Samples were analyzed by Western blot to confirm equal loading of samples across all experiments and to evaluate capture of the expressed proteins (Figure 84). Both UL56 constructs were expressed in cells grown in both labeled media, although slightly higher expression levels were observed when heavy medium (H) was used. All three proteins (eGFP, eGFP-UL56(1-234) and UL56(1-207)-eGFP) were captured efficiently. However, for all proteins, some degradation could be observed and this was most pronounced for the truncated UL56 construct. Nevertheless, across all three biological repeats, non-degraded protein was visible for all constructs.

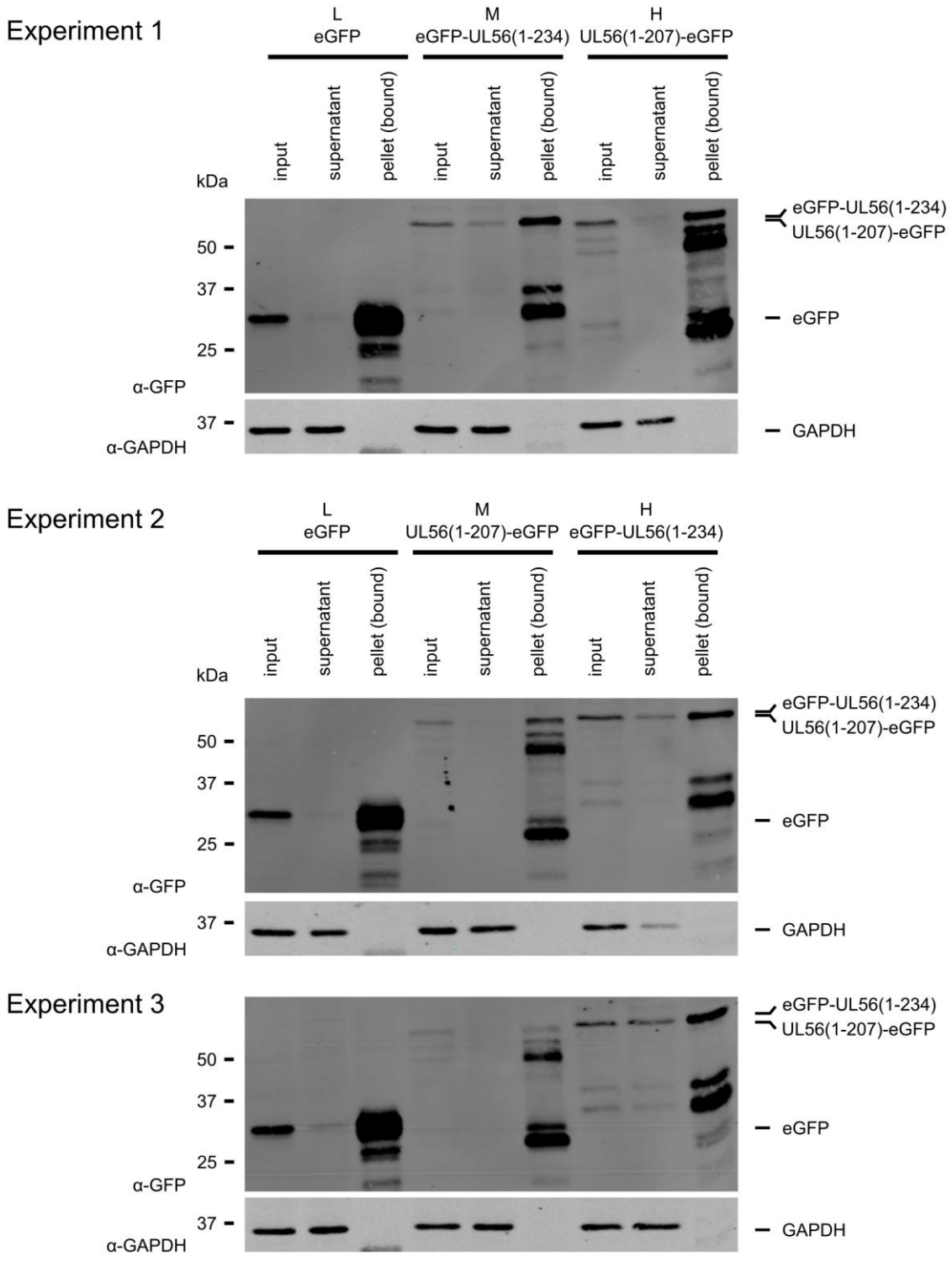


Figure 84 Western blot evaluation of SILAC samples used for quantitative mass spectrometry. Three biological replicates of the experiment with labels swapped in experiments 2 and 3 vs. experiment 1 were performed. Constructs were expressed in HEK293T cells growing in light (L), medium (M) and heavy (H) medium as indicated. Cells were lysed for 70 min and co-immunoprecipitation using GFP-Trap A beads was performed using equal protein concentrations per sample. After 40 min incubation, beads were washed twice and bound proteins were eluted by boiling. Samples were then separated on 12% SDS-PAGE gels and visualized using anti-GFP and anti-GAPDH antibodies. Loading of samples (for all experiments): 0.5% of total input and supernatant, 22% of total pellet sample.

7.1.3 Analysis of mass spectrometry data

Raw mass spectrometry data were analyzed using MaxQuant and statistical analysis of the MaxQuant output was performed using Perseus. Normalized SILAC ratios (eGFP-UL56(1-234)/eGFP, UL56(1-207)-eGFP/eGFP, eGFP-UL56(1-234)/UL56(1-207)-eGFP were used for the analysis. Non-specifically binding proteins, which are enriched in both control and protein of interest samples to equal amounts, should have $\log_2$ SILAC ratios of 0. Proteins enriched for the control, e.g. environmental contaminants, have negative $\log_2$ SILAC ratios, whereas true interaction partners of the protein of interest have positive $\log_2$ SILAC ratios. Therefore, in an ideal experiment, the logarithmized ratios should have a normal distribution centered around 0. Such distribution of ratios was observed for all three biological repeats (Figure 85A).

In total 1409 proteins were identified, of which 918 (for eGFP-UL56(1-234) vs. eGFP) or 924 (for UL56(1-207) vs. eGFP) were also quantified (Figure 85B, C). Close to 50% (454 proteins) of all quantified proteins were found in all three biological repeats, and 201 proteins (eGFP-UL56(1-234) vs. eGFP) or 200 proteins (UL56(1-207) vs. eGFP) were found in exactly two of three of the experiments. Taken together, 71% of all quantified proteins were identified in at least two biological repeats, indicative of robust transfection, co-immunoprecipitation and little bias due to SILAC labeling of the cell growth medium.

For the identification of cellular proteins binding to UL56, $\log_2$ SILAC ratios of UL56(1-234) or UL56(1-207) over eGFP were analyzed separately. Since this study is not concerned with proteins binding differentially to full-length vs. truncated UL56, SILAC ratios for UL56(1-234)/UL56(1-207) were not analyzed further. The samples were filtered for proteins quantified in at least two biological repeats. Significantly enriched (or depleted) proteins were then identified by performing a two-sided one-sample t-test against a $\log_2$ ratio of 0 (no change in abundance) with a threshold p-value of 0.05. Of the proteins identified to be significantly more or less abundant when co-captured with UL56, those with a greater than two-fold increase in abundance ($\log_2$ fold-change = 1) were selected for further investigation ("hits") (Figure 86). In total, 60 such proteins were identified for UL56(1-207) (Table 19) and 73 hits for UL56(1-234) (Table 20).

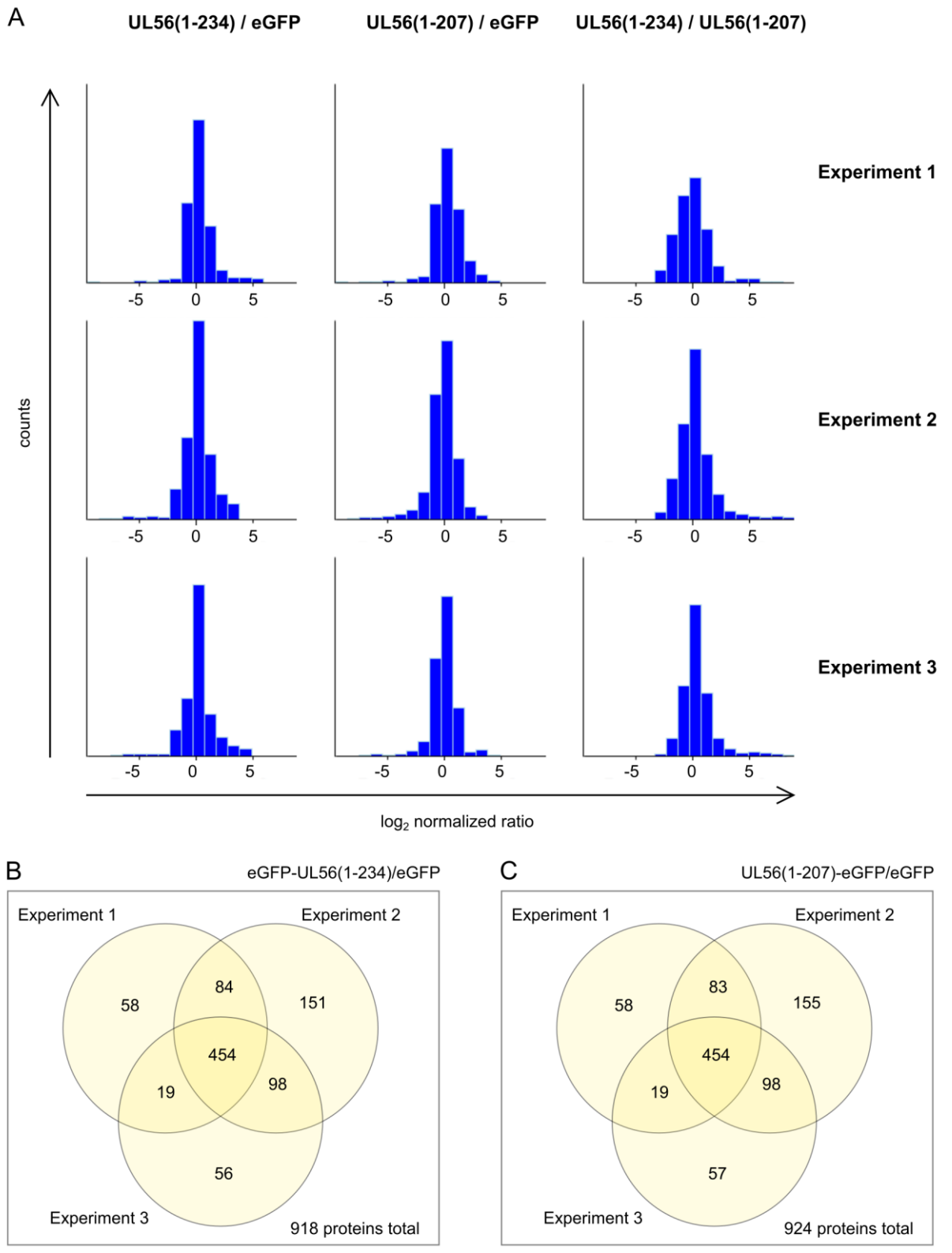


Figure 85 SILAC ratio distributions and protein quantification statistics of quantitative proteomics screen with UL56. (A) Frequency histograms of SILAC ratio distributions for all three biological repeats. (B, C) Venn diagrams showing numbers of proteins quantified across the three biological repeats for eGFP-UL56(1-234) (B) and UL56(1-207)-eGFP (C) when compared to the eGFP control.

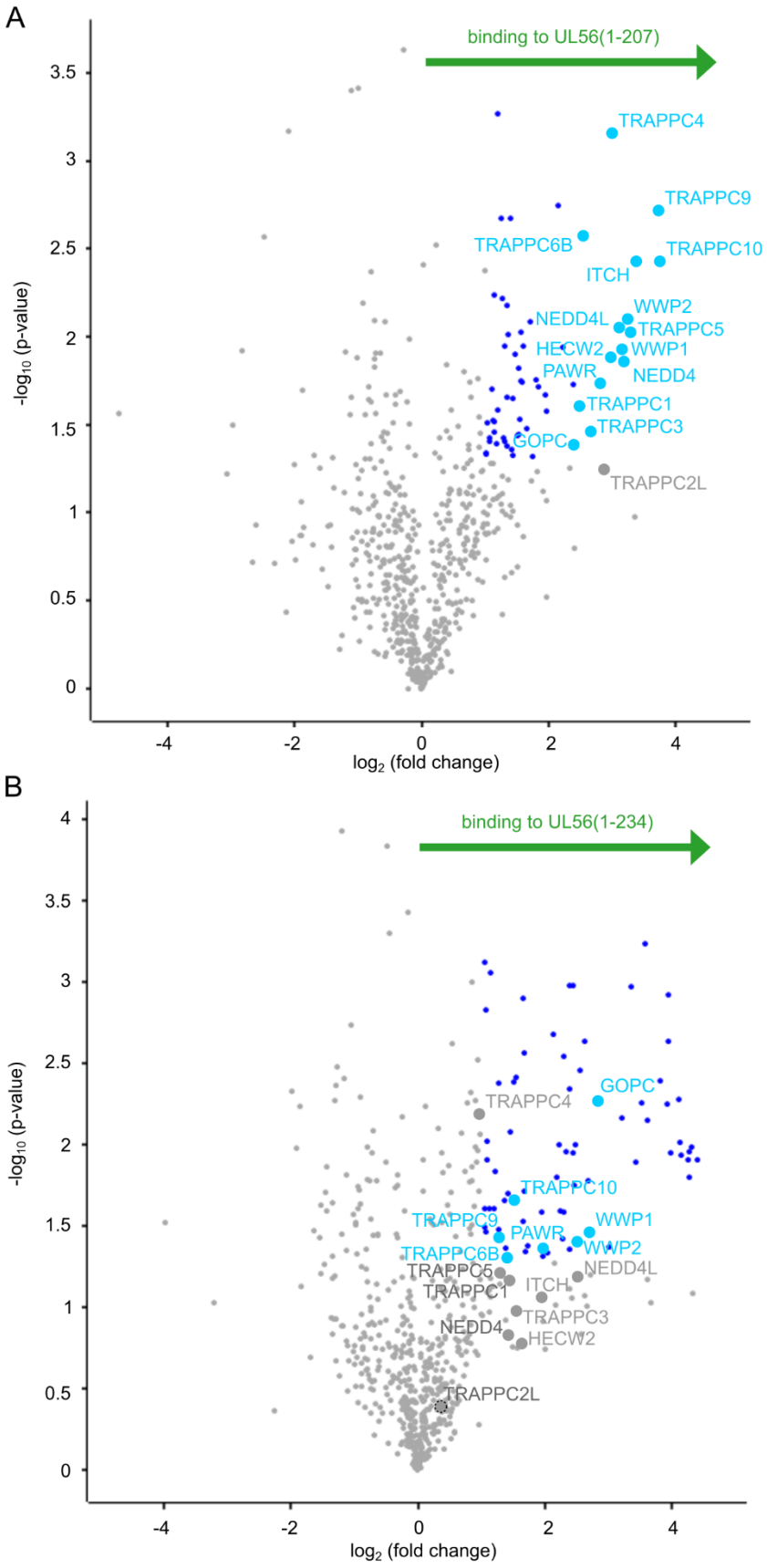


Figure 86 Potential interaction partners of UL56 identified by SILAC quantitative proteomics. Volcano plots of proteins quantified for UL56(1-207) (A) and UL56(1-234) (B) showing the $\log_2$ fold-change versus the eGFP control and significance of the observed fold-change. Proteins with $p < 0.5$ and a fold-change > 2 ("hits") are represented as dark blue dots. (A) Top hits of cellular proteins binding to UL56(1-207) are shown as light blue dots and include GOPC, PAWR, multiple components of the human trafficking particle complex (TRAPPC), as well as several E3 ubiquitin ligases

(ITCH, NEDD4, NEDD4L, HECW2, WWP1 and WWP2). (B) Some hits identified as binding partner of truncated UL56(1-207) also bind to full-length UL56 (light blue dots), whereas other hits found for UL56(1-207) are still quantified for full-length UL56, but do not bind strongly (fold-change < 2) or the quantification of their change in binding is not significant ($p > 0.5$) (grey labeled dots).

Table 19 Hits from SILAC interaction screen for construct UL56(1-207)-eGFP. Entries are sorted by descending $\log_2$ fold-change and identified by gene name. Peptides: Total number of peptide sequences of the protein group identified by the gene name. Intensity: Total intensity (extracted ion current) of all isotopic patterns in the label cluster (light, medium, heavy).

Gene name	**$\log_2$ fold-change UL56(1-207)/eGFP**	**-$\log_{10}$ p-value**	**Peptides**	**Sequence coverage [%]**	**Intensity**
TRAPPC10	3.77181	2.42358	17	14.9	225090000
TRAPPC9	3.72775	2.70467	21	21.5	332970000
ITCH	3.38884	2.41839	35	42.4	4536700000
TRAPPC5	3.28276	2.02546	4	21.8	99442000
WWP2	3.24321	2.09336	16	19.3	614250000
NEDD4	3.18504	1.86529	23	28.3	711050000
WWP1	3.15448	1.93177	18	28.8	555620000
NEDD4L	3.12068	2.04798	24	29.2	1185800000
HECW2	3.00386	1.88542	4	3.1	15690000
TRAPPC4	2.9981	3.15367	8	48.3	111880000
PAWR	2.79235	1.73384	4	15.6	100950000
TRAPPC3	2.67302	1.45742	7	36.1	342780000
TRAPPC6B	2.52889	2.56846	3	22.8	81040000
TRAPPC1	2.4762	1.61463	4	47.4	92898000
PNO1	2.38741	1.73025	5	26.6	107790000
GOPC	2.36628	1.38787	21	50.2	3748600000
BYSL	2.23174	1.93835	13	34.3	212960000
USP9X	2.15427	2.74903	30	13.2	264580000
TSR1	1.96613	1.57454	12	17	309340000
DIMT1	1.95207	1.66987	7	24	37168000
DHX30	1.84721	1.71361	26	23.7	459040000
MRPS22	1.813	1.75616	5	14.4	48798000
RIOK1	1.75466	1.31555	7	14.4	46410000
CDK9	1.71634	2.08813	5	14	31085000
DDX21	1.66047	1.48017	11	17.6	183310000
RPL36	1.60469	1.94632	4	30.5	189020000
DDX54	1.58594	1.73856	2	2.4	28127000
RPS24	1.56747	1.75006	4	26.7	435890000
GTPBP4	1.56556	2.02761	7	12.6	89367000
MRPS9	1.55785	1.52691	2	7.1	20887000
RBM14	1.53345	1.82328	5	9.3	64183000
MRPS23	1.5334	1.44443	4	26.3	41567000
STAU1	1.51998	1.43822	11	24.9	127290000

Gene name	log_2 fold-change UL56(1-207)/eGFP	$-log_{10}$ p-value	Peptides	Sequence coverage [%]	Intensity
RPL10	1.47912	1.90155	7	33.7	364300000
MYBBP1A	1.43811	1.65202	11	9.9	56666000
YBX1	1.43702	1.32725	8	37.3	150080000
GNL3	1.42608	1.35997	9	18.1	278030000
RPL4	1.39922	2.67209	15	33	1531100000
RANGAP1	1.37107	2.01127	13	26.7	152720000
TEX10	1.35393	1.38096	8	10.5	84316000
RPL15	1.35243	1.65689	7	32.8	450940000
RANBP2	1.3482	2.17456	19	9.4	66019000
DHX9	1.31738	1.40615	23	20.2	523540000
NOL9	1.31268	1.94549	3	4.3	21375000
RPS10	1.28764	1.4276	5	29.1	541040000
WRNIP1	1.27162	2.21692	9	15.6	81170000
RPS5	1.25632	2.67364	9	37	1541000000
RBM39	1.20554	1.58387	5	26.9	20068000
ZC3HAV1	1.1936	3.2709	3	6.3	18426000
CCDC124	1.18977	1.39304	3	16.1	92298000
RPL6	1.1482	1.45907	10	28.5	645190000
RPL27	1.14646	1.51497	6	43.4	1591800000
RPS21	1.14461	2.23833	5	46.9	665200000
RPLP2	1.12552	1.5246	7	85.2	720360000
LAS1L	1.10959	1.6991	9	14.6	73968000
EIF6	1.08143	1.42702	5	27.3	60972000
RPL18	1.06842	1.40468	6	39.6	776320000
RPL37A	1.03157	1.51324	3	55.9	187750000
XRN2	1.02247	1.33109	4	6.2	56612000
AAAS	1.0119	1.33797	3	9.8	18001000

Table 20 Hits from SILAC interaction screen for construct eGFP-UL56(1-234). Entries are sorted by descending log_2 fold-change and identified by gene name. Peptides: Total number of peptide sequences of the protein group identified by the gene name. Intensity: Total intensity (extracted ion current) of all isotopic patterns in the label cluster (light, medium, heavy).

Gene name	log_2 fold-change UL56(1-234)/eGFP	$-log_{10}$ p-value	Peptides	Sequence coverage [%]	Intensity
KPNB1	4.41734	1.90192	27	34.9	9014400000
UQCRC1	4.31352	1.98578	13	30	2281200000
PHB2	4.27652	1.80143	14	52.8	2240200000
COX4I1	4.27441	1.95436	6	34.3	343440000
TMED1	4.27049	1.90421	5	21.1	128250000
TNPO1	4.15358	1.93158	20	26.3	976510000
PHB	4.14521	2.00886	13	39.7	2345100000
CYC1	4.11828	2.27416	7	22.8	750100000
WWOX	3.98478	1.9504	6	20.1	156960000
MT-CYB	3.95801	2.63729	2	5	59593000

Gene name	$\log_2$ fold-change UL56(1-234)/eGFP	$-\log_{10}$ p-value	Peptides	Sequence coverage [%]	Intensity
UQCRC2	3.95516	2.91984	14	43.4	1803600000
UQCR10	3.93325	2.24894	2	38.1	181070000
ARL10	3.82793	2.38864	4	20.5	155370000
UQCRB	3.63145	2.15069	6	43.2	100310000
IPO7	3.58905	3.23512	15	15.7	867880000
COX7A2	3.52512	2.25616	2	25.3	136660000
MT-CO2	3.44424	1.88969	4	18.9	385890000
ALDH3A2	3.36515	2.96846	10	21.4	251750000
IMMT	3.21062	2.15946	14	22.6	263760000
CHCHD3	3.0157	1.36959	5	21.1	76796000
GOPC	2.8161	2.25916	21	50.2	3748600000
WWP1	2.71276	1.45259	18	28.8	555620000
STT3A	2.68715	1.77382	4	7	44226000
WRNIP1	2.62749	2.63414	9	15.6	81170000
DAD1	2.5505	2.45879	3	37.6	57983000
HACD3	2.48524	1.99769	5	14.5	85620000
WWP2	2.47765	1.39459	16	19.3	614250000
STOML2	2.45931	1.74747	4	17.1	50611000
C1QBP	2.45106	2.97973	4	14.9	119690000
RPN1	2.4471	1.95131	12	21.6	189810000
CSNK2B	2.40186	2.97505	5	21.4	272750000
SUMO1	2.39966	2.33911	2	48.3	41911000
NDUFB10	2.38538	1.35303	3	21.7	22655000
MYO6	2.34566	1.9564	9	8.9	38571000
ABCD3	2.30195	1.58147	6	10.8	54400000
NDUFV1	2.29795	2.5435	3	19.7	13856000
RPS27A	2.27493	1.41959	9	57.1	2673400000
DDOST	2.23935	1.59218	5	12.8	119110000
VDAC2	2.22001	1.99828	6	27.3	166000000
STXBP4	2.19043	1.797	7	16.4	119680000
RANGAP1	2.13547	2.67354	13	26.7	152720000
ERLIN2	2.04269	1.33227	3	10.7	40915000
PAWR	2.00666	1.34616	4	15.6	100950000
VDAC1	1.98067	1.31172	6	23.7	36838000
NUP205	1.94441	1.58142	7	4	39493000
ST13	1.73882	1.37986	3	25.3	24902000
IPO8	1.69987	1.34077	6	6.8	20932000
ATAD3A	1.67302	2.56402	16	30.5	446420000
EMC2	1.67285	1.7094	3	11.8	18871000
AAAS	1.65487	1.52818	3	9.8	18001000
EMD	1.65475	2.90134	8	42.5	406380000
NUP93	1.55587	2.41341	10	14.3	77899000
DNAJA2	1.5044	2.3819	2	17.2	69187000
TRAPPC10	1.5022	1.65765	17	14.9	225090000
ODR4	1.45589	2.07608	4	14.9	13829000

Gene name	log_2 fold-change UL56(1-234)/eGFP	$-log_{10}$ p-value	Peptides	Sequence coverage [%]	Intensity
GEMIN4	1.41626	1.69861	3	3.5	34905000
TRAPPC6B	1.39524	1.3109	3	22.8	81040000
HSPA8	1.38941	1.36462	28	42.3	9221200000
TECR	1.37276	1.6586	4	13.3	52244000
FAR1	1.2795	1.47952	4	10.1	25210000
SEC61A1	1.27118	2.37332	4	8.3	111450000
TRAPPC9	1.26775	1.43908	21	21.5	332970000
HUWE1	1.21086	1.83048	9	3	37770000
ATP2A2	1.21021	1.60618	5	7.2	23960000
MATR3	1.15383	3.05603	10	14.5	130400000
DNAJA1	1.12478	1.6032	9	28.2	277060000
SLC16A1	1.08843	2.01927	3	9	53345000
USMG5	1.08801	1.90459	2	43.1	56495000
NAP1L4	1.0801	2.82832	4	43.9	72881000
CTNNBL1	1.0797	1.46157	4	7.9	14397000
DHCR7	1.05879	3.11947	3	6.5	41909000
SLC25A3	1.05107	1.48897	2	5.8	251410000
IPO4	1.04857	1.60721	6	6	48616000

7.1.4 Bioinformatic analysis

In order to predict potential direct and functional associations between the hits identified as binding partners of UL56, a bioinformatic interaction network analysis using STRING was carried out for both experiments separately. The putative interaction partners of UL56 found using the truncated construct, UL56(1-207), clustered into three distinct sub-networks (Figure 87). The biggest cluster was formed by components of the ribosome and proteins interacting with ribosomal proteins. A second, isolated cluster consisted of subunits of the trafficking particle complex (TRAPPC). The third, weaker cluster was formed by four E3 ubiquitin ligases and the deubiquitinase USP9X.

For full-length UL56, clustering was less discrete and no ribosomal protein cluster could be observed. Nevertheless, three interconnected sub-networks were found (Figure 88). The most distinct cluster consisted of proteins associated with mitochondrial membranes and proteins involved in cellular respiration. Co-chaperones, a subunit of the translocon (SEC61A1), and subunits of the N-oligosaccharyl transferase (OST) complex, an ER-localized protein complex associated with the translocon and involved in co-translational glycosylation protein synthesis (360), formed a second cluster. Lastly, proteins involved in nuclear pore complex formation and

nuclear protein import clustered together. A very small fourth, separated cluster consisting of three subunits of the human TRAPP complex was also observed.

Hits were also screened against a database containing contaminants commonly found in studies combining bead-based affinity purification with mass spectrometry (AP-MS), the CRAPome (Contaminant Repository for Affinity Purification) database. During the selection of putative interaction partners for follow-up studies (section 7.1.5), the CRAPome profiles were taken into account and hits with few occurrences and low spectral counts were preferred.

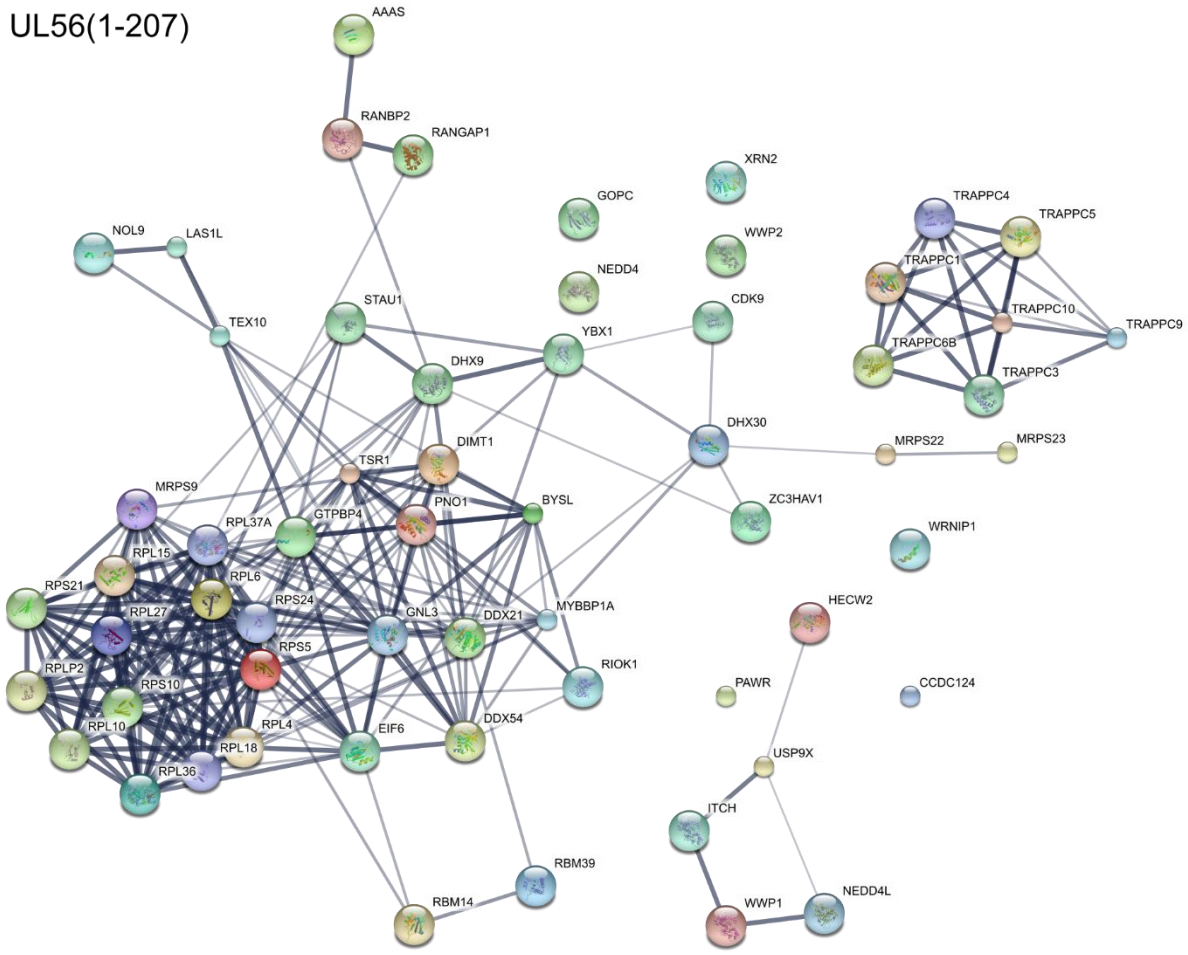


Figure 87 Analysis of the potential interaction network of proteins identified as binding partners of UL56(1-207) using STRING. Sub-networks observed were built by proteins forming or related to the ribosome (left), the human trafficking particle complex TRAPPC (top right), and proteins involved in ubiquitination (bottom center).

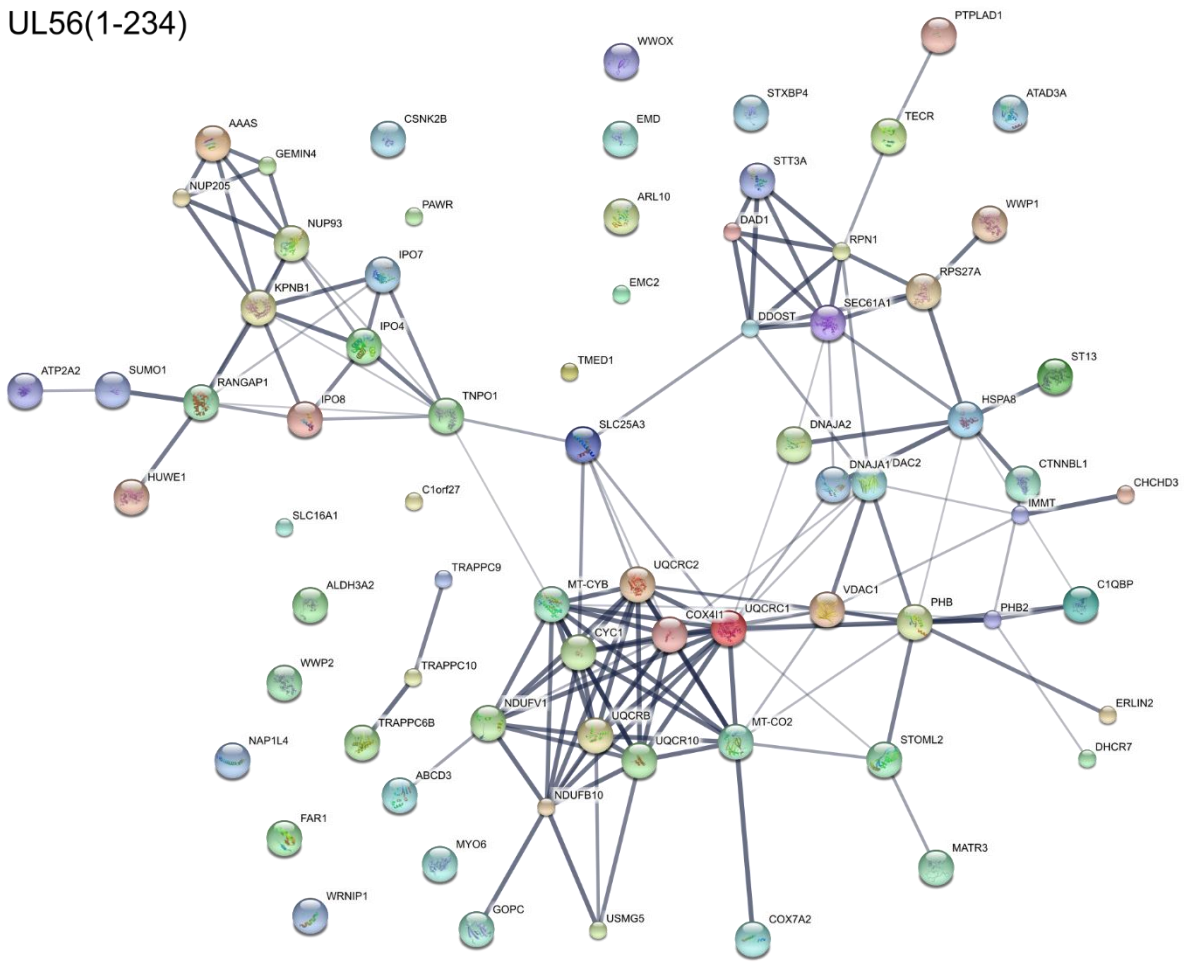


Figure 88 Analysis of the potential interaction network of proteins identified as binding partners of UL56(1-234) using STRING. Proteins clustered into three sub-networks: mitochondrial proteins or proteins related to cellular respiration (bottom center), proteins forming the nuclear pore complex or associated with nuclear transport (top left) and proteins associated with protein re-folding or subunits of the ER-associated N-oligosaccharyl transferase (OST) complex (right top and center). Three components of the human TRAPP complex form a small fourth cluster.

7.1.5 Selection of hits for follow-up

In the SILAC proteomics screen, numerous cellular proteins were identified as potential interaction partners of UL56. Therefore, hits were examined to find proteins which UL56 could exploit in order to facilitate envelopment or trafficking of HSV-1 viral particles. The proteins described in the following paragraphs and selected for biochemical validation are highlighted in Figure 86.

Multiple subunits of the human TRAPP complex (TRAPPC) were found as top hits in the SILAC screen using UL56(1-207). In addition, three TRAPPC subunits, TRAPPC6B, TRAPPC9 and TRAPPC10, were identified as potential binding partners of full-length UL56. Three multi-subunit TRAPP complexes have been characterized extensively in yeast and are known to be

involved in intracellular vesicular trafficking by acting as a tether and as a Ypt/Rab guanine nucleotide exchange factor (361). So far, one mammalian Golgi TRAPP complex (TRAPPII) has been identified (362). It consists of four core subunits, TRAPPC1, TRAPPC3, TRAPPC4 and TRAPPC5, two accessory subunits TRAPPC2 and TRAPPC6, and two TRAPP II-specific subunits TRAPPC9 and TRAPPC10. All eight subunits are homologs of known yeast TRAPP complex constituents (361). The mammalian TRAPPII complex acts as a Rab1 guanine nucleotide exchange factor, localizes to COPI-coated vesicles and is reported to bind to COPI adaptor proteins (363). Thus, the human TRAPP complex seems to be a plausible cellular target of HSV-1 during envelopment or egress.

Interestingly, TRAPPC subunits have rarely been quantified in affinity purification mass spectrometry experiments before, as revealed by querying the CRAPome database. Amongst the 411 affinity purification experiments that constitute the CRAPome database, TRAPPC1, TRAPPC4, TRAPPC9 and TRAPPC10 have been observed 0/411, TRAPPC5 6/411, TRAPPC6B 7/411 and TRAPPC3 10/411 times, supporting the identification of the TRAPP complex as a genuine interaction partner of UL56. Since the TRAPPII-specific subunits TRAPPC9 and TRAPPC10 were the strongest hits in the SILAC screen with UL56(1-207)-eGFP showing fold-changes vs. eGFP of 13.2 and 13.7, respectively, and were also identified as significant hits for full-length UL56 (Figure 86), they were taken forward for further biochemical validation.

Most of the top hits identified in the SILAC screen with full-length UL56 were mitochondrion-associated or related to nuclear import (Table 20). The two strongest hits with links to intracellular membrane trafficking were TMED1 (Transmembrane emp24 domain-containing protein 1) and GOPC with 19-fold and 7-fold enrichment compared to eGFP, respectively. Protein GOPC was also found as a hit for truncated UL56(1-207) with 5-fold enrichment, and showed the second highest intensity (combined light, medium and heavy channels) of all quantified proteins in the SILAC experiment (Table 19 and Table 20). Both TMED1 and GOPC are likely to be non-contaminants since the CRAPome database shows that they have been rarely identified in previous AP-MS experiments (TMED: 2/411; GOPC: 1/411).

The transmembrane protein TMED1 localizes mainly to the ER, but can also be found at the Golgi (364). It possesses a C-terminal transmembrane domain and a very small cytoplasmic

domain, although it has been suggested that the internal GOLD (Golgi dynamics) domain of TMED1 projects into the cytosol and can therefore interact with cytosolic protein (364). The cytosolic protein GOPC localizes to the trans-Golgi network and is involved in intracellular trafficking of both plasma membrane- and lysosome-targeted receptors, membrane proteins, ion channels and adhesion molecules (365). Typically, GOPC binds to C-terminal cytosolic peptides of target proteins via its PDZ (PSD-95, Dlg, and ZO-1) domain (366). Protein GOPC has also been reported to regulate protein trafficking to tight junctions and adherens junctions, leading to increased paracellular permeability (365). Since very little is known about TMED1 and the exact size of its cytosolic domain is unclear, and GOPC was not only detected in the SILAC screen with full-length, but also with truncated UL56 as a top hit, GOPC was chosen for further biochemical follow-up study.

Several E3 ubiquitin ligases were identified as strong hits for both UL56 constructs. For UL56(1-207), six ligases were significantly enriched: ITCH, NEDD4, NEDD4L, HECW2, WWP1 and WWP2. Proteins ITCH, NEDD4, NEDD4L, WWP1 and WWP2 belonged to the hits with highest measured intensity, whereas HECW2 was quantified with the lowest intensity of all hits (Table 19). All six ligases are infrequently identified in AP-MS experiments as catalogued in the CRAPome database (ITCH: 2/411, NEDD4: 3/411, NEDD4L: 3/411, WWP1: 2/411, WWP2: 3/411 and HECW2: 1/411). For UL56(1-234), only WWP1 and WWP2 were identified as hits. The other four ligases were still quantified in the SILAC experiment, but did not reach the set significance threshold due to variability of their fold increase across the biological repeats.

Interestingly, Ushijima et al. conducted a yeast two-hybrid screen with HSV-2 UL56 and observed interaction of UL56 with these ubiquitin ligases and a seventh one, HECW1 (221). They also confirmed binding of HSV-2 UL56 to NEDD4 and ITCH by co-immunoprecipitation (221, 222). All of the identified ubiquitin ligases are part of the Nedd4-family of the HECT- (homologous to E6-associated protein C terminus-) type E3 ubiquitin enzymes, one of two groups of E3 ubiquitin ligases, the other one being the RING- (really interesting new gene-) type. Only about 5% of all cellular ubiquitin ligases belong to the HECT group but, nevertheless members of this group fulfill a diverse range of functions (118). For example, multiple Nedd4-family members have been associated with cancer (367), NEDD4 and ITCH are known to influence signaling pathways, NEDD4 plays an important role or normal embryonic heart development,

NEDD4 and ITCH contribute to the regulation of the immune system, and NEDD4 and NEDD4L promote endocytosis of the epithelial Na^+ channel ENaC (118, 368).

In addition, the ESCRT (endosomal sorting complexes required for transport) system is connected with Nedd4 ligases (36). All Nedd4-family ligases carry WW domains through which they interact with the short peptide motif PPXY, a so-called late domain motif. Proteins containing late domains are thought to be ubiquitinated by Nedd4 ligases and subsequently targeted to the ESCRT system, even though the precise mechanisms connecting these pathways are not very well understood. The ESCRT machinery is required for budding of a plethora of enveloped viruses, including HSV-1 (218), via a process related to the formation of multivesicular bodies (36, 37). Therefore, study of the interaction of HSV-1 UL56 with Nedd4-ubiquitin ligases identified in this SILAC screen could elucidate a promising link to the envelopment of HSV-1 particles by the cellular ESCRT system. Due to time constraints, only two of the six ubiquitin ligases identified, NEDD4 and NEDD4L, could be selected for follow-up validation experiments. Both proteins have been previously reported to interact with UL56 from HSV-2 (220, 221).

Remarkably, protein PAWR, one of the putative interaction partners of UL56 identified in the yeast two-hybrid screen described in the previous chapter, was found as a significant putative interaction partner for both full-length and truncated UL56 in the SILAC screen. In the CRAPome database, 27/411 compiled AP-MS experiments quantified PAWR, indicating that it could be a common contaminant. Although no direct binding between PAWR and GST-UL56(1-207) could be demonstrated by GST pull-down previously (Chapter 6) and no obvious link of PAWR to intracellular trafficking exists, PAWR was taken forward for repeated biochemical validation because of its occurrence in two completely independent screens using different techniques.

7.2 Biochemical validation of interactions between SILAC hits and UL56

7.2.1 Confirmation of UL56-interacting proteins by co-immunoprecipitation

After selection of proteins TRAPPC9, TRAPPC10, GOPC, NEDD4, NEDD4L and PAWR as promising potential interaction partners of UL56, co-immunoprecipitation experiments were performed in order to demonstrate binding of overexpressed UL56 to endogenously expressed

proteins (Figure 89). Full-length or C-terminally truncated UL56 were expressed as eGFP fusion proteins in HEK293T cells. For both constructs, strong enrichment of the tagged protein on GFP-Trap A beads could be observed, whilst no enrichment of the housekeeping protein GAPDH was seen (α-GFP and α-GAPDH blots, Figure 89). Strinkingly, strong bands were detected for GOPC with both constructs, indicating that HSV-1 UL56 and GOPC indeed interact in HEK293T cells (Figure 89A).

For PAWR, TRAPPC9 and TRAPPC10, no bands running at the correct predicted molecular weight were found in the bound (pellet) fractions. Whilst the α-PAWR antibody detected a band at the predicted molecular weight in input and supernatant samples, the antibody staining of input and supernatant samples for both TRAPPC9 and TRAPPC10 revealed many unspecific or background bands (Figure 89A). A second antibody was tested for TRAPPC10 (Figure 89B). Unfortunately, similar, non-specific staining was observed, implying that binding between UL56 and the two components of the mammalian TRAPPII complex could neither be confirmed nor refuted because target specificity of currently commercially available antibodies against these targets is questionable.

Full-length UL56 was also found to interact with endogenously expressed NEDD4, but not NEDD4L (Figure 89C). Interestingly, two NEDD4 bands differing by only a few kDa were detected in the bound sample lane, whereas only the band with higher electrophoretic mobility running at the predicted molecular weight for NEDD4 was observed in input and supernatant lanes. The band with lower electrophoretic mobility, not visible in input and supernatant lanes, was more intense than the lower molecular weight band. Taken together, these results indicate that HSV-1 UL56 interacts with NEDD4 in cells. It has been previously suggested that HSV-2 UL56 post-translationally phosphorylates NEDD4 (221), and the multiple bands observed in this study could be attributed to similar modification processes.

Since no reliably working antibodies against TRAPPC9 and TRAPPC10 were available, an alternative approach using tagged proteins to monitor the interaction was taken. Full-length UL56 and myc-tagged TRAPPC9 or TRAPPC10 were co-transfected into HEK293T cells and co-immunoprecipitation was performed as for endogenous proteins. Although background staining of the used α-myc antibody around the expected molecular weight of TRAPPC9 and TRAPPC10 was observed (Figure 90, eGFP only and eGFP-UL56(-1234) only lanes), none of

this non-specifically recognized protein was captured in co-immunoprecipitation experiments with eGFP alone. In addition, overexpressed TRAPPC9 and TRAPPC10 could be detected above this background antibody staining. Both proteins were found in bound sample fractions when co-transfected with eGFP-UL56(1-234), but not eGFP alone, indicating that these components of the human TRAPPII complex may indeed interact, albeit weakly, with HSV-1 UL56 in cells.

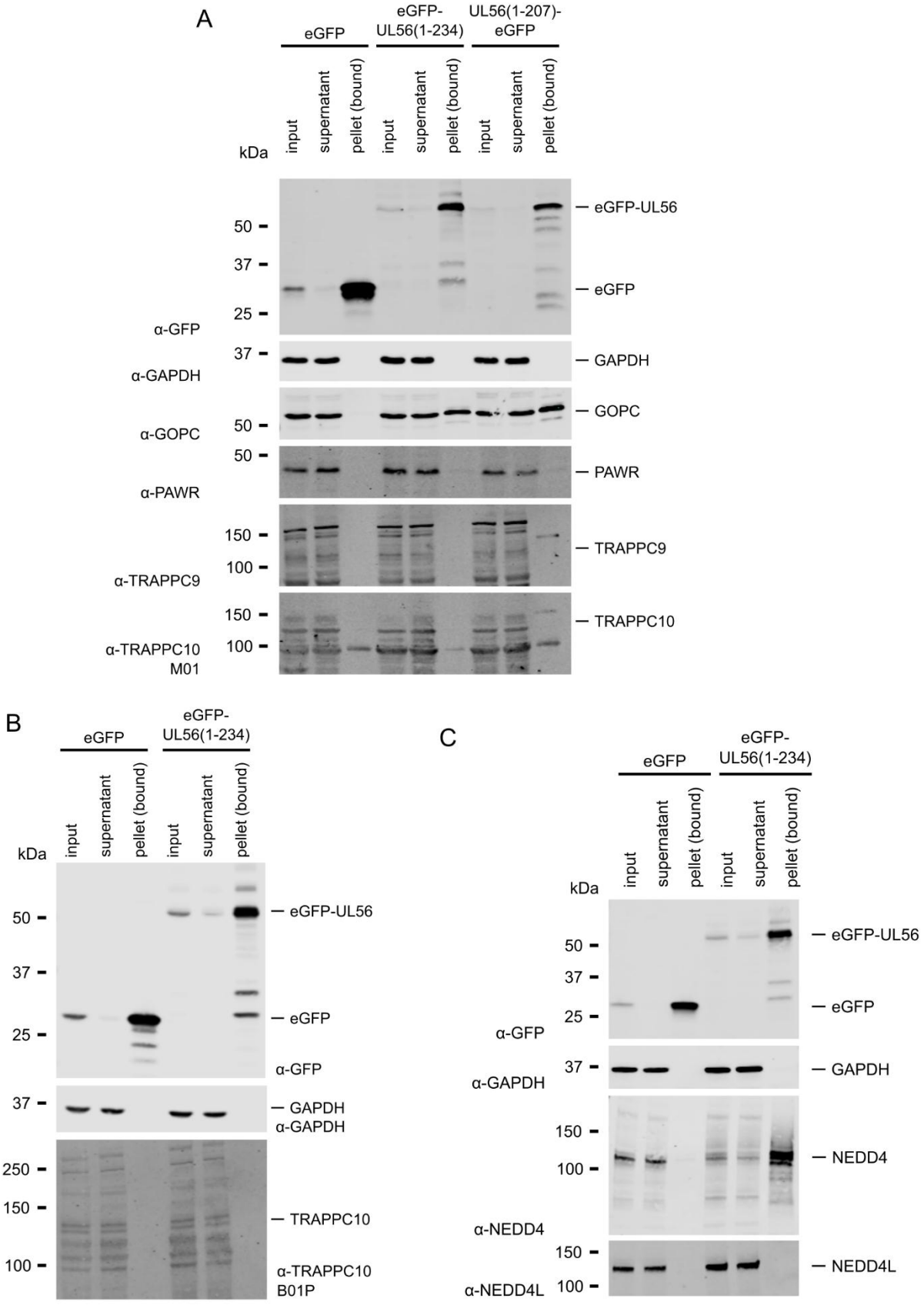


Figure 89 Co-immunoprecipitation experiments investigating interactions between UL56 and endogenously expressed proteins identified as potential UL56-binding proteins in the SILAC screen. Full-length or truncated UL56 was transfected into HEK293T cells, cells were lysed (A: 90 min, B: 75 min, C: 85 min) and the lysate incubated with GFP-Trap A beads (A: 45 min, B: 75 min, C: 50 min). The beads were then washed twice and proteins bound to the beads were eluted by boiling at 95 °C for 10 min. Samples were separated on SDS-PAGE gels (A: 9% and 12%, B, C: 4-15% gradient gels)

and analyzed by Western blot using antibodies as indicated. Two different antibodies were tested for TRAPPC10 (A and B). Loading of samples: (A) eGFP, GAPDH: 0.32% of total input and supernatant, 13.3% of total pellet sample; GOPC, PAWR, TRAPPC9, TRAPPC10: 0.46% of total input and supernatant, 18.9% of total pellet sample. (B) eGFP, GAPDH: 0.28% of total input and supernatant, 11.1% of total pellet sample; TRAPPC10: 0.66% of total input and supernatant, 26.7% of total pellet sample. (C) eGFP, GAPDH: 0.26% of total input and supernatant, 11.1% of total pellet sample; NEDD4, NEDD4L: 0.63% of total input and supernatant, 26.7% of total pellet sample. Note: The Western blots displayed in (B) and (C) were performed using samples from the co-immunoprecipitation experiments presented in Figure 94. Blotting for eGFP and GAPDH was only performed once and therefore, the α-GFP and α-GAPDH blots in this figure are cropped from the blots shown in Figure 94.

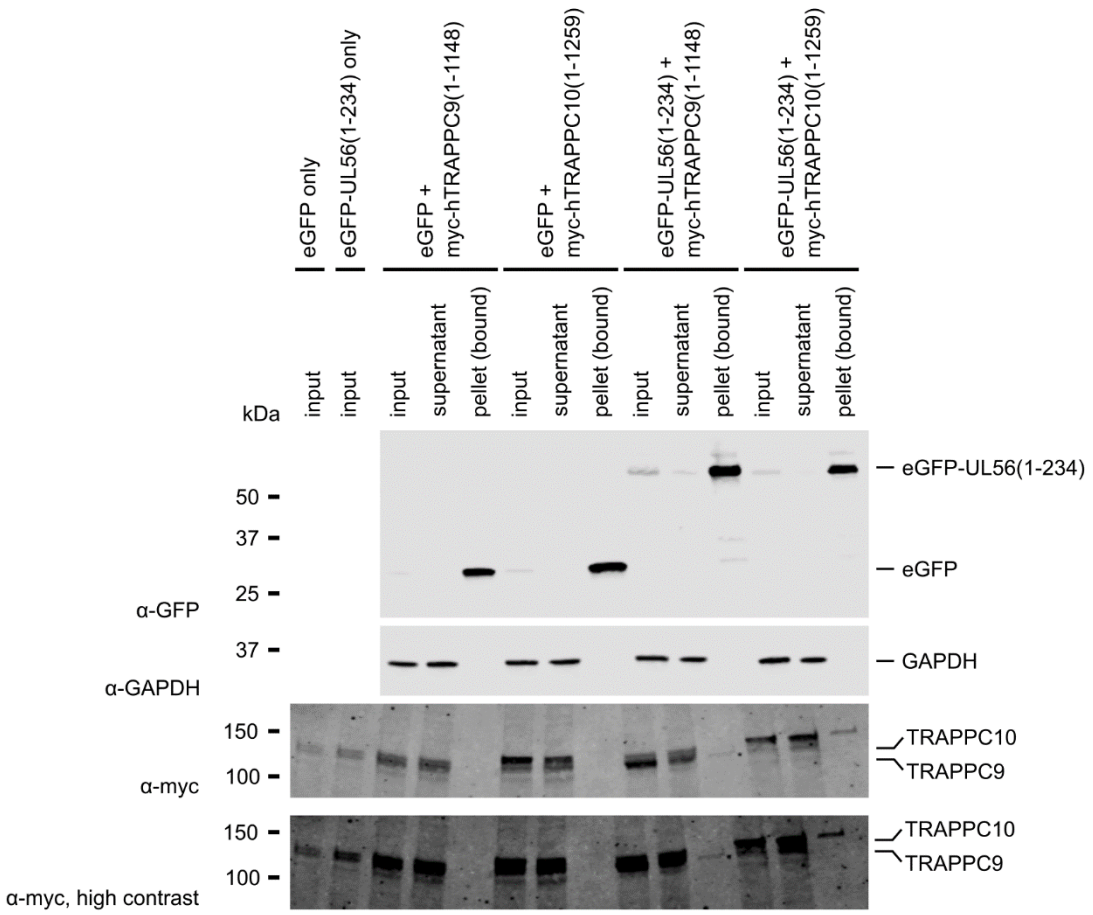


Figure 90 Co-immunoprecipitation of UL56 with co-transfected myc-tagged TRAPPC9 or TRAPPC10. Full-length eGFP-tagged UL56 or eGFP alone were co-transfected with myc-tagged TRAPPC9 or TRAPPC10 into HEK293T cells. Cells were lysed for 65 min and the lysates were incubated with GFP-Trap A beads for 60 min. Bound samples were eluted by boiling at 95 °C for 10 min and samples were then separated on 4-15% SDS-PAGE gradient gels. Western blotting against eGFP, GAPDH and the myc-tag was performed to assess the interaction between UL56 and tagged TRAPPC9 or TRAPPC10. Two samples of HEK293T cell lysate expressing only eGFP or eGFP-UL56(1-234) without co-transfected TRAPPC9 or TRAPPC10 were also analyzed in order to assess non-specific binding of the α-myc antibody (first two lanes). The α-myc blot is displayed at normal and high contrast to improve the visibility of the weak bands in the bound protein lanes. Loading of samples: eGFP, GAPDH: 0.26% of total input and supernatant, 11.1% of total pellet sample; myc: 0.63% of total input and supernatant, 26.7% of total pellet sample.

7.2.2 Detection of direct binding between UL56 and SILAC hits using GST pull-downs

Proteins GOPC, TRAPPC9, TRAPPC10 and NEDD4 were identified as putative UL56 interaction partners in the SILAC quantitative proteomics screen and were demonstrated to bind to UL56 by co-immunoprecipitation. Since the interaction between UL56 and NEDD4 has been characterized extensively for HSV-2 UL56 (220, 221), further validation experiments were performed with GOPC, TRAPPC9 and TRAPPC10 only. The interaction between UL56 and these proteins could potentially be mediated by "bridging" proteins present in cell lysates. To investigate whether binding between UL56 and GOPC, TRAPPC9 and TRAPPC10 is direct or mediated by any third factors, GST pull-downs were performed.

Full-length GOPC, as well as mammalian TRAPPC9 and TRAPPC10 were cloned into a wheat germ expression vector encoding an N-terminal myc-tag (pF3A-myc). Sequencing of multiple constructs for GOPC revealed that only the 454 amino acid long isoform 2 of GOPC could be cloned. Isoforms 1 and 2 of GOPC only differ by a small eight residue deletion between residues 150-157 (LSGPSVEE). Since this region is predicted to form a random coil and work on isoform 2 of GOPC has been published previously (369), all work presented in this thesis was carried out using this isoform of GOPC. The myc-tagged constructs were expressed using a wheat germ expression system and incubated with GST-UL56(1-207), or GST as a negative control, in pull-down experiments (Figure 91). Excitingly, GOPC was found to interact with UL56 strongly, whilst not binding to GST alone (Figure 91A). Both TRAPPC9 and TRAPPC10 also bound to UL56 (Figure 91B), albeit less strongly than observed for GOPC.

In the previous chapter, PAWR was identified as a putative interaction partner of UL56 in the yeast two-hybrid screen. Binding between UL56 and PAWR was assessed by GST pull-down, but a truncated construct (residues 217-340) corresponding to the fragment sequenced from the positive yeast two-hybrid clones was used (Chapter 6). For this version of PAWR, no direct interaction with UL56 was observed. In this chapter, PAWR was identified in the SILAC screen as an interaction partner of UL56, but no binding between UL56 and PAWR could be demonstrated by co-immunoprecipitation (Figure 89). Nevertheless, since PAWR appeared to be an UL56 interaction partner in two independent high-throughput screens, a GST pull-down was conducted again for PAWR. To ensure that a potential direct interaction between UL56 and

PAWR was not obscured by any truncations, full-length PAWR (amino acids 1-340) was used as the prey protein. However, as before, no direct binding between GST-tagged UL56(1-207) and PAWR could be detected (Figure 91A).

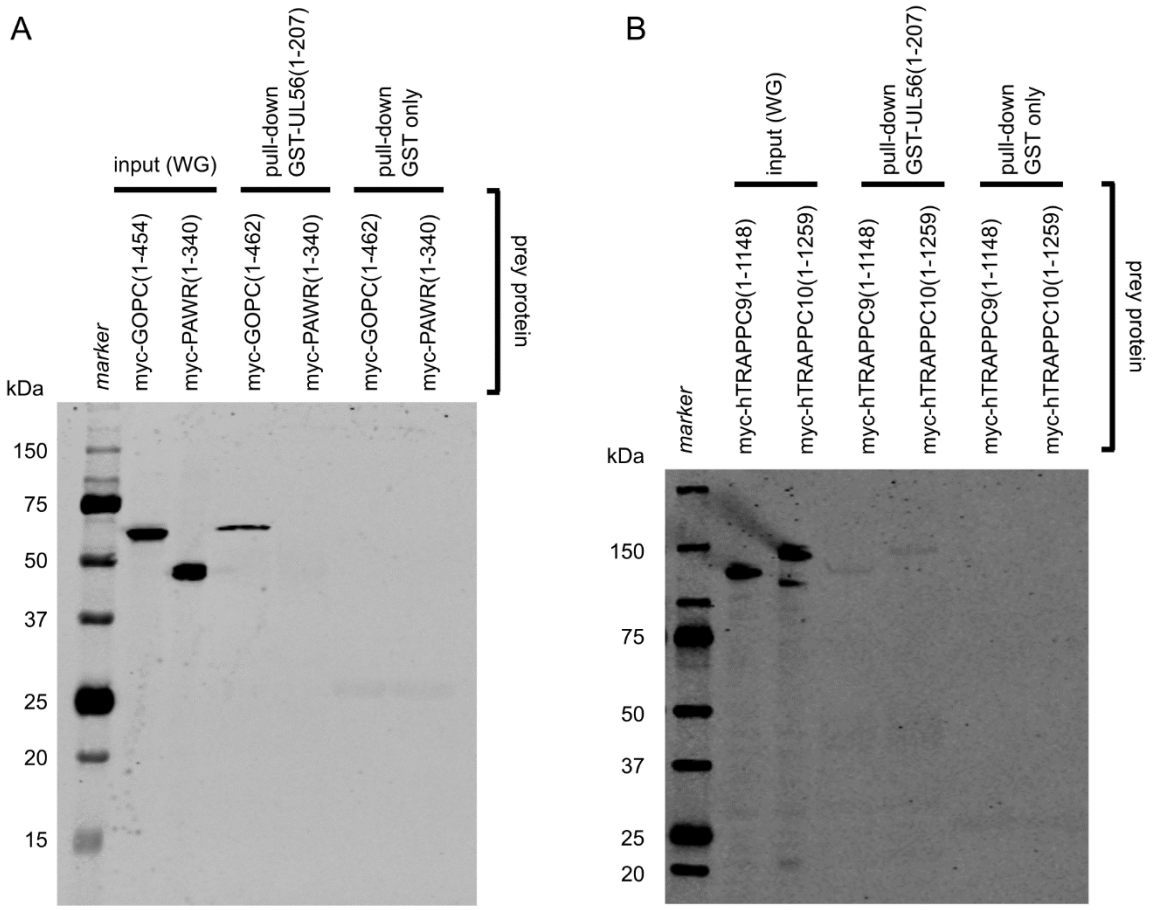


Figure 91 Confirmation of direct binding between UL56 and putative interaction partners identified in the SILAC interaction screen by GST pull-down. Full-length GOPC and PAWR (A), or TRAPPC9 and TRAPPC10 (B) were expressed as myc-tagged fusion proteins for two hours in wheat germ extract (input WG). Magnetic glutathione beads pre-loaded with purified GST-UL56(1-207) or GST were then incubated with the expression extract for 85 min (A) or 55 min (B), the beads were washed and the protein competitively eluted. Input and bound samples (pull-down) were separated on a 12% (A) or a 4-15% gradient (B) SDS-PAGE gel. Western blotting using an α-myc antibody was performed to visualize the results.

The interaction observed between UL56 and TRAPPC9 or TRAPPC10 was very weak in both co-immunoprecipitation and GST pull-down experiments. These two human TRAPPII complex components are known to co-elute during size exclusion chromatography (370) and are thought to promote dimerization of the TRAPPII complex (371), although this model is contested (361). Even though no direct interaction between TRAPPC9 and TRAPPC10 has been reported, it is conceivable that they could cooperatively bind to UL56. Therefore, it was tested whether co-

expression of TRAPPC9 and TRAPPC10 in wheat germ extract would enhance their binding to GSt-UL56(1-207) in the GST pull-down assays (Figure 92). However, TRAPPC9 and TRAPPC10 were again observed to both bind to UL56 weakly even when used as co-expressed prey proteins, suggesting no cooperativity in their interaction with UL56.

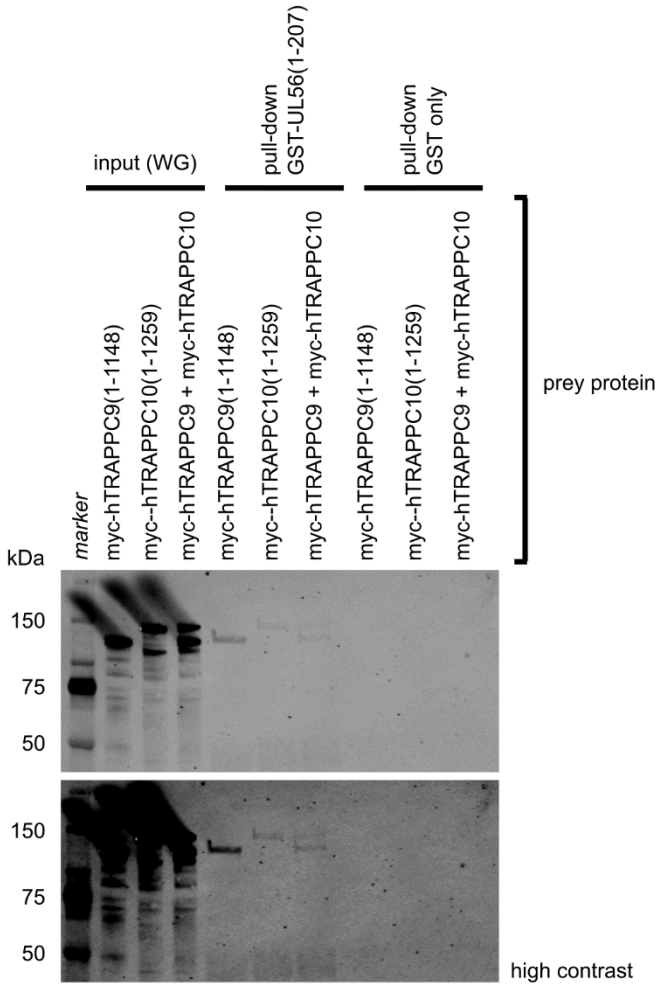


Figure 92 Assessment of potential cooperative binding between TRAPPC9 and TRAPPC10 to UL56 by GST pull-down. Proteins TRAPPC9 and TRAPPC10 were expressed as myc-tagged constructs either separately or together in wheat germ extract for two hours (input WG) and then incubated with magnetic glutathione beads pre-loaded with purified bait proteins GST-UL56(1-207) or GST. After 70 min, the beads were washed and bound protein was competitively eluted. Samples were separated on 4-15% gradient SDS-PAGE gels. Western blotting was performed and membranes probed with an α-myc antibody to visualize binding between UL56 and separately or co-expressed TRAPPC9 and TRAPPC10. The blot is shown at a higher contrast level (lower panel) to enable visualization of weak signals for interacting TRAPPC9 and TRAPPC10.

7.3 Determination of the binding region between UL56 and GOPC

Protein GOPC was not only identified as a top hit for both full-length and truncated UL56 in the SILAC proteomics screen (section 7.1.5), its interaction with UL56 could also be validated

by co-immunoprecipitation and GST pull-downs (section 7.2). The binding between GOPC and UL56 seemed strong and could be reproducibly demonstrated by these techniques. Therefore, in order to establish the domains or protein regions of UL56 and GOPC mediating binding, truncations of both proteins were designed and tested in either GST pull-down experiments (GOPC truncations, Figure 93) or co-immunoprecipitation assays (UL56 truncations, Figure 94).

The PDZ domain of GOPC serves as an interaction hub for its association with various cellular proteins and PDZ domains are exploited by a multitude of viruses during various points of their life cycle (366, 372). It is therefore plausible that HSV-1 UL56 could bind the PDZ domain of GOPC. Consequently, two constructs of GOPC were designed, one encompassing the PDZ domain alone (residues 276-362) and one consisting of the N-terminal domain of the protein up to the PDZ domain (residues 1-275) (Figure 93A). These constructs were expressed as myc-tagged proteins in wheat germ extract and then used as prey in GST pull-downs (Figure 93B). However, whilst good expression of full-length GOPC and the N-terminal domain was observed, the PDZ domain construct was expressed at extremely low levels in multiple independent experiments (Figure 93B).

To test whether low expression of the GOPC PDZ domain arose due to an unfortunate choice of domain boundaries, three new constructs encompassing the PDZ domain of GOPC were designed. Two of these constructs consisted of the PDZ domain with a few flanking residues at both termini (residues 270-363 or 272-367), and one encompassed the whole C-terminus of GOPC starting from the PDZ domain (residues 276-454) (Figure 93A). Surprisingly, only the construct containing all of the C-terminus of GOPC, myc-GOPC(276-454), was found to express, whereas both PDZ domain-only constructs expressed only very low levels of protein in the used wheat germ system (Figure 93C).

Binding of full-length GOPC to GST-UL56(1-207) was reproducibly observed, but no interaction between UL56 and the C-terminal region of GOPC (residues 276-454) containing the PDZ domain could be demonstrated (Figure 93C). A very weak band was detected in the GST-UL56(1-207) pull-down sample for the N-terminal domain of GOPC (residues 1-275), but not with GST alone (Figure 93B). These results suggest that not the PDZ domain, but maybe the N-terminal domain of GOPC is mediating the interaction with UL56. However, more work is needed to accurately map the region of GOPC binding to UL56 (see section 7.4).

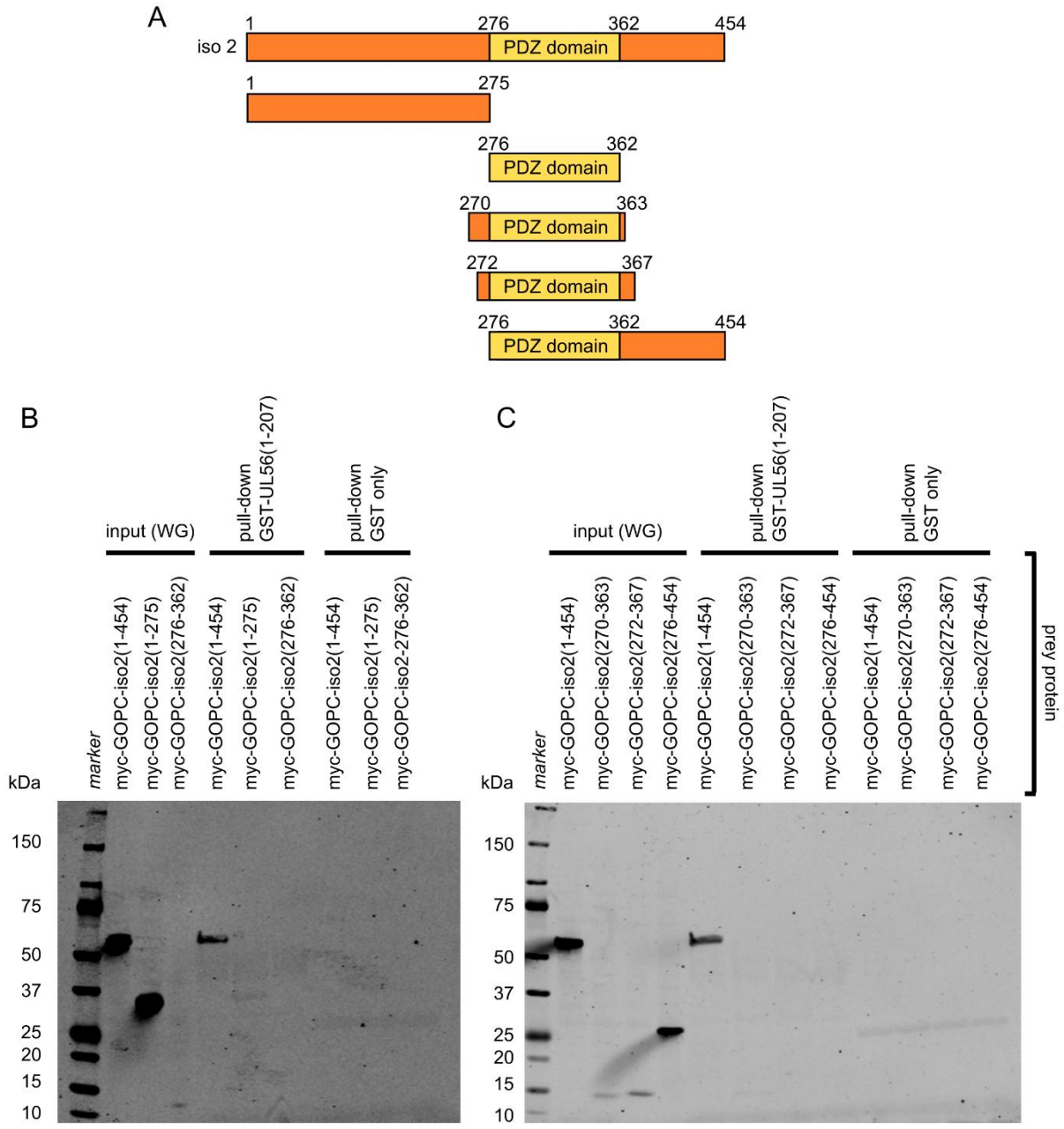


Figure 93 Identification of the region of GOPC interacting with UL56 by GST pull-downs. (A) Different truncation constructs of GOPC (isoform 2, 454 amino acids) were designed to encompass either the N-terminal domain (1-275) or the PDZ domain (276-362). (B, C) Constructs were expressed as myc-tagged proteins in wheat germ extract for two hours (input WG). Magnetic glutathione beads were first loaded with either GST-tagged UL56(1-207) or GST as a negative binding control, and subsequently incubated with the wheat germ-expressed prey proteins for 55 min (A) or 70 min (B). The beads were then washed and bound proteins eluted using reduced glutathione. Samples were separated on 4-15% gradient SDS-PAGE gels, blotted and analyzed by Western blot using an α-myc antibody.

In addition to attempting to define the region of GOPC that binds UL56, I also sought to identify the region of UL56 essential for the interaction. Since UL56 is not predicted to be folded, design of truncations cannot be guided by any domain boundaries (besides the C-terminal transmembrane domain). Nonetheless, a notable feature of the protein sequence is a central,

arginine-rich region (between residues 75-100). Deletion of this stretch of residues does not alter the cellular localization of UL56 in transfected cells (224), but the exact function of the arginine-rich region is not known.

First, C-terminal truncations of UL56 were cloned into a mammalian expression vector adding an N-terminal eGFP-tag. The chosen constructs encoded the N-terminal residues of UL56 up to the start of the arginine-rich region (residues 1-74), the N-terminus of the protein including the arginine-rich region plus a few flanking amino acids (residues 1-104), the N-terminal two-thirds of the whole protein (residues 1-157), as well as the previously used, but now N-terminally eGFP-tagged construct lacking the transmembrane domain plus the cysteine doublet (residues 1-207) (Figure 94A). Co-immunoprecipitation experiments using these constructs expressed in HEK293T cells were performed to assess binding of endogenous GOPC (Figure 94B). All truncated proteins expressed well and were enriched on GFP-Trap A beads. Some degradation could be detected, but strong bands corresponding to the expected sizes of the eGFP-tagged deletion proteins were found for all constructs tested. A very clear-cut result was observed: all constructs besides eGFP-UL56(1-74) and eGFP-UL56(1-104) bound endogenously expressed GOPC strongly (Figure 94B).

The region of UL56 common to all constructs that bound GOPC, but absent from those that did not interact with GOPC, spans residues 105-157. In order to test whether residues 105-157 of HSV-1 UL56 are not just necessary, but sufficient for the interaction between UL56 and GOPC, three more eGFP-tagged truncation constructs of UL56 starting at amino acid 105 and terminating at residues 234, 207 or 157 were designed (Figure 94A). All of these proteins expressed well in HEK293T cells and could be enriched on GFP-Trap A beads with very little degradation (Figure 94C). Fusion proteins spanning residues 105-234 or 105-207 were found to co-immunoprecipitate with GOPC, albeit the bands observed in the bound sample lanes were weaker than the one observed for full-length eGFP-tagged UL56. Surprisingly, the proposed minimal binding region, UL56(105-157), was not able to bind to endogenous GOPC. Taken together, these results indicate that residues 105-157 of HSV-1 UL56 are necessary, but not sufficient for its interaction with GOPC (Figure 94A).

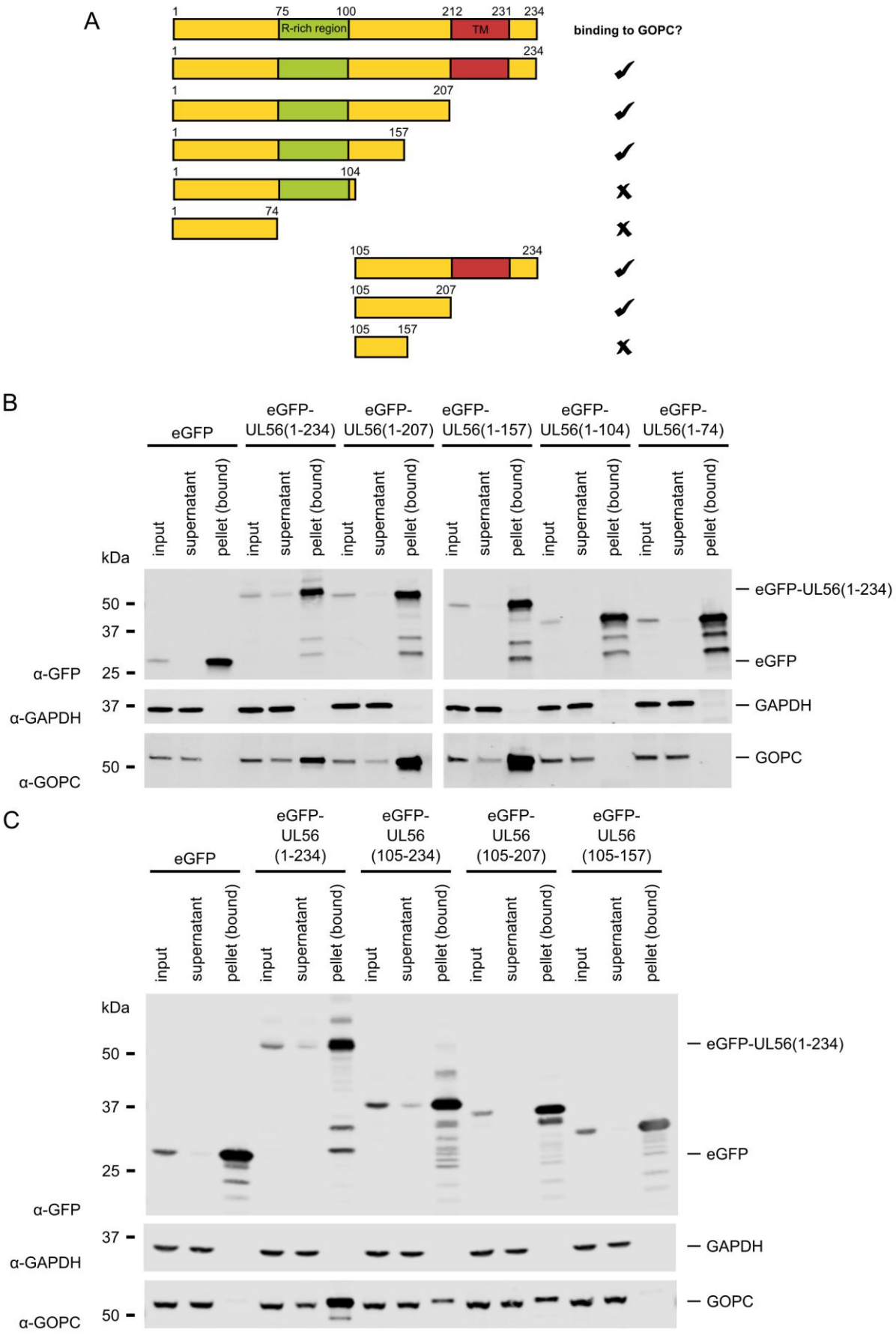


Figure 94 Co-immunoprecipitation of endogenous GOPC with truncated constructs of N-terminally eGFP-labeled UL56 to identify region of UL56 binding to GOPC. (A) Overview of different truncations constructs used to define region of UL56 interacting with GOPC. The arginine-rich and transmembrane (TM) region of UL56 are shown. Qualitative binding

of the different proteins to GOPC as observed in co-immunoprecipitation experiments (B, C) is indicated. (B, C) Full-length and truncated versions of UL56 were expressed as N-terminal eGFP fusion proteins in HEK293T cells. Both C-terminal truncations (B) and N-terminal or C- and N-terminal truncations (C) were explored. Cells were lysed for 85 min (A) or 75 min (B) and the lysates incubated with GFP-Trap A beads for 50 min (A) or 75 min (B). After two washes, proteins bound to the GFP-Trap beads were eluted by boiling for 10 min at 95 °C. Samples were separated on 4-15% gradient SDS-PAGE gels and subsequently visualized by Western blot using antibodies against GFP, GAPDH and GOPC. Loading of samples: (B) eGFP, GAPDH: 0.26% of total input and supernatant, 11.1% of total pellet sample; GOPC: 0.63% of total input and supernatant, 26.7% of total pellet sample. (C) eGFP, GAPDH: 0.28% of total input and supernatant, 11.1% of total pellet sample; GOPC: 0.66% of total input and supernatant, 26.7% of total pellet sample.

7.4 Discussion

Herpes simplex virus 1 relies on cellular machinery to facilitate numerous phases in its life cycle, for example transport to the nucleus, transcription and translation. Importantly, budding of virus into intracellular membranes and egress of enveloped viral particles must also be mediated by cellular factors, but these processes are not well understood and very few interactions between HSV-1 and cellular proteins regulating these steps have been characterized (210, 214, 328). Proteomic studies based on mass spectrometry have greatly enhanced our knowledge of the composition of virions, of post-translational modifications of viral and cellular proteins, of changes in the host proteome and of the adaptation of the immune response upon virus infection (373). However, so far, no proteomic studies were performed to find cellular interaction partners of HSV-1 proteins during envelopment and egress. This uses a SILAC quantitative proteomics approach to identify novel interactions between UL56, an HSV-1 protein with multiple roles during the late phase of infection, and cellular proteins.

7.4.1 Putative binding partners of full-length and truncated UL56 form different interaction networks

Two different eGFP-tagged constructs of the envelope protein UL56 were used in a SILAC proteomic interaction screen, one encompassing the full-length protein including its C-terminal transmembrane domain (eGFP-UL56(1-234)), the second one lacking this region of the protein (UL56(1-207)-eGFP). For both proteins, numerous significantly enriched putative interaction partners were discovered (Figure 86). Bioinformatic interaction network analysis showed minor

overlap between the identified hits, but largely, clustering of hits was distinct for the two constructs.

For the truncated protein UL56(1-207)-eGFP, a large cluster of ribosomal proteins was found. Additional ribosomal proteins were quantified in the screen, but not significantly enriched. This observation can probably be attributed to the association of the basic regions of UL56 with nucleic acids (see Chapter 5) and was therefore considered an artifact.

Subunits of the human TRAPP complex were identified as a distinct cluster. All subunits forming the core of TRAPP complexes, TRAPPC1, TRAPPC3, TRAPPC4 and TRAPPC5, one accessory subunit, TRAPPC6B, and two subunits thought to be specific for TRAPPII, TRAPPC9 and TRAPPC10 (362), formed this cluster. Only subunit TRAPPC2, the eight and last subunit of human TRAPPII, was not enriched more than 2-fold and/or did not reach the significance threshold, although it was quantified in the screen. Since almost all subunits of the human TRAPP complex were found and formed such a discrete cluster, it is likely that UL56 binds to one or more TRAPPC proteins and thus pulls down the whole complex. In a yeast two-hybrid screen performed by Ushijima and colleagues, multiple E3 ubiquitin ligases were found to bind to UL56 (221). Even though none of these interactions was recapitulated in the yeast two-hybrid described in this thesis (Chapter 6), six of the seven E3 ligases first described by Ushijima et al. were found as top hits in the SILAC screen and constituted the third cluster observed for UL56(1-207)-eGFP.

Although subunits TRAPPC6B, TRAPPC9 and TRAPPC10 of the human TRAPP complex were also identified as a separate cluster of top hits for full-length UL56, the clustering of other hits was remarkably different to the one observed for the truncated construct. The enrichment of mitochondrial proteins is particularly puzzling since transfected UL56 is not reported to localize to mitochondrial membranes (220) and not expected to see proteins associated with the inner mitochondrial membrane at any point during its expression. Therefore, this result most likely represents non-specific enrichment of mitochondrial membranes occurring during the immunoprecipitation of UL56(1-234).

The binding of full-length UL56 to SEC61A and proteins of the OST complex is less surprising. Protein UL56 is a type II transmembrane protein, which means that it is targeted to the ER

during translation and inserted into the ER membrane via the translocon (374, 375). Therefore, it is conceivable that all UL56 molecules encounter the translocon component SEC61A. Since UL56 does not possess a luminal domain, it probably does not directly interact with the N-oligosaccharyl transferase complex OST, which resides in the ER lumen (360). More likely, components of the OST complex co-immunoprecipitate with UL56 because of their association with SEC61A (360).

It is interesting that full-length UL56 seemed to interact with multiple protein associated with nuclear import or the nuclear pore complex. It has been suggested that the arginine-rich region of UL56 can serves as an atypical nuclear localization signal (NLS). However, it is thought that the alternative NLS is only exposed in UL56 constructs lacking the transmembrane domain (224), and no nuclear pore complex proteins or importins were identified as SILAC hits with truncated UL56 (residues 1-207). Therefore, the physiological relevance of the potential association between UL56 and this cluster of proteins remains to be investigated.

The interactomes of full-length and truncated UL56 differ considerably. A large proportion of UL56 lacking its transmembrane domain is mislocalized to the nucleus and the cytosol instead of being directly inserted into cytoplasmic membranes (224, 233). Therefore, some of the differences like binding of full-length, but not truncated UL56 to components of the translocon machinery can be easily explained. Other than affecting the cellular localization of the constructs, deletion of the transmembrane domain of UL56 also leads to differential exposure of residues close to the transmembrane domain, potentially affecting interactions with cellular proteins. Lastly, the position of the eGFP tag, N-terminal for UL56(1-234) and C-terminal for UL56(1-207), could also influence binding. Despite the differences observed in the interactomes, it should be noted that several proteins such as GOPC, PAWR, components of the mammalian TRAPP complex and some E3 ubiquitin ligases were identified as putative binding partners of both UL56 constructs, suggesting that these cellular proteins constitute bona fide interaction partners of UL56.

7.4.2 UL56 binds directly to components of the human TRAPP complex

Components of the human TRAPP complex were identified in the SILAC screen with both full-length and truncated UL56. For UL56(1-207), seven of the eight reported subunits of human TRAPPII were found as top hits, whereas for UL56(1-234), only subunits TRAPPC6B,

TRAPPC9 and TRAPPC10 were significantly enriched, even though all other TRAPPII components were still quantified (Figure 86). It was demonstrated by co-immunoprecipitation that in HEK293T cells, co-transfected eGFP-tagged full-length UL56 interacts with myc-tagged TRAPPC9 or TRAPPC10. In addition, binding between these proteins seems to be direct, albeit weak, as shown by GST pull-downs.

The human TRAPPII complex is reported to localize to transitional ER structures as well as COPI-coated vesicles and functions as a tether during the formation of vesicular tubular clusters (363, 376). Human TRAPPII is also pivotal for the maintenance of Golgi architecture, and may function both during ER to Golgi and Golgi to plasma membrane trafficking (361, 362, 376). The architecture of the human TRAPPII complex is not clear and our current understanding is based on structural studies of yeast TRAPPC. Single particle electron microscopy suggests that yeast TRAPPII forms a dimer, with the accessory subunits TRAPPC9 and TRAPPC10 residing on opposite sites of the macromolecule (371). However, the exact placement of these two subunits is ambiguous since crystal structures of TRAPPC9 and TRAPPC10 have not been determined. In particular, Kim et al. refute the electron microscopy structure presented by Yip and colleagues and speculate that the observed dimer of TRAPPII might be functionally irrelevant (361). It should be noted that the TRAPP complex is not thought to be anchored directly into membranes, but rather attaches to them via the flat and highly positively charged surfaces of subunits TRAPPC3 and TRAPPC5 (377).

Taken together, a hypothetical model of the interaction between the human TRAPPII complex and UL56 can be proposed. Proteins TRAPPC9 and TRAPPC10 were top hits for the two UL56 constructs investigated in the SILAC screen and their interaction with UL56 is likely to be direct. Therefore, it seems most likely that the largely unfolded cytoplasmic domain of UL56 acts as a "fishing line", binding to both TRAPPC9 and TRAPPC10. The transmembrane domain of UL56 does not seem to be required for the interaction given that more TRAPPC subunits were detected for truncated UL56(1-207). Since TRAPPC9 and TRAPPC10 are thought to embrace the core of the TRAPP complex, such a binding mode of UL56 could explain why almost all subunits co-immunoprecipitated together.

Since the exact function of the mammalian TRAPP complex is disputed, it can only be speculated why HSV-1 might exploit this cellular machinery. One option could be that UL56 hijacks

TRAPPC during infection to facilitate the transport of other viral envelope or tegument proteins. It has recently been reported that HSV-1 glycoproteins get delivered from the plasma membrane to the site of envelopment (215). Therefore, the interaction of UL56 with TRAPPC could streamline the trafficking of viral proteins from the ER or the Golgi network to the plasma membrane. Alternatively, binding of UL56 to TRAPPC could lead to fusion of membranes enriched in viral proteins, thus enhancing the tegument assembly process. Further work on the UL56:TRAPPC interaction could not only help to understand its role during virus replication, but also contribute to the delineation of the physiological function of the human TRAPPII complex.

7.4.3 The elusive interaction between UL56 and PAWR

One of the proteins identified in the SILAC screen with both full-length and truncated UL56 was PAWR. Curiously, this protein was also the most promising hit from the yeast two-hybrid screen (Chapter 6). However, the interaction between UL56 and PAWR could not be recapitulated in either GST pull-downs or by co-immunoprecipitation. Compared to other hits followed up from the SILAC screen presented in this thesis, PAWR was found relatively often in the CRAPome database, indicating that it might be a false positive hit. Alternatively, the binding between PAWR and UL56 might be weak and/or mediated by an additional cellular protein, explaining why it can be detected using highly sensitive mass spectrometry, but not by Western blot. Clearly, the interaction needs to be studied further in order to confidently in- or exclude PAWR from the list of genuine UL56 binding partners.

7.4.4 Discovery of members of the NEDD4-family of E3 ubiquitin ligases as potential binding partners of HSV-1 UL56

Several HECT-type E3 ubiquitin ligases were identified as putative interaction partners of UL56 in the SILAC screen. This observation agrees with a previously published yeast two-hybrid screen of HSV-2 UL56 (221). In total, six E3 ubiquitin ligases, NEDD4, NEDD4L, HECW2, ITCH, WWP1 and WWP2 were detected as significantly enriched proteins (Figure 86). Two of these hits, NEDD4 and NEDD4L, were tested for binding to UL56 in co-immunoprecipitation experiments (Figure 89).

The interaction between UL56 and NEDD4 could be clearly demonstrated. Interestingly, the α-NEDD4 antibody detected a band running at a slightly higher molecular weight in the bound sample than in input or supernatant samples, which could potentially be attributed to post-translational modification of NEDD4 in the NEDD4 population that binds UL56. It has been demonstrated previously that NEDD4 gets ubiquitinated in the presence of HSV-2 UL56 and that during HSV-2 infection, NEDD4 is not only degraded, but also gets phosphorylated (221). However, during HSV-1 infection, no such downregulation of NEDD4 takes place (222), explaining why an interaction between UL56 and NEDD4 could easily be detected in the SILAC screen and the subsequent co-immunoprecipitation experiment.

It is intriguing that although an interaction between NEDD4L and UL56 was found in the SILAC screen, NEDD4L was not found to co-immunoprecipitate with UL56. Since full-length NEDD4L is easily detected in the cell lysate, this lack of binding cannot be ascribed to potential complete degradation of NEDD4L. It is possible that the interaction between NEDD4L and UL56 is very transient and was therefore only detected by mass spectrometry, not by Western blotting. Alternatively, the epitope of NEDD4L recognized by the used antibody could be altered upon interaction of NEDD4L with UL56, for example by ubiquitination.

In summary, the discovery of multiple E3 ubiquitin ligases belonging to the NEDD4 family as putative interaction partners of HSV-1 UL56 is highly interesting and could present a link to viral envelopment via the cellular ESCRT machinery, Additionally, it serves as an independent validation of earlier results obtained in a yeast two-hybrid screen, a completely different experimental system.

7.4.5 UL56 hijacks the cellular protein GOPC to promote envelopment and trafficking of viral particles or to regulate the expression of cell surface proteins

In the SILAC screen, the Golgi-localized protein GOPC was found as a putative interaction partner of both full-length and truncated UL56. The binding of GOPC to UL56 was confirmed by co-immunoprecipitation and further validated in GST pull-downs. Narrowing down the region of GOPC mediating this interaction proved to be difficult due to problems expressing the PDZ domain of GOPC on its own in wheat germ extract. Only a construct containing the whole C-terminal domain of GOPC (including the PDZ domain), expressed well, but did not interact with

purified GST-UL56(1-207). Very weak binding was demonstrated between the N-terminal domain of GOPC and UL56 (Figure 93). Since the interaction was so much weaker than the one observed between full-length GOPC and UL56, it is possible that the chosen point of truncation (residue 275 of GOPC) lies within the binding region. Since UL56 is largely disordered, it is plausible that multiple contacts between UL56 and GOPC are made. For example, UL56 could bind to both the PDZ domain and a stretch of residues of the N terminus of GOPC cooperatively.

The hypothesis that UL56 binds to GOPC via multiple distinct epitopes is potentially further supported by the results of the co-immunoprecipitation experiments using truncated constructs of UL56 to determine the region interacting with endogenous GOPC. It was clearly demonstrated that residues 105-157 of UL56 are necessary for binding. However, these residues alone were not able to pull down GOPC. One interpretation of these findings could be that multiple regions of UL56 bind to GOPC, with stretch 105-157 being the most important one, and certain residues in regions 1-104 and/or 157-207 forming additional contacts to GOPC in a cooperative manner. However, another possible explanation for the lack of binding between UL56(105-157) and GOPC could be the position or size of the eGFP tag. Given that eGFP has a molecular weight of around 27 kDa on its own, it is conceivable that a short construct like UL56(105-157) folds back onto the tag, whereas longer constructs like UL56(105-207), UL56(105-234) or UL56(1-157) are not disturbed. Therefore, co-immunoprecipitation experiments with an eGFP-tag C-terminal to residues 105-157, or experiments with a smaller tag like a hemagglutinin- or His-tag could be performed to avoid such potential steric hindrance.

Protein GOPC regulates the trafficking of various receptors and membrane proteins such as ion channels and adhesion molecules to both the plasma membrane and the lysosome (365). Notable targets of GOPC are for example the ion channel cystic fibrosis transmembrane regulator (CFTR) (378, 379), the metabotropic glutamate receptor subtype-5a (mGluR5a) (380) and the voltage-gated potassium channel KV10.1 (381). Association of GOPC with these proteins is mediated by the PDZ domain of GOPC, which binds to extreme C-terminal cytosolic peptides. There is no distinct peptide motif bound by GOPC, even though it is known that the very C-terminal four to five residues are most important for the interaction and bind to the surface of the PDZ domain adjacent to the second beta-sheet (366, 382, 383). On the other hand, interactions between GOPC and other Golgi-localized proteins such as Syntaxin 6 or Golgin-160 are facilitated by the coiled-

coil regions of the N-terminus (residues 1-275) of GOPC, potentially contributing to Golgi structure and function (369, 384). It is not possible that UL56 binds to the PDZ domain of GOPC in the canonical fashion since the C terminus of UL56 contains the transmembrane domain and is inserted into the viral envelope. The data presented in this chapter hint towards UL56 interacting with GOPC via multiple sites.

Cellular proteins containing PDZ domains are targeted by viruses across a broad range of virus families to influence tight junction formation, tumorigenesis, cell signaling, apoptosis and many other processes (372). Based on the suggested functions of GOPC, the interaction between GOPC and UL56 could influence various cellular mechanisms. Overall, three models of the purpose of HSV-1 hijacking GOPC shall be proposed here.

First, it could be possible that UL56 binds to GOPC to exploit the cellular trafficking machinery. It has been previously demonstrated that HSV-1 glycoproteins, tegument proteins and capsids localize to the trans-Golgi network, suggesting that secondary envelopment might take place there (212). Via UL56 binding to GOPC, HSV-1 might access the cellular transport machinery and facilitate trafficking of vesicles containing mature virus particles to the plasma membrane. Second, interaction of UL56 with GOPC might also interfere with Golgi architecture, subsequently redistributing membranes of Golgi origin for viral envelopment (214). Campadelli and colleagues showed that in cells infected with HSV-1, the Golgi apparatus is fragmented and drastically rearranged, losing its well-stacked, organized nature and leading to the presence of Golgi-derived vacuoles containing virions in the cytoplasm (216). Third, it is conceivable that UL56 does not interact with GOPC to influence trafficking of viral proteins or particles, but to regulate the expression of cellular targets of GOPC such a cell surface receptors or ion channels. Interestingly, for equine Herpesvirus 1 and 4 (EHV-1/4), UL56-dependent downregulation of the cell surface proteins MHC-I, CD29, CD105 and CD172 has been described (235, 237). None of these proteins contain an obvious GOPC-binding motif at their C terminus, but given the degeneracy of PDZ motifs, interaction of these proteins with GOPC has to be evaluated experimentally. The relatively low sequence conservation of UL56 between equine and human herpesviruses (around 20% identity and 30% similarity for HSV-1 and EHV-1 UL56) should warrant caution in translating these results directly to HSV-1.

The latter two hypotheses are especially compelling in the light of recent results obtained by Timothy Soh and Colin Crump[2] (unpublished data). In time-course mass spectrometry experiments, they observed rapid degradation of GOPC in the presence of HSV-1 infection, starting at around two hours post infection and leading to a 90% loss of endogenous GOPC compared to mock-infected cells at 18 hours post infection. In this thesis, potential interactions of UL56 with numerous E3 ubiquitin ligases was demonstrated. There are hundreds of cellular E3 ubiquitin ligases, which facilitate the transfer of ubiquitin from an E2 ubiquitin-conjugating enzyme onto target proteins and thus mark them for degradation by the proteasome (for reviews of the ubiquitin system, see (385-387)). Therefore, a model integrating the interaction of UL56 with GOPC and the NEDD4-family E3 ubiquitin ligases, such as NEDD4 or ITCH, emerges (Figure 95). Cellular E3 ubiquitin ligases are recruited to UL56 via its PPPY motifs, located in the N terminus of the largely unfolded protein chain. GOPC interacts with residues 105-157 and potentially, additional regions of UL56, bringing GOPC into proximity with the E3 ubiquitin ligase. An E2 ubiquitin-conjugating enzyme transfers ubiquitin onto GOPC, leading to its degradation via the proteasome. Loss of GOPC could then lead to disruption of the Golgi or to faulty trafficking of cellular GOPC-interacting proteins.

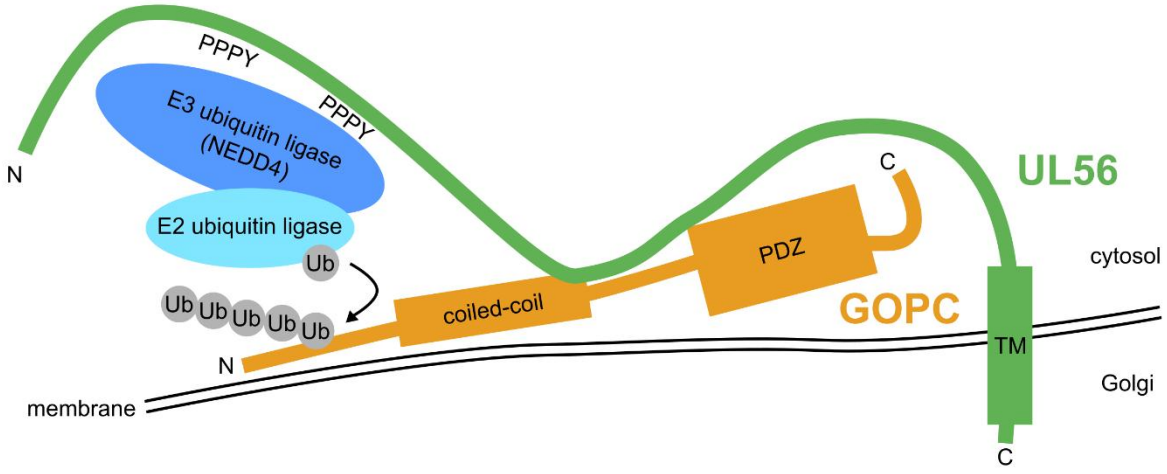


Figure 95 Model of UL56 interacting with GOPC and E3 ubiquitin ligases to promote degradation of GOPC.

[2] Department of Pathology, University of Cambridge

Interestingly, a similar mechanism of GOPC regulation has been reported by Jeong and colleagues for human papillomavirus 16 (HPV16). The C-terminus of protein E6 of human high-risk HPV16 interacts with the PDZ domain of GOPC, leading to the ubiquitination of GOPC by the cellular ubiquitin ligase E6AP and subsequent degradation by the proteasome (388). However, Jeong et al. did not address the biological significance of their observations. Therefore, more work is needed to elucidate the role of the UL56:GOPC interaction during HSV-1 infection.

7.4.6 Summary

This chapter describes a SILAC quantitative proteomics experiment conducted to find cellular interaction partners of an HSV-1 envelope protein. I used two different constructs of UL56, with or without its transmembrane domain, in the SILAC screen and performed stringent data analysis to identify significantly enriched hits. Bioinformatic analysis and manual curation of these hits enabled me to select the most promising candidates for follow-up studies. Co-immunoprecipitation and GST pull-down experiments with the chosen putative interaction partners of UL56 proved to be a suitable way of validating the SILAC screen hits. I demonstrated interactions between UL56 and GOPC, the human TRAPPII complex and the E3 ubiquitin ligase NEDD4 and presented models providing rationales for the exploitation of these proteins by HSV-1.

8 Conclusion

Viruses have evolved to exploit host cells to facilitate replication and subsequent release of virus progeny to spread infection. In this thesis, I examined interactions between viral and host cell proteins involved in envelopment or egress of hepatitis D virus and herpes simplex virus 1. For HDV, I investigated binding between the large viral antigen and clathrin N-terminal domain, which had been previously reported to be necessary for the egress of virus-like particles in cell culture. Using structural and biochemical techniques, I inspected the molecular details of this interaction and related it to binding of cellular proteins to NTD. For HSV-1, I performed unbiased screens to identify novel binding partners of the virus protein UL56.

The crystal structures of NTD in complex with viral and cellular peptides demonstrated that all investigated peptides can bind to multiple sites on clathrin NTD promiscuously. The binding modes of viral and cellular peptides were largely similar: conserved hydrogen bonds between the backbone of the bound peptide and NTD determined the overall peptide conformation, and key hydrophobic residues occupied hydrophobic pockets at the clathrin box, the arrestin box and a third binding site on NTD. Excitingly, the crystal structures enabled the structural characterization of this third binding site, the "Royle box", which corresponds to the interaction site most recently identified in functional studies. Moreover, it was discovered that linear peptides can bind to the arrestin box on NTD, an interaction site that was previously thought to be contacted only by tightly bent peptide forming part of a surface loop of arrestin 2. The binding affinities of cellular and viral peptides for NTD were too low to be quantified reliably, but nevertheless, biochemical experiments using wild-type NTD and NTD containing mutations at a single or multiple binding sites demonstrated that the interactions observed in the crystal structures contribute to binding in vitro.

Since HDV-derived peptides contact the same sites on NTD as cellular peptides, albeit more weakly, clathrin is probably not a suitable target for compound design aimed at interfering with the egress of HDV particles. Nonetheless, because binding of these viral peptides to multiple sites on NTD was observed, the interaction between cellular peptides and NTD was re-assessed. These studies yielded the unanticipated result that binding motifs for clathrin are degenerate. The

clathrin box-binding motif present in many cellular proteins involved in endocytosis, such as adaptor protein complexes or amphiphysin, was previously thought to mediate binding specifically to the clathrin box on NTD, but this study clearly demonstrated that peptides containing the canonical CBM can bind to both the clathrin box and the arrestin box. This observation provides a compelling explanation for the redundancy of clathrin recruitment observed in vivo, suggesting that even if clathrin NTD carries mutations at one binding site, association with its adaptor proteins is not abolished. Interactions between clathrin and adaptor proteins are generally very weak, with the unstructured protein regions containing clathrin-binding motifs dissociating from NTD quickly. If CBM peptides can bind to NTD promiscuously, the avidity of clathrin:adaptor interactions would be increased, especially for adaptor proteins like amphiphysin which contain more than one NTD-binding motif.

In future studies, the physiological relevance of the four different binding sites of NTD should be investigated. It has been shown previously that each binding site on its own is sufficient for functional clathrin-mediated endocytosis, but it has not been investigated whether the same applies to intracellular clathrin trafficking. It is also not known whether all interaction sites of NTD contribute equally to the recruitment of clathrin-binding proteins in vivo — potentially, due to steric constraints or competition with other proteins, clathrin-binding proteins could favor certain binding sites on NTD over others. The stably folded versions of NTD mutated at the different binding sites described in this study could greatly improve the design of new studies addressing these questions. In addition, even though the Royle box was structurally characterized in this thesis using two viral and one non-native cellular peptide, to date, no cellular proteins binding this site on NTD are known. Clearly, future work is needed to find genuine interaction partners of clathrin with occupy the Royle box and to identify a consensus binding motif, keeping in mind that it might be rather degenerate.

Herpes simplex virus 1 membrane protein UL56 had been previously implicated in the late stages of the viral life cycle. The work presented in this thesis demonstrated that UL56 is a largely disordered protein and therefore not suitable for crystallization in isolation. Nevertheless, it was demonstrated that UL56 interacts with a multitude of host cell proteins. Proteins identified as potential UL56 binding partners in yeast two-hybrid assays could not be validated by co-

immunoprecipitation experiments. However, an unbiased mass spectrometry-based screen led to the discovery of various promising putative interaction partners of UL56. Of particular interest are proteins GOPC, the human TRAPPII complex, and several Nedd4 ubiquitin ligases, all of which were confirmed as genuine binding partners of UL56 by biochemical techniques. UL56 could potentially interact with Nedd4 ubiquitin ligases to facilitate sorting of other viral proteins or whole virions to the site of envelopment, or to recruit the cellular ESCRT machinery. On the other hand, degradation or sequestration of GOPC or the TRAPPII complex could interfere with Golgi architecture, repurposing large quantities of cellular membranes for viral envelopment and therefore, focusing the whole cell on the production of viral progeny. Another compelling hypothesis would be that UL56 downregulates GOPC to regulate the trafficking of cell surface molecules, potentially influencing host cell ion transport or the host immune response to infection.

The identification of host cells proteins interacting with HSV-1 envelope protein UL56 provides ample possibilities for future work. Studies should investigate both functional aspects and the molecular details of the demonstrated interactions. Co-localization experiments could support the assumption that UL56 interacts with GOPC, TRAPPII and Nedd4 ubiquitin ligases during infection. Infection of cells depleted of the UL56 interaction partners, or comparative infection of cells with wild-type HSV-1 and UL56-deleted HSV-1, and subsequent assessment of Golgi architecture, distribution of the interaction partners, or the efficiency of viral envelopment could contribute to our understanding of the roles of these proteins during the HSV-1 life cycle. Investigation of the levels of cell surface proteins upon UL56 expression by membrane profiling mass spectrometry, by flow cytometry, or by fluorescence microscopy, could demonstrate whether the interaction between UL56 and GOPC influences trafficking of these proteins. For many of these proposed experiments, a UL56-specific monoclonal antibody would be highly valuable.

The interacting regions of UL56 and GOPC identified in this thesis require validation by further biochemical and biophysical experiments. Likewise, it would be interesting to know how UL56 binds to TRAPPC9 and TRAPPC10. After identifying the domains or regions mediating binding, GOPC, TRAPPC9 or TRAPPC10 could be co-crystallized with full-length UL56 or, more likely, peptides derived from UL56. Such structures could elucidate the molecular details of the interactions and therefore, provide insight into the mechanism of protein recruitment by UL56. The suggested functional and biochemical or structural experiments could also further our

understanding of the physiological roles of the U56-interacting proteins, which, at least for GOPC and TRAPPII, are currently only rudimentarily described.

In summary, in this thesis I demonstrated using two rather different viruses how detailed structural investigations and broad, unbiased screening can help to further our understanding of mechanisms involved viral envelopment or egress. Moreover, my work highlights how study of virus:host interactions can shed new light on the physiological function or structural features of cellular proteins.

References

1. Kang DS, Kern RC, Puthenveedu MA, von Zastrow M, Williams JC, Benovic JL. Structure of an arrestin2-clathrin complex reveals a novel clathrin binding domain that modulates receptor trafficking. J Biol Chem. 2009;284(43):29860-72.

2. Willox AK, Royle SJ. Functional analysis of interaction sites on the N-terminal domain of clathrin heavy chain. Traffic. 2012;13(1):70-81.

3. Laskowski RA, Swindells MB. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. J Chem Inf Model. 2011;51(10):2778-86.

4. Miele AE, Watson PJ, Evans PR, Traub LM, Owen DJ. Two distinct interaction motifs in amphiphysin bind two independent sites on the clathrin terminal domain beta-propeller. Nat Struct Mol Biol. 2004;11(3):242-8.

5. Lorizate M, Krausslich HG. Role of lipids in virus replication. Cold Spring Harb Perspect Biol. 2011;3(10):a004820.

6. Chen BJ, Lamb RA. Mechanisms for enveloped virus budding: can some viruses do without an ESCRT? Virology. 2008;372(2):221-32.

7. Brandenburg B, Zhuang X. Virus trafficking - learning from single-virus tracking. Nat Rev Microbiol. 2007;5(3):197-208.

8. Diekmann Y, Pereira-Leal JB. Evolution of intracellular compartmentalization. Biochem J. 2013;449(2):319-31.

9. Doherty GJ, McMahon HT. Mechanisms of endocytosis. Annu Rev Biochem. 2009;78:857-902.

10. Lippincott-Schwartz J, Roberts TH, Hirschberg K. Secretory protein trafficking and organelle dynamics in living cells. Annu Rev Cell Dev Biol. 2000;16:557-89.

11. Schlacht A, Herman EK, Klute MJ, Field MC, Dacks JB. Missing pieces of an ancient puzzle: evolution of the eukaryotic membrane-trafficking system. Cold Spring Harb Perspect Biol. 2014;6(10):a016048.

12. Bonifacino JS, Glick BS. The mechanisms of vesicle budding and fusion. Cell. 2004;116(2):153-66.

13. Cai H, Reinisch K, Ferro-Novick S. Coats, tethers, Rabs, and SNAREs work together to mediate the intracellular destination of a transport vesicle. Dev Cell. 2007;12(5):671-82.

14. Hurley JH, Boura E, Carlson LA, Rozycki B. Membrane budding. Cell. 2010;143(6):875-87.

15. Li G, Marlin MC. Rab family of GTPases. Methods Mol Biol. 2015;1298:1-15.

16. Stenmark H, Olkkonen VM. The Rab GTPase family. Genome Biol. 2001;2(5):REVIEWS3007.

17. Zerial M, McBride H. Rab proteins as membrane organizers. Nat Rev Mol Cell Biol. 2001;2(2):107-17.

18. Jahn R, Scheller RH. SNAREs--engines for membrane fusion. Nat Rev Mol Cell Biol. 2006;7(9):631-43.

19. Baumgart T, Hammond AT, Sengupta P, Hess ST, Holowka DA, Baird BA, et al. Large-scale fluid/fluid phase separation of proteins and lipids in giant plasma membrane vesicles. Proc Natl Acad Sci U S A. 2007;104(9):3165-70.

20. McMahon HT, Gallop JL. Membrane curvature and mechanisms of dynamic cell membrane remodelling. Nature. 2005;438(7068):590-6.

21. Richnau N, Fransson A, Farsad K, Aspenstrom P. RICH-1 has a BIN/Amphiphysin/Rvsp domain responsible for binding to membrane lipids and tubulation of liposomes. Biochem Biophys Res Commun. 2004;320(3):1034-42.

22. Peter BJ, Kent HM, Mills IG, Vallis Y, Butler PJ, Evans PR, et al. BAR domains as sensors of membrane curvature: the amphiphysin BAR structure. Science. 2004;303(5657):495-9.

23. Cuello LG, Cortes DM, Perozo E. Molecular architecture of the KvAP voltage-dependent K+ channel in a lipid bilayer. Science. 2004;306(5695):491-5.

24. Unwin N. Refined structure of the nicotinic acetylcholine receptor at 4A resolution. J Mol Biol. 2005;346(4):967-89.

25. Stachowiak JC, Schmid EM, Ryan CJ, Ann HS, Sasaki DY, Sherman MB, et al. Membrane bending by protein-protein crowding. Nat Cell Biol. 2012;14(9):944-9.

26. Chen Z, Atefi E, Baumgart T. Membrane Shape Instability Induced by Protein Crowding. Biophys J. 2016;111(9):1823-6.

27. Busch DJ, Houser JR, Hayden CC, Sherman MB, Lafer EM, Stachowiak JC. Intrinsically disordered proteins drive membrane curvature. Nat Commun. 2015;6:7875.

28. Rao Y, Haucke V. Membrane shaping by the Bin/amphiphysin/Rvs (BAR) domain protein superfamily. Cell Mol Life Sci. 2011;68(24):3983-93.

29. Faini M, Beck R, Wieland FT, Briggs JA. Vesicle coats: structure, function, and general principles of assembly. Trends Cell Biol. 2013;23(6):279-88.

30. Letourneur F, Gaynor EC, Hennecke S, Demolliere C, Duden R, Emr SD, et al. Coatomer is essential for retrieval of dilysine-tagged proteins to the endoplasmic reticulum. Cell. 1994;79(7):1199-207.

31. Barlowe C, Orci L, Yeung T, Hosobuchi M, Hamamoto S, Salama N, et al. COPII: a membrane coat formed by Sec proteins that drive vesicle budding from the endoplasmic reticulum. Cell. 1994;77(6):895-907.

32. Robinson MS. Forty Years of Clathrin-coated Vesicles. Traffic. 2015;16(12):1210-38.

33. Brodsky FM. Living with clathrin: its role in intracellular membrane traffic. Science. 1988;242(4884):1396-402.

34. Schoneberg J, Lee IH, Iwasa JH, Hurley JH. Reverse-topology membrane scission by the ESCRT proteins. Nat Rev Mol Cell Biol. 2017;18(1):5-17.

35. Katzmann DJ, Odorizzi G, Emr SD. Receptor downregulation and multivesicular-body sorting. Nat Rev Mol Cell Biol. 2002;3(12):893-905.

36. Votteler J, Sundquist WI. Virus budding and the ESCRT pathway. Cell Host Microbe. 2013;14(3):232-41.

37. Ren X, Hurley JH. Proline-rich regions and motifs in trafficking: from ESCRT interaction to viral exploitation. Traffic. 2011;12(10):1282-90.

38. Christ L, Raiborg C, Wenzel EM, Campsteijn C, Stenmark H. Cellular Functions and Molecular Mechanisms of the ESCRT Membrane-Scission Machinery. Trends Biochem Sci. 2017;42(1):42-56.

39. Mitsunari T, Nakatsu F, Shioda N, Love PE, Grinberg A, Bonifacino JS, et al. Clathrin adaptor AP-2 is essential for early embryonal development. Mol Cell Biol. 2005;25(21):9318-23.

40. Ferguson SM, Raimondi A, Paradise S, Shen H, Mesaki K, Ferguson A, et al. Coordinated actions of actin and BAR proteins upstream of dynamin at endocytic clathrin-coated pits. Dev Cell. 2009;17(6):811-22.

41. McMahon HT, Boucrot E. Molecular mechanism and physiological functions of clathrin-mediated endocytosis. Nat Rev Mol Cell Biol. 2011;12(8):517-33.

42. Bonifacino JS, Rojas R. Retrograde transport from endosomes to the trans-Golgi network. Nat Rev Mol Cell Biol. 2006;7(8):568-79.

43. Zhao Y, Keen JH. Gyrating clathrin: highly dynamic clathrin structures involved in rapid receptor recycling. Traffic. 2008;9(12):2253-64.

44. Ungewickell E, Branton D. Assembly units of clathrin coats. Nature. 1981;289(5796):420-2.

45. Heuser J. Three-dimensional visualization of coated vesicle formation in fibroblasts. J Cell Biol. 1980;84(3):560-83.

46. Dannhauser PN, Ungewickell EJ. Reconstitution of clathrin-coated bud and vesicle formation with minimal components. Nat Cell Biol. 2012;14(6):634-9.

47. Pearse BM. Coated vesicles from pig brain: purification and biochemical characterization. J Mol Biol. 1975;97(1):93-8.

48. Ybe JA, Brodsky FM, Hofmann K, Lin K, Liu SH, Chen L, et al. Clathrin self-assembly is mediated by a tandemly repeated superhelix. Nature. 1999;399(6734):371-5.

49. Brodsky FM. Diversity of clathrin function: new tricks for an old protein. Annu Rev Cell Dev Biol. 2012;28:309-36.

50. Fotin A, Cheng Y, Sliz P, Grigorieff N, Harrison SC, Kirchhausen T, et al. Molecular model for a complete clathrin lattice from electron cryomicroscopy. Nature. 2004;432(7017):573-9.

51. Fotin A, Kirchhausen T, Grigorieff N, Harrison SC, Walz T, Cheng Y. Structure determination of clathrin coats to subnanometer resolution by single particle cryo-electron microscopy. J Struct Biol. 2006;156(3):453-60.

52. Hirst J, Robinson MS. Clathrin and adaptors. Biochim Biophys Acta. 1998;1404(1-2):173-93.

53. Borner GH, Antrobus R, Hirst J, Bhumbra GS, Kozik P, Jackson LP, et al. Multivariate proteomic profiling identifies novel accessory proteins of coated vesicles. J Cell Biol. 2012;197(1):141-60.

54. Owen DJ, Collins BM, Evans PR. Adaptors for clathrin coats: structure and function. Annu Rev Cell Dev Biol. 2004;20:153-91.

55. Hirst J, Barlow LD, Francisco GC, Sahlender DA, Seaman MN, Dacks JB, et al. The fifth adaptor protein complex. PLoS Biol. 2011;9(10):e1001170.

56. Collins BM, McCoy AJ, Kent HM, Evans PR, Owen DJ. Molecular architecture and functional model of the endocytic AP2 complex. Cell. 2002;109(4):523-35.

57. Owen DJ, Evans PR. A structural explanation for the recognition of tyrosine-based endocytotic signals. Science. 1998;282(5392):1327-32.

58. Jackson LP, Kelly BT, McCoy AJ, Gaffry T, James LC, Collins BM, et al. A large-scale conformational change couples membrane recruitment to cargo binding in the AP2 clathrin adaptor complex. Cell. 2010;141(7):1220-9.

59. Kelly BT, McCoy AJ, Spate K, Miller SE, Evans PR, Honing S, et al. A structural explanation for the binding of endocytic dileucine motifs by the AP2 complex. Nature. 2008;456(7224):976-79.

60. Lundmark R, Carlsson SR. The beta-appendages of the four adaptor-protein (AP) complexes: structure and binding properties, and identification of sorting nexin 9 as an accessory protein to AP-2. Biochem J. 2002;362(Pt 3):597-607.

61. Owen DJ, Vallis Y, Pearse BM, McMahon HT, Evans PR. The structure and function of the beta 2-adaptin appendage domain. EMBO J. 2000;19(16):4216-27.

62. ter Haar E, Harrison SC, Kirchhausen T. Peptide-in-groove interactions link target proteins to the beta-propeller of clathrin. Proc Natl Acad Sci U S A. 2000;97(3):1096-100.

63. Edeling MA, Mishra SK, Keyel PA, Steinhauser AL, Collins BM, Roth R, et al. Molecular switches involving the AP-2 beta2 appendage regulate endocytic cargo selection and clathrin coat assembly. Dev Cell. 2006;10(3):329-42.

64. Robinson MS. 100-kD coated vesicle proteins: molecular heterogeneity and intracellular distribution studied with monoclonal antibodies. J Cell Biol. 1987;104(4):887-95.

65. Robinson MS. Immunofluorescent localization of 100K coated vesicle proteins. The Journal of Cell Biology. 1986;102(1):48-54.

66. Chang MP, Mallet WG, Mostov KE, Brodsky FM. Adaptor self-aggregation, adaptor-receptor recognition and binding of alpha-adaptin subunits to the plasma membrane contribute to recruitment of adaptor (AP2) components of clathrin-coated pits. EMBO J. 1993;12(5):2169-80.

67. Gaidarov I, Keen JH. Phosphoinositide-AP-2 interactions required for targeting to plasma membrane clathrin-coated pits. J Cell Biol. 1999;146(4):755-64.

68. Kelly BT, Graham SC, Liska N, Dannhauser PN, Honing S, Ungewickell EJ, et al. AP2 controls clathrin polymerization with a membrane-activated switch. Science. 2014;345(6195):459-63.

69. Lichte B, Veh RW, Meyer HE, Kilimann MW. Amphiphysin, a novel protein associated with synaptic vesicles. EMBO J. 1992;11(7):2521-30.

70. Prokic I, Cowling BS, Laporte J. Amphiphysin 2 (BIN1) in physiology and diseases. J Mol Med (Berl). 2014;92(5):453-63.

71. McMahon HT, Wigge P, Smith C. Clathrin interacts specifically with amphiphysin and is displaced by dynamin1. FEBS Letters. 1997;413(2):319-22.

72. David C, McPherson PS, Mundigl O, de Camilli P. A role of amphiphysin in synaptic vesicle endocytosis suggested by its binding to dynamin in nerve terminals. Proc Natl Acad Sci U S A. 1996;93(1):331-5.

73. Owen DJ, Wigge P, Vallis Y, Moore JD, Evans PR, McMahon HT. Crystal structure of the amphiphysin-2 SH3 domain and its role in the prevention of dynamin ring formation. EMBO J. 1998;17(18):5273-85.

74. ter Haar E, Musacchio A, Harrison SC, Kirchhausen T. Atomic structure of clathrin: a beta propeller terminal domain joins an alpha zigzag linker. Cell. 1998;95(4):563-73.

75. Dell'Angelica EC, Klumperman J, Stoorvogel W, Bonifacino JS. Association of the AP-3 adaptor complex with clathrin. Science. 1998;280(5362):431-4.

76. Krupnick JG, Goodman OB, Jr., Keen JH, Benovic JL. Arrestin/clathrin interaction. Localization of the clathrin binding domain of nonvisual arrestins to the carboxy terminus. J Biol Chem. 1997;272(23):15011-6.

77. Goodman OB, Jr., Krupnick JG, Gurevich VV, Benovic JL, Keen JH. Arrestin/clathrin interaction. Localization of the arrestin binding locus to the clathrin terminal domain. J Biol Chem. 1997;272(23):15017-22.

78. Dell'Angelica EC. Clathrin-binding proteins: got a motif? Join the network! Trends Cell Biol. 2001;11(8):315-8.

79. Drake MT, Traub LM. Interaction of two structurally distinct sequence types with the clathrin terminal domain beta-propeller. J Biol Chem. 2001;276(31):28700-9.

80. Ramjaun AR, McPherson PS. Multiple amphiphysin II splice variants display differential clathrin binding: identification of two distinct clathrin-binding sites. J Neurochem. 1998;70(6):2369-76.

81. Slepnev VI, Ochoa GC, Butler MH, De Camilli P. Tandem arrangement of the clathrin and AP-2 binding domains in amphiphysin 1 and disruption of clathrin coat function by amphiphysin fragments comprising these sites. J Biol Chem. 2000;275(23):17583-9.

82. Collette JR, Chi RJ, Boettner DR, Fernandez-Golbano IM, Plemel R, Merz AJ, et al. Clathrin functions in the absence of the terminal domain binding site for adaptor-associated clathrin-box motifs. Mol Biol Cell. 2009;20(14):3401-13.

83. Lemmon SK, Traub LM. Getting in touch with the clathrin terminal domain. Traffic. 2012;13(4):511-9.

84. Raiborg C, Bremnes B, Mehlum A, Gillooly DJ, D'Arrigo A, Stang E, et al. FYVE and coiled-coil domains determine the specific localisation of Hrs to early endosomes. J Cell Sci. 2001;114(Pt 12):2255-63.

85. Wollert T, Hurley JH. Molecular mechanism of multivesicular body biogenesis by ESCRT complexes. Nature. 2010;464(7290):864-9.

86. Mageswaran SK, Dixon MG, Curtiss M, Keener JP, Babst M. Binding to any ESCRT can mediate ubiquitin-independent cargo sorting. Traffic. 2014;15(2):212-29.

87. Kostelansky MS, Sun J, Lee S, Kim J, Ghirlando R, Hierro A, et al. Structural and functional organization of the ESCRT-I trafficking complex. Cell. 2006;125(1):113-26.

88. Teo H, Gill DJ, Sun J, Perisic O, Veprintsev DB, Vallis Y, et al. ESCRT-I core and ESCRT-II GLUE domain structures reveal role for GLUE in linking to ESCRT-I and membranes. Cell. 2006;125(1):99-111.

89. Yorikawa C, Shibata H, Waguri S, Hatta K, Horii M, Katoh K, et al. Human CHMP6, a myristoylated ESCRT-III protein, interacts directly with an ESCRT-II component EAP20 and regulates endosomal cargo sorting. Biochem J. 2005;387(Pt 1):17-26.

90. Teis D, Saksena S, Emr SD. Ordered assembly of the ESCRT-III complex on endosomes is required to sequester cargo during MVB formation. Dev Cell. 2008;15(4):578-89.

91. Saksena S, Wahlman J, Teis D, Johnson AE, Emr SD. Functional reconstitution of ESCRT-III assembly and disassembly. Cell. 2009;136(1):97-109.

92. Hanson PI, Roth R, Lin Y, Heuser JE. Plasma membrane deformation by circular arrays of ESCRT-III protein filaments. J Cell Biol. 2008;180(2):389-402.

93. Lata S, Schoehn G, Jain A, Pires R, Piehler J, Gottlinger HG, et al. Helical structures of ESCRT-III are disassembled by VPS4. Science. 2008;321(5894):1354-7.

94. Yang B, Stjepanovic G, Shen Q, Martin A, Hurley JH. Vps4 disassembles an ESCRT-III filament by global unfolding and processive translocation. Nat Struct Mol Biol. 2015;22(6):492-8.

95. Henne WM, Stenmark H, Emr SD. Molecular mechanisms of the membrane sculpting ESCRT pathway. Cold Spring Harb Perspect Biol. 2013;5(9).

96. DeTulleo L, Kirchhausen T. The clathrin endocytic pathway in viral infection. EMBO J. 1998;17(16):4585-93.

97. Rust MJ, Lakadamyali M, Zhang F, Zhuang X. Assembly of endocytic machinery around individual influenza viruses during viral entry. Nat Struct Mol Biol. 2004;11(6):567-73.

98. Cureton DK, Massol RH, Whelan SP, Kirchhausen T. The length of vesicular stomatitis virus particles dictates a need for actin assembly during clathrin-dependent endocytosis. PLoS Pathog. 2010;6(9):e1001127.

99. Ren X, Park SY, Bonifacino JS, Hurley JH. How HIV-1 Nef hijacks the AP-2 clathrin adaptor to downregulate CD4. Elife. 2014;3:e01754.

100. Lubben NB, Sahlender DA, Motley AM, Lehner PJ, Benaroch P, Robinson MS. HIV-1 Nef-induced down-regulation of MHC class I requires AP-1 and clathrin but not PACS-1 and is impeded by AP-2. Mol Biol Cell. 2007;18(9):3351-65.

101. Ivanovic T, Boulant S, Ehrlich M, Demidenko AA, Arnold MM, Kirchhausen T, et al. Recruitment of cellular clathrin to viral factories and disruption of clathrin-dependent trafficking. Traffic. 2011;12(9):1179-95.

102. Humphries AC, Way M. The non-canonical roles of clathrin and actin in pathogen internalization, egress and spread. Nature Reviews Microbiology. 2013;11(8):551-60.

103. Benedicto I, Gondar V, Molina-Jimenez F, Garcia-Buey L, Lopez-Cabrera M, Gastaminza P, et al. Clathrin mediates infectious hepatitis C virus particle egress. J Virol. 2015;89(8):4180-90.

104. Camus G, Segura-Morales C, Molle D, Lopez-Verges S, Begon-Pescia C, Cazevieille C, et al. The clathrin adaptor complex AP-1 binds HIV-1 and MLV Gag and facilitates their budding. Mol Biol Cell. 2007;18(8):3193-203.

105. Wen X, Ding L, Hunter E, Spearman P. An siRNA screen of membrane trafficking genes highlights pathways common to HIV-1 and M-PMV virus assembly and release. PLoS One. 2014;9(9):e106151.

106. Huang C, Chang SC, Yang HC, Chien CL, Chang MF. Clathrin-mediated post-Golgi membrane trafficking in the morphogenesis of hepatitis delta virus. J Virol. 2009;83(23):12314-24.

107. Huang C, Chang SC, Yu IC, Tsay YG, Chang MF. Large hepatitis delta antigen is a novel clathrin adaptor-like protein. J Virol. 2007;81(11):5985-94.

108. Humphries AC, Dodding MP, Barry DJ, Collinson LM, Durkin CH, Way M. Clathrin potentiates vaccinia-induced actin polymerization to facilitate viral spread. Cell Host Microbe. 2012;12(3):346-59.

109. Schmidt FI, Mercer J. Vaccinia virus egress: actin OUT with clathrin. Cell Host Microbe. 2012;12(3):263-5.

110. McDonald B, Martin-Serrano J. No strings attached: the ESCRT machinery in viral budding and cytokinesis. J Cell Sci. 2009;122(Pt 13):2167-77.

111. Gottlinger HG, Dorfman T, Sodroski JG, Haseltine WA. Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release. Proc Natl Acad Sci U S A. 1991;88(8):3195-9.

112. Huang M, Orenstein JM, Martin MA, Freed EO. p6Gag is required for particle production from full-length human immunodeficiency virus type 1 molecular clones expressing protease. J Virol. 1995;69(11):6810-8.

113. Martin-Serrano J, Zang T, Bieniasz PD. HIV-1 and Ebola virus encode small peptide motifs that recruit Tsg101 to sites of particle assembly to facilitate egress. Nat Med. 2001;7(12):1313-9.

114. VerPlank L, Bouamr F, LaGrassa TJ, Agresta B, Kikonyogo A, Leis J, et al. Tsg101, a homologue of ubiquitin-conjugating (E2) enzymes, binds the L domain in HIV type 1 Pr55(Gag). Proc Natl Acad Sci U S A. 2001;98(14):7724-9.

115. Demirov DG, Ono A, Orenstein JM, Freed EO. Overexpression of the N-terminal domain of TSG101 inhibits HIV-1 budding by blocking late domain function. Proc Natl Acad Sci U S A. 2002;99(2):955-60.

116. Garrus JE, von Schwedler UK, Pornillos OW, Morham SG, Zavitz KH, Wang HE, et al. Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. Cell. 2001;107(1):55-65.

117. Puffer BA, Parent LJ, Wills JW, Montelaro RC. Equine infectious anemia virus utilizes a YXXL motif within the late assembly domain of the Gag p9 protein. J Virol. 1997;71(9):6541-6.

118. Rotin D, Kumar S. Physiological functions of the HECT family of ubiquitin ligases. Nat Rev Mol Cell Biol. 2009;10(6):398-409.

119. Martin-Serrano J, Eastman SW, Chung W, Bieniasz PD. HECT ubiquitin ligases link viral and cellular PPXY motifs to the vacuolar protein-sorting pathway. J Cell Biol. 2005;168(1):89-101.

120. Bouamr F, Melillo JA, Wang MQ, Nagashima K, de Los Santos M, Rein A, et al. PPPYEPTAP Motif Is the Late Domain of Human T-Cell Leukemia Virus Type 1 Gag and Mediates Its Functional Interaction with Cellular Proteins Nedd4 and Tsg101. Journal of Virology. 2003;77(22):11882-95.

121. Han Z, Sagum CA, Bedford MT, Sidhu SS, Sudol M, Harty RN. ITCH E3 Ubiquitin Ligase Interacts with Ebola Virus VP40 To Regulate Budding. J Virol. 2016;90(20):9163-71.

122. Harty RN, Paragas J, Sudol M, Palese P. A proline-rich motif within the matrix protein of vesicular stomatitis virus and rabies virus interacts with WW domains of cellular proteins: implications for viral budding. J Virol. 1999;73(4):2921-9.

123. Jadwin JA, Rudd V, Sette P, Challa S, Bouamr F. Late domain-independent rescue of a release-deficient Moloney murine leukemia virus by the ubiquitin ligase itch. J Virol. 2010;84(2):704-15.

124. Rauch S, Martin-Serrano J. Multiple interactions between the ESCRT machinery and arrestin-related proteins: implications for PPXY-dependent budding. J Virol. 2011;85(7):3546-56.

125. Rossman JS, Jing X, Leser GP, Lamb RA. Influenza virus M2 protein mediates ESCRT-independent membrane scission. Cell. 2010;142(6):902-13.

126. Krupovic M, Kuhn JH, Fischer MG. A classification system for virophages and satellite viruses. Arch Virol. 2016;161(1):233-47.

127. Lempp FA, Ni Y, Urban S. Hepatitis delta virus: insights into a peculiar pathogen and novel treatment options. Nat Rev Gastroenterol Hepatol. 2016;13(10):580-9.

128. Ramia S, Bahakim H. Perinatal transmission of hepatitis B virus-associated hepatitis D virus. Ann Inst Pasteur Virol. 1988;139(3):285-90.

129. Yurdaydin C, Idilman R, Bozkaya H, Bozdayi AM. Natural history and treatment of chronic delta hepatitis. J Viral Hepat. 2010;17(11):749-56.

130. Smedile A, Farci P, Verme G, Caredda F, Cargnel A, Caporaso N, et al. Influence of delta infection on severity of hepatitis B. Lancet. 1982;2(8305):945-7.

131. Negro F. Hepatitis D virus coinfection and superinfection. Cold Spring Harb Perspect Med. 2014;4(11):a021550.

132. Ciancio A, Rizzetto M. Chronic hepatitis D at a standstill: where do we go from here? Nat Rev Gastroenterol Hepatol. 2014;11(1):68-71.

133. Pascarella S, Negro F. Hepatitis D virus: an update. Liver Int. 2011;31(1):7-21.

134. Poovorawan Y, Chatchatee P, Chongsrisawat V. Epidemiology and prophylaxis of viral hepatitis: a global perspective. J Gastroenterol Hepatol. 2002;17 Suppl:S155-66.

135. Schilsky ML. Hepatitis B "360". Transplant Proc. 2013;45(3):982-5.

136. Hughes SA, Wedemeyer H, Harrison PM. Hepatitis delta virus. The Lancet. 2011;378(9785):73-85.

137. Chen PJ, Kalpana G, Goldberg J, Mason W, Werner B, Gerin J, et al. Structure and replication of the genome of the hepatitis delta virus. Proc Natl Acad Sci U S A. 1986;83(22):8774-8.

138. Kos A, Dijkema R, Arnberg AC, van der Meide PH, Schellekens H. The hepatitis delta (delta) virus possesses a circular RNA. Nature. 1986;323(6088):558-60.

139. Wang KS, Choo QL, Weiner AJ, Ou JH, Najarian RC, Thayer RM, et al. Structure, sequence and expression of the hepatitis delta (delta) viral genome. Nature. 1986;323(6088):508-14.

140. Tsagris EM, Martinez de Alba AE, Gozmanova M, Kalantidis K. Viroids. Cell Microbiol. 2008;10(11):2168-79.

141. Flores R, Owens RA, Taylor J. Pathogenesis by subviral agents: viroids and hepatitis delta virus. Curr Opin Virol. 2016;17:87-94.

142. Bergmann KF, Gerin JL. Antigens of hepatitis delta virus in the liver and serum of humans and animals. J Infect Dis. 1986;154(4):702-6.

143. Hwang SB, Lee CZ, Lai MM. Hepatitis delta antigen expressed by recombinant baculoviruses: comparison of biochemical properties and post-translational modifications between the large and small forms. Virology. 1992;190(1):413-22.

144. Ryu WS, Netter HJ, Bayer M, Taylor J. Ribonucleoprotein complexes of hepatitis delta virus. J Virol. 1993;67(6):3281-7.

145. Lin BC, Defenbaugh DA, Casey JL. Multimerization of hepatitis delta antigen is a critical determinant of RNA binding specificity. J Virol. 2010;84(3):1406-13.

146. Zuccola HJ, Rozzelle JE, Lemon SM, Erickson BW, Hogle JM. Structural basis of the oligomerization of hepatitis delta antigen. Structure. 1998;6(7):821-30.

147. Wang CJ, Chen PJ, Wu JC, Patel D, Chen DS. Small-form hepatitis B surface antigen is sufficient to help in the assembly of hepatitis delta virus-like particles. J Virol. 1991;65(12):6630-6.

148. Hwang SB, Lai MM. Isoprenylation mediates direct protein-protein interactions between hepatitis large delta antigen and hepatitis B virus surface antigen. J Virol. 1993;67(12):7659-62.

149. Blanchet M, Sureau C. Analysis of the cytosolic domains of the hepatitis B virus envelope proteins for their function in viral particle assembly and infectivity. J Virol. 2006;80(24):11935-45.

150. Komla-Soukha I, Sureau C. A tryptophan-rich motif in the carboxyl terminus of the small envelope protein of hepatitis B virus is central to the assembly of hepatitis delta virus particles. J Virol. 2006;80(10):4648-55.

151. Sureau C, Guerra B, Lanford RE. Role of the large hepatitis B virus envelope protein in infectivity of the hepatitis delta virion. J Virol. 1993;67(1):366-72.

152. Sureau C, Guerra B, Lee H. The middle hepatitis B virus envelope protein is not necessary for infectivity of hepatitis delta virus. J Virol. 1994;68(6):4063-6.

153. Lamas Longarela O, Schmidt TT, Schoneweis K, Romeo R, Wedemeyer H, Urban S, et al. Proteoglycans act as cellular hepatitis delta virus attachment receptors. PLoS One. 2013;8(3):e58340.

154. Yan H, Zhong G, Xu G, He W, Jing Z, Gao Z, et al. Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis B and D virus. eLife. 2012;1.

155. Ni Y, Lempp FA, Mehrle S, Nkongolo S, Kaufman C, Falth M, et al. Hepatitis B and D viruses exploit sodium taurocholate co-transporting polypeptide for species-specific entry into hepatocytes. Gastroenterology. 2014;146(4):1070-83.

156. Urban S, Bartenschlager R, Kubitz R, Zoulim F. Strategies to inhibit entry of HBV and HDV into hepatocytes. Gastroenterology. 2014;147(1):48-64.

157. Konig A, Doring B, Mohr C, Geipel A, Geyer J, Glebe D. Kinetics of the bile acid transporter and hepatitis B virus receptor Na+/taurocholate cotransporting polypeptide (NTCP) in hepatocytes. J Hepatol. 2014;61(4):867-75.

158. Schieck A, Schulze A, Gahler C, Muller T, Haberkorn U, Alexandrov A, et al. Hepatitis B virus hepatotropism is mediated by specific receptor recognition in the liver and not restricted to susceptible hosts. Hepatology. 2013;58(1):43-53.

159. Sureau C, Negro F. The hepatitis delta virus: Replication and pathogenesis. J Hepatol. 2016;64(1 Suppl):S102-16.

160. Chou HC, Hsieh TY, Sheu GT, Lai MM. Hepatitis delta antigen mediates the nuclear import of hepatitis delta virus RNA. J Virol. 1998;72(5):3684-90.

161. Tavanez JP, Cunha C, Silva MC, David E, Monjardino J, Carmo-Fonseca M. Hepatitis delta virus ribonucleoproteins shuttle between the nucleus and the cytoplasm. RNA. 2002;8(5):637-46.

162. Chang J, Nie X, Chang HE, Han Z, Taylor J. Transcription of hepatitis delta virus RNA by RNA polymerase II. J Virol. 2008;82(3):1118-27.

163. Wu HN, Lin YJ, Lin FP, Makino S, Chang MF, Lai MM. Human hepatitis delta virus RNA subfragments contain an autocleavage activity. Proc Natl Acad Sci U S A. 1989;86(6):1831-5.

164. Shih IH, Been MD. Catalytic strategies of the hepatitis delta virus ribozymes. Annu Rev Biochem. 2002;71:887-917.

165. Sharmeen L, Kuo MY, Taylor J. Self-ligating RNA sequences on the antigenome of human hepatitis delta virus. J Virol. 1989;63(3):1428-30.

166. Reid CE, Lazinski DW. A host-specific function is required for ligation of a wide variety of ribozyme-processed RNAs. Proc Natl Acad Sci U S A. 2000;97(1):424-9.

167. Modahl LE, Lai MM. Transcription of hepatitis delta antigen mRNA continues throughout hepatitis delta virus (HDV) replication: a new model of HDV RNA transcription and replication. J Virol. 1998;72(7):5449-56.

168. Wong SK, Lazinski DW. Replicating hepatitis delta virus RNA is edited in the nucleus by the small form of ADAR1. Proc Natl Acad Sci U S A. 2002;99(23):15118-23.

169. Hwang SB, Lai MM. Isoprenylation masks a conformational epitope and enhances trans-dominant inhibitory function of the large hepatitis delta antigen. J Virol. 1994;68(5):2958-64.

170. Chao M, Hsieh SY, Taylor J. Role of two forms of hepatitis delta virus antigen: evidence for a mechanism of self-limiting genome replication. J Virol. 1990;64(10):5066-9.

171. Hsieh SY, Chao M, Coates L, Taylor J. Hepatitis delta virus genome replication: a polyadenylated mRNA for delta antigen. J Virol. 1990;64(7):3192-8.

172. Gudima S, Wu SY, Chiang CM, Moraleda G, Taylor J. Origin of hepatitis delta virus mRNA. J Virol. 2000;74(16):7204-10.

173. Macnaughton TB, Lai MM. Genomic but not antigenomic hepatitis delta virus RNA is preferentially exported from the nucleus immediately after synthesis and processing. J Virol. 2002;76(8):3928-35.

174. Lambert C, Doring T, Prange R. Hepatitis B virus maturation is sensitive to functional inhibition of ESCRT-III, Vps4, and gamma 2-adaptin. J Virol. 2007;81(17):9050-60.

175. Patient R, Hourioux C, Roingeard P. Morphogenesis of hepatitis B virus and its subviral envelope particles. Cell Microbiol. 2009;11(11):1561-70.

176. Patzer EJ, Nakamura GR, Simonsen CC, Levinson AD, Brands R. Intracellular assembly and packaging of hepatitis B surface antigen particles occur in the endoplasmic reticulum. J Virol. 1986;58(3):884-92.

177. Wang YC, Huang CR, Chao M, Lo SJ. The C-terminal sequence of the large hepatitis delta antigen is variable but retains the ability to bind clathrin. Virol J. 2009;6:31.

178. Radjef N, Gordien E, Ivaniushina V, Gault E, Anais P, Drugan T, et al. Molecular Phylogenetic Analyses Indicate a Wide and Ancient Radiation of African Hepatitis Delta Virus, Suggesting a Deltavirus Genus of at Least Seven Major Clades. Journal of Virology. 2004;78(5):2537-44.

179. Le Gal F, Gault E, Ripault MP, Serpaggi J, Trinchet JC, Gordien E, et al. Eighth major clade for hepatitis delta virus. Emerg Infect Dis. 2006;12(9):1447-50.

180. Huang CR, Lo SJ. Evolution and diversity of the human hepatitis d virus genome. Adv Bioinformatics. 2010:323654.

181. Davison AJ, Eberle R, Ehlers B, Hayward GS, McGeoch DJ, Minson AC, et al. The order Herpesvirales. Arch Virol. 2009;154(1):171-7.

182. Steiner I, Benninger F. Update on herpes virus infections of the nervous system. Curr Neurol Neurosci Rep. 2013;13(12):414.

183. Azher TN, Yin XT, Tajfirouz D, Huang AJ, Stuart PM. Herpes simplex keratitis: challenges in diagnosis and clinical management. Clin Ophthalmol. 2017;11:185-91.

184. Sauerbrei A. Herpes Genitalis: Diagnosis, Treatment and Prevention. Geburtshilfe Frauenheilkd. 2016;76(12):1310-7.

185. Ramanathan J, Rammouni M, Baran J, Jr., Khatib R. Herpes simplex virus esophagitis in the immunocompetent host: an overview. Am J Gastroenterol. 2000;95(9):2171-6.

186. World Health Organization fact sheet: Herpes simplex virus. 2017 [updated January 2017. Available from: http://www.who.int/mediacentre/factsheets/fs400/en/.

187. Looker KJ, Magaret AS, May MT, Turner KM, Vickerman P, Newman LM, et al. First estimates of the global and regional incidence of neonatal herpes infection. Lancet Glob Health. 2017;5(3):e300-e9.

188. Whitley RJ, Roizman B. Herpes simplex virus infections. The Lancet. 2001;357(9267):1513-8.

189. Friedman HM. Keratin, a dual role in herpes simplex virus pathogenesis. J Clin Virol. 2006;35(1):103-5.

190. Smith G. Herpesvirus transport to the nervous system and back again. Annu Rev Microbiol. 2012;66:153-76.

191. Nicoll MP, Proenca JT, Efstathiou S. The molecular basis of herpes simplex virus latency. FEMS Microbiol Rev. 2012;36(3):684-705.

192. Balfour HH, Jr. Antiviral drugs. N Engl J Med. 1999;340(16):1255-68.

193. Fields BN, Knipe DM, Howley PM. Fields Virology. Sixth edition ed: Lippincott, Williams & Wilkins; 2013.

194. Zhou ZH, Dougherty M, Jakana J, He J, Rixon FJ, Chiu W. Seeing the herpesvirus capsid at 8.5 A. Science. 2000;288(5467):877-80.

195. Grunewald K, Desai P, Winkler DC, Heymann JB, Belnap DM, Baumeister W, et al. Three-dimensional structure of herpes simplex virus from cryo-electron tomography. Science. 2003;302(5649):1396-8.

196. Loret S, Guay G, Lippe R. Comprehensive characterization of extracellular herpes simplex virus type 1 virions. J Virol. 2008;82(17):8605-18.

197. Connolly SA, Jackson JO, Jardetzky TS, Longnecker R. Fusing structure and function: a structural view of the herpesvirus entry machinery. Nat Rev Microbiol. 2011;9(5):369-81.

198. Spear PG. Herpes simplex virus: receptors and ligands for cell entry. Cell Microbiol. 2004;6(5):401-10.

199. Herold BC, Visalli RJ, Susmarski N, Brandt CR, Spear PG. Glycoprotein C-independent binding of herpes simplex virus to cells requires cell surface heparan sulphate and glycoprotein B. J Gen Virol. 1994;75 (Pt 6):1211-22.

200. Kopp SJ, Banisadr G, Glajch K, Maurer UE, Grunewald K, Miller RJ, et al. Infection of neurons and encephalitis after intracranial inoculation of herpes simplex virus requires the entry receptor nectin-1. Proc Natl Acad Sci U S A. 2009;106(42):17916-20.

201. Nicola AV, Hou J, Major EO, Straus SE. Herpes simplex virus type 1 enters human epidermal keratinocytes, but not neurons, via a pH-dependent endocytic pathway. J Virol. 2005;79(12):7609-16.

202. Maurer UE, Sodeik B, Grunewald K. Native 3D intermediates of membrane fusion in herpes simplex virus 1 entry. Proc Natl Acad Sci U S A. 2008;105(30):10559-64.

203. Wolfstein A, Nagel CH, Radtke K, Dohner K, Allan VJ, Sodeik B. The inner tegument promotes herpes simplex virus capsid motility along microtubules in vitro. Traffic. 2006;7(2):227-37.

204. Dohner K, Nagel CH, Sodeik B. Viral stop-and-go along microtubules: taking a ride with dynein and kinesins. Trends Microbiol. 2005;13(7):320-7.

205. Aggarwal A, Miranda-Saksena M, Boadle RA, Kelly BJ, Diefenbach RJ, Alam W, et al. Ultrastructural visualization of individual tegument protein dissociation during entry of herpes simplex virus 1 into human and rat dorsal root ganglion neurons. J Virol. 2012;86(11):6123-37.

206. Ojala PM, Sodeik B, Ebersold MW, Kutay U, Helenius A. Herpes simplex virus type 1 entry into host cells: reconstitution of capsid binding and uncoating at the nuclear pore complex in vitro. Mol Cell Biol. 2000;20(13):4922-31.

207. Mettenleiter TC, Klupp BG, Granzow H. Herpesvirus assembly: an update. Virus Res. 2009;143(2):222-34.

208. Zeev-Ben-Mordehai T, Hagen C, Grunewald K. A cool hybrid approach to the herpesvirus 'life' cycle. Curr Opin Virol. 2014;5:42-9.

209. Mettenleiter TC. Budding events in herpesvirus morphogenesis. Virus Res. 2004;106(2):167-80.

210. Owen DJ, Crump CM, Graham SC. Tegument Assembly and Secondary Envelopment of Alphaherpesviruses. Viruses. 2015;7(9):5084-114.

211. Hollinshead M, Johns HL, Sayers CL, Gonzalez-Lopez C, Smith GL, Elliott G. Endocytic tubules regulated by Rab GTPases 5 and 11 are used for envelopment of herpes simplex virus. EMBO J. 2012;31(21):4204-20.

212. Sugimoto K, Uema M, Sagara H, Tanaka M, Sata T, Hashimoto Y, et al. Simultaneous tracking of capsid, tegument, and envelope protein localization in living cells infected with triply fluorescent herpes simplex virus 1. J Virol. 2008;82(11):5198-211.

213. Turcotte S, Letellier J, Lippe R. Herpes simplex virus type 1 capsids transit by the trans-Golgi network, where viral glycoproteins accumulate independently of capsid egress. J Virol. 2005;79(14):8847-60.

214. Johnson DC, Baines JD. Herpesviruses remodel host membranes for virus egress. Nat Rev Microbiol. 2011;9(5):382-94.

215. Albecka A, Laine RF, Janssen AF, Kaminski CF, Crump CM. HSV-1 Glycoproteins Are Delivered to Virus Assembly Sites Through Dynamin-Dependent Endocytosis. Traffic. 2016;17(1):21-39.

216. Campadelli G, Brandimarti R, Di Lazzaro C, Ward PL, Roizman B, Torrisi MR. Fragmentation and dispersal of Golgi proteins and redistribution of glycoproteins and glycolipids processed through the Golgi apparatus after infection with herpes simplex virus 1. Proc Natl Acad Sci U S A. 1993;90(7):2798-802.

217. Wisner TW, Johnson DC. Redistribution of cellular and herpes simplex virus proteins from the trans-golgi network to cell junctions without enveloped capsids. J Virol. 2004;78(21):11519-35.

218. Crump CM, Yates C, Minson T. Herpes simplex virus type 1 cytoplasmic envelopment requires functional Vps4. J Virol. 2007;81(14):7380-7.

219. Pawliczek T, Crump CM. Herpes simplex virus type 1 production requires a functional ESCRT-III complex but is independent of TSG101 and ALIX expression. J Virol. 2009;83(21):11254-64.

220. Ushijima Y, Goshima F, Kimura H, Nishiyama Y. Herpes simplex virus type 2 tegument protein UL56 relocalizes ubiquitin ligase Nedd4 and has a role in transport and/or release of virions. Virol J. 2009;6:168.

221. Ushijima Y, Koshizuka T, Goshima F, Kimura H, Nishiyama Y. Herpes simplex virus type 2 UL56 interacts with the ubiquitin ligase Nedd4 and increases its ubiquitination. J Virol. 2008;82(11):5220-33.

222. Ushijima Y, Luo C, Kamakura M, Goshima F, Kimura H, Nishiyama Y. Herpes simplex virus UL56 interacts with and regulates the Nedd4-family ubiquitin ligase Itch. Virol J. 2010;7:179.

223. Kratchmarov R, Taylor MP, Enquist LW. Making the case: married versus separate models of alphaherpes virus anterograde transport in axons. Rev Med Virol. 2012;22(6):378-91.

224. Koshizuka T, Goshima F, Takakuwa H, Nozawa N, Daikoku T, Koiwai O, et al. Identification and Characterization of the UL56 Gene Product of Herpes Simplex Virus Type 2. Journal of Virology. 2002;76(13):6718-28.

225. Brown J. Effect of gene location on the evolutionary rate of amino acid substitutions in herpes simplex virus proteins. Virology. 2004;330(1):209-20.

226. Rosen-Wolff A, Darai G. Identification and mapping of the UL56 gene transcript of herpes simplex virus type 1. Virus Res. 1991;19(1):115-26.

227. Rosen-Wolff A, Lamade W, Berkowitz C, Becker Y, Darai G. Elimination of UL56 gene by insertion of LacZ cassette between nucleotide position 116030 to 121753 of the herpes simplex virus

type 1 genome abrogates intraperitoneal pathogenicity in tree shrews and mice. Virus Res. 1991;20(3):205-21.

228. Nash TC, Spivack JG. The UL55 and UL56 genes of herpes simplex virus type 1 are not required for viral replication, intraperitoneal virulence, or establishment of latency in mice. Virology. 1994;204(2):794-8.

229. Kehm R, Rosen-Wolff A, Darai G. Restitution of the UL56 gene expression of HSV-1 HFEM led to restoration of virulent phenotype; deletion of the amino acids 217 to 234 of the UL56 protein abrogates the virulent phenotype. Virus Res. 1996;40(1):17-31.

230. Koshizuka T, Kawaguchi Y, Goshima F, Mori I, Nishiyama Y. Association of two membrane proteins encoded by herpes simplex virus type 2, UL11 and UL56. Virus Genes. 2006;32(2):153-63.

231. Kehm R, Gelderblom HR, Darai G. Identification of the UL56 protein of herpes simplex virus type 1 within the virion by immuno electron microscopy. Virus Genes. 1998;17(1):49-53.

232. Kehm R, Lorentzen E, Rosen-Wolff A, Darai G. In vitro expression of UL56 gene of herpes simplex virus type 1; detection of UL56 gene product in infected cells and in virions. Virus Res. 1994;33(1):55-66.

233. Koshizuka T, Kawaguchi Y, Nishiyama Y. Herpes simplex virus type 2 membrane protein UL56 associates with the kinesin motor protein KIF1A. J Gen Virol. 2005;86(Pt 3):527-33.

234. Daniel GR, Sollars PJ, Pickard GE, Smith GA. The pseudorabies virus protein, pUL56, enhances virus dissemination and virulence but is dispensable for axonal transport. Virology. 2016;488:179-86.

235. Claessen C, Favoreel H, Ma G, Osterrieder N, De Schauwer C, Piepers S, et al. Equid herpesvirus 1 (EHV1) infection of equine mesenchymal stem cells induces a pUL56-dependent downregulation of select cell surface markers. Vet Microbiol. 2015;176(1-2):32-9.

236. Said A, Azab W, Damiani A, Osterrieder N. Equine herpesvirus type 4 UL56 and UL49.5 proteins downregulate cell surface major histocompatibility complex class I expression independently of each other. J Virol. 2012;86(15):8059-71.

237. Ma G, Feineis S, Osterrieder N, Van de Walle GR. Identification and characterization of equine herpesvirus type 1 pUL56 and its role in virus-induced downregulation of major histocompatibility complex class I. J Virol. 2012;86(7):3554-63.

238. Johnson DC, Frame MC, Ligas MW, Cross AM, Stow ND. Herpes simplex virus immunoglobulin G Fc receptor activity depends on a complex of two viral glycoproteins, gE and gI. J Virol. 1988;62(4):1347-54.

239. Johnson DC, Feenstra V. Identification of a novel herpes simplex virus type 1-induced glycoprotein which complexes with gE and binds immunoglobulin. J Virol. 1987;61(7):2208-16.

240. Sprague ER, Wang C, Baker D, Bjorkman PJ. Crystal structure of the HSV-1 Fc receptor bound to Fc reveals a mechanism for antibody bipolar bridging. PLoS Biol. 2006;4(6):e148.

241. Awasthi S, Friedman HM. Molecular association of herpes simplex virus type 1 glycoprotein E with membrane protein Us9. Arch Virol. 2016;161(11):3203-13.

242. Dingwell KS, Brunetti CR, Hendricks RL, Tang Q, Tang M, Rainbow AJ, et al. Herpes simplex virus glycoproteins E and I facilitate cell-to-cell spread in vivo and across junctions of cultured cells. J Virol. 1994;68(2):834-45.

243. Johnson DC, Webb M, Wisner TW, Brunetti C. Herpes simplex virus gE/gI sorts nascent virions to epithelial cell junctions, promoting virus spread. J Virol. 2001;75(2):821-33.

244. Dingwell KS, Johnson DC. The herpes simplex virus gE-gI complex facilitates cell-to-cell spread and binds to components of cell junctions. J Virol. 1998;72(11):8933-42.

245. Howard PW, Howard TL, Johnson DC. Herpes simplex virus membrane proteins gE/gI and US9 act cooperatively to promote transport of capsids and glycoproteins from neuron cell bodies into initial axon segments. J Virol. 2013;87(1):403-14.

246. Snyder A, Polcicova K, Johnson DC. Herpes simplex virus gE/gI and US9 proteins promote transport of both capsids and virion glycoproteins in neuronal axons. J Virol. 2008;82(21):10613-24.

247. Kramer T, Greco TM, Taylor MP, Ambrosini AE, Cristea IM, Enquist LW. Kinesin-3 mediates axonal sorting and directional transport of alphaherpesvirus particles in neurons. Cell Host Microbe. 2012;12(6):806-14.

248. Kratchmarov R, Kramer T, Greco TM, Taylor MP, Ch'ng TH, Cristea IM, et al. Glycoproteins gE and gI are required for efficient KIF1A-dependent anterograde axonal transport of alphaherpesvirus particles in neurons. J Virol. 2013;87(17):9431-40.

249. Diefenbach RJ, Davis A, Miranda-Saksena M, Fernandez MA, Kelly BJ, Jones CA, et al. The Basic Domain of Herpes Simplex Virus 1 pUS9 Recruits Kinesin-1 To Facilitate Egress from Neurons. J Virol. 2015;90(4):2102-11.

250. Polcicova K, Goldsmith K, Rainish BL, Wisner TW, Johnson DC. The extracellular domain of herpes simplex virus gE is indispensable for efficient cell-to-cell spread: evidence for gE/gI receptors. J Virol. 2005;79(18):11990-2001.

251. Howard PW, Wright CC, Howard T, Johnson DC. Herpes simplex virus gE/gI extracellular domains promote axonal transport and spread from neurons to epithelial cells. J Virol. 2014;88(19):11178-86.

252. DuRaine G, Wisner TW, Howard P, Johnson DC. Herpes simplex virus gE/gI and US9 promote both envelopment and sorting of virus particles in the cytoplasm of neurons, two processes that precede anterograde transport in axons. J Virol. 2017.

253. Kelly BJ, Fraefel C, Cunningham AL, Diefenbach RJ. Functional roles of the tegument proteins of herpes simplex virus type 1. Virus Res. 2009;145(2):173-86.

254. Metrick CM, Heldwein EE. Novel Structure and Unexpected RNA-Binding Ability of the C-Terminal Domain of Herpes Simplex Virus 1 Tegument Protein UL21. J Virol. 2016;90(12):5759-69.

255. Metrick CM, Chadha P, Heldwein EE. The unusual fold of herpes simplex virus 1 UL21, a multifunctional tegument protein. J Virol. 2015;89(5):2979-84.

256. Mbong EF, Woodley L, Frost E, Baines JD, Duffy C. Deletion of UL21 causes a delay in the early stages of the herpes simplex virus 1 replication cycle. J Virol. 2012;86(12):7003-7.

257. Muto Y, Goshima F, Ushijima Y, Kimura H, Nishiyama Y. Generation and Characterization of UL21-Null Herpes Simplex Virus Type 1. Front Microbiol. 2012;3:394.

258. Le Sage V, Jung M, Alter JD, Wills EG, Johnston SM, Kawaguchi Y, et al. The herpes simplex virus 2 UL21 protein is essential for virus propagation. J Virol. 2013;87(10):5904-15.

259. Klupp BG, Lomniczi B, Visser N, Fuchs W, Mettenleiter TC. Mutations affecting the UL21 gene contribute to avirulence of pseudorabies virus vaccine strain Bartha. Virology. 1995;212(2):466-73.

260. Curanovic D, Lyman MG, Bou-Abboud C, Card JP, Enquist LW. Repair of the UL21 locus in pseudorabies virus Bartha enhances the kinetics of retrograde, transneuronal infection in vitro and in vivo. J Virol. 2009;83(3):1173-83.

261. Han J, Chadha P, Starkey JL, Wills JW. Function of glycoprotein E of herpes simplex virus requires coordinated assembly of three tegument proteins on its cytoplasmic tail. Proc Natl Acad Sci U S A. 2012;109(48):19798-803.

262. Takakuwa H, Goshima F, Koshizuka T, Murata T, Daikoku T, Nishiyama Y. Herpes simplex virus encodes a virion-associated protein which promotes long cellular processes in over-expressing cells. Genes Cells. 2001;6(11):955-66.

263. Michael K, Klupp BG, Karger A, Mettenleiter TC. Efficient incorporation of tegument proteins pUL46, pUL49, and pUS3 into pseudorabies virus particles depends on the presence of pUL21. J Virol. 2007;81(2):1048-51.

264. Harper AL, Meckes DG, Jr., Marsh JA, Ward MD, Yeh PC, Baird NL, et al. Interaction domains of the UL16 and UL21 tegument proteins of herpes simplex virus. J Virol. 2010;84(6):2963-71.

265. Klupp BG, Bottcher S, Granzow H, Kopp M, Mettenleiter TC. Complex formation between the UL16 and UL21 tegument proteins of pseudorabies virus. J Virol. 2005;79(3):1510-22.

266. Zhuo Y, Cano KE, Wang L, Ilangovan U, Hinck AP, Sousa R, et al. Nuclear Magnetic Resonance Structural Mapping Reveals Promiscuous Interactions between Clathrin-Box Motif Sequences and the N-Terminal Domain of the Clathrin Heavy Chain. Biochemistry. 2015;54(16):2571-80.

267. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, et al. ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res. 2012;40(Web Server issue):W597-603.

268. Towbin H, Staehelin T, Gordon J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc Natl Acad Sci U S A. 1979;76(9):4350-4.

269. Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. 2015.

270. Birnboim HC, Doly J. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 1979;7(6):1513-23.

271. Graham SC, Wartosch L, Gray SR, Scourfield EJ, Deane JE, Luzio JP, et al. Structural basis of Vps33A recruitment to the human HOPS complex by Vps16. Proc Natl Acad Sci U S A. 2013;110(33):13345-50.

272. Muenzner J, Traub LM, Kelly BT, Graham SC. Cellular and viral peptides bind multiple sites on the N-terminal domain of clathrin. Traffic. 2017;18(1):44-57.

273. Deng J, Davies DR, Wisedchaisri G, Wu M, Hol WG, Mehlin C. An improved protocol for rapid freezing of protein samples for long-term storage. Acta Crystallogr D Biol Crystallogr. 2004;60(Pt 1):203-4.

274. von Kleist L, Stahlschmidt W, Bulut H, Gromova K, Puchkov D, Robertson MJ, et al. Role of the clathrin terminal domain in regulating coated pit dynamics revealed by small molecule inhibition. Cell. 2011;146(3):471-84.

275. Winter G. xia2: an expert system for macromolecular crystallography data reduction. Journal of Applied Crystallography. 2009;43(1):186-90.

276. Kabsch W. Xds. Acta Crystallogr D Biol Crystallogr. 2010;66(Pt 2):125-32.

277. Evans P. Scaling and assessment of data quality. Acta Crystallogr D Biol Crystallogr. 2006;62(Pt 1):72-82.

278. Sauter NK, Grosse-Kunstleve RW, Adams PD. Robust indexing for automatic data collection. J Appl Crystallogr. 2004;37(Pt 3):399-409.

279. Zhang Z, Sauter NK, van den Bedem H, Snell G, Deacon AM. Automated diffraction image analysis and spot searching for high-throughput crystal screening. Journal of Applied Crystallography. 2006;39(1):112-9.

280. Evans PR. An introduction to data reduction: space-group determination, scaling and intensity statistics. Acta Crystallogr D Biol Crystallogr. 2011;67(Pt 4):282-92.

281. Collaborative Computational Project N. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr. 1994;50(Pt 5):760-3.

282. Waterman DG, Winter G, Gildea RJ, Parkhurst JM, Brewster AS, Sauter NK, et al. Diffraction-geometry refinement in the DIALS framework. Acta Crystallogr D Struct Biol. 2016;72(Pt 4):558-75.

283. Gildea RJ, Waterman DG, Parkhurst JM, Axford D, Sutton G, Stuart DI, et al. New methods for indexing multi-lattice diffraction data. Acta Crystallogr D Biol Crystallogr. 2014;70(Pt 10):2652-66.

284. Powell HR, Johnson O, Leslie AG. Autoindexing diffraction images with iMosflm. Acta Crystallogr D Biol Crystallogr. 2013;69(Pt 7):1195-203.

285. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr. 2011;67(Pt 4):355-67.

286. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr. 1997;53(Pt 3):240-55.

287. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr. 2007;40(Pt 4):658-74.

288. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010;66(Pt 4):486-501.

289. Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr. 2010;66(Pt 1):12-21.

290. Hooft RW, Vriend G, Sander C, Abola EE. Errors in protein structures. Nature. 1996;381(6580):272.

291. Celniker G, Nimrod G, Ashkenazy H, Glaser F, Martz E, Mayrose I, et al. ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function. Israel Journal of Chemistry. 2013;53(3-4):199-206.

292. Santoro MM, Bolen DW. Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. Biochemistry. 1988;27(21):8063-8.

293. Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 2005;33(Web Server issue):W244-8.

294. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7:539.

295. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25(9):1189-91.

296. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305(3):567-80.

297. Yang ZR, Thomson R, McNeil P, Esnouf RM. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. Bioinformatics. 2005;21(16):3369-76.

298. Buchan DW, Minneci F, Nugent TC, Bryson K, Jones DT. Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res. 2013;41(Web Server issue):W349-57.

299. Jones DT, Cozzetto D. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. Bioinformatics. 2015;31(6):857-63.

300. Petersen B, Petersen TN, Andersen P, Nielsen M, Lundegaard C. A generic method for assignment of reliability scores applied to solvent accessibility predictions. BMC Struct Biol. 2009;9:51.

301. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol. 1999;292(2):195-202.

302. Jones DT. Improving the accuracy of transmembrane protein topology prediction using evolutionary information. Bioinformatics. 2007;23(5):538-44.

303. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008;26(12):1367-72.

304. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res. 2011;10(4):1794-805.

305. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods. 2016;13(9):731-40.

306. Mellacheruvu D, Wright Z, Couzens AL, Lambert JP, St-Denis NA, Li T, et al. The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. Nat Methods. 2013;10(8):730-6.

307. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015;43(Database issue):D447-52.

308. Drake MT. Epsin Binds to Clathrin by Associating Directly with the Clathrin-terminal Domain. Evidence for cooperative binding through two discrete sites. Journal of Biological Chemistry. 2000;275(9):6479-89.

309. Lafer EM. Clathrin-protein interactions. Traffic. 2002;3(8):513-20.

310. Shih W, Gallusser A, Kirchhausen T. A clathrin-binding site in the hinge of the beta 2 chain of mammalian AP-2 complexes. J Biol Chem. 1995;270(52):31083-90.

311. Kim YM, Benovic JL. Differential roles of arrestin-2 interaction with clathrin and adaptor protein 2 in G protein-coupled receptor trafficking. J Biol Chem. 2002;277(34):30760-8.

312. Lo YC, Lin SC, Rospigliosi CC, Conze DB, Wu CJ, Ashwell JD, et al. Structural basis for recognition of diubiquitins by NEMO. Mol Cell. 2009;33(5):602-15.

313. Rossi AM, Taylor CW. Analysis of protein-ligand interactions by fluorescence polarization. Nat Protoc. 2011;6(3):365-87.

314. Zhou M, Li Q, Wang R. Current Experimental Methods for Characterizing Protein-Protein Interactions. ChemMedChem. 2016;11(8):738-56.

315. Simeonov A. Recent developments in the use of differential scanning fluorometry in protein and small molecule discovery and characterization. Expert Opin Drug Discov. 2013;8(9):1071-82.

316. Layton CJ, Hellinga HW. Quantitation of protein-protein interactions by thermal stability shift analysis. Protein Sci. 2011;20(8):1439-50.

317. Velazquez-Campoy A, Leavitt SA, Freire E. Characterization of protein-protein interactions by isothermal titration calorimetry. Methods Mol Biol. 2015;1278:183-204.

318. Leavitt S, Freire E. Direct measurement of protein binding energetics by isothermal titration calorimetry. Curr Opin Struct Biol. 2001;11(5):560-6.

319. Pierce MM, Raman CS, Nall BT. Isothermal titration calorimetry of protein-protein interactions. Methods. 1999;19(2):213-21.

320. Mills IG, Praefcke GJ, Vallis Y, Peter BJ, Olesen LE, Gallop JL, et al. EpsinR: an AP1/clathrin interacting protein involved in vesicle trafficking. J Cell Biol. 2003;160(2):213-22.

321. Willox AK, Sahraoui YM, Royle SJ. Non-specificity of Pitstop 2 in clathrin-mediated endocytosis. Biol Open. 2014;3(5):326-31.

322. Dutta D, Williamson CD, Cole NB, Donaldson JG. Pitstop 2 is a potent inhibitor of clathrin-independent endocytosis. PLoS One. 2012;7(9):e45799.

323. Agou F, Traincard F, Vinolo E, Courtois G, Yamaoka S, Israel A, et al. The trimerization domain of NEMO is composed of the interacting C-terminal CC2 and LZ coiled-coil subdomains. J Biol Chem. 2004;279(27):27861-9.

324. Huang GJ, Zhang ZQ, Jin DY. Stimulation of IKK-gamma oligomerization by the human T-cell leukemia virus oncoprotein Tax. FEBS Lett. 2002;531(3):494-8.

325. Tegethoff S, Behlke J, Scheidereit C. Tetrameric oligomerization of IkappaB kinase gamma (IKKgamma) is obligatory for IKK complex activity and NF-kappaB activation. Mol Cell Biol. 2003;23(6):2029-41.

326. Ivins FJ, Montgomery MG, Smith SJ, Morris-Davies AC, Taylor IA, Rittinger K. NEMO oligomerization and its ubiquitin-binding properties. Biochem J. 2009;421(2):243-51.

327. Zhuo Y, Ilangovan U, Schirf V, Demeler B, Sousa R, Hinck AP, et al. Dynamic interactions between clathrin and locally structured elements in a disordered protein mediate clathrin lattice assembly. J Mol Biol. 2010;404(2):274-90.

328. Guo H, Shen S, Wang L, Deng H. Role of tegument proteins in herpesvirus assembly and egress. Protein Cell. 2010;1(11):987-98.

329. Lyman MG, Curanovic D, Enquist LW. Targeting of pseudorabies virus structural proteins to axons requires association of the viral Us9 protein with lipid rafts. PLoS Pathog. 2008;4(5):e1000065.

330. Koshizuka T, Kawaguchi Y, Nozawa N, Mori I, Nishiyama Y. Herpes simplex virus protein UL11 but not UL51 is associated with lipid rafts. Virus Genes. 2007;35(3):571-5.

331. Lyman MG, Feierbach B, Curanovic D, Bisher M, Enquist LW. Pseudorabies virus Us9 directs axonal sorting of viral capsids. J Virol. 2007;81(20):11363-71.

332. Brideau AD, Card JP, Enquist LW. Role of pseudorabies virus Us9, a type II membrane protein, in infection of tissue culture cells and the rat nervous system. J Virol. 2000;74(2):834-45.

333. Brideau AD, Eldridge MG, Enquist LW. Directional transneuronal infection by pseudorabies virus is dependent on an acidic internalization motif in the Us9 cytoplasmic tail. J Virol. 2000;74(10):4549-61.

334. McGraw HM, Awasthi S, Wojcechowskyj JA, Friedman HM. Anterograde spread of herpes simplex virus type 1 requires glycoprotein E and glycoprotein I but not Us9. J Virol. 2009;83(17):8315-26.

335. Baines JD, Roizman B. The UL11 gene of herpes simplex virus 1 encodes a function that facilitates nucleocapsid envelopment and egress from cells. J Virol. 1992;66(8):5168-74.

336. Chouljenko DV, Kim IJ, Chouljenko VN, Subramanian R, Walker JD, Kousoulas KG. Functional hierarchy of herpes simplex virus 1 viral glycoproteins in cytoplasmic virion envelopment and egress. J Virol. 2012;86(8):4262-70.

337. Kopp M, Granzow H, Fuchs W, Klupp BG, Mundt E, Karger A, et al. The Pseudorabies Virus UL11 Protein Is a Virion Component Involved in Secondary Envelopment in the Cytoplasm. Journal of Virology. 2003;77(9):5339-51.

338. Cunningham C, Davison AJ, MacLean AR, Taus NS, Baines JD. Herpes simplex virus type 1 gene UL14: phenotype of a null mutant and identification of the encoded protein. J Virol. 2000;74(1):33-41.

339. Yeh PC, Han J, Chadha P, Meckes DG, Jr., Ward MD, Semmes OJ, et al. Direct and specific binding of the UL16 tegument protein of herpes simplex virus to the cytoplasmic tail of glycoprotein E. J Virol. 2011;85(18):9425-36.

340. Baines JD, Roizman B. The open reading frames UL3, UL4, UL10, and UL16 are dispensable for the replication of herpes simplex virus 1 in cell culture. J Virol. 1991;65(2):938-44.

341. Yeh PC, Meckes DG, Jr., Wills JW. Analysis of the interaction between the UL11 and UL16 tegument proteins of herpes simplex virus. J Virol. 2008;82(21):10693-700.

342. Starkey JL, Han J, Chadha P, Marsh JA, Wills JW. Elucidation of the block to herpes simplex virus egress in the absence of tegument protein UL16 reveals a novel interaction with VP22. J Virol. 2014;88(1):110-9.

343. Duffy C, Lavail JH, Tauscher AN, Wills EG, Blaho JA, Baines JD. Characterization of a UL49-null mutant: VP22 of herpes simplex virus type 1 facilitates viral spread in cultured cells and the mouse cornea. J Virol. 2006;80(17):8664-75.

344. Li B, Fields S. Identification of mutations in p53 that affect its binding to SV40 large T antigen by using the yeast two-hybrid system. FASEB J. 1993;7(10):957-63.

345. Serebriiskii IG, Golemis EA. Two-hybrid system and false positives. Approaches to detection and elimination. Methods Mol Biol. 2001;177:123-34.

346. Hebbar N, Wang C, Rangnekar VM. Mechanisms of apoptosis by the tumor suppressor Par-4. J Cell Physiol. 2012;227(12):3715-21.

347. D'Souza-Schorey C, Chavrier P. ARF proteins: roles in membrane traffic and beyond. Nat Rev Mol Cell Biol. 2006;7(5):347-58.

348. Nayak A, Muller S. SUMO-specific proteases/isopeptidases: SENPs and beyond. Genome Biol. 2014;15(7):422.

349. Munz C. Beclin-1 targeting for viral immune escape. Viruses. 2011;3(7):1166-78.

350. Lima CD, Reverter D. Structure of the human SENP7 catalytic domain and poly-SUMO deconjugation activities for SENP6 and SENP7. J Biol Chem. 2008;283(46):32045-55.

351. Wu F, Xing J, Wang S, Li M, Zheng C. Screening and identification of host factors interacting with UL14 of herpes simplex virus 1. Med Microbiol Immunol. 2011;200(3):203-8.

352. Vittone V, Diefenbach E, Triffett D, Douglas MW, Cunningham AL, Diefenbach RJ. Determination of interactions between tegument proteins of herpes simplex virus type 1. J Virol. 2005;79(15):9566-71.

353. Luo S, Rubinsztein DC. Apoptosis blocks Beclin 1-dependent autophagosome synthesis: an effect rescued by Bcl-xL. Cell Death Differ. 2010;17(2):268-77.

354. Zhu B, Mandal SS, Pham AD, Zheng Y, Erdjument-Bromage H, Batra SK, et al. The human PAF complex coordinates transcription with events downstream of RNA synthesis. Genes Dev. 2005;19(14):1668-73.

355. Hasumi H, Baba M, Hong SB, Hasumi Y, Huang Y, Yao M, et al. Identification and characterization of a novel folliculin-interacting protein FNIP2. Gene. 2008;415(1-2):60-7.

356. Bruckner A, Polge C, Lentze N, Auerbach D, Schlattner U. Yeast two-hybrid, a powerful tool for systems biology. Int J Mol Sci. 2009;10(6):2763-88.

357. Van Criekinge W, Beyaert R. Yeast Two-Hybrid: State of the Art. Biol Proced Online. 1999;2:1-38.

358. Chen X, Wei S, Ji Y, Guo X, Yang F. Quantitative proteomics using SILAC: Principles, applications, and developments. Proteomics. 2015;15(18):3175-92.

359. Emmott E, Goodfellow I. Identification of protein interaction partners in mammalian cells using SILAC-immunoprecipitation quantitative proteomics. J Vis Exp. 2014(89).

360. Cherepanova N, Shrimal S, Gilmore R. N-linked glycosylation and homeostasis of the endoplasmic reticulum. Current Opinion in Cell Biology. 2016;41:57-65.

361. Kim JJ, Lipatova Z, Segev N. TRAPP Complexes in Secretion and Autophagy. Front Cell Dev Biol. 2016;4:20.

362. Barrowman J, Bhandari D, Reinisch K, Ferro-Novick S. TRAPP complexes in membrane traffic: convergence through a common Rab. Nat Rev Mol Cell Biol. 2010;11(11):759-63.

363. Yamasaki A, Menon S, Yu S, Barrowman J, Meerloo T, Oorschot V, et al. mTrs130 is a component of a mammalian TRAPPII complex, a Rab1 GEF that binds to COPI-coated vesicles. Mol Biol Cell. 2009;20(19):4205-15.

364. Connolly DJ, O'Neill LA, McGettrick AF. The GOLD domain-containing protein TMED1 is involved in interleukin-33 signaling. J Biol Chem. 2013;288(8):5616-23.

365. Lu R, Stewart L, Wilson JM. Scaffolding protein GOPC regulates tight junction structure. Cell Tissue Res. 2015;360(2):321-32.

366. Amacher JF, Cushing PR, Bahl CD, Beck T, Madden DR. Stereochemical determinants of C-terminal specificity in PDZ peptide-binding domains: a novel contribution of the carboxylate-binding loop. J Biol Chem. 2013;288(7):5114-26.

367. Chen C, Matesic LE. The Nedd4-like family of E3 ubiquitin ligases and cancer. Cancer Metastasis Rev. 2007;26(3-4):587-604.

368. Flores SY, Debonneville C, Staub O. The role of Nedd4/Nedd4-like dependant ubiquitylation in epithelial transport processes. Pflugers Arch. 2003;446(3):334-8.

369. Charest A, Lane K, McMahon K, Housman DE. Association of a novel PDZ domain-containing peripheral Golgi protein with the Q-SNARE (Q-soluble N-ethylmaleimide-sensitive fusion protein (NSF) attachment protein receptor) protein syntaxin 6. J Biol Chem. 2001;276(31):29456-65.

370. Choi C, Davey M, Schluter C, Pandher P, Fang Y, Foster LJ, et al. Organization and assembly of the TRAPPII complex. Traffic. 2011;12(6):715-25.

371. Yip CK, Berscheminski J, Walz T. Molecular architecture of the TRAPPII complex and implications for vesicle tethering. Nat Struct Mol Biol. 2010;17(11):1298-304.

372. Javier RT, Rice AP. Emerging theme: cellular PDZ proteins as common targets of pathogenic viruses. J Virol. 2011;85(22):11544-56.

373. Engel EA, Song R, Koyuncu OO, Enquist LW. Investigating the biology of alpha herpesviruses with MS-based proteomics. Proteomics. 2015;15(12):1943-56.

374. Dudek J, Pfeffer S, Lee PH, Jung M, Cavalie A, Helms V, et al. Protein transport into the human endoplasmic reticulum. J Mol Biol. 2015;427(6 Pt A):1159-75.

375. Devaraneni PK, Conti B, Matsumura Y, Yang Z, Johnson AE, Skach WR. Stepwise insertion and inversion of a type II signal anchor sequence in the ribosome-Sec61 translocon complex. Cell. 2011;146(1):134-47.

376. Yu S, Satoh A, Pypaert M, Mullen K, Hay JC, Ferro-Novick S. mBet3p is required for homotypic COPII vesicle tethering in mammalian cells. J Cell Biol. 2006;174(3):359-68.

377. Kim YG, Raunser S, Munger C, Wagner J, Song YL, Cygler M, et al. The architecture of the multisubunit TRAPP I complex suggests a model for vesicle tethering. Cell. 2006;127(4):817-30.

378. Piserchio A, Fellows A, Madden DR, Mierke DF. Association of the cystic fibrosis transmembrane regulator with CAL: structural features and molecular dynamics. Biochemistry. 2005;44(49):16158-66.

379. Cheng J, Wang H, Guggino WB. Modulation of mature cystic fibrosis transmembrane regulator protein by the PDZ domain protein CAL. J Biol Chem. 2004;279(3):1892-8.

380. Cheng S, Zhang J, Zhu P, Ma Y, Xiong Y, Sun L, et al. The PDZ domain protein CAL interacts with mGluR5a and modulates receptor expression. J Neurochem. 2010;112(3):588-98.

381. Herrmann S, Ninkovic M, Kohl T, Pardo LA. PIST (GOPC) modulates the oncogenic voltage-gated potassium channel KV10.1. Front Physiol. 2013;4:201.

382. Amacher JF, Cushing PR, Brooks L, 3rd, Boisguerin P, Madden DR. Stereochemical preferences modulate affinity and selectivity among five PDZ domains that bind CFTR: comparative structural and sequence analyses. Structure. 2014;22(1):82-93.

383. Vouilleme L, Cushing PR, Volkmer R, Madden DR, Boisguerin P. Engineering peptide inhibitors to overcome PDZ binding promiscuity. Angew Chem Int Ed Engl. 2010;49(51):9912-6.

384. Hicks SW, Machamer CE. Isoform-specific interaction of golgin-160 with the Golgi-associated protein PIST. J Biol Chem. 2005;280(32):28944-51.

385. Pickart CM. Mechanisms underlying ubiquitination. Annu Rev Biochem. 2001;70:503-33.

386. Yu H, Matouschek A. Recognition of Client Proteins by the Proteasome. Annu Rev Biophys. 2017.

387. Zheng N, Shabek N. Ubiquitin Ligases: Structure, Function, and Regulation. Annu Rev Biochem. 2017.

388. Jeong KW, Kim HZ, Kim S, Kim YS, Choe J. Human papillomavirus type 16 E6 protein interacts with cystic fibrosis transmembrane regulator-associated ligand and promotes E6-associated protein-mediated ubiquitination and proteasomal degradation. Oncogene. 2007;26(4):487-99.

Appendix

Received: 28 June 2016 | Revised and accepted: 31 October 2016 | Uncorrected manuscript published: 3 November 2016 | Published on: 14 December 2016

DOI 10.1111/tra.12457

WILEY Traffic

ORIGINAL ARTICLE

Cellular and viral peptides bind multiple sites on the N-terminal domain of clathrin

Julia Muenzner[1] | Linton M. Traub[2] | Bernard T. Kelly[3] | Stephen C. Graham[1]

[1]Department of Pathology, University of Cambridge, Cambridge, UK

[2]Department of Cell Biology, University of Pittsburgh School of Medicine, Pittsburgh, PA

[3]Cambridge Institute for Medical Research, Department of Clinical Biochemistry, University of Cambridge, Cambridge, UK

Corresponding Authors: Bernard T. Kelly, Cambridge Institute for Medical Research, Department of Clinical Biochemistry, University of Cambridge, Cambridge, UK (btk1000@cam.ac.uk).
and
Stephen C. Graham, Department of Pathology, University of Cambridge, Cambridge, UK (scg34@cam.ac.uk).

Funding information
EU FP7 infrastructure grant BIOSTRUCT-X, Grant/Award number: 283570; Sir Henry Dale Fellowship; Royal Society; Wellcome Trust, Grant/Award number: 098406/Z/12/Z; NIH R01, Grant/Award number: GM106963; Wellcome grant, Grant/Award number: 090909/Z/09/Z; Wellcome Trust Strategic Award, Grant/Award number: 079895.

Short peptide motifs in unstructured regions of clathrin-adaptor proteins recruit clathrin to membranes to facilitate post-Golgi membrane transport. Three consensus clathrin-binding peptide sequences have been identified and structural studies show that each binds distinct sites on the clathrin heavy chain N-terminal domain (NTD). A fourth binding site for adaptors on NTD has been functionally identified but not structurally characterised. We have solved high resolution structures of NTD bound to peptide motifs from the cellular clathrin adaptors β2 adaptin and amphiphysin plus a putative viral clathrin adaptor, hepatitis D virus large antigen (HDAg-L). Surprisingly, with each peptide we observe simultaneous peptide binding at multiple sites on NTD and viral peptides binding to the same sites as cellular peptides. Peptides containing clathrin-box motifs (CBMs) with the consensus sequence LΦxΦ[DE] bind at the 'arrestin box' on NTD, between β-propeller blades 4 and 5, which had previously been thought to bind a distinct consensus sequence. Further, we structurally define the fourth peptide binding site on NTD, which we term the Royle box. In vitro binding assays show that clathrin is more readily captured by cellular CBMs than by HDAg-L, and site-directed mutagenesis confirms that multiple binding sites on NTD contribute to efficient capture by CBM peptides.

KEYWORDS

amphiphysin, arrestin, assembly polypeptide 2 (AP2), clathrin-mediated endocytosis, endocytosis, hepatitis D virus

1 | INTRODUCTION

Clathrin mediates vesicular transport between post-Golgi membranes in eukaryotes and is targeted to specific membranes by interactions with clathrin adaptor proteins.[1–3] Individually these interactions are weak,[4] but because clathrin polymerization drives the growth of a network of available binding sites, a wide range of adaptors and accessory factors may be recruited and retained at sites of coated pit formation.[2,5–7]

Clathrin:adaptor interactions are typically driven by linear peptide motifs in the unstructured regions of clathrin adaptors that bind the N-terminal β-propeller domain (NTD) of the clathrin heavy chain at several distinct sites (reviewed in reference 8): the "clathrin-box motif" (CBM), consensus sequence LΦxΦ[DE] (where x denotes any amino acid, Φ denotes a bulky hydrophobic residue and [DE] is a glutamate or aspartate), binds in a groove between blades 1 and 2 of the NTD β-propeller[3,9,10]; and the "W box" consensus PWxxW, binds the cleft near the centre of the NTD β-propeller.[11,12] Thirdly, an extended surface loop of the arrestin 2 long isoform (arrestin2L) has been shown to occupy the "arrestin box", a site lying between blades 4 and 5 of the NTD that binds peptides with consensus [LI][LI]GxL.[13] More recently, a fourth adaptor binding site on the clathrin NTD, between blades 6 and 7, was defined by Willox and Royle[14] on the basis of functional experiments in HeLa cells expressing clathrin heavy chain mutated in the NTD. This last study found that even a single functional NTD site was sufficient to sustain transferrin uptake.

The observation that any individual binding site on NTD is sufficient to sustain clathrin-mediated endocytosis of the transferrin receptor raises several questions. Does it reflect promiscuity in the binding of clathrin-interaction motifs, such that an individual

clathrin-binding motif can bind to different sites on NTD, or does it instead suggest intrinsic redundancy in the recruitment of adaptors such that endocytosis still proceeds even when an entire 'class' of clathrin-binding motif is prevented from binding clathrin? Previous studies suggest the latter, as each peptide binding site on clathrin characterized to date has a distinct consensus binding motif.[10,12,13] However, recent studies have suggested that the binding of peptides to clathrin may be promiscuous.[14,15] Promiscuity of CBM peptides for multiple sites on clathrin is relevant in the context of host:pathogen interactions, as it has previously been observed that viruses contain motifs resembling cellular CBMs that interact with clathrin in cells. These viral proteins have the ability to sequester clathrin, thus preventing endocytosis.[16,17] In the case of hepatitis D virus (HDV), which harbours a putative CBM sequence in the C-terminal region of the large antigen protein (HDAg-L), the presence of the clathrin binding motif seems essential for production of virus particles.[18] If viral CBM peptides bind only to the clathrin box on NTD it would be possible to blockade this site using a small molecule inhibitor.[19] This blockade would prevent virus hijacking of clathrin without perturbing cellular endocytosis, which can proceed when binding to the NTD "clathrin box" site is disrupted.[14] However, if viral CBMs bind multiple sites on NTD with comparable affinities then small molecule interventions are unlikely to succeed. We thus sought to investigate the relative affinity of cellular vs viral peptides for clathrin NTD and to compare their modes of binding. Further, we sought to investigate the potential degeneracy of clathrin binding that had been suggested by previous studies.[14,15]

Here we present high-resolution structures of clathrin NTD bound to cellular and viral peptide motifs. Surprisingly, in all cases we observe peptide binding at multiple sites on NTD. We use clathrin-binding assays and site-directed mutagenesis to qualitatively assess the binding of these peptide motifs to the different sites on NTD. Further, we provide the first structural characterization of the putative "fourth" adaptor binding site on clathrin NTD.

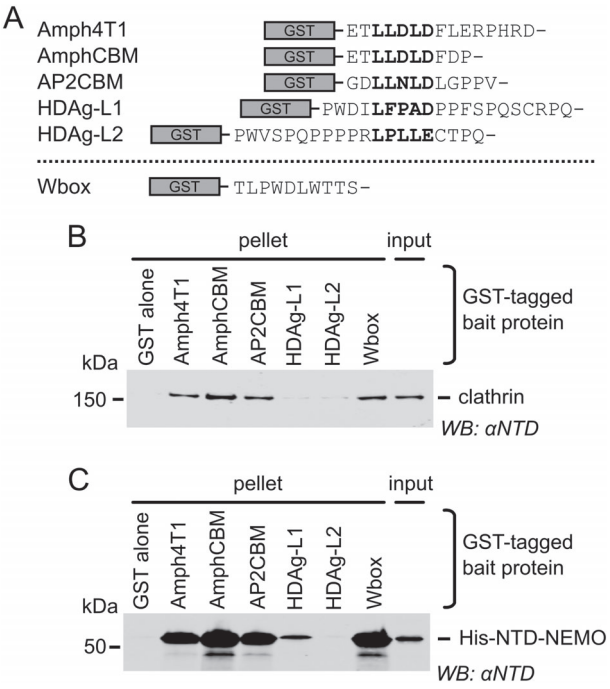


FIGURE 1 Cellular and viral peptide motifs bind clathrin N-terminal domain (NTD) to different degrees. A, Glutathione S-transferase (GST) fusions of clathrin-binding peptides used in this study. Clathrin-box motifs (CBMs) are aligned in bold. Amph4T1, human amphiphysin I CBM plus additional residues derived from the expression vector[22]; AmphCBM, human amphiphysin I CBM; AP2CBM, CBM from flexible hinge of β2 adaptin subunit of human AP2; HDAg-L1, putative CBM from clade I hepatitis D virus large antigen; HDAg-L2, putative CBM from clade II hepatitis D virus large antigen; Wbox, human amphiphysin W box binding motif. B, Capture ("GST pull-down") of purified clathrin by GST-tagged clathrin-binding peptides. Clathrin (input) was incubated with glutathione sepharose pre-loaded with GST-tagged "bait" proteins. After washing, proteins bound to the beads (pellet) were subjected to SDS-PAGE and immunoblotting (WB) using an antibody that recognizes clathrin NTD (αNTD). C, Capture of His-NTD-NEMO by GST-tagged clathrin binding peptides. Purified recombinant His-NTD-NEMO was used in GST pull-down experiments as in (B).

2 | RESULTS

2.1 | Cellular clathrin-binding motifs recruit clathrin more efficiently than those from hepatitis D virus

The CBMs from the cellular proteins β2 adaptin (AP2CBM) and amphiphysin (AmphCBM),[9,20] the W box motif of amphiphysin (Wbox),[11,12,20,21] and the C-terminal extensions of HDAg-L from 2 different HDV genotypes containing putative clathrin binding motifs (HDAg-L1 and HDAg-L2, respectively)[16,18] were fused to glutathione S-transferase (GST) (Figure 1A) and immobilized on glutathione resin for use in "GST pull-down" experiments to capture clathrin purified from pig brain. In addition, to aid comparison with previous biochemical studies,[11] an extended amphiphysin CBM construct (termed Amph4T1) was used in which the clathrin-binding motif is followed by the amino acids "LERPHRD" arising from the XhoI cloning site and subsequent vector-derived nucleotides.[22] Consistent with previous studies,[11,12] clathrin was efficiently captured by GST fused to AP2CBM, AmphCBM, Amph4T1 or the amphiphysin Wbox, while it was not significantly captured by GST alone (Figure 1B). Interestingly, clathrin was not efficiently captured by GST fused to either of the HDAg-L C-terminal extensions tested (Figure 1B).

To facilitate subsequent mutational work the N-terminal domain (NTD) of bovine clathrin heavy chain (100% amino acid identity to human clathrin heavy chain NTD), fused to a His_6 affinity tag at the N terminus and to the dimerization domain of NF-κB essential modulator (NEMO)[23] at the C terminus, was purified following expression in *Escherichia coli*. This His-NTD-NEMO protein binds GST-AP2CBM much more efficiently than His-NTD lacking the NEMO oligomerization domain (Figure S1A). We ascribe this to increased avidity of binding, as His-NTD-NEMO is capable of oligomerizing at higher concentrations whereas His-NTD is monomeric (Figure S1). As observed using purified clathrin, recombinant His-NTD-NEMO is efficiently pulled-down by GST-AP2CBM, GST-AmphCBM, GST-Amph4T1 and GST-Wbox (Figure 1C). His-NTD-NEMO is weakly pulled down by GST-HDAg-L1, whereas GST-HDAg-L2 does not pull down His-NTD-NEMO any more efficiently than does GST alone.

2.2 | CBMs of cellular and viral peptides bind multiple sites on the clathrin NTD

To gain structural insight into the binding of cellular and viral peptides, recombinant NTD was crystallized in the presence of peptides corresponding to CBMs of β2 adaptin and amphiphysin, the CBM region of the non-natural Amph4T1 sequence, and the putative CBMs of HDAg-L1 and HDAg-L2. These co-crystallization experiments were performed using high concentrations (3.4-3.6 mM) of clathrin-binding peptide to ensure saturation of the peptide binding sites. The structures of all co-crystals were solved and refined to high resolution (Table 1) and, surprisingly, in all cases electron density consistent with the presence of peptide bound to NTD could be observed at more than 1 locus on NTD (Figure 2).

In all the structures presented a peptide could be observed at the "clathrin box", lying between blades 1 and 2 of the NTD β-propeller fold ("clathrin box", Figure 2). For the cellular peptides ($AP2CBM_{pep}$, $AmphCBM_{pep}$ and $Amph4T1_{pep}$) the binding is similar to that previously described,[10] with the 3 leucine side chains of the CBM LΦxΦ[DE] consensus sequence occupying the hydrophobic pocket formed at the groove between the 2 blades. The viral peptides ($HDAg\text{-}L1_{pep}$ and $HDAg\text{-}L2_{pep}$) bound at a similar site on NTD. However, in both cases only 2 consecutive hydrophobic side chains could be observed in the hydrophobic pocket ("IL" and "LL" in the cases of $HDAg\text{-}L1_{pep}$ and $HDAg\text{-}L2_{pep}$, respectively). Interestingly, for both $HDAg\text{-}L1_{pep}$ and $HDAg\text{-}L2_{pep}$ the residues bound at the clathrin box do not match predictions based on alignments to the CBM consensus sequence (Figure 1)[16,18]: in $HDAg\text{-}L1_{pep}$ residues "ILFPA" occupy the positions corresponding to the LΦxΦ[DE], whereas in $HDAg\text{-}L2_{pep}$ the residues "LLES", including a C-terminal serine residue that is non-native and was added to the peptide to aid solubility, occupy the positions equivalent to the first 4 residues of the LΦxΦ[DE] consensus.

In addition to binding at the clathrin box, we observed significant binding of the cellular peptides ($AP2CBM_{pep}$, $AmphCBM_{pep}$ and $Amph4T1_{pep}$) at the "arrestin box", which lies between blades 4 and 5 of the NTD β-propeller fold ("arrestin box", Figure 2). While all 3 CBM peptides bind in the same general orientation at the arrestin box (Figure 3), the molecular details of these interactions differ from the interaction seen between NTD and the extended surface loop of the arrestin 2 long isoform (arrestin2L).[13] Most notably, the directionality of the peptide chain is reversed. The first 2 leucine residues of each CBM (**LΦ**xΦ[DE]) bind in a hydrophobic cavity lined by the side chains of NTD residues W164, L183, S185, R188, V190, I194, F216, I231 and V233 plus the peptide backbones of Y184 and S191 (Figure 3B). The position occupied by these 2 peptide leucine side chains is very similar to that occupied by arrestin2L residues L338 and L335 (residues 5 and 2 of the [LI][LI]GxL arrestin box consensus motif, respectively) in the complex with NTD, despite the fact that in the arrestin2L:NTD complex the peptide backbone adopts a vastly different conformation to accommodate the 2 intervening amino acid residues (Figure 3C). In the structures presented here the side chain oxygen atom of NTD residue Q192 forms hydrogen bonds with the backbone amide protons of these 2 leucine residues (**LΦ**xΦ[DE]) and the side chain nitrogen atom of Q192 forms a hydrogen bond with the carbonyl oxygen of the second leucine (L**Φ**xΦ[DE]). While the electron density was not sufficiently well-resolved to allow unambiguous modelling in the case of $Amph4T1_{pep}$, for both the $AP2CBM_{pep}$ and $AmphCBM_{pep}$ structures the leucine side chain of the third CBM motif residue (LΦx**Φ**[DE]) occupies a similar location to the side chain of arrestin2L residue L334 ([**LI**][LI]GxL), binding at a hydrophobic surface patch formed by NTD residues I194, F218, H229 and I231. In the $AmphCBM_{pep}$ structure the side chain of the phenylalanine residue that follows the CBM forms an additional hydrophobic interaction with NTD, binding in a hydrophobic cleft formed by side chains of residues H229, I231 and A247, the aryl side chain region of K245, and the peptide backbones of residues 245-247. The $HDAg\text{-}L2_{pep}$ peptide could also be observed binding at the arrestin site, although the interaction was less extensive (Figure 2). The binding was largely similar to that observed for the cellular CBM peptides: the side chains of the 2 consecutive leucine residues bound at the hydrophobic cleft and their backbone atoms interacted with the side chain of Q192 as described above. There was not strong evidence for the $HDAg\text{-}L1_{pep}$ peptide bound at the arrestin box (Figure 2).

2.3 | Structural identification of the "fourth" peptide binding site on the clathrin NTD

In addition to binding at the clathrin and arrestin boxes, in 3 of the NTD:peptide co-crystal structures solved ($Amph4T1_{pep}$, $HDAg\text{-}L1_{pep}$ and $HDAg\text{-}L2_{pep}$) a bound peptide could be observed lying across the interface between blades 6 and 7 of the NTD β-propeller (Figure 2). This peptide binding site overlaps with the region identified in the functional studies of Willox and Royle[14] as the fourth and final clathrin adaptor binding site on NTD, and we thus henceforth refer to it as the "Royle box" In comparison to surrounding residues, peptides were generally less well-ordered when bound at this site than when bound at the clathrin box (Table S1). However, in all cases a single orientation of the peptide could be modelled with good stereochemistry and an acceptable fit to electron density (Figure 4A).

The surface residues bound by peptides at the Royle box are highly conserved amongst eukaryotic clathrin sequences (Figure 4B). Binding at the Royle box (Figure 4C) centres on a deep hydrophobic pocket lying at the interface of blades 6 and 7, formed by the side chains of NTD residues L5, I7, F9, I282, N296 and V327. In the co-structure with $Amph4T1_{pep}$ a phenylalanine side chain projects deep into this pocket while in the $HDAg\text{-}L1_{pep}$ and $HDAg\text{-}L2_{pep}$ structures the side chains of a leucine residue and proline residue, respectively, bind less deeply. In all structures a hydrophobic side chain (leucine in $Amph4T1_{pep}$ and $HDAg\text{-}L2_{pep}$, isoleucine in $HDAg\text{-}L1_{pep}$) covers a surface hydrophobic patch formed by the hydrophobic portion of the R297 side chain and the hydrophobic faces of the peptide bonds between NTD residues 298-300 on the surface of β-propeller blade 6, and in all structures the backbone carbonyl oxygen of R297 forms a hydrogen bond with a backbone amide nitrogen of the bound peptide. In each structure 3 consecutive amino acid residues wrap around the side chain of F9 on the surface of blade 7, forming

TABLE 1 Crystallographic data collection and refinement. Values for the highest resolution shell are shown in parentheses

NTD:	AP2CBM$_{pep}$	AmphCBM$_{pep}$	Amph4T1$_{pep}$	HDAg-L1$_{pep}$	HDAg-L2$_{pep}$
Data collection					
Space group	$C222_1$	$C2$	$C2$	$C2$	$C2$
Cell dimensions					
a, *b*, *c* (Å)	108.1, 133.2, 77.9	140.0, 134.1, 78.0	137.8, 131.0, 79.1	136.2, 131.2, 77.9	136.9, 131.2, 78.5
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 115.1, 90.0	90.0, 116.2, 90.0	90.0, 115.6, 90.0	90.0, 115.9, 90.0
Resolution (Å)	57.1–1.8 (1.81–1.76)	67.1–1.9 (1.93–1.88)	33.6–1.7 (1.74–1.70)	48.4–2.2 (2.21–2.15)	39.2–2.0 (2.00–1.96)
R_{merge}	0.053 (1.538)	0.149 (1.092)	0.055 (0.751)	0.101 (0.930)	0.081 (0.581)
< I/σI>	15.8 (1.2)	6.5 (1.3)	13.1 (1.6)	9.6 (1.5)	7.2 (1.5)
$CC_{1/2}$	1.000 (0.672)	0.993 (0.568)	0.999 (0.564)	0.996 (0.501)	0.996 (0.563)
Completeness (%)	100.0 (100.0)	100.0 (100.0)	98.2 (94.5)	99.7 (99.0)	94.6 (95.9)
Redundancy	7.5 (7.0)	5.1 (4.4)	3.9 (3.4)	4.5 (4.3)	2.5 (2.3)
Refinement					
Resolution (Å)	57.1–1.8 (1.81–1.76)	67.1–1.9 (1.93–1.88)	33.6–1.7 (1.74–1.70)	48.4–2.2 (2.21–2.15)	39.2–2.0 (2.01–1.96)
No. of reflections (work/free)	52,951/2740	99,868/5291	128,743/6564	63,339/3381	79,976/4224
R_{work}/R_{free}	0.176/0.205	0.204/0.234	0.158/0.185	0.175/0.207	0.169/0.193
Ramachandran favoured regions (%)	98.7	98.6	98.8	98.2	99.1
Ramachandran outliers (%)	0.0	0.0	0.0	0.0	0.0
No. of atoms					
Protein	2836	5845	5796	5634	5634
Glycerol	—	18	6	6	6
Peptide ligands	104	268	340	200	240
Water	403	803	1019	437	583
B-factors					
Protein	36.1	22.0	26.2	39.8	32.0
Glycerol	—	36.5	21.8	32.7	20.9
Peptide ligands	52.4	45.1	48.7	70.2	61.0
Water	53.5	37.8	42.3	45.6	38.2
r.m.s. deviations					
Bond lengths (Å)	0.016	0.010	0.014	0.020	0.019
Bond angles (°)	1.719	1.396	1.619	1.897	1.917
PDB ID	5M5R	5M5S	5M5T	5M5U	5M5V

Abbreviations: NTD, N-terminal domain; PDB, Protein data bank.

hydrophobic interactions with both faces of the phenylalanine residue's hydrophobic side chain benzyl group. In addition, in the co-structures with HDAg-L1$_{pep}$ and Amph4T1$_{pep}$ the backbone carbonyl oxygen of F9 forms a hydrogen bond with a backbone amide nitrogen of the bound peptide, and backbone atom(s) of E11 form hydrogen bond(s) with the bound peptide (1 bond in the case of HDAg-L1$_{pep}$, 2 in the case of Amph4T1$_{pep}$). Despite the similar molecular interactions made between the Royle box and the bound peptides, we note that the direction of the peptide chain differs between HDAg-L1$_{pep}$ and Amph4T1$_{pep}$/HDAg-L2$_{pep}$ (Figure 4C). In addition, unlike the binding of peptides to the clathrin and arrestin boxes, where the same side chains of the CBM sequence form key interactions, we note that residues outside the CBM consensus sequence also form extensive interactions at the Royle box.

To investigate whether the absence of AmphCBM$_{pep}$ binding at the Royle box arose from an absence of stabilizing residues C-terminal to the CBM motif, we solved the structure of NTD in complex with a longer sequence containing the human amphiphysin I CBM (AmphCBMlong$_{pep}$, sequence ETLLDLDFDPFK; Table S2). As observed for AmphCBM$_{pep}$, AmphCBMlong$_{pep}$ bound NTD at the CBM and arrestin boxes but not at the Royle box (Figure S2).

2.4 | Multiple interaction sites on clathrin NTD contribute to peptide binding in vitro

A series of His-NTD-NEMO constructs with mutations at each of the 4 peptide binding sites on NTD were generated to probe whether all the interactions observed in the crystal structures contribute to binding in a biochemical context, or whether binding can be explained by a single dominant binding interaction. The mutations introduced at each site were informed by the crystal structures presented above and by previous studies (Figure 5A and Table 2). To confirm that these mutations did not introduce defects in NTD folding, the secondary structure and thermal stability of these mutants was probed by circular dichroism (CD) spectroscopy and differential scanning fluorimetry (DSF, a.k.a. Thermofluor), respectively. The CD spectra of

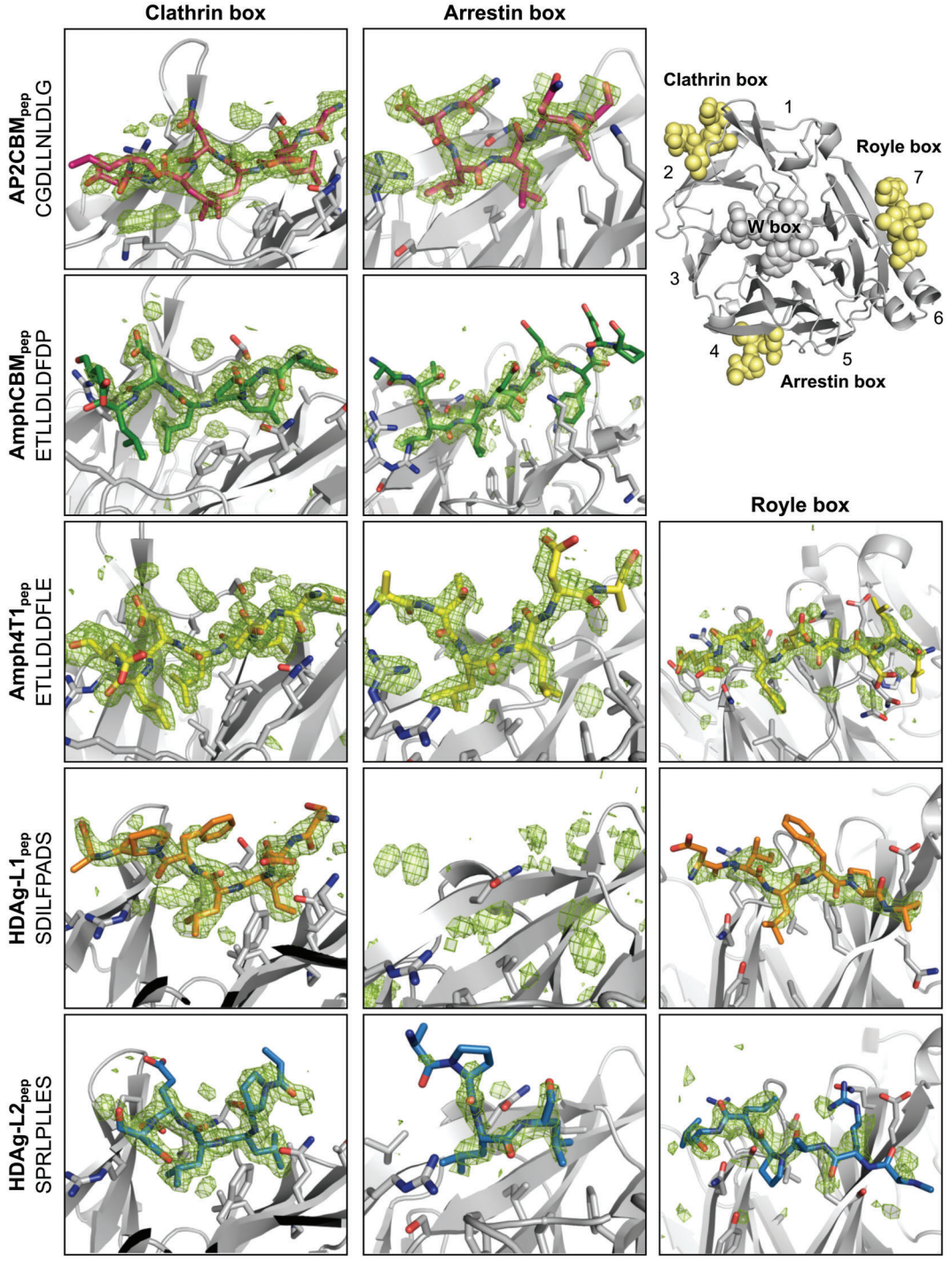


FIGURE 2 The clathrin-box motifs (CBMs) of cellular and viral proteins bind multiple sites on clathrin N-terminal domain (NTD). Unboxed image shows the β-propeller fold of clathrin NTD (grey ribbons) with numbers enumerating the 7 β-stranded blades. Spheres represent peptides bound at the 4 peptide-interaction sites on NTD. Boxed images show CBM-containing peptides (sticks, carbon atoms coloured as follows: AP2CBM$_{pep}$, magenta; AmphCBM$_{pep}$, dark green; Amph4T1$_{pep}$, yellow; HDAg-L1$_{pep}$, orange; HDAg-L2$_{pep}$, light blue) bound at the clathrin box, arrestin box and Royle box sites on clathrin NTD (grey ribbons). Unbiased F_O-F_C electron density (3 σ) used to place peptides into the structures is shown as a green mesh. Selected side chain atoms of clathrin NTD are shown (sticks with grey carbon atoms).

mutants were very similar to that of wild-type NTD (Figure S3), confirming that they had the correct secondary structure composition. However, DSF showed a number of mutants to have melting temperatures significantly lower than the wild-type protein, consistent with disrupted folding, and these mutants were thus not considered further (Figure 5B).

A selection of the correctly folded mutants was tested for ability to bind the cellular clathrin-binding motifs in GST pull-down experiments. Given the modest ability of GST-HDAg-L1 and -L2 to capture His-NTD-NEMO (Figure 1C) we limited our analysis to the cellular peptide sequences (GST-AP2CBM, GST-AmphCBM, GST-Amph4T1 and GST-Wbox). Figure 5C shows that mutations at the W box

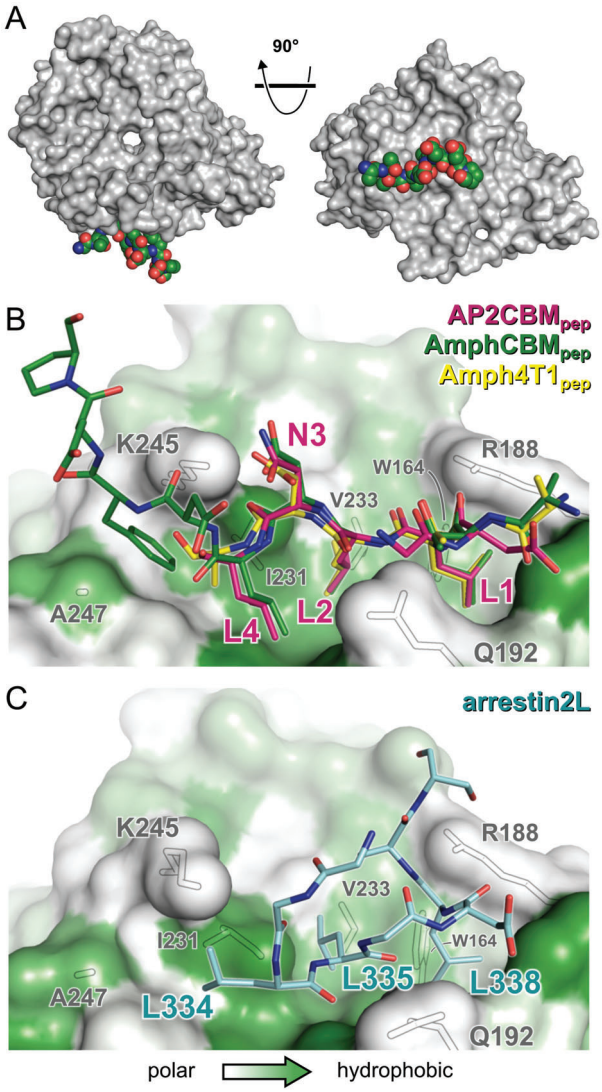


FIGURE 3 Cellular clathrin-box motifs (CBMs) bind in a different conformation than arrestin2L at the arrestin box. A, The surface of clathrin N-terminal domain (NTD) is shown (grey) oriented as in Figure 2 (left) and rotated by 90° around the horizontal axis (right). The $AmphCBM_{pep}$ peptide bound at the arrestin box is shown as coloured spheres. B, Close-up view of cellular CBM-containing peptides bound at the arrestin box. Peptides are shown as sticks coloured as in Figure 2. The surface of clathrin NTD is shown, coloured from high (green) to low (white) surface residue hydrophobicity, with outlines of selected surface side chains shown in grey. Bound $AP2CBM_{pep}$ residues are numbered by their position in the LΦxΦ[DE] CBM consensus sequence. C, The extended surface loop of arrestin 2 long isoform (arrestin2L) bound at the arrestin box (PDB 3GD1).[13] NTD is shown as in (B) and arrestin2L residues 332-340 are shown as sticks with cyan carbon atoms. Note that in (B) the direction of the bound peptides is right (N terminus) to left (C terminus), whereas in (C) the peptide chain between residues L334-L338 runs in the opposite direction.

(I154Q and I152L + I154Q) severely disrupt the ability of GST-Wbox to capture His-NTD-NEMO, consistent with previous studies.[12] His-NTD-NEMO with mutations at the clathrin box was less efficiently captured by GST-AP2CBM, the defect being most pronounced for the Q89A + F91K mutant, but capture of these mutants by GST-AmphCBM or GST-Amph4T1 was largely unperturbed (Figure 5C). Similarly, His-NTD-NEMO mutated at the arrestin box (Q192Y) was captured less efficiently by GST-AP2CBM but the capture of this mutant by GST-AmphCBM or GST-Amph4T1 was not significantly changed (Figure 5C).

To test whether mutation at more than 1 peptide binding site further reduced capture by cellular CBM peptides, His-NTD-NEMO constructs with mutations at multiple binding loci were generated. All these 'compound mutants' had CD spectra similar to that of the wild-type protein (Figure S3) and melting temperatures within 3 K of the wild-type protein (Figure 5B), consistent with the compound mutants being well-folded. As shown in Figure 5D, while His-NTD-NEMO mutated at either the clathrin box (Q89A + F91K) or arrestin box (Q192Y) can still be captured by GST-AP2CBM, combining the mutations (Q89A + F91K + Q192Y) reduces the binding to the level of the GST control. Similarly, His-NTD-NEMO mutated at both the clathrin and arrestin boxes is captured less efficiently by GST-Amph4T1 than is the wild-type protein or protein with mutants at either site individually. The decrease in capture of His-NTD-NEMO with mutated clathrin and arrestin boxes by GST-Amph4T1 is more pronounced than is the capture of clathrin and Royle box or arrestin plus Royle box mutants. However, when NTD is mutated at all 3 sites, namely the arrestin, clathrin and Royle boxes, the extent of capture by GST-Amph4T1 is further decreased and approaches the levels seen for GST alone. GST-AmphCBM captures His-NTD-NEMO mutated at both the clathrin and arrestin boxes less efficiently than it does wild-type protein or His-NTD-NEMO with either site mutated individually. However, none of the His-NTD-NEMO mutations tested completely abolished binding to GST-AmphCBM.

2.5 | The arrestin motif of AP2 can also bind multiple sites on clathrin NTD

The hinge region of the assembly polypeptide 2 (AP2) complex β2 adaptin subunit contains 2 overlapping clathrin-binding motifs, a CBM and an arrestin-box motif (Figure 6A). However, the CBM motif can bind at the arrestin box in addition to binding at the clathrin box (Figure 2) and both such interactions contribute to NTD recruitment (Figure 5D). We therefore sought to compare the NTD binding of the β2 adaptin arrestin-box motif to that of the β2 adaptin CBM.

Two GST-tagged peptide constructs containing the arrestin-box motif of β2 adaptin (GST-AP2arrL and GST-AP2arrS) were generated (Figure 6A). Both contained the β2 adaptin arrestin-box motif (LLGDL) but, to avoid the potentially confounding issue of a carboxylate group immediately following the final residue of the motif, the sequences of their C-terminal residues differed: AP2arrS had the subsequent "L" residue of β2 adaptin appended to the arrestin-box motif, while AP2arrL had the residues "ASS" appended, corresponding to the residues that follow the LLGDL arrestin-box motif of arrestin2L.[13] We compared the ability of GST-AP2CBM, GST-AP2arrS and GST-AP2arrL to capture either wild-type His-NTD-NEMO or a mutant (Q98A + F91K + F9W) where the clathrin and Royle boxes, but not the arrestin box, had been disrupted. Figure 6B shows that GST-AP2CBM and GST-AP2arrS capture wild-type and Q98A + F91K + F9W His-NTD-NEMO more efficiently than does AP2arrL, suggesting that the arrestin-box motif (LLGDL) alone binds the arrestin box more

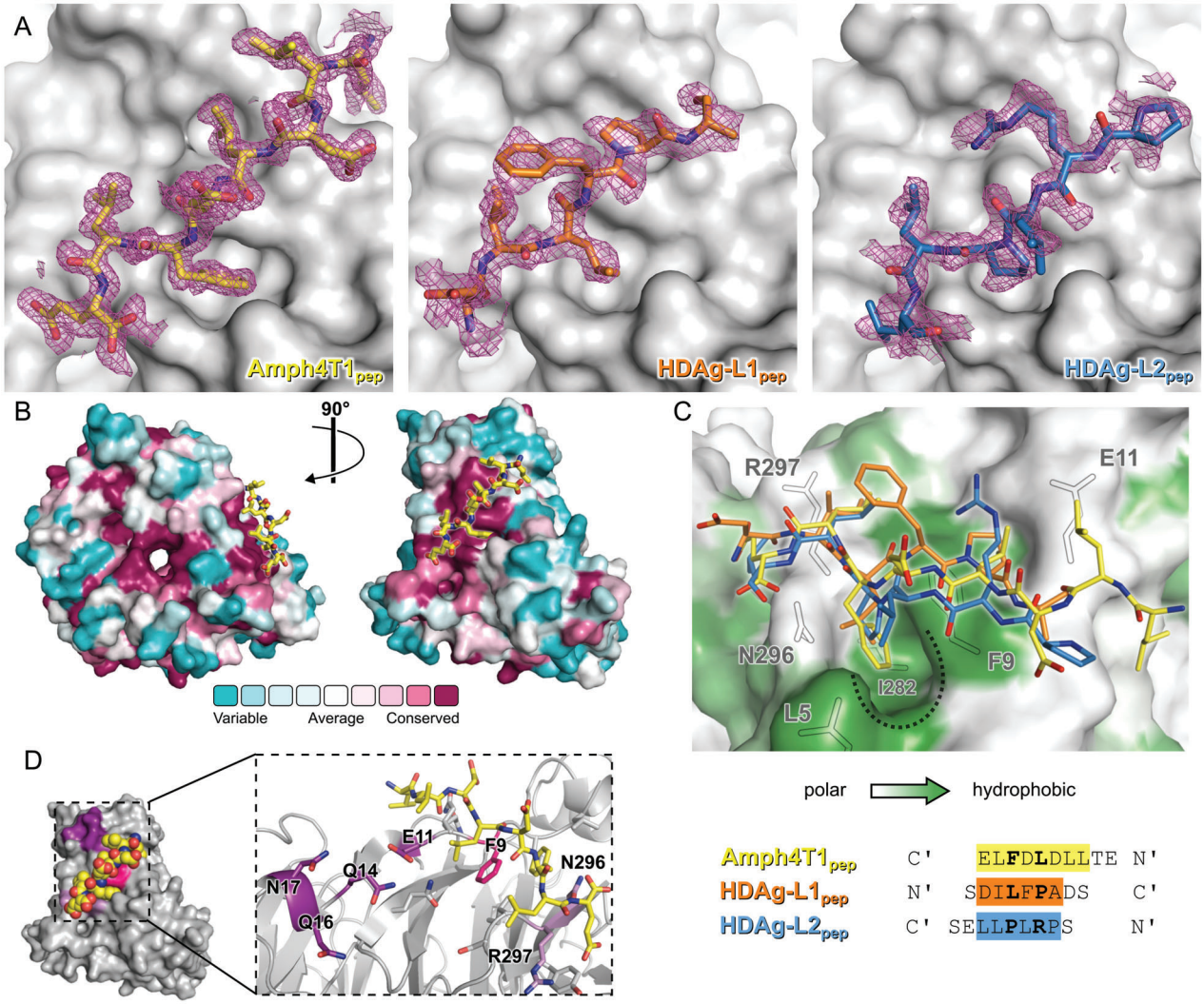


FIGURE 4 Localization and characterization of the fourth peptide binding site on clathrin N-terminal domain (NTD): The Royle box. A, Amph4T1$_{pep}$ (left), HDAg-L1$_{pep}$ (middle) and HDAg-L2$_{pep}$ (right) peptides bound at the Royle box in feature-enhanced maps[24] calculated using the final refined model (2σ, magenta). For clarity, maps are shown only within 2 Å of the bound peptides. Peptides are shown as sticks, coloured as in Figure 2, and clathrin NTD is shown as a grey surface. B, The surface of clathrin NTD, coloured by amino acid conservation from conserved (magenta) to variable (blue), is shown oriented as in Figure 2 (left) and rotated by 90° around the vertical axis (right). The Amph4T1$_{pep}$ peptide bound at a conserved surface patch between NTD β-propeller blades 6 and 7 (which we term the Royle box) is shown as sticks with yellow carbon atoms. C, Close-up view of cellular and virus peptides bound at the Royle box. Peptides are shown as sticks coloured as in Figure 2. The surface of clathrin NTD is shown, coloured from high (green) to low (white) surface residue hydrophobicity, with outlines of selected surface side chains shown in grey. A hydrophobic NTD surface pocket that is occupied by hydrophobic residues of all three peptides is marked by a dotted line. The peptide sequences used for co-crystallization are structurally aligned at the bottom of the panel. The directionality of the bound peptides is indicated. Residues that could be confidently modelled in the structures are highlighted and residues that form side chain interactions with NTD surface pockets are printed in bold type. D, Cellular and viral peptides bind near NTD residues functionally implicated in clathrin-mediated endocytosis. The surface of NTD (grey) is oriented as in the right image of (A) with residues mutated by Willox and Royle[14] (light and dark purple) or in this study (pink) highlighted. The Amph4T1$_{pep}$ peptide is shown as spheres. Inset shows the Amph4T1$_{pep}$ peptide (sticks with yellow carbon atoms) bound to NTD (grey, ribbon with selected side chains shown as sticks). The carbon atoms of residues substituted in clathrin mutants that disrupt transferrin uptake,[14] a proxy for clathrin-mediated endocytosis, are dark purple while those of residues where substitution does not affect transferrin uptake are light purple. The side chain of F9, mutated in this study, is coloured bright pink.

weakly than does the CBM or an extended arrestin motif (LLGDLL) containing the first 2 residues of the overlapping CBM. However, these experiments also show that GST-AP2arrL and GST-AP2arrS capture wild-type His-NTD-NEMO much more efficiently than they do the Q98A + F91K + F9W mutant, consistent with the β2 adaptin arrestin-box motif binding to either the clathrin or Royle boxes in addition to the arrestin box.

3 | DISCUSSION

This study presents the structure of the clathrin heavy chain NTD solved in the presence of cellular and viral clathrin-binding peptides. In all cases, we observe that these peptides bind promiscuously to more than one site on the clathrin NTD surface. This differs from previous high-resolution structural characterizations of peptide binding

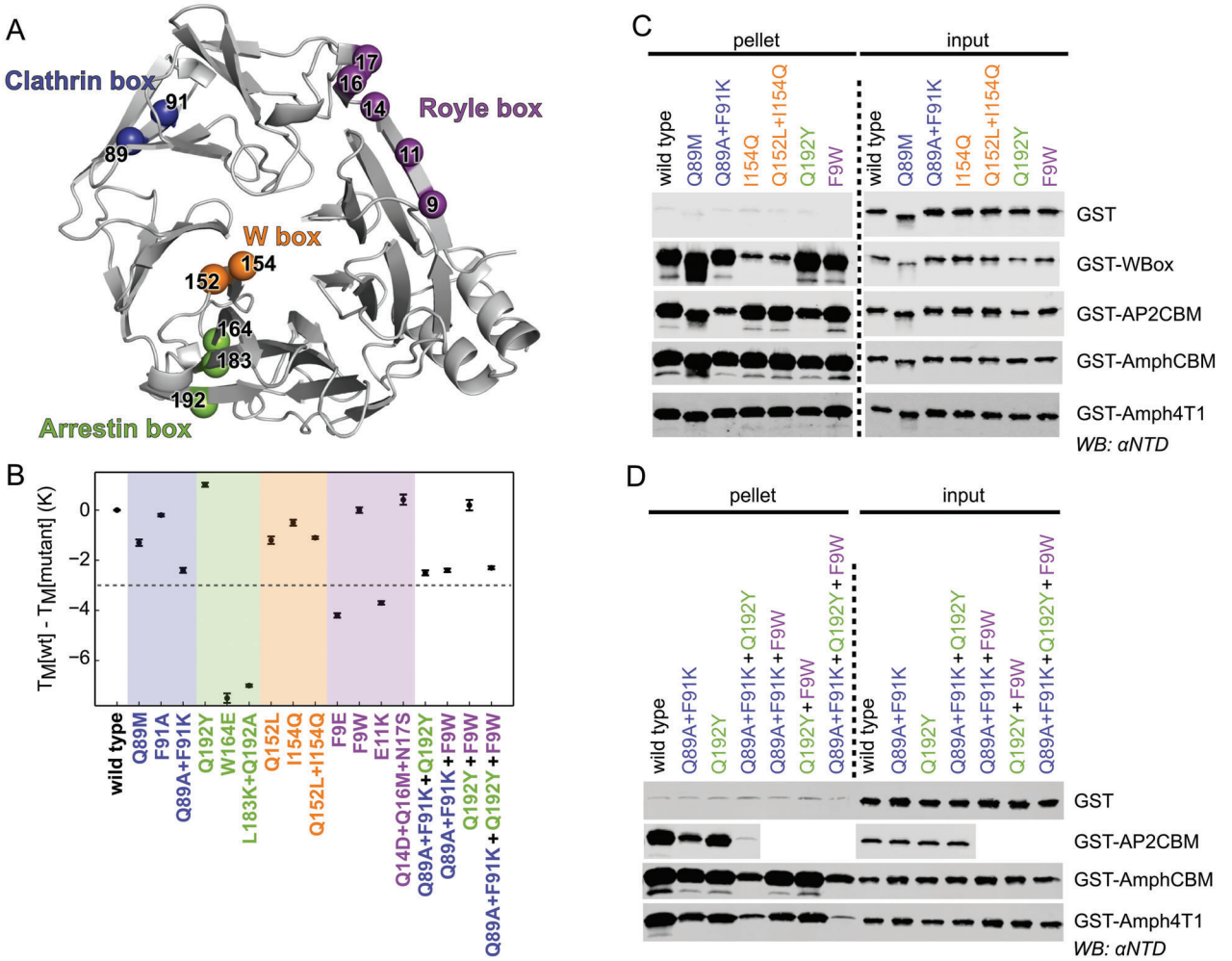


FIGURE 5 Mutation at single or multiple sites on clathrin N-terminal domain (NTD) disrupts binding to peptide motifs. A, Ribbon representation of NTD (grey) showing the location of residues that were mutated on their own or in combination to disrupt peptide binding at the clathrin box (blue), arrestin box (green), W box (orange) and Royle box (purple). B, Thermal stability of single- or multiple-site mutations of NTD as determined by differential scanning fluorimetry. The melting temperatures (T_M) of mutants relative to that of wild-type His-NTD are shown (error bars represent the standard deviation of measurement in triplicate). Mutations that perturb the T_M by more than 3 K (dotted line) were not considered further. C and D, Capture of NTD mutants by glutathione S-transferase (GST)-tagged clathrin-binding peptides. Purified recombinant wild-type or mutant His-NTD-NEMO was incubated with glutathione sepharose pre-loaded with GST-tagged "bait" proteins. After washing, proteins bound to the beads (pellet) were subjected to SDS-PAGE and immunoblotting (WB) using an antibody that recognizes clathrin NTD (αNTD).

TABLE 2 Clathrin heavy chain N-terminal domain mutations

Mutation	Site	Reference
Q89M	Clathrin box	25
F91A	Clathrin box	25
Q89A + F91K	Clathrin box	This study
Q192Y	Arrestin box	This study
W164E	Arrestin box	13
L183K + Q192A	Arrestin box	This study
Q152L	W box	12
I154Q	W box	12
Q152L + I154Q	W box	14
F9E	Royle box	This study
F9W	Royle box	This study
E11K	Royle box	14
Q14D + Q16M + N17S	Royle box	14

to clathrin NTD: structures of NTD solved in the presence of β3 adaptin and β-arrestin 2 CBMs showed binding only at the clathrin box,[10] only the W box site is occupied in the structure solved in the presence of a peptide derived from the "second" (PWDLW) clathrin-binding motif of amphiphysin,[12] and the structure of NTD in complex with a long splice form arrestin 2 (arrestin2L) shows binding of 2 different peptide motifs at the clathrin and arrestin box sites. We observe that the CBMs of β2 adaptin, amphiphysin and HDAg-L2 bind to both the clathrin and arrestin box sites. Further, we provide the first structural characterization of the putative fourth and final peptide binding site on clathrin NTD,[14] which we term the Royle box.

3.1 | The arrestin box binds linear CBM peptides

The structure of arrestin2L bound to clathrin NTD revealed two different peptide epitopes bound at the clathrin and arrestin boxes.[13]

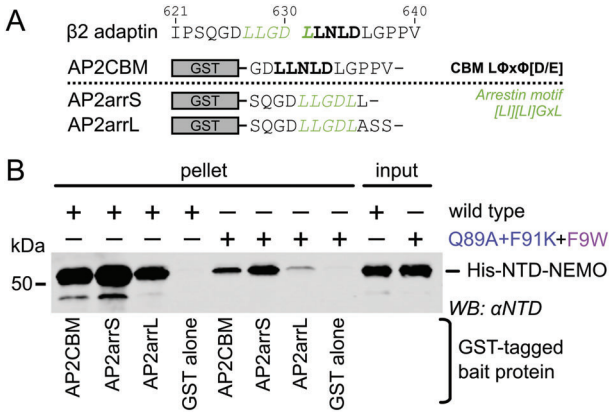


FIGURE 6 The overlapping β2 adaptin arrestin-box and clathrin-box motifs both bind multiple sites on clathrin N-terminal domain (NTD). A, Glutathione S-transferase (GST) fusions of the clathrin-box motif (GST-AP2CBM) and arrestin-binding motif (GST-AP2arrS and GST-AP2arrL) from the hinge region of β2 adaptin, the arrestin-box motif constructs having either the next residue of β2 adaptin ("L", GST-AP2arrS) or the sequence that follows the LLGDL motif of arrestin2L ("ASS", GST-AP2arrL) appended at their C termini. B, Capture of wild-type NTD or a mutant with disrupted clathrin and Royle boxes (Q89A + F91K + F9W) by GST-tagged β2 adaptin clathrin-binding motifs. Purified recombinant wild-type or mutant His-NTD-NEMO was incubated with glutathione sepharose pre-loaded with GST-tagged "bait" proteins. After washing, proteins bound to the beads (pellet) were subjected to SDS-PAGE and immunoblotting (WB) using an antibody that recognizes clathrin NTD (αNTD).

The epitope bound at the arrestin box comprised an 8 amino acid surface loop that connects 2 adjacent anti-parallel beta strands of arrestin2L: this loop thus necessarily forms a relatively tight turn at its apex. The key molecular interactions are formed by three leucine side chains (L334, L335 and L338) from the arrestin2L loop that are adjacent in space (Figure 3C), but not consecutive in sequence, and biochemical studies defined the consensus binding sequence of this loop as [LI][LI]GxL.[13]

The structures presented here show that CBM peptides, matching the LΦxΦ[DE] CBM consensus sequence, also bind NTD at the arrestin box (Figure 2) and that this interaction contributes to binding in vitro (Figure 5). Interestingly, the molecular determinants of binding are conserved between the arrestin2L loop and the CBM peptides: hydrophobic leucine or isoleucine side chains bind the groove between NTD β-propeller blades 4 and 5, occupying roughly equivalent spatial positions (Figure 3). However, the peptide backbone topology differs substantially as does the spacing between the crucial leucine/isoleucine residues. We therefore propose that the consensus sequence for binding at the arrestin box is likely to be context-dependent. In the case of arrestin2L, the amino acids that mediate binding are partly determined by the constrained nature of the surface loop. However, when presented as linear motifs, as is likely to be the biological context of the β2 adaptin, epsin 1 and amphiphysin CBM epitopes,[3,4,9,21,26] peptides that conform to the LΦxΦ[DE] CBM consensus sequence can also bind the arrestin box.

Our structural characterization of CBM peptides bound promiscuously at multiple sites on clathrin NTD is largely consistent with a recent biophysical study showing promiscuous binding of long peptides derived from β2 adaptin and AP180 at the clathrin, arrestin and W boxes of NTD.[15] However, while the long β2 adaptin peptide used in the biophysical study harboured both CBM (LLNLD) and arrestin2L (LLGDL) consensus sequences, the β2 adaptin CBM (AP2CBM) peptide used here contains only a CBM. We observe that combining mutations at the arrestin and clathrin boxes completely abolishes the ability of the AP2CBM to capture NTD in GST pull-down assays, confirming that CBM peptides bind promiscuously to both sites. Unlike the previous biophysical study, peptide binding at the W box was not observed in any of the crystal structures presented here. We ascribe this to the use of much longer peptides in the biophysical study that may harbour additional, non-canonical W box binding motifs.

Given the similar molecular determinants of linear peptide binding at the clathrin and arrestin boxes, it is perhaps surprising that binding at the arrestin box was not observed in the previous co-crystallization study that used the β3 adaptin and β-arrestin 2 CBM peptides.[10] However, we note the prior study used a lower molar excess of peptide (4-fold excess versus 7- to 10-fold excess used here). Further, the extended cryo-protection protocols employed in the prior study, using cryo-preservative solutions without added peptide, may have facilitated dissociation of bound peptides from lower-affinity sites and thus removed evidence of their binding.

3.2 | Structural characterization of the Royle box

As observed at the arrestin box (above), the molecular determinants of binding at the Royle box are conserved despite a difference in peptide orientation observed for the bound HDAg-L1$_{pep}$ peptide vs the bound HDAg-L2$_{pep}$ and Amph4T1$_{pep}$ peptides. However, unlike at the arrestin box, residues that form the molecular interactions at the Royle box do not correspond to those conserved in the LΦxΦ[DE] consensus CBM sequence and we note that several peptides containing a CBM sequence (AP2CBM$_{pep}$, AmphCBM$_{pep}$ and AmphCBMlong$_{pep}$) do not bind at the Royle box (Figures 2 and S2). Together, this suggests that a distinct consensus sequence mediates binding of cellular proteins at the Royle box. While the HDAg-L1$_{pep}$ and HDAg-L2$_{pep}$ peptides that bind the site occur naturally in hepatitis D virus, the Amph4T1$_{pep}$ sequence used in this study contains amino acids corresponding to those introduced when cloning the amphiphysin CBM into the pGEX-4T1 vector.[22] Attempts to define a consensus sequence and screen in silico for genuine cellular binding peptide motifs were unsuccessful due to the degeneracy of the peptide sequences bound in our structures. Identification of the Royle box consensus binding motif therefore awaits experimental elucidation.

In accordance with previous functional studies,[14] all three structures of peptides bound at the Royle box presented here show binding at a conserved surface patch that lies between blades 6 and 7 of the NTD β-propeller (Figure 4B). The bound peptides all wrap around the hydrophobic side chain of NTD residue F9, which also lines the hydrophobic pocket central to the interaction of peptides at the Royle box (Figures 4C and 4D). Mutation of F9 to the bulkier residue tryptophan does not destabilize NTD (Figure 5B) but is able to diminish binding to GST-Amph4T1 when combined with mutations at the

arrestin and clathrin boxes, confirming the importance of F9 for peptide binding. Previous functional experiments showed that 2 sets of mutations at the Royle box, E11K and Q14D + Q16M + N17S (Figure 4D), were sufficient to prevent transferrin uptake (a readout for clathrin-mediated endocytosis) when combined with mutations at the clathrin, arrestin and W boxes. While the side chain of E11 does not interact directly with bound peptides, in the HDAg-L1$_{pep}$ and Amph4T1$_{pep}$ structures E11 backbone atoms form hydrogen bonds with the bound peptide. Further, we find that the E11K mutation reduces the T_M of NTD by 3.7 K, consistent with some destabilization of the protein fold. We propose that destabilization of the local fold in the E11K NTD mutant prevents its binding to Royle box binding epitopes in cells. Residues of the second disrupting mutation, Q14D + Q16M + N17S, do not directly contact peptides bound at the Royle box in our structures but are in close proximity to peptide-binding residues (Figure 4D). Residue Q14 lies on the same short stretch of β-sheet as E11, the side chains of these two residues forming a hydrogen bond, while Q16 and N17 lie on the surface of a short α-helix immediately following Q14 (Figure 4D). The thermal stability of purified Q14D + Q16M + N17S NTD is slightly higher than the wild-type NTD (Figure 5B), which may indicate perturbation of the protein fold in the vicinity of bound peptide. Alternatively, one could speculate that the N-terminal residues of a bona fide cellular Royle box binding motif could bind the Royle box in an extended conformation and interact with these residues, although identification of such a motif remains elusive. A third set of mutations at the Royle box, N296A + R297E, did not seem to affect the ability of NTD to promote transferrin uptake,[14] yet both residues are in close proximity to the peptide bound at the Royle box in the structures presented here. It is possible that the precise nature of substitutions introduced at residues N296 and R297 led to sustained transferrin uptake: N296 lines the rim of the deep hydrophobic pocket and its mutation to alanine, conferring a short hydrophobic side chain, should not prevent binding. Similarly, interactions with R297 are mediated primarily by the backbone and hydrophobic C^{β} and C^{γ} side chain atoms, all of which would remain when the residue was mutated to glutamate.

3.3 | HDV peptides bind the same sites on NTD as cellular peptides

Previous studies showed that GST fused to residues 198-210 of HDAg-L, comprising the majority of the HDAg-L C-terminal extension that is expressed following editing of the HDV RNA antigenome,[27] is capable of capturing clathrin heavy chain from cell lysates.[16] Further, mutation of a putative CBM sequence in this C-terminal extension prevented both co-immunoprecipitation of clathrin heavy chain from transfected cells and the formation of virus-like particles (VLP).[18] It was therefore concluded that HDAg-L is a novel viral clathrin adaptor-like protein.[16,18] We sought to extend this observation by probing whether cellular and HDAg-L peptides bind the same or different sites on clathrin NTD, with a view to developing small-molecule inhibitors of the HDAg-L interaction with clathrin NTD that would restrict HDV replication.

Our structural results show that peptides containing the putative CBMs of two distinct HDAg-L sequences (HDAg-L1$_{pep}$ and HDAg-L2$_{pep}$) bind the same sites on clathrin NTD as cellular CBM peptides, binding at the clathrin box, the Royle box and (for HDAg-L2$_{pep}$) the arrestin box (Figure 2). However, GST pull-down experiments performed with either purified clathrin or an oligomerized form of the NTD (His-NTD-NEMO) showed that these viral peptides capture NTD much less efficiently than do cellular CBM epitopes (Figure 1B,C). Previous cell-based studies showed that mutation of HDAg-L1 L199 to alanine severely reduced VLP production and clathrin co-immunoprecipitation.[18] This is consistent with our structure, as L199 forms extensive hydrophobic contacts at both the Royle and clathrin boxes (Figure 2). HDAg-L1 mutation D203A also diminished VLP production and co-immunoprecipitation with clathrin, but in the structures presented here this residue is consistently disordered. It is thus unclear whether this mutation directly affects binding of HDAg-L to clathrin or has some secondary effect. Previous biochemical studies of HDAg-L peptides binding to purified NTD also suggested that both L199 and D203 are important for the interaction.[16,28] However, these experiments should be viewed with caution as they utilized an extremely short NTD construct (residues 1-107) that not only lacks both the Royle and arrestin box sites but also spans only the first two blades of the NTD β-propeller and is thus highly unlikely to be correctly folded. Taken together, our results confirm that the putative CBM peptides from HDAg-L can directly bind clathrin NTD, but do so weakly. It is therefore unclear whether these viral CBM-like epitopes directly promote recruitment of clathrin heavy chain in vivo or whether they act synergistically with other clathrin-adaptor proteins.

3.4 | Degeneracy of clathrin-binding peptide motifs

Our structural (Figure 2) and biochemical (Figure 5) studies show that 2 distinct sequence motifs can bind the arrestin box of NTD: the arrestin-box motif ([LI][LI]GxL) and the CBM (LΦxΦ[DE]). Further, Figure 6B shows that, when presented as linear peptides, either the β2 adaptin CBM motif (GST-AP2CBM) or an extended arrestin-box motif (GST-AP2arrS, where the arrestin-box motif is followed by a leucine residue) bind the arrestin box more strongly than does the arrestin-box motif alone (GST-AP2arrL). This experiment also suggests that the β2 adaptin arrestin-box motif is capable of binding the clathrin or Royle boxes (compare capture of wild-type vs Q98A + F91K + F9W His-NTD-NEMO), despite this arrestin-box motif (LLGDL) not conforming to the CBM consensus sequence. Similarly, we note that the sequences capable of binding the Royle box in crystallo are also rather degenerate, precluding the identification of a consensus binding sequence. Together, this suggests that the model of "1 consensus binding motif per peptide-binding site on clathrin NTD" might need revisiting, as the binding of these short peptides to the NTD surface is degenerate and may depend on the structural context in which the peptides are presented.

3.5 | Dynamics of association between clathrin terminal domain and adaptor peptides in coated pits

The results presented in this study underline the dynamic nature of clathrin:adaptor interactions and suggest that each clathrin terminal domain is capable of simultaneously binding multiple adaptors, even those containing only CBM (LΦxΦ[DE]) or arrestin-box ([LI][LI]GxL) motifs. Individually, these interactions are of low affinity.[4,15] Proteins typically bind to both specific and non-specific sites with similar association rates (k_{on}), with differential affinity conferred by differing rates of dissociation (k_{off}).[29] Weak bimolecular interactions in the micromolar affinity range, such as those between clathrin and its adaptors, typically correspond to dissociation rates of about 1 second or a half-time of dissociation of ~0.7 seconds.[29] This is significantly shorter than the timescale of productive clathrin-coated pit assembly and deformation, which occurs over the course of ~90 seconds.[30] Thus, we would expect adaptors to display rapid cycles of binding to and dissociation from individual binding sites on clathrin, allowing recruitment of a multitude of different adaptor molecules to any given clathrin terminal domain. It is also possible that plasticity in clathrin motif binding allows individual adaptors harbouring multiple clathrin-interaction motifs, such as epsin,[22] to bind multiple sites on the clathrin N-terminal domain simultaneously, thereby increasing their apparent affinity. However, as we are not yet able to define a consensus binding sequence for the Royle box, and considering the degenerate sequence requirements for binding at the clathrin or arrestin boxes, it is unclear how frequently clathrin adaptors might be able to employ such a mode of binding.

In summary, we have shown that cellular CBM peptides bind degenerately to multiple sites on clathrin, we define a set of NTD mutations at each of the 4 peptide binding sites that do not disrupt the NTD fold, and show in biochemical assays that multiple sites contribute to binding of NTD by cellular clathrin-binding peptides. In addition, we find that viral CBM peptides bind the same sites on NTD as cellular peptides, albeit much more weakly. Finally, we present the first structural characterization of the Royle box, the fourth and final functional peptide binding site on the clathrin NTD.

4 | MATERIALS AND METHODS

4.1 | Constructs and mutagenesis

Wild-type bovine clathrin heavy chain N-terminal domain (1-363) (NTD) with an N-terminal glutathione S-transferase (GST) purification tag and thrombin cleavage site was used as described previously.[11] For binding assays, an oligomeric construct was designed by fusing clathrin NTD (1-363) to the NF-κB essential modulator (NEMO) oligomerization domain (246-365) and cloning into pET-28(a) to add an N-terminal His_6 purification tag (His-NTD-NEMO). Mutated constructs encoding His-NTD-NEMO with single amino acid substitutions at residues F9, E11, Q14, Q16, N17, Q89, F91, Q152, I154, W164, L183 and Q192 were generated by QuikChange site-directed mutagenesis (Agilent) and introduction of the desired mutations was confirmed by Sanger sequencing. Clathrin-binding motifs were cloned into pOPT3G,[31] encoding GST followed by a human rhinovirus 3C protease cleavage site and the peptide of interest, by ligation of phosphorylated annealed oligonucleotide primers as follows: GST-HDAg-L1, residues 195-214 of hepatitis D virus large antigen (HDAg-L) clade I (UniProt P0C6L6); GST-HDAg-L2, residues 194-213 of HDAg-L clade II (UniProt A4ZNG7); GST-AP2CBM, residues 629-640 of the β2 adaptin subunit of human AP2 (UniProt P63010); GST-Wbox, residues 379-388 of human amphiphysin I (UniProt P49418); GST-AmphCBM, residues 349-358 of human amphiphysin I (Uniprot P49418). Two constructs containing the arrestin-box motif (LLGDL) of the β2 adaptin subunit of human AP2 (UniProt P63010) were generated the in the same manner: GST-AP2arrS, residues 623-632 and GST-AP2arrL, residues 623-631 followed by the sequence "ASS" that corresponds to the residues C-terminal to the LLGDL arrestin-box motif of arrestin2L.[13] An additional construct (GST-Amph4T1) encoding residues 349-356 of rat amphiphysin I (UniProt O08838; rat and human amphiphysin I residues 349-356 are identical), inserted into pGEX-4T1 after EcoRI/XhoI digestion and thus encoding 7 additional amino acids (LERPHRD) C-terminal to the amphiphysin sequence, was described previously.[22]

4.2 | Protein expression and purification

A clathrin-coated vesicle fraction was isolated from pig brains essentially as described.[32] Coat proteins were stripped from the vesicles[33] and clathrin was purified from the coat protein mixture by gel filtration in 0.5 M Tris pH 7.0, 1 mM DTT (Superose 6, GE Healthcare), dialyzed into 10 mM Tris pH 8.0 and stored at 4°C. All other proteins were expressed in *E. coli* strains BL21(DE3) pLysS (GST-tagged constructs) or B834(DE3) (wild-type and mutant His-NTD-NEMO). Bacteria were grown in 2×TY medium with appropriate selection antibiotics to an optical density $(OD)_{600}$ of 0.8-1.0, the temperature was reduced to 22°C and protein expression was induced by addition of 0.2 mM isopropyl β-D-thiogalactopyranoside. After 12-18 h cells were harvested by centrifugation (6000 × g, 15 min, 4°C) and stored at −80°C.

Bacterial pellets containing GST-NTD were resuspended in lysis buffer (20 mM Tris pH 7.5, 200 mM NaCl, 0.05% TWEEN-20, 0.5 mM $MgCl_2$, 1.4 mM β-mercaptoethanol) supplemented with 200-400 U bovine DNase I (Sigma) and 200 μL ethylenediaminetetraacetic acid (EDTA)-free protease inhibitor cocktail (Sigma). Cells were lysed at 24 kpsi using a cell disruptor (Constant Systems) and the lysate was cleared by centrifugation (40 000*g*, 30 min, 4°C). Cleared lysate was incubated with glutathione Sepharose 4B (GE Healthcare) for 60 minutes at 4°C, the beads were washed (20 mM Tris pH 7.5, 200 mM NaCl, 1 mM DTT), equilibrated in thrombin cleavage buffer (20 mM Tris pH 7.5, 200 mM NaCl, 1 mM $CaCl_2$), and the GST tag removed by overnight incubation at room temperature with thrombin (125 U, Serva). Following incubation with fresh glutathione resin to capture liberated GST and uncleaved GST-NTD fusion, NTD was further purified using a HiLoad Superdex 200 size exclusion chromatography column (GE Healthcare) equilibrated in 10 mM Tris pH 7.5, 50 mM NaCl, 4 mM DTT. Following concentration, small aliquots (20-100 μL) of purified NTD were

snap-frozen in liquid nitrogen and stored at −80°C.[34] For other GST-tagged proteins, cell pellets were resuspended in lysis buffer (20 mM HEPES pH 7.5, 300 mM NaCl, 0.05% TWEEN-20, 0.5 mM $MgCl_2$, 1.4 mM β-mercaptoethanol) supplemented with DNase I and protease inhibitors as above. Cells were lysed, and lysates were clarified and incubated with glutathione resin as above. The glutathione resin was washed (20 mM HEPES pH 7.5, 300 mM NaCl, 1 mM DTT) and bound proteins were eluted using wash buffer supplemented with 25 mM reduced glutathione. Following size exclusion chromatography using HiLoad Superdex 75 or 200 columns (GE Healthcare) equilibrated in 20 mM HEPES pH 7.5, 200 mM NaCl, 1 mM DTT, fusion proteins were mixed 1:1 with 100% (v/v) glycerol and stored at −20°C.

For wild-type and mutant His-NTD-NEMO, cell pellets were resuspended in lysis buffer (20 mM Tris pH 7.5, 500 mM NaCl, 0.05% TWEEN-20, 0.5 mM $MgCl_2$, 1.4 mM β-mercaptoethanol) supplemented with 2-20 mg hen egg white lysozyme (Sigma), 400 U bovine DNase I (Sigma) and 200 µL EDTA-free protease inhibitor cocktail (Sigma). The cells were lysed and the lysate clarified as described above. The cleared lysate was applied to a 1 mL HisTrap excel Ni affinity column (GE Healthcare), the column was washed (20 mM Tris pH 7.5, 500 mM NaCl, 12.5 mM imidazole pH 7.5) and the protein eluted (20 mM Tris pH 7.5, 500 mM NaCl, 250 mM imidazole pH 7.5). The Ni affinity column eluate was injected onto a HiLoad Superdex 200 column (GE Healthcare) equilibrated in 20 mM Tris pH 7.5, 200 mM NaCl, 1 mM DTT and eluted in distinct peaks, which were collected and concentrated separately to yield His-NTD-NEMO and His-NTD (Figure S1), which were both were snap-frozen in liquid nitrogen as described above.

4.3 | Crystallization, data collection, structure determination and analysis

Crystals were grown at 20°C by sitting drop vapour diffusion. Clathrin NTD (1-363) was co-crystallized in complex with the following peptides (peptide sequences are listed with residues not present in the GST-tagged constructs underlined): $Amph4T1_{pep}$ (ETLLDLDFLE); $AmphCBM_{pep}$ (ETLLDLDFDP); $AP2CBM_{pep}$ (CGDLLNLDLG); $HDAg\text{-}L1_{pep}$ (SDILFPADS); $HDAg\text{-}L2_{pep}$ (SPRLPLLES); $AmphCBMlong_{pep}$ (ETLLDLDFDPFK). Peptides were purchased from Genscript ($Amph4T1_{pep}$, $AmphCBM_{pep}$ and $AmphCBMlong_{pep}$) or Designer Bioscience ($AP2CBM_{pep}$, $HDAg\text{-}L1_{pep}$ and $HDAg\text{-}L2_{pep}$). All peptides were prepared as 10 mM stock solutions in 10 mM Tris pH 7.5, 50 mM NaCl, 4 mM DTT and stored at −20°C. NTD was mixed 2:1 with peptide to give final concentrations of 14 mg/mL NTD and 3.4 mM peptide for all crystallization experiments except NTD:$HDAg\text{-}L2_{pep}$, where 20 mg/mL NTD and 3.6 mM peptide was used. Crystals for structure determination were obtained under the following conditions (P and R indicate peptide:protein and reservoir volumes in sitting drops, respectively): NTD:$Amph4T1_{pep}$, 1 µL P plus 1 µL R equilibrated against a 200 µL reservoir of 1.1 M sodium malonate pH 8.0 (Hampton Research); NTD:$AmphCBM_{pep}$, 1 µL P plus 2 µL R equilibrated against a 200 µL reservoir of 0.85 M sodium malonate pH 7.5; NTD:$AP2CBM_{pep}$, 1 µL P plus 2 µL R equilibrated against a 200 µL reservoir of 0.94 M sodium malonate pH 6.7; NTD:$HDAg\text{-}L1_{pep}$, 400 nL P plus 200 nL R equilibrated against a 80 µL reservoir of 1.21 M sodium malonate pH 7.0; NTD:$HDAg\text{-}L2_{pep}$, 200 nL P plus 400 nL R equilibrated against a 80 µL reservoir of 1.75 M sodium malonate pH 7.0; NTD:$AmphCBMlong_{pep}$, 1 µL P plus 2 µL R (1.04 M sodium malonate pH 7.1, 0.2 M sodium perchlorate [Jena Bioscience]) equilibrated against a 200 µL reservoir of 1.15 M sodium malonate pH 7.1. All crystals were cryoprotected by rapid transfer into a drop comprising 55% reservoir solution, 25% (v/v) glycerol and 20% (v/v) 10 mM peptide stock solution, the peptide being added to prevent dissociation from NTD of bound peptides. Crystals were then immediately flash-cooled in liquid nitrogen and stored in liquid nitrogen.

Diffraction data were recorded at 100 K on Diamond Light Source beamlines I02 (NTD:$HDAg\text{-}L1_{pep}$, NTD:$HDAg\text{-}L2_{pep}$ and NTD:$AmphCBMlong_{pep}$) and I04-1 (NTD:$Amph4T1_{pep}$, NTD:$AmphCBM_{pep}$ and NTD:$AP2CBM_{pep}$). Data were processed using XDS, XSCALE and Aimless (NTD:$Amph4T1_{pep}$, NTD:$HDAg\text{-}L1_{pep}$ and NTD:$AmphCBMlong_{pep}$), or DIALS and Aimless (NTD:$AmphCBM_{pep}$ and NTD:$AP2CBM_{pep}$), as implemented by the xia2 automated processing pipeline,[35–41] or using iMOSFLM[42] and Aimless interactively (NTD:$HDAg\text{-}L2_{pep}$). The structures of the NTD:$HDAg\text{-}L1_{pep}$, NTD:$HDAg\text{-}L2_{pep}$, NTD:$Amph4T1_{pep}$, NTD:$AmphCBM_{pep}$ and NTD:$AmphCBMlong_{pep}$ complexes were solved by isomorphous replacement in REFMAC5[43,44] using a high-resolution ligand-free model of NTD (PDB 1C9I)[10] as a starting model. The structure of the NTD:$AP2CBM_{pep}$ complex was solved by molecular replacement with a single chain of NTD (PDB 1C9I, chain A) as a search model using Phaser.[45] Manual model building was performed using COOT[46] and the models were refined using REFMAC5. In all structures, the peptides were modelled after initial improvement of the peptide-free structure. The geometric quality of the models was improved by consulting the validation tools in COOT as well as the programmes MolProbity[47] and WHAT_CHECK.[48] Structure factors and final refined models have been deposited with the Protein Data Bank as listed in Table 1 and Table S2. Feature-enhanced maps, which have reduced model bias and optimized scaling to ease comparison of strong and weak features, were calculated using phenix.fem.[24,49] Evolutionary conservation of amino acids was estimated using ConSurf[50] with default parameters and chain A of the refined NTD:$Amph4T1_{pep}$ structure as an input model. PyMOL (Schrodinger) was used to generate molecular graphics and figures were assembled using Inkscape (https://inkscape.org/).

4.4 | Capture (GST pull-down) assays and immunoblotting

All steps of the clathrin or His-NTD-NEMO capture assays were performed at 4°C using a previously published protocol[11] adapted to enable detection of the very weak interactions investigated in this work. A total of 40 µL of glutathione sepharose 4B beads (GE Healthcare) pre-equilibrated in assay buffer (25 mM HEPES-KOH, 125 mM potassium acetate, 5 mM magnesium acetate, 2 mM

EDTA, 2 mM EGTA, 1 mM DTT, pH 7.2) were incubated for 2 h with 20 μg (for purified clathrin pull-down experiments) or 500 μg (for recombinant NTD pull-downs) of GST or GST fusion proteins in assay buffer to a final volume of 400 μL. Non-immobilized bait protein was removed following centrifugation (10 000*g*, 2 min) and the resin was washed thrice with assay buffer. The protein-loaded resin was then incubated with 300 μL of 0.1 mg/mL His-NTD-NEMO or His-NTD, or 0.4 μM purified clathrin, for 2 hours. Following centrifugation, supernatants containing uncaptured protein were retained for analysis and the resins were washed 4× with phosphate-buffered saline. After the final wash, the resin pellet was resuspended in 80 μL SDS-PAGE buffer and the samples were eluted by boiling for 5 min at 95°C. Input and supernatant samples were prepared in SDS-PAGE buffer. Samples (0.33% of the prey input samples, 11.25% of the eluted pellet samples, 0.6% of the supernatant kept after the prey incubation) were separated by SDS-PAGE and transferred to nitrocellulose membranes before immunoblotting using a mouse anti-clathrin N-terminal domain primary antibody (ab11221, Abcam) and fluorescently labelled goat anti-mouse secondary antibody (925-32210, LI-COR). Dried membranes were visualized using an Odyssey scanner (LI-COR).

4.5 | Biophysical assays: CD spectroscopy, differential scanning fluorimetry and multi-angle light scattering

Circular dichroism spectra were recorded on a Jasco J-810 spectropolarimeter at 20°C. Protein samples were diluted to 1 mg/mL in 50 mM phosphate buffer, pH 7.4. A total of 8 spectra per sample were recorded (50 nm/min, 1 nm bandwidth, 260-190 nm), averaged, and smoothed (Savitzky and Golay method, second order smoothing, 5 nm sliding window) using PRISM 5 (GRAPHPAD Software).

Differential scanning fluorimetry experiments to determine the melting temperature (T_M) of wild-type or mutant His-NTD were performed using a MiniOpticon real-time PCR system (BioRad) with 10× SYPRO Orange dye (Molecular Probes) or Viia7 real-time PCR system (Applied Biosystems) using 1× Protein Thermal Shift dye (Applied Biosystems). Multiple experiments confirmed that the difference between the T_M of wild-type His-NTD and mutants (T_M[wt]-T_M[mutant]) is measured consistently using either platform. In all experiments, assay buffer (20 mM HEPES-KOH, 120 mM potassium acetate, pH 7.5) was mixed with dye stock solution and protein solution in an 8:1:1 ratio, to give 10 ng protein in a final volume of 50 μL (MiniOpticon) or 2 ng protein in a final volume of 20 μL (Viia7). Samples (measured in triplicate) were heated from 20°C to 90°C at 1 K/minutes (MiniOpticon) or 25 to 95°C at 1 K/20 seconds (Viia7) and fluorescence was monitored at 1 K increments. Curve fitting, melting temperature calculations and plotting were performed using MATLAB (MathWorks).

Multi-angle light scattering (MALS) experiments[51] were performed at 22°C using a Superdex 200 10/300 gel filtration column (GE Healthcare) equilibrated in 20 mM Tris pH 7.5, 200 mM NaCl, 1 mM DTT. Samples (100 μL) were injected at a flow rate of 0.5 mL/min and size exclusion was followed by inline measurement of static light scattering (DAWN 8+, Wyatt Technology) and differential refractive index (Optilab T-rEX, Wyatt Technology). The data were analysed using Astra6 (Wyatt Technology).

ACKNOWLEDGMENTS

We thank Janet Deane for access to MALS equipment and helpful comments. We thank Diamond Light Source for access to beamlines I02 and I04-1 (mx8547 and mx11235), this access being supported in part by the EU FP7 infrastructure grant BIOSTRUCT-X (contract no. 283570). This work was supported by a Sir Henry Dale Fellowship, jointly funded by the Royal Society and the Wellcome Trust, to S.C.G. (098406/Z/12/Z), by an NIH R01 grant (GM106963; L.M.T.) and by a Wellcome grant (090909/Z/09/Z; B.T.K.). J.M. holds a Wellcome Trust studentship. CIMR is supported by a Wellcome Trust Strategic Award (079895).

The Editorial Process File is available in the online version of this article.

REFERENCES

1. Unanue ER, Ungewickell E, Branton D. The binding of clathrin triskelions to membranes from coated vesicles. *Cell*. 1981;26(3 Pt 1):439-446.
2. Kirchhausen T. Clathrin. *Annu Rev Biochem*. 2000;69:699-727.
3. Owen DJ, Collins BM, Evans PR. Adaptors for clathrin coats: structure and function. *Annu Rev Cell Dev Biol*. 2004;20:153-191.
4. Shih W, Gallusser A, Kirchhausen T. A clathrin-binding site in the hinge of the beta 2 chain of mammalian AP-2 complexes. *J Biol Chem*. 1995;270(52):31083-31090.
5. Keyel PA, Mishra SK, Roth R, Heuser JE, Watkins SC, Traub LM. A single common portal for clathrin-mediated endocytosis of distinct cargo governed by cargo-selective adaptors. *Mol Biol Cell*. 2006;17(10):4300-4317.
6. Blondeau F, Ritter B, Allaire PD, et al. Tandem MS analysis of brain clathrin-coated vesicles reveals their critical involvement in synaptic vesicle recycling. *Proc Natl Acad Sci USA*. 2004;101(11):3833-3838.
7. Borner GH, Harbour M, Hester S, Lilley KS, Robinson MS. Comparative proteomics of clathrin-coated vesicles. *J Cell Biol*. 2006;175(4):571-578.
8. Lemmon SK, Traub LM. Getting in touch with the clathrin terminal domain. *Traffic*. 2012;13(4):511-519.
9. Dell'Angelica EC, Klumperman J, Stoorvogel W, Bonifacino JS. Association of the AP-3 adaptor complex with clathrin. *Science*. 1998;280(5362):431-434.
10. ter Haar E, Harrison SC, Kirchhausen T. Peptide-in-groove interactions link target proteins to the beta-propeller of clathrin. *Proc Natl Acad Sci USA*. 2000;97(3):1096-1100.
11. Drake MT, Traub LM. Interaction of two structurally distinct sequence types with the clathrin terminal domain beta-propeller. *J Biol Chem*. 2001;276(31):28700-28709.
12. Miele AE, Watson PJ, Evans PR, Traub LM, Owen DJ. Two distinct interaction motifs in amphiphysin bind two independent sites on the clathrin terminal domain beta-propeller. *Nat Struct Mol Biol*. 2004;11(3):242-248.
13. Kang DS, Kern RC, Puthenveedu MA, von Zastrow M, Williams JC, Benovic JL. Structure of an arrestin2-clathrin complex reveals a novel clathrin binding domain that modulates receptor trafficking. *J Biol Chem*. 2009;284(43):29860-29872.
14. Willox AK, Royle SJ. Functional analysis of interaction sites on the N-terminal domain of clathrin heavy chain. *Traffic*. 2012;13(1):70-81.
15. Zhuo Y, Cano KE, Wang L, et al. Nuclear magnetic resonance structural mapping reveals promiscuous interactions between clathrin-box motif sequences and the N-terminal domain of the clathrin heavy chain. *Biochemistry*. 2015;54(16):2571-2580.

16. Huang C, Chang SC, Yu IC, Tsay YG, Chang MF. Large hepatitis delta antigen is a novel clathrin adaptor-like protein. *J Virol*. 2007;81(11):5985-5994.
17. Ivanovic T, Boulant S, Ehrlich M, et al. Recruitment of cellular clathrin to viral factories and disruption of clathrin-dependent trafficking. *Traffic*. 2011;12(9):1179-1195.
18. Huang C, Chang SC, Yang HC, Chien CL, Chang MF. Clathrin-mediated post-Golgi membrane trafficking in the morphogenesis of hepatitis delta virus. *J Virol*. 2009;83(23):12314-12324.
19. von Kleist L, Stahlschmidt W, Bulut H, et al. Role of the clathrin terminal domain in regulating coated pit dynamics revealed by small molecule inhibition. *Cell*. 2011;146(3):471-484.
20. Slepnev VI, Ochoa GC, Butler MH, De Camilli P. Tandem arrangement of the clathrin and AP-2 binding domains in amphiphysin 1 and disruption of clathrin coat function by amphiphysin fragments comprising these sites. *J Biol Chem*. 2000;275(23):17583-17589.
21. Ramjaun AR, McPherson PS. Multiple amphiphysin II splice variants display differential clathrin binding: identification of two distinct clathrin-binding sites. *J Neurochem*. 1998;70(6):2369-2376.
22. 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(9):6479-6489.
23. Ivins FJ, Montgomery MG, Smith SJ, Morris-Davies AC, Taylor IA, Rittinger K. NEMO oligomerization and its ubiquitin-binding properties. *Biochem J*. 2009;421(2):243-251.
24. Afonine PV, Moriarty NW, Mustyakimov M, et al. FEM: feature-enhanced map. *Acta Crystallogr D Biol Crystallogr*. 2015;71(Pt 3):646-666.
25. Goodman OB Jr, Krupnick JG, Gurevich VV, Benovic JL, Keen JH. Arrestin/clathrin interaction. Localization of the arrestin binding locus to the clathrin terminal domain. *J Biol Chem*. 1997;272(23):15017-15022.
26. Kalthoff C, Alves J, Urbanke C, Knorr R, Ungewickell EJ. Unusual structural organization of the endocytic proteins AP180 and epsin 1. *J Biol Chem*. 2002;277(10):8209-8216.
27. Casey JL. RNA editing in hepatitis delta virus. *Curr Top Microbiol Immunol*. 2006;307:67-89.
28. Wang YC, Huang CR, Chao M, Lo SJ. The C-terminal sequence of the large hepatitis delta antigen is variable but retains the ability to bind clathrin. *Virol J*. 2009;6:31.
29. Pollard TD. A guide to simple and informative binding assays. *Mol Biol Cell*. 2010;21(23):4061-4067.
30. Loerke D, Mettlen M, Yarar D, et al. Cargo and dynamin regulate clathrin-coated pit maturation. *PLoS Biol*. 2009;7(3):e57.
31. Graham SC, Wartosch L, Gray SR, et al. Structural basis of Vps33A recruitment to the human HOPS complex by Vps16. *Proc Natl Acad Sci USA*. 2013;110(33):13345-13350.
32. Smith CJ, Grigorieff N, Pearse BM. Clathrin coats at 21 A resolution: a cellular assembly designed to recycle multiple membrane receptors. *EMBO J*. 1998;17(17):4943-4953.
33. Pearse BM, Robinson MS. Purification and properties of 100-kd proteins from coated vesicles and their reconstitution with clathrin. *EMBO J*. 1984;3(9):1951-1957.
34. Deng J, Davies DR, Wisedchaisri G, Wu M, Hol WG, Mehlin C. An improved protocol for rapid freezing of protein samples for long-term storage. *Acta Crystallogr D Biol Crystallogr*. 2004;60(Pt 1):203-204.
35. Winter G. Xia2: an expert system for macromolecular crystallography data reduction. *J Appl Crystall*. 2009;43:186-190.
36. Kabsch W. XDS. *Acta Crystallogr D Biol Crystallogr*. 2010;66(Pt 2):125-132.
37. Evans P. Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr*. 2006;62(Pt 1):72-82.
38. Sauter NK, Grosse-Kunstleve RW, Adams PD. Robust indexing for automatic data collection. *J Appl Crystallogr*. 2004;37(Pt 3):399-409.
39. Zhang Z, Sauter NK, van den Bedem H, Snell G, Deacon AM. Automated diffraction image analysis and spot searching for high-throughput crystal screening. *J Appl Crystallogr*. 2006;39:112-119.
40. Collaborative Computational Project. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr*. 1994;50(Pt 5):760-763.
41. Evans PR. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr D Biol Crystallogr*. 2011;67(Pt 4):282-292.
42. Powell HR, Johnson O, Leslie AG. Autoindexing diffraction images with iMosflm. *Acta Crystallogr D Biol Crystallogr*. 2013;69(Pt 7):1195-1203.
43. Murshudov GN, Skubak P, Lebedev AA, et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr*. 2011;67(Pt 4):355-367.
44. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*. 1997;53(Pt 3):240-255.
45. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Crystallogr*. 2007;40:658-674.
46. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*. 2010;66(Pt 4):486-501.
47. Chen VB, Arendall WB 3rd, Headd JJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*. 2010;66(Pt 1):12-21.
48. Hooft RW, Vriend G, Sander C, Abola EE. Errors in protein structures. *Nature*. 1996;381(6580):272.
49. Adams PD, Afonine PV, Bunkoczi G, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*. 2010;66(Pt 2):213-221.
50. Celniker G, Nimrod G, Ashkenazy H, et al. ConSurf: using evolutionary data to raise testable hypotheses about protein function. *Isr J Chem*. 2013;53:199-206.
51. Wyatt PJ. Light-scattering and the absolute characterization of macromolecules. *Anal Chim Acta*. 1993;272(1):1-40.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Muenzner J, Traub LM, Kelly BT and Graham SC. Cellular and viral peptides bind multiple sites on the N-terminal domain of clathrin, *Traffic*, 2017;18(1):44–57.

Cellular and viral peptides bind multiple sites on the N-terminal domain of clathrin

Julia Muenzner[1], Linton M. Traub[2], Bernard T. Kelly[3]*, Stephen C. Graham[1]*

[1]Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK

[2]Department of Cell Biology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15261, USA

[3]Cambridge Institute for Medical Research, Department of Clinical Biochemistry, University of Cambridge, Hills Road, Cambridge CB2 0XY, UK

*Corresponding authors: Bernard T. Kelly, btk1000@cam.ac.uk, and Stephen C. Graham, scg34@cam.ac.uk

Table S1: Average atomic displacement parameters (ADPs) of NTD peptide-binding residues and of bound peptides. In crystal structures at near-atomic resolution the ADP of an atom is an indication of its degree of order. Peptide-binding residues were defined as any NTD residue within 5 Å of a bound peptide, not including molecules related by crystallographic symmetry. Average isotropic ADPs for were calculated using phenix.pdbtools. ΔADP is the difference between mean ADPs of the peptide and the residues to which it binds.

		Clathrin box			Arrestin box			Royle box		
Structure	**Chain**	**Peptide**	**NTD**	**ΔADP**	**Peptide**	**NTD**	**ΔADP**	**Peptide**	**NTD**	**ΔADP**
AP2CBM$_{pep}$	A	48.4	33.2	15.2	57.9	29.2	28.7	–	–	–
AmphCBM$_{pep}$	A	34.7	19.6	15.1	53.6	20.9	32.7	–	–	–
	B	38.3	22.2	16.1	49.1	18.2	30.9	–	–	–
Amph4T1$_{pep}$	A	31.0	21.0	10.0	50.8	23.5	27.3	53.5	32.7	20.8
	B	36.8	24.2	12.6	63.8	22.6	41.2	56.2	31.7	24.5
HDAg-L1$_{pep}$	A	57.5	37.4	20.1	–	–	–	76.0	45.3	30.6
	B	75.4	46.5	29.0	–	–	–	72.8	46.7	26.1
HDAg-L2$_{pep}$	A	42.3	28.9	13.4	64.8	28.1	36.7	73.7	38.4	35.3
	B	49.6	34.8	14.8	55.8	26.1	29.8	75.5	39.7	35.9

Table S2: Crystallographic data collection and refinement of NTD co-crystallised with AmphCBMlong$_{pep}$. AmphCBMlong$_{pep}$ corresponds in sequence to residues 349–360 of human amphiphysin I (UniProt P49418). Values for the highest resolution shell are shown in parentheses.

	NTD:AmphCBMlong$_{pep}$
Data collection	
Space group	*C*2
Cell dimensions	
a, *b*, *c* (Å)	138.1, 131.2, 77.8
α, β, γ (°)	90.0, 115.4, 90.0
Resolution (Å)	39.7–1.8 (1.89–1.84)
R_{merge}	0.084 (1.232)
< I/σI>	12.8 (1.5)
$CC_{1/2}$	0.999 (0.528)
Completeness (%)	99.9 (99.5)
Redundancy	6.8 (6.2)
Refinement	
Resolution (Å)	39.7–1.8 (1.89–1.84)
No. of reflections (R_{work}/R_{free})	102,809/5312
R_{work}/R_{free}	0.162/0.182
Ramachandran favored regions (%)	98.4
Ramachandran outliers (%)	0.0
No. of atoms	
Protein	5772
Glycerol	6
Peptide ligands	336
Water	785
B-factors	
Protein	31.9
Glycerol	27.1
Peptide ligands	57.9
Water	46.4
r.m.s. deviations	
Bond lengths (Å)	0.012
Bond angles (°)	1.549
PDB ID	5M61

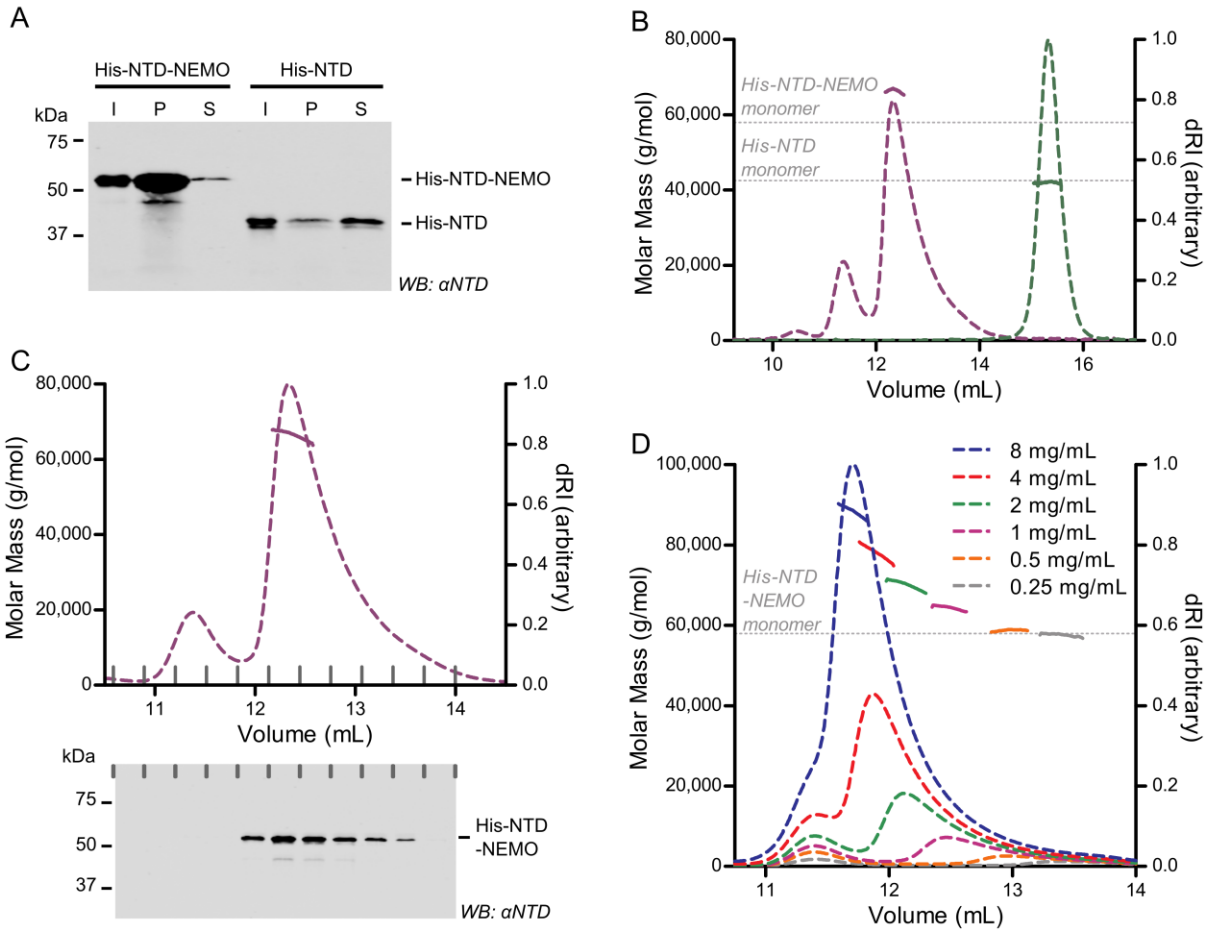


Figure S1: His-NTD-NEMO forms oligomers and is captured more efficiently than His-NTD by GST-tagged clathrin-binding peptides (A) Capture ("GST pull-down") of His-NTD-NEMO and His-NTD by GST-AP2CBM. His-NTD was produced by taking advantage of a proteolytic cleavage event that occurred during the purification of wild-type and mutant His-NTD-NEMO. During gel filtration chromatography a significant amount of protein eluted at a volume consistent with it lacking the NEMO domain. As this protein reacted with an anti-NTD antibody and its experimental mass was as expected for His-NTD alone (B) we assumed it to be His-NTD alone, the NEMO oligomerisation domain having been liberated by proteolysis during the bacterial expression or subsequent lysis and affinity purification. Glutathione sepharose beads loaded with GST-AP2CBM bait protein were incubated with His-NTD-NEMO or His-NTD, washed, and the beads were collected. The input NTD samples (I), supernatant following prey incubation (S) and bound protein sample following washing (P) were subjected to SDS-PAGE and immunoblotting (WB) using an antibody that recognizes clathrin NTD (αNTD). His-NTD-NEMO is more readily captured than His-NTD in this assay. (B) Determination of the mass of His-NTD (green) and His-NTD-NEMO (purple) by size-exclusion chromatography with inline multi-angle light scattering (SEC-MALS). The elution profiles of each protein, monitored using the solvent differential refractive index (dRI), are shown as dashed curves. Weight-averaged molar masses, determined directly from the dRI and light scattering of the samples, are shown as solid lines across the elution profiles. The expected molar masses for a His-NTD monomer and a His-NTD-NEMO monomer are shown as dotted grey lines. (C) Immunoblot analysis of the two main elution peaks observed during SEC-MALS of His-NTD-NEMO. The weight-averaged molar mass is shown as a line across the elution profile (upper panel, purple lines). Fractions collected throughout the experiment (grey ticks on horizontal axis) were subjected to SDS-PAGE and immunoblotting as in (A). Only the larger peak, eluting between 12–14 mL, is recognised by the anti-clathrin NTD antibody and thus the smaller peak, eluting between 11–12 mL, is presumed to be a co-purified contaminant. (D) Concentration-dependent oligomerisation of His-NTD-NEMO. SEC-MALS was performed using His-NTD-NEMO injected at six different concentrations. While the weight-averaged molar mass of the eluted protein (solid lines) matches the expected molar mass of monomeric His-NTD-NEMO (grey dotted line) at low concentrations of injected His-NTD-NEMO, the protein exhibits lower elution volumes (monitored using dRI, dashed line) and increased weight-averaged molar mass when the concentration of injected protein increases. This is consistent with homo-oligomerisation of His-NTD-NEMO, presumably mediated by the NEMO oligomerisation domain.

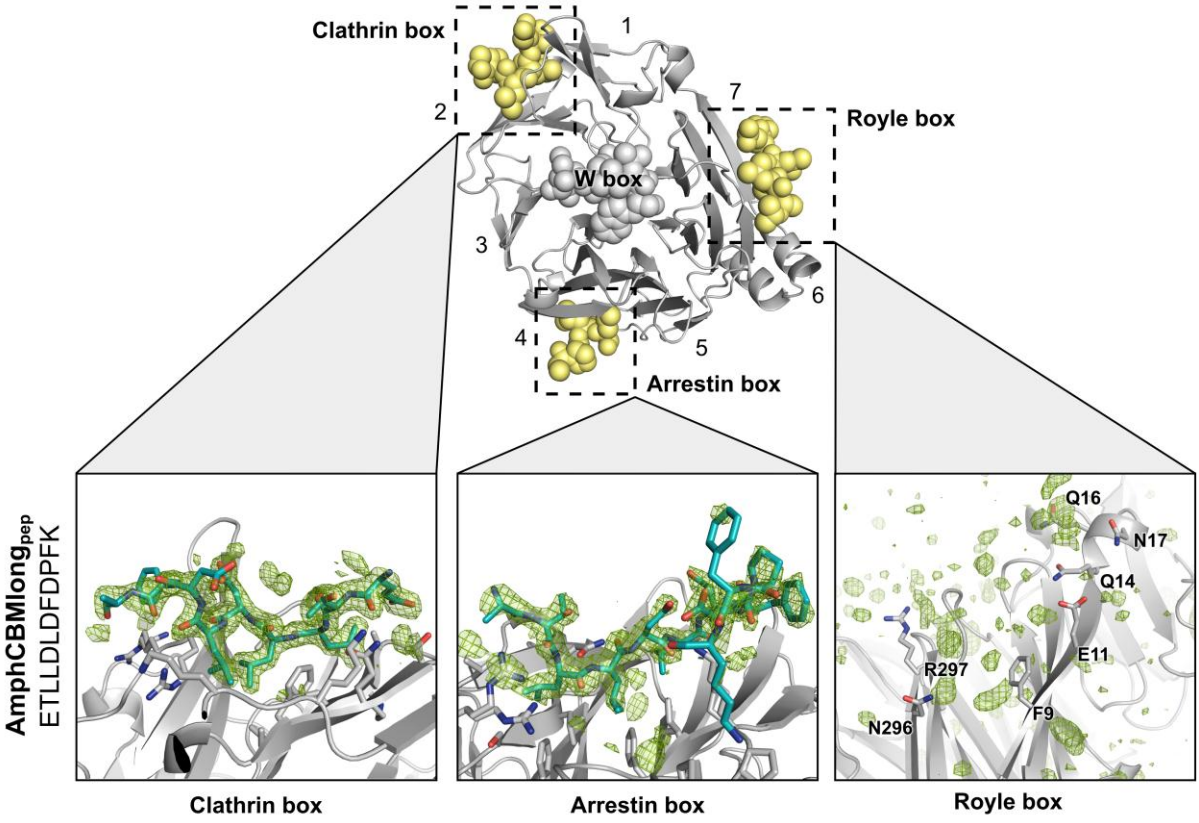


Figure S2: AmphCBMlong$_{pep}$ binds NTD at the clathrin and arrestin boxes but not the Royle box. The β-propeller fold of clathrin NTD (grey ribbons) is shown with numbers enumerating the seven β-stranded blades. Spheres represent peptides bound at the four peptide-interaction sites on NTD. Insets show unbiased F_O-F_C electron density (3 σ), calculated before the addition of peptide residues to the structural model, that is consistent with binding of AmphCBMlong$_{pep}$ at the clathrin and arrestin boxes but not at the Royle box. The final refined model of AmphCBMlong$_{pep}$ (sticks, carbon atoms cyan) bound at the clathrin and arrestin sites is shown with selected NTD side chain atoms also displayed (sticks, carbon atoms grey).

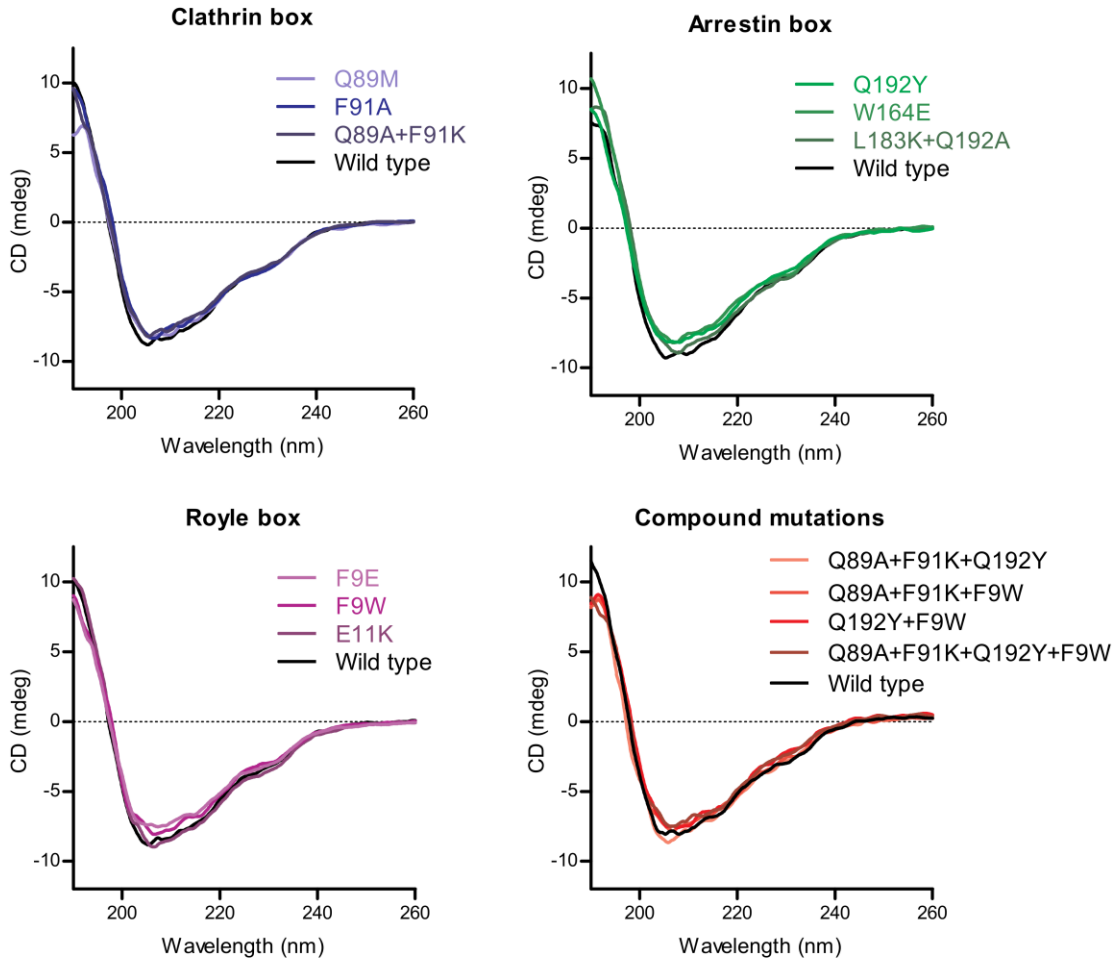


Figure S3: Circular dichroism of wild-type and mutant clathrin NTD. Circular dichroism (CD) spectra of wild-type (black) or mutant (coloured) His-NTD. The spectra are consistent with His-NTD having a predominantly β-sheet composition, as expected from the clathrin NTD crystal structure. None of the His-NTD mutants have significantly different CD spectra, consistent with them all having secondary structure content similar to wild-type His-NTD.