Experimental and in silico Alphafold2 derived structures of the SNX-RGS proteins suggest a new class of lipid transfer protein

Blessy Paul[1,2]*, Saroja Weeratunga[1]*, Vikas A. Tillu[1], Hanaa Hariri[2,3], W. Mike Henne[2] and Brett M. Collins[1#]

[1]The University of Queensland, Institute for Molecular Bioscience, Queensland, 4072, Australia; [2]Department of Cell Biology, University of Texas Southwestern Medical Center, Dallas, TX 75390, United States; [3]Current address: Department of Biological Sciences, College of Liberal Arts and Sciences, Wayne State University, Detroit, MI, United States.

Running Title: Structural prediction of a novel lipid transfer domain

*These authors contributed equally.
#Corresponding author:
Brett Collins, +61 (0)7 33462043, Email: b.collins@imb.uq.edu.au

Keywords – sorting nexin (SNX), regulator of G-protein signaling (RGS), lipid droplets (LDs), AlphaFold2, X-ray crystallography, lipid transfer protein (LTP), phox homology (PX) domain.
Running title – SNX-RGS lipid transfer proteins

Summary

Recent advances in protein structure prediction using machine learning such as AlphaFold2 and RosettaFold presage a revolution in structural biology. Genome-wide predictions of protein structures are providing unprecedented insights into their architecture and intradomain interactions, and applications have already progressed towards assessing protein complex formation. Here we present detailed analyses of the sorting nexin proteins that contain regulator of G-protein signalling domains (SNX-RGS proteins), providing a key example of the ability of AlphaFold2 to reveal novel structures with previously unsuspected biological functions. These large proteins are conserved in most eukaryotes and are known to associate with lipid droplets (LDs) and sites of LD-membrane contacts, with key roles in regulating lipid metabolism. Previous studies indicate they possess five domains, including an N-terminal transmembrane domain that anchors them to the endoplasmic reticulum, an RGS domain, a lipid interacting phox homology (PX) domain and two additional domains named the PXA and PXC domains of unknown structure and function. Here we report the crystal structure of the RGS domain of sorting nexin 25 (SNX25) and show that the AlphaFold2 prediction closely matches the experimental structure. Analysing the full-length SNX-RGS proteins across multiple homologues and species we find that the distant PXA and PXC domains in fact fold into a single unique structure that notably features a large and conserved hydrophobic pocket. The nature of this pocket strongly suggests a role in lipid or fatty acid binding, and we propose that these molecules represent a new class of conserved lipid transfer proteins.

Introduction

The sorting nexins (SNXs) are a large family of proteins with diverse structures and functions. They are found in all eukaryotes and their defining feature is the presence of a phox homology (PX) domain, which most commonly associates with phosphatidylinositol phospholipids (PtdInsPs) to mediate interactions with membranes of the endolysosomal system [1-3]. SNX proteins are typically grouped into sub-families based on the presence of additional functional domains such as membrane tubulating bin/amphiphysin/rvs (BAR) domains, protein interacting SH3 domains, and GTPase activating protein (GAP) domains among many others [3].

One of the most highly conserved sub-families are the SNX proteins with regulator of G-protein signalling (RGS) domains or SNX-RGS proteins [4, 5]. In humans there are four genes encoding SNX-RGS proteins, SNX13, SNX14, SNX19 and SNX25. Other well-characterised family members include Mdm1 from *Saccharomyces cerevisiae* and Snazarus (Snz) from *Drosophila melanogaster*. These proteins have relatively low sequence similarity across homologues and across species but share a conserved architecture with an N-terminal hydrophobic membrane anchor, and a central RGS domain and PX domain (except for SNX19 orthologues which lack the RGS domain). In addition, the RGS and PX domains are flanked by two sequences referred to as the PX-associated domains PXA and PXC at the N- and C-terminus respectively (**Fig. 1A**). The N-terminal membrane anchor mediates localisation to the endoplasmic reticulum (ER) [6-10], and the proteins typically localise to sites of new lipid droplet synthesis and can enhance membrane tethering of the ER to other membrane compartments via their PX domains including the vacuole in yeast [8, 11, 12], the plasma membrane in *Drosophila* [10] and endolysosomal compartments in mammalian cells [6, 7, 9, 13].

The precise function of the SNX-RGS proteins is still unclear although their mutation or depletion has significant impacts on cellular physiology, particularly with respect to lipid homeostasis and endolysosomal function. Mutations in human SNX14 cause the autosomal-recessive cerebellar ataxia and intellectual disability syndrome SCAR20, with a common cellular phenotype of autophagic structures containing undigested material and increased cholesterol accumulation in endolysosomal organelles [6, 14, 15]. In mice SNX13, SNX14 and SNX25 are all essential for normal development [16-18], while altering the expression of Snz significantly extends the lifespan of flies [10, 19]. Previous work has defined the structures and lipid binding properties of the PX domains of these proteins; the human SNX13 and SNX19 PX domains bind the endosomal lipid PtdIns3P and SNX25 associates with multiple phosphorylated PtdInsP species, while SNX14 has little affinity for PtdInsP lipids due to an altered binding pocket [1, 5, 10]. The RGS domains of SNX13 and SNX14 associate with the $G\alpha_s$ subunit of trimeric G-proteins [20, 21]. While SNX13 was originally discovered as a GTPase activating protein (GAP) for $G\alpha_s$ [21], SNX14 can bind $G\alpha_s$ but has so far been found to not stimulate its GAP activity[20]. SNX13 and SNX14 have also been found to associate with the endosomal Rab GTPase Rab9A, although the functional significance of this is unclear [22]. Proteomics studies have identified association of yeast Mdm1 with Faa1 long-chain-fatty-acid--CoA ligase, while *Drosophila* Snz and human SNX14 were associated with fatty acid desaturases Desat1 and SCD1 respectively [10, 23],

suggesting a role in lipid and fatty acid regulation. In line with this, members of the SNX-RGS protein family have almost universally been shown localize to ER-lipid-droplet (LD) contact sites with other organelles (**Fig. 1B**). Yeast Mdm1 localizes to nucleus-LD-vacuole junctions[8, 12]. Human SNX14 localizes to ER-LD contacts[7], and SNX19 has recently been shown to localize to ER-lysosome contacts, where it also contacts LDs [9].

A better understanding of the function(s) of the SNX-RGS proteins requires further knowledge of their structures and molecular interactions. Here we have solved the crystal structure of the human SNX25 RGS domain, revealing a typical α-helical RGS fold. Building on this we took advantage of the recently developed AlphaFold2 machine learning (ML)-based structural prediction algorithm [24] and the related AlphaFold2 database [25] to examine the structures of the full-length SNX-RGS proteins. These analyses revealed an intriguing structural fold not previously seen, that is formed by intramolecular association of the distal PXA and PXC domains. A remarkable feature of this structure is the presence of a large hydrophobic cavity that is likely to mediate binding of lipids or other fatty acid derived molecules, and we propose that the SNX-RGS proteins may represent a previously undefined class of lipid transfer proteins (LTPs), based on their structures and known localisation and interactions. Further, using AlphaFold2 we identify an additional yeast PX domain-containing protein Lec1/Ypr097w that contains another likely lipid-binding structure that shares superficial resemblance but is not directly related to the SNX-RGS proteins. This work provides a new understanding of the likely function of the SNX-RGS proteins, presents a structural bioinformatics comparison of novel lipid binding domains across various eukaryotic species, and further demonstrates the utility of new ML-based structure prediction programs such as AlphaFold2 to develop novel structural and functional insights.

Results

Crystal structure of the human SNX25 RGS domain

The SNX-RGS family members are typically represented as modular proteins with four to five distinct conserved domains (**Fig. 1A and 1B**) [4, 5, 8, 9, 12, 14, 15, 19, 21, 23, 26]. Our previous studies examined the structures and phosphoinositide binding properties of the PX domains of SNX14, SNX19 and SNX25 using NMR and X-ray crystallography [1, 5, 10]. Here we have also solved the structure of the human SNX25 RGS domain by X-ray crystallography in two crystal forms (**Fig. 1C**; **Fig. S1; Table S1**). The wildtype sequence of the RGS domain crystallises with a non-native disulfide bond formed by Cys526 between adjacent monomers in the crystal lattice which distorts the orientation of the helix α5. Mutation of Cys526 to alanine (C526A) prevents this bond formation and allows the domain to crystallise with a more typical α5 conformation similar for example to a canonical RGS domain such as RGS16 (PDB ID 2IK8)[27] (**Fig. 1D**).

We next investigated how similar the experimentally determined SNX25 RGS domain was to the predicted structure from the AlphaFold2 database [25]. The structure predicted by AlphaFold2 is remarkably close to the SNX25 crystal structure, with an overall root-mean-squared-deviation (RMSD) of 0.8 Å over 107 C_α atoms (**Fig. 1E**). Side chain conformations in general were likewise closely aligned between the experimental and predicted structures (**Fig. S2**). The structures of the human SNX13 and SNX14 RGS domains from the AlphaFold2 database are generally similar in structure to the experimental SNX25 RGS domain structure as expected, but with some notable differences (**Fig. 1F**). In general, the three domains have relatively low sequence homology, and this is reflected for example in a significantly longer α3-α4 extension in SNX13 compared to the other two proteins and (**Fig. 1F and 1G**). The experimental structures of the PX domains of SNX14, SNX19 and SNX25 were also compared with their respective AlphaFold2 predictions and in general, the AlphaFold2 models were found to be highly similar with only minor differences in sidechain positions and in some flexible loop regions. (**Fig. S2**). Overall, these comparisons confirm that the AlphaFold2 predictions of selected SNX-RGS domains are accurate models comparable to their known experimental structures, including the RGS domain of SNX25 for which no detailed structural information was available prior to this study.

AlphaFold2 predictions of the yeast, fly and human SNX-RGS proteins reveal conserved PXA and PXC domain interactions

Despite these structural insights into the SNX-RGS proteins, the structures of the N-terminal PXA and PXC domains bear no obvious sequence similarity to other proteins and their structures have not been determined, and the overall architecture of the proteins is unknown. We have so far been unsuccessful in expressing and purifying either full-length SNX-RGS proteins or any of their individual PXA and PXC domains in suitable quantities for structural or biochemical analyses. As the recently released AlphaFold2 database [25] includes predictions of the proteomes of multiple eukaryotic species from yeast to humans, we therefore took advantage of this to examine the

predicted structures of full-length SNX-RGS proteins across different species (**Fig. 2; Fig. S2; Fig. S3**).

The plots of the predicted alignment errors (PAE) of the AlphaFold2 models provide important insights into the organisation of the full-length proteins (**Fig. 2A**). The PAE measures AlphaFold's expected position error at residue 'x' if the predicted and true structures were aligned on residue 'y'. Square regions of the plot along the diagonal indicate the presence of globular domains where residues are structurally correlated with each other, while off-diagonal low PAE regions for residue pairs from two different domains is highly indicative of stable inter-domain interactions. The PAE plots for the human SNX-RGS proteins display expected low PAE regions for each of the four cytoplasmic domains downstream of the N-terminal ER membrane anchor (except for SNX19 which lacks an RGS domain). Notably, even though the N-terminal PXA domain and C-terminal PXC domains are around 500 amino-acids apart in the primary sequence, these two domains are invariably predicted to be tightly correlated with each other. Further, *Drosophila* Snz and yeast Mdm1p homologues show similar high degree of structural correlation between their PXA and PXC domains.

The AlphaFold2 predicted atomic structures of the human SNX-RGS proteins are shown in **Fig. 2B** and **Fig. S3**. Consistent with secondary structure predictions [5] both the PXA and PXC domains are entirely α-helical, and as indicated by the PAE plots they are tightly associated with each other. Indeed, they form an overall structure with a highly entwined topology, which is structurally similar in all four proteins. Relative to this 'core' structure the N-terminal membrane anchor, and intervening RGS and PX domains adopt diverse orientations pointing to a substantial degree of structural dynamics in the full-length proteins. Note that the human SNX25 sequence in the AlphaFold2 database (Uniprot ID Q9H3E2) lacks the N-terminal membrane anchor as previously identified [5, 28] (Uniprot A0A494C0S0). We also performed predictions of the full SNX25 sequence using the online ColabFold pipeline [29], which show a similar structure but with the additional N-terminal α-helical membrane (**Fig. S4**). As for predictions across the family, the three separate models of human SNX25 generated by ColabFold suggest that the transmembrane anchor, RGS and PX domains adopt relatively flexible orientations with respect to the core PXA-PXC structure. Although not shown, both full-length Mdm1p and Snz have analogous architectures with a core structure composed of the PXA and PXC domains that adopt a flexible orientation with respect to the transmembrane, RGS and PX domains. Overall, the predicted structures of the RGS-SNX proteins are highly consistent across homologues and species and reveal a novel intramolecular association between the previously uncharacterised PXA and PXC domains.

The PXA and PXC domains form an interwoven domain with a large conserved hydrophobic channel

For a closer examination of the PXA-PXC structure we focus on these domains from the SNX13 model (**Fig. 3A; Movie S1**). Both the PXA and PXC domains are entirely α-helical as previously

predicted [5] and do not share any obvious structural similarity with previously characterised proteins. A notable feature of their structures is that they are intimately intertwined in such a way that neither domain would be expected to form their correct fold independently of the other. The threading of their α-helical structures together forms an overall elongated structure with a length of around 80 Å. The same α-helical topology and pattern of threading between the PXA and PXC domains is highly conserved and is seen in all SNX-RGS family members (**Fig. 3B**). Although the yeast protein Nvj3p (Ydr179w-a) is not found in higher eukaryotes, it was previously proposed to share weak sequence similarity with the PXA domain of the SNX-RGS proteins [8]. Interestingly the AlphaFold2 prediction of Nvj3p shows a very similar structure to the combined PXA and PXC fold of the SNX-RGS proteins (**Fig, 3C**). Although it lacks a transmembrane, RGS or PX domains, and has only a minimal linker sequence between the two halves of its structure, it shares the same overall topology indicating it is a bona fide PXA-PXC structure.

A visual inspection of the core PXA-PXC domains reveals the presence of a large channel that runs through much of the structure with an exposed surface composed of side chains of both domains (**Fig. 4; Fig. S5; Movie S1**). The channel is extremely hydrophobic, as seen from mapping surface hydrophobicity (**Fig. 4A; Fig. S5**) or from inspecting the details of the side chains lining the channel of human SNX13 as an example (**Fig. 4B; Fig. S5**). The entrance to this cavity is highly conserved indicating it is likely to be a critical functional element of the proteins (**Fig. 4A; Fig. S5**). The presence of such a hydrophobic channel or pocket is a common feature of the class of proteins known as lipid transfer proteins (LTPs) which can bind, extract and transport hydrophobic lipids and fatty acids from various compartments for non-vesicular trafficking [30-33] (**Fig. S5**). Given this structural similarity, coupled with the known localisation of the SNX-RGS proteins at ER membrane contact sites and their effects on lipid homeostasis, we propose that this channel in the SNX-RGS proteins is very likely to be a lipid binding pocket. Manual placement of a phosphatidylethanolamine lipid in the channel of SNX13 shows that the dimensions of this hydrophobic pocket in the SNX-RGS proteins is well suited for binding lipid acyl chains (**Fig. 4C; Movie S1**).

Several previous studies have shown that the SNX-RGS proteins are rapidly recruited to the surface of newly formed lipid droplets. This has been observed for Mdm1 in yeast [11], Snz in flies [10], and SNX13, SNX14, and SNX19 in mammalian cells [6, 7, 9, 13]. This provides strong support for a role in direct transfer of lipids or lipid metabolites at ER-LD contact sites. We tested whether the ability to localise to the LD surface is conserved in all human family members by analysing the localisation of SNX25 (**Fig. 5**). In A431 cells in the absence of stimulation SNX25 shows a relatively diffuse localisation, however on addition of oleic acid SNX25-FLAG is rapidly redistributed and shows a strong degree of localisation to the surface of newly formed LDs. Further work will be needed to examine if these ER-associated LDs are also in contact with other membrane compartments, but this data confirms that all SNX-RGS family members from yeast to humans has a strong affinity for the LD monolayer.

Structural bioinformatics of the SNX-RGS proteins

The DALI webserver [34] is now capable of searching the entire database of AlphaFold2 predictions [25] for structurally similar proteins. We performed a search using the core PXA-PXC structure of SNX25 to determine of any other proteins share similar folds **(Table S2**). This revealed several interesting observations. The PXA-PXC structure is found in proteins across metazoans, plants and yeast. Interestingly we did not find any similar structures in either *D. dictyostelium* or *P. falciparum* and they therefore do not seem to be present in all eukaryotes. We identified three proteins in the fission yeast *S. pombe* with this fold. Meiotically up-regulated gene 122 protein (Mug122) (O74444) lacks the RGS domain but has both transmembrane and central PX domains, while snx12 (Q9USN1) has each of the domains common to the SNX-RGS family. The protein annotated as Pxa1 (O14200) appears to be most similar to Nvj3, as it lacks transmembrane, PX and RGS domains. We also performed a BLAST search of the *Chaetomium thermophilum* genome, a thermophilic filamentous yeast popular with structural biologists because of the generally higher stability of the encoded proteins. The *C. thermophilum* genome encodes three proteins with putative PXA-PXC structures (**Table S2**), and these were confirmed by performing ColabFold predictions of their structures (not shown). One of these has all domains common to the SNX-RGS family, a second lacks the RGS domain like human SNX19, and the third has only the PXA-PXC domains like *S. cerevisiae* Nvj3.

Lec1/Ypr097w is another putative LTP found in yeast and ameoba

While examining the AlphaFold2 predicted structures of PX domain containing proteins in various species, and primed by the discovery of the PXA-PXC lipid-binding fold, we noticed that another yeast protein called Lec1/Ypr097w (Q06839) also possessed a superficially similar architecture. Lec1/Ypr097w is 1073 residues in length and is annotated to have a central PX domain between an N-terminal PXB domain and C-terminal domain of unknown function (DUF3818) (**Fig. 6A**). Analogously to the SNX-RGS proteins, the N- and C-terminal domains of Lec1/Ypr097w form α-helical structures with previously unknown folds that come together to form an intimately interwoven structure (**Fig. 6B; Fig. S6**). However, neither domain is directly related in sequence or structure to the PXA or PXC domains, and the way they interact with each other is also entirely different. Note that the term ‘PXB domain’ is also used for a C-terminal domain found in human SNX20 and SNX21 which is also completely unrelated in sequence or structure [35]. For clarity herein we will refer to the N-terminal Lec1/Ypr097w PXB domain as ‘PXYn’ and the C-terminal DUF3818 domain as ‘PXYc’, where PXY denotes ‘PX-associated domains from yeast’.

Most interestingly, the PXYn and PXYc domains of Lec1/Ypr097w also form a barrel-like structure that has a very large cavity surrounded by hydrophobic residues (**Fig. 6C**). The structure is conserved in homologues from *Schizosaccharomyces pombe* (Q9Y7N9), and *Candida albicans* (A0A1D8PQ52) (**Fig. 6E**). Although structurally unrelated to the SNX-RGS proteins, the general similarities to Lec1/Ypr097w and its homologues are striking; a central phosphoinositide binding PX domain flanked by N-terminal and C-terminal α-helical domains that come together to form a co-

folded intramolecular structure. Like the SNX-RGS proteins it is very likely based on its structural properties that the hydrophobic cavity formed by the PXYn-PXYc domains is also involved in binding to lipids or fatty acids. In each of the predicted structures the PX domain forms a loose plug over the large hydrophobic cavity suggesting a potential regulatory role for this domain in lipid binding (**Fig. S6C**). This may be relatively dynamic as several independent predictions of Lec1/Ypr097w using ColabFold suggest the PX domain can adopt orientations that either block the pocket or leave the cavity exposed (**Movie S2**).

The PXYn-PXYc structure appears to be much more restricted through evolution, as we only found examples in yeast, including two proteins in *S. pombe*, as well as a protein in *D. dictyostelium. S. pombe* PX domain-containing protein C1450.12 (Q9Y7N9) is highly akin to Lec1/Ypr097w from *S. cerevisiae,* while uncharacterized protein C663.15c (O74521) appears to lack a central PX domain between the two halves of the PXYn-PXYc structure (**Fig. S6**; **Fig. S7**). The *D. discoideum* DUF3818 domain-containing protein (Q54JB8) also lacks a PX domain, and while its C-terminal PXYc domain is like Lec1/Ypr197w, its N-terminal region is structurally divergent from the PXYa domain (**Fig. S7**). Despite this it also possesses a predicted large hydrophobic cavity and therefore may share a functional relationship. Again, we performed a BLAST search of the *C. thermophilum* protein sequences using both *S. pombe* sequences as inputs and found one protein homologous to the Lec1/Ypr097w-like Q9Y7N9 (G0SCP5) and a protein lacking the PX domain like O74521 (G0S997). These discoveries further demonstrate the utility of AlphaFold2 for predicting novel structures, and its power in aiding evolutionary analyses of protein homology.

Discussion

With the recent advances made by the AlphaFold2 machine learning algorithm and public release of the AlphaFold2 database [24, 25], it is now possible to perform detailed structural analyses and evolutionary comparisons of protein families across diverse species. In this work we have solved the crystal structure of the human SNX25 RGS domain and gone on to examine the previously unexplored structures of the full-length SNX-RGS proteins. Although the overall structures of these proteins display conformational flexibility with respect to their N-terminal transmembrane ER tethers, in contrast to previous pictures based on primary and secondary structure predictions the disparate N- and C-terminal PXA and PXC domains form a highly entwined intramolecular interaction generating a compact cylindrical α-helical structure. The most notable feature of a central hydrophobic cavity with a highly conserved entrance and surface lining. This leads us to conclude that the likely function of the PXA-PXC structure will be to bind and transport lipids or fatty acids, and that the proteins represent a distinct class of ER membrane-anchored LTPs.

Broader analysis of other PX domain-containing protein family members indicate that the yeast protein Lec1/Ypr097w and related molecules share a superficially similar LTP structure and function. Here the central PX domain is found between N-terminal PXYn and C-terminal PXYc domains that also become conjoined to form an all α-helical barrel-like structure with a large hydrophobic cavity, although their sequences are not related, and their structural details are entirely different. As we were performing these studies, a preprint was published by the Schuldiner lab that identified novel proteins that reside at or regulate formation of membrane contact sites in yeast [36]. In this work Ypr097w was identified to localise to contacts between lipid droplets and the cell surface and renamed Lec1 (Lipid Droplet Ergosterol Cortex 1) because of a potential role in ergosterol transport. This team also analysed the AlphaFold2 predicted structure of Lec1/Ypr097w and came to similar conclusions about its potential role as a novel lipid transfer protein (LTP), as well as identifying structural similarities to the SNX-RGS proteins.

What do these new structural insights suggest about the function(s) of the SNX-RGS proteins? The hydrophobic nature and conserved surface of the large interior cavity found in the PXA-PXC domain structure indicates they are very likely to mediate the binding of lipids, fatty acids or other lipid-derived metabolites. A general finding is that the SNX-RGS proteins reside in the ER at steady state via their N-terminal membrane anchor, and colocalise with LDs particularly under conditions of new LD synthesis. They are typically found to be enriched in regions of the ER and ER-associated LDs that are in contact with other membranes where they can play a tethering role via their phosphoinositide-binding PX domain. These membrane contacts include the cell surface in the fat body cells of fruit flies [10], the nuclear-vacuolar junction in budding yeast [8, 11, 12], and with endolysosomal compartments in mammalian cells [6, 7, 9, 13]. This localisation to ER-LD-membrane contact sites is dramatically altered under conditions of lipid flux (e.g. addition of excess free fatty acids to cells [7, 9]), and the SNX-RGS proteins play important roles in lipid metabolism and homeostasis.

The effects of mutations, knockouts or overexpression of SNX-RGS proteins are complex, but at the cellular level their depletion commonly leads to higher levels of neutral triacylglycerol (TAG) and accumulation of cholesterol in late endosomal compartments [6, 11, 13, 16]. Yeast Mdm1 and human SNX14 have been associated with Faa1 and ACSL3 respectively, long-chain-fatty-acid--CoA ligases that activate fatty acids for synthesis of cellular lipids and degradation via β-oxidation [7, 11]. SNX14 was also found to associate with stearoyl-CoA desaturase (SCD1) that catalyzes the insertion of a cis double bond at the Δ-9 position of fatty acyl-CoA substrates [23]. Knockout of SNX14 causes cells to be sensitive to saturated fatty acid (SFA)-induced toxicity and increased levels of SFAs are incorporated into phospholipids, which can be rescued by SCD1 overexpression. A recent CRISPR screen for genes that influence cholesterol homeostasis found SNX13 (and SNX14) as a prominent hit [13]. Niemann Pick type C (NPC) disease is caused by genetic defects in the lysosomal cholesterol transport system, Niemann Pick C1 and C2 proteins (NPC1 and NPC2) leading to endolysosomal accumulation of cholesterol and Bis(monoacylglycero)phosphate (BMP, also known as LBPA) [37, 38], which is essential for normal cholesterol transport out of the endolysosomal compartment [39-43]. In screens using the NPC1 inhibitor U18666A cholesterol accumulation could be reversed by depletion of SNX13, and cholesterol was transported in an NPC1-independent manner to other compartments including the plasma membrane. Surprisingly SNX13 depletion led to a significant increase in total lysosomal BMP, which may be important for the observed NPC1-independent cholesterol transport [44, 45]. Although highly speculative one possibility is that the PXA-PXC domains are important for transport and regulation of lipids such as BMP at contacts between the ER and endolysosomal compartments. Identifying specific ligand(s) and whether the different family members have alternative binding activities will require further experimentation.

The PXA-PXC domains were previously considered as conserved but essentially independent structures in the SNX-RGS family of proteins. From the analyses presented here it is clear that these two domains are intimately associated in a single structural unit and that neither the PXA or PXC sequences are likely to fold correctly in the absence of the other. Previously, several studies have examined the impact of deleting these regions on SNX-RGS protein function. While deletion of either domain generally renders the proteins non-functional as expected, there is evidence that the two sequences contain regions that are important for LD targeting and potentially protein interactions. In yeast, it was found that the PXA domain of Mdm1 could define sites of LD formation, and could recruit the Faa1 fatty-acid—CoA ligase [12]. In contrast, several reports have shown that the PXC domain by itself is required and sufficient for targeting to the surface of newly formed LDs in both flies and mammals [7, 10, 13]. Deleting an amphipathic helical region in SNX14 PXC (residues 801-819), which is predicted to sit at the entrance of the hydrophobic cavity of the assembled PXA-PXC structure (**Fig. S5**), prevents LD attachment. These studies suggest that even in the absence of the PXA domain, and thus proper folding of the PXA-PXC structure, there are sequences in the PXC domain capable of direct association with the LD monolayer.

In conclusion, although further experiments are required to define the functional lipid interactions of the SNX-RGS proteins, and their distantly related Lec1/Ypr097w cousins, these structural studies provide strong evidence for a novel lipid transport activity mediated by these conserved molecules. As a final general observation, this work further confirms the ability of AlphaFold2 to predict not only already known folds, but to define completely new structures without previously recognised topologies and thus provide new insights into biological function.

Materials and methods

Cloning, expression purification of the SNX25 RGS domain

The genes encoding the human SNX25 RGS domain (residues 446-569; A0A494C0S0) and the SNX25 RGS(C526A) mutant were synthesised by Gene Universal (USA) and cloned into the expression plasmid of pGEX-4T-2 for bacterial expression with an N-terminal GST tag and thrombin cleavage site. was construct was transformed into *Escherichia coli* BL21-CodonPlus (DE3)-RIL cells and plated on lysogeny-broth (LB) agar plates supplemented with ampicillin (0.1 mg/mL). Single colonies were then used to inoculate 50 mL of LB medium containing ampicillin and the culture was grown overnight at 37°C with shaking at 180 rpm. The following day 1 L of LB medium containing ampicillin (0.1 mg/mL) was inoculated using 10 ml of the overnight culture. Cells were then grown at 37°C with shaking at 200 rpm to an optical density of 0.8-0.9 at 600 nm and the protein expression was induced by adding 0.5 mM IPTG (isopropyl-β-D-thiogalactopyranoside). Expression cultures were incubated at 20°C overnight with shaking and the cells were harvested the next day by centrifugation at 4000 rpm for 15 min using a Beckman rotor JLA 8.100. Cell pellets were resuspended in 20 mL (for cell pellet from 1 L culture) of lysis buffer (50 mM HEPES, (pH 7.5), 500 mM NaCl, 5% glycerol, Benzamidine (0.1 mg/mL), and DNAse (0.1 mg/mL)). Resuspended cells were lysed by using the cell disrupter (Constant systems, LTD, UK, TS-Series) and the soluble fraction containing the protein was separated from cell debris by centrifugation at 18,000 rpm for 30 min at 4°C. The soluble fraction was first purified by affinity chromatography using Glutathione Sepharose 4B resin (GE Healthcare) and the GST tag was cleaved on column by incubating the protein with Thrombin (Sigma Aldrich) overnight at 4°C. The next day the protein was eluted using 50 mM HEPES (pH 7.5), 200 mM NaCl. The eluted protein was then concentrated and further purified by gel filtration chromatography (Superdex 75 (16/600), GE Healthcare) using 50 mM HEPES, pH 7.5, 200 mM NaCl, 0.5 mM TCEP (tri(2-carboxyethyl)phosphine) and the fractions corresponding to SNX25 RGS domains were analysed by SDS PAGE.

Crystallisation and structure determination of the SNX25 RGS domain

Purified proteins were concentrated to 12 mg/mL for crystallisation. Initially 96 well crystallisation screens were set up using the hanging drop method with a Mosquito Liquid Handling robot (TTP LabTech). The optimised diffraction-quality crystals were obtained in 24-well plates using hanging drop method. The SNX25 RGS domain was crystallised in 30% PEG 4000, 0.2 M Sodium acetate, 0.1 M Tris pH 8.5, while the SNX25 RGS(C526A) domain was crystallised in 18% PEG 8000, 0.1 M calcium acetate, 0.1 M sodium cacodylate, pH 6.5. Diffraction data was collected at the Australian Synchrotron MX2 beamline (Clayton, VIC, Australia), integrated with XDS [46] and scaled with AIMLESS software [47]. The SNX25 RGS structure was solved by molecular replacement using the RGS1 structure as input (PDB ID 2BV1; [27]) in PHASER [48] and the resulting model was rebuilt and refined by using COOT [49] and PHENIX [50] respectively. The SNX25 RGS(C526A) was solved similarly by molecular replacement using the wild-type SNX25 RGS domain as input.

Protein structural prediction, modelling and visualisation

The structural predictions of the full-length proteins from different species including *S. cerevisiae, D. melanogaster, C. elegans, P. falciparum, D. discoideum, A. thaliana, S. pombe, M. janaschii, D. rerio* and *H. sapiens* analysed in this study were obtained from the AlphaFold2 database (https://alphafold.ebi.ac.uk; [24, 25]). To generate predicted models of human SNX25 including the previously annotated transmembrane domain [5, 28] and the *S. cerevisiae* Lec1/Ypr097w to assess the flexibility of its PX domain, we used the AlphaFold2 neural-network [24] implemented within the freely accessible ColabFold pipeline [29]. For each modelling experiment ColabFold was executed using default settings where multiple sequence alignments were generated with MMseqs2 [51] and structural relaxation of final peptide geometry was performed with Amber [52] to generate three models per structure. Sequence conservation was mapped onto the modelled SNX27 FERM domain structure with Consurf [53]. All structural images were made with Pymol (Schrodinger, USA; https://pymol.org/2/).

Localisation of SNX25 in A431 cells

The human SNX25 open reading frame (ORF) cloned into the pcDNA3.1+-C DYK was obtained from Genscript (NM_031953). This expresses a C-terminally tagged SNX25 (SNX25-FLAG) that has the same sequence as the AlphaFold2 model shown in **Fig. 2B**. Note however, that it lacks the predicted N-terminal ER anchor found in the SNX25 isoform A0A494C0S0 that is modelled in **Fig. S4**. A431 cells were maintained in DMEM medium (Gibco Life technologies) supplemented with 10% foetal bovine serum (FBS) and Penicillin/Streptomycin. A431 cell line was sourced from ATCC and tested regularly for mycoplasma contamination. Cells (2 X 10^5) were plated in 6 well culture dishes (Nunc™, Cat. No. 140675, Culture area—9.6 cm^2) and grown to 50% confluency before transfection of SNX25-FLAG construct using lipofectamine 3000 (Invitrogen) as per manufacturer's protocol. Cells were shifted to DMEM media containing oleic acid (50 µg/ml) 4 h post transfection and incubated for another 20 h. DMEM with oleic acid was prepared by mixing oleic acid (Sigma Aldrich, Cat. No. O-1008) in DMEM medium containing bovine serum albumin (fatty acid free) to final oleic acid concentration (50 µg/ml) with rigorous shaking at 37°C to avoid precipitation. Cells were fixed 24 h post-transfection with 4% paraformaldehyde in phosphate buffered saline (PBS) at 4°C and subsequently permeabilised with 0.1% Triton X-100 in PBS for 7 min. Cells were probed with FLAG antibody (Sigma Aldrich Cat. No. F7425, 2 µg/ml) and anti-Rabbit secondary antibody Alexa Fluor®488 conjugate (4 µg/ml) to visualise SNX-25 protein. Cells were finally labelled with nile red solution (100 nM) to visualise lipid droplets. Confocal images (1024 X 1024) were acquired on a Zeiss inverted LSM 880 microscope coupled with fast airyscan detector (Carl Zeiss, Inc) equipped with 63X oil immersion objective, NA 1.4.

Acknowledgements

We thank Prof. Rob Parton for the gift of A431 cells, and antibodies for LD staining. This work was supported by funding from the Australian Research Council (ARC) (DP200102551). BMC is supported by an NHMRC Senior Research Fellowship (APP1136021). WMH is supported by funds from the Welch Foundation (I-1873), the NIH NIGMS (GM119768), NIDDK (DK126887), Ara Parseghian Medical Research Fund, and the UT Southwestern Endowed Scholars Program.

Author contributions

Conceptualization, WMH, BMC.; Methodology, SW, BP, VT, BMC; Investigation, SW, BP, VT, BMC; Writing – Original Draft, SW, BMC; Writing – Review & Editing, SW, BP, VT, BMC; Funding Acquisition, WMH, BMC; Supervision, WMH, BMC.

Conflict of interest

Authors declare that they have no conflict of interest.

References

1. Chandra, M. *et al.* Classification of the human phox homology (PX) domains based on their phosphoinositide binding specificities. *Nat Commun* **10**, 1528 (2019).

2. Chandra, M. & Collins, B.M. The Phox Homology (PX) Domain. *Adv Exp Med Biol* **1111**, 1-17 (2019).

3. Teasdale, R.D. & Collins, B.M. Insights into the PX (phox-homology) domain and SNX (sorting nexin) protein families: structures, functions and roles in disease. *Biochem J* **441**, 39-59 (2012).

4. Amatya, B. *et al.* SNX-PXA-RGS-PXC Subfamily of SNXs in the Regulation of Receptor-Mediated Signaling and Membrane Trafficking. *Int J Mol Sci* **22** (2021).

5. Mas, C. *et al.* Structural basis for different phosphoinositide specificities of the PX domains of sorting nexins regulating G-protein signaling. *J Biol Chem* **289**, 28554-28568 (2014).

6. Bryant, D. *et al.* SNX14 mutations affect endoplasmic reticulum-associated neutral lipid metabolism in autosomal recessive spinocerebellar ataxia 20. *Hum Mol Genet* **27**, 1927-1940 (2018).

7. Datta, S., Liu, Y., Hariri, H., Bowerman, J. & Henne, W.M. Cerebellar ataxia disease-associated Snx14 promotes lipid droplet growth at ER-droplet contacts. *J Cell Biol* **218**, 1335-1351 (2019).

8. Henne, W.M. *et al.* Mdm1/Snx13 is a novel ER-endolysosomal interorganelle tethering protein. *J Cell Biol* **210**, 541-551 (2015).

9. Saric, A. *et al.* SNX19 restricts endolysosome motility through contacts with the endoplasmic reticulum. *Nat Commun* **12**, 4552 (2021).

10. Ugrankar, R. *et al.* Drosophila Snazarus Regulates a Lipid Droplet Population at Plasma Membrane-Droplet Contacts in Adipocytes. *Dev Cell* **50**, 557-572 e555 (2019).

11. Hariri, H. *et al.* Lipid droplet biogenesis is spatially coordinated at ER-vacuole contacts under nutritional stress. *EMBO Rep* **19**, 57-72 (2018).

12. Hariri, H. *et al.* Mdm1 maintains endoplasmic reticulum homeostasis by spatially regulating lipid droplet biogenesis. *J Cell Biol* **218**, 1319-1334 (2019).

13. Lu, A., Hsieh, F., Enrich, C. & Pfeffer, S.R. CRISPR screens for lipid regulators reveal a role for ER-bound SNX13 in lysosomal cholesterol export. *bioRxiv*, 2021.2005.2010.443492 (2021).

14. Akizu, N. *et al.* Biallelic mutations in SNX14 cause a syndromic form of cerebellar atrophy and lysosome-autophagosome dysfunction. *Nat Genet* **47**, 528-534 (2015).

15. Thomas, A.C. *et al.* Mutations in SNX14 cause a distinctive autosomal-recessive cerebellar ataxia and intellectual disability syndrome. *Am J Hum Genet* **95**, 611-621 (2014).

16. Bryant, D. *et al.* Diverse species-specific phenotypic consequences of loss of function sorting nexin 14 mutations. *Sci Rep* **10**, 13763 (2020).

17. Bult, C.J. *et al.* Mouse Genome Database (MGD) 2019. *Nucleic Acids Res* **47**, D801-D806 (2019).

18. Zheng, B. *et al.* Essential role of RGS-PX1/sorting nexin 13 in mouse development and regulation of endocytosis dynamics. *Proc Natl Acad Sci U S A* **103**, 16776-16781 (2006).

19. Suh, J.M. *et al.* An RGS-containing sorting nexin controls Drosophila lifespan. *PLoS One* **3**, e2152 (2008).

20. Ha, C.M. *et al.* SNX14 is a bifunctional negative regulator for neuronal 5-HT6 receptor signaling. *J Cell Sci* **128**, 1848-1861 (2015).

21. Zheng, B. *et al.* RGS-PX1, a GAP for GalphaS and sorting nexin in vesicular trafficking. *Science* **294**, 1939-1942 (2001).

22. Gillingham, A.K., Bertram, J., Begum, F. & Munro, S. In vivo identification of GTPase interactors by mitochondrial relocalization and proximity biotinylation. *Elife* **8** (2019).

23. Datta, S. *et al.* Snx14 proximity labeling reveals a role in saturated fatty acid metabolism and ER homeostasis defective in SCAR20 disease. *Proc Natl Acad Sci U S A* (2020).

24. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).

25. Tunyasuvunakool, K. *et al.* Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590-596 (2021).

26. Li, J. *et al.* SNX13 reduction mediates heart failure through degradative sorting of apoptosis repressor with caspase recruitment domain. *Nat Commun* **5**, 5177 (2014).

27. Soundararajan, M. *et al.* Structural diversity in the RGS domain and its interaction with heterotrimeric G protein alpha-subunits. *Proc Natl Acad Sci U S A* **105**, 6457-6462 (2008).

28. Lauzier, A. *et al.* Snazarus and its human ortholog SNX25 modulate autophagic flux. *J Cell Sci* (2021).

29. Mirdita, M., Ovchinnikov, S. & Steinegger, M. ColabFold - Making protein folding accessible to all. . *bioRxiv* (2021).

30. Malinina, L., Patel, D.J. & Brown, R.E. How alpha-Helical Motifs Form Functionally Diverse Lipid-Binding Compartments. *Annu Rev Biochem* **86**, 609-636 (2017).

31. Wong, L.H., Copic, A. & Levine, T.P. Advances on the Transfer of Lipids by Lipid Transfer Proteins. *Trends Biochem Sci* **42**, 516-530 (2017).

32. 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).

33. Chiapparino, A., Maeda, K., Turei, D., Saez-Rodriguez, J. & Gavin, A.C. The orchestra of lipid-transfer proteins at the crossroads between metabolism and signaling. *Prog Lipid Res* **61**, 30-39 (2016).

34. Holm, L. Using Dali for Protein Structure Comparison. *Methods Mol Biol* **2112**, 29-42 (2020).

35. Clairfeuille, T., Norwood, S.J., Qi, X., Teasdale, R.D. & Collins, B.M. Structure and Membrane Binding Properties of the Endosomal Tetratricopeptide Repeat (TPR) Domain-containing Sorting Nexins SNX20 and SNX21. *J Biol Chem* **290**, 14504-14517 (2015).

36. Castro, I.G. *et al.* Systematic analysis of membrane contact sites in Saccharomyces cerevisiae uncovers modulators of cellular lipid distribution. *bioRxiv*, 2021.2010.2017.464712 (2021).

37. Pentchev, P.G. Niemann-Pick C research from mouse to gene. *Biochim Biophys Acta* **1685**, 3-7 (2004).

38. Wheeler, S. & Sillence, D.J. Niemann-Pick type C disease: cellular pathology and pharmacotherapy. *J Neurochem* **153**, 674-692 (2020).

39. Chevallier, J. *et al.* Lysobisphosphatidic acid controls endosomal cholesterol levels. *J Biol Chem* **283**, 27871-27880 (2008).

40. Gallala, H.D. & Sandhoff, K. Biological function of the cellular lipid BMP-BMP as a key activator for cholesterol sorting and membrane digestion. *Neurochem Res* **36**, 1594-1600 (2011).

41. Gruenberg, J. Life in the lumen: The multivesicular endosome. *Traffic* **21**, 76-93 (2020).

42. Kobayashi, T. *et al.* Late endosomal membranes rich in lysobisphosphatidic acid regulate cholesterol transport. *Nat Cell Biol* **1**, 113-118 (1999).

43. McCauliff, L.A. *et al.* Intracellular cholesterol trafficking is dependent upon NPC2 interaction with lysobisphosphatidic acid. *Elife* **8** (2019).

44. Ilnytska, O. *et al.* Enrichment of NPC1-deficient cells with the lipid LBPA stimulates autophagy, improves lysosomal function, and reduces cholesterol storage. *J Biol Chem* **297**, 100813 (2021).

45. Moreau, D. *et al.* Drug-induced increase in lysobisphosphatidic acid reduces the cholesterol overload in Niemann-Pick type C cells and mice. *EMBO Rep* **20**, e47055 (2019).

46. Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132 (2010).

47. Evans, P.R. & Murshudov, G.N. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**, 1204-1214 (2013).

48. McCoy, A.J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).

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

50. Adams, P.D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221 (2010).

51. Mirdita, M., Steinegger, M. & Soding, J. MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinformatics* **35**, 2856-2858 (2019).

52. Hornak, V. *et al.* Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **65**, 712-725 (2006).

53. Ashkenazy, H. *et al.* ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* **44**, W344-350 (2016).

54. Samygina, V.R. *et al.* Enhanced selectivity for sulfatide by engineered human glycolipid transfer protein. *Structure* **19**, 1644-1654 (2011).

55. Schauder, C.M. *et al.* Structure of a lipid-bound extended synaptotagmin indicates a role in lipid transfer. *Nature* **510**, 552-555 (2014).

551

Figure legends

Figure 1. The SNX-RGS protein family.

(A) Domains of human, *Drosophila*, and *S. cerevisiae* SNX-RGS family members. Putative ligands for each domain are indicated above. SNX14 mutations causing cerebellar ataxia are shown with dashed lines [14, 15]. Boxes indicate domains with experimental structures determined. (**B**) The current working model for the architecture, localisation and function of the SNX-RGS protein family and their potential roles at the interface between the ER, LDs, and membranes of the endolysosomal system. (**C**) Crystal structure of human SNX25(C362A) RGS domain in cartoon diagram. Helical secondary structure elements are indicated. (**D**) The crystal structure of human SNX25(C362A) RGS domain (light blue) is aligned with the canonical RGS domain of RGS16 (pink)(PDB ID 2IK8)[27] and shows the expected topology. (**E**) Cα ribbon diagram showing the overlay of the SNX25(C362A) RGS domain (light blue) with the same sequence predicted in the AlphaFold2 database (red) (Q9H3E2). (**F**) Cartoon representation of SNX25(C362A) RGS domain compared with the SNX13 and SNX14 RGS domains from the AlphaFold2 database (Q9Y5W8; Q9Y5W7). (**G**) Sequence alignment of human SNX13, SNX14 and SNX25 RGS domains based on structural comparisons. The secondary structure of SNX25 derived from its crystal structure is shown schematically above the alignment, while the α-helical regions of the three proteins from their AlphaFold2 predictions are shaded blue within the alignment.

Figure 2. Structures of the SNX-RGS proteins predicted by AlphaFold2.

(A) Predicted Alignment Error (PAE) plots from the AlphaFold2 database [25] are shown for human fly and yeast SNX-RGS proteins. In these plots all SNX-RGS proteins show a strong degree of correlation between the PXA and PXC domain suggesting these two domains are physically associated. (**B**) The predicted structures of human SNX-RGS proteins from the AlphaFold2 database. The PXA domain is coloured green, RGS domain in light blue, PX domain in blue and PXC domain in orange. The predicted TM domain and any unstructured linker regions are coloured grey. The structures are shown in the same orientation after alignments based on the PXA and PXC core region. This shows that the two domains are intimately entwined with each other, whereas the TM, RGS and PX domains are predicted to have flexible orientations relative to these domains.

Figure 3. The PXA and PXC domains combine form an intertwined α-helical structure.

(**A**) Structure of the PXA and PXC domains of human SNX13 predicted by AlphaFold2 in green and orange respectively, shown with cylinders for α-helices. The two domains are predicted to be tightly interwoven. (**B**) An overlay of the core PXA-PXC domains of human, yeast and fly SNX-RGS proteins shows that all predicted structures have the same topology. (**C**) Predicted structure of *S. cerevisiae* Nvj3 with the regions expected to be similar to PXA and PXC domains coloured green and orange and the linker shown in grey.

Figure 4. The PXA-PXC structure forms a conserved hydrophobic cavity with potential for lipid binding.

(**A**) The predicted structures of PXA-PXC domains from indicated species are shown in ribbon representation (left), surface coloured by hydrophobicity (middle), and sequence conservation (right). All PXA-PXC structures have a highly conserved hydrophobic tunnel. (**B**) Structure of PXA and PXC domains of SNX13 with sidechains of the putative lipid binding pocket shown. (**C**) The PXA and PXC domains of SNX13 is shown in close-up with its surface coloured for hydrophobicity as in (**A**). A phosphatidylethanolamine lipid has been docked manually in the putative lipid binding pocket to give perspective on the dimensions of the channel.

Figure 5. Human SNX25 is localises to newly synthesised LDs following oleic acid addition

A431 cells were transfected with SNX25-FLAG and either left untreated or treated with 50 µg/ml oleic acid to stimulate LD formation. After oleic acid addition SNX25-FLAG undergoes rapid redistribution to the periphery of newly generated LDs LD periphery. Scale bars represent 10 μm.

Figure 6. The yeast Lec1/Ypr097w protein forms a unique structure with a hydrophobic cavity.

(**A**) Domain structure of the *S. cerevisiae* Lec1/Ypr097w protein. (**B**) AlphaFold2 structural prediction of the Lec1/Ypr097w protein with the PXYn domain in green, PX domain in blue and PXYc domain in orange. (**C**) The PXYn and PXYc domains are shown with α-helices in cylinder representation and the PX domain removed for clarity. The domains encompass a large conserved hydrophobic cavity and like the SNX-RGS proteins may be a potential lipid binding protein.

Supplementary Information

Figure S1. Crystal structure of the SNX25 RGS domain.

(**A**) Crystal structure of the wild-type human SNX25 RGS domain. (**B**) Alignment of the two chains in the asymmetric unit of wild-type SNX25 RGS domain. (**C**) The position of the non-native disulfide bond formed between adjacent chains in the crystal lattice. This causes the change in orientation of the α5 helix. (**D**) Crystal structure of the SNX25(C526A) mutant protein. (**E**) Alignment of the two chains in the asymmetric unit of SNX25(C526A) RGS domain. (**F**) Structural alignment of the wild-type and C526A SNX25 RGS domain structures.

Figure S2. Comparison of experimental SNX-RGS structures with their AlphaFold2 predictions.

(**A**) Crystal structure of the human SNX25(C326A) RGS domain (red, this study) aligned with the AlphaFold2 prediction (light blue). (**B**) Crystal structures of human SNX14 PX domain (green, PDB 4PQP; magenta, 4PQO) with the AlphaFold2 prediction (blue). (**C**) Crystal structure of mouse SNX19 PX domain (pink, PDB 4P2I) aligned with the AlphaFold2 prediction (blue). (**D**) NMR structure of human SNX25 PX domain (cyan, PDB 4PQP) aligned with the AlphaFold2 prediction (blue).

Figure S3. Additional details of the SNX13 predicted full length structure.

(**A**) AlphaFold2 prediction of full-length human SNX13 protein with individual domains coloured as in **Fig. 2B** with the TM domain and linker regions (grey), PXA domain (green), RGS domain (light blue), PX domain (blue), and PXC domain (orange). (**B**) AlphaFold2 prediction of full-length human SNX13 protein coloured according to the pLDDT score. (**C**) The AlphaFold2 predictions of human, mouse and zebrafish SNX13 proteins were aligned based on the core PXA-PXC structure.

Figure S4. AlphaFold2 prediction of full-length human SNX25 protein.

(**A**) Top ranked structure of full length human SNX25 [5] (Uniprot ID A0A494C0S0) predicted by ColabFold coloured according to pLDDT score. (**B**) Top ranked structure of human SNX25 predicted by ColabFold coloured according to the indicated domains. (**C**) Alignment based on the core PXA-PXC structure of the five structural predictions of human SNX25 from ColabFold.

Figure S5. Conserved hydrophobic cavities in all human, yeast and fly PXA-PXC domain.

Structures of indicated PXA-PXC domains in ribbon (left) surface coloured by hydrophobicity (middle), and sequence conservation (right). A region in SNX14 is highlighted that when deleted prevents LD recruitment [7]. The two bottom structures are examples of lipid transfer proteins E-Syt2 and GLTP that show how similar kinds of conserved hydrophobic pockets can serve as lipid binding cavities. The PXA-PXC structures therefore predict a role in lipid transfer.

Figure S6. AlphaFold2 predicted structure of Lec1/Ypr097w.

(**A**) Plot of the Predicted Alignment Error (PAE) from the AlphaFold2 database. There is a strong degree of correlation between the N-terminal PXYn and C-terminal PXYc domains 22uggesting these two domains are physically associated. (**B**) The predicted structure of Ypr097w/Lec1 from S. cerevisiae coloured according to the pLDDT score. (**C**) Overlay of the PXY proteins from S. cerevisiae, S. pombe and C. albicans.

Figure S7. AlphaFold2 predicted structure of PXY domain proteins found in *S. pombe*, *C. thermophilum* and *D. dictyostelium*.

(**A**) Structural prediction of *S. pombe* Lec1/Ypr092w orthologue C1450.12. (**B**) Structural prediction of *S. pombe* Lec1/Ypr092w homologue C663.15c. (**C**) Structural prediction of *C. thermophilum* Lec1/Ypr092w orthologue EGS19171. (**D**) Structural prediction of *C. thermophilum* Lec1/Ypr092w homologue EGS20008. (**E**) Structural prediction of *D. discoideum* DUF3818 domain-containing protein.

Table S1. Statistics for X-ray crystallographic data collection and structure refinement[a]

	Human SNX25 RGS domain	Human SNX25 RGS domain (C526A)
Wavelength (Å)	1.006	0.95364
Resolution range (Å)	45.84 -2.4 (2.486-2.4)	48.25-2.42 (2.52-2.42)
Space group	P 1 21 1	P 1 21 1
Unit cell	48.6 57.1 67.0 90 109.5 90	44.0 75.9 51.1 90.00 109.3 90.00
Total reflections	47931 (3290)	81322 (7264)
Unique reflections	13200 (1003)	11898 (1111)
Multiplicity	3.6 (3.1)	6.8 (6.5)
Completeness (%)	96.5 (74.1)	97.9 (89.2)
Mean I/sigma(I)	11.9 (2.1)	7.2 (1.4)
R-merge	0.065 (0.506)	0.171 (0.976)
R-meas	0.089 (0.690)	0.185 (1.062)
R-pim	0.061 (0.467)	0.070 (0.410)
CC1/2	0.998 (0.718)	0.992 (0.631)
R-work	0.2164 (0.2984)	0.2108 (0.3258)
R-free	0.2421 (0.2906)	0.2497 (0.4051)
Number of non-hydrogen atoms	2157	2026
macromolecules	2080	1991
solvent	77	30
RMS(bonds)	0.003	0.007
RMS(angles)	0.55	0.87
Ramachandran favored (%)	97.08	97.38
Ramachandran allowed (%)	2.92	2.62
Ramachandran outliers (%)	0.00	0.00
Clashscore	2.66	6.07
Average B-factor	51.51	52.68
macromolecules	51.69	52.76
solvent	46.74	46.51
PDB ID	7SR1	7SR2

a. Statistics for the highest-resolution shell are shown in parentheses.

Table S2. Homologues of the PXA-PXC and PXYn-PXYc domain-containing proteins based on structural similarity in the AlphaFold2 database[a].

	PXA-PXC domains	Uniprot ID	PXYn-PXYc domains	Uniprot ID
H. sapiens **(human)**	SNX13 SNX14 SNX19 SNX25 (SNX25 full length)	Q9Y5W8 Q9Y5W7 Q92543 Q9H3E2 A0A494C0S0	-	-
D. melanogaster **(fly)**	Snz	Q9W3N0	-	-
C. elegans **(worm)**	snx-14 snx-13	G5EF63 Q9U2U6	-	-
A. thaliana **(plant)**	At1g15240 At2g15900	F4HZJ6 F4IJE1	-	-
D. discoideum **(amoeba)**	-	-	DDB0187809	Q54JB8
S. cerevisiae **(budding yeast)**	Mdm1/Yml104c Nvj3/Ydr179W-a	Q01846 Q03983	Lec1/Ypr097w	Q06839
S. pombe **(fission yeast)**	snx12 mug122 pxa1	Q9USN1 (all domains) O74444 (no RGS) O14200 (like Nvj3)	SPCC663.15c SPCC1450.12	O74521 (no PX) Q9Y7N9 (all domains)
C. thermophilum[b] **(thermophilic fungi)**	CTHT_0043790/EGS19887 CTHT_0051170/EGS18515 CTHT_0024100/EGS20576	G0S8X6 (all domains) G0SDB4 (no RGS) G0S5A3 (like Nvj3)	CTHT_0045050/EGS20008 CTHT_0057960/EGS19171	G0S997 (no PX) G0SCP5 (all domains)
M. janaschii **(archaea)**	-	-	-	-
P. falciparum **(protozoan)**	-	-	-	-

a. The DALI webserver (http://ekhidna2.biocenter.helsinki.fi/dali/) was used to search the AlphaFold2 database against the indicates species using either the SNX25 PXA-PXC structure or the Lec1/Ypr097w PXYn-PXYc structure.
b. *C. thermophilum* proteins were identified via BLAST sequence searches and their domain structures were validated using ColabFold structural predictions.

Table S3. Key Resources Table.

REAGENT or RESOURCE	SOURCE or REFERENCE	IDENTIFIER
Bacterial Strains		
E. coli DH5α	Invitrogen	18265017
BL21-CodonPlus (DE3)-RIL	Agilent Technologies	230245
Mammalian cell Lines		
A431 (human squamous carcinoma cell line)	Merck (ATCC)	85090402
Antibodies, mammalian cell culture and imaging		
FLAG	Sigma Aldrich	F7425
Nile red	Sigma Aldrich	72485
Donkey anti-Rabbit secondary antibody Alexa Fluor®488 conjugate	Life Technologies Australia	A21206
Culture dishes	Nunc	140675
DMEM, high glucose, pyruvate, no glutamine	Thermo Fisher Scientific Aus Pty Ltd	10313039
Foetal bovine serum	Cytiva (Global Life Sciences Solutions)	SH30084.03
Trypsin-EDTA 0.05%	Thermo Fisher Scientific Aus Pty Ltd	25300054
lipofectamine 3000	Thermo Fisher Scientific Aus Pty Ltd	L3000015
Penicillin/Streptomycin	Thermo Fisher Scientific Aus Pty Ltd	15070063
Chemicals and enzymes		
Thrombin	Sigma Aldrich	T6634-1KU
Recombinant DNA		
pGEX4T-2 human SNX25 RGS domain (residues 446-569; A0A494C0S0)	This study	N/A
pGEX4T-2 human SNX25 RGS domain (residues 446-569; A0A494C0S0) C526A mutation	This study	N/A
pcDNA3.1+-C DYK (FLAG) human SNX25	Genscript	NM_031953
Deposited Data		
human SNX13 AlphaFold2 model	[24, 25]	Q9Y5W8
human SNX14 AlphaFold2 model	[24, 25]	Q9Y5W7
human SNX19 AlphaFold2 model	[24, 25]	Q92543
human SNX25 AlphaFold2 model	[24, 25]	Q9H3E2
human SNX25 sequence including an N-terminal hydrophobic ER anchor sequence.	Uniprot	
fly Snz AlphaFold2 model	[24, 25]	Q9W3N0
yeast Mdm1 AlphaFold2 model	[24, 25]	Q01846
yeast Nvj3 AlphaFold2 model	[24, 25]	Q03983
yeast Lec1/Ypr097w AlphaFold2 model	[24, 25]	Q06839
S. pombe SNX14 AlphaFold2 model	[24, 25]	Q9Y7N9

C. albicans SNX14 AlphaFold2 model	[24, 25]	A0A1D8PQ52
human SNX14 PX domain crystal structure	[5]	4PQP
human SNX14 PX domain crystal structure	[5]	4PQO
mouse SNX19 PX domain crystal structure	[5]	4P2I
human SNX25 PX domain NMR structure	[1]	4PQP
human GLTP crystal structure	[54]	3S0K
human E-Syt2 crystal structure	[55]	4P42
human SNX25 RGS domain (residues 283-405; Q9H3E2) crystal structure	This study	7SR1
human SNX25 RGS domain (residues 283-405; Q9H3E2) (C326A) crystal structure	This study	7SR2
human RGS1 RGS domain crystal structure	[27]	2BV1
Software		
Coot	[49]	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
Pymol	Schrodinger, USA.	https://pymol.org/2/
AlphaFold2	[24]	N/A
ColabFold	[29]	https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/beta/AlphaFold2_advanced.ipynb
Consurf	[53]	https://consurf.tau.ac.il
Other		
HiLoad™ Superdex75 PG	GE Healthcare	Catalogue: 28989333

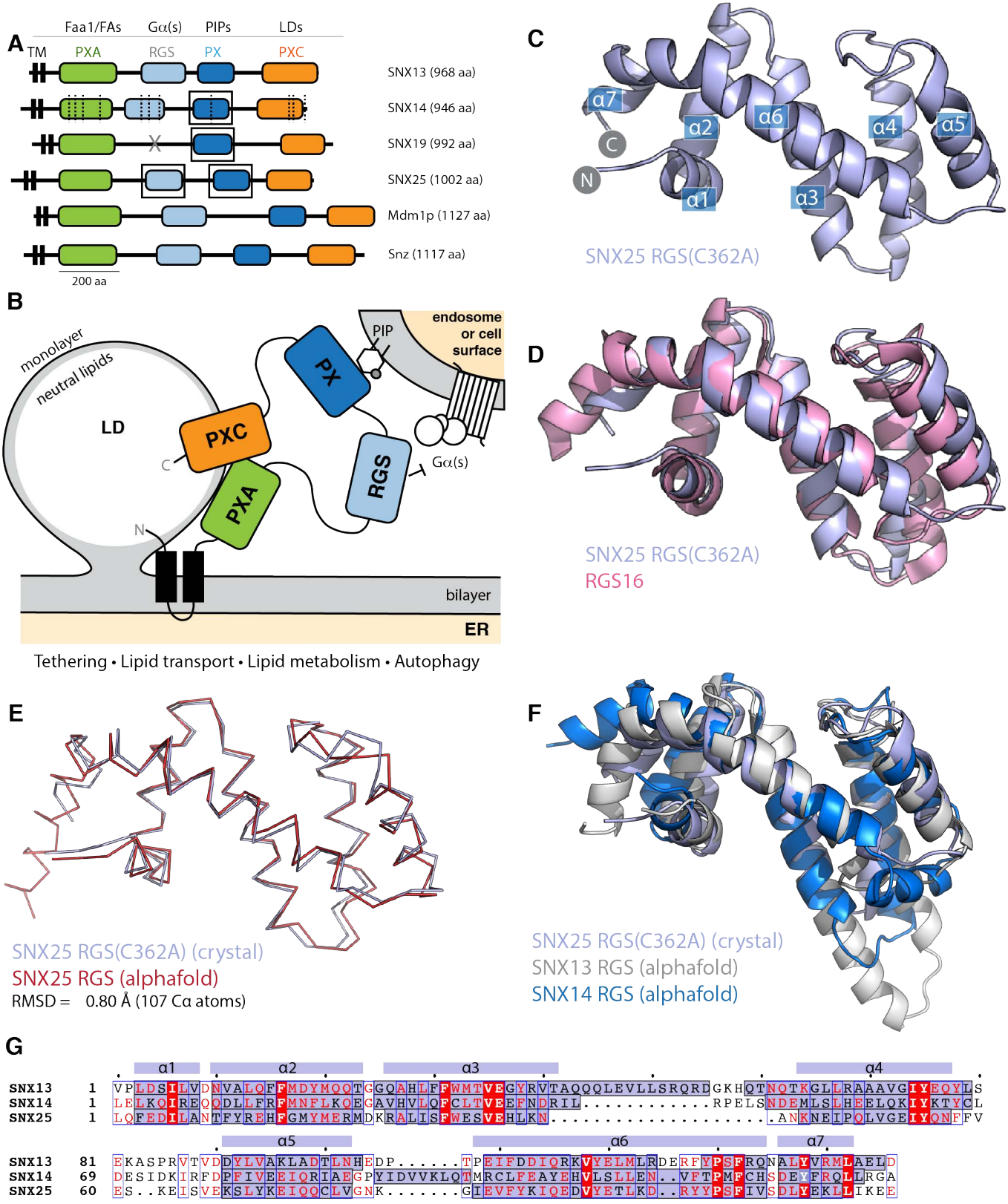

Figure 1. The SNX-RGS protein family.
(A) Domains of human, Drosophila, and S. cerevisiae SNX-RGS family members. Putative ligands for each domain are indicated above. SNX14 mutations causing cerebellar ataxia are shown with dashed lines 14, 15. Boxes indicate domains with experimental structures determined. (B) The current working model for the architecture, localisation and function of the SNX-RGS protein family and their potential roles at the interface between the ER, LDs, and membranes of the endolysosomal system. (C) Crystal structure of human SNX25(C362A) RGS domain in cartoon diagram. Helical secondary structure elements are indicated. (D) The crystal structure of human SNX25(C362A) RGS domain (light blue) is aligned with the canonical RGS domain of RGS16 (pink)(PDB ID 2IK8)27 and shows the expected topology. (E) Cα ribbon diagram showing the overlay of the SNX25(C362A) RGS domain (light blue) with the same sequence predicted in the AlphaFold2 database (red) (Q9H3E2). (F) Cartoon representation of SNX25(C362A) RGS domain compared with the SNX13 and SNX14 RGS domains from the AlphaFold2 database (Q9Y5W8; Q9Y5W7). (G) Sequence alignment of human SNX13, SNX14 and SNX25 RGS domains based on structural comparisons. The secondary structure of SNX25 derived from its crystal structure is shown schematically above the alignment, while the α-helical regions of the three proteins from their AlphaFold2predictions are shaded blue within the alignment.

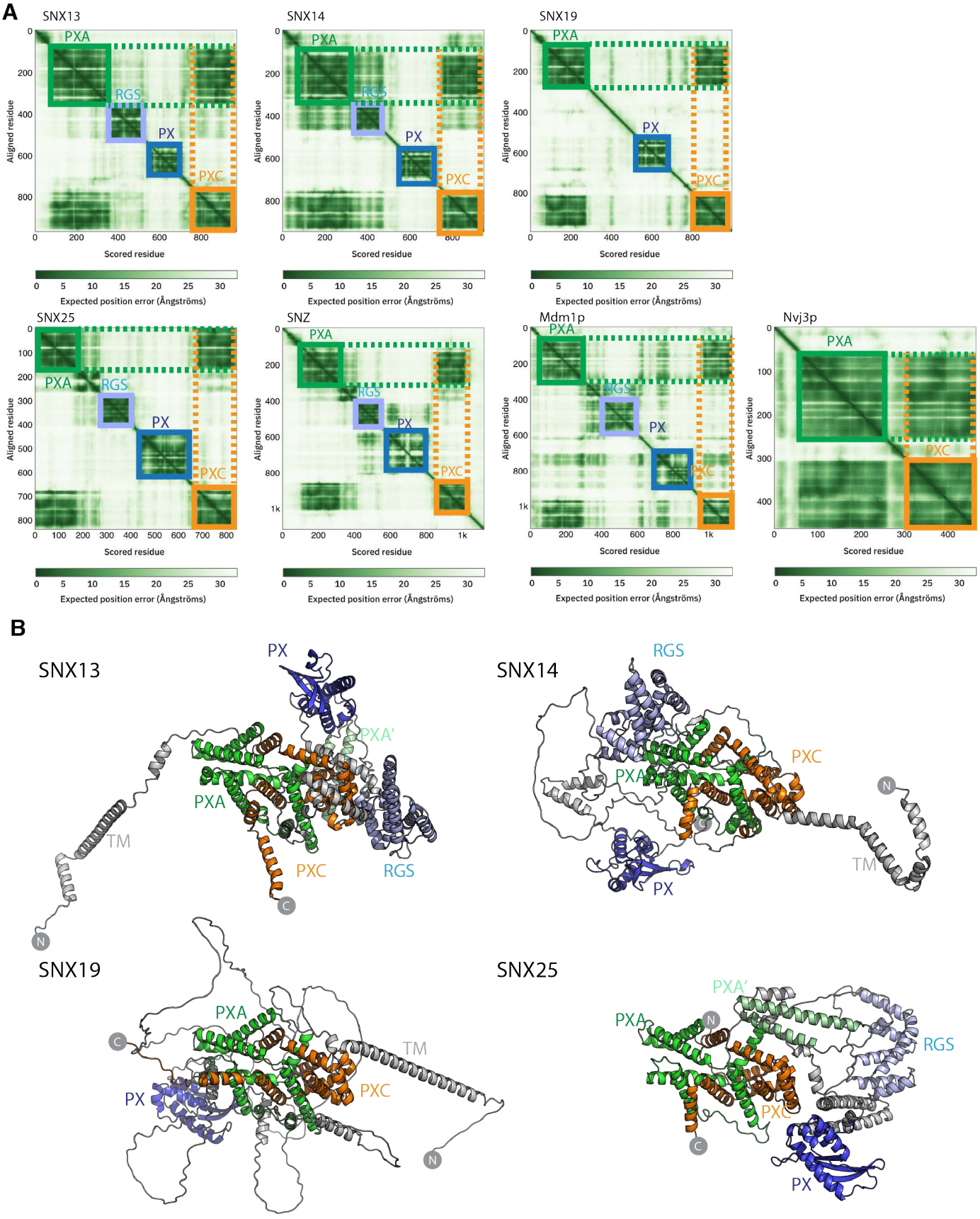

Figure 2. Structures of the SNX-RGS proteins predicted by AlphaFold2.
(A) Predicted Alignment Error (PAE) plots from the AlphaFold2database 25 are shown for human fly and yeast SNX-RGS proteins. In these plots all SNX-RGS proteins show a strong degree of correlation between the PXA and PXC domain suggesting these two domains are physically associated. (B) The predicted structures of human SNX-RGS proteins from the AlphaFold2database. The PXA domain is coloured green, RGS domain in light blue, PX domain in blue and PXC domain in orange. The predicted TM domain and any unstructured linker regions are coloured grey. The structures are shown in the same orientation after alignments based on the PXA and PXC core region. This shows that the two domains are intimately entwined with each other, whereas the TM, RGS and PX domains are predicted to have flexible orientations relative to these domains.

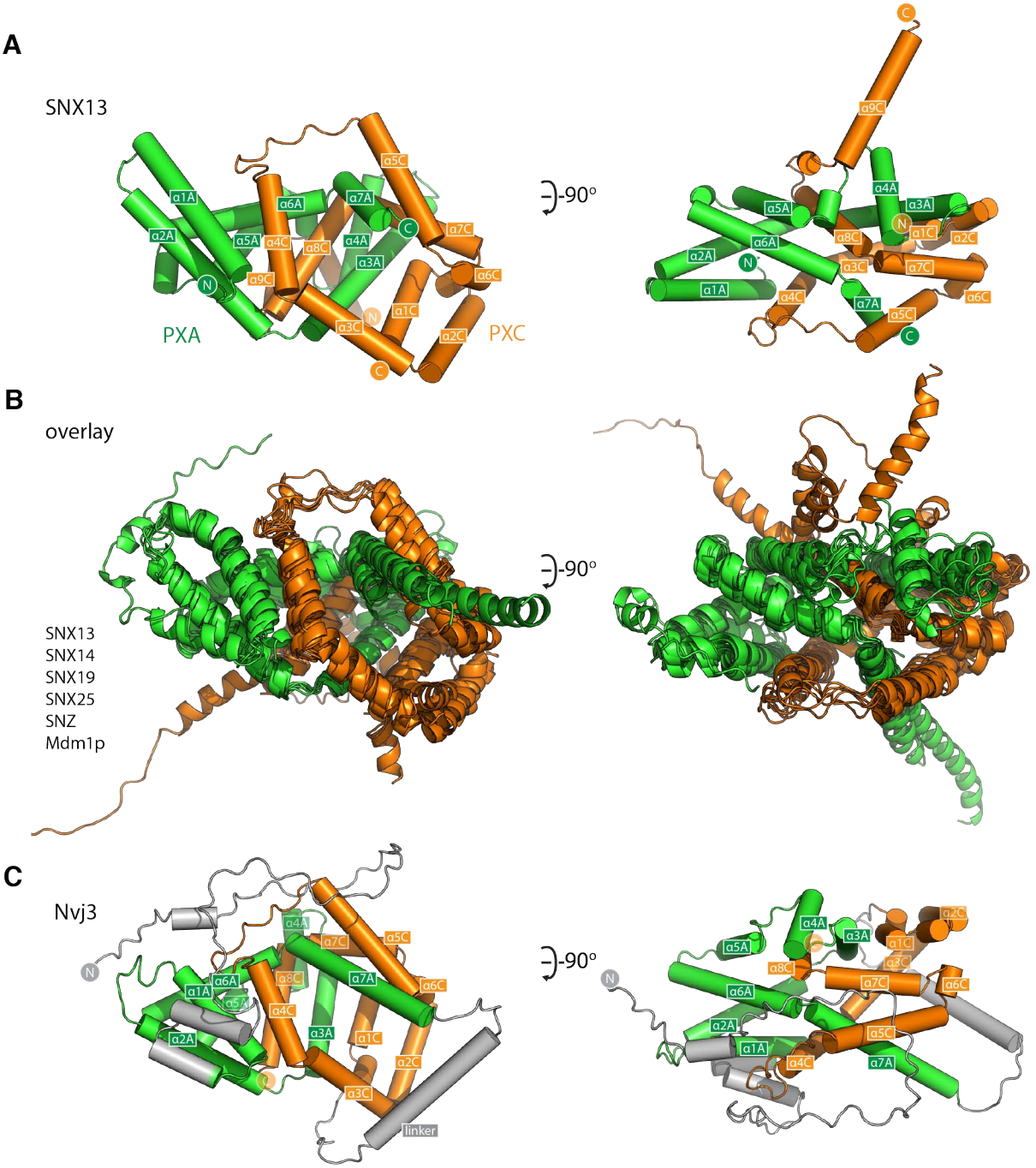

Figure 3. The PXA and PXC domains combine form an intertwined α-helical structure.
(A) Structure of the PXA and PXC domains of human SNX13 predicted by AlphaFold2 in green and orange respectively, shown with cylinders for α-helices. The two domains are predicted to be tightly interwoven. (B) An overlay of the core PXA-PXC domains of human, yeast and fly SNX-RGS proteins shows that all predicted structures have the same topology. (C) Predicted structure of S. cerevisiae Nvj3 with the regions expected to be similar to PXA and PXC domains coloured green and orange and the linker shown in grey.

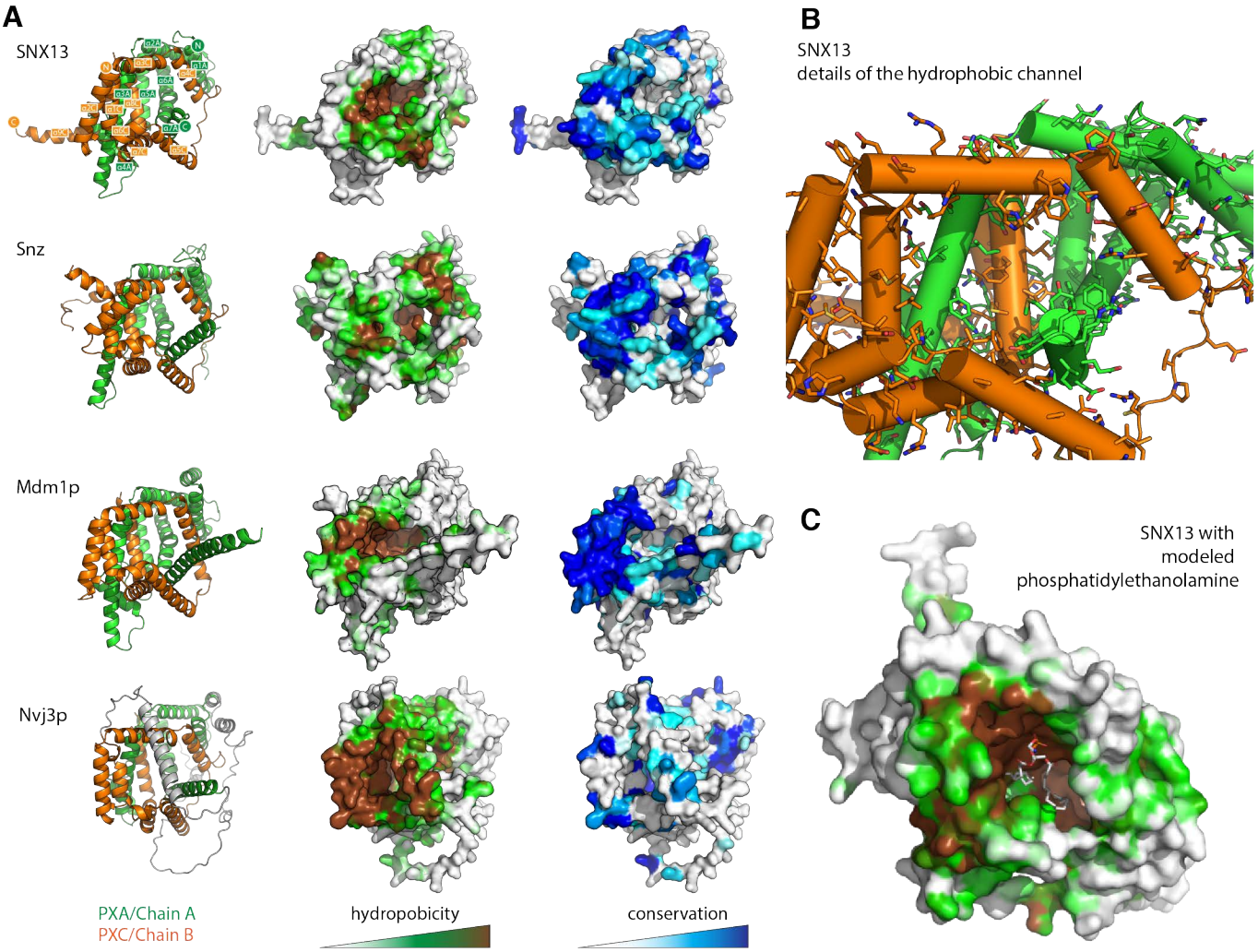

Figure 4. The PXA-PXC structure forms a conserved hydrophobic cavity with potential for lipid binding.
(A) The predicted structures of PXA-PXC domains from indicated species are shown in ribbon representation (left), surface coloured by hydrophobicity (middle), and sequence conservation (right). All PXA-PXC structures have a highly conserved hydrophobic tunnel. (B) Structure of PXA and PXC domains of SNX13 with sidechains of the putative lipid binding pocket shown. (C) The PXA and PXC domains of SNX13 is shown in close-up with its surface coloured for hydrophobicity as in (A). A phosphatidylethanolamine lipid has been docked manually in the putative lipid binding pocket to give perspective on the dimensions of the channel.

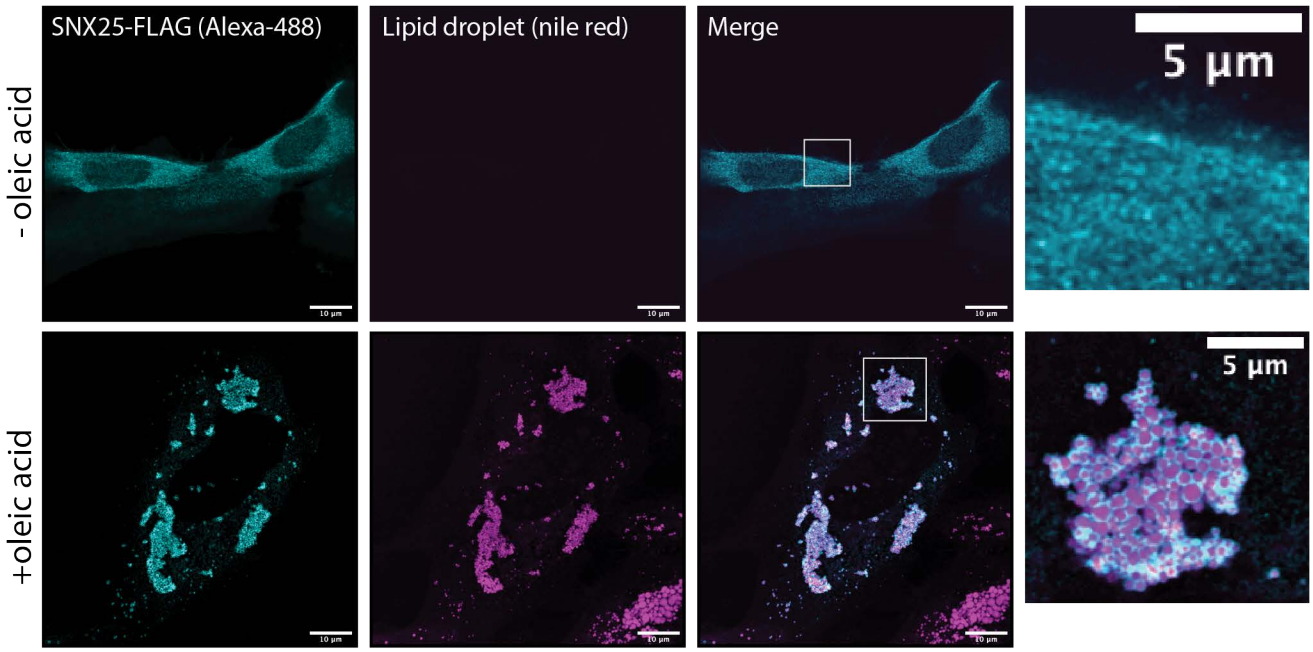

Figure 5. Human SNX25 is localises to newly synthesised LDs following oleic acid addition
A431 cells were transfected with SNX25-FLAG and either left untreated or treated with 50 µg/ml oleic acid to stimulate LD formation. After oleic acid addition SNX25-FLAG undergoes rapid redistribution to the periphery of newly generated LDs LD periphery. Scale bars represent 10 µm.

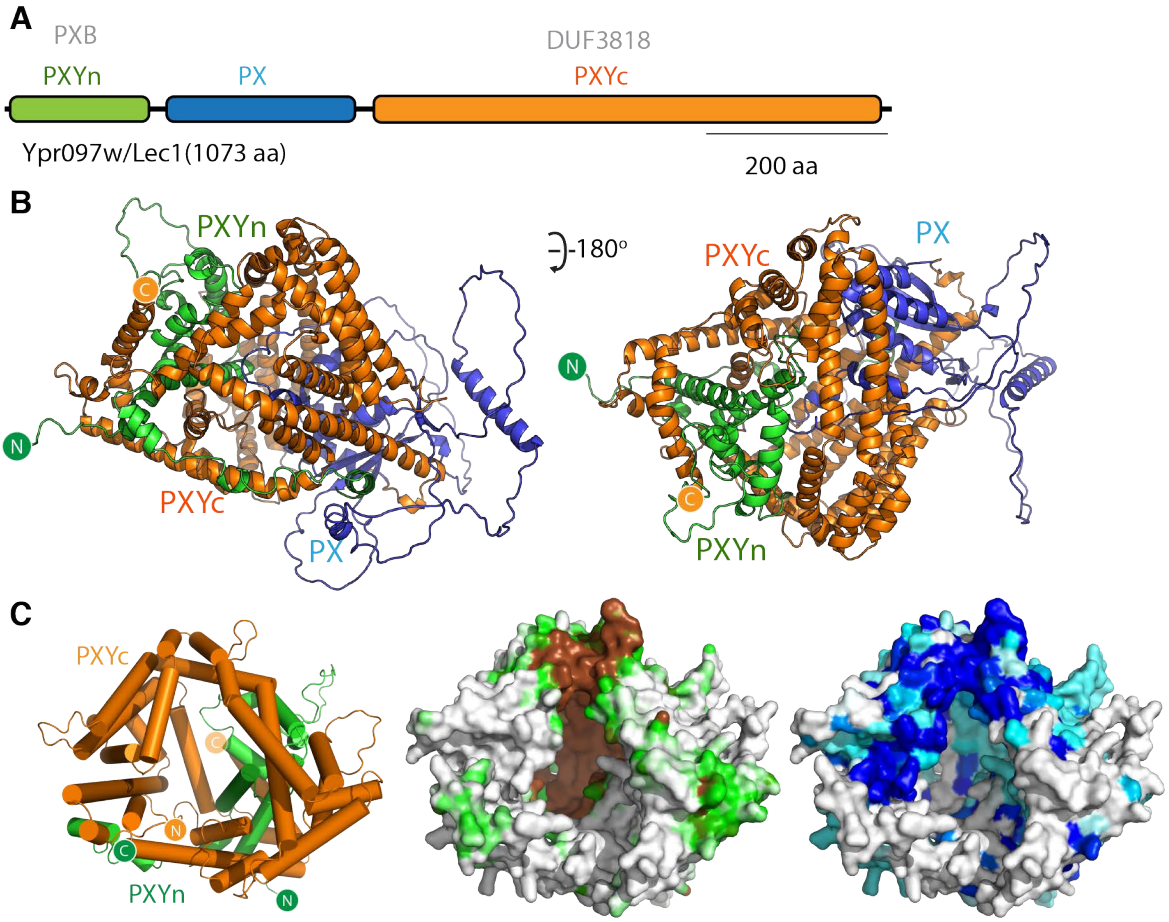

Figure 6. The yeast Lec1/Ypr097w protein forms a unique structure with a hydrophobic cavity.
(A) Domain structure of the S. cerevisiae Lec1/Ypr097w protein. (B) AlphaFold2 structural prediction of the Lec1/Ypr097w protein with the PXYn domain in green, PX domain in blue and PXYc domain in orange. (C) The PXYn and PXYc domains are shown with α-helices in cylinder representation and the PX domain removed for clarity. The domains encompass a large conserved hydrophobic cavity and like the SNX-RGS proteins may be a potential lipid binding protein.

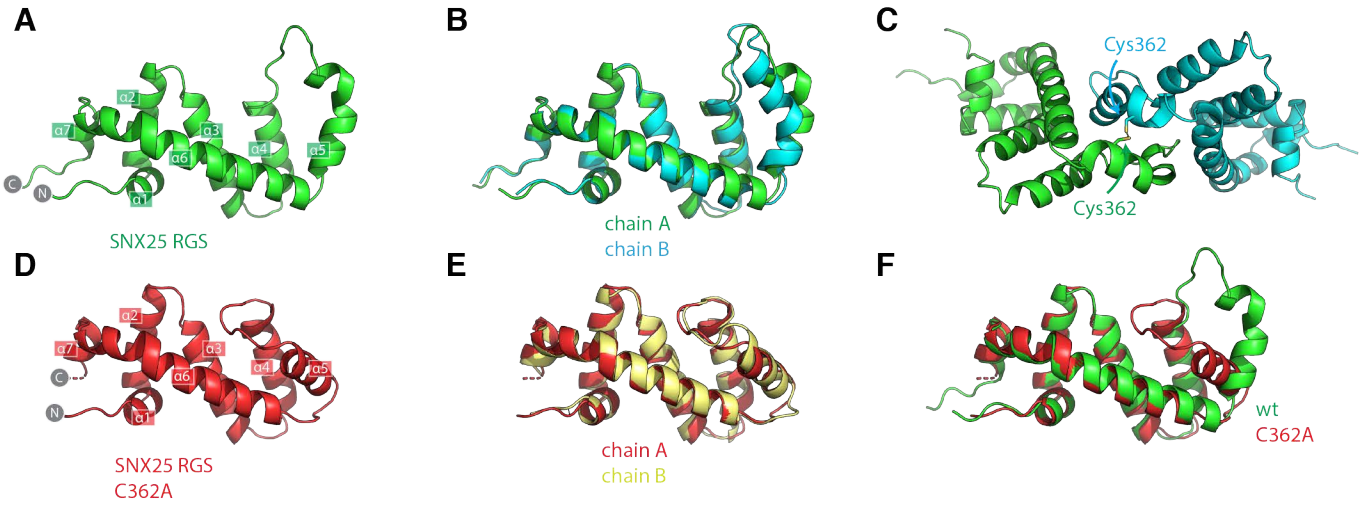

Figure S1. Crystal structure of the SNX25 RGS domain.
(A) Crystal structure of the wild-type human SNX25 RGS domain. (B) Alignment of the two chains in the asymmetric unit of wild-type SNX25 RGS domain. (C) The position of the non-native disulfide bond formed between adjacent chains in the crystal lattice. This causes the change in orientation of the α5 helix. (D) Crystal structure of the SNX25(C526A) mutant protein. (E) Alignment of the two chains in the asymmetric unit of SNX25(C526A) RGS domain. (F) Structural alignment of the wild-type and C526A SNX25 RGS domain structures.

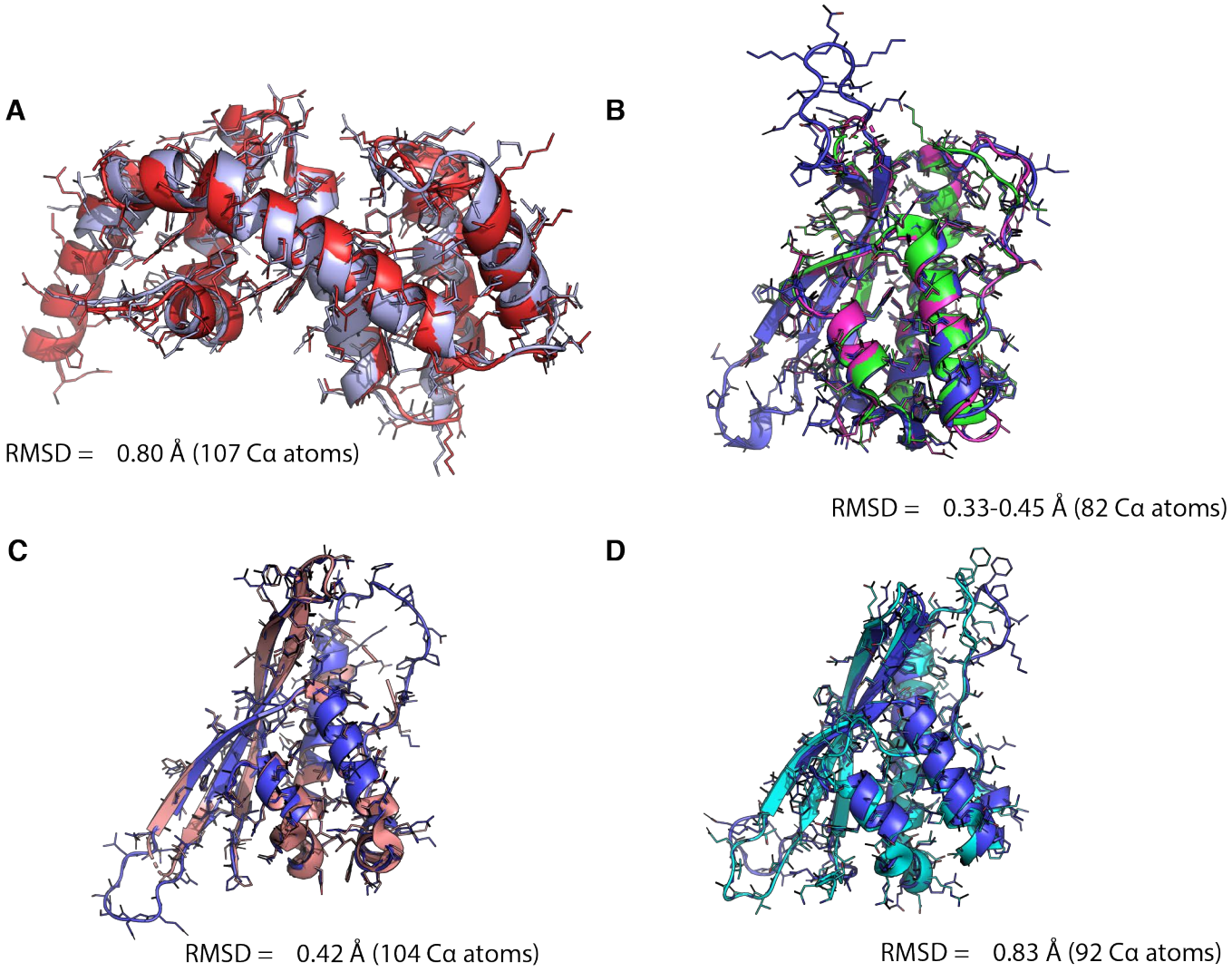

Figure S2. Comparison of experimental SNX-RGS structures with their AlphaFold2 predictions.
(A) Crystal structure of the human SNX25(C326A) RGS domain (red, this study) aligned with the AlphaFold2 prediction (light blue). (B) Crystal structures of human SNX14 PX domain (green, PDB 4PQP; magenta, 4PQO) with the AlphaFold2 prediction (blue). (C) Crystal structure of mouse SNX19 PX domain (pink, PDB 4P2I) aligned with the AlphaFold2 prediction (blue). (D) NMR structure of human SNX25 PX domain (cyan, PDB 4PQP) aligned with the AlphaFold2 prediction (blue).

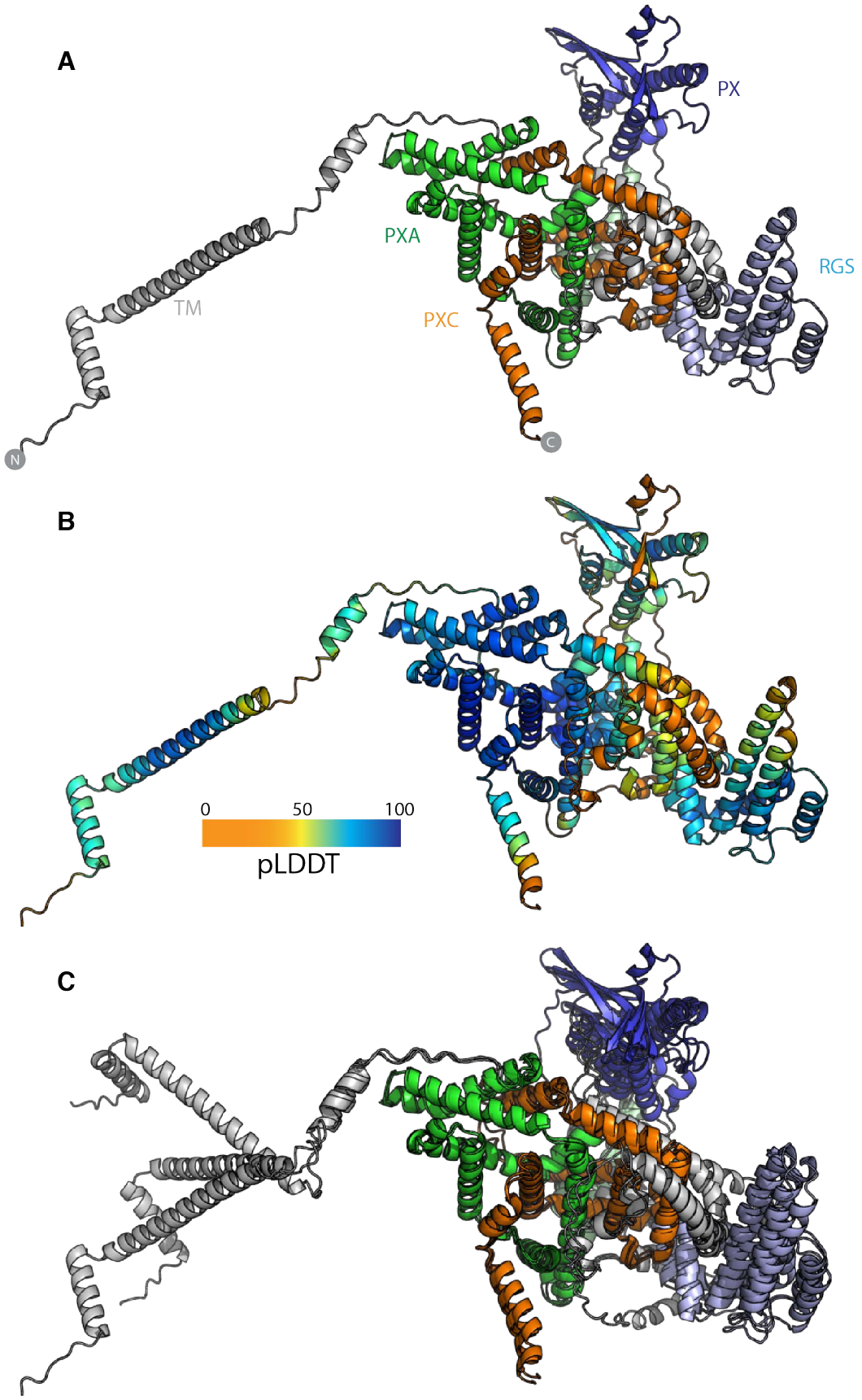

Figure S3. Additional details of the SNX13 predicted full length structure.
(A) AlphaFold2 prediction of full-length human SNX13 protein with individual domains coloured as in Fig. 2B with the TM domain and linker regions (grey), PXA domain (green), RGS domain (light blue), PX domain (blue), and PXC domain (orange). (B) AlphaFold2 prediction of full-length human SNX13 protein coloured according to the pLDDT score. (C) The AlphaFold2 predictions of human, mouse and zebrafish SNX13 proteins were aligned based on the core PXA-PXC structure.

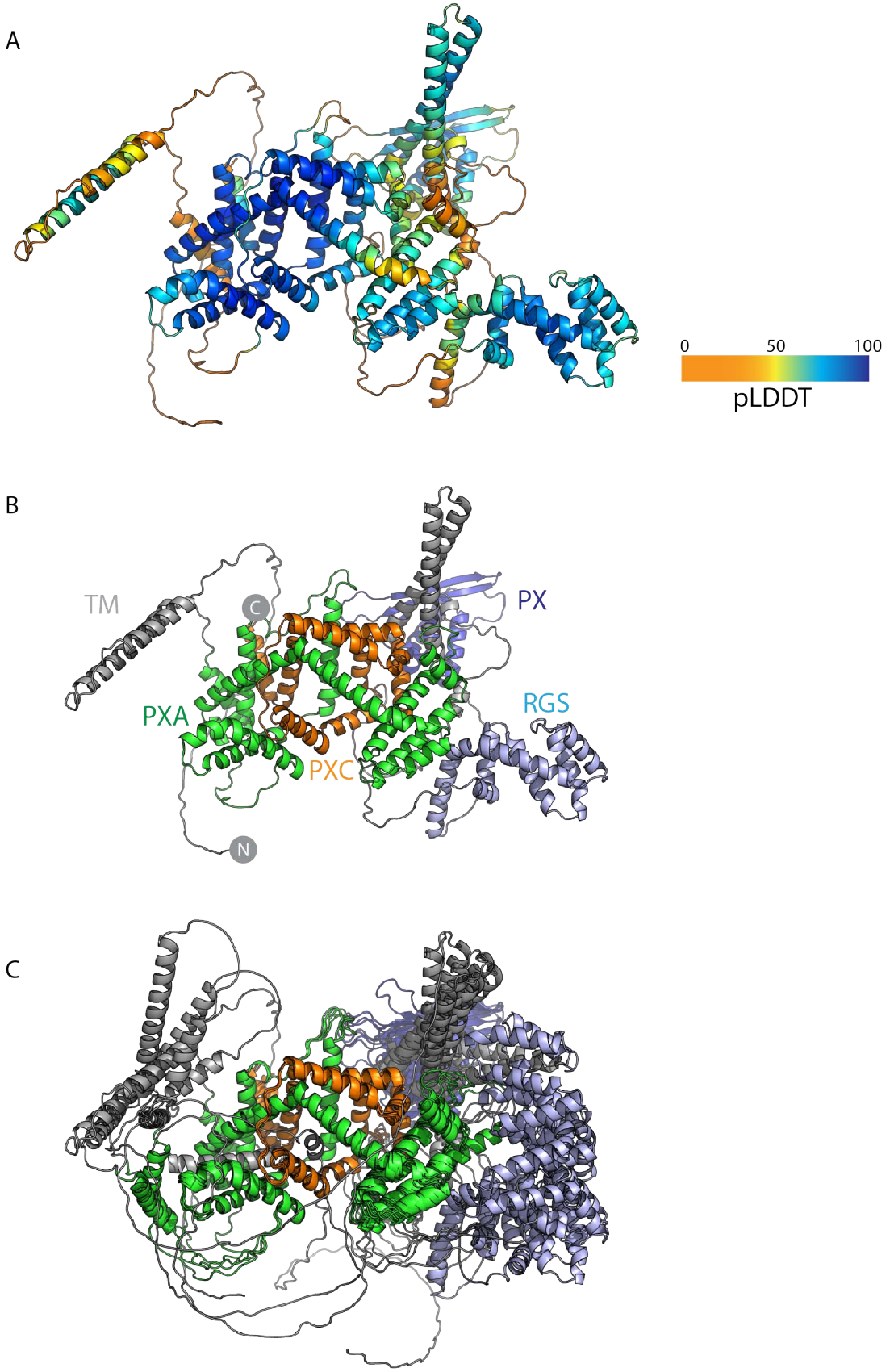

Figure S4. AlphaFold2 prediction of full-length human SNX25 protein.
(A) Top ranked structure of full length human SNX25 5 (Uniprot ID A0A494C0S0) predicted by ColabFold coloured according to pLDDT score. (B) Top ranked structure of human SNX25 predicted by ColabFold coloured according to the indicated domains. (C) Alignment based on the core PXA-PXC structure of the five structural predictions of human SNX25 from ColabFold.

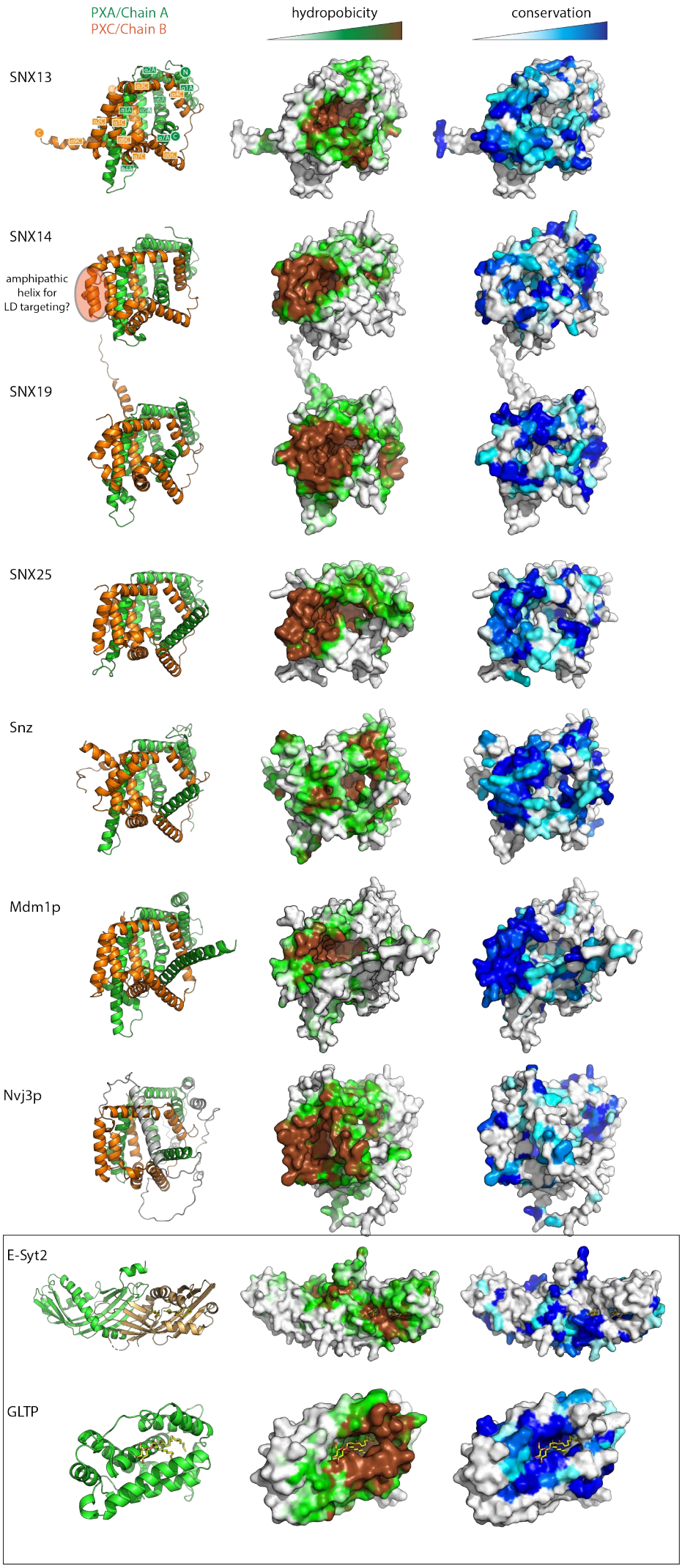

Figure S5. Conserved hydrophobic cavities in all human, yeast and fly PXA-PXC domain.
Structures of indicated PXA-PXC domains in ribbon (left) surface coloured by hydrophobicity (middle), and sequence conservation (right). A region in SNX14 is highlighted that when deleted prevents LD recruitment 7. The two bottom structures are examples of lipid transfer proteins E-Syt2 and GLTP that show how similar kinds of conserved hydrophobic pockets can serve as lipid binding cavities. The PXA-PXC structures therefore predict a role in lipid transfer.

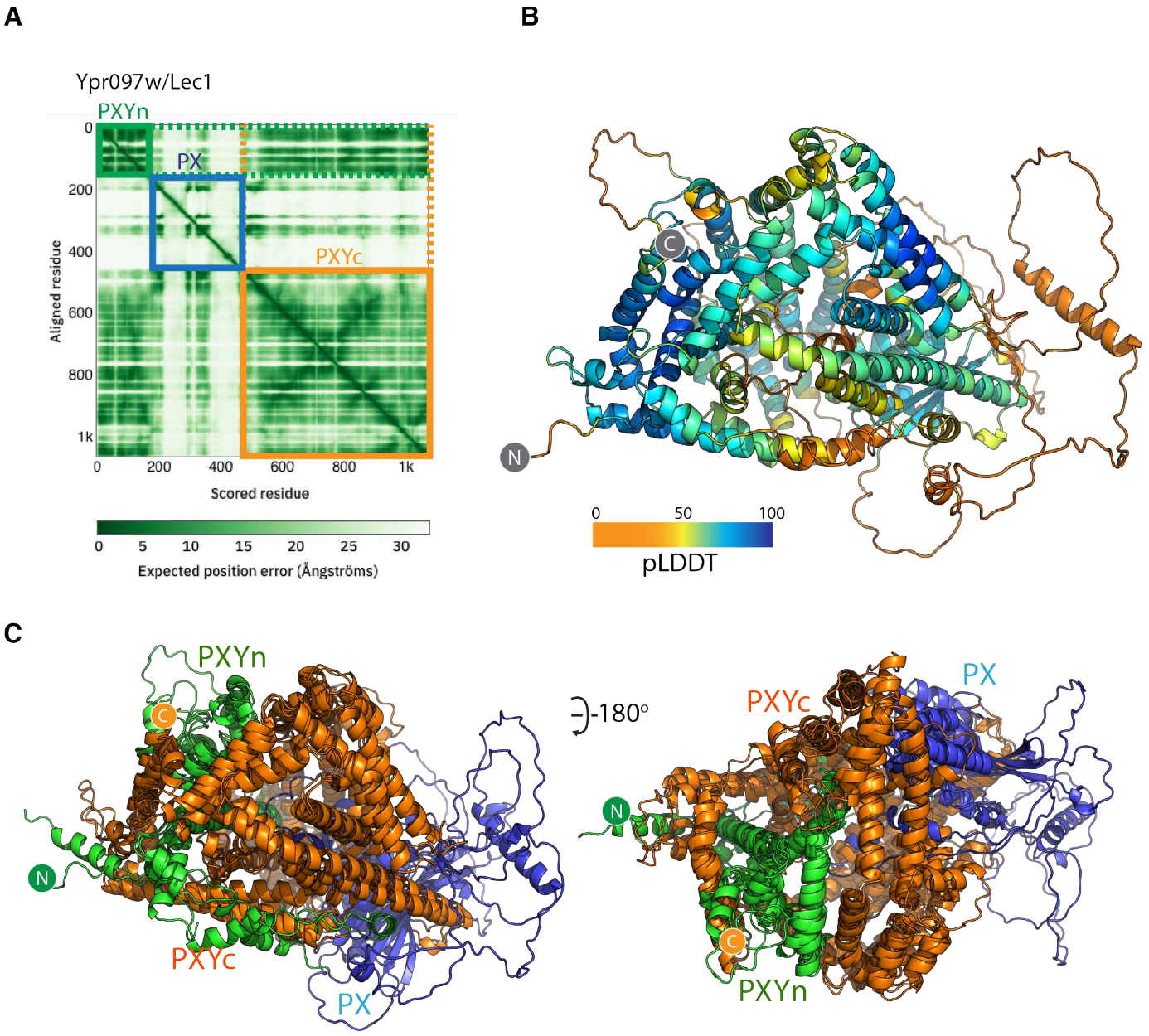

Figure S6. AlphaFold2 predicted structure of Lec1/Ypr097w.
(A) Plot of the Predicted Alignment Error (PAE) from the AlphaFold2 database. There is a strong degree of correlation between the N-terminal PXYn and C-terminal PXYc domains uggesting these two domains are physically associated. (B) The predicted structure of Ypr097w/Lec1 from S. cerevisiae coloured according to the pLDDT score. (C) Overlay of the PXY proteins from S. cerevisiae, S. pombe and C. albicans.

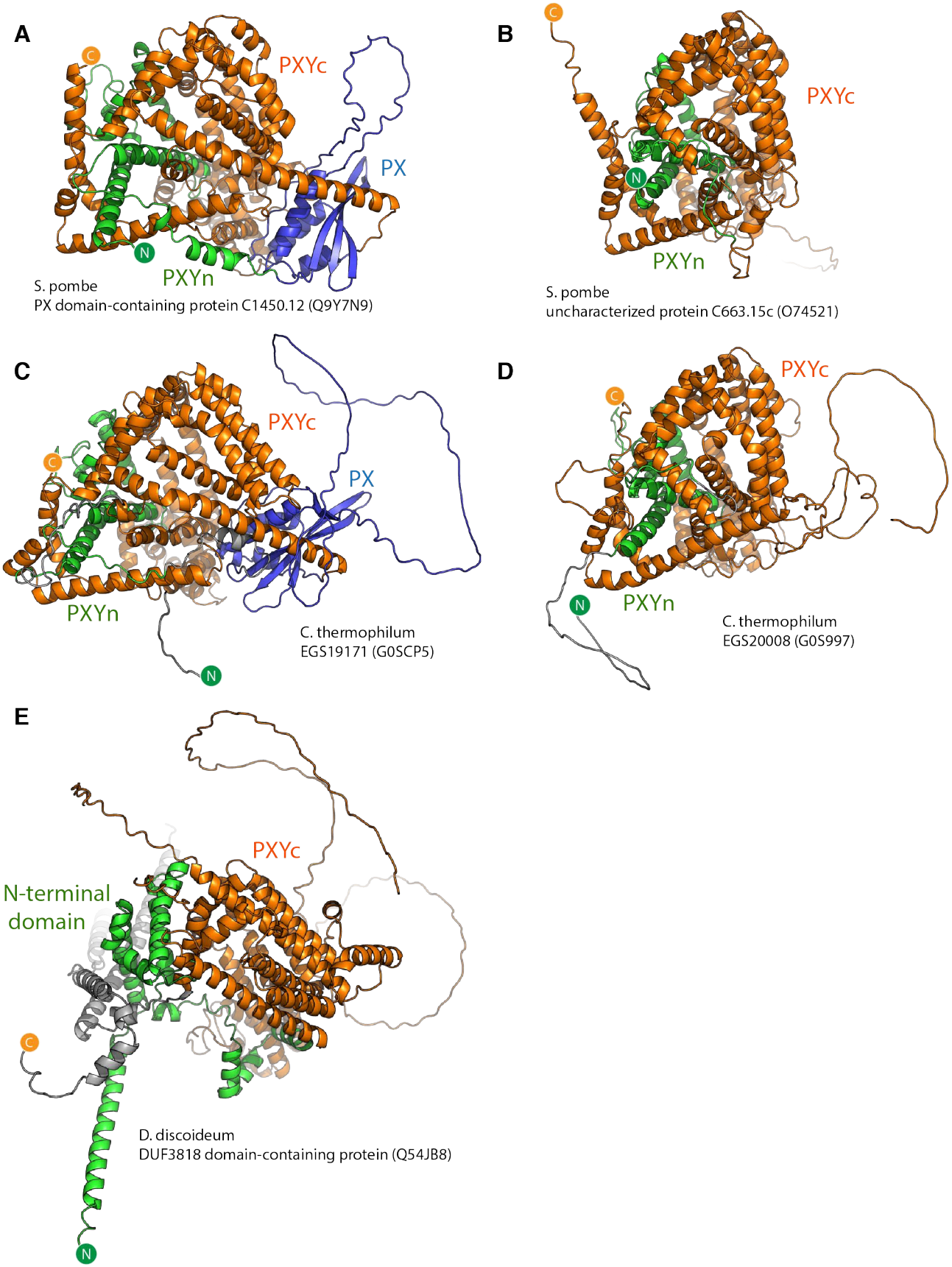

Figure S7. AlphaFold2 predicted structure of PXY domain proteins found in S. pombe, C. thermophilum and D. dictyostelium. (A) Structural prediction of S. pombe Lec1/Ypr092w orthologue C1450.12. (B) Structural prediction of S. pombe Lec1/Ypr092w homologue C663.15c. (C) Structural prediction of C. thermophilum Lec1/Ypr092w orthologue EGS19171. (D) Structural prediction of C. thermophilum Lec1/Ypr092w homologue EGS20008. (E) Structural prediction of D. discoideum DUF3818 domain-containing protein.