Article

An intramembrane chaperone complex facilitates membrane protein biogenesis

https://doi.org/10.1038/s41586-020-2624-y

Received: 28 January 2020

Accepted: 21 May 2020

Published online: 19 August 2020

Check for updates

Patrick J. Chitwood[1] & Ramanujan S. Hegde[1]✉

Integral membrane proteins are encoded by approximately 25% of all protein-coding genes[1]. In eukaryotes, the majority of membrane proteins are inserted, modified and folded at the endoplasmic reticulum (ER)[2]. Research over the past several decades has determined how membrane proteins are targeted to the ER and how individual transmembrane domains (TMDs) are inserted into the lipid bilayer[3]. By contrast, very little is known about how multi-spanning membrane proteins with several TMDs are assembled within the membrane. During the assembly of TMDs, interactions between polar or charged amino acids typically stabilize the final folded configuration[4–8]. TMDs with hydrophilic amino acids are likely to be chaperoned during the co-translational biogenesis of membrane proteins; however, ER-resident intramembrane chaperones are poorly defined. Here we identify the PAT complex, an abundant obligate heterodimer of the widely conserved ER-resident membrane proteins CCDC47 and Asterix. The PAT complex engages nascent TMDs that contain unshielded hydrophilic side chains within the lipid bilayer, and it disengages concomitant with substrate folding. Cells that lack either subunit of the PAT complex show reduced biogenesis of numerous multi-spanning membrane proteins. Thus, the PAT complex is an intramembrane chaperone that protects TMDs during assembly to minimize misfolding of multi-spanning membrane proteins and maintain cellular protein homeostasis.

Multi-spanning membrane proteins have wide-ranging functions in transmembrane transport and intercellular interactions, and as cell surface receptors[2]. Many of these functions require the TMDs of these proteins to serve as more than simple membrane anchors. For example, receptors or transporters need specific hydrophilic surfaces in the plane of the membrane to execute their functions[2]. For this reason, the TMDs of multi-spanning membrane proteins are highly diverse in length, sequence and biophysical properties.

When folded, the TMDs of multi-spanning membrane proteins typically pack such that their hydrophilic regions are facing each other, presenting a mostly uninterrupted hydrophobic surface to the surrounding lipid bilayer[4–8]. How semi-hydrophilic TMDs are temporarily stabilized within the membrane until their assembly with distal TMDs is not well understood. ER-resident intramembrane chaperones have been postulated but are poorly defined[3,9,10]. Chaperones that act to facilitate membrane protein biogenesis are likely to be important because impaired folding can cause loss- and gain-of-function phenotypes in many experimental systems and various diseases in humans[11,12].

Detection of a potential TMD chaperone

We sought to identify interacting partners of early-stage intermediates during biogenesis of G-protein-coupled receptors (GPCRs). Tagged constructs coding for the first two TMDs of the GPCR bovine rhodopsin (Rho TM1 + 2) were inserted in the appropriate orientation in ER-derived rough microsomes isolated from HEK293 cells (Extended Data Fig. 1). Stalled translation complexes containing different lengths of nascent polypeptide were crosslinked to adjacent proteins through a single cysteine introduced into TMD1 (Fig. 1a, Extended Data Fig. 2a, b). Early intermediates (gel lanes 1–3) crosslinked to well-characterized ribosomal and ribosome-associated proteins, including the SRP54 subunit of the targeting factor SRP[13] and the Sec61α subunit of the translocation channel[14]. Although the ER membrane protein complex (EMC) facilitates TMD1 insertion[15], this interaction was not captured by cysteine crosslinking because it is thought to act transiently[16].

Intermediates in which TMD2 was emerging from the ribosome (lanes 4–6) showed progressively less intense TMD1–Sec61α crosslinks and the appearance of crosslinks to a protein of approximately 10 kDa. We presume this protein is identical to a previously observed crosslinking partner that was denoted PAT10, for 'protein associated with the ER translocon of 10 kDa'[10]. At all subsequent lengths, the PAT10 crosslink persisted at approximately the same efficiency and the Sec61α crosslink remained low in abundance. Thus, PAT10 engages TMD1 as it departs the vicinity of the Sec61 complex[17].

Substrate crosslinking to PAT10 did not depend on glycosylation and crosslinks were observed from other cysteine positions within TMD1 (Extended Data Fig. 2c–e), consistent with previous observations[10]. The natively solubilized ribosome-released TMD1–PAT10 complex migrated at an apparent molecular mass of more than 100 kDa by sucrose gradient sedimentation (Fig. 1b). This indicates that PAT10 is part of a larger assembly—the PAT complex—that is adjacent to a membrane-inserted TMD. Because efficient maleimide-mediated crosslinking requires an aqueous environment, the PAT complex seems to provide a membrane-embedded partially hydrophilic environment in which TMD1 binds.

[1]MRC Laboratory of Molecular Biology, Cambridge, UK. ✉e-mail: rhegde@mrc-lmb.cam.ac.uk

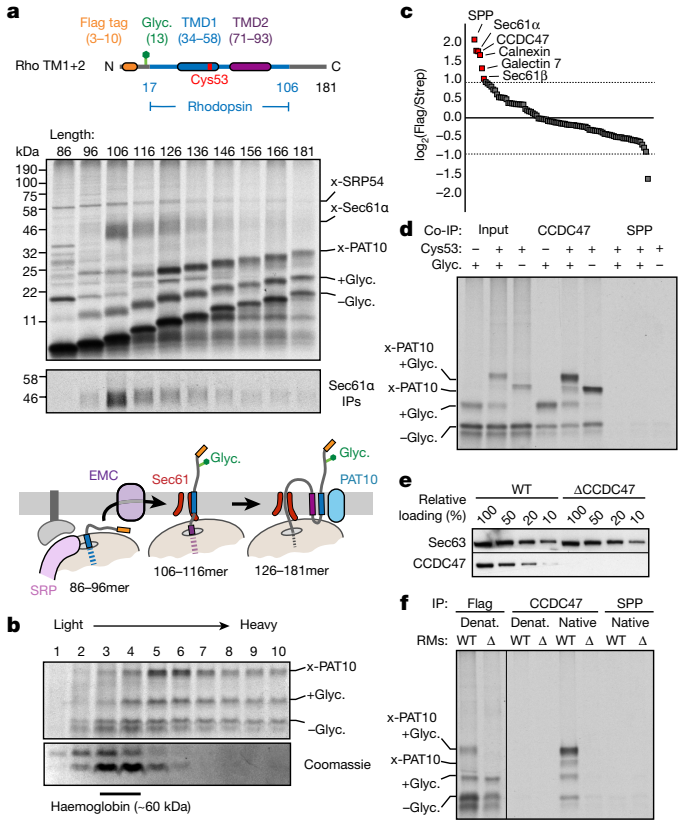


Fig. 1 | A protein complex containing CCDC47 engages nascent membrane proteins. a, Cysteine-based crosslinking of ^{35}S-labelled ribosome nascent chain complexes (RNCs) of the indicated length representing intermediates during targeting and insertion of mammalian rhodopsin. The construct is shown in the top diagram and a schematic of the results in the bottom diagram, beneath the gels. RNCs were produced by in vitro translation containing ER-derived rough microsomes from HEK293 cells. The top gel shows the translation products and all of their crosslinks (x) as visualized by autoradiography of immunoprecipitations (IPs) via the N-terminal Flag tag of the nascent chain. The non-glycosylated (–glyc.) and glycosylated (+glyc.) translation products and the crosslinks to PAT10, Sec61α and SRP54 are indicated. The bottom autoradiograph shows the immunoprecipitation products using antibodies against Sec61α. **b**, Sucrose gradient separation of the ^{35}S-labelled membrane-targeted 146-mer RNC after bismaleimidohexane (BMH) crosslinking, native solubilization and release from the ribosome by RNase digestion. Endogenous haemoglobin (around 60 kDa) from the translation extract was visualized by Coomassie staining of the same gel. **c**, BMH-crosslinked 146-mer RNCs containing a Flag or a Strep tag were released from the ribosome by RNase digestion, subjected to native Flag IPs and analysed by quantitative mass spectrometry. Proteins enriched twofold or more in the Flag-tagged RNCs are indicated. **d**, 146-mer RNCs containing or lacking either a glycosylation site (glyc.) or a cysteine at position 53 (Cys53) were crosslinked with BMH and analysed directly (input) or after native IP using antibodies against CCDC47 or SPP. Nascent chains were released from the ribosome with RNase A before immunoprecipitation. **e**, Rough microsomes prepared from wild-type (WT) or CCDC47-knockout (ΔCCDC47) HEK293 cells were analysed for CCDC47 and Sec63 by immunoblotting. **f**, RNCs (146-mer) targeted to rough microsomes (RMs) from wild-type or ΔCCDC47 (Δ) cells were treated with BMH, released from the ribosome with RNase, and immunoprecipitated under denaturing (denat.) or native conditions with the antibodies indicated. For gel source data, see Supplementary Fig. 1.

CCDC47 and Asterix form the PAT complex

To identify the components of the PAT complex, we solubilized the crosslinking reactions under non-denaturing conditions, released the nascent chain from the ribosome, purified the protein by affinity purification of the Flag tag on the nascent chain and identified the co-purifying proteins by mass spectrometry (Fig. 1c). Six proteins were enriched more than twofold relative to a control substrate containing a Strep tag: Sec61α and Sec61β, the lectins calnexin and galectin 7, signal peptide peptidase (SPP) and CCDC47.

Sec61α, Sec61β and SPP were apparently recovered because of their direct crosslinking to the nascent chain (Fig. 1a, Extended Data Fig. 3a). Of the remaining candidates, we focused on CCDC47 because—unlike the others—the PAT10 crosslink was strongly enriched in native immunoprecipitations using anti-CCDC47 antibodies (Fig. 1d, Extended Data Fig. 3b, c). Even without crosslinking, the inserted population of nascent chains was enriched by CCDC47 native immunoprecipitations (but not denaturing immunoprecipitations) from wild-type microsomes, but not from CCDC47-knockout microsomes (Fig. 1e, f). TMD1 crosslinking to PAT10 was lost in microsomes that lacked CCDC47 (Fig. 1f, Extended Data Fig. 3d). We therefore conclude that CCDC47 interacts with membrane proteins and is a stable subunit of the PAT complex.

Affinity columns prepared with two different anti-CCDC47 antibodies each recovered CCDC47 and a prominent protein of approximately 10 kDa observed on silver stained gels (Fig. 2a). Whereas mass spectrometric analysis of tryptic digests did not identify this protein—explaining why it was not identified in the initial mass spectrometry experiment—digestion with other proteases produced peptides consistent with Asterix (encoded by *WDR83OS*). No other stoichiometric and specific interaction partners were observed in the purifications or identified by mass spectrometry experiments. Antibodies against Asterix specifically recovered the PAT10 crosslink by denaturing immunoprecipitation (Fig. 2b). It was established that CCDC47 and Asterix form an obligate complex, because knockdown or knockout of either protein results in substantial loss of the other (Fig. 2c) and immunodepletion of CCDC47 from ER extracts efficiently depletes Asterix (Fig. 2d). We therefore identify the components of the PAT complex as CCDC47 and Asterix, two poorly studied proteins that are widely expressed in all tissues and are widely conserved across eukaryotes.

As depicted in Fig. 2e, CCDC47 is an ER-resident single-pass membrane protein with a well-conserved cytosolic domain[1]. Sequence and topology analysis of Asterix suggests that it is a three-TMD membrane protein, the N terminus of which faces the cytosol and the C terminus faces the ER lumen (Extended Data Fig. 4). The three putative TMDs are each only around 15 amino acids in length, are populated by numerous hydrophilic amino acids, and contain several cysteine residues that explain the crosslinking of Asterix to nascent TMDs via bismaleimidohexane.

We used a recently optimized site-specific photo-crosslinking approach[18] (Extended Data Fig. 5a, b) to introduce the UV-activated crosslinking amino acid benzoyl-phenylalanine at 11 positions within or near TMD1 of β1-adrenergic receptor (β1AR), a GPCR in the same family as rhodopsin. At a nascent chain length 75 amino acids downstream of TMD1—when TMD1 has been inserted and can diffuse away from Sec61α—several UV-dependent crosslinking products were observed (Fig. 2f). The most prominent of these were SRP54 (bound to non-translocated nascent chains) and UBQLN2, a cytosolic quality control factor for TMDs that fail targeting[19]. The most prominent remaining crosslink was to Asterix, as verified by immunoprecipitation. Crosslinking to Asterix was observed from most but not all positions within the TMD, and was not seen from control positions at the boundaries or outside the TMD (Fig. 2g, Extended Data Fig. 5c). Crosslinks to CCDC47 were not observed. We therefore conclude that Asterix is the substrate-interacting subunit of the PAT complex, whereas CCDC47 is required to maintain the stability of Asterix. This conclusion is consistent with the chemical crosslinking experiments.

The PAT complex affects membrane proteins

Using a dual-colour ratiometric assay for protein stability[15], we found that acute depletion of the PAT complex using short interfering RNAs

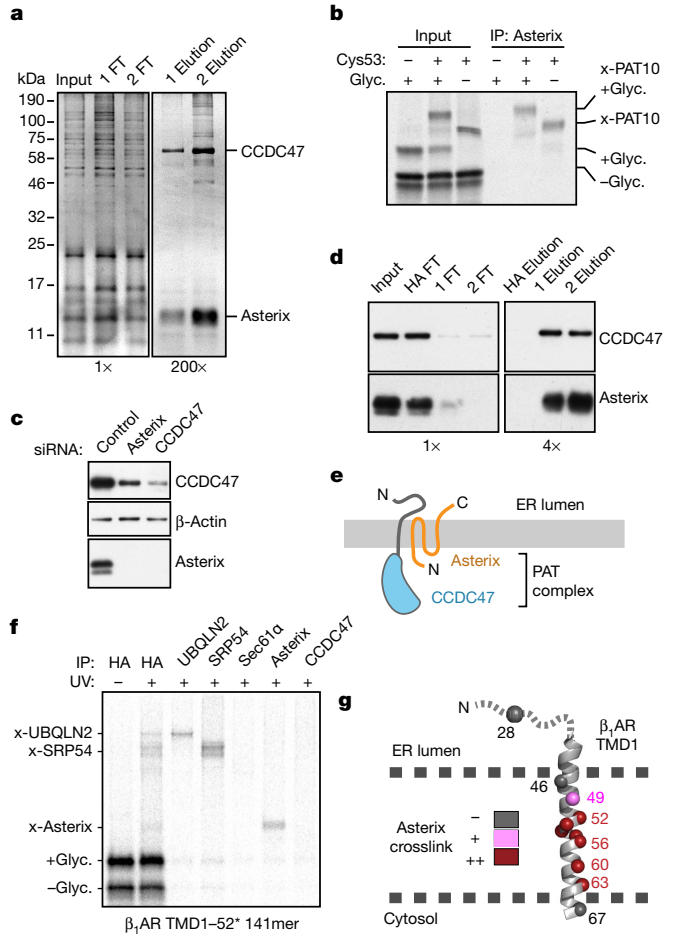


Fig. 2 | Asterix is the substrate-binding subunit of the PAT complex. **a**, Affinity purification of CCDC47 from natively solubilized rough microsomes using two unrelated CCDC47 antibodies (1 and 2). The elution samples are derived from approximately 200-fold more rough microsomes than the input and flow-through (FT) samples. **b**, ^{35}S-radiolabelled 146-mer RNCs of Rho TM1 + 2 (Fig. 1a) containing or lacking Cys53 or a glycosylation site were targeted to rough microsomes, treated with BMH, digested with RNase, and analysed directly (input) or after denaturing immunoprecipitation with anti-Asterix antibodies. **c**, Asterix or CCDC47 was depleted from HEK293 cells using siRNAs, and control samples were treated with scrambled siRNA. The indicated antigens were detected by immunoblotting. **d**, Affinity purification via CCDC47 as in **a**, but with a negative control using anti-HA antibodies. Asterix and CCDC47 were detected by immunoblot. The elution samples are derived from fourfold more rough microsomes than the input and FT samples. **e**, Cartoon depicting topology of the PAT complex as deduced from predictions and direct analysis (Extended Data Fig. 4). **f**, Site-specific photo-crosslinking of a 141-mer RNC containing the UV-activated photo-crosslinking amino acid benzoyl-phenylalanine at position 52 within TMD1. Total translation products recovered by immunoprecipitation via an haemagglutinin (HA) tag on the nascent chain are shown adjacent to parallel immunoprecipitations using the indicated antibodies. RNCs that did not engage SRP crosslink to UBQLN2, a quality control factor that binds exposed TMDs. A subset of RNCs do not release from SRP and crosslink to SRP54. Of the membrane-inserted RNCs, the main crosslink is to Asterix. At this length, the TMD has moved away from Sec61α, so crosslinks to this factor are minimal. **g**, Summary of Asterix crosslinks observed (or not) from different positions in or near TMD1 of β1AR (Extended Data Fig. 5). No CCDC47 crosslinks were seen from any of these positions. For gel source data, see Supplementary Fig. 1.

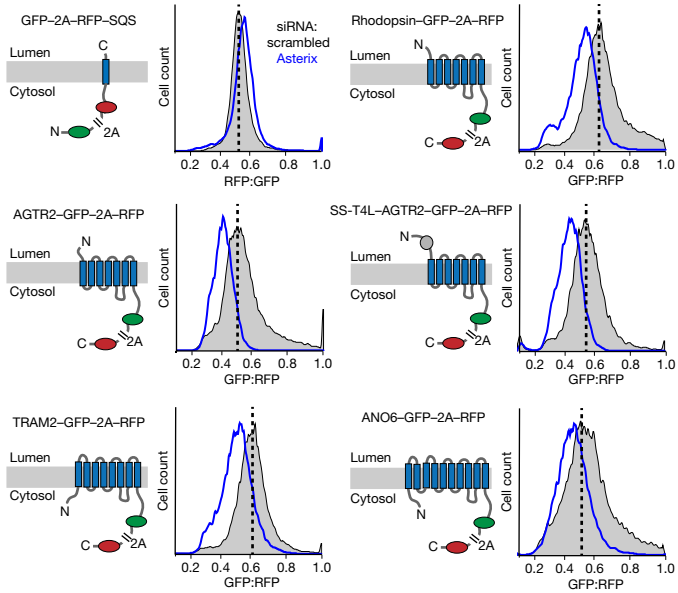


Fig. 3 | The PAT complex facilitates biogenesis of multi-spanning membrane proteins. Stable cell lines containing the indicated inducible reporters were treated with scrambled or Asterix-targeting siRNAs, reporter expression was induced for around 6 h, and the cells were analysed by flow cytometry. A cartoon depicting the topology, number of TMDs and fluorescent proteins for each of the membrane protein reporters is shown to the left of its respective flow cytometry data. The viral P2A peptide sequence results in two proteins from a single translation reaction as indicated. The plots show histograms of fluorescent protein ratios in control cells (grey) and Asterix-knockdown cells (blue). The dashed black line indicates the mode for the control population.

(siRNAs) against Asterix—which co-deplete CCDC47—had minimal effects on the tail-anchored membrane protein squalene synthase (SQS) but resulted in reduced biogenesis of five GPCRs, the ER resident protein TRAM2 and the cation channel ANO6 (Fig. 3, Extended Data Fig. 6). Similar phenotypes were seen after CCDC47 knockdown and after CRISPR-mediated disruption of *CCDC47* or *WDR83OS* (Extended Data Fig. 7). The PAT complex is therefore required for optimal biogenesis of various multi-spanning membrane proteins. This explains why both genes were hits in a recent screen for factors that impair surface expression of a mutant TRP6 channel[20].

Neither SQS nor the type II single-pass membrane protein asialyloglycoprotein receptor (ASGR) were affected in CCDC47- or Asterix-knockout cells using the same assay (Extended Data Fig. 7) under the same conditions. We therefore conclude that the TMD insertase EMC (used by SQS[21] and the first TMD of many GPCRs[15]) and the Sec61 complex (used by ASGR[22]) function correctly in cells that lack the PAT complex. In such cells, the levels of EMC, Sec61 and other biogenesis factors are not reduced (Extended Data Fig. 8).

Appending a signal sequence and T4 lysozyme to the N terminus of AGTR2, which bypasses its requirement for EMC[15], does not affect its dependence on the PAT complex (Fig. 3, Extended Data Fig. 7). Similarly, the EMC-independent substrate TRAM2 is dependent on the PAT complex, which suggests that EMC and the PAT complex facilitate membrane protein biogenesis in different ways. As expected for central ER factors, loss of either EMC or the PAT complex results in an activated unfolded protein response, as determined by increased levels of the ER chaperones BiP and PDI (Extended Data Fig. 8).

Hydrophilic TMDs engage the PAT complex

Analysis of substrate insertion and crosslinking in microsomes that lacked the PAT complex showed that the timing of TMD1 insertion, and interaction with and release from Sec61α, was not appreciably different from that in wild-type microsomes (Extended Data Fig. 9a). This contrasts with the failed TMD1 insertion phenotype seen when microsomes lack the EMC insertase[15]. The PAT complex therefore does

not participate in targeting and is neither an insertase nor a facilitator of TMD release from Sec61α, explaining why single-pass membrane proteins such as SQS and ASGR are unaffected by depletion of the PAT complex. Instead, the PAT complex acts after a TMD accesses the lipid bilayer.

Replacement of a polar amino acid (asparagine) in TMD1 with leucine markedly reduced the Asterix–TMD1 interaction at all nascent chain lengths examined (Fig. 4a, Extended Data Fig. 9b). Reintroduction of polar or charged residues at other positions in the TMD partially restored crosslinking to Asterix (Fig. 4b). Analysis of substrate interaction with the PAT complex via CCDC47 native immunoprecipitation experiments led to the same conclusion (Fig. 4c). These data indicate that Asterix engages TMDs that expose hydrophilic residues within the lipid bilayer in a manner that is mostly independent of the position of these residues. This explains observations that PAT10 can crosslink with TMDs of either topology and of different unrelated sequences[10].

The PAT complex releases folded substrates

Although the PAT complex initially engages TMDs co-translationally as a peptidyl-tRNA, the complex remains associated even after termination releases the nascent chain from the ribosome–Sec61 complex (Fig. 4d). This observation—together with the finding that exposed hydrophilic residues within a TMD influence interaction with the PAT complex—suggested that dissociation of the PAT complex might be triggered when TMDs pack correctly to shield exposed hydrophilicity. To test this, we determined whether PAT complex interaction is selective for immature rather than folded membrane proteins. Exploiting the fact that an engineered β_1AR folds efficiently after in vitro translation[15], we compared its interaction with the PAT complex relative to that of truncation products that were intended to mimic biogenesis intermediates. A pooled mixture of intermediates truncated after each TMD was produced by in vitro translation and subjected to native immunoprecipitation using anti-CCDC47 antibodies. The amount of full-length β_1AR recovered was substantially lower than that of each of the intermediates, suggesting that the PAT complex dissociates after the correct folding of β_1AR (Fig. 4e). A similar conclusion was reached using a site-specific photo-crosslinker installed within TMD1 of β_1AR (Extended Data Fig. 10).

Discussion

Our findings indicate that the PAT complex is a widely conserved ER-resident complex that directly interacts with nascent TMDs, releases its substrates upon correct folding, and is needed for optimal biogenesis of membrane proteins. These features define the PAT complex as an intramembrane chaperone. Its preference for TMDs with exposed hydrophilic amino acids within the lipid bilayer is analogous to the preference of soluble chaperones for hydrophobic patches that are exposed to the aqueous environment[23,24]. The more favourable TMD–TMD interactions that accompany correct folding probably displace the PAT complex, which must necessarily interact with TMDs in a more generic and therefore weaker manner. Just as the binding clefts of chaperones such as Hsp70 are hydrophobic[25], the substrate-interaction domain of Asterix might be hydrophilic, which would be consistent with a large number of conserved hydroxylated amino acids in its TMDs.

Analogous intramembrane interactions with hydrophilic TMDs have been proposed to facilitate recognition of misfolded membrane proteins during quality control[26–28]. By binding to and shielding similar motifs in a nascent protein, the highly abundant PAT complex[29] probably prevents promiscuous degradation. This would explain why its depletion leads to post-translational reduction of various multi-spanning membrane proteins. Loss of CCDC47 in mice causes a wide range of developmental problems that lead to lethality during embryogenesis[30]. Similarly, *CCDC47* mutations in humans cause numerous developmental

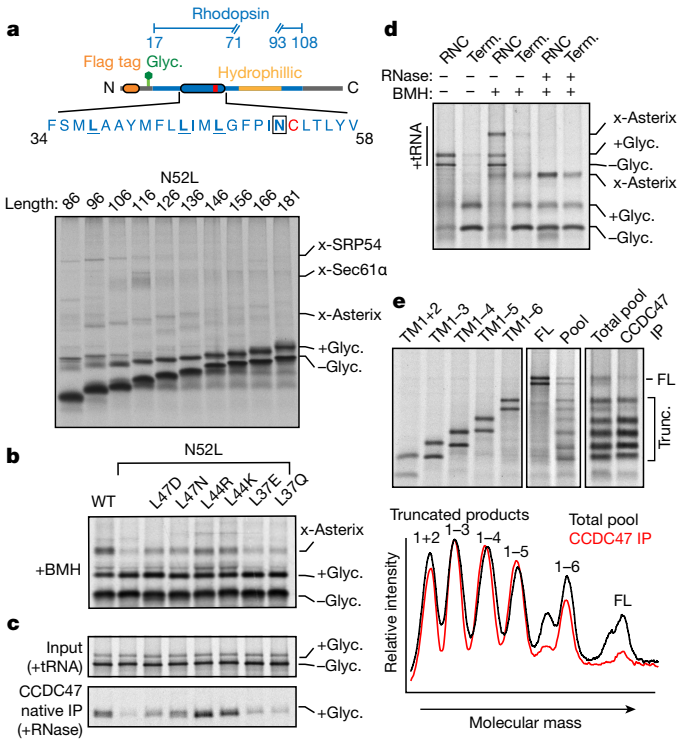


Fig. 4 | The PAT complex engages TMDs via exposed polar residues. a, The Rho TM1 construct (diagram) containing the asparagine-to-leucine mutation at residue 52 (N52L) was analysed for crosslinking as in Fig. 1a. Compared to the identical construct without the N52L mutation (Extended Data Fig. 9b), crosslinks to Asterix are markedly diminished and strong crosslinks to Sec61α are seen only for the 116-mer. **b**, Crosslinking reactions of 146-mer RNCs of Rho TM1 containing the indicated mutations. **c**, Translation reactions as in **b** (but without crosslinking) were either analysed directly (input) or subjected to native immunoprecipitation using anti-CCDC47 antibodies. Nascent chains were released from the ribosome with RNase A before immunoprecipitation. The glycosylated substrate recovered with CCDC47 was visualized by autoradiography. The efficiencies of crosslinking to Asterix in **b** correlate to the efficiencies of recovery with CCDC47. **d**, Terminated (Term.) or truncated (RNC) Rho TM1+2 was inserted into rough microsomes and treated with BMH where indicated. Before SDS–PAGE, some of the samples were digested with RNase A as indicated to remove the tRNA. **e**, Full-length β1AR (FL) or constructs truncated after each TMD were translated in the presence of rough microsomes and analysed either individually (left) or pooled before analysis. All constructs contained a stop codon and are terminated. The membrane-targeted population from the pooled reaction was isolated by sedimentation and divided in two aliquots. One aliquot was set aside ('Total pool') and the other was used for a native immunoprecipitation with anti-CCDC47 antibody. Both samples were then subjected to denaturing pulldown via the C-terminal His tag to ensure that only completed translation products were visualized. The graph presents the results of scanning densitometry on the last two lanes. Substantially less full-length β1AR is recovered with CCDC47 relative to each of the truncation products. For gel source data, see Supplementary Fig. 1.

phenotypes with early death[31]. In several mammalian cell lines, cells that lack either subunit of the PAT complex show reduced fitness[32]; however, they are viable, show activated ER stress responses and have several otherwise unrelated membrane-associated phenotypes[30]. These observations are consistent with loss of an intramembrane chaperone radiating to numerous cellular and organismal functions.

Despite such pleiotropic phenotypes, disruption of individual chaperones often has only modest consequences at a cellular level owing to their considerable redundancy[32]. There could therefore be many other intramembrane chaperones that could partially compensate

for the PAT complex in its absence. Although several substrate-specific chaperones have been proposed for particular membrane proteins[33–35], our findings suggest that the PAT complex is an abundant generalist for semi-hydrophilic TMDs that are found in nearly all of the more than 2,500 multi-spanning human membrane proteins. Furthermore, the PAT complex—or factors like it—might have a similar role in chaperoning integral membrane subunits of multi-protein complexes.

Online content

Any methods, additional references, Nature Research 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/s41586-020-2624-y.

1. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **46**, D158–D169 (2018).
2. von Heijne, G. The membrane protein universe: what's out there and why bother? *J. Intern. Med.* **261**, 543–557 (2007).
3. Shao, S. & Hegde, R. S. Membrane protein insertion at the endoplasmic reticulum. *Annu. Rev. Cell Dev. Biol.* **27**, 25–56 (2011).
4. Lu, P. et al. Accurate computational design of multipass transmembrane proteins. *Science* **359**, 1042–1046 (2018).
5. Harrington, S. E. & Ben-Tal, N. Structural determinants of transmembrane helical proteins. *Structure* **17**, 1092–1103 (2009).
6. Zhou, F. X., Cocco, M. J., Russ, W. P., Brunger, A. T. & Engelman, D. M. Interhelical hydrogen bonding drives strong interactions in membrane proteins. *Nat. Struct. Biol.* **7**, 154–160 (2000).
7. Venkatakrishnan, A. J. et al. Molecular signatures of G-protein-coupled receptors. *Nature* **494**, 185–194 (2013).
8. Choma, C., Gratkowski, H., Lear, J. D. & DeGrado, W. F. Asparagine-mediated self-association of a model transmembrane helix. *Nat. Struct. Biol.* **7**, 161–166 (2000).
9. Shurtleff, M. J. et al. The ER membrane protein complex interacts cotranslationally to enable biogenesis of multipass membrane proteins. *eLife* **7**, e37018 (2018).
10. Meacock, S. L., Lecomte, F. J. L., Crawshaw, S. G. & High, S. Different transmembrane domains associate with distinct endoplasmic reticulum components during membrane integration of a polytopic protein. *Mol. Biol. Cell* **13**, 4114–4129 (2002).
11. Park, P. S.-H. Rhodopsin oligomerization and aggregation. *J. Membr. Biol.* **252**, 413–423 (2019).
12. Estabrooks, S. & Brodsky, J. L. Regulation of CFTR biogenesis by the proteostatic network and pharmacological modulators. *Int. J. Mol. Sci.* **21**, 452 (2020).
13. Keenan, R. J., Freymann, D. M., Stroud, R. M. & Walter, P. The signal recognition particle. *Annu. Rev. Biochem.* **70**, 755–775 (2001).
14. Rapoport, T. A. Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* **450**, 663–669 (2007).
15. Chitwood, P. J., Juszkiewicz, S., Guna, A., Shao, S. & Hegde, R. S. EMC is required to initiate accurate membrane protein topogenesis. *Cell* **175**, 1507–1519.e16 (2018).
16. Chitwood, P. J. & Hegde, R. S. The role of EMC during membrane protein biogenesis. *Trends Cell Biol.* **29**, 371–384 (2019).
17. Ismail, N., Crawshaw, S. G. & High, S. Active and passive displacement of transmembrane domains both occur during opsin biogenesis at the Sec61 translocon. *J. Cell Sci.* **119**, 2826–2836 (2006).
18. Lin, Z. et al. TTC5 mediates autoregulation of tubulin via mRNA degradation. *Science* **367**, 100–104 (2020).
19. Itakura, E. et al. Ubiquilins chaperone and triage mitochondrial membrane proteins for degradation. *Mol. Cell* **63**, 21–33 (2016).
20. Talbot, B. E., Vandorpe, D. H., Stotter, B. R., Alper, S. L. & Schlondorff, J. S. Transmembrane insertases and *N*-glycosylation critically determine synthesis, trafficking, and activity of the nonselective cation channel TRPC6. *J. Biol. Chem.* **294**, 12655–12669 (2019).
21. 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).
22. Görlich, D. & Rapoport, T. A. Protein translocation into proteoliposomes reconstituted from purified components of the endoplasmic reticulum membrane. *Cell* **75**, 615–630 (1993).
23. Döring, K. et al. Profiling Ssb-nascent chain interactions reveals principles of Hsp70-assisted folding. *Cell* **170**, 298–311.e20 (2017).
24. Stein, K. C., Kriel, A. & Frydman, J. Nascent polypeptide domain topology and elongation rate direct the cotranslational hierarchy of Hsp70 and TRiC/CCT. *Mol. Cell* **75**, 1117–1130. e5 (2019).
25. Mayer, M. P. & Bukau, B. Hsp70 chaperones: cellular functions and molecular mechanism. *Cell. Mol. Life Sci.* **62**, 670–684 (2005).
26. 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).
27. Sato, K., Sato, M. & Nakano, A. Rer1p, a retrieval receptor for ER membrane proteins, recognizes transmembrane domains in multiple modes. *Mol. Biol. Cell* **14**, 3605–3616 (2003).
28. Natarajan, N., Foresti, O., Wendrich, K., Stein, A. & Carvalho, P. Quality control of protein complex assembly by a transmembrane recognition factor. *Mol. Cell* **77**, 108–119.e9 (2020).
29. Itzhak, D. N., Tyanova, S., Cox, J. & Borner, G. H. Global, quantitative and dynamic mapping of protein subcellular localization. *eLife* **5**, e16950 (2016).
30. Yamamoto, S. et al. Contribution of calumin to embryogenesis through participation in the endoplasmic reticulum-associated degradation activity. *Dev. Biol.* **393**, 33–43 (2014).
31. Morimoto, M. et al. Bi-allelic CCDC47 variants cause a disorder characterized by woolly hair, liver dysfunction, dysmorphic features, and global developmental delay. *Am. J. Hum. Genet.* **103**, 794–807 (2018).
32. Wang, T. et al. Identification and characterization of essential genes in the human genome. *Science* **350**, 1096–1101 (2015).
33. Kota, J. & Ljungdahl, P. O. Specialized membrane-localized chaperones prevent aggregation of polytopic proteins in the ER. *J. Cell Biol.* **168**, 79–88 (2005).
34. Gu, S. et al. Brain α7 nicotinic acetylcholine receptor assembly requires NACHO. *Neuron* **89**, 948–955 (2016).
35. Brechet, A. et al. AMPA-receptor specific biogenesis complexes control synaptic transmission and intellectual ability. *Nat. Commun.* **8**, 15910 (2017).

Methods

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Constructs

Sequences encoding 1×Flag-Rho TM1+2 and 1×Flag-Rho TM1+2 (F53C) were ordered as gene blocks (IDT) and inserted into a parent vector containing an SP6 promoter and the flexible N terminus of Sec61β with a 6×His tag appended to the C terminus. A glycosylation acceptor site (NGT) was introduced at amino acids 13–16. Variants of these parent constructs with a TwinStrep tag, HA tag, or lacking a glycosylation site were generated by standard subcloning methods. Constructs encoding Rho TM1, in which TM2 is replaced by a flexible hydrophilic linker (SGSGSGSGSGGSSGGMGGSGS), were ordered as gene blocks containing a 5′ SP6 promoter and transcribed directly for in vitro translation. Similar methods were used for all Rho TM1+linker polar residue point mutants. The in vitro expression plasmid encoding HA-β_1AR containing a C-terminal His tag has been described previously[15]. Truncated versions and variants containing amber codons at specific sites were created by PCR-based cloning methods and site-directed mutagenesis, respectively, and verified by sequencing. Constructs encoding the mutant *Escherichia coli* tyrosyl-tRNA synthetase in the pET21 vector and *Bacillus stearothermophilus* suppressor $tRNA^{Tyr}$ sequence in the pRSET-A vector have been described[18]. Templates for in vitro translation of tagged Asterix constructs were ordered as gene blocks containing a 5′ SP6 promoter and used directly for transcription and translation. Mammalian expression constructs for C-terminally Flag-tagged human Asterix and the various cysteine variants were produced as gene blocks, sub-cloned into pcDNA3.1-based vectors, and verified by sequencing. For the creation of stable cell lines expressing various membrane protein reporters, the coding sequences for CHRM1 (NP_000729.2), bovine rhodopsin (NP_001014890.1), Anoctamin-6 (NP_001020527.2), and MTR1L (NP_004215.2) were PCR-amplified and placed into a parent pcDNA5/FRT/TO vector backbone with a C-terminal GFP-P2A-RFP using Gibson Assembly (NEB). Analogous constructs encoding ASGR, SQS, TRAM2, β_1AR, AGTR2 and SS-T4L-AGTR2 have previously been described[15]. CRISPR–Cas9 mediated gene disruptions of *CCDC47* and *WDR83OS* were performed using the pSpCas9(BB)-2A-Puro (PX459) plasmid (Addgene) containing the guide RNAs 5′-GTATGGACTG CCGGACTCTT-3′ (CCDC47) and 5′-AAGGCCGGGTT ACATTCGCT-3′ (Asterix).

Antibodies

CCDC47 antibody 1 (A305-100A) and CCDC47 antibody 2 (A305-101A) were obtained from Bethyl Laboratories. Flag immunoprecipitations were performed using Flag-M2 affinity gel (Sigma). Anti-HA antibody was generated in house[19]. Signal peptide peptidase (SPP) antibody was purchased from Bethyl Laboratories (A304-404A). Anti-Asterix (WDR83OS) antibody was purchased from Sigma/Human Protein Atlas (HPA065685). Anti-UBQLN2 antibody was clone 5F5 obtained from Sigma (WH0029978M3). Anti-SRP54 was from BD Biosciences (610940). Custom antibodies against Sec61α, Sec61β, TRAPα and TRAM have previously been described and characterized[36]. All antibodies for western blotting in Extended Data Fig. 8 have previously been described[15].

Cell lines

Flp-In TRex 293 cells (Invitrogen) were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal calf serum (FCS). The authentication of this cell line was ensured by the antibiotic resistance markers within its genome and by its unique FRT site downstream of a doxycycline-inducible promoter. Cell lines were tested monthly for mycoplasma contamination and verified to be negative. For cell lines containing a stably expressed doxycycline-induced reporter, tetracycline-free FCS was used in conjunction with 15 μg ml^{-1} blasticidin and 100 μg ml^{-1} hygromycin. Stable cells lines were generated using the FRT Flp-In system according to the manufacturer's protocol (Invitrogen). Stably expressed cell lines included GFP-2A-RFP-SQS, β_1AR-GFP-2A-RFP, AGTR2-GFP-2A-RFP, SS-T4L-AGTR2-GFP-2A-RFP, TRAM2-GFP-2A-RFP, Ano6-GFP-2A-RFP, CHRM1-GFP-2A-RFP, MTR1L-GFP-2A-RFP and Rhodosin-GFP-2A-RFP. CCDC47- and Asterix-knockout cells were generated by transient transfection with the px459 plasmids containing the appropriate guide RNAs. After 24 h, transfected cells were selected for 48 h with 1 μg ml^{-1} puromycin then trypsinized and diluted into 96-well plates at a concentration of 0.5 cells per well to select for single-cell colonies. After around 2 weeks, single colonies were expanded and clones displaying undetectable CCDC47 or Asterix were chosen for further analysis.

Flow cytometry analysis

For experiments using knockout cell lines, ΔCCDC47 and ΔAsterix cells were transiently transfected with 1 μg ml^{-1} of the appropriate pcDNA5 expression constructs approximately 24 h before induction with 100 ng ml^{-1} doxycycline. All transfections were performed using Mirus-Transit 293 according to the manufacturer's instructions. For experiments using stably expressed reporter cell lines, siRNA depletion was performed over period of approximately 96 h using the Lipofectamine RNAiMAX reagent according to the manufacturer's instructions (Thermo). In brief, a first round of siRNA treatment was performed in the presence of DMEM and 10% tetracycline-free FCS. Cells were incubated for 48 h before a second round of siRNA treatment was performed under the exact same conditions. After a second incubation of approximately 48 h, expression of fluorescent reporter constructs was induced with 100 ng ml^{-1} doxycycline for 6 h before analysis by flow cytometry. Acute expression of reporters was essential to accurately monitor degradation of reporter constructs and avoid saturation of degradation pathways. In all experiments the cells were collected by trypsinization, washed once in ice-cold PBS, then resuspended in the equivalent culture volume of PBS and 1 μg ml^{-1} DAPI stain (Thermo). Cells were passed through a 70-μm filter before flow cytometry analysis using a Beckton Dickinson LSRII instrument. A total of 20,000 GFP-positive (or RFP-positive for SQS and ASGR) cells were collected. Further gating for live cells (negative for DAPI stain) and relatively high levels of the soluble fluorescent protein reporter was used to focus on the population of cells with productive translation of reporter constructs.

Preparation of ER-derived rough microsomes

Rough microsomes (RMs) derived from adherent wild-type, ΔCCDC47 or ΔAsterix HEK293 cells were prepared as previously described[15]. Wild-type RMs were also obtained from HEK293 cells grown in suspension (Expi293F cells) and prepared similarly to RMs from adherent cells with a few minor modifications to adjust for the larger scale. In brief, around 2 l of cells were grown to a concentration of 5×10^6 cells per ml and then collected by centrifugation. Cell pellets were washed twice with ice-cold PBS and pooled as necessary. A pellet (approximately 30 ml) was resuspended in 60 ml of sucrose buffer (50 mM HEPES, pH 7.4, 50 mM KOAc, 6 mM $Mg(OAc)_2$, 1 mM EDTA, 250 mM sucrose, 1 mM DTT) and lysis was carried out in a glass dounce homogenizer. Lysate was cleared twice by centrifugation at 3,500*g* for 30 min at 4 °C. Supernatant was recovered and underlaid with one-third the volume of sucrose cushion (1.3 M sucrose, 50 mM HEPES, pH 7.4, 50 mM KOAc, 6 mM $Mg(OAc)_2$, 1 mM EDTA, 1 mM DTT) and centrifuged for 1 h at 371,000*g* (60,000 rpm) and 4 °C for 1 h in a Type 70Ti rotor (Beckman). The supernatant was removed by aspiration, and the pellets were resuspended and pooled by manual homogenization in a dounce using 6 ml resuspension buffer (250 mM sucrose, 50 mM HEPES, pH

7.4, 1 mM DTT). The final preparation was adjusted to an absorbance of 75 when measured at 280 nm in 1% SDS. Rough microsomes derived from canine pancreas (cRMs) were prepared as previously described[37], and were used in very few experiments where explicitly stated in the figure legends. All microsome preparations were flash-frozen in liquid nitrogen and stored at −80 °C.

In vitro transcription and translation

In vitro transcription was performed with SP6 polymerase using PCR products as the template[38,39]. The transcription reactions were conducted with 5–20 ng ml^{-1} PCR product at 37 °C. In vitro translation reactions were performed in rabbit reticulocyte lysate (RRL) as previously described in detail[38,39]. All translation reactions were carried out at 32 °C for 15–30 min. For translation reactions in the presence of RMs, 0.25–1.0 µl RMs (at an absorbance of 75) were added to a 10 µl translation reaction. All batches of microsomes were individually titrated to achieve optimal translation and insertion as monitored by glycosylation efficiency. For pooled translation reactions in Fig. 4e, initial experiments provided the relative translation efficiency of each intermediate. The levels of each transcript were titrated accordingly to achieve roughly equal translation of each individual intermediate and the full-length product.

Generation of ribosome-nascent chain complexes

Truncated mRNA templates were generated by PCR using a 5′ primer that annealed upstream of the SP6 promoter and a 3′ primer that annealed at the site of truncation to generate the desired nascent chain length. The 3′ primer also encoded the residues MLKV at the very C terminus to improve radiolabelling (via the added methionine) and minimize hydrolysis of the tRNA-peptidyl bond during electrophoresis[40]. For all experiments, RMs were included during translation, which was carried out at 32 °C for 15–20 min, then returned to ice for subsequent crosslinking and protease-protection analysis.

Site-specific photo-crosslinking

Incorporation of benzoyl-phenylalanine (BPA) at specific positions during in vitro translation was accomplished by amber suppression as described[18]. In short, an amber codon containing template was translated in RRL as above but with 5 µM *B. stearothermophilus* tRNA^{Tyr}, 0.25 µM BPA tyrosyl-tRNA synthetase and 0.1 mM BPA. UV irradiation (12 min, 254 nm) was carried out with a hand-held UV lamp (Thermo Fisher Scientific, 95034) positioned about 2.5 cm above the sample sitting on ice. As previously described[18], *E. coli* BPA tyrosyl-tRNA synthetase was purified via the C-terminal His tag on a 5 ml HiTrap Ni-NTA column (GE Healthcare), desalted using a gel filtration column by fast protein liquid chromatography and concentrated using an Amicon Ultra centrifugal filter (Millipore, Z717185-8EA). *B. stearothermophilus* tRNA^{Tyr} was synthesized by in vitro transcription as previously described[18]. The pRSET-based construct encoding the tRNA was digested with BstN1, yielding a DNA fragment containing the exact tRNA^{Tyr} sequence under a T7 promoter. The transcription reaction (on a 5 ml scale) was carried out with 1.2 mg DNA template, 1 mM spermidine, 5 mM DTT, 0.1% Triton, 5 mM NTPs, 25 µM $MgCl_2$, 20 µg ml^{-1} *E. coli* pyrophosphatase, 20 µg ml^{-1} T7 polymerase and 125 U Recombinant RNasin (Promega) for 4 h at 37 °C. The reaction product was digested with Turbo DNase (Ambion) and extracted by acid phenol/chloroform extraction to yield purified tRNA.

Cysteine-based crosslinking

Unless explicitly stated, the crosslinking reactions were performed on isolated RMs. The translation reactions were placed over a 20% sucrose cushion in physiological salt buffer (PSB; 100 mM KOAc, 50 mM HEPES pH 7.4, 2.5 mM $Mg(OAc)_2$), centrifuged in a TLA-55 rotor for 20 min at 4 °C, and the pelleted RMs were resuspended in one half of the original translation reaction volume of PSB. BMH was added to a final concentration of 250 µM and the reaction was incubated on ice for 1 h to allow the crosslinking reaction to occur. BMH was quenched with 5 mM 2-mercaptoethanol. Aliquots of the reaction were removed at different stages of the process for analysis, as indicated in the individual figure legends. The samples were either used directly for downstream applications (primarily immunoprecipitations, as described in 'Immunoprecipitation') or flash-frozen in liquid nitrogen and stored at −80 °C for later analysis.

Large-scale affinity purification of crosslinked substrate complexes

Translation reactions (1 ml) in the presence of 0.8 mM cold methionine were carried out to produce 1×Flag-tagged or TwinStrep-tagged (as a control) Rho TM1+2 146mer RNCs. Membranes were isolated and cysteine-based crosslinking was carried out as described in the previous section. Isolated membranes resuspended in PSB were solubilized on ice by addition of an equal volume of 2× solubilization buffer (300 mM KOAc, 2% deoxyBigChap, 20 mM EDTA). RNase A was added to a final concentration of 10 ng ml^{-1} and RNA digestion was carried out on ice for 20 min before the solubilized extracts were cleared by centrifugation at 100,000 rpm (434,513*g*) in the TLA120.2 (Beckman). Cleared extracts were directly immunoprecipitated with Flag M2 affinity resin (Sigma) in batch at 4 °C with end-over-end rotation for 2.5 h. The unbound fraction was removed and beads were washed 4 times in 1× native solubilization buffer (NSB) (200 mM KOAc, 1% DBC, 10 mM EDTA, 25 mM HEPES, 1 mM $Mg(OAc)_2$), then twice in 1× NSB without detergent. Beads were resuspended in 1.5 times the bead volume of 200 mM KOAc and 25 mM HEPES pH 7.4 for direct trypsinization and analysis by quantitative mass spectrometry.

Quantitative mass spectrometry using tandem mass tag labelling

Proteins samples on beads were reduced with 5 mM DTT at 56 °C and alkylated with 10 mM iodoacetamide in the dark at 22 °C. The alkylation reaction was quenched by the addition of DTT and the samples were digested overnight with trypsin (Promega, 0.1 µg) at 37 °C. After digestion, each supernatant was transferred to a fresh Eppendorf tube, the bead samples were extracted once with 50% acetonitrile/0.1% TFA and combined with the corresponding supernatant. The peptide mixtures were then partially dried in a SpeedVac vacuum concentrator and desalted using a homemade C18 (3M Empore) stage tip filled with 2 µl of poros R3 (Applied Biosystems) resin. Bound peptides were eluted sequentially with 30%, 50% and 80% acetonitrile in 0.1% TFA and lyophilized. Dried peptide mixtures from each condition were re-suspended in 20 µl of 7% MeCN and 1 M triethyl ammonium bicarbonate was added to a final concentration of 200 mM. TMT10plex reagents (0.8 mg, Thermo Fisher Scientific) were re-constituted in 41 µl anhydrous MeCN. TMT (10 µl, 130C or 131) reagent was added to each peptide mixture and incubated for 1 h at 20 °C. The labelling reactions were terminated by incubation with 2.5 µl of 5% hydroxylamine for 15 min. The labelled samples were pooled into one Eppendorf tube and the SpeedVac vacuum concentrator was used to evaporate acetonitrile. Peptides were separated on an Ultimate 3000 RSLC nano System (Thermo Scientific), using an acetonitrile gradient, consisting of buffer A (2% MeCN, 0.1% formic acid) and buffer B (80% MeCN, 0.1% formic acid). Eluted peptides were introduced directly via a nanospray ion source into a Q Exactive Plus hybrid quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific). The mass spectrometer was operated in standard data-dependent mode, a survey full-scan (MS, m/z = 380–1,600) was performed with a resolution of 70,000, followed by MS2 acquisitions of the 15 most intense ions with a resolution of 35,000 and normalized collision energy of 33%. MS target values of 3×10^6 and MS2 target values of 1×10^5 were used. The isolation window was set as 0.7 m/z and dynamic exclusion was enabled for 40 s. The acquired MS/MS raw files were processed using Proteome Discoverer (v.2.1, Thermo Scientific).

MS/MS spectra were searched against a human-reviewed, UniProt Fasta database (downloaded in 2016), using Mascot (v.2.4, Matrix Science) search engine. Carbamidomethylation of cysteines, TMT6plex (N-term) and TMT6plex (K) were set as fixed modifications, while methionine oxidation and N-terminal acetylation (protein) were selected as variable modifications. For reporter ion quantification, the co-isolation threshold is 30 and the average reporter signal-to-noise threshold is 10. An output file was generated from Proteome Discoverer, the protein table was filtered to a protein false discovery rate of 1% and exported as an Excel file to produce the plot in Fig. 1c.

Preparation of anti-CCDC47 affinity columns

A 50-µl bead volume of protein A agarose was diluted in 750 µl PBS, placed in 1-ml Pierce Spin columns, then pre-equilibrated with two 800-µl PBS washes. Then, 100 µg CCDC47 antibody 1 (A305-100A), CCDC47 antibody 2 (A305-101A) or an anti-HA antibody were diluted in 1 ml of PBS then placed over the prepared protein A columns and allowed to pass through the resin by gravity flow. Flow through was collected and passed over the column a second time. Columns were equilibrated with 800 µl of 0.1 M Na borate pH 9.0, then antibodies were conjugated to resin by passing through a 1 ml solution by gravity flow of 0.1 M Na borate pH 9.0 containing 5 mg ml^{-1} dimethyl pimelimidate (DMP). DMP was quenched with 1 ml of 0.2 M ethanolamine pH 8.0 and columns were re-equilibrated in 1× PBS and 0.02% NaN_3 for storage at 4 °C until use.

Native affinity purification of CCDC47

The purification was performed at the bench on ice. CCDC47 and HA affinity columns were pre-washed with 300 µl of 0.1 M glycine-HCl pH 2.3, neutralized with 800 µl of 1× PBS, then equilibrated in 800 µl of 1× NSB: 200 mM KOAc, 10 mM EDTA, 1% DBC, 25 mM HEPES pH 7.4. RMs (3 ml), prepared as described above (at an absorbance of 75 at 280 nm), were solubilized by addition of an equal volume of 2× NSB. RNase A was added to a final concentration of 10 ng ml^{-1} and the samples were incubated on ice for 30 min during column preparation. The solubilized extract was cleared by centrifugation at 100,000 rpm (539,511*g*) in the TLA100.3 (Beckman) for 1 h at 4 °C. Supernatant was removed, divided into three equal parts, and passed twice over each column (anti-CCDC47 antibody 1, antibody 2 or an anti-HA control). Each column was washed once with 1 ml 1× NSB. Recovered proteins were eluted with 300 µl of 0.1 M glycine pH 2.3 and elutions were immediately neutralized with 1 M Tris-Cl pH 8.8. For initial small-scale experiments using only CCDC47 antibodies (Fig. 2a), the protocol was the exact same except 100 µl of starting RMs was used and final elutions were precipitated with trichloroacetic acid using standard procedures before loading onto a gel.

Identification of Asterix by mass spectrometry

Protein samples eluted from the CCDC47 affinity resin were reduced with DTT and alkylated with iodoacetamide. Because initial efforts to identify the co-purifying 10 kDa protein via analysis of tryptic digests failed, we re-did the analysis using other proteases. The samples were digested overnight either with trypsin, Glu-C, chymotrypsin or elastase (Promega). Digest mixtures were acidified with formic acid and a portion of each of these samples was analysed by nano-scale capillary liquid chromatography coupled to tandem mass spectrometry (LC–MS/MS) (Ultimate U3000 HPLC, Thermo Scientific Dionex) at a flow of 300 nl min^{-1}. A C18 Acclaim PepMap100 5 µm, 100 µm × 20 mm nanoViper (Thermo Scientific Dionex), trapped the peptides before separation on an EASY-Spray column with an acetonitrile gradient. The eluted peptides were introduced directly via an EASY-Spray ion source into a Q Exactive mass spectrometer (Thermo Scientific). Data-dependent analysis was performed using a resolution of 35,000 for the full mass spectrum, followed by ten MS/MS spectra in the Orbitrap. Mass spectra were collected over an m/z range of 350–1,600. LC–MS/MS data were searched against the UniProt KB database using Mascot, with a precursor tolerance of 10 ppm and a fragment ion mass tolerance of 0.1 Da. Two missed enzyme cleavages (or no enzyme for elastase) and variable modifications for oxidized methionine, carbamidomethyl cysteine and pyroglutamic acid were included. MS/MS data were validated using Scaffold (Proteome Software).

Immunoprecipitations

Unless otherwise indicated, all CCDC47 immunoprecipitations were performed with CCDC47 antibody 1. For immunoprecipitations under denaturing conditions, samples of interest were first denatured in 1% SDS and 100 mM Tris pH 8.0 and boiled for 2–5 min. Samples were diluted tenfold in IP buffer (100 mM NaCl, 50 mM HEPES pH 7.4, 1% TritonX-100) then immunoprecipitated in batch with desired antibodies at 4 °C rotating end-over-end for 2–4 h. Unbound supernatant was removed by aspiration and beads were washed three times in IP buffer before elution in 2× SDS–PAGE sample buffer. Native immunoprecipitations followed similar protocols but were performed in the presence of NSB (200 mM KOAc, 25 mM HEPES pH 7.4, 1% DBC). EDTA and RNase A was added before solubilization where RNase treatment is indicated in the figure legends. Pull-downs of His-tagged β_1AR intermediates and full-length β_1AR were performed using Ni-NTA affinity resin (Invitrogen) in 1× PBS +250 mM NaCl, 0.5% TritonX-100, 10 mM imidazole.

Sucrose gradient separation

The products of the quenched crosslinking reaction were solubilized in NSB (200 mM KOAc, 25 mM HEPES pH 7.4, 1% DBC). EDTA and RNase A were added to release the nascent chain from the ribosome and digest the attached tRNA. The sample (typically 10 or 20 µl volume) was loaded onto 200 µl micro sucrose gradient (5–25% sucrose in NSB), centrifuged at 55,000 rpm (258,488*g*) in the TLS-55 rotor (with suitable adaptors) for 2 h 20 min, and fractionated manually into 11 fractions from the top. The final fraction, which contains aggregates and non-solubilized material, was not analysed. Fractions 1–10 were analysed by SDS–PAGE and autoradiography (or Coomassie staining to detect endogenous proteins).

Protease protection assays

Immediately after the translation reaction, the samples were placed on ice and 10% of the reactions were set aside for analysis by SDS–PAGE and autoradiography of total products. The remainder was subjected to protease digestion by the addition of proteinase K (PK) to a final concentration of 0.5 mg ml^{-1} and incubated on ice for 50 min. To stop the digestion reaction, PMSF was added to 5 mM, incubated on ice for 2–5 min, and the entire reaction transferred to 10 volumes of boiling 1% SDS, 100 mM Tris-Cl, pH 8.0. A 10% portion of the reaction was set aside for analysis by SDS–PAGE and autoradiography of total digestion. The remainder of the sample was prepared for immunoprecipitation as described in the section 'Immunoprecipitations'.

Topology mapping by cysteine accessibility

Asterix-knockout HEK293 T-REx cells were transfected with constructs encoding human Asterix tagged at the C terminus with a Flag tag, including variants lacking all cysteines or containing single cysteines as described in Extended Data Fig. 5. After 48 h, the cells from each well of a 6-well plate were washed once and collected in ice-cold PBS, pelleted by centrifugation, resuspended on ice in 500 µl permeabilization buffer (100 mM KOAc, 10 mM Tris, pH 7.4, 10 mM $MgCl_2$), and adjusted to 0.01% digitonin. After 5 min on ice, the cells were re-sedimented by centrifugation, washed once in permeabilization buffer lacking digitonin, sedimented again, and resuspended in 30 µl of PSB on ice. An aliquot was removed for the untreated sample, and the remainder was adjusted to 1 mM final concentration of 5,000 Da PEG-maleimide prepared in DMSO. After 1 h on ice, the reaction was quenched with 10 mM 2-mercaptoethanol, the cells recovered by centrifugation, washed once with 600 µl PSB, and finally prepared for analysis by SDS–PAGE and immunoblotting.

Statistics and reproducibility

No statistical analyses were applied to any of the data. The following figure panels show representative data from at least two independent biological replicates that showed similar results: Figs. 1b, f, 2a, d, 4a–d, Extended Data Figs. 1b, c, 2b, d, e, 3c, d, 4c, 6b, 8, 9b, 10b. The following figure panels show representative data from at least three independent biological replicates that showed similar results: Figs. 1a, d, e, 2b, c, f, 4e, Extended Data Figs. 2c, 3a, b, 5b, c, 7a, b, 9a. The following figure panels show representative data from at least four independent biological replicates that showed similar results: Fig. 3, Extended Data Fig. 6a. No attempts at replication failed. The mass spectrometry experiment in Fig. 1c was performed once but all identified proteins indicated in the figure were directly validated in follow-up experiments that are shown within the paper. The experiment in Extended Data Fig. 4e was performed once, but is internally controlled for both positive and negative results.

Reporting summary

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

Data availability

All data supporting the findings of this study are available within the Article files. Uncropped images of all gels and autoradiographs in the figures are provided in Supplementary Fig. 1.

36. 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).
37. Walter, P. & Blobel, G. Preparation of microsomal membranes for cotranslational protein translocation. *Methods Enzymol*. **96**, 84–93 (1983).
38. Sharma, A., Mariappan, M., Appathurai, S. & Hegde, R. S. In vitro dissection of protein translocation into the mammalian endoplasmic reticulum. *Methods Mol. Biol*. **619**, 339–363 (2010).
39. Feng, Q. & Shao, S. In vitro reconstitution of translational arrest pathways. *Methods* **137**, 20–36 (2018).
40. Shao, S., von der Malsburg, K. & Hegde, R. S. Listerin-dependent nascent protein ubiquitination relies on ribosome subunit dissociation. *Mol. Cell* **50**, 637–648 (2013).

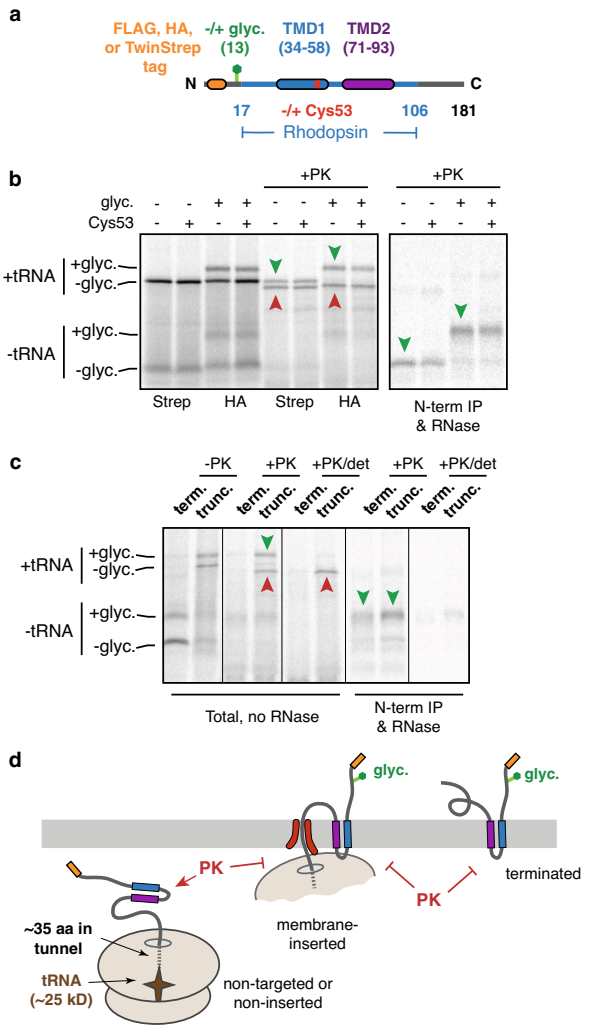
Acknowledgements We thank S.-Y. Peak-Chew, F. Begum and M. Skehel for mass spectrometry analysis; S. Juszkiewicz, J. O'Donnell and E. Zavodszky for discussions and advice; and N. Peters and H. Damstra for initial characterization of photo-crosslinking methods. This work was supported by the UK Medical Research Council (MC_UP_A022_1007 to R.S.H.) and a studentship (to P.J.C.) from the MRC International PhD Programme.

Author contributions P.J.C. and R.S.H. conceived the project, analysed the results and wrote the manuscript. P.J.C. performed all of the experimental work with advice from R.S.H.

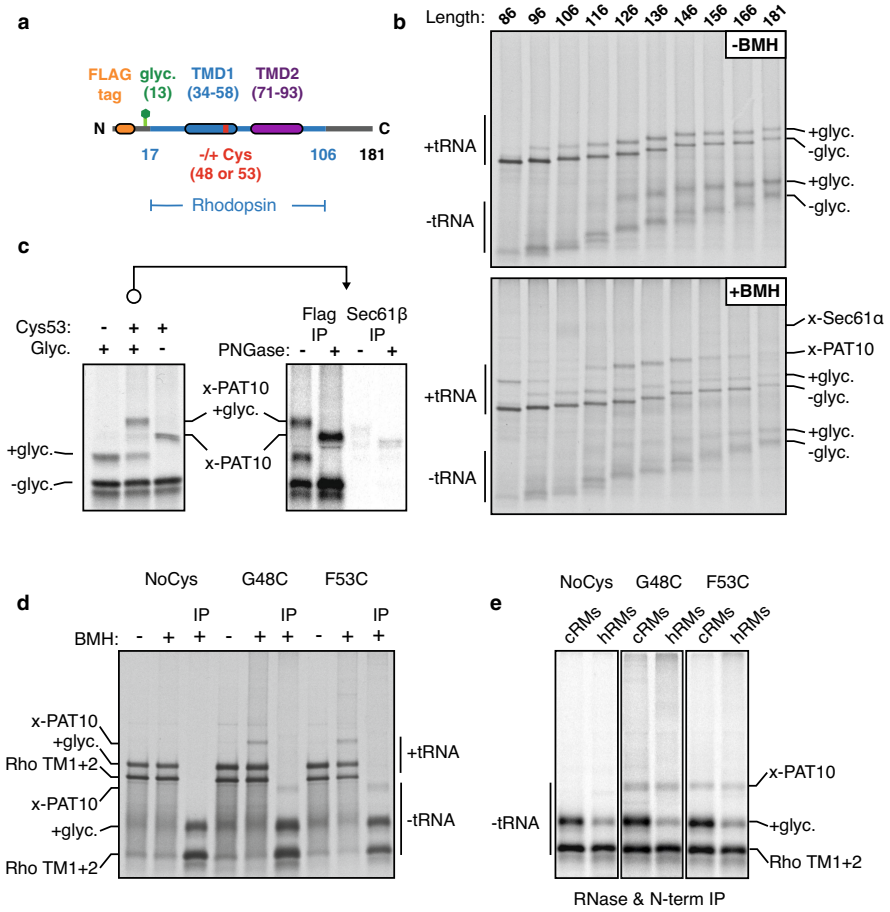
Competing interests The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2624-y.

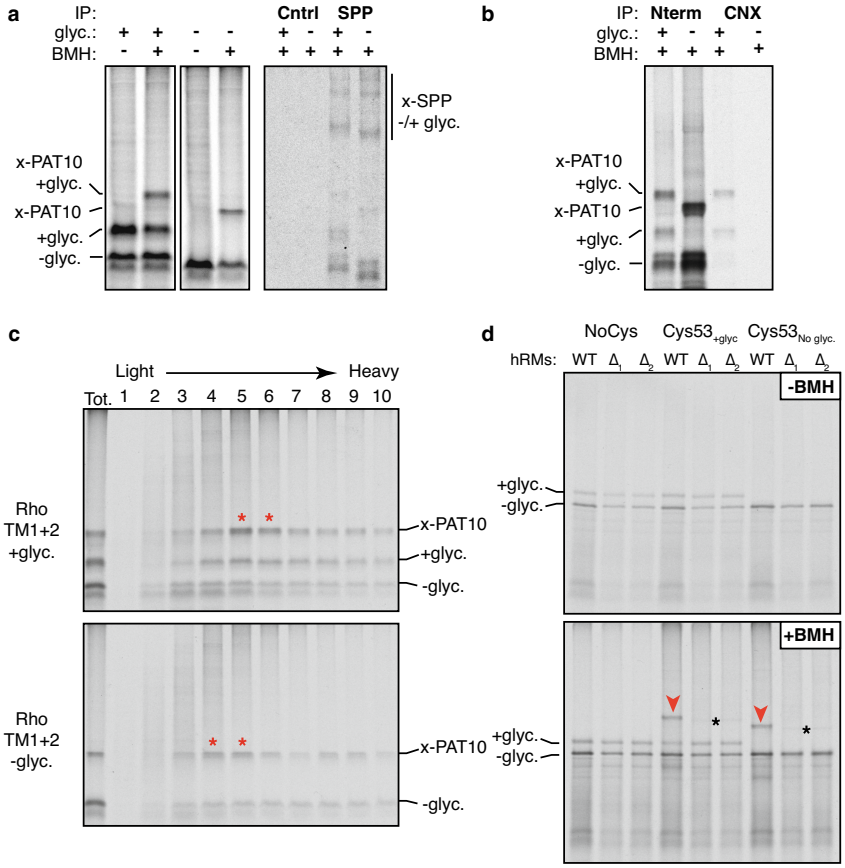
Correspondence and requests for materials should be addressed to R.S.H.



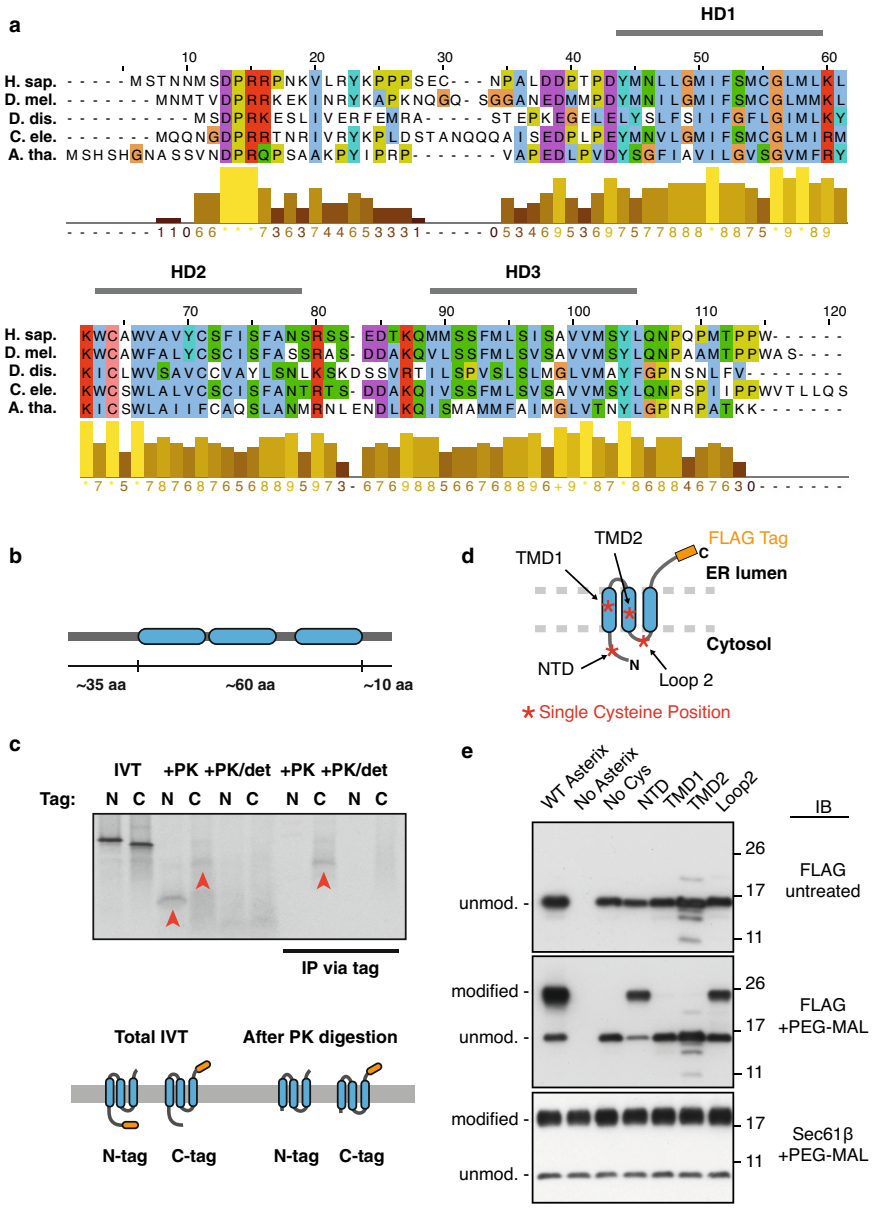
Extended Data Fig. 1 | Characterization of Rho TM1+2 insertion. a, Diagram of Rho TM1+2 constructs used throughout this study. Variations on this construct include different N-terminal epitope tags, the presence or absence of a glycosylation site near the N terminus, the presence, absence, or position of a cysteine within TMD1, various mutations within TMD1, and the presence or absence of TMD2. All of the constructs were tested either by protease protection or glycosylation to verify that no appreciable differences were observed in their insertion efficiencies. Although the exact amino acid numbering varies depending on the N-terminal tag, the numbering system corresponding to the Flag-tagged version is used throughout. Thus, the 146mer refers to a truncation at the 146th codon in the numbering scheme indicated in Fig. 1a and in this diagram, even in constructs containing a different tag. **b**, Representative example of insertion assays on two different tagged versions of Rho TM1+2. The TwinStrep tagged version (Strep) lacking a glycosylation site was compared to an HA-tagged version containing a glycosylation site. Identical constructs containing either the wild-type Rho TM1 sequence or a point mutant (F53C) were tested in parallel to confirm no insertion defects result from insertion of a cysteine in TM1 (used for BMH-mediated crosslinking in later experiments). In this experiment, [^{35}S] methionine labelled ribosome nascent chain complexes (RNCs) of 181 amino acids were produced by in vitro translation using rabbit reticulocyte lysate (RRL) in the presence of ER-derived RMs after which the microsomes were isolated and resuspended. Aliquots of the reactions were left untreated or digested with PK and analysed directly by SDS–PAGE and visualized by autoradiography (left). Green arrowheads represent the fully inserted and PK-protected population and red arrowheads denote the non-inserted and proteolytically cleaved products. The cleaved product contains the region of polypeptide protected by the ribosomal tunnel and the attached tRNA. Aliquots of the PK-digested sample were treated with EDTA and RNase to release the polypeptide from the ribosome and tRNA and immunoprecipitated (IP) via the N-terminal tag. Only the fully inserted products are recovered by IP (green arrowheads). **c**, Comparison of the topology of truncated RNCs and terminated Rho TM1+2. In this experiment, the Flag-tagged Rho TM1+2 containing a glycosylation site with (term.) or without (trunc.) a stop codon was translated in the presence of RMs, after which the RMs were isolated by centrifugation. Aliquots of the isolated RMs were analysed directly (−PK), after PK digestion (+PK), or after PK digestion in the presence of detergent (+PK/det). Where indicated, the +PK and +PK/det samples were released from the attached tRNA and immunoprecipitated via the N-terminal tag. Note comparable glycosylation near the N terminus and complete protection from PK for both the truncated and terminated products. **d**, Diagram representing the interpretation of the experiments in **b** and **c**. The relatively short cytosolic loop between TMD1 and TMD2 is not accessible to PK digestion either as an RNC or a terminated product. For gel source data, see Supplementary Fig. 1.



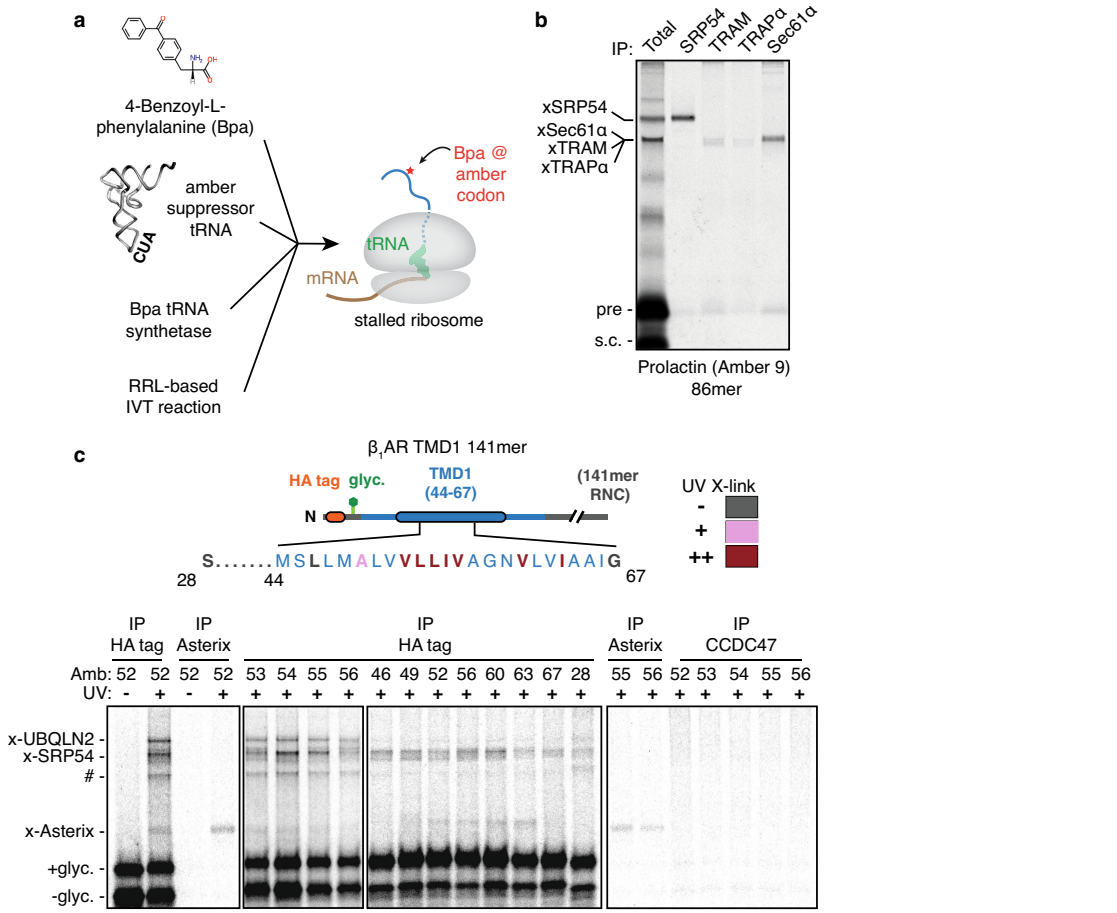
Extended Data Fig. 2 | Additional characterization of crosslinking to PAT10. a, Diagram of constructs either lacking or containing a cysteine in place of Gly48 or Phe53 within TMD1 of a Rho TM1+2 cassette. **b**, ^{35}S-labelled Flag-tagged Rho TM1+2 RNCs of varying lengths (as described in Fig. 1a) containing a Cys at position 53 were generated by in vitro translation in the presence of RMs. Membranes were isolated by centrifugation through a sucrose cushion and resuspended in physiological salt buffer (PSB). An equivalent amount of each translation reaction was taken before (−BMH) and after (+BMH) the addition of BMH for analysis by SDS–PAGE. The tRNA-linked nascent chains and free nascent chains are indicated. The free nascent chains arise from partial hydrolysis of the tRNA during electrophoresis under moderately basic conditions. Glycosylation is first observed at the 96mer length, which is 38 amino acids downstream of the end of TMD1. This matches the length of the ribosome tunnel, and indicates that membrane insertion and glycosylation occurs only after TMD1 is fully exposed outside the ribosome. Crosslinks to Sec61α are most prominent for the 106mer. Crosslinks to PAT10 are most prominent from the 126mer onwards, after the Sec61α crosslinks diminish. All of the crosslinked adducts are seen to the tRNA-attached nascent chain, verifying that they are co-translational. The Sec61α crosslink and others are not as visible when total translation products are analysed, which is why we typically immunoprecipitate the sample via the nascent chain (for example, in Fig. 1a). This reduces the background, allowing otherwise obscured crosslinks to be visualized clearly. Furthermore, we usually digest the samples with RNase after the experiment but before SDS–PAGE to remove the tRNA, thereby avoiding the heterogeneity that results from partial tRNA hydrolysis during sample handling and SDS–PAGE. All of the indicated crosslinking adducts observed were completely dependent on the presence of BMH. **c**, The indicated 'Input' crosslinking sample from Fig. 2b (reproduced here on the left) was subjected to immunoprecipitation using anti-Flag or anti-Sec61β antibodies under denaturing conditions. The IP samples were either left untreated or digested with PNGase F to remove N-linked glycans. Equivalent amounts were loaded in each lane. Sec61β is not an appreciable crosslinking partner of these RNCs, and to the extent a crosslink is observed, it migrates slightly faster than the PAT10 crosslink. **d**, The indicated Rho TM1+2 variants were translated in vitro in the presence of RMs and treated with BMH as indicated. In this experiment, the crosslinking was performed directly on total translation reactions, not after isolation of the microsome fraction. Instead, translation reactions were diluted fivefold with buffer to dilute the reduced glutathione and minimize quenching of BMH. An aliquot of each reaction was analysed directly by SDS–PAGE. Cross-linking efficiency is reduced compared to other experiments because membranes were not isolated by centrifugation through a sucrose cushion to remove reduced glutathione from translation extract. One aliquot of the BMH-treated translation reactions were treated with RNase A and EDTA, denatured, and IPed via the N-terminal Flag tag (IP). **e**, As in **d**, except Rho TM1+2 variants were generated in the presence of RMs derived from either canine pancreas (cRMs) or HEK293 cells (hRMs). In this experiment, the microsomes were isolated by centrifugation through a sucrose cushion before BMH crosslinking (note the higher crosslinking efficiency). Although the PAT10 crosslink is seen in both cRMs and hRMs, crosslinking efficiency of the inserted (glycosylated) product is appreciably lower in cRMs, which is one reason why we used HEK293-derived RMs for most of the experiments in this study. For gel source data, see Supplementary Fig. 1.



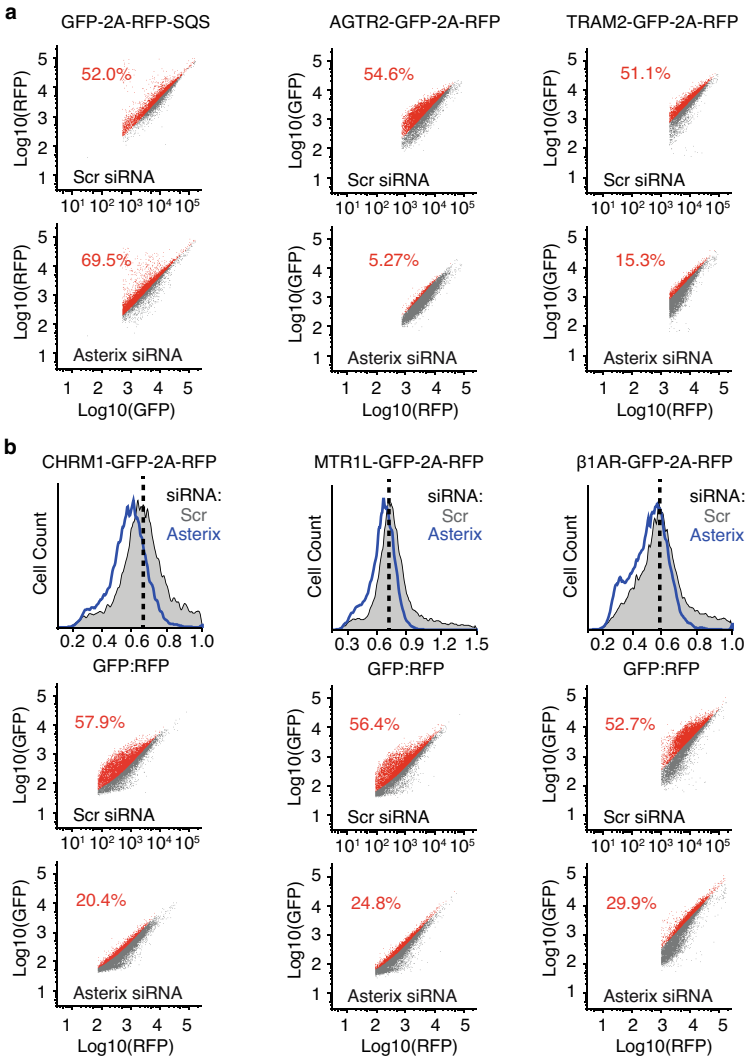
Extended Data Fig. 3 | Analysis of PAT complex candidates. a, Radiolabelled 146mer RNCs of Rho TM1+2 constructs containing or lacking a glycosylation site were generated by in vitro translation in the presence of RMs. Membranes were isolated by centrifugation through a sucrose cushion, subjected to BMH crosslinking where indicated, treated with RNase A, then analysed by SDS–PAGE. One aliquot of the BMH treated reactions were solubilized under native conditions and immunoprecipitated with either an antibody raised against the HA epitope tag (Cntrl) for a specificity control, or an antibody against signal peptide peptidase (SPP). Direct crosslinks to SPP are observed as two distinct adducts seen on very long exposures (compare to Fig. 1d, which is a shorter exposure), but native IPs do not enrich for a PAT10 engaged substrate. **b**, Radiolabelled 146mer RNCs of Flag-tagged Rho TM1+2 containing (+) or lacking (–) a glycosylation site (glyc.) were generated by in vitro translation with RMs, crosslinked with BMH, then immunoprecipitated under native conditions via the N-terminal Flag tag on the substrate (Nterm) or with an antibody recognizing the C terminus of calnexin (CNX). Only the glycosylated substrate is recovered with CNX, consistent with its binding via the glycan. **c**, Aliquots of the crosslinked and natively solubilized samples from **b** were run on a 5–25% sucrose gradient before analysis by SDS–PAGE and autoradiography. Red asterisks denote peak fractions containing the PAT complex as detected by the PAT10 crosslinking product. The PAT complex crosslinked to unglycosylated Rho TM1+2 migrates slightly smaller on the gradient than glycosylated Rho TM1+2, probably the result of CNX (around 90 kDa) no longer being associated with the nascent chain. **d**, The insertion and BMH-mediated crosslinking for 146mer RNCs of the parent Rho TM1+2 construct or versions lacking a cysteine (NoCys) or lacking a glycosylation site ($Cys53_{No\,glyc.}$). The radiolabelled RNCs were produced by in vitro translation in the presence of RMs isolated from wild-type (WT) or two different CCDC47 knockout cell lines (Δ1 and Δ2) generated from two different guide RNAs. Aliquots of the reaction before (–BMH) and after (+BMH) addition BMH were analysed by SDS–PAGE. No appreciable difference in insertion efficiency was observed in knockout microsomes for Rho TM1+2 as monitored by glycosylation efficiency. Red arrowheads indicate the PAT10 crosslink which is lost upon CCDC47 knockout. The faint crosslinked adduct observed in the knockout samples (black asterisks) migrates slightly faster on the gel and probably represents weak Rho TM1+2 crosslinks to Sec61β (Extended Data Fig. 2c). For gel source data, see Supplementary Fig. 1.



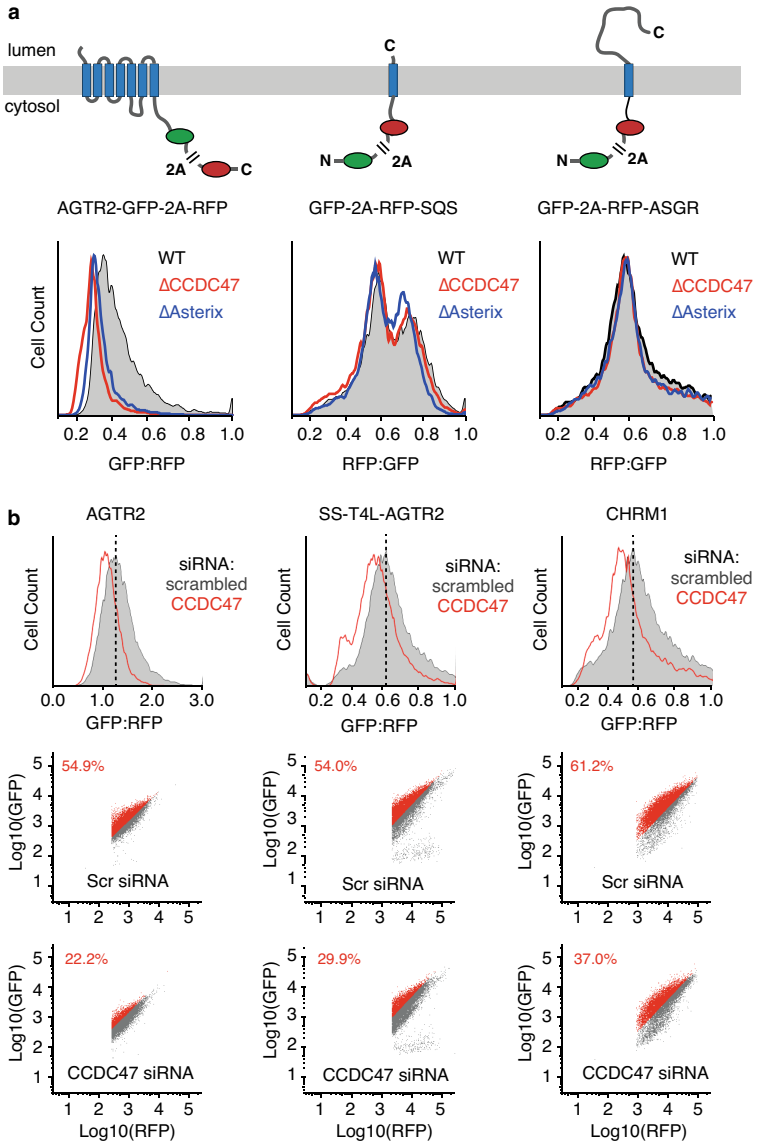
Extended Data Fig. 4 | Conservation and topology of Asterix. a, Alignment of Asterix homologues for five divergent species with a bar chart representing conservation scores of each amino acid. Indicated are three hydrophobic domains (HD1 to HD3), each consisting of approximately 15 amino acids, that are candidate TMDs. **b**, Representation of human Asterix amino acid sequence and the relative lengths of hydrophilic (grey bar) and hydrophobic (blue) regions. **c**, Two matched human Asterix constructs containing either an N- or C-terminal Flag tag were generated by in vitro translation in the presence of RMs. One aliquot of the reaction was set aside for analysis by SDS–PAGE of total translation products (IVT). The remainder of the reaction was treated with proteinase K (+PK) without or with detergent (det). These protease-digested samples were either analysed directly or after immunoprecipitation via the Flag tag. Red arrowheads indicate fragments protected from PK in the absence, but not in the presence, of detergent. The PK-protected fragment from the C-terminally tagged Asterix was recovered by IP, suggesting that the C terminus is located within the ER lumen and the N terminus is located in the cytosol. The relative size difference between the N-and C-terminally tagged constructs observed after PK digestion can be attributed to digestion or protection of the Flag tag. Below the gel is a cartoon depiction of one possible topology based on the results and the protease protected fragments that remain after digestion with PK. The other possible topology is a single-spanning orientation with HD2 and HD3 in the lumen. **d**, Schematic of human Asterix with a C-terminal Flag tag in its predicted 3-TMD topology based on the protease digestion results in **c**. To test this prediction, a cysteine-free version of Asterix (No Cys) was modified with single cysteines at the position indicated by the red-asterisks. If the topology prediction is correct, only the N-terminal domain (NTD) cysteine and the loop 2 cysteine should be accessible to sulfhydryl modifying reagents added to the cytosolic side of them membrane. If the protein spans the membrane only once with the N terminus facing the cytosol, then the loop 2 cysteine should not be modified. As shown in **a**, wild-type Asterix naturally has four cysteines, only one of which should be exposed to the cytosol because it is in the NTD. **e**, Asterix knockout HEK-293 cells were transiently transfected with the indicated Asterix–Flag constructs, semi-permeabilized in 0.01% digitonin, washed to remove digitonin, and treated with PEG-maleimide (average molecular weight 5 kDa) in order to modify any cytosolically exposed cysteine residues. Wild-type Asterix contains four native cysteine residues, one in the N terminus preceding TMD1 and three others within the putative TMD regions. Modification was observed only for the NTD cysteine and the cysteine in loop 2, supporting a 3-TMD topology as depicted in **d**. The single cysteine present in the cytosolic domain of Sec61β was used as a positive control demonstrating equal modification efficiency in all samples, and the no-Cys construct verifies sulfhydryl-dependent modification. Protection of the cysteines in TMD1 and TMD2 from modification verifies that membrane integrity was maintained in the experiment. For gel source data, see Supplementary Fig. 1.



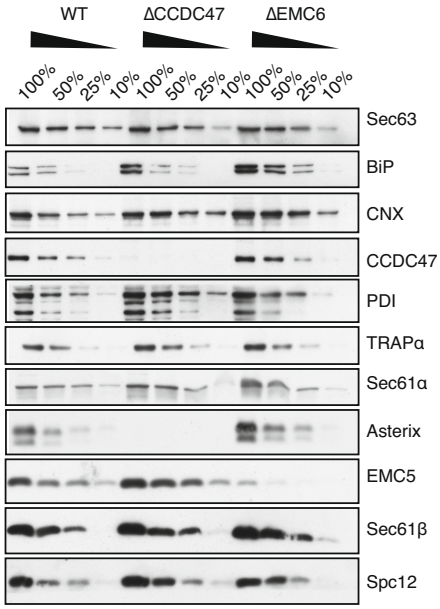
Extended Data Fig. 5 | Characterization of site-specific photo-crosslinking. **a**, Schematic of the strategy for site-specific incorporation of the photo-crosslinking amino acid BPA during in vitro translation (IVT). BPA, a synthetic amber-suppressor tRNA, and recombinant BPA tRNA synthetase are added to an IVT reaction. The nascent protein that is produced incorporates BPA at an amber codon. UV irradiation results in activation of BPA and crosslinking to adjacent proteins. **b**, The photo-crosslinking strategy was tested using a well-validated translocation intermediate: the 86mer of the secretory protein prolactin. The amber codon was installed at position 9, within the hydrophobic core of the signal sequence. At this length, the majority of the nascent chain is precursor (pre), with a small amount that is signal-cleaved (s.c.). The primary crosslinks to SRP54 and components of the translocation site (Sec61α, TRAM, and TRAPα) were verified by immunoprecipitation. **c**, Site-specific photo-crosslinking of a 141mer RNC containing the UV-activated photo-crosslinking amino acid BPA at the indicated amber positions (Amb). The diagram above the autoradiographs shows a schematic of the construct with the appropriate amino acid numbering. Amino acids in red show the strongest crosslinks to Asterix, pink show reduced crosslinks, and grey no detectable crosslinks. Total translation products recovered by IP via an HA tag on the nascent chain are shown adjacent to parallel IPs of selected samples using the indicated antibodies. Although not all IPs are shown, each position was tested for crosslinking to Asterix and CCDC47. RNCs that failed to engage SRP crosslink to UBQLN2, a quality control factor that binds exposed TMDs. This crosslink diminishes markedly when increased RMs are used in the reaction (lanes 9–16, compared to lanes 5–8), presumably because the RMs contribute SRP, which is otherwise limiting in the reaction. A subset of RNCs fail to release from SRP and crosslink to SRP54. The crosslink indicated by the hashtag (#) is likely to be a mixture of similarly migrating crosslinks. Because this crosslink diminishes substantially with increased RMs (similar to the UBQLN2 crosslink), it is likely to be SGTA, another TMD-binding factor in the cytosol of this size. A small proportion of this crosslink could be the similarly sized Sec61α or TRAM. Of the membrane-inserted RNCs, the main crosslink is to Asterix, seen prominently for residues 52 to 63. At this length, the TMD has moved away from Sec61α, so crosslinks to this factor are minimal. No crosslinking to CCDC47 were ever observed. By testing five sequential positions in the centre of the TMD, all sides of the helix have been sampled.



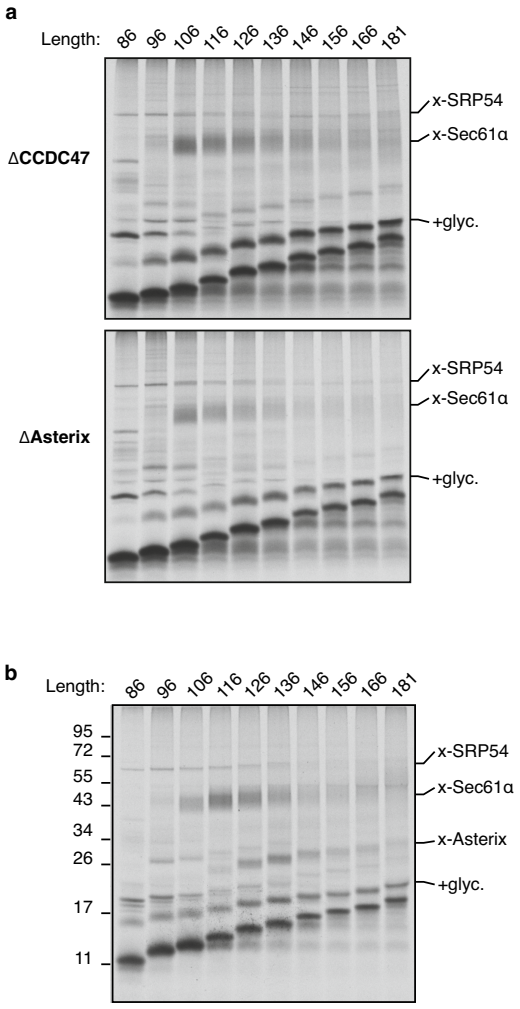
Extended Data Fig. 6 | Effect of Asterix depletion on multi-spanning membrane proteins. a, The raw data for three of the histograms of the GFP:RFP ratio (or RFP:GFP ratio in the case of GFP-2A-RFP-SQS) shown in Fig. 3. The mode of the control histogram (dotted black line in Fig. 3) was used to determine the statistical mode of GFP:RFP (or RFP:GFP) ratio. This mode was used as a gate to colour the dot plots shown below the histograms such that all cells above the gate were coloured red and those below the gate were coloured grey. The percent of cells above the gate for each plot is indicated. **b**, Flow-cytometry analysis of the indicated GPCR reporters using the dual-colour assay system exactly as in Fig. 3. Cell lines containing the reporter stably integrated at a single FRT site located downstream of a doxycycline-inducible reporter were used for these assays. This allows assay of cells using a single transfection (which proved to be less toxic than sequential transfections with both the siRNA and the reporter), and provided control of the length of time of reporter expression. Each reporter cell line was treated with scrambled (Scr) or Asterix-targeting siRNAs, then at the time of effective knockdown (verified in separate experiments using immunoblotting), the reporter was induced for approximately 6–8 h. Induction only after knockdown allows us to monitor the reporter that was produced in the absence of Asterix rather than a heterogeneous mixture of reporter expressed during the knockdown. The histograms of the GFP:RFP ratio in the scrambled- versus Asterix-siRNA cells are shown in grey and blue, respectively, in the upper plot for each construct. The two dot plots below the histogram are the corresponding raw data plotted as described in **a**. Each reporter shows a distribution of lower GFP:RFP ratio, with some reporters being more impacted than others. This is not seen with the tail-anchored protein SQS using the same assay format.



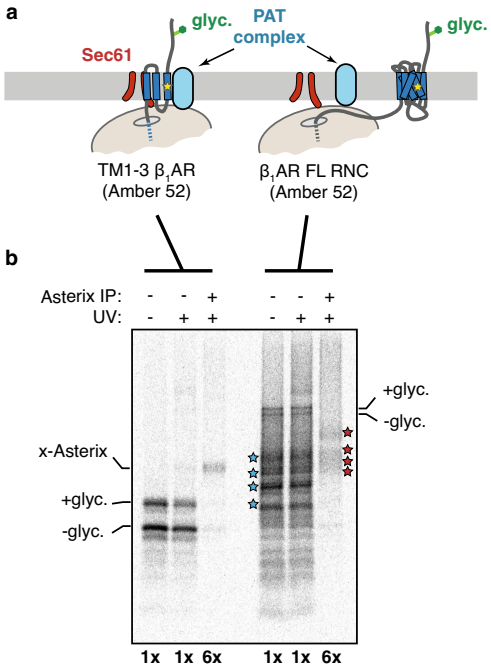
Extended Data Fig. 7 | Effect of CCDC47 depletion on membrane protein biogenesis. a,l The diagrams depict dual-colour fluorescent reporters for protein stability as an indirect measure of successful biogenesis. The membrane protein of interest is tagged with one fluorescent protein (FP), which is separated from a second FP by the viral 2A peptide sequence. When the 2A sequence is translated, peptide bond formation is skipped without perturbing elongation by the ribosome. Thus, translation results in two separate proteins made in a 1:1 stoichiometry that are separated at the 2A sequence. If biogenesis of the membrane protein is impaired, it will be degraded along with its tagged FP, resulting in an altered ratio of the two FPs. Thus, treatment conditions that impair biogenesis of the membrane protein will be reflected as a relative change in the ratio of FPs. The three reporters encoding angiotensin type-2 receptor II (AGTR2), squalene synthase (SQS) and Asialglycoprotein receptor (ASGR) were transiently transfected into wild-type (WT), CCDC47 knockout (ΔCCDC47) or Asterix knockout (ΔAsterix) HEK293 cells and analysed by dual-colour flow cytometry. Histograms represent the distribution of FP ratio in wild-type (grey), ΔCCDC47 (red) and ΔAsterix (blue) cells. A biogenesis defect is only seen for the multi-spanning membrane protein AGTR2, but not for the tail-anchored protein SQS or the signal-anchored single pass protein ASGR. **b**, Assays similar to those in Fig. 3, but for cell lines treated with scrambled versus CCDC47 siRNAs as indicated. We find that the phenotypes for Asterix and CCDC47 knockdowns are very similar for all reporters (three are shown here), with CCDC47 consistently being somewhat more modest. The reason for this seems to be that CCDC47 knockdown is slower and less efficient than Asterix knockdown. Similar phenotypes are seen for AGTR2 and SS-T4L-AGTR2, a version that contains an N-terminal signal sequence and T4 lysozyme preceding TMD1. In earlier studies, we found that initiating translocation with a signal sequence completely bypasses the requirement for EMC-mediated TMD1 insertion. The fact that SS-T4L-AGTR2 remains sensitive to PAT complex depletion (as judged by either Asterix or CCDC47 knockdowns) argues that the PAT complex acts independently of EMC.



Extended Data Fig. 8 | Expression of ER biogenesis factors in ΔCCDC47 and ΔEMC6 cells. ER rough microsomes were isolated from wild-type, ΔCCDC47 and ΔEMC6 HEK293 cells and normalized to an absorbance of 75 at 280 nm. Serial dilutions of each sample were analysed by SDS–PAGE and immunoblotting for the indicated antigens. BiP levels are increased in both knockout cell lines, consistent with an activated UPR caused by altered ER homeostasis. For gel source data, see Supplementary Fig. 1.



Extended Data Fig. 9 | TMD1 insertion does not require the PAT complex or TMD2. a, Rho TM1+2 RNCs of varying nascent chain lengths (indicated at top of gels) were translated in vitro in the presence of RMs prepared from ΔCCDC47 or ΔAsterix HEK293 cells as indicated. Membranes were isolated by centrifugation through a sucrose cushion and treated with the chemical cross linking reagent BMH. The samples were denatured in 1% SDS and immunoprecipitated via the N-terminal Flag tag of the substrate. Notice that the glycosylation of substrate is very similar in efficiency and timing as that seen in RMs prepared from wild-type HEK293 cells (see Fig. 1a for comparison). Furthermore, the appearance and disappearance of the SRP54 and Sec61α cross linking adducts are not appreciably altered from the results seen in RMs prepared from wild type cells. Thus, the early steps of rhodopsin biogenesis are not impaired appreciably in the absence of Asterix or CCDC47. As expected, the crosslink to Asterix/PAT10 is not seen (verified by anti-Asterix immunoprecipitation; not shown). Crosslinking products seen at the approximate size of the Asterix crosslink are therefore other protein(s). **b**, The Rho TM1 construct in which TMD2 is replaced with a hydrophilic linker sequence (see diagram in Fig. 4a) was analysed for crosslinking as in Fig. 1a. The absence of TMD2 does not affect the crosslinking between TMD1 and Asterix. By contrast, mutation of the most polar residue in TMD1 (N52) markedly reduces Asterix crosslinking and reduces TMD1 proximity to Sec61α (Fig. 4a).



Extended Data Fig. 10 | Analysis of Asterix interaction with TMD1 by photo-crosslinking. a, Experimental strategy for comparing Asterix interaction with a membrane protein intermediate versus full-length product. In this experiment, the photo-crosslinking amino acid BPA (yellow star) is incorporated into position 52 within TMD1 of β1AR by in vitro translation. The intermediate is represented by the TM1-3 product containing the first three TMDs of β1AR. The full-length (FL) β1AR contains all seven TMDs followed by a long flexible linker. TM1-3 is stalled 35 amino acids downstream of TMD3 (with TMD4 inside the ribosomal tunnel), allowing TMD3 to be outside the ribosome. β1AR FL is stalled 152 amino acids downstream of TMD7, providing a sufficiently long tether for all seven TMDs to have emerged, inserted into the membrane, and assembled together. The translation products are then irradiated with UV light to activate the BPA and any crosslinking to Asterix is subsequently detected by denaturing IP via Asterix. **b**, Results from a photo-crosslinking experiment as depicted in **a**. The microsomes from the IVT reaction were isolated, resuspended, irradiated with UV light (or left untreated), and denatured. The samples were then divided in two and immunoprecipitated via the nascent chain or via Asterix. Sixfold more of the Asterix IPs were loaded on the gel relative to the nascent chain IPs. As expected, the BPA in TMD1 crosslinks to Asterix in the TM1–3 intermediate. The crosslinked band in the 6× Asterix IP is the same intensity as the glycosylated band in the 1× nascent chain IP. Although elongation to the full-length product was somewhat inefficient, clear glycosylated and non-glycosylated products are observed in the nascent chain IPs. No band is seen in the 6× Asterix IP sample that is of comparable intensity to the glycosylated band in the 1× nascent chain IP. This argues that the proximity of TMD1 to Asterix has diminished substantially in the full-length nascent chain relative to the TM1-3 intermediate. Of note, a heterogeneous set of crosslinks (marked by red stars) are seen at a lower part of the gel in the 6× Asterix IP. These correspond to the sizes expected for Asterix crosslinks (that is, shifted by about 10 kDa) to the major incomplete translation products (marked by blue stars). These crosslinks provide an internal control and further supports the conclusion that incomplete products engage Asterix, while a complete 7-TMD product does not.

nature research

Corresponding author(s): Hegde

Last updated by author(s): Apr 19, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist .

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
✘	☐ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
✘	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
✘	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
✘	☐ A description of all covariates tested
✘	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
✘	☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
✘	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
✘	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
✘	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
✘	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Mass spectrometry data were collected using a Q Exactive Plus hybrid quardrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific) equipped with a nanospray ion source. Phosphorimaging of radioactive samples was acquired on Typhoon FLA7000 (GE Healthcare). Flow cytometry data were collected using Becton Dickinson LSRII instrument.
Data analysis	ImageJ (version 2.0.0-rc-69/1.52p) was used to display densitometry of autoradiographs. FlowJo (version 10.6.2 for Mac OS X operating system) was used to plot flow cytometry data. Mass spectrometry raw files were processed using Proteome Discoverer (version 2.1, Thermo Scientific). MS/MS spectra were searched against mammals, UniProt Fasta database using Mascot (version 2.4, Matrix Science) search engine. Scaffold 4 Version 4.8.9 (www.proteomesoftware.com) was used to analyze mass spectrometry data.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

No datasets were generated or analysed during the current study.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculations were performed. Biochemical experiments were repeated on independent days to verify reproducibility with no further statistical testing.
Data exclusions	No data were excluded from the analysis.
Replication	Biochemical experiments were replicated to verify reproducibility. No attempts at replication failed. Details are provided in the "Statistics and Reproducibility" section of the Methods.
Randomization	Randomisation is not relevant to this study which reports biochemical analyses.
Blinding	Blinding is not relevant to this study which reports biochemical analyses.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	CCDC47 antibody #1 (A305-100A-1) and CCDC47 antibody #2 (A305-101A-1) were obtained from Bethyl-laboratories and each was used at 1:5,000 for immunoblotting and 1:500 for IPs. FLAG immunoprecipitations (IPs) were performed using FLAG-M2 affinity gel (Sigma, Lot # SLBW6125) and used at 1:100 dilution for IPs. Anti-HA antibody (1:500 for IPs) was generated in house and has been described (Itakura et al., 2016, Mol. Cell, 63:21-33). Signal Peptide Peptidase (SPP) antibody was purchased from Bethyl Laboratories (Lot # A304-404A-1) and used at 1:250 dilution for IPs. Anti-Asterix (WDR83OS) antibody was purchased from Sigma/Human Protein Atlas (HPA065685, Lot # A304-404A-1) and used at 1:1,000 for immunoblotting and 1:250 for IPs. Anti-UBQLN2 antibody was clone 5F5 obtained from Sigma (WH0029978M3, Lot # FB261-5F5) and used at 1:250 for IPs. Anti-SRP54 was from BD Biosciences (610940, Lot # 7157596) and used at 1:200 for IPs. Custom antibodies against Sec61-beta (1:10,000 for immunoblotting, 1:500 for IPs), TRAP-alpha (1:5,000 for immunoblotting, 1:500 for IPs), and TRAM (1:5,000 for immunoblotting, 1:500 for IPs) have been described previously (Fons et al., 2003, J. Cell Biol.,160:529-30). Custom antibody against Sec61-alpha (1:5,000 for immunoblotting, 1:500 for IPs) has been described previously (Song et al., 2000, Cell, 100:333-343). Custom antibody against SPC12 (1:5,000 for immunoblotting) has been described previously (Gorlich and Rapoport, 1993, Cell, 75:615-630). Custom antibody against Sec63 (1:2,000 for immunoblotting) has been described previously (Garrison et al., 2005, Nature, 436:285-289). Calnexin antibody (1:5,000 for immunoblotting) was from Enzo Lifesciences (#ADI-SPA-865). BiP antibody (1:1,000 for immunoblotting) was from BD Biosciences (#610978). PDI antibody (1:1,000 for immunoblotting) was from Enzo Lifesciences (#ADI-SPA-890). EMC5 antibody (1:1,000 for immunoblotting) was from Abcam (#Ab174366).
Validation	Antibodies were not validated. Specificity controls are displayed within the figures of the manuscript.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK293 FRT/TO TRex cells from Invitrogen.
Authentication	The validity of this cell line was ensured by the antibiotic resistance markers within its genome and by its unique FRT site downstream of a doxycycline-inducible promoter.
Mycoplasma contamination	Cell lines were negative for mycoplasma. They are tested monthly.
Commonly misidentified lines (See ICLAC register)	None used.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Cells were washed with PBS and trypsinised. After inhibiting trypsin, cells were collected into 1.5 ml tubes, spun at 5,000 rpm for 3 min at 4°C, washed with 1 ml of ice cold PBS, spun again and resuspended in 500 µl of ice cold PBS for immediate analysis by flow cytometry.
Instrument	LSRII instrument (Becton Dickinson).
Software	FlowJo software (version 10.6.2 for Mac OS X operating system).
Cell population abundance	No staining was performed and no sorting was performed. All transfected live cells were counted in the analysis. Transfection was judged by GFP reporter expression. Cell viability was judged by the absence of DAPI staining.
Gating strategy	No gating or sorting was performed. All transfected live cells were counted in the analysis. Transfection was judged by GFP reporter expression. Cell viability was judged by the absence of DAPI staining. Examples of dot plots representing the raw data used to generate the histograms are shown in Extended Data Fig. 6 and 7.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.