

RESEARCH ARTICLE

Evolution and insights into the structure and function of the DedA superfamily containing TMEM41B and VMP1

Fumiya Okawa^{1,‡}, Yutaro Hama^{1,‡}, Sidi Zhang^{1,‡}, Hideaki Morishita^{1,*}, Hayashi Yamamoto¹, Tim P. Levine² and Noboru Mizushima^{1,§}

ABSTRACT

TMEM41B and VMP1 are endoplasmic reticulum (ER)-localizing multi-spanning membrane proteins required for ER-related cellular processes such as autophagosome formation, lipid droplet homeostasis and lipoprotein secretion in eukaryotes. Both proteins have a VTT domain, which is similar to the DedA domain found in bacterial DedA family proteins. However, the molecular function and structure of the DedA and VTT domains (collectively referred to as DedA domains) and the evolutionary relationships among the DedA domain-containing proteins are largely unknown. Here, we conduct a remote homology search and identify a new clade consisting mainly of bacterial proteins of unknown function that are members of the Pfam family PF06695. Phylogenetic analysis reveals that the TMEM41, VMP1, DedA and PF06695 families form a superfamily with a common origin, which we term the DedA superfamily. Coevolution-based structural prediction suggests that the DedA domain contains two reentrant loops facing each other in the membrane. This topology is biochemically verified by the substituted cysteine accessibility method. The predicted structure is topologically similar to that of the substrate-binding region of Na⁺-coupled glutamate transporter solute carrier 1 (SLC1) proteins. A potential ion-coupled transport function of the DedA superfamily proteins is discussed.

This article has an associated First Person interview with the joint first authors of the paper.

KEY WORDS: DedA, VMP1, TMEM41B, Reentrant loop, Autophagy

INTRODUCTION

TMEM41B and VMP1 are endoplasmic reticulum (ER)-localizing multi-spanning membrane proteins essential for autophagosome formation, lipid droplet homeostasis, membrane contact, lipoprotein secretion and replication of RNA viruses including SARS-CoV-2 (Demignot et al., 2014; Hoffmann et al., 2021; Moretti et al., 2018; Morishita et al., 2019; Morita et al., 2018; Ropolo et al., 2007; Schneider et al., 2021; Shoemaker et al., 2019; Tabara and Escalante, 2016; Van Alstyne et al., 2018; Zhao et al., 2017). Because the formation of lipid droplets (Walther et al., 2017),

lipoproteins (Demignot et al., 2014), viral replication complexes (Romero-Brey and Bartenschlager, 2016) and autophagosomes involves the ER (Nakatogawa, 2020), TMEM41B and VMP1 are considered to play fundamental roles in the ER. Elucidation of the molecular functions of these proteins would provide important insights into our understanding of the role of the ER in autophagy and other pathways, but their functions, structure and even membrane topology are largely unknown.

VMP1 and TMEM41B contain a conserved transmembrane domain that is also found in TMEM64 and its homolog Tvp38 in metazoans, yeasts (Inadome et al., 2007), amoebozoans (Tabara and Escalante, 2016), chloroplasts and cyanobacteria (Keller and Schneider, 2013). We previously termed this domain the VTT (VMP1, TMEM41 and Tvp38/TMEM64) domain (also known as SNARE_assoc domain; Pfam PF09335) (Morita et al., 2018, 2019). The VTT domain is similar to the bacterial downstream (of *hisT*) *Escherichia coli* DNA gene A (DedA) domain (Doerrler et al., 2013; Inadome et al., 2007; Khafizov et al., 2010; Nonet et al., 1987; Thompkins et al., 2008). The DedA domain is present in a set of bacterial proteins that constitute the DedA family (Thompkins et al., 2008). YqjA and YghB are the best-characterized members of this family and are known to regulate temperature sensitivity, cell division (Thompkins et al., 2008), the export of periplasmic amidases (Sikdar and Doerrler, 2010), drug resistance (Kumar and Doerrler, 2014; Panta et al., 2019), pH sensitivity (Kumar and Doerrler, 2015), lipid A modification (Panta and Doerrler, 2021) and lipid composition of the cell membrane (Boughner and Doerrler, 2012; Thompkins et al., 2008). However, although putative transporter functions have been hypothesized based on genetic studies (Doerrler et al., 2013; Kumar and Doerrler, 2014), the molecular functions of these bacterial DedA family proteins are unknown. The VTT and DedA domains of most proteins in this family contain the conserved sequence motifs [F/Y]XXX[R/K] and GXXX[V/I/L/M]XXXX[F/Y] (Doerrler et al., 2013; Keller and Schneider, 2013; Tabara et al., 2019). Although the VTT and DedA domains are evolutionarily related, previous phylogenetic analyses of VTT and DedA domain-containing proteins have been conducted with relatively small numbers of these proteins, excluding potential remote homologs (Boughner and Doerrler, 2012; Doerrler et al., 2013; Keller and Schneider, 2013; Thompkins et al., 2008). Thus, the exact definition of the VTT and DedA domains and their evolutionary relationships remain unclear.


Little is known about the structure of the VTT and DedA domains. The VTT domain is predicted to form a complicated structure containing several transmembrane helices (TMHs), some of which may be discontinuous (Morita et al., 2018, 2019). The DedA domain has been proposed to adopt a structure similar to one half of the LeuT fold (Keller et al., 2014; Khafizov et al., 2010), leading to the hypothesis that the DedA domain serves as a half transporter module. Consistently, the self-interaction of YqjA has

¹Department of Biochemistry and Molecular Biology, Graduate School of Medicine, The University of Tokyo, Tokyo 113-0033, Japan. ²UCL Institute of Ophthalmology, University College London, London EC1V 9EL, UK.

*Present address: Department of Physiology, Juntendo University Graduate School of Medicine, Tokyo 113-8421, Japan.

[‡]These authors contributed equally to this work

[§]Author for correspondence (nmizu@m.u-tokyo.ac.jp)

 T.P.L., 0000-0002-7231-0775; N.M., 0000-0002-6258-6444

been reported (Keller et al., 2015). However, there is no experimental evidence supporting this structural prediction.

Here, we provide novel insights into the evolution and molecular functions of the VTT and DedA domains from both an expanded phylogenetic analysis that includes the remote homologs and a coevolution-based structural prediction. We found that the VTT and DedA domain-containing proteins, including newly identified remote homologs of the Pfam PF06695 family, constitute a large superfamily with a common origin, which we term the DedA superfamily. The new phylogenetic tree suggests that the prokaryotic species already had several ancestral proteins of the superfamily, some of which evolved into the present eukaryotic homologs. Structure predictions and accompanying biochemical verifications define the membrane topology of the VTT/DedA domain, which contains two canonical TMHs and two reentrant loops that face each other in the membrane. Such structures are observed in transporters, ion channels and a lipid dephosphorylating enzyme, suggesting a potential ion-coupled transporter-like function and a lipid-binding property for the DedA domain.

RESULTS

VTT/DedA domain-containing proteins form the DedA superfamily

To expand the phylogenetic analysis of the VTT and DedA domain-containing proteins, remote homology search was conducted using HHsearch, a hidden Markov model (HMM)-based method suitable for identifying homologous genes over long evolutionary distances (Steinegger et al., 2019). HHsearch considers both primary sequences and secondary structures, making it more sensitive when primary sequences have diverged among distantly related taxa, such as between eukaryotes and prokaryotes. Using human VMP1 as a query, we identified 125 homologous sequences in 25 species comprising representatives of eukaryotes, bacteria and archaea (34 species in total counting additional sequences included in the phylogenetic analysis; Fig. S1). These sequences include all known proteins containing the VTT domain (TMEM41A, TMEM41B, TMEM64, Tvp38, YdjX and YdjZ, the last two of which are also in the DedA family; Morita et al., 2018) and all eight *E. coli* DedA family proteins (YdjX, YdjZ, YabI, DedA, YohD, YghB, YqjA, and YqaA) (Boughner and Doerrler, 2012; Doerrler et al., 2013), and this is consistent with a previous report that Tvp38 and the DedA family proteins are homologs (Keller and Schneider, 2013). The search also identified a new Pfam family, PF06695, consisting of putative small multi-drug export proteins. The majority of the members in this Pfam family are from bacteria, and their functions are unknown. A similar result (that PF06695 family proteins are remote homologs) was obtained using a PSI-BLAST–HHsearch combination (instead of HHblits–HHsearch), and by turning off the secondary structure scoring option in HHsearch (data not shown). Indeed, even PSI-BLAST alone revealed the remote homology (data not shown).

Alignment of representative sequences from the VTT and DedA domain-containing proteins, including the PF06695 proteins (Fig. 1), shows that the homologous region extends beyond the previously suggested VTT domain (Morita et al., 2018) towards the N terminus by ~30 amino acids. The extended region is predicted to form a helix-loop-helix structure ('reentrant loop 1', as described later). Of all the homologous sequences identified by HHsearch, we found that five bacterial proteins (YqaA in *E. coli*, NP_388110 in *Bacillus subtilis*, YP_002348660 and YP_002348198 in *Yersinia pestis*, and NP_273579 in *Neisseria meningitidis*) and one archaeal protein (WP_048046344 in *Methanosarcina mazei*) consisted almost entirely of

this aligned region alone, suggesting that this region could be a functional unit. Consistent with previous reports (Keller and Schneider, 2013; Tabara et al., 2019), the two motifs [F/Y]XXX[R/K] (motif 1) and GXXX[V/I/L/M]XXXX[F/Y] (motif 2) are conserved in VMP1, Tvp38 and most of the *E. coli* DedA family proteins, but not in TMEM41A, TMEM41B, TMEM64 or the newly identified PF06695 proteins. For *Homo sapiens* (Hs)TMEM41A and HsTMEM41B, motif 1 ends with tyrosine or serine and motif 2 starts with proline. For HsTMEM64, motif 1 starts with histidine and motif 2 starts with serine. For the representative *Clostridium* sp. PF06695 protein (R7M7P8), motif 2 starts with asparagine and ends with alanine. Taken together, these results show that, despite minor differences, these proteins may form a large superfamily with a common origin, and we will hereafter refer to this superfamily as the DedA superfamily and to the shared domain including the extended region as the DedA domain (an extended version of the previously defined VTT domain) (Fig. 1).

To establish the evolutionary relationships among the DedA superfamily proteins, including the newly identified remote homologs, we reconstructed a phylogenetic tree using the Graph Splitting method (Matsui and Iwasaki, 2020) (Fig. 2A). This method outperforms classical methods such as maximum likelihood and Bayesian inference (Felsenstein, 1981; Rannala and Yang, 1996) when sequences are divergent, as it relies on all-to-all pairwise alignment instead of multiple sequence alignment, which shrinks significantly when sequence similarity is low. In the resulting phylogenetic tree, there are four families: the VMP1 family; a family including TMEM41A, TMEM41B, TMEM64 and Tvp38 (referred to as the TMEM41 family hereafter); the DedA family, except for YdjX and YdjZ; and the PF06695 family. Note that Tvp38, which contains the two aforementioned sequence motifs, resides with TMEM41A, TMEM41B and TMEM64, which are devoid of the two motifs, in the TMEM41 family. Bacterial YdjX and YdjZ are in the TMEM41 family, suggesting that they may be evolutionarily closer to the eukaryotic TMEM41 family proteins than to other DedA proteins, in agreement with a previous report (Keller and Schneider, 2013). The VMP1 family is the outmost family in the eukaryotic cluster and is surrounded by the DedA and PF06695 families. Most eukaryotic proteins were found only in the TMEM41 and VMP1 families, but a few plant proteins that possibly localize to the chloroplast and a protein from the SAR (Stramenopiles, Alveolata and Rhizaria) supergroup were also found in the DedA (*Arabidopsis thaliana* NP_193051 and *Solanum lycopersicum* XP_004247084; numbers 66 and 78, respectively, in Fig. 2B) and PF06695 (*Arabidopsis thaliana* NP_178363, *Solanum lycopersicum* XP_004238763 and *Thalassiosira oceanica* K0TKX5; numbers 5, 73 and 111, respectively, in Fig. 2B) families. Bacterial and archaeal proteins were found in all four families. Among homologous sequences from *Candidatus Prometheoarchaeum syntrophicum*, an archaeon very close to the branching point of archaea and eukaryotes (Imachi et al., 2020), one sequence (referred to here as seq2) lies at the center of the TMEM41 family, another sequence (referred to here as seq1) appears in the VMP1 family (Fig. 2A), and a third sequence (referred to here as seq3) is at the periphery of the TMEM41 family, suggesting that there were probably different prokaryotic ancestors for the two eukaryotic families. These patterns can also be seen directly from the sequence similarity network underlying the phylogenetic tree (Fig. 2B; Table S1), where similar sequences are clustered together. In summary, the phylogenetic analysis further supports the existence of the DedA superfamily and suggests that the PF06695 family probably branched out early, with ancestral DedA proteins splitting into three groups and developing into the DedA, VMP1 and TMEM41 families.

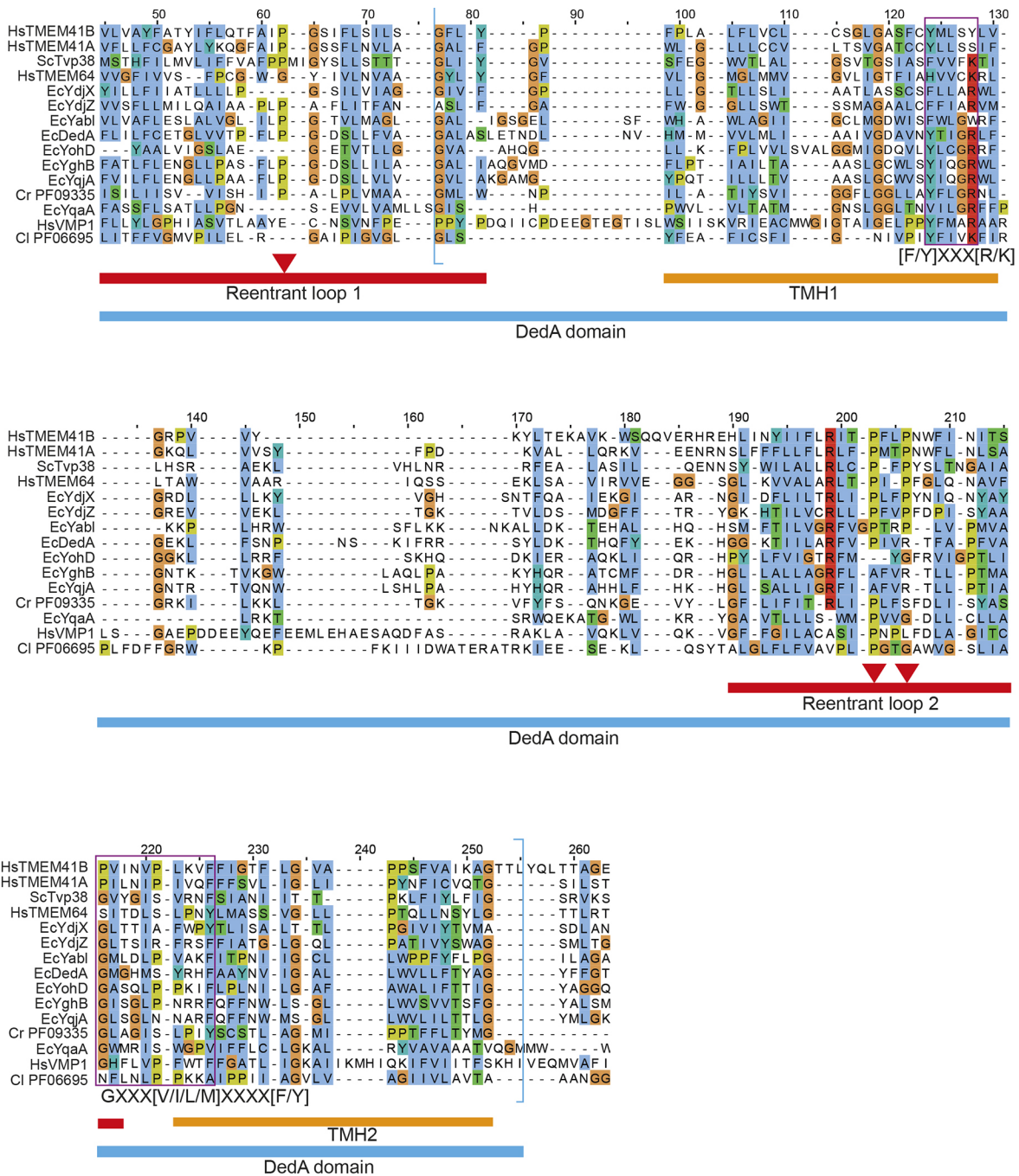


Fig. 1. Multiple sequence alignment of representative sequences from the DedA superfamily. The numbers at the top indicate column numbers in the alignment. At the bottom, the blue bar shows the range of the DedA domain, consisting of reentrant loop 1 and the previously proposed VTT domain (indicated by blue brackets around the sequences) (Morita et al., 2018), red bars show the ranges of reentrant loops with positions of conserved prolines flagged as triangles, and orange bars show the ranges of canonical transmembrane helices (TMHs), as predicted by TMHMM. Purple boxes around the sequences show conserved motifs (Keller and Schneider, 2013; Tabara et al., 2019), with the motif consensus indicated beneath each box. The Clustal X color scheme was used to color residues in the alignment. Hs, *Homo sapiens*; Sc, *Saccharomyces cerevisiae*; Ec, *Escherichia coli*; Cr, *Crocospaera subtropica*; Cl, *Clostridium* sp. Cr PF09335 (B1WQI7) and Cl PF06695 (R7M7P8) are representative sequences from PF09335 and PF06695, respectively.

TMEM41A, TMEM64 and Tvp38 are not required for autophagy

The human genome encodes four DedA superfamily proteins: three TMEM41 family proteins (TMEM41A, TMEM41B and TMEM64) and VMP1, whereas the genome of *Saccharomyces cerevisiae* encodes only one DedA superfamily protein, Tvp38, which belongs to the TMEM41 family. Among these proteins, VMP1 and TMEM41B are known to be required for autophagy (Moretti et al.,

2018; Morita et al., 2018; Ropolo et al., 2007; Shoemaker et al., 2019; Zhao et al., 2017). To determine whether the other three proteins are required for autophagy, we generated TMEM41A- and TMEM64-knockout (KO) HeLa cells and obtained a *tp38Δ* yeast strain (BY4741). In wild-type (WT) unstarved HeLa cells, the amount of the autophagosomal-localizing phosphatidylethanolamine-conjugated LC3 (also known as MAP1LC3B), referred to as LC3-II, increased upon treatment with bafilomycin A₁, an inhibitor of

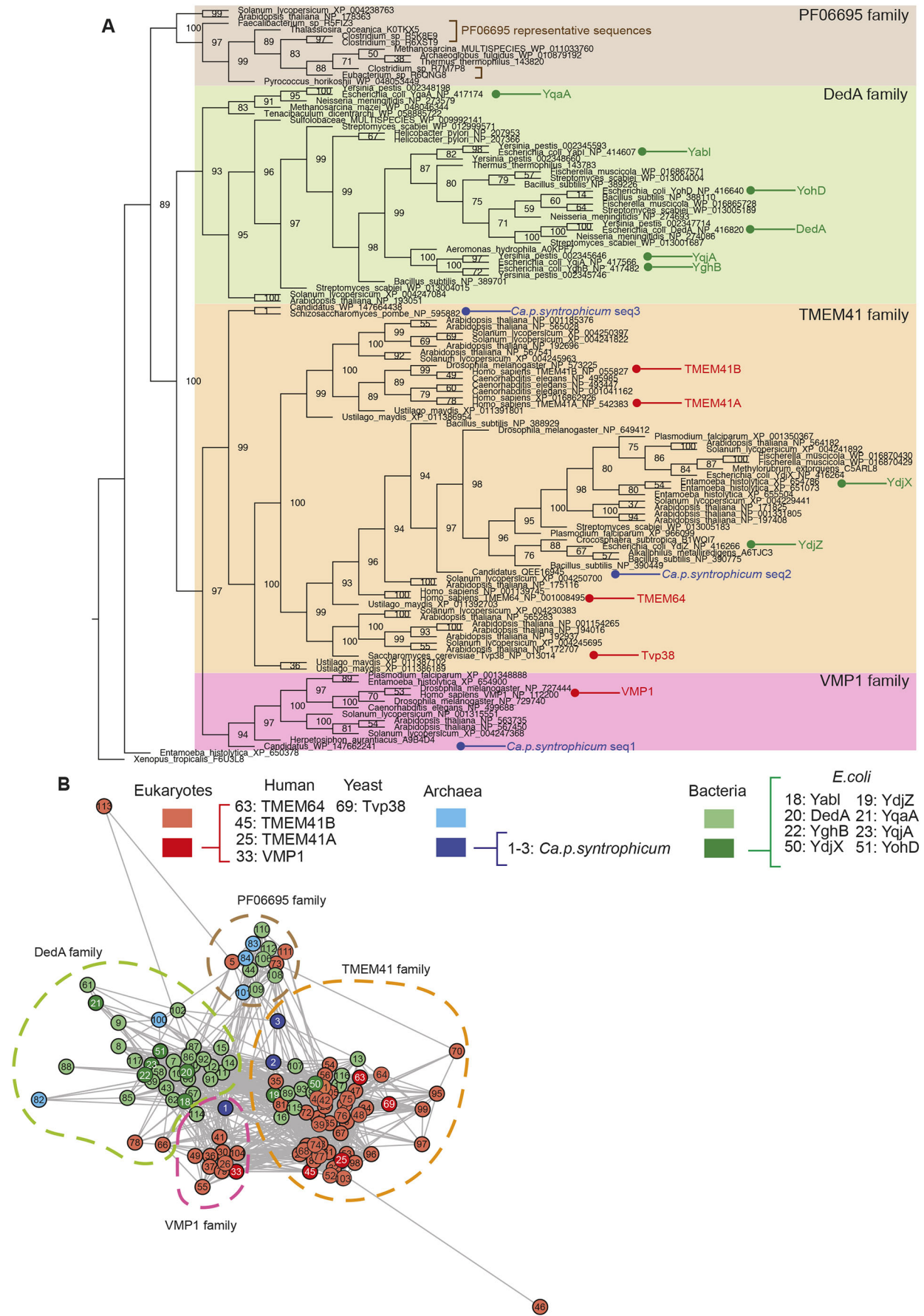


Fig. 2. See next page for legend.

Fig. 2. Evolutionary relationships among the PF06695, DedA, TMEM41 and VMP1 family proteins within the DedA superfamily. (A) Phylogenetic tree of the DedA superfamily proteins in 34 species reconstructed using the Graph Splitting method, with bootstrap values ($n=100$ replications) shown at the nodes. Representative sequences from PF06695 are labeled with brown brackets. (B) Sequence similarity network underlying the phylogenetic tree. Each node represents one sequence, with colors denoting whether each node corresponds to a eukaryotic (red), archaeal (blue) or bacterial (green) protein. Nodes corresponding to the representative sequences highlighted in A are presented in darker colors. Families within the DedA superfamily are circled with dashed lines of the same colors as in A. The distance between nodes is inversely proportional to the degree of sequence similarity (greater distance indicating less similar sequences). A list of the node numbers and the corresponding sequence names and family names is shown in Table S1.

vacuolar ATPase, indicating that autophagosomal LC3 was degraded in lysosomes by basal autophagy (Fig. 3A). Under starvation conditions, further accumulation of LC3-II was observed upon bafilomycin A₁ treatment, suggesting an increase in autophagic flux during starvation. By contrast, in VMP1-KO cells, LC3-II accumulated even under nutrient-rich conditions, and the accumulation was not further increased by starvation or bafilomycin A₁ treatment, suggesting that autophagic flux was blocked. Consistently, p62 (also known as SQSTM1) and its phosphorylated form, which are selective substrates of autophagy, accumulated in VMP1-KO cells. Similarly, in TMEM41B-KO cells, LC3-II and phosphorylated p62 accumulated under both nutrient-rich and starvation conditions compared with their levels in WT cells, suggesting that autophagic activity was defective, although less severely than in VMP1-KO cells. The lysosomal turnover of LC3-II and the expression level of p62 were normal in TMEM41A-KO cells, as previously shown in TMEM41A-knockdown cells (Morita et al., 2018), and in TMEM64-KO cells (Fig. 3A). We also determined autophagic flux by a quantitative method using the autophagic flux reporter GFP-LC3-mRuby3, a variation of GFP-LC3-RFP (Kaizuka et al., 2016). This reporter is cleaved into GFP-LC3 and mRuby3 by endogenous ATG4 proteases, and GFP-LC3, but not mRuby3, is degraded by autophagy. Thus, a reduction in the GFP:mRuby3 ratio represents autophagic flux. We measured autophagic flux upon treatment with Torin 1, an inhibitor of mTOR. Compared with that of WT cells, autophagic flux was significantly reduced in VMP1-KO and TMEM41B-KO cells but not in TMEM41A-KO and TMEM64-KO cells (Fig. 3B). Thus, we concluded that TMEM41A and TMEM64 are not required for autophagy.

Autophagic flux in yeast was determined by monitoring the cleavage of GFP-Atg8, which was expressed in the cytosol and is degraded after delivery to the vacuole by autophagy (Klionsky et al., 2016). In WT *Saccharomyces cerevisiae* (BY4741), cleaved GFP accumulated upon treatment with autophagy-inducible rapamycin, an inhibitor of TORC1. By contrast, in *atg1Δ* cells, cleaved GFP did not accumulate even after rapamycin treatment, indicating that autophagic activity was deficient. In *tpv38Δ* cells, cleaved GFP normally accumulated after rapamycin treatment, suggesting that autophagic activity was maintained (Fig. 3C). Thus, TMEM41A, TMEM64 and Tvp38 are not required for autophagy.

Prediction of the DedA domain structure

Structural information of the DedA superfamily has been limited to secondary structure predictions, which suggest that members might contain 5–8 TMHs (Morita et al., 2018). However, the assignment of TMHs may not be accurate, because even the conserved DedA domain has been suggested to carry different numbers of TMHs depending on protein and species. To gain more reliable structural

information and functional insights, we conducted *ab initio* structural prediction using trRosetta (Yang et al., 2020). Building on the assumption that coevolving residues are often in contact, trRosetta predicts distance and orientation between residues from sequence coevolution using deep learning. The accuracy of trRosetta prediction relies on the number and depth of the homologous sequences collected. In our case, 65,535 homologous sequences (the default upper limit) were used for TMEM41A, TMEM41B, TMEM64 and YdjX, and 24,330 homologous sequences were used for YdjZ (including overlapping sequences between them), yielding reliable structural predictions. Conversely, predictions for VMP1 yielded results of low or medium quality owing to a relatively small number of homologous sequences. We therefore focused on the TMEM41 family in further analyses.

The prediction for TMEM41B generated a distance map with two ring-like patterns in the N- and C-terminal regions (Fig. 4A). Each of these ring-like patterns translates into a reentrant loop that enters the lipid bilayer but turns inside the membrane to exit from the same side. Notably, the first third of each ring was predicted to have contacts (i.e. predicted distances less than 8 Å) between the two halves of the reentrant loops (e.g. L124–L135, Y121–L135 and Y117–S139 in reentrant loop 1, and L204–I215 and I201–S219 in reentrant loop 2) (Fig. 4B). Furthermore, the reentrant loops contain helix-breaking proline and glycine residues between the halves. Reentrant loop 1 turns roughly at the conserved proline-glycine residues (P130 and G131 in TMEM41B), and reentrant loop 2 turns at two conserved prolines separated by one or two other residues (P208 and P211 in TMEM41B) (Fig. 1; Fig. 4C). In addition to the contact areas within the reentrant loops, the contact map also suggests interactions between each of the reentrant loops and the TMHs, as well as between the TMHs – for example, contacts between the first half of reentrant loop 1 and TMH1 (the pink rectangle in Fig. 4A) and contacts between TMH1 and TMH2 (the orange rectangle in Fig. 4A). As shown, the two hairpin-shaped reentrant loops and two additional TMHs together form into a compact fold, suggesting that the DedA domain could be an independent structural domain. The predictions for TMEM41A, TMEM64, YdjX and YdjZ all yielded similar contact maps (Fig. S2) and structures (Fig. 4C). Along this line, we found that the GREMLIN structural prediction server (<https://gremlin2.bakerlab.org/structures.php>), which also uses coevolution information, lists similar contact maps and structures for the *Clostridium sp.* R6BJC6 protein used as a representative of the PF06695 family (Ovchinnikov et al., 2014). We were also able to obtain similar contact maps and predicted structures using a different prediction method, EVfold (Hopf et al., 2019) (Fig. S3).

The predicted structure reveals a characteristic organization of two reentrant loops facing each other in the membrane. In order to gain insights into its molecular function, we searched the PDBTM database (Kozma et al., 2013), a database of annotated transmembrane proteins with solved structures, for other proteins with two reentrant loops (using an advanced search with the keywords ‘0 [type] AND 2 [n_loop]’). Such reentrant loops were found in transporters and ion channels such as aquaporins (AQPs), chloride channels (CLCs), solute carrier family 1 (SLC1) proteins, solute carrier family 13 (SLC13) proteins, solute carrier family 28 (SLC28) proteins and bacterial undecaprenyl pyrophosphate phosphatase (UppP) (Chang et al., 2014; Forrest, 2015; Kanai et al., 2013; Screpanti and Hunte, 2007) (Fig. S4A; Table S2). Among them, the topology of the substrate-binding region of SLC1 proteins is most similar to that of the DedA domain: both consist of two repeats of a reentrant loop and a succeeding TMH, and the

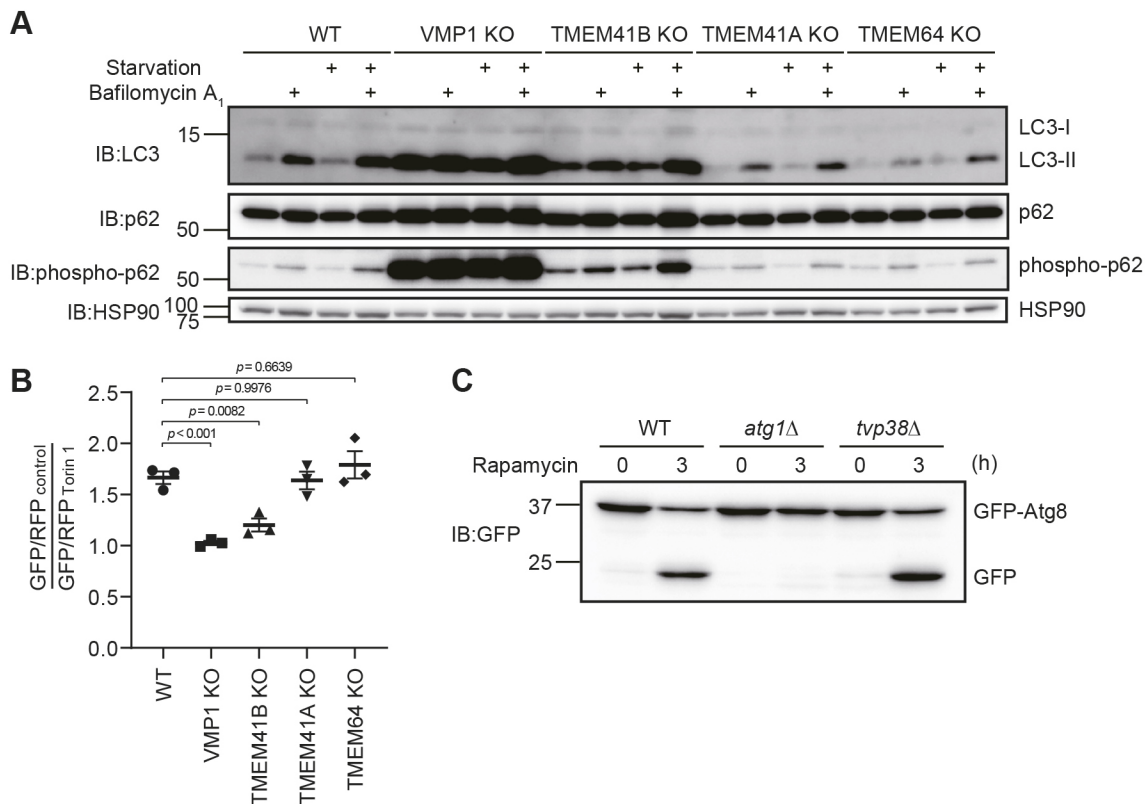


Fig. 3. Autophagic activity in TMEM41A-KO, TMEM64-KO and *tvp38Δ* cells. (A) Autophagic flux of TMEM41A-KO and TMEM64-KO HeLa cells. WT and the indicated mutant cells were cultured under nutrient-rich or starvation conditions with or without bafilomycin A₁ for 2 h. Cell lysates were immunoblotted (IB) using antibodies against the indicated proteins (LC3-I, unlipidated LC3). Data are representative of three independent experiments. (B) Autophagic flux determined using the GFP–LC3–mRuby3 reporter. The indicated cell lines expressing GFP–LC3–mRuby3 were cultured with and without Torin 1 for 6 h. Fluorescence intensities were determined by flow cytometry (2500 cells for each experiment). The ratio of GFP:mRuby3 (GFP/RFP) under control (without Torin 1) and Torin 1 treatments is presented. Data represent the mean ± s.e.m. of three independent experiments. *P*-values were calculated by Dunnett's multiple comparisons test. (C) GFP–Atg8 cleavage assay using WT, *atg1Δ* and *tvp38Δ* yeast cells. Cells were cultured with or without rapamycin for 3 h. Data are representative of two independent experiments. Size markers in A and C indicate molecular mass in kDa.

membrane topologies of the two repeats are inverted (Kanai et al., 2013). Consistently, SLC1 shows a coevolution pattern similar to that of the DedA domain (Fig. S4B,C). Furthermore, the two facing reentrant loops are directly involved in substrate binding in several transporters, including SLC1 proteins (Johnson et al., 2014; Kanai et al., 2013; Mancusso et al., 2012; Workman et al., 2018; El Ghachi et al. 2018) (Fig. S4C). Thus, the two facing reentrant loops of the DedA domain might also serve as a substrate-binding site for potential ion-coupled transporters.

Biochemical verification of the topology of TMEM41B

Structural prediction by trRosetta and EVfold suggests that the DedA domain contains reentrant loop 1, TMH1, an extra-membrane region, reentrant loop 2 and TMH2 (from the N to C terminus) (Fig. 4). In addition, TMHMM (Krogh et al., 2001) predicted that TMEM41B has two more TMHs outside of the DedA domain, at the N- and C-terminal ends (Fig. 5A; Fig. S4A) (Moller et al., 2001). To verify the predicted topology of TMEM41B experimentally, we performed substituted cysteine accessibility method (SCAM) analysis (Bogdanov et al., 2005). Cysteine has a thiol group that can be conjugated with maleimide or maleimide-containing molecules such as methoxypolyethylene glycol maleimide (PEG-maleimide) and *N*-ethylmaleimide (NEM). Once a protein is conjugated with PEG-maleimide, it becomes larger and can be separated from an unconjugated form by SDS–PAGE (Fig. 5B) (Davis et al., 2019). As PEG-maleimide is cell-impermeable,

specific labeling is achieved only after membrane permeabilization with a detergent (Fig. 5C).

We prepared cells expressing cysteine-less TMEM41B or its variants in which one of the amino acids was replaced with cysteine (yellow residues in Fig. 5A). Upon PEG-maleimide (molecular mass 5000 Da) treatment, cysteine-less TMEM41B did not show any band shift, even in the presence of the mild detergent digitonin (Fig. 5D). By contrast, the single-cysteine TMEM41B mutants S35C, S187C and 292C (in which cysteine was added to the C terminus) showed an additional high molecular weight band in the presence of both PEG-maleimide and digitonin, which permeabilized the plasma membrane. The intensity of these high molecular weight bands was unchanged upon Triton X-100 treatment, which permeabilizes organellar membranes as well as the plasma membrane (Fig. 5C,D). Formation of these bands was inhibited by pretreatment with NEM, which blocked the conjugation between PEG-maleimide and the thiol group of cysteine. These results suggest that these high molecular weight bands represent PEG-conjugated TMEM41B and that S35, S187 and the C terminus are in the cytosol. Conversely, the single-cysteine mutants S79C and A257C did not produce TMEM41B–PEG bands in the presence of digitonin, but did so in the presence of Triton X-100, suggesting that these residues are present in the lumen of the ER (Fig. 5D).

We also tested whether these single-cysteine TMEM41B mutants retained their original topologies by assessing their function in

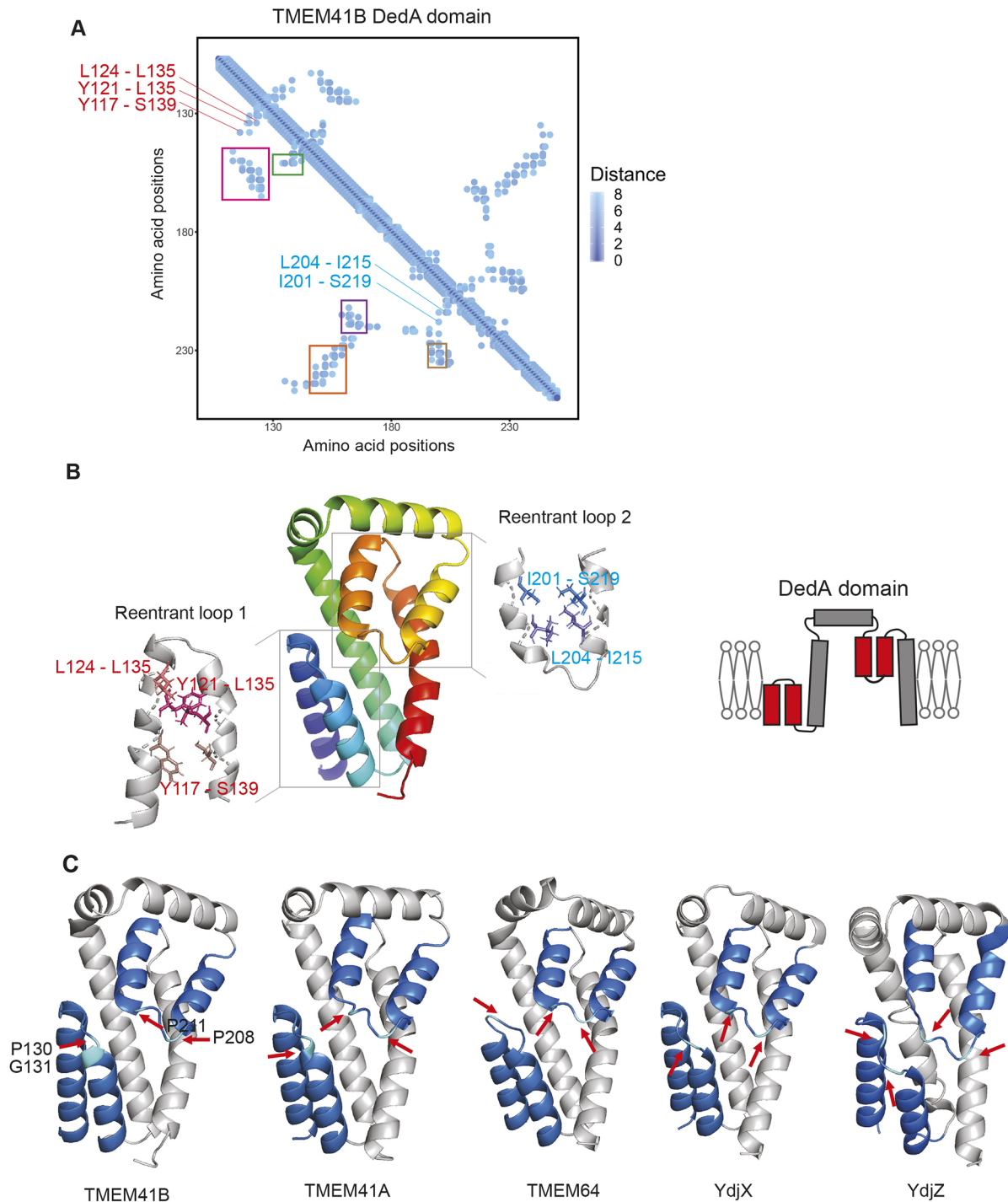


Fig. 4. *Ab initio* structure prediction of the DedA superfamily proteins using trRosetta. (A) Distance map of the DedA domain of TMEM41B. The x- and y-axes show amino acid positions in TMEM41B, and the color gradient indicates the predicted distances (in Å) between residue pairs. Examples of predicted interactions between the two halves of reentrant loops 1 and 2 are labeled in red and blue, respectively. The rectangles indicate contacts between the first half of reentrant loop 1 and TMH1 (pink), between the second half of reentrant loop 1 and TMH1 (green), between the second half of reentrant loop 2 and TMH1 (purple), between TMH1 and TMH2 (orange), and between the first half of reentrant loop 2 and TMH2 (brown). (B) Top-ranking model of TMEM41B predicted by trRosetta, with reentrant loops 1 and 2 (boxes) enlarged. Different colors in the enlargements indicate example pairs of residues predicted to be in contact (i.e. with predicted distance less than 8 Å) by coevolution. A model of membrane topology of the DedA domain is also shown (right), with reentrant loops in light blue. (C) Predicted models of TMEM41B, TMEM41A, TMEM64, YdjX and YdjZ, with reentrant loops in dark blue and conserved proline residues in light blue. Conserved proline residues at which the reentrant loops turn are labeled with red arrows.

autophagy. In TMEM41B-KO cells, we observed accumulation of LC3-II, representing a block in autophagic flux (Moretti et al., 2018; Morita et al., 2018; Shoemaker et al., 2019). This defect was restored by exogenous expression of WT TMEM41B or of the

individual single-cysteine TMEM41B mutants, suggesting that these mutants are correctly integrated into the membrane (Fig. S5). Thus, these results verified the topology predicted by trRosetta and EVfold; the presence of S79 and S187 on opposite sides of the

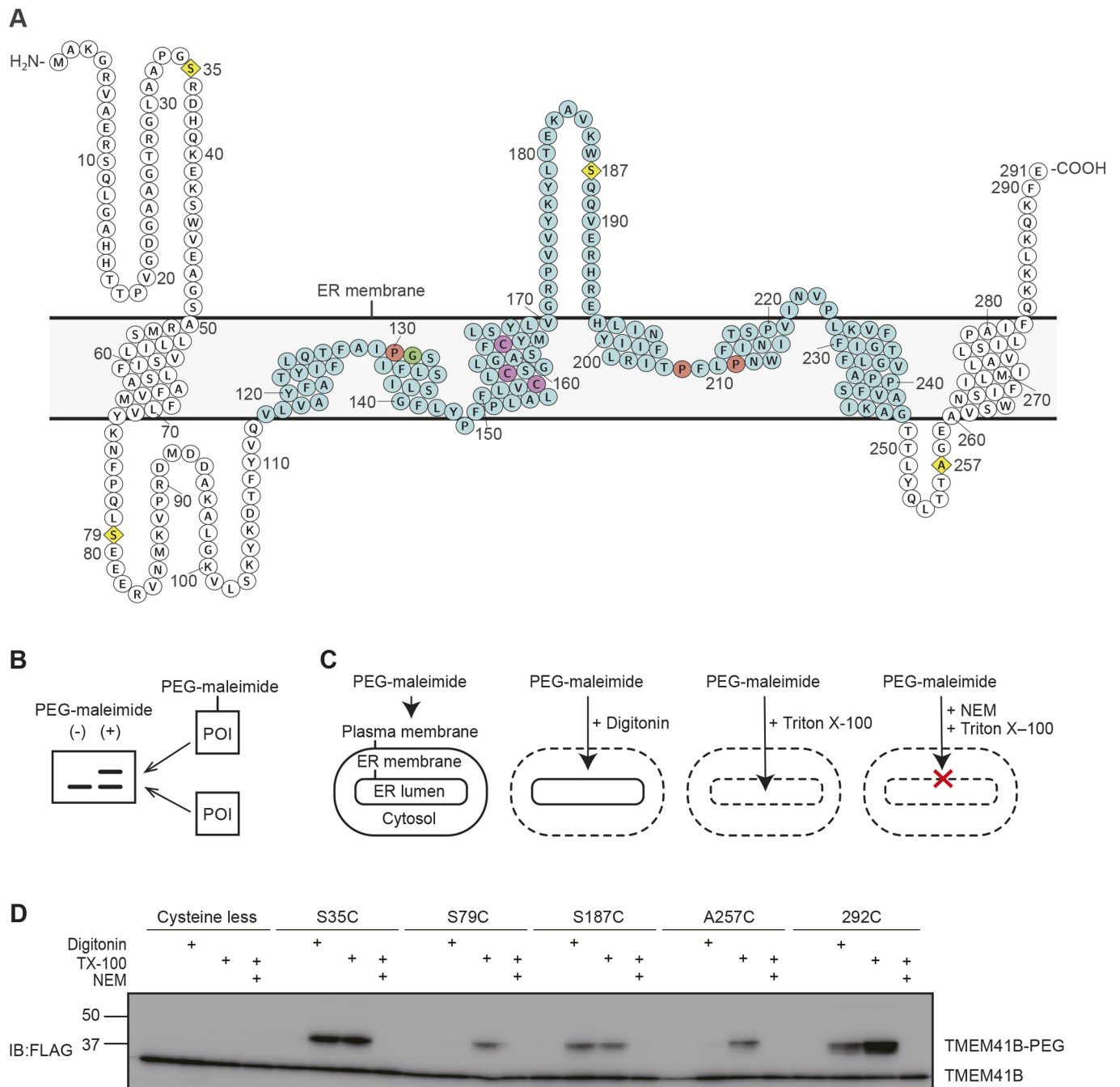


Fig. 5. SCAM-based topology analysis of TMEM41B. (A) Residues in the DedA domain of TMEM41B are shown as blue circles. Residues mutated in the SCAM analysis (yellow diamonds), conserved prolines (orange circles) and glycine (green circle) located in the reentrant loops, and endogenous cysteines (magenta circles) are marked. (B) Schematic representation of the effect of PEG-maleimide treatment in SDS-PAGE. If PEG-maleimide is conjugated with a protein-of-interest (POI), a higher molecular weight band will appear. (C) Experimental design for protein topological analysis of ER proteins. Dotted lines indicate permeabilized membranes. When the plasma membrane is permeabilized by digitonin, PEG-maleimide can be conjugated with cysteine residues located in the cytosol. When the ER membrane is permeabilized by Triton X-100, PEG-maleimide is able to penetrate into the ER lumen and be conjugated with cysteine residues there. Conjugation between PEG-maleimide and cysteine is inhibited by NEM treatment. (D) SCAM-based topology analysis of TMEM41B. HeLa cells expressing each single-cysteine mutant were treated with indicated reagents (TX-100, Triton X-100), and lysates were immunoblotted (IB) with an antibody against FLAG to detect TMEM41B. Mutants were constructed in a cysteine-less TMEM41B background. The 292C mutant is cysteine-less TMEM41B with an additional cysteine on the C terminus. Size markers indicate molecular mass in kDa. Data are representative of two independent experiments.

membrane suggests that reentrant loop 1 indeed turns back in the membrane rather than penetrating the membrane, and the presence of S187 and A257 on opposite sides suggests that reentrant loop 2 is indeed reentrant. Collectively, these results suggest that TMEM41B is composed of four TMHs and two reentrant loops facing each other, and both the N and C terminus are in the cytosol.

DISCUSSION

Evolution of the DedA superfamily proteins and acquisition of autophagic function

In this study, we conducted a phylogenetic analysis of the DedA superfamily, containing the DedA, TMEM41, VMP1 and PF06695 families, all of which possess the DedA domain. Although the

DedA and PF06695 families and the TMEM41 and VMP1 families primarily contain prokaryotic and eukaryotic proteins, respectively, each of these four families contains both prokaryotic and eukaryotic proteins (Fig. 2), indicating that the origins of these four families pre-date the split between prokaryotes and eukaryotes. Among eukaryotic proteins, a few proteins from plants and the SAR supergroup were grouped with the DedA and PF06695 families. Because these proteins appear to exist in only limited lineages, they might have been transferred from the chloroplast to the nuclear genome after these lineages separated from other eukaryotes.

Among the DedA superfamily members, only VMP1 and TMEM41B have a role in autophagosome formation, whereas TMEM41A, TMEM64 and Tvp38 do not (Fig. 3) (Moretti et al., 2018; Morita et al., 2018; Shoemaker et al., 2019). Similarly, flaviviral replication requires VMP1 and TMEM41B, but not TMEM41A and TMEM64 (Hoffmann et al., 2021). Notably, although prokaryotes do not have an autophagy system or lysosomes, they do have ancestors of both VMP1 and TMEM41. There are several possible scenarios of how the prokaryotic DedA ancestors acquired autophagic functions in the course of evolution. One is that the prokaryotic ancestors of VMP1 and TMEM41 had a common function at the plasma membrane, which was later directly used in autophagy in eukaryotes. In this case, these proteins might have spontaneously acquired their autophagic function after translocation to the ER membrane. However, this hypothesis cannot explain why most TMEM41 family members do not have an autophagic function. Even TMEM41A, the closest homolog of TMEM41B, does not play a role in autophagy. Also, Tvp38, the only TMEM41 family protein in yeast, is dispensable for autophagy.

An alternative scenario is that a VMP1 ancestor acquired autophagic function during evolution first, probably in a eukaryotic ancestor. Accordingly, the autophagic function of VMP1 is conserved broadly in eukaryotes, such as in Metazoa and Amoebozoa (Calvo-Garrido et al., 2008) and probably also in green algae (Tenenboim et al., 2014). Later, probably after diverging from TMEM41A, TMEM41B became involved in autophagy, with this new function of TMEM41B being dependent on the preexisting autophagic function of VMP1, for example, through binding to VMP1 (Morita et al., 2018). This hypothesis can explain why only TMEM41B is involved in autophagy among the TMEM41 family proteins. It is also consistent with the previous observation that the role of TMEM41B is rather accessory; the phenotype of TMEM41B-KO cells is milder than that of VMP1-KO cells, and overexpression of VMP1 can rescue the phenotype of TMEM41B-KO cells, whereas overexpression of TMEM41B cannot rescue the VMP1-KO phenotype (Morita et al., 2018; Shoemaker et al., 2019; Hoffmann et al., 2021). A more comprehensive analysis of the function of VMP1 and TMEM41 family proteins in non-metazoan eukaryotes will provide further crucial information on how DedA family proteins acquired their autophagic function.

Potential functions of the DedA superfamily proteins based on the predicted structure

The next fundamental unresolved issue is the function of the evolutionarily conserved DedA superfamily proteins. Starting with the HMM-based alignment, we first specified the core domain conserved in these proteins, which was rather ambiguously specified previously (Doerrler et al., 2013; Keller and Schneider, 2013; Morita et al., 2018), and we defined it as the DedA domain. Our prediction shows that the DedA domain forms a characteristic structure with two reentrant loops and two TMHs. This topology

was verified experimentally (Fig. 5). Thus, the structure of the proposed DedA domain differs from the previous speculation that the DedA family proteins adopt half of a LeuT fold-like structure (Keller et al., 2014; Khafizov et al., 2010). During the preparation of this article, a similar structural prediction was reported by Mesdaghi et al. (2020).

The reentrant loops in transporters and ion channels often directly interact with substrates (Johnson et al., 2012; Kanai et al., 2013; Mancusso et al., 2012; Tornroth-Horsefield et al., 2010; Workman et al., 2018; El Ghachi et al. 2018). Among them, the local architecture around the substrate-binding site of the Na⁺-coupled glutamate transporter SLC1 (Fig. S4A) is highly similar to that of the DedA domain (Kanai et al., 2013). Therefore, the DedA domain might have an ion-coupled transport function. It is tempting to speculate that VMP1 and TMEM41B are Ca²⁺-coupled transporters, because VMP1 physically interacts with and is functionally related to SERCA, a Ca²⁺ transporter in the ER (Zhao et al., 2017). With regard to the potential substrate, we note that bacterial UppP uses a similar pair of reentrant loops to bind the head groups of membrane lipids and catalyze their dephosphorylation within the membrane (Workman et al., 2018; El Ghachi et al. 2018). Although the catalytic residues are not conserved in the DedA domain, and the overall topologies are not identical between UppP and the DedA domain (additional elements, including two TMHs, are inserted between the two internal repeats in UppP), the linkage between this structural feature and the lipid binding of UppP suggests that the DedA domain may recognize the head groups of membrane lipids as substrates. This hypothesis aligns well with the lipid-related phenotypes observed in VMP1- and TMEM41B-deficient eukaryotic cells (Calvo-Garrido et al., 2008; Kang et al., 2020; Moretti et al., 2018; Morishita et al., 2019; Morita et al., 2018; Ropolo et al., 2007; Shoemaker et al., 2019; Zhao et al., 2017), and YqjA- and YghB-deficient bacterial cells (Boughner and Doerrler, 2012; Thompkins et al., 2008). The slight phenotypic difference between VMP1 and TMEM41B deficiency may represent a difference in their substrates. Furthermore, several genetic studies suggest ion-dependent solute exporting functions for YqjA and YghB (Boughner and Doerrler, 2012; Doerrler et al., 2013; Keller et al., 2015; Kumar et al., 2016; Kumar and Doerrler, 2014; Ledgham et al., 2005; Panta et al., 2019). Collectively, predicted structural similarities suggest that the DedA superfamily proteins could have ion-dependent lipid or solute transport functions.

Reentrant loops are generally not hydrophobic enough to be stably embedded in membranes; they are often stabilized by surrounding TMHs (Yan and Luo, 2010) and/or participate in the subunit interface in a complex (Table S2). Thus, DedA superfamily proteins may also form similar complexes. Determining the actual structure of the DedA superfamily will eventually reveal the function of this broadly conserved family of proteins.

MATERIALS AND METHODS

Remote homology search

Remote homology search was conducted using HHsearch (Steinegger et al., 2019), part of the MPI Bioinformatics Toolkit (Zimmermann et al., 2018). The 'local:realign' option was used to enable the maximum accuracy algorithm for more accurate alignment. Full-length VMP1 sequence (NP112200.2) was used as the query, and Pfam (El-Gebali et al., 2019) as well as the proteomes of *Homo sapiens*, *Saccharomyces cerevisiae*, *Escherichia coli* and twenty randomly selected eukaryotic, bacterial and archaeal species were used as the search database. A total of 125 homologous sequences (E-value cutoff=1), including one representative sequence from each of the Pfam PF09335 SNARE-associated Golgi protein

family and PF06695 putative small multi-drug export protein family, were identified. PF09335 was renamed as the VTT family (Morita et al., 2018). A pool of 3172 TMEM41B homologs, 2624 VMP1 homologs and 185 PF06695 family proteins were collected, and 500 sequences were randomly selected from this pool after eliminating redundant sequences at 95% sequence similarity threshold. Then, representative sequences shown in Fig. 1 were added and aligned together with these 500 sequences using MUSCLE v3.8.1551 (Edgar, 2004). Only the representative sequences are shown in Fig. 1. The TMEM41B and VMP1 homologs were collected using the GREMLIN server (gremlin.bakerlab.org/submit.php) and the PF06695 family sequences were seed sequences of this Pfam family downloaded from Pfam in January 2020.

Phylogeny reconstruction

The phylogenetic tree of the DedA superfamily was reconstructed using the Graph Splitting method (Matsui and Iwasaki, 2020) with default parameters. In addition to the homologous sequences identified, five additional sequences from PF09335 and PF06695 each, as well as three sequences from the archaeon *Ca. P. syntrophicum*, were also included. After eliminating redundant sequences with 95% sequence similarity, 117 sequences (from 34 species in total; Fig. S1) remained (Table S1). Randomly selecting different sets of sequences from the two Pfam families did not change the result. While Graph Splitting does not estimate branch length, bootstrap values (number of replicates=100) are displayed at the nodes of the phylogenetic tree.

Structural prediction based on coevolution

Structural prediction was conducted using trRosetta (Yang et al., 2020) and EVfold (Hopf et al., 2019) with default parameters. For trRosetta, residue pairs with predicted distance less than 8 Å were used to make the distance plot. As the five predicted models produced for each protein were highly similar, the first model was plotted. For EVfold, at the multiple-sequence-alignment-building step, significance thresholds of 0.2, 0.3 and 0.4 (bitscore/sequence length) were tested, and 0.4 was ultimately chosen because it produced the clearest ring-like pattern on the N-terminal side (Fig. S3). For each protein, the top-ranking model was selected. The models were plotted using Pymol (<https://pymol.org/2/>).

Antibodies

The following antibodies were used for immunoblotting: mouse monoclonal antibodies against HSP90 (1:1000; 610419; BD), FLAG tag (1:1000; F1804; Sigma-Aldrich) and GFP (1:1000; 11814460001; Roche) and rabbit polyclonal antibodies against p62 (SQSTM1; 1:1000; PM045; MBL) and phospho-p62 (1:1000; PM074; MBL). The rabbit polyclonal antibody against LC3 was described previously (Hosokawa et al., 2006). Peroxidase-conjugated anti-mouse IgG (315-035-003; Jackson ImmunoResearch Laboratories, Inc.) and anti-rabbit IgG (111-035-144; Jackson ImmunoResearch Laboratories, Inc.) were used as secondary antibodies.

Plasmids

The pMRXIP-TMEM41B-3×FLAG plasmid encoding human TMEM41B was described previously (Morita et al., 2018) and used as a template for making the single-cysteine mutants. Three endogenous cysteines at 153, 155 and 163 were mutated to serines, and the product was utilized to make each single-cysteine mutant. PrimeSTAR Max DNA Polymerase (R045A; Takara Bio Inc.) was used for mutagenesis. Preparations of primers and mutagenesis steps followed the manufacturer's instructions. Each generated construct was confirmed by sequencing (Eurofins Genomics JP). For the generation of knockout cell lines, guide RNA (gRNA) targeting TMEM41B (5'-GTCCG-CCGAACGATCGCAGTT-3'), VMP1 (5'-CTTTTGTATGCTACTGGAT-3'), TMEM41A (5'-GCCGAGAAGCGGGCGCATGT-3') and TMEM64 (5'-CCGCGCTGGGCGGAGGCATG-3') were cloned into pSpCas9(BB)-2A-GFP (Addgene 48138; deposited by Dr Feng Zhang, Broad Institute of Massachusetts Institute of Technology and Harvard, USA). Additionally, pRS416-GFP-ATG8 (Addgene 49425; deposited by Dr Daniel J. Klionsky, Life Sciences Institute and Department of Molecular, Cellular and

Developmental Biology, University of Michigan, Ann Arbor, MI, USA) was used for the GFP-Atg8 assay. pMRXIP-GFP-LC3-mRuby3 was generated by inserting human codon-optimized mRuby3 (Addgene 74252; deposited by Dr Michael Z. Lin, Department of Bioengineering, Stanford University, CA, USA) into the pMRX-IP vector (Saitoh et al., 2002) along with EGFP and *Rattus norvegicus* LC3B (Q62625), and used for flow cytometry analysis.

Cell culture

Using Dulbecco's modified Eagle's medium (DMEM; D6546; Sigma-Aldrich) supplemented with 10% fetal-bovine-serum (FBS) and 2 mM glutamine (25030-081; Gibco), HeLa cells, authenticated by RIKEN, were cultured in a 5% CO₂ incubator. To impose starvation conditions, cells were washed with phosphate-buffered saline (PBS) and cultured in amino acid-free DMEM (048-33575; Wako) without FBS. For vacuolar ATPase inhibition, cells were cultured with 100 nM bafilomycin A₁ (B1793; Sigma-Aldrich) for 2 h. For flow cytometry, cells were treated with 500 nM Torin 1 (4247, Tocris Bioscience) in DMEM for 6 h.

Generation of stable cell lines

HEK293T cells were transiently transfected with a target plasmid, pCG-VSV-G and pCG-gag-pol (gifts from Dr T. Yasui, Osaka University, Japan) using Lipofectamine 2000 (11668019; Thermo Fisher Scientific). Two days after transfection, culture medium including retrovirus was collected through a 0.45 µm syringe filter unit (SLHV033RB; Merck Millipore). Retrovirus was mixed with 8 µg/ml polybrene (H9268; Sigma-Aldrich), and host cells were transfected with the mixture. After 24 h, the medium was exchanged to DMEM containing 2 µg/ml puromycin (P8833; Sigma-Aldrich) for selection.

Establishment of TMEM41B-KO, VMP1-KO, TMEM41A-KO and TMEM64-KO HeLa cells

HeLa cells were transfected with pSpCas9(BB)-2A-GFP encoding gRNAs using FuGENE HD Transfection Reagent (E2311; Promega). Two days after transfection, GFP-positive cells were isolated using a cell sorter (MoFlo Astrios EQ; Beckman Coulter), and single clones were obtained. Clones containing knockout mutations were selected by immunoblotting and sequencing of genomic DNA.

Immunoblotting

Cells were collected in ice cold PBS using a cell scraper and centrifuged at 5000 g for 3 min. They were then treated with lysis buffer [1% Triton X-100, 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA and protease inhibitor cocktail (03969; Nacalai Tesque)] and incubated on ice for 15 min. Then, lysed samples were centrifuged at 12,000 g for 15 min, and the resulting supernatants were collected for analysis. SDS-PAGE sample buffer [46.7 mM Tris-HCl pH 6.8, 5% glycerol, 1.67% sodium dodecyl sulfate, 1.55% dithiothreitol (DTT) and 0.02% Bromophenol Blue] was added to samples and boiled. SDS-PAGE was conducted to separate proteins, which were transferred to a polyvinylidene difluoride (PVDF) membrane. Appropriate antibodies were applied to the membrane after blocking with Tris-buffered saline with Tween 20 (TBST) containing 5% skim milk. Membranes were incubated with primary antibodies at 4°C overnight, followed by incubation with secondary antibodies at room temperature for an hour. After washing and reacting the membranes with Super-Signal West Pico Chemiluminescent substrate (1856135; Thermo Fisher Scientific), signals were detected using a FUSION Solo S imaging system (Vilber-Lourmat). Contrast and brightness adjustments were performed using Fiji software (Schindelin et al., 2012).

Modification of cysteine residues

HeLa cells were transiently transfected with plasmids encoding the TMEM41B single-cysteine mutants using Lipofectamine 2000 (11668019; Thermo Fisher Scientific). After 24 h, the plasma membrane was permeabilized using DMEM containing 100 µg/ml digitonin (12333-51; Nacalai Tesque) for 3 min at 37°C. To permeabilize both the plasma and

ER membranes, cells were treated with 0.1% Triton X-100 in PBS containing 0.1 mM CaCl₂ and 1 mM MgCl₂ (PBSCM) for 3 min at room temperature. Detergents were removed, and cells were washed with PBS. Then, *N*-ethylmaleimide (NEM; 15512-11; Nacalai Tesque) was diluted to 5.0 mM using PBS, and cells were incubated for 1 h on ice using a rocker before permeabilization with 0.1% Triton X-100. Methoxypolyethylene glycol maleimide (PEG-maleimide; 63187; SIGMA) was diluted to 1.5 mM using PBSCM. After membrane permeabilization, cells were incubated in PEG-maleimide solutions for 30 min on ice using a rocker. PEG-maleimide modification was stopped by a solution containing 10 mM DTT (14112-52; Nacalai Tesque) in PBSCM with 2% bovine serum albumin (BSA) for 10 min on ice using a rocker. After removing the solution, cells were collected in ice cold PBS and centrifuged at 5000 *g* for 3 min. Cells were broken by passing through a 26-gauge needle 20 times with lysis buffer [0.1% Triton X-100, 250 mM sucrose, 25 mM Tris-HCl (pH 7.5), 2 mM DTT and protease inhibitor cocktail (03969; Nacalai Tesque)] and incubated on ice for 15 min. Finally, lysates were centrifuged at 100 *g* for 10 minutes, and supernatants were collected as samples.

Yeast cells and GFP-Atg8 cleavage assay

The yeast knockout haploid MATA collection (TKY3502; TOT) was obtained from Funakoshi Co., Ltd. After confirmation of knockout by PCR, the *typ38Δ* and *atg1Δ* strains were used for experiments. Cells were transformed with pRS316-GFP-Atg8 as previously described (Gietz and Woods, 2002). The GFP-Atg8 cleavage assay was performed as previously described (Cheong and Klionsky, 2008).

Flow cytometry

Trypsinized wild-type, VMP1-KO, TMEM41B-KO, TMEM41A-KO and TMEM64-KO HeLa cells expressing GFP-LC3-mRuby3 were harvested and centrifuged at 2000 *g* for 2 min. After washing, the cells were diluted with ice-cold PBS and analyzed using a cell analyzer (EC800, SONY) equipped with 488 nm and 561 nm lasers. Data were processed using Kaluza Analysis 2.1 software (Beckman Coulter).

Statistical analysis

For statistical analysis, GraphPad Prism 8 software (GraphPad Software) was used. The statistical method is specified in the figure legends.

Acknowledgements

We would like to thank Drs Wataru Iwasaki, Motomu Matsui and Euki Yazaki for helpful discussions on the phylogenetic analysis conducted, and Dr Norito Tamura for help with establishment of TMEM41B-KO and VMP1-KO cells.

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: Y.H., T.P.L., N.M.; Methodology: T.P.L.; Formal analysis: F.O., Y.H., S.Z.; Investigation: F.O., Y.H., S.Z., T.P.L.; Data curation: F.O., Y.H., S.Z.; Writing - original draft: F.O., Y.H., S.Z., N.M.; Supervision: H.M., H.Y., N.M.; Funding acquisition: N.M.

Funding

This work was supported by a grant for Exploratory Research for Advanced Technology from the Japan Science and Technology Agency (JPMJER1702 to N.M.).

Peer review history

The peer review history is available online at <https://journals.biologists.com/jcs/article-lookup/doi/10.1242/jcs.255877>

References

Bogdanov, M., Zhang, W., Xie, J. and Dowhan, W. (2005). Transmembrane protein topology mapping by the substituted cysteine accessibility method (SCAMTM): application to lipid-specific membrane protein topogenesis. *Methods* **36**, 148-171. doi:10.1016/j.yemeth.2004.11.002

Boughner, L. A. and Doerrler, W. T. (2012). Multiple deletions reveal the essentiality of the DedA membrane protein family in *Escherichia coli*. *Microbiology* **158**, 1162-1171. doi:10.1099/mic.0.056325-0

Calvo-Garrido, J., Carilla-Latorre, S. and Escalante, R. (2008). Vacuole membrane protein 1, autophagy and much more. *Autophagy* **4**, 835-837. doi:10.4161/auto.6574

Chang, H.-Y., Chou, C.-C., Hsu, M.-F. and Wang, A. H. J. (2014). Proposed carrier lipid-binding site of undecaprenyl pyrophosphate phosphatase from *Escherichia coli*. *J. Biol. Chem.* **289**, 18719-18735. doi:10.1074/jbc.M114.575076

Cheong, H. and Klionsky, D. J. (2008). Biochemical methods to monitor autophagy-related processes in yeast. *Methods Enzymol.* **451**, 1-26. doi:10.1016/S0076-6879(08)03201-1

Davis, D., Suemitsu, J. and Wattenberg, B. (2019). Transmembrane topology of mammalian ORMDL proteins in the endoplasmic reticulum as revealed by the substituted cysteine accessibility method (SCAMTM). *Biochim. Biophys. Acta (BBA) Proteins Proteom.* **1867**, 382-395. doi:10.1016/j.bbapap.2019.01.005

Demignot, S., Beilstein, F. and Morel, E. (2014). Triglyceride-rich lipoproteins and cytosolic lipid droplets in enterocytes: key players in intestinal physiology and metabolic disorders. *Biochimie* **96**, 48-55. doi:10.1016/j.biochi.2013.07.009

Doerrler, W. T., Sikdar, R., Kumar, S. and Boughner, L. A. (2013). New functions for the ancient DedA membrane protein family. *J. Bacteriol.* **195**, 3-11. doi:10.1128/JB.01006-12

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792-1797. doi:10.1093/nar/gkh340

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A. et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427-D432. doi:10.1093/nar/gky995

El Ghachi, M., Howe, N., Huang, C. Y., Olieric, V., Warshamanage, R., Touze, T., Weichert, D., Stansfeld, P. J., Wang, M., Kerff, F. et al. (2018). Crystal structure of undecaprenyl-pyrophosphate phosphatase and its role in peptidoglycan biosynthesis. *Nat. Commun.* **9**, 1078. doi:10.1038/s41467-018-03477-5

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* **17**, 368-376. doi:10.1007/BF01734359

Forrest, L. R. (2015). Structural symmetry in membrane proteins. *Annu. Rev. Biophys.* **44**, 311-337. doi:10.1146/annurev-biophys-051013-023008

Gietz, R. D. and Woods, R. A. (2002). Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.* **350**, 87-96. doi:10.1016/S0076-6879(02)50957-5

Hoffmann, H.-H., Schneider, W. M., Rozen-Gagnon, K., Miles, L. A., Schuster, F., Razoooky, B., Jacobson, E., Wu, X., Yi, S., Rudin, C. M. et al. (2021). TMEM41B is a Pan-flavivirus Host Factor. *Cell* **184**, 133-148.e20. doi:10.1016/j.cell.2020.12.005

Hopf, T. A., Green, A. G., Schubert, B., Mersmann, S., Schärfe, C. P. I., Ingraham, J. B., Toth-Petroczy, A., Brock, K., Riesselman, A. J., Palmedo, P. et al. (2019). The EVcouplings Python framework for coevolutionary sequence analysis. *Bioinformatics* **35**, 1582-1584. doi:10.1093/bioinformatics/bty862

Hosokawa, N., Hara, Y. and Mizushima, N. (2006). Generation of cell lines with tetracycline-regulated autophagy and a role for autophagy in controlling cell size. *FEBS Lett.* **580**, 2623-2629. doi:10.1016/j.febslet.2006.04.008

Imachi, H., Nobu, M. K., Nakahara, N., Morono, Y., Ogawara, M., Takaki, Y., Takano, Y., Uematsu, K., Ikuta, T., Ito, M. et al. (2020). Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature* **577**, 519-525. doi:10.1038/s41586-019-1916-6

Inadome, H., Noda, Y., Kamimura, Y., Adachi, H. and Yoda, K. (2007). Tvp38, Tvp23, Tvp18 and Tvp15: novel membrane proteins in the Tlg2-containing Golgi/endosome compartments of *Saccharomyces cerevisiae*. *Exp. Cell Res.* **313**, 688-697. doi:10.1016/j.yexcr.2006.11.008

Johnson, Z. L., Cheong, C.-G. and Lee, S.-Y. (2012). Crystal structure of a concentrative nucleoside transporter from *Vibrio cholerae* at 2.4 Å. *Nature* **483**, 489-493. doi:10.1038/nature10882

Johnson, Z. L., Lee, J.-H., Lee, K., Lee, M., Kwon, D.-Y., Hong, J. and Lee, S.-Y. (2014). Structural basis of nucleoside and nucleoside drug selectivity by concentrative nucleoside transporters. *eLife* **3**, e03604. doi:10.7554/eLife.03604

Kaizuka, T., Morishita, H., Hama, Y., Tsukamoto, S., Matsui, T., Toyota, Y., Kodama, A., Ishihara, T., Mizushima, T. and Mizushima, N. (2016). An autophagic flux probe that releases an internal control. *Mol. Cell* **64**, 835-849. doi:10.1016/j.molcel.2016.09.037

Kanai, Y., Cléménçon, B., Simonin, A., Leuenberger, M., Lochner, M., Weisstanner, M. and Hediger, M. A. (2013). The SLC1 high-affinity glutamate and neutral amino acid transporter family. *Mol. Aspects Med.* **34**, 108-120. doi:10.1016/j.mam.2013.01.001

Kang, Z. B., Moutsatsos, I., Moretti, F., Bergman, P., Zhang, X., Nyfeler, B. and Antczak, C. (2020). Fluopack screening platform for unbiased cellular phenotype profiling. *Sci. Rep.* **10**, 2097. doi:10.1038/s41598-020-58861-3

Keller, R. and Schneider, D. (2013). Homologs of the yeast Tvp38 vesicle-associated protein are conserved in chloroplasts and cyanobacteria. *Front. Plant Sci.* **4**, 467. doi:10.3389/fpls.2013.00467

Keller, R., Ziegler, C. and Schneider, D. (2014). When two turn into one: evolution of membrane transporters from half modules. *Biol. Chem.* **395**, 1379-1388. doi:10.1515/hsz-2014-0224

- Keller, R., Schleppei, N., Weikum, J. and Schneider, D. (2015). Mutational analyses of YqjA, a Tvp38/DedA protein of *E. coli*. *FEBS Lett.* **589**, 842-848. doi:10.1016/j.febslet.2015.02.019
- Khafizov, K., Staritzbichler, R., Stamm, M. and Forrest, L. R. (2010). A study of the evolution of inverted-topology repeats from LeuT-fold transporters using AlignMe. *Biochemistry* **49**, 10702-10713. doi:10.1021/bi101256x
- Klionsky, D. J., Abdelmohsen, K., Abe, A., Abedin, M. J., Abeliovich, H., Acevedo Arozena, A., Adachi, H., Adams, C. M., Adams, P. D., Adeli, K. et al. (2016). Guidelines for the use and interpretation of assays for monitoring autophagy (3rd edition). *Autophagy* **12**, 1-222. doi:10.1080/15548627.2015.1100356
- Kozma, D., Simon, I. and Tusnády, G. E. (2013). PDBTM: protein data bank of transmembrane proteins after 8 years. *Nucleic Acids Res.* **41**, D524-D529. doi:10.1093/nar/gks1169
- Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567-80. doi:10.1006/jmbi.2000.4315
- Kumar, S. and Doerfler, W. T. (2014). Members of the conserved DedA family are likely membrane transporters and are required for drug resistance in *Escherichia coli*. *Antimicrob. Agents Chemother.* **58**, 923-930. doi:10.1128/AAC.02238-13
- Kumar, S. and Doerfler, W. T. (2015). *Escherichia coli* YqjA, a member of the conserved DedA/Tvp38 membrane protein family, is a putative osmosensing transporter required for growth at alkaline pH. *J. Bacteriol.* **197**, 2292-2300. doi:10.1128/JB.00175-15
- Kumar, S., Bradley, C. L., Mukashyaka, P. and Doerfler, W. T. (2016). Identification of essential arginine residues of *Escherichia coli* DedA/Tvp38 family membrane proteins YqjA and YghB. *FEMS Microbiol. Lett.* **363**, fnw133. doi:10.1093/femsle/fnw133
- Ledgham, F., Quest, B., Vallaes, T., Mergeay, M. and Covès, J. (2005). A probable link between the DedA protein and resistance to selenite. *Res. Microbiol.* **156**, 367-374. doi:10.1016/j.resmic.2004.11.003
- Mancusso, R., Gregorio, G. G., Liu, Q. and Wang, D.-N. (2012). Structure and mechanism of a bacterial sodium-dependent dicarboxylate transporter. *Nature* **491**, 622-626. doi:10.1038/nature11542
- Matsui, M. and Iwasaki, W. (2020). Graph splitting: a graph-based approach for superfamily-scale phylogenetic tree reconstruction. *Syst. Biol.* **69**, 265-279. doi:10.1093/sysbio/sy049
- Mesdaghi, S., Murphy, D. L., Sánchez Rodríguez, F., Burgos-Mármol, J. J. and Rigden, D. J. (2020). In silico prediction of structure and function for a large family of transmembrane proteins that includes human Tmem41b. *F1000Research* **9**, 1395. doi:10.12688/f1000research.27676.1
- Moller, S., Croning, M. D. R. and Apweiler, R. (2001). Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **17**, 646-653. doi:10.1093/bioinformatics/17.7.646
- Moretti, F., Bergman, P., Dodgson, S., Marcellin, D., Claerr, I., Goodwin, J. M., DeJesus, R., Kang, Z., Antczak, C., Begue, D. et al. (2018). TMEM41B is a novel regulator of autophagy and lipid mobilization. *EMBO Rep.* **19**, e45889. doi:10.15252/embr.201845889
- Morishita, H., Zhao, Y. G., Tamura, N., Nishimura, T., Kanda, Y., Sakamaki, Y., Okazaki, M., Li, D. and Mizushima, N. (2019). A critical role of VMP1 in lipoprotein secretion. *eLife* **8**, e48834. doi:10.7554/eLife.48834
- Morita, K., Hama, Y., Izume, T., Tamura, N., Ueno, T., Yamashita, Y., Sakamaki, Y., Mimura, K., Morishita, H., Shihoya, W. et al. (2018). Genome-wide CRISPR screen identifies TMEM41B as a gene required for autophagosome formation. *J. Cell Biol.* **217**, 3817-3828. doi:10.1083/jcb.201804132
- Morita, K., Hama, Y. and Mizushima, N. (2019). TMEM41B functions with VMP1 in autophagosome formation. *Autophagy* **15**, 922-923. doi:10.1080/15548627.2019.1582952
- Nakatogawa, H. (2020). Mechanisms governing autophagosome biogenesis. *Nat. Rev. Mol. Cell Biol.* **21**, 439-458. doi:10.1038/s41580-020-0241-0
- Nonet, M. L., Marvel, C. C. and Tolan, D. R. (1987). The hisT-purF region of the *Escherichia coli* K-12 chromosome. Identification of additional genes of the hisT and purF operons. *J. Biol. Chem.* **262**, 12209-12217. doi:10.1016/S0021-9258(18)45338-0
- Ovchinnikov, S., Kamisetty, H. and Baker, D. (2014). Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *eLife* **3**, e02030. doi:10.7554/eLife.02030
- Panta, P. R. and Doerfler, W. T. (2021). A *Burkholderia thailandensis* DedA family membrane protein is required for proton motive force dependent lipid A modification. *Fron. Microbiol.* **11**, 618389. doi:10.3389/fmicb.2020.618389
- Panta, P. R., Kumar, S., Stafford, C. F., Billiot, C. E., Douglass, M. V., Herrera, C. M., Trent, M. S. and Doerfler, W. T. (2019). A DedA family membrane protein is required for *Burkholderia thailandensis* colistin resistance. *Front. Microbiol.* **10**, 2532. doi:10.3389/fmicb.2019.02532
- Rannala, B. and Yang, Z. (1996). Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *J. Mol. Evol.* **43**, 304-311. doi:10.1007/BF02338839
- Romero-Brey, I. and Bartenschlager, R. (2016). Endoplasmic reticulum: the favorite intracellular niche for viral replication and assembly. *Viruses* **8**, 160. doi:10.3390/v8060160
- Ropolo, A., Grasso, D., Pardo, R., Sacchetti, M. L., Archange, C., Lo Re, A., Seux, M., Nowak, J., Gonzalez, C. D., Iovanna, J. L. et al. (2007). The pancreatitis-induced vacuole membrane protein 1 triggers autophagy in mammalian cells. *J. Biol. Chem.* **282**, 37124-37133. doi:10.1074/jbc.M706956200
- Saitoh, T., Nakano, H., Yamamoto, N. and Yamaoka, S. (2002). Lymphotoxin-beta receptor mediates NEMO-independent NF-kappaB activation. *FEBS Lett.* **532**, 45-51. doi:10.1016/S0014-5793(02)03622-0
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B. et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676-682. doi:10.1038/nmeth.2019
- Schneider, W. M., Luna, J. M., Hoffmann, H.-H., Sánchez-Rivera, F. J., Leal, A. A., Ashbrook, A. W., Le Pen, J., Ricardo-Lax, I., Michailidis, E., Peace, A. et al. (2021). Genome-scale identification of SARS-CoV-2 and Pan-coronavirus host factor networks. *Cell* **184**, 120-132.e14. doi:10.1016/j.cell.2020.12.006
- Scrapanti, E. and Hunte, C. (2007). Discontinuous membrane helices in transport proteins and their correlation with function. *J. Struct. Biol.* **159**, 261-267. doi:10.1016/j.jsb.2007.01.011
- Shoemaker, C. J., Huang, T. Q., Weir, N. R., Polyakov, N. J., Schultz, S. W. and Denic, V. (2019). CRISPR screening using an expanded toolkit of autophagy reporters identifies TMEM41B as a novel autophagy factor. *PLoS Biol.* **17**, e2007044. doi:10.1371/journal.pbio.2007044
- Sikdar, R. and Doerfler, W. T. (2010). Inefficient Tat-dependent export of periplasmic amidases in an *Escherichia coli* strain with mutations in two DedA family genes. *J. Bacteriol.* **192**, 807-818. doi:10.1128/JB.00716-09
- Steinegger, M., Meier, M., Mirdita, M., Vöhringer, H., Haunsberger, S. J. and Söding, J. (2019). HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* **20**, 473. doi:10.1186/s12859-019-3019-7
- Tábara, L.-C. and Escalante, R. (2016). VMP1 Establishes ER-microdomains that regulate membrane contact sites and autophagy. *PLoS ONE* **11**, e0166499. doi:10.1371/journal.pone.0166499
- Tábara, L.-C., Vincent, O. and Escalante, R. (2019). Evidence for an evolutionary relationship between Vmp1 and bacterial DedA proteins. *Int. J. Dev. Biol.* **63**, 67-71. doi:10.1387/ijdb.180312re
- Tenenboim, H., Smirnova, J., Willmitzer, L., Steup, M. and Brotman, Y. (2014). VMP1-deficient *Chlamydomonas* exhibits severely aberrant cell morphology and disrupted cytokinesis. *BMC Plant Biol.* **14**, 121. doi:10.1186/1471-2229-14-121
- Thompkins, K., Chattopadhyay, B., Xiao, Y., Henk, M. C. and Doerfler, W. T. (2008). Temperature sensitivity and cell division defects in an *Escherichia coli* strain with mutations in yghB and yqjA, encoding related and conserved inner membrane proteins. *J. Bacteriol.* **190**, 4489-4500. doi:10.1128/JB.00414-08
- Törnroth-Horsefield, S., Hedfalk, K., Fischer, G., Lindkvist-Petersson, K. and Neutze, R. (2010). Structural insights into eukaryotic aquaporin regulation. *FEBS Lett.* **584**, 2580-2588. doi:10.1016/j.febslet.2010.04.037
- Van Alstyne, M., Lotti, F., Dal Mas, A., Area-Gomez, E. and Pellizzoni, L. (2018). Stasimon/Tmem41b localizes to mitochondria-associated ER membranes and is essential for mouse embryonic development. *Biochem. Biophys. Res. Commun.* **506**, 463-470. doi:10.1016/j.bbrc.2018.10.073
- Walther, T. C., Chung, J. and Farese, R. V. Jr. (2017). Lipid droplet biogenesis. *Annu. Rev. Cell Dev. Biol.* **33**, 491-510. doi:10.1146/annurev-cellbio-100616-060608
- Workman, S. D., Worrall, L. J. and Strynadka, N. C. J. (2018). Crystal structure of an intramembranal phosphatase central to bacterial cell-wall peptidoglycan biosynthesis and lipid recycling. *Nat. Commun.* **9**, 1159. doi:10.1038/s41467-018-03547-8
- Yan, C. and Luo, J. (2010). An analysis of reentrant loops. *Protein J.* **29**, 350-354. doi:10.1007/s10930-010-9259-z
- Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S. and Baker, D. (2020). Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl. Acad. Sci. USA* **117**, 1496-1503. doi:10.1073/pnas.1914677117
- Zhao, Y. G., Chen, Y., Miao, G., Zhao, H., Qu, W., Li, D., Wang, Z., Liu, N., Li, L., Chen, S. et al. (2017). The ER-localized transmembrane protein EPG-3/VMP1 regulates SERCA activity to control ER-isolation membrane contacts for autophagosome formation. *Mol. Cell* **67**, 974-989.e6. doi:10.1016/j.molcel.2017.08.005
- Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A. N. and Alva, V. (2018). A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol.* **430**, 2237-2243. doi:10.1016/j.jmb.2017.12.007