Published in final edited form as:
Nature. 2022 November 01; 611(7934): 161–166. doi:10.1038/s41586-022-05336-2.

Mechanism of an intramembrane chaperone for multipass membrane proteins

Luka Smalinskaitė[1], **Min Kyung Kim**[1], **Aaron J. O. Lewis**[1], **Robert J. Keenan**[2], **Ramanujan S. Hegde**[1,3]

[1]MRC Laboratory of Molecular Biology, Cell Biology Division, Francis Crick Avenue, Cambridge, CB2 0QH, UK

[2]Department of Biochemistry and Molecular Biology, The University of Chicago, 929 East 57th Street, Chicago, IL 60637, USA

Abstract

Multipass membrane proteins play numerous roles in biology and include receptors, transporters, ion channels, and enzymes[1,2]. How multipass proteins are co-translationally inserted and folded at the endoplasmic reticulum (ER) is not well understood[2]. The prevailing model posits that each transmembrane domain (TMD) of a multipass protein successively passes into the lipid bilayer through a front-side lateral gate of the Sec61 protein translocation channel[3–9]. The PAT complex, an intramembrane chaperone comprising Asterix and CCDC47, engages early TMDs of multipass proteins to promote their biogenesis by an unknown mechanism[10]. Here, biochemical and structural analysis of intermediates during multipass protein biogenesis revealed that the nascent chain is not engaged with Sec61, which is occluded and latched closed by CCDC47. Instead, Asterix binds to and redirects the substrate behind Sec61, where the PAT complex contributes to a multipass translocon surrounding a semi-enclosed lipid-filled cavity[11]. Detection of multiple TMDs in this cavity after their emergence from the ribosome suggests that multipass proteins insert and fold behind Sec61. Accordingly, biogenesis of several multipass proteins was unimpeded by inhibitors of the Sec61 lateral gate. These findings elucidate the mechanism of an intramembrane chaperone and suggest a new framework for multipass membrane protein biogenesis at the ER.

The first transmembrane domain (TMD1) of most membrane proteins inserts into the lipid bilayer by passing through a lateral gate in the Sec61 protein translocation channel[3,8,12,13]. TMD1 of membrane proteins with short exoplasmic N-terminal domains, such as most G-protein coupled receptors (GPCRs), can be inserted independently of the Sec61 lateral gate by the ER membrane protein complex (EMC)[14,15], a member of the Oxa1 family of TMD

[3]Correspondence: rhegde@mrc-lmb.cam.ac.uk.

Author contributions: LS designed and performed nearly all biochemical experiments and associated analysis; MK prepared and characterised samples for cryo-EM, collected and processed cryo-EM data, and contributed to model building; AJOL and RJK analysed structural data and contributed to model building; RSH and RJK provided funding and guidance; RSH conceived the project, oversaw its implementation, and wrote the paper with input from all other authors.

Competing interests: The authors declare that they have no competing interests.

Additional information: Reagents from this work are available from RSH by request.

Europe PMC Funders Author Manuscripts

insertases[16,17]. With either route, the ribosome-nascent chain complex (RNC) is ultimately docked tightly at Sec61[4,7,18,19] where the later steps of multipass protein biogenesis are thought to occur by yet unclear mechanisms[2].

The long-standing paradigm envisions iterative TMD insertion via the Sec61 lateral gate[3,13] followed by intramembrane packing of TMDs into the correct tertiary structure[20,21]. Recently, optimal biogenesis of multipass but not single-pass membrane proteins was found to require a protein complex termed PAT (for **p**rotein **a**ssociated with **t**ranslocon), a heterodimer of CCDC47 and Asterix[10,22]. The PAT complex interacts co-translationally with semi-hydrophilic TMDs throughout multipass protein biogenesis[10,23,24], suggesting an intramembrane chaperone function. The PAT complex probably engages the ribosome as part of a poorly characterised multipass translocon with Sec61 and other factors[2,25], but its role and mechanism during membrane protein biogenesis are unclear.

Recruitment of the PAT complex to Sec61

We analysed a series of early insertion intermediates of rhodopsin (Rho), a 7-TMD G protein-coupled receptor (GPCR) and model multipass membrane protein, produced by *in vitro* translation using cytosol from reticulocyte lysate and ER microsomes from canine pancreas. These experiments used a construct (termed Rho^{ext}) containing the first two TMDs of Rho preceded by an extended N-terminus and epitope tag (Fig. 1a; characterised in Extended Data Fig. 1a-d). Immunoblotting of sequential Rho^{ext} insertion intermediates affinity purified via the nascent chain revealed stable PAT complex recruitment to Sec61-associated ribosome-nascent chain-complexes (RNCs) when ~70 residues had been synthesised downstream of TMD1 (Fig. 1b; Extended Data Fig. 1e).

TRAM2, an unrelated multipass membrane protein whose topology is opposite to Rho, also recruited the PAT complex in a nascent chain length-dependent manner (Extended Data Fig. 1f). Recruitment of the PAT complex to Rho^{ext} RNCs was reduced by increasing TMD1 hydrophobicity (Fig. 1c), consistent with the hydrophilicity requirement for substrate crosslinking to Asterix[10]. Thus, the PAT complex is stably recruited to the ribosome-Sec61 complex by membrane proteins of either topology at roughly the same tether length beyond TMD1 (after accounting for their opposite orientations; see Extended Data Fig. 1g).

Architecture of the PAT complex at Sec61

The Rho^{ext} RNC at the point of initial PAT complex engagement was analysed by single-particle electron cryomicroscopy (cryo-EM; Extended Data Fig. 2a,b). The overall resolution of the PAT complex-containing map was 3.25 Å with local resolutions from 2.8 Å for most of the ribosome to ~7 Å for poorly-resolved regions of associated membrane proteins (Extended Data Fig. 2c). The map contained well-resolved density for the Sec61 complex, the cytosolic domain of CCDC47, and the membrane domains of TMEM147 and Nicalin at locations seen in earlier work[11] (Fig. 2a; Extended Data Fig. 3). The cytosolic domain of TMCO1 and the lumenal domain of the translocon-associated protein (TRAP) complex could be seen at their previously observed locations[11,26] at high contour levels.

TMEM147 and Nicalin interact with NOMO to form an obligate heterotrimer[25,27] whose function is not known. We term this complex BOS (for **b**ack **o**f **S**ec61) due to its location. TMCO1 is a member of the Oxa1 insertase family[16] and forms an obligate complex with C20Orf24, which we name OPTI (**o**bligate **p**artner of **T**MCO1 **i**nsertase)[25]. The TMCO1-OPTI complex is homologous to the GET1-GET2 complex and the EMC3-EMC6 subcomplex within EMC[28], both of which mediate TMD insertion[17,29]. We term the TMCO1-OPTI complex GEL (for **G**ET-and **E**MC-**l**ike). As shown elsewhere[25], nascent membrane proteins trigger not only PAT complex recruitment (Fig. 1), but also the BOS and GEL complexes. We focus primarily on the PAT complex, with the BOS and GEL complexes described briefly later and in the Supplementary Discussion.

High-confidence AlphaFold2 predictions of the PAT-, BOS-, and Sec61-complexes[30,31] (Extended Data Fig. 4) were fitted into the map and adjusted conservatively to generate a model of RNCs engaged with these complexes (Fig. 2a,b; Extended Data Fig. 3). The GEL and TRAP complexes were omitted from the model due to their low occupancy and poor resolution, respectively. The Sec61 channel is occluded by its plug helix and the lateral gate is closed, similar to the ‘primed’ state of Sec61 bound to non-translating ribosomes[32]. The PAT complex is located behind the Sec61 complex (Fig. 2a), whose front side is defined by its lateral gate used for signal sequence and TMD insertion[33–35].

The globular domain of CCDC47 contacts ribosomal protein eL6 and several sites along 28S rRNA. C-terminal to this globular domain, CCDC47 contains three helices that run along the surface of the ribosome making contacts with uL22 and various parts of the 28S rRNA. The last two of these helices, which we term the latch helices, are separated by a sharp bend at the point of contact with Sec61α. The final helix culminates inside the mouth of the ribosomal exit tunnel, substantially narrowing the otherwise wide opening and abutting the emerging nascent chain (Extended Data Fig. 3e,f).

Site-specific photo-crosslinking was used to validate our placement of Asterix and identify its substrate-binding surface (Fig. 2c-e). Semi-permeabilised Asterix knockout cells were reconstituted by *in vitro* translation of Asterix variants containing the photo-crosslinking amino acid benzoyl-phenylalanine (BPA) at a variety of desired positions (Extended Data Fig. 5a-c). The sum of crosslinking data verifies the AlphaFold2-predicted CCDC47-interacting region, which we term the back of Asterix (Fig. 2e; Extended Data Fig. 5d), and define Asterix’s front-side substrate-binding surface (Fig. 2e; Extended Data Fig. 5e).

The substrate binding domain of Asterix

The substrate-binding surface of Asterix is ~70 Å away from the back of Sec61 and the ribosome exit tunnel (Fig. 2a,b).This position means that ~25-35 residues of unstructured polypeptide is needed to span the distance from the exit tunnel to Asterix. Accounting for another ~35 residues inside the ribosome, a TMD cannot reach Asterix until ~60-70 downstream residues have been synthesised, explaining earlier crosslinking data[10] and the length at which stable PAT complex recruitment is seen (Fig. 1b; Extended Data Fig. 1g).

Europe PMC Funders Author Manuscripts

The substrate-binding domain is highly conserved and amphiphilic (Fig. 3a). Several hydrophilic side chains (mostly serine) are interspersed with the long and flexible hydrophobic side chains of eight methionine residues. Mutants that disrupt the amphiphilic surface by changing several polar amino acids to leucine reduced substrate photo-crosslinking to Asterix (Fig. 3b). Similarly, replacement of multiple methionine residues, particularly those in TMD1 of Asterix, for the less flexible but similarly hydrophobic valine also impaired substrate photo-crosslinking.

Both hydrophilicity and side-chain flexibility are therefore important for an optimal interaction of Asterix with substrate TMDs. This arrangement is well suited to bind and stabilise partially hydrophilic TMDs in the membrane. Furthermore, the flexibility of methionine side chains would help accommodate diverse substrates. Methionine-rich TMD binding sites are used widely in the cytosol by targeting factors[36,37], chaperones[38–40], and quality control factors[41]. That Asterix employs a similar strategy suggests that TMD recognition in the membrane shares mechanistic principles with recognition in the cytosol.

CCDC47 impedes opening of Sec61

The sharp bend between the latch helices of CCDC47 is wedged between the ribosome and the N-terminal half (N-half) of Sec61α (Fig. 3c). CCDC47 makes close contacts with the cytosolic loop between Sec61α lateral gate helices TMD2 and TMD3. When substrates pass through Sec61's lateral gate, its N-half must rotate away from the ribosome-bound C-half[34]. With CCDC47 in position, this rotation would be impeded by clashes with the latch helices (Extended Data Fig. 6a). Hence, CCDC47 constrains Sec61 to a closed conformation. This constraint probably explains why the cryo-EM map containing CCDC47 shows relatively strong density for the N-half of Sec61, whereas in maps lacking CCDC47 the N-half density is typically weak, indicating flexibility (Extended Data Fig. 6b).

Another consequence of CCDC47 binding is that the latch helices are positioned between the mouth of the ribosome exit tunnel and the cytosolic vestibule of Sec61. This configuration markedly obstructs nascent chain access to the vestibule and lateral gate on the opposite side (Fig. 3c). By disfavouring vestibule access, the nascent chain instead would be directed behind Sec61 toward the multipass translocon components. Routing of the nascent chain in this direction is also favoured by TMD1 being held ~70 Å behind Sec61 by Asterix. Thus, after the PAT complex has been engaged, downstream hydrophobic sequences would be disfavoured from insertion via the Sec61 lateral gate.

We tested this prediction by determining whether a hydrophobic element that emerges at a PAT complex-engaged translocon is impaired in accessing Sec61 (Fig. 3d). Cleavable N-terminal signal sequences are known to function even when placed in an internal location[42], allowing us to use cleavage as an assay for Sec61 engagement. Sec61-dependent cleavage of the secretory protein preprolactin placed downstream of an artificial TMD (containing 22 leucines), which does not recruit the PAT complex (see Extended Data Fig. 7h), was ~64%. Cleavage did not occur in the absence of ER, was prevented by an inhibitor of the Sec61 lateral gate, and was unchanged when CCDC47 was knocked out. When preprolactin was preceded by a Rho membrane domain that recruits the PAT complex, Sec61-dependent

cleavage efficiency was notably lower (39%, p=0.01). Importantly, this impairment was completely reversed (66% cleavage) when CCDC47 was deleted. Thus, PAT complex recruitment by an upstream membrane domain impedes a downstream hydrophobic domain from engaging the Sec61 complex, consistent with the structural observations in Fig. 3c.

Multipass translocon bound to substrate

After the PAT complex has initially engaged TMD1 of Rho, the next step upon further elongation would be insertion of TMD2 and TMD3 accompanied by translocation of the intervening loop. To investigate how this step might work in the context of a closed Sec61, we affinity-purified a Rho^{ext} insertion intermediate at a length just after TMD3 has emerged from the ribosome, when TMD4 is still inside the ribosome (see diagram, Fig. 4a), and analysed it by quantitative mass spectrometry, immunoblotting, and cryo-EM.

This Rho-4TMD intermediate (and an analogous late TRAM2 insertion intermediate) showed enrichment of the PAT, GEL and BOS complexes (Fig. 4a,b; Extended Data Fig. 7a-c). Cryo-EM analysis of the Rho-4TMD intermediate (Extended Data Fig. 8) combined with high-confidence predictions of each multipass translocon subcomplex (Extended Data Fig. 4) allowed us to place them relative to Sec61 (Fig. 4c). Compared to the earlier intermediate (Fig. 2a), the occupancy of the GEL complex was higher, whereas the PAT complex and the BOS complex were largely unchanged. The other major difference was the visualisation of a membrane-spanning helix, assigned to TMD3 of Rho based on subsequent crosslinking data, immediately outside the exit tunnel behind the closed Sec61 complex.

Site-specific photocrosslinking using probes in the Rho-4TMD intermediate showed that TMD1 is adjacent to Asterix, TMD2 is adjacent to TMCO1, and TMD3 is at the back of Sec61 (Fig. 4d; Extended Data Fig. 7d,e). Thus, the three TMDs of this Rho intermediate reside in the cavity behind the Sec61 complex surrounded by the other multipass translocon components. This intermediate represents the post-insertion state in which the TMD2-TMD3 unit has just been translocated into the cavity between the back of Sec61 and the multipass translocon subcomplexes. The proximity of TMD2 to the GEL complex suggested that it facilitated insertion of the TMD2-TMD3 unit. Consistent with this idea, this insertion step is reduced when the GEL complex is deleted (Extended Data Fig. 7f), which does not appreciably impair recruitment of the PAT and BOS complexes to Sec61 (Extended Data Fig. 7g) or TMD1 engagement of the PAT complex (Extended Data Fig. 7h).

Protein biogenesis with Sec61 inhibitors

Our findings suggest that after assembly of the multipass translocon, TMD insertion does not rely on the Sec61 lateral gate. For GPCRs, whose first TMD can use EMC for insertion[14], biogenesis would occur completely independent of the Sec61 lateral gate. Consistent with this idea, Rho biogenesis *in vitro* (Fig. 4e) and in cells (Fig. 4f) was completely unimpaired by Apratoxin A, a broadly potent Sec61 inhibitor. Asialoglycoprotein receptor 1 (ASGR1), a single-pass membrane protein known to use Sec61 for insertion[5], was strongly inhibited, whereas the post-translationally inserted EMC substrate squalene synthase (SQS)[17] was refractory (Fig. 4e,f). Although Sec61 activity was

not needed, Sec61 protein was required for Rho biogenesis (Fig. 4f), presumably because it is needed as a ribosome receptor[43] and helps nucleate assembly of the multipass translocon.

Lack of reliance on Sec61 activity was seen for two other GPCRs (β1AR and AGTR2) and with two other structurally unrelated Sec61 inhibitors *in vitro* and in cells (Extended Data Fig. 9, 10). By contrast, cleavable signal sequences and signal anchors that are known to pass through the Sec61 lateral gate are partially or completely blocked by the same inhibitors. β1AR and AGTR2 were rendered sensitive to Sec61 inhibitors by appending an N-terminal extracellular domain preceded by a signal sequence (Extended Data Fig. 10c).

Similarly, C3AR1, a GPCR that contains a long translocated domain between TMD4 and TMD5, requires Sec61 activity specifically at this step of biogenesis (Extended Data Fig. 9d). C3AR1 translocation becomes refractory to Sec61 inhibition when this loop is shortened. Thus, Sec61 activity is dispensable for translocation of short domains, whether at the N-terminus or between internal TMDs, but becomes essential when long domains need to be translocated across the ER membrane. Long inter-TMD segments of polypeptide might trigger displacement of CCDC47's latch helices, whose C-terminal region probes the ribosome exit tunnel (Extended Data Fig. 3f; Supplementary Discussion). This displacement would temporary relieve Sec61 occlusion by CCDC47, allowing the immediately preceding TMD to initiate translocation of its downstream soluble domain through the Sec61 channel.

Model for multipass protein biogenesis

Our findings lead to a model for multipass membrane protein biogenesis that differs from the long-standing paradigm of Sec61-mediated sequential TMD insertion (see Supplementary Discussion for more details). Once a semi-hydrophilic early TMD has a ~70 Å downstream tether to the ribosome exit tunnel, the TMD can engage Asterix's amphiphilic substrate-binding surface. This interaction favours stable residence at the ribosome of the PAT complex, whose latch helices then disfavour Sec61 access and opening. At the same time, the GEL and BOS complexes are recruited concomitant with OST displacement to form the multipass translocon[25]. The next two TMDs then emerge from the ribosome behind a closed Sec61 and insert as a pair into the multipass translocon's lipid-filled cavity. Translocation of the loop between these two TMDs may be facilitated by a thinned membrane adjacent to the conserved semi-hydrophilic interior surface of the GEL complex (Extended Data Fig. 8d).

As with other Oxa1 family members[2,44], translocation by the GEL complex is probably limited to loops of less than ~50 amino acids, explaining why translocation of long loops requires Sec61. When long loops do emerge from the ribosome, a mechanism might exist to temporarily displace CCDC47's latch helices to allow Sec61 engagement (see Supplementary Discussion). Long translocated loops between internal TMD pairs are far less common than short loops[45], suggesting that the majority of multipass protein biogenesis occurs independently of Sec61's lateral gate. The semi-shielded lipid cavity of the multipass translocon would provide a site for accumulation and packing of around seven TMDs, which can egress into the bulk membrane via the large gap between Asterix and the back of Sec61. Because most larger membrane proteins are built of modules comprising seven or fewer

TMDs[46,47], the multipass translocon should be capable of accommodating a wide variety of diverse substrates.

Methods

Plasmids, GeneBlocks, and antibodies

Constructs for *in vitro* translation (IVT) in rabbit reticulocyte lysate were cloned into a pSP64-based vector or ordered as gene blocks (from Integrated DNA Technologies) containing a 5' SP6 promoter for transcription and are described in Supplementary Table 2. Antibodies were either from commercial sources or were custom antibodies that have been described previously as detailed in Supplementary Table 3.

Cell culture

HEK293 Flp-In TRex cells (Invitrogen) were grown in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal calf serum. The HEK293-derived Asterix, CCDC47 and TMCO1 knockout (KO) cell lines have been described[10,11]. CRISPR-Cas9-mediated disruption of TMEM147 was performed using pSPCas9(BB)-2A-Puro (PX459) plasmid (Addgene) encompassing the gRNA 5'-CGACGATGTAATGGACCTAG-3'. Transfected cells were selected for 48 h with 1 μg/ml puromycin. Remaining cells were sorted into 96-well plates at 1 cell/well concentration to select for single-cell colonies. Single colonies were expanded and screened for successful gene disruption. Cell lines with various stably expressed doxycycline-inducible reporters have been described[10,14,17]. These reporter cell lines were grown in DMEM was supplemented with tetracycline-free FCS (Biosera) and 15 μg/ml blasticidin and 100 μg/ml hygromycin.

Flow cytometry analysis

For knockdown experiments in reporter cell lines, siRNAs were transfected using the Lipofectamine RNAiMAX reagent according to manufacturer's instructions (Thermo Fisher Scientific). After 48 h, a second round of siRNA transfection was performed, and the cells were analysed 24 h later. Fluorescent reporter expression was induced with 1 μg/ml doxycycline in DMEM supplemented with 10% fetal calf serum for 6 h prior to analysis by flow cytometry. Sec61 inhibitors (or DMSO for control samples) were included during the induction. The inhibitors were used at following concentrations: Apratoxin – 200 nM; Ipomoeassin F – 200 nM; CT8 – 1000 nM, as described earlier[15,48–50]. The cells were collected in ice-cold PBS, washed and resuspended in PBS supplemented with 2% FCS and 1 μg/ml DAPI (Thermo Fisher Scientific). Cells were passed through 70-μm filter immediately prior to analysis using Beckton Dickinson LSRII or LSRFortessa instrument. A total of at least 20,000 fluorescent and live (negative for DAPI stain) cells were collected (see Supplementary Fig. 2 for an example of how gating was performed). For experiments with a transiently transfected reporter (N_{ext}-β_1AR), the plasmid was transfected into wild-type HEK293 Flp-In TRex cells using Transit 293 reagent (Mirus) according to manufacturer's instructions. Twelve hours post-transfection, the cells were treated with 200 nM Apratoxin (or DMSO) in DMEM supplemented with 10% fetal calf serum and incubated for another 12 hours before analysis by flow cytometry as above.

Preparation of semi-permeabilised cells

Semi-permeabilised (SP) cells were prepared by modification of earlier protocols[41]. All steps of SP-cell preparation were performed at 0-4°C on cells at ~70% confluency, typically from a 10 cm dish. After removing the growth media, the cells were washed once with ice-cold PBS, collected by gentle pipetting in 1 ml PBS, and counted using Scepter™ 2.0 Cell Counter (Merck Millipore) with the 60 μM sensor (Merck Millipore, PHCC60050). The cells were recovered by centrifugation for 2 min at 5000 rpm in a microcentrifuge, washed once with ice-cold PBS, then resuspended in 1 ml of 1X "physiologic salt buffer" [PSB: 50 mM HEPES-KOH, pH 7.5, 100 mM KOAc, 2.5 mM $Mg(OAc)_2$] supplemented with 0.01% digitonin. Following a 10 min incubation on ice, the cells were collected by centrifugation, washed twice with 1X PSB, then resuspended in 0.5X PSB to a concentration of 4×10^7 cells/ml. The SP cells were used immediately without freezing at a final concentration in translation reactions of 4×10^6 cells/ml.

In vitro translation

All *in vitro* transcription reactions used PCR-generated templates containing the SP6 promoter[51,52]. The transcription reactions were for 1 hour at 37°C. The resulting transcript was used without further purification and was diluted 1:20 in the IVT reaction, which was carried out in rabbit reticulocyte lysate (RRL) as described earlier[51,52]. Where indicated in the figure legends, the reaction was supplemented with either canine rough microsomes (cRMs) prepared and used according to the method of Walter and Blobel[53] or SP cells prepared as above. Labelling of nascent proteins was achieved by including ^{35}S-methionine (500 μCi/ml). Site-specific incorporation of the photo-crosslinkable amino acid benzoyl-phenylalanine (BPA) was achieved via amber suppression as described previously[54]. In brief, amber codon(s) were suppressed by supplementing translation reactions with 0.1 mM BPA, 5 μM *B. Stearothermophilus* $tRNA^{Tyr}$ with a CUA anti-codon, and 0.25 μM BPA-tRNA synthetase. In experiments with Sec61 inhibitors, the drugs (dissolved in DMSO) were added to a final concentration of 2 μM. Control samples without the inhibitor included an equivalent volume of DMSO. All translation reactions were incubated for 30 min at 32°C, then halted by transferring the samples to ice. All further steps were performed at 0-4°C, unless stated otherwise. Prior to SDS-PAGE analysis, the tRNA on RNCs was removed by adjusting the sample to 50 μg/ml RNaseA, 10 mM EDTA, 0.05 % SDS and incubating 10-15 min at room temperature.

Affinity purification of RNCs

Biochemical analysis of proteins associated with defined RNC intermediates (Figs. 1 and 4b; Extended Data Figs. 1e-g, 7a-c, g) was done by immunoblotting of products affinity purified via an epitope tag on the nascent chain. In the first step, microsomes from the IVT reactions were recovered by centrifugation at 4°C in the TLA55 rotor (Beckman) for 20 min at 55,000 rpm. The pellet was washed three times with 1XRNC buffer [50 mM HEPES-KOH, pH 7.5, 200 mM KOAc, 5 mM $Mg(OAc)_2$] then resuspended in one-fourth the volume of the original translation reaction. The resuspended microsomes were diluted 8-fold in solubilisation buffer (1XRNC buffer supplemented with 1.5% digitonin) and incubated for 10-30 min on ice. Insoluble material was sedimented for 15 min at 20,000 × g at 4°C in

a microcentrifuge and the supernatant was transferred to 20-50 µl anti-FLAG-M2 affinity resin (Sigma-Aldrich) or Streptactin sepharose (IBA Lifesciences) that had been equilibrated in 1XRNC buffer supplemented with 0.25% digitonin (wash buffer). After 2 h with gentle end-over-end rotation at 4°C, the beads were washed three times with wash buffer, then transferred to a new tube. The anti-FLAG resin was eluted with 0.25 mg/ml 3xFLAG peptide (Sigma-Aldrich) in wash buffer at 22°C for 30 min with agitation. Streptactin sepharose was eluted with 50 mM biotin in wash buffer on ice for 1 h. The eluates were analysed by SDS-PAGE and immunoblotting with the antibodies indicated in the figures.

Substrate architecture in the ribosome-translocon complex

Estimates of the position of substrate relative to the ribosome and translocon are based on established dimensions of the mammalian ribosome-Sec61 complex[32]. The exit tunnel of the ribosome accommodates a minimum of ~30 residues for an extended polypeptide and ~40 residues if the chain has some degree of secondary structure. Approximately 10 and 35 residues are needed to span the distance from the mouth of the exit tunnel to the cytosolic and lumenal sides of the membrane, respectively, assuming that the membrane-embedded part of the nascent chain is an alpha helical TMD. The distance spanned by an unstructured coil was assumed to be an average of ~ 2 Å per residue. Hence, the ~50 Å distance from the exit tunnel to the lumenal end of the Sec61 channel would be spanned by ~25 residues of unstructured polypeptide. This means a total of 55 to 65 residues is needed to span from the P site tRNA to the lumen through Sec61, consistent with the structure of an early secretory protein translocation intermediate[34]. Similarly, ~35 residues are needed to span the ~70 Å distance from the exit tunnel to the substrate-binding surface of Asterix. Crosslinks can be seen at shorter tether lengths[10], presumably because the nascent chain can sample more extended conformations that become trapped upon crosslinking.

PAT complex reconstitution in Asterix knockout SP cells

Functional analysis of Asterix and various mutants relied on replenishment of Asterix KO cells by *in vitro* translated Asterix (Fig. 2c-d, 3b; Extended Data Figs. 5b-f). SP cells prepared as described above were included during IVT of Asterix. The reaction contained ^{35}S-methionine in experiments where Asterix was radiolabeled (Fig. 2c; Extended Data Figs. Extended Data Figs. 5b, 5d, 5f) or cold methionine (at 40 µM) in experiments where unlabeled Asterix was subsequently tested for interaction with radiolabeled substrate (Fig. 2d, 3b; Extended Data Figs. 5c, 5e). Where desired, amber suppression reagents (see above) were included during the translation to incorporate BPA into specified sites in Asterix. Following translation, the reaction was cooled on ice and the SP cells were isolated by centrifugation through a 100 µl sucrose cushion at 10,000 rpm for 5-10 minutes in a microcentrifuge at 4°C. The cells were resuspended in 1X PSB prior to immediate use (without freezing) in downstream analyses as described below.

In separate parallel experiments, native CCDC47 immunoprecipitations (IP) were performed to verify successful reconstitution of each Asterix variant into the PAT complex (Extended Data Figs. 5b, 5f). In these analytical-scale experiments, ^{35}S-methionine labelled Asterix variants were translated in the presence of Asterix knockout SP cells in a total reaction volume of 20 µl. After translation, the SP cells were isolated by centrifugation through a

100 µl sucrose cushion at 10,000 rpm for 5-10 minutes in a microcentrifuge. The pellet was resuspended in 50 µl 1XRNC buffer and diluted 8-fold in 1XRNC buffer supplemented with 1.5% digitonin. After 10-30 min on ice, the insoluble material was removed by centrifugation for 15 min at 20,000 × g at 4°C in a microcentrifuge. The supernatant was incubated with end-over-end rotation for 3 h at 4°C with 1 µl anti-CCDC47 antibody and 5 µl of CaptivA Protein A sepharose (Repligen). The resin was washed thrice with wash buffer, transferred to a new tube, and eluted in SDS-PAGE sample buffer by heating to 95°C. Co-immunoprecipitated Asterix was detected by SDS-PAGE and autoradiography.

Preparation of substrate RNCs for downstream insertion assays

^{35}S-labelled substrate RNCs used for subsequent insertion assays were prepared by IVT. To isolate these RNCs, a 200 µl translation reaction of ^{35}S-methionine labelled truncated Rhodopsin was layered on a 2 ml 10-50% (w/v) sucrose gradient. The gradient was prepared in 11 × 34 mm centrifuge tubes (Beckman coulter, 347357) by successively layering 400 µl each of 50%, 40%, 30%, 20% and 10% sucrose solution in 1X PSB. Gradients were equilibrated for 30 minutes at 4°C prior to use. The samples were centrifuged in a TLS-55 rotor (Beckman coulter) at 55,000 rpm for 1 h, at 4°C with slow acceleration and deceleration settings. Eleven 200 µL fractions were successively collected from the top and analysed by SDS-PAGE and autoradiography. The two fractions containing the majority of RNCs (generally fractions 6 and 7) were pooled and supplemented with 1 mM GTP. The isolated RNCs were used directly in insertion reactions with PAT complex-reconstituted SP cells or snap-frozen in liquid nitrogen and stored at -80°C for later use in translocation reactions.

Photo-crosslinking via probes in Asterix

Site-specific Asterix-CCDC47 interactions (Fig. 2c; Extended Data Fig. 5d) were analysed in SP cells reconstituted of the PAT complex with ^{35}S-labelled Asterix variants containing BPA installed at defined sites by amber suppression as described above. The reconstituted resuspended SP cells were transferred to a 384-well plate on ice. The samples were irradiated for 10 min using UVP B-100 series lamp (UVP LLC) positioned 10 cm above the sample. The samples were either analysed directly by SDS-PAGE and autoradiography, or after denaturing IPs as described below.

Probing substrate-Asterix interactions in PAT complex reconstituted SP cells (Fig. 2d; Extended Data Fig. 5c, e) was accomplished by resuspending the sedimented SP cells in the isolated RNC mix at a concentration of 4×10^6 cells/ml. After an incubation for 10 min at 32°C to allow substrate insertion, the samples were placed on ice for either photo-crosslinking as described above (Extended Data Fig. 5e), or chemical crosslinking (Extended Data Fig. 5c) with 250 µM bismaleimidohexane (BMH, Thermo Fisher Scientific) for 15 min on ice. Chemical crosslinking reactions were quenched with 1 mM DTT. The samples were either analysed directly by SDS-PAGE and autoradiography, or after denaturing IPs (described below) as indicated in the figure legends.

Photo-crosslinking via probes in the substrate

In experiments shown in Figs. 3b, 4d and Extended Data Fig. 7d, photo-crosslinking utilised probes in the substrate. In Fig 3b, the ^{35}S-methionine labelled RNC containing BPA was generated in a membrane-free IVT reaction and isolated as described above. The isolated RNC was incubated with the desired SP cells for post-translational insertion as described above. The suspension after post-translational insertion was UV-irradiated, digested of tRNA as described above and either analysed directly or after IP as indicated in the figure legends. In Figs 4d and Extended Data Fig. 7d, the Rho-4TMD substrate RNC was translated with ^{35}S-methionine and the desired SP cells. Amber codons and amber suppression were used as described above to incorporate BPA into the substrate at defined positions as indicated in the figure. The SP cells were isolated, resuspended, and subjected to UV crosslinking as described above. To separate RNCs from prematurely terminated products, the RNCs were recovered by centrifugation. Here, the SP cells were lysed for 15 min on ice in 50 mM HEPES, pH 7.4, 100 mM KOAc, 5 mM $Mg(OAc)_2$, 0.5% Triton X-100, 1 mM DTT, and 1X EDTA-free protease inhibitor cocktail (Roche). The lysate was clarified by spinning for 10 min at 15,000 x *g* in a microcentrifuge at 4°C. The supernatant was layered onto 20% sucrose cushion in 50 mM HEPES, pH 7.4, 100 mM KOAc, 5 mM $Mg(OAc)_2$, 1 mM DTT in 8 × 34 mm centrifuge tubes (Beckman coulter, 343776). The sample was centrifuged at 100,000 rpm in TLA-120.1 rotor (Beckman coulter) at 4°C for 1 h. The supernatant, containing the prematurely terminated translation products, was discarded. The RNC pellet was resuspended and digested of tRNA as described above prior to either direct analysis by SDS-PAGE or denaturing IP using antibodies to Asterix or subunits of the Sec61 complex as indicated in the figure.

Cysteine-based chemical crosslinking

For experiment in Extended Data Fig. 7h ^{35}S-methionine labelled RNCs containing a cysteine in the first TMD were generated by IVT in the presence of different SP cells as indicated in the figure. The cells were then pelleted through a sucrose cushion and resuspended in 1X PSB. The suspension was supplemented with BMH to 250 μM concentration and incubated for 15 min on ice. The reaction was quenched with 1 mM DTT on ice for 5 min. The sample was then digested of tRNA as described above and denatured by heating to 95 °C for 10 minutes in SDS-PAGE sample buffer. The samples were either directly analysed by SDS-PAGE and autoradiography or subjected to denaturing anti-Asterix IP as indicated in the figure legend.

Protease protection assays

The topology of reconstituted Asterix in Asterix KO SP cells was assessed by a protease protection assay using trypsin (Extended Data Fig. 5b). After IVT of wild-type and mutant Asterix in the SP cells, the cells were isolated through a sucrose cushion and resuspended in 1X PSB. Trypsin was added to 0.25 mg/ml and the reaction was incubated for 1 hour on ice. The reaction was stopped by addition of 10-fold excess trypsin inhibitor. The SP-cells were re-isolated by centrifugation through a sucrose cushion and denatured in SDS-PAGE sample buffer. The samples were analysed directly or processed for denaturing anti-FLAG IP.

Europe PMC Funders Author Manuscripts

Proteinase K (PK) protection assays for different integral membrane or secreted proteins in Fig. 4e and Extended Data Fig. 9 was done directly following the translation reaction as described before[52,55]. In brief, translation reactions performed in the absence or presence of cRMs and Sec61 inhibitors (as described above) were put on ice, then divided into aliquots and adjusted to 0.5 mg/ml PK (or 10 μg/ml in Extended Data Fig 9e) and 1% Triton X-100 as indicated in the figure. After 1 h on ice, 5 mM of freshly-prepared PMSF in DMSO was added from a 250 mM stock and incubated for 2-5 min on ice to stop the reaction. The entire reaction volume was transferred to 10 volumes of boiling 1% SDS, 100 mM Tris-HCl, pH 8.0. The samples were then analysed by SDS-PAGE and autoradiography or western-blotting (for TNFα, which lacks methionine residues in PK-digested fragments). All Sec61 inhibitors were used at 2 μM as characterised before[15,48–50].

For the protease protection assay of Rho-4TMD substrate RNC (Extended Data Fig. 7f), the ^{35}S-methionine labelled translation intermediate was generated in the presence of WT or ΔTMCO1 SP cells. After translation, cells were pelleted through a sucrose cushion and resuspended in 1X PSB. Digestion with PK (0.5 mg/ml) in the presence or absence of detergent was done as described above. The reaction was quenched with PMSF as above and transferred to 10 volumes of boiling 1% SDS, 100 mM Tris-HCl, pH 8.0. The sample was then diluted and subjected to denaturing IP as described below. RNase (10 ng/ml) and EDTA (5 mM) were included during the IP to digest the peptide-associated tRNA.

Immunoprecipitations

Denaturing IPs were performed on samples denatured in SDS-PAGE sample buffer by heating for 10 minutes at 95°C. After cooling, the samples were diluted 10-fold in IP buffer [50 mM HEPES pH 7.5, 100 mM NaCl, 2.5 mM $Mg(OAc)_2$, 1% Triton X-100] and incubated for 2-3 hours at 4°C with either 5 μl of anti-FLAG-M2 affinity resin (Sigma-Aldrich), Streptactin sepharose (IBA Lifesciences), or CaptivA Protein A sepharose (Repligen) plus the desired antibody. The resin was washed three times with 0.5 ml each of IP buffer and eluted in SDS-PAGE sample buffer by heating to 95°C.

Mass spectrometry

Translation reactions containing transcripts coding for early or late insertion intermediates were subjected to affinity purification of RNCs as described above. Instead of eluting the RNCs from the resin, they were washed twice in RNC buffer without digitonin. The samples were subjected to on-bead digestion with trypsin followed by peptide tandem mass tag (TMT) labelling (Thermo Fisher Scientific, cat. #90110) and protein identification by mass spectrometry as described previously[54].

In brief, the trypsin digested samples were desalted using home-made C18 stage tips (3M Empore) filled with poros R3 (Applied Biosystems) resin. The stage tips were equilibrated with 80% acetonitrile (MeCN)/0.5% formic acid (FA) followed by 0.5% FA. Bound peptides were eluted with 30-80%MeCN/0.5% FA and lyophilized. Dried peptide mixtures from each condition were resuspended in 40 μl of 200 mM HEPES (pH 8.5). 20 μl of TMT 10plex reagent (Thermo Fisher Scientific), reconstituted in 41 μl anhydrous MeCN, according to manufacturer's instructions, was added and incubated at room temperature for 1 hr. The

labelling reaction was then terminated by incubation with 4 μL 5% hydroxylamine for 30 min. Labelled samples were subsequently pooled into a single sample. The acetonitrile was evaporated in a SpeedVac, desalted (using stage tips prepared according to the method described above) and then fractionated using 10 mM ammonium bicarbonate and acetonitrile gradients. The eluate was acidified, concentrated in a SpeedVac, and subjected to Liquid chromatography (LC)-MS/MS using a fully automated Ultimate 3000 RSLC nano System (Thermo Fisher Scientific) fitted with a 100 μm × 2 cm PepMap100 C18 nano trap column and a 75 μm × 25 cm reverse-phase nanoEase M/Z HSS C18 T3 column (Waters). Peptides were separated using a binary gradient consisting of buffer A (2% MeCN, 0.1% formic acid) and buffer B (80% MeCN, 0.1% formic acid). Peptides were eluted at 300 nL/min flow rate with an increasing acetonitrile gradient. The HPLC system was coupled to a Q Exactive Plus hybrid quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific) equipped with a nanospray ionisation source.

The acquired MS/MS raw files were processed using Proteome Discoverer (version 2.1, Thermo Scientific). MS/MS spectra were searched against *Canis lupus familiaris* (downloaded on 30/03/2021) and *Oryctolagus cuniculus* (downloaded on 09/02/2021) UniProt Fasta databases using Sequest search engine. MS/MS hits were analysed using Scaffold (Proteome Software Inc). The complete data plotted in Fig. 4a is provided in Supplementary Table 1.

Sample preparation for electron microscopy

In vitro transcription and translation reactions were performed as described above using cRMs. A 2 ml translation reaction was chilled on ice, divided in four, and each aliquot layered on a 500 μl cushion of 20% sucrose in 1XRNC buffer. The microsomes were sedimented by centrifugation at 4°C in the TLA-55 rotor (Beckman) at 55,000 rpm for 20 min. The cRM pellets were each resuspended in 25 μl of 1X RNC buffer and pooled. The Rho-2TMD sample was incubated with 250 μM BMH on ice for 15 min and quenched with 5 mM 2-mercaptoethanol. The Rho-4TMD sample was not crosslinked.

The microsomes were diluted with 400 μl of solubilisation buffer (1X RNC buffer containing 1.5% digitonin) and incubated for 10 min on ice. The digitonin was obtained from Calbiochem and further purified as described previously[5]. The sample was centrifuged at 20,000 × *g* and 4°C for 15 min. The supernatant was transferred to a tube containing 20 μl of StrepTactin High Performance Sepharose beads (GE Healthcare) and incubated for 1.5 h at 4°C. The resin was then washed five times with 0.5 ml 1X RNC buffer containing 0.25% digitonin and eluted by incubation for 1 h on ice with 40 μl of 1X RNC buffer containing 0.25% digitonin and 50 mM biotin. The absorbance of the eluate for both samples was 3.4 at 260 nm. They were applied to grids immediately without further dilution or adjustments.

Electron microscopy

The affinity-purified RNCs were vitrified on UltrAuFoil R 1.2/1.3 300-mesh grids (Quantifoil) coated with graphene oxide (Sigma-Aldrich). In a Vitrobot Mark IV (Thermo Fisher Scientific) at 4°C and 100% ambient humidity, each grid was loaded with 3 μl of

sample, blotted 4 sec with Whatman filter papers at a blot force of -15, and plunge frozen in liquid ethane at 92 K.

Automated data collection was performed on a Titan Krios microscope (Thermo Fisher Scientific) equipped with an XFEG source operating at an accelerating voltage of 300 kV. Defocus was programmed to range between 2.7 and 1.9 μm. Movies were captured using a K3 Bioquantum direct electron detector (Gatan) operating in super-resolution mode. Movies were dose-fractionated into 54 frames covering a total dose of 54 $e^-/Å^2$. One dataset was collected per sample. For the Rho-2TMD sample, 17,540 images were collected at 105,000× magnification (0.83 Å/px, or 0.415 Å/px in super-resolution). For the Rho-4TMD sample, 13,755 images were recorded using Gatan K3 Bioquantum direct electron detector in super-resolution mode at 53,000× magnification (1.39 Å/px, or 0.69 Å/px in super-resolution).

Image processing

All following steps except particle picking were performed in RELION-3.1. Movie frames were motion-corrected using MotionCor2 with 7×5 patches and dose-weighting. Contrast transfer function parameters were estimated using CTFFIND-4.1. Particles were picked using either cryoSPARC 3.2 (for the Rho-2TMD dataset) or crYOLO (for the Rho-4TMD dataset). 2,444,921 Rho-2TMD and 1,454,565 Rho-4TMD particles were picked. Rho-2TMD particles were then extracted in 412-px boxes, and Rho-4TMD particles in 386-px boxes. These particles were downsampled to 3 Å/px and subjected to 2-D classification. Classes containing 1,665,551 Rho-2TMD and 1,445,986 Rho-4TMD particles were retained. 3-D refinement was then performed with a mammalian ribosome lowpass filtered at 70 Å as a reference.

Focused classification with partial signal subtraction (FCwSS) was then used to identify the subset of particles containing the PAT complex. Signal outside a soft mask encompassing the translocon and detergent micelle was subtracted from the particle images, which were then subjected to 3-D classification using fixed alignment parameters determined during the 3-D refinement. Division into 7 classes yielded one class with strong density for the PAT complex. For the Rho-2TMD dataset this class comprised 9% or 148,833 particles, and for the Rho-4TMD, 9.7% or 136,812 particles. These particles were then re-extracted without downsampling and 3-D refined to obtain Rho-2TMD and Rho-4TMD maps at overall resolutions of 3.25 Å and 3.88 Å, respectively.

We attempted to improve different regions of the map by further FCwSS around individual translocon sub-complexes. For the PAT and BOS sub-complexes, this strategy yielded moderately improved density, particularly for the TMD of Nicalin and the ribosome-proximal regions of CCDC47, but no qualitatively new information beyond what could be derived from the Rho-2TMD reference map. Attempts to improve the very weak density for the GEL complex by FCwSS were unsuccessful, so this was not included in the final model.

Model building, refinement and validation

The 60S subunit and P-site tRNA from PDB ID 6T59 were used as a starting model for the rabbit ribosome. Models for the four canine translocon subcomplexes were generated using the ColabFold implementation (AlphaFold2_advanced.ipynb notebook)[56] of Alphafold[30].

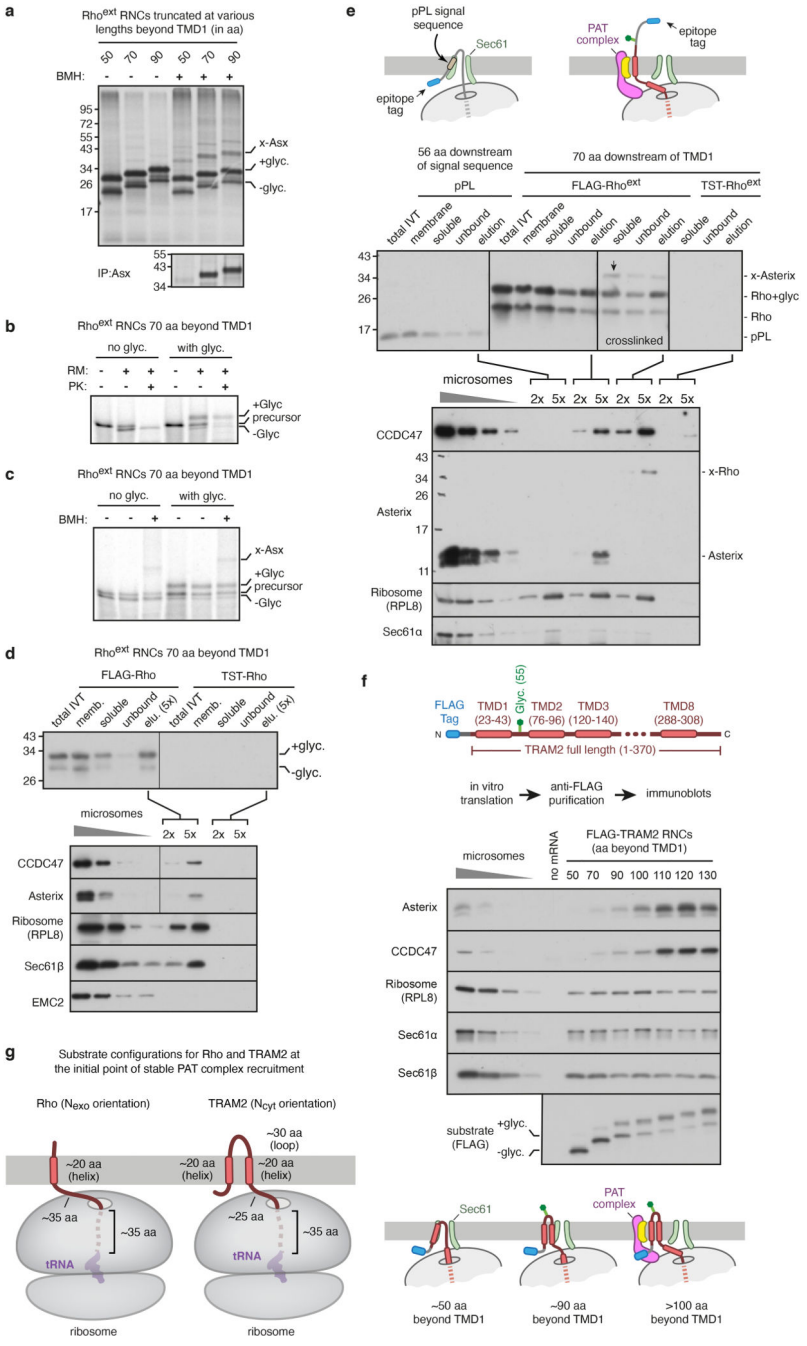
Default settings were used, except that the top-scoring models in each case were refined with Amber-Relax. The quality of each complex was assessed by the predicted local-distance difference test (pLDDT), which provides a per-residue confidence score for each subunit, and the Predicted Aligned Error (PAE), which provides a confidence measure of the predicted protein-protein interface. Models were further validated by their fit to the cryo-EM maps, site-specific photo-crosslinking or comparison with previously determined structures.

For the Rho-2TMD model, the 60S ribosomal subunit and P-site tRNA were placed as rigid bodies into Rho-2TMD reference map. The C-terminal 38 residues of the stalled nascent chain were built into density in the ribosome exit tunnel. Residues 219-250 of eL6 (RPL6), which pack against a portion of the CCDC47 cytosolic domain, were built into density, and a portion of the 5.8S rRNA loop that contacts TMEM147 was manually adjusted. The Sec61 and PAT complex models were placed into Rho-2TMD submap 2 and each complex was adjusted as a single unit (to maintain the predicted interchain contacts) using tightly restrained real-space refinement in COOT[57]. The BOS complex (excluding NOMO) was placed into Rho-2TMD submap 1 and adjusted similarly. Density for the GEL complex was very weak in the Rho-2TMD sample, and therefore not included in the final model. Additional low resolution density in the lumen in the approximate region previously ascribed to the TRAP complex was not assigned.

The Rho-4TMD model was built similarly. After placing the Rho-2TMD model into Rho-4TMD map, the GEL complex was fit into density and adjusted as a single unit using tightly restrained real-space refinement in COOT. The C-terminal 70 residues (including TMD3) of the stalled Rho-4TMD intermediate were built into density that extends nearly continuously from the ribosomal peptidyl transferase centre into the membrane. The register of the TMD3 helix is approximate, and is based on site-specific photo-crosslinking results showing F115 adjacent to Sec61β and W126 adjacent to Sec61α.

Real-space refinement of the Rho-2TMD and Rho-4TMD models was done in PHENIX[58,59] against the Rho-2TMD reference map and the Rho-4TMD map, respectively. Three rounds of global minimisation and group B-factor refinement were performed, with tight secondary structure, reference model, rotamer, and Ramachandran restraints applied. Secondary structure- and reference model restraints were determined from the starting models. Hydrogen-bonding and base-pair and stacking parallelity restraints were applied to the rRNA. Final model statistics are provided in Extended Data Table 1. Structure figures were generated with UCSF ChimeraX[60].

Extended Data

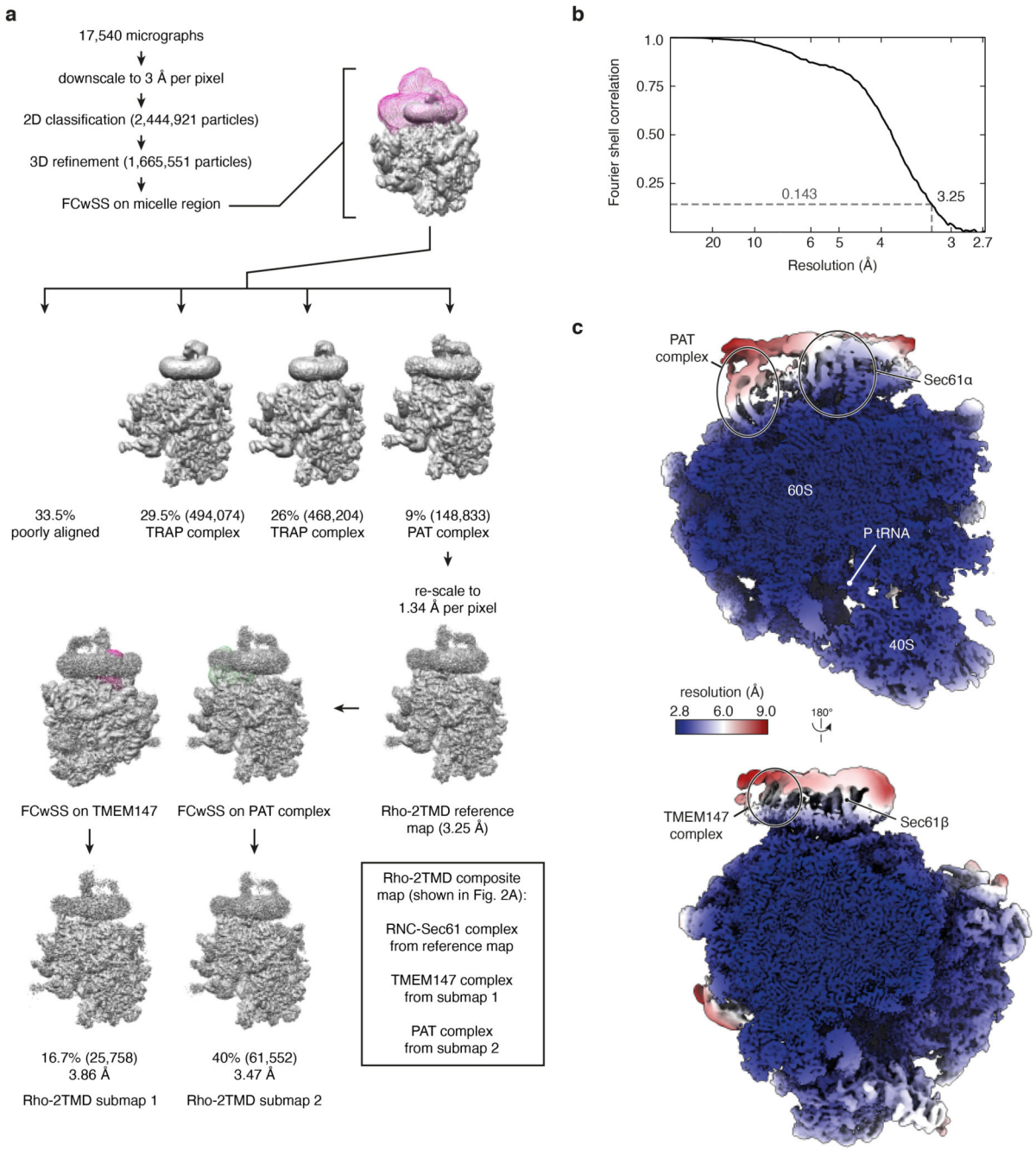


Extended Data Fig. 1. Characterisation of the Rho^{ext} construct.

a, ^{35}S-methionine labelled Rho^{ext} ribosome nascent chain complexes (RNCs) of indicated length were synthesised in rabbit reticulocyte lysate (RRL) in the presence of canine pancreas-derived rough microsomes (RMs). Where indicated, RNCs were chemically crosslinked using bismaleimidohexane (BMH). The positions of the glycosylated (+glyc.) and non-glycosylated (-glyc.) translation products are indicated. The crosslink to Asterix (indicated by x-Asx) was verified by immunoprecipitation under denaturing conditions using

anti-Asterix antibody (bottom panel). The pattern of PAT complex recruitment for Rho^{ext} construct is similar to the non-extended Rho construct construct described previously[10]. **b**, ^{35}S-methionine labelled Rho^{ext} RNC of the indicated length, lacking or containing the glycosylation site, was synthesised in RRL containing or lacking RMs. The translation products were then digested with proteinase K (PK). The population of polypeptides inserted into the membrane is protected from PK. **c**, Translation products were produced as in panel b (lanes 1 and 4), after which the RMs were isolated by sedimentation (lanes 2 and 5) and subjected to crosslinking with BMH (lanes 3 and 6). The crosslink to Asterix is indicated. **d**, FLAG or twin-Strep tag (TST) containing membrane-inserted RNCs of Rho^{ext} truncated 70 amino acids (aa) beyond TMD1 were subjected to fractionation and affinity purification. The total IVT products were centrifuged to obtain a membrane fraction, which was then solubilised under non-denaturing conditions (soluble fraction). The soluble fraction was then subjected to anti-FLAG affinity purification. Aliquots of the purification (with five-fold more loaded for the elution fraction) were analysed by anti-FLAG immunoblot (top panel). The bottom panel shows the elution fractions of the two purifications immunoblotted for CCDC47, Asterix, RPL8, Sec61b, and EMC2. Serial dilutions of ER microsomes were analysed in parallel. RNCs of Rho^{ext} can be efficiently affinity-purified, recovering the associated PAT complex. Note that the left and right lanes of the CCDC47 and Asterix blots are from the same gel and taken from the same exposure, with the vertical line indicating the point where the lanes were spliced together. **e**, Membrane-inserted RNCs of bovine preprolactin (pPL) truncated 56 residues beyond the signal sequence[34] and Rho^{ext} truncated 70 residues beyond TMD1 were subjected to anti-FLAG affinity purification as in panel d. The TST-tagged Rho^{ext} served as a specificity control. Where indicated, the sample was crosslinked with BMH just prior to solubilisation of RMs. The top panel shows the anti-FLAG immunoblot of the steps of affinity purification (as in panel d), and the bottom panel shows the elution fractions of each purification immunoblotted for CCDC47, Asterix, RPL8, and Sec61α. The blots indicate that the pPL translation intermediate does not associate with the PAT complex, while the PAT complex is recovered with Rho^{ext} with comparable efficiency without or with crosslinking, which proved to be nearly quantitative as judged by the near absence of non-crosslinked Asterix. This indicates that association of the PAT complex with Rho^{ext} RNCs is salt- and detergent-resistant under the purification conditions employed. The faint band in the final lane of the CCDC47 is IgG heavy chain contamination. **f**, The top panel shows a diagram depicting the construct encoding full-length N-terminal FLAG-tagged TRAM2, whose topology is opposite to that of Rho. The bottom panel shows an experiment analogous to that for Rho^{ext} shown in Fig. 1b. The purified samples were analysed by immunoblot adjacent to serial dilutions of RMs. Equal translation levels and recovery of RNCs is reflected by both the substrate blot and by similar levels of Sec61 subunits and the ribosome. A reaction lacking mRNA served as a negative control for non-specific binding to the affinity resin. Note that as for Rho^{ext} (Fig. 1b), recruitment of the PAT complex to TRAM2 translation intermediates is dependent on nascent chain length. Although a small amount of recruitment is detectable at lengths of 70-100 aa beyond TMD1, the initial point of stable maximal recruitment is observed at 110 aa beyond TMD1. **g**, Diagrams showing the substrate configurations, approximately to scale, for Rho and TRAM2 at the point of initial stable recruitment of the PAT complex. This corresponds to ~70 aa beyond TMD1 for Rho and ~110 aa beyond TMD1 for TRAM2. The

difference in length can be accounted for by the different topology of TMD1. The translocon is not shown for simplicity.



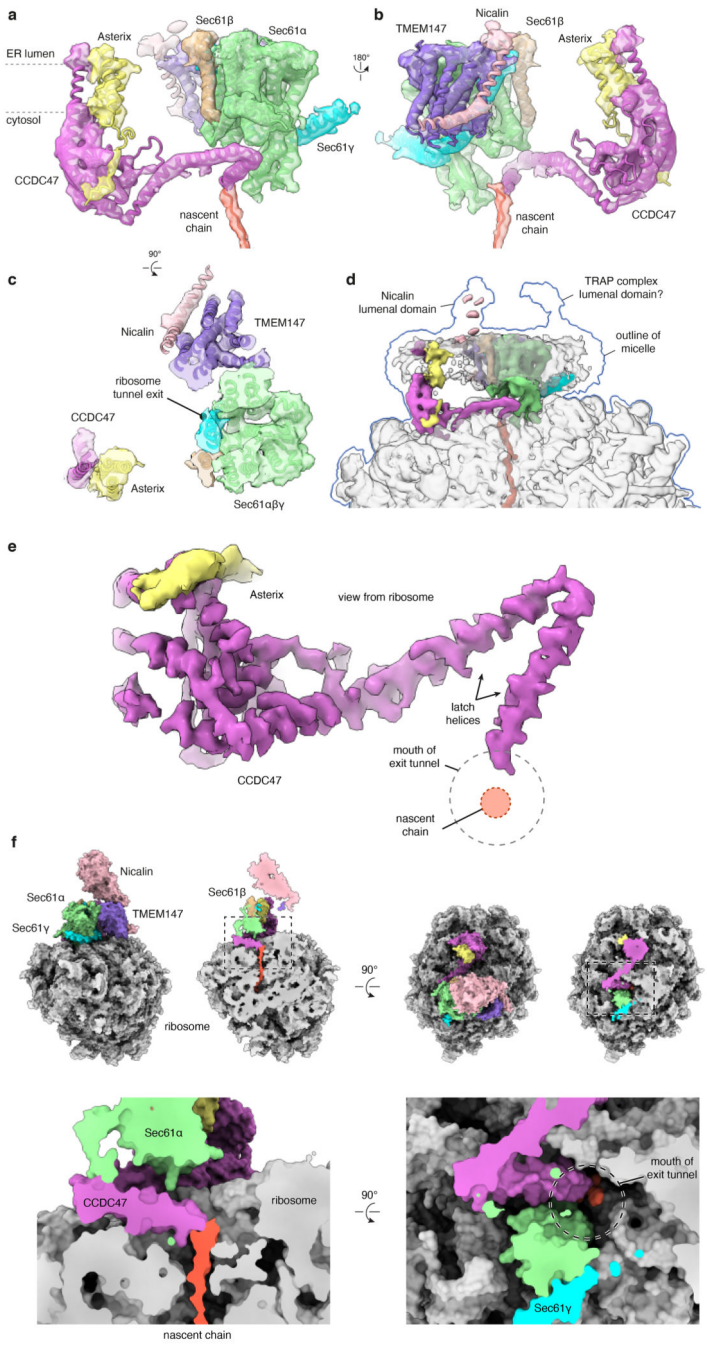
Extended Data Fig. 2. Processing scheme and resolution information of the Rho-2TMD map.

a, Flowchart depicting the classification scheme and analysis of cryo-EM micrographs of the Rho-2TMD RNC sample. The classes denoted as “TRAP complex” appear to only contain the core translocon comprised of the Sec61 and TRAP complexes, evidently in somewhat different conformations, as characterised previously in cryo-EM and cryo-electron

Europe PMC Funders Author Manuscripts

tomography work[26,61–63]. These were not pursued further. Note that model building relied almost entirely on the Rho-2TMD reference map, with a few regions employing the two submaps for verification. The composite map shown in Fig. 2a is solely for illustration purposes as a means of simultaneously displaying the best regions contributed by each map. FCwSS indicates focused classification with signal subtraction. **b**, Fourier shell correlation (FSC) curve for the Rho-2TMD reference map illustrating an overall resolution of 3.25 Å by the gold-standard method[64]. **c**, Two views of the Rho-2TMD reference map coloured by local resolution. Key structural elements are indicated.

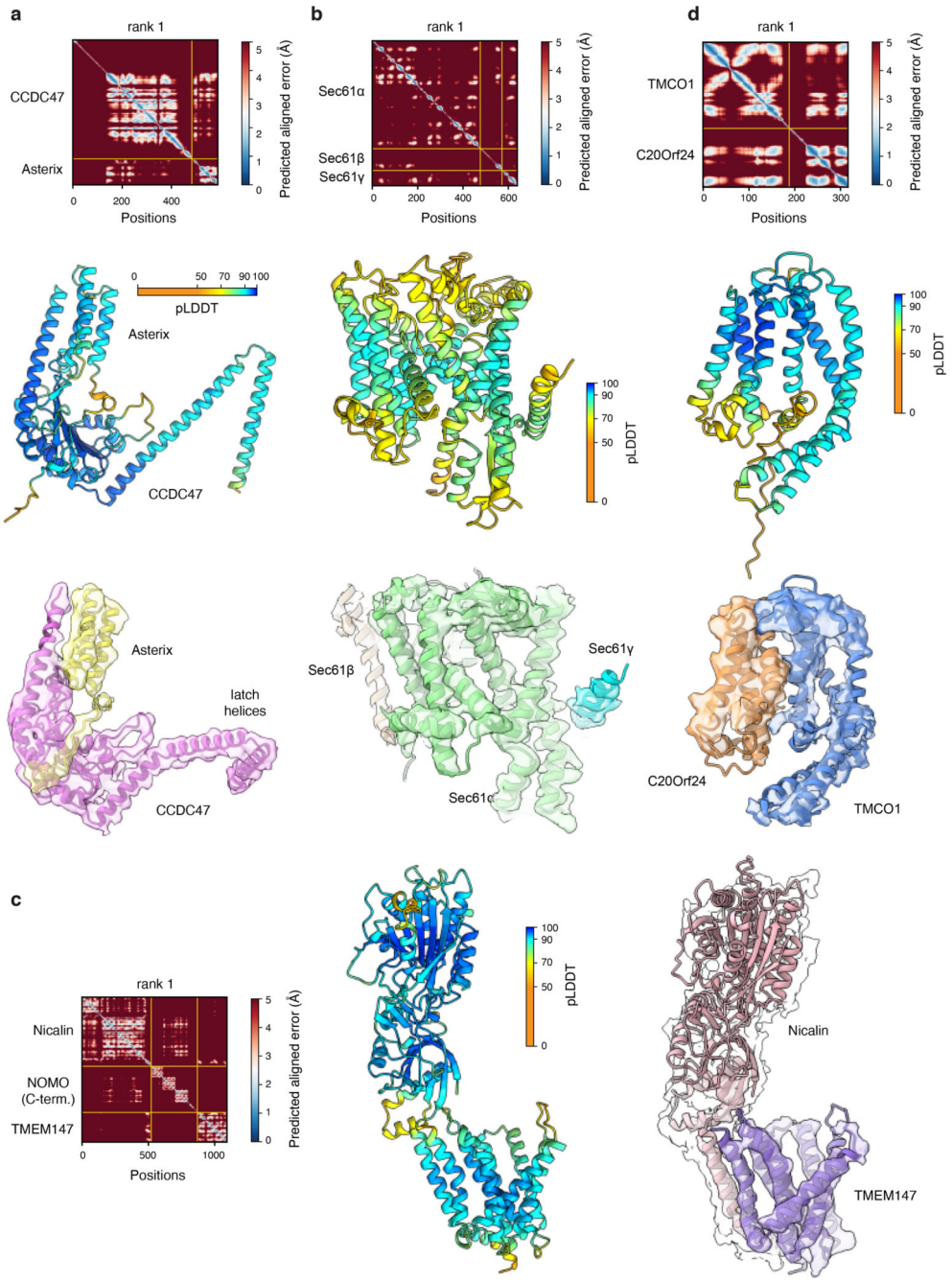
Europe PMC Funders Author Manuscripts



Extended Data Fig. 3. Features of the Rho-2TMD map and model.

a, b, Two views from the plane of the membrane of the Rho-2TMD reference map, low-pass filtered to 7 Å resolution, fitted with the model. The approximate position of the membrane is indicated. The detergent micelle and ribosome are omitted for clarity. **c**, View from the ER lumen of the low-pass filtered Rho-2TMD reference map fitted with the model. The clipping plane is within the membrane close to the ER lumen. The detergent micelle and ribosome are omitted for clarity, with the position of the ribosome tunnel exit indicated. **d**, Overview of the low-pass filtered reference map in the context of the detergent micelle

and ribosome. The map is shown at two contour levels. The blue line shows a contour level at which lumenal density and the complete micelle is visible. The opaque density is shown at a lower contour level to visualise the translocon, coloured as in panels a-c. The ribosome and micelle are in transparent grey. **e**, The PAT sub-classified map (submap 2 in Extended Data Fig. 2) filtered by local resolution viewed from the ribosome to illustrate that the regions closest to (and stabilised by) the ribosome are particularly well resolved. The position of the nascent chain density and the mouth of the ribosome exit tunnel are indicated. **f**, Space-filling depictions of the Rho-2TMD model illustrating how the C-terminal region of CCDC47 enters the mouth of the ribosome exit tunnel, abuts the nascent chain, and narrows the exit tunnel dimensions. The top shows overviews before and after clipping, with dashed boxes indicating the regions shown at higher magnification below.



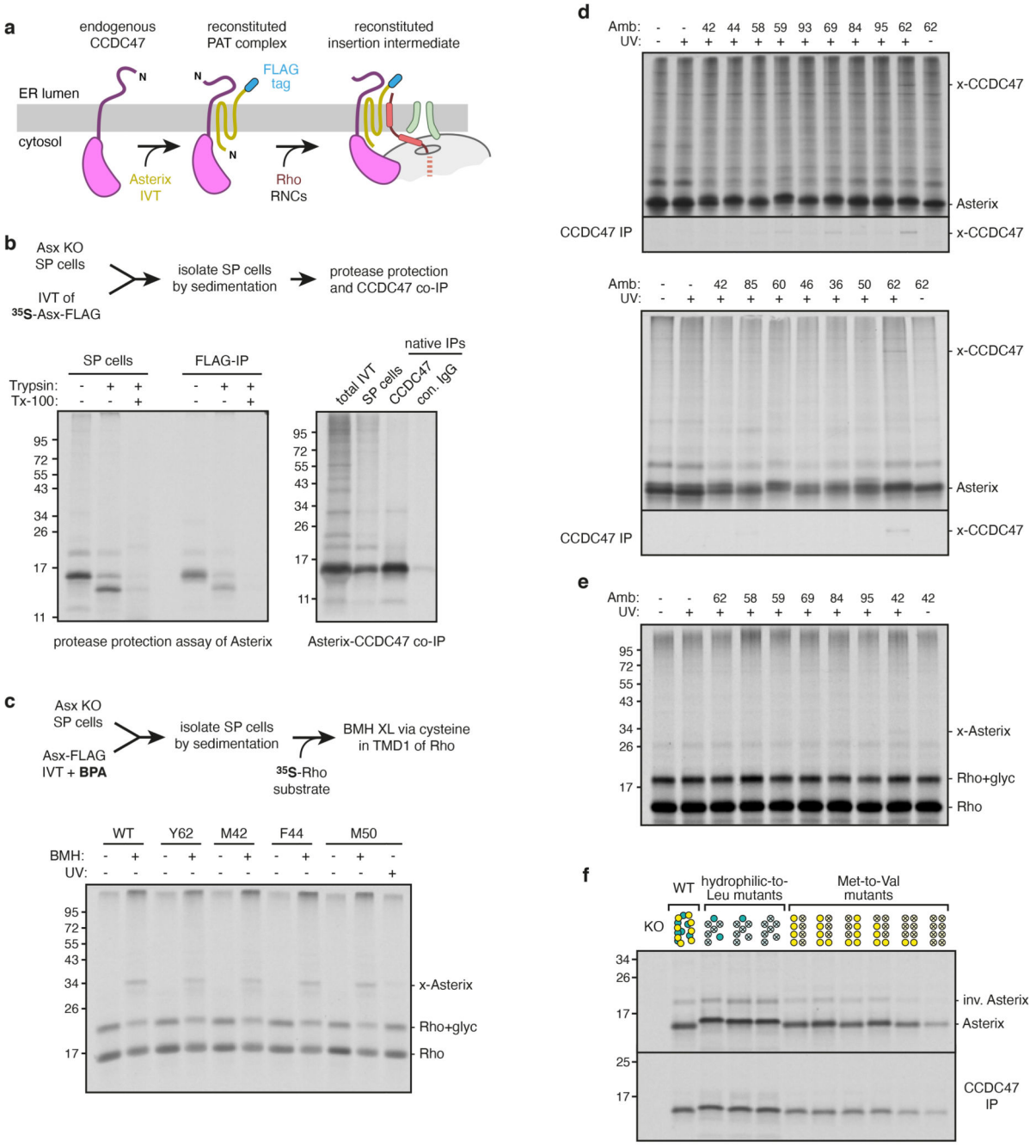
Extended Data Fig. 4. AlphaFold2 predictions of the PAT, Sec61, GEL and BOS complexes.

All AlphaFold2 predictions were obtained using ColabFold. For all complexes, predicted alignment error (PAE) matrices, structural models coloured by the predicted Local Distance Difference Test (pLDDT) and the final models fitted into respective densities are shown. Note that the PAE matrix scale is from 0-5 Å, not the default 0-30 Å output, to emphasise the very high-confidence interactions. **a,** Human PAT complex, comprised of CCDC47 and Asterix. The final model is fitted into the respective density taken from the PAT sub-classified map (submap 2 in Extended Data Fig. 2) filtered by local resolution. **b**,

Europe PMC Funders Author Manuscripts

Canine Sec61 complex. The final model is fitted into the respective density taken from the Rho-2TMD reference map filtered by local resolution. **c**, Human BOS complex, comprised of TMEM147, Nicalin, and NOMO. The final model (which omitted NOMO because it was not visualised in the map) is fitted into the respective density taken from the TMEM147 sub-classified map (submap 1 in Extended Data Fig. 2) filtered by local resolution. The density for Nicalin is also shown at a very low threshold in transparent white to visualise the lumenal domain density. **d**, Human GEL complex, comprised of TMCO1 and OPTI (C20orf24). The final model is fitted into the respective density taken from the Rho-4TMD map filtered by local resolution.

Europe PMC Funders Author Manuscripts

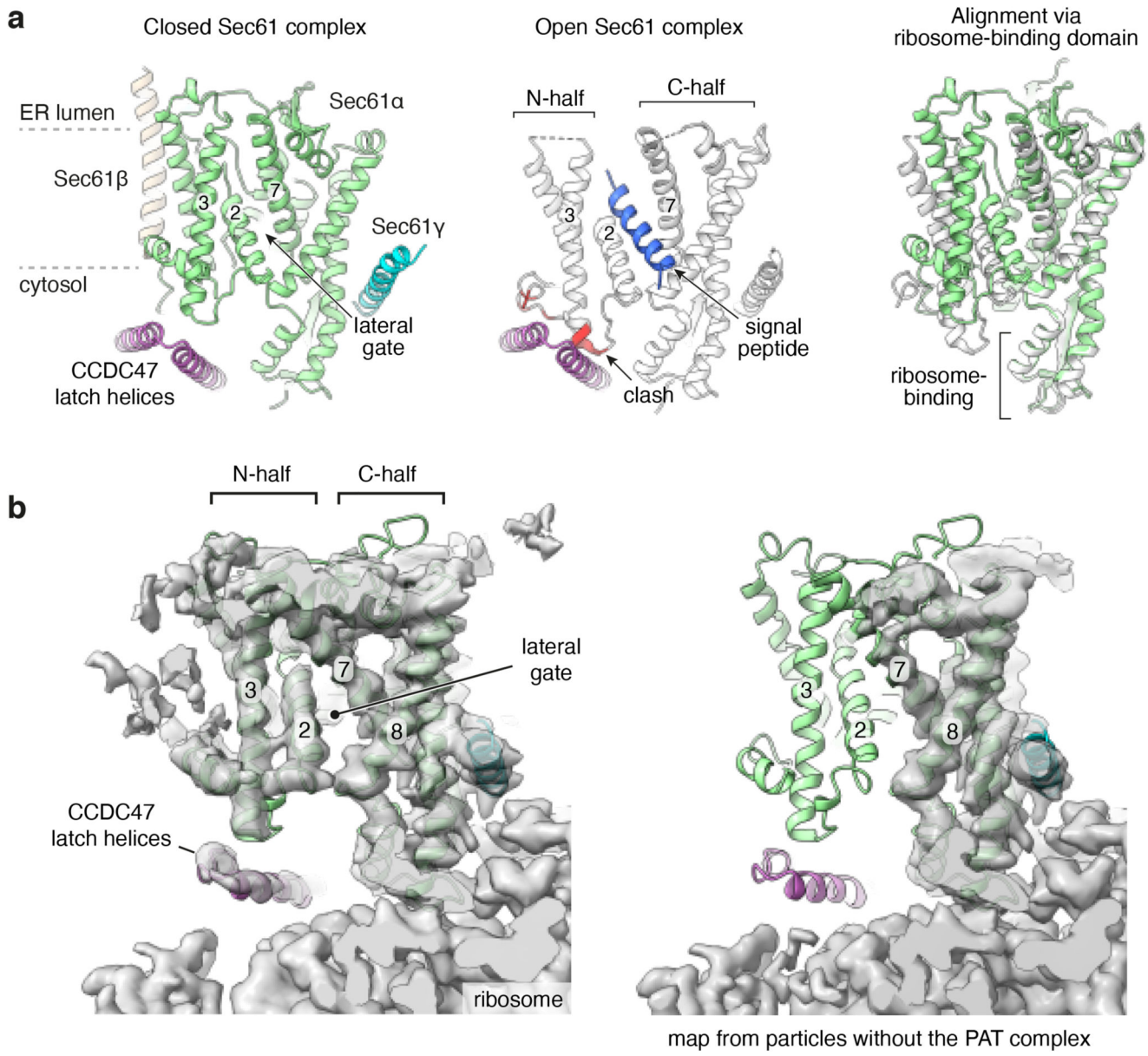


Extended Data Fig. 5. Characterisation of PAT complex reconstitution in SP cells.

a, Diagram illustrating the strategy for PAT complex reconstitution with Asterix variants followed by analysis of substrate interaction. Asterix KO cells contain residual levels of endogenous CCDC47. When Asterix is *in vitro* translated in RRL, supplemented with Asterix KO semi-permeabilised (SP) cells, it is inserted and interacts with CCDC47 to reconstitute the PAT complex. When substrate RNCs are subsequently introduced, their interaction with the PAT complex can be tested. **b**,^{35}S-labelled Asterix was translated in the presence of Asterix KO SP cells. The cells were then isolated by centrifugation

Europe PMC Funders Author Manuscripts

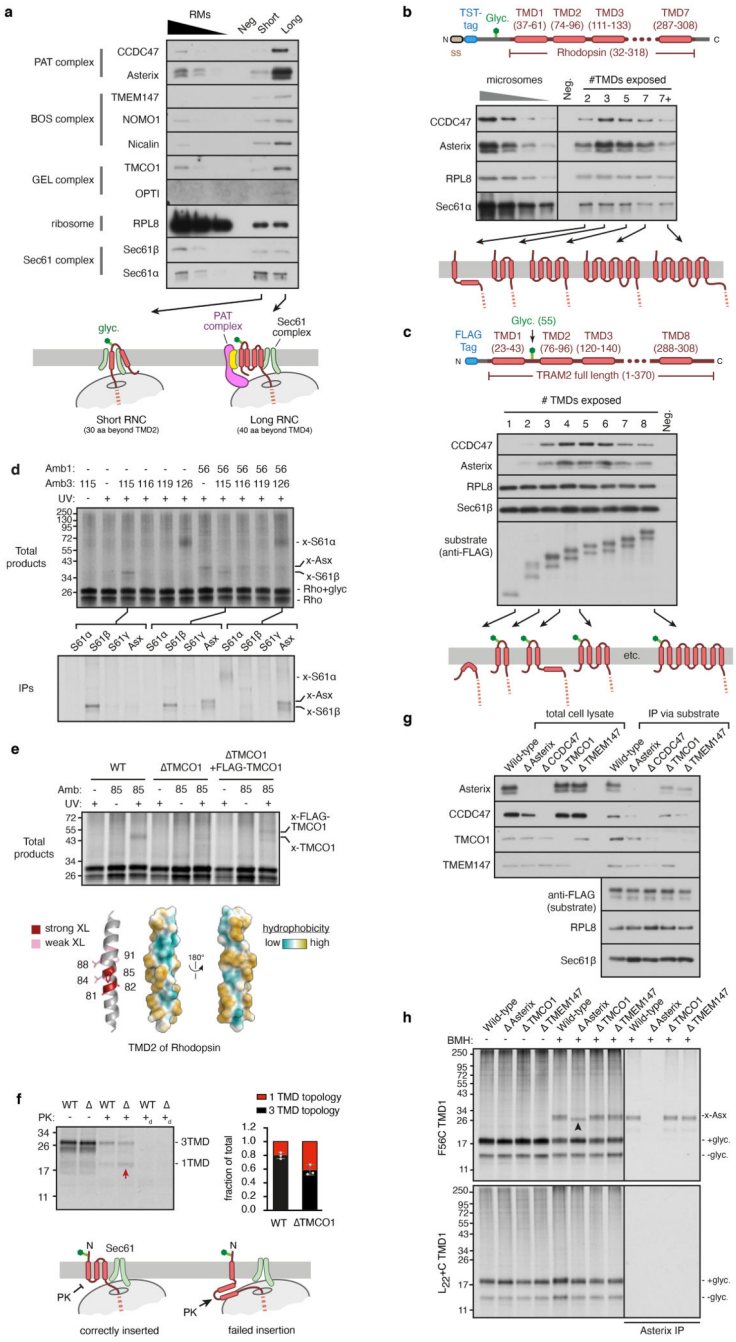
and analysed by protease protection to check Asterix topology (left) and interaction with CCDC47 by co-immunoprecipitation (IP; right). The cytosolic N-terminus of Asterix is accessible to protease (trypsin), and the remaining protected fragment is recoverable by immunoprecipitation via the C-terminal FLAG-tag. The right panel shows the total *in vitro* translated products (lane 1), the sedimented SP cells (lane 2), native IP with CCDC47 antibody (lane 3) or native IP using control antibody (lane 4). **c**, Wild-type and amber codon-containing Asterix variants (at the indicated positions) were translated in RRL (with non-radioactive methionine) containing Asterix KO SP cells and amber suppression reagents for site-specific incorporation of the photocrosslinking amino acid BPA. The cells were isolated and incubated with ^{35}S-labelled Rho RNCs truncated at 70 residues beyond TMD1. The samples were then crosslinked using bismaleimidohexane (BMH) where indicated. The only cysteine of the substrate is located in the first TMD. Note that all Asterix variants form BMH-mediated crosslinks with the substrate, indicating successful reconstitution of the substrate-Asterix interaction. For comparison, the sample containing BPA at M50 of Asterix was UV-irradiated, illustrating that photo-crosslinks between Asterix and substrate can also be visualised (see also Fig. 2d and panel e). As expected, the efficiency of photo-crosslinking is lower than chemical crosslinking. **d**, Autoradiographs of photocrosslinking experiments between Asterix and CCDC47. ^{35}S-labelled Asterix variants with BPA at the indicated positions were reconstituted into Asterix KO SP cells as in panel c. The cells were isolated, irradiated with UV light, and analysed directly or after denaturing IP using anti-CCDC47 antibody. **e**, Non-radioactive asterix variants with BPA at indicated positions were reconstituted into Asterix KO SP cells as in panel c. The reconstituted SP cells were then incubated at 32°C for 10 min with isolated ^{35}S-labelled Rho RNCs truncated 70 residues beyond TMD1. The samples were irradiated with UV light and analysed by SDS-PAGE and autoradiography. The positions of non-glycosylated and glycosylated Rho, and the crosslink to Asterix (x-Asterix) are indicated. Only position 42 showed a strong crosslink among those tested in this experiment. **f**, Native CCDC47 IPs for the Asterix mutants analysed in Fig. 3b. ^{35}S-labelled Asterix variants were translated in RRL supplemented with Asterix KO SP cells. The SP cells were isolated by centrifugation, solubilised under native conditions, and subjected to immunoprecipitation using anti-CCDC47 antibody. Aliquots of the isolated SP cells (top panel) and products of CCDC47 IP (bottom panel) are shown. Note that a small population of Asterix is evidently inserted in the inverted orientation (“inv. Asterix”) and becomes glycosylated. This population does not co-IP with CCDC47, providing an internal specificity control. The scheme for depicting the positions that are mutated is the same as shown in Fig. 3.



Extended Data Fig. 6. CCDC47 latch helices disfavour Sec61 opening.

a, The Sec61 complex in the Rho-2TMD model (left) is in a closed conformation with the latch helices of CCDC47 abutting the cytosolic loop between TMD2 and TMD3 of Sec61α. The open Sec61 complex (middle, PDB 3JC2) would clash with CCDC47 because the N-half of Sec61 would need to rotate away from the ribosome-bound C-half to accommodate a substrate at the lateral gate (such as a signal peptide in this structure). The right panel shows an alignment of the two structures by their ribosome-binding domains, illustrating how the N-half of Sec61 needs to rotate during lateral gate opening. **b**, Experimental cryo-EM density of reconstructions from the Rho-2TMD dataset from particles containing (left) or lacking (right) the PAT complex. At comparable contour levels where the ribosome and the C-half of Sec61 are essentially the same between the two maps, the N-half density

is markedly better in the CCDC47-containing map. At more generous contour levels, the N-half can be seen in the CCDC47-lacking map, albeit at lower resolution due to presumed heterogeneity of positions. The differences in Sec61 density in particles without and with the PAT complex cannot be ascribed to refinement differences. The particle alignments during refinement are dominated by the strong signal provided by the ribosome, which does not differ between the two classes. No realignment was performed after focused classification. The fact that the density for the ribosome and many parts of Sec61 is very similar between the two maps provides post hoc validation that the refinement procedure was consistent. Differences in overall flexibility of the translocon-micelle complex relative to the ribosome can be excluded because significant density differences are seen only in the parts of Sec61 known to move during gating. By contrast, parts of Sec61 that are distal to the RBD but unmoved by gating remain unchanged. For these reasons, we ascribe the difference in Sec61 density between the PAT-containing and PAT-free maps to the former being stabilised in the closed state.

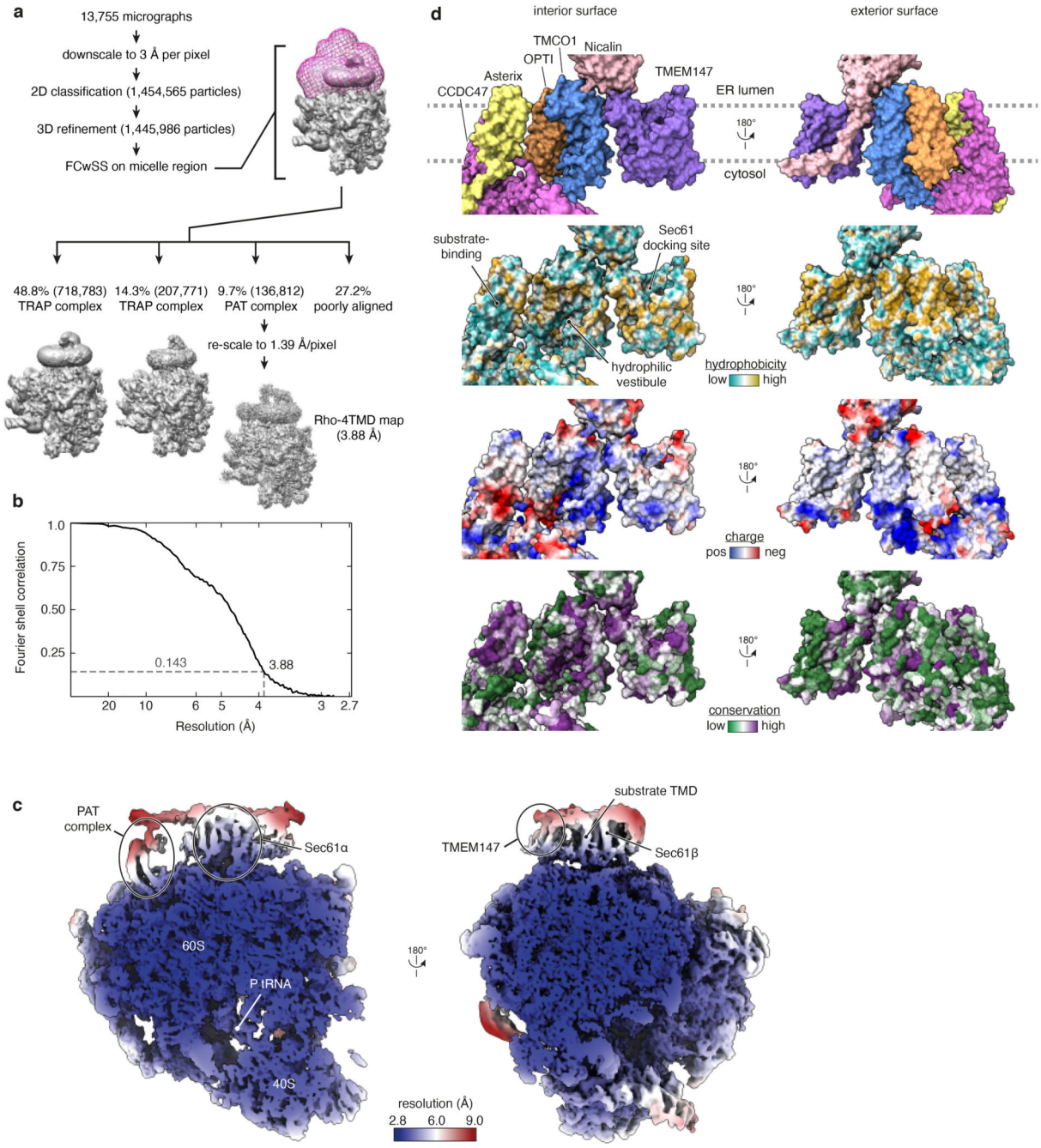


Extended Data Fig. 7. Analysis of the substrate-engaged multipass translocon.

a, RNCs of FLAG-TRAM2 truncated at the indicated lengths (see diagram) were synthesised in RRL supplemented with HEK293 SP cells. The RNCs were affinity-purified under native conditions via the N-terminal FLAG-tag (as in Extended Data Fig. 1f) and analysed by immunoblotting for the indicated proteins. The negative control (neg) is a translation reaction without mRNA. **b**, RNCs of Twin-Strep Tag (TST)-Rho^{ext} (see diagram) truncated 40 residues downstream of TMD2, TMD3, TMD5 or TMD7 were analysed for PAT complex association as in Fig. 1b. The RNC of 7+ length represents a nascent chain

Europe PMC Funders Author Manuscripts

truncated 70 residues beyond TMD7. **c**, RNCs of FLAG-TRAM2 (see diagram) truncated 40 residues downstream of indicated TMD were analysed for PAT complex association as in Extended Data Fig. 1f. **d**, Photo-crosslinking analysis of ^{35}S-methionine-labelled Rho-4TMD translocation intermediate (as in Fig. 4d) via BPA installed at amber codons in TMD1 (Amb1), TMD3 (Amb3), or both. The top panel shows the autoradiograph of total RNCs after isolation by ultracentrifugation. Three of the key samples were also subjected to immunoprecipitation using the indicated antibodies and analysed in the bottom panel by autoradiography. Non-glycosylated and glycosylated substrate (Rho and Rho+glyc., respectively) and the adducts to Asterix (x-Asterix), Sec61β (x-S61β) and Sec61α (x-S61α) are indicated. **e**, Photo-crosslinking analysis of ^{35}S-methionine-labelled Rho-4TMD translocation intermediate via BPA installed at TMD2 (position 85). Three types of semi-permeabilised cells were compared: wild type (WT), TMCO1 knockout (ΔTMCO1) and ΔTMCO1 cells transiently transfected with FLAG-tagged TMCO1. The major crosslink seen with WT SP cells is lost in ΔTMCO1 SP cells, where other weak crosslinks to similar-sized unidentified proteins are seen. When a subset of the ΔTMCO1 cells now express FLAG-TMCO1, a new crosslink is seen that migrates slightly slower than the major product seen in WT cells. These results verify that the major crosslink in WT cells is TMCO1 and the slightly larger crosslink in the reconstituted ΔTMCO1 cells is FLAG-TMCO1. Crosslinking with BPA at different positions in TMD2 show that its more hydrophilic face interacts with TMCO1 (summarised in the diagram). **f**, Protease protection assay of Rho-4TMD RNC in WT or ΔTMCO1 SP cells. The autoradiograph shows equal glycosylation indicative of equal efficiencies of TMD1 insertion, but different amounts of fully-protected product indicative of successful 3-TMD insertion. In ΔTMCO1 cells, a larger population of proteins failed to insert TMDs 2 and 3, leading to protease-accessibility (see diagram). PK digestion in the presence of detergent (subscripted d) leads to complete digestion. Graph shows quantification of three independent experiments showing the mean and standard deviation. **g**, RNCs of Rho^{ext} stalled 70 residues beyond TMD1 were assembled in WT, ΔAsterix, ΔCCDC47, ΔTMCO1 and ΔTMEM147 SP cells and affinity-purified via the FLAG-tag under native conditions as in Fig. 1b. Total cell lysates and purified RNCs were analysed by immunoblotting for the indicated proteins. **h**, Bismaleimidohexane (BMH) mediated crosslinking via cysteine in place of F56 in TMD1 of Rhodopsin stalled 70 residues downstream of TMD1 (top panel), or the equivalent RNC in which TMD1 is replaced with 22 leucine residues and a similarly positioned cysteine residue (bottom panel). ^{35}S-methionine-labelled RNCs were assembled with WT, ΔAsterix, ΔTMCO1, and ΔTMEM147 SP cells, subjected to BMH crosslinking where indicated, and either directly analysed by SDS-PAGE and autoradiography or after denaturing anti-Asterix IP. Equal glycosylation efficiency in all samples without BMH crosslinking indicate TMD1 insertion is unaffected by the loss of MPT components. Asterix crosslinking (marked as x-Asx) is unimpaired in ΔTMCO1 and ΔTMEM147 SP cells. In ΔAsterix cells, TMD1 of Rho crosslinks to an unidentified product that migrates slightly faster than Asterix (upward arrowhead). The 22L TMD does not crosslink to Asterix or the unidentified product, suggesting these interactions need partial TMD hydrophilicity.

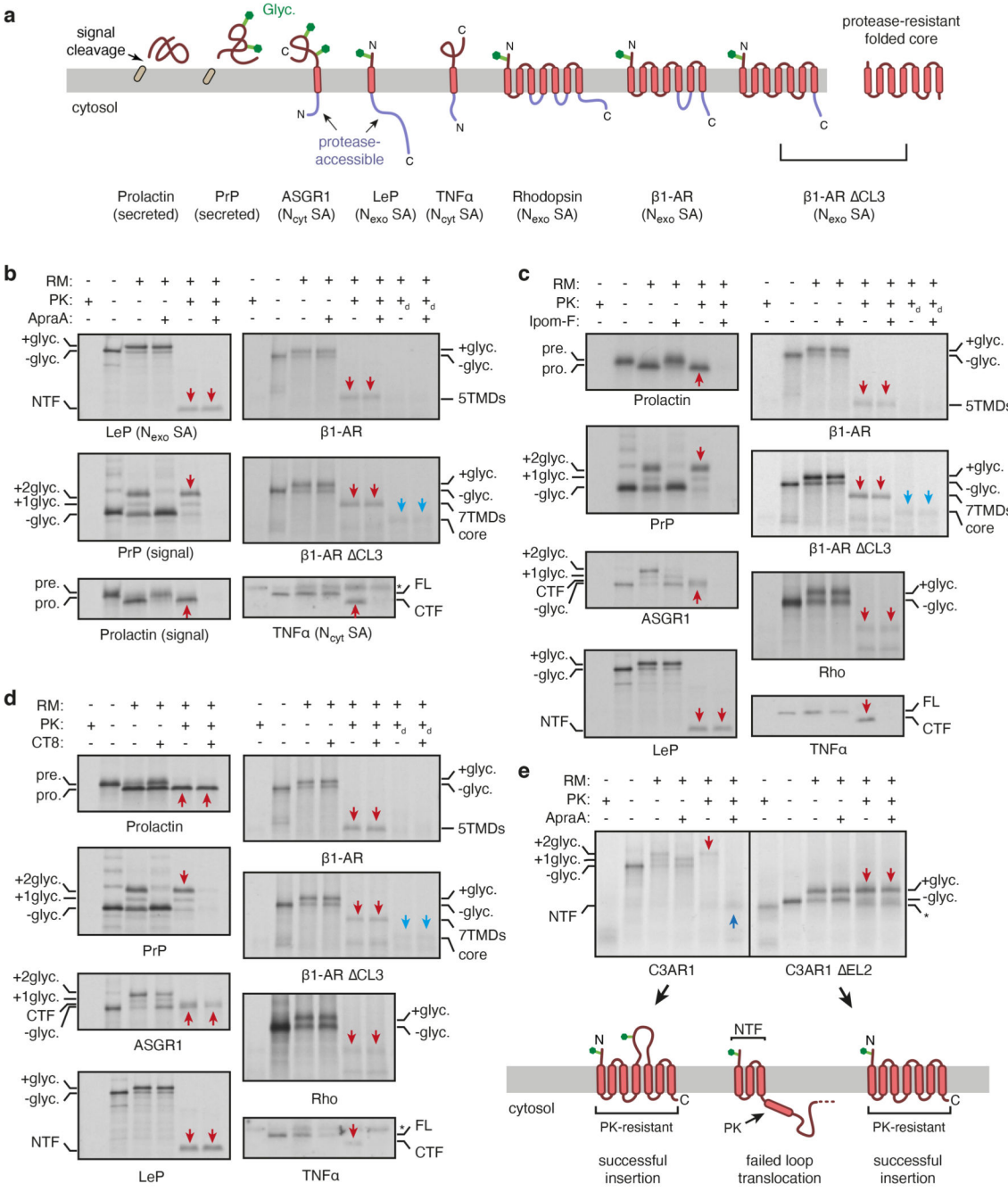


Extended Data Fig. 8. Features of the Rho-4TMD map and model.

a, Flowchart depicting the classification scheme and analysis of cryo-EM micrographs of the Rho-4TMD RNC sample. The classes denoted as "TRAP complex" appear to only contain the core translocon comprised of the Sec61 and TRAP complexes, evidently in somewhat different conformations, as characterised previously in cryo-EM and cryo-electron tomography work[26,61–63]. These were not pursued further. **b**, Fourier shell correlation (FSC) curve for the Rho-4TMD map illustrating an overall resolution of 3.88 Å by the gold-standard method[64]. **c**, Two views of the Rho-4TMD reference map coloured by

Europe PMC Funders Author Manuscripts

local resolution. Key structural elements, including the density assigned to TMD3 of the substrate, are indicated. **d**, Space-filling models showing the multipass translocon's highly conserved interior surface that faces the lipid-filled cavity (*left*) and poorly conserved exterior surface that faces the surrounding membrane (*right*). The models are coloured successively (from top to bottom) by protein, hydrophobicity, charge and conservation (calculated by ConSurf[65]). The conserved amphiphilic substrate-binding domain of the PAT complex, the conserved positively-charged hydrophilic vestibule of the GEL complex, and the conserved Sec61-docking site on the BOS complex are each indicated.



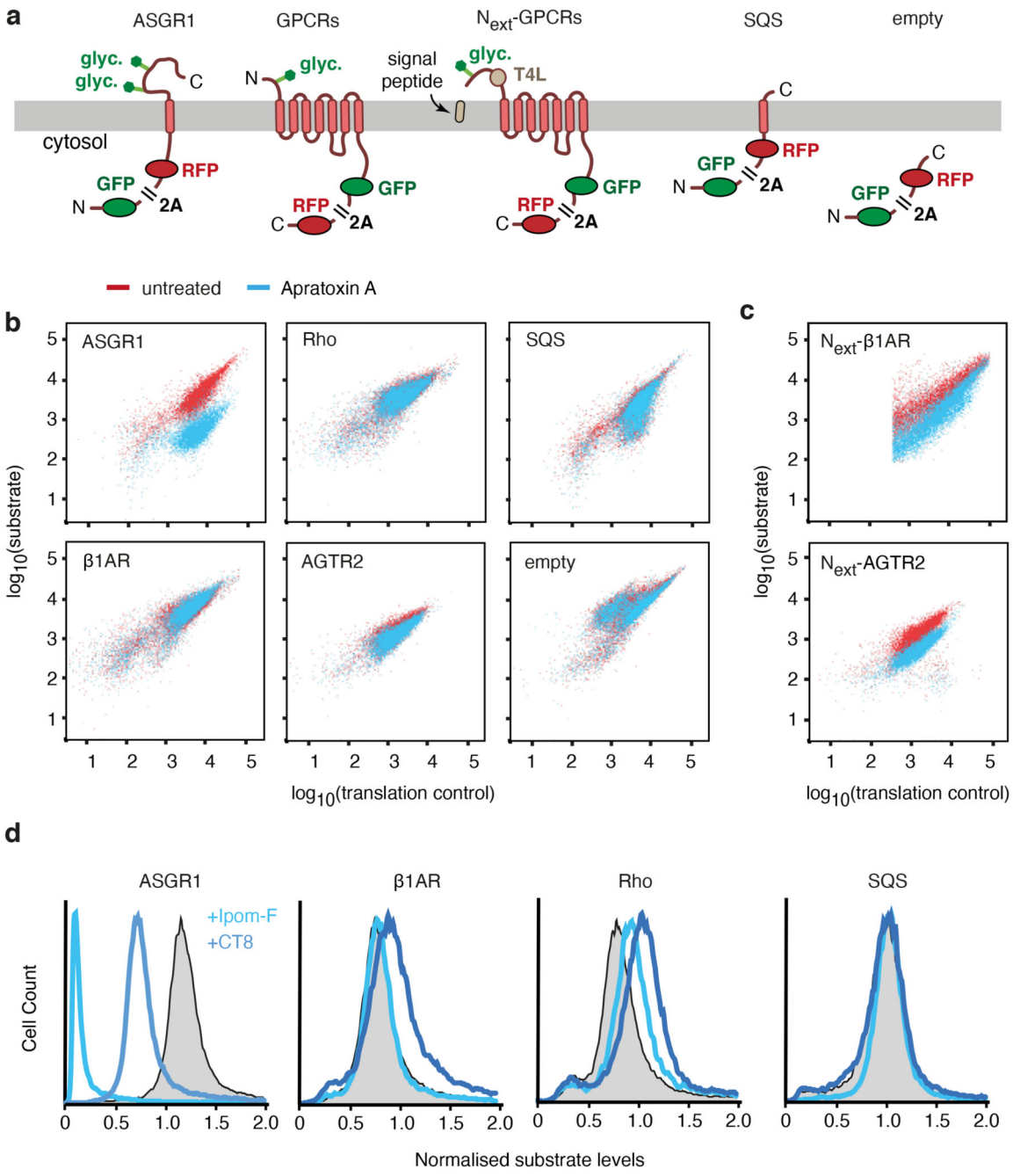
Extended Data Fig. 9. Effect of Sec61 inhibitors on various substrates *in vitro*.

a, Diagram of substrates tested for their sensitivity to Sec61 inhibitors. Cleavable signal sequences (tan), glycosylation sites (green), and experimentally determined sites of protease accessibility (lavender) are indicated. The core of β1ARΔCL3 that is resistant to protease digestion is also indicated. PrP is prion protein; LeP is leader peptidase; ASGR1 is asialoglycoprotein receptor 1; TNFα is tumor necrosis factor alpha. **b-d**, ^{35}S-methionine containing RRL translation reactions in the absence or presence of RMs and 2 μM of the indicated Sec61 inhibitor were analysed by protease protection. Panel b employed Apratoxin

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

A (ApraA, as in Fig. 4e), panel c employed Ipomeassin F (Ipom-F) and panel d used cotransin-8 (CT8). An equivalent volume of DMSO was used as a negative control. The translation products were either left untreated or digested with proteinase K (PK) without or with detergent (subscripted d) as indicated. The samples were analysed directly by SDS-PAGE and autoradiography or immunoblotting via a C-terminal FLAG tag (in the case of TNFa, which lacks methionines in the protease-protected fragment). Red arrows represent the fragments protected from PK digestion that are indicative of successful translocation. The protease-resistant folded core of β1ARΔCL3 is marked by blue arrows. NTF and CTF stand for N-terminal fragment and C-terminal fragment, respectively. Asterisk in the TNFa blot marks a background band that is sometimes detected by the FLAG antibody. **e**, Protease-protection analysis of C3AR1 (a GPCR) or a variant in which the ~180 amino acid long second extracellular loop (EL2) was shortened by deleting 61 residues (from T166 to D326). For simplicity, the majority of the C-terminal cytosolic domain was omitted from the constructs. Translation reactions of C3AR1 (left) and C3AR1ΔEL2 (right) were performed in the presence or absence of 2 μM ApraA and analyzed by protease protection. At the protease concentration used, all of the short cytosolic loops are resistant to digestion if the protein is inserted correctly, but become accessible if any insertion step fails (see diagram). Successfully inserted and PK-resistant products are indicated with red downward arrows. Failed insertion of TMD4 in the presence of ApraA has two consequences: the downstream loop is not glycosylated and the protein is protease-accessible downstream of TMD3. Thus, the doubly-glycosylated product is not observed and a protected N-terminal fragment (NTF, marked with dark blue upward arrow) representing the first three TMDs is generated upon PK digestion. Note that by these criteria, some failed insertion is seen even without ApraA. C3AR1ΔEL2 is unaffected by ApraA and the majority of it is correctly inserted. The asterisk marks the position of the C3AR1ΔEL2 fragment generated by PK digestion of the noninserted (and hence, non-glycosylated) population.



Extended Data Fig. 10. Effect of Sec61 inhibitors on various substrates in cells.

a, Diagram of fluorescent protein (FP)-tagged substrates tested for their sensitivity to Sec61 inhibitors in cells. 2A is the viral sequence at which peptide bond formation fails without disrupting elongation. Thus, each translation cycle generates two products: an FP-tagged substrate and a different coloured FP that serves as an internal control for translation levels. For RFP-tagged ASGR1 and SQS, GFP, which precedes the viral 2A sequence, serves as the translation control, while RFP fluorescence levels report on the substrate. The other substrates are GFP-tagged, with RFP serving as the translation control. **b**, Scatter plots of

Europe PMC Funders Author Manuscripts

the indicated reporters, each expressed from an inducible promoter in stable cell lines. Cells treated with 200 nM Apratoxin A (ApraA) during reporter induction are shown in blue, while cells treated with vehicle are shown in red. The histograms corresponding to some of these constructs are shown in Fig. 4f. **c**, Scatter plots of N-terminally extended (N_{ext}) β1AR and AGTR2 (see diagram in panel a) from an experiment similar to that shown in panel b. N_{ext}-AGTR2 is expressed stably whereas N_{ext}-b1-AR was expressed for 24 h by transient transfection with ApraA included during the final 12 hours. **d**, Histograms of normalised substrate levels after 6 hours of expression in the presence of vehicle (grey), 200 nM Ipomoeassin F (Ipom-F, light blue) or 1 μM CT8 (dark blue).

Extended Data Table 1

Cryo-EM data collection, refinement and validation statistics

	Rho-2TMD (EMDB-25994) (PDB 7TM3)	Rho-4TMD (EMDB-26133) (PDB 7TUT)
Data collection and processing		
Magnification	105,000	53,000
Voltage (kV)	300	300
Electron exposure (e-/Å^2)	54	54
Defocus range (μm)	-2.7 to -1.9	-2.7 to -1.9
Pixel size (Å)	1.34	1.39
Symmetry imposed	C1	C1
Initial particle images (no.)	1,665,551	1,445,986
Final particle images (no.)	148,833	136,812
Map resolution (Å)	3.25	3.88
FSC threshold	0.143	0.143
Refinement		
Model resolution (Å)	3.2	4.1
FSC threshold	0.5	0.5
Model composition		
Non-hydrogen atoms	149,944	152,599
Protein residues	8213	8542
Nucleotide bases	3895	3895
Ligands	Mg^{2+}: 220	Mg^{2+}: 220
	Zn^{2+}: 5	Zn^{2+}: 5
B factors (Å^2)		
Protein	183	205
Nucleotide	139	199
Ligand	104	161
R.m.s. deviations		
Bond lengths (Å)	0.004	0.004
Bond angles (°)	0.681	0.673
Validation		
MolProbity score	1.79	1.85

	Rho-2TMD (EMDB-25994) (PDB 7TM3)	Rho-4TMD (EMDB-26133) (PDB 7TUT)
Clashscore	6.88	7.36
Poor rotamers (%)	0.86	1.13
Ramachandran plot		
Favored (%)	93.9	94.0
Allowed (%)	6.0	5.8
Disallowed (%)	0.1	0.2

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank V. O. Paavilainen and K. McPhail for providing Apratoxin A, J. Taunton for providing CT8, W.Q. Shi for providing Ipomoeassin F, S.-Y. Peak-Chew for mass spectrometry, J. O'Donnell and V. Chandrasekaran for advice on structural modeling and figures, H. Wang for comments on the manuscript, and Hegde lab members for productive discussions. This work was supported by the UK Medical Research Council (grant MC_UP_A022_1007 to RSH), the MRC International PhD Programme (LS and AJOL), a National Research Foundation of Korea Fellowship (MKK), and the US National Institutes of Health (R01 GM130051 and R01 GM086487 to RJK).

Data availability

Data are available in the main article, supplementary materials, or public repositories (EMD-25994 and EMD-26133 of the EMDB; 7TM3 and 7TUT of the PDB).

References

1. von Heijne G. The membrane protein universe: what's out there and why bother? J Intern Med. 2007; 261: 543–57. [PubMed: 17547710]
2. Hegde RS, Keenan RJ. The mechanisms of integral membrane protein biogenesis. Nature Reviews Molecular Cell Biology. 2022; 23: 107–124. [PubMed: 34556847]
3. Blobel G. Intracellular protein topogenesis. Proc Natl Acad Sci U S A. 1980; 77: 1496–500. [PubMed: 6929499]
4. Heinrich SU, Mothes W, Brunner J, Rapoport TA. The Sec61p complex mediates the integration of a membrane protein by allowing lipid partitioning of the transmembrane domain. Cell. 2000; 102: 233–44. [PubMed: 10943843]
5. Görlich D, Rapoport TA. Protein translocation into proteoliposomes reconstituted from purified components of the endoplasmic reticulum membrane. Cell. 1993; 75: 615–30. [PubMed: 8242738]
6. Martoglio B, Hofmann MW, Brunner J, Dobberstein B. The protein-conducting channel in the membrane of the endoplasmic reticulum is open laterally toward the lipid bilayer. Cell. 1995; 81: 207–14. [PubMed: 7736572]
7. High S, et al. Sec61p is adjacent to nascent type I and type II signal-anchor proteins during their membrane insertion. J Cell Biol. 1993; 121: 743–50. [PubMed: 8491769]
8. Rapoport TA, Li L, Park E. Structural and Mechanistic Insights into Protein Translocation. Annu Rev Cell Dev Biol. 2017; 33: 369–390. [PubMed: 28564553]
9. Van den Berg B, et al. X-ray structure of a protein-conducting channel. Nature. 2004; 427: 36–44. [PubMed: 14661030]
10. Chitwood PJ, Hegde RS. An intramembrane chaperone complex facilitates membrane protein biogenesis. Nature. 2020; 584: 630–634. [PubMed: 32814900]

11. McGilvray PT, et al. An ER translocon for multi-pass membrane protein biogenesis. Elife. 2020; 9 e56889 [PubMed: 32820719]
12. Hegde RS, Lingappa VR. Membrane Protein Biogenesis : Endoplasmic Reticulum. Cell. 1997; 91: 575–582. [PubMed: 9393851]
13. Shao S, Hegde RS. Membrane protein insertion at the endoplasmic reticulum. Annu Rev Cell Dev Biol. 2011; 27: 25–56. [PubMed: 21801011]
14. Chitwood PJ, Juszkiewicz S, Guna A, Shao S, Hegde RS. EMC Is Required to Initiate Accurate Membrane Protein Topogenesis. Cell. 2018; 175: 1507–1519. e16 [PubMed: 30415835]
15. O'Keefe S, et al. An alternative pathway for membrane protein biogenesis at the endoplasmic reticulum. Commun Biol. 2021; 4: 828. [PubMed: 34211117]
16. Anghel SA, McGilvray PT, Hegde RS, Keenan RJ. Identification of Oxa1 Homologs Operating in the Eukaryotic Endoplasmic Reticulum. Cell Rep. 2017; 21: 3708–3716. [PubMed: 29281821]
17. Guna A, Volkmar N, Christianson JC, Hegde RS. The ER membrane protein complex is a transmembrane domain insertase. Science. 2018; 359: 470–473. [PubMed: 29242231]
18. Do H, Falcone D, Lin J, Andrews DW, Johnson AE. The cotranslational integration of membrane proteins into the phospholipid bilayer is a multistep process. Cell. 1996; 85: 369–78. [PubMed: 8616892]
19. Mothes W, et al. Molecular mechanism of membrane protein integration into the endoplasmic reticulum. Cell. 1997; 89: 523–533. [PubMed: 9160744]
20. Mravic M, et al. Packing of apolar side chains enables accurate design of highly stable membrane proteins. Science. 2019; 363: 1418–1423. [PubMed: 30923216]
21. Cymer F, von Heijne G, White SH. Mechanisms of integral membrane protein insertion and folding. J Mol Biol. 2015; 427: 999–1022. [PubMed: 25277655]
22. Talbot BE, Vandorpe DH, Stotter BR, Alper SL, Schlondorff JS. Transmembrane insertases and N-glycosylation critically determine synthesis, trafficking, and activity of the nonselective cation channel TRPC6. J Biol Chem. 2019; 294: 12655–12669. [PubMed: 31266804]
23. Meacock SL, Lecomte FJL, Crawshaw SG, High S. Different Transmembrane Domains Associate with Distinct Endoplasmic Reticulum Components during Membrane Integration of a Polytopic Protein. Mol Biol Cell. 2002; 13: 4114–4129. [PubMed: 12475939]
24. Ismail N, Crawshaw SG, High S. Active and passive displacement of transmembrane domains both occur during opsin biogenesis at the Sec61 translocon. J Cell Sci. 2006; 119: 2826–36. [PubMed: 16787949]
25. Sundaram A, et al. Substrate-driven assembly of a modular translocon for multipass membrane protein topogenesis. Submitted. 2022.
26. Pfeffer S, et al. Dissecting the molecular organization of the translocon-associated protein complex. Nat Commun. 2017; 8
27. Dettmer U, et al. Transmembrane protein 147 (TMEM147) is a novel component of the Nicalin-NOMO protein complex. J Biol Chem. 2010; 285: 26174–26181. [PubMed: 20538592]
28. Lewis AJO, Hegde RS. A unified evolutionary origin for the ubiquitous protein transporters SecY and YidC. BMC Biol. 2021; 19
29. Mariappan M, et al. The mechanism of membrane-associated steps in tail-anchored protein insertion. Nature. 2011; 477: 61–6. [PubMed: 21866104]
30. Jumper J, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021; 596: 583–589. [PubMed: 34265844]
31. Tunyasuvunakool K, et al. Highly accurate protein structure prediction for the human proteome. Nature. 2021; 596: 590–596. [PubMed: 34293799]
32. Voorhees RM, Fernández IS, Scheres SHW, Hegde RS. Structure of the Mammalian Ribosome-Sec61 Complex to 3.4 Å Resolution. Cell. 2014; 157: 1632–1643. [PubMed: 24930395]
33. Plath K, Mothes W, Wilkinson BM, Stirling CJ, Rapoport TA. Signal sequence recognition in posttranslational protein transport across the yeast ER membrane. Cell. 1998; 94: 795–807. [PubMed: 9753326]
34. Voorhees RM, Hegde RS. Structure of the Sec61 channel opened by a signal sequence. Science. 2016; 351: 88–91. [PubMed: 26721998]

Europe PMC Funders Author Manuscripts

35. Li L, et al. Crystal structure of a substrate-engaged SecY protein-translocation channel. Nature. 2016; 531: 395–399. [PubMed: 26950603]
36. Keenan RJ, Freymann DM, Walter P, Stroud RM. Crystal structure of the signal sequence binding subunit of the signal recognition particle. Cell. 1998; 94: 181–91. [PubMed: 9695947]
37. Mateja A, et al. The structural basis of tail-anchored membrane protein recognition by Get3. Nature. 2009; 461: 361–6. [PubMed: 19675567]
38. Fry MY, Saladi SM, Clemons WM. The STI1-domain is a flexible alpha-helical fold with a hydrophobic groove. Protein Sci. 2021; 30: 882–898. [PubMed: 33620121]
39. Wang F, Brown EC, Mak G, Zhuang J, Denic V. A chaperone cascade sorts proteins for posttranslational membrane insertion into the endoplasmic reticulum. Mol Cell. 2010; 40: 159–71. [PubMed: 20850366]
40. Shao S, Hegde RS. A calmodulin-dependent translocation pathway for small secretory proteins. Cell. 2011; 147: 1576–88. [PubMed: 22196732]
41. Itakura E, et al. Ubiquilins Chaperone and Triage Mitochondrial Membrane Proteins for Degradation. Mol Cell. 2016; 63: 21–33. [PubMed: 27345149]
42. Perara E, Lingappa VR. A former amino terminal signal sequence engineered to an internal location directs translocation of both flanking protein domains. J Cell Biol. 1985; 101: 2292–2301. [PubMed: 3864782]
43. Kalies KU, Görlich D, Rapoport TA. Binding of ribosomes to the rough endoplasmic reticulum mediated by the Sec61p-complex. J Cell Biol. 1994; 126: 925–34. [PubMed: 8051212]
44. Hennon SW, Soman R, Zhu L, Dalbey RE. YidC/Alb3/Oxa1 family of insertases. J Biol Chem. 2015; 290: 14866–14874. [PubMed: 25947384]
45. UniProt Consortium, T. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2018; 46: 2699. [PubMed: 29425356]
46. von Heijne G. Membrane-protein topology. Nature Reviews Molecular Cell Biology. 2006; 7: 909–918. [PubMed: 17139331]
47. Forrest LR. Structural Symmetry in Membrane Proteins. Annu Rev Biophys. 2015; 44: 311–337. [PubMed: 26098517]
48. Maifeld SV, et al. Secretory Protein Profiling Reveals TNF-α Inactivation by Selective and Promiscuous Sec61 Modulators. Chem Biol. 2011; 18: 1082–1088. [PubMed: 21944747]
49. Paatero AO, et al. Apratoxin Kills Cells by Direct Blockade of the Sec61 Protein Translocation Channel. Cell Chem Biol. 2016; 23: 561–566. [PubMed: 27203376]
50. Zong G, et al. Ipomoeassin F Binds Sec61α to Inhibit Protein Translocation. J Am Chem Soc. 2019; 141: 8450–8461. [PubMed: 31059257]
51. Feng Q, Shao S. In vitro reconstitution of translational arrest pathways. Methods. 2018; 137: 20–36. [PubMed: 29277545]
52. Sharma A, Mariappan M, Appathurai S, Hegde RS. In vitro dissection of protein translocation into the mammalian endoplasmic reticulum. Methods Mol Biol. 2010; 619: 339–63. [PubMed: 20419420]
53. Walter P, Blobel G. Preparation of microsomal membranes for cotranslational protein translocation. Methods Enzymol. 1983; 96: 84–93. [PubMed: 6656655]
54. Lin Z, et al. TTC5 mediates autoregulation of tubulin via mRNA degradation. Science. 2020; 367: 100–104. [PubMed: 31727855]
55. Fons RD, Bogert BA, Hegde RS. Substrate-specific function of the translocon-associated protein complex during translocation across the ER membrane. J Cell Biol. 2003; 160: 529–39. [PubMed: 12578908]
56. Mirdita M, et al. ColabFold - Making protein folding accessible to all. bioRxiv. 2021; 2021.08.15.456425 doi: 10.1101/2021.08.15.456425
57. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010; 66: 486–501. [PubMed: 20383002]
58. Afonine PV, et al. Real-space refinement in PHENIX for cryo-EM and crystallography. Acta Crystallogr Sect D, Struct Biol. 2018; 74: 531–544. [PubMed: 29872004]

59. Liebschner D, et al. Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. Acta Crystallogr Sect D Struct Biol. 2019; 75: 861–877. [PubMed: 31588918]

60. Pettersen EF, et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. 2021; 30: 70–82. [PubMed: 32881101]

61. Ménétret JF, et al. Single Copies of Sec61 and TRAP Associate with a Nontranslating Mammalian Ribosome. Structure. 2008; 16: 1126–1137. [PubMed: 18611385]

62. Ménétret JF, et al. Architecture of the ribosome-channel complex derived from native membranes. J Mol Biol. 2005; 348: 445–457. [PubMed: 15811380]

63. Pfeffer S, et al. Structure of the mammalian oligosaccharyl-transferase complex in the native ER protein translocon. Nat Commun. 2014; 5

64. Henderson R, et al. Outcome of the first electron microscopy validation task force meeting. Structure. 2012; 20: 205–214. [PubMed: 22325770]

65. Ashkenazy H, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. 2016; 44: W344–50. [PubMed: 27166375]

Europe PMC Funders Author Manuscripts

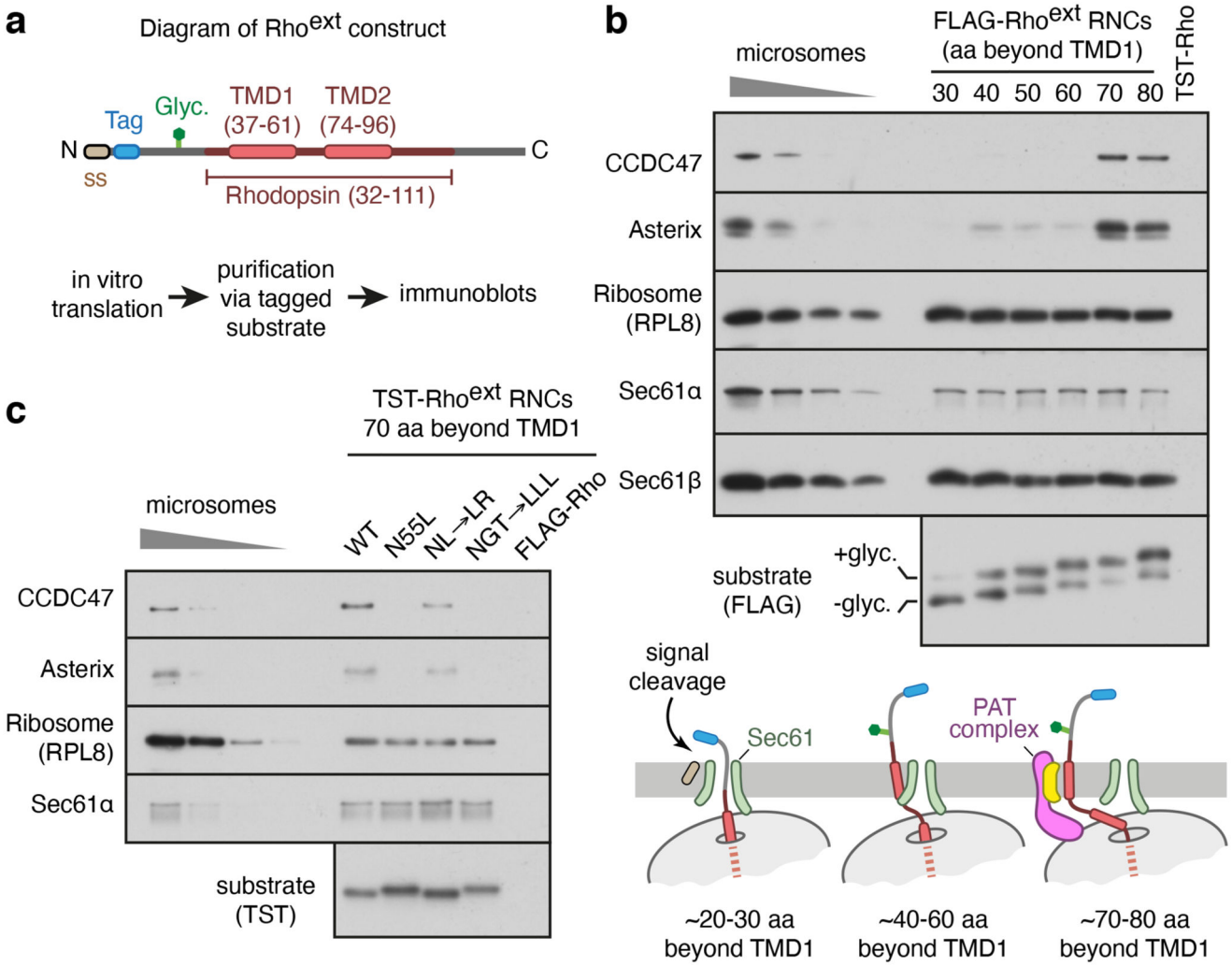


Fig. 1. Substrate-mediated recruitment of the PAT complex to Sec61.

a, Diagram of the Rho^{ext} construct (not to scale) containing the first two TMDs and flanking regions of bovine rhodopsin (amino acids 32-111, Uniprot ID P02699). The Rho region is preceded by a signal sequence (ss) and a polypeptide of 52 amino acids containing a site for N-linked glycosylation (Glyc.) to monitor translocation. The experimental strategy to analyse intermediates of Rho^{ext} is indicated. **b**, Ribosome nascent chain complexes (RNCs) of FLAG-tagged Rho^{ext} truncated at the indicated lengths (in amino acids, aa) beyond TMD1 were assembled by *in vitro* translation in rabbit reticulocyte lysate (RRL) containing canine pancreas rough microsomes (RMs). The RNCs were affinity-purified via the FLAG tag and analysed by immunoblotting relative to serial two-fold dilutions of RMs. Rho^{ext} tagged with the twin-Strep tag (TST) and truncated 70 amino acids beyond TMD1 served as a specificity control for anti-FLAG affinity purification. The diagram below (not to scale) depicts key stages of biogenesis as deduced from the results. **c**, Affinity purification of TST-tagged Rho^{ext} RNCs similar to panel B for constructs in which TMD1 contains mutations: NL→LR stands for N55L and L47R mutations; NGT→LLL changes N55, G51, T58 to leucines. The FLAG-tagged wt Rho^{ext} served as a specificity control for the purifications via

the TST. The substrates in this experiment do not contain a glycosylation site. Source data for all gels in this and other figures can be found in Supplementary Fig. 1.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

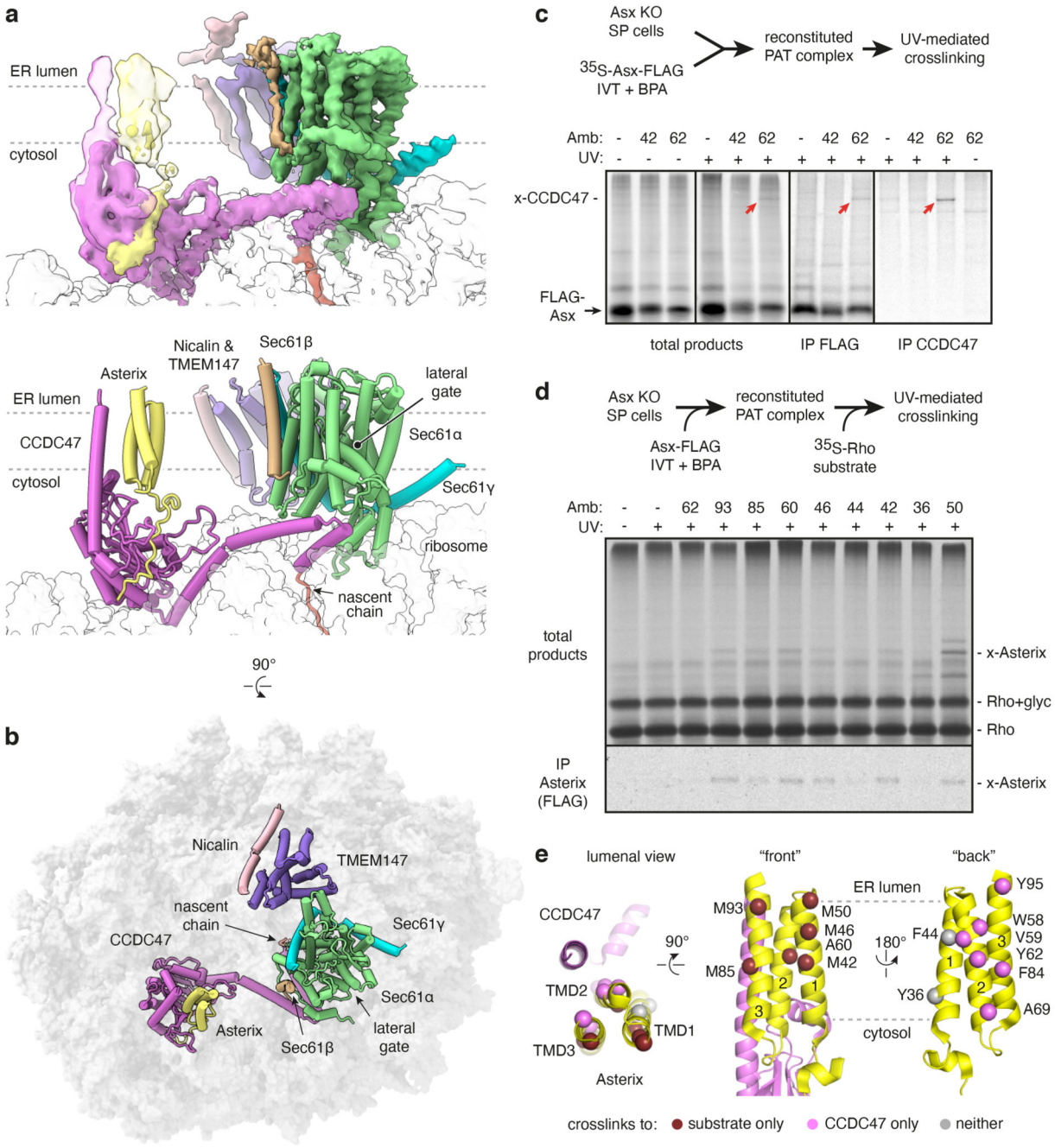


Fig. 2. Structure of the PAT complex recruited to Sec61.

a, Composite map (top) and model (bottom) of the Rho-2TMD translocation intermediate stalled 70 amino acids beyond TMD1 (see Fig. 1b). The density for the PAT and BOS complexes derives from submaps obtained after focussed classification on each complex (see Extended Data Fig. 2). The PAT complex is shown at two contour levels to visualise the lower-resolution membrane domain simultaneously with the higher resolution ribosome-proximal regions. The GEL and TRAP complexes, visible only at very low contour levels, were not modelled. The detergent micelle has been removed for clarity. **b**, View

from the ER lumen of the Rho-2TMD translocation intermediate model. **c**, Site-specific photo-crosslinking from benzoyl-phenylalanine (BPA) installed in Asterix by amber (Amb) suppression at residue 42 or 62. Semi-permeabilised (SP) Asterix knockout (KO) cells were reconstituted with ^{35}S-methionine labelled Asterix-FLAG variants (see Extended Data Fig. 5d), irradiated with UV, and analysed directly (total products) or after immunoprecipitation (IP). The red arrow indicates the Asterix-CCDC47 adduct. **d**, Analysis of substrate-Asterix photo-crosslinking from BPA installed at amber codons (Amb) at the indicated positions in Asterix. Asterix KO cells are reconstituted with non-radioactive Asterix-FLAG variants as in panel c, then incubated with ^{35}S-methionine labelled RNCs of Rho-2TMD. After UV irradiation, the samples were analysed directly or after IP. Non-glycosylated and glycosylated substrate (Rho and Rho+glyc., respectively), and the adduct to Asterix (x-Asterix) are indicated. **e,** Summary of all site-specific photo-crosslinking results (see Fig. 2d and Extended Data Fig. 5d-e) mapped onto the PAT complex structure shown in three different views.

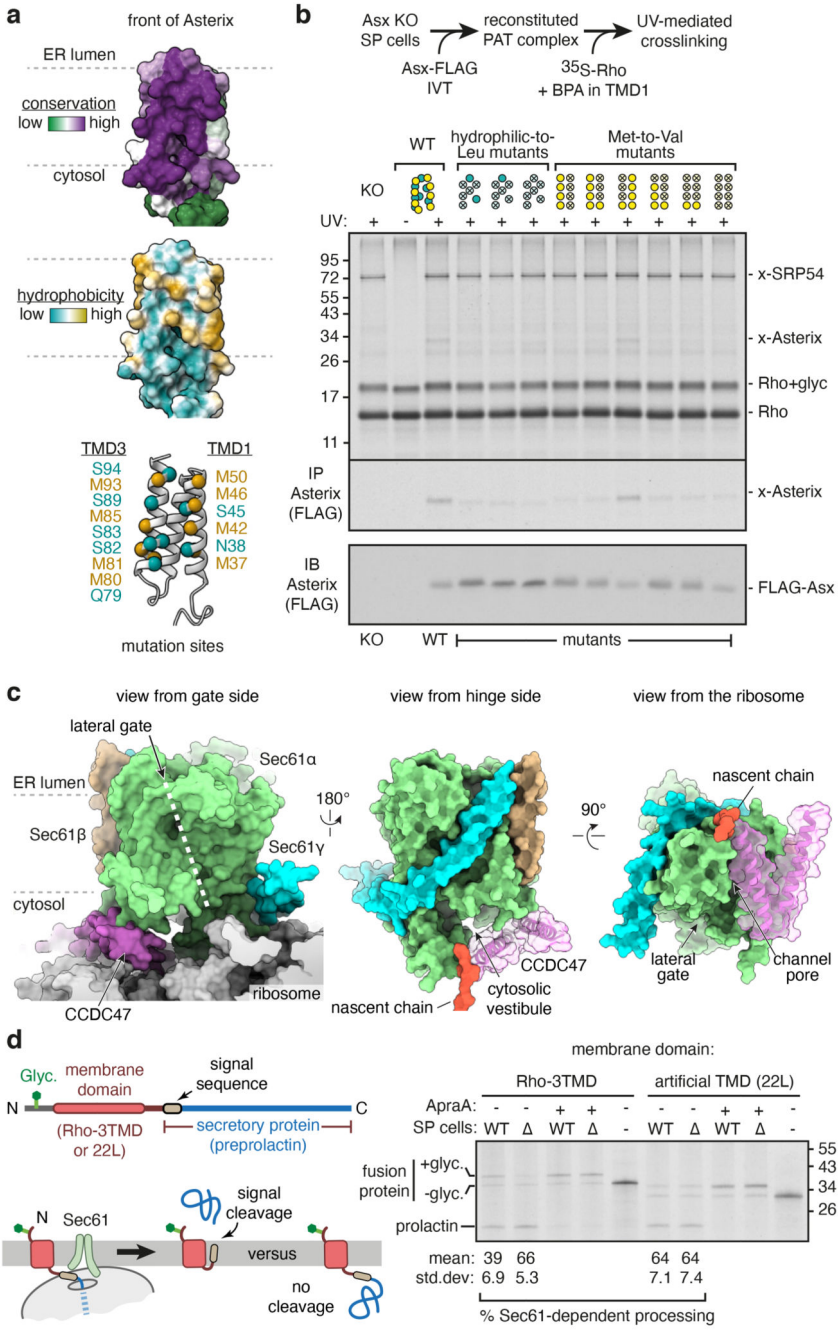


Fig. 3. The PAT complex binds the substrate and redirects it to the back of Sec61.

a, The substrate-binding surface of Asterix showing conservation (top), hydrophobicity (middle) and positions subject to mutagenesis (bottom). **b**, Photo-crosslinking from a BPA probe in place of F56 in TMD1 of the Rho-2TMD substrate. Semi-permeabilised (SP) Asterix KO cells were reconstituted with either wild-type or mutant Asterix-FLAG variants and tested for photo-crosslinking with ^{35}S-methionine labelled Rho-2TMD RNCs similar to Fig. 2d. The adducts to Asterix (x-Asterix) and SRP54 (x-SRP54, from the non-translocated population of RNCs) are indicated. The bottom panel shows an immunoblot for Asterix

to verify comparable levels of reconstitution for each mutant. The diagram above the gel indicates which positions were mutated (indicated by an 'X'), using the positional scheme in the bottom diagram of panel a. Hydrophilic (teal) and methionine (yellow) residues were mutated to leucine and valine, respectively. **c**, Three views of the Rho-2TMD structure showing that CCDC47's latch helices are wedged between the ribosome and a closed Sec61. **d**, Experimental strategy (left) to measure hydrophobic domain access to the Sec61 lateral gate downstream of a membrane domain. Either the first three TMDs of Rho (Rho-3TMD) or an artificial poly-leucine TMD (22L), is fused ahead of preprolactin. If the preprolactin signal sequence successfully engages Sec61, the glycosylated fusion protein is cleaved; otherwise, the glycosylated fusion protein remains intact. ^{35}S-methionine labelled fusion proteins were translated with wild type (WT) or CCDC47 KO (Δ) semi-permeabilised cells in the absence or presence of 2 μM Apratoxin A (ApraA, a Sec61 inhibitor). The pelleted cells were analysed by SDS-PAGE and autoradiography. Glycosylated (+glyc) and non-glycosylated (-glyc) fusion protein and the cleaved prolactin domain are marked. Quantification from three experiments (mean ± standard deviation) is shown below the gel.

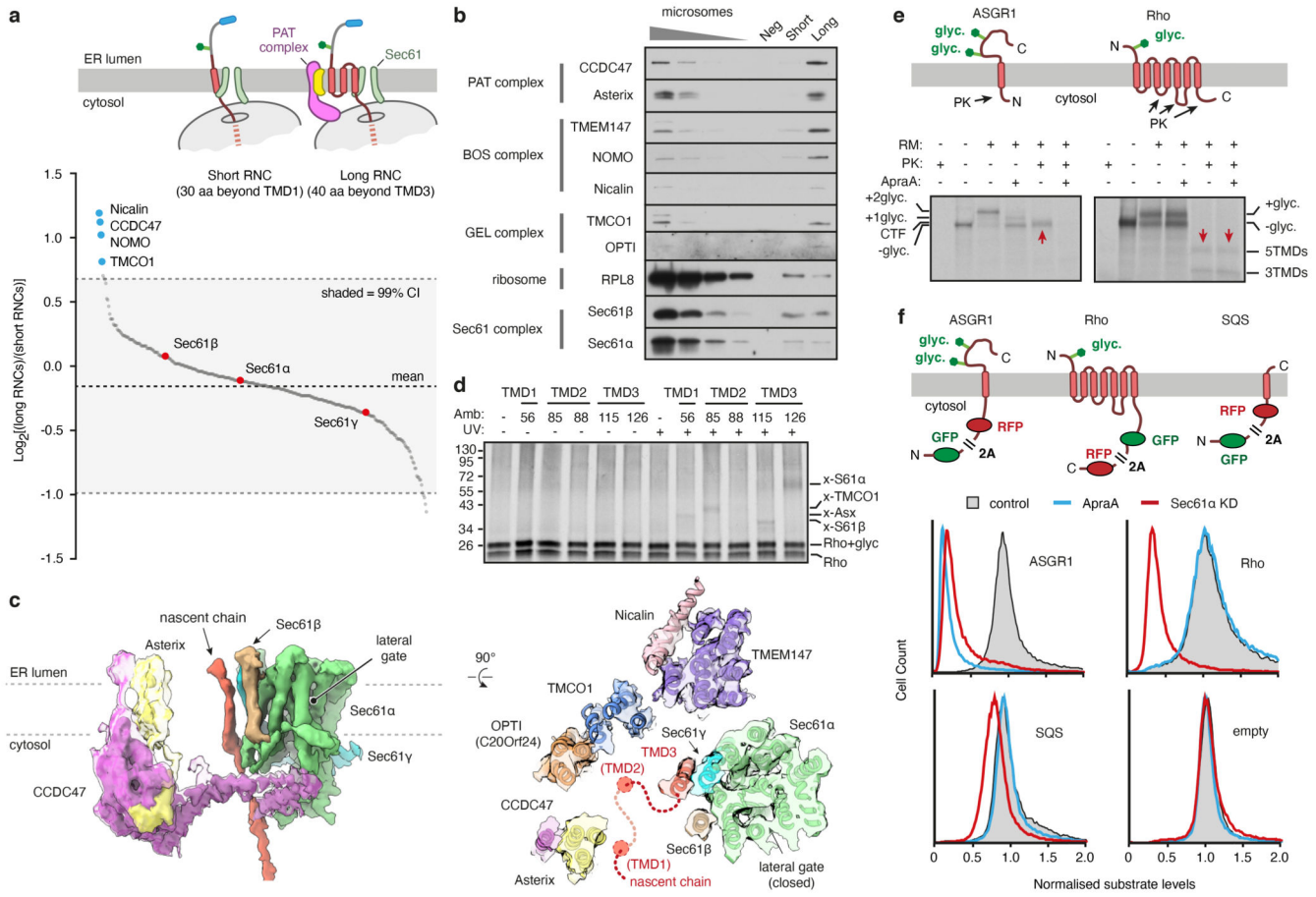


Fig. 4. Insertion of multipass proteins independent of Sec61 activity.

a, Tandem-mass-tag (TMT) mass spectrometry of proteins copurified with early and late Rho^{ext} insertion intermediates (see diagram). The top four proteins differentially recovered with the late intermediate are indicated. The raw data are in Supplementary Table 1. **b**, Samples prepared as in panel a and a negative control lacking substrate were analysed by immunoblotting for the indicated proteins. **c**, Cryo-EM map (top) and map overlaid with model (bottom) of the Rho-4TMD translocation intermediate (stalled 40 amino acids beyond TMD3) containing the PAT, BOS, GEL and Sec61 complexes. The PAT complex is shown at two contour levels. The detergent micelle was removed for clarity. The position of the substrate is based on crosslinking data showing that TMD1, TMD2, and TMD3 are adjacent to Asterix, TMCO1 and the closed Sec61 complex, respectively. **d**, Photo-crosslinking of ^{35}S-methionine-labelled Rho-4TMD translocation intermediates via BPA at the indicated positions of each TMD. Non-glycosylated and glycosylated substrate (Rho and Rho+glyc., respectively) and the adducts to Asterix (x-Asterix), TMCO1 (x-TMCO1), Sec61β (x-S61β) and Sec61α (x-S61α), verified in Extended Data Fig. 7d, and 7e, are indicated. **e**, In vitro translation reactions of ASGR1 and Rho in the absence or presence of RMs and 2 μM Apratoxin A were analysed by protease protection. The translation products were digested with proteinase K (PK) without or with detergent (subscripted d) where indicated. Red arrows represent PK-protected fragments indicative of successful translocation. **f**, Cell lines containing the indicated inducible fluorescent protein reporters (see Extended Data Fig. 10)

Europe PMC Funders Author Manuscripts

were treated with control or Sec61A1-targeting siRNAs. Reporter expression was induced in the presence or absence of 200 nM Apratoxin A and the cells were analysed by flow cytometry. Histograms of normalised reporter expression are shown. SQS is a tail-anchored protein inserted by EMC.

Europe PMC Funders Author Manuscripts