EMC is required for biogenesis and membrane insertion of Xport-A, an essential chaperone of rhodopsin-1 and the TRP channel

Catarina J. Gaspar^{1,2}, Lígia C. Vieira^{1#}, John C. Christianson³, David Jakubec⁴, Kvido Strisovsky⁴, Colin Adrain^{*2,5}, Pedro M. Domingos^{*1}

1 - Instituto de Tecnologia Química e Biológica da Universidade Nova de Lisboa (ITQB-NOVA), Oeiras, Portugal. 2 - Membrane Traffic Lab, Instituto Gulbenkian de Ciência (IGC), Oeiras, Portugal. 3- Nuffield Dept. of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Botnar Research Centre, University of Oxford, UK. 4- Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, Prague, Czech Republic. 5 – Patrick G Johnston Centre for Cancer Research, Queen's University, Belfast, N. Ireland. # present address – Center for Genomics and Systems Biology, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates. *co-senior and corresponding authors (domingop@itqb.unl.pt or c.adrain@qub.ac.uk)

SUMMARY

Insertion of hydrophobic transmembrane domains (TMDs) into the endoplasmic reticulum (ER) lipid bilayer is an essential step during eukaryotic membrane protein biogenesis. The ER membrane complex (EMC) functions as an insertase for TMDs of low hydrophobicity and is required for the biogenesis of a subset of tail-anchored (TA) and polytopic membrane proteins, including rhodopsin-1 (Rh1) and the TRP channel. To better understand the physiological implications of membrane protein biogenesis dependent on the EMC, we performed a bioinformatic analysis to predict TA proteins present in the *Drosophila* proteome. From 254 predicted TA proteins, subsequent genetic screening in *Drosophila* larval eye discs led to the identification of 2 proteins that require EMC for their biogenesis: farinelli (fan) and Xport-A. Fan is required for sperm individualization and male fertility in *Drosophila* and we now show that EMC is also required for these important biological processes. Interestingly, Xport-A is essential for the biogenesis of both Rh1 and TRP, raising the possibility that disruption of Rh1 and TRP biogenesis in EMC loss of function mutations is secondary to the Xport-A defect. We show that EMC is required for Xport-A TMD membrane insertion and increasing the hydrophobicity of Xport-A TMD rendered its membrane insertion to become EMC-independent. Moreover, these EMC-independent Xport-A mutants rescued Rh1 and TRP biogenesis in EMC mutants. Our data establish that EMC can impact the biogenesis of polytopic membrane proteins indirectly, by controlling the biogenesis and membrane insertion of an essential protein co-factor.

INTRODUCTION

Membrane proteins comprise approximately 30% of the eukaryotic proteome and confer many essential functions to biological membranes ^{1,2}. These proteins contain hydrophobic transmembrane domains (TMDs) that must be inserted into the membrane of the endoplasmic reticulum (ER) through evolutionarily conserved molecular machineries. The vast majority of membrane proteins are inserted into the ER membrane through the co-translational pathway, in which proteins containing signal peptides or TMDs are recognized by the SRP (signal recognition particle) as they emerge from the ribosomes. Eventually, the ribosome-nascent chain complex is delivered to the Sec61 translocon, where membrane insertion takes place ³.

The co-translational insertion pathway is the predominant mechanism for membrane protein insertion, but it is unable to deal with the biogenesis of a specific class of membrane proteins called tail anchored (TA) proteins, which lack a signal peptide and contain a single TMD at their C-terminus ⁴.

Consequently, their TMD remains shielded inside the ribosomal tunnel until the termination codon is reached, and TMD recognition can only occur post-translationally ⁵. Accordingly, TA proteins utilize a dedicated conserved TMD recognition complex (TRC) pathway, which facilitates the targeted release of these proteins into the ER in eukaryotic cells, in a post-translational manner ⁶. The crucial component of this pathway is the ATPase TRC40 (or in yeast Get3), which captures and shields TMDs of TA proteins ^{6–8}, until they are released to a receptor complex composed by WRB-CAML (yeast Get1-Get2), which mediates TMD insertion into the ER membrane ^{9,10}. Structural and biochemical studies have shown that this pathway displays a preference for TMDs of high hydrophobicity ^{11–13}. Recently, the ER membrane protein complex (EMC) was identified as an insertase for TA proteins with TMDs of moderate to low hydrophobicity ^{13–15}. The EMC is a highly conserved, multi-subunit protein complex, with 9 subunits in mammals ¹⁶. The EMC was initially described in yeast in a high throughput genetic screen for genes required for protein folding ¹⁷ and

its mammalian counterpart was later identified as part of the interaction network of the ER-associated protein degradation (ERAD) machinery¹⁸.

The EMC was subsequently shown to serve as an insertase for specific polytopic membrane proteins that contain a signal anchor sequence (SAS), including a subset of G-protein coupled receptors (GPCR)¹⁹. A variety of experiments suggested that the EMC coordinates the insertion of the first TMD of the GPCRs, after which subsequent TMD insertions are EMC-independent and require the Sec61 translocon, via a “handing-off” mechanism that remains to be fully elucidated¹⁹.

The structure of both the human and yeast EMC have recently been determined using cryo-electron microscopy (cryo-EM)^{20–23} and a model has been proposed for EMC-mediated co- and post-translational insertion of client proteins²⁰. According to this model, a captured client protein is directed towards the membrane by the flexible cytosolic loop of EMC3, the insertase subunit of EMC; the EMC then presumably reduces the energetic cost of insertion by inducing a local thinning of the membrane and by arranging polar and positively charged residues within the bilayer. The client would then dissociate from EMC3, encountering EMC1’s β propellers, which could act as a scaffold for co-factor recruitment^{20,22}.

The requirement for the EMC in the biogenesis of some TA proteins and the first TMD of some polytopic membrane proteins containing a SAS has been dissected *in vitro*, leading to a rationalization of how EMC can act mechanistically. However, the EMC has also been shown to be required for the biogenesis of multi-pass membrane proteins enriched with “challenging” TMDs²⁴. Indeed, the majority of proteins found to date to be affected by loss of EMC are neither TA nor SAS-containing membrane proteins, suggesting that much remains to be understood about EMC specificity for its client proteins^{15,24}. These include the ABC transporter Yor1 in yeast^{25,26}, acetylcholine receptor in *C. elegans*²⁷, rhodopsins^{28–31}, the transient receptor potential (TRP) channel in *Drosophila*²⁹, ABCA1 in mice³² and mutant connexin32³³. Furthermore, EMC loss of function has also been associated with defects in phospholipid trafficking^{34,35}, cholesterol homeostasis¹⁴, autophagosome formation^{36,37}, viral pathogenesis^{38–42}, neurological defects⁴³ and male fertility⁴⁴. Although this diversity in phenotypes is still an area of investigation, several of these examples impact unrelated membrane proteins with multiple TMDs^{45,46} and many of these candidate EMC clients are neither TA nor SAS-containing proteins. An alternative possibility to EMC acting as an insertase for a broader range of TMD protein topologies is that some of these proteins may be “indirect” or “secondary clients” of

the EMC, for example, proteins whose biogenesis or stability depends on a direct EMC client protein.

In this study, we interrogated the *Drosophila* proteome in an effort to bioinformatically identify TA proteins that could have a dependency on the EMC for their biogenesis. Using a *Drosophila* larval eye imaginal disc assay, we identified two EMC clients: fan (farinelli), which controls sperm individualization⁴⁷ and Xport-A (exit protein of rhodopsin and TRP-A)^{48,49}, an essential chaperone for the biogenesis of both Rhodopsin-1 (Rh1) and the TRP (Transient Receptor Potential) channel, and their targeting to the rhabdomere, the light sensitive compartment of the photoreceptors. Interestingly, the biogenesis of Rh1 and TRP has been shown to be deficient in EMC mutant clones^{29–31}. We generated a mutant of Xport-A (Xport-A4L) whose biogenesis proceeds independently of the EMC. Xport-A4L is able to rescue the expression of Rh1 and TRP in EMC mutant tissue, suggesting that the latter proteins are not direct clients of the EMC but rather, depend on EMC indirectly via Xport-A. Overall, our results suggest that EMC is required for sperm and photoreceptor differentiation in *Drosophila*, due to EMC’s role in the biogenesis of fan and Xport-A, respectively. Crucially, our data establish that EMC impacts the biogenesis of some multi-pass membrane proteins indirectly, by governing insertion of the cofactors they require for assembly and deployment. This paradigm expands the potential governance of the EMC to a greater portion of the membrane proteome.

RESULTS

EMC is required for the biogenesis of a subset of tail-anchored proteins

The EMC has been shown to function as an insertase for TA proteins with TMDs of moderate to low hydrophobicity¹³, but the identities of its clients have not been fully determined. To ascertain the range of EMC-dependent TA proteins, we began by screening the *Drosophila melanogaster* proteome for all possible TA proteins. Tail anchored proteins are defined by their cytosolic N-terminal domain that is anchored to the lipid bilayer by a single hydrophobic TMD proximal to the C-terminus⁵⁰. We bioinformatically interrogated the proteome using prediction algorithms for: signal peptide, TMD and topology⁵¹ as well as subcellular localisation⁵² (Figure 1A). This analysis yielded a total of 254 candidate TA proteins in the *Drosophila* proteome, with predicted membrane localizations (Figure 1A, S1-A, supplementary file1). We subsequently characterized the distribution of different biophysical features such as TMD length, TMD hydrophobicity, tail length and charge (Figure S1-C-F). These analyses show that within the list of predicted TA proteins, those that are ER-localized

Figure 1. EMC is required for the biogenesis of a subset of tail-anchored proteins. (A) The *Drosophila* proteome was screened to identify TA proteins using the PolyPhobius⁵¹ algorithm within the TOPCONS⁶¹ web server. Proteins with a single TMD within 35 AAs or less upstream of the C terminus and lacking a signal peptide or mitochondrial targeting peptide were selected for further interrogation. The DeepLoc⁵² predictor of protein sub-cellular localization was used to eliminate soluble proteins from the list. These procedures yielded 254 proteins, which were scored for their tail charge and hydrophobicity using the “transmembrane tendency” score of Zhao & London⁵⁵. (B) A subset of the predicted TA proteins was screened by overexpression in *Drosophila* larval eye imaginal discs containing clones of a previously isolated EMC3 mutant allele, EMC3^{ΔΔ}²⁹. The ratio of fluorescence intensity for the tested TA proteins in EMC3 homozygous mutant cells over WT cells (EMC3^{ΔΔ}/WT) was measured and plotted. Proteins were classified into four groups according to the ratio measured: normal (ratio 0.8-1.2; bars in blue), increased (ratio>1.2; bars in dark blue), reduced (ratio 0.8-0.5; bar in pink) and strongly reduced (ratio<0.5; bars in red). Error bars correspond to standard deviation (stdev). (C-J) All panels show *Drosophila* 3rd instar larval eye imaginal discs with eyeless-Flippase-induced clones of cells homozygous for EMC3^{ΔΔ}²⁹, labelled by the absence of ubiGFP (green). The red channel shows ELAV, which labels the nuclei of photoreceptor cells. All UAS constructs were expressed under the control of GMR-GAL4. (C) Expression of UAS-Rh1 (4C5, in blue) and (D) Na⁺K⁺ATPase (A5-C, in blue) is strongly reduced in EMC3^{ΔΔ} homozygous mutant cells. (E) Expression of UAS-Xport-A-HA (anti-HA, in blue) and (F) UAS-fan-HA (anti-HA, in blue) is strongly reduced in EMC3^{ΔΔ} homozygous mutant cells. (G) Expression of UAS-rtv-HA (anti-HA, in blue) and (H) hid (anti-hid, in blue) is increased in EMC3^{ΔΔ} homozygous mutant cells. (I-J) Normal expression of two TA proteins in EMC3^{ΔΔ} homozygous mutant cells. (I) UAS-CG8814-HA (anti-HA, in blue) is a TA protein containing a TMD of low hydrophobicity (J) UAS-Syx13-HA is a TA protein with high hydrophobicity TMD. Scale bars represent 10 μm.

tend to have higher TMD hydrophobicity and length than their mitochondrial and peroxisomal counterparts (Figure S1-C-F), which is consistent with published predictions for both human⁵³ and *Arabidopsis*⁵⁴ proteomes.

The TMDs of most predicted TA proteins (~67%) exhibited low hydrophobicity¹³ (hydrophobicity value below 22 in the Zhao & London hydrophobicity scale⁵⁵). This ratio was maintained in most sub-cellular localizations including the ER but with the exception of the Golgi membrane (75% proteins with high hydrophobicity) and peroxisome and mitochondrion membrane (100% of proteins with low hydrophobicity).

Next, we selected 23 candidate TA proteins to test for EMC dependency by monitoring their expression in EMC3 mutant clones in *Drosophila* larval eye imaginal discs (Figure 1B, 1C-J). This model enables candidate TA protein expression levels in WT and adjacent EMC3 mutant territories to be compared side-by-side, within the same tissue. Of the 23 proteins tested, approximately 87% had low predicted TMD hydrophobicity, were from within all predicted sub-cellular membrane localizations and had similar distribution of localization to the overall pool of predicted TA proteins (Figure S1-A-B). Candidate TA proteins were overexpressed under the control of GMR-GAL4 in larval eye imaginal discs containing EMC3 mutant (EMC3^{Δ4} 29) clones. EMC3 is considered a “core” subunit of the complex, with mutations in EMC3 and other core subunits (EMC1, 2, 5 and 6) causing significant impairment of EMC functionality^{13,14,46}. Candidate TA protein expression was assayed by immunofluorescence using antibodies raised against the respective proteins or, when no suitable antibodies existed, against an appended HA (Hemagglutinin) tag. Rh1 and the Na⁺K⁺ATPase α subunit were also evaluated (Figure 1C, D), as their expression has been reported previously to be EMC dependent^{29–31}.

We evaluated TA candidate expression by determining the ratio of fluorescence intensity between the signal associated with the candidate TA protein in EMC3 mutant clones versus non-mutant (WT) clones (EMC3^{Δ4}/WT) (Figure 1B). TA candidate expression was classified into four groups according to the EMC3^{Δ4}/WT fluorescence intensity ratio: normal (ratio 0.8-1.2, bars in blue), increased (ratio>1.2, bars in dark blue), reduced (ratio 0.8-0.5, bar in pink) and strongly reduced (ratio<0.5, bars in red). Rh1 (Figure 1C) and the Na⁺K⁺ATPase α subunit (Figure 1D) exhibited strongly reduced expression in EMC3^{Δ4} mutant clones, consistent with previously published results in the adult/pupal eye^{29–31}. Expression of Dpck was reduced in EMC3^{Δ4} mutant clones while rrv (Figure 1G), hid (Figure 1H) and PIG-X showed

increased expression. Notably, two TA candidates, Xport-A (Figure 1E) and fan (Figure 1F), showed strongly reduced expression in EMC3^{Δ4} mutant clones compared to WT. Both proteins are predicted to localize to the ER membrane and contain TMDs of low hydrophobicity.

Of the 23 TA proteins tested, 17 exhibited normal ratios of expression, demonstrating that the GMR-GAL4 driven transcription of the candidate proteins is not affected by the loss of EMC function. Examples of TA proteins with TMDs of low hydrophobicity (CG8814 – Figure 1I) and high hydrophobicity (Syx13 – Figure 1J) showing normal EMC3^{Δ4}/WT fluorescence intensity ratios, could also be found.

EMC is required for sperm differentiation in *Drosophila*

Our screen identified the predicted TA protein fan, whose expression was defective in EMC3^{Δ4} homozygous mutant clones (Figure 1-F). As fan reportedly plays an important role in sperm individualization and male fertility⁴⁷, we asked whether the EMC was also required for male fertility in *Drosophila*. We selected RNAi lines targeting different EMC subunits (EMC1, EMC3, EMC4, EMC5, EMC6, EMC7, EMC8/9) and mated adult males in which individual EMC subunits were knocked down, with wild-type virgin females (Figure S2A-B). We counted the mean number of progeny obtained from these crosses and observed a statistically significant reduction for all RNAi lines of EMC5 and EMC6, and for one RNAi line of EMC1 (Figure S2-A). We analysed sperm vesicle size and observed that, in accordance with the reduced male fertility, all EMC5 and EMC6 RNAi lines tested exhibited a statistically significant reduction in sperm vesicles (Figure S2-C), which correlates with a reduction in sperm production^{47,56}. Furthermore, the seminal vesicles of flies expressing EMC5_2 RNAi line appeared devoid of mature sperm, as the needle shaped nuclei of mature sperm were absent (Figure S2-D). Finally, we tried to address the cause of male fertility in an EMC5 RNAi line, observing that while staining with an antibody against active Caspase-3 was present in multiple cystic bulges (cb) in wild-type testis, this staining was deficient in EMC5 RNAi testis (Figure S2-E), demonstrating sperm differentiation defects, as previously shown⁵⁷.

EMC3 is required for the expression of Xport-A, but not Xport-A4L

Although the specificity of EMC for its clients remains to be fully delineated, the EMC exhibits a preference for TA proteins with low TMD hydrophobicity and/or containing polar/charged amino acid residues^{13–15}. As the Xport-A TMD has a reduced hydrophobicity due to the presence of one polar and multiple charged residues, an Xport-

Figure 2. EMC3 is required for the expression of Xport-A, but not Xport-A4L. (A) The TMD of Xport-A was identified using TOPCONS⁵¹ and is in bold/underlined, together with the immediate flanking residues. Red residues indicate mutations made in the Xport-A TMD to increase its hydrophobicity, creating the Xport-A4L mutant. To the right of each sequence is the “transmembrane tendency” score, calculated according to Zhao & London⁵⁵. (B-D) – Immunostaining of 3rd instar larval eye imaginal discs with eyeless-Flippase-induced clones of cells homozygous for EMC3^{Δ4}, labelled by the absence of ubiGFP (green) shows reduced expression of (B) UAS-HA-Xport-A (anti-HA, in blue) but normal expression of (C) UAS-HA-Xport-A4L (anti-HA, in blue) or (D) UAS-Xport-A4L-HA (anti-HA, in blue) in EMC3 mutant homozygous cells. ELAV (red in B and C) marks photoreceptor cells and Na⁺K⁺ATPase (red in D) acts as a positive control for the presence of EMC3^{Δ4} homozygous mutant clones. The UAS constructs were expressed under the control of GMR-GAL4. (E-F) Immunostaining of mosaic adult retina with eyeless-Flippase-induced clones of cells homozygous for EMC3^{e02662}, labelled by the absence of RFP (red), show loss of (E) UAS-HA-Xport-A (anti-HA, in blue) in EMC3 mutant homozygous cells. (F) Expression of UAS-HA-Xport-A4L (anti-HA, in blue) is normal in EMC3^{e02662} homozygous mutant clones. The rhabdomeres are stained for actin (Phalloidin, in green). Scale bars represent 10 μm. (G) The ratio of fluorescence intensity in EMC3 homozygous mutant cells over that of WT cells (EMC3^{mut}/WT) was measured and plotted. Proteins were classified into two groups according to the ratio measured: normal (ratio 0.8-1.2; bars in blue) and strongly reduced (ratio < 0.5; bars in red). Error bars correspond to standard deviation (stdev). Significance was determined by Welch's t-test: ****p ≤ 0.0001.

A TMD with increased hydrophobicity would be expected to relieve its EMC dependency. To that end, we performed site-directed mutagenesis on the Xport-A TMD to substitute the four most hydrophilic amino acids with leucine, creating the mutant Xport-A4L (Figure 2A). As hypothesized, whereas WT HA-Xport-A was dependent on the EMC (Figure 2B), expression of HA-Xport-A4L (Figure 2C) or Xport-A4L-HA (Figure 2D) no longer required EMC3 in larval eye imaginal discs. Importantly, the levels of the Na⁺K⁺ATPase control protein remained defective in EMC3 mutant clones (Figure 2D), confirming the presence of “bona fide” EMC3 mutant clones in the eye disc. A similar impact of the Xport-A4L mutant was observed in experiments carried out in adult retinas (Figure 2 E,F). Altogether, these results indicate that increasing the hydrophobicity of the Xport-A TMD renders its expression independent of the EMC and that tagging Xport-A and Xport-A4L at either the N- or C-terminus yields identical results.

EMC is required for the biogenesis of Xport-A in mammalian cells

We next sought to confirm the dependence of Xport-A biogenesis on the EMC using an assay independent of *Drosophila*. We cloned the TMDs of Xport-A, Xport-A4L and the validated EMC client squalene synthase (SQS)^{13,14}, individually into the dual fluorescent reporter GFP-P2A-RFP (Figure 3A). When translated in mammalian cells, this mRNA will generate two different products due to peptide bond skipping induced by the P2A site: GFP and RFP-SQS, RFP-Xport-A or RFP-Xport-A4L. A stable GFP serves as a reporter for translation of the construct that enables quantitative read-out for protein stability through fluorescence ratios¹⁹. When the RFP:GFP ratio is reduced, it reflects post-translational degradation of the RFP-client protein fusion, presumably due a failure to integrate into the ER membrane. In human U2OS cells lacking the core subunit EMC5 (ΔEMC5)^{13,14}, transiently expressed GFP-P2A-RFP-Xport-A exhibited a reduced RFP:GFP ratio when compared to WT U2OS cells (Figure 3C). Induced expression of EMC5 in ΔEMC5 cells restored the EMC and resulted in a RFP:GFP ratio that was comparable to WT cells (Figure 3C). Similar results were obtained from cells lacking the core subunit EMC6 (ΔEMC6, Figure 3F). In fact, Xport-A behaved comparably to SQS (Figure 3B and E), a TA protein involved in cholesterol homeostasis whose biogenesis has been established as EMC dependent^{13,14}. Importantly, there was no difference in the measured RFP:GFP ratios following expression of GFP-P2A-RFP-Xport-A4L in ΔEMC5 or ΔEMC6 cells, when compared to either WT or EMC rescue cell lines (Figure 3D-G). These results indicate that the EMC is important for post-translational stability of Xport-

Figure 3. EMC is required for the biogenesis of Xport-A in mammalian cells.

(A) Schematic representation of the reporter constructs used for analysis of protein biogenesis by flow cytometry. All constructs contain an N-terminal GFP and a C-terminal RFP, separated by a viral 2A peptide that mediates peptide bond skipping. Changes in the stability of SQS, Xport-A and Xport-A4L fused to RFP change the RFP:GFP fluorescence ratio.

(B-D) Histograms of flow cytometry data monitoring the fluorescence protein ratio in the indicated U2OS cell lines for each construct. "EMC5KO+EV" indicates a knockout of EMC5 stably harbouring empty vector, while "EMC5KO+EMC5" indicates EMC5KO cells rescued by inducible re-expression of a stably integrated EMC5.

(E-G) Histograms of flow cytometry data monitoring the fluorescence protein ratio in the indicated U2OS cell lines for each construct. "EMC6KO+EV" indicates a knockout of EMC6 stably harbouring empty vector, while "EMC6KO+EMC6" indicates EMC6KO cells rescued by re-expression of a stably integrated EMC6.

Figure 4. EMC6 is required for membrane insertion of Xport-A TMD.

(A) Schematic representation of the constructs used to monitor insertion efficiency in the ER membrane. All constructs contain a HA-tagged SQS cytosolic domain fused to a varying TMD sequence, which is C-terminally fused to an opsin tag (sequence indicated). Glycosylation of the C-terminal opsin tag acts as an insertion readout.

(B) Sequences of the TMD regions used for cloning and the immediate flanking residues are represented. The regions defined as TMDs were determined using TOPCONS⁶¹ and are in bold/underlined. Red residues indicate mutations made in the Xport-A TMD to increase its hydrophobicity. To the right of each sequence is the "transmembrane tendency" score of Zhao & London⁵⁵.

(C-D, F) Western-Blot of WT and EMC6KO (6KO) U2OS cell lines transiently transfected with the opsin-tagged constructs probed for HA. When indicated, denatured samples were treated with PNGase to confirm glycosylation. Glycosylated proteins shift upwards in the gel and are indicated by arrows (□).

(E-G) Western-Blot of WT and EMC6KO (6KO) U2OS cell lines transiently transfected with the opsin-tagged constructs probed for Actin and EMC6.

A and that this effect relies on the low hydrophobicity of its TMD.

EMC6 is required for membrane insertion of Xport-A TMD

Xport-A is a tail-anchored protein and its TMD must be inserted post-translationally into the ER membrane. To monitor insertion efficiency in the ER membrane, we cloned the TMDs from Sec61β, SQS, and Xport-A into a common cassette containing a C-terminal opsin tag (Figure 4A). The opsin tag, if exposed to the ER lumen, is able to accept N-linked oligosaccharides, serving as an indicator of successful TMD insertion into the membrane⁵⁸. As predicted by our *Drosophila* results (Figures 1 and 2), the Xport-A TMD reporter is glycosylated in WT cells, but fails to be glycosylated in ΔEMC6 cells (Figure 4D), behaving similarly to HA-SQS-opsin (Figure 4C), as previously reported^{13,14}. In contrast, the Sec61β reporter remains fully glycosylated in ΔEMC6 cells (Figure 4C), consistent with its EMC-independent insertion pathway¹³.

Next we sequentially mutagenized the TMD of Xport-A to substitute the four most hydrophilic AAs with leucine (Figure 4B). By progressively increasing the hydrophobicity of Xport-A TMD, we attempted to identify a TMD hydrophobicity threshold for EMC recognition. A single point mutant of the Xport-A TMD (Xport-A1L) retained its dependence on the EMC for membrane insertion (Figure 4F), but subsequent mutations (Xport-A2L, -A3L) were increasingly glycosylated in ΔEMC6 cells (Figure 4F). With four leucine mutations in the Xport-A TMD (Xport-A4L), insertion became entirely EMC independent, as indicated by the glycosylation pattern equivalent to what is observed

Figure 5. Expression of Xport-A4L can rescue Rhodopsin-1 and TRP biogenesis defects in EMC3 mutant clones. (A-B) Immunostaining of 3rd instar larval eye imaginal discs with eyeless-Flippase-induced clones of cells homozygous for EMC3^{Δ4}, labelled by the absence of ubiGFP (green) show loss of (A) UAS-Xport-A-HA (anti-HA, in blue) and UAS-Rh1 (4C5, in red). (B) Expression of UAS-Xport-A4L (anti-HA, in blue) and UAS-Rh1 (4C5, in red) is observed in EMC3^{Δ4} homozygous cells. The UAS constructs were expressed under the control of GMR-GAL4. Scale bars represent 10 μm. (C-D) Immunostaining of mosaic adult retinas with eyeless-Flippase-induced clones of cells homozygous for EMC3^{e02662}, labelled by the absence of RFP (red), show loss of (C) TRP (Mab83F6, in blue) when UAS-HA-Xport-A is expressed. (D) Expression of TRP (Mab83F6, in blue) is observed in EMC3^{e02662} homozygous cells, when HA-Xport-A4L is expressed. The rhabdomeres are stained for actin (Phalloidin, in green). The UAS constructs were expressed under the control of Rh1-GAL4. Scale bars represent 10 μm. (E) The ratio of fluorescence intensity in EMC3 homozygous mutant cells over that of WT cells (EMC3^{mut}/WT) was measured and plotted. Proteins were classified into three groups according to the ratio measured: normal (ratio 0.8-1.2; bar in dark blue), rescue (ratio 0.5-0.8, bars in light blue) and strongly reduced (ratio <0.5; bars in red). Error bars correspond to standard deviation (stdev). Significance was determined by Welch's t-test: **p ≤ 0.01, ****p ≤ 0.0001.

in WT cells (Figure 4D). This finding supports the existence of a TMD “hydrophobicity threshold” that directs clients to insert in the ER membrane via the EMC.

Expression of Xport-A4L can rescue Rhodopsin-1 and TRP biogenesis defects in EMC3 mutant clones

Xport-A serves as an essential chaperone during the biogenesis and maturation of both Rh1 and TRP; proteins that are essential for light sensing within the rhabdomere^{48,49}. The protein levels of both Rh1 and TRP are significantly reduced in

Xport-A loss of function mutations^{48,49}, as well as in EMC mutations²⁹⁻³¹. Our demonstration that Xport-A is a client protein of the EMC, raised the possibility that loss of TRP and Rh1 observed in EMC mutants could instead be attributable to failed membrane insertion and biogenesis of Xport-A. To test this hypothesis, we compared the capacity of Xport-A and Xport-A4L to rescue expression of Rh1 and TRP in EMC mutant clones. While Xport-A overexpression in EMC3 mutant clones was unable to rescue Rh1 (Figure 4-A), we did find that Xport-A4L expression was sufficient to restore Rh1 protein levels in EMC3^{Δ4} clones (Figure 4-B). Likewise, Xport-A4L rescued TRP expression in cells homozygous for a hypomorphic mutation of EMC3 (EMC3^{e02662}, Figure 4D). Together, these data suggest that Rh1 and TRP biogenesis and maturation are not directly dependent on the EMC,

but rather appear to be indirectly dependent on it, via the membrane insertion of Xport-A.

DISCUSSION

Understanding the full scope of client proteins that depend on the EMC for biogenesis and maturation is yet to be fully appreciated. Here we describe an approach to predict, screen and validate candidate EMC client proteins which we apply to the *Drosophila* tail-anchored (TA) proteome. From a subset of TA candidate proteins with low predicted hydrophobicity scores, we identified fan and Xport-A as novel clients of the EMC. Taken together with previous studies that identified SQS, a key enzyme in sterol biogenesis^{13,14} and ZFPL1, a zinc-finger containing protein required for efficient ER to Golgi transport/integrity of the *cis-Golgi*^{15,59}, our observations emphasize the importance of the EMC as an insertase for a subset of TA proteins whose insertion is independent of the GET/TRC40-complex. Taken together with previous studies^{13–15}, our data also demonstrate conservation of the EMC's ability to recognize TA proteins with reduced TMD hydrophobicity caused by the presence of charged/polar residues. Notably, the EMC appears to have additional levels of selectivity TA client proteins. Even though our predictions identified multiple candidate proteins with reduced TMD hydrophobicity, only a small fraction of these were confirmed as clients in our genetic screen in *Drosophila*. Future studies will be required to dissect the precise basis whereby the presence, nature and position of non-hydrophobic residues confers EMC dependency.

Fan is a member of VAP (VAMP associated protein) family of proteins that bind OSBPs (Oxysterol binding proteins) and play a role in sterol trafficking, including in the “individualization complex”, a cellular structure that promotes the membrane reorganization process (sperm individualization) that is crucial for the production of mature, functional sperm. Consistent with the role for fan in sperm individualization, depletion of several independent EMC subunits in testis (particularly EMC5, EMC6) render male flies infertile, with individualization defects similar to those reported for ablation of fan or OSBP family proteins⁴⁷. Our identification of Fan as an EMC client, further strengthens the connections between the EMC and sterol homeostasis, established by the identification of SQS and SOAT1 as EMC clients^{13,14}.

The identification of Xport-A as an EMC client raises some interesting questions. The Xport locus, consisting of the bicistronic operon separately encoding the Xport-A and Xport-B proteins, is required for the stable expression and localization of Rh1 (a GPCR) and TRP (a calcium ion channel) to the rhabdomeres^{48,49}. Loss of Xport

proteins results in depletion of Rh1 and TRP, causing photoreceptor degeneration and visual impairment. Intriguingly, as highlighted above, several studies have reported the loss of Rh1, TRP and accompanying photoreceptor degeneration in EMC-mutant flies^{29–31}. As polytopic membrane proteins, Rh1 and TRP could directly engage the EMC during their biogenesis. However, our identification of Xport-A as an EMC client suggested the possibility that EMC governance over Rh1 and TRP expression could be indirect. Such a model is supported by the results obtained with the Xport-A mutant engineered to be EMC independent (Xport-A4L), which is sufficient to rescue expression of both Rh1 and TRP in mutant clones lacking functional EMC (Figure 5). Moreover, in the heterologous setting of mammalian cells where neither Rh1 nor TRP are expressed, Xport-A remains an EMC client with no collateral impact on its biogenesis. Conceptually, a model where “third party” proteins are lost as an indirect consequence of impaired biogenesis of a bona fide EMC client protein, could help to explain the pleiotropic impact of loss on multiple classes of membrane proteins when EMC is ablated.

While an attractive hypothesis, some important caveats could apply to the model above. First, Rh1 is a GPCR, a class of proteins that contain an N-terminal SAS which have recently been proposed to be direct EMC clients¹⁹. Chitwood and colleagues showed that the EMC can mediate the insertion of the 1st TMD of some SAS-containing GPCRs, after which insertion proceeds co-translationally in a Sec61-translocon-dependent manner¹⁹. However, this study did not investigate EMC requirement for *Drosophila* Rh1 membrane insertion. Moreover, expression levels of a Rh1 C-terminal truncation containing only the first TMD were uncompromised in EMC mutant *Drosophila* retinal cells³¹. In fact, only a truncation containing the first 5 of the 7 Rh1 TMDs was expressed at lower levels in EMC mutant tissue³¹. This study did, however, assessed expression levels of Rh1 mutants by immunofluorescence, only, and did not investigate either Rh1 TMD membrane insertion or topogenesis.

TRP contains 6 TMDs and a cytoplasmic N-terminus, which does not currently conform to the elucidated criteria for EMC client candidacy. An outstanding question is whether such non-canonical membrane protein clients are in fact engaged by the EMC for their biogenesis and if so, how? Several studies have reported the loss of multi-pass membrane proteins with loss of EMC functionality^{15,23,24}. Often, these studies correlate EMC-dependence with abundance changes at steady-state, rather than monitoring EMC-dependent biogenesis directly. Consequently, it remains possible that loss of some of these putative EMC clients will occur indirectly, as we have

demonstrated for Rh1 and TRP through Xport-A. Moreover, metabolic changes arising from disruption to EMC-dependent processes, such as sterol regulation, may also indirectly impact protein levels post-translationally through changes in intrinsic stability or homeostatic mechanisms linked to transcription. Care must be exercised when interpreting EMC-dependent biogenesis of complex multi-pass clients to ensure that, in fact, what is observed is a direct impact linked to TMD engagement.

Recently, studies from the Weissman lab identified features within the EMC structure that differentially impact TA protein clients when compared to multi-pass membrane proteins including TMEM97, a 4-TMD protein with a similar cytoplasmic N-terminal topology to TRP²³. One interpretation is that EMC may be able to handle distinct classes of clients through different features. Miller-Vedam and colleagues proposed that the EMC may function not only as an insertase for TA proteins, but also as a “holdase” factor, that participates in the biogenesis of a wider class of polytopic membrane proteins via distinct mechanisms, which may also include the recruitment of client-specific chaperones²⁴. It will be important to determine whether the impact of EMC on TRP is fully indirect, through its impact on Xport-A, or whether TRP directly engages the “holdase” function of EMC. The latter case could also involve recruitment of Xport-A to the EMC, to serve as a client-specific chaperone for Rh1/TRP, to aid in their maturation and/or release from the EMC and subsequent trafficking. Clearly, the biogenesis of membrane protein via the EMC has far reaching impact on organismal physiology and defining precisely the protein clients of EMC is indispensable to understand the selectivity of this complex.

METHODS

Prediction of TA proteins in *Drosophila* proteome

The *Drosophila melanogaster* proteome (UP000000803, release 2018_10_26) was downloaded from the UniProt website⁶⁰. Canonical forms of protein sequences longer than 49 amino acids (AAs) and shorter than 6,001 AAs were selected. Topology (*i.e.*, localization of TMDs and non-membrane regions) and presence of the signal peptide were predicted for each protein using the PolyPhobius⁵¹ predictor implemented within the TOPCONS consensus predictor⁶¹. The presence of the mitochondrial targeting peptide was predicted for each protein using TargetP 1.1⁶². Proteins without predicted signal and mitochondrial targeting peptides, lacking the keywords “SIGNAL” and “TRANSIT” in their UniProt annotation, and containing only a single predicted TMD localized close to the C-terminus (*i.e.*, followed by 35 or fewer flanking AAs) were selected. All thus selected protein sequences were analysed using the DeepLoc⁵² cellular localization predictor in order to further filter out soluble proteins. The TMD hydrophobicity was determined using the transmembrane tendency values for individual amino acids as

defined by Zhao & London⁵⁵. The C-terminal tail charge was calculated by adding +1 for arginine and lysine residues and -1 for aspartic and glutamic acid residues. The prediction of TA proteins was validated using experimentally identified proteins^{13,63}. Characterization of the distribution of cell localizations and biophysical features was performed in GraphPad Prism v.9 and significance was evaluated using the Kolmogorov-Smirnov test.

Fly stocks and crosses

Flies and crosses were raised with standard cornmeal, at 25°C under 12hr light/ 12 hr dark cycles.

Fly stocks used were obtained from multiple sources and are listed in Supplementary Table 1 and 2.

To determine if expression of predicted tail anchored proteins was dependent on EMC3, fly stocks carrying C-terminally HA-tagged constructs on the third chromosome (obtained from FlyOrf – Zurich ORFeome Project) were crossed with flies with an EMC3 mutant allele EMC3^{Δ4}FRT40A²⁹ on the second chromosome. The resulting flies (w; EMC3^{Δ4}FRT40A; UAS-TA-HA) were then crossed with “eyFLP, GMRGAL4; ubiGFPRT40A” to obtain EMC3^{Δ4} mosaic eye imaginal discs, expressing the predicted TA proteins. In the case of VAP33A, hid and Na⁺K⁺ATPase α-subunit, there was no need to express tagged fly lines, as these genes are endogenously expressed in eye imaginal discs and antibodies were available.

Immunostaining and imaging

Larval eye imaginal discs and adult retina were dissected in 1x PBS, fixed in 4% PFA (paraformaldehyde) at room temperature (15min for eye imaginal discs, 30min for adult retina) and washed three times in PBT (1x PBS + 0.3% Triton X-100), 10 min each. For eye imaginal discs, primary antibodies were diluted in blocking buffer (1x PBS, 0.1% TX-100, 0.1% BSA, 250mM NaCl) and for adult retina, primary antibodies were diluted in 5% Fetal Calf Serum in PBT. Primary antibody incubations were performed overnight at 4°C under gentle agitation. Primary antibodies were as follow: mouse anti-ELAV (1:50) (9F8A9, Developmental Studies Hybridoma Bank (DSHB)), rat anti-ELAV (1:50) (7E8A10, DSHB), rat anti-HA (1:400) (7C9, Chromotek), mouse anti-Rh1 (1:50) (4C5, DSHB), mouse anti-Na⁺K⁺ATPase α subunit (1:25) (A5-C, DSHB), anti-VAP33A (1:50)(kind gift of Hugo Bellen), anti-hid (kind gift from Hyung Don Ryoo), rabbit anti-HA (1:40) (9110, Abcam), mouse anti-TRP (1:25) (Mab83F6, DSHB). Samples were washed three times with PBT and incubated with appropriate secondary antibodies (Jackson Immuno Research) for 2 hours at room temperature. Control staining of actin in the rhabdomeres of adult retina was performed with Phalloidin-FITC (1:500) (Abcam 235137). After rinsing three times with PBT, samples were mounted in Vectashield (Vector Laboratories). Adult eye samples were mounted in a bridge formed by two coverslips to prevent the samples from being crushed and analysed on a Zeiss LSM 880 confocal microscope. Larval eye imaginal discs were analysed on a Leica SP5Live confocal microscope.

Adult testis were dissected in 1x PBS and fixed in 4% PFA overnight at 4°C under gentle agitation. Following this, samples were washed three times in PBS, 5 min each and then treated with PBT at room temperature for 30 min. After three more washes with PBS (5 min each), testis were incubated with rabbit anti-cleaved caspase 3 (1:200) (Asp175, Cell Signaling) diluted in 0.1% PBT, overnight at 4°C. Samples were washed

three times with PBS, 5 min each, and incubated with anti-rabbit Cy3 (1:200) (Jackson Immuno Research). After three more washes with PBS (5 min each), samples were incubated with DAPI, diluted in 0.1% PBT for 10 min. After rinsing three more times with PBS, samples were mounted in Vectashield and image acquisition was performed in Leica SP5Live.

For measurement of seminal vesicle size, testis were stained with DAPI, as described above and image acquisition was performed in Leica SP5Live. Imaging was performed at the widest portion of the seminal vesicle and area measurement was performed in Fiji. Characterization of distribution of seminal vesicle size was performed in Prism and significance was determined by Welch's t-test.

Quantification of protein expression in EMC3^{Δ4} mutant clones

Quantification was executed in Fiji⁶⁴. Maximum intensity projections were performed on the raw confocal data to visualize EMC3^{Δ4} mutant clones. The three coloured images were split into single colour images. Channel 1 corresponded to ubiGFP, in which homozygous mutant clones were identified by the absence of fluorescence. Channel 2 corresponded to the fluorescence signal for the protein of interest, while channel 3, which showed ELAV staining, marked the photoreceptor cells. Segmentation was first performed on channel 3 to define the GMRGAL4 domain, and then on channel 1 to highlight patches with no detectable ubiGFP fluorescence, which coincided with EMC3^{Δ4} homozygous mutant cells. Both segmentations were performed using the Otsu thresholding method⁶⁵. Following segmentation, the mathematical process "AND" was performed to intersect mutant patches delimited in the second segmentation with the GMRGAL4 domain identified in the first segmentation. EMC3^{Δ4} mutant patches in the GMRGAL4 domain were defined as ROIs. Mean grey values for the ROIs in channel 2 were measured (EMC3^{Δ4} mean) and compared to mean grey values of WT patches of comparable area in channel 2 (WT mean). A ratio of fluorescence intensity in EMC3^{Δ4} mutant cells/WT cells was calculated for each protein. The Fiji macro developed for this analysis can be found on GitHub (<https://github.com/zeserrado-marques/Quantification-of-protein-expression-inside-dark-patches>). For quantification, at least 6 mutant patches were quantified per eye imaginal disc, and at least three eye imaginal discs derived from distinct flies were used. Significance was evaluated using the Welch's t-test in Graphpad Prism 9 software.

Generation of HA tagged Xport-A transgenic flies

To generate transgenic flies expressing HA tagged Xport-A we PCR amplified the full open reading frame of Xport-A from plasmid pAC5.1-FLAG::XPORT-A⁴⁹ (kind gift of Craig Montell). For N-terminally tagged Xport-A, we used the following primers: 5'-CCGGATTACGCCAAGCCGAAGAAATCGGCC-3' and 5'-CGCAGATCTGTTAACGAATTCTTACTGCTGATGC TCCTCGC-3'. For C-terminally tagged Xport-A we used the following primers: 5'-GAATAGGGAATTGGGAATTCATG AAGCCGAAGAAATCGGCC-3' and 5'-TCAGGGACGTCGTA CGGTACTGCTGATGCTCCTCGCGAAAG-3'. For amplification of 3XHA tag, we PCR amplified 3XHA with following primers: 5'-TGAATAGGGAATTGGGAATTCATGTAC CCGTACGACGTCCTGA-3' and 5'-TCTTCGGCTTGGC GTAATCCGGCACATCA-3' for HA-Xport-A. For PCR amplification of 3XHA intended for Xport-A-HA, we used the following primers: 5'-GAAAGCTTTCGCGAGGAGCATCAGCA GTACCCGTACGACGTCCTGA -3' and 5'-CGCAGATC GTTAAACGAATTCTTAGCCGGCGTAATCCGGCACATCA-3'.

After amplification of Xport-A and 3XHA inserts, we proceeded to perform a Gibson assembly® (New England Biolabs) with plasmid pUASTattb (previously linearized with EcoRI).

Xport-A4L constructs were made by mutating hydrophilic AAs to Leucine, using KOD polymerase site-directed mutagenesis (Merck Millipore). Xport-A4L insert was then amplified by PCR with primers listed above and Gibson assembly was used to create N-terminal and C-terminal HA tagged pUASTattb constructs.

The resulting constructs were inserted into the *attp2* site by PhiC31 integrase mediated transgenesis at the Champalimaud Foundation transgenics facility.

Cell culture

U2OS Flp-In™ TRex™ cell lines describe previously¹⁴ were maintained in Dulbecco's Modified Eagle Medium (DMEM)(Biowest) + 10% v/v fetal bovine serum (FBS) (Biowest) + L-glutamine (2mM) (GRiSP) + 1% v/v Penicillin-Streptomycin (Biowest) in 5% CO₂ at 37°C. Stable expression cell lines, generated by flippase mediated site-specific integration, were continuously cultured in 100 µg/ml hygromycin B (Merck Millipore). Cells containing a stably expressed tetracycline-induced constructs were cultured in tetracycline-free FBS (Biowest), 100 µg/ml hygromycin, with their expression induced by addition of 1 µg/ml doxycycline (Sigma Aldrich).

Plasmids and expression constructs

Xport-A and XportA4L dual fluorescent reporters were generated by first amplifying the TMD and flanking residues of Xport-A (AFEMMKLFVFANTIMLIVTMAWPHIKEQFYM) and Xport-A4L (AFEMMKLFVFALLIMLIVLMAWPLIKEQFYM) using the primers PFmcherry-xportTMD (GGCGGCATGGACGAGCGTTACAAGGCATTTGAAATGATG AAATCTC) and PRmcherry-xportTMD (TTATGATCAGTTATCTAGATCCGGTTTACATGTAGAATTGC TCCTTGAT). Reporter plasmids were constructed from BamHI-linearized, dephosphorylated pcDNA5/FRT/TO-GFP-P2A-mcherry vector (kind gift from Manu Hegde, LMB-Cambridge) using Gibson assembly® (New England Biolabs). The SQS fluorescent reporter GFP-P2A-mcherry-SQS has been reported previously¹³. TMD insertion reporters pcDNA5/FRT/TO-HA:SQS and pcDNA5/FRT/TO-HA:SQS-Sec61□TMD^{opsin} have been reported previously¹⁴. To generate XportATMD^{opsin} (pcDNA5/FRT/TO-HA:SQS-XportATMD^{opsin}) and hydrophobic TMD variants (1L/2L/3L/4L), geneblocks (IDT) containing each TMD sequence fused to the opsin tag were purchased. Each gene block contained flanking restriction sites for BamHI and XhoI that enabled subcloning into linearized pcDNA5/FRT/TO-HA:SQS.

Flow cytometry

Analysis of protein post-translational stability was performed according to described methods¹³. Cells were transfected with fluorescent reporter constructs (1 µg total DNA/well, 6 well plate) using GenJet (SignaGen Laboratories) according to manufacturer's instructions. The amount of fluorescent reporter for each protein of interest was titrated according to transfection efficiency and a control plasmid was co-transfected in order to maintain the total amount of DNA (1 µg/well). The U2OS FlpIn Trex EMC5KO rescue cell lines were cultured in the presence of doxycycline (1µg/ml) for 24 hours

prior to transfection, to rescue EMC5 expression. After transfection, cells were trypsinized, washed twice with PBS +2%FCS and stained with DAPI (to determine cell viability). Flow cytometry was performed on a BD LSR Fortessa X20 instrument, where 20,000 RFP-positive cells were collected. Data analysis was performed in FlowJo (v10). For quantification, cells were first gated for GFP (translation reporter) and mean RFP and GFP fluorescence intensities were determined. Then, RFP:GFP ratios were calculated and normalized to the ratio observed in WT cells.

Transient transfection of TMD insertion reporters

1×10⁶ cells/6cm² tissue culture plate were transfected with TMD insertion reporters (2.5 µg total DNA/6cm² plate) using GenJet (SigmaGen Laboratories) according to manufacturer's instructions. The following day, cells were lysed.

Immunoblotting

Cells were washed twice in ice-cold PBS and lysed in TX-100 lysis buffer (50mM Tris-HCl pH 7.4, 150mM NaCl, 1% Triton X-100, protease inhibitors) + 5mM EDTA. Following incubation on ice (10 min), lysates were centrifuged (top speed, 10min, 4°C). Protein concentrations were determined by Bradford and protein denaturation was performed by adding 3x LDS buffer + DTT and incubating the samples at 65°C (15 min). After denaturation, samples were digested with PNGaseF (New England Biolabs) for 2 hours at 37°C, according to manufacturer's instructions.

Samples were fractioned by SDS-PAGE, transferred to PVDF membranes (Amersham Hybond) and probed with following antibodies: mouse anti-HA-HRP (1:1000) (12013819001, Roche) rabbit anti-EMC6 (1:300) (84902, Abcam) and rabbit anti-actin (1:2000) (8227, Abcam).

Fertility assay

To examine fertility, individual adult males (3-days old) were mated to three wild-type virgin females in separate vials. All RNAi lines were expressed with bam-GAL4 driver. For each RNAi line, thirty individual males were mated to virgin females. The females were transferred after 7 days at 25°C to fresh vials. Progeny from the original vial and the first transfer vial were counted. Characterization of distribution of progeny/female/day was performed in Prism and significance was determined by Welch's t-test.

Acknowledgements

We thank the flow cytometry facility at IGC, the microscopy facility at IGC (more specifically José Marques), the microscopy facility at ITQB-NOVA, the fly facility from IGC and the transgenics facility at Champalimaud Foundation – Centre for the Unknown. We thank the Developmental Studies Hybridoma Bank (DSHB) at the University of Iowa for antibodies and the Bloomington Drosophila Stock Centre for fly stocks. We thank Craig Montell for flies, plasmids and antibodies, Akiko Satoh for the EMC3 mutant flies, Manu Hegde for plasmids, Norbert Volkmar and Miguel Cavadas for discussions and suggestions, Nansi Jo Colley and Steve Britt for antibodies, Hugo Bellen for VAP33A antibody, Hyung Don Ryoo for UAS-Rh1 flies and hid antibody. We thank Sebastien Alfaiate for help with figure design. The project leading to these results has received funding from 'la Caixa' Foundation (ID 100010434), under the agreement <LCF/PR/HR17/52150018>.

Competing interests

The authors declare no competing interests.

Author contributions

Research design: all authors. TA protein prediction was done by DJ and KS. All experiments were performed by CJG, except the *Drosophila* male fertility experiments (LCV). Paper writing and editing: all authors.

Data availability

The macro method developed for the quantification of protein expression in EMC3^{Δ4} mutant clones can be found on Github (<https://github.com/zeserrado-marques/Quantification-of-protein-expression-inside-dark-patches>).

Abbreviations

AAs – amino acids
cb – cystic bulges
Cryo-EM – cryo-electron microscopy
EMC – ER membrane protein complex
ER – endoplasmic reticulum
ERAD – ER-associated protein degradation
fan - farinelli
GPCR – G-protein coupled receptor
HA- hemagglutinin
OSBP – oxysterol binding protein
PFA - paraformaldehyde
Rh1 – Rhodopsin1
SAS – signal anchor sequence
SOAT1 – sterol O-acyltransferase 1
SQS – squalene synthase
SRP – signal recognition particle
TA – tail anchored
TMD – transmembrane domain
TRC – TMD recognition complex
TRP – transient receptor potential
VAP – VAMP associated protein
Xport-A – exit protein of rhodopsin and TRP - A

References

1. Wallin, E. & Von Heijne, G. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.* **7**, 1029–1038 (1998).
2. Fagerberg, L., Jonasson, K., Von Heijne, G., Uhlén, M. & Berglund, L. Prediction of the human membrane proteome. *Proteomics* **10**, 1141–1149 (2010).
3. Guna, A. & Hegde, R. S. Transmembrane Domain Recognition during Membrane Protein Biogenesis and Quality Control. *Current Biology* **28**, R498–R511 (2018).
4. Kutay, U., Hartmann, E. & Rapoport, T. A. A class of membrane proteins with a C-terminal anchor. *Trends Cell Biol.* **3**, 72–75 (1993).
5. Hegde, R. S. & Keenan, R. J. Tail-anchored membrane protein insertion into the endoplasmic reticulum. *Nat. Rev. Mol. Cell Biol.* **12**, 787–798 (2011).
6. Stefanovic, S. & Hegde, R. S. Identification of a Targeting Factor for Posttranslational Membrane Protein Insertion into the ER. *Cell* **128**, 1147–1159 (2007).
7. Schuldiner, M. *et al.* The GET Complex Mediates Insertion of Tail-Anchored Proteins into the ER Membrane. *Cell* **134**, 634–645 (2008).

8. Mateja, A. *et al.* Structure of the Get3 targeting factor in complex with its membrane protein cargo. *Science* **347**, 1152–1155 (2015).
9. Mariappan, M. *et al.* The mechanism of membrane-associated steps in tail-anchored protein insertion. *Nature* **477**, 61–69 (2011).
10. Stefer, S. *et al.* Structural Basis for Tail-Anchored Membrane Protein Biogenesis by the Get3-Receptor Complex. *Science* **333**, 758–762 (2011).
11. Wang, F., Brown, E. C., Mak, G., Zhuang, J. & Denic, V. A chaperone cascade sorts proteins for posttranslational membrane insertion into the endoplasmic reticulum. *Mol. Cell* **40**, 159–171 (2010).
12. Rao, M. *et al.* Multiple selection filters ensure accurate tail-anchored membrane protein targeting. *Elife* **5**, 1–24 (2016).
13. Guna, A., Volkmar, N., Christianson, J. C. & Hegde, R. S. The ER membrane protein complex is a transmembrane domain insertase. *Science* **359**, 470–473 (2018).
14. Volkmar, N. *et al.* The ER membrane protein complex promotes biogenesis of sterol-related enzymes maintaining cholesterol homeostasis. *J. Cell Sci.* **132**, (2019).
15. Tian, S. *et al.* Proteomic Analysis Identifies Membrane Proteins Dependent on the ER Membrane Protein Complex. *Cell Rep.* **28**, 2517–2526.e5 (2019).
16. Wideman, J. G. The ubiquitous and ancient ER membrane protein complex (EMC): tether or not? *F1000Research* **4**, 624 (2015).
17. Jonikas, M. C. *et al.* Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. *Science* **323**, 1693–7 (2009).
18. Christianson, J. C. *et al.* Defining human ERAD networks through an integrative mapping strategy. *Nat. Cell Biol.* **14**, 93–105 (2011).
19. Chitwood, P. J., Juszkiwicz, S., Guna, A., Shao, S. & Hegde, R. S. EMC Is Required to Initiate Accurate Membrane Protein Topogenesis. *Cell* **175**, 1–13 (2018).
20. Pleiner, T. *et al.* Structural basis for membrane insertion by the human ER membrane protein complex. *Science* **369**, 433–436 (2020).
21. O'Donnell, J. P. *et al.* The architecture of EMC reveals a path for membrane protein insertion. *Elife* **9**, 1–30 (2020).
22. Bai, L., You, Q., Feng, X., Kovach, A. & Li, H. Structure of the ER membrane complex, a transmembrane-domain insertase. *Nature* **584**, 475–478 (2020).
23. Miller-Vedam, L. E. *et al.* Structural and mechanistic basis of the EMC-dependent biogenesis of distinct transmembrane clients. *Elife* **9**, e62611 (2020).
24. Shurtleff, M. J. *et al.* The ER membrane protein complex interacts cotranslationally to enable biogenesis of multipass membrane proteins. *Elife* **7**, 1–23 (2018).
25. Louie, R. J. *et al.* A yeast phenomic model for the gene interaction network modulating CFTR-ΔF508 protein biogenesis. *Genome Med.* **4**, 103 (2012).
26. Lakshminarayan, R. *et al.* Pre-emptive Quality Control of a Misfolded Membrane Protein by Ribosome-Driven Effects. *Curr. Biol.* **30**, 854–864.e5 (2020).
27. Richard, M., Boulon, T., Robert, V. J. P., Richmond, J. E. & Bessereau, J.-L. Biosynthesis of ionotropic acetylcholine receptors requires the evolutionarily conserved ER membrane complex. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E1055–63 (2013).
28. Taylor, M. R. *et al.* The zebrafish pob gene encodes a novel protein required for survival of red cone photoreceptor cells. *Genetics* **170**, 263–273 (2005).
29. Satoh, T., Ohba, A., Liu, Z., Inagaki, T. & Satoh, A. K. dPob/EMC is essential for biosynthesis of rhodopsin and other multi-pass membrane proteins in Drosophila photoreceptors. *Elife* **4**, 1–21 (2015).
30. Xiong, L. *et al.* ER complex proteins are required for rhodopsin biosynthesis and photoreceptor survival in Drosophila and mice. *Cell Death Differ.* **27**, 646–661 (2020).
31. Hiramatsu, N., Tago, T., Satoh, T. & Satoh, A. K. ER membrane protein complex is required for the insertions of late-synthesized transmembrane helices of Rh1 in Drosophila photoreceptors. *Mol. Biol. Cell* **30**, 2890–2900 (2019).
32. Tang, X. *et al.* EMC3 coordinates surfactant protein and lipid homeostasis required for respiration. *J. Clin. Invest.* **127**, 4314–4325 (2017).
33. Coelho, J. P. L. *et al.* A network of chaperones prevents and detects failures in membrane protein lipid bilayer integration. *Nat. Commun.* **10**, 1–10 (2019).
34. Lahiri, S. *et al.* A Conserved Endoplasmic Reticulum Membrane Protein Complex (EMC) Facilitates Phospholipid Transfer from the ER to Mitochondria. *PLoS Biol.* **12**, (2014).
35. Janer, A. *et al.* SLC 25 A 46 is required for mitochondrial lipid homeostasis and cristae maintenance and is responsible for Leigh syndrome. *EMBO Mol. Med.* **8**, 1–20 (2016).
36. Li, Y. *et al.* A novel ER-localized transmembrane protein, EMC6, interacts with RAB5A and regulates cell Autophagy. *Autophagy* **9**, 150–163 (2013).
37. Shen, X. *et al.* EMC6/TMEM93 suppresses glioblastoma proliferation by modulating autophagy. *Cell Death Dis.* **7**, e2043 (2016).
38. Bagchi, P., Inoue, T. & Tsai, B. EMC1-dependent stabilization drives membrane penetration of a partially destabilized non-enveloped virus. *Elife* **5**, 1–23 (2016).
39. Savidis, G. *et al.* Identification of Zika Virus and Dengue Virus Dependency Factors using Functional Genomics. *Cell Rep.* **16**, 232–246 (2016).
40. Barrows, N. J. *et al.* Dual roles for the ER membrane protein complex in flavivirus infection: viral entry and protein biogenesis. *Sci. Rep.* **9**, 1–16 (2019).
41. Lin, D. L. *et al.* The ER Membrane Protein Complex Promotes Biogenesis of Dengue and Zika Virus Non-structural Multi-pass Transmembrane Proteins to Support Infection. *Cell Rep.* **27**, 1666–1674.e4 (2019).
42. Bagchi, P., Torres, M., Qi, L. & Tsai, B. Selective EMC subunits act as molecular tethers of intracellular organelles exploited during viral entry. *Nat. Commun.* **11**, 1–15 (2020).
43. Harel, T. *et al.* Monoallelic and Biallelic Variants in EMC1 Identified in Individuals with Global Developmental Delay, Hypotonia, Scoliosis, and Cerebellar Atrophy. *Am. J. Hum. Genet.* **98**, 562–570 (2016).
44. Zhou, Y. *et al.* EMC 10 governs male fertility via maintaining sperm ion balance. *J. Mol. Cell Biol.* **6**, 503–514 (2018).
45. Chitwood, P. J. & Hegde, R. S. The Role of EMC during Membrane Protein Biogenesis. *Trends Cell Biol.* **29**, 1–14 (2019).
46. Volkmar, N. & Christianson, J. C. Squaring the EMC - How promoting membrane protein biogenesis impacts cellular functions and organismal homeostasis. *Journal of Cell Science* **133**, 1–14 (2020).
47. Ma, Z., Liu, Z. & Huang, X. OSBP- and FAN-mediated sterol requirement for spermatogenesis in Drosophila. *Development* **137**, 3775–3784 (2010).
48. Rosenbaum, E. E. *et al.* XPORT-Dependent Transport of TRP and Rhodopsin. *Neuron* **72**, 602–615 (2011).
49. Chen, Z., Chen, H. C. & Montell, C. TRP and Rhodopsin Transport Depends on Dual XPORT ER Chaperones Encoded by an Operon. *Cell Rep.* **13**, 573–584 (2015).
50. Borgese, N., Colombo, S. & Pedrazzini, E. The tale of tail-anchored proteins: Coming from the cytosol and looking for a membrane. *J. Cell Biol.* **161**, 1013–1019 (2003).

51. Käll, L., Krogh, A. & Sonnhammer, E. L. L. An HMM posterior decoder for sequence feature prediction that includes homology information. *Bioinformatics* **21**, 251–257 (2005).
52. Almagro Armenteros, J. J., Sønderby, C. K., Sønderby, S. K., Nielsen, H. & Winther, O. DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics* **33**, 3387–3395 (2017).
53. Costello, J. L. *et al.* Predicting the targeting of tail-anchored proteins to subcellular compartments in mammalian cells. *J. Cell Sci.* **130**, 1675–1687 (2017).
54. Kriechbaumer, V. *et al.* Subcellular distribution of tail-anchored proteins in arabidopsis. *Traffic* **10**, 1753–1764 (2009).
55. Zhao, G. & London, E. An amino acid “transmembrane tendency” scale that approaches the theoretical limit to accuracy for prediction of transmembrane helices: Relationship to biological hydrophobicity. *Protein Sci.* **15**, 1987–2001 (2006).
56. Chen, Y. N. *et al.* Knockdown of ATPsyn-b caused larval growth defect and male infertility in Drosophila. *Arch. Insect Biochem. Physiol.* **88**, 144–154 (2015).
57. Arama, E., Agapite, J. & Steller, H. Caspase activity and a specific cytochrome C are required for sperm differentiation in Drosophila. *Dev. Cell* **4**, 687–697 (2003).
58. Brambillasca, S. *et al.* Transmembrane topogenesis of a tail-anchored protein is modulated by membrane lipid composition. *EMBO J.* **24**, 2533–2542 (2005).
59. Chiu, C. F. *et al.* ZFPL1, a novel ring finger protein required for cis-Golgi integrity and efficient ER-to-Golgi transport. *EMBO J.* **27**, 934–947 (2008).
60. Bateman, A. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).
61. Tsirigos, K. D., Peters, C., Shu, N., Käll, L. & Elofsson, A. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res.* **43**, W401–W407 (2015).
62. Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* **2**, 953–971 (2007).
63. Coy-Vergara, J., Rivera-Monroy, J., Urlaub, H., Lenz, C. & Schwappach, B. A trap mutant reveals the physiological client spectrum of TRC40. *J. Cell Sci.* **132**, (2019).
64. Schindelin, J. *et al.* Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
65. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man. Cybern.* **C**, 62–66 (1979).

Figure S1. Characterization of the predicted TA proteins.

(A) Pie chart showing the distribution of the cell localization in predicted TA proteins. The predicted TA proteins were predominately localized to the ER membrane (35%), mitochondrial membrane (~20%) and cell membrane (~20%). Other subcellular localizations predicted were Golgi membrane (~13%), nuclear membrane (9%) and peroxisome membrane (~3%).

(B) Pie chart showing the distribution of the cell localization in tested TA proteins. TA candidates with all predicted sub-cellular membrane localizations were tested. The tested proteins had the following sub-cellular localizations: ER membrane (~30%), mitochondrion membrane (~30%), cell membrane (~13%), nuclear membrane (~8,7%), Golgi membrane (~8,7%) and peroxisome membrane (~4,3%).

(C-F) Violin plots showing the distribution of biophysical features such as (C) TMD length, (D) average TMD hydrophobicity (TMD hydrophobicity in Zhao & London score, averaged to the number of AAs in each TMD), (E) tail length and (F) tail charge in TA proteins, categorized by different cellular localization. The mid-line in the violin plots represents the median of the population and the extremities below and above the median correspond to the minimum and maximum values. Significance was determined by Kolmogorov-Smirnov test: * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$.

Figure S2. EMC is required for sperm differentiation in *Drosophila*.

(A-B) Box-and-whisker plots of progeny per female per day for RNAi lines. Individual adult males of the represented RNAi lines were mated to wild-type virgin females in separate vials and the mean number of progeny for crosses of the different genotypes was counted. The mid-line in the box plots represents median progeny (per female/day), and whiskers below and above the box indicate the sample range. All tested EMC5 and EMC6 RNAi lines, and one EMC1 RNAi line (EMC1_1) have a significant reduction of progeny produced. Significance was determined by Welch's t-test: **** $p \leq 0.0001$.

(C) Box-and-whisker plots of seminal vesicle size measured for 15 males per RNAi line. The box represents the 25-75th percentiles, and the median is indicated. Whiskers below and above the box indicate the minimum and maximum seminal vesicle size. All tested EMC5 and EMC6 RNAi samples (Control_2) but not when lines show a significant reduction in sperm production compared to controls. Significance was determined by Welch's t-test: **** $p \leq 0.0001$.

(D) Detection of sperm by nuclear staining with DAPI in seminal vesicles. Boxed areas are magnified below. The characteristic needle-like nuclei of mature sperm can be seen in control, but not when UAS-EMC5_2 RNAi is expressed (EMC5_2).

(E) Immunostaining of testis with an antibody against active caspase-3 (red), showing presence of multiple cystic bulges in Control_2 (left panel) and absence of cystic bulges when UAS-EMC5_2 RNAi is expressed (right panel). Nuclear staining was performed with DAPI (blue).

Figure S3. EMC3 is required for the expression of Xport-A in the adult retina, but not Xport-A4L.

Immunostaining of mosaic adult retina with eyeless-Flippase-induced clones of cells homozygous for EMC3^{e02662}, labelled by the absence of RFP (red), show loss of (top panel) UAS-Xport-A-HA (anti-HA, in blue) in EMC3 mutant homozygous cells. By contrast, expression of UAS- Xport-A4L-HA (bottom panel) is not lost in EMC3^{e02662} homozygous mutant clones.

	T R A N S M E M B R A N E D O M A I N																			
Xport-A	A	F	E	M	M	K	L	F	V	F	A	N	T	I	M	L	I	V	T	M
<i>D. simulans</i>	A	F	E	M	M	K	L	F	V	F	A	N	T	I	M	L	I	V	T	M
<i>D. sechellia</i>	A	F	E	M	M	K	L	F	V	F	A	N	T	I	M	L	I	V	T	M
<i>D. yakuba</i>	A	F	E	M	M	K	L	F	V	F	A	N	T	I	M	L	I	V	T	M
<i>D. erecta</i>	A	F	E	M	M	K	L	F	V	F	A	N	T	I	M	L	I	V	T	M
<i>D. persimilis</i>	A	F	E	M	M	K	L	F	V	F	A	N	T	I	M	L	I	V	T	M
<i>D. pseudoobscura</i>	A	F	E	M	M	K	L	F	V	F	A	N	T	I	M	L	I	V	T	M
<i>D. ananassae</i>	A	F	E	M	M	K	L	F	V	F	A	N	T	I	M	L	I	V	T	M
<i>D. mojavensis</i>	A	F	E	M	M	K	L	F	V	F	A	N	T	I	M	L	I	V	T	M
<i>D. virilis</i>	A	F	E	M	M	K	L	F	V	F	A	N	T	I	M	L	I	V	T	M
<i>D. grimshawi</i>	A	F	E	M	M	K	L	F	V	F	A	N	T	I	M	L	I	V	T	M
<i>D. willistoni</i>	A	F	E	M	M	K	L	F	V	F	A	N	T	I	M	L	I	V	T	M

Figure S4. XportA TMD protein alignment across *Drosophila*.

Xport-A TMD sequences were aligned from 16 different *Drosophila* species, using Uniprot Blastp tool. Pink shading indicates hydrophobic residues, blue shading indicates positively charged residues, and grey shading indicates negatively charged residues. Uniprot accession numbers obtained from Blastp alignment are as follows: B4QSS5_DROSI, B4IHQ4_DROSE, B4GL95_DROYA, B3P066_DROER, B4GL95_DROPE, A0A6I8USC0_DROPS, B3LXT7_DROAN, B4KE44_DROMO, B4M4S9_DROVI, B4K0W3_DROGR, Q81176_DROWI.

Table S1 – Fly Lines used to test TA proteins		
FlyBase ID	Uniprot ID	Fly Line
CG8814	Q9VQI6_DROME	Flyorf F002670
CG3566	Q8IRRO_DROME	FlyOrf F003667
Xport-A	Q9VDS3_DROME	uasXportA-3xHA ATP2 (this study)
rtv	M9PJI2_DROME	FlyOrf F003933
fan	Q9VSD3_DROME	FlyOrf F003648
MIC10	Q8SY69_DROME	FlyOrf F002914
CG7130	Q9VNW0_DROME	FlyOrf F002684
PIG-X	A8DY62_DROME	FlyOrf F002949
CG11327	Q9VM98_DROME	FlyOrf F001719
CG43321	M9NEV9_DROME	FlyOrf F002322
CG5050	Q9VJD8_DROME	FlyOrf F004285
CG30269	Q9W217_DROME	FlyOrf F002892
zda	Q7K3D4_DROME	FlyOrf F003205
Tango11	A0A0B4LG59_DROME	FlyOrf F002828
CG34242	A1A6X2_DROME	FlyOrf F002579
CG12012	Q1ECB1_DROME	FlyOrf F003820
Dpck	Q9VI57_DROME	FlyOrf F003172
Syx18	Q9VC58_DROME	FlyOrf F003037
CG17580	Q7JW78_DROME	FlyOrf F000885
CG43210	A0A0B4K684_DROME	FlyOrf F002648
Syx13	Q9VU45_DROME	FlyOrf F003008

Table S2 – Fly Lines used for fertility experiments		
Fly Line number	Genotype	Fly Line VDRC #
Control 1	Landing site of RNAi constructs	60100
Control 2	Dicer- bamGAL4	
EMC1_1	EMC1 RNAi	107165
EMC1_2	EMC1 RNAi	8477
EMC3_1	EMC3 RNAi	105637
EMC3_2	EMC3 RNAi	9408
EMC4_1	EMC4 RNAi	108256
EMC4_2	EMC4 RNAi	8464
EMC5_1	EMC5 RNAi	12248
EMC5_2	EMC5 RNAi	48614
EMC6_1	EMC6 RNAi	100613
EMC6_2	EMC6 RNAi	49583
EMC7_1	EMC7 RNAi	108875
EMC7_2	EMC7 RNAi	8141
EMC8/9_1	EMC8/9 RNAi	26025
EMC8/9_2	EMC8/9 RNAi	26026