Membrane protein sequence features direct post-translational insertion

Ilya A. Kalinin[1]*, Hadas Peled-Zehavi[1]*, Alon B.D. Barshap[1], Shai A. Tamari[1], Yarden Weiss[1], Reinat Nevo[1], Nir Fluman[1#]

[1] Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot, Israel.

* authors contributed equally to this work

Adress correspondence to: nir.fluman@weizmann.ac.il

Abstract

The proper folding of multispanning membrane proteins (MPs) hinges on the accurate insertion of their transmembrane helices (TMs) into the membrane. Predominantly, TMs are inserted during protein translation, via a conserved mechanism centered around the Sec translocon. Our study reveals that the C-terminal TMs (cTMs) of numerous MPs across various organisms bypass this cotranslational route, necessitating an alternative posttranslational insertion strategy. We demonstrate that evolution has refined the hydrophilicity and length of these proteins' C-terminal tails to optimize cTM insertion. Alterations in the C-tail sequence disrupt cTM insertion in both *E. coli* and human, leading to protein defects, loss of function, and genetic diseases. In *E. coli*, we identify YidC, a member of the widespread Oxa1 family, as the insertase facilitating cTMs insertion, with C-tail mutations disrupting the productive interaction of cTMs with YidC. Thus, MP sequences are fine-tuned for effective collaboration with the cellular biogenesis machinery, ensuring proper membrane protein folding.

Main

Membrane proteins (MPs) perform a variety of essential cellular functions and constitute at least a quarter of every proteome[1]. The largest group of MPs, called multispanning MPs, consists of proteins with multiple transmembrane helices (TMs) that span the membrane. A crucial step in the biogenesis and folding of these proteins is the correct insertion of their TMs into the membrane[2]. Inaccurate TM insertion can lead to protein misfolding, loss of function, and in some cases, disease[3].

The insertion of the TMs of multispanning MPs typically occurs co-translationally at the endoplasmic reticulum (ER) membrane in eukaryotes and plasma membrane in prokaryotes and is centered around the universally conserved Sec translocon[4]. In bacteria, the translating ribosome docks on the SecYEG translocon, which inserts the TMs as they emerge out of the ribosome[5,6]. The Oxa1 family insertase YidC and the auxiliary SecDF-YajC complex can cooperate with the SecYEG translocon to facilitate insertion[7,8]. The insertion of multipass proteins in eukaryotes is more complex, and was recently shown to involve several additional complexes in the vicinity of Sec61, the eukaryotic SecYEG homolog[9-11]. However, the fundamental principles of cotranslational insertion are conserved across organisms. Importantly, when each TM finishes its synthesis, it still resides inside the ribosome exit tunnel, inaccessible to the insertion machinery. Geometric considerations, along with experimental studies suggest that the translocon inserts TM only once an additional stretch of about 45 amino acids has been synthesized C-terminally to the TM (Fig.

1a)[12,13]. This immediately suggests a hurdle – if a TM is too close to the C-terminus and not followed by at least 45 amino acids, it will not be able to insert co-translationally. Indeed, recent studies have highlighted that post-translational insertion of C-terminal TMs (cTMs) can be inefficient, posing a challenge to biogenesis in yeast and vertebrates[14,15]. Extending a protein's C-tail, either by evolution[14] or artificially[14,15], can allow co-translational insertion and improve biogenesis. However, the C-tails of MPs are often short, necessitating post-translational insertion.

The challenging insertion of cTMs has received much attention in the case of proteins that harbor a *single* TM close to the C-terminus. Such proteins, called Tail-Anchored proteins, complete their translation in the cytosol and must be chaperoned, targeted, and inserted into the membrane post-translationally[16,17]. Several cellular pathways mediate Tail-Anchored protein biogenesis, all converging at post-translational insertases distinct from the Sec translocon. Insertases of Tail-Anchored proteins identified to date include the ER membrane protein complex (EMC) and the GET complex in the ER membrane, Mtch2 and mim1 in the mitochondrial membrane, and likely YidC in bacteria[18-23]. However, in contrast to the level of detail we have for Tail-Anchored proteins, we know relatively little about the insertion of cTMs in multispanning proteins. In *S. cerevisiae*, the Sec62-Sec63 complex was suggested to play a role in cTM insertion, but this complex is absent from bacteria[24].

Multispanning MPs occupy a vastly different state at the end of translation compared to Tail-Anchored proteins, presenting a somewhat distinct challenge. These proteins are efficiently targeted and anchored to the membrane by their more N-terminal TMs, which insert cotranslationally. Moreover, until translation ends, they are stably docked to the translocon by the ribosome-translocon junction. Their N-terminal, inserted segments may interact with the translocon and its auxiliary subunits. At the moment translation ends, the detachment of the translating proteins from the ribosome may allow them to diffuse away from the translocon. Despite this freedom, studies suggest that multispanning proteins might dwell in the translocon vicinity even after translation[15,25]. However, how common this phenomenon is, and whether the Sec translocon plays a direct role in cTM insertion was not clear.

Here we set out to elucidate the mechanism of posttranslational insertion of cTMs. We show that the sequences around cTMs, and particularly the hydrophilicity of the C-terminal tails, are constrained by evolution across organisms to optimize insertion. We reveal that for *E. coli* proteins having an extracytosolic C-terminus, a C-tail of low hydrophilicity allows the cTMs to get inserted by the YidC insertase. Increasing the hydrophilicity of the C-terminal tail impairs cTM insertion, leading to protein misfolding and loss of function. A study published during the preparation of this manuscript revealed that cTM insertion in human cells is mediated by EMC, a distant homolog of YidC[26], suggesting that the mechanism of cTM insertion is ancient and conserved. In accordance, we identify disease-causing mutations in human MPs that increase C-tail hydrophilicity, causing misinsertion and mistrafficking consistent with misfolding. Our results indicate that cTMs are inserted by a mechanism distinct from the Sec translocon, and suggest that membrane protein sequences have evolved to optimize their biogenesis through the cellular insertion machinery.

Results

Membrane protein C-termini display distinct sequence constraints.

Recent studies suggest that C-terminal TMs (cTMs) can insert into the membrane either co- or post-translationally. The main factor determining the route taken by a protein is the length of the C-terminal tail (C-tail) following the cTM. C-tails shorter than ~45 residues will likely require posttranslational insertion, whereas longer C-tails may allow cotranslational insertion and were shown to improve biogenesis efficiency in some cases[12,13,15,27].

We first asked how many multispanning MPs depend on each route across organisms. To this end, we predicted the topologies of the membrane proteomes of *E. coli*, yeast (*S. cerevisiae*), and human. To improve the prediction accuracy, we combined data from several topology and structure prediction methods (see methods). We then focused on translocon-inserted proteins, by excluding proteins with TMs fewer than 3, and excluding mitochondrial and peroxisomal proteins from the eukaryotic datasets. The resulting dataset contains 673, 785 and 2483 multispanning proteins from *E. coli*, yeast and human, respectively. The distribution of C-tail lengths among these proteins indicates that many membrane proteins have C-tail shorter than 45 residues (Fig. 1b). While in *E. coli* these are the majority (567 proteins), in humans, they are roughly a half (1385 proteins). Thus, post-translational insertion of the cTM is a pervasive necessity during biogenesis.

Since cTM insertion might be challenging, we hypothesized that cTMs may have evolved sequence adaptations to optimize their post-translational insertion. We first focused on the hydrophobicity of the cTM, since hydrophobicity is a major driving force for insertion. The hydrophobicity of posttranslationally inserting cTMs was only slightly different from other TMs. Furthermore, the differences were not consistent across organisms, with cTMs having slightly increased hydrophobicity than other TMs in *E. coli* and decreased hydrophobicity in human (Fig, S1a). Similar, slight and organism-specific differences were observed when comparing cTMs that require post- or co-translational insertion (Fig. S1b). Thus, we see no evidence for evolution acting on cTM hydrophobicity to optimize insertion.

To explore other sequence adaptations, we focused on another main barrier for insertion – the translocation of the TM's flanking loop/tail across the membrane. The Sec translocon enables the translocation of practically any loop including large polypeptide stretches. However, other insertases struggle with translocating very hydrophilic or long polypeptides[2,28]. Depending on the cTM orientation, the translocated loop may precede or follow the TM. In multispanning proteins that possess a cytosolic C-terminus (C_{cyt}), the C-tail remains in the cytosol and the loop preceding the cTM will have to get translocated (Fig. 1c). In contrast, in proteins having an extracytosolic, C_{ext} topology, the C-tail gets translocated (Fig. 1d). We focused on proteins that require posttranslational insertion, possessing C-tails shorter than 45 residues. Our analysis revealed significant differences in the hydrophilicity of their TM-flanking regions, especially in the C-tail following the cTM. In C_{cyt} proteins, the C-tails tend to be longer and more hydrophilic, while in C_{ext} proteins, C-tails are significantly shorter and less hydrophilic (Fig 1e,f, Fig. S2a). Since we calculated tail hydrophilicity by summing the ΔG^{app} values of all tail residues, much of the tail hydrophilicity is explained by the mere length of the tail. However, the differences in the tail hydrophilicity were sharper than the length differences, suggesting that hydrophilicity is the main adaptive determinant (Fig. 1e,f). Notably, cytosolic loops tend to be longer and more hydrophilic than extracytosolic ones, even when they do not flank the cTM (Fig S2a). However, this trend is dramatically sharper in the C-tail, as compared to other loop locations or to the N-terminal tail (Fig. S2). Collectively, these observations suggest that the C-tails are under unique

evolutionary pressure, which keeps them hydrophobic or hydrophilic depending on the tail location.

Based on these results, we hypothesized that the hydrophilicity of the C-tail facilitates the correct orientation of cTMs in membrane proteins. In C_{cyt} proteins, a hydrophilic C-tail may present a barrier preventing its erroneous translocation across the membrane, thereby anchoring the C-tail to the cytosol (Fig. 1c, Fig. S2b). By contrast, in C_{ext} proteins, a short and relatively hydrophobic C-tail may enable posttranslational insertion, by presenting a lower barrier for translocation (Fig. 1d).

Hydrophilic C-tails perturb the insertion of cTM in *E coli*.

The distinct sequence attributes of the C-tails suggest an evolutionary constraint, but is this constraint related to cTM insertion? To study this experimentally, we focused on C_{ext} proteins. At the end of translation, the more N-terminal TMs of C_{ext} proteins should already be inserted and correctly oriented (Fig. 1d), making their cTM insertion a relatively isolated and simple process. We chose two *E. coli* proteins to study experimentally: UbiA and RcnA. The two proteins are structurally unrelated; UbiA is a 4-hydroxybenzoate octaprenyltransferase and has 9 TM, while RcnA is a Nickel/cobalt efflux pump with 6 TMs.

We studied their insertion *in vivo* by a Cys accessibility assay[29] as illustrated in Fig. 2a. Briefly, if the cTM is inserted, the C-tail will translocate to the periplasm. The degree of periplasmic localization of the C-tail is quantified from the relative labeling of a cysteine engineered at the C-tail by AMS (4-Acetamido-4'-Maleimidylstilbene-2,2'-Disulfonic Acid), a Cys-reactive maleimide that labels only periplasmic cysteines as it cannot cross the plasma membrane. As controls, a membrane-permeable agent, N-ethylmaleimide (NEM), labeling both cytosolic and periplasmic cysteines is included, as well as a mock treatment labeling no cysteines. To quantify AMS labeling, the membrane is disrupted, and an additional agent, maleimide-polyethyleneglycol (mal-PEG) is used to label the remaining free cysteines that were not blocked by AMS or NEM. The level of PEGylation blocking by AMS, therefore, reports on the level of periplasmic accessibility.

To simplify the assay's interpretation, we generated cysteine-less variants of these proteins, harboring a single Cys at the C-tail. The C-tails of both UbiA and RcnA were labeled by AMS, blocking the tail's PEGylation (Fig. 2c-e, C-tail extension: 'C'). The accessibility of the C-tails to the periplasmic AMS suggests that the cTM is well inserted. We then tested if the length or hydrophobicity of the C-tail play a role in insertion. The tails of UbiA and RcnA were extended by one to four asparagines (Fig. 2b), increasing their hydrophilicity and length gradually. C_{ext} proteins in *E. coli* typically do not exceed a hydrophilicity of $\Delta G^{app}\approx 5\text{-}6$ in their C-tails (Fig. 1e, Fig. S2), as calculated by the Biological Hydrophobicity scale[30]. Remarkably, as soon as the asparagine additions to the C-tails crossed the hydrophilicity threshold, the insertion of the cTMs was abolished (Fig. 2c-e). To examine whether length or hydrophilicity affected insertion, we compared the effect of four aspargines with two other four-residue extensions. Extending the C-tail with an HSDS peptide, having similar hydrophilicity to 4xN, similarly abolished insertion. By contrast, a less hydrophilic GSGS extension maintained a wild-type level of insertion (Fig. 2c-e). Thus, sequence hydrophilicity, rather than length, poses the main barrier for insertion. Finally, we investigated the effect of an opsin tag, by appending the sequence GPNFYVPFSNKTG to the C-tail. This tag is a 13-residue-long hydrophilic tag that undergoes glycosylation in eukaryotes and is often used to study MP

insertion in the ER membrane[31]. While it is assumed that the Opsin tag does not affect MP insertion, our results show that it abolished insertion in *E. coli* (Fig. 2c-e).

We verified our insertion measurements by testing the effect of the C-terminal extensions on protein function. Since accurate TM insertion is crucial for protein folding and function, we reasoned that mutations perturbing cTM insertion should impair functionality. RcnA is a Ni^{2+}/Co^{2+} exporter, conferring resistance to *E. coli* against these heavy metals. Indeed, extensions to the C-tail that perturbed insertion impaired the ability of mutants to support bacterial growth on Co^{2+}-rich medium (Fig. 2f). Collectively, our results suggest that evolution has constrained the hydrophilicity of the C-tails to enable post-translational cTM insertion, allowing the proteins to fold and function.

YidC mediates post-translational cTM insertion in *E. coli*.

The observation that hydrophilicity perturbs cTM insertion suggests that the Sec translocon is an unlikely candidate to mediate its insertion, since this machinery should be able to handle nearly any loop or tail sequence. We, therefore, set out to identify the alternative insertion mechanism. Beyond the Sec translocon, *E. coli* possesses the YidC insertase, which can insert TMs co- or post-translationally. YidC can work independently or as a part of the holo-translocon complex along with the SecYEG translocon and SecDF-YajC[7,8]. The cytosolic protein SecA was also shown to participate in targeting and insertion of some MPs[32-34]. Notably, each of these factors may mediate the insertion of TMs other than the cTM, complicating the study of these multispanning proteins. We, therefore, first simplified our insertion assay to include only a single TM at the C-terminus. The cTMs and C-tails of RcnA and UbiA, along with a C-terminal Cys, were fused to an N-terminal His_6-sfGFP-SUMO protein (Fig. 3a). Since these constructs harbor a single TM at the C-terminus, they cannot insert cotranslationally, similar to Tail-Anchored proteins. Due to technical problems in Cys labeling of the RcnA-cTM construct, we switched to the construct with GSGS-extended C-terminus, which was properly inserted in the context of the full length protein (Fig. 2d,e). These chimeric proteins were properly inserted when expressed in *E. coli*, and a four-asparagines C-tail extension impaired insertion, suggesting that the insertion mechanism and its sequence dependence were preserved (Fig. 3b,c, 'mock'). We then tested the effect of depleting YidC, SecY, SecA or SecD on the insertion of these simplified constructs. We utilized CRISPRi to deplete the expression levels of the factors to 10% or less (Fig. 3d), and confirmed that SecY and SecA were sufficiently depleted to impair their activity (Fig. S3). SecY and SecA were dispensable for insertion, confirming that the Sec translocon indeed plays no role in cTM insertion (Fig. 3b,c). By contrast, YidC depletion severely impaired the insertion of the cTMs of RcnA and UbiA, whereas SecD had a minor but consistent effect (Fig. 3b,c). As a control, we tested a similar chimeric construct with a synthetic cTM, termed TM-D, which was shown to be YidC and Sec-independent[35]. Indeed, the depletion of YidC or the Sec components had no effect on TM-D insertion (Fig. S4), highlighting the specificity of the effect of YidC on the cTMs of UbiA and RcnA. This suggests that the YidC insertase is the main factor mediating cTM insertion of these multispanning proteins.

To explore the insertion mechanism further, we tested if the His_6-sfGFP-SUMO-cTM chimeras physically interact with these insertion factors. The chimeric constructs were expressed in *E. coli,* extracted from the membrane using synthetic nanodiscs, and purified on metal-affinity chromatography. YidC, SecD and, to some extent, SecY, copurified with the cTM chimeras of both UbiA and RcnA (Fig. 3e). This suggests that YidC interacts with these cTMs, possibly in the context of the holotranslocon. No SecA was detected in the copurified

fractions. Surprisingly, however, extending the C-tail by four asparagines had contrasting effects on UbiA and RcnA.for RcnA, the extension abolished its interaction with YidC and SecD, explaining why its insertion was impaired (Fig. 3e). By contrast, for UbiA, the 4N extension intensified the interaction with the factors (Fig. 3e).

Since the 4N extension strengthened the interaction of the UbiA cTM with the factors, we wondered if the UbiA-cTM-4N could still be inserted. We hypothesized that the extension may slow down the insertion through YidC, thereby prolonging its transient interaction with the cTM before full insertion. To examine this, we followed the insertion kinetics, by allowing insertion to proceed for 2 or 20 minutes in the presence of AMS. In the absence of the C-tail extensions, both the UbiA and RcnA cTMs inserted rapidly in a YidC-dependent manner, without appreciable increase over 20 minutes (Fig. 3f,g). The RcnA-cTM-4N, which did not copurify with YidC, was not appreciably inserted over this time (Fig. 3f,g). By contrast, the chimeric UbiA-cTM harboring the 4N extension displayed slower insertion, inserting poorly at 2 min but well at 20. The insertion was partially YidC dependent (Fig. 3f,g), suggesting that YidC mediates the slow insertion. The different behavior of UbiA and RcnA is likely rooted in the composition of their natural cTM and tail sequence. Nevertheless, our results suggest that a short and relatively hydrophobic C-tail optimizes the productive interaction of the cTM with YidC, either by regulating the physical interaction, or by affecting the insertion rate.

Disease-causing mutations in C-tails of human MPs perturb cTM insertion

Several elements in the insertion of multispanning proteins are conserved across organisms. The Sec translocon is universally conserved, and both the ER and bacterial inner membranes house Oxa1-family insertases, such as YidC in *E. coli* and EMC, TMCO, and the GET complex in the mammalian ER[36]. Our bioinformatic analysis suggests that the mechanism of cTM insertion might be conserved as well, since similar evolutionary constraints act on the C-tails of proteins from *E. coli* to human, keeping them short and hydrophobic in the case of C_{ext} proteins (Fig. 1e,f). We therefore wondered if increasing the hydrophilicity of C-tails, which perturbs cTM insertion *E. coli*, may have similar consequences in human.

MP misfolding has been linked to several diseases[3]. We therefore hypothesized that mutations causing cTM mis-insertion, and subsequent misfolding, may be pathogenic[14,15]. To test this, we mined the ClinVar database[37] for pathogenic mutations affecting the C-tails of multispanning MPs, concentrating on proteins whose wild-type C-tail length requires posttranslational insertion. We focused on nonstop mutations, where the stop codon is mutated, allowing the translation of an additional polypeptide stretch until the next in-frame stop codon. Notably, such extensions can be detrimental irrespective of insertion. Indeed, while we find pathogenic nonstop mutations in C_{ext} proteins, we also find them in C_{cyt} proteins, though C_{cyt} proteins do not rely on short C-tails for insertion (Fig. 4a). Moreover, in C_{cyt} proteins the pathogenic mutations typically extend the C-tail sufficiently to allow cotranslational rather than posttranslational insertion (Fig. 4a), further suggesting that the mechanism of pathogenicity is likely unrelated to posttranslational insertion. By contrast, the extended tails of most C_{ext} mutants remain shorter than 45 residues (Fig. 4a,b). These mutants would still rely on posttranslational insertion, but with a substantially more hydrophilic C-tail that may impair insertion.

We next studied the effects of nonstop mutations experimentally. The genes of sodium glucose co-transporter1 (SGLT1), subunit c of the integral V0 domain of vacuolar H+-ATPase

(ATP6V0C), and human equilibrative nucleoside transporter 3 (ENT3), along with their nonstop mutants, were transiently expressed in HEK293 cells. The nonstop mutations did not dramatically affect the expression of SGLT1 and ENT3. By contrast, the ATP6V0C mutant showed a drastically lowered expression, hinting at a potential degradation of the misfolded mutant (Fig. 4c). To study membrane insertion we introduced an NFT glycosylation site to the extended C-tails of the ATP6V0C, ENT3 and SGLT1 mutants. Correct insertion of the C-terminus will enable its efficient glycosylation in the ER lumen, resulting in an increase in the apparent size of the protein. SGLT1 and ENT3 harbor an endogenous glycosylation site[38,39]. To simplify interpretation, the asparagine in these sites was mutated to glutamine to eliminate native glycosylation (Fig. 4d, Q mutants). When expressed in HEK293 cells, the C-tail of SGLT1 was efficiently glycosylated, suggesting that the cTM is well inserted and the C-tail successfully enters the ER lumen (Fig. 4d). Thus, the nonstop mutation did not abolish SGLT1 insertion and the mutation's pathogenicity may stem from another reason. By contrast, neither the ATP6V0C nor the ENT3 mutants were glycosylated (Fig. 4d), suggesting that the nonstop mutation prevented the C-tail translocation to the ER lumen. Thus, the nonstop mutations abolished the cTM insertion of both these proteins.

Finally, we examined the consequences of cTM mis-insertion on the biogenesis of the ATP6V0C and ENT3 nonstop mutants. Misfolded membrane proteins typically display a trafficking defect and are retained in the endoplasmic reticulum (ER) by ER quality control. ENT3 is an organellar transporter, partially localized to lysosomes and mitochondria depending on cell type[38,40]. ATP6V0C is a subunit of vacuolar ATPase, a membrane-embedded ATP hydrolysis-driven proton pump that is responsible for acidifying and maintaining the pH of intracellular compartments including lysosomes[41]. Indeed, confocal microscopy shows that wild-type ENT3 and ATP6V0C localize to the lysosome in HEK293 cells (Fig. 5a-c). No significant co-localization of ENT3 with mitochondria was observed in HEK293 cells (Fig. S5). In contrast, lysosome localization of both ENT3 and ATP6V0C nonstop mutants is significantly reduced, and their pattern of staining is diffused throughout the cytoplasm, resembling the ER pattern (Fig. 5a-c). These findings are consistent with the expectation that cTM mis-insertion would result in mistrafficking due to the misfolding of these mutant proteins.

Collectively, our results suggest that the machinery of post-translational cTM insertion has a limited capacity to translocate hydrophilic C-tails. This limitation generates an evolutionary constraint, limiting the hydrophilicity of the C-tail in C_{ext} proteins. Mutations that increase the C-tail hydrophilicity perturb this sequence attribute, which in turn abolishes the insertion and folding of the protein, and may lead to disease.

Discussion

Membrane proteomes encompass a vast array of TMs, approximately 23,000 TMs in the human proteome and 6,000 in *E. coli*. This diversity poses a significant challenge for accurate TM insertion, essential for proteome functionality. Nature has evolved several insertion machineries to handle this diversity, however, the full spectrum of these factors and their client specificities remains incompletely understood[2,3,5,17].

For multispanning MPs, the best-characterized insertion systems are the bacterial membrane's Sec translocon (SecYEG)[6] and its eukaryotic homolog in the ER, Sec61[2]. The Sec translocons can function independently or in concert, forming complexes like the holotranslocon (SecYEG-YidC-SecDF-YajC) in bacteria[8,42,43] or the recently characterized

multipass translocon in the ER[10,11]. These complexes allow for the concerted action of the Sec protein-conducting channel, Oxa1-family insertases like YidC in bacteria and TMCO in the ER, and additional factors, all collaborating to efficiently chaperone and insert the multispanning protein[2,7,10,11,34,43-45] The ER also hosts other Oxa1 family members in the GET and EMC complexes[2,16,36], handling single-spanning proteins[20,22,23] and cotranslational insertion of some N-terminal TMs[46-48].

Our study sheds light on the posttranslational insertion process of a large subset of cTMs in multispanning proteins. We found that in *E. coli*, YidC, an Oxa1-family insertase, mediates the insertion of cTMs that are close to the C-terminus of the protein (Fig. 3). This finding adds a large number of potential clients to YidC, explaining why it is so abundant[49] and essential. It also suggests a potential explanation to the role of YidC in the insertion of proteins like SecE and ATP synthase subunit c[50,51]. Our results align with recent findings that EMC, another Oxa1-family member, performs a similar role in the mammalian ER[26].

YidC, like other Oxa1 family members, does not possess a transmembrane pore capable of translocating long polypeptide stretches. Rather, it contains a hydrophilic groove in the membrane, open from the cytoplasm, that locally thins the membrane, providing a route for TMs to enter while translocating their flanking loops across[52-57]. In line with this mechanism of insertion, the substrates of YidC and other Oxa1 family members typically have short and relatively hydrophobic translocated loops and tails[2,20,28,46,58]. This reliance on Oxa1-family insertases has exerted evolutionary pressure on the C-tails of multispanning MPs, requiring C_{ext} proteins to have short, hydrophobic C-tails to facilitate their translocation. By contrast, the longer, hydrophilic tails of C_{cyt} proteins, might prevent their translocation and anchor them in the cytosol. Thus, by modulating the ability or inability of the C-tail to translocate, evolution seemingly adapted MP sequences for optimal topogenesis. Similar evolutionary principles may apply to the correct orientation of other TMs or TM pairs that bypass the translocon. This suggests that once translation ends, the Sec translocon can no longer insert most MPs, since this machinery can easily translocate long hydrophilic loops and would not constrain the C-tail sequence.

Comparing the bacterial and human mechanisms reveals notable differences. *E. coli* C_{ext} proteins have uniformly hydrophobic C-tails, whereas human C_{ext} tails exhibit diverse and higher hydrophilicity (Fig. 1e). This suggests a greater capacity of eukaryotic insertases to handle hydrophilic C-tails, as suggested recently for EMC[26]. This increased hydrophilic capacity is also reflected in C_{cyt} proteins, with human proteins requiring a more hydrophilic C-tail to anchor the tail to the cytosol and avoid translocation (Fig. 1e). Yeast proteins populate intermediate hydrophilicity, suggesting that the capacity of Oxa1 members evolved gradually. These conclusions are supported by our experimental results, showing that the cTMs of the bacterial proteins are unable to insert correctly once C-tail hydrophilicity crosses a threshold of $\Delta G^{app} \approx 5\text{-}6$. By contrast, a similar increase in SGLT1 C-tail hydrophilicity does not abolish its insertion (Figs. 2e, 4d). Only when the mutations result in a longer and much more hydrophilic C-tail, as is the case for ATP6V0C and ENT3, the threshold hydrophilicity is reached, and C-tail insertion is disturbed (Figs. 4d, S6).

The mechanism of handover of multispanning proteins from the Sec translocon to Oxa1-familiy members is still unclear. In *E coli*, a direct transfer is plausible due to the ability of

YidC to function as part of the holotranslocon[8,42]. Indeed, we find that cTMs can interact with additional holotranslocon components, though the functional relevance of this interaction remains unclear (Fig. 3e). The human multipass translocon houses the Oxa1-family member TMCO[10,11], yet whether this protein contributes to the insertion of some cTM remains unknown. By contrast, the human EMC operates independently of Sec61[59,60], suggesting that the protein may need to diffuse away from the translocon to be engaged by EMC. Notably, the N-terminal, already inserted parts of multispanning proteins may interact with EMC cotranslationally, potentially facilitating a swift handover[26,61].

The physiological importance of this insertion mechanism, which likely serves hundreds of MPs in every proteome, is underscored by its implications in disease. Misinsertion of MPs, leading to misfolding and loss of function[14,62], has been implicated in various genetic diseases[15,63,64]. Our findings suggest that human systems are relatively tolerant to hydrophilic C-tails, making it less likely that single-residue mutations will increase the hydrophilicity sufficiently to impair insertion. However, nonstop mutations pose a greater risk, often resulting in cTM misinsertion and misfolding. Moreover, C_{ext} membrane proteins with naturally hydrophilic C-tails might be more prone to misinsertion and misfolding, depending on environmental stresses and the physiological state of the cell.

In summary, our research sheds light on the conserved mechanisms of cTM insertion across species, shaped by evolutionary forces and with direct implications for protein misfolding diseases.

Materials and Methods

Bacterial strains

E. coli DH5α was used for plasmid propagation. MG1655(DE3) strain was used for all protein expression experiments. BW25113 ΔRcnA knockout strain (Keio strain JW2093)[65] was used for Co^{2+} resistance assays.

Bacterial DNA constructs

Gene fragments of cysteine-less UbiA and RcnA were cloned into the pET19b and pBAD24 vectors, respectively, using Gibson assembly. The protein ORFs contained an N-terminal 3 x HA tag and a C-terminal cysteine. C-terminal extension variants were generated by either site-directed mutagenesis or Gibson assembly. Chimeric GFP proteins were produced by incorporating a SUMO tag followed by the last TM and C-tail of either UbiA or RcnA, along with a C-terminal cysteine, into pBAD/HisB-sfGFP (TIR^{STD})[66] (a kind gift from Daniel O. Daley). Guide RNAs (gRNAs) specific to SecY (tagtAAGCAGTTTGGCAAGTACAG), SecA(tagtCAGACGTGCACGAAACTCTG), YidC (tagtTCTTTTTCAACGTTATACAG), and SecD (tagtGGAGTCGGTGATATGGTCAC) were integrated into the pFD152 vector[67] (Addgene #125546) using Gibson assembly. All constructs were verified by sequencing.

Bacterial Cell Growth and Protein Expression

For full length proteins, MG1655 (DE3) cells harboring plasmids encoding the proteins of interest were grown in LB supplemented with 100 µg/ml ampicillin at 37 °C to mid-log phase (OD_{600} of ~0.5). The cultures were then induced with 0.025 mM IPTG for UbiA variants or 0.01% arabinose for RcnA variants for 30 min induction period, then placed on ice. For

CrisprI experiments, MG1655 (DE3) cells doubly transformed with plasmids encoding the appropriate gRNAs and GFP-fusion proteins of interest were grown in LB supplemented with 50 µg/ml ampicillin and 40 µg/ml spectinomycin. After 50 min, CAS9 expression was induced by the addition of 25 µg/ml anhydrotetracycline, and the cultures were allowed to grow for an additional 4 hours. Protein expression was then induced with 0.02% arabinose for 30 mins before the cultures were placed on ice. For pull-down experiments, MG1655 (DE3) cells transformed with plasmids encoding the GFP-chimeric proteins of interest were grown to mid-log phase, then induced with 0.02% arabinose for 15 min before the cultures were placed on ice.

Cysteine blocking in whole cells

The protocol was adapted from[68]. Cells were collected by centrifugation (4 °C, 4,700 rpm, 5 min), resuspended in labeling buffer (5 mM MgSO4, 1 mM (tris(2-carboxyethyl)phosphine (TCEP) in PBS) and divided into three microcentrifuge tubes for blocking with AMS (Setareh Biotech 6508), NEM or labeling buffer. The cells were equilibrated at 37°C with gentle shaking for 4 min and the sulfhydryl reagents (dissolved freshly in labeling buffer) or labeling buffer were added. Final concentration was 5 mM for AMS and 10 mM for NEM. The cells were allowed to react for 2 or 20 min at 37°C before quenching for 20 min on ice with dithiothreitol (DTT) at a final concentration of 40 mM. For CrisprI depletion experiments, a fourth sample of cells was set aside to assess depletion levels of SecY, SecA, YidC or SecD.

Cell Lysis and Cysteine PEGylation:

For full length proteins, proteins were precipitated by adding trichloroacetic acid (10% final concentration) and incubation for 30 min on ice, followed by centrifugation at (4 °C, 20000g, 15 min). Supernatant was removed and the pellet was washed with ice cold acetone followed by centrifugation (4 °C, 20000g, 5 min). Acetone was removed and the pellet was completely solubilized in 50 µl of pegylation buffer (100 mM Tris-HCl pH 7.5, 2% SDS, 10% glycerol, 1 mM TCEP, trace amount of bromophenol blue, and protease inhibitor). 13 µl of freshly dissolved 27 mM 5 kDa methoxypolyethylene glycol maleimide (mal-PEG) (Sigma 63187) were added, and the samples were PEGylated with continuous mixing at 30 °C for 2-60 min. The reaction was quenched by the addition of equal volume of stop buffer (2% SDS, 10% glycerol, 40 mM DTT).

For CrisprI experiments, cells were collected by centrifugation (4 °C, 10000g, 1 min), resuspended in 150 µl of lysozyme buffer (1 mg/ml lysozyme in 150 mM NaCl, 30 mM Tris-HCl, pH 8, 10 mM EDTA, 1 mM TCEP, and protease inhibitor), then frozen at −20 °C overnight. The next day, cells were disrupted by thawing at 25 °C for 5 min followed by shaking at 37 °C for 10 min. Then, 0.9 ml of turbo DNase solution (15 mM $MgSO_4$, 10 µg ml^{-1} turbo DNase (Jena Bioscience, EN-180), 1 mM TCEP and 0.5 mM phenylmethanesulfonyl fluoride) was added and the samples were allowed to shake at 37 °C for 10 min before transferring them to ice. Crude membranes were collected by centrifugation at 4 °C, 20,000*g* for 20 min and the pellet was resuspended in 50 µl of PBS supplemented with 1 mM TCEP and cOmplete protease inhibitor (Roche, 11836170001). Membrane proteins were solubilized and PEGylated by adding 7 µl of 10% β-d-dodecyl maltoside (Anatrace 329370010) and 13 µl of freshly dissolved 27 mM 5 kDa mal-PEG. The samples were incubated with continuous mixing at 30 °C for 2-60 min and the reaction was quenched by mixing with 70 µl of GFP sample buffer (50 mM Tris pH 8, 10% glycerol, 2% SDS, 20 mM DTT, and a trace amount of bromophenol blue).

Pull-downs:
Following induction, 15 OD units of cells (approximately 25.5 mg) were harvested, pelleted, and incubated on ice with lysozyme buffer supplemented with DNase to remove the outer wall. Following a 2 min centrifugation at 10,000g, the resulting spheroplasts were washed, resuspended in 80 µl of 5% (w/v) SMALP 140 (Cube Biotech) in purification buffer (20mM Tris pH 8.0, 500mM NaCl, 5% Glycerol, 1 mM mercaptoethanol, 5 mM Imidazole, 0.5mM PMSF), and incubated overnight at 4°C with shaking (1800 rpm). The sample was then diluted with 900µl of purification buffer and incubated with shaking for another 30 min of, followed by pelleting of unsolubilized debris. The supernatant was incubated with 35µl of Talon metal affinity resin (TaKaRa) at 4°C for 3 hours. Subsequently, the resin was washed four times with purification buffer, and proteins were eluted with 400mM imidazole in purification buffer.

Bacterial protein analysis and Western blotting
All PEGylated samples were resolved on 12% SDS–PAGE gels run with a modified running TRIS-glycine buffer containing only 0.05% SDS. PEGylated full-length proteins were transferred to nitrocellulose membrane and probed with mouse anti-HA antibodies (Sigma-Aldrich, H9658, 1:5000) overnight at 4°C. PEGylated GFP-chimeras were visualized by Typhoon biomolecular imager (Cytiva) using a 488 nm excitation wavelength and Cy2 filter. For analysis of protein depletion by CriprI, samples were seperated on 4-20% gradient Bis-Tris SDS-PAGE in MES running buffer, then transferred to nitrocellulose. The membrane was probed with polyclonal antibodies against SecY (1:3000), SecA (1:10000), YidC (1:10000), SecD (1:2500)[69], and OmpA (1:10000, Bioorbyt orb862303). Pull-down GFP-chimera samples were separated on 12% SDS-PAGE gels in Tris-Glycine containing 0.05% SDS, followed by in-gel fluorescence detection. Subsequently, the samples were transferred to nitrocellulose and probed with polyclonal antibodies against SecY, SecA, YidC (1:5000), and SecD as above. ImageQuant™ TL 10.2 analysis software (Cytiva) was used for band quantification. To ensure maximal accuracy of PEGylation level quantification, a calibration curve was constructed for each experiment using a series of 2-fold dilutions of sample.

Co^{2+} resistance assay

ΔRcnA knockout strain was transformed with pBAD24 plasmid with and without the different RcnA C-tail mutants. Cells were grown to mid-log on M63 minimal medium supplemented with 0.2% glycerol and 0.4% glucose, and tenfold serial dilutions were prepared with the highest density having an OD_{600} of 0.1. The serial dilution was spotted (3 µL) on supplemented M63 agar–ampicillin plates containing 0.1% arabinose and either 0 or 50 mM CoCl2. The plates were allowed to grow for one night at 37 °C. As mutation of the conserved cysteine187 of RcnA reduces its activity, the complementation assay was done with RcnA C-tail mutants that retain this cysteine.

Mammalian cell culture and transient transfections
HEK-293 cells (ATCC, CRL-1573) were subcultured every 3-4 days in Dulbecco's modified Eagle's medium (Thermo Fisher Scientific, 41965-039) supplemented with 10% v/v fetal bovine serum (Sigma Aldrich, F7524), 1% l-glutamine (Biological Industries – Sartorius, 03-020-1B), and 1% Penicillin Streptomycin Solution (Biological Industries – Sartorius, 03-031-

1B) under standard conditions (37 °C and 5% CO_2 in a humidified incubator). Cells were routinely tested by PCR for the absence of mycoplasma contamination. Cells were transiently transfected with constructs using the JetPRIME reagents (Polyplus Transfection 101000015) according to the manufacturers' instructions. Typically, 1.0-1.5 µg DNA was used to transfect cells 1n a 6-well plate (0.4 µg DNA for wild type ATP6V0C). Cells were harvested 44-48 h after transfection.

Mammalian DNA constructs

Gene fragments of ATP6V0C, ENT3 (SLC29A3) and SGLT1 (SLC5A1) and their stop codon mutants (ClinVar variants NM_001694.4(ATP6V0C):c.467A>T (p.Ter156Leu), NM_018344.6 (SLC29A3):c.1427A>G (p.Ter476Trp) and NM_000343.4(SLC5A1):c.1993T>C (p.Ter665Arg)) were cloned into pCDNA3.1-HA (Addgene 128034) using Gibson assembly. Glycosylation site mutagenesis of ENT3 (N84Q) and SGLT1 (N248Q) and addition of an NFT glycosylation site 6-9 amino acids from the C-terminal of all three proteins was done by site-directed mutagenesis PCR. All constructs were verified by sequencing.
ER-mScarletI[70] (Addgene 137805) was used as an ER organellar marker.

Mammalian protein extraction, deglycosylation assays and immunoblotting

For expression levels analysis, cells were washed once with ice cold PBS and then lysed with SDS lysis buffer (2% SDS, 50mM TRIS-HCl pH6.8, 10% glycerol, turbo nuclease and cOmplete protease inhibitor for 20 min. at 37^0C. The lysate was quantified using BCA protein assay (Pierce, 23227) and supplemented with a trace of bromophenol blue and DTT (0.1 M final concentration). Equal amounts of proteins were loaded on gels. For deglycosylation experiments, cells were lysed 30 min on ice in triton lysis buffer (50 mM Tris–HCl pH 8, 150 mM NaCl, 1% triton, 1mM EDTA and protease inhibitor). The lysate was centrifuged for 15 min. at 20,000 g at 4^0C, and the supernatant was digested for 1 h at 37°C with PNGase F (New England Biolabs, P0704) according to the manufacturer's instructions. Negative controls were prepared in the same manner without the addition of enzymes. Following digestion, the samples were supplemented with 4× reducing Laemmli buffer.
Cell lysates were analyzed by SDS–PAGE on 12% Tris-glycine gels. SDS–PAGE gels were transferred to a nitrocellulose membrane (for ATP6V0C samples) or PVDF membranes (for ENT3 and SGLT1). Membranes were probed with mouse anti-HA antibodies (Sigma-Aldrich, H9658, 1:5000) overnight at 4^0C.

Immunofluorescence staining

HEK-293 cells were seeded on poly-L-lysine (Sigma Aldrich, P4707) coated cover slips in a 24-well plate and transiently transfected the following day with 50-250 ng DNA. The cells were fixed 44-48 h after transfection. If required, MitoTracker® Deep Red FM (Invitrogen, M22426) at a final concentration of 125 nM was added to the growth media of cells prior to fixation. Following 10 min incubation with MitoTracker®, the cells were washed once, growth media was replaced, and the cells were incubated for another 30min. The cells were fixed with 4% PFA in PBS for 15 minutes at room temperature, permeabilized with 0.1% saponine in PBS for 10 min, and blocked in 10% FBS in PBS (blocking buffer) for 1h. The cells were then incubated with primary antibodies in blocking buffer for 2 h, and with secondary antibodies for 1 h, followed by 10 min incubation with DAPI at 0.5 µg/ml. Slides were mounted in Fluoromount-G (SouthernBiotech Associates, 0100-01). The following

antibodies were used: mouse anti-HA (BioLegend, 901501, 1:750), rabbit anti-HA (Sigma-Aldrich, H6908, 1:750), mouse anti-LAMP1 (DSHB, H4A3, 1:200), goat Anti-Mouse IgG H&L Alexa Fluor® 488 (Abcam, 150113, 1:500) and donkey Anti-Rabbit IgG H&L Alexa Fluor® 568 (ab175692, 1:600).

Confocal microscopy and image analysis

The cells were imaged with a DMi8 Leica confocal laser-scanning microscope (Leica Microsystems, Mannheim, Germany), using ×60/1.4 oil-immersion objective. Fluorescence of DAPI, Alexa488, and Alexa568 or mScarlet was acquired using 405 nm, 488 nm and 561 nm lasers, with emission collection width set to 415-460, 500-550 and 570-630, respectively. Fluorescence of MitoTracker® Deep Red FM was acquired by white-light laser using a 641 nm laser line with emission collection width set to 650-750.

The images were analyzed using Fiji software. Background subtraction was performed using control samples treated only with secondary antibodies (for LAMP1) or by using mock transfection controls (for HA-tagged proteins or the ER-mScarletl marker). Thresholds were determined using Otsu's threshold clustering algorithm, followed by pixel- to-pixel measurement of co-occurrence in the region of interest.

Topology Predictions

The entire reviewed proteomes of E. coli K-12 (taxon identifier 83333), S. cerevisiae (559292) and human (9606) were downloaded from Uniprot (as of January 2023) and were parsed using custom Python scripts. To analyze the topologies of the proteomes, topologies were downloaded from three sources: CCTOP[71], TMAlphaFold[72], and PolyPhobius[73]. From TM-AlphaFold, we kept only proteins that passed at least 5 tests[72]. To obtain Polyphobius predictions, we utilized the TOPCONS server. We crossed data from all predictors to check the level of agreement in the number of TMs and the position of their core residues (within a 14-residue window). In cases where the three predictors agree, we labeled the topology as 'high' quality. When disagreements were present, we prioritized topologies where CCTOP and TMAlphaFold agree, followed by agreements between TM-AlphaFold and PolyPhobius. When there was no agreement, or when data were missing, we preferred Polyphobius predictions over CCTOP and the topologies were labeled as 'low' quality.

The membrane topology (TMH boundaries and membrane orientation) and biological hydrophobicity (ΔG^{app} values for TMH insertion) of the entire proteome were analyzed by PolyPhiobius and the ΔG-predictor, as implemented in the Topcons webserver[74]. Proteins longer than 10,000 or shorter than 18 residues were removed. The amino acid 'U' (selenocysteine) was substituted by cysteine. To assign the hydrophobicity ΔG^{app} value for each TMH, the minimal ΔG^{app} given by the ΔG predictor value within five to six residues of the TMH center was kept. Signal peptides were ignored. To analyze loop hydrophobicity, the ΔG^{app} values of all residues within the loops were summed. The ΔG^{app} values for loops were taken from the biological hydrophobicity scale[30], corresponding to residues positioned in the middle of the membrane. The topology predictions are provided in the Supplementary Table.

In order to preferentially analyze co-translationally inserted proteins, proteins containing less than 3 TMHs were discarded. In addition, we discarded mitochondrial and peroxisomal proteins. The identities of yeast mitochondrial and peroxisomal proteins were provided as a

personal communication from the lab of Maya Schuldiner. Human mitochondrial proteins were taken from MitoCarta3.0 on March 2022[75]. human peroxisomal proteins were from[76].

References

1. Wallin, E. & von Heijne, G. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci* **7**, 1029-38 (1998).
2. Hegde, R.S. & Keenan, R.J. The mechanisms of integral membrane protein biogenesis. *Nature Reviews Molecular Cell Biology* **23**, 107-124 (2022).
3. Marinko, J.T. et al. Folding and Misfolding of Human Membrane Proteins in Health and Disease: From Single Molecules to Cellular Proteostasis. *Chem Rev* **119**, 5537-5606 (2019).
4. Cymer, F., von Heijne, G. & White, S.H. Mechanisms of integral membrane protein insertion and folding. *J Mol Biol* **427**, 999-1022 (2015).
5. De Geyter, J., Smets, D., Karamanou, S. & Economou, A. Inner Membrane Translocases and Insertases. in *Bacterial Cell Walls and Membranes* (ed. Kuhn, A.) 337-366 (Springer International Publishing, Cham, 2019).
6. Oswald, J., Njenga, R., Natriashvili, A., Sarmah, P. & Koch, H.G. The Dynamic SecYEG Translocon. *Front Mol Biosci* **8**, 664241 (2021).
7. Shanmugam, S.K. & Dalbey, R.E. The Conserved Role of YidC in Membrane Protein Biogenesis. *Microbiol Spectr* **7**(2019).
8. Troman, L. & Collinson, I. Pushing the Envelope: The Mysterious Journey Through the Bacterial Secretory Machinery, and Beyond. *Front Microbiol* **12**, 782900 (2021).
9. Bai, L. & Li, H. Structural insights into the membrane chaperones for multi-pass membrane protein biogenesis. *Current Opinion in Structural Biology* **79**, 102563 (2023).
10. Smalinskaite, L., Kim, M.K., Lewis, A.J.O., Keenan, R.J. & Hegde, R.S. Mechanism of an intramembrane chaperone for multipass membrane proteins. *Nature* **611**, 161-166 (2022).
11. Sundaram, A. et al. Substrate-driven assembly of a translocon for multipass membrane proteins. *Nature* **611**, 167-172 (2022).
12. Mercier, E., Wintermeyer, W. & Rodnina, M.V. Co-translational insertion and topogenesis of bacterial membrane proteins monitored in real time. *EMBO J* **39**, e104054 (2020).
13. Nicolaus, F. et al. Residue-by-residue analysis of cotranslational membrane protein integration in vivo. *Elife* **10**(2021).
14. Chamness, L.M. et al. Molecular basis for the evolved instability of a human G-protein coupled receptor. *Cell Rep* **37**, 110046 (2021).
15. Sun, S. & Mariappan, M. C-terminal tail length guides insertion and assembly of membrane proteins. *J Biol Chem* **295**, 15498-15510 (2020).
16. Aviram, N. & Schuldiner, M. Targeting and translocation of proteins to the endoplasmic reticulum at a glance. *J Cell Sci* **130**, 4079-4085 (2017).
17. Guna, A., Hazu, M., Pinton Tomaleri, G. & Voorhees, R.M. A TAle of Two Pathways: Tail-Anchored Protein Insertion at the Endoplasmic Reticulum. *Cold Spring Harb Perspect Biol* **15**(2023).

18. Doan, K.N. et al. The Mitochondrial Import Complex MIM Functions as Main Translocase for α-Helical Outer Membrane Proteins. *Cell Rep* **31**, 107567 (2020).
19. Guna, A. et al. MTCH2 is a mitochondrial outer membrane protein insertase. *Science* **378**, 317-322 (2022).
20. Guna, A., Volkmar, N., Christianson, J.C. & Hegde, R.S. The ER membrane protein complex is a transmembrane domain insertase. *Science* **359**, 470-473 (2018).
21. Peschke, M. et al. SRP, FtsY, DnaK and YidC Are Required for the Biogenesis of the E. coli Tail-Anchored Membrane Proteins DjlC and Flk. *J Mol Biol* **430**, 389-403 (2018).
22. Schuldiner, M. et al. The GET complex mediates insertion of tail-anchored proteins into the ER membrane. *Cell* **134**, 634-45 (2008).
23. Stefanovic, S. & Hegde, R.S. Identification of a Targeting Factor for Posttranslational Membrane Protein Insertion into the ER. *Cell* **128**, 1147-1159 (2007).
24. Jung, S.J., Kim, J.E., Reithinger, J.H. & Kim, H. The Sec62-Sec63 translocon facilitates translocation of the C-terminus of membrane proteins. *J Cell Sci* **127**, 4270-8 (2014).
25. Potter, M.D. & Nicchitta, C.V. Endoplasmic reticulum-bound ribosomes reside in stable association with the translocon following termination of protein synthesis. *J Biol Chem* **277**, 23314-20 (2002).
26. Wu, H., Smalinskaite, L. & Hegde, R.S. EMC rectifies the topology of multipass membrane proteins. *Nat Struct Mol Biol* (2023).
27. Voorhees, R.M. & Hegde, R.S. Structure of the Sec61 channel opened by a signal sequence. *Science* **351**, 88-91 (2016).
28. Hennon, S.W., Soman, R., Zhu, L. & Dalbey, R.E. YidC/Alb3/Oxa1 Family of Insertases*. *Journal of Biological Chemistry* **290**, 14866-14874 (2015).
29. Bogdanov, M. Exploring Uniform, Dual, and Dynamic Topologies of Membrane Proteins by Substituted Cysteine Accessibility Method (SCAM™). *Methods Mol Biol* **2715**, 121-157 (2024).
30. Hessa, T. et al. Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* **450**, 1026-1030 (2007).
31. Favaloro, V., Spasic, M., Schwappach, B. & Dobberstein, B. Distinct targeting pathways for the membrane insertion of tail-anchored (TA) proteins. *J Cell Sci* **121**, 1832-40 (2008).
32. Rawat, S., Zhu, L., Lindner, E., Dalbey, R.E. & White, S.H. SecA drives transmembrane insertion of RodZ, an unusual single-span membrane protein. *J Mol Biol* **427**, 1023-37 (2015).
33. Wang, S., Yang, C.I. & Shan, S.O. SecA mediates cotranslational targeting and translocation of an inner membrane protein. *J Cell Biol* **216**, 3639-3653 (2017).
34. Zhu, Z., Wang, S. & Shan, S.-o. Ribosome profiling reveals multiple roles of SecA in cotranslational protein export. *Nature Communications* **13**, 3393 (2022).
35. Peschke, M. et al. Distinct Requirements for Tail-Anchored Membrane Protein Biogenesis in Escherichia coli. *mBio* **10**(2019).
36. Anghel, S.A., McGilvray, P.T., Hegde, R.S. & Keenan, R.J. Identification of Oxa1 Homologs Operating in the Eukaryotic Endoplasmic Reticulum. *Cell Reports* **21**, 3708-3716 (2017).
37. Landrum, M.J. et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* **46**, D1062-d1067 (2018).

38. Baldwin, S.A. et al. Functional Characterization of Novel Human and Mouse Equilibrative Nucleoside Transporters (hENT3 and mENT3) Located in Intracellular Membranes*. *Journal of Biological Chemistry* **280**, 15880-15887 (2005).
39. Hediger, M.A., Mendlein, J., Lee, H.S. & Wright, E.M. Biosynthesis of the cloned intestinal Na+/glucose cotransporter. *Biochim Biophys Acta* **1064**, 360-4 (1991).
40. Govindarajan, R. et al. Facilitated mitochondrial import of antiviral and anticancer nucleoside drugs by human equilibrative nucleoside transporter-3. *American Journal of Physiology-Gastrointestinal and Liver Physiology* **296**, G910-G922 (2009).
41. Vasanthakumar, T. & Rubinstein, J.L. Structure and Roles of V-type ATPases. *Trends Biochem Sci* **45**, 295-307 (2020).
42. Botte, M. et al. A central cavity within the holo-translocon suggests a mechanism for membrane protein insertion. *Scientific Reports* **6**, 38399 (2016).
43. Komar, J. et al. Membrane protein insertion and assembly by the bacterial holo-translocon SecYEG-SecDF-YajC-YidC. *Biochem J* **473**, 3341-54 (2016).
44. McDowell, M.A., Heimes, M. & Sinning, I. Structural and molecular mechanisms for membrane protein biogenesis by the Oxa1 superfamily. *Nat Struct Mol Biol* **28**, 234-239 (2021).
45. McGilvray, P.T. et al. An ER translocon for multi-pass membrane protein biogenesis. *eLife* **9**, e56889 (2020).
46. Chitwood, P.J., Juszkiewicz, S., Guna, A., Shao, S. & Hegde, R.S. EMC Is Required to Initiate Accurate Membrane Protein Topogenesis. *Cell* **175**, 1507-1519.e16 (2018).
47. Leznicki, P., Schneider, H.O., Harvey, J.V., Shi, W.Q. & High, S. Co-translational biogenesis of lipid droplet integral membrane proteins. *J Cell Sci* **135**(2022).
48. O'Keefe, S. et al. An alternative pathway for membrane protein biogenesis at the endoplasmic reticulum. *Communications Biology* **4**, 828 (2021).
49. Urbanus, M.L. et al. Targeting, insertion, and localization of Escherichia coli YidC. *J Biol Chem* **277**, 12718-23 (2002).
50. van der Laan, M., Bechtluft, P., Kol, S., Nouwen, N. & Driessen, A.J. F1F0 ATP synthase subunit c is a substrate of the novel YidC pathway for membrane protein biogenesis. *J Cell Biol* **165**, 213-22 (2004).
51. Yi, L. et al. YidC is strictly required for membrane insertion of subunits a and c of the F(1)F(0)ATP synthase and SecE of the SecYEG translocase. *Biochemistry* **42**, 10537-44 (2003).
52. Chen, Y. et al. YidC Insertase of Escherichia coli: Water Accessibility and Membrane Shaping. *Structure* **25**, 1403-1414.e3 (2017).
53. He, H., Kuhn, A. & Dalbey, R.E. Tracking the Stepwise Movement of a Membrane-inserting Protein In Vivo. *Journal of Molecular Biology* **432**, 484-496 (2020).
54. Kumazaki, K. et al. Structural basis of Sec-independent membrane protein insertion by YidC. *Nature* **509**, 516-20 (2014).
55. Kumazaki, K. et al. Crystal structure of Escherichia coli YidC, a membrane protein chaperone and insertase. *Sci Rep* **4**, 7299 (2014).
56. Shimokawa-Chiba, N. et al. Hydrophilic microenvironment required for the channel-independent insertase function of YidC protein. *Proc Natl Acad Sci U S A* **112**, 5063-8 (2015).
57. Wickles, S. et al. A structural model of the active ribosome-bound membrane protein insertase YidC. *Elife* **3**, e03035 (2014).

58. Soman, R., Yuan, J., Kuhn, A. & Dalbey, R.E. Polarity and Charge of the Periplasmic Loop Determine the YidC and Sec Translocase Requirement for the M13 Procoat Lep Protein*. *Journal of Biological Chemistry* **289**, 1023-1032 (2014).
59. O'Donnell, J.P. et al. The architecture of EMC reveals a path for membrane protein insertion. *Elife* **9**(2020).
60. Wu, H. & Hegde, R.S. Mechanism of signal-anchor triage during early steps of membrane protein insertion. *Mol Cell* **83**, 961-973.e7 (2023).
61. Shurtleff, M.J. et al. The ER membrane protein complex interacts cotranslationally to enable biogenesis of multipass membrane proteins. *Elife* **7**(2018).
62. Marshall, S.S. et al. A Link between Integral Membrane Protein Expression and Simulated Integration Efficiency. *Cell Rep* **16**, 2169-2177 (2016).
63. Coelho, J.P.L. et al. A network of chaperones prevents and detects failures in membrane protein lipid bilayer integration. *Nat Commun* **10**, 672 (2019).
64. Roushar, F.J. et al. Contribution of Cotranslational Folding Defects to Membrane Protein Homeostasis. *J Am Chem Soc* **141**, 204-215 (2019).
65. Baba, T. et al. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, 2006.0008 (2006).
66. Shilling, P.J., Khananisho, D., Cumming, A.J., Söderström, B. & Daley, D.O. Signal amplification of araC pBAD using a standardized translation initiation region. *Synthetic Biology* **7**(2022).
67. Depardieu, F. & Bikard, D. Gene silencing with CRISPRi in bacteria and optimization of dCas9 expression levels. *Methods* **172**, 61-75 (2020).
68. Seurig, M., Ek, M., von Heijne, G. & Fluman, N. Dynamic membrane topology in an unassembled membrane protein. *Nature Chemical Biology* **15**, 945-948 (2019).
69. Mayer, M. et al. Co-Translational Membrane Targeting and Holo-Translocon Docking of Ribosomes Translating the SRP Receptor. *Journal of Molecular Biology* **434**, 167459 (2022).
70. Chertkova, A. et al. Robust and Bright Genetically Encoded Fluorescent Markers for Highlighting Structures and Compartments in Mammalian Cells. (bioRxiv, 2017).
71. Dobson, L., Reményi, I. & Tusnády, G.E. CCTOP: a Consensus Constrained TOPology prediction web server. *Nucleic Acids Res* **43**, W408-12 (2015).
72. Dobson, L. et al. TmAlphaFold database: membrane localization and evaluation of AlphaFold2 predicted alpha-helical transmembrane protein structures. *Nucleic Acids Res* **51**, D517-d522 (2023).
73. Käll, L., Krogh, A. & Sonnhammer, E.L. An HMM posterior decoder for sequence feature prediction that includes homology information. *Bioinformatics* **21 Suppl 1**, i251-7 (2005).
74. Tsirigos, K.D., Peters, C., Shu, N., Käll, L. & Elofsson, A. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res* **43**, W401-7 (2015).
75. Rath, S. et al. MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations. *Nucleic Acids Res* **49**, D1541-d1547 (2021).
76. Yifrach, E. et al. Defining the Mammalian Peroxisomal Proteome. *Subcell Biochem* **89**, 47-66 (2018).

Acknowledgements

We thank Ofer Shoshani and Karen Hakeny for invaluable help with establishing human cell culture facility and methods; Yael Alon, Ivan Milenkovic, Natali Muskat and Inna Goliand for their helpful advice on immunofluorescence staining and confocal imaging, Yoav Peleg for consultations on the cloning, and Eitan Bibi for kindly providing SecY, SecA, SecD and YidC antibodies. This research was generously supported by research grants from the ISRAEL SCIENCE FOUNDATION (grants no. 2207/21 and 2208/21), the Kekst Family Institute for Medical Genetics, the Center for New Scientists at the Weizmann Institute of Science, and the Weizmann SABRA - Yeda-Sela - WRC Program, the Estate of Emile Mimran, and The Maurice and Vivienne Wohl Biology Endowment. Illustrations were Created with BioRender.com.

Author Contributions

I.A.K., H.P-Z. and N.F. designed the study and the experiments.
I.A.K. and H.P-Z. conducted the experiments.
I.A.K., A.B.D.B., S.A.T and Y.W performed the computational analysis.
H.P-Z. and R.N. performed the image analysis.
N.F., I.A.K. and H.P-Z. conducted the majority of data analysis.
N.F. and H.P-Z. wrote the manuscript.
All authors discussed the results and contributed to editing the manuscript.
H.P-Z and N.F. conceived the study.
N.F. supervised the work.

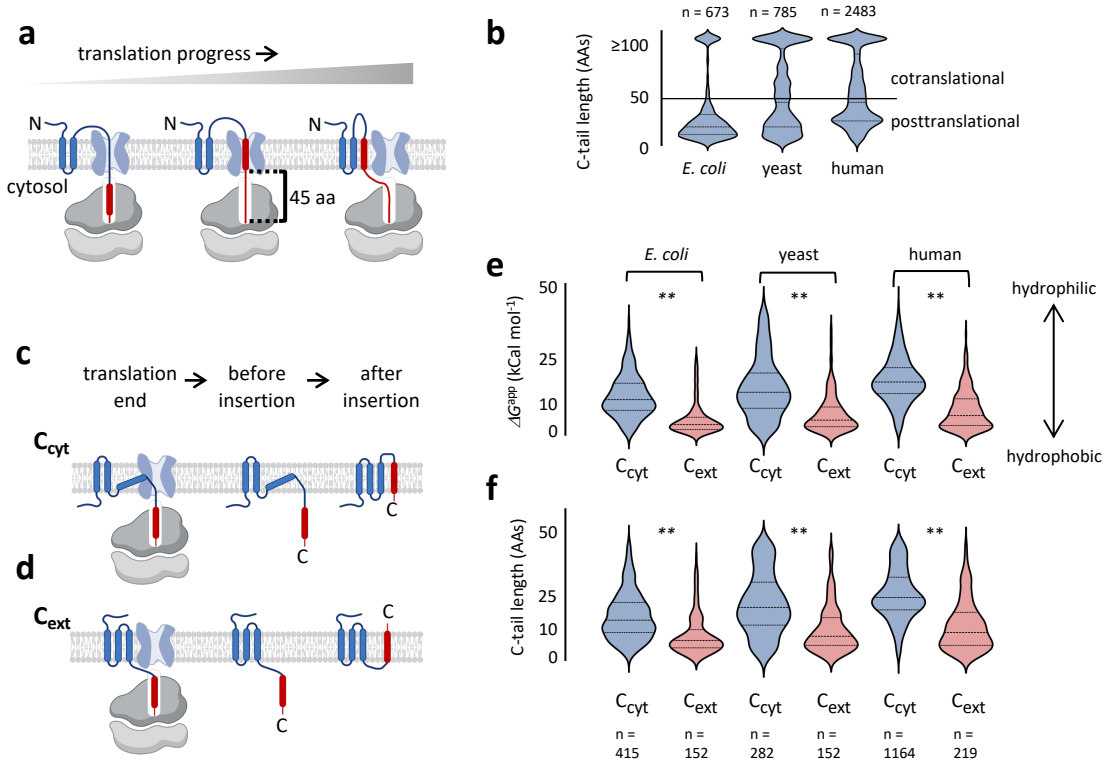


Fig. 1. The insertion route of cTMs correlates with the properties of their C-tails
(a) Illustration of cotranslational insertion, in which the inserted cTM (red) is accessible to the translocon (light blue) only after 45 amino acids have been synthesized following the TM.
(b) The distribution of C-tail length among proteins from different organisms suggests that many proteins depend on posttranslational cTM insertion.
(c,d) Illustration of the events leading to posttranslational insertion of cTM, in C_{cyt} (c) and C_{ext} (d) proteins. The events include the end of translation and posttranslational topological rearrangements. C_{cyt} proteins translocate the penultimate loop, whereas C_{ext} proteins translocate the C-tail.
(e,f) The hydrophilicity (e) and length (f) of posttranslational C-tails are dramatically different between C_{cyt} and C_{ext} MPs across organisms. C_{ext} proteins have shorter and less hydrophilic C-tails. ** indicates $p < 10^{-18}$, Mann-Whitney test.

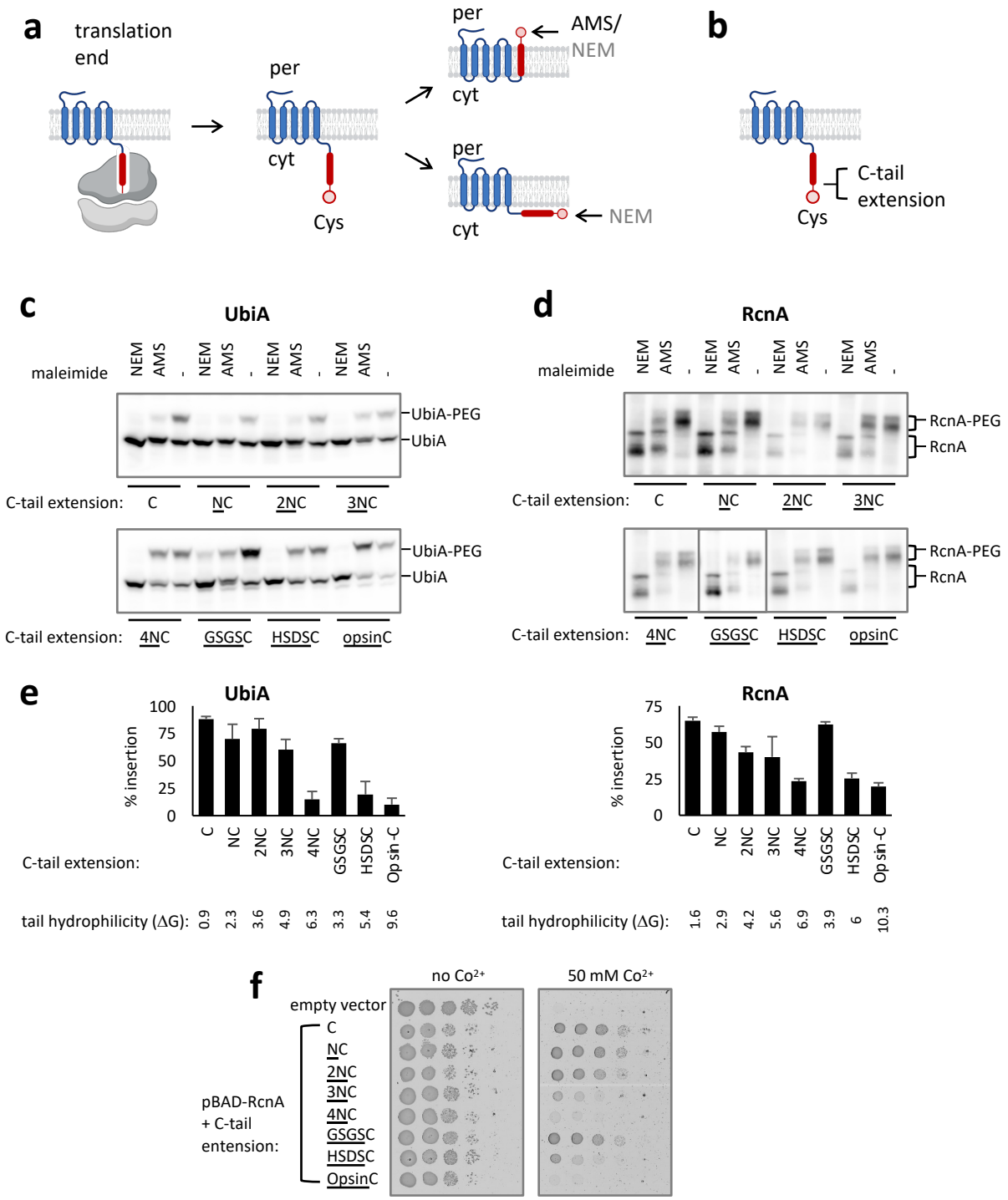


Fig. 2. Increasing the hydrophilicity of the C-tail in C_{ext} proteins perturbs insertion.
(a) Measuring cTM insertion *in vivo* by the accessibility of a C-terminal Cys to the periplasmic reagent AMS. Following translation termination, a Cys located in the C-tail lies in the cytosol (cyt). Upon insertion, it becomes accessible to periplasmic (per) AMS. NEM labels the C-tail regardless of its insertion status, since it can access both cytosolic and periplasmic Cys. The degree of labeling is quantified from the extent to which AMS blocks Cys PEGylation.
(b) Location of C-terminal extensions used to assess how the properties of the C-tail affect insertion.
(c,d) Periplasmic accessibility assay, measuring the localization of the C-tail by its ability to be labeled by AMS. AMS blocks PEGylation when the protein is properly inserted. Cys PEGylation causes a shift in SDS-PAGE migration, as visualized by Western blotting. Several C-tail extensions perturb cTM insertion and C-tail access to the periplasm. Representative blots of three biological repeats are shown for UbiA (c) and RcnA (d).
(e) Quantifications of the experiments shown in panels c and d, showing that C-terminal hydrophilicity perturbs cTM insertion. Shown are means ± SEM of at least three biological replicates.
(f) Co^{2+}-resistance conferred by various RcnA C-tail mutants suggests that mutations perturbing cTM insertion impair RcnA function. A serial 10-fold dilution of ΔRcnA *E. coli* cells harboring an empty vector, or pBAD-RcnA encoding the indicated version was grown on solid media. Growth In the absence of $Co^{2+}$is independent of RcnA, whereas growth on Co^{2+}-containing media depends on the resistance conferred by active RcnA.

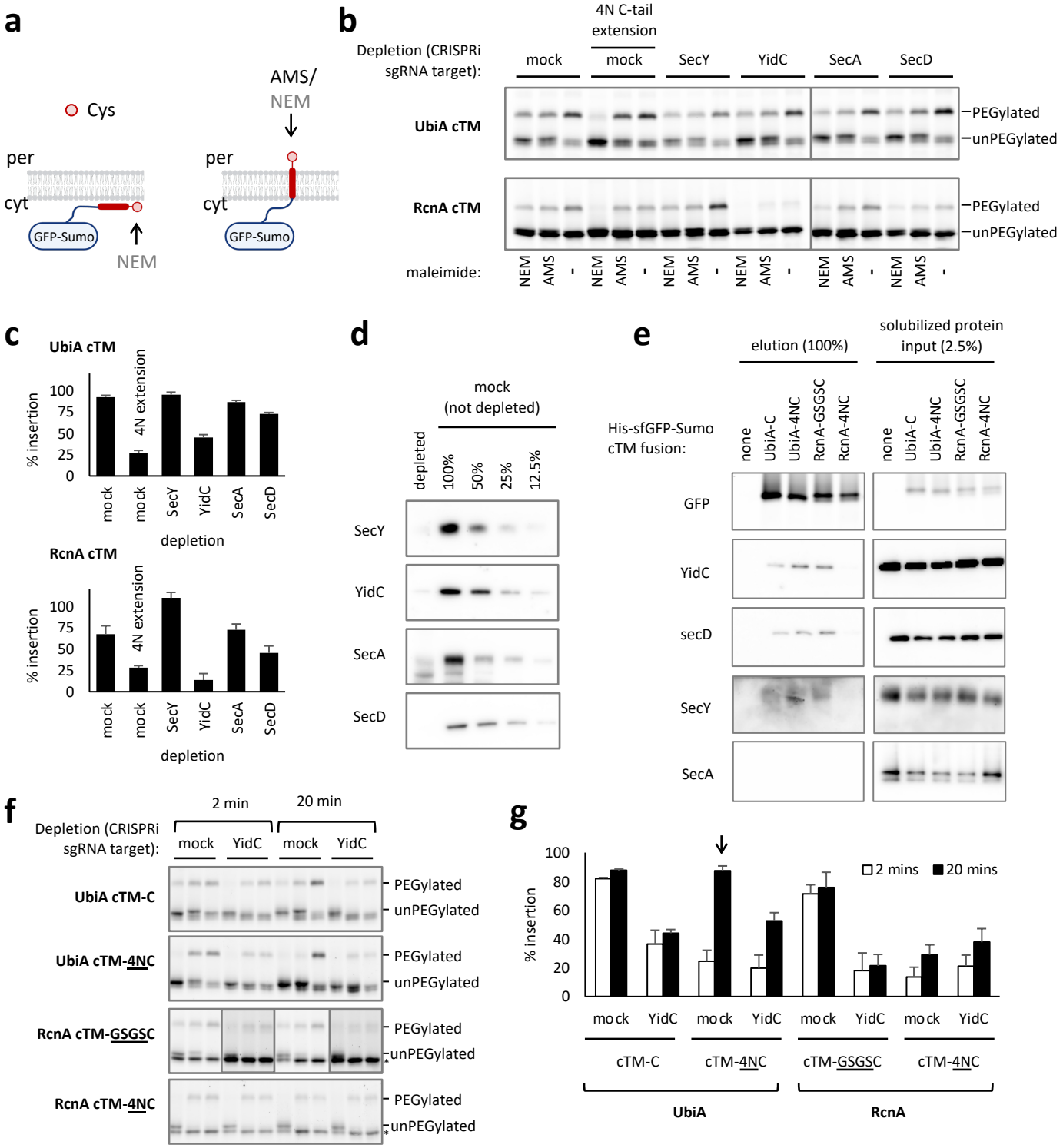
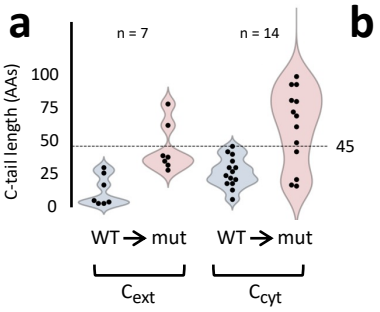


Fig. 3. *E. coli* YidC is required for cTM insertion and interacts with cTMs.
(a) Illustration of the simplified assay for cTM insertion. The cTMs of UbiA or RcnA, along with their C-tails and a C-terminal Cys, were appended to N-terminal His_6-GFP-Sumo. If inserted, the C-terminal Cys will be accessible to both NEM and extracellular AMS, whereas an uninserted species will only be accessible to NEM.
(b) YidC depletion inhibits insertion of UbiA and RcnA cTMs. The *in vivo* insertion of His_6-GFP-Sumo-cTM constructs was assessed by their reaction with AMS and NEM, which blocks PEGylation. CRISPRi was used to deplete the indicated factor using sgRNAs inhibiting their genes. An inert sgRNA was used as control (mock).
(c) Quantification of the level of insertion shows that a 4N extention, YidC depletion, and to a lesser extent SecD depletion impair cTM insertion. Three biological replicates of the experiments shown in (b) were quantified. Shown are means ± SEM.
(d) Quantitative Western blotting showing that the depletion of various factors reduces their level to below 12%. Total proteins from the depleted cells were run alongside a serial 2-fold dilution of proteins from undepleted cells and detected with an antibody against the depleted factors.
(e) YidC, SecD, and SecY copurify with the chimeric cTM constructs. Synthetic nanodiscs were used to solubilize membranes from whole cells expressing the indicated His_6-GFP-Sumo-cTM constructs. Subsequently, the cTM chimeric proteins were purified on Co^{2+} chromatography. The eluted fractions, as well as diluted total protein input samples, were detected for various proteins using the indicated antibodies by Western blotting, or GFP fluorescence.
(f) Insertion of UbiA and RcnA Sumo-GFP-cTM variants at 2 or 20 min, and its inhibition by YidC depletion (YidC) or no depletion (Mock). The experiments were performed as in (b), but allowing more labeling time with AMS for the 20 min samples. (*) Marks a cleaved GFP-SUMO that does not contain the C-terminal Cys and was not quantified.
(g) Quantification of the level of insertion from three biological repeats of the representative experiment shown in (f). The UbiA-cTM-4NC is inserted over time, in a manner that can be inhibited by YidC depletion.



b

Protein	Phenotype	Wild type / mutant seq	Mut C-tail ΔG
SGLT1	Congenital glucose-galactose malabsorption	FARVLPFAVDLPWLDSYSPSFSLVLWC	8.93
ENT3	H syndrome	VHLISKGGHKDIGASEPLKMRRECRRAGGHGGKA	26.29
ATP6V0C	Seizure	TKLTLSEPTSHRILCKDHPSSFQNEQPDTYARGRRPQ	32.84
CHRNE	Congenital myasthenic syndrome	YFNRVPDLPYAPCIQPQLAPTSISHPSPVGN	19.58
CHRND	Lethal multiple pterygium syndrome	VYNQPPPQPFPGDPYSYNVQDKRFIYGGPVGEPGDSRV	31.61

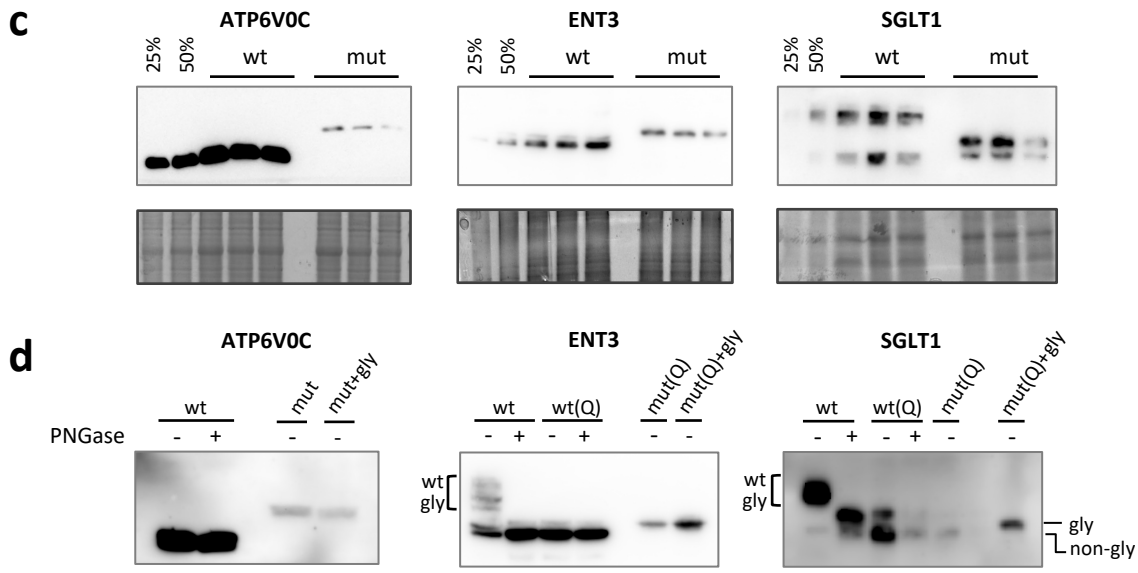


Fig. 4. Disease-causing mutations in the C-tails of human MPs impair insertion.
(a) Clinvar-annotated nonstop mutations in posttranslationally-inserting MPs. Pathogenic Mutations were identified in 7 C_{ext} and 14 C_{cyt} multispanning MPs. The length of the C-tail before and after the mutation is shown. Most pathogenic mutations in C_{ext} MPs maintain a C-tail shorter than 45 residues, suggesting that their cTMs would still rely on posttranslational insertion.
(b) C-tail sequences of 5 C_{ext} proteins before and after the pathogenic nonstop mutation.
(c) Relative expression of ATP6V0C, ENT3 , SGLT1 and their nonstop mutants. Lysate from triplicate transient transfections of HEK293, transfected with either the wildtype (wt) or mutant (mut) proteins, were analyzed by immunoblotting with anti-HA antibodies. Equal amounts of total proteins were loaded and Coomassie staining of the gels is shown as a loading control. Serial dilutions of wt sample were used to estimate relative expression. The ATP6V0C mutant displays reduced expression, while the ENT3 and SGLT1 mutants are comparable to wild-type. Representative blots of 3-4 biological repeats are shown.
(d) Glycosylation analysis demonstrates that the C-tails of ATP6V0C and ENT3 mutants are not translocated to the ER lumen. The endogenous glycosylation site of ENT3 and SGLT1 was eliminated by mutation (Q) and an NFT glycosylation acceptor site was added to the C-tail of the nonstop mutants of ATP6V0C, ENT3 and SGLT1 (mut(Q)+gly). Treatment of the lysate from HEK293 cells transfected with wildtype (wt) or glycosylation mutants (wt(Q)) with PNGase F verified the elimination of endogenous glycosylation for ENT3 and SGLT1 (elimination of wt glyc bands). Addition of the NFT site to either ATP6V0C or ENT3 nonstop mutants did not affect the apparent size of the protein (compare mut(Q)+gly to mut(Q), left and middle panels), suggesting that their C-tails are not translocated to the ER lumen. In contrast, addition of the NFT site to the SGLT1 mutant resulted in an increase in its apparent size, suggesting that it is efficiently translocated and glycosylated (right panel).

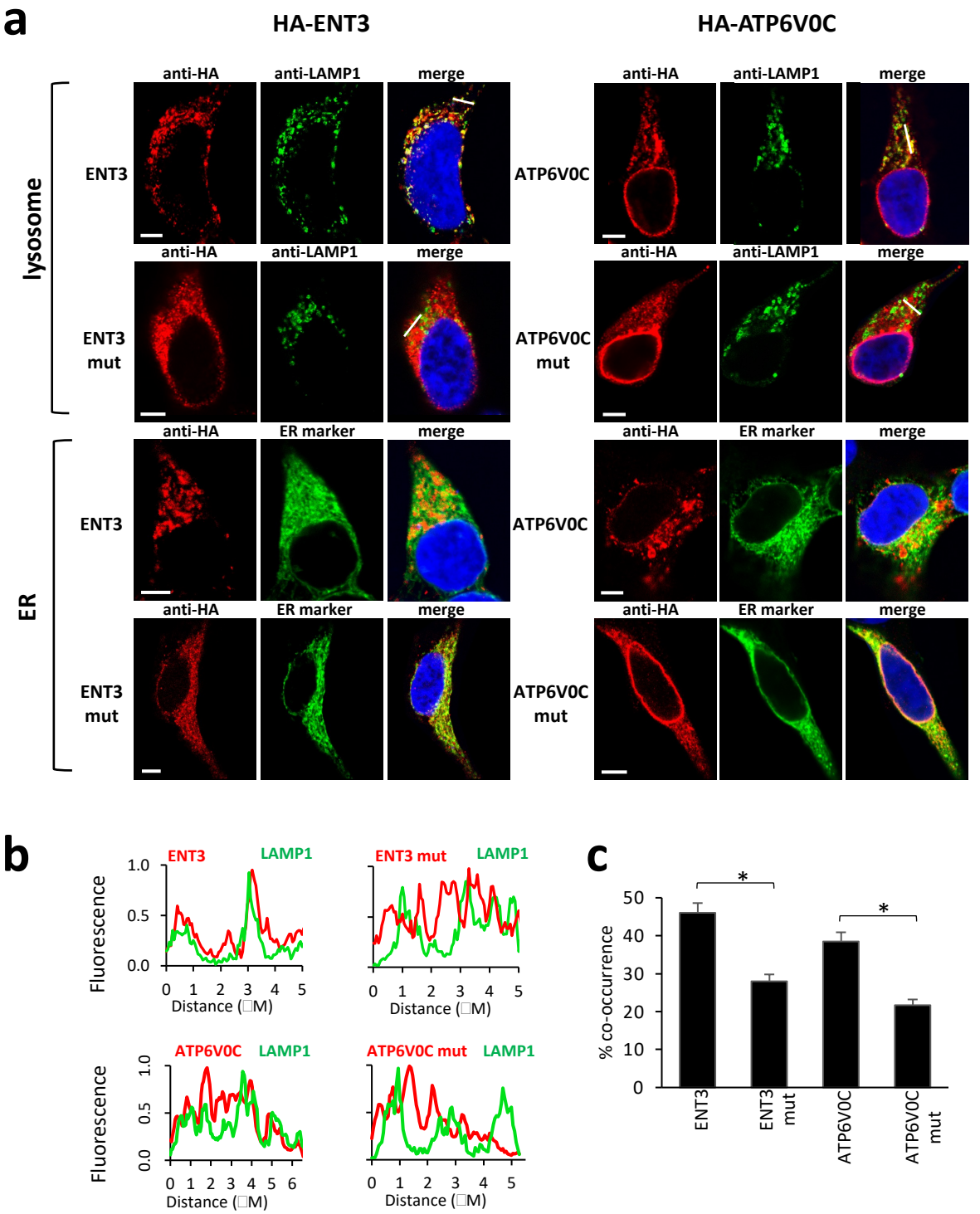


Fig. 5. Disease-causing mutations in the C-tails of human MPs display a trafficking defect.
(a) Immunofluorescence confocal microscopy demonstrates reduced localization of ENT3 and ATP6V0C C-tail mutants to lysosomes. HEK293 cells were transfected with ENT3 and ATP6V0C wildtype or mutant proteins and treated with primary antibodies against HA to label the protein, or LAMP1 to label the lysosome as indicated. A vector expressing ER-mScarletI was used to label the ER (the mScarletI visualization was digitally swapped to red for ease of viewing). Colocalization of the proteins with the organellar markers results in a yellow signal in the merge panel. DAPY nuclear staining is shown in blue. Scale bar, 5 □m.
(b) Intensity profiles of the fluorescence signals along the white lines indicated in the merge images of the proteins with LAMP1 lysosome marker in (a). Fluorescence intensity on the Y axis was normalized to the maximal intensity along the line. Red, ENT3 or ATP6V0C immunofluorescence signal; green, LAMP1 immunofluorescence signal. The signal of the mutants is not correlated with the lysosomal LAMP1.
(c) Quantification of the co-occurrence of the different proteins with lysosomal LAMP1. Bars show averages ± SEM of 36-45 cells. * indicate $P < 0.01$, Student's t-test.

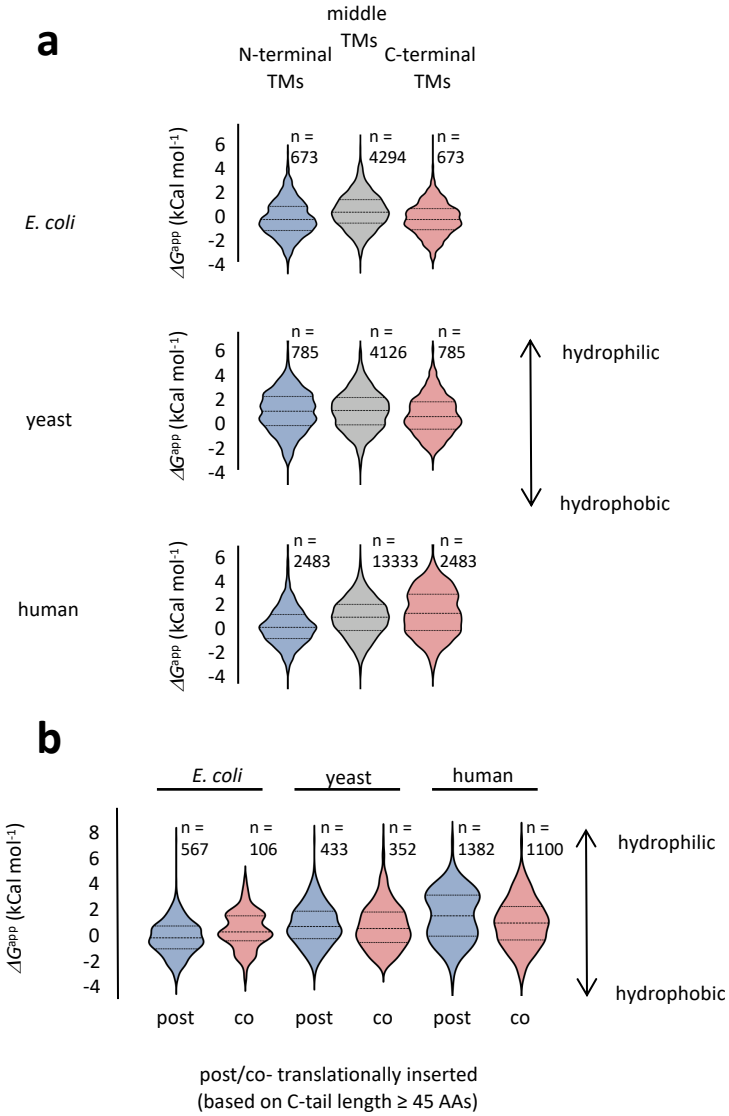


Fig. S1. Hydrophobicities of TMs placed in different positions in multispanning proteins.
The hydrophobicities of all TMs were calculated according to the biological hydrophobicity scale (Hessa et al., 2007), which predicts the ΔG^{app} for membrane insertion. Lower values indicate higher hydrophobicity.
(a) Hydrophobicities of TMs placed at the terminal positions, either N- or C-terminal, compared to middle TMs. The cTMs display similar hydrophobicities.
(b) The Hydrophobicities of cTMs that precede a C-tail of ≥ 45 amino acids, which can theoretically get inserted cotranslationally, are similar to cTMs with short C-tails that necessitate posttranslational insertion.

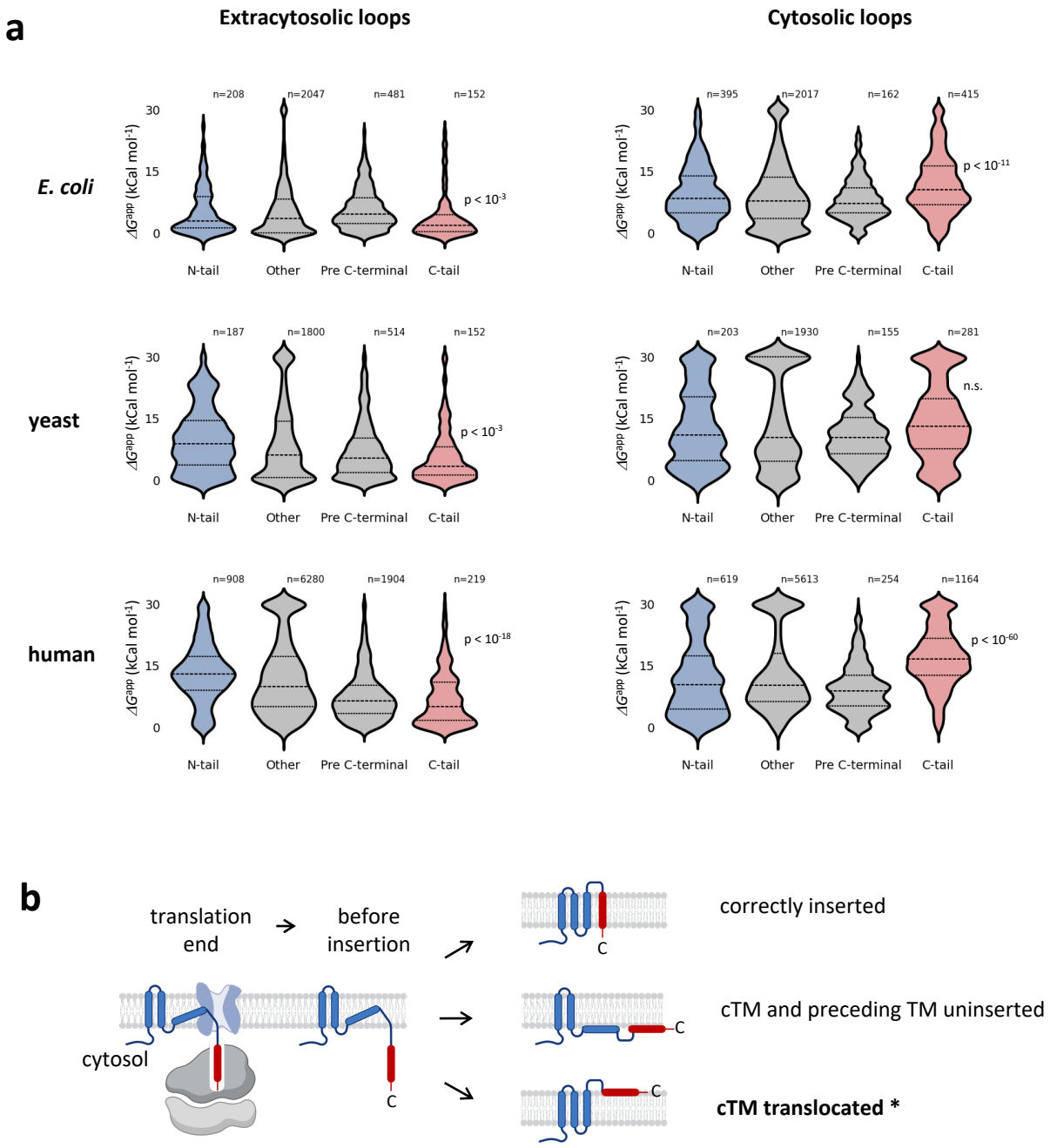


Fig. S2. Hydrophilicities of loops and tails in different positions along the protein and different membrane sides. (a) C-tails possess significantly different hydrophilicities compared to other loops or tails found on the same side of the membrane. 'Pre C-terminal' refers to the loop preceding the cTM. The indicated P-values are for comparison between the C-tail and 'Other' loops, Mann-Whitney test. Only loops/tails shorter than 45 residues were taken for analysis. **(b)** Possible topologies after correct or erroneous posttranslational cTM insertion in C_{cyt} proteins. * marks a potential insertion error, in which the cTM, along with the C-tail, is translocated to the extracytosolic side. This error might be alleviated by a hydrophilic C-tail anchoring the C-tail to the cytosol.

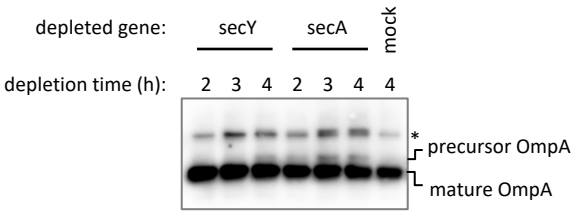


Fig. S3. Four hours of CRISPRi depletion is sufficient to inactivate SecY and SecA. *E. coli* cells were grown under CRISPRi depletion conditions that inhibit the transcription of the indicated genes for 2-4 hours, and the OmpA protein was detected by Western blotting. The inactivation of SecY and SecA after four hours of depletion, which is the duration of depletion in our experiments, is indicated by the appearance of an OmpA precursor due to secretion impairment. Note that since OmpA is a stable protein, most of the OmpA in these samples has been synthesized and matured before the depletion, and hence the mature form is dominant.

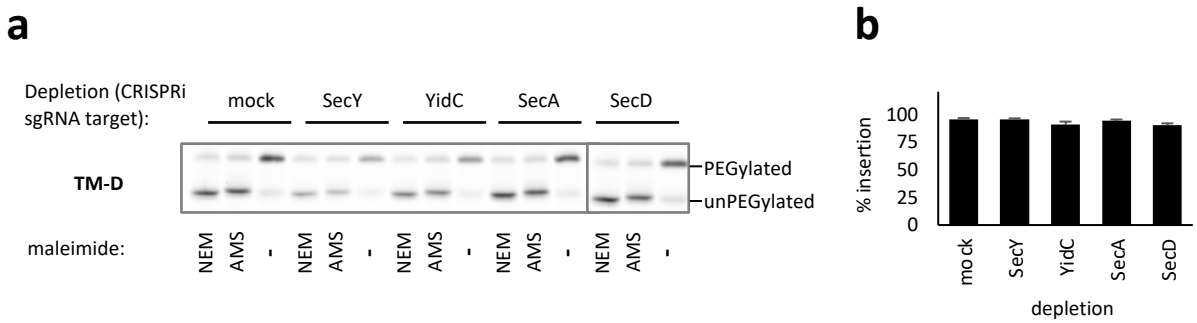


Fig. S4. Depletion of Sec components or YidC does not impair the insertion of TM-D. Panels a and b are similar to Fig. 3 b and c, respectively, except that the chimeric His_6-GFP-Sumo-TM construct harbored the TM-D transmembrane helix.

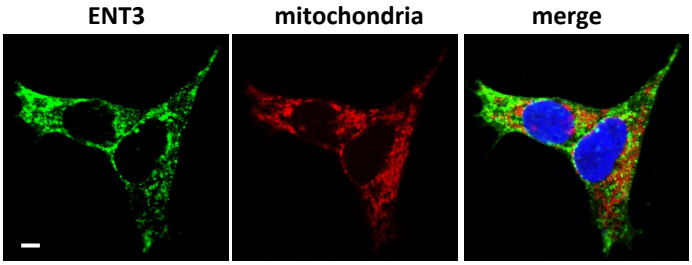


Fig. S5. ENT3 is not localized to mitochondria in HEK293 cells.
HEK293 cells transfected with HA-ENT3 were treated with MitoTracker® Deep Red FM and then immunostained with anti-HA antibodies. DAPY nuclear staining is shown in blue. Scale bar, 5 □m.

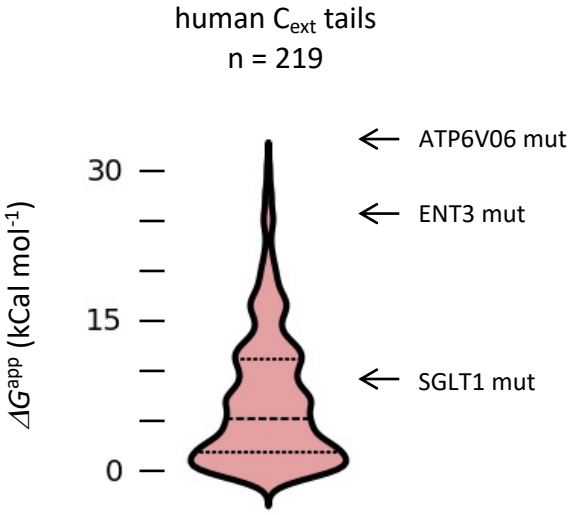


Fig. S6. The distribution of C-tail hydrophilicities in natural human proteins compared to the nonstop mutants of SGLT1, ENT3 and ATP6V06