



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Alisaraie, L; Rouiller, I

Title:

Full-length structural model of RET3 and SEC21 in COPI: Identification of binding sites on the appendage for accessory protein recruitment motifs

Date:

2012-07-01

Citation:

Alisaraie, L. & Rouiller, I. (2012). Full-length structural model of RET3 and SEC21 in COPI: Identification of binding sites on the appendage for accessory protein recruitment motifs. *Journal of Molecular Modeling*, 18 (7), pp.3199-3212. <https://doi.org/10.1007/s00894-011-1324-9>.

Persistent Link:

<https://hdl.handle.net/11343/270393>

License:

[CC BY-NC](#)

# Full-length structural model of RET3 and SEC21 in COPI: identification of binding sites on the appendage for accessory protein recruitment motifs

Laleh Alisaraie · Isabelle Rouiller

Received: 14 July 2011 / Accepted: 25 November 2011 / Published online: 14 January 2012  
© The Author(s) 2012. This article is published with open access at Springerlink.com

**Abstract** COPI, a 600 kD heptameric complex (consisting of subunits  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ ,  $\zeta$ , and  $\beta'$ ) “coatamer,” assembles non-clathrin-coated vesicles and is responsible for intra-Golgi and Golgi-to-ER protein trafficking. Here, we report the three-dimensional structures of the entire sequences of yeast Sec21 ( $\gamma$ -COPI mammalian ortholog), yeast Ret3 ( $\zeta$ -COPI mammalian ortholog), and the results of successive molecular dynamics investigations of the subunits and assembly based on a protein–protein docking experiment. The three-dimensional structures of the subunits in complexes indicate the residues of the two subunits that impact on assembly, the conformations of Ret3 and Sec21, and their binding orientations in the complexed state. The structure of the appendage domain of Sec21, with its two subdomains—the platform and the  $\beta$ -sandwich, was investigated to explore its capacity to bind to accessory protein recruitment motifs. Our study shows that a binding site on the platform is capable of binding the Eps15 DPF and epsin DPW2 peptides, whereas the second site on the platform and the site on the  $\beta$ -sandwich subdomain were found to selectively bind to the amphiphysin FXDXF and epsin DPW1 peptides, respectively. Identifying the regions of both the platform and sandwich subdomains involved in binding each peptide motif clarifies the mechanism through which the appendage domain of Sec21 engages with the accessory proteins during the trafficking process of non-clathrin-coated vesicles.

**Keywords** Protein trafficking · Coatamer subunits · Homology modeling · Molecular dynamics · Docking · Accessory protein motifs

## Introduction

Cellular membranes serve as barriers to the external environment and organize the interior of the eukaryotic cell into biochemically distinct compartments in which specific cellular functions are performed. The transport of proteins (cargo) between discontinuous subcellular compartments relies on the assembly and disassembly of multilayered coat protein complexes onto cellular membranes. The assembly of these complexes induces vesicle budding, in which cargo proteins are captured and transported in-between cellular compartments. To date, three major types of vesicles have been described: clathrin-coated vesicles, COPI (coat protein complex I)-coated vesicles, and COPII (coat protein complex II)-coated vesicles. Clathrin-coated vesicles bud from the plasma membrane and from the trans-Golgi network (TGN) to the endosome. COPI-coated vesicles traffic primarily from the cis-Golgi to the endoplasmic reticulum (ER) and between Golgi cisternae. COPII-coated vesicles traffic from the ER to the Golgi apparatus [1].

The COPI-coated vesicles form by recruitment on the membrane of two cytosolic components: an ARF-family G protein (such as ARF-1) and the ~600 kDa coatamer complex composed of seven coat proteins ( $\zeta$ ,  $\beta$ ,  $\beta'$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ , and  $\zeta$ ) [2, 3]. Prior to fusion with its target membrane, each COPI vesicle loses its coat; ARF1 and COPI are released back into the cytosol. This is initiated when ARF1 hydrolyses its bound GTP [3]. GTP hydrolysis is greatly increased by GTPase-activating proteins (ARF-GAPs), such as yeast Gcs1p (ARFGAP1 mammalian ortholog) and Gls3p

**Electronic supplementary material** The online version of this article (doi:10.1007/s00894-011-1324-9) contains supplementary material, which is available to authorized users.

L. Alisaraie · I. Rouiller (✉)  
Department of Anatomy and Cell Biology, McGill University,  
Strathcona Building, 3640 University Street,  
Montreal QC H3A 2B2, Canada  
e-mail: isabelle.rouiller@mcgill.ca

(ARFGAP2 mammalian ortholog). Under high salt conditions, the COPI complex can be disassembled into two subcomplexes, F (composed of the  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$  subunits) and B (composed of the  $\alpha$ ,  $\beta'$ , and  $\epsilon$  subunits) [4].

Proteins in the coat complexes can generally be subdivided into (i) the cage-forming proteins, which induce curvatures into the membrane that lead to the formation of coated buds, and (ii) the adaptor proteins, which bind to the scaffolding proteins and membrane-bound cargo receptors, thereby mediating selective recruitment into the vesicles. For clathrin-coated vesicles, the cage-forming proteins are the clathrin light and heavy chains [2]. The predominant adaptor protein (AP) complex found in clathrin coat is AP2, which belongs to the family of heterotetrameric AP complexes [5]. The AP2 complex, a typical member of AP complex, is composed of two large subunits (AP2- $\alpha$  and AP2- $\beta$ 2, 100–130 kDa), which contain an N-terminal trunk domain, a linker, a C-terminal appendage domain, a medium subunit called AP2- $\mu$ 2 (~50 kDa), and a small subunit called AP2- $\sigma$ 2 (~20 kDa). The N-terminal trunk domains of AP2- $\alpha$  and AP2- $\beta$ 2, together with the subunits AP2- $\mu$ 2 and AP2- $\sigma$ 2, assemble into a compact “core,” while the two appendages of the two large subunits are joined to the core through flexible linkers [6, 7]. To fulfill its function as an adaptor complex, AP2 interacts in a dynamic fashion (i) with internalization motifs on the cargo proteins via its  $\mu$ 2 and  $\delta$ 2 subunits, (ii) with the plasma membrane PIP (phosphatidyl inositol-4,5-bisphosphate) via the  $\alpha$ ,  $\beta$ 2, and  $\mu$ 2 subunits [8], (iii) with clathrin via its  $\beta$ 2 appendage [9] and via a clathrin box [10] found in the AP2- $\beta$ 2 linker [11], and (iv) with the sequence motifs DP[FW], FxDxF, and Wxx[FW], which are found in many accessory/regulatory proteins (such as Eps15, epsin, AP180, and Dab2, among many others) in the vesicle assembly/disassembly and endocytosis pathway via the appendages of the  $\alpha$ -AP2 and  $\beta$ 2-AP2 [12]. There are three different types of binding sites on the appendages [13]. The first one is a hydrophobic pocket around a conserved tryptophan within the platform subdomain in  $\alpha$ -AP2,  $\beta$ 2-AP2,  $\alpha$ -adaptin, and  $\beta$ 2-adaptin which recognizes the DPF/W or FXDXF peptides [9, 14, 15] and [DE]nX1-2FXX[FL]XXR [16] from accessory proteins [13]. The two other binding sites are on the sandwich (immunoglobulin-like) subdomain: one, found on the  $\alpha$ -AP2 appendage, recognizes WXX[FW]X[DE]n; the other, on the opposite side of the sandwich subdomain of  $\beta$ 2-AP2, recognizes FXn[FW]n [13, 17].

The function of the subunits of COPI and COPII complexes is less understood than those found in clathrin-coated vesicles, but the general principles are starting to emerge. For COPII, cryoelectron microscopy identified the subunits Sec13–31 as the lattice-forming complexes and the subunits Sec23–24 as the cargo adaptors [18, 19]. For COPI, recent X-ray crystallography studies have shown that the  $\alpha$ -COPI

and  $\beta'$ -COPI subunits crystallize as a triskelion [20], suggesting that they are the cage components of the COPI-coated vesicles. The heterotetrameric arrangement of the AP complexes is believed to be mirrored by that of the subunits of the F subcomplex ( $\zeta$ ,  $\beta$ ,  $\gamma$ , and  $\delta$ ) in COPI [21, 22]. In the F subcomplex,  $\zeta$ -COP is the only subunit that interacts with the large subunit  $\gamma$ -COP and is required for the interaction of  $\gamma$ -COP with the other large subunit  $\beta$ -COP [22]. An NMR study shows that the truncated human  $\zeta$ -COP (residues 1–149) presents general folding that is very similar to the crystal structure of the  $\sigma$ -subunit in AP1 and AP2 [23]. X-ray crystallography studies of the carboxyl-terminal region (residues 555–874) showed that the appendage of  $\gamma$ -COPI is structurally similar to the appendage in  $\alpha$ -AP2 and  $\beta$ 2-AP2, and that the yeast ARFGAP Glo3p (mammalian ortholog ARFGAP2) binds in the hydrophobic pocket of the platform subdomain of the  $\gamma$ -COPI appendage [24, 25]. A second site of interaction containing a patch of highly hydrophobic potential that mediates the interaction with the B subcomplex via  $\epsilon$ -COPI has been suggested to exist in the  $\beta$ -sandwich subdomain of the  $\gamma$ -COPI appendage [25, 26].

In this study, we present the homology model of the entire sequence (935 residues) of Sec21 (yeast homolog to mammalian  $\gamma$ -COPI), including the trunk and the appendage domains as well as the linker connecting the appendage to the trunk domain. We also present the full-length (189 residues) Ret3 (yeast homolog to mammalian  $\zeta$ -COPI), molecular dynamics simulations of Sec21 and Ret3 for 6 ns of simulation time, and protein–protein docking studies of the complexed Sec21 and Ret3 subunits. The complex of these two subunits, obtained from a docking experiment, was further studied by performing 20 ns of molecular dynamics simulations. The model of Sec21 was examined to explore the possible role of the appendage in the protein-trafficking mechanism. Three binding sites for accessory protein recruitment motifs were identified: (i) a site on the platform capable of binding the Eps15 DPF (GSDPFK) and epsin DPW2 (FSDPWC) peptides [27], (ii) another site also on the platform that was found to selectively bind to the amphiphysin FXDXF peptide, and (iii) a third site, in the  $\beta$ -sandwich subdomain, which was found to selectively bind to the epsin DPW1 (GSDPWK) peptide. Ent1 (Q12518, FGSENCVLWC), Ent1p (C8Z6E1, GSENCVLWCR), Pan1p (C8ZAQ5, GSSNLVEPRATPFQ), and Ent2 (Q05785, GSNNPFSDMDNLERQK) all belong to epsin family [28], which contain the epsin N-terminal homology domain, and are the yeast homologs of epsin DPW2, epsin DPW1, and Eps DPF. The sequences of the peptides in yeast and mammals show similarities of 50–80%, with 20–50% identical residues. Thus, analyzing the interaction profiles of each accessory peptide, the role of the highly conserved motifs at their binding site in Sec21, and the dynamic

conformational variations of the involved protein residues has provided useful insight into the mechanism of action in non-clathrin-coated vesicles against accessory protein motif in the trafficking process.

## Materials and methods

### Structure prediction and homology modeling

The complete sequences of the Ret3 and Sec21 subunits for *Saccharomyces cerevisiae* were retrieved from the Swiss-Prot-TREMBL database [28, 29] (identification codes: P53600 and P32074, respectively). The NMR structure of the truncated human  $\zeta$ -COP149 (2HF6.pdb) [23] and the crystal structures of the clathrin adaptor protein 1 core (1W63.pdb) [28], the adapter-related protein complex 2 (2JKR.pdb) [30], and the human  $\gamma$ -COP appendage domain structure (1R4X.pdb) [25] were employed as template structures to model the trunk and appendage of each homolog subunit in the yeast coatomer.

For the docking experiment, the atomic coordinates of the peptides were retrieved from the crystal structures of the complexes of the AP2 clathrin adaptor  $\alpha$ -appendage with epsin DPW1 peptide (1KYD.pdb) [27], epsin DPW2 peptide (1KY6.pdb) [27], Eps15 DPF (1KYU.pdb) [27], and amphiphysin FXDXF (1KY7.pdb) [27].

Protein structure predictions were obtained via the fold-recognition servers FUGUE [31], GenTHREADER [32], and 3D-PSSM [33], and the secondary structure prediction servers PSIPRED [34] and Phyre [35], as well as the Pcons consensus predictor [36] for fold recognition and to rank the best templates. The suggested templates were applied to model target proteins using the SWISS-MODEL workspace [37].

We first built structural models of the regions of Ret3 and Sec21 using the available template structures. In particular, the crystal structure of the truncated human  $\zeta$ -COP (2HF6.pdb) [23] was used for residues 1–150 of Ret3. The crystal structures of the clathrin adaptor protein 1 core (1W63.pdb chain A, H) [28] and the adapter-related protein complex 2 (2JKR.pdb) [30] were used for residues 1–267 and 275–584 of Sec21. The crystal structure of the human  $\gamma$ -COP appendage domain (1R4X.pdb) [25] was employed as the template for residues 600–874 of Sec21 (Figs. 1a and 2a).

We then modeled the missing fragments in the crystal structures of truncated human  $\zeta$ -COP (residues 150–189) [23], in the clathrin adaptor protein (residues 268–274), in the C-terminal region of human  $\gamma$ -appendage (residues 875–935) [25], as well as residues 584–599, part of the linker connecting the appendage to the trunk domain. Possible template structures for each missing string were found by searching the structure prediction servers mentioned above. From among the search results obtained for each missing fragment, the template with the highest available certainty,

sequence identity and similarity, and the best  $P$  and  $E$  values, was selected for model building. For residues 150–189 of  $\zeta$ -COP, the appropriate part from 3CJH.pdb was utilized, the appropriate part from 2DB0.pdb was used for 268–274 of Sec21, the appropriate part from 1Z65.pdb was obtained for 584–599, and the appropriate part from 1R4X.pdb was used for 875–935. In the third step, the qualified models of the missing parts were merged into the major parts of the modeled structures of their corresponding subunits via covalent bonds (Figs. 1a–b and 2a–c).

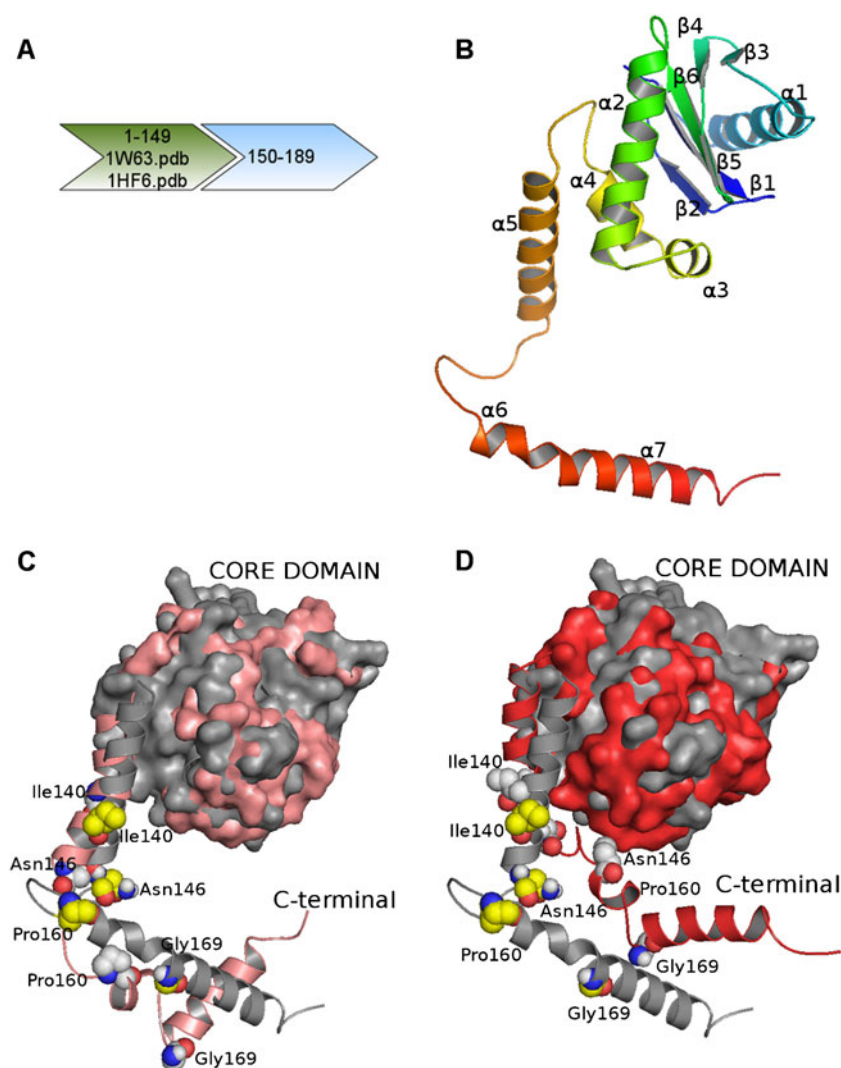
Following the removal of geometric strain and the elimination of special restraints on the modeled structures of the subunits with energy minimization and MD, the developed models of the whole of Ret3 as well as the trunk, linker, and appendage of Sec21 were evaluated by PROCHECK [38]. The PROCHECK [38] results showed that 94.8% and 96.9% of the backbone angles were in allowed regions, with  $G$  factors of  $-0.12$  and  $-0.7$  obtained for Ret3 and Sec21, respectively. ERRAT [39] is a program for calculation the “overall quality factor” of nonbonded atomic interactions. The cut-off value from ERRAT is 50; higher scores indicate high model quality. The ERRAT score for the model of Sec21 was 83.77 and that for the model of Ret3 was 85.08, which therefore indicate that they are high-quality models. Since the backbone angles and the nonbonded interactions of the models were all within their normal ranges, they can be safely used for further experiments in protein–protein docking and MD simulations.

A 2D representation of the assigned secondary structure of each subunit was prepared using STRIDE [40].

### Molecular dynamics simulations and setup

For the molecular dynamics (MD) simulations that were performed within the two individual experiments, each protein structure (Sec21 and Ret3) was embedded into a pre-equilibrated solvent box with water molecules consisting of simple point charges [41]. MD simulations were performed on the Krylov cluster of CLUMEQ using four computational nodes with eight Opteron 2.3 GHz CPUs. All simulations were performed using the GROMACS package, v.4.0.5. [42], periodic boundary conditions, and the GROMOS96 (G43a1) force-field to set up the parameters. The number of water molecules included in each box depended on the size of the protein structure. For Lennard–Jones and electrostatic interactions, a cutoff distance of 1.0 nm was assigned. The particle mesh Ewald algorithm [43–45] was used to calculate the electrostatic contributions to energies and forces. Bond lengths were constrained using the LINCS algorithm [46], which was also used to constrain hydrogen-bond lengths. The simulation was performed under normal pressure and conditions. Pressure and compressibility ( $\tau_p$ ) were set to constant values of 0.5 ps and  $4.5 \times 10^{-5} \text{ bar}^{-1}$ , respectively. The water and protein molecules were coupled separately to the thermal condition of 300 K with

**Fig. 1** Ret3 structural features and frames as extracted from MD trajectories. **a** Residues of Ret3 present in the template structure (*green block*), and the missing residues of the template (*blue block*). **b** Full-length structural model of Ret3. **c, d** Structural conformation of Ret3 at 0 ps (*gray*) superimposed onto the conformations at **c** 1000 ps (*pink*) and **d** 4200 ps (*red*)



a coupling constant ( $\tau_T$ ) of 0.1 ps [47]. The MD simulation of each protein system of our study was first energy minimized using a steepest descent algorithm followed by a conjugated gradient algorithm. All-bond progression of position restraint was then performed for 1500 ps. Dynamics simulation was used in the next phase to gradually release all of the constraints within 6 ns for the individual subunits of Ret3 and Sec21, and within 20 ns for the complex of the two subunits.

The atomic coordinates of Ret3 and Sec21 obtained from the individual MD experiments were utilized in docking experiments to simulate the three-dimensional structures of the complexed subunits.

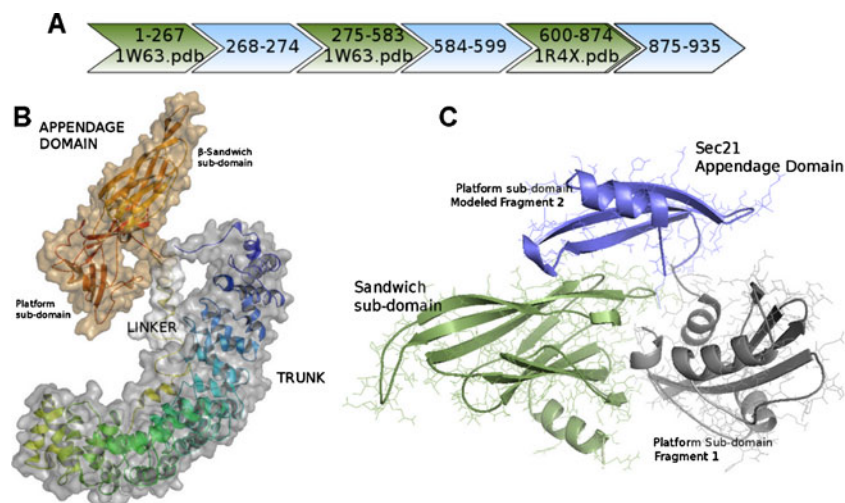
#### Protein–protein docking

HEX [48], a rigid-body docking program, was utilized to scan the conformations of the complementary shapes of the two subunits Ret3 and Sec21 in order to identify the orientations of the structures upon binding. It should be noted

that the first-rank solution from this experiment was chosen as the structure of the complex of the two subunits. The docking was repeated using ZDOCK [49], where the top five solutions had RMSDs of less than 2.5 Å with respect to the first-rank hit obtained from HEX [48]. To further evaluate the selected binding mode, docking experiments were followed by a 20 ns long MD simulation of the selected conformational pose of the complex structure.

The Sec21 structure was studied to determine the individual binding modes of the peptide ligands Eps15 DPF, epsin DPW2, epsin DPW1, and FXDXF when complexed with the appendage domain of Sec21. The regions equivalent to the binding sites of the same ligands on the  $\alpha$ -AP2,  $\beta$ 2-AP2, and  $\gamma$ -AP1 appendages [17, 27, 50] were identified by sequence alignment of the corresponding proteins [51].

The ZDOCK docking program [49] was employed to dock the peptide ligands and predict their putative binding pockets. The GHECOM grid-based HECOMi finder [52] was utilized as an additional approach to find the best binding site for each individual peptide.



**Fig. 2** Tertiary structure of Sec21. **a** Residues of Sec21 present in the template structure (*green block*), and the missing residues of the template (*blue block*). **b** Cartoon representation of the modeled structure of full-length Sec21, with trunk domain (residues 1–600), linker (residues 601–645), and appendage (646–935). **c** Model of the Sec21

appendage, with the sandwich subdomain and the fragmented platform subdomain. Fragment 1 is similar to the  $\gamma$ -appendage in  $\gamma$ -COP (*gray*), while the additional fragment of the appendage, residues 875–935, is fragment 2 (*blue*)

The potential binding sites on Sec21 include a  $\sim 14$  Å deep pocket consisting of the antiparallel sheets  $\beta 9$ – $\beta 14$  and  $\alpha 33$ , as well as a  $\sim 20$  Å-deep pocket consisting of  $\beta 3$ ,  $\beta 4$ , the  $\beta 8$ – $\beta 9$  loop,  $\beta 12$ ,  $\beta 13$ , and a  $3_{10}$  helix. The latter is the effect of the additional fragment in the modeled structure of the Sec21 appendage with amino acids 874–935 in the platform subdomain. In addition, a third pocket was identified that is located on the surface of the sandwich subdomain, consisting of the  $\alpha 32$ – $\beta 3$  loop,  $\beta 6$  and  $\beta 7$ , the  $\beta 6$ – $\beta 7$  loop, and  $\beta 10$ .

## Results and discussion

### Ret3 subunit

The entire Ret3 was modeled as described in the “**Materials and methods**” section. Despite the low sequence identities (<21%) between the subunit of the COPI and the APs (Fig. S1), phylogenetic analysis of both the large and the small/medium subunits indicated a common ancestor for all components of the heterotetrameric adaptor complexes AP1, 2, 3, including  $\zeta$ -COP. Accordingly, the subunits of the F subcomplex of COPI show size similarities as well as N-terminal homology with the subunits of the APs [21, 53].

Our model of the smallest subunit of the F subcomplex, Ret3, consists of five beta sheets and seven alpha helices (Figs. 1b and S1), where the beta sheets and the first five alpha helices (N-terminal domain, residues 1–140) assemble into a core domain, and the last two alpha helices (C-terminal region, residues 141–189) assemble into the tail fragment (Figs. 1b and S1).

Upon minimizing the entire structure, the N-terminal domain (residues 1–140) of the subunit equilibrated after nearly 3250 ps of MD simulation, with an RMS deviation of <5 Å, suggesting a very stable structure in the core region (Fig. S2). This observation, and the fact that 15 of the 16 conserved amino acids [including six polar residues (Ser5, Lys25, Asn92, Asn123, Asp132, and Glu133) and all of the aliphatic amino acids] all accumulate in the N-terminal domain (Fig. S3), support the pivotal role of the N-terminal region in the stability and function of the subunit, as reported by Yu et al. [23].

In contrast, the C-terminal region (residues 141–189), with its folded helix structure, was found to equilibrate after 4,250 ps, with an RMSD of nearly 9 Å from the starting structure (Fig. S2). Monitoring the conformational changes of the Ret3 subunit during the MD simulation showed that the folding of amino acids 145–180 (including  $\alpha 6$  and  $\alpha 7$ ) changed; the  $\alpha 6$  helix unfolded at Pro160–Gly169 into a simple coil (Fig. 1c–d). The corresponding RMS deviation of the tail (Fig. S2) shows a climax at 4100–4400 ps, with an increase of nearly 2 Å due to the unfolding stage at  $\alpha 5$ , an isolated  $\beta$ -bridge (residues 128–144), and at Ile140–Asn146, which unfolds to a simple coil (Fig. 1b–d). The analysis of the MD simulation shows high root mean square (RMS) fluctuations of 4.0–10.0 Å in the atomic coordinates of the amino acids 145–158, 163–168, and 183–189 (Fig. S2), so the tail sweeps a large conformational space, causing destabilization of the structure in this region, as seen at 1000 ps and 4200 ps (Fig. 1c–d).

The C-terminal fragment contains several semiconserved residues (Fig. S3). The electrostatic potential surface on the residues 107–171 located between  $\alpha 3$  and  $\alpha 6$  and in the tail is mainly positive, causing electrostatic repulsion between

$\alpha 6$  in the C-terminal region and the  $\beta$ -bridge,  $\beta 2$ , in the  $\alpha 3$  region. This force drives the C-terminal region away, preventing it from joining to the well-packed core domain (residues 1–140) of the protein. The conformational variability of the C-terminal region, as well as its highly solvent-exposed folding, result in unsteadiness and eventually the detachment of the tail fragment.

In addition, *Saccharomyces cerevisiae* (P53600, Fig. S1) has the highest number of hydrophobic amino acids in  $\alpha 6$  and  $\alpha 7$  [with 20 hydrophobic residues out of the total of 40 in the C-terminal tail fragment (Ala149–Leu189)] among its homologous proteins: adapter-related protein complex 1 $\sigma$  (P61967, six of nine in total), adapter-related protein complex 2 $\sigma$  subunit in mouse (P62743, zero out of two) and the coatomer subunit- $\zeta 1$  in human (P61923, 13 of a total of 29). The MD simulation of Ret3 shows unsteadiness in the  $\alpha 6$  and  $\alpha 7$  regions, indicating weak structural folding. Since, in proteins, hydrophobic residues characteristically improve the stability of buried side chains, the weak structural folding of the  $\alpha 6$  and  $\alpha 7$  regions may be weakened further in the homologous proteins by the presence of even fewer hydrophobic residues in the equivalent region, which similarly leads to the disconnection of the highly flexible C-terminal region. Thus, the structure of the tail is unsolved for human  $\zeta$ -COP [23]. It should be noted that  $\zeta$ -COP is the only subunit to exist independently, separate from the COPI heptameric complex, and that free  $\zeta$ -COP and the heptamer are both stable, with similar half-lives in the cell of about 30 h [4, 54]. Similar to  $\zeta$ -COP, the N-terminal of Ret3 may be the only structurally stable domain, as it is shown by MD (Fig. S2).

### Sec21 subunit

The Sec21 subunit with 935 amino acids has three major components: a trunk domain, an appendage domain, and a linker connecting the trunk and the appendage domains. The subunit was modeled as described in the “Materials and methods” section. Briefly, first the major parts of the trunk and the appendage, including residues 1–267, 275–583, and 600–874, were modeled. The linker, the missing amino acids of the trunk, as well as the complementary region of the appendage were then modeled and added to the Sec21 structure (Fig. 2a–c).

Sec21 contains 34  $\alpha$ -helices, 17  $\beta$ -sheets, and seven  $3_{10}$  helices (Fig. S4). The trunk domain is a  $\alpha$ -solenoid domain consisting of 30  $\alpha$ -helices and five  $3_{10}$  helices. The linker is mainly folded into a simple coil, but it possesses a short  $\alpha$ -helix at residues 585–592. Similar to the  $\gamma$ -appendage, the Sec21 appendage has a bilobal structure, including platform (C-terminal) and the  $\beta$ -sandwich (or immunoglobulin-like or N-terminal) subdomains (Fig. 2b–c).

Compared to the platform subdomain in the  $\gamma$ -appendage (1R4X.pdb) [25], Sec21 has additional residues: 875–935

(Figs. 2c and S4b). The structure of the  $\gamma$ -appendage (1R4X.pdb) [25] was found to be the best template to model this fragment. This is folded into three  $\beta$ -sheets and one  $\alpha$ -helix at residues 920–928 (Fig. 2c). Sequence alignment for residues 875–935 with the amino acids in the  $\gamma$ -appendage yields 18% sequence identity as well as 46% similarity between the sequences of Sec21 and the  $\gamma$ -appendage in COPI (Fig. S4).

Sequence alignment was carried out on *Saccharomyces cerevisiae* Sec21 (P32074), coatomer subunit gamma of *Homo sapiens* (Q9Y678), AP1 complex subunit gamma-1 in mouse (P22892), and the alpha subunit of the AP2 complex in rat (Q66HM2), using the highly accurate alignment algorithm of the Clustal program [51, 55]. We tested the alignment using T-Coffee [56] and found very high similarity in the multi-sequence alignment results. In particular, for the sections relating to the conserved and semiconserved residues, the results from T-coffee and Clustal were found to be identical. We also searched for other possible protein families using Pfam [57] and no additional template structure was found. Pfam also suggested the same structure of adaptin and  $\gamma$ -COP for Sec21, and clathrin adaptor for the Ret3.

Clustal alignment shows 24 conserved amino acids, including Leu (30%), Ala (25%), Gly (16%), Asp (8%), as well as Tyr, Lys, Met, Arg, and Phe (each 4%), where the majority of the conserved residues are found on the trunk domain. Semiconserved amino acids are present in both the appendage and trunk domains, but are absent from the linker. Except for Asp119, Asp257, and Arg322, the conserved amino acids are all hydrophobic residues.

MD simulation of Sec21 demonstrates that stability is attained at about 4,300 ps for all three fragments of the subunit (Fig. S5). The linker and the trunk have more rigid structures (RMSD 7 Å) than the appendage (RMSD 8.2 Å) with respect to the average structure obtained from MD (Fig. S5). The RMS fluctuation is low,  $\sim 4$  Å, for amino acids 100–425 in the trunk, whereas the residues in the linker and the appendage fluctuate by up to 7.5 Å. The most unstable residues occur mainly in the linker (residues 575–610), in appendage  $\beta 8$  (residues 710–735), in  $\beta 13$  and  $\beta 14$  (residues 840–855), and in  $\beta 16$ ,  $\beta 17$ , and  $\alpha 34$  (residues 900–945). The  $3_{10}$  helix,  $\alpha 1$  (residues 1–20), the  $\alpha 1$ – $\alpha 2$  loop (residues 45–55),  $\alpha 26$ ,  $\alpha 27$ , and  $\alpha 28$  (residues 475–525), the coil between  $\alpha 30$ , and the isolated  $\beta$ -bridge are the residues of the trunk that also show considerable fluctuations, reaching as high as 5 Å (Fig. 2b, Figs. S4, and S5). The high quality of the trunk model was confirmed by PROCHECK [38], with an overall  $G$  factor of  $-0.26$ , 95.5% of the residues in the core, and 4.2% in the allowed regions.

### Three-dimensional structure of the Sec21–Ret3 complex

In the three-dimensional structure of the Ret3–Sec21 complex, the two subunits are linked through nonbonding

interaction forces, mainly between the core N-domain of Ret3 (1–149) and residues 1–601 of Sec21, forming four binding subsites (nodes) (Fig. 3a–b).

The C-terminal fragment of Ret3 (residues 150–189), including the  $\alpha 5$ – $\alpha 6$  loop,  $\alpha 6$ , and  $\alpha 7$ , are more than 29 Å from the interface of the subunits—too far to influence Ret3–Sec21 complex formation. The core domain of Ret3 binds to the  $\alpha$ -solenoid domain of Sec21, forming a four-node-like interaction site. The first interaction subsite (node 1) consists of networking between the aromatic amino acids Tyr26 and Tyr27 located at the  $3_{10}$  helix of Ret3 and Asp595, Ala598, and Thr599 from Sec21. At the second subsite (node 2) in Ret3, Leu54 at the  $\alpha 3$ – $\alpha 4$  loop, and Val73 and Leu74 at the  $\alpha 5$ – $\beta 6$  loop interact directly with Ser581, Leu582, Leu586, Tyr589, Ile590, and Ser596 in Sec21. The third subsite (node 3) is formed by Tyr78 at the  $\alpha 2$  and Glu122 at the  $\alpha 3$  from Ret3 and Lys562, Asp563, Ile566, Ala567, and Gln568 from Sec21. Met126 and Val127 at the  $\beta 7$ – $\alpha 4$  loop, and Leu128 and Leu129 at  $\alpha 4$  in Ret3 establish the fourth interaction subsite (node 4) at the interface of the two subunits with Ser289, Phe290, Arg293, and Arg296 in Sec21 (Fig. 3b). The complex obtained from docking was evaluated with a 20 ns long molecular dynamics simulation. The simulation reached equilibrium after approximately 10 ns of MD (Fig. S6).

Despite the individual conformational changes in the subunits during MD, a comparison of the different frames at various time steps demonstrates that the interactions between Ret3 and the  $\alpha$ -solenoid domain of Sec21 are stable. The amino acids of each subsite (node) remain involved in the interaction network, and are similar to those observed in the docking results. This is evident upon monitoring the variations in the distances between the centers of mass of the paired groups of amino acids from each subunit in each of the four nodes during 20 ns of simulation and comparing them to the corresponding distances in the docking result (Fig. 4a–d).

At around 8 ns, the distance between the centers of mass of the paired groups in node 1 decreases, as it also does in

node 2, remaining very close to the corresponding values for the reference structure in the docking result. The distances between the paired groups of amino acids in nodes 3 and 4 increase at around 10 ns, although they are very close to the corresponding docking results at about 12 ns and 17.5 ns. However, the distance between the centers of mass reduces in node 4 after 18 ns such that it is nearly the same as in the docking structure (Fig. 4a).

Since node 3 is located between node 4 and the highly stable node 2, the four-node-like binding site of Ret3 and Sec21 retains its complex pose during MD, and is similar to that predicted by docking. The structure of the complex and the interaction network at the binding site remain valid even after 10 ns and 16 ns of MD, when the largest distances between the centers of mass for the groups in node 3 and node 4 are observed (Fig. 4a–d).

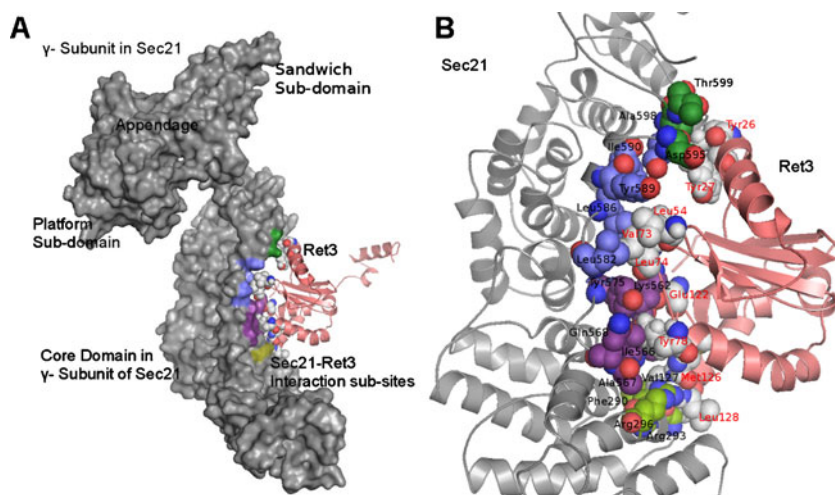
The information obtained from 20 ns long MD simulations shows the stability of the predicted binding mode of the subunits in a complex, and the reliability of the identification of the amino acids that are involved at the interface of the core and  $\alpha$ -solenoid domains in Ret3 and Sec21, respectively (Coordinate data in ESM1 Supplementary).

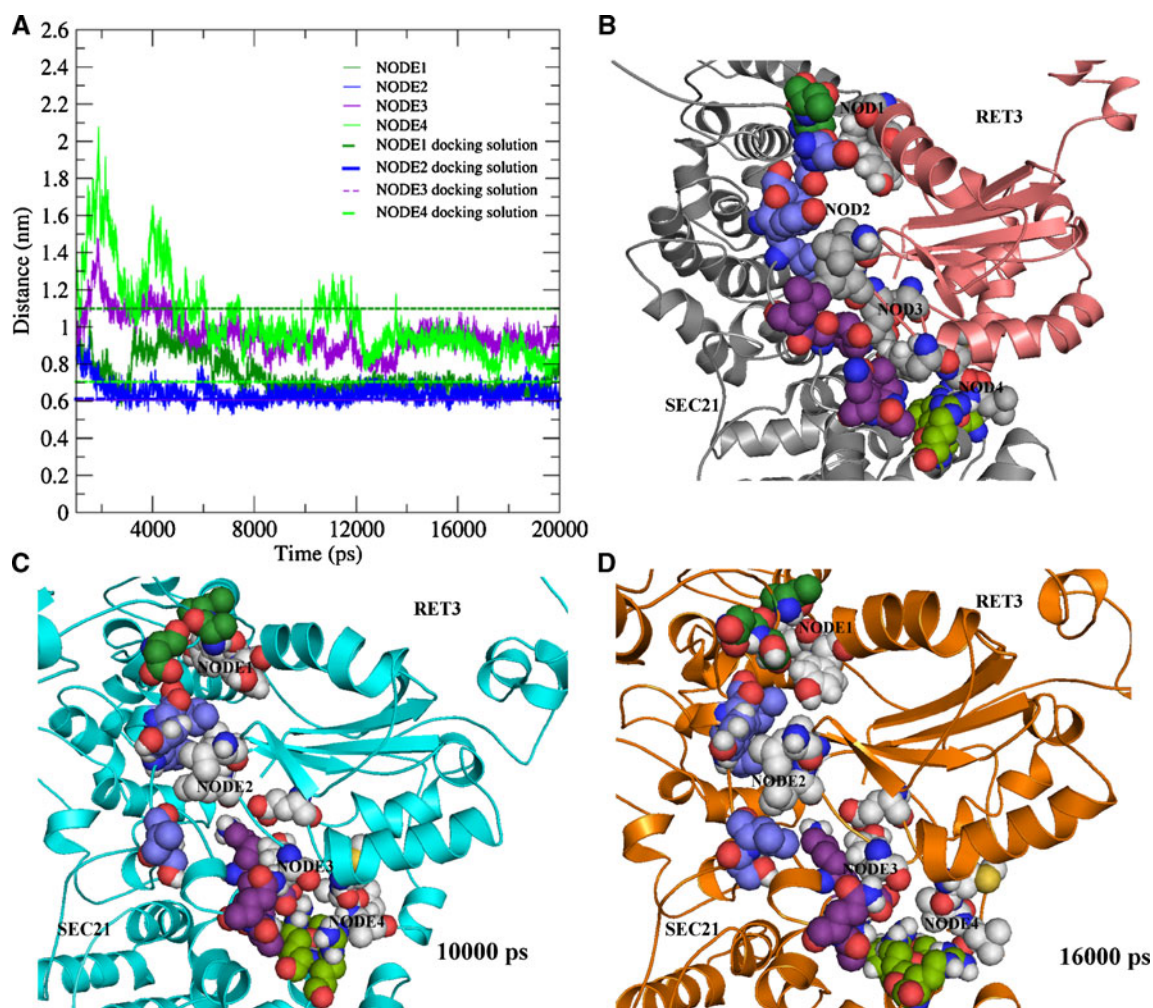
#### Potential binding sites in the appendage subdomain of Sec21 for accessory proteins

Based on sequence alignment between Sec21 and  $\alpha$ -AP2,  $\beta 2$ -AP2, and  $\gamma$ -COP, and using the peptide docking method, three regions in the Sec21 appendage were identified as being similar to the regions on  $\alpha$ -AP2,  $\beta 2$ -AP2 and  $\gamma$ -COP, and were shown to bind the recruitment peptides Eps15 DPF, epsin DPW2, epsin DPW1, and FXDXF [17, 27, 50].

The first binding site on the Sec21 subunit is similar to the very shallow pocket on the surfaces of the proteins  $\alpha$ -AP2,  $\beta 2$ -AP2, and  $\gamma$ -COP, which have been known to play a key role in protein–protein interactions with ARFGAP2 and ARFGAP3, the two mammalian isoforms of the yeast

**Fig. 3** The complex of Ret3 and Sec21 as part of the tetrameric F subcomplex obtained from docking. **a** The complex of Ret3 and Sec21 obtained from protein–protein docking; Ret3 is shown in pink and Sec21 in gray. **b** A close-up view of the interacting amino acids of the subunits (spheres); amino acids of Sec21 are shown in green, blue, violet, and olive, while those of Ret3 are shown in white)



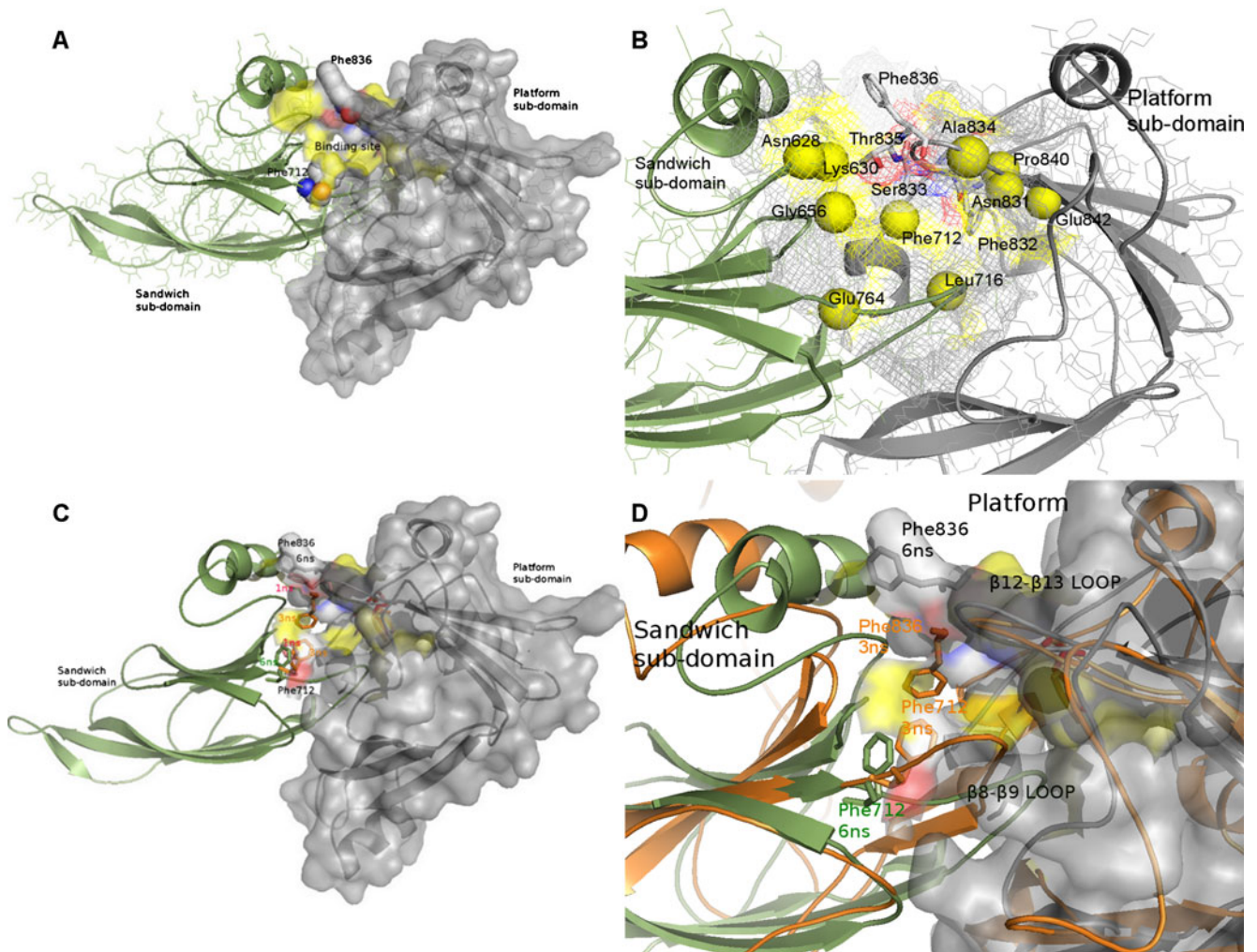


**Fig. 4** Molecular dynamics simulations for the Ret3–Sec21 complex. **a** Variations in the distances between the centers of mass of the paired groups of amino acids from each of the four interaction nodes (subsites) at the interface of the two subunits during 20 ns of MD simulation. The *dashed lines* represent the distances between centers of mass for each node in the docking structure (the reference), and the *fluctuating lines* show the distance variations during MD. The *dashed lines* relating

to node 2 (*blue*) and node 3 (*violet*) overlap with the fluctuating MD plot of node 2 (*blue*). **b** The amino acids of the two subunits at the binding site in the structure of the complex obtained from docking. The conformation at the interface of the subunits and the interacting amino acids from each node are shown at time steps of **c** 10 ns and **d** 16 ns. In both the plot and the figures, node 1 is *dark green*, node 2 is *blue*, node 3 is *violet*, and node 4 is *light green*

ARFGAP Glo3p [25]. In the  $\gamma$ -COP subunit, the residues Leu829, Leu841, Arg843, Arg859, Phe772, Glu773, Ala774, Ala775, and Trp776 create a shallow binding pocket, with the last five residues forming a hydrophobic patch that is essential for the interactions with ARFGAP2 and ARFGAP3 [25] (Fig. S7). In Sec21, Phe832 and Phe836 are the equivalent residues to Phe772 and Trp776 in  $\gamma$ -COP, which are also known as the F/W motif in the appendage domains of the three proteins  $\gamma$ -COP,  $\alpha$ -AP2, and  $\beta$ -AP2 (the F/W motif forms a shallow, solvent-exposed hydrophobic patch in the proteins homologous to  $\gamma$ -COP) [24]. In Sec21, an equivalent hydrophobic patch is formed by Phe832, Ser833, Ala834, Thr835 and Phe836 of the  $\beta$ 12– $\beta$ 13 loop. This patch is in a binding pocket  $\sim$ 14 Å

deep, which also contains Asn628, Ly630, Gly656, Phe712, Leu716, Gly656, Glu764, Asn831, Phe832, Ser833, Ala834, Thr835, Phe836, Pro840, and Glu842 (Fig. 5a–b). This binding site is suitable for binding ligands with long and flexible structures, and Phe712 from the  $\beta$ 8– $\beta$ 9 loop and Phe836 from the  $\beta$ 12– $\beta$ 13 loop are located at its entrance. Monitoring alterations to this region during 6 ns of MD simulations reveals that the aromatic rings of Phe712 and Phe836 move toward and away from each other. In the frame obtained at 3 ns of MD simulation, the Phe836 and Phe712 residues are at their closest positions to one another, while at 6 ns they are at the most distant positions (Fig. 5c–d). At 3 ns, they are close to one another at the entrance of the pocket, blocking access to the binding site. At 4 ns, they



**Fig. 5** Binding site on the platform subdomain of the Sec21 appendage. **a** The deep pocket at the interface of the platform subdomain (gray surface) and the sandwich (green ribbon) on Sec21. **b** Phe836 and Phe712 are on two opposite sides of the pocket entrance. **c** Superposition of the conformation of the appendage obtained at 3 ns

(orange ribbon) onto the one at 6 ns (green ribbon). **d** A close up view of the frames at 3 ns and 6 ns. Phe712 and Phe836 block the entrance to the binding pocket at 3 ns by displacing the  $\beta 8$ – $\beta 9$  loop of the sandwich subdomain and the  $\beta 12$ – $\beta 13$  loop of the platform

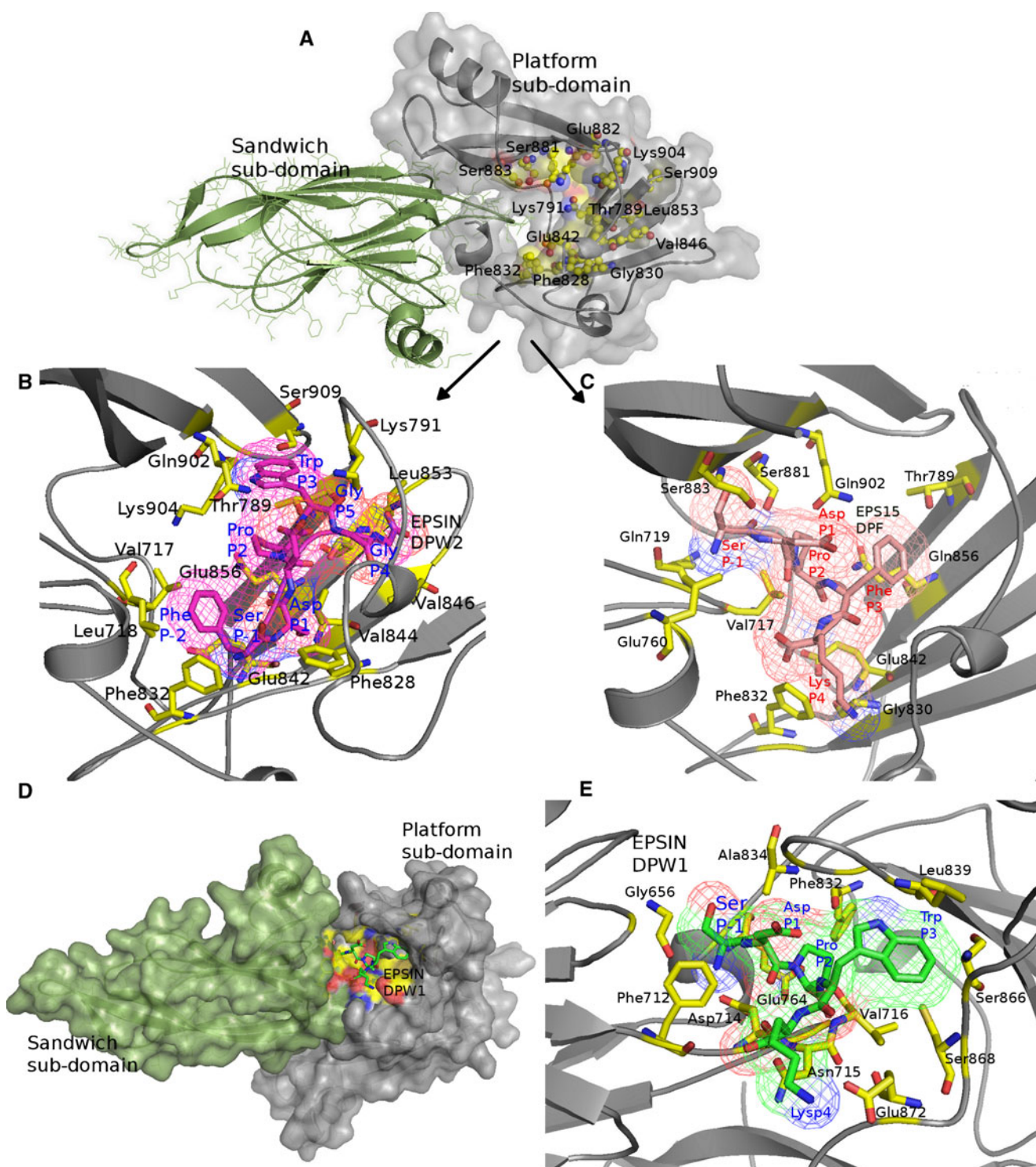
gradually move apart, allowing access to the binding site. The MD results show that the distance between Phe712 and Phe836 ranges from 12 Å to 4.3 Å. This indicates that they may play a major role in the availability of this pocket to incoming ligands by controlling the entrance to the pocket. Watson et al. [25] have shown that a serine mutation of Phe836 results in the elimination of the interaction between the appendage and Glo3p, though it does not affect the maturation of the soluble vacuolar hydrolase carboxypeptidase Y (CPY), a common marker for membrane transport in yeast, while deleting the whole appendage leads to a CPY defect, suggesting the presence of at least one other binding site on the appendage for this function [25].

The second binding site is a ~20 Å deep pocket on the platform subdomain of Sec21, with Phe832 and Ser883 at

the entrance to the pocket. Lys791, Val846, Leu853, and Ile855 are located on the antiparallel beta sheets deep at the bottom of this pocket, whereas Val844, Phe828 and Glu856 are in the middle of the pocket, where they form a negatively charged environment that is appropriate for holding positively charged ligands or those with aromatic moieties capable of  $\pi$ – $\pi$  stacking interactions (Fig. 6a–b).

A third binding site was found on the sandwich subdomain, which was equivalent to the binding site of DPW2 on the  $\beta$ -sandwich subdomain of AP2- $\alpha 2$ . It consists of Lys622, Gln623 on the  $\beta 17$ – $\alpha 34$  loop, Ala672–Ile678 on  $\beta 6$ , as well as Ala693–Val704 on the  $\beta 7$ – $\beta 8$  loop, and Asn733 on  $\beta 9$  (Fig. 7a–b).

We then studied the interactions between Sec21 and the peptides that have been previously shown to interact with the

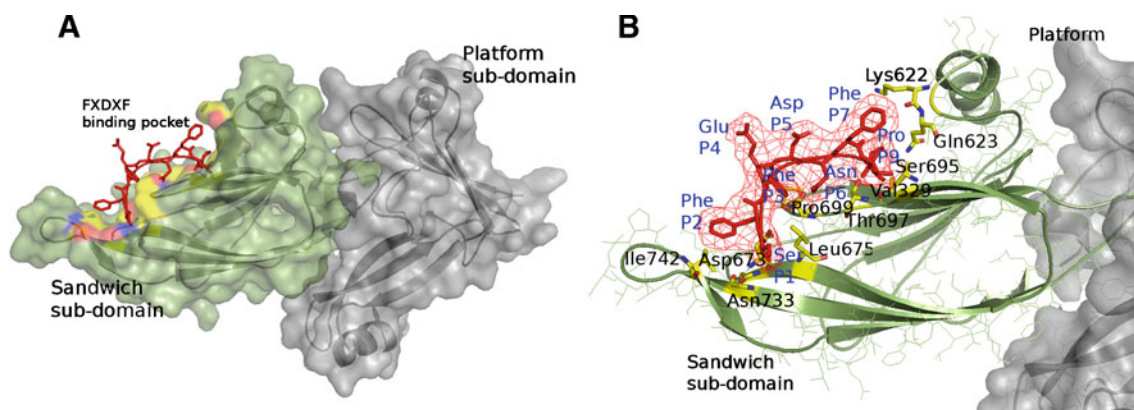


**Fig. 6** Peptide binding sites on the platform subdomain of the Sec21 appendage. **a** The ~20 Å deep binding pocket of the platform sub-domain and the residues for accommodating a ligand. **b** Residues of Sec21 that interact with DPW2 (*magenta sticks*). **c** Residues of the

binding site that interact with DPF (*pink sticks*). **d** DPW1 (*green sticks*) binds to a second binding site on the platform. **e** Residues of the ~14 Å deep binding pocket that interact with DPW1

platform and sandwich subdomains of the  $\alpha$ -AP2,  $\beta$ 2-AP2, and  $\gamma$ -COPI appendages. The peptides are found in accessory proteins such as epsin and Eps15. Epsin is involved in creating membrane curvature and Eps15 in a scaffolding protein.

Although the docking experiments for each ligand provide detailed profiles for the interactions of the ligands with their surrounding environment, we further validated them using the GHECOM grid-based HECOMi finder [52] as an additional



**Fig. 7** FXDXF binds to the sandwich subdomain of the Sec21 appendage. **a** The binding site of FXDXF (red sticks) on the surface of the sandwich subdomain. **b** Interaction profile of FXDXF with binding site residues (yellow sticks)

binding site detector approach. The first-rank cluster of GHECOM grids, representing the best match for the binding site, partially covers the docking site of FXDXF, while it completely overlaps with those of DPF, DPW1, and DPW2 (Fig. S8).

The high level of agreement between the first-rank clusters of pocket grids suggested by GHECOM [52] for each peptide and those obtained from docking further supports the predictions regarding the regions that accommodate the ligands.

#### Binding of Eps15 DPF and epsin DPW2 to Sec21

Two individual docking experiments confirm the binding of two peptide ligands, Eps15 DPF (1KYU.pdb) [27] and epsin DPW2 (1KY6.pdb) [27], in a ~20 Å deep pocket. Both interaction networks engage the residues Val717, Phe832, Glu842, and Gln902 (Fig. 6a–c). This binding site has not been described previously, as it is shaped by the association of the newly modeled fragment of appendage residues 875–935 (Fig. 2c), which is missing from the structure of the  $\gamma$ -appendage [25] (Figs. 2a and S4b).

The conformations of the peptides in their binding pockets result from the direct involvement of Val717, Phe832, Glu842, and Gln902 with Pro<sup>P2</sup>, Lys<sup>P4</sup>, Phe<sup>P3</sup>, and Asp<sup>P1</sup> of DPF (Fig. 6c), and with the Phe<sup>P-2</sup>, Ser<sup>P-1</sup>, and Trp<sup>P3</sup> motifs in DPW2 (Fig. 6b). Trp<sup>P3</sup> in DPW2 is surrounded by Gln902 and Lys791. Asp<sup>P1</sup> in DPF interacts directly with Thr789 and Gln902 at a similar location close to the pocket entrance, while Asp<sup>P1</sup> in DPW2 is involved with Phe828, Glu842, and Val844 near to the bottom of the site. Phe<sup>P-2</sup> of DPW2 is accommodated by Leu718, Phe832, and Glu842. However, Lys<sup>P4</sup> and Phe<sup>P3</sup> from DPF occupy similar positions at the bottom of the binding pocket, surrounded by Phe832, Glu842, and Gln856 (Fig. 6b–c).

#### Binding of epsin DPW1 to Sec21

It has been shown that  $\alpha$ -AP2 accommodates DPW1 [27]. In contrast, no data has been published to show that Sec21

binds to DPW1 in a similar fashion to AP2- $\alpha$ 2. The in silico experiment revealed that DPW1 can bind to Sec21 (Fig. 6d–e). It is accommodated among the beta sheets at the second identified binding site with a depth of nearly 14 Å on the platform subdomain. The Ser<sup>P-1</sup> moiety of the peptide interacts with the  $\beta$ 4– $\beta$ 5,  $\beta$ 8– $\beta$ 9, and  $\beta$ 12– $\beta$ 13 loops of Sec21, involving Ala834, Gly656, and Phe712. Asp<sup>P1</sup> and Pro<sup>P2</sup> are located close to the  $\alpha$ 10 helix and the  $\beta$ 12– $\beta$ 13 loop, interacting with Glu764 and Phe832, respectively. Trp<sup>P3</sup> is surrounded by Phe832, Leu839, and Ser866 at the  $\beta$ 12– $\beta$ 13 and  $\beta$ 14– $\beta$ 15 loops, as Lys<sup>P4</sup> interacts electrostatically with the two negatively charged residues Asn715 and Glu872. DPW1 is the only peptide among the four peptides we investigated (DPW1, DPW2, DPF, and FXDXF) that we found to bind in this pocket. Sequence alignment of Sec21,  $\gamma$ -COP, and AP2- $\alpha$ 2 shows that amino acids at the binding site of DPW1 on Sec21 (Gly656, Phe712, Asn715, Glu764, Phe832, Leu838, Ser 866, and Glu872) are 50% identical to their equivalent amino acids in  $\gamma$ -COP and in AP2- $\alpha$ 2, so Sec21 is expected to treat the ligand similarly to either of these proteins (Fig. S9).

#### Binding of FXDXF to Sec21

FXDXF binds at a similar region to that of DPW2 on the  $\beta$ -sandwich subdomain in AP2- $\alpha$ 2, with the direct involvement of the conserved amino acids Leu676, Ala672, and Asn733. FXDXF interacts with the amino acids Lys622–Gln623 on the  $\beta$ 17– $\alpha$ 34 loop via the Phe<sup>P7</sup> moiety, with Ala672–Ile678 at  $\beta$ 6, as well as with Ala693–Val704 on the  $\beta$ 7– $\beta$ 8 loop via Val<sup>P8</sup>, and Ser<sup>P1</sup> interacts with Asn733 on  $\beta$ 9 (Fig. 7a–b).

As in  $\alpha$ 2-AP2, wherein DPW2 is the only peptide that binds at the  $\beta$ -sandwich surface [27], docking results show that FXDXF is the only peptide among DPW1, DPW2, and DPF that binds at the  $\beta$ -sandwich subdomain: with Lys622–Gln623 on the  $\beta$ 17– $\alpha$ 34 loop via the Phe<sup>P7</sup> moiety, with Ala672–Ile678 at  $\beta$ 6, as well as with Ala693–Val704 on the

$\beta 7$ – $\beta 8$  loop via Val<sup>P8</sup>, and Ser<sup>P1</sup> interacts with Asn733 on  $\beta 9$  of the appendage (Fig. 7a–b). This peptide shares no residues with other ligands on the platform subdomain.

## Conclusions

Models of the entire Sec21 coatomer subunit (P32074) and of the entire Ret3 coatomer subunit (P53600) from COPI non-clathrin-coated vesicle have been built using the crystallographic structures of the trunk domain of clathrin adaptor protein 1 (AP1), the human  $\gamma$ -appendage domain, and the structure of the truncated human  $\zeta$ -COPI, respectively. Analysis of the molecular dynamics simulations for Ret3 reveals that the subunit consists of a stable bulky domain connected to a flexible C-terminal domain that consists of two highly flexible helices,  $\alpha 6$  and  $\alpha 7$ . The interaction between the C-terminal domain and the stable bulky domain is relatively weak, due to the positive electrostatic region between the C-terminal domain and the neighboring regions of the bulky domain ( $\beta$ -bridge,  $\beta 2$ , and  $\alpha 3$ ). This electrostatic repulsion keeps the C-terminal helices away from the bulky domain, preventing it from folding with the stable bulky “core” domain. The highly solvent-exposed C-terminal domain explores a large conformational space during MD, with high RMS fluctuations of its residues. This observation explains why this region is sensitive to proteolysis. Indeed, the full-length  $\zeta$ -subunit (20 kDa) cleaves into a smaller 17 kDa fragment, resulting in the truncated structure of the human  $\zeta$ -subunit that is yielded by X-ray crystallography [23]. Analysis of the MD simulation of the entire Sec21 subunit, which has three domains (a trunk, a linker, and an appendage), reveals that the appendage undergoes large conformational variations, particularly within the first 3.5 ns of MD, though it achieves stability after 4.5 ns of simulation. The protein–protein docking solution for Ret3 and Sec21 shows a four-node-like binding site at the interface of the trunk and core region of Ret3. The interaction profile for the two subunits does not exactly match with the equivalent sequences of the involved amino acids, as reported by Yu et al. [23], where the subcomplex was simulated by superposing the N-terminal domain of each subunit onto adaptin core protein 1.

In our study, the protein–protein docking technique was employed to simulate the complex—taking into account the structures determined from the full sequences of the subunits—in order to predict the most energetically favorable conformation and the amino acids involved at the site of interaction between the Ret3 core and Sec21  $\alpha$ -solenoid domains. The 20 ns long MD simulation of the structure of the Ret3–Sec21 complex obtained from docking demonstrated that the interaction between Ret3 and Sec21 is stable and that the amino acids at the interaction sites predicted by

docking were remained valid during MD, despite conformational fluctuations, thus indicating that the predicted binding mode by docking is reliable. Compared to the previous model [23], it is evident that the missing domains and amino acids have an impact on the overall interaction profile of the subunits, which explains some of the differences between the two models.

The simulated structure of the appendage contains potential binding sites for accommodating the ligands on both the  $\beta$ -sandwich and the platform subdomains. These were studied using docking in addition to a grid-based binding site predictor method. A  $\sim 14$  Å deep pocket on the platform subdomain was identified with the potential to selectively accommodate the epsin DPW1 peptide. Sequence alignment of Sec21,  $\gamma$ -COP, and AP2- $\alpha 2$  showed that amino acids at the binding site of DPW1 on Sec21 are 50% identical to the amino acids in the equivalent regions of  $\gamma$ -COP and AP2- $\alpha 2$ . Thus, there is an almost equal probability of Sec21 treating DPW1 in a similar manner to the either of the proteins. However, the observed binding mode of DPW1 at its binding site on Sec21 shows that it most likely acts in a similar fashion to AP2- $\alpha 2$ .

Another binding site is a narrow  $\sim 20$  Å deep pocket formed by an association of the newly modeled fragment of the appendage (residues 875–935) on the platform subdomain of Sec21. This fragment is missing from the crystal structure of the  $\gamma$ -appendage [25]. The conformations and orientations of Phe712 and Phe836 at the rim of the deep pocket, along with those of Asn715, appear to control the entry of ligands into the pocket. These amino acids cover the entrance to the binding site within the first 3.5 ns of MD, but move away after 4 ns, opening the entrance to the pocket for the incoming ligand. The docking results for Eps15 DPF and epsin DPW2 at this site revealed detailed binding profiles of the peptides when complexed with the appendage.

The binding pocket on the surface of the  $\beta$ -sandwich subdomain in the Sec21 appendage was shown to be in close proximity to the DPW2 binding site on  $\alpha 2$ -AP2. Docking experiments on DPW1, DPW2, epsin DPF, and FXDXF demonstrate that the latter binds to the  $\beta$ -sandwich subdomain of Sec21 selectively, like DPW2, which binds to  $\alpha 2$ -AP2.

Modeling the three-dimensional structures of the full-length sequences of the subunits and then performing MD simulations has allowed us to find the structural properties of the domains involved, as well as their impact on the stabilization of the tertiary structures of the two-subunit complex. Analysis of the docked peptide binding modes in the appendage subdomains has pointed to the existence of potential sites on the appendage that may facilitate the coatomer’s function as a coordinator in the recruitment of accessory proteins.

**Acknowledgments** The authors thank Drs. John Bergeron, John Presley, Tommy Nilsson, and Oliver Stueker for their critical comments on the manuscript. I.R. is grateful to the Natural Sciences and Engineering Research Council of Canada (NSERC 355873–08) and the Canada Foundation for Innovation (CFI) for financial assistance. I. R. is the recipient of a Canadian Institutes of Health Research (CIHR) New Investigator award.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Langer JD, Stoops EH, Béthune J, Wieland FT (2007) Conformational changes of coat proteins during vesicle formation. *FEBS Lett* 581:2083–2088. doi:10.1016/j.febslet.2007.03.020
- Fotin A, Cheng Y, Sliz P, Grigorieff N, Harrison SC, Kirchhausen T, Walz T (2004) Molecular model for a complete clathrin lattice from electron cryomicroscopy. *Nature* 432:573–579. doi:10.1038/nature03079
- Rothman JE (2002) The machinery and principles of vesicle transport in the cell. Lasker Basic Medical Research Award *Nat Med* 8:1059–1062. doi:10.1038/nm770
- Lowe M, Kreis TE (1995) In vitro assembly and disassembly of coat-omer. *J Biol Chem* 270:31364–31371. doi:10.1074/jbc.270.52.31364
- Robinson MS (2004) Adaptable adaptors for coated vesicles. *Trends Cell Biol* 14. doi:10.1016/j.tcb.2004.02.002
- Collins BM, McCoy AJ, Kent HM, Evans PR, Owen DJ (2002) Molecular architecture and functional model of the endocytic AP2 complex. *Cell* 109:523–535. doi:10.1016/S0092-8674(02)00735-3
- Heuser JE, Keen J (1988) Deep-etch visualization of proteins involved in clathrin assembly. *J Cell Biol* 107:877–886. doi:10.1083/jcb.107.3.877
- Jackson LP, Kelly BT, McCoy AJ, Gaffry T, James LC, Collins BM, Höning S, Evans PR, Owen DJ (2010) A large-scale conformational change couples membrane recruitment to cargo binding in the AP2 clathrin adaptor complex. *Cell* 141:1220–1229. doi:10.1016/j.cell.2010.05.006
- Owen DJ, Vallis Y, Pearce BMF, McMahon HT, Evans PR (2000) The structure and function of the [beta]2-adaptin appendage domain. *EMBO J* 19:4216–4227. doi:10.1093/emboj/19.16.4216
- Dell'Angelica EC, Klumperman J, Stoorvogel W, Bonifacino JS (1998) Association of the AP-3 adaptor complex with clathrin. *Science* 280:431–434. doi:10.1126/science.280.5362.431
- ter Haar E, Harrison SC, Kirchhausen T (2000) Peptide-in-groove interactions link target proteins to the beta-propeller of clathrin. *Proc Natl Acad Sci USA* 97:1096–1100
- Slepnev VI, Ochoa G-C, Butler MH, De Camilli P (2000) Tandem arrangement of the clathrin and AP-2 binding domains in amphiphysin 1 and disruption of clathrin coat function by amphiphysin fragments comprising these sites. *J Biol Chem* 275:17583–17589. doi:10.1074/jbc.M910430199
- Traub LM (2009) Tickets to ride: selecting cargo for clathrin-regulated internalization. *Nat Rev Mol Cell Biol* 10:583–596. doi:10.1038/nrm2751
- Owen DJ, Vallis Y, Noble MEM, Hunter JB, Dafforn TR, Evans PR, McMahon HT (1999) A structural explanation for the binding of multiple ligands by the alpha-adaptin appendage domain. *Cell* 97:805–815. doi:10.1016/S0092-8674(00)80791-6
- Traub LM, Downs MA, Westrich JL, Fremont DH (1999) Crystal structure of the alpha appendage of AP-2 reveals a recruitment platform for clathrin-coat assembly. *Proc Natl Acad Sci USA* 96:8907–8912
- Edeling MA, Smith C, Owen D (2006) Life of a clathrin coat: insights from clathrin and AP structures. *Nat Rev Mol Cell Biol* 7:32–44. doi:10.1038/nrm1786
- Traub LM, Wendland B (2010) Cell biology: how to don a coat. *Nature* 465:556–557. doi:10.1038/465556a
- Stagg SM, Gurkan 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. doi:10.1038/nature04339
- Stagg SM, LaPointe P, Razvi A, Gürkan C, Potter CS, Carragher B, Balch WE (2008) Structural basis for cargo regulation of COPII coat assembly. *Cell* 134:474–484. doi:10.1016/j.cell.2008.06.024
- Lee C, Goldberg J (2010) Structure of coatmer cage proteins and the relationship among COPI, COPII, and clathrin vesicle coats. *Cell* 142:123–132. doi:10.1016/j.cell.2010.05.030
- Schledzewski K, Brinkmann H, Mendel RR (1999) Phylogenetic analysis of components of the eukaryotic vesicle transport system reveals a common origin of adaptor protein complexes 1, 2, and 3 and the F subcomplex of the coatmer COPI. *J Mol Evolution* 48:770–778. doi:10.1007/PL00006521
- Takatsu H, Futatsumori M, Yoshino K, Yoshida Y, Shin H-W, Nakayama K (2001) Similar subunit interactions contribute to assembly of clathrin adaptor complexes and COPI complex: analysis using yeast three-hybrid system. *Biochem Biophys Res Commun* 284:1083–1089. doi:10.1006/bbrc.2001.5081
- Yu W, Lin J, Jin C, Xia B (2009) Solution structure of human [zeta]-COP: direct evidences for structural similarity between COP I and clathrin-adaptor coats. *J Mol Biol* 386:903–912. doi:10.1016/j.jmb.2008.12.083
- Hoffman GR, Rahl PB, Collins RN, Cerione RA (2003) Conserved structural motifs in intracellular trafficking pathways: structure of the [gamma]COP appendage domain. *Mol Cell* 12:615–625. doi:10.1016/j.molcel.2003.08.002
- Watson PJ, Frigerio G, Collins BM, Duden R, Owen DJ (2004) gamma-COP appendage domain—structure and function. *Traffic* 5:79–88. doi:10.1111/j.1600-0854.2004.00158.x
- Eugster A, Frigerio G, Dale M, Duden R (2000) COP I domains required for coatmer integrity, and novel interactions with ARF and ARF-GAP. *EMBO J* 19:3905–3917. doi:10.1093/emboj/19.15.3905
- Brett TJ, Traub LM, Fremont DH (2002) Accessory protein recruitment motifs in clathrin-mediated endocytosis. *Structure* 10:797–809. doi:10.1016/S0969-2126(02)00784-0
- Heldwein EE, Macia E, Wang J, Yin HL, Kirchhausen T, Harrison SC (2004) Crystal structure of the clathrin adaptor protein 1 core. *Proc Natl Acad Sci USA* 101:14108–14113. doi:10.1073/pnas.0406102101
- Bairoch A, Apweiler R (1999) The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. *Nucl Acids Res* 27:49–54. doi:10.1093/nar/27.1.49
- Kelly BT, McCoy AJ, Spate K, Miller SE, Evans PR, Honing S, Owen DJ (2008) A structural explanation for the binding of endocytic dileucine motifs by the AP2 complex. *Nature* 456:976–979. doi:10.1038/nature07422
- Shi J, Blundell TL, Mizuguchi K (2001) FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J Mol Biol* 310:243–257. doi:10.1006/jmbi.2001.4762
- Jones DT (1999) GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *J Mol Biol* 287:797–815. doi:10.1006/jmbi.1999.2583
- Kelley LA, MacCallum RM, Sternberg MJE (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J Mol Biol* 299:501–522. doi:10.1006/jmbi.2000.3741
- McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. *Bioinformatics* 16:404–405. doi:10.1093/bioinformatics/16.4.404

35. Kelley LA, Sternberg MJE (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protocols* 4:363–371
36. Lundström J, Rychlewski L, Bujnicki J, Elofsson A (2001) Pcons: a neural-network-based consensus predictor that improves fold recognition. *Protein Sci* 10:2354–2362. doi:10.1101/ps.08501
37. Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T (2008) Protein structure homology modeling using SWISS-MODEL workspace. *Nat Protocols* 4:1–13. doi:10.1038/nprot.2008.197
38. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallograph* 26:283–291. doi:10.1107/S002188982009944
39. Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci* 2:1511–1519. doi:10.1002/pro.5560020916
40. Heinig M, Frishman D (2004) STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucl Acids Res* 32(2):W500–502. doi:10.1093/nar/gkh429
41. Berendsen HC, Postma JP, van Gunsteren WF, Hermans J (1981) Interaction models for water in relation to protein hydration. In: Pullman B (ed) *Intermolecular Forces*. D Reidel, Dordrecht
42. Berendsen HJC, van der Spoel D, van Drunen R (1995) GROMACS: a message-passing parallel molecular dynamics implementation. *Comput Phys Commun* 91:43–56. doi:10.1016/0010-4655(95)00042-E
43. Daren T, York D, Pedersen L (1993) Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. *J Phys Chem* 98:10089. doi:10.1063/1.464397
44. Hess B, van der Vegt NFA (2006) Hydration thermodynamic properties of amino acid analogues: a systematic comparison of biomolecular force fields and water models. *J Phys Chem B* 110:17616–17626. doi:10.1021/jp0641029
45. Essmann U, Perera L, Berkowitz M, Darden T, Lee H, Pedersen L (1995) A smooth particle mesh Ewald method. *J Phys Chem* 103:17. doi:10.1063/1
46. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: a linear constraint solver for molecular simulations. *J Comput Chem* 18:1463–1472. doi:10.1002/(SICI)1096-987X(199709
47. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* 81:3684–3690. doi:10.1063/1.448118
48. Ritchie DW, Kozakov D, Vajda S (2008) Accelerating and focusing protein–protein docking correlations using multi-dimensional rotational FFT generating functions. *Bioinformatics* 24:1865–1873. doi:10.1093/bioinformatics/btn334
49. Chen R, Li L, Weng Z (2003) ZDOCK: an initial-stage protein-docking algorithm. *Prot Struct Funct Genetics* 52:80–87. doi:10.1002/prot.10389
50. Kent HM, McMahon HT, Evans PR, Benmerah A, Owen DJ (2002) [gamma]-adapin appendage domain: structure and binding site for Eps15 and [gamma]-synergins. *Structure* 10:1139–1148
51. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003) Multiple sequence alignment with the Clustal series of programs. *Nucl Acids Res* 31:3497–3500. doi:10.1093/nar/gkg500
52. Kawabata T (2010) Detection of multiscale pockets on protein surfaces using mathematical morphology. *Prot Struct Funct Bioinf* 78:1195–1211. doi:10.1002/prot.22639
53. Lewin DA, Mellman I (1998) Sorting out adaptors. *Biochim Biophys Acta* 1401:129–145. doi:10.1016/S0167-4889(97)00130-4
54. Kuge O, Hara-Kuge S, Orci L, Ravazzola M, Amherdt M, Tanigawa G, Wieland FT, Rothman JE (1993) zeta-COP, a subunit of coatamer, is required for COP-coated vesicle assembly. *J Cell Biol* 123(6 Pt 2):1727–1734
55. Raghava G, Searle S, Audley P, Barber J, Barton G (2003) OXbench: a benchmark for evaluation of protein multiple sequence alignment accuracy. *BMC Bioinforma* 4:47. doi:10.1186/1471-2105-4-47
56. Notredame C, Higgins DG, Heringa J (2000) T-coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 302:205–217
57. Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer ELL, Eddy SR, Bateman A (2010) The Pfam protein families database. *Nucl Acids Res* 38(suppl 1):D211–D222. doi:10.1093/nar/gkp985