

# Rules for the recognition of dilysine retrieval motifs by coatomer

Wenfu Ma and Jonathan Goldberg\*

Memorial Sloan-Kettering Cancer Center, Howard Hughes Medical Institute and the Structural Biology Program, New York, NY, USA

**Cytoplasmic dilysine motifs on transmembrane proteins are captured by coatomer  $\alpha$ -COP and  $\beta'$ -COP subunits and packaged into COPI-coated vesicles for Golgi-to-ER retrieval. Numerous ER/Golgi proteins contain K(x)Kxx motifs, but the rules for their recognition are unclear. We present crystal structures of  $\alpha$ -COP and  $\beta'$ -COP bound to a series of naturally occurring retrieval motifs—encompassing KKxx, KxKxx and non-canonical RKxx and viral KxHxx sequences. Binding experiments show that  $\alpha$ -COP and  $\beta'$ -COP have generally the same specificity for KKxx and KxKxx, but only  $\beta'$ -COP recognizes the RKxx signal. Dilysine motif recognition involves lysine side-chain interactions with two acidic patches. Surprisingly, however, KKxx and KxKxx motifs bind differently, with their lysine residues transposed at the binding patches. We derive rules for retrieval motif recognition from key structural features: the reversed binding modes, the recognition of the C-terminal carboxylate group which enforces lysine positional context, and the tolerance of the acidic patches for non-lysine residues.**

*The EMBO Journal* (2013) 32, 926–937. doi:10.1038/emboj.2013.41; Published online 12 March 2013

**Subject Categories:** membranes & transport; structural biology

**Keywords:** coatomer; COPI; dilysine signal; vesicular transport

## Introduction

Many resident proteins of the endoplasmic reticulum (ER) are localized to that compartment by continuous retrieval from the Golgi complex. Luminal ER proteins may be retrieved by means of a C-terminal KDEL sequence that binds to the KDEL receptor to trigger retrograde transport (Lewis and Pelham, 1992). Type I transmembrane proteins—including ER residents and itinerant proteins of the ER/Golgi system—contain a C-terminal dilysine motif that signals their retrieval from post-ER membranes (Jackson *et al.*, 1993; Gaynor *et al.*, 1994; Townsley and Pelham, 1994).

The dilysine motif, initially identified on the adenoviral E3 19 kDa (E19) protein, consists of a pair of lysine residues at the –3 and –4 (or –5) positions relative to the cytoplasmic

C terminus of the cargo molecule—termed as KKxx and KxKxx motifs (Nilsson *et al.*, 1989; Jackson *et al.*, 1990). Variants of the dilysine motif have been characterized, in particular a KxHxx retrieval motif in the tail of the spike protein of group I coronaviruses (Lontok *et al.*, 2004), and the RKxx motif present in the Golgi-localized Scyl1 protein (Burman *et al.*, 2008). The dilysine motif is recognized by coatomer subunits (Cosson and Letourneur, 1994), and dilysine-tagged transmembrane cargo is thereby packaged into COPI(coatomer)-coated vesicles for retrograde transport to the ER (Letourneur *et al.*, 1994).

The biochemical basis for recognition of dilysine motifs by coatomer has been controversial, but one path of research has led in recent years to a clear delineation of the dilysine motif-binding sites. These were mapped initially to the  $\alpha\beta'\epsilon$ -COP subcomplex of coatomer (Cosson and Letourneur, 1994; Lowe and Kreis, 1995; Fiedler *et al.*, 1996; Schröder-Köhne *et al.*, 1998). Subsequent yeast genetic studies implicated the N-terminal  $\beta$ -propeller domains of  $\alpha$ -COP and  $\beta'$ -COP, the sequence-related large subunits of  $\alpha\beta'\epsilon$ -COP (Eugster *et al.*, 2004). A recent structural analysis of the  $\beta'$ -COP N-terminal domain crystallized with a KxKxx peptide pinpointed the binding site near to the ‘pore’ of the  $\beta$ -propeller protein, and a corresponding binding site was mapped via functional tests to the pore region of  $\alpha$ -COP (Jackson *et al.*, 2012). However, the mode of KxKxx motif recognition by coatomer was not resolved in this study because the dilysine-binding site on  $\beta'$ -COP was partially occluded by a crystal contact (Jackson *et al.*, 2012).

The location of dilysine-binding sites on the cage-forming  $\alpha\beta'\epsilon$ -COP subcomplex (and not the adaptin-like  $\beta\delta/\gamma\zeta$ -COP subcomplex) of coatomer is interesting in view of the finding that dilysine-tagged cargo molecules are required for efficient COPI-coat assembly and vesicle formation (Bremser *et al.*, 1999; Aguilera-Romero *et al.*, 2008). Thus, the dilysine motif binds to  $\beta'$ -COP at a site close to the vertex of the triskelion formed by three  $\alpha\beta'\epsilon$ -COP assembly units (Lee and Goldberg, 2010), suggesting how COPI coat nucleation may be coupled to the packaging of dilysine-tagged cargo molecules, in particular oligomeric species such as p24-family proteins and the Mst27p/Mst28p complex (Bremser *et al.*, 1999; Sandmann *et al.*, 2003; Aguilera-Romero *et al.*, 2008).

Collectively, these findings highlight the importance of establishing the binding modes and specificity of dilysine-type motifs for coatomer subunits. Here, we describe 11 high-resolution crystal structures that show how  $\alpha$ -COP and  $\beta'$ -COP bind to a comprehensive series of motifs, including the canonical KKxx and KxKxx classes as well as the variant KxHxx and RKxx signals. We show that the recognition modes of KKxx and KxKxx motifs are distinct, a non-intuitive observation that helps to rationalize the binding mode of the RKxx motif. We demonstrate that  $\alpha$ -COP and  $\beta'$ -COP subunits have comparable specificity towards KKxx and KxKxx motifs, contrary to previous reports. And we show that coatomer subunit specificity diverges instead at the level of K(x)Kxx

\*Corresponding author. Memorial Sloan-Kettering Cancer Center, Howard Hughes Medical Institute and the Structural Biology Program, 1275 York Avenue, New York, NY 10065, USA. Tel.: +1 212 639 6712; E-mail: jonathan@ximpackt4.ski.mskcc.org

Received: 22 January 2013; accepted: 7 February 2013; published online: 12 March 2013

sequence context and in the recognition of the RKxx motif, which binds exclusively to  $\beta'$ -COP.

## Results

### **Specific interactions of $\alpha$ -COP and $\beta'$ -COP with KKxx and KxKxx retrieval motifs**

The results of yeast genetic experiments led to the proposal that  $\alpha$ -COP and  $\beta'$ -COP are selective towards KKxx and KxKxx motifs, respectively (Eugster *et al*, 2004). The reporter molecule used in these studies was fused to the C terminus of yeast Wbp1 (a component of the ER N-oligosaccharyl-transferase complex), terminating in the sequence TFKKTN. A recent analysis demonstrated that this sequence binds to  $\beta'$ -COP with 10-fold lower affinity than do KxKxx peptides, and that while  $\beta'$ -COP has some capacity to transport KKxx-tagged cargo, it has a distinct binding preference for KxKxx motifs (Jackson *et al*, 2012). In preliminary binding experiments, we noted, however, that  $\alpha$ -COP and  $\beta'$ -COP displayed highly similar selectivity towards a panel of KKxx and KxKxx sequences (Figure 1).

Figure 1A shows the results of an initial test for binding of dilysine motifs to the complete  $\alpha\beta'\epsilon$ -COP subcomplex of coatomer (see Materials and methods). The dilysine motifs were fused to the C terminus of glutathione S-transferase (GST), then immobilized on glutathione sepharose beads and probed for binding to  $\alpha\beta'\epsilon$ -COP. Purified recombinant *Bos taurus*  $\alpha\beta'\epsilon$ -COP was added to a mix of 'background' *E. coli* proteins, and the specificity of binding was assessed by the enrichment of COPs from the input mixture (Figure 1A).  $\alpha\beta'\epsilon$ -COP binds to the KKxx motifs of the mammalian proteins p25 and Ergic-53 and the adenoviral E19 protein, and to a KKxx motif nested in a poly-serine sequence (labelled KKSS in Figure 1; the polyS-KKSS sequence confers ER residence to a CD8 reporter molecule in HeLa cells; Jackson *et al*, 1990). No specific binding was observed to the p24-family protein p24 $\beta$ , which does not contain a dilysine motif at its C terminus. For control sequences, we added two C-terminal serine residues, KKxxSS (labelled p25 + SS in Figure 1), since this addition was shown to abolish the ER retention of a CD8 reporter molecule (Jackson *et al*, 1990).

Next, we tested the binding of *S. cerevisiae*  $\alpha$ -COP and  $\beta'$ -COP individually to the dilysine motifs. Utilizing the sequence homology of  $\alpha$ -COP and  $\beta'$ -COP, we designed two protein complexes of the form shown in Figure 1H, namely the propeller-propeller-solenoid array of one COP (containing the dilysine-binding site in the N-terminal  $\beta$ -propeller domain) plus a short interacting  $\alpha$ -solenoid region of the other COP, which aids the stability of the dimer (Lee and Goldberg, 2010). Pull-down experiments with the two complexes, performed in parallel, show a highly similar pattern of binding to the KKxx motifs (compare Figures 1B and E, lane by lane). Likewise, the two complexes bind specifically to the KxKxx motif of Emp47p (Figures 1C and D).

These data reveal that  $\alpha$ -COP and  $\beta'$ -COP bind with comparable specificity to a panel of KKxx and KxKxx motifs, and we conclude that  $\beta'$ -COP does not disfavour KKxx motifs in general. Instead, we speculate that  $\beta'$ -COP may bind weakly to yeast Wbp1 owing to the particular sequence context of its dilysine motif, an issue we return to below.

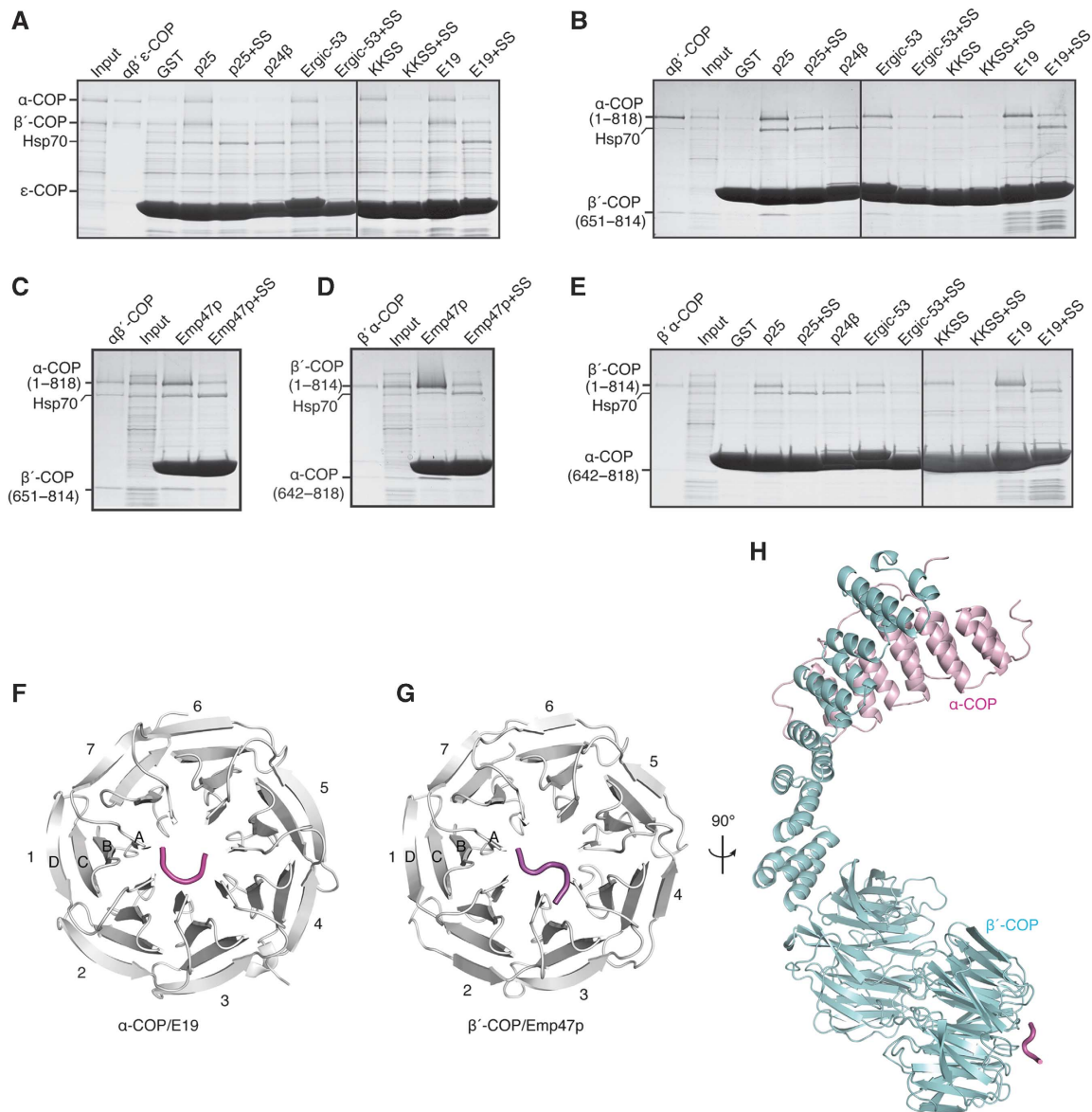
### **Crystal structures of $\alpha$ -COP and $\beta'$ -COP bound to KxKxx retrieval motifs**

Efforts to crystallize  $\alpha$ -COP and  $\beta'$ -COP (specifically their N-terminal  $\beta$ -propeller domains) bound to a comprehensive set of naturally occurring retrieval motifs met with two main challenges. First, the problem of poor solubility of yeast and mammalian  $\alpha$ -COP proteins that hampered previous studies (Lee and Goldberg, 2010) was solved by using the *Schizosaccharomyces pombe* N-terminal  $\beta$ -propeller domain, engineered to improve its solubility by mutating five hydrophobic residues to lysine in a flexible loop—L181K, L185K, I192K, L196K and F197K. Second, a fraction of the crystals we obtained had intimate crystal contacts in the vicinity of the dilysine-binding site that occlude all or part of the site. Six crystal structures were discarded for this reason, yielding a final set of eleven structures: eight *S. cerevisiae*  $\beta'$ -COP/peptide complexes, two *S. pombe*  $\alpha$ -COP/peptide complexes and one structure of *S. pombe*  $\alpha$ -COP with a vacant binding site. The crystal structures were refined to resolutions ranging from 1.45 to 1.9 Å (see Table I for crystallographic data and a list of the dilysine sequences used).

Figures 1F and G show the closely related folds of the N-terminal  $\beta$ -propeller domains of  $\alpha$ -COP and  $\beta'$ -COP (the *S. pombe*  $\alpha$ -COP and *S. cerevisiae*  $\beta'$ -COP sequences share 30% sequence identity in this ~300 residue domain). Both coatomer subunits bind to dilysine motifs via a highly conserved surface region of the  $\beta$ -propeller domain, near to the central 'pore' of the propeller, as mapped previously by Jackson *et al* (2012).

To facilitate a direct structural comparison of the dilysine-binding sites of  $\alpha$ -COP and  $\beta'$ -COP, we solved crystal structures of both proteins bound to the KxKxx motif of yeast Emp47p, with the sequence IKTKLL (Figures 2B and D, and see Table I). These are the first views of dilysine motifs bound to coatomer unobstructed by crystal contacts, and they define the three 'primary' points of interaction between the retrieval motif and coatomer subunit: the C-terminal carboxylate group of the KxKxx peptide binds to a basic patch involving residues Arg15, Lys17 and Arg59 (*S. cerevisiae*  $\beta'$ -COP numbering), and the two lysine side chains interact with two acidic patches. These primary interactions can be seen more clearly in Figures 3B and D, which show the crystal structure of  $\beta'$ -COP bound to the KxKxx motif of human Wbp1 (hWbp1 in Table I; also called OST48). We refer to the acidic patches as patch 1, involving  $\beta'$ -COP residue Asp206 that binds to the -3 lysine side chain, and acidic patch 2, involving residues Asp98 and Asp117 binding to lysine -5 of the peptide (labelled accordingly in Figure 3B). In the crystal structure of *S. cerevisiae*  $\beta'$ -COP reported by Jackson *et al* (2012; PDB identifier 2YNN), a KxKxx peptide with sequence CTFKTKTN interacts with the basic patch and with patch 1 via the -3 lysine, but patch 2 is occluded by a crystal contact and the -5 lysine side chain of the motif projects away from the binding site as a result.

In addition to the three primary contacts, important secondary contacts involve the backbone carbonyl oxygen atoms of the -4, -3 and -2 residues, which interact with charged side chains Arg59 and Arg101 ( $\beta'$ -COP numbering) at the base of the dilysine-binding site (see, e.g., KEKSD of hWbp1 in Figure 3D). The primary and secondary interactions made by the KxKxx motif, as defined above, are common to the  $\alpha$ -COP/Emp47p,  $\beta'$ -COP/Emp47p and  $\beta'$ -COP/hWbp1



**Figure 1**  $\alpha$ -COP and  $\beta'$ -COP bind with comparable specificity to a series of KKxx and KxKxx retrieval motifs. (A) Binding of *B. taurus*  $\alpha\beta'\epsilon$ -COP subcomplex to KKxx sequences. GST-dilysine fusion proteins immobilized on glutathione sepharose beads were incubated with the  $\alpha\beta'\epsilon$ -COP subcomplex plus 'background' protein mix (see Materials and methods). Specific binding was assessed by the enrichment of  $\alpha\beta'\epsilon$ -COP proteins from the input, relative to the GST control. *E. coli* Hsp70 is an innocuous component of the background mix. Bound proteins were analysed by 4–20% SDS-PAGE and Coomassie blue staining. Control dilysine sequences contained two additional serine residues at their C termini (labelled, e.g., p25 + SS). The label KKSS refers to a dilysine sequence nested in a poly-serine background. (B) The *S. cerevisiae*  $\alpha$ -COP dilysine-binding site displays similar specificity towards KKxx motifs as full-length  $\alpha\beta'\epsilon$ -COP. These binding experiments were performed with the same panel of dilysine motifs as (A), using in this case the *S. cerevisiae*  $\alpha$ -COP(1–818)/ $\beta'$ -COP(651–814) protein dimer. (C)  $\alpha$ -COP binds specifically to the KxKxx motif of Emp47p. This experiment used the  $\alpha$ -COP(1–818)/ $\beta'$ -COP(651–814) protein dimer. (D)  $\beta'$ -COP likewise binds specifically to the KxKxx motif of Emp47p. This experiment used the *S. cerevisiae*  $\beta'$ -COP(1–814)/ $\alpha$ -COP(642–818) dimer. (E)  $\beta'$ -COP binds to the panel of KKxx motifs with the same specificity as  $\alpha$ -COP. Here, we used the  $\beta'$ -COP(1–814)/ $\alpha$ -COP(642–818) dimer. Note the comparison with (B), lane by lane. (F) Crystal structure of the N-terminal  $\beta$ -propeller domain of *S. pombe*  $\alpha$ -COP bound to the KKxx motif of adenoviral E19 protein (magenta worm). This view towards the dilysine-binding site indicates the seven blades (or WD-repeat elements) of the  $\beta$ -propeller. (G) Crystal structure of the N-terminal  $\beta$ -propeller domain of *S. cerevisiae*  $\beta'$ -COP bound to the KxKxx motif of Emp47p. (H) Structure of the Emp47p KxKxx motif bound to the  $\alpha\beta'$ -COP fragment of the COPI assembly unit. This is a composite model comprising the crystal structure from the present study, as in (G), together with the  $\alpha\beta'$ -COP fragment from a previous study (Lee and Goldberg, 2010).

complexes and, as this implies, the conformations of the bound peptides are highly similar (compare, e.g., Figures 2B and D). This despite the significant difference in the character of the bound peptides, namely the highly charged hWbp1 sequence EKEKSD versus the more hydrophobic Emp47p sequence IKTKLL. The constancy of the KxKxx motif conformation and interactions among the four high-resolution crystallographic observations (the  $\beta'$ -COP/hWbp1 crystals

contain two distinct complexes in the asymmetric unit) implies that this is the bona fide binding mode.

#### Equivalence of the dilysine-binding sites of $\alpha$ -COP and $\beta'$ -COP

In hindsight, the close correspondence of KxKxx binding to  $\alpha$ -COP and  $\beta'$ -COP is not surprising in view of the very high degree of conservation of the binding-site residues

**Table 1** Data collection and refinement statistics

Coatmer subunit	$\beta'$ -COP	$\beta'$ -COP	$\beta'$ -COP	$\beta'$ -COP	$\beta'$ -COP	$\beta'$ -COP	$\beta'$ -COP	$\beta'$ -COP	$\beta'$ -COP	$\beta'$ -COP	$\beta'$ -COP	$\beta'$ -COP	$\beta'$ -COP	$\beta'$ -COP	$\alpha'$ -COP	$\alpha'$ -COP
Peptide source	p25	yWbp1	hWbp1	Emp47p	Scyl1	PEDVspike	Insig-1	Insig-2	E19	Emp47p						
Peptide sequence	FEAKKLV	TFKKTN	EKEKSD	IKTKLL	GARKLD	FEKVHVQ	PEKPHSD	AEKSHQE	RRSFIDE- KKMP	IKTKLL						
Protein Data Bank ID	4J73	4J86	4J77	4J78	4J84	4J79	4J81	4J82	4J8G	4J8B						
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub>	P2 <sub>1</sub>	C2	P1	C2	P1	C2	P2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>						
Copies per asymmetric unit	1	2	2	1	2	1	2	2	2	1						
Cell parameters <i>a</i> , <i>b</i> , <i>c</i> (Å)	36.6, 56.1, 132.7	74.0, 50.6, 83.6	75.7, 50.9, 82.4	75.0, 51.0, 85.1	42.5, 46.1, 84.3	75.1, 51.2, 85.2	48.7, 48.7, 75.2	83.8, 51.4, 76.4	64.9, 62.6, 77.3	49.7, 92.0, 87.2						
$\alpha$ , $\beta$ , $\gamma$ (deg)	90, 90, 90	90, 99.5, 90	90, 90.2, 90	90, 101.6, 90	81.5, 82.0, 69.5	90, 101.7, 90	89.7, 89.7, 63.9	90, 91.1, 90	90, 93.6, 90	90, 90, 90						
<i>Data processing</i>																
Resolution (Å)	44–1.45	43–1.5	43–1.75	42–1.5	50–1.5	42–1.55	50–1.75	42–1.7	50–1.9	42–1.9						
$R_{\text{merge}}$ (%) <sup>a</sup>	4.5 (37.8) <sup>b</sup>	4.7 (48.3)	7.4 (39.2)	4.0 (44.7)	11.6 (49.5)	5.4 (52.4)	5.0 (24.5)	10.4 (38.8)	8.5 (49.1)	5.2 (52.9)						
<i>I</i> / $\sigma$	15.6 (2.3)	17.4 (2.6)	8.4 (2.5)	16.5 (2.4)	20.4 (2.5)	13.3 (2.5)	21.9 (3.4)	7.2 (2.2)	11.7 (2.4)	14.9 (2.4)						
Completeness (%)	96.8 (83.1)	99.4 (96.6)	97.8 (95.3)	97.6 (95.9)	93.5 (93.0)	97.2 (96.6)	95.9 (93.4)	95.9 (87.3)	99.8 (99.9)	98.8 (98.3)						
Redundancy	4.4 (3.2)	3.6 (3.5)	4.1 (3.8)	3.7 (3.6)	6.1 (3.6)	3.7 (3.7)	2.2 (1.8)	3.9 (3.9)	4.1 (4.1)	4.2 (4.2)						
<i>Refinement statistics</i>																
Data range (Å)	44–1.45	43–1.5	43–1.75	42–1.5	50–1.5	42–1.55	50–1.75	42–1.7	50–1.9	42–1.9						
Reflections	48 785	102 029	60 204	51 789	93 130	43 890	60 400	34 056	47 040	29 390						
Non-hydrogen atoms	4863	5618	5487	2785	5597	2800	5461	5410	5421	2645						
Water molecules	301	705	599	342	797	343	588	535	468	149						
R.m.s. $\Delta$ bonds (Å) <sup>c</sup>	0.011	0.008	0.013	0.006	0.006	0.007	0.007	0.006	0.007	0.008						
R.m.s. $\Delta$ angles (deg) <sup>c</sup>	1.4	1.3	1.3	1.2	1.2	1.2	1.2	1.1	1.3	1.3						
R factor (%) <sup>d</sup>	17.7	16.3	20.2	14.2	18.8	17.3	15.8	16.9	18.0	18.6						
$R_{\text{free}}$ (%) <sup>d,e</sup>	20.0	18.8	24.5	16.3	20.8	20.3	19.8	21.2	21.9	21.8						

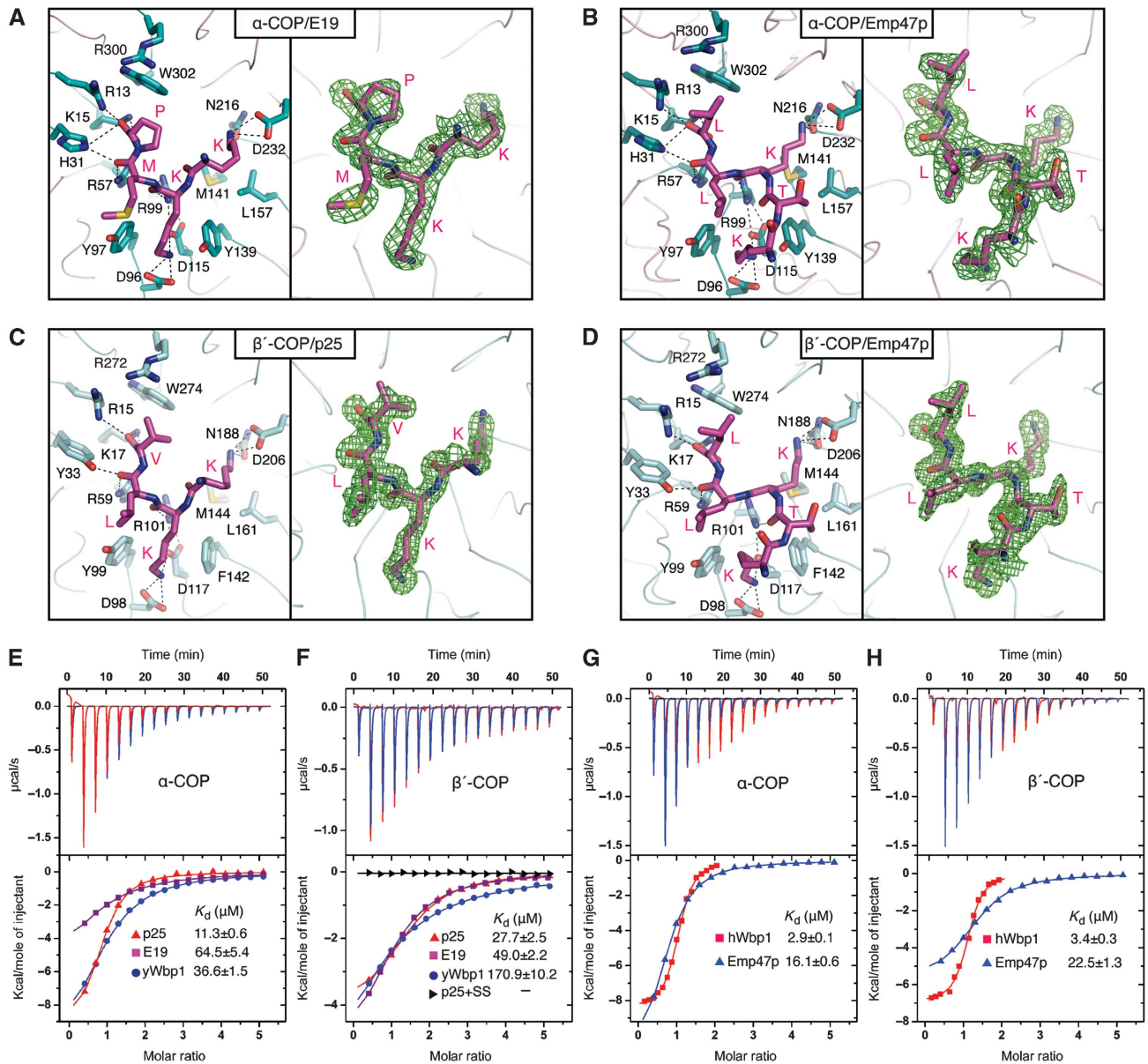
<sup>a</sup> $R_{\text{merge}} = 100 \times \sum_i \sum_h |I_i(h) - \langle I_i(h) \rangle| / \sum_h \langle I_i(h) \rangle$ , where  $I_i(h)$  is the *i*th measurement and  $\langle I_i(h) \rangle$  is the weighted mean of all measurements of  $I_i(h)$  for Miller indices *h*.

<sup>b</sup>Highest resolution shell is shown in parenthesis.

<sup>c</sup>Root-mean-squared deviation (r.m.s.  $\Delta$ ) from target geometries.

<sup>d</sup>R factor =  $100 \times \sum |F_o - F_p| / \sum F_o$ .

<sup>e</sup> $R_{\text{free}}$  was calculated with 5% of the data.

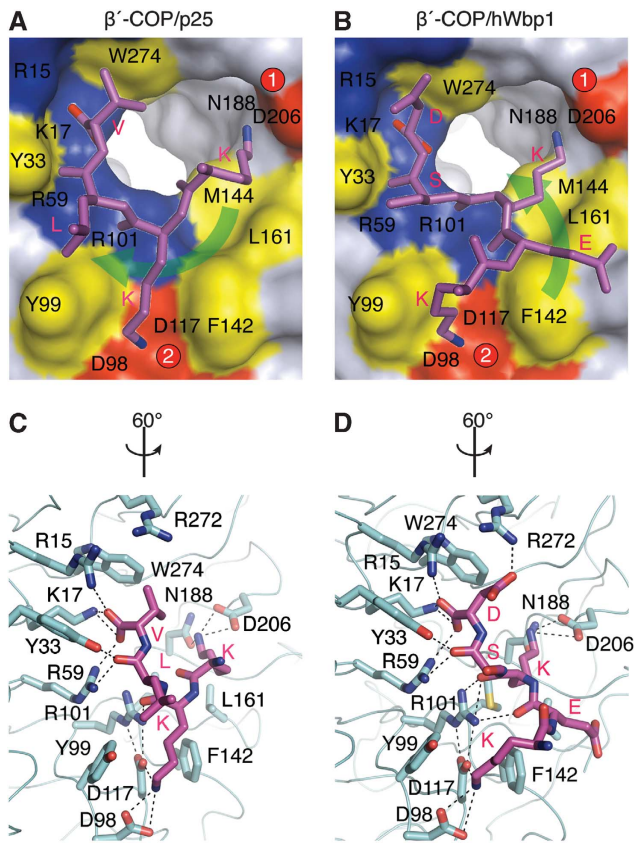


**Figure 2** Molecular recognition of KKxx and KxKxx retrieval motifs. (A) Two views of the KKMP motif of adenoviral E19 protein bound to  $\alpha$ -COP. The left panel shows a close-up view of the dilysine-binding site. All the  $\alpha$ -COP residues that contact the dilysine motif are drawn in the picture. The right panel shows the KKMP peptide together with a difference electron density map calculated following a simulated annealing protocol (SA-omit map; Brünger, 1998), calculated at 1.9 Å resolution and contoured at  $3.0\sigma$  (Table 1). Electron density has been truncated for N-terminal residues outside the binding site. (B) View of the KTKLL motif of Emp47p bound to  $\alpha$ -COP. In the right panel, the SA-omit map is calculated at 1.9 Å resolution and contoured at  $3.0\sigma$ . (C) Structure of the KKLK motif of the p24-family protein p25 (from human) bound to  $\beta'$ -COP. The SA-omit map is calculated at 1.45 Å resolution and contoured at  $3.0\sigma$ . (D) Structure of the KTKLL motif of Emp47p bound to  $\beta'$ -COP. The SA-omit map is calculated at 1.5 Å resolution and contoured at  $3.0\sigma$ . (E) Isothermal calorimetry measurements for  $\alpha$ -COP binding to KKxx motifs. The upper panel shows examples of the primary data. The lower panel shows data fitted to binding isotherms to obtain affinities (and standard errors) for the p25 (KKLV), E19 (KKMP) and yWbp1 (KKTN) sequences. (F) Affinity of  $\beta'$ -COP for KKxx motifs. (G) Affinity of  $\alpha$ -COP for the KxKxx motifs of hWbp1 (KEKSD) and Emp47p (KTKLL). (H) Affinity of  $\beta'$ -COP for the KxKxx motifs.

(Figures 2 and 3). Only two residues differ, and these are conservative alterations: Tyr33 of  $\beta'$ -COP changes to His31 in  $\alpha$ -COP and Phe142 of  $\beta'$ -COP changes to Tyr139 in  $\alpha$ -COP. The close structural equivalence of the binding sites accords with the results of pull-down experiments showing the highly similar specificity of  $\alpha$ -COP and  $\beta'$ -COP towards the dilysine motifs (Figure 1).

To extend the results of the pull-down experiments, we measured more precisely the affinities of dilysine motifs using isothermal calorimetry (ITC; Figures 2E–H, and see

Materials and methods). All the functional dilysine motifs we tested caused a significant heat change upon mixing with  $\alpha$ -COP and  $\beta'$ -COP (see Table 1 for a list of the sequences of peptides used). A control peptide KKxxSS caused no heat change upon mixing with  $\beta'$ -COP, confirming that the assay measures the protein/peptide interaction (black curve in Figure 2F, lower panel). Calorimetry data on the KxKxx motifs (Figures 2G and H) reveal that the Emp47p sequence IKTKLL binds with very similar affinity to  $\alpha$ -COP ( $K_d = 16.1 \mu\text{M}$ ) and  $\beta'$ -COP ( $K_d = 22.5 \mu\text{M}$ ); the hWbp1 peptide



**Figure 3** Comparison of the distinct binding modes of the KKxx and KxKxx motifs. (A) Conformation of the KKxx motif. Picture shows the KKLVL sequence of human p25 bound to  $\beta'$ -COP (refined at 1.45 Å resolution). The molecular surface of  $\beta'$ -COP is coloured according to residue properties: acidic residues coloured red, basic side chains blue and hydrophobic side chains yellow. Labels indicate acidic patch 1 (red circle, Asp206) and acidic patch 2 (Asp98 and Asp117). The binding mode of the KKxx motif is indicated by the clockwise arrow coloured green. The KKLVL peptide is coloured magenta, except for the side-chain nitrogen atoms of the lysine residues (blue) and the terminal carboxylate oxygen atoms (red); for complete atom colouring of the peptide see (C). (B) Conformation of the KxKxx motif. Picture shows the KEKSD sequence of hWbp1 bound to  $\beta'$ -COP (refined at 1.75 Å resolution). Colouring is as in (A), and the anticlockwise arrow indicates the binding mode of the KxKxx motif. (C) Related to (A), showing the KKxx motif bound to  $\beta'$ -COP. The view is rotated 60° around a vertical axis to provide a clearer view of the bonding network to the two lysine residues and to the terminal carboxylate group. (D) Related to (B), showing the KxKxx motif bound to  $\beta'$ -COP.

EKEKSD binds somewhat more tightly to both proteins, but again with almost equal affinity to  $\alpha$ -COP ( $K_d = 2.9 \mu\text{M}$ ) and  $\beta'$ -COP ( $K_d = 3.4 \mu\text{M}$ ).

Collectively, the results of the ITC and crystallography experiments on KxKxx motifs define a common and specific mode of binding to the  $\alpha$ -COP and  $\beta'$ -COP subunits of coatmer.

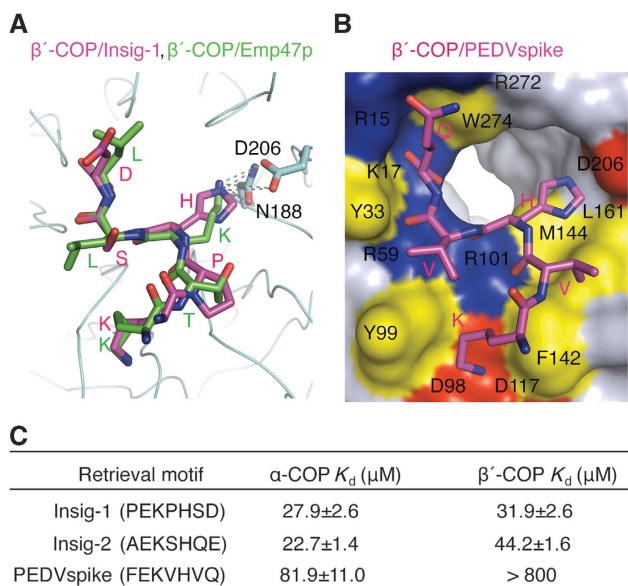
#### Distinct binding modes of KKxx and KxKxx motifs

Intuitively, one might expect the two motifs to bind in a similar fashion, with the intermediary residue of KxKxx looped away from the binding site relative to KKxx. But this is not the case, as our crystallographic analysis shows that the two motifs adopt different binding modes. Figures 2A and C show the structure of  $\alpha$ -COP bound to the KKMP sequence of adenoviral E19 protein and of  $\beta'$ -COP bound to

the KKLVL sequence of p25, a p24-family protein (see Table I for full peptide sequences). The KKxx motifs adopt a helical conformation at the binding sites. The comparison with the KxKxx binding mode shows some shared features: the terminal carboxylate group is bound in the same manner to the basic patch, and the backbone carbonyl oxygen atoms of the  $-3$  and  $-2$  (though not the  $-4$ ) residues of the KKxx motif interact with the side chains of Arg59 and Arg101 in a similar fashion to KxKxx (see Figure 3 for a side-by-side comparison of motifs). The lysine side chains of the KKxx motifs are recognized by the acidic patches 1 and 2 but, surprisingly, they bind in a reversed fashion relative to KxKxx. In Figures 3A and B, we emphasize the alternate binding modes with the clockwise green arrow indicating the helical conformation of the  $-4$  and  $-3$  residues of the KKxx motif and the anticlockwise arrow denoting the direction of the  $-5$ ,  $-4$  and  $-3$  residues of the KxKxx motif.

These crystallographic results indicate that the KKxx and KxKxx motifs adopt distinctive low energy binding modes to accommodate their lysine residues in the acidic patches. How certain can we be that this represents the bona fide binding mode for KKxx motifs? First, a comparison of the peptide conformations in the  $\beta'$ -COP/p25,  $\alpha$ -COP/E19 (Figures 2 and 3) and  $\beta'$ -COP/yWbp1 (Supplementary Figure S1) complexes shows highly similar conformations and sets of contacts at the binding sites. This involves a total of five distinct crystallographic observations of KKxx peptides, as both the  $\alpha$ -COP/E19 and  $\beta'$ -COP/yWbp1 crystals contain two complexes in the asymmetric unit (Table I). Second, the peptide conformations were defined by difference Fourier analysis of 1.45 Å (p25), 1.9 Å (E19) and 1.5 Å (yWbp1) resolution X-ray data sets. Third, the KKxx binding mode is common to the  $\alpha$ -COP and  $\beta'$ -COP binding sites. Fourth, the binding mode is independent of the sequence context, unaffected by a preponderance of hydrophobic (p25; KKLVL), hydrophilic (yWbp1; KKTN) or terminal proline (E19; KKMP) residues. On this basis, we conclude that the two dominant classes of dilysine motifs, KKxx and KxKxx, bind to a common site on  $\alpha$ -COP and  $\beta'$ -COP proteins, as shown previously (Eugster *et al*, 2004; Jackson *et al*, 2012), but adopt distinct modes of interaction.

The measurements by calorimetry of KKxx motif interactions are shown in Figures 2E and F. As expected, both  $\alpha$ -COP and  $\beta'$ -COP bind to the p25 and adenoviral E19 sequences, in accord with the results of the pull-down experiments, but there is somewhat more variation in the specificity of the two COPs (compared with their highly similar affinities towards KxKxx motifs). The proline-containing E19 motif, KKMP, binds with comparable affinity to  $\alpha$ -COP ( $K_d = 64.5 \mu\text{M}$ ) and  $\beta'$ -COP ( $K_d = 49.0 \mu\text{M}$ ), while the more conventional KKLVL sequence of p25 binds somewhat more tightly to  $\alpha$ -COP ( $K_d = 11.3 \mu\text{M}$ ) than to  $\beta'$ -COP ( $K_d = 27.7 \mu\text{M}$ ). The results reinforce the view that  $\beta'$ -COP does not disfavour KKxx motifs in general. The outlier in this analysis is the dilysine motif of yeast Wbp1 (yWbp1; KKTN), which binds with significantly lower affinity to  $\beta'$ -COP ( $K_d = 170.9 \mu\text{M}$ ) than to  $\alpha$ -COP ( $K_d = 36.6 \mu\text{M}$ ). These measurements are consistent with a previous study showing a 10-fold lower affinity of the KKTN sequence for  $\beta'$ -COP compared to KxKxx motifs (Jackson *et al*, 2012). And they offer a possible explanation for the results of trafficking experiments that revealed a selective retrieval defect in coatmer mutants lacking a



**Figure 4** Structural basis for recognition of the KxHxx motif. (A) Close-up view of the  $\beta'$ -COP binding site showing the structure of the KxHxx motif from Insig-1 (sequence KPHSD; coloured magenta). Overlapped is the structure of the KxKxx motif of Emp47p, coloured green. The  $\beta'$ -COP residue Asp206 (from acidic patch 1) is included, and bonding interactions are drawn to the histidine and lysine side chains. (B) Picture shows the KVHVQ sequence of the spike protein from porcine epidemic diarrhoea virus bound to  $\beta'$ -COP (PEDVspike in Table 1, refined at 1.55 Å resolution). The molecular surface of  $\beta'$ -COP is coloured according to residue properties: acidic residues coloured red, basic side chains blue and hydrophobic side chains yellow. (C) Summary of binding measurements by isothermal calorimetry of KxHxx motif interactions with  $\alpha$ -COP and  $\beta'$ -COP.

functional  $\alpha$ -COP dilysine-binding site that presumably are reliant on the  $\beta'$ -COP subunit for dilysine retrieval. Specifically, the  $\alpha$ -COP mutants were unable to retain a KKTN-tagged reporter molecule in the ER, but were able to maintain the localization of Emp47p, dependent on its KTKLL motif (Schröder-Köhne *et al.*, 1998; Eugster *et al.*, 2004).

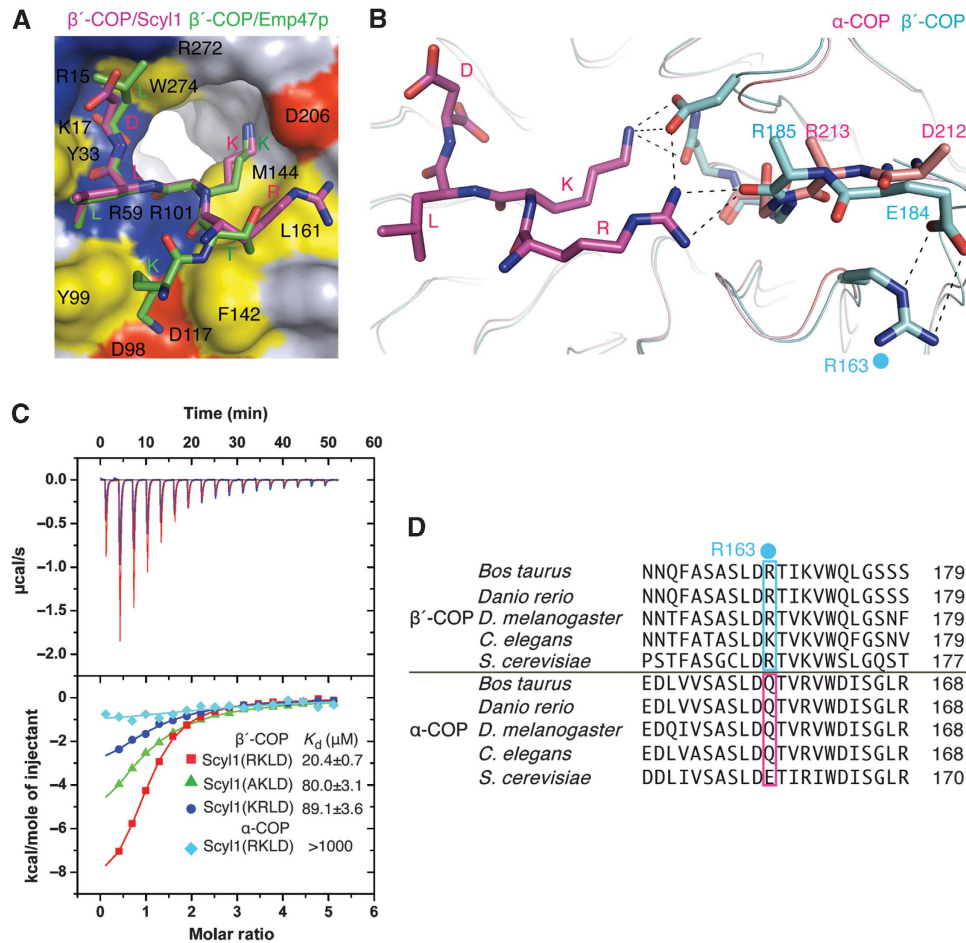
In the final stage of our analysis of KKxx motifs, we sought a molecular explanation for this, the difference in specificity of  $\alpha$ -COP and  $\beta'$ -COP towards the yWbp1 KKTN signal. Presumably, an explanation will be found in the interplay between the terminal residues of the motif KKTN (but not the upstream residues, TFKKTN, since these lie outside the binding site) and specific residue differences between the  $\alpha$ -COP and  $\beta'$ -COP binding sites. As described previously, the binding sites differ at only two positions, and only one of the two is in the vicinity of the terminal KKTN side chains: namely Tyr33 ( $\beta'$ -COP) and its equivalent His31 ( $\alpha$ -COP). Supplementary Figure S1A shows the structure of the  $\beta'$ -COP/yWbp1 complex (the millimolar concentrations of peptide and  $\beta'$ -COP protein used for crystallization are more than sufficient to saturate the binding site;  $K_d = 170.9 \mu$ M). In the picture, we have overlapped the His31 side chain of  $\alpha$ -COP with Tyr33 of  $\beta'$ -COP. The juxtaposition of the Tyr33/His31 side chains and the threonine residue of the KKTN motif lead us to propose that an unfavourably close approach of the Tyr33 hydroxyl group impinges on the  $\beta$ -branched threonine residue (and hinders a favoured side-chain rotamer) and is responsible for lowering the affinity of the  $\beta'$ -COP/KKTN complex. This invokes a rather subtle interplay, such that the shorter His33 side chain of  $\alpha$ -COP is

more compatible with a  $\beta$ -branched residue at the  $-2$  position, all the more so because position 33 is conserved in  $\alpha$ -COP as either a histidine or phenylalanine residue across species (and is a Phe in *S. cerevisiae*  $\alpha$ -COP). Nevertheless, a peptide in which the threonine residue is replaced with serine, TFKKTN to TFKKSN, was tested for its affinity to  $\alpha$ -COP and  $\beta'$ -COP by ITC (Supplementary Figure S1B). Strikingly, the KKSNN peptide binds more tightly than KKTN to  $\beta'$ -COP by a factor of 1.8-fold, and a comparison of the (ratio of) affinities of  $\alpha$ -COP and  $\beta'$ -COP for KKxx motifs shows that the KKTN motif of yWbp1 favours  $\alpha$ -COP by a factor of 4.7, the KKSNN motif favours  $\alpha$ -COP by a factor of 2.2, and the KKLTV motif of p25 favours  $\alpha$ -COP by a factor of 2.5. The results confirm the interplay between the Tyr33/His31 residue at the binding site and the  $-2$  position of the dilysine motif, and they reveal that threonine is disfavoured by  $\beta'$ -COP at the  $-2$  position, hence the reliance of the yWbp1 retrieval motif on a functional  $\alpha$ -COP subunit for trafficking *in vivo* (Eugster *et al.*, 2004). In our subsequent analysis of the variant KxHxx motif, we found that  $\beta'$ -COP binds with far lower affinity than  $\alpha$ -COP to the KVHVQ sequence (Figure 4C; peptide PEDVspike in Table I), and the crystal structure of the  $\beta'$ -COP/KVHVQ complex (Figure 4B) shows the close approach of the  $-2$  valine residue to Tyr33. Collectively, these observations are consistent with the proposal that  $\beta'$ -COP disfavours dilysine motifs that contain a  $\beta$ -branched residue (I, T, or V) at the  $-2$  position.

#### Coatomer interaction with the non-canonical KxHxx retrieval motif

The KxHxx motif was identified as a functional retrieval signal present on the cytoplasmic tail of the spike protein of group 1 coronaviruses and of the SARS coronavirus (Lontok *et al.*, 2004). We determined the crystal structure of  $\beta'$ -COP bound to the KVHVQ sequence of the spike protein from porcine epidemic diarrhoea virus, abbreviated as PEDVspike in Table I (Shirato *et al.*, 2011). The PEDVspike motif binds very weakly to  $\beta'$ -COP, as described above, but crystallization drives together the complex (Figures 4B and C). The KxHxx sequence adopts a highly similar binding mode to the KxKxx motif. A slight positional shift of the entire histidine residue enables its side-chain  $\epsilon_2$  nitrogen atom to bond Asp206 of  $\beta'$ -COP at acidic patch 1.

To assess whether the shift is a conserved feature of the KxHxx binding mode, we solved the structures of additional  $\beta'$ -COP/KxHxx complexes. For this purpose, we identified conserved KxHxx sequences in the cytoplasmic C termini of the Insig-1 (KPHSD) and Insig-2 (KSHQE) proteins. The Insig proteins are ER localized and facilitate the retention of SCAP/sterol/SREBP complexes in the ER (Yang *et al.*, 2002), but a role for the Insig C terminus in active retrieval of the protein from post-ER membranes has not been considered; we leave this as an open question. ITC measurements showed that the KxHxx motifs of the Insig proteins bind to  $\alpha$ -COP and  $\beta'$ -COP (Figure 4C). Structures of  $\beta'$ -COP bound to the Insig-1 and Insig-2 sequences reveal a highly similar binding mode to the KxHxx motif of PEDVspike (Figure 4; Table I). Importantly, the position of the histidine residue is conserved in the three complexes. Figure 4A shows an overlap of the  $\beta'$ -COP/Insig-1 complex with the KxKxx sequence of Emp47p, and highlights the subtle shift in the position of the histidine residue relative to the  $-3$  lysine of Emp47p (note, e.g., the displacement of



**Figure 5** Distinctive binding mode of the RKxx motif and its specificity for  $\beta'$ -COP. **(A)** Close-up view of the  $\beta'$ -COP binding site showing the structure of the RKxx motif from Scyl1, coloured magenta (sequence RKLD; refined at 1.5 Å resolution). Overlapped is the structure of the KxKxx motif of Emp47p, coloured green. The molecular surface of  $\beta'$ -COP is coloured according to residue properties: acidic residues coloured red, basic side chains blue and hydrophobic side chains yellow. **(B)** Picture of the Scyl1 RKLD motif bound to  $\beta'$ -COP in a similar orientation to **(A)**, showing the bonding interactions made by the arginine residue of the RKLD motif (coloured magenta) to the backbone carbonyl oxygen atom of  $\beta'$ -COP residue Arg185 ( $\beta'$ -COP coloured cyan). The corresponding carbonyl group of  $\alpha$ -COP residue Arg213 points in the opposite direction, as shown in the overlapped structure (coloured pink). Side-chain atoms of Arg185, Arg213 (and Asp212) are omitted for clarity. A key salt bridge that maintains the local conformation of  $\beta'$ -COP is drawn between residues Glu184 and Arg163 (indicated with blue dot). **(C)** Isothermal calorimetry measurements for  $\alpha$ -COP and  $\beta'$ -COP binding to the GARKLD sequence of Scyl1 and variants thereof. The upper panel shows examples of the primary data; lower panel shows data fitted to binding isotherms. **(D)** Alignment of  $\alpha$ -COP and  $\beta'$ -COP sequences in the vicinity of  $\beta'$ -COP residue Arg163 (blue dot), the key salt-bridge residue drawn in **(B)**. This basic residue is conserved in  $\beta'$ -COP sequences, but changes to glutamine or glutamic acid in  $\alpha$ -COP and the salt bridge is lost as a result.

the  $\alpha$  position) that enables the histidine side-chain  $\epsilon 2$  nitrogen atom to occupy a very similar position to the lysine side-chain nitrogen atom, both then bonding to Asp206 of acidic patch 1.

These results indicate that acidic patch 1 of the binding site can accommodate histidine residues, and as a consequence the KxHxx motif mimics the KxKxx binding mode in interactions with coatomer.

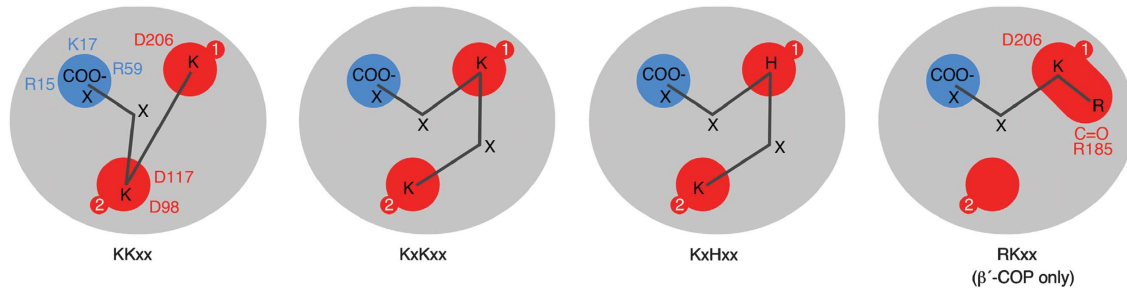
#### Specificity and unique binding mode of the RKxx motif

A CD8 reporter molecule tagged with the KKMP motif of adenoviral E19 protein is dependent on both lysine residues for ER retention, and among a series of substitutions of lysine with basic residues—HK, KH, RK and KR—only RKMP was shown to be a functional motif (Jackson *et al.*, 1990). Figure 5 illustrates the structural analysis of  $\beta'$ -COP bound to the RKLD motif of the Golgi-localized protein Scyl1 (Burman *et al.*, 2008). Surprisingly, the conformation of the RKxx motif does not mimic KKxx, and we suggest that an

arginine residue is slightly too long for optimal binding to acidic patch 1 in the KKxx mode of interaction. Instead, the RKxx motif mimics the KxKxx binding mode from position  $-4$  to the C terminus (Figure 5A). The arginine side chain adopts a unique binding position, outside the circumscribed dilysine-binding site (shown schematically in Figure 6). Figure 5B illustrates how the arginine residue is extended to form bonds via its guanidinium group to the backbone carbonyl oxygen atom of  $\beta'$ -COP residue Arg185. Interestingly, the corresponding carbonyl oxygen atom is not available to form this bond in  $\alpha$ -COP; instead, in all the  $\alpha$ -COP crystal structures the carbonyl group of Arg213 points away from the binding site ( $\psi$  angle of Arg213 is rotated  $\sim 165^\circ$  relative to the conformation in  $\beta'$ -COP; Figure 5B).

These structural observations suggest that the RKxx motif should favour  $\beta'$ -COP over  $\alpha$ -COP. ITC measurements show that the RKLD sequence binds to  $\beta'$ -COP ( $K_d = 20.4 \mu\text{M}$ ) with comparable affinity to KKxx and KxKxx motifs (Figure 5C). The AKLD and KRLD test sequences each lose  $\sim 4$ -fold





**Figure 6** Binding modes of the four classes of dilysine-type retrieval signals. Schematic diagram shows the binding modes of the different classes of retrieval signals related to the dilysine motif. The acid patches 1 and 2 are labelled and indicated in red; these patches are defined in the text and in Figure 3A. The basic patch that recognizes the terminal carboxylate group is coloured blue.  $\alpha$ -COP and  $\beta'$ -COP recognize all the signals, except for the RKxx motif which is specific for  $\beta'$ -COP.

affinity, indicative of suboptimal binding relative to the RKLD motif. Strikingly, the RKLD sequence binds exceedingly weakly to  $\alpha$ -COP ( $K_d > 1$  mM), revealing a particular incompatibility of this motif with the  $\alpha$ -COP binding site, the structural correlate of which we infer is the clash of the arginine side chain with the backbone amide group of  $\alpha$ -COP residue Arg213 (Figure 5B).

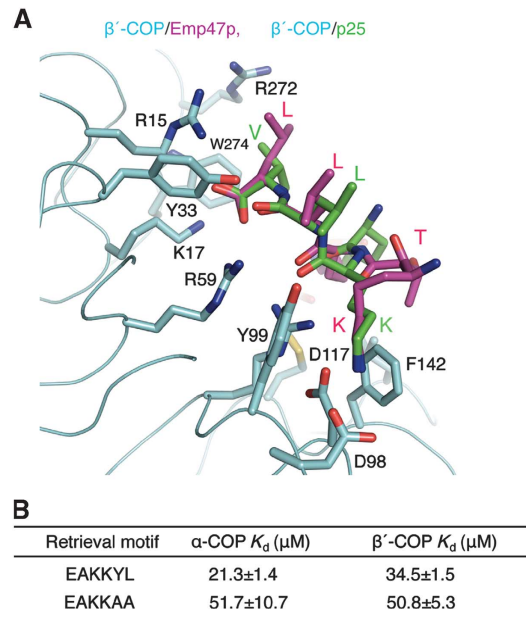
The concordance of the binding measurements and structural observations indicate that the  $\beta'$ -COP/RKxx crystal structure, and in particular the bonding arrangement of the  $-4$  arginine residue, corresponds to the binding mode in solution. And it is of note that the RKxx motif adopts the same conformation in the two distinct complexes in the crystal asymmetric unit (Table I).

An analysis of structure and sequence conservation in this region of the  $\alpha$ -COP and  $\beta'$ -COP molecules suggests that the specificity of the RKxx motif for  $\beta'$ -COP is a phylogenetically conserved feature (Figures 5B and D). Binding of the RKxx motif relies on the orientation of the carbonyl group of Arg185 (*S. cerevisiae*  $\beta'$ -COP numbering), and this backbone orientation is imposed in turn by the salt bridge between residues Glu184 and Arg163 (Figure 5B). The salt-bridge pairing is highly conserved in  $\beta'$ -COP sequences: Glu184 is conserved as a glutamic acid, and Arg163 as an arginine, or as lysine in *Caenorhabditis* species. By contrast, the salt bridge is lost in  $\alpha$ -COP sequences: the position corresponding to  $\beta'$ -COP Glu184 is highly conserved as an aspartic acid (or threonine in *Saccharomyces* spp.  $\alpha$ -COP); however, as shown in Figure 5D, the position corresponding to  $\beta'$ -COP Arg163 changes in  $\alpha$ -COP to a conserved glutamine or glutamic acid residue.

Thus, the RKxx motif adopts a distinctive binding mode, and utilizes an expanded region of the binding site that is unique to the  $\beta'$ -COP subunit.

### Sequence context of the dilysine motif

A comparative assessment of the crystal structures and ITC data suggests that the sequence context of the KKxx and KxKxx motifs plays a relatively minor role in coatomer recognition. The analysis of yWbp1 revealed one exception to this, namely the incompatibility of  $\beta$ -branched residues at the  $-2$  position with tight binding to  $\beta'$ -COP. Additionally, we speculate that aspartic and glutamic acid residues may be favoured at the  $-1$  position, to facilitate a charge interaction with Arg272 ( $\beta'$ -COP numbering), perhaps explaining the tight binding of the hWbp1 KEKSD motif (Figure 3D;  $K_d[\beta'$ -COP] = 3.4  $\mu$ M). These exceptions aside, the binding



**Figure 7**  $\alpha$ -COP and  $\beta'$ -COP are designed to bind dilysine motifs with little regard for sequence context. (A) Close-up view of the  $\beta'$ -COP binding site showing an overlap of the KKLK motif of p25 (coloured green) and the KTKLL motif of Emp47p (magenta). The picture is meant to emphasize how the  $-1$  and  $-2$  residues point away from the dilysine-binding site. (B) Summary of binding measurements by isothermal calorimetry of KKYL and KKA motifs versus  $\alpha$ -COP and  $\beta'$ -COP.

site seems to be designed to bind dilysine motifs without regard for sequence context. As illustrated in Figure 7A, the side chains of the  $-1$  and  $-2$  residues (KTKLL of Emp47p coloured magenta; KKLK of p25 coloured green) are directed away from the binding site; recognition of these residues is instead via their backbone carbonyl groups. Likewise, the  $-4$  position of the KxKxx motif is directed away from the binding site (see, e.g., KEKSD in Figure 3D).

A study of ER retrieval signals in mammalian cells took a combinatorial approach to generate a series of K(x)Kxx motifs, and obtained a wealth of information on the trafficking itinerary of reporter molecules bearing these C-terminal sequences (Zerangue *et al.*, 2001). The results showed that certain dilysine-containing sequences strongly enforced ER localization (e.g., KKYL and KKTN), whereas others did not (e.g., KKYY, KKTA and KKSP). With hindsight, it seems likely that many of the K(x)Kxx sequences generated

in the combinatorial library encode additional trafficking information, such as the C-terminal (di)hydrophobic signal that links to the COPII coat machinery (Nufer *et al*, 2002; Miller *et al*, 2003). For example, the screen of Zerangue *et al* (2001) reported that proline at position  $-1$  is strongly disfavoured in a functional retrieval motif, yet the adenoviral E19 KKMP sequence is a bona fide retrieval signal. We tested the affinity of two additional sequences described by Zerangue *et al* (2001): the KKYL sequence reported as an especially potent retrieval signal binds to both  $\alpha$ -COP and  $\beta'$ -COP, with comparable affinity to other K(x)Kxx sequences; however, the KKAA motif reported not to function as a retrieval signal does also bind to both COP subunits in our ITC assay (Figure 7B).

We surmise that residues surrounding the lysines do not play a dominant role in binding to  $\alpha$ -COP and  $\beta'$ -COP subunits, and we suggest that this may allow cargo C termini to nest additional topogenic information in their K(x)Kxx sequences without compromising the interaction with coatomer.

## Discussion

In this study, we establish how the four classes of dilysine-type retrieval motifs are recognized by coatomer subunits. Our focus was on naturally occurring motifs, rather than on K(x)Kxx sequence permutations that yield the highest affinity for coatomer. Consequently, there are no doubt rules and restrictions on dilysine-motif recognition not accounted for in our analysis. Nevertheless, the range of structures and sequences we have explored offer a fairly thorough description of the binding modes, recognition bonding interactions and certain sequence restrictions, to provide a heuristic basis on which to further explore the role of coatomer and dilysine signals in retrograde trafficking.

Here, we briefly summarize these rules with reference to Figure 6. A dilysine-type motif adopts one of three possible binding modes (the KxKxx and KxHxx motifs are considered as equivalent), dependent on the combination of the two basic residues—KxK, KK, RK—and independent of sequence context. For all binding modes, there is an absolute requirement for a basic residue at the  $-3$  position of the motif, because the recognition of the C-terminal carboxylate group controls the distance to the acidic patches 1 and 2 on  $\alpha$ -COP and  $\beta'$ -COP.

The general preference for lysine residues in the retrieval signal does not arise from an absolute discrimination of the acidic patches against arginine and histidine side chains, as demonstrated by the finding that ER localization of hWbp1 protein is maintained when the KxKxx motif is replaced by the non-natural RxKxx, KxRxx, HxHxx, or the coronavirus KxHxx motif (Hardt and Bause, 2002). Favouritism towards lysine is also due to the binding modes of the motifs (enforced by interactions with peptide backbone carbonyl groups), which position arginine and histidine side chains suboptimally at the acidic patches. This effect seems to be the most pronounced for KKxx-family motifs, as reporter-based analysis shows that KRxx, KHxx and HKxx motifs are not functional retrieval signals (Jackson *et al*, 1990); only KKxx and RKxx motifs function as such, and as we have demonstrated RKxx does so by adopting a KxKxx-like binding mode and positioning its arginine side chain beyond the acidic patches.

Our results shed light on the dilysine-binding specificity of the coatomer subunits. They demonstrate that  $\alpha$ -COP and

$\beta'$ -COP do not discriminate between KKxx and KxKxx motifs as originally proposed by a study of Emp47p and Wbp1 trafficking in yeast (Eugster *et al*, 2004). Instead, discrete structural differences at the binding sites allow only  $\alpha$ -COP to bind tightly to dilysine motifs with a  $\beta$ -branched residue at the  $-2$  position, and allow only  $\beta'$ -COP to accept the rare RKxx signal.

The major unresolved issue in this context is the functional organization of  $\alpha$ -COP and  $\beta'$ -COP in the polymerized COPII coat. Our previous structural analysis of  $\alpha\beta'$ -COP implicated a triskelion vertex formed by  $\beta$ -propeller contacts that are highly similar to those in the COPII cage (Stagg *et al*, 2006; Fath *et al*, 2007; Lee and Goldberg, 2010). However, as we emphasize in Supplementary Figure S2A, the  $\alpha\beta'$ -COP crystal structure included only a fragment of  $\alpha$ -COP so that the location of the  $\beta$ -propeller domain of  $\alpha$ -COP remains unclear. A direct attack on this problem will require electron microscopy, and a recent study of reconstituted COPI-coated vesicles using cryo-electron tomography (cryo-ET) has pointed the way forward (Faini *et al*, 2012). This study identified a three-fold symmetry centre in the density distribution for the COPI cage, consistent with the  $\alpha\beta'$ -COP triskelion nucleus. However, fitting of additional solenoid regions of  $\alpha\beta'$ -COP and of the AP2 crystal structure (as a model for the  $\beta\delta/\gamma\zeta$ -COP subcomplex) was unsatisfactory, and clearer insights into the COPI architecture await an advance in the resolution of the cryo-ET analysis. A key insight from the cryo-ET study is that protein features near to the three-fold symmetry centre reside close to the membrane surface (Faini *et al*, 2012). This is important, because most dilysine motifs are presented on cytoplasmic tail sequences that are intriguingly short. To assess this issue, we modelled the binding of a dilysine motif to  $\beta'$ -COP in the  $\alpha\beta'$ -COP triskelion on a membrane surface. As Supplementary Figure S2 illustrates, the dilysine motif binds very close to the contact surfaces at the triskelion centre: the KKxx motif fits snugly between adjacent  $\beta'$ -COP surfaces, whereas the KxKxx motif clashes somewhat with a neighbouring surface (however, we did not evaluate these contacts in detail because the triskelion interface is distorted from proper three-fold symmetry by crystal packing interactions). Importantly, the first lysine residue of the KKxx motif resides  $\sim 25$  Å from the membrane surface, according to this model (Supplementary Figure S2C). This is just compatible with the lengths of cargo cytoplasmic tails: for example, the number of residues between the membrane surface and the first lysine residue is 13 in yWbp1, 10 in p25 and 12 in adenoviral E19. This striking juxtaposition of coat proteins and cargo signals has implications for the mechanism, still unresolved, through which coatomer couples cargo packaging to COPI-vesicle formation (Bremser *et al*, 1999).

## Materials and methods

### Protein and peptide production

The expression and purification of full-length *B. taurus*  $\alpha\beta'\epsilon$ -COP subcomplex was carried out as described previously (Lee and Goldberg, 2010; Yu *et al*, 2012). Two dimeric forms of *S. cerevisiae*  $\alpha\beta'$ -COP complex—one comprising  $\alpha$ -COP(1–818)/ $\beta'$ -COP(651–814) and the other  $\beta'$ -COP(1–814)/ $\alpha$ -COP(642–818)—were prepared from baculovirus-infected Hi-5 insect cells, with the proteins encoded on individual vectors (pFastBac HTB vector from Invitrogen) and the baculoviruses constructed using the DH10Bac bacmid procedure (Invitrogen). Insect cells were harvested 48 h post infection, lysed by sonication and centrifuged at  $100\,000 \times g$  for 90 min. Protein

complexes were purified by Nickel-affinity (NI-NTA, Invitrogen), ion-exchange (HiTrap Q, Invitrogen), and size-exclusion chromatography. All proteins were brought to 20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 5 mM dithiothreitol (DTT), and were flash frozen in liquid nitrogen.

Proteins for crystallization were produced in *E. coli*. The N-terminal  $\beta$ -propeller domains of *S. cerevisiae*  $\beta'$ -COP (residues 1–301) and quintuple-mutant *S. pombe*  $\alpha$ -COP (1–327; L181K, L185K, I192K, L196K and F197K) were expressed with an N-terminal hexa-histidine tag plus Smt3. Bacterial cells expressing target protein were lysed by French press and purified on a nickel-affinity column. The Smt3 tag was removed by Ulp1 protease cleavage and protein purified by anion exchange and size-exclusion chromatography. Mutant protein constructs were generated using the Phusion PCR method (New England Biolabs) and purified in the same manner as wild-type protein.

For pull-down experiments, the cytoplasmic regions of dilysine-containing proteins were fused to the C terminus of GST. Constructs were made in the pGEX-4T-1 vector using the Phusion mutagenesis kit.

High purity synthetic peptides were purchased from the Tufts University Core Facility. The sequences of the peptides used in crystallographic analyses and calorimetry experiments are listed in Table 1.

#### Dilysine motif pull-down assay

The GST pull-down experiments were performed as described in Mossessova *et al.* (2003), with several modifications. A saturating quantity of GST-dilysine fusion protein was incubated with 10  $\mu$ l of a 50% (v/v) slurry of glutathione sepharose 4B beads (GE Healthcare) for 30 min at 4°C. Beads were washed once with 500  $\mu$ l Buffer A (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM DTT and 0.1% Triton X-100), leaving ~60  $\mu$ g GST fusion protein bound to the beads. Beads were mixed with 0.1 mg/ml coatomer proteins and 5 mg/ml *E. coli* proteins (prepared as described in Mossessova *et al.*, 2003) in 100  $\mu$ l buffer A. The assay mix was incubated at 4°C for 45 min and washed once with 100  $\mu$ l buffer A. Proteins were eluted with 60  $\mu$ l SDS sample buffer, separated by 4–20% gradient SDS-PAGE and stained with Coomassie blue.

#### Isothermal calorimetry

Binding experiments were performed on a microCal ITC200 instrument (GE Healthcare) operated at 25°C. The protein and synthetic dilysine peptide samples were dialysed overnight against separate volumes of the same buffer: 20 mM Tris-HCl pH 7.5, 150 mM NaCl (peptides were dialysed using a 100–500 Da cutoff dialysis membrane). Measurements were taken over ~17 injections and data were fitted using Origin 7.0 software to estimate the equilibrium dissociation constant, stoichiometry and enthalpy of binding. Binding of peptides to  $\alpha$ -COP was tested using the quintuple-mutant *S. pombe*  $\alpha$ -COP(1–327), and binding to  $\beta'$ -COP was tested using the *S. cerevisiae*  $\beta'$ -COP(1–814)/ $\alpha$ -COP(624–818) dimer.

#### Crystallization and structure determination

For crystallization purposes, purified *S. cerevisiae*  $\beta'$ -COP(1–301) was concentrated to 30 mg/ml in 20 mM Tris-HCl pH 7.4, 150 mM NaCl and 5 mM DTT and flash frozen in liquid nitrogen. Crystals of protein/peptide complexes were grown by the hanging-drop method. For example, a 1- $\mu$ l solution containing 1 mM  $\beta'$ -COP and 1.5 mM p25 peptide was mixed with an equal volume of well solution comprising 0.1 M MES pH 6.5, 30% PEG-400. Crystals grew at room temperature in 4 days in the space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> (see Table 1 for peptide sequence and crystallographic details). Crystals were transferred into a cryoprotectant buffer containing well solution plus 20% glycerol, and frozen in liquid nitrogen. The  $\beta'$ -COP/p25 crystals diffracted synchrotron X-rays beyond 1.5 Å resolution.

## References

- Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D* **66**: 213–221
- Aguilera-Romero A, Kaminska J, Spang A, Riezman H, Muñiz M (2008) The yeast p24 complex is required for the formation of COPI retrograde transport vesicles from the Golgi apparatus. *J Cell Biol* **180**: 713–720
- Bremser M, Nickel W, Schweikert M, Ravazzola M, Amherdt M, Hughes CA, Söllner TH, Rothman JE, Wieland FT (1999)
- Diffraction data for the  $\beta'$ -COP/p25 crystals were collected at NE-CAT beamline 24ID-C and E of the Advanced Photon Source at the Argonne National Laboratory. Data (for all the crystals) were collected using an X-ray wavelength of 0.979 Å, and processed using the HKL2000 suite of programs (Otwinowski and Minor, 1997). The structure was solved by molecular replacement with the program Phaser within the Phenix program suite (Adams *et al.*, 2010) using the N-terminal  $\beta$ -propeller of  $\beta'$ -COP as the search model (Lee and Goldberg, 2010). The model was refined using Phenix. Subsequently, the p25 peptide model was built manually into high-resolution difference Fourier maps. The density for the p25 peptide (FEAKKLV) is excellent other than the first two residues, FE, which are outside the binding site and disordered in the maps. Crystallographic data and refinement results are summarized in Table 1.
- All the other structures of  $\beta'$ -COP/peptide complexes were determined in a similar manner. However, they almost all grew as distinct crystal forms, from different crystallization conditions. In brief, the composition of the crystallization well solutions was as follows: yWbp1, 0.1 M Hepes pH 7.5 and 20% PEG-10000; hWbp1, 0.1 M Hepes pH 6.3 and 23% PEG-6000; Emp47p, 0.1 M MES pH 6.0 and 20% PEG-6000; Scyl1, 0.1 M HEPES pH 7.1 and 23% PEG-3350; PEDVspike, 0.1 M MES pH 6.2 and 19% PEG-4000; Insig-1, 0.1 M MES pH 6.5 and 15% PEG-20000; Insig-2, 0.2 M NaSCN and 20% PEG-3350.
- Crystallization of  $\alpha$ -COP/peptide complexes used the *S. pombe* N-terminal  $\beta$ -propeller domain (residues 1–327) with five point mutations in a loop region (L181K, L185K, I192K, L196K and F197K). This 28-residue element was defined previously as a flexible surface loop based on proteolytic analysis of the mammalian protein (Lee and Goldberg, 2010). The purified protein was concentrated to 3.2 mg/ml in 20 mM Tris-HCl pH 7.4, 150 mM NaCl, 5 mM DTT and flash frozen. Synthetic peptides (adenoviral E19 and Emp47p) were added in a 1.5-fold molar excess to the  $\alpha$ -COP protein, and crystals were grown by the hanging-drop method. For  $\alpha$ -COP/E19 crystals, the well solution contained 0.2 M K Na tartrate plus 20% PEG-3350, and for the  $\alpha$ -COP/Emp47p crystals it contained 0.2 M ammonium acetate plus 20% PEG-3350. The apo form of *S. pombe*  $\alpha$ -COP was crystallized from 0.2 M tri-sodium citrate plus 20% PEG-3350. Crystals of  $\alpha$ -COP/Emp47p and apo- $\alpha$ -COP were cryoprotected using well solution plus 25% glycerol. The  $\alpha$ -COP/E19 crystals were transferred to paraffin oil prior to freezing, and diffracted X-rays to 1.9 Å resolution.  $\alpha$ -COP crystal structures were solved using the  $\beta'$ -COP(1–301) structure as the search model. Model building for protein and peptide was carried out as for  $\beta'$ -COP complexes.

#### Supplementary data

Supplementary data are available at *The EMBO Journal* Online (<http://www.embojournal.org>).

## Acknowledgements

We thank staff of the NE-CAT beamlines at the Advanced Photon Source of the Argonne National Laboratory for access to synchrotron facilities. We thank all the members of our laboratory for advice and insights on coatomer behaviour and protein biochemistry experiments.

*Author contributions:* JG and WM designed the experiments and analysed the data. WM performed the experiments. JG and WM prepared the manuscript.

## Conflict of interest

The authors declare that they have no conflict of interest.

- Coupling of coat assembly and vesicle budding to packaging of putative cargo receptors. *Cell* **96**: 495–506
- Brünger AT (1998) Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D* **54**: 905–921
- Burman JL, Bourbonniere L, Philie J, Stroh T, Dejgaard SY, Presley JF, McPherson PS (2008) Scyl1, mutated in a recessive form of spinocerebellar neurodegeneration, regulates COPI-mediated retrograde traffic. *J Biol Chem* **283**: 22774–22786
- Cosson P, Letourneur F (1994) Coatomer interaction with di-lysine endoplasmic reticulum retention motifs. *Science* **263**: 1629–1631
- Eugster A, Frigerio G, Dale M, Duden R (2004) The alpha- and beta'-COP WD40 domains mediate cargo-selective interactions with distinct di-lysine motifs. *Mol Biol Cell* **15**: 1011–1023
- Faini M, Prinz S, Beck R, Schorb M, Riches JD, Bacia K, Brügger B, Wieland FT, Briggs JA (2012) The structures of COPI-coated vesicles reveal alternate coatomer conformations and interactions. *Science* **336**: 1451–1454
- Fath S, Mancias JD, Bi X, Goldberg J (2007) Structure and organization of coat proteins in the COPII cage. *Cell* **129**: 1325–1336
- Fiedler K, Veit M, Stamnes MA, Rothman JE (1996) Bimodal interaction of coatomer with the p24 family of putative cargo receptors. *Science* **273**: 1396–1399
- Gaynor EC, te Heesen S, Graham TR, Aebi M, Emr SD (1994) Signal-mediated retrieval of a membrane protein from the Golgi to the ER in yeast. *J Cell Biol* **127**: 653–665
- Hardt B, Bause E (2002) Lysine can be replaced by histidine but not by arginine as the ER retrieval motif for type I membrane proteins. *Biochem Biophys Res Commun* **291**: 751–757
- Jackson LP, Lewis M, Kent HM, Edeling MA, Evans PR, Duden R, Owen DJ (2012) Molecular basis for recognition of dilysine trafficking motifs by COPI. *Dev Cell* **23**: 1255–1262
- Jackson MR, Nilsson T, Peterson PA (1990) Identification of a consensus motif for retention of transmembrane proteins in the endoplasmic reticulum. *EMBO J* **9**: 3153–3162
- Jackson MR, Nilsson T, Peterson PA (1993) Retrieval of transmembrane proteins to the endoplasmic reticulum. *J Cell Biol* **121**: 317–333
- Lee C, Goldberg J (2010) Structure of coatomer cage proteins and the relationship among COPI, COPII and clathrin vesicle coats. *Cell* **142**: 123–132
- Letourneur F, Gaynor EC, Hennecke S, Demolliere C, Duden R, Emr SD, Riezman H, Cosson P (1994) Coatomer is essential for retrieval of dilysine-tagged proteins to the endoplasmic reticulum. *Cell* **79**: 1199–1207
- Lewis MJ, Pelham HR (1992) Ligand-induced redistribution of a human KDEL receptor from the Golgi complex to the endoplasmic reticulum. *Cell* **68**: 353–364
- Lontok E, Corse E, Machamer CE (2004) Intracellular targeting signals contribute to localization of coronavirus spike proteins near the virus assembly site. *J Virol* **78**: 5913–5922
- Lowe M, Kreis TE (1995) In vitro assembly and disassembly of coatomer. *J Biol Chem* **270**: 31364–31371
- Miller EA, Beilharz TH, Malkus PN, Lee MC, Hamamoto S, Orci L (2003) Multiple cargo binding sites on the COPII subunit Sec24 ensure capture of diverse membrane proteins into transport vesicles. *Cell* **114**: 497–509
- Mossessova E, Bickford LC, Goldberg J (2003) SNARE selectivity of the COPII coat. *Cell* **114**: 483–495
- Nilsson T, Jackson M, Peterson PA (1989) Short cytoplasmic sequences serve as retention signals for transmembrane proteins in the endoplasmic reticulum. *Cell* **59**: 707–718
- Nufer O, Guldbrandsen S, Degen M, Kappeler F, Paccaud JP, Tani K, Hauri HP (2002) Role of cytoplasmic C-terminal amino acids of membrane proteins in ER export. *Cell* **59**: 707–718
- Otwinowski W, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**: 307–326
- Sandmann T, Herrmann JM, Dangjel J, Schwarz H, Spang A (2003) Suppression of coatomer mutants by a new protein family with COPI and COPII binding motifs in *Saccharomyces cerevisiae*. *Mol Biol Cell* **14**: 3097–3113
- Schröder-Köhne S, Letourneur F, Riezman H (1998) Alpha-COP can discriminate between distinct, functional di-lysine signals in vitro and regulates access into retrograde transport. *J Cell Sci* **111**: 3459–3470
- Shirato K, Maejima M, Matsuyama S, Ujike M, Miyazaki A, Takeyama N, Ikeda H, Taguchi F (2011) Mutation in the cytoplasmic retrieval signal of porcine epidemic diarrhea virus spike (S) protein is responsible for enhanced fusion activity. *Virus Res* **161**: 188–193
- Stagg SM, Gürkan C, Fowler DM, LaPointe P, Foss TR, Potter CS, Carragher B, Balch WE (2006) Structure of the Sec13/31 COPII coat cage. *Nature* **439**: 234–238
- Townsley FM, Pelham HR (1994) The KKXX signal mediates retrieval of membrane proteins from the Golgi to the ER in yeast. *Eur J Cell Biol* **64**: 211–216
- Yang T, Espenshade PJ, Wright ME, Yabe D, Gong Y, Aebersold R, Goldstein JL, Brown MS (2002) Crucial step in cholesterol homeostasis: sterols promote binding of SCAP to INSIG-1, a membrane protein that facilitates retention of SREBPs in ER. *Cell* **110**: 489–500
- Yu X, Breitman M, Goldberg J (2012) A structure-based mechanism for Arf1-dependent recruitment of coatomer to membranes. *Cell* **148**: 530–542
- Zerangue N, Malan MJ, Fried SR, Dazin PF, Jan YN, Jan LY, Schwappach B (2001) Analysis of endoplasmic reticulum trafficking signals by combinatorial screening in mammalian cells. *Proc Natl Acad Sci USA* **98**: 2431–2436