Mechanism for ESCPE-1-mediated membrane remodelling and cargo selection in endosomal sorting

Carlos López Robles

2020

Table of Contents

Abbreviations

AD	Alzheimer´s disease
AP	Adaptor protein
APP	Amyloid precursor protein
ATP	Adenosine triphosphate
BAR	Bin/Amphiphysin/Rvs domain
CIMPR	Cation-independent mannose-6-phosphate receptor
CORVET	Class C core vacuole/endosome tethering complex
CTF	Contrast transfer function
EDTA	Ethylenediaminetetraacetic acid
EE	Early endosomes
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
ER	Endoplasmic reticulum
ESCRT	Endosomal sorting complex
CIE	Clathrin-independent endocytosis
DMT1	Divalent metal transporter 1
DNA	Deoxyribonucleic acid
DOPC	1,2-dioleoyl-sn-glycero-3-phosphocholine
DOPE	1,2-dioleoyl-sn-glycero-3-phosphoethanolamine
DOPS	1,2-dioleoyl-sn-glycero-3-phosphoserine
GA	Golgi apparatus
GAP	GTPase-activating proteins
GEF	GTP exchange factors
GPI-AP	Glycosylphosphatidylinositol-anchored protein
GDP	Guanosine diphosphate
GTP	Guanosine triphosphate
HEPES	4-(2-hydroxyethyl)-1 piperazineethanesulfonic acid
ILV	Intra-luminal bodies
IPTG	Isopropyl-B-D-thiogalactopyranoside
MALS	Multi-angle light scattering
MCS	Membrane contact sites

DTT	Threo/1,4/Dimercapto/2,3/butanediol, DL-ithiothreitol
HOPS	Homotypic fusion and vacuole protein sorting
LB	Luria Bertani
LE	Late endosomes
LUV	Large unilamellar vesicles
GEF	Guanine exchange factor
GST	Glutathione-S-transferase
MAD	Multiple anomalous dispersion
MVB	Multi-vesicular bodies
NMR	Nuclear magnetic resonance
PCR	Polymerase chain reaction
PD	Parkinson´s disease
PDB	Protein data bank
PH	Pleckstrin homology
PI3K	Phosphatidylinositol 3-kinase
$PI(4,5)P_2$	phosphatidylinositol 4,5-bisphosphate
PI(3)P	Phosphatidylinositol 3-phosphate
PM	Plasma membrane
PMSF	Phenylmethylsulfonyl fluoride
PX	Phox homology
OD	Optical density
SNARE	N-ethylmaleimide-sensitive factor attachment protein receptor
TEN	Tubular endosomal networks
TEV	Tobacco each virus
TGN	*Trans* Golgi network
Tris/HCl	Tris(hydroxymethyl)aminomethane hydrochloride
RE	Recycling endosomes
RT	Room temperature
RILP	Rab-interacting lysosomal protein
SAD	Single anomalous dispersion
SEC	Size-exclusion chromatography

SDS	Sodium dodecyl sulphate
SENP2	Sentrin specific protease 2
SNX	Sorting Nexin
TCEP	Tris(2-carboxyethyl)phosphine hydrochloride
TRAPP	Transport protein particle
VPS	Vacuolar protein sorting
WASH	Wiskott-Aldrich syndrome protein and scar homolog
WASP	Wiskott-Aldrich syndrome protein

1. INTRODUCTION

1. INTRODUCTION

1.1. Membrane trafficking system

The defining feature of eukaryotic cells is their compartmentalisation into specialised membrane-separated structures called organelles. Many cellular processes, from DNA storage to energy production, take place inside these structures. This complexity could have arisen from endosymbiotic events such as the formation of mitochondria and chloroplasts[1] but also by gene duplication, producing similar proteins (paralogues) with different subcellular locations[2]. Compartmentalisation allows a high degree of specialisation, but it limits the communication within the cell and (in multicellular organisms) with other cells. Hence, the cell needs a transport system regulating the traffic of molecules[3]. Some organelles are interconnected directly or through membrane contact sites (MCSs)[4]. The MCSs are the areas where the membranes of two organelles are in close proximity, typically within less than 30 nm, with micro-domains that favour the exchange of molecules and ions[5]. The endoplasmic reticulum (ER) is the most extensive membrane network in the cell; it is the organelle that contains most MCSs. It interacts with endosomes, lysosomes, mitochondria, peroxisomes and the plasma membrane (PM)[6]. For instance, the MCSs between the ER and PM allow the recovery of intra-ER calcium levels after calcium depletion (during cell signalling response). The calcium channels in the PM open and the Ca^{2+} level in the ER is restored without affecting its level in the cytoplasm[7]. However, molecule transport to distal organelles or other cells cannot be achieved using the MCSs; it is mediated by vesicle trafficking. The endosome is the principal repository of vesicles. It receives various molecules from the PM and the Golgi apparatus (GA) and passes them to other organelles, to the PM, or sends them to lysosomes for degradation.

There are many routes of vesicle trafficking in the cell (Fig. 1). They can be grouped into two main categories: the first, the inward flux of vesicles from the exterior of the cell to the cytosol (endocytosis/endocytic pathway) and the second, the outward flux to the extracellular media (exocytosis/secretory pathway)[8].

The secretory (or biosynthetic) pathway (Fig. 1, blue lines) is responsible for the transport of lipids and proteins from the ER to other organelles or to the extracellular space, using the vesicles and MCSs. Lipid transport can also take place within the

membrane. Intra-membrane movement has been explained by liquid phase separation[9] and molecular dynamics studies[10] showing how lipid rafts and micro-domains with particular properties are created. Lipids can also be transported between organelles or even cells. In the latter case, the vesicle transfer is involved, although the transfer of specific lipids between membranes is accomplished mainly by lipid transfer proteins (LTP). They can engulf the aliphatic chain in hydrophobic cavities to transfer a particular lipid, or bridge two organelles to transfer lipids through the MCSs[11]. For proteins, there is bulk transportation not dependent on signal sequences or specific organelles[12]. However, the most common procedure of transporting proteins involves a signal peptide that directs the molecule to the correct destination[13]. Soluble proteins are directly secreted to the cytosol by Sec translocases[14], but the transmembrane proteins (cargos) are sent to their correct destinations in the vesicles. The cargos are packed in the vesicles and transported to the ER-Golgi intermediate compartment (between the ER and GA), which functions as a sorting station[15]. They travel through the different GA sub-compartments (*cis, medial* and *trans),* where they may be further modified (e.g. glycosylated), up to the *trans* Golgi network (TGN), to reach their destination[16,17]. Besides its function in cargo distribution, the TGN receive recycled membrane and cargo from the endosomes; it prevents these specific molecules from entering the lysosomal degradative pathway, which would be metabolically expensive for the cell[17].

The endocytosis process (Fig. 1, grey lines), coupled to vesicle trafficking, is the movement of molecules from the plasma membrane into the cell. This route is usually used for nutrient uptake; however, some pathogens take advantage of this process to infect the cell[18]. Endocytosis will be explained in more detail in Chapter 1.1.5.

In any vesicle-trafficking event, there are four distinct events. The first one is vesicle budding, the formation of the vesicle in the original compartment. Many factors participate in this process, controlling the cargo selection and membrane deformation. The next step is the vesicle displacement to the target organelle, using motor proteins and other elements of the cytoskeleton. Then, the vesicle is recognised and tethered to the target. Finally, the coat proteins, tethering complexes, and N-ethylmaleimide-sensitive factor attachment protein receptors (SNAREs) facilitate the eventual membrane fusion. Many of these processes are thermodynamically unfavourable and are tightly regulated.

1.1.1. Vesicle budding

Vesicle budding is a tightly controlled process ensuring the correct cargo loading. Many different effectors participate in the different budding processes (Fig. 1); those effectors are recruited mostly by small GTPases from the Rab family[19] and phosphoinositides[20]. Many of the effectors involved in vesicle budding are the coat proteins, which will cover the vesicle. The main functions of coat proteins are cargo selection and budding of the vesicle. Their importance has been demonstrated by several groups; silencing those proteins impairs endocytosis and cell development and is lethal in high-order vertebrates[21].

Figure 1 shows a subset of proteins facilitating vesicle formation. Some of the most studied have been the proteins COP I, COP II and clathrin[22]. COP I is a heteroheptameric protein complex important in the retrograde trafficking from the GA to ER, in the intra-Golgi transport and the maintenance of this organelle[23]. COP II is formed by a stable association of two heterodimers and takes part in the transport from the ER to the GA, ensuring that only the properly folded proteins are exported[24]. Clathrin is the main coat protein involved in the endocytosis. It is essential for signal transduction and neurotransmission and regulates many of the activities in the plasma membrane. Its structural unit is the triskelion, a complex formed by three heavy and three light polypeptide chains. The triskelia associate forming a polyhedral lattice, which covers the vesicle[25]. This structure binds to the membrane through adaptor complexes (APs), forming the characteristic clathrin coat helping in cargo recognition[22]. There are five adaptor proteins, AP1, AP2, AP3, AP4 and AP5, but only the first three interact with clathrin[26]. APs form heterotetramers, found at different levels of vesicle trafficking, helping in cargo selection and concentration. AP1 is located in the TGN and endosomes. It is implicated in the bidirectional transport between these structures. AP2 participates in clathrin-dependent endocytosis and AP3 is found at the endosomal and lysosomal levels. AP4 mediates trafficking between the TGN and endosomes. AP5, seen at the late endosome level, may act by recycling the proteins from the late endosome to TGN, serving as a “backup” when other recycling systems fail[27].

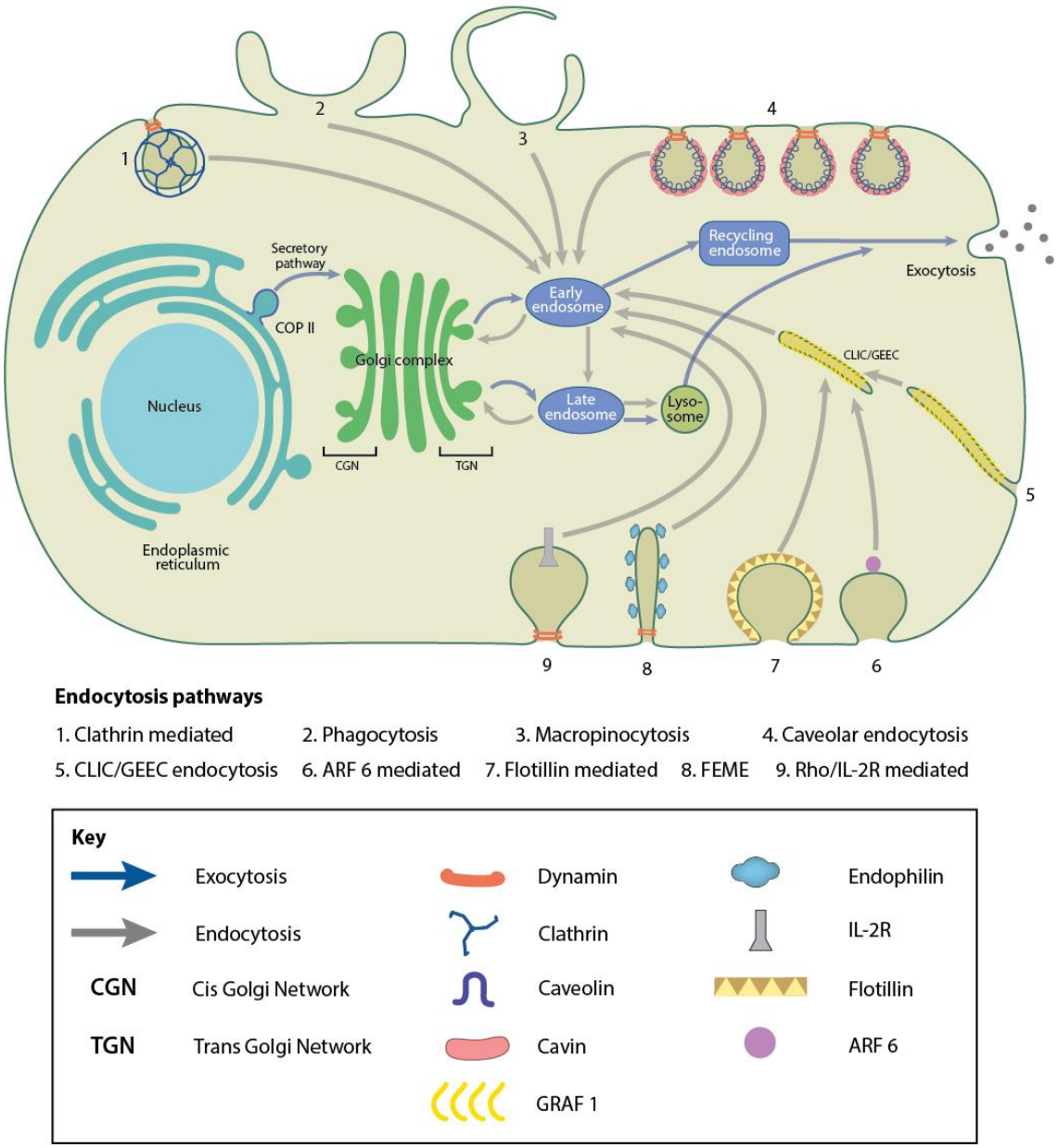


Figure 1: Overview of membrane trafficking systems, with particular attention to the endocytic routes. Used with the permission of MBInfo: Mechanobiology Institute, National University of Singapore.

Apart from the described coat proteins, there are other proteins involved in membrane budding, mainly at the PM level, where the endocytosis takes place (Fig.1). These are the clathrin-independent endocytic routes (CIE).

Phagocytosis and macropinocytosis are specialised endocytic systems for engulfing large particles (like bacteria) and solutes, respectively. In these systems, the coat proteins are not involved, and vesicle formation depends on actin polymerisation[28].

Caveolar endocytosis mainly uses caveolins as coat proteins. These are the membrane proteins required (as well as the cavins) for the formation of membrane invaginations called caveolae. This membrane trafficking system is restricted to metazoans. It is important for the regulation and spatial compartmentalisation of several signalling cascades, participating in cell polarisation, apoptosis, cell cycle and extracellular matrix remodelling[29]. CLIC/GEEC endocytosis refers to clathrin-independent carriers (CLIC) that arise from the PM and mature into tubular endocytic compartments (the glycosylphosphatidylinositol-anchored protein (GPI-AP) enriched compartments (GEEC)). Apart from the internalisation of GPI-APs, this route is used in membrane recycling and PM repair[30]. Arf6 endocytosis is another CIE regulated by the protein Arf6, a member of the Arf family of small GTPases. In this case, internalisation of cargos, such as the major histocompatibility complex I, is facilitated by Arf6-GTP. Hydrolysis of GTP triggers the internalisation of cargo. The exchange of GDP for GTP is facilitated by the guanine nucleotide exchange factor (GEF)[31]. The nucleotide exchange allows the recycling of Arf6 back to the PM. Flotillins are the proteins found in specific micro-domains, or lipid rafts, in the PM; they form another CIE route internalising different proteins, such as CD59[32]. FEME, another CIE pathway, performs the fast endophilin-mediated endocytosis. This is regulated by endophilin, a Bin/Amphiphysin/Rvs (BAR) domain-containing protein. Endophilin forms tubulo-vesicular carriers within seconds and is involved in the transport of G-protein-coupled receptors (GPCRs)[33]. Finally, the Rho-dependent IL-2 receptor endocytosis works with the Rho small GTPase RhoA, and, like the FEME route, is regulated by PI3K[33, 34].

There are also some coat proteins not required for vesicle budding but needed for cargo selection. Retromer[35], the recently discovered retriever[36] and CCC complexes[37] are the coat proteins involved in cargo recycling from the endosomes to PM or GA. They will be described in the next chapters.

Budding of vesicles, initiated by coat proteins, requires additional factors that stabilise the membrane curvature. They affect the lipid composition, insertion of transmembrane proteins or amphipathic helices into the membrane and oligomerisation of proteins with specific domains (e.g., BAR domains). They can change the mechanical involvement of cytoskeleton elements, like actin filaments or microtubules[38]. Scission of the budding vesicle needs dynamin, with oligomerisation

along the neck of the nascent vesicle. Then, the dynamin triggers the fusion of the lipid bilayer, employing the hydrolysis of GTP, resulting in the detachment of the vesicle[39].

Finally, the coat proteins are removed, involving inactivation of GTPases (GTP to GDP state) or the recruitment of enzymes that aid the coat removal[40]. However, the complete disassembly of coat proteins does not take place until the final recognition of the vesicle by the target compartment. Nevertheless, these proteins might have some functions in transport, tethering and fusion of the vesicle; one example is the reported interaction of COP I with the tethering factor Dsl1p at the ER[41].

1.1.2. Vesicle transport

Vesicle movement through the cytosol needs coordination between the vesicle, motor proteins, and the cytoskeleton, mainly actin filaments and microtubules. The filaments are formed by long chains of polymerised subunits of actin. These structures are distributed in the vicinity of the PM and unevenly dispersed in the cytosol[42]. Microtubules are built using heterodimers of α- and β-tubulin, creating a polarised structure initiated at the centrosome, starting from the centre of the cell and spreading towards the PM in a radial structure[43]. Both the microtubules and actin filaments require motor proteins for their activity. These proteins exert a mechanical force, employing ATP hydrolysis, which drives the movement of vesicles. They consist of two globular subunits binding to the cytoskeleton elements and hydrolysing ATP and a flexible tail adhering to the transported entity (in this case, the vesicle). Proteins of the myosin family move along the actin filaments, transporting molecules to the plus (+) end where actin is polymerising[44]. Conversely, the kinesins use the microtubules to transport cargo from the centre to the periphery of the cell, while the dyneins are in charge of retrograde transport, from the PM to the interior of the cell[45].

Vesicle transport through the cytoskeleton is tightly regulated. First, the polarisation of the cytoskeleton elements directs the transport to a specific destination. Second, the coordination of motor proteins with the vesicles requires the activation of the binding site of the vesicle. These binding sites are formed by adaptor proteins, mainly proteins of the Rab family[19]. Rab11, Rab6 and Rab7 proteins have been extensively studied. Rab11 interacts with myosin V, directing vesicles to the surface of the cell[46]. Rab6 associates with kinesin-like protein Rabkinesin-6, participating in the

retrograde trafficking between endosomes and TGN, in intra-Golgi transport and the transport to the ER[47]. Dynein is recruited to late endosomes through a Rab7a-mediated interaction with RILP (Rab-interacting lysosomal protein) and ORP1L. In the late endosome, the tripartite complex Rab7-RILP-ORP1L allows the movement towards the minus end[48].

1.1.3. Vesicle tethering to the target compartment

Once the vesicle has been transported to the target compartment, it needs to be close to the membrane. First, the vesicles and target compartment must recognise each other; then, the vesicles bind and fuse to the compartment. Recognition of specific phosphoinositides and the presence of small GTPases are crucial for vesicle fusion with the target organelle. They combine with soluble N-ethylmaleimide-sensitive factor attachment protein receptors, or SNAREs, to ensure that the cargo is delivered to the correct destination. The vesicle docks to the appropriate target compartment, and the fusion takes place, completing the delivery of the cargo.

However, the SNAREs, phosphoinositides and GTPases are small molecules that cannot penetrate far into the cytoplasm; bigger tethering factors with longer range are also needed[49]. These components are specific for each organelle. The first tethering complex discovered was the homotypic fusion and vacuole protein sorting (HOPS) complex[50]. The complex is composed of six subunits and participates in the fusion of the late endosomes with lysosomes through interaction between Rab7 and the HOPS subunits, Vps39 and Vps41[51]. Strikingly, the two subunits are at the distal locations in the complex; this suggests that the HOPS complex can tether to different membranes containing Rab7, as has been shown in *in vitro* studies[52]. Another tethering factor is the class C core vacuole/endosome tethering complex (CORVET). This is also a heterohexamer, sharing four of its six subunits with the HOPS complex. The tethering specificity is conveyed by the two distinct subunits (Vps8 and Vps3 in CORVET), which can interact with Rab5 instead of Rab7. This suggests that the CORVET aids in the tethering of endocytosed vesicles with early endosomes, where Rab5 is found in abundance[53]. At the ER level, Dsl1 (NRZ in metazoans) tethering system forms a stable complex with SNAREs to mediate the retrograde transport of COP I vesicles from GA to ER[54]. The GA contains several tethering factors, each with different directionality and specificity. There are also some coiled-coil tethers, like p115, GM130 or giantin,

which facilitate transport between ER and Golgi, both retrograde and anterograde[55]. Large tether complexes are also found in the GA. The TRAPP (transport protein particle) complex is involved in the ER-to-GA transport, the conserved oligomeric Golgi complex takes part in the *intra*-Golgi transport, and the Golgi associated retrograde protein complex, in the endosome to *trans*-Golgi retrograde transport[55]. Last but not least, the exocyst, a hetero-octameric protein, is the tethering factor allowing the recognition of vesicles that reach the PM (its architecture has been unravelled using *state of the art* fluorescence techniques[56]).

This multi-layered recognition system (tethering complexes and/or coiled-coil proteins, GTPases, specific phosphoinositides and SNAREs) ensures the correct delivery of cargos to the target organelle. These molecules are also implicated in the coat complex disassembly and vesicle fusion[40,57].

1.1.4. Vesicle fusion with the target compartment

The final step in vesicle trafficking is the vesicle fusion with the membrane of the target compartment. It is believed that most membrane fusion processes go through an intermediate hemifusion stage, where the first step is the merging of the closest monolayers[58]. The process involves the fusion of two membranes, which implies overcoming high thermodynamic barriers. The first barrier is the proximity of two negatively charged membranes, repelling each other, and the second, a strong curvature deformation that can occur during hemifusion. These barriers are overcome by the proteins involved in this process[57], such as the proteins from the SNARE family, essential for membrane fusion[59]. These proteins are present on the surface of the vesicle and the target membrane; they dimerise to facilitate membrane fusion. Most SNARE proteins associate with the membrane via a transmembrane domain. Some members of the family lack this domain but can be prenylated or farnesylated to be anchored to the membrane[60]. In their monomeric state (i.e. not priming the membrane fusion) they are unstructured. Once they are close to another SNARE protein, they can form very stable oligomeric complexes, largely hydrophobic, but with three highly conserved glutamine residues (Q-SNAREs) or one conserved arginine (R-SNAREs). In this “zipper” model, the SNAREs of the vesicle and the target compartment oligomerise from the N-terminal to the C-terminal domain, clamp the membranes together and trigger the fusion[59]. As an example, the fusion of synaptic vesicles requires oligomerisation of synaptobrevin from

the surface of those vesicles and the SNAREs SNAP25 and syntaxin1 in the plasma membrane. They form a four-helix bundle, which draws together the two membranes. The positively charged Q and R residues form an "ionic zero layer", essential for the disassembly of the complex by the N-ethylmaleimide-sensitive factor (NSF)[61].

1.1.5. The endolysosomal system

There are several membrane trafficking routes in the cell, and most of them have a common checkpoint, the endosome. Endosomes are highly heterogeneous in shape and composition. As shown in Figure 2, there are several types of endosomes (early endosomes, recycling endosomes, late endosomes, and multi-vesicular bodies), depending on the maturation stage. They form, together with the lysosomes, the endolysosomal system. These compartments are essential for nutrient acquisition, lipid and protein turnover, protection from pathogens and cell debris (e.g., protein aggregates) and as a reservoir of membranes.

Endocytosis starts with the budding of a fragment of the PM towards the interior of the cell, thus entering the endocytic pathway. Endocytosed vesicles can contain nutrients, solutes or other molecules (e.g., hormones or neurotransmitters) that are recognised by transmembrane proteins (cargos). Internalised vesicles are 60–120 nm in size. Although there are several systems of vesicle formation (described in Chapter 1.1.1), around 95% of endocytosed vesicles are clathrin-dependent[62]. The assembly of coat proteins for clathrin-mediated endocytosis requires several different subunits, apart from clathrin. The adaptor proteins (AP2), clathrin-assembly lymphoid myeloid leukaemia (CALM) protein and the epsins facilitate the interaction of clathrin with the cargo and the membrane, controlling the vesicle size[63]. After this initial coating, actin filaments are formed, contributing to the budding of the vesicle. Actin polymerisation is regulated by the proteins from the Wiskott-Aldrich syndrome protein family (WASP), key activators of myosin motors and actin filament polymerisation. Afterwards, the constriction of vesicles is mediated by endophilin and amphiphysin, BAR domain-containing proteins that cooperate with dynamin in separating the vesicles from the PM (Fig. 2).

It has been suggested that endocytosis may initiate at random sites on the plasma membrane, where the presence of cargo would prevent the collapse of the newly formed coated tip[64]. However, this is not usually the case. There are some preferred areas for

vesicle formation, such as the synaptic region in a neuron, and particular sites on the PM where endocytic events occur repeatedly. These regions are enriched with specific lipids, such as phosphatidylinositol 4,5-bisphosphate ($PI(4,5)P_2$). This is a PM-specific phospholipid, which drives the recruitment of several proteins involved in endocytosis, e.g., epsin or AP2. These two proteins recruit kinases that produce more $PI(4,5)P_2$, feeding forward the system for coat-protein recruitement[65]. It has been suggested that the cargo might not be just a molecule to be ferried. It might also serve as a driving force in vesicle generation, or even as a checkpoint in budding processes, allowing the final bud of the vesicle only when sufficient cargo is loaded[64,66].

Endocytic vesicles can undergo homotypic fusion, creating early endosomes (EE), or they can fuse with an existing EE in a different, currently poorly understood process[67]. The EEs are pleomorphic structures where all internalised material converges. They are characterised by mild levels of acidification ($pH \approx 6 – 6.5$)[68], and they contain Rab5, among other markers[19,69]. Nevertheless, endosomes are highly heterogeneous. It has been shown that different maturation states can coexist in the same endosome, forming micro-domains that sort cargos for different destinations[70]. Cargos can travel in the endosome membrane until they finally reach lysosomes for degradation or they can be recycled. Some cargos, like transferrin receptor (TfR) or the major histocompatibility complex (MHC), are recycled fast (Fig. 2, light blue lines) by the recycling endosomes (REs), a specific subset of endosomes enriched in Rab4 and Rab11[71] proteins (Fig. 2). Although the EEs and REs have distinct surface markers and levels of acidification, the differences are blurred as they both form reticular networks[67]. The EE mature to late endosomes (LE) and multi-vesicular bodies (MVB) before they finally fuse with lysosomes. Endosome maturation involves lumen acidification (by the proton pump V-ATPase), displacement from the periphery of the cell towards the perinuclear region (by the mechanical forces of microtubules), the exchange of Rab proteins and the appearance of intra-luminal vesicles (ILVs) inside the MVB (Figure 2)[72]. During maturation, ligands are released from their cargos due to the drop in pH. The molecules and cargos in the degradative pathway are internalised in the ILVs of the MVB and degraded when the MVB and lysosomes fuse (Figure 2, red line).

Degradation of some cargos is essential for signal downregulation. For instance, the epidermal growth factor receptor (EGFR), which is internalised upon binding the epidermal growth factor (EGF), has to be degraded in the lysosomes to be

downregulated[73]. However, some other cargos undergo recycling from the early and late endosomes, mainly to the PM or TGN. The glucose transporter GLUT1 is recycled back to PM from PI(3)P-containing endosomes, in the so-called direct recycling (Figure 2, blue line and the blue box)[74]. Other cargos, like cation-independent mannose-6-phosphate receptor (CIMPR), reach the LE transporting acid hydrolases from GA. They release the hydrolases and travel back to the GA in the so-called retrograde trafficking (Figure 2, green line and the green box). Recycling happens mostly inside the tubules emanating from the endosome and forming the tubular endosomal network (TEN, Figure 2, green line). This retrograde transport is mediated by recycling complexes, such as the retromer and accessory proteins, in a process whose details are now becoming clear[75,76,77].

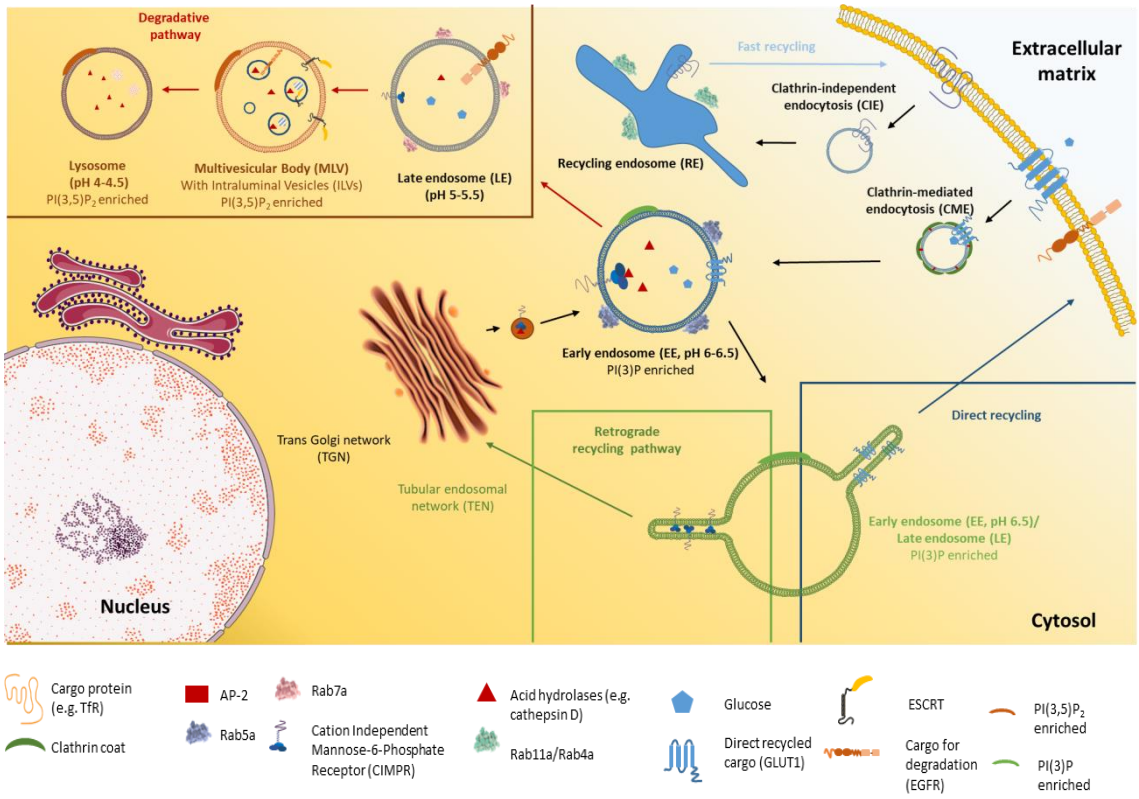


Figure 2: Schematic view of the endolysosomal pathway. In this example, three internalised cargos are shown. The EGFR is internalised and directed to endosomes, where it joins the degradation pathway (red lines and box). Some cargos, like transferrin receptor (TfR), undergo fast recycling from the REs (above, light blue line) without the intervention of any other regulator. Other receptors are also recycled back to the PM (direct recycling, blue box and arrow) with the help of recycling complexes. Some cargos from the interior of the cell, like the CIMPR, are recycled from the endosome to

GA in the so-called retrograde recycling pathway (green box and arrow). Each type of endosome has characteristic markers (Rab proteins) and phosphoinositides.

The endosome is essential for molecule sorting; dysfunctions in different endolysosomal pathways have been linked to several diseases. Malfunctions of lysosome degradation are associated with over 50 different diseases[78], and defective protein recycling leads to severe neurodegenerative disorders, such as Alzheimer´s and Parkinson´s diseases (PD) [79].

1.2. Endosomal cargo recycling

The endolysosomal system is very important in nutrient uptake, but it also has a key role in the protein homeostasis in the cell. Most of the molecules that enter this pathway are internalised as a cargo ferrying them for degradation. However, those cargos, as well as the internalised membrane, cannot be degraded or retained every time they are internalised as it would be metabolically unsustainable. Instead, they are recycled back to their respective origins, whether these are the PM, GA, or the distal organelles like the ER.

The cargos face two possible destinies once they enter the endolysosomal system. They can be targeted for degradation or recycled back. The proteins to be degraded are ubiquitinated and then recognised by the endosomal sorting complex (ESCRT). This is a multi-protein system that recognises ubiquitinated cargos on the endosome membrane, invaginating the endosomal membrane to form the ILV for degradation during later phases[72]. In contrast, the proteins can be recycled by any of the protein complexes involved in the cargo retrieval, such as the retromer, retriever, or the CCC complex (Figure 3)[80]. These are multi-protein complexes that, together with the WASP and SCAR homologue (WASH) and other accessory proteins, can recycle the cargo and avoid lysosomal degradation[80].

These antagonistic activities of ESCRT and retrieval complexes are segregated in the endosomes[80]. The cargos are separated on the endosomal membrane and pre-selected for degradation or recycling. It is possible that an erroneous placement of cargo in the degradative micro-domains on the endosomal membrane can cause degradation of cargos that should be recycled. One of the main players in the micro-domain formation is the WASH complex[81], which will be described in the following chapters.

1.2.1. Retriever

The recently discovered multimeric retriever complex is essential for the recycling of many cargos[36]. A study of $\alpha_5\beta_1$ integrin recycling has reported its retrieval to the PM by the sorting nexin 17 (SNX17). SNX17 does not belong to the canonical (retromer) recycling route. Further studies have found that SNX17 is an accessory protein of retriever. The complex is formed by three subunits: VPS35-like protein (VPS35L), VPS26C, also called DSCR3, and VPS29 (Figure 3B). The predicted structure of VPS35L (previously named C16Orf62, the protein coded by the open frame 62 of chromosome 6) includes HEAT (from Huntington/EF3/PP2A/TOR1) repeats, with a predicted folding similar to that of retromer subunit VPS35. VPS26C is a protein expected to have an arrestin-like fold and is a paralogue of the retromer subunit VPS26[75]. VPS29 is found in both the retriever and retromer complexes[82]. This heterotrimer has been purified by *in vitro* techniques and it has been observed in cell culture[36]. The retriever and retromer complexes are found on the membrane of endosomes sharing a "recycling" micro-domain on their surfaces.

The retromer is thought to be recruited by the active (i.e., GTP-bound) Rab7a and SNX3, in a cooperative manner[83]. However, the retriever is recruited neither by Rab7 nor SNX3[36] but by the WASH complex (Figure 3D). This complex recruits the CCC through the interaction of FAM21 and CCDC93[84], and, at the same time, helps in retriever recruitment. Afterwards, the retriever encounters SNX17 at the endosome. The SNX17 is an accessory protein of the family of sorting nexins (SNX). It contains a PX domain, directly interacting with phosphatidylinositol phosphates (PIPs), and a FERM (F for 4.1 protein, E for ezrin, R for radixin and M for moesin) domain, which interacts with the cargo at the membrane level[85]. The retriever interacts with SNX17 through the C-terminus of SNX17 and the VPS26C retriever subunit, serving as an adaptor to achieve the cargo transport to the PM[36].

A SILAC proteomic analysis has established that the retriever complex is important for recycling of PM cargos involved in cell adhesion and nutrient transport and signalling receptors with the canonical sequence NxxY (N: asparagine, x: any amino acid and Y: tyrosine). This signal sequence is recognised by FERM domain-containing proteins (such as SNX17). This recently discovered recycling protein complex should be further characterised to analyse the different recycling routes and their mutual relationships.

1.2.2. The CCC complex

The CCC is a multi-protein complex formed by up to 12 subunits. It contains up to 10 COMMD proteins (from COMMD1 to COMMD10) in association with CCDC22 and CCDC93[80] (Figure 3C), which can form a large complex with the retriever subunits[37]. It has been identified for the first time as a part of the recycling machinery of ATP7A, a copper transporter recycled to the PM by COMMD1, CCDC22 and CCDC93[84].

The CCC complex is highly conserved in the metazoans, although not all the COMMD subunits are present in all invertebrates, and it is expressed in nearly all tissues. Its primary function is the endosomal protein sorting during development. Defects in the CCC complex are linked to Ritscher-Schinzel syndrome and other developmental abnormalities[37]. However, other functions of these complex proteins have also been reported. It has been shown that COMMD proteins (and especially COMMD1) are important for NF-κβ signalling, ubiquitinating this protein and promoting its degradation and therefore reducing pro-inflammatory response[86]. Another study has shown that the CCC and the WASH complex are essential for the low-density lipoprotein receptor (LDLR) recycling, thus helping to regulate cholesterol levels and reducing hypercholesterolemia[87].

Although several subunits are needed to form the CCC complex, not all are necessary for cargo recycling. Li and co-workers have shown that the Notch receptors (involved in developmental programmes) are recycled to the PM by COMMD5-COMMD9-CCDC22 and CCDC93, while the other COMMD subunits are not necessary[88]. The exact structure of the CCC complex is still unclear. *In vitro* results suggest that the CCC complex and the retriever, pooled together to form a big recycling complex termed Commander[37], are separate entities as silencing of the components of one complex does not affect the other. Nevertheless, it is possible that they co-evolved[36]. The CCC complex is very important for cargo recycling processes, from development to metabolism regulation. Silencing COMMD subunits causes lethality in mouse embryos[88]. It constitutes another non-canonical recycling pathway; before its discovery, the retromer had been the only example.

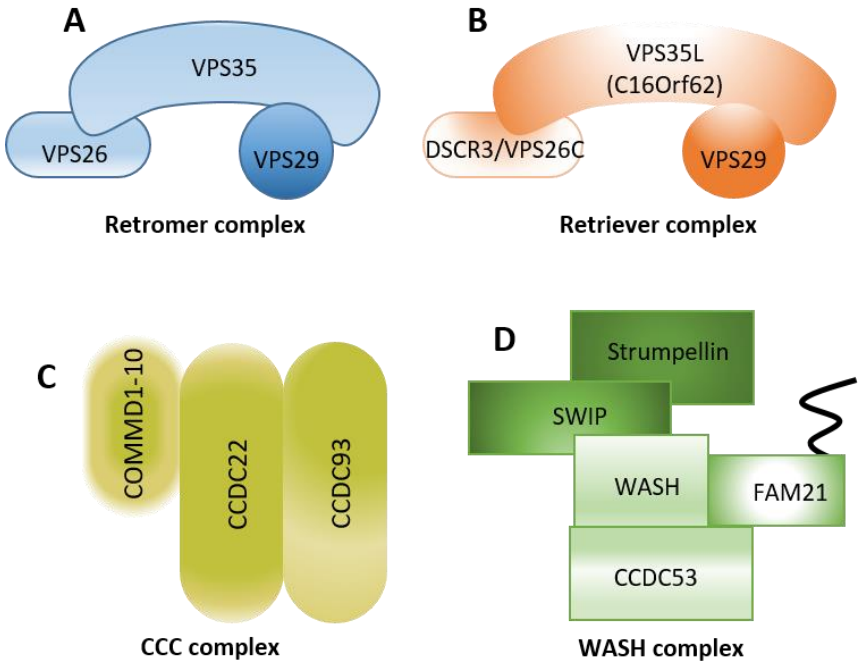


Figure 3: Schematic structures of retromer (A), retriever (B), the CCC complex (C) and the WASH complex (D), with the different subunits highlighted. It can be seen that VPS29 subunit is found both in the retromer and the retriever.

1.2.3. Retromer

The retromer was first described in yeast in 1998 as a system for recycling of Vps10, a cargo that transports acid hydrolases from TGN to yeast lysosome (vacuole)[89]. It is essential for protein transfer from the endosome to the PM (direct recycling, see Figure 2) or to the TGN (retrograde recycling, see Figure 2). It is conserved from yeast to humans and consists of a heterotrimer of VPS26, VPS35 and VPS29 (Figure 3A) (Vps26p, Vps35p and Vps29p in yeast). This heterotrimer is also called the cargo-selection complex (CSC)[90].

Despite a high degree of conservation, there are notable differences between the yeast and human retromers. The yeast retromer is formed by the CSC bound to two auxiliary proteins, Vps5 and Vps17[89]. Vps5 and Vps17 form SNX heterodimers containing PX and BAR domains (SNX-BAR, Figure 4C). SNX-BAR heterodimers can contribute to the tubular endosomal network (TEN, Figure 2, green lines). Their interaction with the retromer favours tubule formation and recycling of cargos in the yeast[90]. Although the heteropentameric yeast complex is not stable *in vitro*[91], the

structure of the yeast retromer complex (Vps26p–Vps29p–Vps35p) together with an homodimer of Vps5 on membrane tubules has been solved recently[92].

In contrast, human retromer does not form stable complexes with SNX-BAR proteins (Figure 4C). Moreover, it has been shown that the human SNX-BAR heterodimer can recycle cargo (CIMPR) in a retromer-independent manner[76, 93]. There is an ongoing controversy over the retromer dependence since a recent study has reported that CIMPR recycling is retromer-dependent[94]. The human retromer has a large number of accessory proteins in comparison with the yeast retromer (Figures 4A and 4B). This increased complexity suggests that human retromer is more specialised than its yeast counterpart. For instance, the yeast retromer has two associated SNX-BAR proteins (Vps5 and Vps17) forming a stable complex. In contrast, the human retromer has been functionally linked to up to five SNX-BAR proteins (SNX1, SNX2, SNX5, SNX6 and SNX32) although these SNXs might also act independently of retromer.

Retromer machinery is used by several intracellular pathogens. *Legionella Pneumophila, Chlamydia trachomatis* and the papillomavirus take advantage of this mechanism to replicate inside the cell, avoiding the immune system[95,96, 97]. Furthermore, retromer malfunction has been linked to severe neurodegenerative diseases[79]. It is important for cargo recycling from endosomes to the TGN or the PM. Malfunctions of this system lead to degradation of cargos as they cannot be recycled, and to protein retention/accumulation in the endosomal compartment. Accumulation of proteins can produce aberrant aggregates which are implicated in Parkinson´s or Alzheimer´s disease (AD). For instance, a specific mutation in the retromer subunit VPS35, D620N, has emerged as a new cause of late-onset, autosomal, dominant Parkinson´s disease[98].

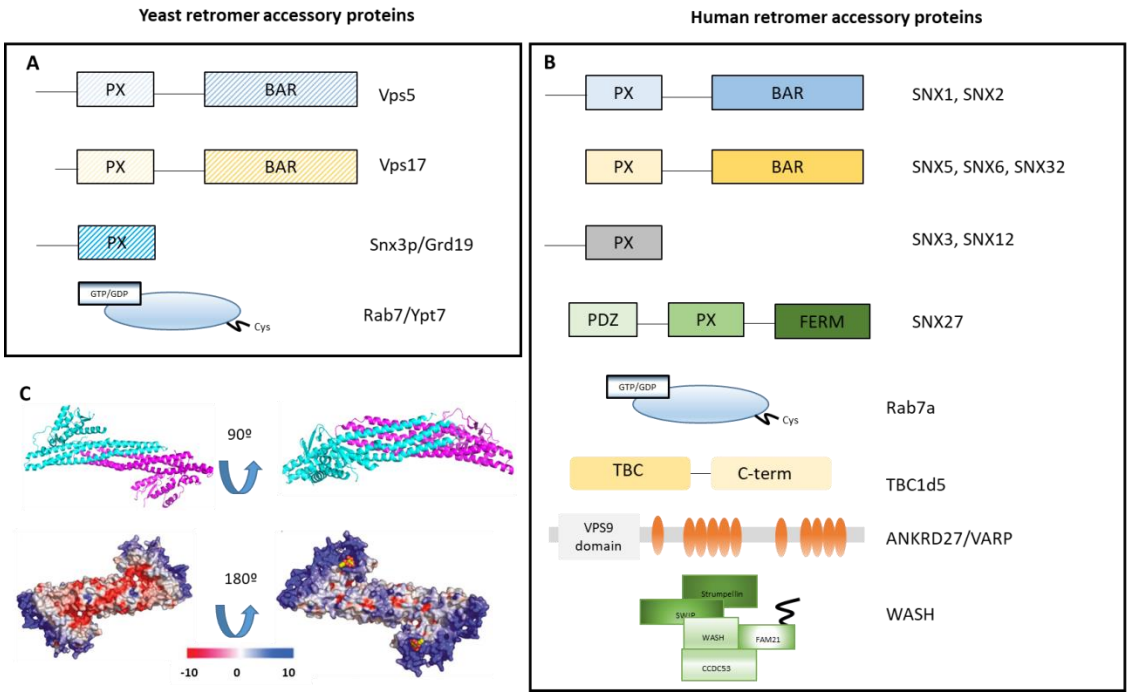


Figure 4: Schematic representation of yeast (A) and human retromer accessory proteins (B). These proteins are expected to interact directly with the retromer, although some of them could act independently. There are other proteins, especially in humans, which can modulate retromer activity via other proteins (e.g., RME-8 can regulate cargo recycling through interaction with SNX1)[76]. (C) Structure of the homodimer of SNX9. The characteristic banana shape of BAR domains can be seen (upper right). The positively charged electrostatic potential of this surface is highlighted in blue (bottom right). Electrostatic scale is shown in Kcal/mol×e. Note that although the picture (C, top) shows a homodimer (i.e., two subunits of SNX9), the subunits are shown in two different colours.

It has also been observed that retromer impairment causes accumulation of amyloid precursor protein (APP) in the endosomal membrane (Figure 5). APP is located on the PM of the cell; it is transported by endocytosis to endosomes and recycled back to the PM in a retromer-dependent manner. The APP can be processed via two proteolytic routes, the amyloidogenic pathway and non-amyloidogenic pathway, depending on the secretase that processes the protein[79, 99]. In the non-amyloidogenic route (Figure 5A), the APP is digested by α-secretase on the PM. The product of this reaction (CTFα) is the substrate of γ-secretase, which is found on the endosomal membrane and produces soluble amyloids with neuroprotective properties[100]. In the amyloidogenic route (Figure

5B), the APP faces β-secretase at the endosomal membrane. It is then digested producing CTFβ, which is, in turn, the substrate of γ-secretase as well. However, one of the products of this last digestion is the β-amyloid, 16 amino acids longer than the product of α-secretase (P3), and one of the hallmarks of AD[101].

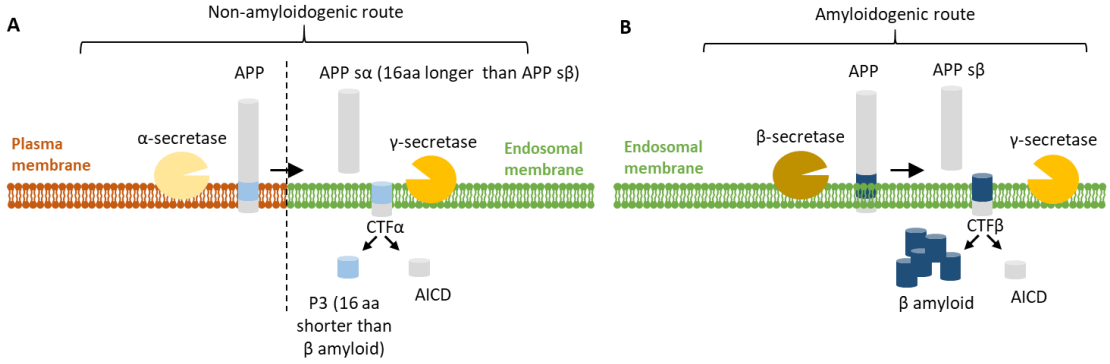


Figure 5: Differential processing of APP renders protective or pathogenic products. (A) The non-amyloidogenic route involves pre-processing of the APP in the plasma membrane. Then, when the product (CTFα) is endocytosed, it is processed by γ-secretase to render neuroprotective peptides (P3). (B) The amyloidogenic route involves the digestion of APP by β secretase at the endosomal membrane. Afterwards, the product (CTFβ) is digested by γ-secretase producing β-amyloid, the causative agent of AD.

It is well known that the patients with AD have enlarged endosomes, which is considered a hallmark of endosomal malfunction. This endosomal anomaly is associated with poor AD prognosis[102]. The amyloidogenic route is favoured by longer retention of the APP at the endosomal level. This retention is caused, at least in part, by retromer dysfunction[79]. Reduced retromer activity slows down APP recycling to the PM, resulting in an increase in β-amyloid production. The importance of retromer complex in AD and PD makes it a possible therapeutic target. In a study published in 2014 in *Nature Chemical Biology*, Mecozzi and colleagues identify a "molecular chaperone" capable of stabilising the retromer under the conditions tested, increasing retromer concentrations and reducing the levels of amyloid-β in neurons[103]. Although the results are preliminary, it suggests new pharmacological strategies to treat the PD or AD. In

any case, further studies of retromer biology are necessary to understand its function and its potential use in medical treatments.

1.2.3.1. Structure of retromer

The structure of a protein gives invaluable information about its shape, function and its role in diseases. Furthermore, 3D structures of therapeutically relevant targets help in drug discovery, as appears to be the case for the retromer[102]. Its complete structure has been solved by several laboratories, using X-ray crystallography,[75, 82] cryo-electron microscopy[104] and cryo-electron tomography[92].

The retromer is a heterotrimeric protein composed of three different subunits: VPS35, VPS29 and VPS26 (Figure 6). VPS35 (Figure 6, red) is a 92-kDa protein; its folding resembles that of other proteins involved in coated vesicle trafficking (e.g., AP and clathrin[82]). It is an elongated α-solenoid structure composed of 33 α-helixes grouped in 16 anti-parallel pairs, forming HEAT repeats. VPS35 wraps on its C-terminal around VPS29 (Figure 6, green), a 20-kDa protein with metallophosphoesterase fold. However, VPS29 has no enzymatic activity, because the catalytic site is occluded by VPS35 and it lacks the critical histidine, which is substituted by a phenylalanine[82]. VPS29 interacts with many retromer effectors regulating the function and localisation of the complex. Finally, VPS26 (Figure 6, cyan) is a 38-kDa protein with arrestin-like fold consisting of two β-sandwich domains. Arrestins are a family of proteins important in signal transduction of G protein-coupled receptors and cargo trafficking[105].

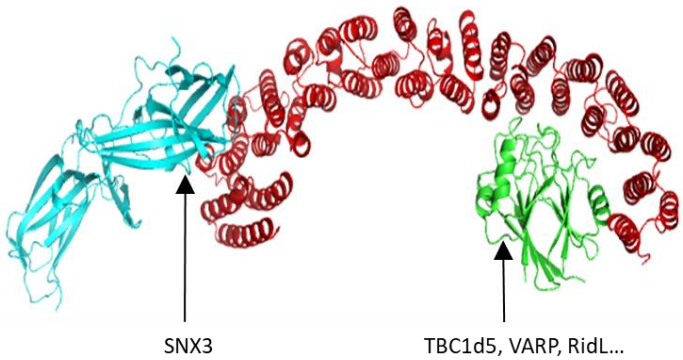


Figure 6: Structure of retromer. VPS26 can be seen at its N-terminal (cyan), VPS35 in the central position (red) and VPS29 (green) at the C-terminal and partially buried in the VPS35 structure. Interaction sites are indicated by arrows.

1.2.3.2. The function of retromer and accessory proteins

The main function of retromer is the recycling of cargos back to the plasma membrane (direct recycling) and/or the retrograde transport of cargos to TGN, (retrograde trafficking). However, the retromer cannot do it on its own. It needs accessory proteins to aid in its recruitment to the endosomes and help in cargo recognition and membrane deformation during recycling.

There are numerous retromer effector proteins; here, I will highlight the most important to date (Figure 4B). First, there is a subset of sorting nexins (SNX), retromer accessory proteins implicated in membrane localisation of the retromer[83], membrane bending and tubulation[106] and even in the direct protein recycling[76]. There are 33 SNX (SNX1 to SNX33), but only 8 are associated with the retromer. SNXs will be described in more detail in the following chapter. Besides SNXs, there are other effectors, such as Rab7a, TBC1d5, VARP and FAM21/WASH complex.

Rab proteins belong to the Ras superfamily of small GTPases[19]. They are specific markers of distinct organelles, where they participate in the recruitment of other proteins. For example, Rab5 identifies the EE, while Rab7 is found on the membrane of the LE and lysosomes[107]. Rab proteins can be inserted into the membrane after the addition of a prenyl group. Rab7, like many other Rabs, contains one or two C-terminal cysteine(s), which can be prenylated. In other words, one or two molecules of farnesyl (C15) or geranylgeranyl (C20) can be added to the terminal cysteine(s), allowing the insertion of Rab into the membrane[19]. Then, Rab7 can be attached to the endosomal membrane. Rab proteins can be found in an active form, bound to GTP, or in an inactive form, bound to GDP. Although they have an intrinsic GTPase activity, this enzymatic function is favoured by GTPase-activating proteins (GAP), such as TBC1d5 (see below). The exchange of GDP for GTP is stimulated by GTP exchange factors (GEF), like Ccz1-Mon1, the GEF of Rab7[108].

Although among the Rab family proteins, Rab7 is the main retromer interactor with nanomolar affinity[109], both Rab5 (indirectly) and Rab7 (directly) are involved in retromer recruitment to the membrane. First, Rab5 is activated (GTP-bound state) and recruited to the membrane of the EE. It engages the phosphatidylinositol 3-kinase (PI3K), which synthesises PI(3)P from phosphatidylinositol. PI(3)P can marshal SNX3, which can assist in retromer recruitment[83]. Later, Rab7 is activated by the exchange of

GDP for GTP (catalysed by the GEF Ccz1-Mon1) and prenylated on the endosomal membrane. There, it works together with SNX3, resulting in the recruitment of retromer. This can be enhanced by the presence of cargo, as has been shown in the yeast[90].

TBC1d5 is a GTPase-activating protein (GAP) of Rab7a[110]. It has a TBC domain at its N-terminal, which probably functions as a GAP, and a presumably disordered C-terminal domain. It can interact directly with the retromer, with nanomolar affinity, via two inserts in the TBC domain, Ins1 and Ins2. Ins1 binds to a hydrophobic pocket of VPS29, and Ins2 to a non-determined area of VPS35[110]. It has been shown that TBC1d5 localises to endosomes due to the interaction with the retromer, and then stimulates the GTPase activity of Rab7. Moreover, this activity is increased by the presence of retromer[110]. This inactivation (i.e., the hydrolysis of GTP) of Rab7 triggers the dissociation of retromer from the LE, probably regulating spatiotemporally the association of retromer to endosomes[110].

VPS9-domain ankyrin-repeat protein, or VARP, is a Rab32/38 effector; it has GEF activity towards Rab21 and binds to VAMP7, an R-SNARE involved in endocytic and secretory pathways[111]. These interactions had been originally expected to be the cause of its endosomal localisation. However, it has been demonstrated later that the VARP interacts directly with retromer subunit VPS29 through the binding site of TBC1d5[112], which is the binding site of pathogen effectors like RidL[95]. Although RidL competes with VARP for VPS29 binding, the VARP protein is still recruited to the endosomes, probably due to interactions with other proteins such as VAMP7. The function of VARP association with the retromer is not clear although this protein is important for cargo recycling from the endosomes to PM (e.g., the retromer cargo GLUT1[112]).

WASH is a multi-protein complex formed by WASH1, strumpellin, CCDC53, SWIP and FAM21. It has been proposed that the WASH works as an obligate complex since a knockout of any of its subunits leads to the degradation of the rest of the components[113]. It is essential for actin nucleation at the endosomal level as a result of recruitment of Arp2/3, a complex that nucleates actin on pre-formed actin filaments[114]. The WASH complex is localised on the endosomal surface in a retromer-dependent manner via interaction with retromer subunit VPS35[98, 114]. This recruitment is effected by the interaction of the long unstructured (~1100 amino acids) C-terminal tail of

FAM21 and VPS35, with VPS29 collaborating in this event[114, 115]. It has been suggested that the WASH complex might be responsible for maintaining the endosome recycling micro-domains[81]. Furthermore, it participates in the formation of endosomal transport carriers, such as the membrane tubules (Figure 2, TEN). The WASH complex triggers the polymerisation of actin, which affords the mechanical force to pull the tubes out of the membrane[116].

There are many, yet uncharacterised, proteins which can interact with the retromer, as it has been shown using quantitative proteomics[115, 117]. For example, the recently discovered ankyrin-repeat-domain-containing protein 50 (ANKRD50) is involved in the direct recycling of nutrient transporters (e.g., GLUT1 and SLC2A1) between endosome and PM. It can interact with retromer accessory proteins, such as SNX27[118]. Nonetheless, no direct interaction of ANKRD50 with the retromer *in vitro* has been reported; however, such interaction might be weak or might need some accessory proteins.

1.2.4. Sorting nexins

Sorting nexins (SNXs) are a group of 33 proteins characterised by the presence of a Phox (PX) domain. PX domains were first identified in 1996 by Ponting in two SH-containing domains of the cytosolic part of NADPH oxidase, called $p47^{phox}$ and $p40^{phox}$[119]. However, the ability of these domains to bind membranes has not been demonstrated until 2002, when the interaction of $p40^{phox}$ with different membrane phosphoinositides, especially PI(3)P, was described[120]. Since then, the domains have been intensely studied, and their interactions with different phosphoinositides, such as the weak interaction of the PX domain of SNX5 with $PI(4,5)P_2$, have been reported[121]. Nevertheless, most PX domains are expected to bind specifically to PI(3)P[122]. A recent study has reported the ability of all known PX domains to bind to different phosphoinositides[123].

All SNXs known to date are shown in Figure 7[124]. They can be classified into three groups. The first group consist of SNXs that, apart from the common PX domain, contain a BAR domain (SNX-BAR, Figure 7, left column). The second group only contains a PX domain (SNX-PX, Figure 7, middle column). The SNXs in the last group contain other domains with a variety of functions (other SNXs, Figure 7, right

column)[124]. Only 8 SNXs have been associated with the retromer: SNX1, SNX2, SNX3, SNX5, SNX6, SNX12, SNX27 and SNX32[75, 125, 76, 117].

SNX1 and SNX2 are the mammalian orthologues of Vps5 from the yeast[125]. They contain PX and BAR domains (described in detail in the next chapters). It has been thought that both PX domains can interact with PI(3)P-containing membranes. However, it has been demonstrated that they bind preferentially to $PI(3,4)P_2$ rather than to PI(3)P[123]. They can form homodimers between themselves and heterodimers with SNX5, SNX6 or SNX32 (see next paragraph). SNX1 and SNX2 have an extended N-terminal domain. It is expected to be unstructured, and its function is not clear although its interactions with other proteins, like DENND5 (also called Rab6IP1), have been described[126].

SNX5, SNX6 and SNX32 are also the SNX-BAR proteins associated with retromer-dependent recycling pathways. In contrast to SNX1/SNX2 and Vps5, it is not clear whether SNX5/SNX6/SNX32 are orthologues of the yeast Vps17[17]. They are expected to form homodimers and heterodimers with SNX1 and SNX2 [93, 127]. Although it seems clear that the heterodimers are the functional units of SNX-BAR proteins[76, 118, 127], there is little information about their formation, structure and biology. SNX1/SNX6 heterodimer was purified for the first time in 2018, during the development of this work[128]; these complexes are now attracting growing attention[76, 77, 93].

SNX3 and the closely related SNX12 only contain a PX domain; they interact exclusively with PI(3)P[123]. It has been reported that the SNX3 is essential for the recycling of some known retromer cargos, such as DMT1-II[75]. The SNX12 is relatively abundant in neuronal tissues[129], and it has been associated with the recycling of CIMPR from the EE to TGN[130].

The SNX27 is directly linked with retromer function, interacting with retromer through VPS26 to recycle cargos from the endosomes to PM[131, 132]. At the N-terminus, it contains a PDZ domain, which can interact with other proteins and cargos[132]. Downstream from the PDZ domain, there is a PX domain, which directly interacts with PI(3)P-containing membranes[123]. Finally, a FERM domain is found at the C-terminus. It is a highly promiscuous protein–protein interaction domain[131].

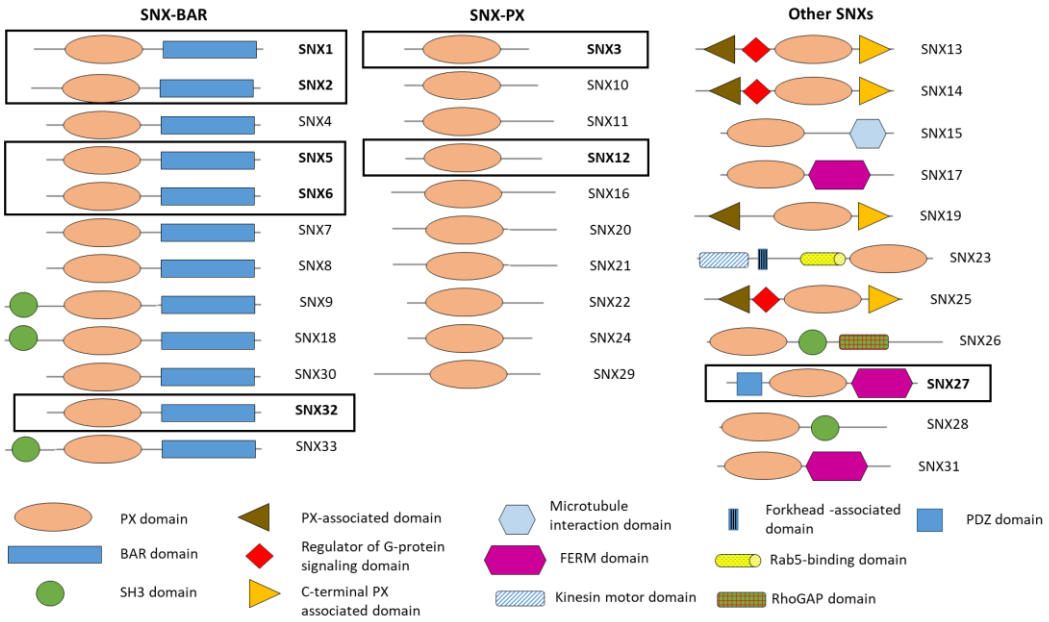


Figure 7: Schematic view of known sorting nexins (SNXs). SNXs associated with retromer pathways are shown in the boxes; SNX1, SNX2, SNX5, SNX6 and SNX32 belong to the SNX-BAR group (left column). SNX3 and SNX12 contain only a PX domain (middle column). SNX27 contains, apart from a PX domain, a PDZ domain and a FERM domain, to interact with cargos or other proteins.

1.2.4.1. Structure and function of sorting nexins

Sorting nexins have essential roles in endocytosis, endosomal sorting and signalling[124]. Thanks to their association with membranes via the PX domain, they mediate membrane localisation of other proteins. They can recruit effectors to trigger cargo recycling[75, 132], create a tubular endosomal network (TEN) by membrane bending and tubulation[116] or even directly recycle the cargos[76, 93, 77]. All these functions are directly related to the domain composition of each protein.

The structure of human SNX3 has been solved using NMR under physiological conditions[133]. It has 162 amino acids arranged in a classical PX fold, with an unstructured N-terminus of about 25 amino acids followed by three anti-parallel β sheets and four α helixes (Figure 8A). Most of the PX domains are very similar (Figure 8A-D), except for PX domains of SNX5 (Figure 8E) and those of the closely related SNX6 and SNX32. The primary function of SNX3 (and of SNX12, see Figure 8B) is the interaction with PI(3)P. Besides this direct interaction, SNX3 inserts a loop in the membrane for stabilisation. Lenoir and colleagues have also shown that phosphorylation

of a specific serine (Ser73) completely abolishes the SNX3–PI(3)P interaction[133]. It suggests a biological switch for the SNX binding to the membrane.

Moreover, the SNX3 structure has been solved in complex with VPS26 (Figure 8F, cyan) the N-terminal part of VPS35 ($VPS35_{14-470,}$ or VPS35N. Figure 7D, red) and a peptide corresponding to the retromer cargo DMT-II[75]. This detailed structure shows that the N-terminal part of SNX3 binds to a groove between VPS26 and VPS35N. The authors also address the question of cargo recruitment by this complex. They show that the cargo is recognised by the interface between VPS26 and SNX3. Furthermore, they demonstrate that the interaction of SNX3 with the retromer depends on the presence of cargo; a tripartite interaction between the retromer, SNX3 and the cargo takes place. The study has helped to formulate a mechanism for cargo recognition and retrograde recycling by the retromer and SNX3.

There is a subset of SNXs that contain a Bis/Amphiphysin/Rvs (BAR) domain, also called SNX-BAR. The BAR domain is composed of three coiled-coil α-helices, which can self-dimerise and sense/induce the membrane curvature[134]. This domain will be further described in the next chapter. The human SNX-BAR proteins have been widely studied; however, it has not been clear until recently whether the functional unit takes a homodimeric or heterodimeric form[125]. In 2017, Simonetti and colleagues showed that the SNX-BAR heterodimers (SNX1/SNX5) can recycle the model cargo CIMPR from the endosomes to TGN (i.e., retrograde recycling) in a retromer-independent manner[76]. Then, later in 2019, they extended their results providing extensive knowledge of what they denominated as “endosomal SNX-BAR sorting complex for promoting exit – 1”, or ESCPE-1 [77]. They demonstrated that the cargo is directly recognised by the PX domain of SNX5 (and SNX6) and the recycling is driven by heterodimers without the intervention of the retromer[77]. However, further biochemical and structural characterisation of SNX-BAR heterodimers is necessary to understand its function.

The described SNXs are involved in retrograde transport from the endosomes to TGN (Figure 9). However, it has been suggested that the SNX1/SNX5 heterodimer also prompts the direct recycling of Sema4C from the endosomes to PM[77]. Nevertheless, the main SNX implicated in direct recycling (i.e., from endosome to PM) is SNX27 (Figure 9). This protein is essential for recycling of several transmembrane cargos that perform their functions in the PM[117]. SNX27 can directly interact with those cargos through a PDZ domain (Figure 8G). The domain is composed of approximately 90 amino acids

and recruits cargos that contain a PDZ binding motif (PDZbm)[132]. Many cargos contain class I PDZbm, for example, the potassium channel Kir3.3, the nutrient transporter GLUT1 and the β2 adrenergic receptor, all of them localised on the PM. PDZbm has an S/T-x-Φ motif (S/T stands for serine or threonine, x for any amino acid and Φ, any hydrophobic amino acid) and is usually located at the C-terminus of proteins. Cargo recognition by SNX27 is improved by the presence of the retromer[135].

Moreover, acidic amino acids at positions -3 and -5 from the PDZbm strengthen the interaction with SNX27 PDZ domain. Phosphorylation of Ser or Thr residues downstream of the canonical recognition sequence adds a negative charge, promoting the interaction with SNX27 and the retromer[135]. This post-translational modification could serve as a switch to regulate cargo transport. Finally, the FERM domain is a widespread domain involved in protein–protein interactions[85]. It has been shown that the FERM domain of SNX17, involved in retriever recycling, can directly recognise cargos that contain the sequence NPxY/NxxY[131]. In the SNX27, this domain seems to be involved in downstream cargo recognition, such as the interaction with FAM21/WASH complex. The interactions of the FERM domain of SNX27 with SNX1 might link the SNX27 with the tubular endosomal network[117]. Interactions of SNX27 FERM with phosphoinositides, especially bi- and tri-phosphoinositides, have also been reported[136]. Therefore, SNX27 can be recruited to PI(3)P-enriched membranes (endosomal membranes) through its PX domain, but also to PI(3,4,5)P_3 membranes, which links its function to cargo recycling from the endosomes to PM.

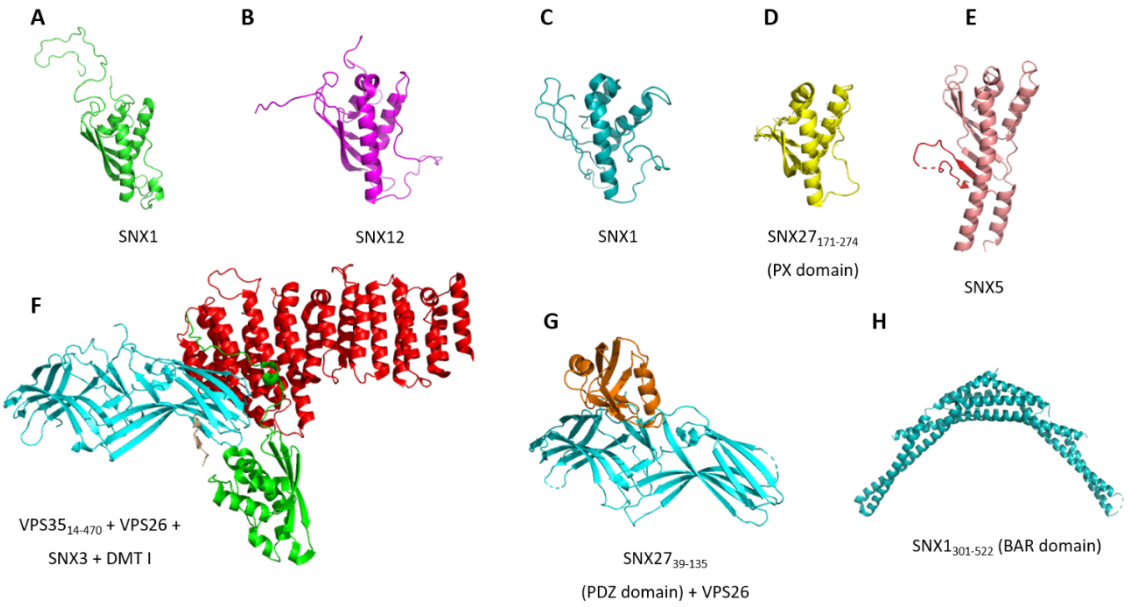


Figure 8: Known structures of different sorting nexins (SNX). (A) The structure (obtained using NMR) of full-length human SNX3 (PDB accession number: 2MXC). (B) Structure of full-length human SNX12 (PDB accession number: 2CSK). (C) PX domain of human $SNX1_{142-269}$ (PDB accession number: 2I4K). (D) Crystal structure of the PX domain of $SNX27_{171-274}$ (PDB accession number: 4HAS). (E) Crystal structure of the PX domain of $SNX5_{22-170}$ (in salmon-pink) bound to a peptide from $CIMPR_{2347-2374}$ (in red). (PDB accession number: 6N5X). (F) Crystal structure of VPS26 (cyan) in complex with the N-terminal part of $VPS35_{14-470}$ (VPS35N, in red) and SNX3 (in green). This complex was solved bound to a peptide corresponding to the retromer cargo DMTI–II (yellow) (PDB accession number: 5F0L). (G) Crystal structure of VPS26 (cyan) in complex with PDZ domain of $SNX27_{31-135}$ (orange) (PDB accession number: 4P2A). (H) Crystal structure of human $SNX1_{K301-522}$ (BAR domain) (PDB accession number: 4FZS).

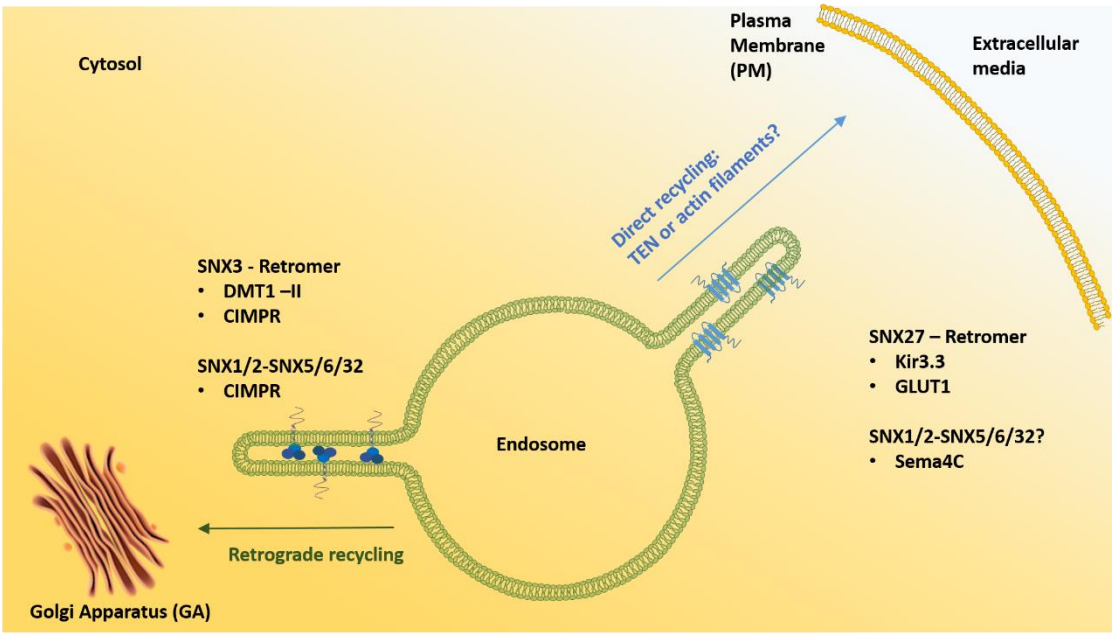


Figure 9: Summary of the participation of different SNXs in cargo recycling. SNX3 collaborates with the retromer in retrograde recycling of cargos such as DMT1-II[75]. SNXs-BAR (SNX1, SNX2, SNX5, SNX6 and SNX32) participate in retrograde recycling of cargos (such as CIMPR) to GA [76, 118] although it has been associated with direct recycling of cargos like Sema4C[77]. SNX27 can recycle the cargos back to the PM in collaboration with the retromer[74]. It is not clear if direct recycling takes place on the tubular endosomal network (TEN) or via actin filaments. Note that the dimensions are not to scale, and the mechanisms are simplified for clarity.

1.2.4.2. Membrane deformation by BAR domains

A very important feature of SNX-BAR proteins is their ability to bind to membranes and, in some cases, deform them producing the tubular endosomal network (TEN), where the cargo is transported to a specific target in tubulated membranes derived from the endosome[116]. Tubular carriers are ideal for transportation of membrane-bound components due to the high surface-to-volume ratio, which allows cargo sorting without much exchange of luminal content. However, to maintain the tubular shape, the membrane needs some configuration-stabilising factors.[137]. BAR domain-containing proteins can bind and stabilise such tubular carriers.

SNX-BAR membrane recognition and binding are made possible by the interaction of the PX domain with phosphoinositides[123] and electrostatic attraction of the negatively charged surface of BAR domains to the membrane[138]. Then, the membrane deformation

is driven by the BAR domains. These domains have been extensively studied as they are found in many proteins involved in membrane binding and deformation. There are five commonly accepted types of BAR domains[134, 139]: (1) the classical BAR domains; the first of this type to be discovered and crystallised was Arfaptin[140]; (2) N-BAR proteins, such as endophilin (Figure 10A), which contain an N-terminal amphipathic helix (Figure 10B) that can be inserted into the membrane and induce curvature; (3) BAR-PH proteins, such as APPLs, which contain a PH (pleckstrin homology) domain at the C-terminus to interact with phosphoinositides[20]; (4) the PX-BAR or SNX-BAR (Figure 8, left column); (5) F-BAR (FCH-BAR) proteins, such as Toca-1, that can induce a negative curvature. This classification has been extensively reviewed, and it is now often reduced to three types: N-BAR (which includes classical BAR, SNX-BAR and PH-BAR domains), F-BAR and I-BAR domains[141, 142].

Despite the differences, these domains share some common features. They are all involved in several membrane remodelling processes, from vesicle budding to neurite outgrowth[143]. Structurally, BAR domains are formed by coiled-coil α-helices. They can make homodimers or heterodimers (with other BAR domain-containing proteins). This dimerisation creates a curved structure with a positively charged concave surface ("banana shape", see Fig. 4C) that can interact directly with negatively charged membranes[139]. After binding to the membrane, they can induce curvature *de novo*, stabilise curvature generated by other sources (e.g., actin cytoskeleton), or recruit other factors for downstream processes. For example, the structure of BAR-containing protein arfaptin was solved while bound to the small GTPase Rac[140]. This binding improves the arfaptin specificity to associate with the GA membrane[143].

Several factors affect membrane bending[138]. One of these is the lipid composition of the membrane. Chemical properties of the different headgroups and acyl chains can promote different membrane curvatures[143]. Furthermore, specific headgroups can recruit other factors that contribute to membrane deformation, e.g., recruitment of SNX through the interaction of the PX domain with PI(3)P. Another important factor is the presence of transmembrane proteins. They can have a conical shape that helps in membrane bending, or they can just cluster together, increasing the membrane curvature. The third factor is the contribution of the cytoskeleton, especially of actin filaments; the microtubules are involved in vesicle transport rather than vesicle budding.

It is well known that actin responses to membrane tension and some processes, like phagocytosis and micropinocytosis, are dependent on the actin cytoskeleton[138].

Furthermore, the BAR domains are linked to the actin cytoskeleton either on their own or through small GTPases[141]. The fourth factor affecting membrane bending is the protein scaffolding around the membrane. Clathrin, COPI or COP II interact with membranes and aid in membrane remodelling. BAR domain-containing proteins on the membrane, like SNX-BAR, are responsible for membrane tubulation[106], although not all SNX-BAR proteins can produce this effect[106]. Another process that facilitates membrane deformation, characteristic for the N-BAR domains, is the insertion of an amphipathic helix into the membrane. A recent study has reported that the amphipathic helix remains unstructured in the absence of membrane binding. When the protein approaches a membrane, the amphipathic helix folds and inserts into the membrane, triggering membrane curvature[144]. In some cases, this helix is not inserted into the membrane but helps in the lattice stabilisation by lateral contacts between adjacent BAR domains[145] (Figure 10B). Moreover, truncation of this helix impairs membrane deformation by proteins like endophilin (Figure 10B) or SNX1[145, 106].

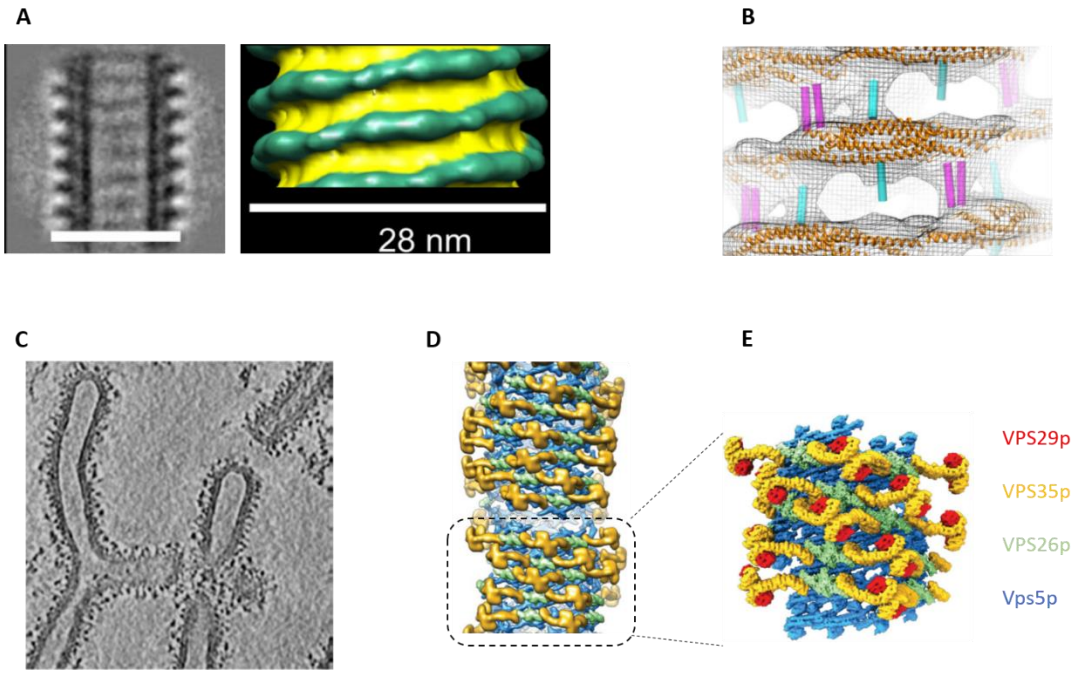


Figure 10: Membrane deformation by BAR domains. (A) Reconstruction of endophilin BAR domain tubulating a membrane. Left, 2D classes of EM images of endophilin bound to a tubule. Right, a model of the BAR domain tubulating a membrane. (B)

Endophilin has an amphipathic helix (blue and pink cylinders) that does not insert itself into the membrane. It assists in membrane deformation by stabilisation of the helix, using lateral contacts. (C) CryoEM image of yeast SNX Vps5p tubulating an artificial membrane in the presence of yeast retromer. (D) Model of tubulation of Vps5p-retromer on a membrane. (E) Detailed view of the helix. The heterogeneity of the model obtained using cryo-electron tomography can be appreciated here. The images in A and B were adapted from Carsten et al.[145]. Images in C–E were adapted from Kovtun et al.[92].

The exact mechanism of deformation of the membrane by the BAR domain-containing proteins are still a hot topic. The first structure of BAR domain producing tubules from artificial membranes dates from 2008[134]. The researchers took advantage of helical geometry of the BAR domain of Toca1 protein, adopted on the membrane, to solve the structure by helical reconstruction of cryo-electron microscopy (cryoEM) images. However, not all the proteins enter a perfect helix state on the membranes. Kovtun and colleagues have shown that yeast SNX VPS5p adopts a "pseudo-helical" arrangement (Figure 10C-E), which makes a helical reconstruction of cryoEM images impossible[92]. Instead, a more versatile technique such as the cryo-electron tomography (cryoET), followed by sub-tomogram averaging, can be used to circumvent the heterogeneity of the sample.

2. OBJECTIVES

2. OBJECTIVES

Although the recent research[92] has shown, for the first time, how the yeast SNX homodimers are arranged on a membrane, the mechanism of this process in mammals is not well understood. Moreover, mammalian SNX-BAR proteins act as heterodimers rather than homodimers. How this asymmetrical arrangement of SNX-BAR heterodimers takes place remains an open question. Moreover, it has been shown that SNXs can recruit the cargo and recycle it in a retromer-independent manner[76,77]. However, there is little biochemical or structural information on SNX heterodimers. Obtaining new data on cargo recruitment by SNX-BAR heterodimers and their distribution on the membrane is essential for understanding different recycling mechanisms orchestrated in the cell.

To determine functional differences between homodimers and heterodimers of sorting nexins, my work was focused on:

- The atomic and functional analyses of sorting-nexin heterodimers.
- Elucidation of the molecular mechanism of cargo recruitment by sorting nexins.
- Clarification of the structural mechanism of tubule formation mediated by sorting nexins.

3. MATERIALS

3. MATERIALS

3.1. Bacterial strains and culture

Heterologous expression of proteins was performed in *Escherichia coli*. Bacterial strains used in this project are shown in Table 1. XL1-Blue was used for cloning and plasmid amplification. BL21 (DE3) and Rosetta (DE3) strains were used for protein expression.

Strain	Genotype	Reference
XL1-Blue	*endA1 gyrA96(nal^R) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB$^+$ lacI^q Δ(lacZ)M15] hsdR17(r_K^- m_K^+)*	Stratagene
BL21 (DE3)	*E. coli B F - dcm ompT hsdS(r m) gal λ (DE3)*	Stratagene
Rosetta (DE3)	*E. coli* B F$^-$ *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm (DE3) pRARE (Cam^R)*	Sigma

Table 1: Bacterial strains used in this study

Bacteria were grown, and protein expression was carried out in Luria Bertani medium (LB, Pronadisa, Spain) or auto-induction medium[146]. Tryptone and yeast extract were obtained from AppliChem (Panreac, Spain). $(NH_4)_2SO_4$, KH_2PO_4, Na_2HPO_4, $MgSO_4$, glucose, glycerol and α-lactose were purchased from Sigma. Bacteria were grown in borosilicate Erlenmeyer 250-ml (for pre-inoculum) or 5-litre flasks (VWR International, LLC, USA), in shaking incubators Innova 44. For plasmid production, 10-ml aliquots of LB culture were grown in 50-ml Falcon tubes (Sarstedt, Germany). For solid-media culture, Petri dishes (VWR, USA) with LB supplemented with 1.5% (w/v) of agar (Pronadisa, Spain) were used. All culture media and materials were sterilised at 134 ºC for 20 minutes unless purchased already sterile (e.g., Petri dishes or Falcon tubes). When needed, ampicillin at 50 mg/ml and kanamycin at 25 mg/ml (Sigma) were prepared in Milli-Q water, filtered through a 0.2-µm filter (Sartorius) and added (diluted 1000 times) to the sterilised media. Chloramphenicol (Sigma) was prepared as a 35 mg/ml solution in ethanol 100% and filtered through a 0.2-µm filter.

3.2. DNA

Template DNAs for SNX1, SNX2, SNX5 and SNX6 were kindly donated by Dr Juan S. Bonifacino, from Cell Biology and Neurobiology Branch, Eunice Kennedy Shriver National Institute of Child Health and Human Development (NIH, Bethesda, USA). Cation independent mannose-6-phosphate receptor (CIMPR) was obtained from the DNASU Plasmid Repository (HsCD00403292).

3.3. Plasmids

Plasmids used in this project are shown in Table 2. The proteins were cloned using expression vectors pGST-Parallel 2 (pGEX derived plasmid, constructed at NIH), and pET HisSUMO-GFP[147]. pHisMBP-Parallel 2 was also used; the clone was constructed by Ander Vidaurrazaga.

In the case of pGST-Parallel 2 (pGST-P2) vector, ampicillin resistance served as a selectable marker for positive clones. Glutathione S-transferase (GST) was used as an affinity tag. The vector contains the TEV (tobacco etch virus) protease site after the GST tag, so the GST can be removed from the expressed protein.

pET HisSUMO-GFP is a pET28-derived bacterial expression vector. It has a gene for resistance to kanamycin (selectable marker) and a poly-histidine affinity tag. This poly-His is flanked downstream with N-terminal (100aa, 11.5 kDa) SUMO3 fusion partner to increase protein solubility. Furthermore, this tag can be removed using a protease specific for SUMO3 structure, the sentrin-specific protease 2 (SENP2). The vector contains a GFP tag as well; however, it was not used in our experiments.

3.4. Oligonucleotides

Oligonucleotides used in this project are shown in Table 3. All oligonucleotides were designed to have a melting temperature between 60 and 70 ºC. They were purchased from Invitrogen, 25 or 50 nanomoles with purity ≥ 98%. When Gibson assembly[148] was used for cloning, the primers were designed to amplify the vector pET28 HisSumo3 from *AgeI* site and *XhoI* site (see Table and Figure 11). In the case of GST-P2 vector, *EheI* site and *XhoI* sites were used. Later, we designed constructs to be inserted instead of the GFP gene (for pET Sumo3 vector) or downstream from GST (for pGST-P2 vector), in frame with the corresponding tags and the stop codon (Figure 11).

Mutagenesis primers were designed to amplify the antibiotic-resistant genes (see Table 3 and Figure 11). For further details of Gibson assembly, see section 4.1.4.

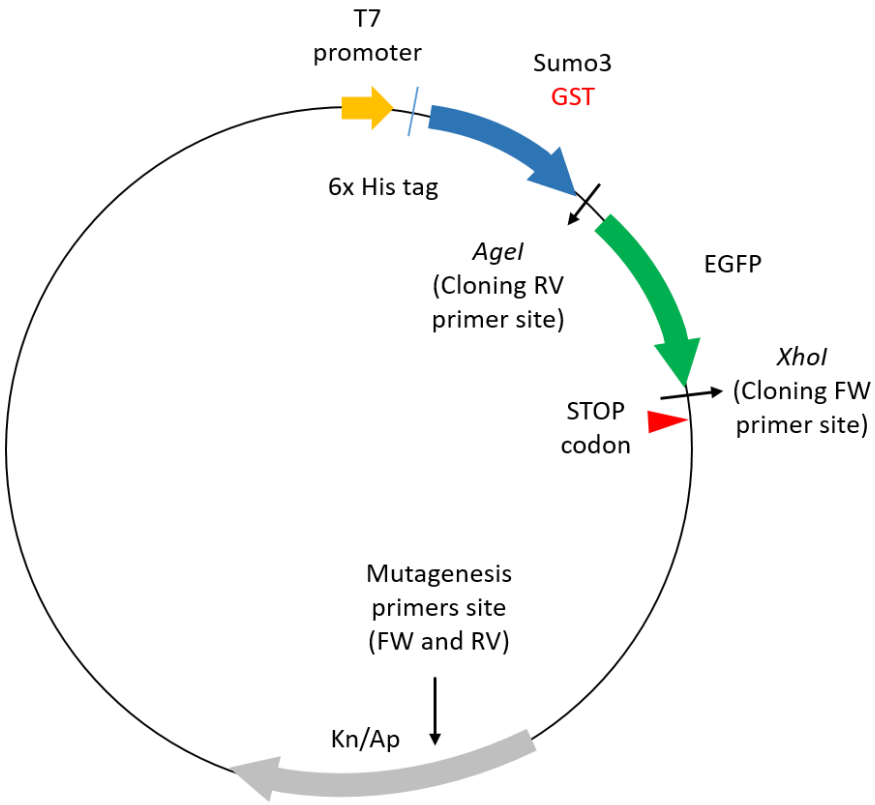


Figure 11: Schematic representation of the vectors (pET HisSumo3 GFP and pGST-P2) used in this work. Both vectors contain a T7 promoter activated by IPTG induction, an affinity tag (HisSumo3 or GST), a STOP codon and an antibiotic-resistance selectable marker. Gibson assembly primers are designed in the indicated sites in frame with the affinity tag and the stop codon. In the case of pET HisSumo3, there is a GFP downstream from the tag (removed for protein expression).

Table 2: Plasmids and constructs used in this project. The already existing constructs are indicated with an asterisk (*).

Plasmid	Characteristics	Resistance
pET HisSumo3 SNX1 FL*	Human full-length SNX1 with His tag at the N-terminus	Kanamycin
pET HisSumo3 SNX2 FL*	Human full-length SNX1 with His tag at the N-terminus	Kanamycin
pGST-P2 SNX5 FL	Human full-length SNX5 with GST tag at the N-terminus	Ampicillin
pGST-P2 SNX6 FL	Human full-length SNX6 with GST tag at the N-terminus	Ampicillin
pET HisSumo3 $SNX1_{(K301-S522)}$	Human BAR domain of SNX1 (K301-S522) with His tag at the N-terminus	Kanamycin
pGST-P2 $SNX5_{(D195-N404)}$	Human BAR domain of SNX5 (D195-N409) with GST tag at the N-terminus	Ampicillin
pET HisSumo3 $SNX1\ FL_{(F347A)}$	Human full-length SNX1 with His tag at the N-terminus, and one mutation in the BAR domain	Kanamycin
pET HisSumo3 $SNX1\ FL_{(W511A)}$	Human full-length SNX1 with His tag at the N-terminus, and one mutation in the BAR domain	Kanamycin
pET HisSumo3 $SNX1\ FL_{(F347A+W511A)}$	Human full-length SNX1 with His tag at the N-terminus, and two mutations in the BAR domain	Kanamycin
pGST-P2 SNX5 FL (Y219A + R368A+M233A+V240A)	Human full-length SNX5 with GST tag at the N-terminus, and four mutations in the BAR domain	Ampicillin
pGST-P2 SNX5 FL (Y219A + R368A + M233A + V240A + I398A + F401A)	Human full-length SNX5 with GST tag at the N-terminus, and six mutations in the BAR domain	Ampicillin
pHisMBP P2 $SNX1_{(142-282)}$*	PX domain of human SNX1	Ampicillin
pGST-P2 $SNX5_{(22-170)}$	PX domain of human SNX5	Kanamycin
pET HisSumo3 CIMPR (2330-2491), C2341A + C2342A	The cytosolic tail of CIMPR (2330-2491), with two cysteines at 2341 and 2342 mutated to alanine	Kanamycin Chloramphenicol
pET HisSumo3 CIMPR (2330-2491), C2341A + C2342A + W2369A + L2370A + M2371A +	The cytosolic tail of CIMPR (2330-2491), with the canonical recognition site (WLM) mutated to alanines, and two cysteines at 2341 and 2342 mutated to alanine	Kanamycin Chloramphenicol
pET HisSumo3 $CIMPR_{(2372-2491)}$, C2341A + C2342A	Part of the cytosolic tail of CIMPR (2330-2491), with two cysteines at 2341 and 2342 mutated to alanine	Kanamycin Chloramphenicol

Table 3: Oligonucleotides used in this project. The already existing constructs are indicated with an asterisk (*).

Plasmid	Oligonucleotide	Cloning method
pGST-P2 SNX5 FL (Y219A+R368A+M233A+V240A)**	GAGCAAGAGAAGAACTTCCTTATTAACGCATACAATAGGATCAAAGATTCTTGTGTGAAAG CTTTCACACAAGAATCTTTGATCCTATTGTATGCGTTAATAAGGAAGTTCTTCTCTTGCTC GAAGAACTGATAAATTTCAAACGGAAGGCAGTGGCAGCATTTAGAAAGAATCTAATTG CAATTAGATTCTTTCTAAATGCTGCCACTGCCTTCCGTTTGAAATTTATCAGTTCTTC CAAAGATTCTTGTGTGAAAGCTGACAAAGCGACCAGATCTCATAAAAATGCAGCCGATGACTATATCCACACCGCAGCCTG CAGGCTGCGGTGTGGATATAGTCATCGGCTGCATTTTTATGAGATCTGGTCGCTTTGTCAGCTTTCACACAAGAATCTTTG + mutagenesis primers of pGST P2	Gibson Assembly
pGST-P2 SNX5 FL (Y219A+R368A+M233A+V240A+ I398A+F401A)	It was done from the previous quadruple mutant GAACAATGTCTCCCTTTTGCAGAGCTGTGCAGACTTGGCAAGAATAACTAAGCAGTCTCGAGCGGC GCCGCTCGAGACTGCTTAGTTATTCTTTGCCAAGTCTGCACAGCTCTGCAAAAGGGAGACATTGTTC + mutagenesis primers of pGST P2	Gibson Assembly
pET HisSumo3 CIMPR (2330-2491), C2341A + C2342A	CGAGGACACCATCGACGTGTTCCAGCAGCAGACCGGTGGAGAGAGGAGGGAAACAGTGATAAG GCAGCCGGATCTCAGTGGTGGTGGTGGTGGTGCTCGAGTTATTTTGGAGGATGGTCGCCCAC + Gibson assembly primers of pET HisSUMO3-GFP	Gibson Assembly
pET HisSumo3 CIMPR (2330-2491), C2341A + C2342A + W2369A + L2370A + M2371A	GAGACAGATGAGAATGAAACAGAGGCTGCCGCAGAAGAGATCCAGCTGCCTCCTC GAGGAGGCAGCTGGATCTCTTCTGCGGCAGCCTCTGTTTCATTCTCATCTGTCTC + mutagenesis primers of pET HisSUMO-GFP	Gibson Assembly
pET HisSumo3 CIMPR(2372-2491), C2341A + C2342A	CGAGGACACCATCGACGTGTTCCAGCAGCAGACCGGTGGAGAAGAGATCCAGCTGCCTCC GCAGCCGGATCTCAGTGGTGGTGGTGGTGGTGCTCGAGTTATTTTGGAGGATGGTCGCCAC	Gibson Assembly
pET HisSumo3 -GFP: Gibson assembly primers for cloning (up) and mutagenesis (bottom)	Gibson assembly primers for cloning: TCCACCGGTCTGCTGCTGGAACACGTCGATGGTGTCCTCG (Reverse primer) CTCGAGCACCACCACCACCACCACTGAGATCCGGCTGC (Forward primer) Gibson assembly primers for mutagenesis: GGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAG CTCTGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCC	-
pGST-P2: Gibson assembly primers for cloning (up) and mutagenesis (bottom)	Gibson assembly primers for cloning: TCCCATGGCGCCCTGAAAATACAGGTT (Reverse primer) TGCAGTCTCGAGCGGCCGCATCGTGAC (Forward primer) Gibson assembly primers for mutagenesis: GTGGAAGCGGCGATGGCGGAGCTGAATTACATTCCCAACC GGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCAC	-

(**) Mutations were introduced one by one, except for M233A and V2340A, which were introduced using the same primer.

3.5. Peptides

$CIMPR_{2348-2375}$ was ordered through GenScript, with purity > 95%.

3.6. TEV and SENP2 proteases

Two proteases were used to remove the purification tags. For the GST tag (pGST-P2 vector), TEV protease was employed. The plasmid encoding TEV protease was obtained from the laboratory of Dr David S. Waugh (NCI, CCR, USA). The protease was expressed in *E. coli* BL21 (DE3), and it was purified as previously described[149].

SENP2 has been cloned into pET28 vector. It was kindly donated by Prof. Karl-Peter Hopfner from the Gene Centre in Munich. SENP2 was expressed in *E. coli* Rosetta (DE3) cells as described previously[150].

3.7. Crystallisation screenings and heavy atoms

Protein crystallisation was carried out at CIC bioGUNE X-ray Platform. Several commercial kits for high throughput screening of protein crystallisation conditions were tested (Table 4).

The heavy-atom compounds screened, all purchased from Hampton Research Corp. can be seen in Table 7 (Methods section).

Table 4: Commercial kits for protein crystallisation screening

Crystallisation screening kits	Source	Catalogue number
PEG-ion 1/2	Hampton Research Corp.	HR2-126/098
Morpheus™	Molecular Dimensions Ltd.	MD1-46
ProPlex™	Molecular Dimensions Ltd.	MD1-38
Index	Hampton Research Corp.	HR2-144
JCGS I	Qiagen	130724
JCGS II	Qiagen	130725
JCGS III	Qiagen	130726
JCGS-plus™	Molecular Dimensions Ltd.	MD1-37
Salt Rx 1/2	Hampton Research Corp.	HR2-107/109
Natrix	Hampton Research Corp.	HR2-116/117
GRID Screen™ Ammonium Sulphate	Hampton Research Corp.	HR2-211
PACT premier	Molecular Dimensions Ltd.	MD1-29
Structure screen 1/2	Molecular Dimensions Ltd.	MD1-01/02
NR-LBD	Molecular Dimensions Ltd.	MD1-24
MIDAS	Molecular Dimensions Ltd.	MD1-59
Crystal screen	Hampton Research Corp.	HR2-110/112

3.8. Lipids

All lipids used for cryo-electron microscopy (cryoEM) reconstitutions were purchased from Avanti® (USA). 1,2 dioleoyl-sn-glycero-3-phosphocoline, (DOPC, ref. 850375), 1,2 dioleoyl-sn-glycero-3-phosphoethanolamine, (DOPE, ref. 850725), 1,2 dioleoyl-sn-glycero-3-phosphoserine, (DOPS, ref. 840035), ovine cholesterol (Chol, ref 700000) and C24:1 Galactosyl (β) Ceramide (d 18:1/24:1(15Z) (GalCer, ref. 860546) were purchased as powder and dissolved in chloroform at 1 mg/ml (0.2 mg/ ml for Liss Rhod PE). 1,2 dioleoyl-sn-glycero-3-phospho-(1′-myo-inositol-3′-phosphate) (ammonium salt), (PI(3)P, ref. 850150) was dissolved at 0.2 mg/ml in a mixture of chloroform, methanol and Milli-Q water (20:9:1). Glassware was used for chloroform-containing mixtures. We used Hamilton pipettes (Sigma, USA) and Durham tubes (Fisher Scientific, USA).

For PX-domain binding tests, we used phosphoinositides purchased from Echelon Biosciences. C16:0 PI (ref. P-0016), C16:0 PI(3)P (ref. P-3016), C16:0 PI(4)P (ref. P-4016), C16:0 PI(5)P (ref. 5016), C16:0 PI(3,4)P (ref. P-3416), C16:0 PI(3,5)P (ref. P-3516), C16:0 PI(4,5)P (ref. P-4516) and C16:0 PI(3,4,5)P (ref. P-3916) were dissolved at 0.2 mg/ml in a mixture of chloroform, methanol and Milli-Q water (20:9:1). All lipids were kept in 2-ml silanised Amber screw-top vials (SUPELCO, ref. 27238), with caps covered with PTFE liner, from Sigma (ref. 27091)

Large unilamellar vesicles (LUVs) were produced using a Mini-Extruder from Avanti® (ref. 610000). For extrusion, filter supports (ref. 610014) and polycarbonate membranes (pore size 0.2 µm, ref. 800281 and 0.4 µm, ref. 800282) from Whatman were used.

4. METHODS

4.1. Cloning methods

Milli-Q water used for cloning was sterilised at 134 ºC for 20 minutes to make sure it was free of nucleases. All DNA was dissolved or eluted in this water.

4.1.1. Amplification of DNA by PCR

DNA amplification was accomplished by the polymerase chain reaction, or PCR, using Phusion High-Fidelity PCR Kit (Thermo Scientific, USA). Template DNA and oligonucleotides used are described in the Materials section. Oligonucleotides were manually designed (see Table 3) and optimised for melting temperature using OligoCalc (http://biotools.nubic.northwestern.edu/OligoCalc.html). All oligonucleotides were designed to have a melting temperature ≥ 62 ºC, and most annealing temperatures were set in this range. A web tool (https://www.bioinformatics.org/sms/rev_comp.html) was used to obtain the reverse complementary sequences. The reaction mixture and conditions of PCR are detailed in Table 5.

Table 5: PCR reaction mixture for 50-µl reaction (A) and PCR conditions (B)

A

Reactive	Volume (µl)
Milli-Q water	37.5 µl
5x Phusion HF buffer	10 µl
dNTPs	200 µM (1 µl of 10 mM)
Oligonucleotides	2.5 µl each (0.5 µM each)
DNA template (1 ng/µl)	1 µl
Phusion HF DNA polymerase	0.5 µl

B

Stage	Temperature	Time (s)	
Initial denaturation	98 ºC	30	1
Cycled denaturation	98 ºC	10	30
Annealing	60-65 ºC	30	
Extension	72 ºC	30 s/kb	
Final extension	72 ºC	600	1

4.1.2. Agarose gel electrophoresis

Separation of DNA fragments was carried out using agarose gel electrophoresis. Agarose D1 Low EEO (Pronadisa) was dissolved at 1–2% (w/v) in TAE buffer (20 mM acetic acid, 1 mM EDTA, 40 mM Tris, pH 8) and boiled in a microwave until the agarose was completely dissolved. When the solution cooled to around 60 ºC, SYBR® Safe stain (Invitrogen) was added at 1:10000 (v/v), and the gel was polymerised in a Mini-Sub Cell GT gel caster (Bio-Rad). Samples were mixed with 6× DNA loading

buffer (Thermo Scientific, USA); 5 µl of GeneRuler 1 kb DNA Ladder (Thermo Scientific, USA) supplied the molecular weight markers. Gels were run in a Mini-Sub Cell GT system (Bio-Rad, USA) at 100 V for 45–60 minutes. DNA was visualised using an ImageQuant™ LAS 4000 imager (GE Healthcare, USA).

4.1.3. Restriction endonuclease digestion, dephosphorylation and ligation

PCR products (45 µl) were digested with the appropriate restriction enzymes (AgeI and NotI, from NEB, for the only clone produced by digestion). In this case, 5 µl of NEB buffer 2.1 10× was added to 44 µl of PCR product or 44 µl of the closed vector (pET HisSUMO-GFP), and 0.5 µl of each enzyme was added. The digestion mixture was left for 16 hours at 37 ºC. After digestion, the vector (but not insert) ends were dephosphorylated using NEB Antarctic Phosphatase, for 2 hours at 37 ºC. The enzymes were inactivated at 80 ºC for 5 minutes. The products were evaluated by agarose gel electrophoresis to check the efficiency of the digestion reaction. Then, they were purified using the Wizard® SV gel and PCR Clean-Up System (Promega). DNA was quantified by measuring absorbance at 260 nm in a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, USA). The quality of DNA was calculated using the ratio of absorbances at 260 and 280 nm. Fifty nanograms of vector DNA was mixed with a three-fold amount of insert (molar ratio). The ligation reaction was conducted overnight at room temperature (RT) with T4 DNA ligase (NEB, USA) in the appropriate ligase buffer.

4.1.4. Gibson assembly

In general, cloning was performed using Gibson isothermal assembly, described by Daniel G. Gibson[148]. Briefly, the DNA of interest was amplified by PCR, as was the vector into which it was going to be introduced. Both PCR products (vector and insert) contained overlapping regions on their flanks. They were digested with DpnI (Thermo Fisher Scientific, USA), a restriction enzyme that can digest methylated DNA, avoiding false-positive colonies caused by the presence of template DNAs. PCR products were purified using the Promega Clean-Up System, and they were incubated with the so-called "isothermal mix" (Table 6A). An aliquot of 50–100 ng of vector DNA was mixed with 3–5 molar excess of insert DNA in a volume of 5 µl. The DNAs were then combined with 15 µl of the isothermal mix (see Table 6). The ratios between insert and

vector concentrations varied depending on the size of the insert; they were increased when the insert was smaller than 500 bp. The mixture was incubated for one hour at 50 ºC. The reaction allowed the introduction of the insert at the amplified region of the vector (as long as both sequences contained overlapping regions).

Protein mutagenesis was also performed using this method. The vector containing the gene to be mutated was amplified in two PCRs producing two overlapping fragments of DNA that contain the mutation. The mutation was introduced in the middle of the primers that amplified the gene (Figure 12). First, the primer FW (1) containing the mutation was used to amplify the DNA fragment with primer RV (1) that follows from the antibiotic-resistant gene of the vector (see Figure 12B). Second, the primer RV (2) containing the mutation amplified the gene with primer FW from the antibiotic-resistant gene of the vector (2). Afterwards, both PCR products (Figure 12C), with the overlapping flanking sequences hybridising with each other, were incubated with the "isothermal mix", as explained above. Finally, the product of the reaction was directly transformed into *E. coli*.

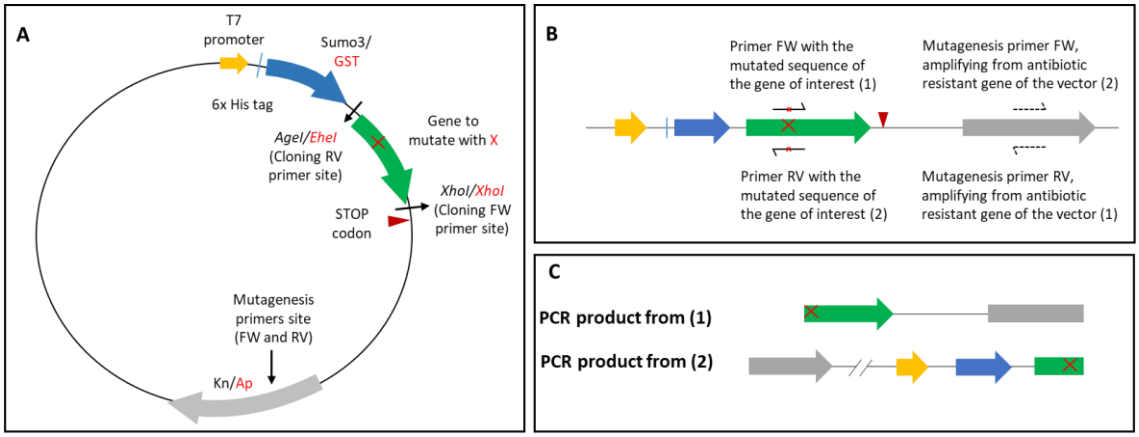


Figure 12: Schematic representation of mutagenesis by Gibson assembly. (A) A circularised vector as pictured in Figure 11. The gene shown in green expresses the protein to be mutated, and the red cross indicates the amino acid to mutate. (B) The linear view of the circularised vector, indicating the primers for amplifying from the gene to be mutated (solid lines) and the primers for amplifying from the antibiotic-resistance marker (dashed lines). The red cross indicate the amino acid to mutate (C) Representation of the PCR product from primers (1), or PCR product (1), and PCR product using primers marked as (2). The mutation (red cross) have been introduced in both PCR fragments.

Table 6: (A) Isothermal reaction mixture. (B) Composition of the isothermal mix.

A

Reactive	Quantity (picomoles)
DNA insert	0.3-0.5 pmoles
DNA vector	0.1 pmoles
+15 µl Isothermal mix	

Incubate 1 hour at 50 ºC

B

Isothermal mix	Isothermal buffer composition
40 µl of isothermal buffer	500 mM Tris-Hcl pH 7.5 25% PEG 8000 50 mM MgCl2 50 mM DTT 5 mM NAD 1mM dNTP (each) Fill up to 6 ml dH_2O
1.5 µl of T5 exonuclease 1 U/µl (NEB)	
2.5 µl of Phusion polymerase (Thermo)	
10 µl of Taq ligase NEB	
96 µl of Milli-Q water	

4.1.5. DNA transformation by heat shock

Isothermal or ligation products were directly transformed into *E. coli* XL1-Blue. A 5-µl aliquot of ligation mixture or Gibson assembly was added to 50 µl of homemade competent cells (made by Ander Vidaurrazaga), and the mixture was incubated on ice for 30 minutes. Cells were then incubated in a heating block at 42 ºC for one minute. Then, the cells were kept on ice for 5 minutes, and 800 µl of LB media was added. This was followed by the incubation for 45–60 minutes at 37 ºC. The cells were centrifuged at 9000 × g for 1 minute at RT, and 600 µl of the supernatant was removed. The resultant pellet was re-suspended in 250 µl of and plated on a Petri dish containing LB agar supplemented with the appropriate antibiotic as a selective marker for the plasmid.

4.1.6. Colony PCR

Colonies obtained after transformation were checked for the presence of the insert. PCR mixture was prepared with one oligonucleotide of the insert and one universal primer that hybridises with the vector (e.g., T7 promoter primer). Ten µl of PCR 2× Master Mix (Thermo) was mixed with 0.5 µl of each primer, 10 µl of Milli-Q water and 1 colony. Then, PCR cycles similar to those in Table 5B were followed, but with the denaturation temperature of 95 ºC, and 1 minute of extension per kb. A negative control, using empty vector, was prepared for each test. PCR products were analysed employing agarose gel electrophoresis as described before. All colonies tested were plated on fresh LB agar supplemented with the appropriate antibiotic. One positive colony was selected and grown for plasmid propagation.

4.1.7. DNA extraction

The positive colonies were grown in 50-ml sterile Falcon tubes containing 10 ml of LB supplemented with the appropriate antibiotic. They were incubated overnight at 37 ºC in an Excella E24 incubator shaker (New Brunswick Scientific). The following day, the cells were harvested by centrifuging at 3900 × g in Allegra X 22-R centrifuge (Beckman Coulter). The plasmid was extracted from the resultant pellet using Wizard® Plus SV Miniprep DNA Purification System (Promega). Quantification of DNA was carried using a NanoDrop spectrophotometer, whereas purity was evaluated by agarose gel electrophoresis.

4.1.8. DNA sequencing

One microgram of DNA, mixed with 3 µl of the oligonucleotide, was sent for sequencing to STAB VIDA (Caparica, Portugal). Universal primers were used (i.e., T7 promoter/terminator). Obtained sequences were aligned with the expected sequence using BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Nucleotides) to confirm the presence of the correct insert.

4.2. Protein characterisation

4.2.1. Physicochemical characterisation

Physicochemical properties of the proteins such as the molecular weight, theoretical pI and extinction coefficient were examined using the ProtParam tool from ExPASy (https://web.expasy.org/protparam).

4.2.2. Protein concentration estimation

Protein concentration was measured using NanoDrop ND-1000 (Thermo Sci.). The molar extinction coefficient was calculated from the amino acid sequence of each protein using the ProtParam tool from ExPASy server[151]. Then, its value was used to obtain protein concentration using the absorbance at 280 nm, according to Lambert-Beer law. The calculation was performed using Equation 1, where ε is the molar extinction coefficient, c is protein concentration, and l is the optical path length in centimetres. Absorption at 280 nm is dominated by aromatic residues, especially tryptophan, but also tyrosine and phenylalanine.

$$Abs_{280\,nm} = \varepsilon \times c \times l$$ **Equation 1**

4.2.3. Denaturing Polyacrylamide gel electrophoresis (SDS-PAGE)

Most of the time, protein quality was assessed by denaturing SDS-PAGE; samples boiled in presence of sodium dodecyl sulphate (SDS), as strong anionic detergent, and dithiothreitol (DTT), as reducing agent, were subjected to electrophoresis. Fifteen-well gels were prepared (8 × 8 cm, 1-cm thick, 12 or 15% of polyacrylamide). Protein samples were prepared by mixing 16 µl of protein solution with 4 µl of Laemmli loading buffer 5× (Bio-Rad) supplemented with 10 mM DTT. Samples were heated at 95 ºC for 5 minutes. Precision Plus Protein Dual Color Standards (Bio-Rad) were used as molecular weight markers. Electrophoresis was carried out in Mini-Protean® Tetra Vertical Electrophoresis Cell (Bio-Rad). Inner and outer chamber was filled with Tris-glycine "running buffer" (192 mM glycine, 0.1% SDS and 25 mM Tris pH 8.3). Gels were run at 150 – 200 V for 1.5 hours. They were stained with Coomassie Brilliant Blue R-250 (0.1% (w/v). Coomassie Brilliant Blue R-250, 45% ethanol, 10% acetic acid and 45% demi-water (v/v)). Gels were stained for 15 to 30 minutes after heating in a microwave (not boiling). For protein visualisation, the gels were washed using a mixture of 40% ethanol, 10% acetic acid, and 50% demi-water (v/v) until background stain disappeared and the protein bands were clearly visible.

4.2.4. Sequence alignment and 3D structure prediction

Secondary structure prediction was carried out employing the Phyre2 web service (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index). Predictions were used to design protein constructs for crystallisation.

Protein sequence alignment was performed employing the EMBL programme Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/). The results were plotted by the ESPript 3.0 tool (https://espript.ibcp.fr/ESPript).

4.3. Heterologous protein expression and purification

All proteins analysed in this project correspond to heterologously expressed human proteins. Heterologous expression of these proteins was performed in *E. coli* BL21

(DE3) and *E. coli* Rosetta (DE3) cells. Briefly, 1% of the pre-inoculum (e.g., 20 ml in 2 litres) grown overnight was added to a 5-litre flask containing 2 litres of fresh LB, supplemented with the appropriate antibiotic. Bacteria were grown until the optical density at 600 nm (OD_{600}) was 0.6. At this point, the temperature of incubators was reduced to 20 ºC, and 30 minutes later, 0.5 mM of isopropyl-B-D-thiogalactopyranoside (IPTG) was added to the culture with a final concentration of 0.5 mM. The incubation continued overnight at 20 ºC. The following day, OD_{600} of the cultures was measured. The cells were harvested in JLA 8.1000 rotor, in an Avanti® J-26 XP centrifuge, at 5000 × g for 20 minutes at 18 ºC. The pellet was directly lysed for protein purification, or frozen in liquid nitrogen and stored at -80 ºC.

The CIMPR constructs were expressed in *E. coli* Rosetta (DE3) cells (Table 2) in auto-induction medium (see Materials). The medium for a 2-litre batch was prepared as follows. First, ZYP medium was prepared by mixing 20 g of tryptone and 10 g of yeast extract, dissolved in 1.718 litres of distilled water in a 5-litre flask. Second, 10× NPS was prepared by mixing 6.6 g of $(NH4)_2SO_4$, 13.6 g of KH_2PO_4 and 8.9 g of Na_2HPO_4, and dissolving the mixture in 200 ml of Milli-Q water. The two solutions were sterilised in a MATACHANA autoclave. Then, another two solutions were prepared and passed through a 0.2-µm filter; 2 ml of 1M $MgSO_4$ heptahydrate and 80 ml of a 25× 5052 solution mix (containing 12.5% glycerol, 1.25% glucose and 5% α-lactose (all percentages in w/v)). Finally, the $MgSO_4$ solution was added to the 5-litre flask containing ZYP medium, followed by the remaining two solutions (10× NPS and 25× 5052).

General purification strategy is shown in Figure 13. All steps were performed at 4 ºC, and the protein sample was kept on ice as much as possible to avoid protease degradation. The process usually followed the same scheme: affinity tag (histidine or GST tag) purification in batch in a cold room, ion-exchange chromatography using HiTrap 5-ml columns (GE Healthcare) and finally a size-exclusion chromatography. The size-exclusion chromatography was conducted using Superdex™ HiLoad 200 (for proteins > 50 kDa) or Superdex™ HiLoad 75 (for proteins < 50 kDa), 16/60 or 10/300, in an ÄKTA pure system (GE Healthcare) in a refrigerated cabinet. The details of purification using different tags are described in the next chapters.

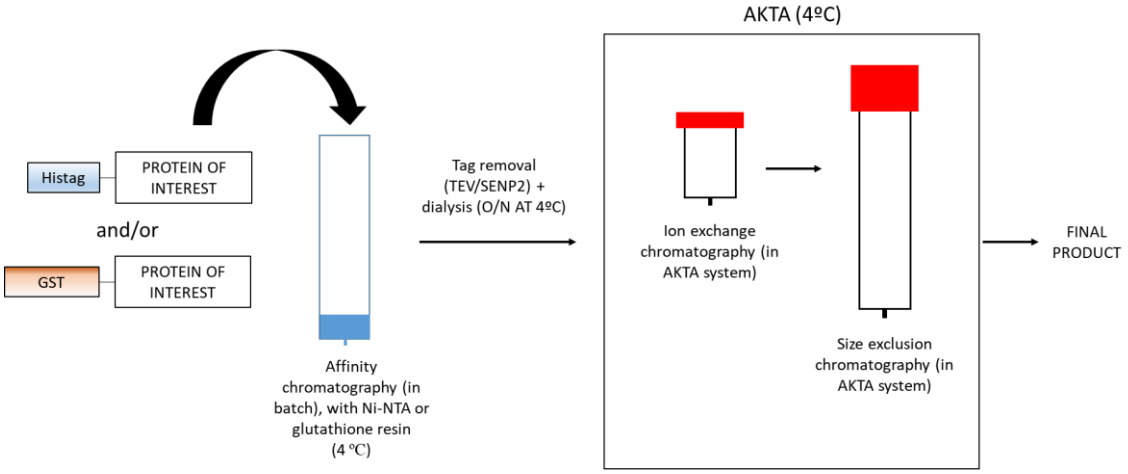


Figure 13: General strategy followed for all the purifications.

After removing the tag, the process was the same in all cases. The protein solution was added to a buffer without NaCl to reduce the ionic strength. Depending on the calculated isoelectric point, anion (Q) or cation (SP) exchange chromatography was carried out. Salt concentration was empirically adjusted for each protein; 90 mM NaCl was sufficient for most proteins to be adsorbed to the resin. Protein was eluted with 20 column volumes of NaCl gradient, from the initial NaCl concentration to 1 M NaCl. SDS-PAGE was run, and the purest fractions were concentrated and loaded onto an S200 (proteins > 50 kDa) or S75 (< 50 kDa) 16/60 column. Size-exclusion columns were run in 300 mM NaCl, 1 mM DTT, 25 mM Tris pH 7.5. In cases of poor peak resolution (protein did not separate from degradation products or aggregates), another size-exclusion procedure was performed (e.g SNX5). After chromatographic isolation, the fractions were analysed by SDS-PAGE, and the purest samples were pooled together and concentrated. If not used directly, the purified proteins were frozen in liquid nitrogen and kept at -80 ºC until further use.

4.3.1. Purification of proteins with histidine tags

Proteins tagged with His-SUMO or His-MBP (see Table 2) were purified using Protino® Ni-NTA (nickel–nitriloacetic acid) agarose, from Macherey-Nagel. The bacterial pellet was suspended in precooled buffer A (300 mM NaCl, 20 mM Imidazole, 1 mM DTT, Tris-HCl 50 mM pH 7.5) supplemented with 5 mM benzamidine and 0.5 mM phenylmethylsulfonyl fluoride (PMSF). The volume of buffer to be added was

calculated using the formula in Equation 2. The pellet was stirred at 4 ºC until it formed a uniform suspension. The cells were lysed in a SONICS Vibra-Cell sonicator with a 13-mm titanium tip, in cycles of 10 seconds ON and 1 minute OFF, for 5 minutes. Afterwards, the lysate was clarified in JA 25.50 rotor, in an Avanti® J-26 XP centrifuge, at 60000 × g for 45 minutes at 4 ºC. Meanwhile, 10 ml of Ni-NTA resin was prepared by washing with 50 ml of distilled water and then equilibrated with 50 ml of buffer A. The supernatant was incubated on a roller with pre-equilibrated Ni-NTA resin in 250 ml bottles for 1 hour at 4 ºC. Then, the mixture of supernatant and resin was loaded onto a DWK Konte™ 50-ml column (Fisher Scientific, USA), and the supernatant was discarded. The resin was washed with 30 column volumes of buffer A, and the protein was eluted with 30 ml of buffer B (300 mM NaCl, 250 mM imidazole, 1 mM DTT, 25 mM Tris-HCl pH 7.5). The protein was dialysed overnight at 4 ºC in buffer C (150 mM NaCl, 1 mM DTT, 25 mM Tris-HCl pH 8), in a 10 kDa cut-off SnakeSkin™ dialysis bag, with gentle stirring. One mg of TEV protease (for His-MBP) or 1 mg of SENP2 protease (for His-SUMO3-tagged proteins) was added to the dialysis bag for tag removal. On the next day, the protein was subjected to ion-exchange chromatography, as described before.

$$Volume\ (ml) = OD_{600} \times litre\ of\ culture\ \times 6$$ **Equation 2**

4.3.2. Purification of proteins with GST tags

Purification protocol of GST-tagged proteins was essentially the same as described for histidine-tagged proteins (previous section), with minor modifications. The bacterial pellet was re-suspended in buffer D (300 mM NaCl, 1 mM DTT, 50 mM Tris-HCl pH 7.5) supplemented with 5 mM benzamidine, 0.5 mM PMSF and 1 mM EDTA. Incubation with Glutathione Sepharose™ 4B resin (GE Healthcare) was carried out for 2 hours instead of 1 hour. After washing the resin with 300 ml of buffer D, the GST tag was removed. The tag can be removed by dialysis after elution with buffer E (150 mM NaCl, 1 mM DTT, 50 mM reduced glutathione, 25 mM Tris-HCl pH 8). It can also be cut in-resin, re-suspending GST resin (with the protein attached to it) in 30 ml of buffer C, adding 1 mg TEV protease, and incubating overnight at 4 ºC with gentle rotation. Once the GST tag was removed, the protein was subjected to ion-exchange chromatography, as described before.

4.3.3. Purification of protein complexes with double tag

SNX1/SNX5 heterodimer was purified using a strategy ensuring that all the obtained protein was indeed a heterodimer of SNX1 and SNX5, not contaminated with homodimers or monomers. To achieve a suitable purification, obtaining the two proteins in 1:1 ratio, a double-tag approach was followed. SNX1 was cloned into pET HisSUMO (kanamycin^R) vector and SNX5 into pGST-P2 (ampicillin^R) vector. After co-expressing the proteins (in BL21 (DE3) strain, in LB with both antibiotics), the cells were harvested as described previously. Then, the pellet was dissolved in buffer D, and the GST tag was cut in-resin overnight at 4 ºC. On the following day, the flow-through containing His-SUMO-SNX1 and SNX5 was collected and incubated with 10 ml of Ni-NTA resin pre-equilibrated with buffer A. Using this approach, all the SNX5 not forming heterodimers was removed. All the His-SUMO-SNX1 not forming heterodimers with GST-SNX5 was rinsed out during the washing of glutathione resin. Then, the purification continued as for histidine-tagged proteins.

4.4. Biochemical characterisation of proteins

4.4.1. Circular Dichroism

Circular dichroism (CD) is a useful technique for protein secondary structure analysis in solution[152]. It is based on the differences between the absorption of left- and right-handed polarised light. Proteins are chiral compounds; the organisation of peptide bonds in the molecule gives rise to a specific CD spectrum. This can be correlated with the proportion of α-helices, β-sheets and random coil structures in the analysed protein (Figure 14A).

The structural comparison of mutant and wild-type (WT) proteins was performed using CD analysis. Proteins were dialysed overnight at 4 ºC in phosphate-buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4 and 1.8 mM KH_2PO_4) without reducing agents. Next, the proteins were analysed using a Jasco 710/810 Spectropolarimeter. The data were obtained at 25 ºC, measuring between 200 – 250 nm, with a quartz cuvette of 0.1-cm path length. PBS was used as a blank. Proteins were measured at 500–1000 nM concentration. Data were blank-subtracted (using PBS). Then, ellipticity per amino acid was obtained, using Equation 3. The data were plotted using GraphPad.

$$\frac{ellipticity}{aminoacid} = optical\ path\ (cm)\ x\ ellipticity\ (mdeg)\ x\ Molarity\ (M)\ x\ (nº\ aa - 1)$$

Equation 3

4.4.2. Pull-down assays

Pull-down assays allows to examine the ability of two proteins to interact. In our case, the pull-down procedure was performed with purified protein, indicating a direct interaction. It was used as a competition assay, where SNX5 could displace SNX1 homodimers. It also allowed us to monitor SNX1/SNX5 heterodimer breakup.

The proteins were dialysed against "pull-down buffer" (150 mM NaCl, 10 mM imidazole, 0.5 mM TCEP and 10 mM HEPES pH 7.5) overnight at 4 ºC. Ni-NTA resin was washed with Milli-Q water and then pre-equilibrated in the same buffer. Next, 50 µl of 5 µM His-SUMO-SNX1 FL and mutants (or His-SUMO SNX27 WT as negative control), and 50 µl of 10 µM SNX5 FL WT and mutants were incubated with 20 µl of Ni-NTA resin slurry for 1 hour at 4 ºC. Before the addition of Ni-NTA resin, 20 µl of the sample was taken as input and mixed with 20 µl of Laemmli buffer 2×, boiled and frozen. After incubation of proteins with the resin, the sample was centrifuged at 1000 × g for 5 minutes at 4 ºC. The supernatant was gently removed, taking care to avoid disturbing the resin at the bottom of the vial. The sample was washed three times, with 500 µl of pull-down buffer. After the last wash, the resin was mixed with 30 µl of Laemmli buffer 2×, boiled, and examined by SDS-PAGE.

4.4.3. Multi-angle light scattering (MALS)

The absolute molar mass of proteins in solution was obtained using MALS coupled to size-exclusion chromatography (SEC-MALS)[153]. In this assay, the protein solution is irradiated with collimated and polarised light, normally a laser (see Figure 14B). The light scattering of the beam is monitored continuously, generating a chromatogram, which shows the different molecules (or different oligomeric states) in the sample. The sample causes the light to scatter into different angles, which allows the determination of molecular weight, while the diffraction index indicates the radius of the molecule.

MALS analysis requires a monodisperse protein sample. Hence, the size-exclusion chromatography (S200 10/300, GE Healthcare) was conducted before SEC-MALS. All

proteins were run in buffer F (150 mM NaCl, 0.5 mM TCEP, 10 mM HEPES pH 7.5). Afterwards, 100 µl of the protein at 1 mg/ml was analysed on an HPLC (Shimazu) coupled with an S200 Increase 10/300 GL column (GE Healthcare). The process was monitored using a DAWN® HELEOS® II and an Optilab™ rEX detectors (Wyatt Technology), measuring the multi-angle light scattering (MALS), and refractive index of the sample, respectively. The results were analysed using ASTRA 6 software (Wyatt Technology). Bovine Serum Albumin (≥ 98% purity; Sigma), 1 mg/ml, was prepared in the same buffer, filtered through an Ultrafree® centrifugal filter of 0.1-µm cut-off (Millipore®) and used as the standard. The results were plotted using GraphPad.

4.4.4. Isothermal titration calorimetry (ITC)

ITC allows the measurement of thermodynamic parameters of interaction in solution[154]. The technique is based on an adiabatic system, where the heat released or absorbed by a reaction in a cell containing the sample (sample cell) is compensated by a reference cell (Figure 14C). The emitted (i.e., endothermic) or released (i.e., exothermic) heat is converted into an electric signal (voltage necessary to keep the reference cell at the same temperature). The electric signal is integrated against time. As concentrations (and hence, the ratios) of the protein and the ligand are known, an affinity curve can be obtained, as well as thermodynamic parameters: enthalpy (ΔH), entropy (ΔS), dissociation constant (K_d) and stoichiometry of the reaction (n). Free energy (ΔG) can be calculated from these data using Equation 4.

$$\Delta G = \Delta H - T\Delta S \qquad \textbf{Equation 4}$$

In this study, the ITC measurements were performed at 25 ºC using a MicroCal PEAQ-ITC titration calorimeter (Malvern Panalytical). Samples were dialysed overnight at 4 ºC against ITC buffer (300 mM NaCl, 0.5 mM TCEP and 25 mM HEPES pH 7.5). Afterwards, the proteins were tempered at 25 ºC and degassed using a ThermoVac. An initial injection of 0.4 µl (not used for the curve calculation) was followed by 18 injections of 2 µL, with a spacing of 150 seconds between injections, allowing the system to go back to equilibrium. The data were analysed employing the MicroCal PEAQ-ITC software, from Malvern Panalytical, using "one set of sites" as the binding model. Final graphs were prepared using Origin ITC (MicroCal) and GraphPad.

The binding analysis was conducted by titrating CIMPR over the SNX1/SNX5 heterodimer or the different subunits (see Results section). Both the CIMPR peptide and $CIMPR_{2330-2491}$ were used for titration, with negligible differences. CIMPR (300 µM) was titrated over 15 µM SNX1/SNX5. To find the sites of interaction with heterodimers, 20 µM CIMPR was titrated over PX domains of 450 µM SNX1 and SNX5, and 800 µM CIMPR peptide was titrated over BAR domains of SNX1/SNX5. All the measurements were performed at least three times.

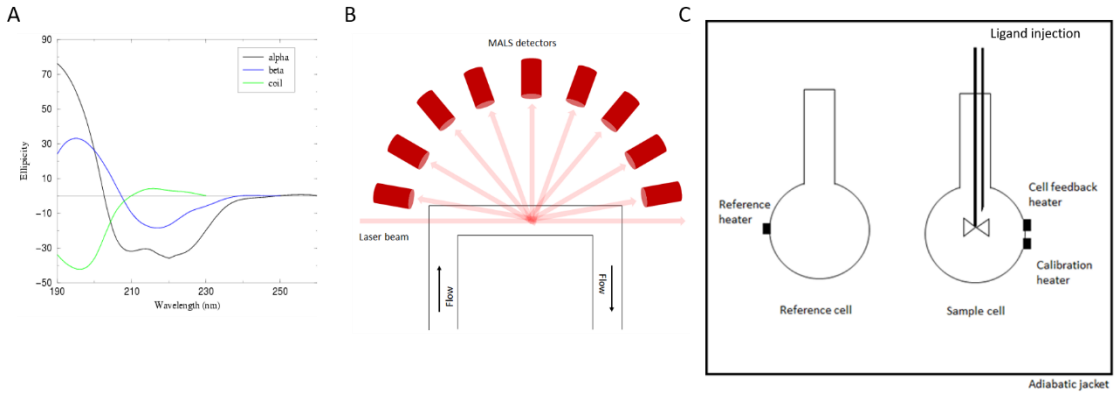


Figure 14: Schematic representation of biochemical techniques used in this study. (A) Circular dichroism spectra for α-helices (black line), β-sheets (blue line) or random coil (green line). Image obtained from www-isbg.fr. (B) Schematic representation of MALS illumination. Image adapted from Malvern Panalytical (https://www.materials-talks.com/blog/2014/08/12/multi-angle-light-scattering-mals/). (C) Schematic representation of an ITC instrument.

4.5. Structural analysis using X-ray crystallography

To solve the structure of a protein using x-ray crystallography first, a crystal of the protein of interest needs to be obtained. A crystal is a three-dimensional arrangement of millions of repetitions of the same molecule, characterized by the periodicity of a fundamental unit, which is known as an asymmetric unit. In a protein crystal, the asymmetric unit can contain one or a few molecules of the protein. A set of asymmetric units can be related by symmetry operations (translations and rotations). All symmetry operations of a crystal conform the space group, when these operators are applied to the asymmetric unit the minimum volume that is generated is called unit cell.

In general, the dimension of the unit cell for a protein crystal varies from tens to hundreds of angstroms, which is in the same order of magnitude of the wavelength of

the x-ray. For this reason, when the x-ray passes through a crystal produces a diffraction pattern. This pattern can be explained considering the x-ray beam as a set of waves of electromagnetic radiation that after interacting with a crystal produce a new set of diffracted waves that cancel one another in most directions through destructive interference, but add constructively in a few specific directions, determined by Bragg's law:

$$2d \sin\theta = n\lambda$$

Where θ is the incident angle, d is the spacing between diffracting planes, n is any integer, and λ is the wavelength of the beam. A detector could be placed after the crystal in the specific directions where constructive interference happens and some spots could be recorded, these spots are called reflexions. The intensity and relative position of these reflexions depend on the position and type of atoms of the molecule that conform the crystal. Then, the diffraction pattern can be used to determine the position of the atoms within a crystal.

Each reflexion in the diffraction pattern may be interpreted as the resulting wave of the x-rays scattered along one direction. The mathematical representation of this wave is called the structure factor and has a magnitude and relative phase. Experimentally is only possible to measure the amplitude of the structure factors since it is directly proportional to the intensities of the spots in the diffraction pattern, but the phase information is lost: this is the 'phase problem' of X-ray crystallography.

There are several methods available to obtain the phase and therefore solve the 3D structure of a molecule. Molecular replacement is the easiest approach. It needs a solved structure that is similar to the studied one, so the phase of this known structure can be used. It is based on positioning the 3D model of the known protein in the crystal packing of the studied protein. Once the studied protein has been oriented, phases can be calculated.

When such a model is not available, or it is not sufficiently homologous, phases can be obtained using isomorphous replacement. This method is based on the introduction of heavy atoms (atoms with a large number of electrons) into the crystal. The analysed crystals can be soaked in a solution containing the selected heavy atom to obtain derivative crystals. Then, when diffraction patterns of native and derivative crystals are compared, differences caused by heavy atoms are used to calculate their

relative positions and their phases. This information allows calculating the phases of the crystal derivate. However, the native and derivative crystals must be isomorphous, i.e., the unit cell and the space group should be the same. When only one derivative is produced, the method is called single isomorphous replacement (SIR) when more than one is used the technique is called multiple isomorphous replacement (MIR). To solve the structure of a protein using this method at least two derivatives plus the native are needed.

The phase can also be obtained by introducing anomalous scatter that produces inelastic scattering of the radiation. When the X-ray photon energy is sufficient to promote an electron from an inner shell, it produces an anomalous scattering. Normally $\lceil F(h,k,l) \rceil$ is equal to $\lceil F(-h,-k,-l) \rceil$ (Friedel's law). When anomalous scattering is produced within a crystal, this leads to a breakdown in Friedel's law, giving rise to anomalous differences ($\lceil F(h,k,l) \rceil \neq \lceil F(-h,-k,-l) \rceil$) that can be used to locate the anomalous scatterers, and its phases can be calculated. This information allows calculating the phases of the crystal derivate. If the diffraction is collected at the absorption edge of the anomalous scatter the method is called single wavelength anomalous dispersion (SAD). Also, it is possible to use the difference between the intensities of the native crystal and the anomalous derivative, when this information is combined the technique is called single isomorphous replacement with anomalous scattering (SIRAS)

The changes in structure-factor amplitudes arising from anomalous scattering are generally small and require accurate measurement of intensities. Non-isomorphism between the native and the derivative crystals could prevent phase determination; to overcomes this problem multiwavelength anomalous diffraction (MAD) method is used. In this case, three data sets are collected for the same crystal; at the peak of absorption, at the inflection point of the absorbance spectra, and a remote wavelength from the absorption peak.

We attempted to utilise the phases of other SNX-BAR domains solved (SNX9, 2RAI; and SNX33, 4AKV). However, the solution was not found. Next, we tried to add heavy atoms or anomalous scatters with the excitation wavelength in the range of X-rays used in the synchrotron facilities that we have access and, proceed with SAD phasing. The habitual approach is to incorporate selenomethionine into the proteins[155]. However, although it was tested for our crystals, this approach did not work (see

Results section). Hence, we used the crystals soaked in solutions containing heavy-atom compounds. First, analysis of the most appropriate heavy atoms for our crystallization conditions (pH 7) were sought using the Heavy Atom Databank (HAD)[156] (http://www.sbg.bio.ic.ac.uk/had/heavyatom.html). Then, several heavy atoms were tested (see Table 7). Soaking crystals in 10 mM K_2PtBr_4 produced anomalous diffraction without breaking the crystalline network; three data sets were collected using the same crystal (peak, inflection and remote), this allowed us to solve the SNX1$^{301-522}$/SNX5$^{195-404}$ heterodimer using the multiple anomalous diffraction (MAD , see below, section 4.5.4).

4.5.1. Crystallization of proteins

Vapor diffusion is the most common method to crystallize proteins. Crystallization is carried out by mixing a solution containing the pure protein with a precipitant solution. This drop is placed in a coverslip over a reservoir solution, which contains the undiluted precipitant solution (see Figure 15). While the equilibrium between the drop and the reservoir is reached the protein becomes supersaturated, allowing nucleation and facilitating crystal growth.

This technique can be performed in two ways; sitting drop and hanging-drop. For the initial high-throughput screening, the sitting drop method was used (see Figure 15A). 50 µl of the precipitant were dispensed into the reservoir of the 96-well plates (MRC 2, Molecular dimensions) using the TECAN Freedom EVO (Tecan Group Ltd.) liquid handling system. Subsequently, the protein and the precipitant were dispensed using the MOSQUITO robot (TTP Labtech). Then, the plates were sealed using Clear Vue ™ sheets (Molecular Dimensions Ltd.) and stored at 18 ºC for 3–5 days. The plates were examined under an MZ12.5 stereomicroscope (Leica).

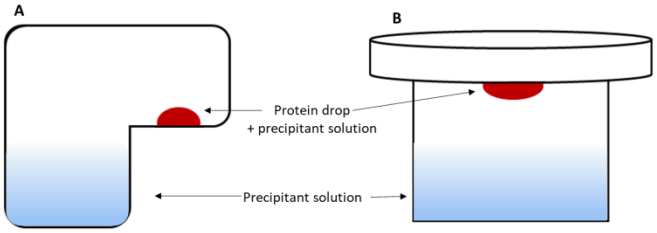


Figure 15: Sitting-drop method (A) and hanging-drop method (B) for crystallization.

Once the crystals were formed in a plate, the conditions were optimized to produce bigger crystals. Here, the hanging-drop method was employed (see Figure 15B). EasyXtal 15-well plates were used (Qiagen). Protein concentration, pH, and concentration of precipitants were optimized. For the crystallization of SNX1/SNX5 BAR domains, the protein at 8 mg/ml was mixed with the same volume of precipitant (11% PVA, 10% 1-propanol, 100 mM HEPES pH 7.0). These conditions were obtained from MIDAS screening and modified to obtained better crystals.

4.5.2. Seeding methods

Protein crystallisation can be improved by adding nucleation seeds from other crystals to increase the size of the crystals[157]. Two different method types are available: micro-seeding and macro-seeding. To use micro-seeding, crystals are taken from the initial incubation, broken in small pieces in the precipitant, and added to a fresh crystallization drop. Alternatively, a small crystal can be washed in the crystallization solution without protein and added to the new crystallization condition with fresh protein, this approach is call macro-seeding. The transfer of the seeds can be done by pipetting from the stock or by streak seeding that consist in using a strong filament, like a cat whisker, as a seeding wand, to touch the crystals. Then, the filament is passed once through a new crystallization drop, leaving the seeds that will help with nucleation.

In this study, micro-seeding was used and the seeds were transfer using streak-seeding, this method rendered bigger and better crystals that diffracted enough to be used for protein structure determination.

4.5.3. Data collection and X-ray diffraction data

Data collection was carried out using the XALOC beamline at ALBA synchrotron (Cerdanyola del Vallès, Spain), and different beamlines at the Diamond Light Source (DLS, IO3, IO4, I24). The parameters of the data collection (the starting angle, angular interval, number of images and the detector distance) were established after testing the diffraction of crystals at two or four different angles (0, 90º, 180, 270). Then, the EDNA[158] programme was employed to evaluate the diffraction patterns, and the parameters for the data collection were set.

4.5.4. Structure determination

Two hundred and one crystals were tested before finding a data set good enough to obtain the structure of the $SNX1^{301-522}/SNX5^{195-404}$ heterodimer. Native crystals diffracted to 2.5 Å resolution. Data reduction was carried out using XDS[159] and molecular replacement using Molrep (CCP4 software suite[160]). Previously known structures from other SNX-BAR domains (SNX9, 2RAI; and SNX33, 4AKV) were used as a search model in MR approaches. However, no solution was found. Then, we purified and crystalized the selenomethionine derivative of $SNX1^{BAR}/SNX5^{BAR}$. However, the crystals diffracted at 3.2 Å and suffered radiation damage, so the phasing power was too weak to solve the structure using SAD or SIRAS approaches. Then, we attempted to obtain heavy-atom derivatives by quick-soak methods. Table 7 shows the heavy-atom compounds assessed. Tests of their binding to the protein crystals (ALBA or DLS) revealed anomalous diffraction produced using 10 mM K_2PtBr_4. Good data sets were obtained employing this technique; however, this was not sufficient for solving the crystallographic structure using SAD. Nevertheless, big crystals were grown using several micro-seeding steps allowing us to collected three data sets from the same crystal; at the peak, inflexion, and remote wavelengths of the L-III of Pt (1.0715, 1.07293 and 1.0447 Å, respectively), then we solved the structure using MAD.

The detector was placed at a distance best suited for the collection of the data at 2.5 Å. It started at 159º, with a rotation of 0.15º per image. Hence, 2400 images were taken. The exposure time was set to 0.1 seconds per image. Data was collected using the Pilatus 6M system, at ALBA synchrotron.

Data processing was performed using the XDS programme[159] (see Table 10 for data collection statistics). We tried different phasing programs (SHELX, and AutoSol), but it was the AutoSHARP[161] that allowed us to obtain the phase and build the initial model. Next, iterative refinement with PHENIX[162] and manual building in COOT[163] yielded a final model with two heterodimers in the asymmetric unit. Data processing, structure determination, refinement and data deposition were done with the help of Dr. Adriana Rojas.

Table 7: Heavy-atom compounds tested for anomalous diffraction of the crystals. In bold K_2PtBr_4 that bound to the heterodimer and provided enough anomalous signal to solve the structure.

Heavy-atom compound	Concentration	Source
$PtCl_2(H_2NCH_2CH_2CH_2NH_2)$	1, 10, 100 mM	Hampton Research Corp.
$(NO_2)_2Pt(NH_3)_2$	1, 10, 100 mM	Hampton Research Corp.
K_2PtI_6	1, 10, 100 mM	Hampton Research Corp.
$(NH_4)_2PtCl_2$	1, 10, 100 mM	Hampton Research Corp.
K_2PtCl_4	1, 10, 100 mM	Hampton Research Corp.
K_2PtBr_4	1, **10**, 100 mM	Hampton Research Corp.
I3C	1, 10, 100 mM	Molecular Dimensions Ltd.

4.5.5. Structure validation

The experimental result from macromolecular crystallography is a electron density map used to model the sequence of the protein that has been crystallized. The interpretation of the density map might introduce some errors during model building. To prevent potential misinterpretations, different programs implement an interactive process of validation of some parameters, that should be accompanied by visual inspection during model building and refinement. The Rfree[164] is generally considered the most useful global measure of the agreement of the model with the experimental data (structure factors), this parameter was calculated in PHENIX during the refinement process. In addition, the stereochemistry of the model should be monitored as well. In this work, the Ramachandran plot calculated by MOLPROBITY[157], as implemented in PHENIX, was used to evaluate the overall stereochemistry of the model. MOLPROBITY was also used to check potential steric clashes. Ideal values for bond lengths, bond angles, and dihedral angles were evaluated during iterative model building and refinement. In addition, during the deposition process in the Protein Data Bank, there is a validation report that provides an assessment of the structure quality.

4.6. Production of large unilamellar vesicles (LUVs)

The LUVs were made using the methods reported elsewhere[123, 83]. Briefly, the lipids were dissolved as described in the Materials section. They were mixed in specific proportions, detailed in Table 8. The mixtures were pipetted into Durham glass test tubes, dried under nitrogen flow for 5–10 minutes, and then dried again in a vacuum chamber for 1 hour. Afterwards, the appropriate buffer (see below) was added to induce multilamellar vesicle (MLV) hydration. "cryoEM buffer" was composed of 100 mM NaCl, 0.5 mM TCEP, 10 mM HEPES pH 7.5. Coflotation experiment was performed in buffer F (150 mM NaCl, 0.5 mM TCEP, 10 mM HEPES pH 7.5).

MLVs were left hydrating for at least 30 minutes, with occasional vortexing. After hydration, they were transferred to a 1.5-ml Eppendorf tube, and frozen and thawed 10 times in liquid nitrogen and a water bath at 37 ºC, respectively. The sample was vortexed between freeze-thaw cycles. Finally, the MLVs were extruded 11 times through a 0.4-µm polycarbonate filter for cryoEM reconstitution, or 0.2-µm polycarbonate filter for coflotation analysis. Liposomes were stored under argon atmosphere for up to three days, in the fridge.

Table 8: Lipid compositions used during this work.

	DOPC	**DOPE**	**DOPS**	**PIP**	**Rh-PE**
mol% for cryoEM reconstitution	45	30	20	5	-
mol% for coflotation experiments	45	28	20	5	2

4.6.1. Reconstitution of artificial membranes (LUV) with SNX-BAR proteins

Reconstitution of artificial membranes *in vitro* was the approach selected to understand how SNX-BAR proteins deform the membranes and help/trigger cargo transport. A big screening of samples was examined, employing cryoEM to find the best conditions. We tested lipid nanotubes induced by galactosylceramide (GalCer)[165] and evaluated tubulation of liposomes of different sizes (0.05 µm, 0.1 µm, 0.2, µm and 0.4 µm). We tried different lipid compositions (e.g., with and without cholesterol), different incubation times and conditions, different buffers and different proteins (SNX1, and SNX1/SNX5). The conditions are briefly described in Table 9, focusing on what has worked so far.

Table 9: Summary of samples screened by cryoEM.

DOPC (%mol)	DOPE (%mol)	DOPS (%mol)	PI3P (%mol)	GalCer (%mol)	Other lipids (%mol)	Proteins	Result
-	-	25	5	70	-	SNX1 300 nM	Insufficient protein
-	-	25	5	70	-	SNX1 1.5 µM	Insufficient protein
-	-	25	5	70	-	SNX1 3 µM	Good protein coating, many nanotubes empty
-	-	35	5	60	-	SNX1 3 µM	Many nanotubes empty: too rigid
-	-	45	5	50	-	SNX1 3 µM	Many nanotubes empty: too rigid
-	-	55	5	40	-	SNX1 3 µM	Decreased rigidity, good coating
-	-	65	5	30	-	SNX1 3 µM	Nanotubes not very homogeneous
30	18	12	5	30	-	SNX1 3 µM	Nanotubes very heterogeneous: GalCer levels too low
35	21	14	5	20	-	SNX1 3 µM	Nanotubes very heterogeneous: GalCer levels too low
40	24	16	5	10	-	SNX1 3 µM	Nanotubes very heterogeneous: GalCer levels too low
38	22	15	5	0	-	SNX1 3 µM	Many nanotubes, well coated, but quite heterogeneous
42	25	18	5	10	-	SNX1 3 µM	Nanotubes well coated, but very heterogeneous
45	30	20	5	-	-	SNX1 3 µM	Inferior sample, thick ice. LUVs of 0.2 µm
28	16	11	5	40	-	SNX1 3 µM	Annealing protocol[134]. Good samples, many tubes, quite homogeneous, and good coating
28	16	11	5	40	-	SNX1 + retromer both 3 µM	A lot of protein background. Retromer was weakly bound to the nanotubes
23	16	11	5	40	5% 18:1 DGS-NAT(Ni	SNX1 + retromer + HisCIMPR, both 3 µM	A lot of protein background. Retromer was weakly bound to the nanotubes
23	16	11	1	40	5% 18:1 DGS-NAT(Ni)	SNX1 + retromer + HisCIMPR, both 3 µM	A lot of protein background. Retromer was weakly bound to the nanotubes
23	6	1	.5	0	5% 18:1 DGS-NAT(Ni)	SNX1 + retromer + HisCIMPR, both 3 µM	Little protein bound to the tubes
23	16	11	0.1	40	5% 18:1 DGS-NAT(Ni)	SNX1 + retromer + HisCIMPR both 3 µM	No protein bound to the tubes
28	16	11	5	40	-	SNX1/SNX5 3 µM	Protein seen on the tubes. However, later, it became clear that it was not well organised

Table 9: continued

DOPC	DOPE	DOPS	PI3P	GalCer	Other lipids	Proteins	Result
28	16	11	5	40	-	SNX1/SNX5 3 µM	Tomography. Insufficient resolution, not clear
28	16	11	2.5	40	2.5% $PI(4,5)P_2$	SNX1/SNX5 3 µM	Similar to previous samples
20	15	10	5	-	50% cholesterol	SNX1/SNX5 3 µM	Few tubes
24	19	12	5	-	40% cholesterol	SNX1/SNX5 3 µM. Liposomes of 0.05 um	Few tubes
29	23	13	5	-	30% cholesterol	SNX1/SNX5 3 µM. Liposomes of 0.05 um	Few tubes
35	25	15	5	-	20% cholesterol	SNX1/SNX5 3 µM. Liposomes of 0.05 um	Few tubes
40	27	18	5	-	10% cholesterol	SNX1/SNX5 3 µM. Liposomes of 0.05 um	Few tubes
50	30	15	5	-	-	SNX1/SNX5 1, 3, 5 and 10 µM. Liposomes of 0.05 um	Few tubes
50	30	15	5	-	-	SNX1/SNX5 1, 3, 5 and 10 µM. Liposomes of 0.2 um	Slightly more tubes than in previous tests, but still not many
50	30	15	5	-	-	SNX1/SNX5 1, 3, 5 and 10 µM. Liposomes of 0.4 µm	Quite a few more tubes than in previous tests. Tubes very heterogeneous
50	30	15	5	-	-	SNX1/SNX5 3µM. Liposomes of 0.4 µm. Optimisation of vitrification conditions	Some tubes, and (after optimising vitrification conditions), tubes were more homogeneous
45	30	20	5	-	-	SNX1/SNX5 3 µM. Liposomes of 0.4 µm	Fewer tubes than under previous conditions
45	30	20	5	-	-	SNX1/SNX5 15 µM. Liposomes of 0.4 um, 300 µM	More tubes than previous under conditions, but it was hard to find them
45	30	20	5	-	-	SNX1/SNX5 15 µM + 15 µM $CIMPR_{2330-2491}$. Liposomes of 0.4 µm, 600 µM	More tubes than under previous conditions. A lot of protein background
45	30	20	5	-	-	SNX1/SNX5 15 µM + 30 µM $CIMPR_{(2330-2491)}$. Sample was centrifuged for 30´at 20000 × g and resuspended in fresh buffer to reach higher tubular liposome concentration	**Selected conditions.**

The final conditions for the reconstitution of protein over liposomes were the following:

- Liposome composition: 45% DOPC, 30% DOPE, 20% DOPS and 5% PI(3)P.
- Liposome size: 0.4 µm; prepared in 100 mM NaCl, 0.5 mM TCEP, 10 mM HEPES pH 7.5. Liposome concentration: 600 µM.
- Liposomes were incubated overnight at RT with 15 µM SNX1/SNX5 and 30 µM $CIMPR_{(2330-2491)}$.
- The following day, the sample was centrifuged at 20000 × g for 30 minutes at 4 ºC. The supernatant was removed, and fresh buffer (half of the supernatant volume) was added to re-suspend the sample, remove the unbound protein, and concentrate the liposomes.
- The grid was vitrified under specific conditions, favouring thick and homogeneous ice: the sample was incubated on the grid for 30 seconds, blotted for 2 seconds, with 0 seconds wait after drying, and offset of -2. Grids were plasma-activated using 6 mA for 30 seconds.

4.6.2. Coflotation analysis

The interaction of liposomes with proteins was studied using coflotation analysis, adapted from the published studies[166,167,168]. The liposomes were made in a buffer containing 10% sucrose. They were incubated with the proteins, with 30% sucrose (although the liposomes still contained 10% sucrose in their interior). Then, a step gradient of 30%, 25% and 0% sucrose was prepared for centrifugation. The sample was ultra-centrifuged for one hour; the liposomes floated to the region between 25% and 0% sucrose. Only the proteins that interact with liposomes would have floated in the same band. A schematic representation of coflotation is shown in Figure 15.

In this study, the liposomes were prepared (1 mM) and filtered through a 0.2 µm cut-off filter (Whatman®) with the Mini-Extruder. The lipid composition is shown in Table 8. The liposomes were dissolved in the coflotation buffer A: 150 mM NaCl, 10% sucrose, 0.5 mM TCEP, 10 mM HEPES pH 7.5. The absorbance at 574 nm (caused by the presence of Rhodamine B) was measured to examine their concentration. Liposomes were usually prepared a day before the experiment and kept at 4 ºC under argon atmosphere. They were always used on the following day.

Next, 100 µl of liposomes (1 mM) was mixed with 100 µl of protein at 25 µM protein, to achieve a 1:1 ratio of PI(3)P to protein. The amount of PI(3)P was calculated by multiplying liposome concentration (1 mM) by the PI(3)P molar percentage (5%, or 0.05), and divided by two (assuming that half of the PI(3)P is facing the inside). When $CIMPR_{2330-2491}$ was added to examine the cargo recruitment by SNXs, a 2-fold molar excess (50 µM of CIMPR) was used. The final volume of liposome and protein mixture was 200 µl. The sample was gently stirred and incubated for 15 minutes at RT. Afterwards, a 15-µl aliquot of the mixture was taken, mixed with 15 µl Laemmli buffer 2×, boiled and kept as an input sample.

After incubation of proteins with liposomes, the sucrose gradient was prepared. First, the solution containing liposomes and proteins was mixed with 100 µl of coflotation buffer B (150 mM NaCl, 80% sucrose, 0.5 mM TCEP, 10 mM HEPES pH 7.5). It was gently mixed, so the final sucrose concentration in the bottom layer would be 30% (see Figure 16). Then, using a glass Pasteur pipette, the second layer (300 µl) of the gradient (coflotation buffer C: 150 mM NaCl, 25% sucrose, 0.5 mM TCEP, 10 mM HEPES pH 7.5) was very gently deposited on top of the sample. Finally, buffer F (100 µl) was carefully added over the second layer. Afterwards, the samples were centrifuged at 240000 × g for 1 hour at 4 ºC, using an Optima L-100 XP Ultracentrifuge (Beckman Coulter), in an SW 55Ti rotor. The samples were centrifuged in the ultra-Clear™ tubes of 5 × 41 mm (Beckman Coulter, ref. 344090), with appropriate adaptors for this rotor (ref. 356860, Beckman Coulter). After centrifugation, the tubes were carefully taken out, and the layer of liposomes (visible due to the presence of Rhodamine B, see Figure 16B) was collected using a Hamilton syringe. To load the same amounts of liposomes in SDS-PAGE wells, absorbance at 574 nm was measured. The samples were normalised against the sample with the smallest amount of liposomes. An aliquot of 20 µl of this sample was mixed with 5 µl of Laemmli buffer 5× and boiled, to be run on PAGE. A volume containing the same amount of liposomes (as the sample with the smallest absorbance) was taken from each of the remaining samples. The proteins were visualised using Coomassie staining.

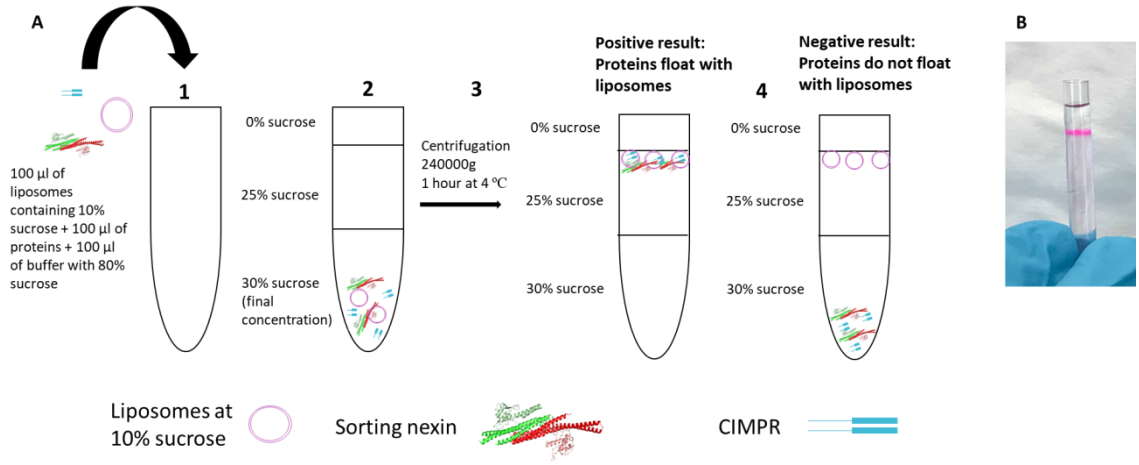


Figure 16: Schematic representation of coflotation. (A) Coflotation process shown step by step. First, liposomes and proteins were mixed and incubated for 15 minutes at RT (1). Then, the gradient was made (2), and the sample (in the bottom layer) was centrifuged (3). After centrifugation, the layer containing liposomes, visible as a pink band between the 25% and 0% sucrose levels, was collected and examined on SDS-PAGE (4). (B) The gradient after centrifugation. A band of liposomes (pink layer) can be seen.

4.7. Cryo-electron microscopy (cryoEM)

Electron microscopy is a technique that allows visualisation of biological (and non-biological) specimens at a high-resolution (nanometre to angstrom range). Biological samples cannot be visualised at the very highest resolutions because a large electron dose causes radiation damage, destroying the sample. However, such a specimen can be imaged thanks to the interaction of accelerated electrons with the sample in a high-vacuum system. The electrons can go through the sample or be scattered by it and are identified by a detector that converts the electrical signals into digital data (image). The devices used in this field can be photographic films, charged-couple device cameras or direct electron detectors.

CryoEM allows the examination of samples in their near-native state, at very low temperatures (-180 ºC), which protect them partially from electron-radiation damage. To visualise a sample accurately, it needs to be vitrified: it has to be fast-frozen to form vitreous (i.e., amorphous) ice, and not ice crystals. This is achieved using liquid ethane at about -180 ºC, cooled in a special chamber surrounded by liquid nitrogen[169].

Furthermore, the ice must be very thin, not thicker than 500 nm, so the electrons can go through the ice and interact with the specimen.

This project used cryoEM to examine the interaction of SNXs with membranes, to find out how these proteins affect tubulation and understand their arrangement on the tubules. A large number of samples were screened (see Table 9) to establish the conditions under which the SNXs are properly attached to membranes and can be visualised. Reconstitution of proteins on artificial membranes (i.e., LUVs) was evaluated on Quantifoil R2/2 300-mesh copper grids. The grids were hydrophilised by plasma treatment using a Glow discharge/carbon evaporation MED 020 system (BALTEC). Then, they were vitrified in liquid ethane in an FEI vitrobot^{TM} Mark III. The samples were stored in a homemade grid box and kept in a Dewar flask filled with liquid nitrogen. They were screened using CIC bioGUNE JEOL 2200FS cryo-electron microscope equipped with an UltraScan 4000 SP (4008 × 4008 pixels) cooled slow-scan CCD camera (GATAN, UK).

However, in later experiments, we used a different cryoEM approach: cryo-electron tomography (cryoET, see section 4.8.2). The CryoET alignment was guided by BSA Gold Tracer 6 nm **(**Electron Microscopy Sciences, Ref. 25484**).** High-resolution cryoET dataset was obtained in an FEI Titan Krios G3, coupled to a Gatan K2 Summit camera, at the University of Leicester.

4.7.1. Sample preparation

Sample preparation for cryoEM was optimised for our reconstitutions (see section 4.6.1, or Table 9). Briefly, 3–4-µl aliquot of the sample was placed on an R2/2 Cu grid, dried by blotting it with filter paper and vitrified by quick introduction into liquid ethane in a cryostat (FEI). The time between blotting and vitrification has to be kept to an absolute minimum. Within milliseconds of sample blotting, before vitrification, the specimen can be affected by the air-water interface[170]. Tubules obtained by the standard vitrification protocol were highly heterogeneous in thin ice, which leads us to think that the air-water interface was impairing the quality of our sample. Hence, we modified the vitrification protocol to obtain thicker ice, resulting in a considerable improvement in the properties of the sample.

First, the grids were subjected to plasma treatment under specific conditions: 6 mA, 2×10^{-1} millibar air pressure, and 30 seconds of plasma. Plasma is used to make the grid more hydrophilic: the more plasma, the higher hydrophilicity, and the better absorption to the carbon (see Figure 17). However, under our conditions, the grid was kept slightly more hydrophobic, reducing sample attachment to the carbon and thickening of the ice layer in the holes. In our case, the average sample thickness of 183 nm was calculated for the high-resolution cryoET samples (see Figure 17B). The sample was incubated with the grid for 30 seconds, with 2-second blotting, (virtually) 0 seconds of waiting time after blotting, and an offset of -2 (which is only meaningful for our specific Vitrobot). For cryo-electron tomography, 6-nm gold fiducials were added at a ratio of 1 to 8 (v/v): 1 volume of fiducials per 8 volumes of sample.

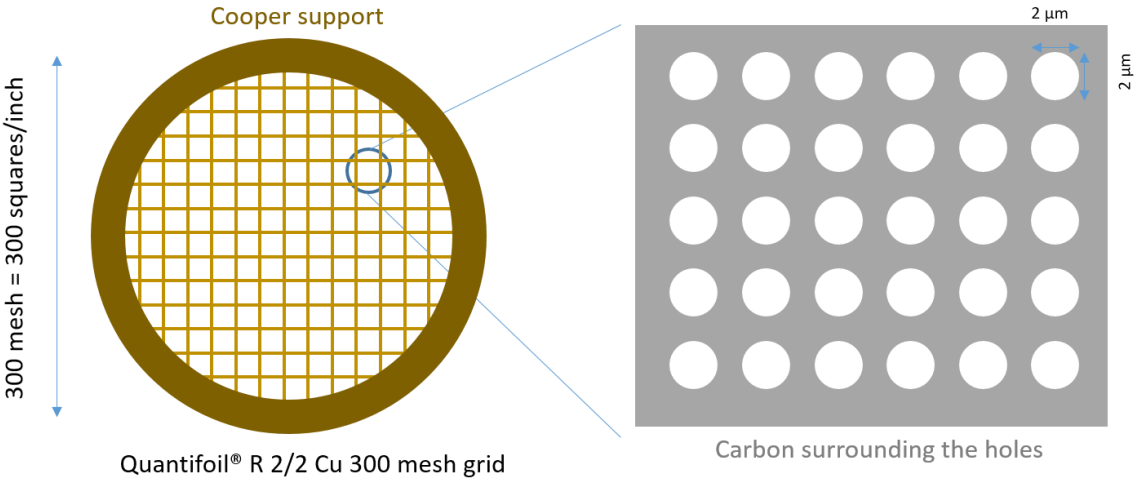


Figure 17: Schematic view of the grids used in this study. Three-mm grids with copper support, 300 squares per inch. There are 2 × 2 µm holes in the squares, surrounded by carbon. The sample is vitrified in these holes.

Sample screening was carried out at CIC bioGUNE using a 200 kV electron microscope (JEOL 2200FS). In general, the screening was conducted at 40000× magnification and defocus ranging from 3 to 5 µm. This allowed us to improve the samples to be used in high-resolution data collection, as well as to quantify the number of tubes per sample. To avoid bias, the photographs were taken using the minimal dose system (MDS): a hole was selected at 4000× (where the particles cannot be distinguished), and then a photograph was taken at 40000× magnification.

4.7.2. Cryo-electron tomography (cryoET)

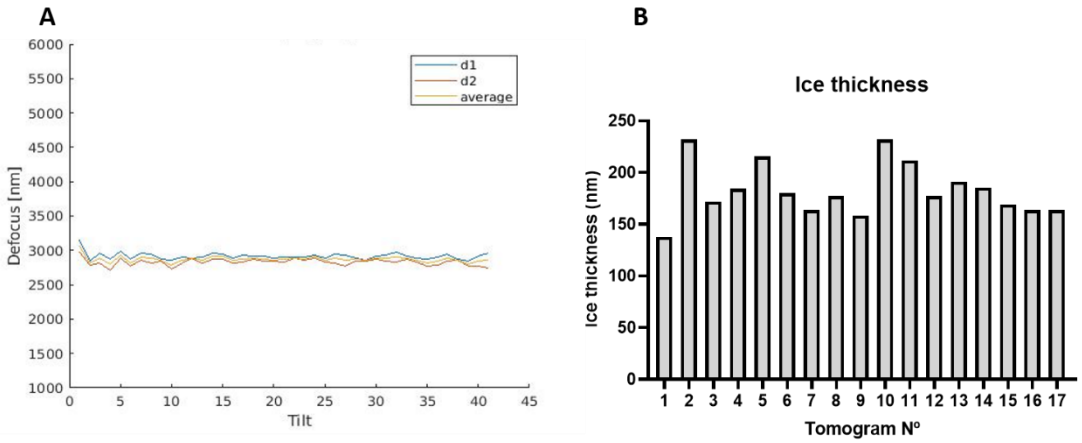
CryoET is a variant of cryoEM where the sample is tilted (usually between -60º and +60º) and imaged typically at each 1–3º. The objective is to combine all collected views and obtain a reconstructed 3D volume, a tomogram. Several tomograms of the same specimen can be averaged (sub-tomogram averaging) to increase the resolution and try to reach the sub-nanometre range. Moreover, the new software and data collection strategies are continuously improving the resolution achieved by this technique[171,172]. It is now widely used to examine the surface of whole cells, thin sections of cells, and complex heterogeneous samples.

The technique is limited to high tilts (> 60º–70º) due to the metallic lattice of the grid and an increase in sample ice thickness during tilting. The region for which the data are not collected, the missing wedge, causes loss of information in Fourier space. This generates artefacts and distortions that have to be corrected during processing. Furthermore, obtaining high-resolution data is hampered by the low signal to noise (SNR) ratio. In cryoET, the specimen is imaged approximately 40 times, so the electron dose has to be carefully distributed between images.

We first attempted to use the lipid nanotubes premade using GalCer (see Table 8) to solve the structure of SNX employing a single-particle approaches. We had hoped to take advantage of the expected helical symmetry of BAR domains on the membrane[134,145] (see results section for a more detailed description). However, a recent publication from John Briggs group has shown that yeast Vps5, an SNX1 orthologue, was organised in a pseudo-helical, not helical, manner[92]. This deviation from the helical symmetry makes it impossible to obtain the parameters needed to solve helices (pitch, rise and rotation angle).

Given the pleiomorphic nature of the tubes, we decided to use cryoET (see Table 8). Although the tubes were not as homogeneous as when GalCer was used, it was clear that SNX induced the tubulation, possibly in a more natural way as there was no order imposed by ceramides[173]. Furthermore, using the cryoET, one can overcome the problems caused by heterogeneity; smaller sub-boxing strategies allow the study of smaller components of the sample despite not being symmetrically distributed. For further information, see section 4.9.2.

High-resolution cryoET was carried out at the cryoEM facility of Leicester University. Fifty-six tomograms were obtained. For each tomogram, 41 images were



taken: from -60º to +60 degrees, in three-degree steps. A dose-symmetric scheme was applied[174]. For each tilt angle, 10 images were taken in a movie with a total electron dose of 114 $e^-/Å^2$ and 0.28 $e^-/Å^2$/frame. Data was acquired with a K2 camera (Gatan Inc.) at 53000× magnification and rendering a pixel size of 2.73 Å. Defocus was set between 2 and 5 µm with defocus steps of 0.5 µm. SerialEM software was used for data collection. The most homogenous tubes were selected manually, by taking photos at low magnification (6000×; with the help of Dr Soledad Baños). The microscope setup was established by Dr Christos Savva (Leicester University), who performed alignment of the electron beam.

4.7.3. CryoET data processing

There are several steps for data processing that have to be performed before any tomogram reconstruction or particle alignment. These were conducted using Matlab software (MathWorks®). Matlab scripts and functions were kindly donated by our collaborators Stefano Scaramuzza and Daniel Castaño-Diez from Biozentrum (Bassel).

Before tomogram reconstruction, images were motion-corrected using MotionCorr2 software. The programme detects and corrects the beam-induced motion, that is, the movement of the specimen due to its interaction with electrons. Samples were dose-weighted by reducing 20% of the used dose (0.28 $e^-/Å^2$/frame) to be more conservative. Afterwards, the drift-corrected movies were merged into tilt series. In this case, two tilt series were obtained: dose-weighted (for sub-tomogram averaging) and unweighted (for CTF correction, so the SNR is increased).

Figure 18: (A), CTF estimation of each tilt (d1 and d2) of tomogram 6 as an example. CTFFIND4 estimated defocus twice (d1 and d2), and then the average (yellow line) was

calculated. (B), ice thickness of the first 17 tomograms. The average thickness is 183 nm.

Then, CTF correction was carried out using CTFFIND4[175] programme, obtaining a defocus for each image. The output of CTFFIND4 was used to correct the CTF of the whole tomogram using Ctfplotter[176] from IMOD software package[172]. An example of the CTF correction of tomogram 6 using CTFFIND4 is shown in Figure 18.

4.7.4. Reconstruction of tomograms

Once the tomograms were corrected for beam-induced motion and optical aberrations, they were merged into a tilt series, and the tomograms were reconstructed. IMOD programme was used for tomogram reconstruction[172]. It is a software that guides the user in a step-by-step process for tomogram building. Briefly, tilt series were corrected for X-ray pixels coming from the detector and an initial fixed stack of the tilt series was obtained. Although direct electron detectors should not have any X-ray pixels, the software found and eliminated several pixels from our tomograms. Then, the tomograms were aligned, and cross-correlation between them was automatically calculated. A coarse alignment performed by the programme was used as a template for manual alignment using the MIDAS software package from the IMOD suite. The third step was to find all gold fiducials, make an automatic alignment and then align them manually. As a rule of thumb, the standard deviation alignment error must be below 1; they were all between 0.35 and 0.85. Afterwards, the sample was positioned in the tomogram. The whole tomogram was produced (binned 3 times for speeding up), and the sample to be analysed was examined. Then, the upper and lower limits were set manually. In the fourth step, a final alignment was performed, and CTF was corrected using Ctfplotter. Defocus values for each tilt, obtained from CTFFIND4, were used as input. Finally, the tomogram was built by the programme and generated by back projection. Although the simultaneous-iteration reconstruction technique (SIRT) would raise the contrast of the tomogram, subsequent processing (i.e., sub-tomogram averaging) gives better results from back-projected tomograms.

4.8. Collaborative methods

4.8.1. Calculation of residue interaction energies

Calculation of residue interaction energies was carried in the group of Dr Juan Fernández-Recio by Miguel Romero, at the Instituto de Ciencias de la Vid y el Vino, Logroño. Once the structures of SNX1/SNX5 BAR domains were solved, they were evaluated using their own developed pyDock software, a computational method which combines electrostatics, desolvation and van der Waals energy scoring functions to provide a systematic analysis of per-residue contribution to the binding energy. The aim was to find the residues important for (hetero) dimerisation.

4.8.2. Computation of per-residue docking energy

Residue contribution to the binding energy was estimated using resEnergy pyDock module[177]. Taking the structure of a complex as input, the module computes pyDock docking energy partitioned at the residue level, giving a detailed view of the energy landscape of the interaction. Residue contribution of SNX1-SNX1 homodimer (PDB id 4FZS) and SNX1-SNX5 heterodimer was calculated using this method. The mean residue docking energy for SNX1-SNX1 homodimer models was computed form an ensemble of 20 models generated in Modeller[178], configured with default parameters and using the SNX1/SNX5 structure as the starting homology model.

4.8.3. Sub-tomogram averaging (StA)

Sub-tomogram averaging was performed by Stefano Scaramuzza and Daniel Castaño-Diez, our collaborators from the Biozentrum (University of Basel, Switzerland). Images were processed using the *Dynamo* software[172] for sub-tomogram averaging.

In the tomograms, the individual particles (SNX1/SNX5 heterodimers with CIMPR on the surface of tubular membranes) were not visible to the naked eye. Only their overall helix-like arrangement on the surface of the tube could be recognised (Figure 19). Particles had to be extracted on a per-tube basis, following a strategy involving oversampling and sub-boxing. First, large-sized sub-volumes were extracted along each tube (Figure 19A), aligned and averaged. A low-resolution map of each tube, where the helical pattern could be seen, was obtained (Figure 19B). Individual molecules were

still not visible at this stage. The helix was automatically parametrised. Based on this parametrisation, new coordinates were defined along the helix, using small and equidistant spacing (oversampling). These coordinates were mapped back to the full tomogram and used for a new sub-volume extraction procedure in which medium-sized sub-volumes were extracted (sub-boxing). The medium-sized particles were aligned and averaged again, using a well-defined alignment mask. The mask excluded neighbouring particles that might affect the alignment (Figure 19C). These sub-volumes were large enough to accommodate 2-3 SNX1/SNX5 heterodimers. The average revealed a low-resolution shape of the particles, allowing to distinguish individual particles for the first time (Figure 19D). Then, a further sub-boxing step was performed. Visible SNX1/SNX5 particles were manually selected, new orientations automatically computed, and their coordinates mapped back to the full tomogram. These new coordinates were then used again to extract small-sized sub-volumes, each containing a single SNX1/SNX5 molecule. The small sub-volumes were aligned and averaged one more time (Figure 19E). After this process was performed for each tube, all small-sized particles were merged into one single large dataset. An outlier exclusion and further alignment gave the current map.

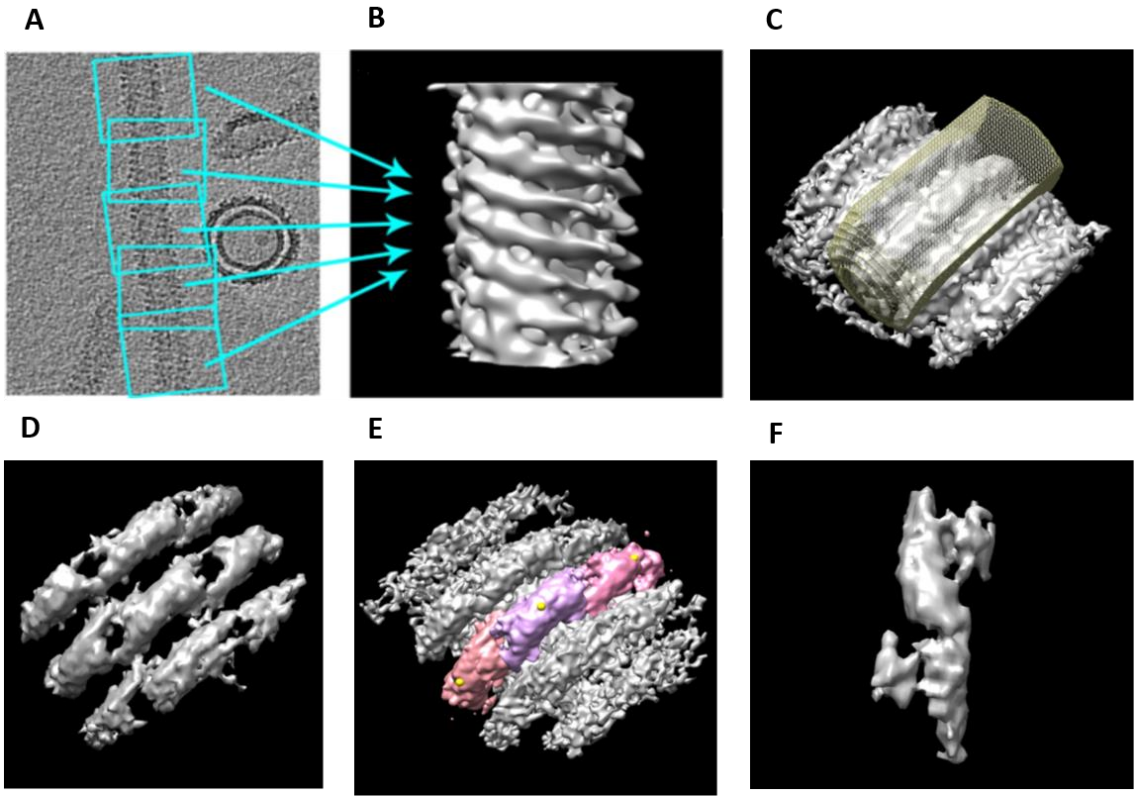


Figure 19: Sub-boxing procedure was used for particle extraction from the tomography data set. (A) Coordinates corresponding to big segments of a single tube in a tomogram were selected. (B) Averages of the previous particles revealed a helical pattern, which was used to get the coordinates of SNX1/SNX5 on the membrane through oversampling. (C) Once helical parameters were established, a more precise medium sub-boxing was computed, obtaining 2–3 SNX1/SNX5 particles. (D) Individual particles were observed and used for a small sub-boxing step. (E) Individual heterodimers can be observed in the reconstruction. (F) A low-resolution volume of SNX1/SNX5 after the last sub-boxing was finally visualised.

5. RESULTS

5. RESULTS

5.1. Purification of SNX1, SNX5 and SNX1/SNX5 heterodimer

The aim of this thesis is the characterisation of SNXs implicated in cargo recycling, with a focus on membrane deformation and cargo selection. To achieve this, the components involved in the process of recycling must be first purified. Although the individual SNXs have been studied[106], little is known about the heterodimerisation of these proteins. Several studies have indicated that the heterodimers, rather than homodimers, might be the functional units[125,127]. Hence, the first step of our research was to purify the SNX homodimers (SNX1, SNX5) and heterodimers (SNX1/SNX5).

5.1.1. Purification of SNX1

First, the purification of individual SNXs was performed. SNX1 was selected as a model because it can form tubes from liposomes, while others, such as SNX5, cannot[106].

For SNX1 purification, the His-tag approach was used (see Methods). The protein tended to degrade during the procedure. Hence, particular care was taken to obtain a full-length and highly pure protein; all the purification steps were performed at 4 ºC and as quickly as possible to minimise degradation. The SDS-PAGE gels (Figure 20) show the level of protein purity after each step. The SNX1 protein is highly soluble (Figure 20A, lane SF) and binds very well to the Ni-NTA resin (Figure 20A, R1 lane). However, after elution from the resin, the protein showed some degradation (Figure 20A, E lane). This was even more pronounced after overnight dialysis at 4 ºC (Figure 20A, lane E2).

Sentrin-specific protease 2 (SENP2) was used to remove the His-SUMO tag during dialysis. After tag removal, SNX1 has a theoretical isoelectric point of 5.8. Hence, anion exchange chromatography was conducted, and the fractions with fewest impurities (lanes 5 to 8, Figure 20B, bottom) were concentrated and injected onto an S200 16/60 size-exclusion column. Note that after the last purification step (size-exclusion chromatography, Figure 20C), some contamination products were still observed. The most pure fractions were pooled (lanes 1 to 5, Figure 20C, lower panel) and used for further experiments.

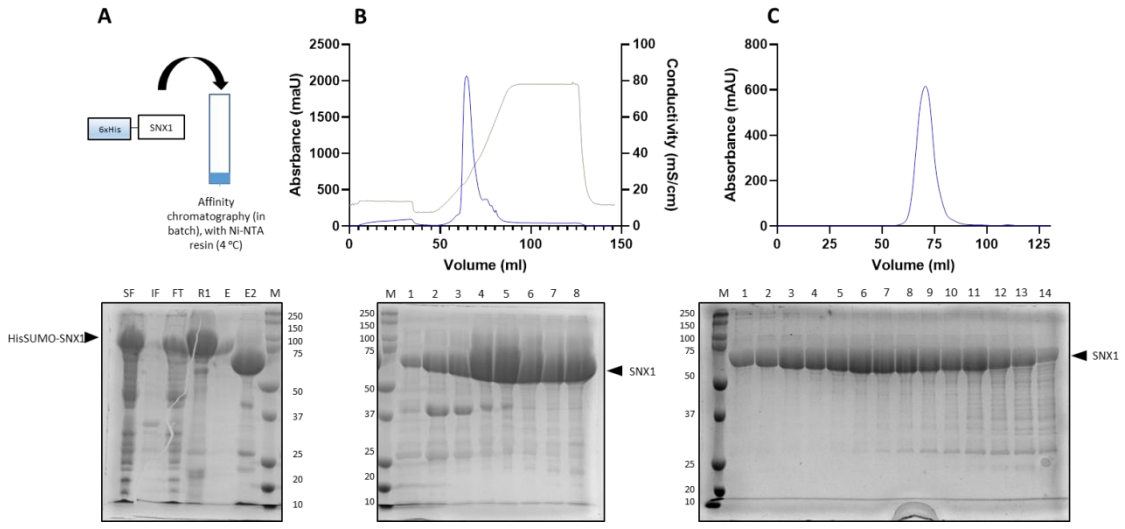


Figure 20: Purification of SNX1. (A) Ni-NTA batch purification. Different lanes on the SDS-PAGE gel show the soluble fraction (SF), the insoluble fraction (IF), flow-through (FT), resin before elution (R1), elution (E), and the protein after tag removal with SENP2 (E2). (B) Anion exchange chromatography of SNX1. Proteins from lanes 5 to 8 were pooled, concentrated and loaded onto S200 16/60 column. (C) Size-exclusion chromatography of SNX1. The fractions corresponding to lanes 1 to 6 were pooled. In all the gels, lane M shows the molecular mass markers. The sizes of marker proteins (in kDa) are given on the left, the fraction numbers of analysed samples are noted above the gel. The position of SNX1 band is indicated with an arrow.

5.1.2. Purification of SNX5

It has been reported that SNX5 and related proteins (SNX6 and SNX32) cannot produce tubes from liposomes *in vitro*[106]. To understand why SNX1 is capable of tubulating liposomes while SNX5 is not, both proteins were purified and characterised.

SNX5 purification followed a slightly different protocol. First, it was purified using a GST-tag. The protein was highly insoluble, prone to aggregation and susceptible to degradation (see Figure 21A, soluble fraction and elute, and chromatogram in Figure 21B). The best strategy to obtain a homogeneous protein entailed multiple size-exclusion chromatography steps. The initial size-exclusion results showed three overlapping peaks (Figure 21B). The fractions were evaluated using SDS-PAGE, and those containing degradation products were discarded (Figure 21B and 21C). Subsequent size-exclusion steps helped to remove aggregates. Note that the aggregates

were not seen on the SDS-PAGE gel (Figure 21, bottom panels), but the non-uniform chromatographic peak showed heterogeneity (i.e., impurity and/or aggregation) of the sample. After several steps, the peak was symmetrical, indicating a homogeneous protein population (Figure 21E).

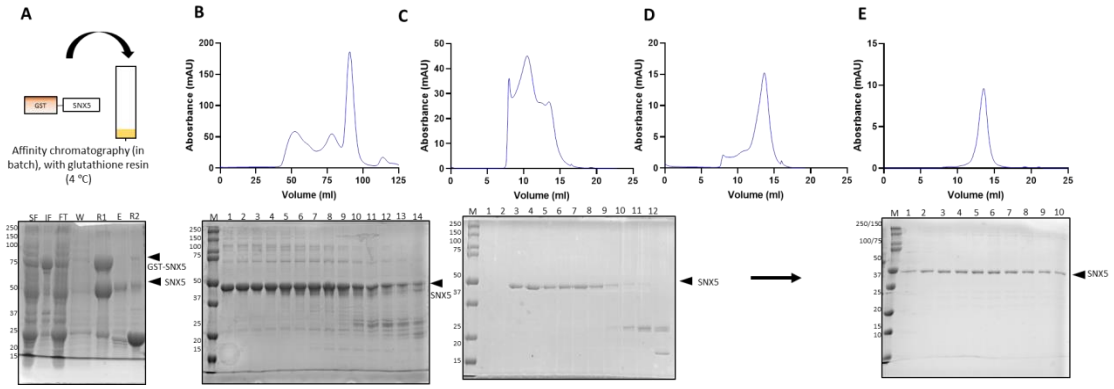


Figure 21: Purification of SNX5. (A) SNX5 was tagged with GST to increase solubility; however, it was still mostly insoluble after tagging (IF). Some of the protein could be recovered from the glutathione resin (A, lane E). Lane R2 shows proteins that were not recovered from the resin after the GST removal. (B) The first size-exclusion showed aggregates (in the void volume of the column, 40 ml). The small peak at 76 ml matched the expected elution volume of a 50-kDa protein. (C–E) Pooled fractions were rerun three times on a size-exclusion S200 10/300 column; the final peak showing monodisperse sample can be seen in (E).

5.1.3. Purification of SNX1/SNX5 heterodimer

SNX heterodimers were expected to be the functional unit in the retromer recycling pathway. However, at the time, there was no protocol for the purification of SNX heterodimer. We devised and implemented a new protocol, described in the Methods chapter. Using this protocol, we were able to purify the SNX1/SNX5 heterodimer.

The SNX1 and SNX5 proteins were co-expressed in *E. coli* BL21 (DE3) strain and tagged with His-SUMO-tag and GST-tag, respectively. SNX1 partly recovered SNX5 from the inclusion bodies (Figure 22A, lane SF and IF). Both proteins bound to the glutathione resin thanks to the GST-tag of SNX5. After washing and subsequent removal of the excess of His-SUMO-SNX1, the proteins bound to the resin were incubated with TEV protease overnight at 4 ºC, so the heterodimer of His-SUMO-SNX1/SNX5 was eluted. The following day, the heterodimer was captured on Ni-NTA

resin and excess of SNX5 was washed out (Figure 22A, lane FTNi) (see section 4.3.2). His-SUMO was removed during overnight dialysis with SENP2 protease. The following day, anion exchange chromatography was done. Degradation products were removed in this step; fractions 3 to 5 (Figure 22B) from anion exchange chromatography were pooled, concentrated and loaded onto an S200 16/60 column, obtaining a well-resolved peak (Figure 22C).

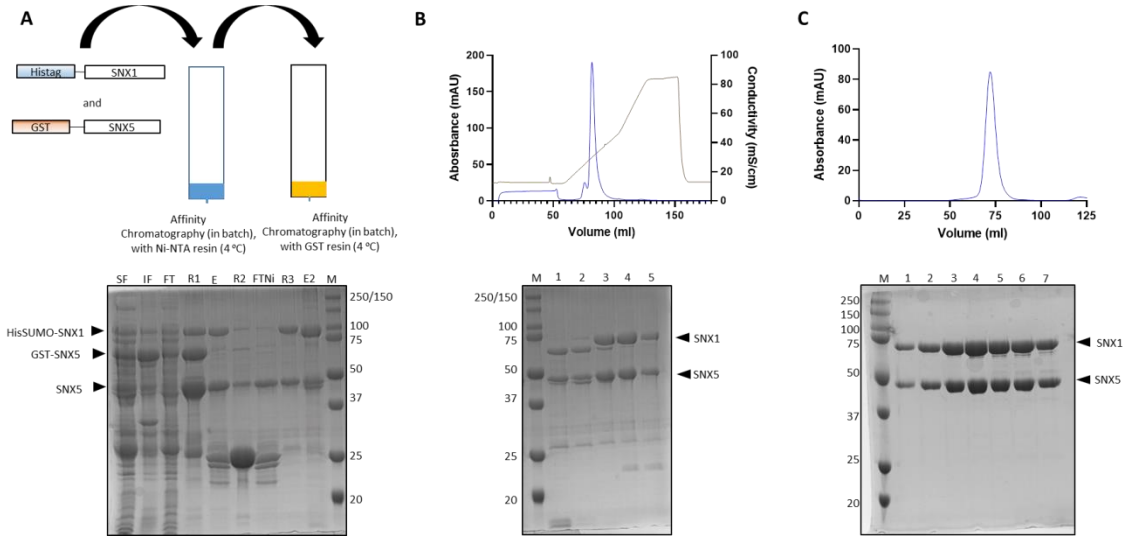


Figure 22: Purification of SNX1/SNX5 heterodimer. (A) Purification began with GST resin, as described in section 4.3.3. The lanes show the soluble fraction (SF), the insoluble fraction (IF), flow-through (FT), GST resin before elution (R1), elution from GST resin (E), the proteins on GST resin after cutting with TEV protease (R2), flow-through after incubation with Ni-NTA resin (FTNi), the proteins bound to Ni-NTA resin before elution (R3) and after elution from Ni-NTA resin (E2). Excess of SNX5, removed by washing, can be seen in the FTNi sample. (B) Anion exchange chromatography of SNX1-SNX5 heterodimer allows the removal of degradation products (see gel in the lower panel). Fractions corresponding to lanes 3 to 5 were pooled, concentrated and loaded onto a size-exclusion column. (C) Size-exclusion of SNX1/SNX5. Proteins seen in lanes 2 to 6 were pooled, concentrated and frozen.

5.1.4. The oligomeric state of SNX1-SNX5

BAR domain-containing sorting nexins can self-dimerise and form dimers with other SNXs[106]. The elution volumes from size-exclusion chromatography of SNX1 and SNX5 differed considerably (Figure 23A). SNX5 (46816 Da) homodimer and SNX1 (59049 Da) homodimer should elute at similar volumes under these conditions. However, the elution volume for SNX1 was 10.3 ml, and for SNX5, 13.5 ml (Figure 23A). SNX1 elution volume was comparable to that obtained for the paralogue SNX2 (purification protocol the same as for SNX1) and the heterodimer SNX1/SNX5. These results suggest a difference between SNX5 and other SNXs (Figure 23B). Therefore, the proteins were analysed using SEC-MALS as this technique provides an accurate estimate of molecular weight. SNX1 and SNX1/SNX5 were found as dimers in solution. SNX1 (Figure 23C) had a mass of 116100 Da, with a deviation of 2% from the theoretical mass of a dimer (118098 Da). SNX1/SNX5 (Figure 23E) had a theoretical mass of 106305 Da, and the mass of 101600 Da was obtained in the experiment (deviation of 4.5%). SNX5 (Figure 23D) showed the mass of 49670 Da, with a deviation of 4.9% from the theoretical mass of a monomer (47274 Da).

Note that the SEC-MALS analysis of SNX1 and SNX1/SNX5 was conducted using a Superdex 200 Increase 10/300 column because the positive surface of SNX1 prolongs the retention in silica-based columns (such as Shodex columns). In contrast, the SEC-MALS analysis of SNX5 rendered good results on a Shodex KW-403 column (B21 beamline, Diamond Light Source).

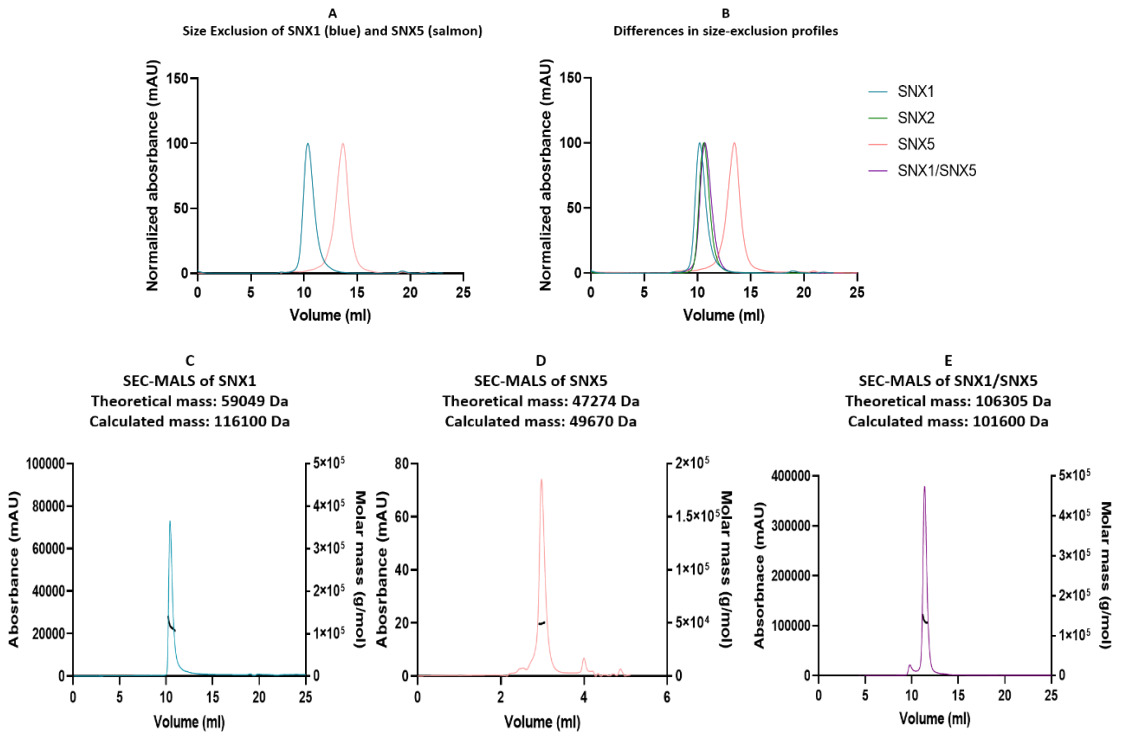


Figure 23: The oligomeric state of SNXs estimated by size-exclusion (A, B) and confirmed by SEC-MALS (C–E). (A) Comparison of size-exclusion profiles of SNX1 (blue line) and SNX5 (salmon-pink line). The significant shift in elution volumes indicated a difference in size, not explained by the theoretical molecular weight. (B) Related SNXs (SNX2 and SNX1/SNX5 heterodimer) eluted at the volumes similar to those for SNX1. (C) SEC-MALS of SNX1, indicating the molecular weight of a dimer (see Table 11). (D) SEC-MALS of SNX5, indicating that it behaves as a monomer in solution. (E) SEC-MALS of SNX1/SNX5 heterodimer, with the mass corresponding to an expected dimer.

5.2. Structural characterisation of the SNX1/SNX5 BAR heterodimer

The crystallisation of the full-length SNX1/SNX5 heterodimer was attempted, but no crystals were obtained. A possible reason could be the flexibility between PX and BAR domains and the presence of flexible regions, like the N-terminus of SNX1, which might preclude the formation of an ordered lattice. Hence, we decided to perform the structural characterisation of SNX1/SNX5 BAR domain, which would allow us to understand SNX dimerisation and the differences between homodimers and heterodimers.

First, *in silico* analysis of the proteins was performed to find out the best constructs for crystallisation. In the case of SNX1, the construct was based on the structure obtained by van Weering and co-workers[106] (Lys301 to Ser522, see Figure 27D). For SNX5, a secondary (using PSIPRED[179], Figure 24B) and tertiary structure prediction (using Phyre2[180], Figure 24C) analyses were conducted to design a construct starting in an unstructured area before the BAR domain. (Figure 24B). An alignment of its sequence versus other SNXs with solved structures was also performed (SNX1 BAR domain, 4FZS[106], SNX9 and 3DYT[181,182], and SNX33 and 4AKV, unpublished data). The structural alignment and secondary structure prediction suggested that the region Lys192 to Asp206 has a random coil conformation (Figure 24A and B). However, the tertiary structure prediction obtained with Phyre2 (Figure 24C) indicates that the region from 181 to 195 might have an α-helix. Considering these data, an SNX5 construct from Asp195 to the C-terminus (Asn 404) was designed and co-expressed with SNX1 BAR domain, as described in section 5.1.3. A charged residue at the N-terminus was chosen to enhance solubility of the terminal hydrophobic residues.

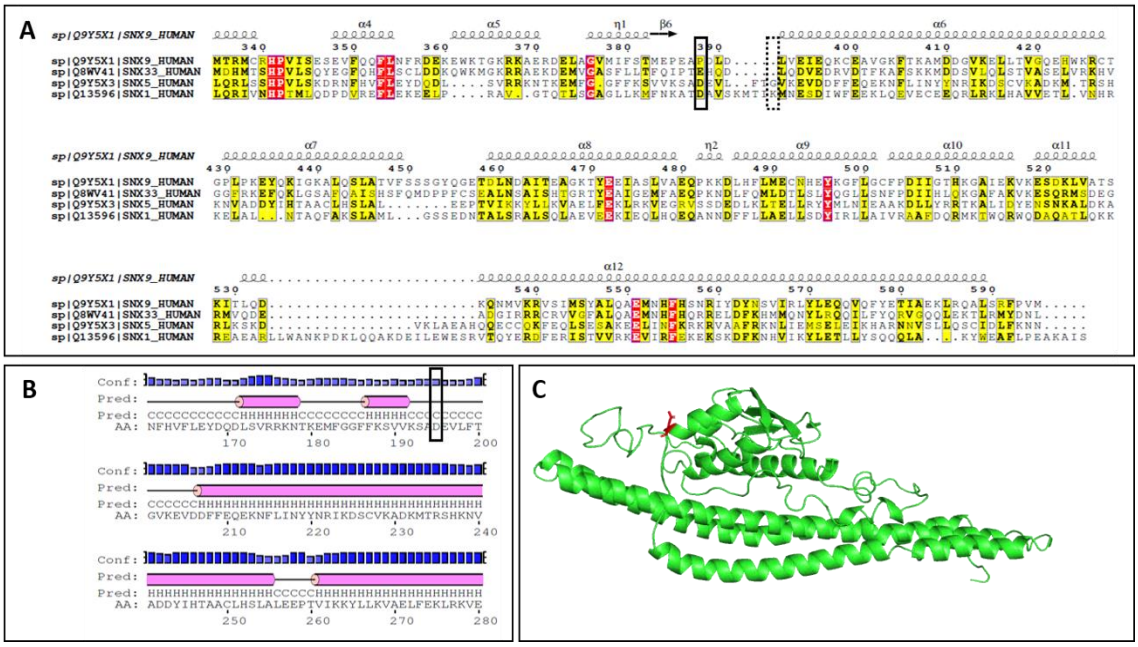


Figure 24: Tools used for SNX5 construct creation. (A) Sequence alignment (Clustal Omega, representation in EsPript 3.0[183], http://espript.ibcp.fr) of the sequence of human SNX9 (Q9Y5X1), human SNX33 (Q8WV41), human SNX5 (Q9Y5X3) and human SNX1 (Q12596). The solid black-line box marks the residue D195 (SNX5) selected as the cloning site. The dotted-line box shows the first residue of the SNX1 construct. The structure shown above the sequence corresponds to SNX9. (B) Secondary structure prediction of SNX5 (from amino acid 160 to 280) produced using PSIPRED. The black

box indicates the residue D195. The prediction expects a random coil at this position. (C) Structure prediction of SNX5 obtained using Phyre2, with the residue D195 shown in red (at the end of a α-helix).

5.2.1. Structure of BAR domains of SNX1/SNX5 heterodimer

SNX1$^{301-522}$/SNX5$^{195-404}$ BAR domains were crystallized using vapor diffusion methods. Upon evaluation of numerous crystallization screens from various commercial sources (3072 crystallisation conditions, table 4), diamond shape crystals were obtained after 2-3 days at 18ºC, in the MIDAS (Molecular Dimensions) condition 1-17 (8% polyvinil alcohol (PVA), 10% 1-propanol, 0.1M HEPES-NaOH pH 7). Those crystals diffracted a low resolution (Figure 25A). This crystallization condition was further refined employing the hanging-drop method and bigger volumes. Good crystals were obtained by mixing 1 µl of the heterodimer at 8 mg/ml with 1 µl of the precipitant solution containing 100 mM HEPES pH 7, 11% w/v Polyvinyl alcohol, and 10% v/v 1-propanol (Figure 25B). However, the best diffracting crystals were obtained after several micro-seeding steps (Figure 25B and C). Native crystals were cryo-protected by quick-soaking into mother liquor supplemented with 20% (v/v) glycerol before being flash-frozen in liquid nitrogen.

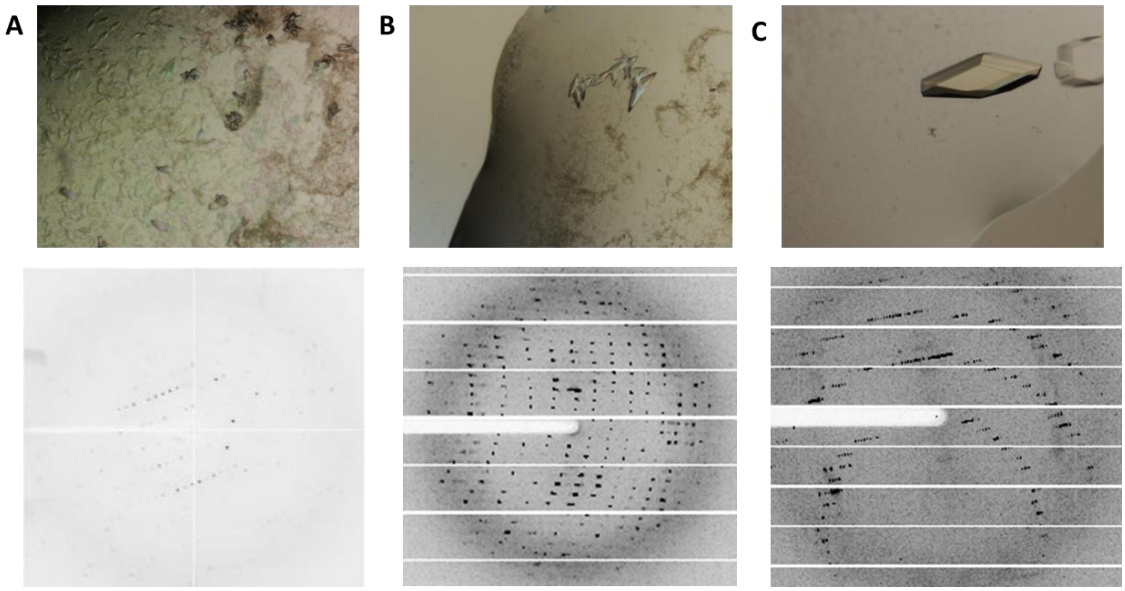


Figure 25: Crystallisation and diffraction patterns of the SNX1/SNX5 BAR domain. (A) Initial crystals (top panel) that diffracted to 10 Å resolution (bottom panel). (B) Crystals obtained after refinement and initial micro-seeding steps (top panel) that

diffracted to 2.8 Å resolution (bottom panel). (C) Crystal obtained after several iacro-seeding steps (top panel), it was soaked in K_2PtBr_4 to performed a Multi-wavelength Anomalous Dispersion (MAD) data collection, and it diffracted up to 2.9 Å resolution (bottom panel).

Data indexing, integration, and scaling of native crystals were performed using the XDS programme [159]. The crystals belonged to space group $P2_12_12$ (see Table 10 for data collection statistics). Matthews coefficient[184] indicated the presence of two copies of the heterodimer in the asymmetric unit. Molrep (CCP4 suite[160]) and Phaser (PHENIX suite[185]) were used to perform Molecular replacement using the BAR domains of known structures of other sorting nexins as search models; SNX9 (PDB code 3DYT), SNX33 (PDB code 4AKV) and SNX1 (PDB code 4FZS). However, no solution was found. Then, we focused on obtaining a selenomethionine (SeMet) derivative. We purified and crystalized SeMet substituted $SNX1^{BAR}/SNX5^{BAR}$, however despite the crystals diffracted to 3.2Å resolution they suffered from serious radiation damage. In this sense, when collecting 360 degrees the data-sets exhibited a completeness of 89.4% and the anomalous signal was very weak, thus being infeasible the structure solution by SAD approaches[155] (see Table 10 for data collection statistics).

In parallel, the isomorphous replacement approach (SIR) was tried. An initial *in silico* analysis[156] of the most adequate heavy atoms for derivatization of heterodimer crystals pointed that platinum salts (among others) could bind to the complex. Stocks solutions of seven different compounds containing heavy atoms (1 M) were prepared in Milli-Q water (see Table 7 in Methods section). Then, individual crystals (Figure 25B) were incubated with each solution diluted at 1, 10 and 100 mM in the precipitant solution for at least 15 minutes (+20% glycerol as cryo-protectant). Suitable derivatives were obtained by 10 minutes-soaking into the cryoprotectant solution supplemented with 10 mM K_2PtBr_4. Afterwards, crystals were fast-frozen in liquid nitrogen and tested for diffraction-quality and the presence of anomalous signal. Diffraction images were indexed and integrated with MOSFLM and scaled with SCALA. The best crystals diffracted up to 3.7 Å resolution and belonged to the space group P2. SHELXC/D as implemented in autoSHARP programme was used to find the positions of heavy atoms in the crystal. At this point, the crystals displayed an anomalous signal that was low (SigAno =1.2 at 4.4 Å) and the phasing power was not enough to find a solution using SAD.

After several micro-seeding steps, bigger crystals (Figure 25C) were obtained that allowed more than one data collection on the same crystal but on different spots, so multi-wavelength anomalous dispersion (MAD) approach was used. A scan around the X-ray absorption edge of the platinum (L-III) was performed to estimate the value of the peak, inflexion and remote wavelengths for the optimal data collection; (remote 1.04473 Å, inflexion point, 1.07293 Å and peak, 1.07158 Å). The three datasets were collected from the same crystal. Data reduction was carried out using MOSFLM program and scaling using SCALA program. The crystals belonged to space group P2 (see Table 10 for data collection statistics). Heavy atoms positions were identified using SHELXC/D as implemented in autoSHARP[161]. After phasing, subsequent density modification using SOLOMON gave a starting map into which 14 chains with 702 residues were automatically built with Buccaneer. Next, iterative refinement with PHENIX[162] and manual building in COOT[163] yield a final model with two heterodimers in the asymmetric unit (4 chains with 792 residues). This model was used as a template for molecular replacement with a native dataset that diffracted to 2.5 Å resolution. The final structure has a R_{fac} and a R_{free} of 22.4% and 27.8% respectively. (Data collection statistics for each dataset are shown in Table 10).

The crystal structure of $SNX1^{BAR}/SNX5^{BAR}$ exhibits the classical curved (banana like) shape determined by the angle of dimerization (Figure 26). Each monomer consists of a coiled-coil formed by three α helices folded onto each other. In the concave face of the protein, there are several patches of positive charge, which most probably contribute to the association with the negatively charged surfaces of membranes (Figure 26B). The structure of $SNX5^{BAR}$ lacks 16 residues, from Asp327 to His342, located at the tip of the BAR domain (Figure 26A, see SNX5 in salmon-pink) since the absence of electron-density on this region precluded model building.

Table 10: Refinement and data collection statistics for SNX1/SNX5 BAR domain.

	SNX1-SNX5	SNX1-SNX5	SNX1-SNX5	SNX1-SNX5	SNX1-SNX5
	Native	**SeMet**	**Pt_PEAK**	**Pt_INF**	**Pt_REM**
Data collection					
Wavelength [Å]	0.9793	0.9789	1.07158	1.07293	1.04473
Space group	$P2_1$	$P2_1$	P2	P2	P2
Resolution [Å]	50-2.5 (2.65-2.50)	50-3.0(3.19-3.01)	50-2.8(2.95-2.80)	50-2.8(2.95-2.8)	50-2.6 (2.74-2.6)
Cell dimensions					
a, b, c [Å]	50.10, 189.30, 52.10	49.77, 189.80, 52.40	52.69, 51.47, 192.32	52.94, 51.49, 192.36	52.77, 51.55, 192.46
α, β, γ [°]	90.0, 90.0, 90.0	90.00, 90.00, 90.000	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
$CC_{1/2}$ (%)	99.3 (63.2)	99.6(20.6)	99.9(70.7)	99.7(39.0)	99.9(12.3)
Completeness (%)	98.5 (94.5)	89.4(86.7)	96.2(96.1)	96.0(96.3)	95.4(95.3)
I/σ	7.15 (1.69)	3.49(0.56)	10.3(1.7)	6.7(1.1)	8.1(1.1)
Number of unique reflexions	32404 (4992)	34550(5409)	24815(3609)	24868(3618)	30770(4469)
Redundancy	5.17 (4.58)	1.8 (1.7)	6.3 (6.2)	6.0 (5.8)	6.2 (6.1)
Refinement					
R-factor (%)	22.4		25.2		
R-free (%)	27.8		30.5		
No. atoms					
Waters	26		4		
Ions/Pt	0		8		
Ligand/propanol	7				
R.m.s deviations					
Bond lengths (Å)	0.001		0.002		
Bond angles (°)	0.293		0.398		
PDB CODE	**6ZB7**		**6ZDD**		

*Highest resolution shell is shown in parenthesis.

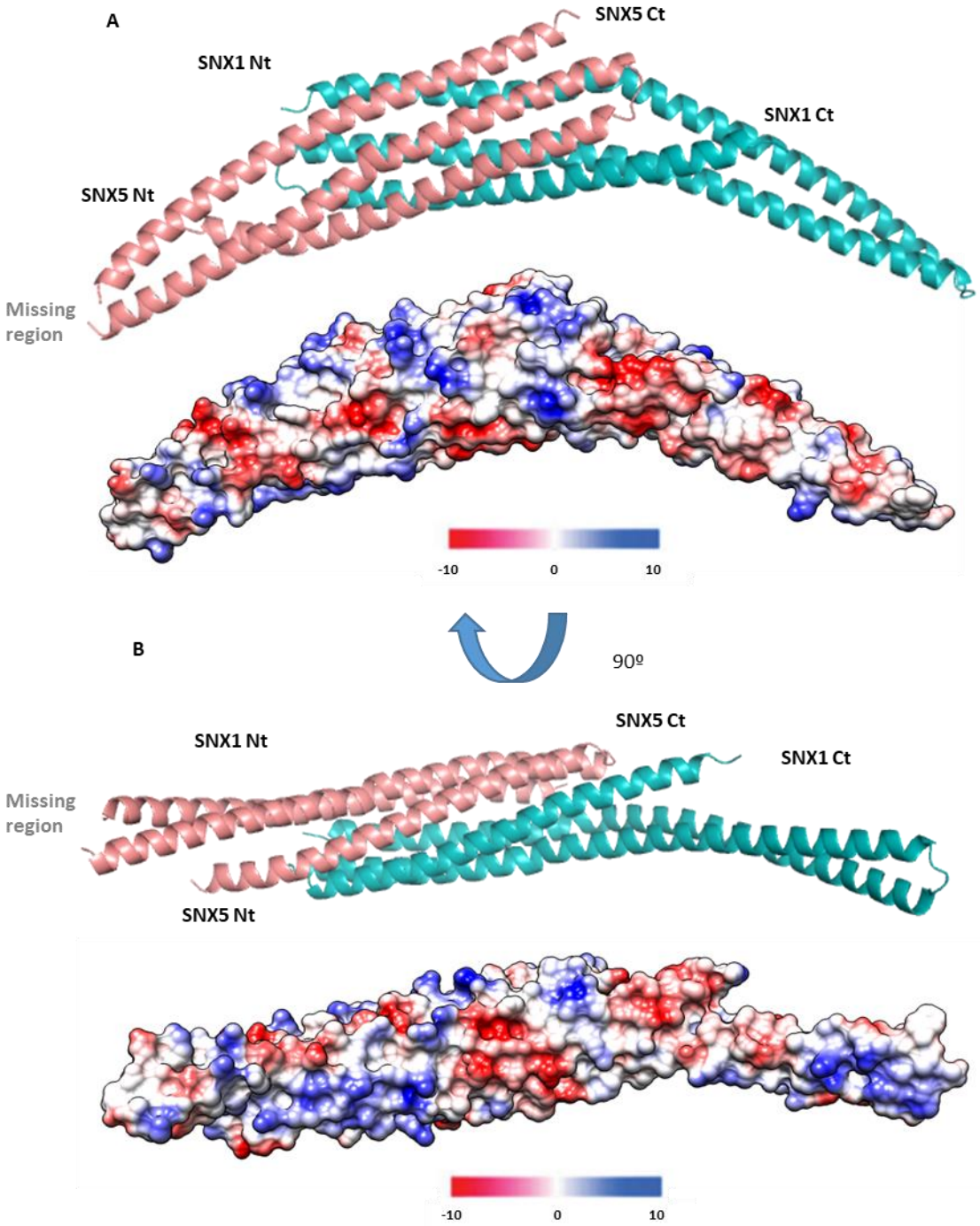


Figure 26: Structure of SNX1/SNX5 BAR domains. (A) Crystal structure (top) of $SNX1_{301-522}$ (cyan) in complex with $SNX5_{195-404}$ (salmon-pink) and electrostatic view (bottom) of the structure. Blue colour indicates positively charged residues and red, negatively charged. (B) 90-degree view of (A), looking at the concave face that would touch the membrane. Cartoon view (top) and electrostatic view (bottom), where the patches of positive charge can be seen. Cartoon views were generated in PyMOL[186], and electrostatic views using Chimera[187]. Electrostatic units scale are in $kcal \cdot mol^{-1} \cdot e^{-1}$, from 10 to -10, as shown at the bottom of both panels.

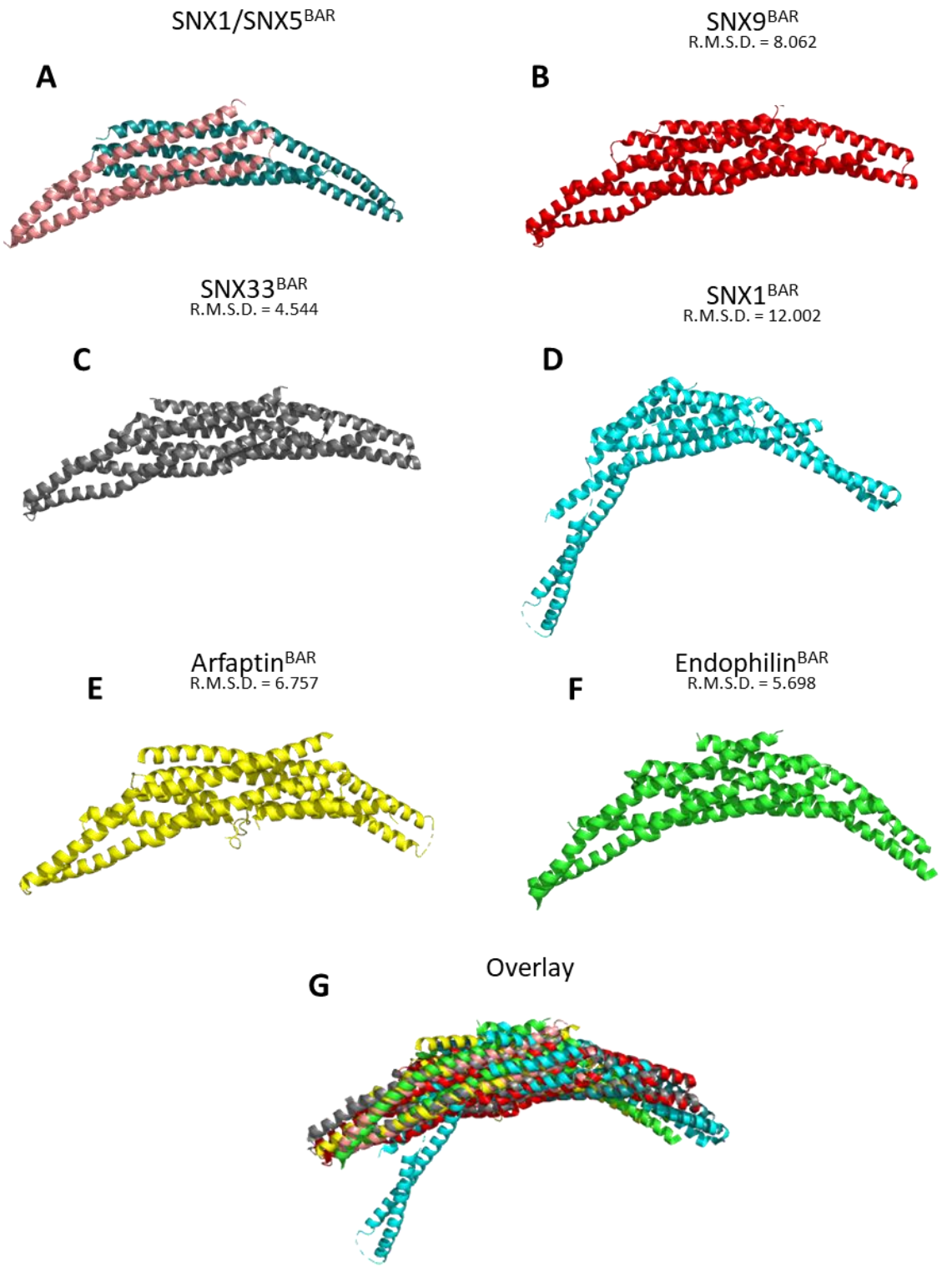


Figure 27: Crystal structures of BAR domains. (A) SNX1/SNX5 BAR domain (this study), (B) SNX9 (PDB code: 2RAI), (C) SNX33 (PDB code: 4AKV), (D) SNX1 (PDB code: 4FZS), (E) Arfaptin (PDB code: 1I4D), (F) endophilin (PDB code: 1ZWW). (G) Overlay of all the structures. The R.M.S.D. of the backbones of each protein compared with $SNX1^{BAR}/SNX5^{BAR}$ is indicated below the name of each protein. R.M.S.D. were calculated using PyMOL[186].

Figure 27 shows the crystal structure of several BAR domains. Some belong to proteins closely related to SNX1 and SNX5: SNX9[181], SNX33 (unpublished) and SNX1[106]. Others come from the distantly related proteins endophilin[188] (N-BAR) and arfaptin[140] (classical BAR, or N-BAR). It can be observed that $SNX1^{BAR}/SNX5^{BAR}$ heterodimer resembles other BAR domains. Root mean square deviation (R.M.S.D.) of the backbone between $SNX1^{BAR}/SNX5^{BAR}$ and the rest of the BAR domains oscillate between 4.5 and 8 Å, except for $SNX1^{BAR}$ (figure 27D), which is 12 Å. $SNX1^{BAR}$ structure has a higher degree of curvature caused by two breaks in the helices α2 and α3. Nonetheless, the crystallographic structure of $SNX1^{BAR}$ homodimer is not consistent with the diameter of the tubules that it forms and therefore might not be representative of an active-state conformation[106].

5.2.2. Analysis of the BAR domain dimer interface

To understand the differences between the homodimers and heterodimers, we attempted to crystallise the $SNX1^{BAR}$ homodimer. However, after screening more than 3000 different conditions, we did not obtain crystals with good diffraction patterns. Therefore, SNX1 homodimer was modelled using MODELLER 9.17[177] (Figure 28A) using our crystal structure of the $SNX1^{BAR}/SNX5^{BAR}$ heterodimer as a model (Figure 28B). The thermodynamic contribution of each residue was measured using the resEnergy module of pyDock[177] for both SNX1 (Figure 29) and SNX1/SNX5 (Figure 30). Then, to complement the analysis, in silico alanine scanning was conducted to determine which residues are important for dimer stability. Two residues in SNX1 (Phe347 and Trp511) and up to 6 residues in SNX5 (Tyr219, Arg368, Met233, Val240, Ile398 and Phe401) show a high energy contribution to dimerisation. As it can be seen in Figure 28C, the residues Phe347 and Trp511 in SNX1 face towards the dimerisation interface.

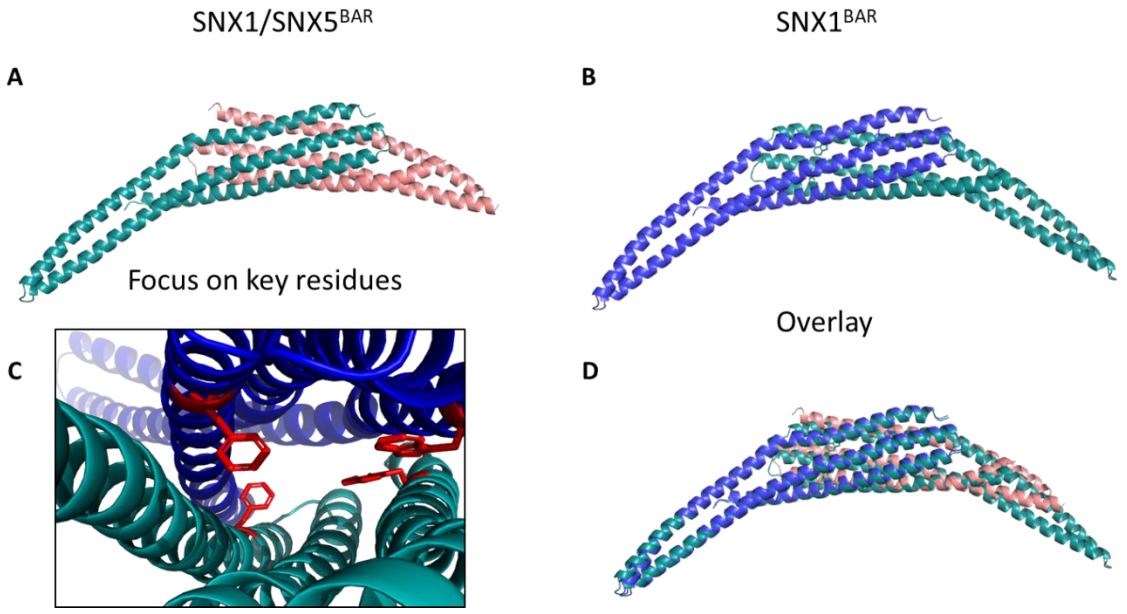


Figure 28: Comparison of modelled $SNX1^{BAR}$ homodimer and solved $SNX1^{BAR}/SNX5^{BAR}$ heterodimer. (A) SNX1 (cyan) + SNX5 (salmon-pink) heterodimer. (B) Model of $SNX1^{BAR}$ homodimer. Note that although both chains are SNX1, they are coloured differently for clarity. (C) Zoom on the key residues (F347 and W511, shown in red) of SNX1, which contribute most to the stabilisation of homodimers. (D) Superimposition of $SNX1^{BAR}/SNX5^{BAR}$ and $SNX1^{BAR}$ domain.

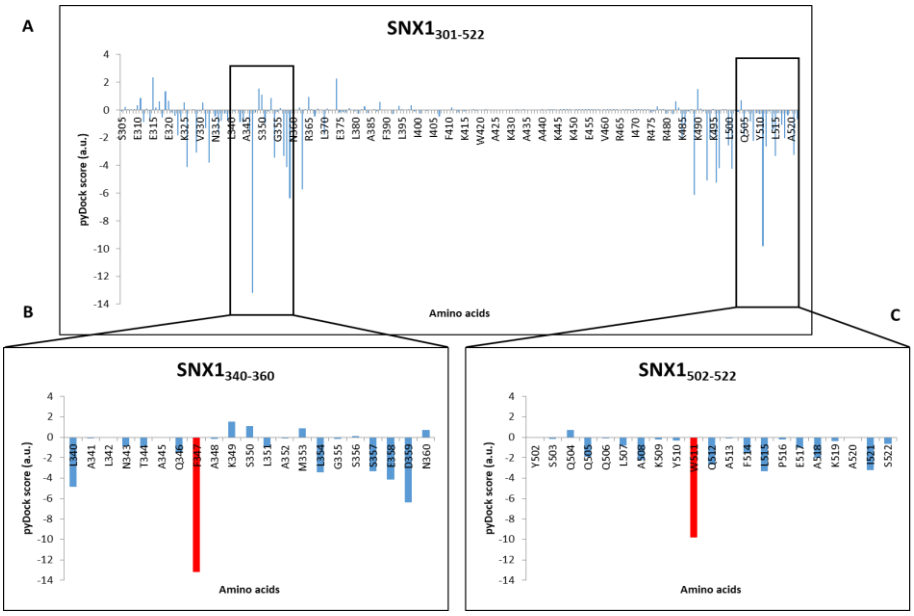


Figure 29: pyDock results for SNX1 model. (A) The energy contribution of each residue in SNX1 to homodimer dimerisation. (B) Zoom window of twenty amino acids, showing Phe347 (highlighted in red). (C) Zoom window of twenty amino acids, showing Trp511 (in red).

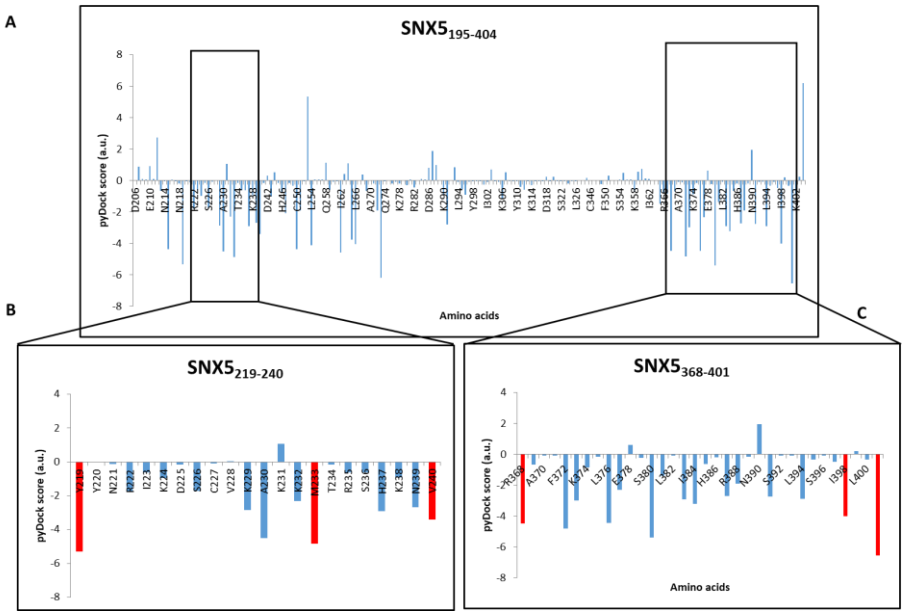


Figure 30: pyDock results for SNX5 residues (A) Energy contribution of each residue in SNX5 to heterodimer dimerisation. (B) Zoom window of 21 amino acids, showing Tyr219, Met233 and Val240, highlighted in red. (C) Zoom window of 33 amino acids, showing Arg368, Ile398 and Phe401 highlighted in red.

The amino acids that contribute energetically more to the heterodimerisation were analysed by sequence alignment between human variants of SNX1 (Figure 32A) or SNX5 (Figure 32B) against different species. They were aligned against rat (*R. novergicus),* chicken (*G. gallus*), zebrafish (*D. rerio*) and yeast (*S. cerevisieae).* SNX1 key residues (Phe347 and Trp511) are conserved from yeast to human. SNX5 key residues (Tyr219, Met233, Val240, Arg368, Ile398, and Phe401) present slightly more variability. Phe401 is substituted by a Leucine in zebrafish, changing the aromatic nature of the amino acid. The other residues are better conserved, with only minor substitutions to similar amino acids, like Ile240 instead of Val240 in chicken. In the case of the yeast orthologue Vps17, there is hardly any conservation (except for Met233). This data is in accordance with previous studies, which suggest that SNX5 and Vps17 might not be orthologues[17]. The high degree of conservation suggests that these residues are essential for the dimerisation of SNX1/SNX5 heterodimer.

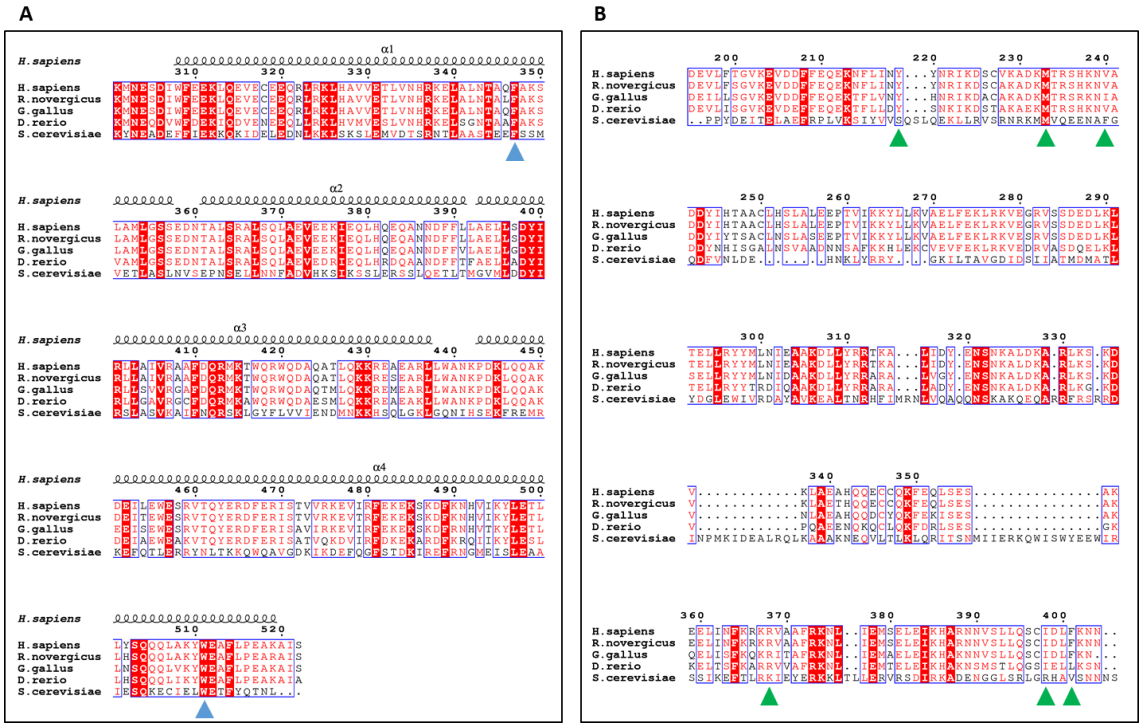


Figure 32: Alignment of the amino acid from different SNXs. (A) SNX1 alignment between *Homo sapiens* (lane 1), *Rattus novergicus* (lane 2), *Gallus gallus* (lane 3), *Danio rerio* (lane 4) and *Saccharomyces cerevisiae* (lane 5). The structure of α-helix of human SNX1 is shown above. The key residues (Phe347 and Trp511) are indicated with a blue triangle. (B) SNX5 alignment between *Homo sapiens* (lane 1), *Rattus novergicus* (lane 2), *Gallus gallus* (lane 1), *Danio rerio* (lane 4) and *Saccharomyces cerevisiae* (lane 5). The key residues (Tyr219, Met233, Val240, Arg368, Ile398, and Phe401) are indicated with a green triangle. Alignments were done in Clustal Omega and data were represented using Espript 3.0.

5.2.3. Mutations in SNX1 break the homodimer

The residues critical for dimerisation were mutated, the proteins purified, and their oligomeric state and tubulation capacity evaluated. The two critical residues in SNX1, stabilising the dimerisation, are Phe347 and Trp511. To examine their importance, they were mutated to alanine. Three constructs were made: $SNX1_{F347A}$, $SNX1_{W511A}$ and $SNX1_{F347A+W511A}$. Their oligomeric state was evaluated using SEC-MALS (Figure 33). The individual mutations did not disrupt dimerisation but destabilised it. In Figure 33B ($SNX1_{F347A}$) and C ($SNX1_{W511A}$), a shoulder in the chromatogram at higher elution volume can be seen, indicating that some of the protein is not a stable dimer and behaves as monomer. Furthermore, the absolute mass did not match the mass of a homodimer, as it did with the wild-type protein (Table 11). However, when the two

mutations were combined ($SNX1_{F347A+W511A,}$ Figure 33C), the dimerisation of SNX1 was abolished. The peak is slightly displaced to the right, indicating a reduced molecular weight. Dimer dissociation was confirmed by MALS analysis, which indicated that the protein had a molecular weight of 61 kDa, matching the mass of a monomer (59049 Da).

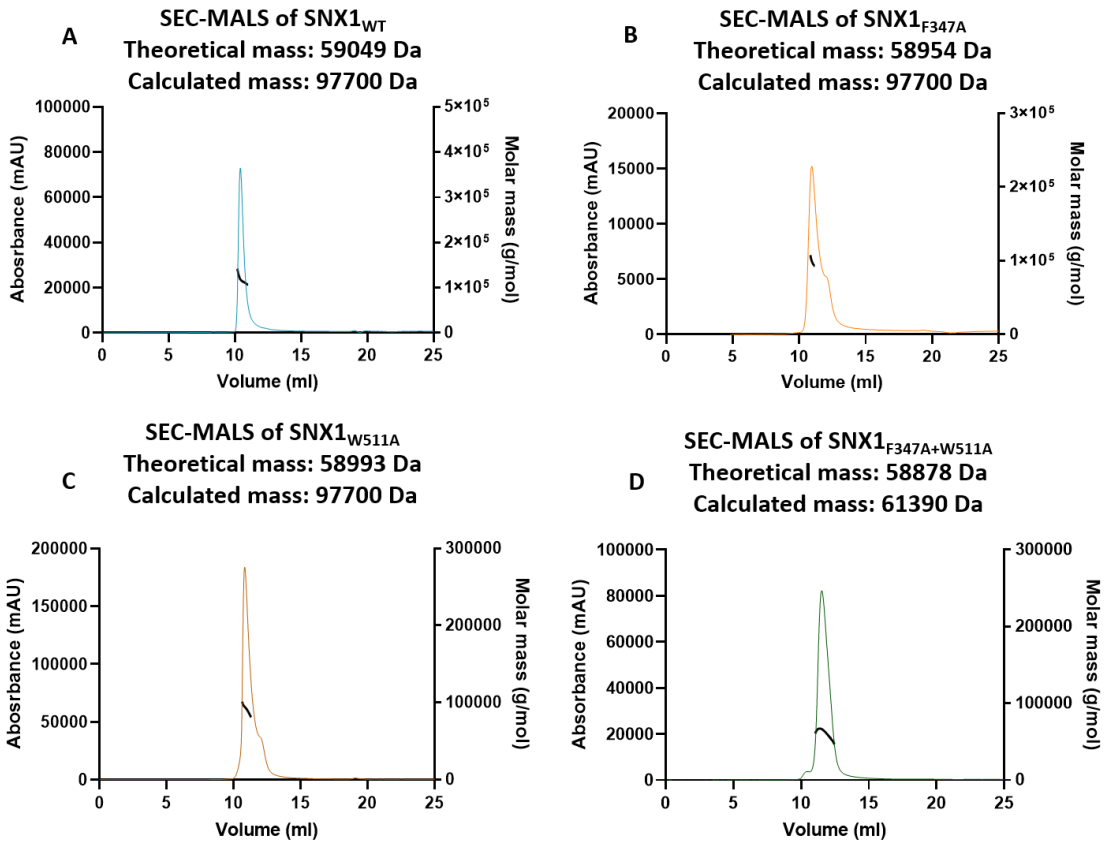


Figure 33: SEC-MALS of wild-type SNX1 (A) and the different mutants analysed (B–D). (A) SEC-MALS results for SNX1 WT. (B) SEC-MALS results for $SNX1_{F347A}$, showing the mass of around 100 kDa (close to a dimer) (C) SEC-MALS results for $SNX1_{W511A}$; the mass of around 100 kDa (close to a dimer) (D) SEC-MALS results for $SNX1_{F347A+W511A}$; the mass of 59 kDa (monomer).

As mentioned above, SNX1 has a human paralogue, SNX2. When the two proteins were aligned, it became clear that the residues of SNX1 critical for maintaining the dimerisation, F347 and W511, were conserved in SNX2 (F344 and W509). Then, both residues were mutated to alanine and the oligomeric state of $SNX2_{F344A+W509A}$ was examined. Strikingly, the dimerisation was not affected (Table 11 and Figure 34). This

means that SNX1 and SNX2, despite being paralogues, contain distinct binding free energy contributions of interfacial residues.

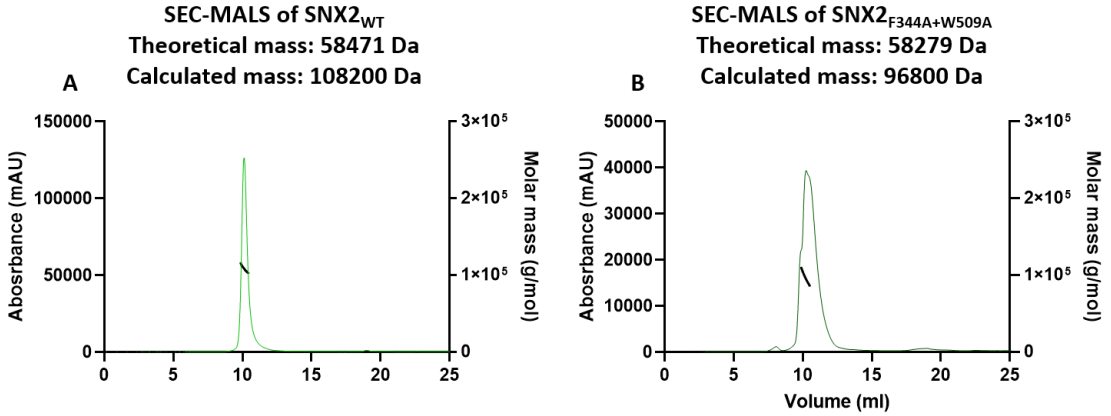


Figure 34: SEC-MALS data for SNX2. (A) Wild-type SNX2. (B) $SNX2_{F344A+W509A}$. The mass calculated from the sequence (theoretical mass) and the mass obtained experimentally (calculated mass) are shown above each graph.

Table 11: Theoretical mass obtained from the amino acid sequence and the calculated mass from SEC-MALS data presented in Figures 22, 33 and 34. Note that $SNX2_{F344A+W509A}$ does not match a dimer, but tends to be closer to a dimer than to a monomer. The asterisk (*) indicates a large deviation from the calculated mass of a dimer.

	SNX1	SNX5	SNX1+SNX5	SNX1 F347A+W511A	SNX2 WT	SNX2 F344A + W509A
Theoretical mass (monomer)	59049 Da	47274 Da	106305 Da	58878	58471	58279
Calculated mass (MALS)	116100 Da	49670 Da	101600 Da	61390 Da	108200 Da	96800 Da
Deviation	1.7%	4.9%	4.5%	4.1%	7.5%	17%
Oligomeric state	Dimer	Monomer	Dimer	Monomer	Dimer	Dimer*

5.2.4. Comparison between the $SNX1^{BAR}$ and $SNX5^{BAR}$ interfaces

The structural alignment of the central region between both BAR domains results in a RMSD of 0.96 Å, indicating that the BAR domains are very similar. The energetic contribution of each amino acid, indicated by orange bars in Figure 31, shows that hydrophobic amino acids are, in general, the ones that contribute more to the binding. Amino acids present at the interface are marked with a grey rectangle atop them. It can be observed that hydrophobic amino acids drive SNX1/SNX5 heterodimerisation, while charged amino acids prevents SNX5 homodimerisation.

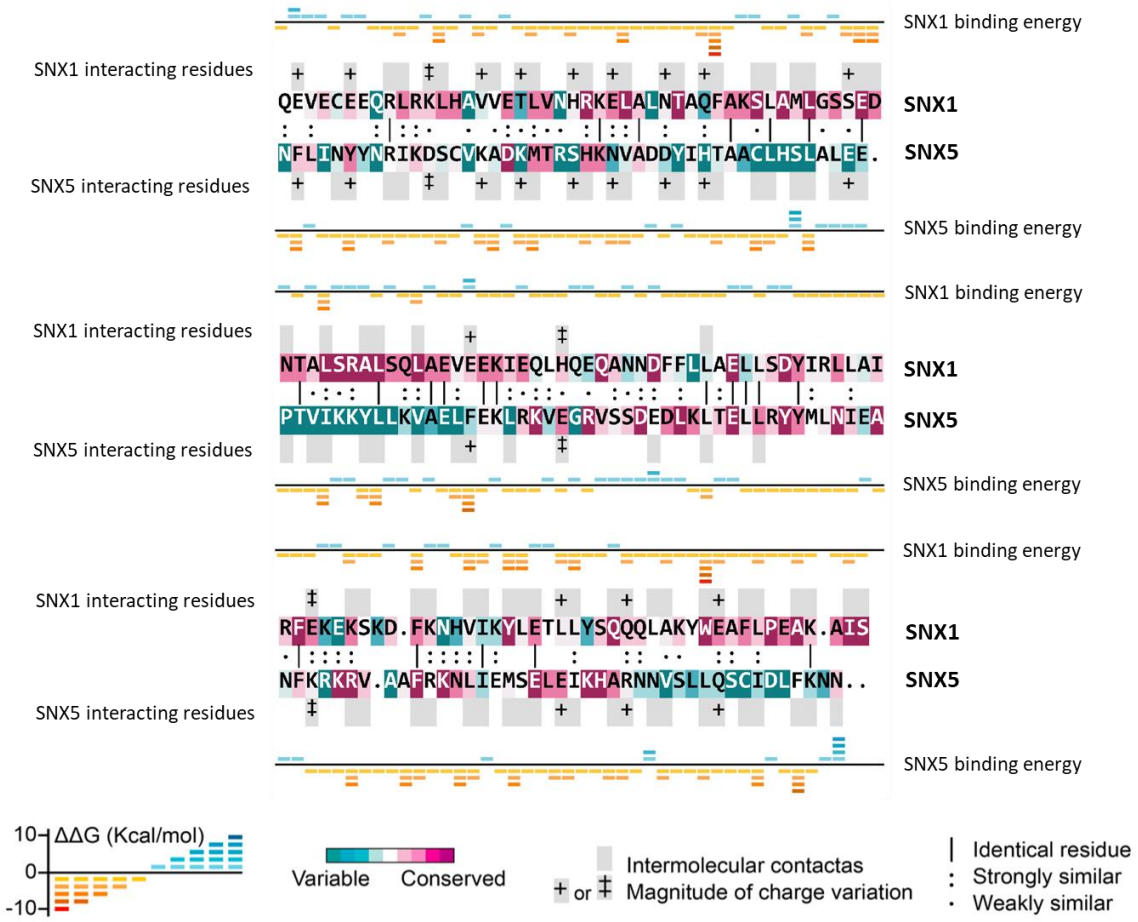


Figure 31: Sequence alignment from the structural superposition of the central regions of the BAR domains from SNX1 and SNX5. Amino acid alignment was done using Clustal W from UCSF Chimera[187]. "Binding energy" depicts for amino acid binding energy using the pyDock scoring function[189], which is based on electrostatics and desolvation energy. It is shown in blue (positive energy) or yellow-orange (negative energy), in Kcal·mol^{-1}. "Interacting residues" (grey bars) indicates the interacting residues defined on the basis of a cut-off distance below 4.5 Å. "+" and "‡" represent the magnitude of charge variation - range of residue formal charges, assuming –1 for D/E, +1 for H/K/R, and 0 for all other types. Color scheme of the amino acids indicates the evolutionary conservation degree of each amino acid based on the ConSurf webserver[190], going from blue (variable) to red (conserved).

5.2.5. SNX5 can displace SNX1 homodimer *in vitro*

We established that SNX1 is a dimer and SNX5 is a monomer in solution (Table 11). Furthermore, when these proteins are co-expressed in *E. coli*, a heterodimer of SNX1/SNX5 with a molar ratio of 1:1 is obtained (Figure 21C). It seems that SNX5 displaces SNX1 homodimer to form a SNX1/SNX5 heterodimer. A pull-down-based competition assay (see section 4.4.2) was designed to confirm that SNX5 can displace SNX1 *in vitro.* In this type of experiment, one protein (known as "bait") is purified with an affinity tag (His-SUMO-SNX1). Then, this protein is incubated with another protein, the "prey" (SNX5), for a specific time (in our case, 1 hour at 4 ºC) with a resin that can trap the bait (Ni-NTA resin to trap His-SUMO tag). Afterwards, the resin is gently centrifuged, washed with buffer, and placed on the gel for SDS-PAGE. If the prey and bait interact, both proteins will be seen in the gel. If not, only the bait protein will be seen.

It was observed that the wild-type SNX5 ("prey") could separate SNX1 homodimer, forming heterodimers (Figure 35B, lane 1). Binding specificity was examined using SNX27 as a negative control. SNX27 is formed by the domains PDZ-PX-FERM, without BAR domain. Therefore, it should not bind SNX5. SNX27 was purified with the His-SUMO tag, like SNX1, and its ability to bind to SNX5 was tested. As it can be seen in Figure 35B (lane 7), SNX5 was unable to bind to SNX27. This negative control confirms that SNX5 is recruited by SNX1, and not by unspecific binding to His-SUMO tag or the Ni-NTA resin.

5.2.6. Simultaneous mutations in SNX1 and SNX5 disrupt SNX1/SNX5 heterodimer *in vitro*

Next, we attempted to confirm the key residues for SNX1/SNX5 heterodimer formation. SEC-MALS analysis shows that SNX1 homodimer is disrupted by four point mutations (F347A + W511A, two mutations per each SNX1 that forms the homodimer). However, the heterodimer-disrupting mutant versions could not be analysed by SEC-MALS, as the sample was not monodisperse. To overcome this problem, pull-down assays were performed.

Five pull-downs were conducted, and their results are shown in Figure 35: inputs (Figure 35A) and "pull-down" samples, after incubation and washing steps (Figure 35B). In all the samples, SNX1 and SNX27 were His-tagged (bound to the Ni-NTA

resin), and SNX5 was not tagged (washed out unless bound to SNX1 or SNX27). SNX5 binding to His-tagged SNX1 was evaluated using SDS-PAGE. Wild-type SNX5 was able to displace SNX1 homodimer (Figure 35B, lane 1). The double mutation that disrupts SNX1 homodimer formation (Phe347A + Trp511A) did not impair SNX1/SNX5 heterodimer formation (Figure 35B, lane 3). Then, six mutations were introduced in SNX5. However, these proteins ($SNX5_{Y219A+M233A+V240A+R368A+I398A+F401A,}$ also called $SNX5_{6xmut}$) were still able to bind to wild-type SNX1 (Figure 35B, lane 2). Only the combination of the SNX1 double mutant and SNX5 sextuple mutant could disrupt the heterodimer formation (Figure 35, lane 4). Note that the combination of SNX1 double mutant and a SNX5 quadruple mutant ($SNX5_{Y219A+M233A+V240A+R368A}$, or $SNX5_{4xmut}$) was not sufficient to impede heterodimerization. As control, we used His-SUMO-SNX27 which lacks a BAR domain and showed that, under the same conditions, it cannot bind SNX5.

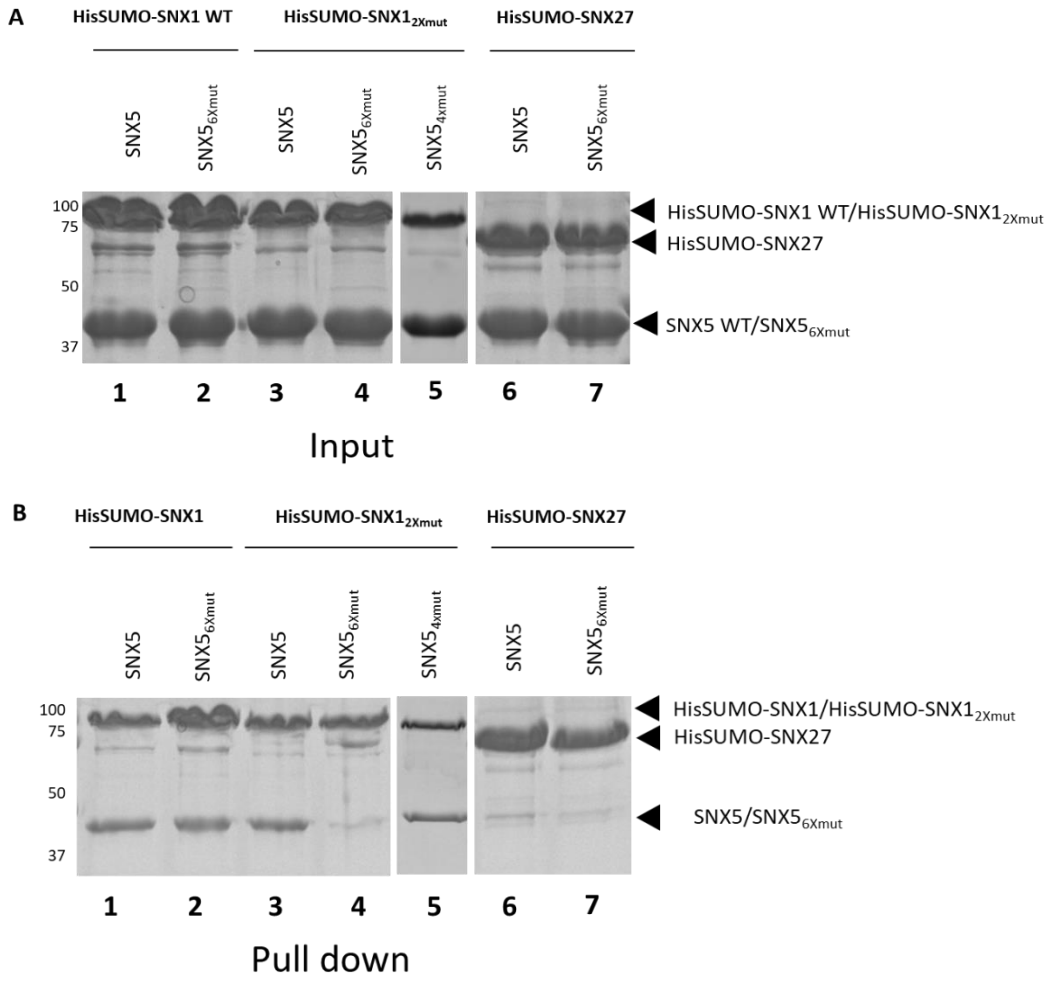


Figure 35: Competition assay between SNX1 and double mutants versus SNX5 and its mutants. (A) SDS-PAGE of samples before competition assay (input). (B) SDS-PAGE

after incubation of proteins and washing of the resin (pull-down). His-SUMO SNX1, either WT or mutant, and His-SUMO SNX27 were captured in the pull-down resin (Ni-NTA). SNX5 (WT or mutants) were incubated with His-tagged SNX1 (or SNX27 as negative control), washed, and the bound proteins were analysed by SDS-PAGE.

To confirm that structure-based point mutations introduced in SNX1 and SNX5 do not disturb the general fold of the proteins which could contribute to the observed binding defects, we used circular dichroism (CD) to compare the spectra of wild-type and mutant proteins. The spectra of SNX1, $SNX1_{F347A,}$ $SNX1_{W511A}$ and $SNX1_{F347A+W511A}$ (Figure 36A) were coincident, indicating that the proportions of secondary structures (α-helices, β-sheets and random coil) were the same. The CD spectra of SNX5, $SNX5_{Y219A+M233A+V240A+R368A}$ and $SNX5_{Y219A+M233A+V240A+R368A+I398A+F401A}$ also matched (Figure 36B), demonstrating that the mutations did not modify the relative content of secondary structures. Hence, we can conclude that the mutations do not alter the protein folding.

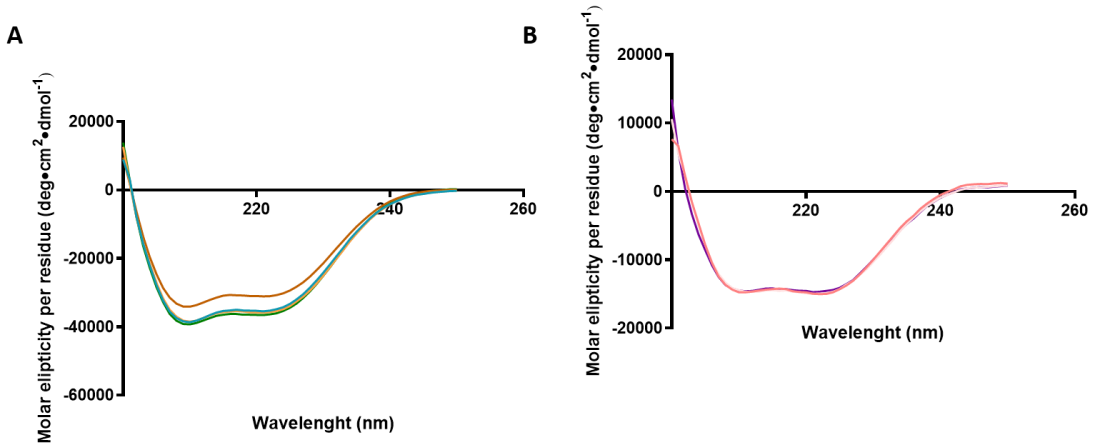


Figure 36: CD spectra of SNX1 (A) and SNX5 (B) and their mutants. (A), WT SNX1 (blue line), $SNX1_{F347A}$ (light brown), $SNX1_{W511A}$ (dark brown) and $SNX1_{F347A+W511A}$ (green). (B) WT SNX5 (salmon-pink), $SNX5_{Y219A+M233A+V240A+R368A}$ (pink) and $SNX5_{Y219A+M233A+V240A+R368A+I398A+F401A}$ (purple).

5.3. Interaction of SNXs with the cation-independent mannose-6-phosphate receptor (CIMPR) cargo

During the development of this project, two new studies reported cargo recycling by SNX1/2 or SNX5/6 heterodimers through retromer independent pathways[76,93]. However, the studies did not analyse the interaction in the context of the full-length heterodimer. The analysis with the cargo was conducted using isothermal titration calorimetry (ITC). To confirm the interaction, CIMPR was injected over the sample cell containing SNX1/SNX5 FL heterodimer or fragments that included the SNX1 PX domain, the SNX5 PX domain, or the SNX1/SNX5 BAR domains. CIMPR and the full-length SNX1/SNX5 heterodimer showed the most robust binding, with an equilibrium dissociation constant K_d of 2-µM (Figure 37A). The construct $CIMPR_{2372-2491,}$ which retained the His-SUMO tag to allow its quantification and to enhance its stability, showed no interaction with the SNX1/SNX5 heterodimer. When CIMPR was titrated against SNX1/SNX5 BAR domains, no binding was detected (Figure 37D). No binding was seen for the PX domain of SNX1 either (Figure 37E). The interaction of CIMPR with SNX1/SNX5 heterodimer was observed at the PX domain of SNX5 (Figure 37C). The affinity of CIMPR to full-length SNX5 (10 µM, Figure 37B) was similar to that for the PX domain of SNX5 (9.2 µM, Figure 37C), indicating that the interaction is restricted to the PX domain.

The signal sequence for the interaction of CIMPR with SNX1/SNX5 (or the previously reported interaction with retromer[191,118]) is formed by residues Trp2369, Leu2370 and Met2371 (WLM). To check whether those residues are implicated in the detected interaction, they were mutated to alanine (WLM → AAA), and the effect of this mutation on the interaction with SNX1/SNX5 was examined using ITC. The mutation abolished this interaction with the wild-type SNX1/SNX5 complex, confirming that these residues play a central role in the interaction (Figure 37G). Importantly, the affinity for the full-length heterodimer was six times higher (Figure 37A or B) that that of the PX domain alone. This observation suggests that another region might cooperate in the binding although it was not detected (Figures 37C and D).

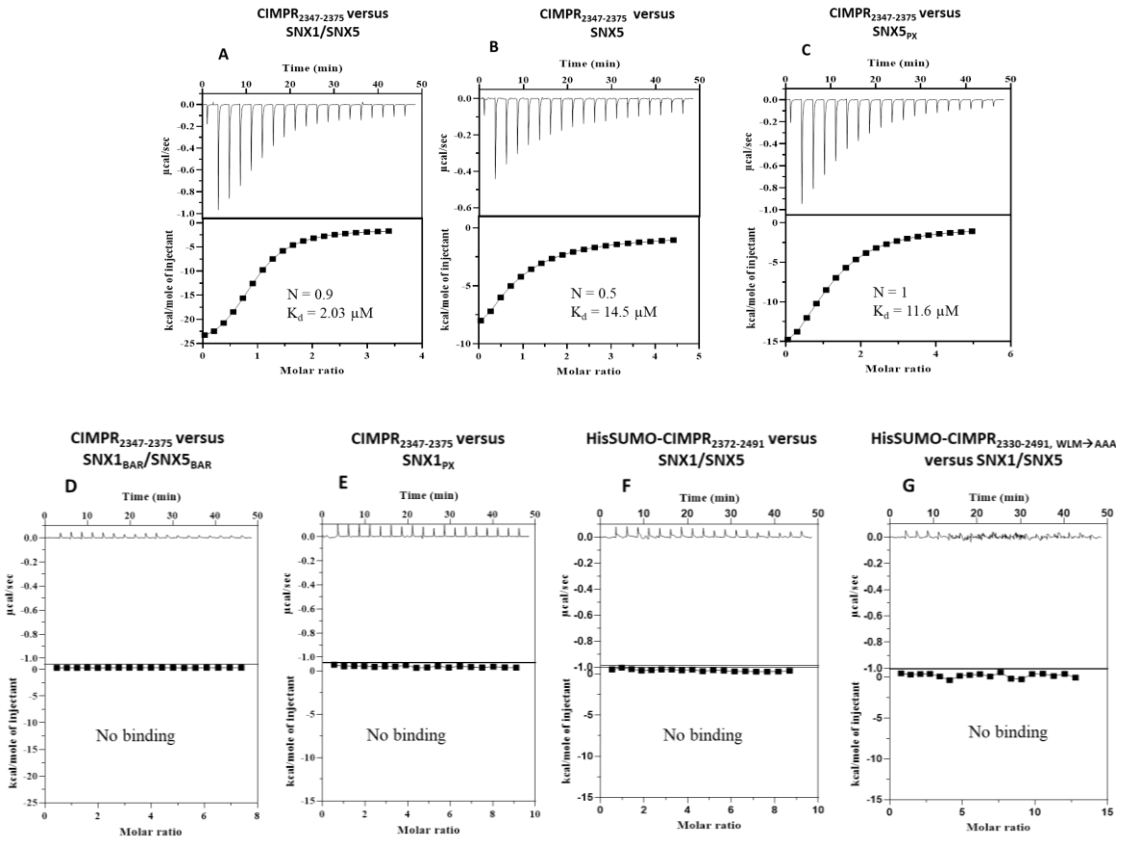


Figure 37: ITC examination of interactions between CIMPR and SNX1/SNX5. (A), 226 µM $CIMPR_{2347-2375}$ was titrated against 13 µM SNX1/SNX5 FL. (B) 330 µM $CIMPR_{2347-2375}$ was titrated against 15 µM SNX5 FL. (C) 418 µM of $SNX5_{PX\ was}$ titrated against 16 µM $CIMPR_{2347-2375}$. (D) 805 µM $CIMPR_{2347-2375}$ was titrated against 20 µM of $SNX1^{BAR}/SNX5^{BAR}$. (E) 746 µM of $SNX1^{PX}$ was titrated against 16 µM $CIMPR_{2347-2375.}$ (F) 350 µM His-SUMO-$CIMPR_{2372-2491}$ was titrated 22 µM SNX1/SNX5. (G) 350 µM $CIMPR_{2330-2491}$ (WLM→AAA) was titrated against 19µM SNX1/SNX5. No binding was detected in D to G titrations.

Table 12: Thermodynamics parameters of interactions of CIMPR with different SNXs and their subunits obtained using ITC. Results are the averages from at least three independent experiments. N.B. means no binding.

	[protein] in syringe (μM)	[protein] in cell (μM)	K_d (μM)	n	ΔH (cal/mol)	- TΔS (cal/mol)	ΔG (kcal/mol)
$CIMPR_{(2347-2375)}$ vs SNX1/SNX5 FL	226	13	2.19 ± 1 (4)	0.79 ± 0.05	$-3.05x10^4 \pm 4.16x10^3$	22.7 ± 4.24	-7.73 ± 33.2
$CIMPR_{(2347-2375)}$ vs SNX1/SNX5 BAR	805	20	N.B.				
SNX1 PX vs $CIMPR_{(2330-2471)}$	746	16	N.B.				
SNX5 PX vs $CIMPR_{(2330-2471)}$	418	16	11.75 ± 0.21 (3)	1	$-2.5x10^4 \pm 1.7x10^3$	18.2 ± 1.62	-6.87 ± 0.21
$CIMPR_{(2347-2375)}$ vs SNX5 FL	330	15	11.4 ± 3.75 (3)	0.49 ± 0.09	$-2.16x10^4 \pm 4.27x10^3$	14.9 ± 0.004	-6.8 ± 0.19
$CIMPR_{(2372-2491)}$ vs SNX1/SNX5 FL	498	22	N. B.				
$CIMPR_{(2330-2491,WLM \rightarrow AAA)}$ vs SNX1/SNX5 FL	454	19	N. B.				

5.3.1. The cytosolic tail of CIMPR is recruited to LUVs in the presence of SNX1 and SNX1/SNX5, but not SNX2

The full-length cytosolic tail of CIMPR (2330–2491) was purified following the His-SUMO strategy (see section 4.3.1). Then, its ability to interact with membranes, on its own or in the presence of SNXs, was evaluated and the results are shown in figure 39. $CIMPR_{2330-2491}$ alone did not interact with membranes (Figure 39). As expected, it was recruited by SNX1/SNX5. Surprisingly, $CIMPR_{2330-2491}$ was also recruited by SNX1, by both dimeric and monomeric SNX1 double mutants F347A+W511A. However, ITC did not detect any interaction between SNX1 and $CIMPR_{2330-2491}$ in solution (data not shown). This might have been a result of a conformational change or other interactions occurring in the presence of membranes. As SNX5 could not recruit itself to the membranes (as shown in section 5.3.2), it would not recruit $CIMPR_{2330-2491}$ either. Even more strikingly, SNX2, although it interacts with the membranes, could not recruit $CIMPR_{2330-2491}$. This suggests that SNX1 and SNX2 could have independent functions in CIMPR recycling.

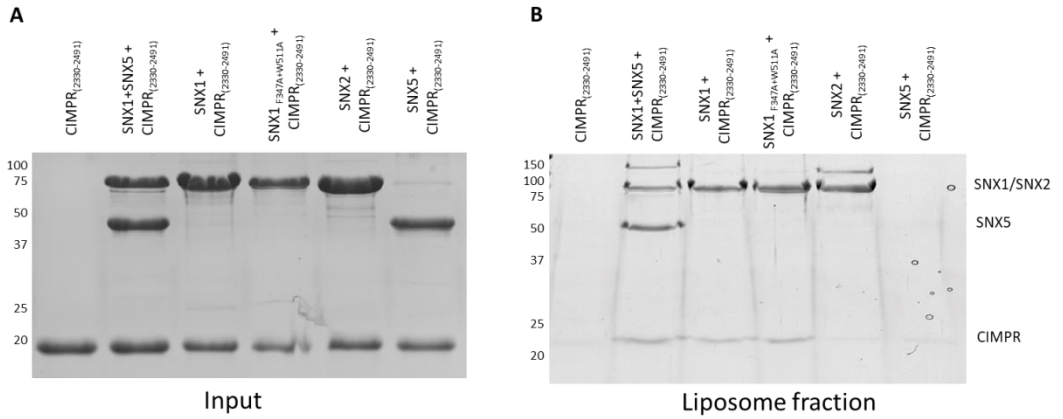


Figure 39: Recruitment of $CIMPR_{2330-2491}$ to membranes by SNXs *in vitro*. (A) Input samples used in the experiment. (B) Liposome fraction after coflotation, showing which protein interacts with LUVs and recruits $CIMPR_{2330-2491}$.

5.3.2 PX domains of SNX1 or SNX5 do not interact with LUVs *in vitro*

The PX domains are expected to interact with phosphatidylinositol phosphates (PIPs)[20]. However, it has been questioned whether all proteins with PX domains share this feature[20, 123]. We decided to examine the ability of PX domains of SNX1 and SNX5 to bind different phosphoinositides. Coflotation analysis of purified PX domain of SNX1 and full-length SNX5 with liposomes containing a screen of phosphoinositides was performed (Figure 38). SNX3 was used as a positive control of binding with PI(3)P-containing liposomes[123].

Neither SNX5 FL nor PX domain of SNX1 were able to bind strongly to phosphoinositide-containing liposomes (Figure 38). The low intensity of the bands of SNX5 seen on the gels (Figure 38B) might have been caused by the BAR domain, which likely associates through electrostatic interactions with the acidic lipids. Surprisingly, no differences in binding to various phosphoinositides were observed. Despite the lack of a clear association with any PIP we decided to keep PI(3)P in our liposome compositions for consistency with other assays.

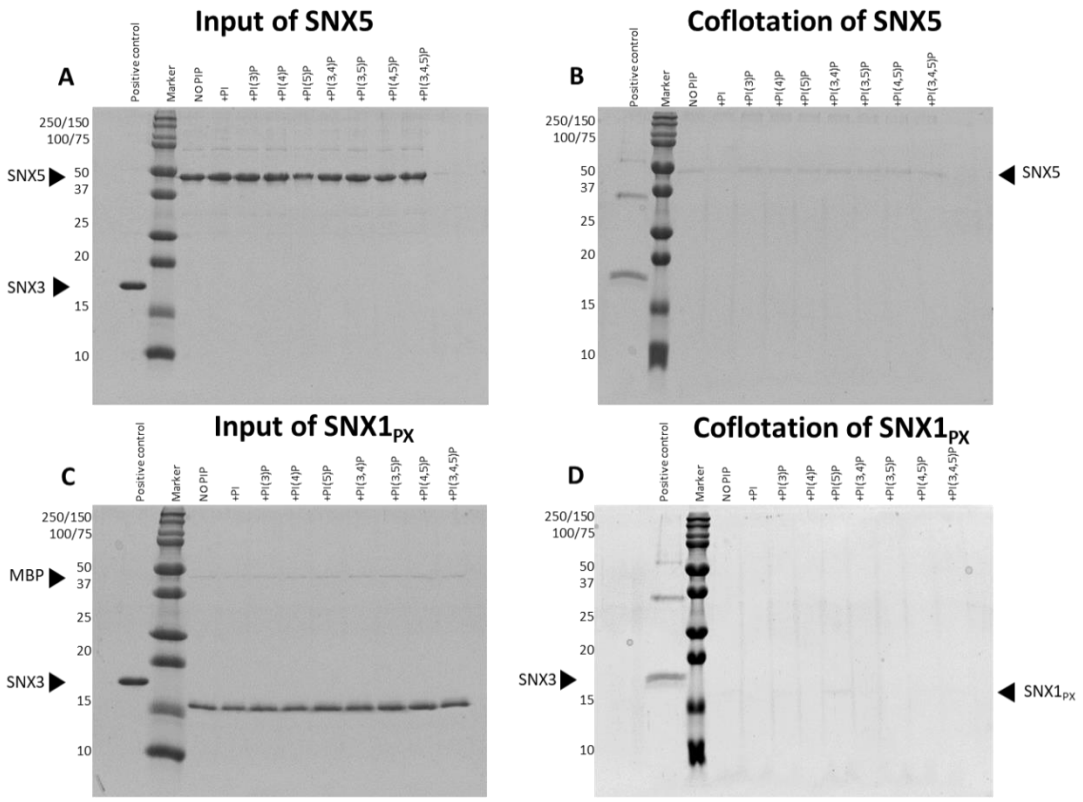


Figure 38: Coflotation analysis of SNX1 PX and SNX5 FL with different PIPs. (A) SNX5 FL input samples. (B) Results of coflotation of the PX domain of SNX5 FL with liposomes containing different PIPs. (C) Input samples for the PX domain of SNX1. (D) Results of coflotation of the PX domain of SNX1 with liposomes containing different PIPs.

5.3.3 Tubulation of LUVs by SNX *in vitro*

Several experiments of membrane tubulation by SNXs have been already reported[106]. However, the heterodimers have never been evaluated. We compared the tubulation ability of SNX1, SNX5 and SNX1/SNX5 heterodimer and studied the contribution of the cargo (CIMPR) to this activity. The tubulation was performed *in vitro* using liposomes and monitored employing cryoEM. The liposomes were incubated with different SNXs, and the percentage of liposomes forming the tubes was registered. These experiments were performed three times.

As can be seen in Figure 40, not all SNXs have the same tubulation capacity. SNX1 gives better results than any other SNX tested (Figure 40A and F), producing tubules in the 88% of the liposomes. However, the double mutation (making it monomeric) reduces its ability to tubulate by half: only 41% of the liposomes were

tubulated (Figure 40E and F). This indicates that SNX1 dimerisation is important for tubulation activity.

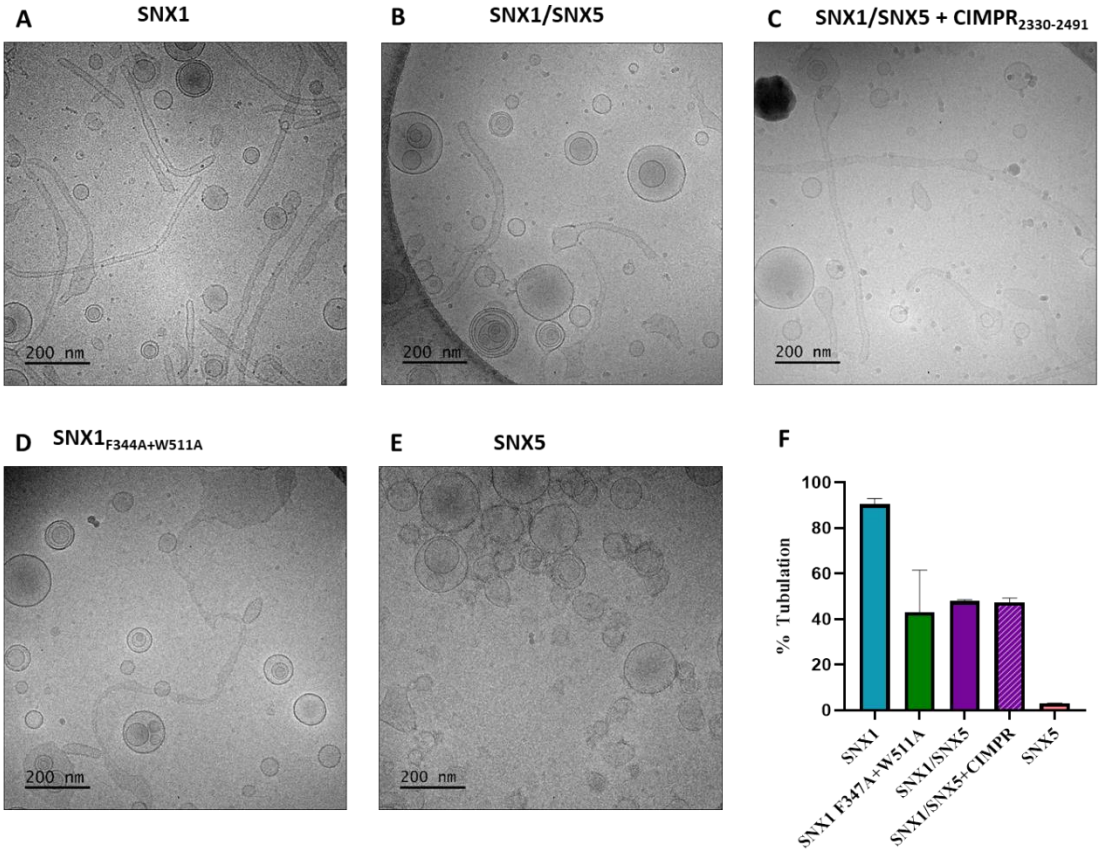


Figure 40: Tubulation using different SNXs. Example micrographs of vitrified liposomes incubated with 10 µM SNX1 (A), SNX1/SNX5 (B), SNX1/SNX5 preincubated with 20 µM $CIMPR_{2330-2491}$ (C), $SNX1_{F344A+W511A}$ (D) and SNX5 (E). (F) Bar graph of tubulation in each sample. The results are the averages of three measurements.

As reported before, we observed that SNX5 did not produce tubes *in vitro* (Figure 40D). SNX1/SNX5 heterodimer made the tubules from liposomes *in vitro*, although the efficiency was lower than for wild-type SNX1 (Figures 40B and F). Indeed, heterodimer tubulation efficiency was similar to the monomeric SNX1 ($SNX1_{F347+W511A}$) (Figure 40F). While SNX1 induces tubulation in 88% of the liposomes, SNX1/SNX5 only tubulates 47%. Addition of the cytosolic tail of CIMPR did not improve the tubulation capacity of this heterodimer (Figure 40C). However, it should be taken into account that CIMPR was not inserted into the membrane as it would be the case in the cellular context. Instead, it was recruited by SNX1/SNX5. To

evaluate a potential increment of tubulation associated with the SNX1/SNX5 recruitment by CIMPR, we believe that the cargo should be incorporated into the membrane.

5.4. Reconstitution of SNX and CIMPR on membranes *in vitro*

Biological activity of SNXs implies its interaction with membranes. Establishing the mechanisms of SNX assembly and cargo distribution on the membrane is crucial for the understanding cargo recycling. Here, we used artificial membrane systems to evaluate the distribution of SNXs.

We initially attempted to obtain the structure using highly homogenous membrane tubules formed by galactosylceramide (GalCer)[173]. Our initial attempts focused on SNX1 decorated tubules and later on SNX1/SNX5 coated tubules. A single-particle dataset was obtained using a Titan Krios microscope at Diamond Light Source. We expected to encounter an helical arrangement similar to other BAR proteins[134,145]. However, no helical diffraction pattern was detected by any of the programmes used (Relion[192], Spring[193]). Therefore, given the pleomorphic nature of the tubules, we decided to skip the use of preassembled tubes and instead use cryo-electron tomography (cryoET) for the analysis of tubules generated from large unilamellar vesicles (LUVs). Various conditions for obtaining tubules from liposomes using SNX heterodimer in complex with CIMPR were tested (see Table 9 in methods section). The structural arrangement of SNX1/SNX5 on tubes was examined by cryo-electron tomography (cryoET). The cryoET is a versatile technique that allows the resolution of heterogeneous/pleomorphic structures[92]. Data processing by averaging regions of the tomograms, known as sub-tomogram averaging, increases the signal-to-noise of repetitive structures allowing 3D reconstructions usually in the range of sub-nanometre resolution.

Initial averaging of tube segments revealed helices around the tubes with different number of helical starts depending on the tubes. Alignment and averaging of sub-volumes lead to an average that includes three clearly visible particles corresponding to three SNX1/SNX5 heterodimers. The estimated resolution is 9 Å, according to Relion[192] (Figure 41A). The structure is asymmetrical because the PX domains have different sizes. Figure 41 shows one heterodimer subunit, with the PX domain of SNX1 in the

upper part and the PX domain of SNX5 in the lower region. Fitting of different domains inside the volume has an average correlation of 0.83 (0.94 for BAR domains, 0.8 for $SNX1^{PX}$ and 0.74 for $SNX5^{PX}$), according to Chimera analysis (Figure 41B). Even though SNX1 is bigger than SNX5, the structured region of SNX1 PX domain is smaller. SNX1 has 142 amino acids at the N-terminus predicted to be unstructured and therefore unlikely to be seen by cryoET. The PX domain of SNX5 has an extended helical hairpin absent in other PX domains[121] (except for the closely related SNX6 and SNX32[77]). Simonetti and colleagues have shown that CIMPR interacts with the PX domain of SNX5 through this extended region[77]. Unfortunately, at the current resolution the cargo binding mode is not visible even though we have proven its presence in co-flotation assays.

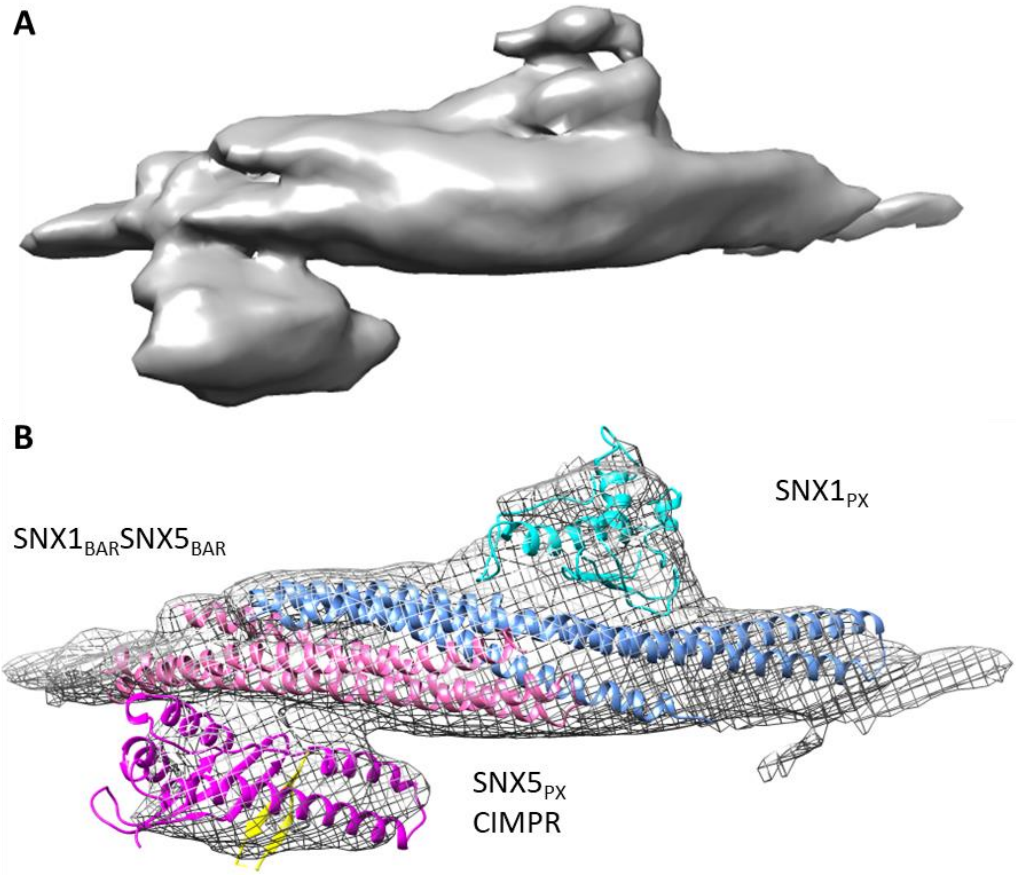


Figure 41: CryoET volume of a single SNX1/SNX5 heterodimer. (A) The volume of SNX1/SNX5 heterodimer. (B) Mesh representation of the same volume with the available 3D structures fitted in. In magenta: the PX domain of SNX5 is shown in complex with CIMPR (yellow, PDB code: 6N5X). In cyan: the PX domain of SNX1 (PDB code: 2I4K). The structures solved in this project: the BAR domains of SNX1 (in blue) and SNX5 (in pink). Images were prepared using UCSF Chimera[187].

The overall distribution of SNXs along the tube does not follow an arrangement with helical symmetry parameters, but instead exhibits a pseudo-helical distribution. Indeed, tubes can accommodate different numbers of helices (Figure 42A). As such, despite the radius of the membrane tube and the pitch remained similar, there are tubes with one, two, three and four helices (Figure 42B and C, respectively). The average radius is 13. 1 ± 1.0 nm and the average pitch is 7.3 ± 0.3 nm. This constant helical pitch indicates that all heterodimers reached the closest lateral distance and therefore the densest packing.

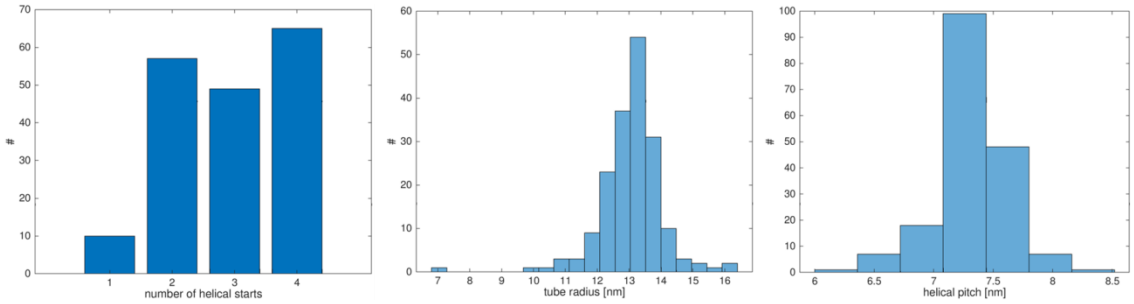


Figure 42: Helix parameters. (A) The number of helices in each tube. One to four helices can be found in the tubes, although the parameters do not change drastically. (B) The radius of the lumen of the membrane tubule. The average radius is 13.1 nm. (C) Pitch of different helices. The main rise is 7.3 nm, ranging between 6.5 and 8 nm.

Even though all tubes have a similar helical pitch and radius, the number of helical starts differs between the tubes. Tubes with 2, 3 and 4 helical starts are equally common while tubes with 1 start represent less than 10% of the occurrence. Since the helical pitch and radius are constant, the helix angle must differ between helices with different numbers of starts. This implies a certain flexibility of the lateral contacts of the heterodimers. A certain flexibility of the protein lattice is expected, since it allows the tubes to accommodate full-sized cargo proteins. The lattice geometry of the particles on the tube surface is revealed by placing a low-resolution average of the heterodimer at the corresponding coordinates of the sub-volumes. The helical arrangement of the particles is confirmed and potential tip-to-tip and lateral contacts are visible. As observed in Figure 43, tip-to-tip contacts between BAR domains (Figure 43B, blue circles) and lateral contacts between PX domains (yellow circles) and between BAR and PX domains (orange circles) stabilise the lattice of the tube.

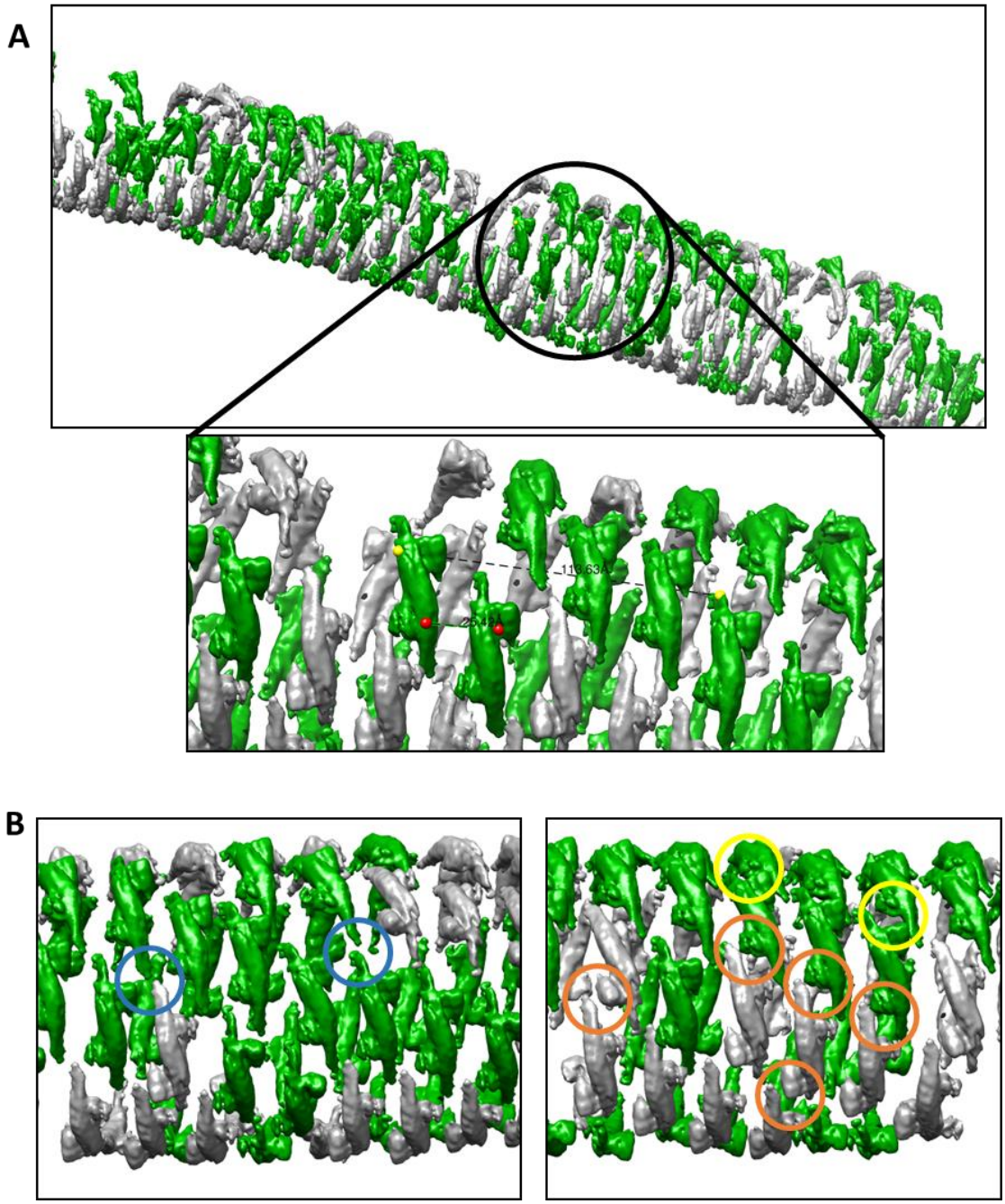


Figure 43: Overall structure of SNX on a tube. (A) This particular tube has 4 helices, with 7 units per turn, a rise of 25 Å and a pitch of 113 Å. SNXs are arranged in the tube with different orientations (top); one orientation is shown in green and the other in grey. Below, there is an inset picturing the area circled in black. The pitch was measured between the subunits marked with yellow dots. The rise is indicated with red dots and green dashed line, and the pitch, with yellow markers and black dashed line. (B) Tip-to-tip contacts (left, blue circles), PX-to-PX (right, yellow circles) and PX-to-BAR (left, orange circles) lateral contacts are observed in the tube. Images were prepared using the UCSF Chimera[187].

6. DISCUSSION

6. DISCUSSION

Eukaryotic cells are characterised by the segregation of their biochemical process into separated membrane-bound compartments called organelles. Their composition is controlled by the transit of biomolecules through transport vesicles. One of the main sorting compartments of vesicles in the cell is the endosome. There, a myriad of transmembrane proteins, also called cargos, can be sent for lysosome degradation or they can be recycled back to their original organelle. Cargo selection for degradation or recycling is fine-tuned by large protein complexes.

The Endosomal Sorting Complex Required for Transport (ESCRT) recruits ubiquitinated cargos for sorting into intra-luminal vesicles and finally to lysosomes for degradation[73]. On the other hand, other multi-protein complexes govern each one of the steps of cargo recycling. Among them, retromer was the first discovered, and it has been extensively studied since its identification in 1998[89]. It regulates cargo recycling from the endosomes to the plasma membrane and *trans*-Golgi network in a system of intracellular membrane transport called the tubular-endosomal network, or TEN.. In addition, two novel complexes involved in protein recycling, termed retriever and CCC complex, have been discovered recently[36,37]. Tubular carriers are optimal for the transport of transmembrane cargos due to the high surface-to-volume ratio, a geometry that aids in cargo transport preventing excessive exchange of intra-luminal content[137]. However, this tubular shape is not thermodynamically favourable, and it needs stabilization induced by several factors. Some proteins, like sorting nexins, present a BAR domain that sense and induce membrane curvature[139]. This domain has a coiled-coil structure composed of three α-helixes that can dimerise with themselves and other BAR domains. They form homo- or heterodimers with a positively charged concave surface that can bind to highly curved membranes and, in some cases, induce tubulation[139, 106] (Figure 38, results section).

Among the BAR domain-containing proteins, the sorting nexins (SNXs) are involved in membrane trafficking events. There are 33 known SNX in mammals, and twelve of them contain a C-terminal BAR domain. While some SNX-BAR such as SNX8, SNX9 and SNX33 are implicated in endocytosis and dynamin recruitment via an additional SH3 domain[182], other SNX-BARs like SNX4, SNX7 and SNX30 have been associated with autophagy. These proteins were able to form homo- and heterodimers between them.[106,194].

Five SNX-BAR proteins, SNX1, SNX2, SNX5, SNX6 and SNX32 (Figure 7, introduction section), are related to cargo retrieval in mammals and have been extensively studied in the last

years[137,106,76,77]. They evolve from the yeast ancestors, Vps5 (SNX1 and SNX2) and Vps17 (SNX5, SNX6 and SNX32), although it is not clear whether they are orthologues or not[17]. They can form homo- or heterodimers with a specific oligomerisation pattern: SNX1 or SNX2 with SNX5, SNX6 or SNX32[137,106] and participate in membrane deformation[106] as well as in cargo recognition[76,77]. Previous work carried by Simonetti and co-workers showed that heterodimers formed by SNX1/2 coupled with SNX5/6, are involved in retromer-independent cargo recycling routes[11]. They characterised the interaction between the model cargo CIMPR and the SNX1/SNX5 heterodimer, and narrowed the binding to the PX domain of SNX5. The complex was defined as the "endosomal sorting complex for promoting exit-1", or ESCPE-1[12].

Our knowledge about how BAR domain-containing proteins interact with membranes and how tubules are formed is scarce. There are few studies describing tubule formation by BAR domain-containing proteins[139,145,195,92]. From these studies it was deduced that protein interactions could stabilise the lattice/coat via tip-to-tip contacts between neighbouring BAR domains, as well as lateral contacts[145]. For example, the interaction of F-BAR domains with the membrane showed how they organise in a helical lattice through tip-to-tip and extensive lateral contacts between them, which are crucial for membrane deformation[134]. This study was later extended to endophilin, a N-BAR containing protein involved in dynamin recuitment[145]. The study showed that an amphipathic helix, characteristic of N-BAR domain-containing proteins[141], was essential for membrane remodelling. Rather than inserting in the membrane, the amphipathic helix of one protein interacts with neighbouring endophilin amphipathic helixes in a promiscuous manner, stabilising the helical lattice. Another study also showed that some BAR domain-containing proteins, like the BAR-PH ACAP1, needed the presence of an auxiliary domain (the PH domain) to bind to the membrane and induce tubulation[195]. Nonetheless, there is currently a big gap in understanding how SNX-BAR proteins concomitantly interact with membranes, recruit cargo and contribute to tubulation.

Accumulating evidence indicates that SNX-BAR heterodimers of the ESCPE-1 contribute to cargo recruitment and membrane tubulation through four key factors: (1) the BAR domain heterodimerisation; (2) the interaction of the PX domain with PIP-containing membranes; (3) the interaction of the PX domain with the model cargo CIMPR; and (4) the formation of SNX lattices through tip-to-tip and lateral contacts. We aimed to structurally and biochemically characterise SNX heterodimers and find out how they coordinate membrane deformation with cargo selection. In this work, we have characterized the BAR dimerisation interface of the SNX1/SNX5 heterodimer. We also have extended the analysis of how heterodimers contribute

to CIMPR recognition. And finally, we have characterised the SNX1/SNX5 membrane coat by cryo-electron tomography (cryoET).

6.1. The BAR heterodimers

The underlying mechanisms for BAR heterodimerisation have been barely studied. Previous research found that SNX1 can form heterodimers with SNX5 using co-immunoprecipitation assays[125,127]; however, an assay showing a direct interaction was missing[106]. The use of co-immunoprecipitation assays allows the detection of protein associations in vivo. However, co-immunoprecipitation is unable to prove direct protein-protein interactions. Given the intrinsic difficulties in the production of PX-BAR heterodimers, the study of these complexes under controllable conditions with individual components has been a challenge for many years.

Our first goal was the purification of SNX1 and SNX5. On one side, SNX1 tended to degrade, but optimization of the purification method allowed us to obtain a homogeneous sample and a relatively good yield (Figure 20, results section). On the other side, SNX5 was very unstable by itself and prone to degrade and form inclusion bodies. The purification required the inclusion of several size exclusion chromatography steps to achieve a homogeneous population (Figure 21, results section). Despite being able to obtain the individual subunits, we next attempted the purification of the SNX1/SNX5 heterodimer. We were able to co-express and purify SNX1/SNX5 heterodimer with a good yield (about 1 mg of pure protein per litre of culture). The purified heterodimer had a one-to-one ratio. Furthermore, we observed that SNX1 precludes the tendency of SNX5 to aggregate (Figure 22, results section). These observations are indicative that SNX5 is unstable by itself and most likely exists as a complex in mammalian cells. Indeed, studies performed with SNX5 shows that it has several possible binding partners (apart from SNXs) liked to various functions[73]. It is worth noting that the expression and purification of SNX1/SNX6 heterodimer have been recently achieved following a similar protocol[128], thus proving this method as an standard for purification of SNX-BAR heterodimers from the ESCPE-1 coat complex.

Previous studies have indicated that most SNX-BARs behave as dimers[106,125,101] although it has been suggested that SNX5 could behave as a monomer[106]. Nonetheless, the oligomeric states were determined by co-immunoprecipitation assays, which give an indication of protein partners in a cellular context but do not demonstrate direct interactions and neither allow the

quantification of the oligomeric state of complexes. Hence, we decided to characterise the oligomeric state of the three SNXs studied in this project using SEC-MALS: SNX1, SNX5, and the SNX1/SNX5 heterodimer. We found that both SNX1 and SNX1/SNX5 behave as dimers in solution at physiological ionic strength (Figure 23C and E respectively, results section). We have proved that SNX5 behaves as a monomer in solution (Figure 23D, results section). Our results confirm the oligomeric state of these proteins, as our analysis was performed from heterologously expressed and purified proteins. This result, combined with previous results of the tubulation ability of each of the proteins (where SNX1 and SNX2 can tubulate whereas SNX5 cannot)[106], indicates that dimerisation is one of the key factors for SNX-BAR-mediated membrane remodelling activity.

Later results in the ESCPE-1 coat prompted us to analyse the SNX1/SNX5 heterodimer from a structural point of view. Retromer recycling routes have been related with severe neurodegenerative diseases[79]. Then, it is important to understand the mechanisms that govern cargo recycling in order to treat the disorders that alter it. Structural information is a highly valuable resource important to understand mechanisms of action and also for drug development. We used X-ray crystallography as it allows obtaining a high-resolution structure of proteins in the range of the size of the heterodimer (105 kDa). NMR techniques are, in general, limited to proteins smaller than about 40 kDa[196]. Current high-resolution cryoEM techniques are very challenging for proteins smaller than 150-200 kDa without distinguishable morphological features as they are hardly observed in the micrographs, making the particle picking very laborious[195]. Given that other PX-BAR structures such as those from SNX9 and SNX33 were solved previously by X-ray crystallography, our initial efforts were directed at obtaining protein crystals from the full-length SNX1/SNX5 heterodimer. Unfortunately, after testing more than 3000 different conditions, we did not obtain any crystal. We speculate that the SNX1/SNX5 heterodimer did not crystallize due to two possible reasons. First, the N-terminal region of SNX1 is structurally disordered. The first 140 amino acids of this protein do not belong to the PX domain and are predicted to have a random coil structure, which might impair crystal lattice formation[197]. According to this, we decided to make a construct lacking these 140 amino acids from the N-terminal region of SNX1. Despite not being able to obtain an homogeneous sample of the $SNX1_{141-522}$-SNX5 complex, we continued with the high-throughput crystallization screening hoping that the crystallization process might favour one subpopulation of particles as it has been previously described[198]. Unfortunately, this was not the case and after some unsuccessful efforts trying to improve the purification, we discontinued working with this construct. The second reason whereby SNX1/SNX5 heterodimer may not crystallize could be the presumably high mobility between the PX and BAR domains. Although, SNX9 and SNX33

have been crystallized with both the PX and BAR domains together (PDB codes 2RAI and 4AKV, respectively), the presence of a "Yoke domain" between the PX and BAR domains may increase the rigidity between each other. However, this Yoke domain is apparently absent in SNX1 and SNX5[199].

Finally, we decided to focus on the BAR domains that form the heterodimer as they represent the core of the interaction. Computational models of SNX5 helped us to design the constructs flanking the BAR domain. The prediction showed that the BAR domain starts at around Val200, so a construct from Asp195 was done. The construct for the BAR domain of SNX1 was inferred from the previous crystal structure[106]. The final construct of the BAR heterodimer was $SNX1_{301-522}/SNX5_{195-404}$. After high-throughput screening good-diffracting crystals were obtained. We obtained native crystals with $P2_12_12$ symmetry, with two screw axis. Molecular replacement methods against solve SNX-BAR proteins (SNX1 PDB code 4FZS, SNX9 PDB code 3DYT and SNX33 PDB code 4AKV) did not render good results. None of the programs used (Molrep from CCP4 suite[160] and Phaser from PHENIX suite[162]) could find the solution as the R-factor indicates that the result was spurious (close to 0.5). We tried to obtain the phase by the production of a selenomethionine derivative of the protein; however, it was not possible as the BAR domain of SNX5 completely degraded during purification. We then decided to obtain a heavy metal derivative crystal. We observed based on derivatisation webservers that Platinum could bind to $SNX1_{301-522}/SNX5_{195-404}$ in our crystallization conditions (pH 7). However, although crystals soaked with K_2PtBr_4 presented anomalous signal up to 3.7 Å, the phase could not be obtained by any dataset obtained using single anomalous dispersion (SAD) methods. Afterwards, we decided to use multiple anomalous dispersion method (MAD). It implied growing bigger crystals, for which micro-seeding crystallization approach were used (see methods for further information). Three datasets were collected at the platinum peak (1.0715 Å), inflexion (1.07293 Å) and remote (1.0447 Å) wavelengths. In order to obtain an adequate phase, the crystal symmetry was reduced (P2). There was a screw axis on the crystal that was not consistent in the whole crystal, hampering the model building. Finally, the program autoSHARP was able to recover the phase and build and initial model that allow structure refinement, as explained in detail in the results section. The model obtained by MAD was solved at 2.8 Å resolution. Then, the obtained phase was extended to native crystals, and the resolution could be reduced to 2.5 Å.

The solved structure was used to investigate the key residues for its interfacial stabilisation and to compare their contribution in homo- and heterodimer configurations. The previously

solved crystal structure of SNX1 is distorted because the angle between the BAR domains does not match with the curvature of a membrane[106]. Thus, it is conceivable that the energetic contribution of individual interface residues using the crystal structure of SNX1 might be biased. For this reason, an *in silico* homology model of SNX1 was obtained using our $SNX1_{301-522}$/$SNX5_{195-404}$ crystal structure as a template. Then, an *in silico* analysis of the interaction was done to identify the key residues that contribute to dimer stabilisation. *In silico* alanine scanning mutagenesis provided an indication of the residues that most contribute to the interfacial stabilization. MD simulations showed that SNX1/SNX5 heterodimers had lower binding energy than SNX1 homodimers indicating that heterodimers are more stable than homodimers. We found that the highest contribution to the binding free energy of SNX1 was from residues Phe347 and Trp511 (figure 29, results section). The computational prediction was then validated introducing point mutants. Mutation of these two amino acids was sufficient to break the SNX1 homodimer. These two residues are conserved from yeast to human (Figure 31, results section), including the SNX1 paralogue SNX2, indicating their importance. However, SNX2 dimer was not broken after alanine mutation of those residues (Phe344 and Trp509), indicating that SNX1 and SNX2 homodimers share similar but not identical pattern of interfacial interactions.

In silico alanine scanning of residues at the SNX5 interface showed several amino acids that have a high energetic contribution for heterodimer formation. Previous results showed that charged amino acids from SNX1 and SNX5 were responsible to prevent dimer formation[106]. Particularly Arg337 from SNX1 favour heterodimer formation because its charge is not neutralised in the homodimer. Second, they also claimed that Glu280 and Glu383 can impair SNX5 homodimer formation, based on a SNX5 homodimer model built from the hyper-curved SNX1 structure. However, it is not possible to deduce which amino acids impair SNX5 homodimer formation, as there is not a confident model. There are not hydrogen bonds between those amino acids, indicating that they are not responsible for heterodimer stabilisation. On the other hand, our MD analysis indicates that the amino acids that drive the heterodimerisation are mainly hydrophobic. Using the competition assay, it was demonstrated that mutations of up to eight amino acids were needed to break the heterodimer, two in SNX1 and six in SNX5. It indicates that heterodimers are preferred to SNX1 homodimers, where only four amino acids (two per chain) are sufficient to impede SNX1 homodimerisation.

The interfacial residues of the $SNX1_{301-522}$/$SNX5_{195-404}$ were aligned. Then, residues were analysed based on the energetic contribution, proximity (below 4.5 Å) in the heterodimer,

evolutionary conservation, electrostatic potential and similarity of the residues (Figure XX). Hydrophobic amino acids of both proteins contribute more energetically to the binding, as previously indicated. What is more, these amino acids are at short distances, allowing stacking interactions. There are only two charged amino acids (Lys325 and His381 from SNX1 versus Asp225 and Glu280 from SNX5, respectively) that stabilise the heterodimer with polar interactions. However, it can be seen that SNX5 presents more charged residues in the interface than SNX1 (17 versus 14 charged amino acids, respectively). Altogether, the results suggest that SNX1/SNX5 heterodimer is stabilised by hydrophobic interactions and that SNX5 may not homodimerise due to electrostatic repulsion in the hydrophobic environment formed by the BAR dimerisation. This result is supported by the fact that mutations of two hydrophobic residues in SNX1 are sufficient to break its homodimer.

To confirm that heterodimers are preferred against SNX1 homodimers, we decided to analyse it using a pull-down based competition assay (Figure 33, results section). The pull-down competition addressed the ability of SNX5 to displace SNX1 homodimers. SNX5, which behaves as a monomer in solution, displaced SNX1 homodimer to form a SNX1/SNX5 heterodimer, which is in agreement with the previously observed co-immunoprecipitation assays[106]. It suggests that SNX heterodimers are the preferential conformation in the cell. However, when mutations of the amino acids that contribute energetically more to the binding were introduced, the binding was lost. It indicates that we have found the key residues to maintain heterodimerisation.

The ability to form distinct SNX heterodimers might contribute to distinct functionalities. Indeed, the concept of context-specific functions associated with alternative subunit composition in protein complexes is widely present in nature. For instance, SNARE proteins allow membrane fusion by the heterodimerisation of different SNAREs present in both compartments. For example, syntaxin-6, a widely expressed Q-SNARE, can bind to more than ten different SNAREs, participating in different membrane fusion events across the cell[200].

6.2. PX interaction with the membrane

The PX domain is expected to bind to PIPs, specially PI(3)P, exemplified in the structure of the PX domain of $p40^{phox}$ in complex with PI(3)P[122]. However, a recent publication reported that the PX domains of SNX1 and SNX5 did not bind strongly to PI(3)P[123]. SNX1 and SNX5 are usually found at the endosomal level, where they were thought to be recruited through the

interaction of their PX domains with PIPs, mainly PI(3)P[201]. Hence, we decided to assay the interaction of PIP-containing liposomes with the PX domains of SNXs with a co-flotation assay to find out if the PX domain is implicated in membrane recruitment. Proteins that oligomerise or aggregate might have a tendency to be overrepresented in sedimentation assays. Liposome flotation assays, on the contrary, mimic closer the physiological conditions the ability of a protein to be recruited to a membrane of defined composition. We have shown that neither PX domain of SNX1 nor SNX5 strongly interacts with PIP containing membranes (Figure 36, results section). Despite these observations, it is evident that SNX1 and SNX1/SNX5 are found on membranes, as we observed by *in vitro* analysis of its tubulation capacity and others have found them on the endosomal surface inside cells[202]. It is reasonable to think that if the PX domain is not sufficient to be recruited by itself to the membrane, other regions may cooperatively contribute to the membrane association. The avidity effect may also be considered. On the one hand, the positively charged concave region of the BAR domains (Figure 26, results section) can bind to the negatively charged surface of the membranes. On the other hand, PX domains contribute to the binding in a lower level, but it may allow SNX clustering in the membrane. Hence, SNX-BAR binding to the membrane is driven mainly by electrostatic interaction rather than the interaction of the PX domains with specific phosphoinositides. Nevertheless, the PX domains interact with cargo therefore contributing to their recruitment to specific membrane locations and also play a role in the lattice/coat organization as discussed below.

6.3. Cargo recruitment by SNXs

Yeast retromer is a heteropentamer formed by two sub-complexes, the so-called “cargo selection complex” (CSC), which is formed by Vsp35-Vps26-Vps29; and the SNX complex, formed by Vps5 and Vps17[89]. In yeast, the VPS subcomplex is responsible for the recruitment of the CSC to the endosomes due to the interaction of their PX domain with PIP-containing membanes[203], while the yeast CSC recruits cargo through Vps35 and Vsp26. Recent research has found that the yeast orthologue of CIMPR, Vps10, is recycled via this heteropentamer through a bipartite signal present in Vps10[204].

However, mammal variants do not follow the same pattern, as mammal retromer does not form a stable complex with the equivalent SNX proteins. Furthermore, a couple of studies in 2017[76,93] demonstrated that CIMPR can be recycled by the heterodimers of SNX1 or SNX2 with SNX5 or SNX6 in a retromer independent manner [76,93]. Later, another study in 2019[77] showed that CIMPR is directly recognised by the PX domain of SNX5, SNX6 or SNX32. This study identified a bipartite sequence in CIMPR, formed by $VSYKYSK_{2349-2455}$ and $WLM_{2369-2371}$,

which interact with the PX domain of SNX5. They form a bipartite signal where VSYKYSK binds to the first alpha-helix following the proline-rich strand called α′ of SNX5, while WLM binds to the second alpha-helix, or α′′. This recognition sequence is also present in another known retromer cargo, Sema4C[77] and is similar to those found in the IncE protein of *C. trachomatis*[96,205]. Remarkably, Sema4C also binds to SNX27 through its PDZ domain. This dual association between SNX-BAR and SNX27 might allow the formation of a supercomplex of SNX-BAR-Sema4C-SNX27-retromer that directs cargo recycling to the plasma membrane[206]. Hence, it seems that SNX-BAR proteins might function in cargo recycling by themselves as the ESCPE-1 coat, or through the assembly of larger complexes. In this sense, even the most basic coat assembly represented by the SNX-BAR proteins has remained unexplored.

First, we focused on the interaction of the cytosolic region of CIMPR and the SNX1/SNX5 heterodimer by isothermal titration calorimetry (ITC). We initially used the whole cytosolic tail of CIMPR ($CIMPR_{2330-2491}$), because the interaction must happen in this area, as the other region is in the lumen of the endosomes. However, the yield of the purification of the $CIMPR_{2330-2491}$ was very low, so we also analysed a peptide encompassing the bipartite binding motif ($CIMPR_{2347-2375}$)[77]. We found no differences in affinity between SNX1/SNX5 and both CIMPR constructs. Hence, we decided to use the peptide in all the ITC experiments.

First, the affinity between $CIMPR_{2347-2375}$ and the full-length SNX1/SNX5 was determined. It has a high affinity (2.2 μM, figure 35 and table 12, results section) when compared with the affinity displayed by other retromer cargos: GLUT1 has an affinity of 154 μM with SNX27[132] (although it is increased up to 15 μM in presence of retromer subunit VPS26) and DMT1-II has an affinity of 146 μM for the complex SNX3-retromer[75]. The thermodynamic[207] of the reaction showed to be in part enthalpy driven. It indicates that the interaction between SNX5 and CIMPR is hydrophilic (i.e. electrostatic and hydrogen bonds, mainly). Nevertheless, there is an important contribution of entropy-driven interactions of hydrophobic nature (see table 12).

We observed that the affinity between CIMPR and full-length SNX1/SNX5 is 2.2 μM, which is about twelve times higher than that reported for the interaction between CIMPR and the PX domain of SNX5[77]. In order to address whether the cargo was recognized solely by the PX domain, we designed several constructs containing the various domain combinations of the SNX1/SNX5 heterodimer: $SNX5_{22-170}$ or $SNX5^{PX}$; $SNX1_{142-269}$ or $SNX1^{PX}$, and the crystallized $SNX1_{301-522}/SNX5_{195-404}$. Then, we initially evaluated the ability of each of them to interact with

CIMPR. We observed that CIMPR is directly recognised by the PX domain of SNX5, (Figure 35, results section), as a recent work has showed[77]. Obtained affinities were higher compared to the previous study (12 μM versus 25 μM[77]). In order to find out if the interaction between CIMPR and SNX5 was restricted to the PX domain of SNX5, we measured the affinity between the cargo and the full-length SNX5. We found no differences in affinity between the PX domain alone and full-length SNX5, indicating that the contribution of SNX5 for CIMPR recognition is restricted to its PX domain. We could not detect any interaction of CIMPR with $SNX1_{301-522}$/$SNX5_{195-404}$ or the PX domain of SNX1 ($SNX1_{142-269}$). According to our results, there is a gap between the affinity of CIMPR with the full-length SNX1/SNX5 heterodimer and the $SNX5^{PX}$. We hypothesize that another region might participate in the interaction, but so far we have not detected it by ITC.

Then, we decided to study CIMPR recruitment in the context of the membrane. We designed a coflotation assay to investigate the recruitment of soluble cytosolic tail of CIMPR ($CIMPR_{2330-2491}$) to membranes in the presence of different SNXs (Figure 37, results section). $CIMPR_{2330-2491}$ was efficiently recruited by SNX1/SNX5 heterodimer, as expected. Furthermore, the recruitment to the membrane was specific, as $CIMPR_{2330-2491}$ alone was not recruited to membranes. SNX5 could not recruit $CIMPR_{2330-2491}$ because it was not able to bind efficiently to the membrane itself under the conditions tested (Figure 36, results section). Surprisingly, SNX1 was able to recruit $CIMPR_{2330-2491}$ to membranes, albeit direct interaction in solution was not detected. Besides, the recruitment was not dependent on the dimerization state, as monomeric SNX1 ($SNX1_{F347A+W511A}$) could recruit $CIMPR_{2330-2491}$ as well. Nonetheless, it should be noted that $SNX1_{F347A+W511A,}$ despite being monomeric, still promotes tubule formation, albeit at a lower rate than wild-type, suggesting that is partially active (Figure 38, results section). These results suggest that SNX1 partially contributes to CIMPR recycling via direct interaction in the context of the membrane. What is even more surprising is the fact that a SNX1 orthologue, SNX2, was unable to recruit $CIMPR_{2330-2491}$ to the membrane, indicating a difference between these proteins. These results, combined with the differences in homodimerisation between SNX1 and SNX2, suggest that although both proteins can form heterodimers with at least SNX5 and SNX6, their interactions exhibit differences in cargo association in the context of membrane and each heterodimer could be responsible for the recycling of different specific cargos.

Once we observed that SNX-BAR interact with CIMPR in the presence of membranes, we wondered if cargo might aid in SNX-BAR tubulation activity. We analysed tubulation ability of the heterodimer in presence of cargo (Figure 38, results section), because previous studies

suggest that cargo may enhance SNX recruitment to the endosomal membrane[206,90]. However, we did not observe an increased activity, remaining its tubulation ability steady. It could be due to the fact that the CIMPR construct used ($CIMPR_{2330-2491}$) did not contain the transmembrane region, but only the cytosolic region. We tried to purify a construct including the transmembrane region of the cargo, but it was not achieved, as protein was insoluble and attempts to recover it from the inclusion bodies were unsuccessful (data not shown).

6.4. Molecular architecture of the ESCPE-1 coat

Interaction of BAR domain containing proteins with the membrane has been previously studied by cryoEM[134,145,195] and cryoET[92]. Initial experiments took advantage of the formation of helical lattices on the membrane to get high resolution structures[134,145,195]. In another case where the helical lattice was not observed but more pleomorphic architectures were formed, cryoET was used[92]. These studies analysed how BAR proteins are able to decorate and deform model membranes. In the case of F-BAR proteins like CIP4[134], the association with the membrane is mediated through positively charged amino acids[134]. The F-BAR lattice is stabilised by extensive tip-to-tip contacts and also lateral contacts between charged and hydrophobic amino acids from one BAR domain and the adjacent one. Interestingly, the lateral contacts regulate the diameter of the tubes, increasing their plasticity. Later, the analysis was extended to endophilin, a N-BAR containing protein involved in dynamin recuitment[145]. The study showed that an amphipathic helix, characteristic of N-BAR domain-containing proteins[141], was essential for membrane remodelling. Rather than inserting in the membrane, the amphipathic helix of one protein interacts with neighbouring endophilin amphipathic helixes in a promiscuous manner, stabilising the helical lattice. Another study also showed that some BAR domain-containing proteins, like the BAR-PH ACAP1, needed the presence of an auxiliary domain (the PH domain) to bind to the membrane and induce tubulation[195]. Of special interest to the current work is the Vps5 homodimer (yeast orthologue of SNX1) in complex with the retromer from the fungus *Chaetomium thermophilum*[92]. In this study, the authors found that Vps5 does not form a perfect helical array, but a pseudo-helical one. This observation agrees with previous *in silico* and *in vitro* studies where BAR proteins decorate tubular membranes in pseudo helical arrangements[208,209]. However, there are several aspects that have not been studied before. First, all the previous studies involved BAR homodimers, not heterodimers. Second, ESCPE-1 SNX-BARs can form membrane coats independently of retromer. Finally, none of the previous studies have included any model cargo, which is indeed the cornerstone of any transport event. Therefore, we decided to study how the mammalian SNX1/SNX5 heterodimer is able to coordinate membrane tubulation with cargo recruitment.

According to previous studies, BAR domains tend to decorate model membranes establishing helical or pseudo helical arrangements[134,145,195]. For this reason we initially attempted to use preformed membrane nanotubes, enriched with galactosyl ceramide, as a template for SNX-BAR assembly. However, we realized that the layer lanes present in the diffraction patterns in the Fourier space of helical samples (which can be clearly observed in other studies of BAR domain coats[195]) were not visible, indicating the absence of a helical lattice. The latest developments in cryo-ET such as the use of direct detection cameras with fast readout speeds and extremely low background noise combined with improved algorithms for sub-tomogram averaging, have greatly boosted the field for structure determination to sub-nanometre resolution. A clear example is the recent visualization of the yeast retromer coat assembled on membrane tubules[210,211]. For these reason, we decided to analyse the binding of SNX1/SNX5 heterodimers and cargo to liposomes and characterise the tubes generated by cryoET.

The tube radii exhibit minor variations between tubes. The average membrane-to-membrane radius over all analysed tubes is 13.1 ± 1 nm (Figure 40B, results section). This radius is similar to the membrane-to-membrane radio generated by the yeast Vps5-retromer coat (radius of 15.5 ± 3 nm) and in the range of those produced *in vivo* (20 – 50 nm)[212], indicating that the tubule geometry produced by SNX-BAR are consistent. The variability in the tubes produced *in vivo* might be caused by the presence of cargos of different sizes in the lumen of the tube. According to this data, a recent study has shown the cryoEM structure of the intraluminal part of CIMPR[213]. The study shows that CIMPR is more packaged at acidic pH (170 Å × 80 Å × 80 Å), which would fit in the tubes we obtained *in vitro* (262 Å diameter).

Same as for the tube radii, also the pitches of the helices formed by the heterodimers show little variation between tubes. The average helical pitch over all analysed tubes is 7.3 ± 0.3 nm. This constant helical pitch indicates that all heterodimers reached the closest lateral distance and therefore the densest packing.

Even though all tubes have a similar helical pitch and radius, the number of helical starts differs between tubes. Tubes with two, three and four helical starts are balanced (31%, 27% and 36%, respectively) while only 5% have one start (Figure 40A, results section). Since the helical pitch and radius are constant, the helix angle differs between helices with different numbers of starts. This implies a certain flexibility of the SNX arrangement on the membrane. This flexibility of the protein lattice is expected, since it also would allow the tubes to accommodate diverse full-sized cargo proteins in the lumen of the tubule. Furthermore, it can deal with different cargo densities, as the more helixes are present in a tube, the more cargo can be recruited to those tubes.

The distribution of SNX particles on the tube surface is revealed by placing a low-resolution average of the heterodimer at the corresponding coordinates of the sub-volumes (see Figure 39, results section). The pseudo-helical arrangement of the particles is confirmed and tip-to-tip and lateral contacts are visible. Furthermore, new contact types are observed along the tubes (see Figure 41, results section), like PX-to-PX contacts and also novel PX-to-BAR contacts. Similar contacts were also observed in the yeast SNX, Vps5. They observed tip-to-tip (i.e., BAR-to-BAR) contacts between adjacent Vps5 and also lateral contacts between the PX domains. Also, they observed that the contacts between Vps5 and the membrane occurs at positive patches in the tip of the BAR domains and the PX domain. However, there are clear differences between the solved yeast structure[92] and our SNX1/SNX5 structure. First, the SNX1/SNX5 heterodimer forms tubes with different number of starts, while the Vps5 tubes do not. It could imply that human SNXs may accommodate many cargos, and perhaps even different ones, in the same tube. Taking into account that mammal's proteome is much more extensive than yeast proteome; it could be thought that the heterodimers requires higher flexibility to accommodate a bigger variety of cargos. Second, SNX1/SNX5 heterodimer exhibits distinct set of lattice contacts compared to the reported Vps5 lattice. They found tip-to-tip contacts between the N-terminal of the α1 helix of the BAR domain and the loop between helixes α2 and α3 of an adjacent BAR domain. Besides, they showed lateral contacts between PX domains. We also observed tip-to-tip contacts between adjacent BAR domains and lateral contacts between PX domains; however, we also found conserved PX-to-BAR lateral contacts that stabilises the lattice. Third, the SNX1/SNX5 heterodimer membrane contacts rely mostly on the BAR domain and not also on the PX domain, as is the case for Vps5. Fourth, the protein density over the membrane is lower for the SNX1/SNX5 heterodimer that the Vps5 homodimer, probably because Vps5 packing is much higher due to the lack of cargo on the tube without steric clashes. Last but not least, while the arrangement of Vps5 does not contain cargo, and Vps5 probably does not participate in its recruitment[204], the Vps5 lattice might not be representative of a coat for cargo transport. Alternatively, SNX1/SNX5 heterodimer must accommodate CIMPR at the PX domain of SNX5. Our reconstitution included the cytosolic tail of CIMPR. Although we cannot observe it at the current resolution, we have proved its presence through co-flotation assays (Figure 37, results section). Nonetheless, this is an ongoing work, and we aim to increase the resolution and observe a density that can match with the cargo. This information would confirm our results, where a contribution of CIMPR recruitment by SNX1 is shown (Figure 37, results section).

In summary, we have described how SNX1 and SNX5 interact between them, how they interact with the membrane, how they interact with the model cargo CIMPR, and how they

assemble on the membrane to create a lattice that stabilises membrane deformation. Our work has established a strategy to obtain highly pure SNX1/SNX5 heterodimers, which has been extended to other heterodimers[128]. Our experiments, combined with recently published research by other laboratories[77, 206], have demonstrated that SNX heterodimers, and not homodimers, are the preferred functional assembly. We have visualized the structure of a SNX-BAR heterodimer and analysed its interface. Its characterisation allowed us to show that heterodimers are far more stable than homodimers, with SNX5 being able to outcompete SNX1 homodimerisation to form SNX1/SNX5 heterodimers. We have seen that heterodimerisation is mainly driven by extensive hydrophobic interactions at the interface (Figure 42, results section). By analysing the interaction of SNXs with membranes, we have established that PX domains do not strongly bind to any PIP. Also, we have confirmed that SNX1, and not SNX5, are able to induce tubulation *in vitro*, and that the SNX1/SNX5 heterodimer keeps the tubulation capacity. We have found that the SNX1/SNX5 heterodimer has twelve times more affinity for the CIMPR than the already reported interaction for the PX domain of SNX5 alone[77], which suggest that other regions of the heterodimer could participate in this interaction. Also, we have observed that SNX1 is able to interact with CIMPR in the presence of membranes, while SNX2 is not. Finally, we have obtained a cryoET model of the heterodimer producing tubules from model membranes. Also, we have characterised the tip-to-tip and tip-to-PX interactions that differ from previously described for VPS5[145]. We also have observed that lateral contacts between adjacent turns of the helical PX-BAR are variable depending on the number of helical starts. Finally, the presence of multiple helical starts may reflect the ability to accommodate different cargo sizes with the axial advance of the helix along the tube.

7. CONCLUSIONS

7. CONCLUSIONS

1. We have stablished a protocol for the purification of SNX1-SNX5 heterodimers.

2. We have solved the crystal structure of the $SNX1^{BAR}/SNX5^{BAR}$ heterodimer showing a banana shaped structure with a Radius of curvature of ~160Å and asymmetrical features.

3. Analysis of the residues present at the interface of the BAR domains indicates that heterodimerisation is mainly driven by electrostatic complementarity and hydrophobic desolvation, and that steric and charge-charge repulsions prevent homodimerisation.

4. Cargo recognition requires the PX domain of SNX5, but the interaction is markedly enhanced by the full length SNX1/SNX5 heterodimer.

5. The lattice coat formed by SNX1/SNX5 heterodimers around model membranes exhibit tip-to-tip contacts that extend the lattice and tip-to-PX lateral contacts that stabilises the formed lattice

6. The SNX1/SNX5 lattice on the membranes presents a homogeneous radii and pitch of 13.1 nm and 7.3 nm respectively while presenting different number of starts that can confer flexibility to the recruitment of different cargoes.

8. BIBLIOGRAPHY

1. Jékely, G. *Eukaryotic membranes and cytoskeleton : origins and evolution*. (Springer Science+Business Media, 2007).

2. Dacks, J. B. & Field, M. C. Evolution of the eukaryotic membrane-trafficking system : origin , tempo and mode. *J. Cell Sci.* **120**, 2977–2985 (2007).

3. Vellai, T. & Vida, G. The origin of eukaryotes: The difference between prokaryotic and eukaryotic cells. *Proc. R. Soc. B Biol. Sci.* **266**, 1571–1577 (1999).

4. Chen S. Valm, A. Lippincot-Schwartz, J. Interacting Organelles. *Curr Opin Cell Biol.* **53**, 84–91 (2018).

5. Helle, S. C. J. *et al.* Organization and function of membrane contact sites. *Biochim. Biophys. Acta - Mol. Cell Res.* **1833**, 2526–2541 (2013).

6. Saheki, Y. & Camilli, P. De. Endoplasmic Reticulum – Plasma Membrane Contact Sites. *Annu. Rev. Biochem.* **86**, 659–684 (2017).

7. Stefan, C. J., Manford, A. G. & Emr, S. D. ER–PM connections: sites of information transfer and inter- organelle communication. *Curr Opin Cell Biol.* **25**, 434–442 (2014).

8. Vassilieva, E. V. & Nusrat, A. Vesicular trafficking: Molecular tools and targets. *Methods Mol. Biol.* **440**, 3–14 (2008).

9. Veatch, S. L. & Keller, S. L. Separation of Liquid Phases in Giant Vesicles of Ternary Mixtures of Phospholipids and Cholesterol. *Biophys. J.* **85**, 3074–3083 (2003).

10. Baoukina, S., Ingolfsson, H. I., Marrink, S. J. & Tieleman, D. P. Curvature-Induced Lipid Sorting in Plasma Membrane Tethers. *Biophys. J.* **112**, 467a (2017).

11. Wong, L. H., Gatta, A. T. & Levine, T. P. Lipid transfer proteins: the lipid commute via shuttles, bridges and tubes. *Nat. Rev. Mol. Cell Biol.* **20**, 85–101 (2019).

12. Nickel, W. & Seedorf, M. Unconventional Mechanisms of Protein Transport to the Cell Surface of Eukaryotic Cells. *Annu. Rev. Cell Dev. Biol.* **24**, 287–308 (2008).

13. Blobel, G., Walter, P. & Gilmore, R. Intracellular protein topogenesis. *Prog. Clin. Biol. Res.* **168**, 3–10 (1984).

14. Osborne, A. R., Rapoport, T. A. & van den Berg, B. Protein Translocation By the Sec61/Secy Channel. *Annu. Rev. Cell Dev. Biol.* **21**, 529–550 (2005).

15. Gadila, S. K. G. & Kim, K. Cargo trafficking from the trans-Golgi network towards the endosome. *Biol. Cell* **108**, 205–218 (2016).

16. Papanikou E., G. B. S. Golgi Compartmentation and Identity. *Curr Opin Cell Biol.* **0**, 74–81 (2014).

17. Bonifacino, J. S. & Rojas, R. Retrograde transport from endosomes to the trans-Golgi network. *Nat. Rev. Mol. Cell Biol.* **7**, 568–579 (2006).

18. Cossart, P. & Helenius, A. Endocytosis of Viruses and Bacteria. (2014).

19. Wandinger-ness, A. & Zerial, M. Rab Proteins and the Compartmentalization of the Endosomal System. *Cold Spring Harb Perspect Biol.* 1–25 (2014) doi:10.1101/cshperspect.a022616.

20. Mayinger, P. Phosphoinositides and vesicular membrane traffic. **1821**, 1104–1113 (2013).

21. Royle, S. J. The cellular functions of clathrin Clathrin. *Cell Mol Life Sci.* **63**, 1823–1832 (2006).

22. Scales, S. J., Gomez, M. & Kreis, T. E. Coat Proteins Regulating Membrane Traffic. *Int. Rev. Citol.* **195**, 67–144 (2000).

23. Lippincott-Schwartz, J. & Liu, W. Online Article: Insights into COPI coat assembly and function in living cells. *Trends Cell Biol.* **16**, e1–e4 (2006).

24. Gürkan, C., Stagg, S. M., LaPointe, P. & Balch, W. E. The COPII cage: Unifying principles of vesicle coat assembly. *Nat. Rev. Mol. Cell Biol.* **7**, 727–738 (2006).

25. Edeling, M. A., Smith, C. & Owen, D. Life of a clathrin coat: Insights from clathrin and AP structures. *Nat. Rev. Mol. Cell Biol.* **7**, 32–44 (2006).

26. Park, S. Y. & Guo, X. Adaptor protein complexes and intracellular transport. *Biosci. Rep.* **34**, 381–390 (2014).

27. Hirst, J., Itzhak, D. N., Antrobus, R., Borner, G. H. H. & Robinson, M. S. Role of the AP-5 adaptor protein complex in late endosome-to-Golgi retrieval. *PLoS Biol.* **16**, 1–28 (2018).

28. Lim, J. P. & Gleeson, P. A. Macropinocytosis: An endocytic pathway for internalising large gulps. *Immunol. Cell Biol.* **89**, 836–843 (2011).

29. Muriel, O., Sánchez-álvarez, M., Strippoli, R. & Angel, M. *Role of the Endocytosis of Caveolae in Intracellular Signaling and Metabolism*. (Endocytosis and signalling, 2018). doi:10.1007/978-3-319-96704-2.

30. Howes, M. T. *et al.* Clathrin-independent carriers form a high capacity endocytic sorting system at the leading edge of migrating cells. *J. Cell Biol.* **190**, 675–691 (2010).

31. Anand Venkataraman*, Daniel J. Nevrivy†, Theresa M. Filtz, and M. L. GRP1-associated scaffold protein (grasp) is a regulator of the adp ribosylation factor 6 (arf6)-dependent membrane trafficking pathway. *Cell Biol. Int.* **36**, 1115–1128 (2012).

32. Glebov, O. O., Bright, N. A. & Nichols, B. J. Flotillin-1 defines a clathrin-independent endocytic pathway in mammalian cells. *Nat. Cell Biol.* **8**, (2006).

33. Boucrot, E. *et al.* Endophilin marks and controls a clathrin- independent endocytic pathway. *Nature* **0**, (2014).

34. Cyril Basquin, Valerie Malarde, Paul Mellor, Deborah H. Anderson, Vannary Meas-Yedid, Jean-Christophe Olivo-Marin, A. D.-V. and N. S. The signalling factor PI3K is a specific regulator of the of IL-2 receptors. *J. Cell Sci.* **126**, 1099–1108 (2013).

35. Seaman, M. N. J., Mccaffery, J. M. & Emr, S. D. A Membrane Coat Complex Essential for Endosome-to-Golgi Retrograde Transport in Yeast. *J. Cell Sci.* **142**, 665–681 (1998).

36. Mcnally, K. E. *et al.* Retriever is a multiprotein complex for retromer-independent endosomal cargo recycling. *Nat. Cell Biol.* (2017) doi:10.1038/ncb3610.

37. Mallam, A. L. & Marcotte, E. M. Systems-wide studies uncover Commander, a multiprotein complex essential to human development. *Cell Syst.* **4**, 483–494 (2017).

38. McMahon, H. T. & Boucrot, E. Membrane curvature at a glance. *J. Cell Sci.* **128**, 1065–1070 (2015).

39. Shawn M. Ferguson, P. D. C. Dynamin, a membrane-remodelling GTPase. *Nat Rev Mol Cell Biol.* **13**, 75–88 (2012).

40. Rappoport, J. Z. *et al.* Dynamics of clathrin and adaptor proteins during endocytosis. *Am J Physiol Cell Physiol* **291**, 1072–1081 (2006).

41. Zink, S., Wenzel, D., Wurm, C. A. & Schmitt, H. D. A Link between ER Tethering and COP-I Vesicle Uncoating. *Dev. Cell* **17**, 403–416 (2009).

42. Blanchoin, L., Boujemaa-Paterski, R., Sykes, C. & Plastino, J. Actin dynamics, architecture, and mechanics in cell motility. *Physiol. Rev.* **94**, 235–263 (2014).

43. Gadadhar, S., Bodakuntla, S., Natarajan, K. & Janke, C. The tubulin code at a glance. *J. Cell Sci.* **130**, 1347–1353 (2017).

44. Vale, R. D. The molecular motor toolbox for intracellular transport. *Cell* **112**, 467–480 (2003).

45. Caviston, J. P. & Holzbaur, E. L. F. Microtubule motors at the intersection of trafficking and transport. *Trends Cell Biol.* **16**, 530–537 (2006).

46. Lapierre, L. A. *et al.* Myosin Vb is associated with plasma membrane recycling systems. *Mol. Biol. Cell* **12**, 1843–1857 (2001).

47. Arnaud Echard, Florence Jollivet, Olivier Martinez, Jean-Jacques Lacapère, Annie Rousselet, Isabelle Janoueix-Lerosey, B. G. Interaction of a Golgi-associated kinesin-like protein with Rab6. *Science (80-.).* **279**, 580–585 (1998).

48. Johansson, M. *et al.* Activation of endosomal dynein motors by stepwise assembly of Rab7-RILP-p150Glued, ORP1L, and the receptor βIII spectrin. *J. Cell Biol.* **176**, 459–471 (2007).

49. Yu, I. & Hughson, F. M. Tethering Factors as Organizers of Intracellular Vesicular Traffic. *Annu. Rev. Cell Dev. Biol.* **26**, 137–156 (2010).

50. Seals, D. F., Eitzen, G., Margolis, N., Wickner, W. T. & Price, A. A Ypt ¯Rab effector complex containing the Sec1 homolog Vps33p is required for homotypic vacuole fusion. *PNAS* **97**, 9402–9407 (2000).

51. Spang, A. Membrane Tethering Complexes in the Endosomal System. *Front. Cel L. Dev. Biol.* **4**, 1–7 (2016).

52. Ho, R. & Stroupe, C. The HOPS / class C Vps complex tethers membranes by binding to one Rab GTPase in each apposed membrane. *Mol. Biol. Cell* **26**, 2655–2663 (2015).

53. Peplowska, K., Markgraf, D. F., Ostrowicz, C. W., Bange, G. & Ungermann, C. The CORVET Tethering Complex Interacts with the Yeast Rab5 Homolog Vps21 and Is Involved in Endo-Lysosomal Biogenesis. *Dev. Cell* 739–750 (2007) doi:10.1016/j.devcel.2007.03.006.

54. Meiringer, C. T. A. *et al.* The Dsl1 Protein Tethering Complex Is a Resident Endoplasmic Reticulum Complex , Which Interacts with Five Soluble NSF (N - Ethylmaleimide-sensitive Factor) Attachment Protein Receptors (SNAREs). *J. Biol. Chem.* **286**, 25039–25046 (2011).

55. Lupashin, V. & Sztul, E. Golgi tethering factors. *Biochim. Biophys. Acta* **1744**, 325–339 (2005).

56. Picco, A., Irastorza-azcarate, I., Specht, T., Devos, D. P. & Pazos, I. The In Vivo Architecture of the Exocyst Provides Structural Basis for Exocytosis Exocyst Secretory vesicle Article The In Vivo Architecture of the Exocyst Provides Structural Basis for Exocytosis. *Cell* **168**, 400–412 (2017).

57. Martens, S. & Mcmahon, H. T. Mechanisms of membrane fusion : disparate players and common principles. *Nat. Rev. Mol. Cell Biol.* **9**, 543–556 (2008).

58. Chernomordik, L. V & Kozlov, M. M. Protein-lipid interplay in fusion and fission of biological membranes. *Annu. Rev. Biochem.* **72**, 175–207 (2003).

59. Jahn, R. & Scheller, R. H. SNAREs — engines for membrane fusion. *Nat. Rev. Mol. Cell Biol.* **7**, 631–643 (2006).

60. Mcnew, J. A. *et al.* Ykt6p , a Prenylated SNARE Essential for Endoplasmic Reticulum-Golgi Transport. *J. Biol. Chem.* **272**, 17776–17783 (1997).

61. Huang, X. *et al.* Mechanistic insights into the SNARE complex disassembly. *Sci. Adv.* 1–12 (2019) doi:10.1126/sciadv.aau8164.

62. Bitsikas, V., Corrêa, I. R. & Nichols, B. J. Clathrin-independent pathways do not contribute significantly to endocytic flux. *Elife* **2014**, 1–26 (2014).

63. Miller, S. E., Mathiasen, S. & Owen, D. J. CALM Regulates Clathrin-Coated Vesicle Size and Maturation by Directly Sensing and Driving Article CALM Regulates Clathrin-Coated Vesicle Size and Maturation by Directly Sensing and Driving Membrane Curvature. *Dev. Cell* **33**, 163–175 (2015).

64. Ehrlich, M. *et al.* Endocytosis by Random Initiation and Stabilization of Clathrin-Coated Pits. *Cell* **118**, 591–605 (2004).

65. Wallroth, A. & Haucke, V. cro Phosphoinositide conversion in endocytosis and the endolysosomal system. *JBC* **293**, 1526–1535 (2018).

66. Mettlen, M., Loerke, D., Yarar, D., Danuser, G. & Schmid, S. L. Cargo- and adaptor-specific mechanisms regulate clathrin-mediated endocytosis. *J. Cell Biol.* **188**, 919–933 (2010).

67. Scott, C. C., Vacca, F. & Gruenberg, J. Endosome maturation, transport and

functions. *Semin. Cell Dev. Biol.* **31**, 2–10 (2014).

68. Wang, C. *et al.* Investigation of endosome and lysosome biology by ultra pH-sensitive nanoprobes. *Adv. Drug Deliv. Rev.* 1–10 (2016) doi:10.1016/j.addr.2016.08.014.

69. Shin, H. *et al.* An enzymatic cascade of Rab5 effectors regulates phosphoinositide turnover in the endocytic pathway. *J. Cell Biol.* **170**, 607–618 (2004).

70. Norris, A. *et al.* SNX-1 and RME-8 oppose the assembly of HGRS-1 / ESCRT-0 degradative microdomains on endosomes. *PNAS* **114**, 307–316 (2017).

71. Cho, K. J., Ishido, S., Eisenlohr, L. & Roche, P. Rab11a Regulates MHC Class II Recycling and Promotes Antigenic Peptide Exchange in Recycling Endosomes of Mature Dendritic Cells. *J. Immunol.* **198**, 146.11 LP-146.11 (2017).

72. Hu, Y., Dammer, E. B., Ren, R. & Wang, G. The endosomal-lysosomal system : from acidification and cargo sorting to neurodegeneration. *Transl. Neurodegener.* 1–10 (2015) doi:10.1186/s40035-015-0041-1.

73. Sun, Y., Hedman, A. C., Tan, X., Schill, N. J. & Anderson, R. A. Endosomal Type I-gamma PIP 5-Kinase Controls EGF Receptor Lysosomal Sorting. *Dev. Cell* **25**, 144–155 (2013).

74. Gallon, M. *et al.* A unique PDZ domain and arrestin-like fold interaction reveals mechanistic details of endocytic recycling by SNX27-retromer. *PNAS* **111**, 3604–3613 (2014).

75. María Lucas, David C. Gershlick, Ander Vidaurrazaga, Adriana L. Rojas, Juan S. Bonifacino, and A. H. Structural Mechanism for Cargo Recognition by the Retromer Complex. *Cell* **167**, 1623–1635 (2016).

76. Simonetti, B., Danson, C. M., Heesom, K. J. & Cullen, P. J. Sequence-dependent cargo recognition by SNX-BARs mediates retromer-independent transport of CI-MPR. *J. Cell Biol.* **216**, 3695–3712 (2017).

77. Simonetti, B. *et al.* Molecular identification of a BAR domain-containing coat complex for endosomal recycling of transmembrane proteins. *Nat. Cell Biol.* **21**, 1219–1233 (2019).

78. Ballabio, A. The awesome lysosome. *EMBO Mol Med* **8**, 73–76 (2016).

79. Petsko, S. A. S. and G. A. Retromer in Alzheimer disease, Parkinson disease and other neurological disorders. *Occup. Environ. Heal.* **16**, 126–132 (2015).

80. McNally, K. E. & Cullen, P. J. Endosomal Retrieval of Cargo: Retromer Is Not Alone. *Trends Cell Biol.* **28**, 807–822 (2018).

81. Derivery, E., Helfer, E., Henriot, V. & Gautreau, A. Actin polymerization controls the organization of WASH domains at the surface of endosomes. *PLoS One* **7**, (2012).

82. Hierro, A. *et al.* Functional architecture of the retromer cargo-recognition complex. *Nature* **449**, 1063–1067 (2007).

83. Harrison, M. S. *et al.* A mechanism for retromer endosomal coat complex

assembly with cargo. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 267–272 (2014).

84. Phillips-Krawczak, C. A. *et al.* COMMD1 is linked to the WASH complex and regulates endosomal trafficking of the copper transporter ATP7A. *Mol. Biol. Cell* **26**, 91–103 (2015).

85. Chishti, A., Kim, A. & Hoover, K. B. The FERM domain: a unique module involved in. *Tibs* **0004**, 281–282 (1998).

86. Bartuzi, P., Hofker, M. H. & Van de Sluis, B. Tuning NF-κB activity: A touch of COMMD proteins. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1832**, 2315–2321 (2013).

87. Bartuzi, P. *et al.* CCC- and WASH-mediated endosomal sorting of LDLR is required for normal clearance of circulating LDL. *Nat. Commun.* **7**, 1–11 (2016).

88. Li, H. *et al.* Endosomal sorting of notch receptors through COM MD9-dependent pathways modulates notch signaling. *J. Cell Biol.* **211**, 605–617 (2015).

89. Seaman, M. N. J., McCaffery, J. M. & Emr, S. D. A membrane coat complex essential for endosome-to-Golgi retrograde transport in yeast. *J. Cell Biol.* **142**, 665–681 (1998).

90. Purushothaman, L. K. & Ungermann, C. Cargo induces retromer-mediated membrane remodeling on membranes. *Mol. Biol. Cell* **29**, 2709–2719 (2018).

91. Purushothaman, L. K., Arlt, H., Kuhlee, A., Raunser, S. & Ungermann, C. Retromer-driven membrane tubulation separates endosomal recycling from Rab7/Ypt7-dependent fusion. *Mol. Biol. Cell* **28**, 783–791 (2017).

92. Kovtun, O. *et al.* Structure of the membrane-assembled retromer coat determined by cryo-electron tomography. *Nature* **561**, 561–564 (2018).

93. Kvainickas, A. *et al.* Cargo-selective SNX-BAR proteins mediate retromer trimer independent retrograde transport. *J. Cell Biol.* **216**, 3677–3693 (2017).

94. Cui, Y. *et al.* Retromer has a selective function in cargo sorting via endosome transport carriers. *J. Cell Biol.* **218**, 615–631 (2019).

95. Romano-Moreno, M. *et al.* Molecular mechanism for the subversion of the retromer coat by the Legionella effector RidL. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E11151–E11160 (2017).

96. Elwell, C. A. *et al.* Chlamydia interfere with an interaction between the mannose-6-phosphate receptor and sorting nexins to counteract host restriction. *Elife* **6**, 1–17 (2017).

97. Popa, A. *et al.* Direct Binding of Retromer to Human Papillomavirus Type 16 Minor Capsid Protein L2 Mediates Endosome Exit during Viral Infection. *PLoS Pathog.* **11**, 1–21 (2015).

98. Williams, E. T., Chen, X. & Moore, D. J. VPS35, the retromer complex and Parkinson's disease. *J. Parkinsons. Dis.* **7**, 219–233 (2017).

99. Vagnozzi, A. N. & Praticò, D. Endosomal sorting and traf fi cking , the retromer complex and neurodegeneration. *Mol. Psychiatry* **24**, 857–868 (2018).

100. Müller, U. C., Deller, T. & Korte, M. Not just amyloid: physiological functions of the amyloid precursor protein family. *Nat. Rev. Neurosci.* **18**, 281–298 (2017).

101. Zhang, H. *et al.* The retromer complex and sorting nexins in neurodegenerative diseases. *Front. Aging Neurosci.* **10**, 1–11 (2018).

102. Berman, D. E., Ringe, D., Petsko, G. A. & Small, S. A. The Use of Pharmacological Retromer Chaperones in Alzheimer ' s Disease and other Endosomal-related Disorders. *Neurotherapeutics* **12**, 12–18 (2015).

103. Vincent J Mecozzi, Diego E Berman, Sabrina Simoes, Chris Vetanovetz, M. R., Awal, Vivek M Patel, Remy T Schneider, G. A. P. D. R. and & Small, S. A. Pharmacological chaperones stabilize retromer to limit APP processing. *Nat Chem Biol* **10**, 443–449 (2014).

104. Amy K. Kendall, Boyang Xie, Peng Xu, Jue Wang, Rodger Burcham, Meredith N. Frazier, Elad Binshtein, Hui Wei, Todd R. Graham, Terunaga Nakagawa, and L. P. J. Mammalian retromer is an adaptable scaffold for cargo sorting from endosomes. *bioRxiv* **53**, 1689–1699 (2019).

105. Moore, C. A. C., Milano, S. K. & Benovic, J. L. Regulation of Receptor Trafficking by GRKs and Arrestins. *Annu. Rev. Physiol.* **69**, 451–482 (2007).

106. Van Weering, J. R. T. *et al.* Molecular basis for SNX-BAR-mediated assembly of distinct endosomal sorting tubules. *EMBO J.* **31**, 4466–4483 (2012).

107. Zhang, M., Chen, L., Wang, S. & Wang, T. Rab7 : roles in membrane trafficking and disease Bioscience Reports. *Biosci. Rep.* **29**, 193–209 (2009).

108. Stroupe, C. This is the end: Regulation of Rab7 nucleotide binding in endolysosomal trafficking and autophagy. *Front. Cell Dev. Biol.* **6**, 1–15 (2018).

109. Priya, A., Kalaidzidis, I. V., Kalaidzidis, Y., Lambright, D. & Datta, S. Molecular Insights into Rab7-Mediated Endosomal Recruitment of Core Retromer: Deciphering the Role of Vps26 and Vps35. *Traffic* **16**, 68–84 (2015).

110. Jia, D. *et al.* Structural and mechanistic insights into regulation of the retromer coat by TBC1d5. *Nat. Commun.* **7**, 1–11 (2016).

111. Chaineau, M., Danglot, L. & Galli, T. Multiple roles of the vesicular-SNARE TI-VAMP in post-Golgi and endosomal trafficking. *FEBS Lett.* **583**, 3817–3826 (2009).

112. Hesketh, G. *et al.* VARP is recruited on to endosomes by direct interaction with retromer, where together they function in export to the cell surface. *Dev. Cell* **29**, 591–606 (2014).

113. Seaman, M. N. J., Gautreau, A. & Billadeau, D. D. Retromer-mediated endosomal protein sorting: All WASHed up! *Trends Cell Biol.* **23**, 522–528 (2013).

114. Helfer, E. *et al.* Endosomal recruitment of the WASH complex: Active sequences and mutations impairing interaction with the retromer. *Biol. Cell* **105**, 191–207 (2013).

115. McGough, I. J. *et al.* Retromer binding to FAM21 and the WASH complex is perturbed by the Parkinson disease-linked VPS35(D620N) mutation. *Curr. Biol.*

24, 1670–1676 (2014).

116. J. Chi, R. Harrison, M. S. & Burd, C. G. Biogenesis of endosome-derived transport carriers. *Cell Mol Life Sci.* **72**, 3441–3455 (2015).

117. Steinberg, F. *et al.* A global analysis of SNX27 – retromer assembly and cargo specificity reveals a function in glucose and metal ion transport relies on a membrane-curvature-sensing domain in Atg3. *Nat. Publ. Gr.* **16**, 821 (2014).

118. Kvainickas, A. *et al.* Retromer- and WASH-dependent sorting of nutrient transporters requires a multivalent interaction network with ANKRD50. *J. Cell Sci.* **130**, 382–395 (2017).

119. Ponting, C. P. Novel domains in NADPH oxidase subunits, sorting nexins, and PtdIns 3- kinases: Binding partners of SH3 domains? *Protein Sci.* **5**, 2353–2357 (1996).

120. Zhan, Y., Virbasius, J. V., Song, X., Pomerleau, D. P. & Wayne Zhou, G. The p40phox and p47phox PX domains of NADPH oxidase target cell membranes via direct and indirect recruitment by phosphoinositides. *J. Biol. Chem.* **277**, 4512–4518 (2002).

121. Koharudin, L. M. I., Furey, W., Liu, H., Liu, Y. J. & Gronenborn, A. M. The phox domain of sorting nexin 5 lacks phosphatidylinositol 3-phosphate (PtdIns(3)P) specificity and preferentially binds to phosphatidylinositol 4,5-bisphosphate (PtdIns(4,5)P2). *J. Biol. Chem.* **284**, 23697–23707 (2009).

122. Ellson, C. D., Andrews, S., Stephens, L. R. and. Hawkins, P. T. The PX domain : a new phosphoinositide-binding module. *J. Cell Sci.* **115**, 1099–1105 (2002).

123. Chandra, M. *et al.* Classification of the human phox homology (PX) domains based on their phosphoinositide binding specificities. *Nat. Commun.* **10**, (2019).

124. Cullen, P. J. Endosomal sorting and signalling: An emerging role for sorting nexins. *Nat. Rev. Mol. Cell Biol.* **9**, 574–582 (2008).

125. Wassmer, T. *et al.* A loss-of-function screen reveals SNX5 and SNX6 as potential components of the mammalian retromer. *J. Cell Sci.* **120**, 45–54 (2007).

126. Fernandes, H., Franklin, E., Jollivet, F., Bliedtner, K. & Khan, A. R. Mapping the interactions between a RUN domain from DENND5/Rab6IP1 and sorting nexin 1. *PLoS One* **7**, 1–9 (2012).

127. Wassmer, T. *et al.* The Retromer Coat Complex Coordinates Endosomal Sorting and Dynein-Mediated Transport, with Carrier Recognition by the trans-Golgi Network. *Dev. Cell* **17**, 110–122 (2009).

128. Yong, X. *et al.* Expression and purification of the SNX1/SNX6 complex. *Protein Expr. Purif.* **151**, 93–98 (2018).

129. Mizutani, R. *et al.* Expression of sorting nexin 12 is regulated in developing cerebral cortical neurons. *J. Neurosci. Res.* **90**, 721–731 (2012).

130. Priya, A. *et al.* Essential and selective role of SNX12 in transport of endocytic and retrograde cargo. *J. Cell Sci.* **130**, 2707–2721 (2017).

131. Ghai, R. *et al.* Structural basis for endosomal trafficking of diverse

transmembrane cargos by PX-FERM proteins. *Proc. Natl. Acad. Sci. U. S. A.* **110**, (2013).

132. Gallon, M. *et al.* A unique PDZ domain and arrestin-like fold interaction reveals mechanistic details of endocytic recycling by SNX27-retromer. *Proc. Natl. Acad. Sci. U. S. A.* **111**, (2014).

133. Lenoir, M. *et al.* Phosphorylation of conserved phosphoinositide binding pocket regulates sorting nexin membrane targeting. *Nat. Commun.* **9**, 1–12 (2018).

134. Adam Frost, Rushika Perera, Aurélien Roux, Krasimir Spasov, O. & Destaing, Edward H. Egelman, Pietro De Camilli, and V. M. U. Structural Basis of Membrane Invagination by F-BAR Domains. *Cell* **132**, 807–817 (2008).

135. Clairfeuille, T. *et al.* A molecular code for endosomal recycling of phosphorylated cargos by the SNX27-retromer complex. *Nat. Struct. Mol. Biol.* **23**, 921–932 (2016).

136. Ghai, R., Tello-lafoz, M., Norwood, S. J., Yang, Z. & Clairfeuille, T. Phosphoinositide binding by the SNX27 FERM domain regulates its localization at the immune synapse of activated T-cells. *J. Cell Sci.* **128**, 553–565 (2015).

137. Weering, J. R. T. Van & Cullen, P. J. Membrane-associated cargo recycling by tubule-based endosomal sorting. *Semin. Cell Dev. Biol.* **31**, 40–47 (2014).

138. Gallop, H. T. M. J. L. Membrane curvature and mechanisms of dynamic cell membrane remodelling. *Nature* **438**, 590–596 (2005).

139. Frost, A., Unger, V. M. & De Camilli, P. The BAR Domain Superfamily: Membrane-Molding Macromolecules. *Cell* **137**, 191–196 (2009).

140. Tarricone, C. *et al.* The structural basis of Arfaptin-mediated cross-talk between Rac and Arf signalling pathways. *Nature* **411**, 215–219 (2001).

141. Salzer, U., Kostan, J. & Djinović, K. Deciphering the BAR code of membrane modulators. *Cell. Mol. Life Sci.* **74**, 2413–2438 (2017).

142. Carman, P. J. & Dominguez, R. BAR domain proteins — a linkage between cellular membranes , signaling pathways , and the actin cytoskeleton. *Biophys. Rev.* **3**, 1587–1604 (2018).

143. Kooijman, E. E. *et al.* Spontaneous Curvature of Phosphatidic Acid and Lysophosphatidic Acid. *Biochemistry* **44**, 2097–2102 (2005).

144. Pick, P. *et al.* An Amphipathic Helix Directs Cellular Membrane Curvature Sensing and Function of the BAR Domain Protein PICK1. *Cell Rep.* **23**, 2056–2069 (2018).

145. Carsten Mim, Haosheng Cui, Joseph A. Gawronski-Salerno, Adam Frost, EdwardLyman, Gregory A. Voth, and V. M. U. Structural Basis of Membrane Bending by the N-BAR Protein Endophilin. *Cell* **149**, 137–145 (2012).

146. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).

147. Marblestone, J. G. Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO. *Protein Sci.* **15**,

182–189 (2006).

148. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).

149. Abramson, J. *et al.* Purification, crystallization and preliminary crystallographic studies of an integral membrane protein, cytochrome bo3 ubiquinol oxidase from Escherichia coli. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **56**, 1076–1078 (2000).

150. Malakhov, M. P. *et al.* SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J. Struct. Funct. Genomics* **5**, 75–86 (2004).

151. Gasteiger, E. *et al.* ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31**, 3784–3788 (2003).

152. Martin, S. R. & Schilstra, M. J. Circular Dichroism and Its Application to the Study of Biomolecules. *Methods Cell Biol.* **84**, 263–293 (2008).

153. Roberts, E. sahin and C. J. Size-Exclusion Chromatography with Multi-angle Light Scattering for Elucidating Protein Aggregation Mechanisms. *Methods Mol Biol.* **899**, 403–423 (2012).

154. Velazquez-Campoy, A., Ohtaka, H., Nezami, A., Muzammil, S. & Freire, E. Isothermal titration calorimetry. *Curr. Protoc. cell Biol.* **Chapter 17**, Unit 17.8 (2004).

155. Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **9**, 1665–1672 (1990).

156. Rould, M. A. Screening for heavy-atom derivatives and obtaining accurate isomorphous differences. *Methods Enzymol.* **276**, 461–472 (1997).

157. Bergfors, T. Seeds to crystals. *J. Struct. Biol.* **142**, 66–76 (2003).

158. Incardona, M. F. *et al.* EDNA: A framework for plugin-based applications applied to X-ray experiment online data analysis. *J. Synchrotron Radiat.* **16**, 872–879 (2009).

159. Kabsch, W. XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125–132 (2010).

160. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 235–242 (2011).

161. Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: Recent developments in and around SHARP 2.0. *Acta Crystallogr. - Sect. D Biol. Crystallogr.* **59**, 2023–2030 (2003).

162. Adams, P. D. *et al.* PHENIX : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 213–221 (2010) doi:10.1107/S0907444909052925.

163. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

164. Brünger, A. T. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nat. Publ. Gr.* **355**, 472–475 (1992).

165. Stowell, M. H. B., Marks, B., Wigge, P. & McMahon, H. T. Nucleotide-dependent conformational changes in dynamin: evidence for a mechanochemical molecular spring. *Nat. Cell Biol.* **1**, 27–32 (1999).

166. Ricarda A. Busse, Andreea Scacioc, Amanda M. Schalk, Roswitha Krick, Michael Thumm, and K. K. Analyzing Protein–Phosphoinositide Interactions with Liposome Flotation Assays. *Lipid Signal. Protoc.* **1376**, 155–162 (2016).

167. Hyland, C., Vuillard, L., Hughes, C. & Koronakis, V. Membrane Interaction of Escherichia coli Hemolysin: Flotation and Insertion-Dependent Labeling by Phospholipid Vesicles. *J. Bacteriol.* **183**, 5364–5370 (2001).

168. Bigay, J. & Antonny, B. Real-time assays for the assembly-disassembly cycle of COP coats on liposomes of defined size. *Methods Enzymol.* **404**, 95–107 (2005).

169. Russo, C. J., Scotcher, S. & Kyte, M. Europe PMC Funders Group Europe PMC Funders Author Manuscripts A precision cryostat design for manual and semi-automated cryo-plunge instruments. **87**, 1–12 (2017).

170. Noble, A. J. *et al.* Reducing effects of particle adsorption to the air–water interface in cryo-EM. *Nat. Methods* **15**, 793–795 (2018).

171. Turoňová, B., Schur, F. K. M., Wan, W. & Briggs, J. A. G. Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4 Å. *J. Struct. Biol.* **199**, 187–195 (2017).

172. Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer Visualization of Three-Dimensional Image DataUsing IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).

173. Wilson-Kubalek, E. M., Brown, R. E., Celia, H. & Milligan, R. A. Lipid nanotubes as substrates for helical crystallization of macromolecules. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 8040–8045 (1998).

174. Hagen, W. J. H., Wan, W. & Briggs, J. A. G. Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *J. Struct. Biol.* **197**, 191–198 (2017).

175. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

176. Xiong, Q., Morphew, M. K., Schwartz, C. L., Hoenger, A. H. & Mastronarde, D. N. CTF determination and correction for low dose tomographic tilt series. *J. Struct. Biol.* **168**, 378–387 (2009).

177. Romero-durana, M., Jiménez-garcía, B. & Fernández-, J. pyDockEneRes : per-residue decomposition of protein-protein docking energy. *Bioinformatics* 2011–2013 (2019) doi:10.1093/bioinformatics/btz884.

178. Sali, A. & Bundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Journal of molecular biology* vol. 234 779–815 (1993).

179. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202 (1999).

180. Lawrence A Kelley, Stefans Mezulis, Christopher M Yates, M. N. W. & M. J. E. S. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).

181. Pylypenko, O., Lundmark, R., Rasmuson, E., Carlsson, S. R. & Rak, A. The PX-BAR membrane-remodeling unit of sorting nexin 9. *EMBO J.* **26**, 4788–4800 (2007).

182. Wang, Q., Kaan, H. Y. K., Hooda, R. N., Goh, S. L. & Sondermann, H. Structure and Plasticity of Endophilin and Sorting Nexin 9. *Structure* **16**, 1574–1587 (2008).

183. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, 320–324 (2014).

184. Matthews, B. W. Solvent Content of Protein. *J. Mol. Biol.* **33**, 491–497 (1968).

185. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr. Sect. D Struct. Biol.* **75**, 861–877 (2019).

186. Delano. The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC. *Ccp4.Ac.Uk* (2002).

187. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

188. Masuda, M. *et al.* Endophilin BAR domain drives membrane curvature by two newly identified structure-based mechanisms. *EMBO J.* **25**, 2889–2897 (2006).

189. Tammy Man-Kuang Cheng, Tom L. Blundell, and J. F.-R. pyDock: Electrostatics and Desolvation for Effective Scoring of Rigid-Body Protein–Protein Docking. *PROTEINS Struct. Funct. Bioinforma.* **68**, 503–515 (2007).

190. Ashkenazy, H. *et al.* ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350 (2016).

191. Seaman, M. N. J. Identification of a novel conserved sorting motif required for retromer-mediated endosome-to-TGN retrieval. *J. Cell Sci.* 2378–2389 (2007) doi:10.1242/jcs.009654.

192. Scheres, S. H. W. RELION : Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).

193. Desfosses, A., Ciuffa, R., Gutsche, I. & Sachse, C. SPRING – An image processing package for single-particle based helical reconstruction from electron cryomicrographs. *J. Struct. Biol.* **185**, 15–26 (2014).

194. Antón, Z. *et al.* A heterodimeric SNX4 : SNX7 SNX-BAR autophagy complex coordinates ATG9A trafficking for efficient autophagosome assembly. *biorXiv* (2020).

195. Pang, X. *et al.* A PH domain in ACAP1 possesses key features of the BAR

domain in promoting membrane curvature. *Dev. Cell* **31**, 73–86 (2014).

196. Dominique P Frueh, Andrew Goodrich, Subrata Mishra, and S. N. NMR methods for structural studies of large monomeric and multimeric proteins. *Curr Opin Struct Biol.* **23**, 734–739 (2013).

197. Qi Zhong, M. J. W., Cheri S. Lazar, A. M. H., Waltho, J. P. & Gill, and G. N. Determinants of the Endosomal Localization of Sorting Nexin 1. *Mol Biol Cell* **16**, 2049–2057 (2005).

198. Pylypenko, O. *et al.* A combinatorial approach to crystallization of PX-BAR unit of the human Sorting Nexin 9. *J. Struct. Biol.* **162**, 356–360 (2008).

199. Chandra, M. & Collins, B. M. The Phox Homology (PX) Domain. in *Advances in Experimental Medicine and Biology* 18 (Springer, 2018). doi:10.1007/5584_2018_185.

200. Wendler, F. & Tooze, S. Syntaxin 6: The Promiscuous behavior of a SNARE protein. *Traffic* **2**, 606–611 (2001).

201. Teasdale, R. D., Loci, D., Houghton, F., Karlsson, L. & Gleeson, P. A. A large family of endosome-localized proteins related to sorting nexin 1. *Biochem. J.* **358**, 7–16 (2001).

202. Rojas, R. *et al.* Regulation of retromer recruitment to endosomes by sequential action of Rab5 and Rab7. *J. Cell Biol.* **183**, 513–526 (2007).

203. Burda, P., Padilla, S. M., Sarkar, S. & Emr, S. D. Retromer function in endosome-to-Golgi retrograde transport is regulated by the yeast Vps34 Ptdlns 3-kinase. *J. Cell Sci.* **115**, 3889–3900 (2002).

204. Suzuki, S. W., Chuang, Y.-S., Li, M., Seaman, M. N. J. & Emr, S. D. A bipartite sorting signal ensures specificity of retromer complex in membrane protein recycling. *J. Cell Biol.* **218**, 2876–2886 (2019).

205. Paul, B. *et al.* Structural basis for the hijacking of endosomal sorting nexin proteins by Chlamydia trachomatis. *Elife* **6**, 1–23 (2017).

206. Yong, X. *et al.* Mechanism of cargo recognition by retromer-linked SNX-BAR proteins. *PLoS Biol.* **18**, e3000631 (2020).

207. Bronowska, A. K. *Thermodynamics of Ligand-Protein Interactions: Implications for Molecular Design*. (IntechOpen, 2011). doi:10.5772/19447.

208. Daum, B. *et al.* Supramolecular organization of the human N-BAR domain in shaping the sarcolemma membrane. *J. Struct. Biol.* **194**, 375–382 (2016).

209. Simunovic, M. *et al.* How curvature-generating proteins build scaffolds on membrane nanotubes. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 11226–11231 (2016).

210. Castaño-Díez, D., Kudryashev, M., Arheit, M. & Stahlberg, H. Dynamo: A flexible, user-friendly development tool for subtomogram averaging of cryo-EM data in high-performance computing environments. *J. Struct. Biol.* **178**, 139–151 (2012).

211. Navarro, P. P., Stahlberg, H. & Castaño-Díez, D. Protocols for subtomogram averaging of membrane proteins in the Dynamo software package. *Front. Mol.*

Biosci. **5**, 1–18 (2018).

212. Mari, M. *et al.* SNX1 defines an early endosomal recycling exit for sortilin and mannose 6-phosphate receptors. *Traffic* **9**, 380–393 (2008).

213. Wang, R., Qi, X., Schmiege, P., Coutavas, E. & Li, X. Marked structural rearrangement of mannose 6-phosphate/IGF2 receptor at different pH environments. *Sci. Adv.* **6**, 1–11 (2020).