The structural basis for the binding of tryptophan-based motifs by δ-COP

Richard J. Suckling\textsuperscript{a,e,1}, Pak P. Poon\textsuperscript{b}, Sophie M. Travis\textsuperscript{c}, Irina V. Majoul\textsuperscript{d}, Frederick M. Hughson\textsuperscript{c}, Philip R. Evans\textsuperscript{a}, Rainer Duden\textsuperscript{d,2}, and David J. Owen\textsuperscript{e,2}

\textsuperscript{a} MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK
\textsuperscript{b} Department of Microbiology and Immunology, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4R2
\textsuperscript{c} Department of Molecular Biology, Princeton University, Princeton NJ 08544, USA
\textsuperscript{d} Institute of Biology, Center for Structural and Cell Biology in Medicine, University of Lübeck, Germany
\textsuperscript{e} CIMR, University of Cambridge, Cambridge Biomedical Campus, WT/MRC Building, Hills Road, Cambridge CB2 0XY, UK

\textsuperscript{1} Current address: Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK

\textsuperscript{2} Correspondence should be addressed to DJO djo30@cam.ac.uk +44 1223 762643 and RD duden@bio.uni-luebeck.de +49 451 500 4116

Keywords:
δ-COP \textsuperscript{1}\textsuperscript{-}homology domain binding motifs
Coatamer
Vesicle coat
Membrane trafficking
COPI
Abstract

Coatomer consists of two subcomplexes: the membrane-targeting, Arf1:GTP binding $\beta\gamma\delta\varepsilon$-COP F-subcomplex, which is related to the AP clathrin adaptors, and the cargo binding $\alpha\beta'\epsilon$-COP B-subcomplex. We present the structure of the C-terminal $\mu$-homology domain of the yeast $\delta$-COP subunit in complex with the WxW motif from its binding partner, the ER-localised Dsl1 tether. The motif binds at a site distinct from that used by the homologous AP $\mu$ subunits to bind Yxx$\Phi$ cargo motifs with its two tryptophan residues sitting in compatible pockets. We also show that the *Saccharomyces cerevisiae* ArfGAP homologue Gcs1p uses a related WxxF motif at its very C-terminus to bind to $\delta$-COP at the same site in the same way. Mutations designed on the basis of the structure in conjunction with ITC confirm the mode of binding and show that mammalian $\delta$-COP binds related tryptophan-based motifs such as that from ArfGAP1 in a similar manner. We conclude that $\delta$-COP subunits bind Wx$_{n[1-6]}$[WF] motifs within unstructured regions of proteins that influence the lifecycle of COPI-coated vesicles, which is supported by the observation that in the context of a sensitising domain deletion in Dsl1p, mutating the tryptophan-based motif binding site in yeast causes defects in both growth and CPY trafficking/processing.
Significance statement:

Eukaryotic cells contain many different organelles between which vesicles traffic material. COPI-coated vesicles mediate essential, evolutionarily conserved retrograde trafficking pathways from the Golgi to the ER and within the Golgi. We have determined structures of the \( \mu \)-homology domain (MHD) of the COPI \( \delta \)-COP subunit in complex with tryptophan-based motifs from the ER-resident vesicle-docking/tethering complex Dsl1. This is the first mechanistic description of a tether/coat interaction; we furthermore demonstrate that this interaction plays a role in facilitating COPI-coated vesicle transport \textit{in vivo}. Our works means that there is now a structure of each eukaryotic cell MHD and shows that MHDs are adaptable scaffolds, which can interact specifically with a range of proteins and phospholipids at different positions.
Introduction

COPI vesicles mediate retrograde trafficking from the Golgi to the ER and within the Golgi (reviewed in (1-3)). The COPI vesicle coat consists of two major components, coatamer and the small GTPase ADP-ribosylation factor 1 (Arf1). Coatamer is a ~600 kDa heteroheptameric complex consisting of two linked sub-complexes, the βγδζ-COP F-subcomplex and the αβε-COP B-subcomplex, all conserved from yeast to humans.

Recent studies have led to a model for COPI-coated vesicle formation in which the F-subcomplex functions as a traditional Arf1:GTP effector but with two membrane attached Arf1:GTP molecules binding to quasi-equivalent sites on a single F-subcomplex, recruiting F-subcomplex and its associated B-subcomplex en bloc onto the membrane (4). Once the vesicle has budded from its donor membrane, an Arf GTPase activating protein (ArfGAP) catalyses the hydrolysis of GTP on Arf1 and the Arf1:GDP dissociates from the membrane, followed later by the dissociation of the coatamer complex that was bound to the transport vesicle (5). ArfGAPs have also been proposed to function at different stages in vesicle biogenesis, as both terminators and, during an earlier ‘cargo editing’ step, as effectors (reviewed in depth in (6)). In the final stages of its life, a COPI-coated vesicle docks with the ER via the Dsl1 tethering complex, completes uncoating and, finally, undergoes SNARE-mediated fusion with the ER (7-9).

The two main ArfGAPs associated with COPI-dependent retrograde transport in yeast are Gcs1p and Glo3p. These proteins can at least partially substitute for one another in COPI-vesicle mediated transport, although Glo3p, which binds the γ-COP appendage domain, has been proposed to play the more significant role (10-12). Cassel and coworkers have demonstrated, for mammalian ArfGAP1, that a di-tryptophan motif (WxxxxW) located within the unstructured C-terminus of the protein can bind to the mammalian δ-COP C-terminal μ-homology domain (MHD) (13). Intriguingly this tryptophan motif resembles a proposed yeast δ-COP binding motif, termed δL, which was identified by yeast-2-hybrid screening and then discovered in a cytosolic protein which at that time was of uncertain function in COPI vesicle trafficking (14). Subsequently similar motifs were identified
in the central low complexity region (the so-called ‘lasso’) of the Dsl1p subunit of the yeast Dsl1 tethering complex (7-9, 15). Here we reveal the structure of S. cerevisiae δ-COP MHD and the basis for its binding to di-tryptophan-based motifs from Dsl1p, and also demonstrate that a related WxxF motif from the very C-terminus of Gcs1p binds at the same site on δ-COP. The functional relevance of this binding site on δ-COP is indicated by the observation that mutations in the site that abolish motif binding can affect CPY trafficking/maturation and cause growth phenotypes.
Results

Structure of δ-COP MHD in complex with Dsl1p WxW motif

Isothermal titration calorimetry (ITC) was used to test the ability of bacterially expressed yeast δ-COP MHD (residues 282-546, see Figure 1A) to bind di-tryptophan peptides derived from the yeast Dsl1p lasso (residues 378-488). Low micromolar K_D’s (~10 µM) were observed for the binding of δ-COP MHD to peptides with spacing WxW and WxxxW (Figure 1B). Di-tryptophan peptides with 4, 5 or 6 residues between the tryptophans also bound, with K_D’s between 10 and 50 µM (Figure S1). These results suggest that the presence of two relatively closely spaced tryptophan side chains is the main determinant for binding.

Yeast δ-COP MHD was co-crystallised with the strongest binding peptide, corresponding to the WxW motif in the Dsl1p lasso (DDWNWE(SeMet)ED) (N.b. the valine in the native sequence was substituted with a selenomethionine to aid identification of the peptide orientation – this peptide bound with a similar affinity to a native sequence peptide (Figure S1)). The structure was determined at 2.8 Å resolution by selenomethionine multiwavelength anomalous dispersion methods (MAD) (Table S1, Figure S2 and Materials and Methods for detailed description of structure determination).

As predicted, yeast δ-COP MHD resembled µ1-4 adaptins and Syp1p, being constructed of 17 β-strands arranged in two inter-linked β-sandwich subdomains (16-20) (Figure 1C). One striking feature of the δ-COP MHD structure is its large negative electrostatic surface potential, which is conserved across species and is therefore presumably functionally important (Figure 1D). This global feature is in marked contrast to the MHDs of AP1 and especially AP2, which are extremely positively charged (Figure 1D). The role of these positively charged patches in AP MHDs is to mediate binding to negatively charged, organelle-specific PIPs: PtdIns(4,5)P_2 in the case of AP2 at the plasma membrane and presumably PtdIns(4)P in the case of the trans-Golgi network (TGN) (21-23). By analogy the δ-COP MHD could also be involved in specificity of COPI membrane recruitment and/or of COPI-coated vesicle docking (see Discussion).
All nine residues of the Dsl1p lasso peptide were clearly visible in electron density maps, binding at a site comprising δ-COP MHD strands 4 and 6 and the loop between strands 5 and 6. The identities of the residues and inter-strand loop length at this site are well conserved from yeast to mammals (see figure 1D and S3). The WxW peptide binding occurs at a position adjacent to and slightly overlapping with the binding site of the YKFFE sequence of APP to µ4 MHD (17), but at a completely different site and in fact on an orthogonal face from the binding of YxxΦ signals to AP MHDs (16) (Figure 1C).

The WxW peptide binds in an extended conformation with the three hydrophobic residues binding into three complementary hydrophobic pockets (Figure 2A, figure S2). The N-terminal tryptophan of the WxW motif in the Dsl1p lasso, Trp-413, is deeply buried in ‘pocket 1’ (Figure 2B). Structure-based superposition shows that pocket 1 in δ-COP is equivalent to the pocket in µ4 in which Phe-690 (YKFFE) in APP is buried (Figure S4) (17). The C-terminal tryptophan, Trp-415 is deeply buried in a hydrophobic pocket (‘pocket 2’) created by the loop between strands 5 and 6, which is significantly longer than in other MHDs (15 residues vs. for example 4 in µ2). The selenomethionine in the DDWNWE(SeMet)ED peptide (valine in the native sequence) is buried in the third hydrophobic pocket (‘pocket 3’), which could be considered as an extension of pocket 2. Aspartate residues, often found adjacent to the tryptophan residues in δ-COP MHD binding sequences, may serve to funnel ligands towards the binding site, which is the only significant non-negatively charged patch on the otherwise negatively charged δ-COP MHD surface (Figure 1D).

We used this structure as a molecular replacement model to solve at 1.8 Å resolution a different (and merohedrally twinned) crystal form of yeast δ-COP MHD in complex with the equivalent non-SeMet peptide DDWNWEVED (Table S2). The structures were essentially identical, with the peptides bound in the same way (Figure S2). We were also able to solve another co-crystal structure of yeast δ-COP MHD in complex with a peptide corresponding to the W456xxxW459 motif in Dsl1p (Table S2). In this structure, the motif also bound at the same site and most likely in the same orientation, although low occupancy of the peptide prevented detailed analysis of the binding (Figure S2). A final crystal form of yeast δ-COP MHD, crystallised without peptide, was determined
by molecular replacement (Table S2). This structure showed that the binding of peptide had no discernable effect on the structure (Figure S5); i.e., binding occurs to an effectively rigid template.

**Mutagenesis of the tryptophan-based motif binding site on δ-COP.**

Most (8 of 12) of the *S. cerevisiae* δ-COP residues that contact the di-tryptophan peptide motif, including His-350 and Arg-384, are conserved across species (Figure S3). We were able to validate the peptide binding mode observed in the crystal structure experimentally by using a double point mutant (H350A, R384S) that, while fully folded as determined by circular dichroism and gel filtration, eliminates the cation-π interactions between His-350 and Arg-384 and the indole rings of Trp-415 and Trp-413, respectively. As predicted, the double point mutant no longer bound the WxW peptide (*K_D > 300 μM*, Figure 2C). In addition, mammalian δ-COP MHD bearing mutations corresponding to yeast H350A, R384S (H330A K363S) no longer bound to the WxxxxW binding peptide from its ligand ArfGAP1, confirming that the mode of di-tryptophan motif binding is conserved from yeast to mammals (Figure S1).

**Studying the binding of tryptophan based motifs to δ-COP *in vivo*.**

The yeast δ-COP H350A, R384S (HARS) mutant provided us with a tool for assessing the functional role of the δ-COP:Dsl1p lasso interaction *in vivo*. When wild-type δ-COP was replaced by the HARS double mutant to abolish the δ-COP:Dsl1p interaction, yeast grew normally. This result is not unexpected based on the previous finding that the Dsl1p lasso is functionally redundant with another region of Dsl1p, the highly-conserved C-terminal E domain (9, 24), for COPI-mediated transport: yeast containing either a deletion of the entire lasso region or a deletion of the E-domain grow normally (24). Strikingly, yeast lacking the Dsl1p E domain – and therefore rendered dependent upon the Dsl1 lasso for COPI-dependent transport and hence viability – displayed marked growth rate defects when lasso binding was compromised by the δ-COP HARS mutations (Figure 3A). This growth defect was accompanied by a major defect in carboxypeptidase Y (CPY) trafficking/maturation at 37°C, and to a lesser degree at 34°C (Figure 3B), suggesting that δ-COP
MHD is involved in an interaction important to maintain correct vesicular trafficking between Golgi and ER.

**Structure of δ-COP µ-homology domain with a WxxF motif from yeast Gcs1p:**

Curiously, mammalian and yeast δ-COP MHDs are thought to bind different di-tryptophan-containing ligands. In mammals, ArfGAP1 but not the Dsl1 (also called NRZ) complex is proposed as the functionally relevant δ-COP MHD ligand; indeed, the mammalian Dsl1p homologue ZW10 lacks a tryptophan-rich lasso. In yeast, Dsl1p but not the ArfGAP1 orthologue Gcs1p has been proposed as the δ-COP MHD ligand. In light of our results, we re-examined the Gcs1p sequence from *S. cerevisiae* and found a WxxF at the very C-terminus that is conserved in most yeasts (Figure 4A).

A WxxF peptide from Gcs1p (residues 345-352; DEDKWDDF) bound to δ-COP MHD with an affinity similar (K\(_D\) = ~10 µM) to that of di-tryptophan peptides (Figure 4C). Moreover, a complex of δ-COP MHD and the same WxxF peptide crystallised in a new crystal form that diffracted to 2.5 Å (Table S1). Electron density representing D\(_{347}\)K\(_{W}\)DDF\(_{352}\) residues in the peptide was clearly visible, showing that the peptide binds at the same site and in the same orientation as the Dsl1p WxW peptide (Figure 4B, figure S2). As expected, the C-terminal Phe residue occupies pocket 2 and the (H350A, R384S) mutation in δ-COP MHD abrogates WxxF peptide binding (Figure 4C). The WxxF peptide is, however, slightly displaced relative to the di-tryptophan peptides (Figure 4D), partly to accommodate the different number of intervening residues but also to allow the formation of a salt bridge between the C-terminus of the peptide and the side chain of Lys-371 (Figure 4B). In the absence of Glo3p, cells expressing Gcs1p with its WxxF mutated grew more slowly than cells expressing wild type Gcs1p, implying an *in vivo* role for the WxxF motif (Figure S6A); however, CPY sorting was compromised to a similar extent when either WT or mutant Gcs1p with its terminal WDDF mutated to ADDA were expressed in the absence of Glo3p (Figure S6B).
**Discussion:**

We have described the structure of the *S. cerevisiae* δ-COP MHD and the basis for its binding to di-tryptophan motifs found in Dsl1p (WxW and WxxW; first reported in (7)) and to a newly identified WxxF motif found at the C-terminus of the ArfGAP protein Gcs1p. Interestingly in the fission yeast *S. pombe* the δ-COP gene is actually missing its MHD altogether and this is the one yeast that lacks a WxxF sequence at the very C-terminus of its Gcs1p homologue (Figure 4A) and also lacks a Dsl1p homologue (25), perhaps reflecting the loss of the evolutionary pressure to maintain one half of an interacting pair of sites when the other is absent.

Unlike AP MHD ligands, neither Gcs1p nor Dsl1p are transmembrane proteins but are vesicle coat assembly/disassembly factors, indicating a fundamentally different role for δ-COP MHD as compared with transmembrane cargo sorting AP MHDs. This observation is in line with the fact that a third yeast protein proposed as a tryptophan-based motif containing δ-COP MHD ligand originally termed δL, and now named Cex1p (14), has a mammalian homologue Scyl1, that has now been shown to be a cytosolic protein implicated in Golgi to ER transport (26). Since binding of similar strength can be detected to WxxxxW, WxxxxxW and WxxxxxxW constructs (Figure S1; note however that the last of these binds less tightly, with $K_D \approx 50 \mu M$), the δ-COP MHD binding motif can be redefined as W$_{n(1-6)}$[WF], which may aid in identifying new accessory factors for COPI-coated vesicle formation/disassembly. Our *in vivo* data indicate that growth defects and CPY trafficking/processing defects can be brought about by disrupting δ-COP MHD W$_{n(1-6)}$[WF] motif binding but only in a sensitised system; e.g., deletion of Dsl1p E-domain or absence of Glo3p. The need for multiple mutations to reveal phenotypic alterations likely reflects high levels of redundancy in this important cellular transport route such as the binding of the other ArfGAP Glo3p by F-subcomplex γ-COP appendage (12) and the binding of the Dsl1p lasso by α-COP (7).

By analogy with AP2 and AP1 (21, 27, 28), membrane recruitment of the coatamer might be expected to trigger an Arf1:GTP-driven conformational change in the F-subcomplex from an inactive closed form to an ‘open’ form (4). However, the open form of F-subcomplexes attached to a
membrane in a coat has recently been demonstrated to be even more extreme than the open conformation of AP complexes (27, 28) and is now referred to as 'hyper-open' (29). The hyper open conformation of coatomer, in which δ-COP MHD is on the very outside of the coat away from the membrane, would be maintained in the vesicle coat so long as coatomer was bound to Arf1:GTP. After nucleotide hydrolysis and Arf1:GDP disengagement, the F-subcomplex would likely revert to a more closed form; however, due to the COPI B-subcomplex binding to KKxx and KxKxx transmembrane cargoes (29-31), the coat should not dissociate from the vesicle surface immediately (5). Assuming that the F-subcomplex adopts a 'closed' conformation similar to that of a cytosolic AP complex, modelling suggests that the Wx_{[1-6]}[WF] binding site on δ-COP should be accessible when the F-subcomplex is closed (Figure S7). This situation contrasts with the unavailability of YxxΦ motif binding site on the μ subunits of closed AP complexes (21, 32). EM tomography suggests that δ-COP MHD Wx_{[1-6]}[WF] binding site remains accessible in the 'hyper open' membrane-associated conformation, indicating it possesses no switching mechanism to regulate ligand binding. This likely reflects the different roles of the MHDs in COPI and AP complexes. APs must bind YxxΦ motifs only when the APs are on the membrane so they do not bind erroneously to YxxΦ motifs in cytosolic proteins. By contrast, COPI needs to bind cytosolic, vesicle-coat assembly/disassembly factors throughout a COPI-coated vesicle’s life cycle: Gcs1p during the early stages of COPI-coated vesicle assembly for a possible role in cargo selection editing, Cex1p at an undefined stage, and Dsl1p in the latter stages for docking and ultimately fusion with target membranes.

The remarkable electronegativity of δ-COP MHD surfaces could, by analogy to the homologous AP MHD surfaces that target APs to PIP-rich and hence negatively charged membranes, function in membrane recruitment specificity. However, since the cis- and medial Golgi membranes are not positively charged but rather much less negatively charged than the late Golgi, such a role would need to be achieved through inhibiting recruitment to the more negatively charged membranes of the late-Golgi and TGN. The reduction in negative charge of the cis- and medial Golgi relative to the TGN and late Golgi is the result of two factors. Firstly, the negatively charged
PtdIns(4)P is concentrated in the later Golgi due to the constant cycling of the PtdIns(4)P phosphatase Sac1 between the cis-Golgi and ER (33, 34). Secondly, phosphatidyl serine, another major negatively charged phospholipid, is mostly located only in the inner leaflet of the ER and cis-Golgi membranes whereas it is found more in the outer leaflet of compartment membranes later in the secretory pathway (35). Interestingly both β'-COP (30) and α-COP (31) juxtamembrane surfaces are similarly highly electronegative, and could thus also contribute to selective membrane recruitment of coatomer.

However, in a recent ground-breaking electron tomographic reconstruction of COPI-coated vesicles (29), Dodonova et al.’s model suggested that δ-COP MHD is actually on the outside of coat, being the furthest part of the COPI coat from the membrane surface. One intriguing possibility is therefore that the high global negative charge of δ-COP may have a role in contributing to specificity of target membrane fusion of COPI-coated vesicles in that δ-COP could inhibit close apposition and hence fusion with more negatively charged membranes such as the late Golgi/TGN, but not inhibit contact with the comparatively uncharged membranes of the cis-Golgi and ER. Such a mechanism could consequently impart retrograde directionality to COPI vesicle transport through the Golgi towards the ER and may also involve α-COPCTD/ε-COP subcomplex, which is similarly both highly negatively charged (36, 37) and located on the outside of the COPI coat (29). It should be noted that for such a scenario to operate, a significant proportion of coatomer must indeed stay associated with vesicles right up to their docking step to allow Dsl1 complex-mediated tethering of COPI-coated vesicles to occur (5, 15).

Finally, comparing the structures of MHDs of the μ subunits in AP complexes and the δ-COP subunit of coatomer indicates that the Wx\textsubscript{m(1-6)}[WF] motif binding is unique to δ-COP since both the residues and length of loops involved in Wx\textsubscript{m(1-6)}[WF] motif binding in δ-COP are not significantly conserved in the MHDs of AP complexes or TSET (38). Further, the proteins containing those motifs are cytosolic coat assembly/disassembly factors rather than membrane embedded cargo and as such δ-COP MHDs may play a role more analogous to that of the MHDs of muniscin family.
members such as Syp1p and its mammalian homologues FCHO1 and FCHO2 (39) than to that the MHDs of AP complexes in CCV formation.
**Materials and Methods:**

**Protein expression and purification:**

δ-COP μ-homology domain (282-546) from *S. cerevisiae* was expressed in BL21(DE3)plysS (for native protein), or in B834(DE3)plysS (for selenomethionine-substituted protein) for 18-20 hr at 20°C after induction with 0.5 mM IPTG as a GST-fusion protein. GST-δ-COP-MHD was affinity purified using glutathione sepharose. Following washing with 20 mM HEPES pH 7.5, 700 mM NaCl, and 5 mM β-ME, δ-COP MHD was eluted through overnight cleavage of the GST-fusion protein with Prescission protease at 4°C. δ-COP-MHD was further purified by size-exclusion chromatography (SEC) using a Superdex SD75 preparative column in 20 mM HEPES pH 7.5, 200 mM NaCl, and 5 mM DTT.

**Isothermal titration calorimetry:**

All ITC experiments were carried out at 20 °C, with proteins and peptides in 50 mM HEPES pH 7.5, 200 mM NaCl, 5 mM β-ME with δ-COP MHD at 100 μM. All experiments were carried out at least three times, with appropriate standard deviations reported.

**Structure determination of δ-COP MHD with Dsl1p WxW and Gcs1p WxF peptides:**

The Dsl1p selenomethionine-substituted WxW peptide (DDWNWE(SeMet)ED) was mixed with selenomethionine-substituted δ-COP (282-546) W404A (30 mg/ml) at 2:1 molar ratio, and crystallised in in 0.1 M HEPES-Na pH 7.0, 0.15 M (NH4)2SO4, 21% (w/v) PEG-3350. Crystals were cryoprotected with 15% glycerol and a three-wavelength MAD dataset was collected at the selenium edge. The best crystal diffracted to 2.8 Å and belonged to space group P1 and the data processed and structure solved as described in Supplementary information and Tables S1 and S2..

The Gcs1p WxF peptide (DEDKWDDDF) was mixed with selenomethionine-substituted δ-COP (282-546) W404A (20 mg/ml) at 2:1 molar ratio, and crystallised in 0.1 M Tris-HCl pH 8.5, 0.15 M magnesium chloride, 26% (w/v) PEG-4000. Crystals were cryoprotected with 15% glycerol and data were collected to 2.5 Å (space group P2₁2₁2₁). The structure was solved by molecular replacement using one molecule of the previously determined δ-COP MHD structure (in complex with the DDWNWE(SeMet)ED peptide).
Yeast Methods:

Heterozygous DSL1/dsl1\textsuperscript{H701Stop}::natMX (24) was transformed with a linearized plasmid containing \textit{ret2}\textsuperscript{H350A,R384S}::URA3. Diploid cells were sporulated and dissected to obtain the double mutant strain.

Growth assays were performed as previously described (8). CPY secretion assays and immunoblots testing CPY trafficking/processing were performed as previously described (40, 41), using rabbit $\alpha$-CPY antibodies generously provided by Karin Römisch (Saarland University, Germany). The following yeast mutants were used as positive controls to assay for trafficking phenotypes: \textit{sec18-1} for p1-CPY accumulation (i.e. an ER block); \textit{vps35}Δ for possible vacuolar protein sorting defects (i.e. p2-CPY accumulation and loss of mCPY).

Acknowledgements:

We should like to thank the beamline scientists at the Diamond Light Source and Mike Lewis (MRC LMB), Gerry Johnston (Dalhousie University), and Mark Rose (Princeton University) for helpful discussions and technical advice. RJS and DJO are funded by a Wellcome Trust fellowship to DJO (090909). PPP was funded by Canadian Institute of Health Research. RD acknowledges support from the DFG Excellence Cluster “Inflammation and Interfaces” (ECX306) and the University of Lübeck. SMT and FMH acknowledge support from NIH (GM071574). PRE is funded by MRC grant U105178845.
References:

Figure 1 – Structure of yeast δ-COP MHD with yeast Dsl1p WxWxφ motif

(A) Schematic representation of interactions between δ-COP and Dsl1p.

(B) ITC experiments showed that yeast δ-COP MHD (282-546) binds to di-tryptophan peptides corresponding to the WxW (DDW413NW415EVED) and WxxxW (ENAW455DEAW459AIDEC) motifs in yeast Dsl1p with K_D’s of 11 ± 2 and 16 ± 1 µM (SD, N=3) respectively at a 1:1 stoichiometric ratio.

(C) Crystal structure of yeast δ-COP MHD with a selenomethionine substituted Dsl1p WxW peptide (DDWNWE(SeMet)ED) showing that di-tryptophan motifs bind to δ-COP at a position similar to the binding of the YKFFE of APP to μ4 MHD (3L81) (17), but at a completely different site from the binding of Yxxφ signals to AP MHDs (1BXX) (16, 18, 19). Final refined 2mF_o-DF_c density contoured at 1.2σ, and mF_o-DF_c difference density at ±3.0σ for the peptide is shown inset. The total buried area of the interface between δ-COP and the WxWx(SeMet) peptide is 1354 Å², as calculated by the PISA server (http://www.ebi.ac.uk/pdbe/pisa/).

(D) The Consurf server (http://consurf.tau.ac.il/) was used to create a surface representation of evolutionary conservation of residues in δ-COP MHD (D), based on an alignment from yeast to humans using ClustalO (see also figure S3). The tryptophan-based motif binding site is the outstanding feature of this conservation surface representation. Surface representations of the electrostatic potential of δ-COP MHD and μ2 MHD (E), highlights that δ-COP MHD is largely negative (contoured from -0.5V (red) to +0.5V (blue)). The di-tryptophan motif binding site on δ-COP
MHD is relatively positive compared to the remainder of the domain. The conservation surface representation shows that the negatively charged surface on the backside of subdomain B is also highly conserved.
Figure 2 – Structural details of Dsl1p WxW peptide bound to δ-COP MHD

(A) Dsl1p WxWx(SeMet) peptide (DDWNWE(SeMet)ED) binds to δ-COP MHD in an extended conformation with the three hydrophobic residues binding into three complementary pockets (labelled).

(B) Surface representation of the di-tryptophan motif binding site in δ-COP MHD coloured from high (dark green) to low (white) hydrophobicity. The three hydrophobic pockets into which the hydrophobic residues in the peptide bind can be clearly seen.

(C) ITC showing that a combination of H350A and R384S mutations (highlighted in dark red in A and B) abolish ($K_D > 300 \, \mu\text{M}$) (red triangles) the interaction of the Dsl1p WxW peptide (DDWNWEVED) with δ-COP MHD (black squares) ($K_D = 11 \pm 2 \, \mu\text{M}$) (SD, N=3). Data for mutant is offset for clarity.
Figure 3 - *ret2*(HARS)/*dsl1*-Δ*E* mutant cells show defects in growth and in CPY trafficking.

(A) Isogenic strains bearing mutations at either the *DSL1* or *RET2* (δ-COP) gene loci, labelled as *dsl1*-Δ*E* and *ret2*(HARS), were serially diluted, spotted and grown at the indicated temperatures for 48 hr. The double-mutant harboring both the *ret2*(HARS) and *dsl1*-Δ*E* allele was lethal at 37°C (no colonies after one week) but viable at lower temperatures, albeit having a slight growth defect. All combinations are shown in biological duplicate.

(B) Cells of strains as indicated at the top of the figure were grown to mid-log at 30°C in YEPD medium, and then shifted for 2 hr to either 34°C or 37°C. Glassbead-extracts of collected cells were separated by 7.5% SDS-PAGE and probed with α-CPY antibodies (40). Positions of p1 (ER), p2 (Golgi), and mature (vacuolar) forms of CPY are indicated. Note p1-CPY accumulation in *ret2*(HARS) *dsl1*-Δ*E* double-mutant cells at 34°C and more strongly at 37°C, but not single mutant cells.
(A) Alignment of the C-terminal region of Gcs1p from different yeasts showing that the WxxF motif at the very C-terminus is conserved in all apart from S. pombe. δ-COP in S. pombe does not contain a MHD (see schematic); lack of selective pressure means the WxxF in the Gcs1p homologue in S. pombe has been lost.

(B) Gcs1p WxxF peptide (DEDKWDDF) binds to δ-COP MHD at the same site as binding of the Dsl1p WxW peptide with Trp-349 in pocket 1 and Phe-352 in pocket 2.

(C) ITC showing that a combination of H350A and R384S mutations (highlighted in dark red in B) also abolish (K_D > 300 µM) (red triangles) the interaction of the Gcs1p WxxF peptide (DEDKWDDF) with δ-COP MHD (black squares) (K_D = 11 ± 1 µM at a 1:1 stoichiometric ratio (SD, N=3)). Data for mutant is offset for clarity.

(D) Superposition of the two δ-COP MHD tryptophan-based peptide co-crystal structures showing that the N-terminal tryptophan of both peptides binds into pocket 1, and the C-terminal tryptophan/phenylalanine into pocket 2. Superposition is within the tryptophan motif binding site (residues 348-355, 363-384). For clarity only the model of δ-COP is shown from the Gcs1p WxxF peptide co-crystal structure, and only the hydrophobic residue side chains within the peptides are shown.
Supplementary Information

The structural basis for the binding of tryptophan-based motifs by δ-COP

Richard J. Suckling\textsuperscript{a,e,1}, Pak P. Poon\textsuperscript{b}, Sophie M. Travis\textsuperscript{c}, Irina V. Majoul\textsuperscript{d}, Frederick M. Hughson\textsuperscript{c}, Philip R. Evans\textsuperscript{a}, Rainer Duden\textsuperscript{d,2}, and David J. Owen\textsuperscript{a,2}

\textsuperscript{a} MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK
\textsuperscript{b} Department of Microbiology and Immunology, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4R2
\textsuperscript{c} Department of Molecular Biology, Princeton University, Princeton NJ 08544, USA
\textsuperscript{d} Institute of Biology, Center for Structural and Cell Biology in Medicine, University of Lübeck, 23562 Lübeck, Germany
\textsuperscript{e} CIMR, University of Cambridge, Cambridge Biomedical Campus, WT/MRC Building, Hills Road, Cambridge CB2 0XY, UK

\textsuperscript{1} Current address: Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford, OX1 3RE, UK

\textsuperscript{2} Correspondence should be addressed to DJO djo30@cam.ac.uk +44 1223 762643 and RD duden@bio.uni-luebeck.de +49 451 500 4116

Keywords:
δ-COP Mu-homology domain binding motifs
Coatamer
Vesicle coat
Membrane trafficking
COPI
Figure S1 - δ-COP MHD binds to W(x)_nW motifs where n = 4, 5 and 6

(A) ITC experiments showed that yeast δ-COP MHD (282-546) (W404A) binds to W(x)_nW peptides where n = 4, 5 and 6 with K_D's of 10, 21 ± 2 and 38 ± 2 µM (SD, N=3) respectively at a 1:1 stoichiometric ratio.

(B) ITC showing that δ-COP MHD (282-546) (W404A) binds to a peptide (DDWNWE(SeMet)ED) corresponding to the WxW motif in the Dsl1p lasso, in which the valine in the native sequence was substituted with a seleno-methionine to aid identification of the peptide orientation. The peptide bound with a similar affinity to the native peptide (Figure 1B) (K_D = 4 ± 1 µM vs. 11 ± 2 µM (SD, N=3)) at a 1:1 stoichiometric ratio. The slightly higher affinity of the seleno-methionine derived...
peptide is presumably due to additional van der Waals interactions of the methionine side chain with hydrophobic pocket 3, compared to the native valine side chain (Figure S2).

(C) ITC showed that a combination of H330A and K363S mutations in human δ-COP (homologous to the H350A and R384S mutations in yeast δ-COP, see figure S3) abolishes (K_D > 300 μM) (red squares) the interaction of the human ArfGAP1 WxxxxW peptide (TDDGWDNQNW) with δ-COP MHD (W381A) (black squares). This shows that the human Gcs1p-orthologue ArfGAP1 WxxxxW peptide (TDDGWDNQNW) binds to human δ-COP MHD at the very same site as yeast Gcs1p WxxF peptide binds to yeast δ-COP MHD. Data for H330A K363S W381A mutant is offset by +0.25 μcal/sec for clarity.
Figure S2 – Structural overview of δ-COP MHD binding to tryptophan-based motifs

δ-COP MHD density

Ds1p WxWxW peptide density

Ds1p WxWxV native peptide density

Gcs1p WxF peptide density
(A-I) Electron density from δ-COP MHD in complex, with the Dsl1p WxWx(SeMet) peptide (A-C), with the Dsl1p WxxxW peptide (D-E), with the Dsl1p WxWxV native peptide (F-G), and finally with the Gcs1p WxxF peptide (H-I). Part of the experimentally phased, solvent flattened electron density map from autoSHARP (42) highlighting the clear protein-solvent boundary (A-B), and the clear density for the peptide (B-C) (1.0σ). (C) An anomalous difference map (solid) (+3.0σ, + = yellow, - = red) (clipped around peptide) shows the large positive peak representing the selenomethionine in the WxWx(SeMet) peptide. 2mFo-Fc (0.8σ) and mFo-Fc (±2.5σ) density maps representing the WxxxW peptide, before peptide model building (D) and after (E), showing that the peptide occupancy is low, and the density difficult to interpret; however the peptide most likely binds in the same orientation as the other peptides. 2mFo-Fc (0.8σ) and mFo-Fc (±3.0σ) density maps representing the native Dsl1p WxWxV peptide before peptide model building (F), and after (G), showing the clear density for the two tryptophans. 2mFo-Fc (1.0σ) and mFo-Fc (±3.0σ) density maps representing the Gcs1p WxxF peptide, before peptide model building (H) and after (I), showing the clear density for the tryptophan and phenylalanine.

Figure S3 – Conservation of the tryptophan-based motif binding site in δ-COP MHD
The Consurf server (http://consurf.tau.ac.il/) was used to create a surface representation of evolutionary conservation of residues in δ-COP MHD (also shown in figure 1D), based on an alignment from yeast to humans using ClustalO (shown here). The region that forms the tryptophan-based motif binding site in this alignment is shown here. Residues directly involved in the binding site are highlighted above the alignment in red. The tryptophan-based motif binding site is the outstanding feature of the conservation surface representation, indicating its importance. The Dsl1p WxW peptide is shown. Note conservation of the length of the loop between strands 5 and 6.
Figure S4 - δ-COP MHD binds to tryptophan-based motifs at a similar site to the binding of μ4 to the YKFFE motif

A

μ4: YKFFE peptide

δ-COP: WxW and WxF peptides

Surface representations of the YKFFE binding site in μ4 MHD (3L81) (17), and the tryptophan-based motif binding site in δ-COP MHD coloured from high (dark green) to low (white) hydrophobicity with the relevant peptides shown (A). This shows that ‘pocket 1’ (pocket between His-350 and Arg-384) in δ-COP is equivalent to the pocket in μ4 in which Phe-690 (YKFFE) in the APP motif is buried. This is evident following superposition of μ4 and δ-COP MHDs; three different angles of the superposed peptides are shown in (B).
Figure S5 – Binding of tryptophan-based motifs to δ-COP MHD through a ‘lock and key’ mechanism

Tryptophan-based motifs bind to δ-COP MHD through a ‘lock and key’ mechanism:

**Peptide-bound & unbound structures**

Superposition of δ-COP MHD bound to peptide (dark green) (of the δ-COP: Dsl1p WxWx(SeMet) peptide structure), with a molecule not bound to peptide (light green), showing that the binding of peptide has no discernable effect on the structure. The only clear difference is the different conformation of leucine 380 side chain, creating a larger hydrophobic pocket into which the second tryptophan binds (pocket 2).
Figure S6 – Compromised Gcs1p binding to δ-COP MHD impairs yeast cell growth.

(A) Yeast cells carrying simultaneous chromosomal deletions of both the ArfGAP genes GCS1 and GLO3 are unviable, unless expression of one of these ArfGAPs is driven from a plasmid. Galactose-dependent expression of Glo3p from the GAL10 promoter can keep double-mutant cells growing in media containing galactose, but not in galactose-free e.g. glucose-containing media. Individual plasmid transformants were serially diluted, spotted onto glucose-containing plates and incubated at 30°C for 48 hours. Note that the mutant Gcs1- AA protein (with the W349A, F352A substitutions) cannot sustain optimal yeast colony growth in galactose-free media, in comparison with and in contrast to wild-type Gcs1p.

(B) Cells of strains as indicated at the top (the RET2 gene encodes yeast δ-COP) were grown to mid-log phase in YEPD liquid culture at 30°C, and then shifted for 2 hr to 34°C or 37°C, resp. Glassbead-extracts of collected cells were separated by 7.5% SDS-Page and probed with anti-CPY antibodies (see Duden et al., 1994). Positions of p1(ER), p2 (Golgi), and mature (vacuolar) forms of
CPY are indicated. \textit{glo3\Delta/gcs1-WF} single- and the \textit{glo3\Delta/gcs1-AA} double-mutant cells display a strong defect in ER-to-Golgi transport of CPY at 37°C, but not 34°C. Also note the absence of a discernable difference between these mutants with regards to their CPY processing ability. Extensive attempts to demonstrate the physical interaction between coatomer and Gcs1p \textit{in vivo} have proved technically impossible, since the levels of detergent needed to reduce non-specific background binding to acceptable levels (>1% IGEPAL) abrogate the interaction \textit{in vitro}. This likely reflects that the interaction is mainly mediated by hydrophobic side chain interactions that will be readily outcompeted by detergent molecules.
Surface representation of the AP2 locked form (2VGL) (21), and docking of the δ-COP MHD (C-δ-COP)-Gcs1p WxxF peptide co-crystal structure into the AP2 locked form based on superposition of C-δ-COP with C-µ2. The Gcs1p WxxF peptide is shown as spheres. Note the significant occlusion of the Yxxφ binding site in µ2 in the AP2 structure, but the fully accessible WxxF binding site in δ-COP.
<table>
<thead>
<tr>
<th>Data Collection</th>
<th>δ-COP MHD W404A SeMet with Dsl1p WxWx(SeMet) peptide</th>
<th>δ-COP MHD W404A SeMet with Gcs1p WxxF peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beamline</td>
<td>Diamond I03</td>
<td>Diamond I04-1</td>
</tr>
<tr>
<td>Space group</td>
<td>P1</td>
<td>P2,2,2,</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.97631</td>
<td>0.97961</td>
</tr>
<tr>
<td></td>
<td>0.97949</td>
<td>0.9173</td>
</tr>
<tr>
<td>Cell dimensions (Å)</td>
<td>a, b, c (Å)</td>
<td>68.4, 79.4, 163.4</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>79.7, 87.5, 83.6</td>
<td>90.0, 90.0, 90.0</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>78-2.80 (2.85-2.80)</td>
<td>55.9-2.45 (2.55-2.45)</td>
</tr>
<tr>
<td>Rmerge&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.126 (1.208)</td>
<td>0.105 (0.990)</td>
</tr>
<tr>
<td>Rmerge&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.149 (2.340)</td>
<td>0.114 (1.086)</td>
</tr>
<tr>
<td>CC&lt;sub&gt;1/2&lt;/sub&gt;</td>
<td>0.996 (0.537)</td>
<td>0.998 (0.508)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>97.8 (97.4)</td>
<td>99.9 (99.9)</td>
</tr>
<tr>
<td>Multiplicity&lt;sup&gt;a&lt;/sup&gt;</td>
<td>7.0 (7.2)</td>
<td>6.5 (5.8)</td>
</tr>
<tr>
<td>Wilson B&lt;sup&gt;a&lt;/sup&gt;</td>
<td>86</td>
<td>55.7</td>
</tr>
<tr>
<td>Anomalous completeness (%)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>96.6 (96.6)</td>
<td>99.7 (98.4)</td>
</tr>
<tr>
<td>CC&lt;sub&gt;1/2&lt;/sub&gt; on Δ anom (inner bin)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.191 (0.000)</td>
<td>0.205 (0.021)</td>
</tr>
<tr>
<td>Phasing</td>
<td>R&lt;sub&gt;c&lt;/sub&gt;, isomorphous - 0.076</td>
<td>0.713</td>
</tr>
<tr>
<td></td>
<td>R&lt;sub&gt;c&lt;/sub&gt;, anomalous 0.945</td>
<td>0.879</td>
</tr>
<tr>
<td>Overall Phasing Power</td>
<td>- 0.470</td>
<td>0.352</td>
</tr>
<tr>
<td></td>
<td>Overall Phasing Power</td>
<td>0.525 0.387</td>
</tr>
<tr>
<td>Difference Flattening</td>
<td>0.293/0.887</td>
<td>0.205/0.258</td>
</tr>
<tr>
<td>Refinement</td>
<td>Resolution range (Å)</td>
<td>77.6-2.80</td>
</tr>
<tr>
<td></td>
<td>No. of reflections</td>
<td>76847</td>
</tr>
<tr>
<td></td>
<td>R&lt;sub&gt;temp&lt;/sub&gt;/R&lt;sub&gt;free&lt;/sub&gt;</td>
<td>0.189/0.234</td>
</tr>
<tr>
<td>Number of atoms</td>
<td>Protein 16445</td>
<td>6156</td>
</tr>
<tr>
<td></td>
<td>Ligand 696</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td>Water 554</td>
<td>257</td>
</tr>
<tr>
<td></td>
<td>B-factors (Å&lt;sup&gt;2&lt;/sup&gt;) 92.3</td>
<td>32.4</td>
</tr>
<tr>
<td>Rmsf from ideal values</td>
<td>Bond lengths (Å)</td>
<td>0.012 0.013</td>
</tr>
<tr>
<td></td>
<td>Bond angles (°) 1.526</td>
<td>1.596</td>
</tr>
<tr>
<td>Ramachandran Plot</td>
<td>Favoured region (%)</td>
<td>97.4 97.5</td>
</tr>
<tr>
<td></td>
<td>Allowed (%) 1.8</td>
<td>2.1</td>
</tr>
<tr>
<td></td>
<td>Outliers (%) 0.4</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>Rotamer outliers (%) 2.9</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td>C-beta outliers 4</td>
<td>1</td>
</tr>
</tbody>
</table>

<sup>a</sup> Values in parentheses are for the highest resolution shell. <sup>b</sup>R<sub>merge</sub> = Σ(|I<sub>h</sub> - <I<sub>h</sub>|)/Σ(I<sub>h</sub>) where <I<sub>h</sub> is the mean intensity of unique reflection h, summed over all reflections for each observed intensity I<sub>h</sub>. <sup>c</sup>R<sub>merge</sub> = Σ(n/n-1)<sup>1/2</sup>(<I<sub>h</sub>-<I<sub>h</sub>)/Σ(I<sub>h</sub>) where n is the number of observations for unique reflection h with mean intensity <I<sub>h</sub>, summed over all reflections for each observed intensity I<sub>h</sub>. <sup>d</sup>CC<sub>1/2</sub> is the correlation coefficient on <I<sub>h</sub> between random halves of the dataset. Δ<sub>anom</sub>, anomalous difference I<sub>+</sub>-I<sub>−</sub>.
Table S2 - Data collection and refinement statistics

<table>
<thead>
<tr>
<th></th>
<th>δ-COP MHD W404A native with Dsl1p WxWxV peptide</th>
<th>δ-COP MHD WT native</th>
<th>δ-COP MHD W404A SeMet with Dsl1p WxxxW peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Collection</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Merohedrally twinned</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Collection</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beamline</td>
<td>Diamond I03</td>
<td>Diamond I02</td>
<td>Diamond I04-1</td>
</tr>
<tr>
<td>Space group</td>
<td>P3,21</td>
<td>P2,2,2,1</td>
<td>P3,21</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.96863</td>
<td>1.0</td>
<td>0.91732</td>
</tr>
<tr>
<td>Cell dimensions (Å)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>72.8, 72.8, 343.3</td>
<td>86.9, 148.6, 222.7</td>
<td>72.9, 72.9, 340.9</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>90.0, 90.0, 120.0</td>
<td>90.0, 90.0, 90.0</td>
<td>90.0, 90.0, 120.0</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>36.4-1.90 (1.94-1.90)</td>
<td>48.4-3.00 (3.16-3.00)</td>
<td>85.2-2.75 (2.9-2.75)</td>
</tr>
<tr>
<td>Rmerge&lt;sub&gt;σ0&lt;/sub&gt;</td>
<td>0.129 (0.713)</td>
<td>0.098 (0.869)</td>
<td>0.117</td>
</tr>
<tr>
<td>Rmeas&lt;sub&gt;σ2&lt;/sub&gt;</td>
<td>0.137 (0.755)</td>
<td>0.115 (1.020)</td>
<td>0.219 (0.598)</td>
</tr>
<tr>
<td>CC&lt;sub&gt;1/2&lt;/sub&gt;</td>
<td>0.994 (0.643)</td>
<td>0.997 (0.448)</td>
<td>0.974 (0.577)</td>
</tr>
<tr>
<td>Mean I/σ&lt;sub&gt;I&lt;/sub&gt;&lt;sup&gt;a&lt;/sup&gt;</td>
<td>11.3 (3.4)</td>
<td>8.5 (1.7)</td>
<td>6.1 (2.8)</td>
</tr>
<tr>
<td>Completeness (%)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>98.3 (98.3)</td>
<td>99.8 (99.9)</td>
<td>97.8 (95.9)</td>
</tr>
<tr>
<td>Multiplicity&lt;sup&gt;a&lt;/sup&gt;</td>
<td>9.1 (9.2)</td>
<td>3.6 (3.6)</td>
<td>5.5 (4.7)</td>
</tr>
<tr>
<td>Wilson &lt;B&gt; (Å&lt;sup&gt;2&lt;/sup&gt;)</td>
<td>19.1</td>
<td>95.1</td>
<td></td>
</tr>
<tr>
<td>No. of twin domains</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Twin fractions (%)&lt;sup&gt;c&lt;/sup&gt; (h,k,l/-h,-k,l)</td>
<td>0.59/0.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>36.4-1.90</td>
<td>48.4-3.00</td>
<td></td>
</tr>
<tr>
<td>No. of reflections</td>
<td>78953</td>
<td>55431</td>
<td></td>
</tr>
<tr>
<td>R&lt;sub&gt;work&lt;/sub&gt;/R&lt;sub&gt;free&lt;/sub&gt;</td>
<td>0.136/0.189</td>
<td>0.195/0.233</td>
<td></td>
</tr>
<tr>
<td>Number of atoms</td>
<td>Protein</td>
<td>8217</td>
<td>16523</td>
</tr>
<tr>
<td></td>
<td>Ligand/ion</td>
<td>240</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Water</td>
<td>589</td>
<td>288</td>
</tr>
<tr>
<td>B-factors (Å&lt;sup&gt;2&lt;/sup&gt;)</td>
<td>28.0</td>
<td>92.6</td>
<td></td>
</tr>
<tr>
<td>Rmsd from ideal values</td>
<td>Bond lengths (Å)</td>
<td>0.020</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>Bond angles (°)</td>
<td>2.11</td>
<td>1.38</td>
</tr>
<tr>
<td>Ramachandran Plot</td>
<td>Favoured region (%)</td>
<td>95.9</td>
<td>95.8</td>
</tr>
<tr>
<td></td>
<td>Allowed (%)</td>
<td>3.4</td>
<td>3.4</td>
</tr>
<tr>
<td></td>
<td>Outliers (%)</td>
<td>0.7</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>Rotamer outliers (%)</td>
<td>2.1</td>
<td>2.1</td>
</tr>
<tr>
<td></td>
<td>C-beta outliers (%)</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>PDB ID code</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Values in parentheses are for the highest resolution shell. <sup>b</sup>R<sub>merge</sub> = Σ(I<sub>h</sub> - <I<sub>-h</sub>>)/Σ(I<sub>h</sub>) where <I<sub>-h</sub>> is the mean intensity of unique reflection h, summed over all reflections for each observed intensity I<sub>h</sub>. <sup>c</sup>R<sub>meas</sub> = Σ(n/n' - 1)1/2(I<sub>h</sub> - <I<sub>-h</sub>>)/Σ(I<sub>h</sub>) where n is the number of observations for unique reflection h with mean intensity <I<sub>-h</sub>>, summed over all reflections for each observed intensity I<sub>h</sub>. <sup>d</sup>CC<sub>1/2</sub> is the correlation coefficient on <I> between random halves of the dataset. Δ<sub>anom</sub>, anomalous difference I' - I.
Supplemental Extended methods:

Cloning and mutagenesis:
Initially δ-COP (282-546) from S. cerevisiae was ligated into pQE-30 (Qiagen, Hilden, Germany) using BamH I and Pst I sites. This incorporated an N-terminal 6x histidine (His6) tag. Initial crystals of the His-tagged wild-type protein were obtained; however, these could not be solved by experimental phasing or molecular replacement (these crystals were subsequently solved by molecular replacement using the δ-COP-W404A structure, see Structure determination). Subsequently δ-COP-282-546 was sub-cloned into pGEX6P.1 (GE Healthcare, Waukesha, WI, USA) using BamH I and Not I sites incorporating a N-terminal, PreScission protease (GE Healthcare) cleavable GST tag. Purification using the GST tag increased the purity of δ-COP MHD. The W404A, and non-binding mutants were created using site-directed mutagenesis by overlap PCR using complementary primers with the relevant mutation(s) across the site of interest. All constructs were sequenced by Beckman Coulter Genomics prior to use.

Protein expression and purification:
All constructs were expressed at high levels in BL21(DE3)plysS cells (Novagen, Merck KGaA, Darmstadt, Germany). Five ml starter cultures were used to inoculate one liter flasks containing 2xTY media and appropriate antibiotics. Cells were grown at 37°C to an OD₆₀₀ between 0.6-1.0, before the temperature was decreased to 20°C for 1 h. After cooling, proteins were expressed overnight at 20°C following induction with 0.5 mM IPTG. After ~20 h of expression cells were harvested and resuspended in 20 mM HEPES pH 7.5, 500 mM NaCl, 5 mM β-ME, plus EDTA-free protease inhibitor tablets (1 tablet/50 ml buffer) (Roche Diagnostics GmbH, Mannheim, Germany). Selenomethionine substituted proteins were expressed in a methionine-auxotroph E. coli strain, B834(DE3)plysS (Novagen). Cells were grown in M9 media supplemented with the following; MgSO₄ at 0.24 g/l (anhydrous), FeSO₄.7H₂O at 0.025 g/l, glucose at 0.4 % (w/v), vitamins at 0.0031 g/l (containing Riboflavin, Niacinamide, Pyridoxine monohydrochloride and thiamine at a 10:10:1:10 mass ratio), amino acids (methionine) at 0.78 g/l (Complete Supplement Mixture Drop-Out: -Met (Formedium, Hunstanton, UK)), seleno-L-methionine at 0.04 g/l (plus appropriate antibiotics). All solutions added to media were sterile filtered, whereas amino acids were added directly as powder. Selenomethionine substituted proteins were expressed as for native proteins; however the cultures grew more slowly as expected.

Cells were lysed using a disruptor (Constant Systems Ltd., Daventry, UK). Lysate containing His6-δ-COP MHD was incubated for one hour at 4°C with 50% Ni-NTA agarose (Qiagen) in buffer, while
GST-tagged proteins were incubated with 50% glutathione sepharose slurry (GE Healthcare) in buffer. Protein bound to either matrix was washed extensively with >500 ml of buffer (20 mM HEPES pH 7.5, 700 mM NaCl, 5 mM β-ME). For His₆-δ-COP-MHD, the wash buffer included 20 mM imidazole to remove contaminants that bind non-specifically to Ni-NTA beads. His₆-δ-COP-MHD was eluted in batch in buffer containing 300 mM imidazole. GST-tagged proteins were cleaved overnight with human rhinovirus 3C protease (Prescission protease (GE Healthcare)) at 4°C and eluted in batch in buffer. Following elution, all proteins were concentrated for gel filtration on a Hi Load Superdex 75 16/60 prep grade size exclusion column (GE Healthcare) pre-equilibrated in the appropriate buffer: in 20 mM HEPES pH 7.5, 200 mM NaCl, 5 mM DTT (if the protein was to be used for crystallisation), or 50 mM HEPES pH 7.5, 200 mM NaCl, 5 mM β-ME (for ITC).

The δ-COP-H350A-R384S-W404A mutant was determined to be folded as judged by circular dichroism and by identical gel filtration elution profiles as compared to W404A.

**Isothermal titration calorimetry (ITC):**

ITC experiments were conducted on a VP-ITC-200 isothermal titration calorimeter (GE Healthcare) at 20°C, using 20 injections of the sample in the syringe into the cell, with a 120 sec interval between injections. Each of these injections had a volume of 2 µl except for the first one, which only consisted of 0.5 µl. All components were in 50 mM HEPES pH 7.5, 200 mM NaCl, 5 mM β-ME. For all experiments the peptide was loaded into the syringe, and δ-COP MHD into the cell. A peptide into buffer run was subtracted from all runs to take into account the effect of peptide dilution. For all runs δ-COP MHD was at 50-150 µM; the concentration of the peptide was adjusted to ensure a significant number of points were on the binding curve, and that saturation was reached. Both protein and peptide (due to presence of tryptophans) concentrations were measured at 280 nm.

Samples in both the cell and syringe were filtered using a Proteus Clarification Mini Spin Column (Generon, Maidenhead, UK), before loading. Titration data were analysed in ORIGIN to obtain values for stoichiometry (N), equilibrium dissociation constant (Kᵤ), and enthalpy of binding. All experiments were carried out at least three times with appropriate standard deviations reported.

N.b. curves labelled as ‘WT’ in the main text were experiments done with W404A mutant. Likewise, ‘H350A R384S’ also contained the W404A mutation. W404A was removed from the labeling to avoid adding complication in the main text.

**Extended structure determination procedures**

**General data collection, data processing, and structure determination methods:**

Programs included in or affiliated with the CCP4 software package (43) were used for most steps during data processing and structure solution. The collected datasets were indexed and integrated
using Mosflm or Xia2, the Laue and point group determined with Pointless and the data scaled, merged and assessed using Scala or Aimless (43). Data were cut when the correlation coefficient (CC$_{1/2}$) between half datasets equalled 0.5. autoSHARP (42) was used to search for heavy atoms and phasing. Solvent flipping was used to determine the hand of the substructure and to improve the phases. The structures could be built using the automated model building software, Buccaneer (44). Phaser (45) was employed for all molecular replacement work. The refinement of the models was done using iterative steps of model building in Coot (46) and in Refmac5 (47). Final model validation was undertaken using MolProbity (http://molprobity.biochem.duke.edu/). Images of the refined structures were prepared with the molecular graphic programs CCP4MG. The structures could be built using automated model building software, supplemented with manual rebuilding and automated refinement.

Crystallisation of δ-COP MHD wildtype:
Initially crystallisation trials were set up with the native N-terminally His-tagged wildtype protein alone, as well as in the presence of the Dsl1p WxxxW peptide. Crystals were only obtained of the protein alone; the best crystal was obtained at a protein concentration of 10 mg/ml in 0.1 M HEPES pH 7.0, 0.05 M CaCl$_2$.2H$_2$O, 2.2% (w/v) PEG-8000, 20% (v/v) Ethylene glycol by vapour diffusion. This crystal was cryoprotected in 30% (v/v) Ethylene glycol and flash cooled in liquid nitrogen. Data were collected to 3.0 Å and belonged to space group P2$_1$2$_1$2$_1$ (Table S2). This crystal form could not be solved by molecular replacement using other known structures of Mu-homology domains, and despite soaking with various heavy atoms, no anomalous signal was visible to enable structure solution by experimental phasing.

This crystal form was eventually solved by molecular replacement using the δ-COP-W404A structure as a model (see below). It has eight molecules in the asymmetric unit packed in a helical arrangement via the interaction between Trp-404 in one molecule, with pocket 1 in the tryptophan-based motif binding site of another molecule. This explained the biophysical data showing that di-tryptophan peptides bind to δ-COP MHD more tightly when this 'back side' tryptophan is mutated (data not shown). The final structure was refined to R$_{work}$/R$_{free}$ values of 0.195/0.233 in Refmac5 using local NCS restraints. Density representing residues 284-546 of δ-COP is visible in all eight molecules in asymmetric unit. Two calcium ions appear to be present (from the crystallisation condition), both of which bridge neighbouring molecules in the crystal.

Crystallisation of native δ-COP-W404A with the Dsl1p WxW peptide:
δ-COP MHD had a tendency to aggregate into larger oligomers, and binding to di-tryptophan peptides prevented this aggregation (highlighted during purification of protein-motif fusion
constructs). Modelling studies suggested that W404 in δ-COP MHD was exposed to solvent, and might be causing this aggregation, so we mutated W404 to alanine, and this appeared to prevent oligomerisation as assessed by gel filtration. The native δ-COP-W404A protein crystallised in the presence of the Dsl1p WxW peptide (DDWNWEVED) with the protein at 10 mg/ml, and the peptide at a 2:1 molar ratio by vapour diffusion, in 0.1 M HEPES pH 6.6, 2.8 M ammonium sulphate, and were cryoprotected in 4 M sodium formate, 0.05 M HEPES pH 6.6, 1.4 M ammonium sulphate, 1 mM DDWNWEVED peptide. Data were collected to 1.8 Å resolution at Diamond Light Source, beamline I03 (Table S2). The L test and cumulative intensity distributions suggested that all crystals were perfect merohedral twins. All crystals belonged to apparent point group P622, with unit cell dimensions of \(a = 73\, \text{Å}, b = 73\, \text{Å}, c = 343\, \text{Å}, \alpha = 90^\circ, \beta = 90^\circ, \gamma = 120^\circ\), with an estimated two molecules in the asymmetric unit (44% solvent), or four molecules for a hemihedral twin. As reflections along the 0,0,l axis were present only when \(l=3n\), the true space group would be one of \(P6_2, P6_4, P3_21, P3_221, P3_12\) or \(P3_212\). Molecular replacement was tried in all six possible space groups, using other known structures of MHDs as search models, however no believable solution was found. We failed to solve the structure by experimental phasing with Hg, presumably due to the twinning. Micro-seeding failed to produce untwinned crystals.

This merohedrally twinned crystal form was eventually solved by molecular replacement using the \(\delta\)-COP-W404A structure, solved by selenomethionine MAD (see below), as a model in space group \(P3_21\). Initially molecular replacement with Phaser (45) was carried out in all six possible space groups searching for one molecule (of an expected four), and plausible solutions were found in both \(P3_21\) and \(P3_12\). Second and third molecules were found in both of these space groups, however a fourth molecule was only found after removing a clashing loop (residues 459-465). The structure was refined with restrained amplitude-based twin refinement using Refmac (47): after one round the \(R\) factors clearly discriminated between space groups \(P3_21\) (\(R_{\text{work}}; R_{\text{free}} = 0.27; 0.33\)) and \(P3_12\) (\(R_{\text{work}}; R_{\text{free}} = 0.34; 0.40\)).

During refinement it became clear that the loop between strands 11 and 12 adopts a different conformation due to contacts with a neighbouring molecule in the crystal. This explained the failure of the molecular replacement to find all of the molecules using the initial unmodified model. It is not clear whether this conformational change has any functional role. Following refinement, the \(mF_o-DF_c\) difference map showed clear electron density map representing the two tryptophans in the peptide (Figure S2F&G). The DDW\(_{413}N\_W\_{415}EVED\) peptide was built (density for DW\(_{413}N\_W\_{415}EV\) residues was visible) in the same orientation as the SeMet Dsl1p DDWNWE(SeMet)ED peptide, with the valine binding into pocket 3, however the density for Val-417 is poor and makes little contribution to binding, unlike the selenomethionine. The interactions between \(\delta\)-COP and the native peptide are largely the same as with the SeMet-derivative peptide, preserving the tryptophan
orientations and the hydrogen bond interactions with side chains His-350, Asn-352, and Gly-381. The final structure was refined to $R_{\text{work}}/R_{\text{free}}$ values of 0.136/0.189 using amplitude-based twin refinement and map sharpening in Refmac5 (47). Residues 412-417 of the peptide, and residues 285-546 of $\delta$-COP of the four molecules in asymmetric unit are visible, however the density for two loops is poor in some molecules (409-412 in chains A and D, 447-452 in chains B and D), and there are two breaks in the main chain at D/409, and B/451-452.

Crystallisation of SeMet-substituted $\delta$-COP-W404A bound to the Dsl1p SeMet-WxW peptide:

Selenomethionine-substituted $\delta$-COP-W404A was crystallised with a Dsl1p WxW peptide with selenomethionine replacing valine (DDWNWE(SeMet)ED) to aid identification of the orientation of the peptide. This peptide bound with an affinity similar to the native peptide (Figure S1). Crystals were obtained by vapour diffusion with 30 mg/ml protein, 2:1 molar ratio peptide:protein, in 0.1 M HEPES-Na pH 7.0, 0.15 M (NH4)$_2$SO$_4$, 21% (w/v) PEG-3350. Crystals grew after 2-3 days before degrading and eventually disappearing after ~7 days, and were therefore flash-cooled on day 3-4. Crystals were cryoprotected in 15% glycerol, 0.1 M HEPES-Na pH 7.0, 0.15 M (NH4)$_2$SO$_4$, 22% (w/v) PEG-3350, 1.5 mM DDWNWE(SeMet)ED peptide, and a three-wavelength MAD dataset was collected to 2.8 Å resolution at the selenium K edge (Table S1). After MAD phasing, the solvent content was optimised to 56% with eight molecules in the asymmetric unit (Figure S2). The initial atomic model was built automatically using Buccaneer (44) and refined against the high remote data set to 2.8 Å resolution using local NCS restraints, experimental phase restraints, and map sharpening in Refmac5 (47), with rebuilding in Coot (46). The two tryptophans on either side of His-350, and the selenomethionine into pocket 3 of the peptide were clear even in the experimental phased maps prior to model building and refinement (Figure S2B&C), and the selenomethionine in the peptide was clear from the anomalous difference map (Figure S2C). The final structure was refined to final $R_{\text{work}}/R_{\text{free}}$ values of 0.189/0.234. All nine residues of the peptide and residues 285-546 of $\delta$-COP are visible in all eight molecules in the asymmetric unit. However due to crystal contacts the density for four of the eight molecules (chains A-D) is better than for the other four (chains E-H); the density for some or all of residues 406-415, 437-466, and 482-493 (chain H) in these chains (E-H) is poor.

Crystallisation of SeMet-substituted $\delta$-COP-W404A with the Dsl1p WxxxW peptide:

Selenomethionine-substituted $\delta$-COP-W404A protein crystallised in the presence of the Dsl1p WxxxW peptide (ENAWEAWAIDEC) with the protein at 15 mg/ml, and the peptide at a 2:1 molar ratio by vapour diffusion, in 0.1 M Bicine pH 9.0, 3.2 M ammonium sulphate, and were cryoprotected in 4 M sodium formate, 0.05 M Bicine pH 9.0, 1.7 M ammonium sulphate, 2 mM ENAWDEAWAIDEC peptide. Data were collected to 2.8 Å resolution at Diamond Light Source,
beamline I03 (Table S2). These crystals belonged to space group P3\(_{1}\)21, the same space group as with the native Dsl1p WxW peptide, but were not twinned. Following refinement the density representing the peptide was discontinuous, presumably due to low occupancy of the peptide, however the motif clearly binds at the same site and most likely in the same orientation (Figure S2).

**Crystallisation of SeMet-substituted \(\delta\)-COP-W404A with the Gcs1p WxxF peptide:**
Following the observation that a WxxF peptide (DEDKWDDF) corresponding to the WxxF motif at the very C-terminus of Gcs1p binds to \(\delta\)-COP MHD (Figure 4C), crystals of selenomethionine-substituted \(\delta\)-COP-MHD W404A were grown, using 20 mg/ml protein with the Gcs1p WxxF peptide (DEDKWDDF) added at a 2:1 molar ratio in 0.1 M Tris-HCl pH 8.5, 0.15 M magnesium chloride, 26% (w/v) PEG-4000. The selenomethionine-substituted protein allowed use of the experimental phase information in refinement, which improved the connectivity of the density (particularly of the peptides). Crystals were obtained by vapour diffusion, cryoprotected with 15% glycerol, and data were collected to 2.5 \(\AA\) resolution. These crystals belonged to space group P2\(_{1}\)2\(_{1}\)2\(_{1}\) (Table S1). This crystal form was solved by molecular replacement using one molecule of the previously determined structure, and had three molecules in the asymmetric unit. Following refinement the \(mF_o-DF_c\) difference map showed clear electron density representing the tryptophan and phenylalanine in the peptide (Figure S2H&I). The DEDKWDDDF peptide binds in the same orientation as the Dsl1p WxWx\(\phi\) peptides, with the tryptophan binding into pocket 1, and the phenylalanine in pocket 2. The tryptophan motif binding site is accessible in all three molecules in the asymmetric unit, however the electron density was better in one molecule (chain B) than in the other two, due to interactions of the peptide with a neighbouring molecule. Despite lower occupancy of the peptide bound to the other two molecules, the peptide appeared to bind in the same orientation. The structure was refined to final R\(_{\text{work}}/R_{\text{free}}\) values of 0.193/0.258 using TLS refinement, local NCS restraints, the SAD data directly, and map sharpening in Refmac5 (47). In the best molecule, all but the first two residues (DE) of the peptide could be built.

**Yeast Strains and Growth Media:**
The Saccharomyces cerevisiae strains Y7092 (MATa, his3\(\Delta\)1 leu2\(\Delta\)0 ura3\(\Delta\)0 met15\(\Delta\)0) and MY15059 (MATa, can1\(\Delta\)::pSTE2-Sp_HIS5 his3\(\Delta\)1 lyp1\(\Delta\) his3\(\Delta\)1 leu2\(\Delta\)0 ura3\(\Delta\)0 met15\(\Delta\)0 dsl1\(\Delta\)E-natMX) (24) were used to study the \(\delta\)-COP:Dsl1p interaction \textit{in vivo}. Yeast strain PPY17:168 (gcs1::URA3 glo3::HIS3 ura3 his3 leu2 trp1 ade2) harboring plasmids pGAL-GLO3 and either pPP16:100-5-GCS1-WF (i.e. wild-type GCS1), pPP16:100-12-gcs1-AA (the gcs1 mutant carrying the W349A, F352A mutations), or empty-vector pRS315 was used to study the \(\delta\)-COP:Gcs1p interaction \textit{in vivo}. Yeast cells were grown aerobically at 30°C, unless otherwise noted. YEPD rich media contained yeast extract (1% w/v), peptone (2% w/v), and glucose (2% w/v). Sporulation
media (SPM) contained potassium acetate (0.3% w/v) and raffinose (0.02% w/v) and was supplemented with histidine, leucine, methionine, and uracil (0.03% w/v each). YEPD+Nat media contained 100 µg/ml nourseothricin (clonNAT).