

# EMC rectifies the topology of multipass membrane proteins

Received: 25 April 2023

Haoxi Wu , Luka Smalinskaitė & Ramanujan S. Hegde  

Accepted: 8 September 2023

Published online: 13 November 2023

 Check for updates

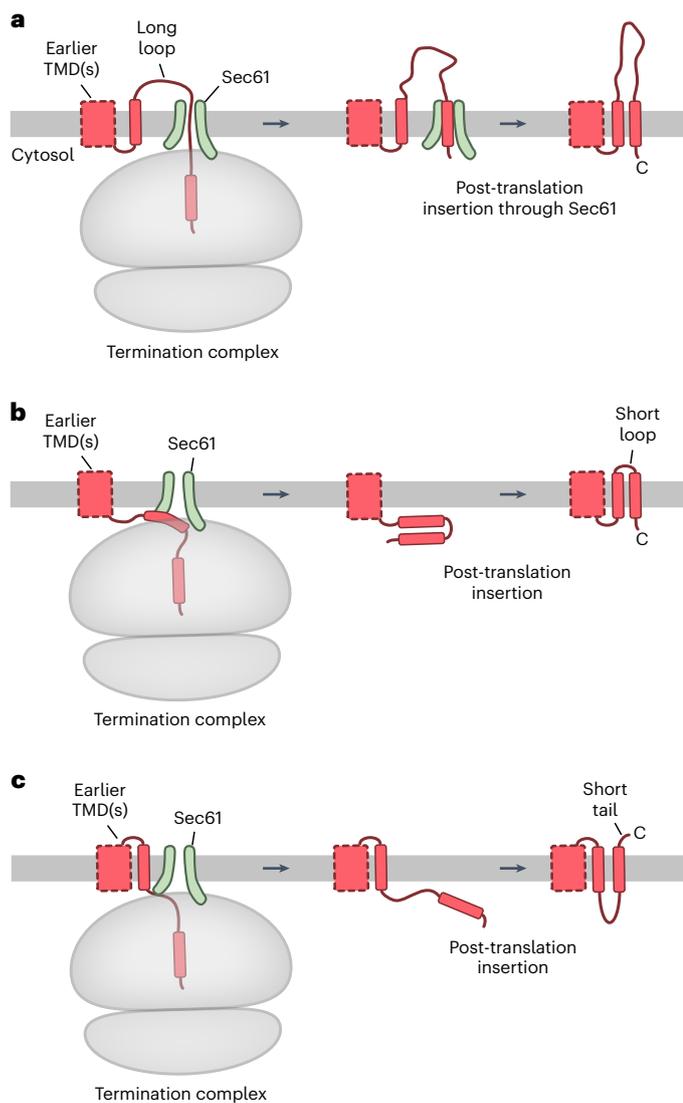
Most eukaryotic multipass membrane proteins are inserted into the membrane of the endoplasmic reticulum. Their transmembrane domains (TMDs) are thought to be inserted co-translationally as they emerge from a membrane-bound ribosome. Here we find that TMDs near the carboxyl terminus of mammalian multipass proteins are inserted post-translationally by the endoplasmic reticulum membrane protein complex (EMC). Site-specific crosslinking shows that the EMC's cytosol-facing hydrophilic vestibule is adjacent to a pre-translocated C-terminal tail. EMC-mediated insertion is mostly agnostic to TMD hydrophobicity, favored for short uncharged C-tails and stimulated by a preceding unassembled TMD bundle. Thus, multipass membrane proteins can be released by the ribosome–translocon complex in an incompletely inserted state, requiring a separate EMC-mediated post-translational insertion step to rectify their topology, complete biogenesis and evade quality control. This sequential co-translational and post-translational mechanism may apply to ~250 diverse multipass proteins, including subunits of the pentameric ion channel family that are crucial for neurotransmission.

Multipass membrane proteins, defined by the presence of more than one TMD, play crucial roles in the transfer of information and molecules across biological membranes<sup>1,2</sup>. Most eukaryotic multipass membrane proteins are inserted co-translationally at the endoplasmic reticulum as individual TMDs emerge from a membrane-bound ribosome<sup>3</sup>. Recent work suggests that at the mammalian endoplasmic reticulum, different parts of a multipass protein are inserted by different factors. The first TMD can be inserted by the EMC or by passing through a lateral gate in the Sec61 translocation channel<sup>4–7</sup>. Subsequent TMDs can also be inserted through the Sec61 lateral gate or into a lipid-filled cavity behind Sec61 created by the multipass translocon (MPT), an assembly of three complexes termed PAT, GEL and BOS<sup>8–10</sup>. GEL is structurally and evolutionarily related to the EMC, suggesting that it may be the insertion factor within the MPT<sup>11–13</sup>.

It is thought that both EMC and MPT insert TMDs only when the flanking translocated domain is shorter than ~50 amino acids, whereas Sec61 can mediate TMD insertion flanked by longer translocated domains<sup>4,8,9,14</sup>. Sec61 can translocate long hydrophilic polypeptides because it houses an aqueous translocation channel<sup>15–18</sup>, a feature lacking in either EMC or any of the MPT complexes<sup>8,19–23</sup>. By using EMC,

Sec61 and MPT in these ways, membrane proteins of widely varying topology and translocated domains can be co-translationally weaved into the lipid bilayer<sup>3,9</sup>. However, the last TMD poses unique problems if it is located within ~50 amino acids of the C terminus (hereafter termed a terminal TMD). As termination occurs before it can engage either Sec61 or the MPT, its insertion is necessarily post-translational. How a terminal TMD of a multipass protein is inserted is not clear.

The most straightforward case is one in which the penultimate TMD and the final TMD are separated by a long translocated loop. In this instance, the translocated loop would already be threaded into the ribosome-bound Sec61 channel at the time of termination (Fig. 1a). The terminal TMD would then necessarily enter the Sec61 channel after termination, from where it would enter the membrane post-translationally, presumably through Sec61's lateral gate. By contrast, the insertion mechanism for a terminal TMD preceded by a short translocated loop (Fig. 1b) or for a terminal TMD followed by a short translocated tail (Fig. 1c) is not clear. In the first situation, the penultimate TMD would not have enough of a tether to have engaged Sec61 before termination. The final two TMDs would be released from the ribosome and both would need to insert post-translationally



**Fig. 1 | Three modes of terminal TMD insertion for multipass membrane proteins.** TMDs located within  $\sim 50$  amino acids of the stop codon of a multipass membrane protein will be partially or completely inside the ribosome exit tunnel at the time of translation termination. This means that they will necessarily be inserted by one of three post-translational mechanisms depending on the preceding membrane domain and intervening loop. **a**, When the penultimate TMD is followed by a long (more than 50 amino acids) translocated loop, its translocation will be Sec61-dependent<sup>8</sup>. Hence, the loop will already be threaded through the Sec61 channel by the time of termination (left diagram). The terminal TMD inside the ribosome will then necessarily enter Sec61 (middle diagram), from where it presumably accesses the membrane through Sec61's lateral gate. **b**, When the final two TMDs are closely spaced, neither one can be inserted co-translationally because the tether downstream of the penultimate TMD is too short to engage Sec61 at the time of termination (left diagram). Therefore, both TMDs will be released and inserted post-translationally by an unknown mechanism. **c**, When the final TMD is located near a translocated C terminus, the TMD and tail are mostly or entirely inside the ribosome at the time of termination (left diagram). Their post-translational insertion mechanism is also unknown. Note that the machinery involved in the insertion and chaperoning of earlier TMDs is not shown for simplicity.

concomitant with translocation of the intervening loop. In the second situation, the terminal TMD would also need to be inserted post-translationally concomitant with translocation of the C-terminal tail (Fig. 1c). The factors that mediate these post-translational insertion reactions are not known.

This problem is exemplified by the large and important family of Cys-loop pentameric ion channels. This family includes acetylcholine receptors,  $\gamma$ -aminobutyric acid type A (GABA<sub>A</sub>) receptors, glycine receptors and others<sup>24</sup>. They play crucial roles in neurotransmission, and their incorrect biogenesis would lead to complex neurologic consequences<sup>25</sup>. Each subunit is composed of four TMDs with both the N terminus and C terminus facing the extracellular environment. The last TMD is followed by a tail that is typically only  $\sim 10$ – $20$  amino acids. This means that TMD4 will be mostly or entirely within the ribosome when translation terminates. The preceding three TMDs would have already been inserted given the  $\sim 100$  amino acid long cytosolic loop between TMD3 and TMD4. Hence, TMD4 would be released from the ribosome and must be inserted post-translationally by the route shown in Fig. 1c. Given the exceptional importance of this class of proteins, we investigated how the final TMD of a GABA<sub>A</sub> receptor subunit is inserted as a model for the general problem of terminal TMD insertion outlined above.

## Results

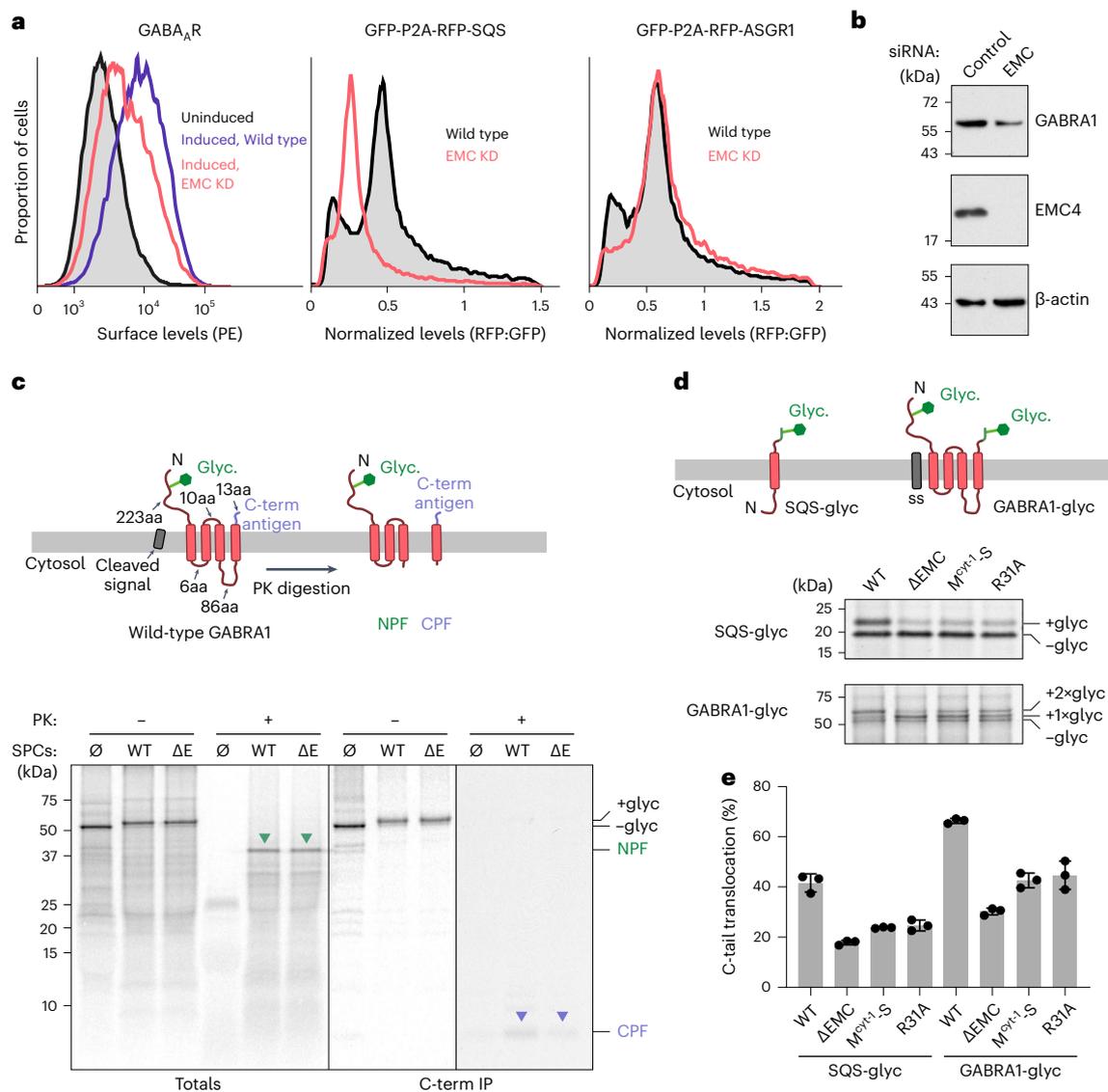
### The C-terminal TMD of GABRA1 is inserted by EMC

In considering factors that might mediate terminal TMD insertion of GABA<sub>A</sub> receptor subunits, we were intrigued by the earlier observation that the loss of EMC impairs the expression of multiple members of Cys-loop pentameric ion channels in worms<sup>26</sup>. As subunits of these channels use an N-terminal signal sequence for endoplasmic reticulum targeting and initiation of N-terminal translocation, it is now appreciated that insertion of the first TMD is not expected to be EMC-dependent<sup>4</sup>. TMD2 and TMD3 would then insert via the MPT given the short translocated loop between them<sup>8,10</sup>. This suggests that the EMC requirement might be at a later stage of insertion, folding or assembly. Among these possibilities, a potential role in the insertion of TMD4 was attractive because the biochemical reaction is similar to the post-translational insertion of tail-anchored proteins, which is an established role for EMC<sup>7,14</sup>.

To examine a potential role for EMC in Cys-loop channels, we focused on GABA<sub>A</sub> receptors, whose reliance on EMC has also been seen in mammals<sup>27</sup>. Using a previously characterized HEK293 cell line with stable inducible expression of a heteropentameric GABA<sub>A</sub> receptor<sup>28,29</sup>, we tested the effect of acute EMC depletion. As seen in earlier studies, induced surface expression of the GABA<sub>A</sub> receptor was reduced to less than  $\sim 50\%$  in cells knocked down for EMC4, which generally phenocopies the insertase deficiency seen with a loss of EMC2, EMC3, EMC5 or EMC6, other core subunits of EMC<sup>30–32</sup> (Fig. 2a). No effect was seen for a dual-color fluorescent reporter of the Sec61-inserted asialoglycoprotein receptor 1 (ASGR1), but a strong reduction was seen for a similar reporter of the known EMC-inserted tail-anchored protein squalene synthase (SQS). Immunoblotting for GABRA1, the  $\alpha 1$  subunit of the GABA<sub>A</sub> receptor, indicated that its lack of surface expression corresponded to its degradation, consistent with a failure in biogenesis (Fig. 2b). Importantly, cells lacking EMC contain the normal complement of all major factors involved in endoplasmic reticulum protein biogenesis (Extended Data Fig. 1 and ref. 4), consistent with normal biogenesis of most secretory pathway proteins<sup>4,14,31–33</sup>. This suggests that the effect of EMC on GABA<sub>A</sub> receptor biogenesis is unlikely to be indirect.

To examine whether the EMC-dependent step involves membrane insertion of a GABA<sub>A</sub> receptor, we reconstituted the insertion of GABRA1 in vitro. Translation of <sup>35</sup>S-labeled GABRA1 in reticulocyte lysate in the presence of semi-permeabilized cells (SPCs) resulted in signal sequence cleavage and glycosylation of the N-terminal domain in the endoplasmic reticulum lumen (Fig. 2c). Identical results were seen using SPCs from EMC knockout ( $\Delta$ EMC) cells, consistent with previous work showing that signal sequences use Sec61 to initiate translocation of the N-terminal domain independent of EMC<sup>4</sup>.

We used a protease protection assay to monitor TMD insertion of GABRA1 downstream of N-terminus translocation. Proteinase K cleaved

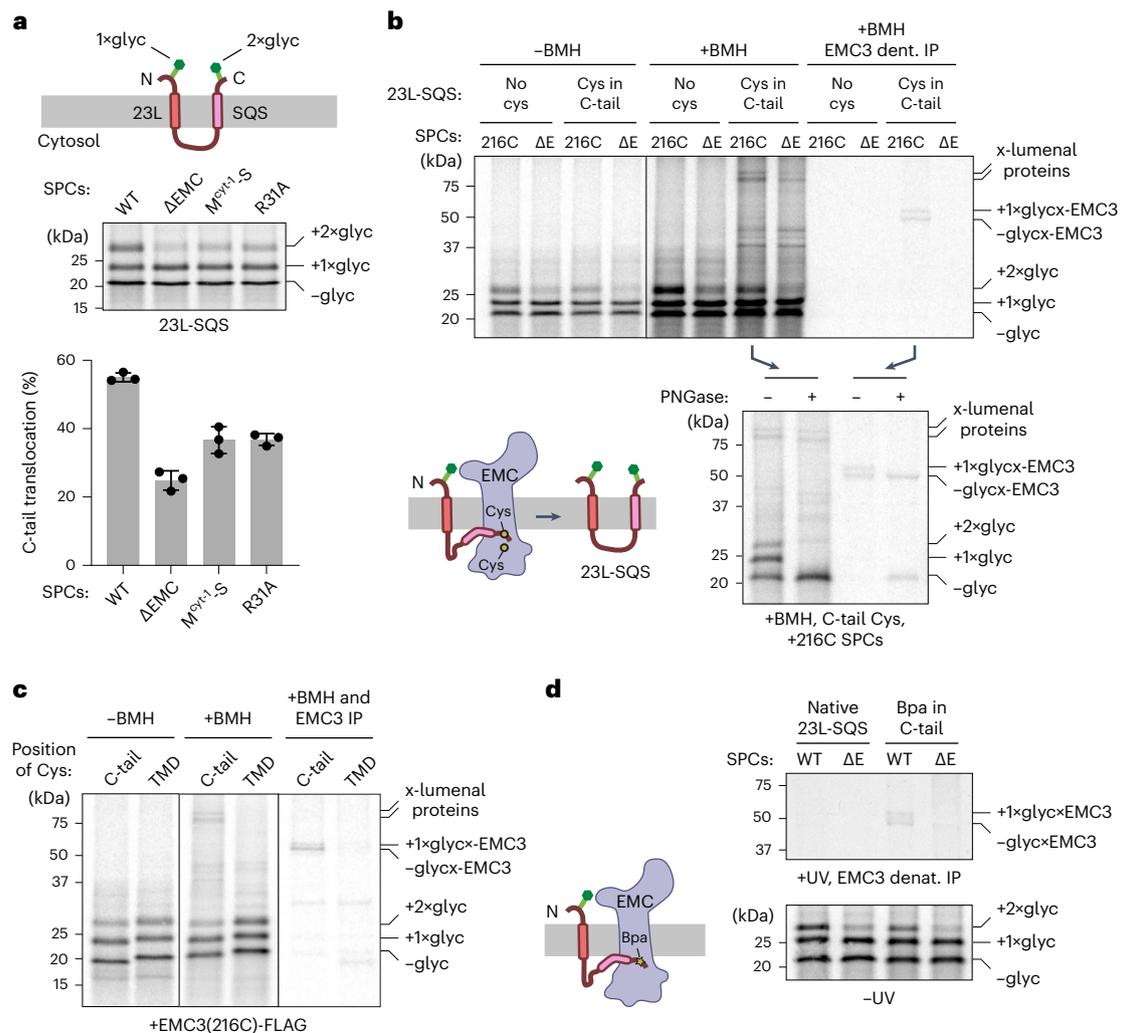


**Fig. 2 | EMC is required for C-terminal TMD insertion of GABRA1.** **a**, Stable cell lines expressing the indicated inducible constructs were treated with non-targeting or EMC4-targeting siRNAs for 3 days, induced for 6 h and analyzed by flow cytometry for surface GABA<sub>A</sub> receptor levels using phycoerythrin (PE) labeled antibody (left), total levels of RFP-SQS (middle) or total levels of RFP-ASGR1 (right). SQS and ASGR1 levels are normalized to GFP, an internal expression control that is separated from the reporter by a ribosome-skipping viral P2A sequence. KD indicates knockdown. **b**, Total levels of GABRA1 and EMC4 were analyzed by blotting in cells treated with control or EMC4 siRNA. β-actin was used as a loading control. **c**, Topology diagram of human GABRA1; the predicted fragments resulting from proteinase K (PK) digestion are shown on top. Tail and loop lengths are indicated. NPF, N-terminal protected fragment; CPF, C-terminal protected fragment. The bottom panel shows the topology analysis of GABRA1 in the endoplasmic reticulum of wild-type (WT) and EMC6-knockout (ΔE) 293 cells. <sup>35</sup>S-methionine-labeled GABRA1 was translated in rabbit reticulocyte lysate in the absence (∅) or presence of SPCs derived from WT or ΔE 293 cells. After

translation, the SPCs were recovered by centrifugation and analyzed directly (–PK) by SDS–PAGE and autoradiography or subjected to PK digestion (+PK). Reactions lacking SPCs were analyzed similarly without centrifugation. Aliquots of both –PK and +PK samples were subjected to immunoprecipitation via an antibody against the C terminus of GABRA1 (antigen labeled in purple). The glycosylated (+glyc) and non-glycosylated (–glyc) products are indicated. NPF and CPF are indicated with green and purple arrowheads, respectively. **d**, <sup>35</sup>S-methionine-labeled SQS and GABRA1 each with a C-terminal glycosylation site were translated in the presence of SPCs derived from ΔEMC cells or cells stably overexpressing either wild-type EMC3-FLAG or insertase-deficient mutants of EMC3-FLAG variants (M<sup>cyt1</sup>-S and R31A). SS indicates a cleavable signal sequence. **e**, Quantification of three independent experiments as shown in **d**, with mean ± s.d. plotted. C-tail translocation was quantified by plotting per cent glycosylation (for SQS) or the per cent of glycosylated products that contain two glycans (For GABRA1). C-term, C-terminal.

the long cytosolic loop between TMD3 and TMD4, generating two protected fragments: an N-terminal protected fragment corresponding to the first three TMDs of GABRA1 and a C-terminal protected fragment corresponding to TMD4 and a short translocated C-terminal tail (Fig. 2c). Wild-type and ΔEMC SPCs generate similar levels of N-terminal protected fragments after proteinase K digestion, indicating that EMC does not participate in the insertion of the first three TMDs of GABRA1.

By contrast, The C-terminal protected fragment (recovered using an antibody against the C-tail of GABRA1) is reduced in ΔEMC SPCs to less than half that seen in wild-type SPCs, suggesting an impairment in TMD4 insertion. To corroborate this conclusion, we introduced a glycosylation site into the GABRA1 C-tail (GABRA1-glyc) to monitor C-tail translocation as a proxy for TMD4 insertion (Fig. 2d). Glycosylation of the C-tail was impaired by more than 50% in the absence of



**Fig. 3 | Substrate C-tail samples EMC before translocation.** **a**, The 23L-SQS reporter (top) consists of a short translocated N-tail, a first TMD made of 23 leucine residues, an -100 amino acid cytosolic loop, a second TMD and flanking sequences from SQS and a short translocated C-tail. Both terminal tails have glycosylation sites to monitor translocation. <sup>35</sup>S-methionine labeled 23L-SQS was translated in the presence of SPCs derived from ΔEMC cells or cells expressing variants of EMC3 (WT, M<sup>91-1</sup>S and R31A). Products with different glycosylation states are indicated. Quantification of three independent experiments (mean ± s.d.) is plotted. C-tail translocation is determined as the per cent of glycosylated products that contains two glycans. **b**, <sup>35</sup>S-methionine-labeled 23L-SQS without or with a single cysteine within the C-tail, translated in the presence of SPCs derived from ΔEMC (ΔE) cells or cells expressing EMC3-216C (in which a single cysteine is introduced into cysteine-free EMC3 at residue 216 located in its cytosolic vestibule)<sup>6</sup>. One aliquot was analyzed directly (-BMH) and another was treated with BMH. Crosslinked samples were analyzed directly

(+BMH) or after EMC3 denaturing immunoprecipitation via FLAG tag (+BMH EMC3 dent. IP). The positions of 23L-SQS with zero, one or two glycans, and crosslinks between 23L-SQS and EMC3 or luminal proteins are indicated. Lanes containing EMC3 crosslinks were digested with PNGase F to confirm that the crosslinks contain either zero or one glycan and hence are not crosslinks to fully translocated double-glycosylated 23L-SQS. **c**, <sup>35</sup>S-methionine-labeled 23L-SQS containing a single cysteine within the C-tail or the SQS TMD, translated in the presence of EMC3-216C SPCs and analyzed similarly to **b**. **d**, <sup>35</sup>S-methionine-labeled 23L-SQS without or with a single Bpa incorporated within the C-tail (through an amber suppression system) was translated in the presence of SPCs derived from WT or ΔE cells. One aliquot was analyzed directly (-UV) and another was subjected to UV crosslinking. Crosslinked samples were analyzed after EMC3 denaturing immunoprecipitation via FLAG tag (+UV EMC3 dent. IP). Labeling is similar to **b**.

EMC. This level of impairment was similar to that seen for SQS-glyc, a well-established substrate for EMC-mediated insertion<sup>14</sup>.

As with SQS, insertion of GABRA1 TMD4 was not entirely eliminated in ΔEMC SPCs, presumably explaining why GABA<sub>A</sub> receptor surface expression is not completely eliminated in ΔEMC cells. We considered whether the residual GABRA1 TMD4 insertion in ΔEMC SPCs could potentially be mediated by the lateral gate of Sec61, the GEL complex or the GET complex, which inserts tail-anchored proteins of high hydrophobicity. Of these, the GET insertase seemed unlikely because later crosslinking experiments did not observe an interaction between a membrane-tethered C-terminal TMD and GET3, the targeting factor required for access to the GET insertase<sup>14,34,35</sup>. As the

Sec61 and GEL complexes are probably used for GABRA1 N-terminal translocation and insertion of the TMD2-TMD3 module, respectively, we devised a simplified reporter to test TMD4 insertion in isolation. This C-tail translocation reporter consists of an artificial signal-anchor comprising 23 leucine residues (23L) followed by a cytosolic loop, TMD4 of GABRA1 and a translocated C-tail (23L-GABRA1). As 23L can be inserted independently of EMC, GEL or Sec61, we can test the role of these factors in TMD4 insertion. Glycosylation sites in the N-terminal and C-terminal tails were used to monitor translocation.

As with full-length GABRA1, C-tail translocation of 23L-GABRA1 was EMC-dependent. Elimination of the GEL complex (by knockout of its TMCO1 subunit) or treatment with Apratoxin A (ApraA), a potent

inhibitor of Sec61's lateral gate<sup>36–38</sup>, had no effect on 23L-GABRA1 insertion (Extended Data Fig. 2). Combining one or both manipulations with the elimination of EMC showed no further impairment of C-tail translocation, suggesting that neither Sec61 nor GEL can contribute to TMD4 insertion. These results indicate that TMD4 insertion is primarily mediated by EMC, with residual EMC-independent insertion occurring unassisted or through an unknown insertase. Given the experimental and theoretical support for unassisted insertion<sup>39,40</sup>, a membrane-tethered TMD4 could readily access this insertion route. Nonetheless, such alternative mechanisms seem to be minor contributors relative to EMC.

EMC uses a cytosol-facing hydrophilic vestibule and membrane-embedded hydrophilic groove, both housed primarily in EMC3, to facilitate translocation of a flanking hydrophilic segment concomitant with TMD insertion<sup>7,19,21,23</sup>. To test whether this established mechanism was used for TMD4 insertion of GABRA1, we analyzed the effect of EMC3 mutations along the translocation route. In these experiments, ~70–90% of endogenous EMC3 is replaced by long-term stable overexpression of FLAG-tagged EMC3, with the excess EMC3 being effectively degraded by cellular quality control<sup>6</sup>. Mutation of EMC3 at either a cytosolic methionine-rich loop at the entry of the hydrophilic vestibule (M<sup>cyt-1</sup>-S) or a charged residue within the hydrophilic groove (R31A) impaired insertion of both SQS and TMD4 of GABRA1 (Fig. 2e). This result shows that EMC's insertase activity is involved in terminal TMD insertion, not some other part of EMC such as its putative chaperone surface<sup>41</sup>. We infer this conclusion because the degree of impairment in these mutants is similar for both SQS and TMD4 of GABRA1, and because these EMC mutants are fully assembled and intact<sup>23</sup>. Thus, the last TMD of GABRA1 is post-translationally inserted into the endoplasmic reticulum membrane through EMC using a similar mechanism as previously known substrates.

### Translocating C-tail samples EMC during insertion

The 23L-TMD strategy provided a simplified reporter to analyze substrate parameters that influence EMC-dependent terminal TMD insertion. However, GABRA1 TMD4 insertion efficiency in this reporter was relatively low (a point we will return to later), so we devised a more efficiently inserted construct containing the TMD of SQS (23L-SQS; Fig. 3a). The linker between 23L and SQS was of the same length (~100 amino acids) as the cytosolic loop preceding the final TMD of GABRA1, so the 23L domain would have been inserted and diffused away from the translocon by the time translation terminates and the terminal TMD emerges from the ribosome<sup>42</sup>. We confirmed that 23L-SQS retains EMC-dependent insertion of the SQS TMD, which was impaired by mutations in EMC's insertase path (Fig. 3a).

With this construct, we used site-specific chemical crosslinking to probe the local environment during C-terminal TMD insertion. A single cysteine in the C-tail of 23L-SQS crosslinked to a single cysteine engineered in the cytosolic vestibule of EMC at position 216 of EMC3 (Fig. 3b). EMC3 crosslinked to non-glycosylated and singly glycosylated substrate, but not to doubly glycosylated substrate. This indicates that EMC's cytosolic vestibule can crosslink to a substrate whose C-terminal TMD has yet to be inserted but whose N-terminal TMD has already been inserted (and is glycosylated). Importantly, a cysteine in the middle of the adjacent SQS TMD showed almost no crosslinking to EMC3 (Fig. 3c). This provides a specificity control for the observed tail-mediated crosslinks and supports a model in which EMC mediates translocation of the hydrophilic tail through its hydrophilic vestibule.

In further support of this idea, two independent EMC3 mutants that partially impair translocation led to a ~1.9-fold increase in crosslinks between the C-tail and EMC's cytosolic vestibule (Extended Data Fig. 3). This observation is consistent with a longer residence time at this pre-translocation step when the translocation reaction is impaired, similar to earlier observations for an N-terminal EMC-dependent TMD<sup>6</sup>. Although less efficient, a UV-activated photo-crosslinker in the C-tail of

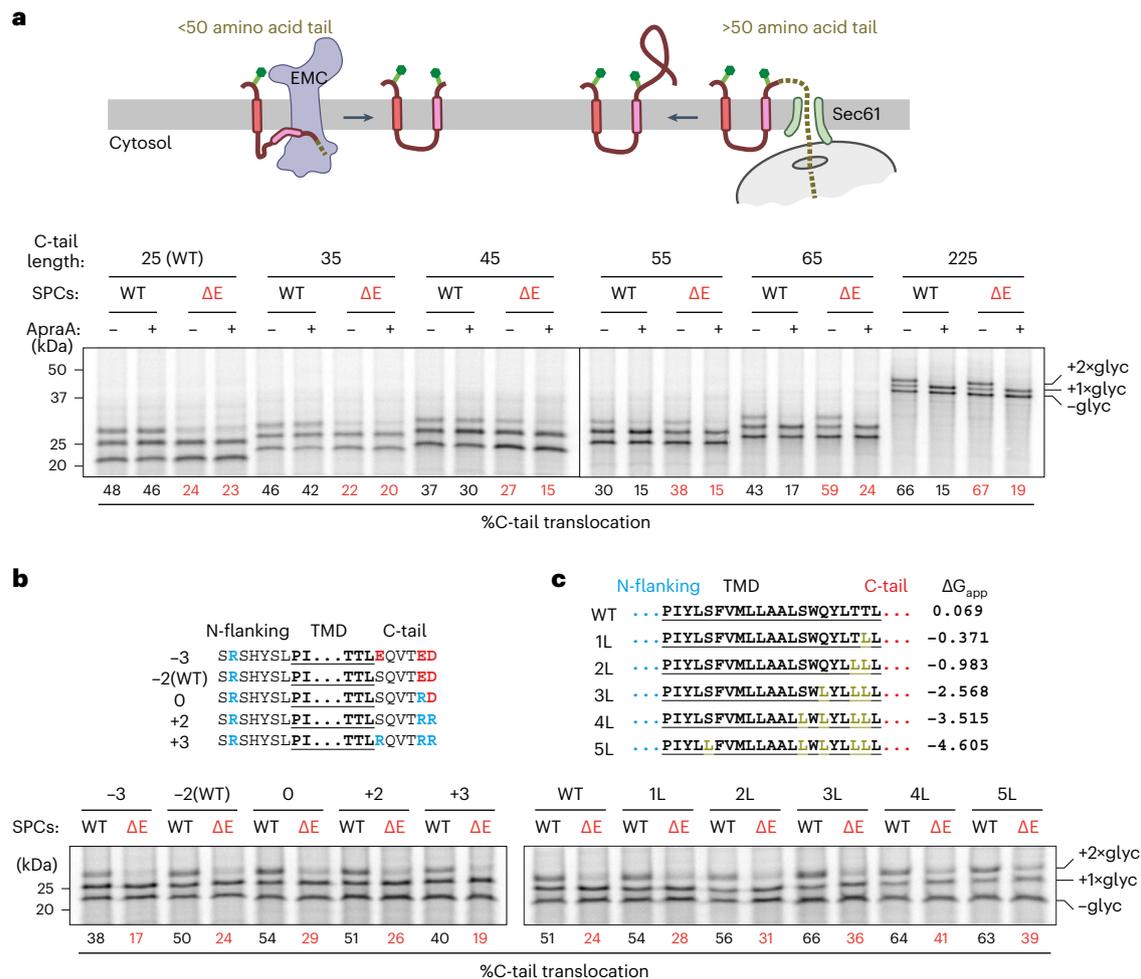
23L-SQS crosslinked to EMC3 in wild-type EMC (Fig. 3d). These results indicate that an incompletely inserted membrane protein encounters EMC, presumably by diffusional sampling within the membrane, and uses EMC's hydrophilic vestibule for translocation of the C-terminal tail concomitant with TMD insertion. In the absence of a functional EMC insertase, the C-terminal TMD presumably binds rapidly to the nearby membrane surface<sup>40</sup> but evidently cannot be inserted as efficiently by an unassisted mechanism or by another insertion factor (Extended Data Fig. 2). Binding to the membrane surface might explain why the TMD does not show any obvious increase in crosslinking to cytosolic chaperones or targeting factors in the absence of EMC as might otherwise be expected for a TMD exposed to the cytosol<sup>14,43–45</sup>.

### Determinants of EMC-mediated C-terminal TMD insertion

The 23L-SQS reporter was then modified to test the substrate features that influence terminal TMD insertion. In the first series of experiments, we found that extending the C-tail shifted the insertion pathway from EMC to Sec61 (Fig. 4a). At lengths of 25 and 35 amino acids, the TMD is insufficiently exposed outside the ribosome to engage the Sec61 lateral gate in the hairpin topology required for C-tail translocation. These constructs were, therefore, unaffected by the Sec61 inhibitor ApraA and were sensitive to the loss of EMC. By contrast, at 55 amino acids or longer, C-tail translocation was unaffected in  $\Delta$ EMC SPCs but completely inhibited by ApraA. The 45 amino acid tail showed an intermediate effect, being partially sensitive to both EMC loss and Sec61 inhibition. Thus, EMC and Sec61 are largely non-redundant in terminal TMD insertion. The critical switchover point between EMC-dependence and Sec61-dependence is ~45 amino acids. This length is close to the translocation limit for the Oxa1 family (of which EMC is a member<sup>46,47</sup>) but long enough for the TMD to reach Sec61's lateral gate in the appropriate looped topology<sup>18</sup>.

Charged residues in the C-terminal tail of 23L-SQS were seen to modestly but clearly reduce tail translocation (Fig. 4b). No difference was seen for tails with net positive or net negative charges. Importantly, all of these charge variants were translocated in an EMC-dependent manner. The charge-imperviousness for C-tail translocation of 23L-SQS was unexpected because EMC-mediated insertion of an N-terminal TMD (that is, a signal-anchor) or a tail-anchored protein is selectively disfavored if the tail to be translocated contains multiple positive charges<sup>4,6,48</sup>. The basis of this different behavior may relate to the length of time available for EMC-mediated insertion for a signal-anchor or tail-anchor versus the terminal TMD of a multipass protein. A ribosome displaying a signal-anchor has limited time for EMC-mediated insertion before docking onto Sec61, at which point EMC becomes inaccessible. Similarly, moderately hydrophobic tail-anchored proteins that are not promptly inserted by EMC can instead be inserted into mitochondria. Consistent with this interpretation, insertion of such signal-anchors and tail-anchors improves when Sec61 is depleted or when mitochondrial targeting is impaired, respectively<sup>6,49</sup>. By contrast, a multipass protein is already committed to endoplasmic reticulum insertion by the time EMC must insert the final tethered TMD, which would have prolonged and repeated access to EMC. Thus, EMC's substrate preference against positive charge translocation can be overcome by simply providing more time, implying that it can accommodate a broader range of substrates as a terminal TMD insertase.

A similar rationale probably explains the finding that in the context of 23L-SQS, EMC is able to accommodate a broad range of TMD hydrophobicity (Fig. 4c). For tail-anchored protein insertion, EMC-dependence is strongly influenced by hydrophobicity, with TMDs of higher hydrophobicity becoming dependent on the GET pathway. For example, SQS can progressively be made EMC-independent and GET-dependent by replacing less-hydrophobic residues in the TMD with leucine residues<sup>14</sup>. The same changes in 23L-SQS not only did not shift EMC-dependence but improved insertion overall. This indicates that EMC's preference for low-hydrophobicity TMDs seen for



**Fig. 4 | Determinants of substrate C-tail translocation.** **a**,  $^{35}\text{S}$ -methionine-labeled 23L-SQS variants with the different C-tail lengths indicated (top) translated in the presence of SPCs derived from WT or  $\Delta\text{E}$  cells. Where indicated,  $2\ \mu\text{M}$  of the Sec61 inhibitor ApraA was included in the translation reaction. Translation reactions were analyzed directly, and substrates with zero, one or two glycans attached are indicated. The two-glycan product is indicative of successful C-tail translocation. C-tail translocation was quantified by calculating the percentage of all glycosylated

products that contain two glycans. **b**,  $^{35}\text{S}$ -methionine-labeled 23L-SQS variants with charged residue mutations in the C-tail flanking the TMD were analyzed as in **a**. The sequences of the variants are shown with the net C-tail flanking charge indicated. **c**,  $^{35}\text{S}$ -methionine-labeled 23L-SQS variants with leucine mutations in the TMD were analyzed as in **a**. The sequences of the variants are shown along with the hydrophobicity of each TMD as a  $\Delta G_{app}$  score<sup>65</sup>, in which negative values indicate favored membrane insertion.

tail-anchored proteins is not a reflection of EMC limitations but rather a consequence of high-hydrophobicity TMDs being captured by the GET pathway in the cytosol. Considered together, these findings illustrate that EMC's substrate specificity for terminal TMDs of multipass proteins is broad with respect to flanking charge and hydrophobicity but is limited by tail lengths shorter than ~50 amino acids.

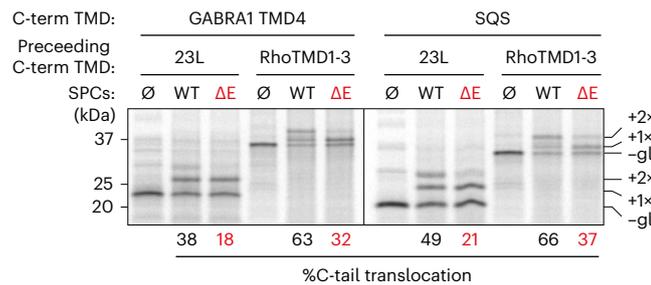
### Immature membrane domains facilitate EMC targeting

As noted above, 23L-GABRA1 shows EMC-dependent insertion of the terminal TMD (Extended Data Fig. 2) but with clearly lower efficiency than that seen in native GABRA1 (~38% compared to ~65%; Fig. 5). In considering possible explanations, we recognized that EMC has been proposed to function as a chaperone in addition to its insertase activity<sup>33,41</sup>. This suggested the possibility that in native GABRA1, TMD4 is brought in proximity to EMC by its preceding TMDs engaging EMC through its putative chaperone function. Replacing the first three TMDs of GABRA1 with 23L would eliminate this 'targeting' activity, explaining the loss of TMD4 insertion efficiency.

As a test of this idea, we asked whether TMD4 insertion could be rescued if TMD4 were preceded by an unrelated membrane domain that

might be a putative chaperone substrate. Earlier studies have shown that until insertion of all of its TMDs, G-protein coupled receptors are targets for intramembrane chaperoning<sup>50</sup>. Therefore, we preceded TMD4 of GABRA1 with the first three TMDs of rhodopsin (termed Rho(1-3)) and tested the efficiency of terminal TMD insertion. To ensure that Rho(1-3) was not an EMC insertion substrate, we initiated its translocation using an N-terminal signal peptide and luminal domain preceding its first TMD. Relative to 23L-GABRA1, terminal TMD insertion of Rho(1-3)-GABRA1 was markedly higher and similar to that seen with native GABRA1 (Fig. 5).

As Rho(1-3) is unrelated to TMD1-3 of GABRA1, these results indicate that the efficient insertion of TMD4 is not due to its complementarity with the earlier TMDs. Instead, the findings support a model in which the earlier TMDs target an otherwise poorly inserted TMD4 to EMC for efficient insertion. Consistent with this idea, the insertion of the SQS TMD improved when preceded by Rho(1-3) relative to 23L-SQS (Fig. 5), which itself was slightly more efficient than SQS inserted as a tail-anchored protein. This suggests that EMC-mediated insertion improves slightly by tethering a TMD close to the membrane (for example, with 23L) and improves further if



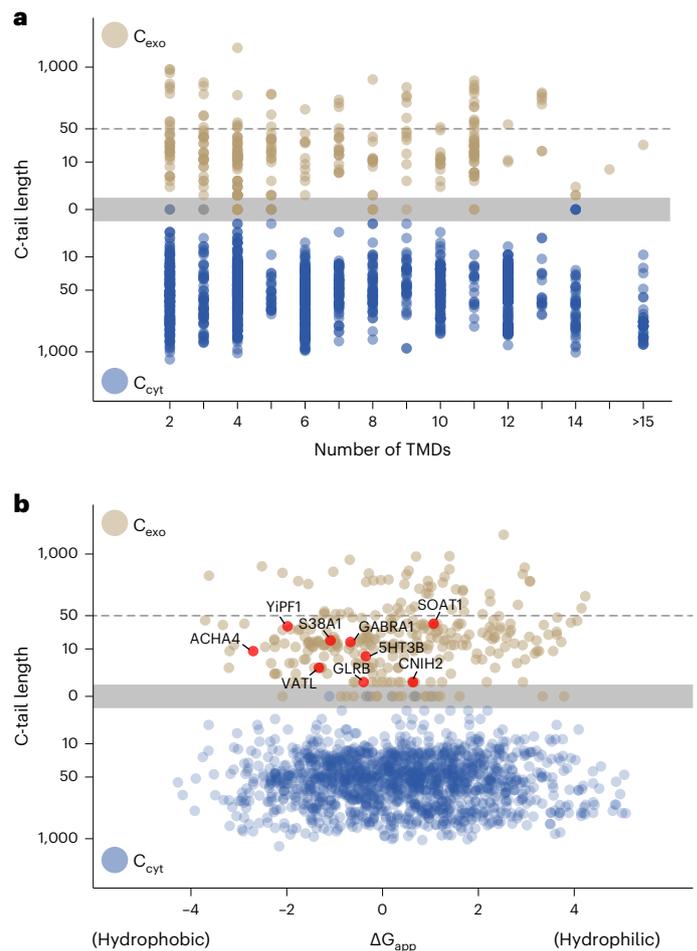
**Fig. 5 | Targeting to EMC facilitates C-terminal TMD insertion.** The GABRA1 TMD4 or SQS TMD was preceded by either the 23L TMD or TMD1–3 of rhodopsin, separated by a 100 amino acid linker. These constructs were translated in the absence ( $\emptyset$ ) or presence of SPCs from WT or  $\Delta E$  cells and analyzed by SDS–PAGE and autoradiography. Substrates with zero, one or two glycans attached are labeled. Per cent C-tail translocation, calculated as in Fig. 4, is indicated.

other parts of the protein are immature (for example, the first three TMDs of Rho or GABRA1).

### An expanded substrate repertoire of EMC

Using the insights gained from GABRA1 and 23L–SQS, we sought to predict all other terminal TMD substrates of EMC in the human genome. Leveraging the structural prediction afforded by AlphaFold2 (ref. 51) combined with the positive-inside rule<sup>52</sup>, we manually curated the orientation of all 1,784 annotated multipass endoplasmic reticulum membrane proteins in the human genome. By plotting the C-tail lengths by topology, number of total TMDs and hydrophobicity of the terminal TMD (Fig. 6a,b), we found that 244 multipass proteins contain a terminal TMD whose downstream tail is non-cytosolic and 50 amino acids or shorter (Supplementary Table 1). From this manually curated list, we chose for analysis six proteins containing terminal TMDs of varying hydrophobicity and C-tail lengths. The functional expression of five of these proteins, or their close homologs, have been observed in earlier studies to be dependent on EMC<sup>26,27,31,33</sup>. Therefore, we tested their terminal TMDs in the 23L–TMD reporter with a C-terminal glycosylation site (Extended Data Fig. 4) to determine whether impaired C-tail translocation could explain their EMC dependence. Each of these showed at least partial dependence on EMC for C-terminal TMD insertion (Fig. 7a,b). The variable levels of insertion in the absence of EMC among the different substrates might reflect their capacity for unassisted insertion. Alternatively, they might be able to access other insertases such as GET, Sec61 or GEL depending on either TMD hydrophobicity or length of the C-tail, but this remains to be investigated.

We also tested two unrelated native proteins, SOAT1 and YIPF1, that also contain a terminal TMD with a translocated C-tail. These were chosen based on the prediction that earlier steps in their insertion would be EMC-independent, thereby allowing us to monitor terminal TMD insertion using a single C-tail glycosylation site (see Fig. 7a). As per our predictions (Fig. 6), both were observed to be partially EMC-dependent for C-tail translocation (Fig. 7c). As the separation of glycosylated from non-glycosylated products of full-length SOAT1 was challenging, we verified our conclusion with a better-resolved SOAT1 construct lacking the N-terminal cytosolic domain. EMC-dependent terminal TMD translocation of full-length and N-terminally deleted SOAT1 was further validated by selective recovery of the glycosylated products using the lectin concanavalin A. Of note, both endogenous and exogenously expressed SOAT1 were shown in earlier studies to be strongly dependent on EMC in cells, but the basis of this dependence was not clear<sup>32</sup>. Our finding that insertion of the final TMD of SOAT1 is EMC-dependent to a comparable level as SOAT1 biogenesis in cells now provides an explanation. The ~250 new putative substrates of EMC are highly diverse in their topology and function and may help to explain

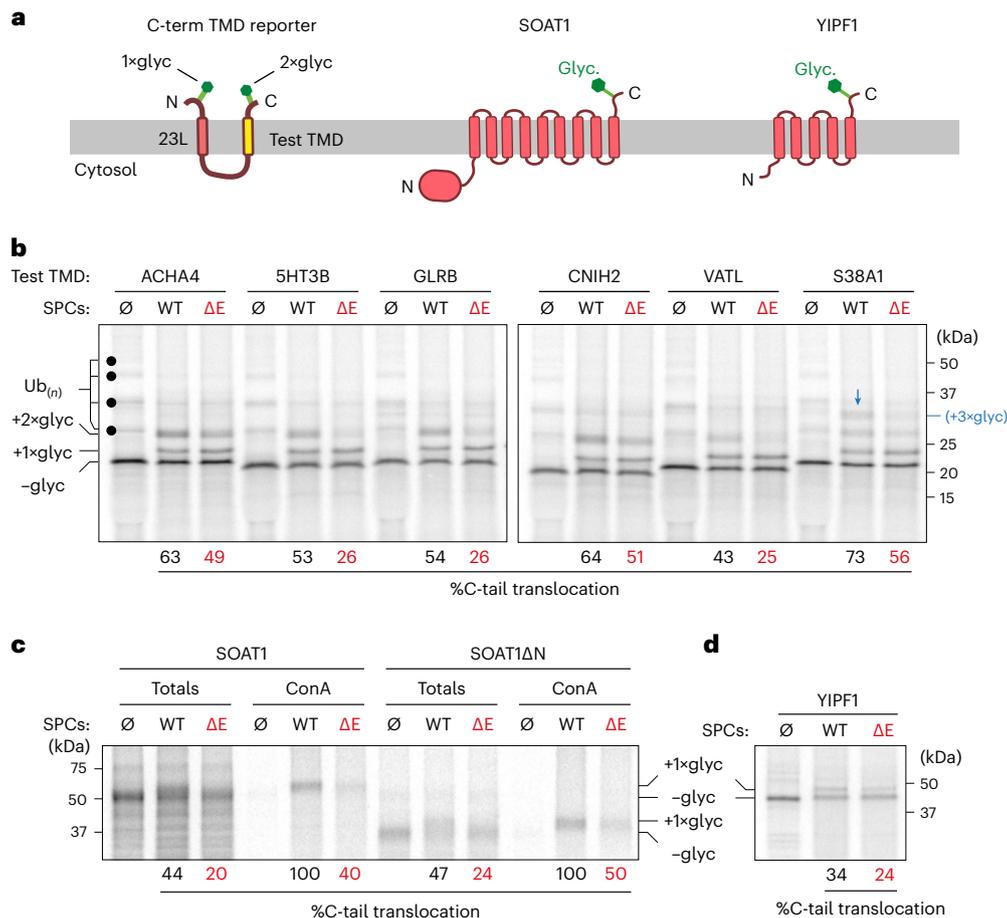


**Fig. 6 | Systematic analysis of multipass membrane protein orientation.** **a**, The topology of all annotated multipass proteins in the human genome (1,784 total), determined using a combination of AlphaFold2 structure prediction to identify TMDs and the positive-inside rule to deduce orientation (see Supplementary Table 1). The C-tail locations ( $C_{\text{exo}}$  or  $C_{\text{cyt}}$  for exoplasmic and cytosolic, respectively) and lengths were tabulated and plotted versus the numbers of TMDs in the protein. **b**, The data from **a** were re-plotted relative to the  $\Delta G_{\text{app}}$  score of the final TMD. Representative substrates analyzed in Fig. 7 for EMC-dependent translocation of their C-terminal TMDs are highlighted in red.

the complex and pleiotropic phenotypes of EMC loss in a wide range of organisms<sup>7,53</sup>.

## Discussion

Our study provides three insights into the problem of membrane protein topogenesis. First, we reveal that unlike previous suggestions<sup>4,14,33</sup>, EMC is not specific for poorly hydrophilic TMDs and does not have a strong discriminatory capacity against the translocation of positive charges. Instead, it seems that EMC's preferences against positive charges and high hydrophobicity that were observed in earlier studies are consequences of competing reactions. Thus, EMC's intrinsic capacity for TMD insertion is broader than had been thought. Second, various substrates that show an EMC requirement in cells, such as pentameric ion channels<sup>26,27</sup> and SOAT1<sup>32</sup>, can probably be ascribed to a failure of terminal TMD insertion. These substrates had been puzzling because they are neither tail-anchored proteins nor initiated with a signal-anchor in the  $N_{\text{exo}}$  topology (in which the N terminus faces the exoplasmic side of the membrane), which previously were the known targets for EMC function. Although we cannot exclude additional roles for EMC



**Fig. 7 | EMC facilitates C-terminal TMD insertion of distinct classes of substrates. a**, Domain organization of a C-terminal TMD reporter, which contains 23L and test TMD (yellow), a cytosolic loop (~100 amino acids) and two glycosylation sites in both N-tail and C-tail (left); topology and domains of full-length SOAT1 (middle) and YIPF1 (right) containing C-terminal glycosylation sites. **b**, <sup>35</sup>S-methionine-labeled C-terminal TMD reporters with indicated TMDs, translated in the absence (∅) or presence of SPCs from WT or ΔE cells. Translation reactions were analyzed directly by SDS-PAGE and autoradiography. The positions of translated reporters with zero, one or two glycans are indicated. In the absence of endoplasmic reticulum membranes, cytosolic quality control factors recognize non-translocated substrates and modify them with mono- or poly-ubiquitin (Ub<sub>(n)</sub>) as indicated by the black dots. Note that S38A1 contains two

C-tail glycosylation sites, resulting in a triply glycosylated product (blue arrow). C-tail translocation was quantified as previous figures. **c**, <sup>35</sup>S-methionine-labeled SOAT1 or an N-terminal domain deletion of SOAT1 (SOAT1ΔN), translated in the absence (∅) or presence of SPCs from WT or ΔE cells. One aliquot was subject to denaturing immunoprecipitation through an N-terminal HA tag (Totals); another was subjected to concanavalin A pull-down (ConA) to recover the glycosylated population. Non-glycosylated and glycosylated products are indicated, and percent glycosylation was calculated. In ConA recovered lanes, glycosylation was normalized to WT. **d**, <sup>35</sup>S-methionine-labeled YIPF1 was translated in the absence (∅) or presence of SPCs from WT or ΔE cells and analyzed directly by SDS-PAGE and autoradiography. Substrate glycosylation state and per cent C-tail translocation are indicated.

in GABRA1 or SOAT1, the magnitude of the defect seen in C-terminal TMD insertion can explain the consequence in cells.

Third, and perhaps of most conceptual importance, is the finding that the topology of multipass proteins can be determined by a combination of co-translational and post-translational reactions. Co-translational events result in committing the protein for endoplasmic reticulum targeting and insertion of most TMDs, whereas other TMDs can be inserted post-translationally after the substrate has presumably departed the ribosome-translocon complex. Diffusion away from the ribosome (or dissociation of the ribosome from the membrane) would be a prerequisite for EMC-mediated insertion because EMC cannot access substrates near the exit tunnel of a Sec61-bound ribosome<sup>6,22</sup>. Hence, our findings show that at least some multipass proteins are released from the ribosome-translocon complex in a topologically immature form, with EMC rectifying the topology at a later step. Targeting to EMC for this post-translational rectification step may be facilitated by EMC's putative chaperone activity<sup>33,41</sup>, which would preferentially engage immature membrane-embedded

domains such as those lacking a terminal TMD. In cells, a failure to rectify the topology promptly would presumably lead to engagement by quality control factors within the membrane of the topologically incomplete protein<sup>54,55</sup> or in the cytosol of the uninserted TMD<sup>45,56,57</sup>, thereby explaining why these EMC substrates are degraded in ΔEMC cells.

Although we have demonstrated post-translational topological rectification for C-terminal TMDs, it is plausible that in other circumstances, pairs of TMDs separated by a short loop are similarly inserted post-translationally (for example, Fig. 1b). The Oxa1 family (which, in eukaryotes, is composed of EMC, GEL and GET complexes) has been shown to be capable of such insertion<sup>8,10,58-60</sup>, which might be needed if some TMDs are skipped during co-translational insertion. Skipping of TMDs followed by post-translational rearrangement has been suggested in earlier work<sup>61</sup>, but the basis of such a mechanism is unclear. Our findings now suggest a plausible mechanism utilizing an Oxa1 family member, which is an idea that warrants experimental analysis in future work.

Our study adds to the emerging principle that there is a straightforward segregation of function between the Oxa1 family and the SecY family<sup>3,9</sup>. Oxa1 family members mediate TMD insertion when the translocated segment of a flanking polypeptide is shorter than ~50 amino acids, whereas Sec61 (or SecY in prokaryotes) mediates TMD insertion when the translocated flanking domain is longer. In eukaryotes, tail-anchored proteins are inserted by the GET and EMC insertases<sup>14,62,63</sup>, signal-anchored proteins with a short translocated N-terminal tail are inserted by EMC<sup>4,6</sup>, internal TMD pairs with short loops are inserted by GEL<sup>8,10</sup> and, as shown here, terminal TMDs with a short translocated C-tail use EMC. In bacteria, the sole Oxa1 family member YidC would perform all of these jobs, perhaps explaining why it is essential<sup>64</sup>. The larger and more diverse membrane proteome in eukaryotes might have driven an expansion and specialization of endoplasmic-reticulum-localized Oxa1 family members, which perhaps also affords a degree of redundancy and robustness to the essential process of membrane protein biogenesis.

### Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41594-023-01120-6>.

### References

- Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
- von Heijne, G. The membrane protein universe: what's out there and why bother? *J. Intern. Med.* **261**, 543–557 (2007).
- Hegde, R. S. & Keenan, R. J. The mechanisms of integral membrane protein biogenesis. *Nat. Rev. Mol. Cell Biol.* **23**, 107–124 (2022).
- Chitwood, P. J., Juszkiwicz, S., Guna, A., Shao, S. & Hegde, R. S. EMC is required to initiate accurate membrane protein topogenesis. *Cell* **175**, 1507–1519.e16 (2018).
- O'Keefe, S. et al. An alternative pathway for membrane protein biogenesis at the endoplasmic reticulum. *Commun. Biol.* **4**, 828 (2021).
- Wu, H. & Hegde, R. S. Mechanism of signal-anchor triage during early steps of membrane protein insertion. *Mol. Cell* **83**, 961–973.e7 (2023).
- Hegde, R. S. The function, structure, and origins of the ER membrane protein complex. *Annu. Rev. Biochem.* **91**, 651–678 (2022).
- Smalinskaitė, L., Kim, M. K., Lewis, A. J. O., Keenan, R. J. & Hegde, R. S. Mechanism of an intramembrane chaperone for multipass membrane proteins. *Nature* **611**, 161–166 (2022).
- Smalinskaitė, L. & Hegde, R. S. The biogenesis of multipass membrane proteins. *Cold Spring Harb. Perspect. Biol.* **15**, a041251 (2022).
- Sundaram, A. et al. Substrate-driven assembly of a translocon for multipass membrane proteins. *Nature* **611**, 167–172 (2022).
- Anghelescu, S. A., McGilvray, P. T., Hegde, R. S. & Keenan, R. J. Identification of Oxa1 homologs operating in the eukaryotic endoplasmic reticulum. *Cell Rep.* **21**, 3708–3716 (2017).
- Lewis, A. J. O. & Hegde, R. S. A unified evolutionary origin for the ubiquitous protein transporters SecY and YidC. *BMC Biol.* **19**, 266 (2021).
- McDowell, M. A., Heimes, M. & Sinning, I. Structural and molecular mechanisms for membrane protein biogenesis by the Oxa1 superfamily. *Nat. Struct. Mol. Biol.* **28**, 234–239 (2021).
- 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).
- Li, L. et al. Crystal structure of a substrate-engaged SecY protein-translocation channel. *Nature* **531**, 395–399 (2016).
- Rapoport, T. A., Li, L. & Park, E. Structural and mechanistic insights into protein translocation. *Annu. Rev. Cell Dev. Biol.* **33**, 369–390 (2017).
- Voorhees, R. M. & Hegde, R. S. Toward a structural understanding of co-translational protein translocation. *Curr. Opin. Cell Biol.* **41**, 91–99 (2016).
- Voorhees, R. M. & Hegde, R. S. Structure of the Sec61 channel opened by a signal sequence. *Science* **351**, 88–91 (2016).
- 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).
- McGilvray, P. T. et al. An ER translocon for multi-pass membrane protein biogenesis. *eLife* **9**, e56889 (2020).
- Miller-Vedam, L. E. et al. Structural and mechanistic basis of the EMC-dependent biogenesis of distinct transmembrane clients. *eLife* **9**, e62611 (2020).
- O'Donnell, J. P. et al. The architecture of EMC reveals a path for membrane protein insertion. *eLife* **9**, e57887 (2020).
- Pleiner, T. et al. Structural basis for membrane insertion by the human ER membrane protein complex. *Science* **369**, 433–436 (2020).
- Lynagh, T. & Pless, S. A. Principles of agonist recognition in Cys-loop receptors. *Front. Physiol.* **5**, 160 (2014).
- Sine, S. M. & Engel, A. G. Recent advances in Cys-loop receptor structure and function. *Nature* **440**, 448–455 (2006).
- Richard, M., Boulin, 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.* **110**, E1055–E1063 (2013).
- Whittsette, A. L., Wang, Y.-J. & Mu, T.-W. The endoplasmic reticulum membrane complex promotes proteostasis of GABA<sub>A</sub> receptors. *iScience* **25**, 104754 (2022).
- Lavery, D. et al. Cryo-EM structure of the human  $\alpha 1\beta 3\gamma 2$  GABA<sub>A</sub> receptor in a lipid bilayer. *Nature* **565**, 516–520 (2019).
- Dostalova, Z. et al. Human  $\alpha 1\beta 3\gamma 2L$  gamma-aminobutyric acid type A receptors: high-level production and purification in a functional state. *Protein Sci.* **23**, 157–166 (2014).
- Coukos, R. et al. An engineered transcriptional reporter of protein localization identifies regulators of mitochondrial and ER membrane protein trafficking in high-throughput CRISPRi screens. *eLife* **10**, e69142 (2021).
- Tian, S. et al. Proteomic analysis identifies membrane proteins dependent on the ER membrane protein complex. *Cell Rep.* **28**, 2517–2526.e5 (2019).
- Volkmar, N. et al. The ER membrane protein complex promotes biogenesis of sterol-related enzymes maintaining cholesterol homeostasis. *J. Cell Sci.* **132**, jcs223453 (2019).
- Shurtleff, M. J. et al. The ER membrane protein complex interacts cotranslationally to enable biogenesis of multipass membrane proteins. *eLife* **7**, e37018 (2018).
- Güngör, B., Flohr, T., Garg, S. G. & Herrmann, J. M. The ER membrane complex (EMC) can functionally replace the Oxa1 insertase in mitochondria. *PLoS Biol.* **20**, e3001380 (2022).
- McDowell, M. A. et al. Structural basis of tail-anchored membrane protein biogenesis by the GET insertase complex. *Mol. Cell* **80**, 72–86.e7 (2020).
- Itskanov, S. et al. A common mechanism of Sec61 translocon inhibition by small molecules. *Nat. Chem. Biol.* **19**, 1063–1071 (2023).

37. Paatero, A. O. et al. Apratoxin kills cells by direct blockade of the Sec61 protein translocation channel. *Cell Chem. Biol.* **23**, 561–566 (2016).
38. Liu, Y., Law, B. K. & Luesch, H. Apratoxin A reversibly inhibits the secretory pathway by preventing cotranslational translocation. *Mol. Pharmacol.* **76**, 91–104 (2009).
39. Brambillasca, S., Yabal, M., Makarow, M. & Borgese, N. Unassisted translocation of large polypeptide domains across phospholipid bilayers. *J. Cell Biol.* **175**, 767–777 (2006).
40. White, S. H. & Wimley, W. C. Membrane protein folding and stability: physical principles. *Annu. Rev. Biophys. Biomol. Struct.* **28**, 319–365 (1999).
41. Chen, Z. et al. EMC chaperone–Ca<sub>v</sub> structure reveals an ion channel assembly intermediate. *Nature* **619**, 410–419 (2023).
42. Heinrich, S. U., Mothes, W., Brunner, J. & Rapoport, T. A. The Sec61p complex mediates the integration of a membrane protein by allowing lipid partitioning of the transmembrane domain. *Cell* **102**, 233–244 (2000).
43. Stefanovic, S. & Hegde, R. S. Identification of a targeting factor for posttranslational membrane protein insertion into the ER. *Cell* **128**, 1147–1159 (2007).
44. Mariappan, M. et al. A ribosome-associating factor chaperones tail-anchored membrane proteins. *Nature* **466**, 1120–1124 (2010).
45. Itakura, E. et al. Ubiquitins chaperone and triage mitochondrial membrane proteins for degradation. *Mol. Cell* **63**, 21–33 (2016).
46. Kumazaki, K. et al. Structural basis of Sec-independent membrane protein insertion by YidC. *Nature* **509**, 516–520 (2014).
47. Shanmugam, S. K. et al. New insights into amino-terminal translocation as revealed by the use of YidC and Sec depletion strains. *J. Mol. Biol.* **431**, 1025–1037 (2019).
48. Pleiner, T. et al. A selectivity filter in the ER membrane protein complex limits protein misinsertion at the ER. *J. Cell Biol.* **222**, e202212007 (2023).
49. Guna, A. et al. MTCH2 is a mitochondrial outer membrane protein insertase. *Science* **378**, 317–322 (2022).
50. Chitwood, P. J. & Hegde, R. S. An intramembrane chaperone complex facilitates membrane protein biogenesis. *Nature* **584**, 630–634 (2020).
51. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
52. Beltzer, J. P. et al. Charged residues are major determinants of the transmembrane orientation of a signal-anchor sequence. *J. Biol. Chem.* **266**, 973–978 (1991).
53. Volkmar, N. & Christianson, J. C. Squaring the EMC—how promoting membrane protein biogenesis impacts cellular functions and organismal homeostasis. *J. Cell Sci.* **133**, jcs243519 (2020).
54. Sato, B. K., Schulz, D., Do, P. H. & Hampton, R. Y. Misfolded membrane proteins are specifically recognized by the transmembrane domain of the Hrd1p ubiquitin ligase. *Mol. Cell* **34**, 212–222 (2009).
55. Wu, X. & Rapoport, T. A. Mechanistic insights into ER-associated protein degradation. *Curr. Opin. Cell Biol.* **53**, 22–28 (2018).
56. Guna, A. & Hegde, R. S. Transmembrane domain recognition during membrane protein biogenesis and quality control. *Curr. Biol.* **28**, R498–R511 (2018).
57. Hessa, T. et al. Protein targeting and degradation are coupled for elimination of mislocalized proteins. *Nature* **475**, 394–397 (2011).
58. Facey, S. J., Neugebauer, S. A., Krauss, S. & Kuhn, A. The mechanosensitive channel protein MscL is targeted by the SRP to the novel YidC membrane insertion pathway of *Escherichia coli*. *J. Mol. Biol.* **365**, 995–1004 (2007).
59. Hell, K., Neupert, W. & Stuart, R. A. Oxa1p acts as a general membrane insertion machinery for proteins encoded by mitochondrial DNA. *EMBO J.* **20**, 1281–1288 (2001).
60. van der Laan, M., Bechtluft, P., Kol, S., Nouwen, N. & Driessen, A. J. M. F<sub>1</sub>F<sub>0</sub> ATP synthase subunit c is a substrate of the novel YidC pathway for membrane protein biogenesis. *J. Cell Biol.* **165**, 213–222 (2004).
61. Lu, Y. et al. Reorientation of aquaporin-1 topology during maturation in the endoplasmic reticulum. *Mol. Biol. Cell* **11**, 2973–2985 (2000).
62. Mariappan, M. et al. The mechanism of membrane-associated steps in tail-anchored protein insertion. *Nature* **477**, 61–66 (2011).
63. Wang, F., Chan, C., Weir, N. R. & Denic, V. The Get1/2 transmembrane complex is an endoplasmic-reticulum membrane protein insertase. *Nature* **512**, 441–444 (2014).
64. Samuelson, J. C. et al. YidC mediates membrane protein insertion in bacteria. *Nature* **406**, 637–641 (2000).
65. Hessa, T. et al. Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* **450**, 1026–1030 (2007).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

## Methods

### Cell culture and cell lines

All cells were cultured at 37°C with 5% CO<sub>2</sub> in DMEM (Gibco, 10569-010) supplemented with 10% fetal bovine serum (Gibco, 10270106). Wild-type and ΔEMC6 HEK293 cells have been previously described<sup>14</sup>. The ΔTMCO1 HEK293 cell line was obtained from R. Keenan<sup>20</sup>. ΔEMC6ΔTMCO1 double-knockout cells were generated by knocking out TMCO1 in ΔEMC6 cells. Ribonucleoprotein complexes were formed between Alt-R S.p. Cas9-GFP V3 (IDT, 10008100) and Alt-R CRISPR-Cas9 sgRNA (5'-ACTTGCTGTCTCTGTAACC-3'; IDT) following the manufacturer's recommendations. Ribonucleoprotein complexes were transfected into ΔEMC6 cells using Lipofectamine RNAiMAX (Invitrogen, 13778150) according to the manufacturer's protocol; 48 h later, green fluorescent protein (GFP)-positive cells were sorted into single colonies and expanded. Knockout cells were screened by immunoblotting. Flp-In T-REx 293 cells stably expressing wild-type EMC3-FLAG (NP\_060917.1) or EMC3-FLAG variants (M<sup>cyt-1</sup>-S: M101, 106, 110, 111S; R31A; F148L; R13E) were generated by integrating each construct into the FRT site and selecting for Flp-mediated recombination through 100 μg ml<sup>-1</sup> hygromycin B for 2 weeks<sup>6</sup>. The tetracycline-inducible 293 cell line expressing the human GABA<sub>A</sub> receptor has been previously described<sup>29</sup>. In this cell line, the α1 subunit (NP\_001178048.1) in the pcDNA4-TO-Zeocin backbone is FLAG-tagged after its 27 amino acid signal sequence; the β3 subunit (NP\_068712.1) in a pcDNA3.1-TO-Hygromycin backbone is untagged; the γ2L subunit (NP\_944494.1) in a pACMV-TO-blasticidin backbone is 1D4-tagged (TETSQVAPA) at the C terminus after a (GGG)<sub>3</sub>GK linker. Tetracycline-inducible 293 cell lines expressing GFP-P2A-RFP-SQS (NP\_004453.3, aa378–410) and GFP-P2A-RFP-ASGR1 (NP\_001662.1) reporters in the pcDNA5-FRT-TO backbone were generated by stably integrating reporter plasmids into the FRT site and have been previously described<sup>4</sup>.

### Recombinant DNA reagents

Plasmids or gBlocks (IDT) used for in vitro transcription and translation assays contained an SP6 promoter and coding sequences. All plasmids were verified by sequencing. Wild-type GABRA1 (NP\_001345964.1) contains two mutations (V436M and L448M) to facilitate the detection of the C-terminal domain by autoradiography without changing the TMD length, hydrophobicity or C-tail charge. GABRA1-glyc was generated by adding an opsin tag (MNGTEGPNFYVPFSNKTVD)<sup>66</sup> to the C terminus of wild-type GABRA1. SQS-glyc (NP\_004453.3, aa378–410) has been previously described<sup>14</sup>. The 23L-GABRA1 contains an N-terminal 9×His tag, a glycosylation sequence, a soluble tail from β1-adrenergic receptor (NP\_001290104.1, residues 29–44), 23 leucine codons, a soluble cytosolic loop (GGSG-mEGFP(I–92)), TMD4 and flanking regions of GABRA1 (NP\_001345964.1, aa 407–455) and an opsin tag. 23L-SQS replaces TMD4 and flanking sequences of 23L-GABRA1 with the TMD and flanking regions of SQS (NP\_004453.3, aa378–410). The following 23L-SQS variants, used in Figs. 3 and 4, were made by site-directed mutagenesis: P202C (cysteine in C-tail for EMC crosslinking); P202Amber (for incorporation of a photoreactive amino acid into the C-tail); S168C (cysteine in TMD); S185E (–3); E189R (0); E189R,D190R (+2); S185R,E189R,D190R (+3); T183L (1L); T182L,T183L (2L); Q179L,T182L,T183L (3L); S177L,Q179L,T182L,T183L (4L); and S168L,S177L,Q179L,T182L,T183L (5L). Extensions of the C-tail length to a total length of 25, 35, 45, 55, 65 and 225 amino acids were generated by inserting part of the coding sequence for mCherry. The Rho(1–3) domain that precedes the cytosolic loop in the GABRA1 and SQS constructs shown in Fig. 5 contains the following: the prolactin signal sequence (NP\_776378, aa1–33), a Twin-Strep-tag (SAWSHPQFEKGGSGGGSSGSSAWSHHPQFEK), a linker (AGGSAGSGGSSAGGSA), the VHP domain (NP\_990773, aa792–826), a glycosylation site, a linker (GGGSAGGSA) and rhodopsin (NP\_001014890, aa32–152). For the reporter constructs shown in Fig. 7b, TMD4 and flanking

regions of 23L-GABRA1 were replaced by the following sequences: CNH2 (NP\_872359, aa126–160); VATL (NP\_001685, aa119–155); S38A1 (NP\_109599, aa440–487); ACHA4 (NP\_000735, aa588–627); 5HT3B (NP\_006019, aa402–441); and GLRB (NP\_001159532, aa463–497). SOAT1 (NP\_003092.4) was appended at the C terminus with the opsin glycosylation tag. The N-terminal deletion of SOAT1 removed amino acids 2–125. YIPF1 (NP\_061855.1) was obtained from R. Keenan.

### Small interfering RNA knockdown and flow cytometry

For monitoring the surface expression of GABA<sub>A</sub> receptors, either negative control small interfering RNA (siRNA) (Invitrogen, 4390843) or EMC4 siRNA (Ambion, s27733) was transfected into the 293 cell line expressing the human GABA<sub>A</sub> receptor using Lipofectamine RNAiMAX according to the manufacturer's protocol. A final concentration of 10 nM of siRNA was used and the total knockdown time was 66 h. At 60 h, doxycycline was added to a final concentration of 0.1 μg ml<sup>-1</sup> to induce expression of the GABA<sub>A</sub> receptor for 6 h. Cells were then collected and subjected to surface labeling. Pelleted cells were resuspended with 100 μl of cold PBS, supplemented with 1 μl of phycoerythrin-labeled FLAG antibody (BioLegend, 637310) and incubated in the dark at 4°C for 1 h. Cells were then washed with cold PBS, passed through a 70 μm filter and then analyzed on a BD LSR II flow cytometer for appropriate fluorescent channels. A total of 30,000 events were analyzed, and phycoerythrin fluorescence, reflective of GABA<sub>A</sub> receptor surface levels, was plotted as a histogram using FlowJo. To analyze the stability of ASGR1 or SQS, 293 cell lines stably expressing GFP-P2A-RFP-ASGR1 or GFP-P2A-RFP-SQS were used as previously described<sup>50</sup>. The P2A sequence in these constructs causes ribosome skipping, resulting in the translation of equimolar amounts of GFP and the RFP-tagged protein. Therefore, a steady-state RFP:GFP ratio reflects the stability of the RFP-tagged protein. Failure in biogenesis will lead to degradation by cellular quality control pathways and a decreased RFP:GFP ratio. Knockdown was performed as for GABA<sub>A</sub> receptor cells, with the GFP and RFP fluorescence monitored by flow cytometry on 30,000 cells. The RFP:GFP ratio was plotted as a histogram using FlowJo.

### Preparation of SPCs

Cells at 95–100% confluency were trypsinized and collected by centrifugation at 4°C, washed once with ice-cold 1×PBS and resuspended in 1×RNC buffer (50 mM HEPES, pH 7.4, 100 mM KOAc, 5 mM Mg(OAc)<sub>2</sub>) containing 0.01% purified digitonin<sup>67</sup>. SPCs were pelleted and washed once with 1×RNC. To digest endogenous mRNAs, SPCs were resuspended in 100 μl of 1×RNC containing 1 mM CaCl<sub>2</sub> and 150 units per ml micrococcal nuclease (Roche, 10107921001). Nuclease digestion was performed for 10 min at room temperature (20°C) and was terminated by adding a final concentration of 2 mM EGTA. Nuclease-digested SPCs were pelleted, washed once with 1×RNC buffer, resuspended in 0.5×RNC buffer to 6,000–10,000 cells per ml and used immediately in translocation assays.

### Preparation of endoplasmic-reticulum-enriched membranes

Approximately 80% of confluence cells were collected by trypsinization. Cells were pelleted, washed once with cold 1×PBS and flash-frozen in liquid nitrogen. Thawed cell pellets were mixed with 5 volumes of 20 mM HEPES, pH 7.4; 5 mM KCl; 1.5 mM MgCl<sub>2</sub>; 2 mM dithiothreitol (DTT); and protease inhibitor (Roche, 10106399001) and incubated on ice for 15 min. Cells were lysed on ice by 35 strokes of dounce homogenization (DWK Life Sciences, 357542). Cell lysates were adjusted to 20 mM HEPES, pH 7.4; 210 mM mannitol; 70 mM sucrose; 0.5 mM EDTA; 2 mM DTT; and protease inhibitor. Cell debris and nuclei were cleared by centrifugation at 4°C for 10 min at 700×g. Membranes were then pelleted by centrifugation at 4°C for 10 min at 8,500×g, washed once and resuspended in a buffer containing 20 mM HEPES, pH 7.4; 210 mM mannitol; 70 mM sucrose; 0.5 mM EDTA; 2 mM DTT; and protease inhibitor to give an OD280 of 20. Different amounts of resuspended

membranes were used to assay the levels of key translocation components by blotting.

### In vitro transcription and translation

Transcription reactions with SP6 polymerase were performed at 37°C for 1 h and contained the following components: DNA that encodes regions of interest for translation reactions (PCR-amplified and purified by Qiagen PCR purification kit, 10 ng  $\mu\text{l}^{-1}$ ); HEPES, pH 7.4 (40 mM); spermidine (2 mM; Sigma, S0266); RNA cap structure analog (0.33 mM; NEB, S1404L); reduced glutathione (10 mM);  $\text{MgCl}_2$  (6 mM); NTPs (0.5 mM each for ATP; Roche, 10519979001), CTP (Sigma, C1506) and UTP (Sigma, U6875), 0.1 mM for GTP (Roche, 10106399001); SP6 RNA polymerase (0.4 U  $\mu\text{l}^{-1}$ ; NEB, M0207L); and RNase inhibitor (0.8 U  $\mu\text{l}^{-1}$ ; Promega, N2515).

Translation reactions were performed at 32°C for 30 min and contained the following components: micrococcal nuclease-digested rabbit reticulocyte lysates (Green Hectares) (34% of the total volume); transcription reaction from the previous step (5% volume); SPCs (10% volume); ATP and GTP (1 mM each); an ATP regeneration system (creatine phosphate (12 mM; Roche, 10621714001); creatine kinase (0.04 mg  $\text{ml}^{-1}$ ; Roche, 10127566001)); spermidine (0.3 mM); HEPES, pH 7.4 (20 mM); KOAc (50 mM);  $\text{Mg}(\text{OAc})_2$  (2 mM); reduced glutathione (1 mM); tRNAs purified from pig liver (0.05 mg  $\text{ml}^{-1}$ ); 19 of the 20 amino acid except for methionine (40  $\mu\text{M}$  each; Promega, L9961); and  $^{35}\text{S}$ -methionine (0.5  $\mu\text{Ci}$   $\mu\text{l}^{-1}$ ; PerkinElmer, NEG709A001MC).

For incorporating the photoreactive amino acid p-benzoyl-l-phenylalanine (Bpa) through amber suppression, the following components are included in the translation reaction<sup>68</sup>: suppressor *Bacillus stearothermophilus* tRNA<sub>CUA</sub><sup>Tyr</sup> (5  $\mu\text{M}$ ); *E. coli* Bpa tRNA synthetase (0.25  $\mu\text{M}$ ); and Bpa (100  $\mu\text{M}$ ). These components were pre-mixed into a 10 $\times$  solution (in 50 mM HEPES, pH 7.4, 100 mM KOAc, 1 mM  $\text{Mg}(\text{OAc})_2$ ) and were pre-incubated at 32°C for 15 min before adding to the translation reaction. Where indicated, the Sec61 lateral gate inhibitor ApraA (obtained from V. Paavilainen and K. McPhail) was included in the translation reaction at 2  $\mu\text{M}$ .

### Protease protection assays

The 60  $\mu\text{l}$  in vitro translation reactions were chilled on ice, and the SPCs were pelleted (20,000 $\times$ g for 2 min), washed once with 1 $\times$ RNC buffer (50 mM HEPES, pH 7.4, 100 mM KOAc, 5 mM  $\text{Mg}(\text{OAc})_2$ ) and resuspended in 30  $\mu\text{l}$  0.5 $\times$ RNC buffer. Samples lacking SPCs were used directly without pelleting. Samples were divided into two aliquots; one aliquot (two-thirds of the total volume) was adjusted to 0.5 mg  $\text{ml}^{-1}$  proteinase K and incubated on ice for 50 min. Proteinase K was quenched by adding 250 mM of PMSF for 2 min, then transferring the entire reaction to a tenfold excess volume of 1% SDS, 100 mM Tris-HCl, pH 8.0 pre-heated to 100°C and heated for 10 min. The samples were either analyzed directly or subjected to immunoprecipitation as indicated in the figure legends.

### Site-specific crosslinking

The 120  $\mu\text{l}$  in vitro translation reactions were used for bismaleimidohexane (BMH) crosslinking experiments. All steps following the translation reaction were at 4°C until the reaction was denatured in SDS. SPCs were pelleted and resuspended in 60  $\mu\text{l}$  0.5 $\times$ RNC buffer. One aliquot was removed as the no-crosslinking control and the remainder of the sample was adjusted to 250  $\mu\text{M}$  BMH (Thermo Scientific, 22330) and incubated on ice for 10 min. The crosslinking reaction was quenched by adjusting the final concentration of DTT to 25 mM. After denaturation in 1% SDS, 100 mM Tris-HCl, pH 8.0, the samples were either analyzed directly or processed further for immunoprecipitation or deglycosylation. SMPH (Succinimidyl 6-((beta-maleimidopropionamido) hexanoate)) and UV crosslinking experiments were performed similarly to the BMH crosslinking experiments, with the following differences: SMPH (Thermo Scientific, 22363) crosslinking (200  $\mu\text{l}$  total reaction)

was at 200  $\mu\text{M}$  final concentration for 30 min, and quenched with 50 mM Tris-HCl pH 7.4 and 5 mM DTT; UV crosslinking (100  $\mu\text{l}$  total reaction) was on ice with UV irradiation by a UVP Blak-Ray B-100AP high-intensity lamp with the bulb positioned -10 cm above the samples.

### Immunoprecipitation and PNGase F treatment

Denaturing immunoprecipitation after crosslinking, proteinase K digestion and glycanase digestion have been previously described<sup>6</sup>. SDS-denatured samples were diluted tenfold in ice-cold immunoprecipitation buffer (1 $\times$ PBS, 250 mM NaCl, 0.5% TX-100, 10 mM imidazole) and mixed with 2.5  $\mu\text{l}$  of anti-FLAG resin (Millipore, A2220), 2.5  $\mu\text{l}$  of Monoclonal Anti-HA resin (Millipore, A2095), or 5  $\mu\text{l}$  of protein A resin (Repligen, CA-HF-0100) along with the appropriate antibody. A total of 1.25  $\mu\text{g}$  of GABRA1 antibody (Invitrogen, PA5-79291) was used per immunoprecipitation. The mixture was rotated end-over-end for 1.5 h (for FLAG or HA) or 3 h (for GABRA1 immunoprecipitation) at 4°C. Beads were washed twice with cold immunoprecipitation buffer and eluted by boiling in 10  $\mu\text{l}$  of 2.5 $\times$  SDS-PAGE sample buffer for 10 min. For deglycosylation experiments, crosslinked samples (Fig. 3) or total translation reactions (Extended Data Fig. 4) were split into two halves after denaturation with 0.5% SDS and 50 mM Tris-HCl, pH 8. One half was untreated and the other was adjusted to 1% NP-40, 1 $\times$  GlycoBuffer 2 and 25 U  $\text{ml}^{-1}$  of PNGase F (NEB, P0704S) and digested at 32°C for 30 min. Both halves were subjected to immunoprecipitation as described above (Fig. 3) or analyzed directly (Extended Data Fig. 4).

### Bioinformatic analysis of membrane proteins

All proteins containing TMDs were retrieved from the UniProt database<sup>69</sup>. The UniProt annotations were used to define the start and end of the TMD helices. Proteins containing a single TMD or multipass membrane proteins localized to mitochondria were manually removed from this set. The AlphaFold2 (ref. 51) predicted structure, available from the UniProt database for each of the remaining 1,784 multipass membrane proteins, was inspected manually to annotate the number of TMD helices and the overall charge of each TMD-flanking side of the structure. The overall basic flank was designated cytosolic as per the positive-inside rule<sup>52</sup>. Then the C-terminal TMD was identified and assigned the appropriate orientation, hydrophobicity (as calculated using the  $\Delta G_{\text{app}}$  predictor<sup>65</sup>) and flanking C-tail length. C-tails facing the cytosol were designated 'C<sub>cyt</sub>' and those facing the opposite orientation were designated 'C<sub>exo</sub>'. The curated list is provided in Supplementary Table 1; information from this table was used to generate the plots in Fig. 6.

### SDS-PAGE and blotting

Cell lysates or endoplasmic-reticulum-enriched membranes were analyzed by SDS-PAGE on 12% Tris-Tricine gels. SDS-PAGE gels were transferred to a nitrocellulose membrane (Biorad, 1620112) and blotting was performed with standard procedures using 5% non-fat dried milk as the blocking agent. The following antibodies and dilutions were used for blotting: CCDC47 (Bethyl Laboratories, A305-100A; 1:5,000); EMC3 (Invitrogen, 711771; 1:5,000); EMC6 (Abcam, ab84902; 1:1,000); Calnexin (Enzo, ADI-SPA-865; 1:5,000); Sec61 $\alpha$  (ref. 70; 1:5,000); TMCO1 (Invitrogen, PA5-43350; 1:500); Sec61 $\beta$  (ref. 71; 1:5,000); CAML (Cell Signaling Technology, 13913S; 1:1,000); FLAG M2-HRP (Sigma, A8592; 1:5,000); EMC4 (Abcam, ab123719; 1:2,000); and  $\beta$ -Actin-HRP (Sigma, A3854; 1:10,000).

### Quantification of C-tail translocation

Quantification was performed on raw phosphorimager files using Fiji. The pixel intensity and area of each band were measured, from which the background intensity was subtracted. For the C-terminal TMD reporters, C-tail translocation was calculated by dividing the value for the C-tail translocated band (typically the 2 $\times$ -glycosylated

product) by the sum of total membrane inserted bands (typically the 1×-glycosylated and 2×-glycosylated bands). In the case of SQS-glyc (Fig. 1d,e), SOAT1 (Fig. 7c) and YIPF1 (Fig. 7d), per cent C-tail translocation is calculated by dividing the intensity of the glycosylated band by the sum of glycosylated and non-glycosylated bands.

### Statistics and reproducibility

This study does not contain any statistical analysis. All data presented in this paper have been reproduced in independent experiments. The number of independent experiments are indicated in parentheses for the following main and extended data figure panels: Fig. 2a (2); Fig. 2b (2); Fig. 2c (2); Fig. 2d (3); Fig. 3a (3); Fig. 3b (3); Fig. 3c (2); Fig. 3d (3); Fig. 4a (2); Fig. 4b (2); Fig. 4c (2); Fig. 5 (2); Fig. 7b (2); Fig. 7c (4 for SOAT1, 2 for SOAT1ΔN); Fig. 7d (6); Extended Data Fig. 1 (2); Extended Data Fig. 2 (2); Extended Data Fig. 3 (2); Extended Data Fig. 4 (2).

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

All data are provided in the Main, Extended Data and Supplementary Information. Publicly available data sets used in this study were obtained from the UniProt database (<http://www.uniprot.org>) and the AlphaFold2 protein structure database (<https://www.alphafold.ebi.ac.uk>). Source data are provided with this paper.

### References

66. Brambillasca, S. et al. Transmembrane topogenesis of a tail-anchored protein is modulated by membrane lipid composition. *EMBO J.* **24**, 2533–2542 (2005).
67. Görlich, D. & Rapoport, T. A. Protein translocation into proteoliposomes reconstituted from purified components of the endoplasmic reticulum membrane. *Cell* **75**, 615–630 (1993).
68. Lin, Z. et al. TTC5 mediates autoregulation of tubulin via mRNA degradation. *Science* **367**, 100–104 (2020).
69. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158–D169 (2017).
70. Song, W., Raden, D., Mandon, E. C. & Gilmore, R. Role of Sec61α in the regulated transfer of the ribosome-nascent chain complex from the signal recognition particle to the translocation channel. *Cell* **100**, 333–343 (2000).
71. Fons, R. D., Bogert, B. A. & Hegde, R. S. Substrate-specific function of the translocon-associated protein complex during translocation across the ER membrane. *J. Cell Biol.* **160**, 529–539 (2003).

### Acknowledgements

We thank H. Wang and other Hegde Lab members for fruitful discussions, Y. Hooda for flow cytometry data, R. Judy for help with data curation, S. Juszkiwicz for plasmids, V. Paavilainen and K. McPhail for providing Apratoxin A and R. Keenan for discussions and reagents. This work was supported by the UK Medical Research Council (grant MC\_UP\_A022\_1007 to RSH) and a European Molecular Biology Organization (EMBO) Postdoctoral Fellowship (ALTF 369-2021 to H.W.).

### Author contributions

H.W. performed all experiments in the study. L.S. generated the ΔEMC6ΔTMCO1 cell line. H.W. and R.S.H. conceived the project and R.S.H. provided guidance and mentoring. H.W. and R.S.H. wrote the paper.

### Competing interests

The authors declare no competing interests.

### Additional information

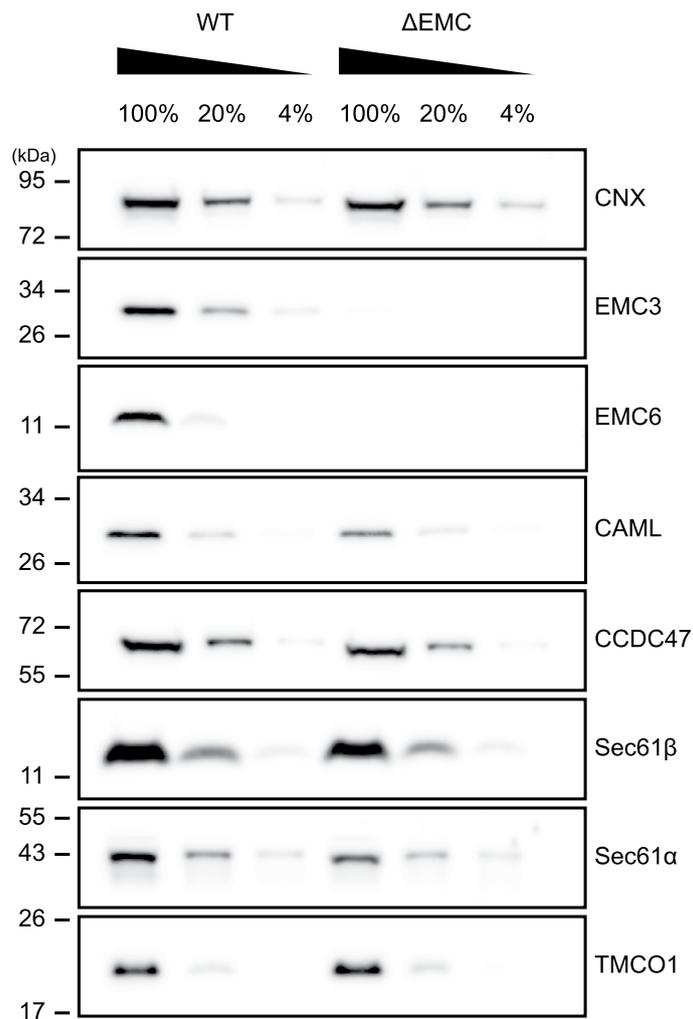
**Extended data** is available for this paper at <https://doi.org/10.1038/s41594-023-01120-6>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41594-023-01120-6>.

**Correspondence and requests for materials** should be addressed to Ramanujan S. Hegde.

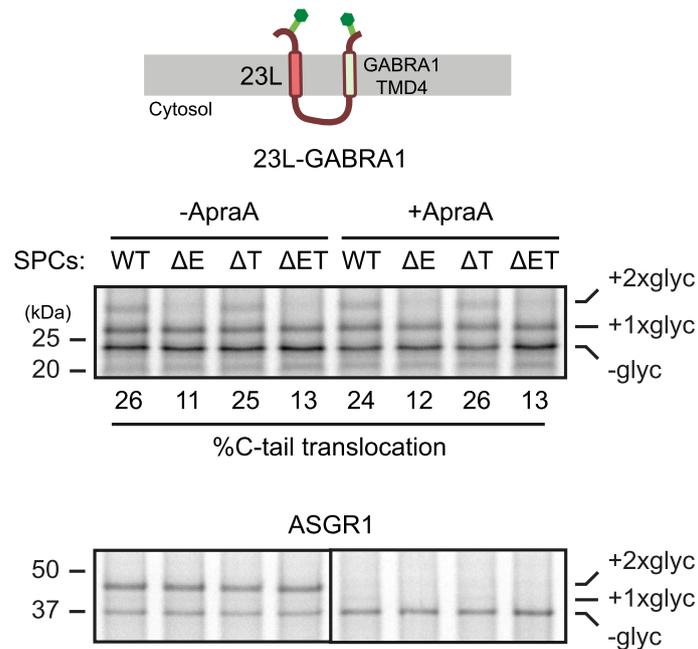
**Peer review information** *Nature Structural & Molecular Biology* thanks Eunyong Park and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Katarzyna Ciazynska, in collaboration with the *Nature Structural & Molecular Biology* team.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



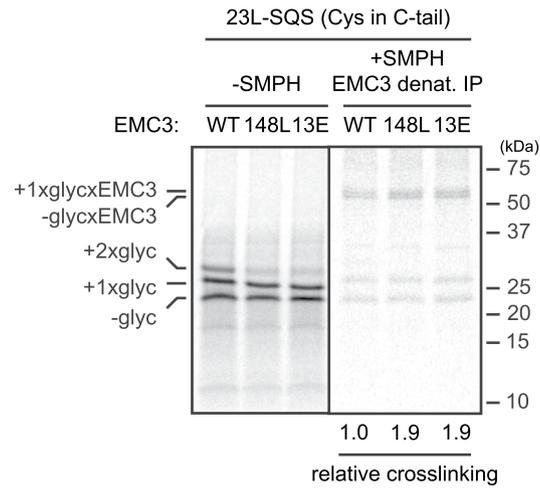
**Extended Data Fig. 1 | Characterization of  $\Delta$ EMC cells.** Serial dilutions of total microsomes from WT or  $\Delta$ EMC cells were analyzed by SDS-PAGE and immunoblotting for the indicated ER-resident proteins. These data complement

earlier analysis for additional ER proteins (including Sec62, Sec63, TRAM, TRAP $\alpha$ , the 12 kD subunit of the signal peptidase complex, and other EMC subunits) using the same parental and  $\Delta$ EMC cells<sup>4</sup>.



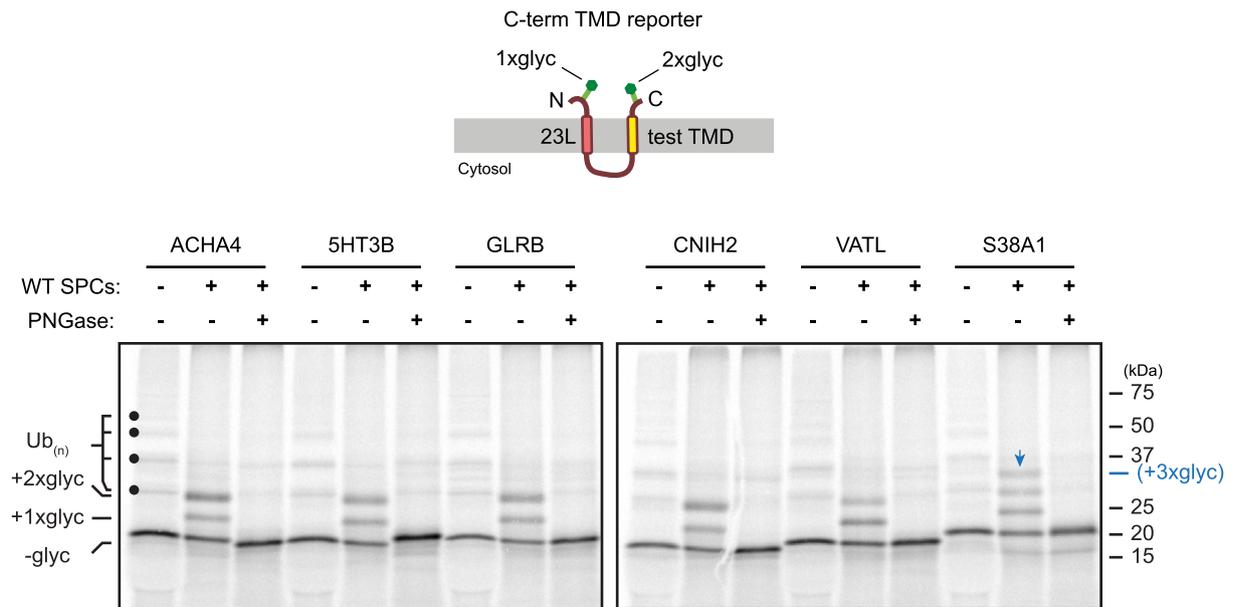
**Extended Data Fig. 2 | Characterization of GABRA1 C-terminal TMD insertion.** Domain diagram of 23L-GABRA1 is shown (top): a glycosylated N-tail is followed by a TMD comprised of 23 leucine residues (23L), a cytosolic loop (~100aa), and TMD4 of GABRA1 that is followed by a translocated and glycosylated C-tail. <sup>35</sup>S-methionine labeled 23L-GABRA1 (middle) or ASGR1 (bottom) were translated in the presence of SPCs from WT, ΔEMC (ΔE), ΔTMC01 (ΔT) or ΔEMCΔTMC01

(ΔET) cells. Translated products were analyzed directly by SDS-PAGE and autoradiography. Where indicated, Sec61 inhibitor ApraA was included in the translation reaction. Substrates with different glycosylation states are indicated. C-tail translocation of 23L-GABRA1 was quantified by calculating the percentage of double glycosylated products among all glycosylated products.



**Extended Data Fig. 3 | Analysis of EMC3-substrate crosslinking in EMC3 mutants.** <sup>35</sup>S-methionine labeled 23L-SQS that contains a single cysteine within its C-tail was translated in the presence of SPCs derived from cells expressing FLAG tagged WT or each of two EMC3 mutants that impair its insertase function (F148L or R13E). Translation products were analyzed directly (-SMPH) or

subject to SMPH crosslinking and FLAG denaturing IP (+SMPH, EMC3 denat. IP). Differentially glycosylated substrates and EMC3 crosslinks are indicated. Relative crosslinking was quantified by normalizing intensity of crosslinked products relative to totals. Note that earlier analysis has shown that these mutants do not affect either the expression level or assembly of EMC<sup>6</sup>.



**Extended Data Fig. 4 | Glycosylation analysis of C-terminal TMD reporters.** The diagram shows the domain organization of the C-terminal TMD reporters consisting of an N-terminal translocated tail, a TMD with 23 leucines, a cytosolic loop (-100aa) and a test TMD followed by a C-terminal translocated tail. If properly inserted, both N- and C-tails are glycosylated. <sup>35</sup>S-methionine labeled reporters were translated in the absence or presence of WT SPCs. Translated

products were either analyzed directly or subject to PNGase F digestion to definitively identify the glycosylated products as indicated. Ubiquitinated non-translocated products [ $Ub_{(n)}$ ] are indicated in black dots. A triply glycosylated form of S38A1 due to a second consensus site in the C-tail is marked by a blue arrow.